Drug Discovery Grid (DDGrid)

Zhang Wenju, Shen Jianhua

Shanghai Institute of Materia Medica, CAS
Shanghai Jiao Tong University
Jiangnan Institute of Computing
The University of Hong Kong

Agenda

1. DDGrid Overview
2. DDGrid Architecture
3. Result Analysis
4. Future Work

Background

Large-scale High-throughput Virtual Screening

- **in Silico**

  The computational analysis of chemical databases to identify compounds appropriate for a given biological receptor

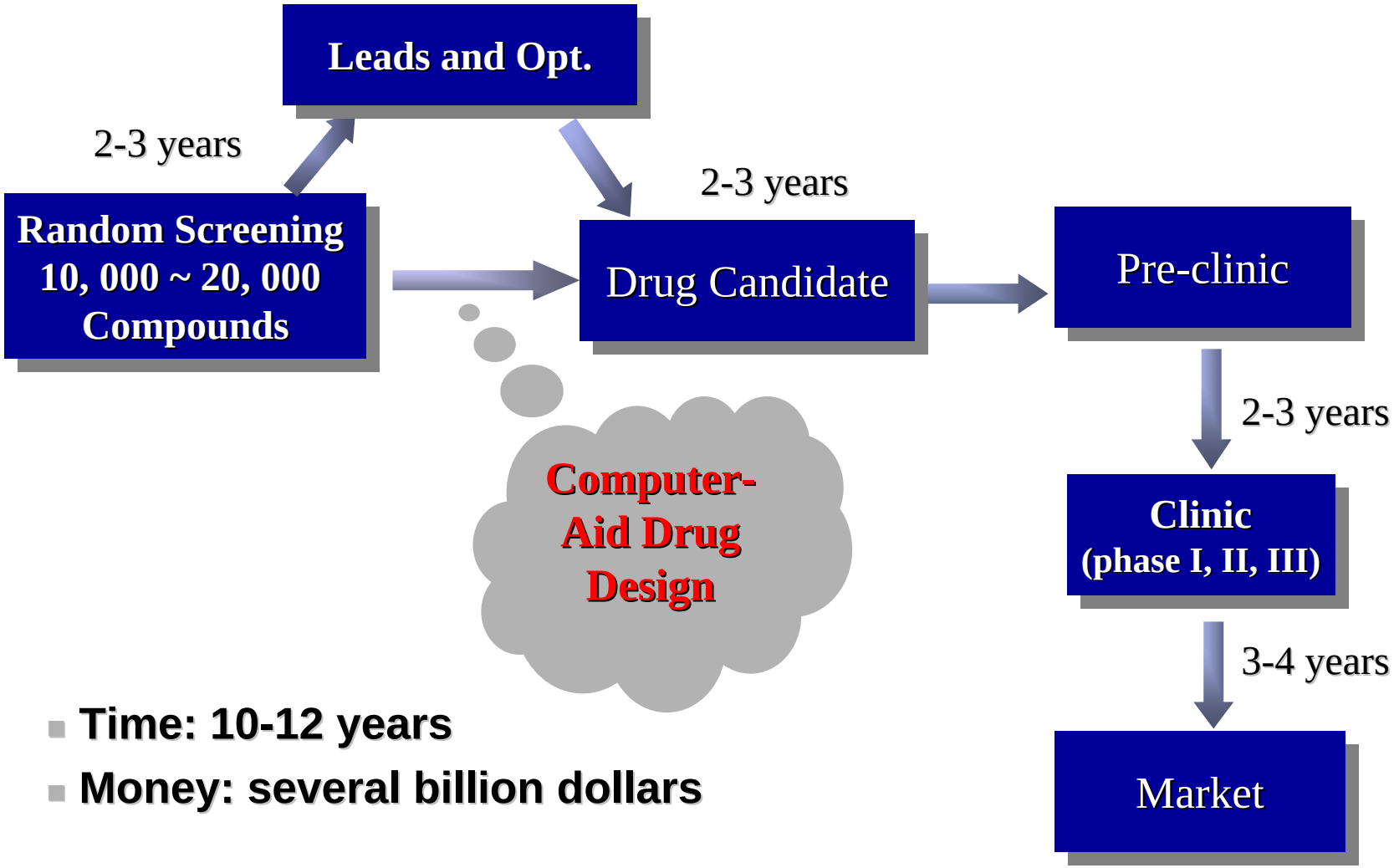
- **in Vitro**

  the progressive optimization of these leads to yield a compound with improved potency and physicochemical properties *in vitro*

