Grant agreement no. 675451

CompBioMed

Research and Innovation Action
H2020-EINFRA-2015-1
Topic: Centres of Excellence for Computing Applications

D2.4 Report on the Impact of Modelling and Simulation within Biomedical Research as enabled by CompBioMed

Work Package: 2

Due date of deliverable: Month 36

Actual submission date: September 30, 2019

Start date of project: October 01, 2016 Duration: 36 months

Lead beneficiary for this deliverable: *BSC*
Contributors: *Bull, UvA, UCL, UNIGE, UPF, USFD, JAN, LTG*

Project co-funded by the European Commission within the H2020 Programme (2014-2020)		
Dissemination Level		
PU	Public	YES
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, as referred to in Commission Decision 2001/844/EC	

Disclaimer

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Table of Contents

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

1. Version Log

Version	Date	Released by	Nature of Change
V0.1	01/07/2019	Mariano Vázquez	Draft of template
V1.0	30/09/2019	Emily Lumley	Final Submission

2. Contributors

Name	Institution	Role
Mariano Vázquez	BSC	Principal Author
Okba Hamitou	Bull	Co-author
Gabor Zavodszky	UvA	Co-author
Peter Coveney	UCL	Co-author
Alfons Hoekstra	UvA	Co-author
Bastien Chopard	UNIGE	Co-author
Jonas Latt	UNIGE	Co-author
Alfonso Santiago	BSC	Co-author
Gavin Pringle	UEDIN	Co-author
Adriá Pérez	UPF	Co-author
Andrea Townsend-Nicholson	UCL	Co-author
David Wright	UCL	Co-author
Alberto Marzo	USFD	Reviewer
Herman Van Vlijmen	JAN	Reviewer
Peter Coveney	UCL	Reviewer
Emily Lumley	UCL	Reviewer

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

3. Acronyms and Definitions

Acronyms	Definitions
AlyaCCM	Alya Cardiac Computational Model
API	Application Progressing Interface
ARP	Architecture Review Board
AVX	Advanced Vector Extension
BAC	Binding Affinity Calculator
CGNS	CFD General Notation System
NetCFD	Ninf computational component for CFD
CFD	Computational Fluid Dynamics
CHASTE	Cancer, Heart and Soft Tissue Environment
CoE	Centre of Excellence
CPU	Central processing unit
CUDA	Compute Unified Device Architecture
DEM	Discrete Element Method
ESMACS	Enhanced Sampling of Molecular Dynamics with Approximation of Continuum Solvent
FSI	Fluid-Structure Interaction
GB	Gigabyte
GPCR	G-protein coupled receptor
GPU	Graphics processing unit
HDF5	Hierarchical Data Format 5
HemoCell	High PErformance MicrOscopic CeLlular Library
HPC	High Performance Computing
HTBAC	High-Throughput Binding Affinity Calculator
HTMD	High-Throughput Molecular Dynamics
I/O	Input / Output
ILP	Instruction-level parallelism
IPC	Instruction per cycle
KPI	Key Performance Indicator
LBM	Lattice Boltzmann Method
LIC	Line Integral Convolution

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

MD	Molecular Dynamics
ML	Machine Learning
MPI	Message Passing Interface
NUMA	Non-uniform Memory Access
OpenCL	Open Computing Language
OpenMM	A high performance toolkit for molecular simulation
OpenMP	Open Multi-processing
PDB	Protein Data Bank
RAM	Random access memory
SIMD	Single Instruction Multiple Data
SPMD	Single Programme Multiple Data
SRAM	Static random access memory
TIES	Thermodynamic Integration with Enhanced Sampling
UMA	Uniform Memory Access
WP	Work Package

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

4. Executive Summary

In this document we summarize the impact of the modelling and simulation research activities in CompBioMed, including:

- High quality research done either by the project full and associate partners within the project or outside the bounds of CompBioMed but under its influence (for instance, research done by a partner as a subproduct of what's done in the project)
- High quality publications produced
- Collaborations: either previous collaborations enforced or new collaborations initiated by the CoE's creation
- Scope widening, especially towards the medical world: clinical, academia, companies, etc.
- Technology transfer to spinoffs and startups from research related to the project.

Since the very conception of the proposal and led by our own long and consolidated experience we were conscious that the impact of our work in the community would be directly related to the way in which we perform our research. We all do research in the biomedical realm with some common characteristics: (a) we are developers of complex computational modelling tools which are (b) very expensive on computational grounds, requiring large amounts of computational power (i.e. supercomputers), (c) within a domain in which multiscale modelling is the norm. In our field, complexity means "closer to reality" which in turn means "multiscale", where complex parallel workflows, huge input and/or output datasets, non-conventional mathematical models, parallel programming techniques and challenging visualisation techniques are essential. We have observed that the simultaneous way in which we address features (a) and (b), from the standpoint of (c), guarantees high impact. Therefore, this document describes our research path and then summarizes its impact.

5. Multi-scale problems and large-scale computational resources

In science [2], we look to convincingly explain the processes at work in phenomena that we observe, as well as to predict what will occur before it does so. Predictions of real-world events all need substantial quantities of data and validated computational models together with the execution of many high-fidelity simulations. In many cases, the models that describe the phenomena are multiscale, as their accuracy and reliability depend on the correct representation of processes taking place on several length and time scales. Multiscale phenomena are everywhere around us [4–10]. If we study the origin and evolution of the Universe [11,12] or properties of materials [13–17], or develop fusion as a potential energy source of the future [18], in all these cases and many more we find that processes on quite disparate length and time scales interact in strong and nonlinear ways. In short, multiscale modelling is ubiquitous and progress in most of these cases is determined by our ability to design and implement multiscale models of the particular systems under study [4,9,19,20].

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Multiscale means a disparity on modelling schemes span along wide space and time ranges. In the biomedical realm, the ranges go from molecules all the way up to healthcare organisations, from nanoseconds to years. This vast space-time region, with all the scales tightly interacting together is the domain where CompBioMed is leaving its footprints.

5.1. The Computational Biomedicine Panorama

CompBioMed strives to model using computational means, ranging from software to hardware and from mathematical grounds to programming strategies, the complex systems involved in biomedical research and its applications.

Figure 1 and Figure 2 display the panorama. Figure 1 depicts the intervening scales and organisational levels in which CompBioMed acts. At the top, there is the Healthcare Organisation, which gathers all our efforts and focuses them. At the bottom there is the molecular level, where the deepest roots of the genome reside. Compared to other complex multiscale systems, in biomedical ones the lowest scales are largely responsible for emergent properties all the way up. Conversely, upper scales feedback to lower ones, creating a non-linear coupling loop. This does not mean that a complete system's understanding is required to simulate at all scales, but it shows that neglecting the effects of any single scale upon the others may lead to erroneous predictions. Considering that the final target of what we do is human healthcare, wrong predictions are highly undesirable.

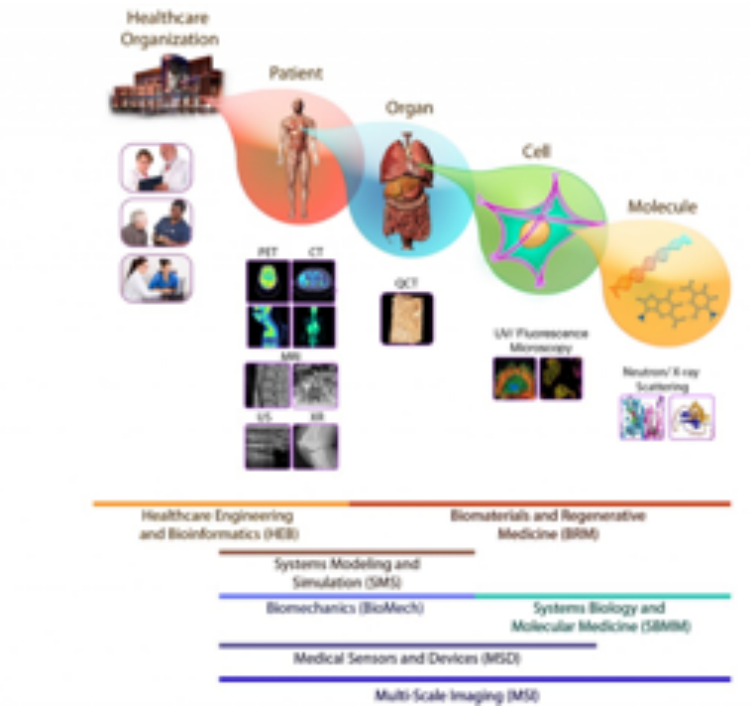

Figure 1. Scales and organisation levels in the biomedical realm (from the IBME - UTK.)

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Error! Reference source not found. shows also another aspect of the problem. In its lower half, scales are grouped by the different scientific specialities dealing with them. The fact that there is no integrative scientific or technical specialty expresses the difficulty in joining efforts to reach a common objective, emphatically conveying the importance of multi-disciplinary projects like ours. From the human point of view, the different and frequently disconnected specialities lead to communication problems among researchers and practitioners. Their background and training are different not just in content but also in orientation, which can even on occasion lead to dismissive behaviour among researchers working at different levels. Perhaps the most transversal of all is the multiscale imaging, which is frequently the glue binding research groups (as the authors of this report have frequently observed). In the middle, the figure shows the wide variety of modes of information acquisition, from purely medical expert observations to X-ray scattering, adding another layer of difficulty when integrating the scales, organisation levels and dedicated researchers.

Figure 2 shows the disparities in the relevant spatial dimenions of each scale. Observe that there are 11 orders of magnitude from the bottom up. Clearly, this shows the impossibility of simulating the full rangeof scales explicit in a single simulation. However, for a given scale, upper and lower influences must be considered, and ultimately, modelled in one way or another.

Figure 2. Size disparities among the scales in the biomedical realm.

The increasing importance of multiscale modelling in many domains of science and engineering is clearly demonstrated. Therefore, we must anticipate that multiscale simulations will become an ever-more important form of scientific application on high-end computing resources, necessitating the development of sustainable and reusable solutions, often in the form of semi-automated workflows or pipelines of individual algorithms, for such applications. That is, we expect to need generic algorithms for multiscale computing.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

We therefore require innovative new ways of computing to face the challenges posed both by multiscale modelling and simulation and by the emerging high-end computing ecosystem. This will contribute to our ability to solve multiscale problems and, as we will argue, can also offer an avenue for new ways to efficiently exploit exascale resources. Multiscale computing could face these challenges by deploying its various single-scale components across heterogeneous architectures of exascale resources, mapped to produce optimal performance and designed to bridge both temporal and spatial scales [2, 21-23]. Therefore, we should embark upon initiatives to efficiently deploy multiscale codes on today's and future high-performance computers, among them and, most notably, HPC-cloud environments. Thereby, we need to establish a new and more effective paradigm for exploiting current and emerging computing resources.

In this respect, the European Centers of Excellence (CoEs) represent a genuine milestone for they have been created to fill this gap. Particularly, in CompBioMed we explore, study and discuss generic multiscale computing on emerging exascale high-performing computing (HPC) environments for biomedicine and biomedical research. Briefly, thanks to our expertise, diversity and integration, the basic action of CompBioMed is to map complex multi-scale simulations to large-scale computer architectures in the most efficient way. The CoEs address not only the technical side with a strong focus on the road to exascale, but also the "as-a-service" character that they must have. All our research impact, as reported in this document, is directed to such objectives.

5.2. CompBioMed action

As described in the DoAand because of the innate multiscale character of computational biomedicine, CompBioMed's WP2 objective is to establish layered application pipelines of simulation and data analysis software, with the goal of mapping them to large-scale computational resources in an efficient way. Due to the different levels of development of the codes at the beginning of the project, we followed a "Fast/Deep Track" strategy, which was streamlined during the project. We regard this strategy as a work paradigm for how to exploit large-scale computer infrastructures in a rapid and efficient way for every realm dominated by multiscale problems. Let us briefly recall the "Fast/Deep Track" concepts.

The **Fast Track (FT)** builds on the existing capabilities of the partners in the use of HPC for modelling and simulation. The FT is focused on HPC-based methods that already handle multi-physics and multi-scale features to produce integrated high fidelity personalised human models. Through the FT we were able to study very rapidly a set of concrete problems with relatively minor adaptations to the simulation codes of the work package developers. The FT has allowed us to address key biomedical challenges from the start of the project using the existing e-infrastructure, focusing firstly on the "low-hanging fruit".

The **Deep Track (DT)** addresses additional medium-term challenges. The DT extended the capabilities available from the FT, providing new capabilities often dictated by end-users requirements as reported in D6.6, addressing more complex requirements that

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

require a deeper and longer period of development. What is studied thanks to the Fast Track was recycled and projected into new areas through the Deep Track. On many cases, the standardised procedures, interfaces and integrative tools designed in the FT extended to meet the needs of different categories of end-users, including clinicians, researchers and industrial users, helping to compose pipelines which can then be executed on HPC and other computing resources. A very interesting example of this is the porting process of some of our tools from supercomputing facilities to cloud infrastructures.

Thanks to this idea, we firstly studied the different phases when developing a multiscale model and simulating it on available computing infrastructure, and analyse where in our expert view we could, and possibly should, continue to further develop generic frameworks and software tools to facilitate multiscale computing. Next, we focused on simulating multiscale models on high-end computing resources, which we call High Performance Multiscale Computing (HPMC), in the face of emerging exascale performance levels. We argue that the problems at hand are so complex that strong and weak parallel scaling of monolithic applications often may reach its limits at the exascale and, therefore, we need to invoke what we would call *multi-scaling* [1]. Note that although our analysis is driven by the needs of modelling multiscale systems, our arguments with respect to the scalability challenge for multiscale systems to the exascale also point to the necessity to consider new approaches to increase concurrency within complex (single-scale) models through new algorithms and corresponding implementations.

In a multiscale simulation, each relevant scale needs its own type of solver and strategy. Accordingly, we can define a multiscale model as a collection of coupled single scale models (loosely defined based on the dominant properties that can be computed reliably with a dedicated, so-called “monolithic” solver). Following the strategy proposed by CompBioMed, one can then identify generic *multiscale computing patterns* (MCPs) [3] that dictate the scope for novel multiscale algorithms at the exascale.

The importance of identifying such MCPs and acting accordingly is clearly deduced from papers such as [1-3] and also in the previous technical report of WP2 [24]. Exascale computing poses a number of key challenges that application developers cannot ignore: scheduling and robustness of algorithms and their implementation on millions of processors, data storage and I/O for extreme parallelism, fault tolerance, reducing energy consumption, diverse data formats and quaility, etc. Even the disparity in backgrounds, trainings or computational resources availability of the stakeholders are a hindrance to further development. For these reasons, an incremental approach that attempts to scale up monolithic solutions from a given level of computing power (i.e. petascale) will not be successful at the following (i.e. exascale), requiring a global analysis and subsequent action. Novel algorithms and simulation strategies are needed across the software stack, bridging between the applications and the hardware environments. These algorithms need to be designed specifically to address these exascale challenges in order to guarantee efficiency and resilience. We believe that,

*“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **675451**“*

drawing on the concept of generic MCPs, we can realise a separation of concerns, where the challenges stated above can be resolved to a large extent on the level of the MCPs, while the multiscale application developers can focus on composing their multiscale simulations. This would then lead to much shorter development cycles for multiscale simulations and much more reliable multiscale computing on exascale machines.

From the purely computational viewpoint, there are also strong computational considerations that dictate a need to shift the paradigm for usage of high performance computers from the conventional promotion of monolithic codes which scale to the full production partition of these computers, to much more flexible computing patterns. To clarify this further, computational scientists including ourselves have worked out numerous effective ways in which to perform spatial domain decomposition. However, petascale and future exascale machines can only reach these performance levels by aggregating a large number of cores whose individual clock speeds are no longer increasing. As a result these high performance computers are becoming "fatter", not faster and speed-up is only achievable by efficient parallelism over all the cores. But because the parallelism is usually applied to the spatial domain, we are increasingly simulating larger slabs of matter, applying weak scaling by using more particles, a higher grid resolution or more finite elements. Yet often it is the temporal behaviour that one is really interested in, and that behaviour is not extended by adopting larger computers of this nature, or by making the problem physically larger.

The smaller the space you want to solve for a given problem, the smaller the time scales you will need to capture: while a hurricane devastates a region in a few days, a leaf falls from a tree in a few seconds. Should we use the same simulation strategy for the leaf and the hurricane? Since the scientific problems of interest usually have timescales which scale as a nonlinear function of the volume of the system under investigation, each temporal update requires more wall clock time for larger physical problems. This is in fact a recipe for disaster: *we are not getting closer to studying large space and long time behaviour with monolithic codes.* To be sure, accelerators (such as GPUs) and special purpose architectures [25] can speed up many floating point calculations in particular cases such as molecular dynamics, often by a factor between one and ten, but this is not sufficient to bridge the vast timescales of concern that range from femtoseconds to seconds, hours and years; nor indeed to quantify the uncertainty in today's still all to prevalent "one-off" simulations.

But this is not all. Problems are not limited to simulation algorithmics and their mapping to architectures. Large differences on resources availability and accessibility, and its connection for different scales, could strongly impair multiscale modelling. For instance, uncontrolled spreading of patients' clinical data very frequently causes concerns against moving it from healthcare institutions. Then, feeding patient-specific data on a simulation pipeline is, in general, much more than a complex technical data, because doctors don't want to move sensitive patient's data from hospitals unless complex (and sometimes uncertain) procedures are followed.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Another problem that goes beyond their technical burden is the different access policies of supercomputing centers and cloud providers. On the one hand, supercomputing centers are almost purely devoted to research runs in which sustained computational efficiency and equal load balance is required from the users (high throughput computing is almost banned there), researchers' accounts are created and resources usage granted under stringent conditions (production runs are only allowed after a committee authorize them), inter-process communications to and from the facilities could be strongly restricted (for security reasons, which makes almost impossible to integrate a supercomputing center in a cloud infrastructure), facilities are in general homogeneous in purpose (when prepared for intense computing they are not for data bases and viceversa) or in architecture (for instance GPU clusters' admins ban CPU-only jobs), quality of service goes against business (whether you pay or not, your runs go always to a general batch system, making very difficult to foresee when you will have your job done), etc. On the other hand, cloud infrastructures, which can solve most of these problems, have been until very recently extremely weak on computational power with respect to supercomputing centers, an issue which is progressively being improved (today most of the commercial cloud providers can make available powerful and efficient instances up to 1000 cores). However, pricing policies of cloud providers are still far from attractive for the kind of problems we are dealing with: computationally intensive, large quantities of data to be transferred and stored, fast interconnects between heterogeneous resources, etc., because providers charge separately for power, cycles, networks and storage. This result is that a relatively complex workflow with all these features can become extremely expensive.

What *is* needed are more innovative ways of bridging the gap. Multiscale computing, as proposed by CompBioMed partners, is progressively able to do this by deploying its various single scale component parts across such heterogeneous ecosystems, mapped so as to produce optimal performance and designed to bridge both time and space scales and, last but not least, being capable of generating a real business model. Thus, we have embarked upon a programme to efficiently deploy componentised multiscale codes on today's and future high performance computers and, thereby, to establish a new and more effective paradigm for exploiting HPC resources, whether they are located in healthcare institutions, supercomputing centres, universities or cloud providers. Computational biomedicine is the ideal melting pot where these problems can be addressed and solved, adding value to every part of the chain.

6. The WP2 organization and strategy: Exemplar Research areas

In CompBioMed we have chosen three biomedical application domains which are highly representative of the whole realm. We call them Exemplar Research. According to the Description of Work, they are the Cardiovascular, Molecularly-based Medicine and Neuro-musculoskeletal Exemplar Research areas. Each Exemplar Research was defined as a task, linked to other smaller six ones which were transversal to all the WP.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

It is important to remark the connections of this work package with others. **WP2 strongly interacts with WP5 Resource and Infrastructure Support**, a WP that is concerned with porting, deployment, and optimisation of the CompBioMed codes for execution on supercomputing resources, and **with WP6 Empowering Biomedical Applications**, which facilitates the development and execution of complex workflows and the interaction with end-users.

The following sections summarise the work carried out within the three exemplar research areas mentioned above, with more details provided for each application together with their related publications reported in Appendix.

6.1. Cardiovascular Exemplar Research

According to the DoA cardiovascular disease accounts for half of sudden deaths in Europe; improvements in patient risk stratification and prediction of clinical intervention are both urgent and challenging. In this area we will consider two critically important disease areas: firstly, cardiovascular diseases having a direct effect on the function of the heart itself, be it on the electrophysiology, mechanics or blood flow (and, ultimately, the combination of all three); and secondly, disorders in the arteries, be it aneurisms in the abdominal aorta or in intracranial arteries, or stenosis in carotid or coronary arteries. These two areas cover the key computational challenges in multiscale modelling of cardiovascular disease (coupling of both loosely and tightly coupled physical processes, reliance on very high-level image processing techniques for geometry extraction and reconstruction, very high computational demand) and together make the ideal test bed for the development of a generic computational framework for biomedical applications requiring multi-scale, multi-physics simulations.

6.1.1. Univesity of Geneva: "Digital Blood" within Palabos

The focus of the Scientific and Parallel Computing Group at the University of Geneva is on high fidelity blood flow simulations. More in detail, we develop a high-performance computational framework for the simulation of fully resolved whole blood (**Digital Blood in Massively Parallel CPU/GPU Systems**) and additionally, we work on continuum models for flow diverting stents in 3d patient-specific intracranial aneurysms.

Digital Blood in Massively Parallel CPU/GPU Systems We propose a novel high-performance computational framework for the simulation of fully resolved whole blood flow. The framework models blood constituents like red blood cells (RBCs) and platelets individually, including their detailed non-linear elastic properties and the complex interactions among them. These kinds of simulation are particularly challenging because the large number of blood cells (up to billions) stand in contrast with the major computational requirement of individual constituents. While classical approaches address this challenge through simplified structural modelling of the deformable bodies (e.g., through mass-spring systems), the present framework guarantees accurate physics, desirable numerical properties through a fully featured FEM model and computational efficiency at the same order as more simplified state-of-the-art models.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

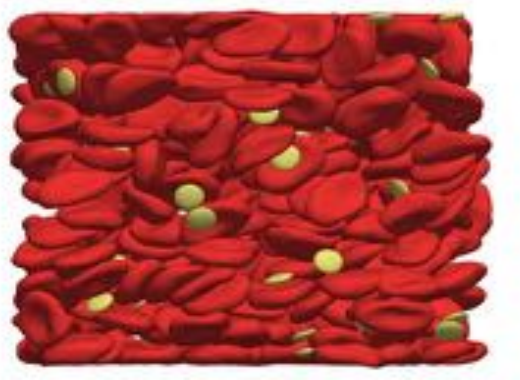

Figure 3. Shear flow with healthy RBCs (in red) and platelets (in yellow) in a domain of 50 μm^3 at 35% hematocrit.

Continuum model for flow diverting stents in 3D patient-specific simulation of intracranial aneurysms

An aneurysm is a weak spot on a blood vessel wall that causes a balloon-like bulging. Aneurysms tend to increase in size, and can be at risk of rupturing, leading to internal bleeding, with severe consequences. In particular, rupture of intracranial aneurysms is an event with a high mortality rate, and about one third of the survivors suffer from permanent neurological or cognitive deficits. For diagnosed aneurysms that are at risk of rupturing, different forms of treatment exist, one of which includes the insertion of a prosthesis into the artery, called a flow diverting stent. It has the shape of a tube, with a surface consisting of a very fine mesh of woven wires, and is inserted to cover the neck of the aneurysm. The role of this device is to divert the main bloodstream from the aneurysm to the artery, while avoiding surgically clipping off the aneurysm from blood supply altogether. The modification in the bloodstream pattern achieved by this procedure is capable of encouraging a blood clotting reaction in the aneurysm, which in its turn fills the aneurysm dome and prevents its rupture.

Computer simulation can be of use in many ways in this field of medical science. For example, the flow mechanics of the blood, and biological processes inside the blood, are simulated to promote a fundamental understanding of the factors involved in the blood clotting process of an aneurysm and propose new means of medical treatment. On the other hand, computer simulation is also a useful tool for day-to-day medical decision making, as it can be applied, for example, whether a given patient aneurysm could be successfully treated through a stent insertion procedure, and which stent model to use for optimal results. The idea is to use medical imagery to construct a virtual model for the artery and the aneurysm of a patient, and introduce, virtually, different flow diverting stents into this artery. Numerical models known under the name of Computational Fluid Dynamics are then used to simulate the mechanical properties of blood flow in the artery and test the ability of the flow diverter to encourage blood clotting.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

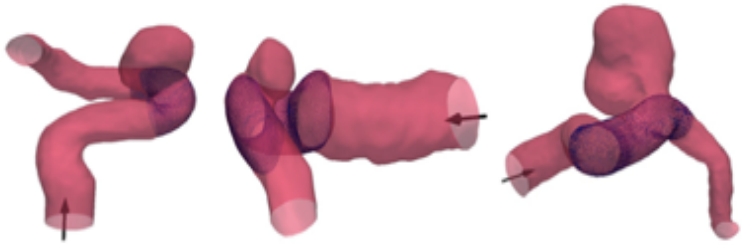

Figure 4. Blood vessel, aneurysm, and stent for the three investigated patient cases.

6.1.2. University of Amsterdam: HemoCell, the microscopic cellular library for supercomputers

The developments were realised in three distinct areas. First and foremost, the advancements and optimisations were carried out on the open-source HemoCell (High pErformance MicrOscopic CELlular Library) simulation code. These allow larger domain sizes and longer simulated physical time-scales. Furthermore, improvements were made to the robustness of the mechanical model, i.e. the mechanical description of flowing cells, using uncertainty quantification (UQ). In tandem, similar UQ was applied to the in-stent restenosis model. Finally, these improvements made new applications possible. The cell level trafficking was explored in both healthy cases, and in cases with rigidified red blood cells, which serve as a model for diabetic disease.

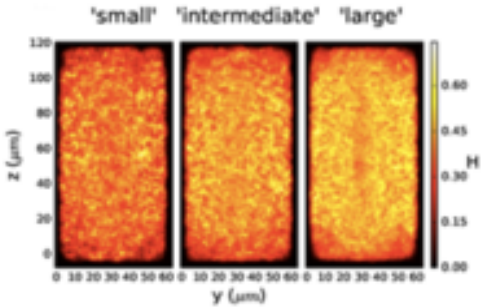

Figure 5. Haematocrit distribution for the channel flow case with different numbers of Red Blood Cells (N), average haematocrit H = 38%. 'small' (left), 'intermediate' (middle), and 'large' (right) systems. Extracted from [26].

6.1.3. University College London: HemeLB for 3D flows in large sparse geometries

Using HemeLB, a highly optimized lattice-Boltzmann solver for haemodynamic flow in large, sparse 3D geometries, our research during the course of the CompBioMed project has focused on two strands - one for automated, validated and patient specific simulation, and the other on the computational optimization work required to best exploit coming exascale infrastructure.

Magnetic Drug Targeting (MDT) capabilities were implemented in HemeLB, allowing the dynamics of paramagnetic iron-oxide particles (in practice often used as a drug delivery method) to be observed in a geometry obtained from an MRI scan of a patient's brain [27]. This model allows exploration of the effect of changes in particle size, magnet

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

strength and placement, and patient physiology (heart rate, etc) on the delivery to a given target site (e.g. a tumour site) for a patient specific vascular system. In this multiscale model, the inlet velocity profiles for the 3D region are generated using a 1D Navier-Stokes solver representing the rest of the human body.

Figure 6. A small section of the simulated Circle of Willis geometry, showing the paramagnetic drug particles passing through a (pink) target region under the influence of a (blue) magnetic dipole.

In another study [28] we collaborated with the National Hospital for Neurology and Neurosurgery (NHNN) in London, to obtain Transcranial Doppler (TCD) measurements of blood velocity in the Middle Cerebral Artery (MCA) of a stroke patient, along with CT scan data allowing the 3D vasculature (of that same patient) to be generated for use in HemeLB simulations. A simple validation study was then carried out by comparing the velocity profiles from simulations against those measured by TCD at multiple points along the MCA. We also considered the sensitivity of the simulations to changes in rheology model (Newtonian vs shear-thinning) and in mesh resolution.

6.1.4. Lifetec: AngioSupport for coronary artery disease

Cardiac teams in the larger hospitals daily discuss the treatment of multiple patients with coronary artery disease (CAD). These patients have one or multiple severe occlusions in the coronary arteries which are complicated cases and requires the expertise of the cardiac team. For each patient a treatment plan is defined, typically consisting of coronary artery bypass graft (CABG) surgery or percutaneous coronary intervention (PCI). The decision between these treatments is currently based on studying coronary angiograms and the experience of the cardiac team. However, in case of multiple occlusions, diffuse coronary disease or complicated vasculature, choices in the position, length or diameter for a CABG or PCI is challenging.

To help the cardiac team in this process, LifeTec Group talked with Pim Tonino**, Intervention Cardiologist** at Eindhoven Catharina Hospital, about a possible numerical model that can assist in this decision making. Together with Frans van de Vosse, Professor at the Biomedical Engineering Department of Eindhoven university of technology, LifeTec Group started the development of a clinical tool that could assist the cardiac team in treatment planning for each patient. Therefore, AngioSupport has been

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

developed; an interactive tool to predict the outcome of CABG or PCI to support **clinical decision making** of coronary interventions.

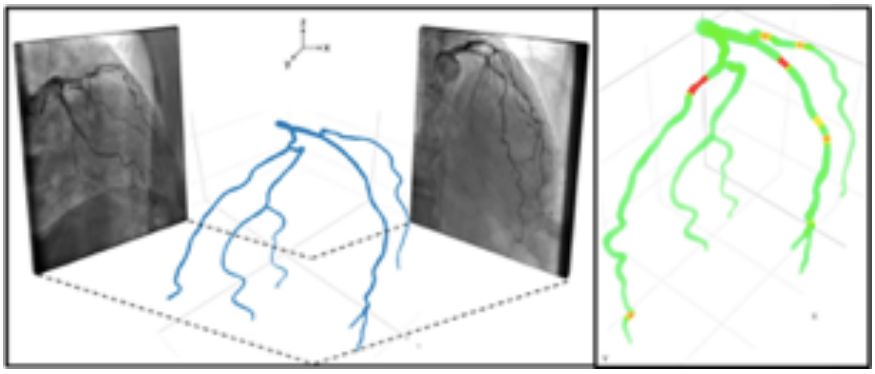

Figure 7: Obtaining the 3D vessels by using the 2D images created during coronary angiography.

6.1.5. University of Sheffield: OpenBF for vascular networks

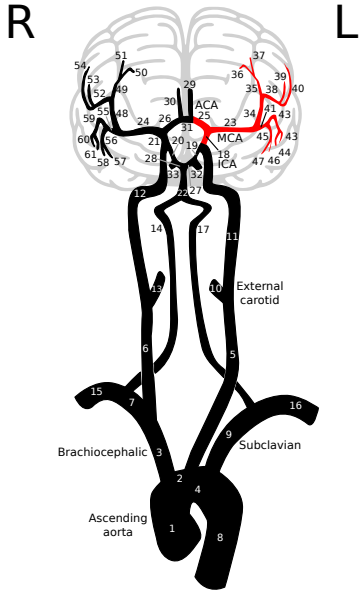

Figure 8. Diagram of the OpenBF cerebral vascular network model.

OpenBF is a 1D hemodynamics network model developed at the University of Sheffield. Cerebral vasospasm (CVS) is a life-threatening condition that occurs in a large proportion of those affected by subarachnoid haemorrhage and stroke. CVS manifests itself as the progressive narrowing of intracranial arteries. It is usually diagnosed using Doppler ultrasound, which quantifies blood velocity changes in the affected vessels, but has low sensitivity when CVS affects the peripheral vasculature. In a recently published study we identified alternative and more effective biomarkers than the ones currebntly used and that could be used to diagnose CVS [29]. For this we used a verified and validated 1D modelling approach, openBF, to describe the properties of pulse waves that propagate through the cardiovascular system, which allowed the effects of different types of

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

vasospasm on waveforms to be characterised at several locations within a simulated cerebral network.

6.1.6. Barcelona Supercomputing Center and University of Oxford: Alya Cardiac Computational Model

Human-based computer models and simulations are a fundamental asset of biomedical research. They augment experimental and clinical research through enabling detailed mechanistic and systematic investigations. Owing to a large body of research across biomedicine, their credibility has expanded beyond academia, with vigorous activity also in regulatory and industrial settings. Thus, human *in silico* trials are now becoming a central paradigm, for example, in the development of medical therapies [30].

Human cardiac physiology is one of the most advanced areas in physiological modelling and simulation. Current human models include detailed information on the ionic processes underlying the action potential such as the sodium, potassium and calcium ionic currents, exchangers such as the Na/Ca exchanger and pumps such as the Na/K pump. They also include representation of the excitation-contraction coupling system, which modulates the calcium transient and, in turn, myocyte contractility. Human cardiac models are also multiscale, both spatially and temporally, and integrate information across the subcellular, cellular, tissue, and organ levels [31].

In this report we showcase the human multiscale models we developed through the integration of multimodality datasets, including: ionic current measurements; action potential and calcium transient recording; active force measurements; magnetic resonance and computed tomography images; electrocardiograms. Human data were used at multiple stages of model development, for calibration and also to perform independent validations at different scales.

Through an intense collaboration BSC and the University of Oxford have developed a cardiac computational model, based on BSC's Alya multi-physics / multi-scale parallel simulation code, which has been used at different scales attacking complex problems, from single cell models up to medical devices clinical trials.

Single Cell Models

Human in Silico Drug Trials to Predict Risk of Torsade de Pointes

During the first year of the project, we demonstrated the predictive power of populations of human ventricular AP models for prediction of drug-induced Torsade de Pointes (TdP) risk based on repolarisation abnormalities occurrence. We were able to achieve a prediction accuracy of 89% for a set of 62 reference compounds. The results of these *in silico* drug trials were published [32], and also led to the award of the International 3Rs prize in 2017. More recently, we performed a similar study, including an additional biomarker: the electro-mechanical window (EMw), defined as the delay between the duration of electrical and mechanical systole, which has been suggested as a promising biomarker to predict clinical risk of Torsade de Pointes (TdP) arrhythmia in several pre-clinical animal models [33-36]. Our single cell surrogate of the *in vivo* EMw

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

was able to predict TdP risk for a dataset of 40 compounds with 90% accuracy, confirming the potential of drug-induced EMw shortening as a biomarker for pro-arrhythmic risk. The results of these *in silico* drug trials are currently under review for publication [37].

In Silico Predictions of Drug-induced Changes in Contractility
These are preliminary results obtained by testing the effect of two reference compounds on our single cell model of cardiac electro-mechanics, to explore drug-induced changes in active tension: dofetilide and verapamil.

Multiscale Models
3D Electro-mechanical Simulations of the Human Heart
The multiscale human cardiac electro-mechanical model with ellipsoidal geometry is able to simulate all the four phases of the cardiac cycle. We use such a model to test pressure - volume functions which help to understand the cycle. We also tested boundary conditions for the pericardium.

Applications towards clinical translation
A common characteristic of the existing human cardiac models is that personalised geometries usually come from in-vivo imaging and the majority of computational meshes consider simplified ventricular geometries with smoothed endocardial (internal) surfaces, due to a lack of high resolution, fast and safe in-vivo imaging techniques. Acquiring human high-resolution images would mean for the patient to undergo long, expensive and impractical scans, in the case of magnetic resonance images (MRI), or could present a risk for the patient's health, in the case of computed tomography (CT), since this process involves a considerable amount of radiation. Smoothed ventricular surfaces are indeed considered by the majority of existing human heart computational models, both when modelling blood flow dynamics and electrophysiology.

However the endocardial wall of human (and other mammals species) cardiac chambers is not smooth at all; it is instead characterised by endocardial sub-structures such as papillary muscles (PMs), trabeculations and false tendons (FTs). Additionally, fundamental anatomical gender differences can be found in cardiac sub-structural heart configuration as female hearts present less amount of FTs [38].

Through collaborations with the University of Minnesota we have created highly detailed human heart models from *ex vivo* high-resolution MRI data, to study the role of cardiac sub-structures and gender phenotype in human cardiac physiology, through computational fluid dynamics (CFD) and electrophysiological high performance computing (HPC) simulations.

