FEP-based large-scale virtual screening for effective drug discovey against COVID-19 and clinical trials

Chengkun Wu, and colleagues

Our multidisciplinary team

- **FEP method development**


Zhe Li


Runduo Liu


Chang-Guo Zhan

- **Experimental validation**


Xin Wang

- **HPC implementation**


Chengkun Wu


Yishui Li


Meng-Xia Mo

- **Clinical studies**


Hai-Bin Luo


Fuling Zhou

- **HPC support**


Kai Lu


Ruibo Wang

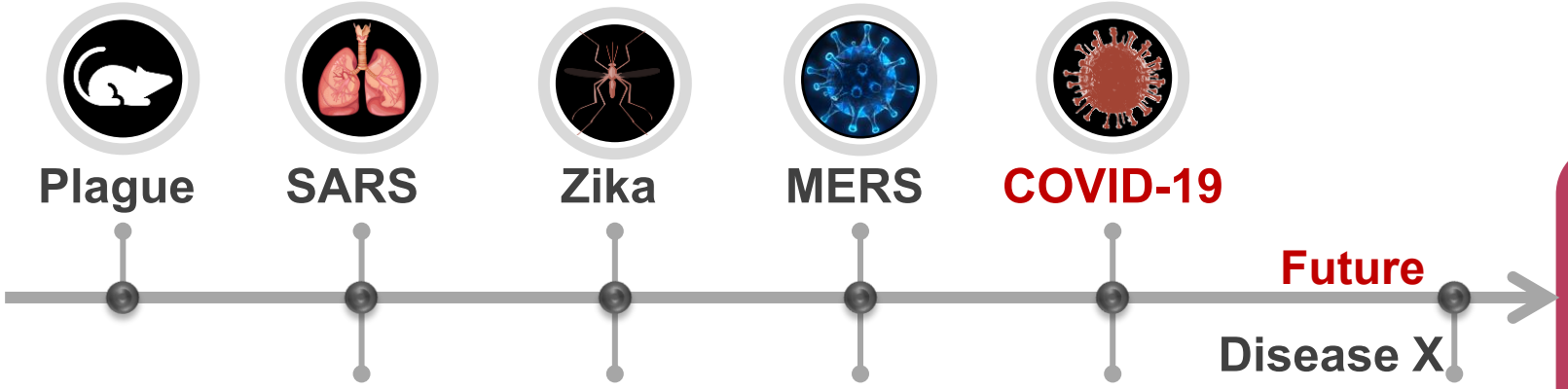

Jie Liu


Chunye Gong


Canqun Yang

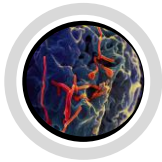
Contents

Emerging and re-emerging infectious diseases are changing our world



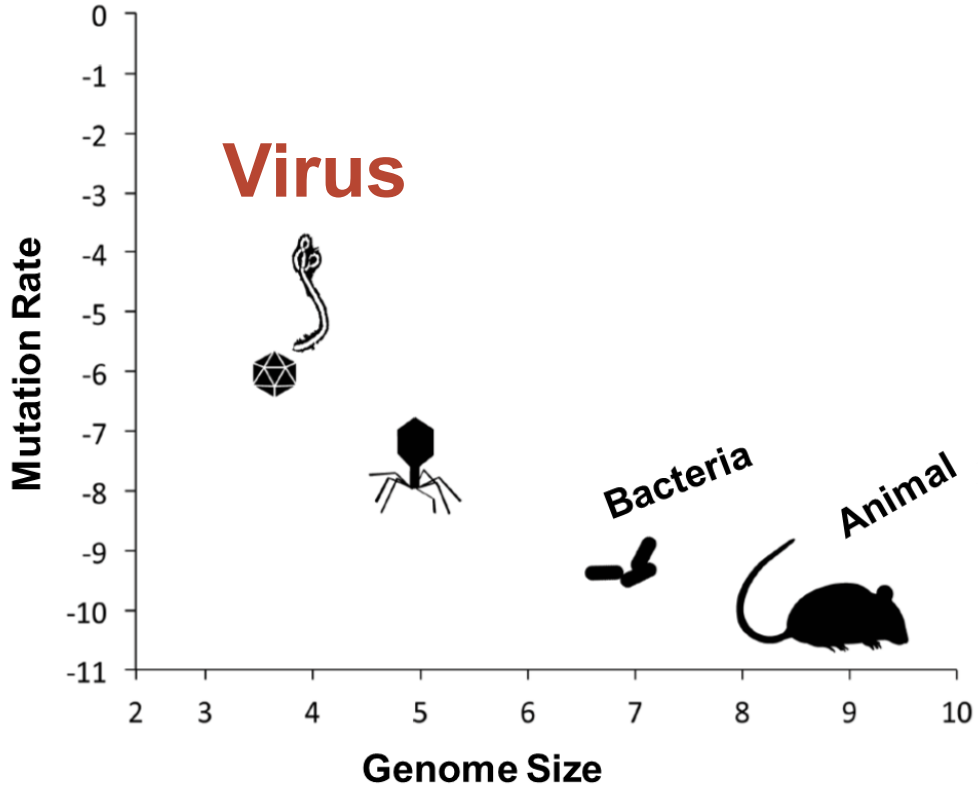
What can we do in fighting against future Disease X with pandemic potential?

Infection	8,096	894,231	2,578	> 250 M
Death	774	6	888	> 5 M
Economic Loss	> $400 B	$7-18B	$2.6 B	7% loss of global GDP in 2020

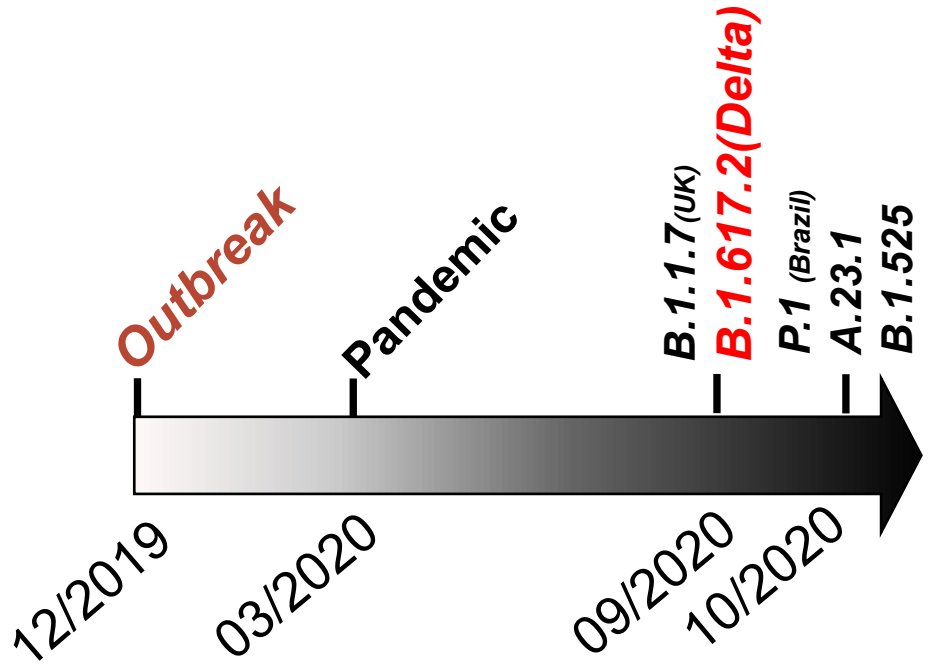


- Lee et al, Estimating the global economic costs of SARS, 2004;
- UNDP report,2017;
- Joo et al, Health Security, 2019;
- World Bank Report “Global Economic Effects of COVID-19”,2020

Virus SARS-CoV-2 has generated hundreds of variants within 0.5 year



Siobain Duffy, PLoS Biology, 2018. with modifications



Within 0.5 year hundreds of variants

https://covid.cdc.gov/covid-data-tracker

High Mutation Rate of Viruses

Emerging New Variants/Viruses

Drug discovery: Not fast enough against viruses with a high mutation rate

Veklury
(Remdesivir)

11 years

Compound screening → lead compound (GS 441524) → Preclinical study → Clinical study → FDA Emergency Use

2006 (SARS) → 2012 → 2017- (Ebola) → 2020- → 10/22/2020 **Redirection to COVID-19**

Tamiflu
(Oseltamivir)

3 years

Target → lead compound (GS 4071) → Preclinical study → Clinical study → FDA approval

1992 → 1995 → 1998 → 1999

CADD

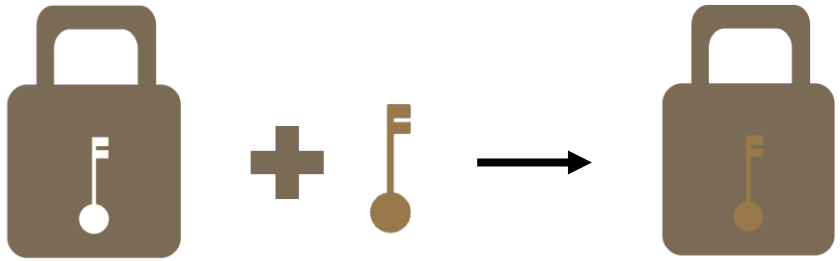
CADD (Computer-Aided Drug Design) has accelerated the drug discovery

BUT not fast enough!

