

# 基于结构的药物设计（一）

——蛋白质的结构与预测

# 上机课内容

## PBD数据库初探

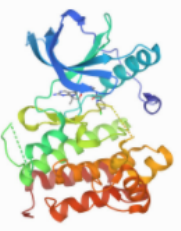
- 大分子PDB选择（选择分辨率高、有小分子配体的结构）
- 结构信息分析

## 同源模建构建蛋白质的三维结构

- 查找目标蛋白序列（Uniprot）
- Blast蛋白序列比对（>30%），找到相似蛋白模板（PDB库）
- 产生序列-结构对齐矩阵  
（align sequence to template）
- 蛋白建模
- 目标结构质量评估

# PDB数据库初探

大分子PDB选择（选择分辨率高、有小分子配体的结构）



3D View

**5JFX**

Crystal structure of TrkA in complex with PF-06273340

Jayasankar, J., Kurumbail, R., Skerratt, S., Brown, D.

(2016) J Med Chem **59**: 10084-10099

**Released** 2017-03-01


**Method** X-RAY DIFFRACTION 1.63 Å

**Organisms** [Homo sapiens](#)

**Macromolecule** [High affinity nerve growth factor receptor \(protein\)](#)

**Unique Ligands** 6K4

[Download File](#) [View File](#) ✓



3D View

**5JFW**

Crystal structure of TrkA in complex with PF-05247452

Jayasankar, J., Kurumbail, R., Skerratt, S., Brown, D.

(2016) J Med Chem **59**: 10084-10099

**Released** 2017-03-01

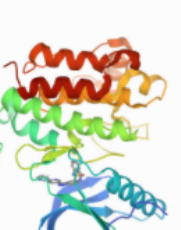
**Method** X-RAY DIFFRACTION 1.52 Å

**Organisms** [Homo sapiens](#)

**Macromolecule** [High affinity nerve growth factor receptor \(protein\)](#)

**Unique Ligands** 6K2

[Download File](#) [View File](#) ✓



3D View

**6J5L**

Crystal structure of Trk-A in complex with the Pan-Trk Kinase Inhibitor, compound 10e

Kensuke, A., Kazutaka, I.

(2019) Bioorg Med Chem Lett **29**: 2320-2326

**Released** 2019-07-17

**Method** X-RAY DIFFRACTION 2.3 Å

**Organisms** [Homo sapiens](#)

**Macromolecule** [High affinity nerve growth factor receptor \(protein\)](#)

**Unique Ligands** B9C

[Download File](#) [View File](#) ✓



# PDB数据库初探

## 结构信息分析

Biological Assembly 1 ?



3D View: Structure | Electron Density | Ligand Interaction

Global Symmetry: Asymmetric - C1 ⓘ  
Global Stoichiometry: Monomer - A1 ⓘ

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

**Macromolecule Content**

- Total Structure Weight: 35.53 kDa ⓘ
- Atom Count: 2706 ⓘ
- Residue Count: 277 ⓘ
- Unique protein chains: 1

**5JFW** PDB号

Crystal structure of TrkA in complex with PF-05247452  
DOI: 10.2210/pdb5JFW/pdb

Classification: **HYDROLASE/HYDROLASE INHIBITOR**  
Organism(s): **Homo sapiens** ← 种属  
Expression System: **Spodoptera frugiperda** ← 蛋白表达系统  
Mutation(s): No ⓘ ← 该蛋白有无突变

Deposited: 2016-04-19 Released: 2017-03-01  
Deposition Author(s): Jayasankar, J., Kurumbail, R., Skerratt, S., Brown, D.

**Experimental Data Snapshot**

Method: X-RAY DIFFRACTION  
Resolution: 1.52 Å  
R-Value Free: 0.209  
R-Value Work: 0.179  
R-Value Observed: 0.180

**wwPDB Validation** ⓘ

Metric	Percentile Ranks	Value
Rfree		0.210
Clashscore		4
Ramachandran outliers		0.7%
Sidechain outliers		4.3%
RSRZ outliers		13.0%

Worse | Better  
■ Percentile relative to all X-ray structures  
□ Percentile relative to X-ray structures of similar resolution

**Literature** Download Primary Citation ▾

The Discovery of a Potent, Selective, and Peripherally Restricted Pan-Trk Inhibitor (PF-06273340) for the Treatment of Pain.  
[Skerratt, S.E.](#), [Andrews, M.](#), [Bagal, S.K.](#), [Bilsland, J.](#), [Brown, D.](#), [Bungay, P.J.](#), [Cole, S.](#), [Gibson, K.R.](#), [Jones, R.](#), [Morao, I.](#), [Nedderman, A.](#), [Omoto, K.](#), [Robinson, C.](#), [Ryckmans, T.](#), [Skinner, K.](#), [Stuppel, P.](#), [Waldron, G.](#)  
(2016) J Med Chem **59**: 10084-10099  
PubMed: [27766865](#) Search on PubMed  
DOI: 10.1021/acs.jmedchem.6b00850  
Primary Citation of Related Structures:  
[5.IFS](#) [5.IFV](#) [5.IFX](#) [5.IFW](#)

该蛋白的实际信息

# PDB数据库初探

## 结构信息分析

Macromolecules

该结构所含大分子信息

Find similar proteins by: Sequence (by identity cutoff) | Structure

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image
High affinity nerve growth factor receptor	A	308	<a href="#">Homo sapiens</a>	Mutation(s): 0 Gene Names: <a href="#">NTRK1</a> , <a href="#">MT</a> , <a href="#">C</a> , <a href="#">TRK</a> , <a href="#">TRKA</a> EC: <a href="#">2.7.10.1</a>	

Find proteins for [P04629](#) ([Homo sapiens](#))

Explore [P04629](#)

Go to UniProtKB: [P04629](#)

NIH Common Fund Data Resources

PHAROS [P04629](#)

GTEX [ENSG00000198400](#)

Protein Feature View

Expand

Reference Sequence [5JFW\\_1](#)

PDB ENTITY SEQ 5JFW\_1

UNIPROT ALIGN P04629

UNMODELED 5JFW.A

ARTIFACT

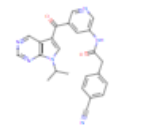
晶体结构中缺失的氨基酸残基及其位置

该结构所含小分子信息

点击此处进入“三维的相互作用视图”

Small Molecules

Ligands 1 Unique

ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
<a href="#">6K2</a> <a href="#">Query on 6K2</a> <a href="#">Download CCD File</a>	A	2-(4-cyanophenyl)-N-[5-[7-(propan-2-yl)-7H-pyrrolo[2,3-d]pyrimidine-5-carbonyl]pyridin-3-yl]acetamide C <sub>24</sub> H <sub>20</sub> N <sub>8</sub> O <sub>2</sub> <a href="#">ZJYCGJNMKTVBQP-UHFFFAOYSA-N</a>		<a href="#">Ligand Interaction</a>

