

長庚大學 / 暑期 / 大數據應用

生物大數據：Protein & AlphaFold

劉 益 忠

清華大學生物資訊與結構生物研究所/國衛院群健所
博士後研究員

2022/07/19

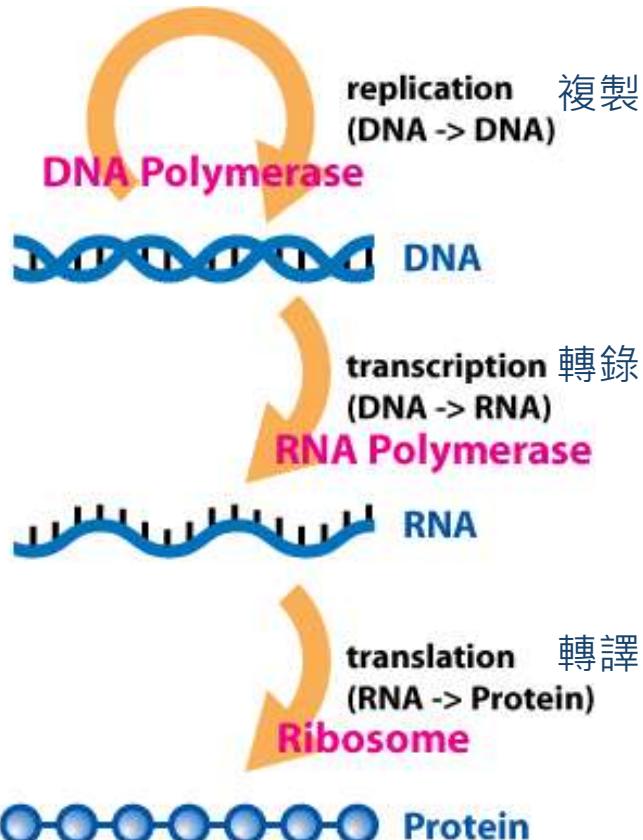
Python vs Protein

- ❖ 蛋白質的四級結構
- ❖ 蛋白質資料庫
Protein Data Bank; PDB file
- ❖ 蛋白質結構視覺化及PyMOL軟體
- ❖ 蛋白質結構模擬與預測
- ❖ AlphaFold & ColabFold

蛋白質的四級結構

Start from Sequence

遺傳訊息的傳遞



遺傳訊息傳遞

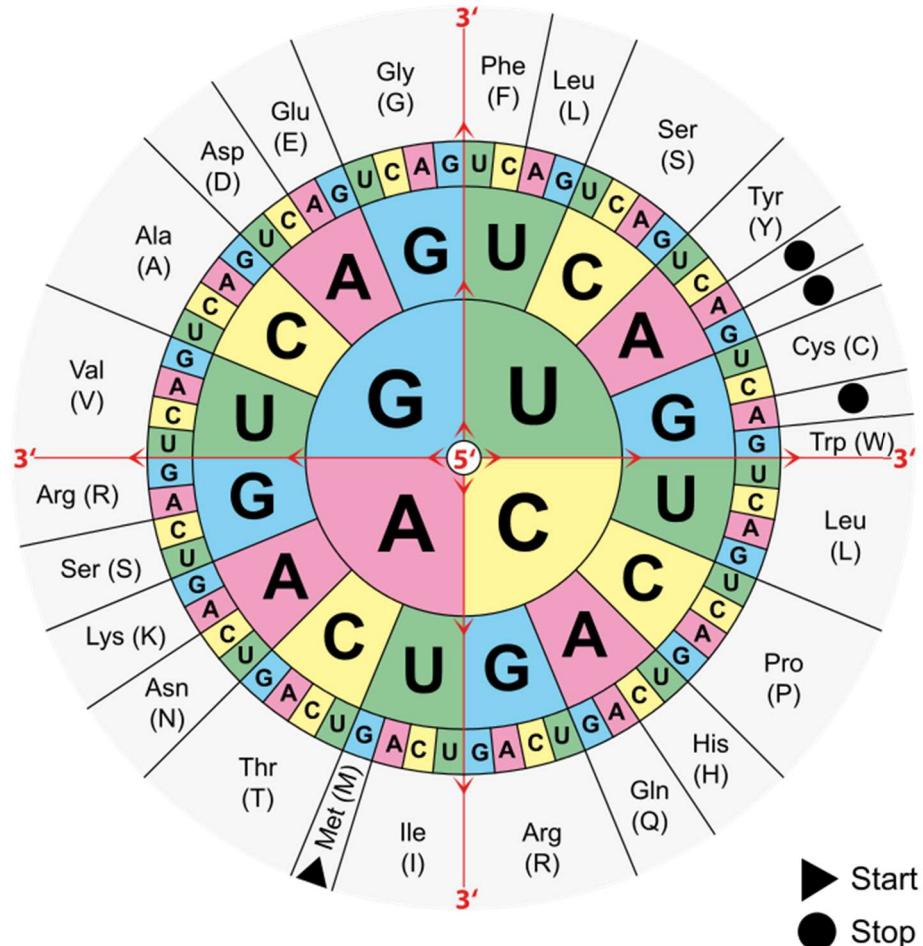
DNA → RNA → Protein

- ❖ **轉錄(Transcription)**是遺傳訊息由DNA轉換到RNA的過程。
- ❖ **轉譯(Translation)**是根據遺傳密碼的中心法則，將成熟的mRNA分子（由DNA透過轉錄而生成）中「鹼基的排列順序」（核苷酸序列）解碼，並生成對應的特定胺基酸序列的過程。

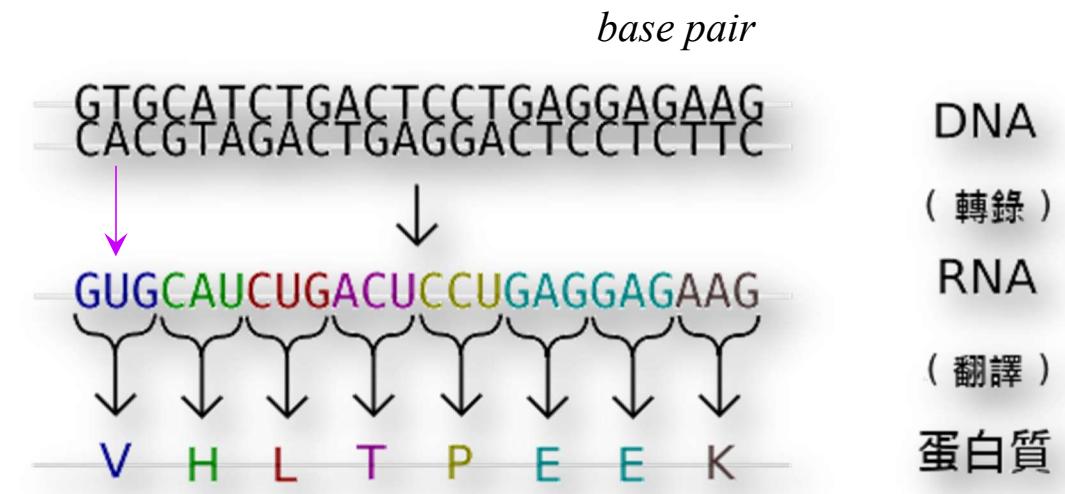
中心法則 (The central dogma)

(https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology)

Start from Sequence



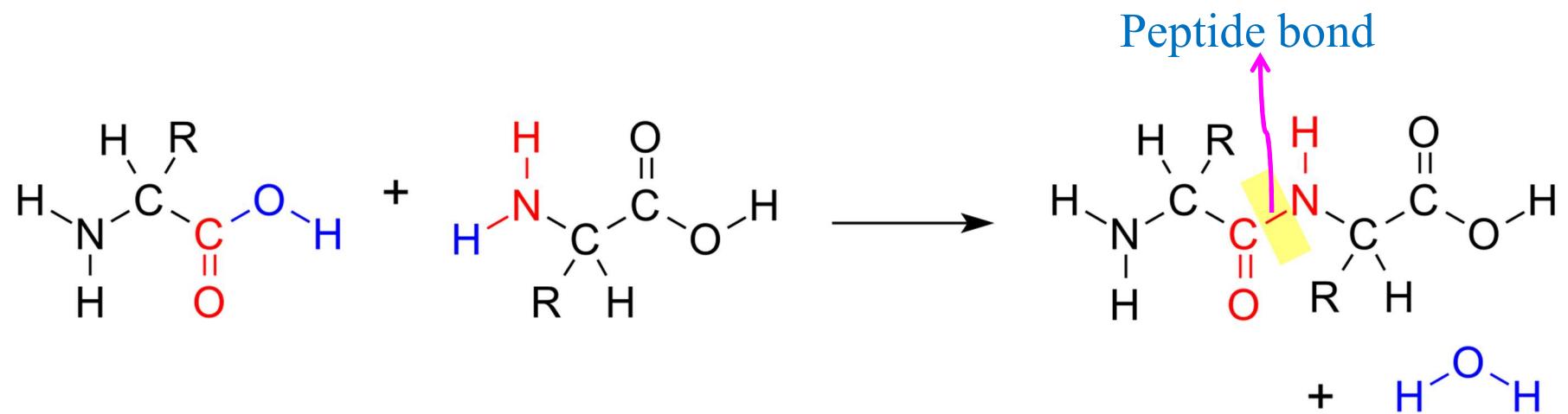
由 Mouagip - Codons aminoacids table.png, 公有領域,
<https://commons.wikimedia.org/w/index.php?curid=5986132>



(編碼一個蛋白質胺基酸序列)

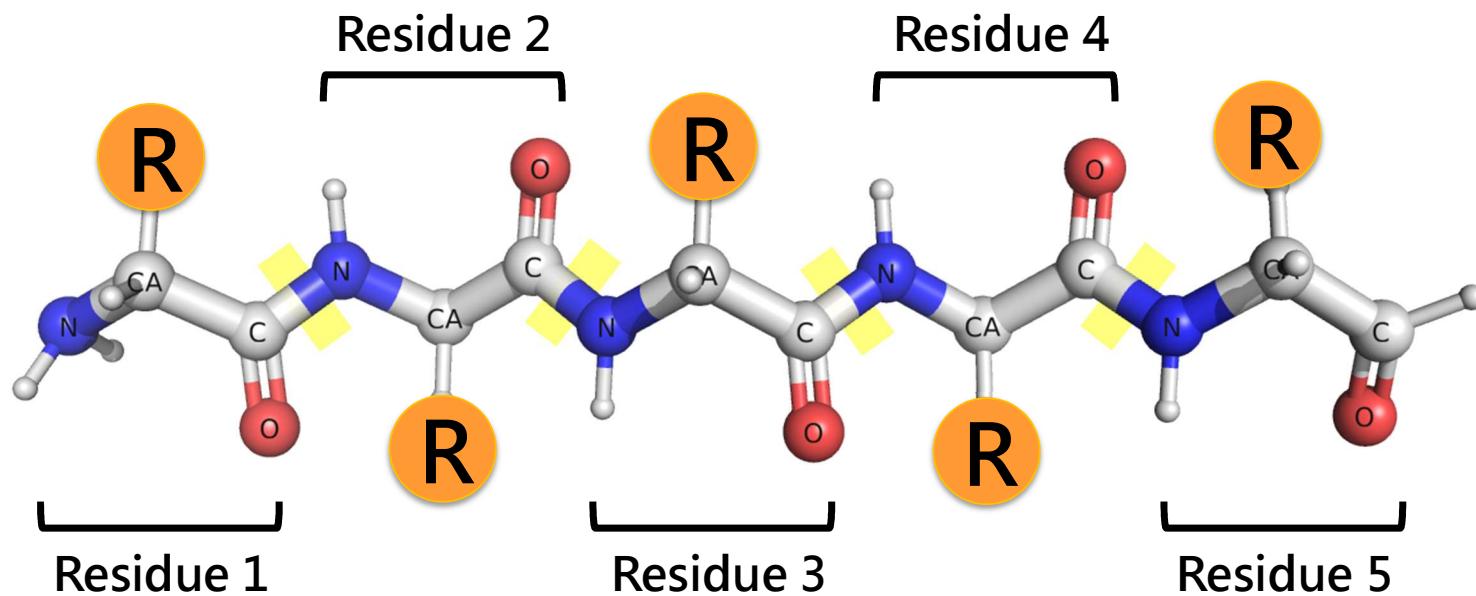
Madprimederivative work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=18365188>

Amino acids are linked via a condensation reaction

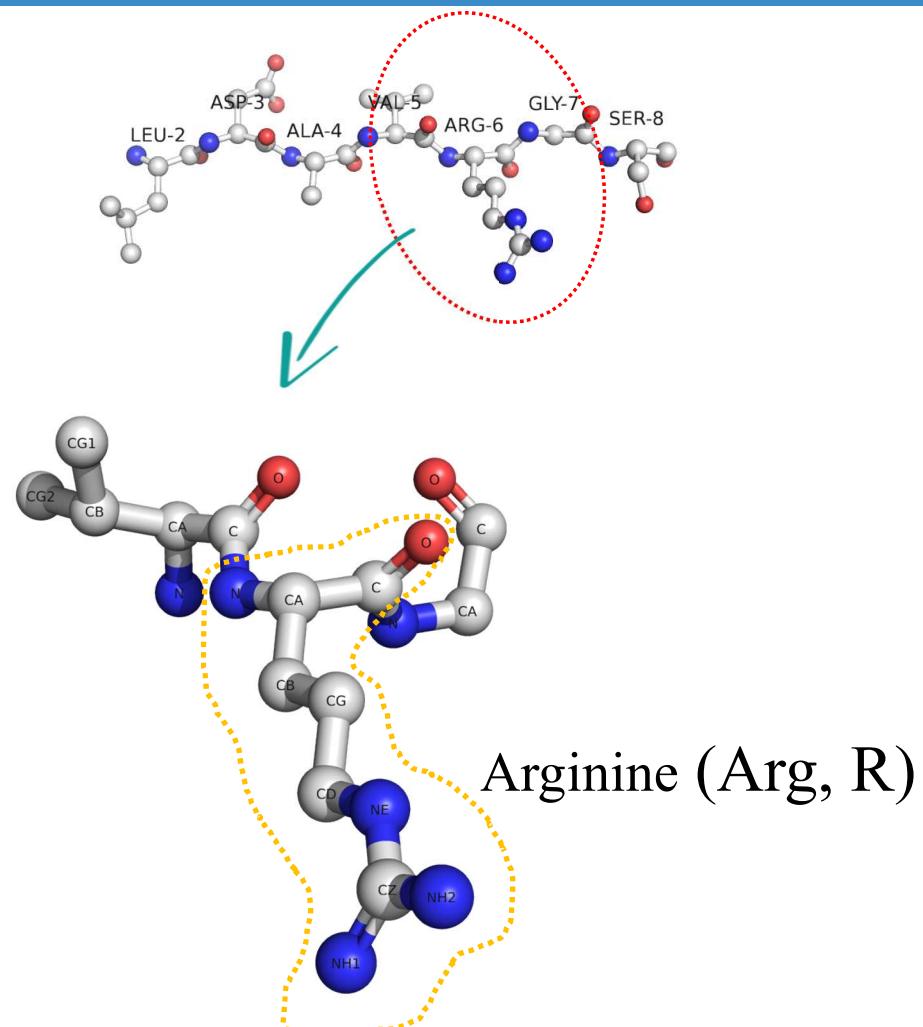
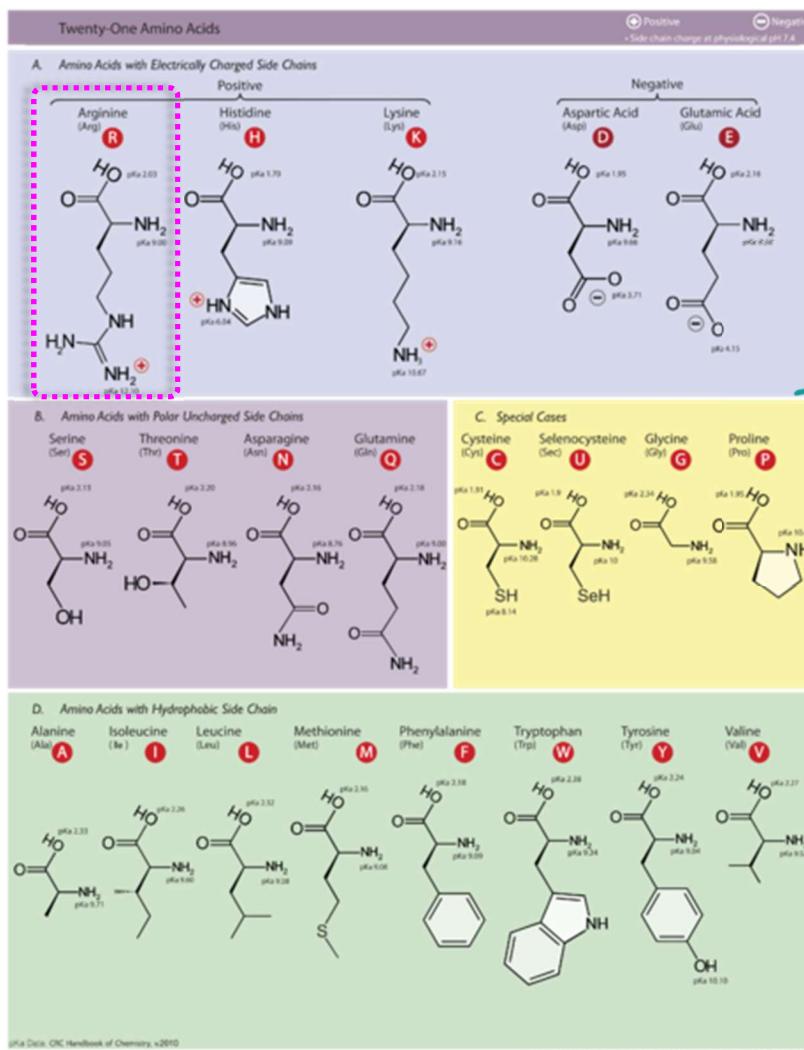


個別胺基酸以肽鍵(Peptide bond)結合

Amino acids are linked by peptide bonds

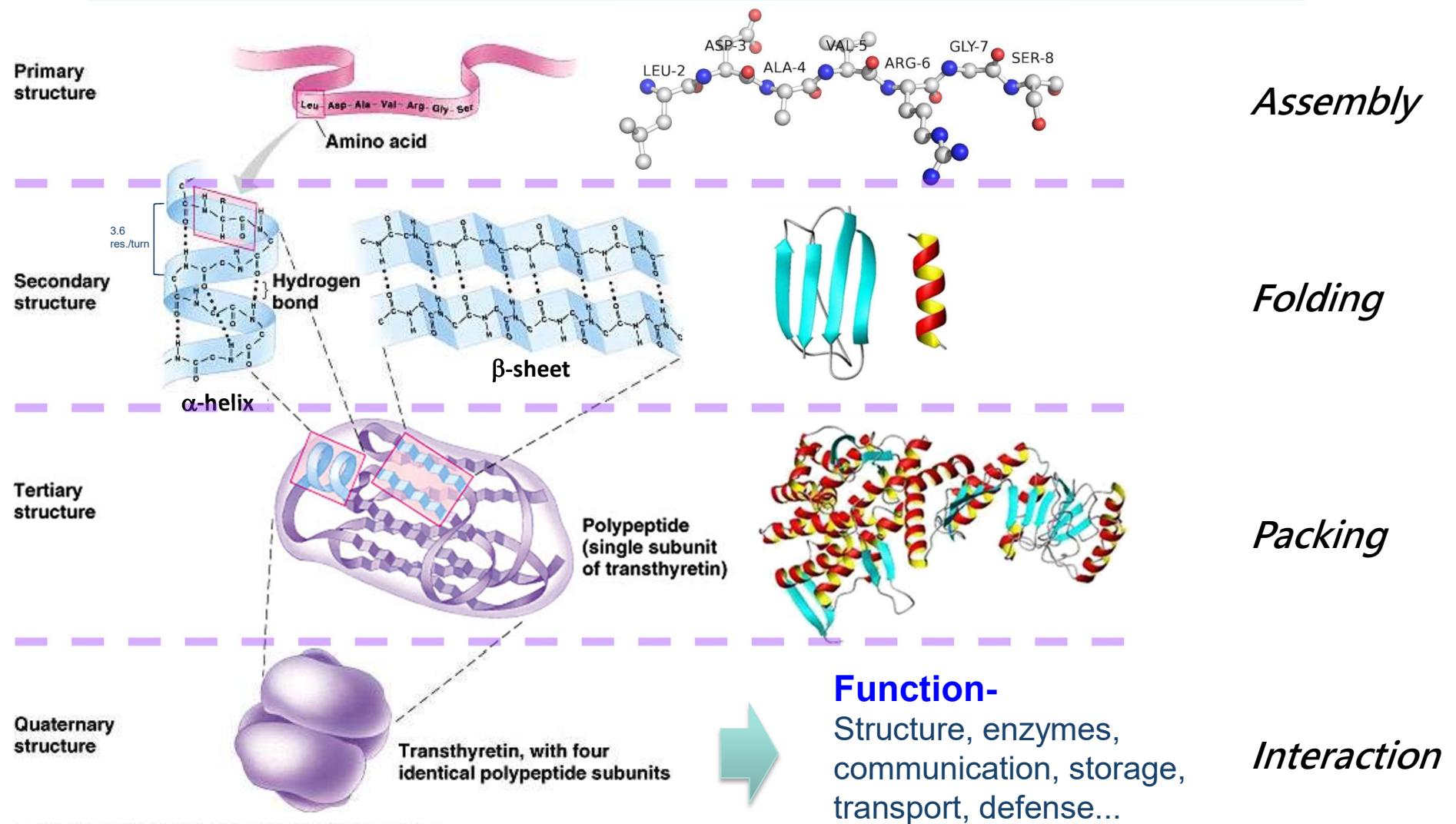


List of amino acid structures

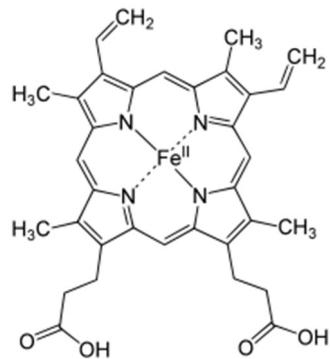


Dan Cojocari, CC BY-SA 3.0, via Wikimedia Commons

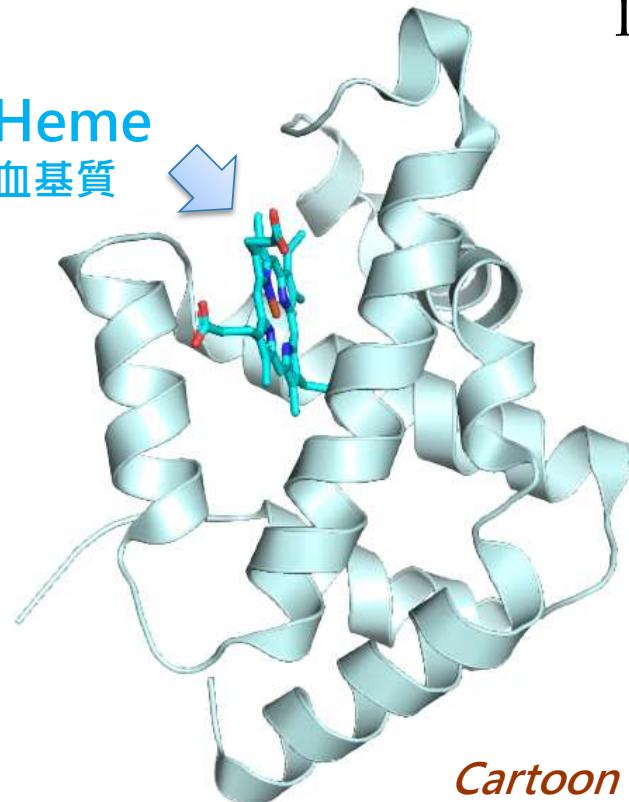
The four levels of protein structure



Hemoglobin alpha chain | *Homo sapiens*

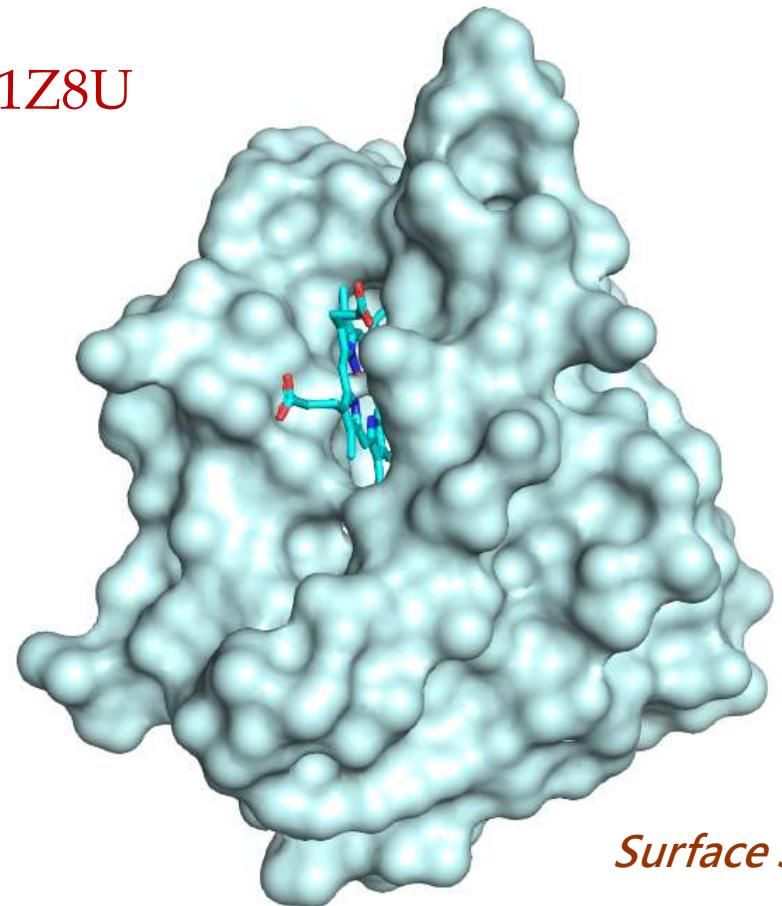


Heme
血基質



Cartoon 3D style

PDB ID: 1Z8U



Surface 3D style

>1Z8U_2|Chains B,D|Hemoglobin (血紅素) alpha chain|*Homo sapiens*

About Protein Structure

蛋白質結構簡介

蛋白質結構簡介

目前主要解析蛋白質結構的實驗方法

Molecular Type	X-ray ↓	NMR ↑	EM ↓↑	Multiple methods ↑↓	Neutron ↑↓	Other ↓↑	Total ↓↑
Protein (only)	147998	11984	7772	187	72	32	168045
Protein/Oligosaccharide	8724	31	1372	5	0	0	10132
Protein/NA	7812	277	2461	3	0	0	10553
Nucleic acid (only)	2453	1416	62	12	2	1	3946
Other	154	31	5	0	0	0	190
Oligosaccharide (only)	11	6	0	1	0	4	22
Total	167152	13745	11672	208	74	37	192888

