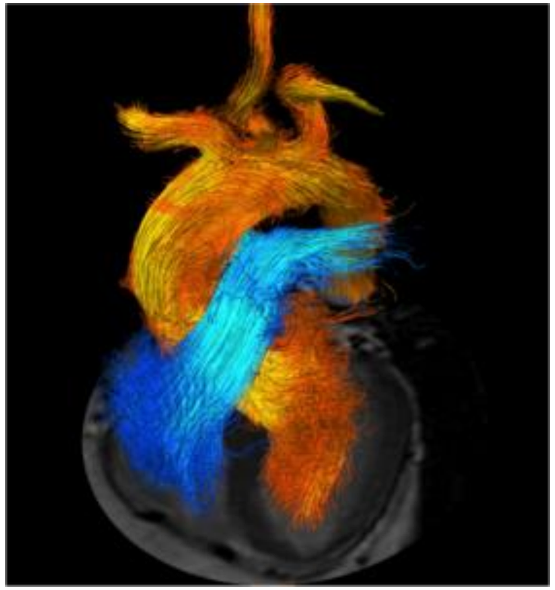
CompBioMed CoE - Addressing Biomedical Challenges with High Performance Computing



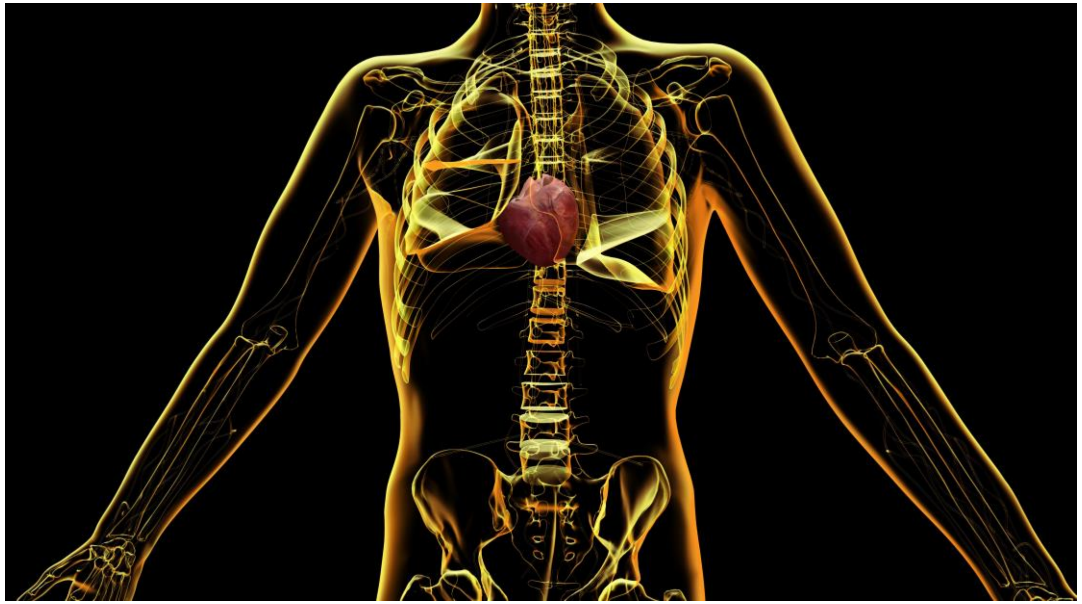
http://www.compbiomed.eu/

Jon McCullough
UCL

This project has received funding from the European Union's Horizon 2020 research and innovation programme.

CompBioMed

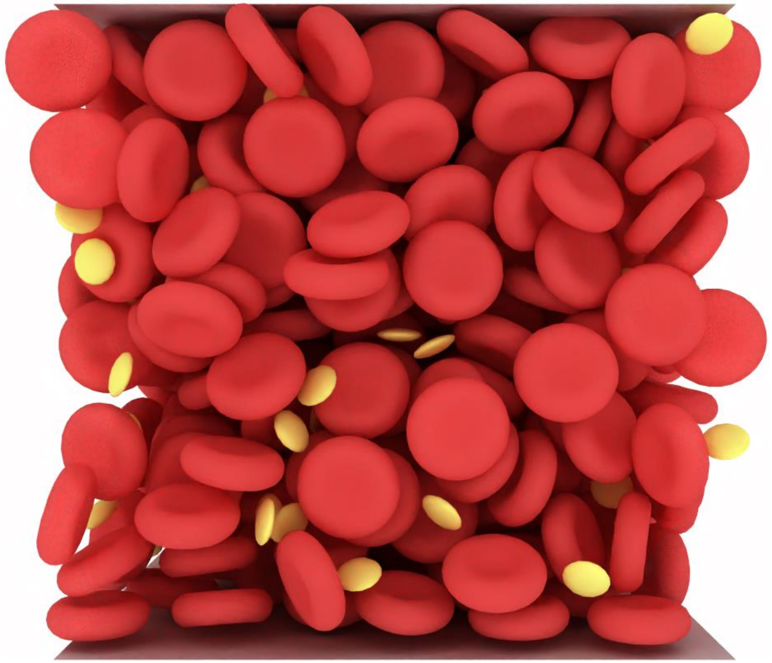
- European Commission H2020 funded Centre of Excellence in two stages: 2016-2019 and 2019-2023
- Consortium of academic institutions, supercomputing centres and industrial partners from across Europe and USA led by Prof. Peter Coveney (UCL)
- Supporting collaborations with associate partners across the globe

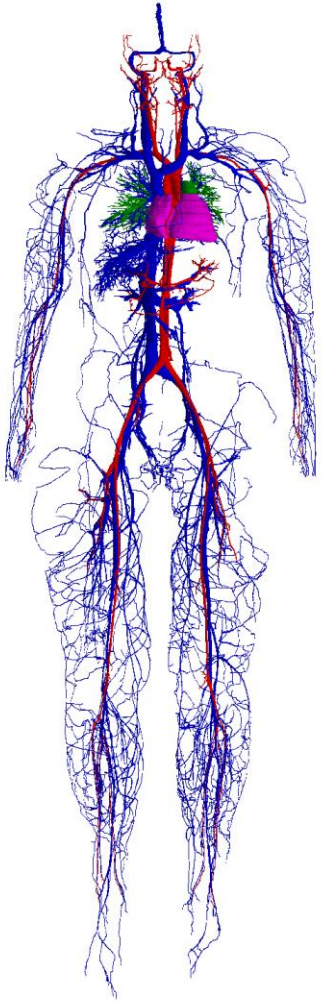


CompBioMed and HPC

- *Modus operandi* – Development and use of HPC for applications in computational biomedicine
- Four national level supercomputing centres are core partners – EPCC (UK), SurfSARA (NED), Barcelona Supercomputing Centre (ESP) and Leibniz Supercomputing Centre (GER)
- Research applications in the fields of cardiovascular simulations, molecular modelling, and neuro-musculoskeletal analysis
- HPC support and innovation – identifying and ensuring exascale readiness of codes in computational biomedicine

Cardiovascular Exemplars

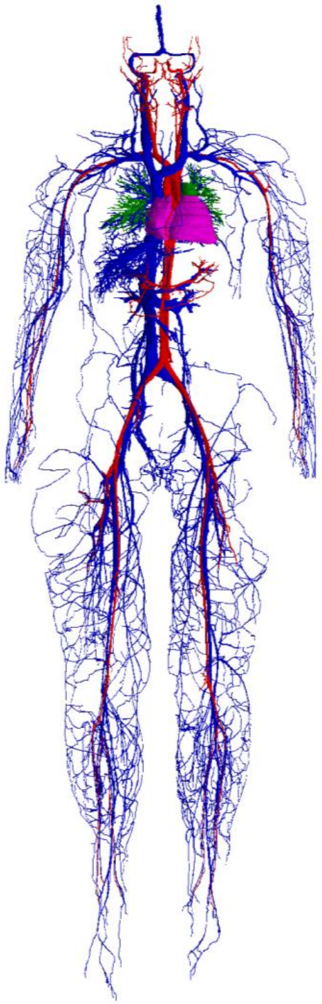




Cardiovascular – HemeLB: Full Human Blood Flow

HemeLB:

- 3D lattice Boltzmann code in C++ with MPI parallelism
- Optimised for sparse geometries seen in vascular networks
- Developing full human flow models of coupled arterial and venous flow
- GPU version in development



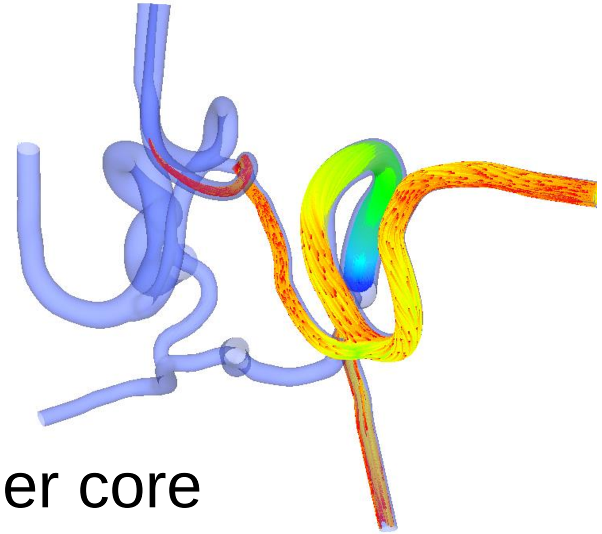
Current model:

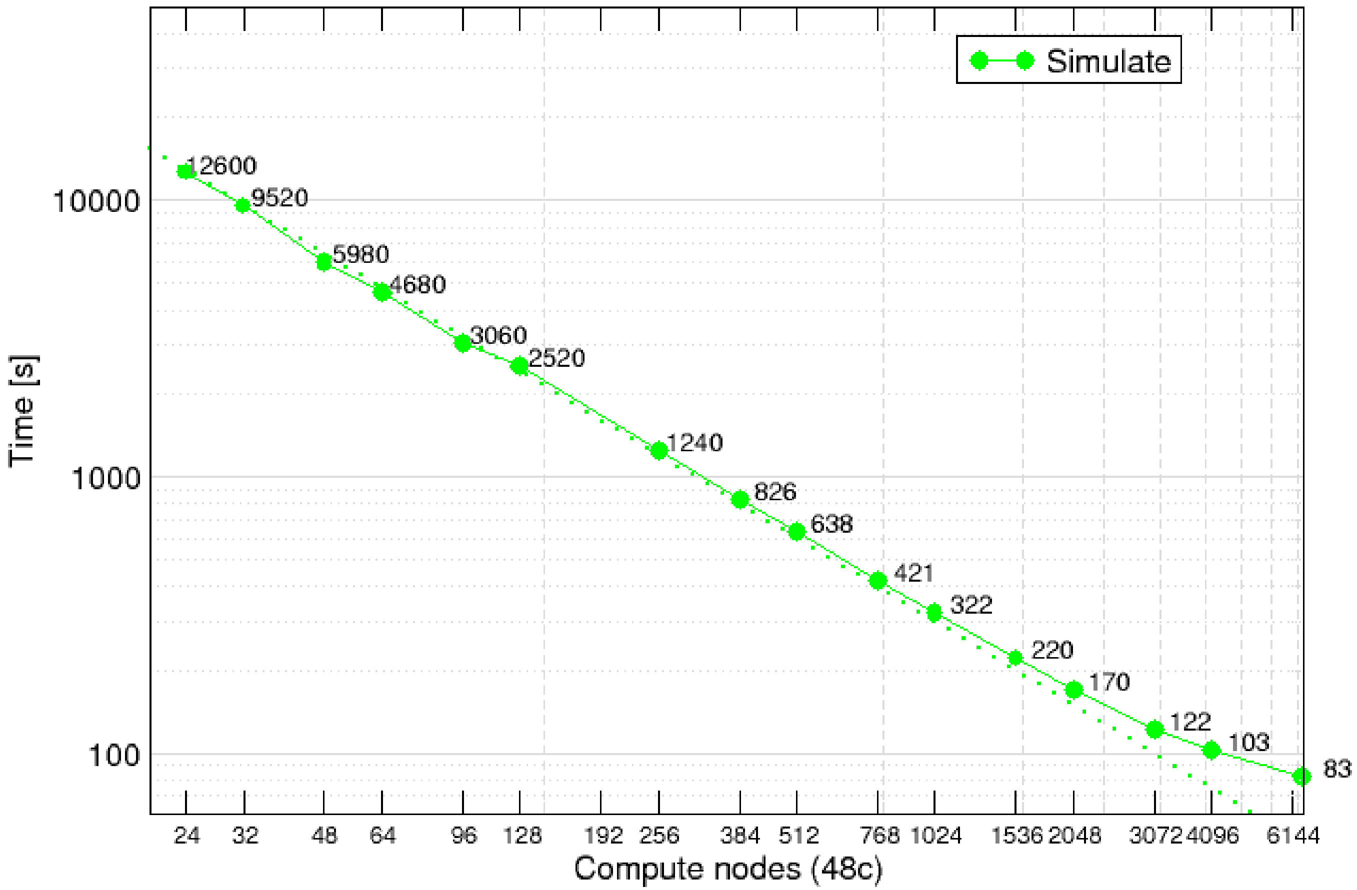
- 60µm resolution
- Systemic arteries ~5e8 sites
- Systemic veins ~1.5e9 sites
- Geometry generation is challenging

Cardiovascular – HemeLB: Full Human Blood Flow

Strong scaling to full machine level on SuperMUC-NG (309,696 cores) – thanks to Dr Brian Wylie (FZJ) via POP

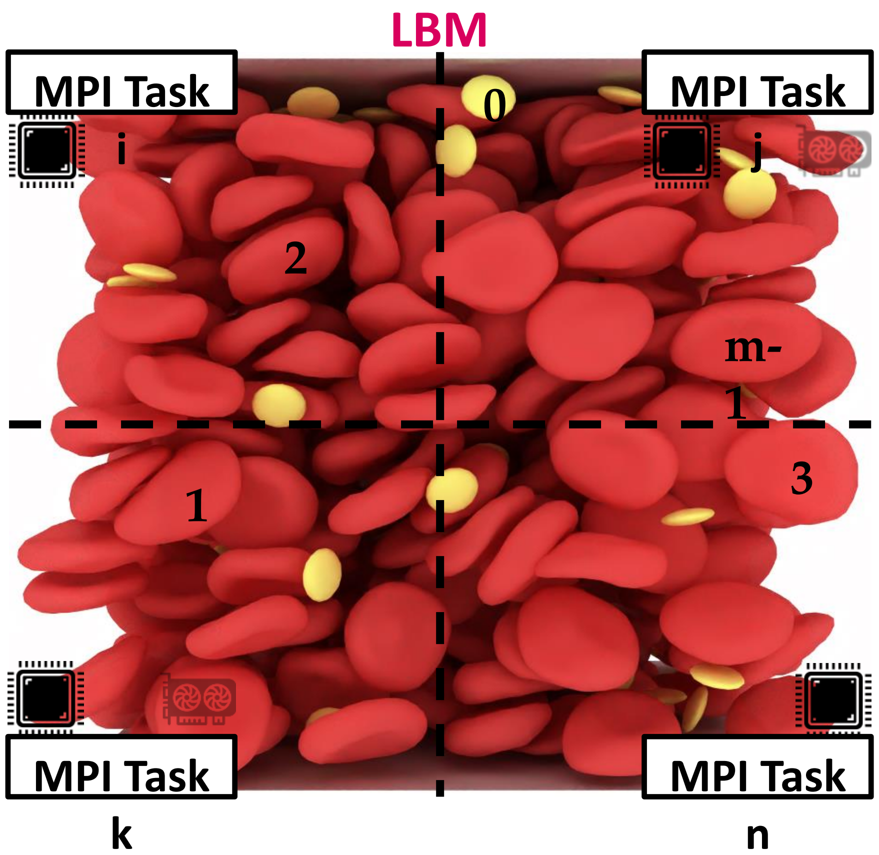
5000 steps of circle of Willis geometry – over 1e10 sites





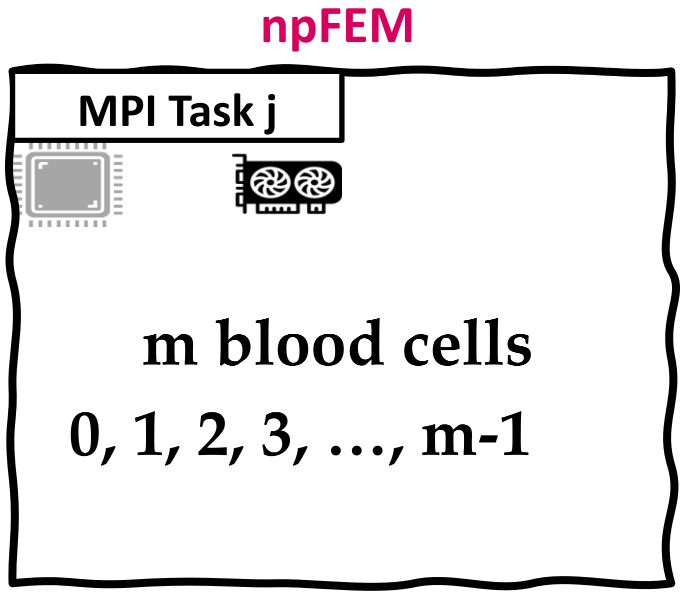
Expanding to higher core counts – geometry generation, scaling for coupled versions

Cardiovascular – Palabos: Resolved Blood Cells



Straightforward to partition a static homogeneous grid

CPUs deal with grid points (LBM) & Lagrangian points (IBM)



The blood cells are distributed once at the beginning to the available GPUs

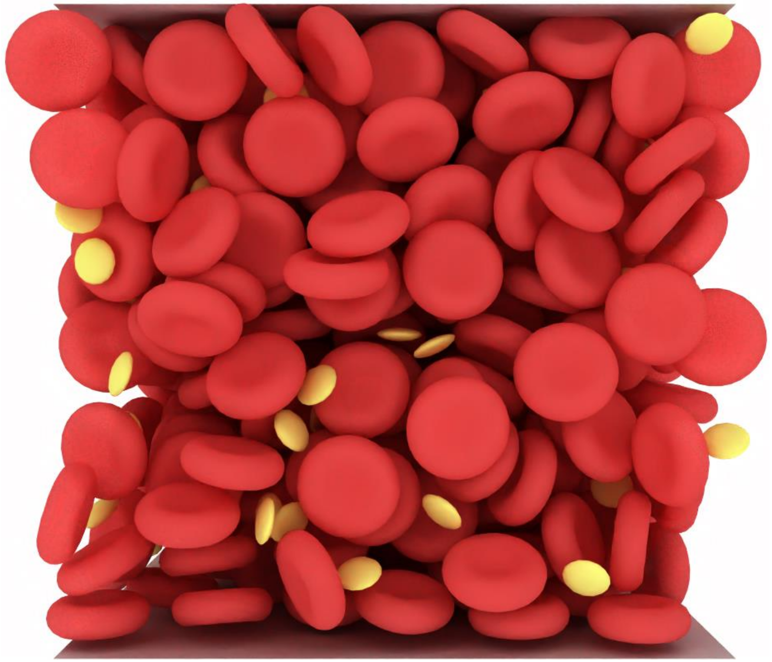
They can be spatially everywhere

MPI point-to-point communication

- The fluid solver sends forces & collision data to the solid solver
- The solid solver communicates the state at t+1