External Ligand Annotations

ID	Binding Affinity (Sequence Identity %)
<a href="#">6K2</a>	<a href="#">IC50: 62 nM</a> <a href="#">Binding MOAD</a>

晶体结构所含小分子的活性值

小分子二维结构

复合物中的小分子名称

小分子所在链

小分子二维结构

# PDB数据库初探

## 结构信息分析

### Experimental Data & Validation

#### Experimental Data

Method: X-RAY DIFFRACTION  
Resolution: 1.52 Å  
R-Value Free: 0.209  
R-Value Work: 0.179  
R-Value Observed: 0.180  
Space Group: [P 6<sub>4</sub>](#)

#### Unit Cell:

Length ( Å )	Angle ( ° )
a = 124.332	α = 90
b = 124.332	β = 90
c = 46.194	γ = 120

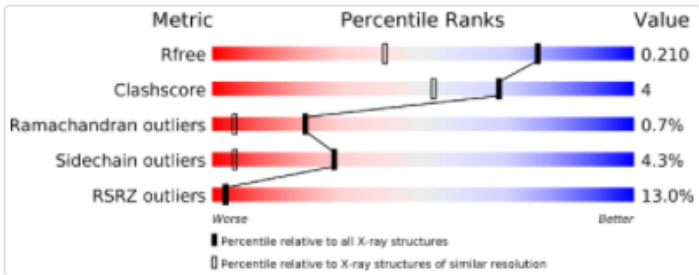
#### Software Package:

Software Name	Purpose
BUSTER	refinement
XDS	data reduction
SCALA	data scaling
PHASER	phasing

[View more in-depth experimental data](#)

#### Structure Validation

[View Full Validation Report](#)



该结构的晶体学数据

晶体解析软件

该结构的数据上传、修改历史记录

### Entry History

#### Deposition Data

Deposited Date: 2016-04-19  
Released Date: 2017-03-01  
Deposition Author(s): Jayasankar, J., Kurumbail, R., Skerratt, S., Brown, D.

#### Revision History

- Version 1.0: 2017-03-01  
Type: Initial release



# PDB数据库初探

## 结构信息分析

点击Ligand Interaction后，可预览配体-受体相互作用图，单击配体可现实/隐藏配体周围氨基酸残基。



显示电子密度图  
显示三维的相互作用图



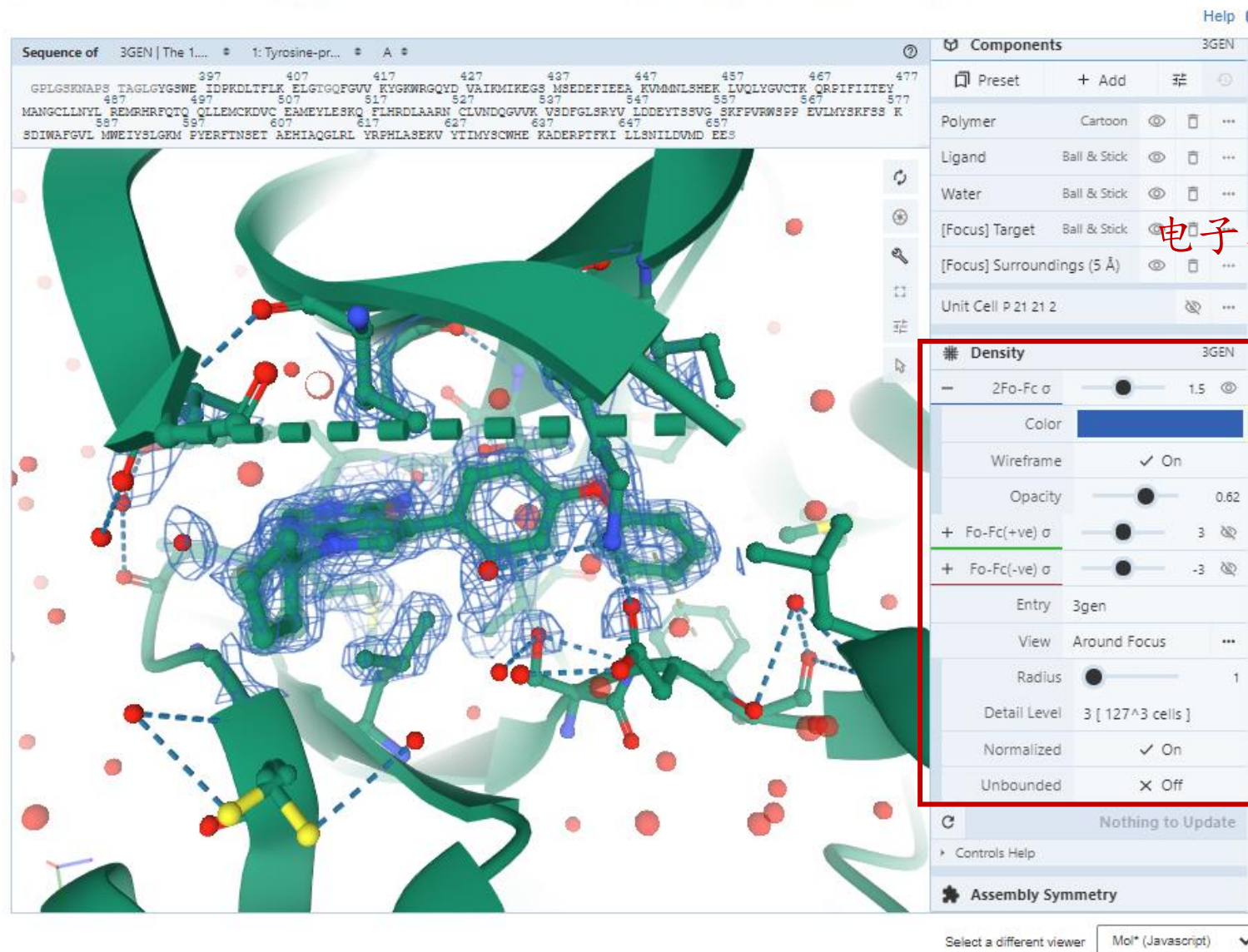
# PDB数据库初探

结构信息分析

开启蛋白晶体的电子  
密度预览模式

3GEN

The 1.6 Å crystal structure of human bruton's tyrosine kinase bound to a pyrrolopyrimidine-containing compound



电子密度图操作区



# 同源模建基本流程：（1）查找目标蛋白序列（P2RY6）

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

The screenshot shows the UniProt website homepage. The browser address bar displays <https://www.uniprot.org/>. The UniProt logo is on the left. The search bar contains 'P2RY6' and has a 'Search' button. Below the search bar, there are navigation links: BLAST, Align, Retrieve/ID mapping (highlighted), Peptide search, and SPARQL. A tooltip for 'Retrieve/ID mapping' says: 'Submit a list of identifiers to retrieve the corresponding UniProt entries, or to map them from or to an external database'. To the right are 'Help' and 'Contact' links. Below the navigation bar, the mission statement reads: 'The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.' The main content area features four large colored boxes: UniProtKB (blue), UniRef (orange), UniParc (pink), and Proteomes (red). UniProtKB is further divided into Swiss-Prot (563,552) and TrEMBL (195,104,019). Below these is a 'Supporting data' section with links to Literature citations, Taxonomy, and Subcellular locations. On the right, there is a 'News' section with a red banner for 'View SARS-CoV-2 Proteins and Receptors' and a 'Forthcoming changes' section stating 'There are currently no changes planned'. A 'UniProt release 2020\_05' section mentions 'PCK1 vacillating between gluconeogenesis and lipogenesis | Cross-references to CPTC, BMRB, PCDDDB and SASDBB'.

