

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 Colleopardi
Davide Marenduzzo

Integration with Experiment

Tom Burnley
Franca Fraternali
Dmitry Nerukh

Research Software Engineer

Rob Welch

Project Support

Pirjo Johnson



UK Research
and Innovation



Science and
Technology
Facilities Council

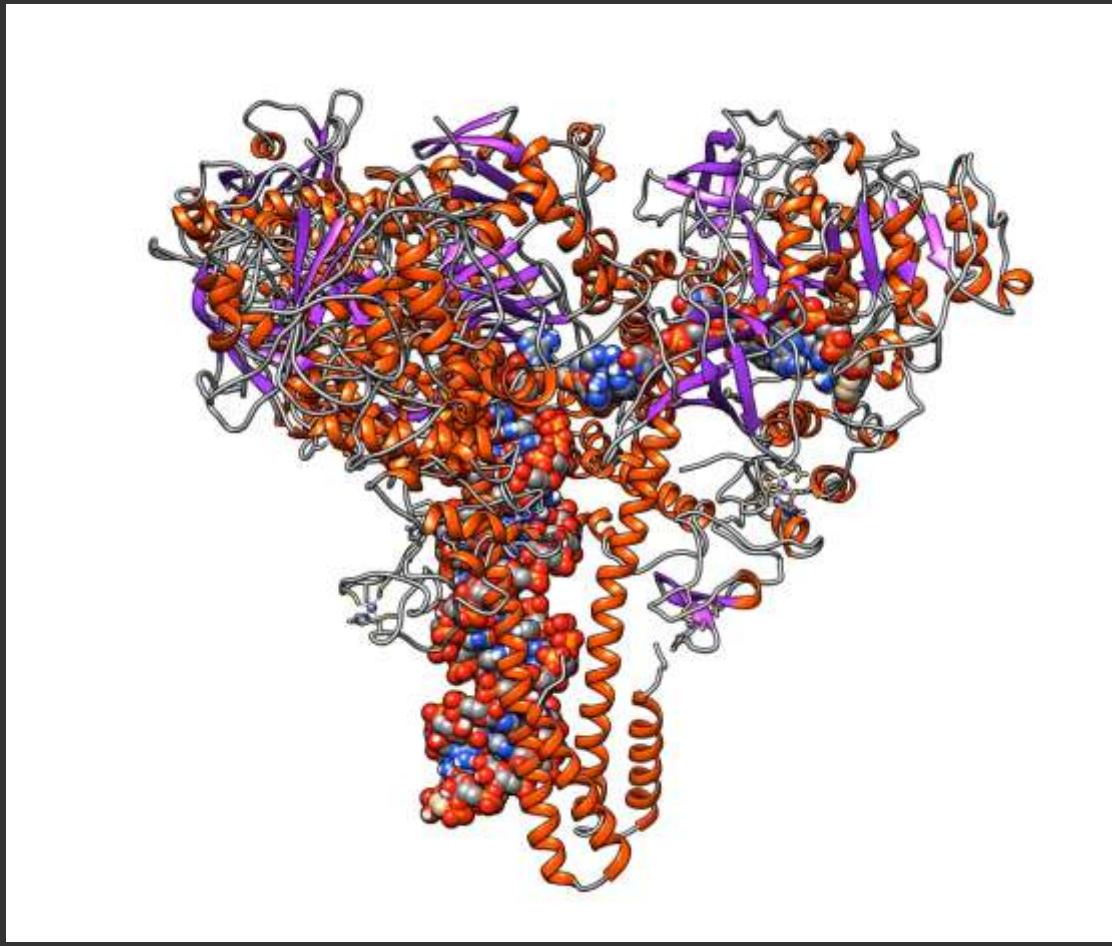


Engineering and
Physical Sciences
Research Council

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

What is a Biomolecular Simulation?

