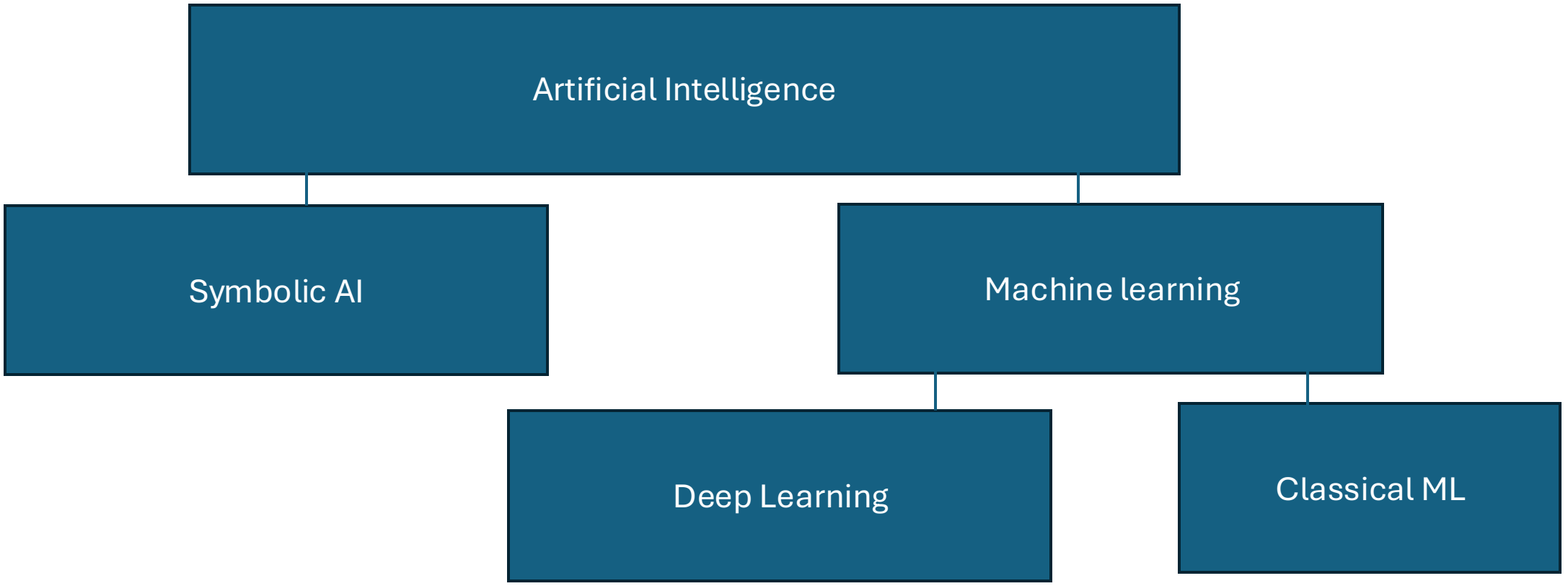
AI/ML for drug discovery

Lou Kohler Voinov

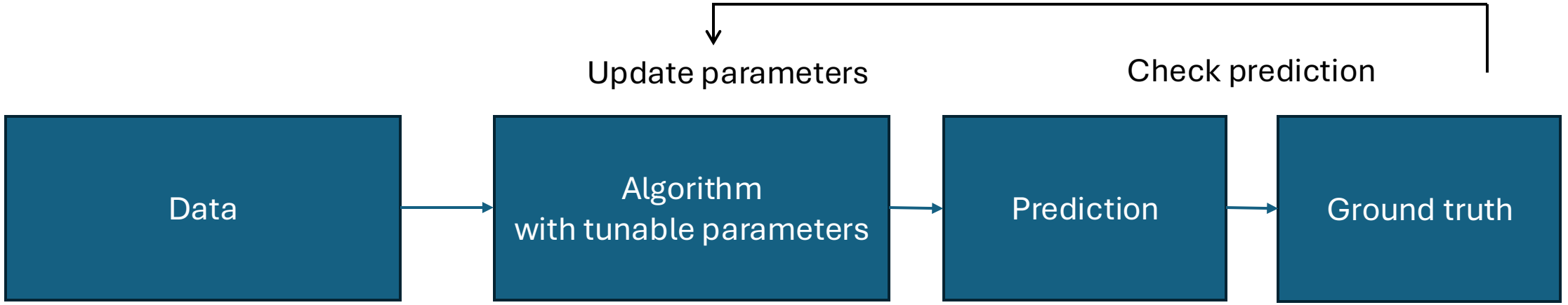
AI vs Machine learning



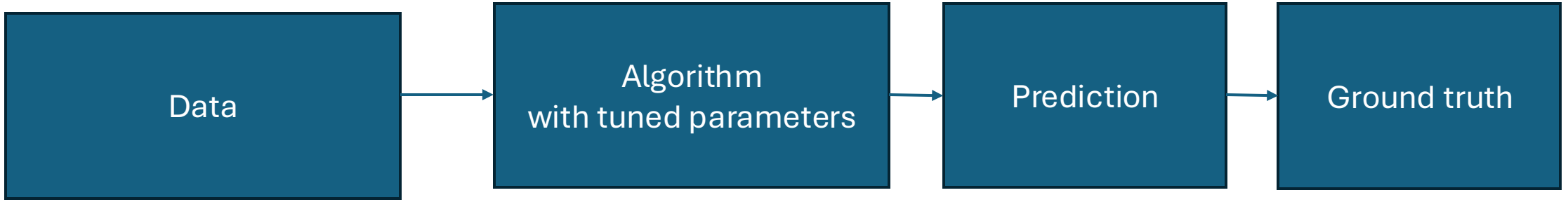
What is machine learning?