UniProtKB  
UniProt Knowledgebase

Swiss-Prot  
(563,552)

Manually annotated and reviewed.

Records with information extracted from literature and curator-evaluated computational analysis.

TrEMBL  
(195,104,019)

UniRef  
Sequence clusters

UniParc  
Sequence archive

Proteomes  
Proteome sets

Supporting data

Literature citations

Taxonomy

Subcellular locations

New UniProt portal for the latest SARS-CoV-2 coronavirus protein entries and receptors, updated independent of the general UniProt release cycle.

**View SARS-CoV-2 Proteins and Receptors**

News

[Forthcoming changes](#)

There are currently no changes planned

[UniProt release 2020\\_05](#)

PCK1 vacillating between gluconeogenesis and lipogenesis | Cross-references to CPTC, BMRB, PCDDDB and SASDBB

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

# 同源模建基本流程：（1）查找目标蛋白序列（P2RY6）

Q15077-1 [UniProt] [↓ FASTA](#) [Add to basket](#)

« Hide

Length: 328  
Mass (Da): 36,429  
Last modified: November 1, 1996 - v1  
Checksum: AAD6C55A43818107

BLAST [GO](#)

```
      10      20      30      40      50
MEWDNGTGQA LGLPPTTCVY RENFKQLLLP PVYSAVLAAG LPLNICVITQ
      60      70      80      90     100
ICTSRRALTR TAVYTLNLAL ADLLYACSLP LLIYNYAQGD HWPFGDFACR
     110     120     130     140     150
LVRFLFYANL HGSILFLTCI SFQRYLGICH PLAPWHKRGG RRAAWLVCVA
     160     170     180     190     200
VWLAVTTQCL PTAIFAATGI QRNRTVCYDL SPPALATHYM PYGMALTVIG
     210     220     230     240     250
FLLPFAALLA CYCLLACRLC RQDGPAEPVA QERRGKAARM AVVVAAAFAI
     260     270     280     290     300
SFLPFHITKT AYLAVRSTPG VPCTVLEAFA AAYKGTRPFA SANSVLDPIL
     310     320
FYFTQKKFRR RPHELLQKLT AKWQRQGR
```

文件名(N): Q15077.fasta.txt

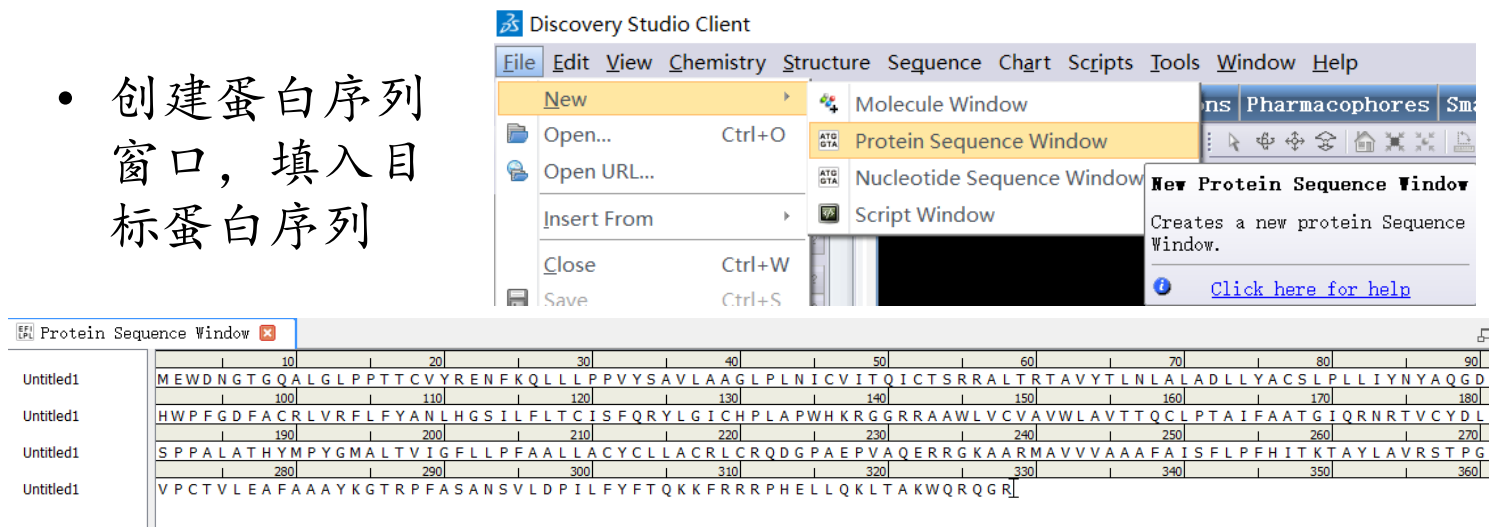
保存类型(T): TXT 文件 (\*.txt)

