Grant agreement no. 823712

CompBioMed2

Research and Innovation Action
H2020-INFRAEDI-2018-1
Topic: Centres of Excellence in computing applications

D2.2 – Second Report on Fast Track Application Readiness

Work Package: 2

Due date of deliverable: Month 24

Actual submission date: 01 October 2021

Start date of project: 01 October 2019 Duration: 48 months

Lead beneficiary for this deliverable: *UvA*
Contributors: *BSC, UNIBO, UNIGE, UCL, UPF, USFD, UOXF, ACE*

Disclaimer
This document's contents are not intended to replace consultation of any applicable legal sources or the necessary advice of a legal expert, where appropriate. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user, therefore, uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view.

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	

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

Table of Contents

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

1 Version Log

Version	Date	Released by	Nature of Change
V0.1	14/08/2021	Gabor Zavodszky	First Draft, outline
V0.2	08/09/2021	Gabor Zavodszky	Submitted draft for internal review
V0.3	22/09/2021	Gabor Zavodszky	Final Draft for review
V1.0	30/09/2021	Emily Lumley	Final version for submission
V1.1	01/10/2021	Emily Lumley	Final submitted version.

2 Contributors

Name	Institution	Role
Gabor Zavodszky	UvA	Principal Author
Jazmin Aguado-Sierra	BSC	Contributor
Alejandro Varela Rial	Acellera	Contributor
Adrià Pérez Culubret	UPF	Contributor
Xinshan Li	USFD	Contributor
Ivan Benemerito	USFD	Contributor
Pinaki Bhattacharya	USFD	Contributor
Daniele Tartarini	USFD	Contributor
Marco Viceconti	UNIBO	Contributor
Antonino Amedeo La Mattina	UNIBO	Contributor
Jon McCullough	UCL	Contributor
Shunzhou Wan	UCL	Contributor
Mat Bieniek	UCL	Contributor
Christos Kotsalos	UNIGE	Contributor
Blanca Rodriguez	UOXF	Contributor
Wang Lei	UOXF	Contributor
Gary Tresaden	JAN	Reviewer
Raimondas Galvelis	ACE	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 **823712**"*

3 Definition and Acronyms

Acronyms	Definitions
0/1/2/3D	0/1/2/3 Dimension
aBMD	Areal Bone Mineral Density
APDL	Algorithmic Processor Description Language
ARFn	Absolute Risk of Fracture at year n
AZM	Azithromycin
BAC	Binding Affinity Calculator
CHU	Centre Hospitalier Universitaire
CMR	Cardiovascular Magnetic Resonance
CNN	Convolutional NeuralNetworks
CoE	Centre of Excellence
COPD	Chronic Obstructive Pulmonary Disease
CPU	Central Processing Unit
CT	Computed Tomography
CT2S	Computed Tomography to Strength
DEXA	Dual Energy X-ray Absorptiometry
DT	Deep Track
ECG	Electrocardiogram
ER	Estrogen receptor
ESMACS	Enhanced Sampling of Molecular dynamics with Approximation of Continuum Solvent
ESMCDLB	Dynamic Load Balancing
EU	European Union
FBDD	Fragment Based Drug Discovery
FE	Finite Element
FPGA	Field Programmable Gate Arrays
FSi	Fluid/Solid interaction
FT	Fast Track
FTI	Fault Tolerance Interface
GPCR	G-protein Coupled Receptor
GPE	Gaussian process emulators
GPU	Graphics Processing Unit

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

HCQ	Hydroxychloroquine
HMM	Heterogeneous Multiscale Model
HPC	High Performance Computing
I/O	Input/Output
IG	Integrated Gradients
IS	Ischaemic stroke
LB	Latttice Boltzmann
LRZ	Leibniz Supercomputing Centre
MD	Molecular Dynamics
MPI	Message Passing Interface
MPIO	MultiPath Input Output
NNP	Neural Network Potentials
PCA	Principal Component Analysis
pFIRE	parallel Framework for Image REgistration
PV loops	Pressure-Volume Loops
RBC	Red Blood Cell
SED	Strain Energy Density
SSA	Sobol's sensitivity analysis
STT	Soft Tissue Thickness
TALP	Termination Analysis of Logic Programs
TB	Tuberculosis
TCD	Transcranial Doppler
TIES	Thermodynamic Integration with Enhanced Sampling
UEABS	The Unified European Application Benchmark Suite
UK	United Kingdom
UNIGE	University of Geneva
UQ	Uncertainty Quantification
USA	United States of America
UvA	University of Amsterdam
vBMD	Volumetric Bone Mineral Density
VVUQ	Validation, Verification, and Uncertainty Quantification
WP	Work Package

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

4 Public Summary

Work Package 2 (WP2) in the CompBioMed2 project aggregates the computational biomedical research activities of the project. The leader of this work package is the University of Amsterdam (UvA). Computational Biomedicine is highly diverse, subsuming a large range of modelling and simulation methods, each applicable to a specific level of human body organisation. State of the art mathematical models in Computational Biomedicine need to take into account the inherent multiphysics and multiscale character of human physiology. This means that, typically, different types of models, representing different scales within a given (patient specific) biomedical system, often need to be taken into account, including the details as to how to correctly connect these different models.

The main objective of this work package is to lay the groundwork for the application of High Performance Computer (HPC)-based Computational Biomedicine approaches to a great number of therapeutic areas. The HPC requirements of our users are as diverse as the communities we represent; therefore, we must support a wide palette of codes potentially scaling to the exascale. These codes can be categorised based on their execution pattern as (I) monolithic code, (II) coupled code, or (III) complex workflow requiring support for advanced execution patterns on a range of diverse and geographically distributed platforms.

This document reports on the scientific progress achieved within our application portfolio and includes a list and description of the publicly available user resources that aim to enable efficient external code adoption and that will collectively form the “best practices” guide for external users.

5 Introduction

In *Deliverable 2.1 - First Report on Fast Track Application Readiness*, we described and presented the deployment of the applications and the related research within CompBioMed2.

Building on these applications, the objectives of WP2 include:

- To advance the state-of-the-art in simulation based biomedical science, from the desktop towards highly scalable, optimised codes for the most powerful existing multi-petaflops HPC systems;
- To prepare appropriate and high impact CompBioMed2 applications for the emerging exascale;
- To develop verification, validation and uncertainty quantification techniques that will permit CompBioMed2 applications to be used in *in silico* clinical trials and personal health forecasting applications;
- To engage with the broader CompBioMed community, advising on deployment of codes, need for parallelization and development of workflows on a range of HPC platforms;
- To develop cross-application systems models that link CompBioMed2 applications in multiscale and multiphysics scenarios, in combination with high performance data analytics where appropriate.

WP2 operates in strong collaboration with the other WPs, in particular, WP3 for data management, long term data storage, and high performance data analytic techniques, and WP4

*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **823712**“*

for compute and data services, to optimize the access and usage of current and emerging e-infrastructures.

For details on the categorisation of the application codes and the description of the software stacks please consult Chpt. 5 and 6.1 of the previous Fast Track Application Deliverable document (D2.1).

In the current deliverable (D2.2) we report advancements on scientific results since the previous report (D2.1). The structure of the document follows the order of the tasks, and for every task it reports the related research results, as well as progress in code development and optimization.

User resources

CompBioMed2 has a broad portfolio of codes and applications, which is necessitated by the extensive palette of research questions that are covered within the project. All of the codes are designed from the ground up for execution in HPC environments, and therefore contain various application specific optimisations. This in practice often implies that the adoption of the codes by external users is a non-trivial task. In order to facilitate the process of adoption, CompBioMed has several ongoing operations. The list of codes, general information, and relevant links are available in the *Software Hub*[1], that is maintained as part of the project website by WP4. Further regular actions include providing trainings, public e-seminars, and hands-on guidance. These are also listed on the website in the *Training Repositor*y[2]. Many of these efforts are carried out within the bounds of WP6; however, there are further relevant ongoing resource developments closely related to the codes reported here. These resources (including installation and user documentation, tutorials and guides) collectively form a first version of "best practices" aimed for external code adopters. The following section also describes the main user resources for every application, which are meant to serve as the starting point for external adopters. An extended full list of publicly available resources is provided in the Appendices (Chpt. 10.1).

Further details on the code developments that relates to the external user experience (i.e., development of the build systems of the codes, and other improvements in user documentation) are reported by WP5 in D5.2.

