e-Seminar #25

Addressing the challenge of drug discovery with Machine Learning and Exascale computing

Presenter:
Agastya Bhati
(University College London)

28 July 2022

The e-Seminar will start at 2pm CEST / 1pm BST

Moderator:
Gavin Pringle
(EPCC, University of Edinburgh)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823712

INSILICO WORLD

https://insilicoworld.slack.com/archives/C0151M02TA4

The e-Seminar series is run in collaboration with:

VPH Institute
Building the Virtual Physiological Human

e-Seminar #25

Addressing the challenge of drug discovery with Machine Learning and Exascale computing

Presenter:
Agastya Bhati
(University College London)

28 July 2022

Welcome!

Moderator:
Gavin Pringle
(EPCC, University of Edinburgh)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823712

INSILICO WORLD

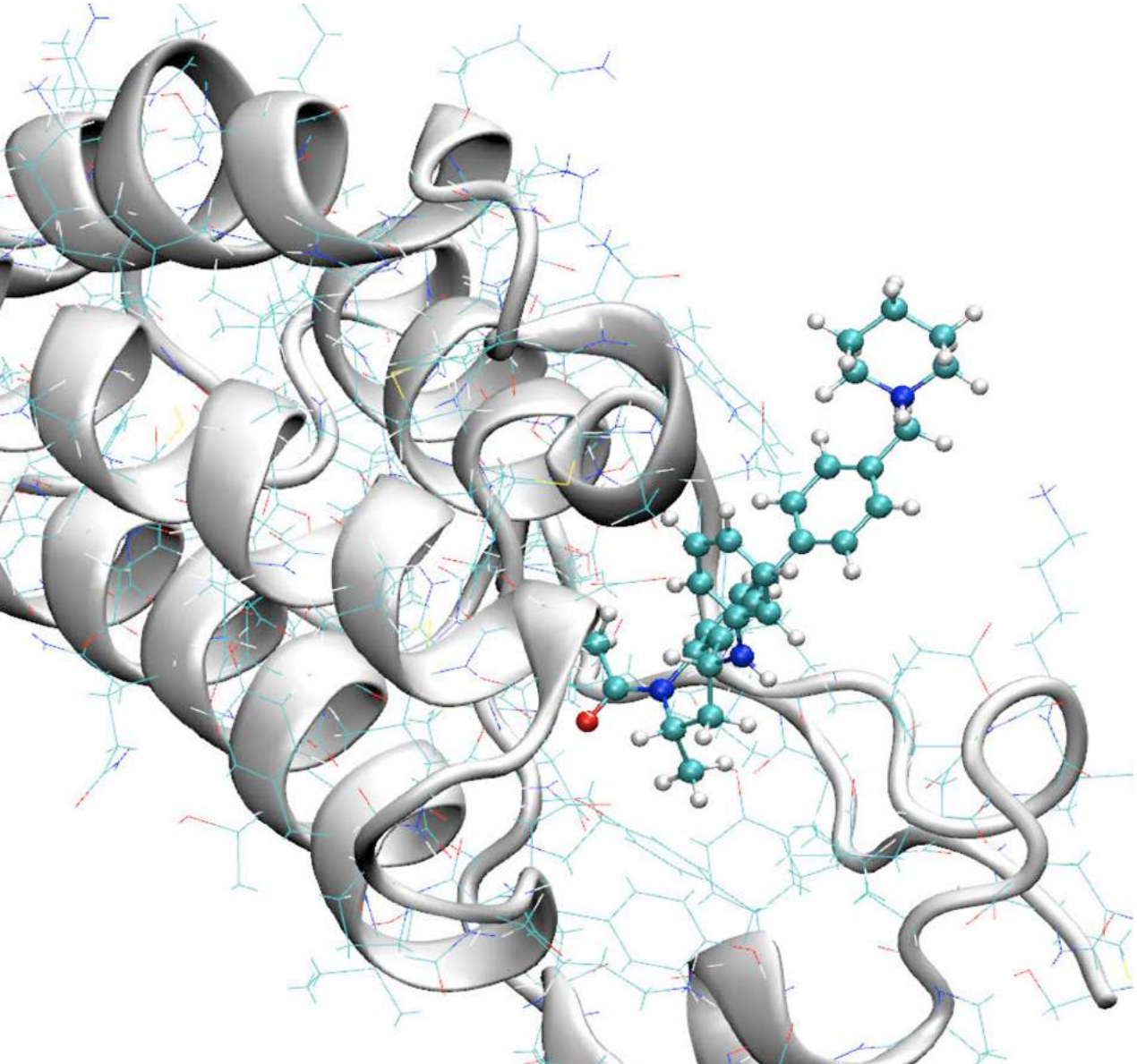
https://insilicoworld.slack.com/archives/C0151M02TA4

The e-Seminar series is run in collaboration with:

VPH Institute
Building the Virtual Physiological Human

Outline

1. Free energy prediction
2. Machine learning approaches
3. Physics-based approaches
4. Hybrid ML and PB approaches



Free Energy Predictions

DRUG DISCOVERY

~10^8 products

PRE CLINICAL

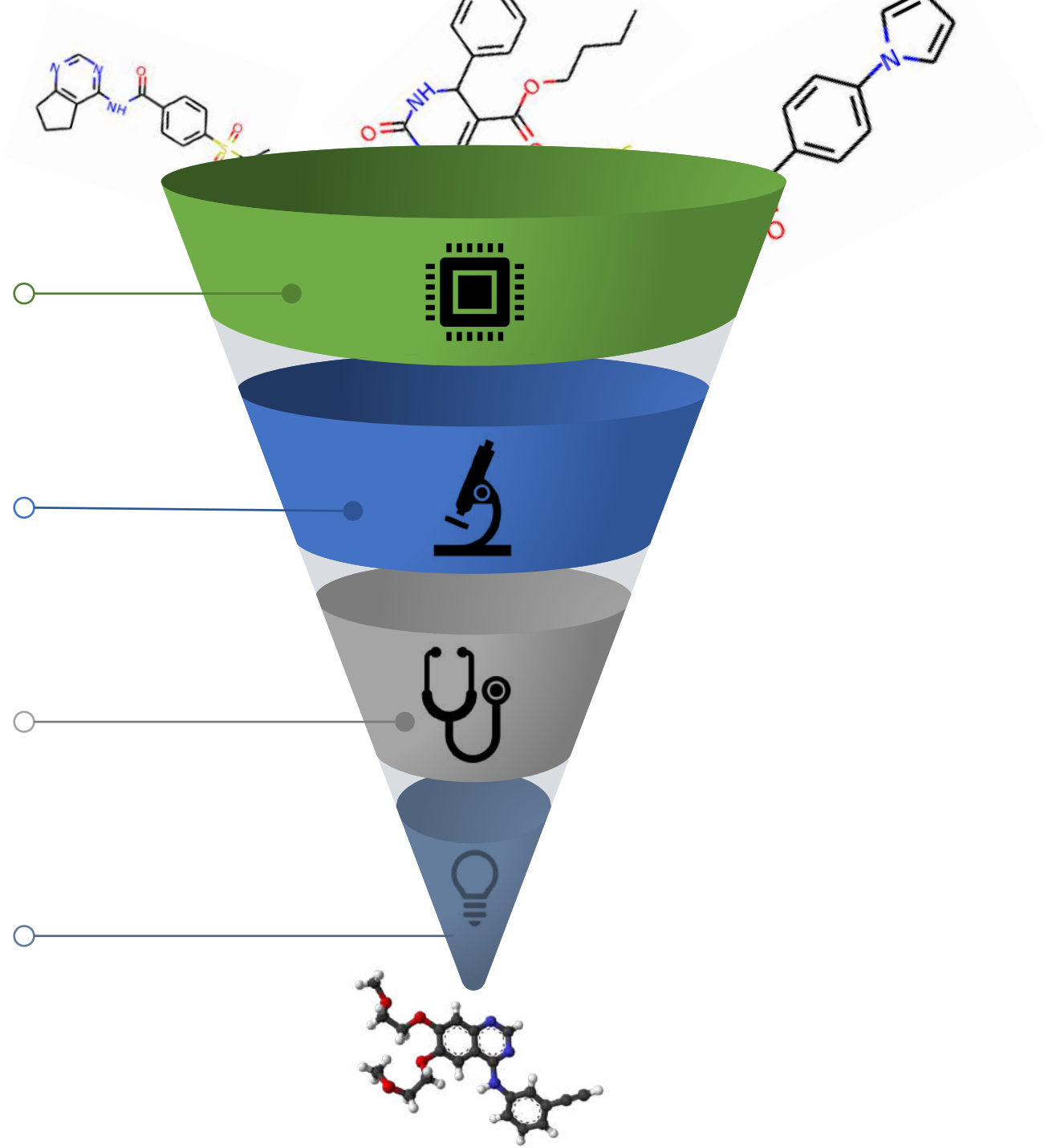
11,000 products

CLINICAL TRIALS

6,300 products

FDA APPROVAL

111 products



Target based compound screening

10^{60} estimated drug-like compounds

COMPOUND DISCOVERY

Mining massive building block or de-novo generated libraries

INTERESTING?

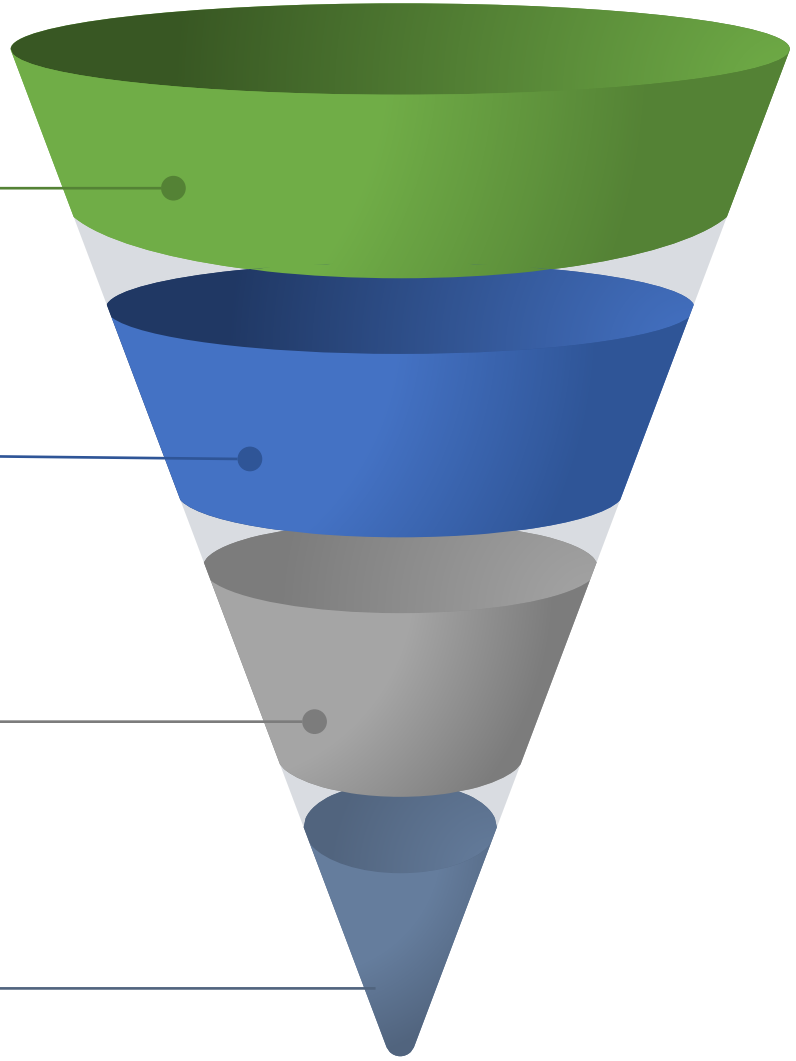
Does this compound inhibit or interact with the target?

TOXICOLOGY

Is this compound reasonably safe?

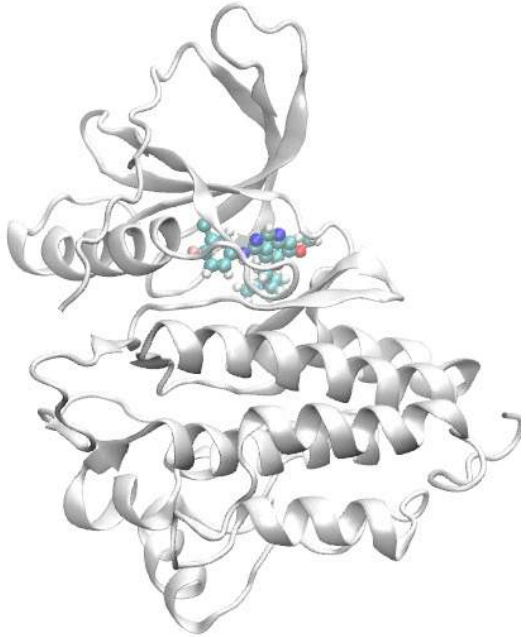
SYNTHESIS

Can we buy it, is it from available building blocks, or do we need to hire a medicinal chemist?

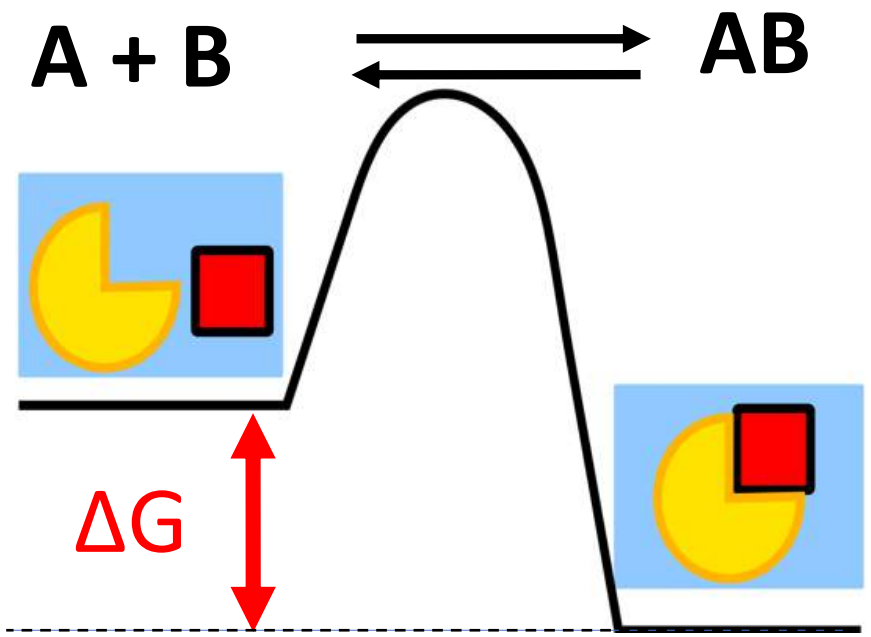


Drug-protein Binding Affinities

- Drugs are designed to act on specific proteins within the body.
- If the wrong drug is used, it cannot interact with the protein.
- The protein shape or configuration can also change between patients.
- We can use computer programs to determine the best drug for a specific protein in a given patient



Scientific motivation: we want to calculate binding free energies



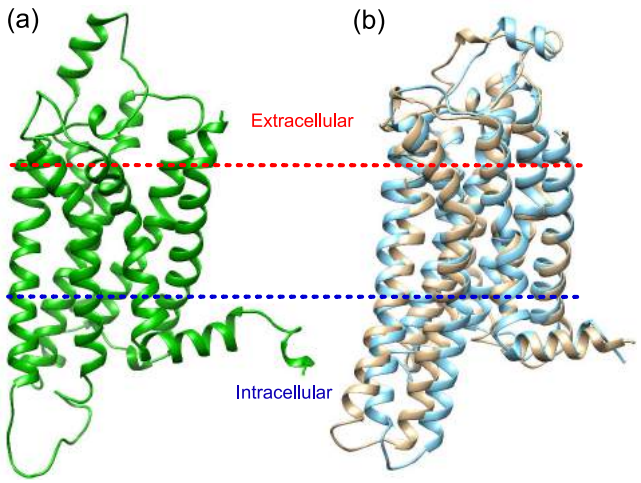
Applications:

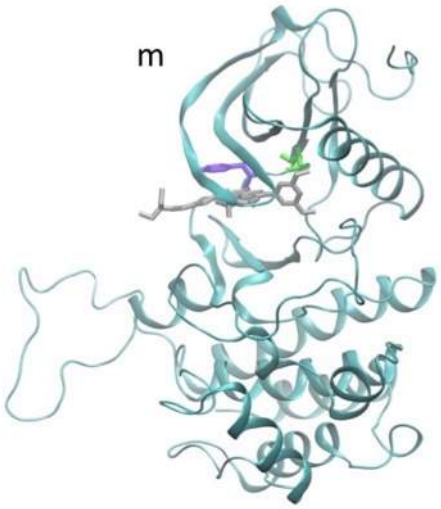
GPCR

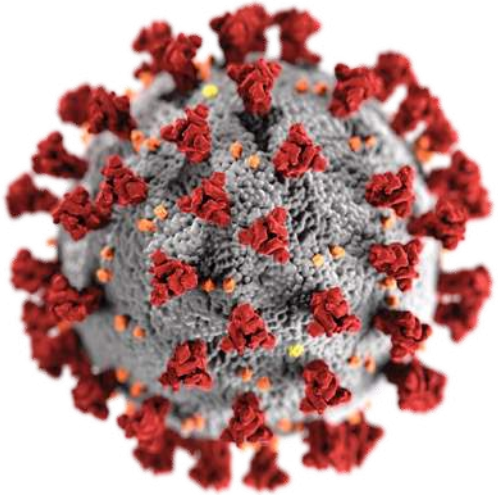
Kinases

Key proteins in SARS-Cov-2

And many more.

