Hewlett Packard
Enterprise

CONVERGED DRUG DISCOVERY WORKFLOWS OF HPC, AI AND DATA IN COVID-19 RESPONSE

Rangan Sukumar, Distinguished Technologist
representing 170+ HPE volunteers

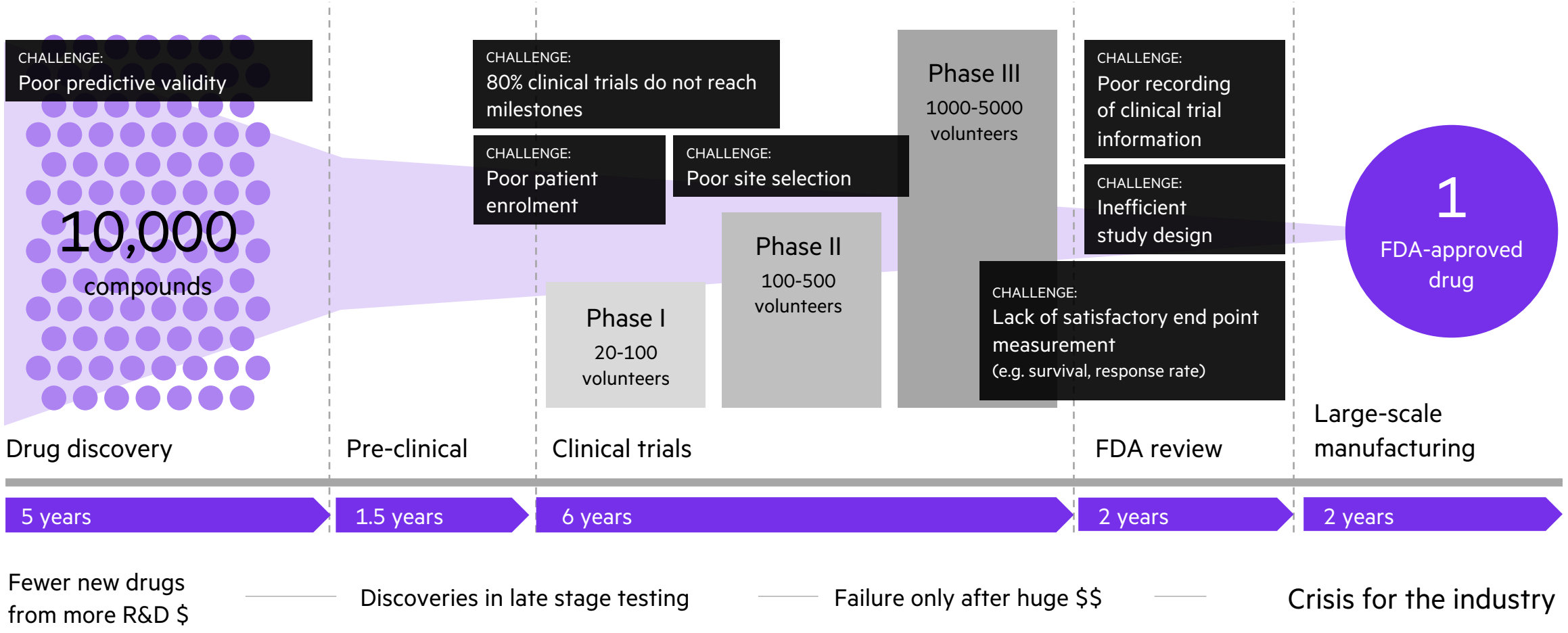
August 10, 2021

CONFIDENTIAL DISCLOSURE AGREEMENT

- The information contained in this presentation is proprietary to Hewlett Packard Enterprise and is offered in confidence, subject to the terms and conditions of a binding Confidential Disclosure Agreement (CDA)
- HPE requires customers and partners to have signed a CDA in order to view this training
- The information contained in this training is HPE confidential
- This presentation is NOT to be used as a 'leave behind' for customers and information may only be shared verbally with HPE external customers under NDA
- This presentation may be shared with Partners under NDA in hard-copy or electronic format for internal training purposes only
- Do not remove any classification labels, warnings or disclaimers on any slide or modify this presentation to change the classification level
- Do not remove this slide from the presentation
- HPE does not warrant or represent that it will introduce any product to which the information relates
- The information contained herein is subject to change without notice
- HPE makes no warranties regarding the accuracy of this information
- The only warranties for HPE products and services are set forth in the express warranty statements accompanying such products and services
- Nothing herein should be construed as constituting an additional warranty
- HPE shall not be liable for technical or editorial errors or omissions contained herein
- Strict adherence to the HPE Standards of Business Conduct regarding this classification level is critical

GETTING FROM TODAY...

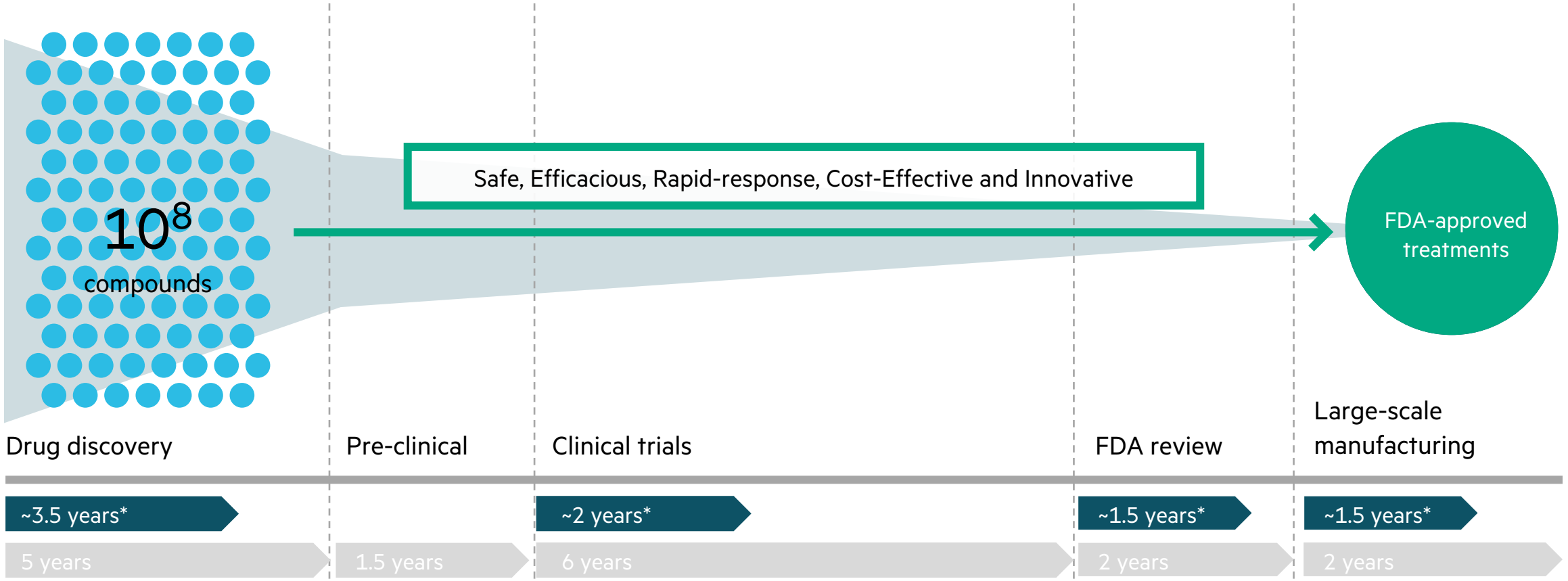
Predominantly an *in vitro* and *in vivo* environment



Source: PhRMA Profile

...TO TOMORROW

An *in silico* HPC, AI and Data augmented environment



* Time reductions are estimates
Source: PhRMA Profile

EARLY SIGNS OF AI CHANGING THE LANDSCAPE

MedTech

Breaking Big Pharma's AI barrier: Insilico Medicine uncovers novel target, new drug for pulmonary fibrosis in 18 months

by Conor Hale | Feb 24, 2021 9:00am

BLOG POST
RESEARCH

30 NOV 2020

AlphaFold: a solution to a 50-year-old grand challenge in biology

JANUARY 29, 2020

Sumitomo Dainippon Pharma and Exscientia Joint Development New Drug Candidate Created Using Artificial Intelligence (AI) Begins Clinical Trial

Artificial intelligence-created medicine to be used on humans for first time

By Jane Wakefield
Technology reporter

30 January 2020

SPOTLIGHT · 30 MAY 2018

How artificial intelligence is changing drug discovery

Machine learning and other technologies are expected to make the hunt for new pharmaceuticals quicker, cheaper and more effective.

Nic Fleming

BACKGROUND: THE GRAND CHALLENGE OF DRUG DISCOVERY

Drug Discovery Today

- Expensive
  - > $1 Billion per drug
- Time-consuming
  - > 10 years per drug
- Consequently, research is market/competition driven
  - Is market size > $1B?
  - Is the patent expiring?
- Suffers from the lack of
  - Knowledge of underlying disease-causing mechanism
  - Personalization and precision to cater a diverse population
  - Translation/Reproducibility of pre-clinical, animal models, etc.
  - Confidence in biomarkers and surrogate endpoints
  - Quality in published assay/experimental data
  - Agility required for rapid response in a pandemic

Let us reimagine “drug discovery”

• Opportunity for a better future

- >6000 rare diseases do not have a cure [1]
  - Affects >25 million Americans
- ~75% of emerging infectious diseases are zoonotic. [2]
  - >2.5 billion affected, ~2.7 million deaths per year
- > 87 new infectious diseases since 1980 [3]
- > 15 new zoonotic disease outbreaks in 15 years [4]
  - Ebola, H5N1, H7N9, SARS, H1N1, COVID-19, etc.
- Pandemics between 1997-2009 cost US$80 billion [5]
  - ~$7 billion per year
- > $8 trillion: Global cost of the COVID-19 pandemic [6]

1. https://www.genome.gov/FAQ/Rare-Diseases
2. Taylor, Louise H., Sophia M. Latham, and Mark EJ Woolhouse. "Risk factors for human disease emergence." Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 356.1411 (2001): 983-989.
3. Gebreyes, Wondwossen A., et al. "The global one health paradigm: challenges and opportunities for tackling infectious diseases at the human, animal, and environment interface in low-resource settings." PLoS Neglected Tropical Diseases 8.11 (2014): e3257.
4. https://www.who.int/emergencies/diseases/en/
5. World Bank (2012) People, Pathogens and Our Planet. Volume 2, The Economics of One Health. World Bank Report #69145-GLB
6. https://www.bbc.com/news/business-52671992

DRUG DISCOVERY IS A WORKFLOW OF "SEARCH" PROBLEMS

Chemical space	Biological Relevance	Synthesizable	Efficacious	Safe
~10^{60}	~10^{7}	~10^{4}	~3000	~800
compounds	molecules	small molecules	potential drugs	approved drugs
https://en.wikipedia.org/wiki/Chemical_space	https://zinc.docking.org/	https://pubchem.ncbi.nlm.nih.gov/	https://go.drugbank.com/drugs	https://www.accessdata.fda.gov/

• Drug discovery workflows search for…

- relevant compounds that modulate a biological system to influence disease development and progression from a chemical space of ~10^{60} potential synthesizable compounds
- similarity across >10^{6} open-science proteins to understand the characteristics of the disease-causing organism
- the appropriate pose and active site interaction between a disease/symptom causing protein against ~10^{7} drug compounds
- evidence from over 10^{6} research documents to explain the mechanism-of-action
- knowledge nuggets from assay and experimental data (over 10^{8} medical facts) to list protein-protein and protein-molecule interactions
- safety and efficacy indicators in siloed public-health data sources with over 10^{8} past patient records and trial data

THE "SEARCH" IS MULTI-MODAL, INTERACTIVE AND ITERATIVE

Data

CryoEM microscopes
- 3D structure of disease targets from multiple slices

Chemical/Molecular Structures
- SMILES sequences
- 2D and 3D structures

Scientific Documents
- Millions of publications, Lab notes
- Results from assays/ instruments

