ETNA BIOTECH
Zydus Cadila Research Center

H2020 Project: 777123

STriTuVaD

Doc. Name: Deliverable 2 WP2

Doc. Num: 006/18 Version: A

Report on the modeling simulation framework able to simulate human immune system dynamics.

30/09/2018

The present document contains the report of Deliverable 2 WP2.

Author: Francesco Pappalardo	replaces version	n.a.	from	n.a.	Page
	valid from date (dd/mm/yyyy)		30/09/2018		1 / 4
	Recipient	STriTuVaD Consortium			

Topic

Partner: UniCT

Deliverable: Summaries, Reports and Information relating to the modeling simulation framework able to simulate human immune system dynamics.

Discussion

Partner: UNICT.

The goal of this deliverable is to show how pre-existing modeling framework (UISS – Universal Immune System Simulator) was assessed to reproduce the immune system dynamics at a human large scale. Pre-existing UISS is a computational framework that makes use of a multi-scale, multi-organ, agent based simulator of the immune system with CrOSSBAR, an attached module able to simulate the dynamics of a biological pathway at the molecular level.

UISS takes into account both cellular and molecular entities. Cellular entities can take up a state from a certain set of suitable states and their dynamics is realized by means of state changes. A state change takes place when a cell interacts with another cell or with a molecule or both of them. We considered the relevant lymphocytes, i.e. B lymphocytes, helper, cytotoxic and regulatory T lymphocytes and natural killer cells. Monocytes are represented as well and we take care of macrophages and dendritic cells. For what concerns molecules, the model distinguishes between simple small molecules like interleukins or signaling molecules in general and more complex molecules like immunoglobulins and antigens, for which we need to represent the specificity.

At the same level of entities, UISS implements immune system activities. They include both interactions and functions. Functions refer to the main immune system tasks. In particular, UISS takes care of the diversity of specific elements, major histocompatibility classes restriction, clonal selection by antigen affinity, thymus education of T cells, antigen processing and presentation (both the cytosolic and endocytic pathways are implemented), cell–cell cooperation, homeostasis of cells created by the bone marrow, hypermutation of antibodies, cellular and humoral response and immune memory.

UISS represents receptors and ligands as bit strings and use a string matching rule to model affinity. This clever idea was introduced by Farmer and Packard as a way to perform calculations for determining molecular complementarity and predicting the optimal size of an epitope. From immunology, we know that binding is a threshold effect consisting of two components: the affinity of a single receptor and ligand, and the total binding, or avidity of multiple binding pairs. Binding is modeled by a string matching rule by counting the number of positions in the string at which the symbols are complementary (known as Hamming distance). Repertoires are represented in the model as sets of strings. By adopting bit strings, many binding events can be simulated quickly, making it

feasible to study large-scale properties of the immune system. Character strings produced accurate models when benchmarked to experiment, suggesting that the abstraction captures important features of receptor/ligand binding.

In particular, specificity is implemented in UISS by a bit-string polyclonal lattice method. Bit-string refers to the way the molecules and the specificity among molecules is represented, polyclonal indicates that more clones of different specificity of lymphocytes are represented and lattice means that we use a discrete lattice to represent the space, that is, the space is discrete. The set of lymphocytes receptors is represented by bit-strings of length *h* which then forms the so called shape space. A clonal set of cells is characterized by the same clonotypic receptor, i.e. by the same bit-string of length l. The potential repertoire of receptors scales as 2^l. The receptor–coreceptor binding among the entities are described in terms of matching between binary strings with fixed directional reading frame. Bit-strings represent the generic binding site between cells (through their receptors) and target molecules (through peptides and epitopes). An interaction between two entities is a complex action which eventually ends with a state change of one or both entities. Specific interactions need a recognition phase between the two entities; recognition is based on Hamming distance and affinity function and is eventually enhanced by adjuvants. When two entities, which may interact, lie in the same lattice site then they interact with a probabilistic law. All entities which may interact and are in the same site have a positive interaction. Physical proximity is modeled through the concept of lattice-site. All interactions among cells and molecules take place within a lattice-site in a single time step, so that there is no correlation between entities residing on different sites at a fixed time. The simulation space is represented as a L *x* L *x* L cubic lattice, with periodic boundary conditions to the left and right side, while the top and bottom are represented by rigid walls. All entities are allowed to move with uniform probability between neighboring lattices in the grid with equal diffusion coefficient (Brownian motion).

