Why drug development takes so long and how to help it?

Zsuzsanna Gáborik, PhD

Head of Research & Development

SOLVO Biotechnology, a Charles River Company

Outline of the presentation

- What could be a drug?
- Introduction to drug discovery
- Computational methods during drug discovery
- CADD (Computer aided drug discovery)
- PBPK models (physiology-based pharmacokinetics)
- Case study of drug repurposing approach for COVID-19 treatment

What is the definition of drug?

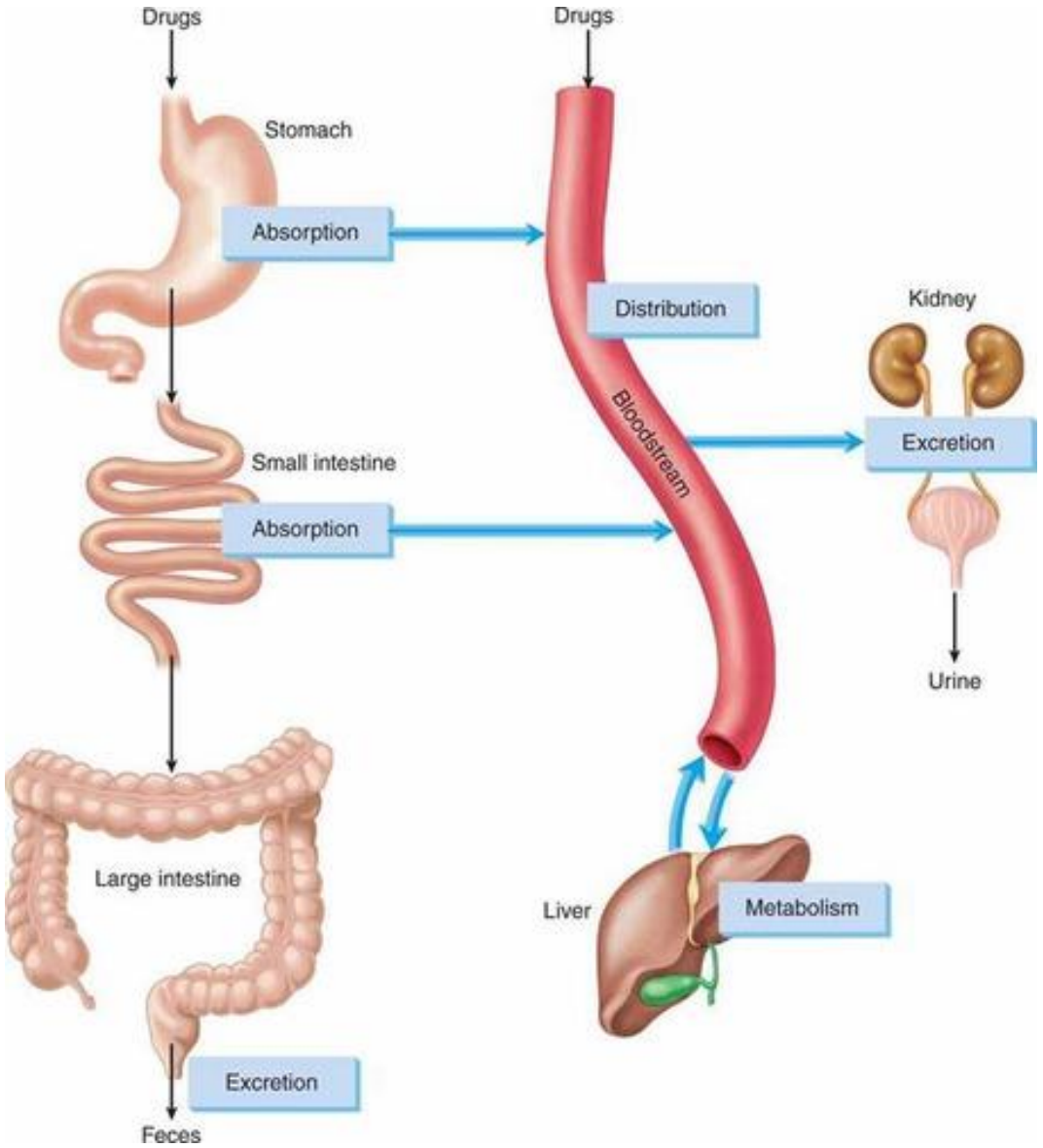
Under the US law, a **drug is any substance** (other than a food or device), which is used in the **diagnosis, cure, relief**, **treatment or prevention of disease**, or intended to **affect the structure or function of the body**. *This comprehensive definition is important for legal purposes.*

- **Simply a drug** can be defined as **any chemical** that **affects the body** and its processes.

What features an “ideal” drug should have?

I. must be **safe** and **effective**
II. should be well absorbed orally and/or **bioavailable**
III. metabolically stable and with a **long half-life**
IV. nontoxic with **minimal** or no **side effects**
V. should have **selective distribution to target tissues**

ADMET properties of drugs:
Absorption
Distribution
Metabolism
Excretion
Toxicity

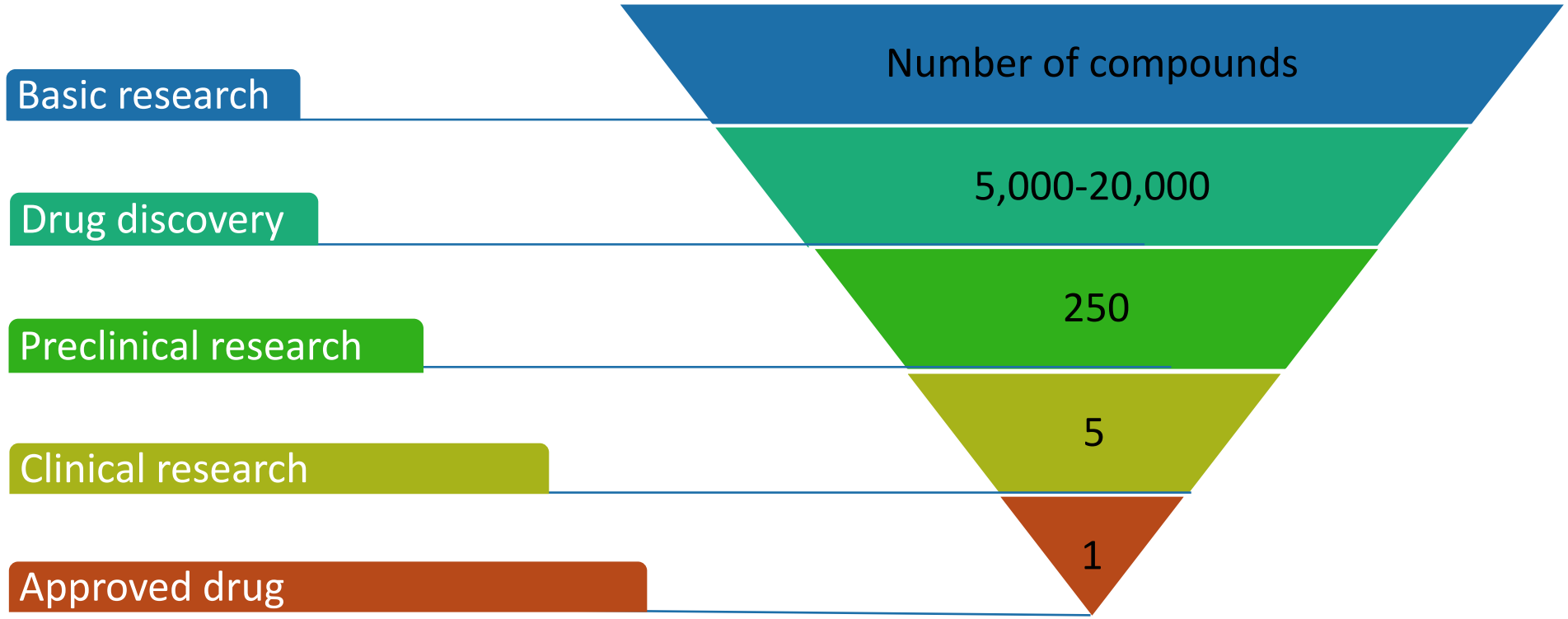


https://quizlet.com/136552549/pharmacology-ch-4-what-happens-after-a-drug-has-been-administered-flash-cards/

Drug discovery and development process

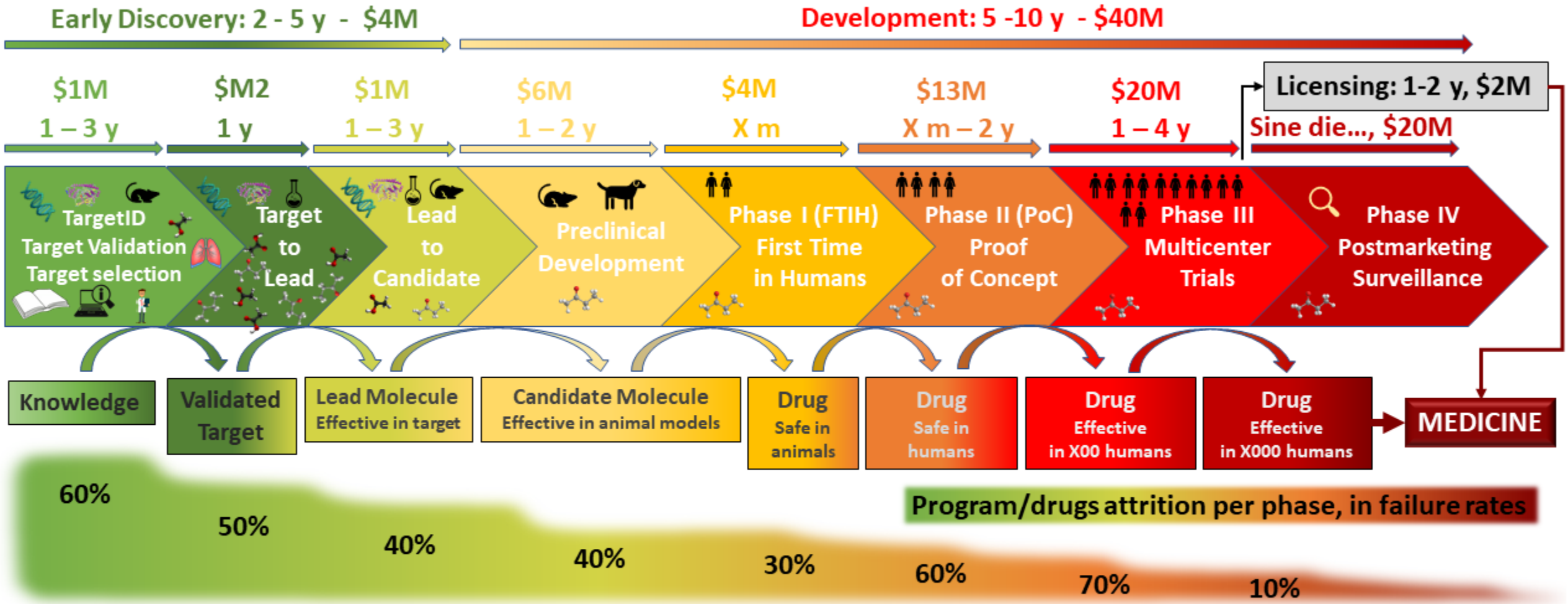
- The **development** of any potential drug **begins with years of scientific study** to determine the **biochemistry behind a disease**, for which pharmaceutical intervention is possible. The result is the determination of **specific receptors (targets)** that must be modulated to **alter their activity** by some means. After **target identification**, the goal then is to **find compounds** that interact with the receptor by mass screening (lead).
- From this point onward, a **cycle of iterative refinement and testing** continues until a drug is developed that undergoes clinical trials. The techniques used to refine drugs are combinatorial and structure-based design.