Genome Sequences
- Sequence of organisms and humans

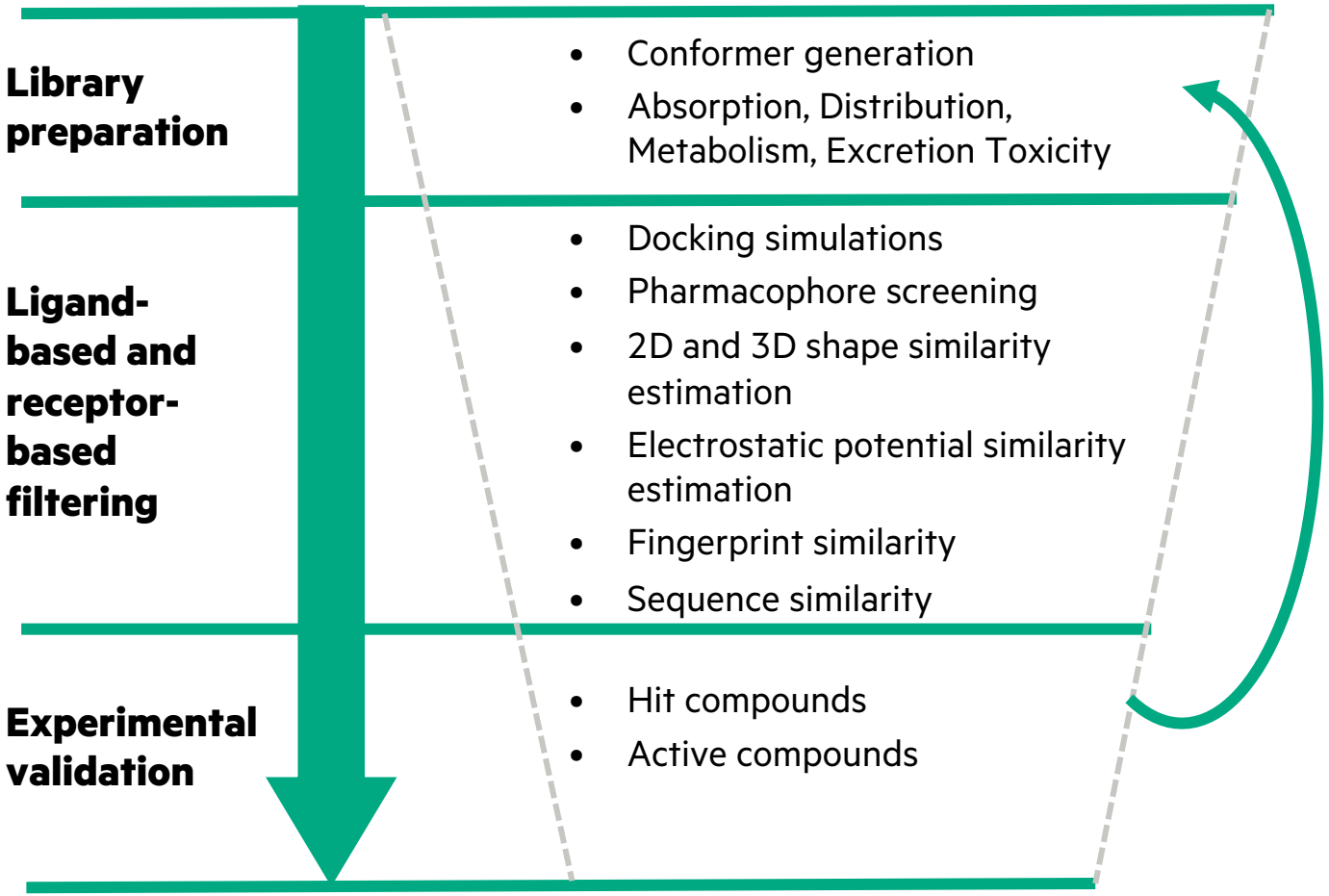
Animal Studies and Clinical Trails
- Longitudinal history – discrete time series
- Cohort studies

Fitness Trackers
- Continuous time-series from sensors

Databases
- Cohorts, Proteins, Medical terms, etc.

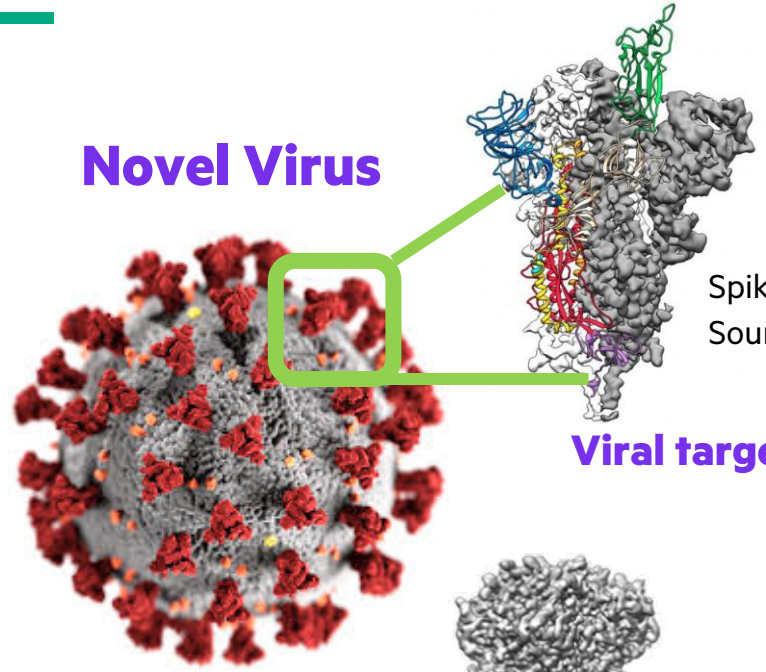
Knowledge Graphs and Ontologies
- Protein-protein interactions, protein-chemical interactions, Drug side effects, etc.

Iterative Search Process



USE CASE: RAPID-RESPONSE TO COVID-19 PANDEMIC

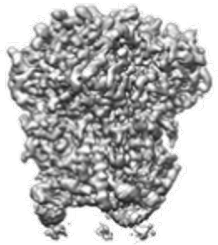
Novel Virus



Source: Wikipedia

Spike protein
Source: LiveScience

Viral target protein

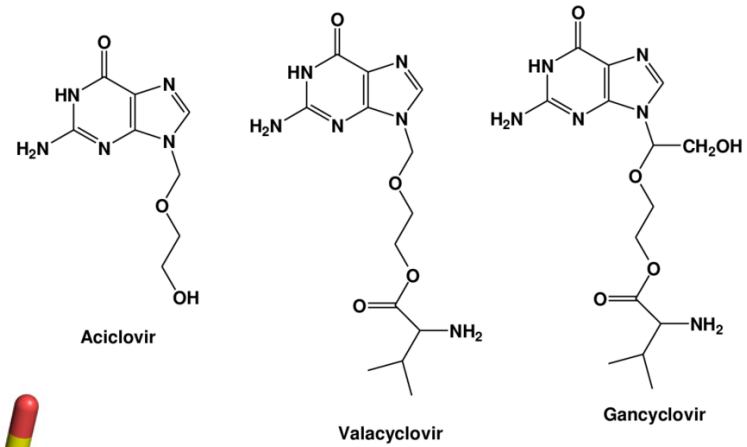


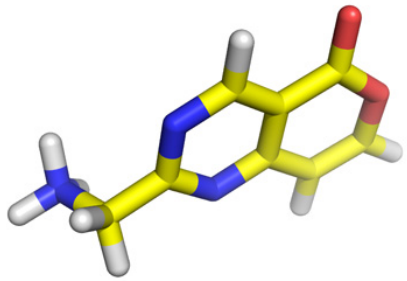
3D Structure from
cryoEM microscopes

```
MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGT
KRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWM
ESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIG
INITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVE
KGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKL
NDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPF
ERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNF
NFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCT
EVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYT
MSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQD
KNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKF
NGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIG
KIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQ
QLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGK
AHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD
VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSC
LKGCCSCGSCCKFDEDDSEPVLKGVKLHYT
```

Source: NCBI; FASTA
Sequence of Spike

Existing Drug Databases





Source: Homology
structure of molecule

Virtual Drug Databases

```
C1[C@@H](CC[C@H](C1)C(=O)[O-])C[NH3+]
C1(C(=O)NC(=O)N1)NC(=O)N
c1ccc2c(c1)C=CC(=O)C2=O
C1C[NH+]2CC[C@H]([C@@H]2[C@H]1CO)O
CC(CC1CCCC[NH+]1C)O
```

USE-CASE: PRIORITIZE COMPOUNDS FOR PRE-CLINICAL TESTING

- **Problem Statement:** A Pharma/academic research lab has a 200,000+ molecule database

- **In-silico Challenge:** Use simulations, Big Data and AI to narrow down the 200,000+ database to prioritize ~10 compounds that can be tested in a wet lab

STATE OF THE PRACTICE: *IN SILICO* DRUG DISCOVERY WORKFLOWS

Different in-silico workflows for drug discovery

Disease Target Structure Estimation/Modeling (e.g. using RELION on CryoEM image data)

Particle picking → 2D classification → 3D classification → Refinement → Sharpening → 3D Reconstruction

Ligand Database Preparation

Bibliographic research (e.g. Uniprot, Brenda) → Activity and structural data collection (e.g. ChEMBL, BindingDB, PubChem) → Conformer generation

Ligand-Based Screening

- Ligand-based pharmacophores
- Fingerprint-based methods
- 3D Shape similarity
- Electrostatic potential similarity

→ Generate conformations

Extract ligand features

Superimpose ligands

Generate pharmacophores

Structure-Based Screening

- Protein-Ligand Docking
- Structure-based pharmacophores

→ Protein preparation

Binding-site definition

Conformational sampling

Scoring

Computational Validation

Classification accuracy studies on actives (e.g. FDA), inactives (e.g. PubChem) and decoys (e.g. DUD-E)

Hit Selection and Absorption, Distribution, Metabolism, Excretion Toxicity Filtering

STATE OF THE PRACTICE: DATA, TOOLS AND SOFTWARE

Purpose	Tool
Drug ligand database	ZincDatabase, Zinc15Database, ChEMBL, Bingo, JChemforExcel, ChemDiff, ProteinDataBank(PDB), BindingMOAD(MotherOfAllDatabase), PDBbind, TTD, STITCH, SMPDB
Chemical structure representation	ChemDraw, MarvinSketch, ACD/ChemSketch, jsMolEditor, Marvinmoleculeeditorandviewer, Ketcher, UCSFChimera, Pymol, OpenStructure, Daylight SMILES, InChI, TriposMol2, OpenBabel, Corina, Indigo, PoseView, PLiP, Ligplot+, E-Babel, Corinaonlinedemo, ChemicalIdentifierResolver, COSMOS, VEGAWE, PDBHydrogenAddition, DG-AMMOS, ChemMobi, ChemSpotlight
Molecular modeling	CHARMM, GROMACS, Amber, SwissParam, CHARMM-GUI, CHARMMing.org
Homology modeling	Modeller, I-TASSER, LOMETS, SWISS-MODEL, SWISS-MODELRepository, Robetta
Binding site prediction	MED-SuMo, TRAPP, CAVER, sc-PDB, CASTp, Pocketome, 3DLigandSite, metaPocket, PockDrug
Docking	Autodock, DOCK, GOLD, SwissDock, DockingServer, 1-ClickDocking
Screening	Pharmer, Catalyst, PharmaGist, SwissSimilarity, Blaster, AnchorQuery
Target prediction	PatchSearch, IXCHEL, CABRAKAN, SwissTargetPrediction, SEA, CSNAP
Ligand design	GANDI, LUDI, BREED, SwissBioisostere, VAMMPIRE, sc-PDB-Frag, e-LEA3D, eDesign, iScreen
Binding free energy estimation	Hyde, X-score, NNScore, DSX^{ONLINE}, BAPPLserver, BAPPL-Zserver, CLiBE
Quantitative structure-activity relationship	cQSAR, SeeSAR, clogP, MOLEdb, ChemDB/Datasets, DatasetsfromtheMilanoChemometricsandQSARResearchGroup, OCHEM, E-Dragon, PatternMatchCounter
Absorption, Distribution, Metabolism, Excretion Toxicity	QikProp, VolSurf, GastroPlus, ALOGPS, OSIRISPropertyExplorer, SwissADME, PACT-F, TOXNET, LeadscopeToxicityDatabase