- **in Vivo**

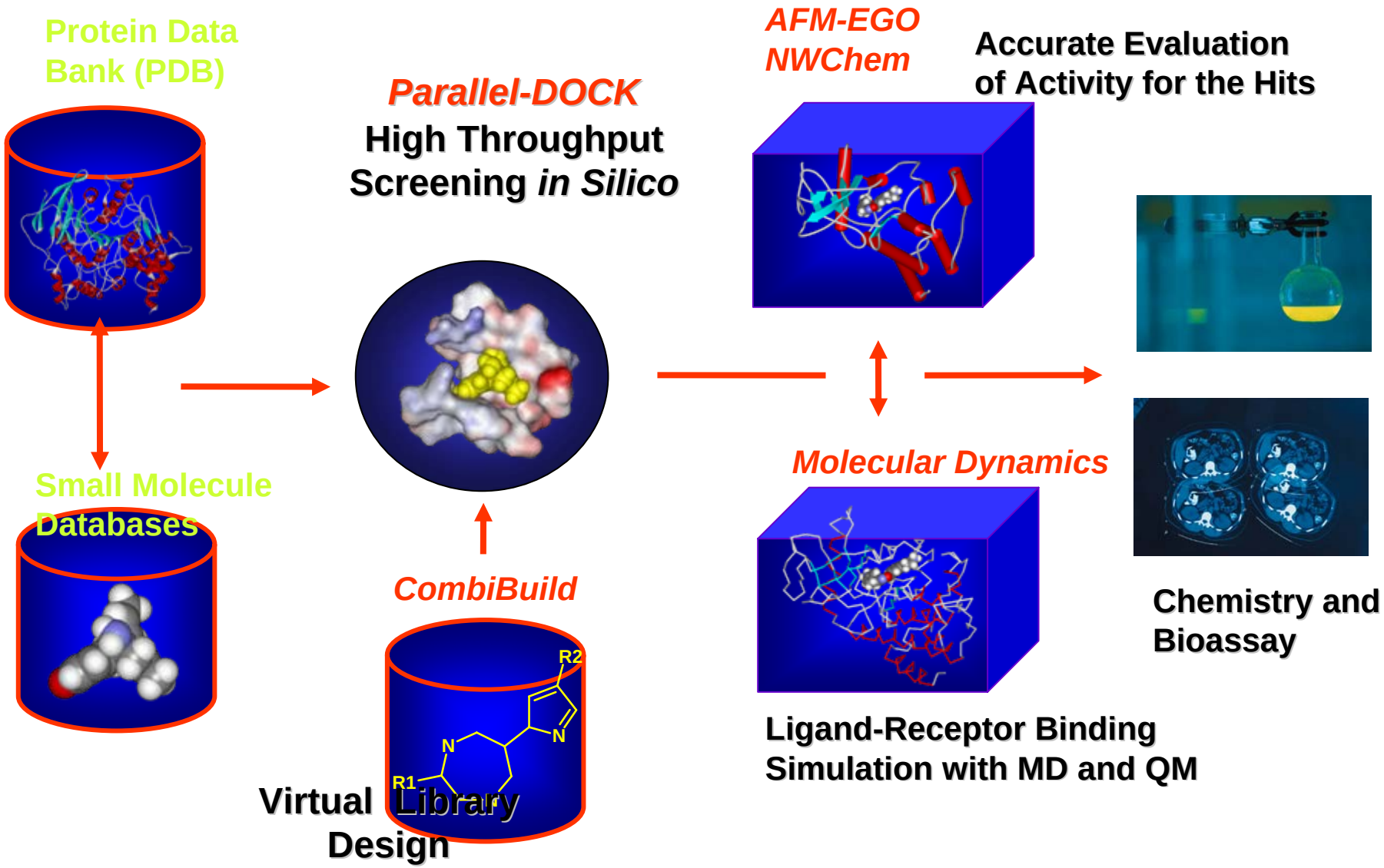
  eventually, improved efficacy, pharmacokinetics, and toxicological profiles *in vivo*.

