



Contents

1 摘要	1
2 前言	1
3 材料和方法	1
3.1 材料	1
3.2 方法	1
4 分析结果	1
5 结论	1
6 附：分析流程	1
6.1 成分	1
6.2 成分靶点	2
6.3 成分靶点网络	3
6.4 疾病靶点	4
6.5 成分-疾病-靶点	6
6.6 分子对接	9
6.6.1 对接可视化	11
6.6.2 对接细节可视化	14
Reference	17



List of Figures

1 Network pharmacology visualization	4
2 Overall targets number of datasets	6
3 Network pharmacology with disease	7
4 Targets intersect with targets of diseases	8
5 Overall combining Affinity	10
6 Docking 643757 into 1tqn	12
7 Docking 99477 into 2dn8	13
8 Docking 5316639 into 4cgv	14

9	Docking 643757 into 1tqn detail	15
10	Docking 99477 into 2dn8 detail	16
11	Docking 5316639 into 4cgv detail	17

List of Tables

1	Targets predicted by Super Pred	3
2	GeneCards used data	5
3	Affinity data	11

1 摘要

```
R input
dic(dic(" 差异表达基因", "Differential Expressed Genes", "DEGs"),
  di(" 非酒精性脂肪肝")
)

# DEGs: Differential Expressed Genes 差异表达基因
# NFLD: Nonalcoholic fatty liver disease 非酒精性脂肪肝
```

2 前言

3 材料和方法

3.1 材料

3.2 方法

Mainly used method:

- Databases of DisGeNet, GeneCards, PharmGKB used for collating disease related targets¹⁻³.
- Website HERB <http://herb.ac.cn/> used for TCM data source⁴.
- R package PubChemR used for querying compounds information.
- Web tool of Super-PRED used for drug-targets relationship prediction⁵.
- The CLI tools of AutoDock vina and ADFR software used for auto molecular docking⁶⁻¹⁰.
- R package ChemmineR used for similar chemical compounds clustering¹¹.
- R version 4.4.0 (2024-04-24); Other R packages (eg., dplyr and ggplot2) used for statistic analysis or data visualization.

4 分析结果

5 结论

6 附：分析流程

6.1 成分

- Aconitic acid
- Dianthoside
- Allo Maltol
- 6-Hydroxycoumarin

- Homovanillyl alcohol 4-O-glucoside
- Epicatechin
- Isoquercetin
- Berberine
- Quinic acid
- 2-Furoic acid
- Adoxosidic acid
- Neochlorogenic acid
- 3-O-Feruloylquinic acid
- Rutin
- Quercetin
- Isorhamnetin
- Gingerol

R input

```
cpds <- c("Aconitic acid", "Dianthoside", "Allo Maltol", "6-Hydroxycoumarin", "Homovanillyl alcohol",
  "Epicatechin", "Isoquercetin", "Berberine", "Quinic acid", "2-Furoic acid", "Adoxosidic acid",
  "Neochlorogenic acid", "3-O-Feruloylquinic acid", "Rutin", "Quercetin",
  "Isorhamnetin", "Gingerol")

infoCpds <- PubChemR::get_cids(cpds)
infoCpds <- dplyr::distinct(infoCpds, Identifier, .keep_all = T)
if (!identical(nrow(infoCpds), length(cpds))) {
  message("Not Found some compounds.")
}
infoCpds <- dplyr::mutate(infoCpds, CID = as.integer(CID))
infoCpds
```

R input

```
pub <- job_pubchemr(nl(infoCpds$Identifier, infoCpds$CID, F))
pub <- step1(pub)
```

6.2 成分靶点

R input

```
sup <- asjob_superpred(pub)
sup <- step1(sup)
sup@tables$step1$targets
```

Table 1 (下方表格) 为表格 Targets predicted by Super Pred 概览。

(对应文件为 **Figure+Table/Targets-predicted-by-Super-Pred.xlsx**)

注：表格共有 827 行 9 列，以下预览的表格可能省略部分数据；含有 17 个唯一 ‘id’。

Table 1: Targets predicted by Super Pred

.id	Target...	ChEMBL-ID	UniPro...	PDB Vi...	TTD ID	Probab...	Model ...	symbols
C1=CC(...	Tyrosy...	CHEMBL...	Q9NUW8	6N0D	Not Av...	99.49%	71.22%	TDP1
C1=CC(...	DNA-(a...	CHEMBL...	P27695	6BOW	T13348	99.11%	91.11%	APEX1
C1=CC(...	Monoam...	CHEMBL...	P21397	2Z5Y	Not Av...	97.49%	91.49%	MAOA
C1=CC(...	DNA to...	CHEMBL...	P11388	6ZY5	T17048	95.68%	89%	TOP2A
C1=CC(...	Arachi...	CHEMBL...	P18054	3D3L	Not Av...	95.66%	75.57%	ALOX12
C1=CC(...	Transt...	CHEMBL...	P02766	6SUG	T86462	93.11%	90.71%	TTR
C1=CC(...	Thyroi...	CHEMBL...	P10827	3ILZ	T79591	93.02%	99.15%	THRA
C1=CC(...	Cathep...	CHEMBL...	P07339	4OD9	T67102	92.28%	98.95%	CTSD
C1=CC(...	Riboso...	CHEMBL...	P51812	4D9T	Not Av...	90.31%	95.64%	RPS6KA3
C1=CC(...	Kruppe...	CHEMBL...	Q13887	Not Av...	Not Av...	88.96%	86.33%	KLF5
C1=CC(...	Pregna...	CHEMBL...	O75469	6TFI	T82702	87.7%	94.73%	NR1I2
C1=CC(...	Dual s...	CHEMBL...	P51452	3F81	Not Av...	87.51%	94%	DUSP3
C1=CC(...	Transc...	CHEMBL...	O15164	4YBM	Not Av...	86.89%	95.56%	TRIM24
C1=CC(...	Dual s...	CHEMBL...	Q9HAZ1	6FYV	Not Av...	85.69%	94.45%	CLK4
C1=CC(...	Serine...	CHEMBL...	O75460	6W39	Not Av...	84.58%	98.11%	ERN1
...