Source: http://click2drug.org/

THE “FORCE-FOR-GOOD” PLEDGE

HPE opens its patents to fight COVID-19

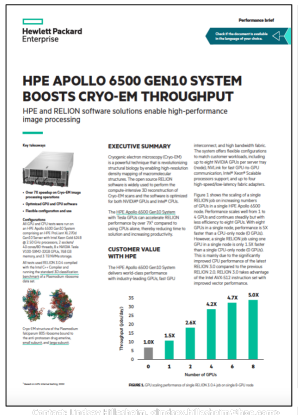
APRIL 20, 2020 • BLOG POST • BRETT ALTEN, CHIEF INTELLECTUAL PROPERTY COUNSEL, HEWLETT PACKARD ENTERPRISE

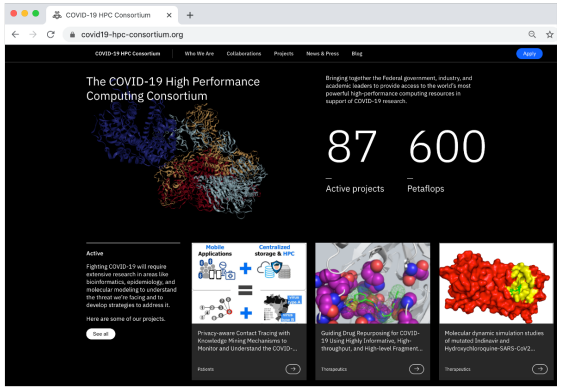
HPE Is Using Its $1.3 Billion Cray Acquisition To Support COVID-19 Research

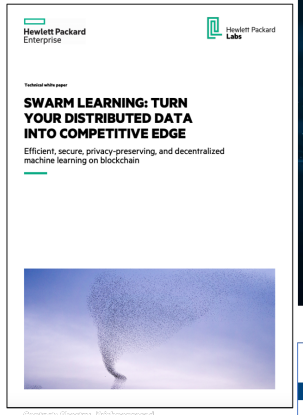
Hewlett Packard Enterprise is arming COVID-19 researchers with the high-performance computing and artificial intelligence capabilities necessary to make scientific breakthroughs on new treatments and vaccines — two fields that are "more critical than ever," according to the vendor's top HPC executive.

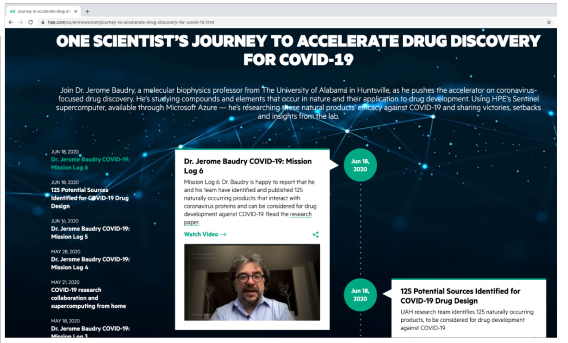
MANY SCIENTIFIC WORKFLOWS IMPLEMENTED…

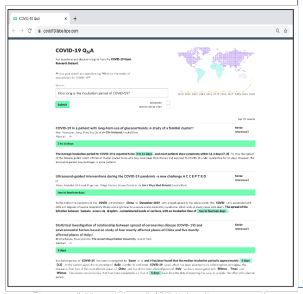
- **Converged simulation and AI workflows**
  - Expertise to simulate molecular dynamics at unprecedented scale (Newtonian to quantum physics fidelity and resolution)
  - Scale-out protein-ligand docking experiments
- **Artificial Intelligence for protein-ligand binding prediction**
  - Open-source AI model (PharML) for virtual screening of novel drug molecule databases
- **Natural Language Processing for rapid literature search**
  - Demonstration of a Q&A platform on the CORD-19 corpus released by the White House
- **Knowledge Graphs for rapid drug repurposing**
  - Integration of 13 popular life science databases (152 billion facts, >30 terabytes on disk)
- **Swarm Learning for disease classification**
  - Enabling federated privacy-preserving machine learning
- **“Drug discovery workflow” acceleration using reference architectures**
  - Accelerating 3D reconstruction of viral protein targets from CryoEM microscopes
  - Architecting data systems to query different shapes of data (sequences, graphs, time-series, documents,2D/3D images).





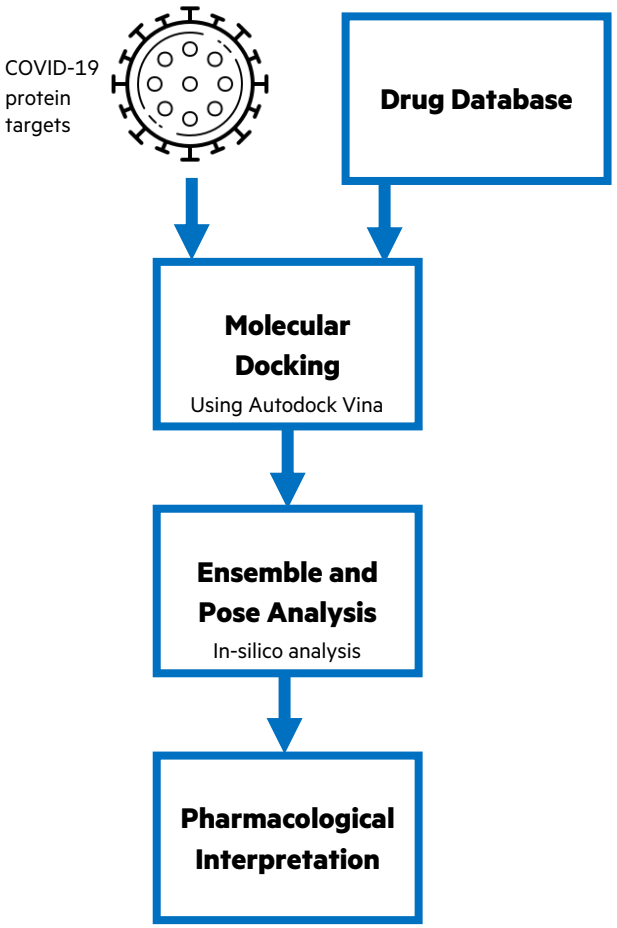






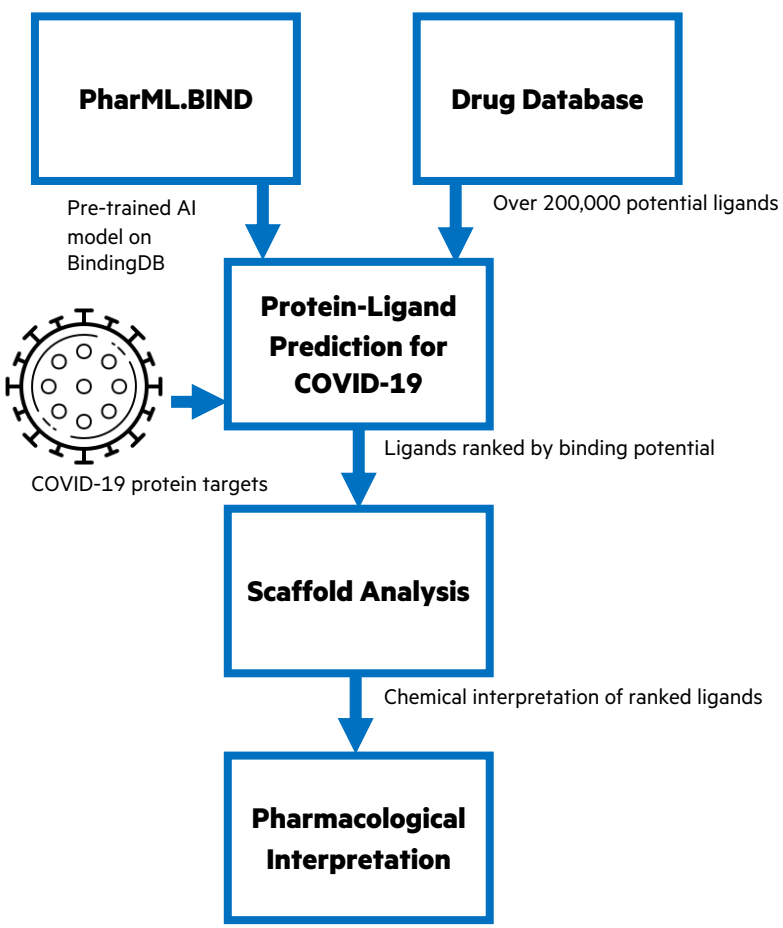
“SEARCHING” FOR THE NEEDLE IN A HAYSTACK: THREE APPROACHES

Simulation



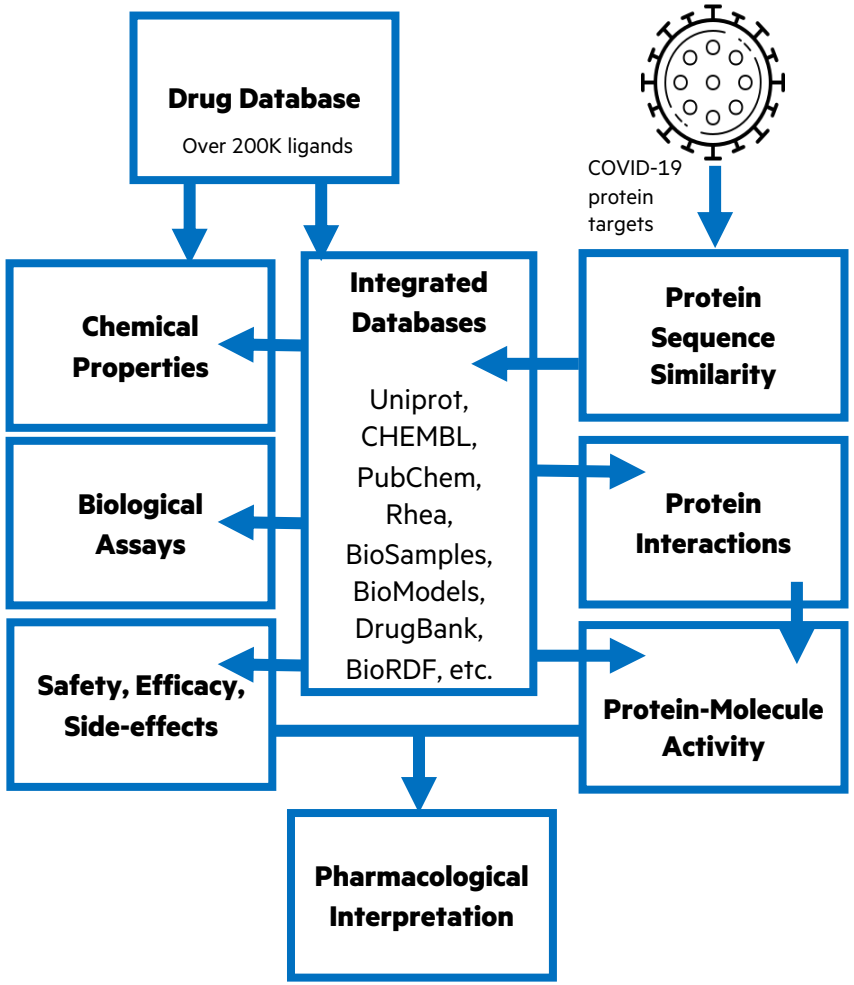
Validation by study of therapeutic implications and mechanisms of action

