IYBSSD2022
International Year of Basis Sciences for Sustainable Development

Moscow, Russia
24-26 May, 2022 | VIRTUAL

IBMC
Institute of Biomedical Chemistry

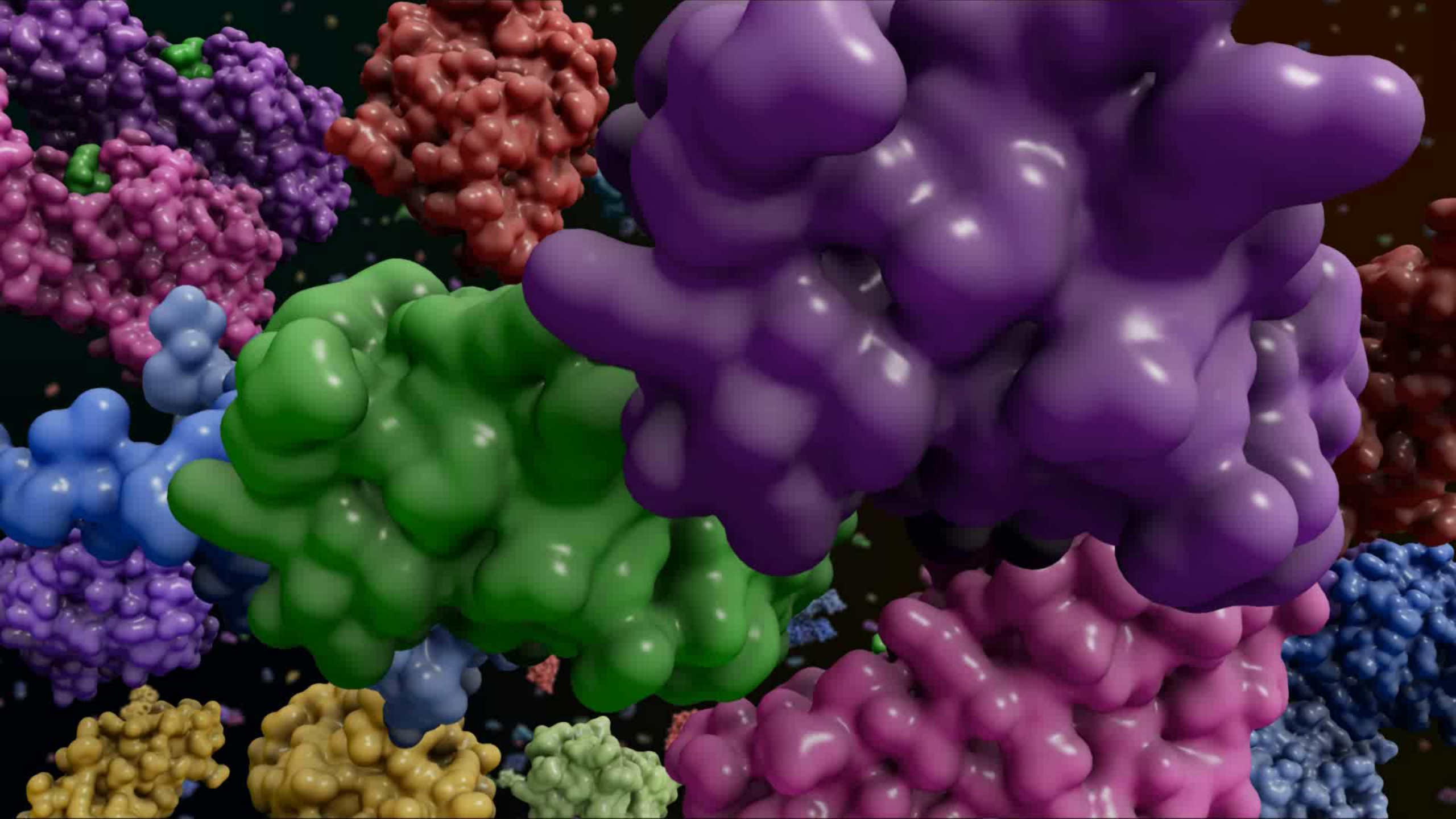
XXVIII Symposium on Bioinformatics and Computer-Aided Drug Discovery

12:00PM 25 May, 2022
6:00PM, SEOUL

Drug Discovery with Fragment Molecular Orbital (FMO) Method

Bioinformatics & Molecular Design Research Center (BMDRC)
Yonsei University, Baobab AiBIO Inc.

Hocheol Lim, Hyeon-Nae Jeon, Jong Wan Kim, Kyoung Tai No



Frontiers in Structure Based Drug Discovery

CADD From Basic Science to Commercial Design Packages

Physics, Phys. Chem., Chemistry, Biology, Computer Science, ..

- Stable Structure
- System Simulation
- Docking Simulation
- Free Energy Perturbation
- ADMET Prediction
- ……

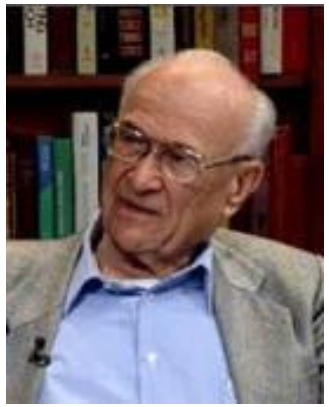
- Thermodynamics
- Statistical Mechanics
- Empirical Force Fields
- Quantum Mechanics
- Dynamic Simulations
- Solvation Free Energy
- Information Science
- Artificial Intelligence

Commercial/ inhouse S/W, DB, …

Drug Design with packaged S/W

Empirical Potential Energy Function (Force Fields)

Energy – Mechanics Based Design



Harold Scheraga

ECEPP/2,3

FF for Protein in Torsion space

Martin Karplus

CHARMMnn

All intra degree of freedom

Noble Prize 2013

Peter Kollman

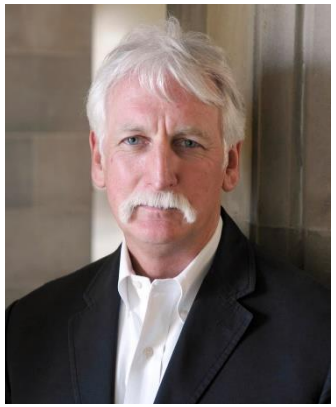
AMBER

Most widely used → open source

Norman Allinger

MM/2&3

Intramolecular FF For smaller molecular structure



William Jorgensen

OPLS,..., OPLS3e..

Good for liquid Sim. Fit to Free Energy Perturbation

BMDRC

PMFF

Since 1987~2020

Physics based FF parameters determination

First in ASIA

PMFF: Development of a Physics-Based Molecular Force Field for Protein Simulation and Ligand Docking

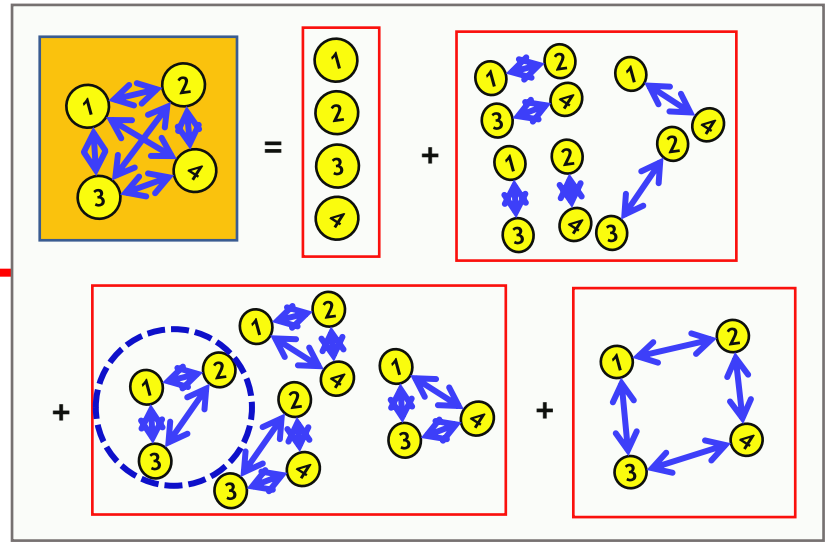
Sung Bo Hwang, Chang Joon Lee, Sehan Lee, Songling Ma, Young-Mook Kang, Kwang Hwi Cho, Su-Yeon Kim, Oh Young Kwon, Chang No Yoon, Young Kee Kang, Jeong Hyeok Yoon, Ky-Youb Nam, Seong-Gon Kim, Youngyong In, Han Ha Chai, William E. Acree, Jr., J. Andrew Grant, Ken D. Gibson, Mu Shik Jhon, Harold A. Scheraga, and Kyoung Tai No*

SB Hwang, et al., J. Phys. Chem. B (2020)

Limitation of Empirical FF in Accuracy of Energy Calculation

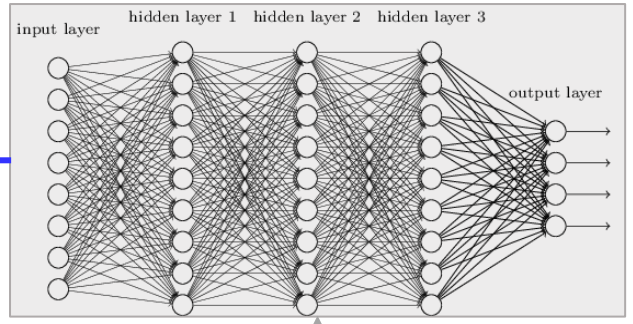
Accuracy in Describing ΔE. ΔΔE

Empirical Force Fields



Where it Comes?

Neural Network Potential (NNP)



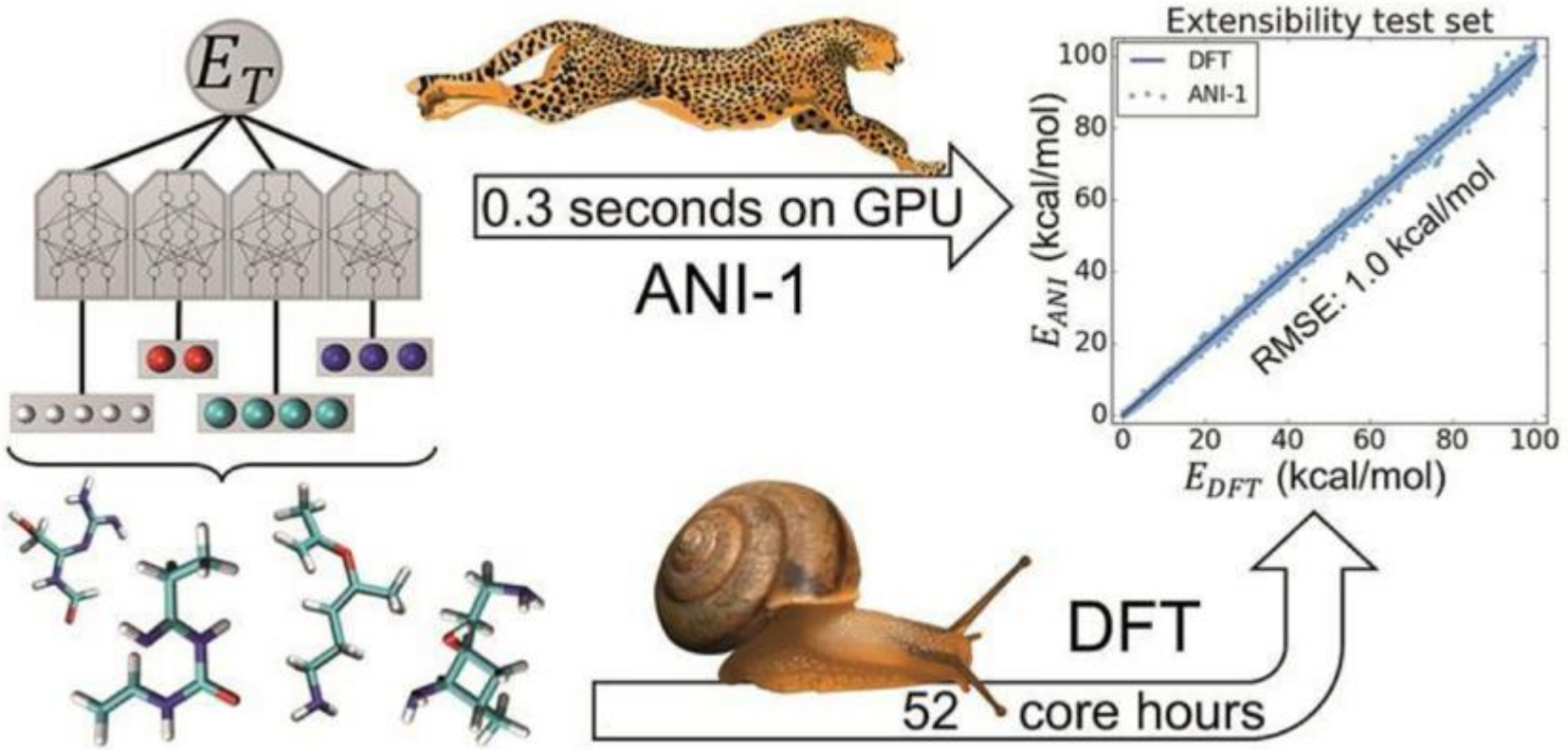
Real, Nature State

$$\nabla^2\Psi + \frac{8\pi^2 m}{h^2}(E - V)\Psi = 0$$

Generate Big Data (PESs)

Neural Potential vs. QM-DFT Calculation

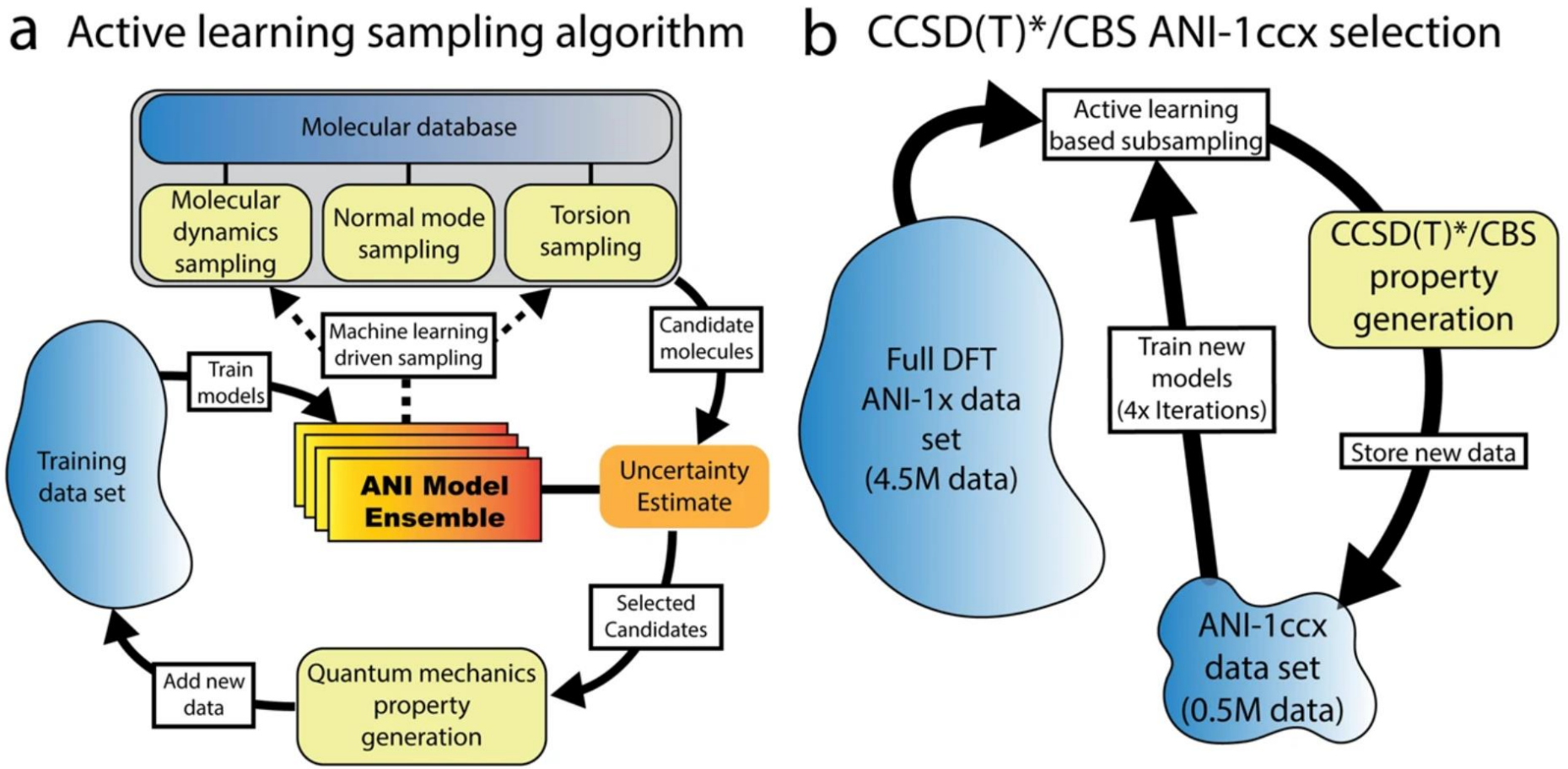
Adrian E. Roitberg



Deep neural network (NN) trained on quantum mechanical (QM) DFT calculations can learn an accurate and transferable potential for organic molecules.