6.3 成分靶点网络

R input

```
hb <- do_herb(pub, sup)
hb@plots$step3$p.pharm
```

Figure 1 (下方图) 为图 Network pharmacology visualization 概览。
 (对应文件为 Figure+Table/Network-pharmacology-visualization.pdf)

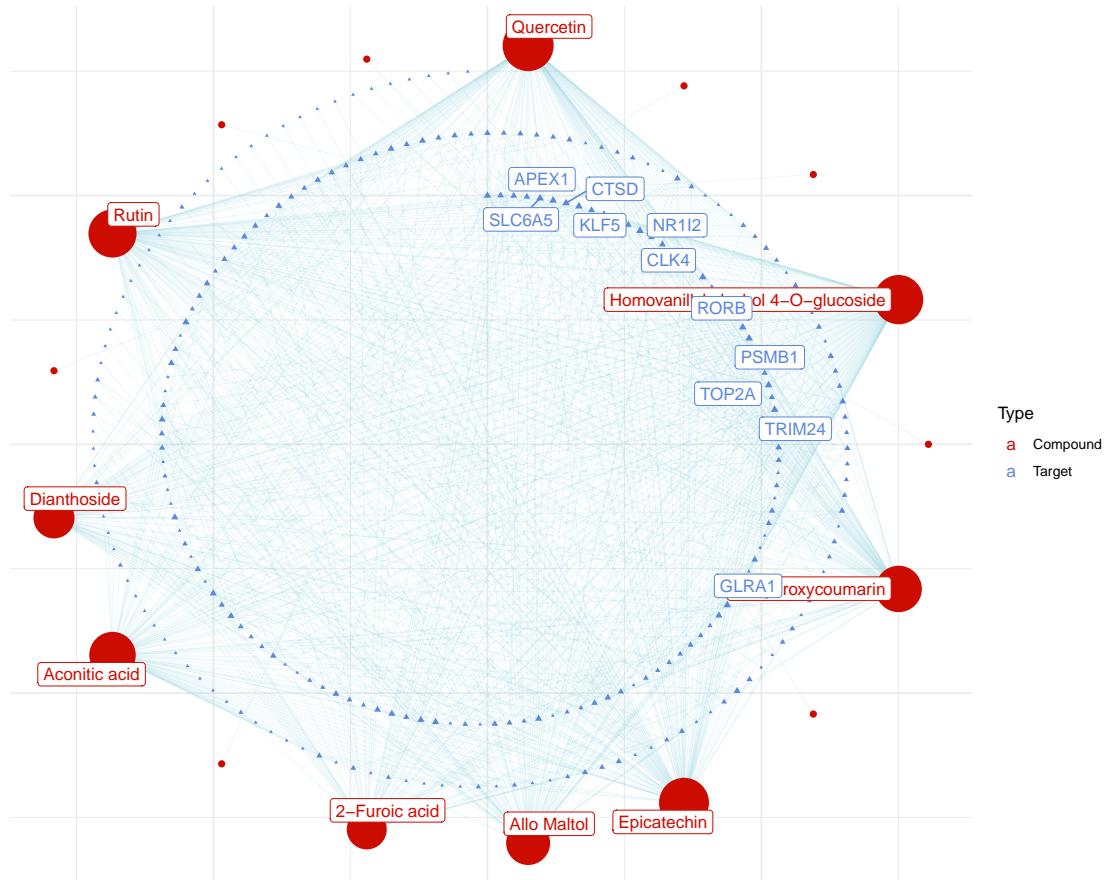


Figure 1: Network pharmacology visualization

6.4 疾病靶点

```
R input
gm <- job_gmix("Nonalcoholic fatty liver disease", "nonalcoholic")
gm <- step1(gm)
gm <- step2(gm, NULL, 1, 3, restrict = T)
gm@tables$step2$t.genecard
gm@plots$step2$p.cols
```

Table 2 (下方表格) 为表格 GeneCards used data 概览。

(对应文件为 **Figure+Table/GeneCards-used-data.xlsx**)

注：表格共有 289 行 7 列，以下预览的表格可能省略部分数据；含有 289 个唯一 ‘Symbol’。

The GeneCards data was obtained by querying :

Nonalcoholic fatty liver disease

Restrict (with quotes) :

TRUE

Filtering by Score: :

Score > 3

Table 2: GeneCards used data

Symbol	Description	Category	UniProt_ID	GIFtS	GC_id	Score
PNPLA3	Patatin Li...	Protein Co...	Q9NST1	51	GC22P043923	36.37
ADIPOQ	Adiponecti...	Protein Co...	Q15848	55	GC03P186842	23.53
INS	Insulin	Protein Co...	P01308	56	GC11M002159	22.46
LEP	Leptin	Protein Co...	P41159	55	GC07P128241	18.65
TNF	Tumor Necr...	Protein Co...	P01375	61	GC06P134820	17.83
PPARA	Peroxisome...	Protein Co...	Q07869	53	GC22P046150	17.77
MIR122	MicroRNA 122	RNA Gene (...)		29	GC18P058451	17.58
GPT	Glutamic...	Protein Co...	P24298	51	GC08P144502	16.73
CYP2E1	Cytochrome...	Protein Co...	P05181	55	GC10P133520	16.28
APOB	Apolipopro...	Protein Co...	P04114	55	GC02M020956	16.1
SREBF1	Sterol Reg...	Protein Co...	P36956	58	GC17M017810	15.94
LIVAR	Liver Cell...	RNA Gene (...)		16	GC18M070336	15.32
MIR27A	MicroRNA 27a	RNA Gene (...)		30	GC19M092209	14.96
MIR34A	MicroRNA 34a	RNA Gene (...)		29	GC01M014460	14.76
NR1H4	Nuclear Re...	Protein Co...	Q96RI1	56	GC12P100473	14.17
...