Artificial Intelligence



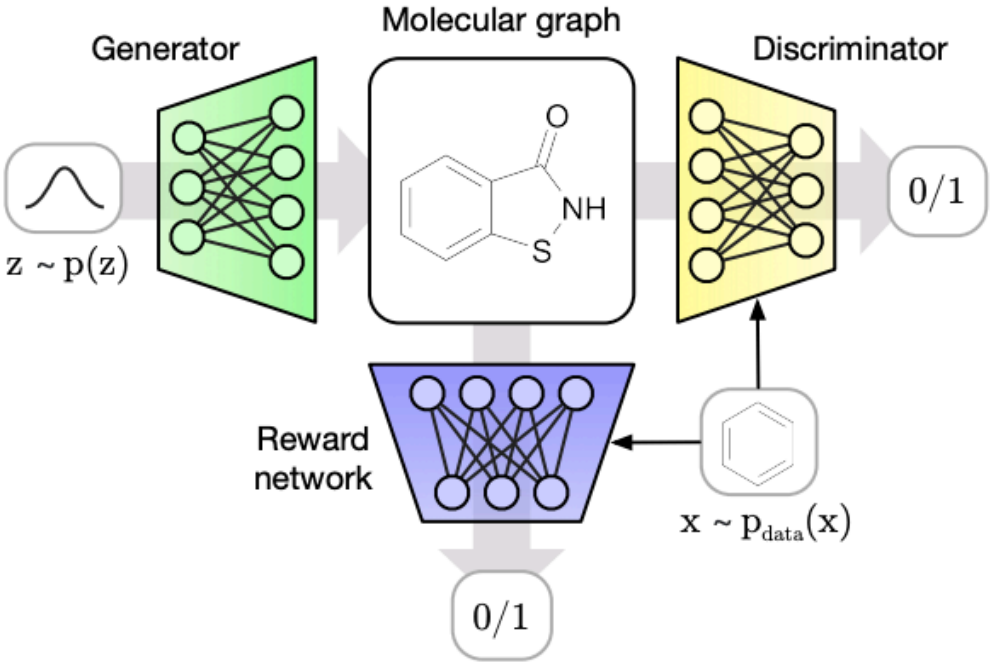
Validation by study of therapeutic implications and mechanisms of action

Knowledge Graphs



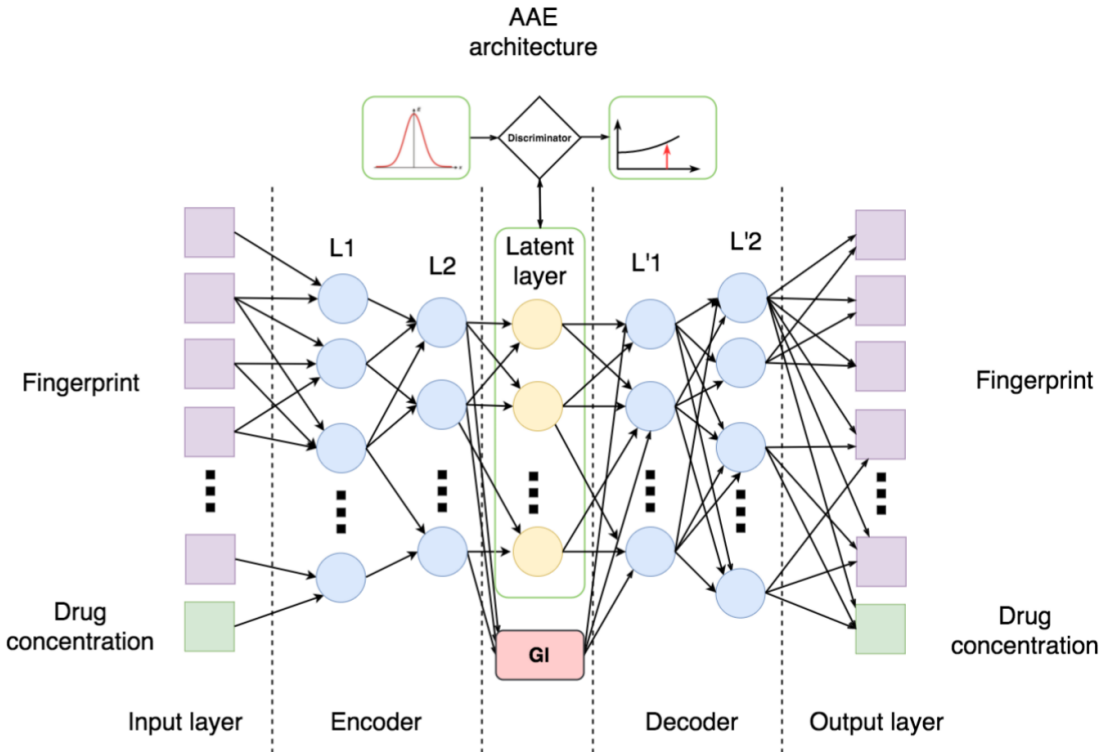
IN-SILICO APPROCH: CREATING NEW MOLECULES

- **Searching the chemical space by synthesizing new virtual molecules**



MolGAN

https://github.com/nicola-decao/MolGAN

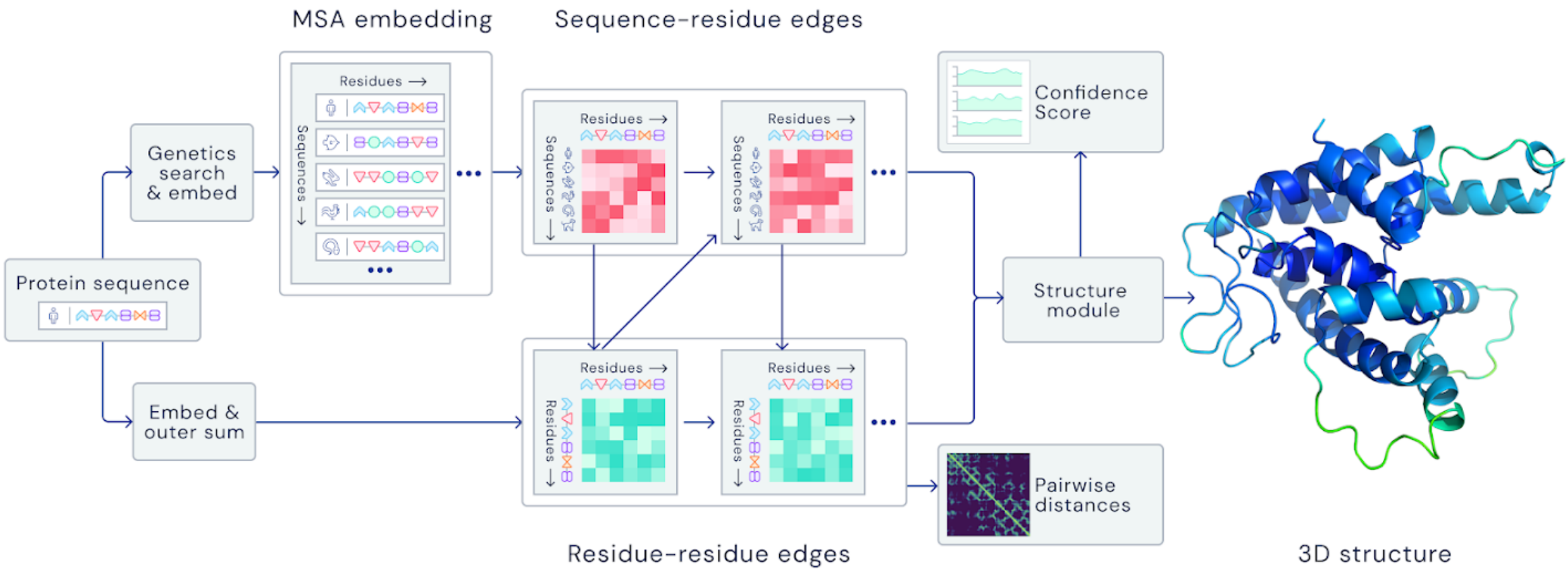


druGAN

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5355231/

IN-SILICO APPROCH: PREDICT 3D STRUCTURE FROM PROTEIN SEQUENCE

- **Searching in 3D to study binding of virtual molecules**

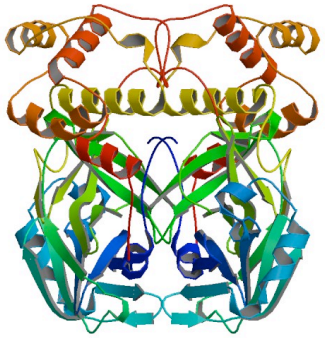


Source: https://github.com/deepmind/alphafold

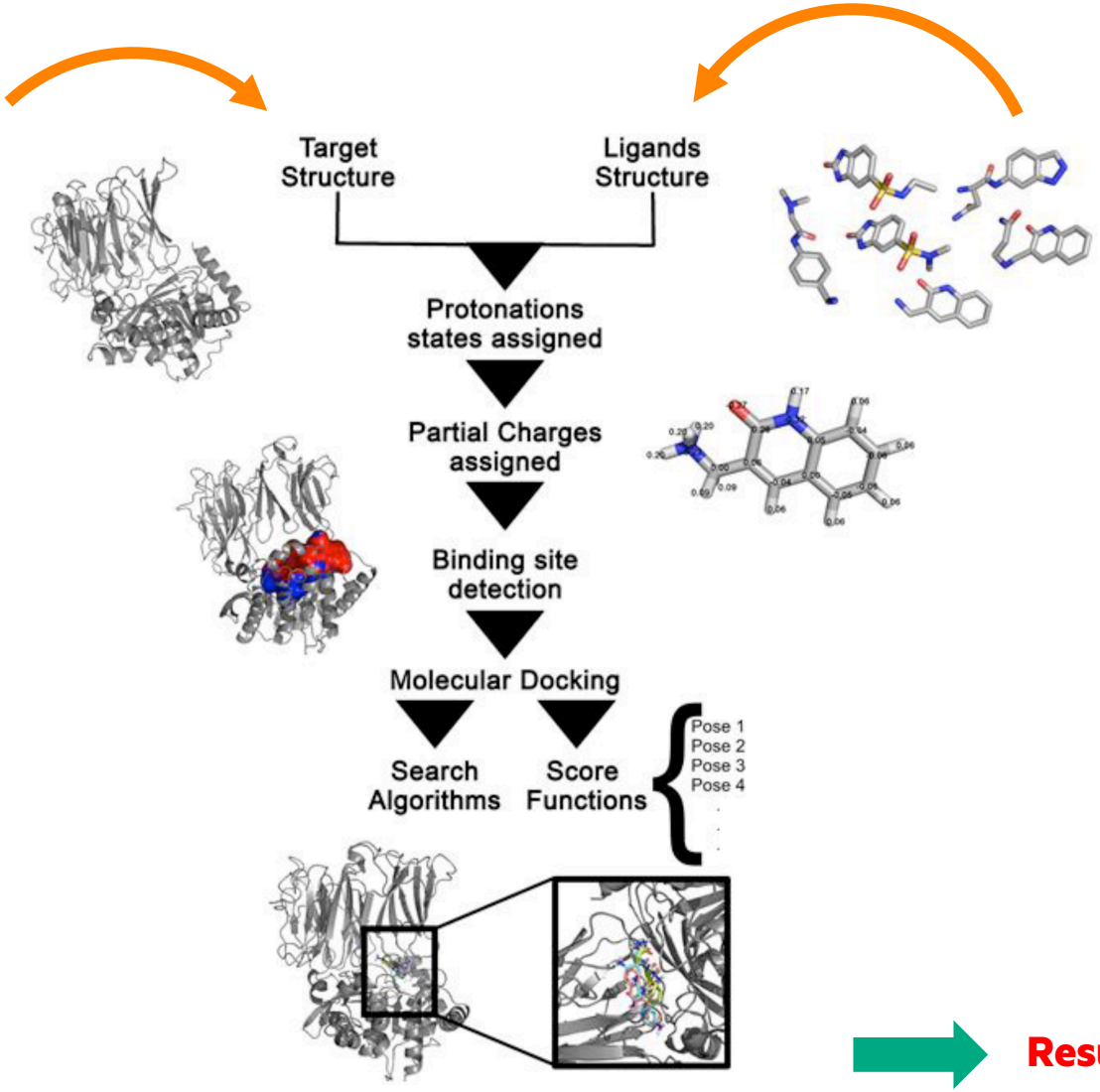
IN-SILICO APPROCH : SIMULATION USING MOLECULAR DYNAMICS

