

实验名称：同源模建法构建蛋白质的三维结构

实验目的：

1. 掌握通过同源模建法进行未知跨膜蛋白 P2RY6 的三维结构的构建。

实验原理：

认为蛋白质三级结构的保守性远超过一级序列的保守性。所以根据待建模序列（目标）与一个或多个已知蛋白结构的序列（模板）间的同源性（序列一致性），直接由目标序列的一级结构预测其三级结构。模板蛋白和目标蛋白的序列一致性需要大于 30%，越大建模准确性越有保障。本实验使用 Discovery Studio 软件通过同源模建法进行未知跨膜蛋白 P2RY6 的三维结构的构建。

本实验所用的软件环境：

DS Version: 19.1.0.18287

PP Version: 19.1.0.1963

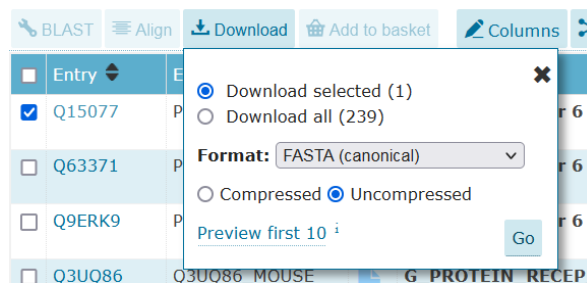
DS Client Version: 19.1.0.18287

OS Distribution: Windows

OS Version: 10.0.19044

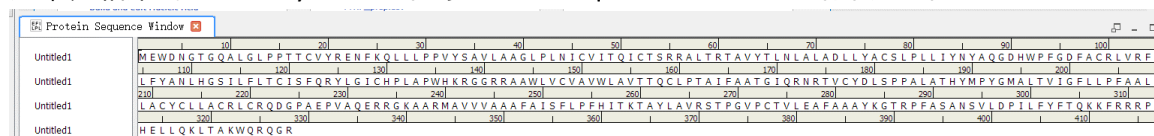
实验步骤：

1. 查找目标蛋白序列：进入 Uniprot 网站 (<https://www.uniprot.org>)，搜索目标蛋白，本实验中搜索 P2RY6。选择 Q15077，点击 Download，再点击 Go，下载目标蛋白序列。



2. Blast 蛋白序列比对：

目标蛋白序列的填入：点击 Discovery Studio 软件上菜单栏上的 File→New→Protein Sequence Window→填入目标蛋白序列，也就是复制刚才下载好的文件中的文本，粘贴到 Discovery Studio 中的 Protein Sequence Window。结果如下图：



Blast 蛋白序列比对：点击 Discovery Studio 软件上的 Macromolecules→ Create Homology Models→ BLAST Search (NCBI Server)来进行蛋白晶体数据库中相似蛋白的查找。

Parameter Name	Parameter Value
Input Sequence	Protein Sequence Window:Untitled1
Input Database	pdbaa
Organism	
Entrez Query	
E-value Cutoff	10
Maximum Hits	100
▼ Advanced	
Scoring Matrix	BLOSUM62
Gap Penalties	Existence: 11 Extension: 1
Gapped Alignment	True
Use Composition Bas...	True
Filter Low Complexity	True
Word Size	3
URL	https://blast.ncbi.nlm.nih.gov/Blast.cgi

☐ Show Parameter Help

Run Options ▼ Cancel Help

3. 选择合适蛋白模板：本实验选择了 4XNV_A (E-value 最小)，右键此选项，点击 Load Selected Structures 进行蛋白模板的载入。

4. 创建序列-结构对齐矩阵：点击 Discovery Studio 软件上的 Macromolecules→ Create Homology Models → Align Sequence to Templates 生成模板蛋白与目标蛋白的结构对齐矩阵。设置参数如下：

