

# Can computers find better proteins?

**Finding more efficient enzymes is critical for chemical industry to become sustainable. In order to efficiently find better enzymes new tools are needed that reliably predict protein properties.**

The chemical industry is on a search for new green technology that can replace catalysts that are often toxic and made of scarce minerals. A promising alternative is catalytic proteins, so called enzymes, that can be produced renewably and pose no toxic threat. However, in order to be a viable alternative, the enzymes must become more effective and be able to withstand harsher conditions.

To design enhanced or entirely new enzymes accurate and fast predictions regarding the protein's properties are needed. One such highly sought after property is protein stability at high temperatures since it usually speeds up chemical reactions. A promising approach for developing predictive models are machine learning techniques, which have proved powerful in numerous fields.

At the department of Biophysical Chemistry at Lund University it was tested if such a model, that predicts the thermal stability of proteins, could be created. The study concluded that the tested models could detect some features of proteins, however for the models to be more useful, they must be trained on more data of high quality.

The limiting factor during development was therefore the amount of relevant and available data. This is a common problem in the machine learning field, where the models are created by *training* on data, the computer is taught how to make predictions by learning from a dataset. The

rule of thumb is that the more data it is trained on, the better performance.

But what would be the benefit of using computers over good old laboratory work? While resulting in reliable values, laboratory work is usually costly and time-consuming and requires well trained personnel. If a relatively accurate predictive model can be constructed an estimate of a protein's stability can be obtained in a few minutes or hours instead of weeks. The benefit is obvious if the stabilities of many proteins are of interest.

Furthermore, the technology would enable researchers to quickly evaluate a large number of enzyme variations at an early stage, therefore enabling selection of only the promising enzymes for further investigation.

Accurate models of protein stability is not only useful for chemical industry, but could also be used to better understand the proteins found in nature, and how some mutations can lead to certain diseases. Diseases such as Alzheimer's and Parkinson's disease have been linked to proteins losing their natural structure and instead forming protein aggregations. By better understanding the properties of these proteins, researchers hope to find new treatments for these diseases that affect roughly 30 million people worldwide, which is about the same population as countries such as Nepal or Venezuela.

By solving the protein stability problem multiple fields could undergo a revolution, leading to a more sustainable and healthier world. However, if computers are the ones to solve it more experimental data is needed.