Abstract. Modern medicine and biology have been transformed into quantitative sciences of high complexity, with challenging objectives. The aims of medicine are related to early diagnosis, effective therapy, accurate intervention, real time monitoring, procedures/systems/instruments optimization, error reduction, and knowledge extraction. Concurrently, following the explosive production of biological data concerning DNA, RNA, and protein biomolecules, a plethora of questions has been raised in relation to their structure and function, the interactions between them, their relationships and dependencies, their regulation and expression, their location, and their thermodynamic characteristics. Furthermore, the interplay between medicine and biology gives rise to fields like molecular medicine and systems biology which are further interconnected with physics, mathematics, informatics, and engineering. Modelling and simulation is a powerful tool in the fields of Medicine and Biology. Simulating the phenomena hidden inside a diagnostic or therapeutic medical procedure, we are able to obtain control on the whole system and perform multilevel optimization. Furthermore, modelling and simulation gives insights in the various scales of biological representation, facilitating the understanding of the huge amounts of derived data and the related mechanisms behind them. Several examples, as well as the insights and the perspectives of simulations in biomedicine will be presented.

1. Introduction

There is a wonderful interplay between Physics, Mathematics, Informatics, Medicine and Biology that drives to a continuous fertile explosion producing other fields like Computational Physics and Mathematics, Biophysics, Medical Physics, Medical Informatics, Bioinformatics as well as Nanomedicine and NanoBiotechnology, AstroBiology and NeuroInformatics. All these scientific fields are interconnected belonging to one or more networks. Their interconnections can be of various types: physical, content-based and concept-based, demonstrating various levels of randomness.

Each of the basic sciences has an initial contribution delineating a large space of new questions and specific needs. To achieve information and data understanding in medicine and biology across the axis prediction/prognosis – diagnosis – therapy – integration, state-of-the-art computational methods are applied. Some of the main objectives of computer-aided Medicine are the early diagnosis of diseases, the effective therapy, the minimization of medical error and the connection to biological data.
towards Molecular Medicine. In parallel, some of the main objectives of computer-aided Biology are the prediction of biomolecules’ properties, the prediction of the biological systems’ properties as well as the investigation of the molecular basis for various diseases.

The related applications to the aforementioned interdisciplinary fields require high performance computing (perhaps several hundred of Petaflops), especially when we are going to systems levels requiring realistic simulations (e.g. simulations of protein machine interactions and molecule-based cell simulations as well). However, up to now, according to the TOP500 list (June 2013), the best available supercomputers present a maximum performance ranging from 2 to 50 Petaflops.

2. Simulations in general
The simulations’ general pipeline starts always by having interesting questions (Figure 1). Afterwards, we have to understand (or believe that we understand) the processes and the phenomena of the systems involved. Then, we make a model that describes our system and we select the proper numerical methods to implement this model. These are translated to code development which is checked for consistency and verification under the prism of a set of predefined expected results. There is a feedback loop here concerning code optimization mainly in technical terms.

![Figure 1. A general pipeline describing simulations’ design and validation](image-url)
Then, the model can be validated via comparisons with experimental results and in the sequel we are able to perform sensitivity analysis and error estimation. Finally, we can proceed to overall optimization by fine tuning the methods, the code and the hardware used for the simulation. By performing scaling of the model we are able to answer whether the more CPUs will proportionally reduce the needed time to complete a long simulation.

Following this pipeline, we reach to results that hopefully help us answer the initial questions. Usually then, we realize the capabilities of the simulation and we haste to put new questions, more exotic than the initial ones.

3. Simulations in Medicine and Biology

Actually, Medicine and Biology are suitable for weird questions. According to Richard Feynman, “everything that living things do can be understood in terms of the jigglings and wigglings of atoms” [1]. As such, simulations in these fields deal with very complex systems, multidimensional feature spaces and multi-scale approaches.

Depending on what is the level of modeling, simulations in medicine and biology can be process driven or data driven. In the first case, the simulations are based on modeling the phenomena/procedures/processes of the real system under investigation. Given a good understanding of the underlying mechanisms, we produce results that should be similar to the experimental ones and vice versa: by tuning our model to fit the experimental results we understand better the underlying mechanisms. In the second case, simulations are based on the experimental data structure and data features, producing similar data to the real ones, preserving the data features and their correlations. In this way, we can populate our samples without necessarily understanding the underlying mechanisms for their production.

