CompBioMed

Newsletter Issue No. 13
May 2024

Welcome

Prof. Peter V Coveney
Principal Investigator & CompBioMed Coordinator

Welcome to the final newsletter of the CompBioMed2 Centre of Excellence. Our focus in 2023-24 has been to conclude the CoE strongly and leave a lasting legacy.

The 2023 CompBioMed Conference in Munich was warmly received, delivering many interesting talks, poster, and discussions. We look forward to continuing the series with future editions.

Notably, the CompBioMed team at UCL secured access to the exascale supercomputer Frontier via the INCITE awards. This will advance projects in drug discovery and stroke simulation, leveraging Frontier's unparalleled computing power.

The CompBioMed2 project, funded by the EU's Horizon 2020 programme, made significant strides in computational biomedicine. We enhanced computational support, optimized biomedical modeling codes, and supported data management with FAIR principles. Strong collaborations with medical professionals highlighted the practical benefits of computational modeling in clinical settings. We also advanced education by integrating HPC training into medical curricula and offering courses in computational biomedicine.

Our outreach efforts, including articles, media appearances, a science book, and an IMAX film, showcased CompBioMed2's impact. We organized numerous events, providing extensive training and seminars, and involved clinicians in dissemination activities. In summary, CompBioMed2 significantly enhanced computational capabilities, fostered clinical collaborations, promoted innovation, and established a sustainable future for computational biomedicine, benefiting both the scientific community and healthcare practices.

The legacy of CompBioMed will live on. Several of our British partners joined forces to successfully land a UK version of our CoE, funded by UKRI (www.compbiomedx.org). Supported by several of our European partners acting as external collaborators, this project will allow us to continue to support the aims of the CompBioMed2 CoE until the end of March 2025.

Deep learning and machine learning protein-ligand absolute binding affinity prediction models

Jose Carlos Gomez Tamayo and Gary Tresadern

At Janssen we set out to explore Deep learning and machine learning (DI/ML) protein-ligand absolute binding affinity prediction models. Limitations of previous work have highlighted the lack of available data in the public domain as well as several biases leading to overfitted models which do not generalize. Our goal was to improve the quantity and the quality of data to overcome these issues, for this we anticipated work on a lager computational scale.

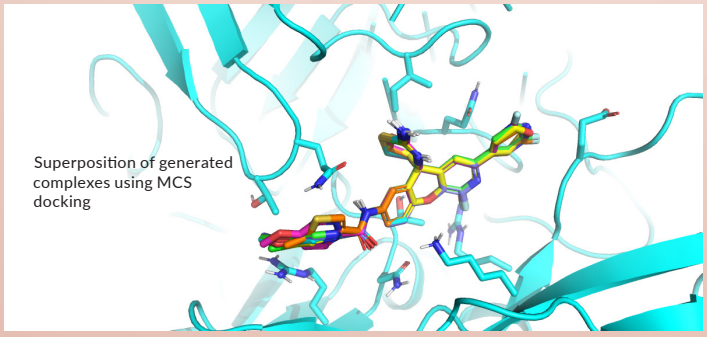
We developed a protocol to augment the number of protein ligand complexes based on careful curation of both public and internal data and very restrictive maximum common substructure (MCS) docking. We identified and chose additional ligands with measured bioactivities and that were like experimentally solved protein ligand complexes. These could be docked with high confidence and solutions that would be as close as possible to a hypothetical experimental solution.

We generated ~30000 high quality complexes representing more dense structure-activity relationships from internal lead-optimization datasets and the ChEMBL database. Additionally, we generated decoy poses for each complex and included inactive compounds to ensure models focus on the learning of interactions and not only ligand information.

We used this data to train in house models as well as state-of-the-art DL architectures and compare the performance training with and without augmented data. For testing, we used targets from LO internal data and ChEMBL to ensure our models are challenged similarly to a real LO scenario. Models trained with internal data showed more generalization power and predictivity than models solely training with only PDBbind data. Comparison to glide docking showed very similar performance in terms of correlation but better hit enrichment. Interestingly, models trained with internal data were also more robust, as they showed predictivity in challenging systems containing flexible loops and coordinated metals where molecular docking was not predictive. The results were encouraging, and a manuscript will be submitted shortly.