X-ray: X-ray crystallography X-光晶體學

NMR: Nuclear magnetic resonance 核磁共振

cryo-EM: cryo-electron microscopy 冷凍電顯

(20220716 updated)

<https://www.rcsb.org/stats/summary>

Convert a PDB file to a Structure

>6ANE_1|Chains A, B, C|Poly(ethylene terephthalate) hydrolase

MQTNPYARGPNPTAASLEASAGPFTVRSFTSRPSGYGAGTVYYPTNAGGTVGAIIAIVPGYTARQSSIK
WWGPRILASHGFVVITIDTNSTLDQPSSRSSQQMAALRQVASLNGTSSSPIYGKVDTARMGVGMWSMGGG
GSLISAANNPSLKAAPQAPWDSSTNFSSVTVP TLIFACENDSIAPVNSSALPIYDSMSRNAKQFLEIN
GGSHSCANSGNNSQALIGKKGVAWMKRFMDNDTRYSTFACEPNSTRVSDFRTANCSLEHHHHHH

FASTA
format

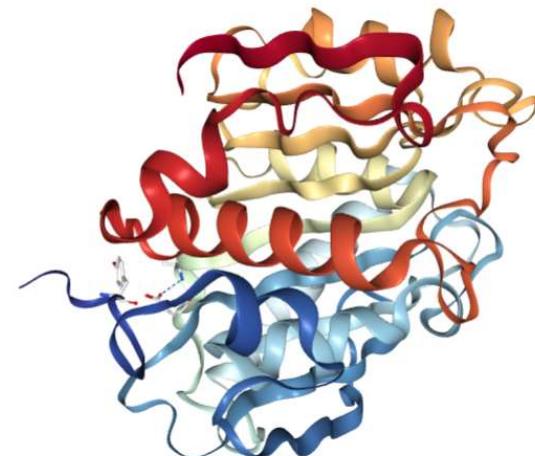
Sequence

PDB structure file

記錄每個原子的3D座標

ORIGX2	0.000000	1.000000	0.000000	0.000000		
ORIGX3	0.000000	0.000000	1.000000	0.000000		
SCALE1	0.018915	0.000000	0.000000	0.000000		
SCALE2	0.000000	0.004276	0.000000	0.000000		
SCALE3	0.000000	0.000000	0.006658	0.000000		
ATOM	1	N	THR	A	2	-38.844 -36.139 26.075 1.00 43.20 N
ATOM	2	CA	THR	A	2	-38.028 -36.236 24.823 1.00 42.15 C
ATOM	3	C	THR	A	2	-37.851 -34.874 24.086 1.00 38.29 C
ATOM	4	O	THR	A	2	-38.695 -34.429 23.294 1.00 39.45 O
ATOM	5	CB	THR	A	2	-38.559 -37.317 23.881 1.00 44.81 C
ATOM	6	O61	THR	A	2	-37.647 -37.461 22.783 1.00 45.87 O
ATOM	7	CG2	THR	A	2	-39.990 -37.009 23.388 1.00 45.59 C
ATOM	8	N	ASN	A	3	-36.718 -34.247 24.370 1.00 31.48 N
ATOM	9	CA	ASN	A	3	-36.367 -32.903 23.903 1.00 26.55 C
ATOM	10	C	ASN	A	3	-36.100 -32.887 22.404 1.00 24.30 C
ATOM	11	O	ASN	A	3	-35.074 -33.415 21.952 1.00 24.90 O
ATOM	12	CB	ASN	A	3	-35.117 -32.423 24.672 1.00 23.36 C
ATOM	13	CG	ASN	A	3	-34.672 -31.001 24.298 1.00 22.22 C
ATOM	14	OD1	ASN	A	3	-35.184 -30.378 23.371 1.00 19.60 O
ATOM	15	ND2	ASN	A	3	-33.706 -30.494 25.041 1.00 22.68 N
ATOM	16	N	PRO	A	4	-36.950 -32.197 21.628 1.00 23.95 N
ATOM	17	CA	PRO	A	4	-36.754 -32.262 20.171 1.00 23.73 C
ATOM	18	C	PRO	A	4	-35.513 -31.536 19.636 1.00 23.47 C
ATOM	19	O	PRO	A	4	-35.170 -31.710 18.468 1.00 22.30 O
ATOM	20	CB	PRO	A	4	-38.030 -31.646 19.613 1.00 24.93 C
ATOM	21	CG	PRO	A	4	-38.510 -30.712 20.682 1.00 26.00 C
ATOM	22	CD	PRO	A	4	-38.061 -31.301 21.999 1.00 25.07 C
ATOM	23	N	TYR	A	5	-34.862 -30.723 20.469 1.00 23.26 N

軟體轉換成
立體結構



Visualized by
Molecular Graphical Tools

Demo: 在 Protein Data Bank, PDB 搜尋 6ANE

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation MyPDB

PDB PROTEIN DATA BANK 蛋白質結構資料庫

179842 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Enter search term(s) **6ANE**

Advanced Search | Browse Annotations Help

Celebrating 50 YEARS OF Protein Data Bank

Welcome Deposit Search Visualize Analyze Download Learn

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.

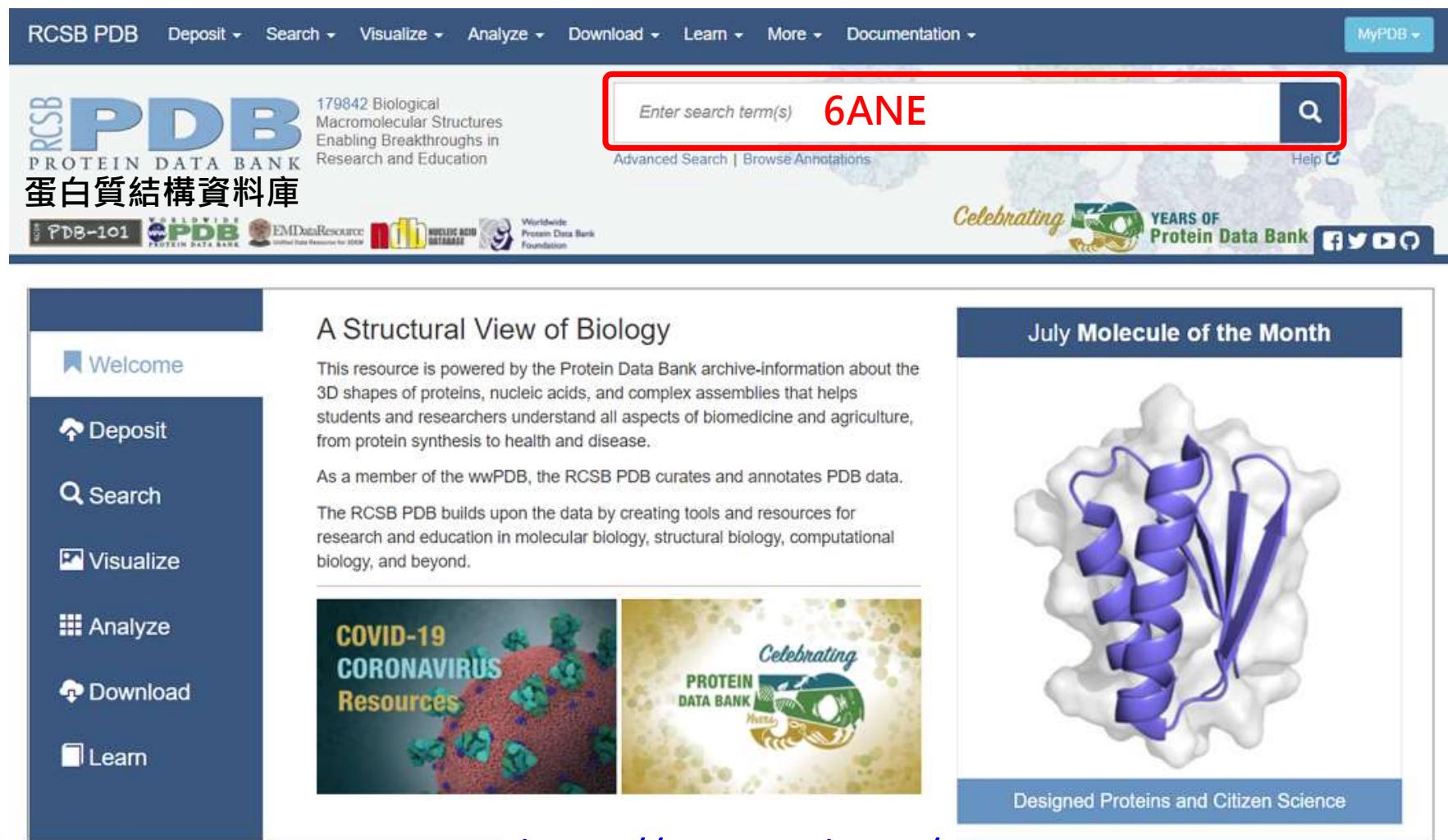
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

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.

COVID-19 CORONAVIRUS Resources

July Molecule of the Month

Designed Proteins and Citizen Science



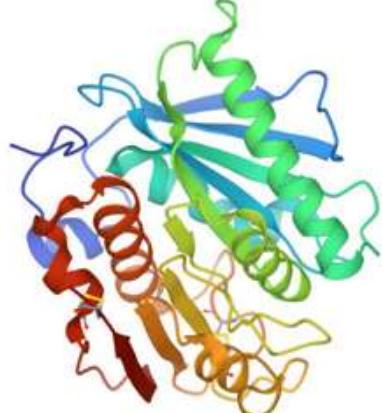
<https://www.rcsb.org/>

Structure Summary : 6ANE

6ANE: Crystal Structure of Ideonella sakaiensis(大阪堺菌) PET Hydrolase(聚對苯二甲酸乙二酯-水解酶)

Structure Summary 3D View Annotations Experiment Sequence Genome Versions 6ANE.pdb

Biological Assembly 1 ?



6ANE PDB ID

Crystal Structure of Ideonella sakaiensis PET Hydrolase

PDB DOI: 10.2210/pdb6ANE/pdb

Classification: HYDROLASE

Organism(s): Ideonella sakaiensis

Expression System: Escherichia coli BL21

Mutation(s): No

Deposited: 2017-08-12 Released: 2018-04-18

Deposition Author(s): Galaz-Davison, P., Sotomayor, M., P.

Funding Organization(s): ASBMB, INACH

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.02 Å

R-Value Free: 0.217

R-Value Work: 0.177

R-Value Observed: 0.179

Display Files Download Files

FASTA Sequence

PDB Format **PDB Format (gz)**

PDBx/mmCIF Format

PDBx/mmCIF Format (gz)

PDBML/XML Format (gz)

wwPDB

Biological Assembly 1

Structure Factors (CIF)

Structure Factors (CIF - gz)

RSRZ outliers

Value

0.223

1

0

0

1.0%

Worse

Percentile relative to all X-ray structures

Percentile relative to X-ray structures of similar resolution

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A1

<https://www.rcsb.org/structure/6ane>

PDB檔案內容



線上顯示PDB檔案內容，
前面部分是對結構的註解

HEADER HYDROLASE 12-AUG-17 6ANE
TITLE CRYSTAL STRUCTURE OF IDEONELLA SAKAIENSIS PET HYDROLASE
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: POLY(ETHYLENE TEREPHTHALATE) HYDROLASE;
COMPND 3 CHAIN: A, B, C;
COMPND 4 FRAGMENT: RESIDUES 28-290;
COMPND 5 SYNONYM: PETASE;
COMPND 6 EC: 3.1.1.101;
COMPND 7 ENGINEERED: YES
SOURCE MOL_ID: 1;
SOURCE 2 ORGANISM_SCIENTIFIC: IDEONELLA SAKAIENSIS (STRAIN 201-F6);
SOURCE 3 ORGANISM_TAXID: 1547922;
SOURCE 4 STRAIN: 201-F6;
SOURCE 5 GENE: ISF6_4831;
SOURCE 6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE 7 EXPRESSION_SYSTEM_TAXID: 511693;
SOURCE 8 EXPRESSION_SYSTEM_STRAIN: BL21;
SOURCE 9 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE 10 EXPRESSION_SYSTEM_PLASMID: PET24B
KEYWDS PETASE, AB HYDROLASE, CUTINASE, PLASTIC, HYDROLASE
EXPDTA X-RAY DIFFRACTION
AUTHOR P.GALAZ-DAVISON,M.SOTOMAYOR,L.P.PARRA,C.A.RAMIREZ-SARMIENTO
REVDAT 2 25-APR-18 6ANE 1 REMARK
REVDAT 1 18-APR-18 6ANE 0
JRNL AUTH T.FECKER,P.GALAZ-DAVISON,F.ENGELBERGER,Y.NARUI,M.SOTOMAYOR,
JRNL AUTH 2 L.P.PARRA,C.A.RAMIREZ-SARMIENTO
JRNL TITL ACTIVE SITE FLEXIBILITY AS A HALLMARK FOR EFFICIENT PET
JRNL TITL 2 DEGRADATION BY I. SAKAIENSIS PETASE.
JRNL REF BIOPHYS. J. V. 114 1302 2018

<http://www.wwpdb.org/documentation/file-format-content/format33/sect2.html>

PDB檔案內容- ATOM



*Text-based format
with 3D information*

CRYST1	52.868	233.852	165.076	90.00	90.00	90.00	C	2	2	21	24
ORIGX1	1.000000	0.000000	0.000000								
ORIGX2	0.000000	1.000000	0.000000								
ORIGX3	0.000000	0.000000	1.000000								
SCALE1	0.018915	0.000000	0.000000								
SCALE2	0.000000	0.004276	0.000000								
SCALE3	0.000000	0.000000	0.006058	x	y	z					
ATOM	1	N	THR	A	2	-38.844	-36.139	26.075	1.00	43.20	N
ATOM	2	CA	THR	A	2	-38.028	-36.236	24.823	1.00	42.15	C
ATOM	3	C	THR	A	2	-37.851	-34.874	24.086	1.00	38.29	C
ATOM	4	O	THR	A	2	-38.695	-34.429	23.294	1.00	39.45	O
ATOM	5	CB	THR	A	2	-38.559	-37.317	23.881	1.00	44.81	C
ATOM	6	OG1	THR	A	2	-37.647	-37.461	22.783	1.00	45.87	O
ATOM	7	CG2	THR	A	2	-39.990	-37.009	23.388	1.00	45.59	C
ATOM	8	N	ASN	A	3	-36.718	-34.247	24.370	1.00	31.48	N
ATOM	9	CA	ASN	A	3	-36.367	-32.903	23.903	1.00	26.55	C
ATOM	10	C	ASN	A	3	-36.100	-32.887	22.404	1.00	24.30	C
ATOM	11	O	ASN	A	3	-35.074	-33.415	21.952	1.00	24.90	O
ATOM	12	CB	ASN	A	3	-35.117	-32.423	24.672	1.00	23.36	C
ATOM	13	CG	ASN	A	3	-34.672	-31.001	24.298	1.00	22.22	C
ATOM	14	OD1	ASN	A	3	-35.184	-30.378	23.371	1.00	19.60	O
ATOM	15	ND2	ASN	A	3	-33.706	-30.494	25.041	1.00	22.68	N
ATOM	16	N	PRO	A	4	-36.950	-32.197	21.628	1.00	23.95	N
ATOM	17	CA	PRO	A	4	-36.754	-32.262	20.171	1.00	23.73	C
ATOM	18	C	PRO	A	4	-35.513	-31.536	19.636	1.00	23.47	C
ATOM	19	O	PRO	A	4	-35.170	-31.710	18.468	1.00	22.30	O
ATOM	20	CB	PRO	A	4	-38.030	-31.646	19.613	1.00	24.93	C
ATOM	21	CG	PRO	A	4	-38.510	-30.712	20.682	1.00	26.00	C
ATOM	22	CD	PRO	A	4	-38.061	-31.301	21.999	1.00	25.07	C
ATOM	23	N	TYR	A	5	-34.862	-30.723	20.469	1.00	23.26	N

PDB檔案內容- HETATM



							X	y	Z	
ATOM	5804	ND2	ASN	C	261		18.723	-11.726	-23.291	1.00 30.12 N
ATOM	5805	N	CYS	C	262		14.285	-9.822	-20.975	1.00 30.45 N
ATOM	5806	CA	CYS	C	262		12.940	-9.294	-20.885	1.00 33.50 C
ATOM	5807	C	CYS	C	262		13.012	-7.773	-20.683	1.00 35.54 C
ATOM	5808	O	CYS	C	262		13.342	-7.323	-19.589	1.00 39.92 O
ATOM	5809	CB	CYS	C	262		12.178	-10.029	-19.765	1.00 32.71 C
ATOM	5810	SG	CYS	C	262		10.381	-10.019	-19.990	1.00 33.77 S
TER	5811		CYS	C	262					
HETATM	5812	MG		MG	C	301	-0.050	-7.496	-40.983	0.50 23.98 MG
HETATM	5813	O	HOH	A	301		-18.774	-8.400	42.837	1.00 28.05 O
HETATM	5814	O	HOH	A	302		-10.532	-34.797	24.956	1.00 27.78 O
HETATM	5815	O	HOH	A	303		-9.357	-5.734	36.827	1.00 18.49 O
HETATM	5816	O	HOH	A	304		-6.470	-18.360	27.628	1.00 21.15 O
HETATM	5817	O	HOH	A	305		-14.156	-17.221	10.176	1.00 25.29 O
HETATM	5818	O	HOH	A	306		-1.366	-5.876	29.725	1.00 29.05 O
HETATM	5819	O	HOH	A	307		-25.001	-35.477	30.661	1.00 27.88 O
HETATM	5820	O	HOH	A	308		-16.346	-28.091	23.272	1.00 13.73 O
HETATM	5821	O	HOH	A	309		-38.083	-24.557	17.452	1.00 33.39 O
HETATM	5822	O	HOH	A	310		-28.481	-26.306	18.284	1.00 15.63 O
HETATM	5823	O	HOH	A	311		-12.427	4.993	16.337	1.00 19.70 O
HETATM	5824	O	HOH	A	312		-20.995	-11.502	41.579	1.00 22.05 O

結構檔案中的金屬離子或化合物會以**HETATM**的形式註解。像本例中的**MG**

ATOM- Record Format

ATOM	5809	CB	CYS	C	262	12.178	-10.029	-19.765	1.00	32.71	C	
ATOM	5810	SG	CYS	C	262	10.381	-10.019	-19.990	1.00	33.77	S	
TER	5811		CYS	C	262							
HETATM	5812	MG		MG	C	301	-0.050	-7.496	-40.983	0.50	23.98	MG
HETATM	5813	O	HOH	A	301	-18.774	-8.400	42.837	1.00	28.05	O	

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ATOM "	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31 - 38	Real(8.3)	x	Orthogonal coordinates for X in Angstroms.
39 - 46	Real(8.3)	y	Orthogonal coordinates for Y in Angstroms.
47 - 54	Real(8.3)	z	Orthogonal coordinates for Z in Angstroms.
55 - 60	Real(6.2)	occupancy	Occupancy.
61 - 66	Real(6.2)	tempFactor	Temperature factor.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

Check protein data file from CoLab

Upload

01_Protein Data Bank.ipynb

To

Colab Notebooks 

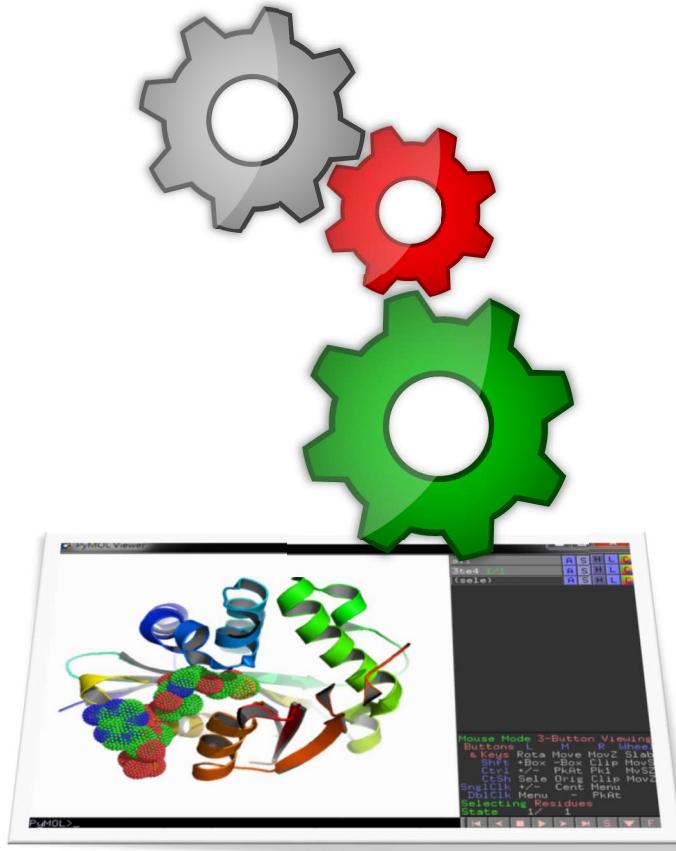
Protein Structure Visualization

蛋白質結構視覺化

蛋白質結構視覺化

工欲善其事 必先利其器

Visualization of Protein Structure



Protein sequence

```
>3TE4:A|PDBID|CHAIN|SEQUENCE
GPLGSPYTIELIQPEDGEAVIAMLKTFKKDEI
PSPDDVPEKAADSCEHPKFKKILSLMDHVEEQI
VYHVLCSSHYSARVMEKLGFHEVFRMQFADYKQ
EKYSLKPLPD
DGKILSVDTN
GIQVMAKEV
```



PDB file format

- 3D structure information



Visualization softwares

- PyMOL
- UCSF Chimera
- RasMol ...

*From Sequence
to Structure*

每個原子的3D座標

轉換成立體結構
結構視覺化軟體工具

Download PyMOL

Download PyMOL 2.5

Version 2.5.2 - Updated August 20th 2021 ([Installation instructions](#))

For previous versions, [see here](#).

These bundles include Python 3.7.



Windows

EXE Installer



Windows

ZIP Archive



macOS

DMG Disk Image



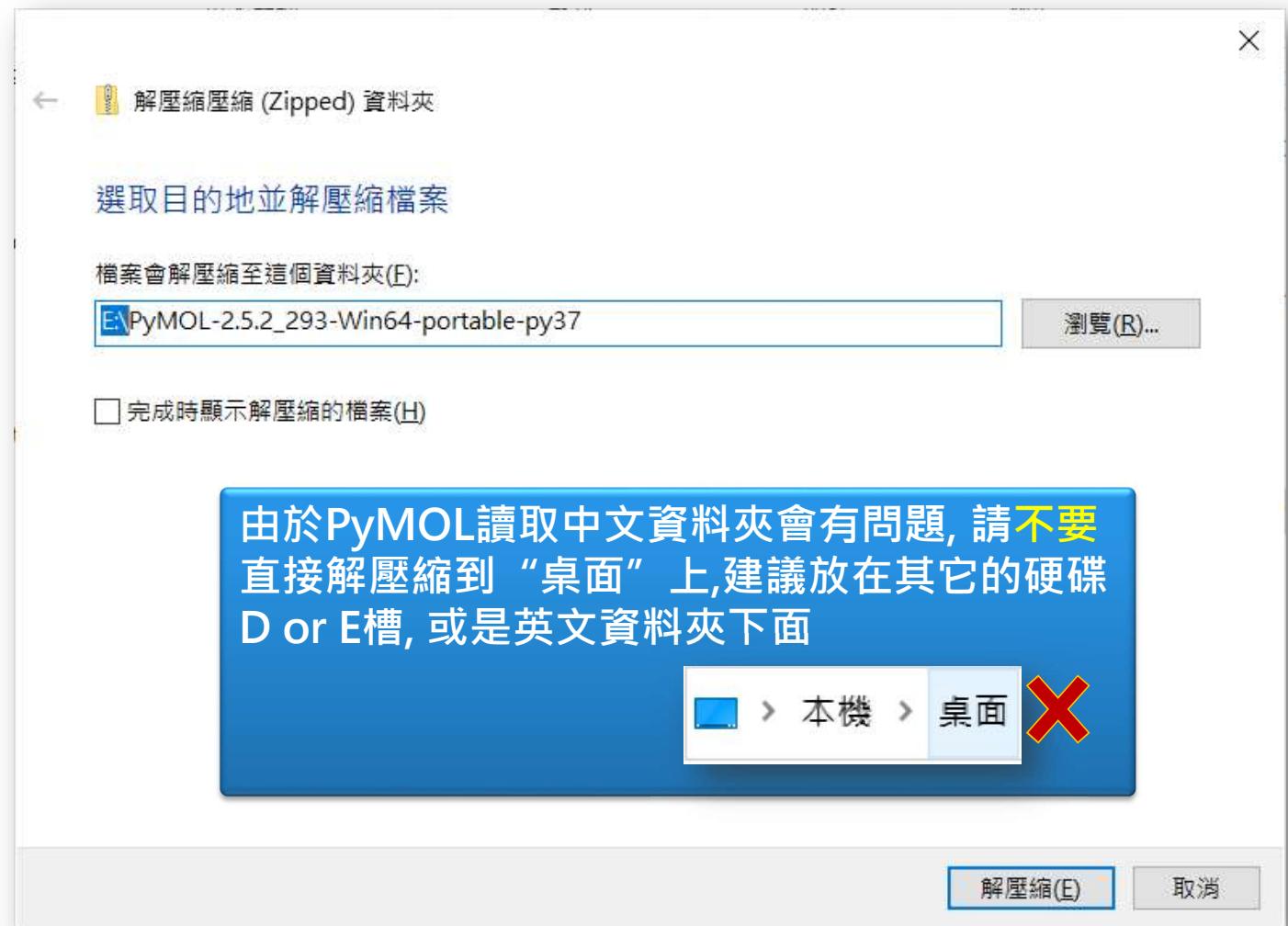
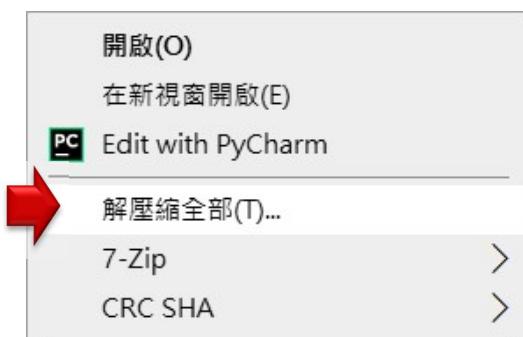
Linux

TAR.BZ2 Archive

<https://pymol.org/2/>

PyMOL-2.5.2_293-Win64-portable-py37.zip

*在zip檔案上按滑鼠右鍵，
選擇
解壓縮全部(T)...