General Process of Drug Discovery and Design

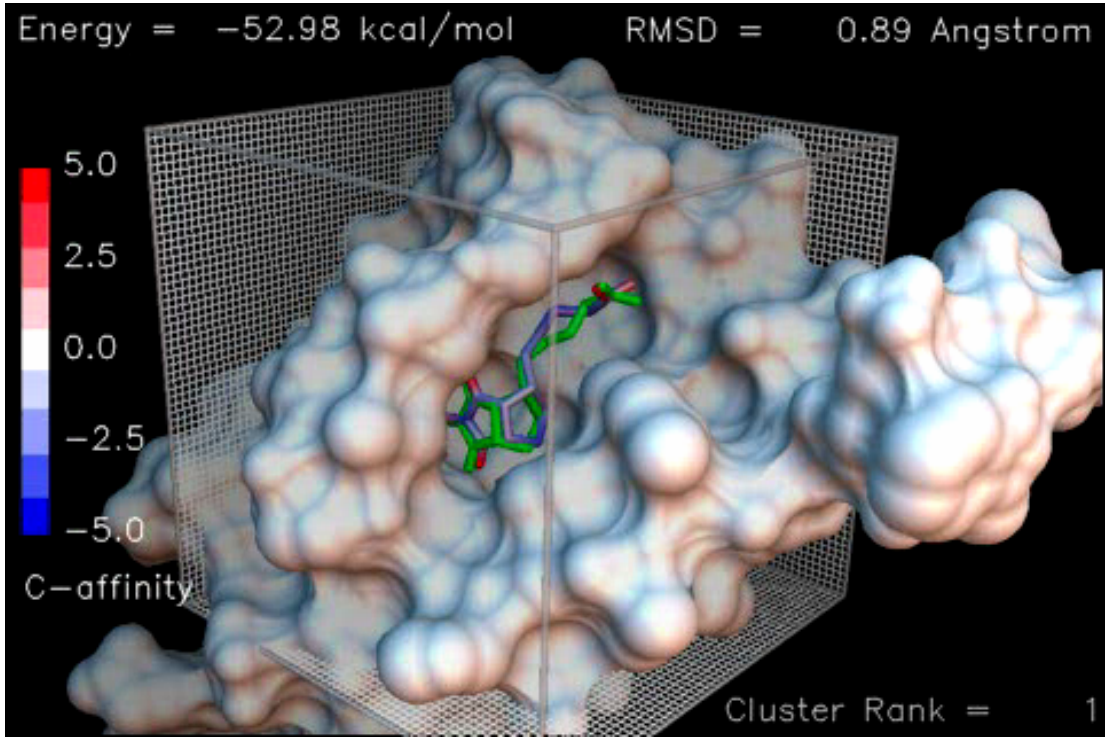


- **Time: 10-12 years**
- **Money: several billion dollars**

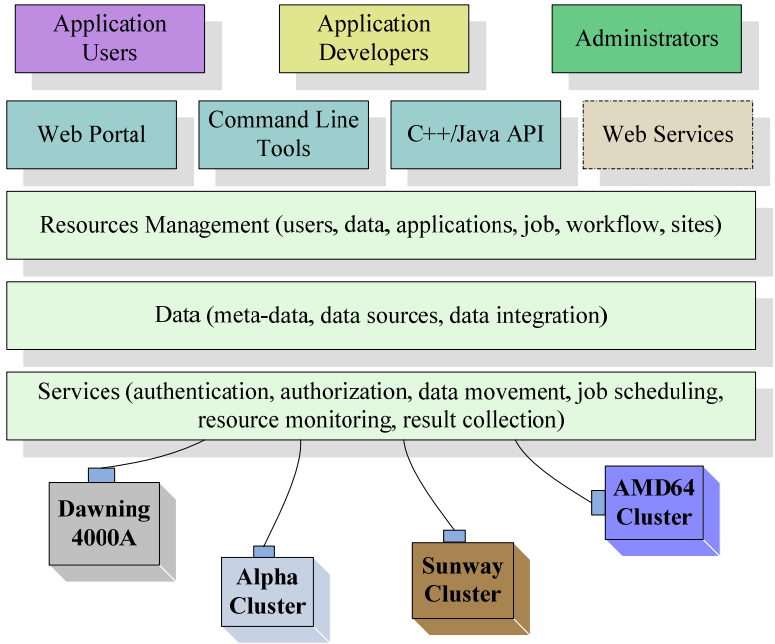
Drug discovery in SIMM

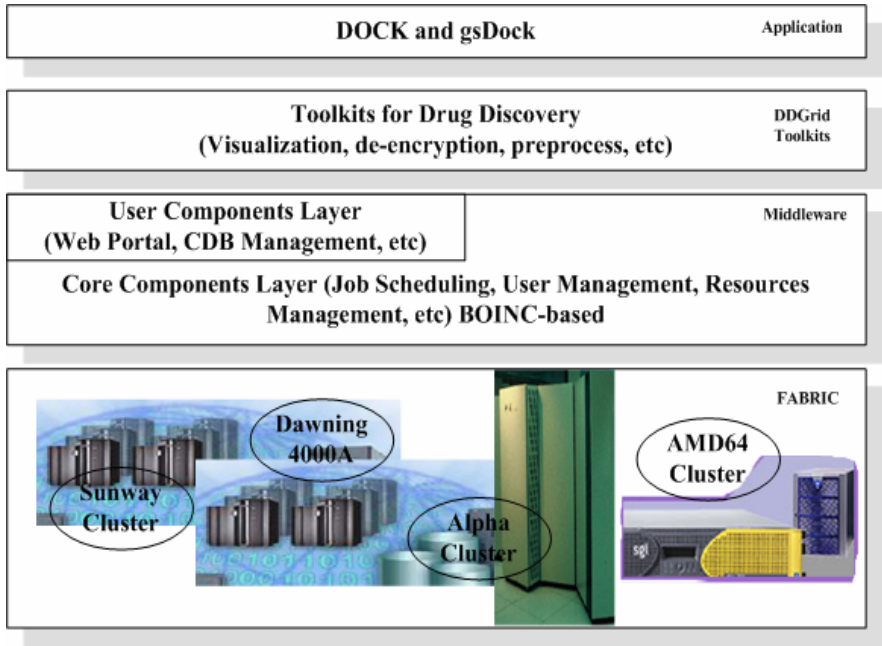


Virtual screening

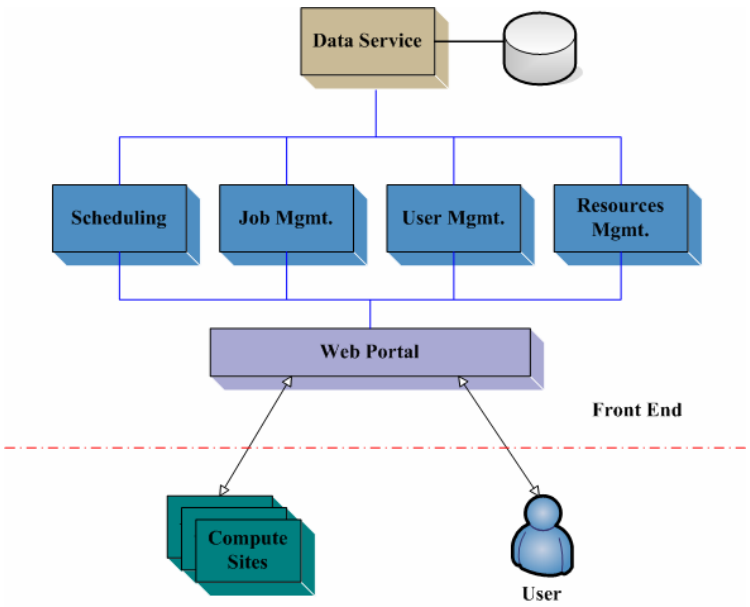


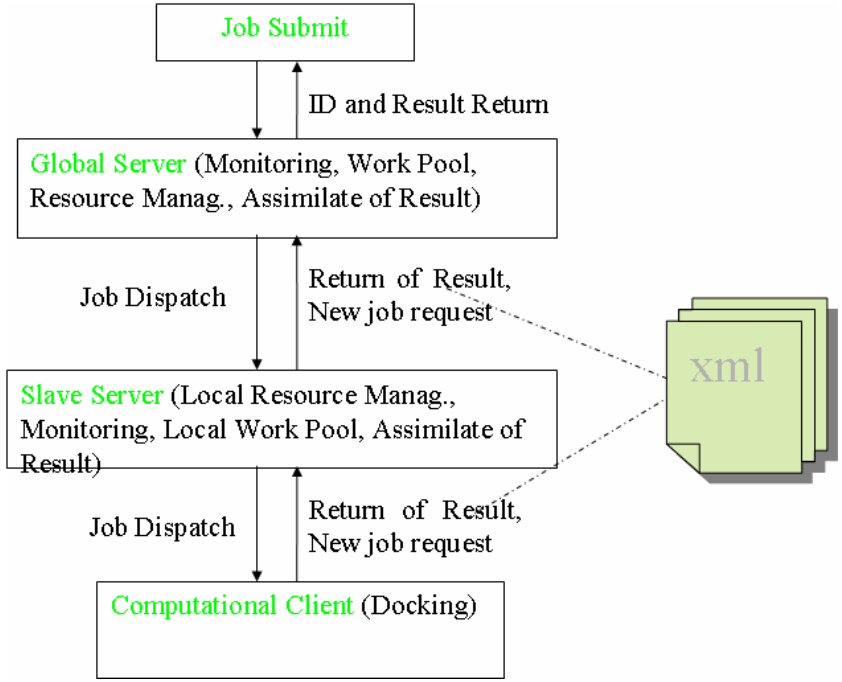
DDGrid Architecture





DDGrid Architecture





Comp. resources aggregated

Computing Resource

No.	Location	Unit	Slave Server	Server Type	Total CPU Number	Avail. CPU Number	Avail. :Busy :Idle	
A	Shanghai	Shanghai Supercomputing Center	SSC	Dawning 4000A	128	0		0% Avail.
B	Shanghai	Shanghai Institute of Materia Medica	SIMM_ALPHA	Sunway 32A				N/A
C	Beijing	Beijing Molecule Company	Molecule	Sunway 256P				N/A
D	Hong Kong	Hong Kong Univ.	HKU	Gideon Cluster	16	16		100% Avail.
E	Shanghai	Shanghai Institute of Materia Medica	SIMM_PC	Sunway 32A	16	0		0% Avail.
F	Dalian	Dalian University of Technology	DLUT	Lenovo Shenteng1800				N/A
G	Shanghai	Shanghai Jiaotong Univ.	sjgcc					N/A
H	Singapore	Singapore Polytechnic Univ.	SP	Sun	32	0		0% Avail.
I	London	London e-Science Centre	LeSC					N/A

