XXXX 同源建模与双底物分子对接

1. 蛋白信息

>XXXX

略

将 XXXX 对 PDB 数据库进行 blast，结果如下：

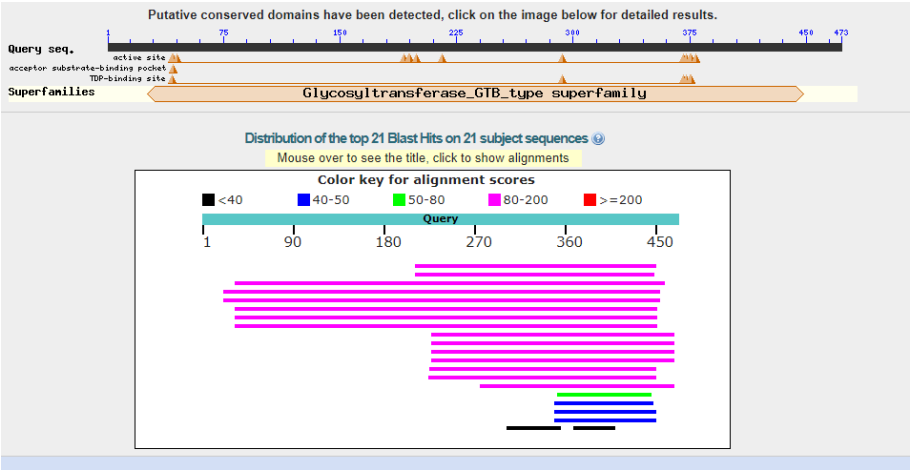


Fig 1.1 XXXX blast 结果，数据库为 PDB

显示 XXXX 蛋白有多个同源结构报道。可以应用同源建模方法来模拟 XXXX 蛋白三维结构。

2. XXXX 三维结构构建

利用 I-TASSER 软件进行同源建模，建模结果见：model_results.tar.bz2

结果说明：https://zhanglab.ccmb.med.umich.edu/I-TASSER/annotation/

共生成 5 个最优结果，我们挑取分数最高的 Model1 进行后续的分子对接。

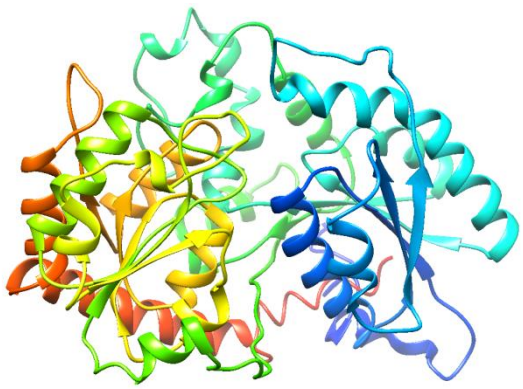


Fig 2.1 XXXX 三维结构 Model1 展示

3. 底物分子准备

Mogroside II E： https://pubchem.ncbi.nlm.nih.gov/compound/24721558#section=Top

UDP-glucose： https://pubchem.ncbi.nlm.nih.gov/compound/8629

从 Pubchem 上下载 SDF 格式的结构文件，通过 openbabel 软件转换成 PDB 格式文件（MogrosideIIe.pdb 和 UDP-glucose.pdb）。通过 AutodockTools 软件将两个底物分子转换成.pdbqt 格式，设定好可扭转的键。

4. 确定 Grid Box 参数

4.1 预测底物结合口袋

利用 COACH 软件来预测 XXXX 的结合位点

配体结合口袋预测结果

Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
1	0.95	73	3otgA	TYD	Rep, Mult	41,42,290,292,293,294,319,353,354,356,371,373,374,375,376
2	0.16	15	4remA	DLM	Rep, Mult	43,108,112,144,165,173,205,207,215,393,394,395
3	0.03	4	2VG8A	2VG8A01	Rep, Mult	164,165,373,374,375,394,395
4	0.01	1	2IYFA	2IYFA00	Rep, Mult	38,39,43,108,109,110,112,113,143,144,145,146,170,173,207,215,218,219,222,394,395
5	0.01	1	2P6PA	2P6PA04	Rep, Mult	18,21,22,24,25,30,57,58,59

