Since the advent of highly efficient genome sequencing techniques, we are flooded with a vast amount of information about genes and proteins. The number of sequences rises much faster than scientists are able to study. Today, there are about 250 million protein sequences. To mitigate this problem sequences of uncharacterized proteins are compared to the sequences of previously studied proteins and based on that comparison the 3D structure and function are being predicted. In this project, we will identify and characterise peptidases, which are enzymes cleaving peptide bonds found mainly in proteins. Peptidases are of great relevance to every aspect of a living cell, from food digestion to cell's shape changing to protein decay. They are widely used in the food and chemical industry, medicine and are also key players in some techniques used in laboratories. Peptidases are promising targets in therapies counterfeiting HIV, diabetes, obesity and many more diseases.

Because peptidases fulfil so diverse and essential functions, it is worth to fully comprehend their structural diversity and evolution as well as relationships between multiple groups of these enzymes. In this project, we will identify, classify and describe the peptidase universe to elucidate the rules governing their function, evolution and adaptation to processing particular substrates.

In this project, we also plan to obtain information about all currently known peptidases what will facilitate developing of their general classification, subsequently annotated with more detailed features. Based on the identification of those features in other protein sequences, we will find new peptidases and incorporate them into our growing dataset. Continuous inflation of the dataset will allow to work out a consistent and comprehensive classification of peptidases, preferably reflecting their structural diversity. Obtained results will be valuable for rational planning of further experimental research and will allow for protein engineering aiming at developing even more efficient peptidases for applications in industry or medicine. They will also enable other computational studies attempting to identify peptidases within wide organism groups, e.g. in fungi. The project will contribute to our understanding of how certain peptidases function at the molecular level and what their roles are in processes going on in cells. A detailed description of protein's function may be a decent starting point to understand the molecular basis of the disease and significantly help in effective drug discovery. In our project, we will also identify human peptidases, what will contribute to worldwide efforts for the human genome and proteome annotation.