Parameter Name	Parameter Value
Input Model Sequence	Protein Sequence Window:Untitled1
Input Template Structures	Blast Structures:Visible
▼ Create Sequence Profile	False
Server	Local
Input Database	Swiss-Prot
Scoring Matrix	BLOSUM62
Gap Penalties	Existence: 11 Extension: 1
Gapped Alignment	True
E-value Cutoff	0.0001
Filter Low Complexity	True
Maximum Hits	500
Align Structures	True

☐ Show Parameter Help

Run Options ▼ Cancel Help

5. 蛋白建模：点击 Discovery Studio 软件上的 Macromolecules→ Create Homology Models → Build Homology Models 进行同源模型的建立。设置参数如下：

Parameter Name	Parameter Value
Input Sequence Alignment	Untitled1_templates:All
Input Template Structures	4XNV
Input Model Sequence	Untitled1
▼ Copy From Templates	
Ligands	
Nucleic Acid Chains	
▼ Protein	
Copy Regions	
Optimize Sidechains	False
Waters	False
▼ Number of Models	3
Start Model Index	1
Optimization Level	High

▼ Advanced	
Cut Overhangs	True
Disulfide Bridges	
Cis-Prolines	
Additional Restraints	
▼ Refine Loops	False
Number of Models	1
Optimization Level	High
Use DOPE Method	True
▼ Parallel Processing	True
▼ Server	localhost
Processes	4
Preserve Order	True

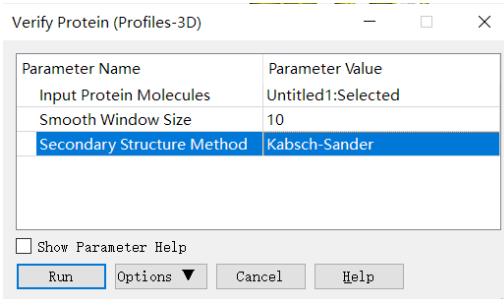
☐ Show Parameter Help

Run Options ▼ Cancel Help

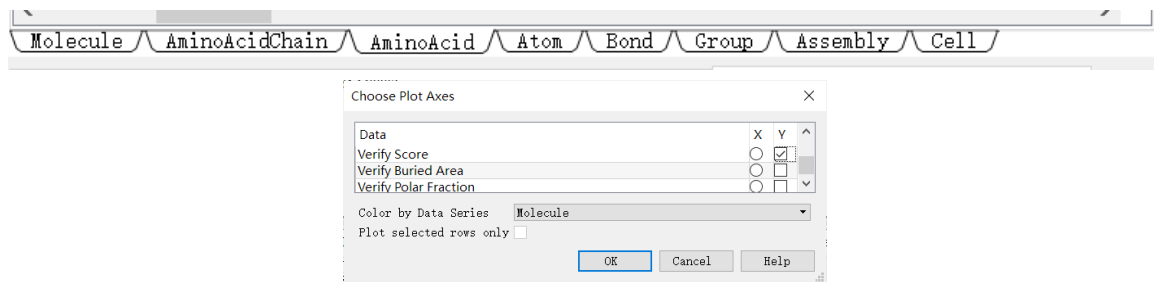
6. 模型评估:

拉氏图的绘制: 点击 Discovery Studio 软件菜单栏上的 Chart → Ramachandran Plot 分别对每个同源模型进行拉氏图的绘制。

Profiles-3D: 点击 Discovery Studio 软件上的 Macromolecules → Create Homology Models → Verify Protein (Profiles-3D) 分别对每个同源模型进行 Profiles-3D 的绘制。设置参数如下:



以残基打分作图: 选择 AminoAcid 选项卡, 点击 Discovery Studio 软件菜单栏上的 Chart → Line Plot, X 轴设为 name, Y 轴设为 verify score, Color by Data Series 设为 Molecule 同时给每个同源模型进行残基打分作图, 如下图所示。



实验结果:

1. Blast 蛋白序列比对的结果:

Status: Success

Elapsed Time: 00:00:28

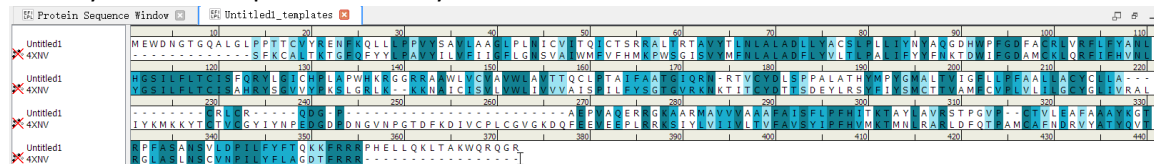
Summary: Untitled1 found 100 NCBI BLAST hits.