ANAKIN-ME: Accurate NeurAl networK engINe for Molecular Energies

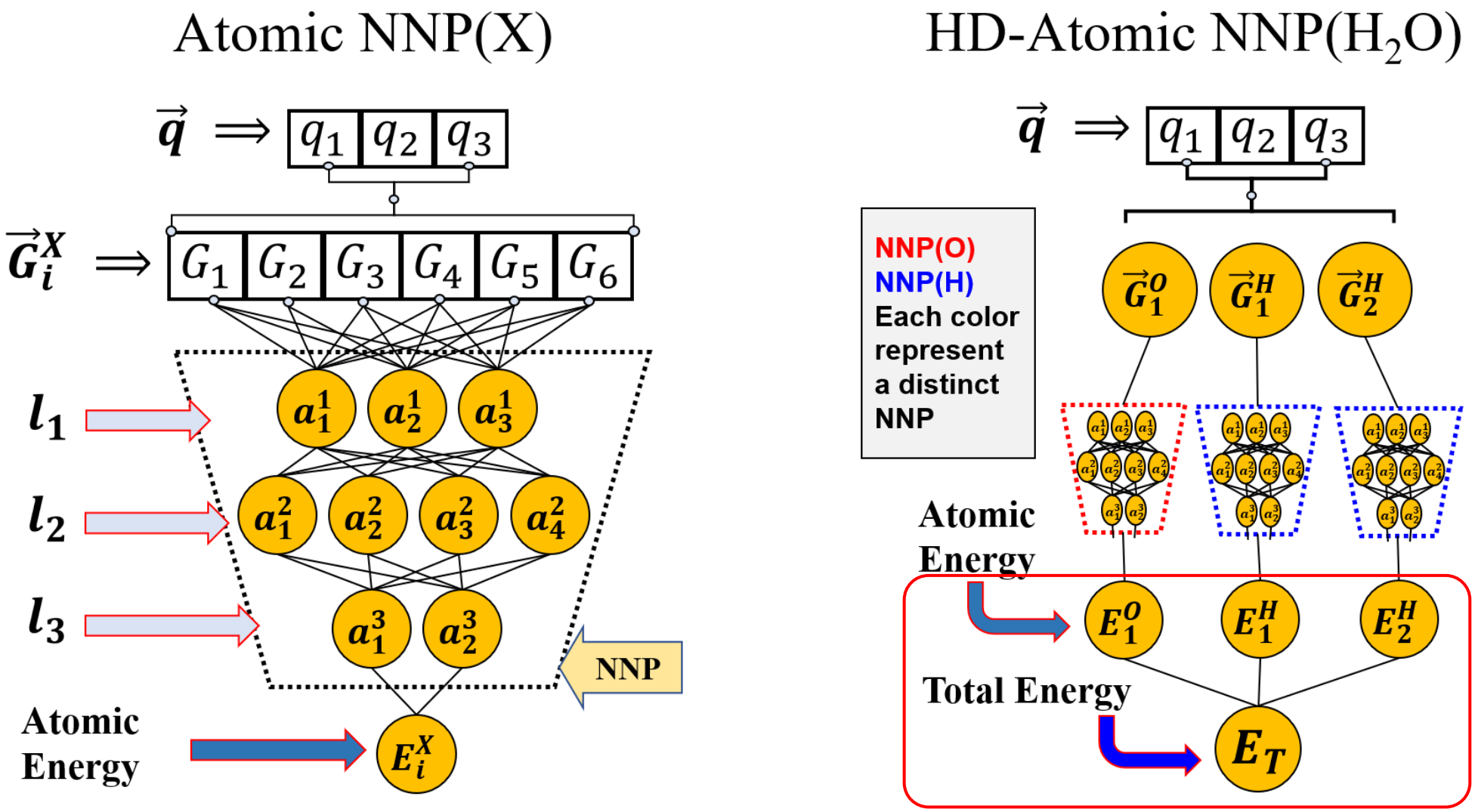
Training of ANI-n Neural Network Potential



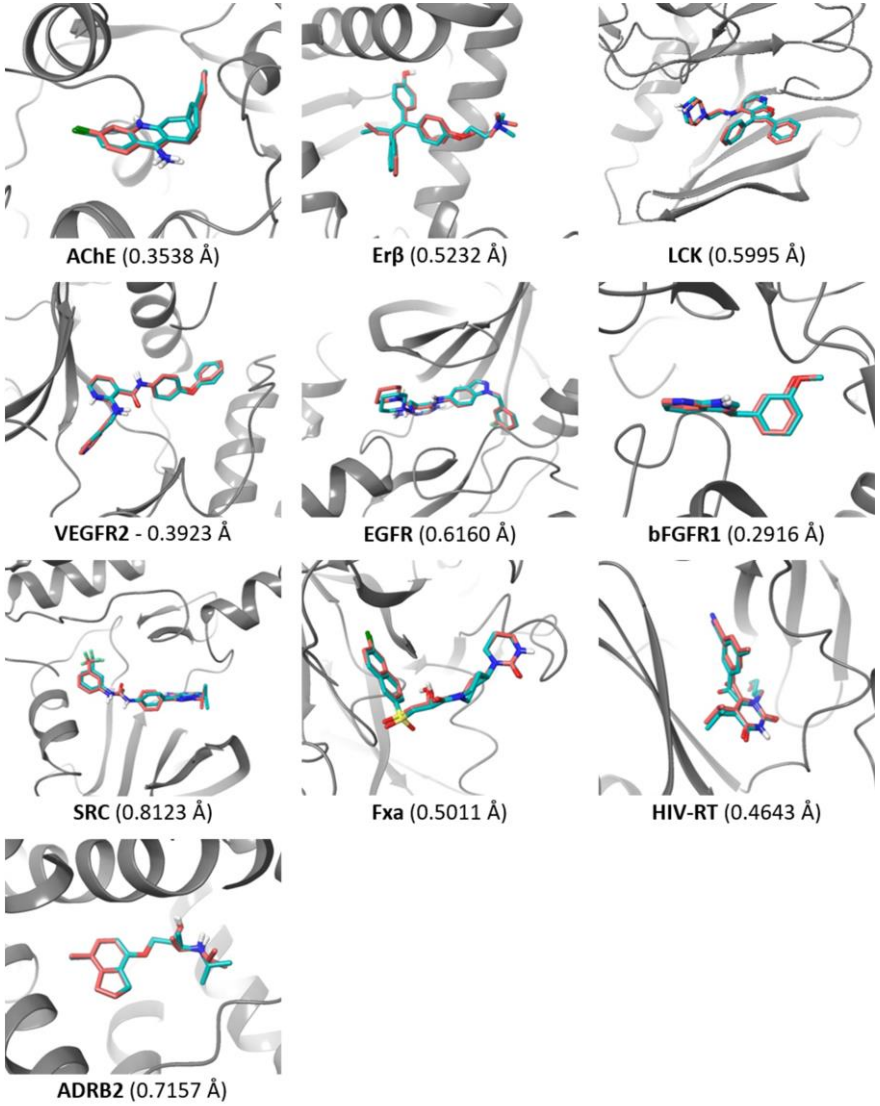
- **ANI-1**: ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost
- **ANI-1x**: Less is more: Sampling chemical space with active learning
- **ANI-1ccx**: Approaching coupled cluster accuracy with a general-purpose NNP through transfer learning
- **ANI-2x**: Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens

Atomic NNP model & It's Application to Molecule (H_2O)

Behler and Parrinello's HDNN or HD-atomic NNP model.



Protein-Ligand Docking Study with ANI-2x

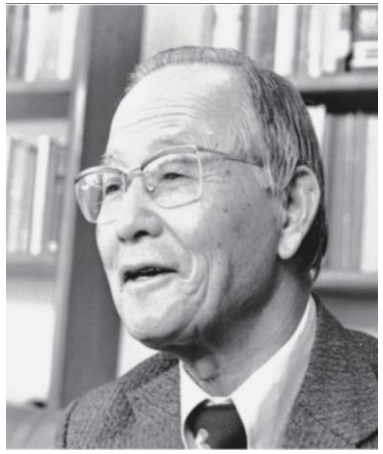


Top ranking of docking poses in 10 X-ray crystal structures					
PDB ID	Generated Poses	Calc. Time (s)	Total Atoms	RMSD (Å) of ANI-2x Top 1	RMSD (Å) of Emodel Top 1
1E66	44	10.23	8247	0.3538	0.3538
2FSZ	512	66.27	3900	0.5232	0.9998
2OF2	133	18.41	4407	0.5995	0.5995
2P2I	333	51.52	4867	0.3923	0.7912
2RGP	955	155.66	5141	0.616	1.4343
3C4F	88	13.35	4871	0.2916	0.2916
3EL8	126	19.25	4462	0.8123	0.8679
3KL6	410	53.65	3688	0.5011	1.2538
3LAN	53	12.74	9120	0.4643	0.4643
3NY8	101	20.68	7180	0.7157	2.511

Neural network potentials have provided accurate results for intra- and intermolecular interactions in protein-ligand complexes. Although ANI-2x was not trained for protein structures and ionic molecules, scoring the docking poses with ANI-2x was reasonable and we showed that ANI-2x can be applied to molecular docking simulations. These methods can be incorporated directly into existing docking scoring methods to select the most favorable binding pose of a ligand. These few applications of NNP would be the start point of how machine learning will create new trends in biosciences.

Manuscript in Preparation

Fragment Molecular Orbital (FMO) for Protein Energy Calculation



Prof. Taikyue REE

Prof. Kenichi FUKUI

Prof. Keiji MOROKUMA

Prof. Kazuo KITAURA

How to calculate the Energy (Electron density) of Proteins with Quantum Chemical Calculation

Fragment Molecular Orbital (FMO)

Institute of Molecular Science, Okazaki

For FMO → Energy Decomposition & Structure Fragmentation

Energy Decomposition (Partitioning): the total dissociation energy D_e is decomposed into a number of physically meaningful components by dividing the interaction process between two or more fragments *A* and *B*.

Kitaura–Morokuma or Ziegler–Rauk schemes

$$-D_e = \Delta E_{elec} + \Delta E_{ex} + \Delta E_{pol} + \Delta E_{CT} + \Delta E_{mix}$$

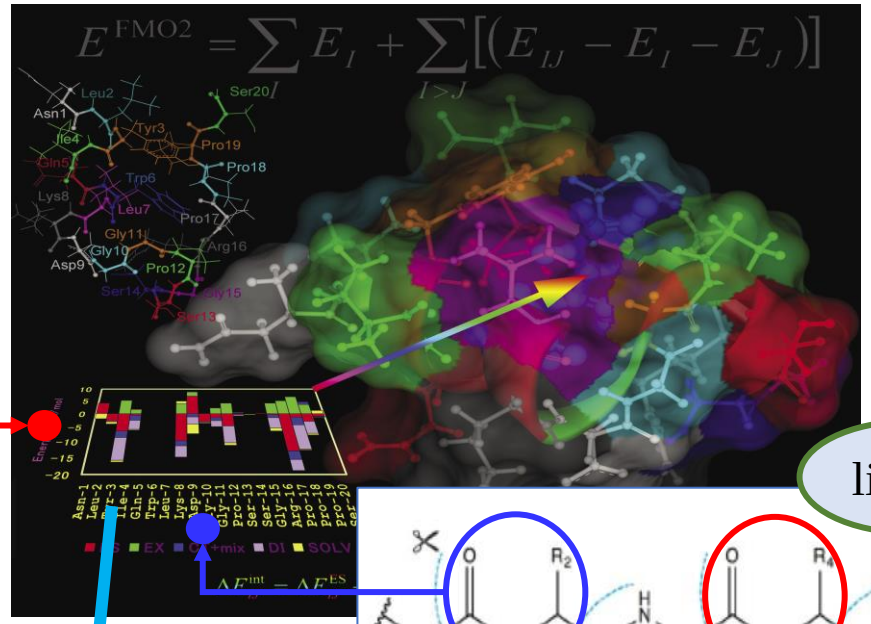
Structure Fragmentation: The bond energetics within a molecule is to *decompose* the total energy of the molecule within a given quantum-chemical method into a sum of monoatomic and diatomic contributions (**fragments**) as follows,

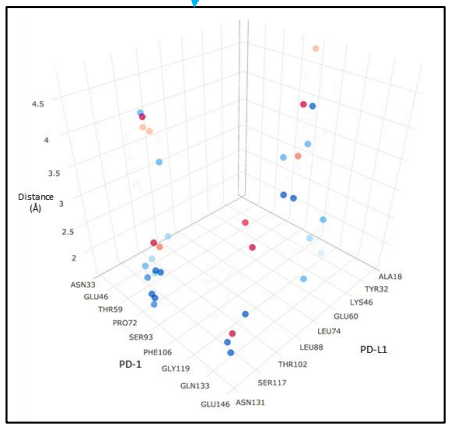
$$E = \sum_{A} E(A) + \sum_{A>B} E(AB) = \sum_{A} E(A) + \frac{1}{2}\sum_{A\neq B} E(AB)$$

