Prediction of Drug Targets by Using a Curated Pharmacophore Database

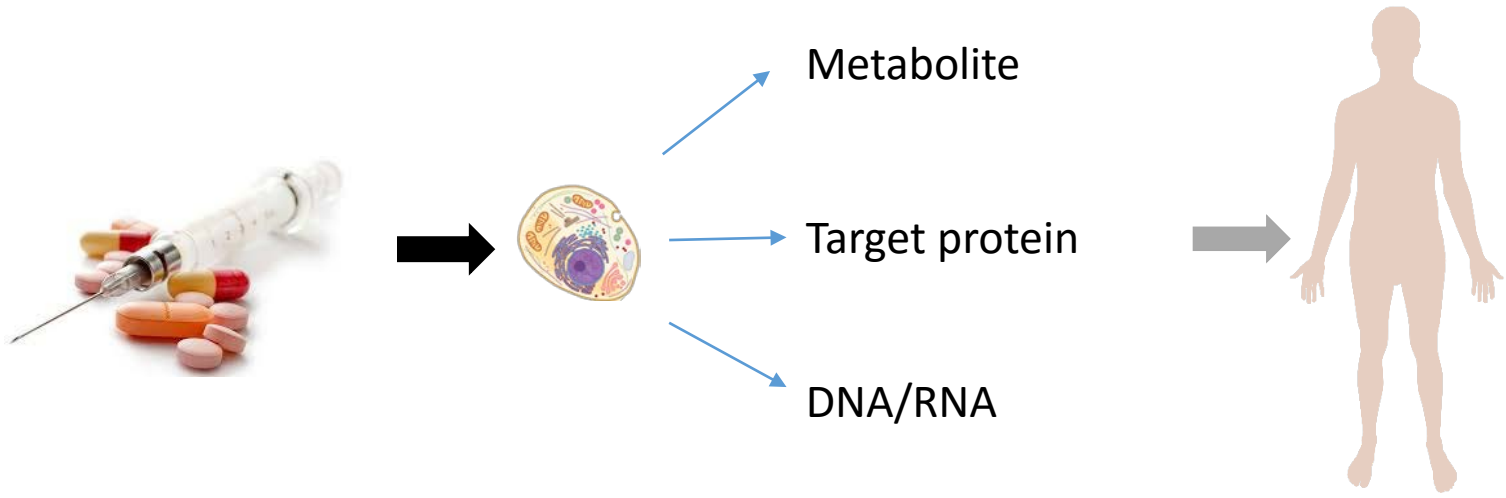
Ying-Ta Wu

The Genomics Research Center, Academia Sinica
Room 3A08 GRC Building, 128 Academia Road Section 2, Nangang, Taipei 115, Taiwan
ywu@gate.sinica.edu.tw

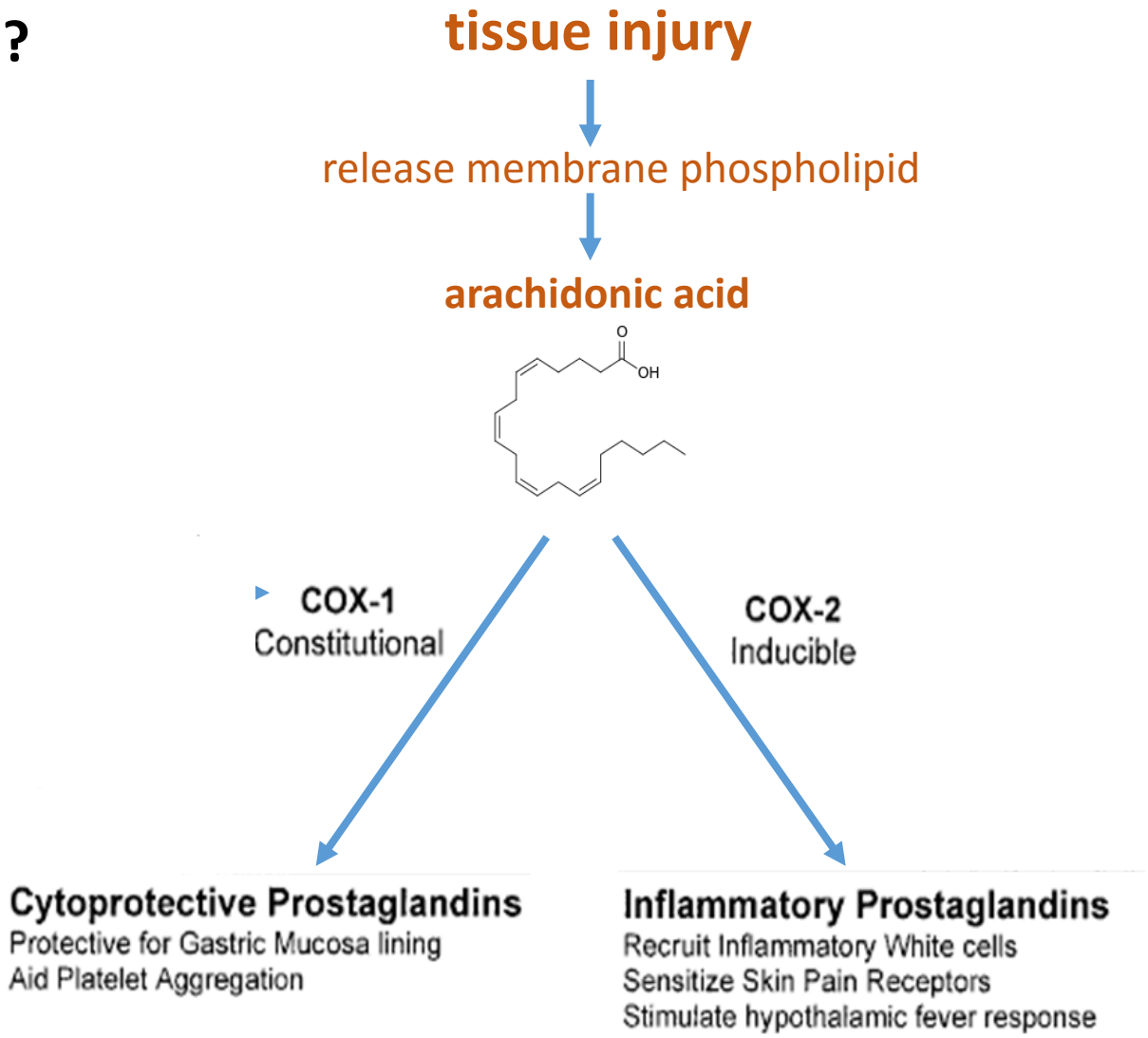
Identification of biological targets for an understanding of their mode of action is crucial for advancing a natural product to a modern medicine, which is effective and safe in treatment of human diseases. A curated 3D-pharmacophore database of therapeutic protein targets of human diseases and pathogens was generated for predicting pharmacological targets of natural products. The features of the pharmacophore models of disease-related proteins were created based on critical contactable residues participating in selectively binding to ligands. The usefulness of this database was demonstrated by a variety of compounds to identify their potential biological counterparts, which could provide useful clues for mode of action.

Good or Bad ?

- Any other targets, resulting to side effects or multiple indications ?



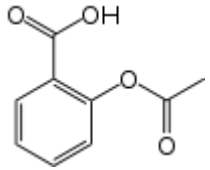
Pain Relief ?



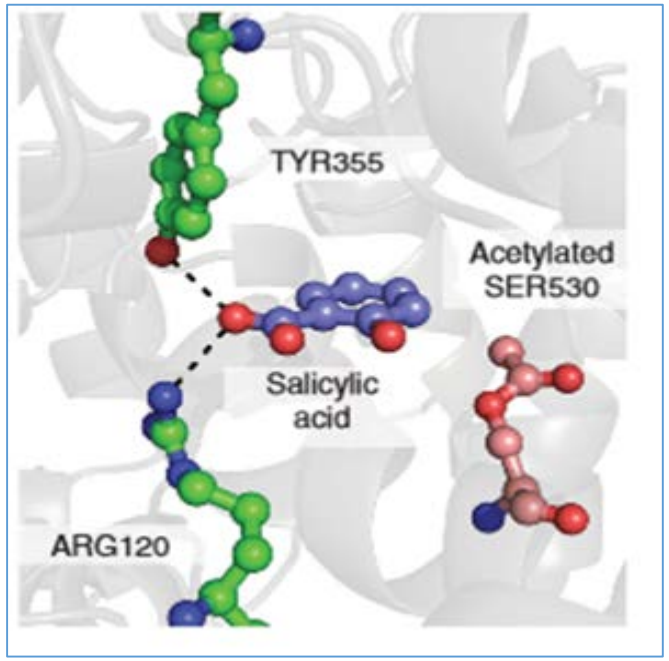
- John Robert Vane (1982 Nobel Prize)

COX-1 mediates the synthesis of prostaglandins responsible for protection of the stomach lining, while COX-2 mediates the synthesis of prostaglandins responsible for pain and inflammation. By creating "selective" NSAIDs that inhibit COX-2, but not COX-1, the same pain relief as traditional NSAIDs is offered, but with greatly reduced risk of fatal or debilitating peptic ulcers.

Acetylsalicylic acid



Aspirin is one of the most widely used medications in the world with an estimated 40K tones of it being consumed each year.



Aspirin ⊣ **COX-1** Constitutional

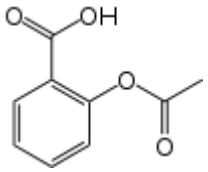
X **Cytoprotective Prostaglandins**
Protective for Gastric Mucosa lining
Aid Platelet Aggregation

