Biomolecular Simulations@ExaScale
The ExaBioSim project

Sarah Anne Harris
Physics and Astronomy, University of Sheffield, UK

Acknowledgements

HPC Development
James Gebbie-Rayet
Christopher Woods
Charlie Laughton
Alan Real

Ensemble Computing
Antonia Mey
Danny Cole
Shozeb Haider
Edina Rosta
Julien Michel
Nicolas Foloppe

Massive Systems
Sarah Harris
Adrian Mulholland
Syma Khalid
Jon Essex

Multiscale Modelling
Oli Henrich
Agnes Noy
Rosana Collepardo
Davide Marenduzzo

Integration with Experiment
Tom Burnley
Franca Fraternali
Dmitry Nerukh

Research Software Engineer
Rob Welch

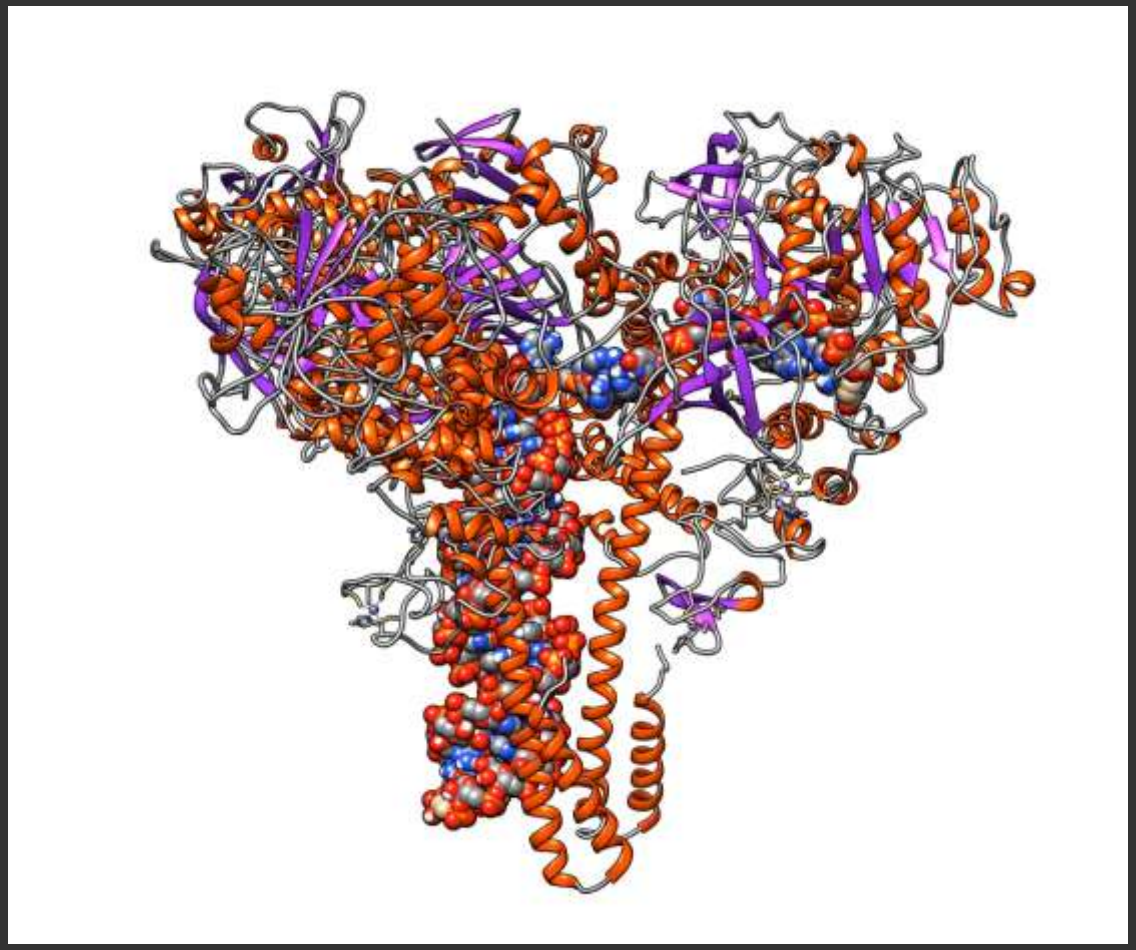
Project Support
Pirjo Johnson

CCPBioSim
HECBioSim
N8 Bede
archer2
ExCALIBUR
UK Research and Innovation
Science and Technology Facilities Council
Engineering and Physical Sciences Research Council

Engineering Supercomputing Platforms for Biomolecular Applications
https://arxiv.org/html/2506.15585v1#S8

What is a Biomolecular Simulation?

SARS-CoV-2
mRTC

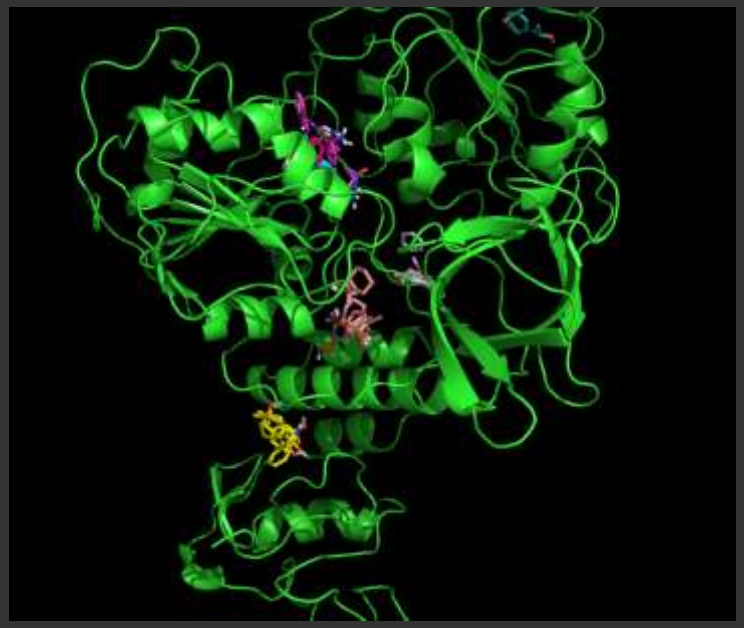


The "computational microscope" `

