# Artificial Intelligence in Drug Design and Synthesis


**Dr. Balakumar Chandrasekaran** **M. Pharm, PhD, PostDoc**

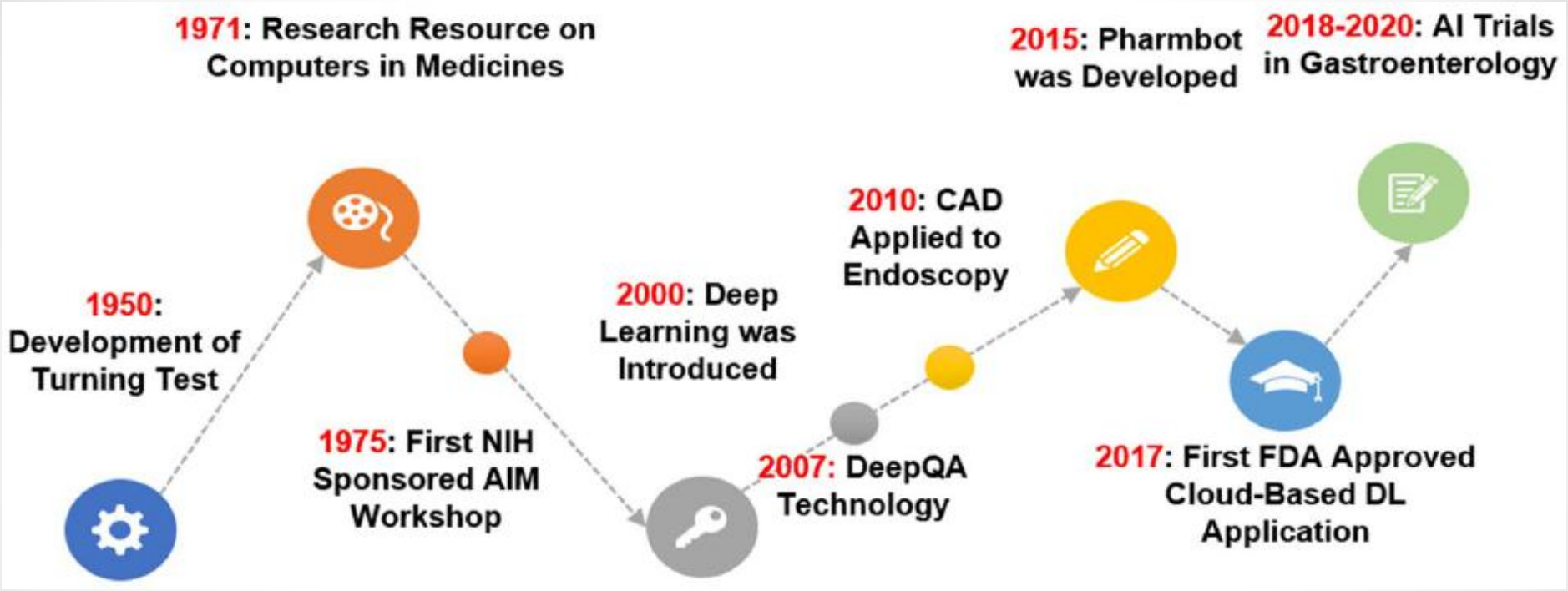
**Professor-Faculty of Pharmacy**

**Philadelphia University-Jordan**

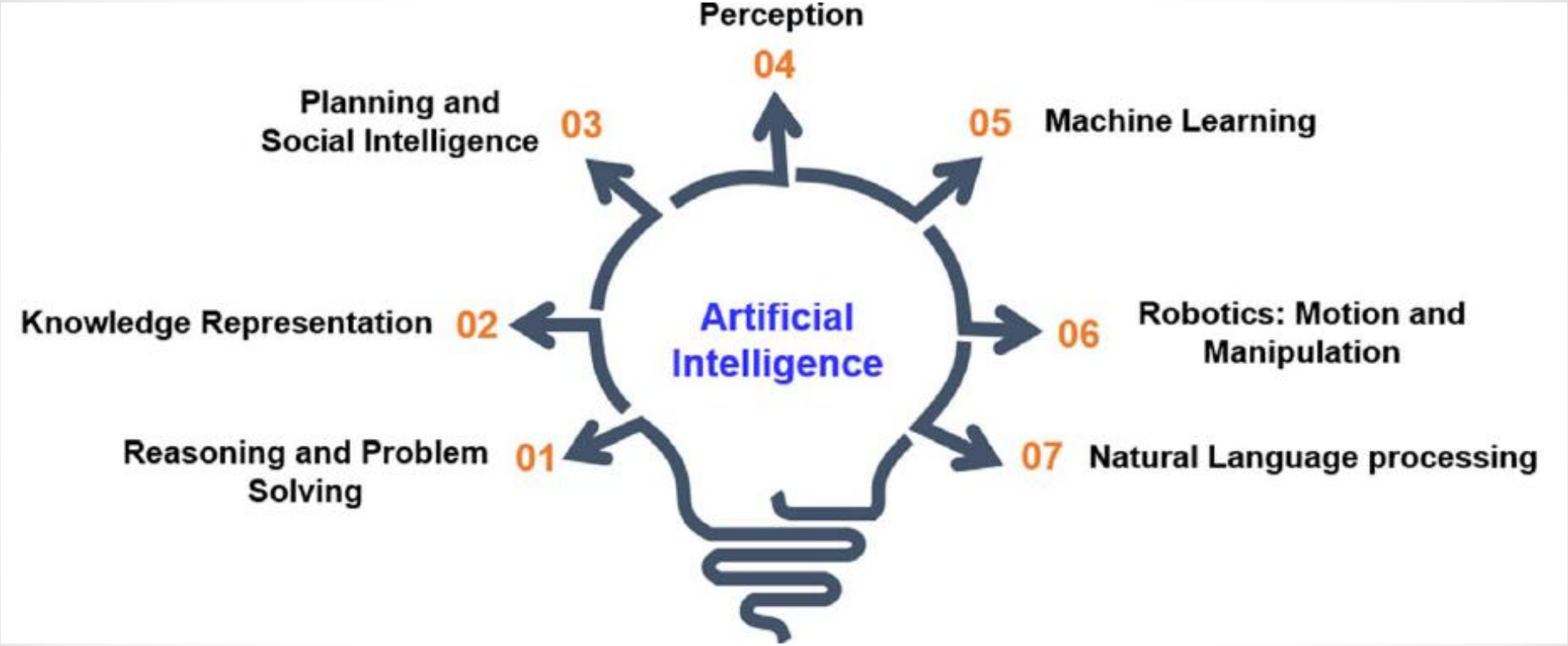
# Artificial Intelligence (AI)

- AI is a technology-based system involving various advanced tools and networks that can mimic human intelligence.
- AI utilizes systems and software that can interpret and learn from the input data to make independent decisions for accomplishing specific objectives.
- AI involves several method domains like reasoning, knowledge representation, solution search, and machine learning (ML).
- A subfield of the ML is deep learning (DL), which engages artificial neural networks (ANNs).