注：C-score 得分越高，结果越可信，取值 0-1。

根据得分，对 3otg 和 4rem 进行了解

3OTG：https://www.rcsb.org/structure/3OTG

Crystal Structure of CalG1, Calicheamicin Glycostyltransferase, TDP bound form

报道了一种糖基转移酶 CalG1 与 THYMIDINE-5'-DIPHOSPHATE 复合物结构

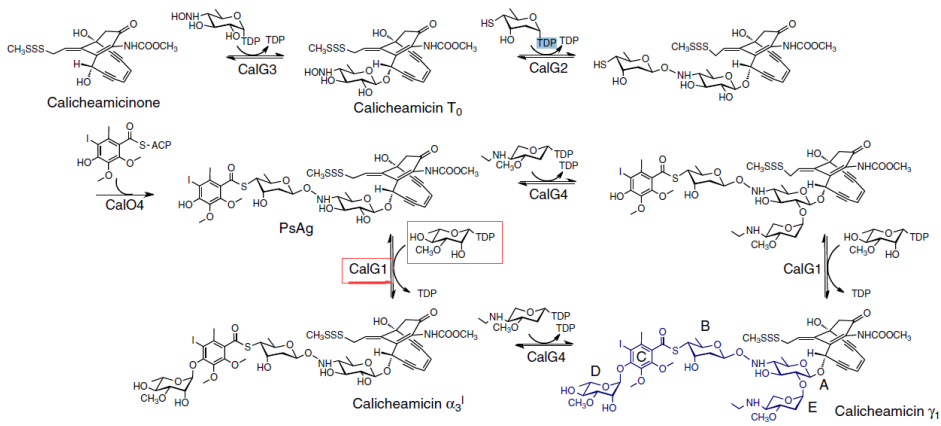


Fig. 1. Proposed calicheamicin glycosylation pathway. CalG3 mediates an internal glycosylation to the aglycon, while CalG2 mediates an internal glycosylation and CalG4 mediates an external glycosylation to the sugar A. CalG1 operates external glycosylation to the orsellinic acid-like moiety (moiety C). The order of the CalG1 and CalG4 reactions are not characterized in vivo. The names of calicheamicin intermediates are indicated below the structure. The calicheamicin γ_1^I chemical structure and sugar nomenclature is in the bottom right. The aryltetrasacchride portion (four sugars and orsellinic acid-like moiety) is colored in blue.

参考文献：10.1073/pnas.1108484108

预示着糖供体可能的结合口袋为：

41,42,290,292,293,294,319,353,354,356,371,373,374,375,376

4REM：https://www.rcsb.org/structure/4REM

Crystal structure of UDP-glucose: anthocyanidin 3-O-glucosyltransferase in complex with delphinidin

该结构解析是 UGT78K6 酶与 delphinidin 复合物结构，能催化 UDP-glucose 上的葡萄糖转移到花青素如 delphinidin

预示着糖受体可能的结合口袋：43,108,112,144,165,173,205,207,215,393,394,395

酶活性位点预测结果：

Rank	Cscore[EC]	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	EC Number	Active Site Residues
1	0.512	2vg8A	0.893	1.48	0.292	0.920	2.4.1.218	40,42,291,293,353,356,373,376,379,391
2	0.470	2c1zA	0.824	2.28	0.233	0.882	2.4.1.115	43,291,321,356,371,373,376,379,395
3	0.249	3dzcB	0.625	3.67	0.095	0.736	5.1.3.14	NA
4	0.249	1fa9A	0.604	5.90	0.046	0.852	2.4.1.1	NA
5	0.249	3beoA	0.611	4.13	0.099	0.748	5.1.3.14	NA

注：C-score 得分越高，结果越可信，取值 0-1。

根据得分，对 2VG8 和 2C1Z 进行了解

2VG8：https://www.rcsb.org/structure/2VG8

Characterization and engineering of the bifunctional N- and O- glucosyltransferase involved in xenobiotic metabolism in plants

该结构是对苯二酚葡糖基转移酶与 URIDINE-5'-DIPHOSPHATE 复合物结构

预示着可能的活性中心为：40,42,291,293,353,356,373,376,379,391

2C1Z：https://www.rcsb.org/structure/2C1Z 该结构是个三元复合物，重点关注

Structure and activity of a flavonoid 3-O glucosyltransferase reveals the basis for plant natural product modification

该结构是 UDP-glucose flavonoid 3-o glycosyltransferase 与
URIDINE-5'-DIPHOSPHATE-2-DEOXY-2-FLUORO-ALPHA-D-GLUCOSE 和
3,5,7-TRIHYDROXY-2-(4-HYDROXYPHENYL)-4H-CHROMEN-4-ONE 的复合物结构