Mapping inputs to outputs

Phase 1: Training

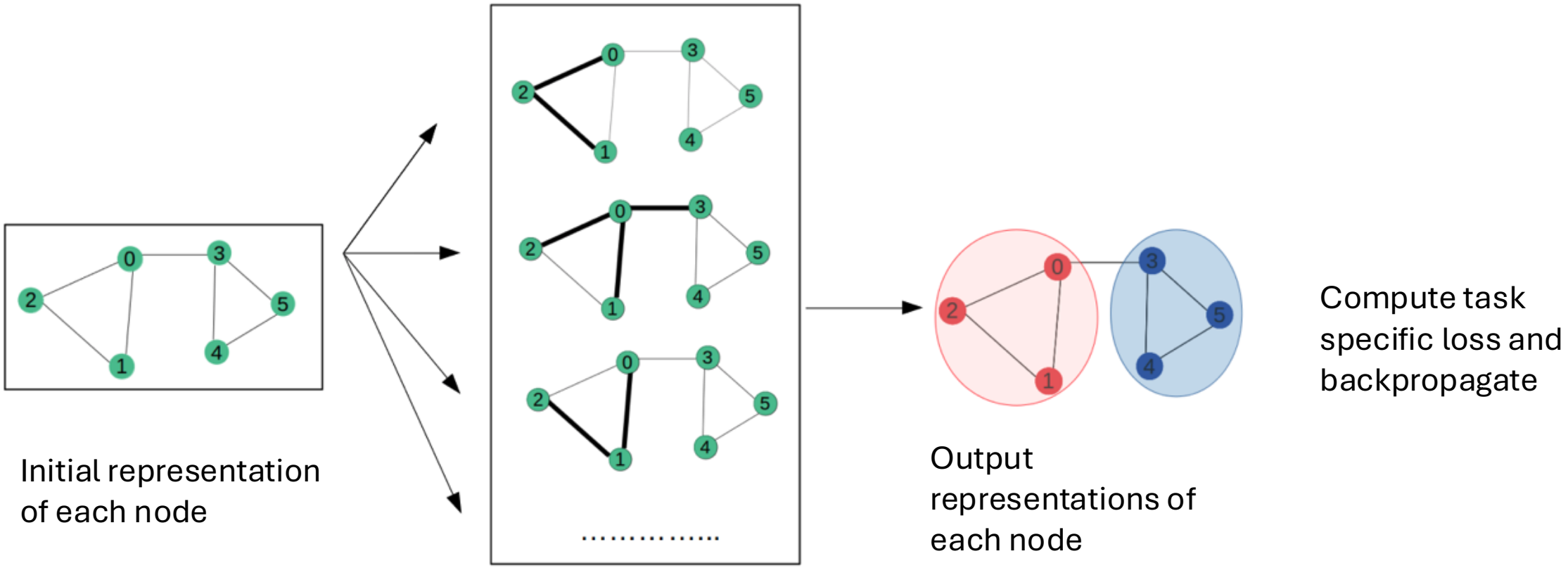


Phase 2: Testing

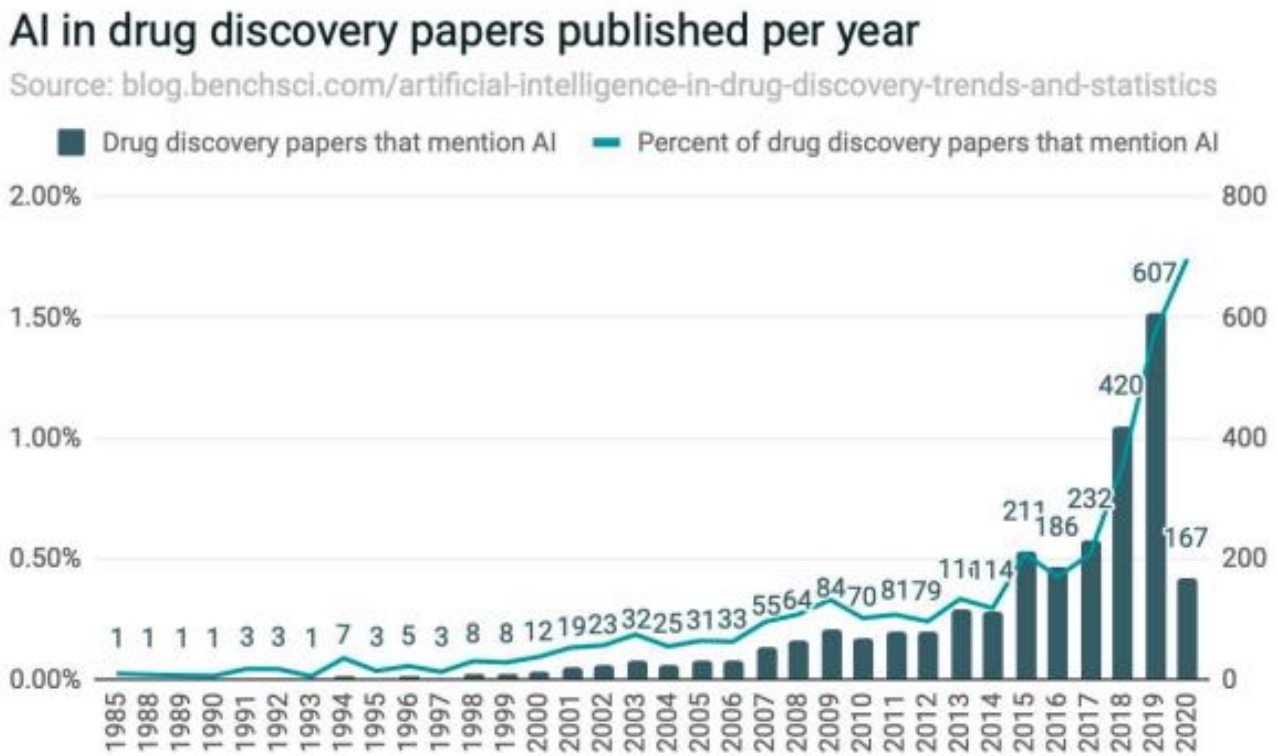


Graph neural networks

Graphs = (Nodes, Edges)



The use of AI in medicine/drug discovery



Article | Open access | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, ... Demis Hassabis

+ Show authors

Nature **596**, 583–589 (2021) | Cite this article

ProteinBERT: a universal deep-learning model of protein sequence and function

