6.S085 Machine Learning for Molecular Design

- We will start at 10:10 am
- Please fill the feedback survey for improving this class 👉

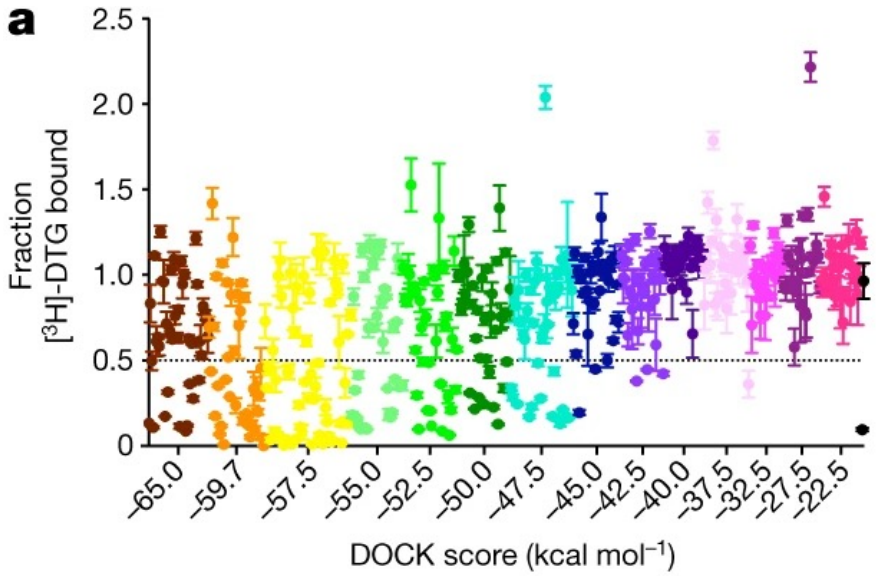
Outline

- Project results announcement
- Introduction to project oracle design
- Presentation from top-3 teams
- Final remark

Project Results Announcement

Course Project: a simulated drug design campaign

- We developed a simulation protocol based on molecular docking. We treat this oracle as ground truth to simulate experimental evaluation.
- We also add drug likeliness and synthesizability into consideration, resulting in a single scalar score, ranging from 0 to 1, to optimize.
- Your task is to design novel molecules that have higher activity under limited evaluation budget.



Course Project: a simulated drug design campaign

- Performance metric: (average score of top-30) + 0.3 * (Internal diversity of top-30)
- We will hold a leaderboard showing top-10 teams.

Internal diversity

We define the *internal diversity* I of a set of molecules A of size $|A|$ to be the average of the Tanimoto-distance T_d of molecules of A with respect to each other. Formally, we have:

$$I(A) = \frac{1}{|A|^2} \sum_{(x,y) \in A \times A} T_d(x, y) \qquad (1)$$

For a sufficiently large set A, any sufficiently large subset $A' \subset A$, sampled with uniform probability, has the same internal diversity as A. This property follows from the law of large numbers. We can thus define the internal diversity of a generative model, by computing the internal diversity of a sufficiently large generated sample. This allows to formalize our challenge:

Benhenda, Mostapha. "ChemGAN challenge for drug discovery: can AI reproduce natural chemical diversity?." *arXiv preprint arXiv:1708.08227* (2017).

Results

Final Project Leaderboard (showing top-10 teams only, finalized!)

🥇	**ChemPandas** (ID: 2)	Score: 1.18
🥈	**Molecular Designer 4** (ID: 4)	Score: 1.163
🥉	**Chem Duo** (ID: 5)	Score: 1.148
4	**Molecular Designer 18** (ID: 18)	Score: 1.13
5	**Molecular Designer 10** (ID: 10)	Score: 1.127
6	**Molecular Designer 8** (ID: 8)	Score: 1.116
7	**Molecular Designer 23** (ID: 23)	Score: 1.114
8	**Molecular Designer 7** (ID: 7)	Score: 1.082
9	**Molecular Designer 6** (ID: 6)	Score: 1.07
10	**Molecular Designer 21** (ID: 21)	Score: 0.979

$300

$200

$100

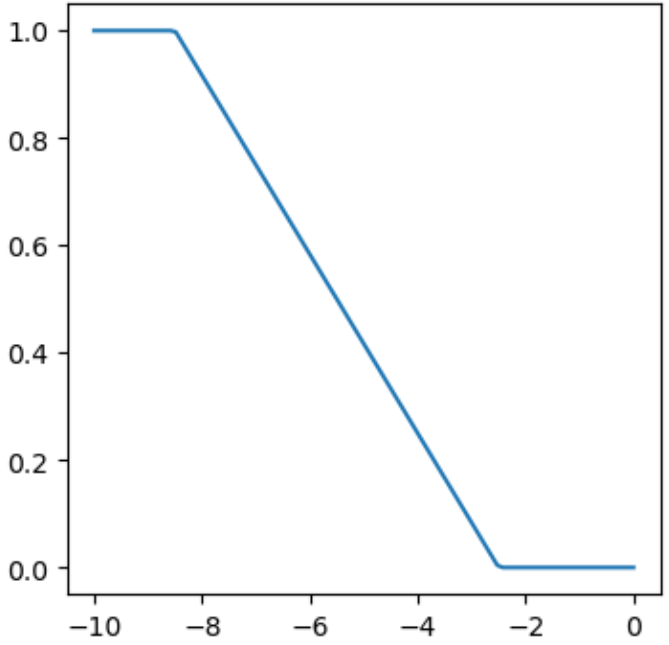
Oracle: a multi-objective oracle for drug discovery