Figure 2 (下方图) 为图 Overall targets number of datasets 概览。

(对应文件为 Figure+Table/Overall-targets-number-of-datasets.pdf)

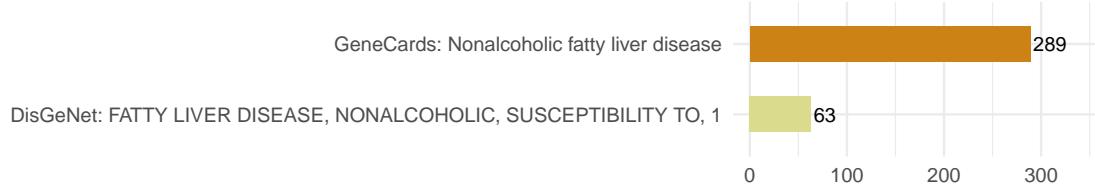


Figure 2: Overall targets number of datasets



6.5 成分-疾病-靶点

R input

```
hb <- map(hb, gm@params$lst.genes, name = "dis", less.label = F)
hb@params$p.pharm2dis
hb@params$p.venn2dis
```



Figure 3 (下方图) 为图 Network pharmacology with disease 概览。

(对应文件为 Figure+Table/Network-pharmacology-with-disease.pdf)

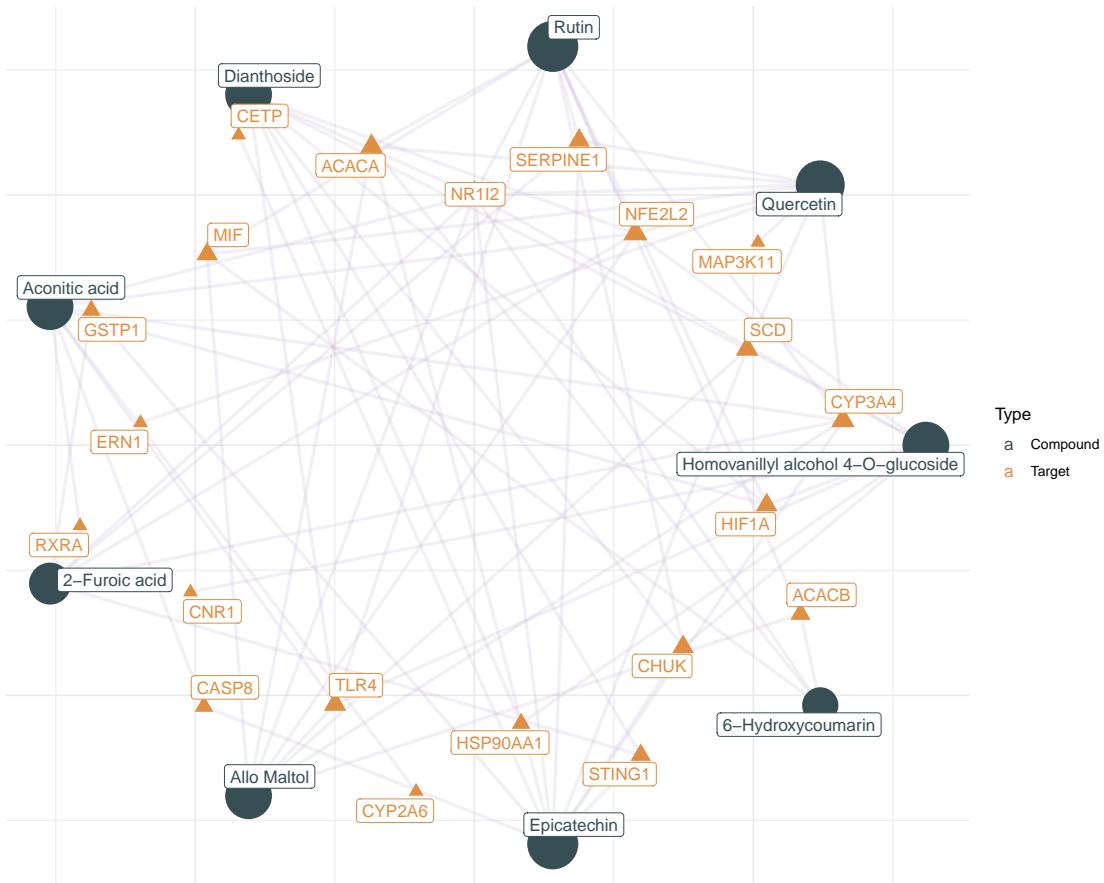


Figure 3: Network pharmacology with disease

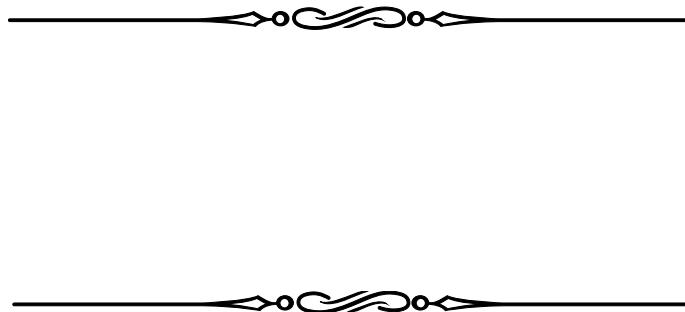


Figure 4 (下方图) 为图 Targets intersect with targets of diseases 概览。

(对应文件为 Figure+Table/Targets-intersect-with-targets-of-diseases.pdf)

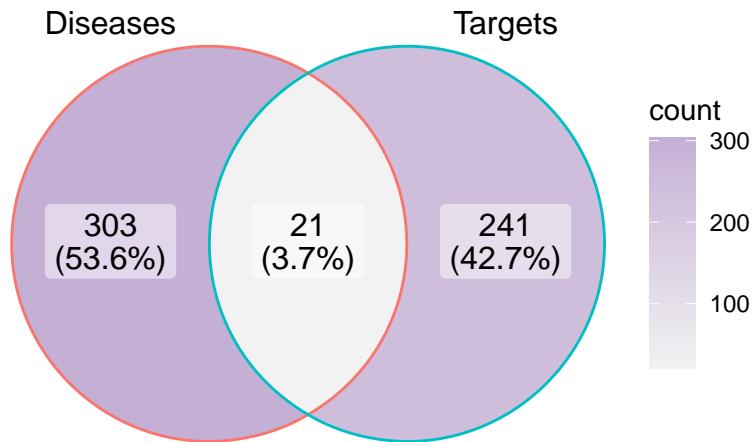


Figure 4: Targets intersect with targets of diseases



All_intersection :