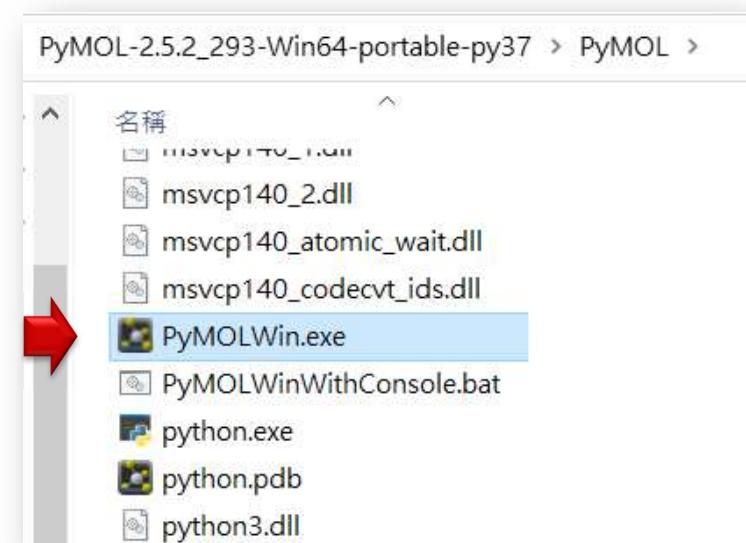
安裝 PyMOL

On Windows

- ZIP Archive: Unzip and run PyMOLWin.exe
(ZIP解壓縮後，直接執行PyMOLWin.exe)

On a Mac

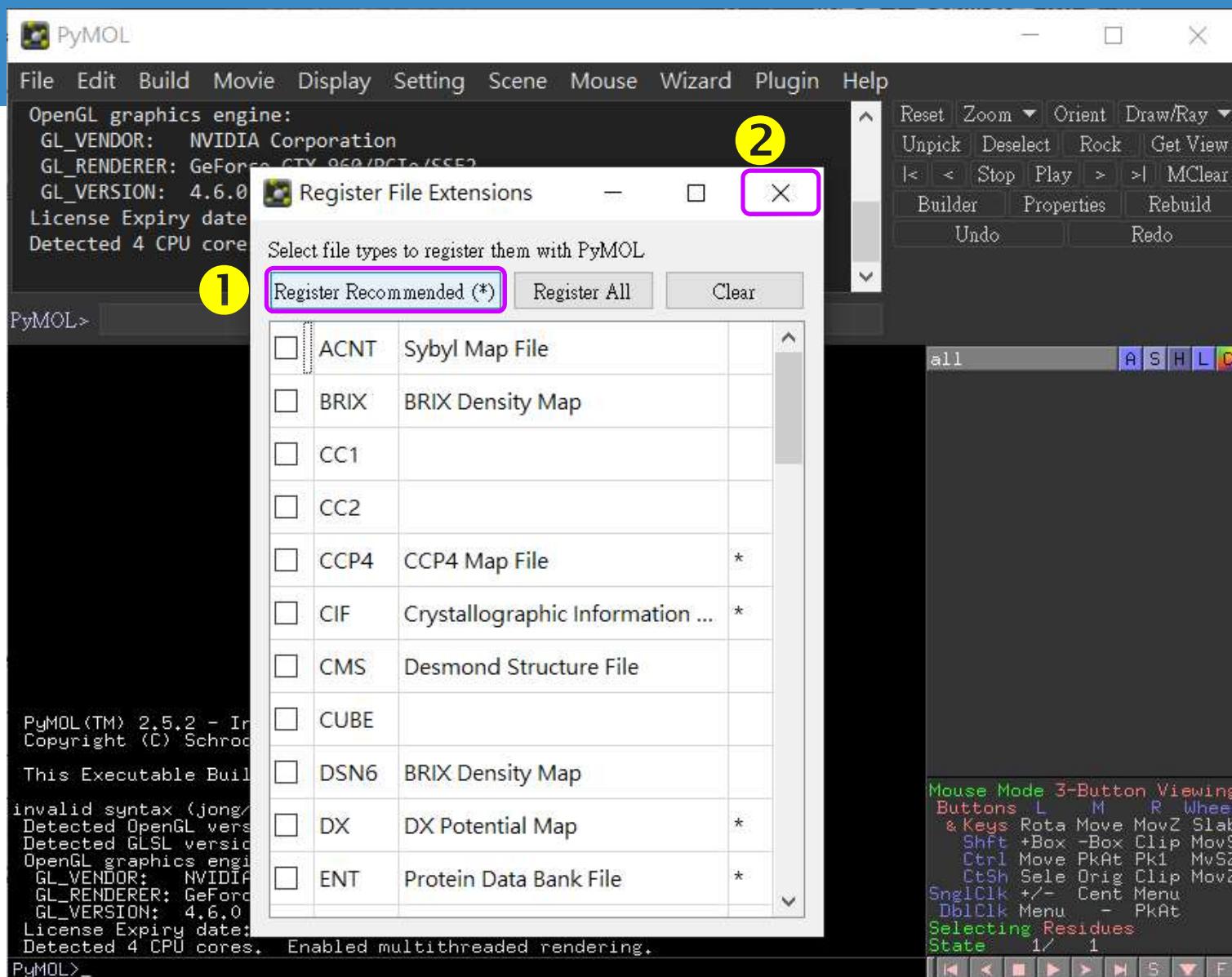
- 1.Double-click on *.dmg file
- 2.Drag PyMOL icon to Applications folder



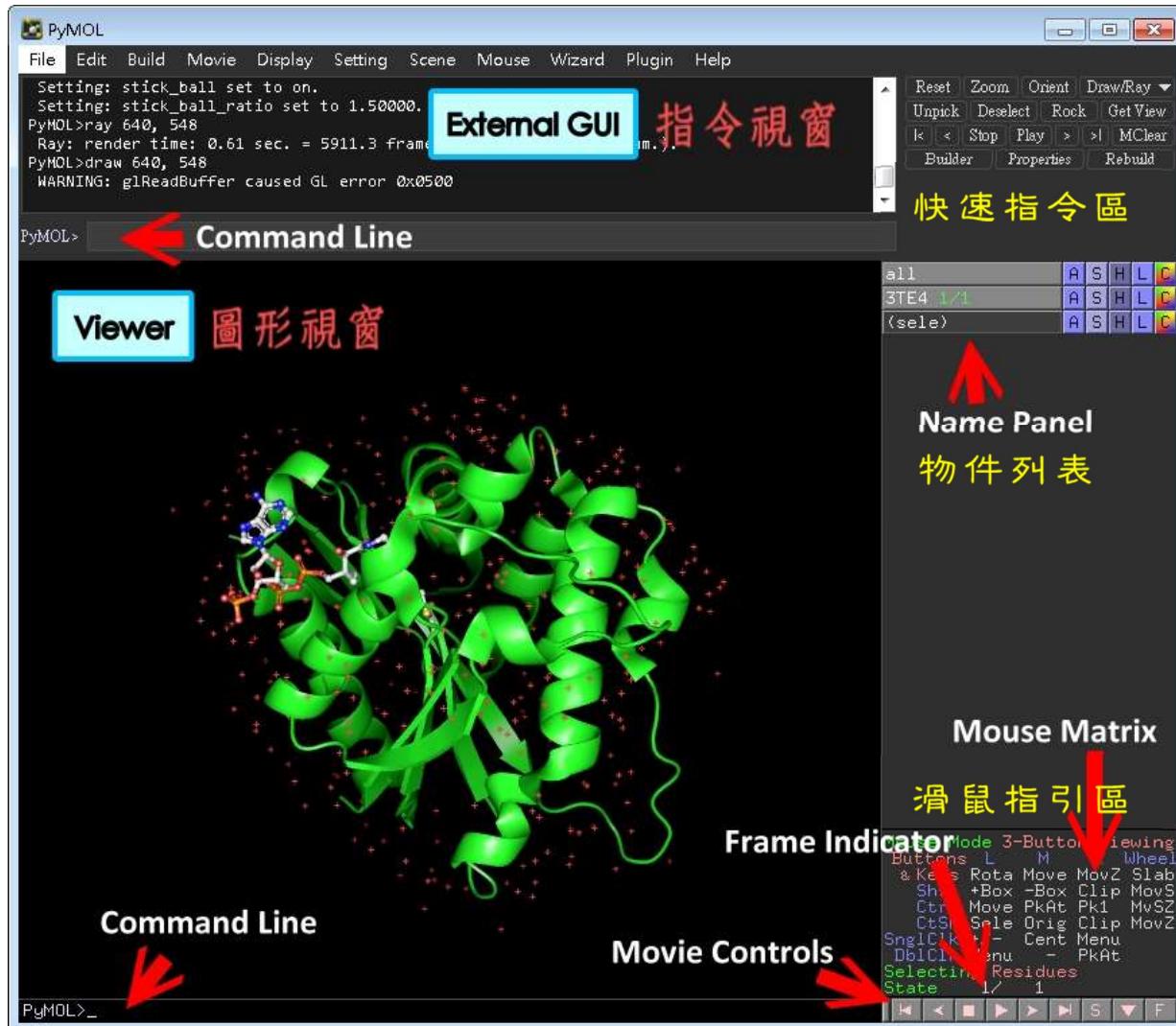
On Linux

- 1.Unpack the .tar.bz2 file, e.g.: tar -jxf PyMOL-2.2.0_0-Linux-x86_64.tar.bz2
- 2.Run PyMOL with: cd pymol; ./pymol

<https://pymol.org/2/support.html?#installation>



PyMOL 軟體操作介面



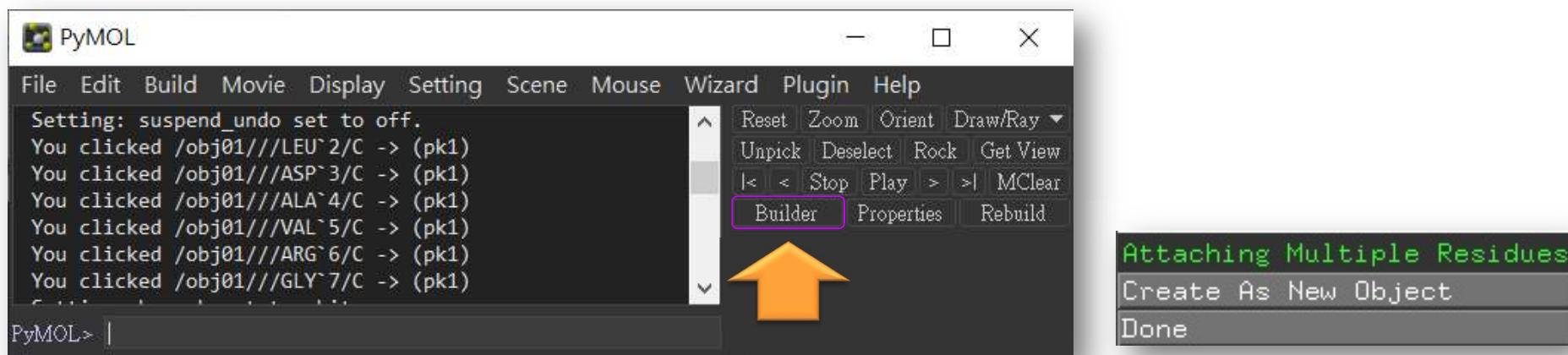
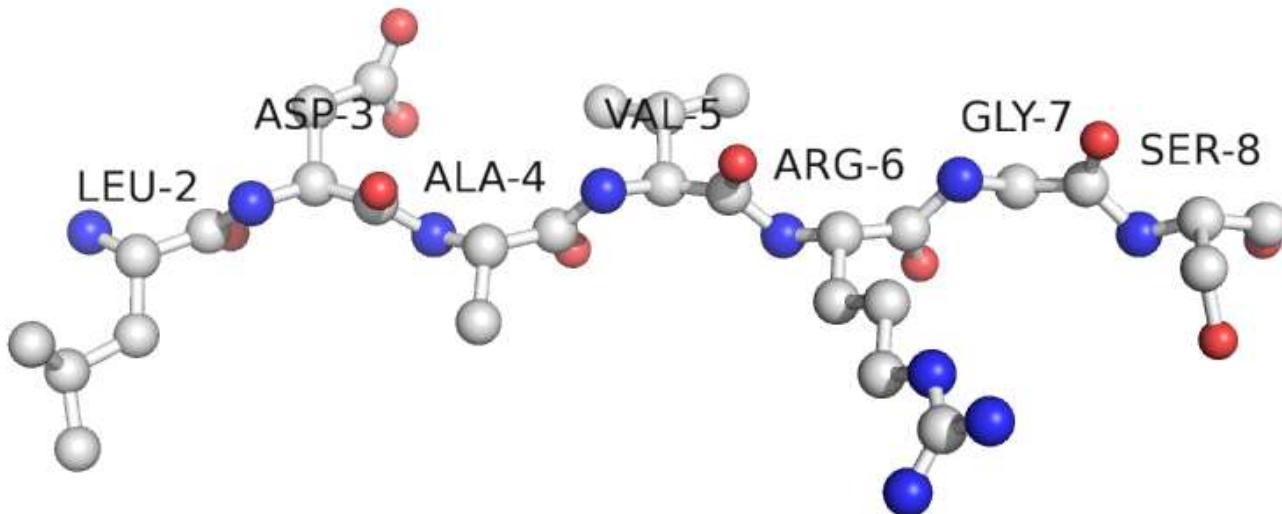
*控制個別物件按鈕



- A: action
- S: show
- H: hide
- L: label
- C: color



Practice 1: Build a peptide from PyMOL



Practice 1: Build a peptide from PyMOL

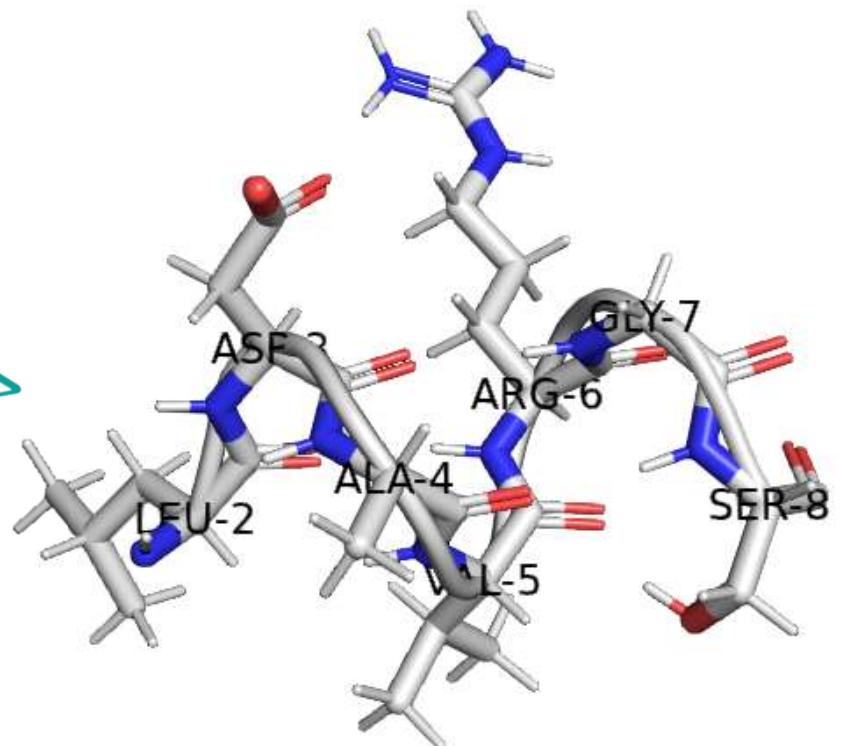
開啟 Builder, 切換 Protein

Builder

Chemical Protein Nucleic Acid

Ace	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu
Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	NMe	NHH	

Secondary Structure: Alpha Helix



選擇Alpha Helix二級結構

*可檢視backbone, side chain

蛋白質的3D立體結構
蛋白質的3D立體結構
可以提供我們什麼訊息？

You got the Protein Structure

what's the **NEXT?**

For example,

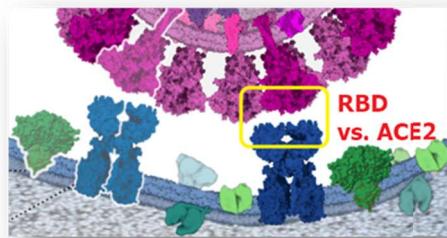
ACE2

Angiotensin-Converting Enzyme 2 (ACE2)

血管緊張素轉換酶



ACE2: The Receptor for SARS-CoV-2



SARS-CoV-2與靶細胞的
表面蛋白相互作用

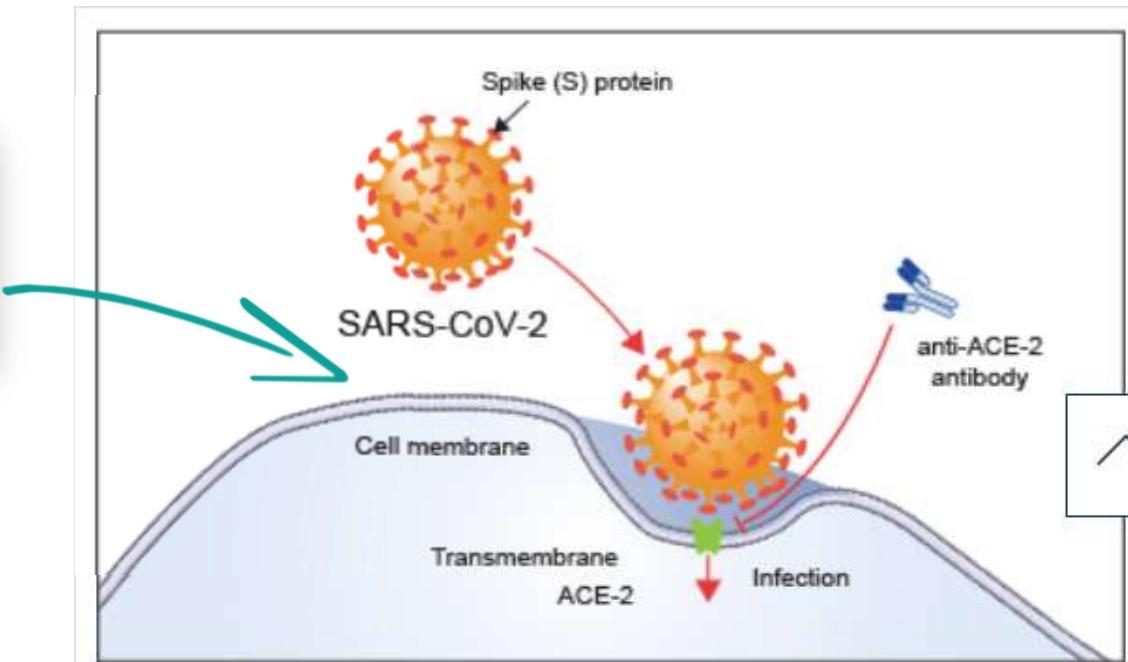
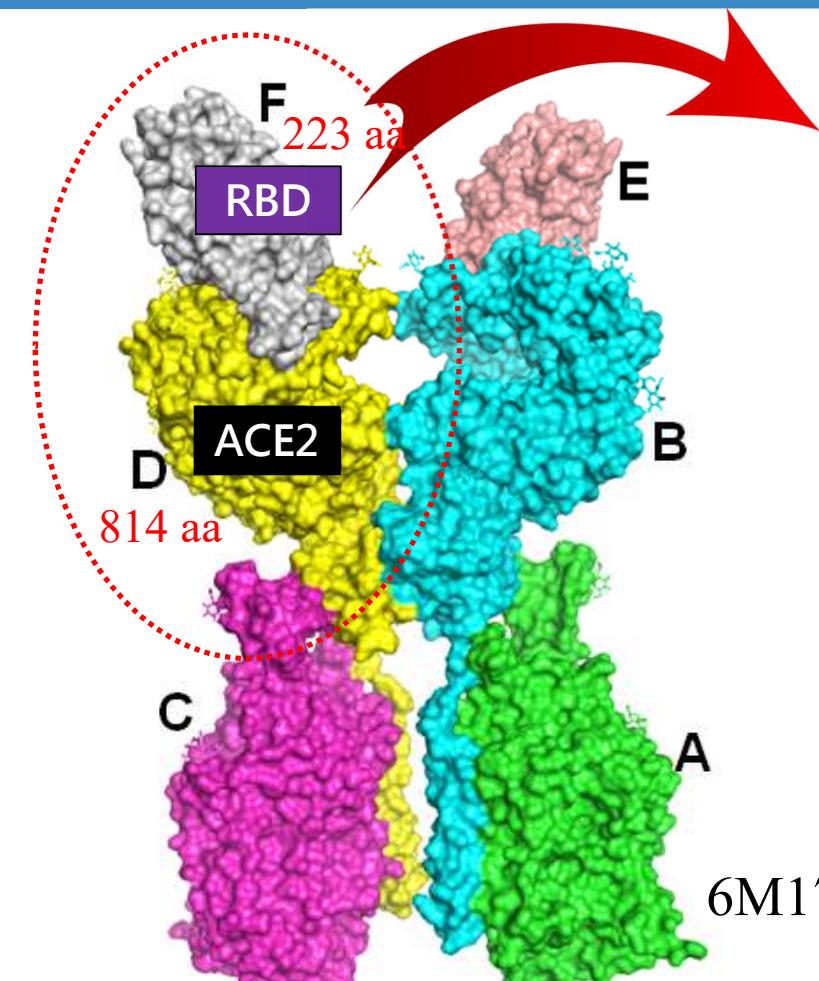


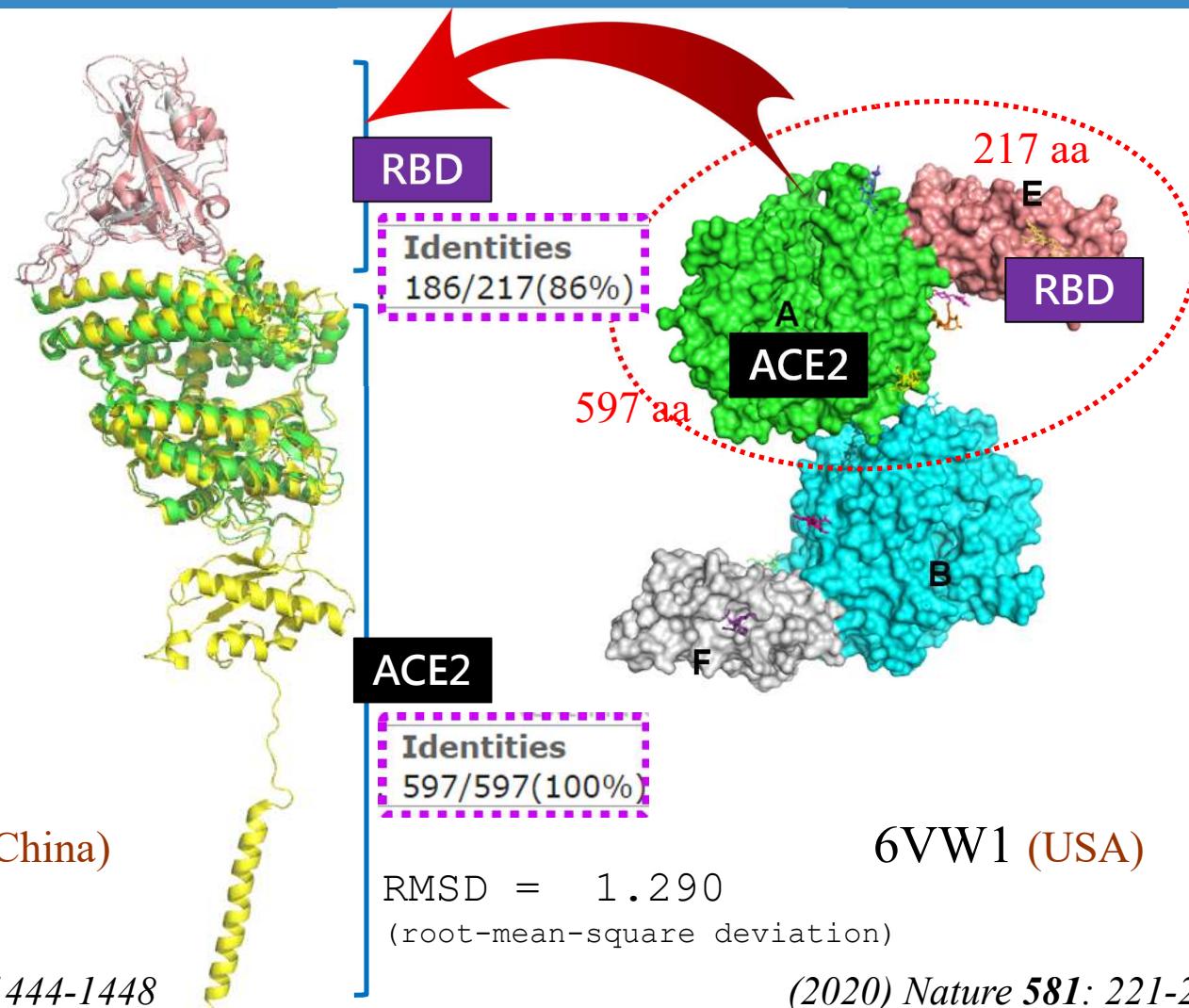
Figure 1. ACE-2 is the host cell receptor responsible for mediating infection by SARS-CoV-2, the novel coronavirus responsible for coronavirus disease 2019 (COVID-19). Treatment with anti-ACE-2 antibodies disrupts the interaction between virus and receptor.