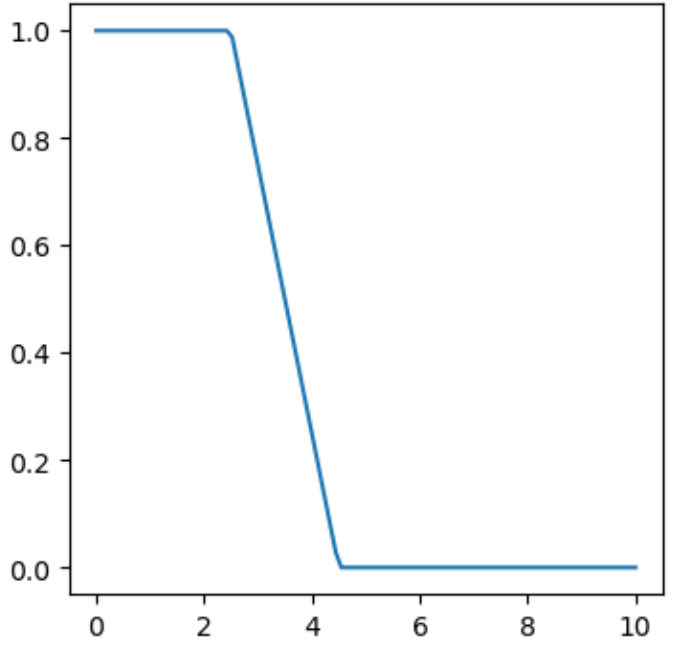
- Score = Vina score * Drug likeliness * Synthetic accessibility