Nadav Brandes, Dan Ofer, Yam Peleg, Nadav Rappoport, Michal Linial Author Notes

Bioinformatics, Volume 38, Issue 8, March 2022, Pages 2102–2110, https://doi.org/10.1093/bioinformatics/btac020

Article | Open access | Published: 11 July 2023

De novo design of protein structure and function with RFdiffusion

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, ... David Baker

+ Show authors

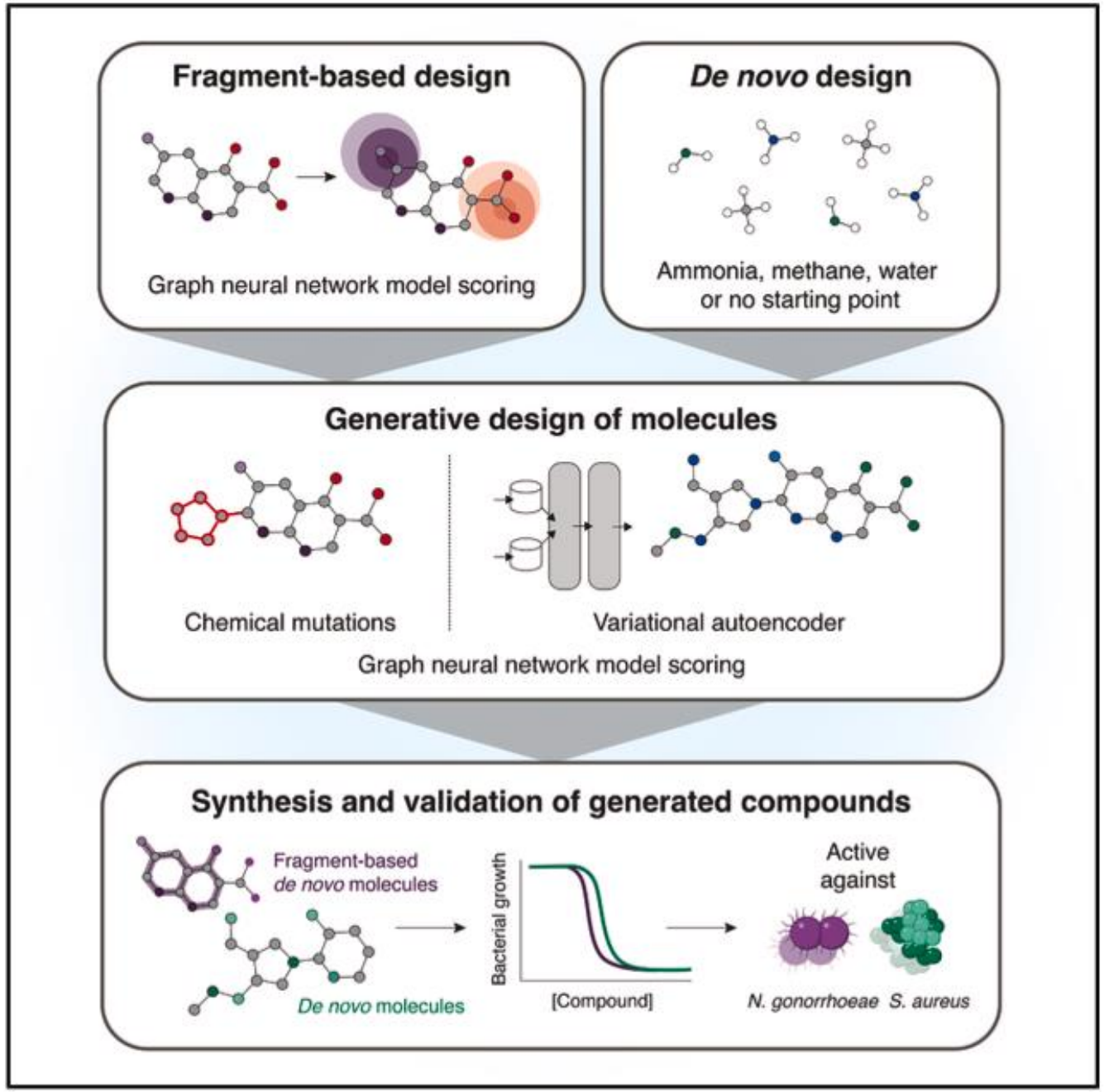
Nature **620**, 1089–1100 (2023) | Cite this article