Number of molecules tested through the developmental process



https://www.technologynetworks.com/drug-discovery/articles/exploring-the-drug-development-process-331894

The Drug Discovery Process



- Each stage output is the input of the next one.
- The system works like a pipeline, each phase feeding the following one with backups in prevention of program failures.
- Individual pipelines represent therapeutic concepts. Failed stages are not replaced by backups when there are no more appropriate molecules available, on target liabilities appear, compound does not prove therapeutic efficacy, or strategic decisions are applied.
- Costs and timelines represent the values for unique iterations of the respective phases.

https://doctortarget.com/machine-learning-applied-drug-discovery/

Why is drug discovery takes so long and so expensive?

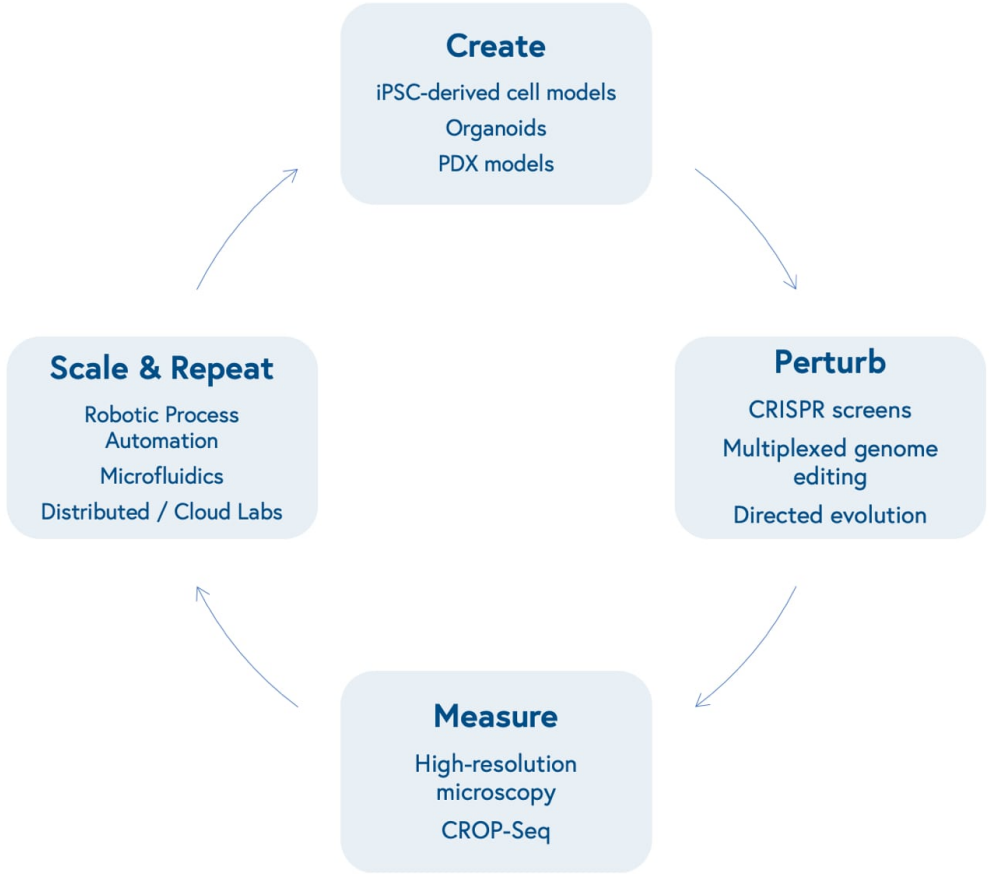
- It is **inherent to the nature of biology**. We **still don't understand biology well enough** to know what proteins to target or how drugging a target of interest will affect an organ system or the entire human body, simply **because of the complexity of the human being**. Drug targets are parts of complex cellular networks leading to unpredictable changes. In addition, biological systems also show a **high degree of redundancy**, which could blunt the effects of even the most specifically-targeted therapeutics.
- **New drugs have to be significantly better** than currently available treatments, so building a new drug for a disease with prior treatments becomes increasingly more difficult.
- **Safety** of new drugs is just as important as efficacy. Whit increasing experience and knowledge of adverse effects and safety concerns, regulatory requirements for safety assessment of new drugs is evolving continuously.
- In the last two decades the industry shifted from **iterative medicinal chemistry coupled with phenotypic assays to the serial filtering of a static compound library** against a given target. However, the ability of high-throughput screening of large chemical libraries to a specific target to tell researchers that a drug candidate will be safe and effective in human clinical trials is very limited.
- **Capturing, representing, and perturbing that complex biological network *in silico* is the holy grail of machine learning for drug discovery.**

https://www.bvp.com/atlas/roadmap-unlocking-machine-learning-for-drug-discovery

The pharmaceutical industry benefits from scientific revolutions in both biomedicine and computer science.

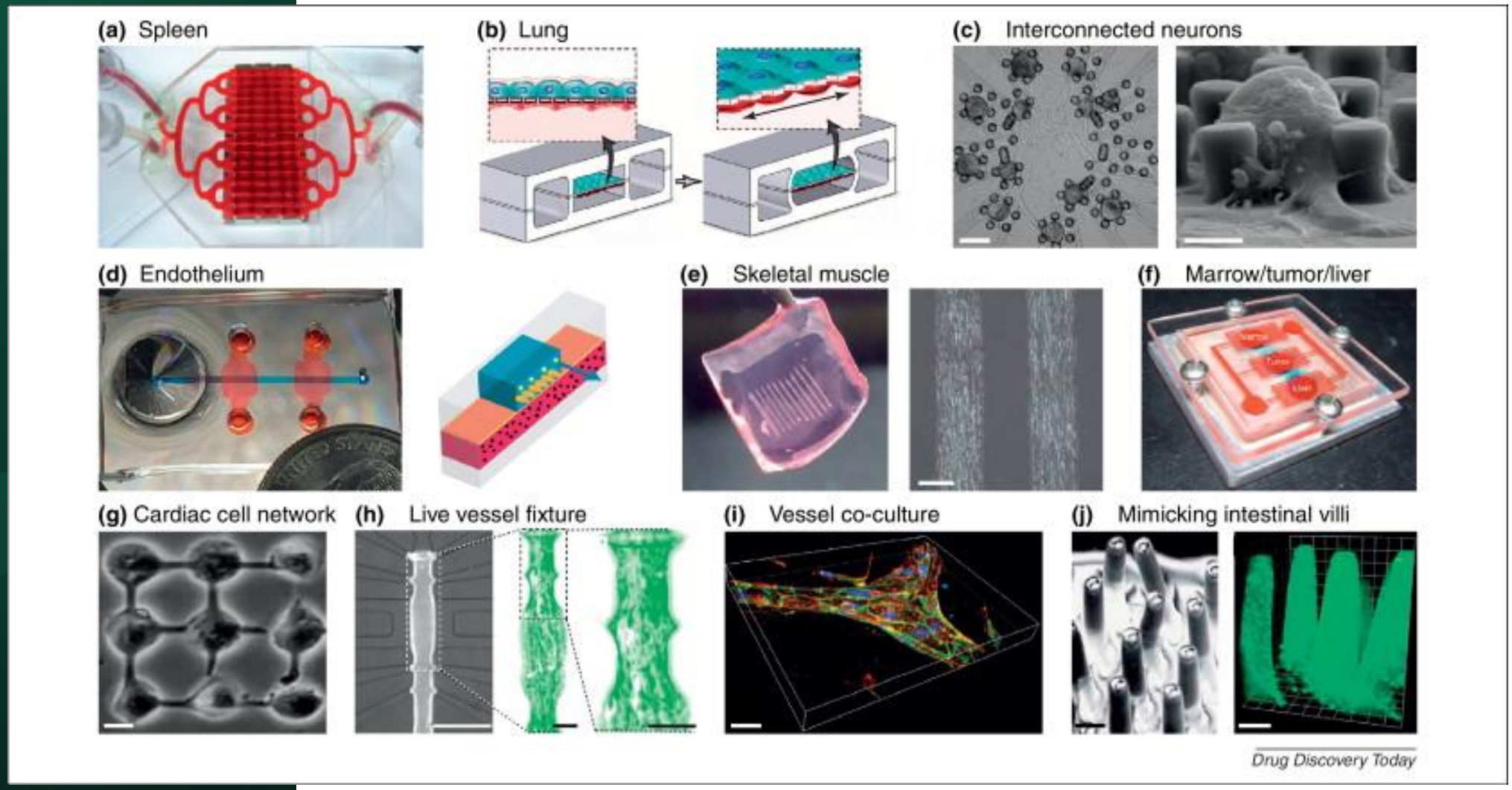
New tools available improved our ability to create, manipulate, and measure biological systems at scale:

