Development and application of a web-based integrated platform D3CARP for target prediction and virtual screening

Weiliang Zhu

wlzhu@simm.ac.cn

Drug Discovery and Design Center (DDDC)

Shanghai Institute of Materia Medica (SIMM), Chinese Academy of Sciences (CAS)

Outline

1. Introduction

2. New Methods for Protein Conformation Sampling

3. D3CARP for Target Prediction and Virtual Screening

4. Applications

Introduction

Approaches for Target Prediction and Virtual Screening

Structure-Based:

Predicts how an active compound binds to a target protein. Software such as AutoDock, Glide, or DOCK simulates the binding interactions and ranks the predicted targets or screened compounds based on their docking scores.

Similarity-based:

A computational approach used to identify potential drug candidates by comparing their chemical similarity to known active compounds. This method relies on the principle that molecules with similar structures or properties are likely to exhibit similar biological activities.

Machine/Deep Learning Based:

Algorithms such as Support Vector Machines (SVMs), Neural Networks and Deep Learning can predict drug targets based on various features, including sequence data and biological interactions.

Questions

- **Should different conformations be considered for docking?**

  Yes, but not included in most approaches (DFG-in vs DFG-out)

- **Should different binding sites be considered for docking?**

  Yes, for example, allosteric sites

- **Is a platform with different approaches useful?**

  Yes, for validation to increase reliability of the predictions

- **Is positive controls important for assessing results?**

  Yes, same score but different activity (1nM vs 1mM)

Outline

1. Introduction

2. New Methods for Protein Conformation Sampling

3. D3CARP for Target Prediction and Virtual Screening

4. Applications

Protein conformation sampling is a challenge

- Protein adopts multiple conformations
- Conformation change is closely related to its biological function

- It is difficult to study the transition pathway experimentally
- Various simulation methods have been developed

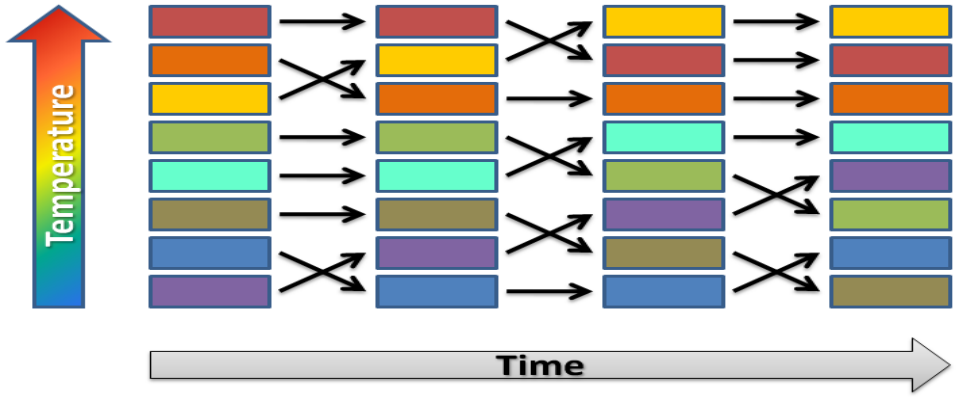
Targeted MD	The free energy calculation is biased force dependent and its connection to function is unclear
Accelerated MD	
REMD	The free energy calculation is time-consuming and computationally expensive
Metadynamics	
Coarse Grained Method	Can predict the transition pathway but can not provide the free energy information

Replica exchange molecular dynamics (REMD)

- **REMD is one of the well recognized MD methods for conformation sampling;**
- **It requires large computational resources**
  - Protein (214 aa) --- 80 replicas --- 21.5 k RMB (3 k US$)
  - Protein (856 aa) --- ~ 450 replicas --- 3 m RMB (400 K US$)

Therefore, highly efficient MD simulation methods are expected.

Conventional REMD



Replica ($N_{replica}$) required for REMD: $N_{replica} \propto \sqrt{n_{freedom}}$

Exchange criteria:

$$\omega(1 \leftrightarrow 2) = min\left(1, exp(\Delta\beta\Delta P)\right)$$

System potential energy

$$P = P_{pp} + P_{pw} + P_{ww}$$

1. Improved REMD method: vsREMD

- The solvent atoms are not considered during replica exchange, the new exchange criteria could be

$$\omega(1 \leftrightarrow 2) = min\left(1, exp\left(\Delta\beta\Delta\left(P_{pp} + P_{pw}\right)\right)\right)$$

$$P = \boxed{(P_{pp}+P_{pw})} + P_{ww}$$

Exchange reference ← **Velocity Scaling**

Difference:

$$H = P_{ww}$$

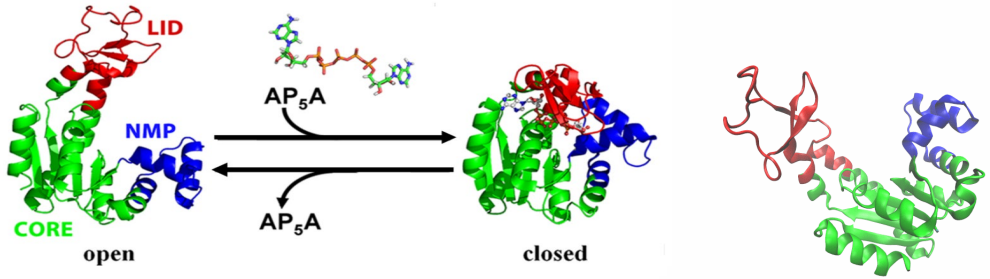
- Velocity scaling (vsREMD):

$$\hat{v}^{(1\to2)} = v^{(2)}\sqrt{\frac{\hat{E}^{(1)}_{kin}}{E^{(2)}_{kin}}} = v^{(2)}\sqrt{\frac{E^{(1)}_{kin} - \Delta H}{E^{(2)}_{kin}}}$$

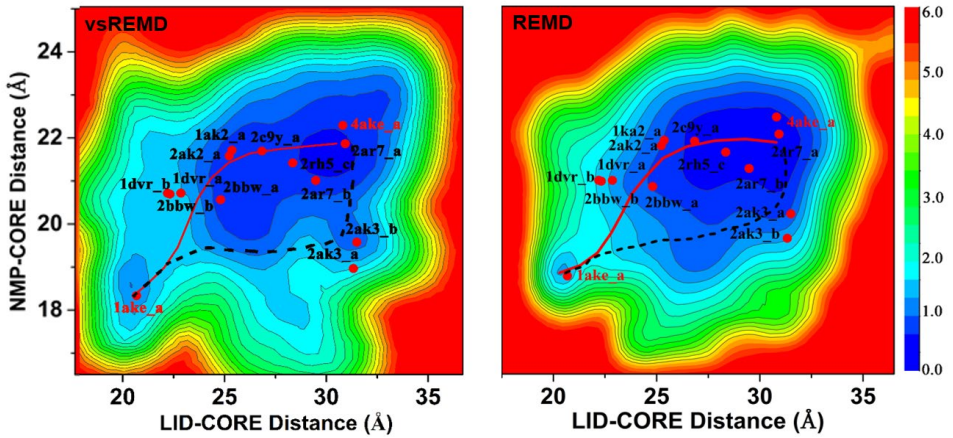
$$\hat{v}^{(2\to1)} = v^{(1)}\sqrt{\frac{\hat{E}^{(2)}_{kin}}{E^{(1)}_{kin}}} = v^{(1)}\sqrt{\frac{E^{(2)}_{kin} + \Delta H}{E^{(1)}_{kin}}}$$

Velocity Scaling REMD (vsREMD, developed based on GROMACS)

Testing vsREMD on Adenylate Kinase (AdK),

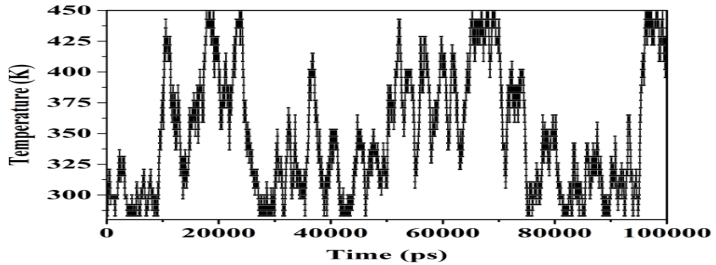


Two X-Ray ADK conformations: open (4AKE), closed (1AKE)



38% replica of REMD required for vsREMD

	副本	温度 (K)											
vsREMD	30	300.0	304.2	308.5	312.8	317.2	321.7	326.2	330.8	335.5	340.2	345.0	349.9
		354.8	359.8	364.9	370.0	375.2	380.5	385.8	391.3	396.8	402.4	408.0	413.8
		419.6	425.5	431.5	437.6	443.8	450.0						
REMD	80	300.0	301.5	303.1	304.6	306.2	307.8	309.4	311.0	312.6	314.2	315.8	317.4
		319.1	320.7	322.4	324.0	325.7	327.4	329.0	330.7	332.4	334.1	335.9	337.6
		339.3	341.1	342.8	344.6	346.4	348.2	349.9	351.7	353.6	355.4	357.2	359.0
		360.9	362.7	364.6	366.5	368.4	370.3	372.2	374.1	376.0	377.9	379.9	381.8
		383.8	385.8	387.7	389.9	391.8	393.8	395.8	397.8	399.9	402.0	404.0	406.1
		408.2	410.3	412.4	414.5	416.7	418.8	421.0	423.1	425.3	427.5	429.7	431.9
		434.1	436.4	438.6	440.9	443.1	445.4	447.7	450.0				



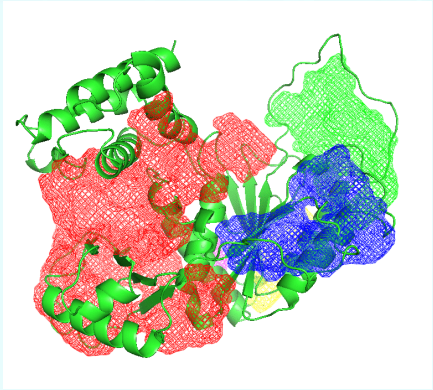
- **Similar transition pathways and free energy landscapes are obtained by vsREMD and REMD;**
- **All the crystal structures are sampled in the area with low free energies.**

Biophys J, 2020, **118**, 1009-1018.