Data resources on DDGrid

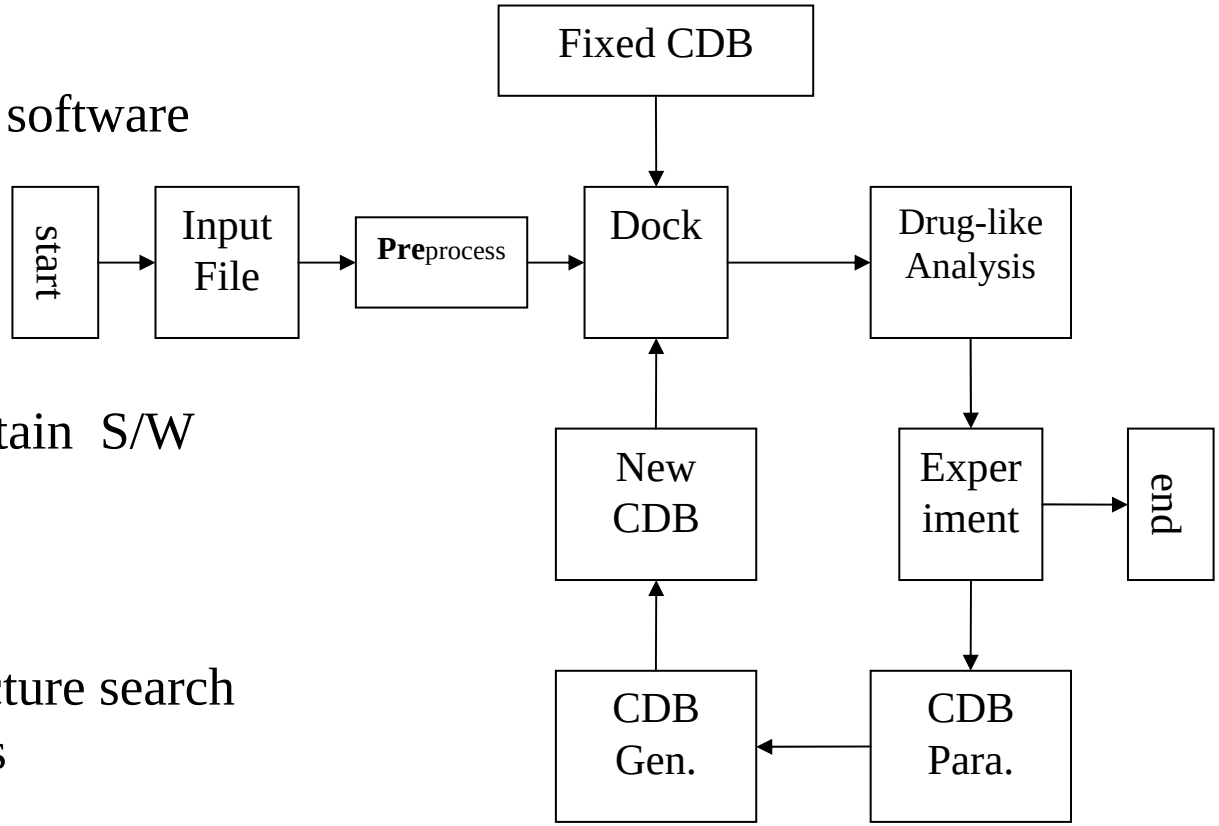
Commerical CDB

Data Resource			
Database Name	Status	Numbers	Description
Specs	Available	230,000+	Chemical available compounds' structure database
MDL Comprehensive Medicinal Chemistry 3D	Available	8,400+	Pharmaceutical compounds (1900- present)
MDL Available Chemicals Directory 3D	Available	200,000+	Commercially available chemical's 3D structure database
National Cancer Institute Databases NCI-3D	Available	213,000+	Structures with corresponding 3D models
China Natural Products Database	Available	12,000+	Chemical, structural and bibliographic data are provided for natural products first isolated in China.
Traditiona Chinese Medicinal Database TCMD	Available		9127 chemical structures, 3922 medicinal herbs or plants
ZINC - ChemBridge	Available	562,624	Chemical structures, web site: www.chembridge.com
ZINC - MayBridge	Available	53,042	Chemical structures, web site: www.maybridge.com

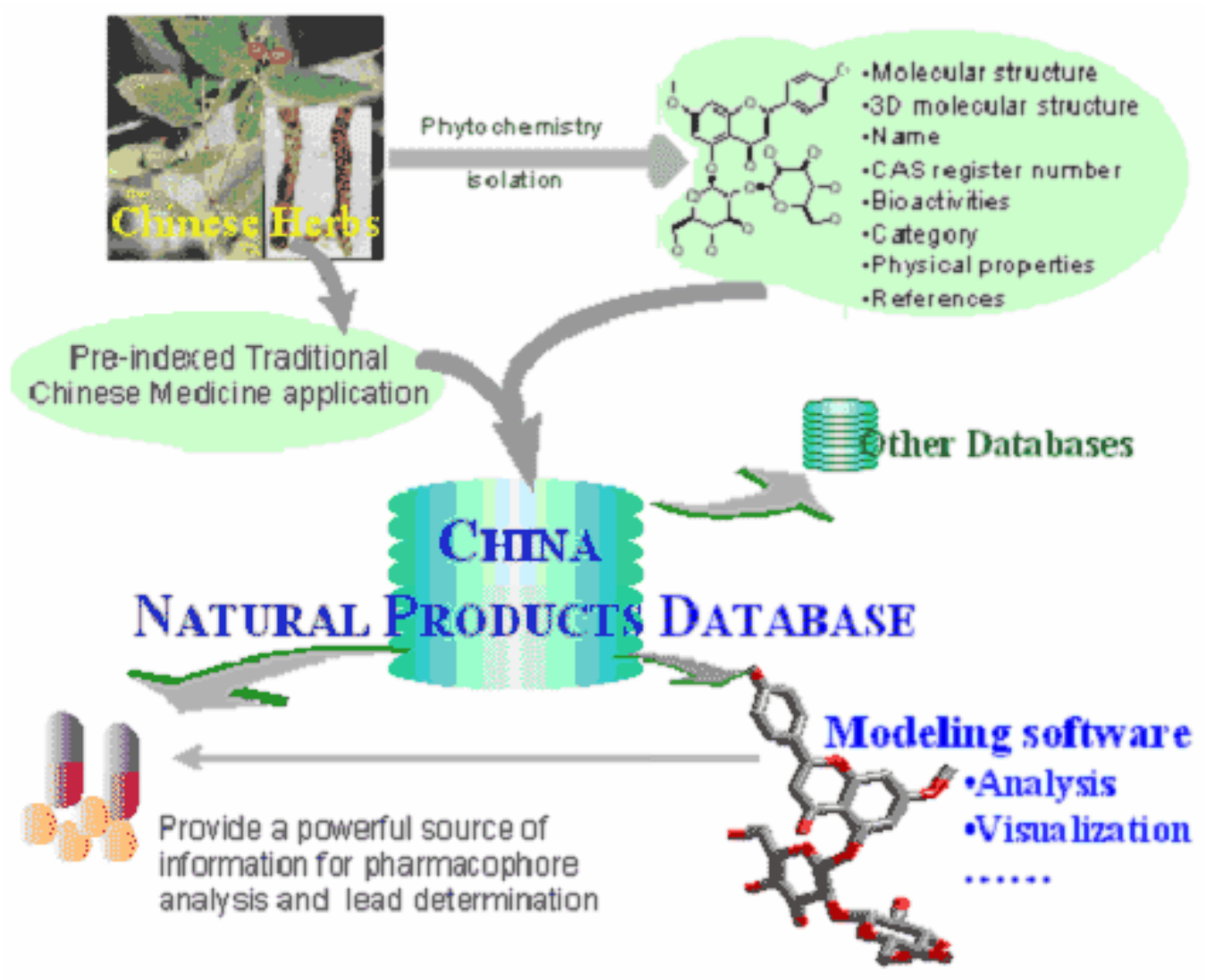
Application softwares

DDGrid Apps.