Dror, Dirks, Grossman, Xu & Shaw, Annu. Rev. Biophys. 2012;41:429-52.

Atomistic MD for Drug Design

Codes are:
AMBER (US)
GROMACS (EU)
NAMD (US)
LAMMPS (US)
openMM (US)



All codes run well on (Nvidia) GPUs but have multi-physics functionality that is less well optimised.

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online

Check for updates

Modelling the active SARS-CoV-2 helicase complex as a basis for structure-based inhibitor design†

Dénes Berta, Magd Badaoui, Sam Alexander Martino, Pedro J. Buigues, Andrei V. Pisliakov, Nadia Elghobashi-Meinhardt, Geoff Wells, Sarah A. Harris, Elisa Frezza and Edina Rosta

Community Support

Our community DO NOT install their own codes.

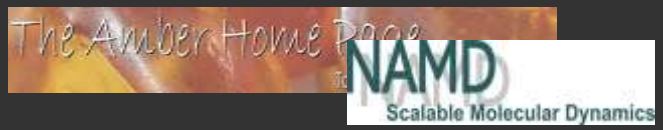
HECBioSim have specialist support who do this on behalf of the community

Advantages:

1. Avoids duplicating effort
2. Code properly installed and optimised for efficiency
3. Core expertise easy to find
4. Prevents multiple copies of the same software
5. HPC configured to benefit community