- models reflecting unique biological dynamics and variation:
  - **induced pluripotent stem cell** (iPSCs)
  - **organoids** that reflect three-dimensional tissue dynamics
  - **patient-derived xenografts** (PDX) to mimic human disease in mouse models.
- ability to **perturb** these models in a genome or protein-targeted fashion.
  - **CRISPR/Cas9-based gene editing**
- **Robotic automation and microfluidics**
- High resolution microscopy – high content screening (HCS)
- New generation sequencing



https://www.bvp.com/atlas/roadmap-unlocking-machine-learning-for-drug-discovery

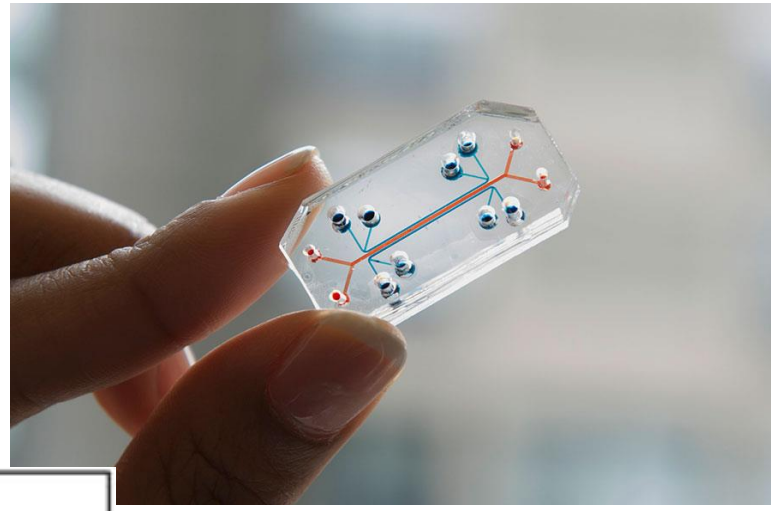
Tissue on a chip



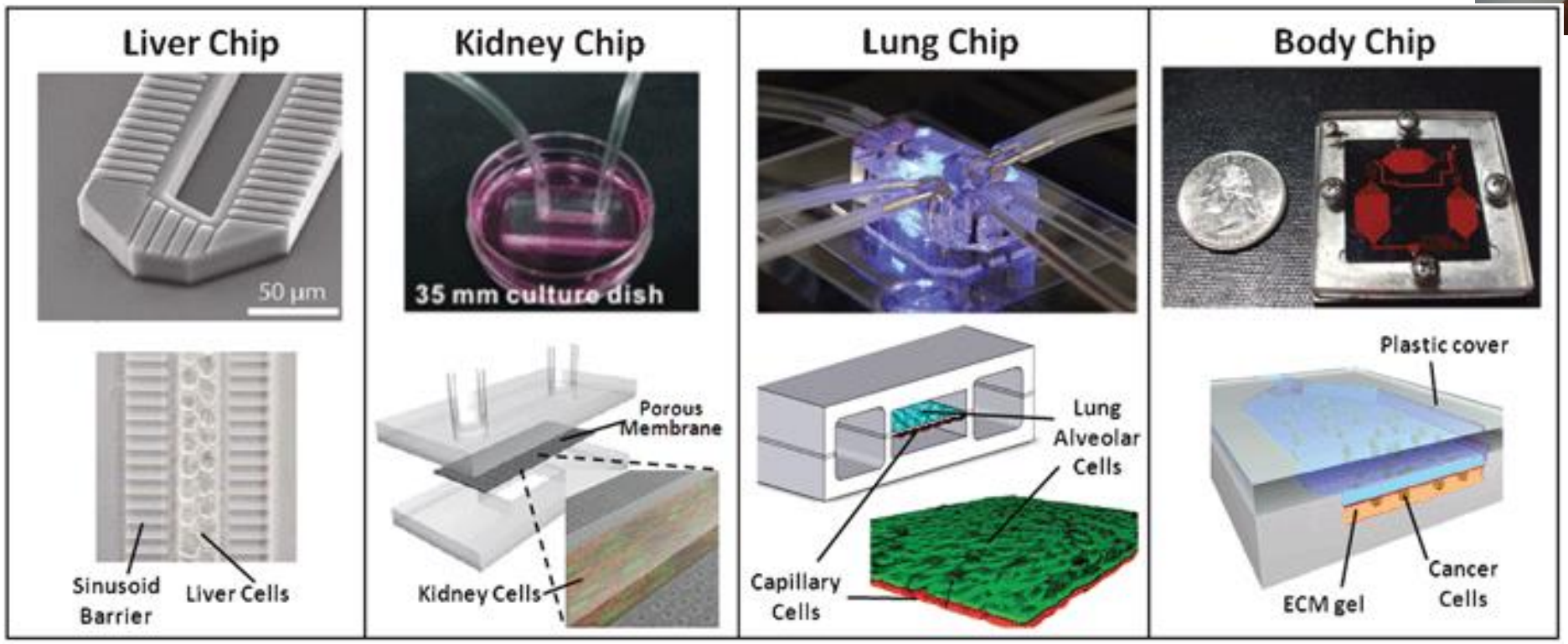
- *Drug Discovery Today, February 2012, Pages 173-181*

Tissue on a chip

This lung-on-a-chip serves as an accurate model of human lungs to test for drug safety and efficacy. (Wyss Institute for Biologically Inspired Engineering, Harvard University Photo)



https://ncats.nih.gov/tissuechip/about



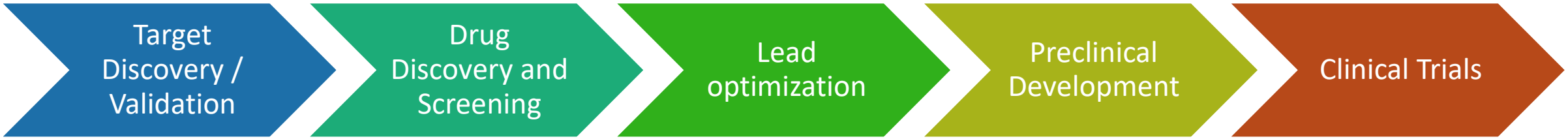
https://pubs.rsc.org/en/content/articlelanding/2012/LC/c2lc40089h

Why is it advantageous to use new computational tools during the process?

With significant improvement in biological models, computational tools and machine learning techniques can be efficiently used in the drug development process:

- Shorten the timeline of drug development– Timeline of drug exclusivity on the market is limited by the patent expiry date, which is 15 years after filing.
- Increase the throughput of screening steps
- 3R – Replace, reduce, refine – minimalize animal testing
- Reduce the risk of failure during clinical studies – save money

Computational and machine learning opportunities in drug discovery



- GWAS /Sequencing
- Crystallography
- *In vitro* and *in vivo* functional assay

- High-throughput screening
- Structure-based drug design
- Molecular dynamics
- Initial *in vitro* safety assessment

- Medicinal chemistry and / or ratinal drug design to optimize drug-like properties
- Absorption, distribution, metabolism, excretion, and toxicity studies (ADMET)

- *In vitro* safety and efficacy in alternate cell lines
- *In vivo* safety and efficacy in mice / non-human primates

- Patient recruitment, engagement, and oversight
- Regulatory documentation and *feedback*

https://www.bvp.com/atlas/roadmap-unlocking-machine-learning-for-drug-discovery

Phases of drug discovery and development

Early Discovery

- **Target identification, validation and selection**:
  - A set benchmark compounds is proved active against a documented target and a disease-relevant experimental model.
- **Hit to Lead**:
  - Active molecules on a selected targets are evaluated for **potency** and **selectivity** and undergo further optimization by chemical improvement.
- **Lead to Candidate**:
  - Lead compounds are tested for **efficacy, pharmacokinetics, pharmacodynamics** and **safety**.

Phases of drug discovery and development

Early Discovery

1. Target identification, validation and selection:

- A **biological target** *can be anything within a living organism to which an endogenous ligand or a drug is directed and/or binds, resulting in a change in its behavior or function*.
  - Proteins
    - G protein-coupled receptors (~700 approved drugs, 35% of total drugs):
      - Examples: drugs to treat allergy, blood pressure, HIV infection, Parkinson
    - enzymes (especially protein kinases, proteases, esterases, and phosphatases),
      - Examples: Cancer treatments, neurological diseases, Paxlovid
    - ion channels
      - Examples: epilepsy, pain
    - nuclear hormone receptors
    - structural proteins such as tubulin
    - membrane transport proteins
  - nucleic acids
    - Examples: siRNA therapy, remdesivir, molnupiravir

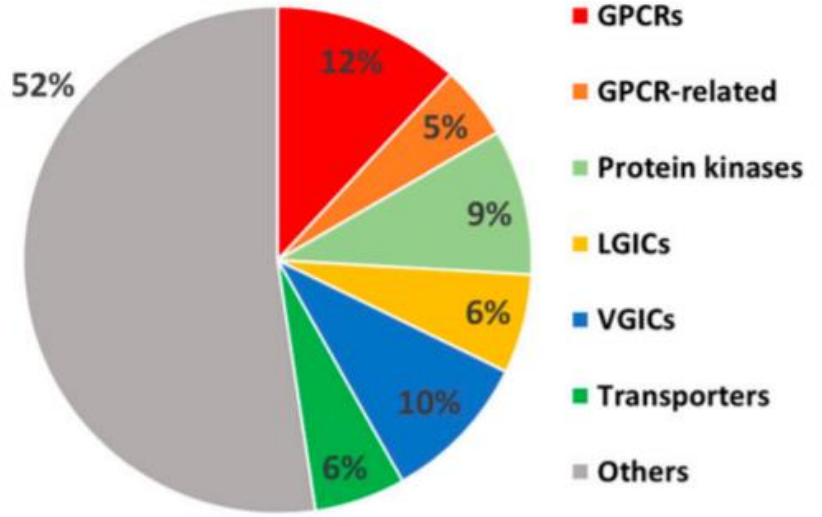


Fig. 2. The estimated proportion of genes from different gene families that are targets for approved drugs. GPCRs comprise the single largest such group. VGICs: voltage-gated ion channels; LGICs: ligand-gated ion channels.

Mol Pharmacol 93:251–258, April 2018 K. Sriram and P. A. Insel G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs?

Role of computation tools in target identification

1. Target identification:

- *bioinformatics can help reveal the key genes from a massive amount of genomic data thus provide possible target proteins for drug screening and design*
- **GWAS: Genome-wide association study:**

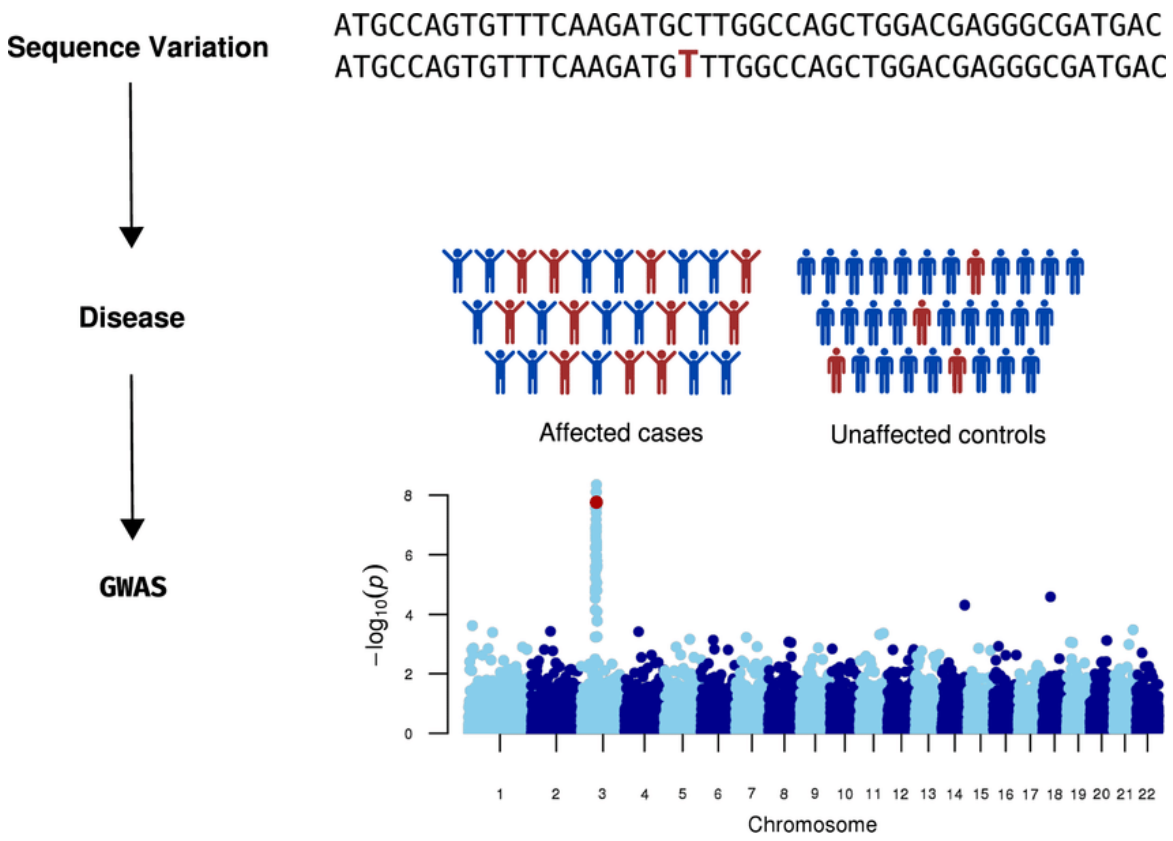
  frequencies of genetic variants are compared between individuals with a disease (cases) and unaffected individuals (controls).