6 Activities Carried out

In this section we summarize the current status of the research tasks and their results for the period since the previous report (D2.1). Furthermore, we report on improvements in code development, deployments, and user resource availability in each subtask.

[1] https://www.compbiomed.eu/services/software-hub/

[2] https://www.compbiomed.eu/training-3/

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

6.1 Task 2.1 – Cardiovascular Research Exemplar

6.1.1 Subtask 2.1.1 – Whole body blood flow modelling

Short context and research progress

The main focus for subtask 2.1.1 is the study of macroscopic blood flow at up to full human scale. In particular, this makes use of the open source lattice Boltzmann code HemeLB [2.1.1/1, 2.1.1/2] to study flows in complex arterial and venous geometries in 3D. HemeLB has been optimized to efficiently simulate blood flow through the sparse and complex geometries that are typical of large-scale vascular structures [2.1.1/3]. Simulating blood flow in 3D at this scale is an ambitious challenge that pushes the limits of current computational capacity. The outcomes of this work will provide deeper insight into cardiovascular flow and disease than can be provided by 0D/1D models.

Scientific developments since the last report have been focused on three key areas: arteriovenous fistula, elastic wall representation and coupling with the Alya heart model. In [2.1.1/4] we discuss refinements to our coupling method and illustrate how our self-coupled HemeLB model is able to capture some key physiological behaviour associated with arteriovenous fistula. We also highlight some of the challenges associated with validating modelling of personalized vessels. Blood vessels are naturally elastic, however incorporating accurate solid mechanics behaviour into a fluid mechanics model such as HemeLB can be expensive both in terms of development and simulation execution time. To overcome this we have developed a simple boundary condition that enables key flow features of elastic walled vessels to be incorporated into a HemeLB simulation more accurately than is achieved with a rigid wall approximation without any loss of computational performance. Finally we have been further developing the coupling of HemeLB and the Alya heart model. Initial tests have been successful and we are continuing to develop the framework to address differences in the simulation requirements of the two codes.

Visualisation of large-scale data remains a significant challenge for whole body blood flow modelling. Working with LRZ, a workflow has been developed that allows HemeLB output data to be efficiently visualized using HPC resources – in particular SuperMUC-NG [2.1.1/5]. This workflow allows for very large data sets to be efficiently rendered and animated (largest consisting of almost 1.5TB of data). It also has the capability of producing immersive visualizations that can be observed in a virtual reality environment. Such techniques will be beneficial in communicating the results of virtual human simulations to end-users and patients. The significance and importance of this work has been recognised through acceptance as a finalist in the SC21 conference Scientific Visualization Showcase.

Code development, deployment and user resources

The GPU enabled version has allowed us to examine the strong scaling performance on the accelerated architectures that are becoming increasingly common in HPC. We have demonstrated excellent scaling results on up to ca 20,000 NVIDIA V100 GPUs on the USA machine Summit as well as being able to execute jobs without further performance loss to ca 25,000 GPUs. Using the Top500 metric of equivalence between GPU streaming multiprocessors and CPU cores this represents strong scaling to the equivalent of almost 1.5 million CPUs. For

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

the CPU version of HemeLB, we have recently updated the intrinsics formulation used in the code and this has resulted in a significant improvement in observed parallel efficiency compared to the previous version (71% vs 58% for a specific test case on SuperMUC-NG using ca 50,000 CPU cores). We are currently planning for a full-machine test using this improved layout. For the self-coupled version of HemeLB we have also demonstrated strong scaling to 96,000 cores on SuperMUC-NG [2.1.1/4].

An introductory guide to running HemeLB simulations and post-processing the results is available at [2.1.1/2] . We are currently working with FocusCoE to develop a more detailed tutorial focused on scaling and performance on HPC with HemeLB at their behest. FocusCoE chose this application precisely because it is a leading exemplar for codes that execute at the emerging exascale. This is expected to become publicly available during Q4 2021.

Further user resources can be found in Table 10/2.1.1 in the Appendix.

6.1.2 Subtask 2.1.2 – Improving diagnosis of cerebral vasospasm

Short context and research progress

Ischaemic stroke (IS) is an occlusion of brain arteries which limits the delivery of blood and nutrients to distal regions of the brain. IS is typically diagnosed using imaging techniques which, besides being expensive, are time consuming [2.1.2/1]. Transcranial Doppler ultrasound (TCD) can measure the features of the blood velocity waveforms and detect noninvasively the presence of an occlusion [2.1.2/1] but it is not able to inform on the status of the distal circulation. We developed a mixed mechanistic-statistic approach to predict the status of distal perfusion. The mechanistic part is based on 1D blood flow modelling, which is used to predict the velocity waveform in the circulatory network [2.1.2/2]. Gaussian process emulators (GPE) and Sobol's sensitivity analysis (SSA) [2.1.2/3] are used in the statistical part to identify the parameters that are both measurable with TCD and sensitive to the network alterations caused by the occlusion. Figure 2.1.2/1 shows an overview of the methodology. Results show that blood pulsatility can predict the distal perfusion in case of IS, and assess the effect of parameter uncertainty on the clinical measurement.

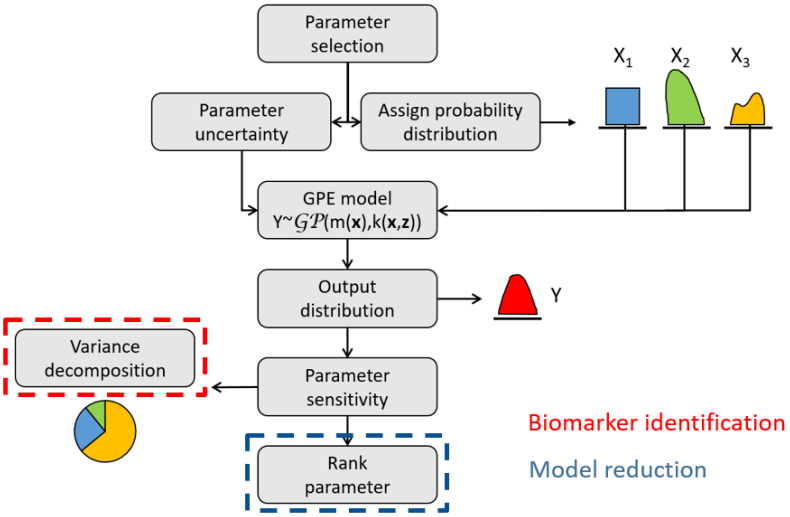


Figure 2.1.2/1 VVUQ pipeline for biomarker identification and model reduction

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

Code development, deployment and user resources

The code of the 1D solver openBF is available open source and documented through GitHub (https://github.com/INSIGNEO/openBF), together with tutorials on the basic usage. Further user resources can be found in Table 10/2.1.2 in the Appendix.

6.1.3 Subtask 2.1.3 – Cell based Blood flow simulations

Short context and research progress

HemoCell is an open-source code developed to simulate cell resolved blood flows in large spatial domains under high sustained shear-rates (typical arterial conditions). It is a coupled code, where the blood plasma (the fluid component) is solved using the lattice Boltzmann method (building on the Palabos library), and the cell mechanics and deformation is solved using the discrete element method. These two components are coupled via the immersed boundary method [2.1.3/1]. The code has been validated for both healthy and diabetic blood flows [2.1.3/2], and the validity and robustness of the red blood cell mechanical model was evaluated with VVUQ methods [2.1.3/3].

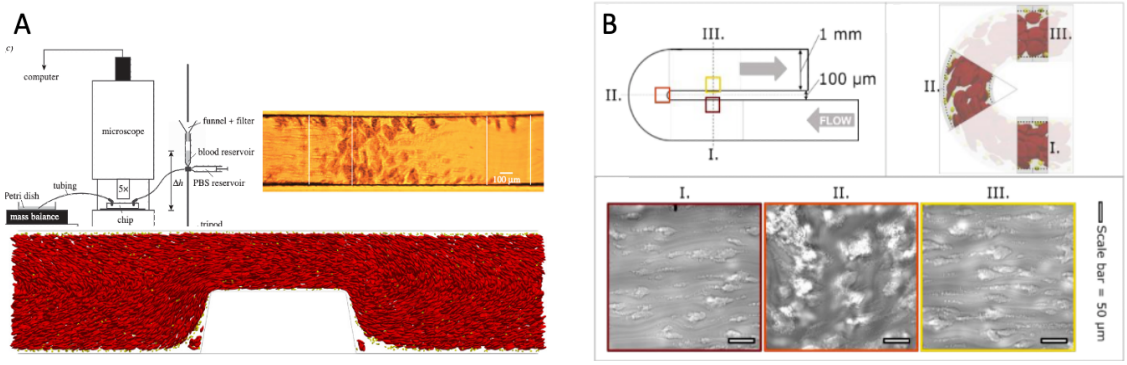


Figure 2.1.3/1 (A) Platelet transport at the initial phase of thrombus formation. In vitro and in silico. (B) Thrombus formation in a curved channel.