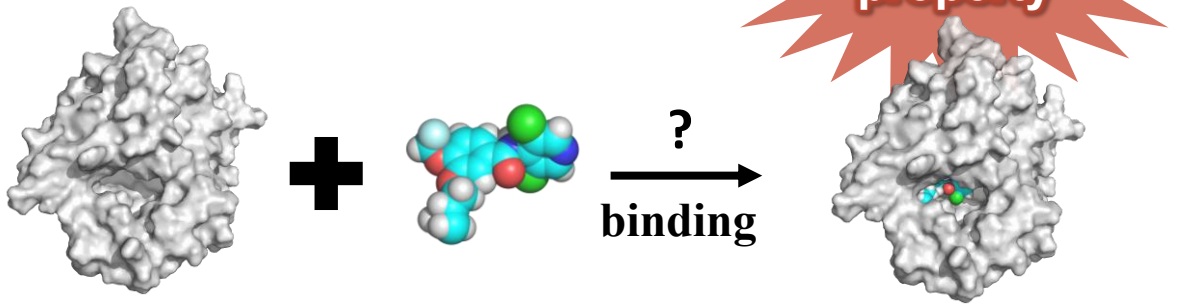
- Chu CK, Antimicrob Agents Chemother, 2006;
- Sheahan TP, Sci Transl Med. 2017;
- Eisenberg EJ, Antimicrob Agents Chemother. 1997;
- https://www.fda.gov/;
- https://clinicaltrials.gov/

CADD: Predicting binding affinity of each potential drug candidate with a given target

Lock and key model for drug-target interaction



A drug bind to its target like lock and key

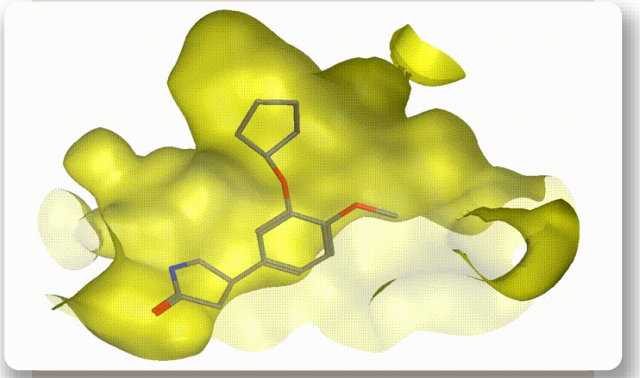


Key point in CADD:

To reliably predict binding free energy of each potential drug candidate with a given drug target

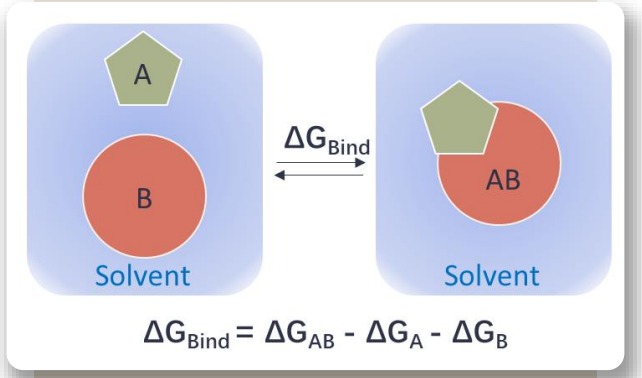
Limitation of traditional computational methods for binding free energy prediction

Scoring function
(Molecular docking)



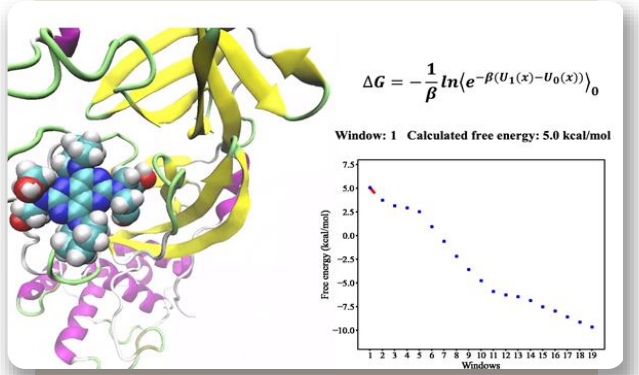
- **Fast** (billions of compounds)
- **Inaccurate**, low hit rate (~2%)[a]

End point methods
(MM-PBSA)



- **Moderate speed**
- **Moderate**, hit rate (<10%)[b]

Statistical mechanical methods
(FEP – Free Energy Perturbation)



- **Theoretically rigorous for relative binding free energy calculation**
- **Not designed for virtual screening**
- **Time consuming**

Unmet need: A truly accurate and efficient computational approach to absolute binding free energy calculations suitable for virtual screening

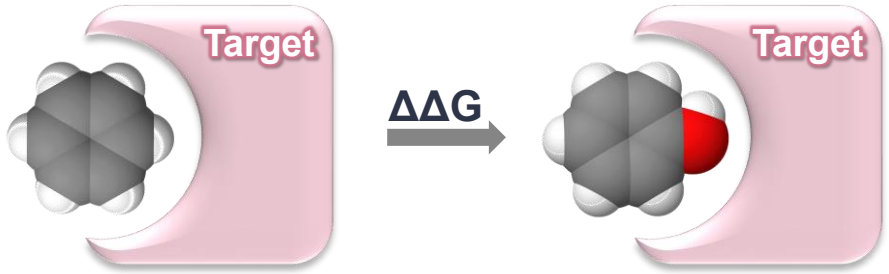
a. Gorgulla C, et al, Nature, 2020; b. Kuhn B, et al, J. Med. Chem. 2019

Contents

Why was FEP difficult for absolute binding free energy (ABFE) calculation?

FEP was designed to simulate a “perturbation” – a minor change of molecular structure;
Computational simulation of the *perturbation* is reliable only for a truly *minor* structural change.

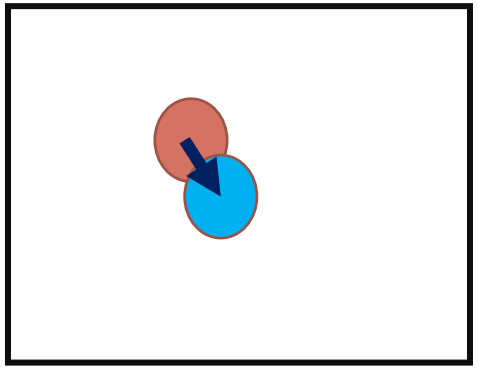
Current FEP: simulate a minor structural change



Relative binding free energy

- Changing < 10 atoms

Easy to calculate



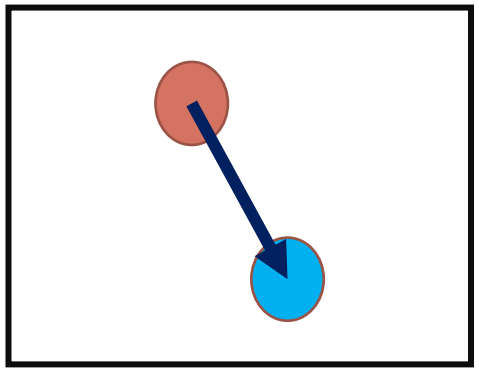
Phase space

Needed FEP: simulate disappearance of an entire molecule

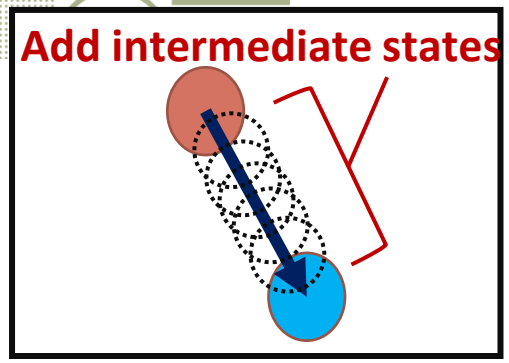
Absolute binding free energy

- Changing **50 ~ 100 atoms**

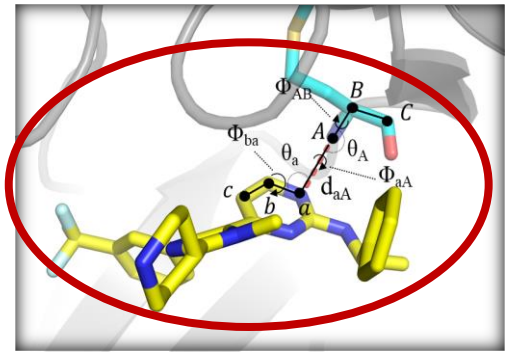
Difficult to calculate



Major problems preventing FEP-ABFE calculations-based virtual screening



- To deal with the large change, one must add many intermediate states, which means that one has to perform many FEP simulations for each FEP ABFE prediction
  --***Computationally time-consuming***



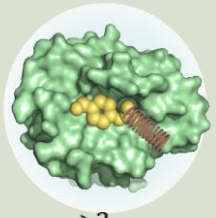
- Technically, to reliably evaluate conformational entropy contribution to ABFE, certain restraints are required. The choice of restraints required is case by case
  --***Difficult for automated virtual screening***

Our solutions:

- A restraint energy distribution (RED) function derived and used to minimize the # of intermediate states required for a converged ABFE calculation.
- A unique algorithm enabling to automatically identify restraints (with three ligand atoms and three target atoms, restrained to their equilibrium).