  T variant (red) is enriched in cases compared to controls.

  tens of millions of sequence variants are tested in this manner and are able to detect small differences in frequency in well-powered studies.

  GWAS results are typically displayed as a “Manhattan plot” that shows significance levels on the vertical axis against the chromosomal positions of sequence variants.

  an association signal typically results in a ‘peak’, arising from regional correlation of genetic variation caused by linkage disequilibrium among SNP (single nucleotide polymorphisms).

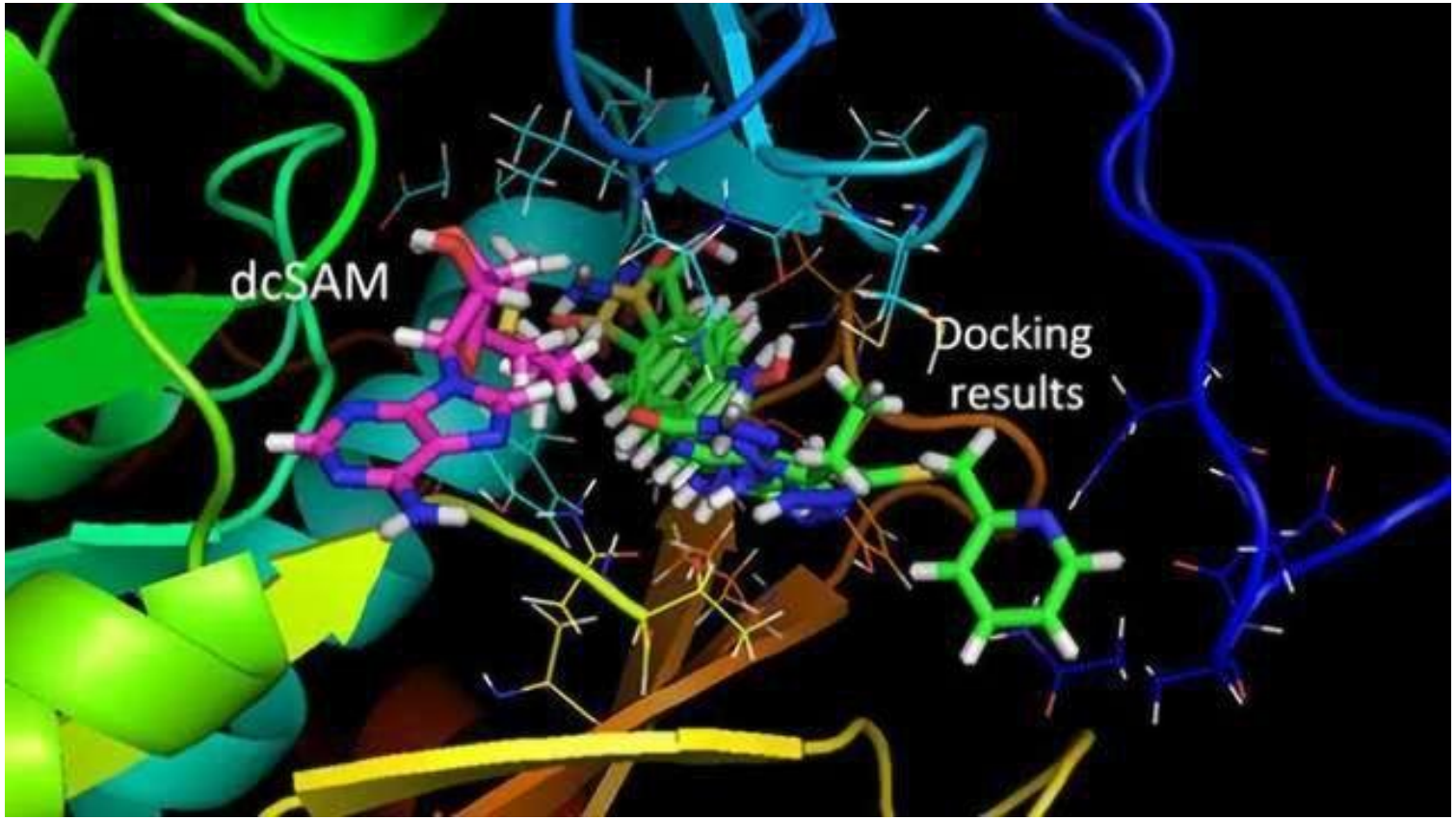


Urolithiasis 2019 Feb;47(1):11-21. Genetics of common complex kidney stone disease: insights from genome-wide association studies. Runolfur Palsson[1,2], Olafur S Indridason[3], Vidar O Edvardsson[4,5], Asmundur Oddsson[6]

CADD: Computer aided drug design

2. Target structure:

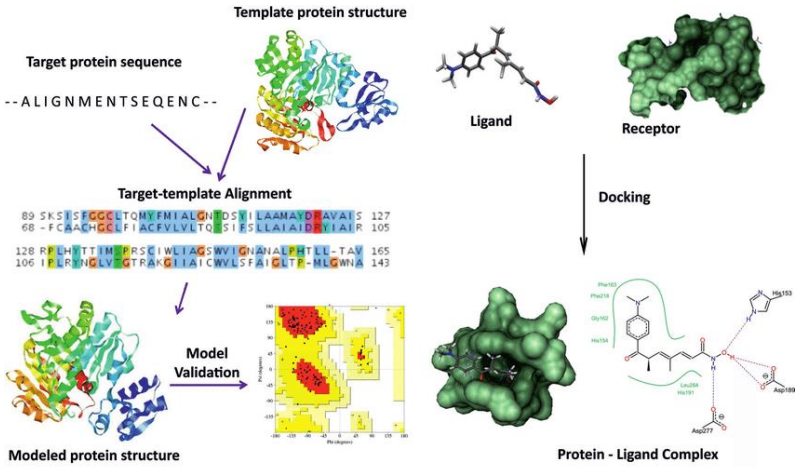
- Protein structure prediction methods can provide protein structures with reasonable precision
- Biomolecular simulations with multiscale models allow for investigations of both structural and thermodynamic features of target proteins on different levels, which are useful for identifying drug binding sites and elucidating drug action mechanisms.

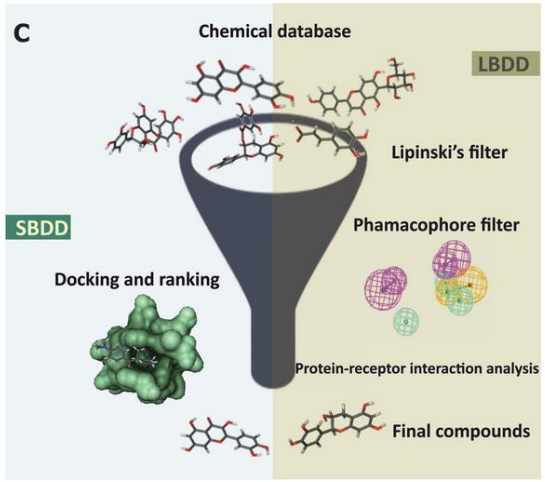


https://www.technologynetworks.com/drug-discovery/news/potential-for-anti-chagas-therapies-using-structure-based-drug-design-292690

CADD Computer-aided drug discovery

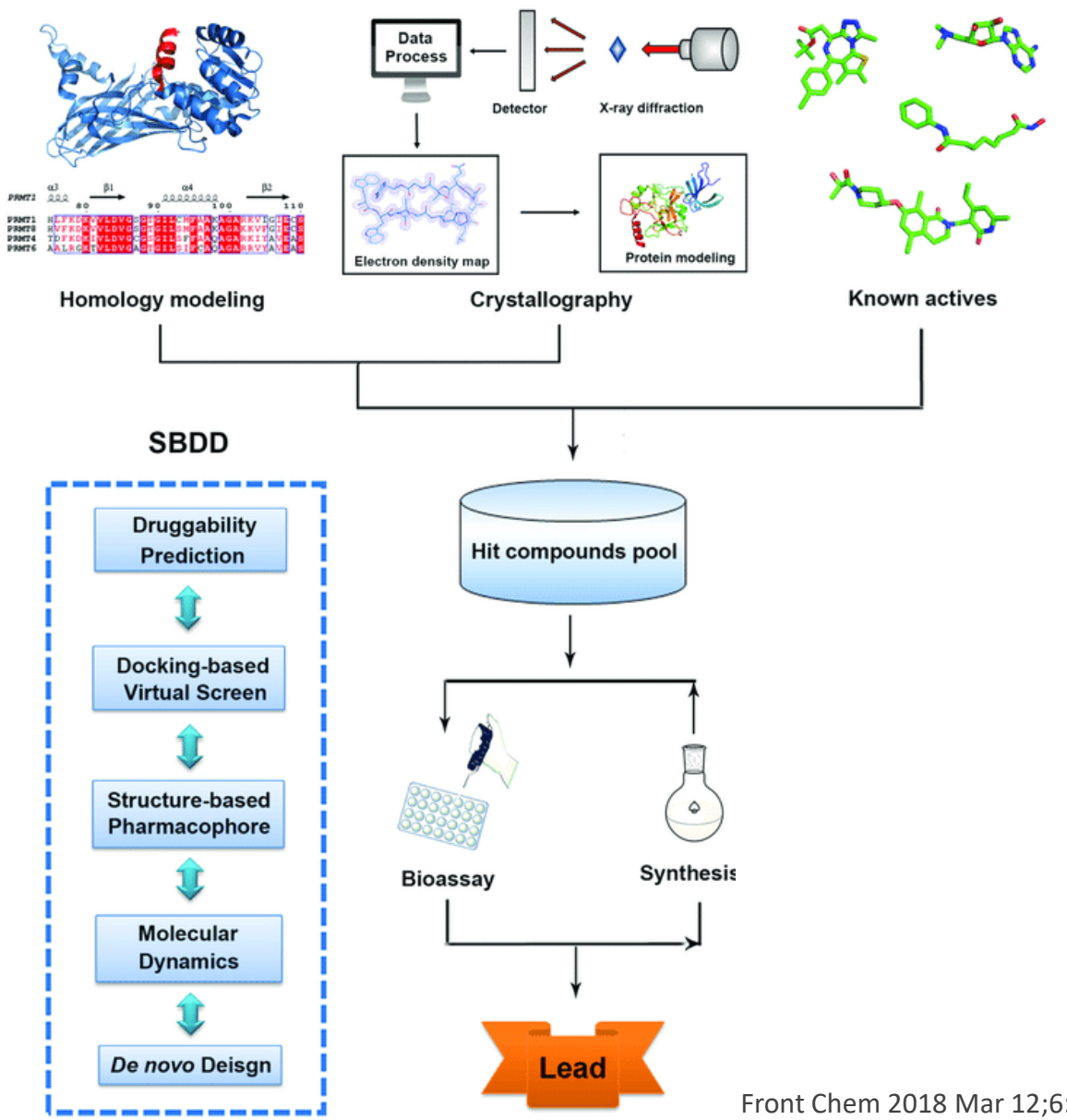
- **Structure-based drug design (SBDD):**
- 3D structure and biological function of target protein are available and large numbers of ligand are screened against it.
- Candidate drugs that are predicted to bind to the target with high affinity and selectivity are designed.
- **Ligand-based drug design (LBDD):**
- 3D structure of the target protein is not known.
- The only information available are ligands and their biological activities, computer-aided LBDD is an effective method to design compounds with improved biological activity.
- Chemical entities of single ligand is used to screen hit compounds and/or screened against various protein targets of interest.