160+ COVID-19 targets

6LU7, 6VSB, 6LVN, 6LXT, 6VW1, 6Y2G, 6Y2F, 6Y2E, 6VXS, 6VWW, 6W02, 6W01, 6Y84, 6VYO, 6VYB, 6VXX, 6M03, 6M17, 5R84, 5R83, 5R7Y, 5R80, 5R82, 5R81, 5R7Z, 6W4H, 6W4B, 6M0J, 6M3M, 6LZG, 5REO, 5REN, 5RFZ, 5RFY, 5RFR, 5RFQ, 5RFT, 5RFS, 5RFV, 5RFU, 5RFX, 5RFW, 5RFJ, 5RFI, 5RFL, 5RFK, 5RFN, 5RFM, 5RFP, 5RFO, 5RG0, 6W41, 6W61, 6W63, 6W75, 6W6Y, 6YB7, 5REA, 5REC, 5REB, 5REE, 5RED, 5REG, 5REF, 5RE9, 5RE8, 5RE5, 5RE4, 5RE7, 5RE6, 5RFB, 5RFA, 5RFD, 5RFC, 5RFF, 5RFE, 5RFH, 5RFG, 5REY, 5REX, 5RF9, 5REZ, 5RF2, 5REP, 5RF1, 5RES, 5RF4, 5RER, 5RF3, 5REU, 5RF6, 5RET, 5RF5, 5REW, 5RF8, 5REV, 5RF7, 5REI, 5REH, 5REK



6Y2E in
10 different poses



https://www.mdpi.com/1422-0067/20/18/4574/htm

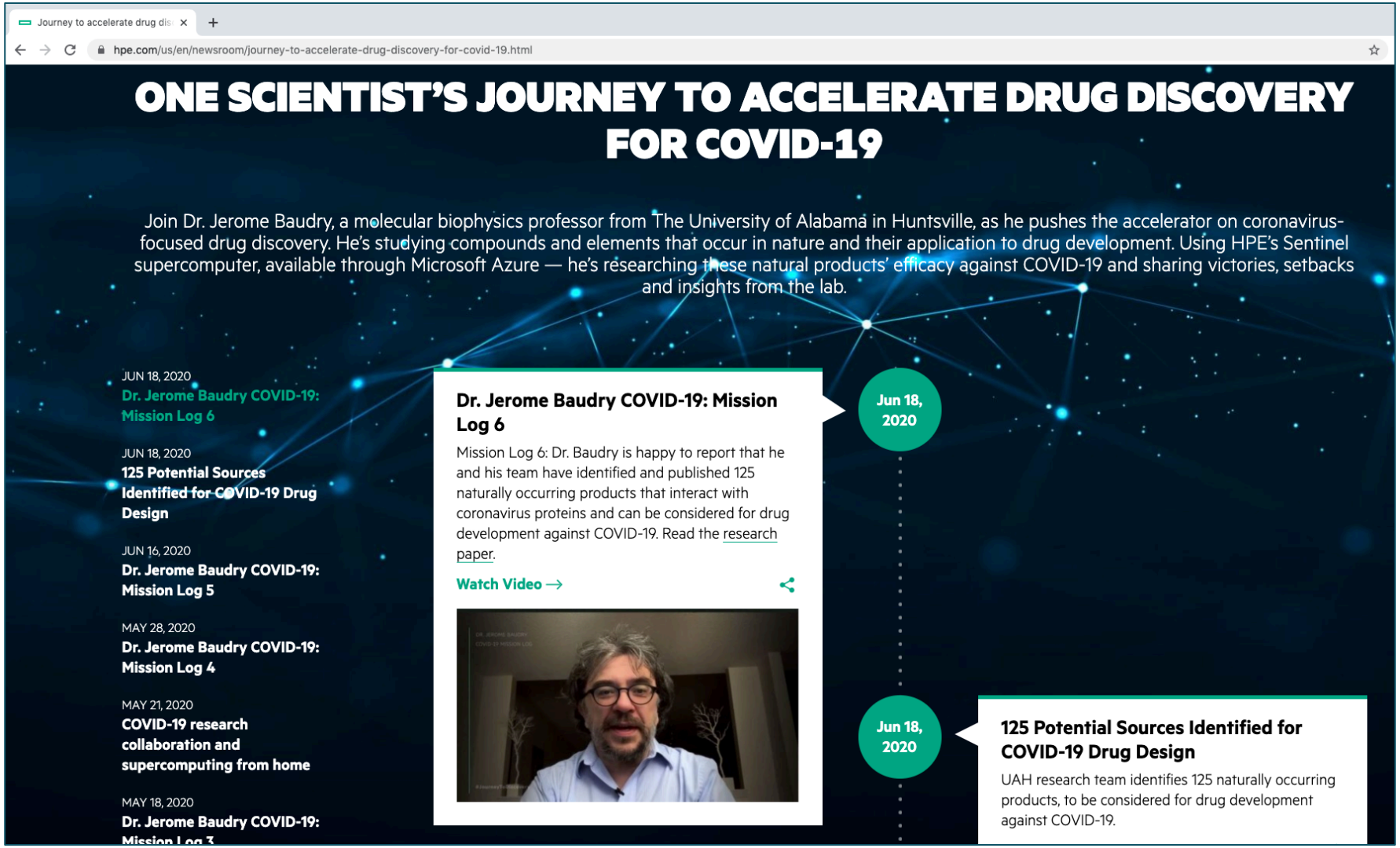
Potential drugs

Drug Databases
CHEMBL **(1.9M molecules)**
SWEETLEAD **(4326 molecules)**
COCONUT **(426K molecules)**

Result: Top 500 molecules

Collaborators: Jerome Baudry (UAH)

AN AWARD-WINNING JOURNEY WITH DR. JEROME BAUDRY



https://www.hpe.com/us/en/newsroom/journey-to-accelerate-drug-discovery-for-covid-19.html

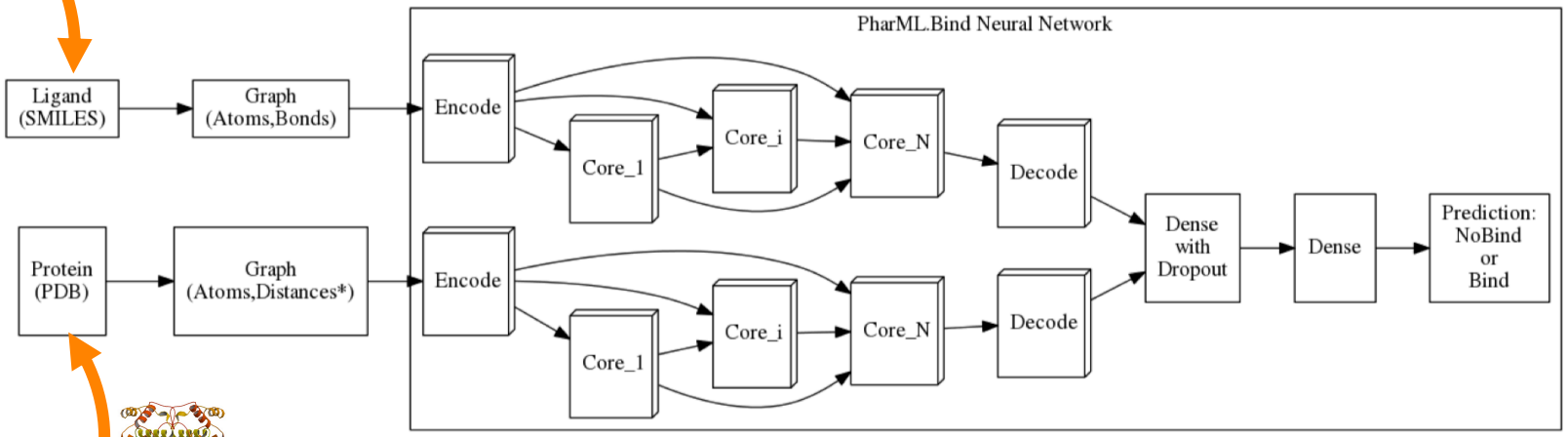
IN-SILICO APPROCH : PROTEIN-LIGAND AFFINITY PREDICTION USING AI

Drug Databases
CHEMBL **(1.9M molecules)**
SWEETLEAD **(4326 molecules)**
COCONUT **(426K molecules)**

Potential drugs

PHARML.BIND – A GRAPH NEURAL NETWORK TRAINED ON BINDINGDB

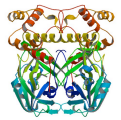
https://arxiv.org/abs/1911.06105



Result: Top 500 molecules

160+ COVID-19 targets

6LU7, 6VSB, 6LVN, 6LXT, 6VW1, 6Y2G, 6Y2F, 6Y2E, 6VXS, 6VWW, 6W02, 6W01, 6Y84, 6VYO, 6VYB, 6VXX, 6M03, 6M17, 5R84, 5R83, 5R7Y, 5R80, 5R82, 5R81, 5R7Z, 6W4H, 6W4B, 6M0J, 6M3M, 6LZG, 5REO, 5REN, 5RFZ, 5RFY, 5RFR, 5RFQ, 5RFT, 5RFS, 5RFV, 5RFU, 5RFX, 5RFW, 5RFJ, 5RFI, 5RFL, 5RFK, 5RFN, 5RFM, 5RFP, 5RFO, 5RG0, 6W41, 6W61, 6W63, 6W75, 6W6Y, 6YB7, 5REA, 5REC, 5REB, 5REE, 5RED, 5REG, 5REF, 5RE9, 5RE8, 5RE5, 5RE4, 5RE7, 5RE6, 5RFB, 5RFA, 5RFD, 5RFC, 5RFF, 5RFE, 5RFH, 5RFG, 5REY, 5REX, 5RF9, 5REZ, 5RF2, 5REP, 5RF1, 5RES, 5RF4, 5RER, 5RF3, 5REU, 5RF6, 5RET, 5RF5, 5REW, 5RF8, 5REV, 5RF7, 5REI, 5REH, 5REK



Binding Database is an open, web-accessible database of measured binding affinities, focusing on the interactions of proteins considered to be drug-targets with small, drug-like molecules. BindingDB contains 910,836 binding data, for over 6,263 protein targets and 378,980 small molecules.

Collaborators: Aaron Vose (NanoSemi), Yuri Peterson (MUSC), Ryan Yates (NCNPR)

SHOULD YOU BELIEVE IN AN AI MODEL ?

Database	# of molecules	Purpose	Key Result
BindingDB	820,433 molecules 7493 targets	Training a molecular-highway graph neural network	98.3% accuracy in under 25 minutes on 2,708,151 protein-ligand pairs
DUD.E	22846 molecules 102 targets 50 decoys /active site	Adversarial testing of the neural network	ROC metrics comparable to docking PharML AUC: 0.669-0.741 DockAlign AUC: 0.717
CHEMBL	13,377 targets (8 COVID-19) 1,950,765 molecules	Identify candidates for drug repurposing	# of molecules in clinical trials (>15%)
FDA	~500 molecules 8 COVID-19 targets	Identify candidates for drug repurposing	# of anti-viral drugs identified # of drugs undergoing clinical trials
<Collaborator>	8 COVID-19 targets 202,046 molecules	Identify products to combat current and future health threats, e.g., COVID-19	Ranked Top 500 (for 8 COVI-19 targets) ~ 1 day

Reference: Balma, J., Vose, A. D., Peterson, Y. K., Chittiboyina, A. G., Pandey, P., Yates, C. R., ... & Sukumar, S. R. (2020, December). Deep Learning Predicts Protein-Ligand Interactions. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 5627-5629)..