Cardiovascular – Palabos: Resolved Blood Cells

Reference case study
Ht 35%, Box 50x50x50 μm^3



RBCs 258 surface vertices
PLTs 66 surface vertices

box50x50x50 : **1**
on 5 GPUs
RBCs: 476
PLTs: 95

box50x100x50 : **2**
on 10 GPUs
RBCs: 953
PLTs: 190

box50x500x50 : **10**
on 50 GPUs
RBCs: 4765
PLTs: 953

box50x1000x50 : **20**
on 100 GPUs
RBCs: 9531
PLTs: 1906

box100x1000x100 : **80**
on 400 GPUs
RBCs: 38126
PLTs: 7625

Piz Daint @CSCS
No. 6 Worldwide
No. 1 Europe

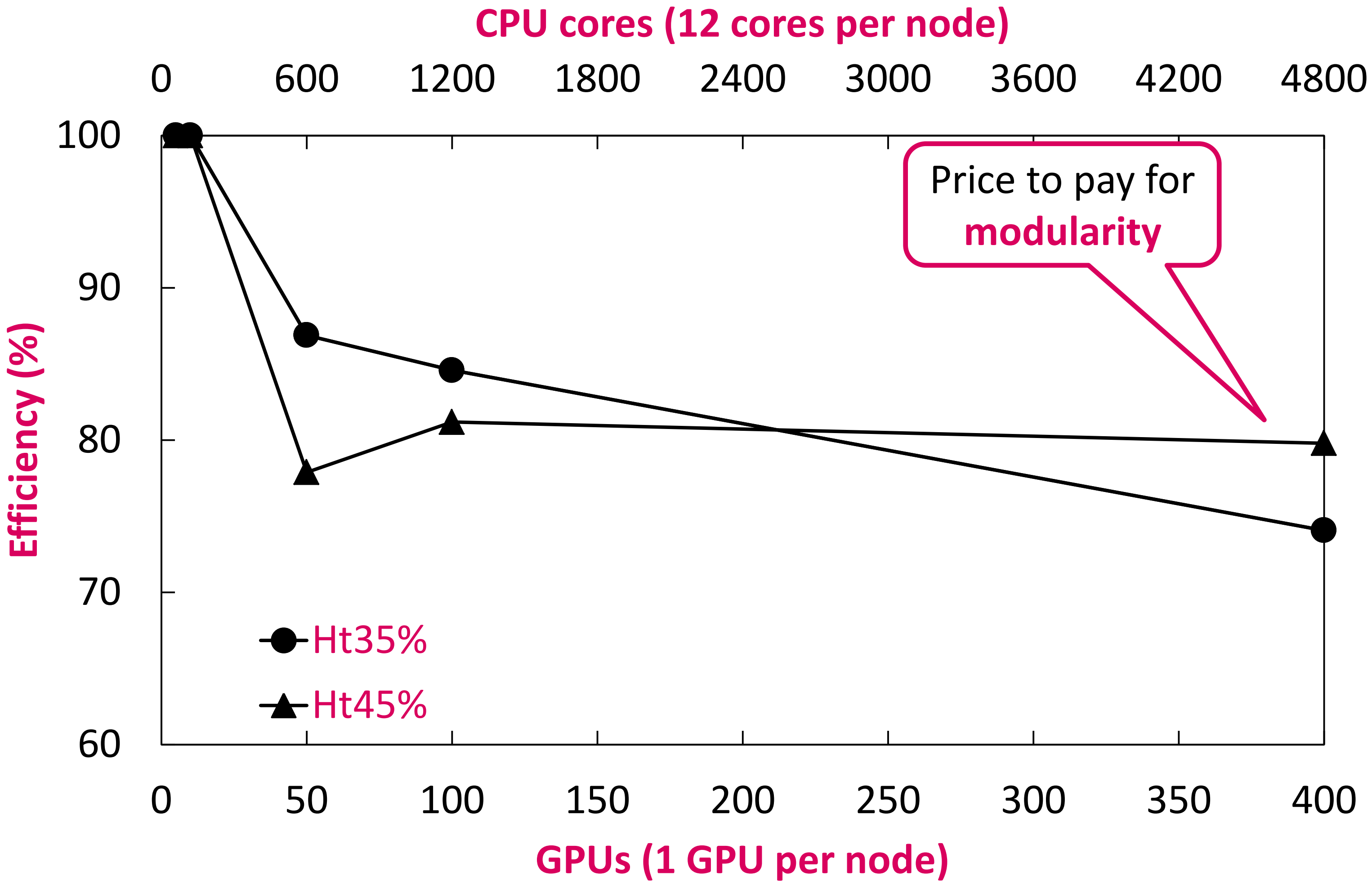
5704 Hybrid GPU+CPU nodes
(NVIDIA P100 + Intel Xeon)

1813 CPU nodes (Intel Xeon)

Cardiovascular – Palabos: Resolved Blood Cells

Weak Scaling –
Hybrid Version on Piz Daint

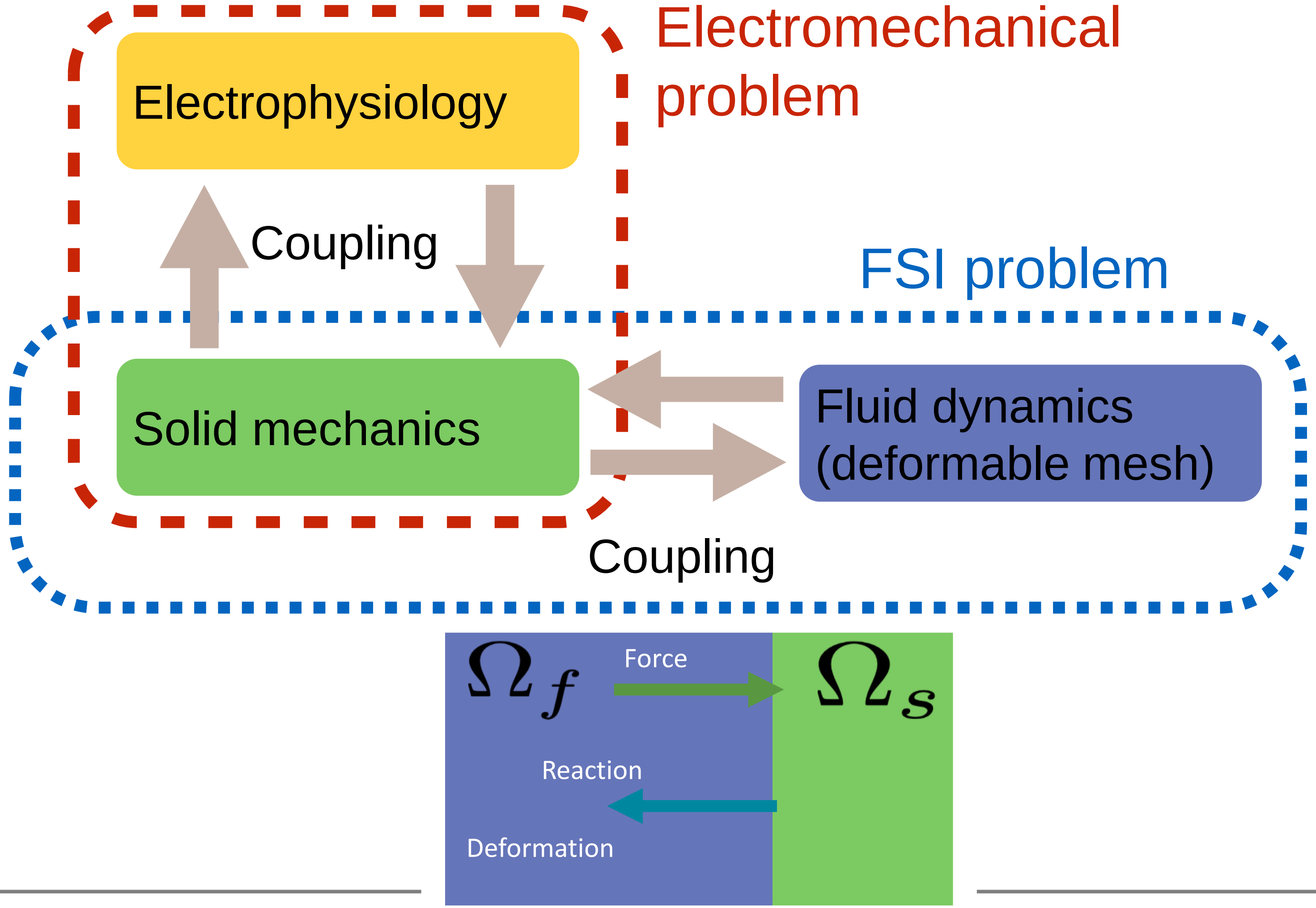
$$\text{Efficiency}_{weak} = \frac{t_{N_0}}{t_N} \times 100\ \%$$



Cardiovascular – Alya: Heart Multiphysics

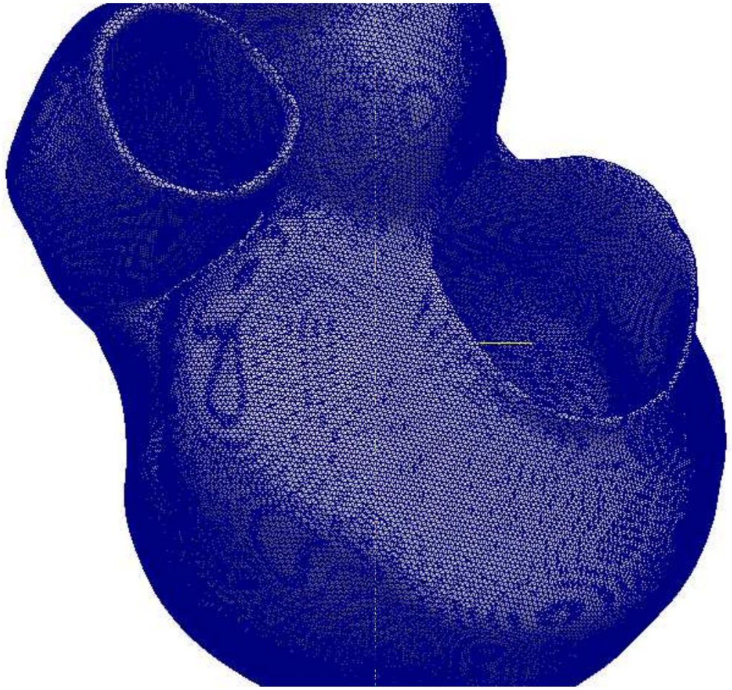
The heart as an electro-mechanic-flow (tri-physics) system:

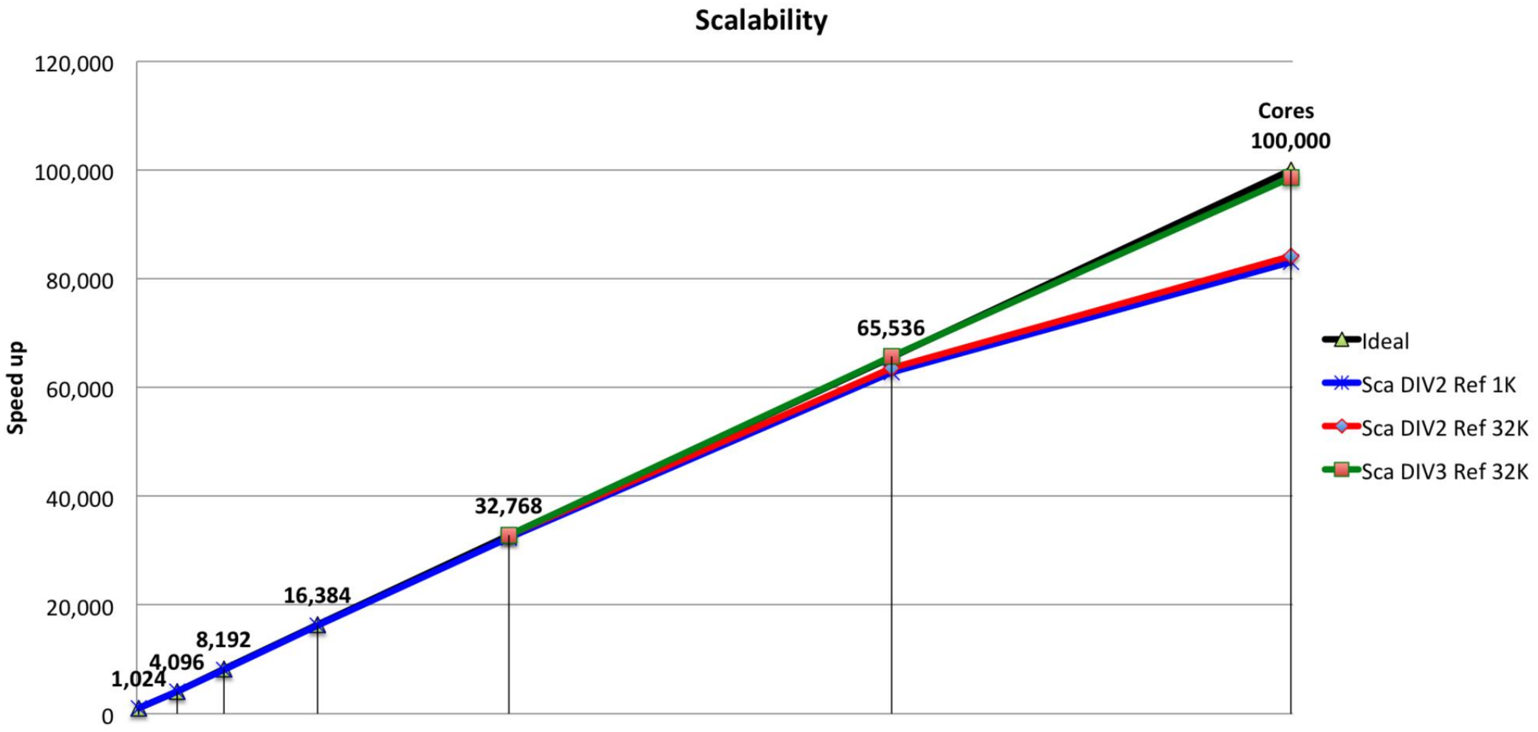
1. Electrophysiology
2. Computational Solid Mechanics
3. Computational Fluid Dynamics



Cardiovascular – Alya: Heart Multiphysics

Results: Electrophysiology simulation scalability





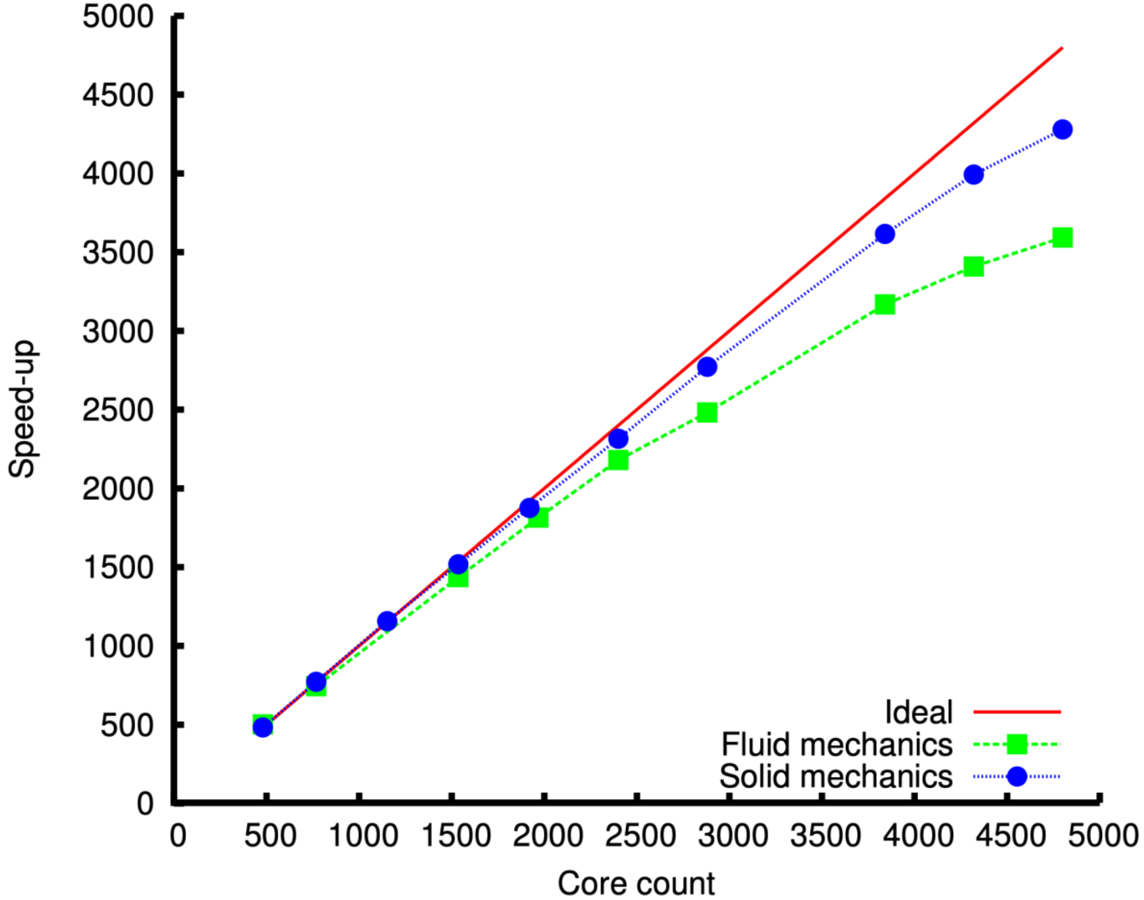
3.46 Billion elements

100,000 cores (Blue Waters, USA)

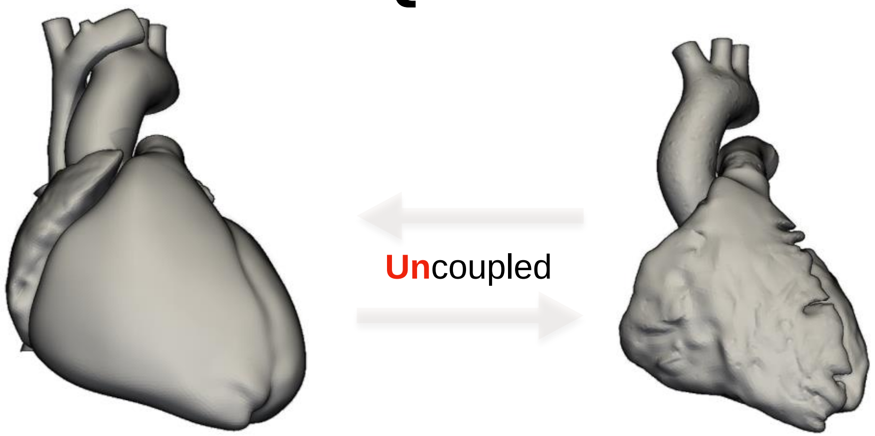
[1] Vázquez, M, et. al. (2016). Alya: Multiphysics engineering simulation toward exascale. *Journal of computational science*, *14*, 15-27.

Cardiovascular – Alya: Heart Multiphysics

Results: Uncoupled parallel performance



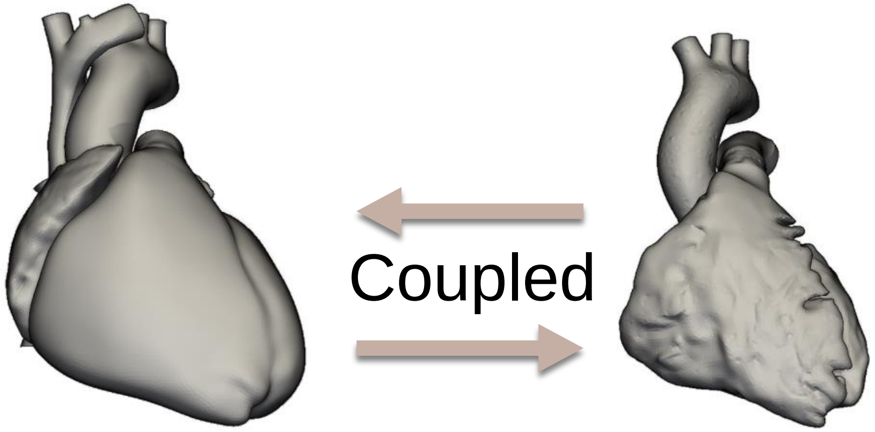
Mesh division { Solid: ~70M elem
Fluid: ~40M elem

