Computational Methods for Drug Design: Traditional & ML

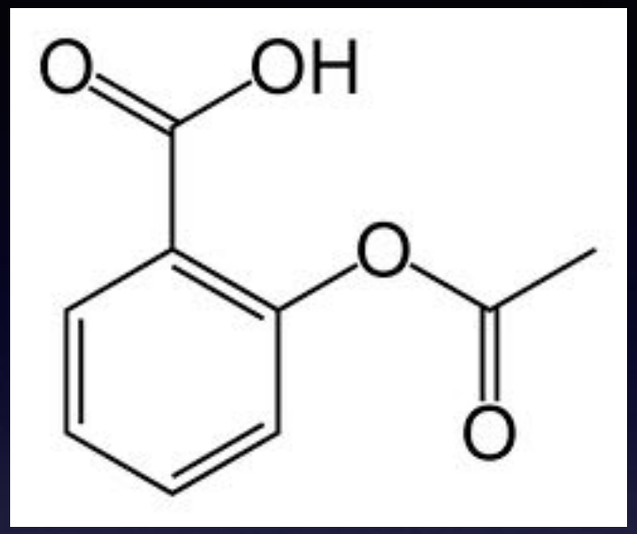
HUBDATA

IIIT, HYDERABAD

CCNSB

@deva_priyakumar

- Back to Chemistry little bit :-)
  - Drug design (as an eg.)
    - Rule (Physics) based methods
    - Role of ML
  - Feature Engineering in Machine Learning
  - Molecular Featurizations (Feature vector)



Story of Aspirin

Most widely used drug in the world - 40,000 tons per year

Aspirin

- Believed to be in use from the 3000 BC
- 400 BC: Hippocrates: Extract from willow (active ingredient: salicylic acid) - pain reliever
- 1853/1859/1869: Synthesis of aspirin (acetyl salicylic acid)
- 1899: Bayer - marketed & patented (Origin of Non-steroidal anti-inflammatory drugs)

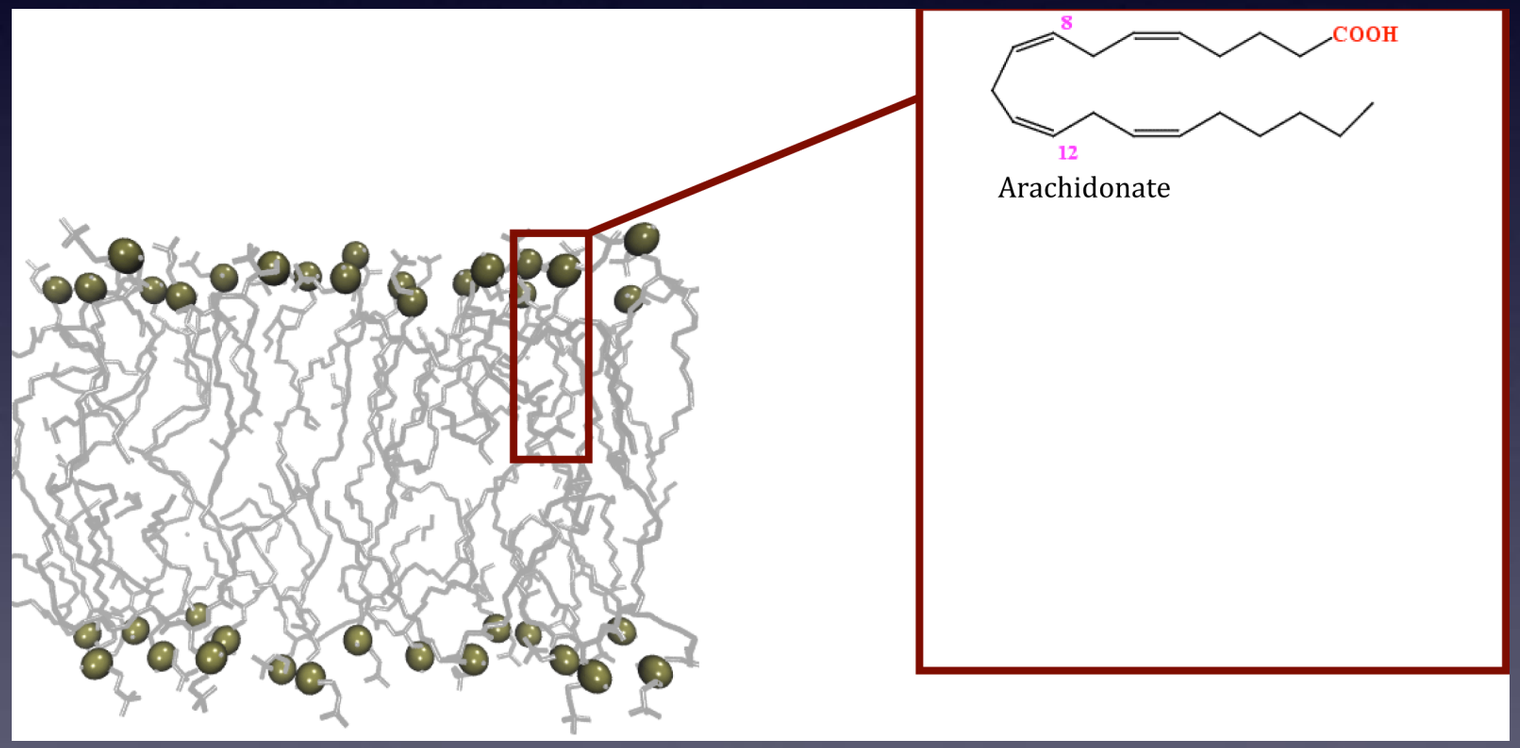
Aspirin

- 1917: Patent expiration
- 1918: Spanish flu
- 1956 & 1969: Acetaminophen & Ibuprofen - downfall of aspirin!
- 1970-1985: Aspirin, an anti-clotting agent - revival! - heart attack & strokes

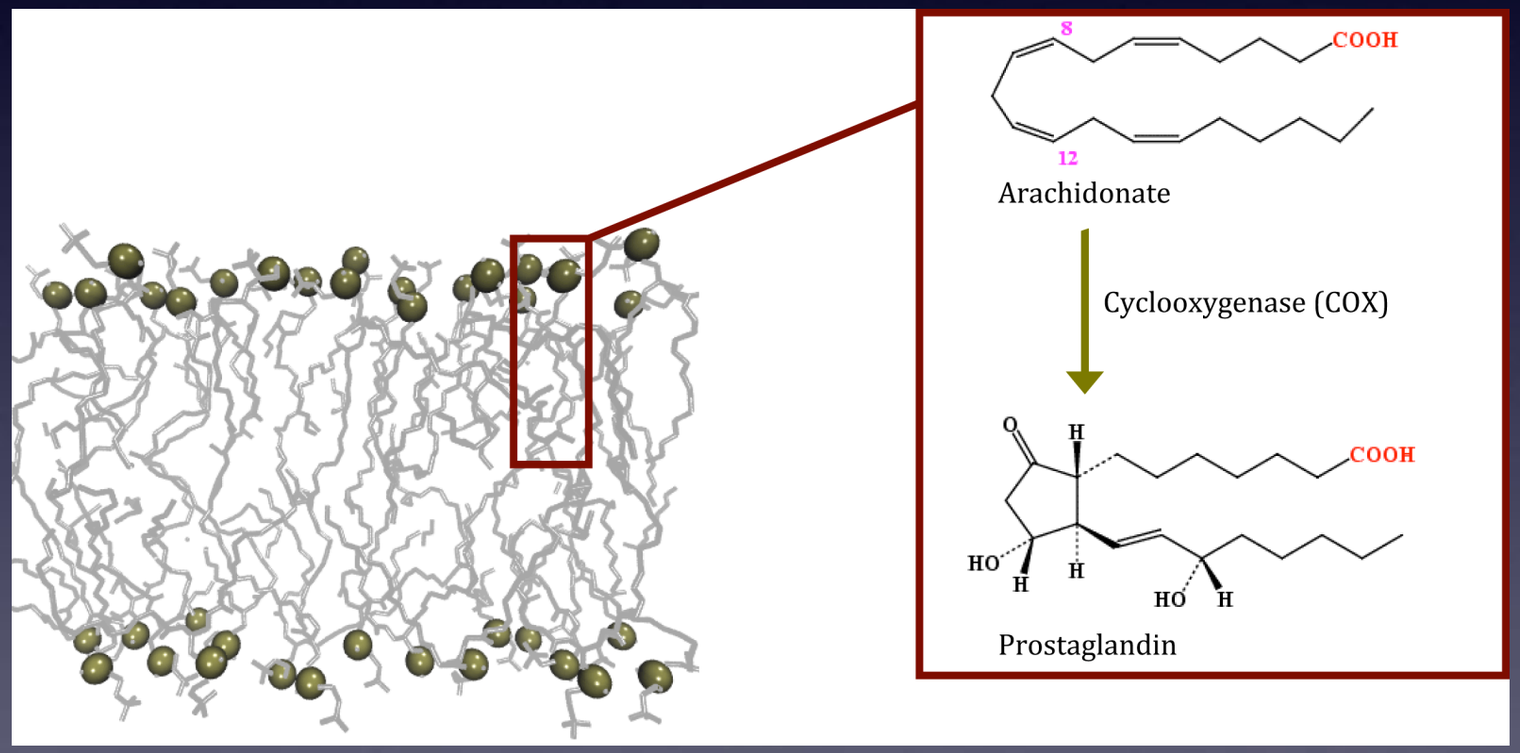
Aspirin - mechanism

- 1970s: Target cyclooxygenase (COX)
- 1980s: Identification of COX1 & COX2
- 1990s: COX1&2 structures
- 1990s - present: Structure Based Design
- 1999: Celebrex & Vioxx marketed by Pfizer & Merck for treating osteoarthritis & rheumatoid arthritis
- 2004: Vioxx withdrawn by Merck
- 90-140k heart attacks & about 40% fatal!
- Present: Celebrex is being sold but with a “black box warning”

