

# Lecture 10: Mathematics for Drug Discovery

Guowei Wei

Mathematics

Michigan State University

<https://users.math.msu.edu/users/wei/>

NSF-CBMS Conference on Mathematical Molecular Bioscience and  
Biophysics

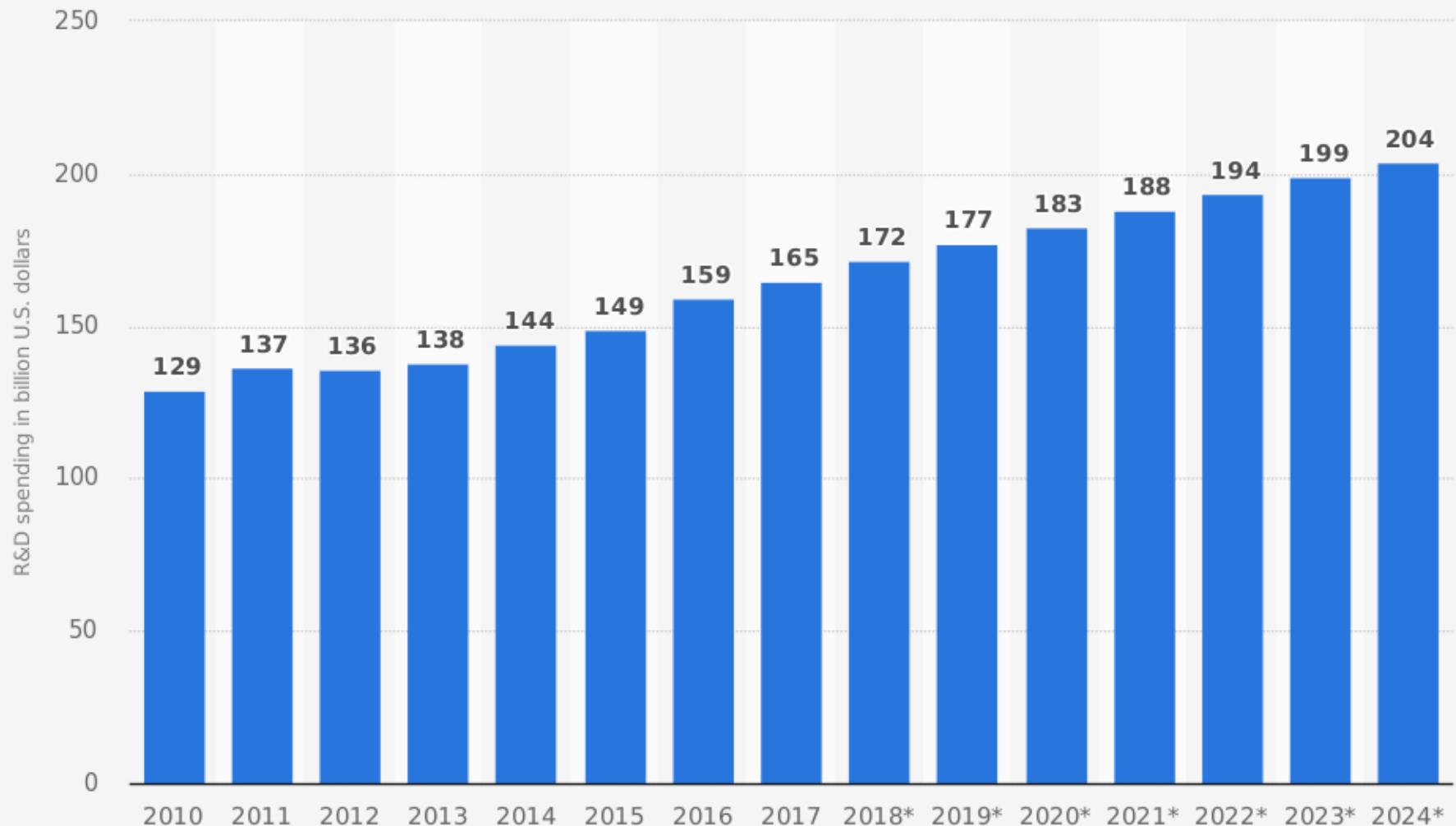
University of Alabama

Tuscaloosa, May, 13-17, 2019

**Grant support: NSF, NIH, MSU, BMS, and Pfizer**



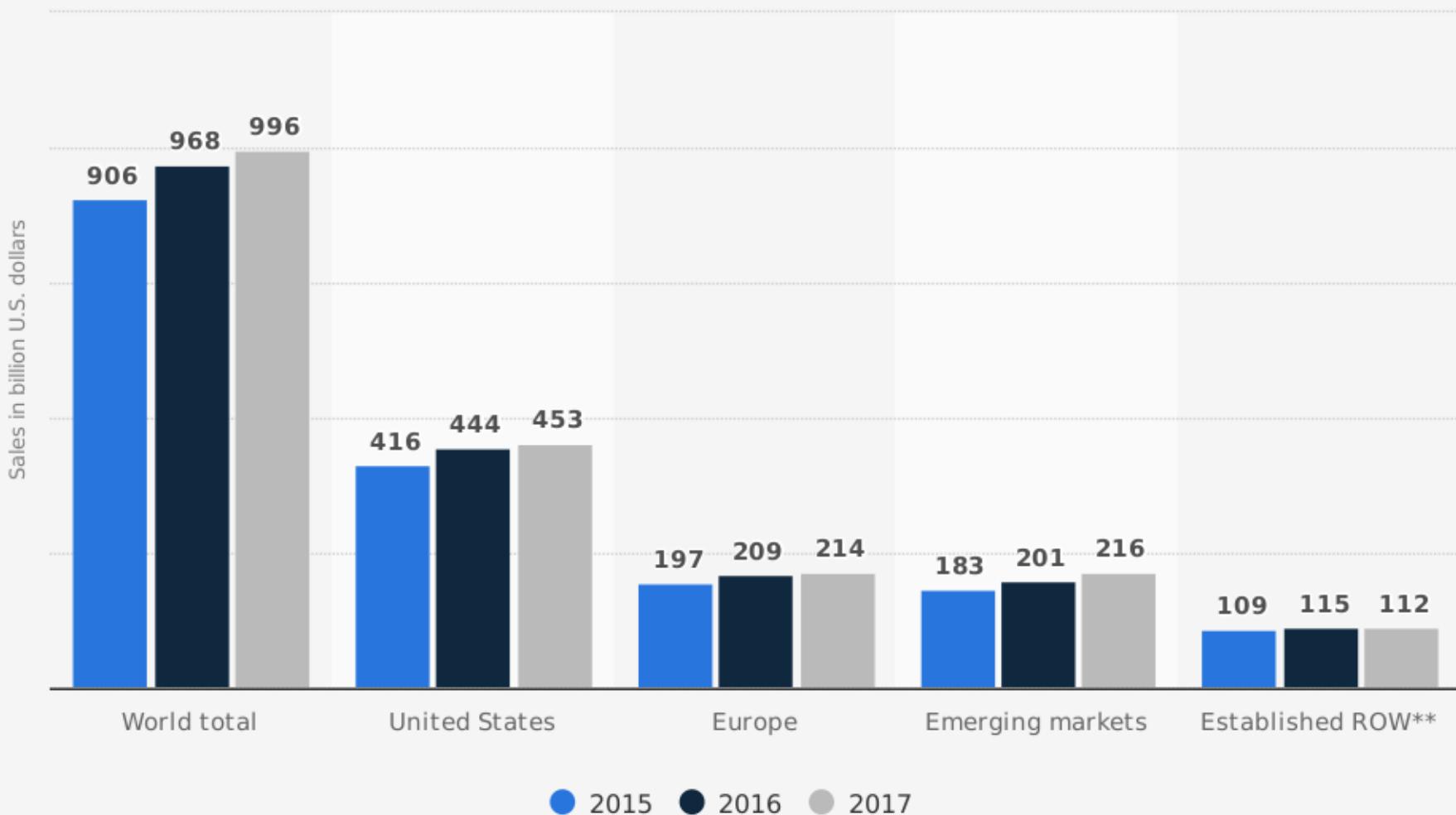
# Total global spending on pharmaceutical research and development from 2010 to 2024 (in billion U.S. dollars)



Source  
Evaluate  
© Statista 2019

Additional Information:  
Worldwide; Evaluate (EvaluatePharma); as of May 2018

# Global pharmaceutical sales from 2015 to 2017, by region (in billion U.S. dollars)\*



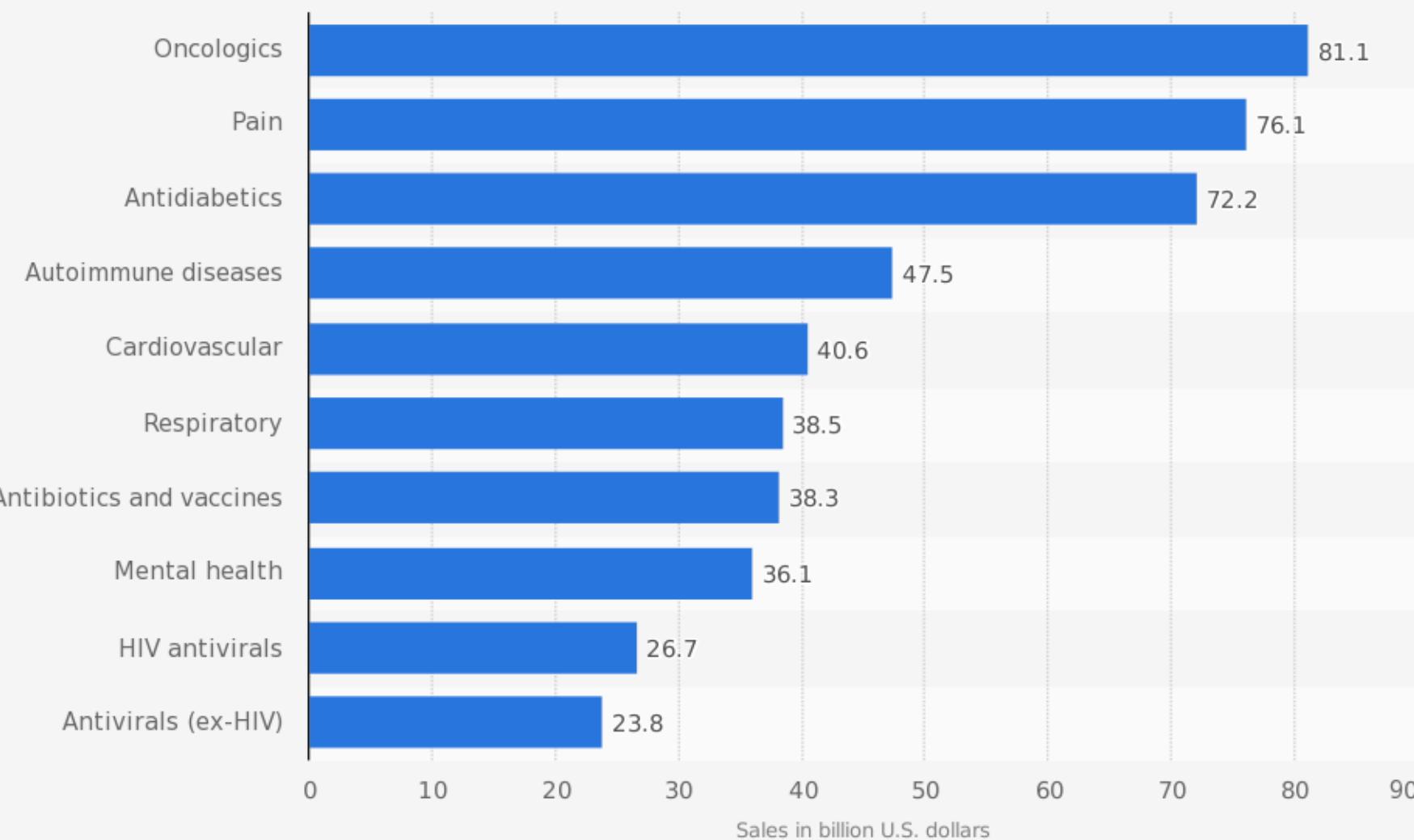
## Sources

AstraZeneca; IQVIA  
© Statista 2019

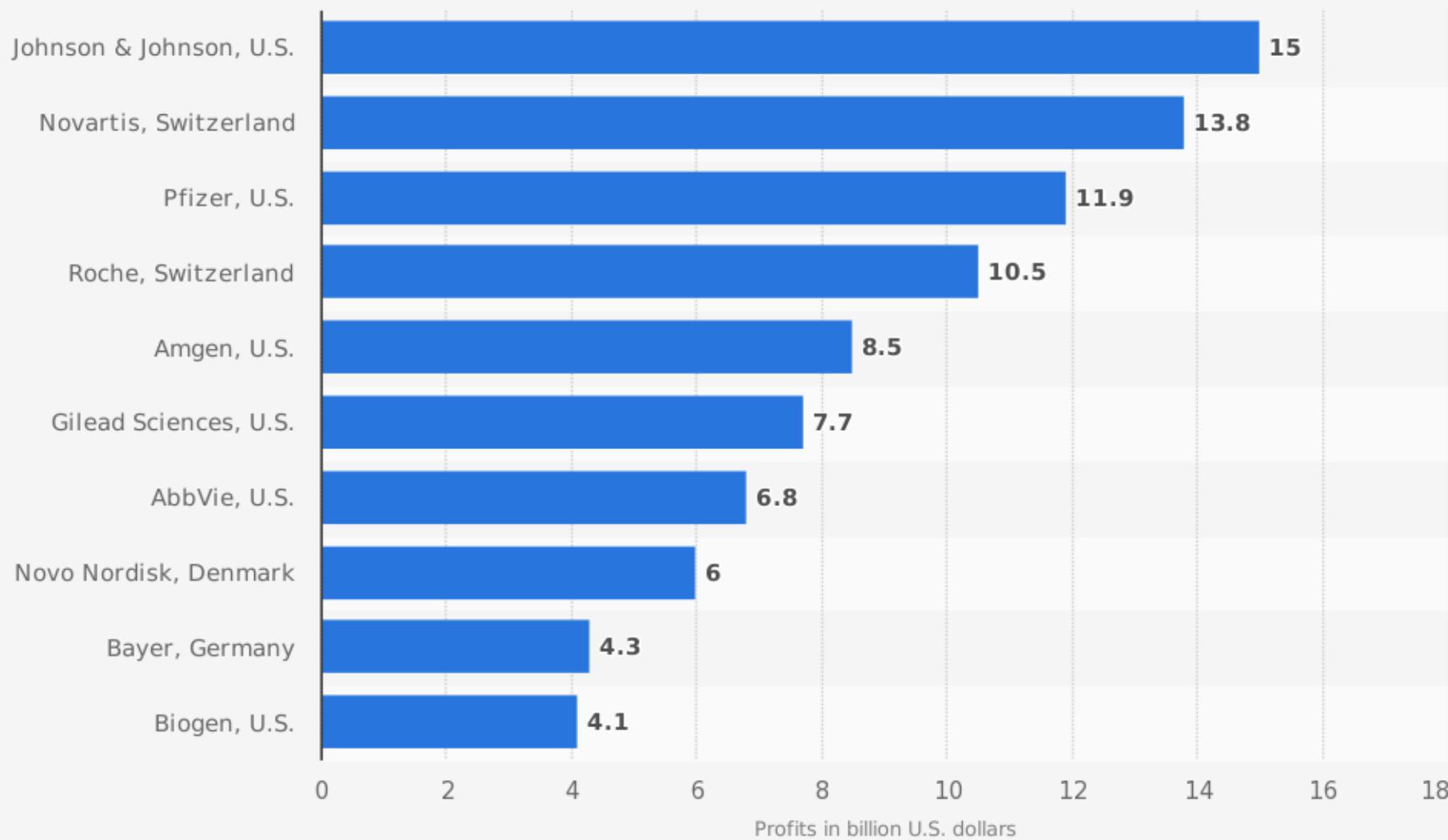
## Additional Information:

Worldwide; IQVIA (Midas Quantum); 2015 to Q3 2017

# Top 10 therapeutic classes by estimated global pharmaceutical sales in 2017 (in billion U.S. dollars)



## 2018 ranking of the global top 10 biotech and pharmaceutical companies based on net income (in billion U.S. dollars)



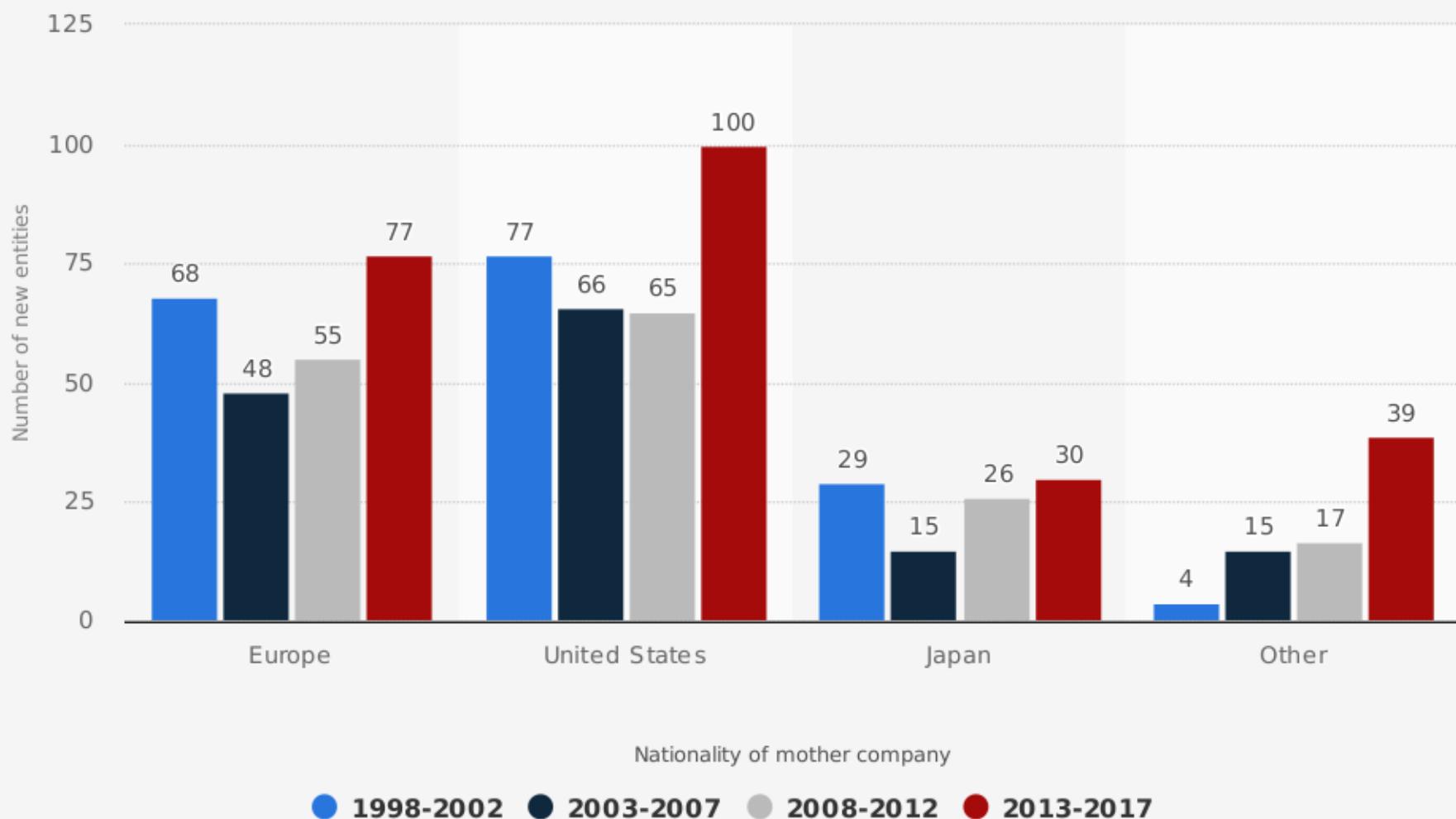
### Sources

Thomson Reuters; Various sources (company data); Financial Times  
© Statista 2019

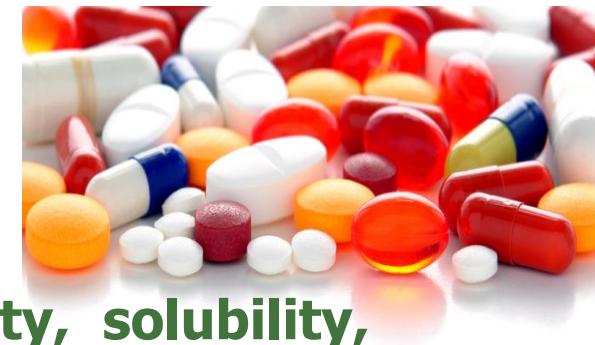
### Additional Information:

Worldwide; Thomson Reuters; Various sources (company data); as of August 2018

## Number of new chemical or biological entities developed between 1992 and 2017, by region of origin

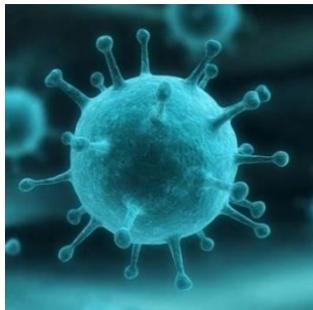


# Drug design and discovery

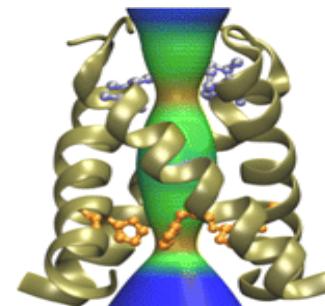


- 1) Disease identification (physiology)
- 2) Target hypothesis (biochem./mole. biol.)
- 3) Virtual screening: drug pose, binding affinity, solubility, partition coefficient, toxicity, and side-effects (biophysics/bioinformatics)
- 4) Drug structural optimization in the target binding site (biochemistry/biophysics/synthetic chem.)
- 5) Preclinical *in vitro* and *in vivo* test
- 6) Clinical trials
- 7) Optimize drug's efficacy, pharmacokinetics, and pharmacodynamics properties (quantitative systems pharmacology)

Influenza -- flu virus



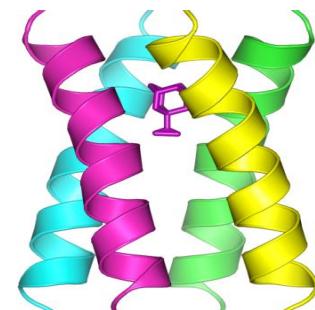
M2 channel



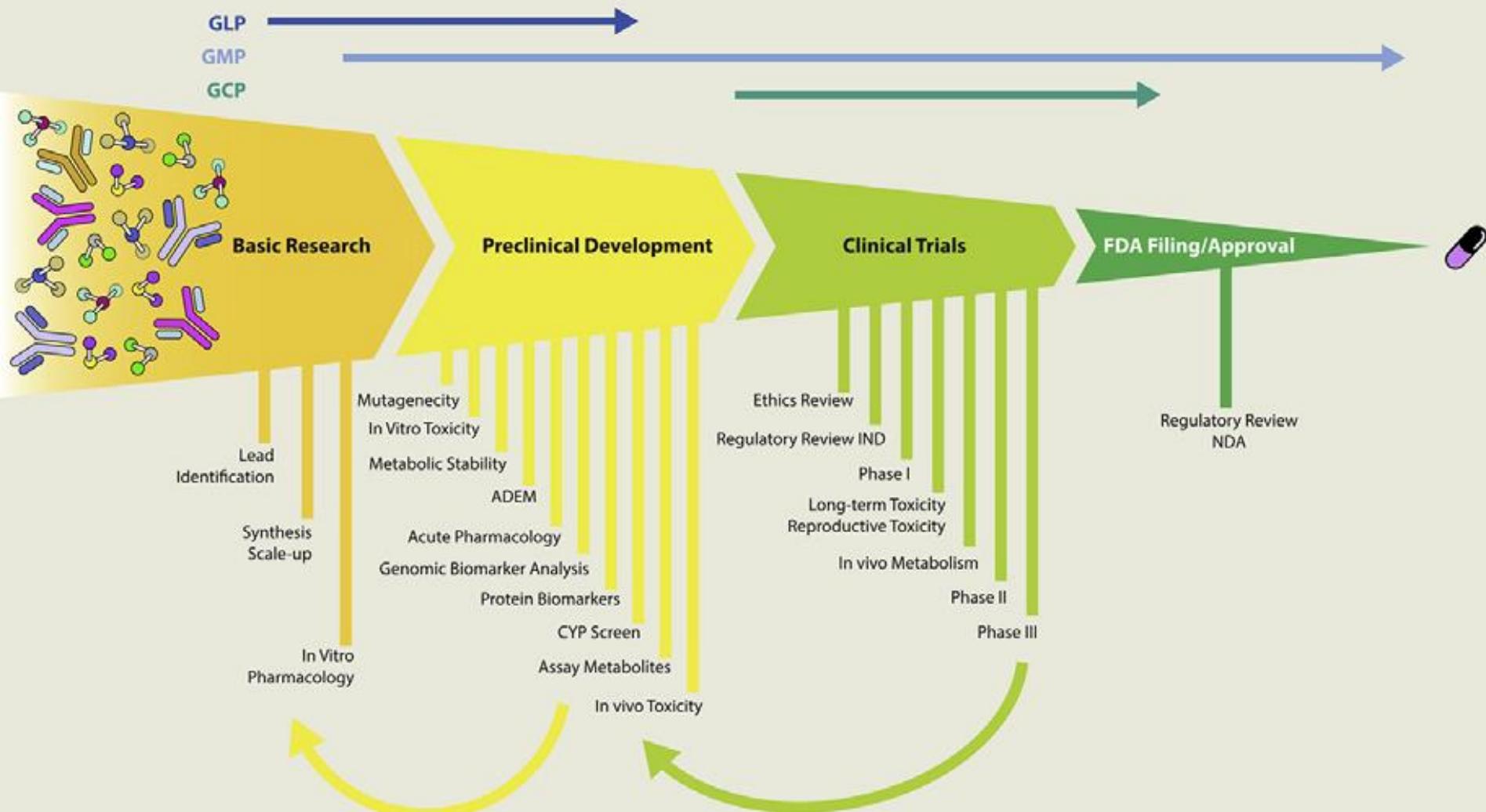
Amantadine



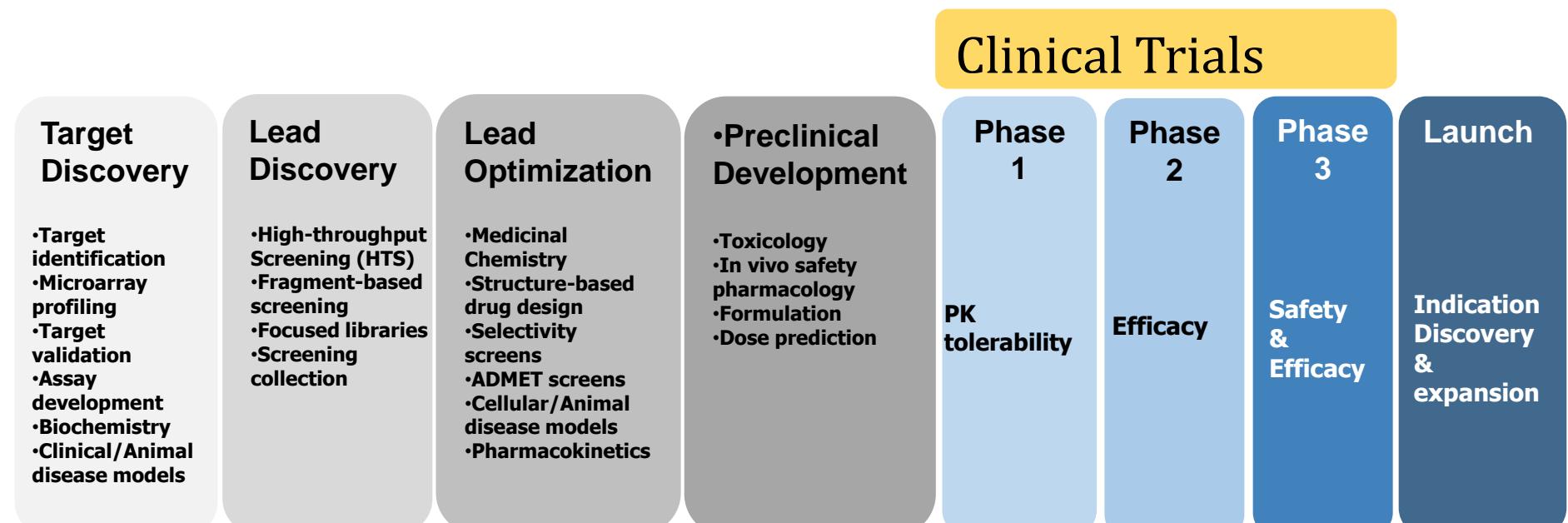
M2-A complex



# Drug Discovery and Development Activities



# Drug Discovery Process (simplified)



>450,000 distinct compounds  
~25,000 distinct lead series

~12,000 candidates

~1,200 drugs

Courtesy: John Overington

# The Drug Development Process: It's Long, Expensive and Risky

	Target to Hit	Hit to Lead	Lead Optim	Non-Clinical	Phase 1	Phase 2	Phase 3	Sub to Launch
# per Launch	24.3	19.4	14.6	12.4	8.6	4.6	1.6	1.1
P(TS)	80%	75%	85%	69%	54%	34%	70%	91%
Cycle time (yrs)	1.0	1.5	2.0	1.0	1.5	2.5	2.5	1.5
Cost/lau nch (\$mil)	\$94	\$166	\$414	\$150	\$273	\$319	\$314	\$48

> 450,000 distinct compounds  
~25,000 distinct lead series

~12,000 candidates

~1,200 drugs

Adapted from: SM Paul et al. Nature Reviews: Drug Discovery, 2010.