2. 创建序列-结构对齐矩阵的结果:

Status: Success

Elapsed Time: 00:00:13

Summary: Model sequence Untitled1 aligned with 1 template. Sequence identity = 31.0%. Sequence similarity = 48.5%.



3. 蛋白建模的结果:

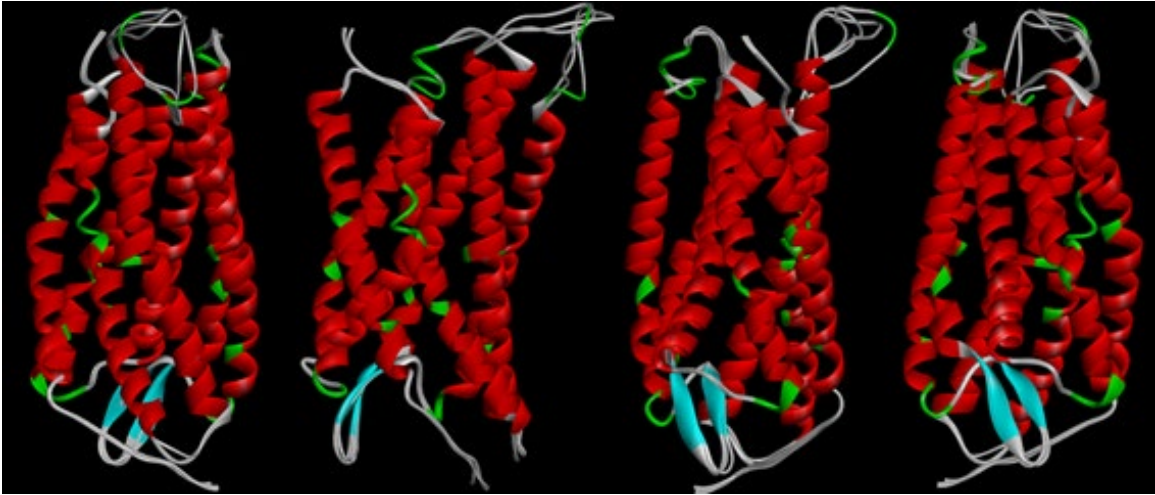
Status: Success

Elapsed Time: 00:01:24

Summary:

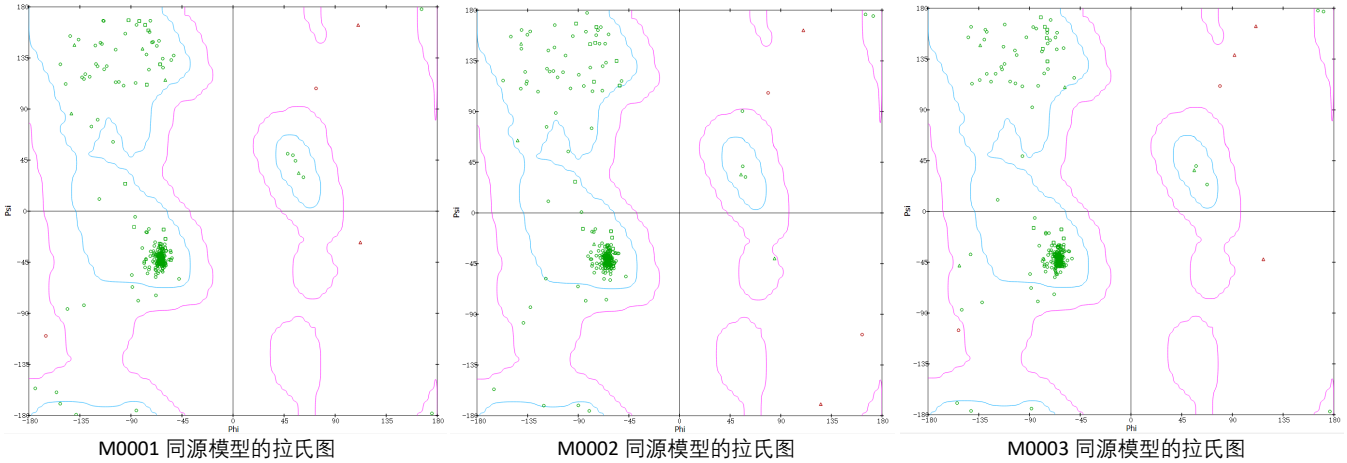
Model Scores			
Name	PDF Total Energy	PDF Physical Energy	DOPE Score
Untitled1.M0003	1219.2933	742.159478000002	-36820.281250
Untitled1.M0001	1238.3477	754.255235000001	-36606.445313
Untitled1.M0002	1412.0177	783.25606	-36696.117188