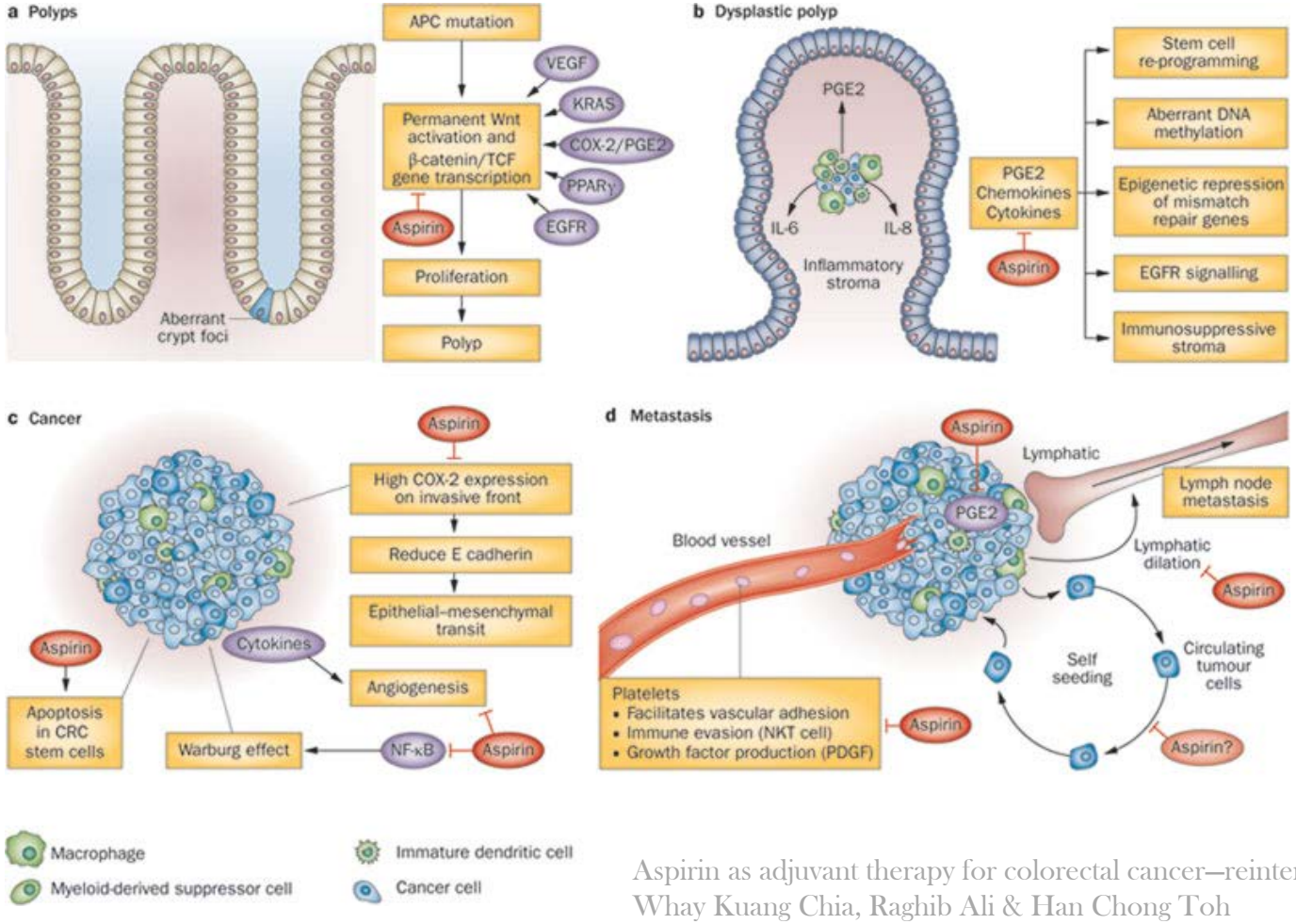
COX-2 Inducible ⊣ Aspirin

X **Inflammatory Prostaglandins**
Recruit Inflammatory White cells
Sensitize Skin Pain Receptors
Stimulate hypothalamic fever response

- Aspirin exerts activities at multiple points along carcinogenesis pathway.

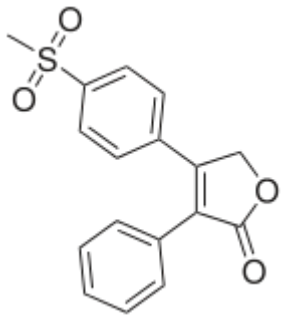
Acetylsalicylic acid





Aspirin as adjuvant therapy for colorectal cancer—reinterpreting paradigms
Whay Kuang Chia, Raghib Ali & Han Chong Toh
Nature Reviews Clinical Oncology 9, 561-570 (October 2012)
doi:10.1038/nrclinonc.2012.137

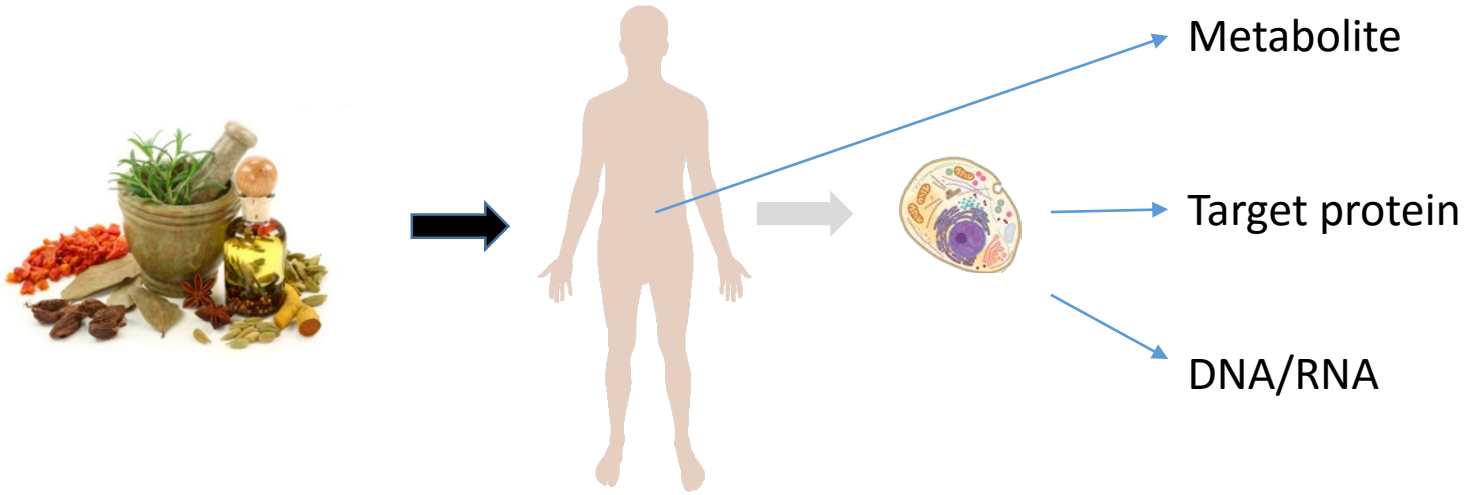
Rofecoxib



- Rofecoxib is **a selective COX-2 inhibitor**, or "coxib“, which offers pain relief effect but could reduced incidence of gastric ulceration.
- a nonsteroidal anti-inflammatory drug (NSAID) marketed under the brand name **Vioxx** by **Merck** to treat osteoarthritis, acute pain conditions, and dysmenorrhea.
- Rofecoxib was approved by FDA in 1999 and market, but was then withdrew in 2004, because of concerns about increased risk of heart attack and stroke associated with long-term, high-dosage use, resulting in between 88,000 and 140,000 cases of serious heart disease.
- In the year before withdrawal, Merck had sales revenue of US$2.5 billion from Vioxx. Merck reserved $970 million to pay for its Vioxx-related legal expenses through 2007, and has set aside $4.85bn for legal claims from US citizens.

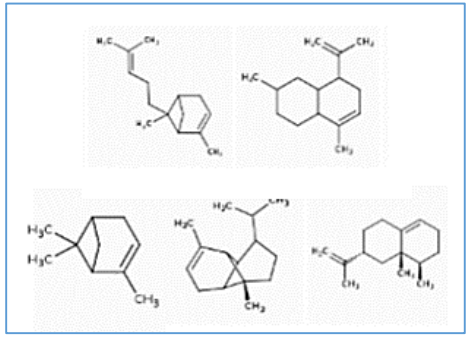
Modern Medicine Aspect

- How much do we know the effectors in the traditional remedy ?
- How do the active components work to cure diseases or “kill” invaders ?