Path to Code Validation
Validation means that the simulation software is correctly reproducing the multiple physics of the question of interest for a determined context of use. This not only requires correctly solving the programmed model, but that the model effectively models the

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

physics. To do so, experimental data is required to compare *ex vivo*, *in vitro* and *in vivo* data against *in silico* results. This stage requires a detailed description of the variables of the physical problem which are, in most of the cases, very difficult to obtain with a high accuracy.

As part of the collaboration with the Centro Nacional de Investigaciones Cardiovasculares (CNIC), the pathway to the validation of the cardiac model against experimental is being underway, as published in one of BSC's PhD thesis, supervised by Mariano Vázquez from BSC.

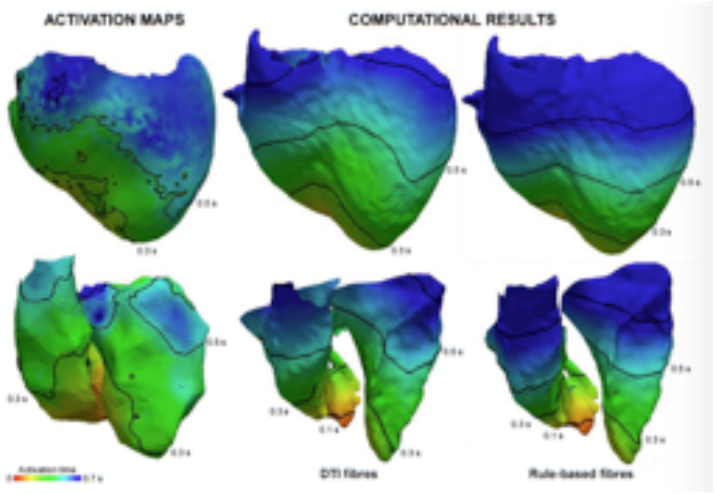

Figure 9. Activation maps including isochrones of the epicardium and the endocardium from experimental measurements and simulation data [39].

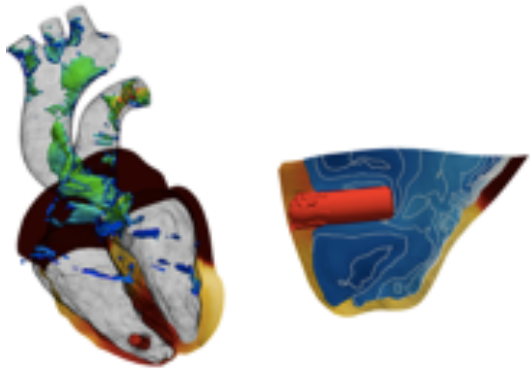

Figure 10. A leadless pacemaker implanted in the virtual heart. Q-criterion and velocity fields can be seen distorted in the soroundings of the implantation spot [40].

Path to medical device testing

During the last three years, through different collaborations, the area of device testing has been exploited in BSC's research group involved in CompBioMed project. Especially thanks to collaborations with the associate partner Medtronic, our model of the heart was used to study devices related with heart diseases such as pacemakers and stents.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

First we study widely known treatments for common pathologies, although not completely understood. These simulations can help to better understand the pathology and the treatment, or at least optimise the device set-up to maximise the performance. The stages of verification and validation are required to confidently translate these results to clinical applications.

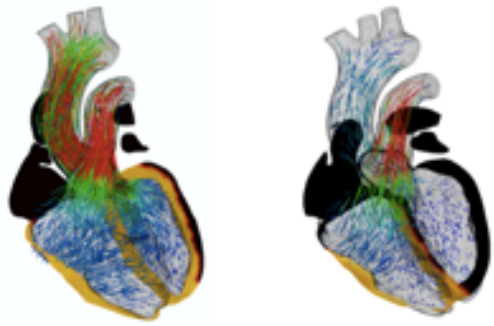

Figure 11. The left side shows a heart beating in normal conditions. The right side shows a heart beating under left bundle branch block. The elecrical dyssynchrony is easily seen. Also the reduction in the velocities on the fluid domain that will lead to a drop in the ejection fraction [40].

6.2. Molecularly-based Medicine Exemplar Research

Quoting the DoW, many areas of medicine rely on a molecular understanding of the underlying human biology. Indeed, the pharmaceutical industry's success has been largely underpinned by such knowledge. Its business model is being seriously challenged today, with rapidly increasing sums of money invested in an attempt to maintain an acceptable pipeline of patentable products. However, the central premise of drug production, namely that one can hope to produce "blockbuster" one-size-fits-all drugs for the entire global population, has proven impractical; most drugs that have been developed for specific disease treatments only applying to subsets of the population. Instead, the industry now needs to think in terms of *multiple* drugs which address any specific disease case, using stratification (based primarily on gene sequencing) as a first step along the way to ultimate personalised drug selection and treatment. Due to advances in gene sequencing, we now have the basic patient specific data to hand that allow us to begin to develop personalised drug treatment.

6.2.1. Universitat Pompeu Fabra: Machine Learning and Molecular Dynamics

Pharmaceutical industry is facing an unprecedented challenge nowadays. Introducing a new drug into the market involves a 15-year-long process and billions of dollars in investment, yet the success rate is pretty low. The probability that a candidate drug in Phase I clinical trials ends up being approved is around 7% [41]. It is in everybody's interest to keep drug discovery as a sustainable model, and therefore it is necessary to reduce its overall cost, speed up the discovery process and improve the success rates.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Figure 12. General scheme of the entire drug discovery process and all the stages required to release new medicines into the market.

Our primary objective within the CompBiomed project has been to develop novel computational methods for the early stages in the drug discovery pipeline in order to accelerate the obtention of drug candidates and reduce the experimental workload and its associated costs. We advance towards the next generation of drug discovery, which relies on computational predictive models that are able to test millions of compounds *in silico*, giving accurate and precise results and reducing the amount of experimental tests needed on the design process. Furthermore, having computational methods that are powerful enough opens up the possibility to drastically reduce (and, on the long term, even remove) animal experiments, a fundamental requirement for a more sustainable and ethically responsible drug discovery process.

With the recent advances in artificial intelligence and deep learning we can leverage the data and use it for novel predictions. Particularly, deep learning can be applied to extract complex patterns from simple representations. In our work, we leverage deep learning methods to extract patterns from three-dimensional representations of molecules and proteins. We developed models inspired by computer vision architectures, where both protein and the ligand are divided in a three dimensional grid with features representing different atomic properties, such as hydrophobicity or aromaticity.

6.2.2. University College London: Supercomputers and binding affinities

Drug development is a lengthy, complex, and costly process (it is estimated that and average of ~€2.2 billion is required to get a drug into the clinic [42]) and involves a high degree of uncertainty that any given candidate will actually succeed. It is increasingly recognised that this is compounded by the variation in response between patients, implying that we can no longer hope to produce "blockbuster" one-size-fits-all drugs for the entire global population [43]. Consequently, new approaches are required that facilitate better targeted treatments for subsets of patients. Our goal is to support this endeavour by developing simulation techniques that allow us to understand how drugs interact with their target proteins and how genetic variation can affect this.

The binding affinity calculator (BAC) software developed within the molecular medicine strand of CompBioMed is at the heart of a research programme which aims to influence both industrial and clinical workflows. The common approach that under pins these goals is the generation of computational protocols which provide reproducible binding

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

free energy estimates for ligand binding from molecular simulation. Our research has sought to address this based upon a theoretical understanding of the utility of ensemble simulations to provide efficient sampling and meaningful uncertainty quantification [44]. This has led us to develop a suite of computational protocols which we call ESMACS (enhanced sampling of molecular dynamics with approximation of continuum solvent) and TIES (thermodynamic integration with enhanced sampling) [45]. The former based on the use of computationally inexpensive end point calculations and the latter more expensive but accurate "alchemical" binding free energy calculations.

On a practical level we have applied this insight to looking into a diverse range of drug discovery targets with the aim of identifying the reasons for differing levels of success for existing methods when using datasets involving different proteins or ligands from different regions of chemical space. Alongside our work focussed on drug discovery problems we have investigated the influence of mutations on drug binding, an issue of direct relevance to the effectiveness of therapies in individual patients.

6.2.3. Janssen: Molecular Dynamics for drug discovery programs

A 'people' definition of industrial computational chemistry includes a group of molecular modelling experts who use computational techniques, mostly to help molecular design in collaboration with medicinal chemists. Whilst other areas of application exist (such as in early target validation) this project will focus on methodologies that can make a significant step forward in the quality of molecular design. It can be argued that the current toolbox of an industrial computational chemist, despite incremental change, has not seen any fundamental improvement for over 10 years. Clearly challenge and investment are needed to learn if new methodologies can provide impact. In this project we studied new methodologies in areas of molecular dynamics calculations, and more accurate binding energy predictions.

Traditional industrial computational chemistry is highly dependent on a small selection of approaches such as virtual screening, molecular docking, ligand similarity etc. Structure-based drug design, where typically an X-ray crystal structure of the target is available, permits docking to help molecular design. Pose prediction with docking, that is correctly placing the ligand in the right orientation in the binding site, is typically considered an achievable task. Docking can also show virtual screening enrichment, which means separating a structurally diverse set (such as a random high throughput screening collection) of actives from inactives. However, it is widely recognized, that for a congeneric series of structural analogues, such as the case in a drug discovery lead optimization program, docking methodologies are unable to differentiate or rank highly active from inactive molecules. Hence, computational structure-based design remains largely qualitative and based on visual assessment and discussion and prioritization of results within project teams. This leads to various limitations, such as the number of molecules which can plausibly be docked and reliably assessed in this labour-intensive way.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Various fundamental computational methodologies have existed for a long time but remain largely unused in an industrial computational chemistry setting. Molecular Dynamics (MD) is one of these. By using Newton's classical equations of motion, computational simulations study the conformational changes of a protein (for instance) with time. Whilst these methods have existed for many decades, only now is it becoming feasible to consider running MD simulations for time scales that are relevant for issues of importance for drug discovery. At the same time, computational chemistry is undergoing significant changes due to access and porting of algorithms to Graphics Processing Unit (GPU) hardware. GPU's provide thousands of cores and offer a cheap highly parallel architecture which is efficient for computational approaches such as MD. During the past several years major steps have also been made in the practical feasibility of calculating free energies of binding of small molecules to proteins from the 3D structure of the protein-ligand complex. Important factors include improved molecular force fields, better conformational sampling, and faster hardware. Free energies of binding can now be calculated to within 1 kcal/mol accuracy, which is much better than commonly used approaches like molecular docking. This level of accuracy has the potential to radically increase the impact of computational design to drug discovery. Development of methods and best practices for free-energy calculations, as carried out in this project, will enable more effective computational design of drug candidates for globular proteins and membrane-bound targets.

Janssen's primary interests in the CompBioMed project are in developing and using advanced molecular simulation methods to optimize lead compounds in discovery programs. Such methods, if proven robust and accurate could have a profound impact on the way drug discovery is performed. They would permit reliable computational triaging of very close analogue molecules greatly improving efficiency. Also, this would lead to high-confidence design of synthetically more challenging molecules leading to better drugs in new chemical space. Also, we envisage the accurate prediction of compound binding for targets that have mutated residues. This latter application can be of value in diagnostics, by predicting the best possible compound for a patient clinically (personalised medicine), but is also of use in discovery, where mutated targets occur regularly in antibacterials, antivirals, and oncology compounds.

We will summarize the CompBioMed impact for Janssen by describing the main collaborations within the project.

Collaboration Janssen - UCL

Janssen collaborated with UCL on calculation of free energies of binding on public and on Janssen internal compound sets (targets BRD4, LDHA and PDE2). A manuscript was co-written and accepted for publication on the BRD4 application. A second manuscript is under review describing the LDHA application. Both cases have led to learnings about the suitable application of MMPBSA, so called ESMACS approach, for the calculation of binding free energies.

Collaboration Janssen – UPF/Acellera

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Janssen collaborated with UPF/Acellera to test machine learning methods derived from protein ligand binding datasets and used them to predict relative binding free energies.

Janssen internal research for CompBioMed
Within Janssen we evaluated the use of GROMACS for Free Energy Perturbation. We streamlined the application of FEP with GROMACS and ran calculations at SURFsara.

6.2.4. EVOTEC: G-protein activated structures modelling for large-scale computers

Evotec (UK) Ltd, as a leading industrial application partner, is responsible for four key objectives: adaptation of hierarchical G-Protein Coupled Receptors (GPCR) modelling protocol (HGMP) to HPC platform, developing of the new HGMP-HPC based tools / plugins that require high scale calculations, testing and application of HGMP-HPC integrated technology in real drug discovery cases within the CoE and to make it available to third parties seeking assistance from the CoE and/or from Evotec, and dissemination of the results of this work to our partners in academia and in pharma & biotech companies in order to stimulate follow-on research. Evotec has also published the outcome of this work in peer-reviewed journals and at scientific conferences. Evotec (UK) Ltd (Dr Alexander Heifetz) has established a close collaboration with UCL (group of Prof Andrea Townsend-Nicholson) [46]. In the framework of this collaboration, they developed computational methodologies for structural exploration and tools for drug design, described as follows:

Rationalizing the receptor-ligand binding and drug-candidates' residence time [47].
Drug-target residence time, the length of time for which a small molecule stays bound to its receptor target, has increasingly become a key property for optimization in drug discovery programs. However, its *in silico* prediction has proven difficult. Here we describe a method, using atomistic ensemble-based steered molecular dynamics (SMD), to observe the dissociation of ligands from their target G protein-coupled receptor in a time scale suitable for drug discovery.

Computational prediction of GPCR oligomerization [48]
There has been a recent and prolific expansion in the number of GPCR crystal structures being solved: in both active and inactive forms and in complex with ligand, with G protein and with each other. Despite this, there is relatively little experimental information about the precise configuration of GPCR oligomers during these different biologically relevant states. While it may be possible to identify the experimental conditions necessary to crystallize a GPCR preferentially in a specific structural conformation, computational approaches afford a potentially more tractable means of describing the probability of formation of receptor dimers and higher order oligomers. Ensemble-based computational methods based on structurally determined dimers, coupled with a computational workflow that uses quantum mechanical methods to analyze the chemical nature of the molecular interactions at a GPCR dimer interface, will generate the reproducible and accurate predictions needed to predict previously unidentified GPCR dimers and to inform future advances in our ability to understand and begin to precisely manipulate GPCR oligomers in biological systems.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

FMO-DFTB tool for rapid analysis of receptor-ligand interactions [49]
The reliable and precise evaluation of receptor-ligand interactions and pair-interaction energy is an essential element of rational drug design. While quantum mechanical (QM) methods have been a promising means by which to achieve this, traditional QM is not applicable for large biological systems due to its high computational cost. Here, the fragment molecular orbital (FMO) method has been used to accelerate QM calculations, and by combining FMO with the density-functional tight-binding (DFTB) method we are able to decrease computational cost 1000 times, achieving results in seconds, instead of hours. We have applied FMO-DFTB to three different GPCR-ligand systems. Our results correlate well with site directed mutagenesis data and findings presented in the published literature, demonstrating that FMO-DFTB is a rapid and accurate means of GPCR-ligand interactions.

FMO-PPi tool of inter-helical interactions of G-protein coupled receptors.
G-protein coupled receptors (GPCRs) are the largest superfamily of membrane proteins. They regulate almost every aspect of cellular activity and are key targets for drug discovery. However, the molecular forces responsible for holding together the seven helices of the GPCR bundle and ensuring receptor stability, ligand binding and activation have not been identified. Even with crystal structures in hand, the strength and chemical nature of these forces cannot be characterised by visual inspection alone and therefore, accurate and reliable computational methods must be employed.

6.3. Neuro-musculoskeletal Exemplar Research

Despite a common perception that most neuro-musculoskeletal diseases are not life threatening, around 30% of the elders who face an osteoporotic fracture of the hip joint will die of related complications within 12 months; Amyotrophic Lateral Sclerosis has a lethal outcome usually within a few years; severe forms of *Osteogenesis Imperfecta* drastically reduce the life expectation of a child, etc. If we consider the quality of life, or the burden of disease that is a combination of quality and quantity of life, of the 10 top causes in Europe, four are neuro-musculoskeletal. The socioeconomic impact of lower back pain, arthritis, and osteoporosis, to name a few common conditions, is larger than that associated to any family of diseases.

Because the primary function of the neuro-musculoskeletal system is mechanical in nature, it is not surprising that computational biomechanics has proved very effective in this field. In many of these applications, the models involved are of considerable complexity, with several nested levels strongly coupled altogether: molecules, cells, tissues, organs, organ systems, living beings and ecosystems. The fact that the scales are strongly coupled means that, when simulating a system, neglecting a scale can lead to inaccurate or plainly incorrect results. One feature of human physiology is its great variability among individuals. Such variability enforces the need for person-specific simulation models.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

6.3.1. University of Sheffield: from tomography to bone simulations

In Sheffield, the **Computer Tomography to Strength (CT2S)** has been successfully rolled out and tested on ShARC (Sheffield Tier-3), and applied to more than 110 patients so far. The service uses hip CT scans to generate personalised finite element models of the femurs, and ran simulations to predict the femoral strength in order to predict the risk of osteoporotic fracture. The service has recently been linked to the Sheffield Teaching Hospital, where a clinical staff can send a request for a set of patient CT scans to be analysed with a report being returned (within 1-2 days) to detail the risk factors. The research work has also resulted in publications [50,51].

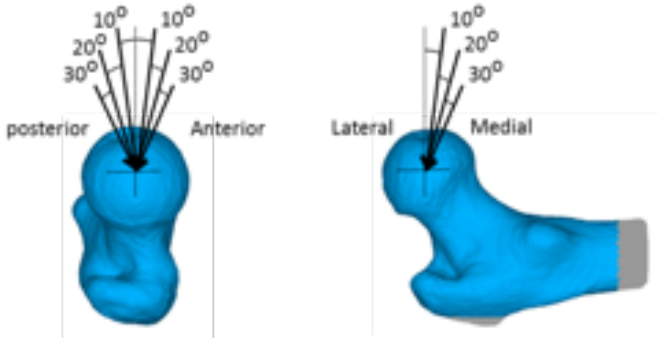

Figure 13. The range of sideways fall loading directions tested in the algorithm. Reproduced from Altai et al. (2019), Clinical Biomechanics.

The CT2S workflow has employed in an in silico study that investigates the effect of ageing on bone strength. A bone loss law has been developed that describes how local volumetric bone mineral density (vBMD) decreases as the areal bone mineral density (aBMD) at the femoral neck (FN) decreases during ageing. Combined with the CT2S workflow, this framework leads to a determination of how fall orientation specific bone strength changes due to ageing, such as following a 5% decrease in FN-aBMD. We oberve that changes in bone strength are much smaller in magnitude and variability when predicted using bone loss law than when predicted by linear regression. This highlights that bone strength loss really depends on the individual's bone shape and size and on how vBMD is distributed spatially within the bone.

The CT2S service is currently being used by the Sheffield Teaching Hospitals to process patient data and predict the risk of osteoporotic fracture. The algorithm is also used by clinicians at the Sheffield Children's Hospitals for research purposes in the application of child abuse. Other research users include the Flinders University (Australia) and the University of Wisconsin (USA).

As part of the BoneDVC service, the Sheffield-based image processing software SHIRT has been rewritten in order to make it easier to parallelise on HPC system. The new software is called **pFIRE**. It has been deployed and tested on ShARC, ARCHER, MareNostrum. The software is available to download via Github with a set of dedicated tutorials to get users started:
(https://insigneo.github.io/pFIRE/tutorial.html).

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

6.4. ELEM Biotech: a Startup Born From Within CompBioMed

After a long period of administrative issues with the government, in July 2018, ELEM Biotech (http://www.elem.bio) was consolidated in Spain. It is a startup spun off from the Barcelona Supercomputing Center, co-founded by M. Vázquez and G. Houzeaux, both researchers of the aforementioned institution. The goal of ELEM is to provide supercomputer-based simulations of the cardiovascular system for medical devices manufacturers, pharmaceutical industry, CROs and academia.

Figure 14. Landing page of ELEM Biotech.

At this moment, ELEM is in the middle of a first funding round. Due to the importance of this round, our company is at the moment "under the radar", unveiling as little as possible information about its purpose and goals.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

7. Conclusion

This document reports the impact of the Modelling and Simulation scientific advances done within CompBioMed's WP2 on the biomedical research realm. It gives just a glimpse of all the research done not thanks just to the individual effort of each partner but to the combined expertise put together by CompBioMed. This is reflected by the variety of institutions taking part as co-authors in our published papers, the great majority in prestigious international journals.

Research done in CompBioMed has covered the full range of the scales of a multi-scale problem as described in the introductory section. As stated above, The Centres of Excellence are there to explore innovative new ways of computing to face the challenges posed both by multiscale modelling and simulation and by the emerging high-end computing ecosystem. In ComBioMed we have worked on multi-scale strategies and their porting to new HPC architectures (among them notably HPC-Cloud infrastructures), specifically for the biomedical realm. We showed all the potential of such effort and how it can impact outside the bounds of the project, especially when planning to offer a service to all kind of healthcare-related stakeholders. We also demonstrate our ability to solve multiscale problems efficiently exploiting exascale resources.

Going through the six original objectives for the project, as written in the Description of Work, this WP has contributed to all of them, especially in objectives 2, 5 and 6. Briefly,

- Objective 2: *"The second objective of our Centre of Excellence therefore will be to promote innovation in the field of computational biomedical modelling and simulation. [...]"*. Our research has been radically innovative and disruptive thanks to the combination of biomedicine, computational and mathematical expertise, together with all the technical advances we studied to pave the road to exascale computing.
- Objective 5: *"The fifth objective of our Centre of Excellence will be to engage with a range of industries across the entire healthcare value chain, from healthcare providers to pharmaceutical and medical device manufacturers, as well as ISVs and HPC system providers, to further the direction, uptake and exploitation of high performance computing within commercial organisations [...]."* Our research deeply involved clinical, academic and industrial collaborators. Pharmaceutical Industry and Medical Devices Manufacturers are not just co-authors of many of our papers but they become customers or pre-customers of the solutions produced in the project. In particular and related to the HPC providers, we have successfully explored the potential of HPC-based cloud computing to boost this realm, not only to do research but also to become a fundamental part in a new business model.
- Objective 6: "*The sixth objective of our Centre will be to engage closely with medical professionals through our partner hospitals and the wider community of*

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

stakeholders, to promote (i) the tools, techniques as well as access mechanisms developed within our Centre; (ii) the wider field of computational biomedicine; (iii) the importance of computational modelling as an integral part of the decision making process within specific clinical fields." All of our research has impact on either clinical or industrial stakeholders (or both) as shown in this report. But regarding clinical, we went beyond, because we make a great effort to evangelise the clinical environment with the ultimate goal of giving access of the tools created directly to medical doctors at healthcare institutions. Of course, this is a very difficult mission, impossible to fully accomplish during the three years of its initial life, but at least we have established fantastic grounds to build things upon.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

8. Bibliography

[1]. Hoekstra A.G., Chopard B., Coster D., Portegies Zwart S., Coveney P.V. (2019). Multiscale computing for science and engineering in the era of exascale performance. Phil. Trans. R. Soc. A 377: 20180144.
[2]. Alowayyed, S., Groen, D., Coveney, P. V., & Hoekstra, A. G. (2017). Multiscale computing in the exascale era. *Journal of Computational Science*, *22*, 15-25.
[3]. Alowayyed, S., Piontek, T., Suter, J.L., Hoenen, O., Groen, D., Luk, O., Bosak, B., Kopta, P., Kurowski, K., Perks, O., Brabazon, K., Jancauskas, V., Coster, D., Coveney, P.V., Hoekstra, A.G. (2019). Patterns for High Performance Multiscale Computing, *Future Generation Computer Systems*, 91, Pages 335-346.
[4]. Hoekstra AG, Chopard B, Coveney PV. 2014 Multiscale modelling and simulation: a position paper. *Phil. Trans. R. Soc. A* **372**, 20130377. (doi:10.1098/rsta.2013.0377)
[5]. Weinan E, Engquist B, Li X, Weiqing R, Vanden-Eijnden E. 2007 Heterogeneous multiscale methods: a review. *Commun. Comput. Phys.* **2**, 367–450.
[6]. Sloot PMA, Hoekstra AG. 2010 Multi-scale modelling in computational biomedicine. *Brief Bioinform.* **11**, 142–152. (doi:10.1093/bib/bbp038)
[7]. Fish J. 2009 *Multiscale methods: bridging the scales in Science and Engineering*. Oxford, UK: Oxford University Press.
[8]. Engquist B, Lötstedt P, Runborg O. 2009 *Multiscale modeling and simulation in Science*. Berlin, Germany: Springer.
[9]. Karabasov S, Nerukh D, Hoekstra AG, Chopard B, Coveney PV. 2014 Multiscale modelling: approaches and challenges. *Phil. Trans. R. Soc. A* **372**, 20130390. (doi:10.1098/rsta.2013.0390)
[10]. Coveney PV, Boon JP, Succi S. 2016 Bridging the gaps at the physics chemistry biology interface. *Phil. Trans. R. Soc. A* **374**, 20160335. (doi:10.1098/rsta.2016.0335)
[11]. Portegies Zwart S *et al.* 2010 Simulating the universe on an intercontinental grid of supercomputers. *IEEE Comput.* **43**, 63–70. (doi:10.1109/mc.2009.419)
[12]. Portegies Zwart S, McMillan S. 2018 *Astrophysical recipes: the art of AMUSE*. Bristol, UK: IOP Publishers.
[13]. Suter JL, Groen D, Coveney PV. 2015 Chemically specific multiscale modeling of clay-polymer nanocomposites reveals intercalation dynamics, tactoid self-assembly and emergent materials properties. *Adv. Mater.* **27**, 966–984. (doi:10.1002/adma.201403361)
[14]. Silani M, Talebi H, Hamouda AM, Rabczuk T. 2016 Nonlocal damage modelling in clay/epoxy nanocomposites using a multiscale approach. *J. Comput. Sci.* **15**, 18–23. (doi:10.1016/j.jocs.2015.11.007)
[15]. Laurini E, Posocco P, Fermeglia M, Pricl S. 2016 MoDeNa nanotools: an integrated multiscale simulation workflow to predict thermophysical properties of thermoplastic polyurethanes. *J. Comput. Sci.* **15**, 24–33. (doi:10.1016/j.jocs.2015.11.006)
[16]. Bin S, Li Z. 2016 Multi-scale modeling and trans-level simulation from material meso-damage to structural failure of reinforced concrete frame structures under seismic loading. *J. Comput. Sci.* **12**, 38–50. (doi:10.1016/j.jocs.2015.11.003)
[17]. Suter JL, Coveney PV, Anderson RL, Greenwell HC, Cliffe S. 2011 Rule based design of clay-swelling inhibitors. *Energy Environ Sci.* **4**, 4572–4586. (doi:10.1039/c1ee01280k)
[18]. Falchetto GL *et al.* 2014 The European Integrated Tokamak Modelling (ITM) effort: achievements and first physics results. *Nuclear Fusion* **54**, 043018. (doi:10.1088/ 0029-5515/54/4/043018)

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

[19]. Bruzzone AG. 2015 Perspectives of modeling; applied simulation: modeling, algorithms and Simulations: advances and novel researches for problem-solving and decision- making in complex, multi-scale and multi-domain systems. *J. Comput. Sci.* **10**, 63–65. (doi:10.1016/j.jocs.2015.06.004)
[20]. Stevens B, Bony S. 2013 Climate change. What are climate models missing? *Science* 340, 1053–1054.
[21]. Borgdorff J, Falcone JL, Lorenz E, Bona-Casas C, Chopard B, Hoekstra AG. 2013 Foundations of distributed multiscale computing: formalization, specification, and analysis. *J. Parallel Distrib. Comput.* **73**, 465–483. (doi:10.1016/j.jpdc.2012.12.011)
[22]. Borgdorff J *et al.* 2014 Performance of distributed multiscale simulations. *Phil. Trans. R. Soc. A* **372**, 20130407. (doi:10.1098/rsta.2013.0407)
[23]. Chopard B, Falcone JL, Kunzli P, Veen L, Hoekstra A. 2018 Multiscale modeling: recent progress and open questions. *Multiscale Multidisc. Model. Exp. Des.* **1**, 57–68. (doi:10.1007/s41939-017-0006-4)
[24]. M. Vazquez et. al. 2018. D2.3 Report on Extreme Scaling of and Porting of Exemplar Applications to Novel Architectures. CompBioMed.
[25]. Houzeaux, G., Borrell, R., Fournier, Y., Garcia-Gasulla, M., Göbbert, J. H., Hachem, E., ... & Vázquez, M. (2018). High-Performance Computing: Dos and Don'ts. In *Computational Fluid Dynamics-Basic Instruments and Applications in Science*. InTech.
[26]. Alowayyed, S., Závodszky, G., Azizi, V. & Hoekstra, A.G. (2018) Load balancing of parallel cell-based blood flow simulations. Journal of Computational Science 24 (2018) 1–7
[27]. Patronis, A., Richardson, R. A., Schmieschek, S., Wylie, B. J., Nash, R. W., & Coveney, P. V. (2018). Modeling Patient-Specific Magnetic Drug Targeting Within the Intracranial Vasculature. *Frontiers in physiology*, *9*, 331.
[28]. Groen, D., Richardson, R. A., Coy, R., Schiller, U. D., Chandrashekar, H., Robertson, F., & Coveney, P. V. (2018). Validation of patient-specific cerebral blood flow simulation using transcranial Doppler measurements. *Frontiers in physiology*, *9*.
[29]. A. Melis et al., "Improved biomechanical metrics of cerebral vasospasm identified via sensitivity analysis of a 1D cerebral circulation model," J. Biomech., 2019.
[30]. Pappalardo, F., Russo, G., Tshinanu, F. M. & Viceconti, M. In silico clinical trials: concepts and early adoptions. Briefings in Bioinformatics (2018). doi:10.1093/bib/bby043
[31]. Carusi, A., Burrage, K. & Rodriguez, B. Bridging experiments, models and simulations: an integrative approach to validation in computational cardiac electrophysiology. Am J Physiol Heart Circ Physiol 303, H144-155 (2012).
[32]. Passini, E. et al. Human In Silico Drug Trials Demonstrate Higher Accuracy than Animal Models in Predicting Clinical Pro-Arrhythmic Cardiotoxicity. Frontiers in Physiology 8, 1–15 (2017).
[33]. Morissette, P. et al. Shortening of the electromechanical window in the ketamine/xylazine-anesthetized guinea pig model to assess pro-arrhythmic risk in early drug development. Journal of Pharmacological and Toxicological Methods 81, 171–182 (2016).
[34]. Guns, P.-J., Johnson, D., Van Op den bosch, J., Weltens, E. & Lissens, J. The electro-mechanical window in anaesthetized guinea pigs: a new marker in screening for Torsade de Pointes risk. British Journal of Pharmacology 166, 689–701 (2012).
[35]. Guns, P.-J., Johnson, D. M., Weltens, E. & Lissens, J. Negative electro-mechanical windows are required for drug-induced Torsades de Pointes in the anesthetized guinea pig. Journal of Pharmacological and Toxicological Methods 66, 125–134 (2012).
[36]. van der Linde, H. J. et al. The Electro-Mechanical window: A risk marker for Torsade de Pointes in a canine model of drug induced arrhythmias. British Journal of Pharmacology 161, 1444–1454 (2010).
[37]. Passini, E. et al. Drug-induced Shortening of the Electromechanical Window is an Effective Biomarker for in Silico Prediction of Clinical Risk of Arrhythmias. (2019).

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

[38] P. H. Luetmer, W. D. Edwards, J. B. Seward, et al. “Incidence and Distribution of Left Ventricular False Tendons: an Autopsy Study of 483 Normal Human Hearts”. In: J. Am. Coll. Cardiol. (1986).
[39] López-Yunta, M. (2018). Multimodal ventricular tachycardia analysis: towards the accurate parameterization of predictive HPC electrophysiological computational models. UPC BarcelonaTech.
[40] Santiago, A. (2018). Fluid-electro-mechanical model of the human heart for supercomputers. PhD Theses, UPC BarcelonaTech.
[41] Dowden, H., & Munro, J. (2019). Trends in clinical success rates and therapeutic focus. Nat. Rev. Drug Discov, 18, 495-496.
[42] Joseph A. DiMasi, Henry G.Grabowski and Ronald W.Hansen, "Innovation in the pharmaceutical industry: New estimates of R&D costs", Journal of Health Economics, 2016, 47, pp 20-33, DOI: 10.1016/j.jhealeco.2016.01.012
[43] Dipak Kalra, "The importance of real-world data to precision medicine", Personalized Medicine, 2019, 16, 2, DOI: 10.2217/pme-2018-0120
[44] P. V. Coveney and S. Wan, "On the calculation of equilibrium thermodynamic properties from molecular dynamics", Phys. Chem. Chem. Phys., 2016, 18, 30236-30240, DOI: 10.1039/C6CP02349E
[45] A. Bhati, S. Wan, D. Wright, P. V. Coveney, "Rapid, accurate, precise and reliable relative free energy prediction using ensemble based thermodynamic integration", Journal of Chemical Theory and Computation, 13 (1), 210–222 (2017), DOI: 10.1021/acs.jctc.6b00979
[46]. Potterton, A., Heifetz, A. & Townsend-Nicholson, A. Synergistic Use of GPCR Modeling and SDM Experiments to Understand Ligand Binding. Methods in molecular biology (Clifton, N.J.) 1705, 335-343, doi:10.1007/978-1-4939-7465-8_15 (2018).
[47]. Potterton, A. et al. Ensemble-Based Steered Molecular Dynamics Predicts Relative Residence Time of A2A Receptor Binders. J Chem Theory Comput, doi:10.1021/acs.jctc.8b01270 (2019).
[48]. Townsend-Nicholson, A., Altwaijry, N., Potterton, A., Morao, I. & Heifetz, A. Computational prediction of GPCR oligomerization. Current opinion in structural biology 55, 178-184, doi:10.1016/j.sbi.2019.04.005 (2019).
[49]. Morao, I. et al. Rapid and accurate assessment of GPCR-ligand interactions Using the fragment molecular orbital-based density-functional tight-binding method. Journal of computational chemistry 38, 1987-1990, doi:10.1002/jcc.24850 (2017).
[50]. Z. Altai, M. Qasim, **X. Li**, M. Viceconti 2019. “The effect of boundary and loading conditions on patient classification using finite element predicted risk of fracture”, Cllinical Biomechanics, 68, pp.137-43. doi: 10.1016/j.clinbiomech.2019.06.004
[51]. M. Viceconti, M. Qasim, P. Bhattacharya, **X. Li** 2018. “Are CT-based finite element model predictions of femoral bone strength clinically useful?”, Current Osteoporosis Reports, 16(3), pp. 216-23. doi: 10.1007/s11914-018-0438-8.

*“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **675451**“*

Appendix

In this Appendix, we describe with deeper and more technical detail what was done at the three different Exemplar Research areas.

Cardiovascular Exemplar Research

Univesity of Geneva: "Digital Blood" on Palabos

The focus of the Scientific and Parallel Computing Group at the University of Geneva is on high fidelity blood flow simulations. In more detail, we develop a high-performance computational framework for the simulation of fully resolved whole blood (**Digital Blood in Massively Parallel CPU/GPU Systems**) and additionally, we work on continuum models for flow diverting stents in 3d patient-specific intracranial aneurysms.

Digital Blood in Massively Parallel CPU/GPU Systems

We propose a novel high-performance computational framework for the simulation of fully resolved whole blood flow. The framework models blood constituents like red blood cells (RBCs) and platelets individually, including their detailed non-linear elastic properties and the complex interactions among them. This kind of simulations are particularly challenging because the large number of blood cells (up to billions) stand in contrast with the high computational requirement of individual constituents. While classical approaches address this challenge through simplified structural modelling of the deformable bodies (e.g., through mass-spring systems), the present framework guarantees accurate physics, desirable numerical properties through a fully featured FEM model and computational efficiency at the same order as the more simplified state-of-the-art models.

The required numerical performance is achieved through a hybrid implementation, using CPUs for the blood plasma and GPUs for the blood cells.