Costs are capitalized based on 11% cost of capital and in 2010 dollars.

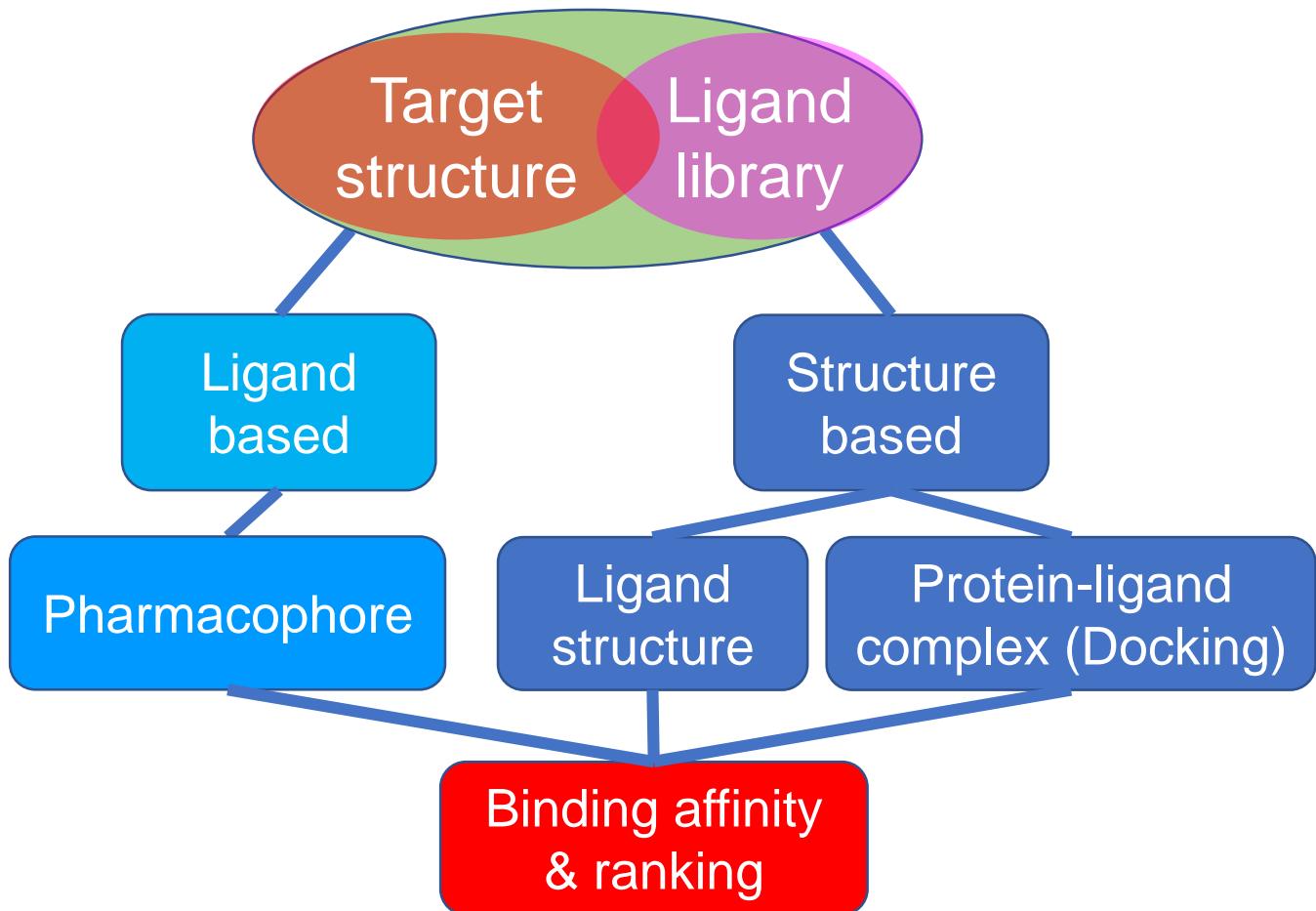
Adapted from John Overington

# **Molecular recognition**

**Molecular recognition** refers to the specific interaction between two or more molecules through noncovalent bonding such as hydrogen bonding, hydrophobic forces, van der Waals forces, metal coordination,  $\pi$ - $\pi$  interactions, halogen bonding, electrostatic and/or electromagnetic effects.

- Geometric complementarity
- Electrostatic complementarity

# Virtual Screening



Force field

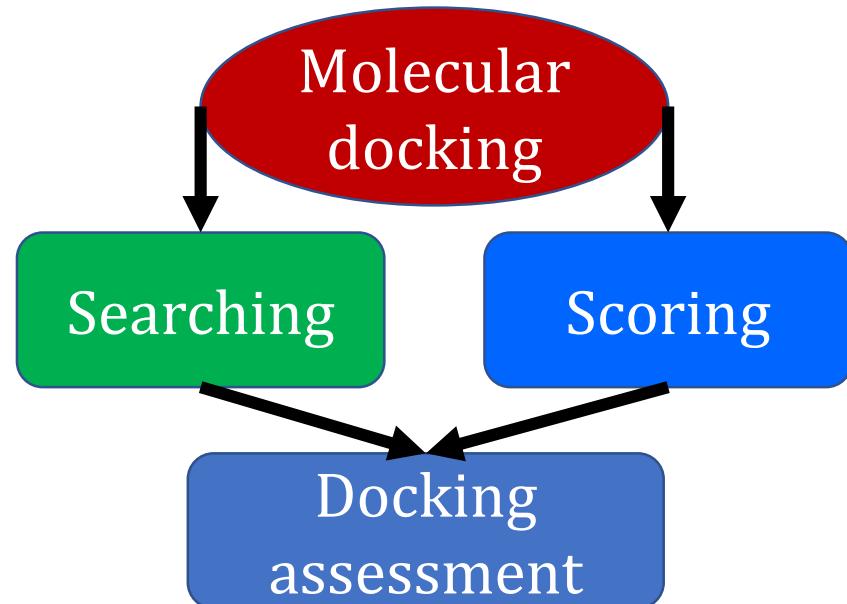
Empirical

Knowledge-based

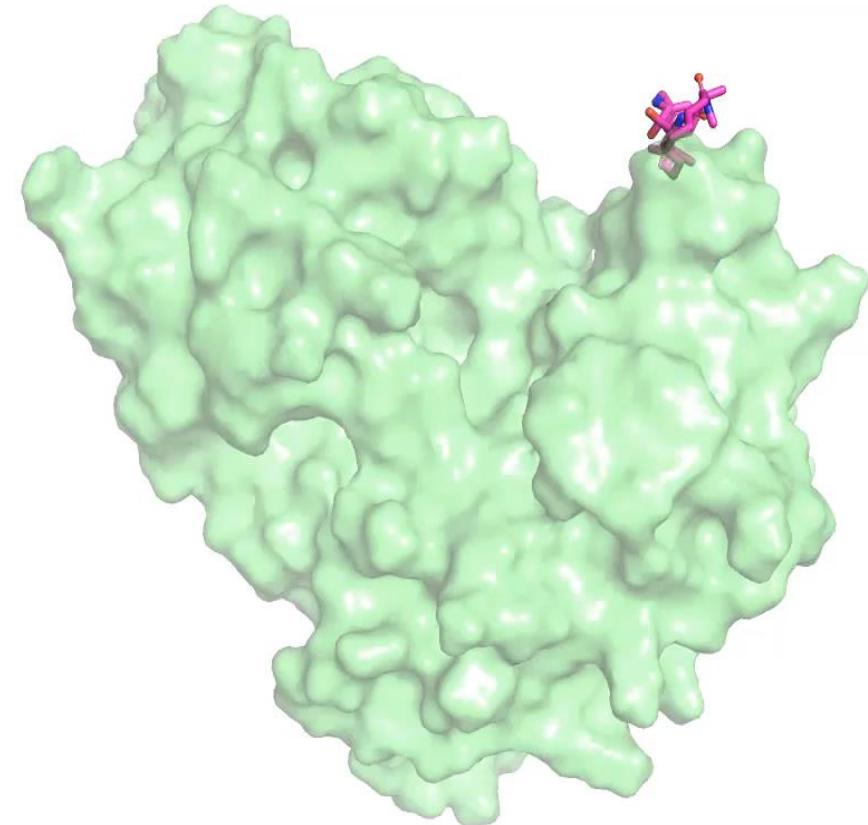
Machine learning

# Molecular docking

Docking is a process for searching the preferred position and orientation of one molecule to another one to form a stable complex.



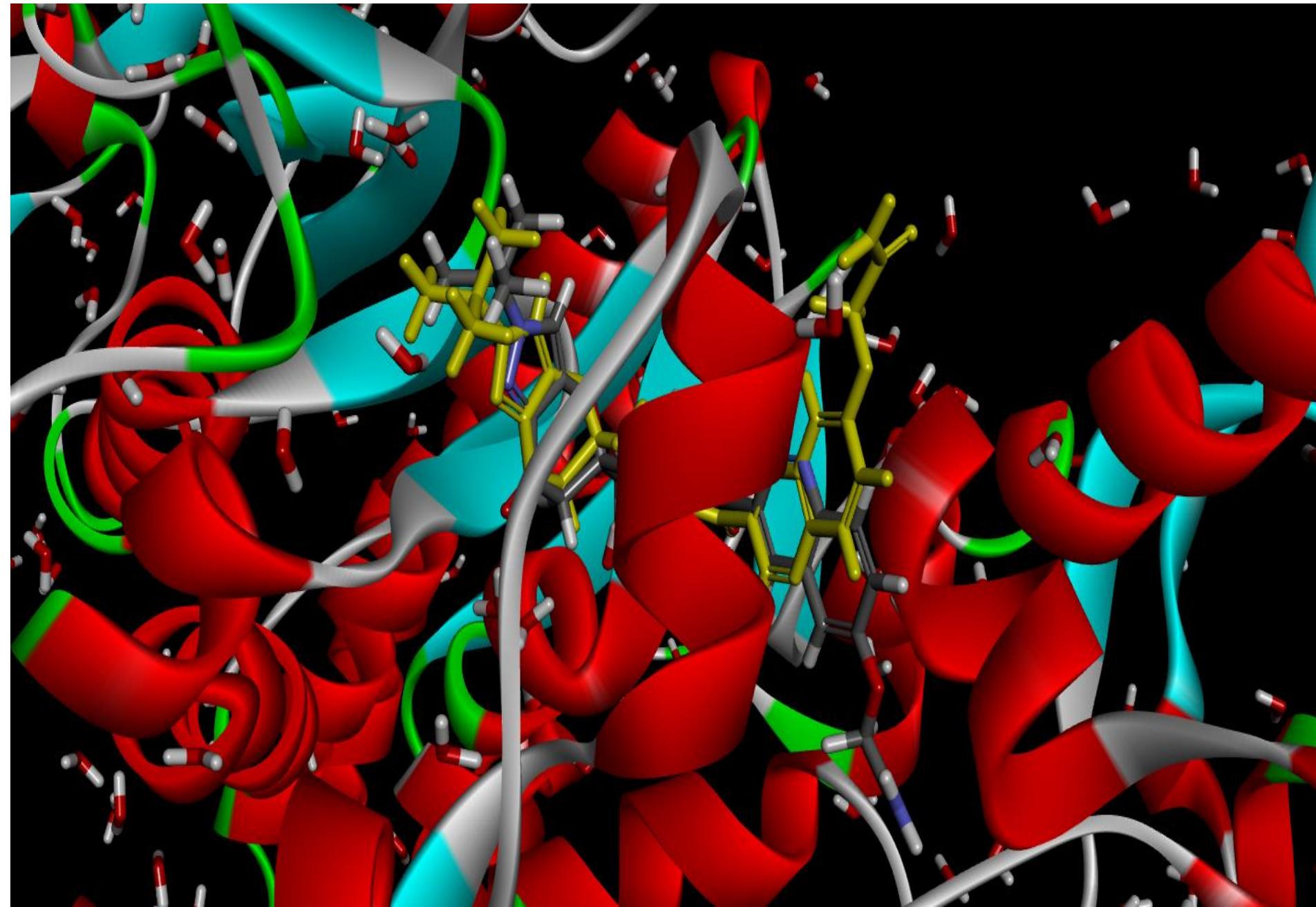
- Accuracy
- Enrichment factor
- Pharmac. prospective



**Docking a ligand to BACE**

(Kaifu Gao, Duc Nguyen, and Wei, 2019)

# Molecular docking



# Evaluation Metrics

- Pearson Correlation:  $R_{xy} = \frac{\sum_i(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i(x_i - \bar{x})^2} \sqrt{\sum_i(y_i - \bar{y})^2}}$
- RMSD:  $\text{RMSD}_{xy} = \sqrt{\frac{1}{n} \sum_i^n \|x_i - y_i\|^2}$
- Kendall's Tau:  $\tau = \frac{\# \text{ of concordeant pairs} - \# \text{ of discordant pairs}}{\# \text{ of all possible pairs}}$
- Area under enrichment curve (AUEC):
  - N active compounds.
  - Y% of screened compounds. It is used as variable in enrichment curve (EC) analysis.
  - Z% of retrieved active compounds =  $\frac{\# \text{ of active compounds among Y \% screened compounds}}{N} \times 100$
  - The pair, {Y,Z}, forms an EC
  - AUC=area under the EC
- Enrichment Factor (EF) =  $\frac{Z}{Y}$

## ***De novo* design and lead optimization**

***De novo* design:** design small/macro molecules that bind to a target (protein or DNA) from scratch. It searches drug chemical space. Generative adversarial networks (GAN) can be designed for the task.