1. Docking pre-process software
   Combimark
2. Docking software
   1) Dock UCSF
   2) gsDock SIMM
3. CDB build and maintain S/W
   Combilib
4. AutoDock
5. AutoGrid
6. Visualisation & structure search
7. Security-related tools



CDB example：CNPD-China Natural Products Database

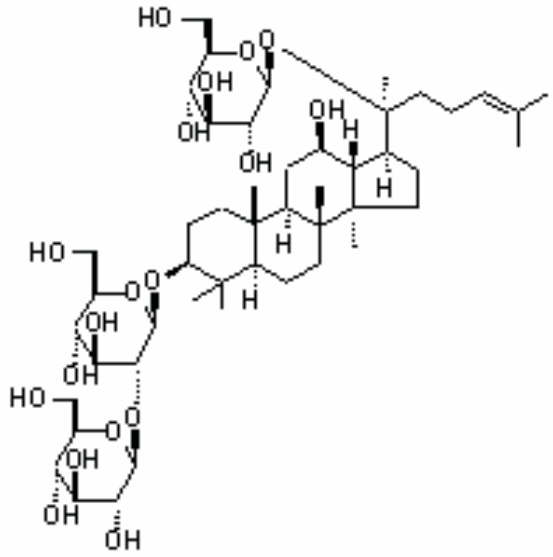


CDB example：TCMD-Traditional Chinese Medicine Database

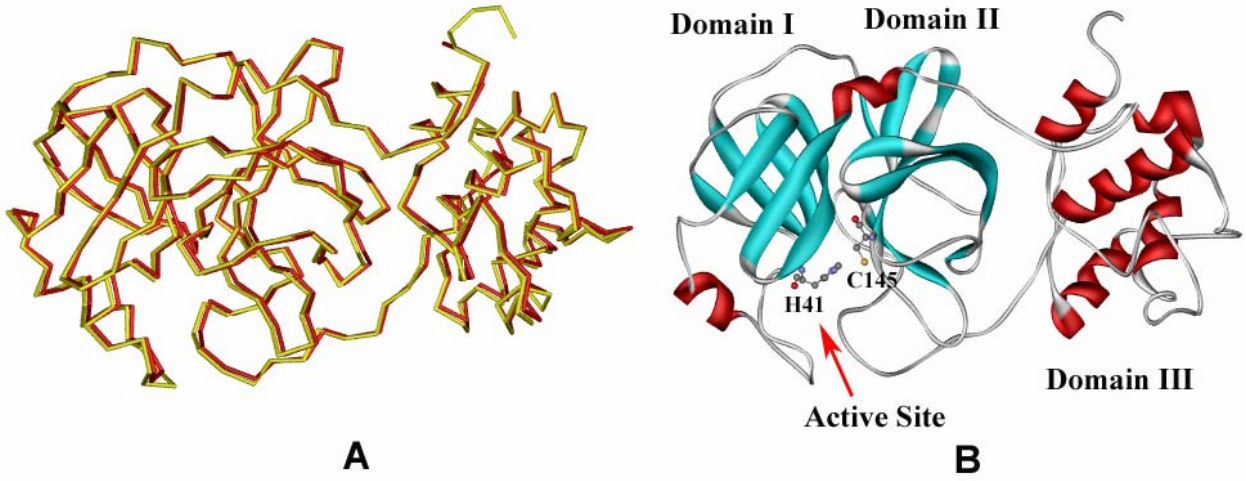
3368 Ginsenoside Rd 人参皂甙Rd

Gypenoside VIII 52705-93-8 C48H82O18 947.18 物化性质： 白色粉末（乙醇-正丁醇 1:1），mp 206-209° C，[α]D22 = +19.38° （c = 1.03，甲醇）。 药理活性：抗心律失常（BaCl2引起的大白鼠心律不齐）；抗病毒；抑制HSV-1的复制；大白鼠肝脏匀浆的抗氧化剂（H2O2所致，IC50 = 12.0±0.8μg/ml，by FeSO4，IC50 = 457.5±15.4μg/ml）；11-β-羟甾脱氢酶抑制剂；cAMP磷酸二酯酶抑制剂（in vitro，IC50 = 84μM）；促进软毛霉素和长春花碱的细胞毒作用；促进皮质酮血浆分泌（ED50 = 112 mumol/kg）；调节肾功能和抑制肾小球再生；血管扩张剂。原植物： 绞股蓝，秦岭珠子参，人参，西洋参。参考文献：