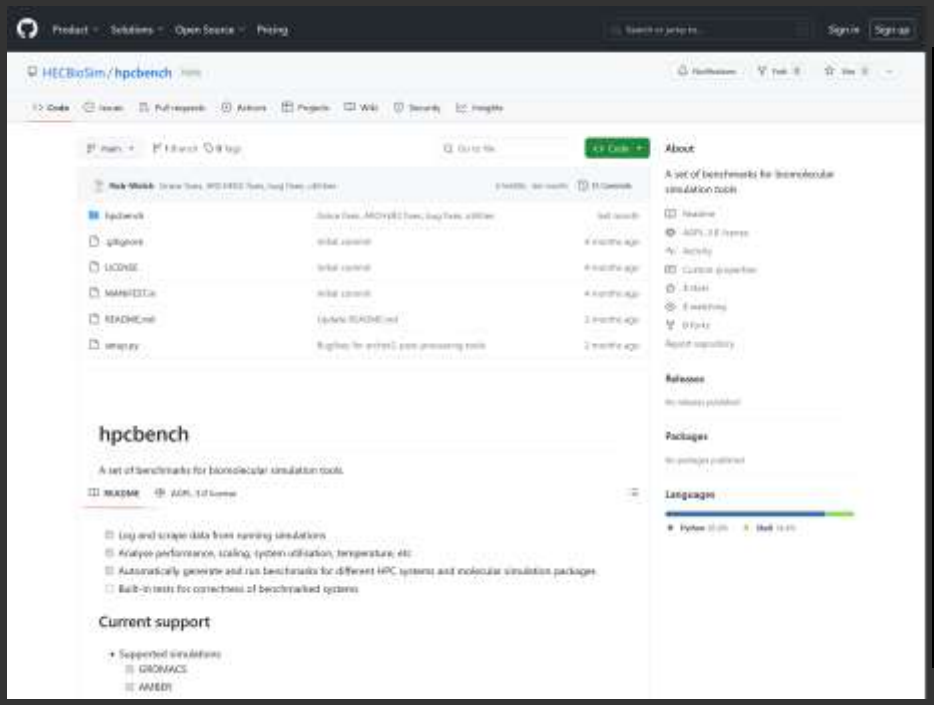
Disadvantages:

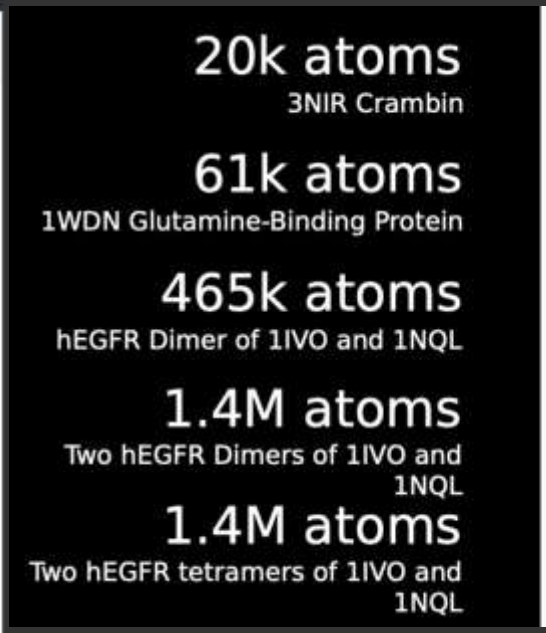
1. Not how all communities work
2. Requires intervention for bespoke codes
3. Support staff require domain knowledge



ExaBioSim Benchmarking Tools

Open Access benchmarking suite incorporating both test systems and performance monitoring tools