TLR4, NFE2L2, CASP8, ACACA, SERPINE1, ERN1, NR1I2, RXRA, MAP3K11, SCD, ACACB, CYP3A4, CNR1, CETP, CHUK, HIF1A, MIF, STING1, HSP90AA1, GSTP1, CYP2A6

(上述信息框内容已保存至 [Figure+Table/Targets-intersect-with-targets-of-diseases-content](#))

6.6 分子对接

R input

```
dockLayout <- dplyr::select(hb@params$p.pharm2dis$.data, -1)
dockLayout <- map(dockLayout, "Ingredient.name", infoCpd, "Identifier", "CID", col = "CID")
dockLayout

vn <- job_vina(.layout = dockLayout)
vn <- step1(vn)
vn <- step2(vn)
vn <- step3(vn)
vn <- set_remote(vn)
vn <- step4(vn)
vn <- step5(vn)
wrap(vn@plots$step5$p.res_vina, 7, 9.5)
vn@tables$step5$res_dock
vn <- step6(vn, top = 3)
vn@plots$step6$Top1_643757_into_1tqn
vn@plots$step6$Top2_99477_into_2dn8
vn@plots$step6$Top3_5316639_into_4cgv
vn <- step7(vn)
vn@plots$step7$Top1_643757_into_1tqn
vn@plots$step7$Top2_99477_into_2dn8
vn@plots$step7$Top3_5316639_into_4cgv
```



Figure 5 (下方图) 为图 Overall combining Affinity 概览。

(对应文件为 Figure+Table/Overall-combining-Affinity.pdf)

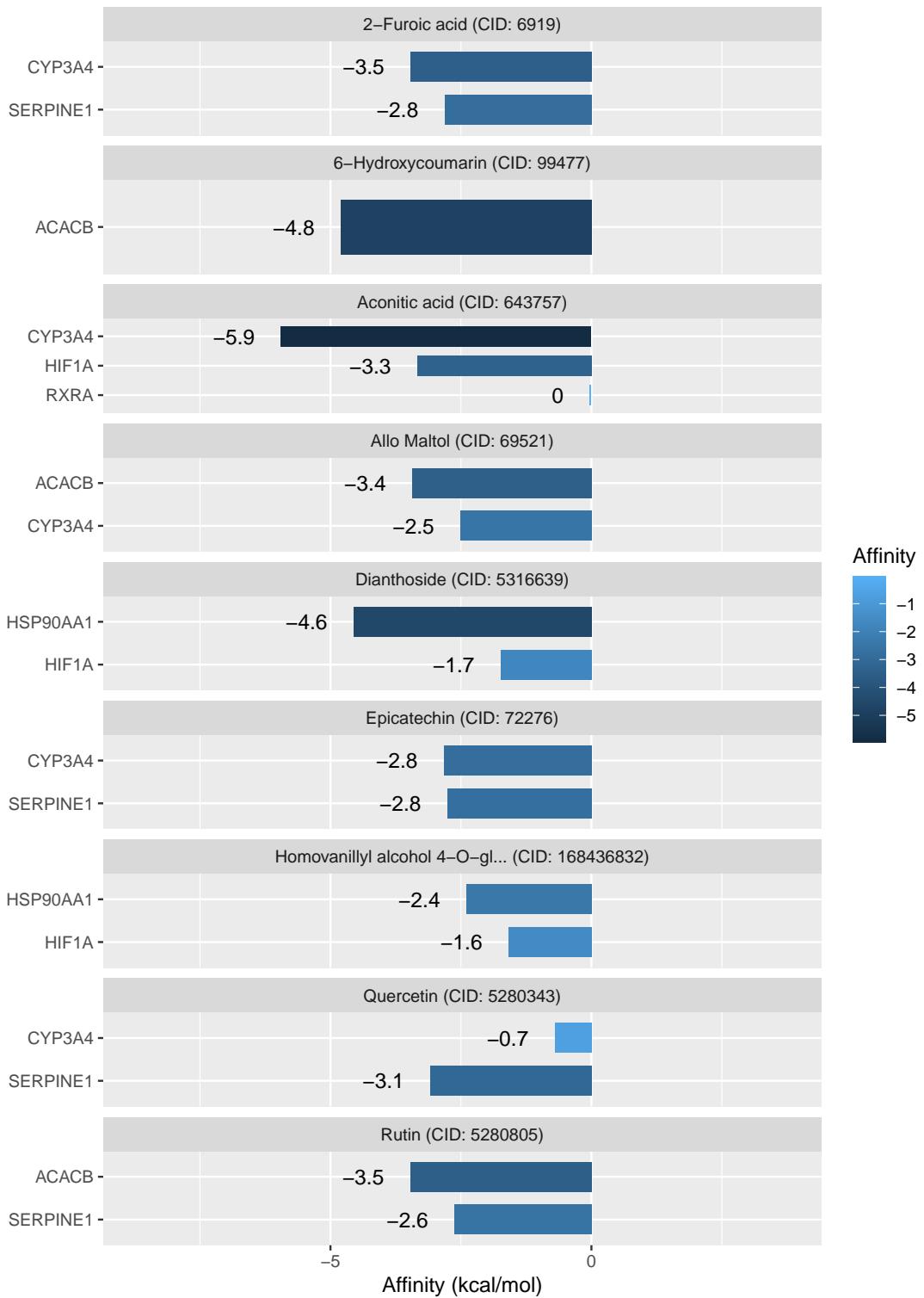


Figure 5: Overall combining Affinity

Table 3 (下方表格) 为表格 Affinity data 概览。

(对应文件为 **Figure+Table/Affinity-data.csv**)

注: 表格共有 36 行 8 列, 以下预览的表格可能省略部分数据; 含有 9 个唯一 ‘PubChem_id’; 含有 13 个唯一 ‘hgnc_symbol’。

1. hgnc_symbol: 基因名 (Human)