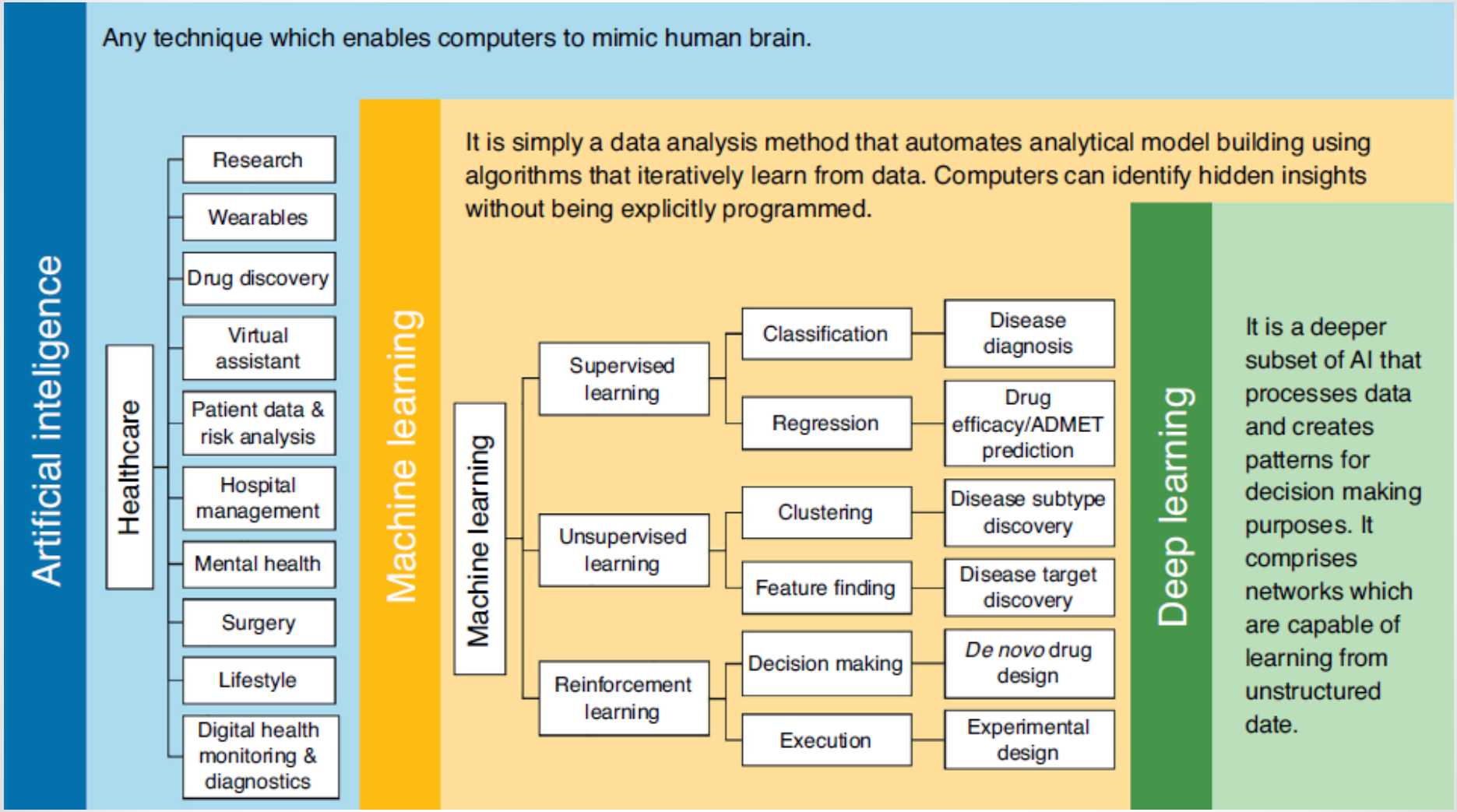
# Artificial Intelligence: History



# Artificial Intelligence: Classification



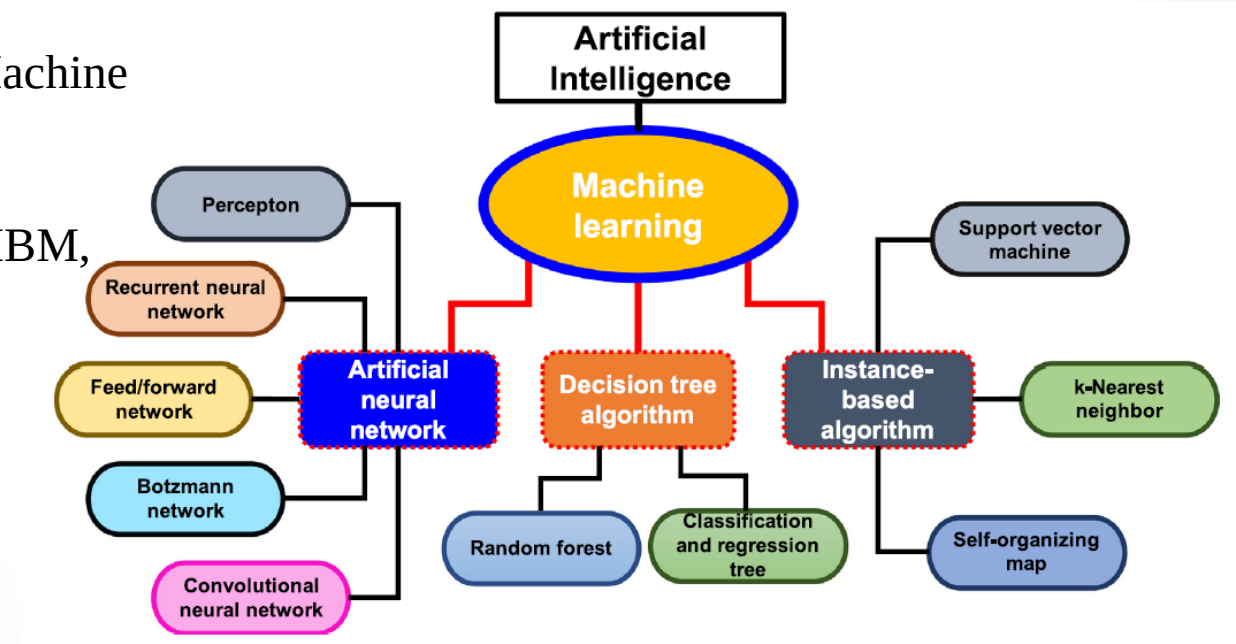
# Artificial Intelligence: Application areas



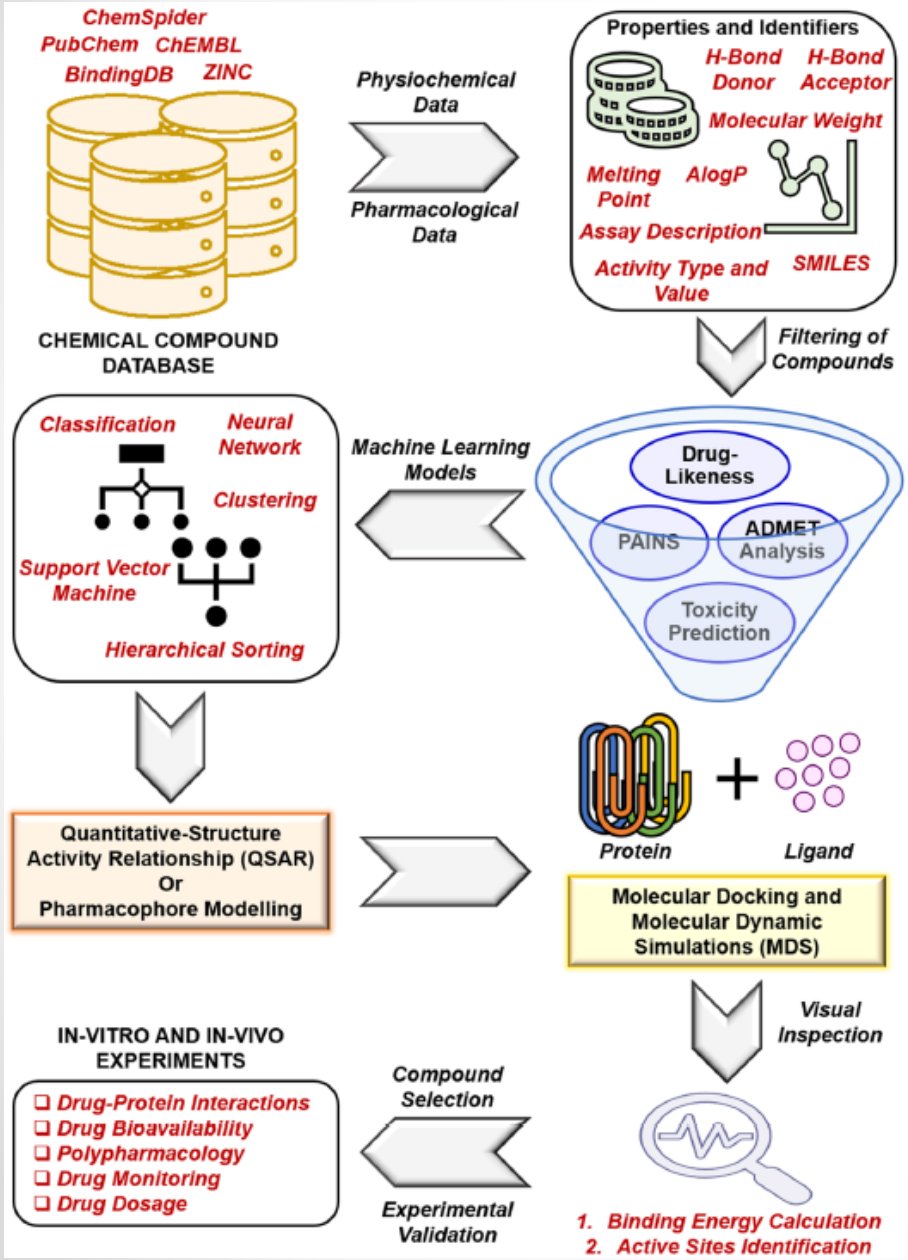
# Artificial Intelligence

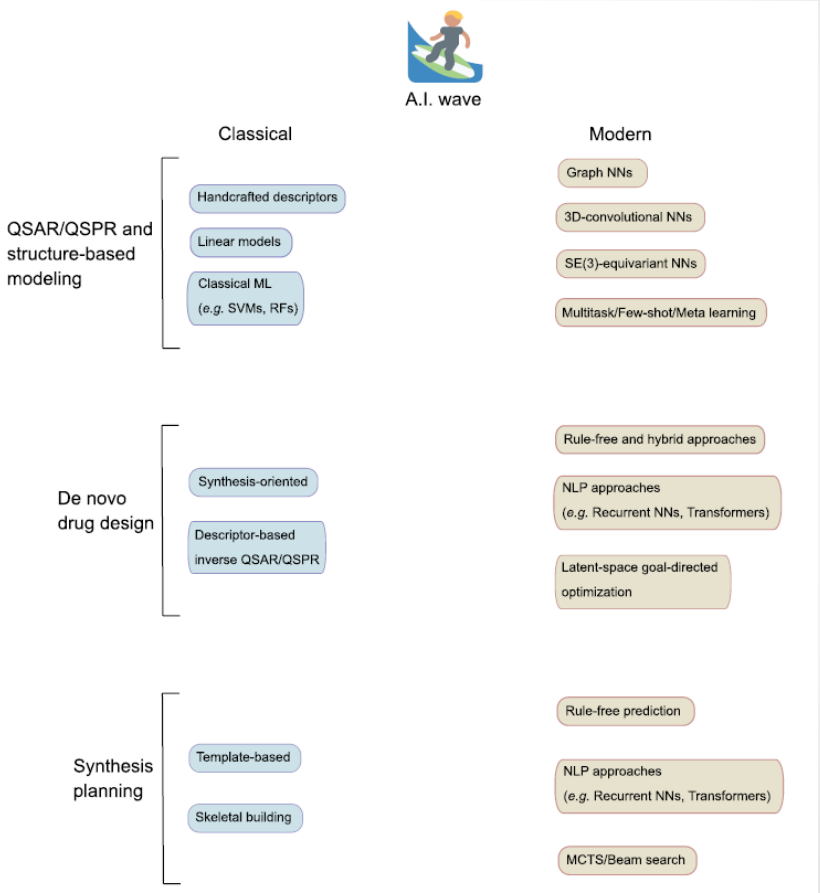
- These comprise a set of interconnected sophisticated computing elements involving 'perceptons' analogous to human biological neurons, mimicking the transmission of electrical impulses in the human brain.