Performance of our novel approach to the conformational entropy estimation

Physical model derivation



$$U_{ini} = k_{ini}\,(r - r_0)^2 \qquad \text{(S1)}$$

$$\Delta U_{i+1,i} = \Delta\lambda_{i+1,i} k_{res} (r - r_0)^2 \qquad \text{(S2)}$$

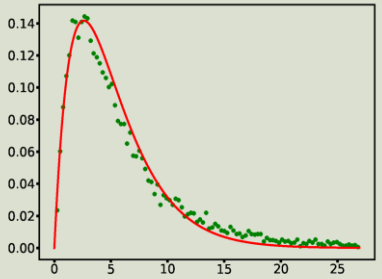
$$P(U_i) = \frac{exp(-\beta U_i)\Omega(U_i)}{Z} \qquad \text{(S3)}$$

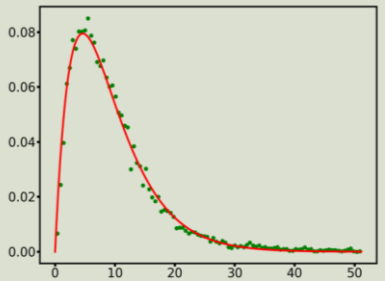
$$P(U_i) = \frac{exp(-\beta(k_{ini}+\lambda_i k_{res})(r-r_0)^2)4\pi(r-r_0)^2}{Z} \qquad \text{(S6)}$$

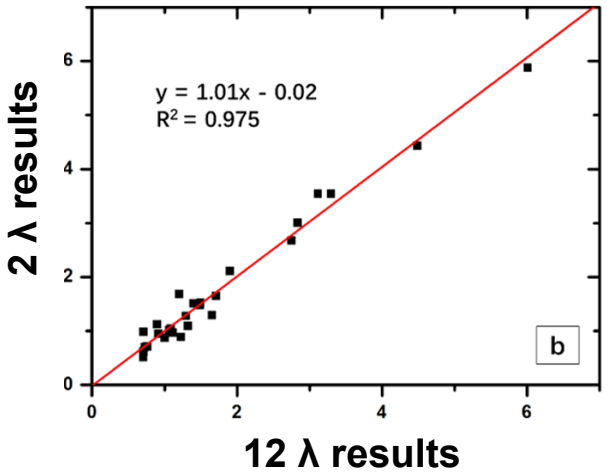
★ **RED function:**

$$P(\Delta U) = b^2 \cdot \exp(-b\Delta U) \cdot \Delta U$$

Finding the best match







- Fast
- Accurate
- Automatic

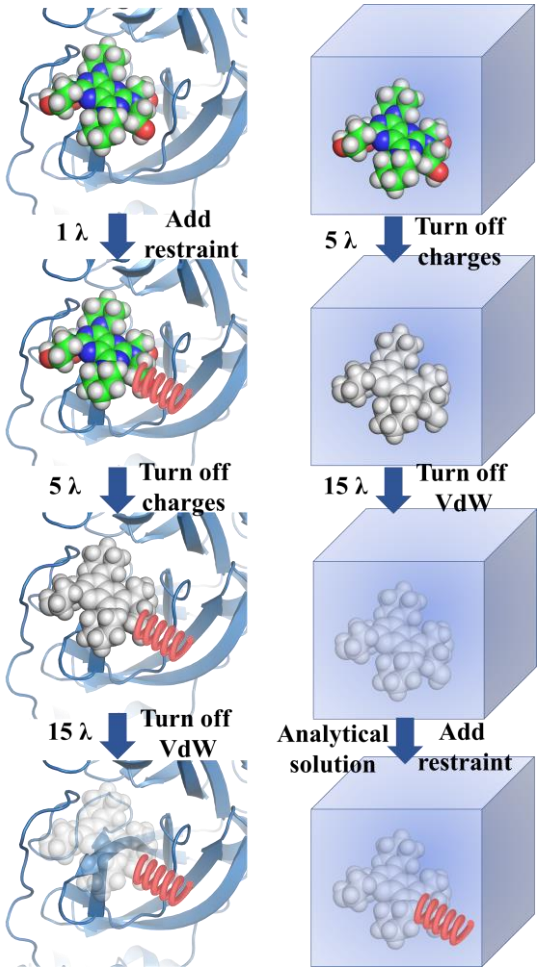
Z. Li, et al. Proc. Natl. Acad. Sci. USA 2020