Blood flow is involved in most of the fundamental functions of living organisms regarding health and disease. It is essential for the transport of oxygen, nutrients, waste products, as well as of infectious parasites and metastasizing tumor cells to tissues and organs. Blood is a complex suspension of RBCs, white blood cells and platelets, submerged in a Newtonian fluid, the plasma. The accurate modelling of the collective transport of the cells in the plasma is of paramount importance since it can help us decipher not well-understood in vivo phenomena, e.g., formation of blood clots and margination of platelets. RBCs are disk-shaped cells, made of a deformable membrane containing a Newtonian solution of hemoglobin, whose role is to transport oxygen in the organism. They account for about 35-45% of the blood volume (this fraction is called the hematocrit), corresponding to roughly 10^6 RBCs per mm^3. The deformability of RBCs is strongly linked to some pathological conditions, e.g., hereditary disorders (like spherocytosis, elliptocytosis, and stomatocytosis), metabolic disorders (like diabetes, hypercholesterolemia, and obesity), malaria, or sickle anemia. Platelets are small blood cells, with a concentration between $250x10^3$ and $500x10^3$ per mm^3, at a ratio about 1 platelet to 10 RBCs.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Most of the simulations at the spatial scale of millimeters ignore the particulate nature of blood because of the tremendous computational cost. On the other hand, in the state-of-the-art fully resolved whole blood simulations, the spatial scale remains very small, of the order of a few tens of micrometers. The suggested HPC framework is built toward the direction of simulating macroscopic flows, of the order of mm^3 of whole blood, and offers to the user the possibility to address a large range of problems with clinical relevance. The constituents of this framework are the fluid solver, the solid body solver and the fluid-structure interaction (FSI) module. Our software is designed to be modular, in the sense that the components above can accommodate any state-of-the-art numerical technique to solve the fluid or solid phase.

As far as the simulation of the blood plasma is concerned, there exists a plethora of mature CFD approaches. For our simulations, we make use of the lattice Boltzmann method (LBM) which indirectly solves the Navier-Stokes equations. LBM uses a static, homogeneous lattice and the advancement of the fluid in time happens through collision and streaming operations. The above two processes alter the fluid populations which reside at every lattice site and constitute the degrees of freedom of the LBM.

Fully resolved blood flow simulations are generally achieved by coupling the fluid solver with a moving boundary condition method. In recent years the Immersed Boundary Method (IBM) boundary condition has become the most widely used technique to model biological moving membranes, like the surface of red blood cells and other blood cells. IBM has become the standard method for biological membranes, thanks to its numerical robustness, simplicity of implementation and its versatility to the deployed fluid/ solid solvers. However, IBM has several drawbacks, mainly regarding accuracy and physical interpretation. On the other side, the classic non-IBM boundary condition for LBM, despite more accurate for non-moving boundaries, shows numerical fluctuations due to the boundary movement across the fluid grid nodes.

To achieve high fidelity simulations more research is needed to gain more stable but theoretically consistent numerical methods. We are working on developing improved boundary conditions to go beyond IBM. The idea is to generalize classic boundary conditions for LBM to handle moving boundaries in a natural way.
In a typical numerical framework for blood flow, the computational time is dominated by the structural solver for the deformable blood cells. We proposed a novel approach for deformable viscoelastic bodies based on the nodal projective finite elements method (npFEM) [1]. The expression “nodal” refers to the mass lumping technique, in which both the masses and the forces are lumped on the vertices of the discretized body, and therefore the finite elements act like generalized viscoelastic springs. The term “projective” stands for the use of specially designed potential energies that help us build a fast solver based on quasi-Newton optimization techniques. Our solver inherits the versatility and robustness of FEM and is almost as fast as plain mass-spring systems (current state-of-the-art solvers). It is characterized by strong mesh independence and

*“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **675451**“*

just one set of parameters, for any mesh resolution, can successfully describe the behavior of blood cells.

For the LB and IBM parts, we use the open source library Palabos (http://www.palabos.org/), which stands for Parallel Lattice Boltzmann Solver. Palabos is a powerful open source high-performance LB solver that utilizes modern C++ and advanced Message Passing Interface (MPI) techniques. The npFEM part is written in C++/ CUDA in order to leverage the massive parallelism offered by the general-purpose GPUs.

The developed numerical framework is intended to grow to be a general-purpose tool for first-principle investigation of blood properties. The current focus of our research is the study of platelet margination [2]. This is a very complex transport phenomenon, where the platelets are pushed toward the vessel walls while the RBCs form a denser structure away from them. While this is a well-reported phenomenon, there is no clear understanding of the real mechanisms behind it. Deciphering this property of blood could help design efficient drugs that prevent clot formation and help doctors detect various cardiovascular diseases at more ease. Our analysis is not only focused on healthy subjects but also on patients with various pathological conditions, e.g., diabetes, obesity and various other hereditary disorders (linked with the RBC/ platelet deformability and shapes). Figure 3 and **Error! Reference source not found.** present a pure shear flow (velocity on top and fixed bottom wall) for healthy and diabetic subjects, respectively, as produced using our HPC framework. In the latter case, the RBCs are more swollen and less deformable, leading to a faster deposition of platelets toward the vessel walls. **Error! Reference source not found.** presents margination after 1 second of physical time.

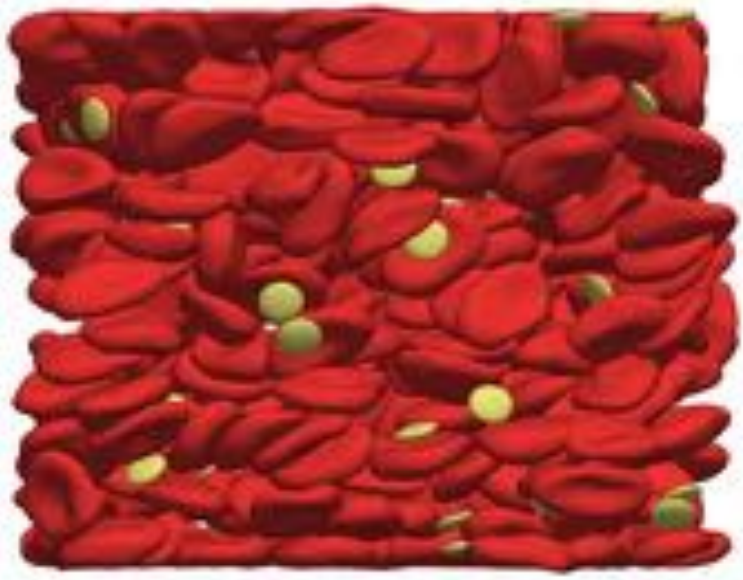

Figure 15. Shear flow with healthy RBCs (in red) and platelets (in yellow) in a domain of 50 μm^3 at 35% hematocrit.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

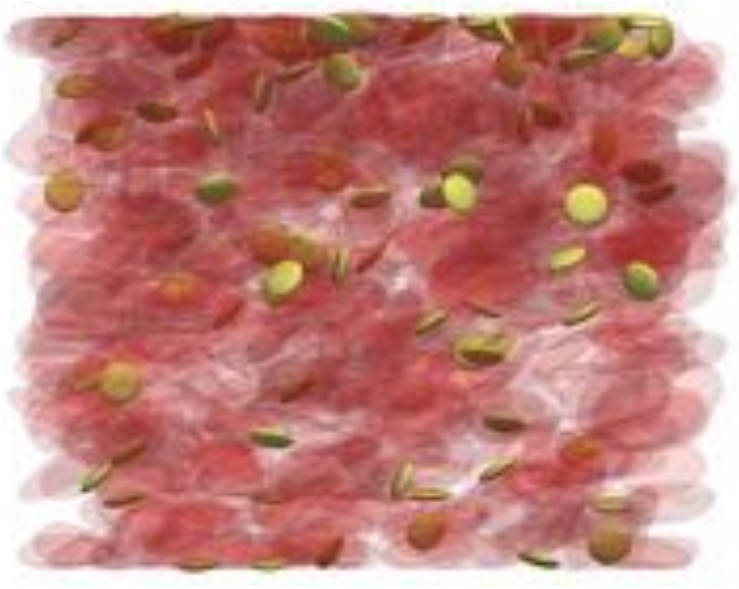

Figure 16. After 1 second of physical time, most of the platelets are closer to the upper wall. This phenomenon is called margination.

In the context of the current scientific goals (toward the simulation of macroscopic flows), the performance metrics of our parallel framework must be considered under the light of weak scaling. Indeed, the purpose of seeking more powerful computational resources is not to improve the resolution or increase the time span of the simulation, but to extend the physical volume of the blood considered in the model. Tests of the hybrid CPU/GPU code have shown that at a hematocrit of 35%, it is reasonable to assign each GPU approximately 500 blood cells, while the CPU cores on the same node treat the corresponding volume of blood plasma. In this case, the computational cost of the different parts of the code are balanced (no single part constitutes a bottleneck), and a single global iteration of the solver is carried out in less than 0.5 seconds, allowing to cover a significant physical time span in a few days of computation. An increase of the number of compute nodes translates into a proportional increase of the number of RBCs and blood plasma volume. Our research product delivers fully resolved whole blood simulations at unprecedented computational efficiency. Current state-of-the-art solvers report that the deformable blood cells solver constitutes over 95% of the total computational time, while our novel computational framework has dropped this time to about 15% of the total computational time. The proposed design deems suitable for the upcoming exascale supercomputers, allowing us to simulate physical domains and time spans that are yet to be explored.

Toward the direction of validating the fidelity of our numerical models, we have developed a tight collaboration with Dr Karim Zouaoui Boudjeltia, biologist at ULB and CHU Charleroi, who is performing flow chamber experiments with whole blood of real patients. These experiments can be compared and designed in synergy with our numerical approach.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Given the high computational cost of fully resolved 3d blood flow simulations, a physical description of platelets deposition was introduced recently in Chopard et al. (2017) [2], by integrating fundamental understandings of how platelets interact in a numerical model, parameterized by five parameters. These parameters specify the deposition process and are relevant for a biomedical understanding of the phenomena. One of the main intuition is that these parameters are precisely the information needed for a pathological test identifying Cardio/cerebrovascular diseases (CVD) captured and that they capture the inter-individual variability.

Following this intuition, we devised a Bayesian inferential scheme for estimation of these parameters, using experimental observations (from Dr Zouaoui Boudjeltia), at different time intervals, on the average size of the aggregation clusters, their number per mm^2, the number of platelets, and the ones activated per µl still in suspension. As the likelihood function of the numerical model is intractable due to the complex stochastic nature of the model, we used a likelihood-free inference scheme approximate Bayesian computation (ABC) to calibrate the parameters in a data-driven manner. As ABC requires the generation of many pseudo-data by expensive simulation runs, we use a high-performance computing (HPC) framework for ABC to make the inference possible for this model. We consider a collective dataset of seven volunteers and use this inference scheme to get an approximate posterior distribution and the Bayes estimate of these five parameters. The mean posterior prediction of platelet deposition pattern matches the experimental dataset closely with a tight posterior prediction error margin, justifying our main intuition and providing a methodology to infer these parameters given patient data. The present approach can be used to build a new generation of personalized platelet functionality tests for CVD detection, using numerical modeling of platelet deposition, Bayesian uncertainty quantification, and high-performance computing. This ongoing project is realized through a tight collaboration with Dr. Ritabrata Dutta who is an Assistant Professor of Statistics in the University of Warwick.

Continuum model for flow diverting stents in 3D patient-specific simulation of intracranial aneurysms

An aneurysm is a weak spot on a blood vessel wall that causes a balloon-like bulging. Aneurysms tend to increase in size, and can be at risk of rupturing, leading to internal bleeding, with severe consequences. In particular, rupture of intracranial aneurysms is an event with a high mortality rate, and about one third of the survivors suffer from permanent neurological or cognitive deficits. For diagnosed aneurysms that are at risk of rupturing, different forms of treatment exist, one of which includes the insertion of a prosthesis into the artery, called a flow diverting stent. It has the shape of a tube, with a surface consisting of a very fine mesh of woven wires and is inserted to cover the neck of the aneurysm. The role of this device is to divert the main bloodstream from the aneurysm to the artery, while avoiding cutting off the aneurysm from blood supply altogether. The modification in the bloodstream pattern achieved by this procedure is capable of encouraging a blood clotting reaction in the aneurysm, which in its turn fills the aneurysm dome and prevents its rupture.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Computer simulation can be of use in many ways in this field of medical science. For example, the flow mechanics of the blood, and biological processes inside the blood, are simulated to promote a fundamental understanding of the factors involved in the blood clotting process of an aneurysm and propose new means of medical treatment. On the other hand, computer simulation is also a useful tool for day-to-day medical decision making, as it can be applied, for example, whether a given patient aneurysm could be successfully treated through a stent insertion procedure, and which stent model to use for optimal results. The idea is to use medical imagery to construct a virtual model for the artery and the aneurysm of a patient, and introduce, virtually, different flow diverting stents into this artery. Numerical models known under the name of Computational Fluid Dynamics are then used to simulate the mechanical properties of blood flow in the artery and test the ability of the flow diverter to encourage blood clotting.

This type of simulation is highly complex and requires powerful computers. The difficulty arises in part from the fact that the stent surface consists of a very fine mesh, which needs to be accurately resolved in order to represent the blood flow through the wired network. The simulations can be computationally intensive to such an extent that several days are required to obtain the results. This is a prohibitively long time in the context of patient-specific, medical decision making: there exists a strong need for more efficient numerical models. We carried out a research activity to develop a faster simulation model for the blood flow across stents, which we have published in a three-article series ([4, 5, 6, 7]).

Model

The idea behind our approach is that the flow diverting stent is modeled by means of a coarse- grained, macroscopic approach. This means in practice that the simulation can be carried out at a resolution which is too coarse to represent the details of the wired stent network. The stent is replaced by a force acting on the bloodstream which, although it has less structure than the stent itself, has largely the same effect on the blood flow than the original stent.

Previous attempts to achieve such a model have been made by other authors, inspired by models in geophysical science to model porous media, such as porous rocks, through a continuum approach. The problem is that, while porous media generally have a certain thickness, stents have a very thin surface, and therefore exhibit properties that cannot be reproduced by conventional porous media models. For example, a stent, as opposed to a classic porous media, can deflect the flow quite sharply, creating a well defined relationship between the angles of the flow that reaches the stent and the flow that leaves it, respectively.

To improve the quality of the model, we chose to base our work on a different theoretical framework, the field of so-called screen models. Screens are thin, porous surfaces of industrial usage of various origins, including for example woven, rigid fabric,

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

or perfo- rated plates. Theoretical models for a continuum modeling of such screens exist, and we adapted them to the field of flow diverting stents. The beginning of our investigations was of theoretical nature and was carried out in 2D simulations ([1, 2]). We showed that screen models had a potential for overcoming the weaknesses of porous media models, and we performed necessary conceptual adaptations and parameter fittings to create a new framework for Screen-model based Stent Models (SSMs). Our modifications of the original theoretical framework of screen models included the extension of these models to inhomogeneous media, and the insertion of the modelled objects into the complex environment of an artery with stent (screens are often placed in a more regular industrial environment, such as a rigid tube).

After this fundamental investigation, we extended the SSM to 3D, and validated the model in real-life scenarios, using data (shape of the artery and aneurysm, profile of the pulsated blood flow, shape of the stent) from actual patients treated for aneurysms.

Results

Figure 4 shows the setup of three simulations that were carried out using real-life data from three different patients. In every case, a simulation was carried out at full resolution, rep- resenting the stent structure accurately (with a corresponding, very large computational cost), and using two continuum approaches, once our SSM, and once porous-media based stent model described in the literature.

Figure 18 shows, for one of the patients, measurements of the blood velocity in four different points inside the aneurysms during one heartbeat, in the fully resolved case, and with the two models. This image illustrates a general observation, namely that our SSM (blue curve) imitates the results of the fully resolved simulation (black curve) much better than the porous media-inspired model (black curve). Furthermore, with the accuracy displayed in these figures, applying the SSM can reduce the computational cost of the simulation by a factor 10 or more, as compared to the fully resolved simulations. This is sufficient to provide a tool with a fast enough response time in the context of medical decision making.

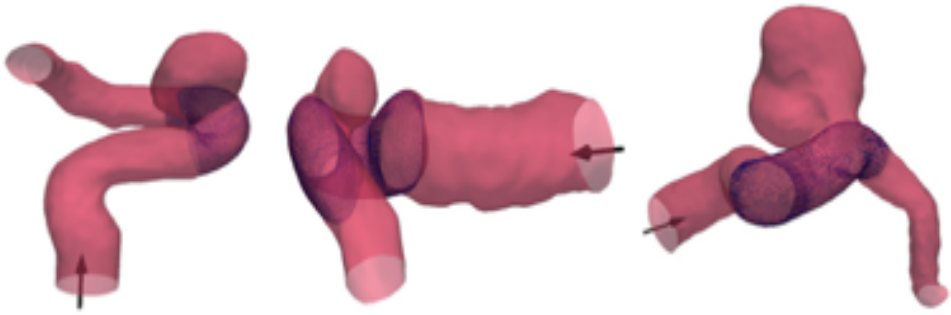

Figure 17. Blood vessel, aneurysm, and stent for the three investigated patient cases.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

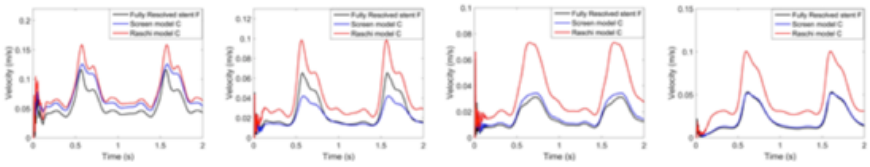

Figure 18. Measurements of the blood velocity in four different points inside the aneurysms during one heartbeat, in the fully resolved case, and with the two models.

Summary of collaborations started or enforced thanks to the project

- Laboratory of Experimental Medicine, Université Libre de Bruxelles & CHU Charleroi: Validation of our numerical frameworks with experiments on whole blood of real patients
- Department of Statistics, University of Warwick: Machine Learning techniques for fitting parameters of our numerical models based on observed data from Dr Karim Zouaoui Boudjeltia

References for this section

[1]. C. Kotsalos, J. Latt, B. Chopard, Bridging the computational gap between mesoscopic and continuum modeling of red blood cells for fully resolved blood flow, 2019, https://arxiv.org/abs/1903.06479
[2]. B. Chopard et al. 2017, A physical description of the adhesion and aggregation of platelets, R. Soc. open sci. 4: 170219. http://dx.doi.org/10.1098/rsos.170219
[3]. J. Beny, C. Kotsalos, J. Latt Toward full GPU implementation of fluid-structure interaction, presented in ISPDC 2019
[4]. S. Li, J. Latt, B. Chopard. Model for pressure drop and flow deflection in the numerical simulation of stents in aneurysms. Int. J. Numer. Meth. Biomed. Engng. 34(3) http://dx.doi.org/10.1002/cnm.2949. 2018
[5]. S. Li S, J. Latt J, B. Chopard. The application of the screen-model based approach for stents in cerebral aneurysms. Comput. Fluids 2018; 172: 651-660
[6]. S. Li, B. Chopard, J. Latt. Continuum model for flow diverting stents in 3D patient-specific simulation of intracranial aneurysms. Submitted for publication.
[7]. S. Li. Continuum model for flow diverting stents of intracranial aneurysms. PhD Thesis (2019), University of Geneva. https://archive-ouverte.unige.ch/authors/view/107024

University of Amsterdam: HemoCell, the microscopic cellular library for supercomputers

The developments were realised in three distinct areas. First and foremost, the advancements and optimisations were carried out on the open-source HemoCell (High pErformance MicrOscopic CELlular Library) simulation code. These allow larger domain sizes and longer simulated physical time-scales. Furthermore, improvements were made to the robustness of the mechanical model, i.e. the mechanical description of flowing cells, using uncertainty quantification (UQ). In tandem, similar UQ was applied to the in-stent restenosis model. Finally, these improvements made new applications possible.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

The cell level trafficking was explored in both healthy cases, and in cases with rigidified red blood cells, which serve as a model for diabetic disease.

Simulation software improvements

The improvements in the computational framework increased the computational performance significantly, and they can be categorised as improvements towards better load-balancing during the simulation execution, and towards more efficient inter-process communication between the computing processes. In the following we explain the steps we took in these two directions.

Load balance strategies for multi-physics problems in large-scale blood flow simulations

The non-homogeneous distribution of computational costs is often challenging to handle in highly parallel applications, especially in multi-physics problems. In [1], the author studied the fractional load imbalance overhead in a high-performance biofluid simulation aiming to accurately resolve blood flow on a cellular level, using a methodology based on fractional overheads. In general, the concentration of particles in such a suspension flow is not homogeneous. Usually, there is a depletion of cells close to walls, and a higher concentration towards the centre of the flow domain, causing a time-dependent and potentially high computational work imbalance. We perform parallel simulations of such suspension flows. The emerging non-homogeneous cell distributions might lead to strong load imbalance, resulting in deterioration of the parallel performance. The authors formulate a model for the fractional load imbalance overhead, validate it by measuring this overhead in parallel lattice Boltzmann based cell-based blood flow simulations, and compare the arising load imbalance with other sources of overhead, in particular the communication overhead. They find a good agreement between the measurements and our load imbalance model. We also find that in our test cases, the communication overhead was higher than the load imbalance overhead. However, for larger systems, we expect load imbalance overhead to be dominant. Thus, efficient load balancing strategies should be further developed.

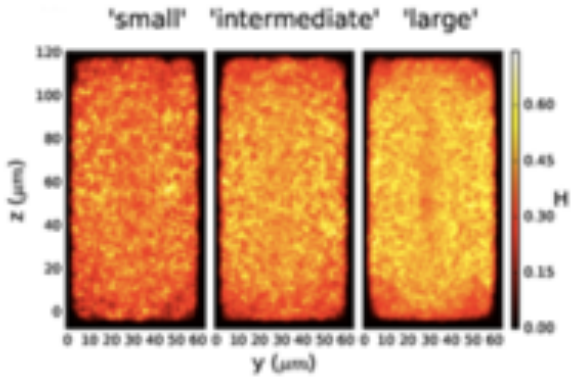

Figure 19. Haematocrit distribution for the channel flow case with different numbers of Red Blood Cells (N), average haematocrit H = 38%. 'small' (left), 'intermediate' (middle), and 'large' (right) systems. Extracted from [1].

Increasing MPI communication efficiency for the HemoCell codebase (UvA)

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

For HemoCell we have also looked into improving the efficiency of the communication between MPI processes. The communication of the cell material information between the processes was one on the major bottlenecks of the simulation, therefore, we targeted this area by restructuring the communication pattern and restricting information exchange to data that is strictly necessary. Improving the efficiency in this context means that we altered the communication structure, but not the resulting computation. We added a new step in the communication where instead of the fixed communication envelope we use a pre-compiled list of necessary information. This presents a minimal computational overhead that is counterweighted by the gain in reduced communication time.

The performance improvement results for simulations like those of Figure 19, are shown in Figure 20 and Figure 21. By reducing the amount of data communicated and improving on the used algorithms and data structures we managed to get an overall improvement of approx. 100% in wall clock time, and in the strongest scaled case we get an improvement of 350%. Furthermore, the strong scaling properties are improved as well.
In practical scenarios this roughly means that if we simulate blood flow in microfluidic chip for 1 s, the computation time is reduced from 10 days to around 3 days.

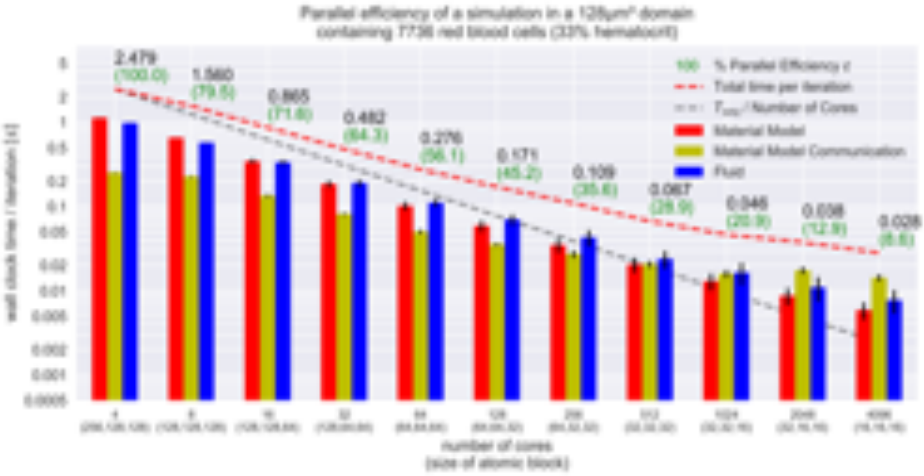

Figure 20. Performance of HemoCell before optimization. The blue fluid bar encompasses communication as well as computation. The total time per iteration is written in black.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

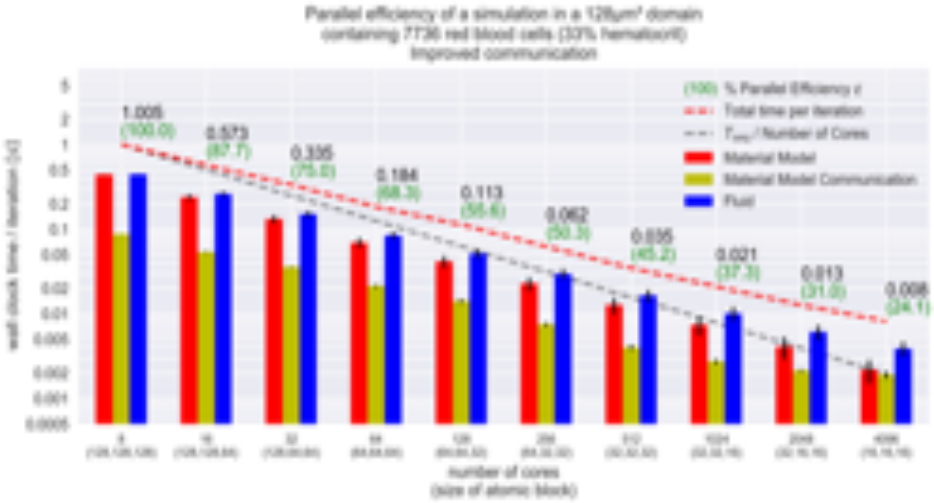

Figure 21. Performance of HemoCell after the optimization of the material communication. Both performance graphs are generated on different supercomputers top: SuperMUC, bottom: Marenostrum, therefore the difference between wall-clock times might not be due to optimization, however, this should not affect the parallel efficiency numbers (green percentages).

Model developments

Inverse Uncertainty Quantification for HemoCell

Uncertainty Quantification (UQ) is an indispensable part of model certification process, where uncertainty in the output of the model is estimated in order to analyse comprehensively the model prediction (forward UQ) and uncertainty in the model input parameters is evaluated in order to access their realistic distributions (inverse UQ) [2]. In the present work, inverse UQ was applied to the HemoCell model of blood flow at the cellular level [3], where uncertainties in the link force coefficient (κ_l), bending force coefficient (κ_b) and viscosity contrast (Λ) were estimated. Additionally, two cases of healthy and treated red blood cells (RBC) were implemented for the inverse UQ study.

In order to analyse the parameters' identifiability, the Sobol sensitivity analysis method [4] was applied. The results show that the link force coefficient and viscosity contrast are both identifiable, whereas identifiability issues arise for the bending force coefficient. This parameter was still included to the inverse UQ analysis, however, in future work must be analysed using an output parameter, which is more sensitive to its value.

The results of the inverse UQ using the Transitional Markov Chain Monte Carlo algorithm [5] predict the following mean values of the analysed parameters: $\kappa_l \approx 40 \pm 5$ kBT, $\Lambda \approx 5.2 \pm 1.8$ and $\kappa_b \approx 275 \pm 6$ kBT for the healthy RBC data, and $\kappa_l \approx 117.5 \pm 2.5$ kBT, $\Lambda \approx 2.6 \pm 1.5$ and $\kappa_b \approx 320 \pm 10$ kBT for the treated RBC data. Forward UQ with obtained distributions of the input parameters is shown in Figure 22 for both healthy and treated RBC cases. It was concluded that the inverse UQ method has yielded values of the model

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

parameters, with which the model predictions of the elongation index match well the available experiential data.

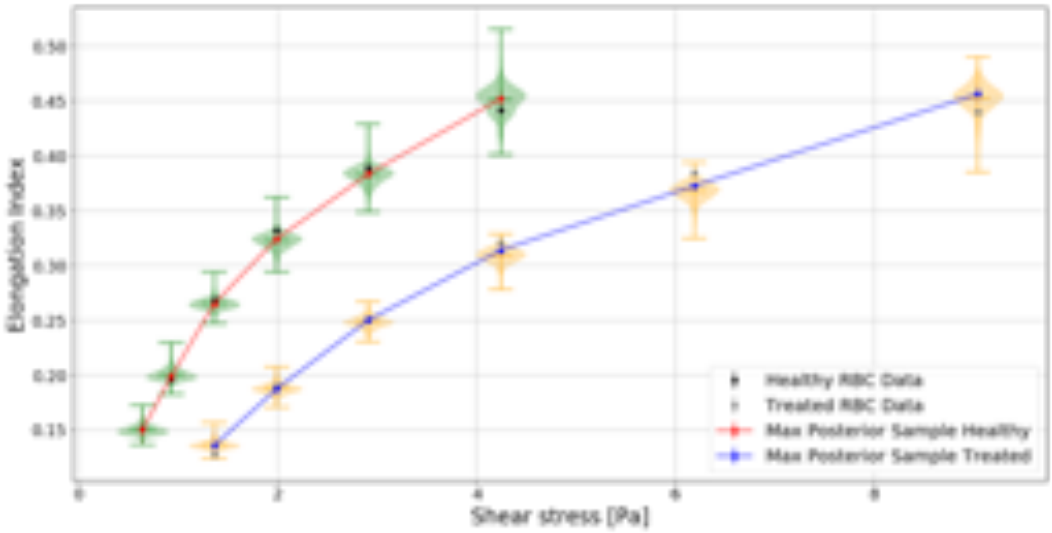

Figure 22. Forward uncertainty propagation in the HemoCell model with the obtained distributions of the input parameters from the inverse Uncertainty Quantification analysis.

Uncertainty Quantification for the in-stent restenosis model

An important aspect of cardiovascular disease is the behaviour of the affected vessel after medical intervention. One important medical treatment is stenting: inflating a balloon inside the affected vessel to open it up and placing a metal mesh called a stent to keep it open. However, this treatment damages the vessel wall, which causes growth and proliferation of smooth muscle cells (SMCs) in the wall. The dynamics of this process are largely governed by the blood flow dynamics in the vessel and by endothelium recovery in the stented area. If the growth is excessive, the vessel can re-narrow and reduce blood flow again, potentially warranting another intervention. This makes it important to study and understand the growth process, and one way to do it is by computational modelling.

A multiscale model of tissue growth in stented arteries, or in-stent restenosis (ISR) model, has been previously built [8,9]. However, to make the model actionable and usable for clinically relevant predictions and in silico clinical trials, it is important to understand the relative importance and effects of the model inputs, based on experimentally measured parameters, on the predictions made by the model. To this end, sensitivity analysis (SA) and uncertainty quantification (UQ) were performed for the ISR model. Endothelium regeneration dynamics, flow velocity and stent deployment depth were studied by performing a quasi-Monte Carlo UQ [6]. To reduce the computational costs, the 2D version of the model was used. The UQ and SA showed that the endothelium regeneration time and the flow velocity in the vessel are the most influential model parameters, so obtaining accurate values for them should be a priority before using the model for in silico clinical trials. Additionally, a methodology for semi-intrusive UQ was proposed, which can be used to reduce computational costs, so that

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

UQ can be performed for the 3D version of the model as well. The results of semi-intrusive UQ were close to the reference quasi-Monte Carlo UQ predictions, but the semi-intrusive UQ results were systematically biased towards lower values of tissue growth (see Figure 23) [7]. This bias can be eliminated, however, by enhancing the metamodels used for semi-intrusive UQ.

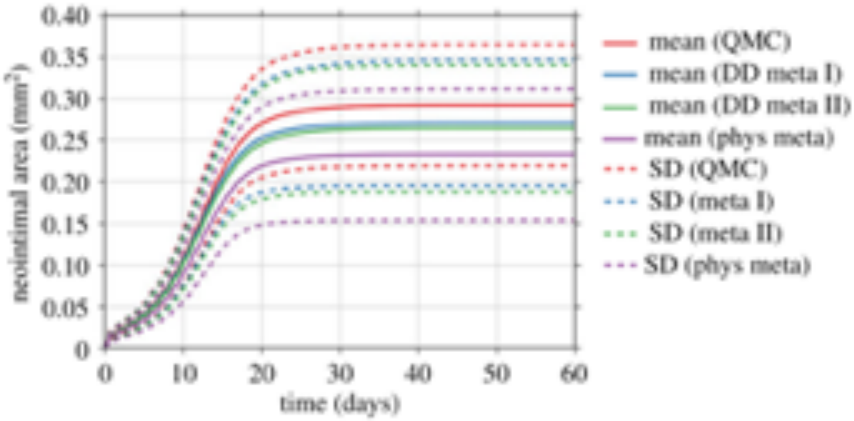

Figure 23. Comparison of the estimated mean and standard deviation (SD) by the quasi-Monte Carlo (QMC) method, semi-intrusive metamodelling methods by data-driven (DD meta I and II) approach and by simplified physics (phys meta).

Framework application

Studying the diffusivity of RBCs and the diffusivity and margination of platelets

The radial distribution of cells in blood flow inside vessels is highly non-homogeneous. This leads to numerous important and unique properties of blood, and has many clinical implications.

The mechanisms shaping these distributions are not fully understood. In this section the investigation of the motion of cells in a straight vessel section is summarised. Such investigation is made possible by the abovementioned improvements in the simulation code and the material model. The motion of cells is governed by a variety of hydrodynamic interactions and cell-deformation mechanics. Properties, such as the effective cell diffusivity, are therefore historically difficult to investigate in flows other than pure shear flows, as the various arising effects are difficult to separate. To this end, several single-cell, cell-pair, and large-scale many-cell simulations were performed using a validated numerical model. Apart from the single-cell mechanical validations, the arising flow profile, cell free layer widths, and cell drift velocities were compared to previous experimental findings. The detailed results of the investigation can be found in reference [10].

The motion of the cells at various radial positions and under different flow conditions was extracted, and evaluated through a statistical approach (Figure 24). The evaluation of particle trajectories showed that the margination of platelets cannot be the net result of gradients in diffusivity, as it was assumed multiple times in the literature. However,

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

the margination mechanism is strongly linked to the gradient of the hematocrit level. Finally, the investigation showed that platelets marginate only until the edge of the red blood cell distribution and they do not fill the cell free layer.

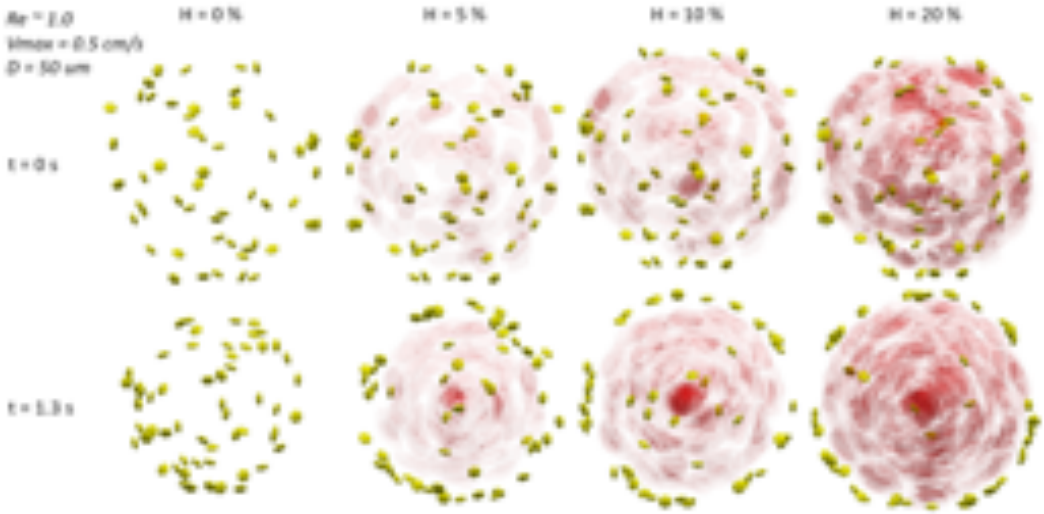

Figure 24. Platelet (yellow disks) distribution projected to a vessel cross-section at different haematocrit levels. The red ghosts show the red blood cells. The top line shows the initial cell positions, while the bottom line shows the positions after 1.3 s of flow. The columns depict various haematocrit levels. The platelets migrate to the outer layer of the flow at a few percent haematocrit already, outlining the importance of cell-cell interactions.