- International Business Machine (IBM)
- Watson supercomputer (IBM, New York, USA).
- Suggesting treatment strategies for cancer and its detection.



# Artificial Intelligence in Drug Discovery

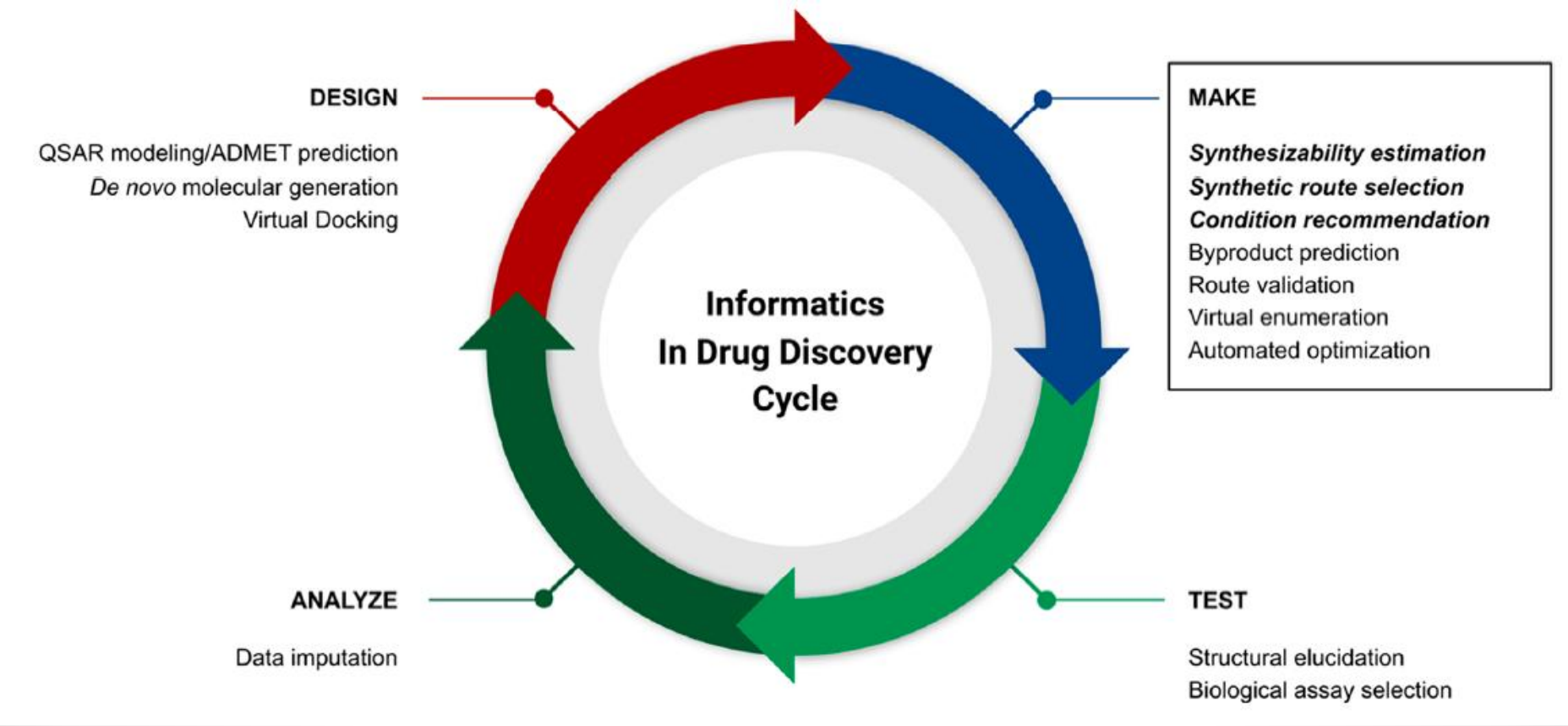


# AI in Drug Discovery

| | | | |
|---|---|---|---|
| Drug Design | Synthesis | Drug repurposing | Pharmacology |
| Predict target prt | Pdt Yield | Predict new target | Predict activity |
| Predict prt-drug interaction | Rxn Mechanism | Predict new use | Predict toxicity |
| Predict activity | Synthetic route | Design new drug | Predict Physico-chem prop |

# AI in Drug Design & Synthesis




FOUR CRUCIAL STAGES: 1. Design, 2. Synthesis, 3. Testing,
4. Analyzing of new drugs.

# AI in Drug Design: Prediction tools for target protein structures

- AlphaFold: https://deepmind.com/blog/alphafold

  It is based on deep neural network (DNN)

  It is used to analyze the distance between the adjacent amino acids and the corresponding angles of the peptide bonds.

- PotentialNet:-

  https://pubs.acs.org/doi/full/10.1021/acscentsci.8b00507

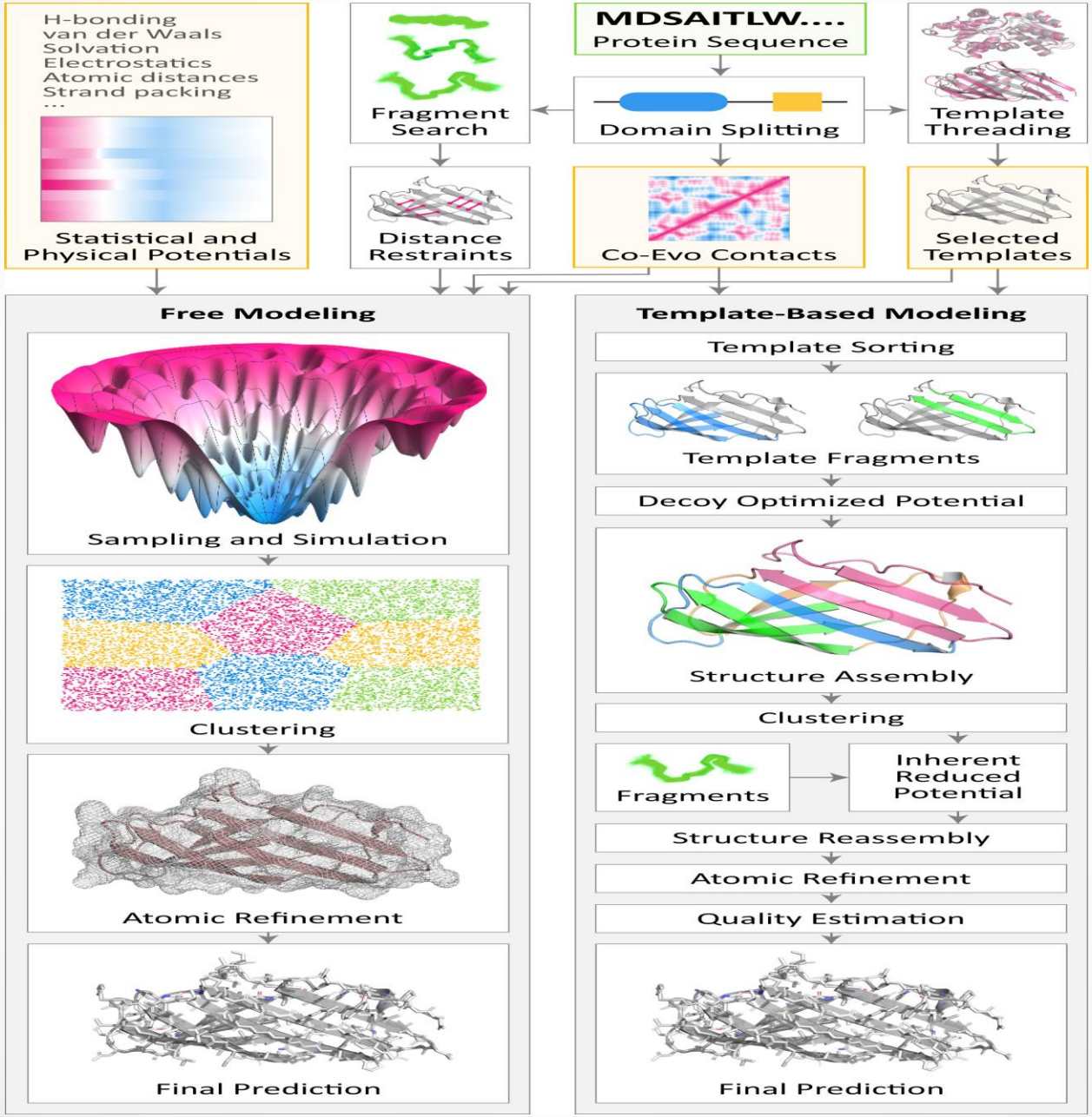
  It is based on neural network (NN).