^ 隐藏文件夹

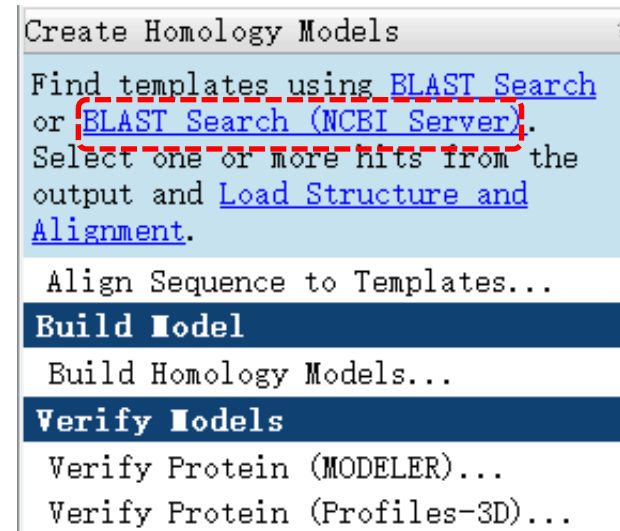
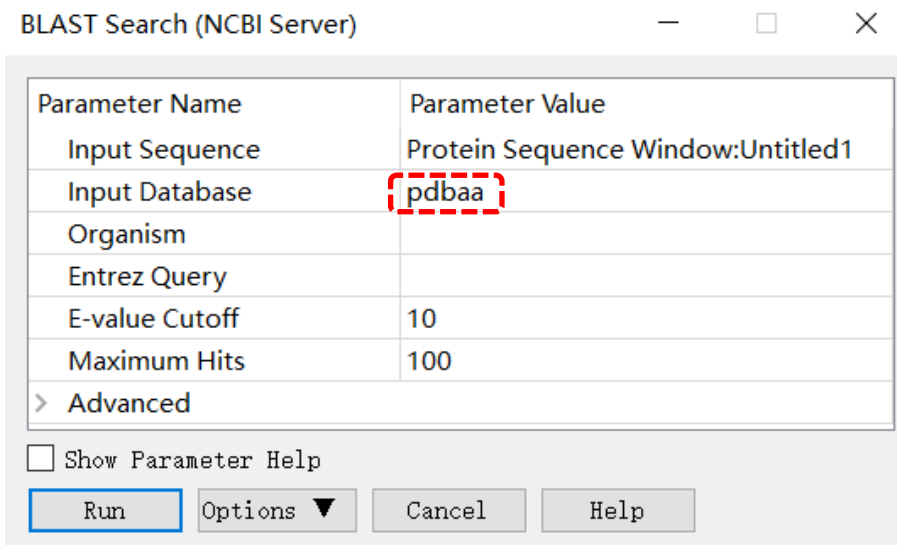
保存(S)

# 同源模建基本流程：（2）Blast蛋白序列比对

- 创建蛋白序列窗口，填入目标蛋白序列



- 使用蛋白晶体数据库找寻相似模板蛋白



# 同源模建基本流程：（2）Blast蛋白序列比对

The screenshot displays the BIOVIA Discovery Studio 2020 interface. A "BLAST Search (NCBI Server)" dialog box is open, showing the search query "sp\_Q15077\_P2RY6\_HUMAN" and the number of hits found (100). The main window shows a table of BLAST results with columns: Title/Description, Accession, Identity, Sequence Length, Alignment Length, Bit Score, E-value, and Posit. Below the table, there are three tabs: "Map View", "Table View", and "Text View". The "Text View" tab is selected, showing the sequence alignment of the query sequence (sp\_Q15077\_P2RY6\_HUMAN) against the top hit (The human P2Y...).

Title/Description	Accession	Identity	Sequence Length	Alignment Length	Bit Score	E-value	Posit
1 The human P2Y...	4XNV_A	39	421	182	110.923	4.98227e-27	52
2 Crystal struc...	5XSZ_A	35	477	191	99.7525	6.92559e-23	52
3 Crystal struc...	6IBB_A	30	342	287	93.9745	2.75504e-21	48
4 Crystal struc...	6RNK_A	29	342	287	92.4337	1.07362e-20	48
5 Crystal struc...	6RZ6_A	37	405	115	86.2705	2.25804e-18	53
6 Crystal struc...	3VW7_A	39	484	106	85.5001	6.54565e-18	51
7 Structure of ...	6D01_A	35	425	118	84.3445	1.19122e-17	54
8 Crystal Struc...	4ZUD_A	35	410	118	83.5741	2.22188e-17	54
9 XFEL structur...	4YAV_A	35	414	118	83.1889	2.54055e-17	54
10 Crystal struc...	5LWE_A	27	331	305	78.1814	8.91568e-16	43
11 Crystal struc...	6RZ4_A	36	423	111	78.5666	9.43064e-16	54
12 Cryo-EM struc...	6WWZ_R	36	529	105	77.7962	2.61796e-15	58
13 Chain A, Lyso...	5NDD_A	36	619	113	77.7962	2.84783e-15	52
14 Crystal struc...	5NJ6_A	36	437	113	77.0258	3.57267e-15	52
15 Synchrotron s...	4RWA_A	40	411	101	76.2554	5.58928e-15	55
16 1.8 Å Structu...	4N6H_A	40	414	101	76.2554	5.70734e-15	55
17 Crystal Struc...	6B73_A	37	418	106	76.2554	6.47652e-15	53
18 Structure of ...	4DJH_A	37	480	106	75.8702	9.05901e-15	53
19 Nanobody-Enab...	6VI4_A	37	307	106	74.7146	1.05973e-14	53
20 Structure of ...	4EJ4_A	40	461	101	75.485	1.178e-14	55
21 Crystal struc...	4DKL_A	39	464	101	73.9442	4.11064e-14	53
22 Chain A, Mu-t...	5C1M_A	39	296	101	72.7886	4.69223e-14	53
23 Chain A, C-X...	3OE6_A	34	508	114	73.559	5.35156e-14	54
24 Chain A, C-X...	3ODU_A	34	502	114	73.559	5.86376e-14	54
25 Chain A, C-X...	3OE0_A	34	499	114	73.559	6.57389e-14	54

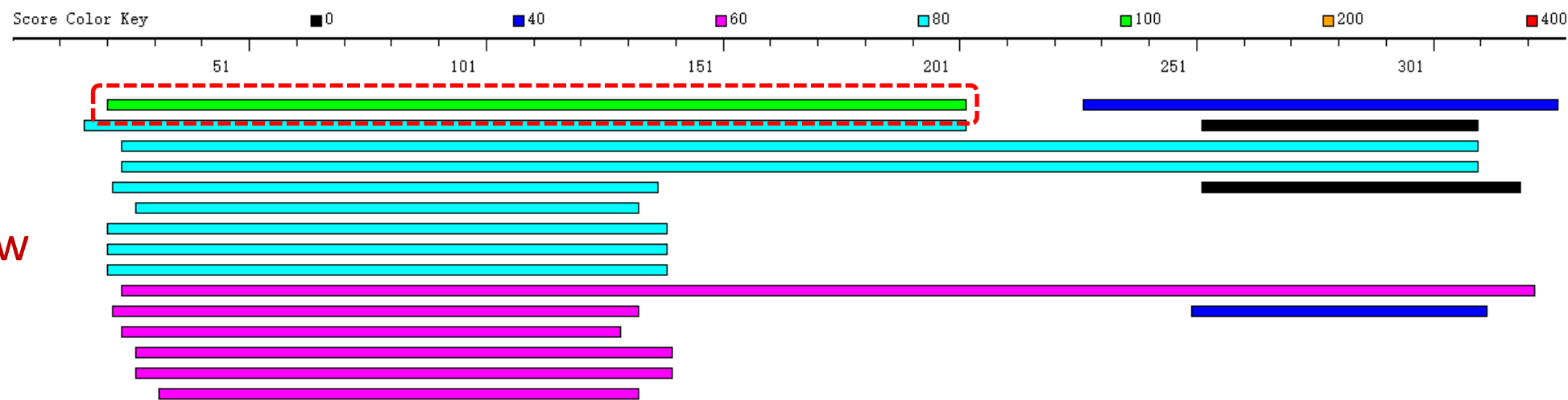
Sequence alignment view (Text View):

```
sp|Q15077|P2RY6_HUMAN MEWDNGTGOALGLPPTTCVYRENFKQLLLPPVYSAVLAAGLPINICVITQICTSRRALTRTAVYTLNLALADLLYACSLPLLIYNYAOGDHWPF GDFACRLVRFIFYANLHGSILFLTCISFORYLGI
sp|Q15077|P2RY6_HUMAN CHPLAPWHKRGGRRAAWLVCAVWLAVTTQCLPTAIFAATGIQRNRTVCYDLSPPALATHYMPYGMALTIVIGFLLPFAALLACYCLLACRLCRODGPAPVQAQERRGKAARMVAVVAFAISFLPFH
sp|Q15077|P2RY6_HUMAN ITKTAYLA VRSTPGVPCTVLEAF AAYKGRTPFASANSVLDPI LFYFTQKKFRRRPHELLQKLTAKWQRDGR
```