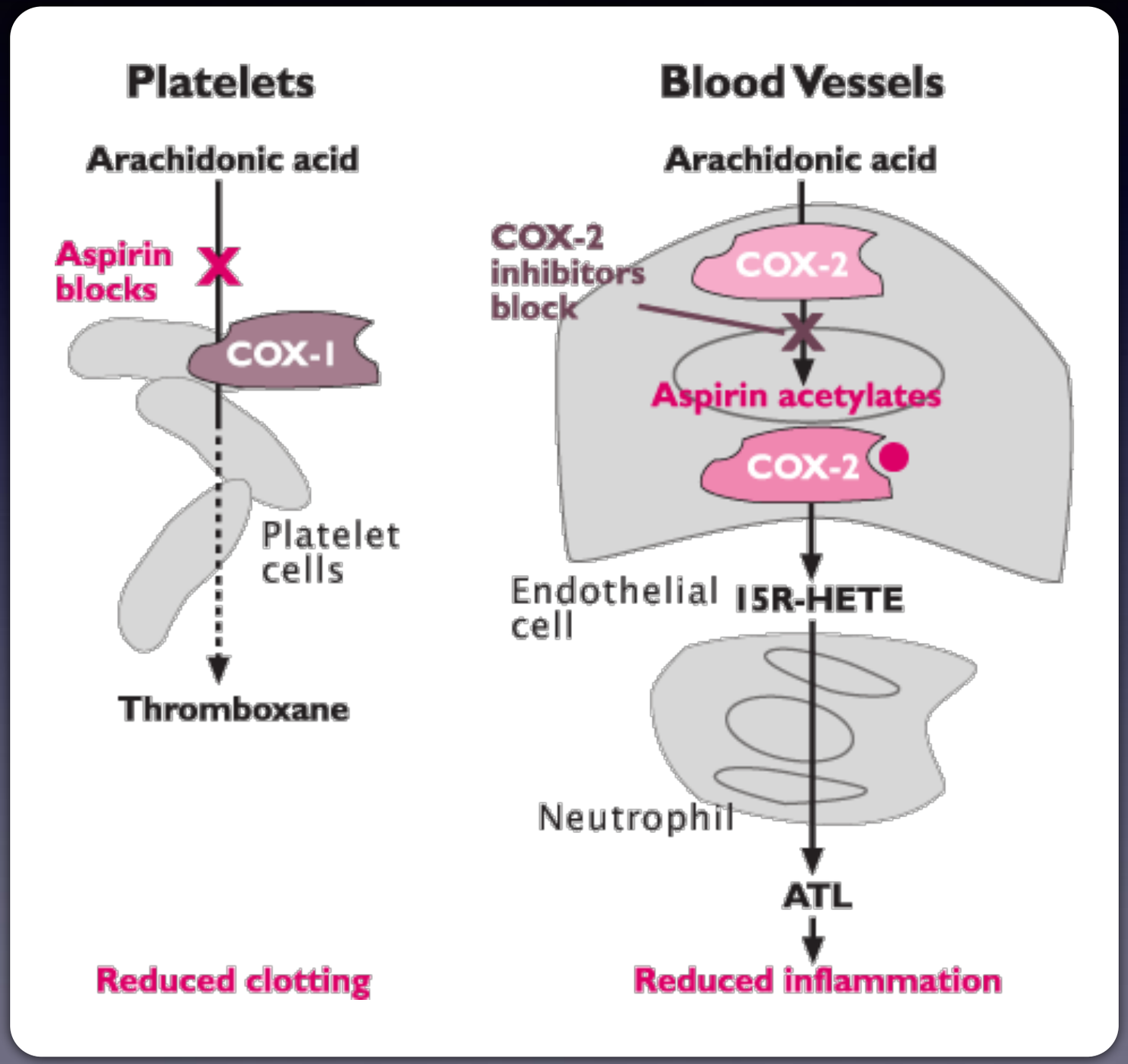
Action of Aspirin



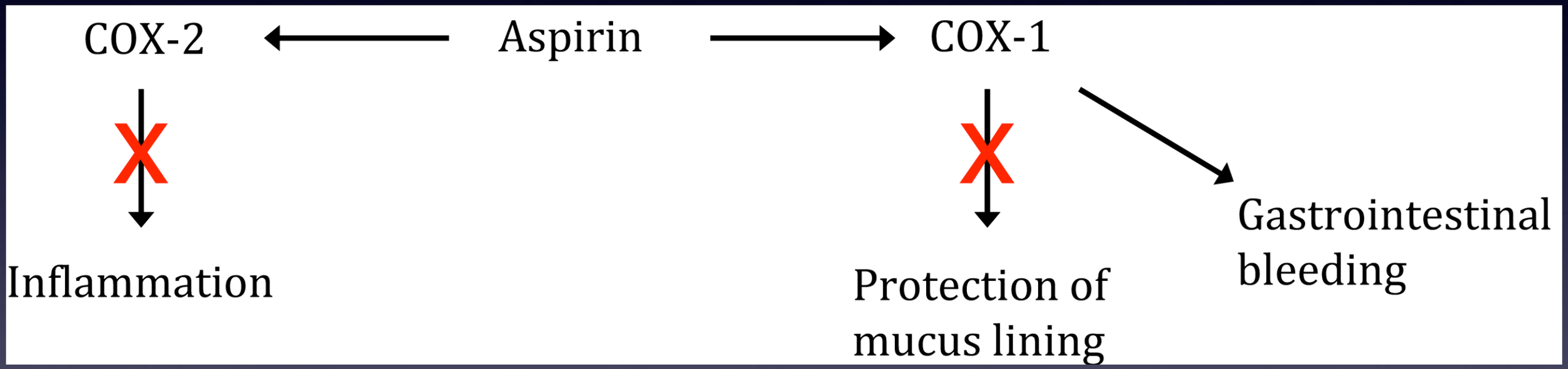
Action of Aspirin



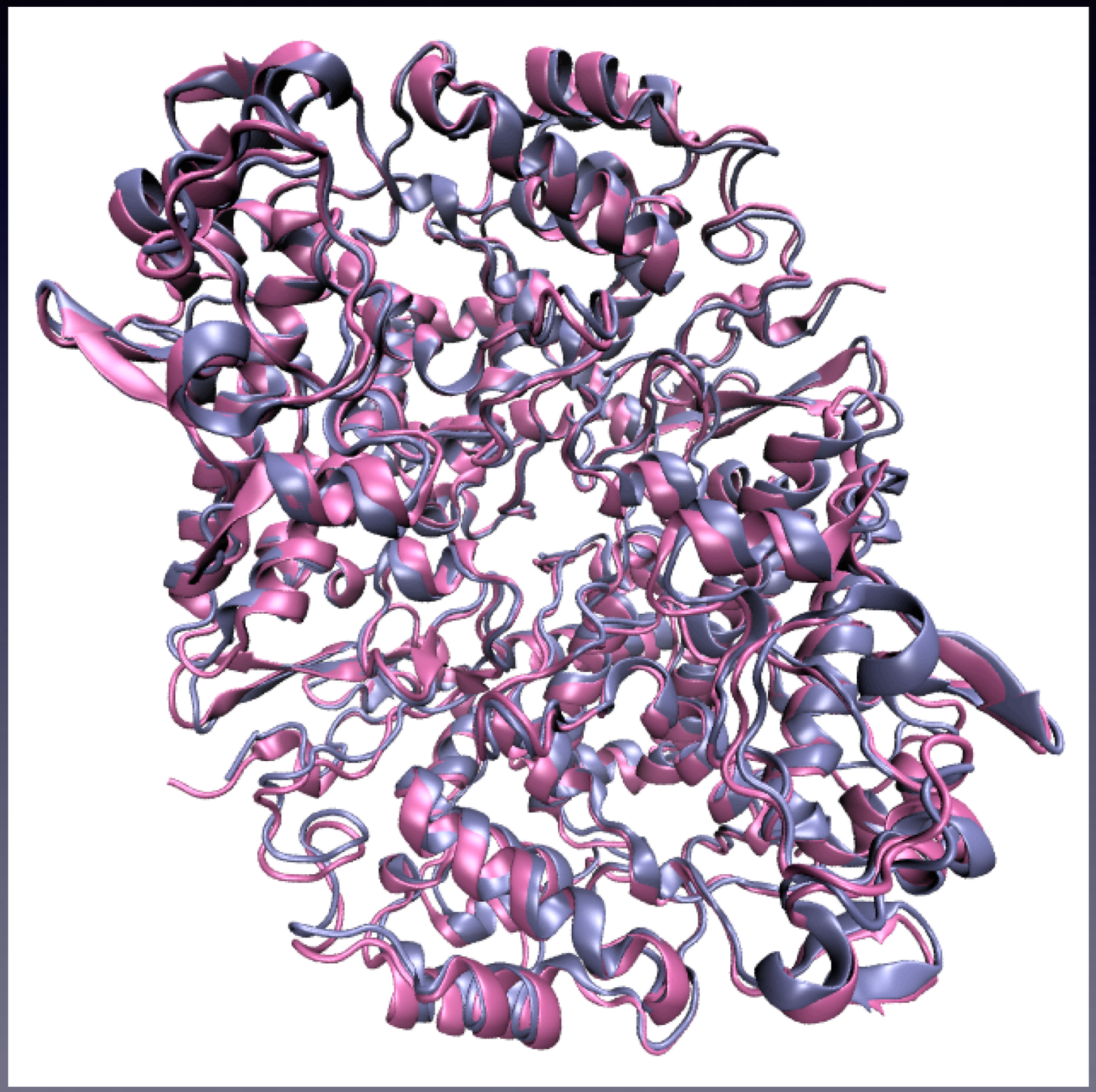
Action of Aspirin



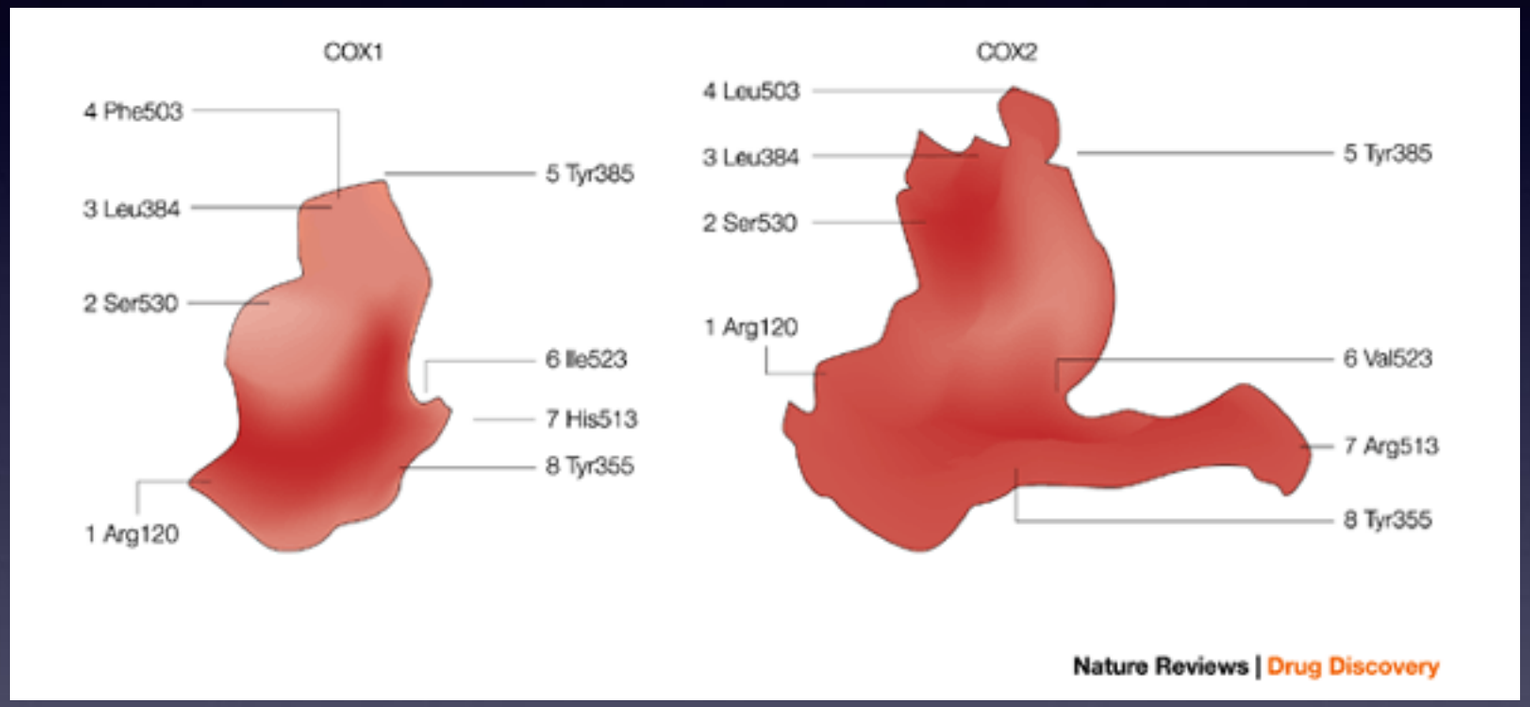
Action of Aspirin



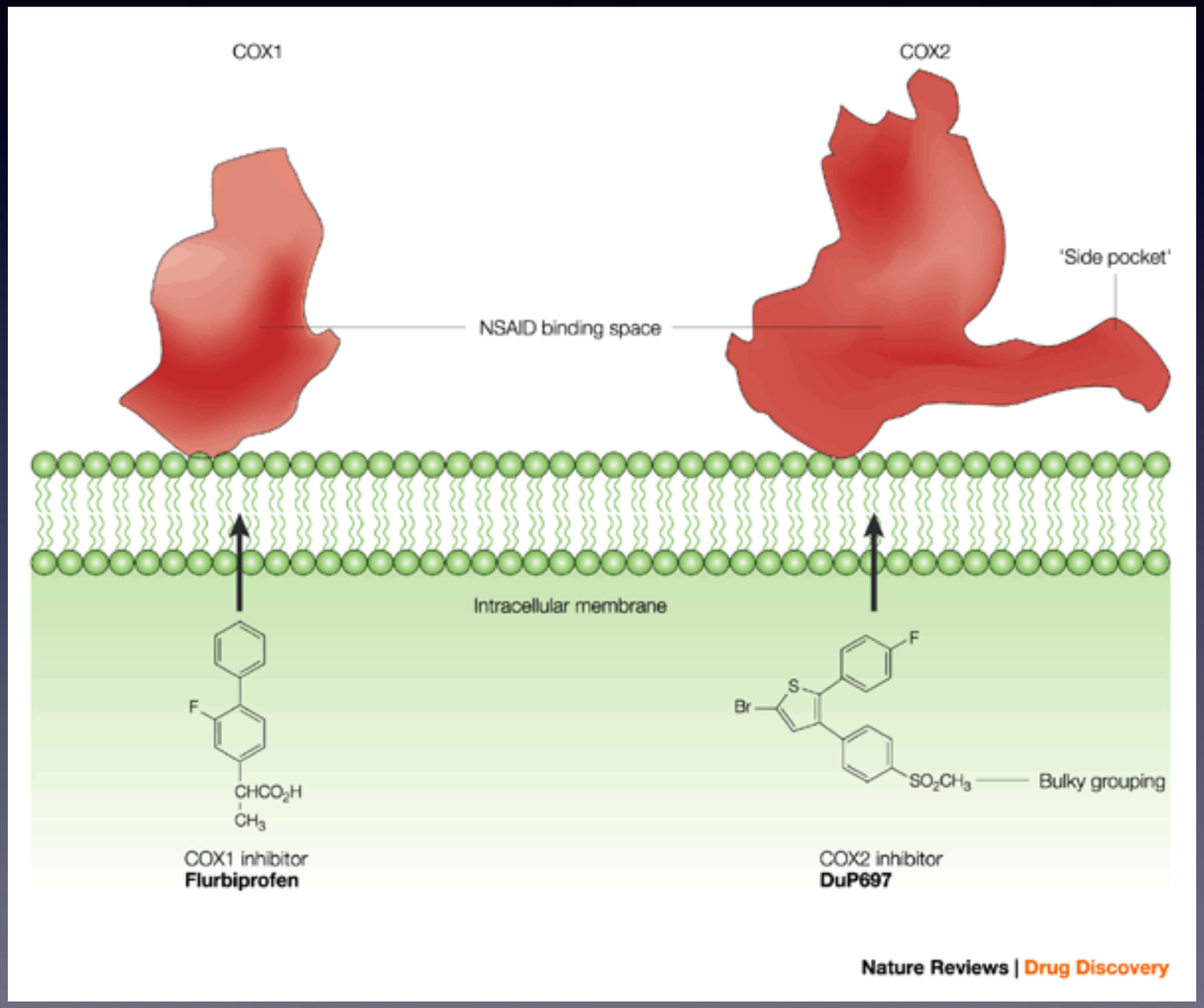
COX-1 and COX-2

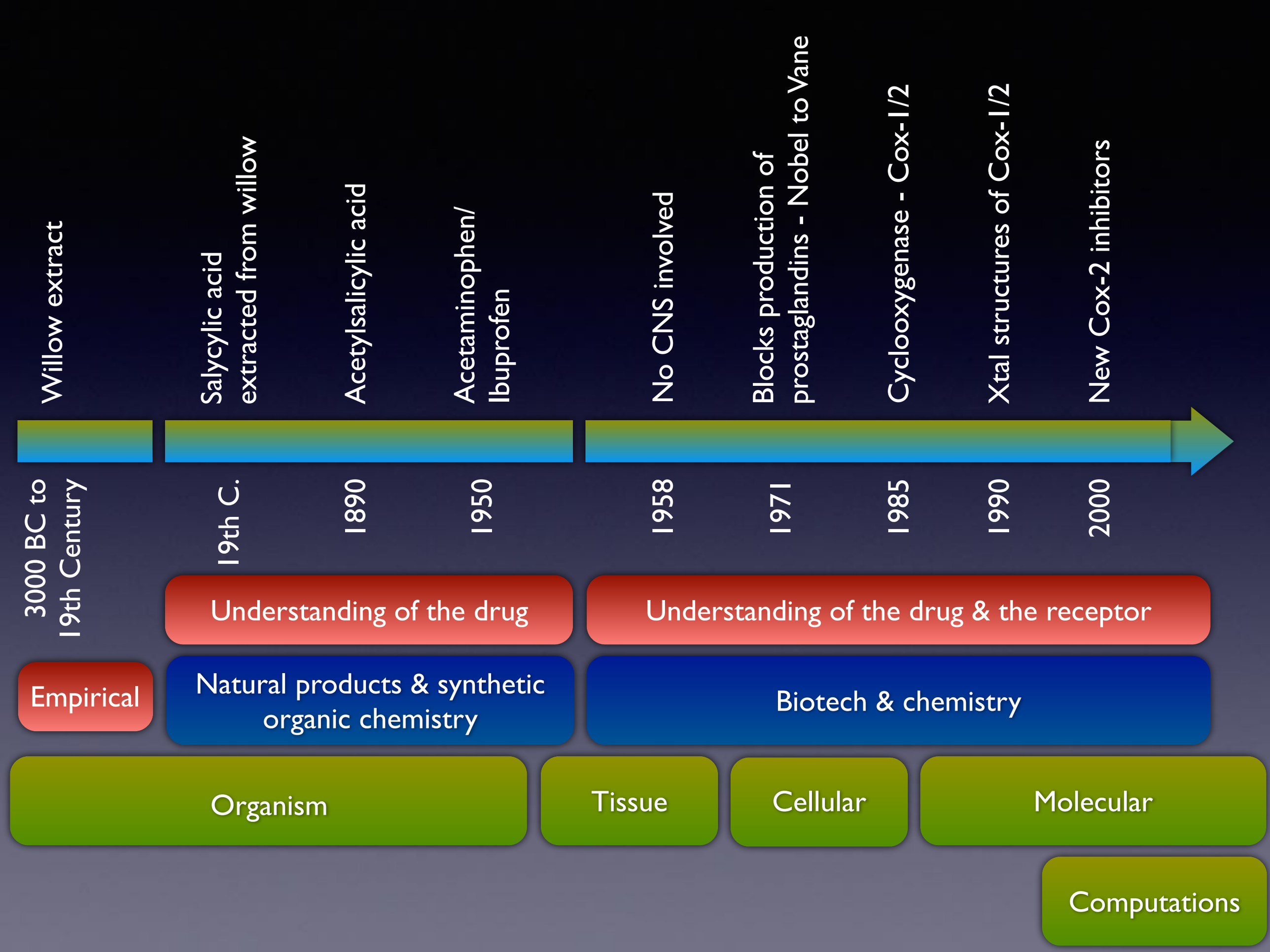


COX-1 and COX-2

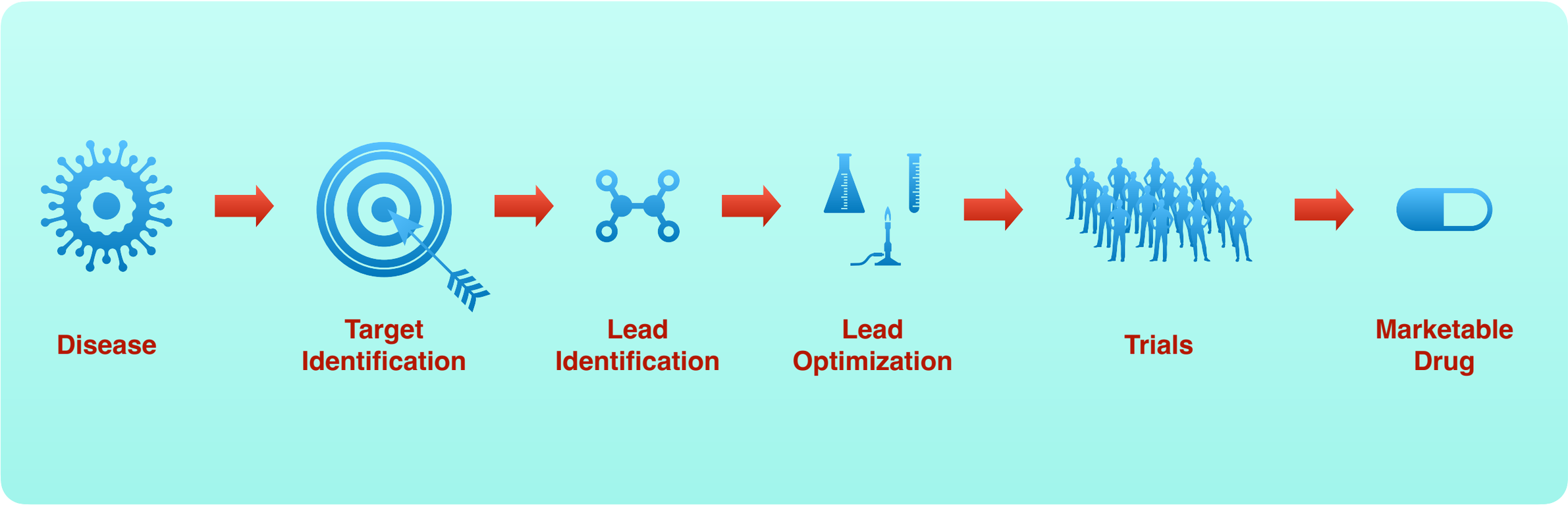


COX-1 and COX-2

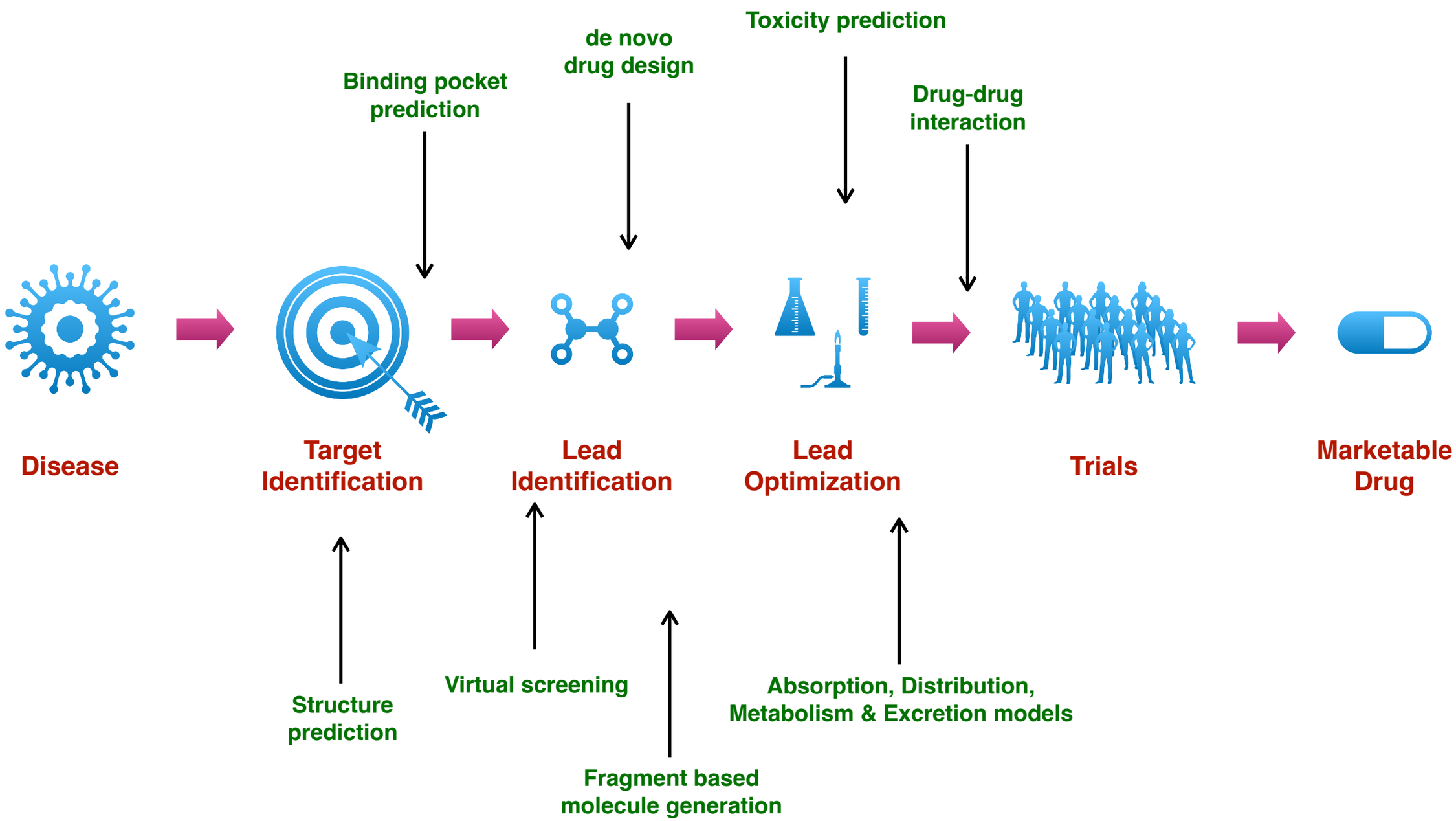



Willow extract
3000 BC to 19th Century
Salycylic acid extracted from willow
19th C.
Acetylsalicylic acid
1890
Acetaminophen/ Ibuprofen
1950
No CNS involved
1958
Blocks production of prostaglandins - Nobel to Vane
1971
Cyclooxygenase - Cox-1/2
1985
Xtal structures of Cox-1/2
1990
New Cox-2 inhibitors
2000
Understanding of the drug
Understanding of the drug & the receptor
Empirical
Natural products & synthetic organic chemistry
Biotech & chemistry
Organism
Tissue
Cellular
Molecular
Computations