Binding affinity

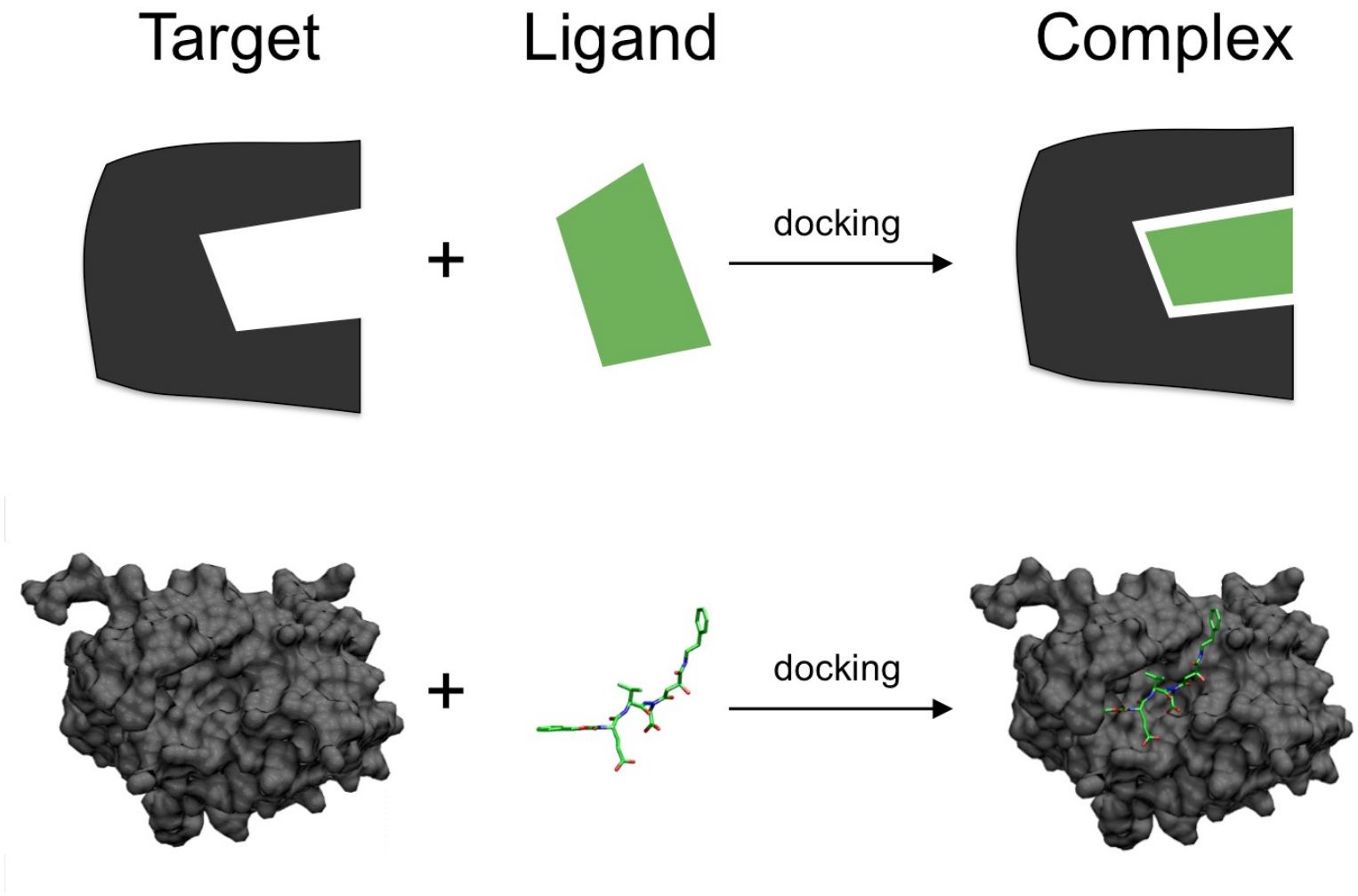
~ADMET

Synthesizability



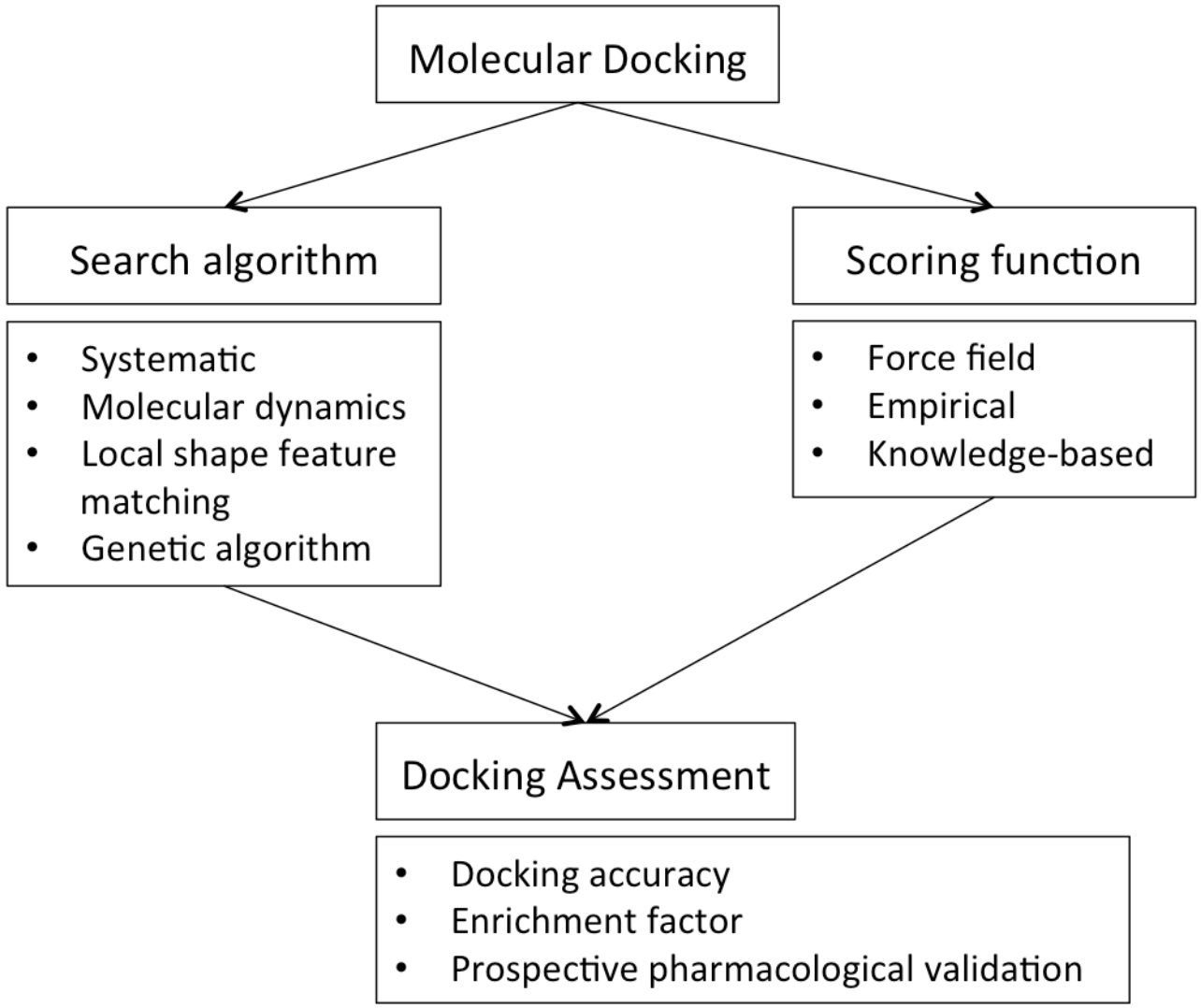


Molecular Docking



https://en.wikipedia.org/wiki/Docking_%28molecular%29

Molecular Docking Algorithm



https://en.wikipedia.org/wiki/Docking_%28molecular%29

How good are they?