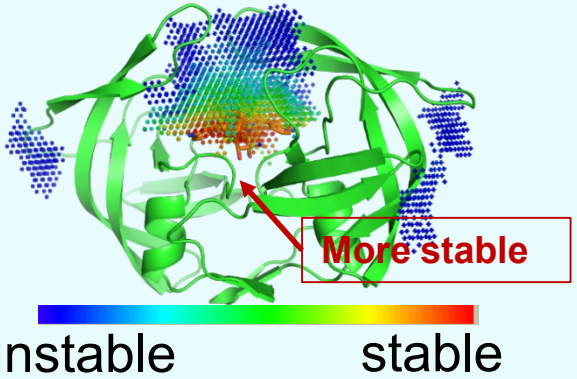
Challenge: it is hard to simulate the transition from DFG-in to DFG-out of kinase.

2. Quantitative Analysis Method for Pocket Dynamics Behavior: D3Pockets

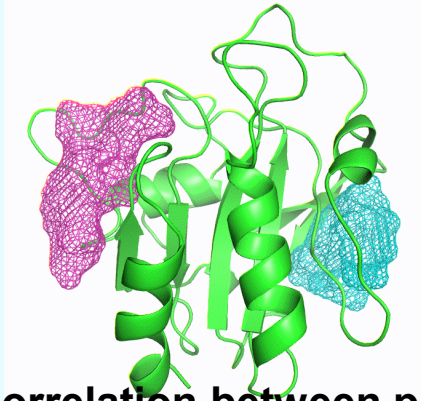
D3Pockets for quantitatively analyzing dynamic behavior



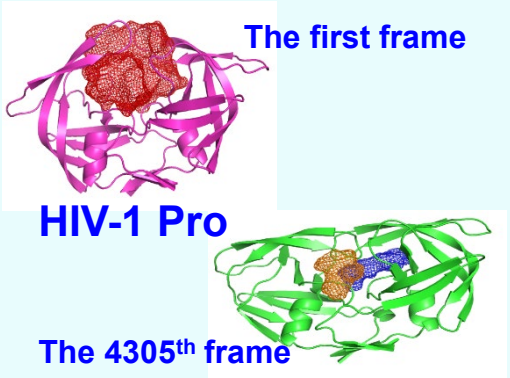
Drug binding pockets are constantly dynamic.



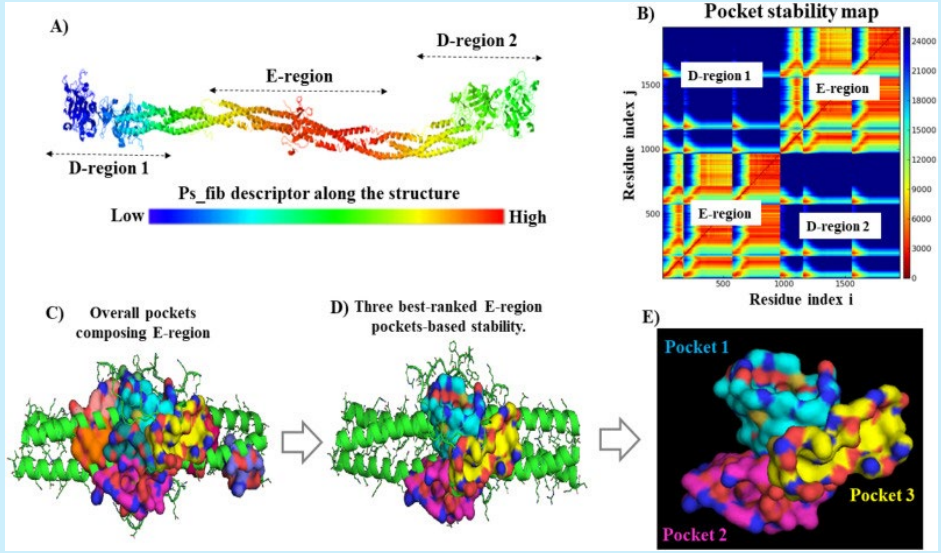
Pocket Stability



Correlation between pockets Allosteric site?



Pocket Continuity



Professor Rusoa *et. al* at the University of Santiago de Compostela in Spain used D3Pockets to compute the dynamic properties of pockets within the fibrinogen protein, which was employed to assess the biochemical correlations between significant regions of the protein. (*Journal of Molecular Liquids* **2021**, *324*)

J. Chem. Inf. Model. **2019, Most read article**

Outline

1. Introduction

2. New Methods for Protein Conformation Sampling

3. D3CARP for Target Prediction and Virtual Screening

4. Applications

3. D3CARP: A Comprehensive Platform for Target Prediction and Virtual Screening

Target Structure Database Construction

PDBbind Database(v2020)

PDBbind

J Med Chem, 2005, 48: 4111-9.

■ **Molecular Docking Software**

AutoDock Vina (v1.2.0)

■ **Pre-docking**

- Docking score<-5 kcal/mol
- RMSD<2 Å
- **9352 3D conformations of 1970 proteins with active ligands**

■ **Output Results**

- Docking scores, atomic efficiency, 2D/3D similarities, score ratio, disease information
- Positive control compounds and literature sources

Ligand Database Construction

BindingDB Database(2021.11.1)

BindingDB

Nucleic Acids Res, 2016, 44: D1045-53.

■ **Ligand Database**

- **1.01 million active compounds，7362 targets associated with 2168 disease types**
- **2.19 million drug-target pairs**

■ **Standard Dataset (K_i/K_d/IC_{50}/EC_{50}<10 μM)**

- 0.79 million ligands，5901 targets
- 1.20 million drug-target pairs

■ **Normalized Dataset (K_i<1 M)**

- 0.19 million ligands，2568 targets
- 0.34 million drug-target pairs
- Standardized inhibition constant nK$_i$: $nK_i = \frac{lgK_i}{lgK_{i_{max}}}$
- Inhibition constant K$_i$: $K_i = 10^{nK_i \times lgK_{i_{max}}}$

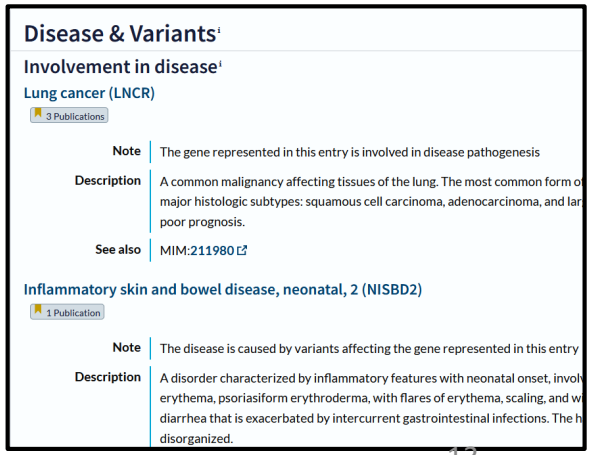
Disease Database Construction

Therapeutic target database

Gene Name	EGFR
Target Type	Successful target
Disease	[+] 9 Target-related Diseases
1	Angina pectoris [ICD-11: BA40]
2	Breast cancer [ICD-11: 2C60-2C6Y]
3	Colorectal cancer [ICD-11: 2B91]
4	Diabetic foot ulcer [ICD-11: BD54]
5	Ischemia [ICD-11: 8B10-8B11]
6	Lung cancer [ICD-11: 2C25]
7	Renal cell carcinoma [ICD-11: 2C90]
8	Solid tumour/cancer [ICD-11: 2A00-2F9Z]
9	Unspecific body region injury [ICD-11: ND56]

Ref: Nucleic Acids Res, 2020, 48: D1031-D1041.

UniProtKB



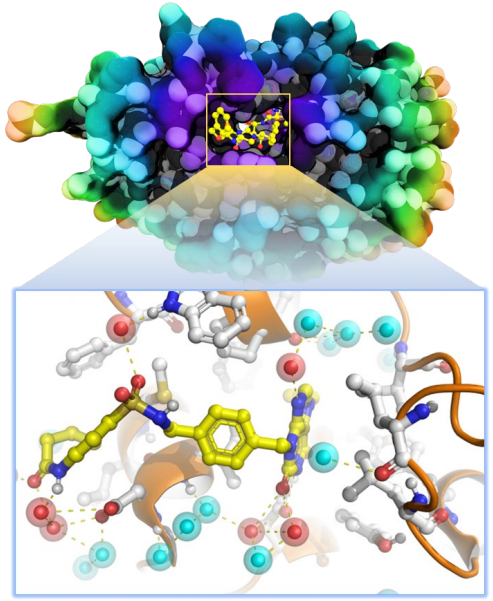
Ref: Nucleic Acids Res, 2021, 49: D480-D489.