可能的活性中心为：43,291,321,356,371,373,376,379,395

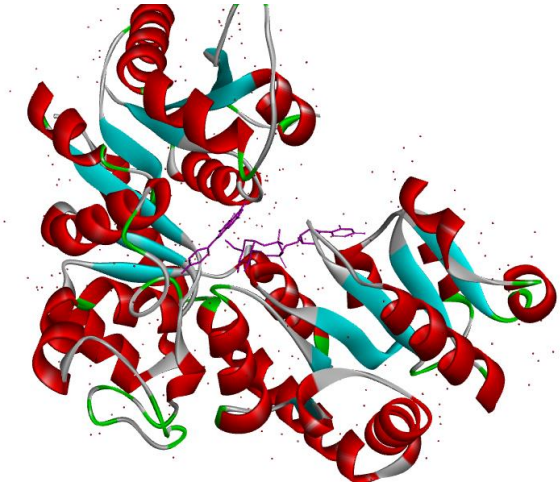


Fig 4.1 2C1Z 结构展示，紫色标记的 2 个分子为底物(类似物)

2C1Z 与供体相互作用

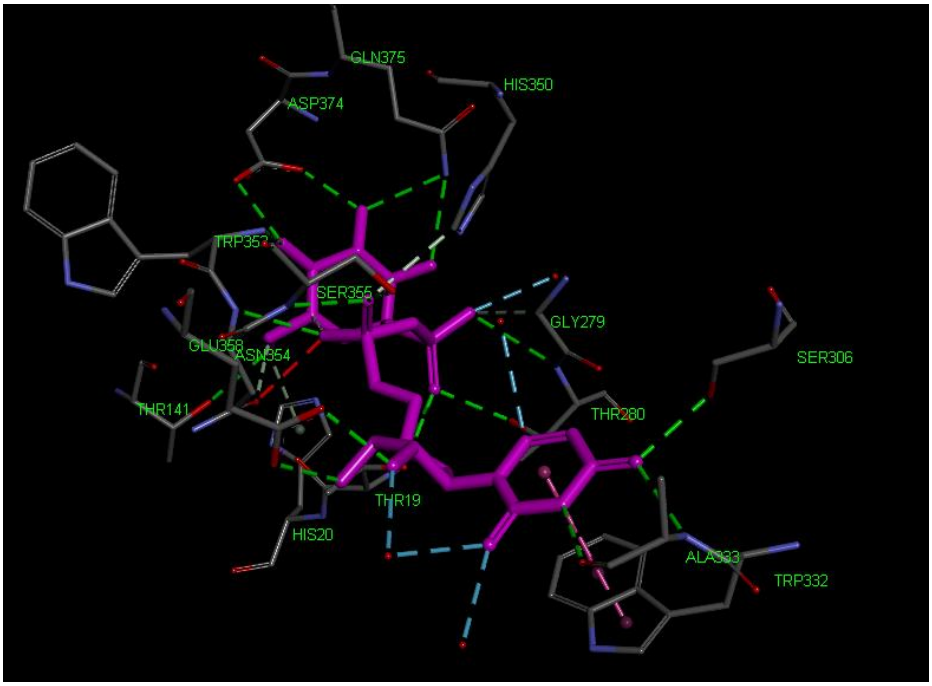


Fig 4.1 2C1Z 与供体相互作用

2C1Z 与 XXXX 结构比对（PyMol），并标记上供体结合口袋

Fig 4.2 2C1Z 与 XXXX 结构比对，其中红色区域为供体的结合口袋

根据比对结果，不难发现对应到 XXXX 上，可能的供体结合口袋为：42,43,164,292,293,319,353,354,371,374,375,376,379,395,396。结构比对文件：alignwith2C1ZA-donor.pse，可用 pymol 软件打开