Plant Pathol J. 2017;33(6):529-542

Structure-based drug design (SBDD)



Front Chem 2018 Mar 12;6:57.

Major SBDD methods:

- **Homology modelling:** homology model based on primary sequence similarity of the target to homologous proteins, of which 3D structure is empirically known, can be created.
- **Molecular docking.** If the model of active sites is available, protein-ligand interactions can be explored through molecular docking, predicting energetically stable ligand binding. Degree of interaction stability is determined by using the mathematical methods of scoring functions. There are four general classes of scoring functions:
  - **Force field** –based on the strength of intermolecular van der Waals and electrostatic interactions between the two molecules in the complex using a force field.
  - **Empirical** –counting the number of various types of interactions between the two binding partners.
  - **Knowledge-based** – statistical observations of intermolecular close contacts in large 3D databases
  - **Machine-learning** –not assuming a predetermined function for the relationship between binding affinity and the structural features of protein-ligand complex.
- **Structure-based virtual screening** (SBVS). The search for new chemical compounds as lead molecules is a critical step during the process of drug discovery. Once the target is selected, the small molecules database are selected for virtual screening and their binding interactions with the selected drug target are explored.
- **De novo ligand design (DnLD)**. Using 3D structure information of the target, ligand can be designed de novo.

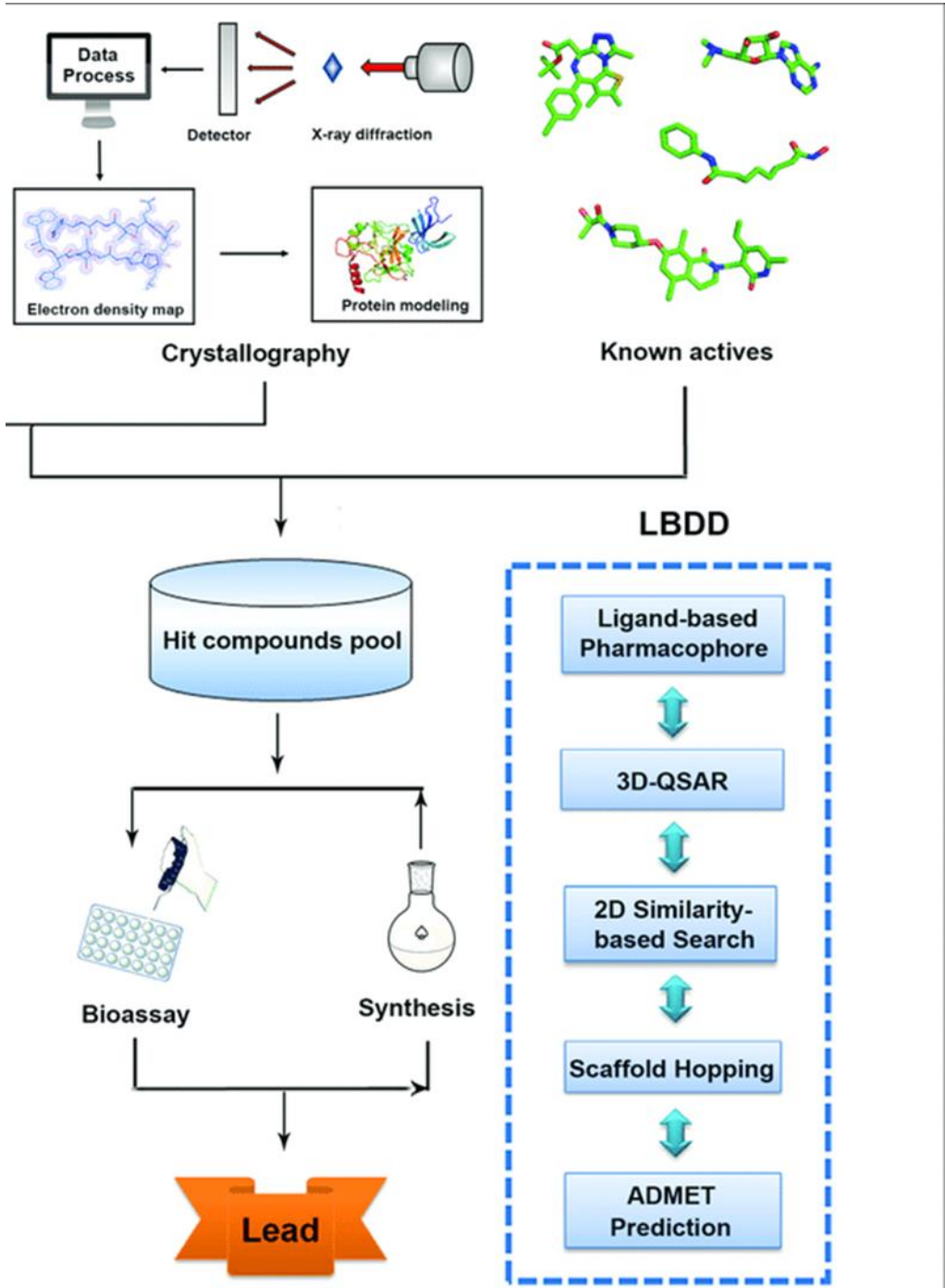
Ligand base drug design (LBDD)

When the 3D structure of the target protein is not known, and the only information available are ligands and their biological activities, computer-aided LBDD is an effective method to design compounds with improved biological activity. Availability of 3D structure of target protein present in cell surface or membrane are very limited due to difficulties in protein crystallization. Since more than 50 % of current FDA-approved drugs targeting **membrane proteins** LBDD methodologies have a significant impact on drug development.

Major LBDD methods:

- **Pharmacophore models** identifying the minimum, steric and electronic features that are required for interaction of target protein with ligand(s).
- **Ligand-based virtual screening (LBVS).** Large diverse set of candidate compounds can be scanned to predict whether candidate ligands are likely to bind to the target through comparison to the pharmacophore model. LBVS works as a chemical database filters.
- **Quantitative structure-activity relationships (QSAR)** yielding quantitative estimates of activities based on physiochemical properties. Quantitative structure–activity relationship (QSAR) models are regression or classification models used to predict activities of new chemical compounds based on their physico-chemical properties.
- **Similarity searching**, which explores compounds with similar properties as well as various combinations of the above.

➢ Overall improvements have been achieved by sophisticated data mining techniques and by more accurate mathematical descriptions of molecules through molecular mechanics (MM) and quantum mechanics (QM) methods.



Front Chem 2018 Mar 12;6:57.

CADD protocol example

(A) Homology models

(B) Model validation using Ramachandran plot for each model.

(C) Ligand selection

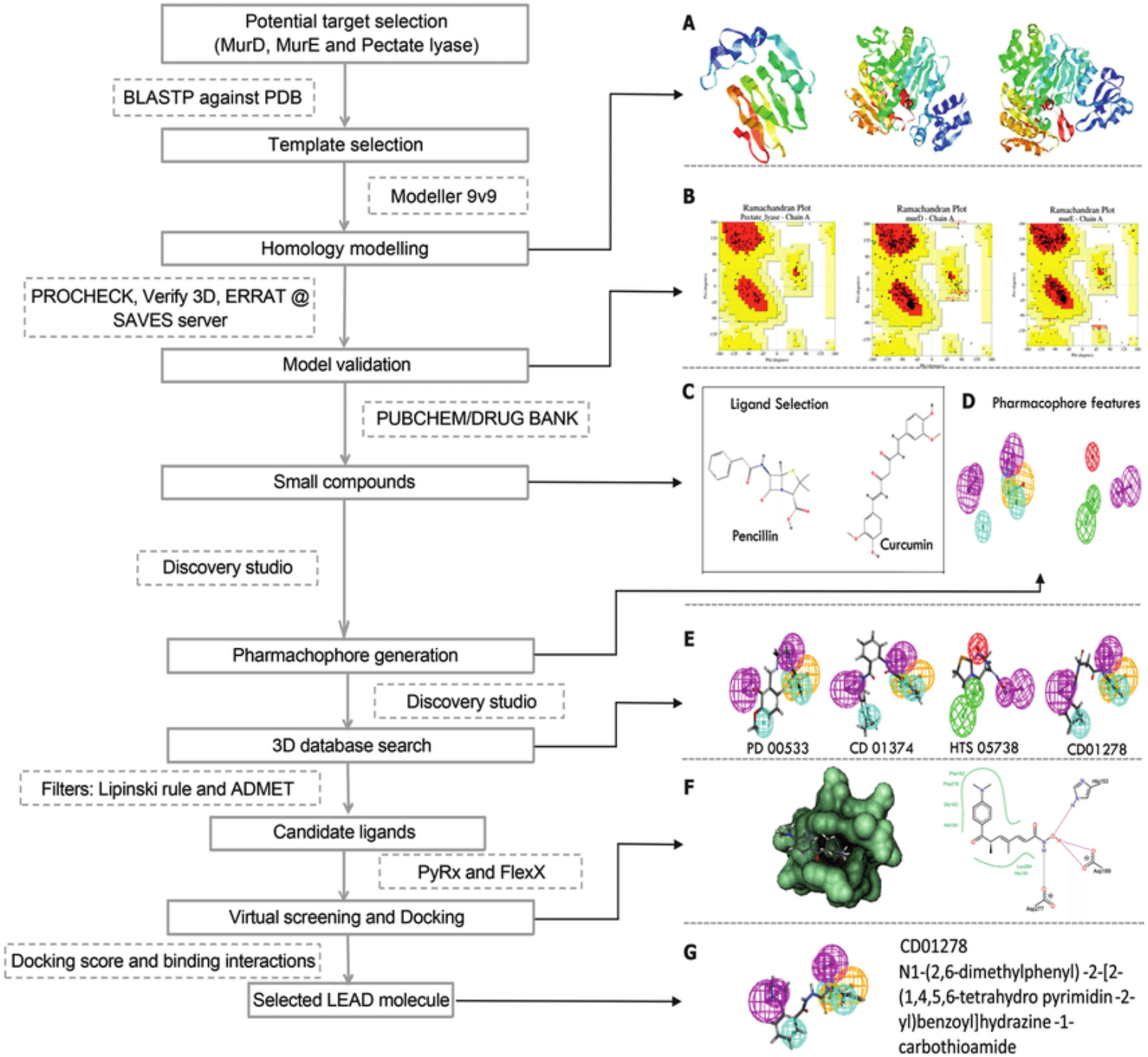
(D) Pharmacophore generation. Hydrogen bond acceptor (green), hydrogen bond donor (magenta), hydrophobic (cyan), and ring aromatic (orange), ionizable positive charge (red) are shown here.