3. D3CARP: A Comprehensive Platform for Target Prediction and Virtual Screening

Docking-Based

(Multi-conformation and binding sites)

9352 docking models for 1970 proteins

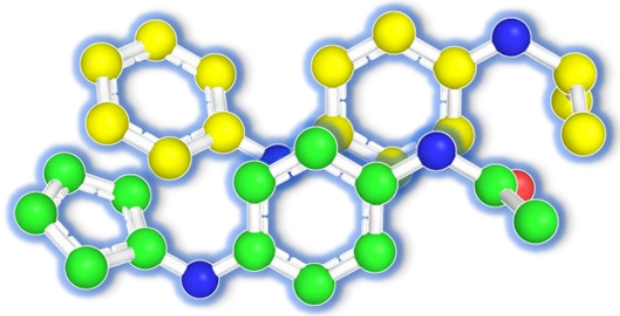


- **Molecular Docking Software**
  AutoDock Vina (v1.2.0)
- **Format Conversion**
  MGLTools program
- **Grid Box**
  Ligand-based extension 5 Å

Ligand-Based

(Similarity search)

2.19 million drug-target pairs



2D Similarity

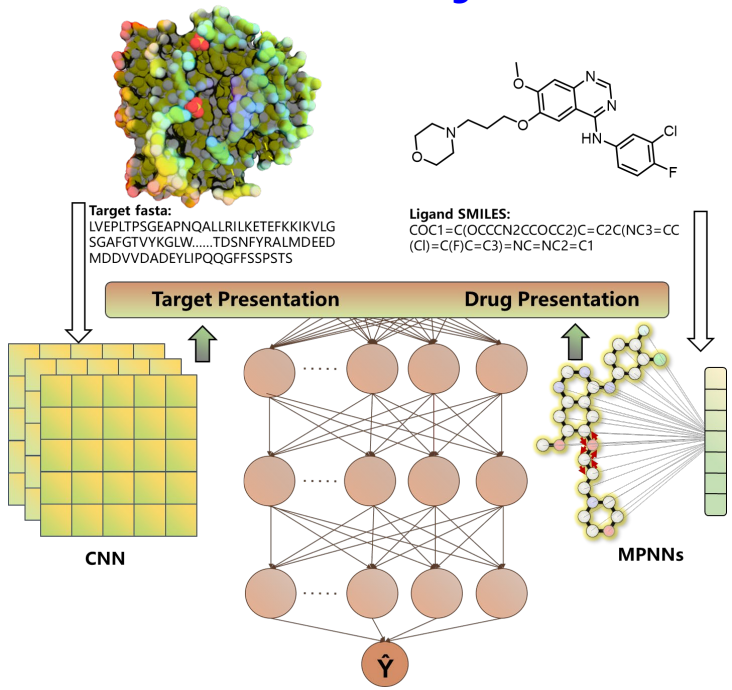
- Software: Open Babel
- Molecular fingerprint: FP2, FP4, MACCS

3D Similarity

- Conformation generation: RDKit
- Software: LS-align
- Rigid-LS-align
- Flexi-LS-align

Deep Learning Based

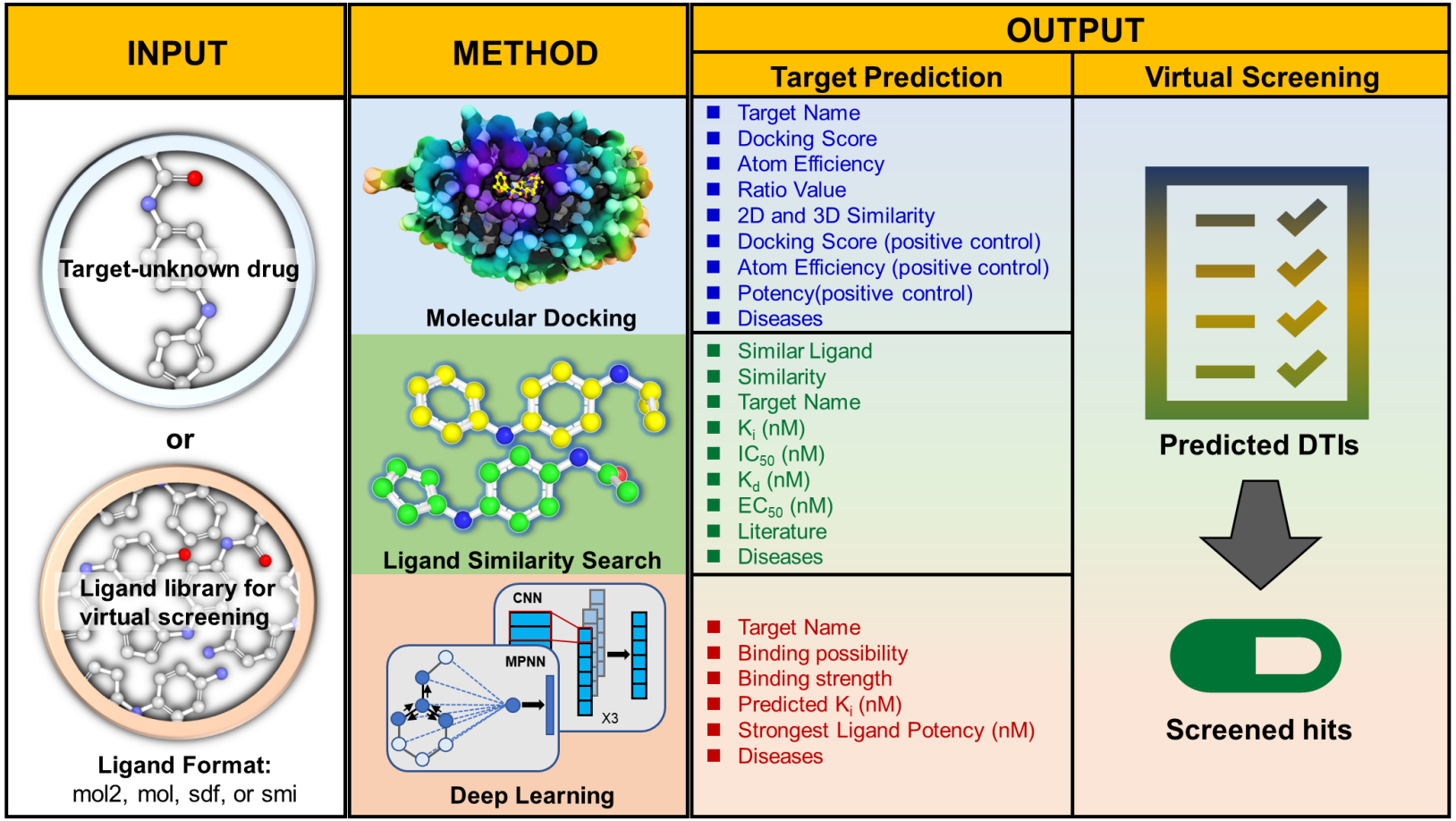
Classification and regression models



- Compound: MPNN
- Target protein: CNN
- Classification model: MPNN-CNN
- Regression model: MPNN-CNN-R
- Partition Dataset partition: 98:1:1

3. D3CARP: A Comprehensive Platform for Target Prediction and Virtual Screening

Input and Output



Platform Features

- **Multiple Drug Design Approaches**
  Docking, ligand similarity and DL
- **Ensemble Docking**
  Multi-target conformations, 9000+
- **Positive Controls as References**
  for 7000+ targets
- **Diverse Disease Types**
  Conduct computational research on specific disease targets (2000+ disease)
- **Cross-Validation of Prediction Results**
- **Wide Range of Application Scenarios**
  Target prediction, virtual screening, drug-target interaction mechanism, molecular scaffold novelty assessment, etc.

Free Application Website D3CARP: https://www.d3pharma.com/D3CARP/index.php

Comput. Biol. Med. **2023,** *164,* 107283

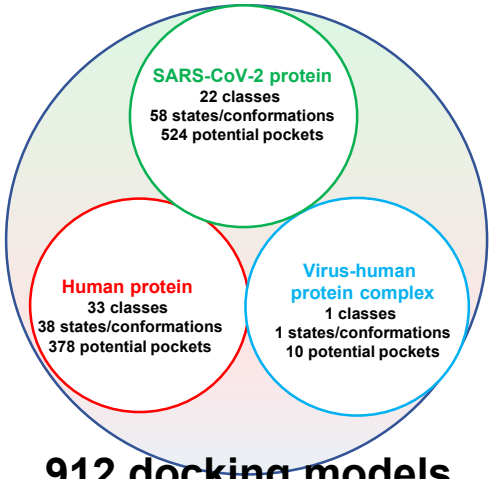
D3CARP-similar platform for COVID-19

4. D3Targets: a Multi-Target and Multi-Conformation based Docking Platform for COVID-19

Research background

Challenges:

- Unclear Targets for many reported "effective drugs"
- Urgent Need for New Drug Development Against Specific Targets