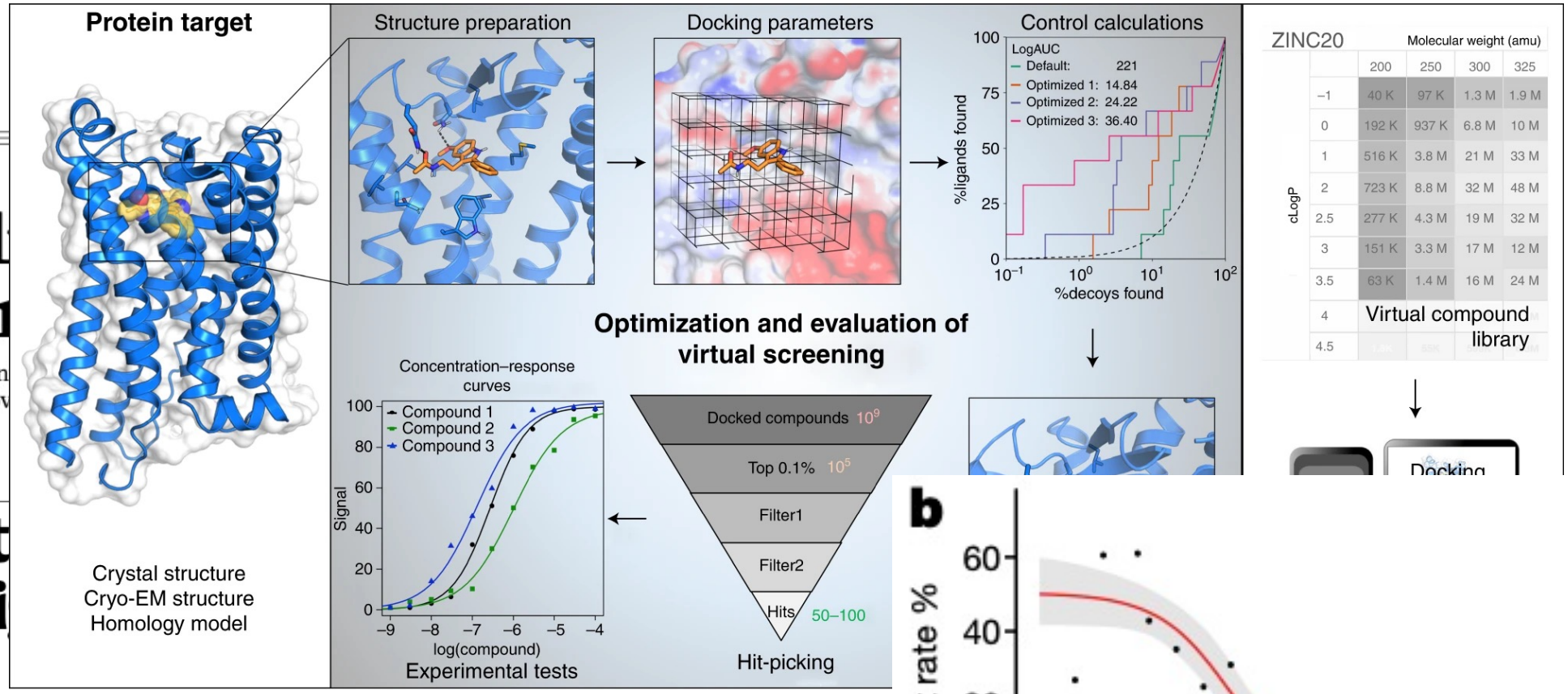
ARTICLE

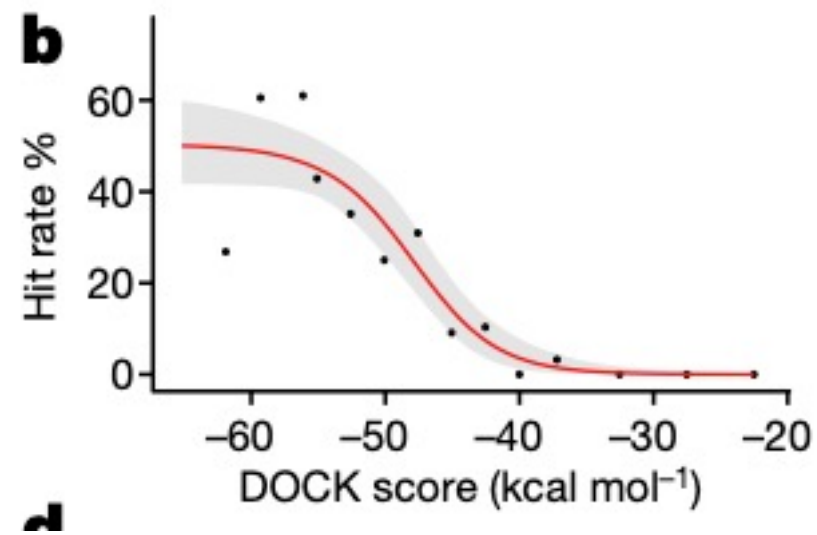
Ultra-large l
discovering

Jiankun Lyu[1,2,10], Sheng Wang[3,4,10], Tren
Enkhjargal Algaa[1], Kateryna Tolmachov

Article

Structures of t
for bioactive li





https://doi.org/10.1038/s41586-021-04175-x
Received: 20 April 2021
Accepted: 19 October 2021

Assaf Alon[1,12], Jiankun Lyu[2,12], Joao M. Braz[3,12], Tia A. Tummino[2,4], Veronica Craik[3], Matthew J. O'Meara[5], Chase M. Webb[2,4], Dmytro S. Radchenko[6,7], Yurii S. Moroz[8], Xi-Ping Huang[9,10], Yongfeng Liu[9,10], Bryan L. Roth[9,10,11], John J. Irwin[2], Allan I. Basbaum[3]✉, Brian K. Shoichet[2]✉ & Andrew C. Kruse[1]✉

Lyu, Jiankun, et al. "Ultra-large library docking for discovering new chemotypes." Nature 566.7743 (2019): 224-229.
Alon, Assaf, et al. "Structures of the σ2 receptor enable docking for bioactive ligand discovery." *Nature* 600.7890 (2021): 759-764.

AutoDock Vina, Main Protease of SARS-Cov-2

CENTER FOR COMPUTATIONAL STRUCTURAL BIOLOGY

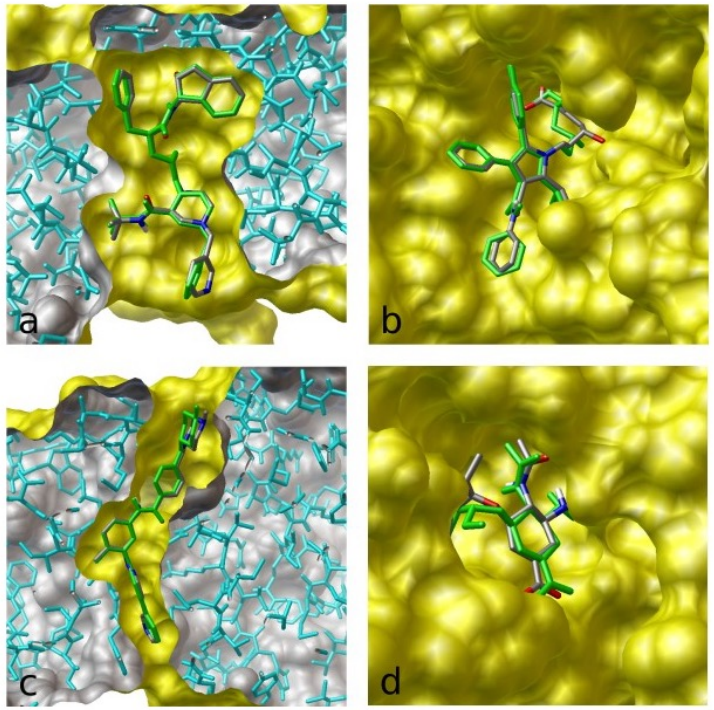
HOME PROJECTS PEOPLE GALLERY OPPORTUNITIES CONTACT

AutoDock Vina

Homepage

Downloads

Documentation >

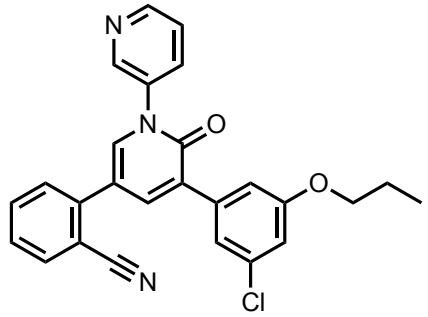


AutoDock Vina is an open-source program for doing molecular docking. It was originally designed and implemented by Dr. Oleg Trott in the Molecular Graphics Lab (now CCSB) at The Scripps Research Institute.

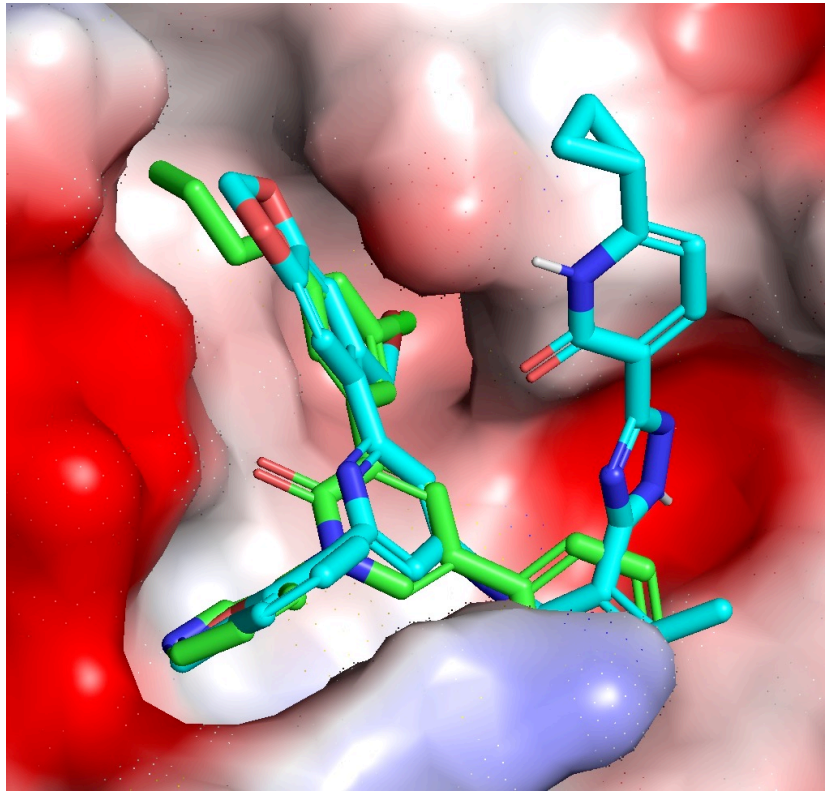
The latest version is available here.