2C1Z 与受体相互作用

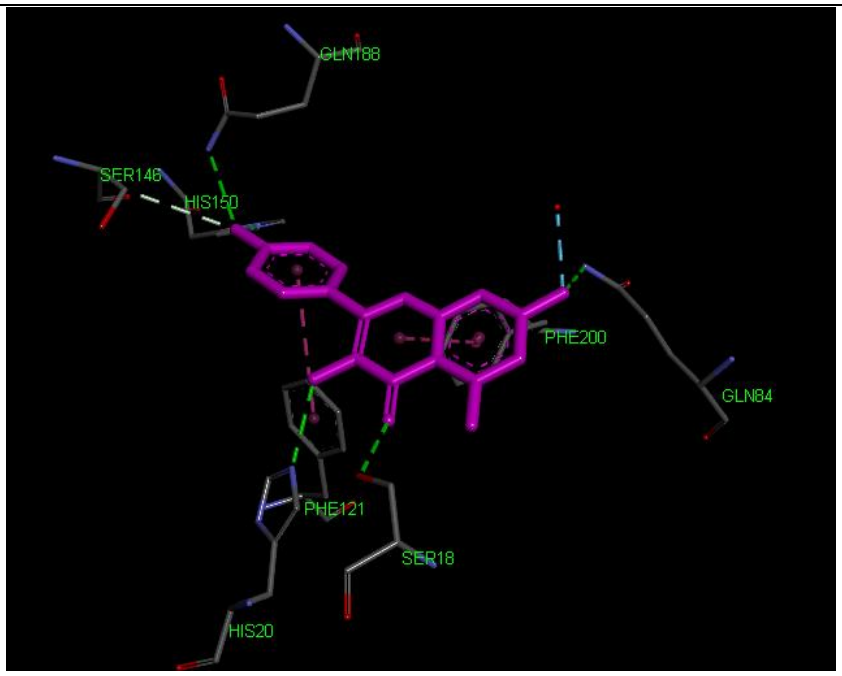


Fig 4.3 2C1Z 与受体相互作用

2C1Z 与 XXXX 结构比对（PyMol），并标记上受体结合口袋

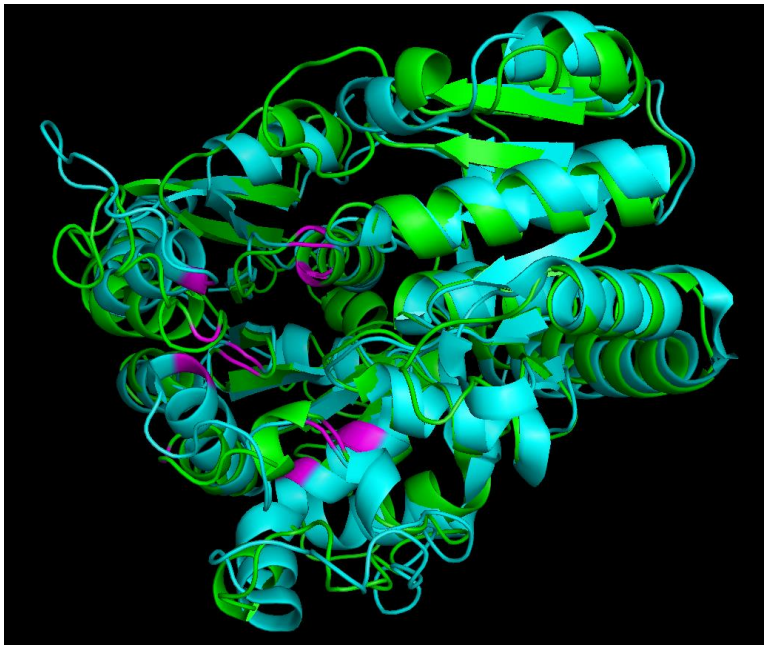


Fig 4.4 2C1Z 与 XXXX 结构比对，其中红色区域为受体的结合口袋

根据比对结果，不难发现对应到 XXXX 上，可能的供体结合口袋为：41,43,107,144,169,173,205,213。比对文件：alignwith2C1ZA-Receptor.pse，可用 pymol 软件打开.