SARS-CoV-2
mRTC



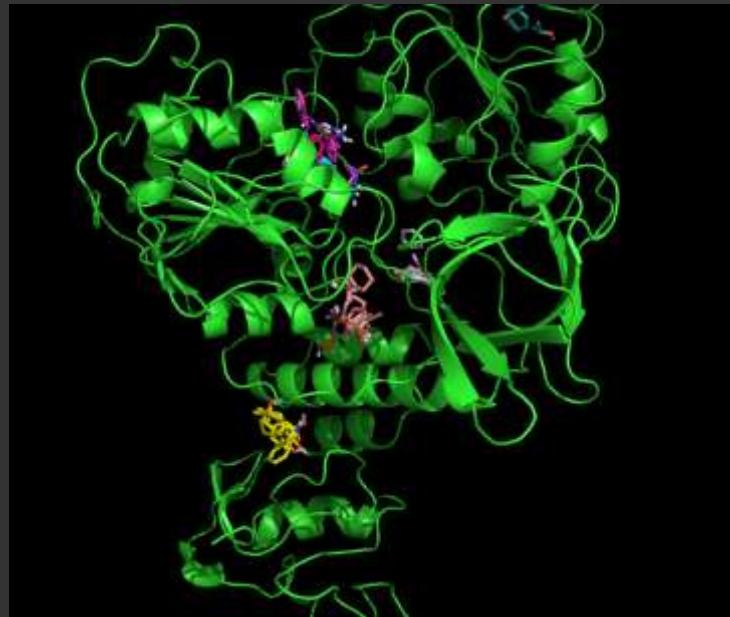
The “computational microscope”

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
EDGE ARTICLE
[Check for updates](#) | [Cite this: DOI: 10.1039/HB00227F](#)
All publication charges for this article have been paid for by the Royal Society of Chemistry.

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

Dániel Berka, Magd Badarou, Sam Alexander Martino, Pedro J. Busque, Andrei V. Pallakov, Nada Elghobashi-Meinhardt, Geoff Wells, Sarah A. Harris, Elsa Frezza and Edina Rosta

ROYAL SOCIETY OF CHEMISTRY
[View Article Online](#) | [View Issue](#)

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

The screenshot shows the GitHub repository page for 'HECBioSim/hpcbench'. The left sidebar lists 'hpcbench' under 'File -> Recent'. The main content area displays a table of benchmarks categorized by system type (HPC, GPU, CPU, MPI, OpenMP) and simulation type (Initial, Equilibration, Production). Each entry includes a brief description, date, and a link to the code. On the right, there are sections for 'About', 'References', 'Packages', and 'Language' (Python 3.6, Python 2.7).

20k atoms
3NIR Crambin

61k atoms
1WDN Glutamine-Binding Protein

465k atoms
hEGFR Dimer of 1IVO and 1NQL

1.4M atoms
Two hEGFR Dimers of 1IVO and 1NQL

1.4M atoms
Two hEGFR tetramers of 1IVO and 1NQL



<https://github.com/HECBioSim/hpcbench>

HECBioSim Benchmarking Tools

The screenshot shows the HECBioSim website with the 'HECTime Calculator' tool. The calculator interface includes dropdown menus for 'System' (JADE2), 'Nodes' (1000), 'Molecules' (100000), and 'Software' (GROMACS 2020.4). Below these are 'Reset' and 'Calculate' buttons. The results section displays various metrics with corresponding icons:

- Estimated Walltime: 11 Days - 16 Hours - 28 Minutes
- Estimated Resource: 280.47 GPU.hrs
- Performance: 85.57 ns/day
- Node Count: 1
- Compute: 1 GPUs
- Power Consumption: 126 kW

Below the results is an illustration of a house with a power plug, labeled "Equivalent to powering the average UK home for 15.89 Days". At the bottom, there is a note about the calculator's purpose and energy equivalence, along with logos for UKRI, CoSeC, and EPSRC.

The information presented in this calculator should only be used to form estimates of resource requests. The figures you obtain in production runs may differ. It is recommended that you use this as a first粗 estimate. You should cross reference the information here with our benchmarks that are publicly available online.

Energy equivalence is provided on figure provided by OVO.