A Modern Medicine Prospect

conventional indication



modern indication

disease codes

Active components

experimental approach

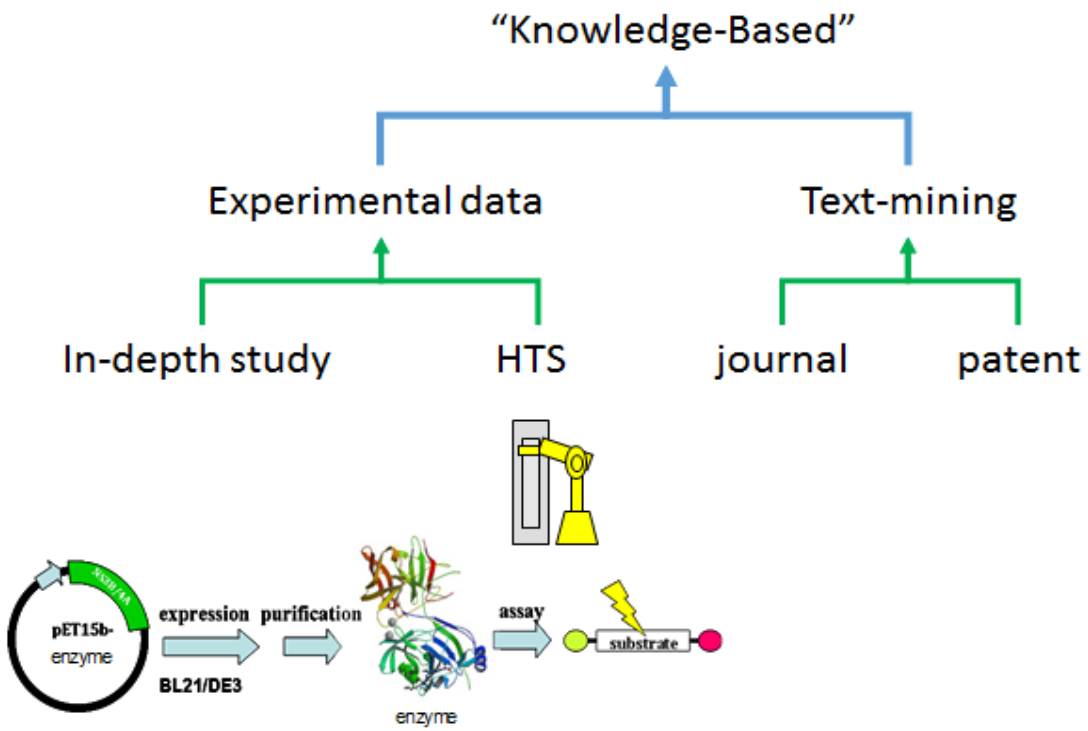
computational approach

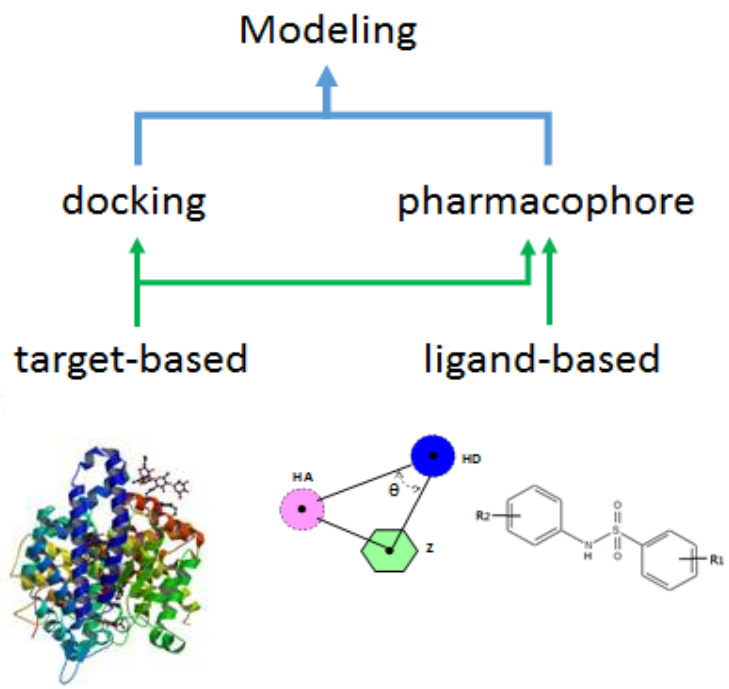
targets

pathways

diseases

Approaches for mapping drug-target relationship



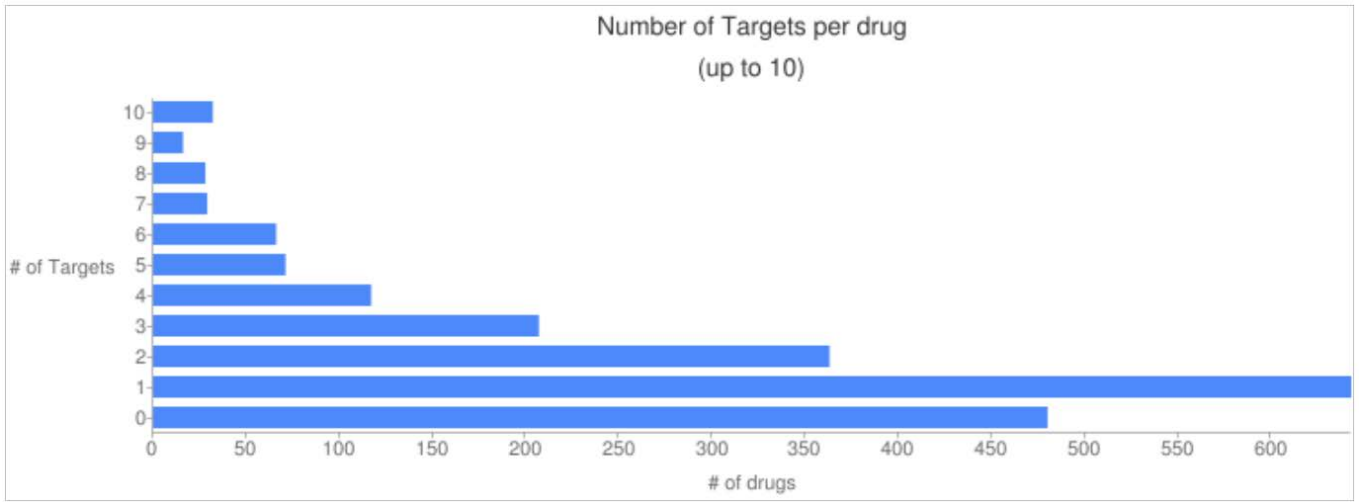


Drug Databases

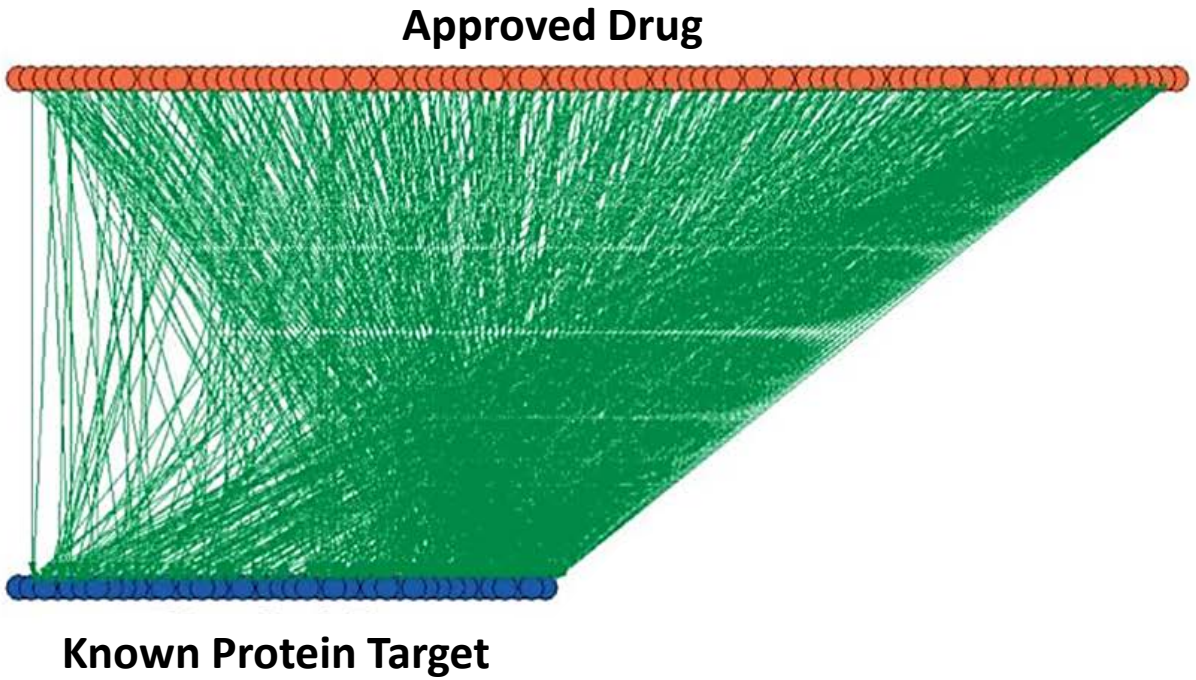
DrugBank	The DrugBank database is a blended bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains nearly 4800 drug entries including >1,350 FDA-approved small molecule drugs, 123 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals and >3,243 experimental drugs. DrugBank also contains extensive SNP-drug data that is useful for pharmacogenomics studies.
Therapeutic Target DB	The Therapeutic Target Database (TTD) is a drug database designed to provide information about the known therapeutic protein and nucleic acid targets described in the literature, the targeted disease conditions, the pathway information and the corresponding drugs/ligands directed at each of these targets. The database currently contains 1535 targets and 2107 drugs/ligands.
PharmGKB	The PharmGKB database is a central repository for genetic, genomic, molecular and cellular phenotype data and clinical information about people who have participated in pharmacogenomics research studies. The data includes, but is not limited to, clinical and basic pharmacokinetic and pharmacogenomic research in the cardiovascular, pulmonary, cancer, pathways, metabolic and transporter domains. Its aim is to aid researchers in understanding how genetic variation among individuals contributes to differences in reactions to drugs. PharmGKB contains searchable data on genes (>20,000), diseases (>3000), drugs (>2500) and pathways (53). It also has detailed information on 470 genetic variants (SNP data) affecting drug metabolism.
STITCH	STITCH ('search tool for interactions of chemicals') is a searchable database that integrates information about interactions from metabolic pathways, crystal structures, binding experiments and drug–target relationships. Text mining and chemical structure similarity is used to predict relations between chemicals. Each proposed interaction can be traced back to the original data sources. The database contains interaction information for over 68 000 different chemicals, including 2200 drugs, and connects them to 1.5 million genes across 373 genomes.
SuperTarget	SuperTarget is a database that contains a core dataset of about 7300 drug-target relations of which 4900 interactions have been subjected to a more extensive manual annotation effort. SuperTarget provides tools for 2D drug screening and sequence comparison of the targets. The database contains more than 2500 target proteins, which are annotated with about 7300 relations to 1500 drugs; the vast majority of entries have pointers to the respective literature source. A subset of 775 more extensively annotated drugs is provided separately through the Matador database (Manually Annotated Targets And Drugs Online Resource)

Target Statistics

Drug Group	Drug Type	Number of drug-target associations	Number of unique targets
Approved	Biotech	556	232
Approved	Small Molecule	5904	1679
Nutraceutical	Small Molecule	994	792
Experimental	Small Molecule	7602	2583
Illicit	Small Molecule	470	87
Withdrawn	Biotech	53	40
Withdrawn	Small Molecule	330	180
Investigational	Biotech	340	186
Investigational	Small Molecule	1868	780
		15475	4157

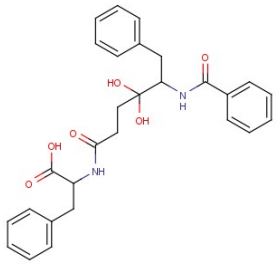


- Map active ingredients of drugs with known human gene targets

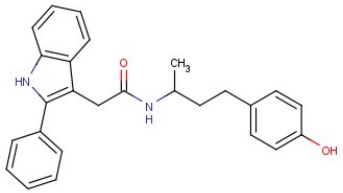


Examples: Query a known drug

Answer to “What could be the targets of a drug?”



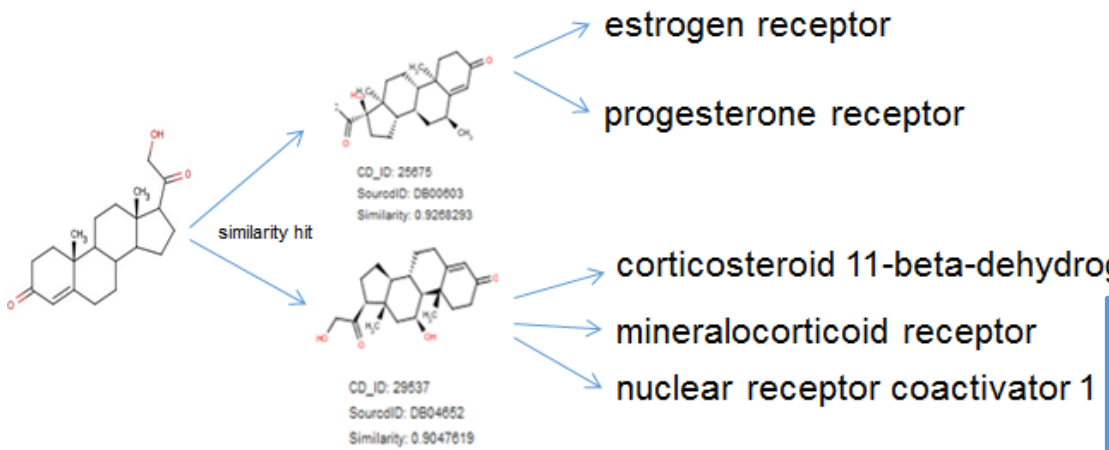
exact hit → angiotensin-converting enzyme

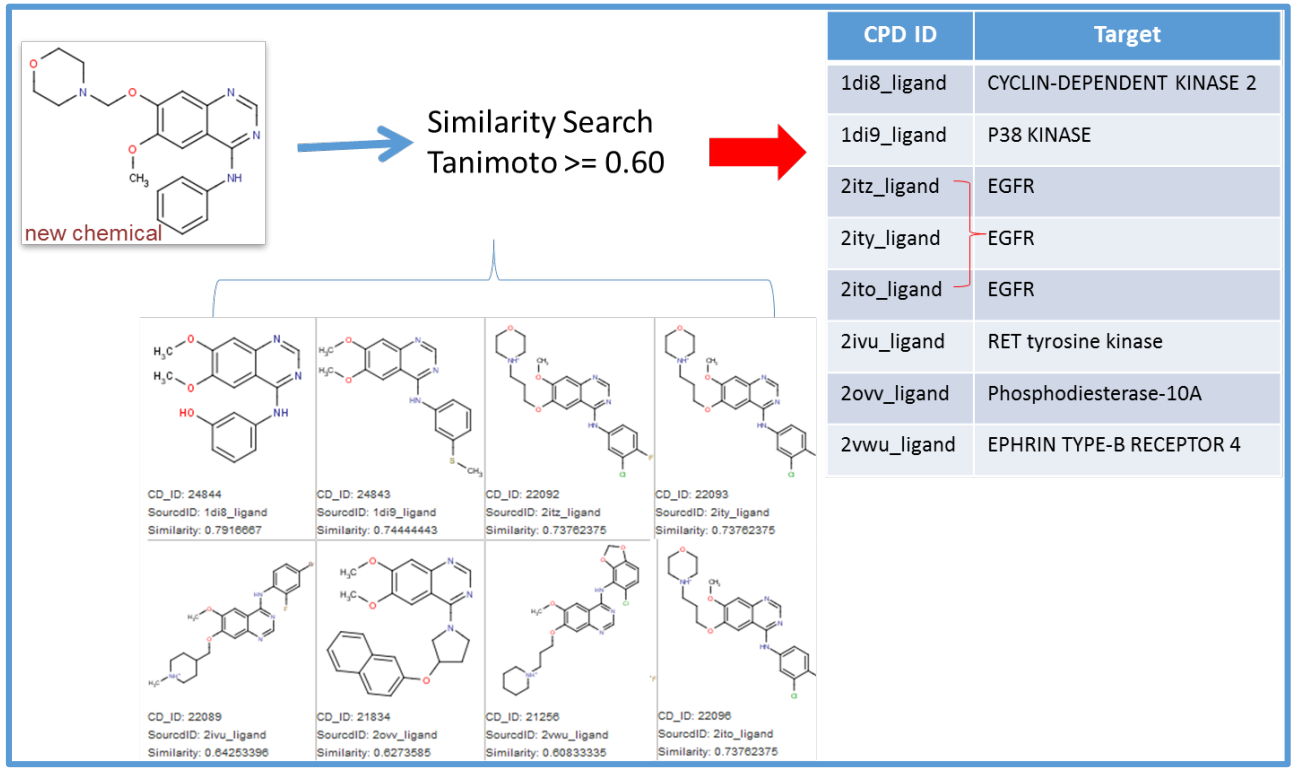


exact hit → estrogen receptor

Examples: Query a new compound

Answer to “What could be the targets of a new compound ?”