(E) 3D database screening.

(F) Virtual screening and docking interactions. Docking interactions of Maybridge database compounds with the models are illustrated.

(G) Identification of lead molecule: the compound showing best docking interaction with the modelled protein.

Softwares, databases and servers used in the case study are given in dotted boxes, while the process is shown in solid boxes.



Plant Pathol J. 2017;33(6):529-542

Role of computation tools in lead generation

Early Discovery

Hit to Lead (Lead generation)

Active molecule hits from the high throughput screen (HTS) with selected target are evaluated for **potency** and **selectivity**, then undergo **limited optimization** to identify promising lead compounds.

- Confirmation of drug mechanism of action *in vitro*
- Physico-chemical parameters, like water solubility, chemical stability
- Pharmacokinetic and pharmacodynamic parameter, drug metabolism
- Initial safety test *in vitro*

- **QSAR**
- **Cheminformatics solutions:** (ChemAxon)
  - **Chemicalize**
  - **Calculators and Predictors:** *High-quality physico-chemical calculations and predictions for drug discovery*

Experimental Properties

PROPERTY	VALUE	SOURCE
water solubility	Sparingly soluble in water	FDA label
logP	0.13	FDA label

Predicted Properties

PROPERTY	VALUE	SOURCE
Water Solubility	0.0886 mg/mL	ALOGPS
logP	1.47	ALOGPS
logP	1.92	ChemAxon
logS	-3.7	ALOGPS
pKa (Strongest Acidic)	4	ChemAxon
pKa (Strongest Basic)	-2.8	ChemAxon
Physiological Charge	-1	ChemAxon
Hydrogen Acceptor Count	8	ChemAxon
Hydrogen Donor Count	3	ChemAxon
Polar Surface Area	140.92 $Å^2$	ChemAxon
Rotatable Bond Count	9	ChemAxon
Refractivity	121.44 $m^3 \cdot mol^{-1}$	ChemAxon
Polarizability	48.55 $Å^3$	ChemAxon
Number of Rings	2	ChemAxon
Bioavailability	1	ChemAxon
Rule of Five	Yes	ChemAxon
Ghose Filter	No	ChemAxon

98#categories-header

Role of computation tools in lead optimization

Early Discovery

- **Lead to Candidate (Lead optimization)**
  - **Chemical modification of the hit structure with the purpose of**
    - Improved efficacy
    - Reduced off-target activities
    - Optimized physiochemical/metabolic properties
  - Prediction of Physicochemical Properties
  - Prediction of ADME Properties
  - Prediction of Toxicity
  - Physicochemical and ADMET Property-based Design

- **SAR**
- **Cheminformatics solutions:**
  - **Calculators and Predictors:** *High-quality physico-chemical calculations and predictions for drug discovery*

Role of computation tools in preclinical development

Preclinical development

- **Preclinical development**
  - The main goals are to
  - Determine a **starting, safe dose** for first-in-human study
  - Assess **potential toxicity**

  - Testing
    - Pharmacodynamic (what the drug does to the body) PD
    - Pharmacokinetic (what the body does to the drug) PK
    - ADME (Absortion, Distribution, Metabolsim, Excretion)
    - Toxicology
    - **In vitro**
    - **Animal tests**

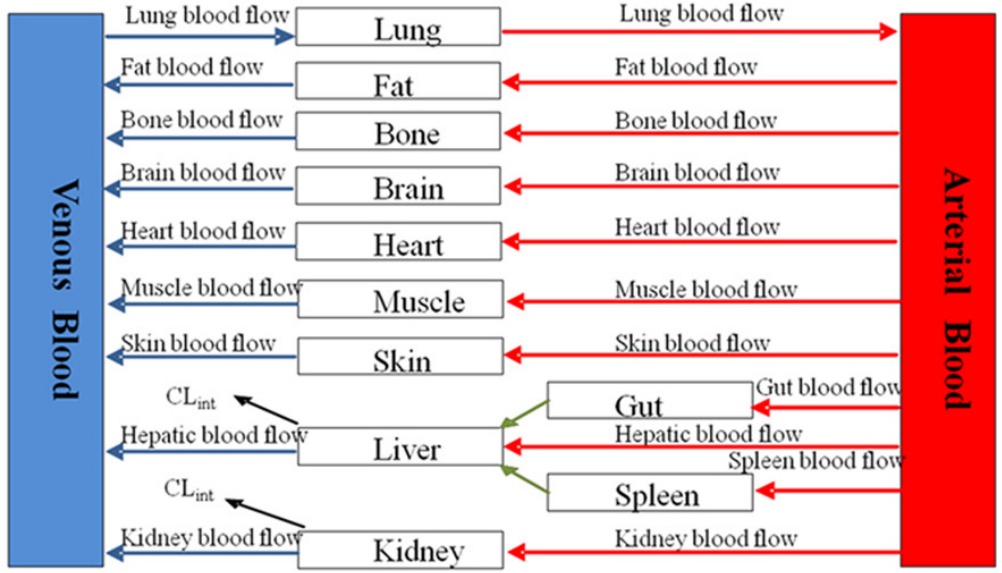
➢**PBPK models and simulators**

Physiological based pharmacokinetic (PBPK) models and simulators

Computational modeling approach that incorporates blood flow and tissue composition of organs to define the pharmacokinetics (PK) of drugs.

- First step is building a reliable model using tons of input data from in vitro, animal and human PK studies and simulations of connection physiology of organs
- Links in vitro data to in vivo absorption, distribution, metabolism, and excretion (ADME) and pharmacokinetic / pharmacodynamic (PK/PD) outcomes
- Enables extrapolation from in vitro to clinical trial data
- Determination of first-in-human dosing
- Provides information for designing clinical trials, to reduce trial size and complexity
- Predicts dosing recommendations for different populations of patients: pediatrics, geriatrics, ethnicities, organ impairment.

Introduction of PBPK M&S



Case Studies

- ✓ Lead optimization or candidate evaluation
- ✓DDI potential prediction
- ✓Human PK and DDI prediction to avoid clinical DDI trials
- ✓Dose guidance for renal impairment patients
- ✓Bridge healthy adults to special populations

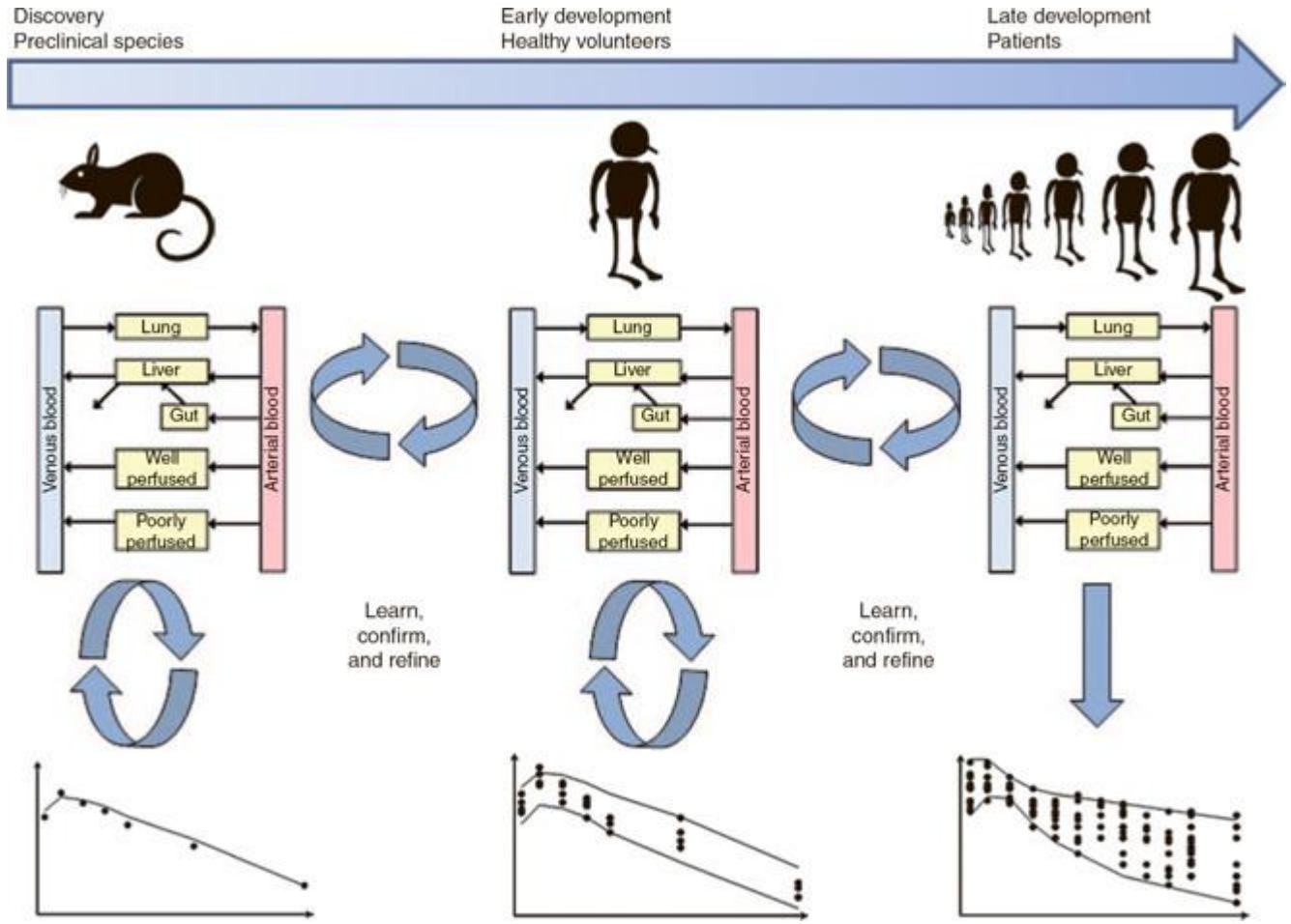
Regulatory acceptance and industrial practices around PBPK M&S

Acta Pharmaceutica Sinica B, Volume 6, Issue 5, September 2016, Pages 430-440

PBPK models and simulators

PBPK modeling strategy:

- iterative “learn, confirm, and refine” approach
- Initially, the PBPK simulation is performed in animals using **animal PBPK models**, animal in vitro data, and compound-specific physicochemical data. The animal simulation is **compared with the in vivo data.**
- If simulation in animals is reasonable then the **healthy volunteer simulation** is performed using a **human PBPK model** built using healthy volunteer physiology, human in vitro data, and compound-specific physicochemical data.
- These simulations can then be extended to various patient populations using relevant physiology.
- If the simulation at any stage is inaccurate, this would indicate a violation of one or more of the model assumptions, in this case further experiments may be performed to understand the mismatch.