**Lead optimization:** optimize screening hits to improve their binding affinity and selectivity.

# Drug optimization (antifungal drugs)

## Imidazole

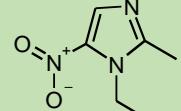


### Prototype



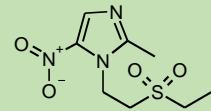
Streptomyces  
natural product  
trichomonacidal  
'toxic'

### 1<sup>st</sup> generation



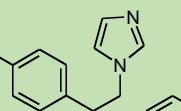
Metronidazole 1962

### 2<sup>nd</sup> generation



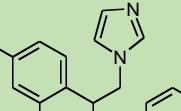
Tinidazole 1970

### 3<sup>rd</sup> generation



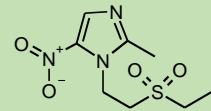
Clotrimazole 1970

### 4<sup>th</sup> generation



Econazole 1972

### 2<sup>nd</sup> generation

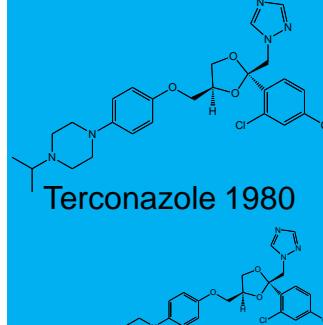


Ketoconazole 1978

## triazole



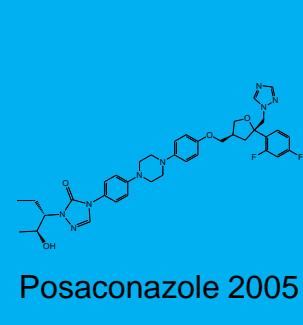
### 3<sup>rd</sup> generation



Terconazole 1980

Itraconazole 1984

### 4<sup>th</sup> generation



Posaconazole 2005

Fluconazole 1988



Voriconazole 2002

Fosfluconazole 2004

Bifonazole 1981

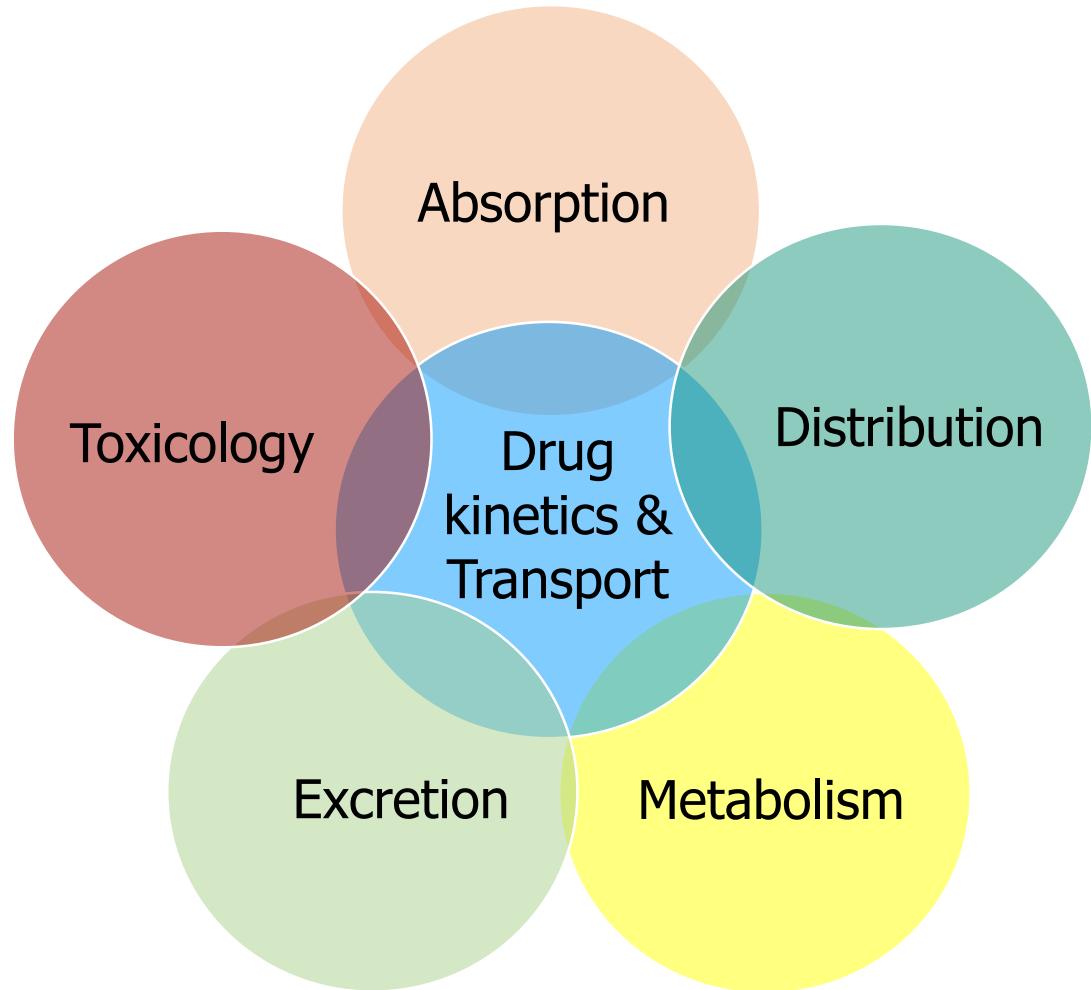
# Pharmacokinetics

Pharmacokinetics (PK) concerns how an organism affects a drug, whereas pharmacodynamics (PD) is the study of how the drug affects the organism. Drug kinetics and transport can be modeled by ``ADMET'' in organism.

Pharmacokinetics is all about mathematical modeling. The drug concentration in the  $\alpha$ th compartment (heart, liver, brain, lung, tissue, etc):

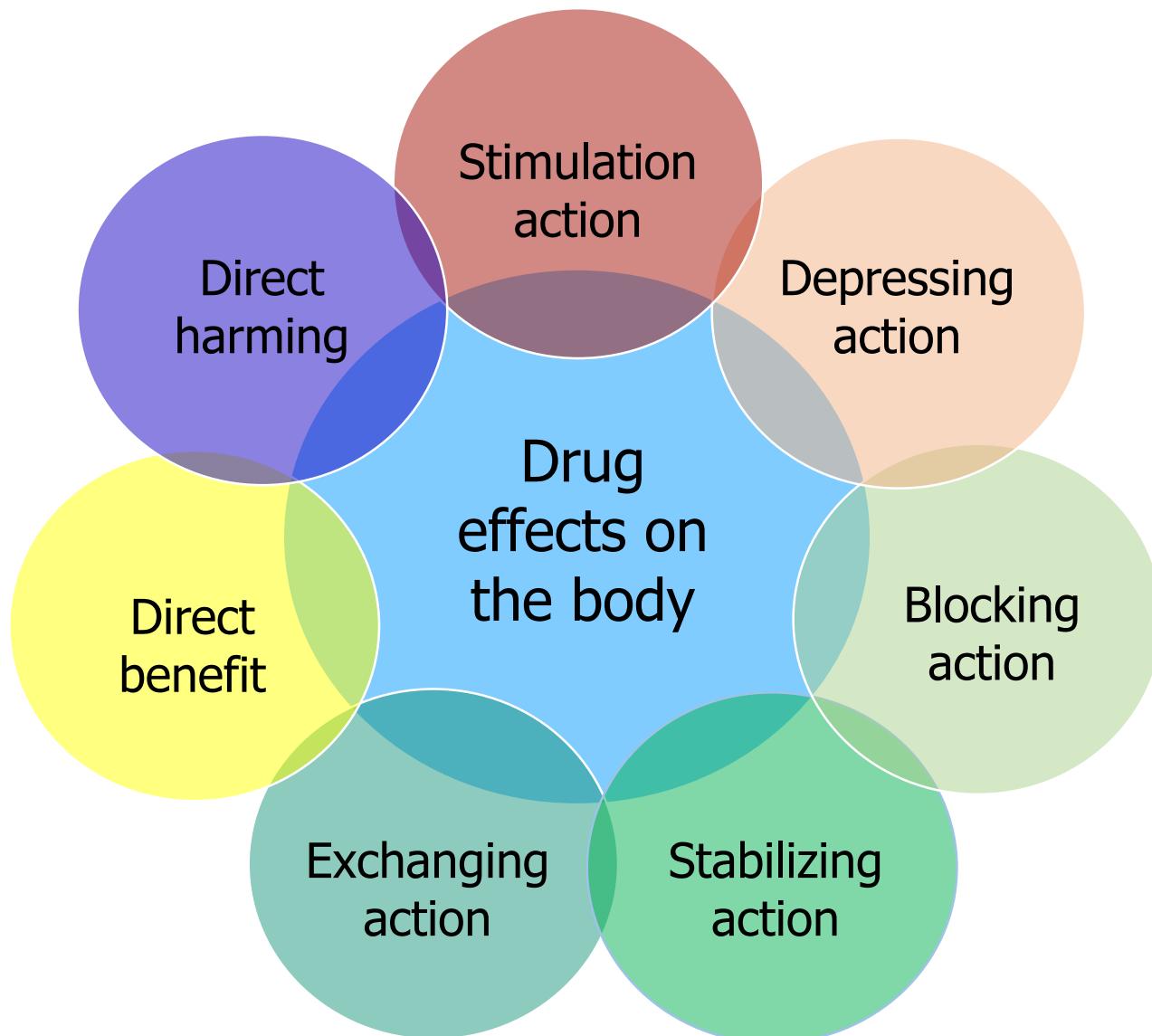
$$\begin{aligned}\frac{\partial C_\alpha}{\partial t} + \nabla \cdot C_\alpha \boldsymbol{v} \\ = \nabla \cdot (D_\alpha \nabla C_\alpha) + R_\alpha\end{aligned}$$

where  $R_\alpha$  includes all production and sink.



# Pharmacodynamics

Pharmacodynamics (PD) is the study of how the drug affects the organism. PK & PD together influence dosing, benefit, and adverse effects.



# Pharmacodynamics



## Desirable activities:

- Cellular membrane disruption
- Chemical reaction with downstream effects
- Interaction with enzyme proteins
- Interaction with structural proteins
- Interaction with carrier proteins
- Interaction with ion channels
- Ligand binding to receptors:
  - Hormone receptors
  - Neuromodulator receptors
  - Neurotransmitter receptors

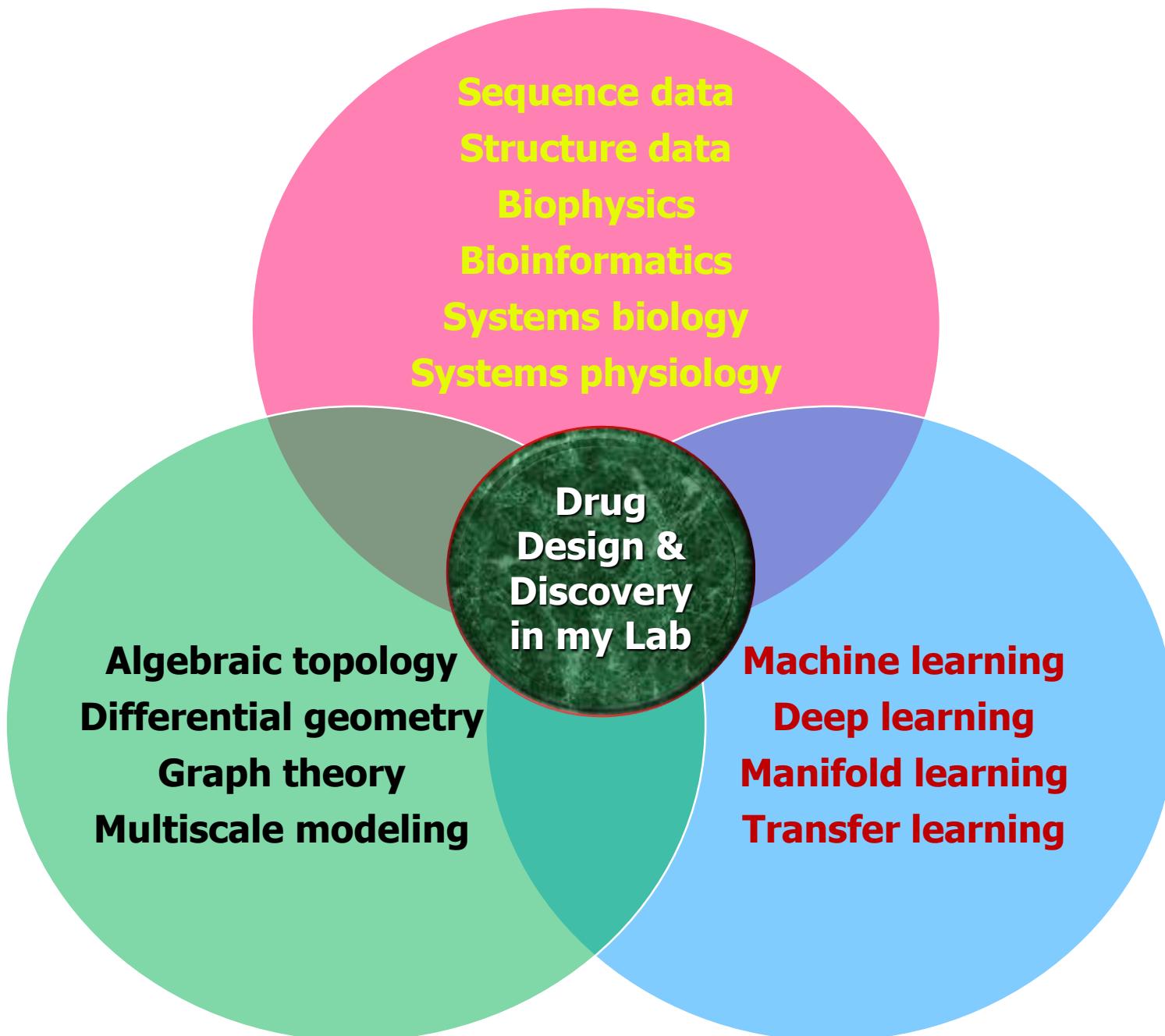
## Undesirable effects

- Increased probability of cell mutation (carcinogenic activity)
- A multitude of simultaneous assorted actions which may be deleterious
- Interaction (additive, multiplicative, or metabolic)
- Induced physiological damage, or abnormal chronic conditions

# **Quantitative systems pharmacology**

**Quantitative systems pharmacology** (QSP) is a new discipline that concerns mathematical, computational, biological, and physiological modeling of organism systems, disease processes and drug pharmacology.