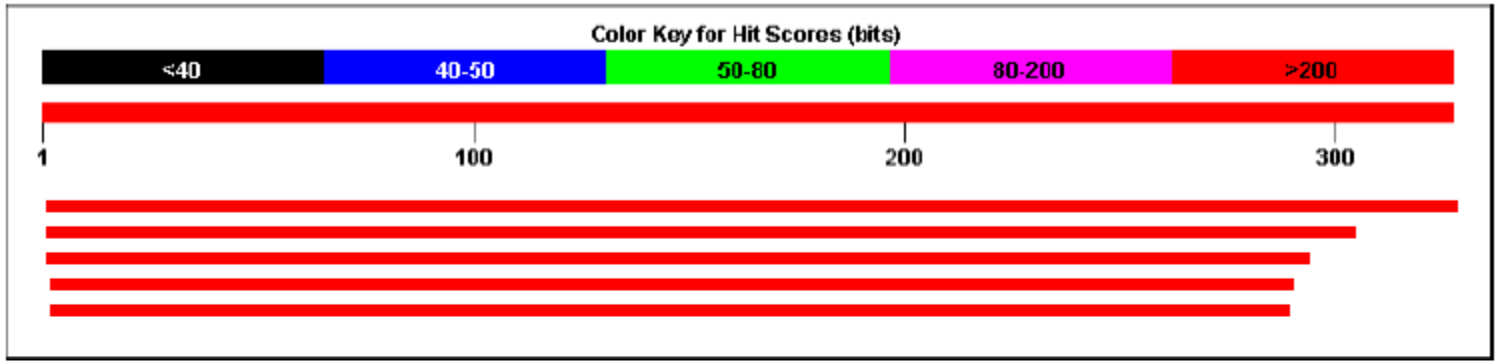
CPD ID	Target
1di8_ligand	CYCLIN-DEPENDENT KINASE 2
1di9_ligand	P38 KINASE
2itz_ligand	EGFR
2ity_ligand	EGFR
2ito_ligand	EGFR
2ivu_ligand	RET tyrosine kinase
2ovv_ligand	Phosphodiesterase-10A
2vwu_ligand	EPHRIN TYPE-B RECEPTOR 4

Examples: Query a protein sequence

Answer to “Any existing drugs affect this protein ?”

GEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVC
RLLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITD
FGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQ
PPICTIDVYMIMVKCWMIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADE

BLAST Search



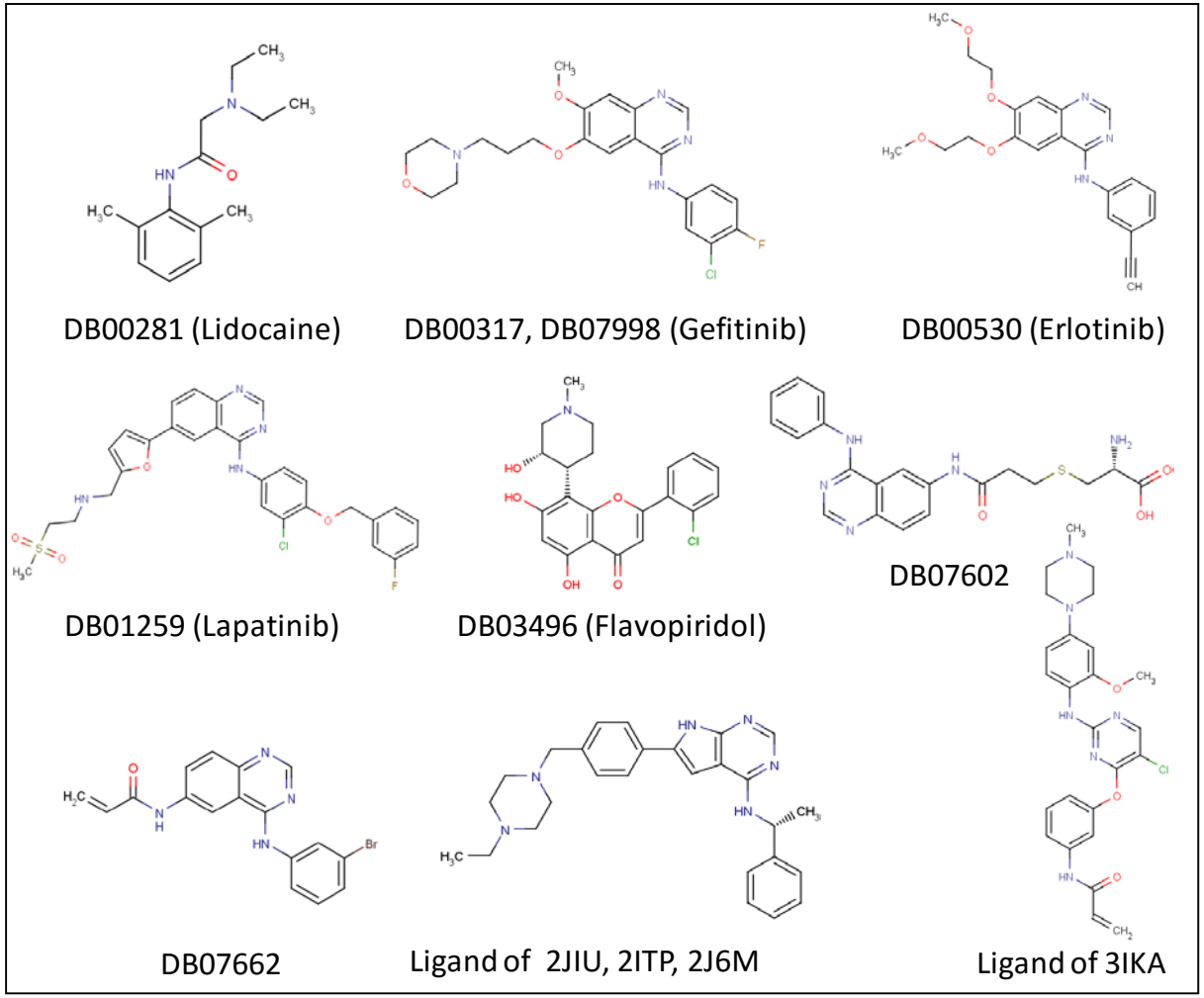
Search result

Hit score (bits) E-value

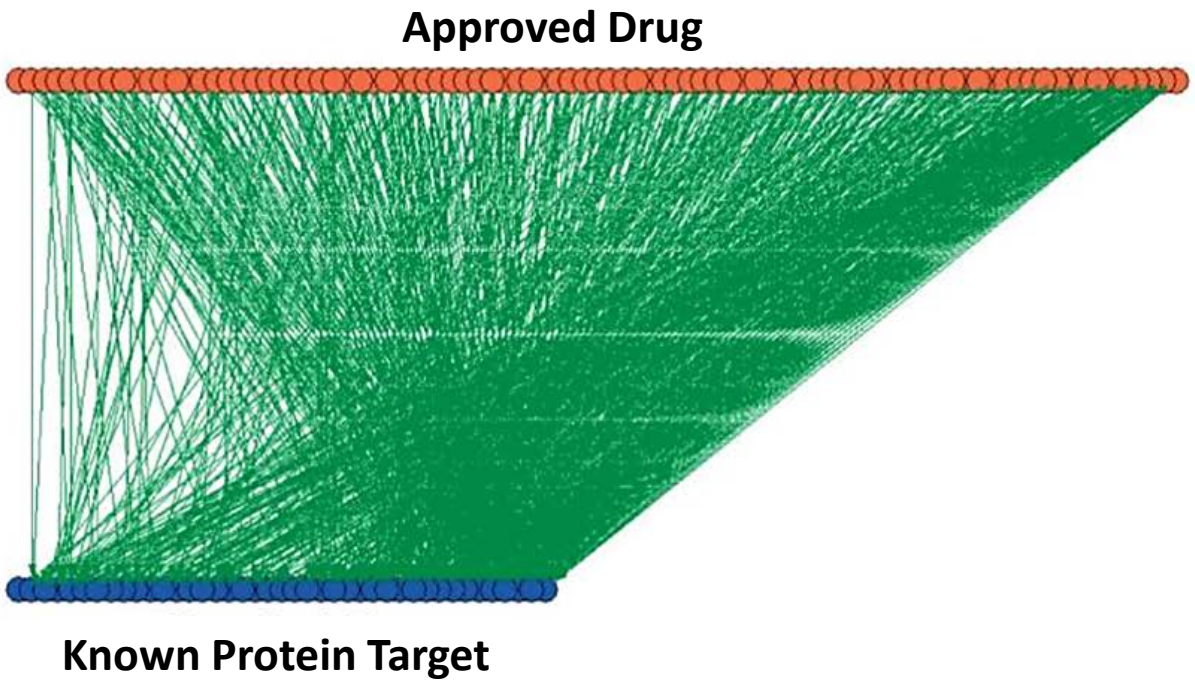
Query Accession Number	Query Description	Accession Number	Description	Hit Score
SEQUENCE	no description available	lcl|720_C:\Users\HTS\Desktop\blastdb\mer	Epidermal growth factor receptor (DB00002; DB00072; DB00281; DB00317; DB00530; DB01259; DB01269; DB03496; DB07602; DB07662; DB07998)	648.662
SEQUENCE	no description available	lcl|Sci_5016	(chain = A) 2JIU_PROTEIN	614.764
SEQUENCE	no description available	lcl|Sci_1105	(chain = A) 3IKA_PROTEIN	605.905
SEQUENCE	no description available	lcl|Sci_5181	(chain = A) 2ITP_PROTEIN	595.504
SEQUENCE	no description available	lcl|Sci_5125	(chain = A) 2J6M_PROTEIN	594.734

- drugs (or known bioactives) resulted from the previous query

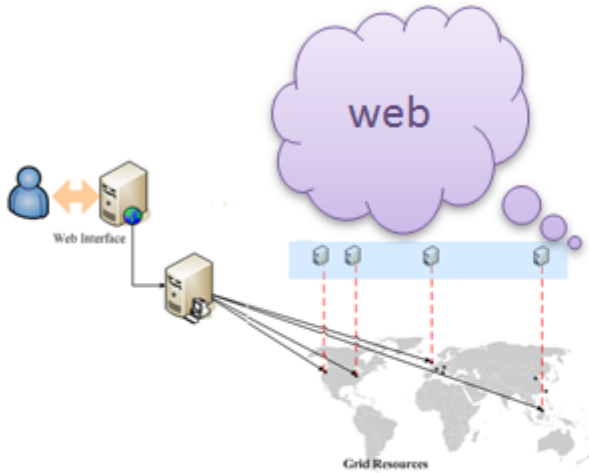
GEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEK
VKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVC
RLLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNW
CVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITD
FGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSD
VWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQ
PPICTIDVYMIMVKCWMIDADSRPKFRELIIEFSKMARDPQ
RYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADE

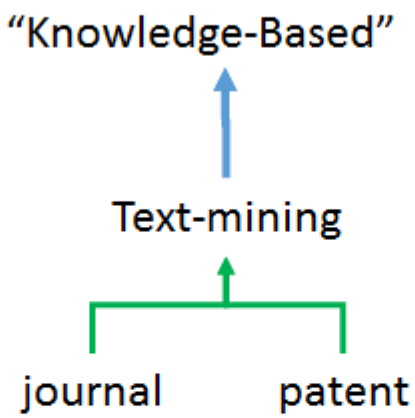


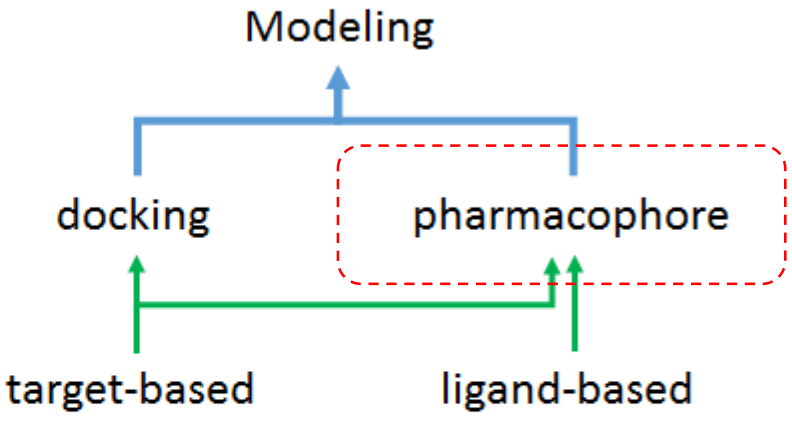
And more …

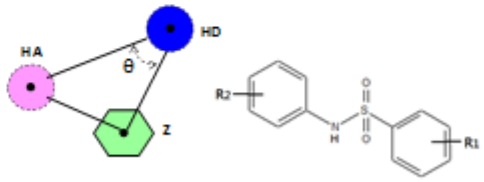


Computational Approach, beside experimental approach



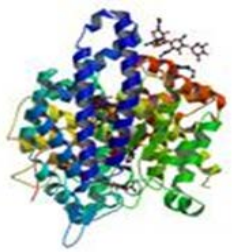


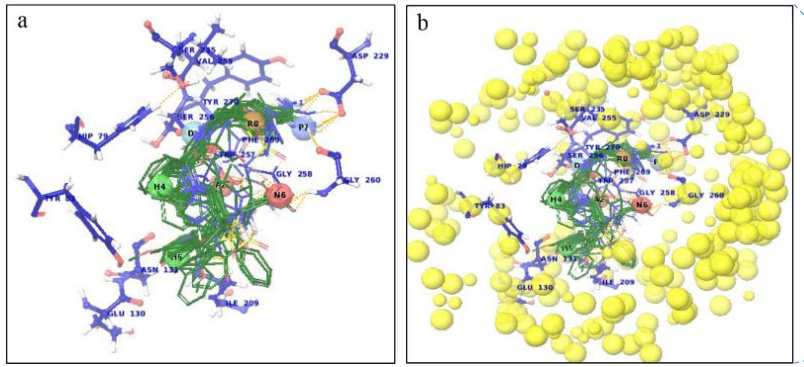




Establish Pharmacophore Database

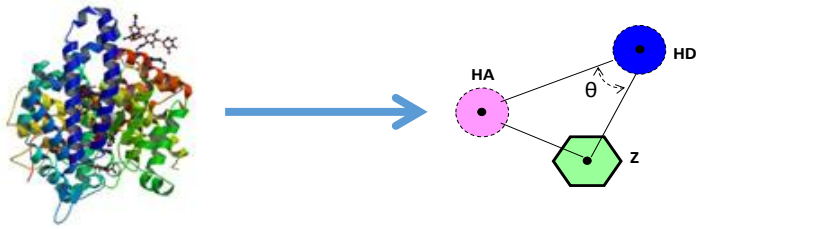
- Procedure

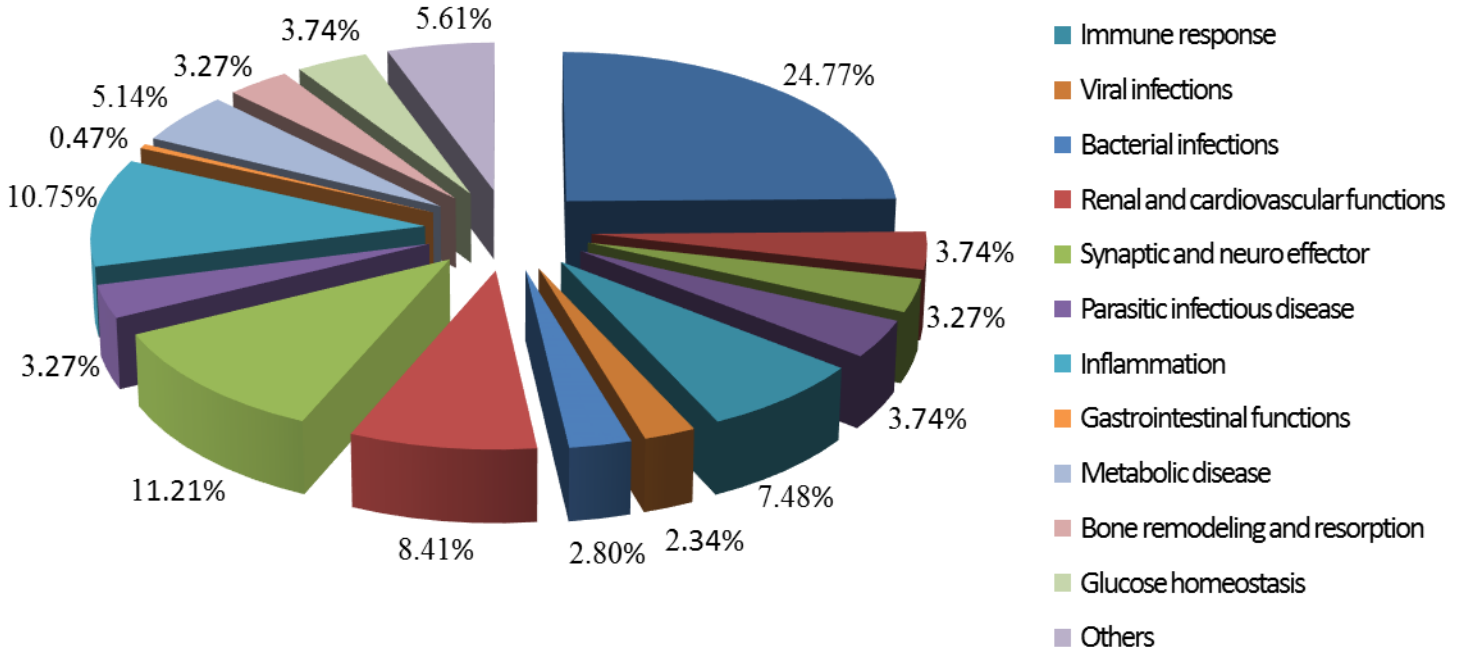




- **The Pharmacophore Database**
  - **Progress**

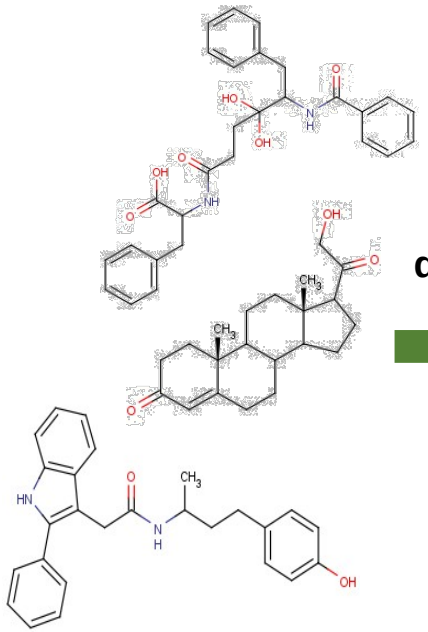
- enzyme
- receptor
- Factors/regulators/hormones
- binding proteins
- transport proteins
- nuclear receptor
- signaling protein
- Lipid binding proteins
- drug binding proteins
- transcriptional factors
- DNA binding
- Cell cycles, DNA repair
- nucleoside metabolism
- transcriptional regulation





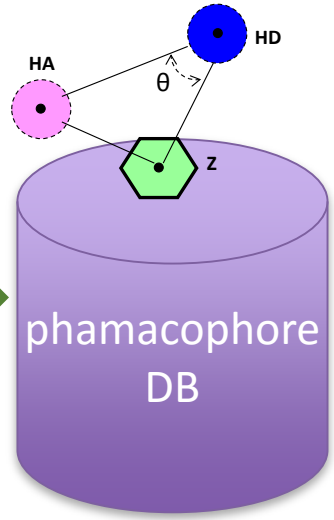
Examples: Query hit compounds

Answer to “What could be the targets of a hit compound ?”