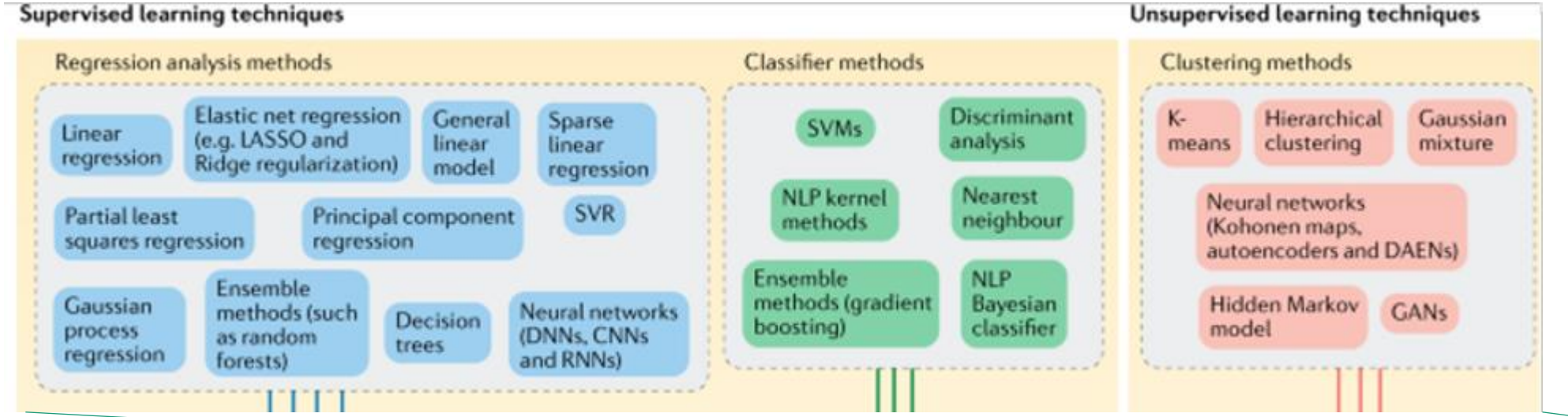


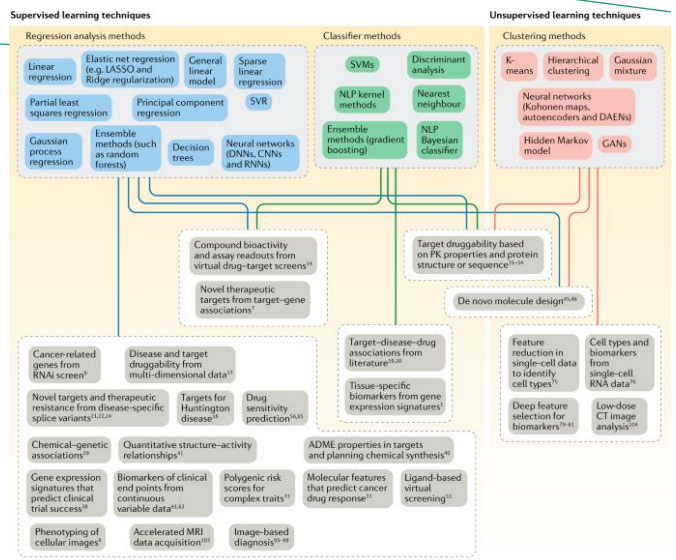
CPT: Pharmacometrics & Systems Pharmacology (2013) 2, e63

Phases of drug discovery and development
Clinical development

- **Phase I**
  - small groups of **healthy volunteers**
  - small ascending **single doses** of the drug
  - in order to obtain information on the **safety, tolerability** and pharmacokinetic properties of the drug in humans. Also, to generate information on **optimal dosage** of the drug candidate.
- **Phase II**
  - **patients** who have the **targeted disease**
  - to evaluate the possible **therapeutic effects**
  - continue Phase I **safety evaluations** in patients with the disease or condition under study.
- **Phase III**
  - **several thousands of patients** in a real-life clinical setting.
  - This phase is essential in determining whether **the drug is safe and effective**.
  - If the drug candidate is successful in Phase III clinical trials, a **new drug application can be submitted to regulatory authorities**
- **Post-marketing surveillance (Phase IV)**
  - to evaluate the **true safety profile** of the drug **in large-scale use**.
  - drug is used by large patient populations, new **rare or long-term side effects** may be detected, and real-world data of the therapeutic value of the drug can be collected

Machine learning tools and their drug discovery applications





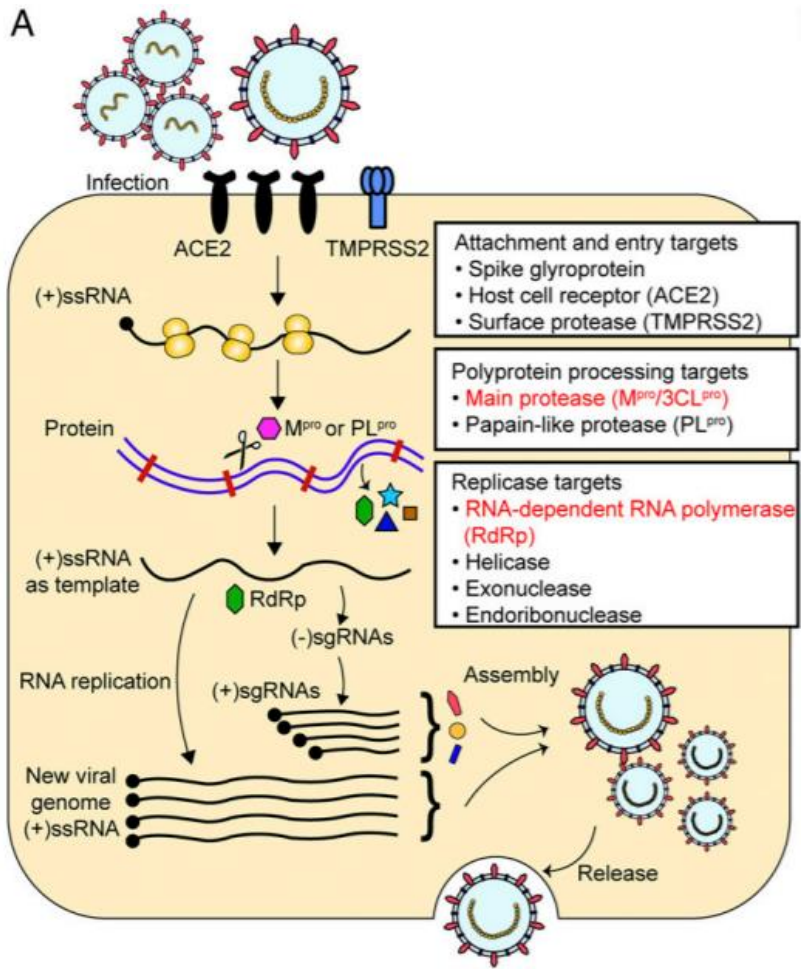
Nature Reviews Drug Discovery **volume 18**, pages463–477 (2019) **Applications of machine learning in drug discovery and development**

Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay

- The COVID-19 pandemic caused by SARS-CoV-2 is an unprecedentedly significant health threat, prompting the need for rapidly developing antiviral drugs for the treatment.
- Drug repurposing is currently one of the most tangible options for rapidly developing drugs for emerging viruses.
- Drug repurposing starts with virtual screening of approved drugs employing various computational methods.
- However, the actual hit rate of virtual screening is very low with many false positives. A strategy for virtual screening with much reduced false positives through incorporating pre-docking filtering based on shape similarity and post-docking filtering based on interaction similarity. This advanced virtual screening approach was applied for 6,218 approved and clinical trial drugs for COVID-19.
- 6,218 compounds were screened against main protease and RNA-dependent RNA polymerase of SARS-CoV-2, resulting in 15 and 23 potential repurposed drugs, respectively.
- Three of these drugs, emodin, omipalisib, and tipifarnib, show anti-SARS-CoV2 activities in human lung cells, Calu-3.
- Activity of omipalisib is 200-fold higher than that of remdesivir in Calu-3.
- Drug combinations of omipalisib/remdesivir, tipifarnib/omipalisib, and tipifarnib/remdesivir, show strong synergistic effects in inhibiting SARS-CoV-2. Such drug combination therapy improves antiviral efficacy in SARS-CoV-2 infection and reduces the risk of each drug's toxicity.

