Computer-Aided Drug Design

Dmitri Kireev
Professor, Chemical Biology and Medicinal Chemistry
Director, Computational Biophysics and Molecular Design
Center for Integrative Chemical Biology and Drug Discovery
UNC Eshelman School Of Pharmacy
University of North Carolina at Chapel Hill

Outline

- Introduction to Drug Discovery
  - Process
  - Deliverables
  - Technologies
- Computer-Aided Drug Design
  - Strategies
  - Techniques
  - Examples

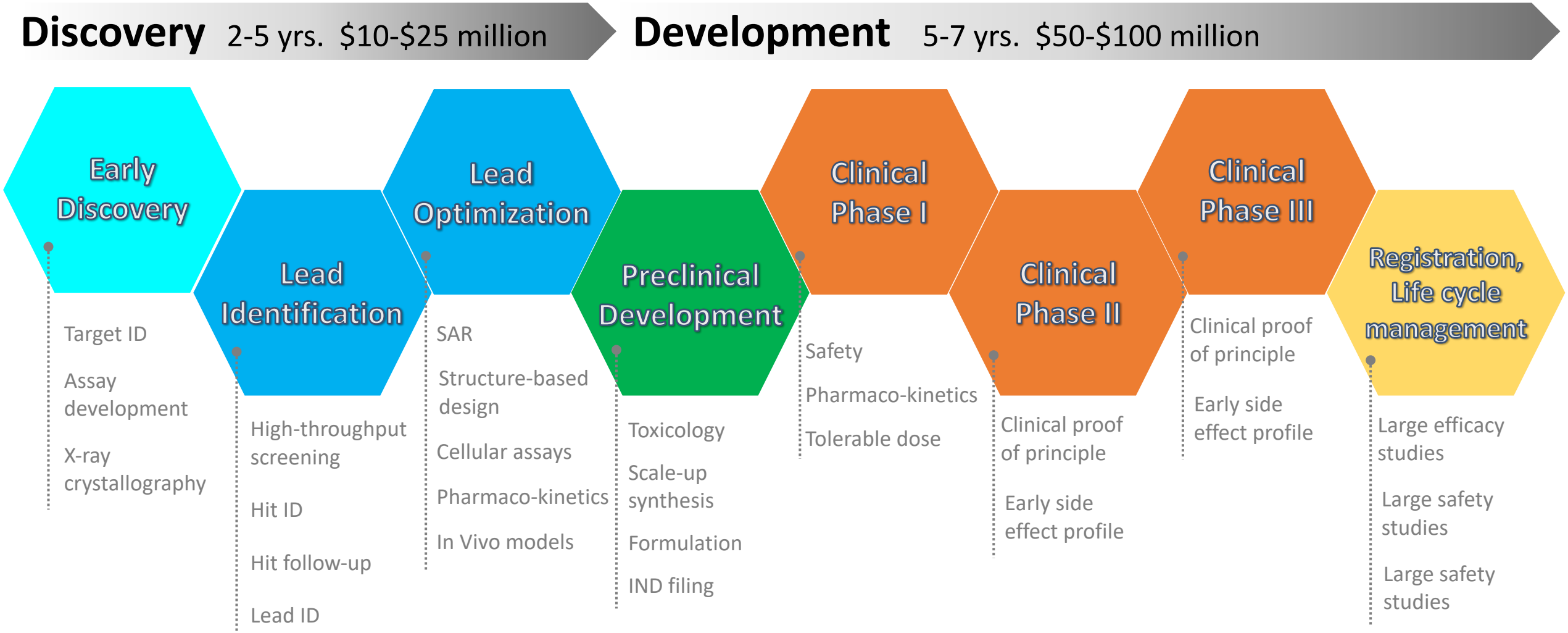
Central premises of modern drug discovery

- Disease is caused by a perturbed function of a protein (target). Modulating the target by a chemical compound (drug) may cure the disease
  - Not a law of nature, but works often enough to be useful
  - This makes of pharma companies a major consumer of latest advances in biomedical research

- A new drug must be safe for intended use
  - ... that is, provide greatest therapeutic benefit without resulting in unacceptable side effects or toxicity
  - This makes drug development a long and highly regulated endeavor (for our own good)

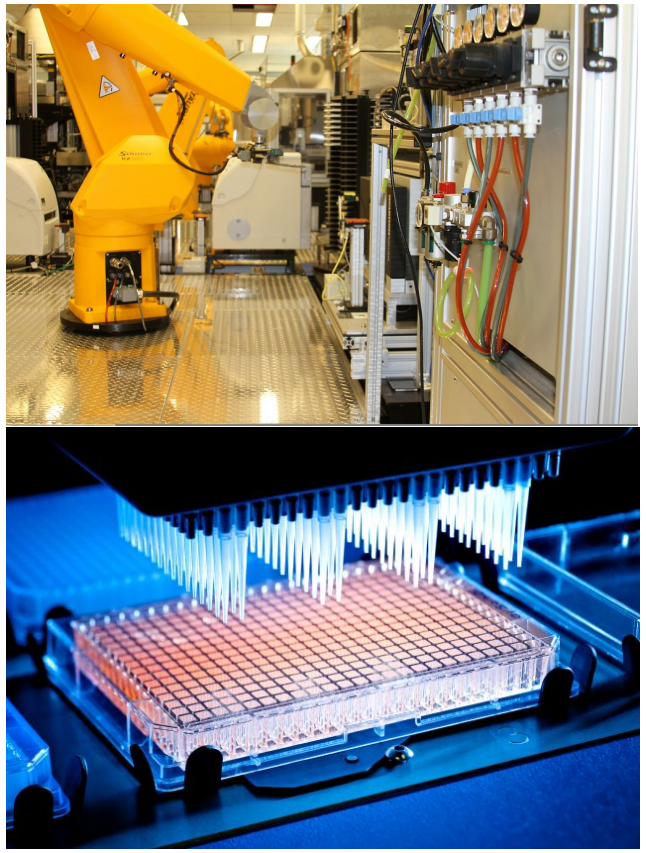
- Drug discovery is a costly endeavor – a drug molecule must be covered by a patent, so that the drug developer could protect its investment
  - This makes pharma companies reluctant to share data and knowledge to expedite the discovery process, for the common good
  - This is changing – there is a rising interest in open discovery models

Drug discovery and development cycle



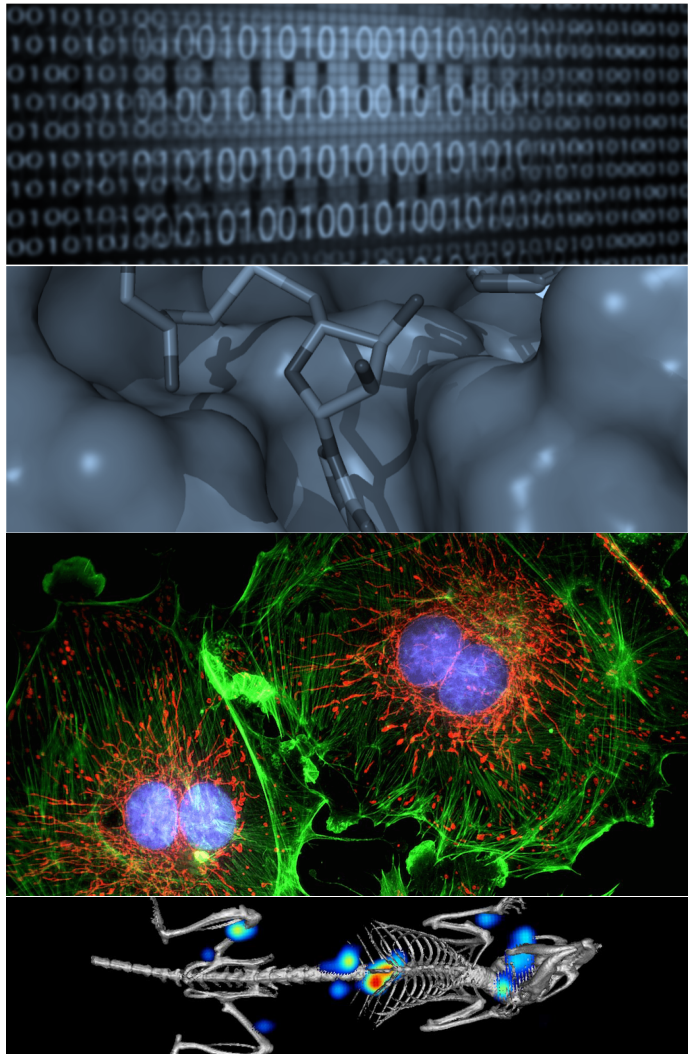
Lead identification

- High-throughput screening (HTS)
  - A typical screen in a big pharma company involves 500K to 3M compounds and takes days to weeks
  - highly automated and miniaturized (384 to 1534 well plates
  - can be functional (agonist/antagonist, inhibitor/activator) or binding
  - most often, a fluorescent readout
- Hit identification and follow-up
  - HTS outcome (usually tens to thousands of compounds showing "positive" signal) is subject to multiple false positives and false negatives
  - need confirmation by "orthogonal" techniques
  - Collect more information on true positives
    - pharmacokinetics, toxicity, intellectual property, synthetic feasibility, etc.
- Lead identification
  - Use information collect to make decision on which 1-3 hit series will be selected as leads for further optimization