综合上述结果

供体(3OTG)	供体(2VG8)	供体(2C1Z)	受体(4REM)	受体(2C1Z)
41	40	42	43	41
42	42	43	108	43
290	291	164	112	107
292	293	292	144	144
293	353	293	165	169
294	356	319	173	173
319	373	353	205	205
353	376	354	207	213
354	379	371	215	
356	391	374	393	
371		375	394	
373		376	395	
374		379		
375		395		
376		396		

根据上述结果大致可以确定供体与受体结合口袋区域，从而确定 Grid box 参数

确定供体的 Grid box 参数为：

spacing 0.375
npts 56 60 72
center 63.492 66.334 57.685

受体的 Grid box 参数为：

spacing 0.375
npts 86 60 72
center 85.764 66.334 64.384

5. 分子对接

常用的分子对接软件有：AutoDock, AutoDock Vina, LeDock, rDock, UCSF DOCK；LigandFit, Glide, GOLD, MOE Dock 和 Surflex-Dock。其中前 5 种是学院版，后五种是商业版。综合测评结果表明 AutoDock, AutoDock Vina, LeDock 表现优异，不劣于商业版分子对接软件。

5.1 酶与供体进行对接

5.1.1 采用 AutodockDockTools 软件进行对接。对接过程可以参考我司的教程：AutoDock 实例教程

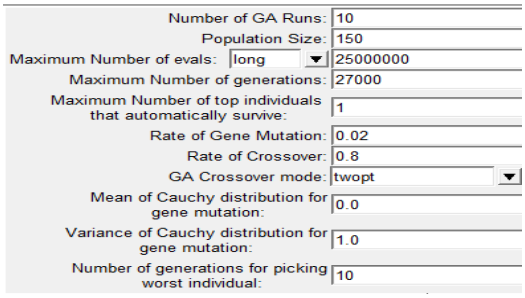


Fig 5.1 Search parameters 设置

保留 100 个最优结合构象

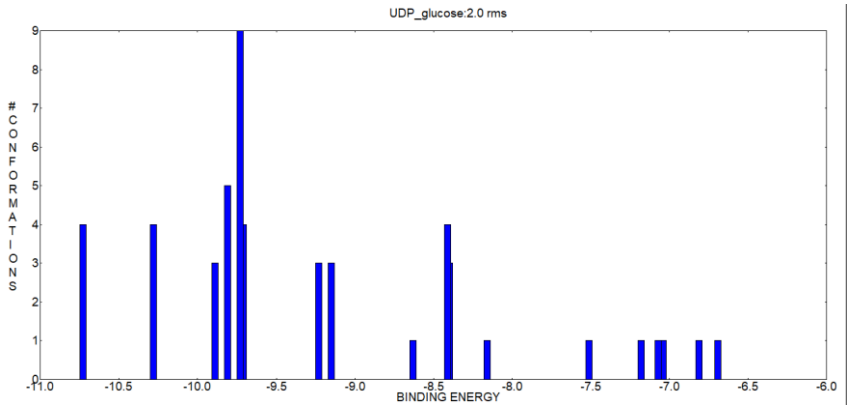


Fig 5.2 100 种结合构象的聚类结果

binding_energy=-10.73
ligand_efficiency=-0.3
inhib_constant=13.61
inhib_constant_units=nM
intermol_energy=-14.01
vdw_hb_desolv_energy=-13.75
electrostatic_energy=-0.26
total_internal=3.22
torsional_energy=3.28
unbound_energy=3.22
filename=ER.dlg
cIRMS=0.0
refRMS=102.83
rseed1=None
rseed2=None