<https://www.rndsystems.com/resources/articles/ace-2-sars-receptor-identified>

Structure Alignment



(2020) *Science* **367**: 1444-1448

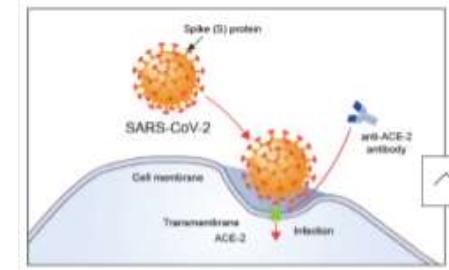


(2020) *Nature* **581**: 221-224

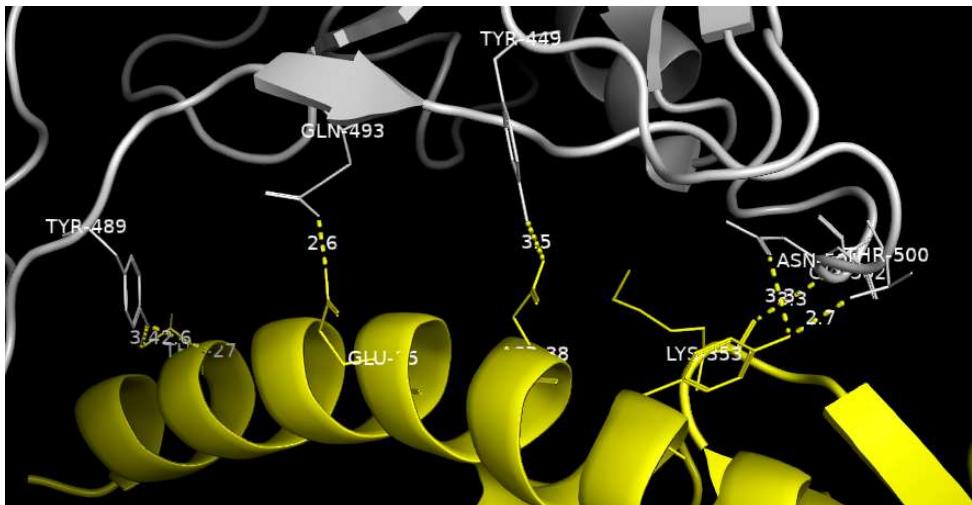
What's the NEXT?

The
coronavirus recognition sites?

識別位點



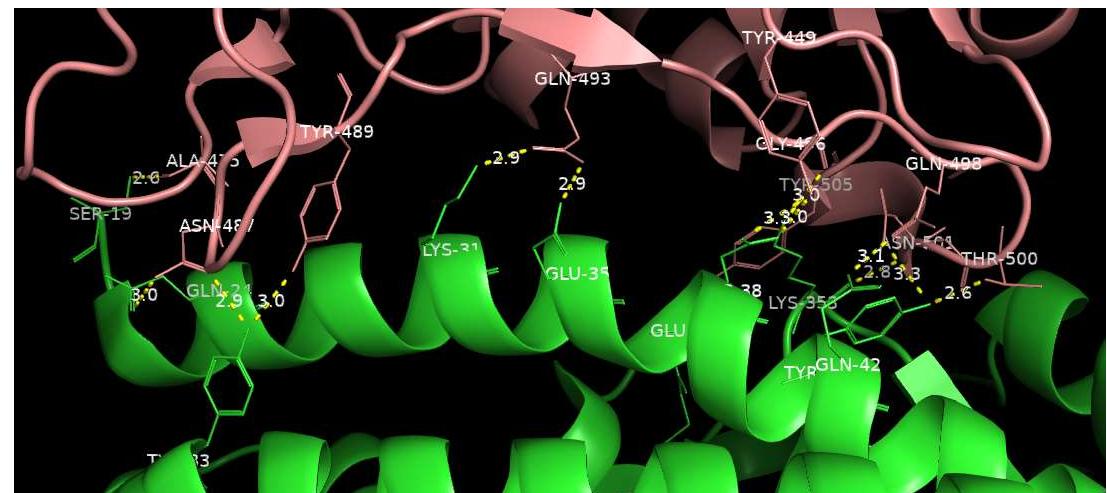
顯示氫鍵結合的位置



6M17 (China)

RBD	Tyr489, Gln493, Tyr449, Thr500, Asn501*, Gly502
ACE2	Thr27, Glu35, Asp38, Tyr41, Lys353

(2020) *Science* **367**: 1444-1448



6VW1 (USA)

RBD	Ala475, Asn487, Tyr489, Gln493, Gly496, Tyr449, Gln498, Thr500, Asn501*, Try505
ACE2	Ser19, Gln24, Tyr83, Lys31, Glu35, Glu37, Asp38, Tyr41, Gln42, Lys353, Arg393

(2020) *Nature* **581**: 221-224

PyMOL 操作環境介紹

PyMOL 操作環境介紹

PyMOL 教學網頁

The screenshot displays a PyMOL tutorial page with a yellow header and sidebar. The header includes the NTHU Bioinformatics Center logo and the title "PyMOL Tutorial". The sidebar lists five parts of the tutorial: Part 1 (Download and Installation), Part 2 (Opening Files and Interface Introduction), Part 3 (Basic Operations) which is expanded to show sub-topics like Secondary Structure, Protein之外的物質, Hydrophobic Amino Acids, Ligand sites, Saving Images, Script and log file, Distance between atoms, Slab, Fixed specific atom rotation, Atom representation, Stereo 3D diagram, Specific atom specification, Transparency settings, and Shortcuts; Part 4 (Advanced Operations), and Part 5 (Practical Exercises). The main content area shows the PyMOL Molecular Graphics System window. The top menu bar includes File, Edit, Build, Movie, Display, Setting, Some, Mouse, Wizard, and Bug. The command line at the bottom shows PyMOL commands like "set stick_radius=3", "set stick_radius=0.5", and "ray". The right panel shows a list of objects: s11, S1E4, (sel0), obj01, S1E4_e_chg, S1E4_e_map, S1E4_e_pot, and (protein01). The central panel displays a 3D ribbon model of a protein structure.

A Tutorial for the PyMOL Basics

<http://140.114.98.75/sbio/pymol/>

PyMOL Tutorial List

▼Part 1- 下載與安裝

- Download PyMOL
- 安裝PyMOL
- 滑鼠的控制 ☆

▼Part 2- 開啟檔案與操作視窗介紹

- 開啟檔案
- 操作視窗介紹
 - A: Action
 - S: Show
 - H: Hide
 - L: Label
 - C: Color

▼Part 3- 基礎操作

- 二級結構
- protein之外的物質
- 疏水性胺基酸
- Ligand sites
- 儲存影像
- Script and log file
- 兩原子間的距離
- 分子內部圖(Slab)
- 固定特定原子旋轉
- 原子的展現方法
- 標示原子
- 分子3D圖(stereo)
- 指定特定原子
- 設定透明surface
- 縮寫表

▼Part 4- 進階操作

- protein/Ligand間的鍵結
- Structure alignment
- 該存與演示結構影像
- PyMOL簡單動畫製作
- Mutagenesis操作
- 蛋白質表面電荷

▼Part 5- 實例練習

- Cavity of DAT

List

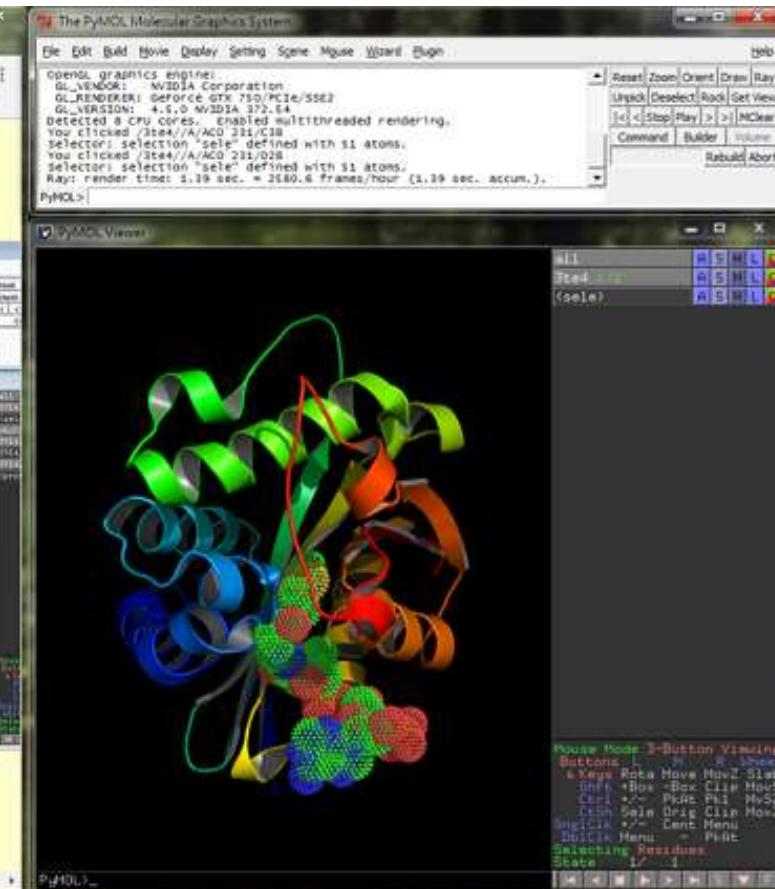


PyMOL Experience

Browse

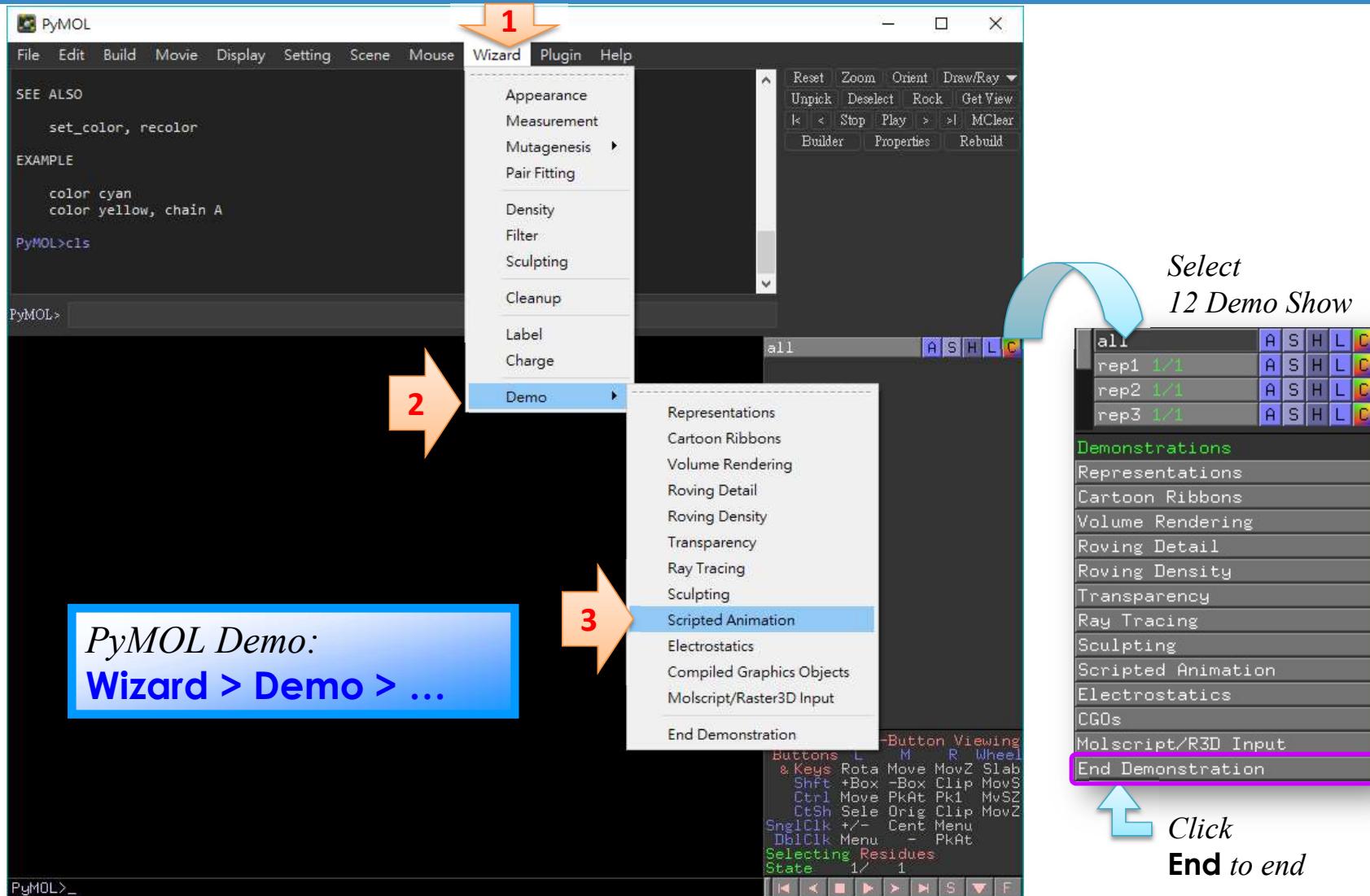
The screenshot shows a web browser window displaying a PyMOL tutorial page. The title is "PyMOL Tutorial". The page content includes a sidebar with links to "Part 1- 下載與安裝", "Part 2- 開啟檔案與操作幾個介紹", "Part 3- 基礎操作", "Part 4- 案例操作", and "Part 5- 實例練習". Below the sidebar is a text box stating "Edit by Yi-Chung Liu" and "PCLyu's Lab, NTHU". The main area features a 3D molecular model of a protein complex, with a color-coded ribbon representation. A status bar at the bottom right says "PyMOL 1.3.0".

PyMOL



*建議可以將教學網頁放在左邊參考, PyMOL程式放在右邊操作

Try PyMOL Demo



輝瑞-新冠抗病毒口服藥 Paxlovid



衛 生 福 利 部
Ministry of Health and Welfare
促進全民健康與福祉

請輸入關鍵字



進階

熱門關鍵字：[COVID-19](#) [防疫補償](#) [疫苗](#) [長期照顧](#)

[本部簡介](#) • [最新消息](#) • [便民服務](#) • [法令規章](#) • [衛教視窗](#) • [重大事件](#) • [本部各單位及所屬機關](#)

首頁 / 最新消息 / 焦點新聞



衛福部食藥署同意核准輝瑞公司COVID-19口服抗病毒藥品 Paxlovid
專案輸入



- 資料來源：食品藥物管理署
- 建檔日期：111-01-15
- 更新時間：111-01-17

<https://www.mohw.gov.tw/cp-16-65592-1.html>

Paxlovid 建議使用對象

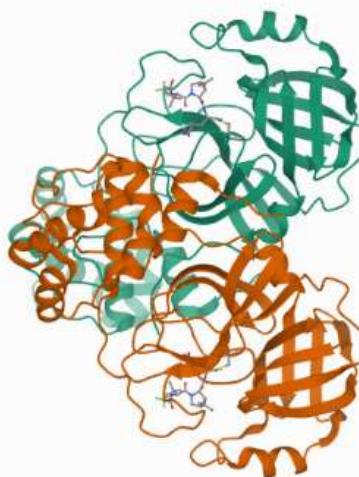
我國診治指引

針對「Paxlovid」建議使用之對象

- ◆具以下任一重症風險因子，未使用
氧氣且於發病五天內之 \geq 12歲且
體重 \geq 40公斤病患。
- ◆重症風險因子：年齡 \geq 65歲、糖尿病、慢性
腎病、心血管疾病(含高血壓)、慢性肺疾、
BMI \geq 25 (或12-17歲兒童青少年BMI 超過
同齡第85百分位)、其他影響免疫功能之疾病
或已知重症風險因子者。

SARS-CoV-2 main protease (Mpro) + Paxlovid

Biological Assembly 1 



7SI9

Room temperature X-ray structure of SARS-CoV-2 main protease (Mpro) in complex with PF-07321332

PDB DOI: [10.22110/pdb7SI9/pdb](https://doi.org/10.22110/pdb7SI9/pdb)

Classification: HYDROLASE

Organism(s): Severe acute respiratory syndrome coronavirus 2

Expression System: Escherichia coli

Mutation(s): No 

Deposited: 2021-10-12 Released: 2021-10-20

Deposition Author(s): Kovalevsky, A., Kneller, D.W., Coates, L.

Funding Organization(s): Not funded

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.00 Å

 Display Files  Download Files

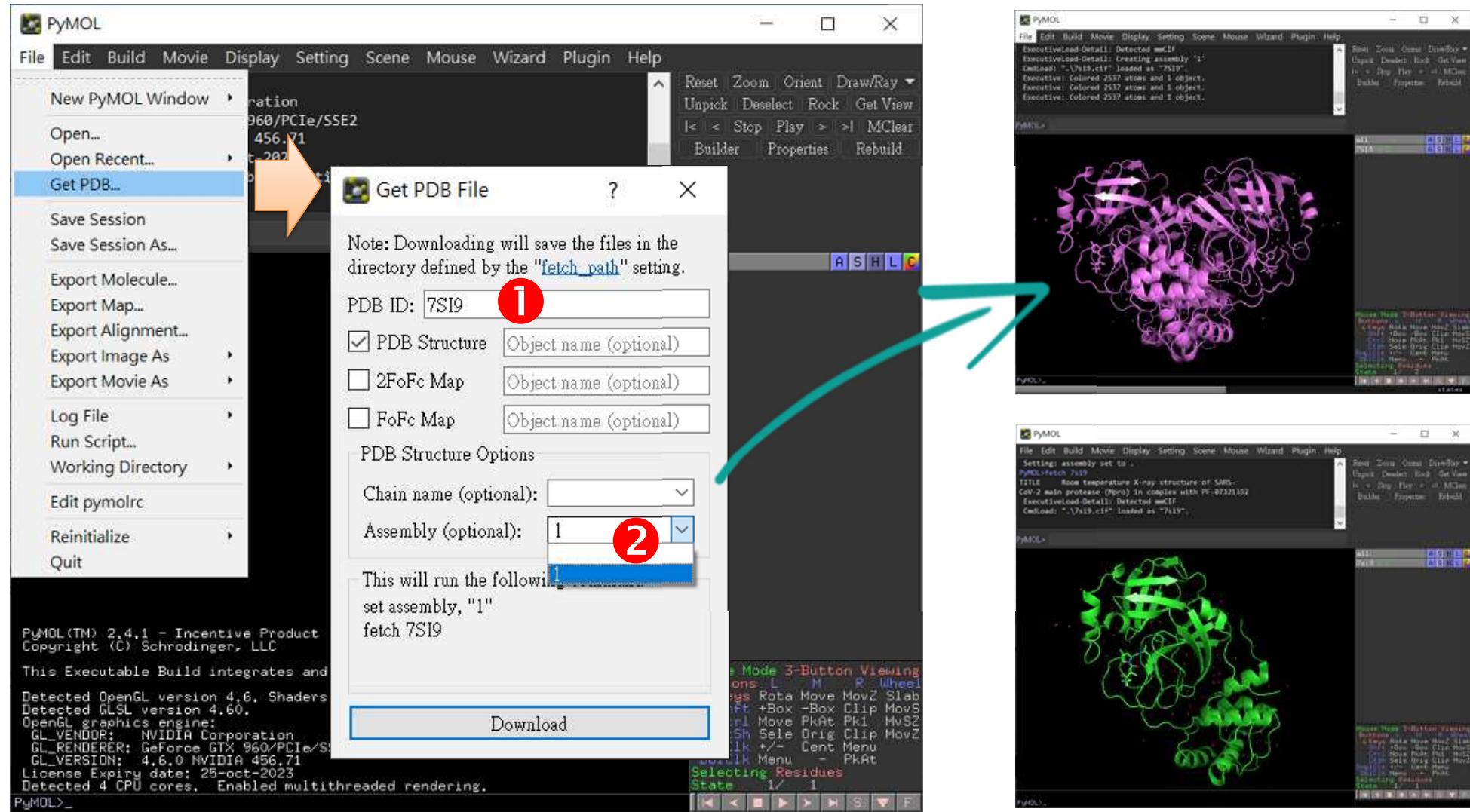
*Paxlovid是Mpro蛋白酶抑制劑

<https://www.rcsb.org/structure/7SI9>

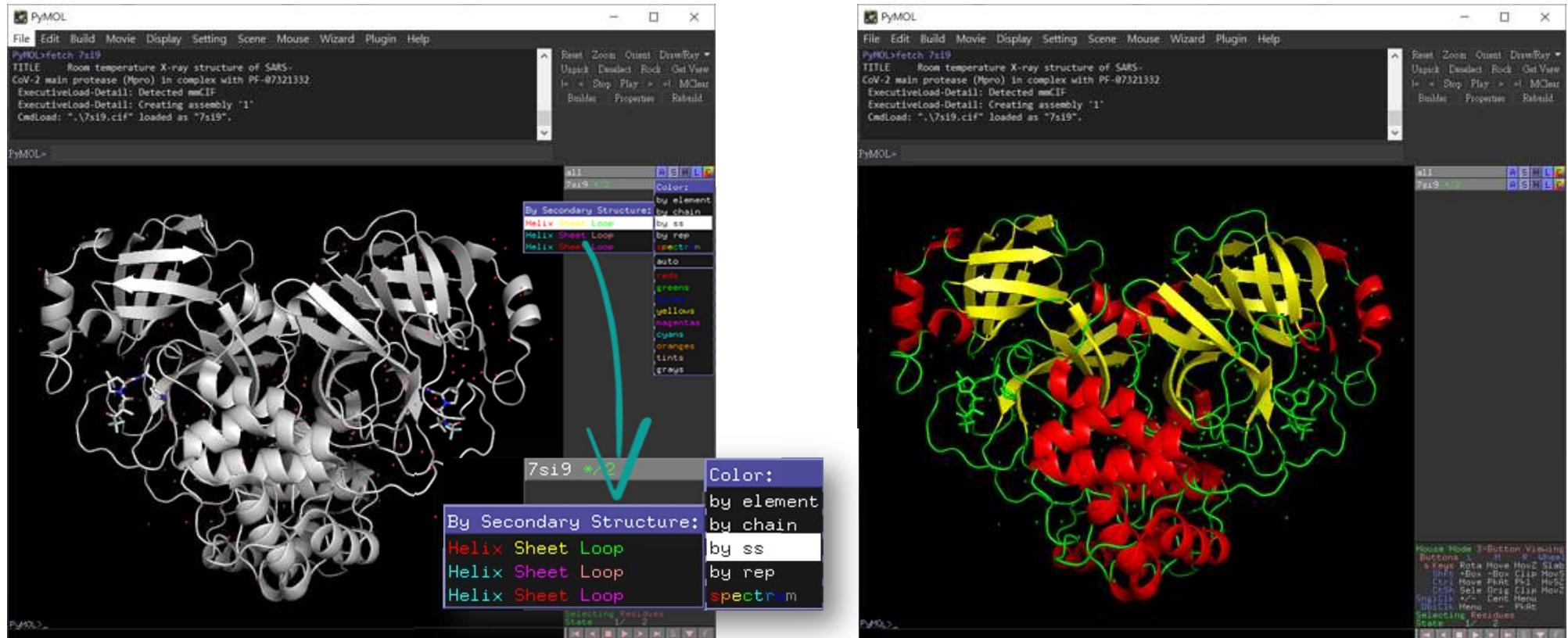
PyMOL 基礎操作

PyMOL 基礎操作

1. 開啟 7SI9

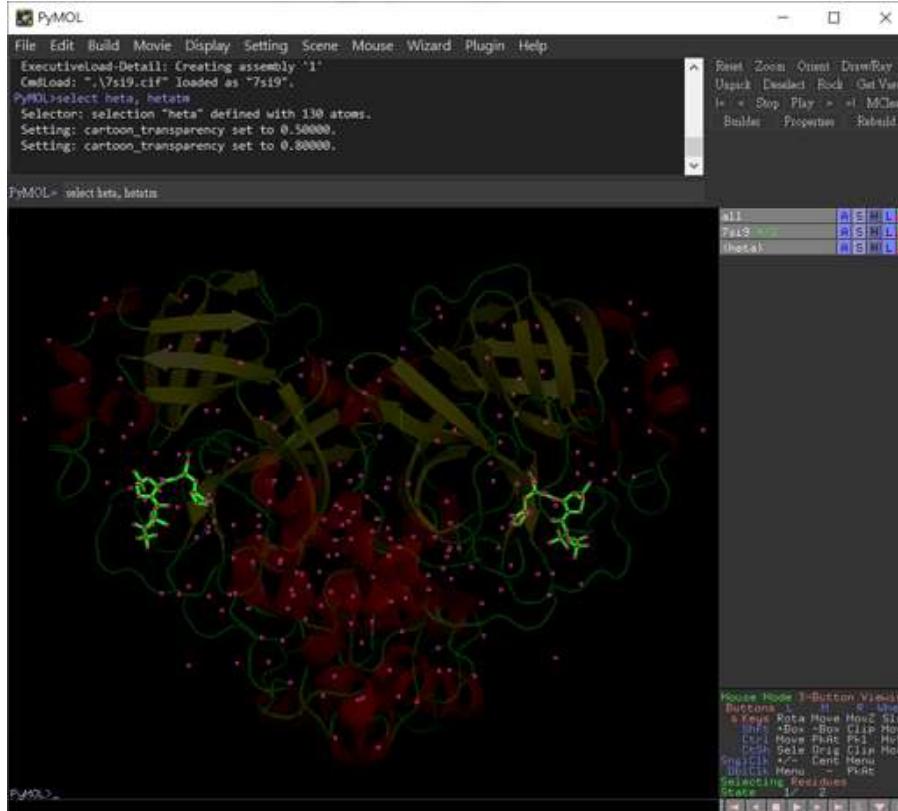


2. 檢視二級結構：哪裡是 α -螺旋和 β -長帶？



依Secondary Structure 著色，Helix, Sheet, Loop 共有3種套色可以選擇(見上圖)
Color by ss (By Secondary Structure)