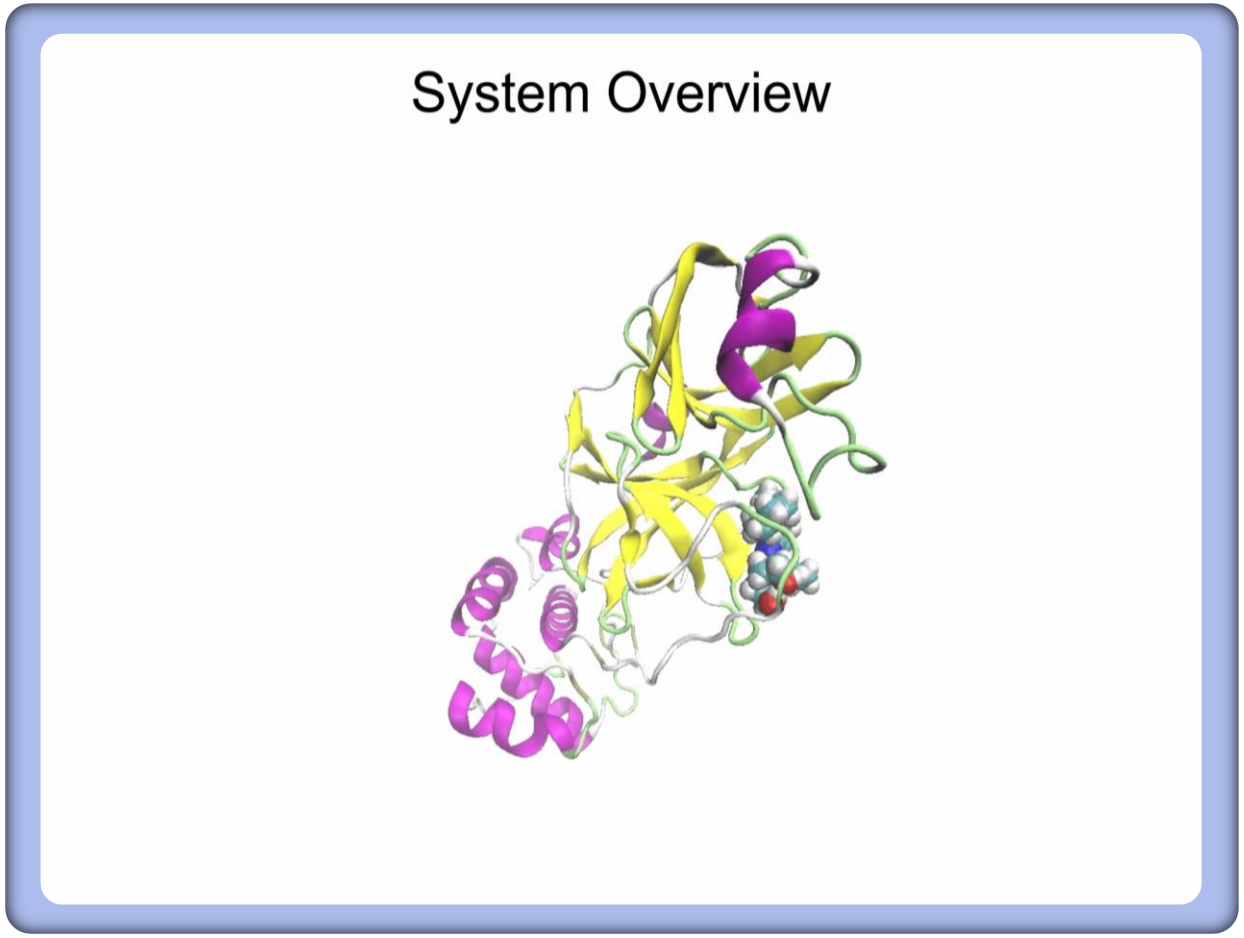
Designed algorithm

Conformational entropy estimation accelerated > 6 times

FEP-ABFE protocol used in this work

1. Pre-equilibrate MD
2. Automatic restraint addition
3. Turn off charges (5 λ)
4. Turn off vdW (15 λ)





42 MD simulations for each FEP-ABFE calculation

Acceleration of FEP-ABFE calculation using the new protocol on Tianhe HPC

A single 8-cores server
Traditional protocol

Intel Xeon E5
~30 days/compound

60,000 times faster

Tianhe supercomputer
New protocol

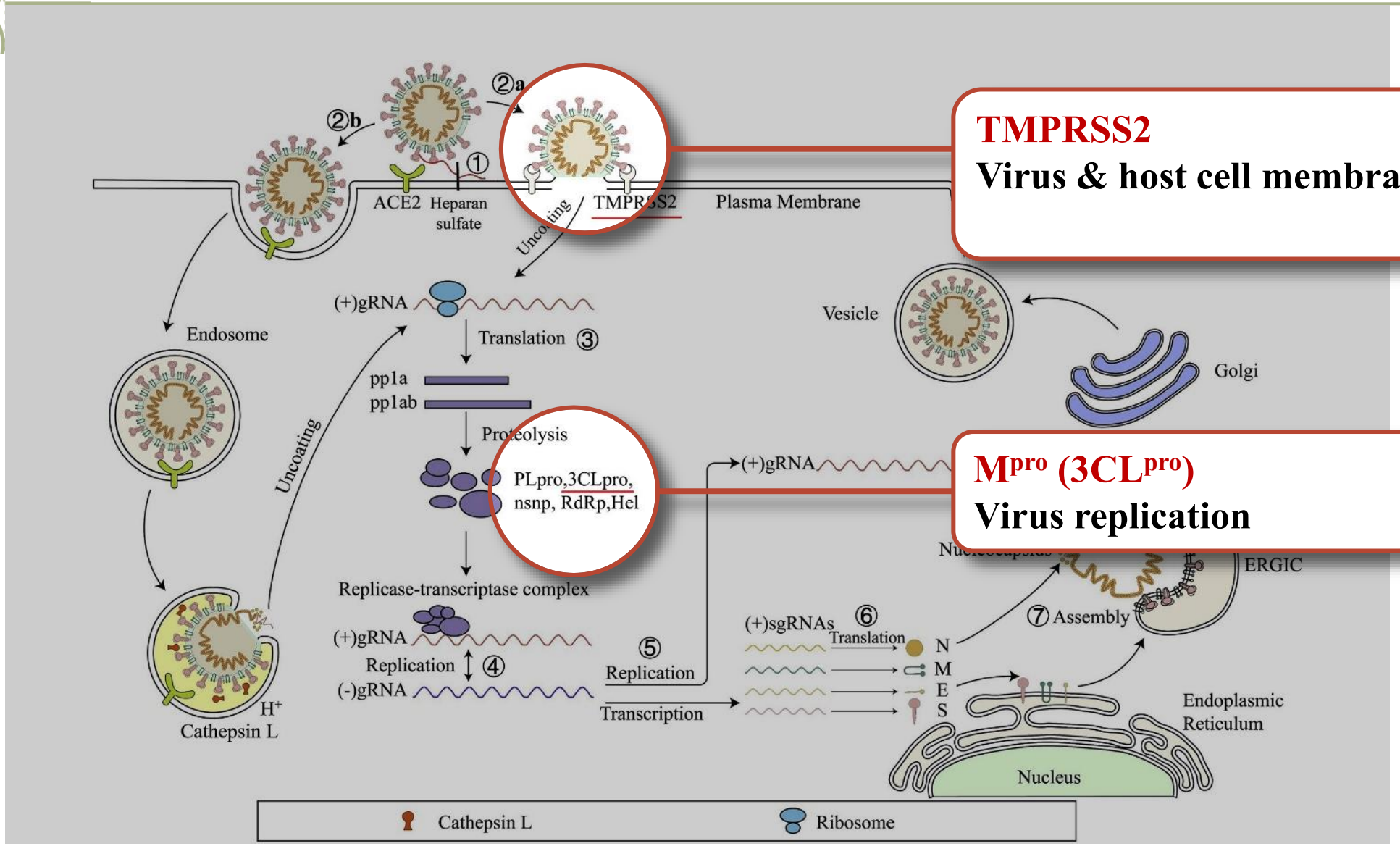
6 days / **12,000** compounds

(or **2,000** compounds per day)

Emergency drug discovery

Contents

Choose key targets for large-scale FEP-ABFE based virtual screening

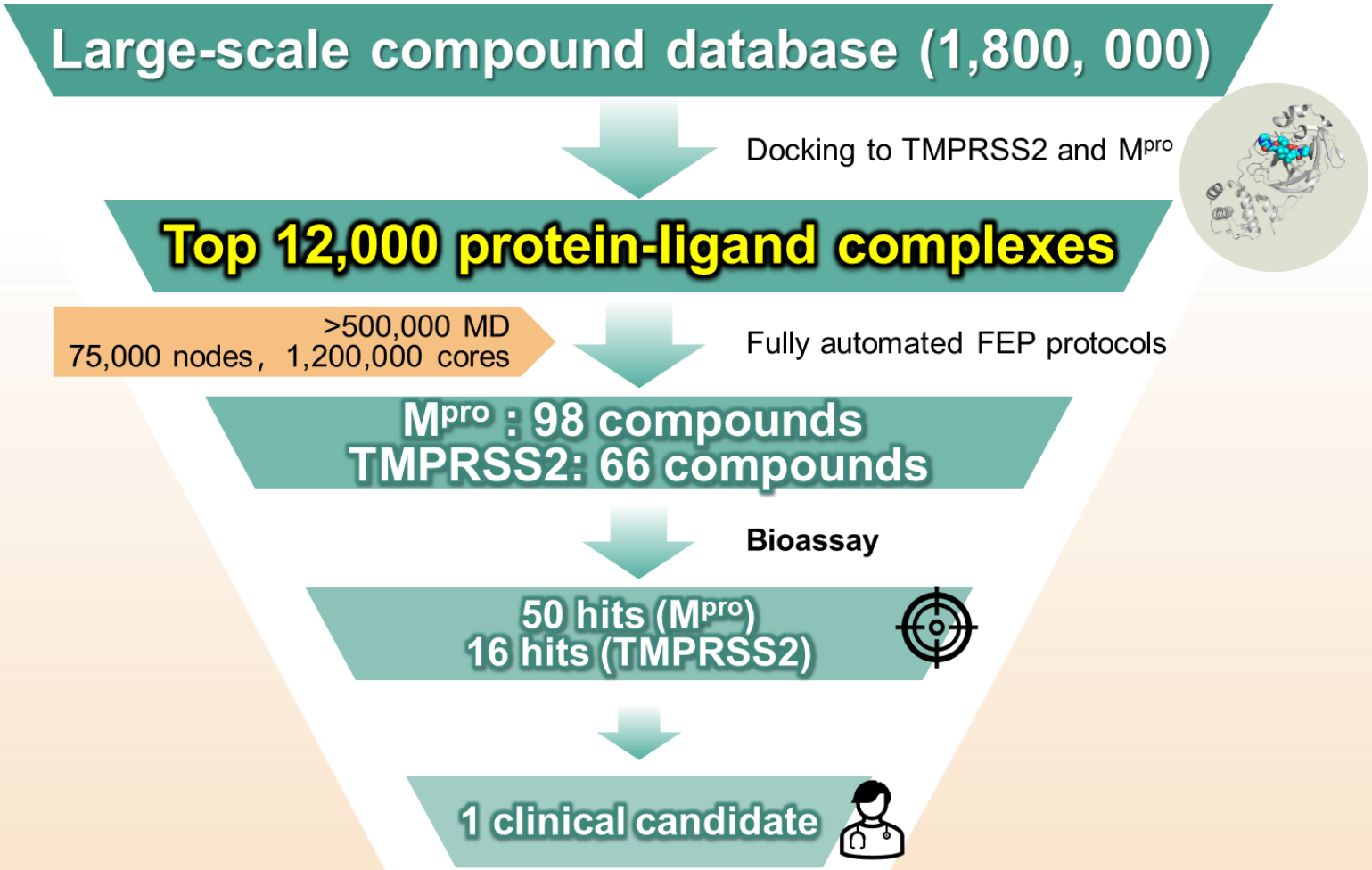


TMPRSS2
Virus & host cell membrane fusion

M^{pro} ($3CL^{pro}$)
Virus replication

Xiang R, et al, Acta Pharm. Sin. B. 2020.