  It is used to predict the binding affinity of the ligand.

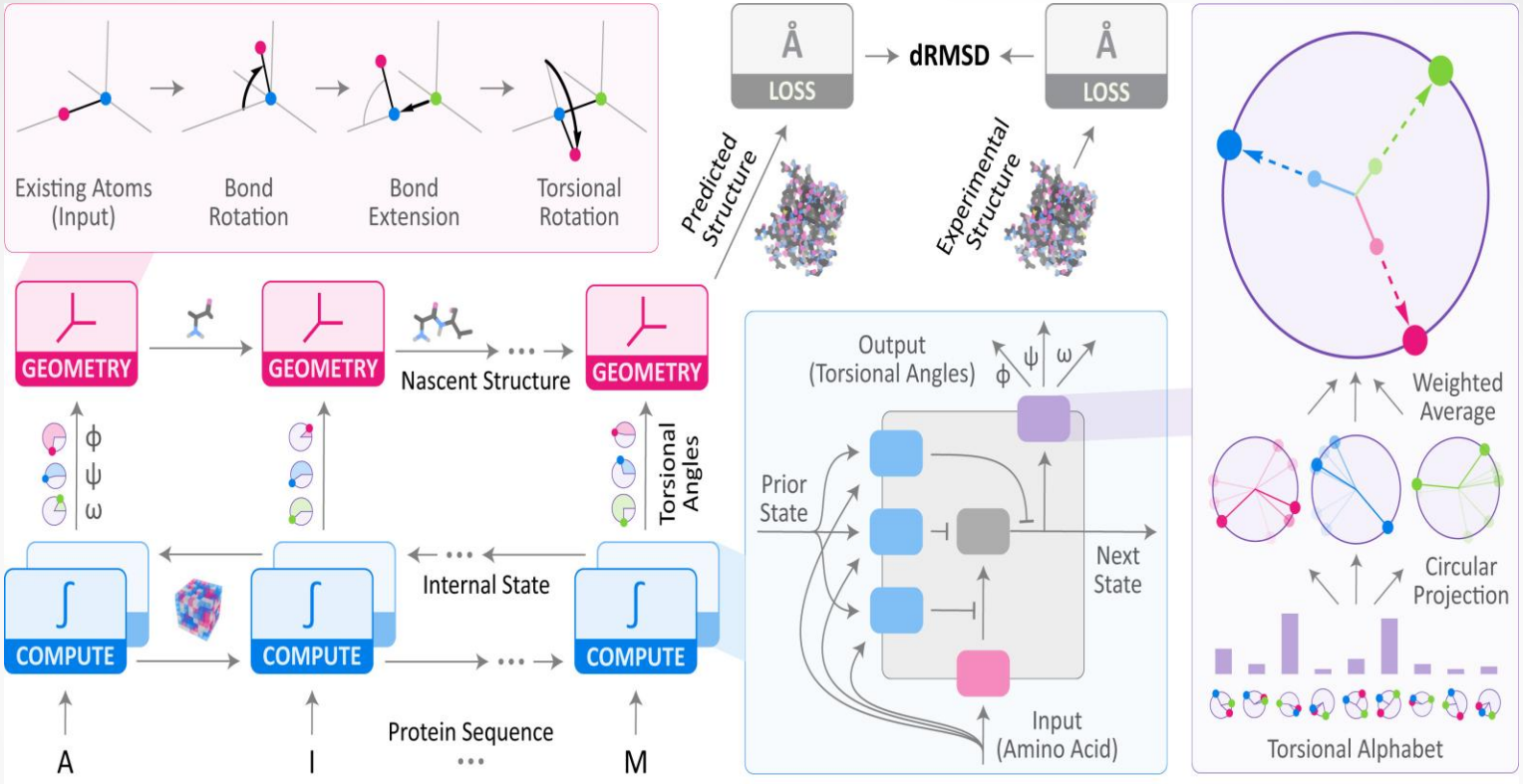
- DeltaVina: https://github.com/chengwang88/deltavina

  It is a scoring function for rescoring drug–ligand binding affinity.

# Prediction tools for target protein structures: Conventional techniques



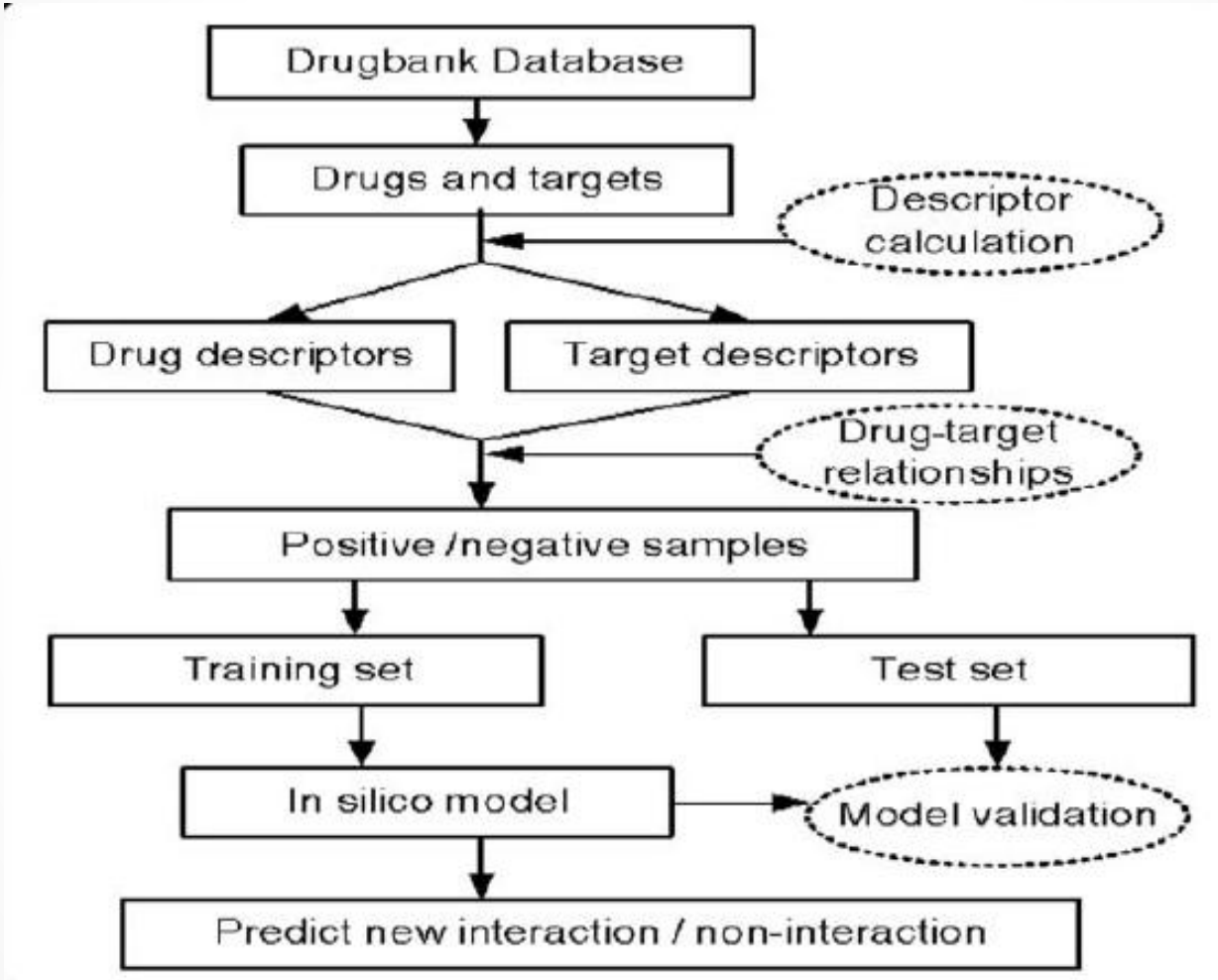
# Prediction tools for target protein structures: Recurrent Geometric network (RGN)




- Protein sequences are fed one residue at a time to the computational units of an RGN. [*Cell Systems* **2019**, *8*, 292].
- Based on these computations, torsional angles are predicted and fed to geometric units.
- dRMSD is used to measure deviation from experimental structures, serving as the signal for optimizing RGN.

**Prediction of drug-protein interactions: (i) Random Forest (RF)**

**(ii) Support Vector Machine (SVM).**




Flow chart of the modeling process.