The field of DL/ML affinity prediction offers much promise, but work is still needed to reach generalizable models that learn the properties that drive bioactivity.

We thank the full team at Janssen, Jose Carlos Gomez Tamayo, Mazen Ahmad, Lili Cao, Gary Tresadern. Jose and Lili were both partly funded via CompBioMed-2 and we thank the wider CompBioMed-2 community.



Superposition of generated complexes using MCS docking

Virtual Heart display at the Engineers Gallery, Science Museum, London. © Science Museum Group.

Science Museum exhibition: simulation of a beating human heart

Jazmin Aguado-Sierra, Barcelona Supercomputing Center

On the 7th of November of 2023 a complex and beautiful simulation of a beating human heart was unveiled in the Engineers gallery at the Science Museum. Created by bioengineer Dr Jazmin Aguado-Sierra using scans of her own heart, it shows the complex interactions between electrical impulses, muscle contraction and blood flow in the heart - a feat only possible using supercomputer power.

Dr Aguado-Sierra used her own data captured from sources including blood tests, electrocardiographs that measured electrical impulses and Magnetic Resonance Imaging (MRI) scans of her heart tissues. The large data sets were then inputted into mathematical equations that described her heart's workings and her virtual heart was brought to life by MareNostrum 4, a high-performance supercomputer based at Barcelona Supercomputing Center, to reproduce her heart in remarkable detail.

The simulation contains 5 billion data points, with each snapshot of her virtual heart changing every hundred thousandth of a second, which works at different length scales using distinct kinds of physics. If a human attempted to calculate them all, it would take almost 57 billion years to complete and analyse or interpret the findings. But using MareNostrum 4 and complex mathematical techniques employed by Dr Aguado-Sierra and her colleagues, it takes just 9 hours to solve.