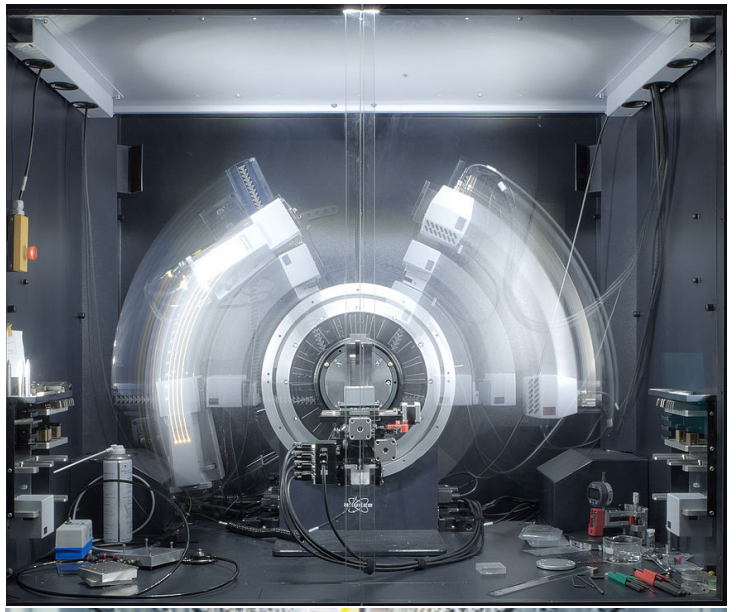
Lead Optimization

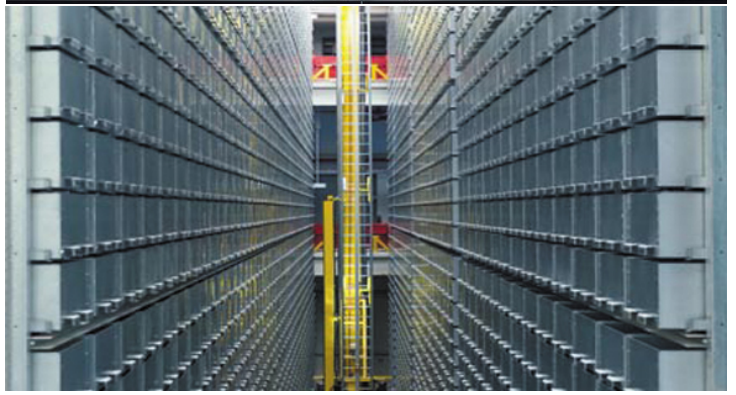
- Structure-Activity Relationships (SAR) analysis
  - Synthesis, testing and analysis of structurally related compounds to "take control" of activity (in any assay)
- Structure-based design
  - Making use of the 3D structure of the protein target to obtain better binders
- Cellular assays
  - Engineer or isolate diseased cells to test the lead's potential as a drug
    - Not always possible (e.g., Alzheimer's disease)
- *In vivo* models
  - Engineer animal organisms with human-like pathologies
    - Not always possible or not faithful enough due to differences between animals and humans at both molecular and systems levels



Drug Discovery Technologies

- Protein purification/production
- X-ray crystallography
- Organic synthesis
- Assay technologies
  - Biophysical
  - Biochemical
  - Cell-based
- Assay miniaturization
- Compound library management
- Transgenic organisms (disease models)





Technological revolutions

“From now on, things will never be the same”

- X-ray crystallography
- Combinatorial chemistry
- High-throughput screening
- Whole genome sequencing
- Fragment-based discovery
- Virtual screening
- Phenotypic screening
- DNA-encoded libraries
- Protein degraders
- Artificial Intelligence

Hype cycle

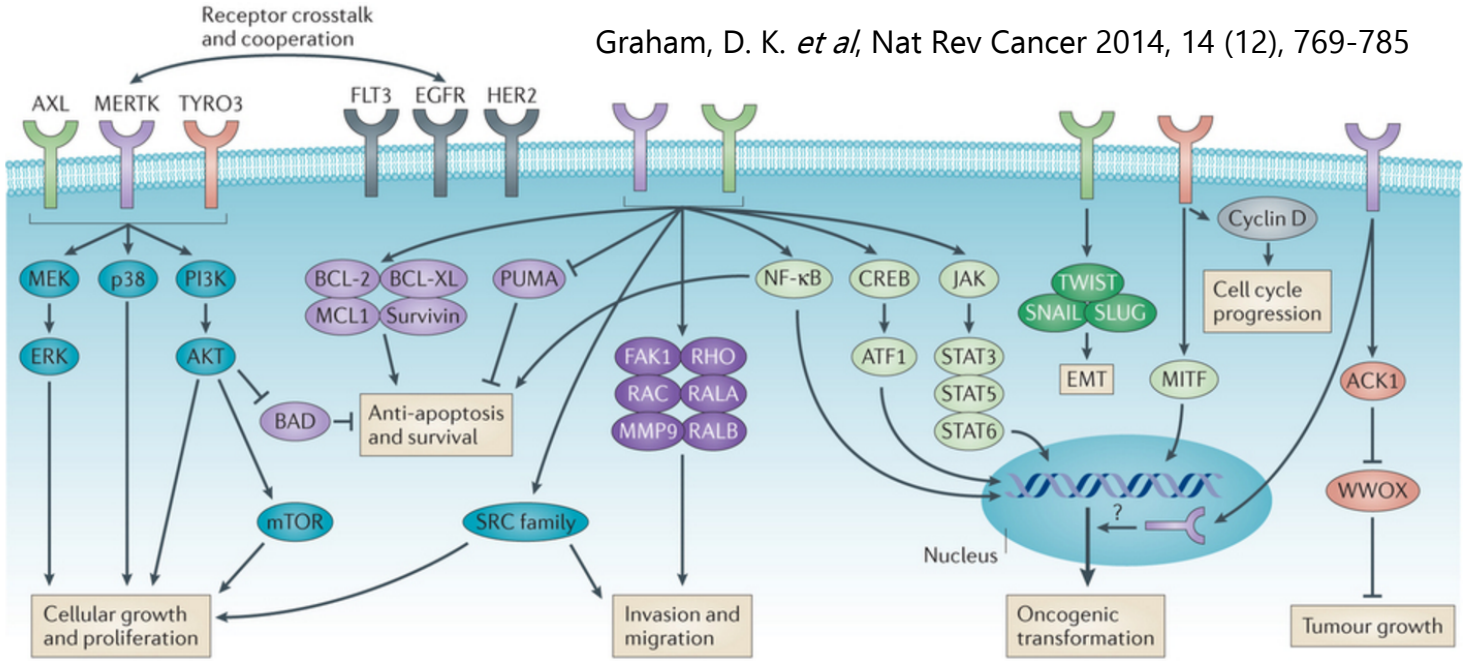
A “typical” discovery project – Part I

TAM-targeted cancer therapeutics

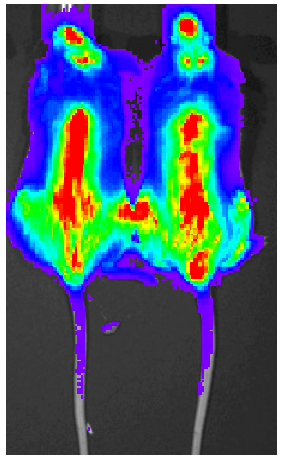
H. Shelton Earp

Douglas K. Graham

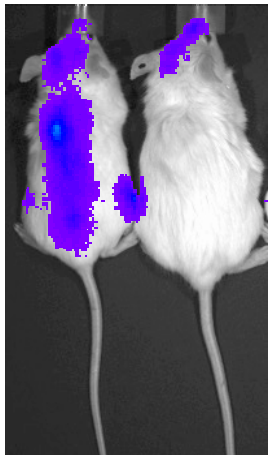
- Tyro3/Axl/Mer (TAM) RTK family
  - Expressed in monocytes to clear apoptotic material; never expressed in normal T or B lymphocytes
- Expressed in human cancer
  - MER: 30-40% T cell Acute Lymphoblastic Leukemia (ALL); MER/AXL: 41% B cell ALL and 68% pediatric AML
- Oncogenic function of ectopic expression
  - Survival signaling – anti-apoptosis
  - Critical for an “immune system” of a cancer cell
- Promising targets for cancer therapeutics