https://github.com/HECBioSim/hpcbench

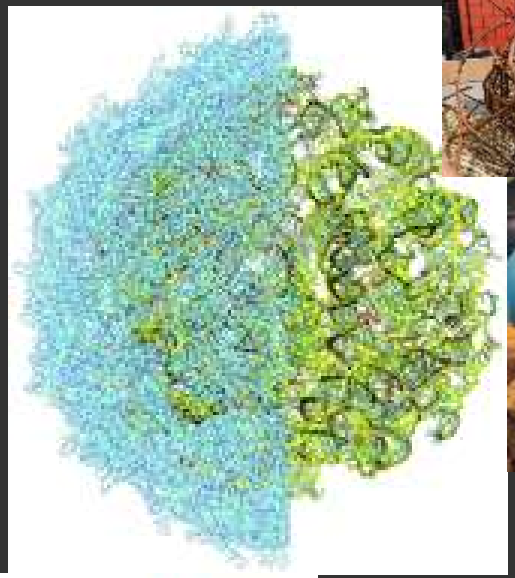
HECBioSim Benchmarking Tools

Benchmark suite helps us support the HECBioSim community to apply for compute time

But our community are diverse….

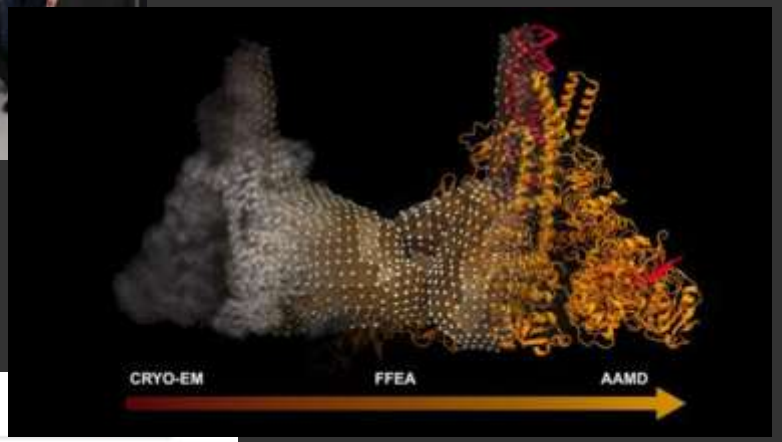
Large systems

> 2 millions atoms
Too big for GPU



Interface with Experiments

Cryo-EM




Faraday Discussions

PAPER

Reconstruction and validation of entire virus model with complete genome from mixed resolution cryo-EM density


Special Issue Paper

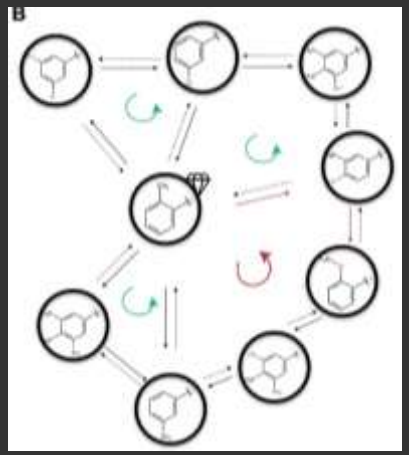
Intelligent resolution: Integrating Cryo-EM with AI-driven multi-resolution simulations to observe the severe acute respiratory syndrome coronavirus-2 replication-transcription machinery in action

Big issues with data transfer

Ensemble Calculations

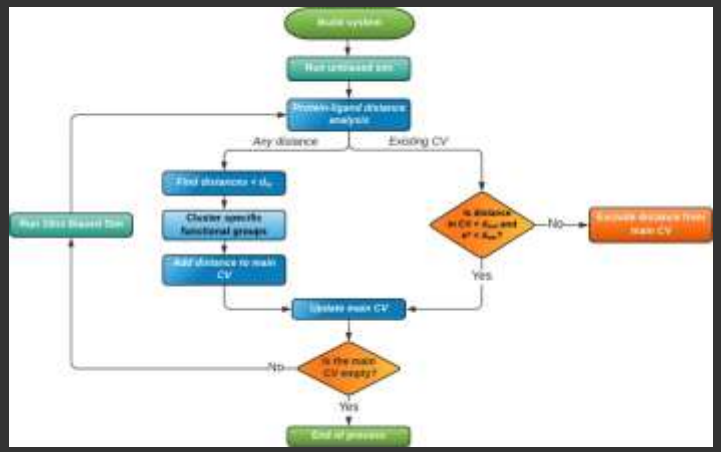
Complex workflows involving *very* large numbers of "small" simulations with complex dependencies and unpredictable in-flight decision-making steps.