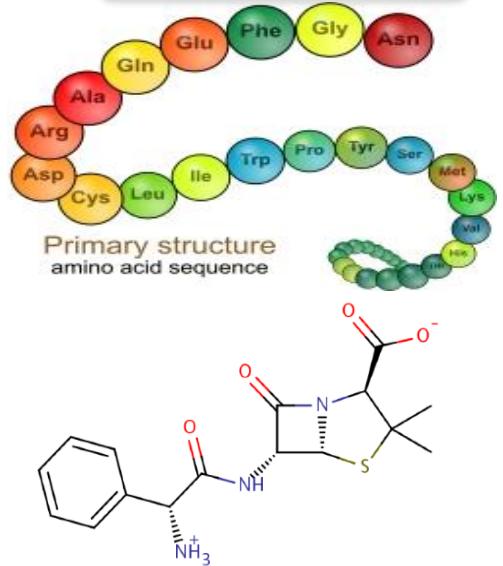
- The mechanisms of drug pharmacokinetics (PK) and pharmacodynamics (PD).
- Using ODEs.
- QSP can generate biological/pharmacological hypotheses *in-silico* to aid the design of in-vitro or in-vivo clinical experiments.
- QSP is a required component for clinical pharmacology review by FDA.



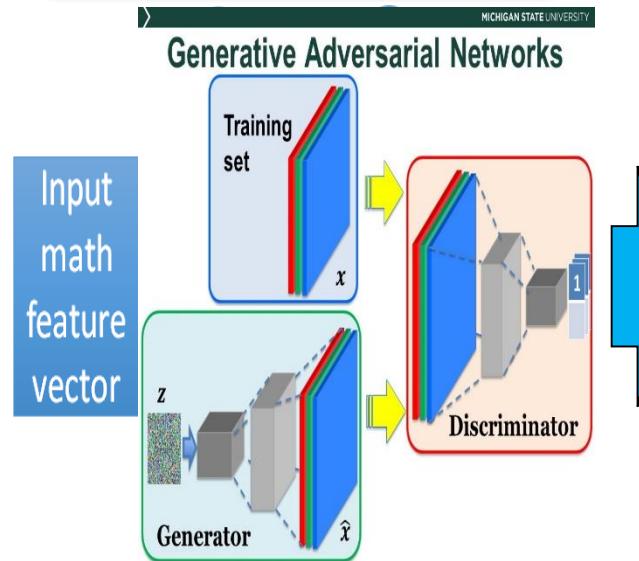
# Drug Design Data Resource (D3R) Grand Challenge



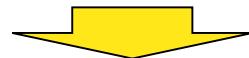
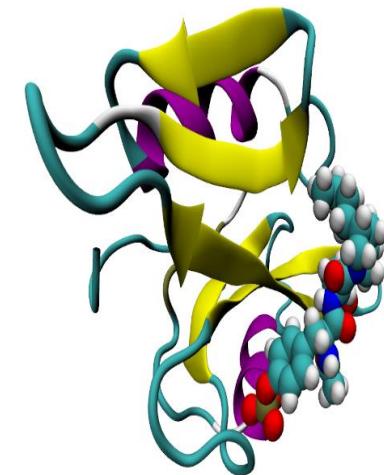
Given data



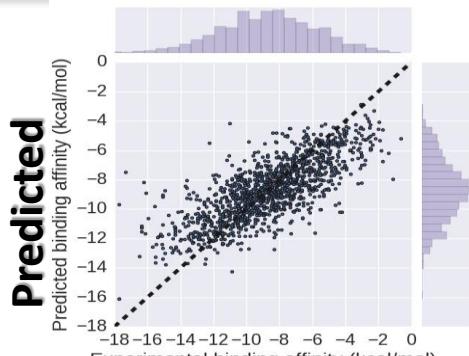
Math based GAN



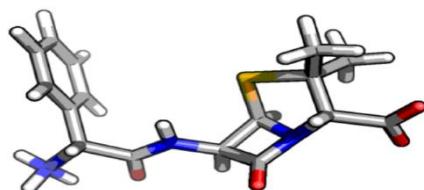
Predicted complex



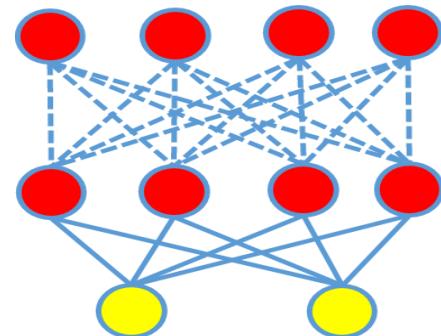
Final predictions to be compared with experiments



Experimental

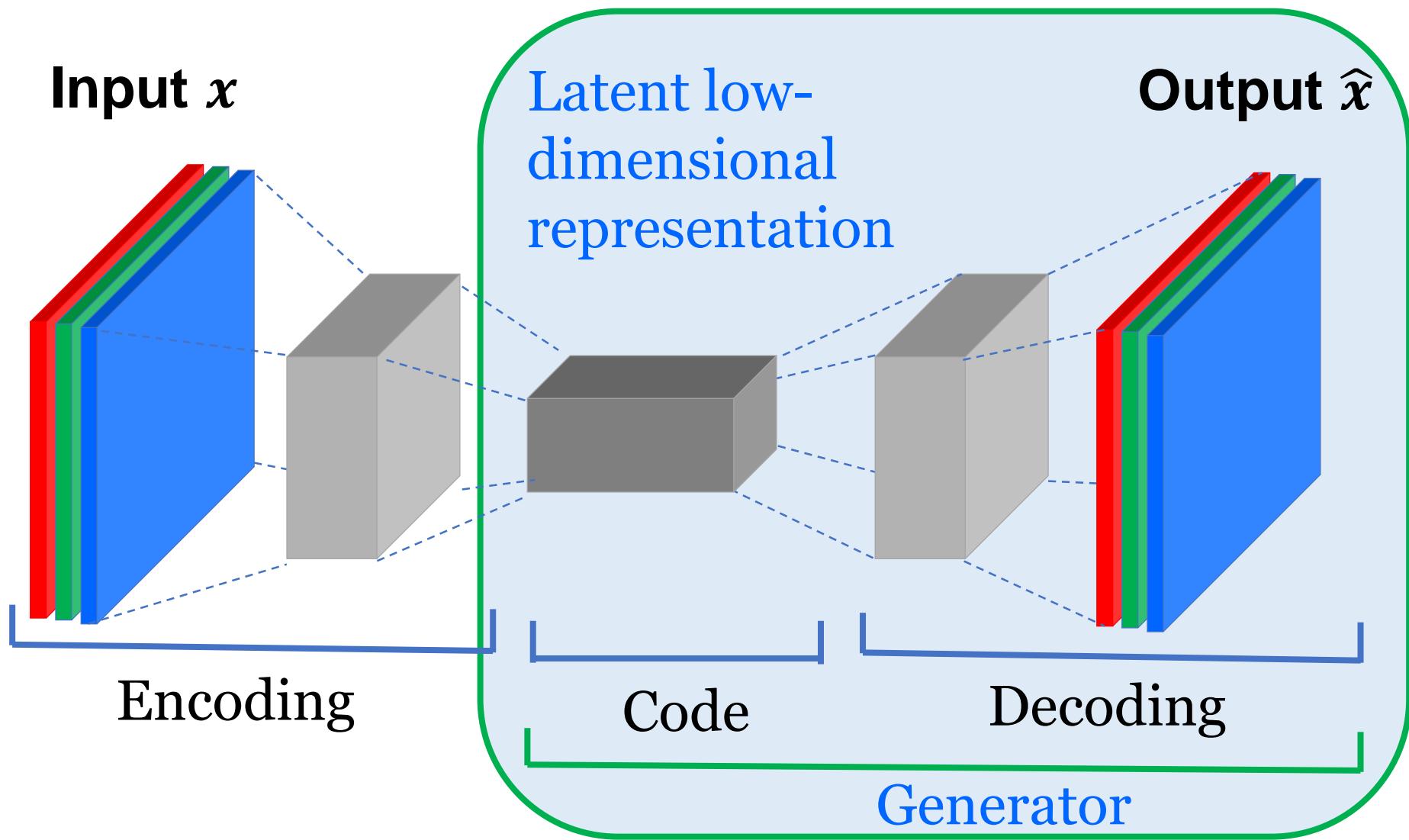


Drug pose



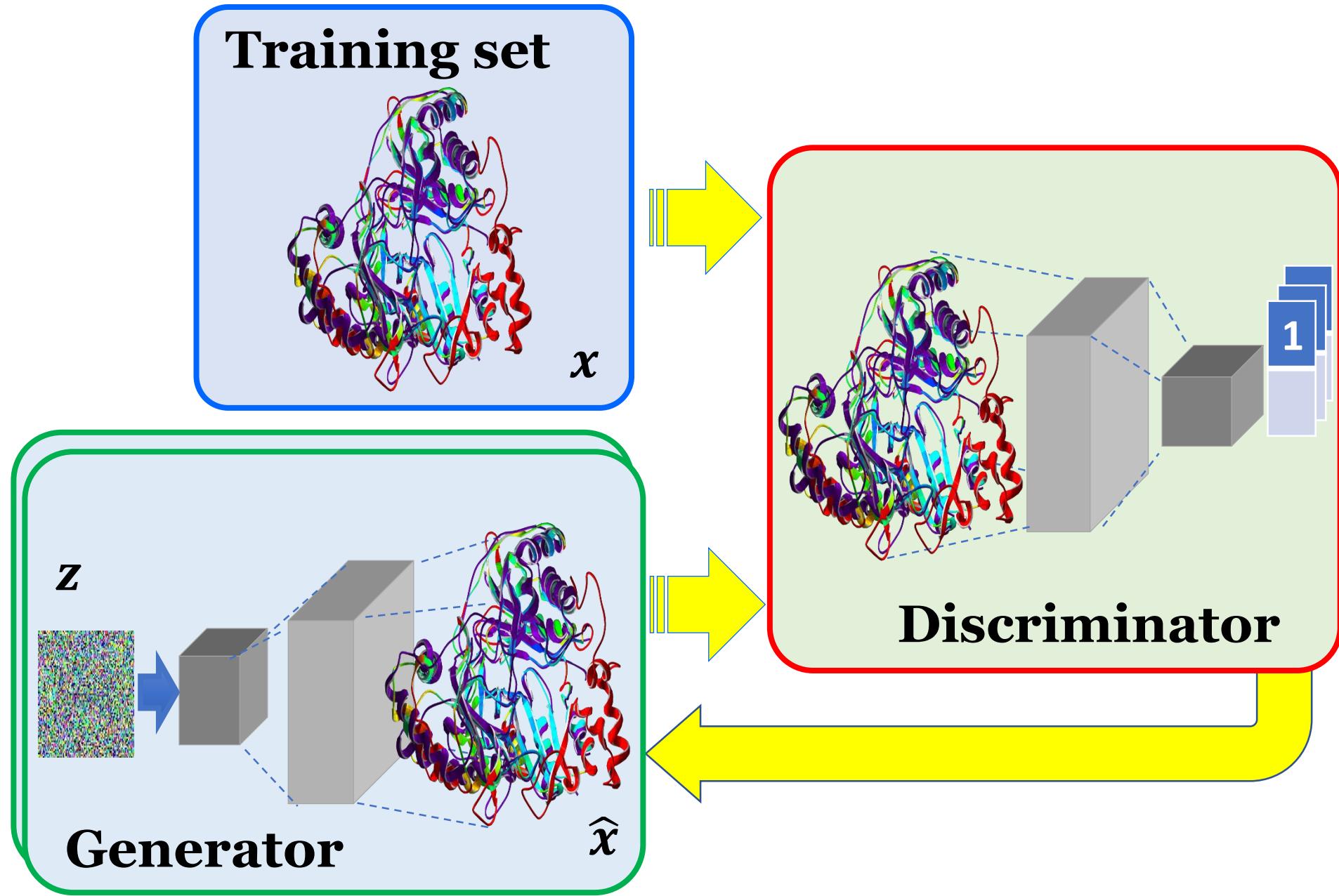
(Nguyen et al, JCAMD, 2018)

# Autoencoder

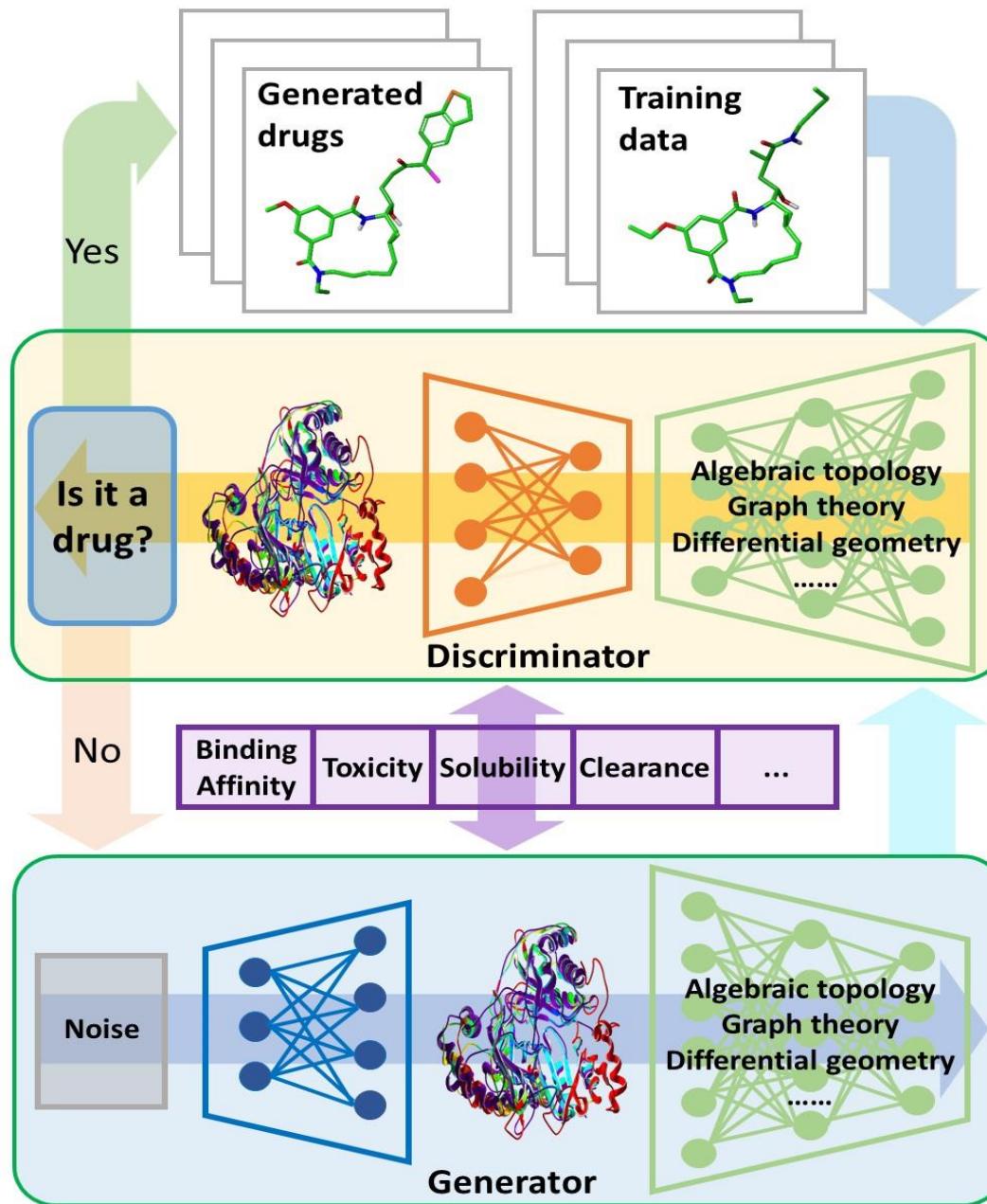


**Lost function:**  $L = \frac{1}{nM} \sum_i^M \sum_j^n \left( x_j^{(i)} - \hat{x}_j^{(i)} \right)^2$

# Generative Adversarial Networks for Drug Design



# Conditional Mathematical GAN for Drug Design



**GAN Variants:**  
Wasserstein GAN  
Conditional GAN  
L2 GAN  
Relaxed WGAN  
Fisher GAN  
Banach GAN  
Gromov-Hausdorff  
...