Studying the influence of red blood cell deformability with HemoCell

Whole blood is a suspension of cells, red blood cells (RBCs), platelets, and white blood cells, suspended in a protein rich plasma that collectively has a non-Newtonian rheology. RBCs are the most numerous blood cells and due to their deformability and bi-concave shape the RBC contributes significantly to the complex rheology of whole blood. Pathologies have been found to affect the deformability of the red blood cell such as Diabetes, Sickle Cell Anemia [11], and HIV. In this research we use HemoCell (High pErformance MicrOscopic CELlular Library) which is a numerical model of blood flow at the cellular level [3] to probe the effects of RBC deformability on flowing whole blood through micro vessels with diameters ≤300 μm.

This study begins at the single cell level with simulations of two colliding RBCs of varying membrane stiffness. RBC membrane stiffness of the HemoCell model is matched with ektacytometry measurements from chemically (tert-Butyl hydroperoxide (TBHP)) stiffened RBCs [12]. We observe from single cell simulations that in a collision between a deformable and stiff RBC that the stiff RBC is pushed away from the original positon more compared to the deformable RBC.

The results of the single cell collisions inform larger scale bulk flow simulations of whole blood (Figure 25). We perform simulations of flowing whole blood through a micro vessel, with different mixtures of stiffened RBCs. We observe with an increase of the

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"

number of stiff RBCs present in flow that platelet populations at the vessel wall decrease. This may have important implications for patients with pathologies that effects the deformability of RBCs as it may hinder their ability to properly form blood clots.

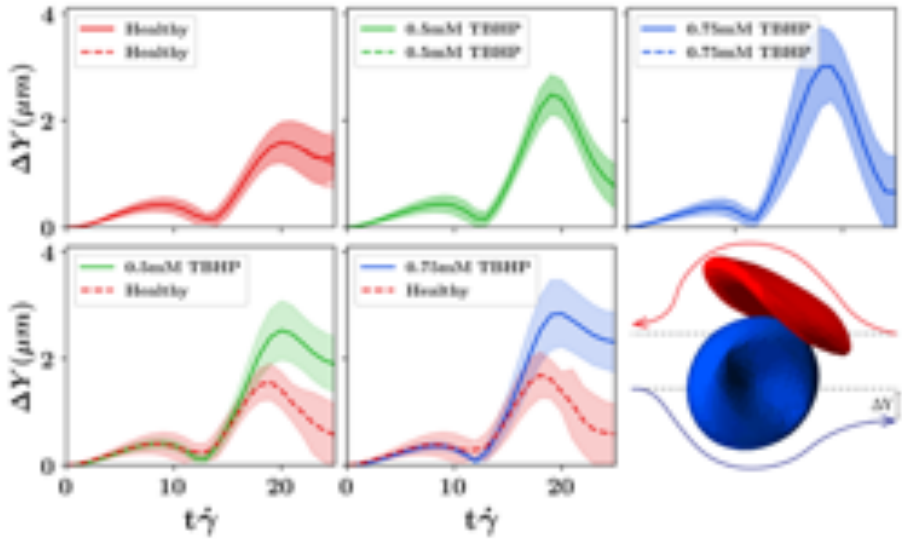

Figure 25. Displacement of RBCs from their original position resulting from single RBC collisions with varying membrane stiffness. The top row are homogenous RBC collisions with membrane stiffness increasing from left to right (red:healthy, green:stiff, blue:stiffest). The bottom row are heterogeneous collisions between a healthy RBC and a stiff RBC.

References for this section

[1]. Alowayyed, S., Závodszky, G., Azizi, V. & Hoekstra, A.G. (2018) Load balancing of parallel cell-based blood flow simulations. Journal of Computational Science 24 (2018) 1–7

[2]. R. C. Smith, “Uncertainty quantification: theory, implementation, and applications”, vol. 12. Siam, 2013.

[3]. G. Zavodszky, B. van Rooij, V. Azizi, and A. Hoekstra, “Cellular level in-silico modeling of blood rheology with an improved material model for red blood cells”, Frontiers in physiology, vol. 8, p. 563, 2017.

[4]. I. M. Sobol, “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates”, Mathematics and computers in simulation, vol. 55, no. 1-3, pp. 271-280, 2001.

[5]. J. Ching and Y.-C. Chen, “Transitional markov chain monte carlo method for Bayesian model updating, model class selection, and model averaging", Journal of engineering mechanics, vol. 133, no. 7, pp. 816-832, 2007.

[6]. Nikishova, A., Veen, L., Zun, P., & Hoekstra, A. G. (2018). Uncertainty Quantification of a Multiscale Model for In-Stent Restenosis. Cardiovascular Engineering and Technology, 9(4), 761–774. https://doi.org/10.1007/s13239-018-00372-4

[7]. Nikishova, A., Veen, L., Zun, P., & Hoekstra, A. G. (2019). Semi-intrusive multiscale metamodelling uncertainty quantification with application to a model of in-stent restenosis. Philosophical Transactions of the Royal Society A: Mathematical, Physical

*“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **675451**“*

and Engineering Sciences, 377(2142), 20180154. https://doi.org/10.1098/rsta.2018.0154
[8]. Tahir, H., Bona-Casas, C., Narracott, A. J., Iqbal, J., Gunn, J. P., Lawford, P. V., & Hoekstra, A. G. (2014). Endothelial repair process and its relevance to longitudinal neointimal tissue patterns: comparing histology with in silico modelling. J. R. Soc. Interface, 11, 20140022. https://doi.org/http://dx.doi.org/10.1098/rsif.2014.0022
[9]. Zun, P. S., Anikina, T., Svitenkov, A., & Hoekstra, A. G. (2017). A Comparison of Fully-Coupled 3D In-Stent Restenosis Simulations to In-vivo Data. Frontiers in Physiology, 8(May), 284. https://doi.org/10.3389/fphys.2017.00284
[10]. Závodszky, G., van Rooij, B., Czaja, B., Azizi, V., de Kanter, D., & Hoekstra, A. G. (2019). Red blood cell and platelet diffusivity and margination in the presence of cross-stream gradients in blood flows. Physics of Fluids, 31(3), 031903.
[11]. Rees, D. Williams,T. Gladwin, M. Sickle cell disease. The Lancet, 376: 2018-2031, 2010.
[12]. Gutierrez, M. Fish, M. Golinski, A. and Eniola-Adefeso. O. Presence of Rigid Red Blood Cells in Blood Flow Interferes with the Vascular Wall Adhesion of Leukocytes. Langmuir, 34: 2363-2372, 2018

University College London: HemeLB for 3D flows in large sparse geometries

Using HemeLB, a highly optimized lattice-Boltzmann solver for haemodynamic flow in large, sparse 3D geometries, our research during the course of the CompBioMed project has focused on two strands - one for automated, validated and patient specific simulation, and the other on the computational optimization work required to best exploit coming exascale infrastructure.

Magnetic Drug Targeting (MDT) capabilities were implemented in HemeLB, allowing the dynamics of paramagnetic iron-oxide particles (in practice often used as a drug delivery method) to be observed in a geometry obtained from an MRI scan of a patient's brain [1] (Figure 26). This model allows exploration of the effect of changes in particle size, magnet strength and placement, and patient physiology (heart rate, etc) on the delivery to a given target site (e.g. a tumour site) for a patient specific vascular system. In this multiscale model, the inlet velocity profiles for the 3D region are generated using a 1D Navier-Stokes solver representing the rest of the human body.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Figure 26. A small section of the simulated Circle of Willis geometry, showing the paramagnetic drug particles passing through a (pink) target region under the influence of a (blue) magnetic dipole.

In another study [2] we collaborated with the National Hospital for Neurology and Neurosurgery (NHNN) in London, to obtain Transcranial Doppler (TCD) measurements of blood velocity in the Middle Cerebral Artery (MCA) of a stroke patient, along with CT scan data allowing the 3D vasculature (of that same patient) to be generated for use in HemeLB simulations. A simple validation study was then carried out by comparing the velocity profiles from simulations against those measured by TCD at multiple points along the MCA. We also considered the sensitivity of the simulations to changes in rheology model (Newtonian vs shear-thinning) and in mesh resolution.

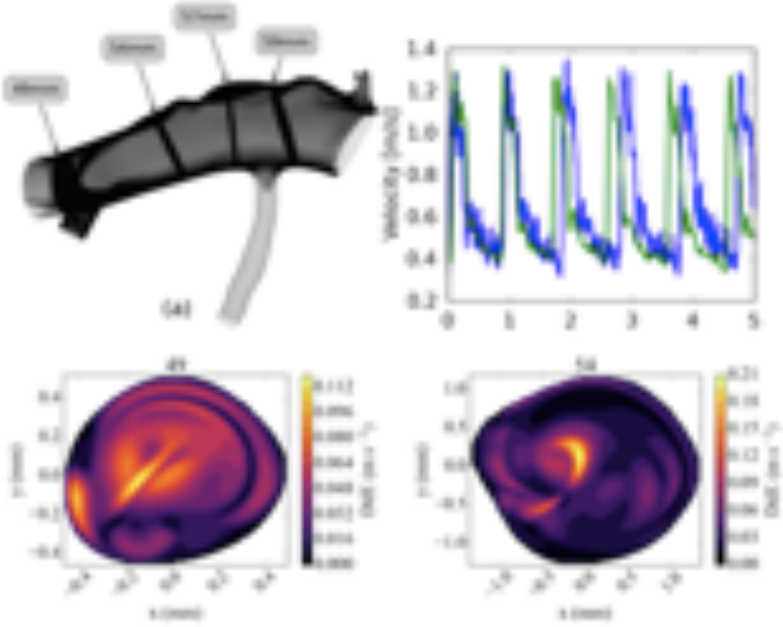

Figure 27. Top left: The Middle Cerebral Artery, showing the several depths of measurement planes carried out with the transcranial doppler device. Top right: A comparison of the peak velocity profile in the simulated and (in vivo) experimentally obtained cases. Bottom: Cross-sectional plots showing deviation in velocity field for differing choice of rheology model.

In order to use HemeLB in the clinic, a large amount of work was dedicated to fully automating the entire pipeline, from CT scan to simulation to visualisation. In particular,

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

real-time visualization is sought after for clinicians, in part due to the short timescales of treatment (Figure 27). We carried out this work as part of a larger collaboration project with Hamad Medical Corporation and Qatar University in Doha, Qatar. This work is still on-going, but some early publications are being produced [3]. These efforts are heavily focused on the patients, in terms of patient-specificity of data, and on interaction with the clinician (such as Interventional Radiologists) to determine the best metrics and visualizations to communicate useful information about the simulation results (Figure 28).

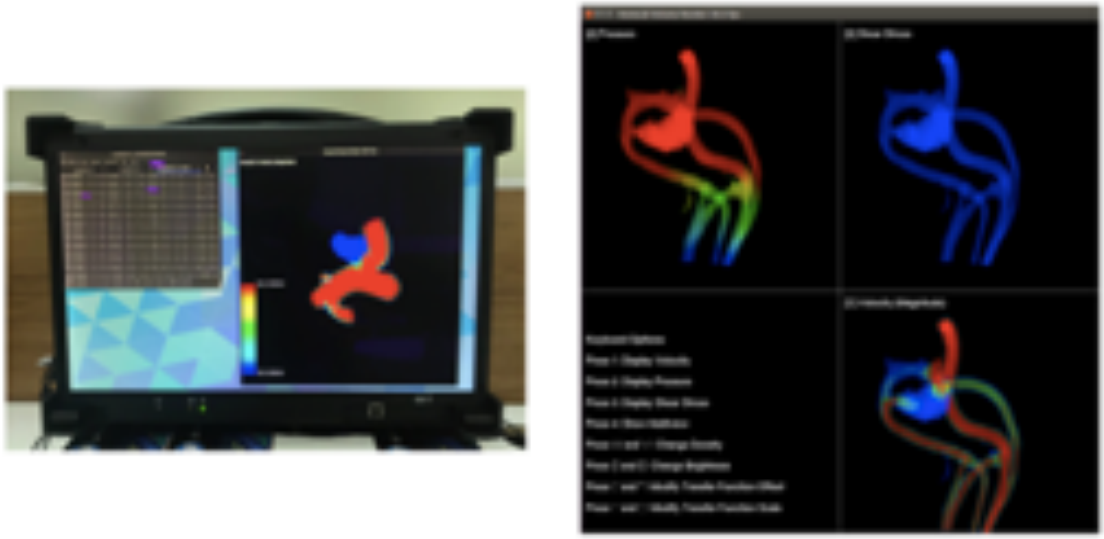

Figure 28. Left: The portable workstation running the automated pipeline. Right: A multi-view visualisation window. The 3D display of the results can be illustrated in two different modes: 1) multi-view window which is depicted in this Figure; 2) full separate window for each lattice property. The multi-view window splits into 4 sub windows; one for the visualization instruction as shown in the bottom left and three for the velocity, pressure and shear stress.

With regards to the computational efficiency of HemeLB, and coming exascale resources, we have expended significant effort towards memory optimization and load balancing, leading to strong scaling up to hundreds of thousands of cores. For example, HemeLB now scales very well up to 256000 cores on the Blue Waters supercomputer (Figure 29).

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

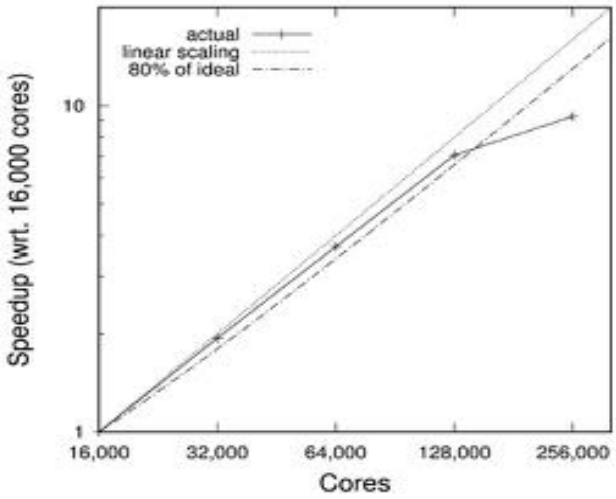

Figure 29. HemeLB strong scalability up to 256000 cores on the Blue Waters supercomputer.

Thanks to these performance improvements, we are now tackling simulations of the entire human vasculature (Figure 30). Our current work concerns coupling of the full human arterial tree to the corresponding venous tree, as well as to a full electrophysiologically detailed simulation of the human heart (the Alya code developed at BSC).

Coupling HemeLB is necessary to allow for the simulation of more complicated anatomical features. The self-coupling of HemeLB has been structured to allow for flow in major vessels to be resolved whilst treating capillaries as a sub-scale feature. In this process, a map is created linking the outlets of the first HemeLB instance and the inlets of the second. Scale factors are applied to both averaged velocity and pressure as they are passed between these locations and used to reconstruct the boundary conditions on the opposing instance. These factors are chosen such that the fundamental flow physics is conserved between the two simulations. It is also believed that this coupling strategy will prove to be advantageous for efficient performance on the architectures of some of the next generation supercomputers. Focusing on SuperMUC-NG in particular, its CPU cores are clustered into separate islands with slower communication channels between them. Being able to couple multiple instances of HemeLB would allow for components of a complex geometry to be simulated on an individual island with inter-island communication being streamlined to only the minimum boundary information required.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

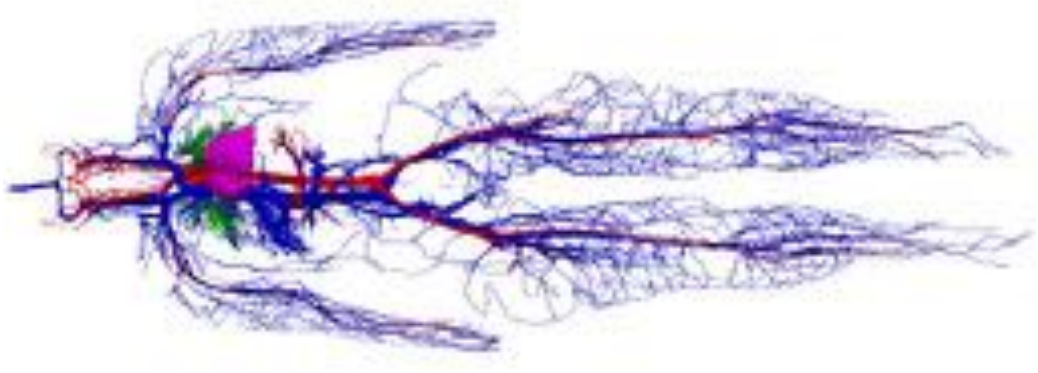

Figure 30. Full human arterial tree and venous tree used in the HemeLB-HemeLB coupling code.

References for this section:

[1] Patronis, A., Richardson, R. A., Schmieschek, S., Wylie, B. J., Nash, R. W., & Coveney, P. V. (2018). Modeling Patient-Specific Magnetic Drug Targeting Within the Intracranial Vasculature. *Frontiers in physiology*, *9*, 331.

[2] Groen, D., Richardson, R. A., Coy, R., Schiller, U. D., Chandrashekar, H., Robertson, F., & Coveney, P. V. (2018). Validation of patient-specific cerebral blood flow simulation using transcranial Doppler measurements. *Frontiers in physiology*, *9*.

[3] Esfahani, S. S., Zhai, X., Chen, M., Amira, A., Bensaali, F., AbiNahed, J., ... & Coveney, P. V. (2019). HEMELB Acceleration and Visualization for Cerebral Aneurysms. *arXiv preprint arXiv:1906.11925*.

Lifetec: AngioSupport for coronary artery disease

Cardiac teams in the larger hospitals daily discuss the treatment of multiple patients with coronary artery disease (CAD). These patients have one or multiple severe occlusions in the coronary arteries which are complicated cases and requires the expertise of the cardiac team. For each patient a treatment plan is defined, typically consisting of coronary artery bypass graft (CABG) surgery or percutaneous coronary intervention (PCI). The decision between these treatments is currently based on studying coronary angiograms and the experience of the cardiac team. However, in case of multiple occlusions, diffuse coronary disease or complicated vasculature, choices in the position, length or diameter for a CABG or PCI is challenging.

To help the cardiac team in this process, LifeTec Group talked with Pim Tonino**,** **intervention cardiologist** at Eindhoven Catharina Hospital, about a possible numerical model that can assist in this decision making. Together with Frans van de Vosse, professor of the biomedical engineering department of Eindhoven university of technology, LifeTec Group started the development of a clinical tool that could assist the cardiac team in treatment planning for each patient. Therefore, AngioSupport is developed; an interactive tool to predict the outcome of CABG or PCI to support **clinical decision making** of coronary interventions.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

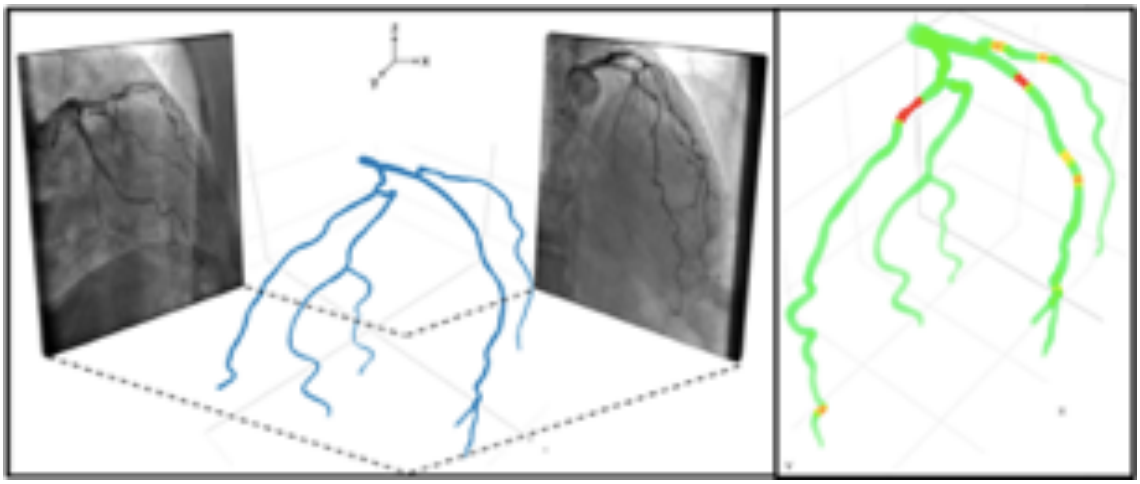

Figure 31. Obtaining the 3D vessels by using the 2D images created during coronary angiography.

For each cardiac team meeting, the patient specific coronary vasculature is needed. AngioSupport will show the full 3D coronary vasculature to assess the stenotic areas. For this step, Lifetec is working together with **Pie Medical Imaging.** Their CAAS software is able to compute the 3D vessels by using the 2D images created during coronary angiography (Figure 31). With their help, we are able to create a full coronary vasculature from these coronary angiograms. This allows us to visualize the CAD in every artery and to show stenotic areas by automatic stenosis recognition.

To compute the blood flow and pressure inside the coronary tree, a 1D wave propagation model is implemented in AngioSupport. The research from the biomedical engineering department at the **Eindhoven University of Technology** is used. We adapted the models to the requirements of AngioSupport, for example to achieve a fast computational time and to use the available patient data. AngioSupport can show the Fractional Flow Reserve (FFR) through the full coronary vasculature (Figure 32). This can help the cardiac team, since the FFR is used to indicate whether a region has insufficient blood supply and therefore needs revascularization.

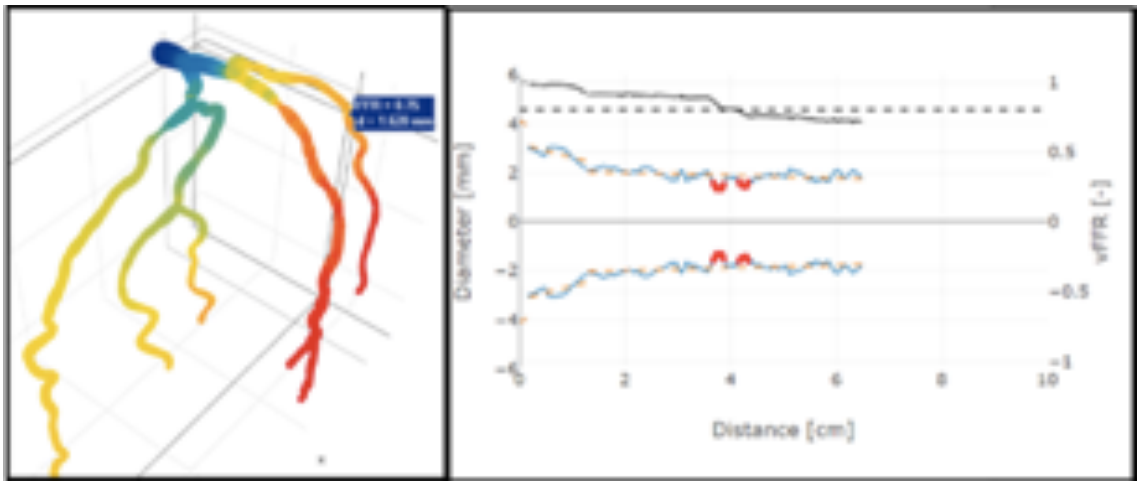

Figure 32. LifeTec Group's AngioSupport interface: computing the virtual FFR.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Since AngioSupport is designed to be used during a cardiac team meeting, the heart team should be able to access the 1D wave propagation simulations themselves. We therefore designed an interface in cooperation with **Dr. Pim Tonino.** This interface allows the cardiac team to see the 3D coronary vasculature and the recognised stenotic areas. The cardiac team can add then the patient data and start the simulation to see the FFR throughout the coronaries.

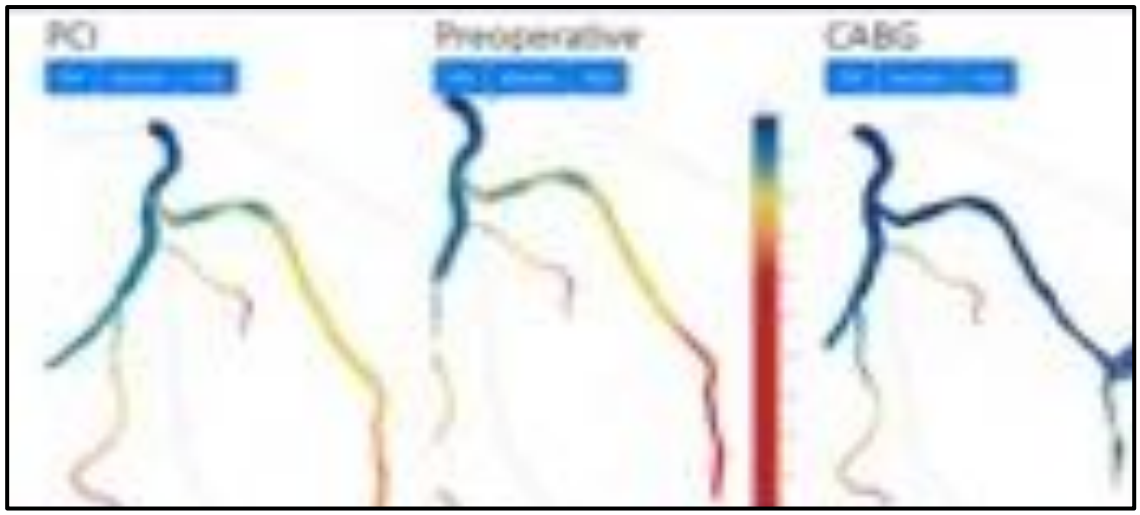

Figure 33. LifeTec Group's AngioSupport interface: the virtual intervention outcome.

AngioSupport then also allows the cardiac team to perform an intervention (Figure 33). The interface enables them to virtually place a stent or place a bypass graft. AngioSupport then shows the result of these interventions within seconds. The cardiac team can then easily compare the results, which helps them in their coronary intervention planning. By allowing the cardiac team to place the stent or bypass themselves in the patient specific coronary arteries, they are able to test the planned intervention virtually. This makes AngioSupport a valuable tool to justify a treatment plan. With the use of AngioSupport, the cardiac team is now able to use the wave propagation models as developed at the Technical University of Eindhoven. Through the AngioSupport interface, the cardiac team can start their own CFD simulations, without the extensive knowledge needed to build these CFD tools.

AngioSupport consists of combined research from the Eindhoven University of Technology**.** This resulted in many elements in de model, such as the 1D line elements, coronary windkessel elements, junction elements and anastomosis elements, which resulted in a large amount of parameters. To investigate the impact of these parameters, a large sensitivity analysis was performed with the help of **SURFsara**. By using their **high performance computing center**, a large amount of simulations were performed to investigate the impact of each parameter. This allowed us to prioritize which parameters needs to be connected to patient data and improve the reliability of the AngioSupport simulations.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

With help from Dr. Simon Dello, **cardiologist in Catharina Hospital**, we received patient data from 75 patients with a complicated coronary artery disease. These patients were all discussed during a heart team meeting. At LifeTec Group, this retrospective data was used to test the feasibility and reliability of AngioSupport. The model was able to create a 3D coronary vasculature of the patient and to automatically recognize the stenotic areas of each patient. The numerical model was also able to calculate the FFR for each patient within seconds and the cardiologist was able to virtually perform placing a stent or a bypass. Although some improvement is still needed in calculating the FFR virtually, the cardiologist showed interest in using the model. LifeTec Group also visited the **Amsterdam Medical Center** and the **Erasmus Medical Center** to give a live demonstration of AngioSupport. Both medical centers showed large interest in the model and the assist it gives for treatment planning. Especially the close relationship between pressure, flow and resistance can be investigated easily with AngioSupport.

AngioSupport could therefore be used during clinical decision making, but also for the training of cardiologists. By being able to virtually perform interventions, the cardiac team can directly see the effect of geometrical changes in pressure and flow distribution. This could increase the insight in coronary hemodynamics, especially for cardiologist still in training. Complicated cases can be simulated in AngioSupport and the result of different interventions can be simulation. LifeTec Group is exploring this possible new business line together with cardiologists in Catharina Hospital in Eindhoven.

The involvement of LifeTec Group in CompBioMed also started the exploration of more of simulation work at LifeTec Group. Since many companies visit LifeTec Group during their preclinical development, the use of numerical simulation could be a useful addition. For instance the use of 3D CFD simulations combined with fluid structure interactions can greatly improve the insight in design choices. As for companies which are for instance developing heart valves, vascular stents or heart assist devices. LifeTec Group is currently starting to build this simulation team at LifeTec Group, to also have 3D computational fluid dynamics simulations and to connect this with fluid structure interaction models. LifeTec Group's close work with experimental work gives them a great advantage, since they these companies visit the experimental set up at LifeTec Group. This gives more insight in numerically simulating these user cases and helping them in preclinical development. For the computational fluid dynamics software, LifeTec Group is exploring the use of Alya developed at **Barcelona Supercomputing Center.** LifeTec Group already is already supervising projects with students from Technical University of Eindhoven which are working on 3D CFD simulations.

For instance, the development of a graft which is used to perform a coronary artery bypass graft around a coronary occlusion could be helped by a combined experimental and numerical platform. This graft consists of a platinum ring which can be attached to the coronary artery without suturing. This procedure can be experimented on the beating heart platform at LifeTec Group. This set up will show the possible downsights of the design. However, with CFD simulations, LifeTec Group could take this a step

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

further. By using the experimental set up and simulating this numerically, small changes in design can be investigated. LifeTec Group could then simulate many design changes which could increase the insight in this device and even give advice for improvements of this device. This would be a great benefit for companies still in preclinical trials.

It is worth to remark the translational character of AngioSupport. It is a clinical decision tool to support the cardiac team with treatment planning for patients with coronary artery disease (CAD). Using AngioSupport, stenoses as well as their severity are easily revealed within seconds by computation of blood flow and pressure inside the coronary arteries. The coronary angiograms are used as input, which are already made for these patients. Clinicians can perform multiple interventions virtually and compare the predicted outcome of each intervention. AngioSupport therefore allows **the heart team** to use the numerical models developed at the **Technical University of Eindhoven**. The interface developed for AngioSupport is created with cardiologists and made for their daily use. Patient data of 75 patients was used in the model and tested for feasibility and reliability in clinical practice.

Finally, it is also worth to be mentioned that CompBioMed has brought an important impact on Lifetec business, especially on three lines: a tool for cardiac team for planning coronary interventions, a tool for training application for cardiologists and a tool for CFD simulation studies, in combination with fluid structure interaction.

Summary of collaborations started or enforced thanks to the project

- **Technical University of Eindhoven:** Building the numerical model used in AngioSupport. Many new possibilities for bachelor, master or PhD students for research to improve AngioSupport.
- **Catharina hospital Eindhoven:** the development of AngioSupport and closer contact with clinical practice.
- **Pie Medical Imaging:** The collaboration in AngioSupport and the segmentation of a full coronary vasculature. A business could be started together with Pie Medical for possible spin-out opportunities.

University of Sheffield: OpenBF for vascular networks

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

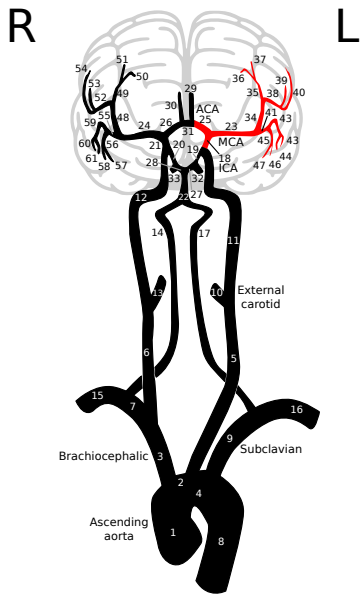

Figure 34. Diagram of the OpenBF cerebral vascular network model.

OpenBF is a 1D hemodynamics network model developed at the University of Sheffield. Cerebral vasospasm (CVS) is a life-threatening condition that occurs in a large proportion of those affected by subarachnoid haemorrhage and stroke. CVS manifests itself as the progressive narrowing of intracranial arteries. It is usually diagnosed using Doppler ultrasound, which quantifies blood velocity changes in the affected vessels, but has low sensitivity when CVS affects the peripheral vasculature. In a recently published study we aimed to identify alternative biomarkers that could be used to diagnose CVS [1]. For this we used a verified and validated 1D modelling approach, openBF, to describe the properties of pulse waves that propagate through the cardiovascular system (Figure 34), which allowed the effects of different types of vasospasm on waveforms to be characterised at several locations within a simulated cerebral network. The model has been previously quantitatively validated with predictions from other models and echo-Doppler velocity measurements from vasospasm patients. A sensitivity analysis empowered by the use of a Gaussian process (GP) statistical emulator was then used to identify waveform features that may have strong correlations with vasospasm. The use of GP statistical approaches was previously analysed and validated against more traditional Monte Carlo analyses using cloud supercomputing systems available in the CompBioMed consortium [2]. In the sensitivity analysis model inputs for lumen radius, length, Young's modulus and peripheral resistances of each vessel were initialised with typical reference values and then changed within ±50% of their reference values to then analyse the effect of these changes on waveform features that would be easy to extract in the clinical context. These consisted of minimum, maximum, and time average in one cardiac cycle of velocity and pressure waveforms, along with their first time derivatives. In Figure 35 each plot shows sensitivity indices for each input. The hatched section of the bars shows the first-order sensitivity indices (measuring contribution to variance in each model output from variance of each model input) , and the plain sections the total-order indices (measuring contribution to variance in each model output from the

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

interaction of one input with other inputs). Thus, the height of the hatched bar shows the biomarker sensitivity to a single input and the plain section the sensitivity to multiple inputs.

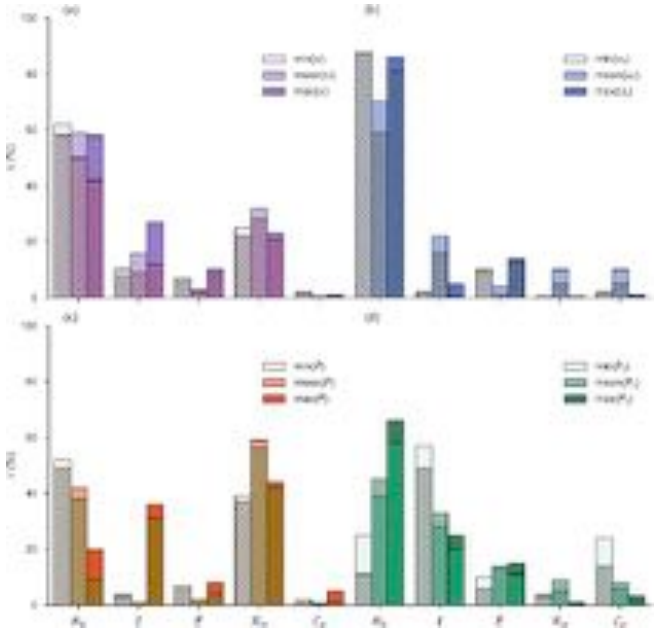

Figure 35. Sensitivity indices for u (velocity) and P (pressure) biomarkers against variations of the model input parameters (R_0 lumen radius, ℓ vessel length, E Young's modulus, R_p peripheral resistance, and C_p peripheral compliance).

A GP emulator can treat inputs and outputs explicitly as uncertain quantities, and so by determining the proportion of output variance that could be accounted for by each uncertain input we were able to calculate variance-based sensitivity indices for each input and output of the model. This was useful to identify those waveform features that are sensitive to vasospasm (changes in vessel radii) but less sensitive to physiological variations in the other model parameters. Using this approach, we showed that the minimum rate of velocity change can detect presence of vasospasm from an early stage (10% in diameter reduction) and be more effective than blood velocity for stratifying typical manifestations of vasospasm and its progression. In the wider context, the present study describes the use of sensitivity indices, combined with modelling, as a way to identify effective biomarkers, which is a novel approach that has the potential to result in clinically useful tools.