Fig 5.3 能量最低模型信息

5.1.2 在 Discovery Studio 软件中分析 XXXX 与糖供体的相互作用

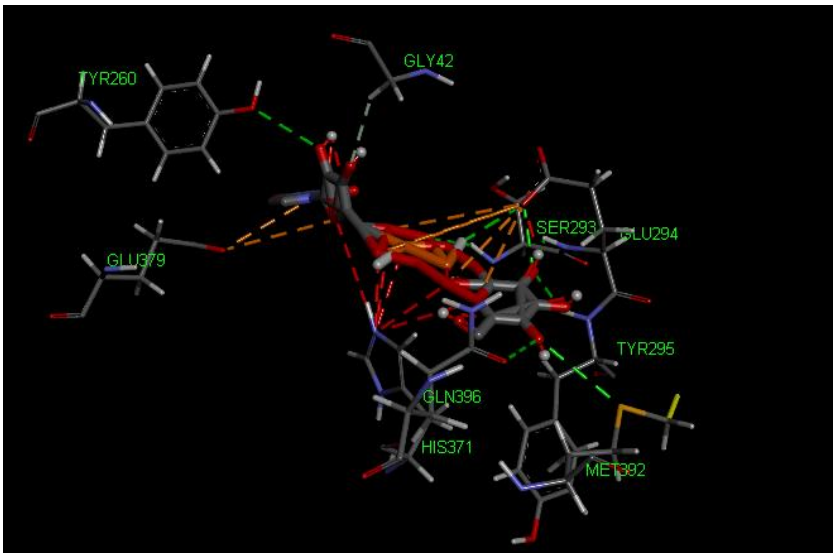


Fig 5.4 XXXX 与糖供体的相互作用

相互作用残基详细信息

Name	Distance	Category	Types	From	From chemistry	To	To chemistry
:UNL1:P - A:GLU294:OE1	4.42701	Electrostatic	Attractive Charge	:UNL1:P	Positive	A:GLU294:OE1	Negative
:UNL1:O - A:GLU294:OE1	4.14772	Electrostatic	Attractive Charge	:UNL1:O	Positive	A:GLU294:OE1	Negative
:UNL1:O - A:GLU294:OE1	4.79942	Electrostatic	Attractive Charge	:UNL1:O	Positive	A:GLU294:OE1	Negative
:UNL1:O - A:GLU294:OE1	4.08212	Electrostatic	Attractive Charge	:UNL1:O	Positive	A:GLU294:OE1	Negative
:UNL1:O - A:GLU294:OE1	5.34253	Electrostatic	Attractive Charge	:UNL1:O	Positive	A:GLU294:OE1	Negative
:UNL1:O - A:GLU379:OE1	4.30026	Electrostatic	Attractive Charge	:UNL1:O	Positive	A:GLU379:OE1	Negative
:UNL1:N - A:GLU379:OE1	3.80098	Electrostatic	Attractive Charge	:UNL1:N	Positive	A:GLU379:OE1	Negative
A:SER293:HN - :UNL1:O	2.78189	Hydrogen Bond	Conventional Hydrogen Bond	A:SER293:HN	H-Donor	:UNL1:O	H-Acceptor
A:SER293:HN - :UNL1:O	2.857	Hydrogen Bond	Conventional Hydrogen Bond	A:SER293:HN	H-Donor	:UNL1:O	H-Acceptor
A:GLU294:HN - :UNL1:O	2.09126	Hydrogen Bond	Conventional Hydrogen Bond	A:GLU294:HN	H-Donor	:UNL1:O	H-Acceptor
A:TYR295:HN - :UNL1:O	2.49065	Hydrogen Bond	Conventional Hydrogen Bond	A:TYR295:HN	H-Donor	:UNL1:O	H-Acceptor
:UNL1:H - A:GLU294:OE1	2.57333	Hydrogen Bond	Conventional Hydrogen Bond	:UNL1:H	H-Donor	A:GLU294:OE1	H-Acceptor
:UNL1:O - A:TYR260:OH	3.26706	Hydrogen Bond	Conventional Hydrogen Bond	:UNL1:O	H-Donor	A:TYR260:OH	H-Acceptor
:UNL1:O - A:GLU294:OE1	2.40303	Hydrogen Bond	Conventional Hydrogen Bond	:UNL1:O	H-Donor	A:GLU294:OE1	H-Acceptor
:UNL1:O - A:MET392:SD	3.55301	Hydrogen Bond	Conventional Hydrogen Bond	:UNL1:O	H-Donor	A:MET392:SD	H-Acceptor
:UNL1:O - A:GLN396:OE1	3.33662	Hydrogen Bond	Conventional Hydrogen Bond	:UNL1:O	H-Donor	A:GLN396:OE1	H-Acceptor
A:GLY42:HA1 - :UNL1:O	2.75749	Hydrogen Bond	Carbon Hydrogen Bond	A:GLY42:HA1	H-Donor	:UNL1:O	H-Acceptor

糖供体结合口袋氨基酸组成

序号	最优模型	次优模型
1	GLY42	GLY42
2	TYR260	TYR260
3	SER293	PHE291
4	GLU294	GLY292
5	TYR295	SER293
6	HIS371	GLU294
7	GLU379	TYR295
8	MET392	ALA357
9	GLN396	HIS371
10		GLU379
11		GLN396

注：最优模型 er-1.pdb，次优模型：er-2.pdb；最优，次优特指能量上最低与次低

糖供体结合口袋的残基组成与我们在第 4 部分的分析结果高度一致。最优模型（er-1.pdb）将用于与糖受体的的对接

5.2 最优模型与受体对接

这一轮对接我们采用 AutoDock Vina 完成，最终保留 9 个最优模型。程序相关参数为：

receptor = e:\autodock\lackR.pdbqt

ligand = e:\autodock\MogrosideIIe.pdbqt

out = e:\autodock\lackR_out.pdbqt

log = e:\autodock\lackR_log.txt

num_modes = 9
exhaustiveness = 9
center_x = 85.764
center_y = 66.334
center_z = 64.384
size_x = 86
size_y = 60
size_z = 72