# D3R Grand Challenge 2 (2016-2017)



**Given:** Farnesoid X receptor (FXR) and 102 ligands

**Tasks:** Dock 102 ligands to FXR, and predict their poses, binding free energies and energy ranking

## Stage 1

[Pose Predictions \(partials\)](#)

[Scoring \(partials\)](#)

[Free Energy Set 1 \(partials\)](#)

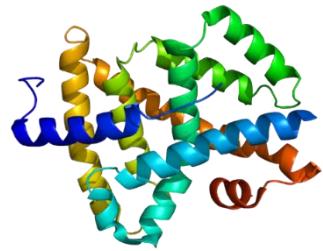
[Free Energy Set 2 \(partials\)](#)

## Stage 2

[Scoring \(partials\)](#)

[Free Energy Set 1 \(partials\)](#)

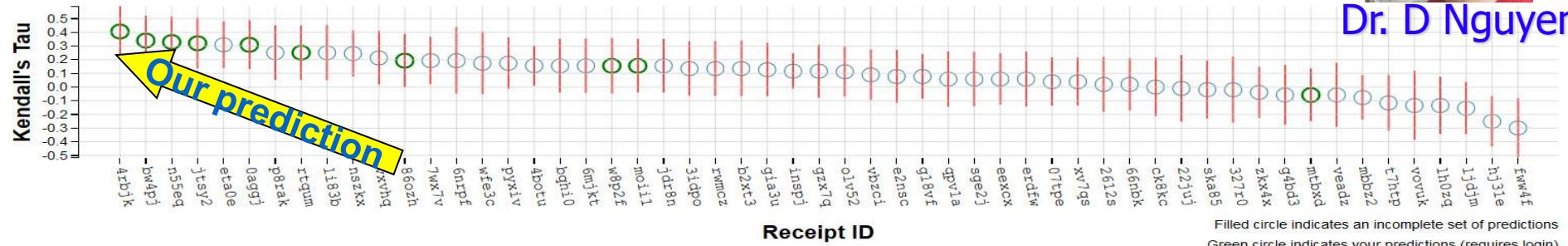
[Free Energy Set 2 \(partials\)](#)



Dr. D Nguyen

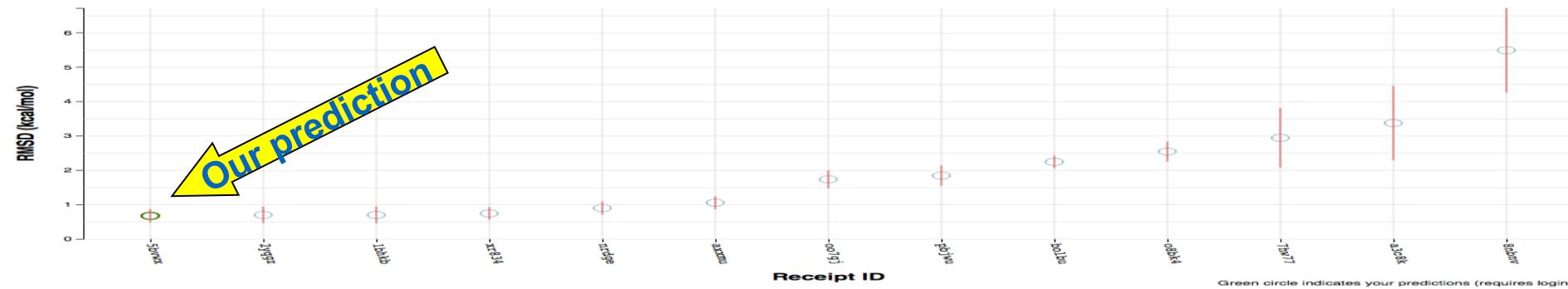
## Grand Challenge 2

Free Energy Set 1 (Stage 2) - Kendall's Tau



## Grand Challenge 2

Free Energy Set 1 (Stage 1) - RMSD



# D3R Grand Challenge 3 (2017-2018)

(Nguyen et al, JCAMD, 2018)



## Pose Prediction

### Cathepsin Stage 1A

[Pose Predictions \(partials\)](#)

### Affinity Rankings excluding Kds > 10 μM

### Cathepsin Stage 1

[Scoring \(partials\)](#)

[Free Energy Set](#)

### VEGFR2

[Scoring \(partials\)](#)

### JAK2 SC3

[Scoring](#)

[Free Energy Set](#)



### Active / Inactive Classification

### VEGFR2

[Scoring \(partials\)](#)

### JAK2 SC3

[Scoring](#)

[Free Energy Set](#)



### Affinity Rankings for Cocrystallized Ligands

### Cathepsin Stage 1

[Scoring \(partials\)](#)

[Free Energy Set](#)



### Cathepsin Stage 1B

[Pose Prediction](#)

### Cathepsin Stage 2

[Scoring \(partials\)](#)

[Free Energy Set](#)

### JAK2 SC2

[Scoring \(partials\)](#)

### TIE2

[Scoring](#)

[Free Energy Set 2](#)



### JAK2 SC2

[Scoring \(partials\)](#)

### TIE2

[Scoring \(partials\)](#)

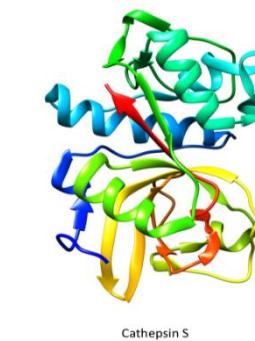
[Free Energy Set 1](#)



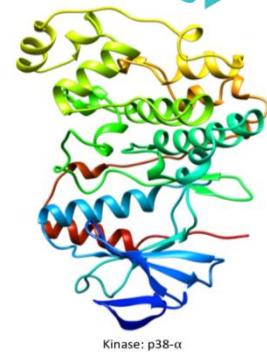
### Cathepsin Stage 2

[Scoring \(partials\)](#)

[Free Energy Set](#)



Cathepsin S



Kinase: p38- $\alpha$

### p38- $\alpha$

[Scoring](#)

### ABL1

[Scoring \(partials\)](#)



### p38- $\alpha$

[Scoring \(partials\)](#)

### ABL1

[Scoring \(partials\)](#)



Zixuan Cang



Dr. D Nguyen

# D3R Grand Challenge 4 (2018-2019)



## Pose Predictions

### BACE Stage 1A

#### Pose Predictions (Partials)



### BACE Stage 1B

#### Pose Prediction (Partials)



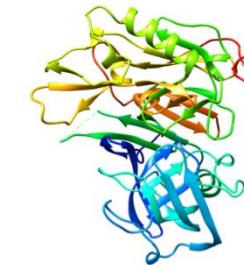
## Affinity Predictions

### Cathepsin Stage 1

#### Combined Ligand and Structure Based Scoring



Dr. Kaifu Gao Dr. D Nguyen



### Structure Based Scoring

### Free Energy Set



### BACE Stage 1

#### Combined Ligand and Structure (No participation)

#### Ligand Based Scoring (Partials) (No participation)

#### Structure Based Scoring (Partials) (No participation)

#### Free Energy Set (No participation)

### BACE Stage 2

#### Combined Ligand and Structure

#### Ligand Based Scoring (No participation)

#### Structure Based Scoring (Partials)

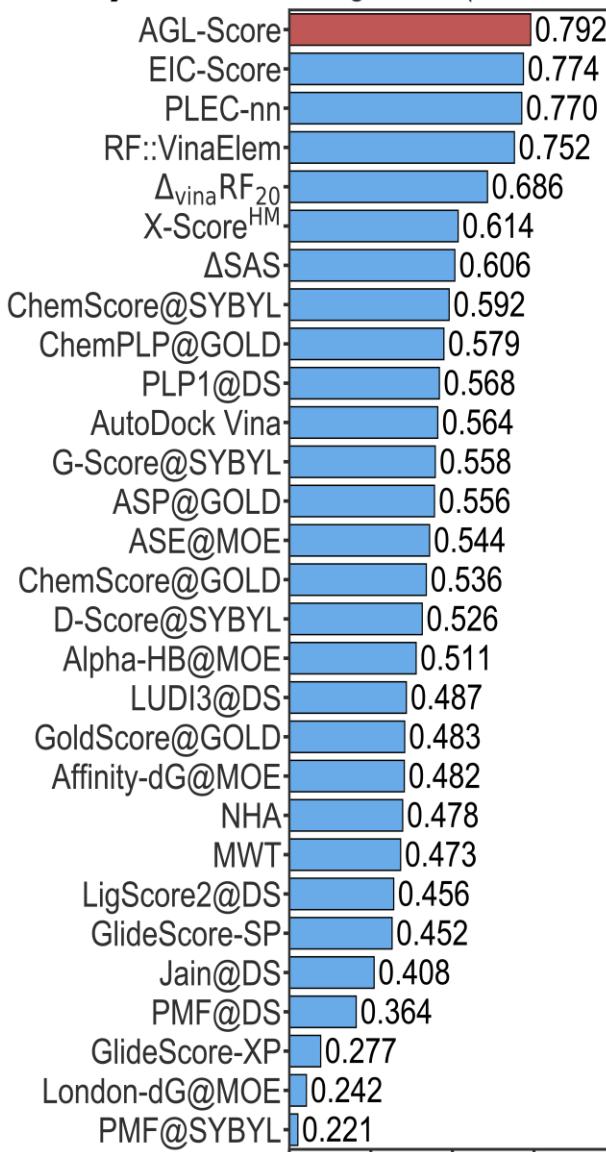


# Performance of algebraic graph learning (AGL-Score)

(Nguyen & Wei, JChem Inf Model, 2019)

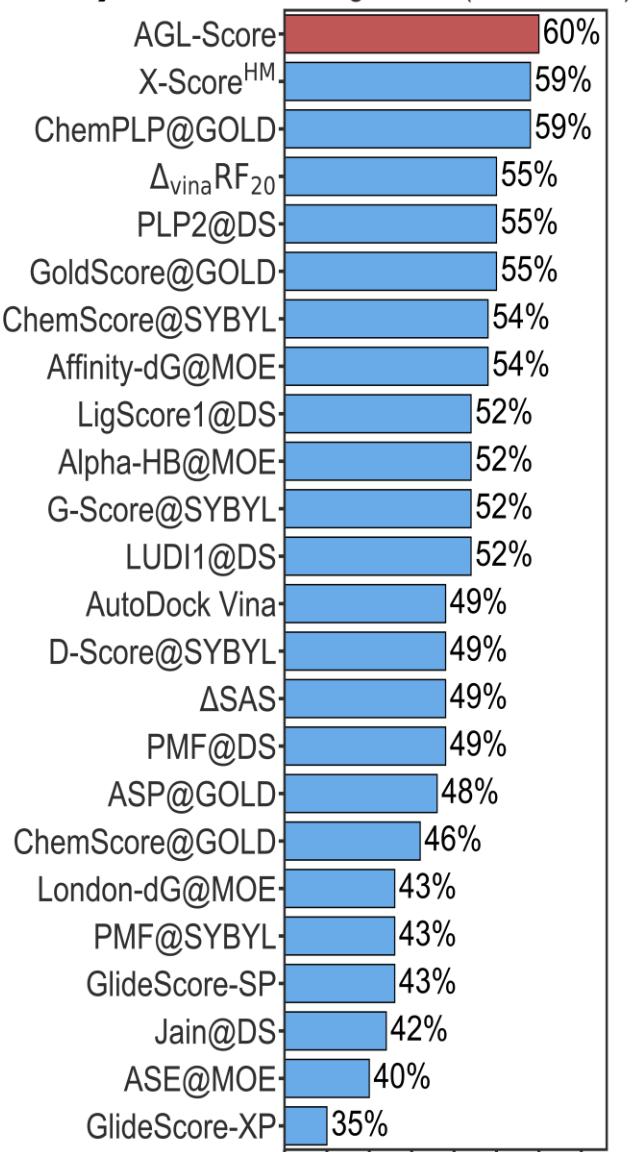
a)

Scoring Power (CASF-2013)

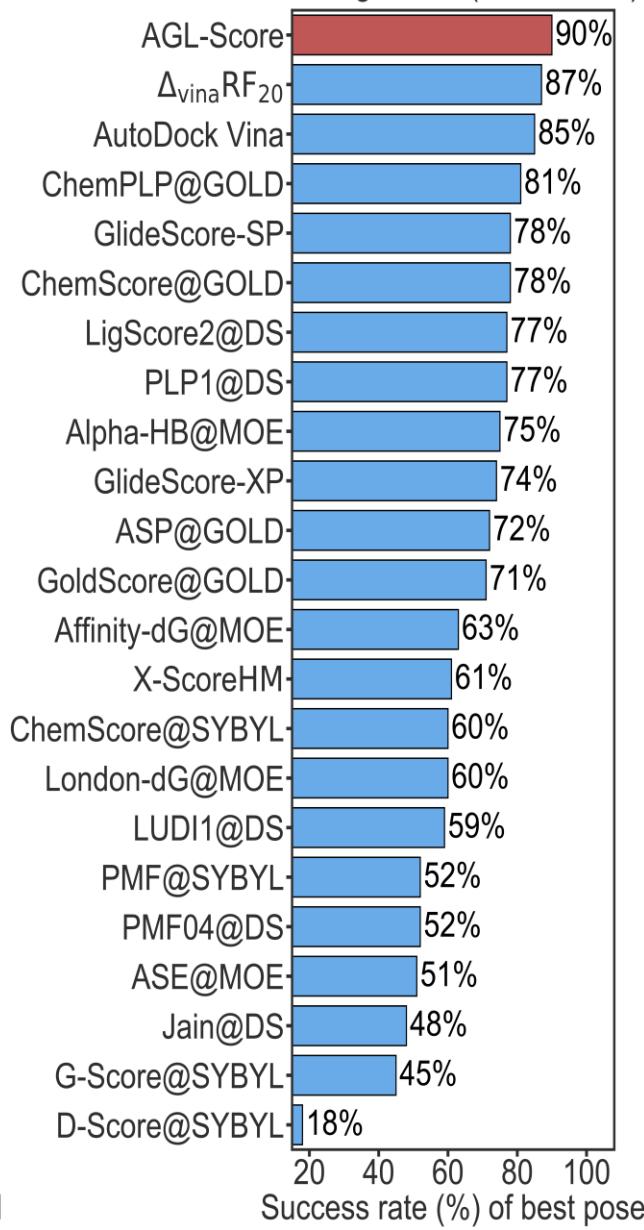


b)