Fragment MO (FMO)

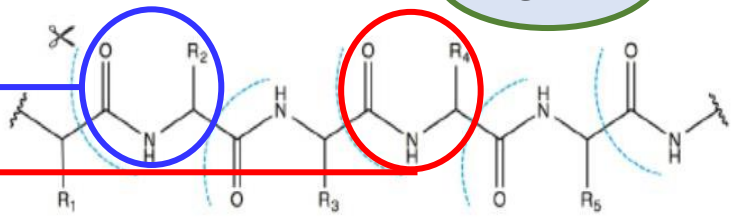
A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation, Kazuo Kitaura & Keiji Morokuma, Int. J. Quantum Chem., 1976, 10, 325~340
On the calculation of bonding energies by the Hartree Fock Slater method, Tom Ziegler & Arvi Rauk, Theoretica Chimica Acta, 1977, 46, 1–10

Structure Fragmentation of Protein





ligand



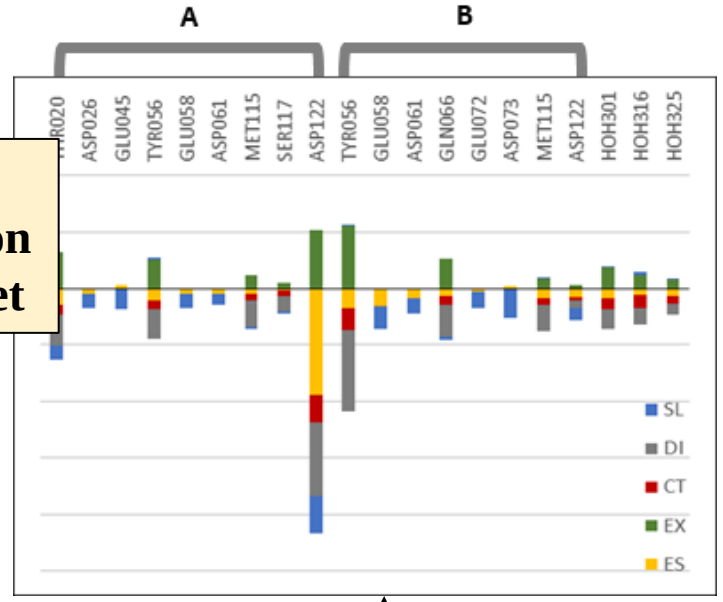
Fragmentation of peptide chain

Fragmentation

$$\Delta E^{int}_{ij} = \Delta E^{es}_{ij} + \Delta E^{ex}_{ij} + \Delta E^{ct}_{ij} + \Delta E^{di}_{ij}$$

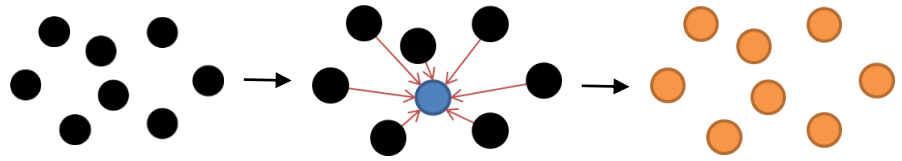
Pair Interaction Energy (PIE)

Pair Interaction E
Energy decomposition
Define binding pocket



ψ_f

Initial electron density → **Electrostatic potential** → **New electron density**



ψ_0 $\quad \hat{H}_0 + \triangle \hat{H}$ $\quad \psi_1$

FMO based Drug Design Platform, AVENGERS

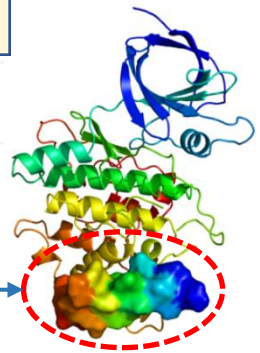
STAGE I

STAGE II

STAGE III

Target Protein Structure Analysis

- Protein-protein complex str.
- Inhibitor complex



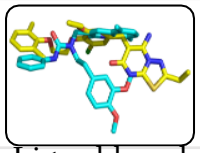
I Binding Pocket Analysis

- Pocket analysis
- FMO Calculation
- Selection of "Hot Spots"

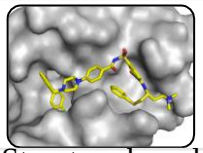
II Customized Library

- **NP3000**, Natural Compd. Lib.
- **10^7** Synth. Compd. Virtual Lib
- Selection of "Hot Spots"

III Virtual Screening

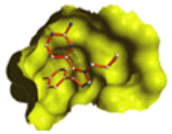


Ligand-based Screening



Structure based Screening

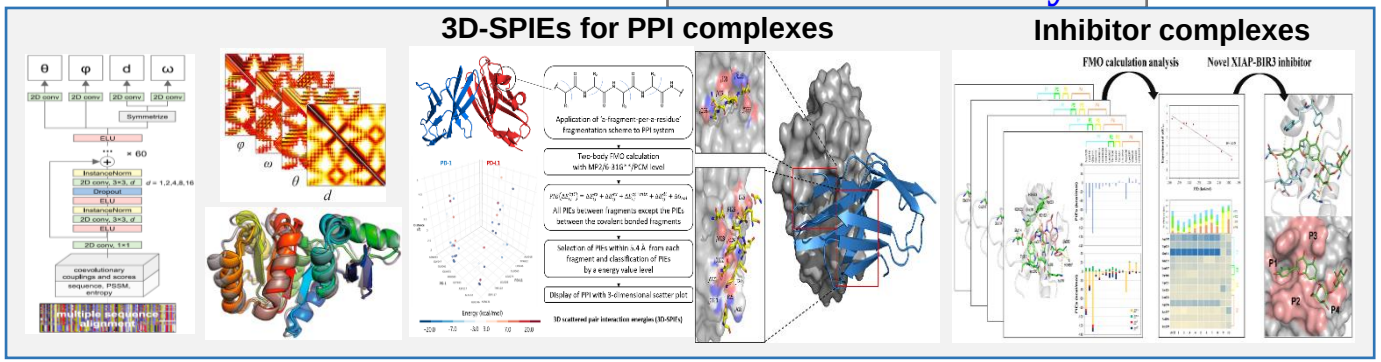
IV Hit Compounds



Screening with *in silico* Toxicity, PK & physical property (QSARs)

VI Hit to Lead Optimization

FMO Based Analysis



A Feature-specific filtering
- Pharmacophore-based screening
- Shape-based ligand clustering

B Initial binding pose generation
- **ANI-1**-based conformational search (NNPot)
- Docking-based binding pose generation

C Initial binding pose filtering
- MM-based pose ranking
- **FMO-DFTB3-based pose ranking**

D Semi-empirical-based pose optimization
- FMO-DFTB3-based pose optimization

E Receptor-specific scoring function
- FMO-PIEDA based enthalpy term
- PCM (MM/GBSA) based solvation term
- Non-polar-SA (N_{rot}) based entropy term

In vitro study – Efficacy
- **Luciferase activity**

***In vitro* study - Binding**
- **Cryo-EM**
- **SPR assay**
- **Thermal shift assay**
- **FRET assay**

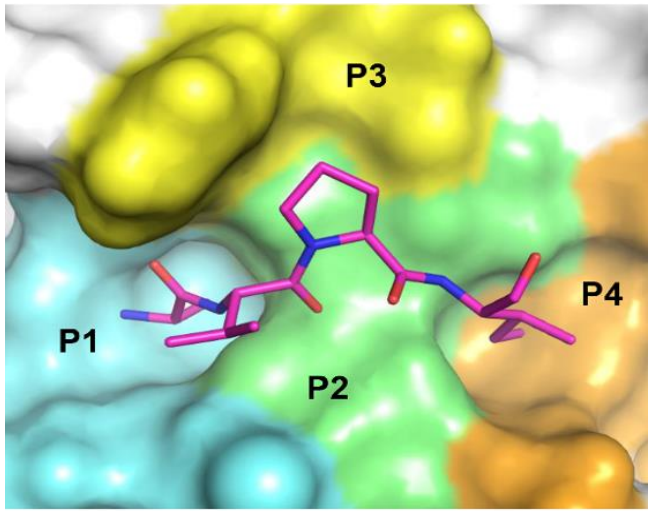
V Hit Compounds Validation

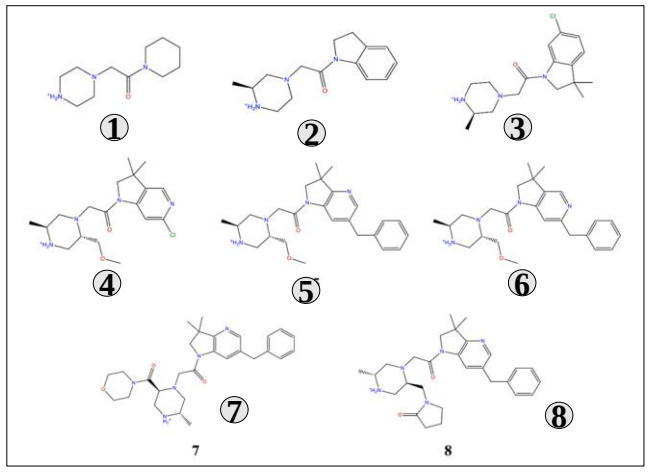
In vitro, in vivo

F Hit-based scoring function optimization
- Classification models (e. g., fluorescence assays)
- Regression models (e. g., binding affinity)

FMO for Protein-Ligand Binding Energy Calculation

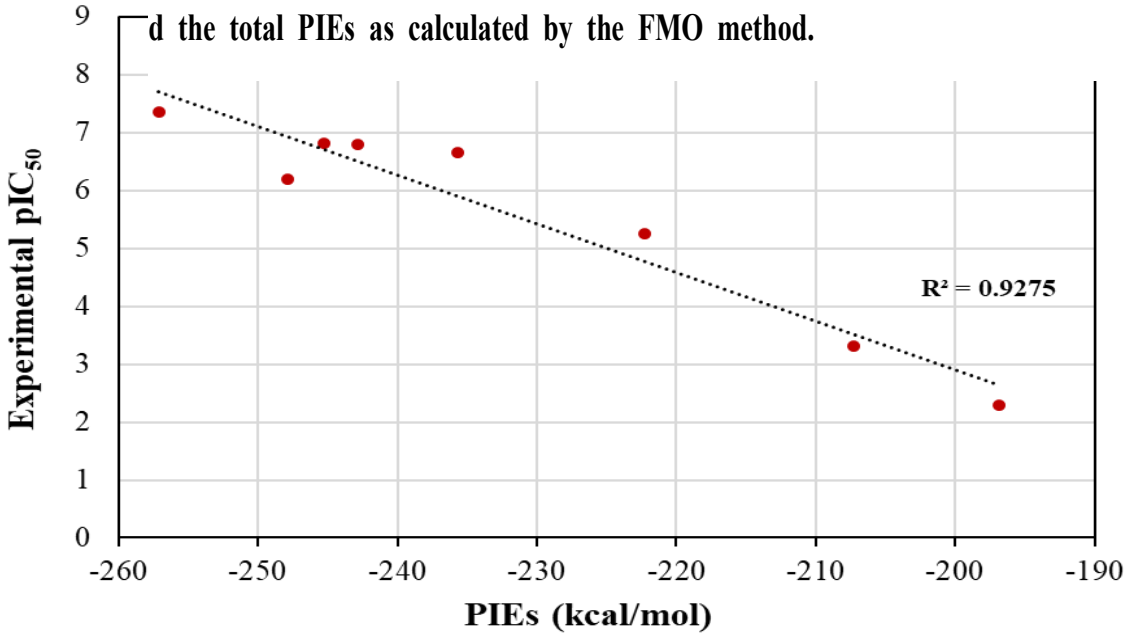
X-linked inhibitor of apoptosis protein (XIAP), inhibits caspases through its (BIR) domains.





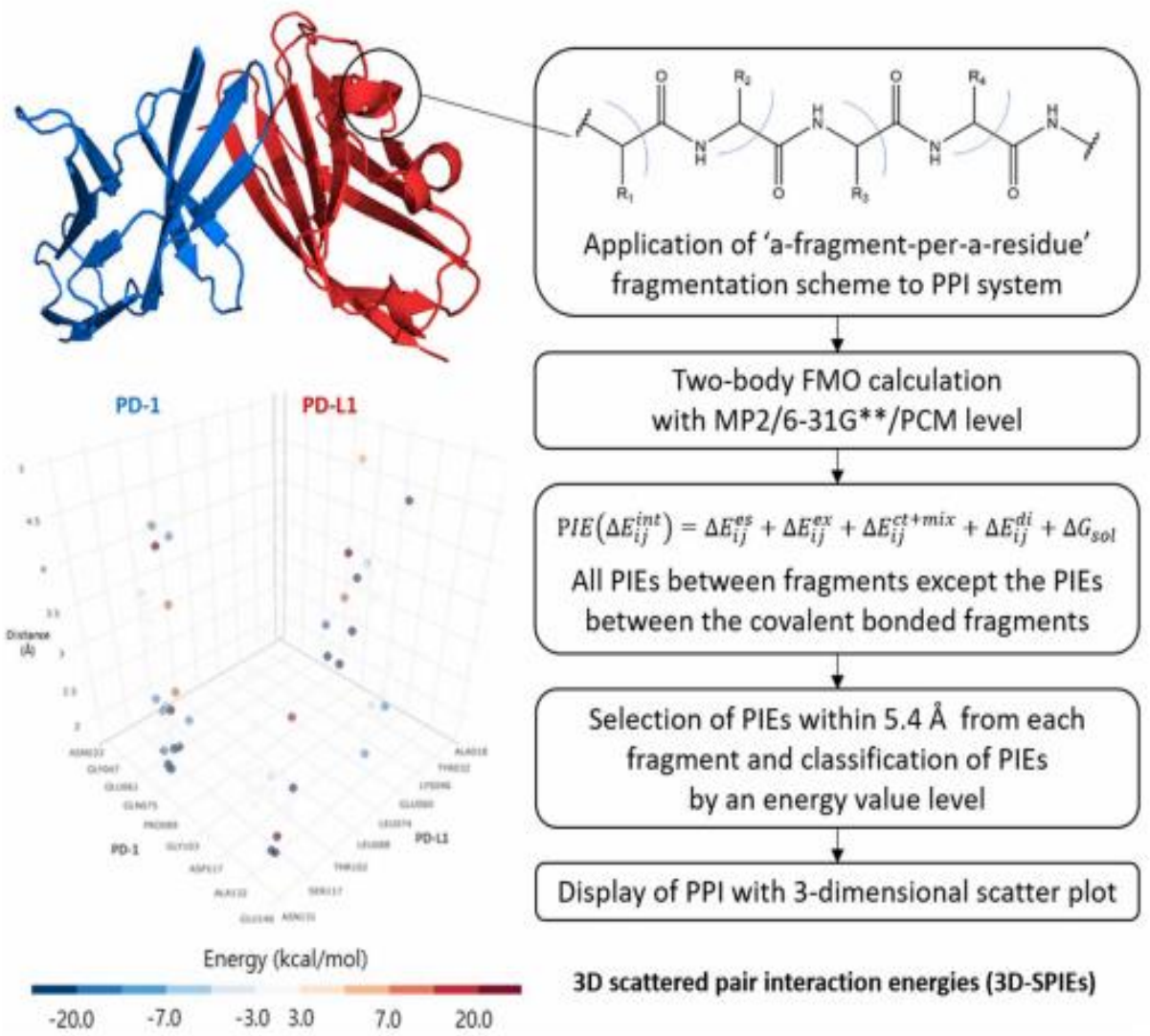
Ligand	IC_{50} (μM)	PDB (resol/A)	PIEs
AVPI	0.32	1G73 (2.00)	-246.078
1	>5,000	5C3H (2.65)	-204.792
2	>495	5C7A (2.36)	-217.849
3	5.5	5C7C (2.32)	-220.880
4	0.64	5C84 (2.36)	-257.064
5	0.22	5M6F (2.39)	-241.754
6	0.16	5C83 (2.33)	-247.424
7	0.15	5M6H (2.50)	-246.405
8	0.044	5M6M (2.37)	-246.913

Figure 1. Correlation plot between the experimentally measured binding affinity pIC_{50} and the total PIEs as calculated by the FMO method.