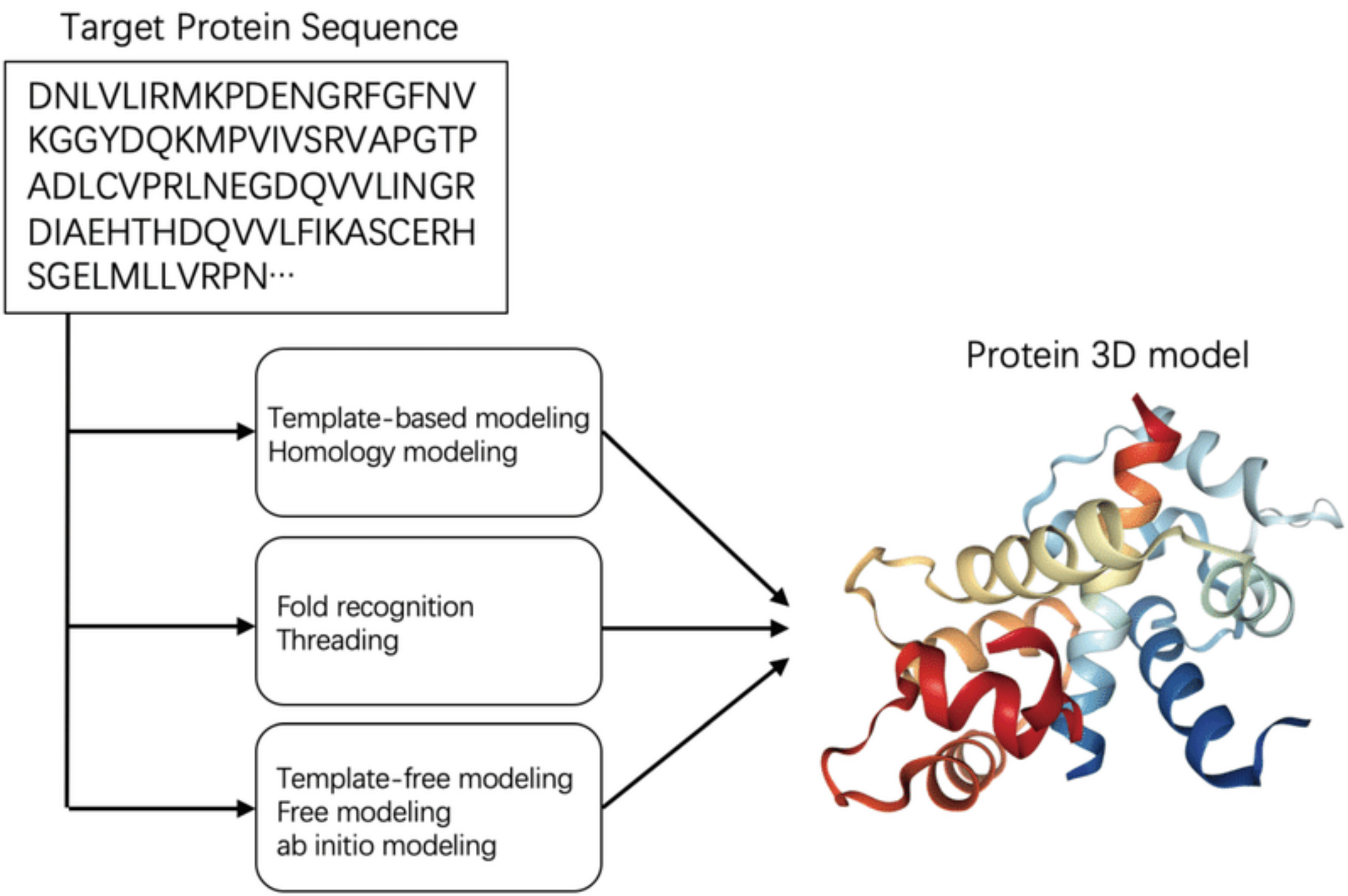
Drug Discovery Process



Drug Discovery Process

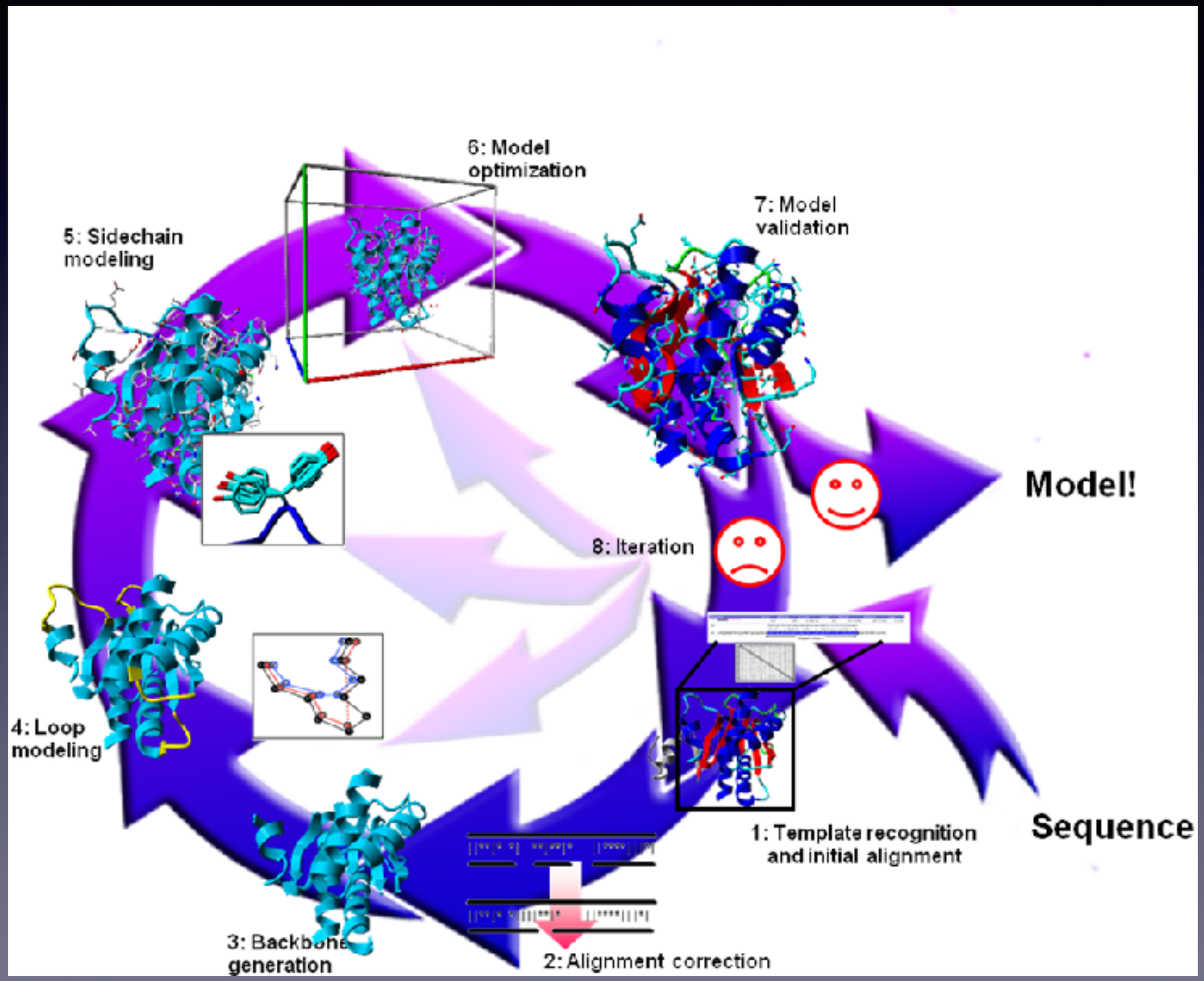


Protein Structure Prediction

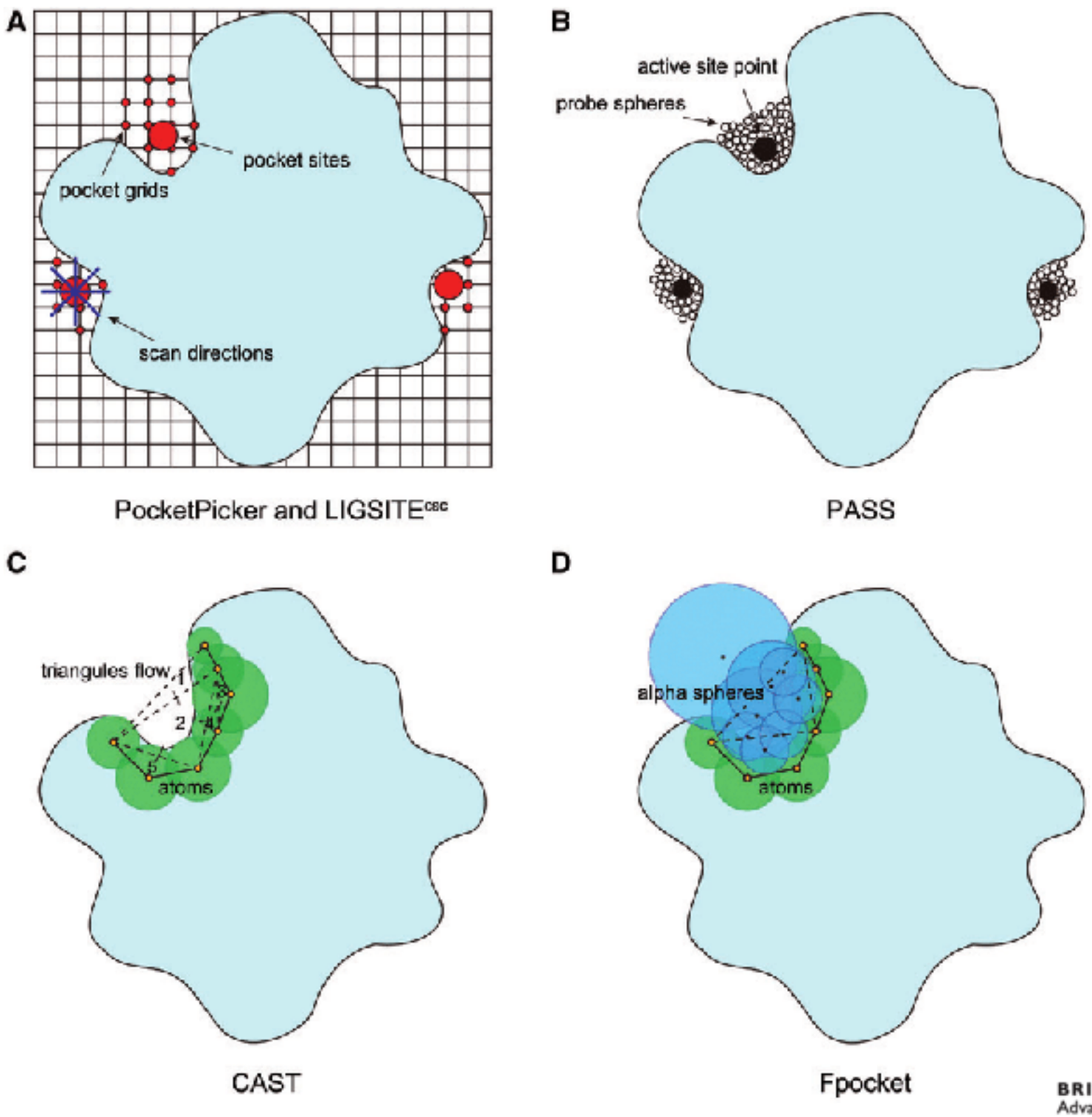


Biomolecules **2020**, *10*, 626; doi:10.3390/biom10040626

Homology Modeling

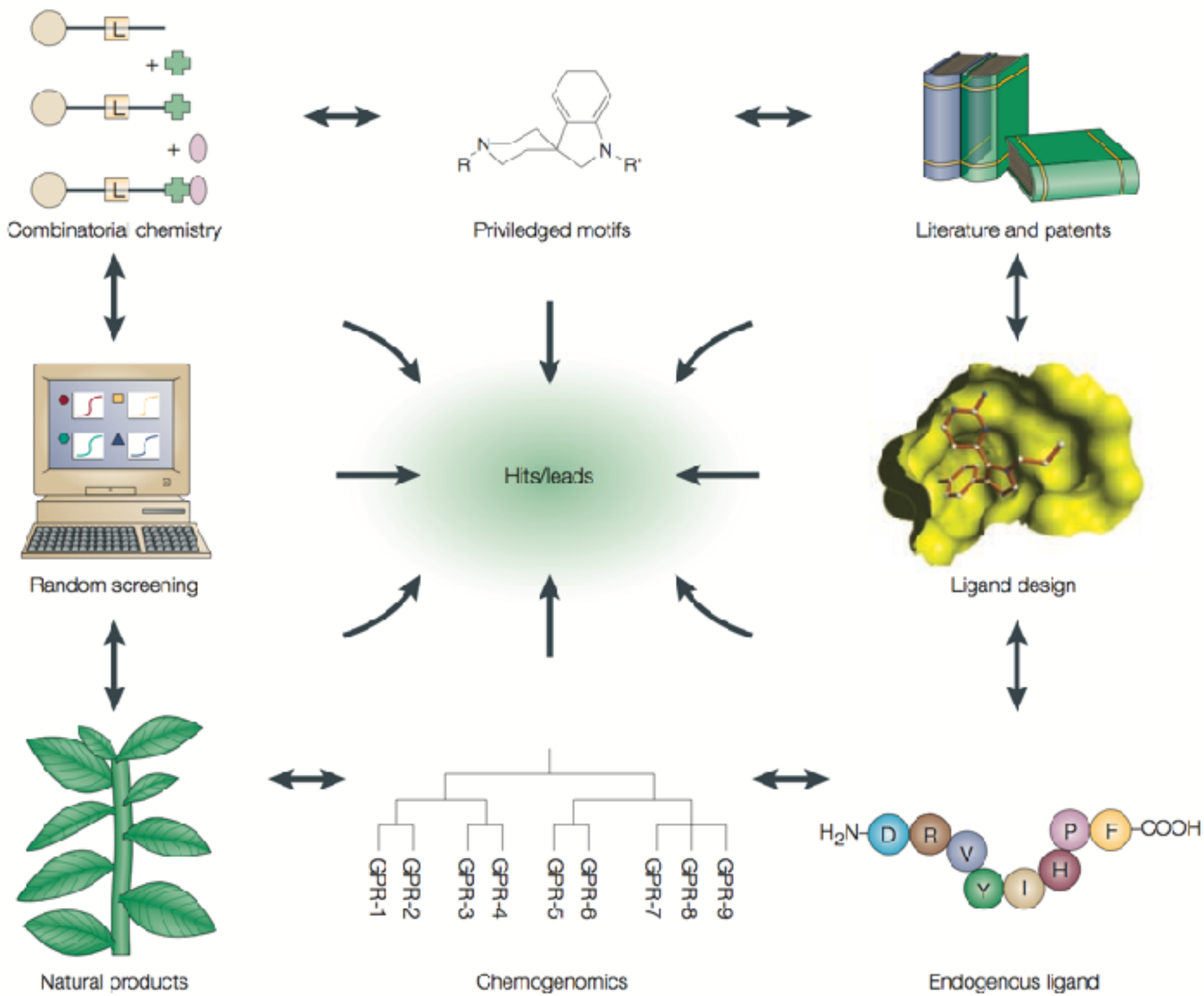


Binding Pocket Identification

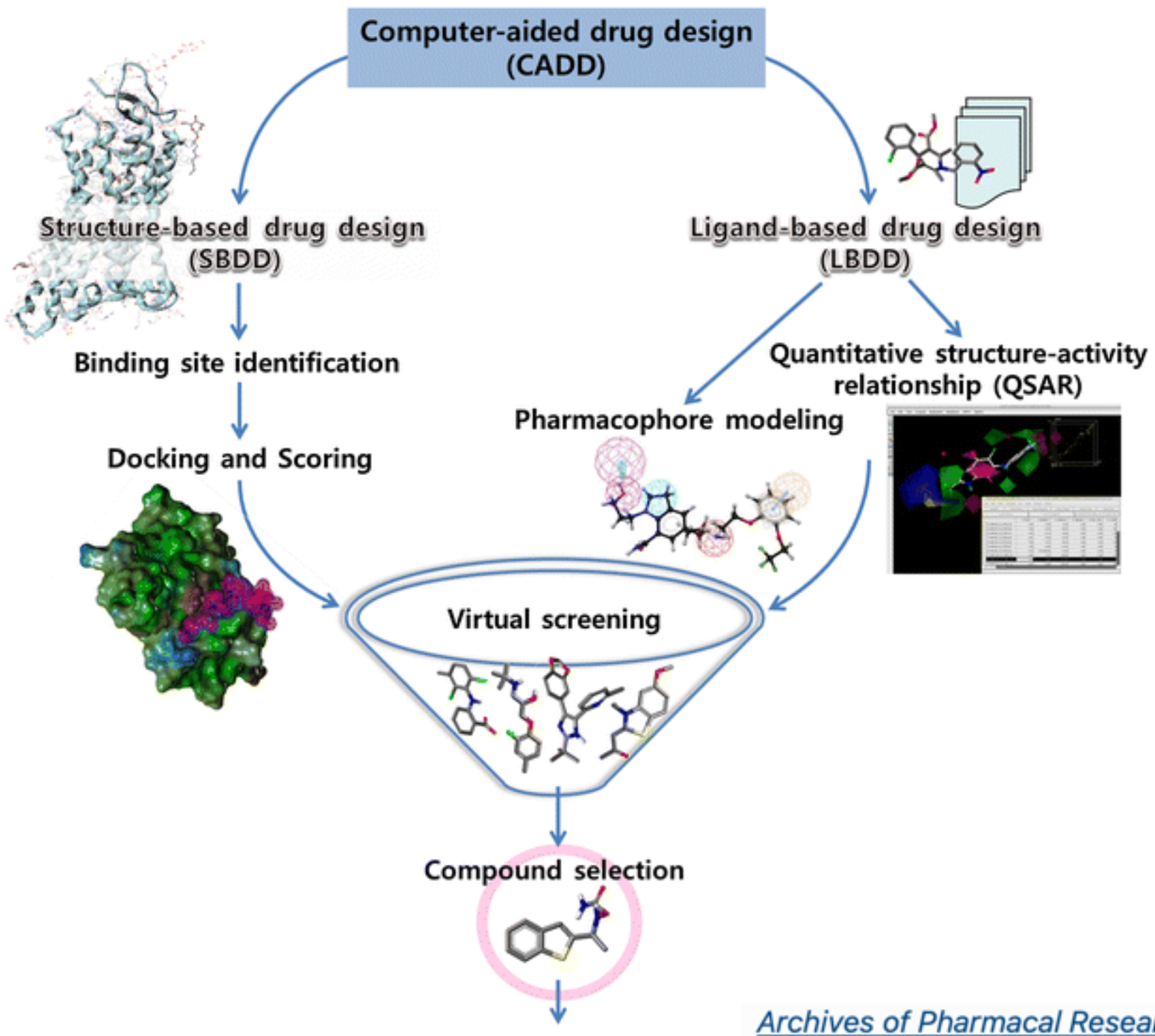


BRIEFINGS IN BIOINFORMATICS. VOL 15 NO 1. 54–64
Advance Access published on 27 November 2012

Hit Identification

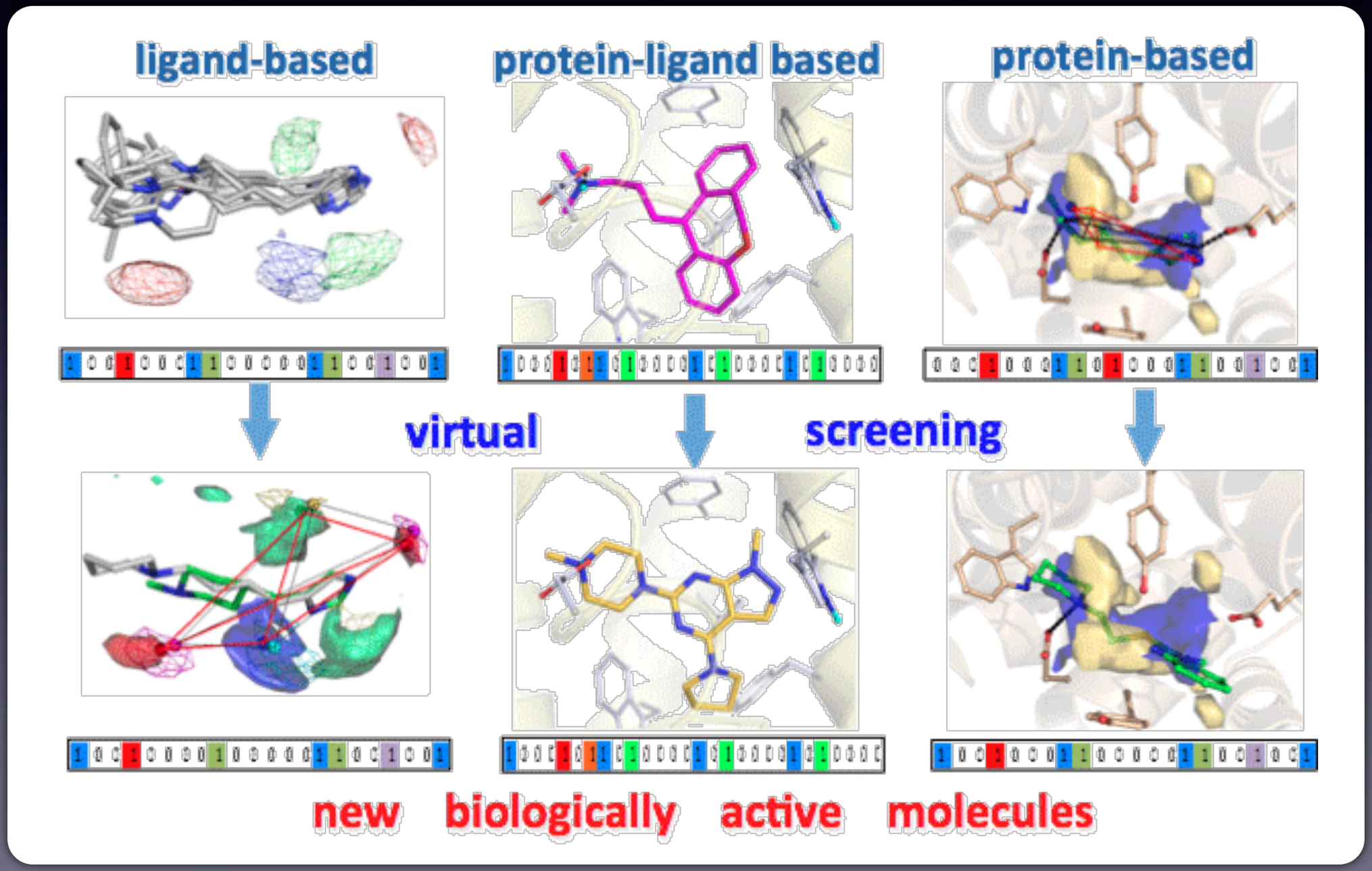


Hit Identification

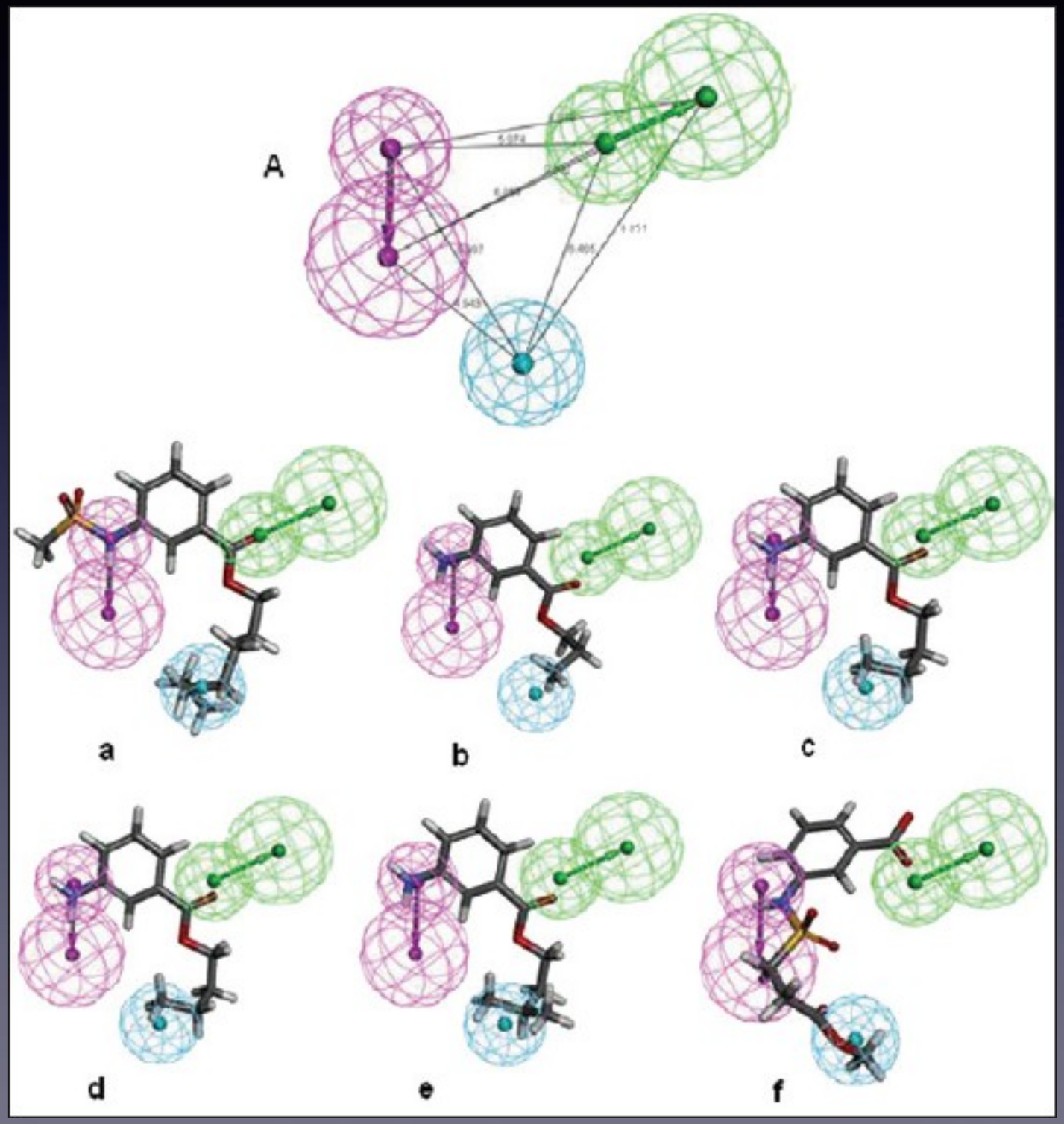


Archives of Pharmacal Research **38**, 1686–1701 (2015)