三种视图可随意切换

将名字改为Q15077等简单的中英文词，否则后面可能报错

# 同源模建基本流程：选择合适蛋白模板



Map View

	Title/Description	Accession	Identity	Sequence Length	Alignment Length	Bit Score	E-value	Positive
1	Chain A, P2Y purino...	4XNV_A	39	421	182	110.923	4.42019e-27	52
2	Chain A, Lysophosph...	5XSZ_A	35	477	191	99.7525	6.14428e-23	52
3	Chain A, Succinate ...	6IBB_A	30	342	287	93.9745	2.44423e-21	48
4	Chain A, Succinate ...	6RNK_A	29	342	287	92.4337	9.52501e-21	48
5	Chain A, Cysteiny...	6RZ6_A	37	405	115	86.2705	2.0033e-18	53
6	Chain A, Proteinase...	3VW7_A	39	484	106	85.5001	5.8072e-18	51
7	Chain A, Type-1 ang...	6D01_A	35	425	118	84.3445	1.05683e-17	54
8	Chain A, Chimera pr...	4ZUD_A	35	410	118	83.5741	1.97122e-17	54
9	Chain A, Soluble cy...	4YAY_A	35	414	118	83.1889	2.25394e-17	54
10	Chain A, C-C chemok...	5LWE_A	27	331	305	78.1814	7.90985e-16	43
11	Chain A, Cysteiny...	6RZ4_A	36	423	111	78.5666	8.36671e-16	54
12	Chain R, C-C chemok...	6WWZ_R	36	529	105	77.7962	2.32261e-15	58
13	Chain A, Lysozyme, P...	5NDD_A	36	619	113	77.7962	2.52655e-15	52
14	Chain A, Proteinase...	5NJ6_A	36	437	113	77.0258	3.16962e-15	52
15	Chain A, Soluble cy...	4RWA_A	40	411	101	76.2554	4.95872e-15	55
16	Chain A, Soluble cy...	4N6H_A	40	414	101	76.2554	5.06346e-15	55
17	Chain A, Soluble cy...	6B73_A	37	418	106	76.2554	5.74586e-15	53
18	Chain A, Kappa-type...	4DJH_A	37	480	106	75.8702	8.03701e-15	53
19	Chain A, Kappa opio...	6VI4_A	37	307	106	74.7146	9.40175e-15	53

- 选择合适蛋白模板并载入结构

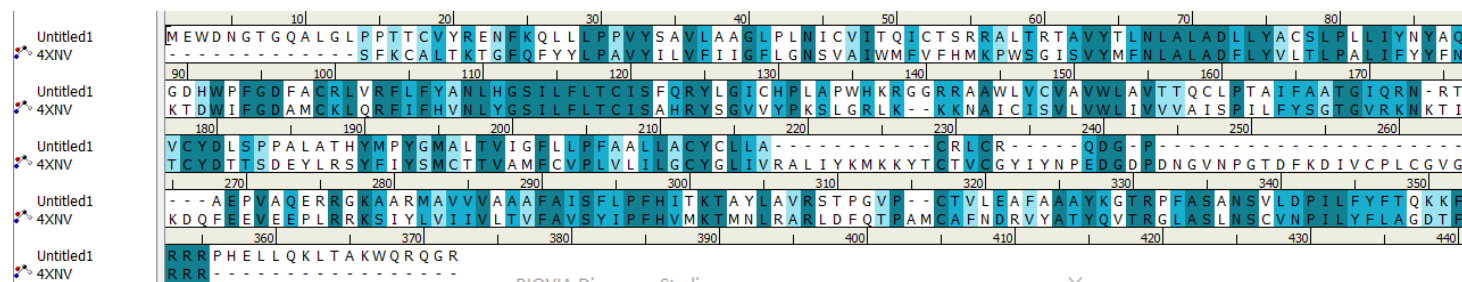
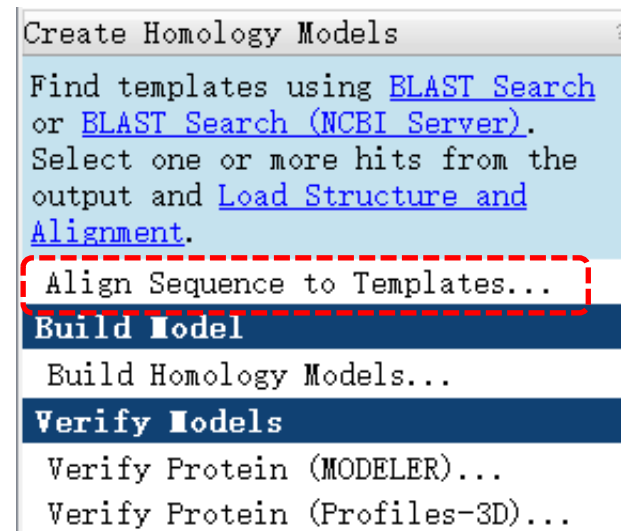
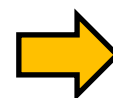
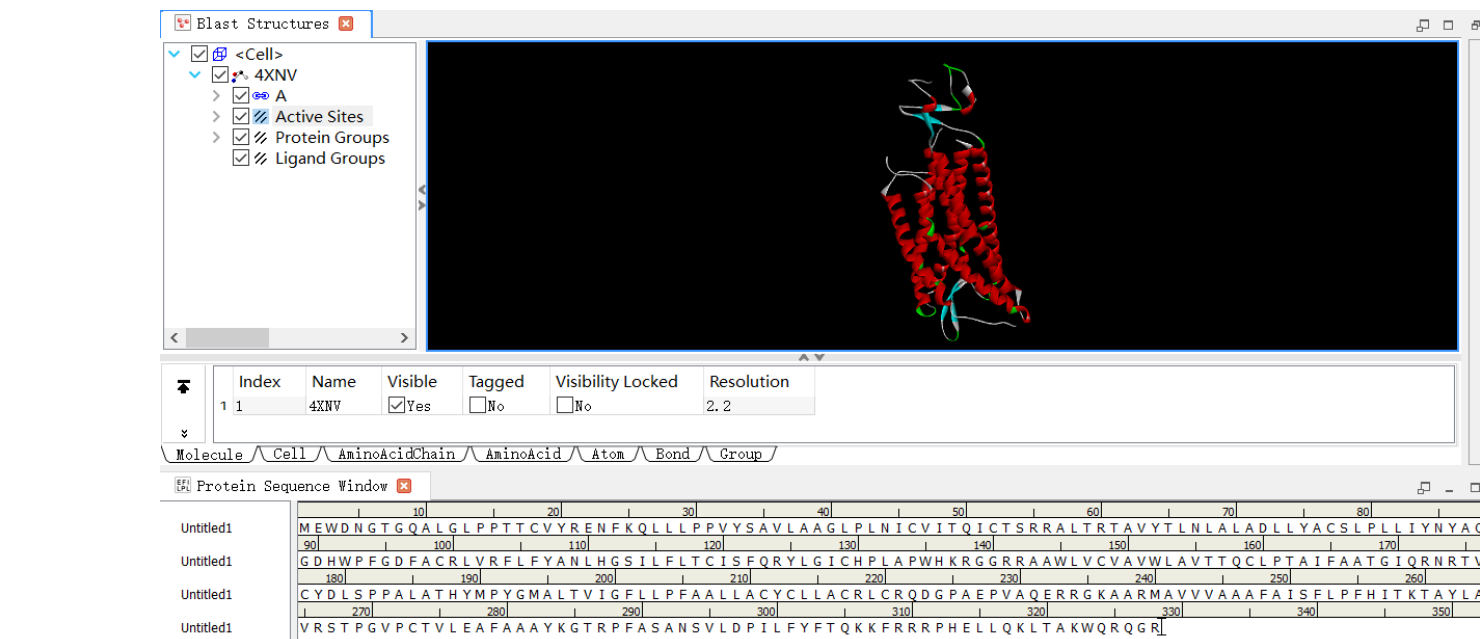
Table View