Computational and Structural Biotechnology Journal 17 (2019) 1217–1225

Analysis of PD1-PDL1 Interaction



Application of 'a-fragment-per-a-residue' fragmentation scheme to PPI system

Two-body FMO calculation with MP2/6-31G**/PCM level

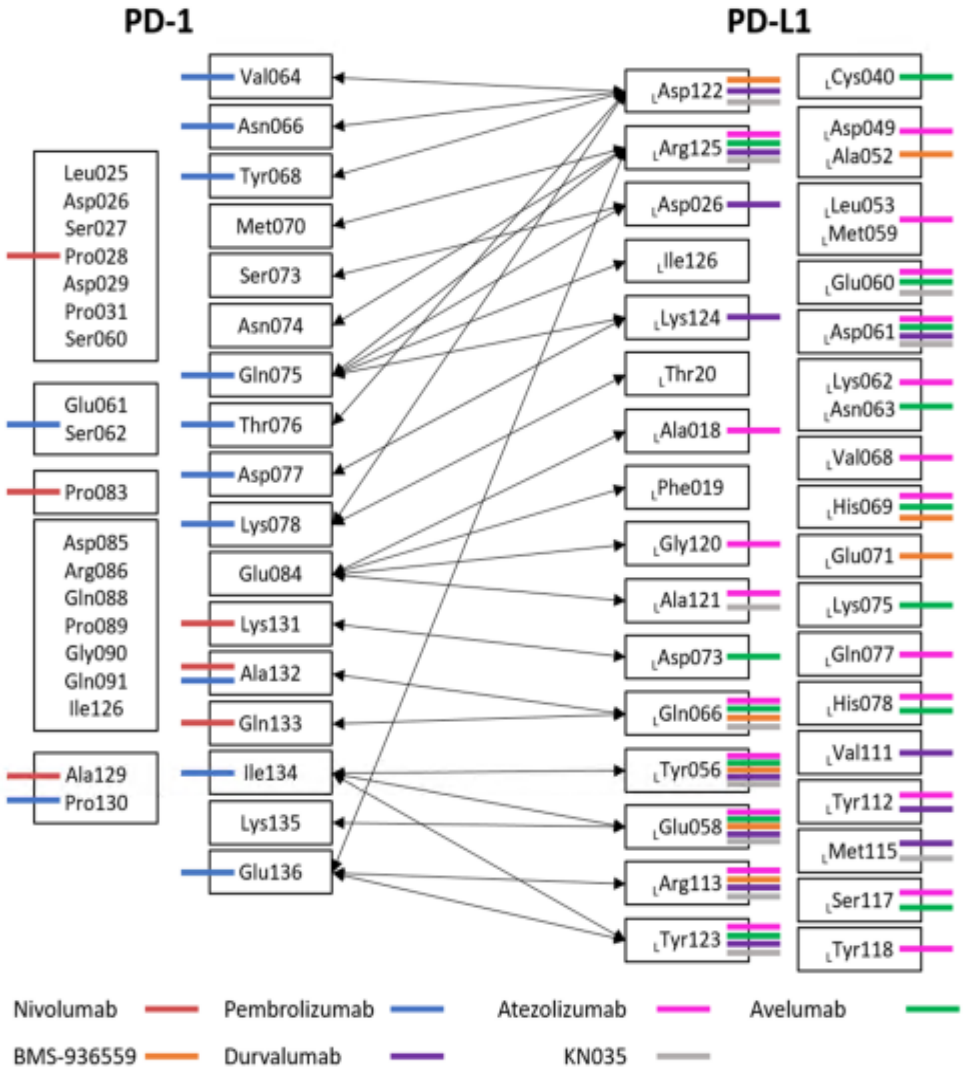
$$PIE\left(\Delta E_{ij}^{int}\right) = \Delta E_{ij}^{es} + \Delta E_{ij}^{ex} + \Delta E_{ij}^{ct+mix} + \Delta E_{ij}^{di} + \Delta G_{sol}$$

All PIEs between fragments except the PIEs between the covalent bonded fragments

Selection of PIEs within 5.4 Å from each fragment and classification of PIEs by an energy value level

Display of PPI with 3-dimensional scatter plot

3D scattered pair interaction energies (3D-SPIEs)

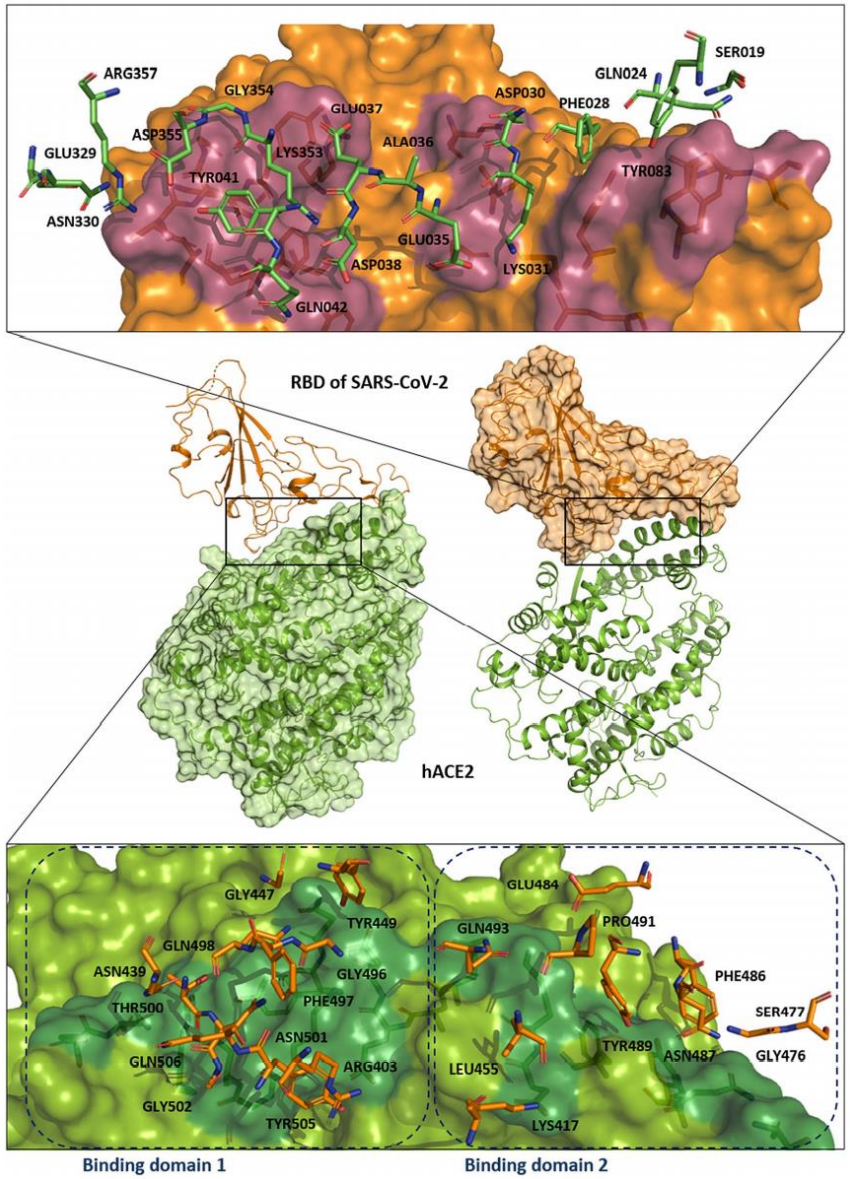


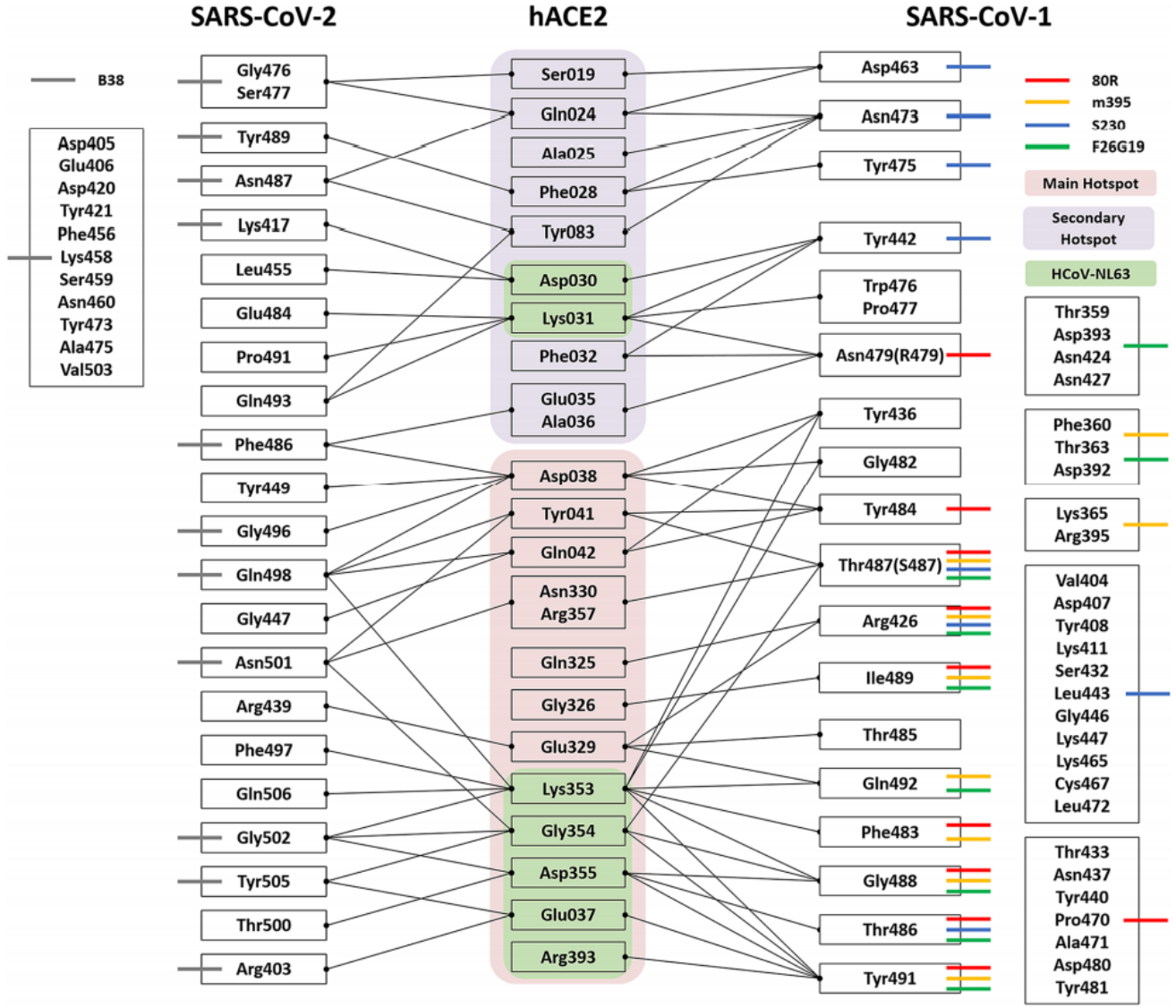
Scientific Reports | (2019) 9:16727 | https://doi.org/10.1038/s41598-019-53216-z

Analysis of SARS-CoV-2 Spike Protein & h-ACEII Interaction

	Jan. 6, 2020	Confirm Coronavirus pathogens with TEM
	In Jan, 2020	How the virus is transmitted is key to prevent & controlling it
	Feb.15, 2020	SAR-CoV2 gene sequence & purified the spike protein
	Feb. 19, 2020	Structure: SARS-CoV2 spike protein bound to h-ACE2: Cryo-EM
	Feb. 21, 2020	Structure, Function, & Antigenicity of the Spike Glycoprotein: Cryo-EM
In two months	Feb. 26, 2020	spike protein structure from SARS-CoV-2 on PDB
	In Mar., 2020	QM calculation on Spike protein-hACE2 with the structure from PDB
	Apr.20 , 2020	Submit information on the import interaction points between Spike protein and h-ACE2 (bioRxiv: April 27): FMO (QM)
	In Apr. , 2020	Virtual Screening for drug repositioning with hot spots: AVENGERS