Table 3: Affinity data

PubChe...	PDB_ID	Affinity	dir	file	Combn	hgnc_s...	Ingred...
643757	1tqn	-5.949	vina_s...	vina_s...	643757...	CYP3A4	Aconit...
99477	2dn8	-4.804	vina_s...	vina_s...	99477_...	ACACB	6-Hydr...
5316639	4cgv	-4.559	vina_s...	vina_s...	531663...	HSP90AA1	Dianth...
5280805	2dn8	-3.473	vina_s...	vina_s...	528080...	ACACB	Rutin
6919	1tqn	-3.464	vina_s...	vina_s...	6919_i...	CYP3A4	2-Furo...
69521	2dn8	-3.435	vina_s...	vina_s...	69521_...	ACACB	Allo M...
643757	4h6j	-3.338	vina_s...	vina_s...	643757...	HIF1A	Aconit...
5280343	5brr	-3.087	vina_s...	vina_s...	528034...	SERPINE1	Quercetin
72276	1tqn	-2.828	vina_s...	vina_s...	72276_...	CYP3A4	Epicat...
6919	5brr	-2.814	vina_s...	vina_s...	6919_i...	SERPINE1	2-Furo...
72276	5brr	-2.762	vina_s...	vina_s...	72276_...	SERPINE1	Epicat...
5280805	5brr	-2.624	vina_s...	vina_s...	528080...	SERPINE1	Rutin
69521	1tqn	-2.508	vina_s...	vina_s...	69521_...	CYP3A4	Allo M...
168436832	4cgv	-2.398	vina_s...	vina_s...	168436...	HSP90AA1	Homova...
5316639	4h6j	-1.74	vina_s...	vina_s...	531663...	HIF1A	Dianth...
...

6.6.1 对接可视化

Figure 6 (下方图) 为图 Docking 643757 into 1tqn 概览。

(对应文件为 Figure+Table/Docking-643757-into-1tqn.png)



Figure 6: Docking 643757 into 1tqn

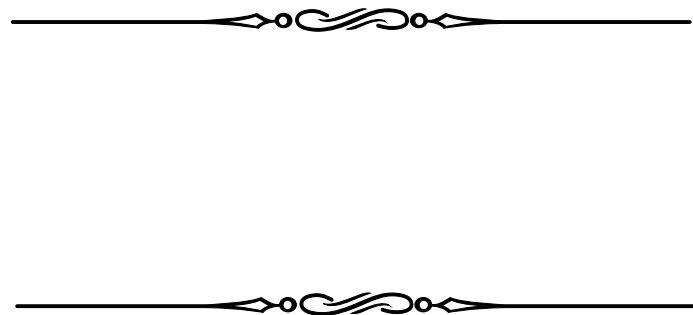


Figure 7 (下方图) 为图 Docking 99477 into 2dn8 概览。

(对应文件为 Figure+Table/Docking-99477-into-2dn8.png)