PHARML VS. STATE-OF-THE-ART ON FDA DATASET

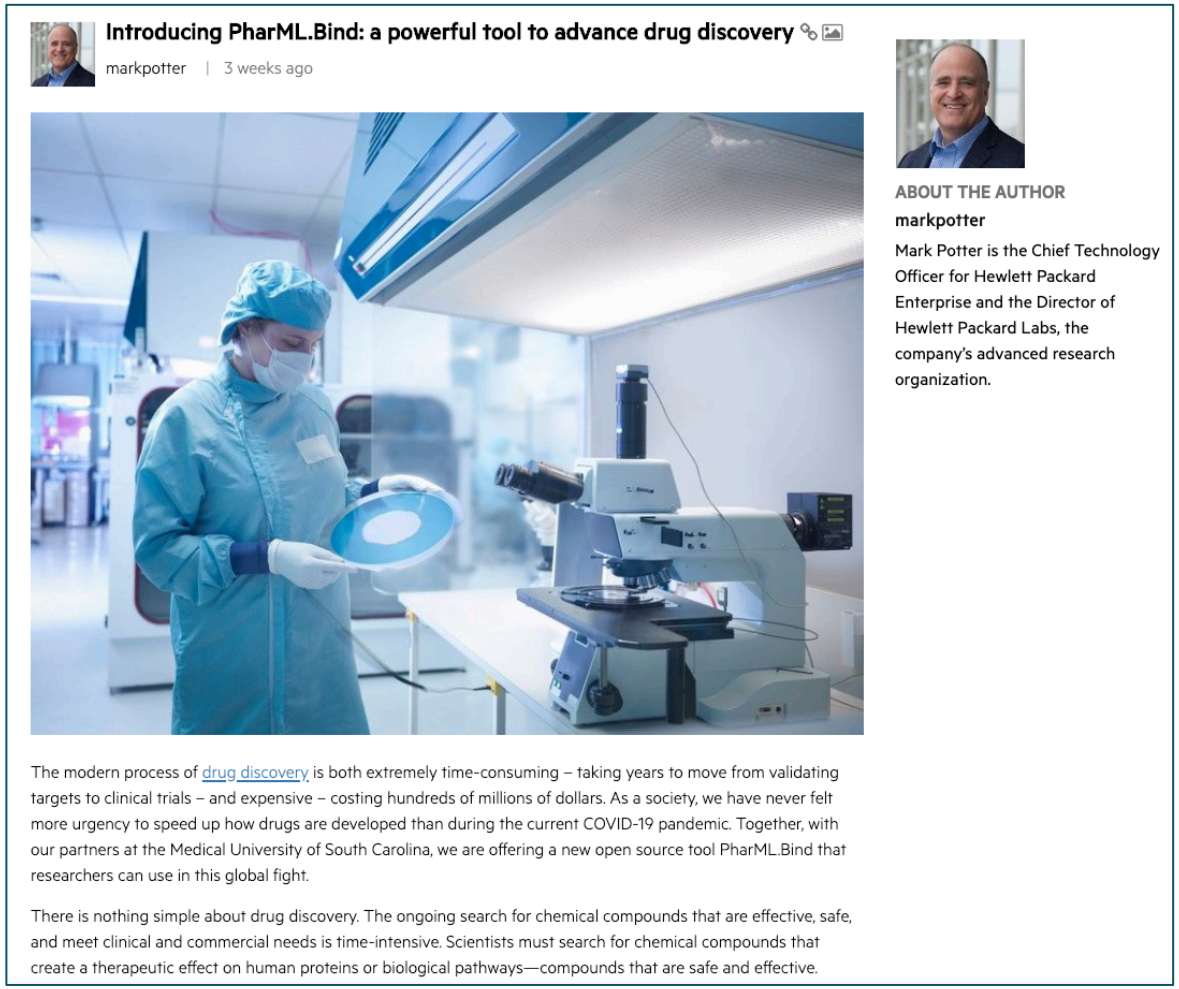
	PharML	Beck et al.	Odhar et al.	Kandeel et al.	Pant et al.
1	Ethoxyzolamide	Rapamycin (Sirolimus)	Perampanel	Chromocarb	Lactulose
2	Bosutinib	Sirolimus	Conivaptan	Ribavarin	Oxytocin
3	Acetazolamide	Tiotropium Bromide	Sonidegib	Telbivudine	Boceprevir
4	Nilotinib	Everolimus	Azelastine	Vitamin B12	Saquinavir
5	Sitagliptin	Tacrolimus (FK506)	Idelalisib	Aminophylline	Adenosine
6	Itraconazole	Etomidate	Suvorexant	Aminophylline	Masoprocol
7	Atazanavir	Daptomycin	Olaparib	Nicotinamide	Doxorubicin
8	Indinavir	Zolmitriptan	Ponatinib	Trifusal	Cromolyn
9	Nelfinavir	Ly2835219	Loxapine	Bemegride	Lopinavir
10	Chembl66274	Batimastat	Tolvaptan	Aminosalicylate	Dibucaine
11	Methazolamide	Batimastat (BB-94)		Pyrazinamide	Ritonavir
12	Atorvastatin	Bacitracin		Temozolomide	Regadenson
13	Efavirenz	Ivermectin		Methazolamide	Cladribine
14	Simvastatin	Methscopolamine		Tioxolone	Daunorubicin
15	Mafenide	Thiostrepton		Propylthiouracil	Albuterol;
16	Lapatinib	Somatostatin		Cysteamine	Dapagliflozin
17	Nelfinavir	Radotinib(iy-5511)		Methoxamine	Pravastatin
18	Zafirlukast	Foxy-5		Zonisamide	Pemetrexed
19	Chembl262022	Scopolamine		Octopamine	Protirelin
20	Dasatinib	Acetylcholine Chloride		Amiloride	Mupirocin

WHY PHARML: PHARML VS. STATE-OF-THE-ART ON FDA DATASET

	PharML.BIND Rank	Beck et al. , 2020 Rank
1	ETHOXYZOLAMIDE	
2	BOSUTINIB	47
3	ACETAZOLAMIDE	294
4	NILOTINIB	488
5	SITAGLIPTIN	2680
6	ITRACONAZOLE	461
7	ATAZANAVIR	66
8	INDINAVIR	2252
9	NELFINAVIR	402/552
10	CHEMBL66274	
11	METHAZOLAMIDE	1918
12	ATORVASTATIN	436/706
13	EFAVIRENZ	116/426
14	SIMVASTATIN	769/2045
15	MAFENIDE	3320
16	LAPATINIB	172/1860
17	NELFINAVIR	402/552
18	ZAFIRLUKAST	1305/1306
19	CHEMBL262022	
20	DASATINIB	126/127

- **Open-source**
  - https://github.com/jbalma/pharml
- **Training dataset**
  - Model trained on 64 GPUs for 1 week on BindingDB
- **Active-site agnostic predictor**
- **Speed: 9 compounds/second on a GPU**
- **Validation**
  - Probability of binding vs. random selection
  - Lipinski's rule-of-5

ARTIFICIAL INTELLIGENCE FOR PROTEIN-LIGAND BINDING PREDICTION



Introducing PharML.Bind: a powerful tool to advance drug discovery

markpotter | 3 weeks ago

ABOUT THE AUTHOR

markpotter

Mark Potter is the Chief Technology Officer for Hewlett Packard Enterprise and the Director of Hewlett Packard Labs, the company's advanced research organization.

The modern process of drug discovery is both extremely time-consuming – taking years to move from validating targets to clinical trials – and expensive – costing hundreds of millions of dollars. As a society, we have never felt more urgency to speed up how drugs are developed than during the current COVID-19 pandemic. Together, with our partners at the Medical University of South Carolina, we are offering a new open source tool PharML.Bind that researchers can use in this global fight.

There is nothing simple about drug discovery. The ongoing search for chemical compounds that are effective, safe, and meet clinical and commercial needs is time-intensive. Scientists must search for chemical compounds that create a therapeutic effect on human proteins or biological pathways—compounds that are safe and effective.

MUSC Medical University of South Carolina

Search

About | Education | Research | Innovation | Patient Care | Resources | Human Resources

MUSC Home › About › MUSC Catalyst News › Drug Discovery

MUSC, HPE make innovative drug discovery software open source

Mikie Hayes | May 15, 2020

Dr. Yuri Peterson, center, with College of Pharmacy students. They chose him as Teacher of the Year in 2019.

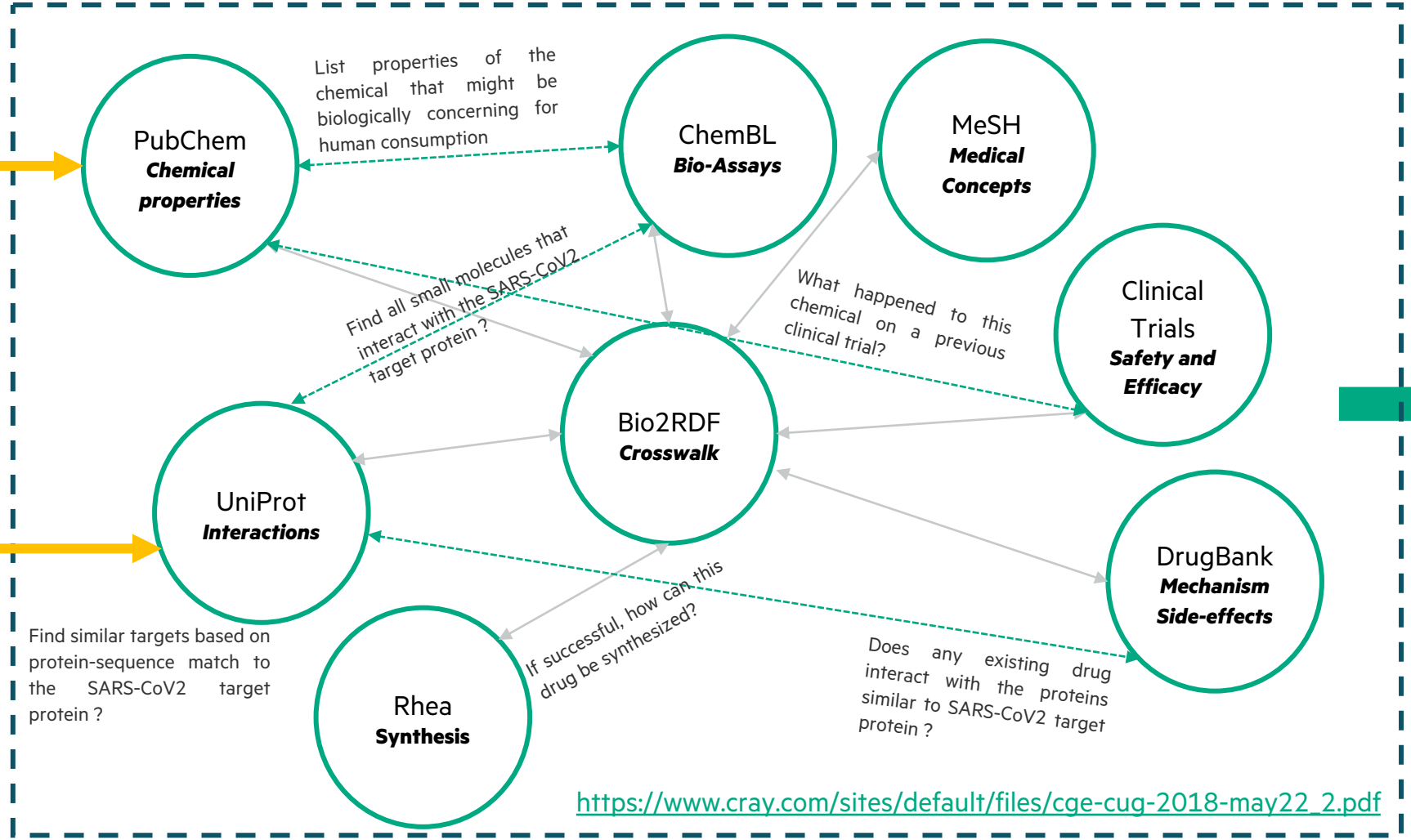
In yet another innovative partnership between academia and industry, MUSC and Hewlett Packard Enterprise (HPE) are making available to researchers worldwide an innovative new drug discovery program they co–developed, PharML.Bind, in an open-source release. Through this release, MUSC and HPE aim to accelerate the search for effective therapies against COVID–19, the disease caused by the SARS–CoV–2 virus.

Open-source code: https://github.com/jbalma/pharml

DATA : SUGGEST MOLECULES BASED ON MECHANISM, SIMILARITY, ETC.