## Prediction of drug-protein interactions: (i) Random Forest (RF)

## (ii) Support Vector Machine (SVM).

### Predicted results for a model in RF

| Dataset | SE (RF/BGL) | SP(RF/BGL) | CO(RF/BGL) | AUC(RF/BGL) |
|---|---|---|---|---|
| Enzyme | 35.82%/57.40% | 82.70%/99.50% | 59.26%/-- | 67.43/90.40 |
| GPCR | 80.31%/23.40% | 55.64%/99.90% | 67.98%/-- | 72.95/89.90 |
| Ion channel | 54.09%/27.10% | 73.38%/99.60% | 63.73%/-- | 66.58/85.10 |
| Nuclear receptor | 91.57%/14.80% | 39.76%/99.90% | 65.66%/-- | 82.29/84.30 |
| Average | 47.51%/— | 74.93%/— | 61.64%/-- | 66.68/-- |

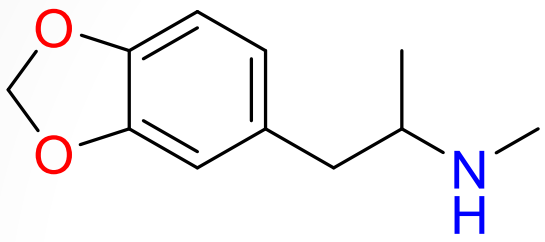
Dataset: DrugBank database
Descriptor: DRAGON
PROFEAT WEBSEVER
Random Forest algorithm
SE: sensitivity
SP: specificity
CO: compound-protein pairs
BGL: Bipartite graph learning

### Predicted results for top 5 scoring novel drug-target interactions

| Protein name (UniProt ID) | Drug generic name (DrugBank ID) | Binding score |
|---|---|---|
| NAD(P)H dehydrogenase [quinone] 1 (P15559) | Flavin-N7 protonated-adenine dinucleotide (DB02332) | 0.996 |
| NAD(P)H dehydrogenase [quinone] 1 (P15559) | NADH (DB00157) | 0.994 |
| Alcohol dehydrogenase [NADP+] (P14550) | NADH (DB00157) | 0.992 |
| Prostaglandin G/H synthase 1 (P23219) | Bromfenac (DB00963) | 0.992 |
| Prostaglandin G/H synthase 1 (P23219) | D-allopyranose (DB03989) | 0.990 |

# Novel Target Prediction for an existing drug: MDMA

3,4-Methylene-dioxy-methamphetamine (MDMA)
Psychoactive drug




Predicted top 5 scoring target proteins for the same drug MDMA

| Protein name | UniProt ID | Binding score |
|---|---|---|
| Beta-1 adrenergic receptor | P08588 | 0.820 |
| Carbonic anhydrase 2 | P00918 | 0.810 |
| Prothrombin | P00734 | 0.804 |
| Alpha-2A adrenergic receptor | P08913 | 0.802 |
| Prostaglandin G/H synthase 2 | P35354 | 0.800 |
| Dipeptidyl peptidase 4 | P27284 | 0.798 |

# Novel Drug Prediction for an existing target:

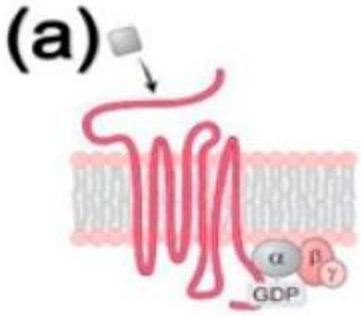
# Thymidine kinase

Predicted top 10 scoring drugs for the same target: Thymidine kinase

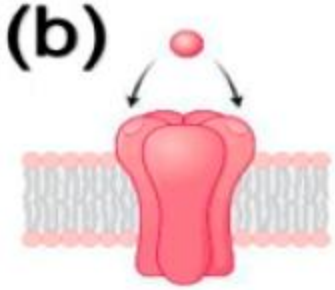
| Drug generic name | DrugBank ID | Binding score |
|---|---|---|
| NADH | DB00157 | 0.870 |
| Nicotinamide-Adenine-Dinucleotide | DB01907 | 0.848 |
| Adenosine-5′-Diphosphate | DB03431 | 0.844 |
| Guanosine-5′-Diphosphate | DB04315 | 0.808 |
| Acetate Ion | DB04184 | 0.792 |
| Mesoheme | DB02577 | 0.786 |
| Heme | DB03014 | 0.786 |
| Idoxuridine | DB00249 | 0.762 |
| Pentostatin | DB00552 | 0.656 |
| 1-Beta-Ribofuranosyl-1,3-Diazepinone | DB03185 | 0.622 |

[*PLoS One* **2012**, *7*, e37608].

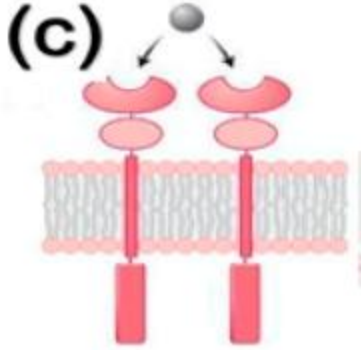
# Software tools for Drug-Receptor Interaction Prediction



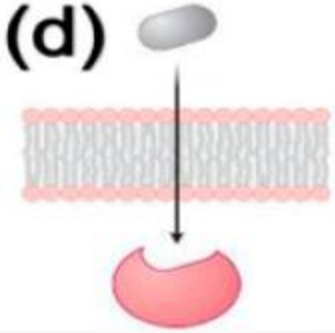

iDrug-GPCR:
The web-server for predicting the interaction between GPCRs and drugs in cellular networking.




iDrug-Chl:
The web-server for predicting the interaction between ion channels and drugs in cellular networking.




iDrug-Ezy:
The web-server for predicting the interaction between enzymes and drugs in cellular networking.




iDrug-NR:
The web-server for predicting the interaction between nuclear receptors and drugs in cellular networking.

[*http://www.jci-bioinfo.cn/iDrug-Target/*]

# Prediction of pharmacological activity/affinity of drugs

- DeepNeuralNetQSAR: https://github.com/Merck/DeepNeuralNet-QSAR

  It utilizes python-based tools and is used to detect molecular activity of a compound.

- Neural graph fingerprint: https://github.com/HIPS/neural-fingerprint

  It is used to predict the properties of novel molecules.

- DeepTox: www.bioinf.jku.at/research/DeepTox

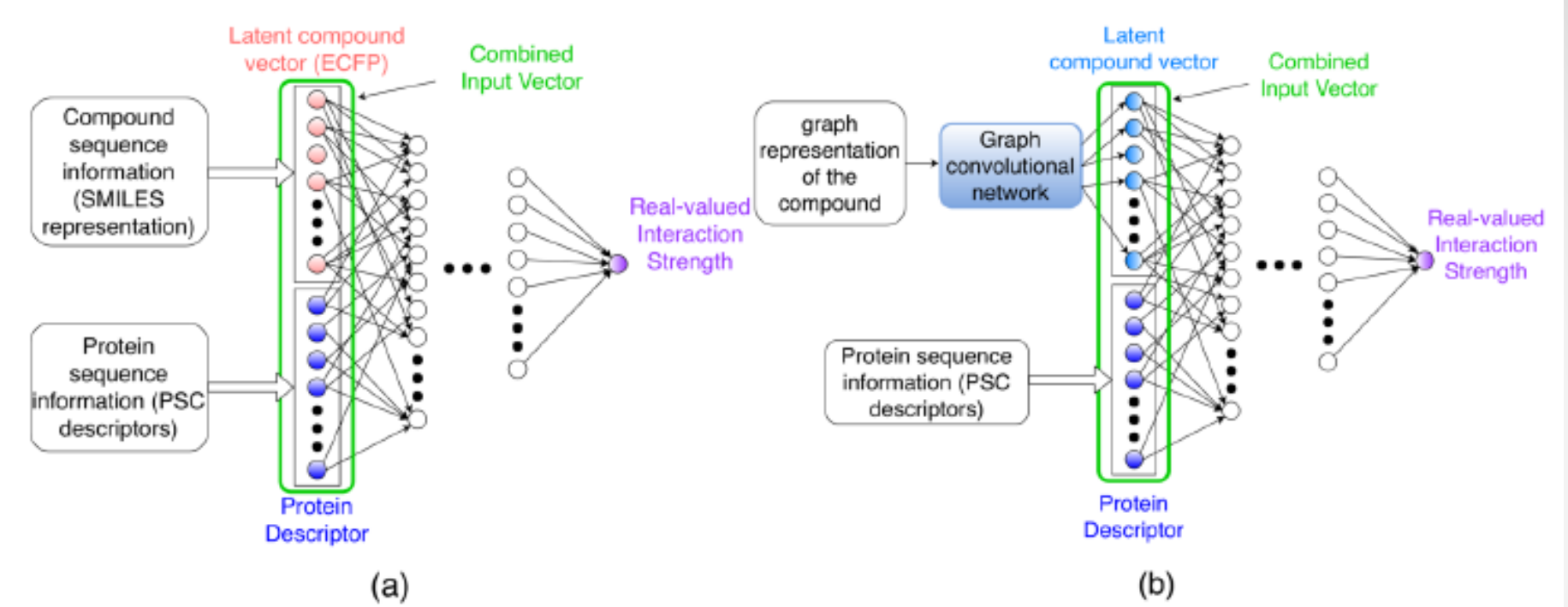
  Software that predicts the toxicity of 12 000 drugs

DeepDTA, PADME, WideDTA, and DeepAffinity are some DL methods used to measure Drug-target binding affinity (DTBA).