During months 1-8 UISS was extended to consider immune system dynamics at human natural scale. This was achieved extending UISS platform in three critical points: *i)* the peripheral blood compartment was designed in order to deal with about 5 liters i.e., 5 cubic decimeters of blood and included immune system entities circulating through it; *ii)* immune system repertoire implementation was extended in order to take into account an immune system potential diversity at human natural scale i.e. about 10^{20} order magnitude of T and B cells clonotypes; *iii)* compartments addition to simulate critical organs targeted by tuberculosis i.e., lungs and near lymph nodes.

The human T-cell receptor (TCR) repertoire i.e., the range of different TCRs expressed ranges between 10^{15} to 10^{20} clonotypes by recombination, random insertion, deletion and substitution; the size of the antibody repertoire has been calculated as 10^{15}. UISS immunological repertoire uses binary strings of 12 bits, reaching a diversity of 2^{12}. In order to simulate the immune system at a human scale it is needed that UISS deals with a repertoire of 10^{20} i.e., having bit-strings of about 60 bits in length. To compute the binding probabilities of all interactions among cells and molecules, UISS use Hamming distance computation i.e., the difference in terms of bits between two different bit strings. The computational effort required to compute the Hamming distance linearly depends on the size of the string. However even a linear effort may be computational heavy if many computations are required. The best known method to compute Hamming distance between two bit strings relies on a pre-computed look-up table kept in computer cache memory; it is computationally fast if the cache misses are rare. With this method, the computation of Hamming distance of binary strings of length l requires a look-up table of $O(2^{2l})$ entries. To manage bit strings of length 60 we developed a new

algorithm that we implemented in UISS. The algorithm is based on magic numbers. Magic numbers in mathematics and computer science are numbers that show special properties when used in certain computations. In particular, they are successfully used in algorithms involving bit handling. They offer faster versions of algorithms over those that can be developed without their use, typically by a factor of n/logn. Our implementation, which is an extension of the well known bit count algorithm implements a branchless computation of the population count. It is based on a O(log(n)) algorithm that successively groups the bits into groups of 2, 4, 8, 16, 32, and 64 (using binary magic numbers) while maintaining a count of the set bits in each group. With this implementation, we were able to deal with binary strings of 64 bits. The method is easily extendable to any bit strings length, but keeping the limit of 12 elementary CPU operations, requires CPUs capable of doing l-bits operations in one time, l>64.

Finally, in UISS it has been set a new compartment i.e., the lung compartment. Lymph nodes compartment is already present and it is extended to take care also of the lymph nodes around lungs. Multi organ ABM simulations, however, require thousands of millions of agents. This clearly represents an issue even for modern CPUs and personal computers. To this end, high-performance computing (HPC) and GPU resources are mandatory to reproduce the natural scale behavior of the immune system and related pathologies. Due to intrinsic nature of the biological and immunological entities that mostly act and interact locally, the simulation of big tissues and/or organs has been split across different computing cores in order to have a parallel run for most of the time, except for the processes that involve entities migration from an organ to another or occasional movement across adjacent tissues fragments belonging to different simulation spaces. This clearly entitles high degrees of scalability in function of the number of available computing cores. To this end, the integration with the FlameGPU software to enable large scale simulations on Graphics Processing Units (GPUs) has been already started and will be completed accordingly to the project schedule.

Conclusion

Partner UniCT: The deliverable for WP2 as required by UniCT has provided the UISS simulation framework that has been extended and enhanced to take care of the immune system at large scale i.e., at human scale, and to include the organs that play important role in TB i.e., lungs and the lymph nodes around them. Large scale simulations would require millions of agents. This clearly represents an issue even for modern CPUs and personal computers. To this end, high- performance computing (HPC) resources are mandatory to reproduce the large scale behavior of the immune system and related pathologies.