截图:



4. 模型评估的结果:

(1) 拉氏图:



(2) Profiles-3D:

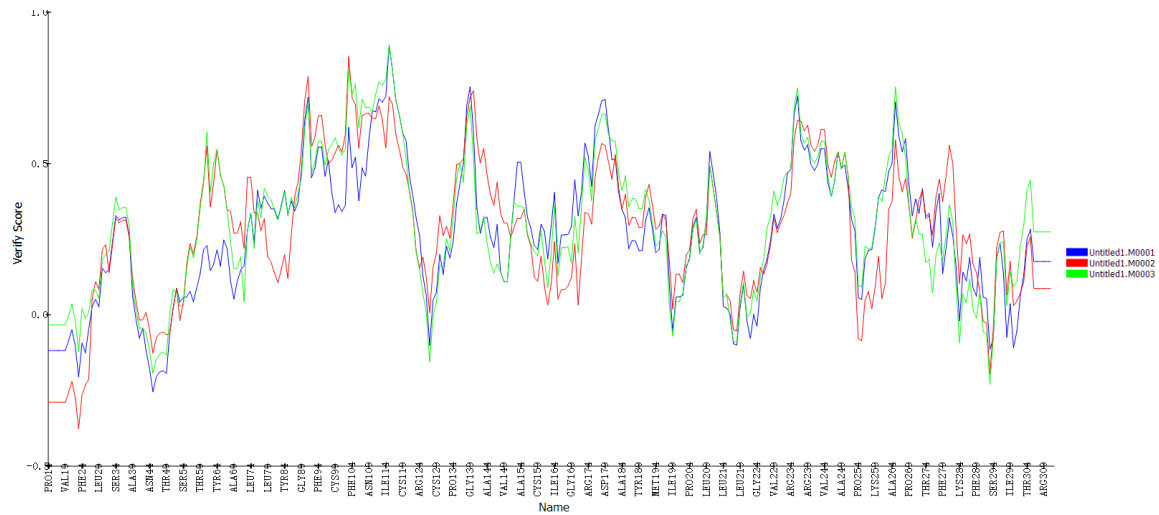
Status: Success

Elapsed Time: 00:00:24

Summary:

Name	Verify Score	Verify Expected High Score	Verify Expected Low Score
Untitled1.M0001	73.93	135.595	61.0176
Untitled1.M0002	76.6	135.595	61.0176
Untitled1.M0003	80.48	135.595	61.0176

(3) 以残基打分作图：



讨论：

可以从各模型的 PDF Total Energy 值看出，M0003 模型的可靠性最好。从拉氏图可以看出，三个模型的质量都很好（判断依据：当落于核心区+允许区的氨基酸残基百分比>95%时，表明模型质量较高）。从各模型的 Profiles-3D 评估可以看出，三个模型的可靠性都很好。