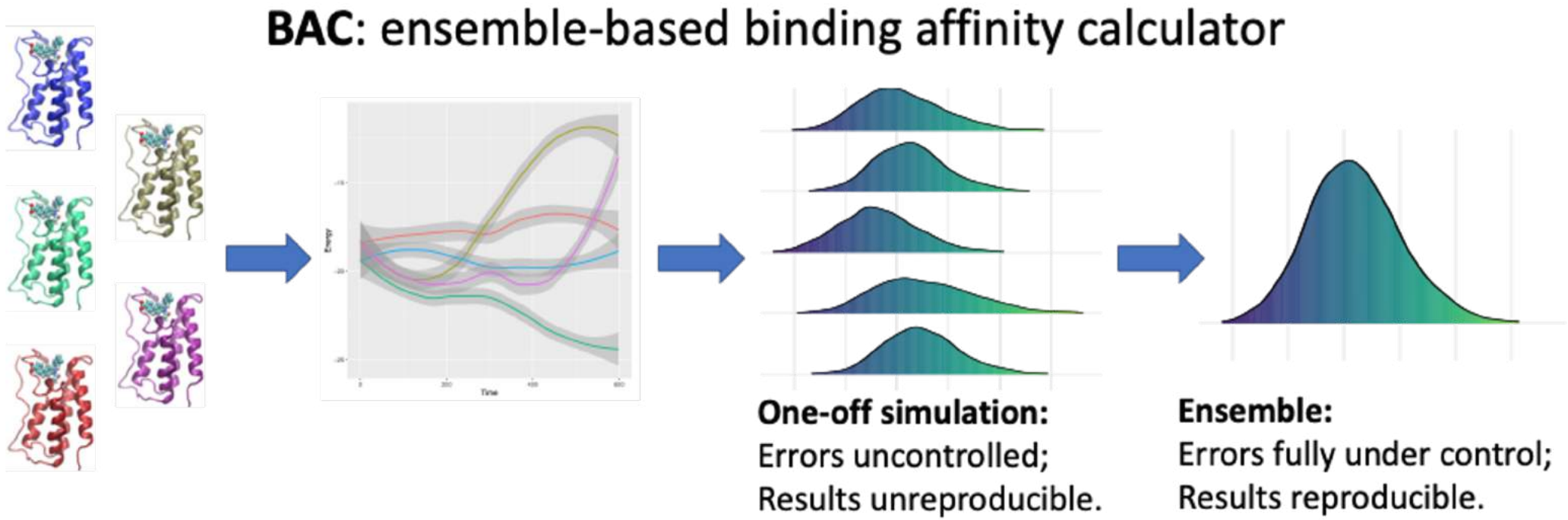






Methods: Ensemble simulations

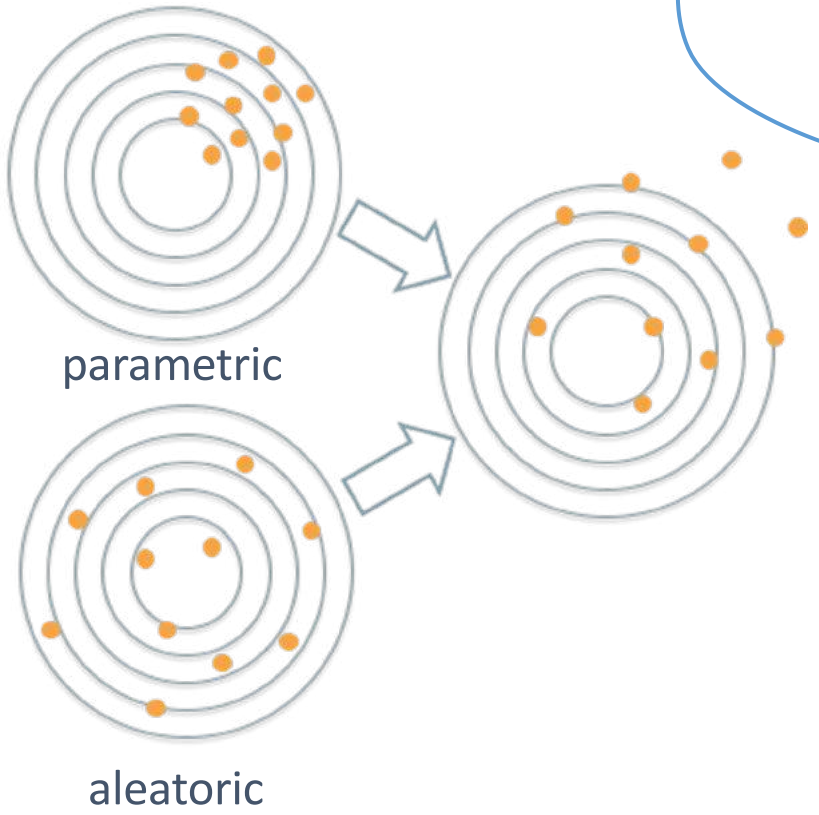
- Performing ensemble simulations and obtaining averages leads to more reliable results
- Ensemble averaging helps to eradicate aleatory errors on stochastic simulations, e.g. in drug affinity ranking

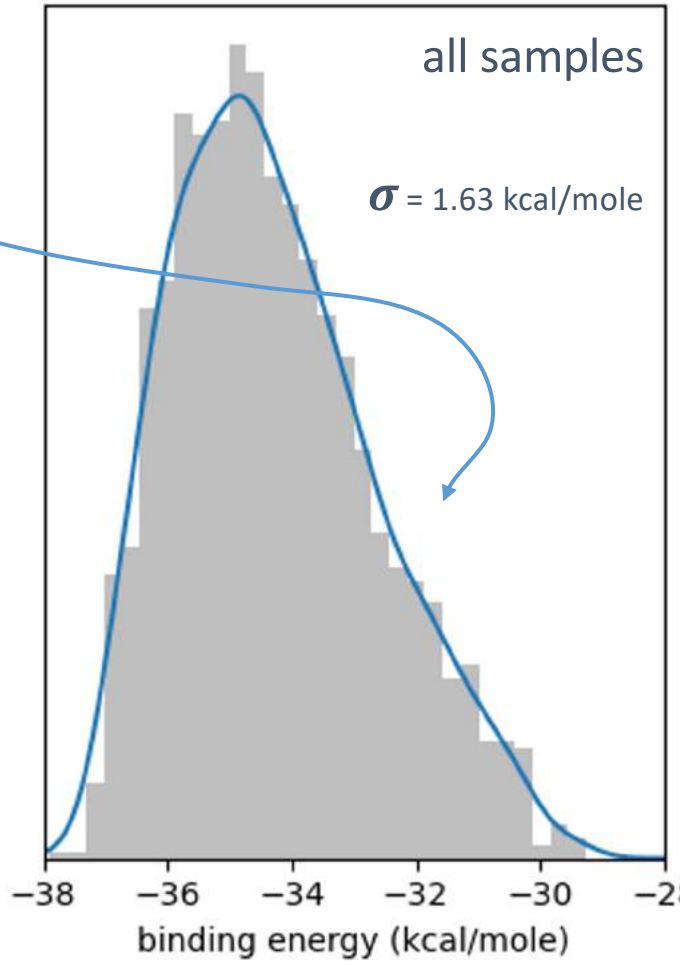


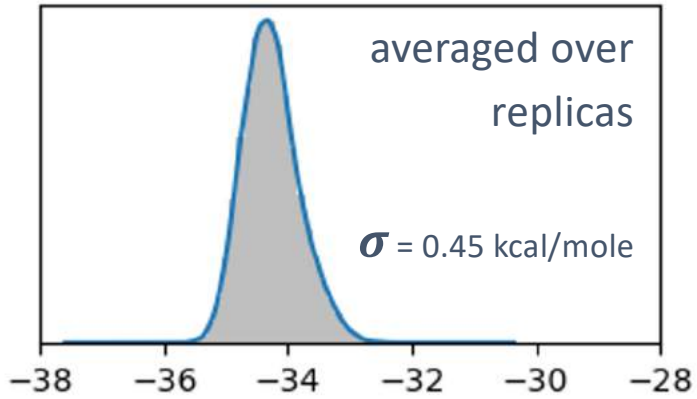
Vassaux et al., *J. Chem. Theory Comp.* (2021) 17 (8), 5187–5197

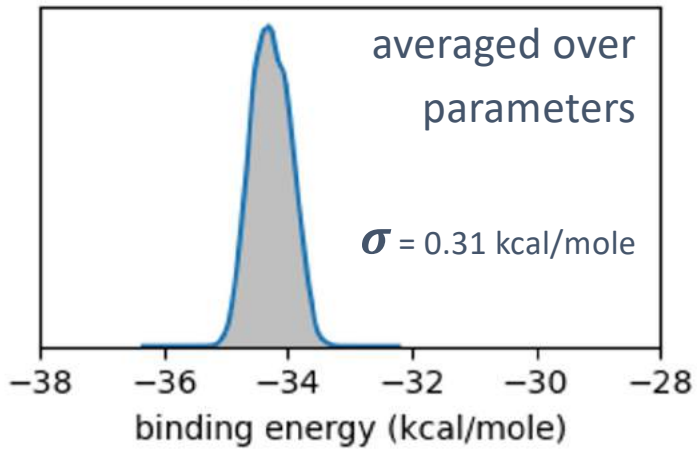
Uncertainty Quantification

- **distribution** of predicted free energies
- influence of ensemble and parametric **averaging**
- distributions often **non-normal**







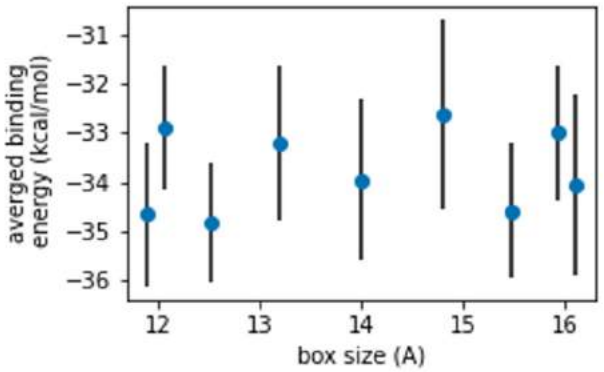


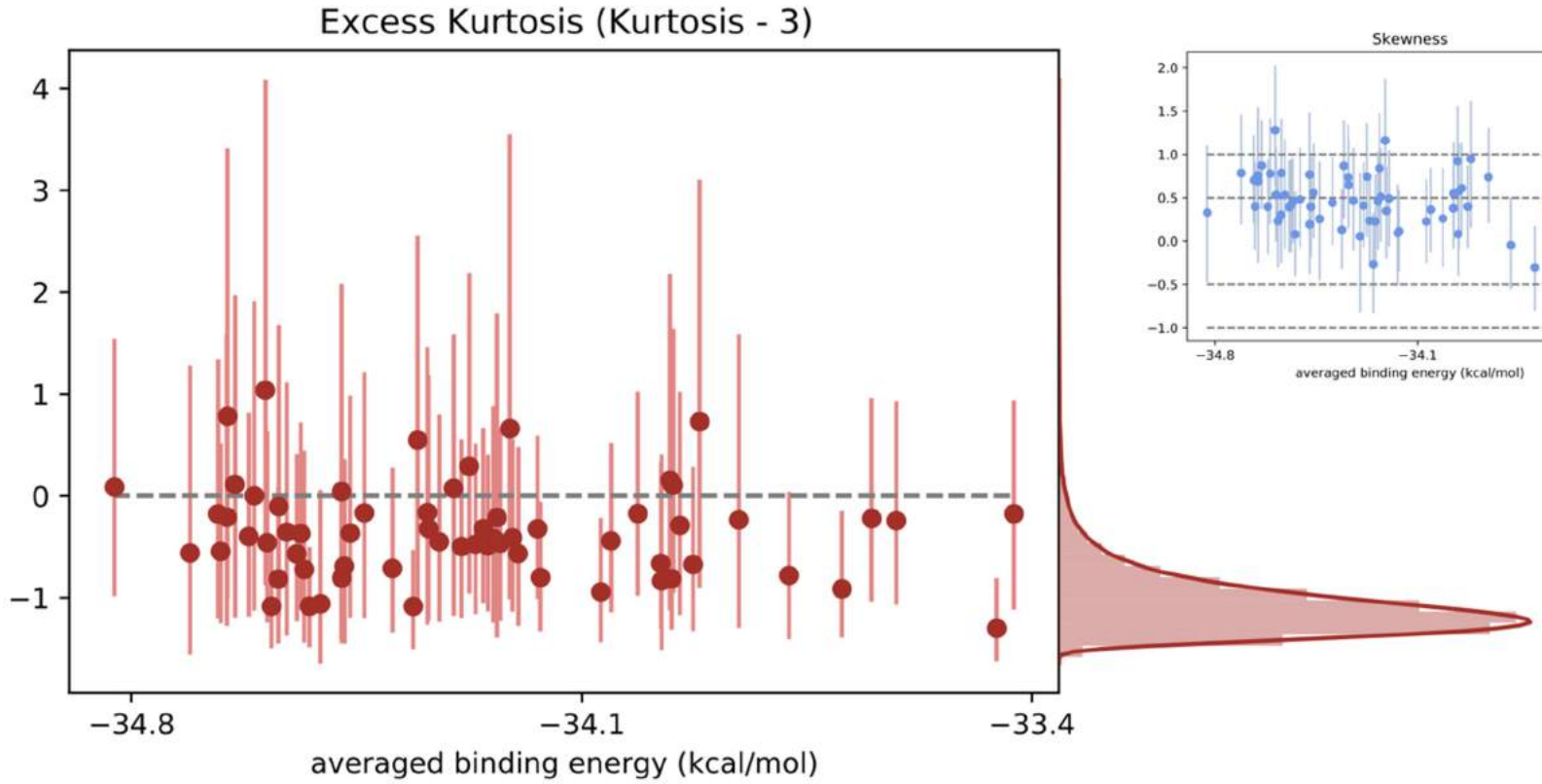
Vassaux et al., *J. Chem. Theory Comp.* (2021) 17 (8), 5187–5197

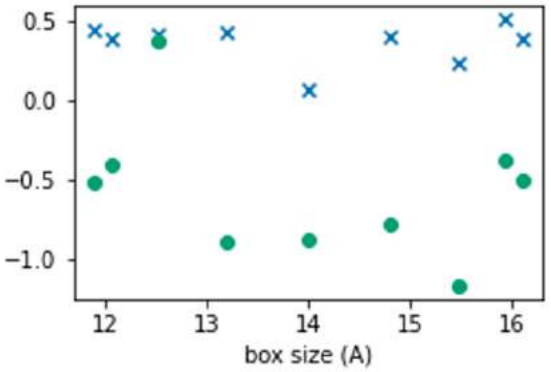
In-depth analysis of the loss of statistical normality

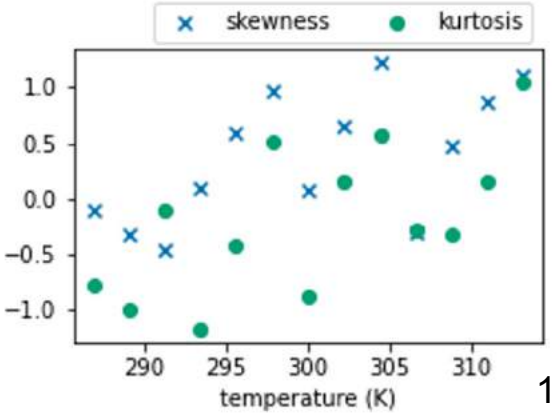
- analysis of **third and fourth statistical moments**
- **dependence of moments on influential parameters**

→ fine-sampling is essential: a **need for ensemble averaging**







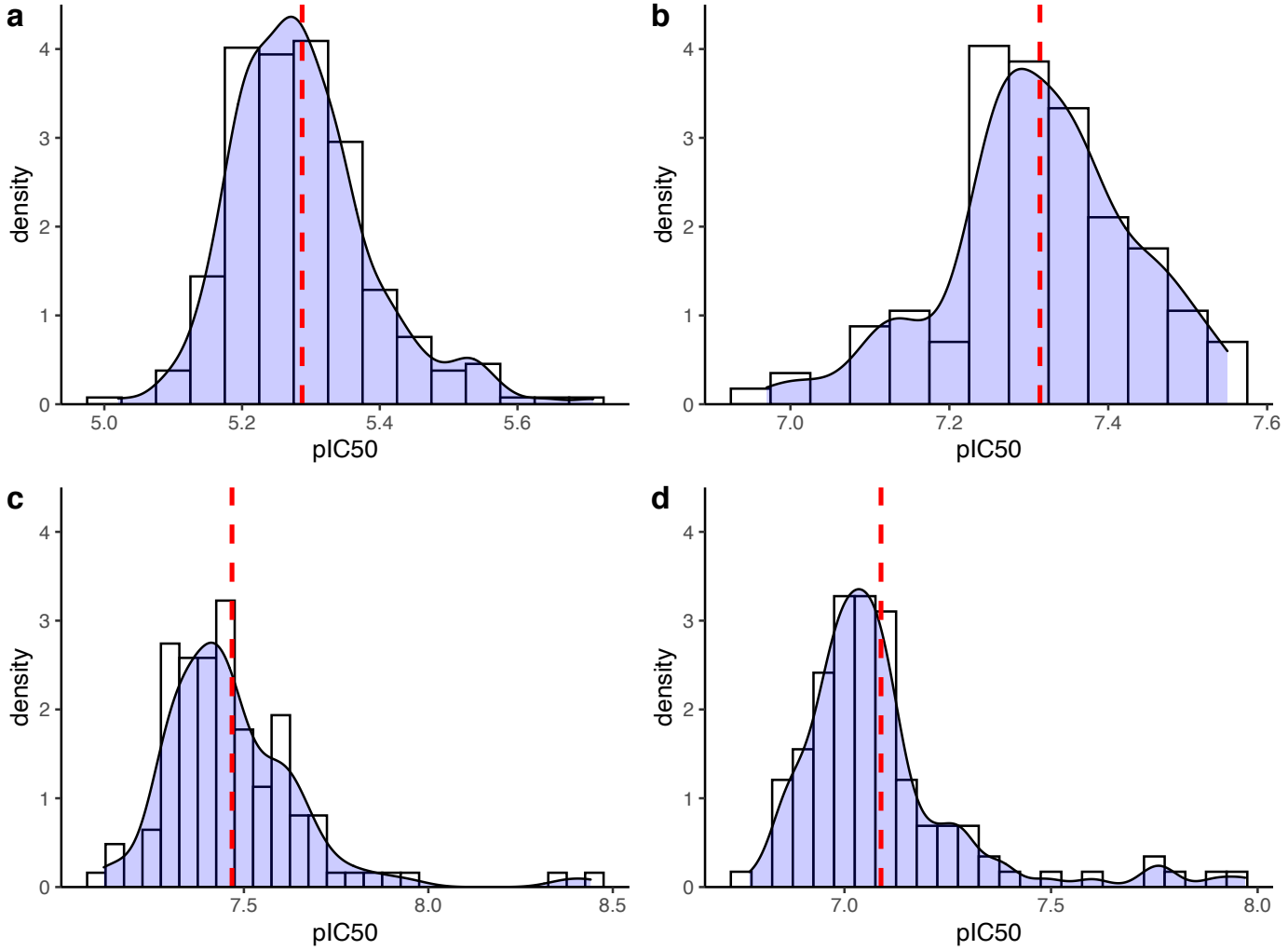


Vassaux et al., *J. Chem. Theory Comp.* (2021) 17 (8), 5187–5197

Non-normal distribution from experimental measurements

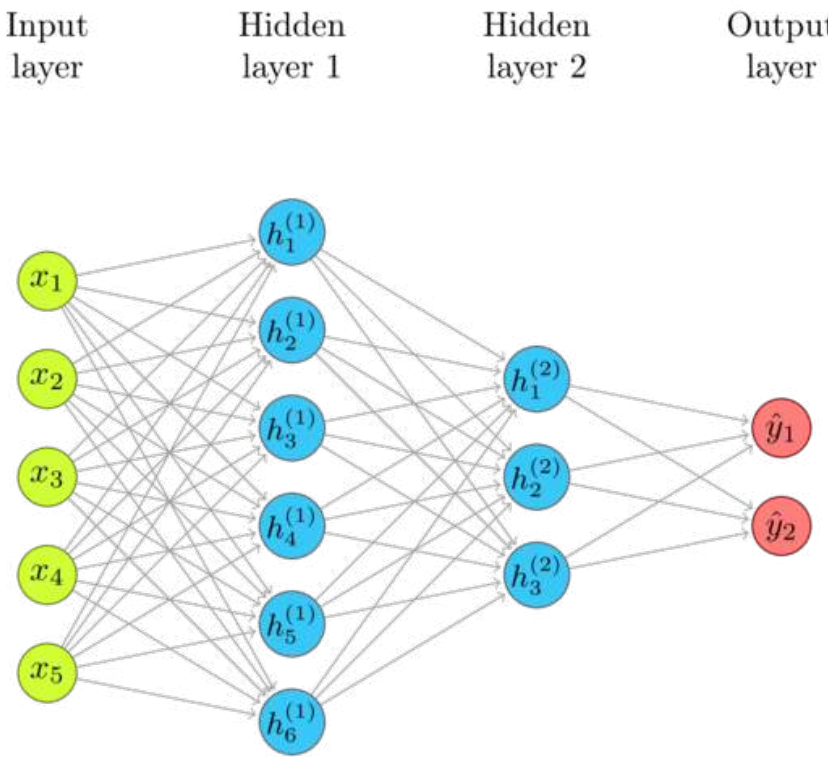
Experimental binding affinity measurements from GSK

- Compounds tested >100 times for their activities to SMYD3.
- Compounds **a** and **b** do not show any drift in the assay over time, compounds c and d show a small amount of time dependency.
- All distributions are skewed from a normal distribution
- The excess kurtoses are all positive, meaning that compared to a normal distribution, the tails are longer and heavier.