AutoDock Vina is one of the docking engines of the AutoDock Suite.

The image on the left illustrates the results of flexible docking (green) superimposed on the crystal structures of (a) indinavir, (b) atorvastatin, (c) imatinib, and (d) oseltamivir bound to their respective targets.



Known Inhibitor
Vina score = -8.96 kJ/mol



Main Protease of SARS-Cov-2 (PDB: 7L11)

https://vina.scripps.edu

Quantitative Estimation of Drug likeliness (QED)

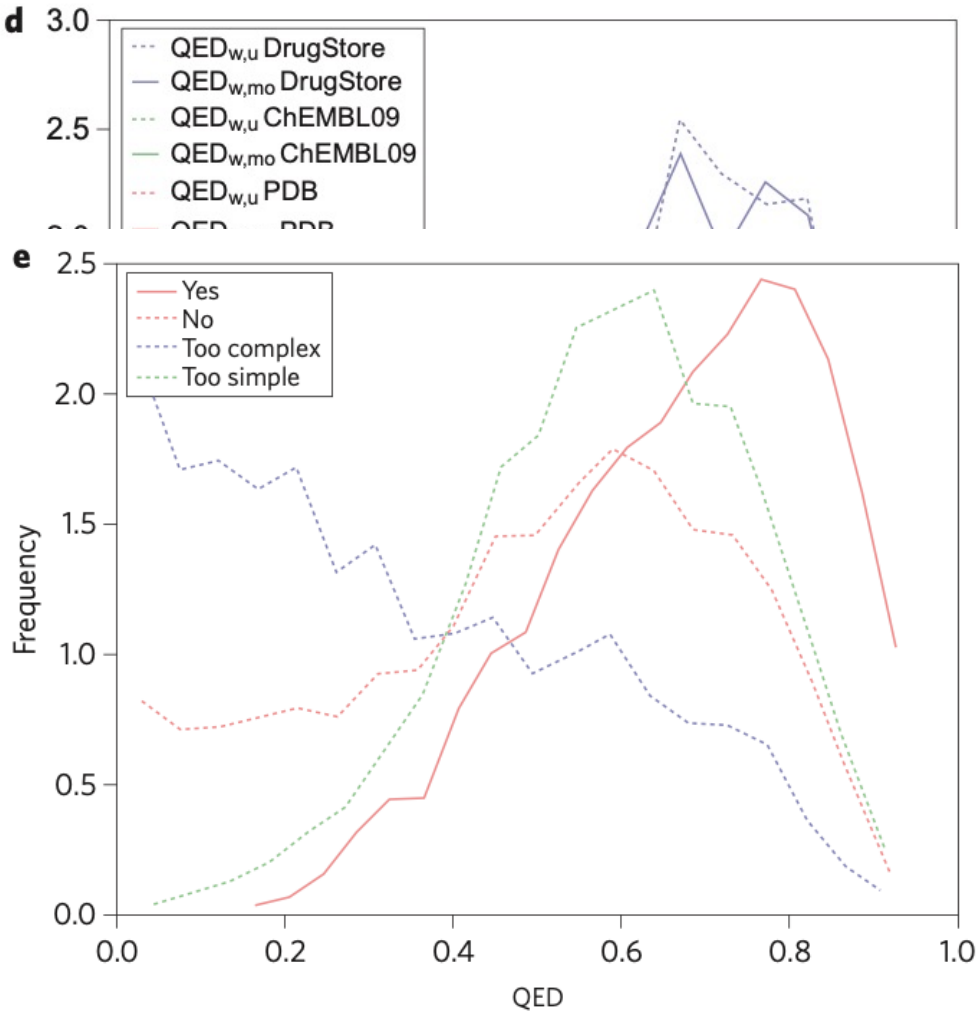
ARTICLES
PUBLISHED ONLINE: 24 JANUARY 2012 | DOI: 10.1038/NCHEM.1243
nature chemistry

Quantifying the chemical beauty of drugs

G. Richard Bickerton[1], Gaia V. Paolini[2], Jérémy Besnard[1], Sorel Muresan[3] and Andrew L. Hopkins[1]*

Drug-likeness is a key consideration when selecting compounds during the early stages of drug discovery. However, evaluation of drug-likeness in absolute terms does not reflect adequately the whole spectrum of compound quality. More worryingly, widely used rules may inadvertently foster undesirable molecular property inflation as they permit the encroachment of rule-compliant compounds towards their boundaries. We propose a measure of drug-likeness based on the concept of desirability called the quantitative estimate of drug-likeness (QED). The empirical rationale of QED reflects the underlying distribution of molecular properties. QED is intuitive, transparent, straightforward to implement in many practical settings and allows compounds to be ranked by their relative merit. We extended the utility of QED by applying it to the problem of molecular target druggability assessment by prioritizing a large set of published bioactive compounds. The measure may also capture the abstract notion of aesthetics in medicinal chemistry.