4,87,451,900



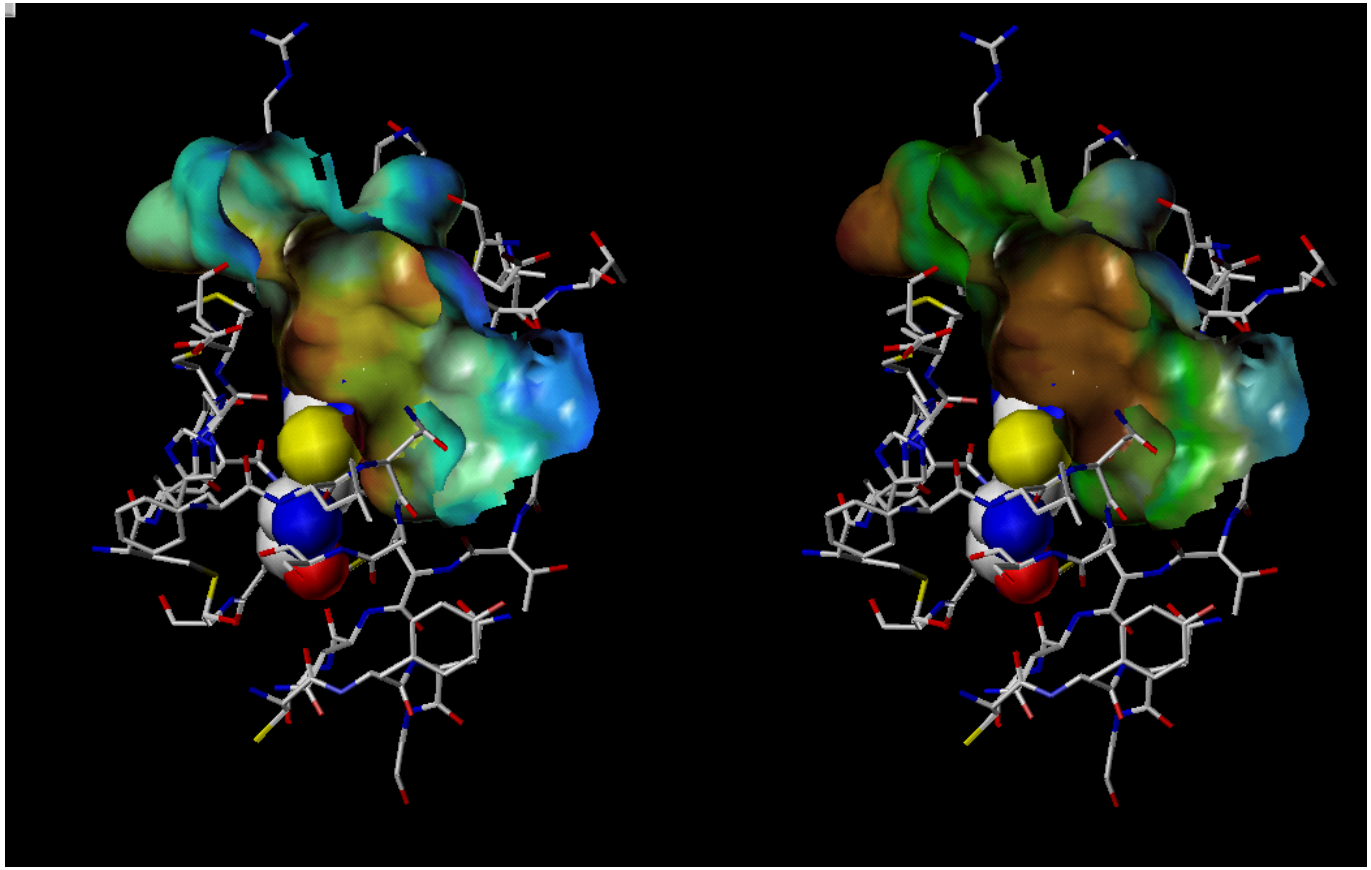
3D Model of SARS-CoV 3CL Proteinase



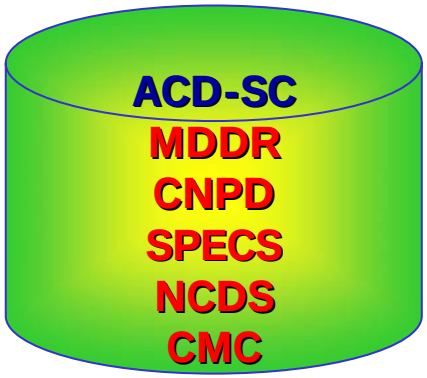
identity: 43%, similirity: 60%, gap: 1%.

(A) Structure superposition of the modeled structure of SARS 3CL proteinase (yellow) with the X-ray crystal structure of the M^{pro} of TGEV (transmissible gastroenteritis virus)(red). Only backbones are shown in this picture. (B) The solid ribbon representation of the structure model of SARS 3CL proteinase. The substrate-binding site is located at the deep cleft between domains I and II, and the active site is situated at the center of the cleft, the catalytic residues H41 and C145 are represented by ball-and-stick.

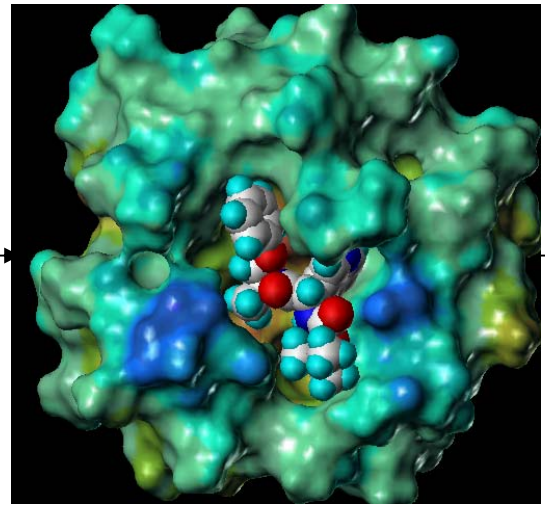
Active site of the SARS 3cl proteinase



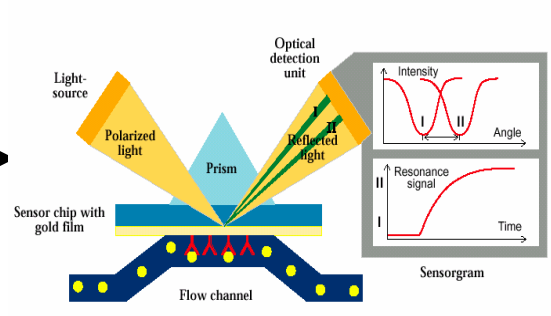
Virtual screening for SARS 3cl proteinase