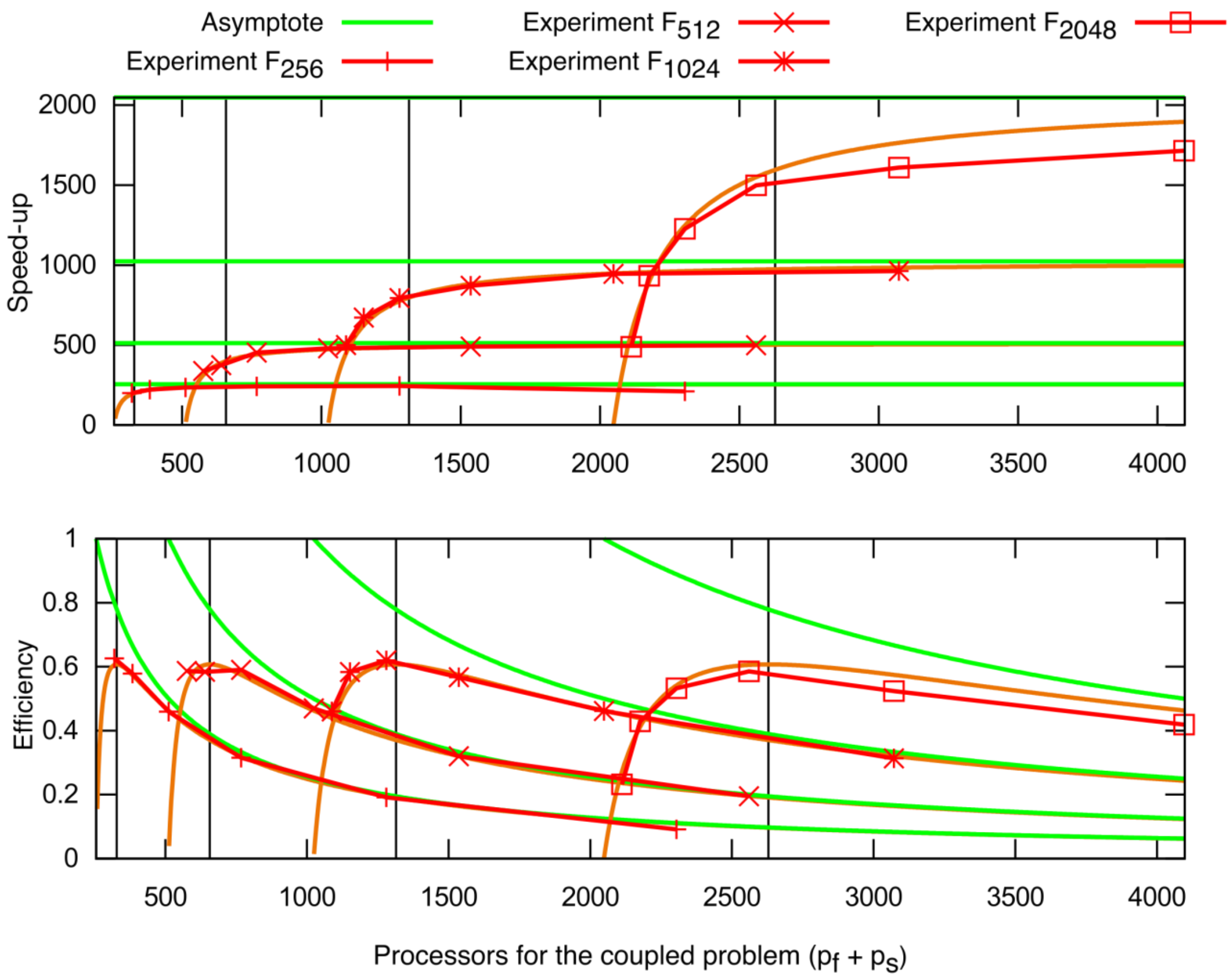


Core	CFD		CSM	
count	elem./core	eff.	elem./core	eff.
768	52k	1.00	91k	1.00
1536	26k	0.97	46k	0.99
2880	15k	0.89	25k	0.96
4800	8k	0.77	14k	0.92

Cardiovascular – Alya: Heart Multiphysics

Results: Coupled parallel performance



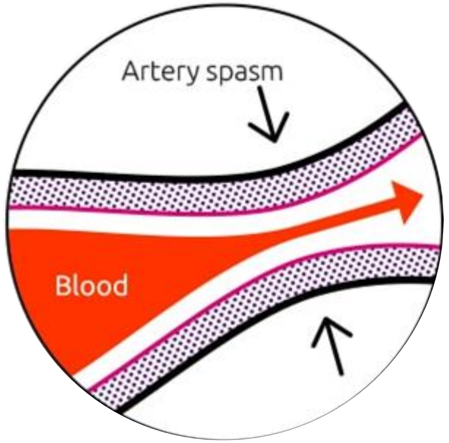


Gauss-Seidel scheme

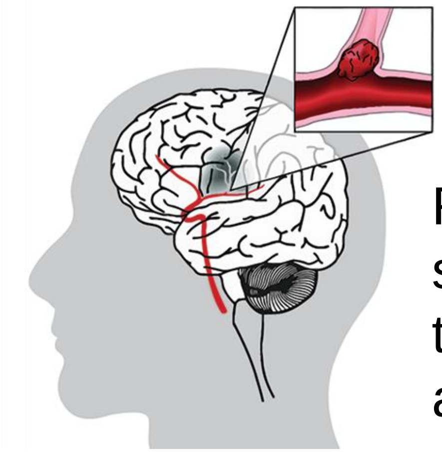
while *Time loop* **do**
  $d_\alpha = d_\alpha^{ini}$
  while *Coupling loop* **do**
    $f_\alpha = \mathrm{NSI}(d_\alpha)$
    $d_\alpha = \mathrm{SLD}(f_\alpha)$
    $d_\alpha^{I+1} = \varphi_{GS}(d_\alpha)$
  end
end

Cardiovascular – 1D models and emulators

Cerebral vasospasm vs Ischaemic stroke



Progressive narrowing of lumen of brain arteries



Reduction of blood supply due to a clots that obstructs the artery

The two pathologies give similar reading on the Transcranial Doppler, but the treatment strategies are considerably different, making an early and clear differentiation between the two fundamental.

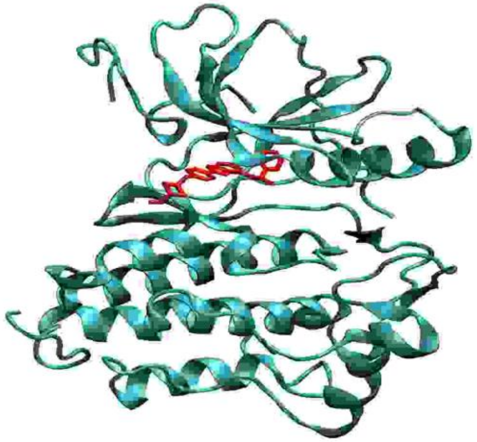
This is done by means of cardiovascular modelling and sensitivity analysis.

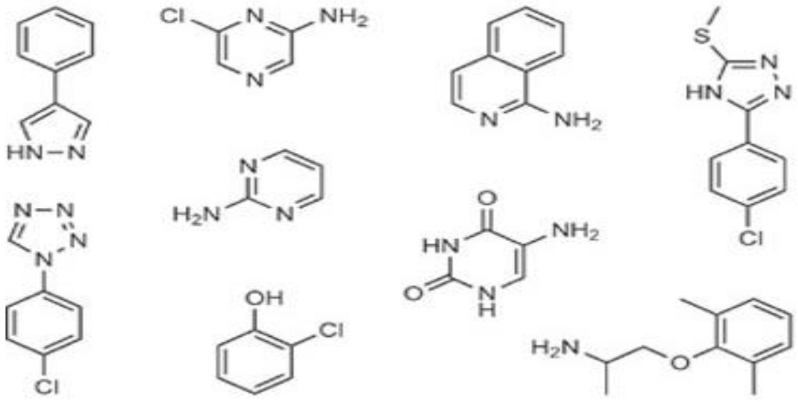
Cardiovascular – 1D models and emulator

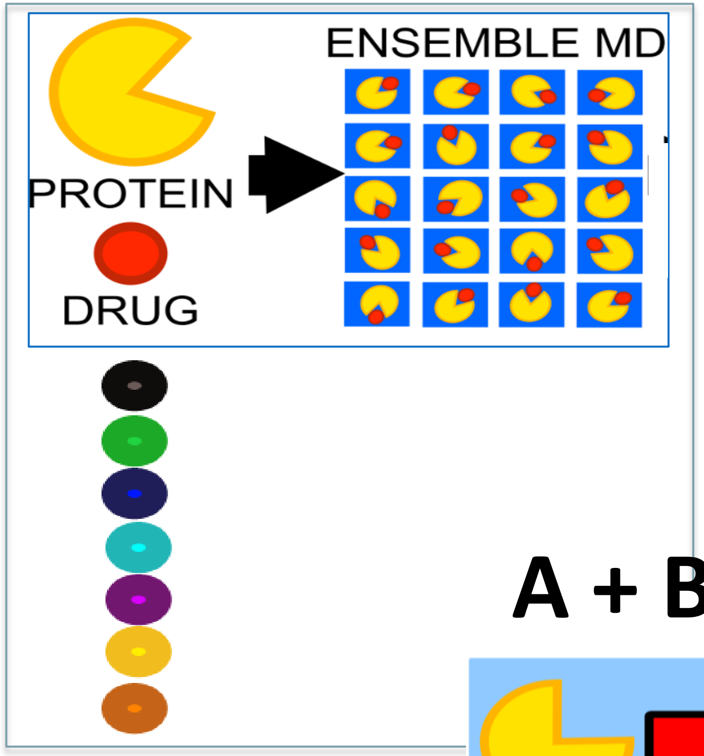
Use of Sobol sensitivity analysis and Gaussian process emulators to identify biomarkers for early differentiation between cerebral vasospasm and ischaemic stroke.

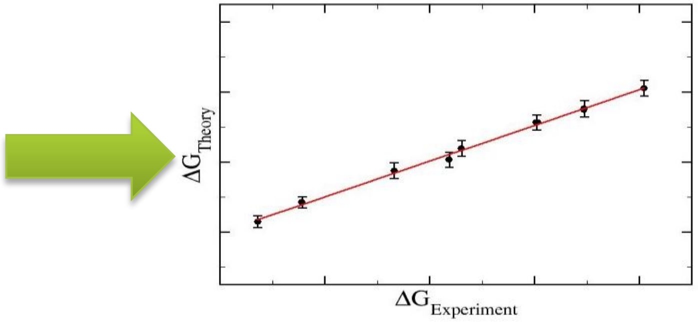
- 1D fluid structure simulations of the brain circulation
- 150 input parameters
- Train and verify emulator: 500 and 5000 simulations, 3 minutes per simulations → Parallel jobs on ShARC, Tier3 HPC system @ USFD
- Sobol sensitivity analysis: 150,000 emulated simulations, few minutes to run