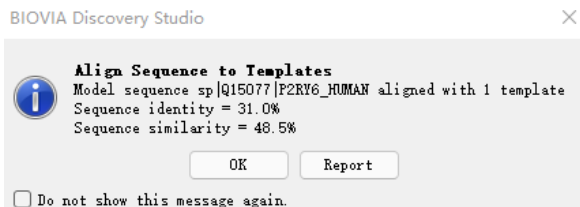
## Map View



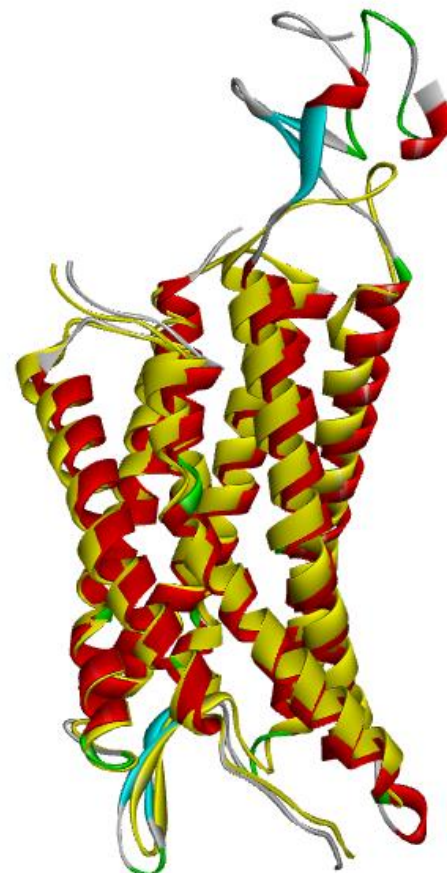
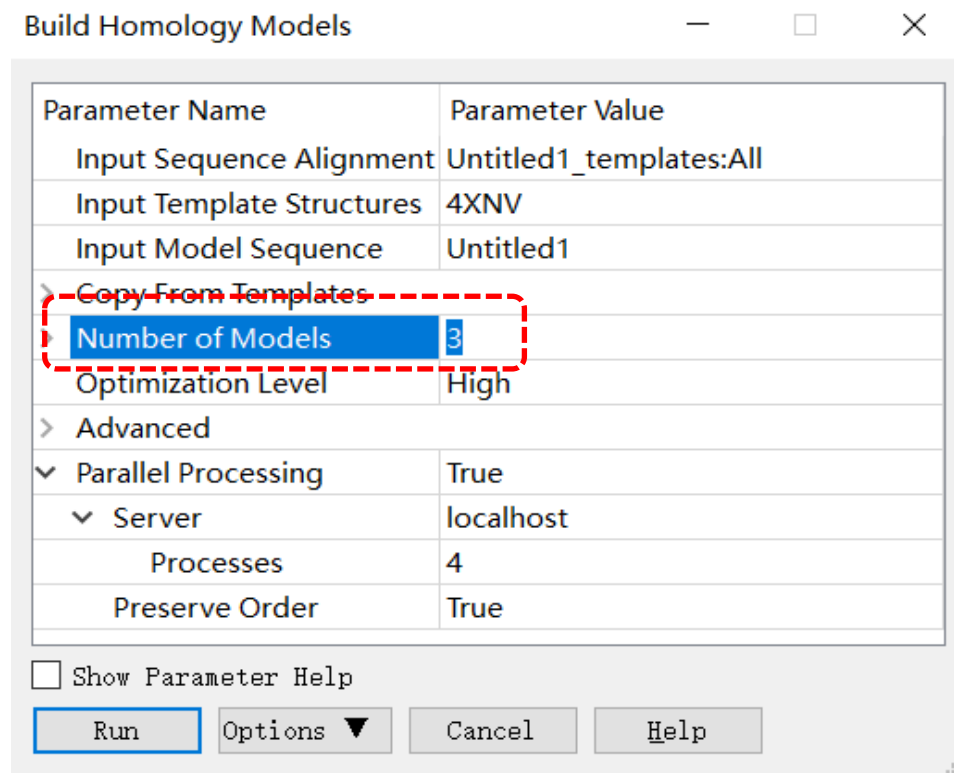
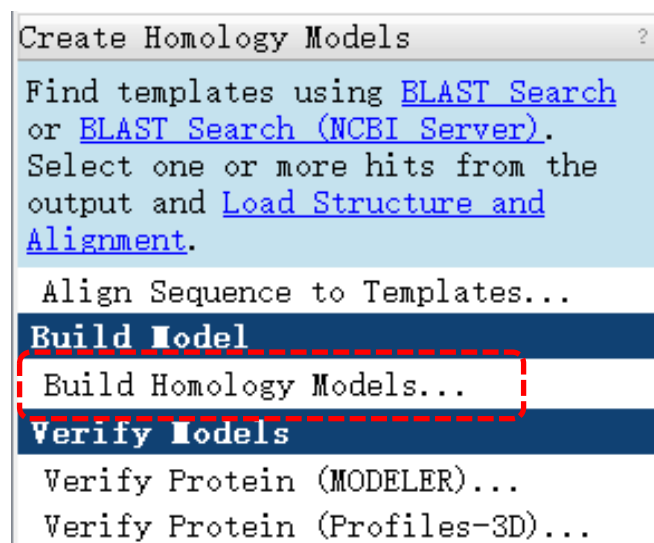
# 同源模建基本流程：（3）创建序列-结构对齐矩阵