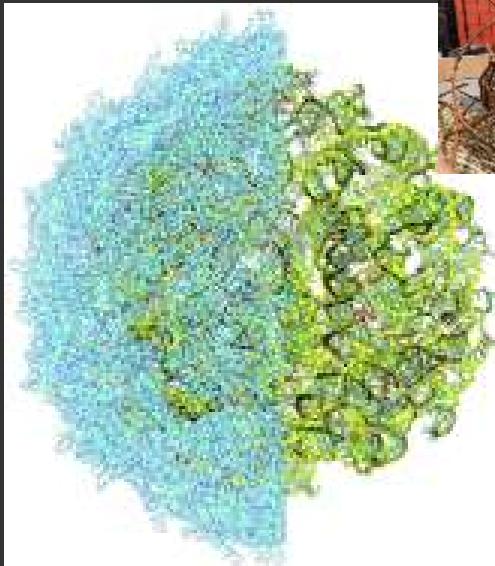
Benchmark suite helps us support the HECBioSim community to apply for compute time

But our community are diverse....

Large systems

> 2 millions atoms

Too big for GPU



Faraday Discussions
Cite this: Faraday Discuss., 2022, 246, 121



View Article Online
<https://doi.org/10.1039/D0FD00012D>

PAPER

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

Vladimir S. Farafonov,¹ Michael Stach,² and Dmitry Nenquin,^{1,✉}

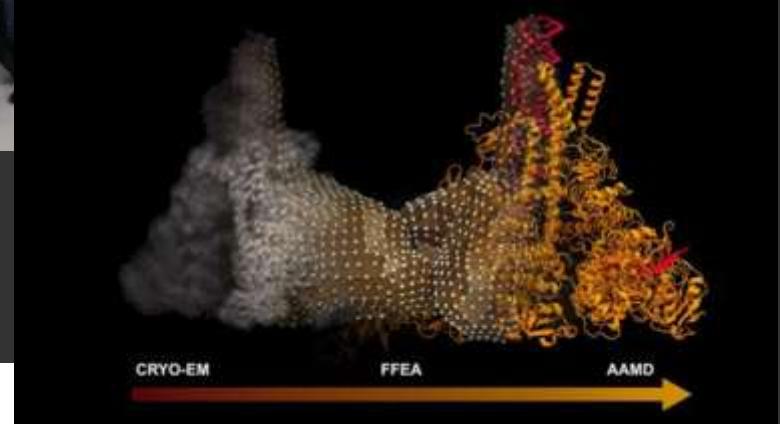


Special issue Paper

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

André Trifunovic^{1,2}, Debin Geng^{1,2}, Michael Salter^{1,2}, Zengqi Li^{1,2}, Alejandro Briseño^{1,2}, Martin Zveroglava¹, Sheng Ma^{1,2}, Anita Chiriac^{1,2}, David Clark¹, David J. Hatch¹, Ian Barford^{1,2}, Lin Huang¹, John McPhee¹, Marcell Esnault¹, Huiyong Yu¹, Junji Xie¹, Antonette Tauris¹, Nikolai Sathules¹, Tatjana Rame¹, Jessica Liu¹, Noah Trottbeck¹, Guelher Weltz^{1,2}, Venkatesh Murugesu², Thomas Gibbons¹, James Phillips¹, S. Chakra Chennapudi¹, Ian Fostier^{1,2}, Rick Stevens^{1,2}, Antonio Amat-Plana^{1,2}, Venkateswaran Vasantha¹, John E. Stone¹, Emad Tajuddin¹, Sarah A. Harris¹, and Arvind Ramanathan¹