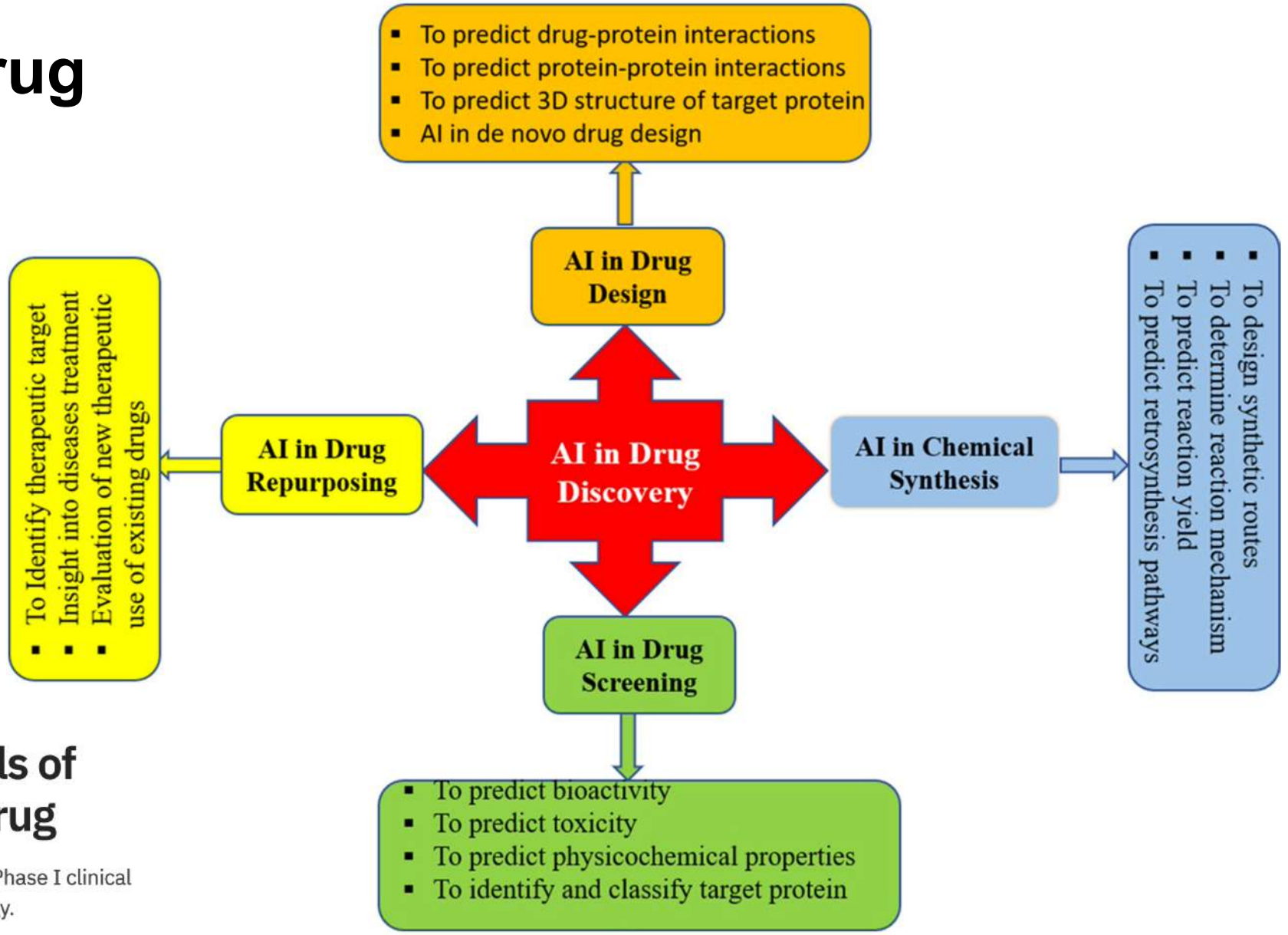


Wan, S. et al. J. Chem. Inf. Model., 62, 10, 2561 (2022) [DOI: 10.1021/acs.jcim.2c00255](https://doi.org/10.1021/acs.jcim.2c00255)

Machine Learning Approaches



What can AI do in drug discovery?



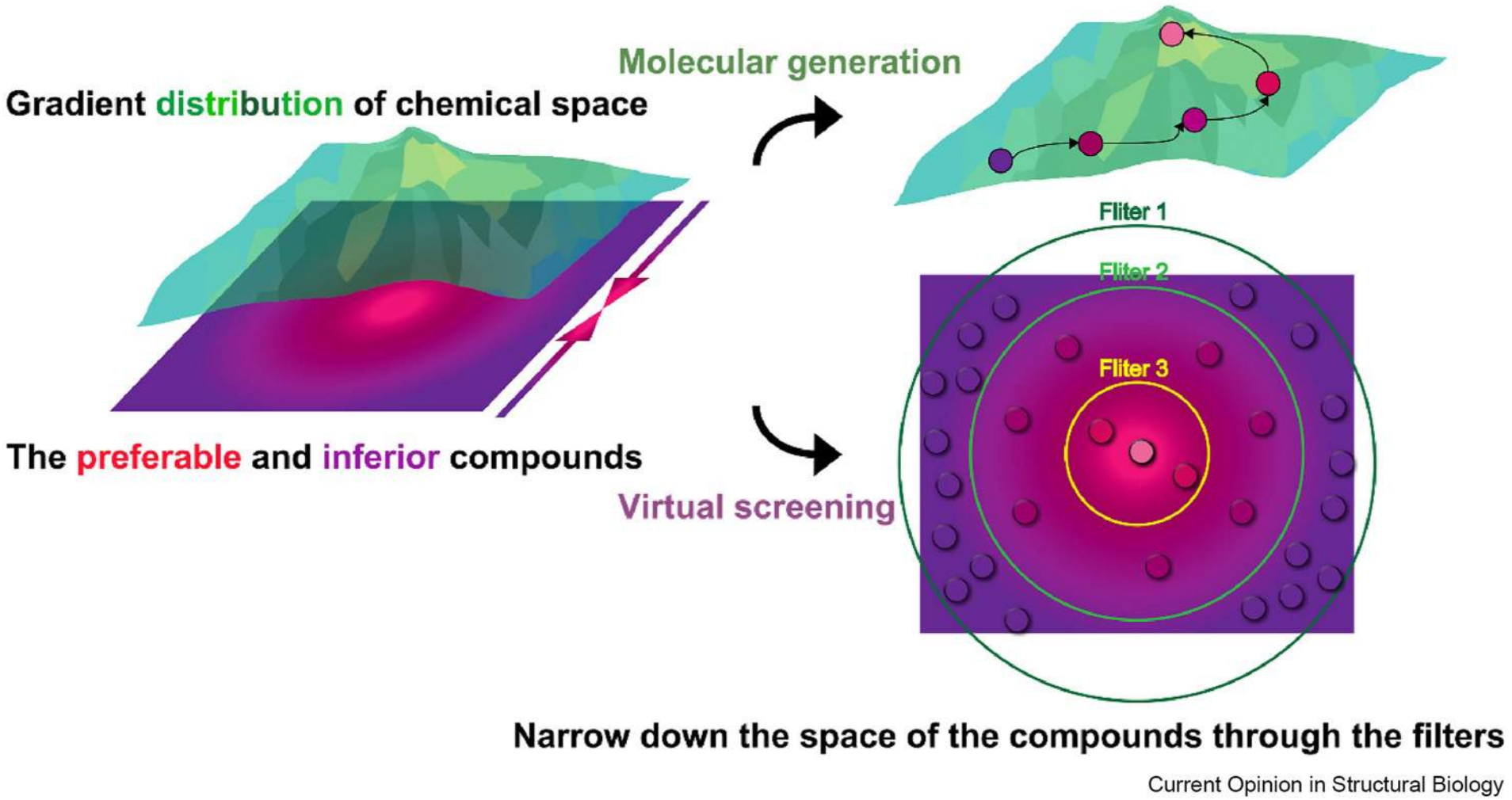
NEWS | April 9, 2021 | updated 16 Apr 2021 10:33am

Exscientia to start world-first trials of AI-designed immuno-oncology drug

Oxford-based AI drug discovery firm Exscientia is preparing to begin Phase I clinical trials with the world's first AI-designed molecule for immuno-oncology.

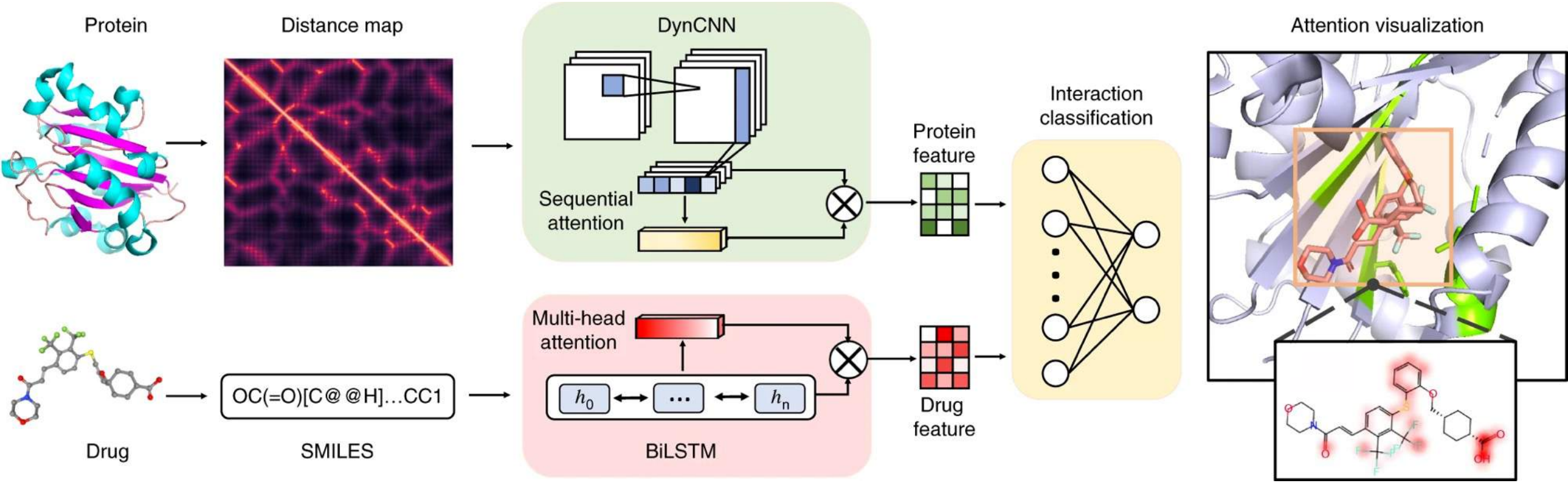
Tripathi, N., *et al.* Mol Divers **25,** 1643–1664 (2021). DOI: 10.1007/s11030-021-10237-z

De Novo Drug Design



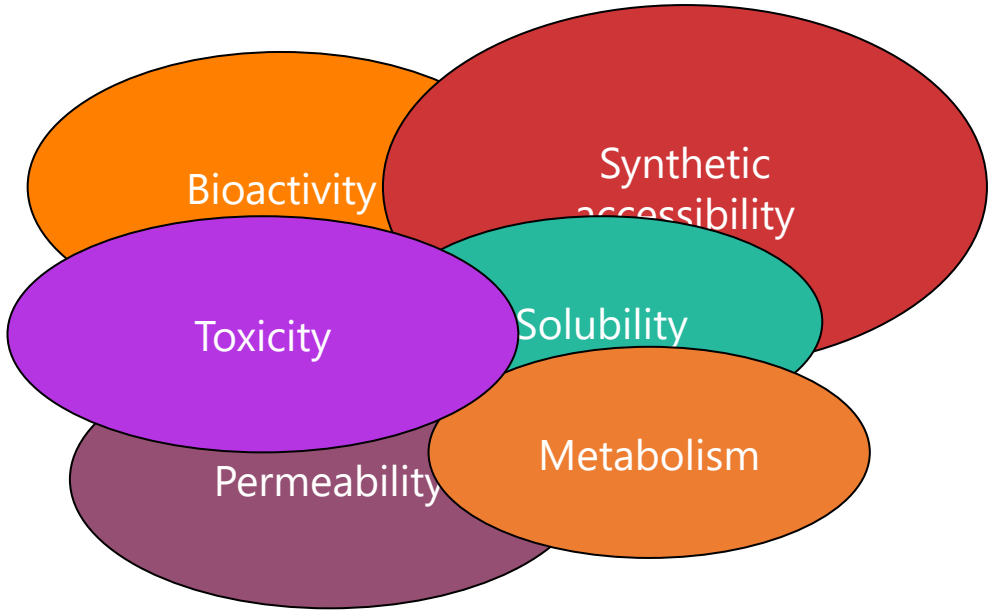
Wang, M., *et al.* Curr. Opin. Struct. Biol., **72**, 135-144 (2022). DOI: 10.1016/j.sbi.2021.10.001

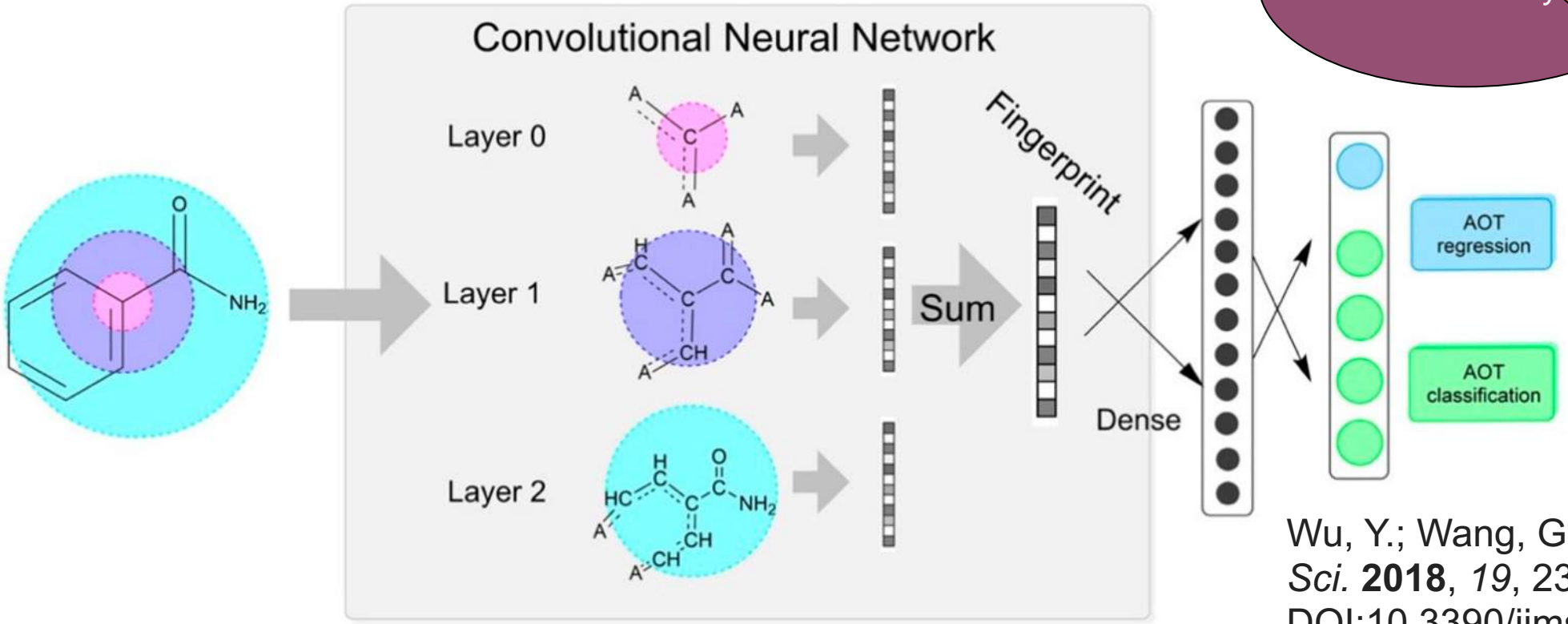
Prediction of Drug-Protein Interactions



Zheng, S., *et al*. Nat Mach Intell 2, 134–140 (2020). DOI: 10.1038/s42256-020-0152-y

Prediction of Bioactivity and Toxicity





Wu, Y.; Wang, G. *Int. J. Mol. Sci.* **2018**, *19*, 2358. DOI:10.3390/ijms19082358

DeepDriveMD: Application to ligand-protein systems

- DDMD is an autoencoder based method which was initially developed for accelerating the rate of protein folding simulations.

- Autoencoders:
  - Deep convolution variational autoencoder (CVAE)
  - Adversarial autoencoder (AAE)

- We are seeking to modify its functionality to extend it to the accelerated, enhanced sampling of ligand binding poses within protein-ligand complexes.

- Accelerated identification of optimal ligand binding poses should increase the precision and accuracy of downstream free energy calculations in the IMPECCABLE workflow, at a reduced computational cost.

Bhati et al. *J.R. Soc. Interface Focus* (2021) 11(6), 20210018
Brace et al. *arXiv preprint* (2021) *2104.04797*

Protein structure prediction

How AlphaFold works:

- Formulate the protein folding problem as a “spatial graph” showing the amino acids localised in space
- Design the structure of the network to learn from training which amino acids lie near each other

In some cases, AlphaFold’s structure predictions are indistinguishable from those determined using experimental methods

News in focus

A protein's function is determined by its 3D shape.