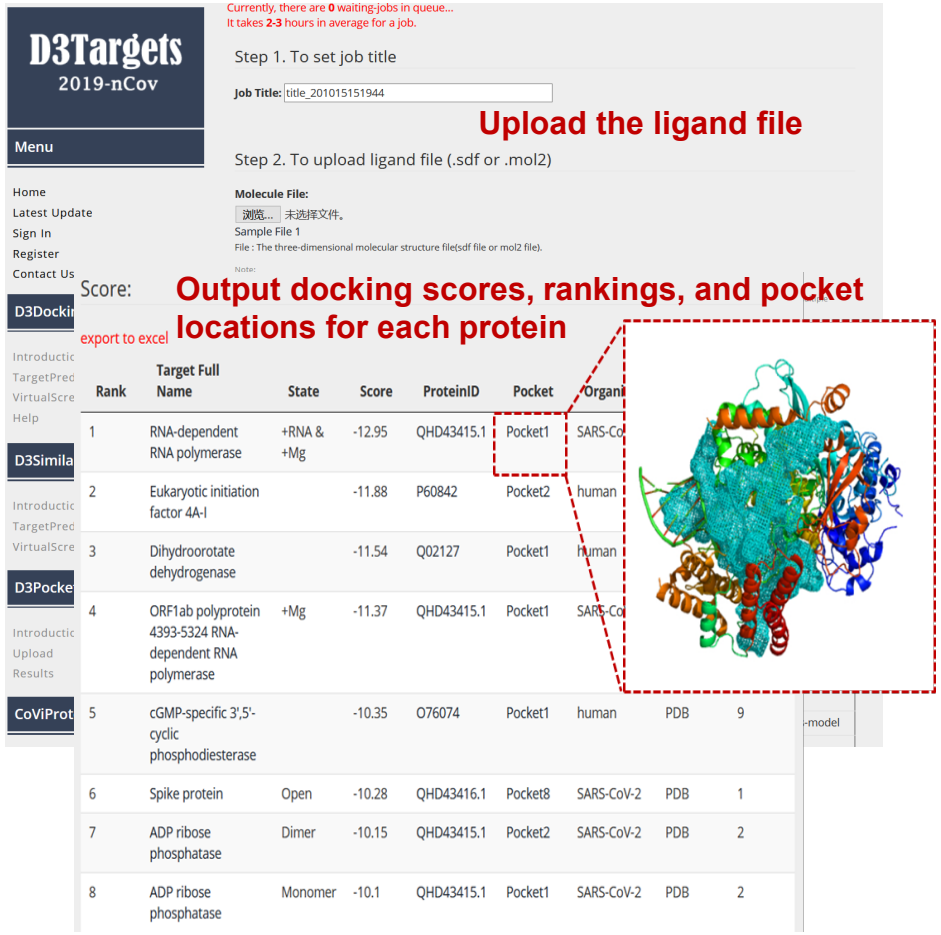


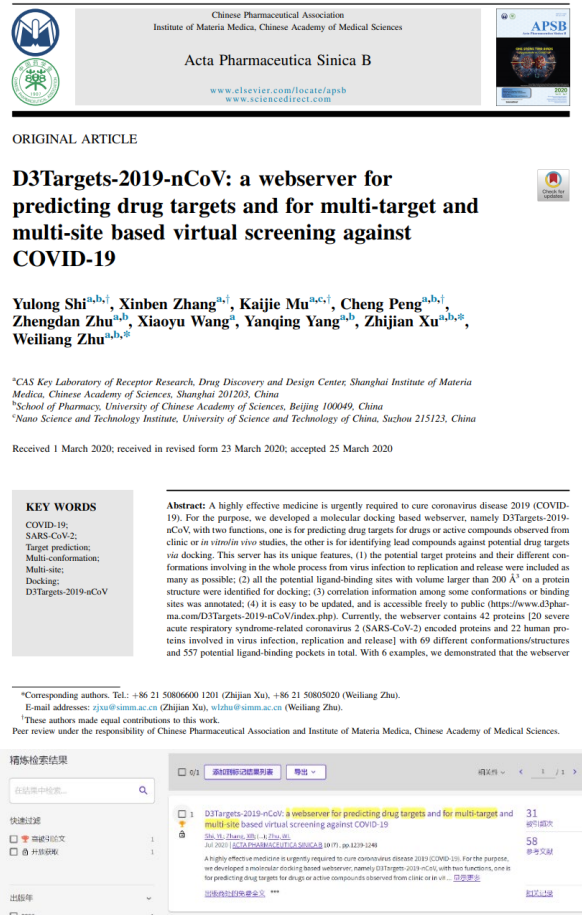
912 docking models for 56 proteins

Functions:

- Prediction of the target for active compounds
- Multi-target, multi-conformation, multi-site-Based Drug Virtual Screening

Usage





The article was recognized as the high-impact outstanding paper of APSB 2020.

Acta Pharm Sin B, 2020, IF: 11.4

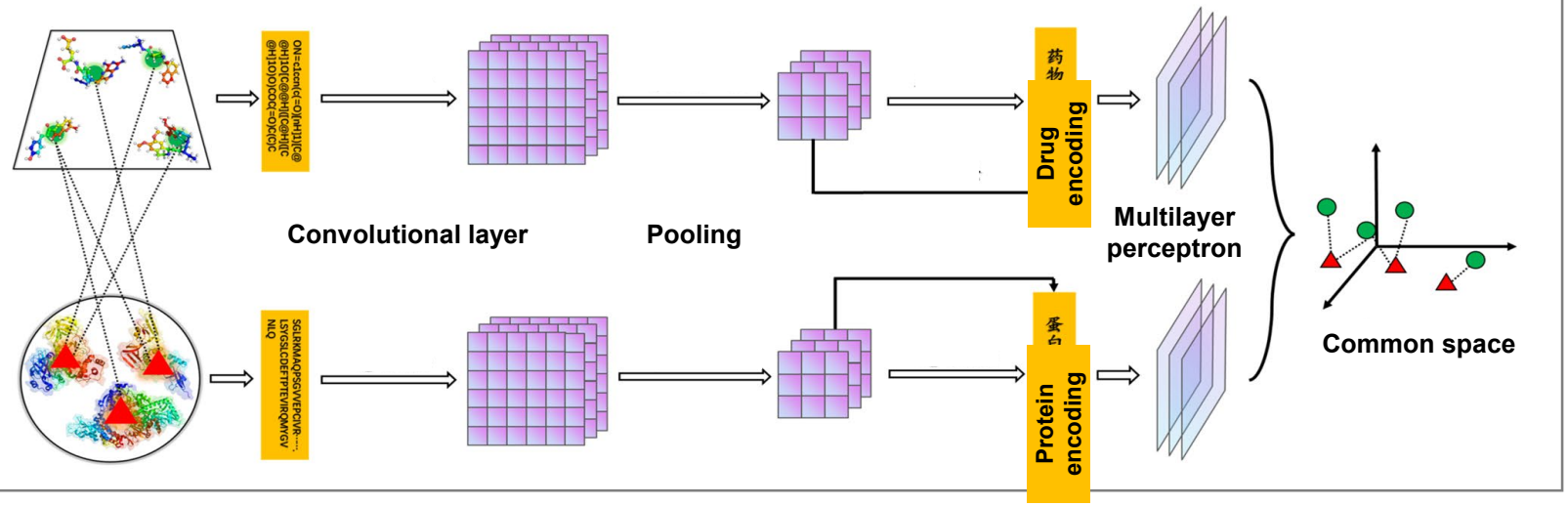
Online server

https://www.d3pharma.com/D3Targets-2019-nCoV/D3Docking/index.php

5. D3AI-CoV: An AI-Based Platform for COVID-19 Target Prediction and Virtual Screening

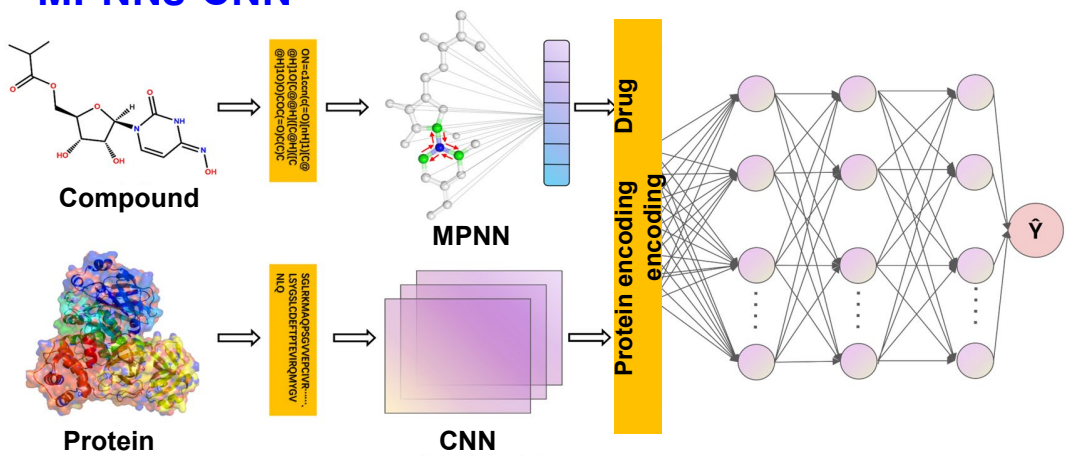
Multi-modal representation learning based approach

MultiDTI



Message Passing Neural Network/Convolutional Neural Network/Regression

MPNNs-CNN



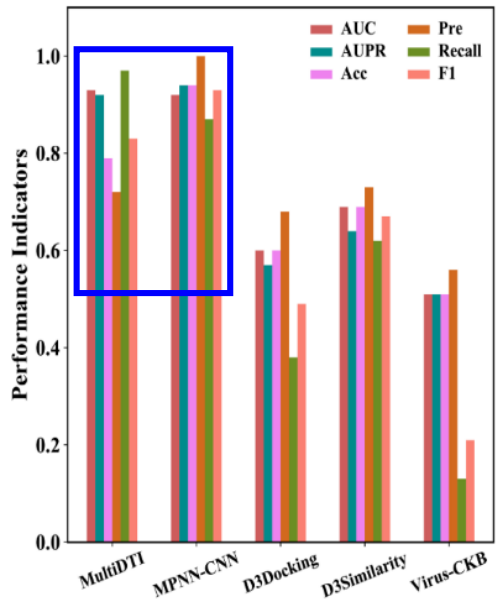
MPNNs-CNN-R

- Normalization of Activity Data.

$$score = 2 - lg(activity)$$

- MPNNs-CNN-R was used to explore the relationship between molecular targets and activity.