# PADME: A Deep Learning-based Framework for Drug-Target Interaction Prediction

- PADME: Protein And Drug Molecule interaction prEdiction.
- It is used to predict real-valued **interaction strength** between compounds and proteins.
- PADME takes both compound and protein information as inputs to solve cold-target (problems involving target protein that never appeared in the training set) or cold-drug problems.
- A study by Feng et al., integrated Molecular Graph Convolution (MGC) for compound featurization with protein descriptors.
- Authors used multiple cross-validation split schemes and evaluation metrics to measure the performance of PADME.
- The success of the PADME tool was demonstrated by taking an example of androgen receptor (AR) against predicting the binding affinity between various compounds.

# Methodology




(a) PADME-ECFP (extended connectivity finger print) architecture.

(b) PADME-GraphConv architecture.

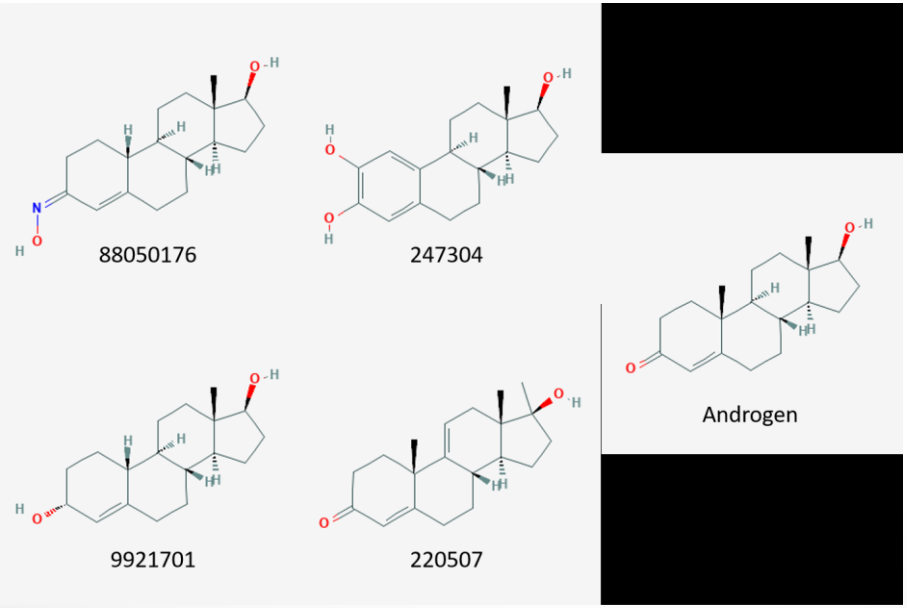
Black dots represent omitted neurons and layers;

CIV: Combined Input Vector.

Compound feature vectors generated using DNN; graph representation of molecules, the atoms are denoted by nodes, while the bonds are denoted by edges.

# A case study: Androgen receptor (AR)

- Ligands: US National Cancer Institute human tumour cell line anticancer drug screen data (NCI60), totaling more than 100000.
- PADME-ECFP and PADME-GraphConv trained on whole ToxCast dataset.
- Average of their predictions were determined, known as averaged model PADME-Ensemble.
- Top 30 compounds are predicted to bind strongly with AR, out of that 4 compounds are selected as active.




[*arXiv:1807.09741*, **2008**].

# AI in Synthesis of drugs

- Chemputer: https://zenodo.org/record/1481731

  It helps to report procedure for chemical synthesis in standardized format.

- DeepChem: https://github.com/deepchem/deepchem

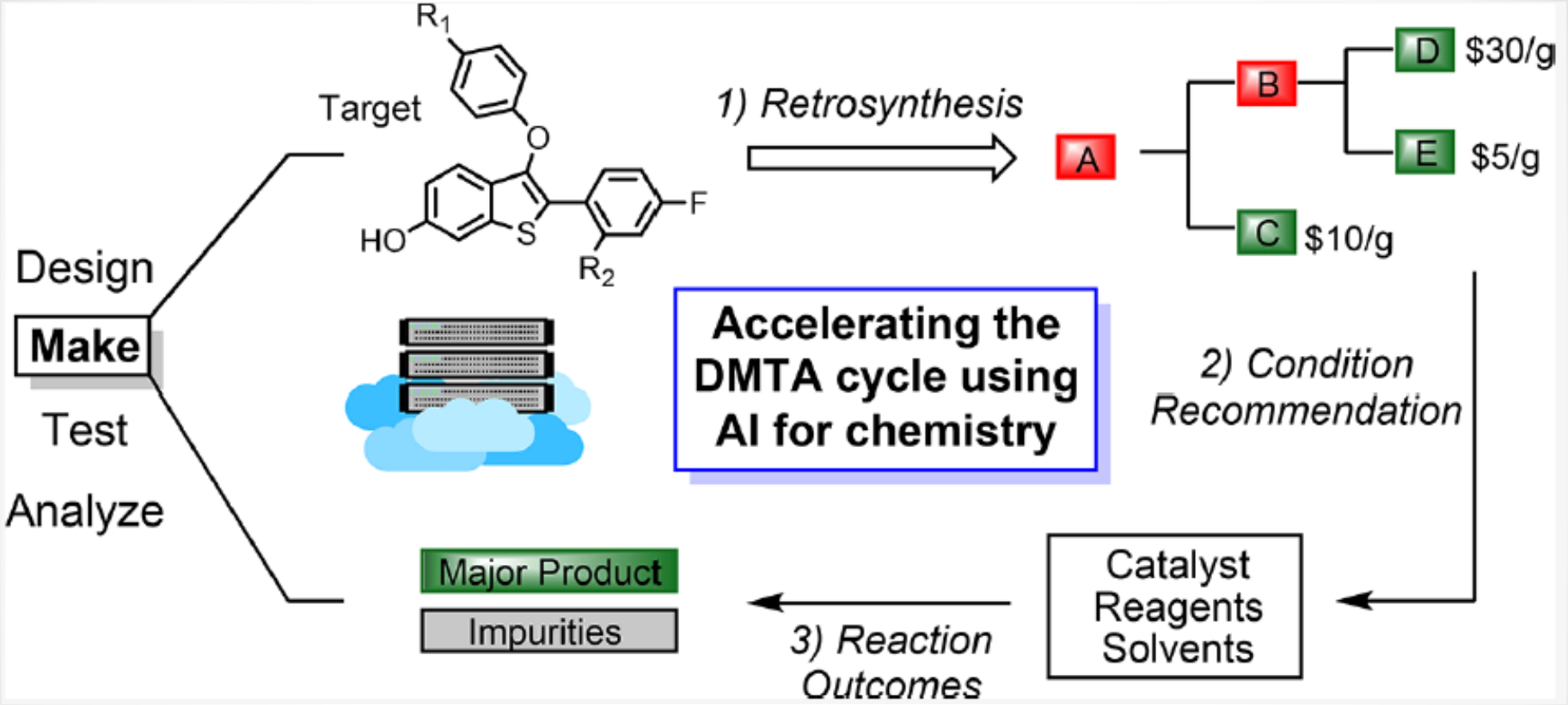
  It is a multilayer perception (MLP) model that uses a python-based AI system to find a suitable candidate in drug discovery program.

- ORGANIC: https://github.com/aspuru-guzik-group/ORGANIC

  A molecular generation tool that helps to create molecules with desired properties.

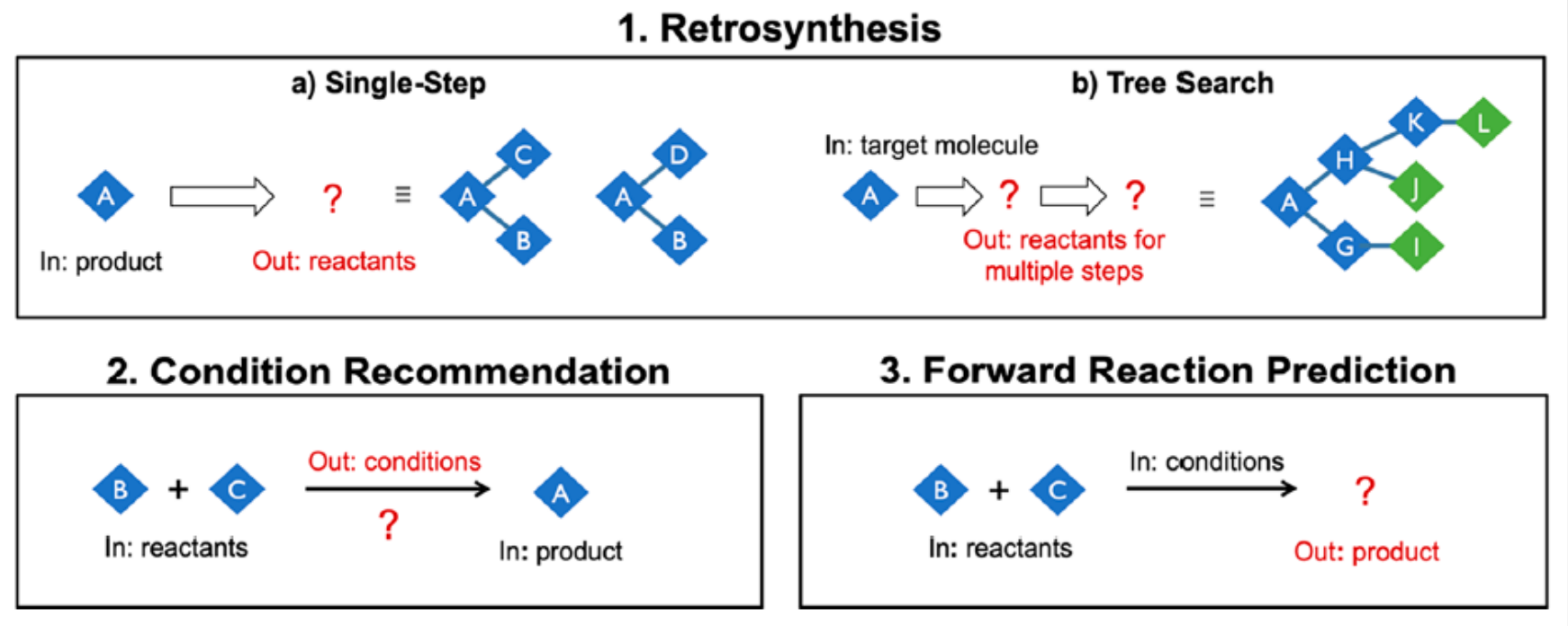
- Chematica and ICSynth: Regularly used software tools in industries.