‘IT WILL CHANGE EVERYTHING’: AI MAKES GIGANTIC LEAP IN SOLVING PROTEIN STRUCTURES

DeepMind’s program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Nature | Vol 588 | 10 December 2020 | 203

E. Callaway (2020), It will change everything: AI makes gigantic leap in solving protein structures, Nature., **588**, 203-204, *doi:10.1038/d41586-020-03348-4*

Limitations

Assumption: The distribution of training data is *identical* to the distribution of test data (including data from future unseen events).

In practice, this assumption is often violated to because of:

- insufficient training data
- noise and errors in real-life data
- skewed distributions
- Etc., etc.

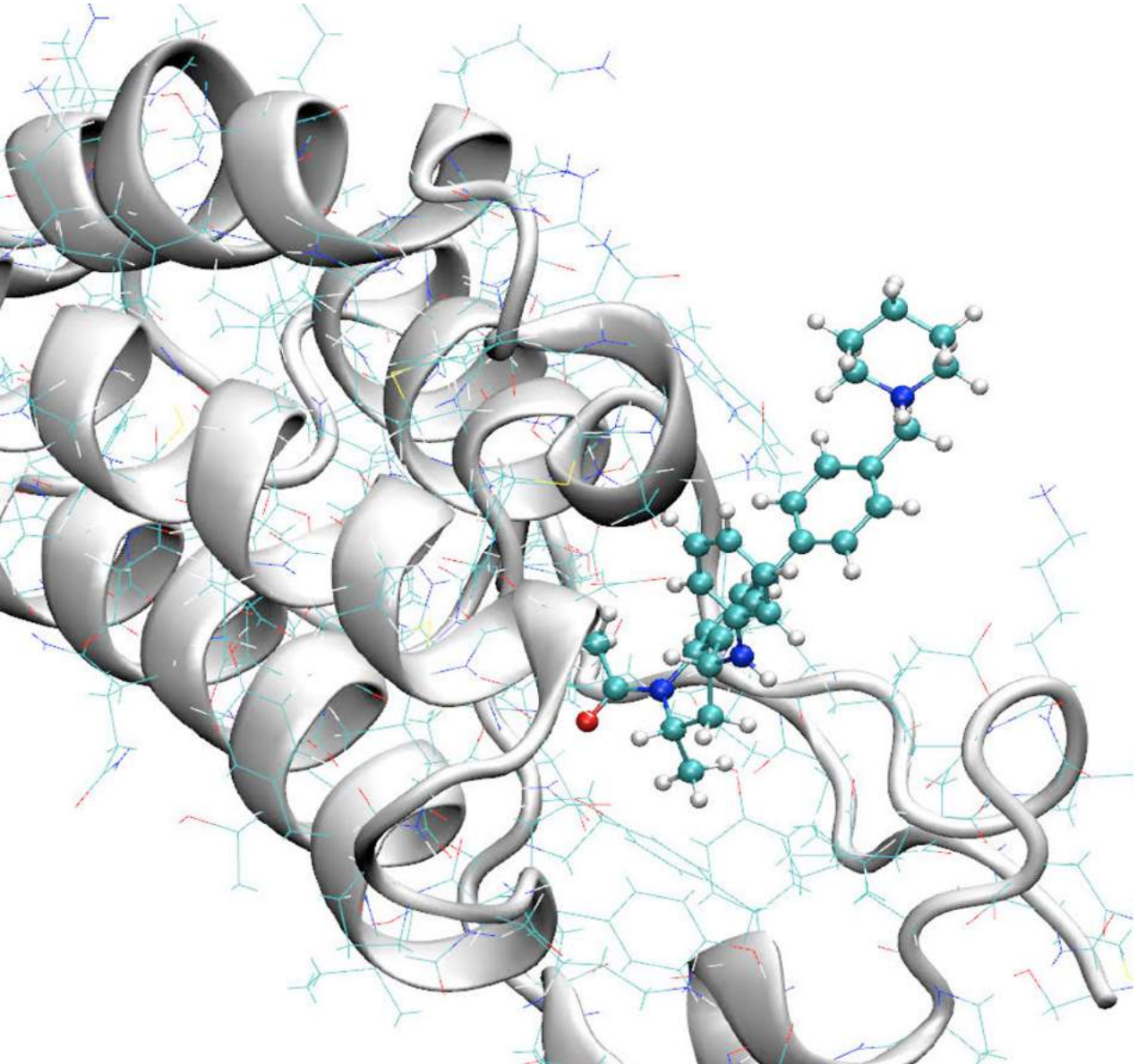
Interpretability:

- Fail to provide conceptual accounts for the processes to which they are applied
- Lack explanatory power and thus no understanding and insight.
- It uses 10^8 parameters to fit the data!

It is vital to use theory as a guide to experimental design for maximal efficiency of data collection and to produce reliable predictive models and conceptual knowledge.

P. V. Coveney & R. Highfield, “Big AI: Blending Big Data With Big Theory to Build Virtual Humans”, in “AI for Science”, Tony Hey *et al*, Editors (forthcoming)

P. V. Coveney, E. R. Dougherty and R. R. Highfield, "Big Data Need Big Theory Too”, *Phil Trans R Soc London A*, 374, 20160153 (2016)



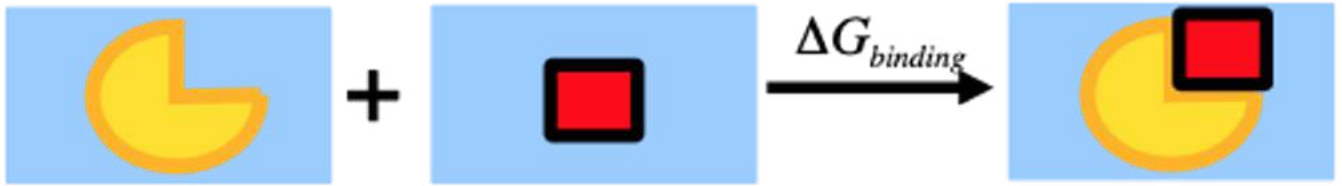
Physics-based Approaches

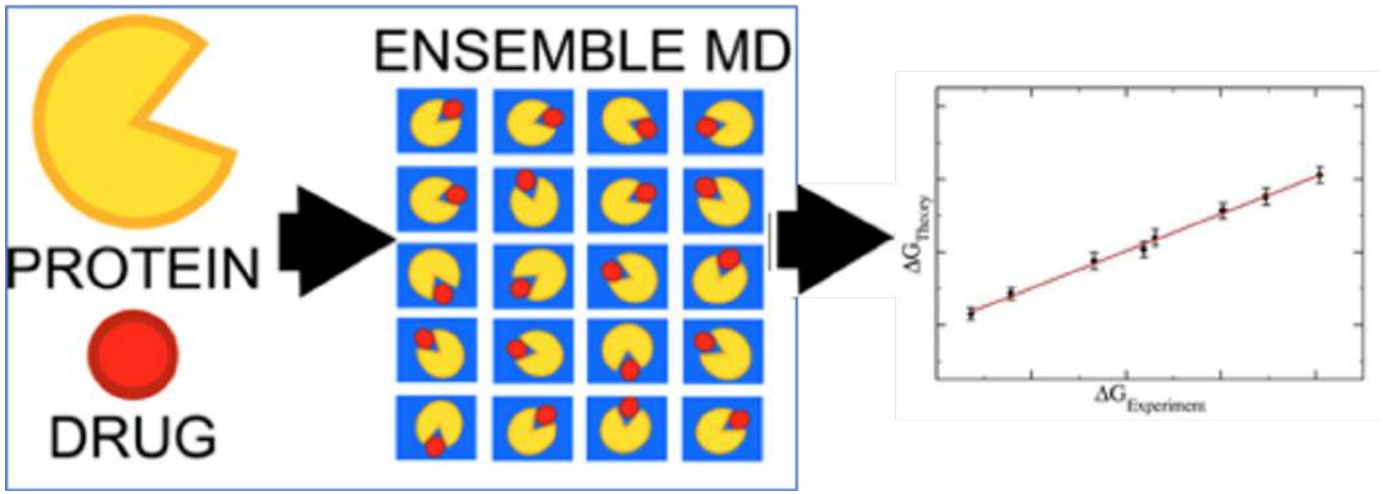
Methods: End-point approach

ESMACS: enhanced sampling of molecular dynamics with approximation of continuum solvent

Ranking binding affinities:

- Evaluate large number of promising compounds
- Structurally and chemically diverse compounds
- Ranking of binding free energies
- Ensemble simulation for reliable predictions
- Lower computational cost

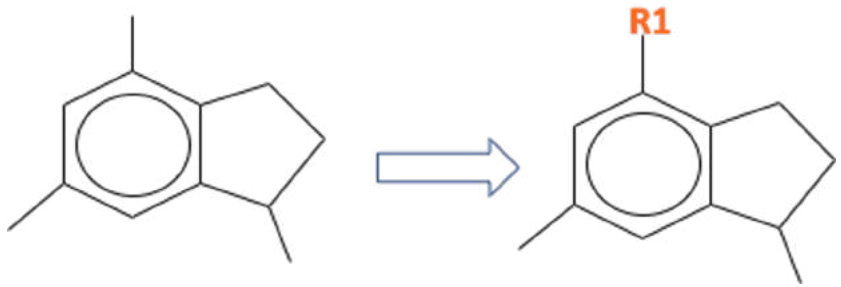




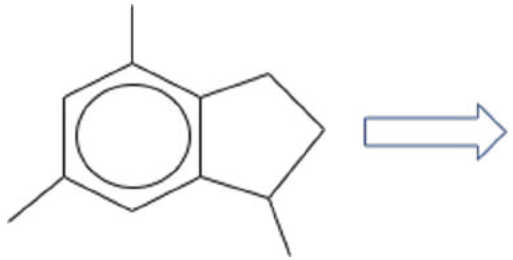
Methods: Alchemical approach

TIES: Thermodynamic integration with enhanced sampling

Relative binding free energy



Absolute binding free energy



TIES20: a software for Flexible Superimposition and Partial Ring Morphing

https://ccs-ties.org/

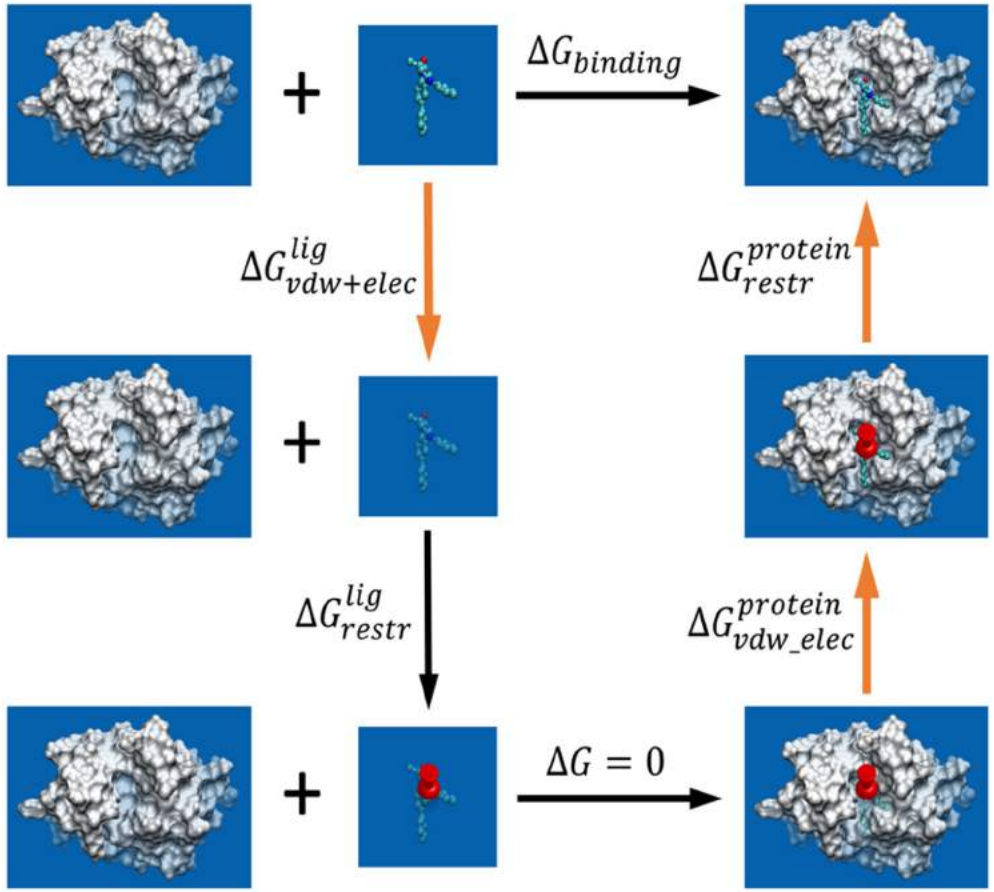
TIES MD: a collection of software packages to calculate protein ligand binding free energies with physics based alchemical methods.

https://ucl-ccs.github.io/TIES_MD/

A. Bhati, et al., J. Chem. Theory Comput. (2017), 13(1), 210–222.
M. Bieniek, et al.,, J. Chem. Theory Comput., (2021) 17(2), 1250–1265.

Absolute binding free energy

- ABFE is capable of comparison of binding affinities of **structurally and chemically unrelated compounds**.
- A thermodynamic cycle is employed, in which the binding process is divided into a series of nonphysical transformations.
- The binding free energy, $\Delta G_{binding}$, is the sum of all ΔG values from the nonphysical step.
- Ensemble simulations, with relatively large number of replicas, are required to attain the desired precision.

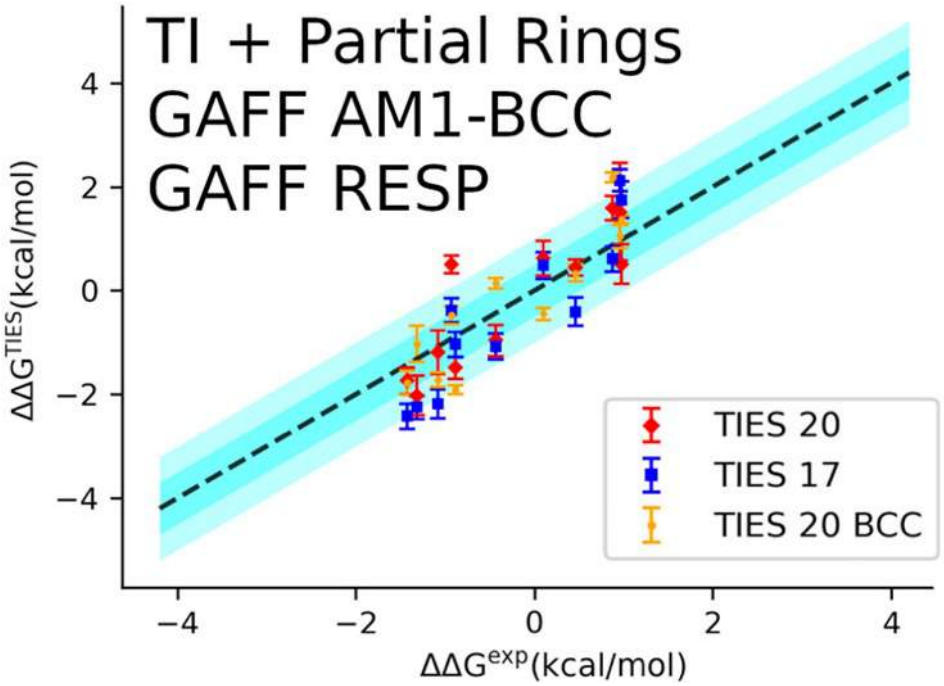


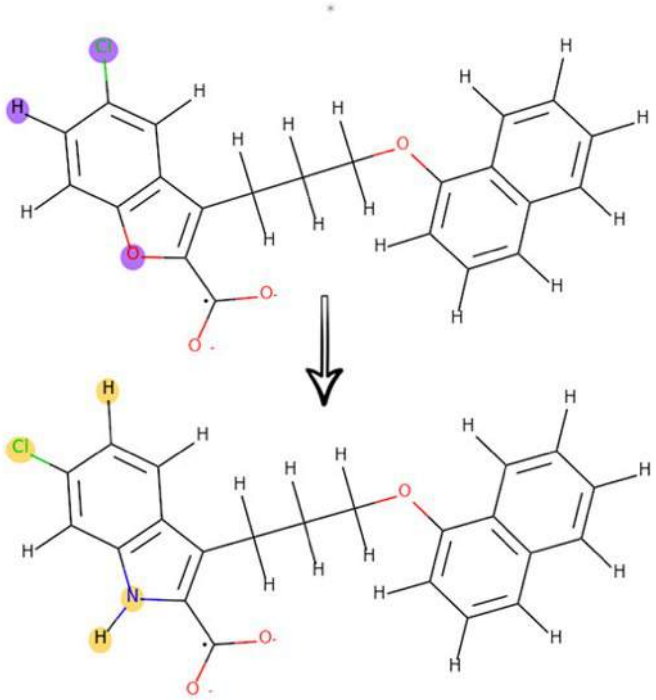
A. Bhati, S. Wan, Y. Hu, B. Sherborne, P. V. Coveney, *J. Chem. Theory Comput.* (2018) 14 (6), 2867–2880.

TIES 20

Relative Binding Free Energy with a Flexible Superimposition Algorithm and Partial Ring Morphing.

TIES 20 implements a flexible topology superimposition algorithm to match drugs and build inputs for relative binding affinity calculations.