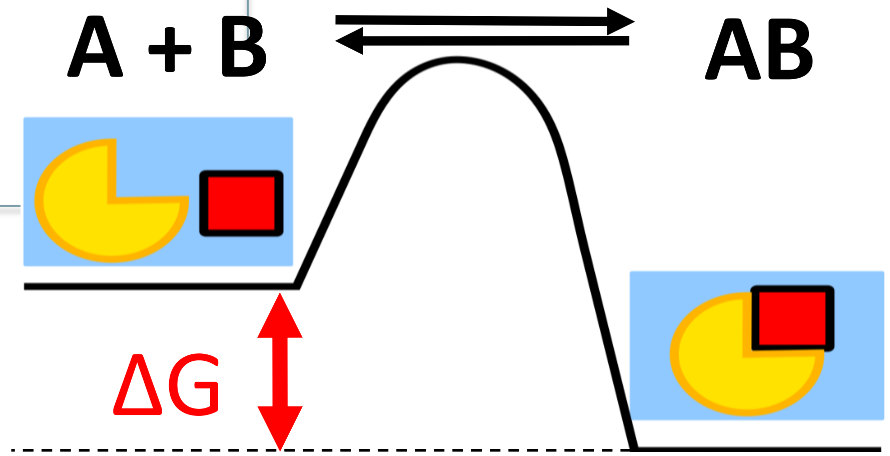
CompBioMed – Molecular Modelling











CompBioMed – Molecular Modelling

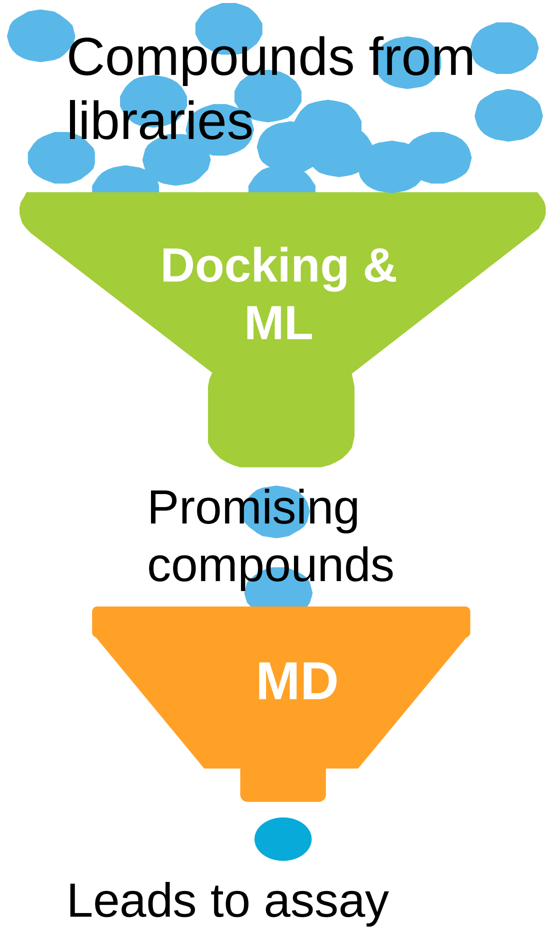
Use **molecular dynamics (MD) simulation** and **machine learning (ML)** to guide drug development.

Large Dataset to Screen

— Existing chemical libraries:
- ZINC: 230 million purchasable compounds
- DrugBank: 13,551 drug entries
- PubChem: 103 million compounds
- Enamine: 188,734 building blocks

— Machine learning generated virtual compound libraries:
vastness of chemical space
all possible compounds that might be synthesised

Quick Turnaround is required, especially in an urgency to find effective drug treatment like Covid-19



CompBioMed – Molecular Modelling

Docking performance on Frontera

- pilot startup (~1 min) + master startup (~3 min) + worker startup (~2 min)

- single node: 2 docks / min / core
  110 docks / min / node ~ 95% efficiency

- scaling:

16 cores:	~ 90% efficiency
32 cores	~ 85% efficiency
64 cores	~ 80% efficiency
128 cores	~ 70% efficiency
256 cores	~ 50% efficiency

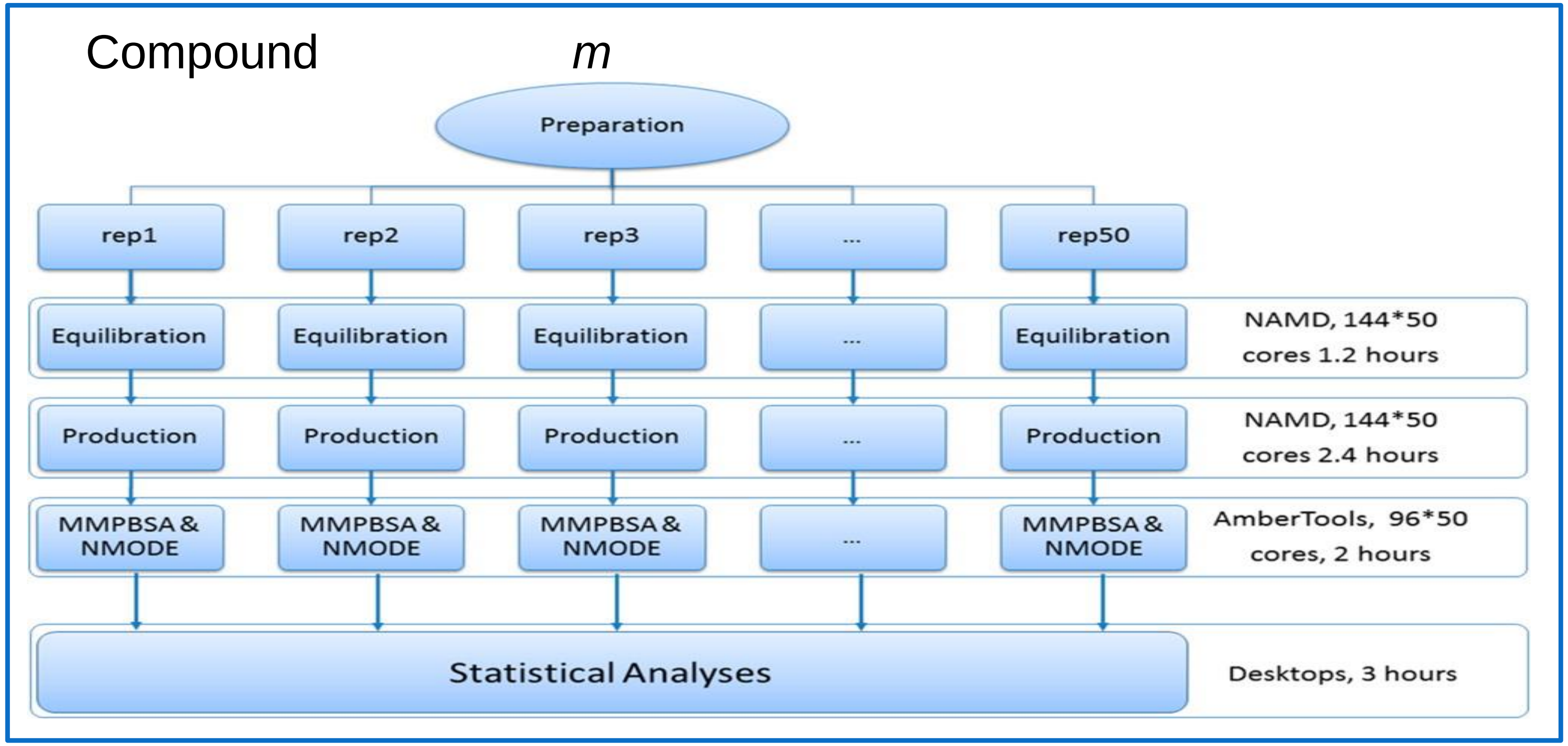
  numbers are rough values (vary with different configuration)

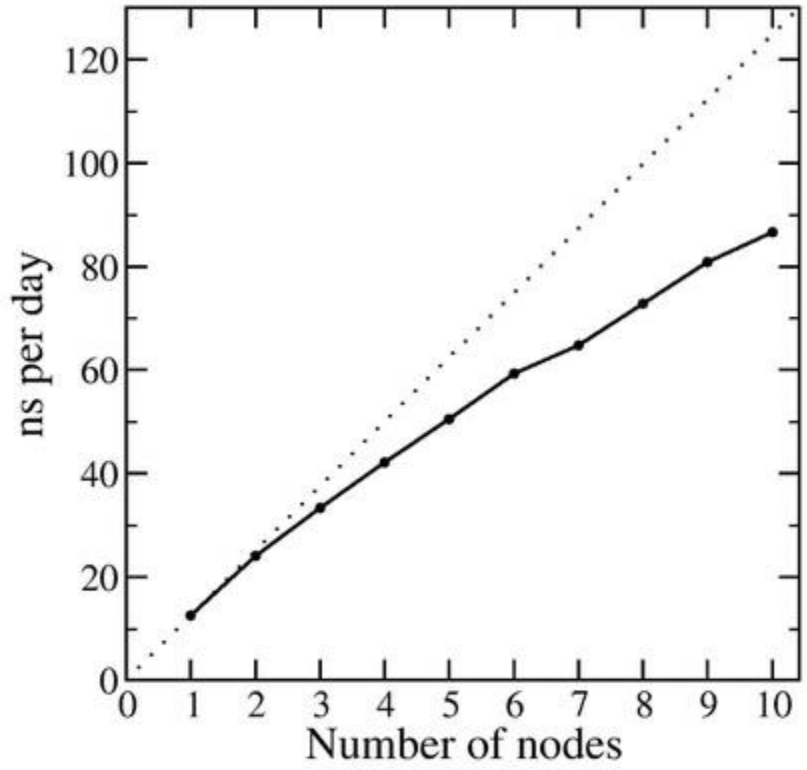
- scaling bottleneck is not yet identified

CompBioMed – Molecular Modelling

Small molecule *hits* from docking and machine learning are further evaluated to identify promising *lead* compounds.