Efficient GPU development using C++ standard language parallelism

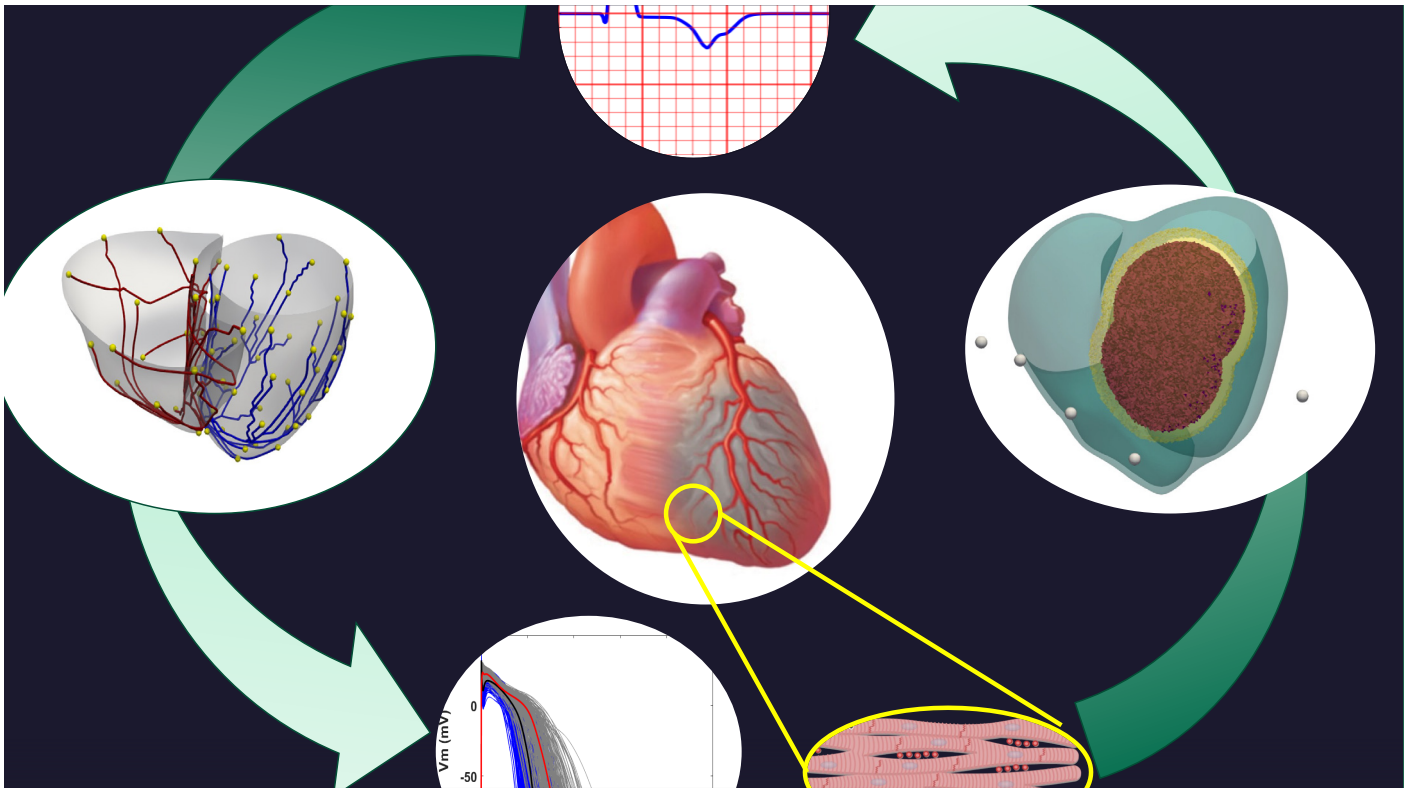
Jonas Latt, Université de Genève

Although general-purpose GPU programming has been practiced for several decades, it is only recently that the use of GPUs has become dominant in standard research codes for scientific computing. This paradigm shift is associated with the emergence of high-level programming techniques such as OpenACC, OpenMP (which enable GPU acceleration), and built-in parallel programming language features that allow for the combination of parallel efficiency and ease of use needed in a scientific team with diverse technical skills.

Within CompBioMed, the University of Geneva focused on standard language parallelism in C++, which, as of C++17, allows for multi-core and GPU parallelism without language extensions. This approach extends basic parallel features, such as parallel loops or sections, with higher-level constructs. Many algorithms from the standard template library have a parallel counterpart, making it easier to parallelize and accelerate scientific code. Decomposing the underlying algorithm into the appropriate parallel components of the language is necessary for this exercise.

To train language users in this new approach to programming, we collaborated with NVIDIA to create a tutorial. The tutorial was taught at several instances of the Supercomputing and ISC conferences and is also available in an open-source format (https://gitlab.com/unigehpfs/cpp-hpc-tutorial). The format includes the full tutorial in a Jupyter notebook and the compilation and execution environment in the form of a Docker container.

For advanced illustrations, the open-source STLBM library (https://gitlab.com/unigehpfs/stlbm) provides reusable code snippets for GPU-based lattice Boltzmann simulation. Additionally, the LEDDS software project demonstrates the use of parallel algorithms in particle simulations (Discrete Element method) and their coupling with fluid simulations.



From clinical data to subcellular processes and back: digital twins of heart attack

Cardiac Digital Twins for Heart Attack Survivors

Zhinuo Jenny Wang and Julia Camps, University of Oxford

In an exciting step towards personalised virtual therapy evaluation, researchers at the University of Oxford recently demonstrated how to model the aftermath of heart attacks from the subcellular to the whole organ mechanisms and developed a next-generation cardiac digital twinning technology. This research will be featured in eLife and Medical Image Analysis in 2024.