Graham, D. K. *et al*, Nat Rev Cancer 2014, 14 (12), 769-785



Jurkat T cells wild type



Jurkat T cells Mer knockdown

A “typical” discovery project – Part II

ACS Medicinal Chemistry Letters Letter

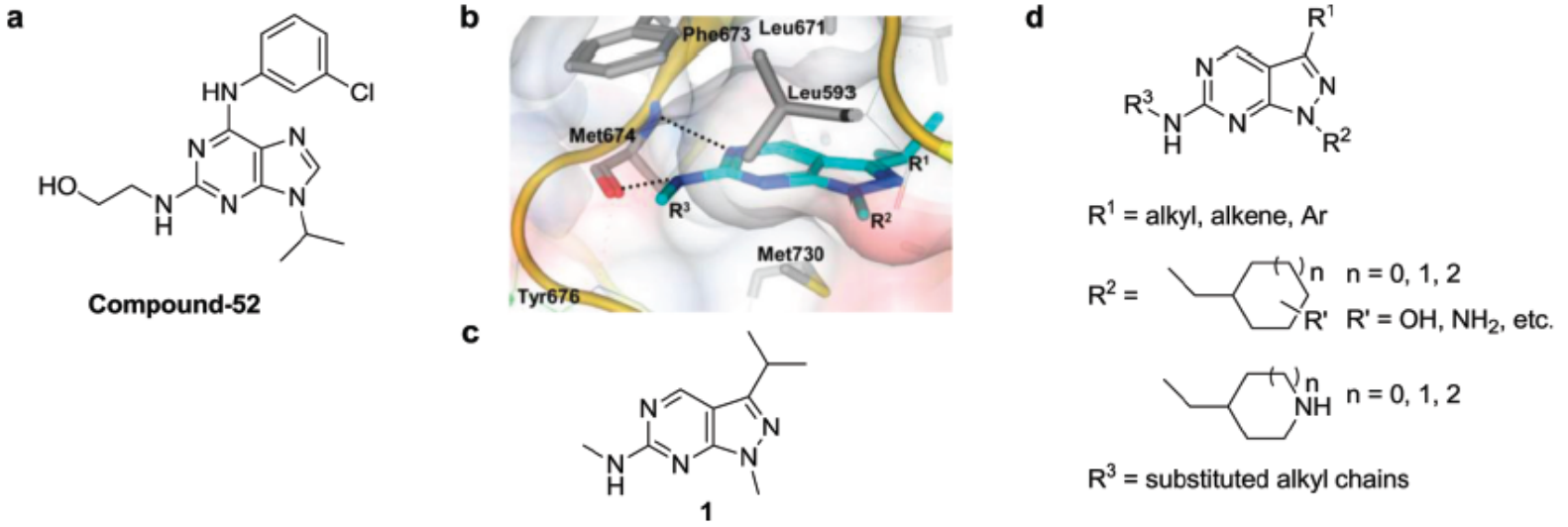


Figure 1. (a) Chemical structure of Compound-52; (b) docking model of **1** in the X-ray structure of Mer; (c) chemical structure of **1**; (d) design of candidate Mer kinase inhibitors based on the pyrazolopyrimidine scaffold.

Stephen V. Frye

Xiaodong Wang

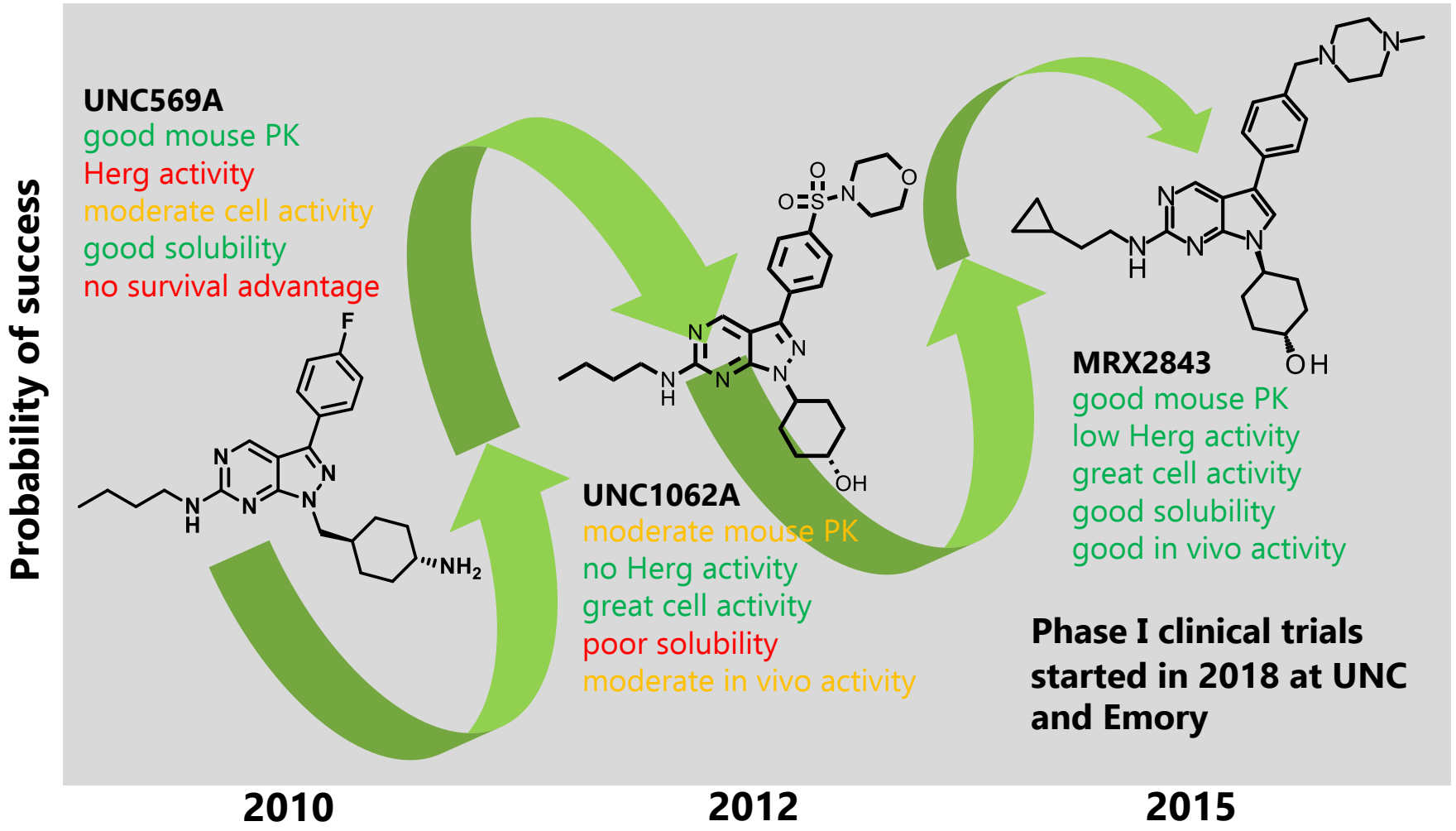
Dmitri Kireev

William Janzen

Huang, X. et al. Structural insights into the inhibited states of the Mer receptor tyrosine kinase. *J Struct Biol* **165**, 88-96 (2009).

Liu, J. *et al.* Discovery of Novel Small Molecule Mer Kinase Inhibitors for the Treatment of Pediatric Acute Lymphoblastic Leukemia. *ACS Med Chem Lett* **3**, 129-134 (2012).