3. 除了 protein chains之外，有沒有其它的東西？



>**select heta, hetatm**

選擇所要的東西，並命名為 *heta* (命名以方便辨識為宜)

>**select resn HOH**

選擇水 (resn= residue name)

或是加以命名:

>**select water, resn HOH**

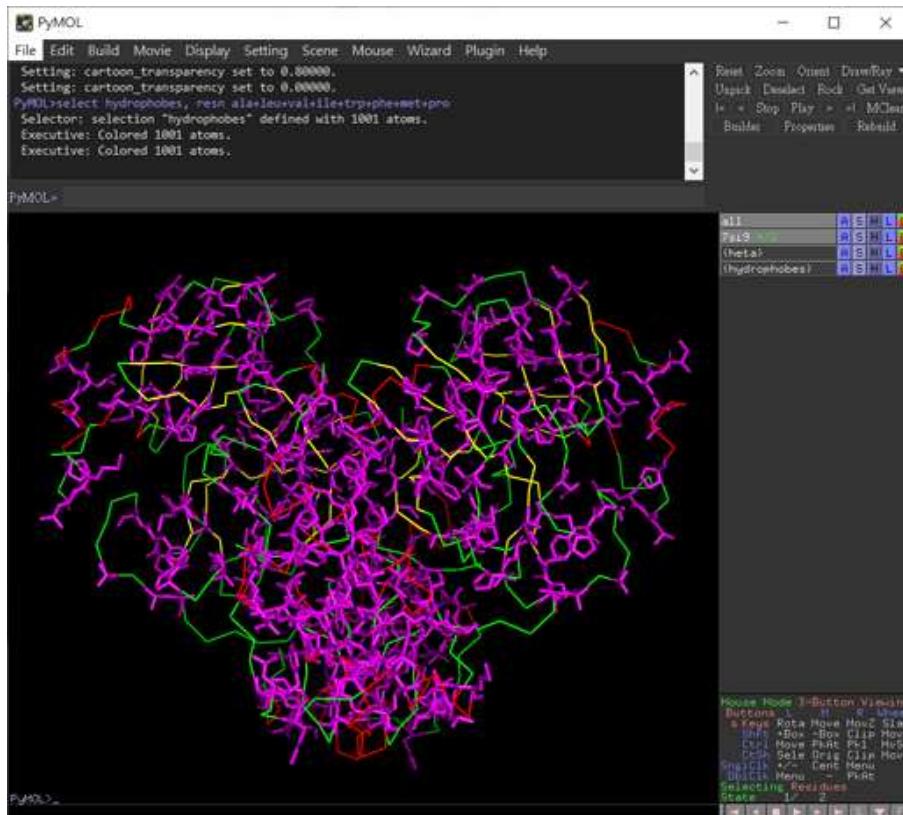
選擇水，並命名為 *water*

>**select hetatm and not resn HOH**

選擇除了水之外的 *hetro atoms*

*也可以直接從視窗介面選取與展示

4. 呈現疏水性的胺基酸



>**select hetat, hetatm**

選擇所要的東西，並命名為 *hetat* (命名以方便辨識為宜)

>**hide everything (或是用 hide all 亦可)**

>**show_as ribbon**

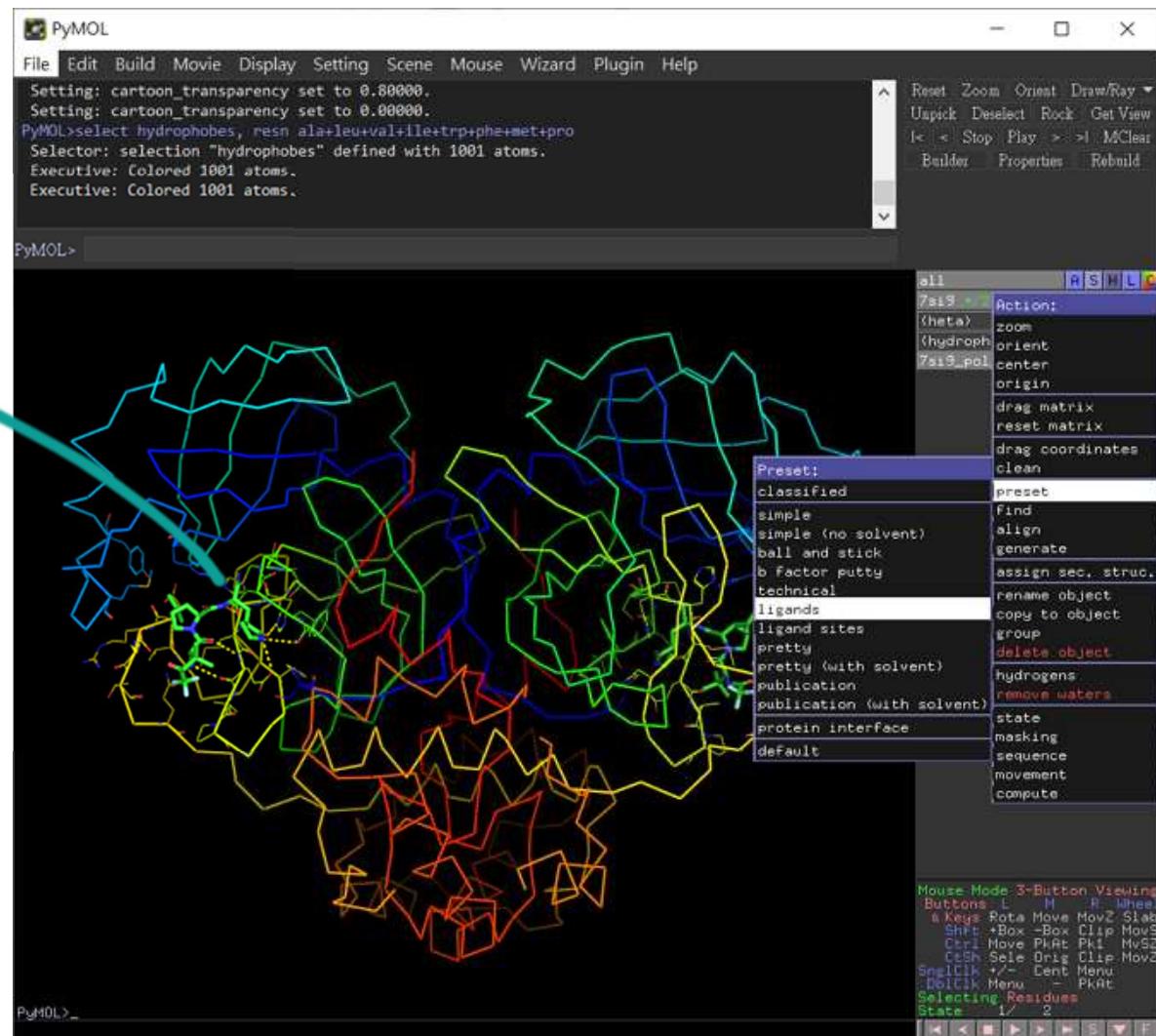
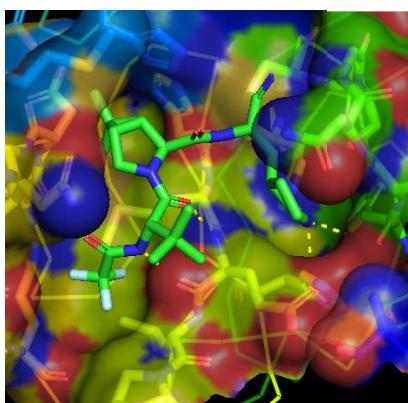
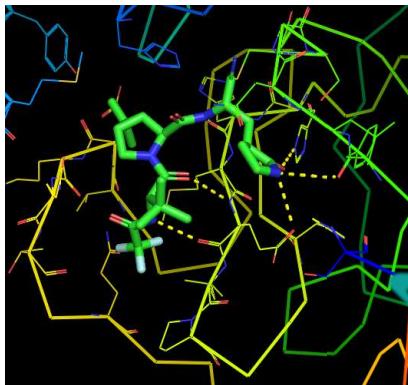
show everything as a ribbon, 這是一個非常好用的指令，常常我們讓結構顯示了多種3D樣式，但是最後只要顯示一種時，只要使用**show_as**，就可以直接完成。

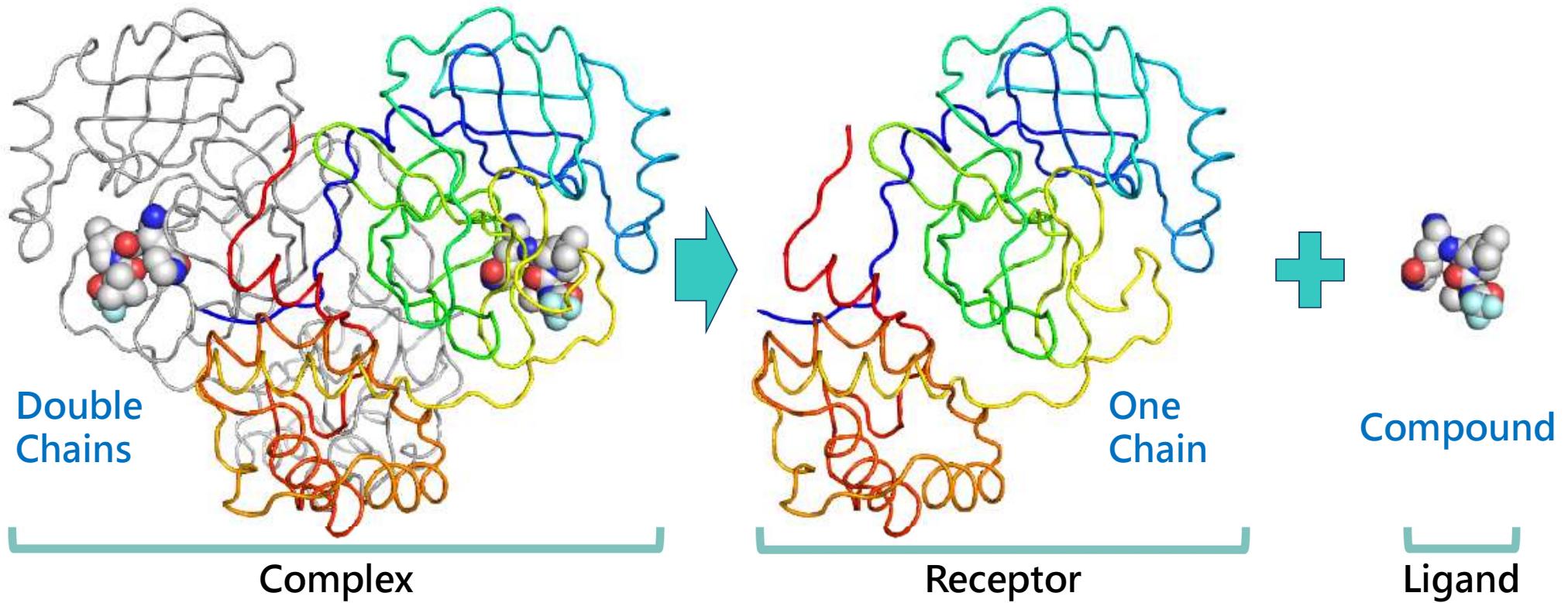
>**select hydrophobes, resn
ala+leu+val+ile+trp+phe+met+pro**

>**show sticks, hydrophobes**

>**color magenta, hydrophobes**

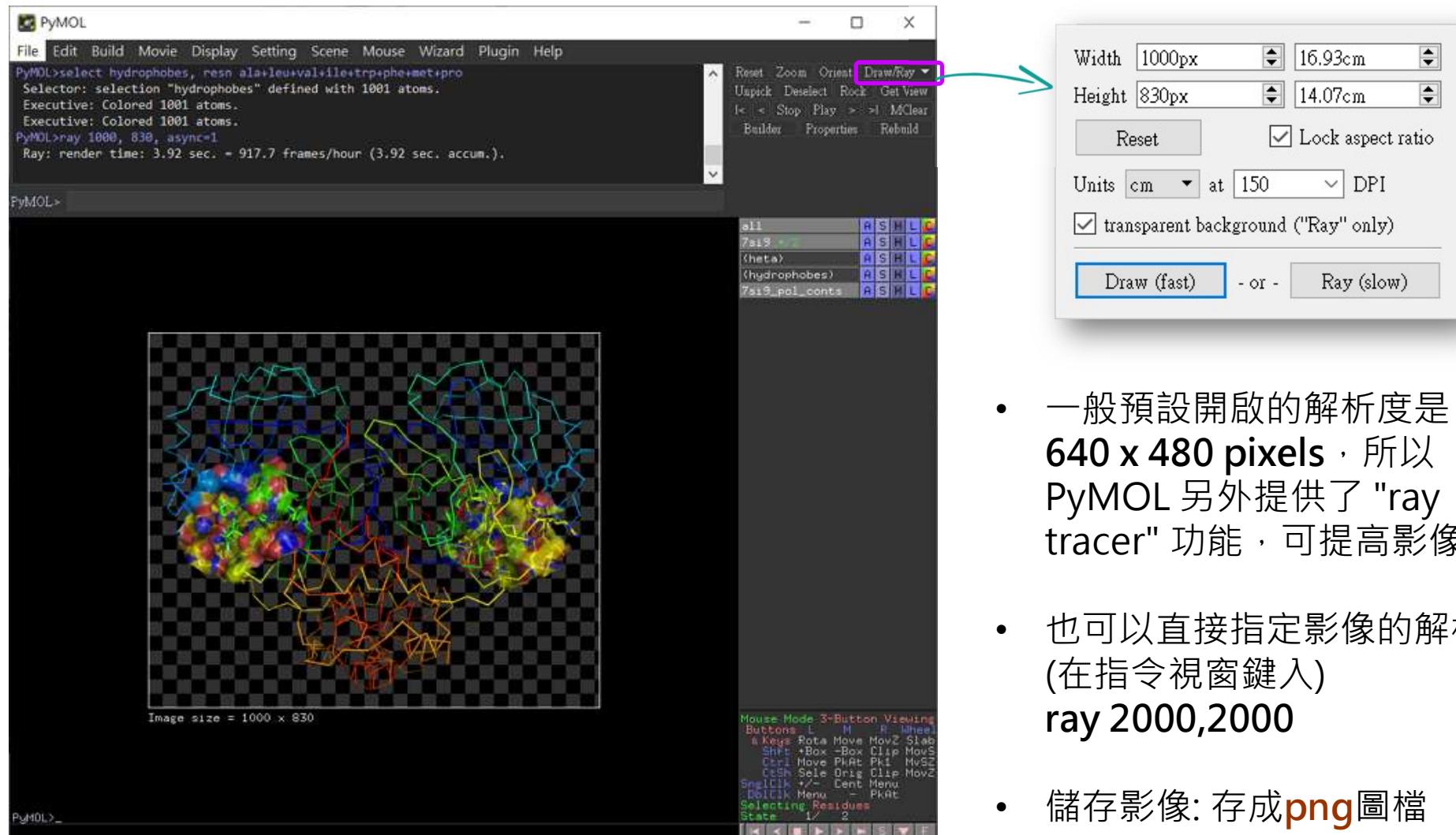
5. 如何呈現Ligand的位置?





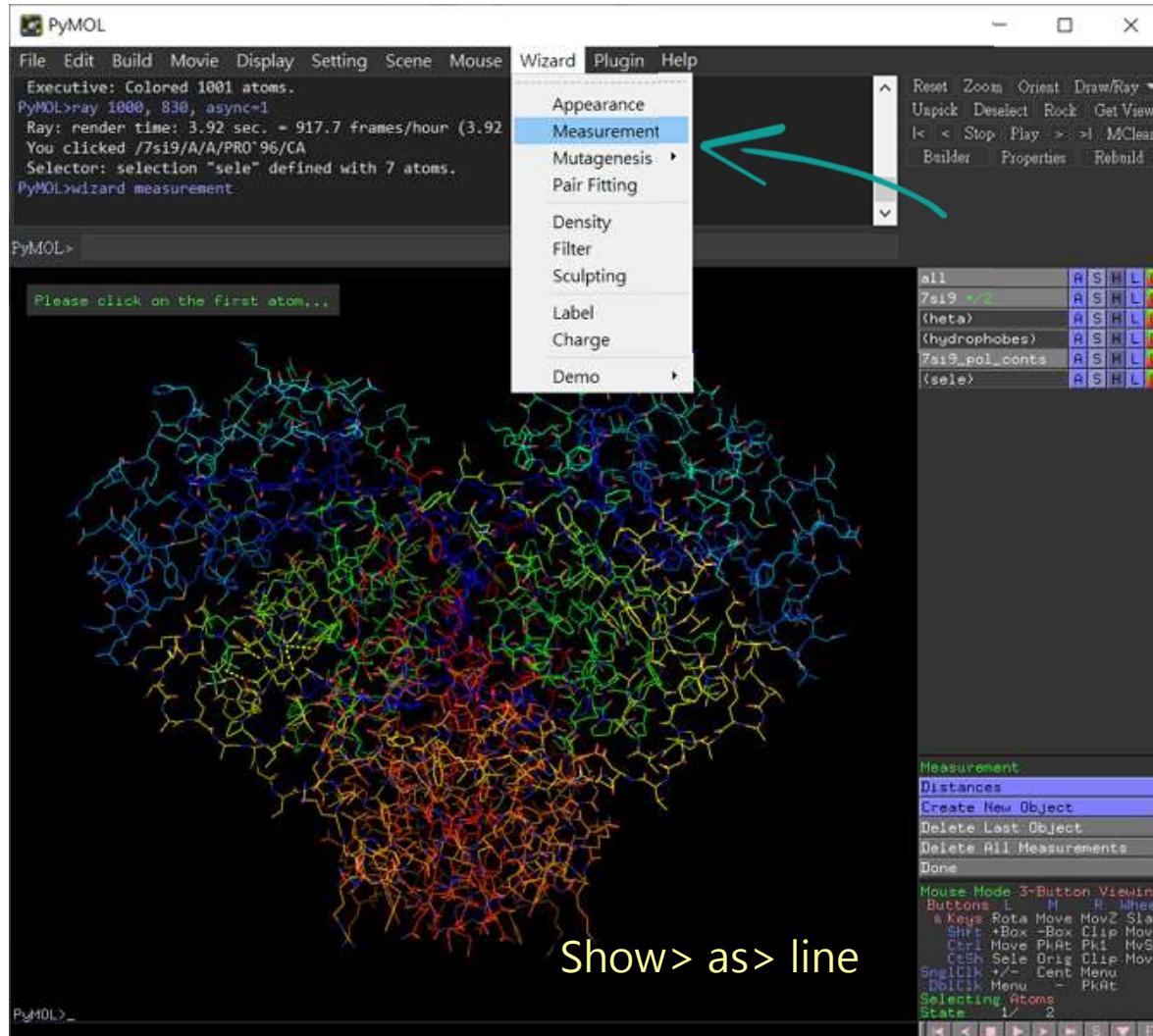
*Receptor與Ligand間主要會探討非共價鍵結作用力，包括靜電力（Ionic Interactions）、偶極力（Dipole Interactions, 如氫鍵）、凡得瓦力（Van Der Waals Interactions）、疏水性（Hydrophobic Effect）

6. 儲存影像

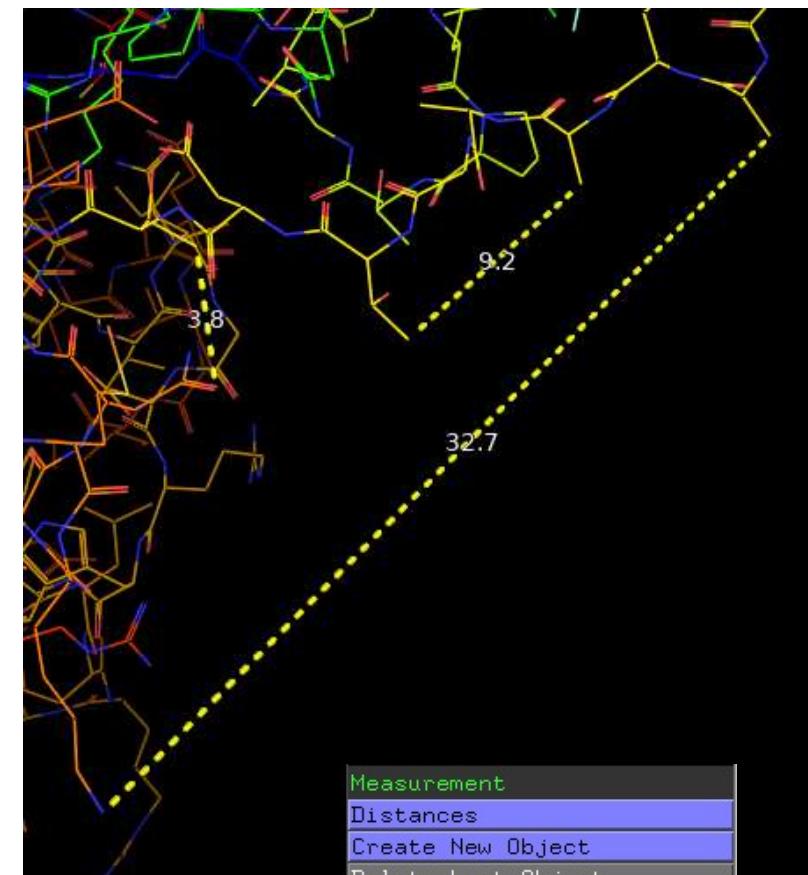


- 一般預設開啟的解析度是 **640 x 480 pixels**，所以 PyMOL 另外提供了 "ray tracer" 功能，可提高影像品質。
- 也可以直接指定影像的解析度：(在指令視窗鍵入)
ray 2000,2000
- 儲存影像：存成**png**圖檔

7. 如何測量兩個原子間的距離?