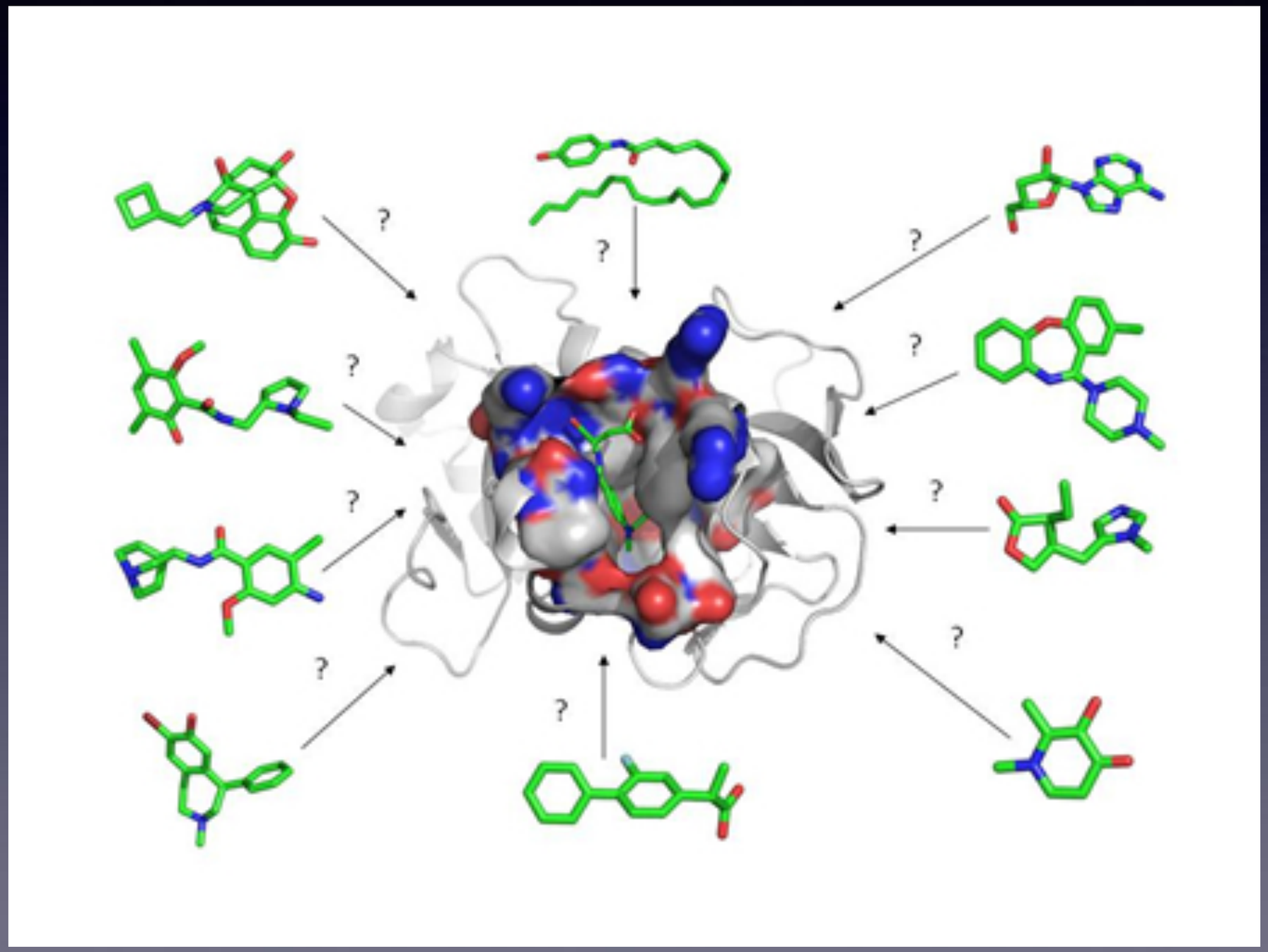
Virtual Screening



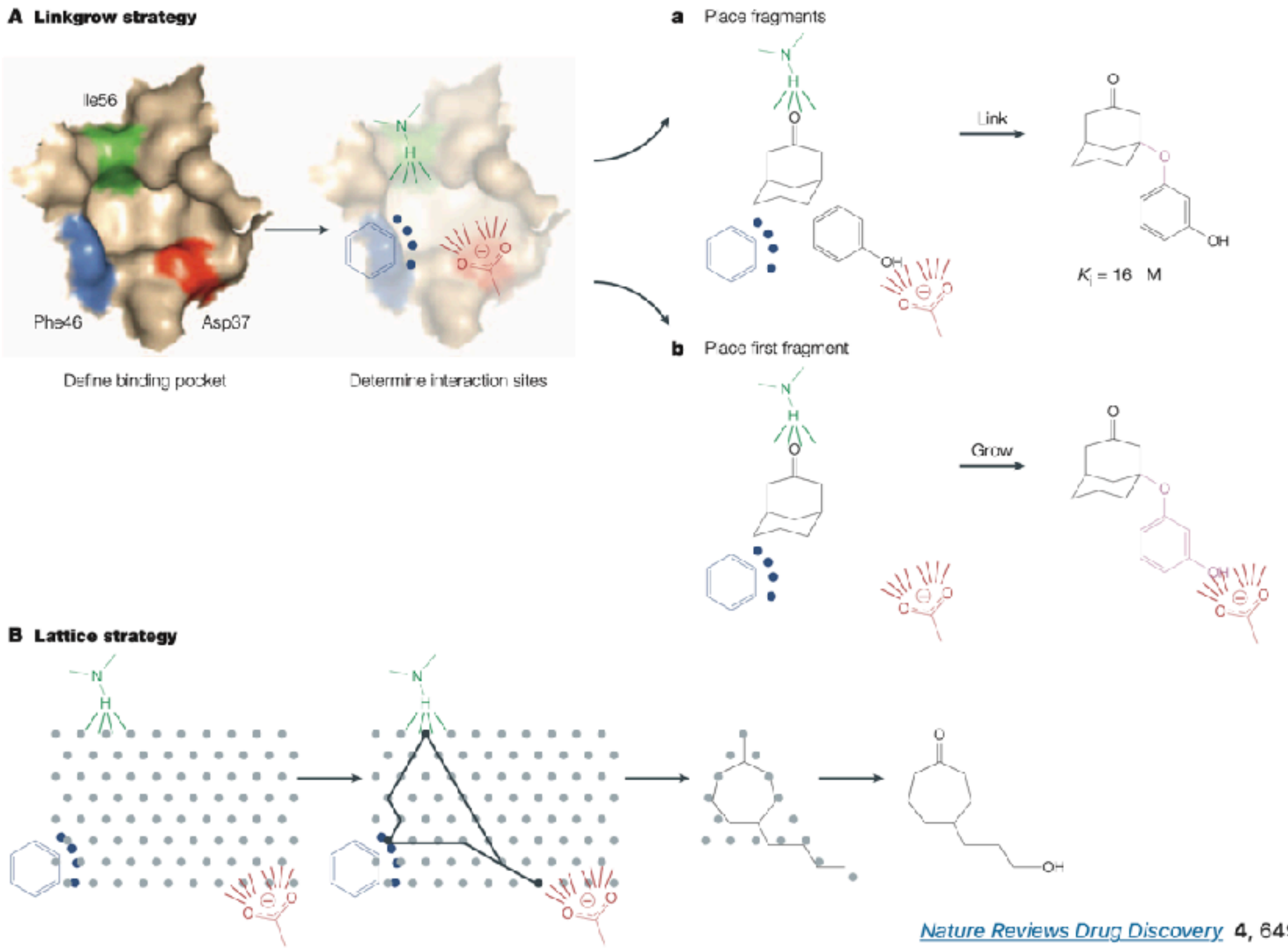
Ligand Based



Protein Based

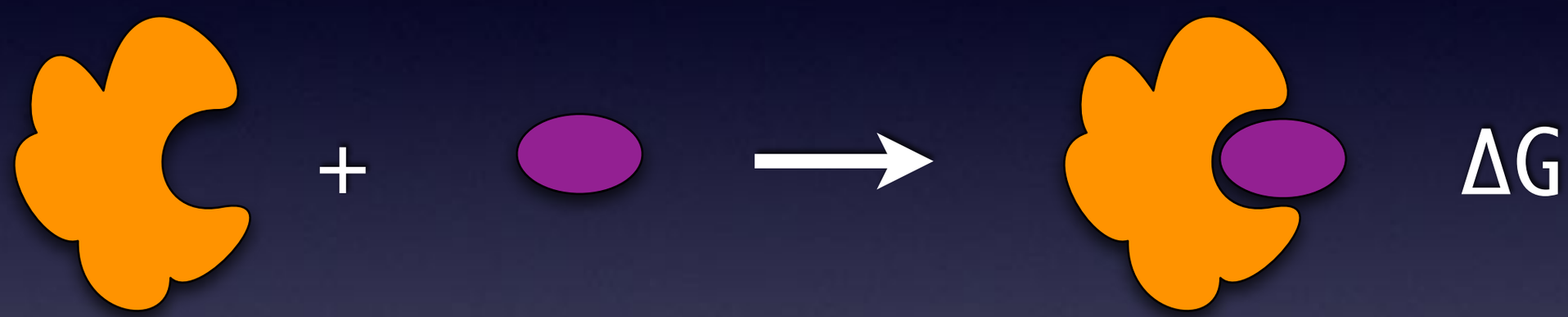


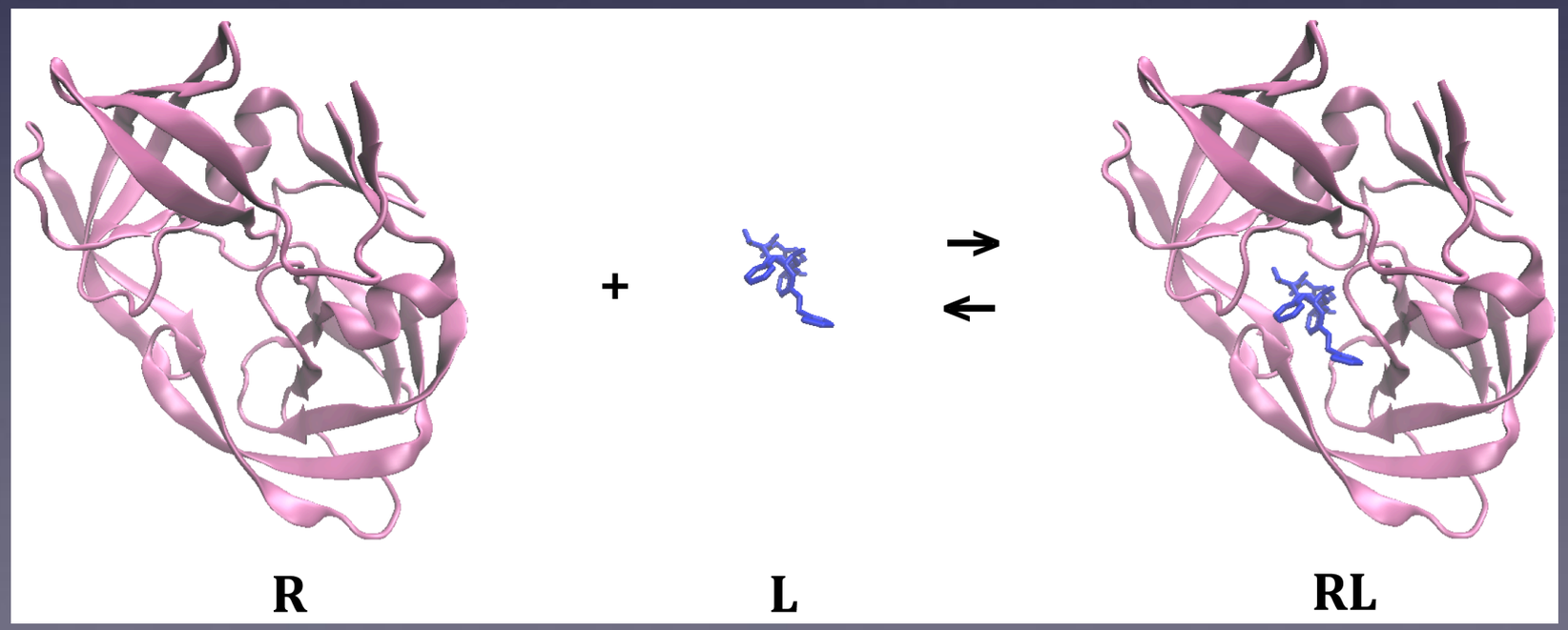
de novo Design

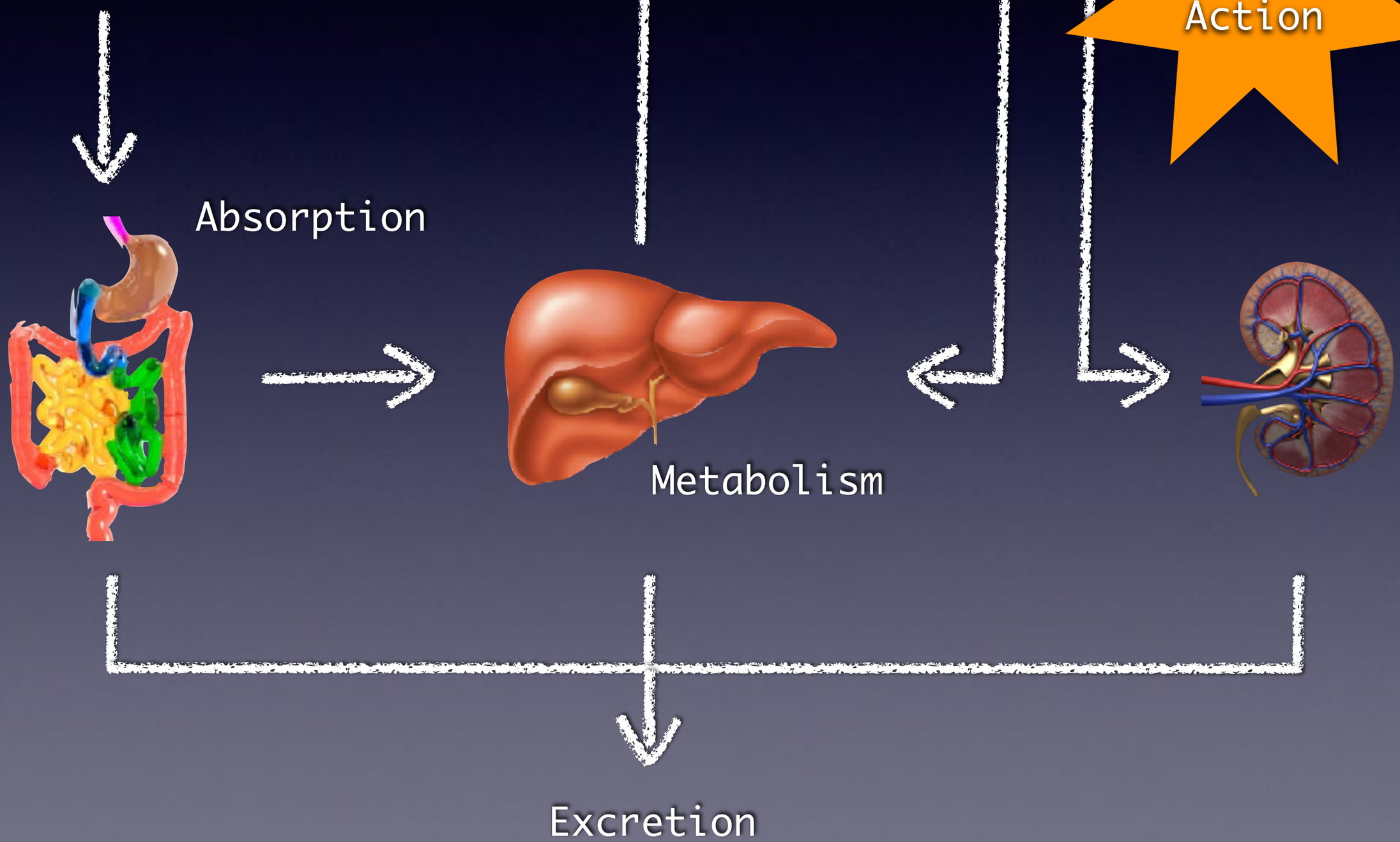


Nature Reviews Drug Discovery **4**, 649–663 (2005)

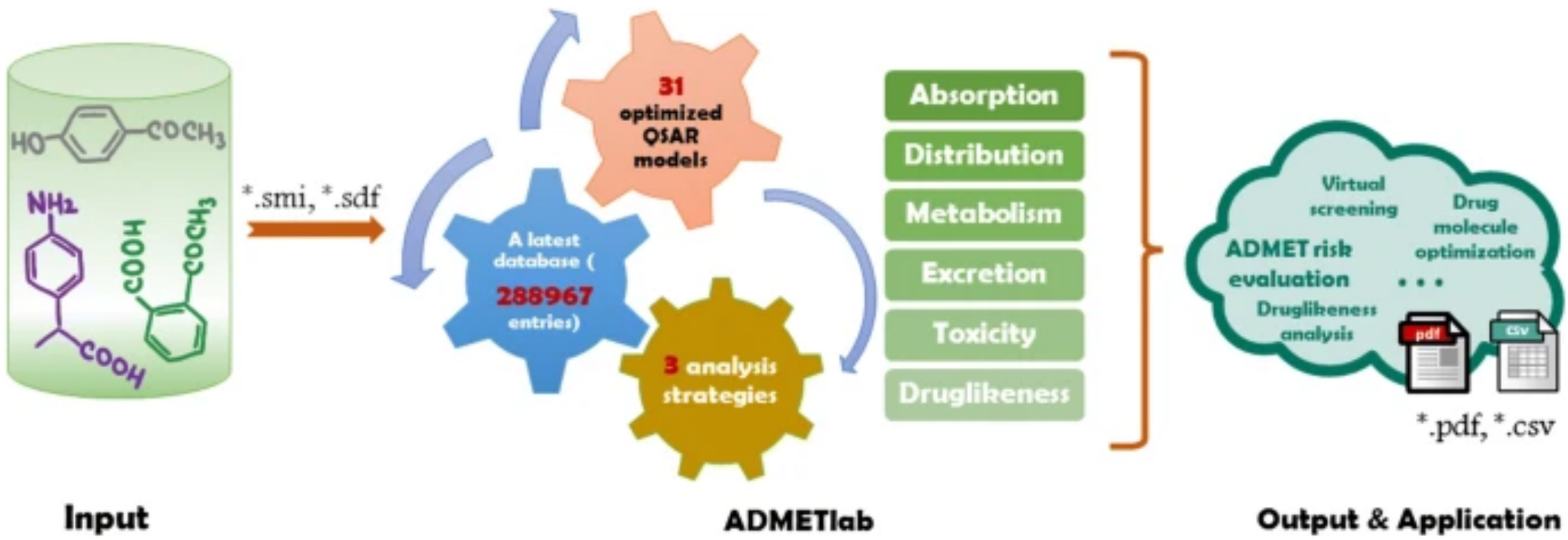
Drug Action



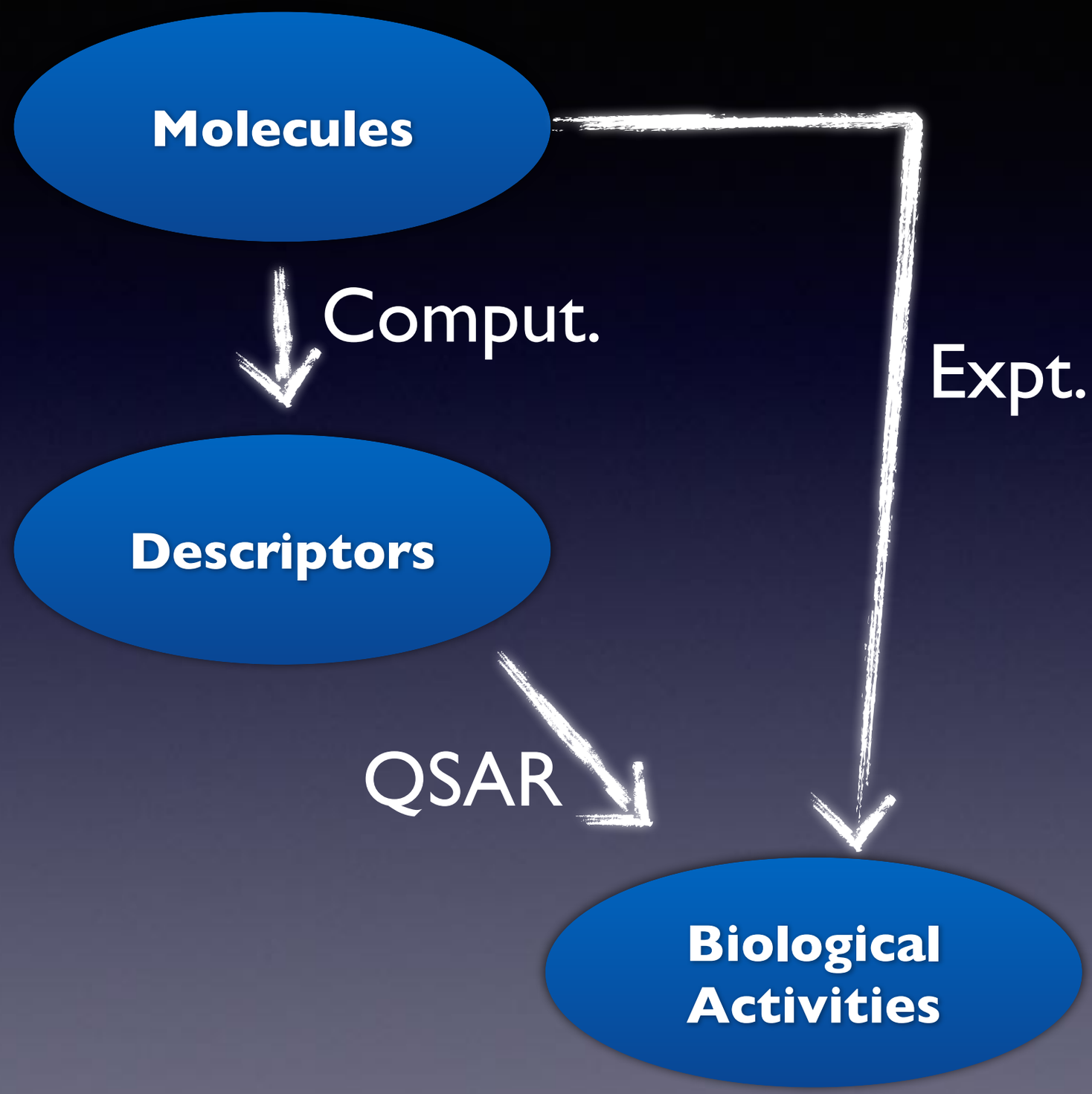



Distribution
Drug Action
Absorption
Metabolism
Excretion

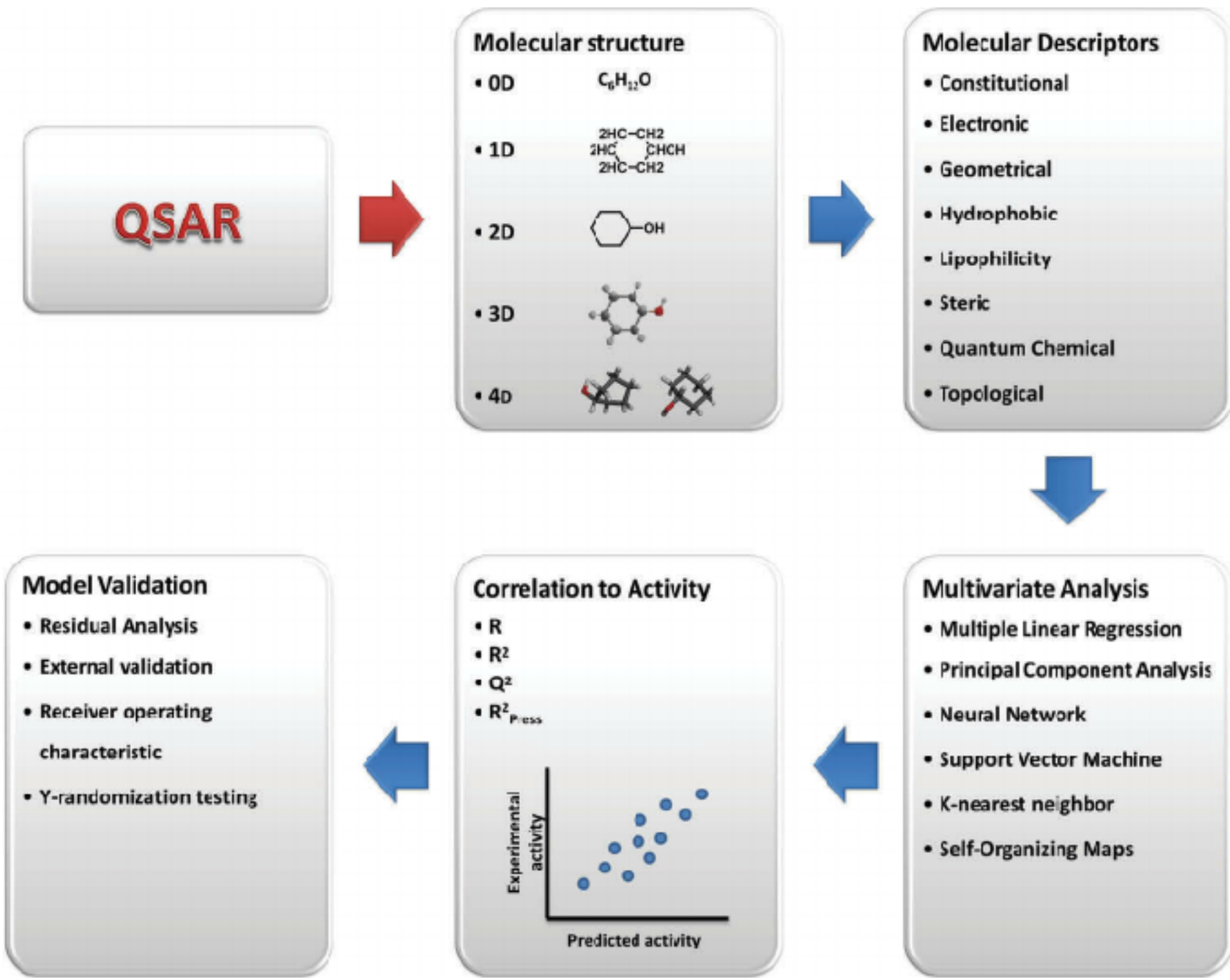
ADMET



Journal of Cheminformatics **10**, Article number: 29 (2018)


Molecules
Comput.
Expt.
Descriptors
QSAR
Biological Activities

QSAR



Expert Opinion on Drug Discovery, 11:2, 197-214,

Role of Machine Learning in Drug Design

nature

Article | Published: 15 January 2020

Improved protein structure prediction using potentials from deep learning