Interface with
Experiments
Cryo-EM

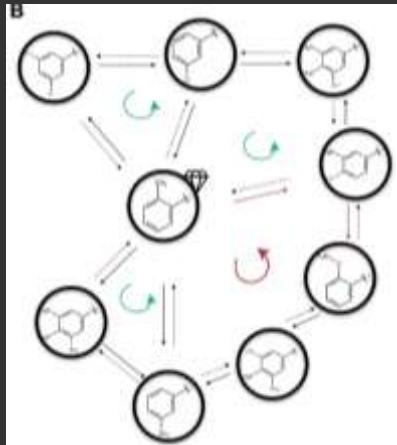


Big issues with
data transfer

Ensemble Calculations

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

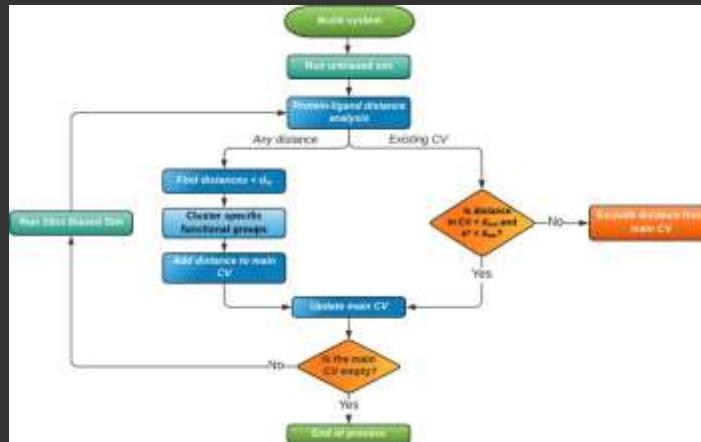
Hardware solutions



Enhanced Sampling

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

FEP
Free energy
perturbation

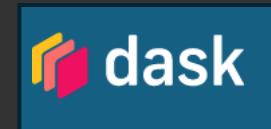


Badaoui et al *J. Chem. Theory Comput.* 2022, 18, 4, 2543–2555

<https://bitbucket.org/claughton/crossflow>



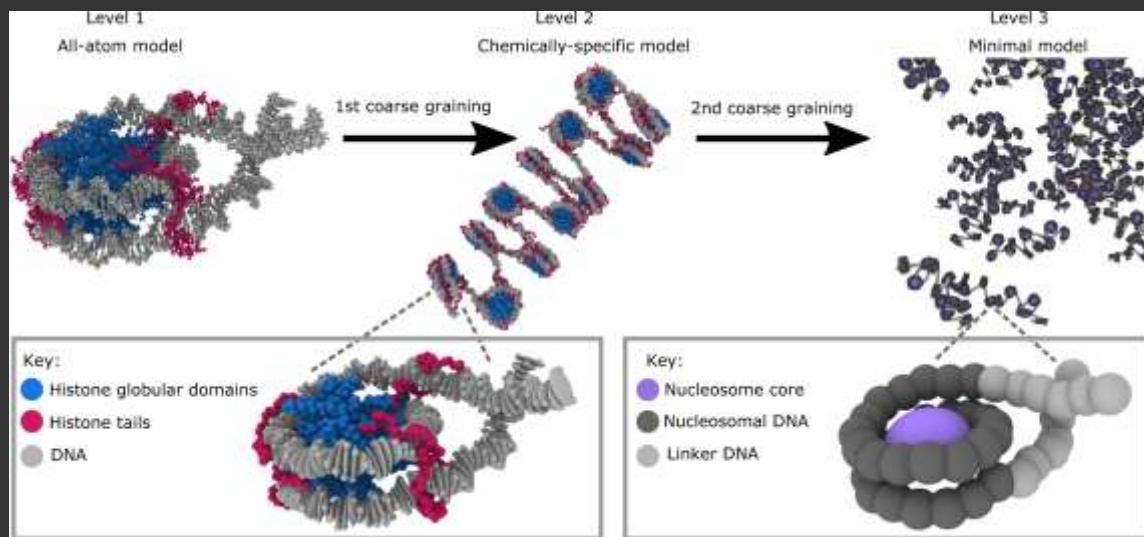
Software solutions



Multiscale: From QM to continuum



QM codes - high memory node requirements – mainly CPU codes



Coarse-grained simulations – CPU only

Supercomputer says “NO”