Cell

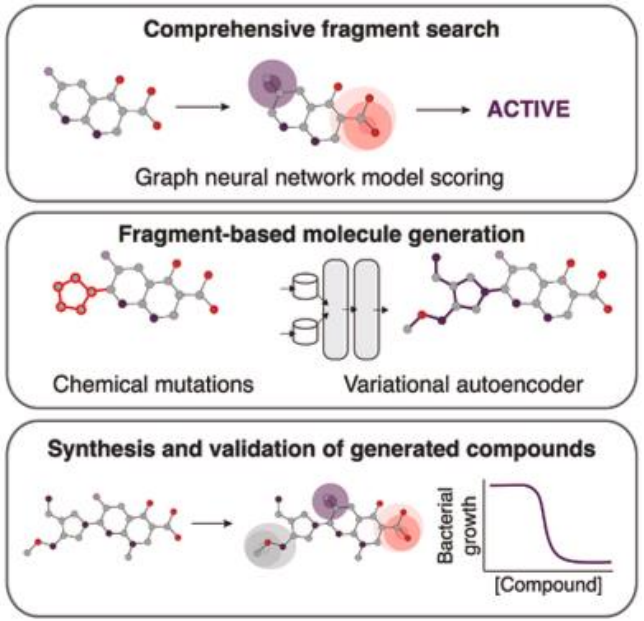
Article

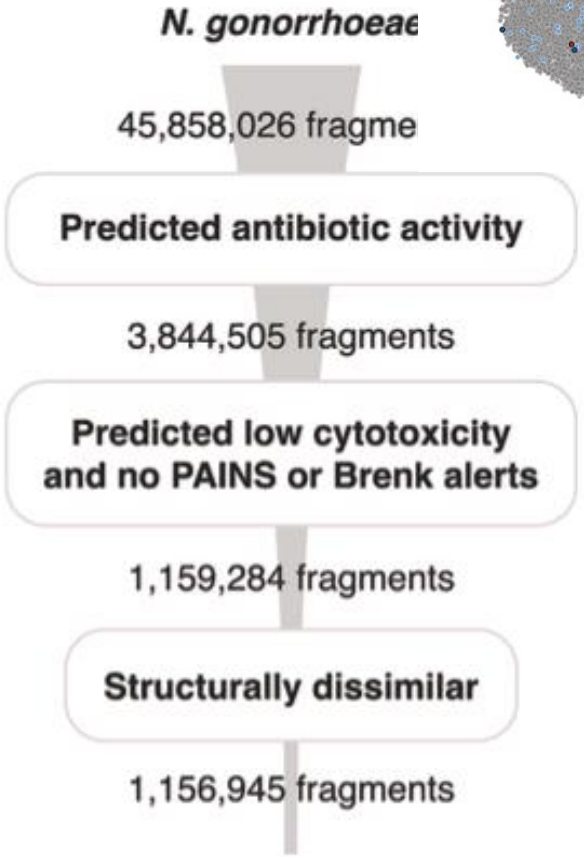
A generative deep learning approach to *de novo* antibiotic design

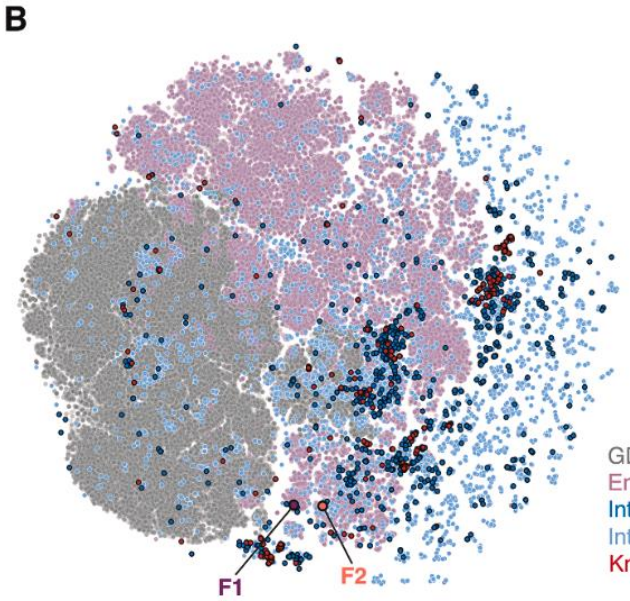
Aarti Krishnan,[1,2,3,4,25] Melis N. Anahtar,[1,2,4,5,25] Jacqueline A. Valeri,[1,2,4,25] Wengong Jin,[6]
Leif Sieben,[1,2,4,8] Andreas Luttens,[1,2] Yu Zhang,[1,2,4] Seyed Majed Modaresi,[1,2,4] Andrew H
Parijat Bandyopadhyay,[1,2] Jonathan C. Chen,[1,2] Danyal Rehman,[10] Ronak Desai,[1,11,12] Paig
Marie-Stéphanie Aschtgen,[14] Margaux Gaborieau,[14] Massimiliano Gaetani,[15,16] Samantha
Lutete Khonde,[18] Yurii S. Moroz,[19,20,21] Bruce Blough,[22] Chunyang Jin,[22] Edmund Loh,[14,2]
Amir Ata Saei,[14] Connor W. Coley,[9,24] Felix Wong,[1,2,13] and James J. Collins[1,2,4,26,*]