RESEARCH ARTICLE

A robotic platform for flow synthesis of organic compounds informed by AI planning

Connor W. Coley[1,*], Dale A. Thomas III[1,2,†], Justin A. M. Lummiss[3,4,†], Jonathan N. Jaworski[3,‡], Christopher P...

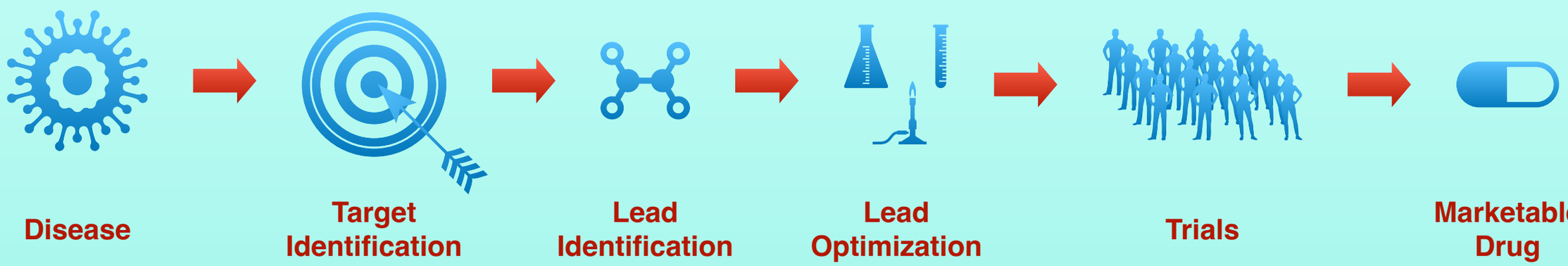
+ See all authors and affiliations

Science 09 Aug 2019:
Vol. 365, Issue 6453, eaax1566
DOI: 10.1126/science.aax1566

nature

Published: 29 March 2018

Planning chemical syntheses with deep neural networks and symbolic AI



nature communications

Article | Open Access | Published: 19 November 2019

A Bayesian machine learning approach for drug target identification using diverse data types

Machine Learning Tools Help Predict Clinical Trial Outcomes

Using machine learning tools to forecast clinical trial outcomes can help stakeholders speed the drug and device approval process.