Align Sequence to Templates	
Parameter Name	Parameter Value
Input Model Sequence	Protein Sequence Window:Untitled1
Input Template Structures	Blast Structures:Visible
Create Sequence Profile	False
Align Structures	True
<input type="checkbox"/> Show Parameter Help	
<div>RunOptions▼CancelHelp</div>	

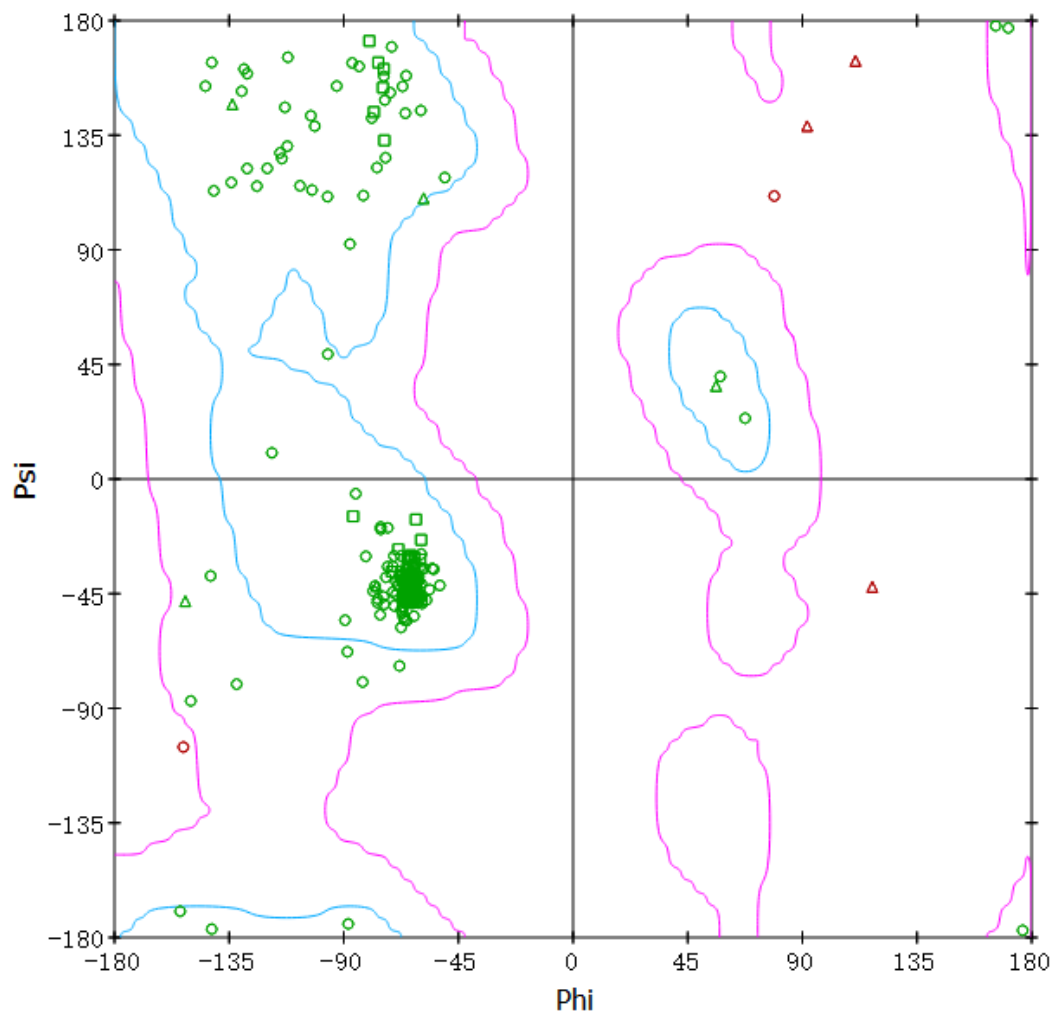
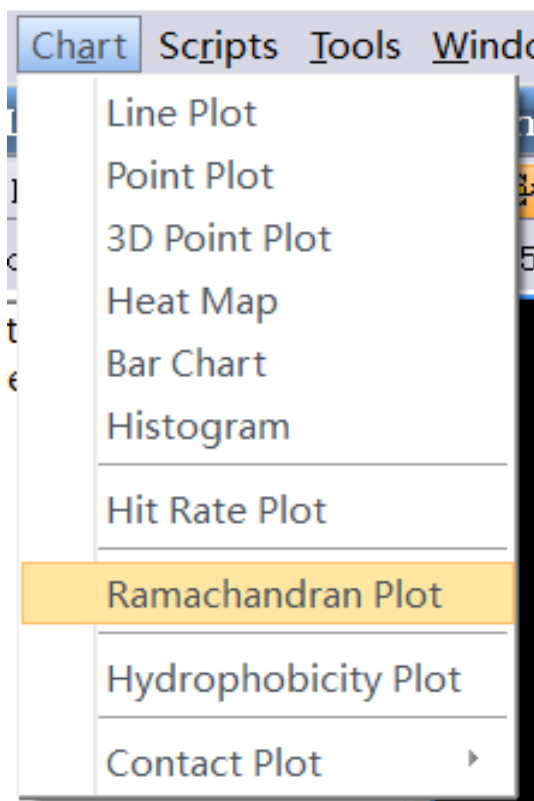


# 同源模建基本流程：（4）蛋白建模



PDF Total Energy	PDF Physical Energy	DOPE Score	RMSD	NOverlap
1, 219. 29	742. 159	-36, 820. 3	3. 744	296
1, 238. 35	754. 255	-36, 606. 4	1. 447	298
1, 412. 02	783. 256	-36, 696. 1	1. 16	298

# 同源模建基本流程： (5) 模建蛋白质质量评估——拉氏图



# 同源模建基本流程： （5） 模建蛋白质质量评估——Profile-3D

Create Homology Models

Find templates using [BLAST Search](#) or [BLAST Search \(NCBI Server\)](#). Select one or more hits from the output and [Load Structure and Alignment](#).

Align Sequence to Templates...

