CENTRE FOR COMPUTATIONAL SCIENCE, ADVANCED RESEARCH COMPUTING CENTRE,
UNIVERSITY COLLEGE LONDON & INSTITUTE FOR INFORMATICS, UNIVERSITY OF AMSTERDAM

Computational Science at the Exascale: the Myriad Uses of Frontier

Peter V. Coveney
Centre for Computational Science & Advanced Research Computing Centre, UCL (UK)
Institute of Informatics, UVA (NL)
Yale School of Medicine (USA)

OLCF User Meeting,
Oak Ridge National Laboratory
Tuesday, 5 August 2025

Contents

- Computing patterns and exascale computers
- Monolithic and coupled codes
- The importance of uncertainty and reproducibility in science
- The drug discovery, development and market pipeline
- Artificial intelligence in drug discovery
- The IMPECCABLE workflow

Much of our work requires big supercomputers

Includes emphasis on ‘VVUQ’ ==> trusting predictions, making them actionable

Increasing opportunities as high performance computing moves from petascale to exascale

Summit

Blue Waters

Titan

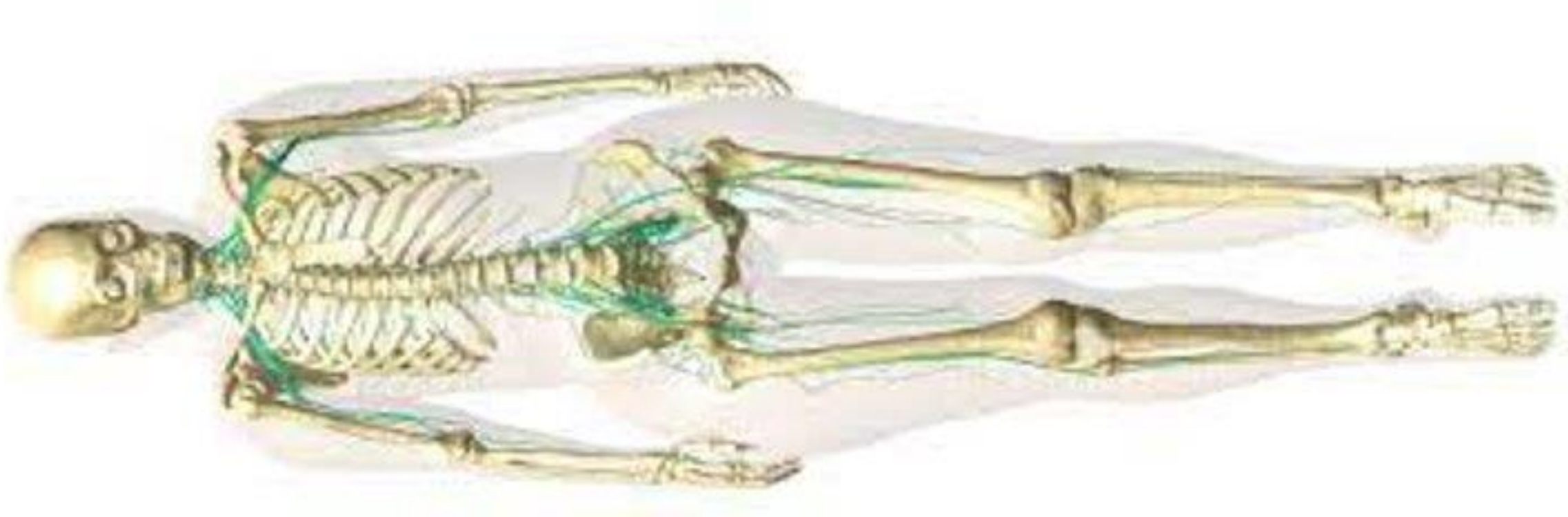
Dawn

SuperMUC-NG 2

Frontier

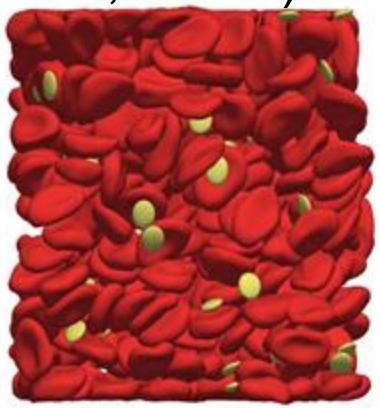
Aurora

Human anatomical data and models

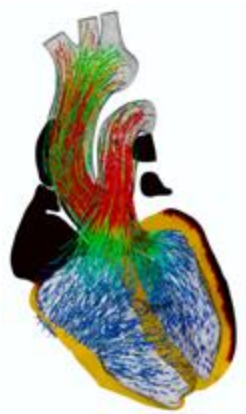


Organ Modelling: Cardiovasculature

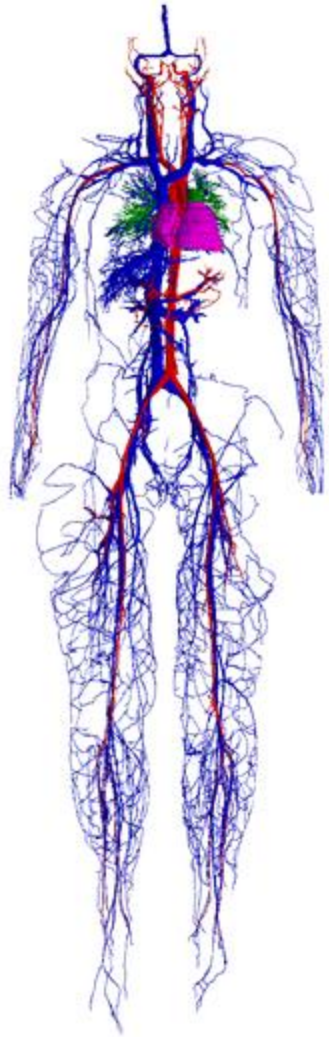
"Digital Blood" and flow diverting stents within Palabos (UvA, UNIGE, ATOS)



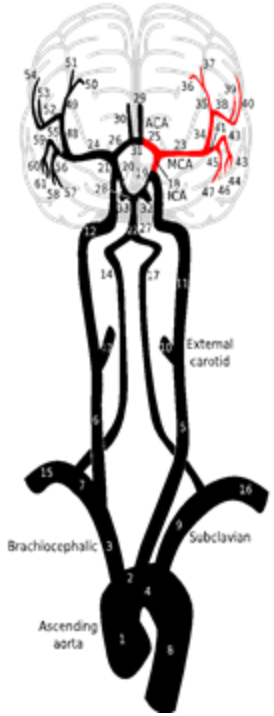
Alya Cardiac Computational Model (BSC, UOXF)



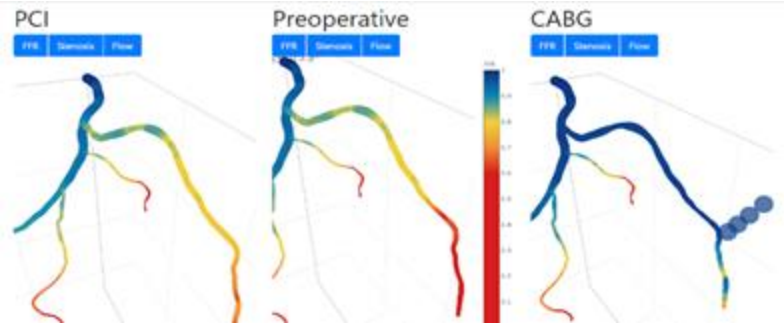
HemeLB-Alya coupling for cardiovascular flow in virtual human scale geometries (UCL, BSC)



OpenBF for vascular networks (USFD/SARA)



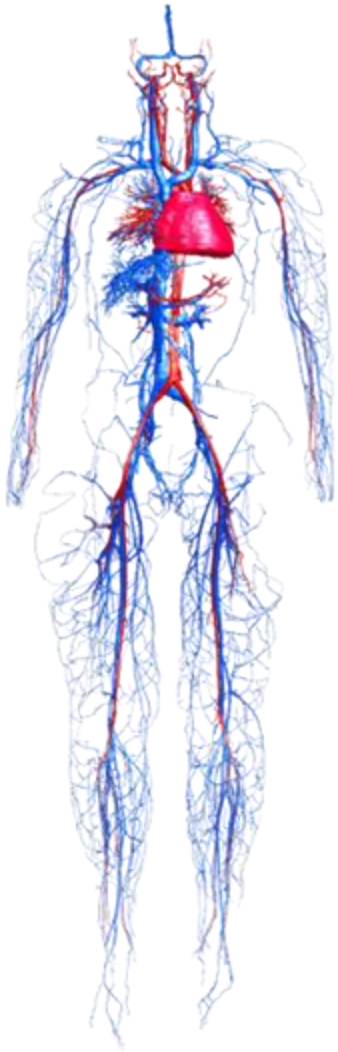
AngioSupport for coronary artery disease (LTG/SARA)



CompBioMed

HemeLB: Codesign of Software and Hardware

HemeLB and the Virtual Human



Blood flow networks through arteries and veins provide a natural basis on which to build a virtual human

Separate organ models can be attached through the blood being transferred to/from major vessels

Computationally requires a framework that can efficiently capture the highly individual nature of vessel structures in high fidelity

HemeLB provides a platform to achieve this

See HemeLB.org

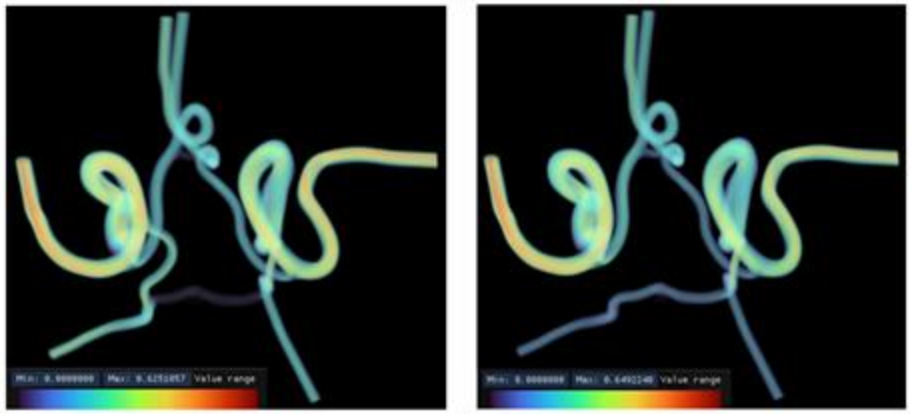
Our software release paper on HemeLB was recently published:

I. Zacharoudiou, J. W. S. McCullough, P. V. Coveney, "Development and performance of HemeLB GPU code for human-scale blood flow simulation", Computer Physics Communications, 282, 108548 (2023) DOI:10.1016/j.cpc.2022.108548

CompBioMed

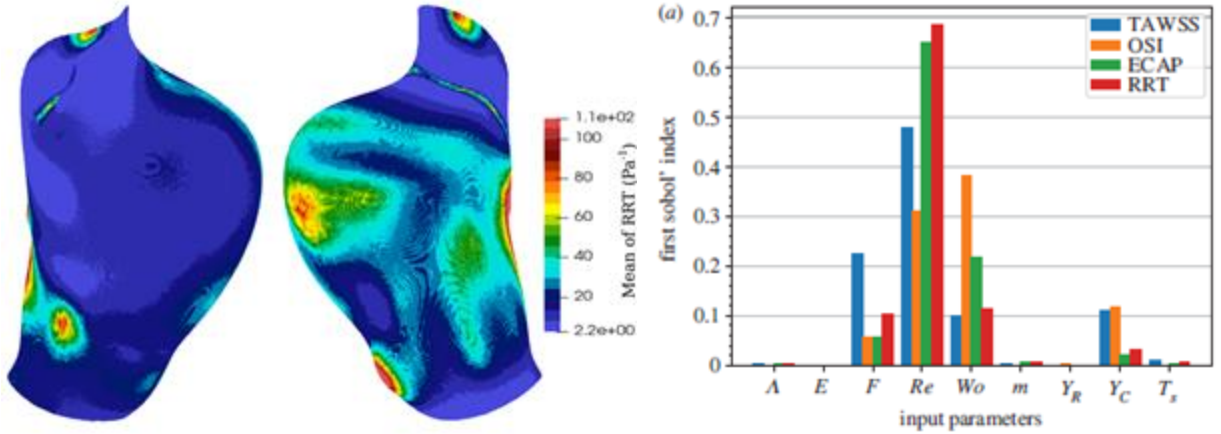
Simulation studies with HemeLB

Stroke in circle of Willis

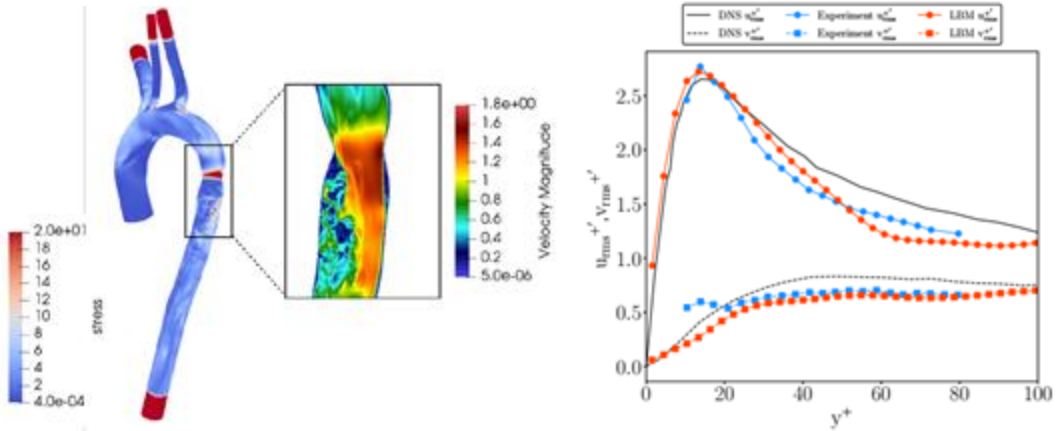


Type O Type Pcl

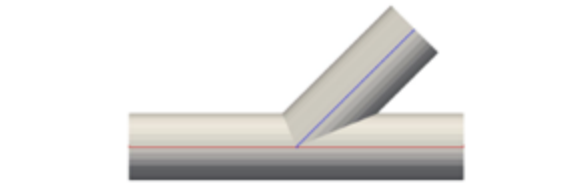
Abdominal aortic aneurysm



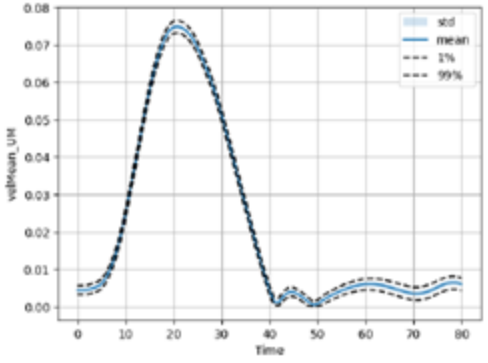
LES turbulence modelling



UQ of flow in bifurcations



(a) Canonical bifurcation



(b) Ulnar - Mean velocity distribution

HemeLB at exascale

HemeLB meets Frontier

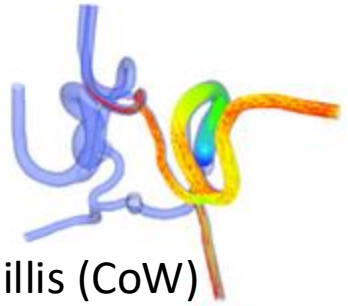
- Lattice Boltzmann solver
- Massively paralleled high-performance code
- Available on both CPU and GPU (CUDA/Hipified)
- Designed for sparse geometry, ideal for hemodynamics simulations

The world's first exascale machine. Our access to it has been via a pathway that has gone via SNG, Titan, BlueWaters, and Summit.

Frontier strong scaling plots

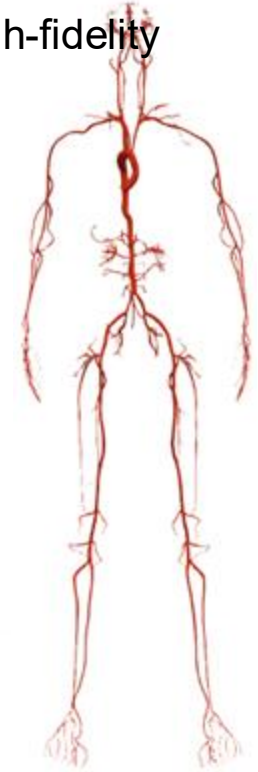
Full-human scale simulation requirements

- 140 billion lattice sites (1.4×10^{10})
- Full deployment on Frontier
- A few cardio cycles produce a high-fidelity simulation



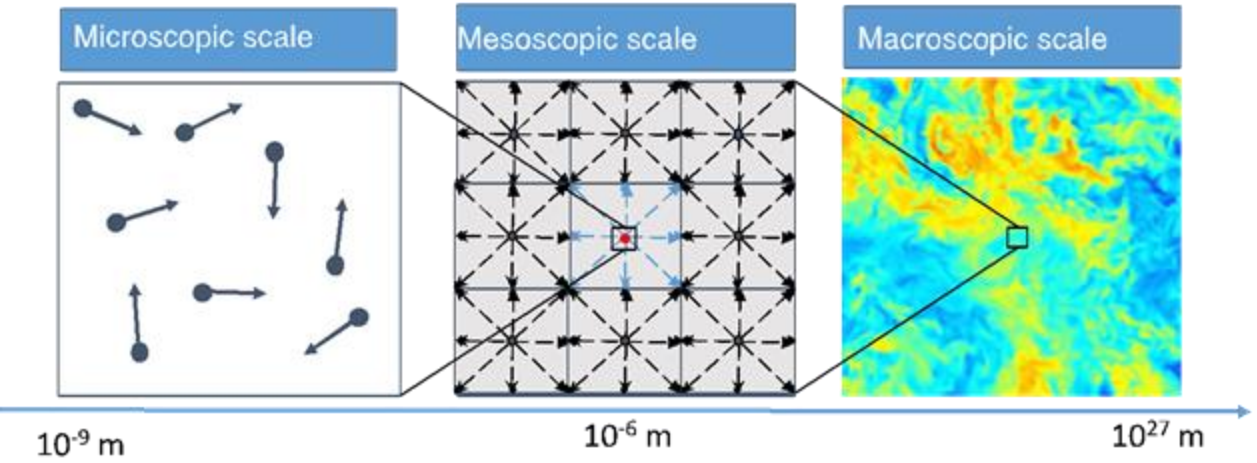
Circle of Willis (CoW)