Ranking Power (CASF-2013)



Docking Power (CASF-2013)



Pearson's R

Success rate (%) at high-level

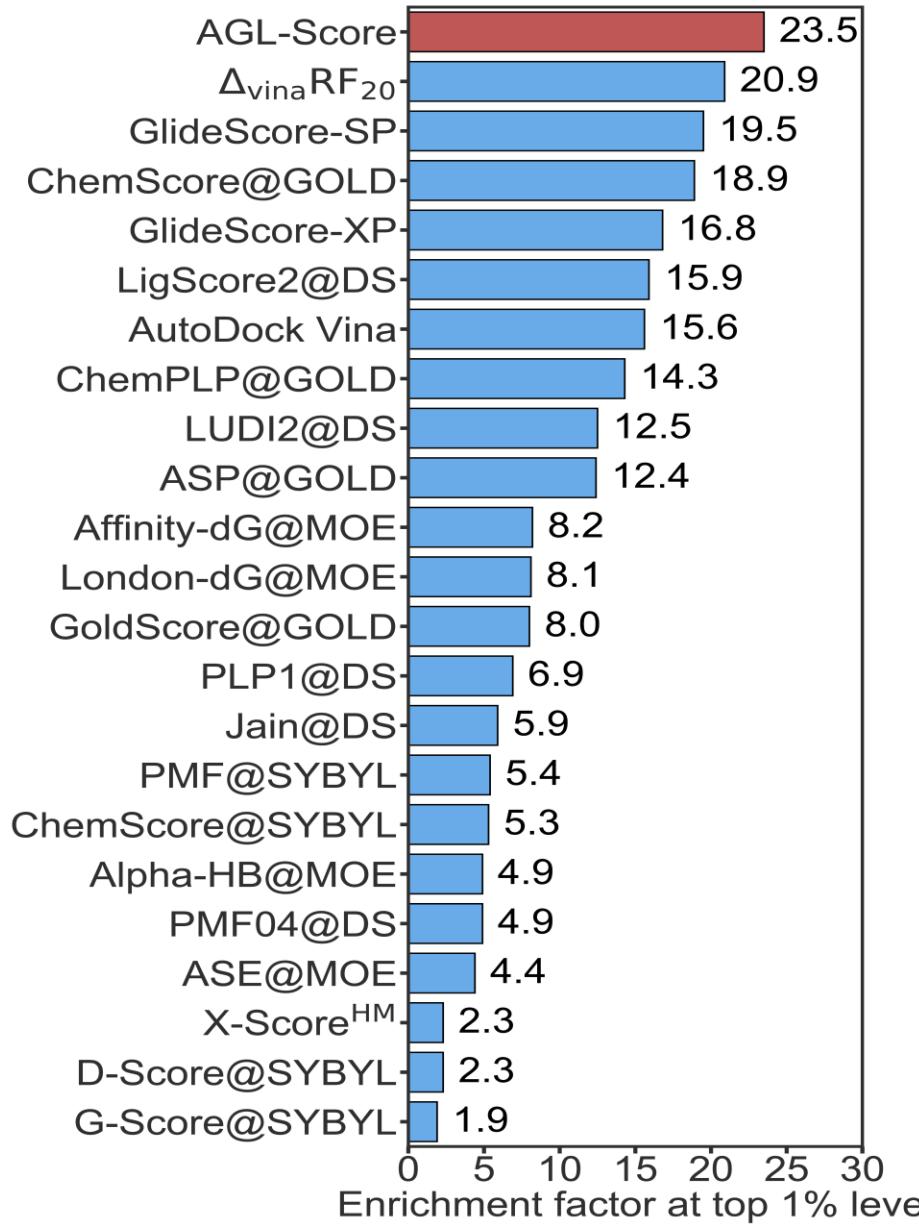
Success rate (%) of best pose

# Performance of algebraic graph learning (AGL-Score)

(Nguyen & Wei, JCIM, 2019)

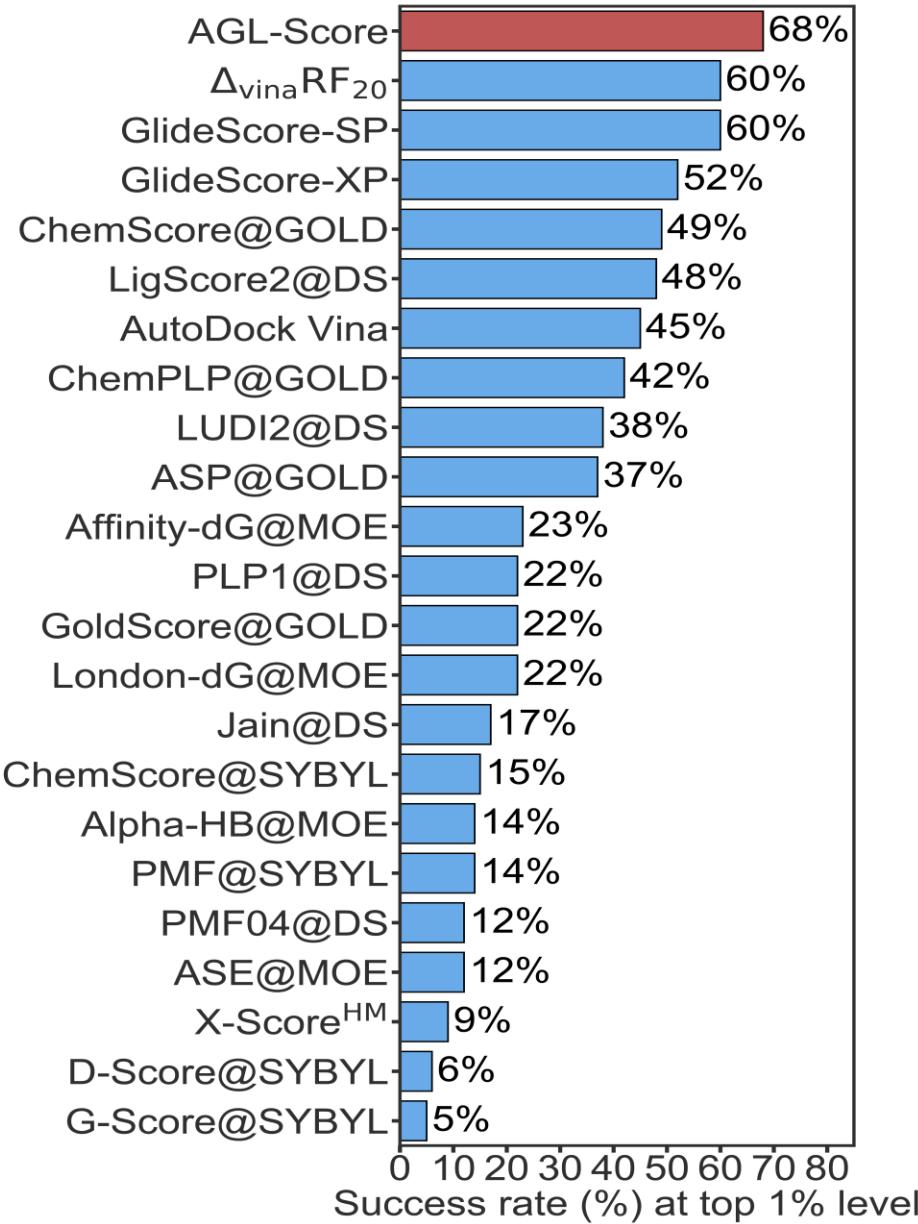
a)

Screening Power (CASF-2013)



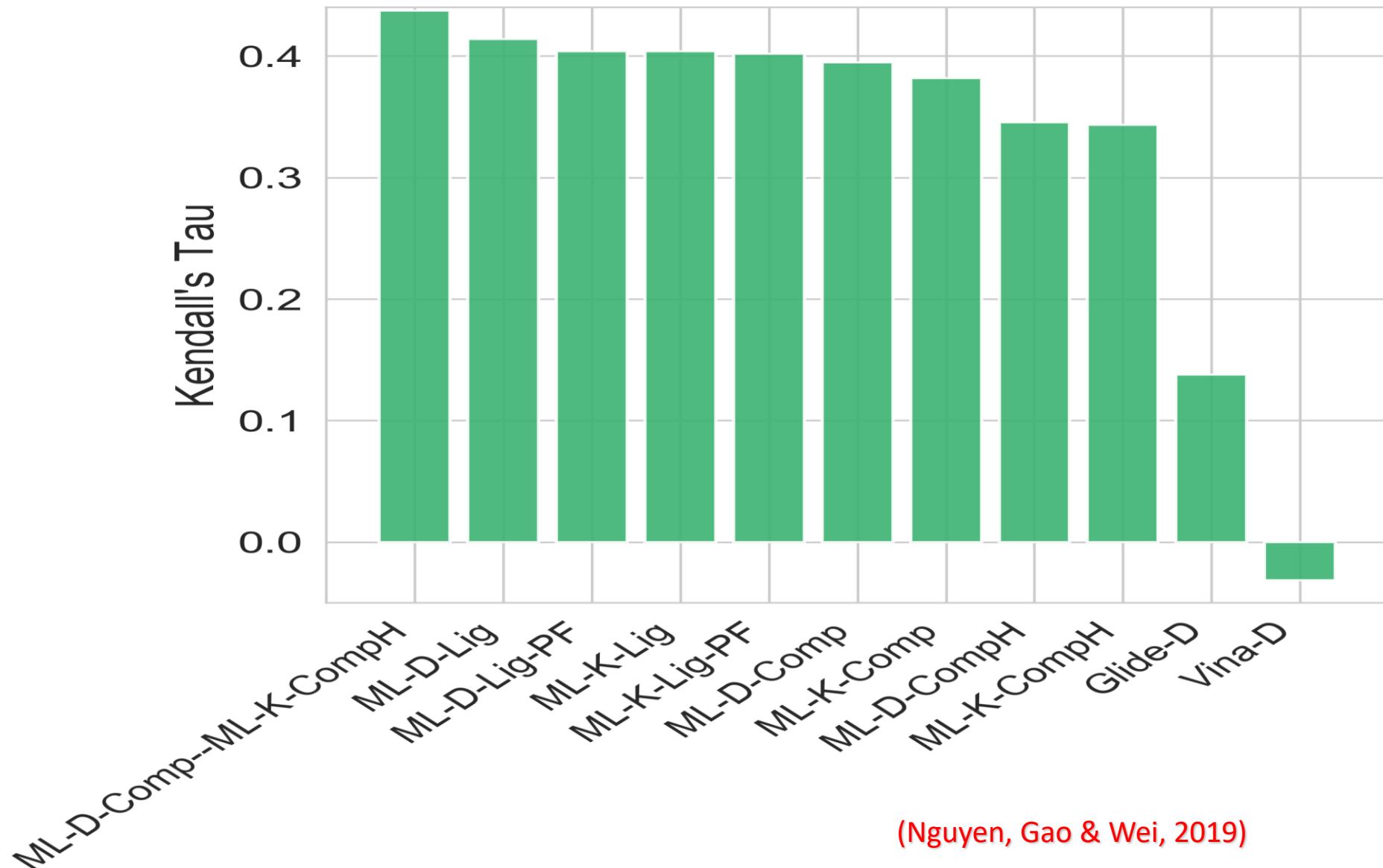
b)

Screening Power (CASF-2013)



# Best performance model vs. single models ( $\tau$ )

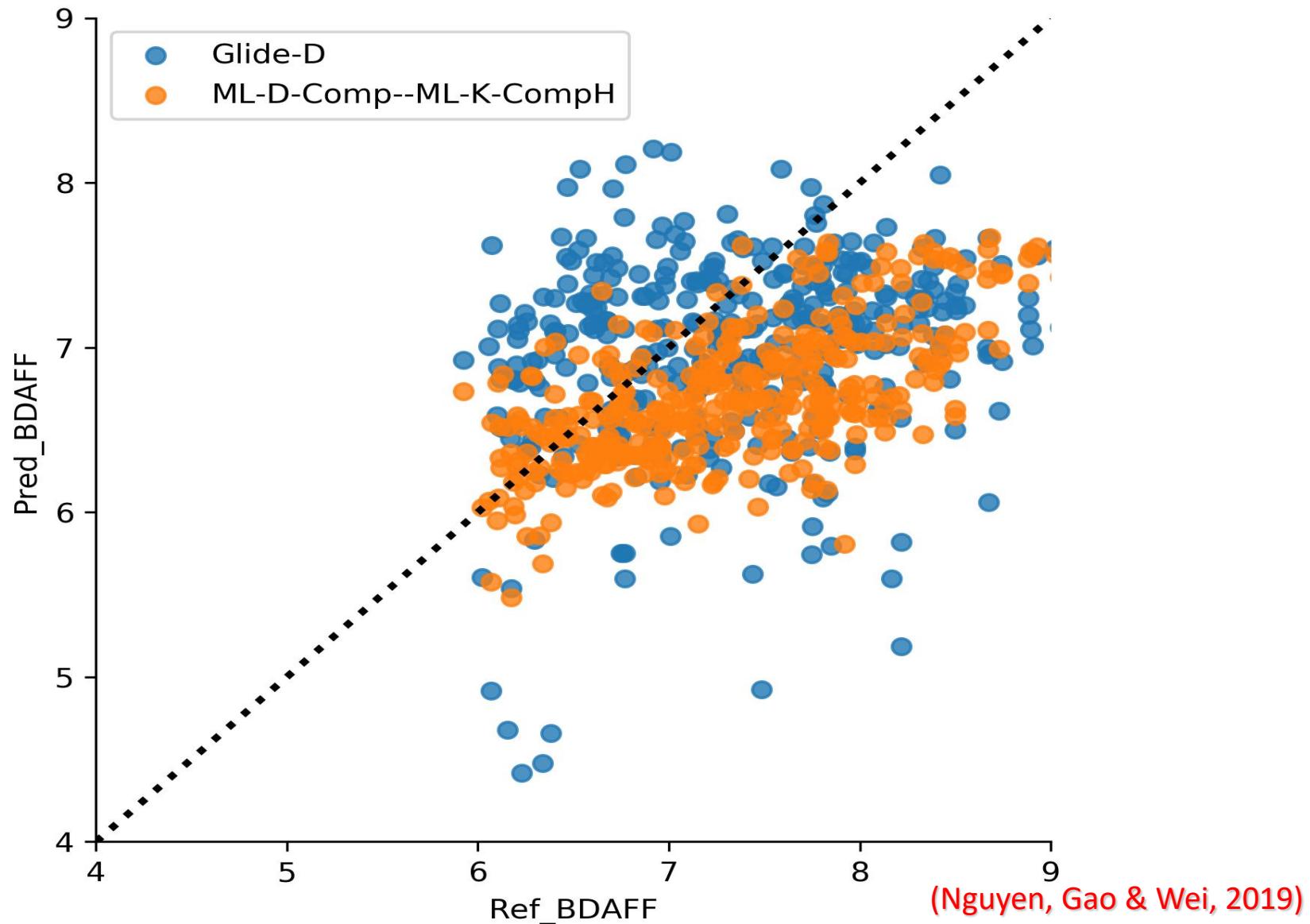
Rank binding affinities of 362 compounds (fully blind)