Hardware solutions

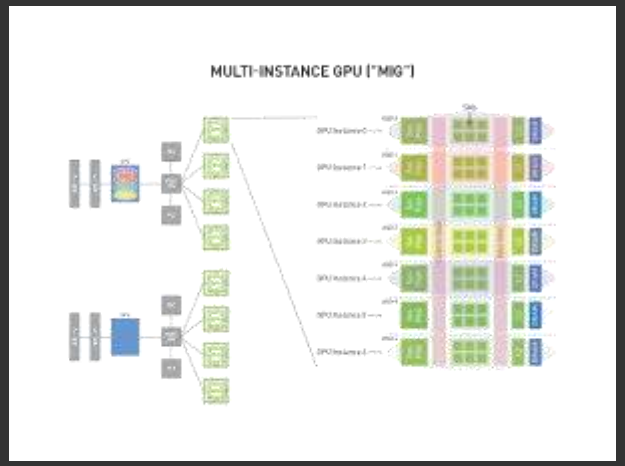


Mey et al., *Living J. Comp. Sci.* **2020**, 2(1), 18378

Enhanced Sampling



Badaoui et al *J. Chem. Theory Comput.* 2022, 18, 4, 2543–2555



Software solutions

dask

FEP

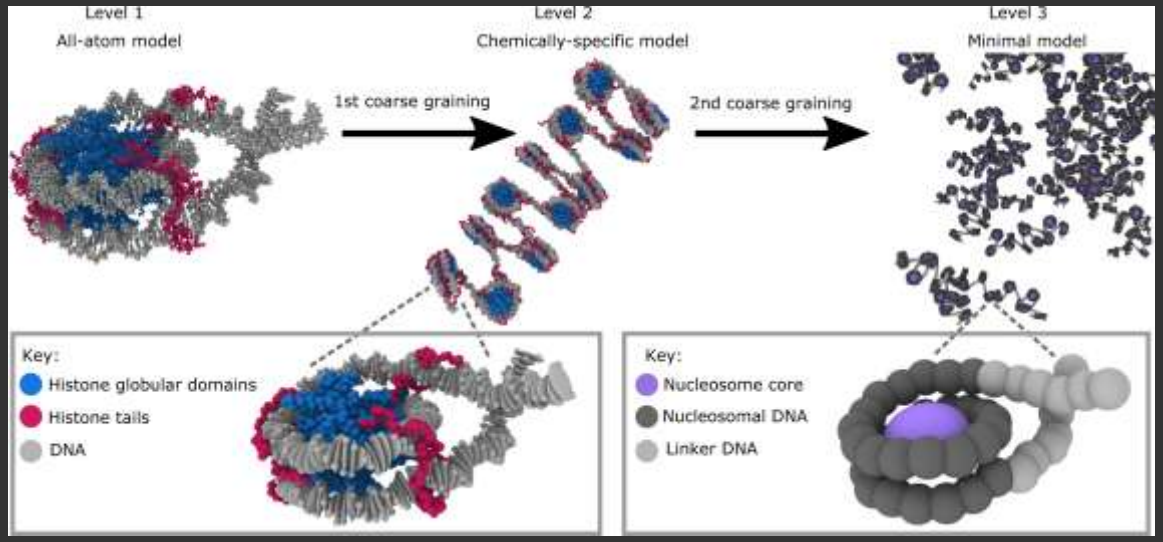
Free energy perturbation

https://bitbucket.org/claughton/crossflow

Multiscale: From QM to continuum

ChemShell
multiscale computational chemistry

QM codes - high memory node requirements – mainly CPU codes



Coarse-grained simulations – CPU only

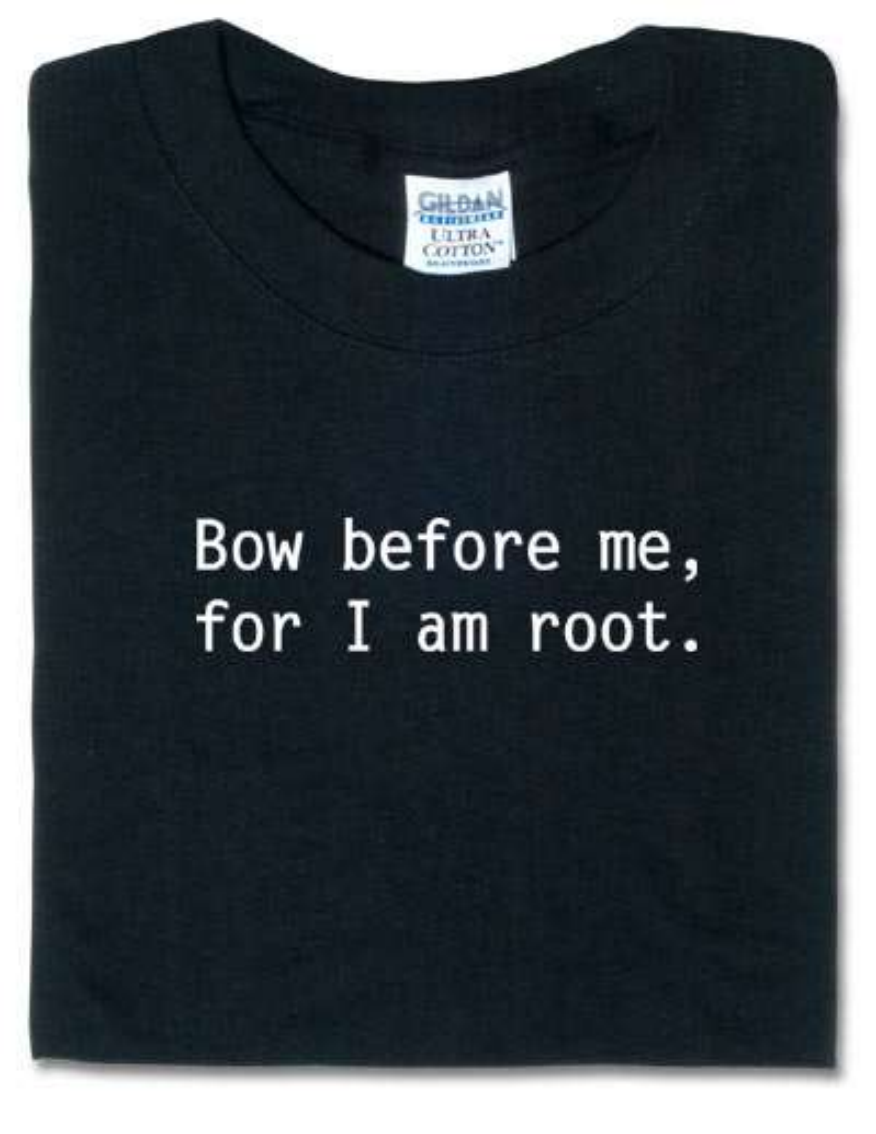
Supercomputer says "NO"

Large systems

> 2 millions atoms
Too big for GPU

Cryo-EM
Transferring large data sets is a problem

SysAdmin says "NO"



Many HPC setups have a limit on the number of jobs in the queue and/or very short queues

ExaBioSim "Wish List" for Exascale

Heterogeneous architectures ~ a bit of everything makes everyone happy (CPU+GPU+BigMemory)

Talk to your users when deciding how to configure your system (e.g. flexible queue times/large job numbers/internet access etc)

RTPs and users should make our conversations more effective.

HPC should be designed for the users you have ~ not those you wish you had.

Anton specialist MD HPC