The same approach has been further developed and applied to the simulation of endovascular removal of blood clots (thrombectomy) as a potential clinical tool to investigate typical clinical scenarios for treatment of ischaemic stroke.

References for this section

[1]. A. Melis et al., "Improved biomechanical metrics of cerebral vasospasm identified via sensitivity analysis of a 1D cerebral circulation model," J. Biomech., 2019.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

[2]. A. Melis, R. H. Clayton, and A. Marzo, "Bayesian sensitivity analysis of a 1D vascular model with Gaussian process emulators," *Int. j. numer. method. biomed. eng.*, vol. 33, no. 12, 2017.

Barcelona Supercomputing Center and University of Oxford: Alya Cardiac Computational Model

Human-based computer models and simulations are a fundamental asset of biomedical research. They augment experimental and clinical research through enabling detailed mechanistic and systematic investigations. Owing to a large body of research across biomedicine, their credibility has expanded beyond academia, with vigorous activity also in regulatory and industrial settings. Thus, human *in silico* trials are now becoming a central paradigm, for example, in the development of medical therapies [1].

Human cardiac physiology is one of the most advanced areas in physiological modelling and simulation. Current human models include detailed information on the ionic processes underlying the action potential such as the sodium, potassium and calcium ionic currents, exchangers such as the Na/Ca exchanger and pumps such as the Na/K pump. They also include representation of the excitation-contraction coupling system, which modulates the calcium transient and, in turn, myocyte contractility. Human cardiac models are also multiscale, both spatially and temporally, and integrate information across the subcellular, cellular, tissue, and organ levels [2].

Here we showcase the human multiscale models we developed through the integration of multimodality datasets, including: ionic current measurements; action potential and calcium transient recording; active force measurements; magnetic resonance and computed tomography images; electrocardiograms. Human data were used at multiple stages of model development, for calibration and also to perform independent validations at different scales.

Examples of how these models have been used to characterise adverse outcome pathways and identify preclinical and clinical biomarkers for cardiotoxicity are included, together with an investigation of the underlying physiological mechanisms relevant to cardiotoxicity in specific (patho)physiological conditions. Thanks to high performance computing facilities we were able to conduct extensive computer simulation studies of cardiac electromechanical activity using the multiscale models for a range of reference compounds leading to a variety of cardiovascular outcomes.

From the engineering point of view, the heartbeat can be decomposed in three different physical problems. In the muscle, the electrical stimuli propagates along the cardiac myocytes, which contract deforming the macroscopic geometry. This produces a change of the volume within the cardiac chambers that are filled with blood. Ventricular computational fluid dynamics (CFD) has to be solved in order to compute the pressure produced by the blood against the endocardium. Since we can decompose the problem in these three sub-problems, we can say that the heartbeat is a fluid-electro-mechanical phenomenon. Each one of these sub-problems is computationally demanding by itself.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

On the one hand, non-linear ordinary differential equations (ODEs) governs the electrical propagation and an exponential orthotropic material models the myocardium solid mechanics. On the other hand, large deformation occurs in the fluid domain with step changes in the velocities. When these problems are coupled, the computational cost multiplies making supercomputer resources a requirement to solve the proposed model.

Physicians evidence this tight three-way coupling on a daily basis. The ECG, which measures electrical potentials in the skin of the patient, can be used to diagnose mechanical pathologies like myocardial hypertrophy. In the same way, a reduced ventricular output can have an electrical or mechanical etiology. Electrophysiology or electromechanical models are useful when analyzing localized events in the heart. But, if a more extensive overview in the heartbeat phenomena is expected, it is required to increase the model complexity including the blood fluid dynamics. Great advances have been achieved in heart modelling from the first electrical wave propagation simulations. Current models are incremental evolutions from previous, slightly simpler stages. Creating a fluid-electro-mechanical model of the heart is a new required stage to improve heart models.

Each one of the independent problems (electrophysiology, solid mechanics and fluid dynamics) is, by itself, computationally demanding. When these problems are coupled, computational costs grows more than the sum of the independent parts. For this reason, efficient and scalable solvers for each problem are required, together with a proven performance for the coupled model.

The Alya Cardiac Computational Model and its numerical and implementation aspects

For the fluid-electro-mechanical problem and from the description in the previous lines, we identify two coupling points. On the one hand we have the electro-mechanical coupling between electrophysiology mechanics. On the other hand, we have the bidirectional structure-fluid coupling, frequently called Fluid-Structure Interaction (FSI). We describe below each model and the two coupling points, all of them implemented in Alya. For the three problems, the space discretisation is based on the Finite Element Method and the time discretisation is based on the Finite Differences Method.

Electrophysiology

The electrical depolarization of the heart is orchestrated by the specialized conduction system, which regulates heart rate and synchronous depolarization. When talking about normal ventricular depolarization, the phenomena starts in the Purkinje network, which drives the electrical impulse to the ventricular endocardium, initiating the depolarization of the myocardium. Once a few myocytes are excited, a so called all-or-nothing process starts, where the induced electrical wave propagates to the whole heart.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

At the scales we are considering, waves of depolarization and repolarization can be modelled under the continuum mechanics hypothesis. In the tissue scale the electrophysiology model can be seen as a transient diffusion PDEs which includes a term that describes the ion cell kinetics, modelled by an ordinary differential equation (ODE) system.

Different approximations have been proposed for cardiac electrophysiology models, such as monodomain or bidomain equations. But, the simplification into monodomain formulation is perhaps the most common one [41, 42, 43, 44] and the one we use.

As said above, at a cell scale, the action potential propagation is a discrete process [45]. The propagation occurs form one myocyte to the neighbouring one through gap junctions. For larger scales, such as tissue scale, the propagation seems to be smooth [46]. For cardiac electrophysiology models, the action potential propagation is considered as continuous. This is a first and general assumption for modelling cardiac electrophysiology [47]. Bidomain model describing the propagation in the myocardium can be derived from the classical cable theory for electricity current along the neurones [48]. Monodomain models assume that the conductivity of extracellular and intracellular regions are proportional.

From the simple models of Hodgkin-Huxley formulation where only Na+ current was modelled, models have became more complex both mathematically and computationally. In the last decade a huge number of phenomenological cell models had been proposed (in chronological order): [49, 50, 51, 52, 53, 54, 55, 56]. On those models the degree of physiological detail and mathematical and computational complexity varies, due to the number of currents, pumps and exchangers describing the cell dynamics included. In the presented examples we used the O'hara-Rudy (ORd) model [56], described below.

Solid mechanics and Excitation Contraction coupling

To model solid mechanics we use the finite elasticity framework. The solid mechanics in the heartbeat problem should include the stresses produced by the material model, the boundary conditions, the fluid that is making pressure in the solid walls, and the active tension induced by the myocytes. The way pressure is imposed in the endocardium and the equations for active tension. The solid problem is governed by the linear momentum balance.

In cardiac tissue, stress is assumed to be a combination of passive and active stress

The passive part is modeled as a slightly compressible invariant-type material and through a transverse isotropic exponential strain energy function. This constitutive relation describe the response of a material to applied loads, which depends on the internal constitution of the material. The active part introduce tension along the fibers which depends on the Ca++ concentration computed from the electrophysiology model. The coupling model is described below.

Fluid dynamics and fluid-structure interaction

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"

The last building block in the heartbeat problem is the set of equations to model fluid dynamics. The blood is modeled as Newtonian and incompressible. This fluid is modeled with the Navier-Stokes equations in deformable domain, with a so-called arbitrary Lagrangian-Eulerian (ALE) formulation.
In the following pages, we briefly explain the methods followed and the results obtained.

The lowest level: Single Cell Models

Electrophysiology
Several human action potential (AP) models have been proposed for ventricular electrophysiology, and amongst them the ORd model [3], developed based on experimental recordings from more than 100 human hearts. Its key strengths are the representation of CaMKII signalling, the capability to manifest arrhythmia precursors such as alternans and early after-depolarisation (EADs), and the good response to simulated drug block and disease remodelling [4-7]. As a consequence, the ORd model was selected by a panel of experts as the model best suited for regulatory purposes [5]. A graphic representation of the ORd model is shown in Figure 36.

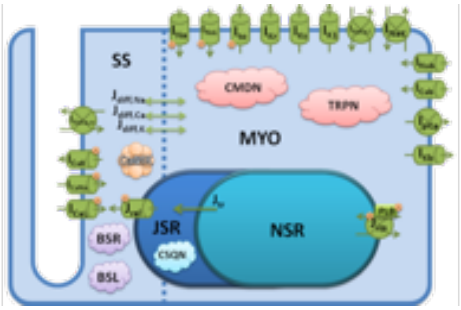

Figure 36. Graphic representation of the biophysically-detailed ORd model, showing all the ionic currents, pumps/exchangers, sub-cellular compartments, buffers and ionic fluxes included in it, and represented by ordinary differential equations.

Electro-mechanics
The electrical activity of the cell is summarised by the action potential (AP), which triggers the calcium transient (CaT), a large release of calcium from the sarcoplasmic reticulum (SR) into the cell. Cardiac contraction is produced by the interaction of filaments containing the proteins myosin and actin, and regulated by tropomyosin and troponin C. The calcium binds to troponin C, thus causing tropomyosin to move from the actin filaments, so that myosin heads can bind to produce contraction. Therefore, the CaT is the main link between excitation and contraction in the heart muscle.
Drugs can affect cardiac electrophysiology and contractility in many different ways. As an example, certain drugs directly affect specific cardiac ion channels (e.g. Dofetilide: hERG channel block), thus inducing changes on the AP and in turn on CaT and contractility, while other drugs can act on different mechanisms (e.g. Pimobendan: myofilament calcium sensitizer), thus still affecting the mechanical function of the heart, but not via electrophysiological changes [8].

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

To investigate drug-induced changes to cardiac contraction, we developed an electro-mechanical model of human ventricular cardiomyocytes, by combining the ORd model with the most recent model of human cardiac contractility [9]. The input of the electrophysiological model is the stimulus current, defined by the specific simulation protocol. The outputs are action potential (AP) and calcium transient (CaT). The latter acts as input of the contractility model, whose output is the iso-sarcometric force developed by the cell. The contractility model also gives back to the electrophysiological model information about the fraction of calcium bound to troponin C (CaTRPN), which will induce electrophysiological changes in the following heartbeat. This mechanism is called excitation-contraction coupling (ECC), as summarised in Figure 37. Calcium constitutes the connection between electrophysiology and mechanics: the CaT computed by the ORd model becomes the input of the Land model (ECC), which in turn computes the fraction of calcium bound to troponin, to be used as input for the next ORd model computation. Cardiac contraction can also, in turn, cause changes to the electrophysiology. This is defined as mechano-electrical feedback (MEF) and a notable example is represented by the stretch-activated channels (SAC) [10], which are also be included in the model.

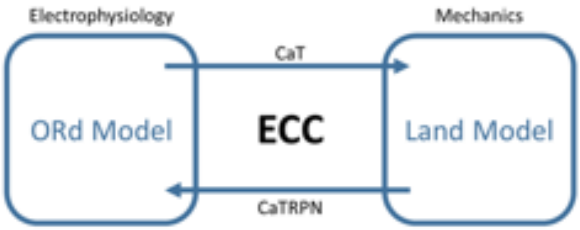

Figure 37. Schematic representation of the computational model of human cardiac electro-mechanics, designed by coupling the ORd model (for electrophysiology) and the Land model (for mechanics).

Moving up: Multiscale Models

Electrophysiology

The 3D computational model consists of a biventricular mesh, embedded in a torso volume with defined lung and bone regions, as shown in Figure 38. The average size of each element in the biventricular mesh is 0.4 mm, to ensure numerical convergence of the numerical algorithms [4, 11].

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

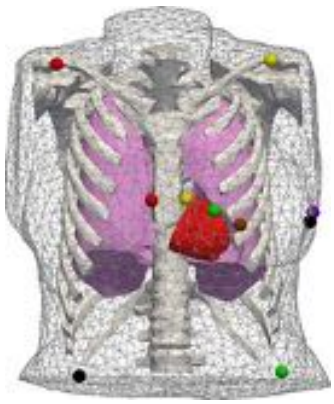

Figure 38. Image extracted from [12]. Visualization of the combined heart-torso mesh. The coloured spheres indicate the location of the virtual electrodes, using standard (European) colour-coding.

Any electrophysiological model (e.g. ORd or ToR-ORd) can be used to represent the membrane kinetics in the heart. Transmural and apex to base cell electrophysiological heterogeneities are defined based on experimental and clinical data [13-16], and incorporated in the biventricular mesh. Transmural heterogeneities are modelled using three layers: endocardial, mid-myocardial and epicardial cells [4,17]. Apex-to-base heterogeneities, known to play a key role in the formation of T waves in the ECG [18] are modelled by including a gradual increase of I_{Ks} conductance from base to apex, resulting in APD_{90} differences of 25-40 ms [19]. Heart fibre directions are generated using the Streeter rule-based method [20], and tissue conductivities are set to generate conduction velocities in line with what measured experimentally in myocardial fibres across the longitudinal, transversal and transmural axis [21,22]. The propagation of the electrical activity in the human ventricles is modelled using bi-domain equations and solved with the Chaste software [23]. Sinus rhythm is simulated at 1 Hz using a phenomenological activation model with early endocardial activation sites and a fast endocardial layer, representing a tightly-packed endocardial Purkinje network, as in [12]. Simulation of the ECG signal is computed by calculating the extracellular potentials in ten nodes in the torso surface [12], corresponding to the standard electrode positions for a 12-lead ECG. When simulating the bi-ventricular mesh only, a pseudo ECG can still be computed using the dipole model [24].

This 3D computational framework has been successfully validated in multiple studies [4, 12, 15], also comparing simulation results against experimental data at tissue and organ level, under healthy and pathological conditions (e.g. acute ischemia).

Electro-mechanics

By integrating the human ventricular electro-mechanical model into the 3D framework described above, we obtained a 3D electro-mechanical model of the human heart. This required also the implementation of non-linear solid mechanics equations, to translate the active force developed at the single cell level into stretch and stretch rate at the whole-organ scale. Figure 39 pictorically describes the model. The main components of the model are: nonlinear solid mechanics (blue), electrical propagation (green), ventricular cell electrophysiology (yellow), and cellular contractility (red), with arrows

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

representing the coupling mechanisms between the different components. Electrophysiology is on the left, mechanics is on the right. Single cell models are at the bottom, while 3D models are at the top.

Figure 39. Schematic representation of the 3D human ventricular electro-mechanical model and simulation framework.

This 3D electro-mechanical model of Alya allows computing additional biomarkers of cardiac function, e.g. ejection fraction (EF) and left ventricular pressure (LVP), in additional to the ECG signal. Importantly, the model was constructed and validated based on human data. The left ventricle is modelled as an ellipsoid, truncated at the base, which provides an end diastolic volume of 160 ml, as shown for human male hearts [26-28]. Myocardial fibre directions are considered parallel to the basal plane in the middle of the cardiac wall, and vary linearly up to forming an angle of ±60° with the basal plane at the endocardium/epicardium, and tangent to the circumferential direction. Figure 40 shows geometry and fibre strcture, where fibre vectors are coloured according to cell type: endocardial (blue), mid-myocardial (red), and epicardial (green).

Sheet directions are parallel to the vectors normal to the endocardial/epicardial surfaces, and do not vary throughout the cardiac wall. The finite element method (FEM) is used for model solving, considering linear tetrahedra with a maximum edge length of ~0.025 cm, resulting in ~33M elements and ~5.6M nodes. Details are included in a manuscript which is submitted for publication [30].

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

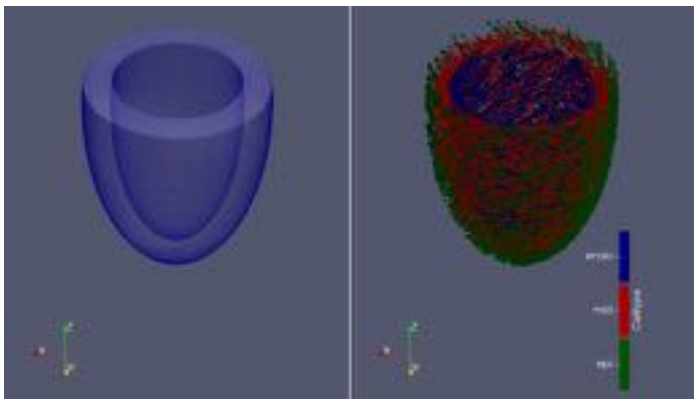

Figure 40. Computational mesh (left) and fibre architecture (right) of the ellipsoid model. The computational mesh is made of linear tetrahedra.

Results summary

Single Cell Models

Human in Silico Drug Trials to Predict Risk of Torsade de Pointes
During the first year of the project, we demonstrated the predictive power of populations of human ventricular AP models for prediction of drug-induced Torsade de Pointes (TdP) risk based on repolarisation abnormalities occurrence. We were able to achieve a prediction accuracy of 89% for a set of 62 reference compounds. The results of these *in silico* drug trials were published [31], and also led to the award of the International 3Rs prize in 2017.

More recently, we performed a similar study, including an additional biomarker: the electro-mechanical window (EMw), defined as the delay between the duration of electrical and mechanical systole, which has been suggested as a promising biomarker to predict clinical risk of Torsade de Pointes (TdP) arrhythmia in several pre-clinical animal models [32-35]. Our single cell surrogate of the *in vivo* EMw was able to predict TdP risk for a dataset of 40 compounds with 90% accuracy, confirming the potential of drug-induced EMw shortening as a biomarker for pro-arrhythmic risk. The results of these *in silico* drug trials are currently under review for publication [36].

All these studies were performed using the ORd human ventricular model. Therefore, after developing our new ToR-ORd model, we are also performing human *in silico* drug trials with a new population of ToR-ORd models. Results obtained with the new model are in agreement with the ones obtained with the original ORd model for most compounds, with an improvement in predictions for sodium blockers. As an example, *in silico* drug trial results for Mexiletine are shown in Figure 41, for the two models. Mexiletine induced early after-depolarisations in the ORd model (right panel), and it is therefore mis-classified as risky drug, while it is generally considered safe. This is due to the non-physiological increase in calcium, following sodium block in the ORd model,

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

which is now corrected in the ToR-ORd model. Indeed, Mexiletine doesn't cause any early after-depolarisation in the ToR-ORd model (left panel). These results are included in manuscript presenting the ToR-ORd model, currenty under review for publication [37].

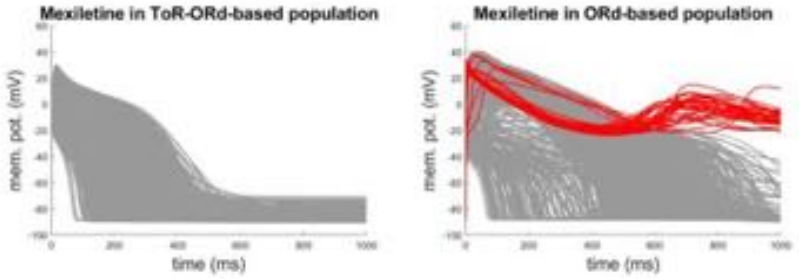

Figure 41. A comparison of in silico drug trial results for a high dose of Mexiletine in the ToR-ORd (left) and ORd (right, based population of human ventricular models. Traces classified as EADs are plotted in red (manifesting only in the ORd population).

In Silico Predictions of Drug-induced Changes in Contractility

These are preliminary results obtained by testing the effect of two reference compounds on our single cell model of cardiac electro-mechanics, to explore drug-induced changes in active tension. Figure 42 and Figure 43 show simulation results for Dofetilide and Verapamil at multiple concentrations.

- **Dofetilide:** Drug-induced changes in AP, CaT and active force, and comparing the results against experimental data obtained in human cardiomyocytes [38]. Dofetilide is a drug with high TdP risk, known to cause early after-depolarisations.

- **Verapamil:** Concentration dependent changes in active force. From the generated dose-response curve, we estimated the IC50 and h for active tension reduction, which was in agreement with what observed experimentally for sarcomere shortening [38]. Verapamil is a drug considered safe, and known to have a negative inotropic effect.

We performed similar test for a variety of reference compounds, and a manuscript with these results in currently in preparation.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

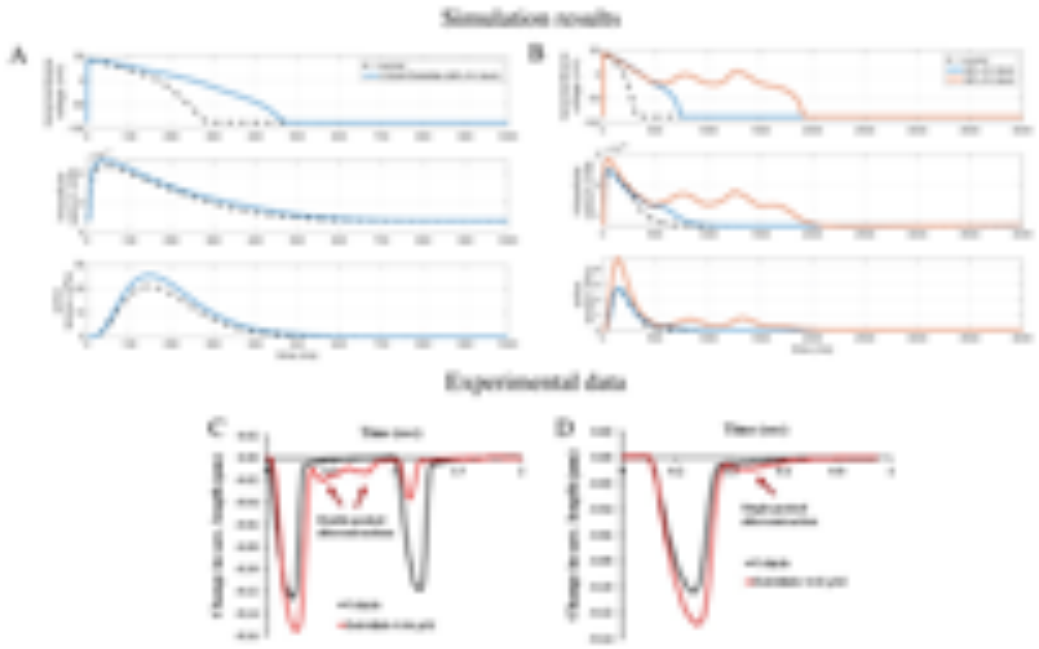

Figure 42. Simulation results for Dofetilide, obtained used our human electro-mechanical model. A: when considering low doses, Dofetilide induces AP prolongation and causes a slight increase in CaT and active force; B: when considering higher dose and slow pacing, Dofetilide induces early after-depolarisations, which also induce after-contractions; C, D: experimental data by [38], showing after-contractions for higher concentrations of Dofetilide.

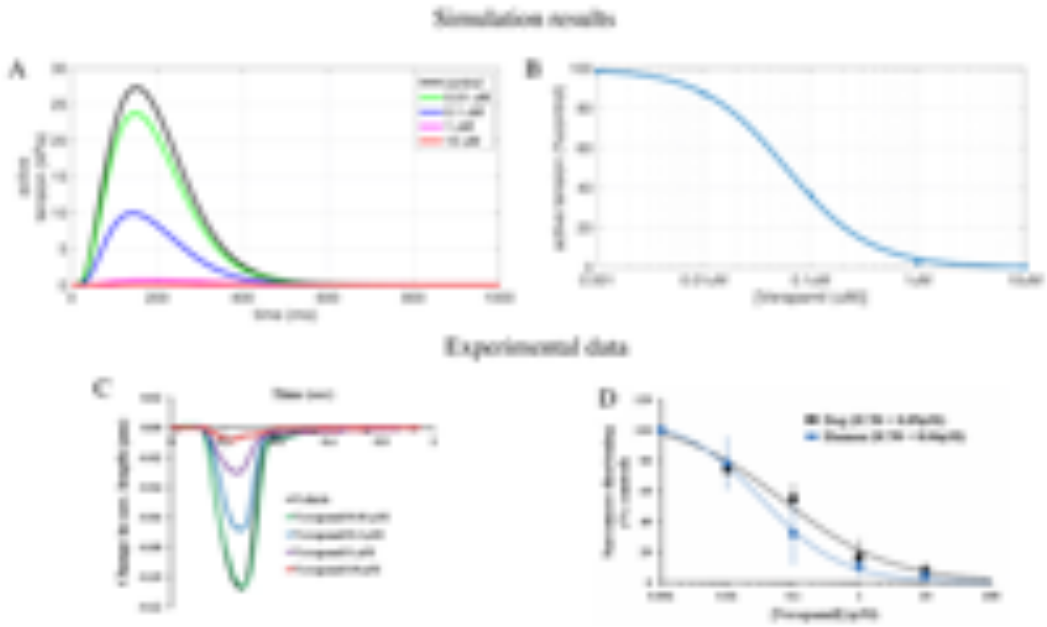

Figure 43. Simulation results for Verapamil, highlighting its negative inotropic effect. A, B: simulation results. C, D: experimental data by Nguyen et al. (2017). Verapamil decreases tension and sarcomere shortening in a concentration-dependent manner.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Multiscale Models

Simulations of the ECG signal in a 3D Model

Figure 44 shows a representative example of the simulated ECG signal, obtained using a biventricular and torso mesh, with the ToR-ORd model [37]. Simulation results are in good agreement with clinical recordings from a healthy subject, PTB database [39].

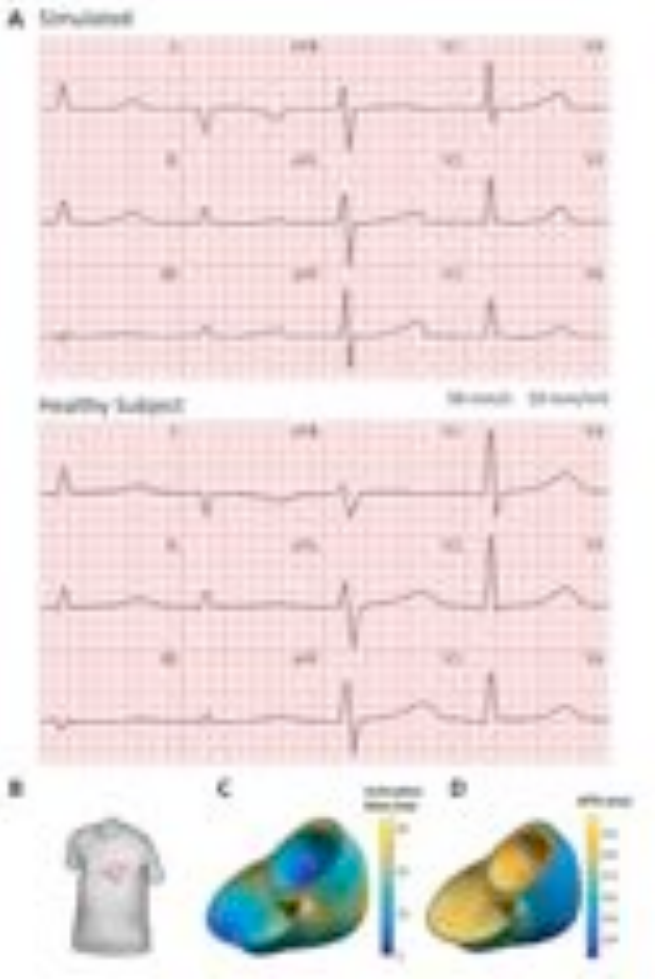

Figure 44. Simulated vs clinical 12-lead electrocardiogram. A) 12-lead ECGs at 1Hz: simulation using the ToR-ORd model in an MRI-based human torso-ventricular model (top panel) and a healthy patient ECG record (bottom panel, https://physionet.org, PTB database, subject 122 [39]). B) Virtual electrode positions on the simulated torso. C) Activation time map. D) APD map.

3D Electro-mechanical Simulations of the Human Heart

The multiscale human cardiac electro-mechanical model with ellipsoidal geometry is able to simulate all the four phases of the cardiac cycle, as illustrated in Figure 45. At the beginning of the cycle, an *initiation* phase brings the system to a physiological end-diastolic configuration, by increasing the ventricular pressure applied to the endocardium up to a physiological value for left ventricular end-diastolic pressure, while keeping the volume constant. The second phase is the *isovolumetric contraction*, where the aortic valve is closed, and the ventricular pressure increases due to contraction of the ventricles caused by electrical activation, while keeping the volume constant. When

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

the ventricular pressure surpasses the arterial pressure, the aortic valve opens and the *ejection* phase begins, eventually leading to a reduction in ventricular volume. Then, when the ventricular flow reverses, the isovolumetric relaxation phase begins. In this phase, the ventricular pressure decreases while keeping the ventricular volume constant at the end-systolic volume. Finally, the isovolumetric relaxation phase ends when the pressure drops below a specified threshold, and blood starts flowing from the atria to the ventricles while the ventricular volume returns to its initial value (filling phase). Illustrative frames acquired during a simulated cardiac cycle are shown in Figure 46.

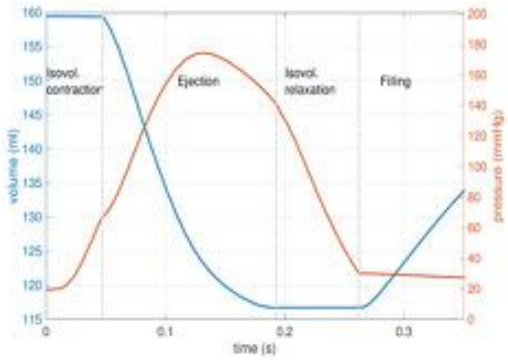

Figure 45. Time evolution of pressure and volume obtained for the ellipsoid electro-mechanical model. The cardiac cycle starts at a physiological end-diastolic volume of 160 ml. All four phases of the cardiac cycle are labelled in the figure.

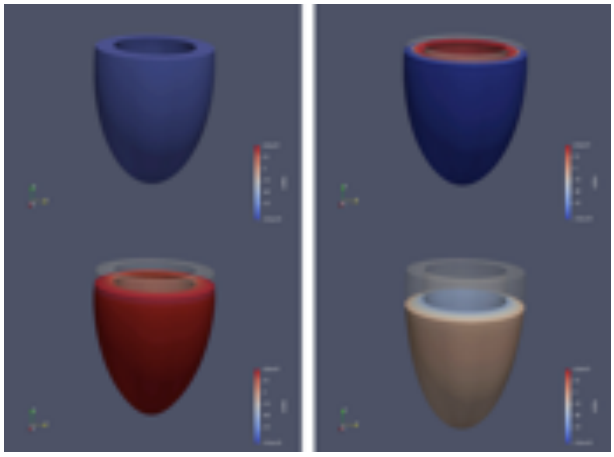

Figure 46. Representative frames captured during the simulation of a cardiac cycle (from left to right, top to bottom). The endocardium is homogeneously activated at the beginning of the cardiac cycle and, as the depolarization wave propagates through the myocardial wall, mechanical contraction is triggered.

Applications towards clinical translation

A common characteristic of the existing human cardiac models is that personalised geometries usually come from in-vivo imaging and the majority of computational meshes consider simplified ventricular geometries with smoothed endocardial (internal) surfaces, due to a lack of high resolution, fast and safe in-vivo imaging techniques. Acquiring human high-resolution images would mean for the patient to undergo long, expensive and impractical scans, in the case of magnetic resonance images (MRI), or

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

could present a risk for the patient's health, in the case of computed tomography (CT), since this process involves a considerable amount of radiation. Smoothed ventricular surfaces are indeed considered by the majority of existing human heart computational models, both when modelling blood flow dynamics and electrophysiology.

However the endocardial wall of human (and other mammals species) cardiac chambers is not smooth at all; it is instead characterised by endocardial sub-structures such as papillary muscles (PMs), trabeculations and false tendons (FTs). Additionally, fundamental anatomical gender differences can be found in cardiac sub-structural heart configuration as female hearts present less amount of FTs [57].
Since there is little information about the role of endocardial substructures in human cardiac function, considering them in the human in-silico cardiac simulations would present a first step towards the understanding of their function. Additionally, comparing simulations results including sub-structural anatomical information with those obtained when considering simplified human cardiac geometries (representing common existing models) would shed a light on the errors introduced when neglecting human endocardial sub-structures.

Another important aspect which is often ignored in in-silico simulations and could influence their outcome is gender phenotype. Female hearts have reduced resources for repolarization due to differences in K+ channels as compared to male phenotypes, leading to longer action potential durations (APDs) [58]. Longer APDs are consistent with clinical observation that females have longer QT intervals (time the heart takes to depolarize and repolarize) than males. Gender specificity can lead then to arrythmogenesis differences and so it may be important to consider different gender phenotypes when running in-silico electrophysiological simulations, in order to obtain results which are of clinical relevance and that can be compared to the subject-specific clinical data.

As shown in [61], we have created highly detailed human heart models from ex-vivo high-resolution MRI data, to study the role of cardiac sub-structures and gender phenotype in human cardiac physiology, through computational fluid dynamics (CFD) and electrophysiological high performance computing (HPC) simulations. The contributions this work can be summarised as follows:

- A pipeline of anatomically detailed cardiac volumetric mesh reconstruction was set up, starting from an *ex vivo* high-resolution human heart MRI database, as shown in Figure 47.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

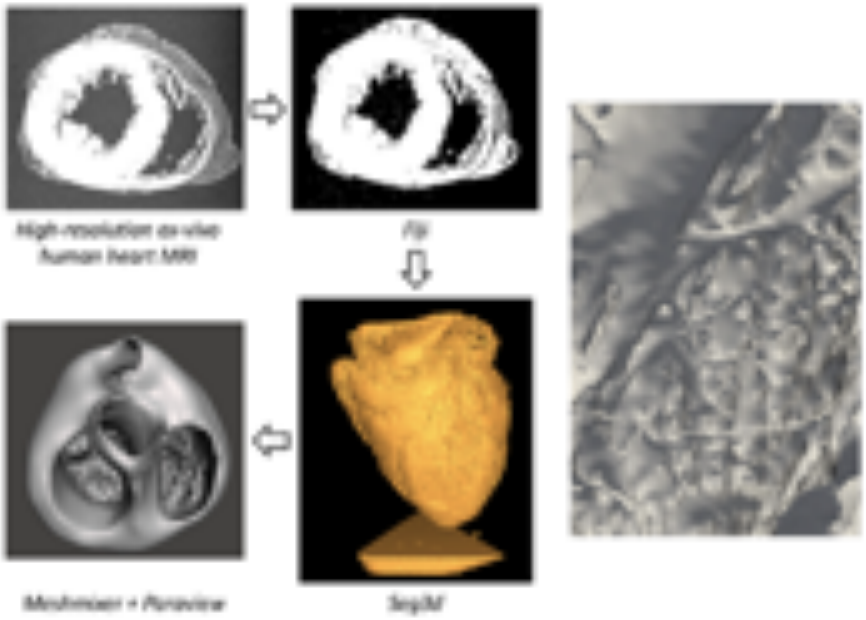

Figure 47. Volumetric mesh generation pipeline and a close up to the amount of endocardial detailed captured in the cardiac models [61].

- The impact of trabeculae and PMs on the blood flow within human left ventricular (LV) chambers was analysed using CFD simulations. This study demonstrated how the presence of trabeculae and PMs increase the intra-ventricular pressure drop, reduce the wall shear stress (WSS) and disrupt the main dominant single vortex, usually present in the smoothed endocardium models, generating secondary small vortices. Moreover, human female LVs were found to be less trabeculated than the male ones (Figure 48, Figure 49).

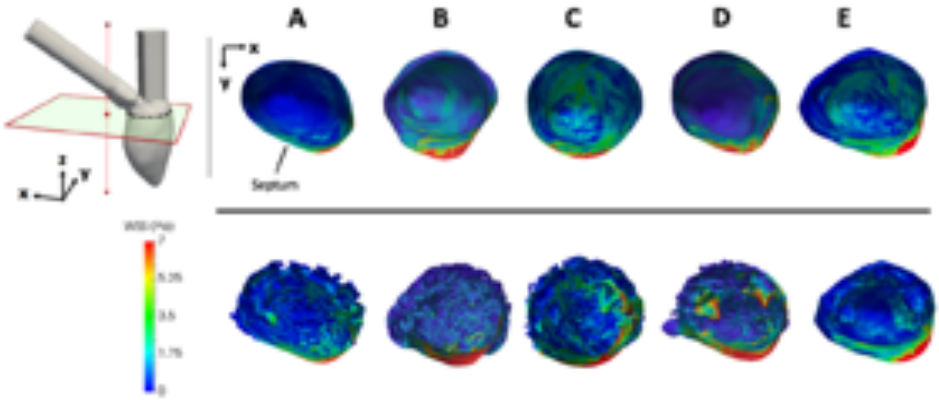

Figure 48. Magnitude of the wall shear stress (WSS) during constant inflow simulations on all cases. All geometries are clipped and aligned to the upper axes in the figure. The septum is pointed out for spatial reference [61].