A “typical” discovery project – Part III



Computational drug discovery

Concepts

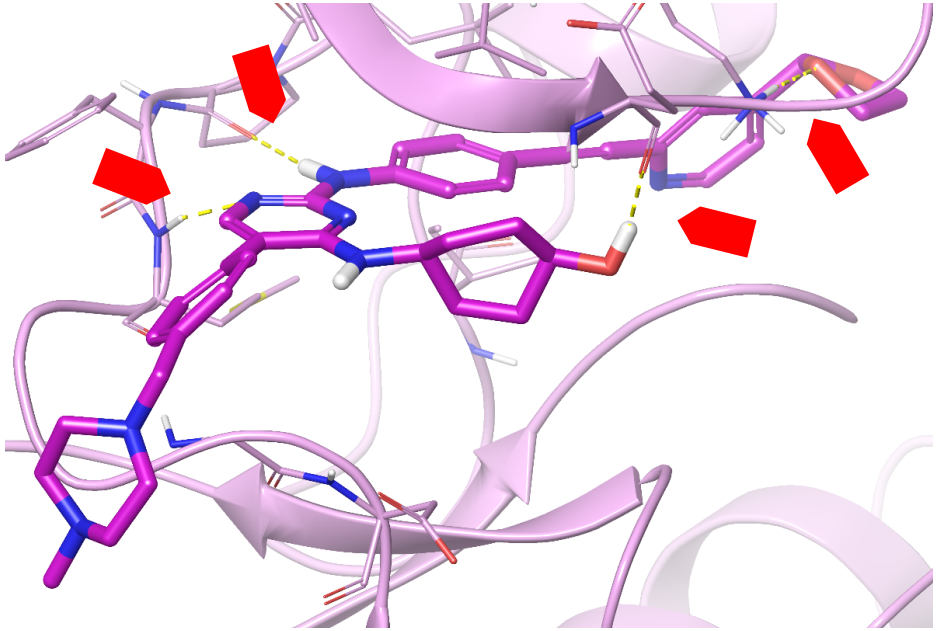
- Structure-based
- Ligand-based

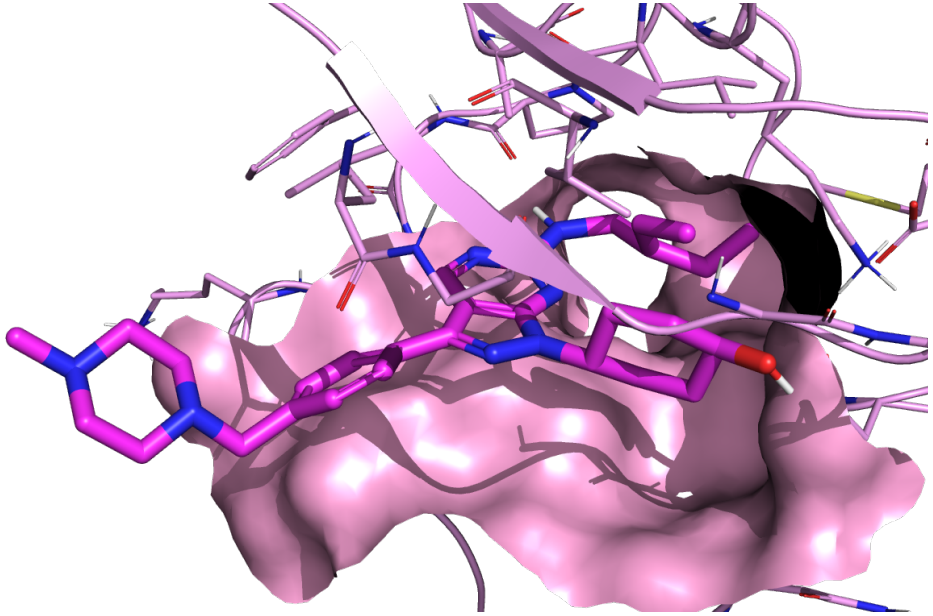
Modes

- Predictive
- Design

Structure-based design

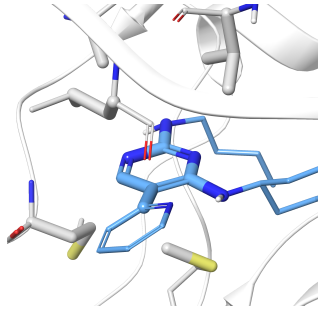
- Goal: Design new small-molecule ligands by making use of 3D structure of the target protein
  - May be used when there are no ligands known to the target protein
- Making incremental changes to a ligand
  - to induce “good” interactions
  - to eliminate “bad” contacts
  - to minimize the entropic penalty
- Intermolecular interactions
  - Electrostatic
  - Hydrogen bonds
  - Van der Waals
  - Cation-p, aromatic stacking, halogen bonds
  - Entropic – solvation
  - Entropic – configurational

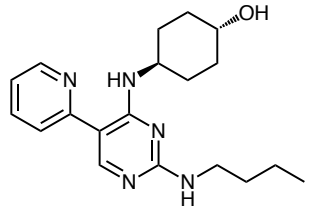




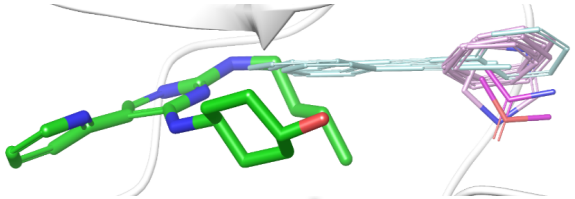
Data-driven design of *in vivo* anti-tumor probes

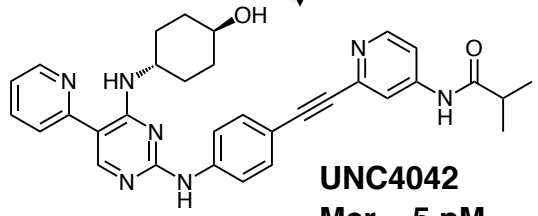
- A new technology to assemble ligands from fragments directly in the protein's binding site was developed
- Unlike most computational techniques, that are "virtual screeners", FRASE-based design is a "virtual medicinal chemist"
- The new concept was used to design a series of TAM inhibitors with better intra-TAM and kinome-wide selectivity by extending the template type I kinase inhibitors toward the back-pocket





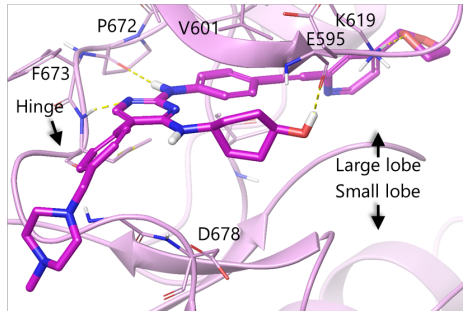
UNC1917 (PDB: 4M3Q)
Mer = 17 nM
Axl = 920 nM
Tyro3 = 570 nM
FLT3 = 38 nM



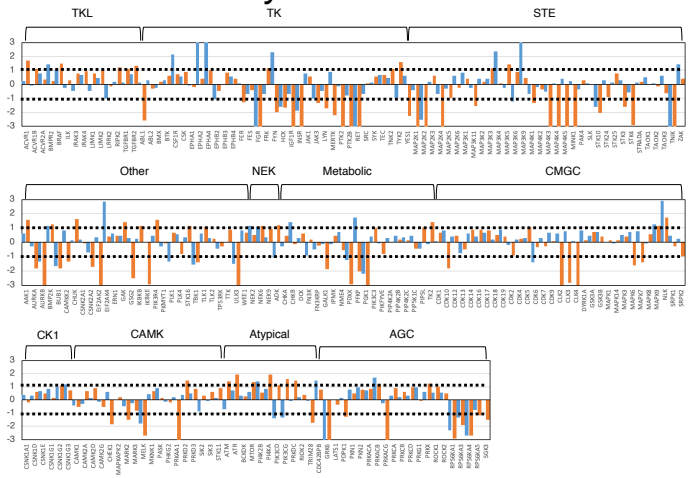


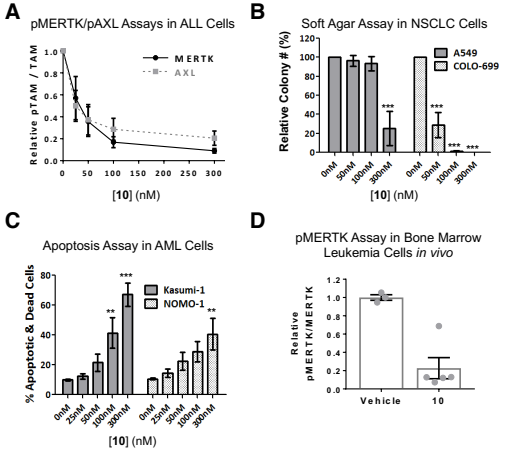
UNC4042
Mer = 5 nM
Axl = 8 nM
Tyro3 = 7 nM
FLT3 = 29 nM