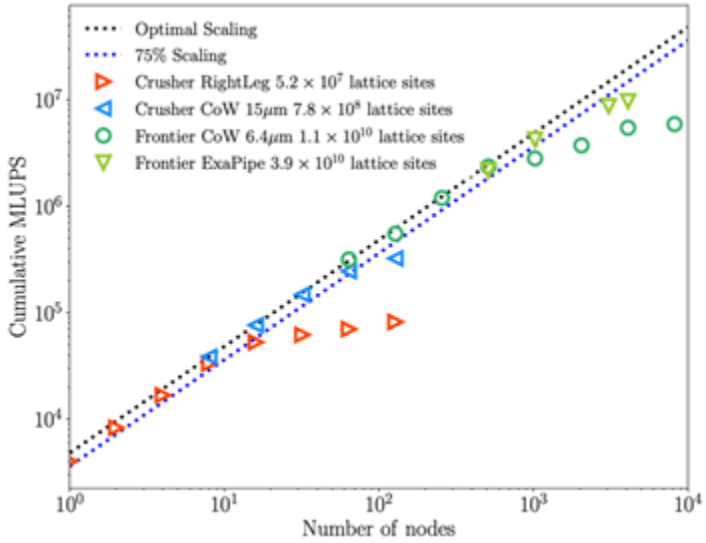
Human Right leg



Full-human scale arterial system

The lattice Boltzmann method





Code Status

HemeLB_GPU: Towards a platform agnostic HemeLB_GPU from CUDA to HIP and oneAPI

HemeLB_GPU code – Port to oneAPI

- Single HemeLB collision-streaming CUDA kernel ported to **Intel's oneAPI** (SYCL/DPC++)
- Ongoing efforts to port the full code to **oneAPI**
  - Deploy HemeLB_GPU on NVIDIA, AMD and Intel GPUs
- Porting to Aurora (ALCF's forthcoming exascale system)

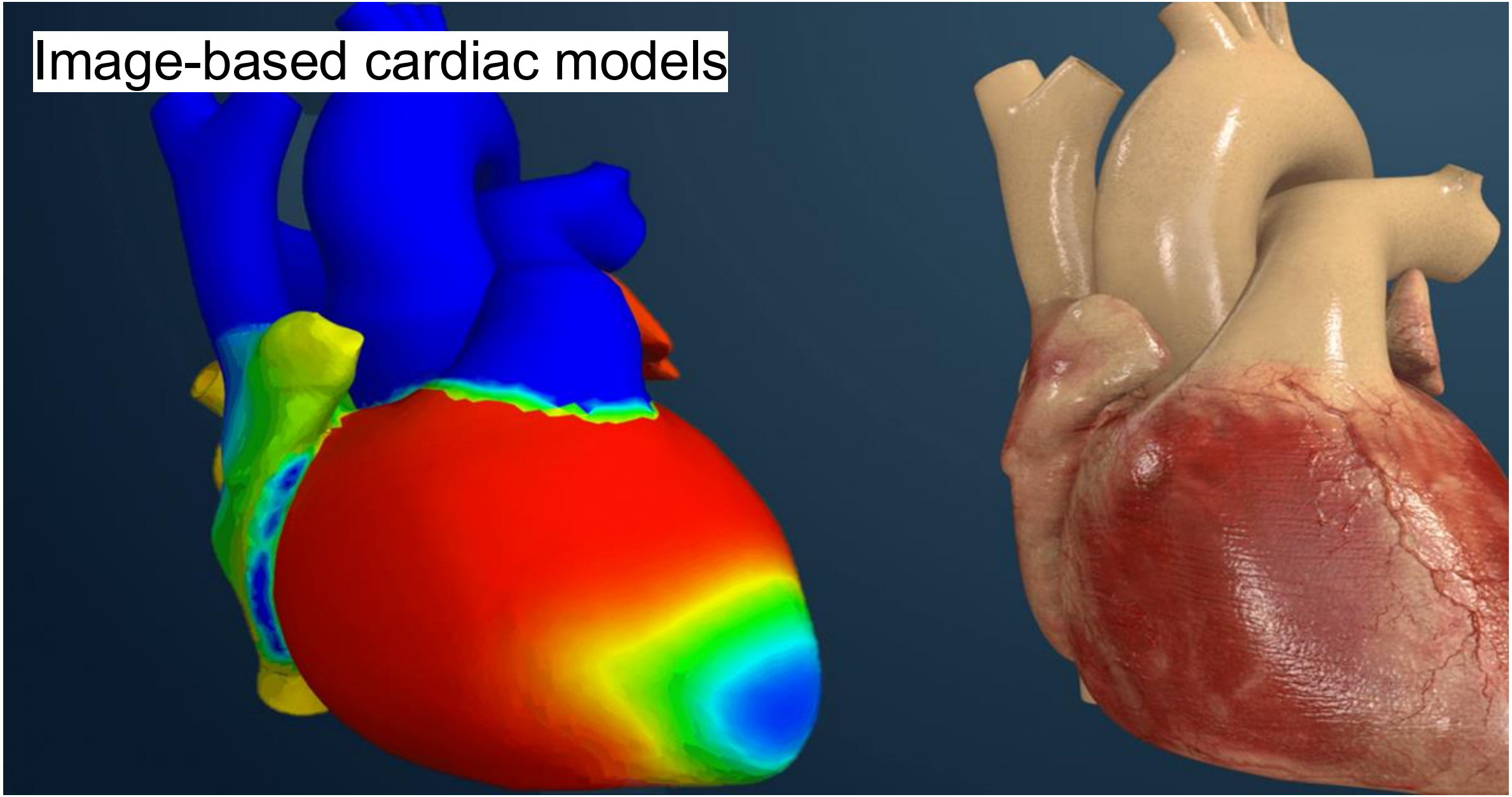
Data **P**arallel C++ **C**ompatibility **T**ool

Port to oneAPI approach:

- **Both with Intel DCPT Porting tool**
  - Testing on Sunspot (**ALCF** - 2x Intel Xeon CPU Max Series (Sapphire Rapids) and 6x Intel Data Center GPU Max Series (codename Ponte Vecchio or PVC).
  - Compiles – Unstable... Debugging... Step by step approach - Focus on Initialisation and then one kernel at a time...
- **Using factor-based approach**
  - Factor out the CUDA-specific parts of the code into back-ends:
  - Working for CUDA, HIP
  - **SYCL/DPC++ up and running on Frontier (OLCF)**
    - Unstable on Sunspot... Debugging in progress...

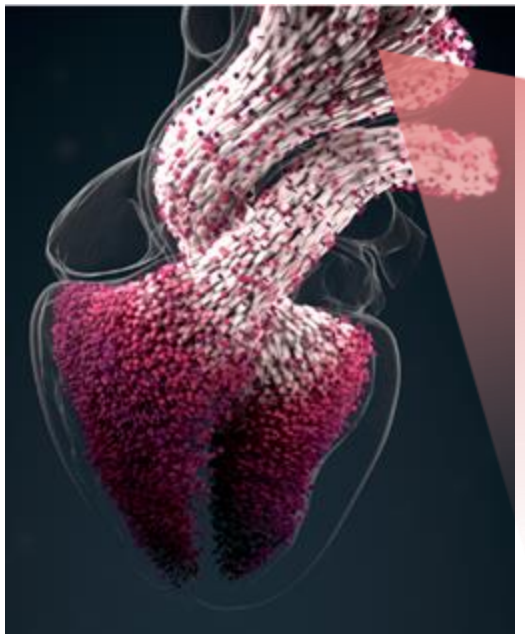
- Single source files – Reduce code duplication
- Simplifies code maintenance

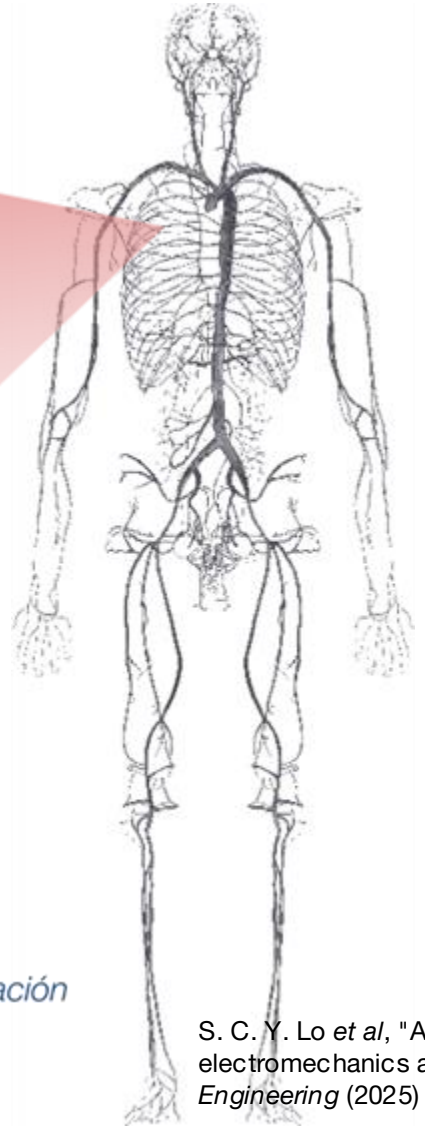
Image-based cardiac models



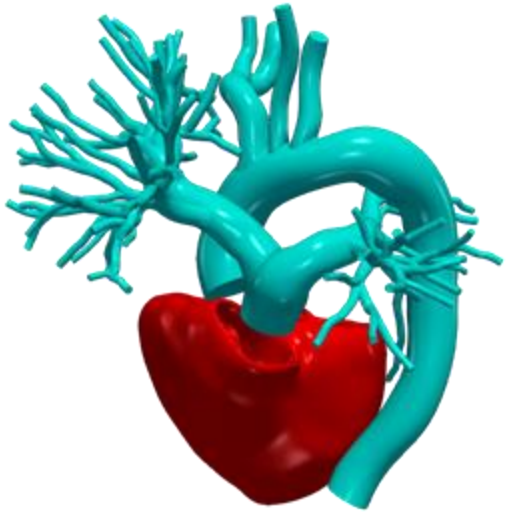
Digital Twin Reality

Coupling the Full Vascular Tree with the Heart





- Coupling with a full electromechanical heart model
- ALYA code is developed at Barcelona Supercomputing Center



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

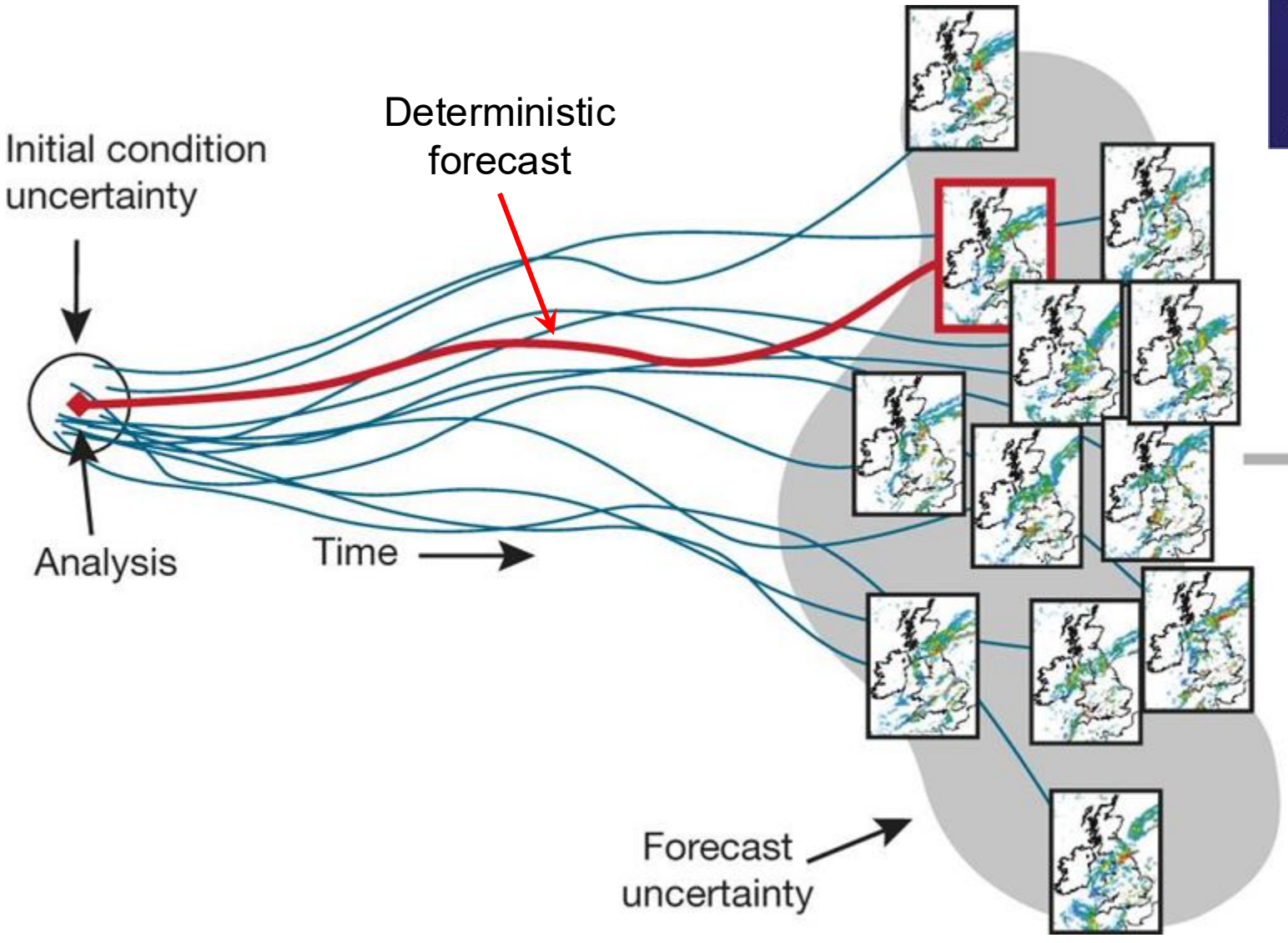
CompBioMed

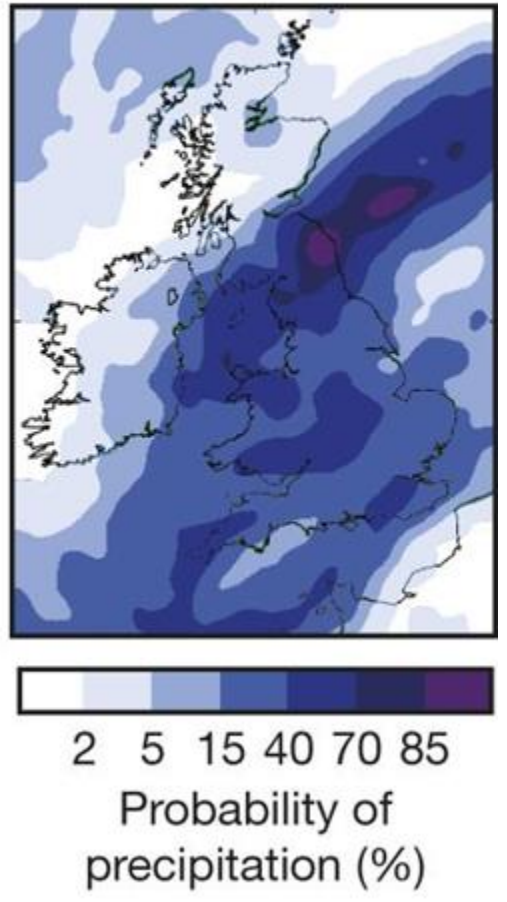
S. C. Y. Lo *et al*, "A multi-component, multi-physics computational model for solving coupled cardiac electromechanics and vascular haemodynamics", *Computer Methods in Applied Mechanics and Engineering* (2025) DOI: 10.1016/j.cma.2025.118185

Uncertainty in weather and climate prediction

UCL

Uncertainty exists in the same model

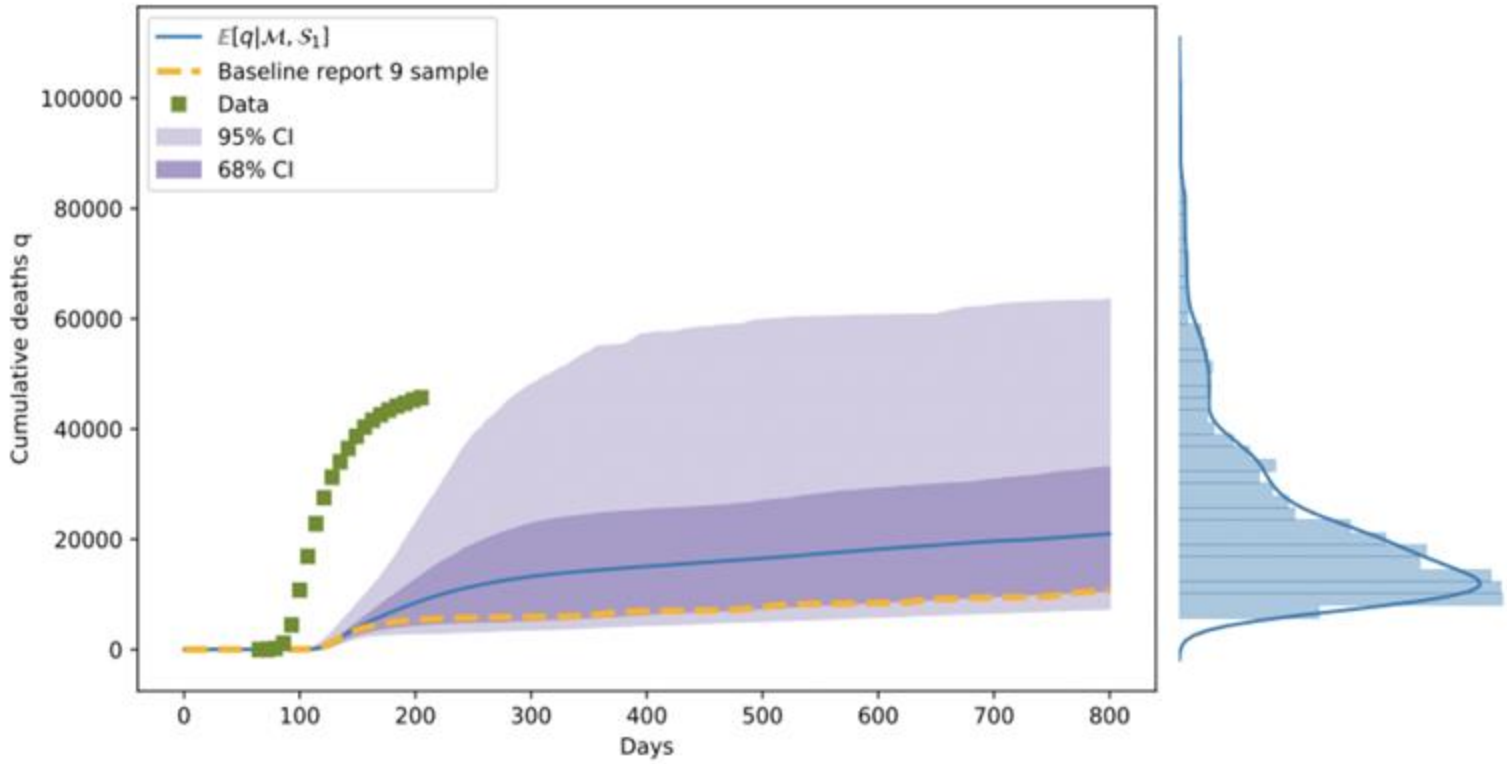




Decision-making with uncertainty

- Why should we care about uncertainty?
  - Model are used to inform high-level decision making.
  - Simulations without error bars paint a very incomplete picture.