Woo Dae Janga, Sangeun Jeonc, Seungtaek Kimc and Sang Yup Leea

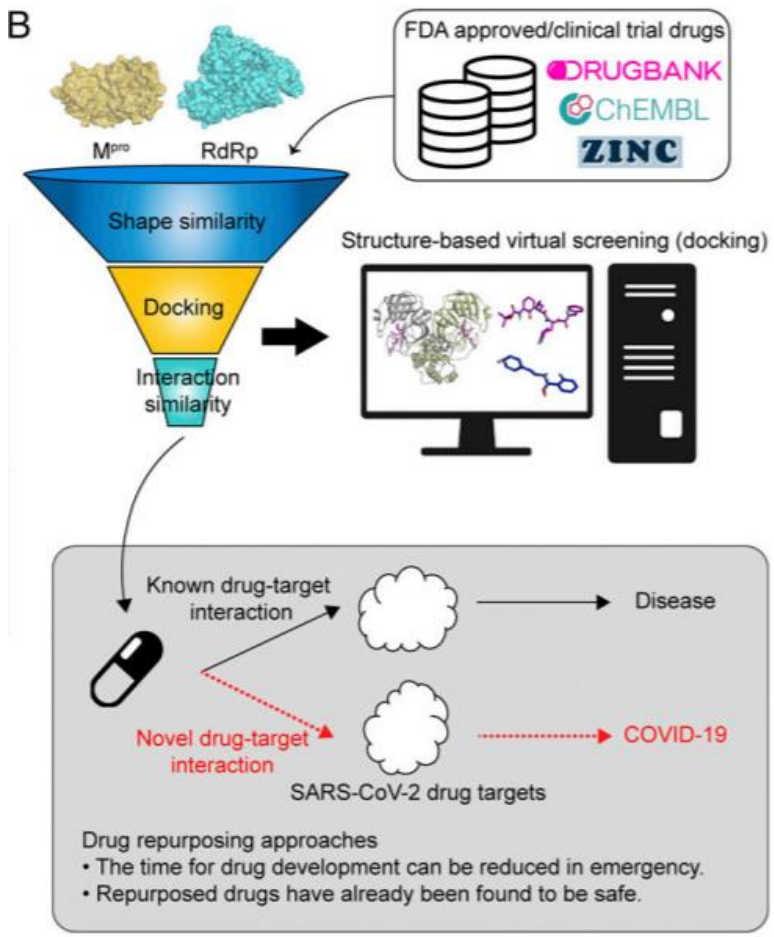
PNAS 2021 Vol. 118 No. 30



Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay

- **Potential drug targets in SARS-CoV-2 replication cycle**.
- Targets for **viral attachment** and entry include the viral spike glycoproteins, host receptors (ACE2), and proteases (TMPRSS2).
- **Polyprotein processing** can be targeted by inhibiting viral proteases such as **main protease Mpro** and papain-like proteases.
- **Viral replicase-related** enzymes are also attractive drug targets for antiviral activity. **RdRp** and helicase are important enzymes involved in the transcription and replication of SARS-CoV-2. Among these, the most important and less variable Mpro and RdRp were selected as drug targets in this study

PNAS 2021 Vol. 118 No. 30

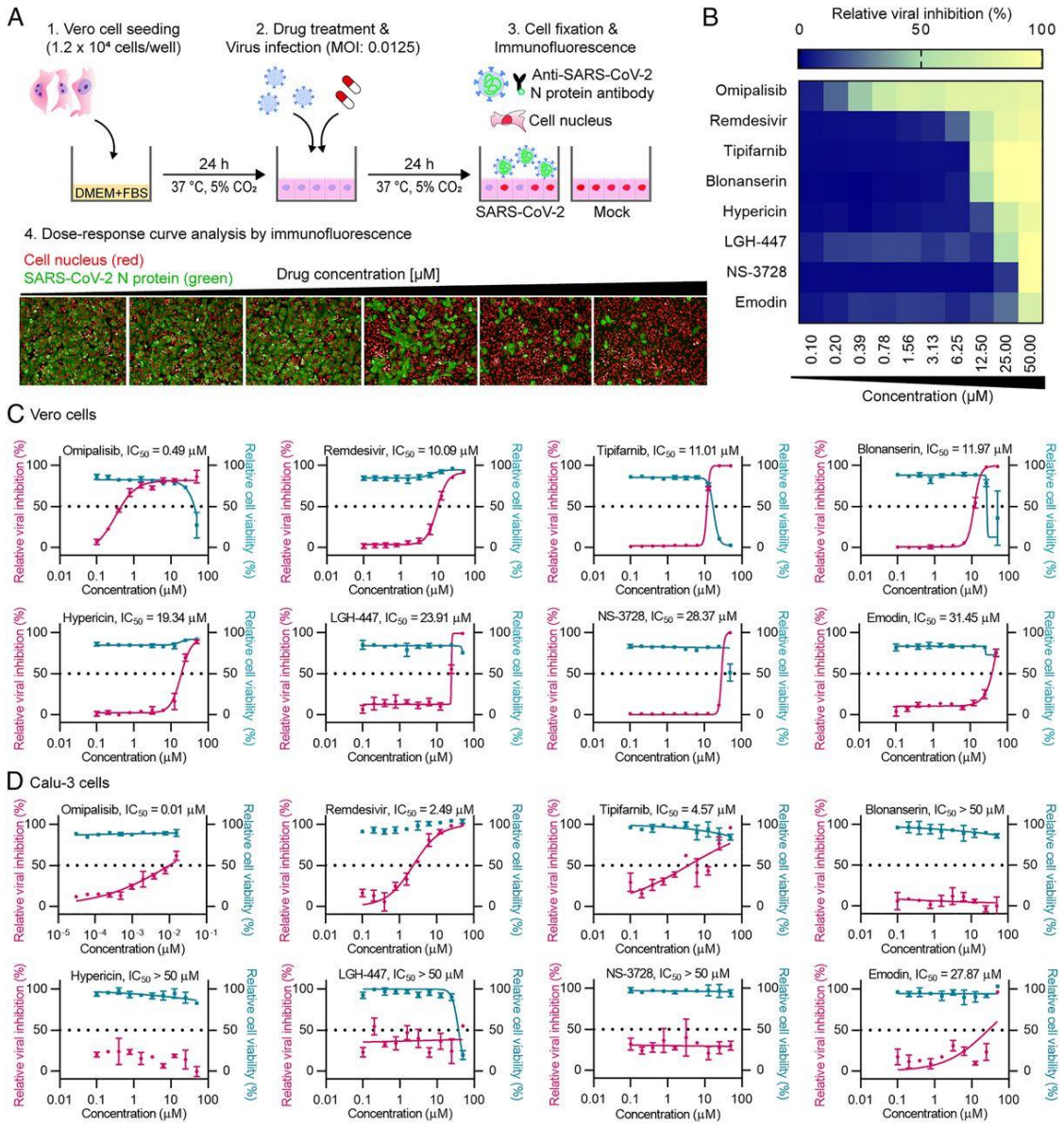


Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay

- **Docking-based virtual screening** can identify novel compounds against targets of SARS-CoV-2 among the collection of approved and clinical trial drugs.

- Computational drug repurposing is an effective approach to identify novel drug-target interactions using the drugs already known to be safe, which provides the advantages of significantly reducing time for drug development and reduced failure rate.

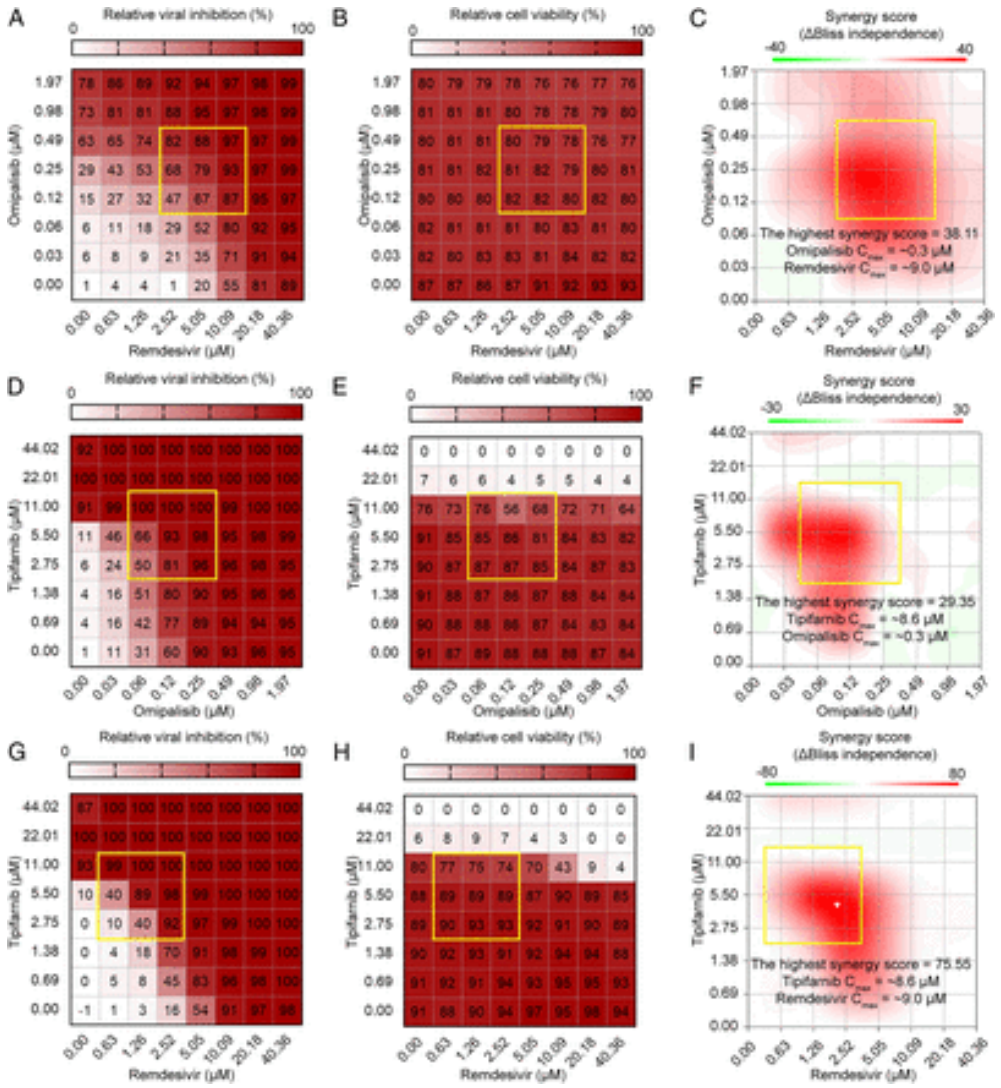
PNAS 2021 Vol. 118 No. 30



Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay

- Fig. 3. Dose–response analysis for the compounds having anti–SARS-CoV-2 activity. (A) A schematic of the immunofluorescence-based assay to examine anti–SARS-CoV-2 activity in Vero cells using the compounds selected from virtual screening.
- (B) A heatmap representing the percentages of normalizedinfection of the eight compounds in dose–response, on a scale from 0 to 100, depicting the average of duplicate independent experiments.
- Dose–response curves of the potent compounds in Vero cells (C) and Calu-3 cells (D).
- Pink line indicates relative viral inhibition and the blue line indicates relative cell viability.
- Data are normalized to the average of DMSO-treated wells and shown as the mean ± SD of duplicate independent experiments.

PNAS 2021 Vol. 118 No. 30



Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay

- Fig. 4. Analyses of drug combinations on anti–SARS-CoV-2 activity, cell viability, and their synergistic effects.
- Two-dimensional matrix of dose–response for relative viral inhibition:
  - omipalisib/remdesivir,
- (D) tipifarnib/omipalisib,
- (G) tipifarnib/remdesivir.
- The heatmap depicts relative viral inhibition scaled to
- the range of 0 to 100%.
- Two-dimensional matrix of dose–response for relative cell viability:
- (B) omipalisib/remdesivir,
- (E) tipifarnib/omipalisib,
- (H) tipifarnib/remdesivir. The heatmap depicts relative cell viability scaled to the range of 0 to 100%.
- Topographic two-dimensional map of synergy scores determined by synergyfinder using the data in A, D, and G, respectively: (C) omipalisib/remdesivir, (F) tipifarnib/omipalisib, and (I) tipifarnib/remdesivir. The synergy map highlights synergistic and antagonistic dose regions in red and green colors, respectively. A yellow box represents the area with the highest synergy score obtained by synergyfinder.

Questions

zsuzsanna.gaborik@crl.com