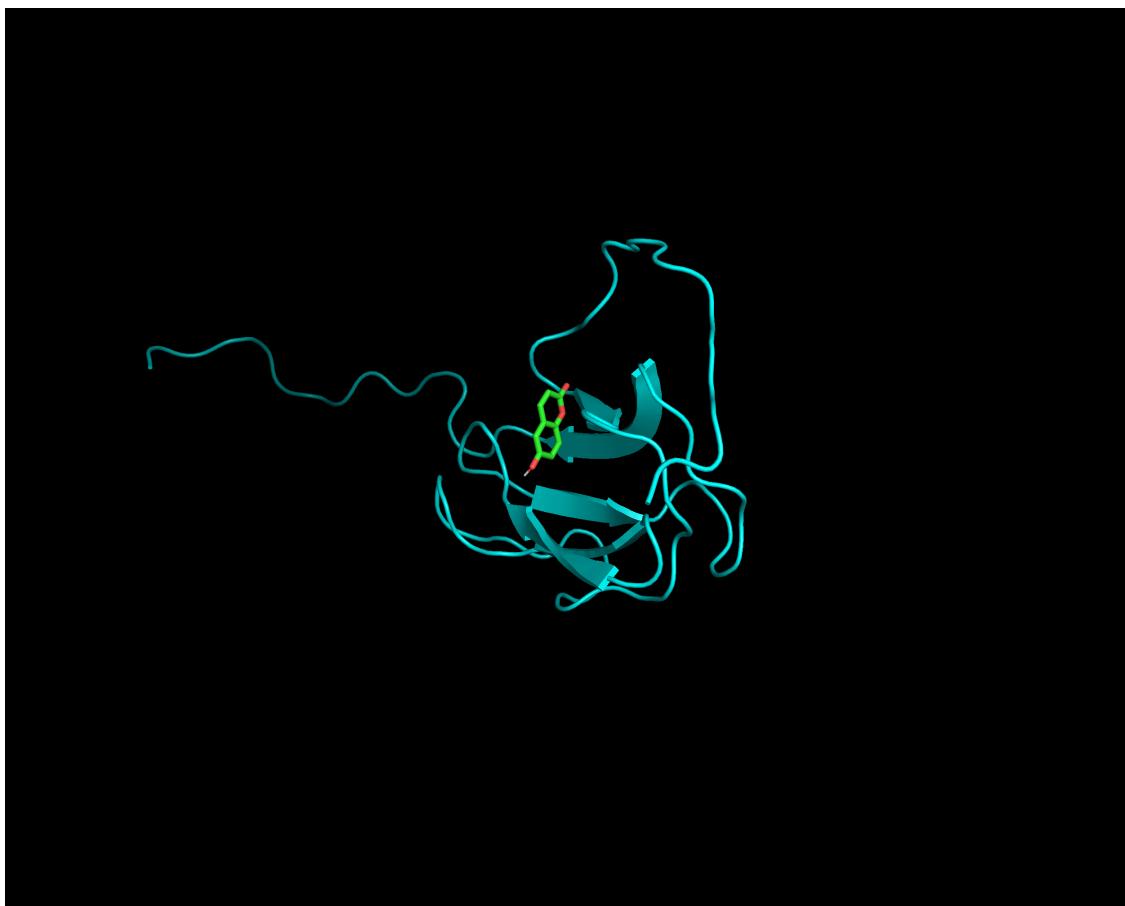


Figure 7: Docking 99477 into 2dn8

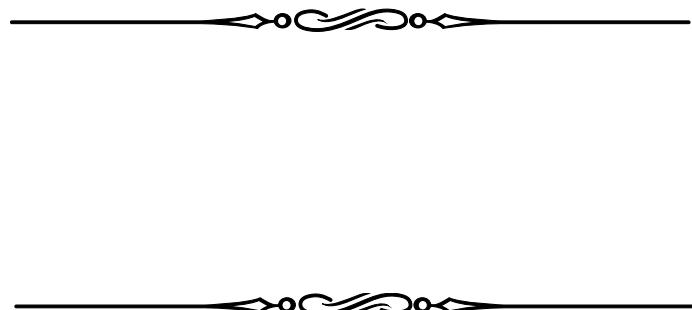


Figure 8 (下方图) 为图 Docking 5316639 into 4cgv 概览。

(对应文件为 Figure+Table/Docking-5316639-into-4cgv.png)

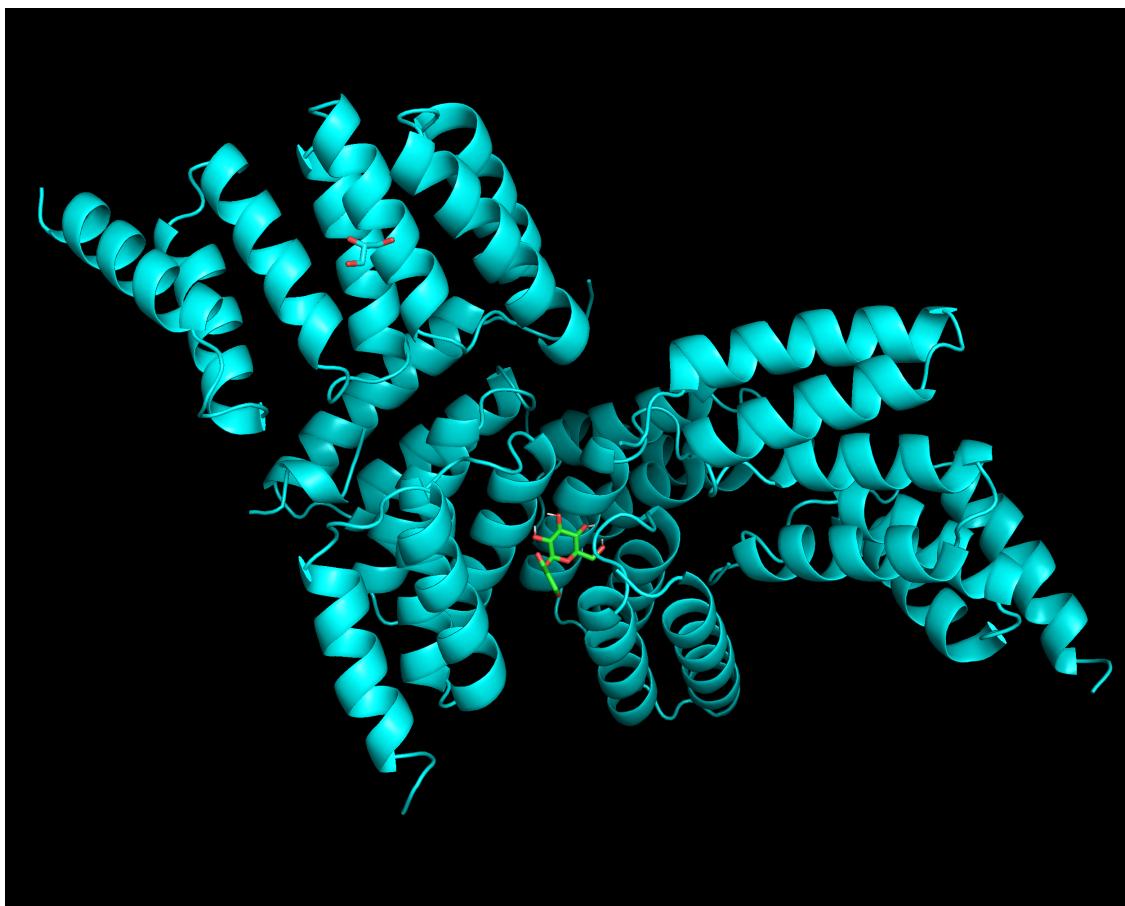


Figure 8: Docking 5316639 into 4cgv



6.6.2 对接细节可视化



Figure 9 (下方图) 为图 Docking 643757 into 1tqn detail 概览。

(对应文件为 Figure+Table/Docking-643757-into-1tqn-detail.png)