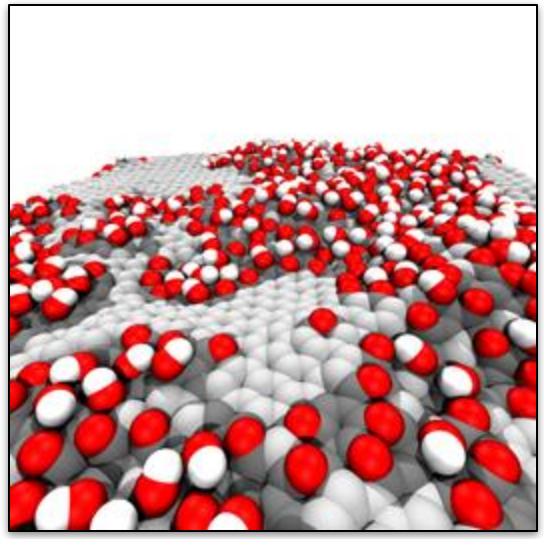
- No error bars: only the yellow line was available to UK government : approx 10,000 deaths after 600 days
- However, the same model applied in the same setting can also predict 10,000 deaths after < 200 days

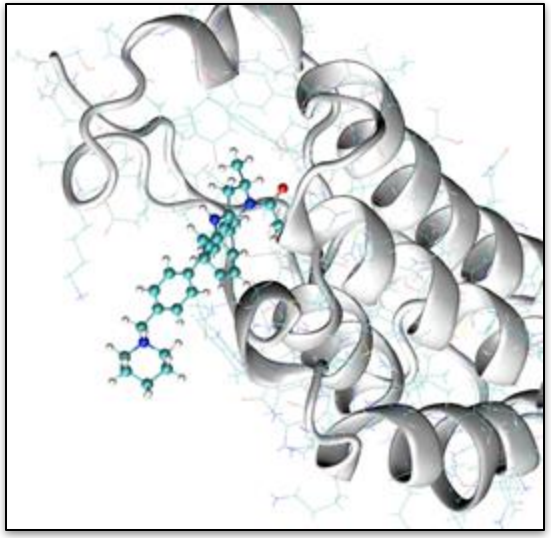


Edeling, Sinclair, Groen, Coveney *et al*. *Nature Computational Science* (2021) **1**, 128–135

Molecular dynamics

- Newton's equations of motion: complex interaction potentials, N-body dynamics, non-linear
- Programming them on a computer
- Linking MD to *thermodynamics*: exploring macroscopic behaviour through molecular interactions.
- Lyapunov instability leads to chaotic systems
  - small errors, large deviations
  - sensitive to uncertainty, need to quantify it!

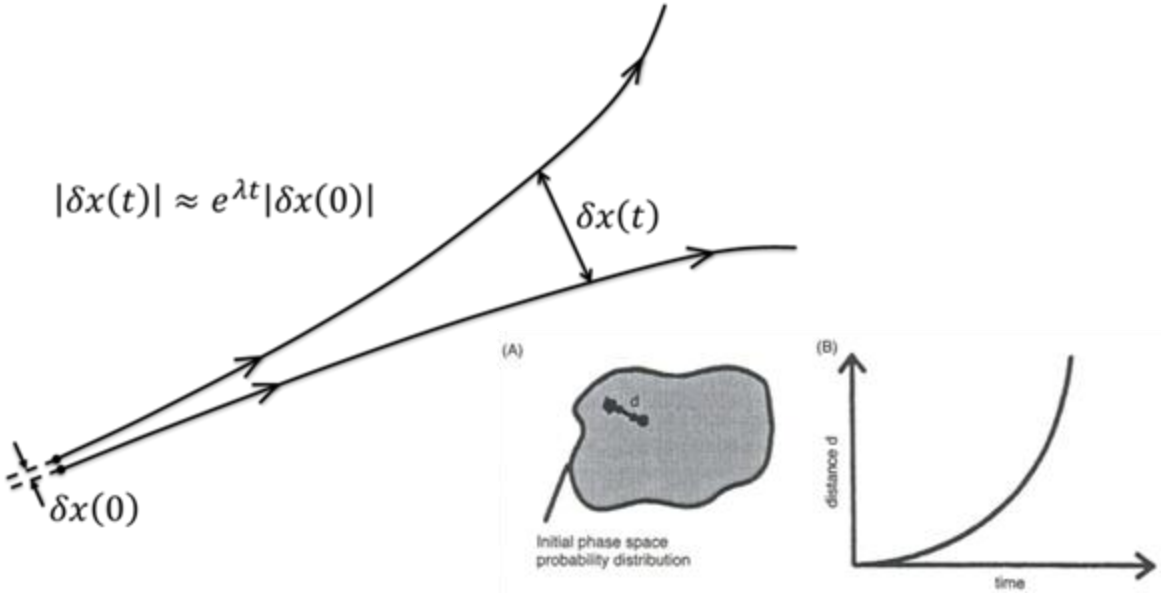




$$\mathbf{F} = m\frac{d^2\mathbf{r}}{dt^2}$$

$$\mathbf{F} = -\nabla V(\{\mathbf{R}\})$$

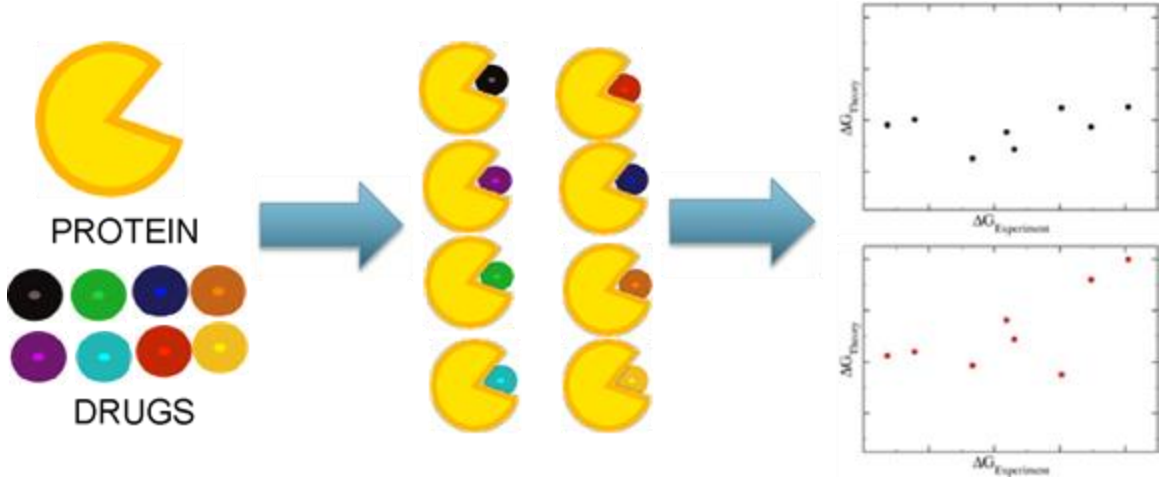
$$V(\{\boldsymbol{R}\}) = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} K_\chi\left(1+cos(n\chi-\chi_0)\right) + \sum_{impropers} K_\psi(\psi-\psi_0)^2 + \sum_{nonbond}\left[\left(\frac{A_{ij}}{r_{ij}}\right)^{12} - \left(\frac{B_{ij}}{r_{ij}}\right)^{6} + \frac{q_i q_j}{\varepsilon r_{ij}}\right]$$



Boghosian, Coveney and Wang. *Adv. Theor. Simul.* (2019) 2.12
Wan, et al. *J. Chem. Theory Comput.* (2015) 11, 7, 3346–3356

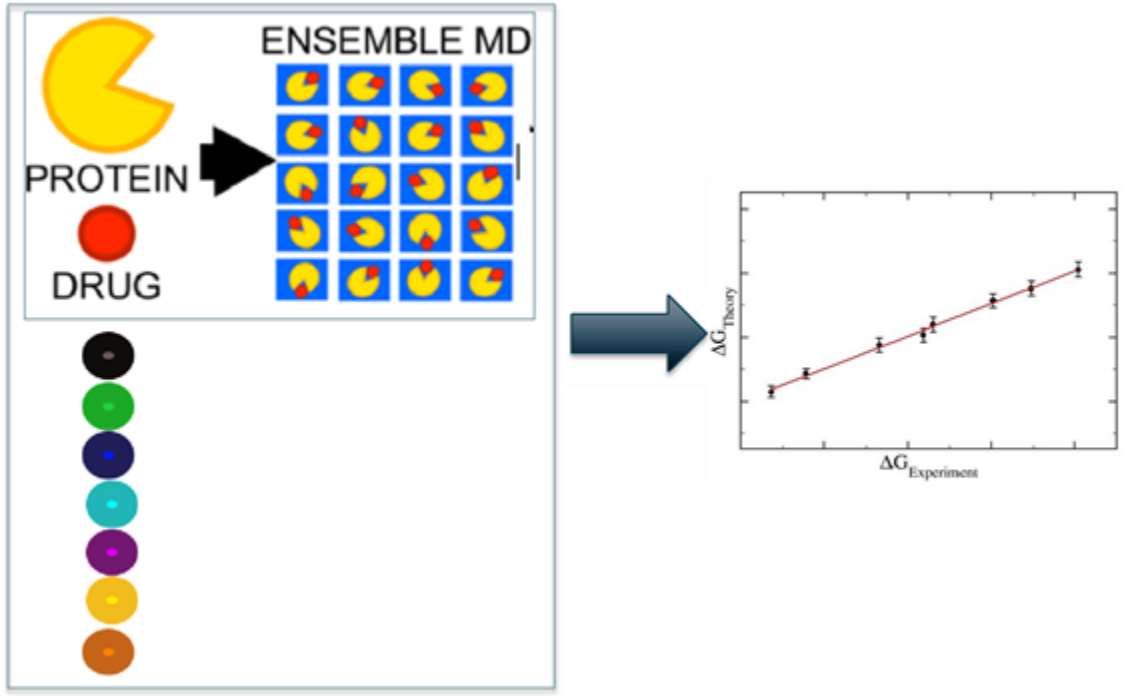
Error, uncertainty and reproducibility

One-off Simulations



Errors uncontrolled;
Results unreliable and unreproducible.

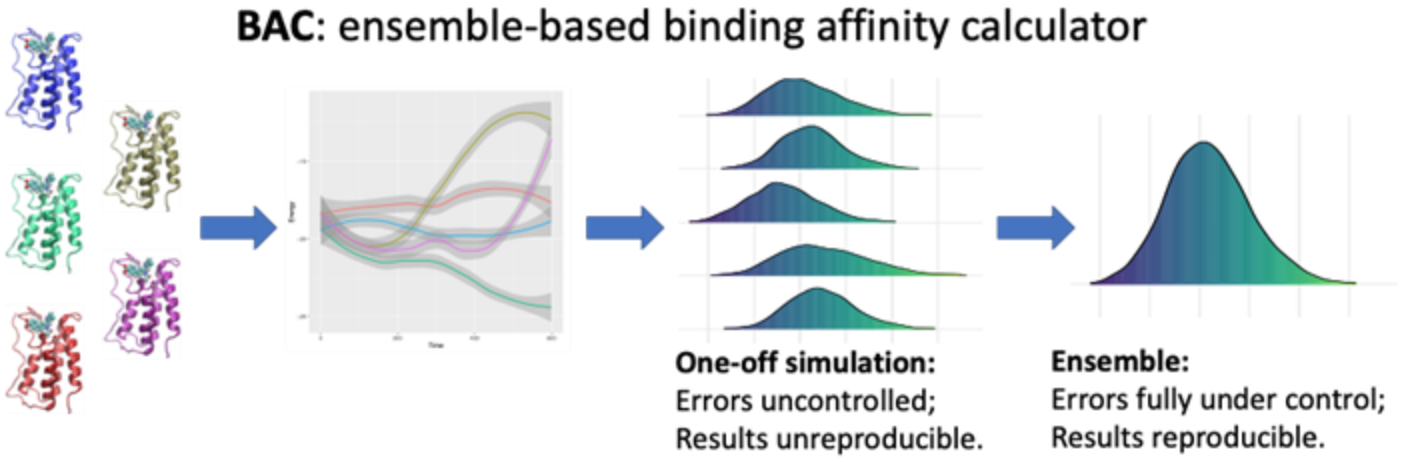
Ensemble Simulations



Errors fully under control;
Results reliable and reproducible.

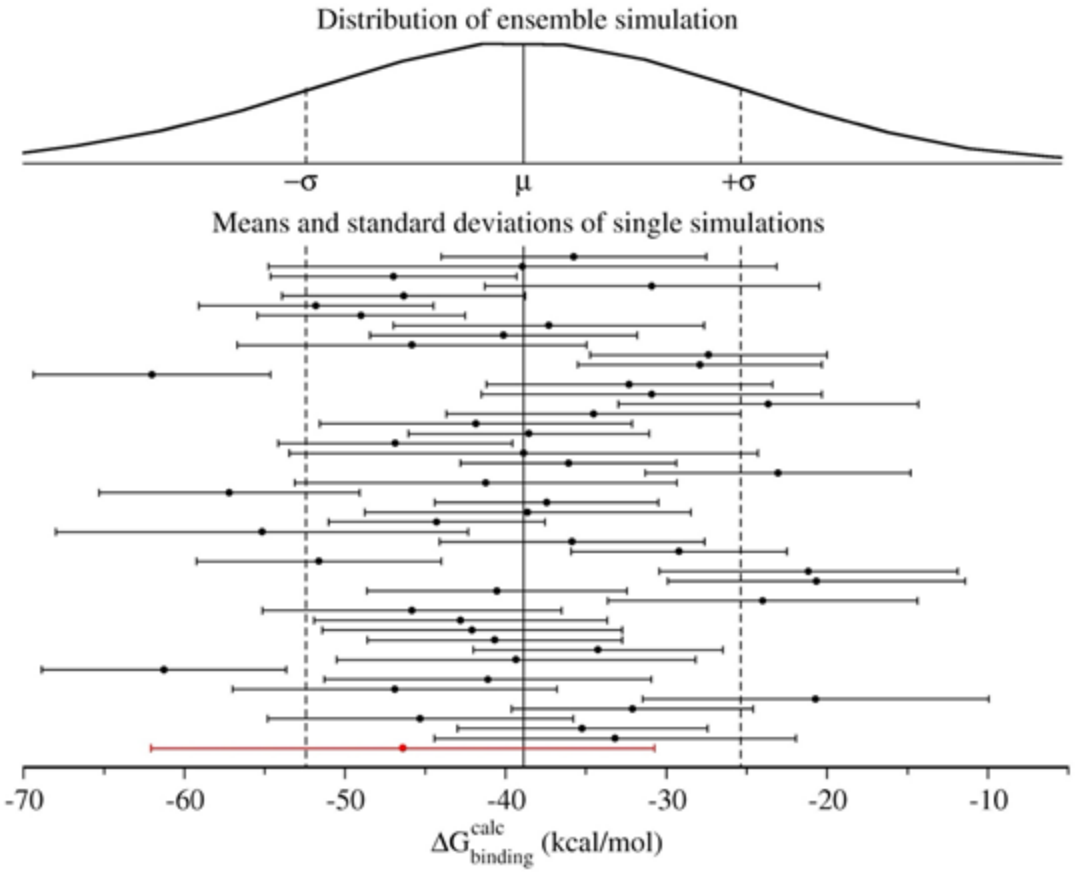
Ensemble simulations

- Performing ensemble simulations and obtaining averages leads to more reliable results
- Ensemble averaging produces robust statistics from chaotic simulations, e.g. in drug affinity ranking



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

An ensemble is far, far more effective and computational efficient than seeking to do it from a single one-off trajectory.



Coveney, Wan. *Phys. Chem. Chem. Phys.* (2016) 18, 30236-30240

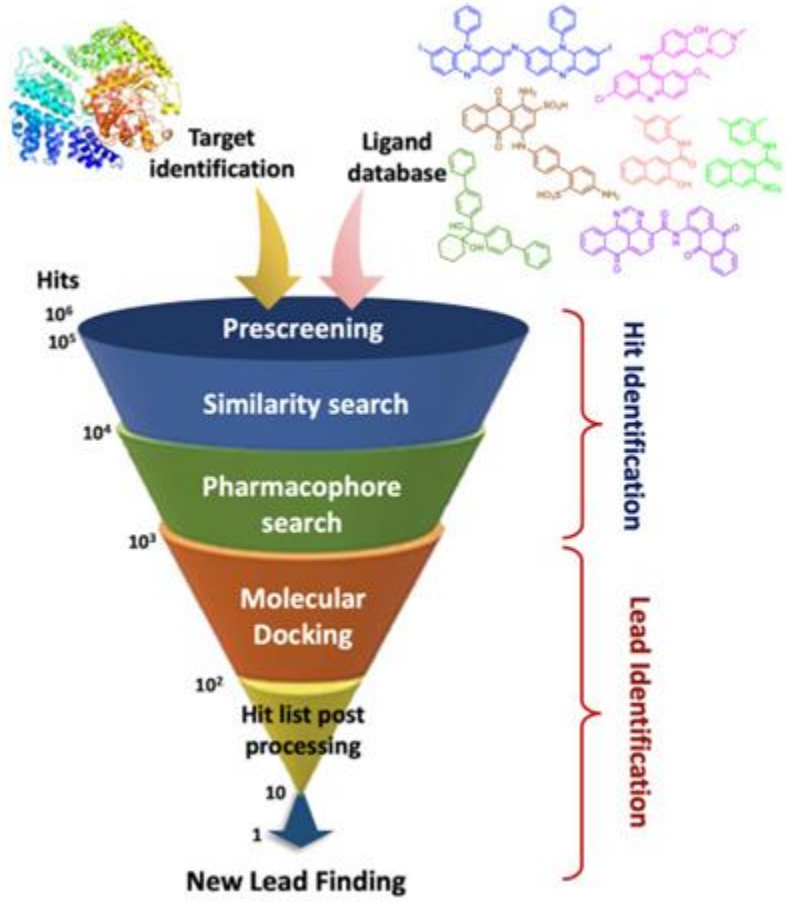
Ensembles ensure actionable predictions

To make routine and actionable predictions, for personalised medicine or for drug development, the predictions need to be *accurate, precise, and time-bound.*