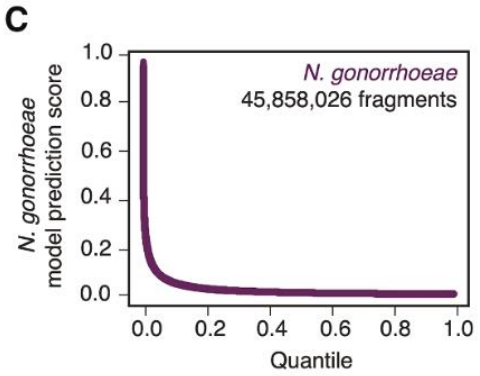


Training the GNN

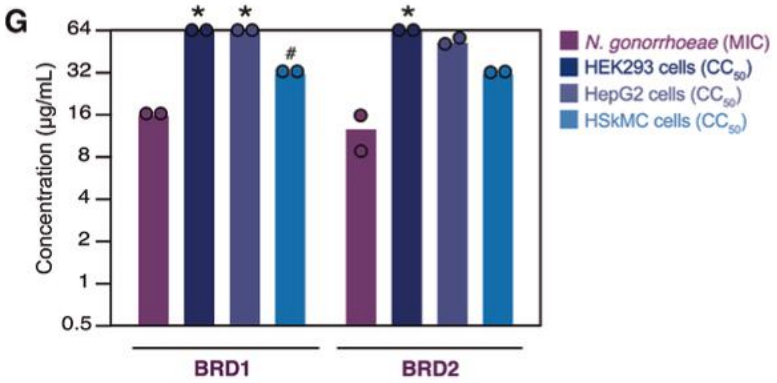




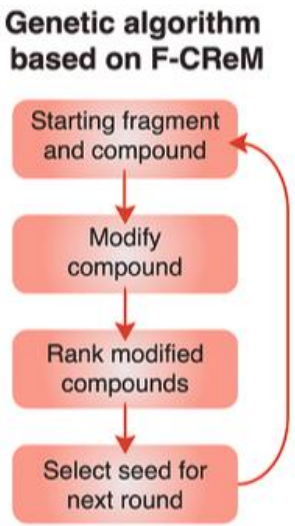


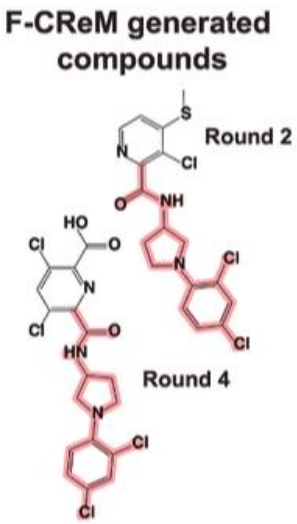


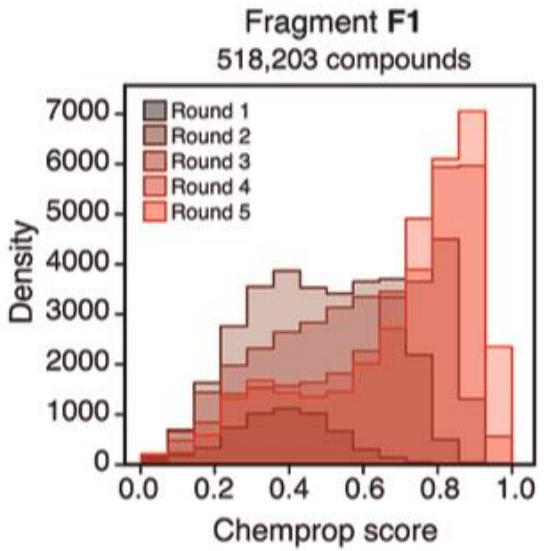
GDB-11: 100,000 fragment sample
Enamine fragments (including **F1** and **F2**)
Internal 39K actives
Internal 39K inactives
Known antibiotics (559 compounds)

