Grand Challenges

The High-End Computing Consortium for Biomolecular Simulation (HECBioSim)

Over the next five to ten years, a key challenge for biomolecular simulation is to bring together Simulation and AI to tackle multi-interfacial problems at larger time and length scales.

The Community

HECBioSim ww.hecbiosim.ac.uk

HECBioSim promotes the use of UK High Performance Computing facilities amongst researchers in the biomolecular simulation domain as well as increasing the uptake by non-traditional users and experimentalist scientists interested in the application of molecular simulations to specific problems. Molecular simulations provide high resolution information about the dynamics of biomolecules and complex systems which is often the link between structure and function – such high resolution and details of dynamics are difficult to obtain by experimental methods alone. HECBioSim since its inception has been responsible for allocation of resources on Tier1 (ARCHER2) and subsequently also several UK TIER2 *via* its Resource Allocation Panels. It develops high performance workflow tooling as well as supporting HPC centres with software installation and optimisation. HECBioSim supports its sister consortium CCPBioSim by providing workshops and training on HPC best practice and run-time optimisation. HECBioSim maintains a benchmark suite that has become the international *de-facto* standard for characterisation of simulation performance for all major codes in the biomolecular simulations domain. We support UK applicants for HPC resources through information on our website, tailored advice and through specialised resource calculators we have developed. HECBioSim is open to new members, and anyone in the remit area with a full-time UK position in academia may submit a proposal to our Resource Allocation Panels for UK HPC time.

The Challenge

Tackling problems at larger scales and integrating next generation technologies into simulation workflows

The landscape of both technology and science has changed significantly over the last 5 years. On the technology front, we have seen the advent and rise of data intensive computing, much of which falls under the AI umbrella. Within biomolecular science, most notably structure prediction technologies, such as AlphaFold3 and machine learnt interatomic potentials offer a wealth of new opportunities. We have also seen an explosion in the diversity of HPC architectures with new vendors entering the HPC market with innovative solutions in both the CPU and GPU hardware spaces. The scientific landscape has experienced a

shift from classic sources of protein structures such as X-Ray diffraction and Nuclear Magnetic Resonance which were somewhat limited to smaller structures, towards imaging-based methods such as Cryo-Electron Microscopy and Electron Tomography - this has brought a vast wealth of new structural data of larger structures and more complex systems.

Addressing the next generation of pressing biological questions in the biomolecular simulation space, is going to require pushing our codes and methods beyond their current capabilities. Accessing the time and length scales required for compatibility with evolving experimental techniques requires not only a step change in how our codes and methods scale but also radical increases in the community accessible computational resources. Furthermore, physics-based simulation and AI are clearly complementary and for understanding the structure-dynamics-function relationships of biomolecules they are usually co-dependent and thus not substitutes for each other, but rather support and enhance each others capabilities and scope.

The integration of AI and simulation such that they augment each other to enable tackling problems that are currently out of scope for both presents an interesting engineering problem. Currently used HPC platforms and software workflows are not engineered for such multimodal approaches, consequently our significant challenges in the future are ones around technology and doing the necessary engineering that will allow us to address the most pressing science questions of tomorrow.

The Solution

Engineering platforms and workflows to greater enable researchers to deploy next generation science

To successfully push our current biomolecular simulations capabilities to enable new discoveries by incorporating AI and scaling to extended spatio-temporal scales we must appropriately evolve the engineering of the whole UK HPC ecosystem. The current generation of HPC are not configured to enable fully multimodal workflows, where simulation and AI can run side-by-side and exchange real-time information. We must engineer the software frameworks to enable researchers to build, configure and deploy reproducibly such complex workflows. We must do this at the next generation of HPC scale and across a diverse array of hardware configurations than has existed in the past decade. Further engineering of our codes and the tooling that support running them is required to enable researchers to define problems which require our codes to talk to next generation technologies such as AI across a wider array of interfaces into the simulation pipeline.

The Outcome

Exploiting next generation platforms and technologies will fertilise new research areas with exceptional potential in healthcare

Grand Challenges

Biology is inherently complex and multidimensional with problems ranging from those that focus on individual molecules up to entire human body, even at the minute scale problems are still highly multi-scale in nature ranging from small molecules up to the whole human cell.

Understanding more about how these biological systems behave and function at each scale leads to direct application in the design and discovery of new therapeutics and beneficial outcomes for our society.

More Information

CoSeC
www.CoSeC.ac.uk
CoSeC@stfc.ac.uk

HECBioSim
www.hecbiosim.ac.uk
hecbiosim@bioch.ox.ac.uk