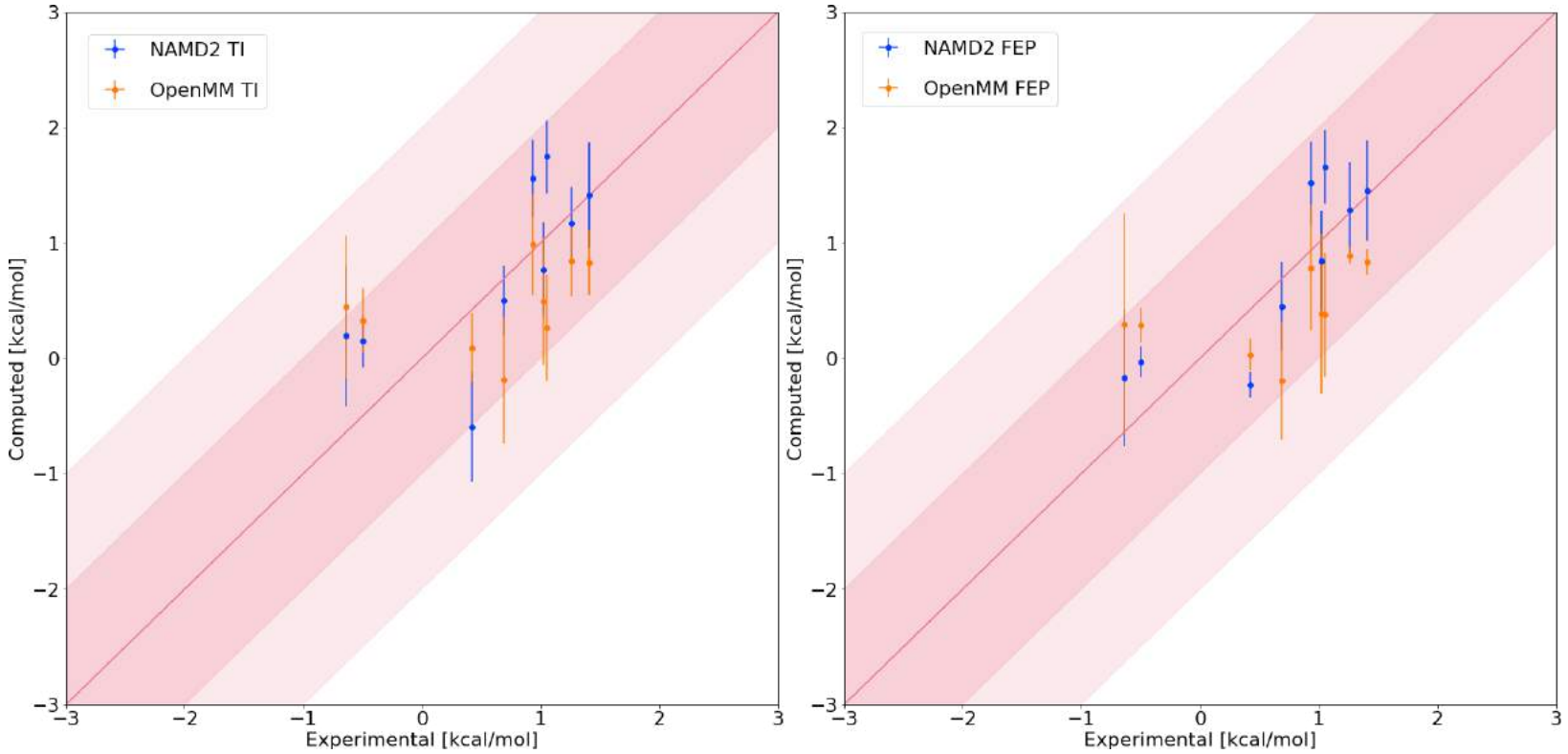


M. K. Bieniek, A. P. Bhati, S. Wan, and P. V. Coveney, J. Chem. Theory Comput. 2021, 17, 2, 1250–1265. DOI:10.1021/acs.jctc.0c01179

TI *vs* FEP

Using ensemble-based simulations

- No statistically significant difference found between the calculated properties for TI and FEP results.
- Both TI and FEP methods achieve comparable accuracy and precision for the systems studied.

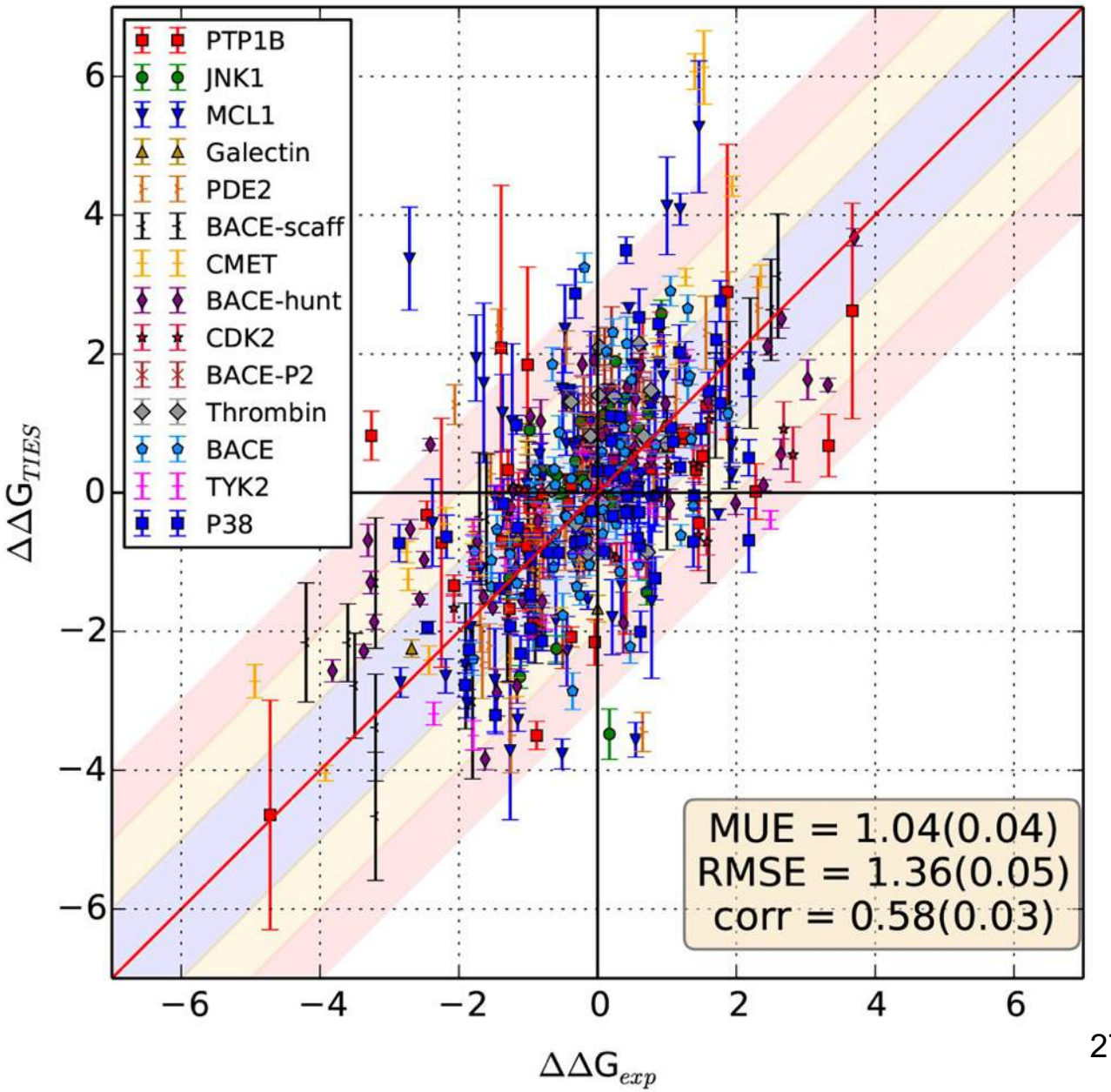


A. Wade, A. Bhati, S. Wan, P. Coveney. *J. Chem. Theory Comput.*, 18, 6, 3972 (2022) DOI: 10.1021/acs.jctc.2c00114

Large Scale Study of Ligand–Protein Relative Binding Free Energy Calculations

Actionable Predictions from Statistically Robust Protocols

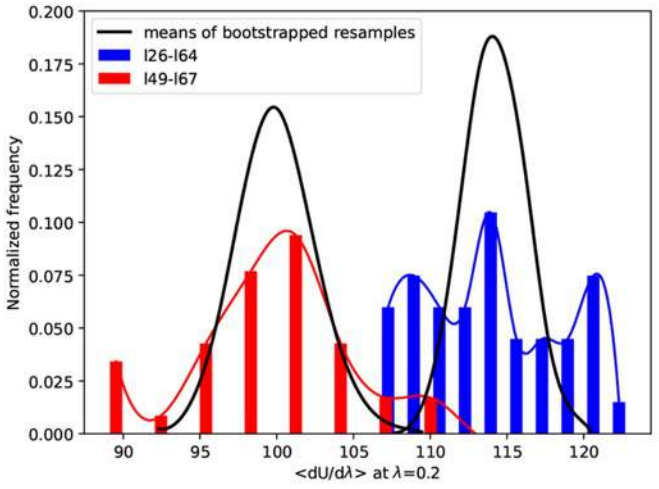
A large data set comprising over 500 ligand transformations spanning over 300 ligands binding to a diverse set of 14 different protein targets.

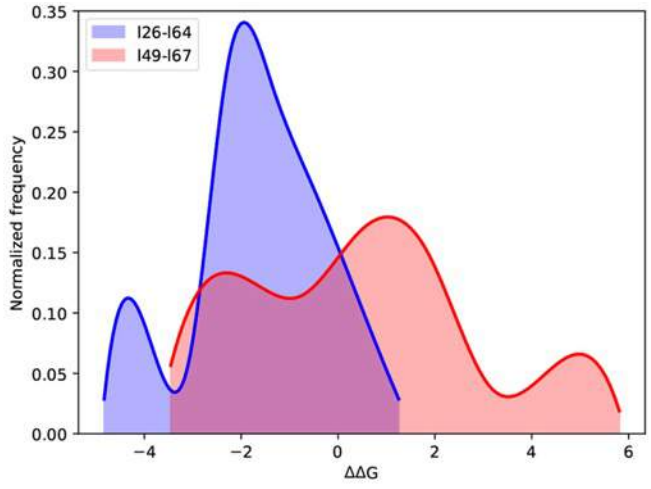


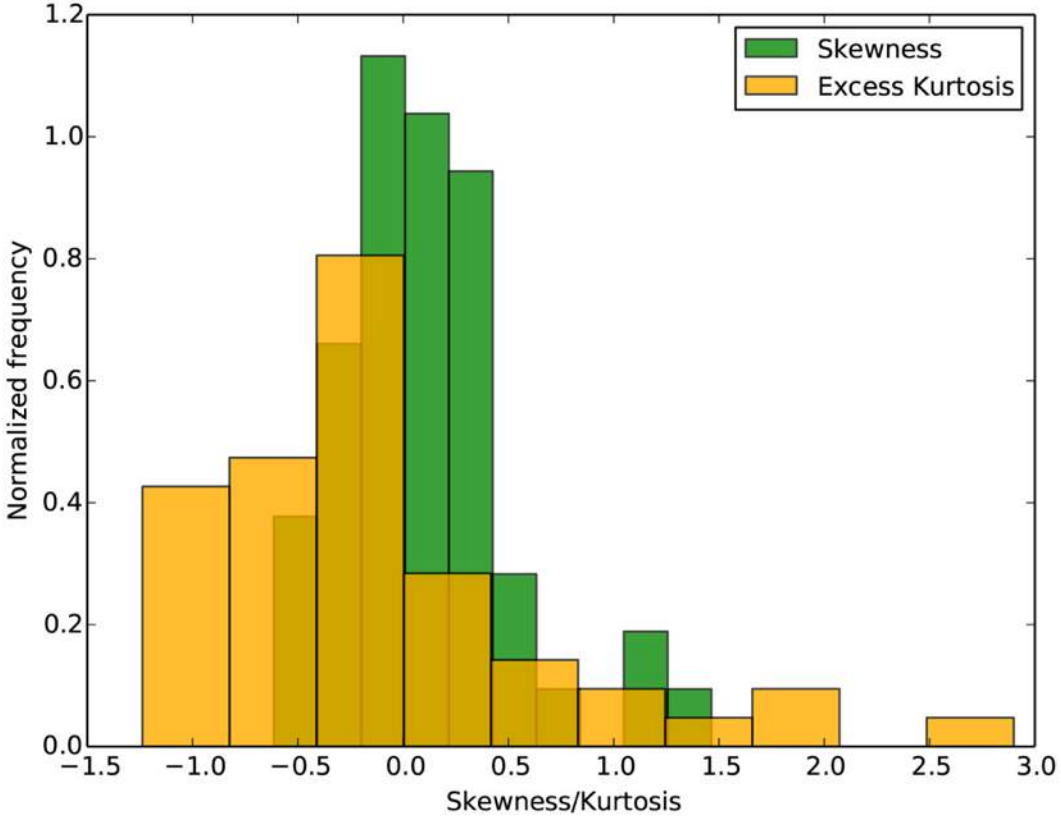
A. P. Bhati and P. V. Coveney, J. Chem. Theory Comput. 2022, 18, 4, 2687–2702. DOI:10.1021/acs.jctc.1c01288

Ensemble simulation-based alchemical approach

Ensemble simulations furnish information on the statistical distributions of the free energy calculations which exhibit non-normal behaviour.



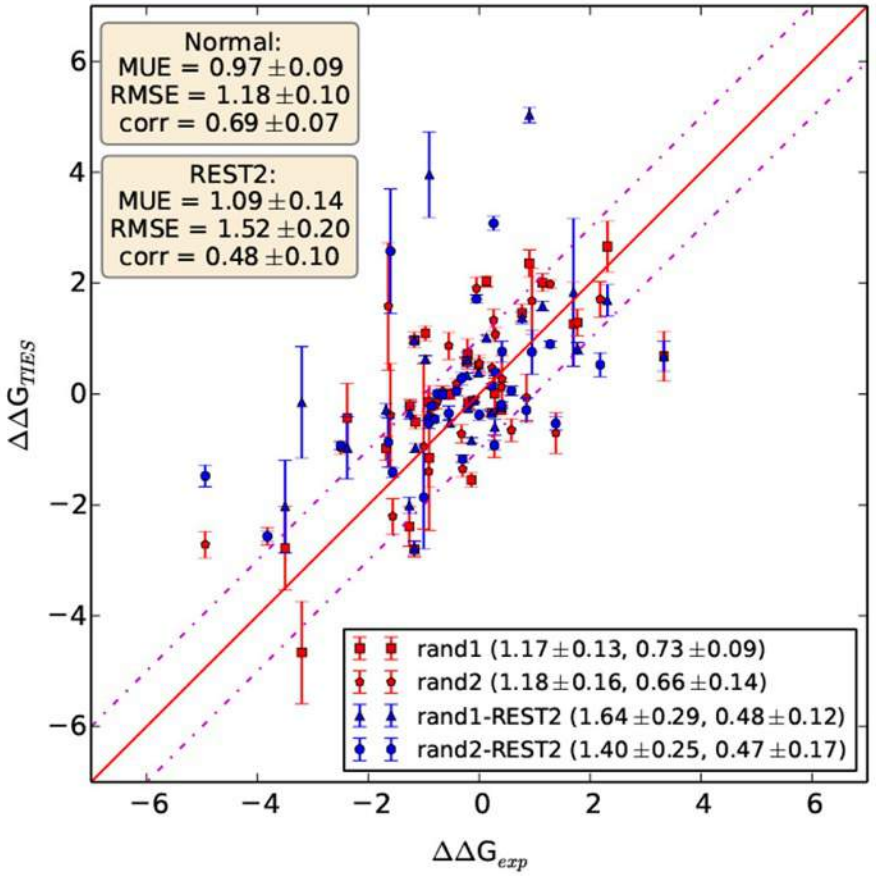


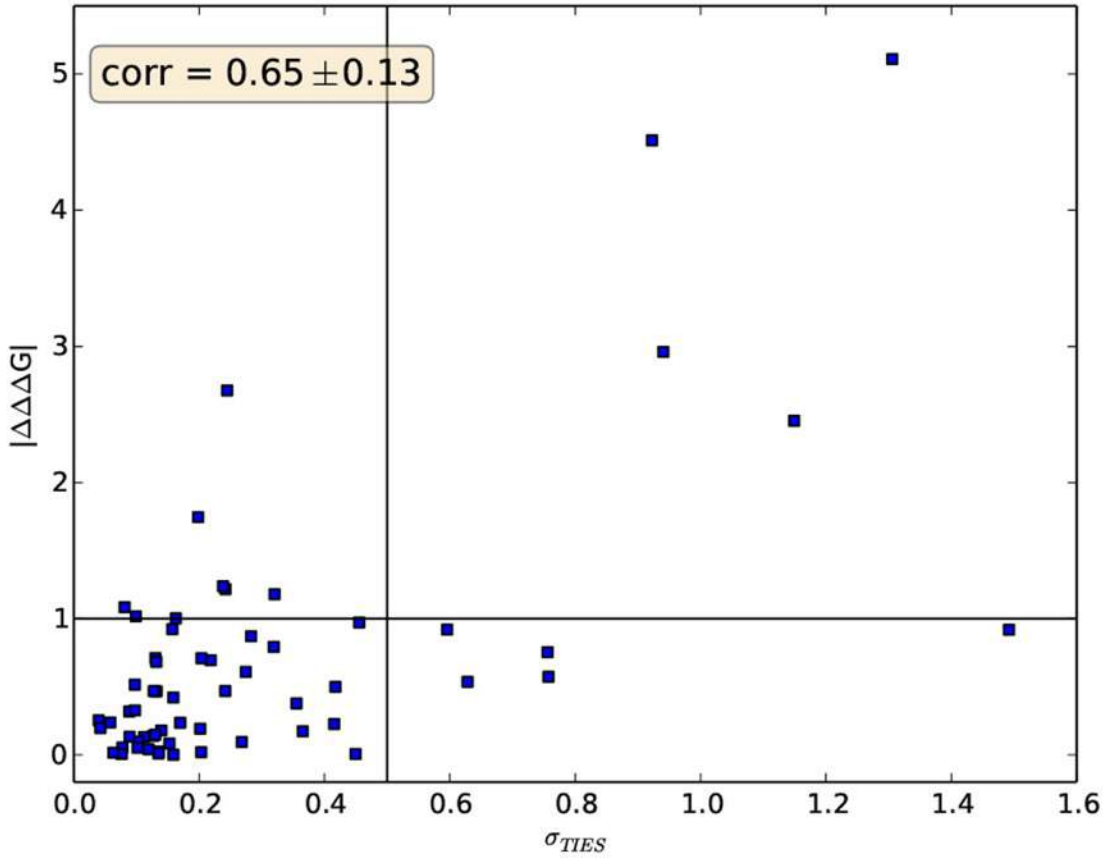


A. P. Bhati and P. V. Coveney, *J. Chem. Theory Comput.* 2022, 18, 4, 2687–2702. DOI:10.1021/acs.jctc.1c01288

Enhanced sampling method

Replica exchange with solute tempering (REST2) protocol degrades the accuracy of results, and hence its routine application is not recommended.