Analysis of SARS-CoV-2 Spike Protein & h-ACEII Interaction

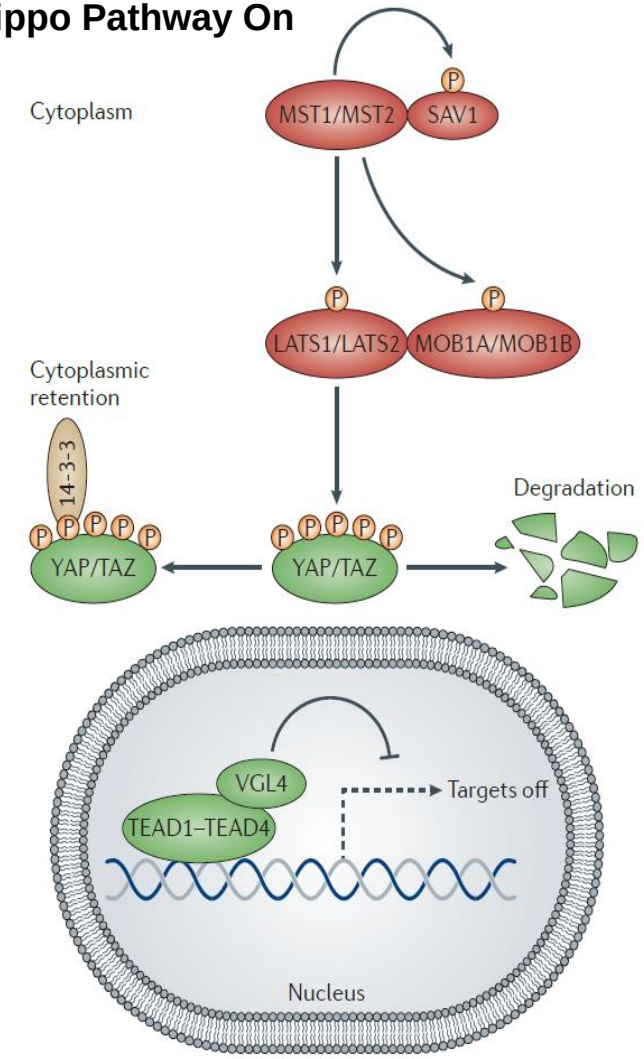




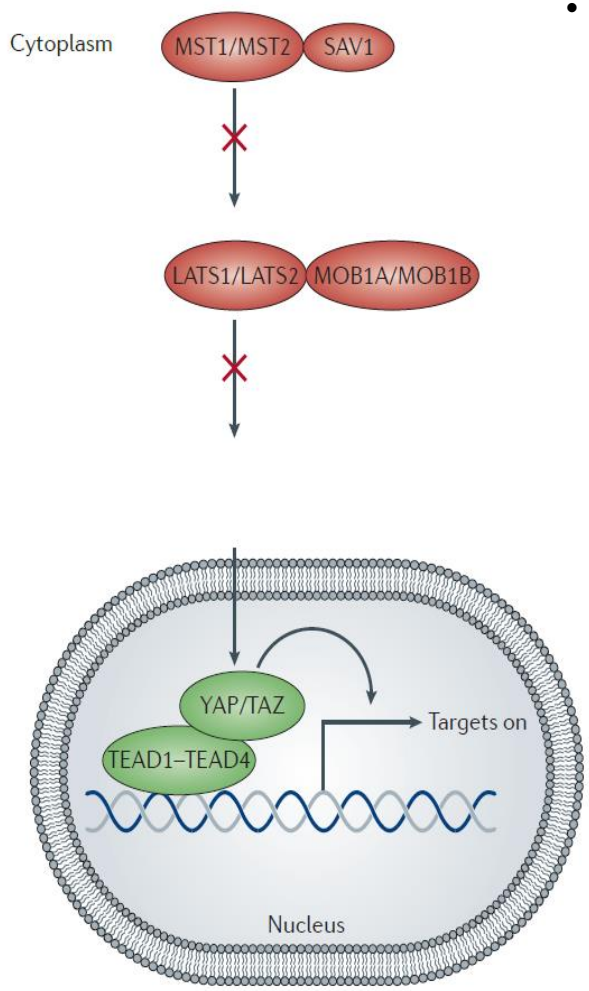
Scientific Reports | (2020) 10:16862 | https://doi.org/10.1038/s41598-020-73820-8

YAP-TEAD PPI Inhibitors Discovery, in Hippo Pathway

Hippo Pathway On

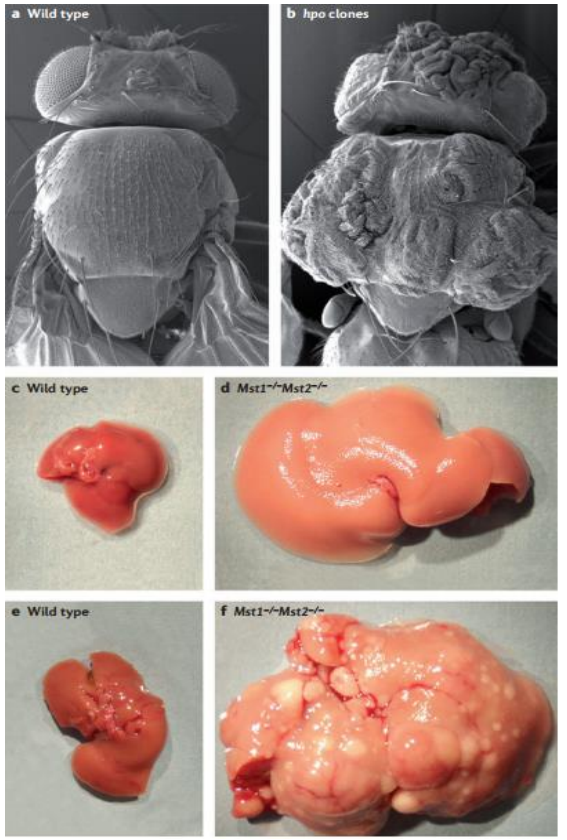


Hippo Pathway Off



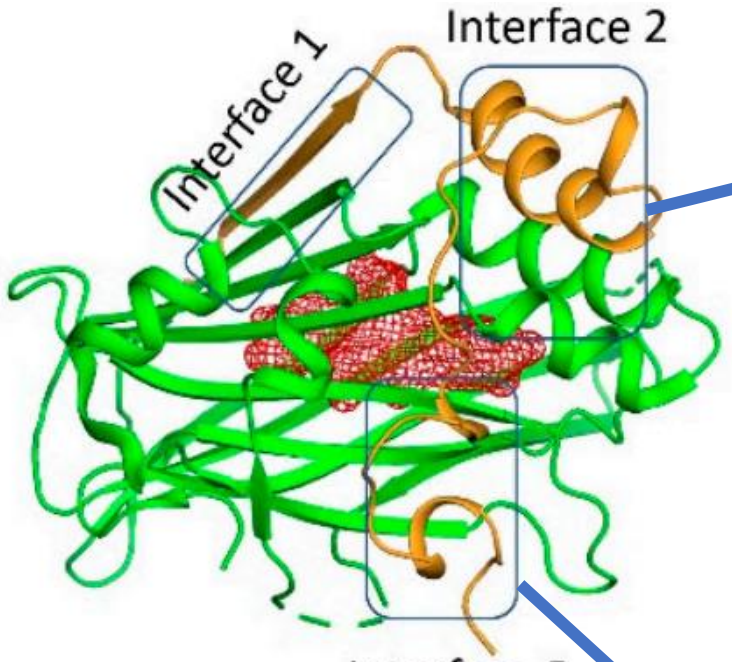
Growth control pathway

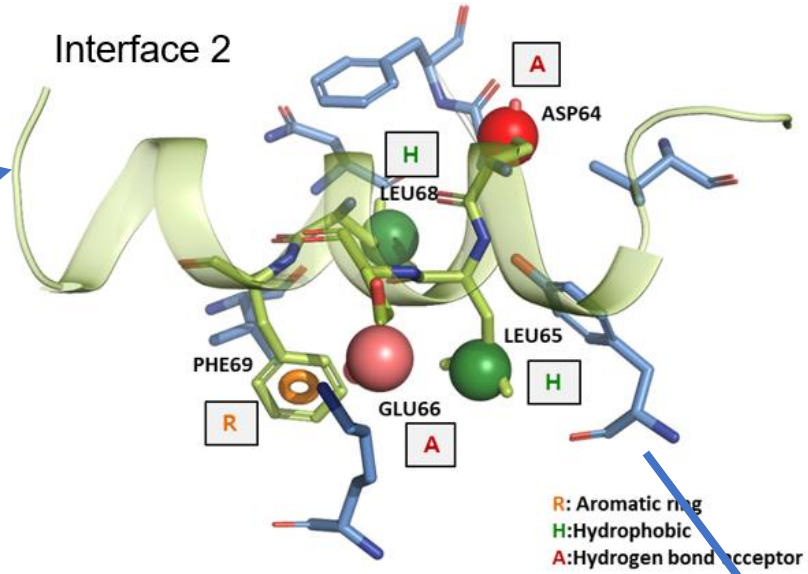
- Organ growth control, stem cell function, regeneration, and tumor suppression
- 2. Deregulated in many cancers → cancer initiation and progression

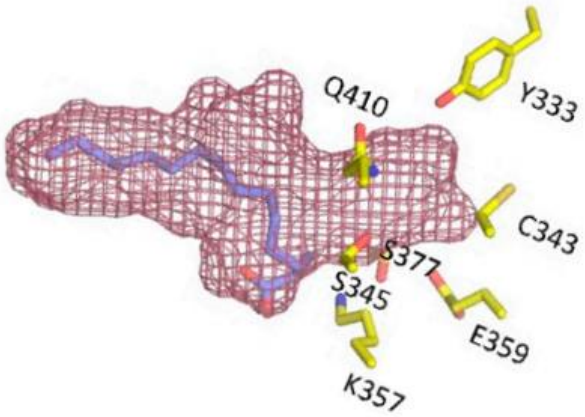


"Loss of Hippo signaling and YAP overactivation are observed in many cancer patient"

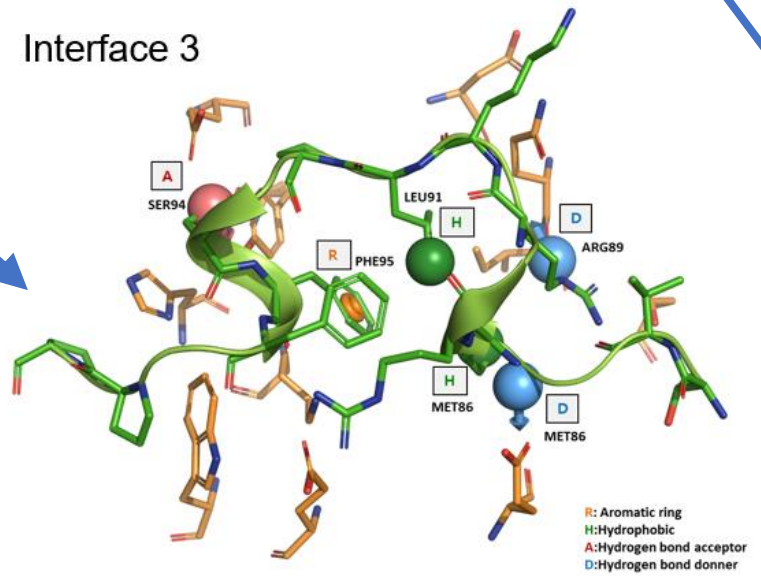
YAP-TEAD PPI Inhibitor Discovery (TEAD Targeting)



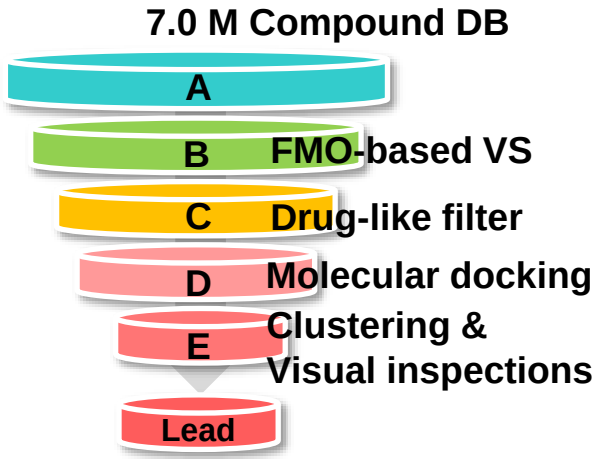




Central Pocket

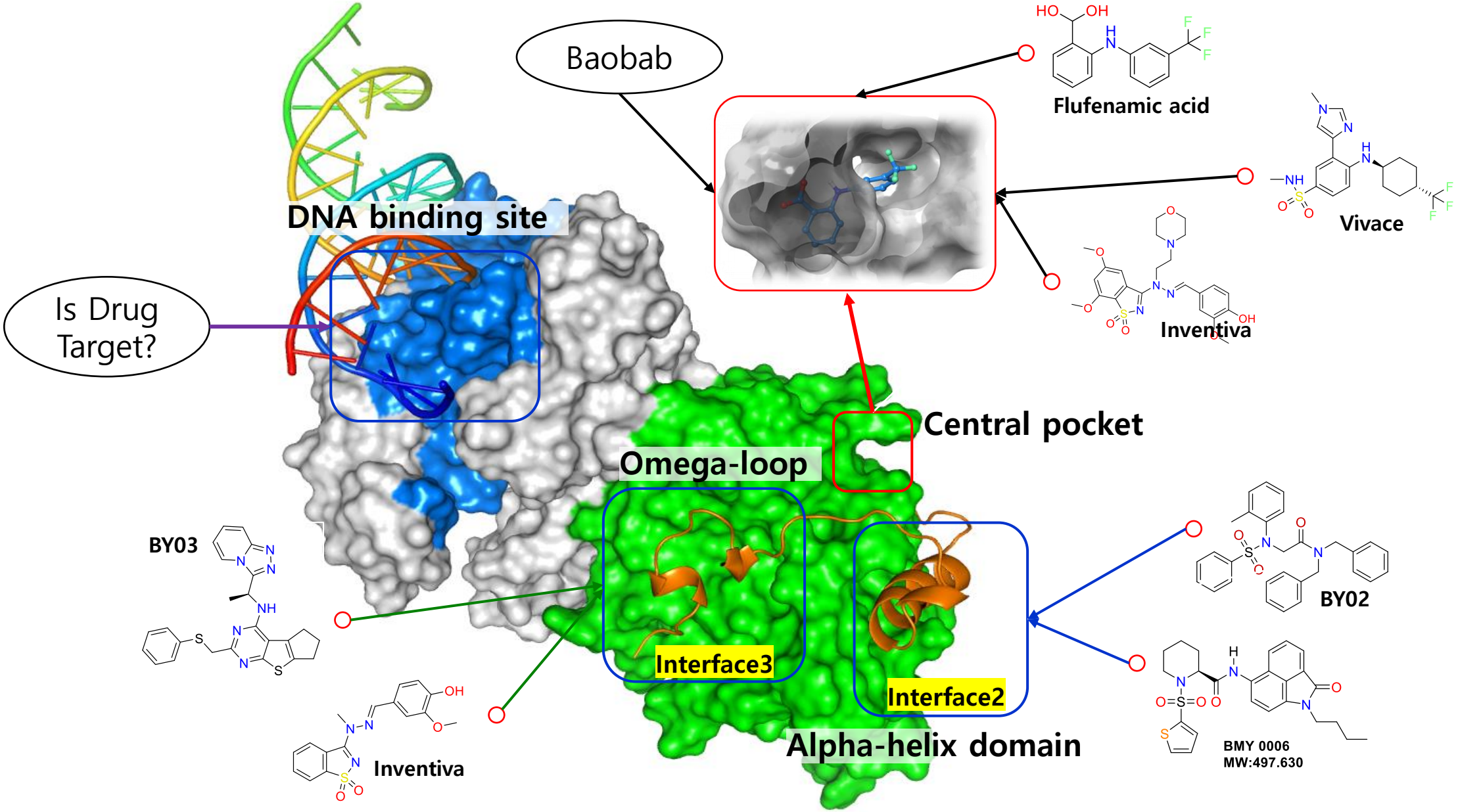


- **Features of pharmacophore were generated from hot spot information obtained with FMO calculations**
- **Features are selected within the surface range that small molecules can cover.**
- **Then Virtual Screening**