The damage caused to the heart muscle tissue during a heart attack causes areas of the heart to lose the ability to contract and form scar tissue. Heart attack survivors can be at great risk of deadly arrhythmias depending on the specific subcellular characteristics of their heart damage and recovery process. The team at Oxford have developed virtual heart models to investigate how heart attacks change the subcellular properties of the human heart. Using virtual simulations, the researchers demonstrated how different regions in the heart develop subcellular abnormalities and how these abnormalities are reflected in the electrocardiogram recordings. Their findings put us one step closer to being able to have digital twins of heart attack survivors to remotely monitor their arrhythmic risk, provide preventive therapies, and improve their chances of recovery.

To enable this vision of patient-specific digital twins, the team at Oxford has also developed a next-generation tool that uses patients' electrocardiograms to build their hearts' digital twins. The tool uses principles from artificial intelligence to characterise an individual's cardiac properties related to deadly arrhythmias, such as the electrical conduction system in the heart and subcellular abnormalities.

This research provides a fast and non-invasive way of detecting heart damage on a personalised basis and conducting virtual therapy evaluation. Moreover, these new tools can change how cardiology is conducted in the clinic and accelerate therapy development by enabling virtual clinical trials.

https://www.biorxiv.org/content/10.1101/2022.02.15.480392v2

https://arxiv.org/abs/2306.13740

https://arxiv.org/abs/2401.10029

Training activities

Carlos Teijeiro Barjas, SURF

Within CompBioMed2 the training activities have been considered as key contributions to engage with our stakeholders and disseminate knowledge within the community. Designing a suitable portfolio of courses and the compromise to keep the efforts alive in the future have been essential tasks that are particularly important now that the project finally approaches its end.

From the beginning of the project, CompBioMed2 has put a very strong focus in connecting with the essential needs of medical students and designed training to complement the content of their taught programmes of study by providing relevant computational upskilling. Considering that the essential computational topics are usually covered by the training provided by other European projects (in the past by PRACE [1], and currently by EuroCC [2]), CompBioMed2 has been continually engaged with these two projects and also other relevant European initiatives (e.g., EOSC-hub [3]) to always keep a connection with the forefront of European training initiatives.

In these years, the "Short Course on HPC-Based Computational BioMedicine" has been considered as a flagship activity for CompBioMed2, because it includes most of the essential topics of interest provided by many representative partner institutes, where the materials are condensed for delivery within a relatively short amount of time. The collaboration with PRACE and EuroCC gave out additional important results by having presence of CompBioMed2 flagship applications in courses designed for application programmers, like "MPI and OpenMP in Scientific Software Development". In addition to it, other courses have been designed to answer some of the challenges that medical students and experts face in their regular work, such as the secure treatment of data, which was addressed in courses like "Sensitive Data Management and EUDAT Services" and the hackathon "Data Management and Publication".

Ongoing discussions are taking place now on how to sketch the future after CompBioMed2. The courses mentioned above, together with regular introductory sessions on the use of compute clusters, are part of the ideal catalogue that should be continued in the coming years. New training calls coming from EuroHPC-JU may come to ensure a framework of collaboration like CompBioMed2 has done so far, but using the existing projects (particularly EuroCC together with the CASTIEL coordination action) should already be a good starting point.

[1] https://prace-ri.eu/training-support/training
[2] https://www.eurocc-access.eu
[3] https://www.eosc-hub.eu/training-material
[4] https://www.bsc.es/

Find CompBioMed online

Our website (www.compbiomed.eu) is full of all the latest news and information about CompBioMed, including further information on our Partners and Associate Partners, and events. We have an active and growing following on Twitter (@bio_comp), LinkedIn (CompBioMed) and we have our own YouTube channel (YouTube Computational Biomedicine), where you can watch various types of video content, including recordings of presentations at previous events and webinars, as well as our *Virtual Humans* and *The Next Pandemic* films.

https://youtu.be/1FvRSJ9W734

CompBioMed Coordinator:
Peter V. Coveney (p.v.coveney@ucl.ac.uk)
Project Manager:
Hugh Martin (h.s.martin@cbkscicon.com)

This Centre of Excellence has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823712.