The recent developments have focused on extending the previous investigations on the movement and trajectory of cells to more complex geometries. On the one hand we have identified the flow conditions present at the initiation of a high-shear platelet aggregate [2.1.3/4]. We will further investigate this to understand how the transport influences the structure and stability of the forming thrombus. We have carried out experiments to compare the simulation predictions to measured data of platelet aggregation intensity in an artificial stenosis (Fig. 2.1.3/1A). On the other hand, we have investigated cellular transport and platelet deposition in a steeply curved channel to go beyond the well-established cases of straight channel setups. We found that the platelet binding activity was significantly altered in the curved section of the vessel [2.1.3/5], where both deposition structure and intensity is influenced.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

Code development, deployment and user resources

We are performing ongoing measurements to evaluate both the strong- and weak-scaling efficiency of the code over new architectures in collaboration with WP4 (Task 4.3 and 4.6). The method used for computing the flow field (lattice Boltzmann method), is a good fit for large parallel deployments, however, it can also become memory bandwidth bound on the newest HPC architectures, which typically present high CPU core counts per node, increasing the computational capacity, however decreasing the available memory bandwidth per core.

The build system has been rewritten to be more modular, and to provide wider compatibility with HPC machines. These efforts have been detailed further in WP5 (under Task 5.3).

Furthermore, the user documentation has been significantly extended based on external user feedback from the last two years and also in relation to new developments that provide more advanced boundary conditions. The detailed list of available user resources can be found in Table 10/2.1.3 in the Appendix.

6.1.4 Subtask 2.1.4 – An *in silico* approach to guide clinical intervention in treatment of stroke

Short context and research progress

UNIGE collaborates with associate partner Prof. Karim Zouaoui Boudjeltia from University Hospital in Charleroi, Belgium, and Prof R. Dutta from the Department of Statistics at the University of Warwick, UK, in order to develop new platelet function tests. Our target is first to develop a new procedure, combining patient data, numerical models and HPC parameter inference tools. In a second step the goal is to design new medical devices with a commercial partner. The main element of our approach is a correct description of the transport of platelets in blood. This requires taking into account the deformation of red blood cells and their interaction with platelets. Only very demanding HPC numerical simulations can answer this question. Indeed, we demonstrated that platelets may not obey the commonly accepted advection-diffusion behavior. In a second step, we developed a numerical model of platelet deposition based of their adhesion and aggregation properties. These *a priori* unknown parameters can be inferred from deposition patterns obtained from patient blood, using the so-called Approximate Bayesian Computation. Our results reveal that we can identify which properties of platelets may be impacted in the cases of various diseases (e.g., diabetes, COPD, ...). Our approach, deeply rooted in the synergy between *in vitro* and *in silico* approaches, opens a new, promising approach to improve substantially our diagnosis and treatment capabilities related to blood diseases.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

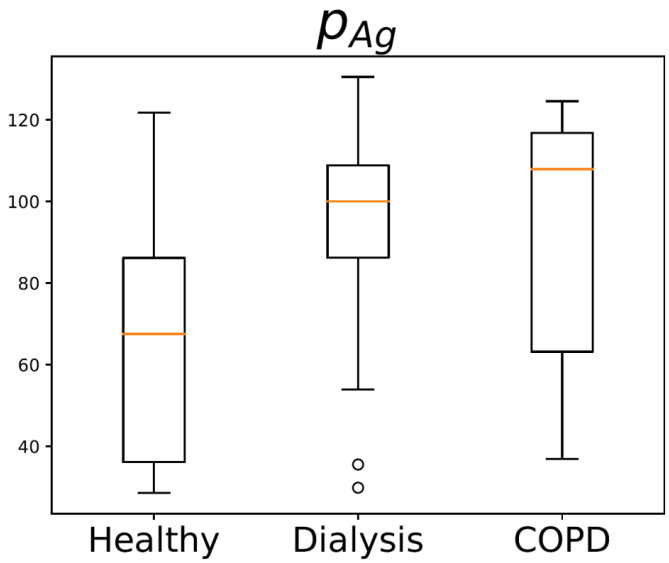


Figure 2.1.4/1 Aggregation probability of healthy and sickle cells.

Code development, deployment and user resources

Improving numerical models for immersed boundary conditions.
Cell-level modeling of blood components requires implementing a coupled liquid-solid framework with fluid-structure interaction. Special attention is devoted to the way the liquid-solid interface is modelled, using an appropriate boundary condition. It is common to use a so-called immersed boundary condition, which is quite robust and relatively simple to deploy, at the cost of the accuracy of the boundary representation. Our research focuses on an alternative, a one-point curved boundary condition, which allows more accurate representation of the boundary at a reduced computational expense, as the algorithm is particularly well adapted to the simulation framework of GPUs.

The detailed list of available user resources can be found in Table 10/2.1.4 in the Appendix.

6.1.5 Subtask 2.1.5 – Developing modelling techniques for *in silico* trials

Short context and research progress

Drug-induced arrhythmias are a major health issue worldwide [2.1.5/1]. This issue was stressed during the COVID-19 pandemic with the use of potentially pro-arrhythmic drugs and the combination of them as an urgent attempt to provide initial treatments to control the disease [2.1.5/2].

Currently, there are no predictors that can provide critical *a priori* information regarding the potential risks for certain patients with normal QTc intervals to develop QT-prolongation after the administration of one or various drugs that may have cardio-toxic side-effects. This need for reliably pro-arrhythmic risk prediction, became relevant during early phases of the Sars-CoV2 pandemic, when it was uncertain whether the uses of hydroxychloroquine (HCQ) and/or azithromycin (AZM) could be more harmful, due to their reported cardiotoxic effects. [2.1.5/3, 2.1.5/4].

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

It is also well documented that males and females present with different risks for drug-induced arrhythmias and QT-interval prolongation due to sex-specific hormones [2.1.5/5, 2.1.5/6]. Furthermore, the function of various cardiac ion channels can be significantly modified by environmental conditions (i.e. hormones, electrolyte concentrations and pH), which can in turn have substantial effects on the overall electrical profile. In the case of COVID-19, hypokalemia has been identified as a prevalent condition in patients, which may increase further the risk of QT prolongation [2.1.5/7]. Further, the combined administrations of several drugs increases the complexities for understanding the associated clinical implications; which is consistent with very limited information relative to drug interactions as drug combinations increase. In many cases, these interactions can be characterized according to most logical biophysical assumptions: potentiation or addition.

A variety of computational methods have been an important component for the study and assessment of drug-induced arrhythmias [2.1.5/8, 2.1.5/9, 2.1.5/10, 2.1.5/11]. However, full human heart biventricular anatomies at a population level, had never been employed.

The primary objective of this work was to create an electrophysiologically "normal" virtual population. No established computational methodology existed to achieve that aim. The second objective was to assess if the administration of a single drug or combinations of drugs within this normal population can reproduce the arrhythmic risk observed clinically. A third objective was to employ existent pharmacodynamic and pharmacokinetic information of the two drugs in question and assess if their measured effects when applied to the human virtual population could provide meaningful clinical information regarding cardiotoxicity. The fourth objective was to develop a reliable methodology to identify the risks of cardiotoxicity and classify the responses within the virtual population to the administered drugs. The work performed by the Barcelona Supercomputing Center describes the creation of a computational framework that employs a gender-specific, cardiac population to assess drug-induced QT-prolongation and arrhythmic risk using Alya. The methodology is shown in Figure 2.1.5/1.

Results were compared to clinical trials recently published, showing remarkably similar results (21% clinical risk vs 21.8% *in silico* trial risk). The *in silico* clinical trial was capable of predicting drug-induced arrhythmic risk effects of HCQ and AZM as single or combined drugs in a normal cohort, with remarkable similar conclusions to clinical trials. The novel methodology developed is capable of providing quantitative evidence of drug-induced arrhythmic risk in a timely manner when no clinical data is available. It can provide evidence of the normal phenotype variants that produce distinct drug-induced arrhythmogenic outcomes. It can also provide the type of arrhythmic risk triggered by drug administration as single or combined doses.