# AI in Synthesis of drugs




**CASP**: Computer-aided synthesis planning

# AI in Synthesis of drugs: 1. Retrosynthesis



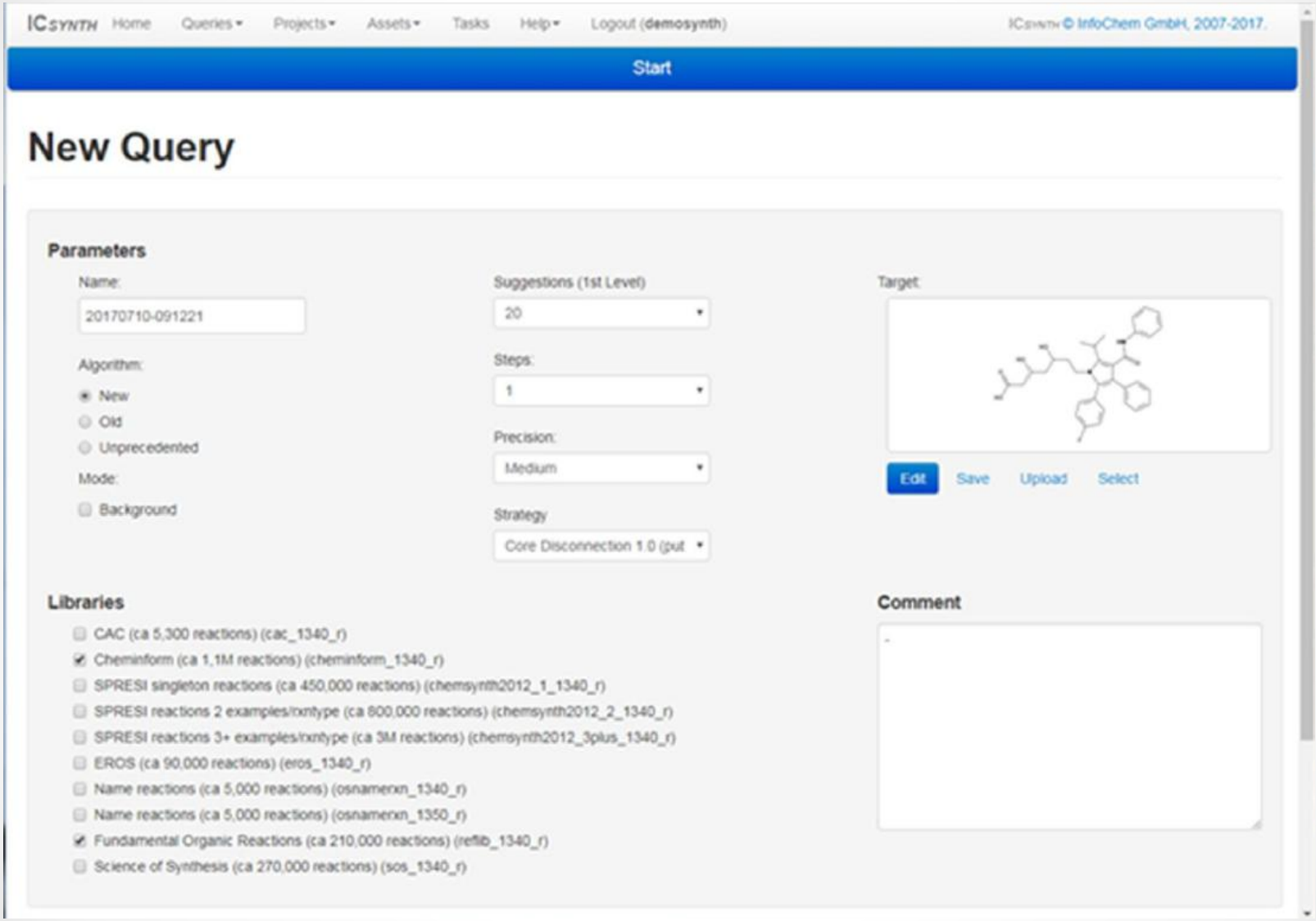

(1) **Retrosynthesis** can be broken into subproblems of (a) the generation of retrosynthetic suggestions one step at a time and (b) the recursive use of the singe step suggestions to identify full, multistep routes.

(2) **Reaction conditions** that will lead to a successful forward reaction must be recommended.

(3) **Reaction prediction** of the possible products from a set of starting materials and conditions for the proposed synthetic steps.

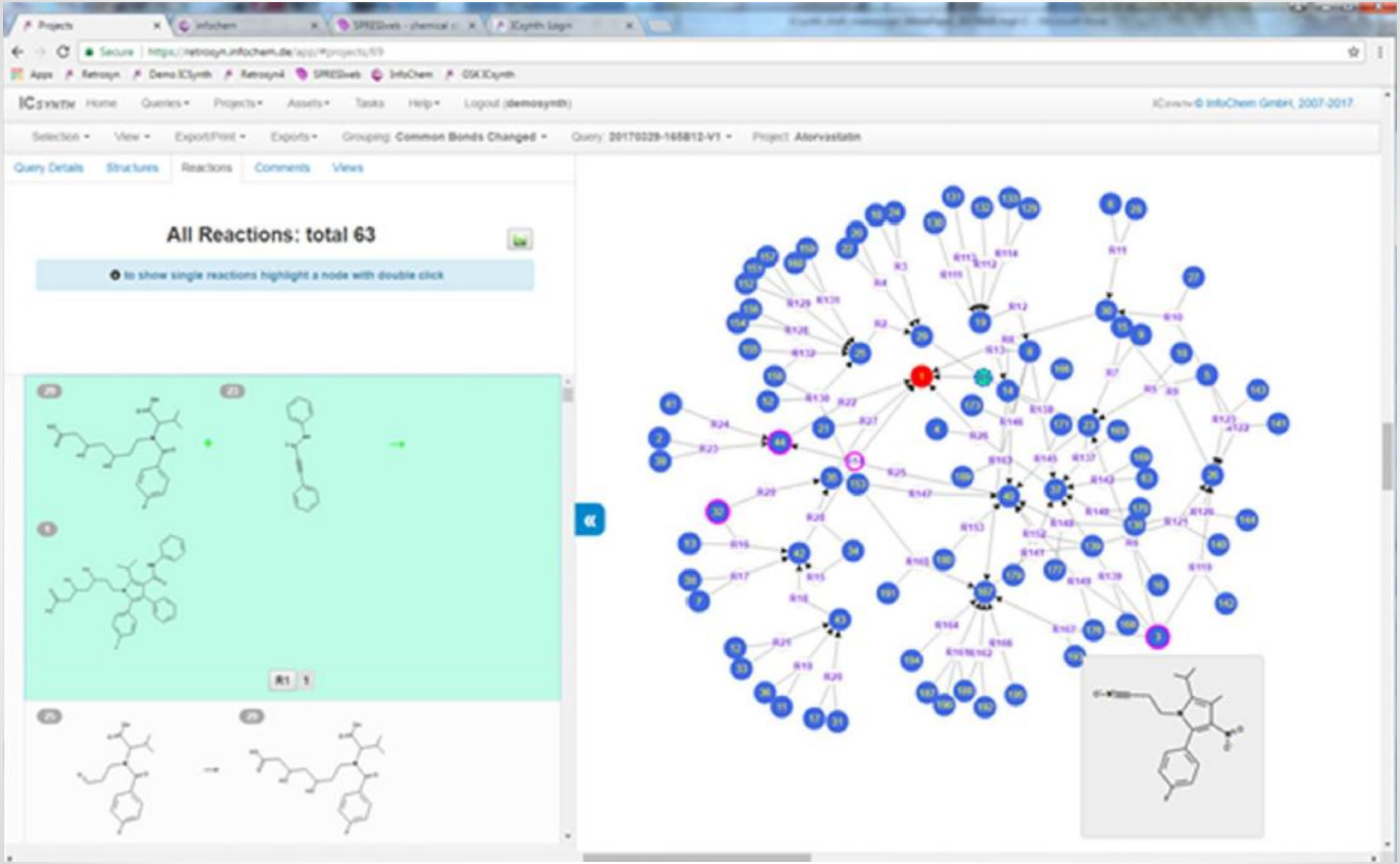
- Expert encoded rules (first wave-AI) and ML methods (second wave-AI) are AI tools in synthesis.
- First wave-AI: Crafted knowledge
- Second wave-AI: Statistical methods
- Two categories of methods for scoring compounds by synthesizability: simplified structure-based heuristics or full retrosynthetic tree expansions.
- A general procedure for the algorithmic extraction of reaction templates from a reaction data set :
  (1) identify the reaction center or changing atoms
  (2) identify atoms adjacent to the reaction center
  (3) add generalized functional groups involved in the reaction.
- A single-step retrosynthetic recommender is sufficient to construct routes for one-step reaction at a time.
- Similarly, it can be extended to derive synthetic route for multistep organic synthesis using a tree search.
- Each step can produce thousands of precursors.