放大局部區域，以便測量兩原子間的距離



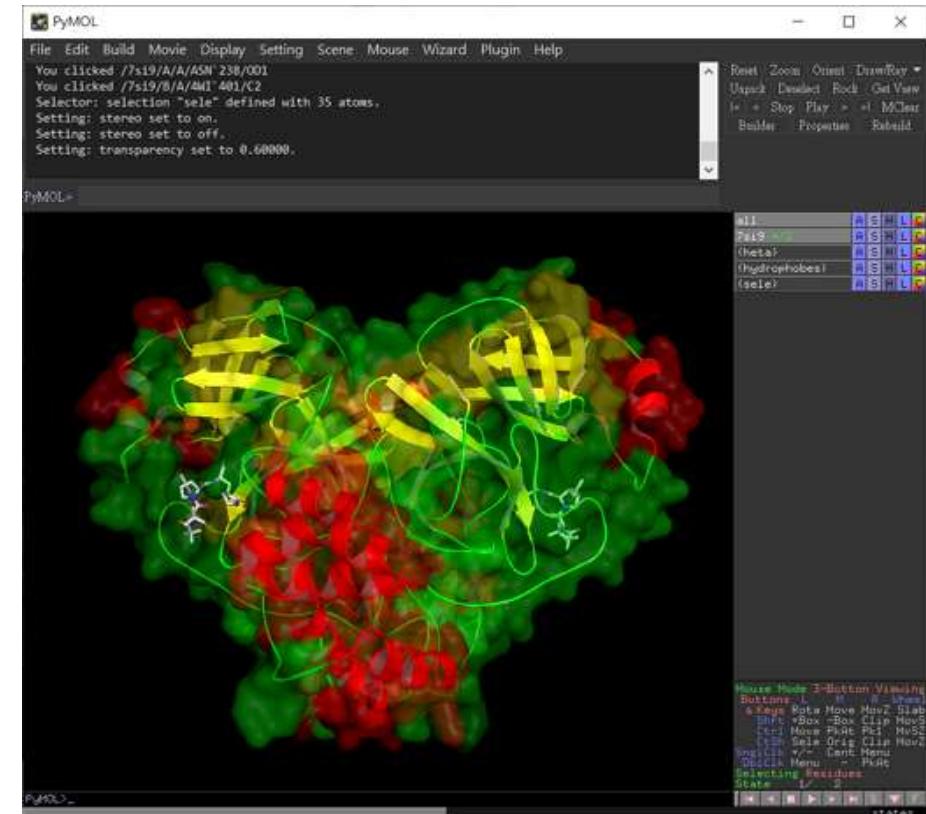
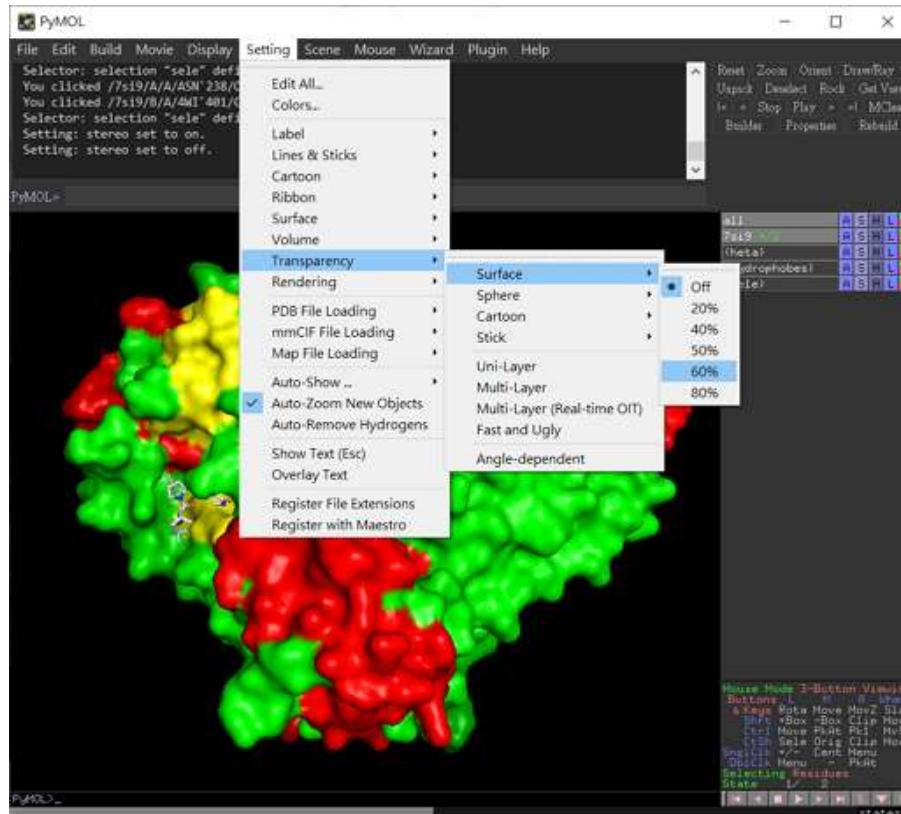
Measurement
Distances
Create New Object
Delete Last Object
Delete All Measurements
Done

8. 分子內部圖 (Slab mode)



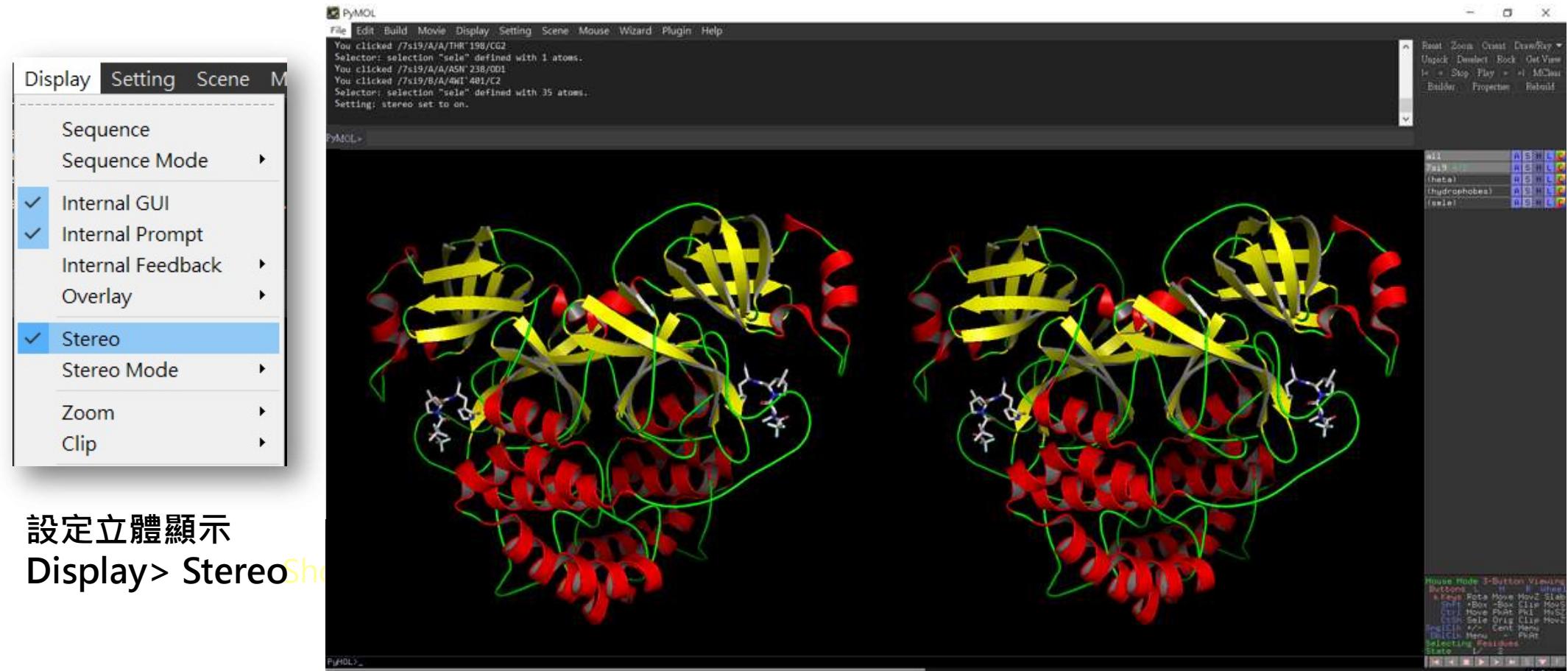
滾動滑鼠中鍵滾輪，執行平板面模式(slab mode)，呈現剖面圖，在剖面之前的原子將看不到，只看到剖面之後的物質。

9. 設定透明surface (Transparency)



- 在功能表Setting的下拉選單中選取Transparency> Surface> 60%
- 這是surface透明度60%的效果，同時也顯露原本被包覆的cartoon樣式

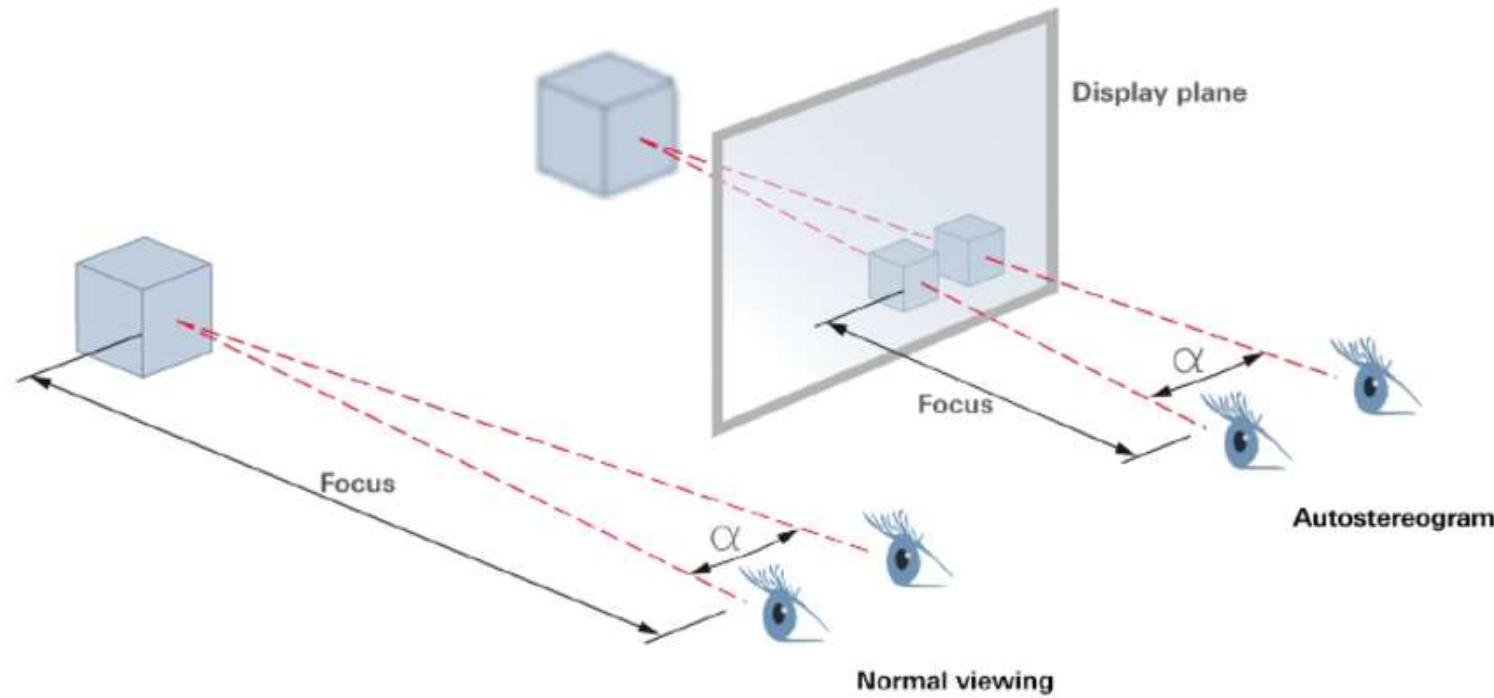
10. 我要如何看到分子的3D立體圖 (stereo view)



設定立體顯示

Display> Stereo

10. 我要如何看到分子的3D立體圖 (stereo view)

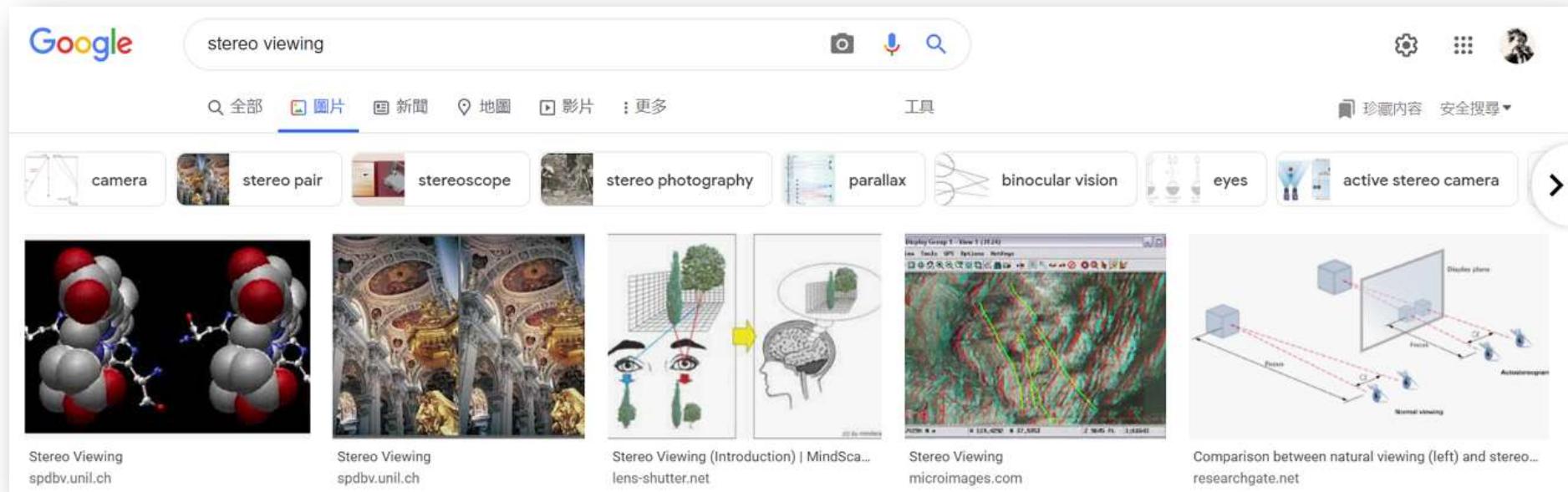


https://www.researchgate.net/figure/Comparison-between-natural-viewing-left-and-stereoscopic-viewing-with-a-3D-stereo_fig1_215478670

10. 我要如何看到分子的3D立體圖 (stereo view)

對於大部分的研究而言，此一步驟並非必要。在沒有立體圖的情形下，旋轉(Rotation)亦有助於了解3D結構的關係。然而，假如你常常看分子圖形，學習如何看立體圖將有助於你對於複合體空間結構的研究。

此外，大家可以利用Google Search去搜尋"stereo viewing"，可找到許多相關資訊參考。



Colab-OpenSource-PyMOLpySnips

<https://github.com/MooersLab/colabOpenSourcePyMOLpySnips>

colabOpenSourcePyMOLpySnips

☞ Library of PyMOL Python snippets for installing and running
Open Source PyMOL on Google Colab via a Colab notebook

The installPyMOL snippet will install PyMOL with one click of a button.



<https://colab.research.google.com/github/Mooerslab/colabOpenSourcePyMOLpySnips/blob/main/colabOpenSourcePyMOLpySnips02.ipynb>

Check protein data file from CoLab

Upload

02_pymol on colab.ipynb

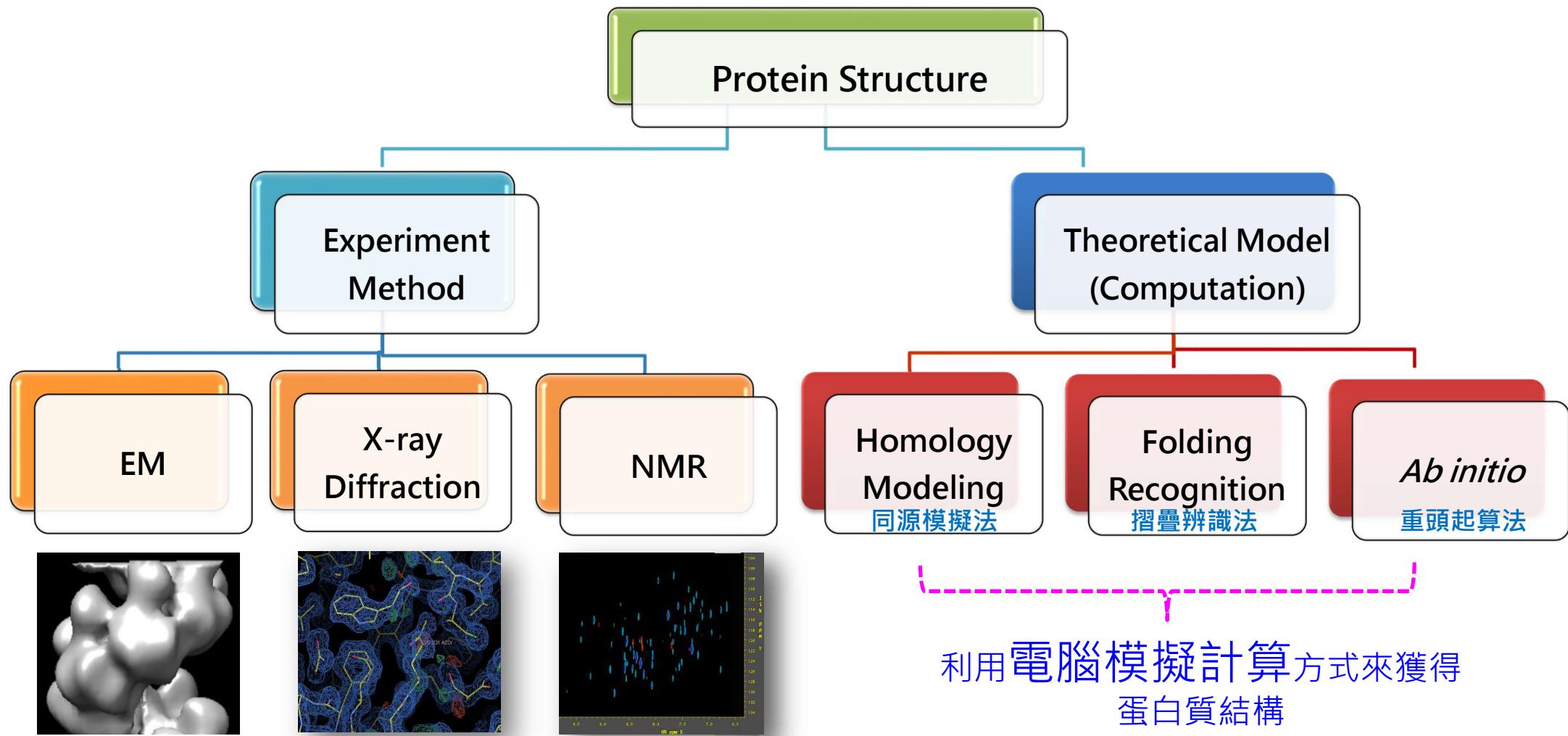
To

Colab Notebooks 

Protein Structure Modeling and Prediction

蛋白質結構模擬與預測

Solutions for Protein Structure



目前主要解析蛋白質結構的實驗方法

Molecular Type	X-ray ↓	NMR ↑	EM ↓↑	Multiple methods ↑↓	Neutron ↑↓	Other ↓↑	Total ↓↑
Protein (only)	147998	11984	7772	187	72	32	168045
Protein/Oligosaccharide	8724	31	1372	5	0	0	10132
Protein/NA	7812	277	2461	3	0	0	10553
Nucleic acid (only)	2453	1416	62	12	2	1	3946
Other	154	31	5	0	0	0	190
Oligosaccharide (only)	11	6	0	1	0	4	22
Total	167152	13745	11672	208	74	37	192888

X-ray: X-ray crystallography X-光晶體學

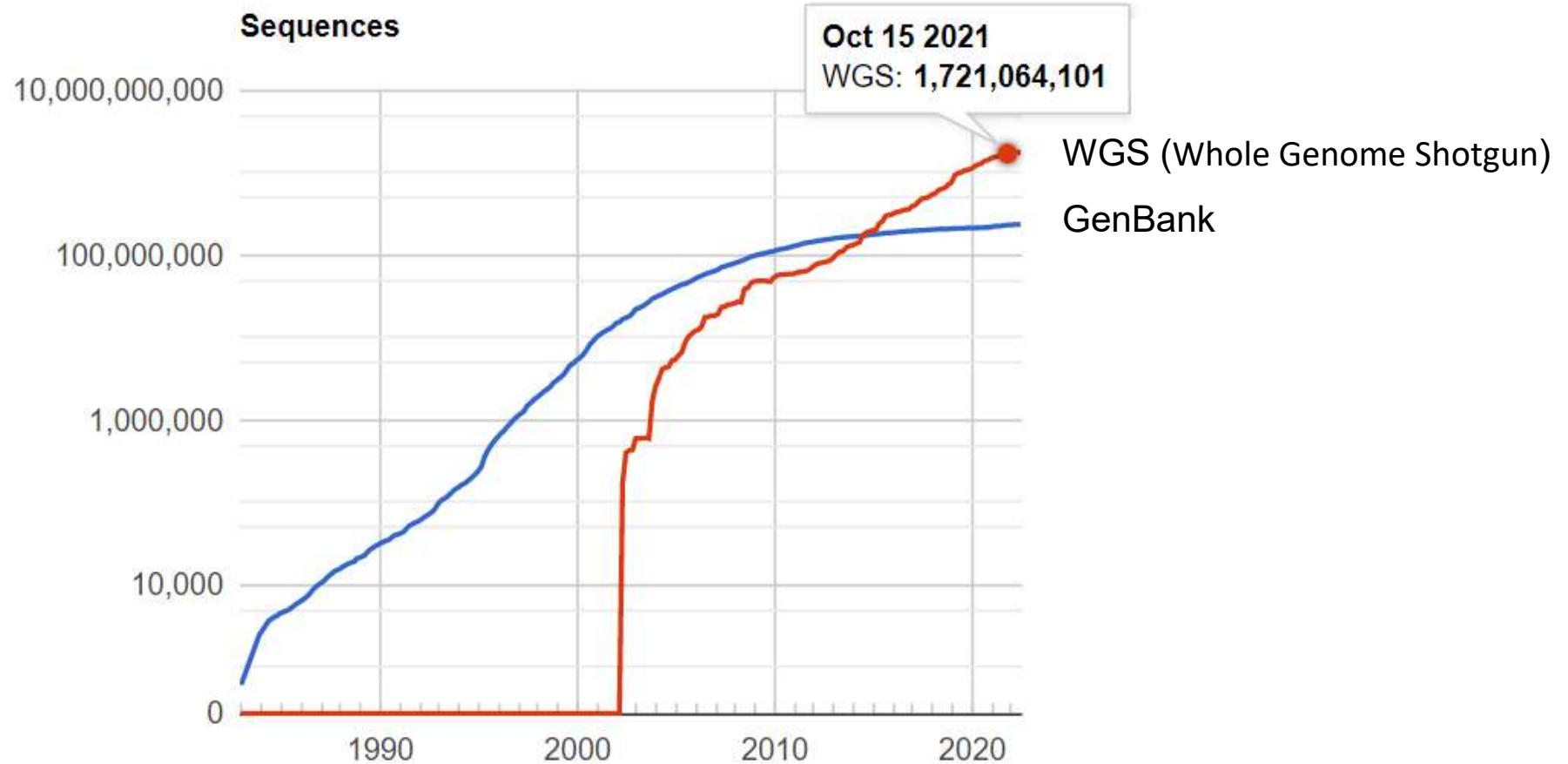
(20220716 updated)

NMR: Nuclear magnetic resonance 核磁共振

cryo-EM: cryo-electron microscopy 冷凍電顯

<https://www.rcsb.org/stats/summary>

GenBank and WGS Statistics



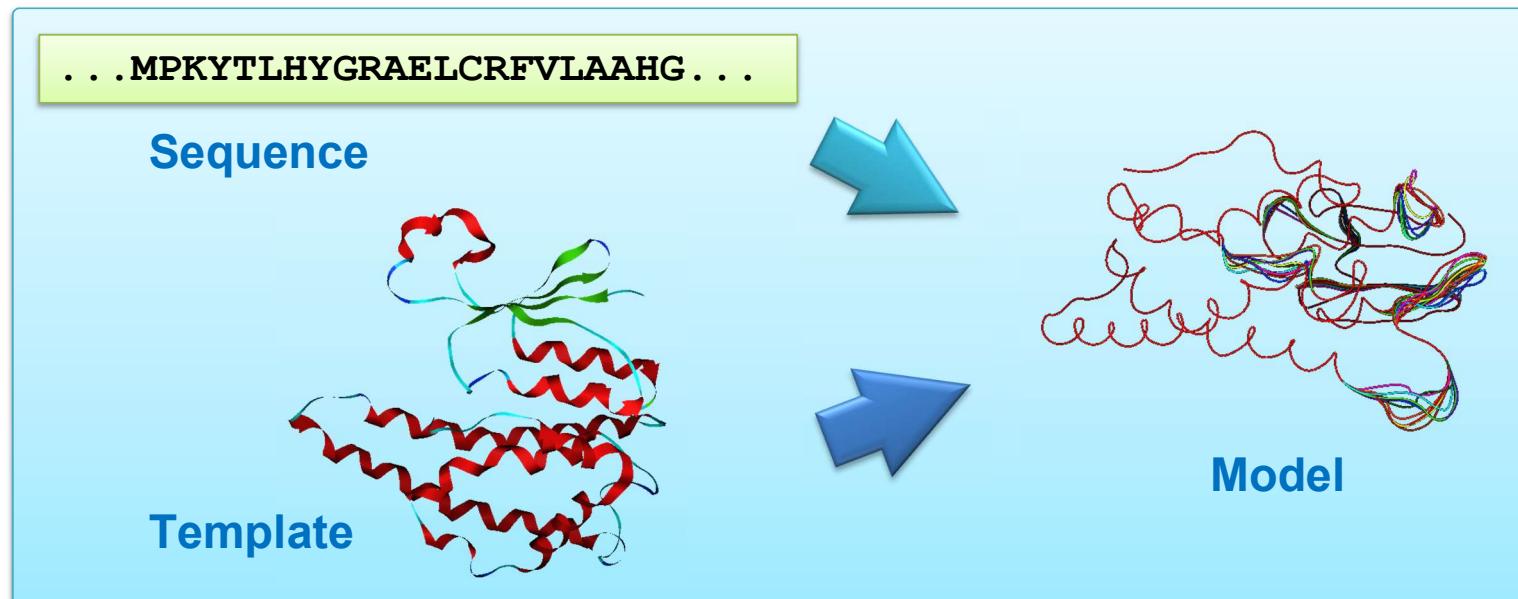
<https://www.ncbi.nlm.nih.gov/genbank/statistics/>

Theoretical Model

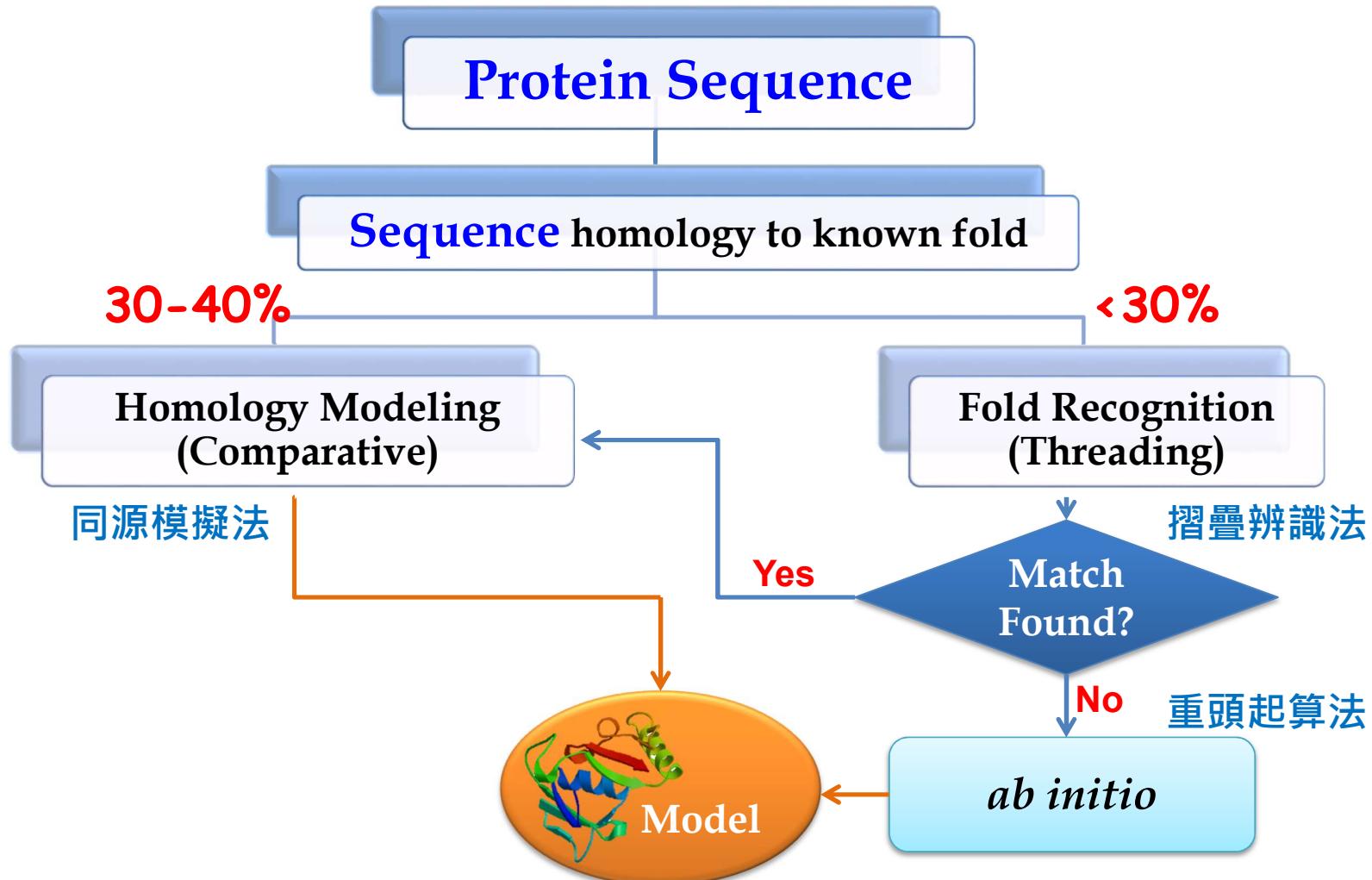
- 由於目前已知資料庫中，完整的蛋白質序列數量已遠遠超過所解出的蛋白質結構數量，而且利用實驗方法解出蛋白質結構不是太耗時間，就是受到一些限制而無法定出結構，於是，科學家想到利用電腦模擬計算方式來獲得蛋白質結構。
- 結構模擬大致可分為下列三種方式：
 - Homology or comparative modeling
 - Fold recognition (threading)
 - *ab initio* techniques.
- Homology or comparative modeling is currently the most accurate method to predict the three-dimensional structure of proteins.