$$\mathrm{QED}_{\mathrm{w}} = \exp\left[\frac{W_{\mathrm{MW}}\ln d_{\mathrm{MW}} + W_{\mathrm{ALOGP}}\ln d_{\mathrm{ALOGP}} + W_{\mathrm{HBA}}\ln d_{\mathrm{HBA}} + W_{\mathrm{HBD}}\ln d_{\mathrm{HBD}} + W_{\mathrm{PSA}}\ln d_{\mathrm{PSA}} + W_{\mathrm{ROTB}}\ln d_{\mathrm{ROTB}} + W_{\mathrm{AROM}}\ln d_{\mathrm{AROM}} + W_{\mathrm{ALERTS}}\ln d_{\mathrm{ALERTS}}}{W_{\mathrm{MW}} + W_{\mathrm{ALOGP}} + W_{\mathrm{HBA}} + W_{\mathrm{HBD}} + W_{\mathrm{PSA}} + W_{\mathrm{ROTB}} + W_{\mathrm{AROM}} + W_{\mathrm{ALERTS}}}\right]$$



Synthetic Accessibility Score (SAscore)

Journal of Cheminformatics

ChemistryCentral

Research article

Open Access

Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions

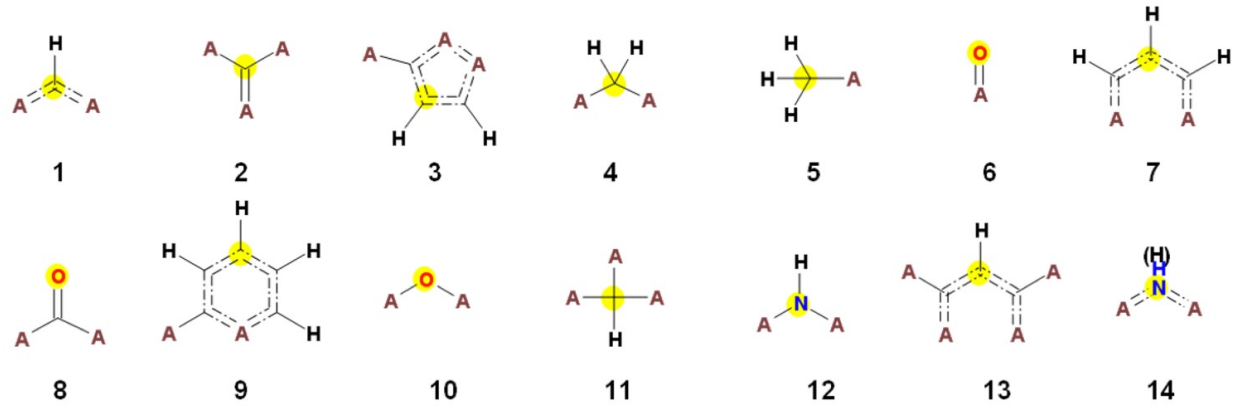
Peter Ertl* and Ansgar Schuffenhauer

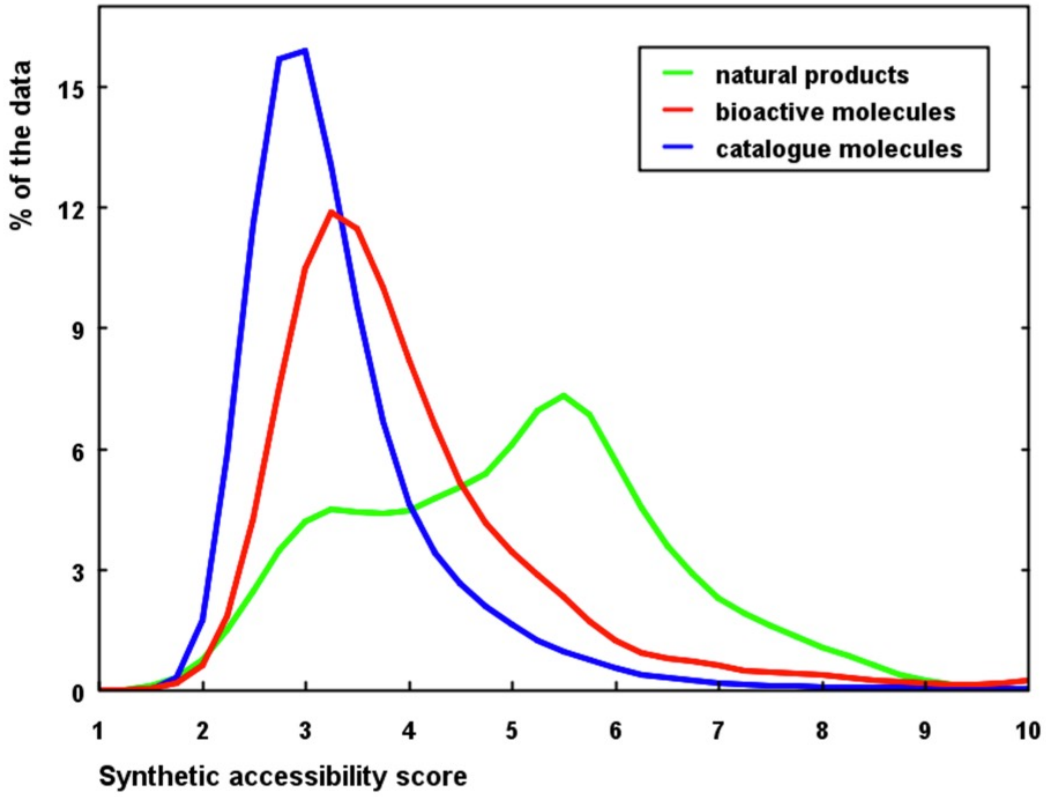
Address: Novartis Institutes for BioMedical Research, Novartis Campus, CH-4002 Basel, Switzerland

Email: Peter Ertl* - peter.ertl@novartis.com; Ansgar Schuffenhauer - ansgar.schuffenhauer@novartis.com