The paper is currently under peer review and can be found as a preprint in MedRxiv [2.1.5/12].

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

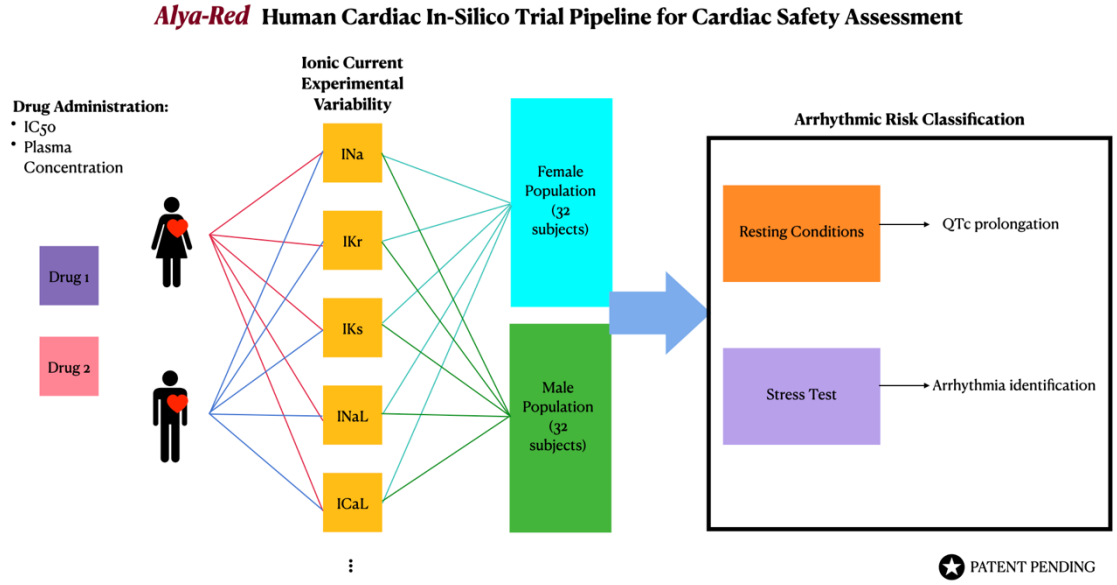


Figure 2.1.5/1. In-silico clinical trial pipeline to assess drug-induced cardiotoxicity

Code development, deployment and user resources

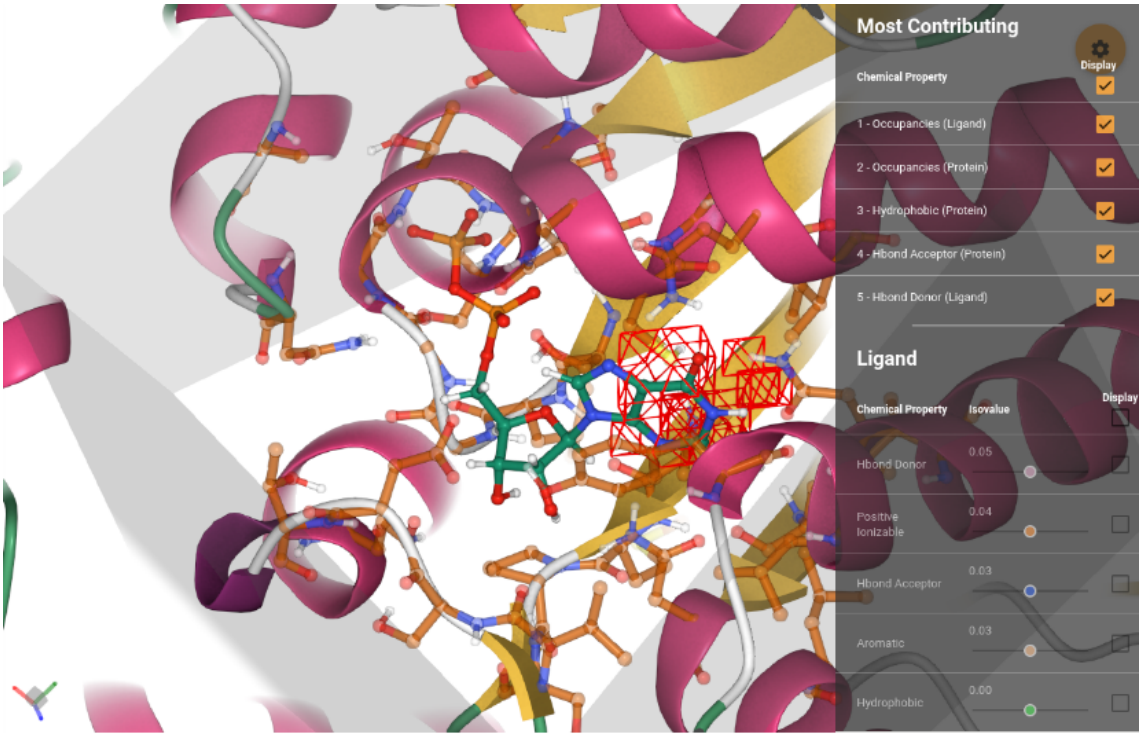
A total of 896 simulations were run using Alya, each one using 640 cores on the Joliot-Curie Rome supercomputer, hosted by GENCI at CEA, France. The population was run as an ensemble composed of 64 simulations of an approximate 2560 core/hours each. Computation time was provided by PRACE via its PRACE-COVID-19 fast track pandemic response (project COVID1933). An approximate total of 2.3 million core/hours were employed for the study of 6 drug interventions, plus the baseline population.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

6.2 Task 2.2 – Molecularly-based Medicine Research Exemplar

6.2.1 Subtask 2.2.1 – Testing virtual candidate libraries against novel targets

Short context and research progress



Virtual screening aims to scan a large library of compounds against a target of interest, ranking them by their binding affinity. In the last few years, modern machine learning algorithms like convolutional neural networks (CNN) have been used to better estimate the binding affinity of a given compound to a protein target. Despite achieving state-of-the-art accuracy at predicting K_D [1], these networks constitute black boxes whose prediction cannot be interpreted, hampering the scientific value of such tools. Recently, a significant amount of effort has been devoted to solving this problem, and new methods have been developed to improve the interpretability of neural networks. One such method is Integrated Gradients (IG) [2], which provides, for each input feature, a value representing how relevant that feature is to the prediction. This leads to a heatmap of attributions where the most contributing features are highlighted.

We applied this same approach to K_{DEEP}, a 3D CNN which takes as input a voxelized representation of a protein-ligand complex and returns a binding affinity prediction. After applying the IG method to the input, we can identify which atoms or interactions are more relevant to the binding affinity estimate. This provides two benefits: (1) improves the interpretability of the model and (2) helps to validate the network, as we can evaluate if the atoms identified as most relevant match the expectations of a chemist (atoms involved in salt bridges, hydrogen bonds, etc.). Our results suggest that CNNs seem to attribute more importance to classical protein-ligand interactions, such as hydrogen bonds or pi-stacking. We call this interpretability method Glimpse and it is available at https://www.playmolecule.com/Glimpse/.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

Code development, deployment and user resources

The website of the application (https://www.playmolecule.com/Glimpse/) provides a functional description and examples to get the users started. Further user resources can be found in Table 10/2.2.1 in the Appendix.

6.2.2 Subtask 2.2.2 – Assessing sequence influence on drug binding using free energy calculations

Short context and research progress

We have recently performed our TIES (thermodynamic integration with enhanced sampling) approach to study the free energy changes caused by protein mutations, a TIES variant we call TIES-PM. TIES-PM calculations involve an alchemical mutation between two amino acids. Estrogen receptor (ER) is expressed in about 70% of breast cancers, and plays a vital role in the initiation and progression of breast cancers. Four residue mutations identified in sequencing study for ER are selected for the TIES-PM simulations: L384V, L387R, K529N and R548P. We have also performed a control study using mutations of which experimental data is available, which confirms the reliability of our TIES-PM predictions. The predicted binding free energies provide a clear explanation for the effects of these mutations. The mutations at the binding site, L384V and L387R, induce resistance to the drugs studied; the mutations L387R and R548P play an important role in the activation of the protein.