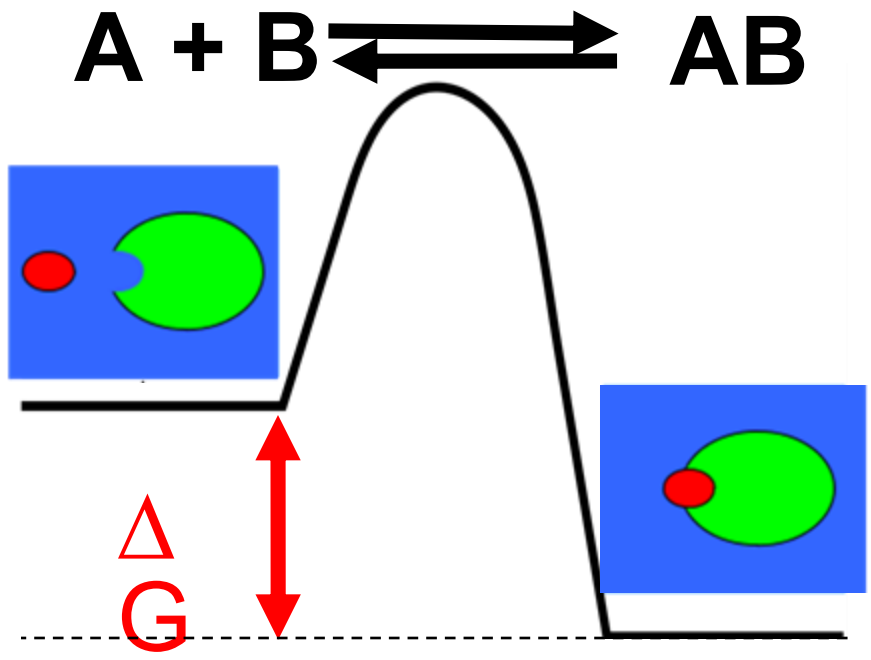
Ensemble approaches ensures accuracy, precision and **quick turnaround.**

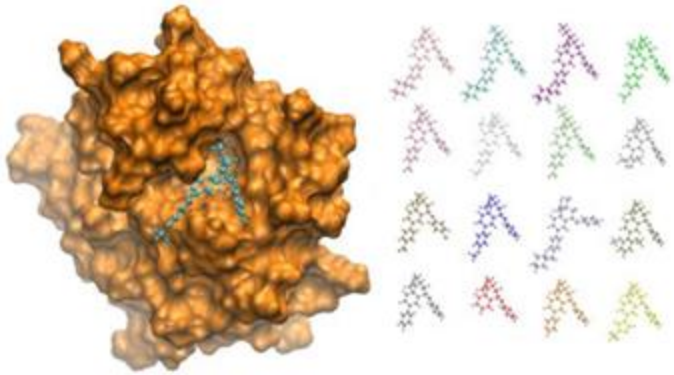
Actionable prediction in drug discovery and personalised medicine

- Better prediction methods, better force fields and better sampling to generate *probabilistic* predictions
- Fast computers to deliver results in a restricted “time-critical” window of time.
- Accurate binding affinity ranking to select the right drug for the right patient
- Choose the best drug candidates for the patient specific target in drug discovery



Binding free energy





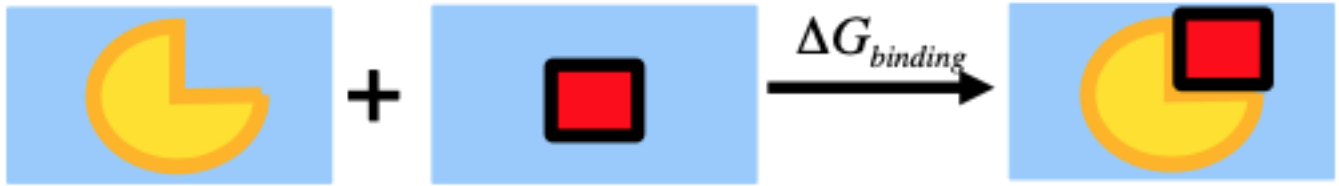
- Ligand binding driven by changes Gibbs free energy
- More negative ΔG = stronger binding

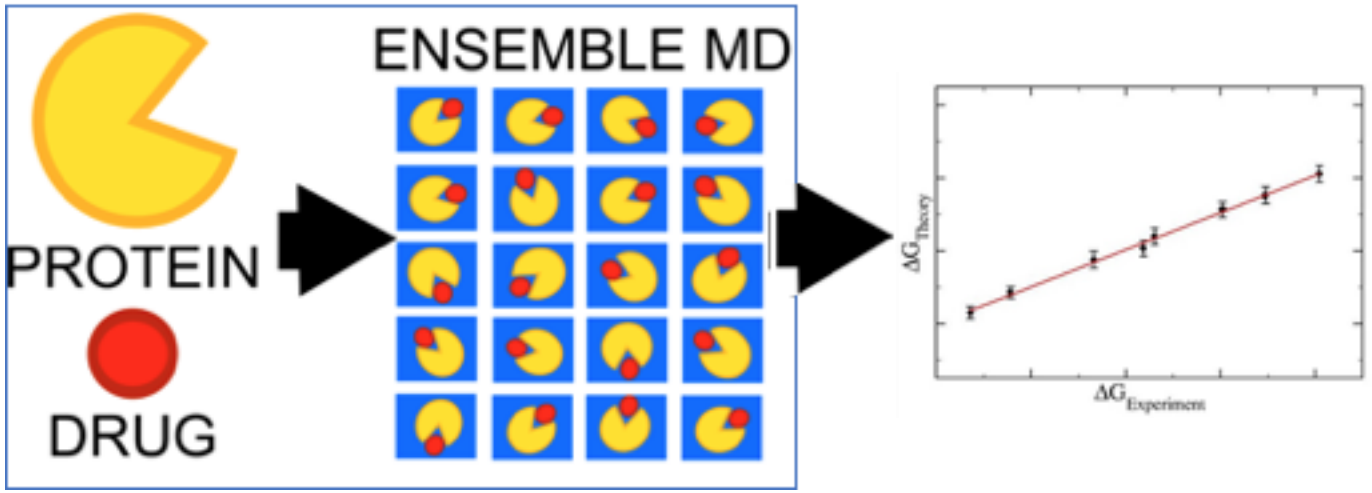
Absolute binding FE: end-point approach

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

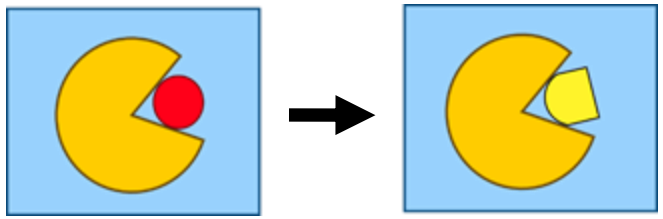


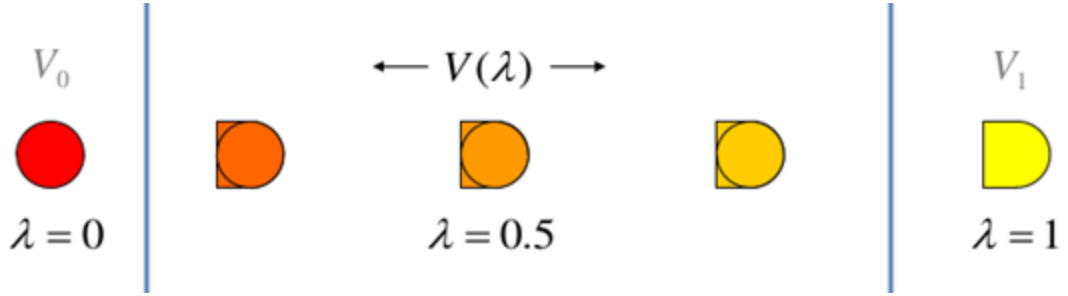


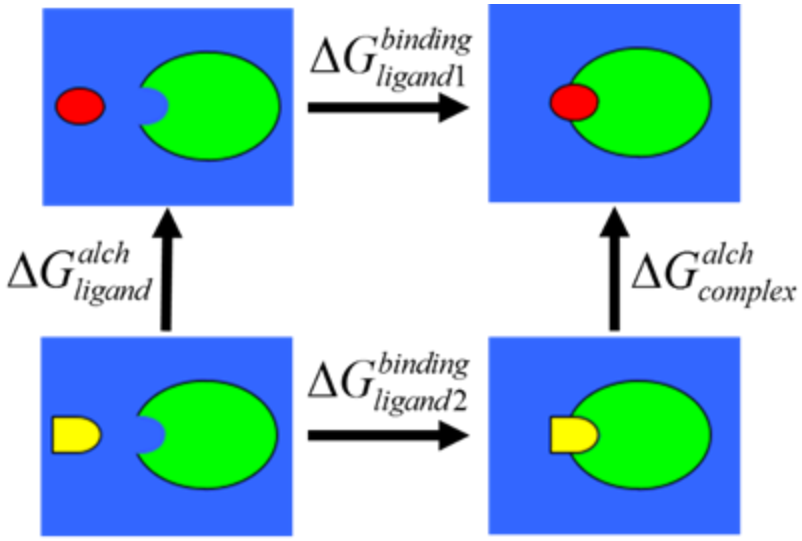
S. Wan, B. Knapp, D. Wright, C. Deane, P. V. Coveney, J. Chem. Theory Comput., 11 (7), 3346-3356 (2015)

Relative binding FE: alchemical approach

Alchemical methods







$$\Delta\Delta G^{binding} = \Delta G^{binding}_{ligand2} - \Delta G^{binding}_{ligand1} = \Delta G^{alch}_{ligand} - \Delta G^{alch}_{complex}$$

Thermodynamic integration (TI) :

$$\Delta G = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Free energy perturbation (FEP) [3]:

$$\Delta G(\lambda_i \rightarrow \lambda_{i+1}) = -k_B T ln \left\langle exp\left(-\frac{V(\lambda_{i+1}) - V(\lambda_i)}{k_B T}\right)\right\rangle_{\lambda_i}$$

TIES: thermodynamic integration with enhanced sampling [1]:

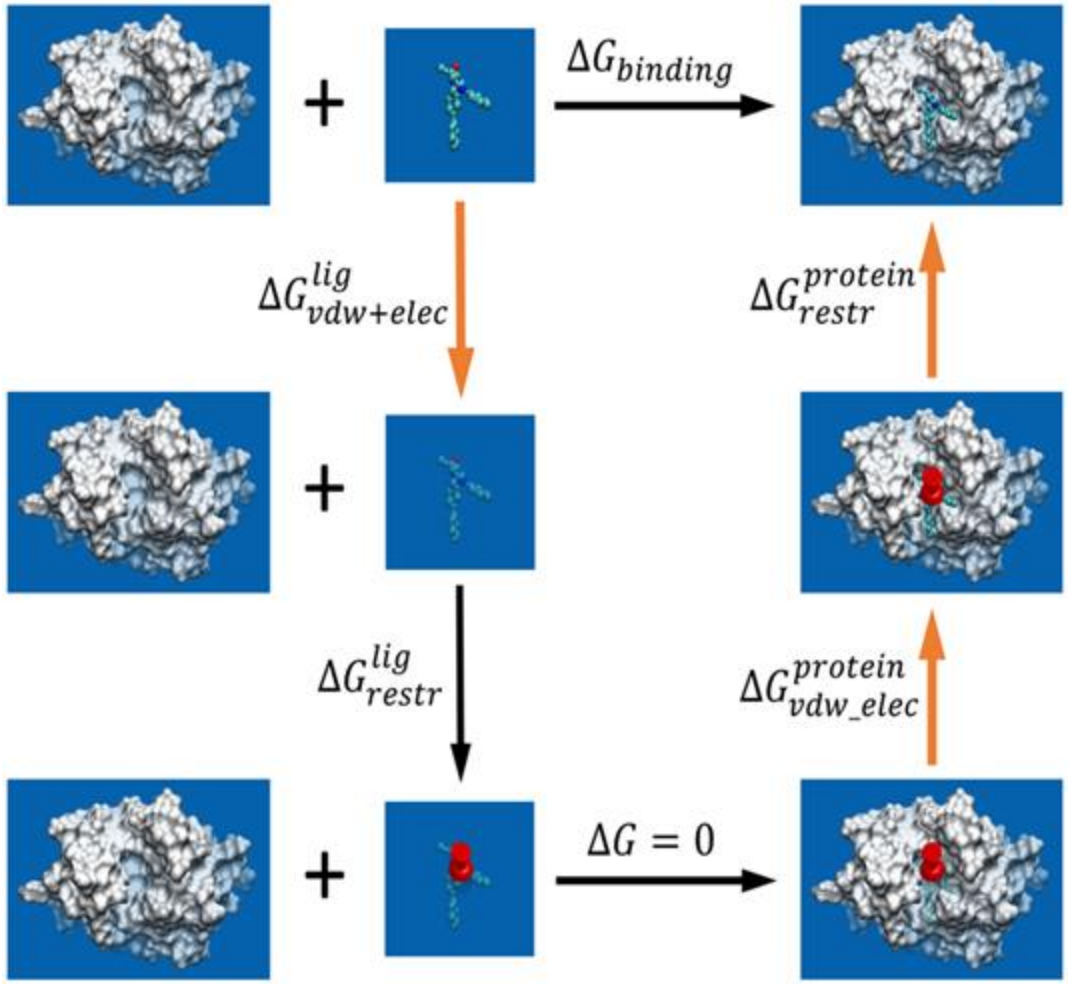
Use of ensemble averaging and the assuming a Gaussian random process (GRP), properties are computed from MD trajectories [2].

[1] A. Bhati, S. Wan, D. Wright, P. V. Coveney, *J. Chem. Theory Comput.*, **2017**, 13, 210–222; [2] P.V. Coveney & S. Wan, Phys. Chem. Chem. Phys., **2016**, 18, 30236-30240; [3] L. Wang, *et al*, *J. Am. Chem. Soc.* **2015**, 137, 2695−2703.

Absolute binding FE: alchemical approach

- 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.

ABFE is ~**10 times more expensive** than RBFE



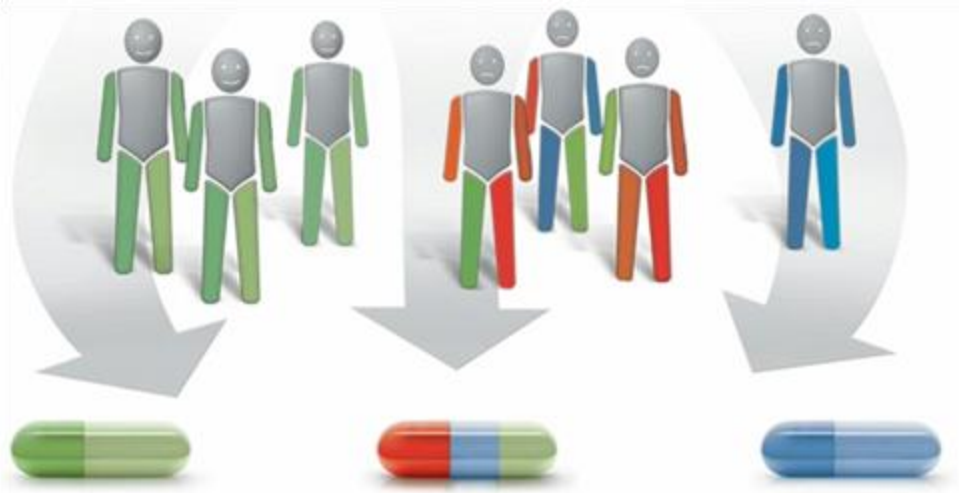
A. P. Bhati, S. Wan, P. V. Coveney, "Equilibrium and Non-equilibrium Ensemble Methods for Accurate, Precise and Reproducible Absolute Binding Free Energy Calculations", ChemRxiv (2024) DOI: 10.26434/chemrxiv-2024-sslzp; *J. Chem Theory & Computation*, in press (2024).

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

Real world applications

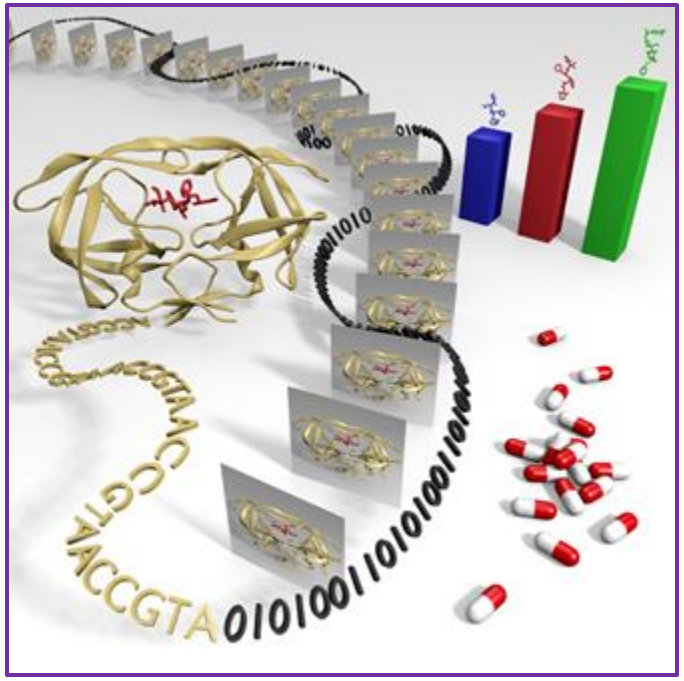
Personalised Medicine

Patient specific computer models are being developed to predict the impact of treatments, improving clinical outcomes. Here I will talk about one such scenario: accurate binding affinity ranking to select the right drug for the right patient.