Performance comparison with other reported methods



- **The First Web Application Platform Applying Deep Learning Models for COVID-19 Target Prediction.**
- **The First AI-based Regression Model for COVID-19 Virtual Screening.**

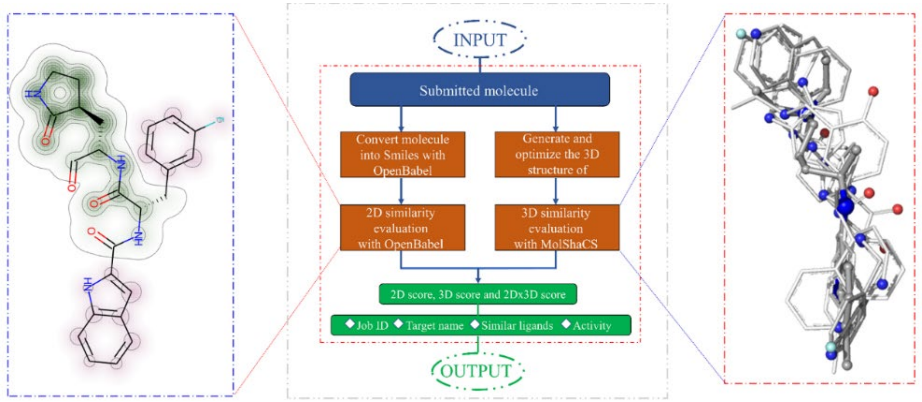
Briefings in Bioinformatics, 2022, IF=11.622

6. D3Similarity: Target Prediction and Virtual Screening Method Based on Ligand Similarity

Database Construction (in 178 Research Papers)

- 7 Pathogenic Coronaviruses
- 32 Target Proteins
- 604 Active Compounds

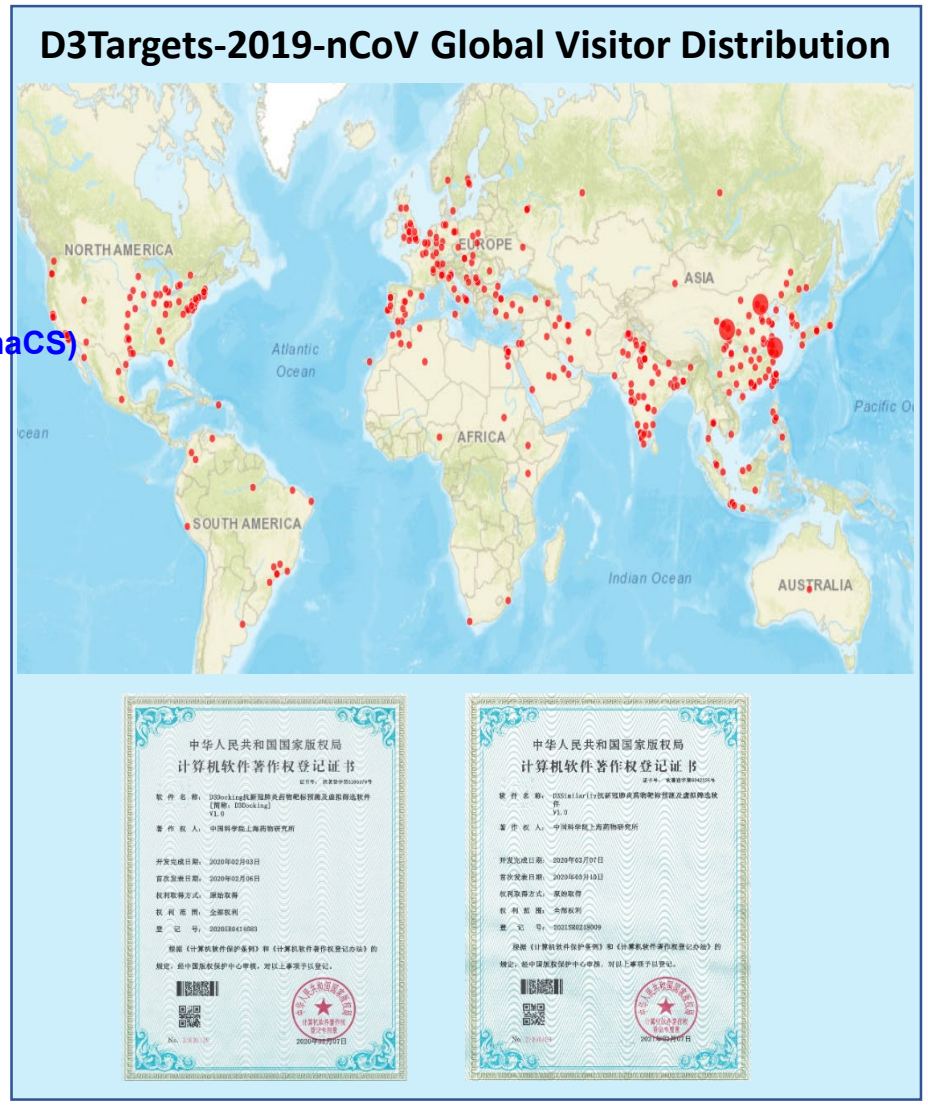
Similarity Comparison and Search (Open Babel, RDKit, MolShaCS)



Briefings in Bioinformatics, 2021, IF=11.622

Online server

http://www.d3pharma.com/D3Targets-2019-nCoV/D3Similarity/index.php



- **>3 Million Pageviews, from Over 60 Countries and Regions Worldwide.**
- **The peak task queue reached 355.**
- **D3Docking was used to study the affinity of bioactive compounds in the Huoshiluo formula with SARS-CoV-2-related proteins by Chen et..**
  (*Ann Palliat Med*, 2021)
- **Researchers from Peking University utilized D3Similarity to evaluate the novelty of their newly discovered main protease inhibitors.**
  (*Brief. in Bioinf.*, 2021)

Outline

1. Introduction

2. New Methods for Protein Conformation Sampling

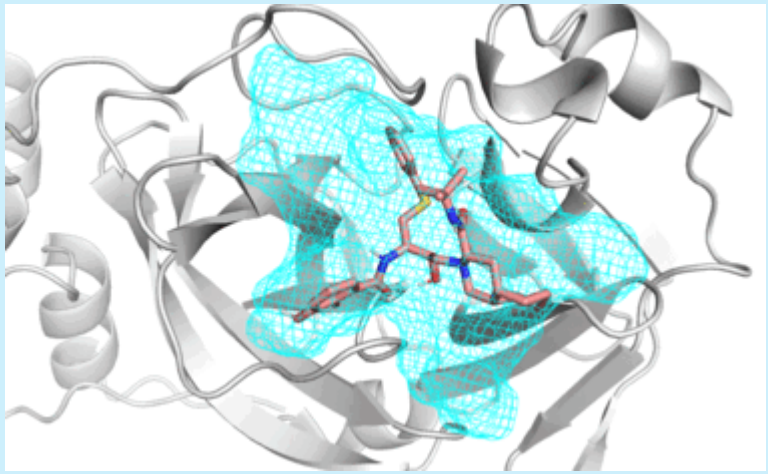
3. D3CARP for Target Prediction and Virtual Screening

4. Applications

7. Repurposing nelfinavir as a drug against SARS-CoV-2

Nelfinavir was predicted to be 3CL^{pro} inhibitor

Sequence with 98% similarity to SARS-CoV-1

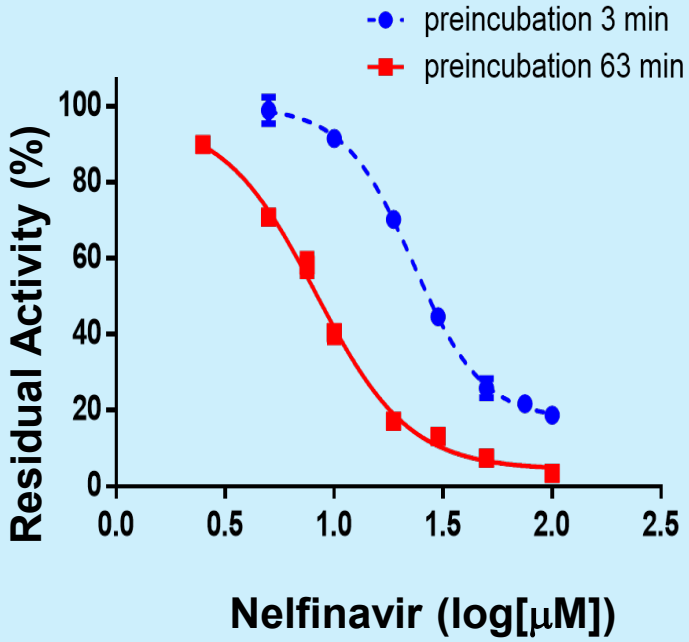


Molecular dynamics simulation

	nelfinavir	Pitavastatin	Perampanel	praziquantel
Docking Score	-9.18	-8.06	-8.63	-7.38
3D Similarity	70.2%	68.1%	66.9%	65.8%
ΔG(MM/GBSA)*	-24.69±0.52	-12.70±0.38	-14.98±0.34	-6.51±0.21
ΔG(SIE method)*	-9.42±0.04	-7.53±0.04	-7.55±0.03	-6.39±0.04

Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation *bioRxiv* **2020. 1. 27**

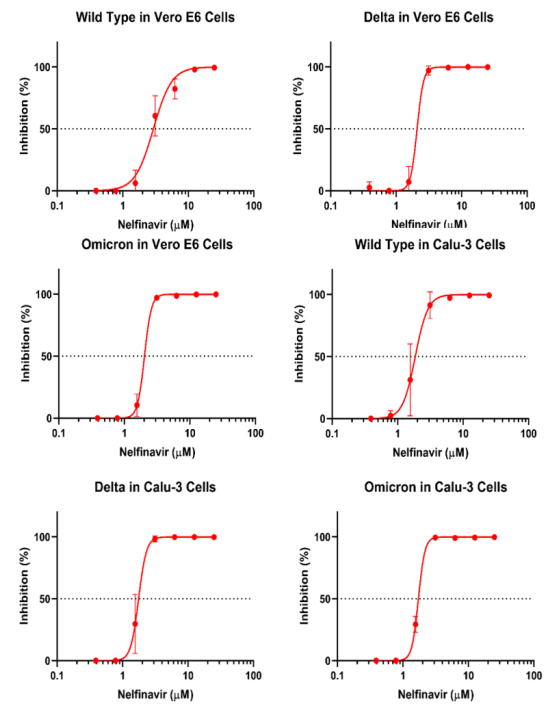
Activity at 3CL^{pro} level



IC_{50} values of 23.04±1.02 μM (3-min incubation) and 8.26±1.04 μM (63-min incubation)

Repurposing nelfinavir as a drug against SARS-CoV-2

Bioassay at cellular level



In vitro inhibition of nelfinavir against wild type, Delta, Omicron SARS-CoV-2 in Vero E6 and Calu-3 cells.

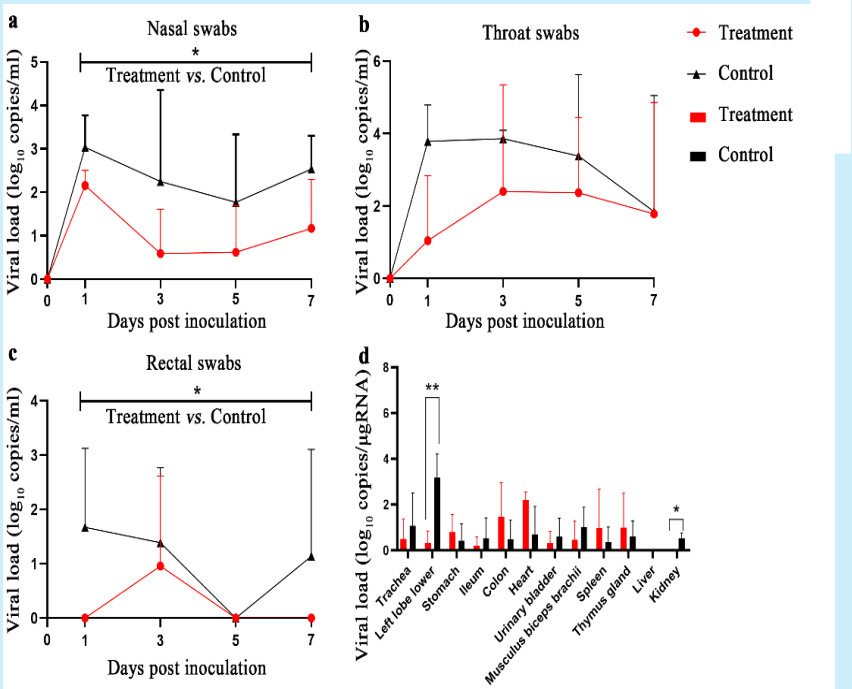
ChemRxiv, 30 Mar 2020

https://doi.org/10.26434/chemrxiv.12039888.v1

Inhibit SARS-CoV-2 in Vero E6 (EC_{50}=2.89 μM)

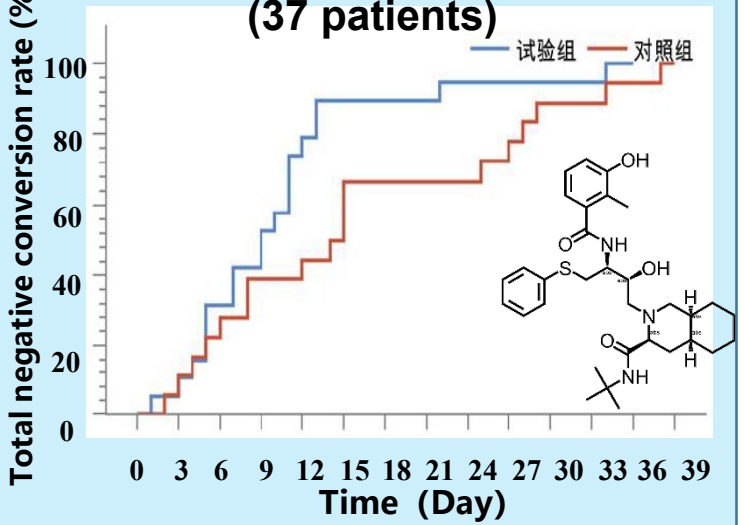
(Remdesivir: ~2 μM)

Preventive benefits in rhesus macaques



- **Significantly reduced virus loads in the lungs and kidneys;**
- **Significantly reduced virus loads in the nasal and anal swabs;**
- **Virus shedding was not reduced by remdesivir. (** ref.: *Nature* **2020,** *585*, 273)

Preliminary clinic trial (37 patients)



Shorten the duration of viral shedding by 5.5 days (9.0 vs. 14.5 days, *P* = 0.055) and the duration of fever time by 3.8 days (2.8 vs. 6.6 days, *P* = 0.014)


Signal Transduction and Targeted Therapy

www.nature.com/sigtrans

ARTICLE OPEN

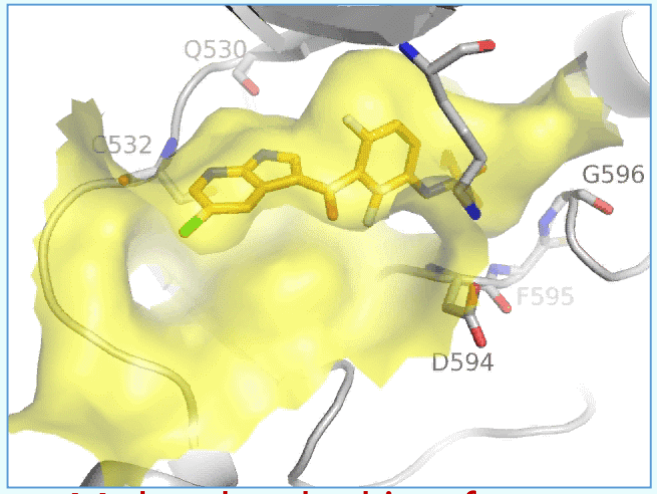
Preventive and therapeutic benefits of nelfinavir in rhesus macaques and human beings infected with SARS-CoV-2

Zhijian Xu[1,2,3], Danrong Shi[4], Jian-Bao Han[5], Yun Ling[6], Xiangrui Jiang[1,3], Xiangyun Lu[4], Chuan Li[1,3], Likun Gong[1,3,7], Guangbo Ge[8], Yani Zhang[9], Yi Zang[1,3], Tian-Zhang Song[5,9], Xiao-Li Feng[5], Ren-Rong Tian[5,9], Jia Ji[4], Miaojin Zhu[4], Nanping Wu[4], Chunhui Wu[1,3], Zhen Wang[1,3], Yechun Xu[1,2,3], Cheng Peng[1,2,3], Min Zheng[4], Junling Yang[1,3], Feifei Du[1,3], Junliang Wu[1,3,7], Peipei Wang[1,3], Jingshan Shen[1,3✉], Jianliang Zhang[6✉], Yong-Tang Zheng[5,9✉], Hangping Yao[4,10✉] and Weiliang Zhu[1,2,3✉]

Effective drugs with broad spectrum safety profile to all people are highly expected to combat COVID-19 caused by SARS-CoV-2. Here we report that nelfinavir, an FDA approved drug for the treatment of HIV infection, is effective against SARS-CoV-2 and COVID-19. Preincubation of nelfinavir could inhibit the activity of the main protease of the SARS-CoV-2 (IC_{50} = 8.26 μM), while its antiviral activity in Vero E6 cells against a clinical isolate of SARS-CoV-2 was determined to be 2.93 μM (EC_{50}). In comparison with vehicle-treated animals, virus [illegible] significantly [illegible] the viral replication in the lungs by nearly three orders of magnitude. A prospective clinic study with 37 enrolled treatment-naive patients at Shanghai Public Health Clinical Center, which were randomized (1:1) to nelfinavir and control groups, showed that the nelfinavir treatment could shorten the duration of viral shedding by 5.5 days (9.0 vs. 14.5 days, P = 0.055) and the duration of fever