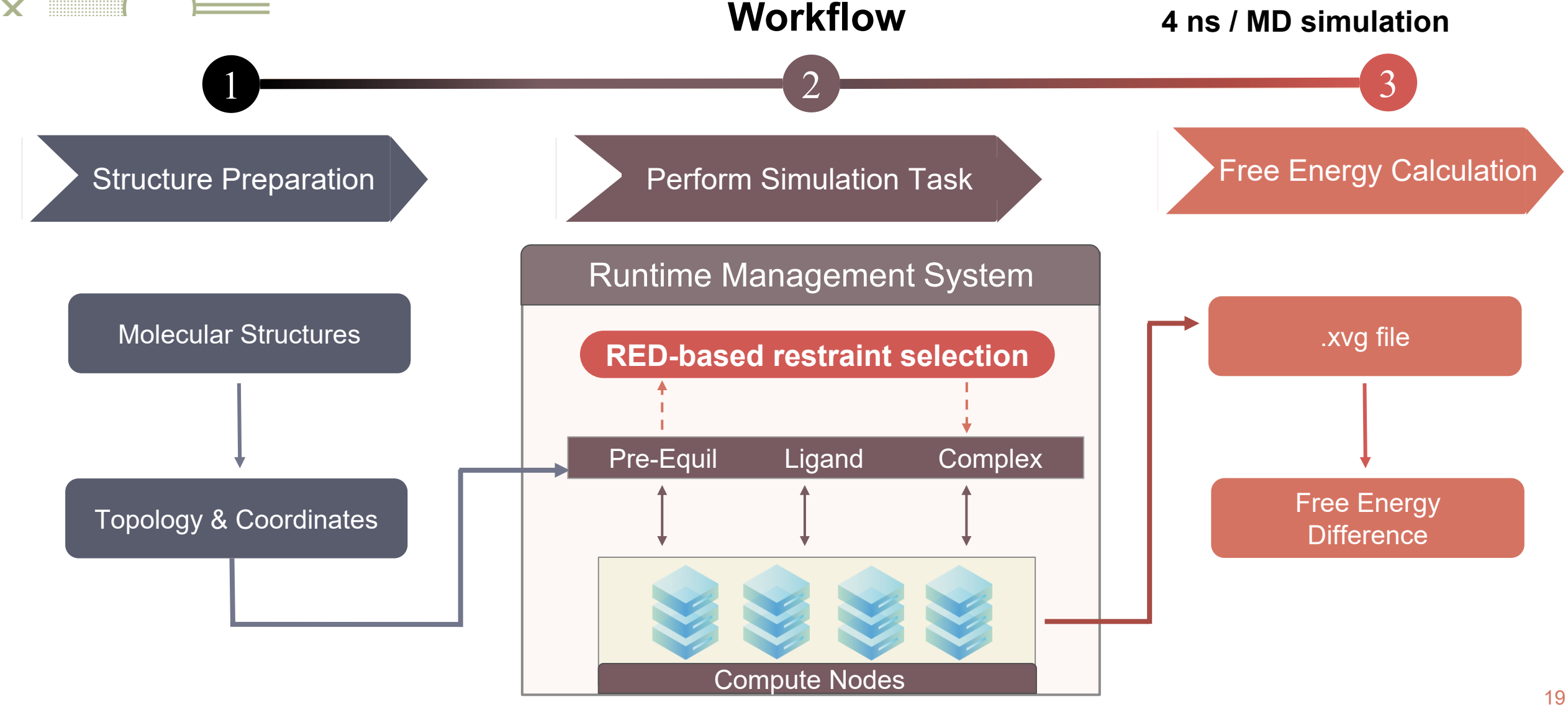
Large-scale virtual screening on Tianhe supercomputer



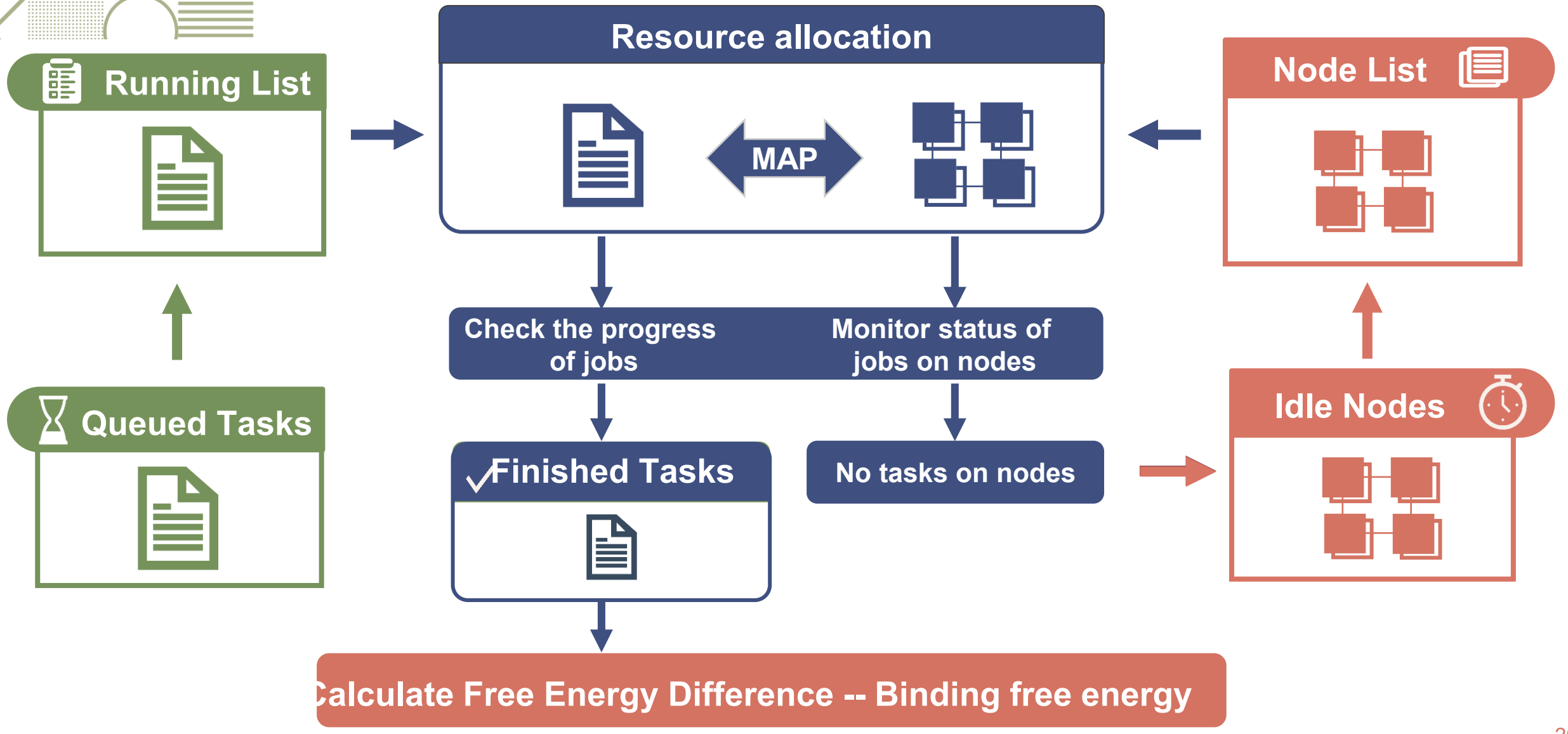
Total number of MD simulation jobs (4 ns / MD simulation)

Target	Ligand DB	Pre-Equilibrate	Ligand	Complex
M^{pro}	FDA	100	2000	2100
	Chemdiv	3143	62860	66003
	SPECS	3027	60540	63567
TMPRSS2	Chemdiv	3004	60060	63084
	SPECS	2825	56500	59325
Total		12099	241960	254079
		508,138		

Intelligent job management system (>500,000 MD simulation tasks)



Intelligent job management system (>500,000 MD simulation tasks)



Computational resource and time used for the large-scale virtual screening

Time for the virtual screening with single precision

Job Type	System used	Time (including IO)
Pre-Equilibrate	12,000 nodes	27.4 h
Ligand	63,000 nodes	23.7 h
Complex	75,000 nodes	114.5 h
Total	**1,200,000 CPU cores 75,000 nodes**	**141.9 h**