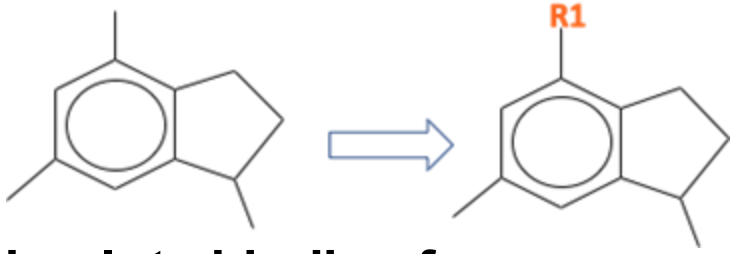
Virtual Drug Discovery

- Testing various candidate drugs against a target protein can be done through computational models
- Candidates can be ranked according to their ability to interact with the target — their binding affinity.
- Best candidates then chosen for the patient specific target

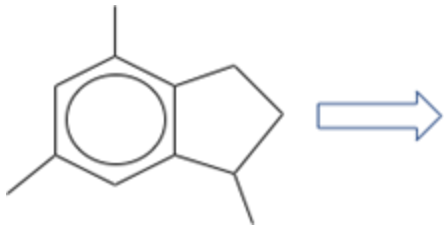


TIES Toolkit (openly available)

Relative binding free energy



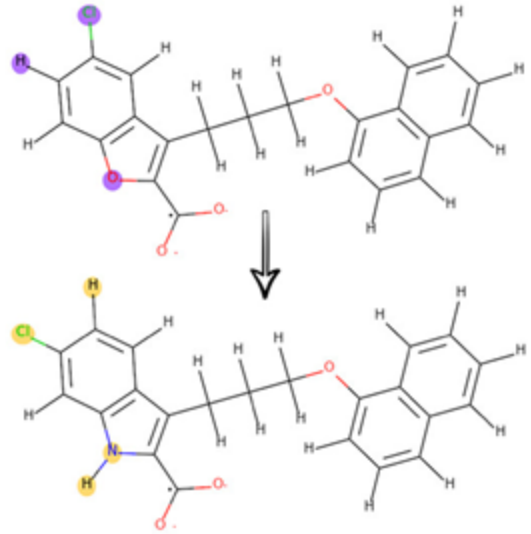
Absolute binding free energy



TIES 20: Relative Binding Free Energy with a Flexible Superimposition Algorithm and Partial Ring Morphing.

https://ccs-ties.org/

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

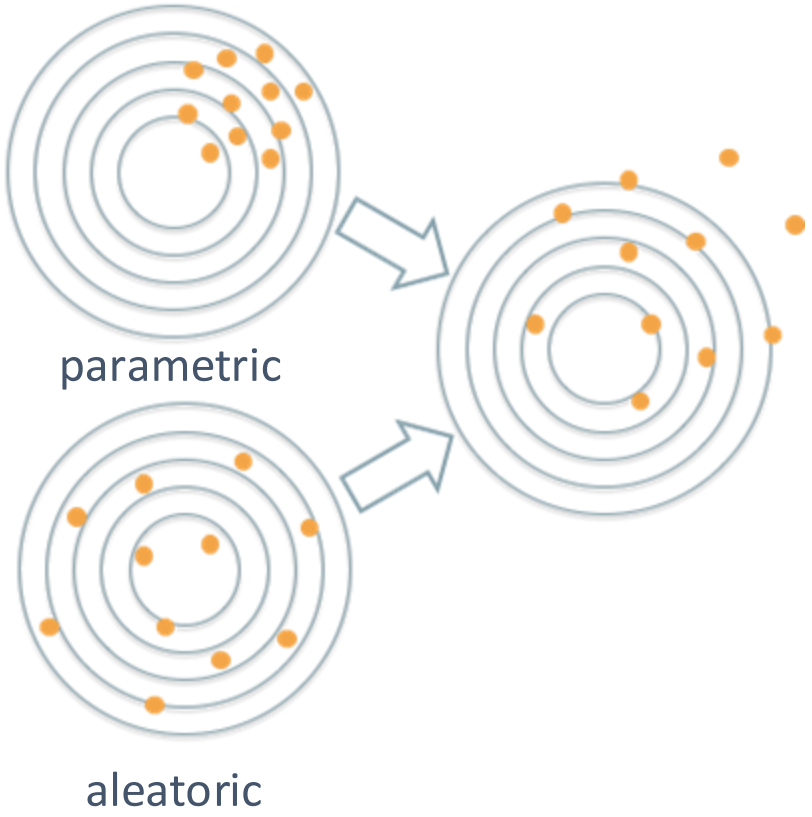


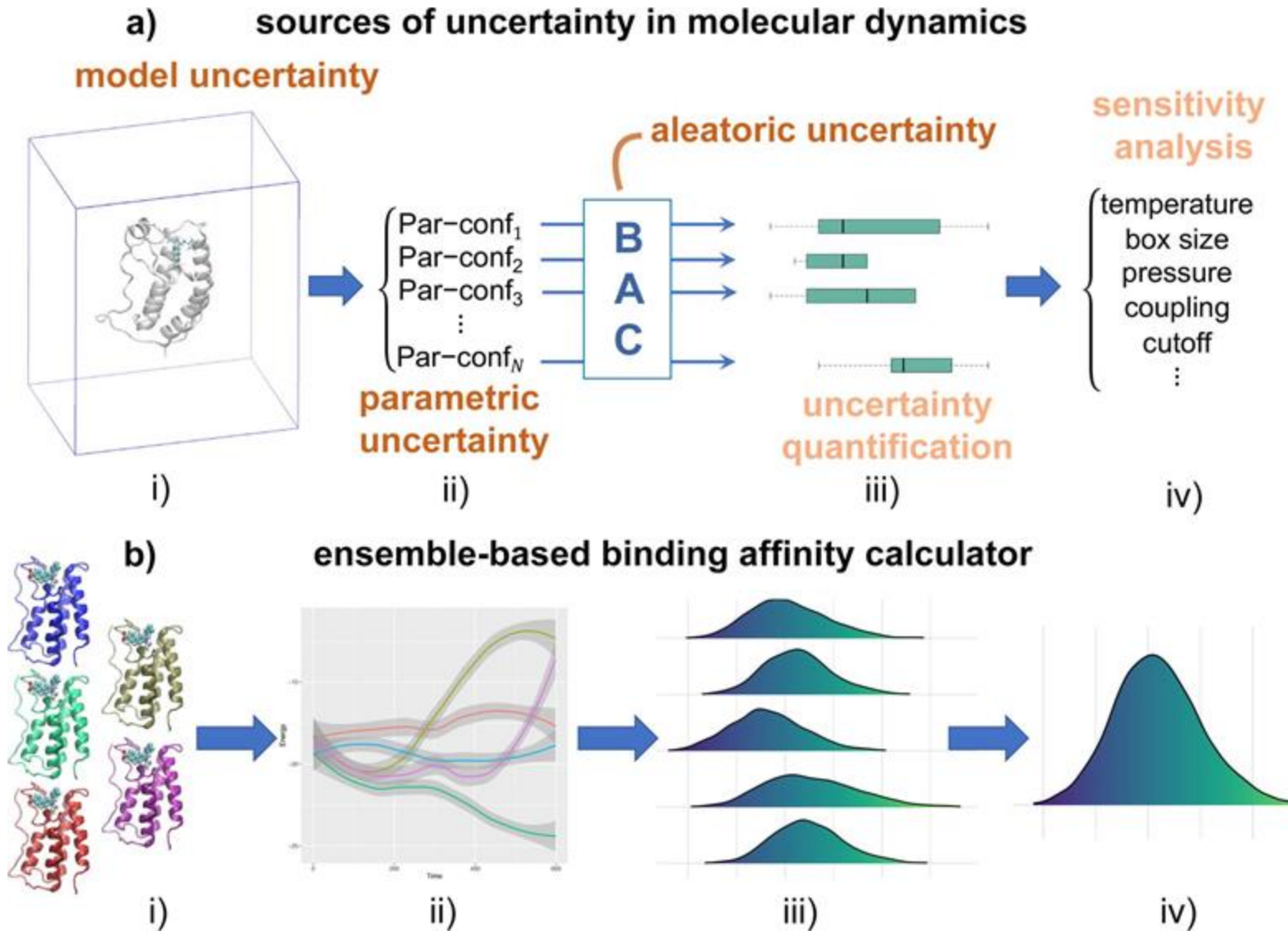
TIES MD: a collection of software packages to calculate protein ligand binding free energies with physics based alchemical methods.

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.

Sources of uncertainty

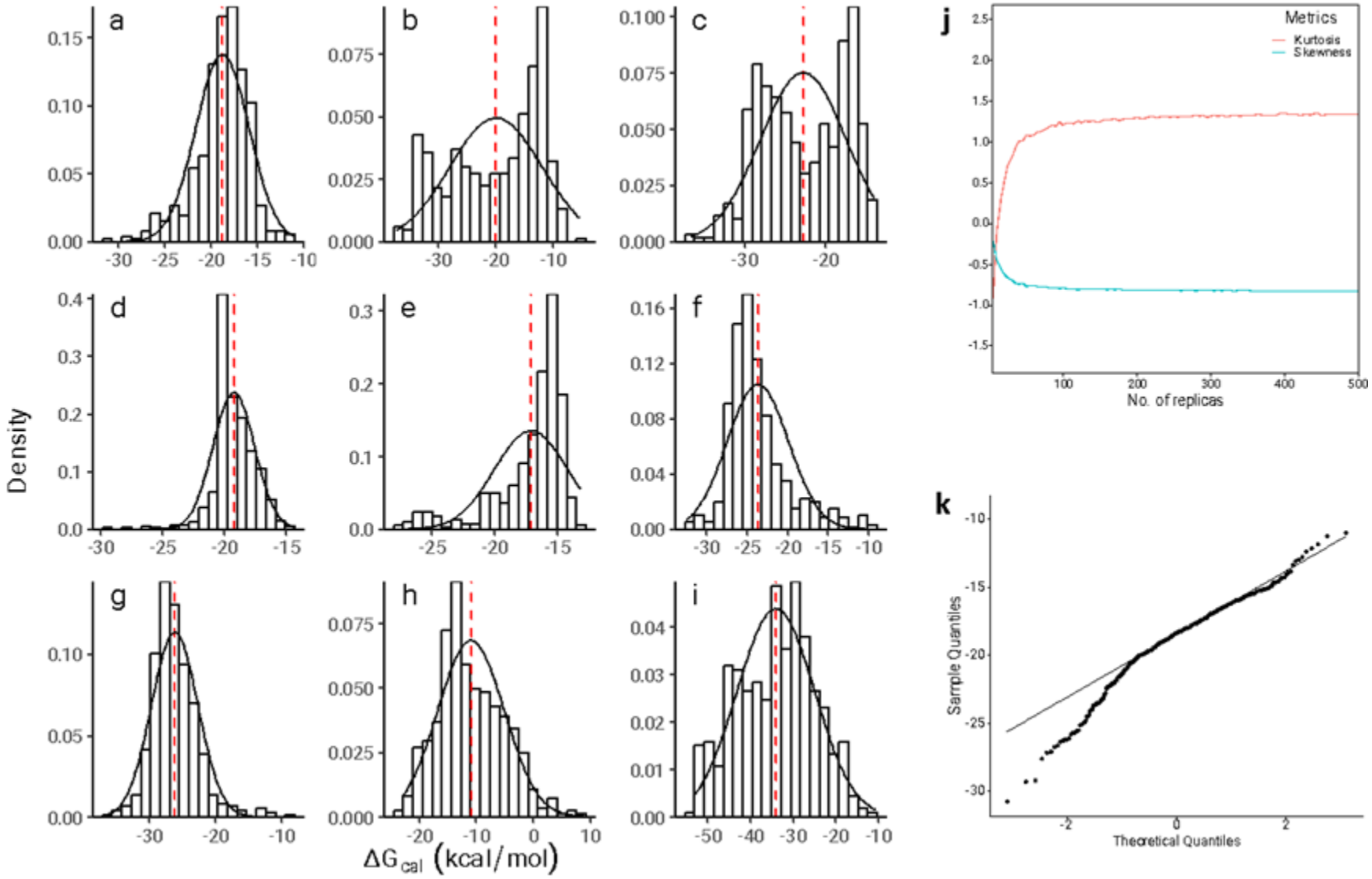




Vassaux et al., *J. Chem. Theory Comp.* (2021) 17 (8), 5187–5197

Statistical probability distributions

Non-Gaussian distributions are common in MD simulation



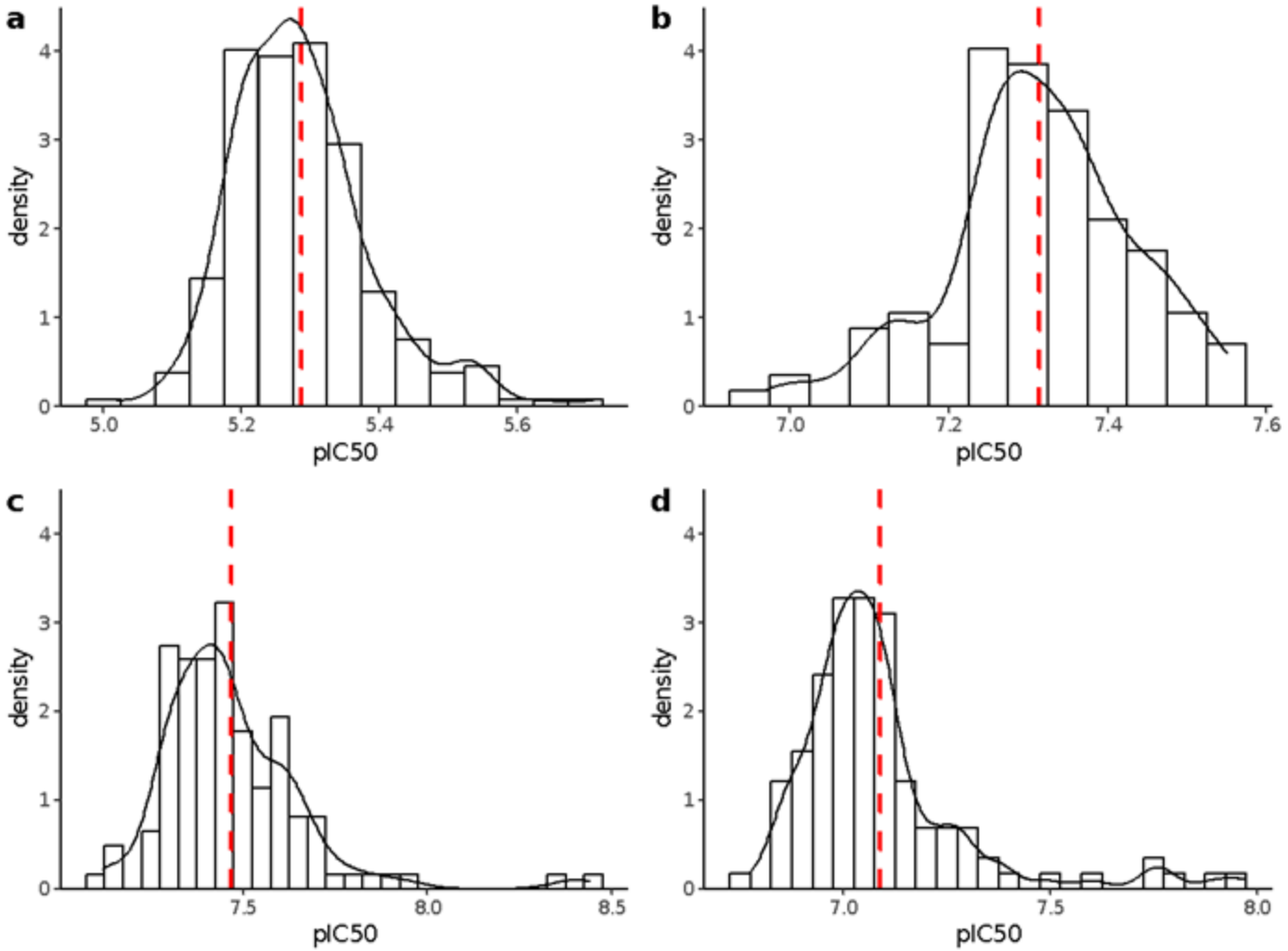
S. Wan, A. Bhati, A. Wade, P. V. Coveney, "Ensemble-Based Approaches Ensure Reliability and Reproducibility", *J. Chem. Inf. Model.*, 2023, 63, 22, 6959–6963.

Statistical probability distribution

… *and now found experimentally*

Experimental binding affinity measurements from GSK

- Compounds measured >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.



S. Wan, *et al*., J. Chem. Inf. Model., **62**, 2561-2570 (2022)

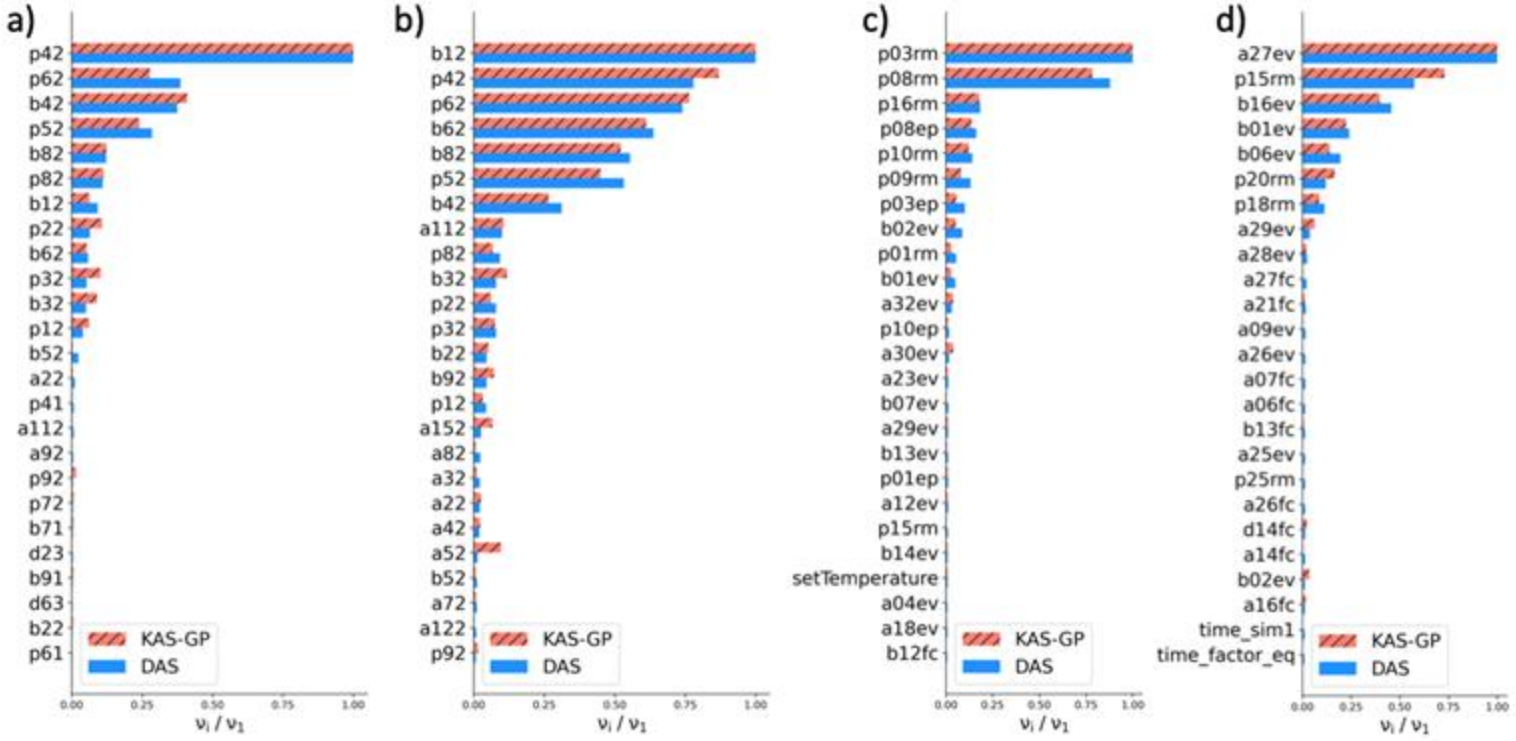
Global ranking of the importance of force field parameters

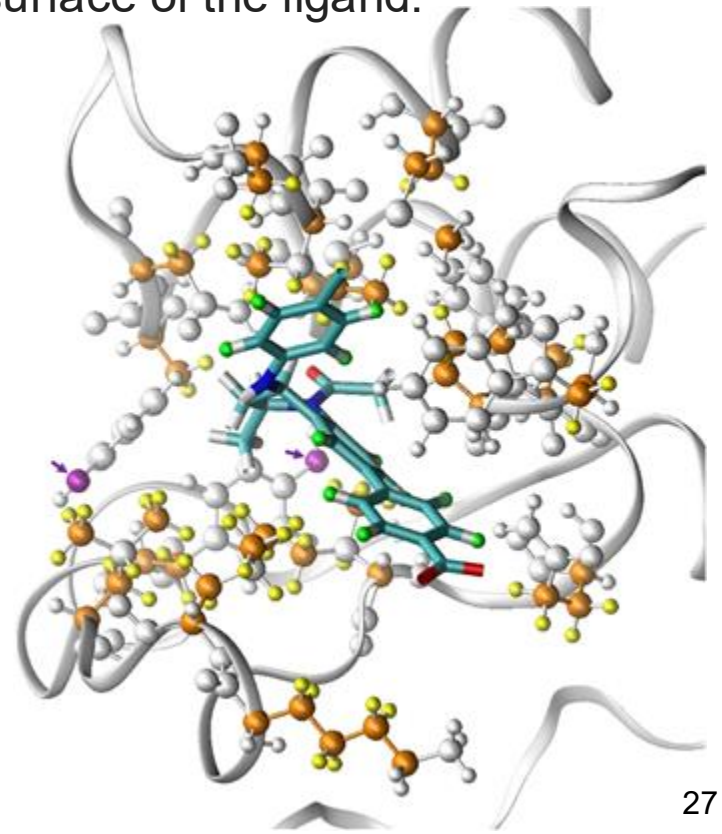
Identify the sources of the dominant contributions to QoIs

Most of the important ESMACS parameters are `pNNrm` and `pNNev`, representing the pairwise equilibrium internuclear distance and well depth of vdW interactions for atoms with index `NN`.