Drug Databases
CHEMBL **(1.9M molecules)**
SWEETLEAD **(4326 molecules)**
COCONUT (426K molecules)

Potential drugs



Result: Top 500 molecules

160+ COVID-19 targets

6LU7, 6VSB, 6LVN, 6LXT, 6VW1, 6Y2G, 6Y2F, 6Y2E, 6VXS, 6VWW, 6W02, 6W01, 6Y84, 6VYO, 6VYB, 6VXX, 6M03, 6M17, 5R84, 5R83, 5R7Y, 5R80, 5R82, 5R81, 5R7Z, 6W4H, 6W4B, 6M0J, 6M3M, 6LZG, 5REO, 5REN, 5RFZ, 5RFY, 5RFR, 5RFQ, 5RFT, 5RFS, 5RFV, 5RFU, 5RFX, 5RFW, 5RFJ, 5RFI, 5RFL, 5RFK, 5RFN, 5RFM, 5RFP, 5RFO, 5RG0, 6W41, 6W61, 6W63, 6W75, 6W6Y, 6YB7, 5REA, 5REC, 5REB, 5REE, 5RED, 5REG, 5REF, 5RE9, 5RE8, 5RE5, 5RE4, 5RE7, 5RE6, 5RFB, 5RFA, 5RFD, 5RFC, 5RFF, 5RFE, 5RFH, 5RFG, 5REY, 5REX, 5RF9, 5REZ, 5RF2, 5REP, 5RF1, 5RES, 5RF4, 5RER, 5RF3, 5REU, 5RF6, 5RET, 5RF5, 5REW, 5RF8, 5REV, 5RF7, 5REI, 5REH, 5REK

https://www.cray.com/sites/default/files/cge-cug-2018-may22_2.pdf

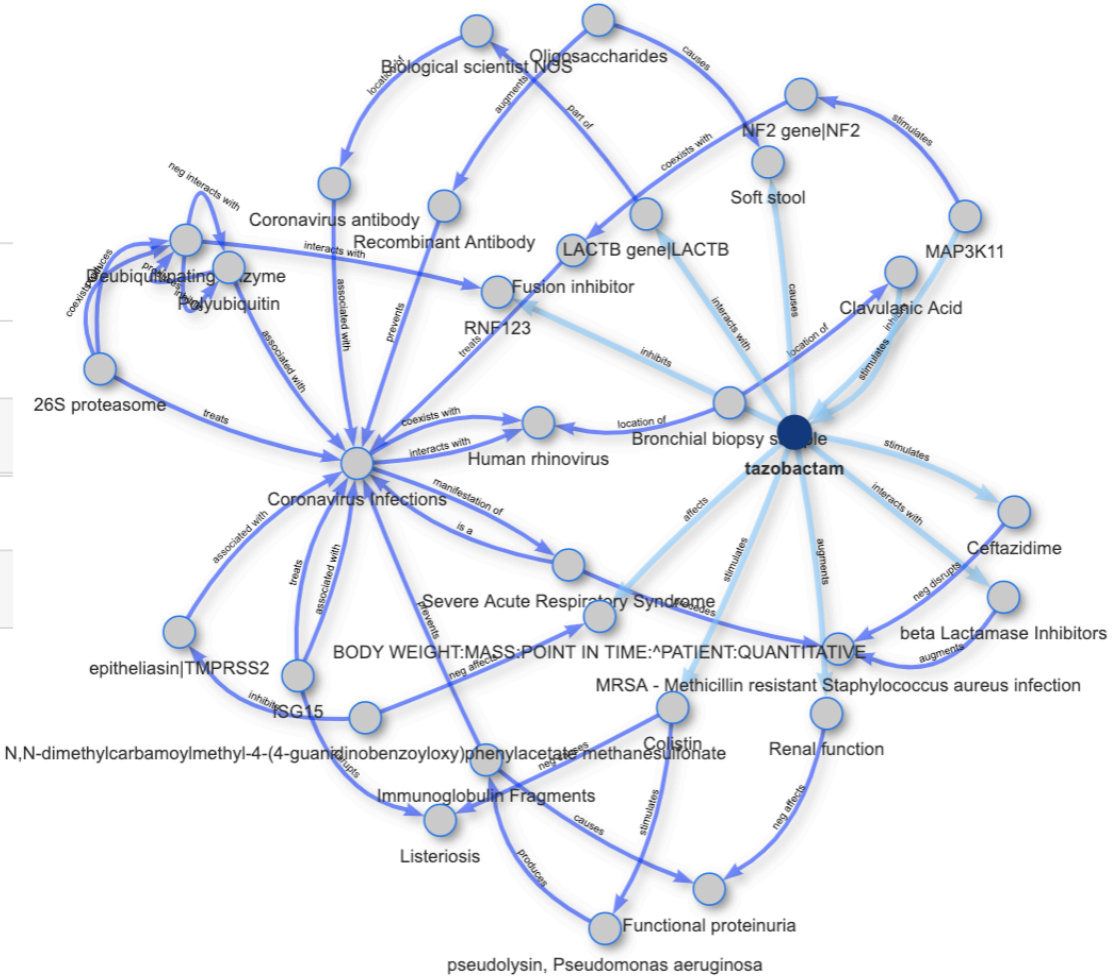
Collaborators: Ryan Yates (NCNPR – Ole Miss)

SEARCHING FOR EMPIRICAL EVIDENCE USING AI

- **Searching for the what-is, what-if, what-else and what–could be?**

How could tazobactam treat Coronavirus infections?

tazobactam	inhibits	RNF123	← interacts with	Deubiquitinating Enzyme	← coexists with	26S proteasome	treats	Coronavirus Infections
tazobactam	← inhibits	MAP3K11	stimulates	NF2 gene|NF2	coexists with	Fusion inhibitor	treats	Coronavirus Infections
tazobactam	inhibits	RNF123	← interacts with	Deubiquitinating Enzyme	← produces	26S proteasome	treats	Coronavirus Infections
tazobactam	augments	Renal function	neg affects	Functional proteinuria	← causes	Immunoglobulin Fragments	prevents	Coronavirus Infections
tazobactam	stimulates	Colistin	stimulates	pseudo lysin	produces	Immunoglobulin Fragments	prevents	Coronavirus Infections



SEARCHING FOR STRUCTURE, FUNCTIONAL AND EMPIRICAL EVIDENCE

Open Data

30 TBs on disk, 150+ billion medical facts

Dataset	Size (on disk)	Size (triples)
UniProt (Mar 2020)	12.7 TB	87.6 Billion
PubChemRDF (v1.6.3 beta)	13.0 TB	80.0 Billion
ChEMBL-RDF (27.0)	81 GB	539 Million
Bio2RDF (Release 4)	2.4 TB	11.5 Billion
OrthoDB (v10)	275 GB	2.2 Billion
Biomodels (r31)	5.2 GB	28 Million
Biosamples (v20191125)	112.8 GB	1.1 Billion
Reactome (r71)	3.2 GB	19 Million

Prototype

Query interface / APIs

Clustering

Ranking

Spectral analysis

Graph traversal

....

Built-in analytics/algorithms

Molecular synthesis
- MolGAN, druGAN

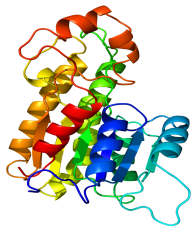
Structure prediction
- AlphaFold

Protein-Ligand interaction
- DBTA, PharML

Sequence Similarity estimation
- Smith-Waterman algorithm

User-defined AI

Data store



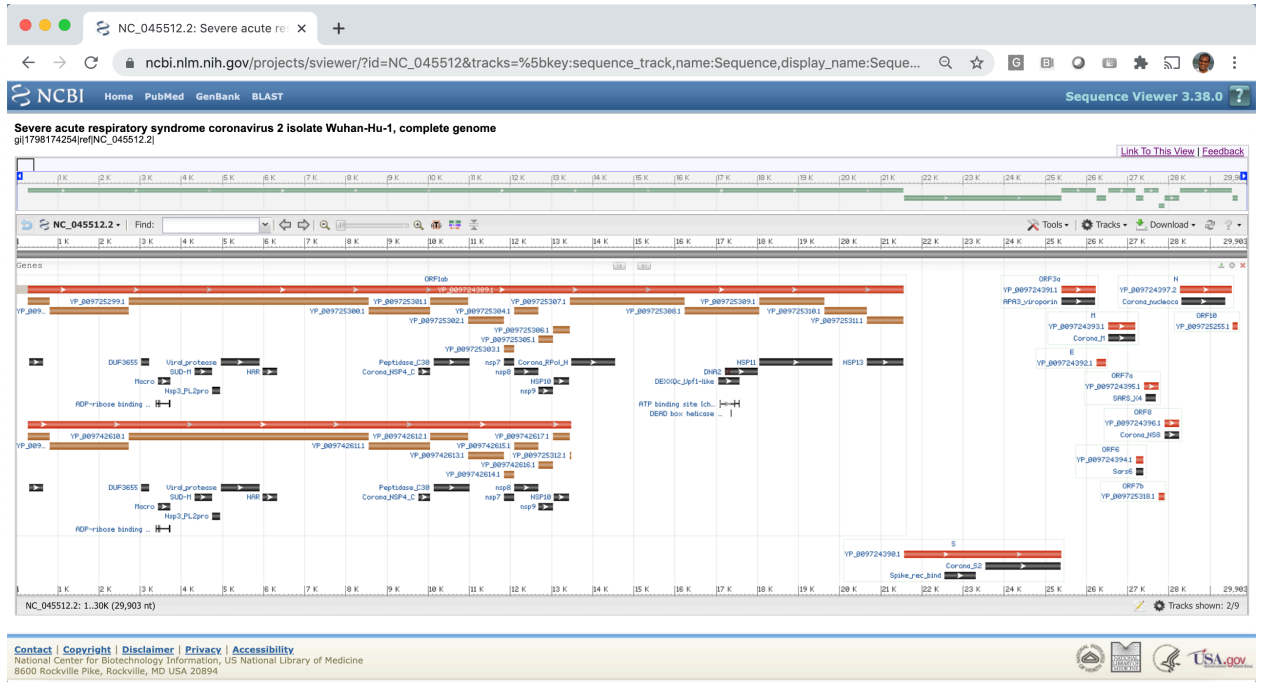
Sequences

Time series

Associations

Reference: Rickett, C. D., Maschhoff, K. J., & Sukumar, S. R. (2020, December). Massively Parallel Processing Database for Sequence and Graph Data Structures Applied to Rapid-Response Drug Repurposing. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 2967-2976).

PROTEIN SIMILARITY SEARCH TO UNDERSTAND COVID-19 EVOLUTION



SARS2 - SAR1 : Similarity

NSP1	NSP2	NSP3	NSP4	NSP5	NSP6	NSP7	NSP8	NSP9	NSP10	NSP11	NSP12	NSP13	NSP14	NSP15	NSP16	E	N	S
0.862	0.73	0.802	0.856	0.972	0.906	0.993	0.981	0.974	0.982	0.888	0.972	1	0.968	0.92	0.956	0.949	0.919	0.807

Searching for the ”what-is”

KNOWING ABOUT PROTEIN EVOLUTION CAN BE SIGNIFICANT

CowPox - Smallpox

Query: U5TAC1_COWPX

Mnemonic	Scientific Name	Score
F13_VACCC	Vaccinia virus (strain Copenhagen)	1
F13_VACCW	Vaccinia virus (strain Western Reserve)	1
F13_VACCP	Vaccinia virus (strain L-IVP)	1
F13_VACCI	Vaccinia virus (strain IHD-J)	0.99
F13_VACCA	Vaccinia virus (strain Ankara)	0.99

"Cowpox virus, a cousin of variola virus, causes a mild smallpox-like disease in cows. The story goes that Jenner was told that milkers who acquired the "cow-version" of smallpox were immune to the human version of the disease. "

https://www.sciencedaily.com/releases/2017/08/170819103656.htm

Zika – Dengue and West Nile

Query: A0A060H177_ZIKV

Mnemonic	Scientific Name	Score
POLG_DEN22	Dengue virus type 2 (isolate Malaysia M2)	0.61
POLG_DEN2H	Dengue virus type 2 (strain Thailand/TH-36/1958)	0.58
POLG_DEN2U	Dengue virus type 2 (strain Thailand/PUO-218/1980)	0.53
POLG_DEN1A	Dengue virus type 1 (strain Thailand/AHF 82-80/1980)	0.5
A0A1V0E2I7_ZIKV	Zika virus	0.47

The structure reveals that Zika virus is similar to other flaviviruses, such as Dengue virus and West Nile virus.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5870307/

Chickenpox- Shingles

Query: SCAF_VZVD

Mnemonic	Scientific Name	Score
SCAF_VZVD	Varicella-zoster virus (strain Dumas)	1
SCAF_VZVD	Varicella-zoster virus (strain Dumas)	1
Q6QCM2_HHV3	Human alphaherpesvirus 3	1
Q6QCM2_HHV3	Human herpesvirus 3	1
A0A4D6F9U7_HHV3	Human alphaherpesvirus 3	1

Shingles, also known as zoster or herpes zoster, is a viral disease characterized by a painful skin rash with blisters in a localized area. Shingles is due to a reactivation of varicella zoster virus (VZV) in a person's body. The disease chickenpox is caused by the initial infection with VZV.