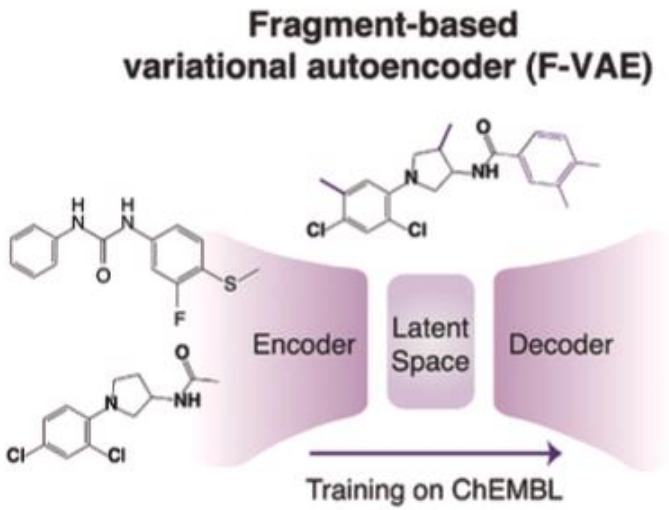


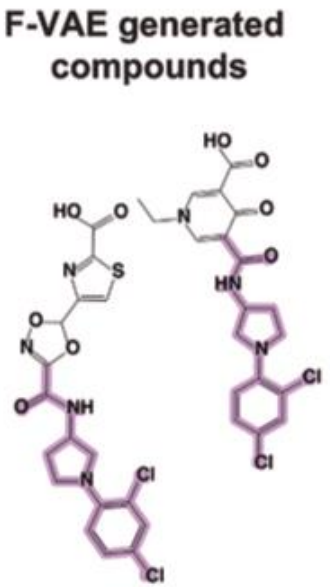
Generative ML algorithms

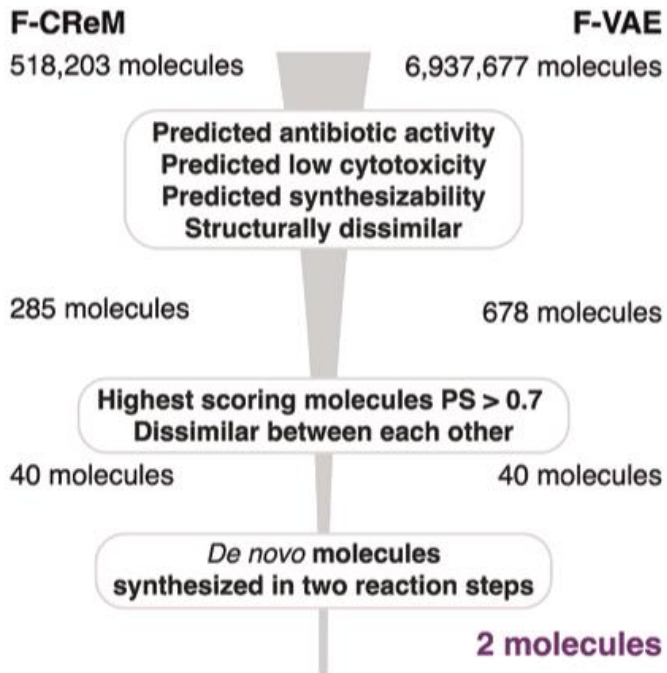




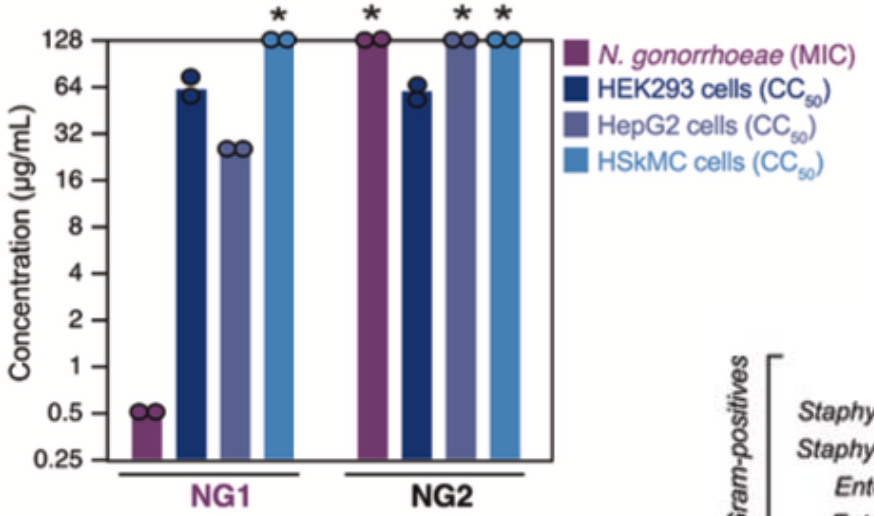


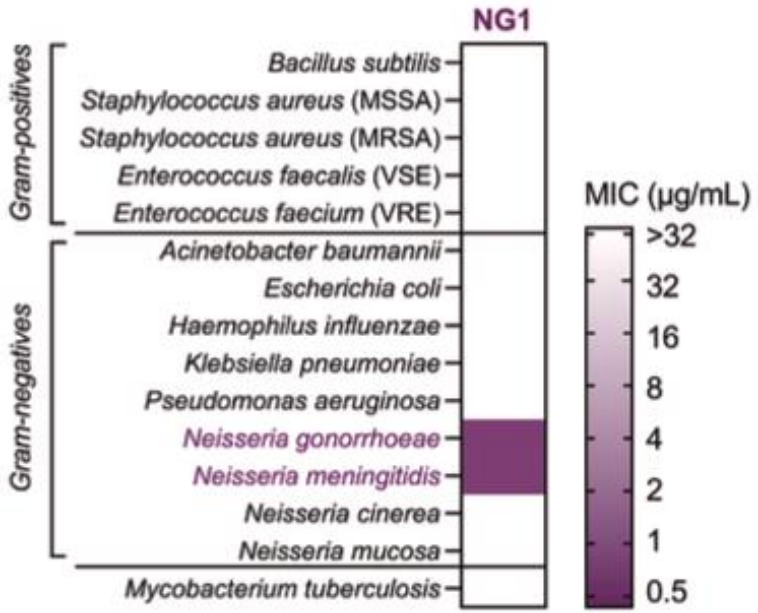






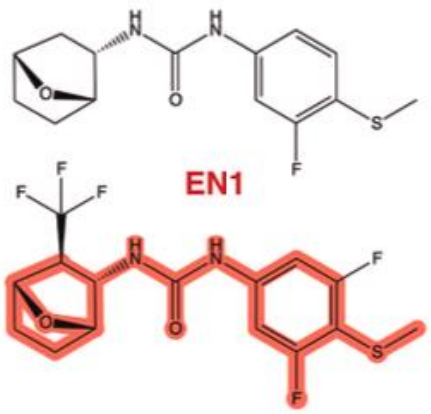
Synthesis and experimental validation of fragment- based designed compounds





Design of compounds with activity against *S. aureus*

Fragment **F2**



45,858,026 fragments

Predicted antibiotic activity

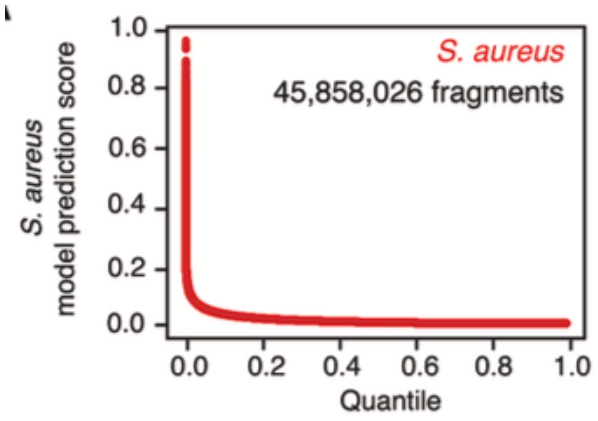
432,919 fragments

Predicted low cytotoxicity and no PAINS or Brenk alerts