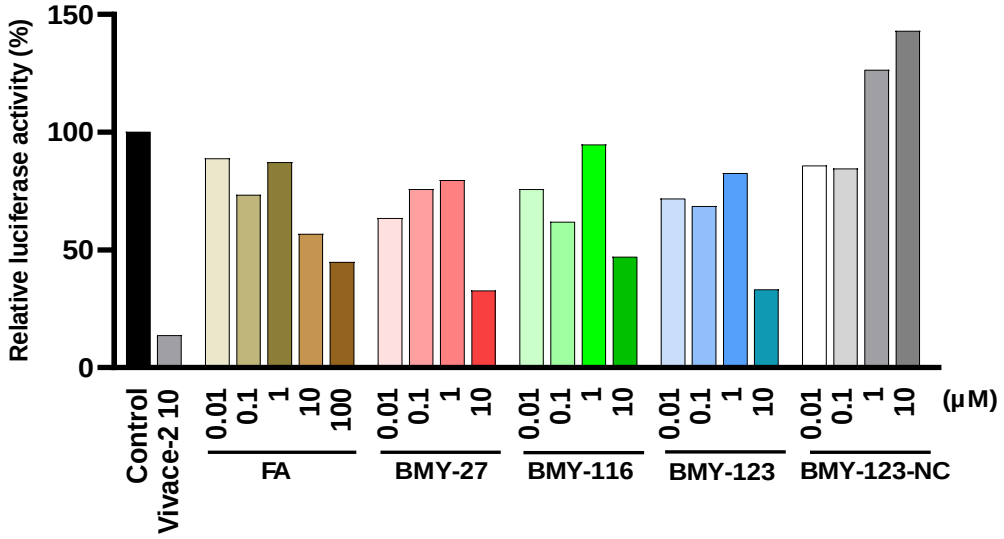


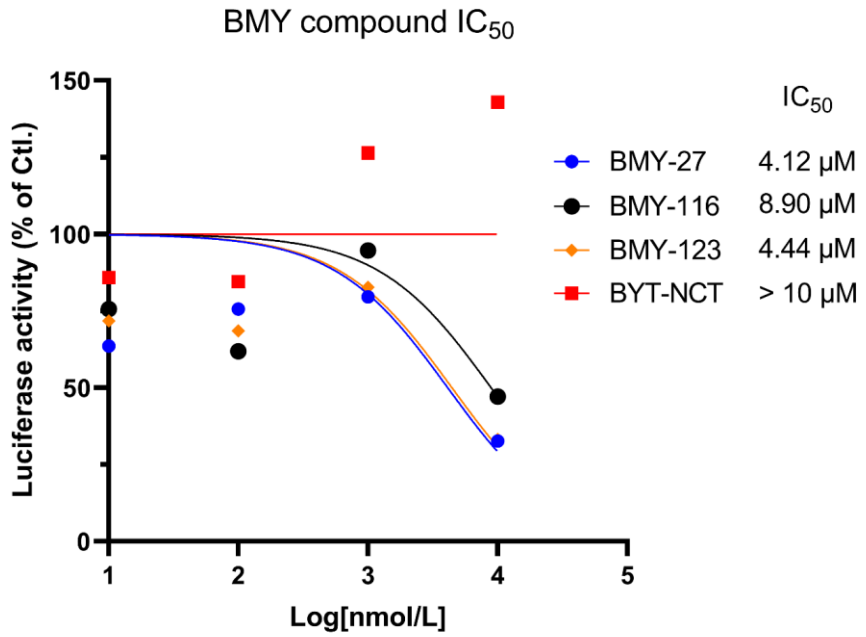
Cancers **2021**, 13, 4246

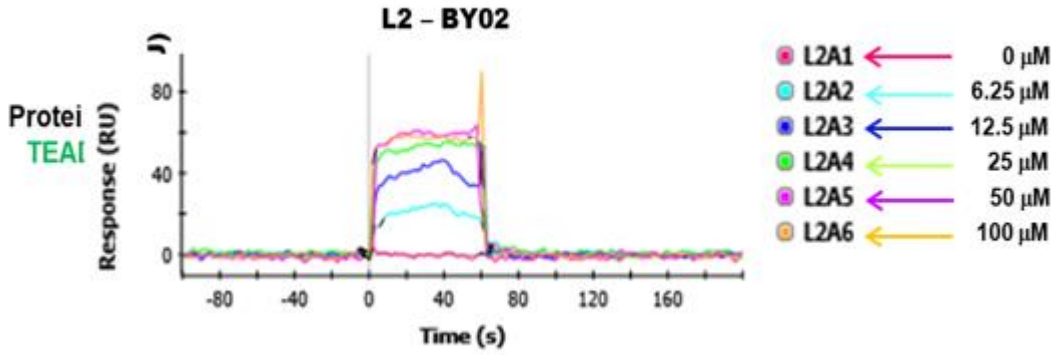
YAP-TEAD PPI Inhibitor Discovery (TEAD Pharmacophores)

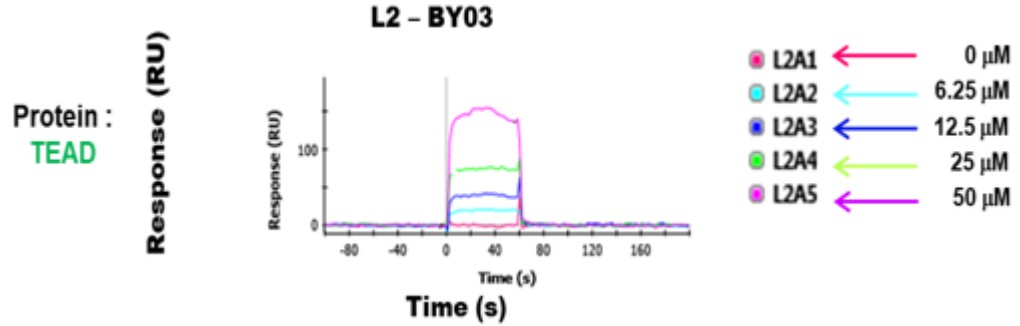


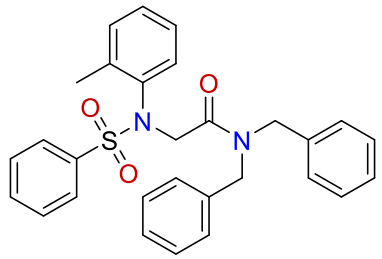
Measure Activity of BY02, BY03 by Luciferase Report Assay



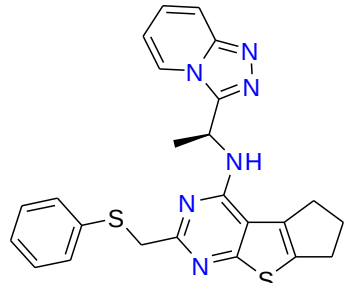






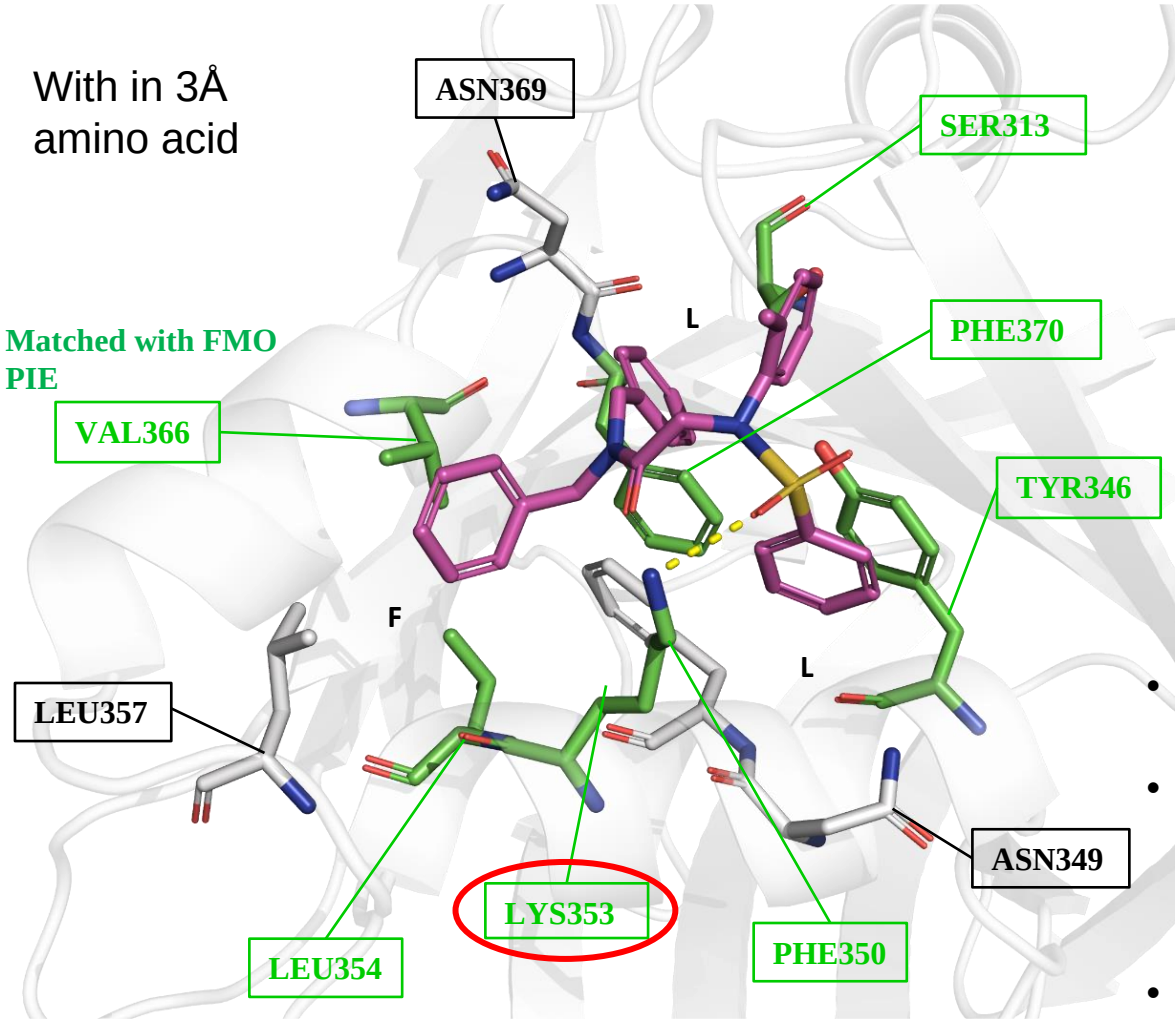


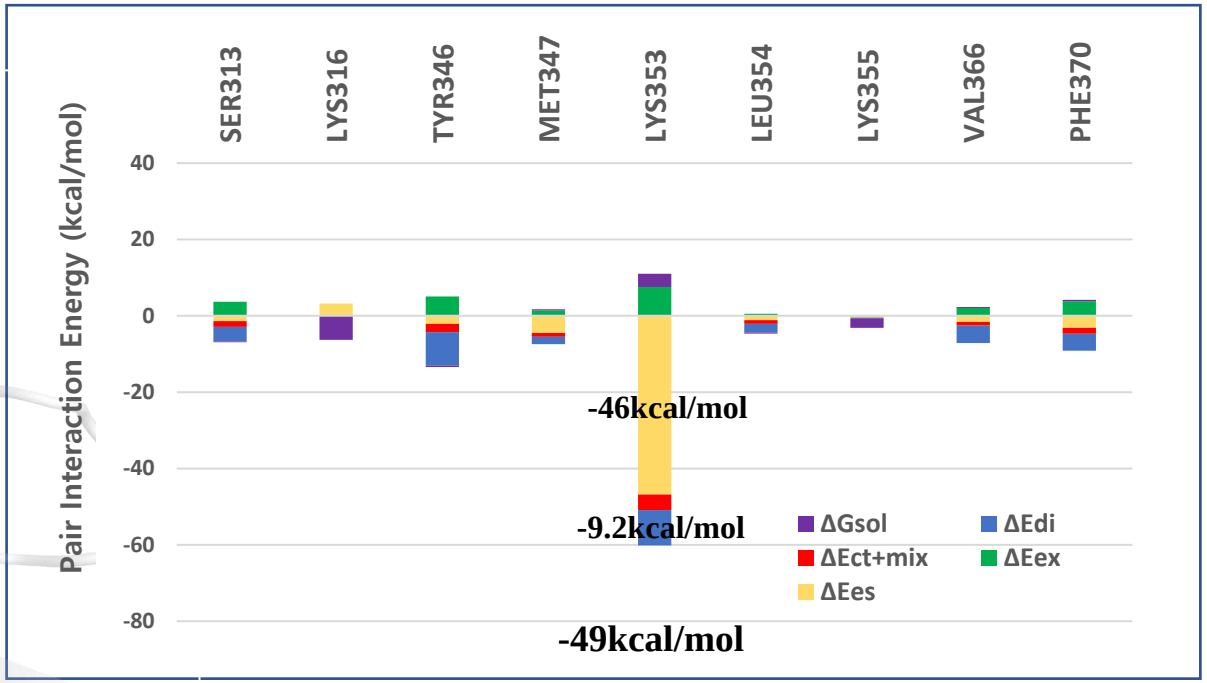
BY-02



BY-03

Binding Energy Analysis of BY-02 with FMO PIE





- TEAD's binding amino acids are represented within 3Å from BY02 docking pose.
- BY02 was performed at FMO-DFTB3/D/PCM level with the third order corrected density functional tight-binding (DFTB3) method with 3OB parameter set, UFF-type dispersion correction (D), and polarizable continuum model (PCM).
- In energy minimization, the residues within 10.4 Å from ligand were included and fixed, while only the ligand was fully flexible.

Cancers **2021**, 13, 4246

FMO based Protein Engineering Platform, CARPET

Phylogenetic tree

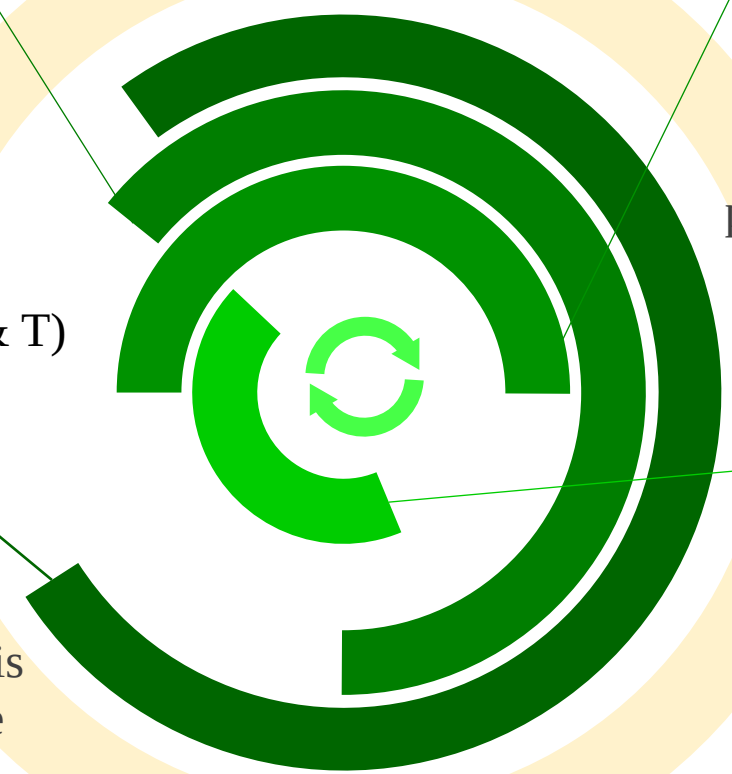
Protein sequence data are used for generating phylogenetic tree and protein mutation pool for binding affinity, thermal stability, and so on.

- NCBI Sequence & Meta data (pH & T)
- Phylogenetic tree with multiple sequence analysis

Cryo EM Structure

Cryo-EM generate scattering image of target protein complexes. The image is converted to the electron density of the target. Then the electron density is converted to protein structure.