X-ray confirmation



X-ray confirmation

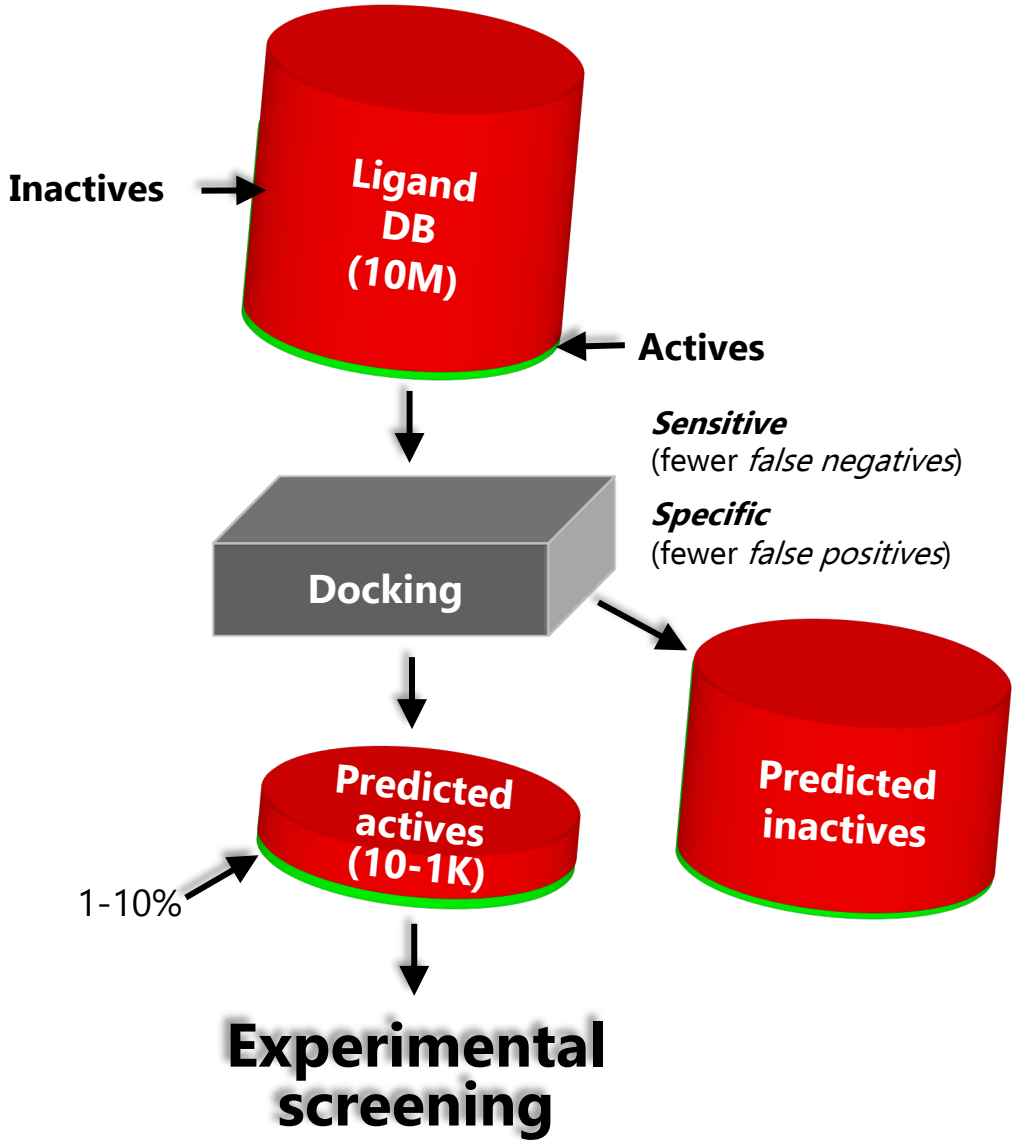




Da et al., *J. Am. Chem. Soc.* 2019, 141, 39, 15700–15709

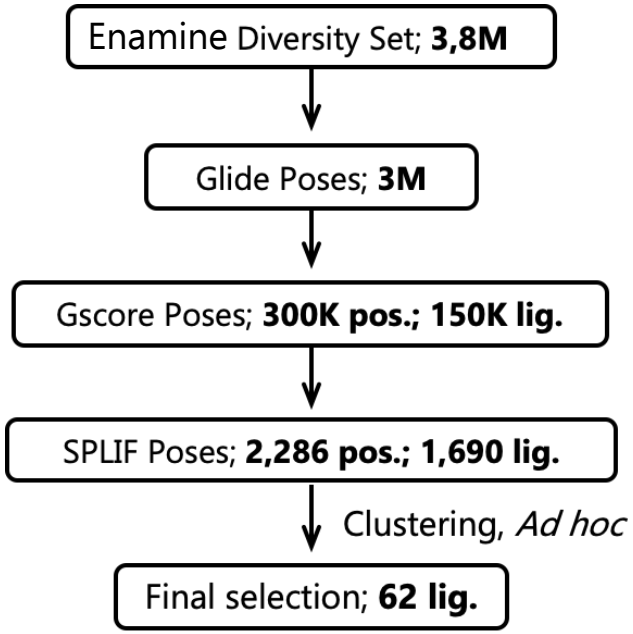
Docking / Virtual screening

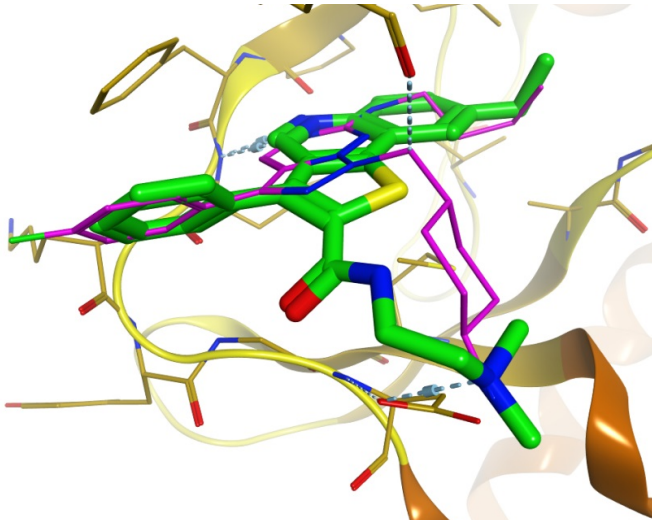
- Docking is an algorithm to determine whether and how a ligand binds to a protein target
  - Scoring function calculates, in a fraction of seconds, a (very) rough estimate of binding free energy that is used to rank docked ligands from the least to most plausible target binder
  - Algorithms:
    - Hundreds developed, but only a few are frequently used
    - **AutoDock**, DOCK, FlexX, FRED, **Glide**, GOLD, ICM, QXP/Flo+, Surflex
  - seconds per ligand
- Often used to perform virtual screening of large compound databases
  - do not need to be an in-house collection
  - can be millions of commercially available or potentially synthesizable compounds
- Output is a compound set enriched in true actives



Discovery of Mer kinase inhibitors by virtual screening

- Techniques
  - Docking / Scoring: Glide
  - Re-scoring: Structural Protein-Ligand Interaction Fingerprints (SPLIF)
- Screening library
  - Enamine (~4 million compounds)
- Experimental testing (62 compounds)
  - Microfluidic Capillary Electrophoresis (MCE) assay
- Hit overview
  - 15 hits with valid dose-response curves (24%)
    - IC50 ranges from **0.46** to 9.9 μM
  - Enrichment = **200-fold** (24% / 0.12%)





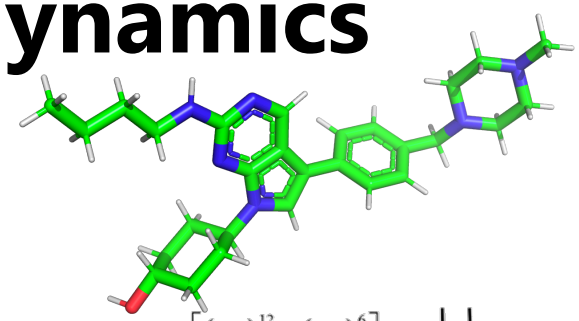
IC50 = **0.46** μM

SPLIF-selected pose:
an intuitive, but chemically non-trivial similarity

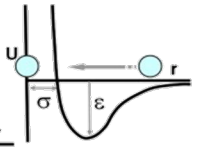
Da *et al.*, *Bioorg. Med. Chem.*, 2015, 23, 1096

Molecular mechanics / Molecular dynamics

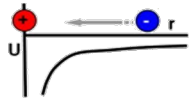
- Molecular Mechanics
  - Goal: to predict 3D molecular structure
  - simple empirical equations – collectively called force fields (FF) – are used to calculate the energy of inter-atomic interactions (instead of solving corresponding Schrodinger equation)
  - a search algorithm (e.g., Newton-Raphson) can be used to find a molecular geometry with a minimal FF energy
- Molecular dynamics
  - Goal: to simulate natural molecular motions
    - ideally, to determine the partition function for the molecule of interest, that is, how much time the molecule spends in each of its microstates
  - Each atom is given an initial momentum; their 3D coordinates are recalculated after each time step *dt* using classical equations of motion and potential energy from force fields



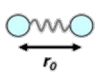
$$U = \sum_{i<j}\sum 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$