These parameters are the most sensitive ones because:

- Many atoms close to the ligand are carbon and hydrogen atoms with these parameters.
- Most rotatable bonds involve sp^3 carbon atoms, which are mainly affected by these parameters.
- The third parameter is for the most common atom type (hydrogen) on the surface of the ligand.

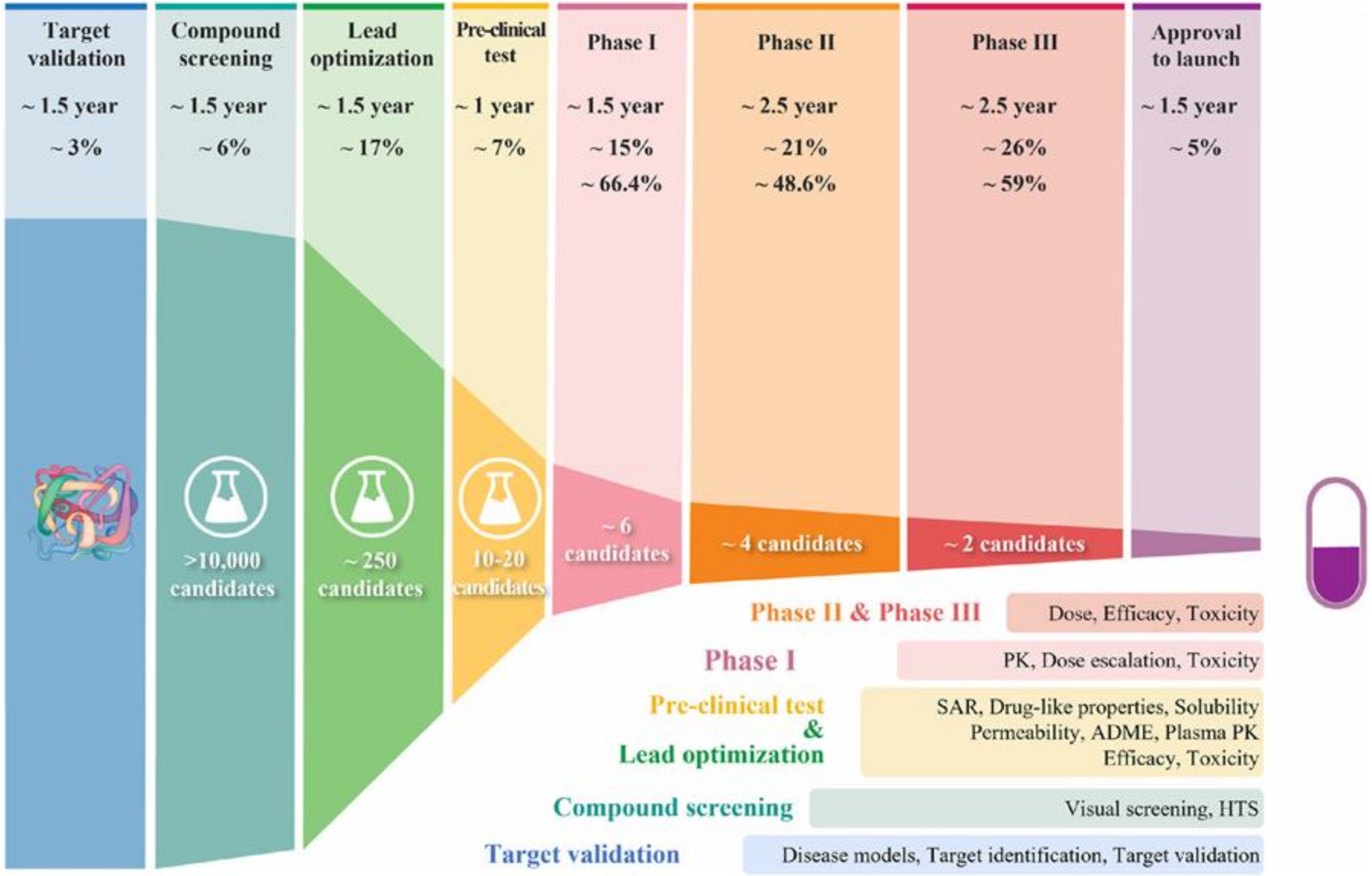




W. Edeling, M. Vassaux, Y. Yang, S. Wan, S. Guillas, P. V. Coveney, "Global ranking of the sensitivity of interaction potential contributions within classical molecular dynamics force fields". npj Comput. Mater. 10, 87 (2024) DOI: 10.1038/s41524-024-01272-z

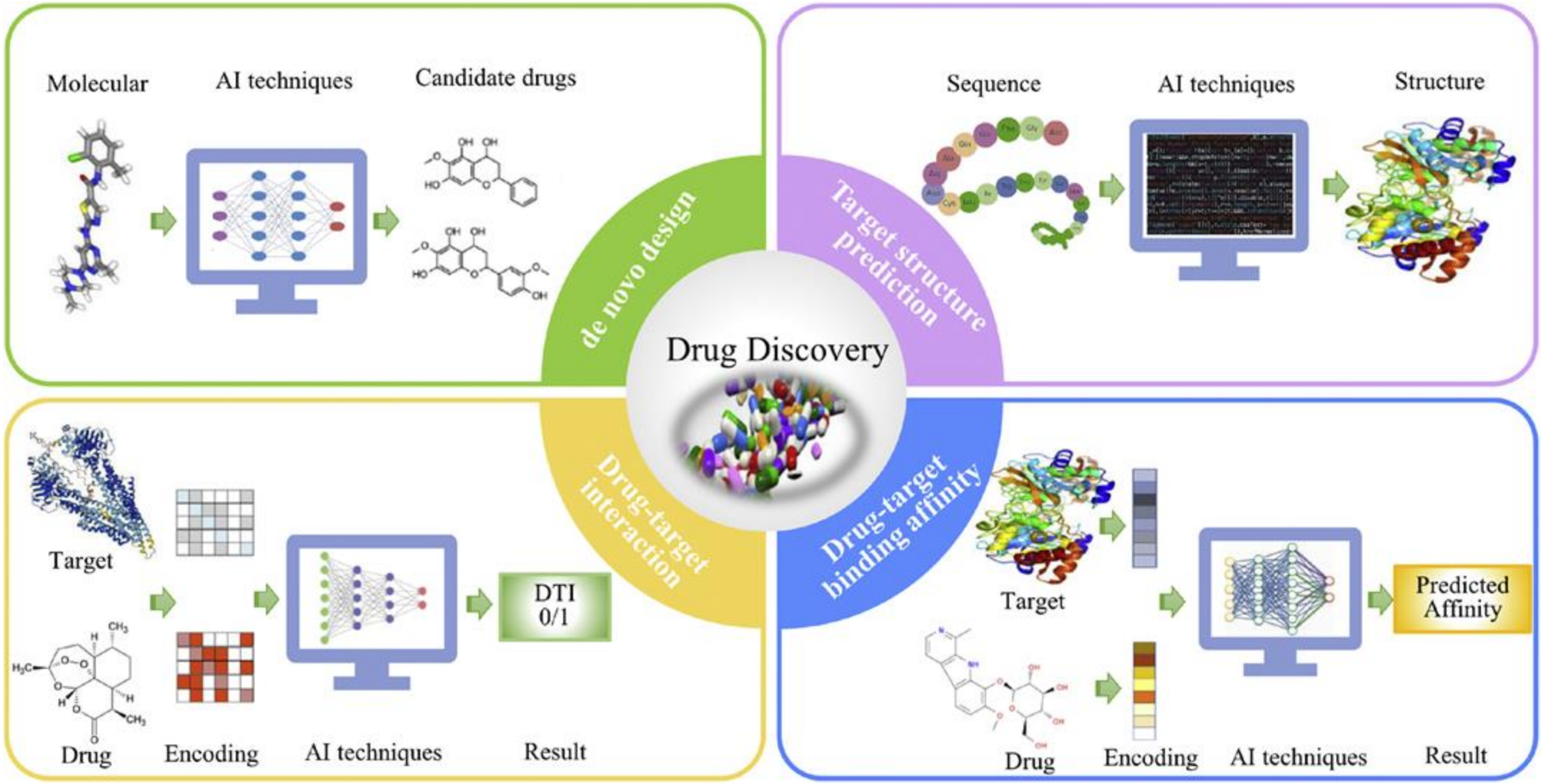
Drug discovery, development and market pipel

FDA classifies a drug as a *New Molecular Entity (NME)* if it contains an active ingredient that has not been previously approved by the FDA.



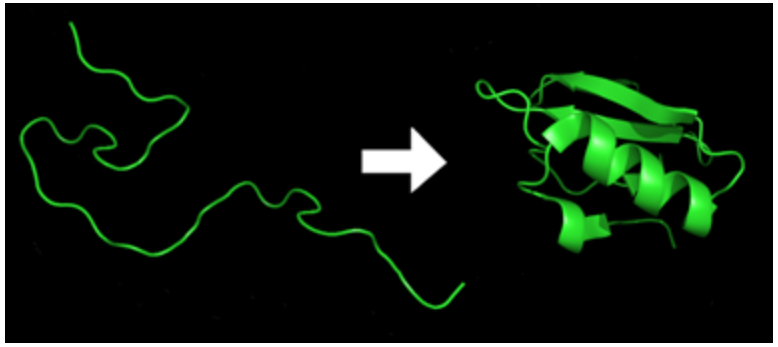
D. Sun, W. Gao, H. Hu, S. Zhou, Why 90% of clinical drug development fails and how to improve it, *Acta Pharmaceutica Sinica B*, 12(7), 3049-3062, 2022.

Artificial intelligence in drug discovery



W. Chen, X. Liu, S. Zhang, S. Chen, "Artificial intelligence for drug discovery: Resources, methods, and applications" *Molecular Therapy Nucleic Acids*, 31, 691–702, **2023**.

AlphaFold for protein structure prediction



Primary (Sequence)

⇩

Secondary

⇩

Tertiary

⇩

Interactions/assemblies

⇩

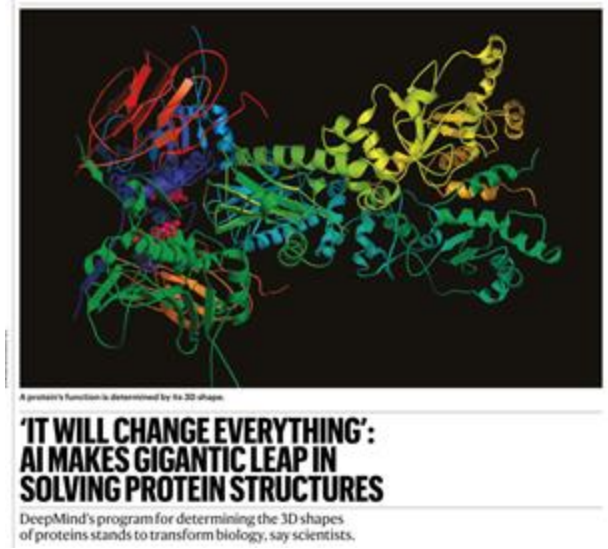
Functions

How AlphaFold works:

- Couch the protein folding problem in mathematical terms – a “spatial graph” showing the amino acids and other molecules close in space
- Design the structure of the network to learn from training what amino acids and other molecules lie near each other

Accurate predictions for primary sequences which are closely similar to the ones AF has been trained on, with no idea of when others will be reliable or simply fail.

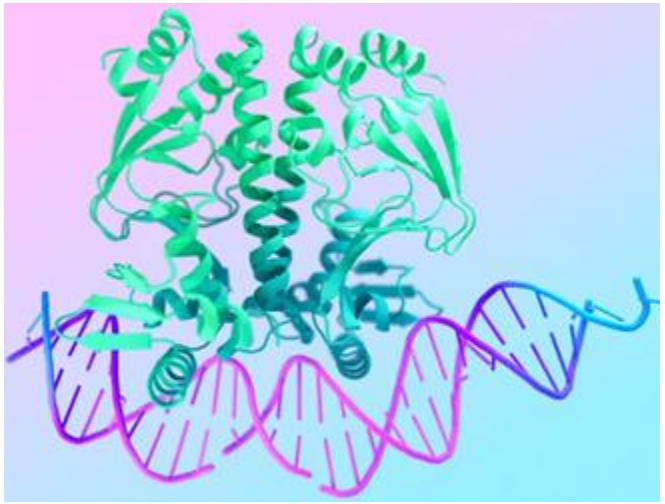
AlphaFold 2



‘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.

AlphaFold 3



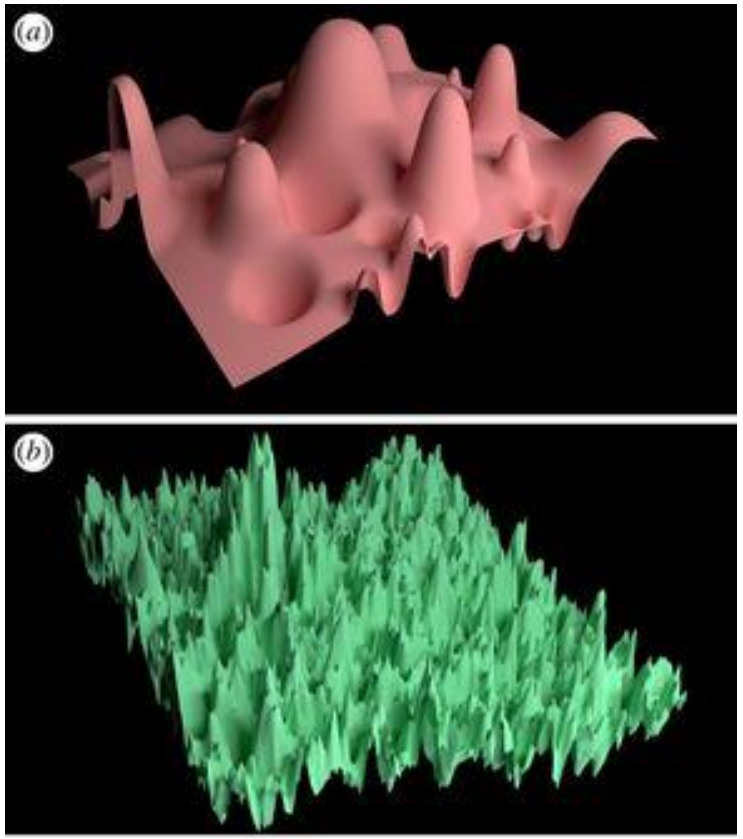
Can AlphaFold3 predict interactions?

AlphaFold3 Server to model e.g. protein-protein structures and interactions:

"AlphaFold3 captures a more global effect of mutations by learning a *smoother energy landscape*, but it lacks the modelling of full atomic details that are better addressed by force field methods, which possess a more *rugged energy landscape"*.

Comparison of $\Delta\Delta G$ estimation results on our SKEMPI test set using three different metrics.

Category	Method	Pearson	Spearman	AUC
Force Field and Profile-based	SSIPe	0.68	0.62	0.78
	FlexddG	0.62	0.58	0.77
	BindProfX	0.56	0.58	0.74
	EvoEF	0.55	0.51	0.72
	FoldX	0.49	0.54	0.74
Structure-based Deep Learning	DSMBind	0.62	0.53	0.73
	ProteinMPNN	0.51	0.45	0.65
AlphaFold	AF3 ranking_score	0.49	0.51	0.71
	AF3 iptm	0.49	0.50	0.72
	AF3 ptm	0.36	0.33	0.63
	AF3 mean_pae	0.32	0.37	0.64
	AF2 ranking_score	0.21	0.23	0.57
	Effective Strain	0.18	0.31	0.61
	AF2 mean_pae	0.05	0.22	0.54
Protein Language-based	ESM2	0.27	0.35	0.68
	ESM1v	-0.02	0.06	0.52
	ProGen2	-0.09	0.01	0.47

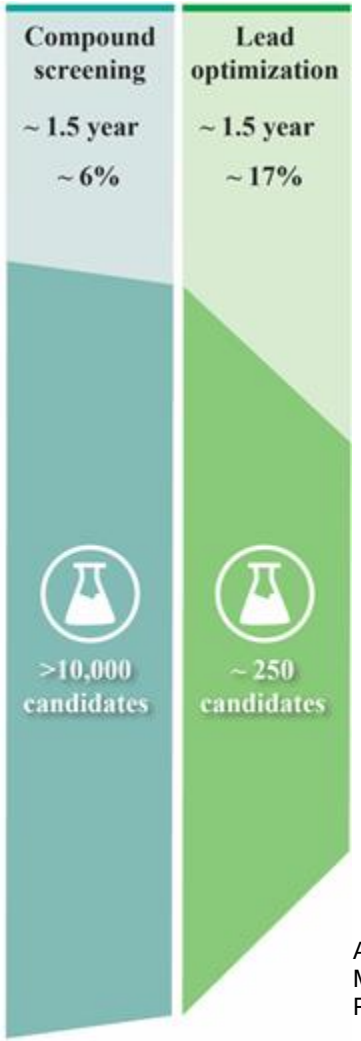


A more complete picture of biomolecular interactions can be captured by physics-based molecular dynamics simulations.