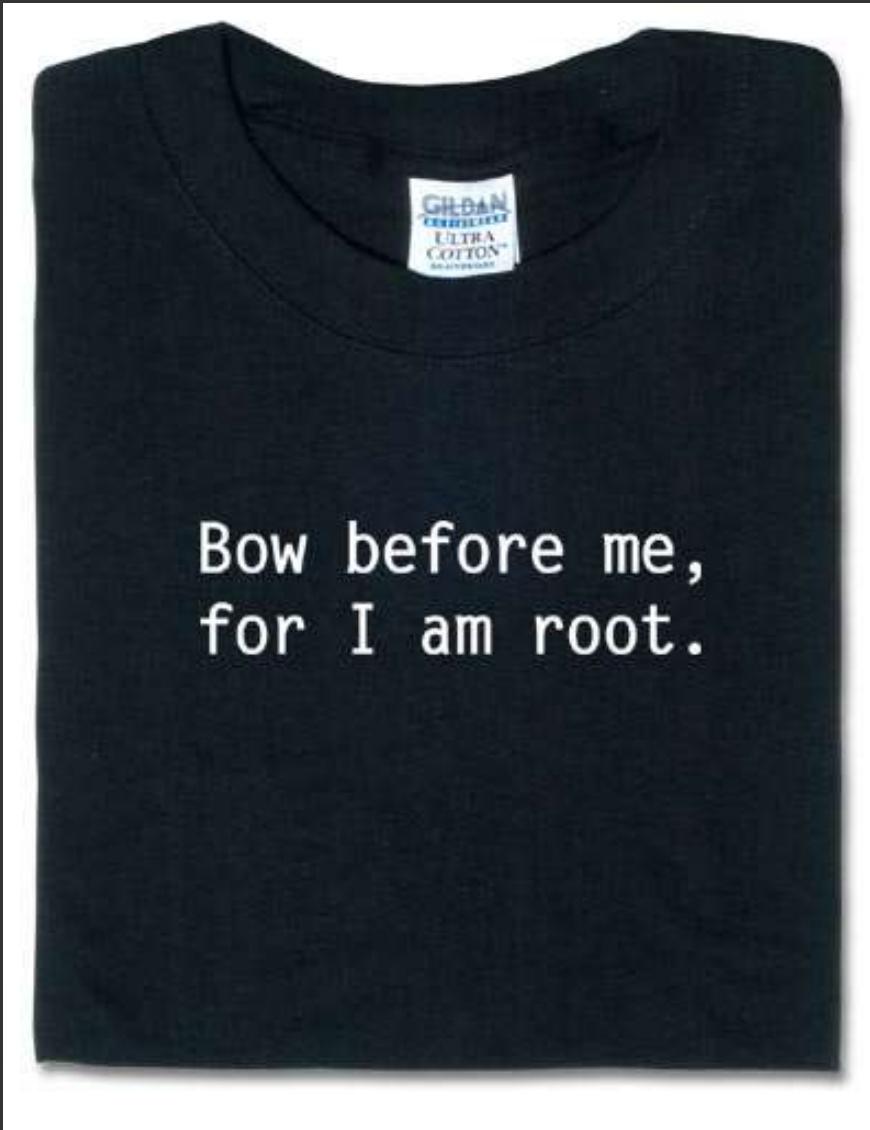


Large systems

> 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

ACM DIGITAL LIBRARY Association for Computing Machinery

Browse About Sign In

Journals Magazines Proceedings Books SIGs Conferences People Search ACM Digital Library

Conference Proceedings Upcoming Events Authors Affiliations Award Winners

Home > Conferences > SC > Proceedings > SC'21 > Anton 3: twenty microseconds of molecular dynamics simulation before lunch

RESEARCH ARTICLE OPEN ACCESS X in ⓘ f =

Anton 3: twenty microseconds of molecular dynamics simulation before lunch

Authors: David E. Shaw, Peter L. Adams, Andrew A. Chod, Joanne A. Domitrovic, ... + 83
Authors Info & Details

SC'21 (Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis) • November 2021 • Article No.: 1 • Pages 1–11 • https://doi.org/10.1145/3487367.3487367

Published: 13 November 2021 Publication History Check for updates