We are aiming to extend our work on the influence of sequence on binding free energies by investigating the effect of the resistance of *Mycobacterium tuberculosis* to rifampin. Rifampin is a critical component of the multi-drug regimens used to treat Tuberculosis (TB). Whilst development of resistance to rifampin is rare, compared to other anti-TB drugs, it often occurs in conjunction with other resistances leading to higher instances of treatment failure and mortality. Understanding of which mutations confer resistance is key to drug susceptibility testing and improving patient outcomes. Rifampin resistance is mediated by mutations in the *rpoB* gene, which codes for the beta-subunit of RNA polymerase, thus our work will study the relative binding free energies associated with amino acid mutations in the *rpoB* protein.

In a drug repurposing study for the main protease ($3CL^{pro}$) in Covid-19, we have investigated a set of drugs which are highly diverse. At the substrate-binding site, the drugs interact with different subsites; residues at the binding site are expected to contribute differently to the bindings of the drugs. A per-residue energy decomposition suggests the amino acids crucial for the binding of drugs at the $3CL^{pro}$ binding sites. The detailed features of the binding sites are important for the building of an energy-based pharmacophore model. The combination of energy-based and structure-based pharmacophore models could provide an improved virtual screening for the initial selection of promising compounds.

Code development, deployment and user resources

We have been working on the extension of the TIES-PM methodology to include the GPU version of GROMACS. Using GROMACS allows for the exploitation of GPU based HPCs such as Longhorn at TACC as well as allow for the efficacy of enhanced sampling methods, such as REST2, to be tested when applied to these sequence influence studies.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

By combining the precision of physics-based methods using MD simulations with machine learning (ML) techniques, we have also worked in collaboration with Rutgers University and Argonne National Laboratory (International Partners) to develop the IMPECCABLE workflow (Integrated Modeling PipelinE for COVID Cure by Assessing Better LEads) [2.2.2/5, 2.2.2/6, 2.2.2/7]. IMPECCABLE provides a scalable, flexible, and extensive infrastructure for campaigns designed to discover improved leads targeted at SARS-CoV-2. The scale of the potential resulting workflow is such that it is dependent on supercomputing to achieve extremely high throughput. We have demonstrated the viability of this workflow for the study of inhibitors for key COVID-19 target proteins and our ability to perform the required large-scale calculations to identify lead antiviral compounds through repurposing on a variety of supercomputers.

A public release of our TIES20 protocol has recently been realised; this includes both an input dual-topology builder and alchemical methodology to perform relative binding free energy calculations with NAMD and OpenMM. These are provided through a website to build the input files and installable packages to run the calculations along with relevant online documentation and tutorials. Further currently available user resources can be found in Table 10/2.2.2 in the Appendix.

6.2.3 Subtask 2.2.3 – Combining machine learning and molecular modelling to assess drug binding

Short context and research progress

The accurate description of molecular potential energy is crucial for computational chemistry and molecular simulations. Machine learning algorithms have been proposed as an alternative for the existing potentials. Particularly, the ones based on neural networks, called neural network potentials (NNP) have been gaining a lot of interest. However, the compatibility of current simulation engines with NNPs is still limited, hindering the development and testing of novel NNPs.

To solve this, we have been developing TorchMD [2.2.3/1], a new molecular simulation application based on the high-performance deep learning python library PyTorch [2.2.3/2]. TorchMD allows performing all the operations to simulate and analyze a machine-learned potential and its potential priors using a Torch backend. All computations are expressed as Torch arrays, including bonds, angles, dihedrals, Lennard-Jones, and Coulomb interactions. Also, it is possible to run standard all-atom using the Amber forcefield.

In combination with TorchMD, we have also been developing TorchMD-net, a PyTorch-based machine learning framework for training and developing NNPs. TorchMD-net is built upon a scalable multi-node training code, using PyTorch Lightning [2.2.3/3]. Currently, TorchMD-net has implemented two types of neural network models to train on, graph-based models and transformer models. Both models are implemented using PyTorch geometric [2.2.3/4] in order to optimize training speed.

By combining both frameworks, we have been able to work on several applications related to NNPs. The first one consists of learning coarse-grained models of atomistic systems, particularly on protein folding, using a graph neural network, based on the Schnet scheme [2.2.3 /5], which is able to correctly fold several fast-folding proteins and to reproduce the thermodynamics of each system. Secondly, we have also started to work on end-to-end differentiable simulations,

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

thanks to the automatic differentiation nature of TorchMD, which allows for an automatic differentiation of molecular dynamics observables with respect to potential energy parameters. End-to-end differentiable simulations could be used to improve, for example, the accuracy of trained NNPs when simulating systems outside of the training dataset.

Code development, deployment and user resources

TorchMD provides several tests and a tutorial. Code and examples are available for free through GitHub: https://github.com/torchmd/torchmd.
TorchMD-net is also available through GitHub: https://github.com/torchmd/torchmd-net. The transformer model is not yet accessible, but will be published in the near future.
Further currently available user resources can be found in Table 10/2.2.3 in the Appendix.

6.3 Task 2.3 – Neuro-musculoskeletal Research Exemplar

6.3.1 Subtask 2.3.1 – Growth and Adaptation

Short context and research progress

The aim of this subtask is to develop an efficient model to describe and simulate bone growth and adaptation, also in response to pharmaceutical treatments.

As a first step, we developed a phenomenological model to describe volumetric bone mineral density (vBMD) variation of the mesh elements in dependence on the proximal femur areal bone mineral density (aBMD), using a simple linear regression model informed by 96 patients from the Sheffield cohort. Briefly, element vBMDs were modified (with 3 different slopes for high, medium, and low mineral density elements) to obtain a predefined proximal femur aBMD value. By simulating aBMD loss due to ageing, we could update the FE material properties of a femur model to estimate its appearance up to 10 years after the actual CT scan recording.

On the other hand, a cell-level agent-based model could be used to generate data for bone remodelling with a physiology-driven approach, but the validation of such models would require processing of a large number of high-resolution images, such as those generated by synchrotron facilities. For this purpose, the parallel Framework for Image REgistration (pFIRE) has been developed, based on the registration algorithm developed by [2.3.1/1, 2.3.1/2] and targeting HPC architectures. Its current development is focused on improving scalability and performance to exploit future computational architectures.

Code development, deployment and user resources

pFIRE is open source and relies on advanced numerical libraries (PETSc, Boost) for parallelization and efficiency.

It is deployed on HPC systems at the University of Sheffield and is available to be used out-of-the-box via Docker containers or Singularity: https://hub.docker.com/r/insigneopfire/pfire

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

The current number of pFIRE users is estimated as 10-50 researchers. In the future, a mechanism will be adopted to register and trace the user base, with a dedicated mailing list.

Instructions on how to install and configure the pFIRE are available on the main Github page: https://github.com/INSIGNEO/pFIRE

6.3.2 Subtask 2.3.2 - ARF10: Multiscale model of hip fracture

Short context and research progress

The aim of this task is the creation of a model to predict the proximal femur fracture risk at 10 years, given a proximal femur CT scan, aBMD, age, height, and weight at time 0. To do so, the idea is to use the original CT scan to create 10 models (one per year), with appropriate bone remodelling effects as estimated by the growth and adaptation model developed in Subtask 2.3.1.

We automated the CT2S/ARF0 pipeline and integrated the growth algorithm into it to simulate bone ageing. The 10 models are then independently simulated, and the resulting ARF0 are eventually combined to estimate ARF10.

We further improved the ARF0 pipeline by including three-dimensional, subject-specific characterization of soft-tissue thickness (STT) in the hip region [2.3.2/1]. This leads to a hip fracture classification accuracy of 87%, up from 85% previously [2.3.2/2] where STT was characterized using a population-based (i.e. non-subject-specific) regression and its three-dimensional variation was neglected.