Contents

Experimental validation of the computational predictions

Hits against M^{pro}

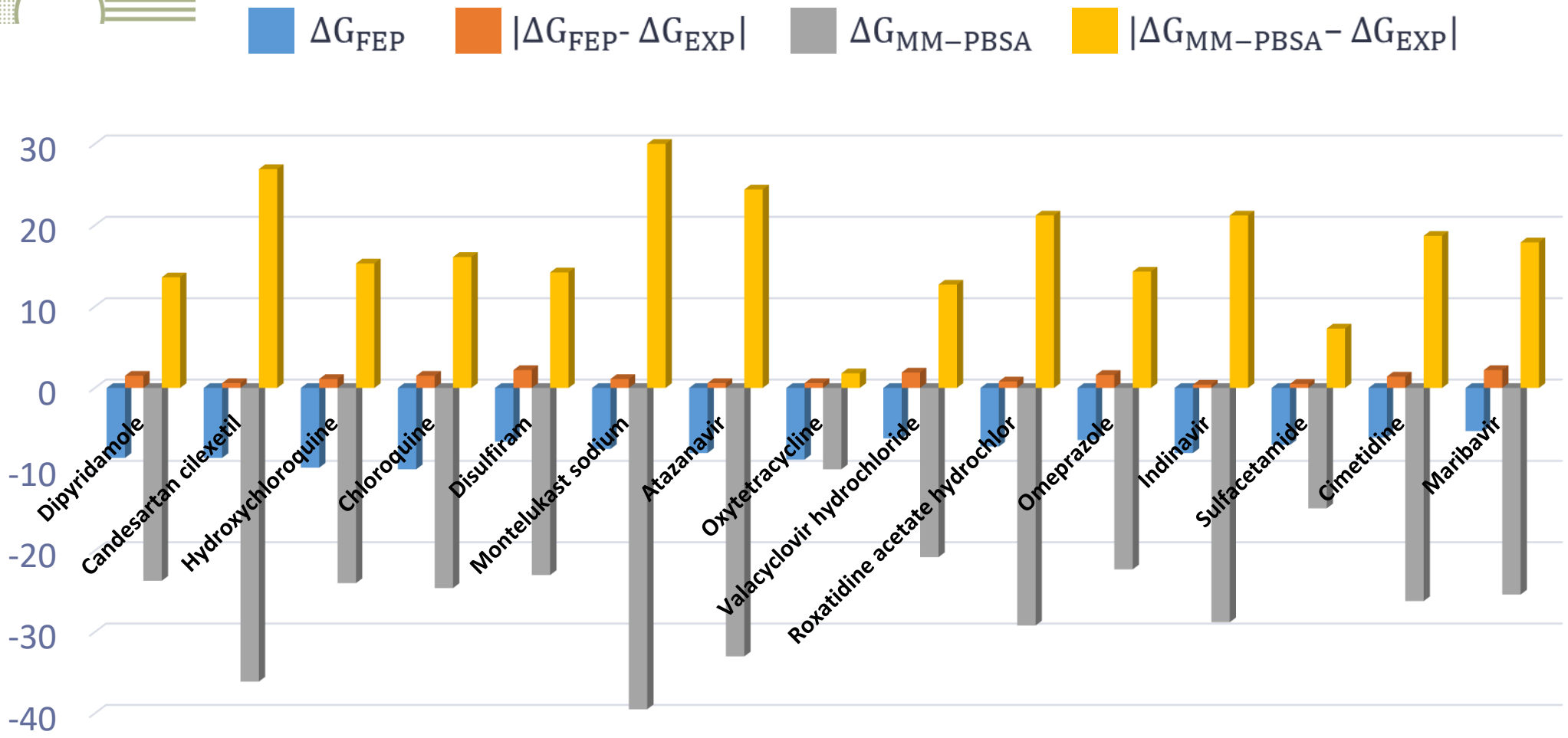
Database	Number of tested compounds	>50% Inhibition at 100 μM	>33% Inhibition at 100 μM
SPECS	38	18	24
ChemDiv	35	16	19
FDA	25	16	20
Total	**98**	**50 (51%)**	**63 (64%)**

Experimental validation of the computational predictions

Hits against TMPRSS2

Database	Number of tested compounds	>50% Inhibition at 100 µM	>33% Inhibition at 100 µM
SPECS	35	9	24
ChemDiv	31	7	20
Total	**66**	**16 (24%)**	**44 (67%)**

Superior performance of FEP-ABFE predictions compared with MM-PBSA

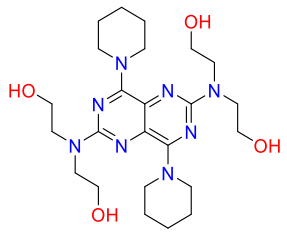


Li Z, et al. Proc. Natl. Acad. Sci. USA. 2020

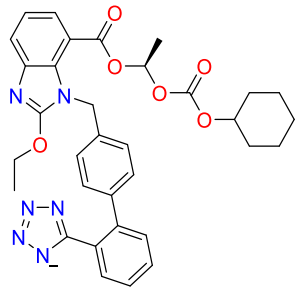
Representative Hits

Further consideration in repurposing a drug for treatment of COVID-19 patients: Known functions of the drug

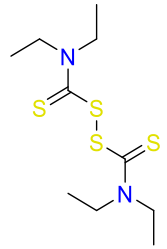
1) Active M^{pro} inhibitors from known FDA-approved drugs



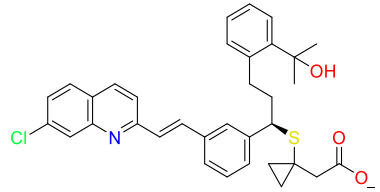
Dipyridamole



Candesartan Cilextil

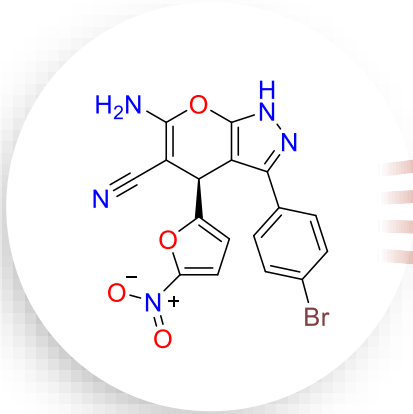


Disulfiram



Montelukast Sodium

2) Active M^{pro} inhibitors from commercial compound libraries



Analogue 1

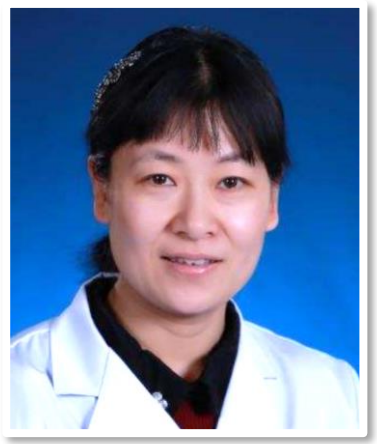
Analogue 2

Analogue 3

……

Structural optimizations are being performed…

Inspiration: Identify a drug with both anti-viral and anti-thrombosis activities



Prof. Fuling Zhou

Head of hematology

Zhongnan Hospital of Wuhan University

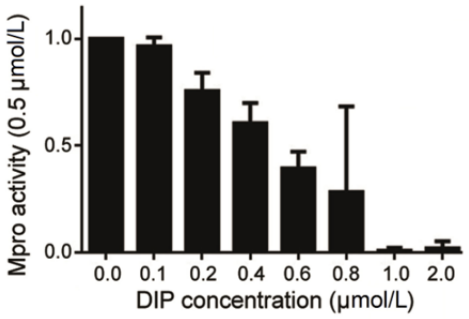
Clinical variables in 124 patients with COVID-19

Variable	Range for normal subjects	Range for COVID-19 patients (Total number = 124)
PLT (109/L)	125-350	191.7 80.0 (54-525)
Lymphocyte (109/L)	1.1-3.2	0.9 0.6 (0.1-5.0)
MPV (fL)	6-12	9.1 1.3 (6.6-12.3)
PT (S)	9.4-12.5	13.0 1.4 (8.6-17.8)
APTT (S)	25.1-36.5	30.3 3.2 (22.4-38.1)
FIB (mg/dL)	238-498	429.8 88.7 (203-750)
D-dimer (μg/L)	**0-500**	**1168.6 3652.7 (35-26315)**

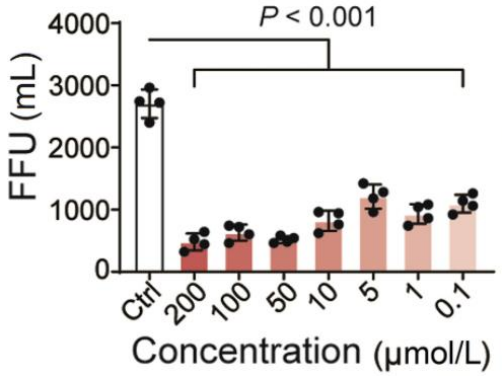
Liu XY, et al, Acta Pharm. Sin. B. 2020.

Hypercoagulability was associated with COVID-19 disease severity.