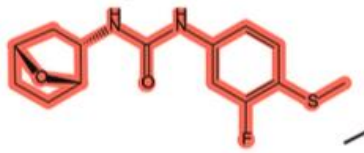
276,349 fragments

Structurally dissimilar

259,807 fragments



Fragment **F2**



Fragment **F2'**

generated molecules

F-CReM
141,109
106,557

F-VAE
3,297,292
18,164,746

Fragment **F2'**

18,271,303 molecules

Predicted antibiotic activity

Predicted low cytotoxicity and no PAINS or Brenk alerts

Synthesizability (RA score)

9,684 molecules

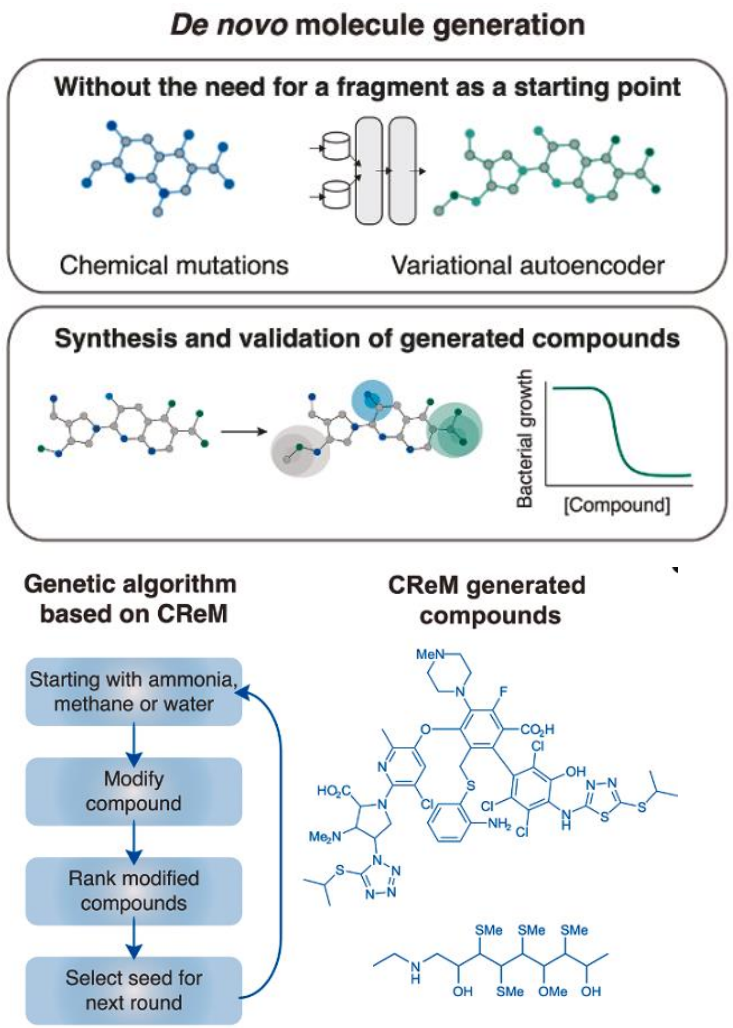
Manually inspected

128 molecules

Received and tested

16 molecules

Generation, synthesis, and experimental validation of *de novo* designed compounds