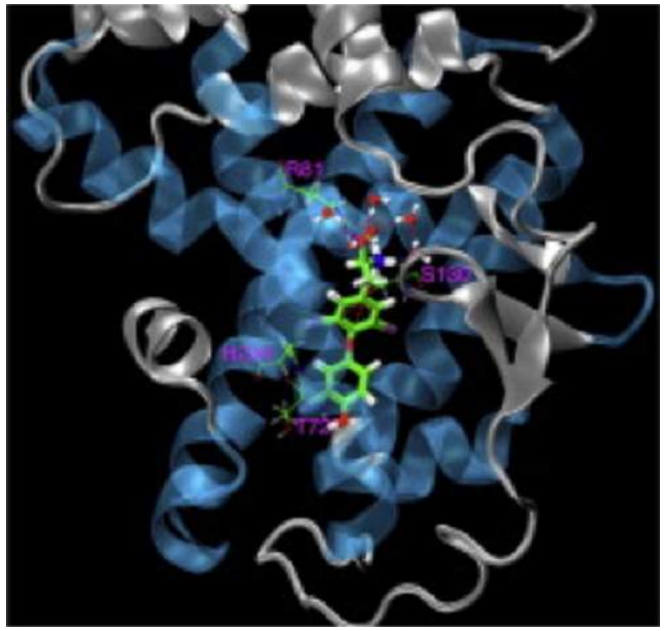
A. P. Bhati and P. V. Coveney, J. Chem. Theory Comput. 2022, 18, 4, 2687–2702. DOI:10.1021/acs.jctc.1c01288

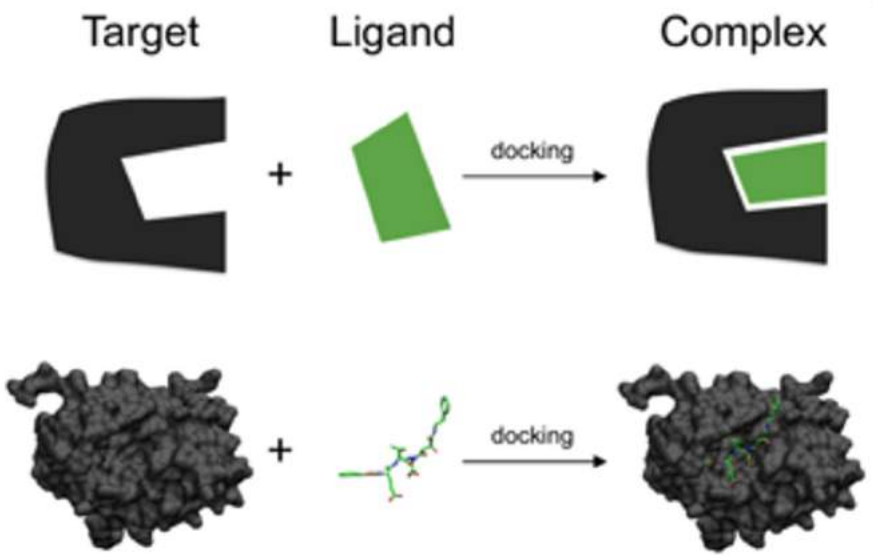
BAC: Molecular simulations of receptor-ligand free energy

- computer-aided drug discovery
- prediction of the dynamics of 100,000s of atoms using Newton's equations of motion on the timescale of nanoseconds
  - → evaluate the free energy of the system using time-averaging and ensemble averaging

- computational cost:
  - 1,000 core hours per replica
  - ensembles of replicas for a robust free energy evaluation

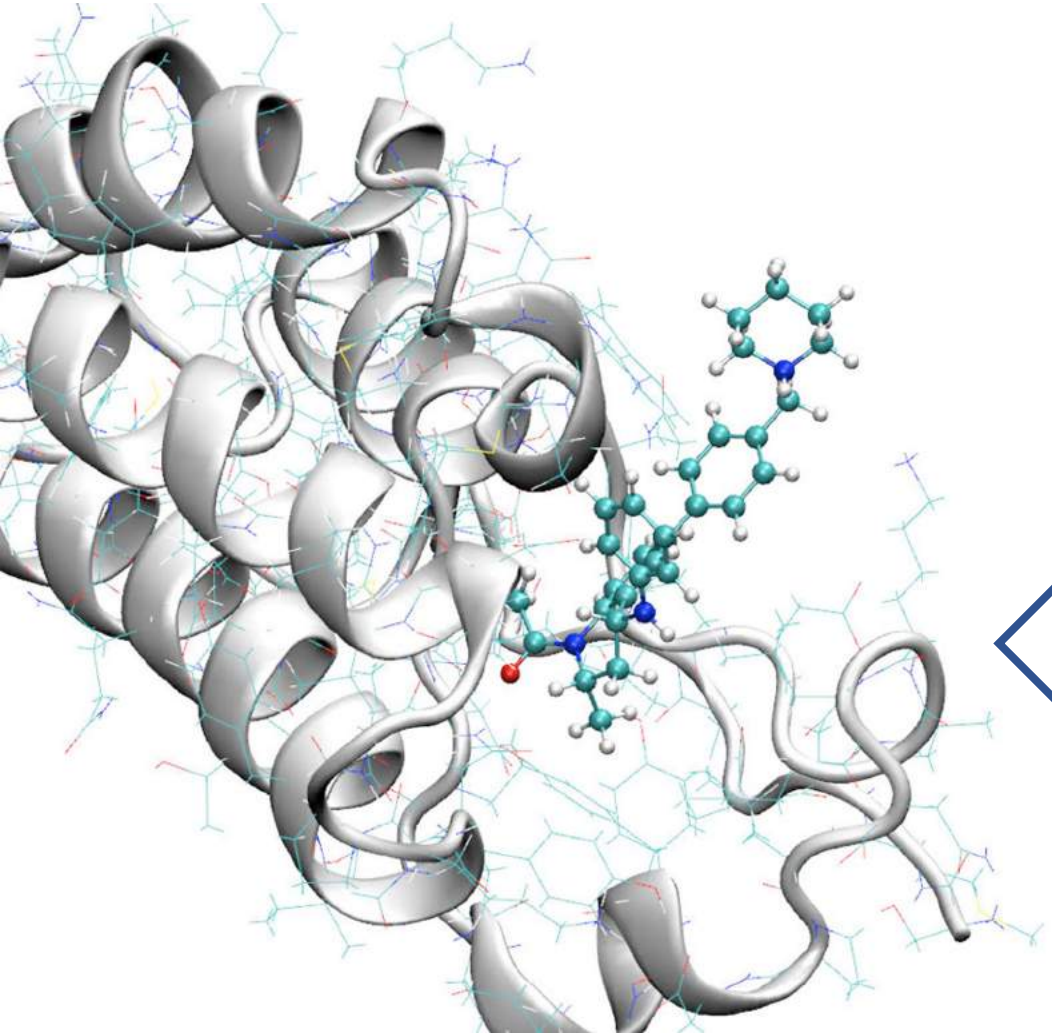
- Approach is being integrated with ML

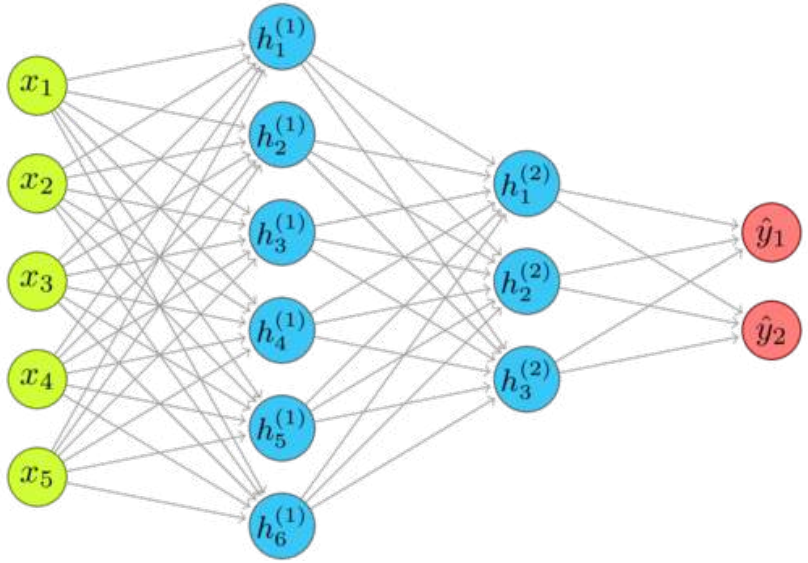




Wright et al., *Sci. Rep.* (2019) 9, 6017

Hybrid physics-based and machine-learning approach





Hybrid physics-based and machine-learning approach

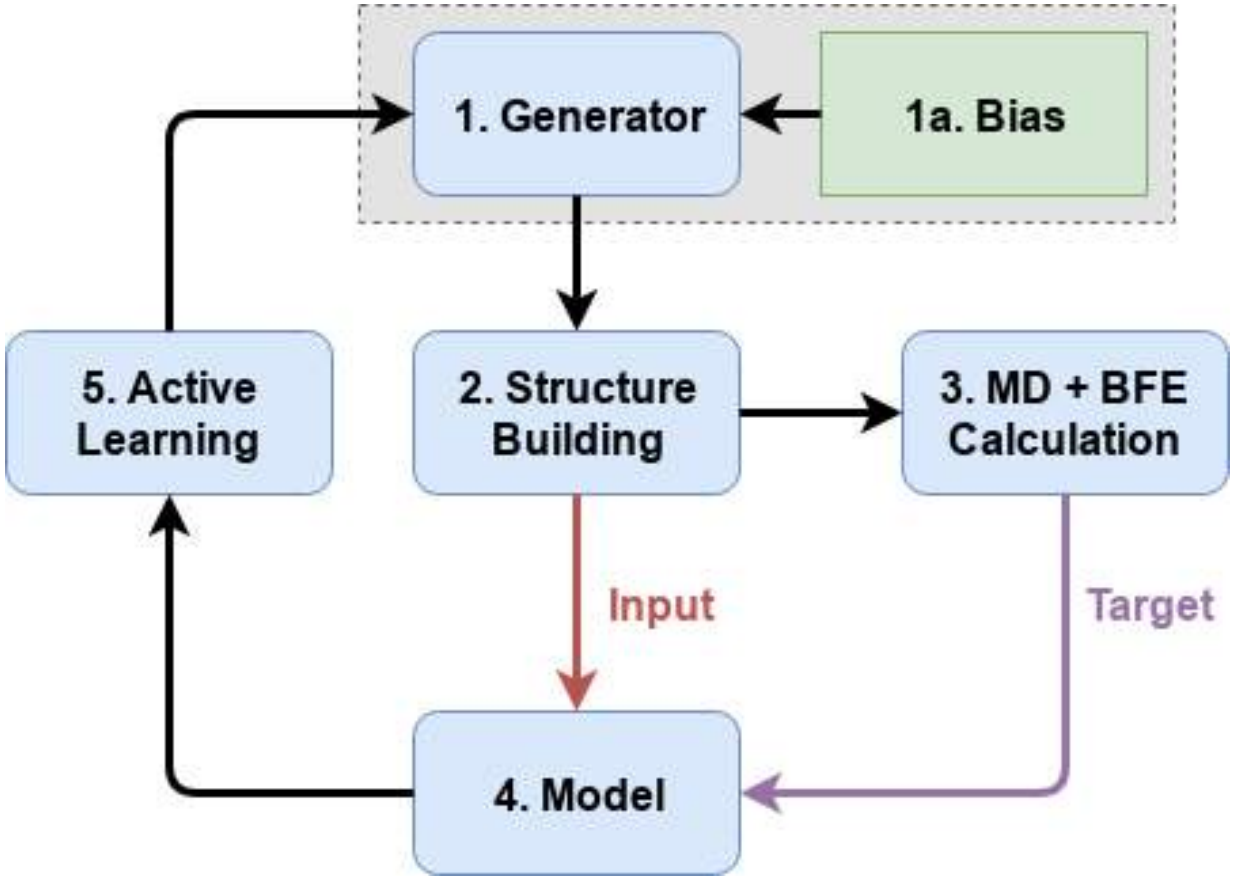
- **Molecular Dynamics** simulations
  - Accurate binding affinity estimation
  - Less accurate methods for virtual screening
  - More accurate methods for lead optimization
  - Computationally expensive and hence limited exploration of chemical space

- **Machine Learning** algorithms
  - Very fast in screening huge libraries of molecules
  - Can explore the chemical space of relevance much widely
  - Generative algorithm predicts useful molecules
  - Heavily depending on training data

Why couple ML and MD?

Molecular dynamics (MD) and **machine learning (ML)** methods are complementary to each other

- MD can be used to screen predictions from ML algorithms
- MD output can be fed back into ML algorithm to train and improve it further
- Quick predictions of potential binders
- Augmenting human intelligence with artificial intelligence



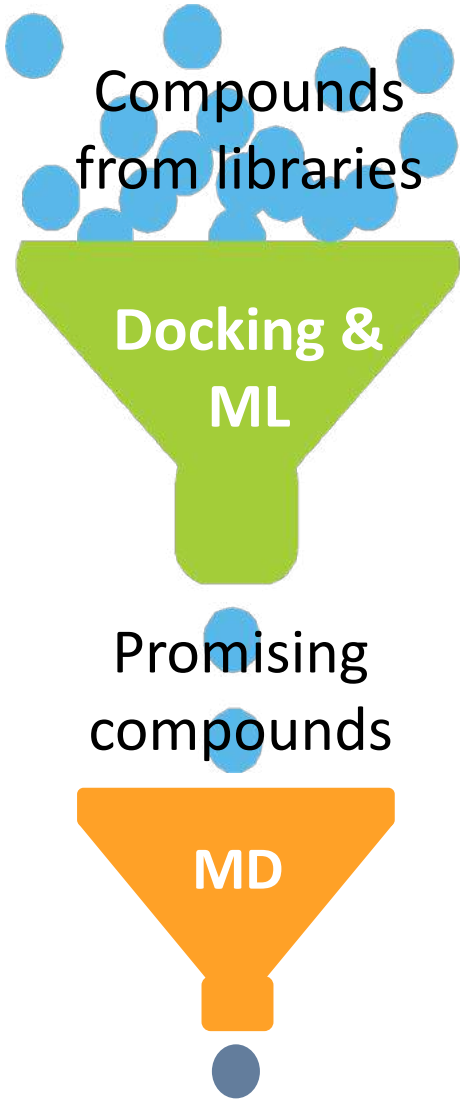
Multiscale modelling and simulation

- Bridging, combining and integrating multiple levels of description and representations of matter
  - combine many theories / methods in our research
  - necessary to solve real-world problems

- **Machine Learning** to screen large datasets
  - *De Novo* drug design
  - ZINC: 230 million purchasable compounds
  - DrugBank: 13,551 drug entries
  - PubChem: 103 million compounds
  - Enamine: 188,734 building blocks

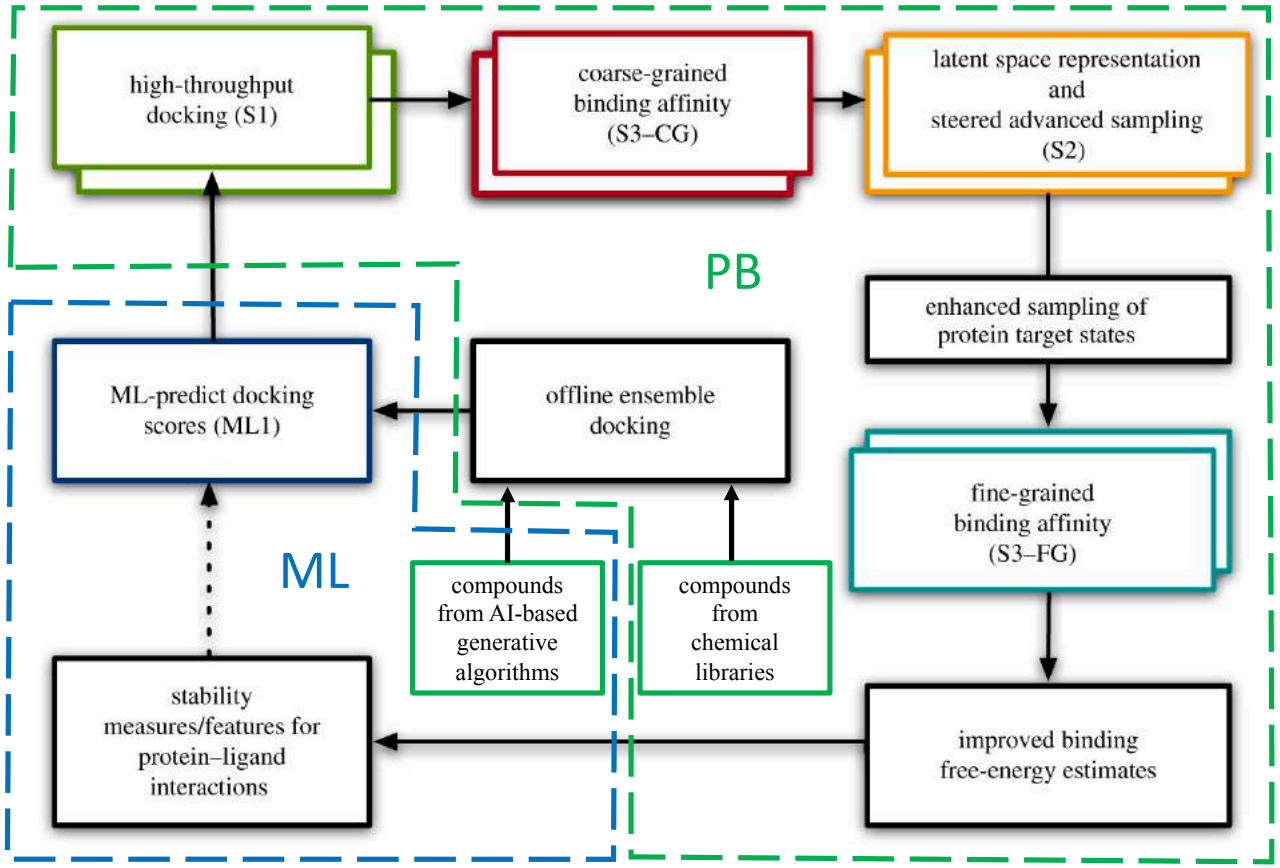
- **Physics-based model** to predict binding affinities
  - **ESMACS** to identify promising lead compounds
  - **TIES** to further optimise the lead compounds

→ **Quick Turnaround** is required, especially urgent to find effective drug treatment for COVID-19



IMPECCABLE Workflow - Accelerating COVID-19 drug discovery

- No single algorithm or method can achieve the necessary accuracy with required efficiency to sample the huge chemical space.
- **Physics and machine learning based methods are used symbiotically** to test drugs with required accuracy and efficiency.
- DeepDriveMD is a deep learning based approach and has been demonstrated for protein folding trajectories, offering at least 2× speedup compared to traditional conformational sampling methods.