- Investigations conducted for *m* compounds, each consisting of *n* replicas
- Simulations on Covid-19 targets show that 300 compounds can be studied within 20 hours using 3000 nodes on SuperMUC-NG

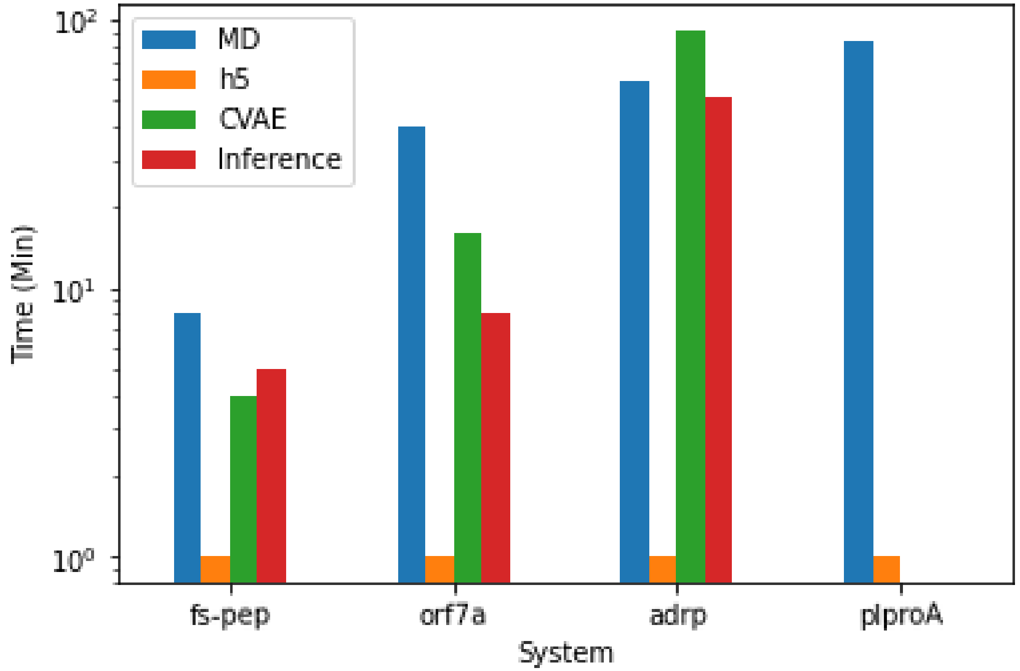




CompBioMed – Molecular Modelling

MD Performance Results

- fs-peptide (66 backbone atoms, 23 residues) : 18 mins
- orf7a (198 backbone atoms, 66 residues): 65 mins
- adrp (495 backbone atoms, 168 residues): 204 mins
- plproA (942 backbone atoms, 314 residues): 84 mins*



Settings:

- Time value is only for 1 iteration
- 10 ns simulation time

* plproA does not include CVAE/Inference elapsed time as it fails loading samples onto GPU memory, horovod-enabled performance results will be added, instead.

CompBioMed – Molecular Modelling

MD Performance Results

Previous slide shows CVAE training becomes a bigger proportion of the workflow execution time for larger systems as the matrix size is increasing

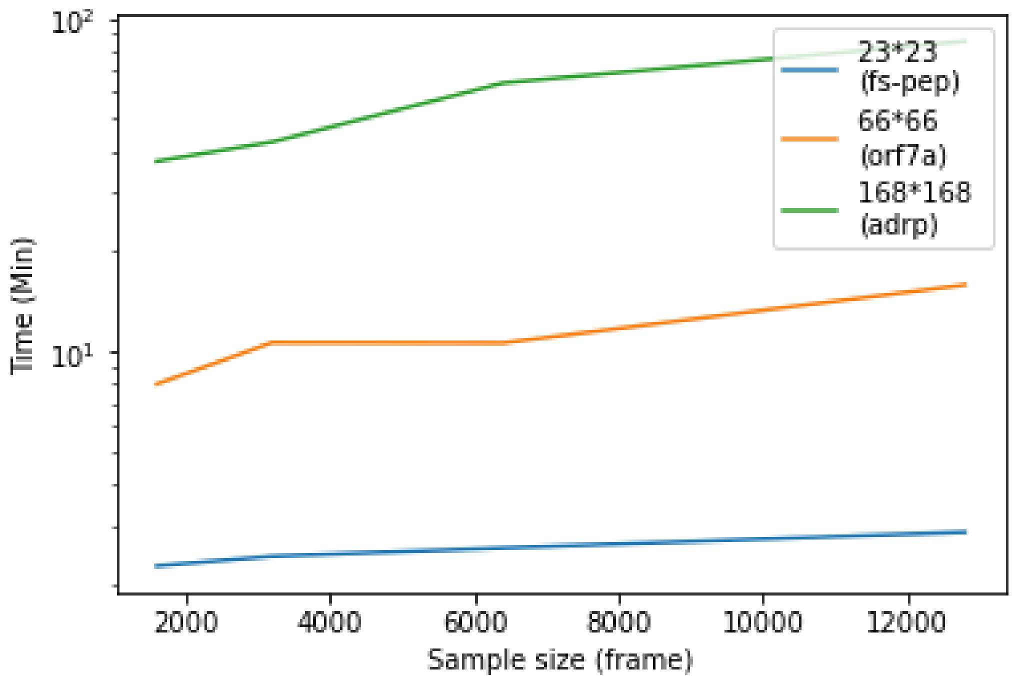
X axis: ensemble size of MD == sample size

Y axis: total time for epochs

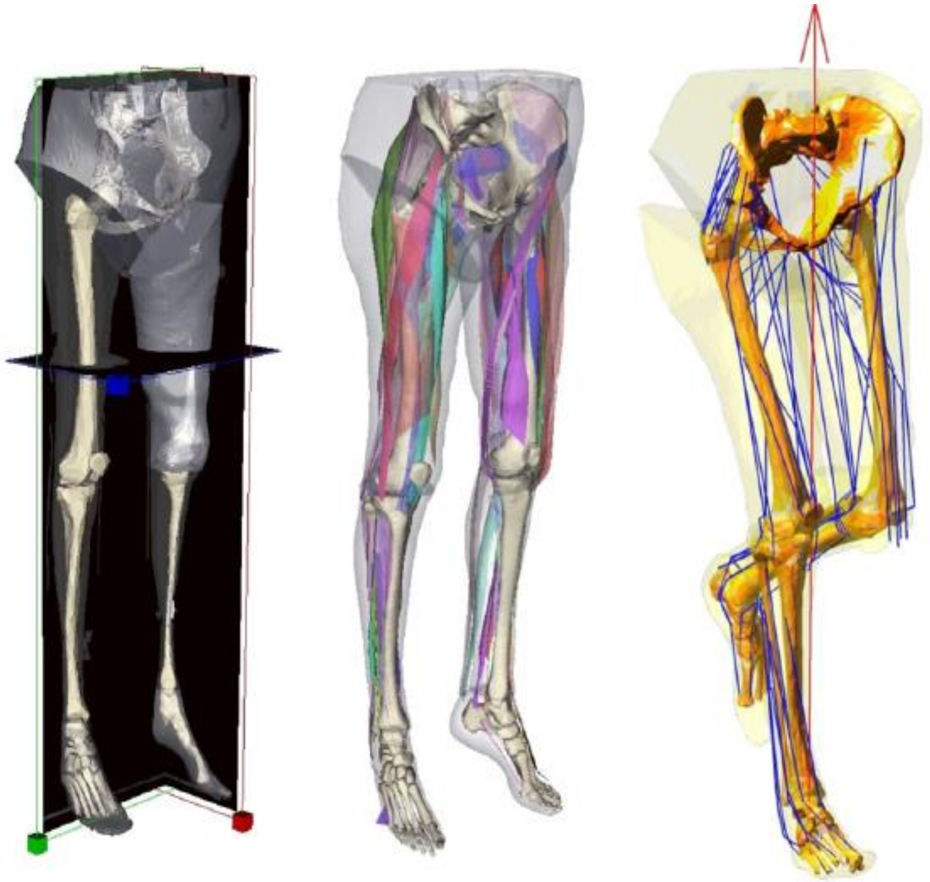
Legend: protein with matrix size multiplied by residue sizes e.g., 23 residues 23*23 matrix

Settings:

- 100 epochs (fixed)
- 10, 20, 40 and 80 MD ensembles == 2000, 4000, 8000 and 16000 samples (frames)



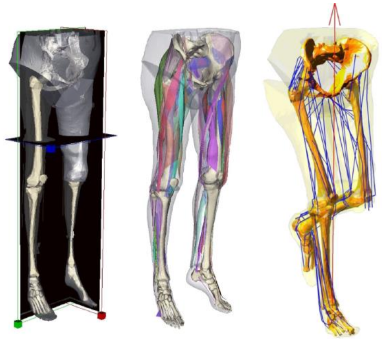
CompBioMed – Neuro-musculoskeletal Analysis



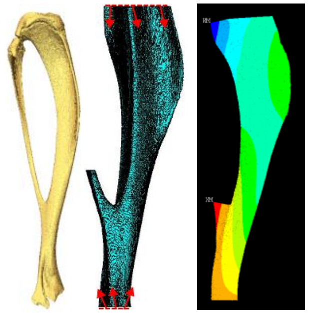
CompBioMed – Hip fracture analysis

Aim

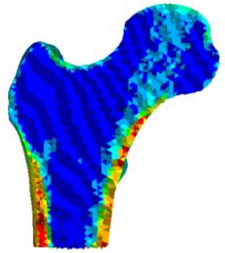
To develop an *in silico* trials platform to evaluate the efficacy of new bone drugs on human cohorts, from limited pharmacodynamics data on the new drug obtained on mice



A **validated** patient-specific model that predicts the risk of hip fracture today