Explainable AI

nature protocols

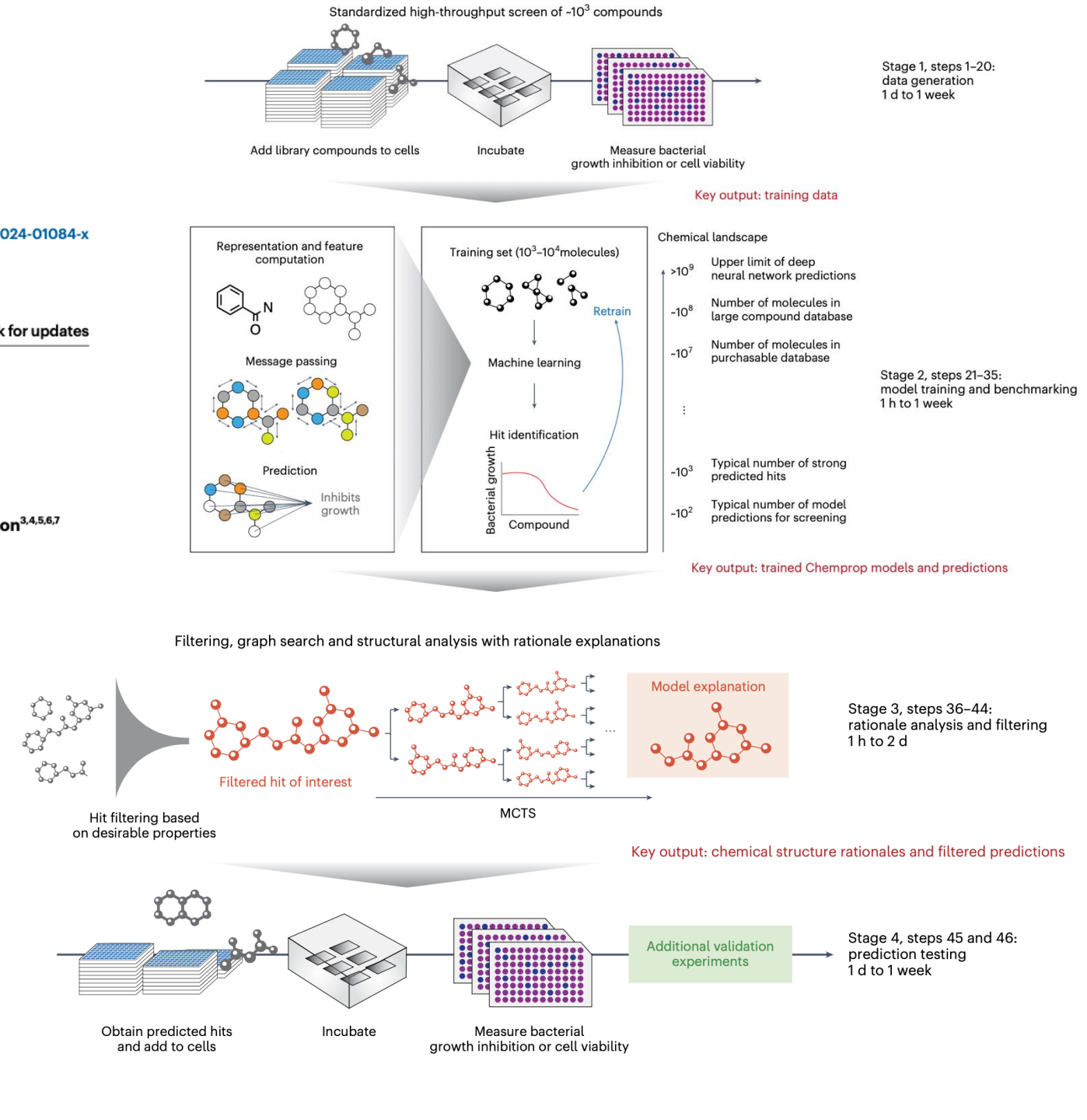
https://doi.org/10.1038/s41596-024-01084-x

Protocol

Check for updates

An explainable deep learning platform for molecular discovery

Felix Wong[1,2,3], Satotaka Omori[1,3], Alicia Li[3], Aarti Krishnan[1,2], Ryan S. Lach[3], Joseph Rufo[4,5,6,7], Maxwell Z. Wilson[3,4,5,6,7] & James J. Collins[1,2,8]



Not without concerns

Data driven Concerns
"Garbage in garbage out"

Technical & Model-Related Concerns
Generalization, overfitting, black box problem

Biological & Clinical Reality Concerns
Oversimplification of biology, AI is trained on what is known, making finding truly novel targets difficult

Practical & Operational Concerns
Experimental validation must follow, efforts must be made to integrate into traditional workflows

Ethical, Legal, and Regulatory Concerns
Regulatory scrutiny, reproducibility, biosafety & biosecurity