- CryoSPARC: Images to electron density
- NNP/QM/MM : with electron density to high resolution structures



ROSETTA

AlphaFold 2

Protein Binding Analysis-QM

QM methods are used for analyzing the crucial interactions in target protein with Cryo-EM structures and construct the hot spot map on the target protein-protein interface.

- 3D-SPIEs with QM Fragment Molecular Orbitals (FMO)

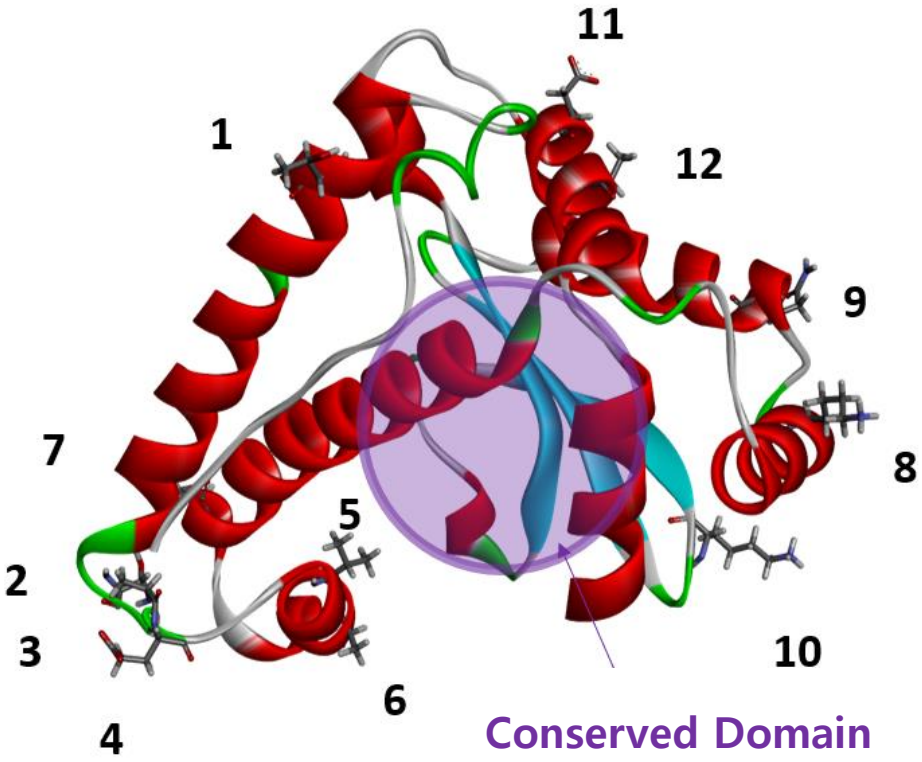
Computational Mutagenesis

Neural Network Potential Guided Statistical Mechanical methods are used in performing computational mutagenesis and confirming the specific properties.

- Free Energy Perturbation (FEP)
- Thermodynamic Integration (TI)
- Neural Network Potentials (NNP)

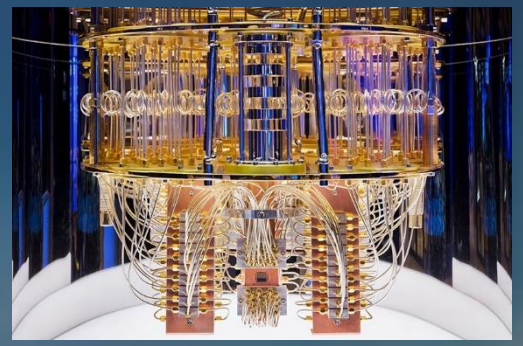
Protein Engineering with CARPET – Superoxide Dismutase

- **Propose 12 Variants from 50% Thermophiles**
- 12 heat-resistant SOD candidates were cloned
- Purified variants were characterized for:
  - ✓ Activity, protein quantification, heat resistance



Conserved Domain

	Specific Activity (U/mg)	Residual activity (%)	BCM_Tm1 (°C)	Tagg 266 (°C)
wild type	2200	47.5	55	64
1	1708	75	65	38
2	2161	92	66	37
3	1381	86.5	67	57
4	868	89.1	51	39
5	2618	80.8	64	51
6	650	55.5	66	36
7	1641	82.7	44	50
8	1934	50.7	NaN	38
9	1857	56	NaN	56
10	1747	104.2	61	52
11	2124	24.1	63	54
12	1503	45.1	46	56



Quantum Computing

Pharma & Chemical Industry-Quantum Computing

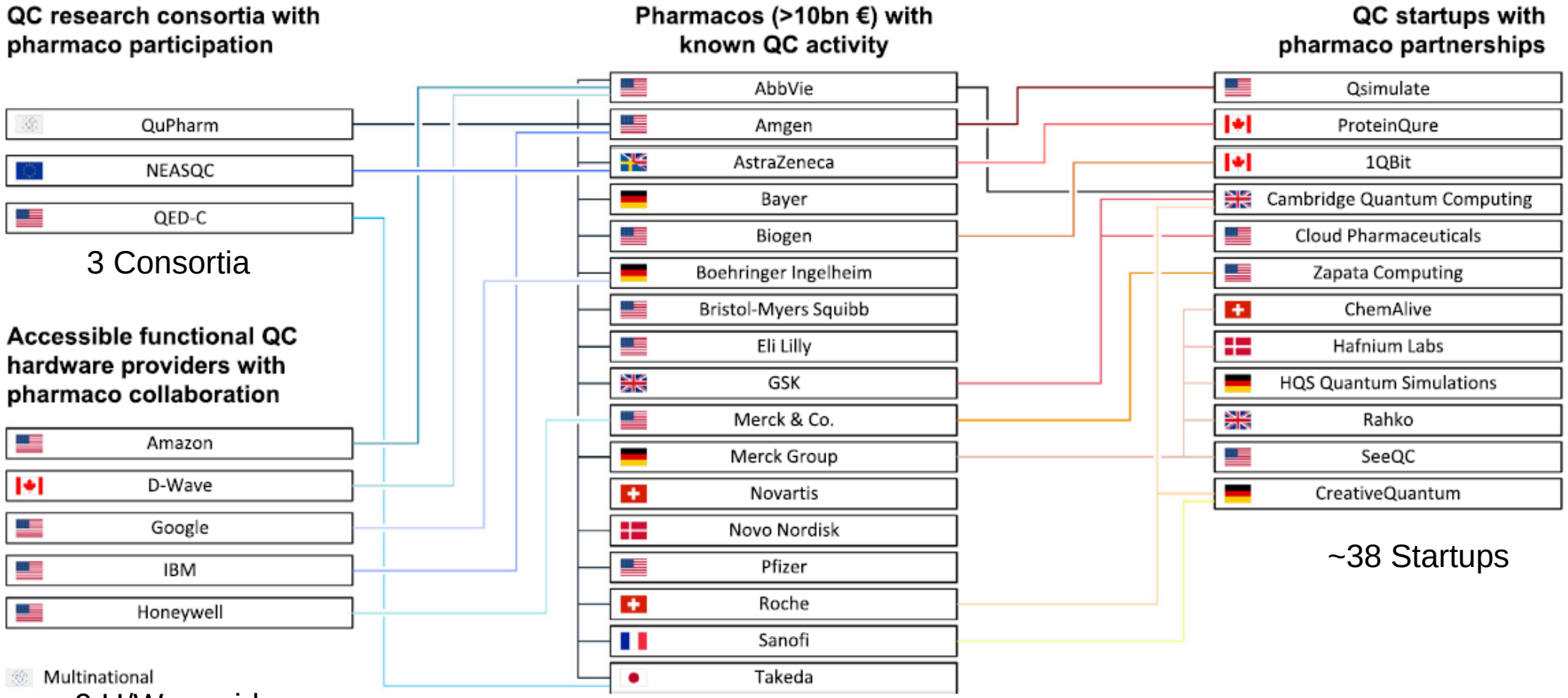
- Quantum computing's ability to simulate larger, more complex molecules could be game changing. Pharmaceutical companies should reflect on their **strategic stance to this promising new technology now**.
- Pharma & chemical industry will be one of the **first industries to benefit** by the impact of Quantum Computing (QC).
- Given its focus on molecular formations, pharma as an industry is a **natural candidate** for QC.
- Since molecules are actually quantum systems, systems that are based on quantum physics, QC is expected to be able to predict and simulate the structure, properties, and behavior of these molecules **more effectively than conventional computing can**.

June 2021, McKinsey & Company

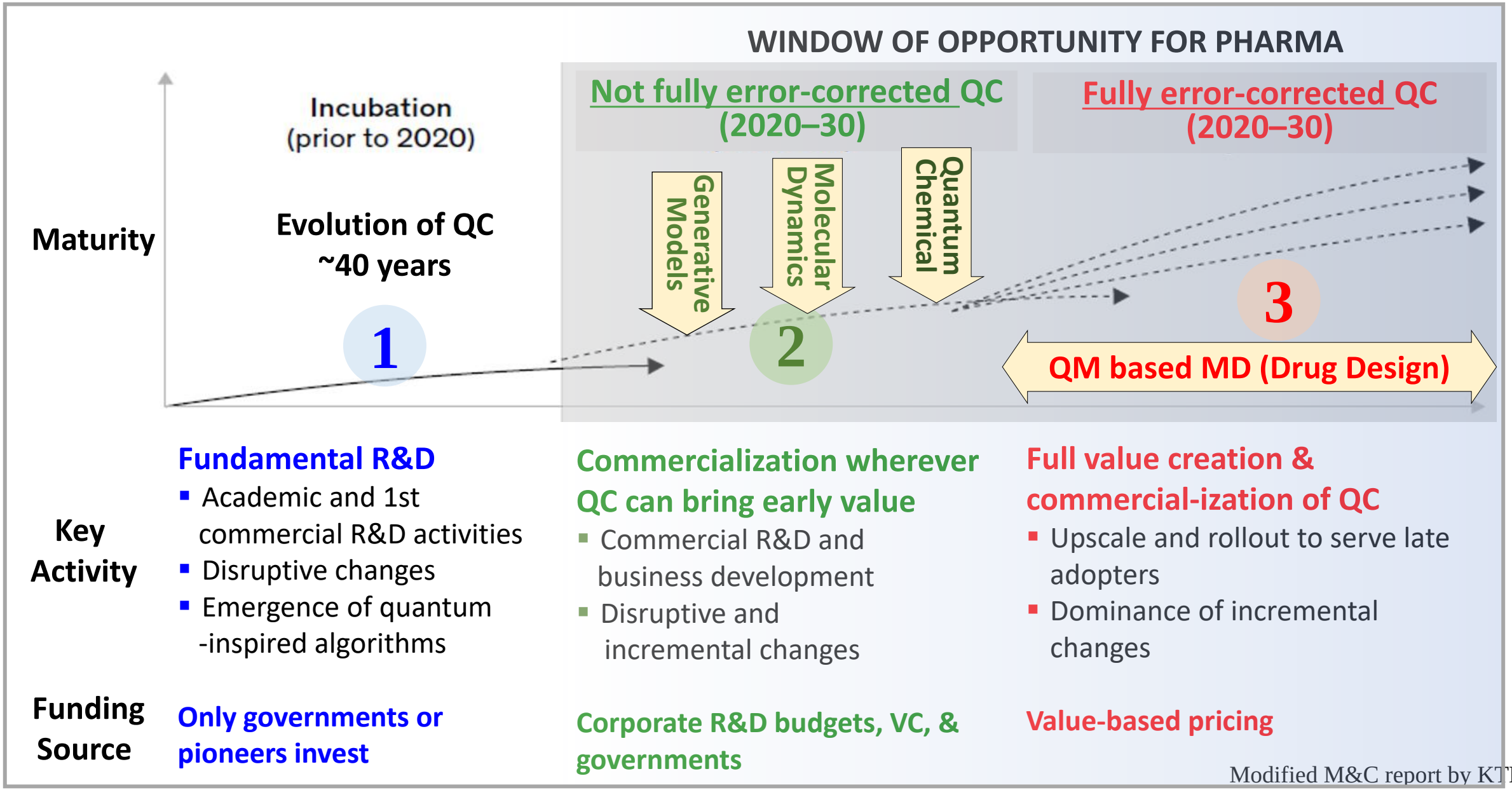
Quantum Computing Ecosystem in Drug Discovery

Pharmacos: For Drug Discovery & Development
Hardware providers: Quantum Computer & Languages