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

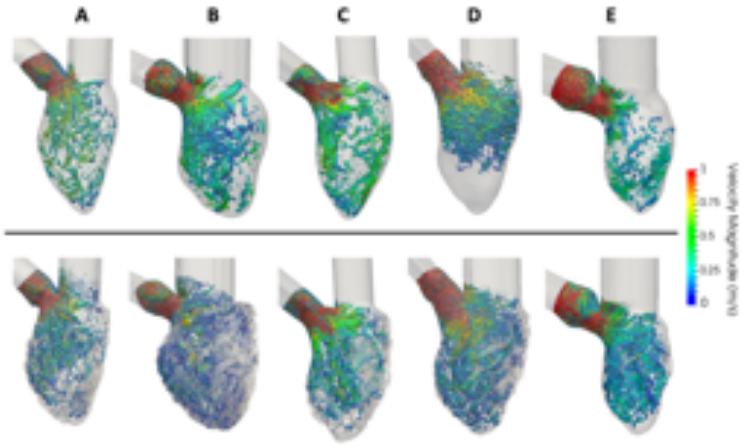

Figure 49. Vorticity estimated using the Q-criterion, thresholded at 5000s-2, for constant flow simulations in the LVs with smoothed (top) and detailed (bottom) geometries. Vortices are coloured by velocity magnitude [61].

- A methodology to incorporate the effect of trabeculations into smoothed ventricular geometries was proposed. By adding a porous layer along the LV endocardial walls, both the intra-ventricular pressure drops and the vorticity, observed in the detailed models, could be reproduced also within smooth-walled LV geometries (Figure 50).

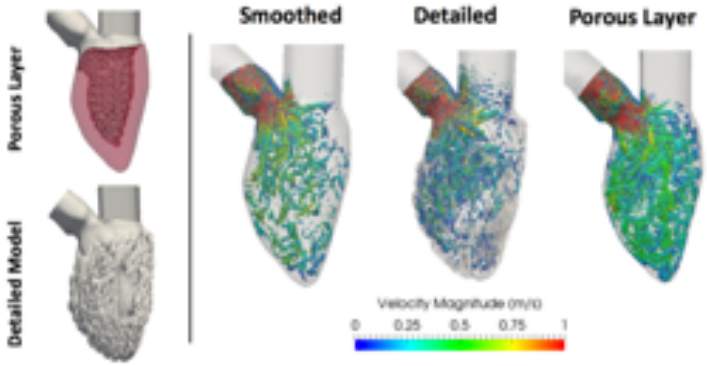

Figure 50. Left: Porous layer (in light red) on subject A and the corresponding detailed anatomy. Right: Vorticity estimated using the Q-criterion, thresholded at 5000 s.2, for constant inflow simulations with smoothed, detailed and smoothed with porous layer geometries. Vortices are coloured by velocity magnitude (m/s) [61].

- The effect of detailed endocardial structures on human right ventricular (RV) haemodynamics was analysed using CFD simulations. RV endocardial walls are even more trabeculated than the LV ones, but even less is known about the effect of the presence of the endocardial structures on RV haemodynamics. In this study, it was shown how detailed endocardial structures increase the degree of RV intra-ventricular pressure drop, decrease the WSS and disrupt the dominant vortex creating secondary small vortices. In addition, turbulent blood flow was observed within the detailed RV chambers. Moreover, human female

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

RV were less trabeculated and presented lower intra-ventricular pressure drops than the male ones (Figure 51, Figure 52).

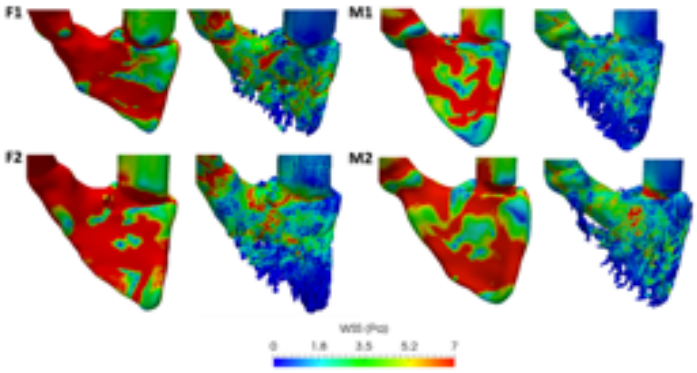

Figure 51. Magnitude of the wall shear stress (WSS) during constant inflow simulations tin RVs with smoothed (left) and detailed (right) geometry [61].

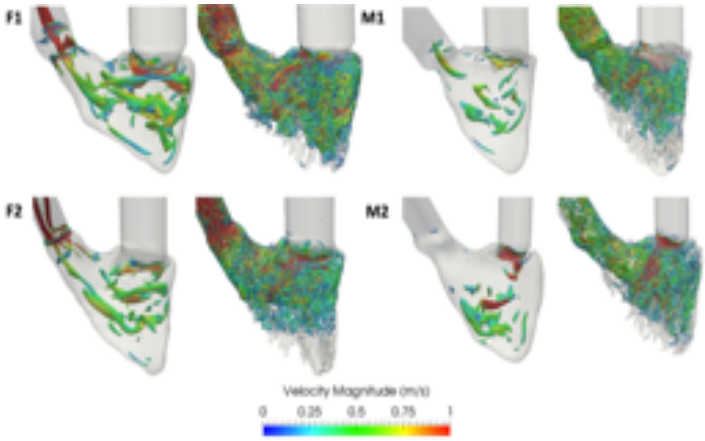

Figure 52. Vortex quantification using the Q-criterion thresholded at 9000 s-2 of constant inflow simulations in all RVs (female, F1 and F2; male, M1 and M2). Vortices are coloured by the velocity magnitude [m/s] [61].

- The influence of both highly detailed anatomical endocardial structures and gender phenotype on the electrophysiology of four biventricular, anatomically normal human heart models was investigated. Furthermore, a comparison to smoothed-endocardium geometries was done to quantify the errors introduced by neglecting such structures (Figure 53). Simulations showed a significant repolarization times increase in the detailed female phenotype cases, coinciding with the observed QT prolongation in the female hearts. Moreover, the simulations suggested that the absence of trabeculations reduces the total cardiac repolarization times due to different activation patterns in the smoothed cases. Finally, the presence of FTs shortcuts the signal propagation leading to faster total activation times

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

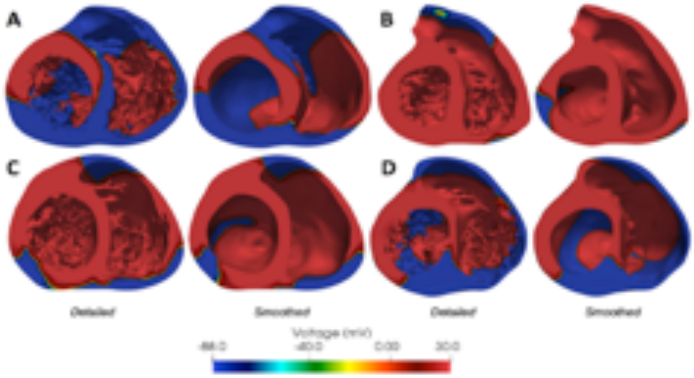

Figure 53. Electrophysiology simulation results: depolarization wave distribution at 60 ms, male phenotype [61].

Path to Code Validation.

Validation means that the simulation software is correctly reproducing the multiple physics of the question of interest for a determined context of use. This not only requires correctly solving the programmed model, but that the model effectively models the physics. To do so, experimental data is required to compare ex-vivo, in-vitro and in-vivo data against in-silico results. This stage requires a detailed description of the variables of the physical problem that is, in most of the cases, complex to obtain with a high accuracy.

As part of the collaboration with the Centro Nacional de Investigaciones Cardiovasculares (CNIC), the pathway to the validation of the cardiac model against experimental is being underway, as published in the thesis [60]. After a myocardial infarction, the affected areas of the cardiac tissue suffer changes in their electrical and mechanical properties. This post-infarction scar tissue has been related with a particular type of arrhythmia: ventricular tachycardia (VT). A thorough study on the experimental data acquired with clinical tools is presented in this thesis with the objective of defining the limitations of the clinical data towards predictive computational models. Computational models have a large potential as predictive tools for VT, but the verification, validation and uncertain quantification of the numerical results is required before they can be employed as a clinical tool.

Swine experimental data from an invasive electrophysiological study and Cardiac Magnetic Resonance imaging is processed to obtain accurate characterizations of the post-infarction scar. Based on the results, the limitation of each technique is described. Furthermore, the volume of the scar is evaluated as marker for post-infarction VT induction mechanisms.

A control case from the animal experimental protocol was employed to build a simulation scenario in which biventricular simulations were done using a detailed cell model adapted to the ionic currents present in the swine myocytes. The uncertainty of the model derived from diffusion and fibre orientation was quantified. Finally, the

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

recovery of the model to an extrastimulus is compared to experimental data by computationally reproducing an S1-S2 protocol.

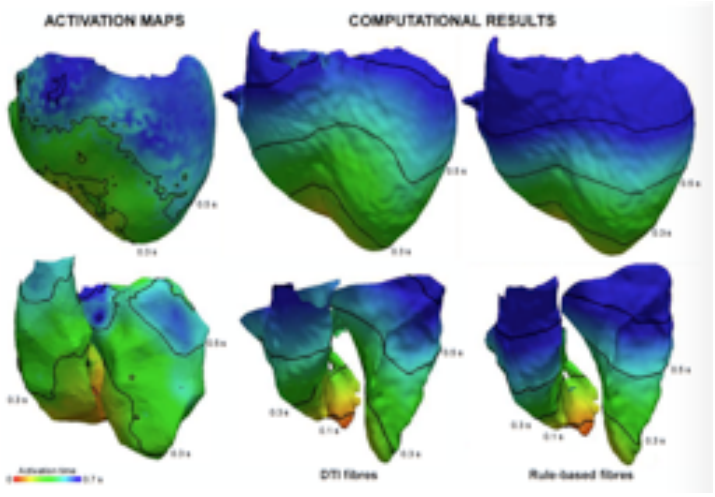

Figure 54. Activation maps including isochrones of the epicardium and the endocardium from experimental measurements and simulation data [60].

Results from the cardiac computational model show that the propagation wave patterns from numerical results very reasonably match the one described by the experimental activation maps if the DTI fibre orientations are used. As the electrophysiological activation is sensitive to fibre orientation, simulations including the fibre orientations from DTI are able to reproduce a physiological wave propagation pattern, as seen in Figure 54. The diffusion coefficients highly determine the conduction velocity. The S1-S2 protocol produced restitution curves that have similar slopes to the experimental curves.

This work is a first step forward towards validation of cardiac electrophysiology simulations. Future work will address the limitations about optimal parametrization of the O'Hara-Rudy cell model to fully validate the cardiac computational model for prediction of VT inducibility.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

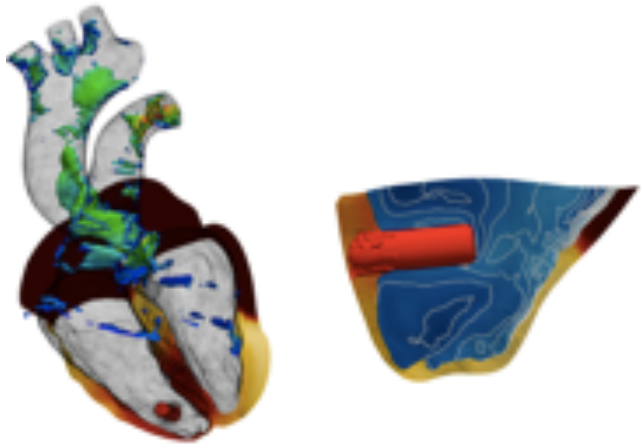

Figure 55. A leadless pacemaker implanted in the virtual heart. Q-criterion and velocity fields can be seen distorted in the soroundings of the implantation spot [40].

Path to medical device testing

During the last three years, through different collaborations, the area of device testing has been exploited in BSC's research group. Especially thanks to collaborations with the associate partner Medtronic, our model of the heart was used to study devices related with heart diseases such as pacemakers and stents (Figure 55, Figure 56). First we study widely known treatments for common pathologies, although not completely understood. These simulations can help to better understand the pathology and the treatment, or at least optimise the device set-up to maximise the performance. The stages of verification and validation are required to confidently translate these results to clinical applications.

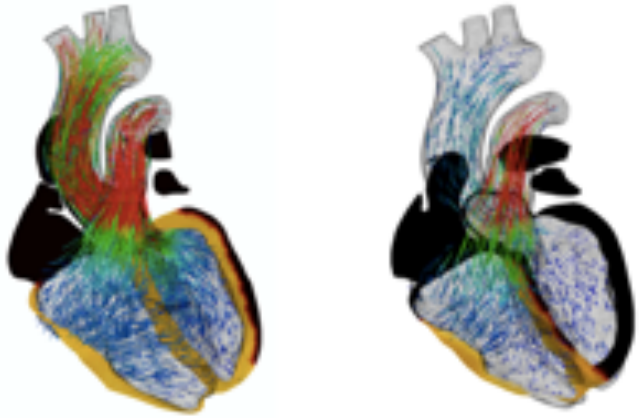

Figure 56. The left side shows a heart beating in normal conditions. The right side shows a heart beating under left bundle branch block. The elecrical dyssynchrony is easily seen. Also the reduction in the velocities on the fluid domain that will lead to a drop in the ejection fraction [40].

Summarizing what was done in ComBioMed, we developed a computational framework to conduct multiscale simulations of human cardiac fluid-electro-mechanics. Our models have been calibrated and validated at different scales against human experimental data, and they can be used to investigate drug-induced changes in ECG signal, EF and LVP and

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

medical devices action. Models can be personalised to be patient specific, incorporating anatomical information and also diseased conditions (e.g. fibrosis, ischemia, myocardial infarction).

Summary of collaborations started or enforced thanks to the project

- **Medtronic:** enforced, use of Alya model in leadless pacemakers.
- **Centro Nacional de Investigaciones Cardiovasculares (CNIC, Madrid)**: enforced, code validation against experiments and human data.
- **Hospital Sant Pau (Barcelona)**: enforced, coronary diseases.
- **Visible Heart Lab, University of Minnesota**: enforced, explanted human hearts data.

References for this section

[1]. Pappalardo, F., Russo, G., Tshinanu, F. M. & Viceconti, M. In silico clinical trials: concepts and early adoptions. Briefings in Bioinformatics (2018). doi:10.1093/bib/bby043

[2]. Carusi, A., Burrage, K. & Rodriguez, B. Bridging experiments, models and simulations: an integrative approach to validation in computational cardiac electrophysiology. Am J Physiol Heart Circ Physiol 303, H144-155 (2012).

[3]. O'Hara, T., Virág, L., Varró, A. & Rudy, Y. Simulation of the undiseased human cardiac ventricular action potential: Model formulation and experimental validation. PLoS Computational Biology 7, e1002061 (2011).

[4]. Dutta, S. et al. Early afterdepolarizations promote transmural reentry in ischemic human ventricles with reduced repolarization reserve. Progress in Biophysics and Molecular Biology 120, 236–248 (2016).

[5]. Dutta, S. et al. Optimization of an In silico Cardiac Cell Model for Proarrhythmia Risk Assessment. Frontiers in Physiology 8, 616 (2017).

[6]. Passini, E. et al. Mechanisms of pro-arrhythmic abnormalities in ventricular repolarisation and anti-arrhythmic therapies in human hypertrophic cardiomyopathy. Journal of Molecular and Cellular Cardiology 96, 72–81 (2016).

[7]. Tomek, J., Rodriguez, B., Bub, G. & Heijman, J. B-Adrenergic Receptor Stimulation Inhibits Proarrhythmic Alternans in Post-Infarction Border Zone Cardiomyocytes: a Computational Analysis. American Journal of Physiology - Heart and Circulatory Physiology ajpheart.00094.2017 (2017). doi:10.1152/ajpheart.00094.2017

[8]. Bers, D. M. Excitation-contraction coupling and cardiac contractile force. (Kluwer Academic Press, 2008).

[9]. Land, S. et al. A model of cardiac contraction based on novel measurements of tension development in human cardiomyocytes. Journal of Molecular and Cellular Cardiology 106, 68–83 (2017).

[10]. Zabel, M., Koller, B. S., Sachs, F. & Franz, M. R. Stretch-induced voltage changes in the isolated beating heart: importance of the timing of stretch and implications for stretch-activated ion channels. Cardiovascular research 32, 120–30 (1996).

[11]. Bernabeu, M. O., Wallman, M. & Rodriguez, B. Shock-induced arrhythmogenesis in the human heart: A computational modelling study. in 2010 Annual International

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Conference of the IEEE Engineering in Medicine and Biology 760–763 (IEEE, 2010). doi:10.1109/IEMBS.2010.5626338
[12]. Cardone-Noott, L., Bueno-Orovio, A., Mincholé, A., Zemzemi, N. & Rodriguez, B. Human ventricular activation sequence and the simulation of the electrocardiographic QRS complex and its variability in healthy and intraventricular block conditions. Europace 120, 179–188 (2016).
[13.]Okada, J. -i. et al. Screening system for drug-induced arrhythmogenic risk combining a patch clamp and heart simulator. Science Advances 1, e1400142–e1400142 (2015).
[14]. Drouin, E., Charpentier, F., Gauthier, C., Laurent, K. & Le Marec, H. Electrophysiologic characteristics of cells spanning the left ventricular wall of human heart: evidence for presence of M cells. Journal of the American College of Cardiology 26, 185–92 (1995).
[15]. Taggart, P. et al. Transmural repolarisation in the left ventricle in humans during normoxia and ischaemia. Cardiovascular research 50, 454–62 (2001).
[16]. Boukens, B. J. et al. Local transmural action potential gradients are absent in the isolated, intact dog heart but present in the corresponding coronary-perfused wedge. Physiological Reports 5, e13251 (2017).
[17]. Lyon, A. et al. Distinct ECG Phenotypes Identified in Hypertrophic Cardiomyopathy Using Machine Learning Associate With Arrhythmic Risk Markers. Frontiers in Physiology 9, 213 (2018).
[18]. Weiss, D. L. et al. Modeling of heterogeneous electrophysiology in the human heart with respect to ECG genesis. in 2007 Computers in Cardiology 49–52 (IEEE, 2007). doi:10.1109/CIC.2007.4745418
[19]. Bishop, M. J., Vigmond, E. J. & Plank, G. The functional role of electrophysiological heterogeneity in the rabbit ventricle during rapid pacing and arrhythmias. American Journal of Physiology-Heart and Circulatory Physiology 304, H1240–H1252 (2013).
[20]. Streeter, D. D., Spotnirz, H. M., Patel, D. P., Ross, J. & Sonnenblick, E. H. Fiber Orientation in the Canine Left Ventricle during Diastole and Systole. Circulation Research 24, (1969).
[21]. Taggart, P. et al. Inhomogeneous transmural conduction during early ischaemia in patients with coronary artery disease. Journal of molecular and cellular cardiology 32, 621–630 (2000).
[22]. Sutton, P. M. et al. Repolarisation and refractoriness during early ischaemia in humans. Heart (British Cardiac Society) 84, 365–9 (2000).
[23]. Pitt-Francis, J. et al. Chaste: A test-driven approach to software development for biological modelling. Computer Physics Communications 180, 2452–2471 (2009).
[24]. Gima, K. & Rudy, Y. Ionic current basis of electrocardiographic waveforms: A model study. Circulation Research 90, 889–896 (2002).
[25]. Martinez-Navarro, H., Mincholé, A., Bueno-Orovio, A. & Rodriguez, B. Septo-apical microreentrant pathways explain variability in arrhythmic risk in acute regional ischemia. Manuscript under review (2019).
[26]. Petersen, S. E. et al. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

population cohort. Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance 19, 18 (2017).
[27]. Clay, S. et al. Normal Range of Human Left Ventricular Volumes and Mass Using Steady State Free Precession MRI in the Radial Long Axis Orientation. Magnetic Resonance Materials in Physics, Biology and Medicine 19, 41–45 (2006).
[28]. Cain, P. A. et al. Age and gender specific normal values of left ventricular mass, volume and function for gradient echo magnetic resonance imaging: a cross sectional study. BMC medical imaging 9, 2 (2009).
[29]. Santiago, A. et al. Fully coupled fluid-electro-mechanical model of the human heart for supercomputers. International journal for numerical methods in biomedical engineering 34, e3140 (2018).
[30]. Levrero-Florencio, F. et al. Sensitivity analysis of a strongly coupled electromechanical cardiac model: effect of mechanical parameters on physiologically relevant biomarkers. (2019).
[31]. Passini, E. et al. Human In Silico Drug Trials Demonstrate Higher Accuracy than Animal Models in Predicting Clinical Pro-Arrhythmic Cardiotoxicity. Frontiers in Physiology 8, 1–15 (2017).
[32]. Morissette, P. et al. Shortening of the electromechanical window in the ketamine/xylazine-anesthetized guinea pig model to assess pro-arrhythmic risk in early drug development. Journal of Pharmacological and Toxicological Methods 81, 171–182 (2016).
[33]. Guns, P.-J., Johnson, D., Van Op den bosch, J., Weltens, E. & Lissens, J. The electro-mechanical window in anaesthetized guinea pigs: a new marker in screening for Torsade de Pointes risk. British Journal of Pharmacology 166, 689–701 (2012).
[34]. Guns, P.-J., Johnson, D. M., Weltens, E. & Lissens, J. Negative electro-mechanical windows are required for drug-induced Torsades de Pointes in the anesthetized guinea pig. Journal of Pharmacological and Toxicological Methods 66, 125–134 (2012).
[35]. van der Linde, H. J. et al. The Electro-Mechanical window: A risk marker for Torsade de Pointes in a canine model of drug induced arrhythmias. British Journal of Pharmacology 161, 1444–1454 (2010).
[36]. Passini, E. et al. Drug-induced Shortening of the Electromechanical Window is an Effective Biomarker for in Silico Prediction of Clinical Risk of Arrhythmias. (2019).
[37]. Tomek, J. et al. Development, Calibration and Validation of a Novel Human Ventricular Myocyte Model in healthy, Diseased, and Drug Block Conditions. Manuscript under review (2019).
[38]. Nguyen, N. et al. Adult human primary cardiomyocyte-based model for the simultaneous prediction of drug-induced inotropic and pro-arrhythmia risk. Frontiers in Physiology 8, 1073 (2017).
[39]. Goldberger, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet. Circulation 101, (2000).
[40] Santiago, A. (2018). Fluid-electro-mechanical model of the human heart for supercomputers. PhD Theses, UPC BarcelonaTech.
[41] Steven A Niederer, Eric Kerfoot, Alan P Benson, Miguel O Bernabeu, Olivier Bernus, Chris Bradley, Elizabeth M Cherry, Richard Clayton, Flavio H Fenton, Alan

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Garny, et al. Verification of cardiac tissue electrophysiology simulators using an n-version benchmark. Phil. Trans. R. Soc. A, 369(1954):4331–4351, 2011.
[42] Mark Potse, Bruno Dubé, Jacques Richer, Alain Vinet, and Ramesh M Gulrajani. A comparison of monodomain and bidomain reaction diffusion models for action potential propagation in the human heart. IEEE Transactions on Biomedical Engineering, 53(12):2425–2435, 2006.
[43] Bradley J Roth. Meandering of spiral waves in anisotropic cardiac tissue. Physica D: Nonlinear Phenomena, 150(1-2):127–136, 2001.
[44] P Colli Franzone, LF Pavarino, and B Taccardi. Simulating patterns of excitation, repolarization and action potential duration with cardiac bidomain and monodomain models. Mathematical biosciences, 197(1):35–66, 2005.
[45] André G Kléber and Yoram Rudy. Basic mechanisms of cardiac impulse propagation and associated arrhythmias. Physiological reviews, 84(2):431–488, 2004.
[46] Dirk Durrer, R Th Van Dam, GE Freud, MJ Janse, FL Meijler, and RC Arzbaecher. Total excitation of the isolated human heart. Circulation, 41(6):899–912, 1970.
[47] RH Clayton, Olivier Bernus, EM Cherry, Hans Dierckx, FH Fenton, L Mirabella, AV Panfilov, Frank B Sachse, G Seemann, and H Zhang. Models of cardiac tissue electrophysiology: progress, challenges and open questions. Progress in biophysics and molecular biology, 104(1-3):22–48, 2011.
[48] Joakim Sundnes, Glenn Terje Lines, Xing Cai, Bjørn Frederik Nielsen, Kent-Andre Mardal, and Aslak Tveito. Computing the electrical activity in the heart, volume 1. Springer Science & Business Media, 2007.
[49] Jinichi Nagumo, Suguru Arimoto, and Shuji Yoshizawa. An active pulse transmission line simulating nerve axon. Proceedings of the IRE, 50(10):2061–2070, 1962.
[50] Go W Beeler and H Reuter. Reconstruction of the action potential of ventricular myocardial fibres. The Journal of physiology, 268(1):177–210, 1977.
[51] Ching-hsing Luo and Yoram Rudy. A model of the ventricular cardiac action potential. depolarization, repolarization, and their interaction. Circulation research, 68(6):1501–1526, 1991.
[52] Flavio Fenton and Alain Karma. Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: Filament instability and fibrillation. Chaos: An Interdisciplinary Journal of Nonlinear Science, 8(1):20–47, 1998.
[53] Jeffrey J Fox, Jennifer L McHarg, and Robert F Gilmour Jr. Ionic mechanism of electrical alternans. American Journal of Physiology- Heart and Circulatory Physiology, 282(2):H516–H530, 2002.
[54] KHWJ Ten Tusscher, Denis Noble, Peter-John Noble, and Alexander V Panfilov. A model for human ventricular tissue. American Journal of Physiology-Heart and Circulatory Physiology, 286(4):H1573–H1589, 2004.
[55] Alfonso Bueno-Orovio, Elizabeth M Cherry, and Flavio H Fenton. Minimal model for human ventricular action potentials in tissue. Journal of theoretical biology, 253(3):544–560, 2008.
[56] Thomas O'Hara, László Virág, András Varró, and Yoram Rudy. Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. PLoS computational biology, 7(5):e1002061, 2011.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

[57] P. H. Luetmer, W. D. Edwards, J. B. Seward, et al. “Incidence and Distribution of Left Ventricular False Tendons: an Autopsy Study of 483 Normal Human Hearts”. In: J. Am. Coll. Cardiol. (1986).
[58] N. Gaborit, A. Varro, S. Le Bouter, et al. “Gender-related Differences in Ion-channel and Transporter Subunit Expression in Nondiseased Human Hearts”. In: J. Mol. Cell. Cardiol. (2010).
[59] Stephenson RS, Atkinson A, Kottas P, et al. High resolution 3-Dimensional imaging of the human cardiac conduction system from microanatomy to mathematical modeling. *Sci Rep*. 2017;7(1):7188. Published 2017 Aug 3. doi:10.1038/s41598-017-07694-8
[60] López-Yunta, M. (2018). Multimodal ventricular tachycardia analysis: towards the accurate parameterization of predictive HPC electrophysiological computational models. UPC BarcelonaTech.
[61] Sacco, F (2019). Quantification of the influence of detailed endocardial structures on human cardiac hemodynamics and electrophysiology using HPC. Universitat Pompeu Fabra & UPC BarcelonaTech.
[62] Gil, D., Aris, R., Borras, A., Ramírez, E., Sebastian, R., & Vazquez, M. (2019). Influence of fiber connectivity in simulations of cardiac biomechanics. International journal of computer-assisted radiology and surgery, 14(1), 63-72.
[63] López-Yunta, M., León, D. G., Alfonso-Almazán, J. M., Marina-Breysse, M., Quintanilla, J. G., Sánchez-González, J., ... & Ibáñez, B. (2018). Implications of bipolar voltage mapping and magnetic resonance imaging resolution in biventricular scar characterization after myocardial infarction. Ep Europace, 21(1), 163-174.
[64] León, D. G., López-Yunta, M., Alfonso-Almazán, J. M., Marina-Breysse, M., Quintanilla, J. G., Sánchez-González, J., ... & Pérez-Villacastín, J. (2019). Three-dimensional cardiac fiber disorganization as a novel parameter for ventricular arrhythmia stratification after myocardial infarction. EP Europace, 21(5), 822-832.
[65] Santiago, A., Aguado-Sierra, J., Zavala-Aké, M., Doste-Beltran, R., Gómez, S., Arís, R., ... & Vázquez, M. (2018). Fully coupled fluid-electro-mechanical model of the human heart for supercomputers. International journal for numerical methods in biomedical engineering, 34(12), e3140.
[66] Sacco, F., Paun, B., Lehmkuhl, O., Iles, T. L., Iaizzo, P. A., Houzeaux, G., ... & Aguado-Sierra, J. (2018). Evaluating the roles of detailed endocardial structures on right ventricular hemodynamics by means of CFD simulations. International journal for numerical methods in biomedical engineering, 34(9), e3115.
[67] Vázquez, M., Houzeaux, G., Koric, S., Artigues, A., Aguado-Sierra, J., Arís, R., ... & Taha, A. (2016). Alya: Multiphysics engineering simulation toward exascale. Journal of computational science, 14, 15-27.
[68] Sacco F., Paun B., Lehmkuhl O., Iles T.L., Iaizzo P.A., Houzeaux G., Vázquez M., Butakoff C., Aguado-Sierra J., Left ventricular trabeculations decrease the wall shear stress and increase the intraventricular pressure drop in CFD simulations, Frontiers in Physiology, 2018
[69] Sacco F., Doste R., Bederián C., Iles T.L., Iaizzo P.A., Houzeaux G., Vázquez M., Camara O., Butakoff C., Aguado-Sierra J., Systematic electrophysiological analysis of

*“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **675451**“*

four biventricular, anatomically normal human heart models, accepted for submission to Interface Focus of The Royal Society, 2019.

Molecularly-based Medicine Exemplar Research

Universitat Pompeu Fabra: Machine Learning and Molecular Dynamics

Pharmaceutical industry is facing an unprecedented challenge nowadays. Introducing a new drugs into the market involves a 15-year-long process and billions of dollars in investment, yet the success rate is pretty low (Figure 57). The probability that a candidate drug in Phase I clinical trials ends up being approved is around 7% [19]. It is in everybody's interest to keep drug discovery as a sustainable model, and therefore it is required to reduce its overall cost, speed up the discovery process and improve the success rates.

Figure 57. General scheme of the entire drug discovery process and all the stages required to release new medicines into the market.

Our primary objective within the CompBiomed project has been to develop novel computational methods for the early stages in the drug discovery pipeline in order to accelerate the obtention of drug candidates and reduce the experimental workload and its associated costs. We advance towards the next generation of drug discovery, which relies on computational predictive models that are able to test millions of compounds *in silico*, giving accurate and precise results and reducing the amount of experimental tests needed on the design process. Furthermore, having computational methods that are powerful enough opens up the possibility to drastically reduce (and, on the long term, even remove) animal experiments, a fundamental requirement for a more sustainable and ethically responsible drug discovery process.

Designing novel drugs is a complex process where multiple parameters have to be optimized, so as the designed compound can be administered to the body and bind strongly to a specific protein target. If the compound will bind with the target is dependent on laws of physics, both on atomic and subatomic scale. As there are many factors affecting the interaction, computational modeling is hard and experimental validation is the most reliable source of data.

However, with the recent advances in artificial intelligence and deep learning we can leverage the data and use it for novel predictions. Particularly, deep learning can be

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

applied to extract complex patterns from simple representations. In our work, we leverage deep learning methods to extract patterns from three-dimensional representations of molecules and proteins. We developed models inspired by computer vision architectures, where both protein and the ligand are divided in a three dimensional grid with features representing different atomic properties, such as hydrophobicity or aromaticity. Deep learning models have already demonstrated an incredible performance (sometimes even above human) in several different tasks, such as image recognition, natural language processing or audio recognition. The next step is drug discovery. Deep learning models can perform *in silico* predictions based on *in vitro/in vivo* experimental data in a matter of seconds, which would directly translate into a faster discovery rate for potential drug candidates.

The efficiency of deep learning methodologies strongly rely on data. For an optimal performance, deep learning models require extensive amounts of data to be trained on. More often than not, the available data is scarce, especially on structural biology. Therefore, to counter this, we have also been working with physics-based methods, such as molecular dynamics (MD) simulations, which do not suffer from strong data dependency. Molecular dynamics is a multi-scale simulation method that simulates biologically relevant molecular systems, such as protein ligand binding scenarios, with atomic resolution. MD simulates atom movements by using Newton's equations of motion, scaling it up to all the atoms in a molecular system. The analysis of biomolecular simulations delivers accurate predictions on drug properties like binding affinities or kinetics. The main drawback of MD simulations is their high computational cost, incompatible with the large molecule screenings required during a drug discovery process [20]. We have working developing novel algorithms, inspired by the fields of reinforcement learning and active learning, to reduce the computational resources needed for MD simulations.

Our research has mainly been focused on applying machine learning methods and Molecular Dynamics (MD) simulations to improve and speed up the drug discovery pipeline. The computational tools we have developed throughout the project are aimed to solve common problems and tasks encountered during the early stages of drug discovery, such as protein binding site prediction, protein-ligand binding affinity prediction, drug selectivity elucidation and molecular generative models. We have also developed some novel tools for ligand generation, based on the recent improvements in generative modelling, as an attempt to increase the automatization of drug design. The research performed by UPF inside the CompBioMed project has materialized in several applications, all made freely available through PlayMolecule or HTMD.

All these novel solutions came directly from bringing the recent advances in machine learning and Deep Convolutional Neural Network (DCNN) for image recognition, which demonstrated superhuman performance on image recognition tasks, into structural biology and molecular medicine. Designing novel drugs is a complex process where multiple parameters have to be optimized, so as the designed compound can be administered to the body and bind strongly to a specific protein target. If the compound

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

will bind with the target is dependent on laws of physics, both on atomic and subatomic scale. As there are many factors affecting the interaction, computational modeling is hard and experimental validation is the most reliable source of data. However, with machine learning methods we can leverage the data and use it for novel predictions. Particularly, deep learning can be applied to extract complex patterns from simple representations. In this work we leverage deep learning to extract patterns from three-dimensional representations of molecules. We developed models inspired by computer vision architectures, where both protein and the ligand are featurized in three dimensions using a grid with features representing different atom types or properties, such as hydrophobicity or aromaticity.

DeepSite

Given the recent success of DCNNs models in several computer vision applications, we started investigating the potential of said models in computational biology and chemistry applications. In a similar way to images, which are two dimensional arrays, usually represented by 3 channels (red, green and blue), we represented a protein / small molecule structure by 3 dimensional arrays, each corresponding to a different pharmacophoric property (i.e aromatic, hydrophobic, ionization…) or atom type.

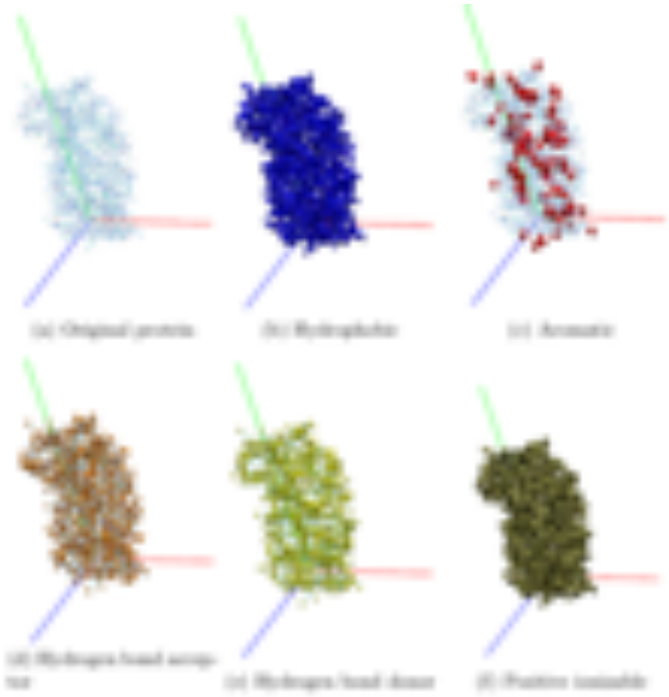

Figure 58. Depiction of the defined protein channels corresponding to different parmacophoric properties.