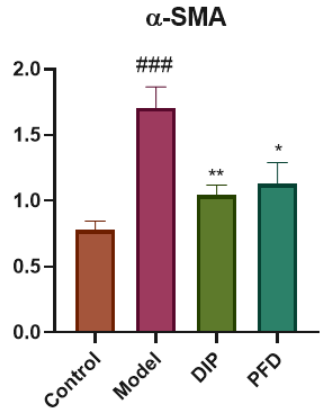
Identified clinical candidate against COVID-19: Persantine (Dipyridamole)



A) M^{pro} inhibition



B) Anti-viral replication



C) Anti-pulmonary fibrosis

D) Anti-thrombosis

New discovery in this work

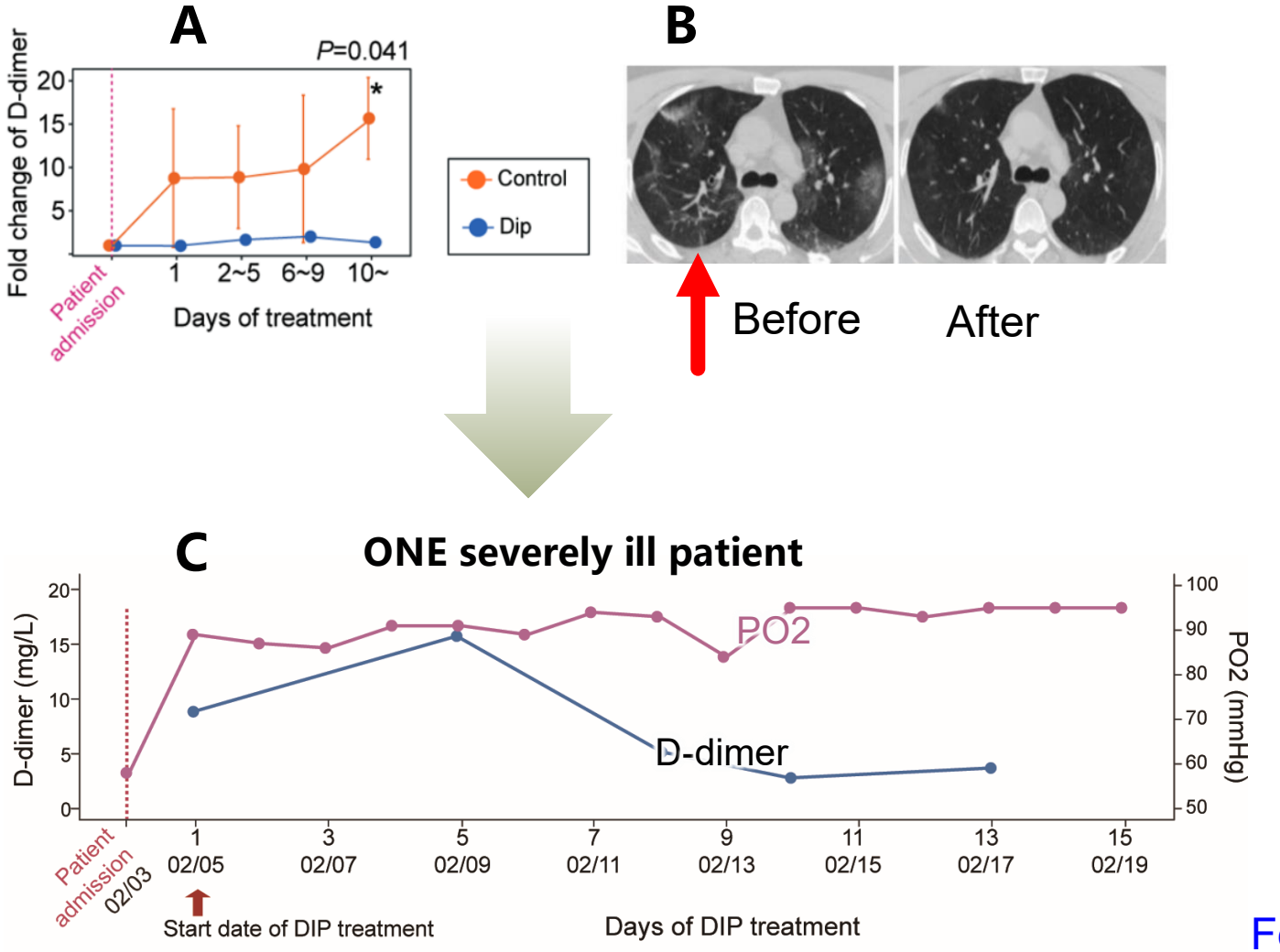
Known function

In vitro and in vivo validation: Emergency drug discovery

Liu XY, et al, Acta Pharm. Sin. B. 2020.

Contents

Dipyridamole adjunctive therapy improved the coagulation profiles and shortened the time for discharging the patients

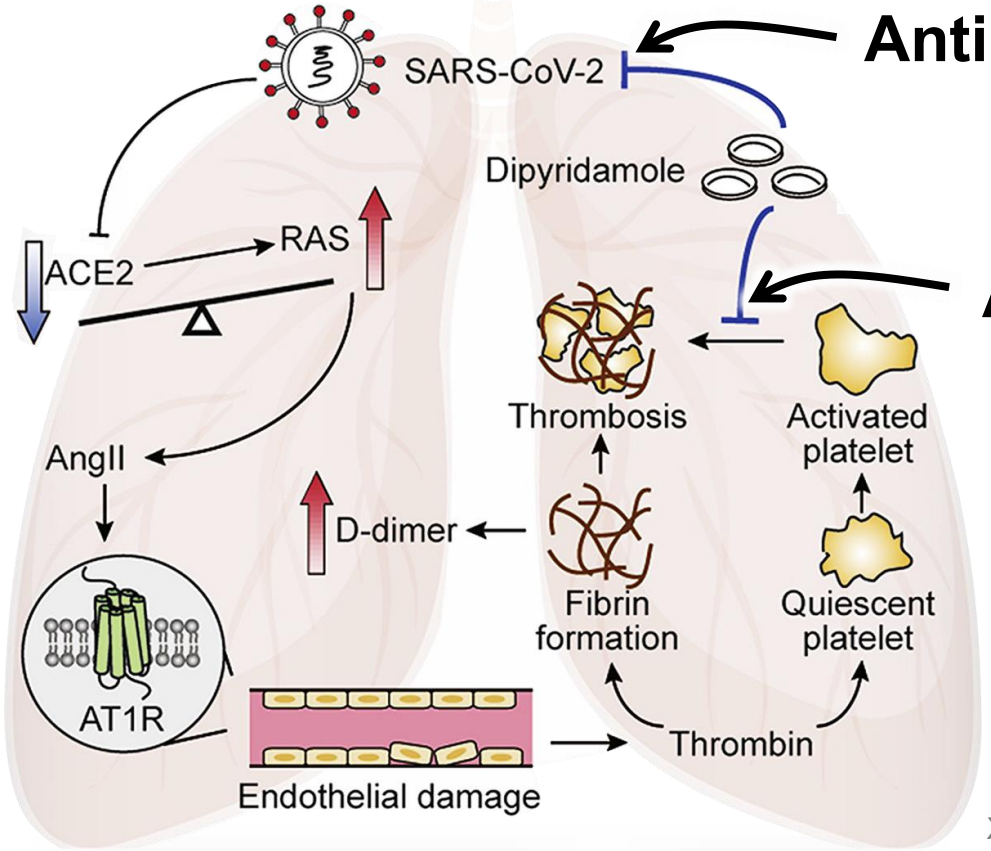


Discharge time for the patients

Group	Patients	Median time from treatment to discharge (days)
Control	86	11
DIP	37	7

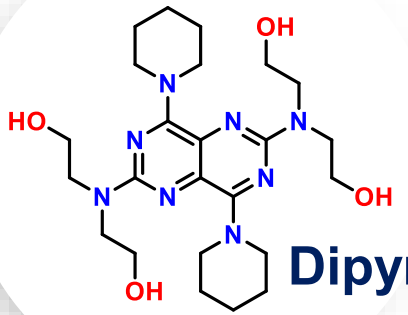
Feb.-Apr., 2020 Jiang M, et al, J. Cell. Mol. Med. 2021.

The mechanism of dipyridamole: Anti-viral and anti-thrombosis



Xiaoyan Liu, et al, 2020.

Clinical Studies



Dipyridamole

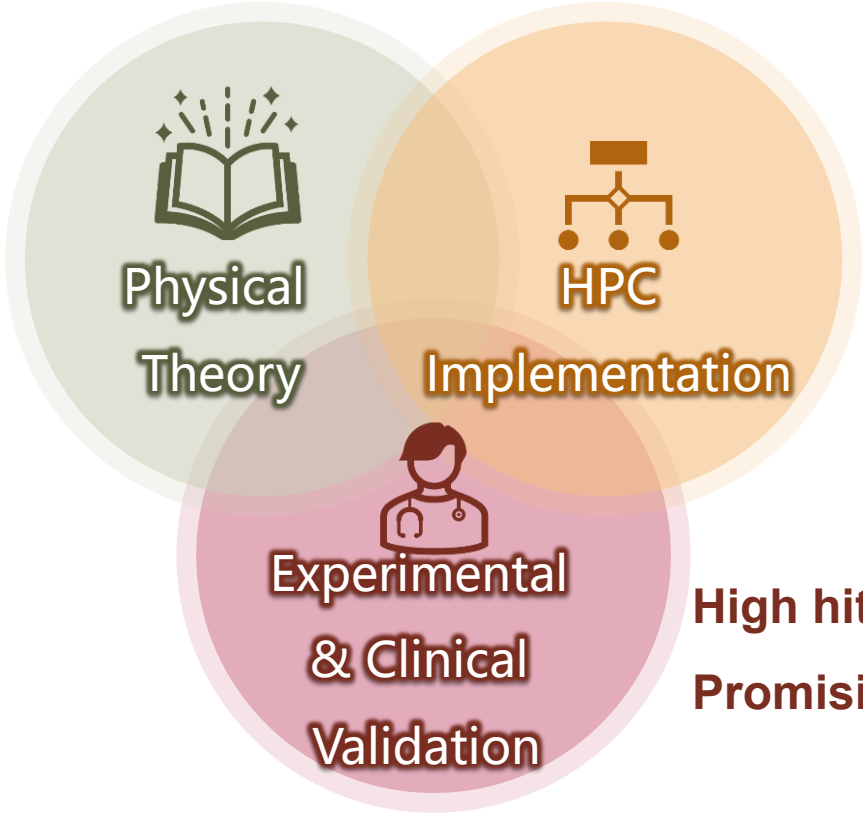
Ongoing clinical trials of dipyridamole against COVID-19 (by other independent groups)