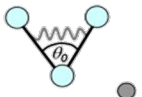
$$+ \sum_{i<j}\sum \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$



$$+ \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2$$

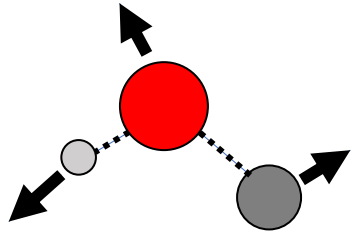


$$+ \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2$$



$$+ \sum_{torsions} k_\phi [1 + \cos(n\phi - \delta)]$$

https://slideplayer.com/slide/8858697/



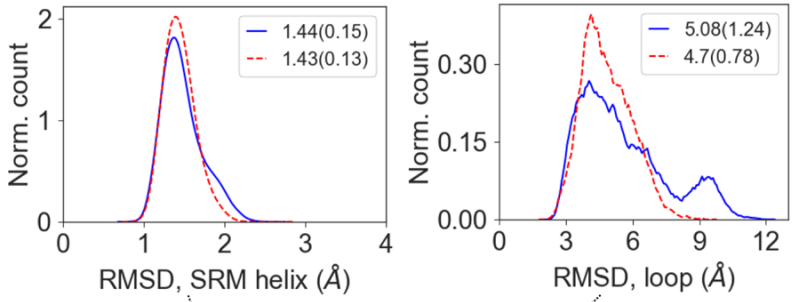
$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{d\mathbf{v}_i}{dt} = \mathbf{a}_i(t) = \frac{\mathbf{F}_i(\mathbf{r}(t))}{m_i} = -\frac{1}{m_i}\nabla_{\mathbf{r}_i} V(\mathbf{r}(t))$$

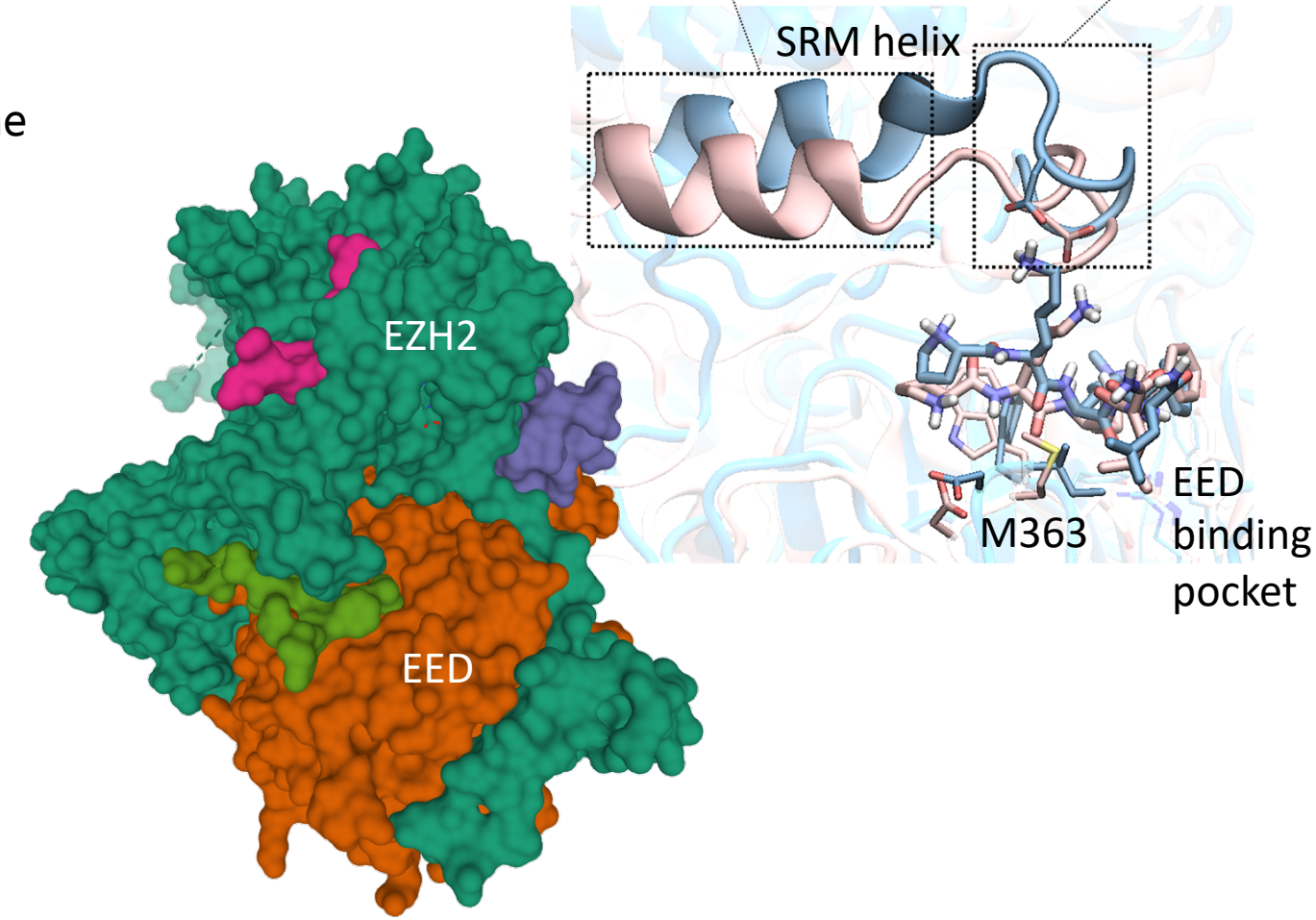
$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 + \ldots$$

https://slideplayer.com/slide/8999257/

Discovery of allosteric activators of PRC2 mutant

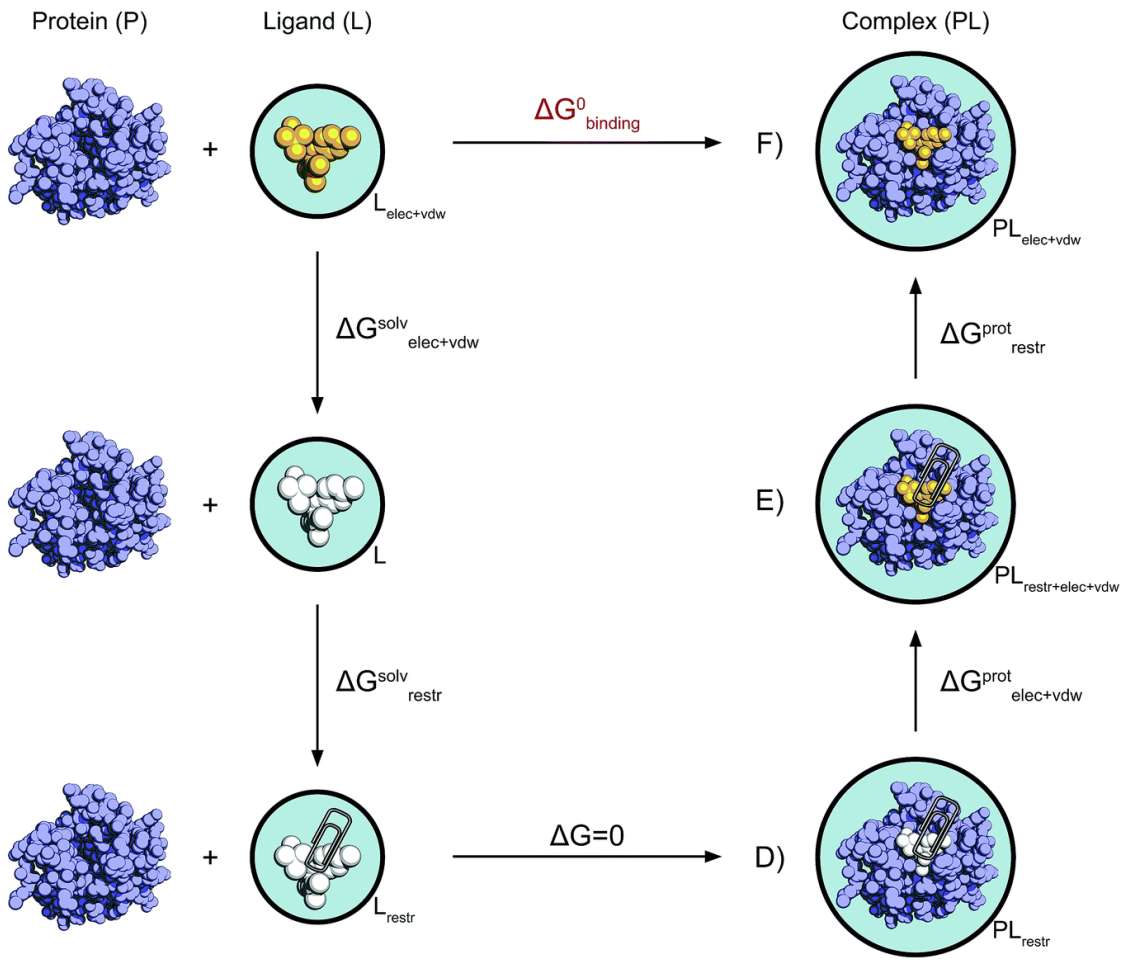
- Goal: Develop selective allosteric activators of the PRC2 mutant EED-I363M
- Computational contribution:
  - Elucidated differences in dynamics of the mutant and wild type complexes
  - Determined that a specific activator for the mutant should satisfy two conditions
    - Fit to a larger pocket of the mutant EED
    - Bind to and stabilize SRM helix of EZH2, a histone lysine methyltransferase
  - Proposed necessary ligand modifications
- Compounds were made, experimentally characterized and confirmed predictions





