Abstract for the general public

When one thinks about the enormous variety and complexity of innumerous processes going on in any type of organism from bacteria to human, one of the best definition of the organism as an entity that can be given, is that a living organism represents a collaborative network of millions and billions of versatile molecular machines. All of them are interconnected, interregulated and dependent on each other, while any small defect could lead to a disruption of their function that is manifested by a disorder or a disease. Amazingly, the ways how the machinery in the organism is able to deal with the irregular work of individual machines, renders it not only extremely complex but less fragile than we could imagine considering all this complexity. The organism specifically "knows" what to do if something wrong happens. If we could obtain and elaborate all this tremendous amount of the information one day, we would be able to deal with any disease or disorder and so to improve the general quality of life. However, despite the progress of the medicine, particularly in the understanding of the fundamental biological processes, the full picture of what is exactly happening is far from being complete. To assemble the full puzzle, one needs first to understand its pieces, which are individual molecules making up cells, tissues, organs and, finally, the organism. Therefore, one of the central objectives in biology-related sciences is to analyze the molecules and their interactions with each other at the individual level. Most of the intermolecular interactions are extremely specific, but what is the source of this specificity? How is it organized? The discovery of the genetic code, one of the key steps in biomolecular sciences, revealed the fundamental concept of the storage and realization of the genetic information by linear sequences of DNA molecules made up of 4 different building blocks, nucleotides. In turn, most of the machinery in the cell works thanks to the proteins, which are the products of this information encoded by DNA. Proteins, which are predominantly made up of 20 amino acid types, also represent linear sequences that depending on their composition determine their 3D structure, dynamic, energetic and functional properties. For a long time it was thought that cellular machinery is almost exclusively dependent on interactions between protein molecules, while two other big classes of biomolecules, namely carbohydrates and lipids are only structural "passive" material of the cell. Hovewer, in the recent decades researchers found out that it is not as simple. Carbohydrates, in comparison to proteins, could be built of hundreds of different building blocks, which are not always linearly connected with each other but also branched, which contributes to the overwhelming complexity of these molecules. One of the carbohydrate classes, glycosaminoglycans (GAG), represents a group of linear periodic negatively charged molecules made up of repetitive blocks, which could be sulfated in a very different way. That makes these molecules particularly complex in terms of their structural and interactional properties. These molecules interact with proteins in the space between the cells determining many biochemical processes underlying such diseases as cancer, Alzheimer's and Parkinson diseases, tissue regeneration abnormalities and the reaction of the cell towards viral infection, as it is in the case of corona virus. How they are sulfated defines their function, which leads to the concept of GAG "sulfation code". Unfortunately, only few attempts have been undertaken to decipher this code. To do this is the goal of our project. We would like to apply molecular modeling techniques such as molecular docking, allowing for the prediction of the structures of complexes established by proteins and GAGs, molecular dynamics, giving insights into the movements of these complexes, and free energy calculations that can predict how strongly the interactions in these complexes are. Doing this, we are going to observe the affects of the sequence on the structure, dynamics, interactions of these molecules, which is inevitable if one would like to understand their functions. In light of ever growing computational power rendering more and more complex computations realistic and the fact that experimental techniques often experience serious challenges to study these complex molecules, theoretical attempts to decode GAG "sulfation code" are especially promising. At the same time, we are going to work hand-in-hand with the experimental groups, which are leading ones in the GAG research, and with whom we have already established fruitful collaborations through the years. We aim to focus on most representative GAG sequences, analyzing their movements and interactions with the surrounding solvent, ions as well as with their experimentally identified protein targets. A systematic plan for doing this is elaborated, and we are very optimistic about the positive results of the project, considering our advances in GAG theoretical results in the last years and strong experimental support from our cooperation partners. Our ultimate goal is not only to understand the fundamental principles underlying the function of these intriguing molecules, but also to be able to propose the novel strategies for the medical sciences suggesting how these molecules could be used in practice for curing a number of diseases, which onset and course they substantially affect.