**Build Model**

Build Homology Models...

**Verify Models**

Verify Protein (MODELER)...

**Verify Protein (Profiles-3D)...**

**Advanced**

When BLAST fails to find templates, use [PSI-BLAST Search](#) to identify remotely related templates.

Search Side-Chain Rotamers

Minimize and Refine Protein

Verify Protein (Profiles-3D)

Parameter Name	Parameter Value
Input Protein Molecules	Q15077:Visible
Smooth Window Size	10
Secondary Structure Method	Kabsch-Sander

☐ Show Parameter Help

Run Options Cancel Help

选中需要计算的结构

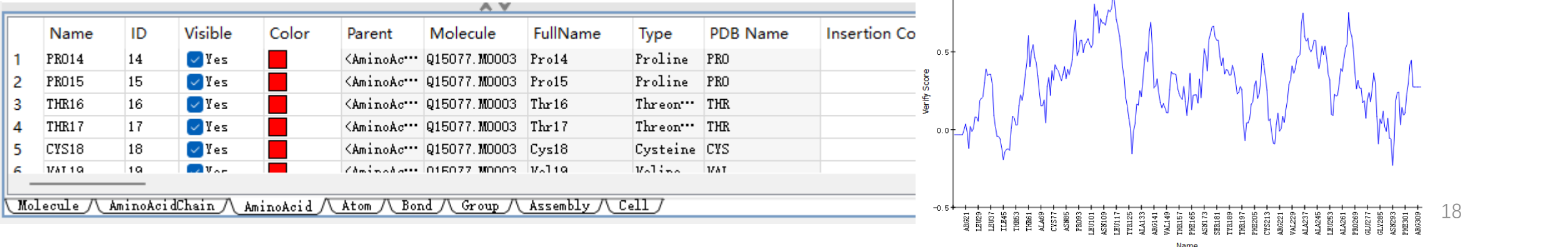
**Verify Protein (Profiles-3D)**

Protein	Verify Score	Expected High Score	Expected Low Score
Q15077.M0003	80.48	135.595	61.0176

OK Report More

☐ Do not show this message again.

- 以残基打分作图，选择Aminoacid选项卡，Chart-Line Plot，X轴设为name，Y轴设为verify score





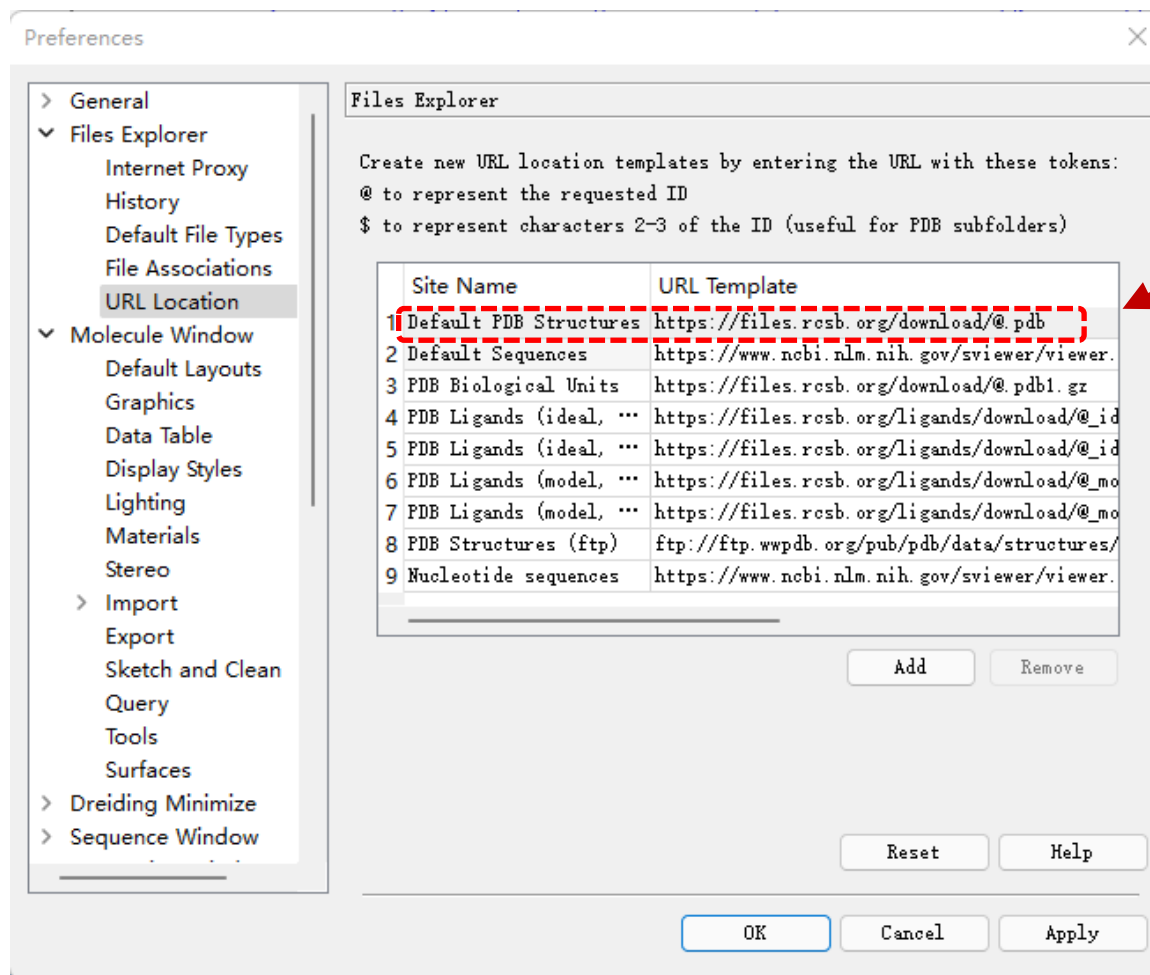
## 任务：

- (1) 从PDB数据库获取感兴趣蛋白结构并挑选优质的蛋白晶体结构  
(如EGFR、c-MET、VEGFR等)
- (2) 从Uniprot数据库获取询问蛋白序列 (P2RY6)
- (3) 通过BLAST序列比对获得PDB库中相似度最高模板蛋白
- (4) 序列对齐，同源模建
- (5) 结果分析 (使用拉氏图和Profile-3D验证模型的合理性)

# 注意事项:

操作前, 需要确保DS软件中, PDB数据库的地址是正确的。

## EDIT—Preferences—Files Explorer



<https://files.rcsb.org/download/@.pdb>

# 实验报告：

## (1) 实验目的

模建未知跨膜蛋白P2RY6三维结构

## (2) 操作流程

Unprot蛋白序列搜索、BLAST蛋白序列比对，确定模板蛋白、序列-模板对齐、蛋白建模

## (3) 结果与讨论

观测蛋白结构，用拉氏图和Profile-3D分析蛋白结构合理性