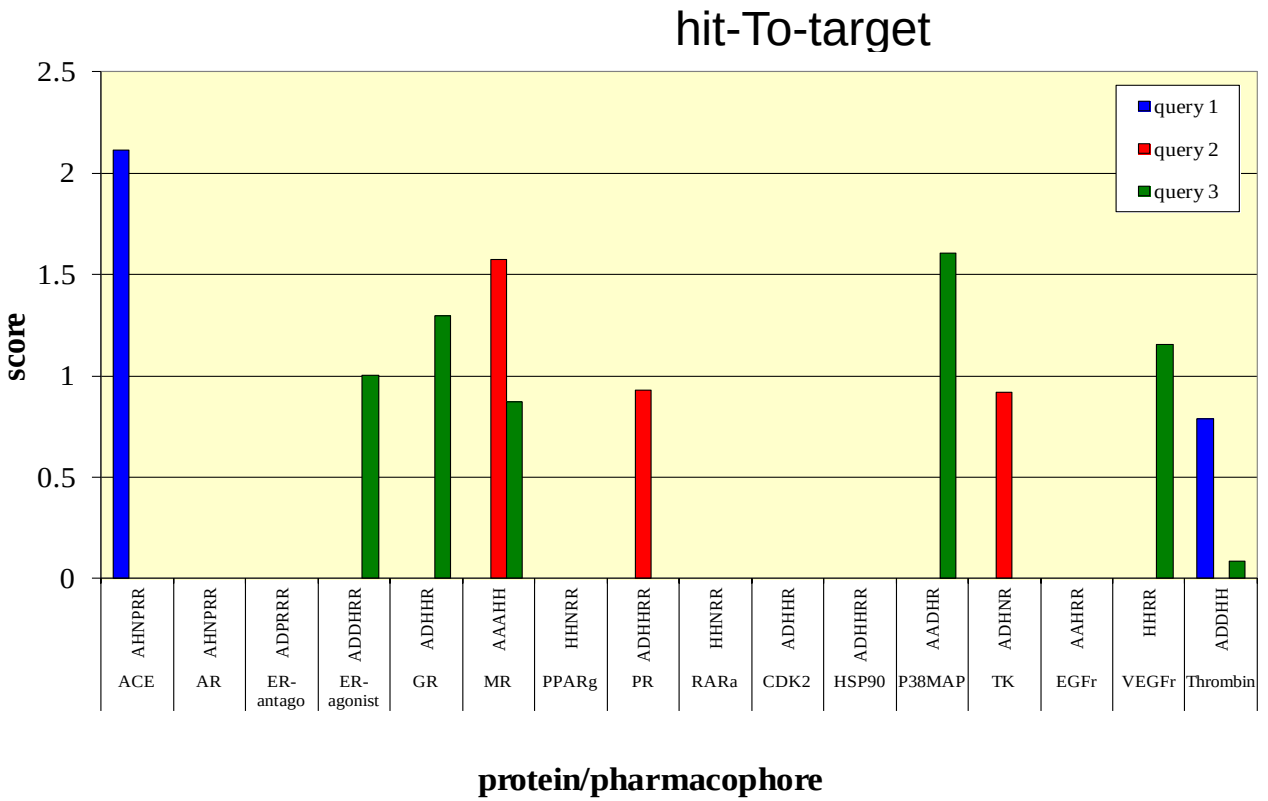


screening hits

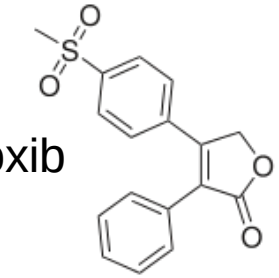
query

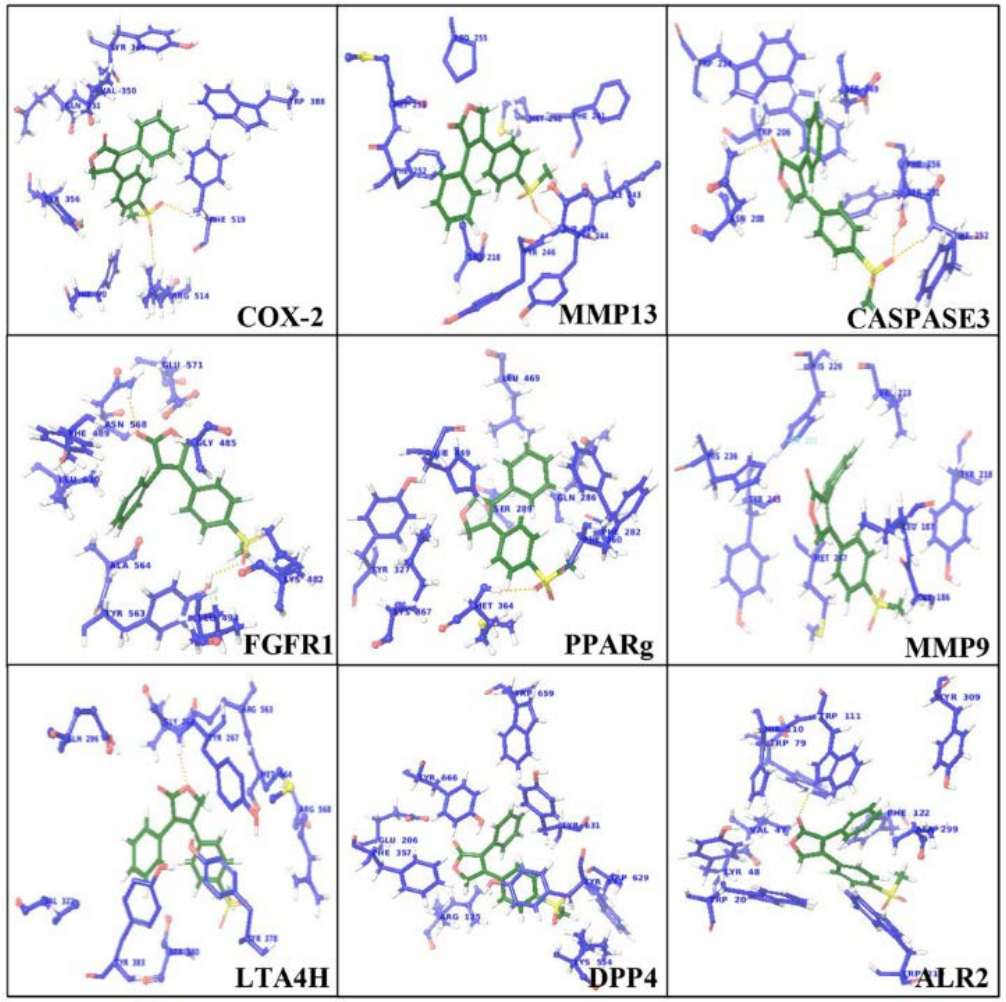


Targets



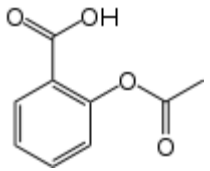
Examples: Query Rofecoxib





Drug target group	Protein	Clinical effect	Ref.
Drug targets	Cyclooxygenase-2 (COX2)	Reduce pain and inflammation	42,43
	Matrix Metalloproteinase 13 (MMP13)	Osteoarthritis treatment	44
Drug targets2	Caspase 3	Alzheimer's disease and neuronal effects	49,50
	Fibroblast growth factor receptor 1 (EGFR1)		56,57
	Transthyretin		51,53
	Peroxisome proliferator-activated receptor gamma (PPARg)		55
Drug side effect targets1	Matrix metalloproteinase 9 (MMP9)	Cardiovascular problem	59,60
	Leukotriene A4 hydrolase (LTA4H)		62,63,64
	Dipeptidyl peptidase 4 (DPP4)		67,68
Drug side effect targets2	Dipeptidyl peptidase 4 (DPP4)	Abnormal of energy glucose metabolism and some abnormal in immune regulation.	65,66
	Peroxisome proliferator-activated receptor gamma		73
	Aldose Reductase (ALR)		76,77
Drug transporter	Human serum albumin	Drug transport	78
Drug metabolize protein	Cytochrome P450 3A4	Drug metabolize	79
	Cytochrome P450 2C9		80
Other proteins	Trypanothione reductase, Acidic chitinase, Androgen receptor, Hepatocyte growth factor receptor c-Met, Enoyl-ACP Reductase, Retinoic X Receptor Alpha, Topoisomerase	Other clinical effects	

Examples: **Query** Aspirin



Comparison

search results from PharmMapper and GOSTAR databases

Our database	COX-2 **	COX-1**	Androgen Receptor	ER-agonist	ER-antagonist	Glucoticoid Receptor	MR	RXRa	CDK2
	EGFr	FGFr1**	P38 MAP**	Protein Kinase SRC **	Tymidine Kinase	VEGFr2	FXa	Thrombin	ACE
	COMT **	DHFR	AChE	GPB	Neuraminidase	InhA	PARP	PNP	HIVRT
	Aldose Reductase **	Alpha-Mannosidase II	Carbonic Anhydrase II	S-AHH	DPP4	MMP9	MMP13	AmpC	Aurora A Kinase
	CHK1	Caspase3	LDHA	Trypanothione reductase	PDHK	SARS 3C-like proteinase	HCV-NS5B	SGLT2	PDE5
	3-Dehydroquinase	AMCase	Human serum albumin **	Trypsin	GluR	Cathepsin K	CYP450 3A4	ERK2	Phospholipase A2 **
	β-Glucuronidase	Topoisomerase	CYP450 2C9	5-Lipoxygenase	B-Raf kinase	c-Met			
PharmMapper database	GTPase HRas	Phosphoenolpyruvate carboxykinase, cytosolic [GTP]	Aldose reductase **	RhoE	Aldo-keto reductase family 1 member C2	Tyr-protein kinase Src **	ACE	Tyr-protein phosphatase non-receptor type 1	Insulin receptor
	Sulfotransferase family cytosolic 2B member 1	Caspase-3	Coagulation factor VII	Stromelysin-1	Cytochrome P450 2C9	3-phosphoinositide dependent protein kinase 1	Uridine-cytidine kinase 2	Glutathione S-transferase P	Cell division protein kinase 7
	Ornithine carbamoyltransferase, mitochondrial	Phospholipase A2 **	Ras-related protein Rap-2a	Dipeptidase 1	Proto-oncogene Tyr-protein kinase LCK	Arginase-1	Eukaryotic tran-slation initiation factor 4E	Cathepsin K	Death-associated protein kinase 1
	Cystathionine beta-synthase	Riboflavin kinase	ADAM 33	Thymidylate kinase	Farnesyl pyrophosphate synthetase	Heat shock protein HSP 90-alpha	MMP9	Inositol monophosphatase	Neprilysin
	Ser/Thr protein kinase 6	ADP-ribosylation factor 4	Cell division control protein 42	6-phosphofructo 2-kinase/ fructose-2,6-biphosphatase	Pyruvate kinase isozymes R/L	Cyclin-A2	Interstitial collagenase	Beta-secretase 1	ADP-ribosylation factor-like protein 5B
	Eukaryotic translation initiation factor 4E	Leukocyte elastase	RAC-α Ser/Thr-protein kinase	3 histone mRNA exonuclease 1	L-serine dehydratase	5(3)-deoxyribo-nucleotidase,	Glu-6-P-isomerase	Ornithine aminotransferase, mitochondrial	UDP-N-acetylhexosamine pyrophosphorylase
	Uridine-cytidine	Tryptase beta-2	Neutrophil						
GOSTAR database	Target/MOA unknown	COX-1**	COX-2 **	COX	Albumin **	Thromboxane A2 receptor	I-keppa-B kinase β	Vanilloid receptor 1	Thromboxane A synthase1
	PRSS1	3-α-HSD2	Phospholipase A2 **	5-Lipoxygenase	NFKB	Postaglandin E recptor 2	Acid sensing ion chanel3	RPS6KA1	Organic anion transporter 1
	Organic anion transporter 3	Tachykinin receptor 1	Tachykinin receptor 2	Thrombin	MAPK8	MAPK14	Histamine H1 receptor	β-Glucuronidase	CAMP phosphodiesterase
	CYP2C9	Gastric H+/K+ ATPase	20-Alpha-HSD	Proto-oncogene c-ros-1	CDK5-P35	CYP 3A4	IL-1 receptor associated kinase 1	Activin receptor-like kinase4	Adenosine deaminase
	Alcohol dehydrogenase	3-α-HSD3	Allene oxide synthase	Anaplastic lymphoma synthase	Arbutin synthase	Arginine kinase	ABC transporter G2	Aurora kinase A	AXL kinase
	B-lymphoid Tyr kinase	BMX Tyr kinase	BRSK1	BRSK2	Brady kinin receptor B2	Bruton's Tyr kinase	Butyrylcholine esterase	ABL1	CAMK2
	CAMK1	Carbonic Anhydrase II	CAMK1-delta						

Examples: **Query** Cistanche tubulosa


- **Cistanche deserticola Y.C. Ma** is widely distributed in the northwest part of China.