Deep reinforcement learning for de novo drug design

Mariya Popova[1,2,3], Olexandr Isayev[1,*] and Alexander Tropsha[1,*]

+ See all authors and affiliations

Science Advances 25 Jul 2018:
Vol. 4, no. 7, eaap7885
DOI: 10.1126/sciadv.aap7885

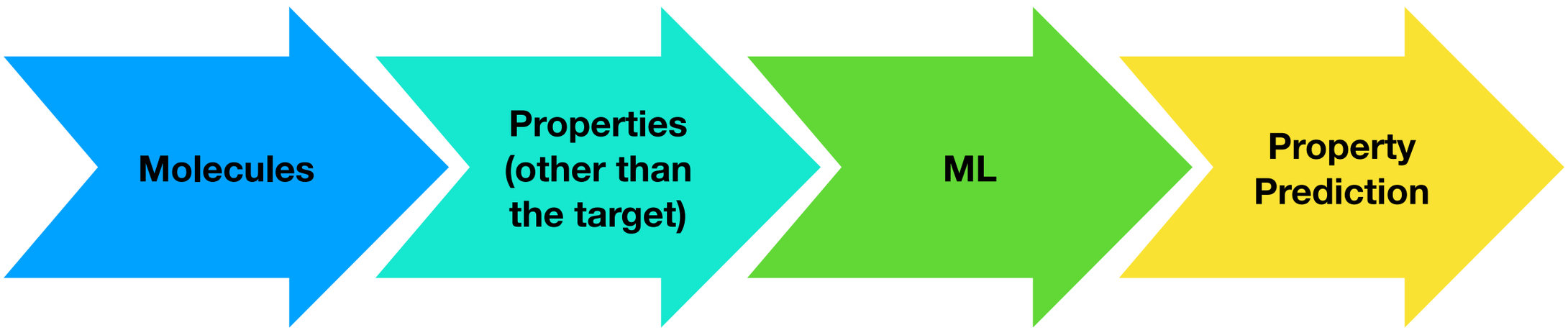
NMR-TS: De Novo Molecule Identification from NMR Spectra

Preprint submitted on 01.04.2020, 12:47 and posted on 03.04.2020, 23:04 by Jinzhe Zhang, Kei Terayama, Masato Sumita, Kazuki Yoshizoe, Kengo Ito, Jun Kikuchi, Koji Tsuda

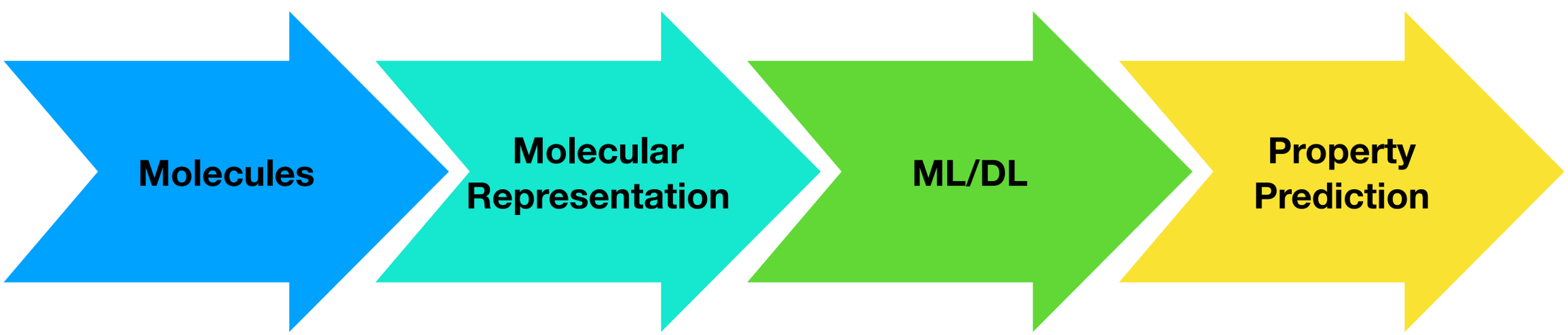
Let us stick to prediction problems for now:
Regression & Classification (mostly)

- As mentioned before, we have been using ML in drug design for a long time

- Typically,

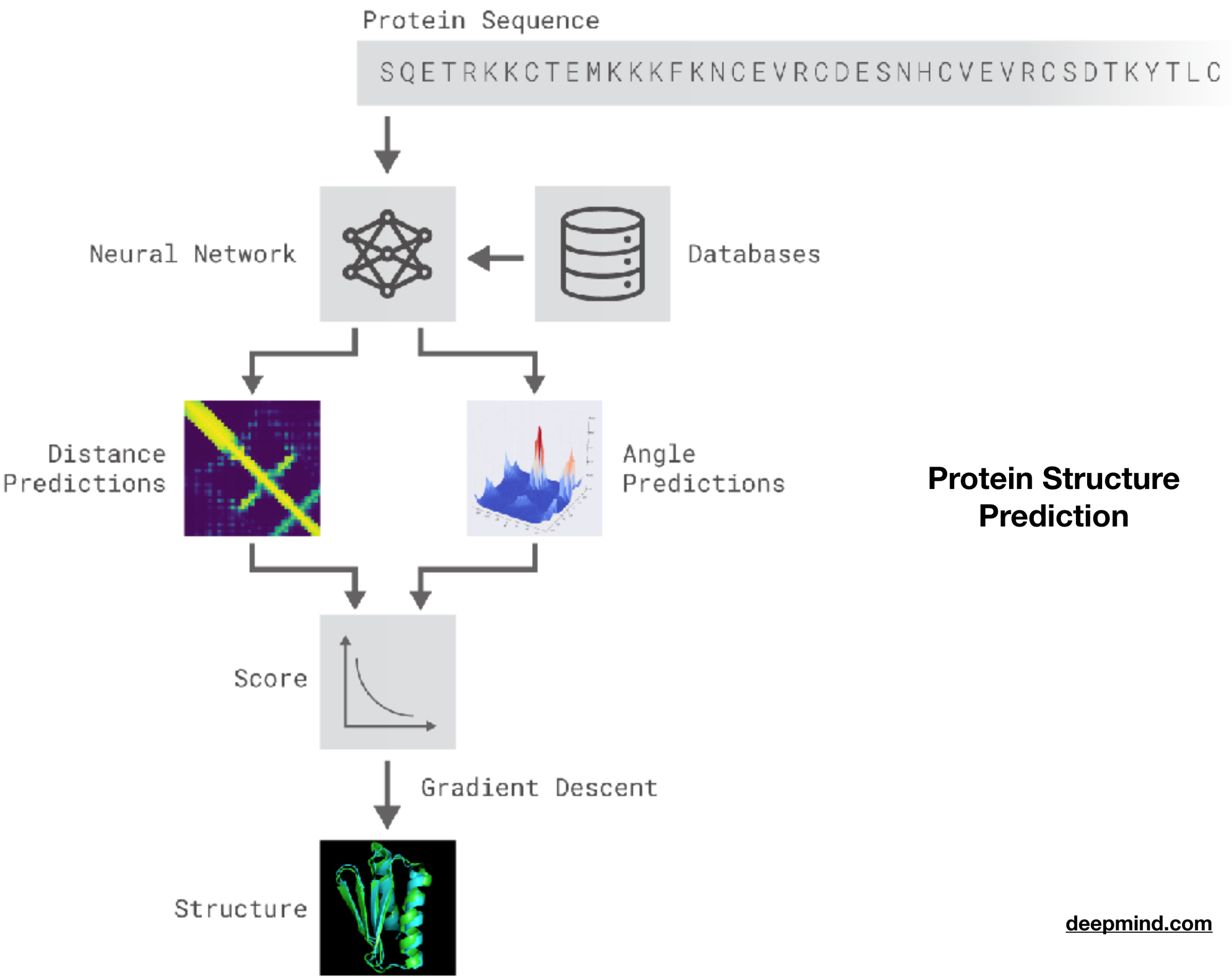


- Now,



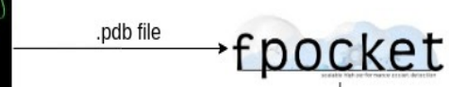
Same tasks different methods

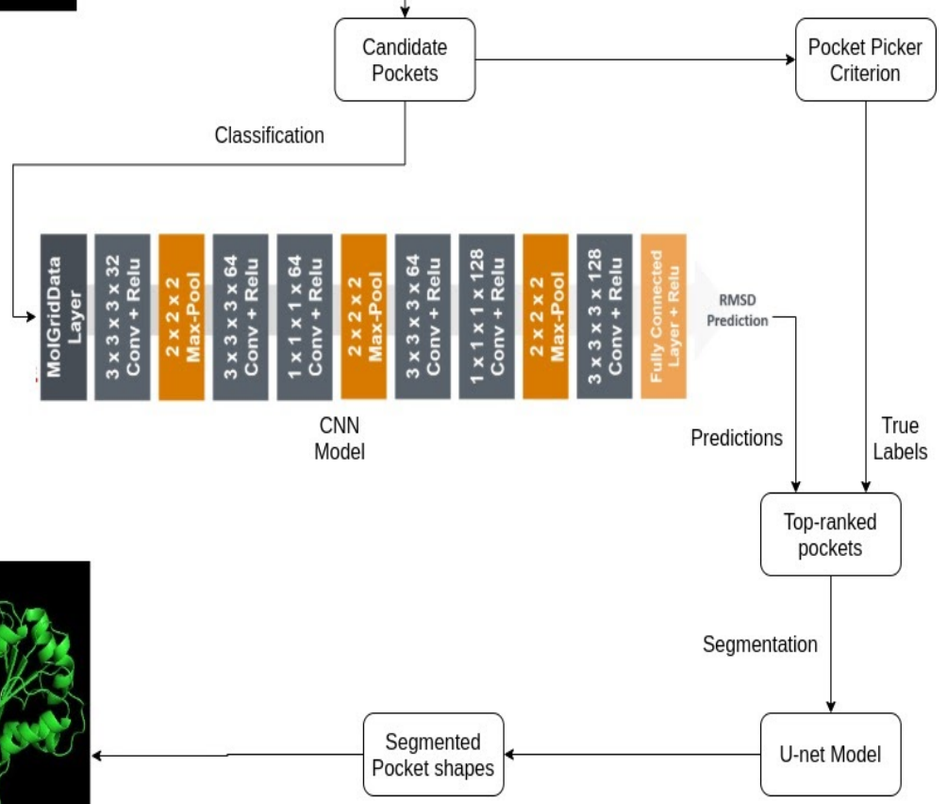
- Protein structure prediction
- Binding site identification
- Hit identification
- de novo design
- ADME/T


Protein Sequence
SQETRKKCTEMKKKFKNCEVRCDESNHCVEVRCSDTKYTLC
Neural Network
Databases
Distance Predictions
Angle Predictions
Score
Gradient Descent
Structure
Protein Structure Prediction
deepmind.com

Binding Pocket Prediction

Detecting the location of Ligand Binding Site on 3D Structure of Proteins





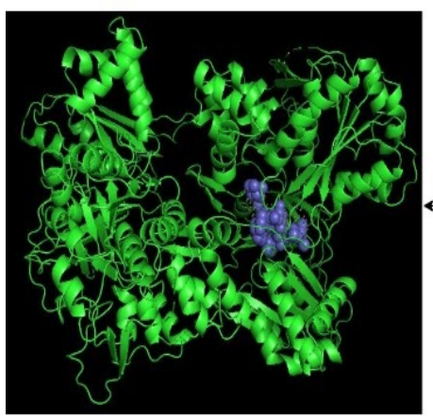
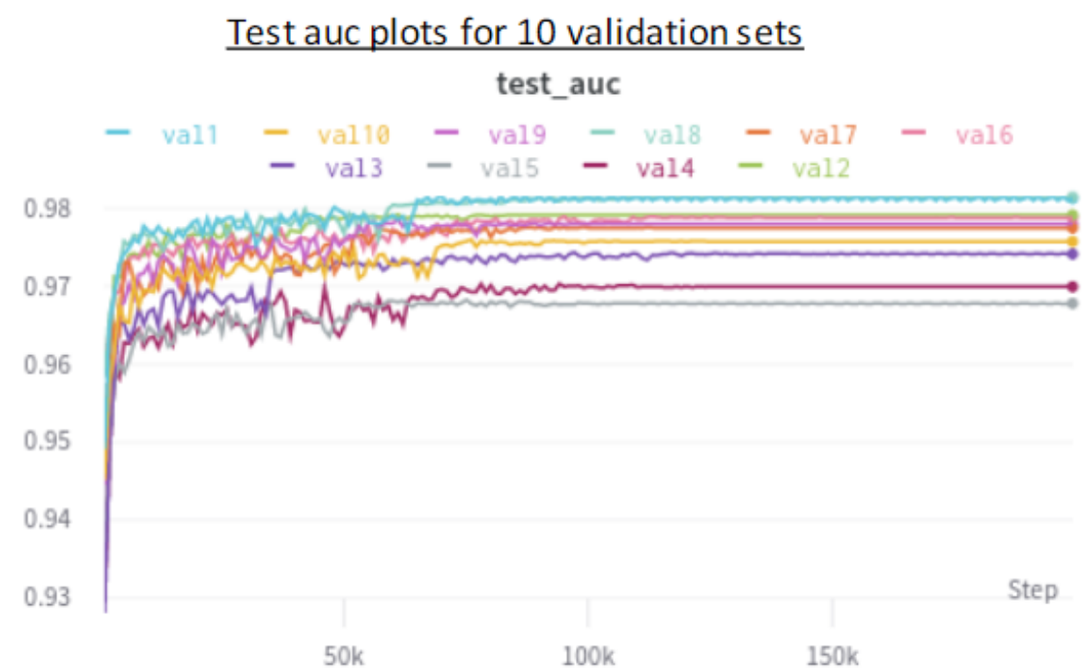