Databases



Docking



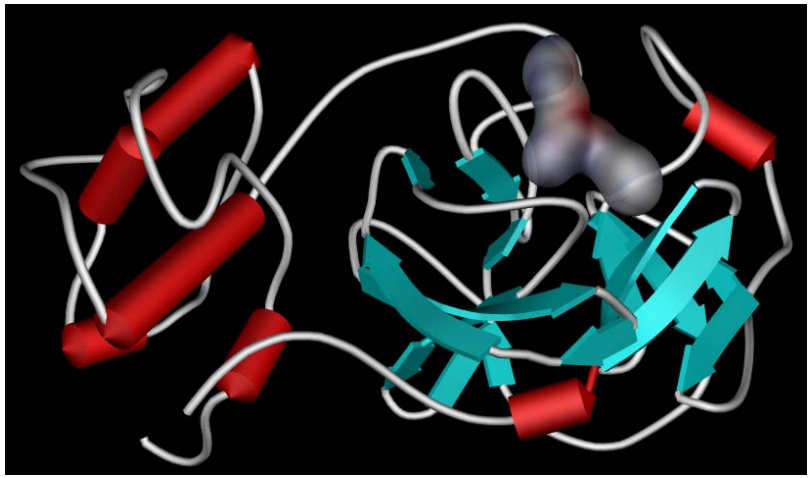
Biacore 3000

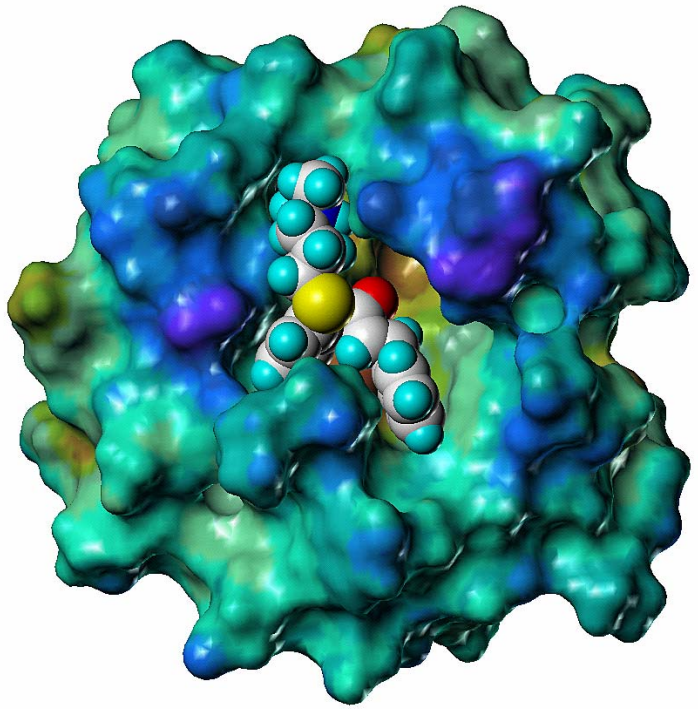
Binding Assay

CMC Virtual Screening Result

	MDLNUMBER	Dock (kcal/mol)	Autodock (kcal/mol)	Cscore
1	DDDCS001	-30.99	-9.68	**5**
2	DDDC001	-37.06	-8.79	**4**
3	DDDCS002	-35.09	-8.60	**5**
4	DDDCS003	-35.52	-8.76	**5**
5	DDDCS004	-32.52	-9.55	**5**
6	DDDCS005	-34.60	-7.57	**5**
7	DDDCS006	-31.98	-9.27	**5**
8	DDDCS007	-33.72	-8.69	**4**
9	DDDCS008	-37.45	-8.07	**4**
10	DDDCS009	-36.06	-7.08	**4**

Result visualization





Selected users of DDGrid

Future work

- **New application and upgrade**
  - Screening considering Protein flexibility,
  - Target Fishing Dock (TarFishDock)
- **Architecture refactoring**
  - adopt to recent common standards,
  - publishing services to remote consumer,
  - swappable screening methods and software,
  - workflow-oriented resource composition
- **Chemical database increment**
  - more self-made CDB
  - more commercial CDB

Acknowledgements

This work has been supported by National High-Tech Research and Development Project of China, under contract No. 2004AA104270

Many thanks to generously resource providers:

SIMM
HKU
SJTU
Molecule Ltd.
SSC
DLUT

Involved Persons:

Shen Jianhua
Ma Fanyuan
Zhang Jun
Zhang Wenju
Chang Yan
Chen Shudong
Du Xuefeng
Li Zhuhua
Liu Fei
Wan Ju
Jiang Maojun
…

Further Cooperation Contact info

Dr. Kunqian YU, SIMM, CAS
kqyu@mail.shcnc.ac.cn
http://www.dddc.ac.cn
Mr. Wenju ZHANG, SJTU
zwj03@sjtu.edu.cn

http://www.ddgrid.ac.cn

Drug Discovery Grid

Q & A

Drug Discovery Grid

Thank you very much!