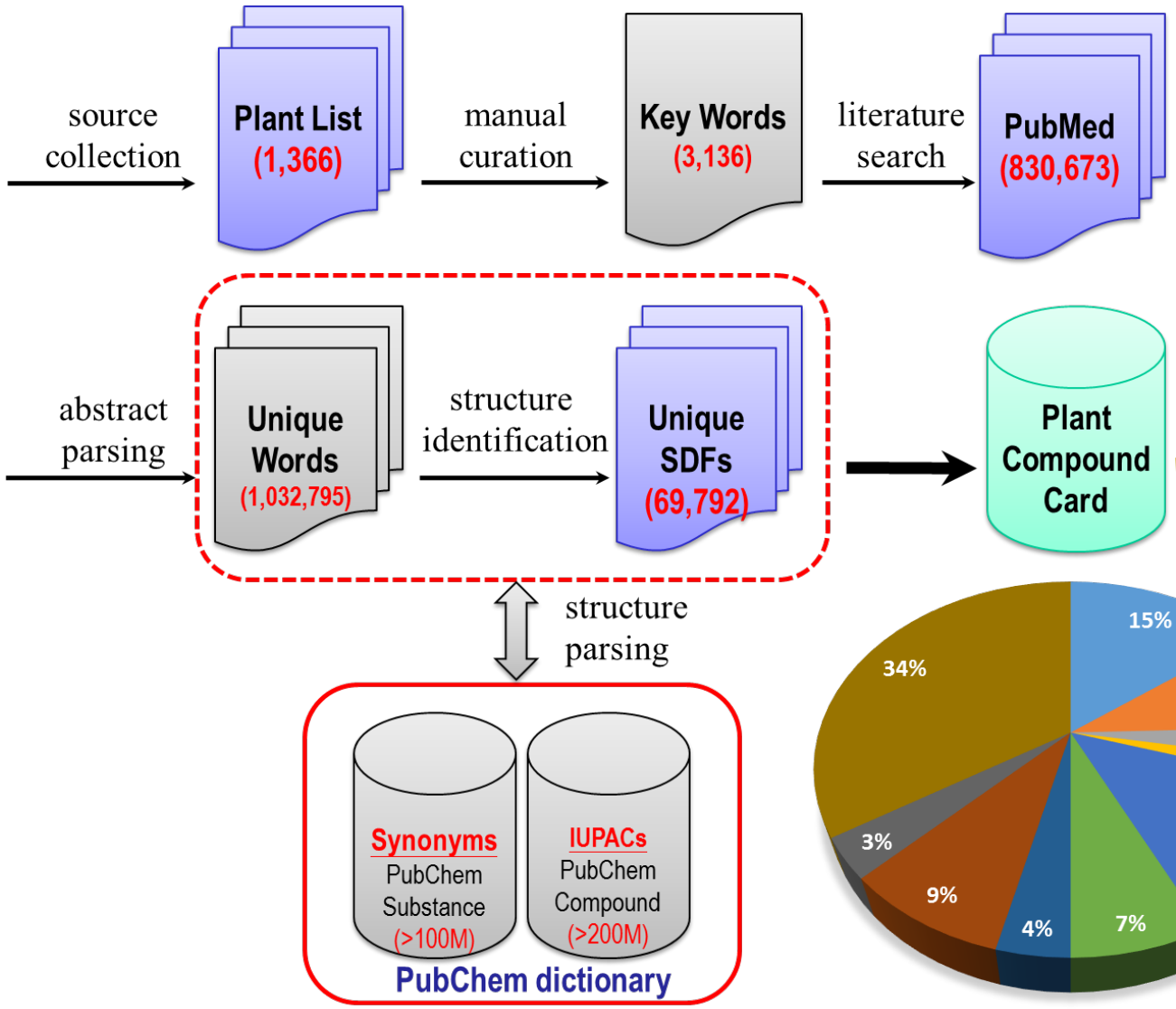
The whole dried parasitic plant (excluding the flower) is known as Cistanchis Herba.

- Cistanche deserticola has been traditionally used in the Chinese medicine

as a tonic, vasodilator and neuroprotective agent.
The plant extract was found to improves the age-related behavioral decline,
which makes it an attractive candidate for treatment of various
neurodegenerative disorders and age-related eye diseases

- Pharmacology studies have been demonstrated that Cistanche deserticola possesses

broad medicinal functions, especially for use in hormone regulation, aperient, immunomodulatory,
neuroprotective, antioxidative, anti-apoptotic, anti-nociceptive, anti-inflammatory,
anti-fatigue activities and the promotion of bone formation.



- **Work parallel track for natural therapeutics**

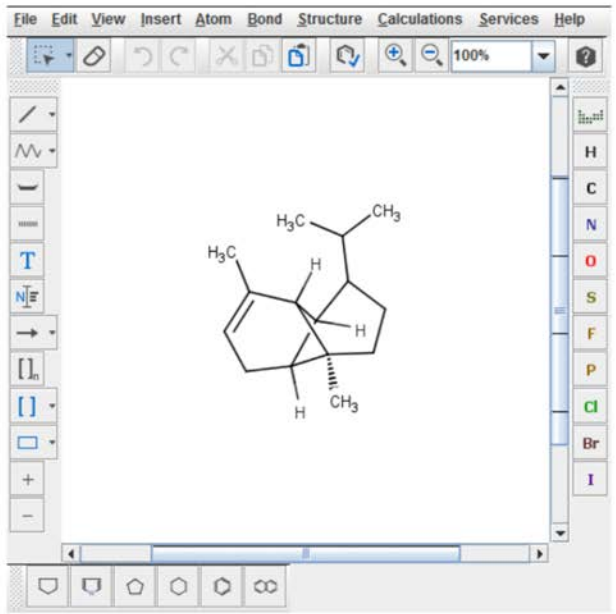
Plant: 1,366

Herb Recipe: 201

Compound: 69,792

Literature: 830,673

CPD Database Search



102181 Get Structure from CPD_REG_ID

OC1=CC=C(CC=C)C=C1

Show Advanced Options

Main Options

Search Type: Similarity

Similarity Threshold: 0.85

Max. Hits: 200

Max. Time: 3 Min.

Return Non-Hits (Inverse Result Set)

Condition Set

Molecular Weight: Low High

Search

Upload CPD structure file for similarity search

Browse... No file selected.

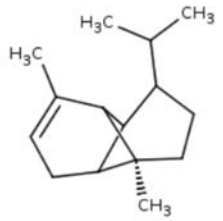
Similarity Threshold:

0.85

Go

CPD Search Result

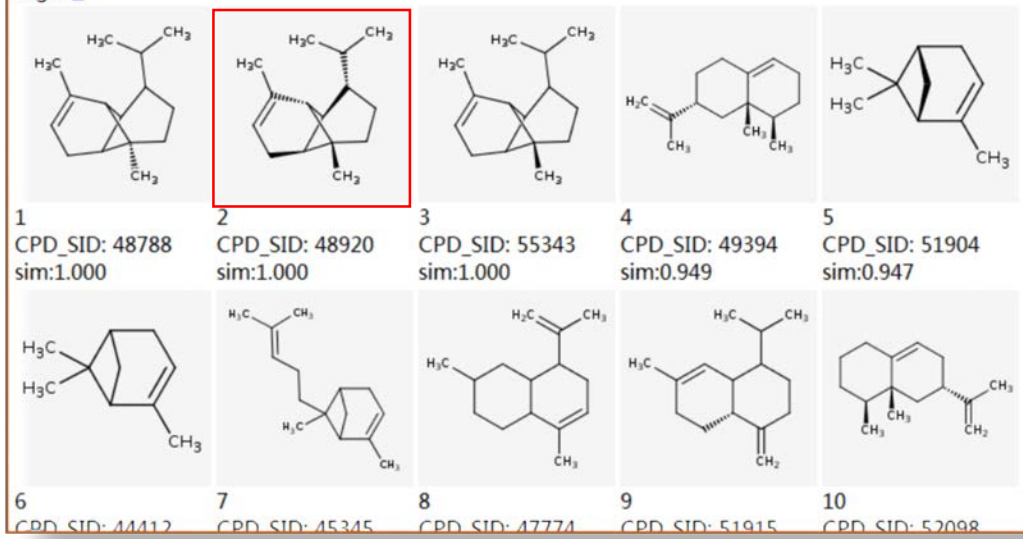
Search for



Total Hits Count: 21

Max Hits Display: 200

Page: 1



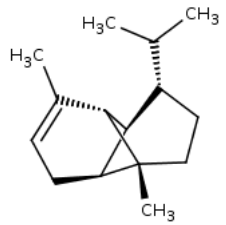
1	2	3	4	5
CPD_SID: 48788	CPD_SID: 48920	CPD_SID: 55343	CPD_SID: 49394	CPD_SID: 51904
sim:1.000	sim:1.000	sim:1.000	sim:0.949	sim:0.947

6	7	8	9	10
CPD_SID: 44412	CPD_SID: 45345	CPD_SID: 47774	CPD_SID: 51915	CPD_SID: 52098

Compound Detail - Mozilla Firefox

192.168.77.1:8080/drugbank/GRC_TCM/CpdDetail.jsp?cpd_sid=48920

Open a new tab (Ctrl+T)



CPD Info

CPD_SID: 48920

SMILES: [H][C@]12CC=C(C)[C@]3([H])[C@@]1([H])[C@@]([H])

Chemical Properties

Molecular Weight: null

Molecular Formula: null

CPD Instance Info

TcmInfo:346

TCM ID

25(TcmInfo)

1026(TcmInfo)