* Corresponding author

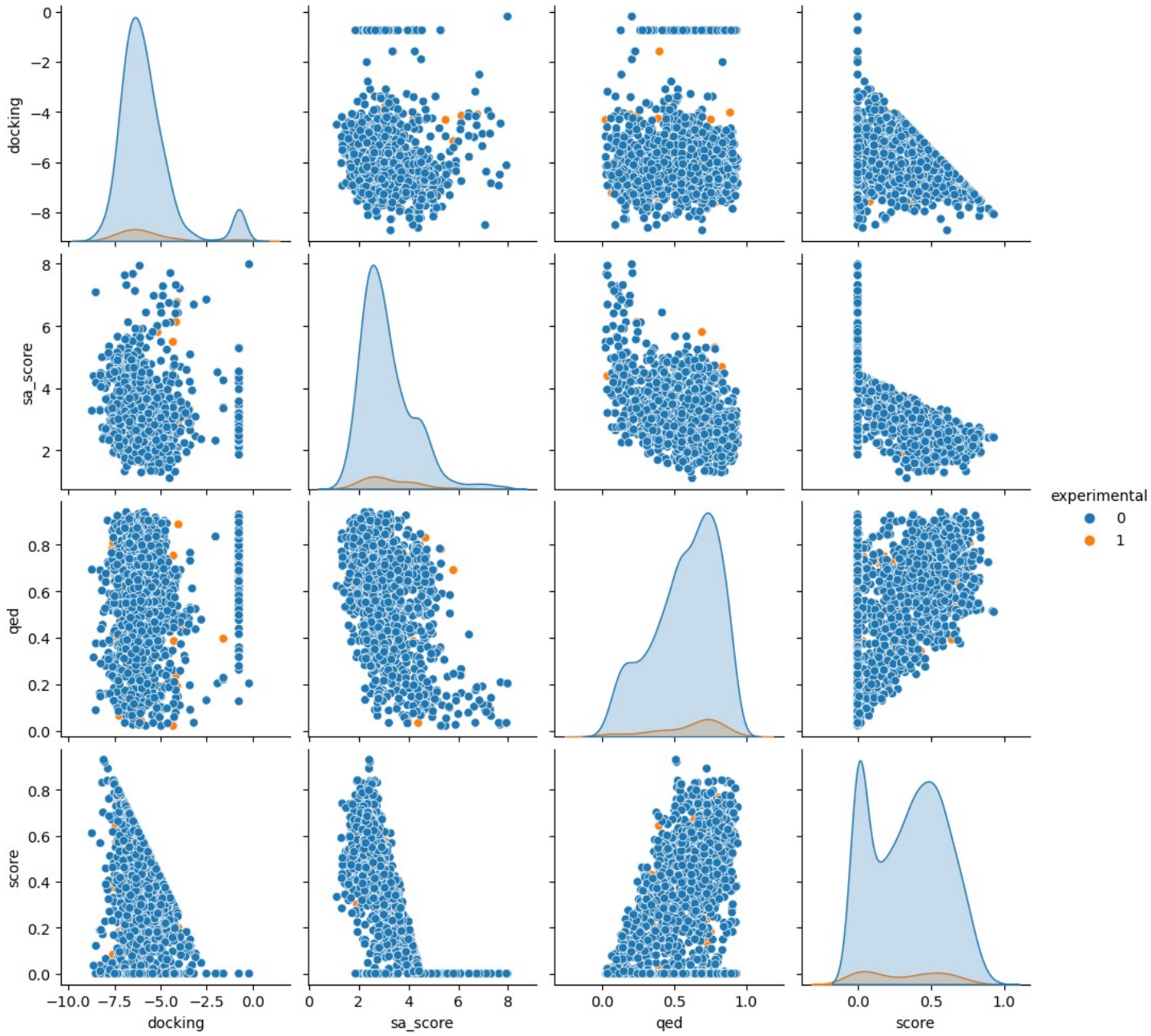
$$\text{SAscore} = \text{fragmentScore} - \text{complexityPenalty}$$





Ertl, Peter, and Ansgar Schuffenhauer. "Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions." Journal of cheminformatics 1 (2009): 1-11.

The score distribution



Team Presentation

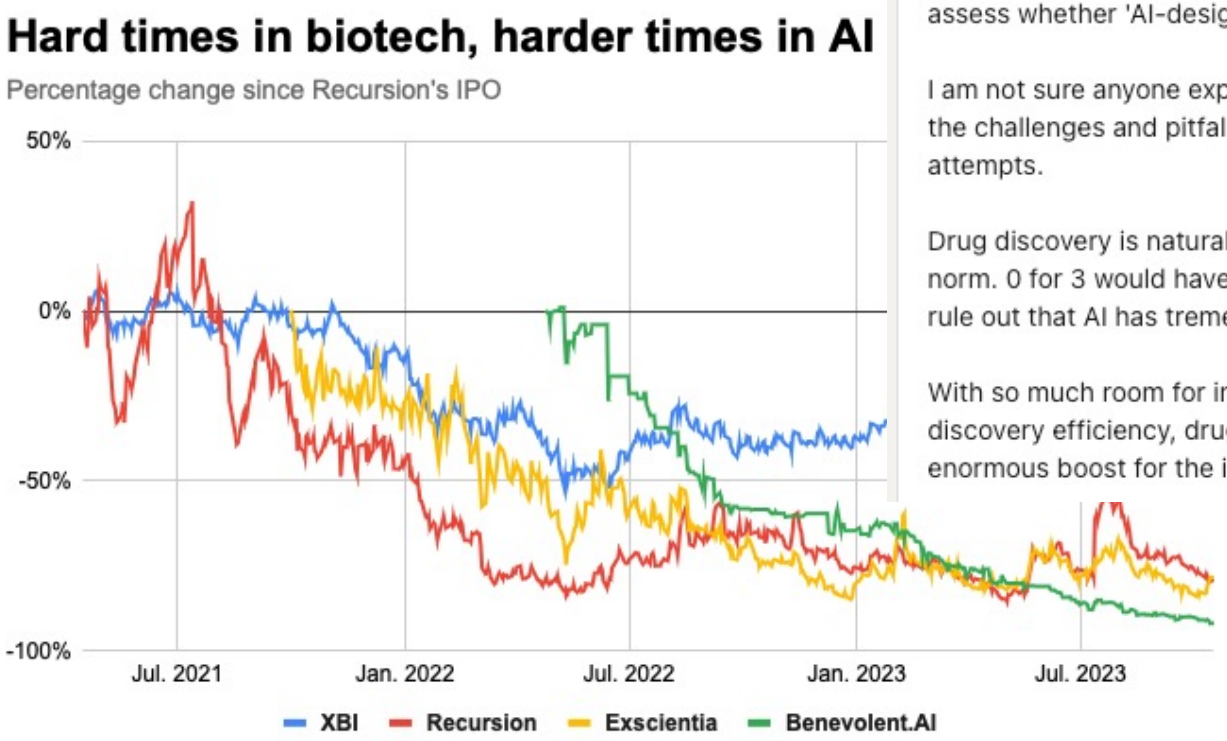
Final Remark

Hype? Valid progress?