Homology Modeling (同源模擬法)

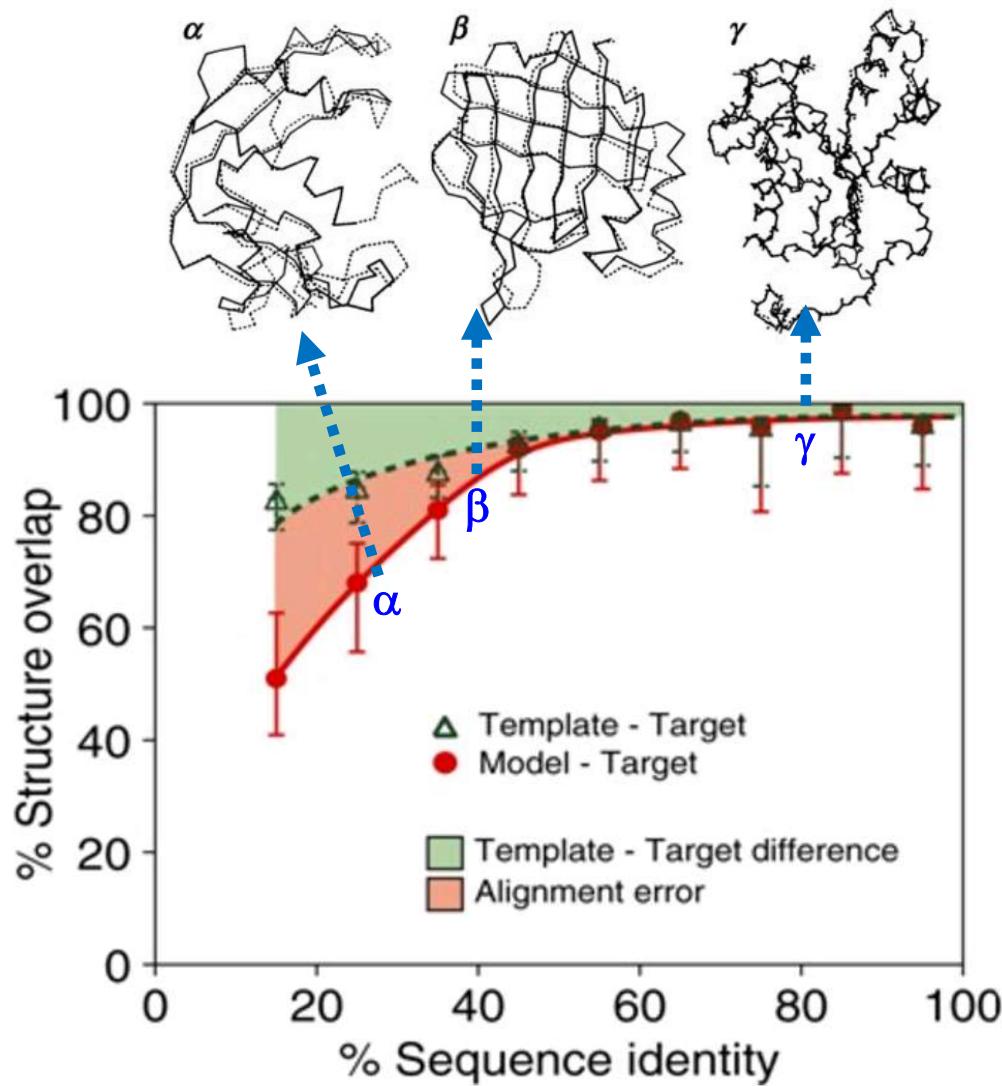
- 同源模擬法的基本假設是特定的胺基酸序列會構成特定的蛋白質結構。
- 主要是利用現存已解出的結構為模板(由NMR或是X-ray diffraction所解出的結構)，模擬出未知結構蛋白質序列(protein sequence)的三度空間結構。
- 一般來說，當所要解的蛋白質序列(target protein)和模板(template)之間的序列相似度愈高，所模擬出來的結構愈正確也愈可信。



Protein Structure Prediction



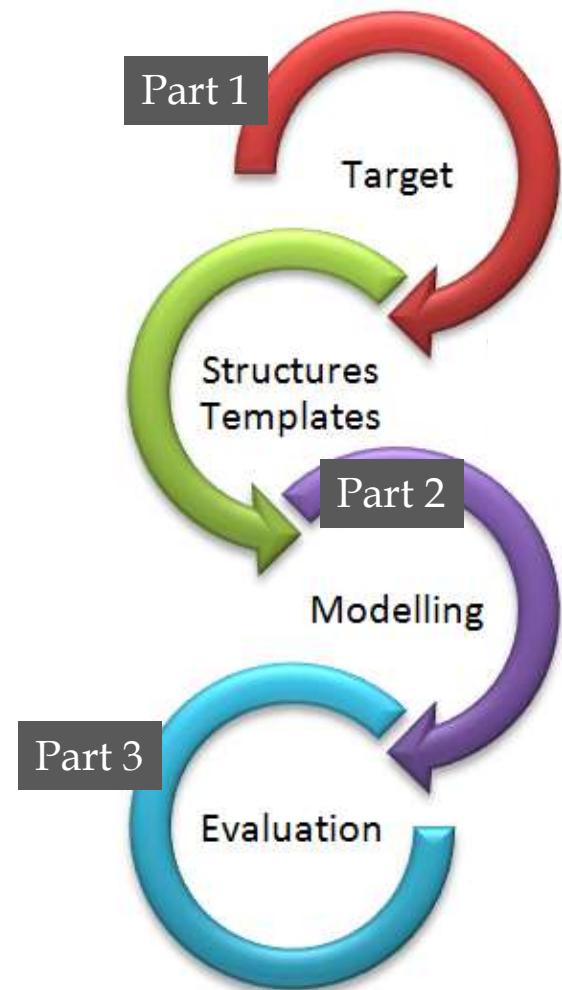
Model Evaluation



The accuracy of the model depends on its **sequence identity** with the template

In general, **30% sequence identity** is required for generating useful models.

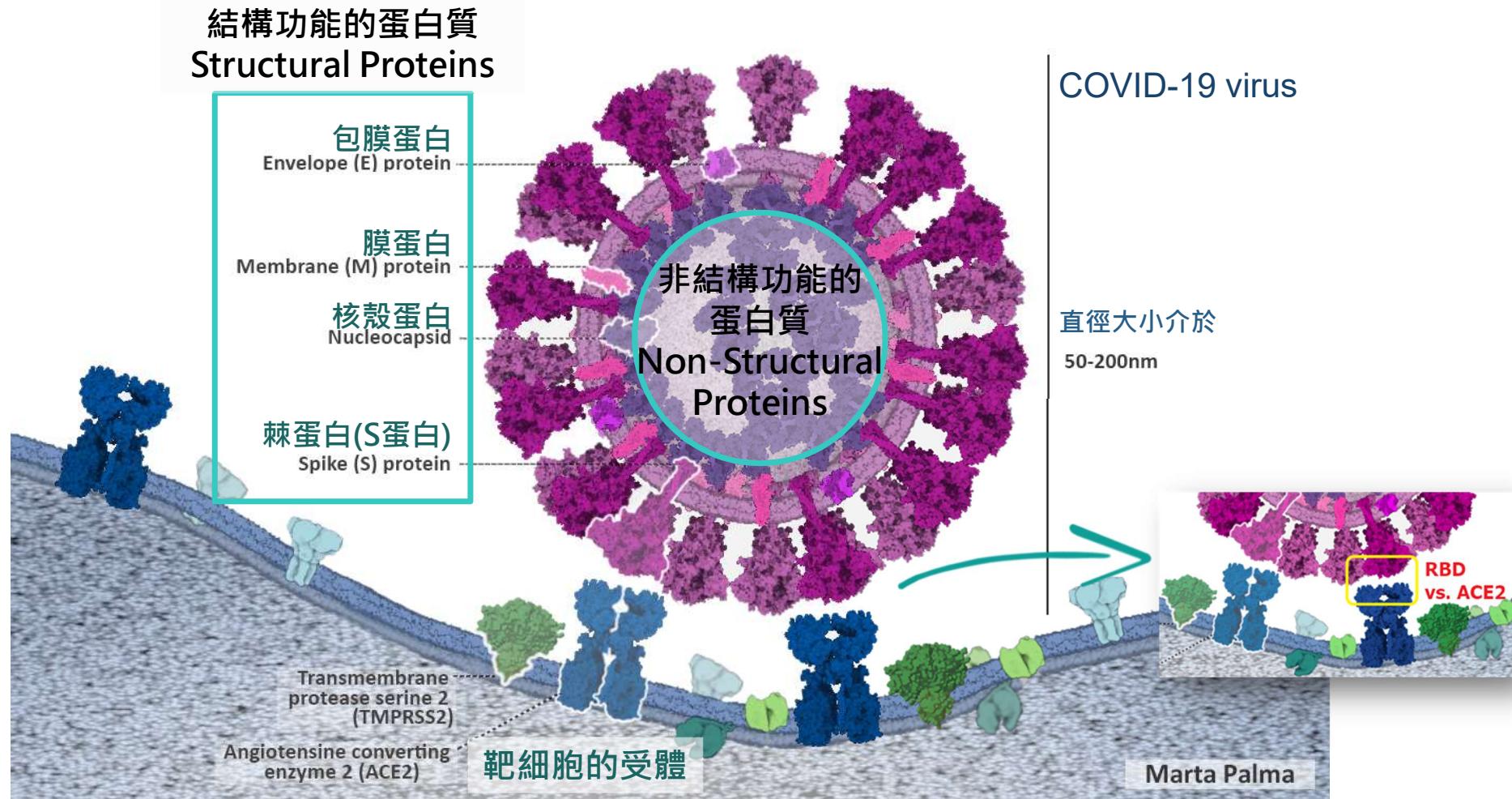
Model Building Steps



- Target sequence (primary sequence)
- Searching for structures of templates in databases
- Construction of the model
- Evaluation of model quality

Start Your Modeling

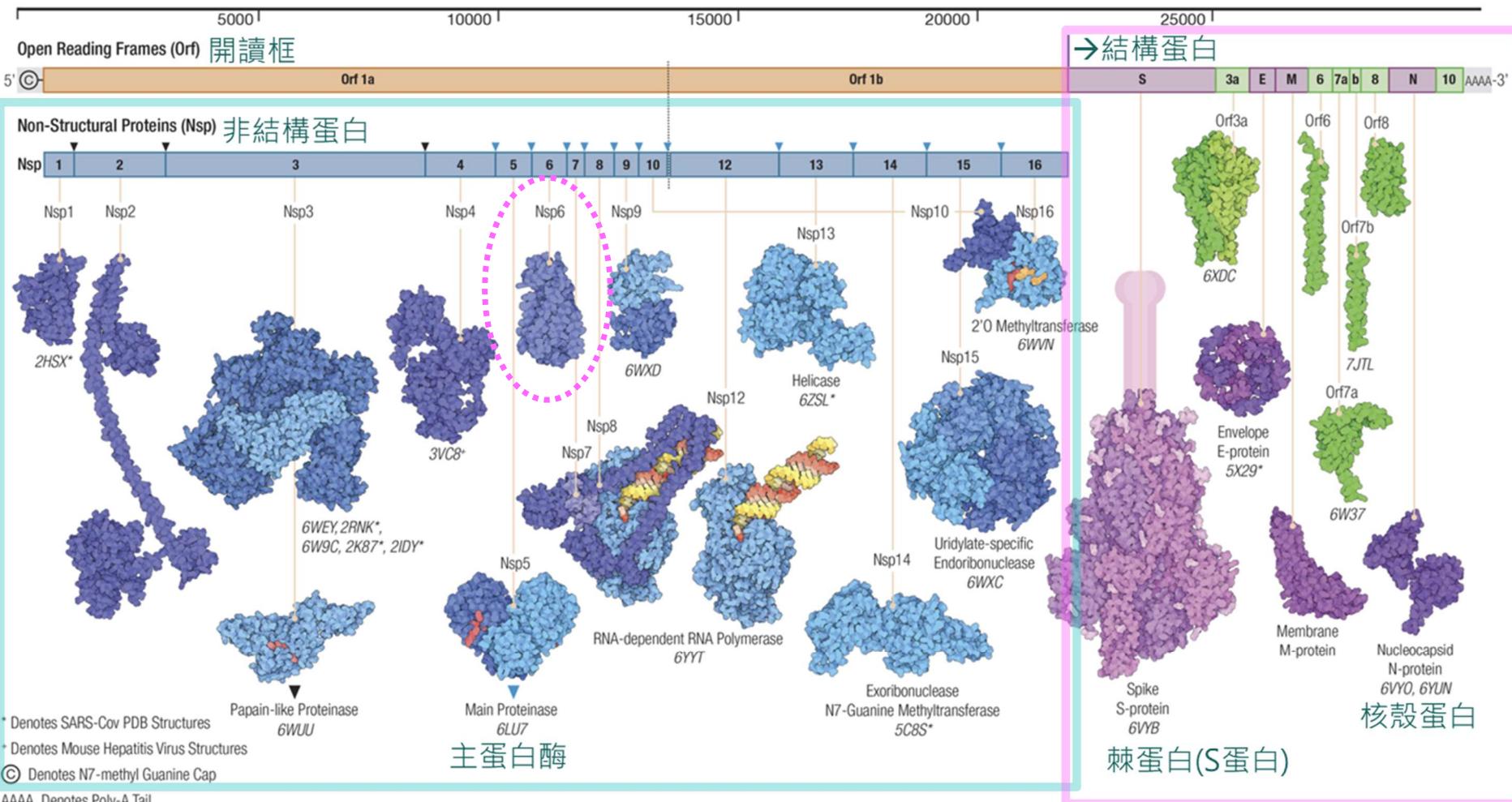
SARS-CoV-2與靶細胞的表面蛋白相互作用



<https://pdb101.rcsb.org/news/2020#5ed7c17cdab5c9354c274f56>

SARS-CoV-2的基因組與蛋白質結構組成

Genome Position 基因組圖



NCBI SARS-CoV-2 Resources

SARS-CoV-2 next-generation sequencing runs in **SRA**

[View in SRA](#)

Coronaviridae family-containing **SRA** runs

[Download from SRA Run Selector](#)

SARS-CoV-2 protein structures, domains, and sequences available through
NCBI Structure

[View Structure Data](#)

SARS-CoV-2 related compounds, substances, pathways, bioassays, and more
in **PubChem**

[View in PubChem](#)

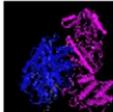
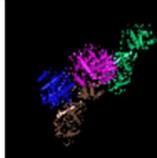
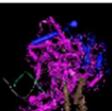
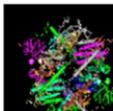
Genome expression studies related to SARS-CoV-2 in **GEO**

[View in GEO](#)

<https://www.ncbi.nlm.nih.gov/sars-cov-2/>

Select modeling target : Nsp6

SARS-CoV-2-related data provided by the **Protein Domains** resource

ORF name	CDD model	Model name (mouse-over name for description)	SARS Cov-2 structure	SARSr-CoV structure	Other coronavirus structure	RefSeq protein	Length
Nsp5	cd21666	betaCoV_Nsp5_Mpro	 7BQY [all]	 1WOF [all]	 3D23 [all]	YP_009725301 [domain architecture]	306
Nsp6	cd21560	betaCoV-Nsp6	<i>No structure data available!</i>				YP_009725302 [domain architecture]
Nsp7	cd21827	betaCoV_Nsp7	 7BV2 [all]	 2AHM [all]	-	YP_009725303 [domain architecture]	83

<https://www.ncbi.nlm.nih.gov/Structure/SARS-CoV-2.html>

Sequence of Nsp6

Key words: Nsp6 SARS-CoV-2

The screenshot shows the National Library of Medicine's search interface. At the top left is the NIH logo and the text "National Library of Medicine" with "National Center for Biotechnology Information" below it. On the right is a "Log in" button. Below this is a search bar with the word "Protein" in a dropdown menu, a search input field containing "Nsp6 SARS-CoV-2", a clear button (x), a "Search" button, and a "Help" link. The main content area displays the protein sequence.

```
>nsp6
SAVKRTIKGTHHWLLLTIITSLLVQSTQWSLFFFYENAFILPFAMGI IAMSAFAMMFVKHKHAFLCLF
LLPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLIILMTARTVYDDGARRVW
TLMNVLTLYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLARGIVFCVEYCPIFFITGNTLQCI
MLVYCFILGYFCTCYFGLFCLLNRYFRLTLGVYDYLVSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGG
KPCIKVATVQ
```

https://www.ncbi.nlm.nih.gov/protein/YP_009742613.1

Model by SWISS-MODEL



SWISS-MODEL

Modelling Repository Tools Documentation Log in Create Account

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.

Start Modelling



Repository

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



<http://swissmodel.expasy.org/>

Start a New Modelling Project



SWISS-MODEL

Modelling Repository Tools

Start a New Modelling Project

Target Sequence(s):
(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

Target	SAVKRTIKGTHHHWLLLTILTSLLVLVQSTQWSLEFFFLYENAFLPFAMGLIAMSASFAMMFVKHKHAFLCLFLPLSATVAY	80
Target	FNMVYMPASWVMRIMTWLDMVDTLSGFKLKDCVMYASAVVLLILMTARTVYDDGARRVWTLMNVTLVYKVYYGNALDQ	160
Target	AISMWALIISVTSNYSGVVTTVMFLARGIVFMCVEYCPPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLG	240
Target	VYDYLVSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQ	290

Add Hetero Target

Reset

Project Title:

nsp6

Email:

Optional



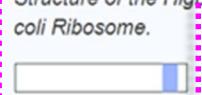
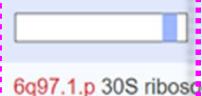
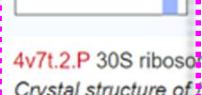
Search For Templates

Build Model

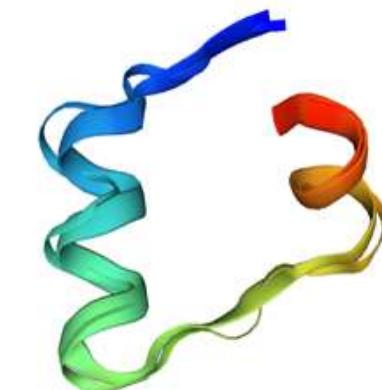
1. Copy & paste your target sequence;
2. Title: *nsp6* (optional)
3. Click “Search For Templates”

Model Results

Template Results

Templates	Quaternary Structure	Sequence Similarity	Alignment	More ▾
↑ Sort	♦ Coverage	♦ GMQE	♦ QSQE	♦ Identity
<input checked="" type="checkbox"/> 7nbu.1.P 30S ribosomal protein S16 Structure of the HigB1 toxin mutant K95A from <i>Mycobacterium tuberculosis</i> (Rv1955) and its target, the <i>cspA</i> mRNA, on the <i>E. coli</i> Ribosome.	 0.00	-	7.69	EM hetero-50-mer ▲ 1 x C-A, 21 x MG, 1 x MS6, 2 x ZN
<input checked="" type="checkbox"/> 5t7v.1.F 30S ribosomal protein S16 <i>Methicillin Resistant, Linezolid resistant Staphylococcus aureus</i> 70S ribosome (delta S145 uL3)	 0.00	-	12.00	EM hetero-38-mer ▲ None
<input type="checkbox"/> 6q97.1.p 30S ribosomal protein S16 Structure of tmRNA-SmpB bound in A site of <i>E. coli</i> 70S ribosome	 0.00	-	7.69	EM hetero-53-mer ▲ 1 x A-U-G-G-U-C, 24 x MG, 2 x ZN
<input type="checkbox"/> 4v5h.1.P 30S RIBOSOMAL PROTEIN S16 <i>E.Coli</i> 70s Ribosome Stalled During Translation Of Trec Leader Peptide.	 0.00	-	8.00	EM, 5.8Å hetero-51-mer ▲ None
<input type="checkbox"/> 5li0.1.P 30S ribosomal protein S16 70S ribosome from <i>Staphylococcus aureus</i>	 0.00	-	11.54	EM hetero-47-mer ▲ 1 x O, 1 x ZN
<input type="checkbox"/> 4v7t.2.P 30S ribosomal protein S16 Crystal structure of the <i>E. coli</i> ribosome bound to chloramphenicol.				