Startups: Provide Quantum Algorithm for Drug Design

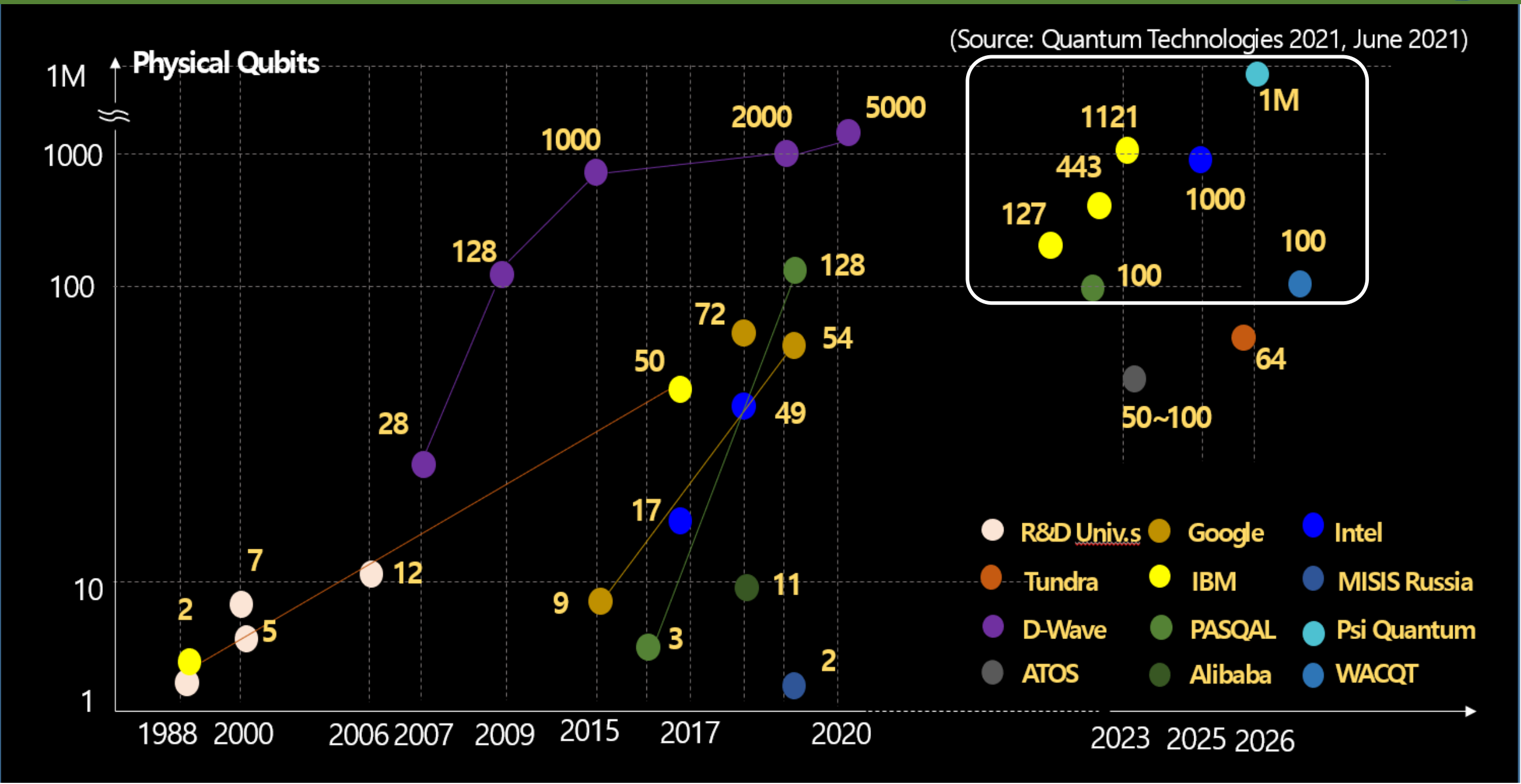


Value Creation through QC in the Pharmaceutical Industry

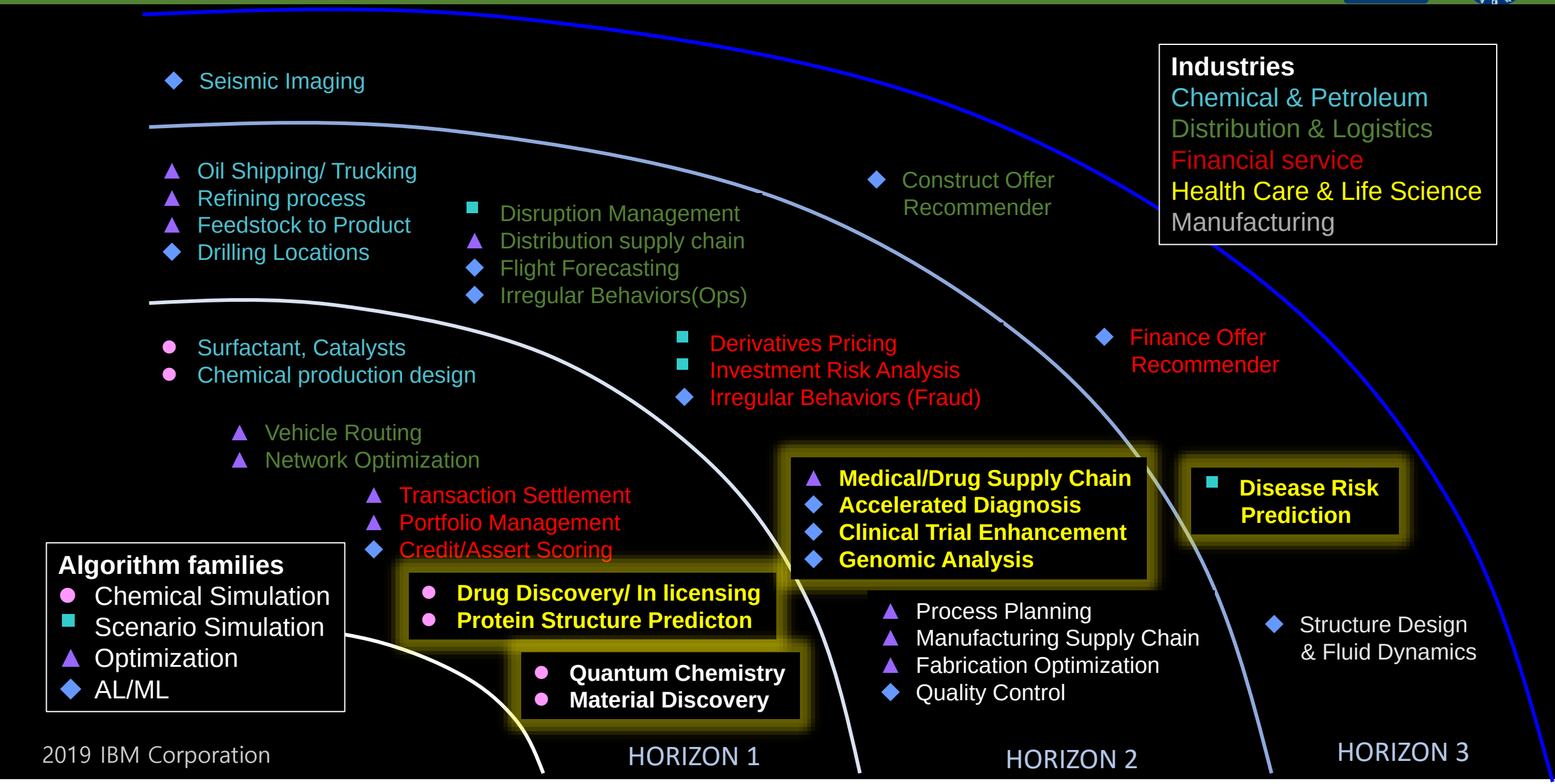


Modified M&C report by KTNo

Physical Qubit Roadmap in Quantum Computer



Quantum Volume and Industrial Application Potential

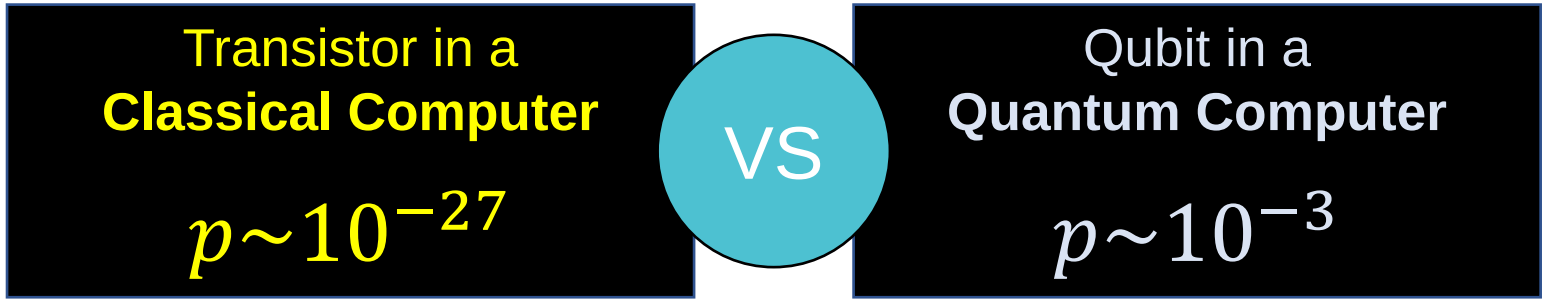


2019 IBM Corporation

Quantum Volume & Quantum Error (Noise) Correction in QC

Quantum Error: High Probability of Hardware Error in Quantum Computer

→ Need Error Correction → Stabilizer



~24 orders of magnitude difference

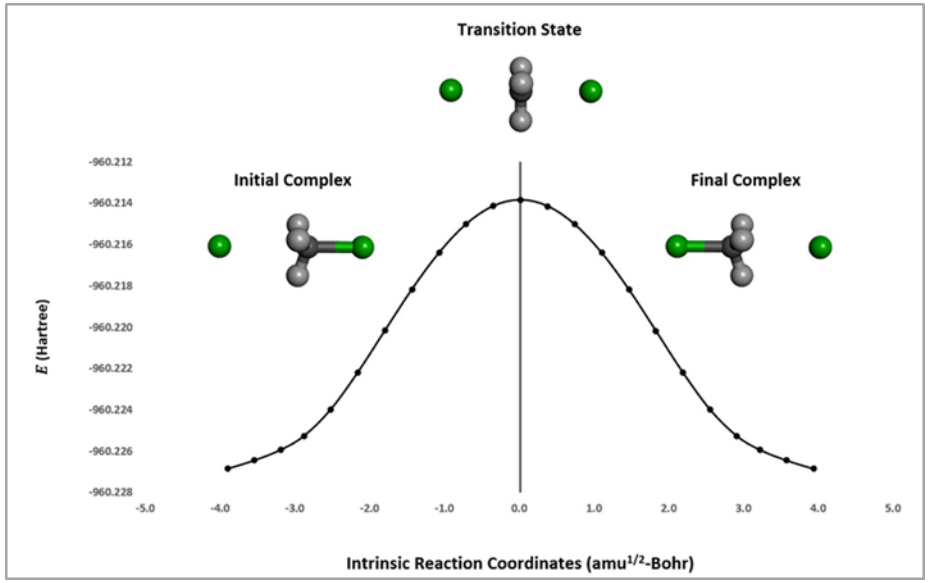
IBM Quantum / © 2021 IBM Corporation

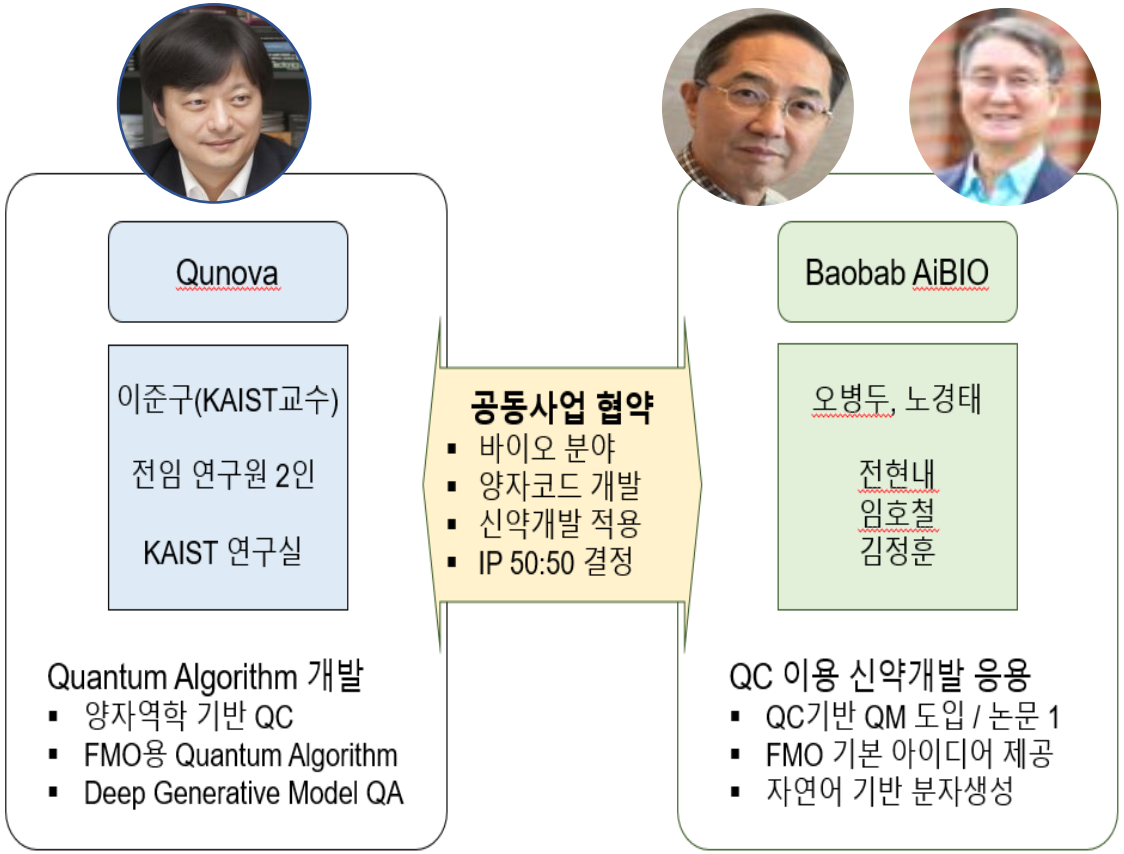
Quantum volume: a metric that measures the capabilities and error rates of a quantum computer. It expresses the maximum size of square quantum circuits that can be implemented successfully by the computer.

다음 2장 1쪽으로

Quantum computational study of chloride ion attack on chloromethane for chemical accuracy and quantum noise effects with UCCSD and k-UpCCGSD ansatzes

- 현재의 Quantum Volume(QV)과 Quantum Error Correction(QEC) 환경에서 가능한 화학 시스템 계산
- QV와 QEC의 개선 후 분자설계 가능성 확인 및 준비
- FMO를 위한 알고리즘의 협력 개발 및 IP 공유
- QC를 사용한 구조기반 모델링 공동 개발 계약
- 국내 최초로 QC를 TS 계산에 적용





SUMMARY

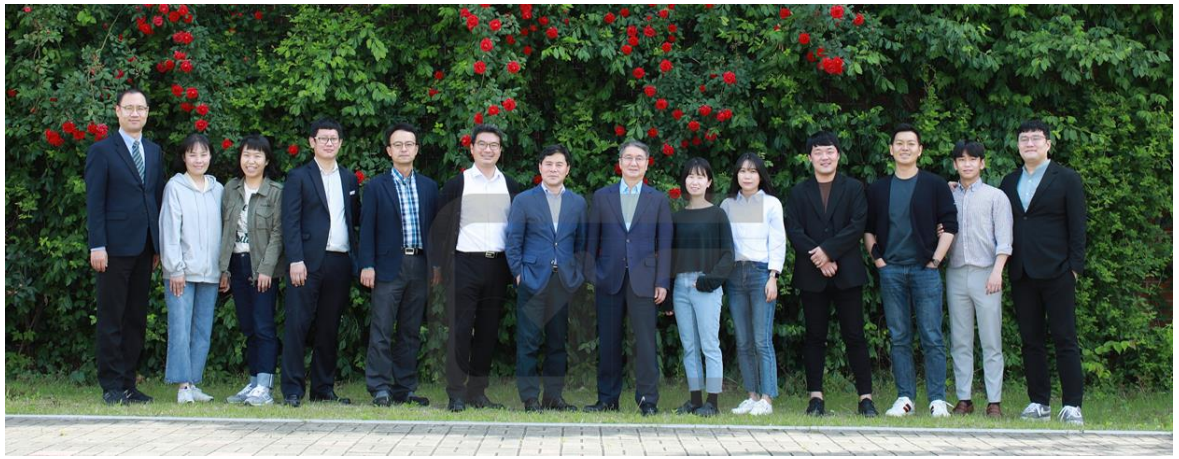
- **FMO is a very useful tool for calculating the electron density and energy of proteins.**
- **Using FMO, it is possible to calculate the interaction energy between a protein and a ligand, and furthermore, the interaction type can be analyzed with the energy-decomposition tool (PIE).**
- **Compared to QM calculation, NNP has a very short calculation time, but the energy can be obtained even at the CCSD MO level depending on the training data set.**
- **We developed AVENGERS for small molecular drug design and CARPET for protein design based on the FMO method. Also, the usefulness of these two platforms was verified through various experiments.**
- **The introduction of quantum computing in Computer-Aided Drug Discovery will allow CADD to lead the entire process of drug discovery within 10 years.**

Acknowledgment

Bioinformatics & Molecular Design Research Center

- Computer-Aided Drug Design
- CADD Tool Development
- AI & Big Data Analysis

BaobabAiBIO Inc.



- Cryo-EM
- Medicinal Chemistry
- Biology

http://www.bmdrc.org