Simulations with biomolecules are usually process driven. They generate representative configurations of the systems providing us with reliable values of structural and thermodynamic properties. They also enable the time-dependent behavior of atomic and molecular systems to be determined, providing a detailed picture of the way in which a system changes from one conformation to another. Such simulations help to predict molecular structures, understand interactions and properties, design of bio-nano materials, experiment on what cannot be studied experimentally, obtain movies of the interacting molecules and many other fascinating capabilities. Molecular modeling is a key methodology for research and development in bionanotechnology, since it provides nanoscale images at atomic and electronic resolution, and it predicts the interaction of biological and inorganic materials. Thus, molecular modeling facilitates studies and in silico experiments with exciting objectives, like the use of carbon nanotubes as in situ biosensors and the use of nanopores for DNA sequencing [2]. For this kind of simulations, the two most popular simulation methods are Molecular dynamics, and Monte Carlo.

Molecular Dynamics is a deterministic method where time dependent properties are calculated, configurations in the past and in the future can be predicted, calculations of kinetic energy and potential energy can be performed, and time average can be calculated from simulation trajectories. On the other hand, Monte Carlo is a stochastic method presenting no temporal relationship and no predictive power among the past and the future. Only the potential energy is used and each configuration depends only on its predecessor. A key function in biomolecular simulations is the energy function that is a target function which describes the interaction energies of all atoms and molecules in the system. The energy function comprises energy terms concerning bonded and non-bonded states. The bonded energy terms describe the capabilities of stretching, bending and rotating around a bond, whereas the non-bonded energy terms include the van der Waals and the electrostatic contributions. The landscape of this energy function is downhill explored when we are investigating stable structural instances of molecules.

Of course, biomolecules are not alone. They are inside a solution and their accurate simulation requires realistic models for the interaction between biomolecules and solvent. There are two types of solvent consideration: the explicit solvent models and the implicit solvent models. In the first category,
the models take into consideration all the solvent molecules presenting high accuracy but also high complexity, slow convergence and high computing requirements. In the second category, solvent is considered as a continuous medium and average estimations of the interactions are produced. In this direction, very useful is the determination of the molecular surface that can be done as a van der Waals, Connolly or Solvent accessible surface by using a rotating probe over the molecule. The choice of the ‘probe radius’ does have an effect on the observed surface area, as using a smaller probe radius detects more surface details and therefore reports a larger surface. A typical probe radius value is the one that approximates the radius of a water molecule [3].

However, there are also efforts in presenting data driven simulations in biology concerning especially biological data modelling like the one presented in the BiDaS Web server we have developed. BiDaS is a web-application that can generate massive Monte Carlo simulated sequence or numerical feature data sets (e.g. dinucleotide content, composition, transition, distribution properties) based on small user-provided data sets. BiDaS server enables users to analyze their data and generate large amounts of: (i) Simulated DNA/RNA and amino acid (AA) sequences following practically identical sequence and/or extracted feature distributions with the original data. (ii) Simulated numerical features, presenting identical distributions, while preserving the exact 2D or 3D between-feature correlations observed in the original data sets [4].

In parallel, at Medicine’s side, we have put a major effort to simulate mammographic imaging, focusing on tasks such as: (i) Simulation of x-ray focal spots with finite size and realistic intensity distribution, (ii) Generation of simulated mammographic images, (iii) Dosimetric calculations, (iv) Investigation of various magnification setup geometries, (v) Simulation of digital detectors for mammographic imaging [5-15]. The objective was to set up the design parameters and the functional parameters of a mammographic unit as well as the properties of a 3-D phantom with inhomogeneities of various geometries and compositions, run a simulation in order to generate the corresponding image and make dosimetric calculations as well. This is a process-driven simulation, thus we have simulated the physical phenomena related to the photons travelling inside material, namely the anelastic and elastic scattering as well as photoelectric absorption. Furthermore, the large number of parameters, along with the need for multiparameter studies and unified result management led us to the development of some simulation environment platforms, named MASTOS and DOSIS as the acronyms for “MAmography Simulation Tool for design Optimization Studies” and “tool for DOsimetry Simulation Studies”. Through such powerful tools we managed to perform various investigations concerning the contrast-to-noise ratio in magnification mammography, the influence of various focal spot sizes and magnification geometries, the influence of x-ray spectra on the image quality and the absorbed dose, the behavior of photoconductor materials as detectors for digital mammography. The Monte Carlo generated mammographic images along with the accompanying information concerning the absorbed dose can be used both for the evaluation of new mammographic imaging setups (geometry, materials, and spectra) and for the training and evaluation of Computer Aided Diagnosis (CAD) schemes.