Build Models 2



7nbu.1.P

5t7v.1.F

AlphaFold2 應用實例

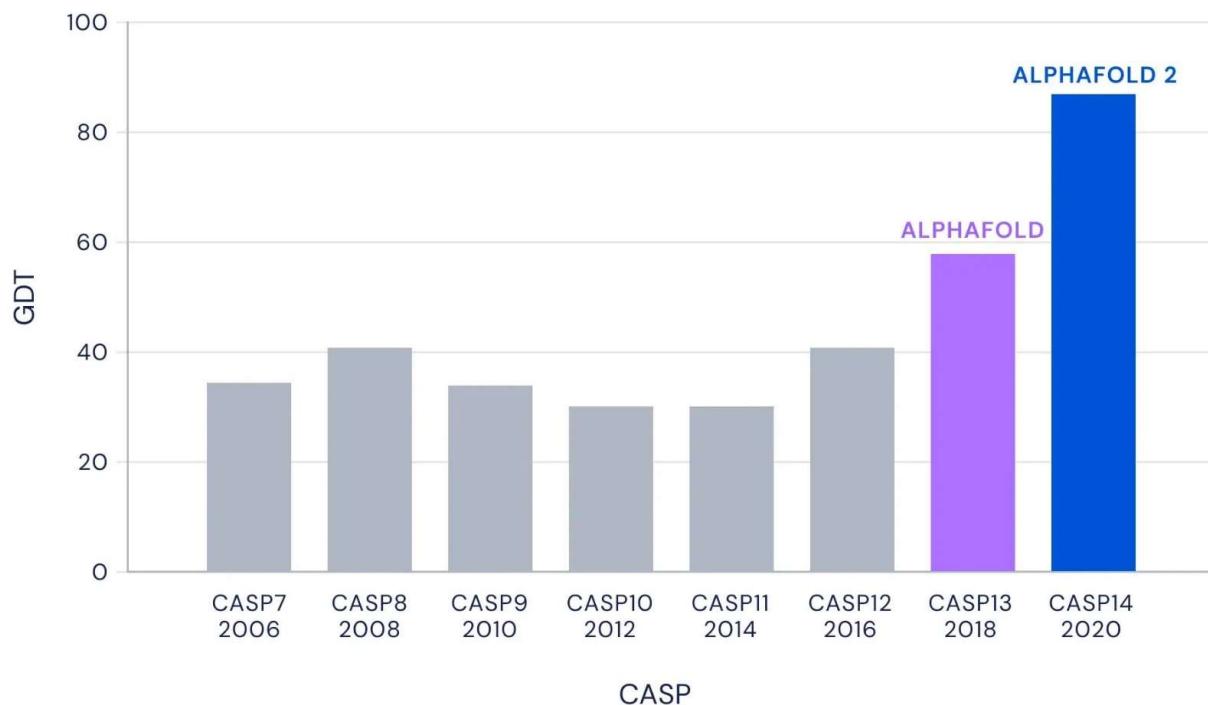
AlphaFold2 應用實例

DeepMind/ AlphaFold

- ❖ AlphaFold是Alphabet/Google旗下DeepMind開發的一款進行蛋白質結構預測的人工智慧程式，AlphaFold有兩個主要版本，AlphaFold 1 (2018)在2018年12月第13屆蛋白質結構預測技術的關鍵測試（**Critical Assessment of protein Structure Prediction**，CASP）的總體排名中名列第一，該程序成功預測了被競賽組織者評為最困難目標的最準確結構，其中沒有來自具有部分相似序列蛋白質的現有模板結構。
- ❖ 在2020年11月的第14屆CASP競賽中，AlphaFold 2 (2020)表現良好，中位分數為92.4（滿分100分），準確度遠遠高於其他任何程式。

AlphaFold in CASP

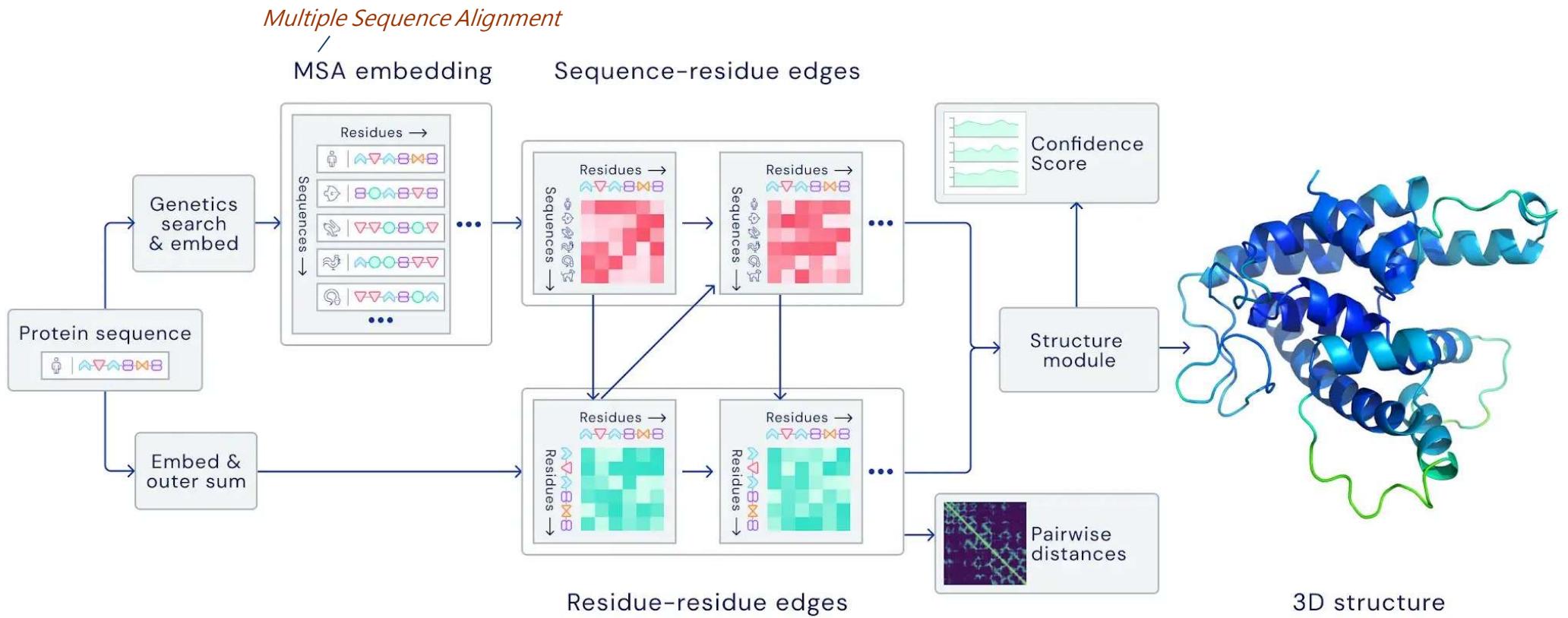
Median Free-Modelling Accuracy



- The main metric used by CASP to measure the accuracy of predictions is the Global Distance Test (GDT) which ranges from 0-100.

- Improvements in the median accuracy of predictions in the free modelling category for the best team in each CASP, measured as best-of-5 GDT.

<https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>



An overview of the main neural network model architecture.

AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism

BETA

Search

Examples: Free fatty acid receptor 2

At1g58602

Q5VSL9

E. coli

Help:

AlphaFold DB search help

Feedback on structure: Contact DeepMind

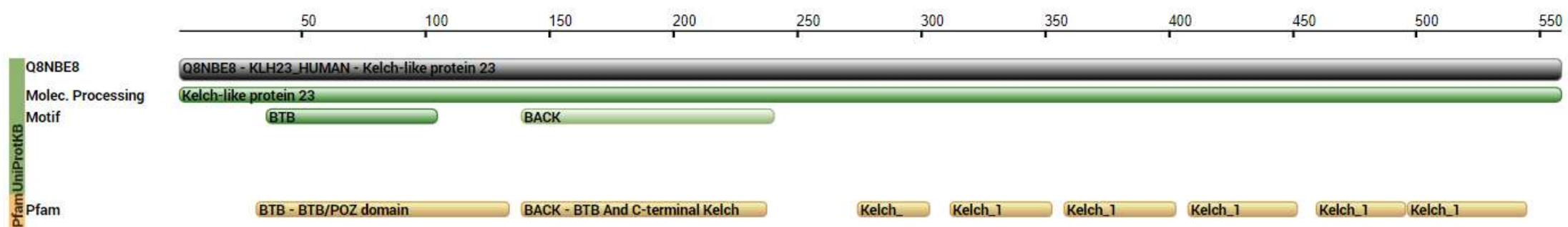
<https://alphafold.ebi.ac.uk/>

AlphaFold Protein Structure Database

- There are **995,411 structures** available on the AlphaFold DB website.
- AlphaFold DB currently provides predicted structures for the **48 organisms**:
 - Model organism proteomes
 - Global health proteomes
 - Swiss-Prot

Length:558
Mass (Da):63,923

KLHL23



KLHL23 structure??

UniProtKB - Q8NBE8 (KLH23_HUMAN)

<https://www.uniprot.org/uniprot/Q8NBE8>

Related template structures of KLHL23



Structureⁱ

KLHL23: Kelch-like protein 23

Model Confidence:

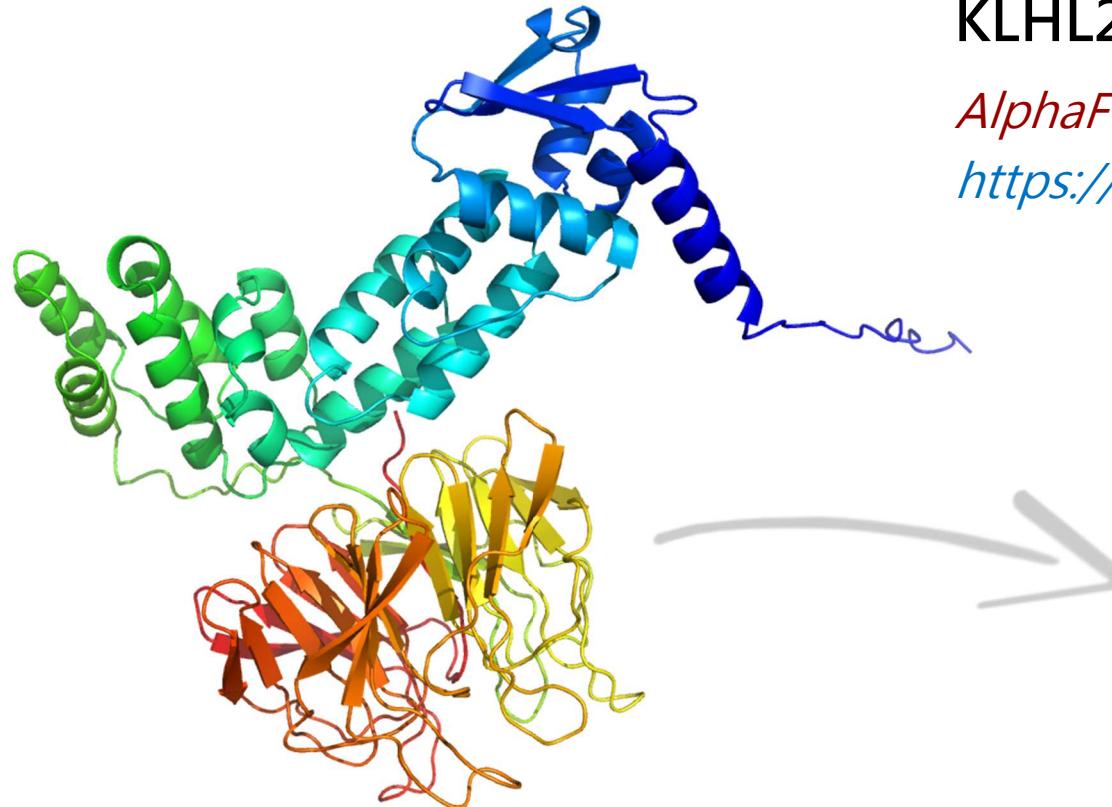
- █ Very high ($p\text{LDDT} > 90$)
- █ Confident ($90 > p\text{LDDT} > 70$)
- █ Low ($70 > p\text{LDDT} > 50$)
- █ Very low ($p\text{LDDT} < 50$)

AlphaFold produces a per-residue confidence score ($p\text{LDDT}$) between 0 and 100. Some regions with low $p\text{LDDT}$ may be unstructured in isolation.

pLDDT corresponds to the model's prediction of its score on the **local Distance Difference Test**



<https://www.uniprot.org/uniprot/Q8NBE8>



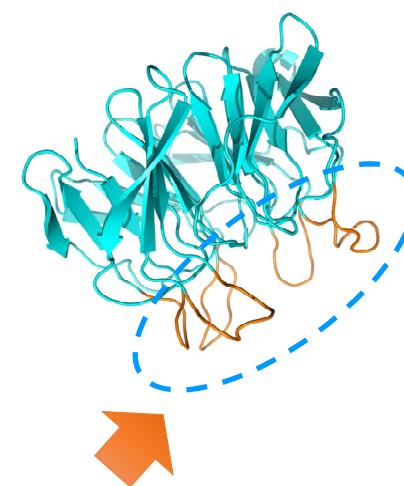
KLHL23

AlphaFold structure prediction

<https://alphafold.ebi.ac.uk/entry/Q8NBE8>

- Potential **ligand** interface residues:

281-286:B,328-336:B,377-383:B,424-433:B,516-523:B



ColabFold

ColabFold

Making protein folding accessible to all

Nature Methods, 2022. 19:679–682

ColabFold

- ❖ ColabFold offers accelerated prediction of protein structures and complexes by combining the fast homology search of MMseqs2 with AlphaFold2 or RoseTTAFold.
Many-against-Many sequence searching
多對多序列搜尋
- ❖ ColabFold's **40–60-fold faster** search and optimized model utilization enables prediction of close to 1,000 structures per day on a server with one graphics processing unit.
- ❖ Coupled with **Google Colaboratory**, ColabFold becomes a free and accessible platform for protein folding.
- ❖ ColabFold is open-source software available at <https://github.com/sokrypton/ColabFold> and its novel environmental databases are available at <https://colabfold.mmseqs.com>.

ColabFold

sokrypton / ColabFold Public

	.gitignore	gitignore pytesttemp	7 months ago
	AlphaFold2.ipynb	Update news that server migration and db update is comp...	4 days ago
	AlphaFold2_complexes.ipynb	Update license from CC-BY-NC to CC-BY of AlphaFold2_co...	6 months ago
	Contributing.md	Update Contributing.md	8 months ago
	LICENSE	Create LICENSE	12 months ago
	README.md	Update README.md	4 days ago
	RoseTTAFold.ipynb	adding link to pyrosetta version	2 months ago
	colabfold_search.sh	Add message to colabfold_search.sh to avoid usage	5 days ago

<https://github.com/sokrypton/ColabFold>

ColabFold/AlphaFold2.ipynb

main ▾ ColabFold / AlphaFold2.ipynb Go to file ...

milot-mirdita Update news that server migration and db update is complete ✓ Latest commit ce2ce65 4 days ago History

6 contributors

546 lines (546 sloc) | 27.4 KB

Open in Colab

Open in Colab

ColabFold: AlphaFold2 using MMseqs2

Easy to use protein structure and complex prediction using AlphaFold2 and Alphafold2-multimer. Sequence alignments/templates are generated through MMseqs2 and HHsearch. For more details, see bottom of the notebook, checkout the ColabFold GitHub and read our manuscript. Old versions: v1.0, v1.1, v1.2, v1.3

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. *Nature Methods*, 2022



<https://github.com/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>



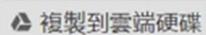
AlphaFold2.ipynb

檔案 編輯 檢視畫面 插入 執行階段 工具 說明

共用

+ 程式碼

+ 文字



ColabFold: AlphaFold2 using MMseqs2

Easy to use protein structure and complex prediction using [AlphaFold2](#) and [AlphaFold2-multimer](#). Sequence alignments/templates are generated through [MMseqs2](#) and [HHsearch](#). For more details, see [bottom](#) of the notebook, checkout the [ColabFold GitHub](#) and read our manuscript. Old versions: [v1.0](#), [v1.1](#), [v1.2](#), [v1.3](#)

[Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. Nature Methods, 2022](#)

- 2022/07/13: The move of the ColabFold MSA server to [KOBIC](#) is now complete. MSA generation was accelerated with multiple threads and databases were updated to UniRef30 2022_02 and PDB/PDB70 220313.

▶ Input protein sequence(s), then hit Runtime -> Run all

```
query_sequence: "PIAQIHLGRSDEQKETLIREVSEAIRSLDAPLTSVRVIITEMAKGHFGIGGELASK"
```

- Use : to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetero-oligomers). For example PI...SK:PI...SK for a homodimer

```
jobname: "test"
```

「AlphaFold2.ipynb」的副本

檔案 編輯 檢視畫面 插入 執行階段 工具 說明 已儲存所有變更

+ 程式碼 + 文字

ColabFold: AlphaFold

Easy to use protein structure alignments/templates are generated in ColabFold. Check out the [ColabFold GitHub repository](#). Mirdita M, Schütze K, Moriyama T, et al. ColabFold: A web server for protein structure prediction that is accessible to all. *Nature Methods*. 2022;19:111–116.

- 2022/07/13: The move to Google Cloud and databases were completed.

[1] Input protein sequence

```
query_sequence: "SAVKKRTIIGTIRHWWLLLTLISLLVLVQSTQWWSLFFFFLYENALPFAMGIIAMSASFAMMFVKHKHAFLCLF LLPSLATVAYFNMVYMPASWVMRIMTWLDMVDT:"
```

- Use : to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetero-oligomers). For example PI...SK:PI...SK for a homodimer

```
jobname: "nsp6"
```

```
use_amber: 
```

```
template_mode: none
```

Input protein sequence,
then hit **Runtime -> Run all**

留言 共用  

RAM 磁碟 



and [AlphaFold2-multimer](#). Sequence alignments/templates are generated in ColabFold. For more details, see [bottom](#) of the notebook. ColabFold has several versions: [v1.0](#), [v1.1](#), [v1.2](#), [v1.3](#).

[ColabFold: Making protein folding](#)

is now complete. MSA generation was accelerated with multiple threads B70 220313.

11

!!



+ 程式碼 + 文字

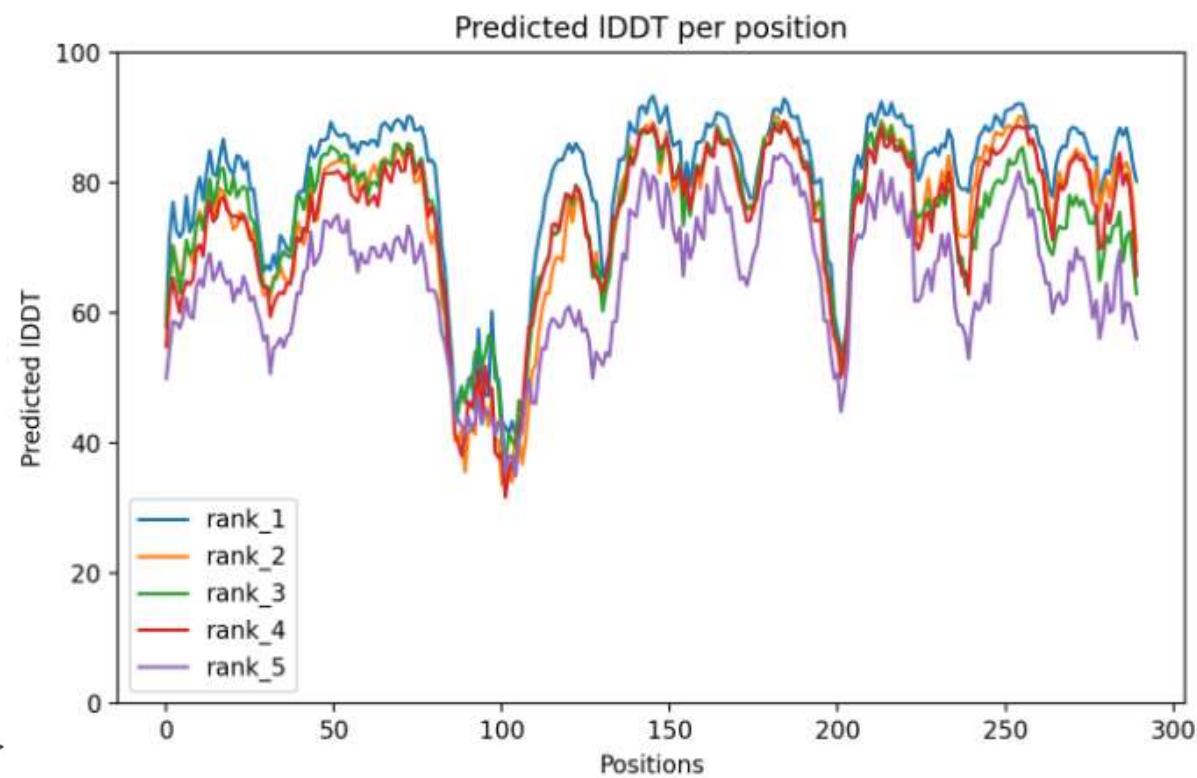
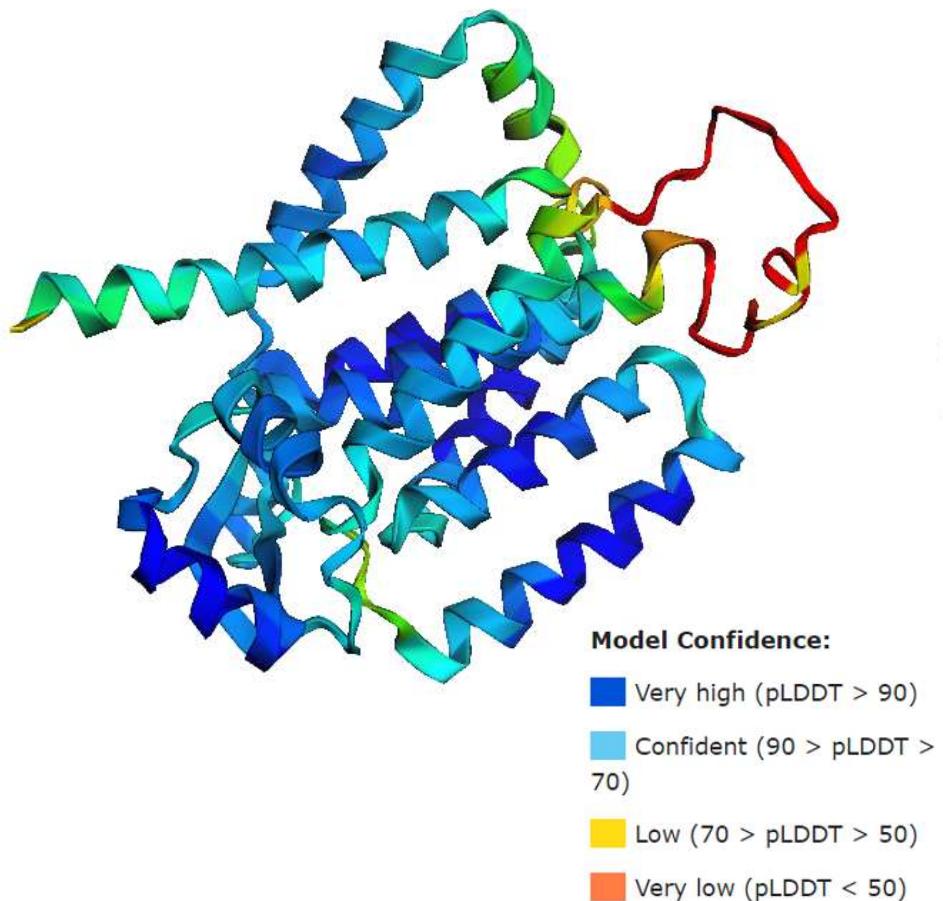


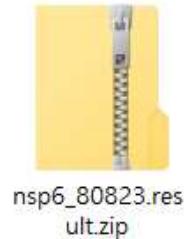
顯示程式碼

```
{x}   ↗ adding: config.json (deflated 48%)
      adding: nsp6_80823_predicted_aligned_error_v1.json (deflated 89%)
      adding: nsp6_80823_unrelaxed_rank_1_model_3_scores.json (deflated 70%)
      adding: nsp6_80823_unrelaxed_rank_2_model_5_scores.json (deflated 70%)
      adding: nsp6_80823_unrelaxed_rank_3_model_4_scores.json (deflated 69%)
      adding: nsp6_80823_unrelaxed_rank_4_model_1_scores.json (deflated 69%)
      adding: nsp6_80823_unrelaxed_rank_5_model_2_scores.json (deflated 70%)
      adding: nsp6_80823.a3m (deflated 75%)
      adding: nsp6_80823_relaxed_rank_1_model_3.pdb (deflated 78%)
      adding: nsp6_80823_relaxed_rank_2_model_5.pdb (deflated 78%)
      adding: nsp6_80823_relaxed_rank_3_model_4.pdb (deflated 78%)
      adding: nsp6_80823_relaxed_rank_4_model_1.pdb (deflated 78%)
      adding: nsp6_80823_relaxed_rank_5_model_2.pdb (deflated 78%)
      adding: nsp6_80823_unrelaxed_rank_1_model_3.pdb (deflated 78%)
      adding: nsp6_80823_unrelaxed_rank_2_model_5.pdb (deflated 78%)
      adding: nsp6_80823_unrelaxed_rank_3_model_4.pdb (deflated 78%)
      adding: nsp6_80823_unrelaxed_rank_4_model_1.pdb (deflated 78%)
      adding: nsp6_80823_unrelaxed_rank_5_model_2.pdb (deflated 78%)
      adding: cite.bibtex (deflated 52%)
      adding: nsp6_80823_coverage.png (deflated 8%)
      adding: nsp6_80823_PAE.png (deflated 1%)
      adding: nsp6_80823_plddt.png (deflated 4%)
2022-07-17 14:12:46,731 file_cache is unavailable when using oauth2client >= 4.0.0 or google-auth
```

nsp6-AlphaFold2.ipynb

nsp6_80823_relaxed_rank_1_model_3.pdb





名稱	修改日期	類型	大小
cite.bibtex	2022/7/17 下午 09:42	BIBTEX 檔案	4 KB
config.json	2022/7/17 下午 09:42	JSON 檔案	1 KB
nsp6_80823.a3m	2022/7/17 下午 09:43	A3M 檔案	93 KB
nsp6_80823_coverage.png	2022/7/17 下午 10:12	PNG 檔案	104 KB
nsp6_80823_PAE.png	2022/7/17 下午 10:12	PNG 檔案	587 KB
nsp6_80823_plddt.png	2022/7/17 下午 10:12	PNG 檔案	221 KB
nsp6_80823_predicted_aligned_error_v1.json	2022/7/17 下午 10:12	JSON 檔案	959 KB
nsp6_80823_relaxed_rank_1_model_3.pdb	2022/7/17 下午 10:12	Protein Data Bank...	373 KB
nsp6_80823_relaxed_rank_2_model_5.pdb	2022/7/17 下午 10:12	Protein Data Bank...	373 KB
nsp6_80823_relaxed_rank_3_model_4.pdb	2022/7/17 下午 10:12	Protein Data Bank...	373 KB
nsp6_80823_relaxed_rank_4_model_1.pdb	2022/7/17 下午 10:12	Protein Data Bank...	373 KB
nsp6_80823_relaxed_rank_5_model_2.pdb	2022/7/17 下午 10:12	Protein Data Bank...	373 KB
nsp6_80823_unrelaxed_rank_1_model_3.pdb	2022/7/17 下午 10:12	Protein Data Bank...	184 KB
nsp6_80823_unrelaxed_rank_1_model_3_scores.json	2022/7/17 下午 10:12	JSON 檔案	512 KB
nsp6_80823_unrelaxed_rank_2_model_5.pdb	2022/7/17 下午 10:12	Protein Data Bank...	184 KB
nsp6_80823_unrelaxed_rank_2_model_5_scores.json	2022/7/17 下午 10:12	JSON 檔案	517 KB
nsp6_80823_unrelaxed_rank_3_model_4.pdb	2022/7/17 下午 10:12	Protein Data Bank...	184 KB
nsp6_80823_unrelaxed_rank_3_model_4_scores.json	2022/7/17 下午 10:12	JSON 檔案	522 KB
nsp6_80823_unrelaxed_rank_4_model_1.pdb	2022/7/17 下午 10:12	Protein Data Bank...	184 KB
nsp6_80823_unrelaxed_rank_4_model_1_scores.json	2022/7/17 下午 10:12	JSON 檔案	521 KB
nsp6_80823_unrelaxed_rank_5_model_2.pdb	2022/7/17 下午 10:12	Protein Data Bank...	184 KB
nsp6_80823_unrelaxed_rank_5_model_2_scores.json	2022/7/17 下午 10:12	JSON 檔案	538 KB

Thank You

Yi-Chung Liu

2022/07/19