9 个模型的相关数据如下：

mode	affinity (kcal/mol)	dist from best mode rmsd l.b.	rmsd u.b.
1	-13.6	0.000	0.000
2	-12.5	1.563	3.015
3	-12.5	1.191	2.016
4	-12.4	2.149	8.807
5	-12.3	2.186	4.075
6	-12.3	2.712	4.171
7	-12.1	1.547	9.179
8	-12.1	2.968	9.314
9	-12.1	1.697	3.045

在 Discovery Studio 软件中分析 XXXX 与糖受体的相互作用

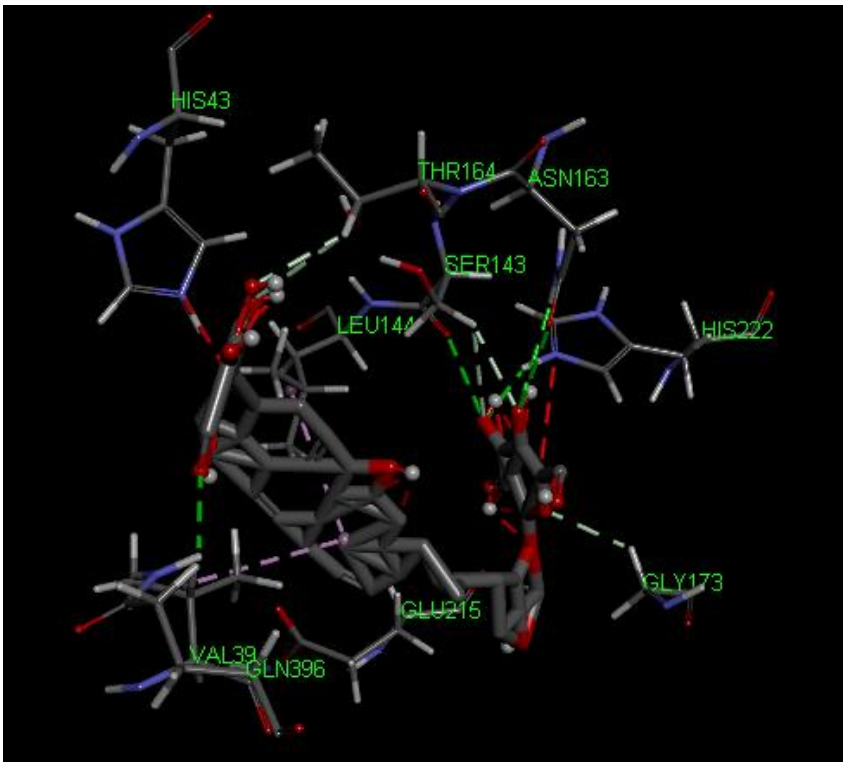


Fig 5.5 XXXX 与糖受体的相互作用

参与相互作用残基详细信息

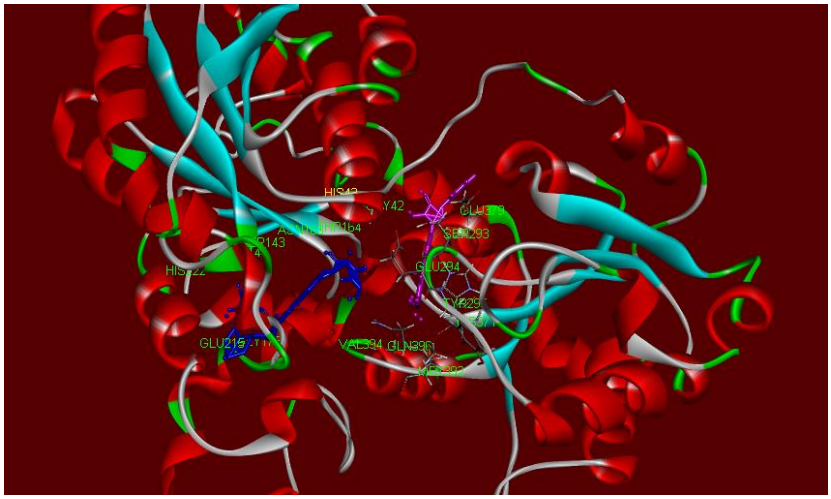
Name	Distance	Category	Types	From	From chemistry	To	To chemistry
A:ASN163:HD21 - :MOG1:O	2.98273	Hydrogen Bond	Conventional Hydrogen Bond	A:ASN163:HD21	H-Donor	:MOG1:O	H-Acceptor
A:HIS222:HE2 - :MOG1:O	2.26834	Hydrogen Bond	Conventional Hydrogen Bond	A:HIS222:HE2	H-Donor	:MOG1:O	H-Acceptor
A:GLN396:HE22 - :MOG1:O	2.53995	Hydrogen Bond	Conventional Hydrogen Bond	A:GLN396:HE22	H-Donor	:MOG1:O	H-Acceptor
:MOG1:O - A:ASN163:OD1	2.93193	Hydrogen Bond	Conventional Hydrogen Bond	:MOG1:O	H-Donor	A:ASN163:OD1	H-Acceptor
:MOG1:O - A:SER143:O	3.0645	Hydrogen Bond	Conventional Hydrogen Bond	:MOG1:O	H-Donor	A:SER143:O	H-Acceptor
A:SER143:HB2 - :MOG1:O	2.48446	Hydrogen Bond	Carbon Hydrogen Bond	A:SER143:HB2	H-Donor	:MOG1:O	H-Acceptor
A:SER143:HB2 - :MOG1:O	2.99572	Hydrogen Bond	Carbon Hydrogen Bond	A:SER143:HB2	H-Donor	:MOG1:O	H-Acceptor
A:THR164:HB - :MOG1:O	2.52663	Hydrogen Bond	Carbon Hydrogen Bond	A:THR164:HB	H-Donor	:MOG1:O	H-Acceptor
A:THR164:HB - :MOG1:O	3.02004	Hydrogen Bond	Carbon Hydrogen Bond	A:THR164:HB	H-Donor	:MOG1:O	H-Acceptor
A:GLY173:HA2 - :MOG1:O	2.40801	Hydrogen Bond	Carbon Hydrogen Bond	A:GLY173:HA2	H-Donor	:MOG1:O	H-Acceptor