ENDPOINTS in FOCUS

After years of hype, the first AI-designed drugs fall short in the clinic

Andrew Dunn
Biopharma Correspondent



Tom Fleming • Following
Co-Founder and COO @ Arctoris | Expertise in Drug Discovery
Visit my website
3mo • Edited •

AI-Designed Drugs...

The term 'AI-designed' drugs, to me, is misleading, since every biotech/pharma has humans in the loop, and most these days have AI in the loop too. We should cut the hype and share the credit!

We'll need to see many more successes/failures before we can even start to assess whether 'AI-designed' drugs are statistically 'better or worse'.

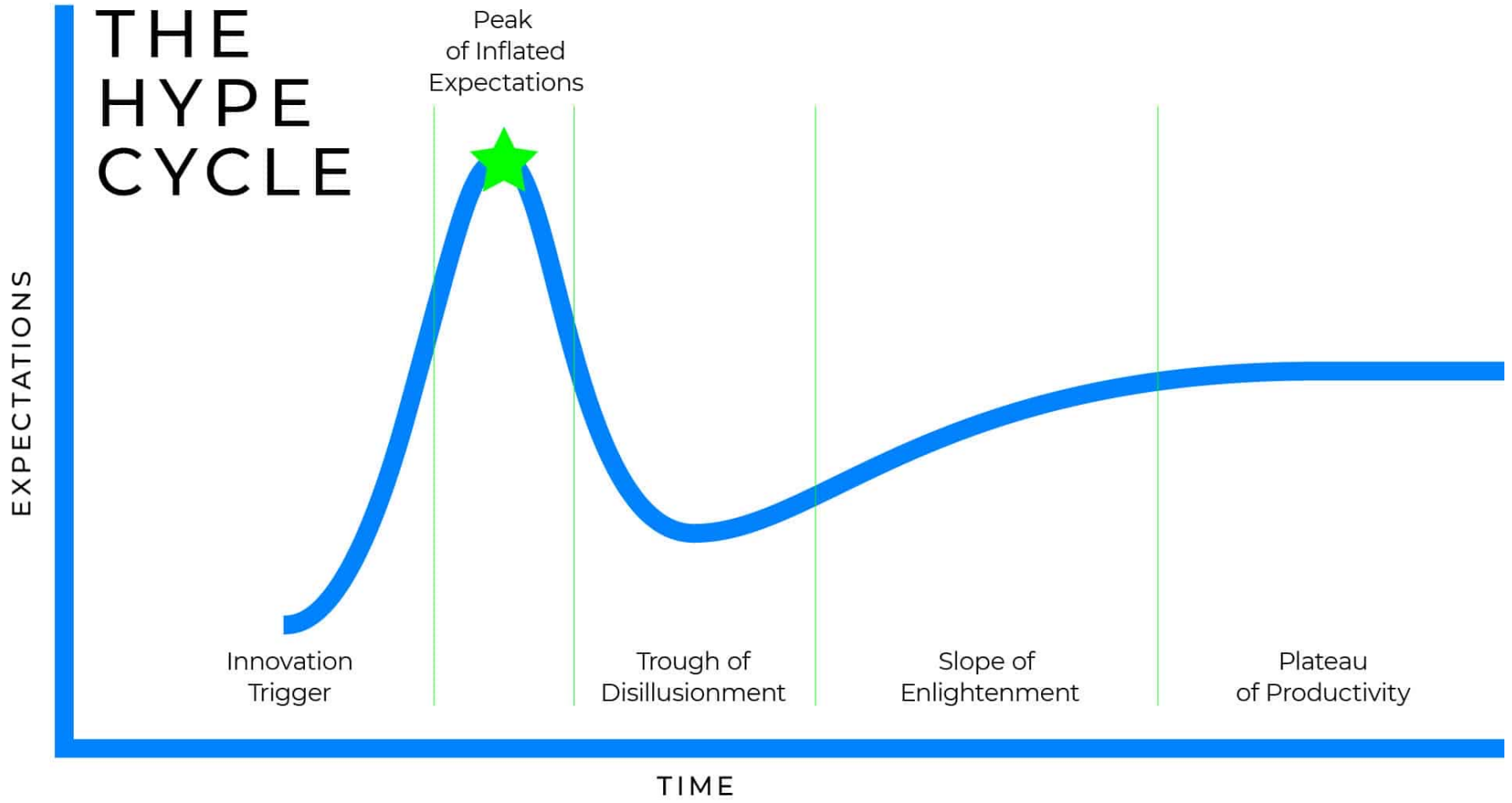
I am not sure anyone expected AI-designed drugs to perfectly overcome all of the challenges and pitfalls of drug discovery, certainly not in the first round of attempts.

Drug discovery is naturally very high attrition. In clinical trials, 90% failure is the norm. 0 for 3 would have been a very plausible prediction, but that does not rule out that AI has tremendous potential.

With so much room for improvement, even incremental gains from AI - in discovery efficiency, drug efficacy, or clinical trial success rates - will be an enormous boost for the industry, and for patients.

https://endpts.com/first-ai-designed-drugs-fall-short-in-the-clinic-following-years-of-hype/
https://www.linkedin.com/feed/update/urn:li:activity:7120808419556487169/

It takes time to reach the impact in CV/NLP



https://viraloctopus.com/magazine/strategy/gartner-hype-cycle-technology-adoption-curve/

If you want to continue in this field

- Class in spring:

Syllabus

Machine Learning for Molecular Engineering

3.C01, 10.C01, 20.C01 (Undergraduate version)

3.C51, 10.C51, 20.C51 (Graduate version)

Co-requisite course: 6.C01/6.C51

- PI: Connor Coley (ChemE+EECS), Rafa Gomez-Bombarelli (DMSE), Heather Kulik (ChemE+Chem), Regina Barzilay (EECS), Tommi Jaakkola (EECS), Bonnie Berger (Math), Tess Smidt (EECS), Alexander Rives (incoming, Broad+EECS), Sergey Ovchinnikov (Biology) ……
- Try to start research early; you could try to submit to workshops at AI conferences.
- Try to be “bilingual:” You need to be able to speak languages from both sides
- If your eventual goal is to solve chemical/biochemical/material problems, scientific knowledge in those fields is the most important

AI is changing the way we do science!

- Welcome to reach out if you have questions!
- Please fill out the feedback survey

Good Luck!