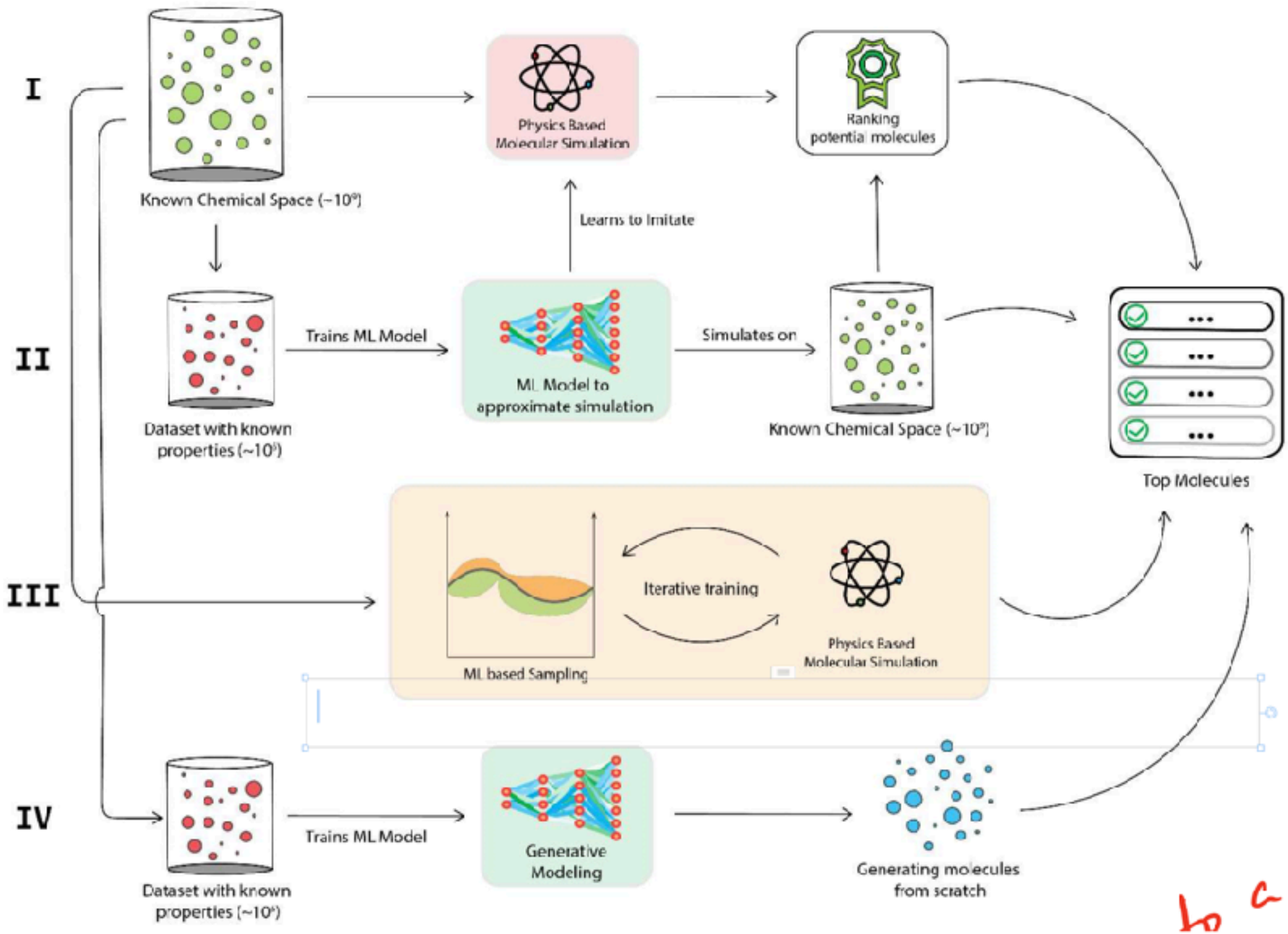
Table 1: PPC results comparison

Methods	COACH420		HOLO4K		2018+	
	Top-n	Top-(n+2)	Top-n	Top-(n+2)	Top-n	Top-(n+2)
Fpocket	35.09%	51.25%	36.34%	51.53%	23.99%	37.23%
Deepsite	53.07%	53.07%	51.65%	51.67%	52.94%	65.41%
P2Rank	68.24%	75.48%	70.6%	80.05%	62.9%	75.74%
Kalasanty	**71.86%**	71.86%	71.5%	71.53%	**69.31%**	73.55%
Deepocket	67.96%	**79.94%**	**73.36%**	**82.97%**	64.58%	**83.01%**



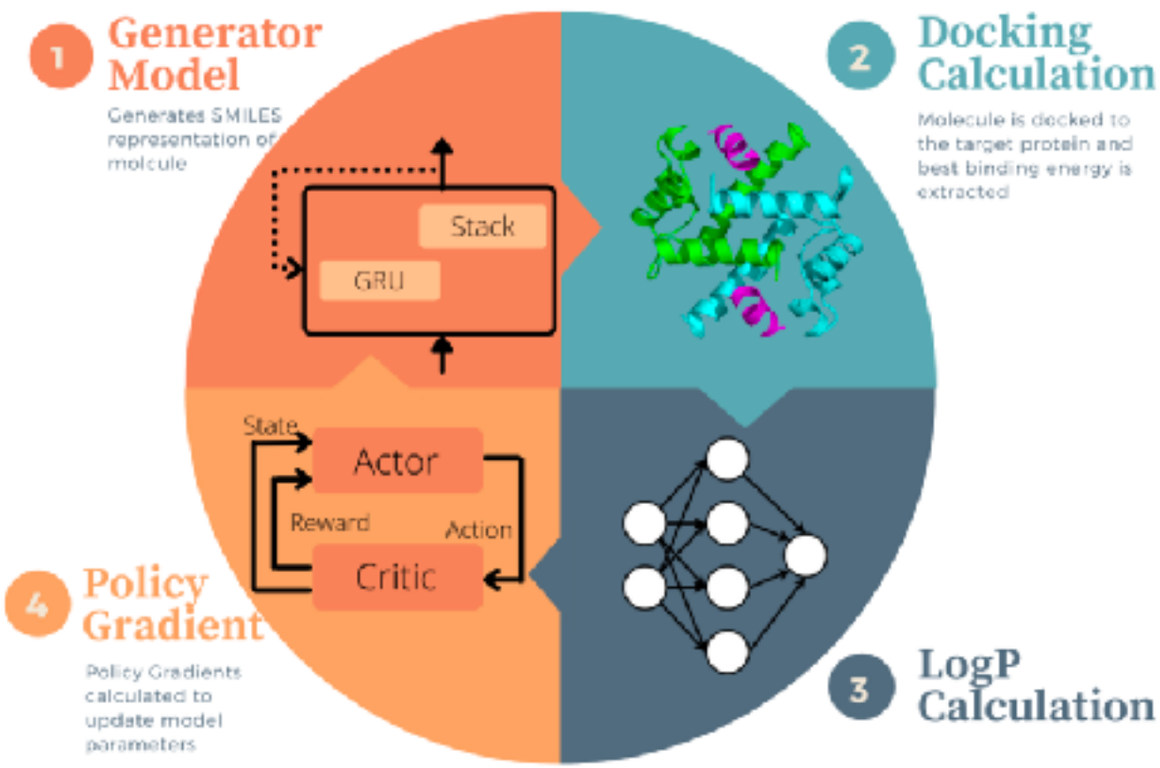
JCIM 2021

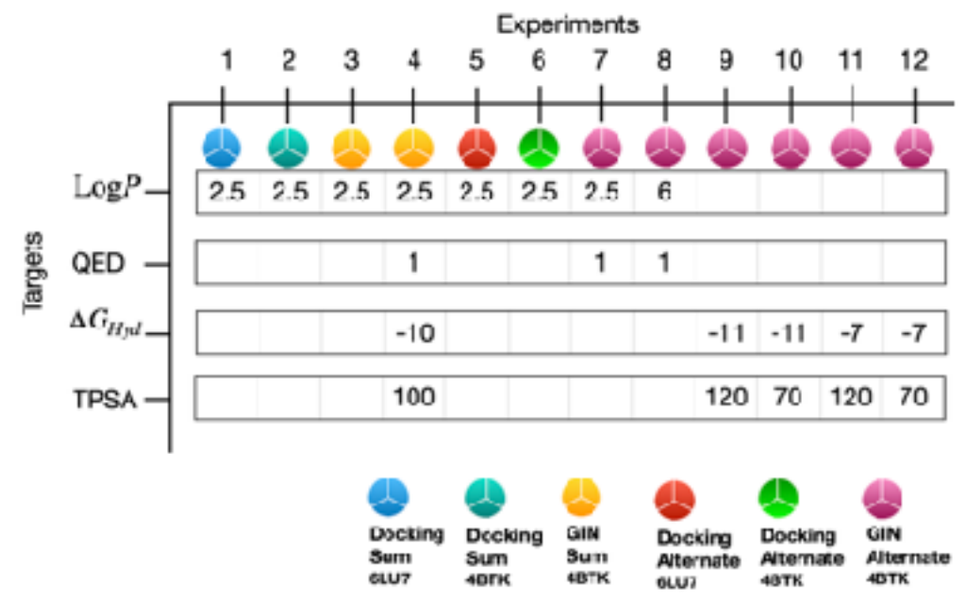
Hit Identification

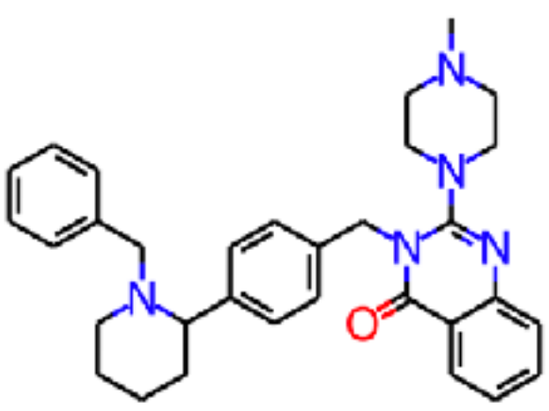


de novo Generation

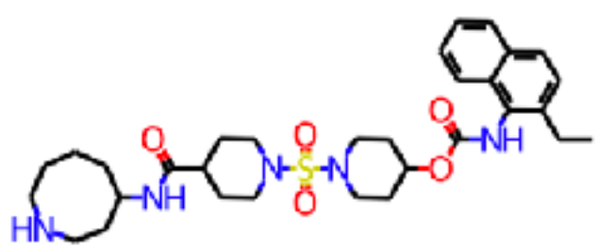
A Reinforcement Learning Guided Method to Generate Molecules using stack augmented RNN that Dock Well to Drug Targets



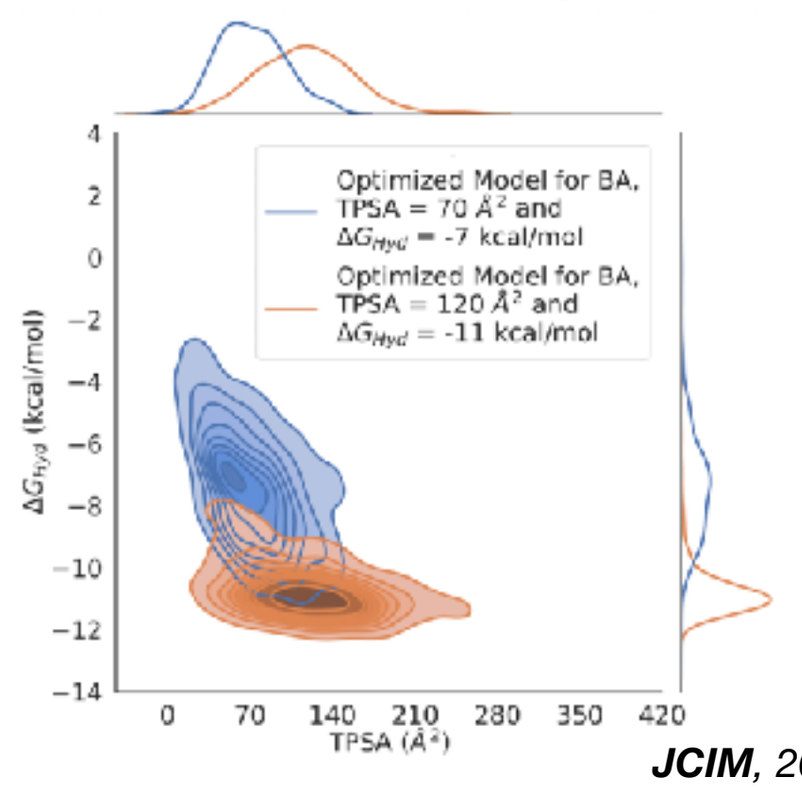




BE: -8.36 LogP: 4.92



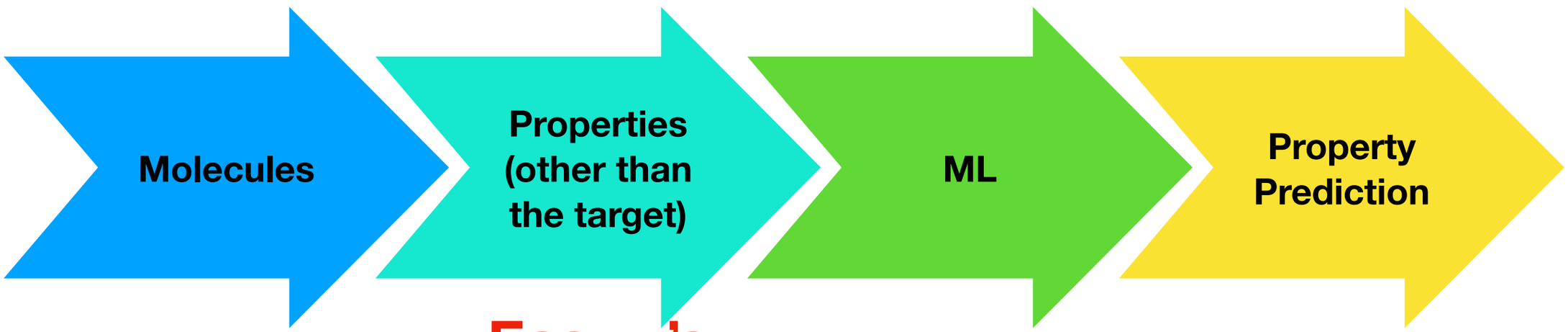
BE: -8.16 LogP: 4.02



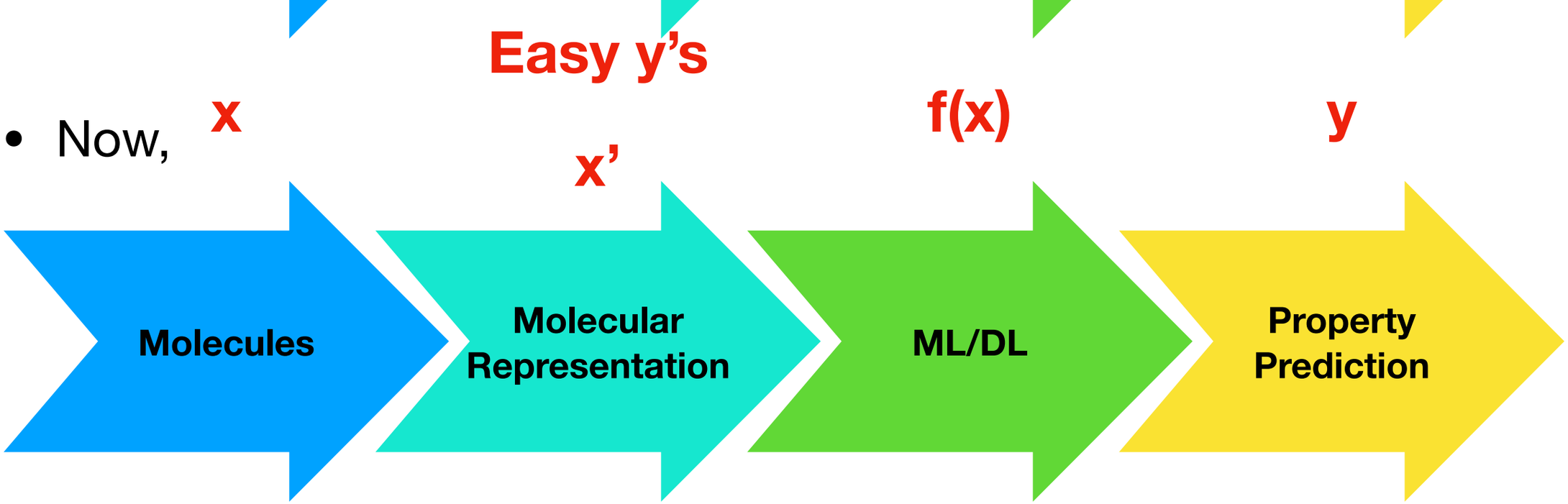
***JCIM**, 2022, xxxx*

- As mentioned before, we have been using ML in drug design for a long time

- Typically,



- Now,



Feature Engineering

Added after the lecture

References on slide 32

- Target Identification: https://www.nature.com/articles/s41467-019-12928-6
- Protein structure prediction: https://www.nature.com/articles/s41586-021-03819-2
- de novo generation: https://pubs.rsc.org/en/content/articlelanding/2022/cc/d1cc07035e/unauth & https://www.science.org/doi/10.1126/sciadv.aap7885
- Retrosynthesis: https://www.nature.com/articles/nature25978
- AI assisted synthesis: https://www.science.org/doi/10.1126/science.aax1566
- Spectra to molecule: https://chemrxiv.org/engage/chemrxiv/article-details/61b9b0a87f367e767d53a9a9 & https://www.tandfonline.com/doi/full/10.1080/14686996.2020.1793382
- Clinical trial outcome prediction: https://hdsr.mitpress.mit.edu/pub/ct67j043/release/10