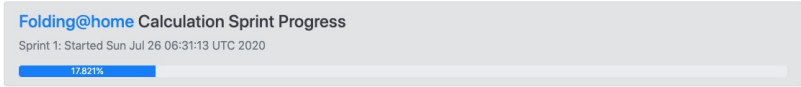
Prediction of ligand-protein binding affinity

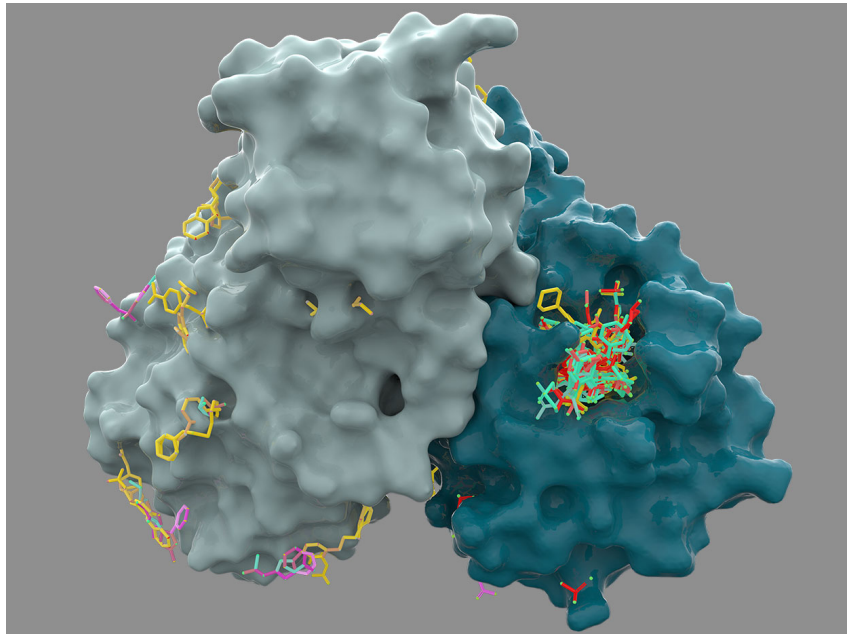
- Accurate prediction of the protein-ligand binding free energy (ΔG) used to be a "holy grail" in the molecular modeling community and there is a recent regain of interest to this field
- The most general and thorough way to calculate ΔG would be to simulate a ligand and a receptor in a box long enough and see how much time would it spend in a protein-bound state
  - would take hundreds of years of calculation on current computers
- Next best option is making use of free energy perturbation theory
  - to calculate differences between binding free energies of two very similar ligands ($\Delta\Delta G$)
- Then, absolute binding free energy can be calculated through progressive waning of the ligand in the binding site



COVID Moonshot project

- COVID Moonshot aims to rapidly develop new therapies against the SARS-CoV-2 main viral protease
- Runs on Folding@Home, a distributed computer system
  - claimed by the protagonists as the largest supercomputer in the world)
- This pro bono initiative crowdsourced 4,500 drug designs, synthesized 311, and is now testing them against viral proteins

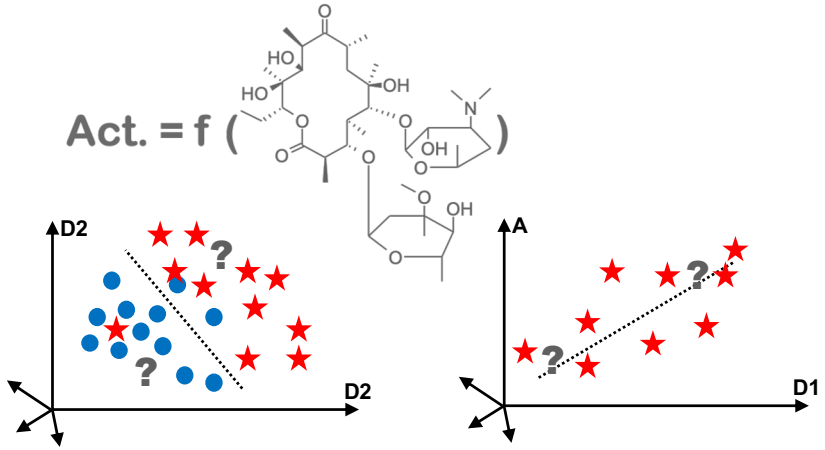




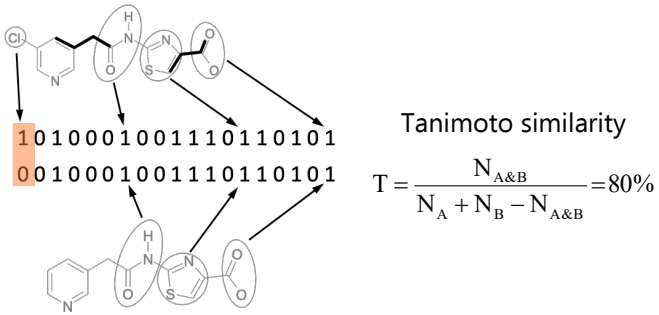
Ligand-based design

- Goal: Design or test new ligands based on the analysis ligand structures
  - No need in 3D protein structure
  - Requires known examples of ligands to the protein target of interest
  - is the approach medicinal chemists use for over a century
- Makes use of the structure-activity relationships (SAR), an assumption that similar compounds show similar biological activity
  - not a law of nature, but works more often than not
- Chemoinformatics quantifies and automates multiple design tasks
  - Q(uantitative)SAR, a machine learning (ML)-based approach that helps to build models for ligand-based virtual screening
  - Similarity/diversity analyses help to more efficiently create compound libraries for experimental screening

QSAR



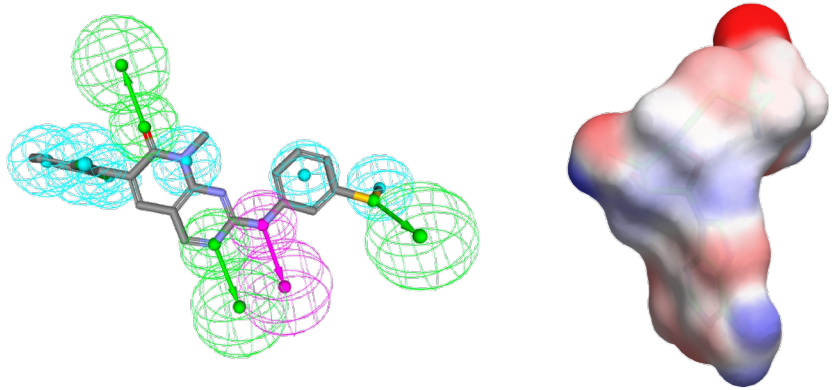
Fingerprint similarity for fast search/analysis



Tanimoto similarity

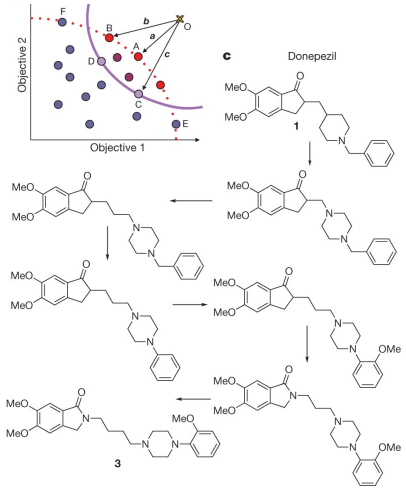
$$T = \frac{N_{A\&B}}{N_A + N_B - N_{A\&B}} = 80\%$$

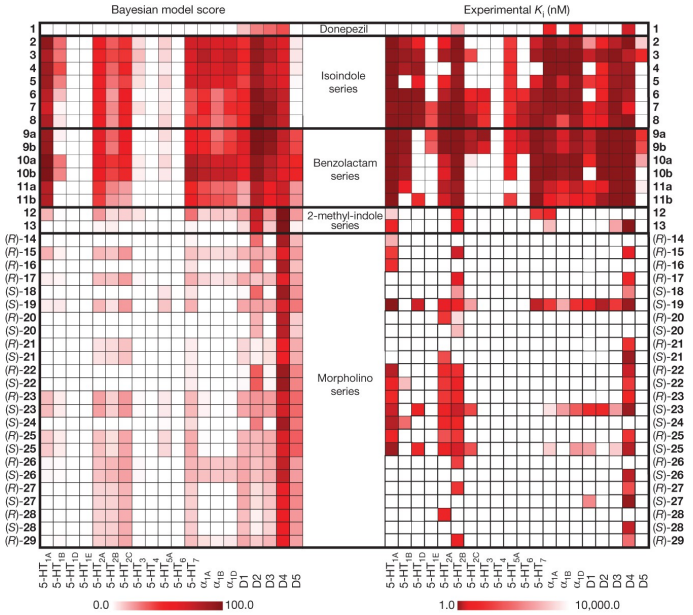
(Less intuitive) pharmacophore/shape similarity