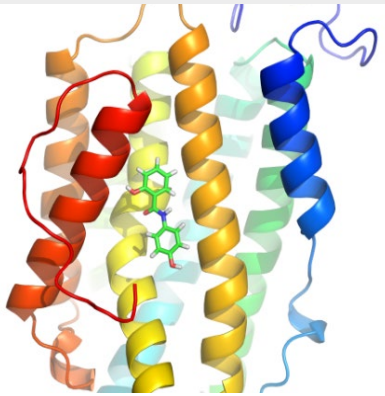
Signal Transduct. Target. Ther. 2023

8. Repurposing Osalmide as a drug against multiple myeloma (MM)



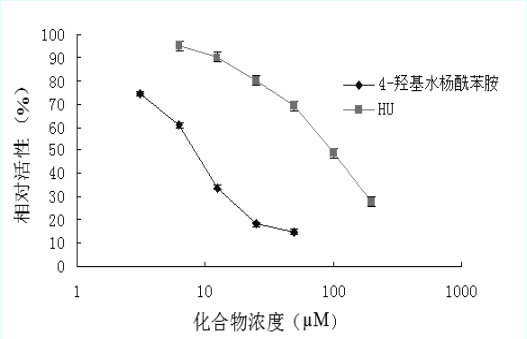
Molecular docking for ribonucleotide reductase RR

Osalmide for cholecystitis

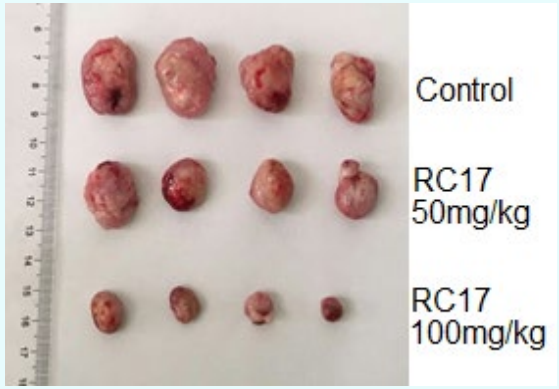


Docked structure of osalmide-RR complex

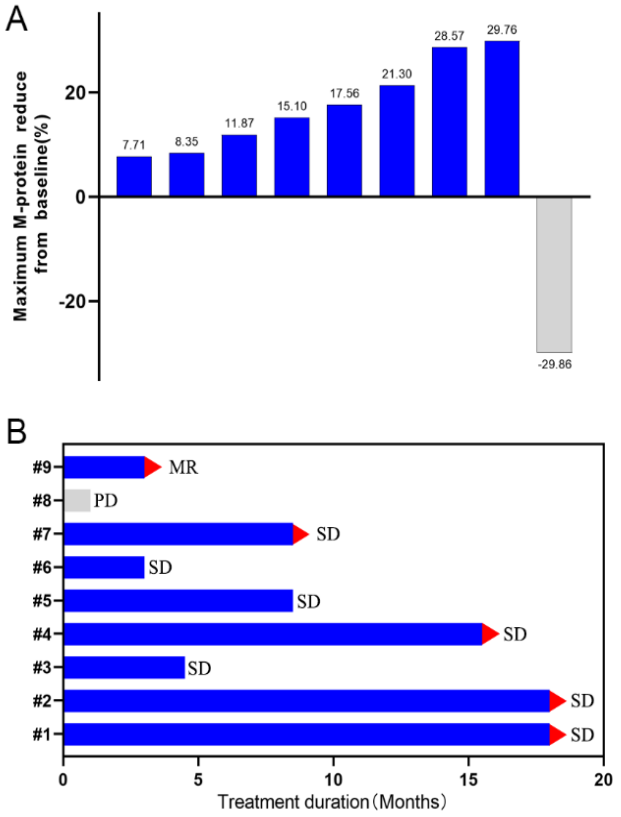
New application?



Much stronger than HU



In vivo activity on mouse model



Clinic study of osalmide on MM (SD: stable disease)

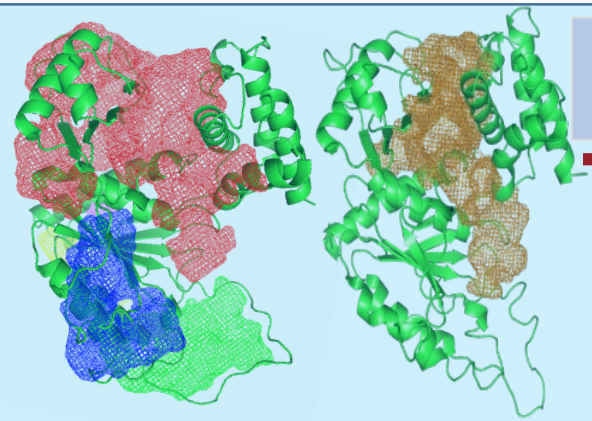
(A) The maximum change from baseline in the level of M-protein after HDS treatment. (B) Swim-lane plot show the treatment response and duration for 9 MM patients after HDS treatment. Arrows: still ongoing at the time of study closure. *J. Biomed. Sci.* **2022,** *29*, 32

9. The first inhibitor of TRIP13

MM patients show high level of TRIP13
(Shanghai Oriental Hospital)

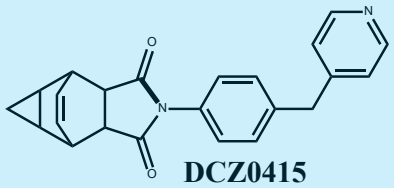
Homology modeling

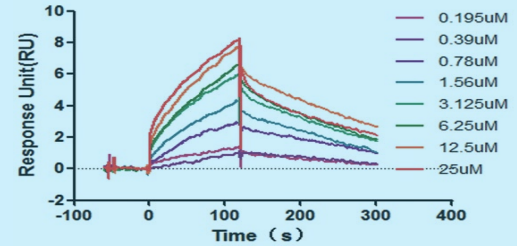
vsREMD simulation
D3Pockets

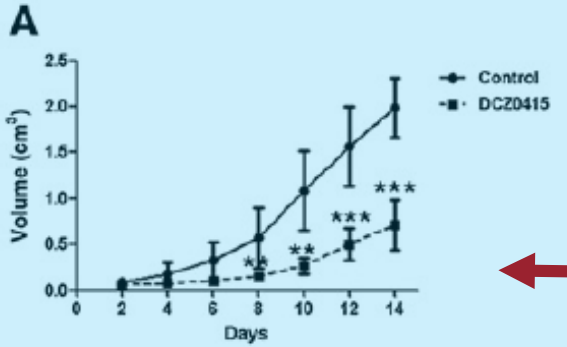


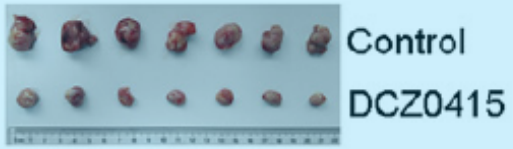
Multiple sites docking

Design and synthesis



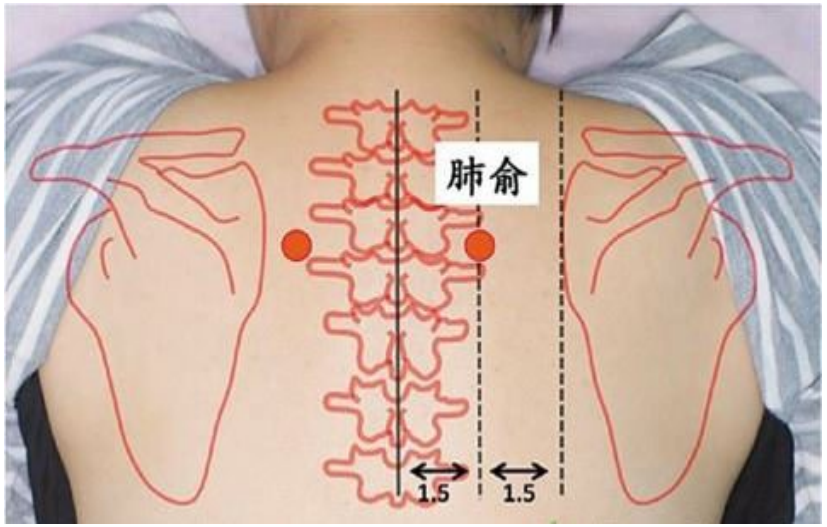






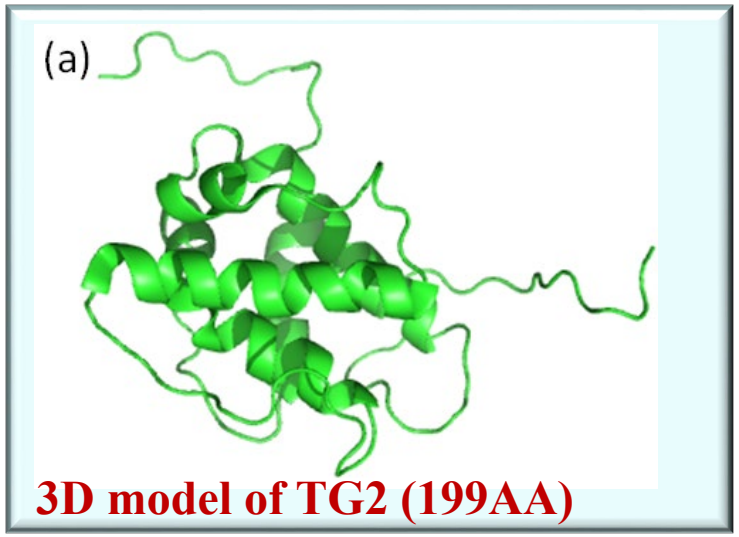
Cancer Res. 2020, *80(3), 536* (IF=9.7)