AlphaFold3, a secret sauce for predicting mutational effects on protein-protein interactions,
bioRxiv preprint https://doi.org/10.1101/2024.05.25.595871

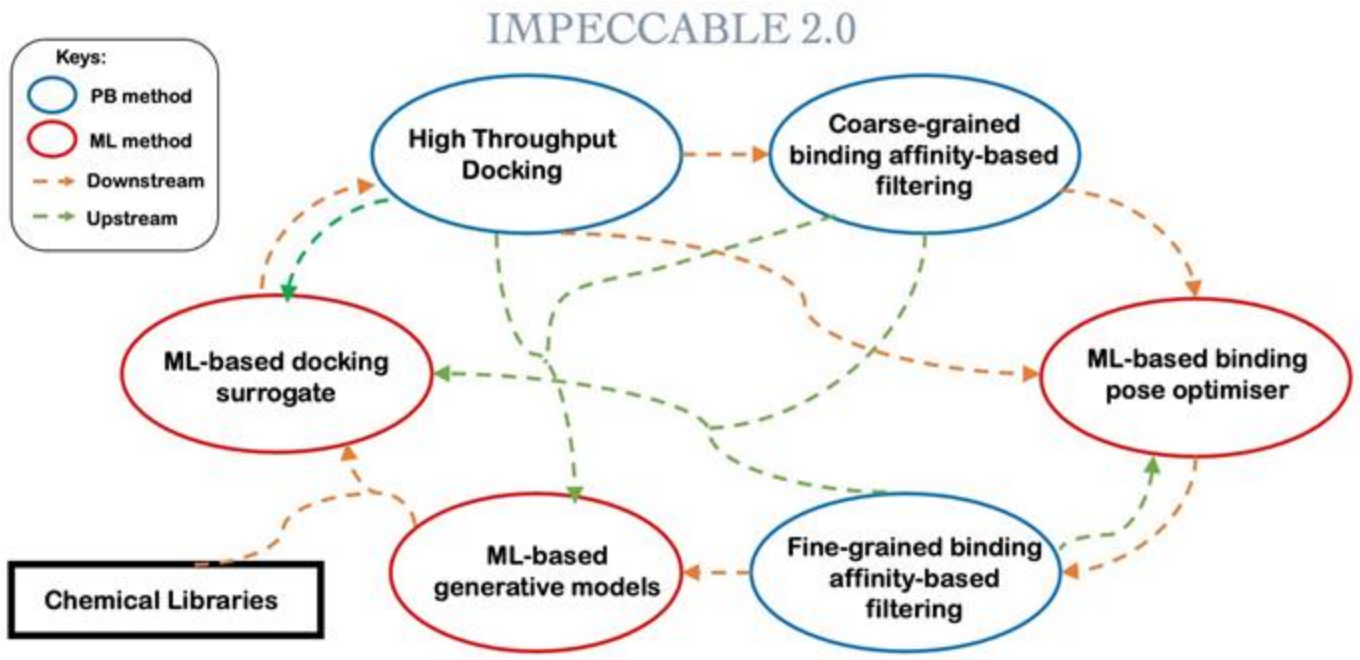
IMPECCABLE workflow

Integrate machine learning and physics-based methods to accelerate drug discovery



The full preclinical workflow

- No single algorithm or method can achieve the necessary accuracy with required efficiency to sample such huge chemical spaces.
- **Physics and machine learning based methods are used symbiotically** to test drugs with required accuracy and efficiency.
- The pipeline combines ML and PB into a unified workflow, allowing both upstream and downstream exchange of information in the iterative loop.



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

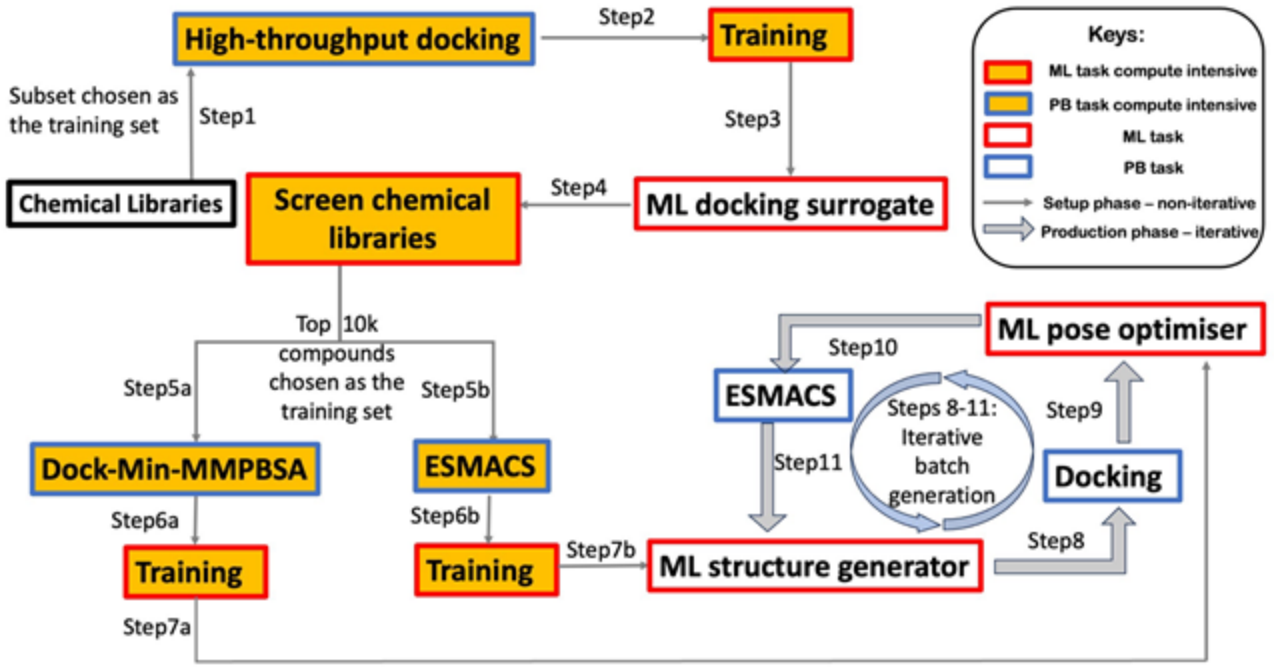
IMPECCABLE

Physics-based methods:

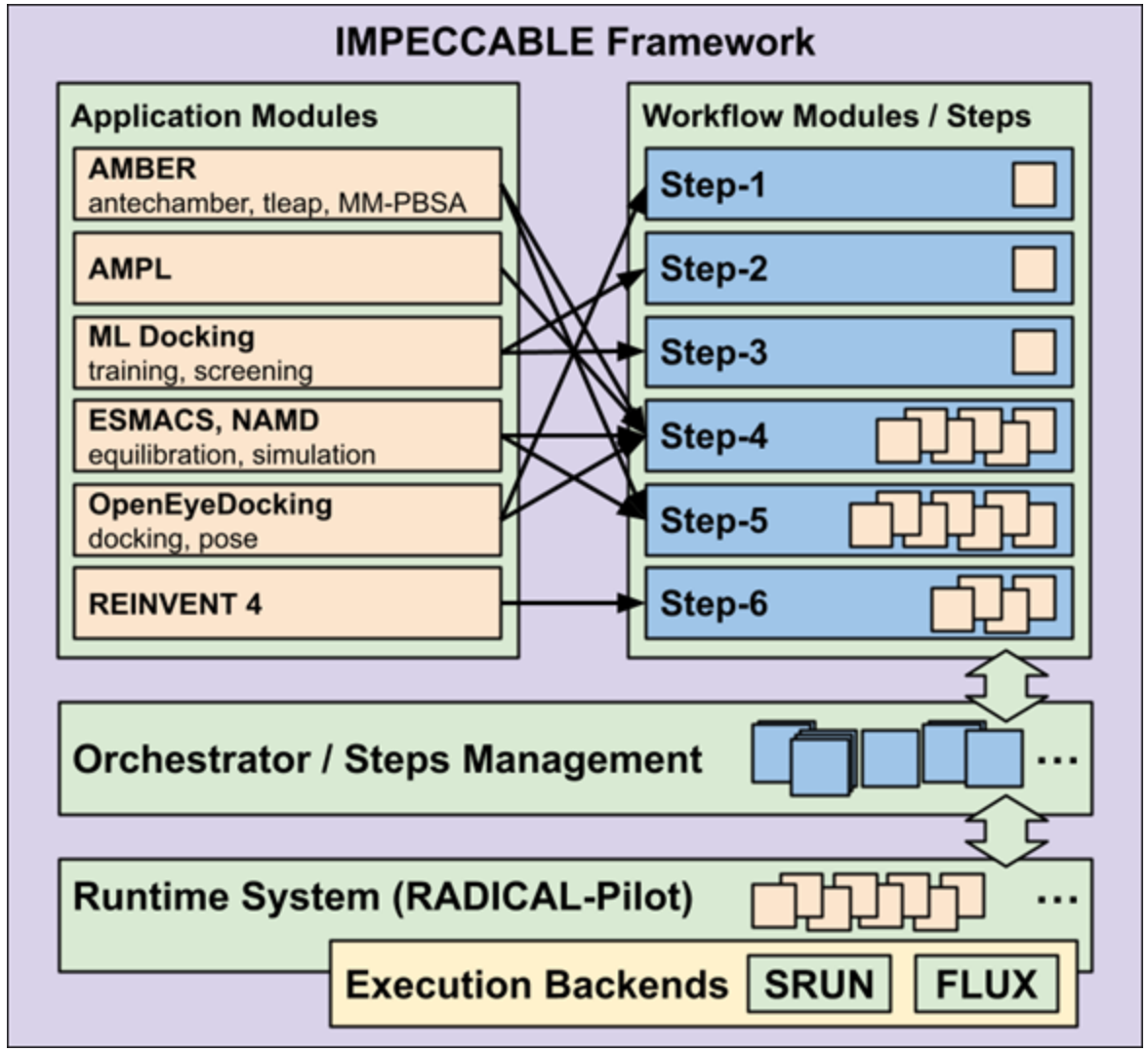
- **Docking**: structure generation for ligand-protein complexes
- **ESMACS**: precise and reliable approximation of absolute binding free energies for *compound screening*
- **LOMAP**: plan efficient relative free energy calculations between potential ligands
- **TIES**: accurate, precise and reliable relative binding free energies for *lead optimisation*

Machine learning methods:

- **REINVENT** 2.0 and ChemProp: generative AI to produce small molecules with optimal binding affinities, synthesisability, toxicity, etc.
- **Docking surrogate model**: AI-accelerated protein-ligand docking
- **AMPL**: ligand pose optimisation model



IMPECCABLE-2



Application Modules: *Application modules* represent application executables, tools, and functions. These are building blocks for IMPECCABLE workflows.

Workflow Modules: *Workflow modules* represent independent workflows as part of the IMPECCABLE campaign. Each workflow corresponds to a particular step of the campaign (e.g., high-throughput docking, training a surrogate model, ensemble simulations).

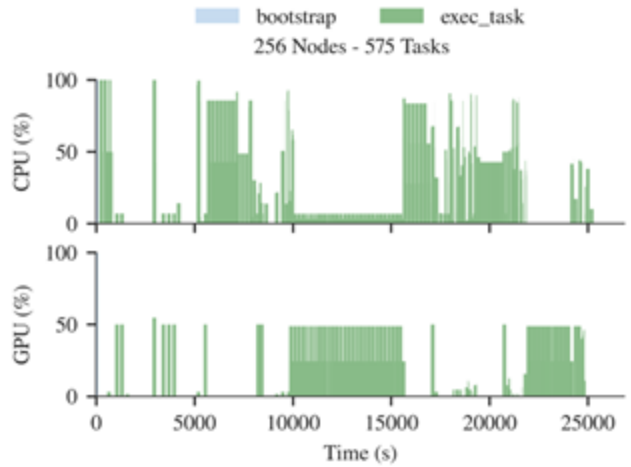
Orchestrator Component: *Orchestrator* governs the execution of a sequence of steps and adapts the initiation of a new sequence based on resources availability. It might launch multiple instances of a particular workflow.

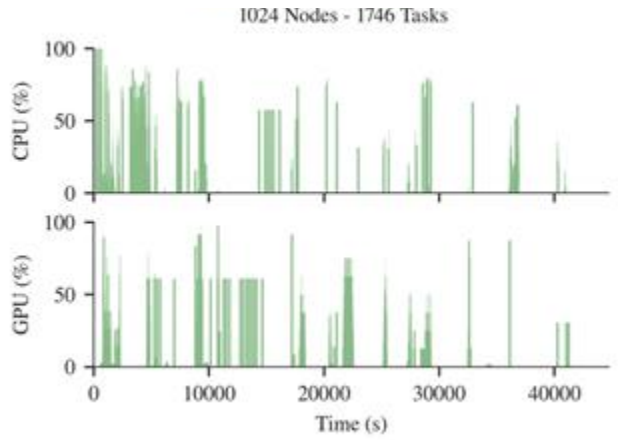
Task Runtime System: The task *runtime system* launches **heterogeneous** tasks, monitors their execution, and gather final states. It addresses limitations of the native resource manager.

IMPECCABLE-2 | **Effective resource occupation**

Resource occupation represents the percentage of resources being assigned to and occupied by tasks as a function of time. *NOTE: the actual resource utilization (CPU/GPU load) should be captured by a corresponding third-party tool.*

Using SRUN and FLUX as execution backends to run the IMPECCABLE campaign with null workloads on 256 and 1024 nodes. Runs with SRUN as a backend shows lower resource occupation (underutilized) in comparison to FLUX.

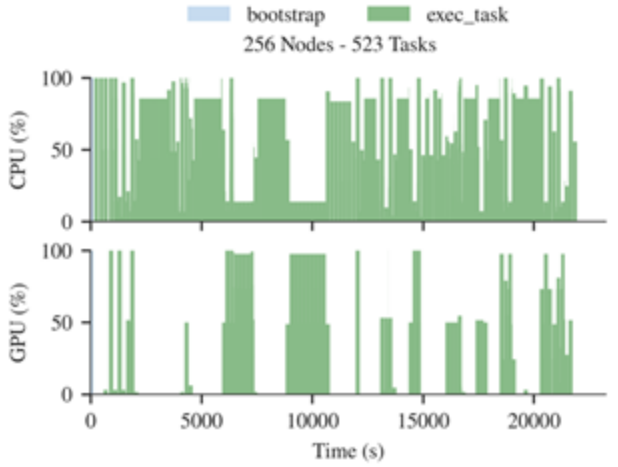


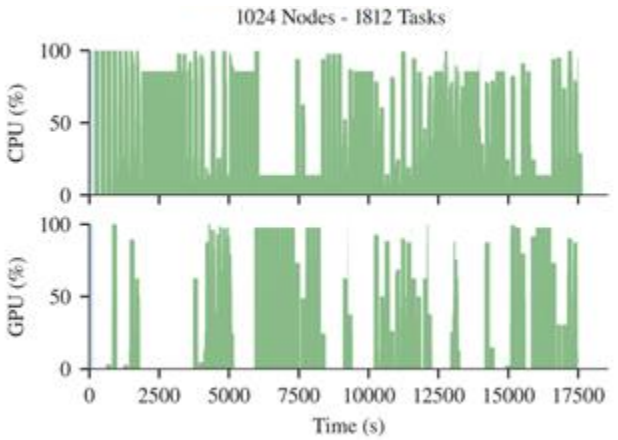


Resources occupation for SRUN runs

256 nodes
- CPU - 30%
- GPU - 20%

1024 nodes
- CPU - 15%
- GPU - 14%





Resources occupation for FLUX runs

256 nodes
- CPU - 68%
- GPU - 33%

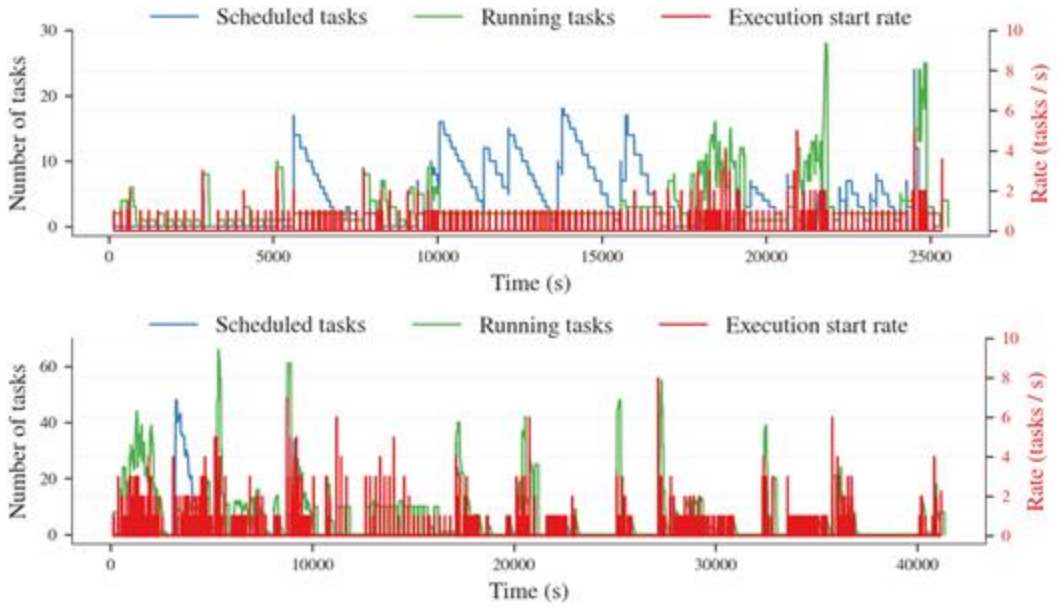
1024 nodes
- CPU - 69%
- GPU - 43%

IMPECCABLE-2 | Concurrency and throughput

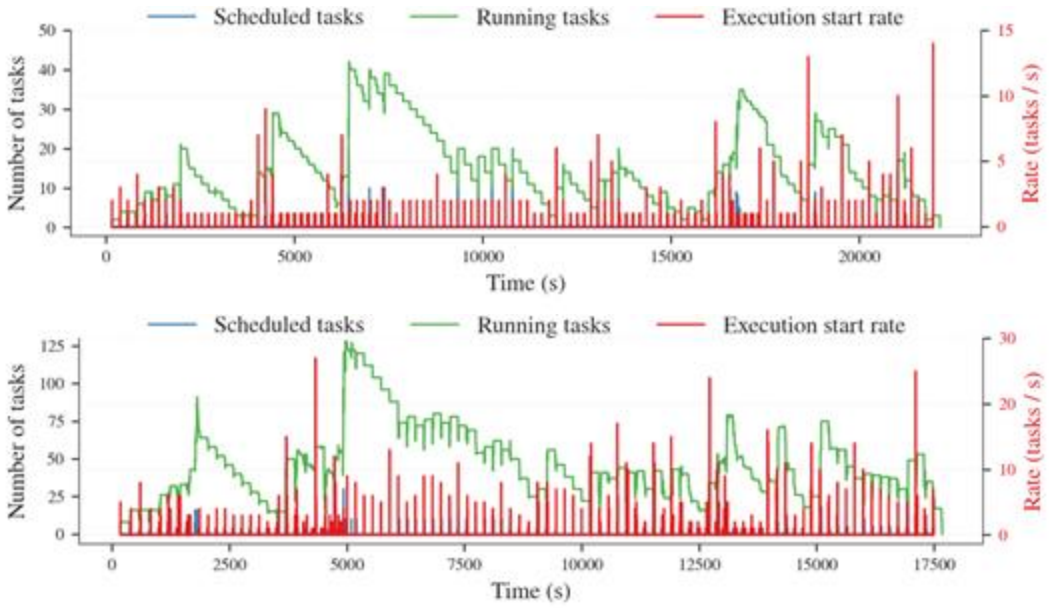
Left Y-axis presents the number of tasks being scheduled (blue) and running (green) concurrently. Right Y-axis presents the launching rate for the tasks execution using SRUN and FLUX as execution backends.