Automated design of ligands to polypharmacological profiles

- Designing drugs with a specific multi-target profile is both complex and difficult
- Authors proposed a new approach for the automated design of ligands against profiles of multiple drug targets
- Demonstrated by the evolution of an approved acetylcholinesterase inhibitor drug into brain-penetrable ligands with either specific polypharmacology or selectivity profiles for G-protein-coupled receptors
- 800 ligand–target predictions of designed ligands were tested experimentally
  - 75% were confirmed to be correct.
- Selected leads demonstrate target engagement *in vivo*





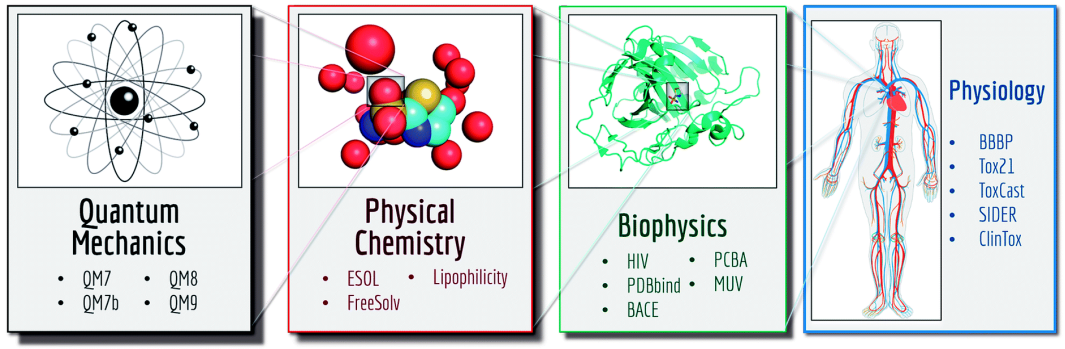
Besnard *et al.*, *Nature*, v. 492, p. 215–220 (2012)

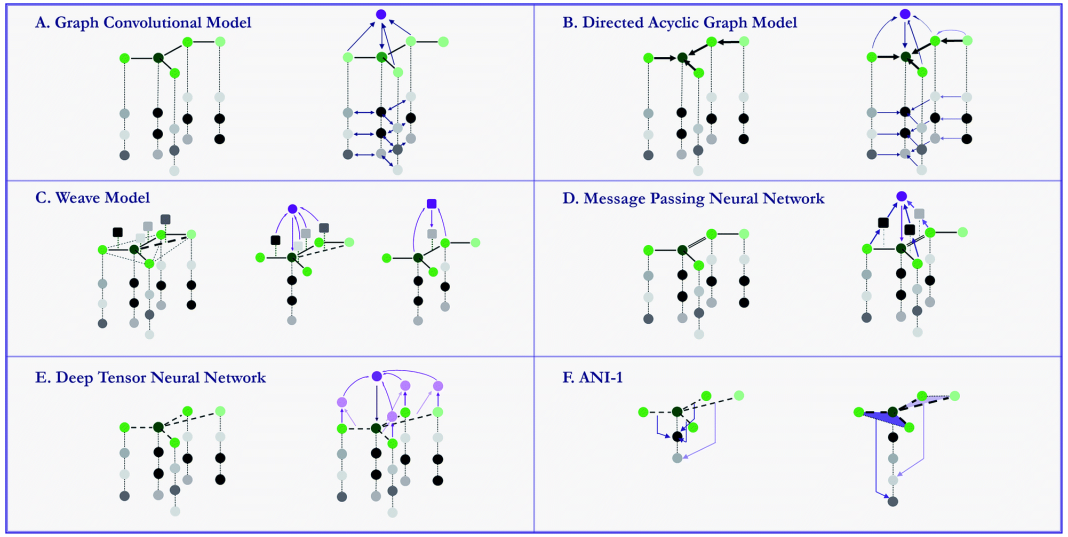
https://www.nature.com/articles/nature11691

Artificial intelligence

- Multiple applications from most obvious to AI-specific
- Deep neural networks as machine learning (ML) techniques for Quantitative Structure-Activity/Property Relationships
- Molecular Mechanics with deep-learned force fields
- Deep-learned quantum chemistry models
- Generative neural networks to enumerate novel compounds with desired properties

MoleculeNet: a benchmark for molecular machine learning



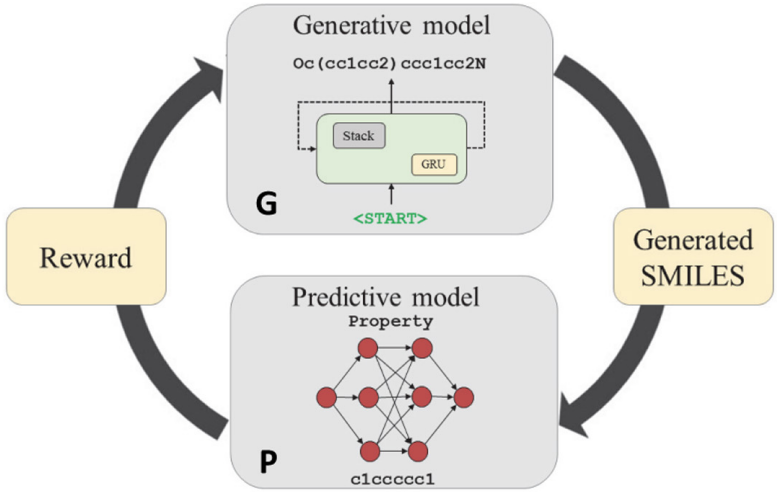


Wu *et al.*, *Chem. Sci.*, 2018, 9, 513-530
https://pubs.rsc.org/--/content/articlehtml/2018/sc/c7sc02664a

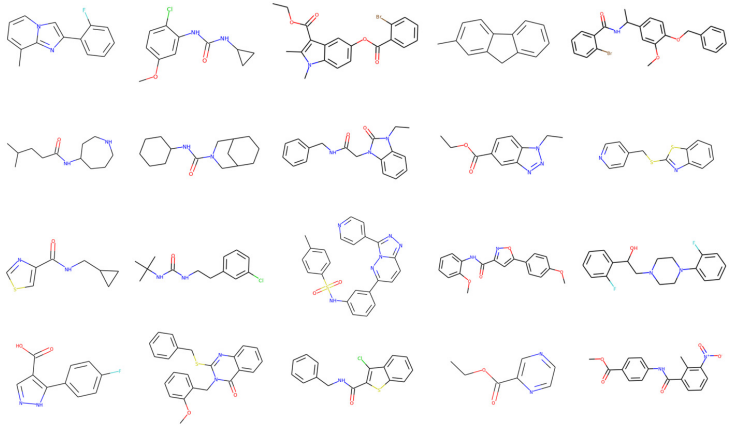
Deep reinforcement learning for de novo drug design

- Authors devised a novel strategy for de novo design of molecules with desired properties termed ReLeaSE (Reinforcement Learning for Structural Evolution)
- integrates two deep neural networks—generative and predictive – trained separately but are used jointly to generate novel targeted chemical libraries
- As a proof-of-concept, ReLeaSE was used to design chemical libraries with a bias toward
  - structural complexity
  - melting point or hydrophobicity
  - inhibitory activity against Janus protein kinase 2

Workflow of deep reinforcement learning for generating new SMILES



A sample of molecules produced by the generative model



Popova *et al., Science Advances,* 4, 7, eaap7885

https://advances.sciencemag.org/content/4/7/eaap7885?intcmp=trendmd-adv

Perspectives

- Increase the pace, cut the cost of developing probes and drugs for new targets
  - There are an estimated 3,000 to 10,000 disease-causing proteins that might be targeted by small molecule drugs. Less than a thousand are targeted by all current drugs. This means that the goal of having a small-molecule probe to every target would take decades (and tens of billion dollars) to accomplish
  - New computational strategies may help
- Need predictions for animal models and humans to decrease the attrition rate
  - Lack of efficacy, poor pharmacokinetics and elevated toxicity account for most of failures in clinics
  - Connect genomics to chemistry to clinical data

https://www.researchgate.net/profile/Dmitri_Kireev
https://twitter.com/deka27516
https://pharmacy.unc.edu/directory/kireev/
https://pharmacy.unc.edu/

https://www.sciencedirect.com/science/article/abs/pii/S0169409X16300370

Alternatives to drug discovery

- Biologics
- Medical devices
- Nanomedicine
- Regenerative medicine