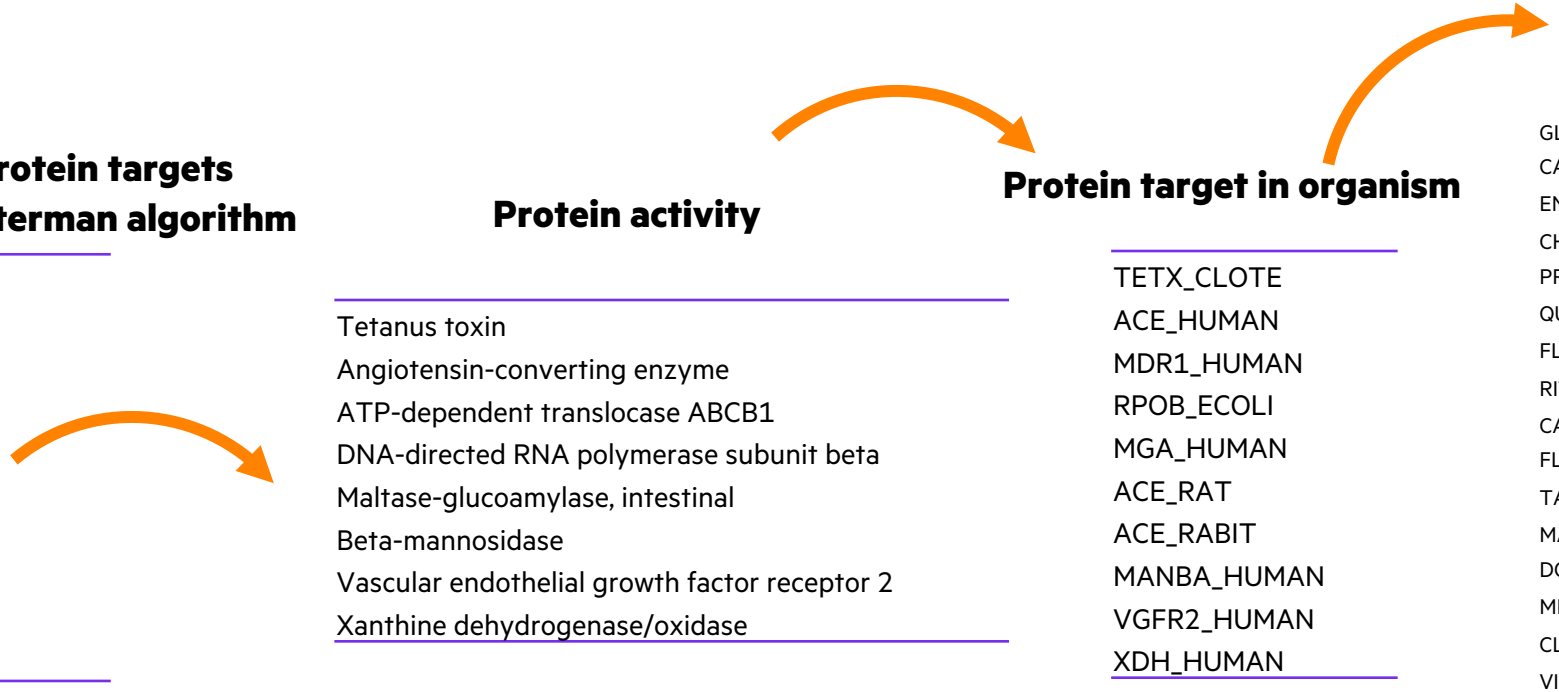
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2563790/

A researcher can now compare against 4+ million proteins per query in less than 60 seconds

Searching for the "what-if"

CONNECTING THE DOTS USING KNOWLEDGE GRAPHS

Given a target protein of interest – search for molecules that have functional evidence



Hypothesis generation with 13 different databases, 30 TBs on disk, 150+ billion medical facts in ~ 60 seconds

Searching for the ”what-could-be”

RESULTS IN RETROSPECTIVE

Searching for the ”what-else”

- **Tetanus toxoid vaccine correlations were more than co-incidence**
  - Similar connections between the tetanus toxin protein sequence and other coronaviruses from the Rhinolophus affinis bat
  - Correlation between tetanus vaccination rates in the US and COVID-19 asymptomatic and mortality rates, especially when compared against countries with lower tetanus vaccination rates and significantly higher COVID-19 mortality rates
  - Correlation between DTP vaccination rates in children worldwide and COVID-19 symptom severity in children
  - Correlation between TDaP vaccination rates in pregnant women and asymptomatic COVID-19 rates
  - Potential for correlation between tetanus vaccination rates in prison inmates and asymptomatic COVID-19 rates
  - Neurological symptoms and TMJ that are typical symptoms of the tetanus disease are being reported in several COVID-19 patients
  - Additional studies also suggesting a protective effect from the DTP/TDaP vaccines against COVID-19

Reference: Rickett, C. D., Maschhoff, K. J., & Sukumar, S. R. (2021). Does tetanus vaccination contribute to reduced severity of the COVID-19 infection?. *Medical Hypotheses*, *146*, 110395.

RESULTS IN RETROSPECTIVE

Protective heterologous T cell immunity in COVID-19 induced by MMR and Tdap vaccine antigens

Vijayashree Mysore, Xavier Cullere, Matthew L. Settles, Xinge Ji, Michael W. Kattan, Michaël Desjardins, Blythe Durbin-Johnson, Tal Gilboa, Lindsey R. Baden, David R. Walt, Andrew Lichtman, Lara Jehi, Tanya N. Mayadas

doi: https://doi.org/10.1101/2021.05.03.441323

ABSTRACT

T cells are critical for control of viral infection and effective vaccination. We investigated whether prior Measles-Mumps-Rubella (MMR) or Tetanus-Diphtheria-pertussis (Tdap) vaccination elicit cross-reactive T cells that mitigate COVID-19. Using co-cultures of antigen presenting cells (APC) loaded with antigens and autologous T cells, we found a high correlation between responses to SARS-CoV-2 (Spike-S1 and Nucleocapsid) and MMR and Tdap vaccine proteins in both SARS-CoV-2 infected individuals and individuals immunized with mRNA-based SARS-CoV-2 vaccines. The overlapping T cell population contained effector memory T cells (TEMRA) previously implicated in anti-viral immunity and their activation required APC-derived IL-15. TCR- and scRNA-sequencing detected cross-reactive clones with TEMRA features among the cells recognizing SARS-CoV-2, MMR and Tdap epitopes. A propensity-weighted analysis of 73,582 COVID-19 patients revealed that severe disease outcomes (hospitalization and transfer to intensive care unit or death) were reduced in MMR or Tdap vaccinated individuals by 38-32% and 23-20% respectively. In summary, SARS-CoV-2 re-activates memory T cells generated by Tdap and MMR vaccines, which may reduce disease severity.

ORIGINAL RESEARCH article

Front. Immunol., 16 October 2020 | https://doi.org/10.3389/fimmu.2020.586984

Potential Cross-Reactive Immunity to SARS-CoV-2 From Common Human Pathogens and Vaccines

Pedro A. Reche*

Department of Immunology & O2, Faculty of Medicine, University Complutense of Madrid, Madrid, Spain

The recently emerged SARS-CoV-2 causing the ongoing COVID-19 pandemic is particularly virulent in the elderly while children are largely spared. Here, we explored the potential role of cross-reactive immunity acquired from pediatric vaccinations and exposure to common human pathogens in the protection and pathology of COVID-19. To that end, we sought for peptide matches to SARS-CoV-2 (identity ≥ 80%, in at least eight residues) in the proteomes of 25 human pathogens and in vaccine antigens, and subsequently predicted their T and B cell reactivity to identify potential cross-reactive epitopes. We found that viruses subject to pediatric vaccinations do not contain cross-reactive epitopes with SARS-CoV-2, precluding that they can provide any general protection against COVID-19. Likewise, common viruses including rhinovirus, respiratory syncytial virus, influenza virus, and several herpesviruses are also poor or null sources of cross-reactive immunity to SARS-CoV-2, discarding that immunological memory against these viruses can have any general protective or pathological role in COVID-19. In contrast, we found combination vaccines for treating diphtheria, tetanus, and pertussis infectious diseases (DTP vaccine) to be significant sources of potential cross-reactive immunity to SARS-CoV-2. DTP cross-reactive epitopes with SARS-CoV-2 include numerous CD8 and CD4 T cell epitopes with broad population protection coverage and potentially neutralizing B cell epitopes in SARS-CoV-2 Spike protein. Worldwide, children receive several DTP vaccinations, including three-four doses the first year of life and one at 4–6 years of age. Moreover, a low antigenic Tdap dose is also given at ages 9–14. Thereby, children may well be protected from SARS-CoV-2 through cross-reactive immunity elicited by DTP vaccinations, supporting testing in the general population to prevent COVID-19.

Most research was focused on the virus ignoring potential cross-immunity from other organisms, viruses, vaccines etc.

RESULTS IN RETROSPECTIVE

Coronavirus: Dexamethasone proves first life-saving drug

By Michelle Roberts
Health editor, BBC News online

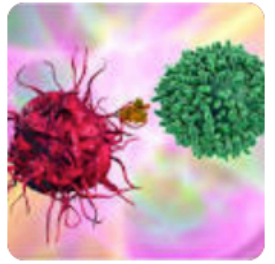
16 June 2020 | Comments

News-Medical.net

Children's DTP vaccine may provide cross-immunity for COVID-19

Reche Pedro A., Potential Cross-Reactive Immunity to SARS-CoV-2 From Common Human Pathogens and Vaccines, Frontiers in Immunology, ...

Nov 17, 2020



MONTREAL HEART INSTITUTE

Source: *Montreal Heart Institute*

January 22, 2021 22:58 ET

Colchicine reduces the risk of COVID-19-related complications

Positive results from COLCORONA trial show that colchicine is the only effective oral medication for treating non-hospitalized patients

SUMMARY : VALUE OF HPC, AI AND DATA SCIENCE

Converged HPC, AI and Data science methods offer unique capabilities

Scenarios	Approach	Metrics
Database: NCNPR Viral target: Spike Protein	Study similarity of the ranked list	Number of common drug molecules suggested
Database: NCNPR, SWETTLEAD Viral target: Spike Protein	Study robustness to molecule diversity	Number of molecules in active trials
Database: CHEMBL Target: 3CLPro, Spike	Evaluate robustness to protein promiscuity	Number of molecules suggested across targets

SUMMARY : VALUE OF HPC, AI AND DATA SCIENCE

Insights based on lessons learned

	AI Generated Drug Molecules	Existing Drug Databases (e.g., CHEMBL)
Emerging Disease Targets (e.g., COVID-19)	**First-principle Design** By simulating molecular dynamics CryoEM Imaging First-principle simulation	**Drug Repurposing** Connecting-the-dots in literature Genome Sequencing Drug Repurposing
Well-Studied Disease Targets (e.g., Influenza)	**Mechanism-driven Design** Learning from prior knowledge Molecular Synthesis Binding Prediction	**Drug Optimization** Side-effect/Dosage Prediction Rapid Literature Search Federated learning

THANK YOU

Email: sreenivas.sukumar@hpe.com

LinkedIn: https://www.linkedin.com/in/rangan/

Twitter: Rangan_Sukumar