One of the first applications using this representation as well as DCNN models was DeepSite [1], a binding site predictor (Figure 58). DeepSite is essentially a classifier trained in the scPDB [2] database. The protein structure is split into 16^3 Å boxes, and then for each we predict whether each one is within 4 Å distance of the annotated

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

binding site. Essentially the procedure gives an entire probabilistic map of the protein, which can be then clustered in order to give point predictions. The project showed early promising results, surpassing the performance of several well-known algorithms.

KDeep

After DeepSite, we focused more on classical drug-discovery problems. In particular, we decided to concentrate our efforts in developing KDeep [3], a deep learning based protein-ligand scoring function (Figure 59). Given the structure of a protein binding site and the corresponding pose of a ligand, we featurize both in a similar way to the previous application, in a fixed sized 24^3 Å box. Once these are computed, the goal is then to build a DCNN-based regression model. We chose PDBbind v.2016 [4] as the main database to train and test our model, since several groups have previously used it as a benchmark in the development of several scoring functions. With KDeep we show that DCNN like models achieve state of the art performance in absolute binding affinity prediction, by testing both on PDBbind, as well as several CSAR and external datasets. In particular, we show that current scoring functions tend to perform well when the ligands considered in a particular set belong to different enough chemical scaffolds, but performance significantly drops when predicting congeneric-like series. In fact, the proposed model has also been tested in the 4th D3R Grand Challenge, obtaining winning solutions for two out of three affinity ranking subchallenges.

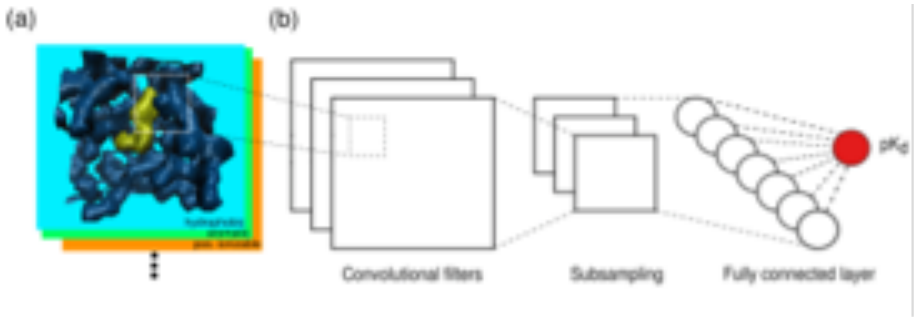

Figure 59. Network architecture description of KDeep.

BindScope

While developing KDeep, we realized that, because it was trained only over bound ligands, it was unable to discriminate between active or inactive ligands. For a complete virtual screening of large small molecule libraries, one needs to be able to discard inactive molecules on their target. Therefore, we developed a new deep learning-based tool for this task, BindScope [5]. The featurization is very similar to the one proposed in KDeep, but their goal is to discriminate between active and inactive ligands for a particular target using the DUDE database [6] of decoys (inactive ligands docked into protein structures). The presence of decoys in the training dataset provides BindScope the capacity to separate active ligands on a specific target from inactive ones, which KDeep is incapable of. The combination of BindScope and KDeep allows researchers to screen large small molecule libraries on a specific protein target, and not only select the active molecules, but also rank them according to their binding affinity.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

DeltaDelta

While we previously had evaluated our models using public databases such as PDBbind, it was clear, after talking to pharmaceutical companies, that their main use for scoring functions such as the ones we were proposing was their application in congeneric series ranking. A congeneric series is a set of molecules with small modifications, with the intent of improving them with respect to some pharmacokinetic parameter. However, the public databases our models were trained on featured a diverse set of protein-ligand complexes, making our models inadequate for predicting the small differences in affinity that were expected in the congeneric series setting.

With that focus, we collaborated with several pharmaceutical companies, such as Janssen, Biogen and Pfizer, to both train and test DeltaDelta, a relative binding affinity predictor to be used on congeneric series. Our results show that said models can accurately interpolate between the binding affinities of a series using only a small fraction of the ligands as training data. Furthermore, they are significantly better than widely popular alternatives such as the empirical scoring functions provided by several docking protocols. Finally, it was shown that the performance of modern deep-learning alternatives can come close to rigorous simulation-based protocols such as free energy perturbation [7], at a fraction of the computational cost, and avoiding known problems such as the treatment of waters or ligand parameterization.

LigVoxel / LigDream / LiGANN

Our work on generative models led to the development of other two applications that we believe are of use in several drug-discovery pipelines. The first one is named LigVoxel [8], a DCNN-based ligand pharmacophore predictor (Figure 60). Contrary to our previous approaches, which output a single value for each structure subarray, this model is trained on pocket voxels and its output is its entire corresponding ligand subarrays, of the same size as the input. This model is able to reproduce important ligand chemical properties, such as hydrogen bonds or aromaticity, making it a valuable tool in structure based drug and pharmacophore design. Several tests were carried out in order to validate the output of the model on several hundreds of unseen structures, showing that in fact the proposed model produces sensible predictions.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

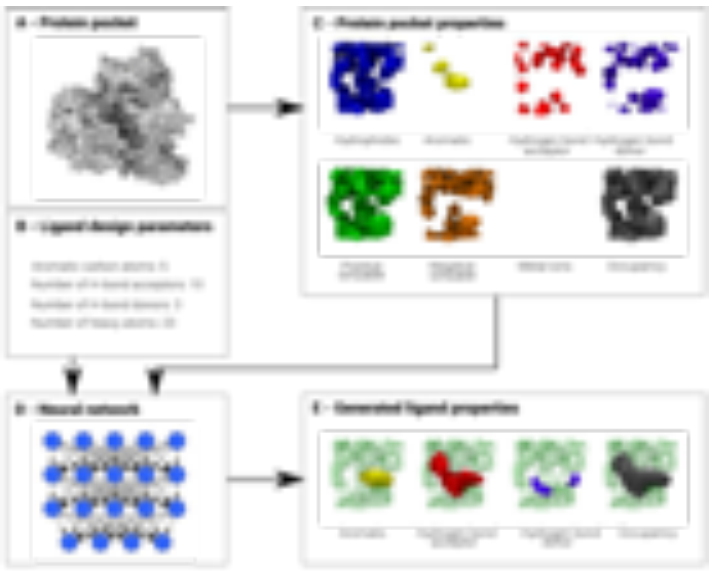

Figure 60. Workflow description for LigVoxel. The protein pocket is featurised with the same pharmacophoric channels used in previous tools, and the network outputs suggested pharmacophoric areas for a ligand bound into the pocket.

Next one, LigDream [9], is a shape decoding tool that decodes a voxelied molecule representation into SMILES strings (Figure 61). The model has been trained on compounds from drug-like ZINC15 [10] database. 3D conformations for the seed ligands are generated, followed by the vozel featurization used in our previous tools . Finally, three-dimensional convolutional and recurrent neural networks are used to to generate sequence of SMILES strings. Variability in SMILES outputs is obtained through distortion of voxelized representation and probabilistically sampling next possible SMILES token.

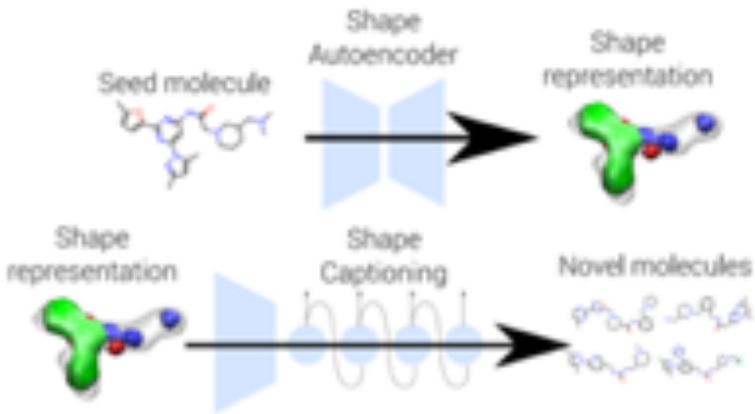

Figure 61. Workflow description for LigDream. A shape representation of the seed molecule is created through a generative model, akin to the pharmacophoric description used in LigVoxel. Afterwards, the shape captioning network is used to generate SMILES sequences according to the same shape representation.

Finally, LiGANN integrates both works into a structure-based de novo drug design tool that produces ligand structures based on protein pocket structures. Given a protein

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

shape, a generative adversarial network produces complementary ligand shapes (similar concept as Ligvoxel) in a multimodal fashion. Finally, a shape captioning network (LigDream) decodes the ligand shapes into SMILES strings.

PathwayMap

Early phases of drug-discovery are critical, since compounds initially deemed as interesting can later show to be problematic, due to lack of affinity towards a target, toxicity, or unspecificity. In this work we show the development of PathwayMap[11], a deep-learning-based model that uses self-normalizing neural networks for predicting the association of a compound towards a particular pathway, with the hopes of helping to reduce high attrition rates in drug discovery.

We use the entire ChEMBL database of compounds [12] as well as the KEGG [13] and Reactome [14] pathway databases, also relying on information from Uniprot. This study, is to the best of our knowledge the most extensive one in terms of dataset size and evaluation, since most previous study ignore the multifunction compound problem. Neural networks can naturally tackle these in a multitask function, providing models that can be trained in hours and deployed with ease. Models were evaluated using a rigorous cross-validation procedure, using both random and scaffold-based splits. A collaboration with the pharmaceutical company Novartis also provided an external dataset in which to train and test our models using an even more realistic temporal split. Results suggest that the models can satisfactorily predict molecular pathway association, and we exemplify one of the uses of such models by identifying dark chemical matter, defined as compounds that seem lack affinity to a lot of targets.

MD simulations and Machine learning

Molecular simulation methods have always been hampered by sampling limitations over such distribution due to their computational cost. The advent of GPUs and GPU molecular dynamics software [15] was a notable improvement, greatly increasing the computational efficiency of simulations. This, combined with Markov state models (MSMs) [16] allowed to reconstruct a complete statistical description of the full dynamical system from many shorter trajectories, obtaining a description that is equivalent to reversible sampling, once at convergence.

Running not one, but hundreds or thousands of simulation trajectories created a new opportunity to decide the starting conditions of these simulations in order to obtain the best equilibrium characterization at the minimal computational cost, i.e. adaptive sampling. Initially, adaptive sampling algorithms [17] were used to reduce statistical uncertainty by choosing conformations that contributed the most to the error in mean first passage time of a MSM, eigenvalues and eigenvectors , or choosing low state populations. In general the adaptive sampling policy was always empirical, not based on any mathematical decision process, even though the problem has been recognized as a multi-armed bandit problem before.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

We have designed AdaptiveBandit, an adaptive sampling algorithm that casts the problem as a multi-armed bandit problem and uses an action value function and an upper confidence bound selection algorithm [18] to improve the performance of adaptive sampling and increase its versatility when faced with different free energy landscapes. AdaptiveBandit formally introduces adaptive sampling algorithms inside a multi-armed bandit frame-work and builds upon it to deliver a novel algorithm with increased performance and flexibility across different energy landscapes. AdaptiveBandit has showed that is able to perform equally or better than previous adaptive sampling algorithms in a diverse set of systems, and has proved its power to learn from the environment to modify its behaviour for an optimal performance. The bandit framework described can also be used in future works to develop novel algorithms based on theoretical work, instead of using simple heuristic policies.

To summarize the collaboration impact of CompBioMed for the UPF, the research performed by UPF has been strongly reenforced by industrial collaborators, comprising CompBioMed Core partners as well as institutions outside the project.
First and foremost, the entire research performed has been done in strong collaboration with Acellera. We have worked together to deliver our applications through a general and accessible platform, PlayMolecule. Our collaboration involved testing and deployment of our applications, making sure that our research and developments do not remain only as publications, but that they also produce solid tools accessible to everyone, including academia, industry and clinicians.
Our industrial collaborations are not limited to Acellera, and we have also been working with another industrial Core partner inside CompBioMed, Jansen, on testing our application, DeltaDelta, with their internal datasets. The same was done with other external industrial partners, such as Pfizer, Novartis and Biogen.

Regarding the commercial impact, most of the developed applications are available either through PlayMolecule or HTMD, solutions provided by Acellera, one of the core partners. Both solutions are freely available, but Acellera also provides personalized customer support on both of them as a service, in addition to custom deployments of PlayMolecule. Therefore, the developed applications have directly improved the quality of the products and services Acellera has to offer.

References for this section

[1] Jiménez, J., Doerr, S., Martínez-Rosell, G., Rose, A. S., & De Fabritiis, G. (2017). DeepSite: Protein-binding site predictor using 3D-convolutional neural networks. Bioinformatics, 33(19), 3036–3042.
[2] Desaphy, J., Bret, G., Rognan, D., & Kellenberger, E. (2014). sc-PDB: a 3D-database of ligandable binding sites—10 years on. Nucleic acids research, 43(D1), D399-D404.
[3] Jiménez, J., Škalič, M., Martínez-Rosell, G., & De Fabritiis, G. (2018). KDEEP: Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. Journal of Chemical Information and Modeling, 58(2), 287–296.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

[4] Liu, Zhihai., Su, Minyi., Han, Li., Liu, Jie., Yang, Qifan., Li, Yan., Wang, Renxiao. (2017). Forging the Basis for Developing Protein-Ligand Interaction Scoring Functions", Accounts of Chemical Research, 50 (2): pp. 302-309.
[5] Skalic, M., Martínez-Rosell, G., Jiménez, J., Gianni De Fabritiis. (2019). PlayMolecule BindScope: large scale CNN-based virtual screening on the web, Bioinformatics, 35 (7), 1237–1238
[6] Mysinger, M. M., Carchia, M., Irwin, J. J., & Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. Journal of medicinal chemistry, 55(14), 6582-6594.
[7] Cournia, Z., Allen, B., & Sherman, W. (2017). Relative binding free energy calculations in drug discovery: recent advances and practical considerations. Journal of chemical information and modeling, 57(12), 2911-2937.
[8] Skalic, M., Varela-Rial, A., Jiménez, J., Martínez-Rosell, G., & De Fabritiis, G. (2018). LigVoxel: inpainting binding pockets using 3D-convolutional neural networks. Bioinformatics, 35(2), 243-250.
[9] Skalic, M., Jiménez, J., Sabbadin, D., & De Fabritiis, G. (2019). Shape-Based Generative Modeling for de Novo Drug Design. Journal of chemical information and modeling, 59(3), 1205-1214.
[10] Sterling, T., & Irwin, J. J. (2015). ZINC 15–ligand discovery for everyone. Journal of chemical information and modeling, 55(11), 2324-2337.
[11] Jiménez, J., Sabbadin, D., Cuzzolin, A., Martínez-Rosell, G., Gora, J., Manchester, J., ... & De Fabritiis, G. (2018). PathwayMap: Molecular pathway association with self-normalizing neural networks. Journal of chemical information and modeling, 59(3), 1172-1181.
[12] Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., ... & Davies, M. (2016). The ChEMBL database in 2017. Nucleic acids research, 45(D1), D945-D954.
[13] Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1), 27-30.
[14] Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., ... & Jassal, B. (2013). The Reactome pathway knowledgebase. Nucleic acids research, 42(D1), D472-D477.
[15] Harvey, M. J., Giupponi, G., & Fabritiis, G. D. (2009). ACEMD: accelerating biomolecular dynamics in the microsecond time scale. Journal of chemical theory and computation, 5(6), 1632-1639.
[16]Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., ... & Noé, F. (2011). Markov models of molecular kinetics: Generation and validation. The Journal of chemical physics, 134(17), 174105.
[17] Doerr, S., & De Fabritiis, G. (2014). On-the-fly learning and sampling of ligand binding by high-throughput molecular simulations. Journal of chemical theory and computation, 10(5), 2064-2069.
[18] Auer, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. Journal of Machine Learning Research, 3(Nov), 397-422.
[19] Dowden, H., & Munro, J. (2019). Trends in clinical success rates and therapeutic focus. Nat. Rev. Drug Discov, 18, 495-496.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

[20] Pérez, A., Martínez-Rosell, G., & De Fabritiis, G. (2018). Simulations meet machine learning in structural biology. Current opinion in structural biology, 49, 139-144.

University College London: Supercomputers and binding affinities

Drug development is a lengthy, complex, and costly process (it is estimated that and average of ~€2.2 billion is required to get a drug into the clinic [13]), and involves a high degree of uncertainty that any given candidate will actually succeed. It is increasingly recognised that this is compounded by the variation in response between patients, implying that we can no longer hope to produce "blockbuster" one-size-fits-all drugs for the entire global population [14]. Consequently, new approaches are required that facilitate better targeted treatments for subsets of patients. Our goal is to support this endeavour by developing simulation techniques that allow us to understand how drugs interact with their target proteins and how genetic variation can affect this.

One example of where such knowledge could be applied in a clinical setting comes from anti-cancer therapy. The rapid drop in cost of next-generation sequencing technologies has led many cancer centres to begin deep sequencing patient tumours to identify the genetic alterations driving individual cancers, with the ultimate goal of making individualized therapeutic decisions based upon this data — an approach termed "precision" cancer therapy. This is approach is attractive as resistance to therapeutics is responsible for more than 90% of deaths in patients with metastatic cancer [15]. While drug resistance can emerge via multiple mechanisms, small changes (mutations) in the drug target protein are responsible for resistance in many patients; for some drug targets this mechanism causes as many as 90% of resistant cases [16]. While several common (recurrent) mutations have been catalogued due to their ability to induce resistance or susceptibility to particular kinase inhibitors, the vast majority of clinically observed mutations are rare, making catalogue-building alone insufficient for making fully informed decisions about the majority of individual patient tumours. There are two major strategies for countering this threat to treatment efficacy: tailoring the drug regimen received by a patient according to the mutations present in their particular cancer, and development of more advanced second- or third-line therapies that retain potency for known resistance mutations. In both cases, future developments require insight into the molecular changes produced by mutations, as well as ways to predict their impact on drug binding on a timescale much shorter than is typically experimentally feasible. Our research is designed to provide the foundations for both these approaches in the future. By basing our approach on physical models of the chemical processes involved our technology should be generally applicable, meaning that it is not limited to just the specific application areas we target initially.

The binding affinity calculator (BAC) software developed within the molecular medicine strand of CompBioMed is at the heart of a research programme which aims to influence both industrial and clinical workflows. The common approach that under pins these goals is the generation of computational protocols which provide reproducible binding free energy estimates for ligand binding from molecular simulation. Our research has

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

sought to address this based upon a theoretical understanding of the utility of ensemble simulations to provide efficient sampling and meaningful uncertainty quantification [1]. This has led us to develop a suite of computational protocols which we call ESMACS (enhanced sampling of molecular dynamics with approximation of continuum solvent) and TIES (thermodynamic integration with enhanced sampling) [2]. The former based on the use of computationally inexpensive end point calculations and the latter more expensive but accurate "alchemical" binding free energy calculations.

On a practical level we have applied this insight to looking into a diverse range of drug discovery targets with the aim of identifying the reasons for differing levels of success for existing methods when using datasets involving different proteins or ligands from different regions of chemical space (Figure 62). Much of this work has been conducted in a collaborative manner with the UCL team who developed BAC working with domain scientists from industrial associate partners Janssen, GSK and Pfiszer [3, 4, 5, 6]. This work has led to now protocols based on the use of different estimates of entropic components in end point calculations and replicas in alchemical free energy methods. Most recently we concluded a study applying our methods to problems in fragment-based lead generation (FBLG) [7], providing computationally efficient extensions to the commonly used MMPBSA free energy calculation approach that allow it to be applied to drugs of varying charge, molecular weight and target binding site. These interactions have led the production of uf-BAC, a user friendly workflow tool designed to allow pharamaceutical collaborators to run BAC protocols on cloud resources. Uf-BAC is supported and developed by the spin out company (and CompBioMed associate partner) EnsembleMD.

Alongside our work focussed on drug discovery problems we have investigated the influence of mutations on drug binding, an issue of direct relevance to the effectiveness of therapies in individual patients. In particular we have develop a variety of our TIES protocol which computes the difference in binding strength produced by point mutations [8]. Our work on protein mutations provides the background for our involvement in the INtegrated and Scalable PredictIon of REsistanace (INSPIRE) project [9]. This project aims to lay the foundations for the use of molecular simulation and machine learning to guide precision cancer therapy, in which therapy is tailored to provide maximum benefit to individual patients based on the genetic information about their particular cancer. It is vital that such an approach is based on predictive methods as the vast majority of clinically observed mutations are rare, essentially ensuring that it will be impossible that catalog-building alone will be sufficient for making therapeutic decisions.

The project is a collaboration effort between UCL, the Memorial Sloan Kettering Cancer Centre (MSKCC), Argonne National Laboratory (ANL) and Rutgers University. Through the US DoE INCITE project INSPIRE was awarded compute time on first the Titan machine at Oakridge and now the world's most powerful supercomputer Summit. A key goal of the project is to develop a computational architecture hat allows the scaling of molecular dynamics and machine-learning workflows to take advantage of future

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

generation of exascale resources. For the machine-learning component of the project we have leveraged ANLs participation in the Exascale Deep Learning and Simulation Enabled Precision Medicine for Cancer (CANDLE) project [10] to develop tools that scale to use thousands of nodes for deep learning tasks. The development of molecular dynamics workflows towards exascale capabilities has built upon the software developed to conduct simulations across the hundreds of thousands of cores of the SuperMUC machine, based at the Leibniz Supercomputing Centre (LRZ) near Munich [11]. The middleware developed, known as HT-BAC was awarded the SCALE 2018 prize at CCGrid 2018 conference [12].

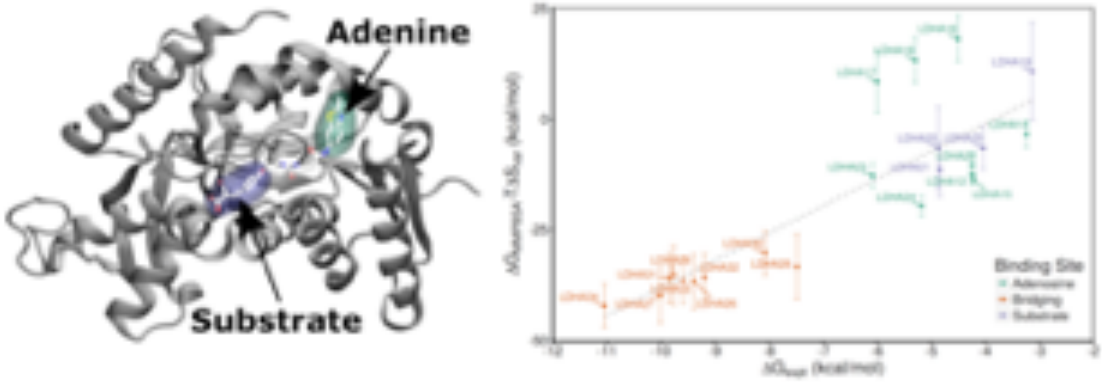

Figure 62. Structure of the lactate dehydogenase A (LDHA) protein with a bound ligand. Two disctint binding sites (adenine and substrate) to which fragments bind are highlighted in blue and cyan. A so-called "bridging ligand" which is built up by joining fragments binding each site is shown in chemical representation. Right: ESMACS simulation results incorporating variational entropy (ΔG_{MMPBSA}-$T\Delta S_{var}$) estimates correlate well with experimental values (ΔG_{expt}) even for drugs binding to different sites. This is typically a challenge for cheap computational methods.

Regarding the collaboration impact for UCL thanks to CompBioMed, UCL has an ongoing research collaboration in which Janssen provide targets and datasets and researchers from UCL and Janssen collaborate to develop and employ new methods in molecular dynamics based free energy calculations. Also, UCL, the Memorial Sloan Kettering Cancer Centre (MSKCC), Argonne National Laboratory (ANL) and Rutgers university collaborate to investigate drug resistance in tyrosine kinases. The collaboration seeks to combine molecular simulation with machine learning to enhance resistance prediction beyond what is possible from clinical and experimental data alone. Then, DNAnexus/UCL/EnsembleMD collaborate together on development support and credit for cloud compute time provided by DNAnexus to allow UCL and EnsembleMD to develop uf-BAC to run BAC workflows in a secure cloud environment. Finally, Azure/UCL/EnsembleMD work together on development support and credit for cloud compute time provided by Azure to allow UCL and EnsembleMD to develop uf-BAC to run BAC workflows on their HPC cloud platform.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

It is worth to mention the commercial impact of CompBioMed for UCL: the work described here has led directly to the formation of the spin out company EnsembleMD. The aim of the company is to provide cloud based simulation solutions based on ensemble molecular simulations and consulting services to the pharmaceutical sector.

References for this section

[1] P. V. Coveney and S. Wan, "On the calculation of equilibrium thermodynamic properties from molecular dynamics", Phys. Chem. Chem. Phys., 2016, 18, 30236-30240, DOI: 10.1039/C6CP02349E

[2] A. Bhati, S. Wan, D. Wright, P. V. Coveney, "Rapid, accurate, precise and reliable relative free energy prediction using ensemble based thermodynamic integration", Journal of Chemical Theory and Computation, 13 (1), 210–222 (2017), DOI: 10.1021/acs.jctc.6b00979

[3] David W. Wright, Shunzhou Wan, Christophe Meyer, Herman van Vlijmen, Gary Tresadern, Peter V. Coveney, "Application of ESMACS binding free energy protocols to diverse datasets: Bromodomain-containing protein 4", Scientific Reports 9, 6017, (2019), DOI:10.26434/chemrxiv.7327019

[4] S. Wan, A. P. Bhati, S. J. Zasada, I. Wall, D. Green, P. Bamborough, and P. V. Coveney, "Rapid and Reliable Binding Affinity Prediction of Bromodomain Inhibitors: a Computational Study", J. Chem. Theory Comput., 13 (2), 784–795 (2017), DOI: 10.1021/acs.jctc.6b00794

[5]S. Wan, A. Bhati, S. Skerratt, K. Omoto, V. Shanmugasundaram, S. Bagal, P. V. Coveney, "Evaluation and Characterization of Trk Kinase Inhibitors for the Treatment of Pain: Reliable Binding Affinity Predictions from Theory and Computation", Journal of Chemical Information and Modelling, 57 (4), 897–909 (2017), DOI: 10.1021/acs.jcim.6b00780

[6] A. Bhati, S. Wan, Y. Hu, B. Sherborne, P. V. Coveney, "Uncertainty Quantification in Alchemical Free Energy Methods", Journal of Chemical Theory and Computation, 14 (6), 2867–2880 (2018), DOI: 10.1021/acs.jctc.7b01143

[7] David W. Wright, Fouad Husseini, Shunzhou Wan, Christophe Meyer, Herman van Vlijmen, Gary Tresadern, and Peter V. Coveney, “Application of the ESMACS binding free energy protocol to a highly varied ligand dataset: lactate dehydogenase A”, Submitted

[8] A. P. Bhati, S. Wan, and P. V. Coveney, "Ensemble-based replica exchange alchemical free energy methods: the effect of protein mutations on inhibitor binding", J. Chem. Theory Comput., Available Online (2018), DOI: 10.1021/acs.jctc.8b01118

[9] US DOE (2018). “INCITE 2018 Fact Sheet”. http://www.doeleadershipcomputing.org/wpcontent/uploads/2019/03/2018INCITEFactSheets.pdf.

[10] CANDLE (2017). “Exascale Deep Learning and Simulation Enabled Precision Medicine for Cancer”. http://candle.cels.anl.gov/.

[11] Gauss Centre for Supercomputing (2016). “SuperMUC Enables Major Finding in Personalised Medicine”. https://www.gauss-centre.eu/news/pressreleases/article/lrzs-supermuc-enables-major-finding-in-personalised-medicine/

*“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **675451**“*

[12] J. Dakka, K. Farkas-Pall, V. Balasubramanian, M. Turilli, S. Wan, D. W. Wright, S. Zasada, P. V. Coveney, S. Jha, "Enabling Trade-offs Between Accuracy and Computational Cost: Adaptive Algorithms to Reduce Time to Clinical Insight", SCALE 2018, 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2018).
[13] Joseph A. DiMasi, Henry G.Grabowski and Ronald W.Hansen, "Innovation in the pharmaceutical industry: New estimates of R&D costs", Journal of Health Economics, 2016, 47, pp 20-33, DOI: 10.1016/j.jhealeco.2016.01.012
[14] Dipak Kalra, "The importance of real-world data to precision medicine", Personalized Medicine, 2019, 16, 2, DOI: 10.2217/pme-2018-0120
[15] D. Longley and P. Johnston (2005) “Molecular mechanisms of drug resistance”. The Journal of pathology, 205 (2), 275–292.
[16] N. P. Shah, J. M. Nicoll, B. Nagar, M. E. Gorre, R. L. Paquette, J. Kuriyan, and C. L. Sawyers (2002) “Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia”. Cancer cell, 2 (2), 117–125

Janssen: Molecular Dynamics for drug discovery programs

A ‘people’ definition of industrial computational chemistry includes a group of molecular modelling experts who use computational techniques, mostly to help molecular design in collaboration with medicinal chemists. Whilst other areas of application exist (such as in early target validation) this project will focus on methodologies that can make a significant step forward in the quality of molecular design. It can be argued that the current toolbox of an industrial computational chemist, despite incremental change, has not seen any fundamental improvement for over 10 years. Clearly challenge and investment are needed to learn if new methodologies can provide impact. In this project we studied new methodologies in areas of molecular dynamics calculations, and more accurate binding energy predictions.

Traditional industrial computational chemistry is highly dependent on a small selection of approaches such as virtual screening, molecular docking, ligand similarity etc. Structure-based drug design, where typically an X-ray crystal structure of the target is available, permits docking to help molecular design. Pose prediction with docking, that is correctly placing the ligand in the right orientation in the binding site, is typically considered an achievable task. Docking can also show virtual screening enrichment, which means separating a structurally diverse set (such as a random high throughput screening collection) of actives from inactives. However, it is widely recognized, that for a congeneric series of structural analogues, such as the case in a drug discovery lead optimization program, docking methodologies are unable to differentiate or rank highly active from inactive molecules. Hence, computational structure-based design remains largely qualitative and based on visual assessment and discussion and prioritization of results within project teams. This leads to various limitations, such as the number of molecules which can plausibly be docked and reliably assessed in this labour-intensive way.

*“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **675451**“*

Various fundamental computational methodologies have existed for a long time but remain largely unused in an industrial computational chemistry setting. Molecular Dynamics (MD) is one of these. By using Newton's classical equations of motion, computational simulations study the conformational changes of a protein (for instance) with time. Whilst these methods have existed for many decades, only now is it becoming feasible to consider running MD simulations for time scales that are relevant for issues of importance for drug discovery. At the same time, computational chemistry is undergoing significant changes due to access and porting of algorithms to Graphics Processing Unit (GPU) hardware. GPU's provide thousands of cores and offer a cheap highly parallel architecture which is efficient for computational approaches such as MD. During the past several years major steps have also been made in the practical feasibility of calculating free energies of binding of small molecules to proteins from the 3D structure of the protein-ligand complex. Important factors include improved molecular force fields, better conformational sampling, and faster hardware. Free energies of binding can now be calculated to within 1 kcal/mol accuracy, which is much better than commonly used approaches like molecular docking. This level of accuracy has the potential to radically increase the impact of computational design to drug discovery. Development of methods and best practices for free-energy calculations, as carried out in this project, will enable more effective computational design of drug candidates for globular proteins and membrane-bound targets.

Within a drug discovery portfolio many targets exist which may be more or less amenable to a structural biology approach, which can then permit structure-based drug design. Kinases and phosphodiesterases for instance contain soluble catalytic domains which can typically be crystallized in the presence of small molecule inhibitors. This provides significant benefit for molecular design and the subsequent computational docking can be performed with high confidence in pose prediction. Our work in this project aimed to create a greater understanding of new computational methods for drug discovery projects. Ultimately, accurate prediction of binding modes and binding energies will reduce the number of compounds required to be synthesized in typical lead optimization drug discovery programs.

Computational docking

Over the last three decades structure-based methods have taken a prime place in the drug discovery process. In the early pre-clinical process, lead optimization is known to be one of the most expensive tasks. Herein virtual screening, including molecular docking, is a tool to reduce costs by computationally differentiating actives from decoys before actual synthesis has been performed. Furthermore, computational methods can be used to steer the huge number (100's to 1000's) of compounds that are tested on one or more targets to yield indications of ligand affinity (e.g. IC_{50}/EC_{50} and K_I values). Hence a prime objective in computational chemistry is therefore the accurate prediction of such affinity values. Still, while useful to differentiate between actives and decoys, docking has been relatively unsuccessful in prediction of binding affinity (**Figure 63**).

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

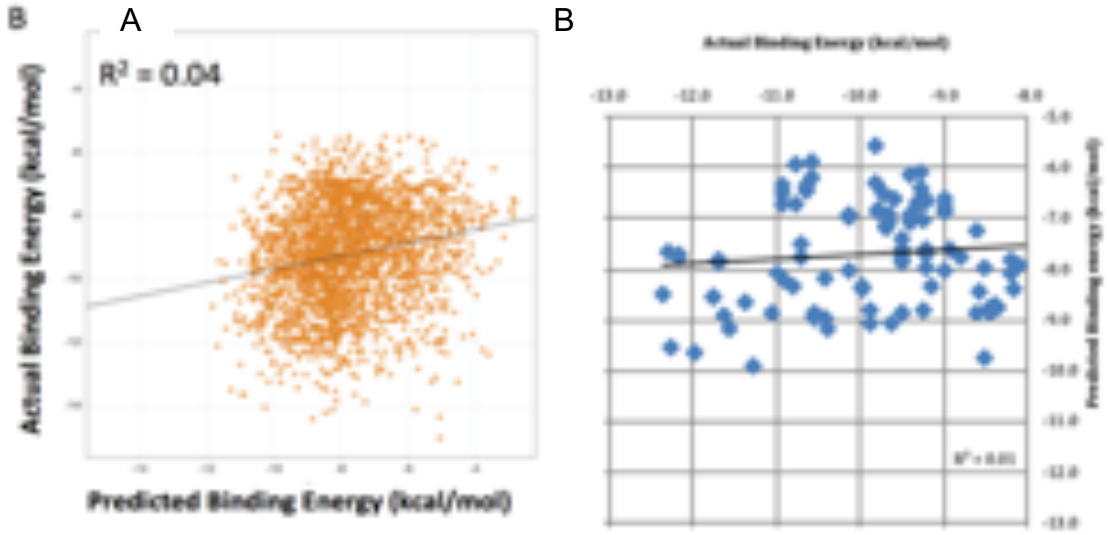

Figure 63. Sketching the problem. Structure-based methods, while generally performing acceptable for ligand pose generation, perform poorly in terms of affinity prediction. This case for the Adenosine A_{2A} receptor *(A unpublished results).* Moreover, a similar effect is observed in a prospective application performed for a soluble protein (PDE2) inhibition project in Janssen *(B unpublished results).* This project proposal will apply more accurate prediction methods and will measure experimental affinities to validate the methods used.

Given the insufficient predictive performance of docking, improving the quality of docking scoring is an active area of research. Ultimately one of the main issues is that scoring functions are deliberately parameterized to keep them as fast as possible to enable large scale virtual screening of hundreds of thousands or millions of molecules. Application in a lead optimization program where one typically deals with a few dozen compounds per week would allow much higher quality but slower approaches.