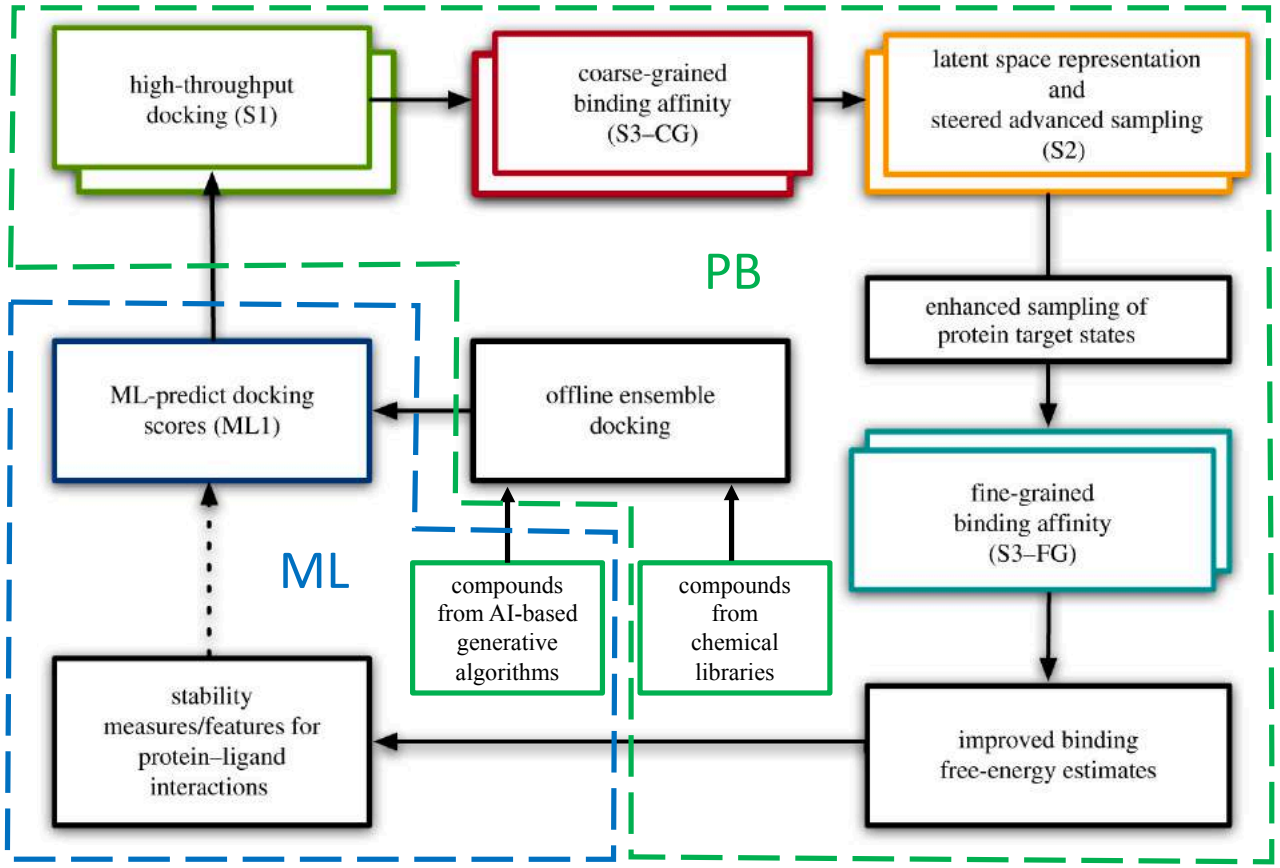


The pipeline combines ML and PB into a unified workflow, allowing both upstream and downstream exchange of information in the iterative loop.

IMPECCABLE: Integrated Modeling Pipeline

Multi-stage campaign employed to select promising drug candidates:

- **Stage-1**: High-throughput ensemble docking to identify small molecules ("hits")
- **Stage-2**: AI-driven Molecular Dynamics modeling specific binding regions and understanding mechanistic changes involving drugs
- **Stage-3**: Binding Free Energy calculation of promising leads ("Hit-to-Lead")



A. Al Saadi, D. Alfe, Y. Babuji, A. Bhati, B. Blaiszik, A. Brace, T. Brettin, K. Chard, R. Chard, A. Clyde, P. V. Coveney, I. Foster, T. Gibbs, S. Jha, K. Keipert, T. Kurth, D. Kranzlmüller, H. Lee, Z. Li, H. Ma, A. Merzky, G. Mathias, A. Partin, J. Yin, A. Ramanathan, A. Shah, A. Stern, R. Stevens, L. Tan, M. Titov, A. Trifan, A. Tsaris, M. Turilli, H. Van Dam, S. Wan, D. Wifling, "IMPECCABLE: Integrated Modeling PipelinE for COVID Cure by Assessing Better LEads", 50th International Conference on Parallel Processing (ICPP '21), August 9-12 (2021), DOI: 10.1145/3472456.3473524

IMPECCABLE: Integrated Modeling Pipeline

No turnkey solutions! Why is this challenging?

- **Heterogeneous:** High-throughput function calls (S1) ensembles of MPI tasks (S3); coupled AI-HPC (S2))
  - Producers of data (S1) and consumers (ML1) *"Supercomputers will become merely rapid generators of data for powerful ML models"*
- **Adaptivity** at multiple levels
  - Workload: Task mix varies over campaign
  - Tasks: Run for varying duration
- **Collective** versus single-task performance
  - Campaigns are "integrated" workflows: S1 and S3-FG differ by 10^7x in computational cost

Table 2: Normalized computational costs on Summit.

Method	Nodes per ligand	Hours per ligand (approx)	Node-hours per ligand
Docking (S1)	1/6	0.0001	~0.0001
BFE-CG (S3-CG)	1	0.5	0.5
Ad. Sampling (S2)	2	2	4
BFE-FG (S3-FG)	4	1.25	5
BFE-TI (not integrated)	64	10	640

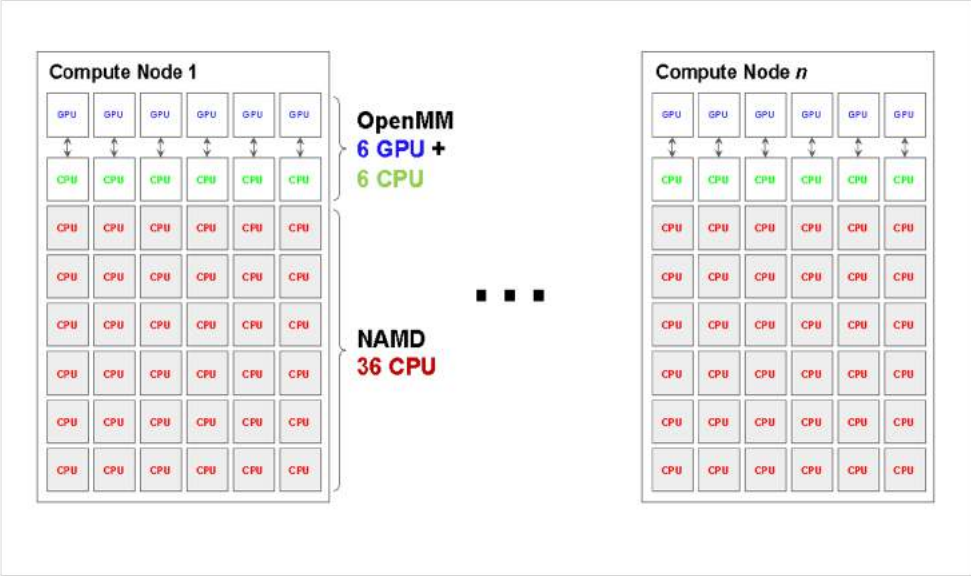
10^7x variation in cost across workflows

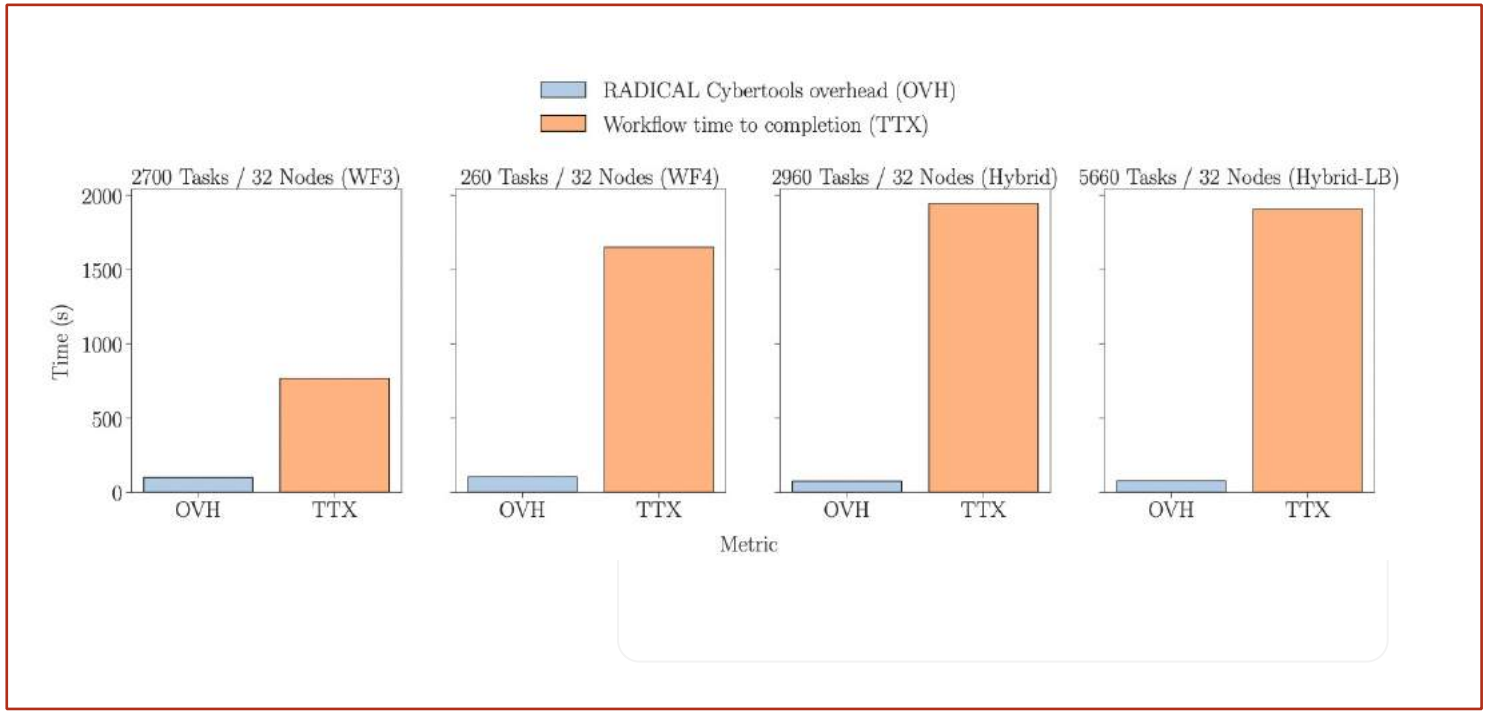
Table 3: Throughput and performance measured as peak flop per second (mixed precision, measured over short but time interval) per Summit node (6 NVIDIA V100 GPU).

Comp.	#GPUs	Tflop/s	Throughput
ML1	1536	753.9	319674 ligands/s
S1	6000	112.5	14252 ligands/s
S3–CG	6000	277.9	2000 ligand/s
S3–FG	6000	732.4	200 ligand/s

1000x variation in workflow throughput

Concurrent Heterogeneous Task Placement





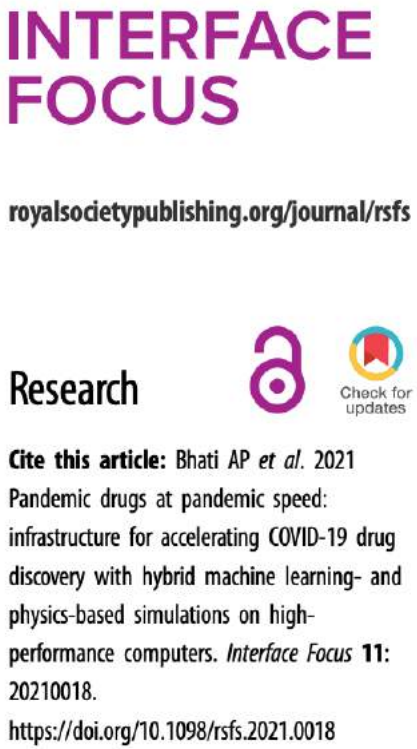
Concurrent Heterogeneous Task placement permits better resource utilization by supporting hybrid workflows.

H. Lee, A. Merzky, L. Tan, M. Titov, M. Turilli, D. Alfe, A. Bhati, A. Brace, A. Clyde, P. V. Coveney, H. Ma, A. Ramanathan, R. Stevens, A. Trifan, H. Van Dam, S. Wan, S. Wilkinson, S. Jha, "Scalable HPC & AI Infrastructure for COVID-19 Therapeutics", Platform for Advanced Scientific Computing Conference (PASC '21), July 5-9 (2021), DOI: 10.1145/3468267.3470573

ML/MD: Accelerating COVID-19 Drug Discovery

IMPECCABLE: Integrated Modeling PipelinE for COVID Cure by Assessing Better Leads

A.A. Saadi, *et al*., **2021**, In ACM Int. Conf. on Parallel Processing (ICPP) 2021, 9–12 August, Chicago, IL. Piscataway, NJ: IEEE Computer Society.



INTERFACE FOCUS

royalsocietypublishing.org/journal/rsfs

Research

Check for updates

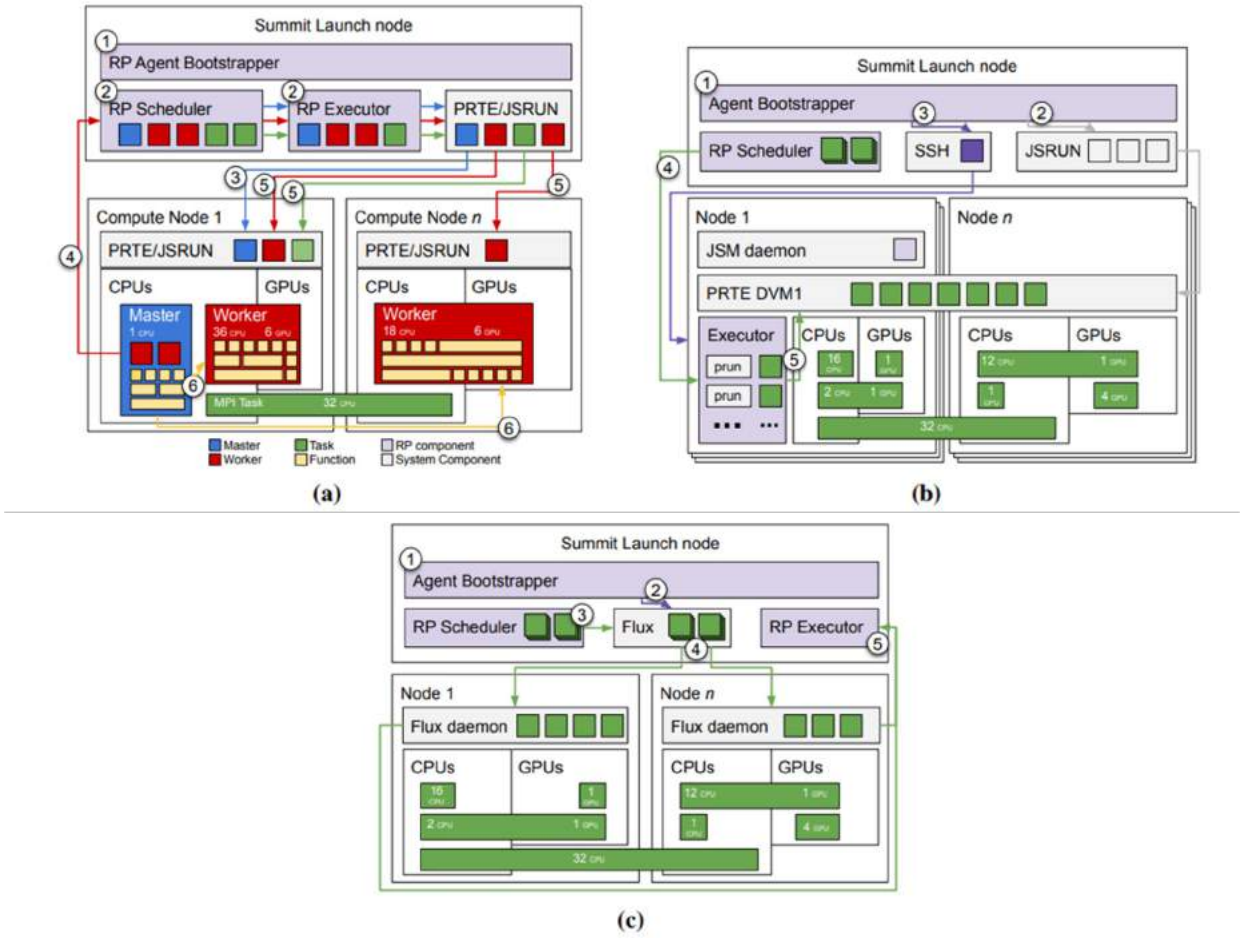
Cite this article: Bhati AP *et al*. 2021 Pandemic drugs at pandemic speed: infrastructure for accelerating COVID-19 drug discovery with hybrid machine learning- and physics-based simulations on high-performance computers. *Interface Focus* **11**: 20210018.
https://doi.org/10.1098/rsfs.2021.0018

Pandemic drugs at pandemic speed: infrastructure for accelerating COVID-19 drug discovery with hybrid machine learning- and physics-based simulations on high-performance computers

Agastya P. Bhati[1], Shunzhou Wan[1], Dario Alfè[2,3], Austin R. Clyde[4], Mathis Bode[5], Li Tan[6], Mikhail Titov[7], Andre Merzky[7], Matteo Turilli[7], Shantenu Jha[6,7], Roger R. Highfield[8], Walter Rocchia[9], Nicola Scafuri[9], Sauro Succi[10], Dieter Kranzlmüller[11], Gerald Mathias[11], David Wifling[11], Yann Donon[12], Alberto Di Meglio[12], Sofia Vallecorsa[12], Heng Ma[13], Anda Trifan[13], Arvind Ramanathan[13], Tom Brettin[14], Alexander Partin[13], Fangfang Xia[13], Xiaotan Duan[4], Rick Stevens[14] and Peter V. Coveney[1,15]

IMPECCABLE Workflow - Performance

- Pilot-based task execution frameworks implemented using RADIAL-Pilot allow for the execution of complex workflows large heterogeneous HPC such as **Summit**.
- The infrastructure has supported a campaign utilizing 2.5×10^6 node-hours on diverse HPC platforms for:
- docking $\sim10^{11}$ ligands with a peak docking rate of $\sim150\times10^6$ docks/hr,
- AI-driven enhanced sampling simulations, which demonstrate 10× scientific improvement over traditional methods; and
- computing binding free energies on $\sim10^5$ ligand-protein complexes, including 10^4 concurrently.
- These methods and infrastructure have enabled the **screening of more than 4.2 billion molecules** against over a dozen drug targets in SARS-CoV-2. So far, **over 1000 compounds have been identified and experimentally validated**, resulting in advanced testing for dozens of hits



Lee, H., Merzky, A., Tan, L., Titov, M., Turilli, M., Alfe, D., Bhati, A., Brace, A., Clyde, A., Coveney, P. and Ma, H., 2021, July. Scalable HPC & AI infrastructure for COVID-19 therapeutics. In *Proceedings of the Platform for Advanced Scientific Computing Conference* (pp. 1-13).