10. Discovering the first agonist of TG2

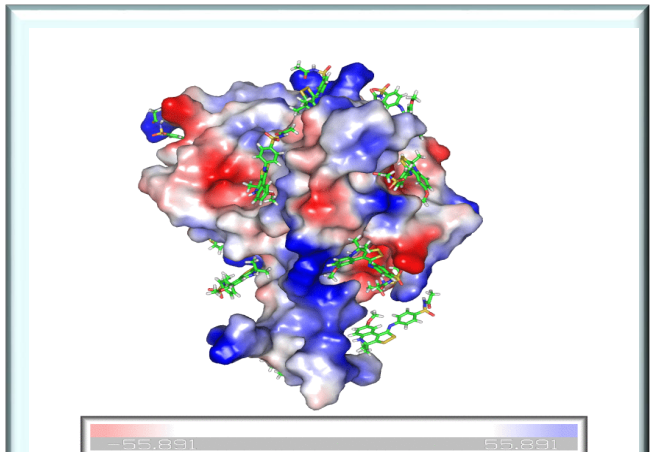


Asthma → **Acupuncture** → Collaborator → TG2 pathway activated

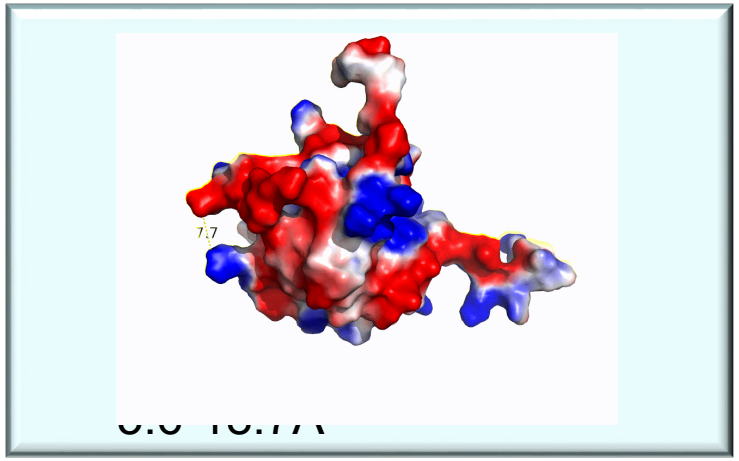
TG2 agonist as anti-asthma drugs?



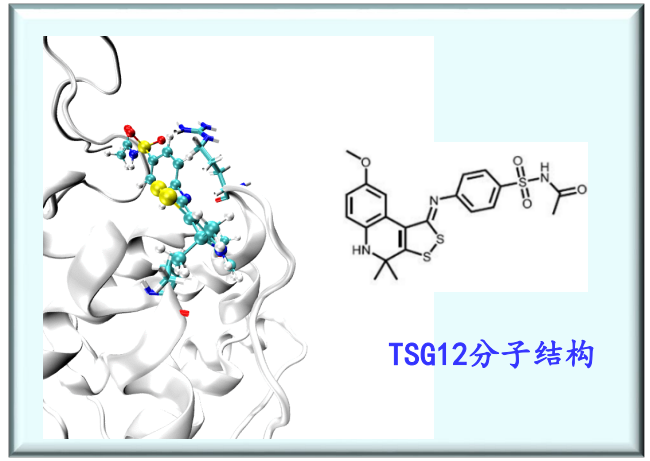
3D model of TG2 (199AA)



Multiple sites screening

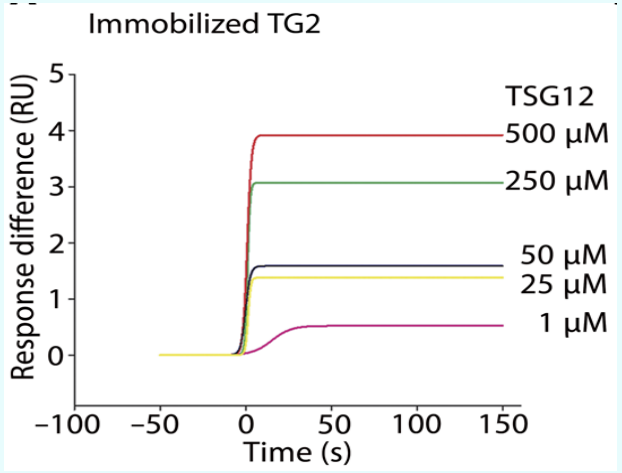


Dynamic profile of TG2

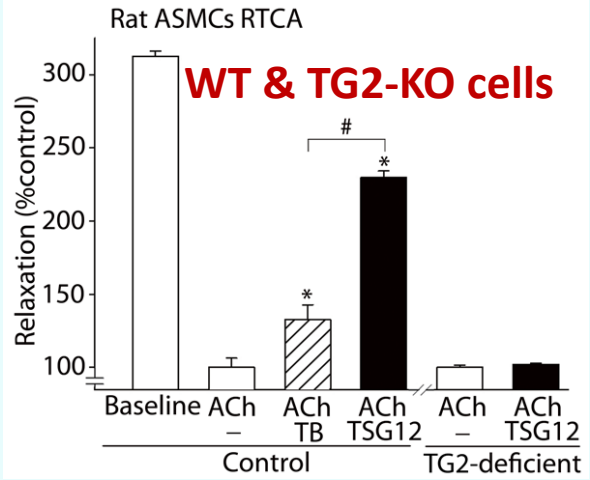


Discovered agonist TSG12 on TG2

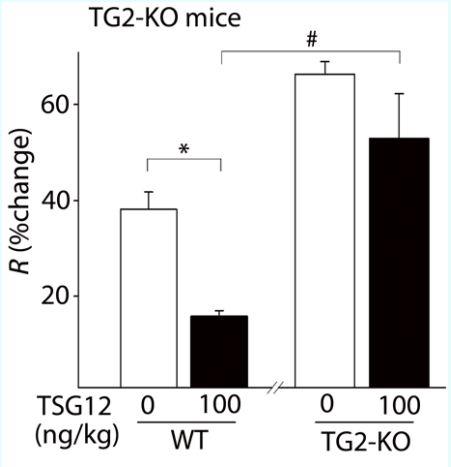
10. Discovering the first agonist of TG2



TG2-TSG12 binding (Biacore)

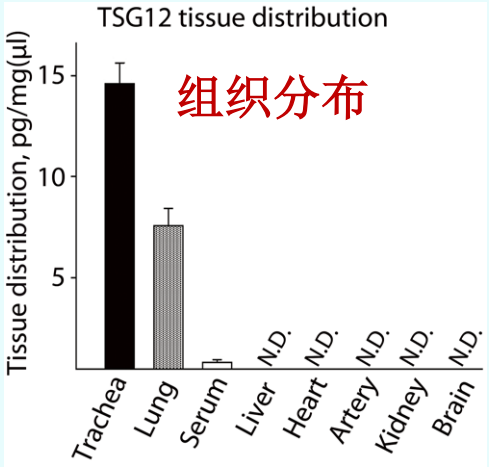


At airway smooth muscle cell (ASMC) level

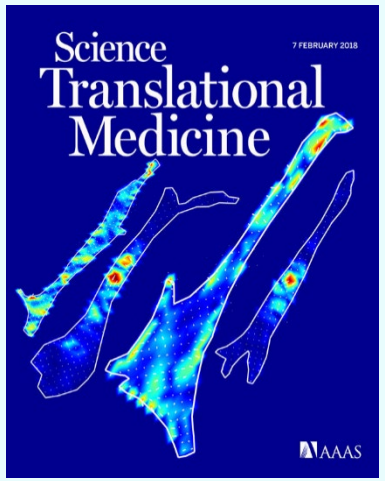


WT & TG2-KO mice

Aerosol administration

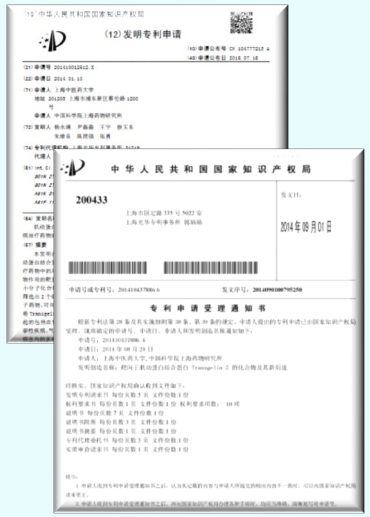


Tissue distribution



Sci. Trans. Med., **2018**, *10*

Cover article



Patents

- **One of the "Top 10 Academic Advances in Traditional Chinese Medicine in 2020"**
- ***Nature* cited it as a model of modernization research in traditional Chinese medicine (special issue of "Focus on Traditional Chinese Medicine" 2021,)**

Research Team

1. Group members

- 3 Professors
- 1 Associate
- 1 Senior technician
- 1 Technician

2. Ongoing Projects Chaired

- 1 National key R&D project
- 1 National Excellent Young Scientists Fund
- 4 NSFC projects
- 6 Other projects

3. Research fields: Drug design and innovative drug development

- Development of new methods and theories for drug design
- Discovery and optimization of lead compounds
- Drug repurposing
- Traditional Chinese medicine and natural products based new drugs

4. Research methods

- Quantum Mechanics
- Molecular Dynamics
- Statistical Mechanics
- Deep Learning


Thank you for your attention!