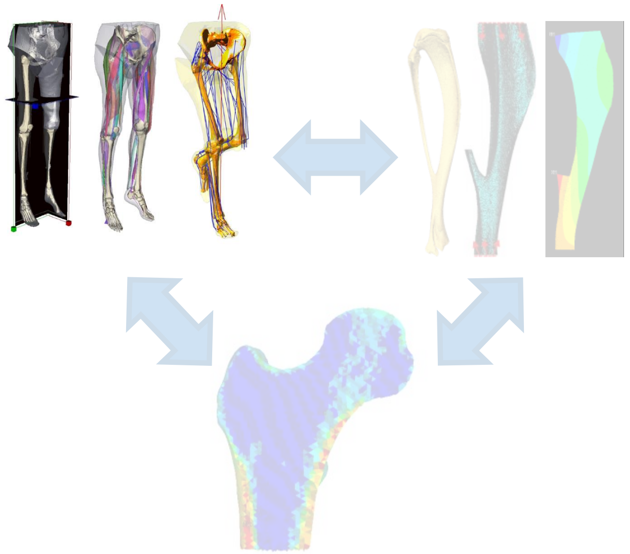


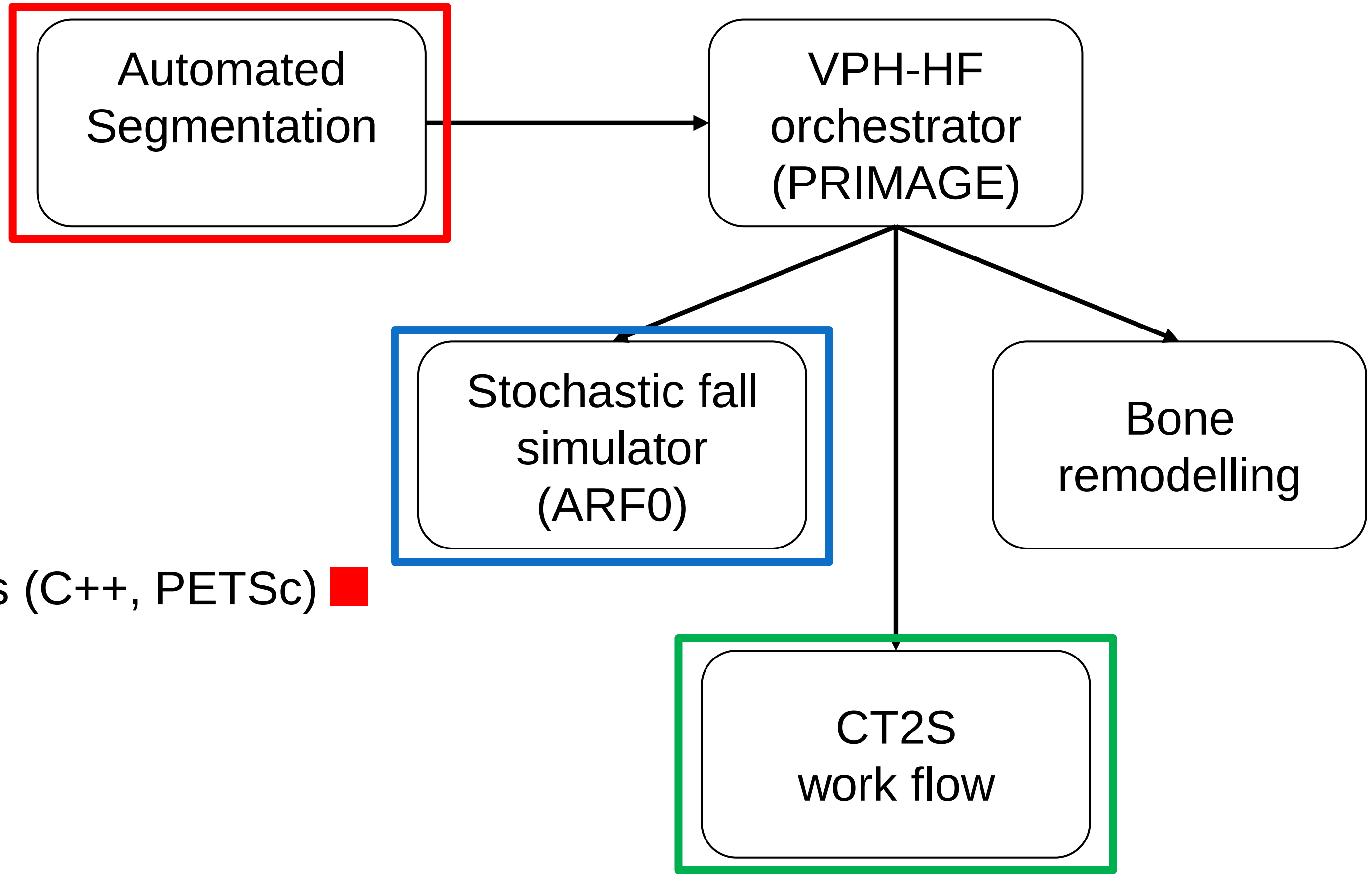
A **validated** mouse-specific model that can provide an estimate of the drug effect



A **validated** stochastic population-specific bone remodelling algorithm model that predicts changes in bone over time

CompBioMed – Hip fracture analysis





Elastic image registration on multiple cores (C++, PETSc)
Subject specific finite element model

SURFsara HPC system
ShARC, Tier3 HPC system @ USFD
UNIBO HPC Cluster

CompBioMed – Other applications

- Teaching
- Data analysis
- Exascale preparedness
- Application incubation

CompBioMed – Teaching and Education

- To create a new category of users of HPC (“future users”) who will be fluent both computationally and experimentally

Academic year	Medical students	Molecular Biosciences students	Total number of students	Core hours consumed	Core hours allocated	Fold difference (consumed/allocated)
2015-2016	0	20	20	0 (local)	0	-
2016-2017	0	29	29	0 (local)	0	-
2017-2018	40	85	125	17,452	10,500	1.66
2018-2019	20	99	119	49,394	10,900	4.53
2019-2020	20	83	103	97,919*	10,830	9.04*

CompBioMed – Teaching and Education

Teaching Outcomes

- 100% success rate for students using HPC as part of their degree (2017-2020)
- Improved diversity:

  >50% female

  >40% BME
- Employability from embedding computation in the Molecular Biosciences Curriculum

Research Outcomes

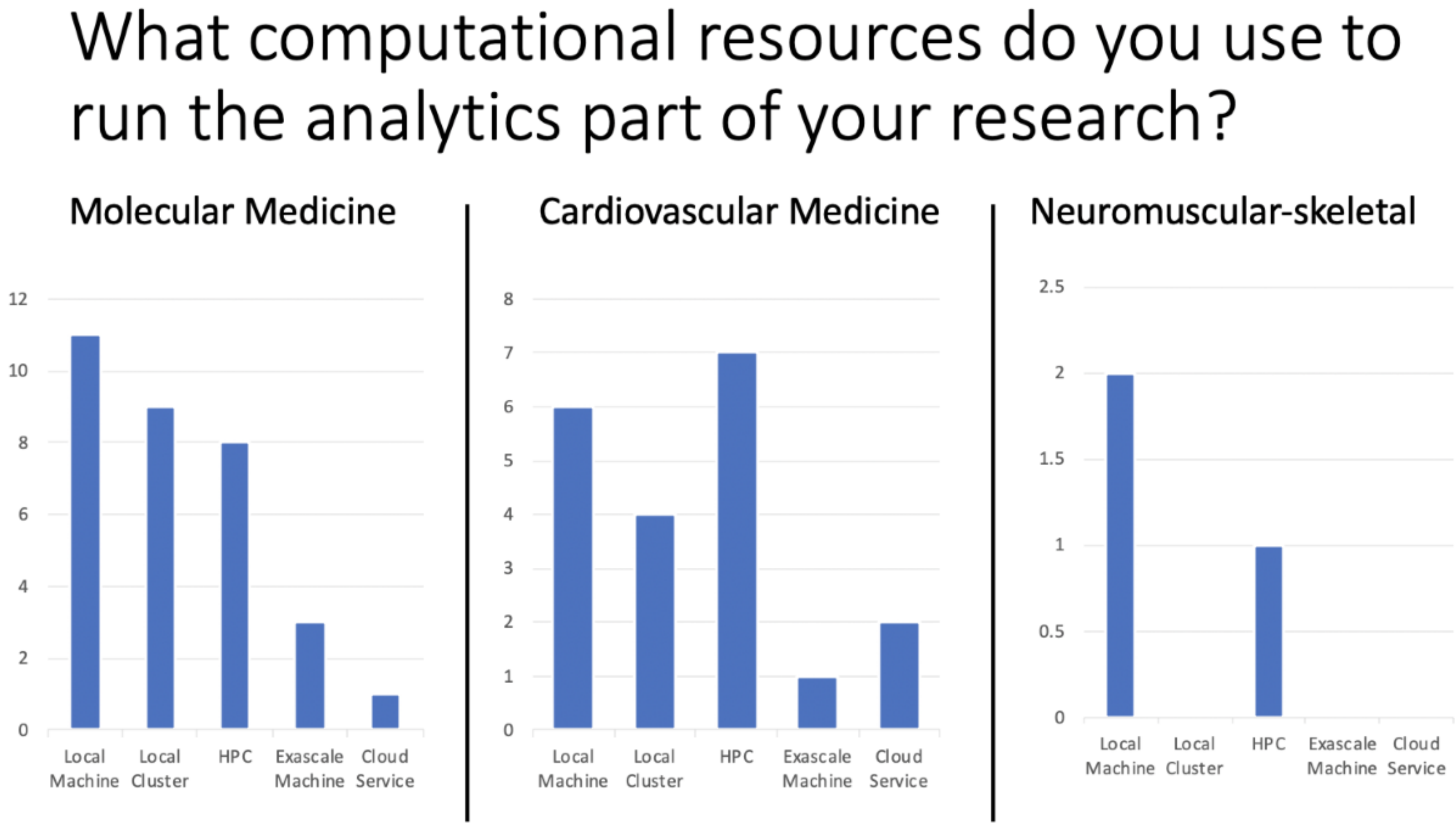
- Expansion of the programme:

  EU funding to deliver medical student HPC teaching at CompBioMed partner institutions
- Engagement with UKRI and EC

  To provide resource to support training HPC users, including students

CompBioMed - Data Analysis

- Evaluate current CompBioMed partners use of HPC/ML for data analytics
- Potentially identify options for streamlining of workflows and integration of VVUQ



CompBioMed - Collaboration

- Many existing applications are continually developing – potential for POP engagement to assist in making these efficient on large HPC infrastructures
- CompBioMed seeks to identify biomedical applications from academia and industry that could benefit from being made HPC-ready

- Webinars on many topics of HPC in computational biomedicine available on CompBioMed website

CompBioMed

- European Commission H2020 funded Centre of Excellence – 2 phases between 2016-23
- Focus on developing applications and exploiting HPC resources for computational biomedicine
- Research applications under main banners of cardiovascular modelling, molecular modelling and neuro-musculoskeletal analysis
  - Several of these are operating on large-scale HPC systems
- Enhancing these and newly identified applications an obvious potential collaboration between POP and CompBioMed

CompBioMed – My thanks

CompBioMed

UNIVERSITÉ DE GENÈVE

BSC Barcelona Supercomputing Center

Centro Nacional de Supercomputación

UCL

INSIGNEO

Institute for *in silico* Medicine

The University Of Sheffield.

Sheffield Teaching Hospitals NHS

NHS Foundation Trust

RUTGERS

THE STATE UNIVERSITY OF NEW JERSEY

ALMA MATER STUDIORUM

UNIVERSITÀ DI BOLOGNA

Argonne

NATIONAL LABORATORY

Thank You

@bio_comp
www.compbiomed.eu
YouTube Channel: Computational Biomedicine