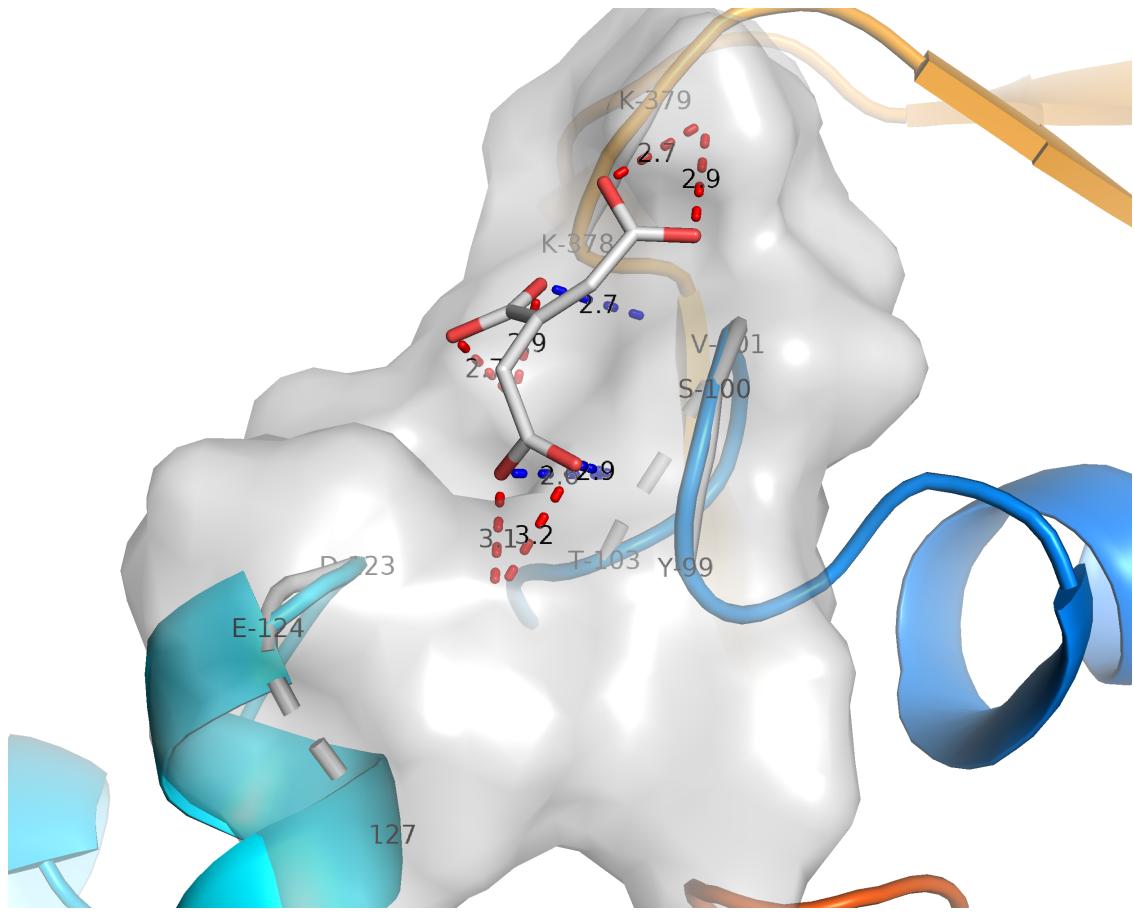


Figure 9: Docking 643757 into 1tqn detail

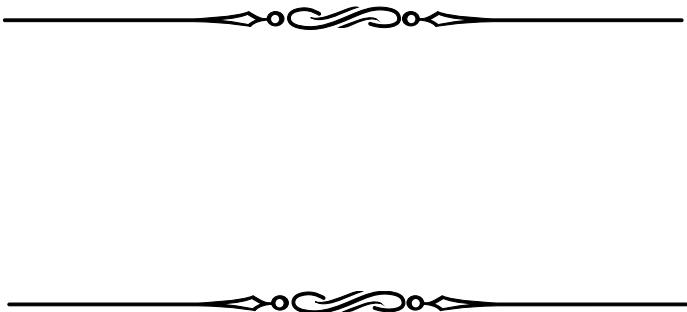


Figure 10 (下方图) 为图 Docking 99477 into 2dn8 detail 概览。

(对应文件为 Figure+Table/Docking-99477-into-2dn8-detail.png)

