



Development of an AI-method for predicting the ability of protein synthesis in microorganisms

Masterthesis, Bachelorthesis, undergraduate assistant

Background:

The production of proteins using microorganisms is an important branch of industrial biotechnology. It ranges from technical enzymes for the production of e.g. bioethanol to the synthesis of pharmacologically active proteins such as therapeutic antibodies. A handful of established production organisms are confronted with a large number of interesting proteins which are "foreign" to the production organism, so-called heterologous proteins. Heterologous protein synthesis often does not achieve the efficiency that is achieved with host-own (homologous) proteins. The information for successful protein synthesis is encoded in the DNA sequence of the homologous proteins, among other things. Gene expression is the first important step on the way from DNA to protein. The decoding of this information is the subject of intensive research worldwide.

Objective:

The aim of the research project is to use deep neural networks to identify patterns in the gene sequences of homologous proteins that are responsible for successful expression. Based on these patterns, a prediction of the producibility of heterologous proteins will be made using their DNA sequence and, if necessary, verified with available experimental data.

Implementation / techniques:

- Preparation of genomic data sets: Genome / Pangenome of a host system based on open access sequence databases
- Preparation of a control data set from the sequence databases
- Design and training of deep neural networks for pattern recognition based on training and control data sets
- Verification / falsification of the discovered patterns with available experimental data from BRAIN AG
- Prediction of the producibility of heterologous proteins of an enzyme class and testing in the laboratory (laboratory work can be performed by the student or employees of Brain AG)

Implementation / Organisation:

The supervision is provided by BCS Lab of TU Darmstadt under leadership of Prof. Köppl. The BRAIN AG provides information on the biological systems and contributes expertise on protein expression in the respective host system to the project. A close exchange of results takes place at regular meetings and conferences.

The (Bachelor's/Master's) student receives a scholarship of 650 € per month from BRAIN AG. An undergraduate assistant is compensated according to the hourly rates of the TU Darmstadt and the agreed number of hours.

If you are interested please send an e-mail to: office@bcs.tu-darmstadt.de

Fachbereich 18
Elektrotechnik und
Informationstechnik

Institut für Nachrichtentechnik

Bioinspirierte
Kommunikationssysteme

www.bcs.tu-darmstadt.de

Prof. Dr. Heinz Köppl

Gebäude S3 | 19, 1.1

Rundeturmstr. 12
64283 Darmstadt

Tel. +49 6151 16 - 57 235
Fax +49 6151 16 - 57 241
Mobil +49 151 61 37 57 19
heinz.koepl@bcs.tu-darmstadt.de

Datum
14. Dezember 20



Entwicklung eines KI-Verfahrens zur Vorhersage der Proteinsynthesefähigkeit in Mikroorganismen

Masterarbeit, Bachelorarbeit, Anstellung als studentische Hilfskraft

Hintergrund:

Die Herstellung von Proteinen mit Mikroorganismen ist ein bedeutender Zweig der industriellen Biotechnologie. Die Spanne umfasst technische Enzyme zur Herstellung von z.B. Bioethanol bis hin zur Synthese pharmakologisch aktiver Proteine, z.B. therapeutischer Antikörper. Einer Hand voll etablierter Produktionsorganismen steht dabei eine Vielzahl interessanter Proteine gegenüber, die für den Produktionsorganismus "fremd" sind, sogenannte heterologe Proteine. Die heterologe Proteinsynthese erreicht oftmals nicht die Effizienz, die mit wirtseigenen (homologen) Proteinen erzielt wird. Die Information zur erfolgreichen Proteinsynthese ist u.a. in der DNA-Sequenz der homologen Proteine codiert. Die Genexpression ist der erste wichtige Schritt auf dem Weg von der DNA zum Protein. Die Decodierung dieser Information ist Gegenstand intensiver Forschung weltweit.

Zielsetzung:

Ziel des Forschungsprojektes ist es, anhand von tiefen neuronalen Netzen Muster in den Gensequenzen homologer Proteine zu erkennen, die für eine gute Exprimierbarkeit verantwortlich sind. Basierend auf diesen Mustern soll nachfolgend eine Vorhersage über die Produzierbarkeit heterologer Proteine anhand deren DNA-Sequenz erfolgen und ggf. mit vorliegenden experimentellen Daten verifiziert werden.

Durchführung / Techniken:

- Aufbereitung genomischer Datensätze: Genom / Pangenom eines Wirtssystems basierend auf öffentlichen Sequenzdatenbanken
- Erstellung eines Kontrolldatensatzes aus öffentlichen Sequenzdatenbanken
- Entwurf und Training von tiefen neuronalen Netzen zur Mustererkennung anhand von Trainings- u. Kontrolldatensätzen
- Verifizierung / Falsifikation der gefundenen Muster mit vorliegenden experimentellen Daten der BRAIN AG
- Vorhersage der Produzierbarkeit heterologer Proteine einer Enzymklasse und Testung im Labor (Laborarbeiten können von dem/der Student/-in oder Mitarbeiter/innen der Brain AG durchgeführt werden)

Durchführung / Organisation:

Die Betreuung findet durch das Fachgebiet BCS der TU Darmstadt unter der Leitung von Prof. Köppl statt. Die BRAIN AG liefert Informationen zu den biologischen Systemen und bringt Expertise zur Proteinexpression im jeweiligen Wirtssystem in das Projekt ein. Hierzu findet ein enger Austausch über Ergebnisse bei regelmäßigen Treffen bzw. Konferenzschaltungen statt.

Der/die (Bachelor-/ Master-) Student/-in erhält seitens der BRAIN AG ein Stipendium von 650 € pro Monat. Eine wissenschaftliche Hilfskraft (HiWi) wird gemäß der Stundensätze der TU Darmstadt und der vereinbarten Stundenzahl entschädigt.

Bei Interesse senden sie bitte eine E-Mail an: office@bcs.tu-darmstadt.de

Fachbereich 18
Elektrotechnik und
Informationstechnik

Institut für Nachrichtentechnik

Bioinspirierte
Kommunikationssysteme

www.bcs.tu-darmstadt.de

Prof. Dr. Heinz Köppl

Gebäude S3 | 19, 1.1

Rundeturmstr. 12
64283 Darmstadt

Tel. +49 6151 16 - 57 235
Fax +49 6151 16 - 57 241
Mobil +49 151 61 37 57 19
heinz.koepl@bcs.tu-darmstadt.de

Datum
14. Dezember 20