Moving towards molecular imaging, we have started studying another imaging modality, namely Multispectral Optoacoustic Tomography (MSOT), in collaboration with the Institute for Biological and Medical Imaging at the Helmholtz Center in Munich, Germany. This is an improved optoacoustic imaging modality of biological tissues and molecular imaging, used mainly for resolving contrast of highly absorbing tissue components or contrast agents. The development of an MSOT simulator includes the optical and the acoustic forward problem formulation. Simulating the optical and acoustic phenomena in concert facilitates the direct and easy comparison with experimental data, providing better understanding of the optical and acoustic phenomena that are present in experimental tissue imaging. The development of the simulation and the initial validation results has been just presented [16].

Switching to the other kind of simulations, the data driven ones, we are running a project concerning Modeling of Solitary Pulmonary Nodules in PET/CT images using Monte Carlo Methods. The assessment of solitary pulmonary nodules remains a difficult diagnostic task despite the rapid
improvement of various imaging modalities. CT and PET are the most common methods for diagnosis of SPN. Physicians and computational systems as well, would gain benefit if trained to a large number of SPN cases with controlled topological and morphological characteristics. The modeling of SPN has been implemented by Monte Carlo methods taking into consideration morphological characteristics, internal features and Standardized Uptake Value activity distribution. Again, the generated images can be used both for the evaluation of the medical experts’ diagnostic performance as well as for the training and evaluation of Computer Aided Diagnosis (CAD) schemes [17].

4. Discussion

Concluding, simulations in Medical and Biological research are a very useful and promising tool that may provide us with invaluable information at various levels and scales. The presented research works are only a drop in a wonderful ocean that we are all welcome to sail.

References

[1] Feynman RP, Leighton RB, and Sands ML 1963 The Feynman Lectures on Physics. Reading, MA: Addison-Wesley Publishing Co. vol. 1, p.3-6
[2] Lu D, Aksimentiev A, Shih AY, Cruz-Chu E, Freddolino PL, Arkhipov A, Schulten K 2006 Phys Biol. 2;3(1):S40-53.
[3] Connolly ML 1983 Science. Aug 19;221(4612):709-13.
[4] Paraskevopoulou MD, Vlachos IS, Athanasiadis E, Spyrou G 2013 Nucleic Acids Res.41:W582-6.
[5] Spyrou G, Tzanakos G, Bakas A, Panayiotakis G 1998 Phys Med Biol. 43(11):3341-57.
[6] Spyrou G, Panayiotakis G, Tzanakos G 2000 Med Inform Internet Med. 25(4):275-93.
[7] Spyrou G, Tzanakos G, Nikiforides G, Panayiotakis G 2002 Phys Med Biol. 47(6):917-33.
[8] Delis H, Spyrou G, Panayiotakis G, Tzanakos G 2005 Eur J Radiol 54(3):371-6.
[9] Delis H, Spyrou G, Costaridou L, Tzanakos G, Panayiotakis G 2006 Med Phys. 33(11):4221-35.
[10] Koutalonis M, Delis H, Spyrou G, Costaridou L, Tzanakos G, Panayiotakis G 2006 Phys Med Biol. 51(21):5539-48.
[11] Koutalonis M, Delis H, Spyrou G, Costaridou L, Tzanakos G, Panayiotakis G. 2007 Phys Med Biol. 52(11):3185-99.
[12] Koutalonis M, Delis H, Spyrou G, Costaridou L, Tzanakos G, Panayiotakis G 2008 Phys Med Biol. 53(5):1369-84.
[13] Sakellaris T, Spyrou G, Tzanakos G, Panayiotakis G 2005 Phys Med Biol. 50(16):3717-38.
[14] Sakellaris T, Spyrou G, Tzanakos G, Panayiotakis G 2007 Phys Med Biol. 52(21):6439-60.
[15] Sakellaris T, Koutalonis M, Spyrou G, Pascoal A 2013 Phys Med. S1120-1797(13)00253-6.
[16] Maneas E, Tzoumas S, Ntziachristos V and Spyrou G 2013 Proceedings of the 13th IEEE International Conference on BioInformatics and BioEngineering – BIBE2013.
[17] Tzanoukos G, Gaitanis A, Georgakopoulos A, Chatziioannou A, Chatziioannou S, Spyrou G 2013 Proceedings of the 13th IEEE International Conference on BioInformatics and BioEngineering – BIBE2013.