Corresponding plots show lower concurrency and launching rate for SRUN in comparison to FLUX.

SRUN (top - 256 nodes, bottom - 1024 nodes)



FLUX (top - 256 nodes, bottom - 1024 nodes)



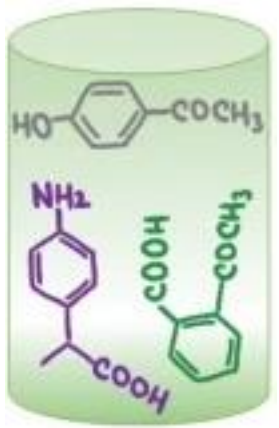
IMPECCABLE

Implement "plug and play" for more modules and functions into IMPECCABLE

REINVENT produces small molecules with optimal properties

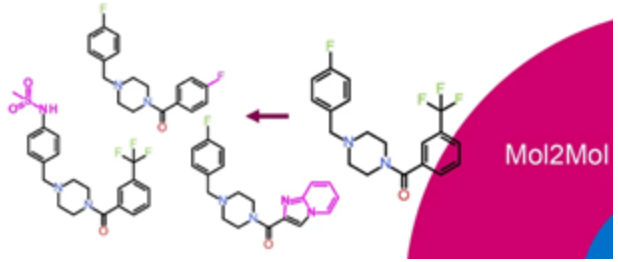
- Binding affinity
- Absorption
- Distribution
- Metabolism
- Excretion
- Toxicity
- Druglikeness
- …

Generative AI

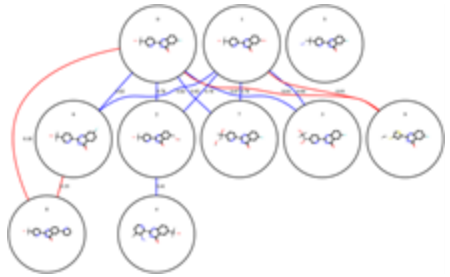


Lead optimisation

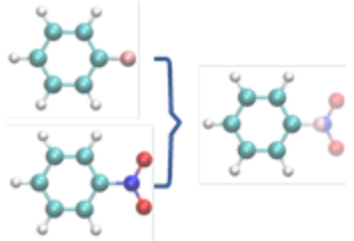
Mol2Mol



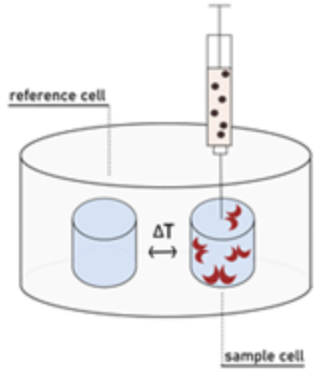
LOMAP

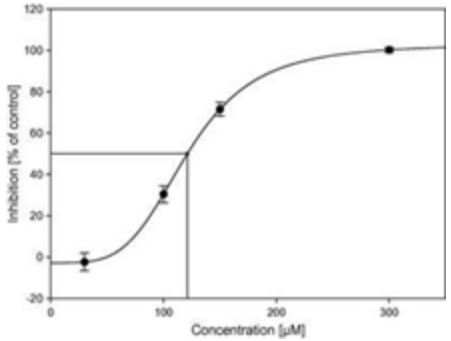


TIES20



Validation





IMPECCABLE: complex and heterogeneous

No turnkey solutions! Why is this challenging?

- **Heterogeneous:** High-throughput function calls, ensembles of MPI tasks, coupled AI-HPC
  - Producers of data (PB) and consumers (ML) *"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 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

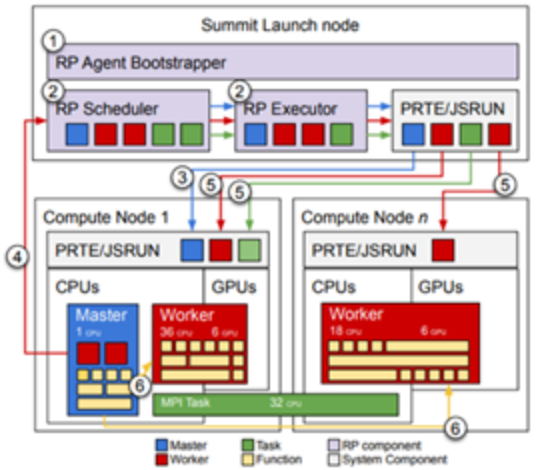
Table 3: Throughput and performance measured as peak flop per second (mixed precision, measured over short but time interval) per Summit node (6 NVIDIA V100 GPU).

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

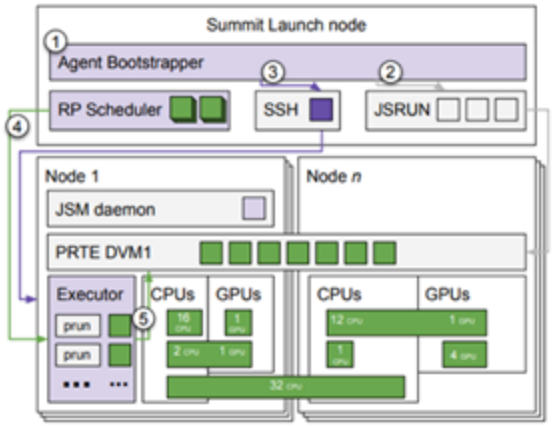
1000x variation in workflow throughput

IMPECCABLE workflow - performance

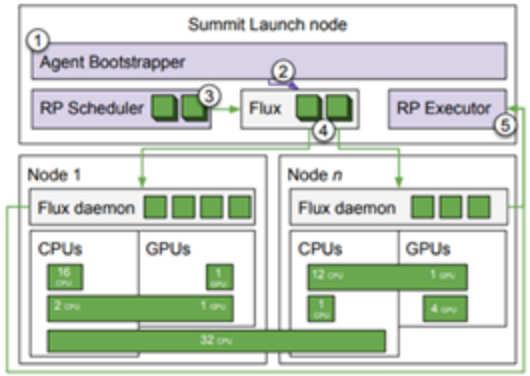
- Pilot-based task execution frameworks implemented using RADIAL-Pilot allow for the execution of complex workflows large heterogeneous HPC.
- The infrastructure has supported a campaign utilizing 2.5×10^6 node-hours on diverse HPC platforms for:
- docking $\sim10^{11}$ ligands with a peak docking rate of $\sim150\times10^6$ docks/hr,
- 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
- Recently implemented using **~8000 nodes on Frontier**



(a)



(b)

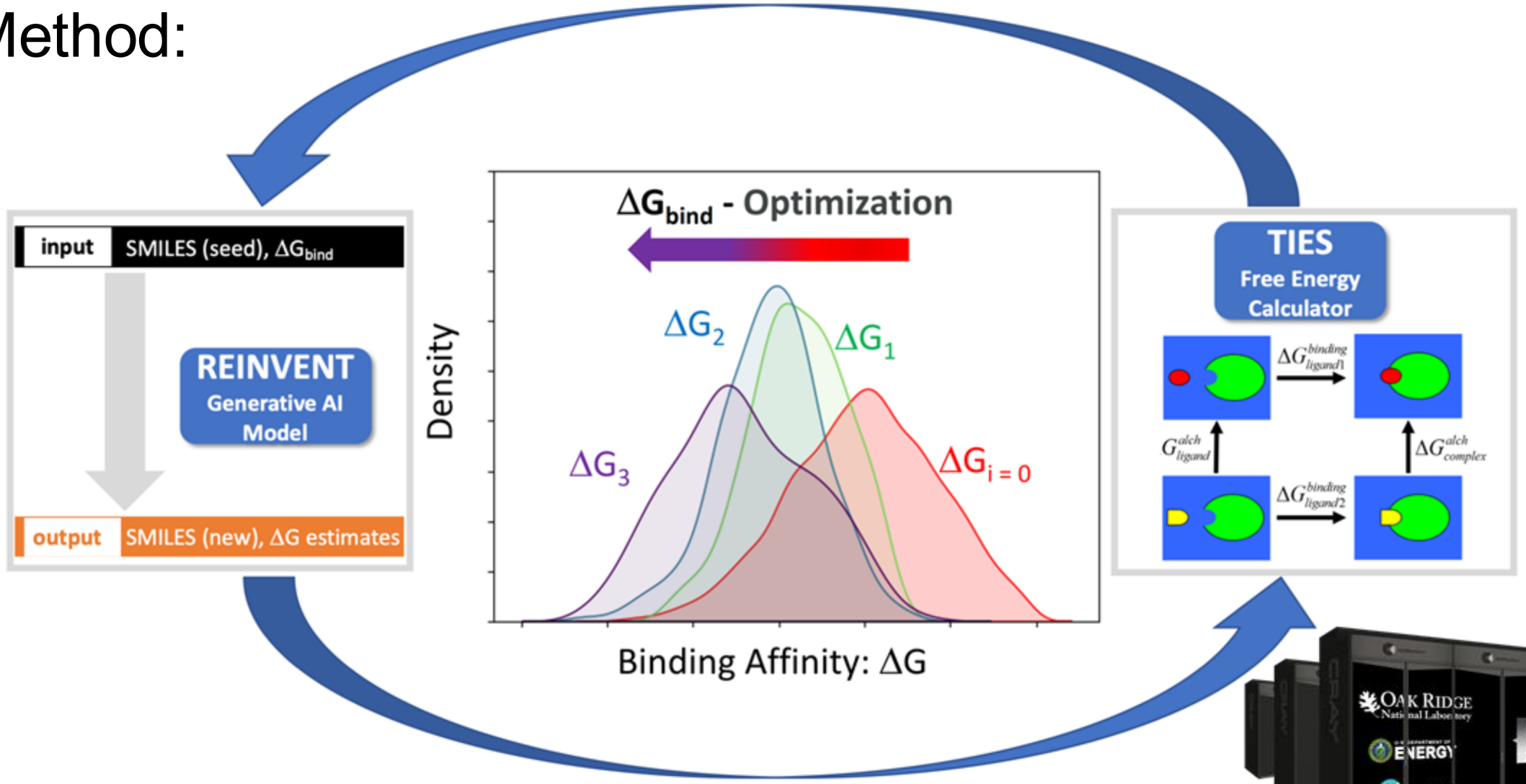


(c)

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. 1-12).

ML and free energy calculation for lead optimisation

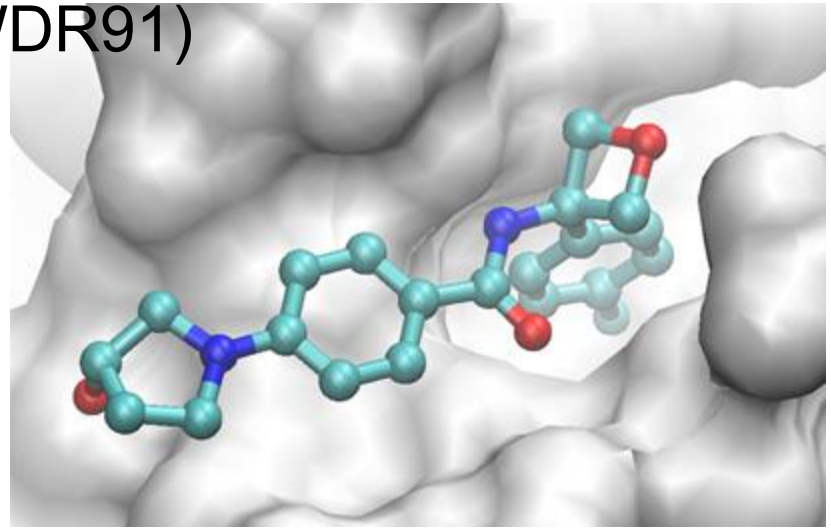
Method:



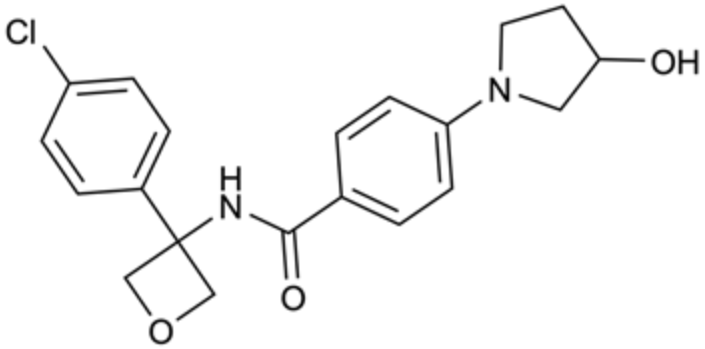
AI and free energy calculations for lead optimisation

Molecular System:

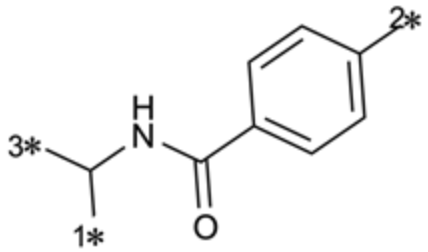
WD40 repeat-containing protein 91 (WDR91)



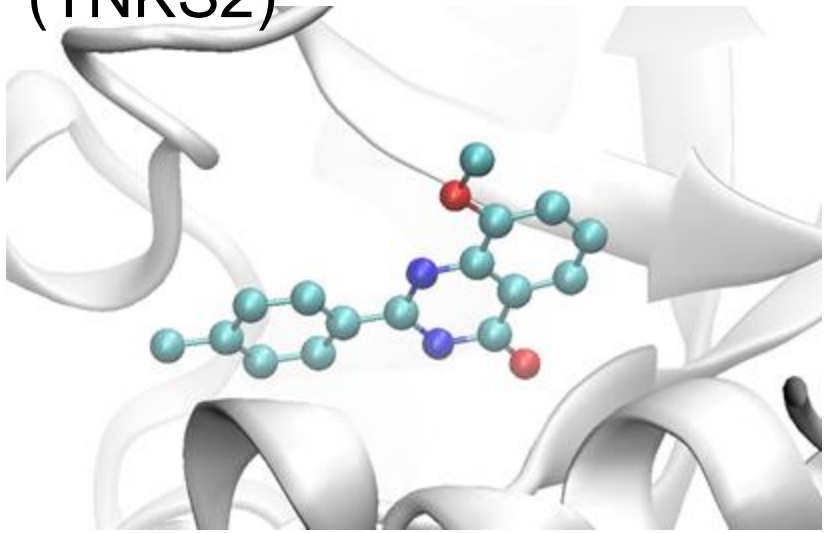
Hit compound, K_d = 6 ± 2 μM



Modifications



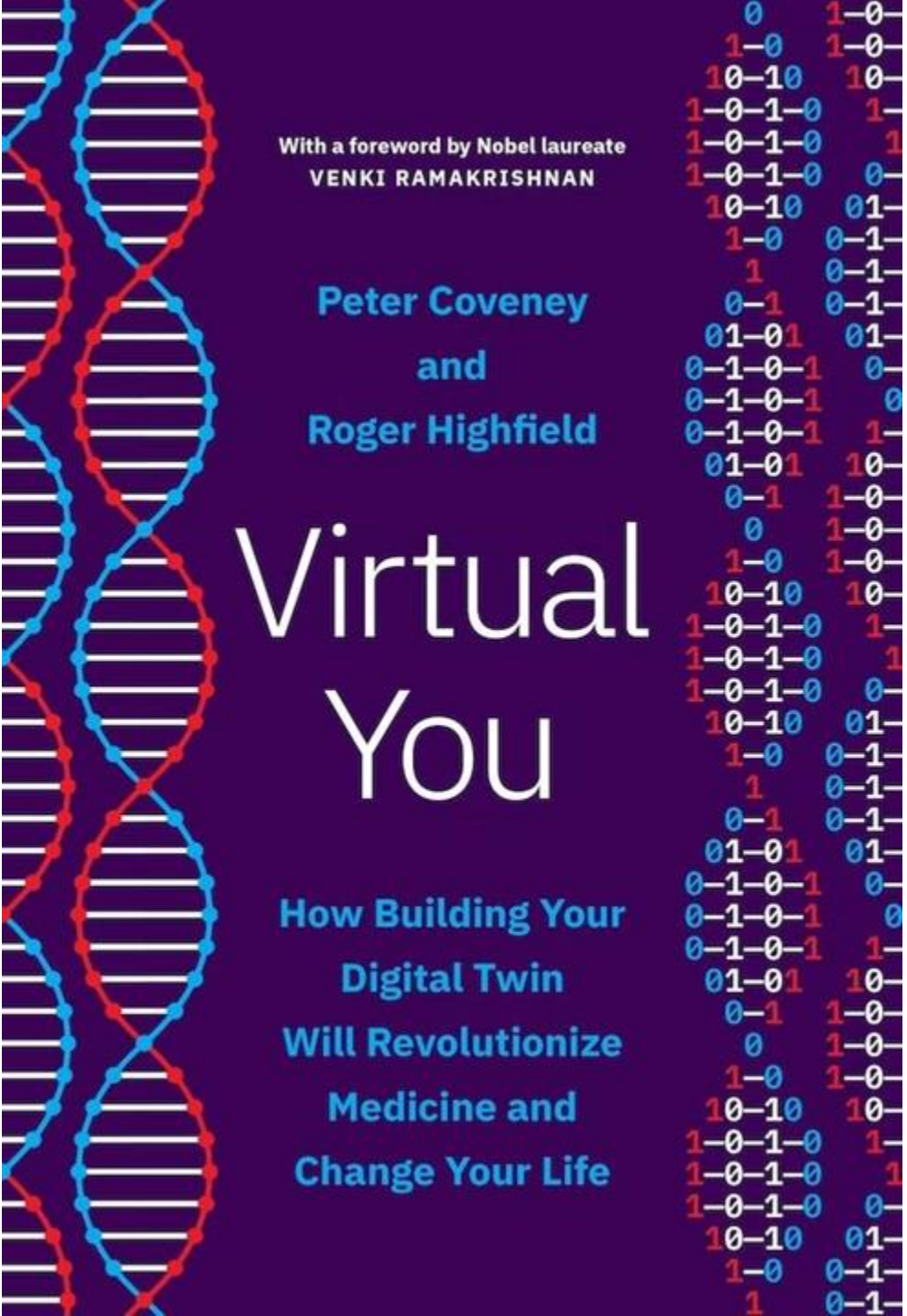
Human tankyrase 2 (TNKS2)



Hit compound, IC50 = 167 ± 60 nM

Conclusions

- Exascale enables all kinds of combinations of computation to take place serially and concurrently
- Combine with artificial intelligence methods
- Particularly suitable for the construction of biomedical digital twins
- Including uncertainty quantification (VVUQ)
- Requires advanced workflow management
- Produces actionable and explainable decisions



Use promotion code **AUFLY30** to save 30% at global.oup.com/academic