Modelling of fragmentation of biomolecules induced by electron transfer in mass spectrometry

The discovery of the electron transfer fragmentation (ETD) has made new discoveries possible on the modern mass spectrometers, extending the analytic repertoire of proteomics. The objective of this study is to better understand the statistical dynamics behind the electron transfer induced reactions inside a mass spectrometer by preparing a simplified, yet consistent statistical description thereof. We base our studies on experimental data obtained by our collaborators from Centre For Proteomics in Antwerp, Belgium. We want to base our statistical framework on the theory of continuous time Markov Chains and in-house developed dedicated proteomic data mining tools.

In our research we focus mainly on proteins, however our results might generalise to other structures as well. Proteins play crucial role in the life cycle of any living cell. They catalyze cell metabolism, perform DNA replication, enable within-cell transport of molecules, and are important in cell signaling pathways. Owing to alternative splicing and post-translational modifications, the number of existing proteins is much larger than that of genes that encode them, making them chief responsible for the diversity of life. The branch of analytical chemistry which attempts to identify and quantify proteins on a large scale, as well as elucidate their structure and function, is known as proteomics and has been ongoing for more than twenty years. The advent of high-throughput analysis of proteins came with breakthrough advances in the design of mass spectrometers that made it possible to ionise biomolecules. A mass spectrometer simultaneously determines the mass and quantity of compounds in the sample. This poses obvious limitations on the obtainable results: one cannot discern in this way isomers - chemical compounds with the same atomic composition but different 3D structure. To bypass this disadvantage one breaks the molecule into constituents (fragments) and based on measurements on those tries to identify the original molecule. One of the possible ways to do it is to use the ETD. The precise mechanism of how the process actually works is still being still hotly debated.

Our research will not try to solve that problem, as it is impossible to do it relying on mass spectrometer data alone. Instead we will approach a different problem. Apart from the ETD, electron transfer chemistry might result in various side reactions. Until today it has been difficult to fully understand their prevalences in the process. Our task is to describe their dynamics based on experimentially acquired spectra.

Precise formulation of the statistical dynamics of the above-mentioned reactions would be highly beneficial. One way, it will help the chemists to optimise their data acquiring protocols. It would also render instrument comparison more transparent. These advantages will result from us better understanding the order of appearance of reactions and their intensities. The number of observed infered parameters will be small, making the analysis much simpler.

Our modelling strategy assumes that the complex coulombic spatial interactions can be simplified. We postulate that the dynamics of the system can be described by counts of subgroups of molecules that compose it. We assume that within the considered system molecules with higher charge should react away faster, and that at all times at most one reaction of a particular type can occur; finally, that this reaction type is itself random. Several things might happen to the ion: more than a year ago our fellow researchers from CFP have shortlisted three types of reactions that could describe the mass spectra. After developing tools specialised for fragment identification (project MassTodon) we are now sure that this list has to be extended to at least four reactions.

We want to further enhance the tools we used for recognising the products of reactions under study in the gathered data. One of the improvements would result from using the newly developed IsoStar algorithm. It is being used for the process for data deisotopifying. What is the meaning of all this? The molecules inside the machine are composed of atoms and atoms contain differern numbers of neutrons inside their nuclei. This feature often doesn’t change their properties too much - it only changes their mass. Unfortunately, mass spectrometer measures precisely the mass, and so the presence of isotopes creates an additional layer of complexity in understanding the nature of the gathered information. There might be many different peaks, i.e. pairs mass-intensity (that is basically what the mass spec sees), that might come one and the same source and by tracing that source we get info on the counts of observed products of various reactions inside the machine.

The information we get will be used in the bayesian analysis of the a posteriori distribution of the parameters describing the reaction dynamics. We assume what drives our reactions is itself random. Describing the a posteriori distribution will amount to saying, how much random is it, given the data describing the phenomenon. We shall simulate our way through to get this information extensively using the computational power of modern day computers.

We want to grasp the essence of the electron transfer reactions and provide chemists with insights valuable in the optimisation of the protocols they use for data acquirement. Also, only by better knowing the process that generates our data will we ever be in position to develop ultra-fast algorithms for protein identification from high-throughput instruments. That is what we believe to be the future of computational methods for proteomics: we want to catch that train.