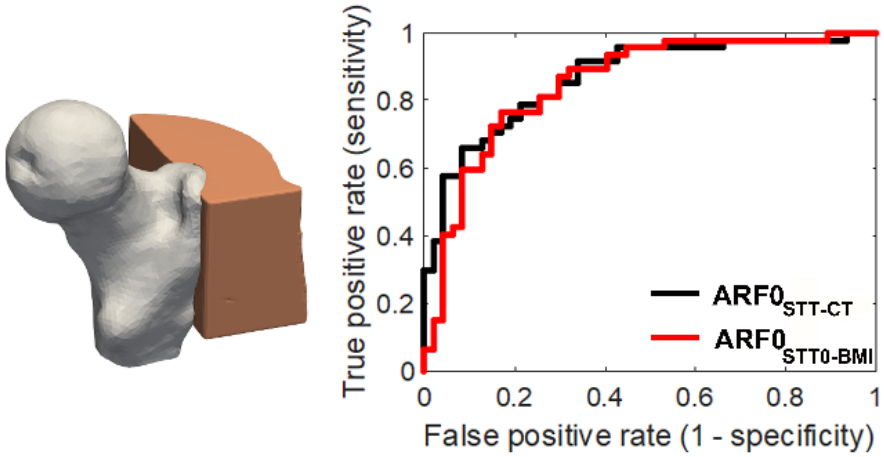


Figure 2.3.2/1: Left: a representative proximal femur with soft tissues overlying the trochanteric region. Right: ROC analysis performance curves using $ARF0_{STT0\text{-}BMI}$ (red) and $ARF0_{STT\text{-}CT}$ (black) as classifiers

Code development, deployment and user resources

The following core documents are provided for the Computed Tomography to Strength (CT2S) application. These are available publicly on the CT2S website (https://ct2s.insigneo.org/ct2s/):

- CT2S service description: includes scientific background, data generation, data transferring and reporting procedure;
- CT scan protocol

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

- CT service presentation: slides with overview of each stage of the pipeline

In addition to the above documentation, we also have detailed tutorials with step by step guide for the pre-processing pipeline. These tutorials are available to students who take BIE6433 module at Sheffield, and students who attend the medics summer school. The tutorial materials are available to share upon request (e.g. via Contact us on the CT2S website). These and further available user resources are listed in Table 10/2.3.2 in the Appendix.

6.3.3 Subtask 2.3.3 – Virtual patients' expansion

Short context and research progress

The aim of this task is the expansion of a physical cohort with synthetic subjects to be added to or to replace the physical ones, in order to obtain a large cohort (~1000 patients) suitable for simulation of phase III clinical trials (*In Silico* Trials).

As a first step, a PCA-based statistical anatomy atlas has been developed starting from 94 proximal femurs (47 fractured and 47 non-fractured) of the Sheffield cohort [2.3.3/1], and it was used to generate over 1000 virtual patients. We applied the CT2S/ARF0 pipeline to the virtual patients and compared the results with the original population, finding that the virtual population was still composed of approximately 50% fractured and 50% non-fractured subjects, although the fractured sub-population was more fragile than the original one [2.3.3/2]. Thus, data from real phase II clinical trial could be used to create synthetic cohorts suitable for phase III *In Silico* Trials.

Also, we developed a patient selection algorithm to create cohorts with custom aBMD distributions, in order to match some desired real population characteristics. We created two 500-patient cohorts to mimic low-risk and high-risk subpopulations of past clinical trials, and used the aforementioned ARF10 pipeline to estimate long-term fracture risk. In order to create a more realistic clinical trial simulation, we are working on a Markov-Chain version of the cohort ageing and falling over 10-years time.

Code development, deployment and user resources

Virtual cohort 10-years simulation code is deployed on Cartesius (SURF, The Netherlands) and Galileo/Galileo100 (CINECA, Italy). The biggest run involved 7200 cores (300 thin nodes) on Cartesius. We are working on the use of a dedicated pilot job manager (QCG-PilotJob) to increase CPU efficiency. This application shares the code components with the previous subtask, and the available user resources are listed in the same table (Table 10/2.3.2 in the Appendix).

7 Risk Management

The following possible sources of risks have been identified during the previous deliverable (D2.1), since then no new potential source has been identified.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

a) The internal budget of HPC core-hours might not be enough to cover all the VVUQ analyses defined in Task 2.5.

Probability	Medium
Impact	Medium
Risk assessment	Medium
Mitigation	The progress of VVUQ application deployment is monitored continuously together with the still available internal budget, and the partners report on the advancements every month. For now, the consortium is evaluating possible ways to provide access to more core-hours for VVUQ, and partners are encouraged to also apply for external budgets individually. The first deliverable on VVUQ strategy (D2.3) has also considered this. There are further allocations on ARCHER2 as part of a new project funded for three years from UCL (SEAVEA (https://www.ucl.ac.uk/news/2021/jul/ps2m-grant-boost-supercomputing)

b) Delays in the proposed work due to COVID-related reduced interactions (e.g. reduced mobility, such as on-site visits between partners).

Probability	Low
Impact	Low
Risk assessment	Low
Mitigation	All research tasks using direct experimental data or relying on lab access and mobility have been coping well with the current situation. A prompt migration to online collaborative tools and the surge in the availability of such tools has helped to avoid any delays so far. We continue to keep an eye on the situation as it evolves.

c) Insufficient or non-timely input from partners delaying the deliverables.

Probability	Low
Impact	Medium
Risk assessment	Medium
Mitigation	The progress of the deliverables is checked internally on a regular basis. The regular work package meetings and the intra-workpackage teleconferences and discussions facilitate the information exchange and allow the WP leader, or if needed the Project Manager, to step in and act to mitigate the problem.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

8 Conclusions

This deliverable reports progress in scientific research performed since the previous deliverable (D2.1) and provides information on the currently available user resources that enable external users to adopt solutions from our application portfolio seamlessly.

The user resources explained at the end of each task description, the publicly available guides, tutorials, and videos listed in the appendix, and finally the user guides already available at the CompBioMed website collectively form the first version of our “best practices”. These documents cover the whole software portfolio developed within the project, and they aim to facilitate the adoption of our codes towards external users.

9 Bibliography/References

Please make use of bibliography and references rather than footnotes, these should be checked rigorously before publication.

[Introduction/1] A. G. Hoekstra, B. Chopard, D. Coster, S. Portegies Zwart, P. V. Coveney, "Multiscale Computing for Science and Engineering in the Era of Exascale Performance", Phil Trans R Soc A, 377 (2019), DOI: 10.1098/rsta.2018.0144

[Introduction/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., DOI: 10.1016/j.jocs.2017.07.004

[2.1.1/1] HemeLB Website – hemelb.org

[2.1.1/2] HemeLB Code Repositories - https://github.com/hemelb-codes (main), https://github.com/UCL-CCS/HemePure,https://github.com/UCL-CCS/HemePure_SelfCoupled

[2.1.1/3] J. W. S. McCullough, R. A. Richardson, A. Patronis, R. Halver, R. Marshall, M. Ruefenacht, B. J. N. Wylie, T. Odaker, M. Wiedemann, B. Lloyd, E. Neufeld, G. Sutmann, A. Skjellum, D. Kranzlmüller and P. V. Coveney, "Towards blood flow in the virtual human: efficient self-coupling of HemeLB", J R Soc Interface Focus 11, 20190119 (2020), DOI:10.1098/rsfs.2019.0119

[2.1.1/4] J. W. S. McCullough and P. V. Coveney, " High fidelity blood flow in a patient-specific arteriovenous fistula", Scientific Reports (UNDER REVIEW), https://arxiv.org/abs/2012.04639v2

[2.1.1/5] E. Mayer, S. Cielo, J. McCullough, J. Gunter, P. Coveney, “Visualization of human-scale blood flow simulation using Intel OSPRay Studio on SuperMUC-NG”, Presentation at Intel Dev Summit – ISC2021, 23 June 2021, https://www.oneapi.io/event-sessions/visualization-of-human-scale-blood-flow-tech-talk/

[2.1.2/1] doi:10.1111/jon.12535

[2.1.2/2] doi:10.1002/cnm.2882

[2.1.2/3] doi:10.1016/j.biomech.2019.04.019

[2.1.3/1] Závodszky, G., van Rooij, B., Azizi, V., & Hoekstra, A. (2017). Cellular level in-silico modeling of blood rheology with an improved material model for red blood cells. Frontiers in physiology, 8, 563.

[2.1.3/2] Czaja, B., Gutierrez, M., Závodszky, G., de Kanter, D., Hoekstra, A., & Eniola-Adefeso, O. (2020). The influence of red blood cell deformability on hematocrit profiles and platelet margination. PLoS computational biology, 16(3), e1007716.

*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **823712**“*

[2.1.3/3] de Vries, K., Nikishova, A., Czaja, B., Závodszky, G., & Hoekstra, A. G. (2020). Inverse Uncertainty Quantification of a cell model using a Gaussian Process metamodel. International Journal for Uncertainty Quantification, 10(4).