Trial and Title	**NCT04424901:** Open Label Dipyridamole- In Hospitalized Patients With COVID-19	**NCT04410328**: Aggrenox To Treat Acute COVID-19	**NCT04391179**: Dipyridamole to Prevent Coronavirus Exacerbation of Respiratory Status (DICER) in COVID-19
Trial Type	**Randomized Phase II Open Label clinical trial**	**Randomized Phase III clinical trial**	**Randomized Phase II clinical trial**
Status	Recruiting	Recruiting	Completed; results not yet disclosed
Conditions	➢ COVID-19 Pneumonia ➢ Vascular Complications	➢ COVID-19	➢ COVID ➢ Corona Virus Infection ➢ COVID-19 ➢ SARS-CoV-2 Infection
Interventions	➢ Drug: Dipyridamole (Standard Care *vs* Standard Care with Dipyridamole)	➢ Drug: Dipyridamole ER 200mg/ Aspirin 25mg orally/enterally and Standard of care ➢ Other: Standard of care	➢ Drug: Dipyridamole 100 Milligram(mg) ➢ Drug: Placebo oral tablet
Locations	UConn Health, Farmington, Connecticut, United States	Rutgers New Jersey Medical School University Hospital, Newark, New Jersey, United States	University of Michigan, Ann Arbor, Michigan, United States

Contents

Summary of the major innovation and an unprecedentedly high hit rate

- A new RED function for accelerated ABFE prediction
- A novel algorithm allowing FEP-ABFE-based virtual screening
- Automatic



A large-scale virtual screening runtime management system

High hit-rate (51%)

Promising clinical outcomes

- Automated high-throughput FEP-ABFE calculation protocol
- **Milestone**: The first time FEP-ABFE was used in large-scale virtual screening
- Efficiency of FEP was greatly increased – applied in emergency drug discovery

SC21

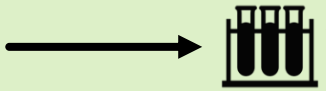
Outlook

A general approach for rapid drug discovery using a supercomputer like Tianhe to make us ready against next breakout or other diseases

FEP algorithm optimization

More reliable and more efficient

Ultra-large scale FEP-ABFE-based virtual screening



Wet lab validation

(FEP-ABFE predictions for millions of compounds)

Accelerate

HPC on a larger scale

Acknowledgement

Hospital:

- Zhongnan Hospital of Wuhan University

HPC support:

- National Supercomputing centers in Tianjing, Shenzhen, and Guangzhou;
- Tencent Cloud

Funding support:

- National Key R&D Program of China
- National Natural Science Foundation of China
- Fundamental Research Funds for Hainan University
- Science Foundation of Guangzhou City
- Guangdong Province Higher Vocational Colleges & Schools Pearl River Scholar Funded Scheme
- The National Science Foundation
- The Taishan Scholars Program
- The Innovative Leader of Qingdao Program
- The special scientific research fund for COVID-19 from the Pilot National Laboratory for Marine Science and Technology
- Open fund from the State Key Laboratory of High Performance Computing

Thank you

Backup Slides

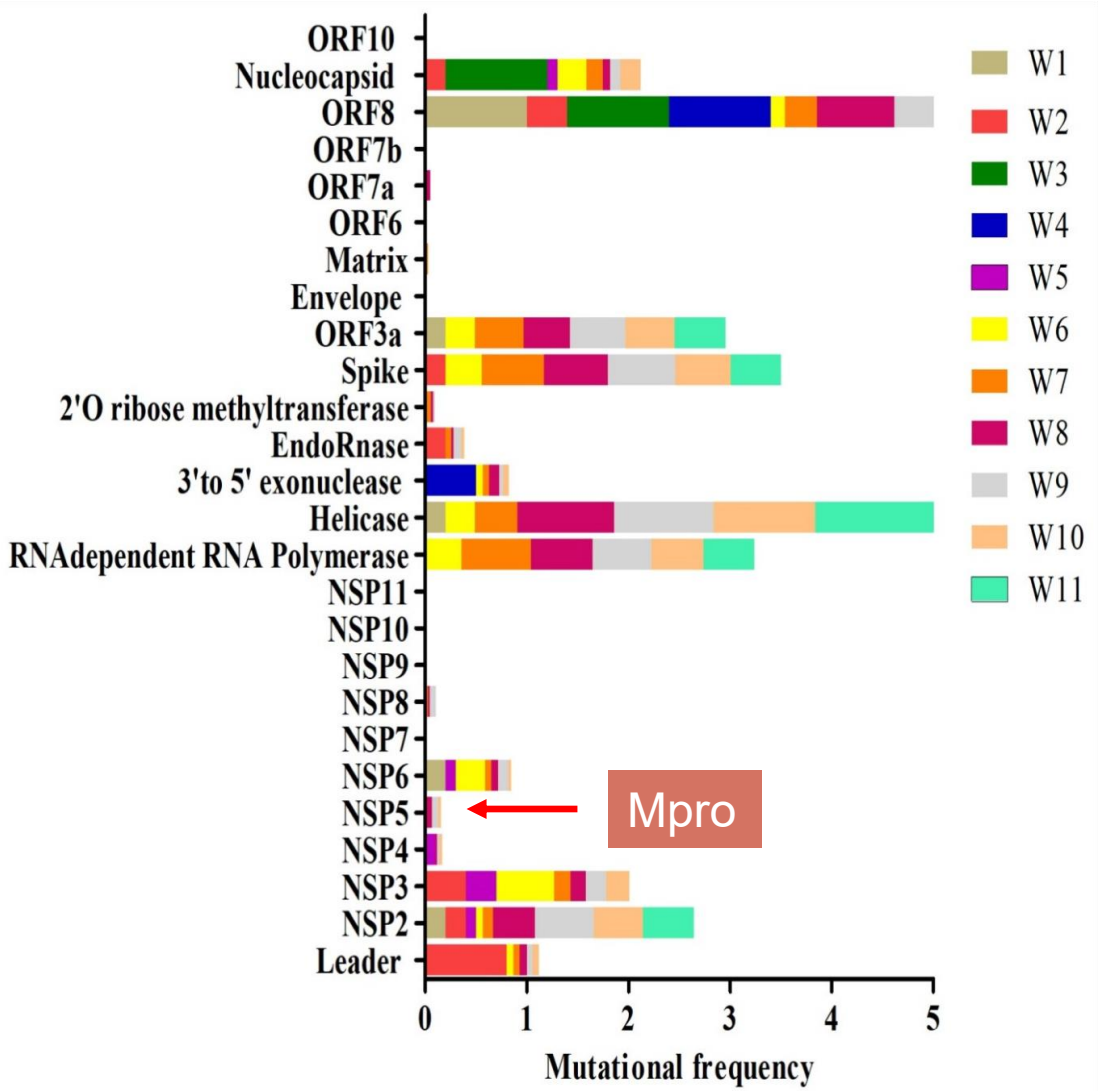
QA of supercomputing

- What system did you use?
  We used the new generation of Tianhe supercomputer.
- Where is the system located?
  In Tianjin.
- What is the scale of the new Tianhe system?
  We used a portion of the new Tianhe system. To be precise, we used over 75,000 nodes.
- What is the architecture of the system?
  We only used the proprietary CPU cores.
- How many cores in the CPU?
  We only used 16 cores per node.
- What is the peak performance of the new Tianhe system?
  Our team focused on finishing the large-scale virtual screening with the system within an acceptable amount of time rather than quantified performance in terms of FLOPS.
- Is there any heterogeneous accelerator in the system?
  We only used CPU cores.
- Why did not screen FDA database against TMPRSS2?
  It has already been done by others.

Backup Slides for page 2 To Dr. Zhan

- COVID 19, SARS, and MERS infection and death data from WHO.
- MERS: between 2012 to 2021.
- Zika: American only, 2015 to 2021, from PAOH.

Mutational frequency of SARS-CoV-2 genes



DOI: 10.3390/pathogens9070565