# Retrosynthesis: ICSYNTH Software by InfoChem



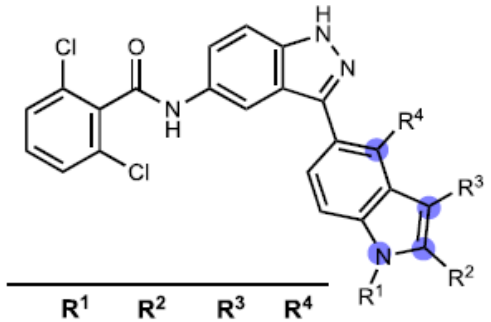

**Search set up**: ICSYNTH (rule-based method)

# Retrosynthesis: ICSYNTH Software by InfoChem



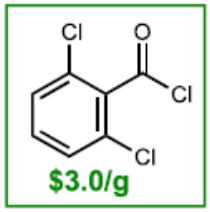

**Results visualization**: Dynamic reaction graph

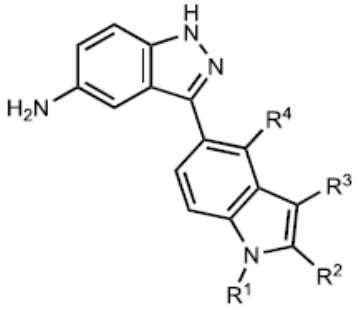
# Retrosynthetic analysis using ASKCOS software



| | $R^1$ | $R^2$ | $R^3$ | $R^4$ |
|---|---|---|---|---|
| **A** | -H | -H | -H | -H |
| **8** | -Me | -H | -H | -H |
| **9** | -H | -Me | -H | -H |
| **10** | -H | -H | -Me | -H |
| **11** | -H | -H | -H | -Me |

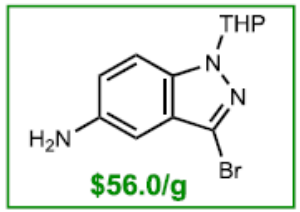
Acylation

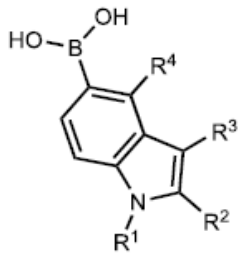


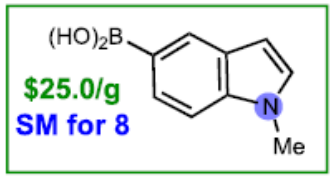



C-C, sp2-sp2 coupling

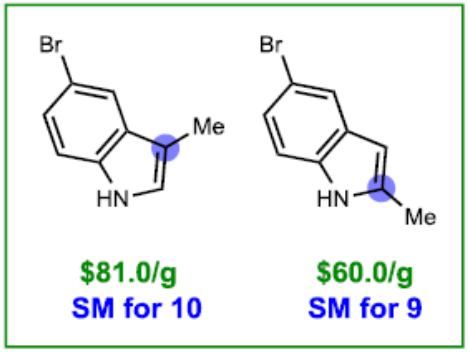
THP deprotection



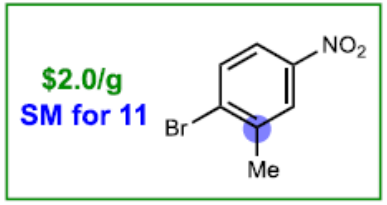


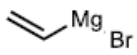

borylation




NOT AVAILABLE

heterocycle synthesis (Bartoli reaction)



**ASKCOS**: It proposes the possible synthetic route based on the availability of the starting material.

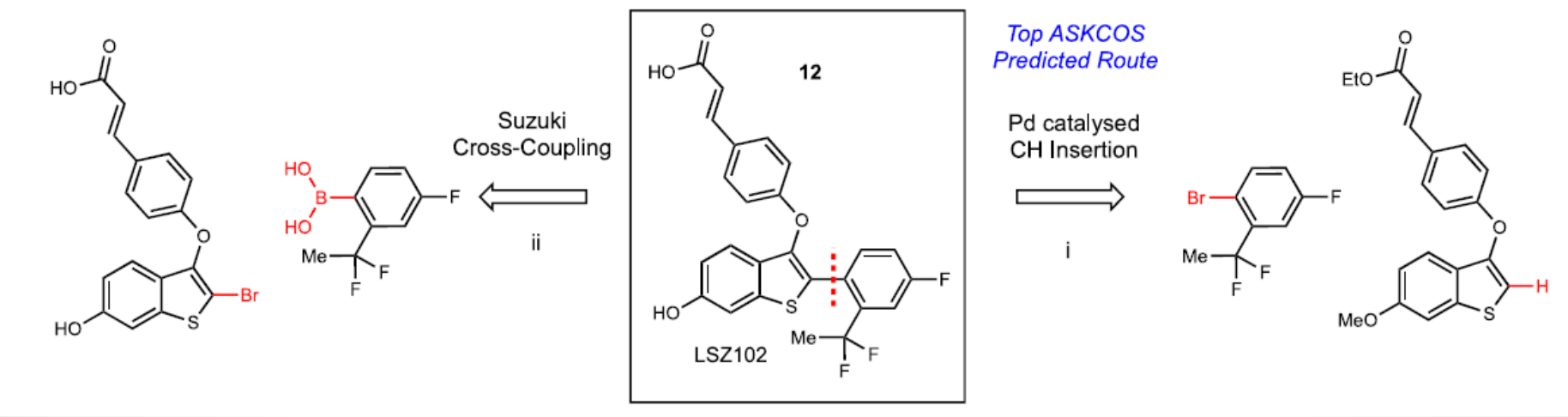
# AI in Synthesis of drugs: 2. Reaction condition planning

Challenges in condition predictions:-

(1) Amounts, volumes, or concentrations

(2) Reaction times or kinetics

(3) Order of the addition of reagents and catalysts




**Optimized condition**: Pd-catalyzed C−H activation, identified the requirement for both high temperature and polar aprotic solvents (DMF/DMA) in the top 3 proposed conditions. This was applied to a diverse range of substrate starting materials and observed yields in the range of 39–97%.

# AI in Synthesis of drugs: 3. Forward Reaction Prediction

- To ensure algorithmic synthesis design are robust and actionable by anticipating, at least qualitatively, reaction products, forward reaction predictive analysis generally conducted.
- The feasibility of a reaction needs to be determined by searching for similar transformations, reading the literature, and determining if the synthetic method will generalize to the substrates of interest.
- Data-driven techniques can learn to perform the same generalization when trained on a broad set of reactions: ML
- Graph convolutional neural networks predict atom and bond changes from starting materials to products.
- Sequence-to-sequence models which predict product SMILES (Simplified Molecular Input Line Entry System).
- Other methods: make-on-demand virtual libraries based on expert-defined reaction templates.

# Challenges in Drug Discovery using AI

- Companies who use AI technology for drug discovery has to go through vigorous process to copyright their work so as to secure patent rights.
- Security is also a major concern, as AI-driven personalized medicine requires person's genetic code (legal issues).
- Faster computation will be required for handling big data and it is said that in future the current supercomputers will be replaced by quantum computers.
- Still no success story where a compound generated through AI made it to the market for public use.
- Insilico medicine, a biotechnology company, proposed a novel target involved in idiopathic pulmonary fibrosis and made its novel inhibitor from scratch, through their AI-based tools.
- The identified small molecule inhibitor has showed good efficacy and applied for investigational new drug (IND) in Dec-2020.
- Expected clinical trials will be late 2022 by that company.

# References


[**1**] Paul et al., *Drug Discov. Today*, **2021**, *26*, 80-93.

[**2**] Jordan A.M., *ACS Med. Chem. Lett*., **2018**, *9*, 1150-1152.

[**3**] Lake F., *Future Drug. Discov.*, **2019**, *1*, FDD19.

[**4**] Mak et al., *Drug Discov. Today*, **2019**, *24*, 773-780.

[**5**] Jose et al., *Expert Opin. Drug Discov*., **2021**, [DOI: 10.1080/17460441.2021.1909567].

[**6**] C. Balakumar et al., *J. Biomol Struc. Dyn*., **2018**, *36*, 3687-3704.

[**7**] Struble et al., *J. Med. Chem. Lett*., **2020**, *63*, 8667-8682.

**[8]** Freedman D.H., *Nature*, **2019**, *576*, S50-S53.

**[9]** Hessler et al., *Molecules*, **2018**, *23*, 2520.