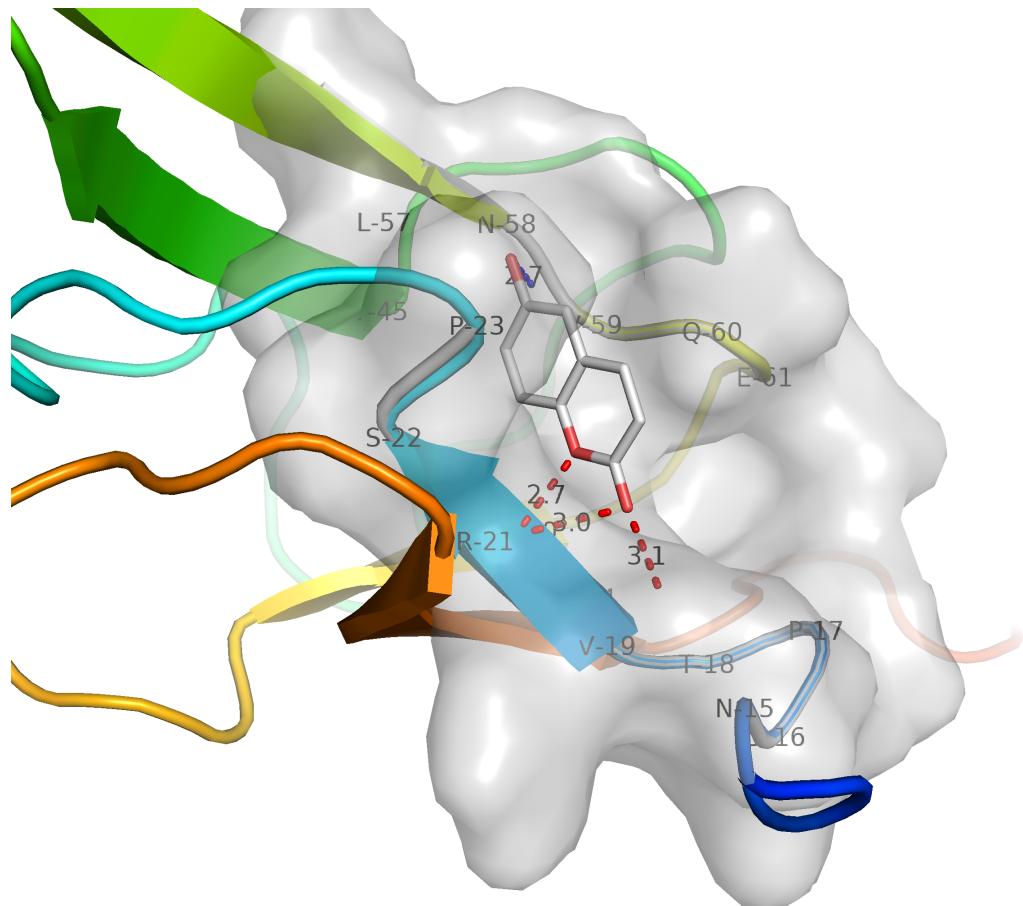


Figure 10: Docking 99477 into 2dn8 detail

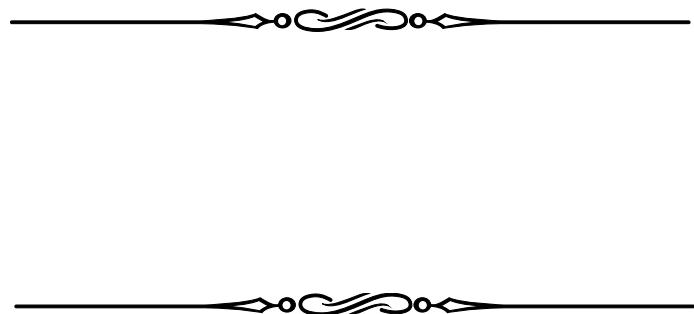


Figure 11 (下方图) 为图 Docking 5316639 into 4cgv detail 概览。

(对应文件为 Figure+Table/Docking-5316639-into-4cgv-detail.png)

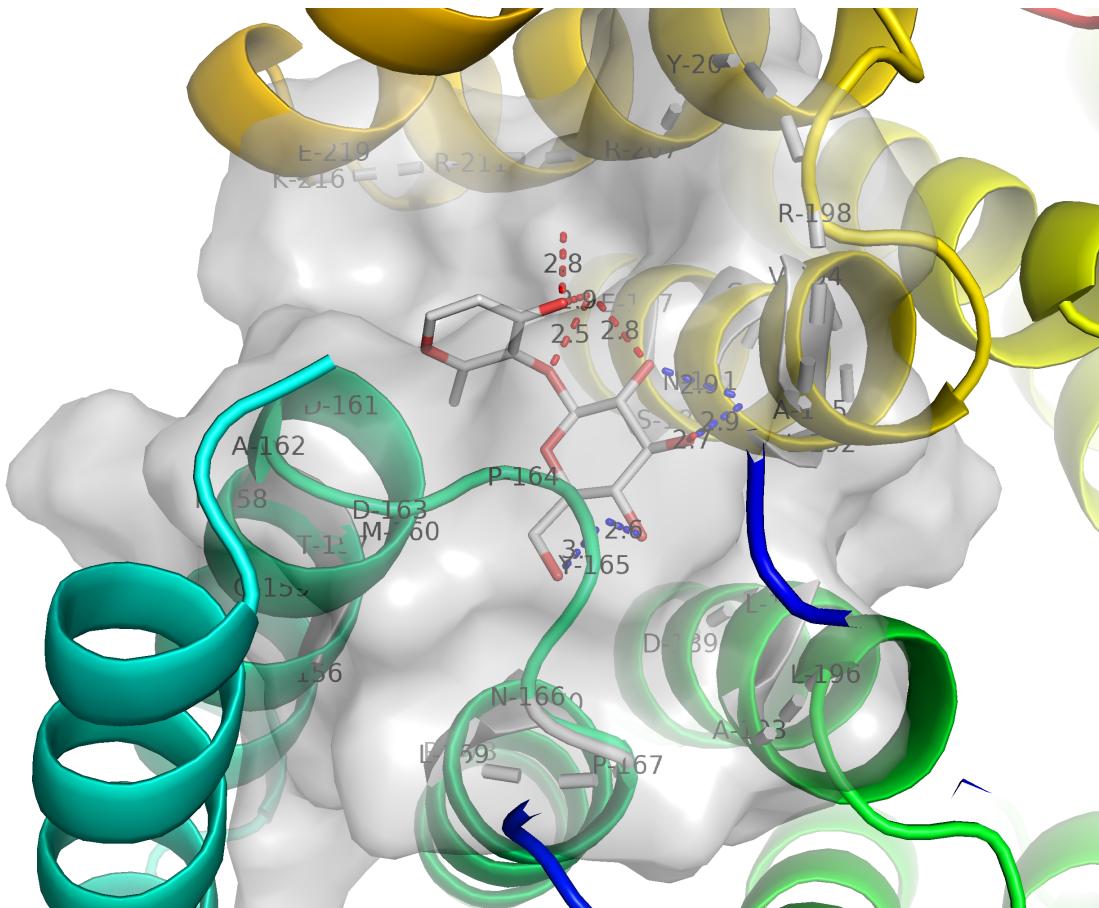


Figure 11: Docking 5316639 into 4cgv detail



Reference

1. Piñero, J. *et al.* The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* (2019) doi:10.1093/nar/gkz1021.
2. Stelzer, G. *et al.* The genecards suite: From gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics* **54**, 1.30.1–1.30.33 (2016).
3. Barbarino, J. M., Whirl-Carrillo, M., Altman, R. B. & Klein, T. E. PharmGKB: A worldwide resource for pharmacogenomic information. *Wiley interdisciplinary reviews. Systems biology and medicine* **10**, (2018).
4. Fang, S. *et al.* HERB: A high-throughput experiment- and reference-guided database of traditional chinese medicine. *Nucleic Acids Research* **49**, D1197–D1206 (2021).
5. Nickel, J. *et al.* SuperPred: Update on drug classification and target prediction. *Nucleic acids research* **42**, W26–W31 (2014).

6. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling* **61**, 3891–3898 (2021).
7. Zhang, Y., Forli, S., Omelchenko, A. & Sanner, M. F. AutoGridFR: Improvements on autodock affinity maps and associated software tools. *Journal of computational chemistry* **40**, 2882–2886 (2019).
8. Zhang, Y. & Sanner, M. F. AutoDock crankpep: Combining folding and docking to predict protein-peptide complexes. *Bioinformatics (Oxford, England)* **35**, 5121–5127 (2019).
9. Ravindranath, P. A. & Sanner, M. F. AutoSite: An automated approach for pseudo-ligands prediction-from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics (Oxford, England)* **32**, 3142–3149 (2016).
10. Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J. & Sanner, M. F. AutoDockFR: Advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS computational biology* **11**, (2015).
11. Cao, Y., Charisi, A., Cheng, L.-C., Jiang, T. & Girke, T. ChemmineR: A compound mining framework for r. *Bioinformatics (Oxford, England)* **24**, 1733–1734 (2008).