(Nguyen, Gao & Wei, 2019)

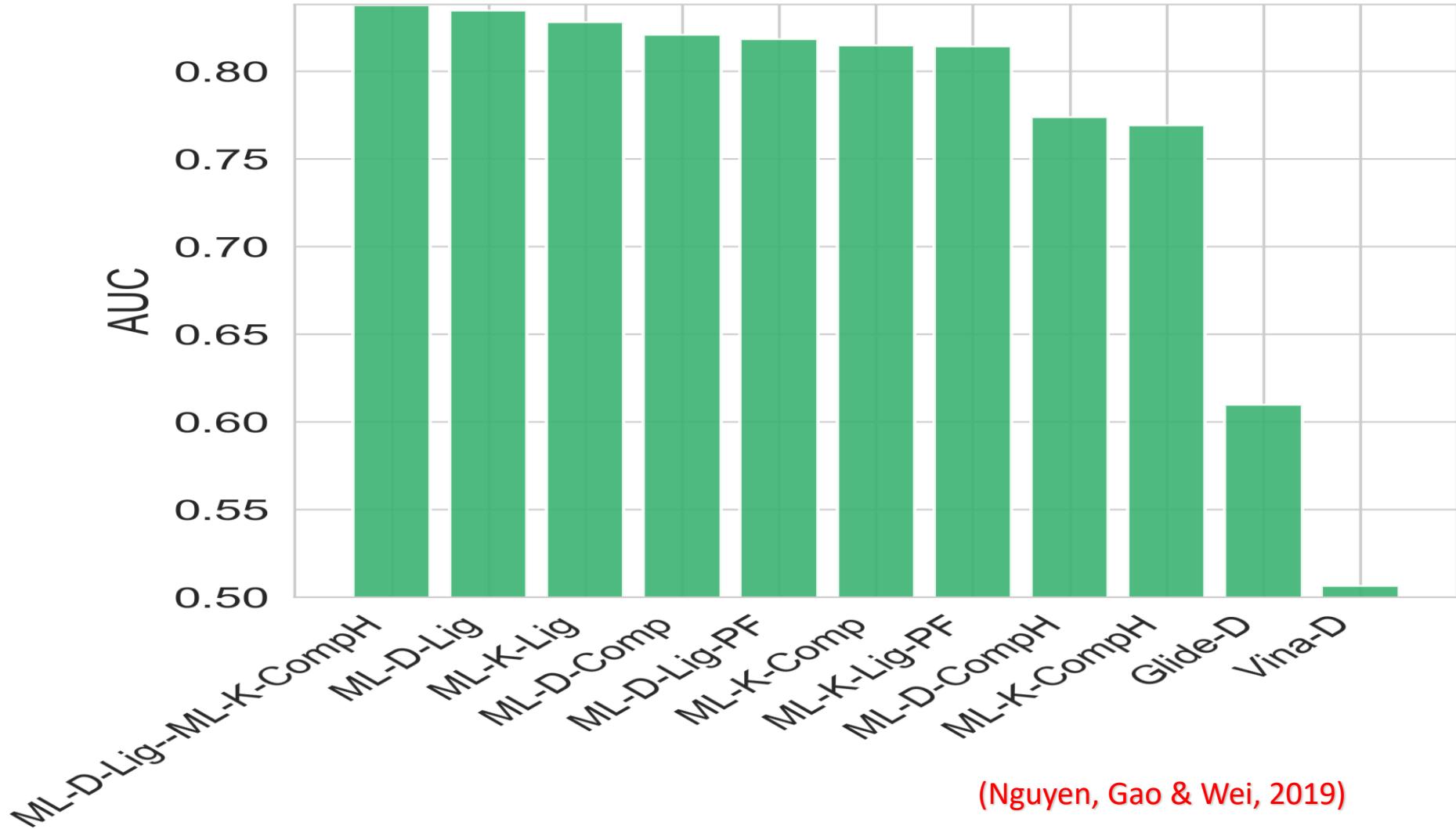
# ML model vs. software

Rank binding affinities of 362 compounds (fully blind)



# Best performance model vs. single models (AUEC)

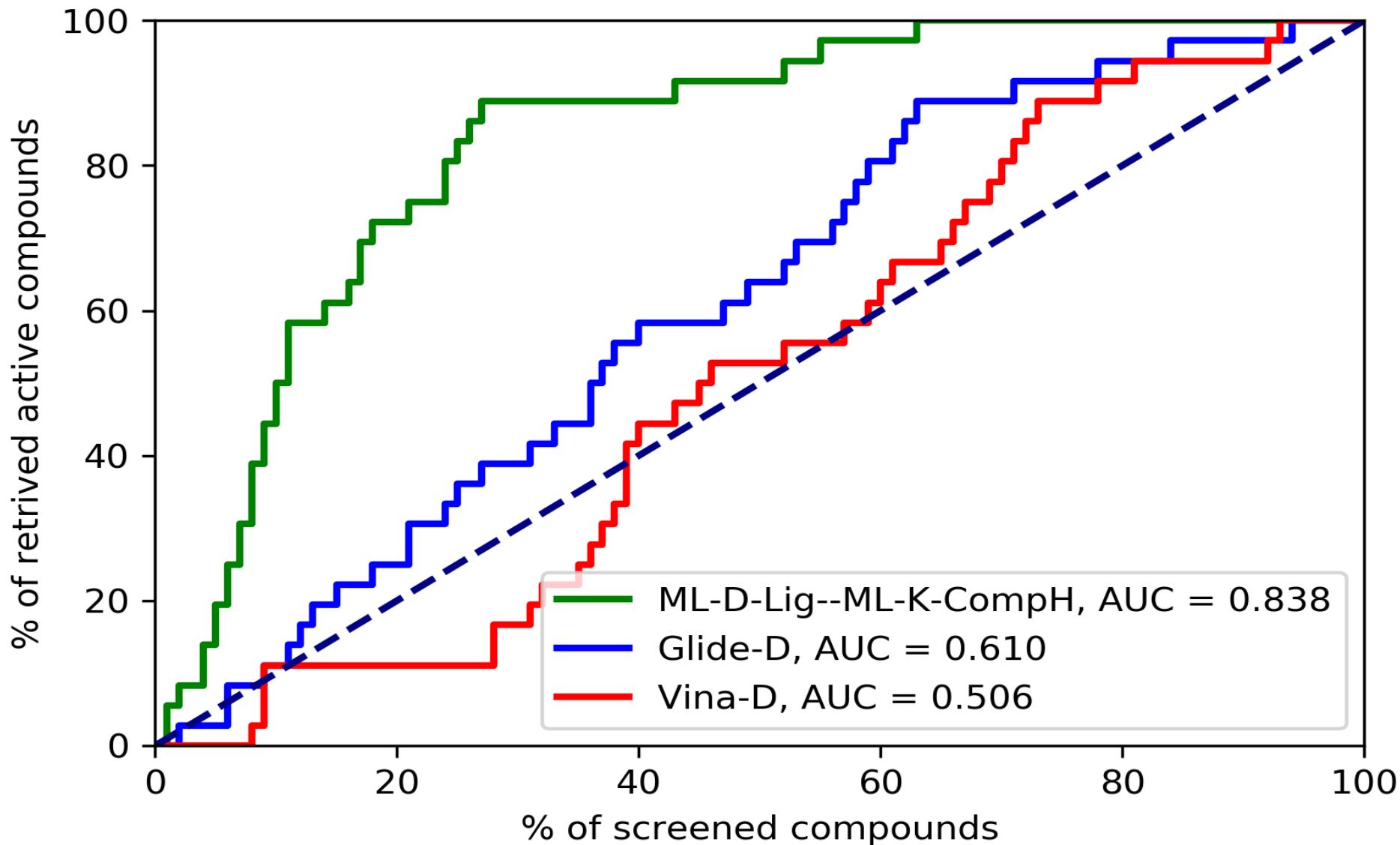
Rank binding affinities of 362 compounds (fully blind, top 10% are assumed to be ``active''.)



(Nguyen, Gao & Wei, 2019)

# Enrichment Curve

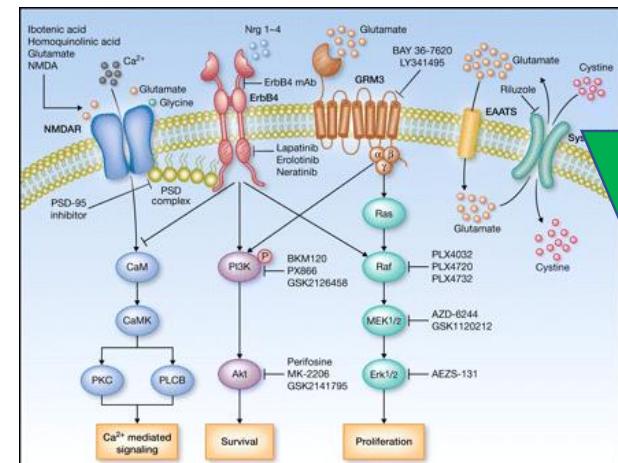
Rank binding affinities of 362 compounds (fully blind, top 10% are assumed to be “active”.)



# Quantitative systems pharmacology modeling

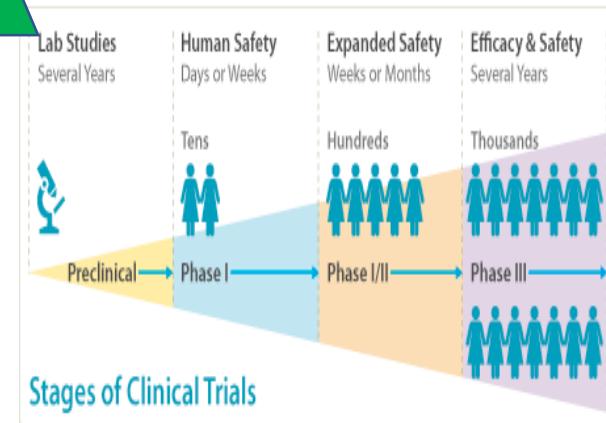
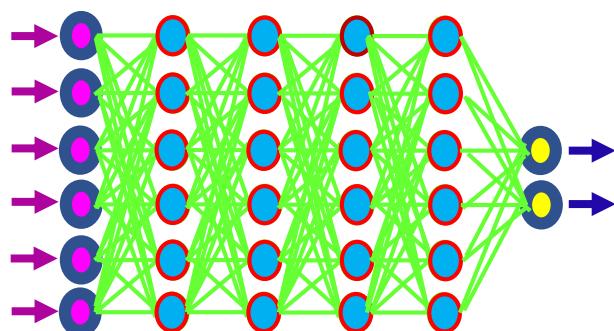
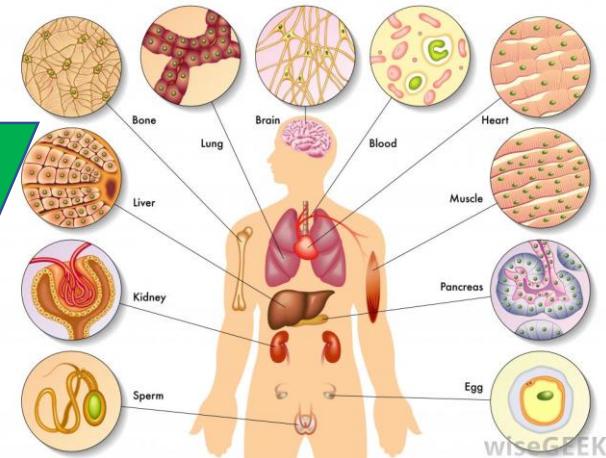
Predicting drug dose efficacy, frequency, range and safety by integrating

- Systems biology, protein networks, signal transduction pathways
- Cellular biology and cellular mechanics
- Systems physiological modeling
- Clinical data and virtual patient simulation
- Deep learning



In collaboration with  
Bristol-Myers Squibb (BMS)

Quantitative  
systems  
pharmacology  
modeling



Differential equation

Algebraic topology

Differential topology

Geometric topology

Algebraic graph

Geometric graph

Topological graph

Linear algebra

Biology became microscopic (i.e., molecular) in 1960s and added an omics dimension around the dawn of the millennium.

Driving by mathematics, biology is transforming from qualitative, phenomenological and descriptive to quantitative, predictive and analytical.

The last frontier of science is biology, while the last frontier of biology is mathematics.

Number theory

Algebraic geometry

Differential geometry

Euclidean geometry

Lie algebra

Complex analysis

Real analysis

Stochastic analysis



Baker Bates Bazil Burton Dickson Dong Hong Hu Lee Munch Tong Ye  
PNNL MSU UKLR



Shan Zhao Alabama Y Zhou CSU W Geng SMU Duan Chen UNCC Zhan Che GSU Lin Mu UGA S Yang CU KH S.N. Yu NY Y.H. Sun Denver L. H. Hu Chicago J. Park Indiana



KL Xia NTU B Wang UCLA ZX Cang UCI K Opron UM Y Gao P.I. KD Wu C. One ZX Zha Foshan CJ Chen HZUST



2019

## **Further topics and future directions**

- Mathematical models for high-throughput screening, hit to lead, and lead optimization.
- New generation of generative adversarial network (GAN) models for high-throughput screening, hit to lead, and lead optimization.
- Mathematical models for drug binding affinity, solubility, partition coefficient, toxicity, permeability, protein plasma binding, clearance, PK, etc.
- Integrating PK, PD, clinic data and AI for quantitative systems pharmacology (QSP).
- Machine learning aided PK modeling.
- Mathematical/transport models for PD.
- Macromolecular drugs.
- Personalized medicines.
- Mathematical modeling of drug delivery.
- Drug design inspired new mathematics for drug discovery.



thank you