[2.1.3/4] van Rooij, B. J. M., Závodszky, G., Hoekstra, A. G., & Ku, D. N. (2021). Haemodynamic flow conditions at the initiation of high-shear platelet aggregation: a combined in vitro and cellular in silico study. Interface focus, 11(1), 20190126.

[2.1.3/5] Spieker, C., Zavodszky, G., Mouriaus, C., van der Kolk, M., Gachet, C., Mangin, P., Hoekstra, A. G., (Under review), The influence of micro-vessel curvature on cellular flow conditions for platelet adhesion, Annals of Biomedical Engineering

[2.4.1/1] Boudjeltia, Karim Zouaoui, Christos Kostalos, Daniel Ribeiro, Alexandre Rousseau, Christophe Lelubre, Olivier Sartenaer, Michael Piagnerelli et al. "Spherization of red blood cells and platelets margination in COPD patients." medRxiv (2020).

[2.4.1/2] Dutta, Ritabrata, Karim Zouaoui-Boudjeltia, Christos Kotsalos, Alexandre Rousseau, Daniel Ribeiro de Sousa, Jean-Marc Desmet, Alain Van Meerhaeghe, Antonietta Mira, and Bastien Chopard. "Interpretable pathological test for Cardio-vascular disease: Approximate Bayesian computation with distance learning." arXiv preprint arXiv:2010.06465 (2020).

[2.4.1/3] Kotsalos, Christos, Karim Zouaoui Boudjeltia, Ritabrata Dutta, Jonas Latt, and Bastien Chopard. "Anomalous platelet transport & fat-tailed distributions." arXiv preprint arXiv:2006.11755 (2020).

[2.4.1/4] Marson, Francesco, Yann Thorimbert, Bastien Chopard, Irina Ginzburg, and Jonas Latt. "Enhanced single-node lattice Boltzmann boundary condition for fluid flows." Physical Review E 103, no. 5 (2021): 053308.

[2.1.5/1] Sarganas G, et al. Epidemiology of symptomatic drug-induced long QT syndrome and torsade de pointes in Germany, *EP Europace*, 2014;16(1):101–108.

[2.1.5/2] Cavalcanti AB, et al; Coalition Covid-19 Brazil I Investigators. Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19. N Engl J Med. 2020 Nov 19;383(21):2041-2052. Erratum in: N Engl J Med. 2020 Nov 19;383(21):e119.

[2.1.5/3] White, Nicholas J. Cardiotoxicity of antimalarial drugs. The Lancet Infectious Diseases; 7(8):549-558.

[2.1.5/4] Ray, Wayne A. et al. Azithromycin and the risk of cardiovascular death. N Engl J Med. 2012. 366(20): 1881-1890.

[2.1.5/5] Salama G, Bett GC. Sex differences in the mechanisms underlying long QT syndrome. Am J Physiol Heart Circ Physiol. 2014 Sep 1;307(5):H640-8.

[2.1.5/6] Vink AS, et al. Effect of age and gender on the QTc-interval in healthy individuals and patients with long-QT syndrome, Trends in Cardiovascular Medicine, Volume 28, Issue 1, 2018, Pages 64-75.

[2.1.5/7] Chen D, et al. Assessment of Hypokalemia and Clinical Characteristics in Patients With Coronavirus Disease 2019 in Wenzhou, China. JAMA Netw Open. 2020 Jun 1;3(6):e2011122.

[2.1.5/8] Yang PC, et al. A Computational Pipeline to Predict Cardiotoxicity: From the Atom to the Rhythm. Circ Res. 2020 Apr 10;126(8):947-964.

[2.1.5/9 Bottino D, et al. Preclinical cardiac safety assessment of pharmaceutical compounds using an integrated systems-based computer model of the heart. Prog Biophys Mol Biol. 2006 Jan-Apr;90(1-3):414-43.

[2.1.5/10] Delaunois A, et al. Applying the CiPA approach to evaluate cardiac proarrhythmia risk of some antimalarials used of-label in the first wave of COVID-19. Clinical and Translational Science. Accepted Author Manuscript. Https://doi.org/10.1111/cts.13011

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

[2.1.5/11] Beattie et al. Evaluation of an in silico cardiac safety assay: using ion channel screening data to predict QT interval changes in the rabbit ventricular wedge. J Pharmacol Toxicol Methods. 2013 Jul-Aug;68(1):88-96.
[2.1.5/12] Aguado-Sierra et al. In-silico clinical trial using high performance computational modeling of a virtual human cardiac population to assess drug-induced arrhythmic risk. medRxiv 2021.04.21.21255870; doi: https://doi.org/10.1101/2021.04.21.21255870
[2.2.1/1] J. Chem. Inf. Model. 2018, 58, 2, 287–296
[2.2.1/2] Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic Attribution for Deep Networks. 2017
[2.2.2/1] 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., 15, 1265–1277 (2019), DOI: 10.1021/acs.jctc.8b01118
[2.2.2/2] S. Wan, D. Kumar, V. Ilyin, U. Al Homsi, G. Sher, A. Knuth, P. V. Coveney, "The effect of protein mutations on drug binding suggests ensuing personalised drug selection", Sci. Rep., 11, 13452 (2021), DOI: 10.1038/s41598-021-92785-w
[2.2.2/3] S. Wan, A. Bhati, A. Wade, D. Alfe, P. V. Coveney. “Thermodynamic and structural insights into the repurposing of drugs that bind to SARS-CoV-2 main protease”. ChemRxiv. Cambridge: Cambridge Open Engage (2021), DOI: 10.33774/chemrxiv-2021-03nrl
[2.2.2/4] https://www.ties-service.org
[2.2.2/5] A. Al Saadi, D. Alfe, Y. Babuji, A. Bhati, B. Blaiszik, A. Brace, T. Brettin, K. Chard, R. Chard, A. Clyde, P. V. Coveney, I. Foster, T. Gibbs, S. Jha, K. Keipert, T. Kurth, D. Kranzlmüller, H. Lee, Z. Li, H. Ma, A. Merzky, G. Mathias, A. Partin, J. Yin, A. Ramanathan, A. Shah, A. Stern, R. Stevens, L. Tan, M. Titov, A. Trifan, A. Tsaris, M. Turilli, H. Van Dam, S. Wan, D. Wifling, “IMPECCABLE: Integrated Modeling PipelinE for COVID Cure by Assessing Better LEads”, 50th International Conference on Parallel Processing (ICPP ’21), August 9-12 (2021), DOI: 10.1145/3472456.3473524
[2.2.2/6] H. Lee, A. Merzky, L. Tan, M. Titov, M. Turilli, D. Alfe, A. Bhati, A. Brace, A. Clyde, P. V. Coveney, H. Ma, A. Ramanathan, R. Stevens, A. Trifan, H. Van Dam, S. Wan, S. Wilkinson, S. Jha, “Scalable HPC & AI Infrastructure for COVID-19 Therapeutics”, Platform for Advanced Scientific Computing Conference (PASC ’21), July 5-9 (2021), DOI: 10.1145/3468267.3470573
[2.2.2/7] A. P. Bhati, S. Wan, D. Alfe, A. R. Clyde, M. Bode, L. Tan, M. Titov, A. Merzky, M. Turilli, S. Jha, R. R. Highfield, W. Rocchia, N. Scafuri, S. Succi, D. Kranzlmüller, G. Mathias, D. Wifling, Y. Donon, A. Di Meglio, S. Vallecorsa, H. Ma, A. Trifan, A. Ramanathan, T. Brettin, A. Partin, F. Xia, X. Duan, R. Stevens, P. V. Coveney, "Pandemic Drugs at Pandemic Speed: Infrastructure for Accelerating COVID-19 Drug Discovery with Hybrid Machine Learning and Physics-based Simulations on High Performance Computers", Interface Focus, in press (2021) arXiv:2103.02843
[2.2.3/1] Doerr, Stefan, et al. "Torchmd: A deep learning framework for molecular simulations." Journal of chemical theory and computation 17.4 (2021): 2355-2363.
[2.2.3/2] Paszke, Adam, et al. "Pytorch: An imperative style, high-performance deep learning library." Advances in neural information processing systems 32 (2019): 8026-8037.
[2.2.3/3] Falcon, W. E. A. "Pytorch lightning." GitHub. Note: https://github. com/williamFalcon/pytorch-lightning, 2019.
[2.2.3/4] Fey, Matthias, and Jan Eric Lenssen. "Fast graph representation learning with PyTorch Geometric." arXiv preprint arXiv:1903.02428 (2019).
[2.2.3/5] Husic, Brooke E., et al. "Coarse graining molecular dynamics with graph neural networks." The Journal of Chemical Physics 153.19 (2020): 194101.