:MOG1:C - A:GLU215:O	3.14837	Hydrogen Bond	Carbon Hydrogen Bond	:MOG1:C	H-Donor	A:GLU215:O	H-Acceptor
A:LEU144 - :MOG1	4.76345	Hydrophobic	Alkyl	A:LEU144	Alkyl	:MOG1	Alkyl
A:VAL394 - :MOG1	5.16888	Hydrophobic	Alkyl	A:VAL394	Alkyl	:MOG1	Alkyl

糖受体结合口袋氨基酸组成

序号	最优结构	次优结构
1	HIS43	HIS43
2	SER143	LEU144
3	LEU144	ILE170
4	ASN163	TYR219
5	THR164	HIS222
6	GLY173	TRP374
7	GLU215	ASP395
8	HIS222	
9	VAL394	
10	GLN396	

注：最优，次优特指能量上最低与次低，模型文件分别为：all-1.pdb 和 all-2.pdb

最终模型展示：



有可能 HIS43 是催化反应的最关键残基

6. 参考文献

1. J Yang, Y Zhang. I-TASSER server: new development for protein structure and function predictions, Nucleic Acids Research, 43: W174-W181, 2015.

2. C Zhang, PL Freddolino, Y Zhang. COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. Nucleic Acids Research, 45: W291-W299, 2017.

3. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-61

4. Michel F. Sanner. Python: A Programming Language for Software Integration and Development. J. Mol. Graphics Mod., 1999, Vol 17, February. pp57-61

5. Wang Z, Sun H, Yao X, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. Phys Chem Chem Phys. 2016;18(18):12964-75

6. Raghavendra S, Aditya Rao SJ, Kumar V, Ramesh CK. Multiple ligand simultaneous docking (MLSD): A novel approach to study the effect of inhibitors on substrate binding to PPO. Comput Biol Chem. 2015

7. Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8, 2015.

8. N M O'Boyle, M Banck, C A James, C Morley, T Vandermeersch, and G R Hutchison. "Open Babel: An open chemical toolbox." J. Cheminf. (2011), 3, 33. DOI:10.1186/1758-2946-3-33