Show TCM Network

Close

TCM Detail - Mozilla Firefox

192.168.77.1:8080/drugbank/GRC_TCM/TcmDetail.jsp?tcm_id=1026&source_db=TcmInfo

TCM Info

TCM ID: 1026

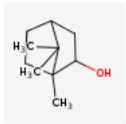
TCM Source DB: TcmInfo

Chinese Name: 青蒿

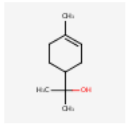
English Name: Herba Artemisiae Annuae

Common Name: Sweet Wormwood Herb

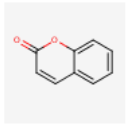
CPD Info



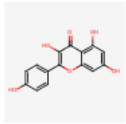
41957



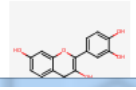
42063



42299



42701



TCM Detail - Mozilla Firefox

192.168.77.1:8080/drugbank/GRC_TCM/TcmDetail.jsp?tcm_id=25&source_db=TcmInfo

TCM Info

TCM ID: 25

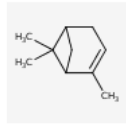
TCM Source DB: TcmInfo

Chinese Name: ??草

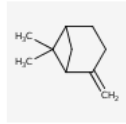
English Name: Herba Patriniae

Common Name: All - grass of Dahurian Patrinia

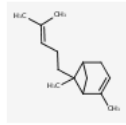
CPD Info



44412

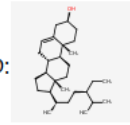


44469

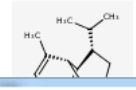


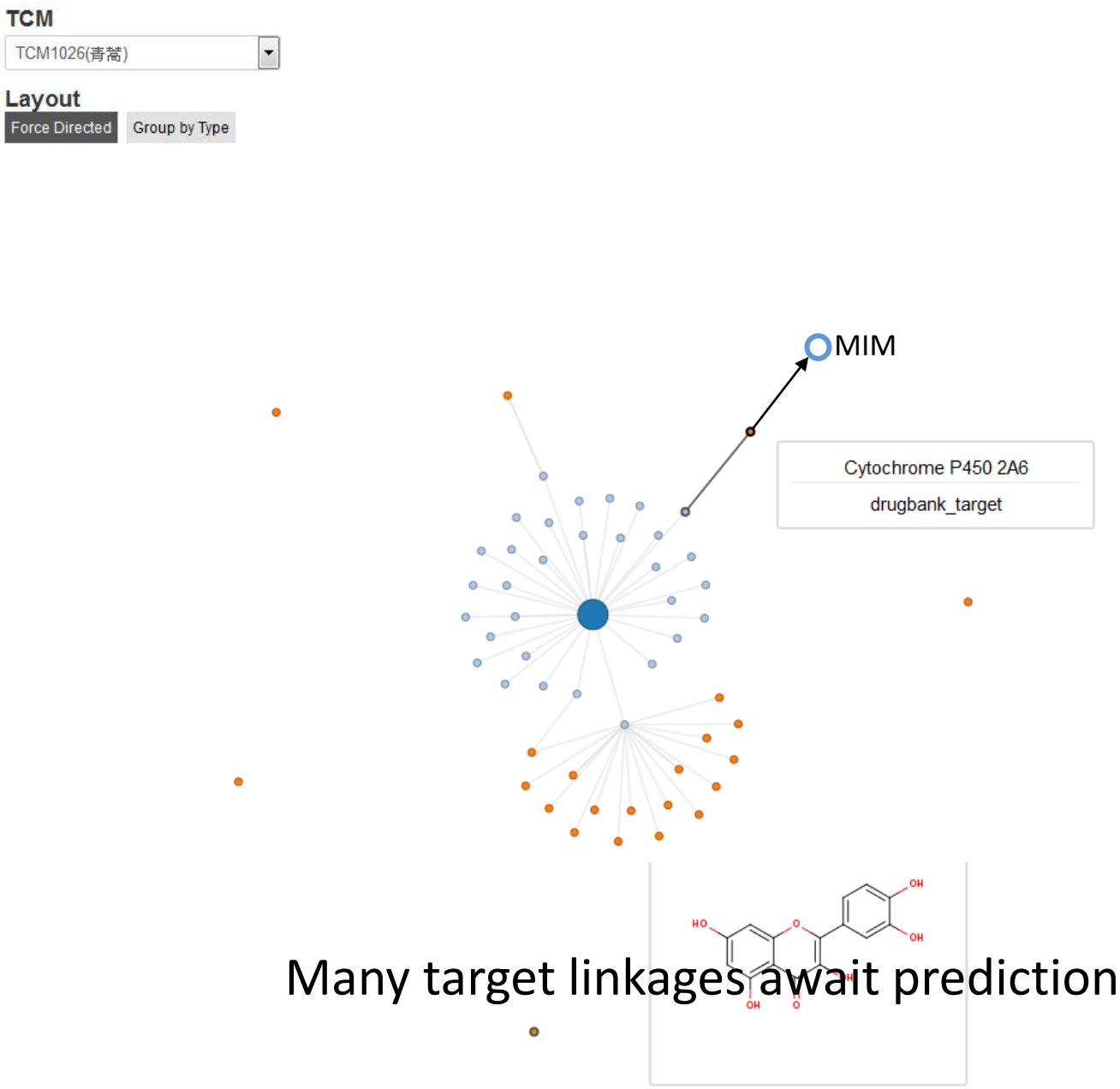
45345

CPD_SID:



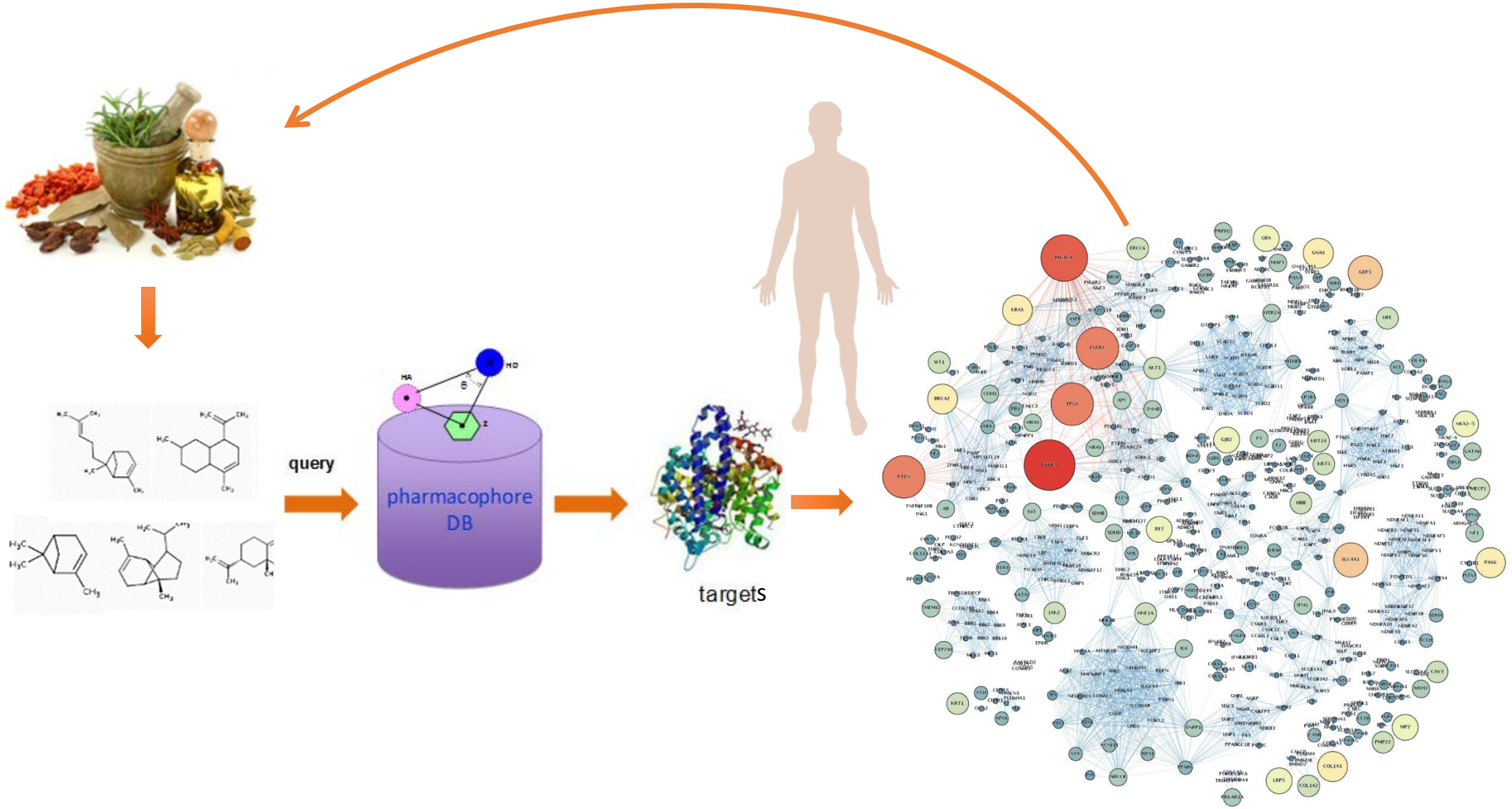
48854





Many target linkages await prediction

Summary



Credits

- **Supports**

  Chi-Huey Wong 翁啟惠
  Chung-Hsuan Chen 陳仲瑄

- **GRC uHTS operation Team**

  Shi-Shan Mao 毛溪山
  Ting-Jen Cheng 鄭婷仁
  Po-Hsun Lin 林柏勳
  Jiraporn Nawarak 林力雯
  Chun-Hung Su 蘇俊宏
  Hua-Ting Hsu 許華亭
  Kuo-Yao Tseng 曾國堯
  Hui-Min Lin 林慧敏

- **Collaborators in application development**

  Jean Lu 呂 仁
  Han-Chung Wu 吳漢忠
  Shiaw-Min Hwang 黃效民

- **Collaborators in library development**

  Jim-Min Fang 方俊民
  Shang-Cheng Huang 洪上程
  Hsu-Shan Huang 黃旭山
  Wen-Shan Li 李文山
  Chung-Ming Sun 孫仲銘
  Wei-Chieh Cheng 鄭偉杰
  Rong-Jie Chein 陳榮傑
  Meng-Yang Chang 張夢揚
  Hui-Ming Yu 余惠敏
  Chien-Tai Ren 任建台

Funding Supports:

中央研究院 基因體研究中心
Genomics Research Center, Academia Sinica

中央研究院--登峰科技計畫

生技醫藥國家型科技計畫
National Research Program for Biopharmaceuticals

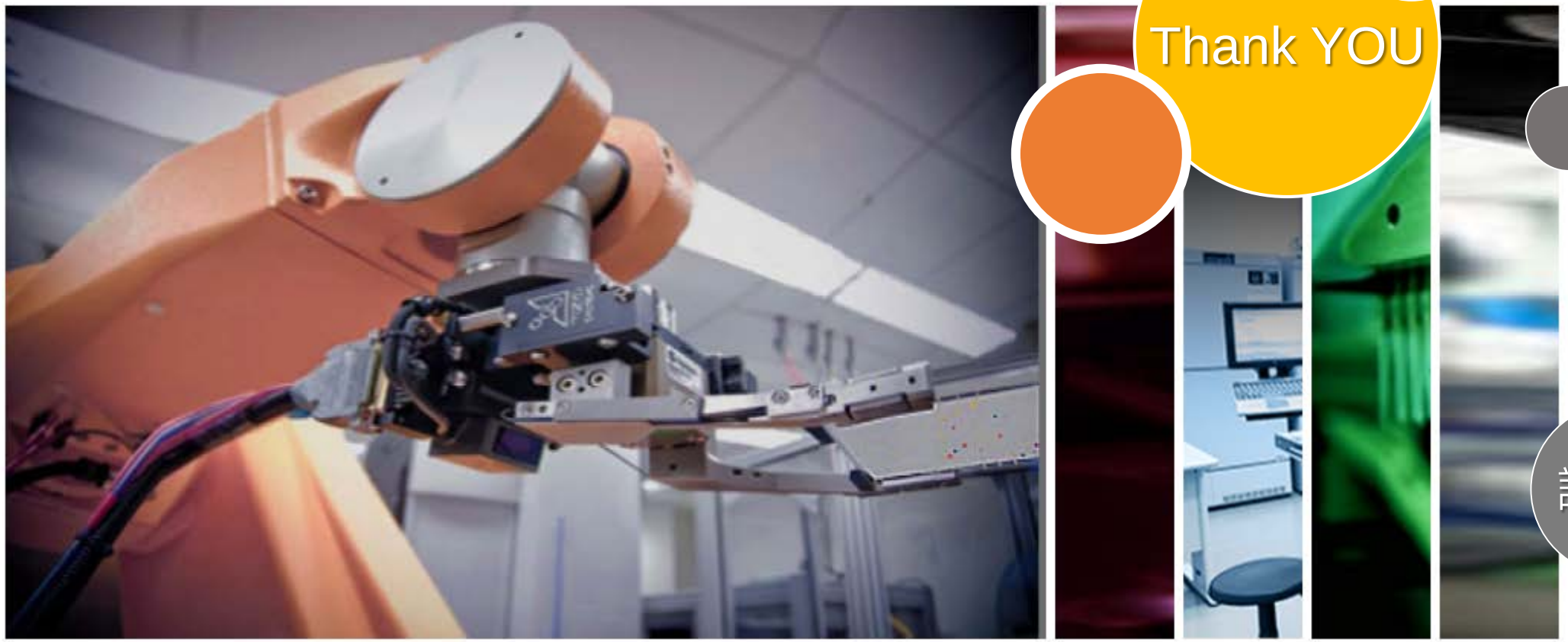
中央研究院 基因體研究中心
Genomics Research Center, Academia Sinica

Ultra High-Throughput Screening System

生技醫藥國家型科技計畫
National Research Program for Biopharmaceuticals

中文版 | English

HOME ABOUT US SERVICES TECHNOLOGY NEWS & PUBLICATIONS arch



Thank YOU

感 恩

謝 謝