*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **823712**“*

[2.3.1/1] D. Barber, D. Hose (2005). Automatic segmentation of medical images using image registration: diagnostic and simulation applications. Journal of medical engineering & technology, 29(2):53-63. DOI: 10.1080/03091900412331289889.
[2.3.1/2] D. C. Barber, E. Oubel, A. F. Frangi, D. Hose (2007). Efficient computational fluid dynamics mesh generation by image registration. Medical image analysis, 11(6):648–662. DOI: 10.1016/j.media.2007.06.011.
[2.3.2/1] A. Aldieri, M. Terzini, A. Audenino, C. Bignardi, M. Paggiosi, R. Eastell, M. Viceconti, P. Bhattacharya (2021). Personalised 3D assessment of trochanteric soft tissues improves hip fracture classification accuracy, submitted. DOI:10.17605/osf.io/dzmkh
[2.3.2/2] P. Bhattacharya, Z. Altai, M. Qasim, M. Viceconti (2019). A multiscale model to predict current absolute risk of femoral fracture in a postmenopausal population. Biomechanics and Modelling in Mechanobiology, 18:301–18. DOI:10.1007/s10237-018-1081-0
[2.3.3/1] M. Taylor, M. Viceconti, P. Bhattacharya, X. Li (2021). Finite element analysis informed variable selection for femoral fracture risk prediction. Journal of the Mechanical Behavior of Biomedical Materials, 118:104434. DOI: 10.1016/j.jmbbm.2021.104434.
[2.3.3/2] A. A. La Mattina, M. Taylor, F. Baruffaldi, M. Viceconti (2021). Statistical properties of virtual cohorts for In Silico Trials generated with statistical anatomy atlas. In submission to CPT: Pharmacometrics & Systems Pharmacology.

10 Appendix

10.1 Available User Resources

The following tables summarize the current publicly available user resources for our application portfolio.

Table 10/2.1.1. HemeLB user resources

Code	Resource	Reference, link, doi
HemeLB	CompBioMed hosted documentation and tutorial	https://compbiomedeu.github.io/
	Tutorial on website	http://hemelb.org/tutorials/
	CompBioMed webinar on HemeLB (including tutorial)	https://www.compbiomed.eu/compbiomed-webinar-10/
	HemeLB GitHub repository (includes various implementations of the code)	https://github.com/hemelb-codes
	Papers describing the main numerical components and	http://dx.doi.org/10.1016/j.cpc.2008.02.013 http://dx.doi.org/10.1103/PhysRevE.89.023303 http://dx.doi.org/10.1016/j.jocs.2013.03.002 https://doi.org/10.1098/rsfs.2019.0119

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

	computational aspects	

Table 10/2.1.2. OpenBF user resources

Code	Resource	Reference, link, doi
openBF	Thesis describing the underlying theory	https://etheses.whiterose.ac.uk/19175/
	Papers describing applications	https://doi.org/10.1002/cnm.2882 https://doi.org/10.1016/j.jbiomech.2019.04.019
	Figshare repository	https://figshare.shef.ac.uk/articles/code/openBF_Julia_software_for_1D_blood_flow_modelling/7166183
	Github repository with tutorials	https://github.com/INSIGNEO/openBF
	Video tutorial	https://www.compbiomed.eu/compbiomed-webinar-2/

Table 10/2.1.3. HemoCell user resources.

Code	Resource	Reference, link, doi
HemoCell	Paper describing the main numerical components	https://doi.org/10.3389/fphys.2017.00563
	Documentation and tutorial on the website	https://www.hemocell.eu/user_guide/index.html
	CompBioMed hosted documentation and tutorial	https://compbiomedeu.github.io/
	YouTube videos	https://www.youtube.com/watch?v=r3hLTNC8vpA https://www.youtube.com/watch?v=364MoJjfNdU https://www.youtube.com/watch?v=tcRMb95yOBA https://www.youtube.com/watch?v=d94P5iG2H9Y https://www.youtube.com/watch?v=Y3XjiSf3MOc
	Terminal recordings	https://asciinema.org/a/185921 https://asciinema.org/a/185922

Table 10/2.1.4. Palabos user resources.

Code	Resource	Reference, link, doi

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

Palabos	Article describing the software	https://doi.org/10.1016/j.camwa.2020.03.022
	User Guide	https://bit.ly/38Kg42x
	Presentations: Community application examples	https://palabos.unige.ch/community/palabos-online-seminar-series/
	Online Videos from the Palabos Summer School 2021	https://palabos.unige.ch/class/summer-school/
	Online Videos from the Palabos Summer School 2020	https://palabos.unige.ch/class/recordings-summer-school-2020/

Table 10/2.2.1 Playmolecule user resources

Code	Resource	Reference, link, doi
PlayMolecule	Tutorials	https://medium.com/playmolecule
	Paper describing the main functionality of the apps	https://doi.org/10.1039/c9sc04606b https://www.nature.com/articles/s41598-018-19345-7 https://doi.org/10.1038/s41598-019-50752-6 https://pubs.acs.org/doi/10.1021/acs.jcim.8b00711 https://pubs.acs.org/doi/10.1021/ct900275y https://pubs.acs.org/doi/10.1021/ct9000685
	YouTube channel and YouTube videos	https://www.youtube.com/c/Acelleralive
	Twitter channel	https://twitter.com/acellera
	Software documentation	https://software.acellera.com/docs/latest/index.html

Table 10/2.2.2 BAC user resources

Code	Resource	Reference, link, doi
BAC	Papers describing the main numerical components	https://doi.org/10.1021/ci8000937 https://doi.org/10.1098/rsfs.2020.0007 https://doi.org/10.1098/rsfs.2019.0133
	CompBioMed hosted	https://www.compbiomed.eu/services/software-hub/compbiomed-software-bac/

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

	documentation and tutorial	
	Documentation and tutorial on the website	https://github.com/radical-cybertools/htbac https://htbac.readthedocs.io/en/latest/ https://adw62.github.io/TIES_MD/index.html
	YouTube videos	https://www.youtube.com/watch?v=XHuHLF7baA4 https://www.youtube.com/watch?v=gHzuF2UfhVw https://www.youtube.com/watch?v=rAbf2Z7FFU8&t=18492s https://www.youtube.com/watch?v=8Gnz99MpBak&t=5s

Table 10/2.2.3 TorchMD user resources

Code	Resource	Reference, link, doi
TorchMD	Main code reference	https://pubs.acs.org/doi/abs/10.1021/acs.jctc.0c01343
	Related references	https://pubs.acs.org/doi/abs/10.1021/acscentsci.8b00913 https://aip.scitation.org/doi/abs/10.1063/5.0026133
	Main repository	https://github.com/torchmd
	MD engine repository	https://github.com/torchmd/torchmd
	MD engine tutorial	https://github.com/torchmd/torchmd/blob/master/examples/tutorial.ipynb
	TorchMD-NET repository	https://github.com/torchmd/torchmd-net
	TorchMD-CG repository	https://github.com/torchmd/torchmd-cg
	TorchMD-CG tutorial	https://github.com/torchmd/torchmd-cg/blob/master/tutorial/Chignolin_Coarse-Grained_Tutorial.ipynb

Table 10/2.3.1. pFIRE user resources.

Code	Resource	Reference, link, doi
pFIRE	Paper describing underlying theory	DC Barber and DR Hose 2005 (https://doi.org/10.1080/03091900412331289889), DC Barber et al. 2007 (https://doi.org/10.1016/j.media.2007.06.011)
	Online Documentation	https://insigneo.github.io/pFIRE/docs.html

Table 10/2.3.2 CT2S user resources.

Code	Resource	Reference, link, doi
C**T2S**	Paper describing the main numerical components	doi: 10.1016/j.clinbiomech.2019.06.004 doi: 10.1007/s00198-016-3597-4.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

	Documentation and tutorial on the website	https://ct2s.insigneo.org/ct2s/

Further general information on user resources aimed for training can be found on the CompBioMed training portal (https://www.compbiomed.eu/training-3/).

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*