Affinity Prediction: Free-Energy Perturbation

Building on the recent advances in MD brings the realm of Free-Energy perturbation (FEP) calculations closer to a reality for application in industrial projects with short timelines. FEP has previously shown to be a promising *in silico* technique to estimate binding affinities. State of the art methods such as docking have proven to perform poorly in this task as mentioned above. Hence, free-energy simulations are a most valuable approach. A variety of free-energy simulation methods including FEP, use molecular dynamics or Monte Carlo (MD/MC) simulations to assess the free-energy difference between two related ligands via either a chemical or alchemical path. In fact, in a typical lead optimization program, the calculation of the relative difference in binding energy between two compounds is a leading principle as touched upon previously. Interestingly, this relative difference is more easily computed than the 'absolute' binding free-energy of a single compound. FEP calculations *per se* are not new. Based on the ideas of Zwanzig, FEP was applied in the 1970s and 1980s, when a number of research groups presented the first concepts of free-energy methods.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

However, lack of compute power, limited parameterization in early force fields, and other reasons impeded substantial progress in the field despite its attractiveness.

Recent advances in a number of areas such as better force fields, novel sampling algorithms, and low-cost (GPU-based) compute power can now deliver the level of accuracy and speed required for a typical drug discovery project. Combined with the large increase in structures available in the public domain and in house this opens the door for routine application of structure-based methods such as FEP, in drug discovery (

Figure 64). Yet before this is a reality, it is imperative that the shortcomings and limits of FEP are assessed systematically. Janssen have invested substantially in bringing FEP protocols to the drug discovery pipeline, in this project we explored the scope of open-source methods and techniques used in leading academic laboratories all in a pre-competitive manner.

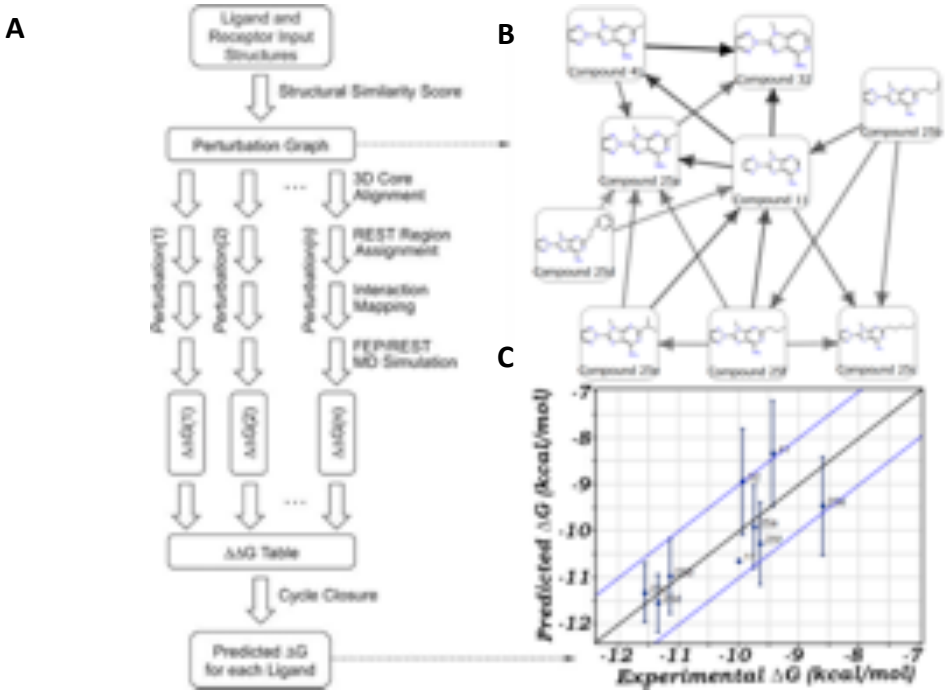

Figure 64. Figure adapted from the publication of Wang et al. (JACS, 2015) (A) FEP workflow for the prediction of protein–ligand binding affinity. A number of computational steps during the perturbations are mentioned. (B) Mapping the perturbations starting from compound 11, an adenosine A_{2A} receptor ligand, onto a set of pathways generated from the workflow for a total of nine ligands. Each arrow represents FEP calculations performed both in the receptor and in solution, for both ligands linked by the arrow. (C) Graph of calculated and experimental binding affinities for the nine adenosine A_{2A} receptor ligands.

Drug Discovery process

The research and development (R&D) process is lengthy and expensive, especially during later clinical stages. During the early phases of drug discovery, a lot of weight is therefore placed on pursuing well-validated pathophysiological mechanisms or targets. 'Validation' can be achieved via tissue expression profiling (including comparisons between healthy and disease states); 2) genetic association studies; 3) phenotypic

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"

analysis of transgenic animals; and 4) conducting in vivo/in vitro pharmacological studies with target-active tool (pre-lead) molecules. Absorption, distribution, metabolism, and excretion (ADME) and pharmacokinetic (PK) studies are employed early on in the drug discovery process to assist in the interpretation of in vivo experimentation. If a drug target survives this initial validation gauntlet, it then proceeds to primary drug screening, where small molecules and/or biological compounds are tested for the desired pharmacological intervention at the specified target. A target-centric *modus operandi* is followed where high throughput screens and follow-up assays during lead optimization are most used to drive molecular design. In this manner a portfolio of many targets is pursued with the best compounds for optimal targets ideally passing through to clinical evaluation.

New computational approaches described above can have a major impact on this drug discovery process. Importantly the work studied here will go to the heart of the design-make-test cycle and contribute higher quality methods for compound prioritization. This is a fundamental issue of drug discovery, because, whilst compounds with acceptable potency can often be found, and found quickly, they do not always come with the desired ADME or PK properties described above. Hence, during a typical lead optimization program the challenge often becomes to maintain potency whilst modifying the chemical structure of the lead molecules to overcome these other issues. In this regard computational tools which can accurately predict binding mode (i.e. high-quality homology modelling and docking) combined with accurate binding affinity prediction (free energy perturbation (FEP) methodologies) will be extremely powerful and reduce the number of 'backwards' steps required to subsequently move forwards in an LO program.

Janssen's primary interests in the CompBioMed project are in developing and using advanced molecular simulation methods to optimize lead compounds in discovery programs. Such methods, if proven robust and accurate could have a profound impact on the way drug discovery is performed. They would permit reliable computational triaging of very close analogue molecules greatly improving efficiency. Also, this would lead to high-confidence design of synthetically more challenging molecules leading to better drugs in new chemical space. Also, we envisage the accurate prediction of compound binding for targets that have mutated residues. This latter application can be of value in diagnostics, by predicting the best possible compound for a patient clinically (personalised medicine), but is also of use in discovery, where mutated targets occur regularly in antibacterials, antivirals, and oncology compounds.

We will summarize the CompBioMed impact for Janssen by describing the main collaborations within the project.

Collaboration Janssen - UCL

Janssen collaborated with UCL on calculation of free energies of binding on public and on Janssen internal compound sets (targets BRD4, LDHA and PDE2). A manuscript was co-written and accepted for publication on the BRD4 application (see Figure 65).

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

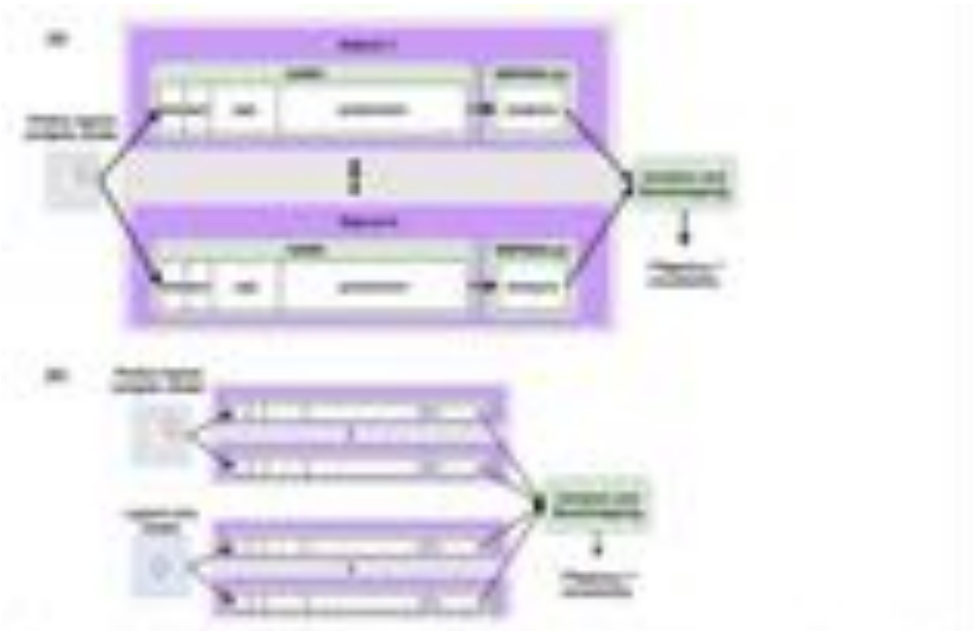

Figure 65. Overview of the ESMACS workflow. The 1traj protocol is shown in (a) consisting of an ensemble of 1 to N (25 in this study) simulations of the protein-ligand complex. Each simulation is made up of (min)imisation and two (eq)uilibration steps and a single production NAMD run which are each analysed independently using the MMPBSA.py script. The output of the analysis is then collated and bootstrap statistics produced. The multiple trajectory approaches, shown in (b) follow a similar outline but with independent trajectories also run of the ligand system alone.

A second manuscript is under review describing the LDHA application. Both cases have led to learnings about the suitable application of MMPBSA, so called ESMACS approach, for the calculation of binding free energies. It clearly is best suited for chemically diverse compounds and can deliver qualitatively useful results. Our work has particularly focused on understanding accurate method to account for explicit water molecules and entropic effects (Figure 65, Figure 66).

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

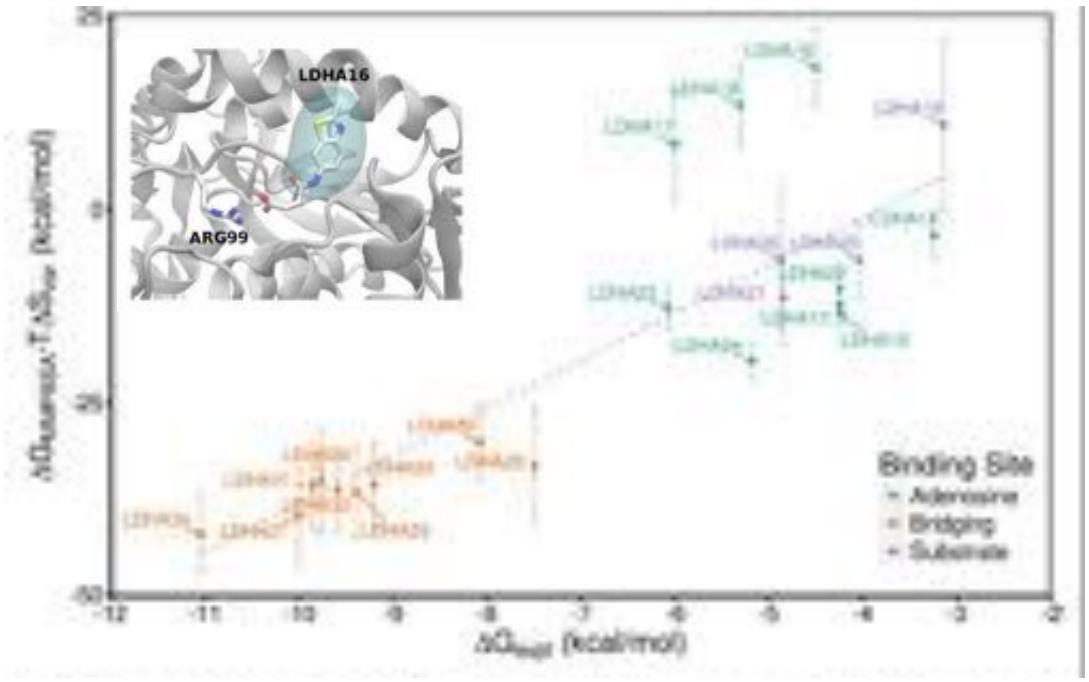

Figure 66. Comparison of binding free energies computed using 1traj MMPBSA based ESMACS protocol incorporating variational entropy. Ligand data points are coloured according to the pocket(s) to which they bind and a dashed grey line indicates the best fit using linear regression.

A third large and extensive study has been completed and a manuscript written and ready for submission to publication. This study has involved the use of alchemical perturbation methods, the TIES approach was performed at the UCL group, and the FEP+ approach at Janssen. Results have been performed for multiple perturbations from various protein targets. Shunzhou Wan from the UCL group also visited the Janssen site in Belgium as part of this project. A particular focus has been to investigate the precision of both methods when submitted to extensive repeated trial calculations. The TIES work has also required new replica exchange methodology recently implemented for TIES in the Coveney group. This work has uncovered that multiple repeats are necessary for improved accuracy and shorter simulations are performing better than individual longer simulations (Figure 67).

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

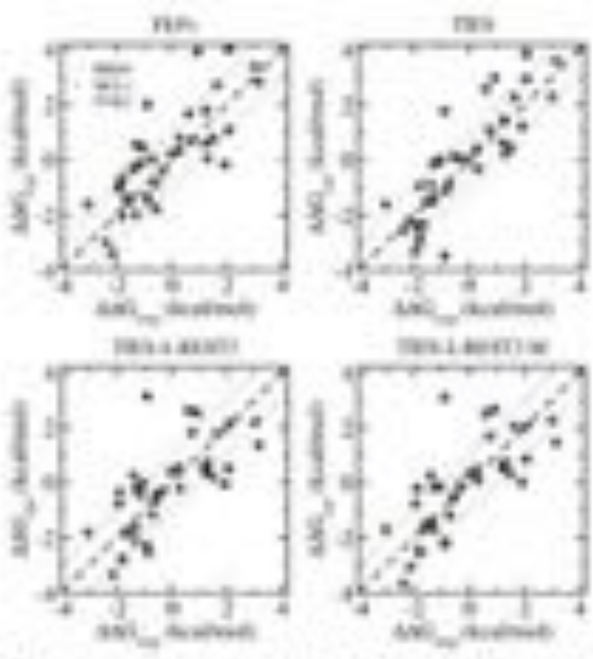

Figure 67. Comparison of the predicted binding free energy differences with the experimental data from the four approaches.

Collaboration Janssen – UPF/Acellera

Janssen collaborated with UPF/Acellera to test machine learning methods derived from protein ligand binding datasets, and used them to predict relative binding free energies. Scientists at Janssen visited the labs of Prof Gianni Fabritiis and worked with scientists there to optimize their approach. UPF/Acellera developed the first generation of the models based on publicly available datasets from PDBbind, and Janssen compiled datasets of PDE2, PDE3 and PDE10, BACE1 and ROS1 kinase bioactivities. The models were tested, retrained and tested again on these drug discovery lead optimization datasets. Performance was studied using chronological ordering of compounds but also in the context of optimal predictions, comparing the number of iterations of predictions that could be required to reveal the most active compound, compared to the established chronological order. The work is part of a manuscript that is under review for publication.

Janssen internal research for CompBioMed

Within Janssen we evaluated the use of GROMACS for Free Energy Perturbation. We streamlined the application of FEP with GROMACS and ran calculations at SurfSARA. Calculations were performed and compared with Schrodinger's FEP+ software. We studied multiple datasets, generating lots of valuable insights on the strengths and limitations of GROMACS FEP. Our work showed qualitative agreement for GROMACS FEP with the commercial FEP+ software, very encouraging results for future work. It also showed that certain protein and ligand systems such as metal containing binding site or large conformational changes in ligands are beyond current FEP calculations. Parts of this work were included in our recent publication [4].

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

Janssen completed a study demonstrating the use of MD for understanding functional activity of allosteric modulators. This work is published online. It is part of our initiative to understand the value of MD simulations in computational drug discovery and was performed by our CompBioMed-funded postdoc at Janssen [5].

Janssen – HPC interactions

During the project we benefited from excellent support by the HPC experts at SURFsara to enable setup and running of our FEP and MD calculations using GROMACS software on the Cartesius HPC infrastructure.

Also, during the project, Janssen has been a partner in two successful bids for HPC resources led by UCL:

- 32M core hours on SuperMUC in 2017-19, and on DoE Titan in an INCITE award worth around 100M core hours between 2018-2019. These grants are for work on binding affinity prediction.

References for this section

[1]. David W. Wright, Shunzhou Wan, Christophe Meyer, Herman van Vlijmen, Gary Tresadern, Peter V. Coveney. Application of ESMACS binding free energy protocols to diverse datasets: Bromodomain-containing protein 4. Nature Sci Rep, 2019, 9:6017.

[2]. David W. Wright, Shunzhou Wan, Christophe Meyer, Herman van Vlijmen, Gary Tresadern, Peter V. Coveney. Application of the ESMACS binding free energy protocol to a highly varied ligand dataset: LDHA. Manuscript under review.

[3]. David W. Wright, Shunzhou Wan, Laura Perez Benito, Herman van Vlijmen, Gary Tresadern, Peter V. Coveney. Study of Precision and Accuracy of Alchemical Relative Free Energy of Binding Predictions. TIES and FEP+. Manuscript under preparation.

[4]. Laura Pérez-Benito, Nil Casajuana Martin, Mireia Jiménez-Rosés, Herman Van Vlijmen, Gary Tresadern. Predicting Activity Cliffs with Free Energy Perturbation. J. Chem. Theory. Comput. 2019, 15, 1884−1895.

[5]. Claudia Llinas del Torrent, Nil Casajuana-Martin, Leonardo Pardo, Gary Tresadern, Laura Pérez-Benito. Mechanisms Underlying Allosteric Molecular Switches of Metabotropic Glutamate Receptor 5. J. Chem. Inf. Model. 2019, 59 (5), 2456-2466.

EVOTEC: G-protein activated structures modelling for large-scale computers

Evotec (UK) Ltd, as a leading industrial application partner, is responsible for four key objectives: adaptation of hierarchical G-Protein Coupled Receptors (GPCR) modelling protocol (HGMP) to HPC platform, developing of the new HGMP-HPC based tools / plugins that require high scale calculations, testing and application of HGMP-HPC integrated technology in real drug discovery cases within the CoE and to make it available to third parties seeking assistance from the CoE and/or from Evotec, and dissemination of the results of this work to our partners in academia and in pharma & biotech companies in order to stimulate follow-on research. Evotec has also published the outcome of this work in peer-reviewed journals and at scientific conferences. Evotec (UK) Ltd (Dr Alexander Heifetz) has established a close collaboration with UCL (group of Prof Andrea Townsend-Nicholson) [1]. In the framework of this collaboration, they

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

developed computational methodologies for structural exploration and tools for drug design, described as follows:

Rationalizing the receptor-ligand binding and drug-candidates' residence time [2]. Drug-target residence time, the length of time for which a small molecule stays bound to its receptor target, has increasingly become a key property for optimization in drug discovery programs. However, its in silico prediction has proven difficult. Here we describe a method, using atomistic ensemble-based steered molecular dynamics (SMD), to observe the dissociation of ligands from their target G protein-coupled receptor in a time scale suitable for drug discovery. These dissociation simulations accurately, precisely, and reproducibly identify ligand-residue interactions and quantify the change in ligand energy values for both protein and water. The method has been applied to 17 ligands of the A_{2A} adenosine receptor, all with published experimental kinetic binding data. The residues that interact with the ligand as it dissociates are known experimentally to have an effect on binding affinities and residence times. There is a good correlation ($R^2 = 0.79$) between the computationally calculated change in water-ligand interaction energy and experimentally determined residence time. Our results indicate that ensemble-based SMD is a rapid, novel, and accurate semi-empirical method for the determination of drug-target relative residence time.

Computational prediction of GPCR oligomerization [3]
There has been a recent and prolific expansion in the number of GPCR crystal structures being solved: in both active and inactive forms and in complex with ligand, with G protein and with each other. Despite this, there is relatively little experimental information about the precise configuration of GPCR oligomers during these different biologically relevant states. While it may be possible to identify the experimental conditions necessary to crystallize a GPCR preferentially in a specific structural conformation, computational approaches afford a potentially more tractable means of describing the probability of formation of receptor dimers and higher order oligomers. Ensemble-based computational methods based on structurally determined dimers, coupled with a computational workflow that uses quantum mechanical methods to analyze the chemical nature of the molecular interactions at a GPCR dimer interface, will generate the reproducible and accurate predictions needed to predict previously unidentified GPCR dimers and to inform future advances in our ability to understand and begin to precisely manipulate GPCR oligomers in biological systems. It may also provide information needed to achieve an increase in the number of experimentally determined oligomeric GPCR structures.

FMO-DFTB tool for rapid analysis of receptor-ligand interactions [4]
The reliable and precise evaluation of receptor-ligand interactions and pair-interaction energy is an essential element of rational drug design. While quantum mechanical (QM) methods have been a promising means by which to achieve this, traditional QM is not applicable for large biological systems due to its high computational cost. Here, the fragment molecular orbital (FMO) method has been used to accelerate QM calculations, and by combining FMO with the density-functional tight-binding (DFTB) method we are

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

able to decrease computational cost 1000 times, achieving results in seconds, instead of hours. We have applied FMO-DFTB to three different GPCR-ligand systems. Our results correlate well with site directed mutagenesis data and findings presented in the published literature, demonstrating that FMO-DFTB is a rapid and accurate means of GPCR-ligand interactions. This action was performed in collaboration with Dr Dmitri Fedorov from National Institute of Advanced Industrial Science and Technology (AIST), Japan.

FMO-PPi tool of inter-helical interactions of G-protein coupled receptors.
G-protein coupled receptors (GPCRs) are the largest superfamily of membrane proteins. They regulate almost every aspect of cellular activity and are key targets for drug discovery. However, the molecular forces responsible for holding together the seven helices of the GPCR bundle and ensuring receptor stability, ligand binding and activation have not been identified. Even with crystal structures in hand, the strength and chemical nature of these forces cannot be characterised by visual inspection alone and therefore, accurate and reliable computational methods must be employed. Quantum mechanics (QM) approaches can be used to provide this information but they are computationally expensive. However, the fragment molecular orbital (FMO) QM method offers an excellent solution that combines accuracy, rigour and speed. FMO run for one full sized receptor took only two hours on 340 CPU cores. Here, we have applied FMO to 35 crystal structures representing different branches of the class A GPCR family to characterise the strength and chemical nature of the inter-helical interactions between the residues of transmembrane (TM) domains in different receptor activation states. Our approach has yielded novel results that are consistent with and help to rationalise experimental data. We have identified 69 topologically-equivalent TM residues of class A GPCRs that form a consensus network of 51 inter-TM interactions. This discovery provides a comprehensive picture of how various molecular forces govern the inter-helical interactions, which in turn support structural stability, ligand binding and activation of GPCRs. Our findings also provide molecular insights into how ligand binding can affect the overall structural properties of these key signalling proteins. At Evotec, we are already intensively applying this information in our internal drug discovery projects and see how important it is. Academic colleagues are doing the same in their research with equal success. This action was performed in collaboration with Dr Dmitri Fedorov from National Institute of Advanced Industrial Science and Technology (AIST), Japan.

These tools are intensively used by Evotec in its internal drug-discovery projects and disseminate it among its clients. The outcome of this work we published in many high profile peer-reviewed journal, presented in various international conferences and summarised in two books two books published by Springer: ‘Computational Methods for GPCR Drug Discovery’ (https://www.springer.com/gp/book/9781493974641) and ‘QM methods for drug discovery’ (ongoing, will be published in 2020), edited by Dr Alexander Heifetz.

Additionally, and to increase the impact, Evotec (Alex Heifetz) provided a training session in the CompBioMed & BioExel Free-Energy Workshop, London, 31 May 2017;

*“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **675451**“*

and at the CompBioMed winter school in Barcelona in February 2018 and in February 2019 on the subject of 'Introduction to Computer-Aided Drug Design (CADD) and GPCR Modelling' illustrated by examples from real drug discovery projects. Videos of both these sessions are available via the CompBioMed YouTube channel.

We further developing and optimising these tools to make them even more user-friendly and integrated.

References for this section

[1]. Potterton, A., Heifetz, A. & Townsend-Nicholson, A. Synergistic Use of GPCR Modeling and SDM Experiments to Understand Ligand Binding. Methods in molecular biology (Clifton, N.J.) 1705, 335-343, doi:10.1007/978-1-4939-7465-8_15 (2018).

[2]. Potterton, A. et al. Ensemble-Based Steered Molecular Dynamics Predicts Relative Residence Time of A2A Receptor Binders. J Chem Theory Comput, doi:10.1021/acs.jctc.8b01270 (2019).

[3]. Townsend-Nicholson, A., Altwaijry, N., Potterton, A., Morao, I. & Heifetz, A. Computational prediction of GPCR oligomerization. Current opinion in structural biology 55, 178-184, doi:10.1016/j.sbi.2019.04.005 (2019).

[4]. Morao, I. et al. Rapid and accurate assessment of GPCR-ligand interactions Using the fragment molecular orbital-based density-functional tight-binding method. Journal of computational chemistry 38, 1987-1990, doi:10.1002/jcc.24850 (2017).

Neuro-musculoskeletal Exemplar Research

University of Sheffield: from tomography to bone simulations

In Sheffield, the **Computer Tomography to Strength (CT2S)** has been successfully rolled out and tested on ShARC (Sheffield Tier-3), and applied to more than 110 patients so far. The service uses hip CT scans to generate personalised finite element models of the femurs, and ran simulations to predict the femoral strength in order to predict the risk of osteoporotic fracture (Figure 68). The service has recently been linked to the Sheffield Teaching Hospital, where a clinical staff can send a request for a set of patient CT scans to be analysed with a report being returned (within 1-2 days) to detail the risk factors. An abstract describing the CT2S workflow has been accepted to the CompBioMed Conference in Sep 2019. The research work has also resulted in publications [1,2].

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

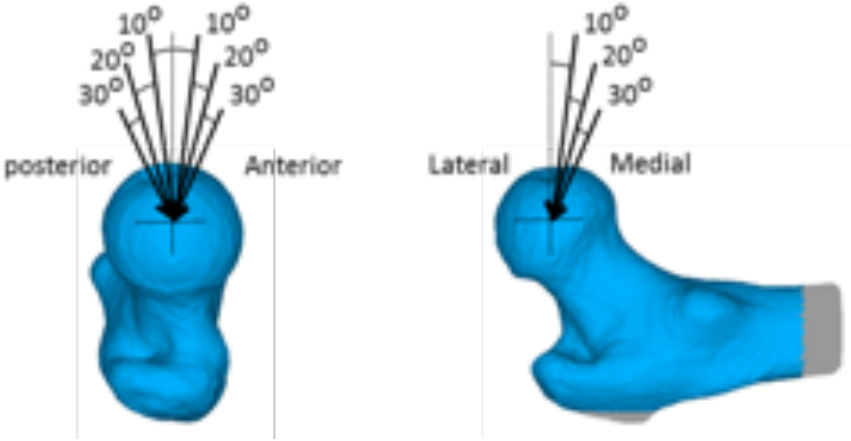

Figure 68. The range of sideways fall loading directions tested in the algorithm. Reproduced from Altai et al. (2019), Clinical Biomechanics.

Currently, the algorithm used in CT2S has been further developed in multi-scale modelling approaches, where the loads obtained from gait analysis were used to study the biomechanical response of the femur during level walking and daily activities (Figure 69). These results are ready to be published, with the paper in preparation.

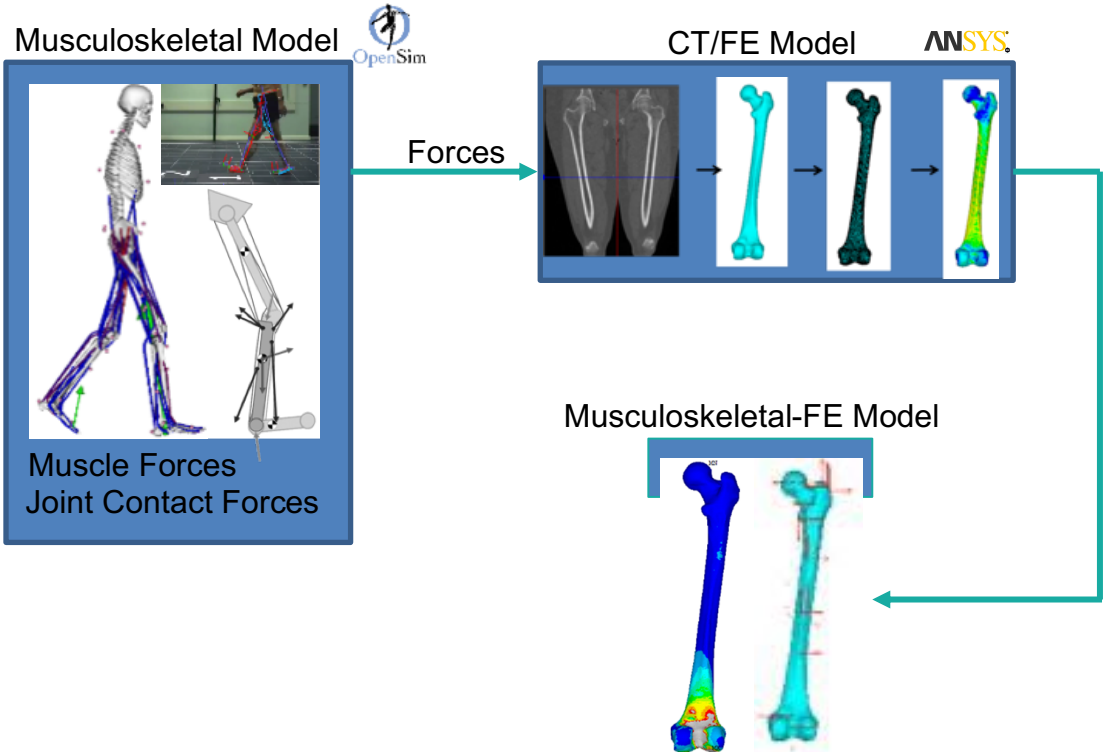

Figure 69. Integrating muscle and joint contact forces predicted from musculoskeletal model with organ-scale finite element model in order to study the biomechanics of the femur during daily activities.

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

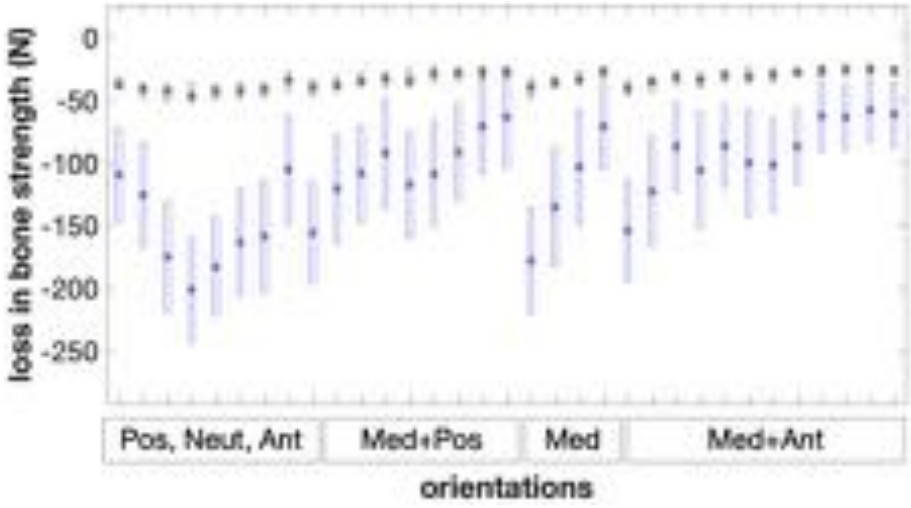

Figure 70. Loss in bone strength (in Newtons) for different fall impact orientations (horizontal axis) following loss in FN-aBMD from 0.66 g/cm^2 to 0.627 g/cm^2. Loss of bone strength predicted by bone loss law (black boxes) is much smaller in magnitude and population variability than that predicted by linear regression relationship between bone strength and FN-aBMD (blue boxes). Median and 25th and 75th percentile values are denoted by the target and the upper and lower edges of the boxes respectively.

The CT2S workflow has employed in an in silico study that investigates the effect of ageing on bone strength. A bone loss law has been developed that describes how local volumetric bone mineral density (vBMD) decreases as the areal bone mineral density (aBMD) at the femoral neck (FN) decreases during ageing. Combined with the CT2S workflow, this framework leads to a determination of how fall orientation specific bone strength changes due to ageing, such as following a 5% decrease in FN-aBMD. Figure 70 shows that changes in bone strength are much smaller in magnitude and variability when predicted using bone loss law than when predicted by linear regression. This highlights that bone strength loss really depends on the individual's bone shape and size and on how vBMD is distributed spatially within the bone.

The CT2S service is currently being used by the Sheffield Teaching Hospitals to process patient data and predict the risk of osteoporotic fracture. The algorithm is also used by clinicians at the Sheffield Children's Hospitals for research purposes in the application of child abuse. Other research users include the Flinders University (Australia) and the University of Wisconsin (USA).

The BoneDVC algorithm has been used to validate micro-finite element models of the mouse tibia for preclinical assessment of the effect of interventions for musculoskeletal diseases [3]. The BoneDVC has been also used to evaluate the reproducibility of a typical approach used to study the effect of mechanical loading on the bone remodelling [4]. The elastic registration library within the BoneDVC algorithm has been used to study the prenatal joint geometrical changes in a mouse model which will be used in the future to evaluate the effect of loading on joint development [5]. In another application the BoneDVC algorithm was used to study the uncertainties in strain measurements in the cortical bone by using high resolution Synchrotron micro computed tomography images

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

[6]. Finally, the BoneDVC has been used to validate computed tomography (CT) based subject specific finite element models of the human scapula by using a combination of CT scans, microCT scans and in situ loading with a six degree-of-freedom hexapod robot [7,8].

As part of the BoneDVC service, the Sheffield-based image processing software SHIRT has been rewritten in order to make it easier to parallelise on HPC system. The new software is called **pFIRE**. It has been deployed and tested on ShARC, ARCHER, MareNostrum. The software is available to download via Github with a set of dedicated tutorials to get the users started (https://insigneo.github.io/pFIRE/tutorial.html).

A training session of pFIRE was provided at the 2019 CompBioMed Winter School in Barcelona, which was well received, with more researchers subscribed to use the software.

References for this section

[1]. Z. Altai, M. Qasim, **X. Li**, M. Viceconti 2019. "The effect of boundary and loading conditions on patient classification using finite element predicted risk of fracture", Cllinical Biomechanics, 68, pp.137-43. doi: 10.1016/j.clinbiomech.2019.06.004

[2]. M. Viceconti, M. Qasim, P. Bhattacharya, **X. Li** 2018. "Are CT-based finite element model predictions of femoral bone strength clinically useful?", Current Osteoporosis Reports, 16(3), pp. 216-23. doi: 10.1007/s11914-018-0438-8.

[3]. Oliviero S, Giorgi M, Dall'Ara E 2018. "Validation of finite element models of the mouse tibia using digital volume correlation", J Mech Behav Mater, 86, pp. 172-84. doi: 10.1016/j.jmbbm.2018.06.022

[4]. Giorgi M, Dall'Ara E 2018. "Variability in strain distribution in the mice tibia loading model: A preliminary study using digital volume correlation", Med Eng Phys, 62, pp. 7-16. doi: 10.1016/j.medengphy.2018.09.001

[5]. Giorgi M, Vivien Sotiriou, Niccolo Fanchini, Simone Conigliaro, Cristina Bignardi, Niamh C. Nomina, Dall'Ara E 2019. "Prenatal growth map of the mouse knee ", PLoS One, 14(1): e0197947. doi: 10.1371/journal.pone.0197947

[6]. Comini F, Palanca M, Cristofolini L, Dall'Ara E 2019. "Uncertainties of synchrotron microCT-based digital volume correlation bone strain measurements under simulated deformation", J Biomech, 86, pp. 232-37. doi: 10.1016/j.jbiomech.2019.01.041

[7]. Kusins J, Knowles N, Ryan M, Dall'Ara E, Ferreira L 2019. "Performance of QCT-derived scapula finite element models in predicting local displacements using digital volume correlation", J Mech Behav Biomed Mat, 97, pp. 339-45. doi: 10.1016/j.jmbbm.2019.05.021

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*

[8]. Knowles NK, Kusins J, Faieghi M, Ryan M, Dall'Ara E, Ferreira LM 2019. "Material mapping of QCT-derived scapular models: A comparison with micro-CT loaded specimens using digital volume correlation", Ann Biomed Eng. doi: 10.1007/s10439-019-02312-2

*"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **675451**"*