Industrial Strength

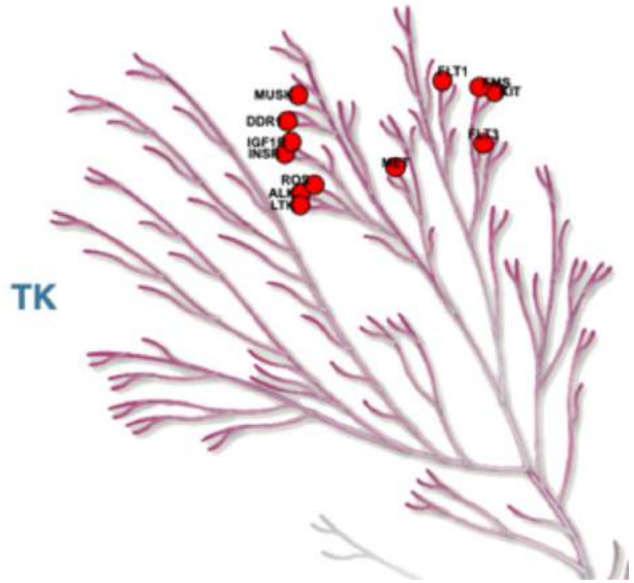
Scale-out block operation on SuperMUC-NG
15–17 November 2021

We are able to use the entire machine, a total of **6444 nodes, for 11 hours, with one single job submission**:

- Pattern #3 simulations
- Embarrassingly parallel workload
- Assembled together via binding affinity calculator (BAC) workflow
- Task farming function on SuperMUC-NG
- Replica implementation in NAMD

***Ca* 150 drug-protein complexes evaluated**

- drug ranking: ranking binding free energies for a series of drugs to a given protein
- drug selectivity: preference of a given drug binding to various protein targets



Industrial Strength

Summit Reservation
23–24 November 2021

We were able to use the entire machine, **a total of 4550 nodes, for 24 hours**:

- Pattern #3 simulations
- MD engines: NAMD3 & OpenMM
- Each individual simulation runs on a single GPU
- Assembled together via binding affinity calculator (BAC) workflow

Scientific simulations:

- Absolute binding free energy calculations: **235 drug-protein complexes**
- Relative binding free energy calculations: **70 protein mutations**
- Markov state models: **500 separate simulations**

Exascale class Problems we encountered during the reservation:

Summit as configured currently cannot distribute the pattern #3 simulations across the entire machine – **very limited number of batch nodes**.

Solution:

Use middleware tools such as **RADICAL-Cybertools** (RCT) for flexible task-level parallelism.

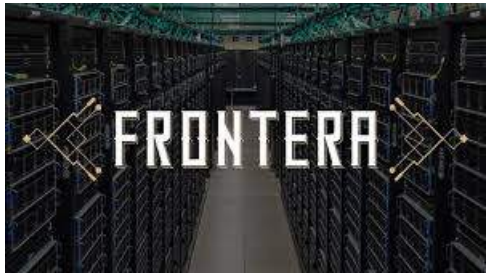
SCIENCE|BUSINESS®

07 Apr 2020 | Network Updates | Update from University College London

These updates are republished press releases and communications from members of the Science|Business Network

UCL researchers are using the world's most powerful supercomputers to tackle COVID-19

By Communication from UCL



Mail Online Science & Tech

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money | Video | Travel | DailyMailTV | Discounts

Latest Headlines Login

Scientists are using the world's most powerful supercomputers to speed up the development of treatments for the deadly coronavirus

Binding Affinity Calculations:

Binding affinity calculations at exascale:

- Applications of binding affinity calculations (BAC) on **Summit** and **SuperMUC-NG** are laying the groundwork to scale to exascale.

Current work has lead to the development of solutions that can facilitate the exascale deployment of BAC:

- Developing middleware tools such as **RADICAL-Cybertools** (RCT) for flexible task-level parallelism.
- Refining ensemble based task farming methodologies.

Current Opportunities to develop exascale solutions:

- **2021 INCITE Award** provided access to Summit; granted unlimited time to run on it through January 2022. Continuing in new award in 2022-23
- Over 150 million core hours awarded on SuperMUC-NG though **GCS awards**.

Preliminary results: High-throughput screening

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

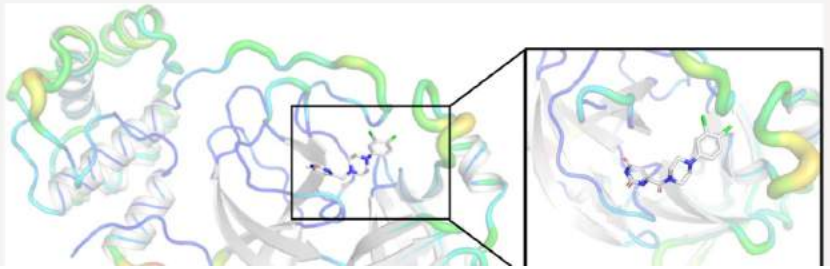
High-Throughput Virtual Screening and Validation of a SARS-CoV-2 Main Protease Noncovalent Inhibitor

Austin Clyde,* Stephanie Galanie, Daniel W. Kneller, Heng Ma, Yadu Babuji, Ben Blaiszik, Alexander Brace, Thomas Brettin, Kyle Chard, Ryan Chard, Leighton Coates, Ian Foster, Darin Hauner, Vilmos Kertesz, Neeraj Kumar, Hyungro Lee, Zhuozhao Li, Andre Merzky, Jurgen G. Schmidt, Li Tan, Mikhail Titov, Anda Trifan, Matteo Turilli, Hubertus Van Dam, Srinivas C. Chennubhotla, Shantenu Jha,* Andrey Kovalevsky,* Arvind Ramanathan,* Martha S. Head,* and Rick Stevens*

Cite This: *J. Chem. Inf. Model.* 2022, 62, 116–128 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Despite the recent availability of vaccines against the acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the search for inhibitory therapeutic agents has assumed importance especially in the context of emerging new viral variants. In this paper, we describe the discovery of a novel noncovalent small-molecule inhibitor, MCULE-5948770040, that binds to and inhibits the SARS-Cov-2 main protease (M^{pro}) by employing a



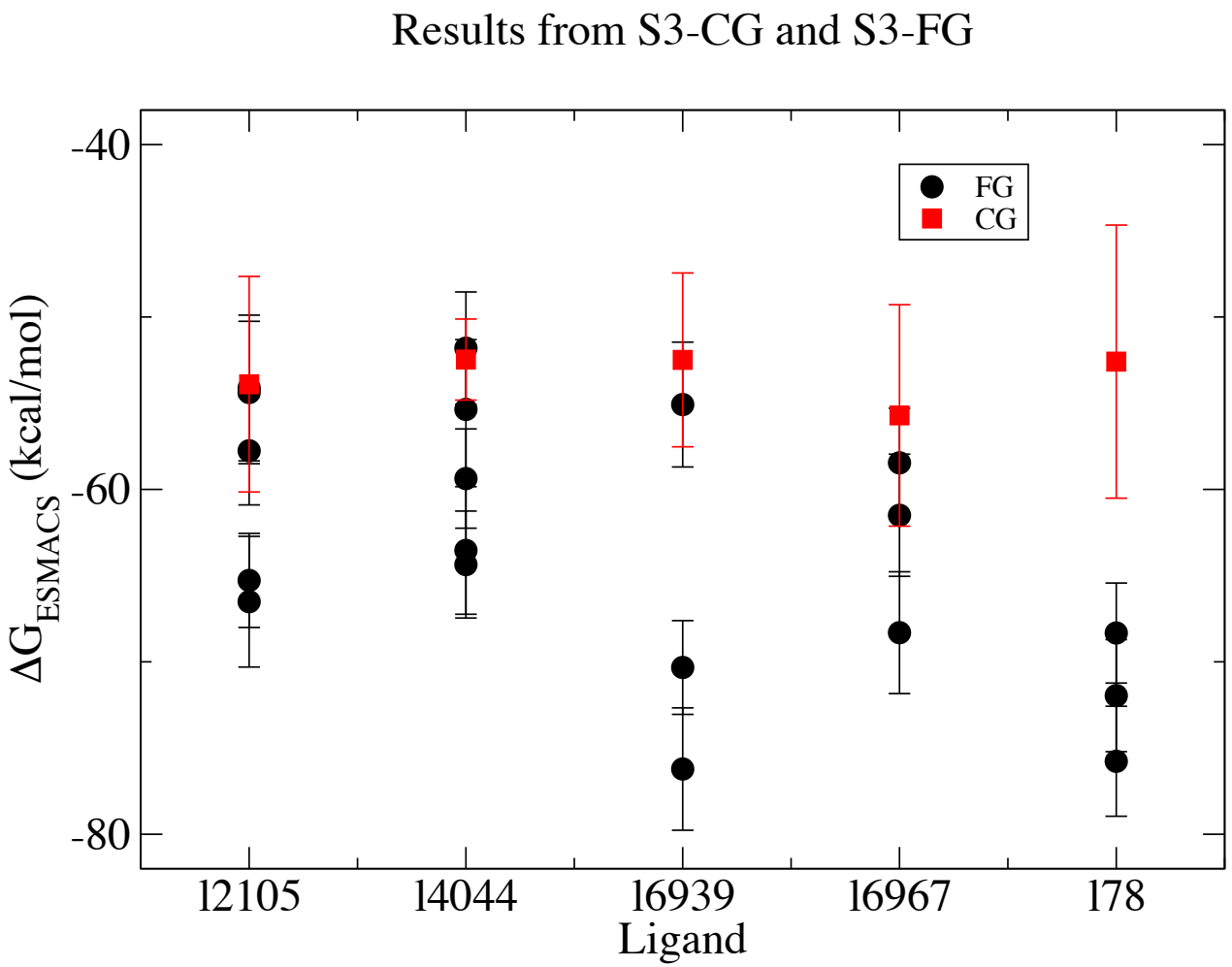
Preliminary results: S3-CG

- Number of the most promising compounds for each of the 4 proteins investigated. For each protein, the top 100 compounds, chosen from 10,000 docked small molecules, are evaluated by ESMACS approach. The number of compounds are listed, which have the most favourable binding free energies, in the ranges corresponding to K_D values of 10 nM (-10.98 kcal/mol), 100 nM (-9.61 kcal/mol), 1 μM (-8.24 kcal/mol).

Energy (kcal/mol)	3CLPro	ADRP	NSP15	PLPro
ΔG < -10.98	1	0	3	6
-10.98 ≤ ΔG < -9.61	2	2	1	8
-9.61 ≤ ΔG < -8.24	1	4	10	5
ΔG < -8.24 Total	4	6	14	19

A. P. Bhati, S. Wan, D. Alfè, A. R. Clyde, M. Bode, L. Tan, M. Titov, A. Merzky, M. Turilli, S. Jha, R. R. Highfield, W. Rocchia, N. Scafuri, S. Succi, D. Kranzlmüller, G. Mathias, D. Wifling, Y. Donon, A. Di Meglio, S. Vallecorsa, H. Ma, A. Trifan, A. Ramanathan, T. Brettin, A. Partin, F. Xia, X. Duan, R. Stevens, P. V. Coveney, "Pandemic Drugs at Pandemic Speed: Accelerating COVID-19 Drug Discovery with Hybrid Machine Learning- and Physics-based Simulations on High Performance Computers", Interface Focus, 11, 20210018, DOI: 10.1098/rsfs.2021.0018

Preliminary results: S3-CG to S2 to S3-FG



Comparison of S3-CG and S3-FG results for the ve best binders for PLPro (PDBID: 6w9c) based on CG-ESMACS results. S2 selected five outlier conformations for each binder and performed FG-ESMACS on them. The provisional results confirm improved binding for the selected conformations in all five compounds, as FG energies are lower than CG.

A. Al Saadi, D. Alfe, Y. Babuji, A. Bhati, B. Blaiszik, A. Brace, T. Brettin, K. Chard, R. Chard, A. Clyde, P. V. Coveney, I. Foster, T. Gibbs, S. Jha, K. Keipert, T. Kurth, D. Kranzlmüller, H. Lee, Z. Li, H. Ma, A. Merzky, G. Mathias, A. Partin, J. Yin, A. Ramanathan, A. Shah, A. Stern, R. Stevens, L. Tan, M. Titov, A. Trifan, A. Tsaris, M. Turilli, H. Van Dam, S. Wan, D. Wifling, "IMPECCABLE: Integrated Modeling PipelinE for COVID Cure by Assessing Better LEads", 50th International Conference on Parallel Processing (ICPP '21), August 9-12 (2021), DOI: 10.1145/3472456.3473524

Conclusions and perspectives

- ML and PB approaches have their own advantages and limitations
- A more productive blend of ML and PB mechanistic understanding
- An iterative cycle is used, where AI hypotheses are tested in physics-based simulations, and the results of PB modelling are used to train AI.
- Equipped with advanced VVUQ to handle high-dimensional parametric spaces
- Use of ML and PB pipelines for actionable predictions to guide drug discovery
- Applications of ML and PB pipelines on emerging exascale architectures

Acknowledgements

Team members:

- Peter V. Coveney
- Shunzhou Wan
- Mat Bieniek
- Alex Wade
- Maxime Vassaux
- Wouter Edeling
- and many others ...

compbiomed.eu/

vecma.eu/

Horizon 2020
European Union Funding
for Research & Innovation

Engineering and
Physical Sciences
Research Council

ucl.ac.uk/mesoscale-modelling-consortium/

Q&A

To pose a question, you can write your question in the “Questions” tab

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 823712

INSILICO WORLD

https://insilicoworld.slack.com/archives/C0151M02TA4

The e-Seminar series is run in collaboration with:

VPH Institute
Building the Virtual Physiological Human

Thank you for participating!

…don't forget to fill in our feedback questionnaire…

Visit the CompBioMed website (www.compbiomed.eu/training)
for a full recording of this and other e-Seminars,
to download the slides
and to keep updated on our upcoming trainings

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823712

INSILICO WORLD

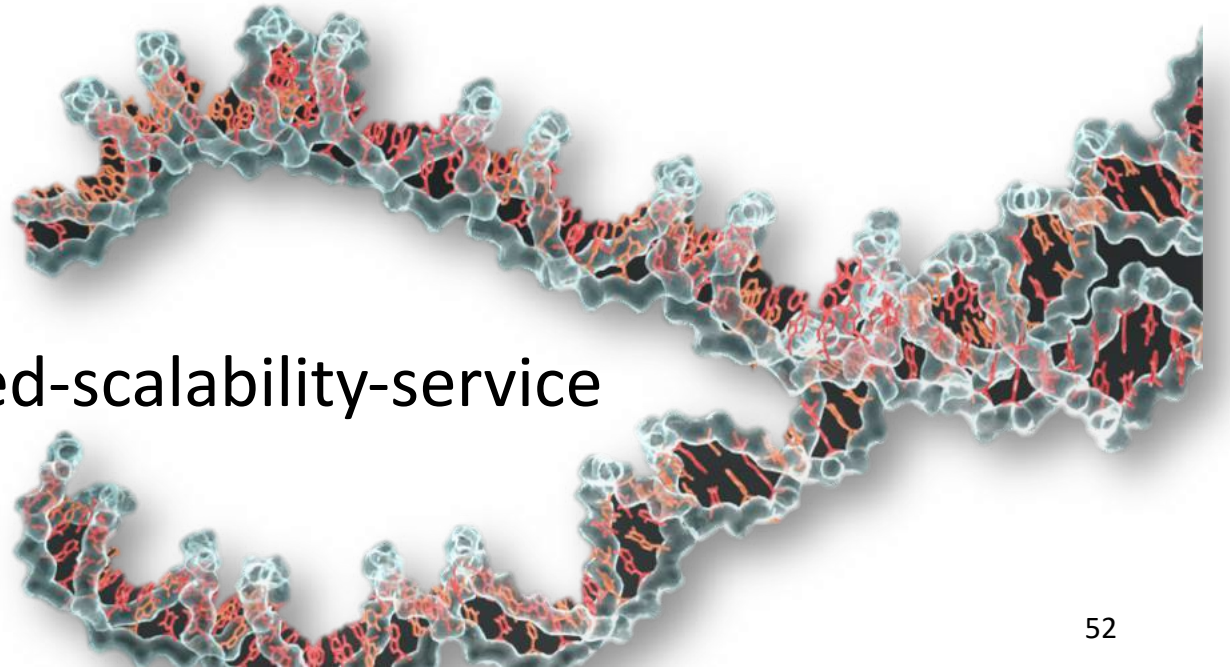
https://insilicoworld.slack.com/archives/C0151M02TA4

The e-Seminar series is run in collaboration with:

VPH Institute
Building the Virtual Physiological Human

CompBioMed's *Free* Scalability Service

- Improves performance of your biomedicine applications on high performance computers
  - Experts in both biomedical applications and high performance computers
  - Make your biomedicine applications run in parallel
  - Improving the scalability of those already parallelised



- www.compbiomed.eu/compbiomed-scalability-service

www.compbiomed.eu/compbiomed-scalability-service

- Contact for *Free* Service
  - General technical questions
    - Slack: #scalability channel of ***the InSilicoWorld Community of Practice***
    - Email: compbiomed-support@ucl.ac.uk
  - Full service
    - Application Form or light-weight web form
      - Formal collaborative relationship with CompBioMed Centre of Excellence

- Application and Data Security
  - Great care when adapting your applications and managing your data
    - Our Data Policies cover Data Privacy, Data Security and Research Data Management

InSilicoWorld Community of Practice

The first community entirely on *in silico* medicine on Slack

www.insilico.world/community

Expertise	• The community is invitation only: in this way we ensure only interested experts have access
Collaboration	• Join teams and collaboratively work on shared goals, projects, concerns, problems or topics
Safe space	• A pre-competitive space where experts from academia, industry, and regulatory agencies can ask for and exchange advices

More than 500 experts have already joined the community and its channels

InSilicoWorld Members

- **Large Biomedical Companies**

  Medtronic, Smith & Nephew, Pfizer, Johnson and Johnson, Innovative Medicine Initiative, CSL Behring, Ambu, RS-Scan, Corwave EN, Zimmer Biomet, Novartis, Bayer, ATOS, Biogen, Agfa, Icon PLC, Amgen, ERT, Exponent, etc.

- **Biomedical SMEs**

  Nova Discovery, Lynkeus, Obsidian Biomedical, Quibim, Mediolanum Cardio Research, Voisin Consulting, CRM-Microport, Mimesis srl, H. M. Pharmacon, MCHCE, etc.

- **Independent Software Vendors**

  Ansys, In Silico Trials Technologies, 3DS, KIT, ASD Advanced Simulation & Design GmbH, Kuano-AI, Aparito, Chemotargets, Digital Orthopaedics, ExactCure, Materialise, Bio-CFD, Matical, FEOPS, 4RealSim, Exploristics, Synopsis, Virtonomy, Cad-Fem Medical, etc.

- **Regulators and Standardisation Bodies**

  FDA, DIN, BSCI China, NICE, Critical Path Institute, ACQUAS, etc.

- **Clinical Research Institutions**

  Istituto Ortopedico Rizzoli, Sloan Kettering Cancer Center, Royal College of Surgeons Ireland, Gratz University Hospital, Charite Berlin, Centre Nacional Invesigaciones Oncologicas, Aspirus Health, Universitätsklinikum des Saarlandes, European Society for Paediatric Oncology, etc.