"Heavy-Light Chain Pair Identification in Antibodies using BERT (work title)"

MASTER THESIS
FACULTY OF SCIENCE, UNIVERSITY OF BERN

HANDED IN BY

Lea Brönnimann

SEPTEMBER 2024

SUPERVISORS:

Prof. Dr. Thomas Lemmin
Chiara Rodella

Abstract

Contents

Glossary

Word Embeddings Mapping of words into vectors with real numbers.

1. Introduction

1.1. Background

1.2. Objectives

1.3. Scope of the Topic

1.4. Research Question

1.5. Nature of the Thesis

1.6. Relevance of the Thesis

2. Literature Review (work title)

2.1. Introduction to Antibodies

Antibodies are Y-shaped proteins that consist of two identical light chains (LCs) and two identical heavy chains (HCs) (Chiu, Goulet, Teplyakov, & Gilliland, 2019).

2.2. Antibody Engineering and Therapeutic Applications

2.3. Deep Learning Methods for Antibodies

Deep learning is a branch of machine learning that focuses on algorithms capable of identifying complex patterns in data by transforming low-level inputs (like pixels in an image) into high-level features (such as object shapes). It utilizes artificial neural networks (ANNs) with multiple layers between the input and output, making them "deep". These networks consist of nodes, or neurons, that process inputs and pass the outputs to subsequent layers, gradually extracting more abstract features. In the context of biochemistry, deep learning can start from basic data, like amino acid sequences, and learn to recognize complex biological structures or functions (Graves et al., 2020).

2.4. BERT and Transformers in Bioinformatics

2.5. Heavy-Light Chain Pair Identification

2.5.1. Gap in the Literature

2.5.2. Conclusion

3. Materials & Methods

4. Results

5. Discussion

6. Conclusion

References

Chiu, M. L., Goulet, D. R., Teplyakov, A., & Gilliland, G. L. (2019). Antibody structure and function: The basis for engineering therapeutics. *Antibodies*, *8*(4). doi: 10.3390/antib8040055

Graves, J., Byerly, J., Priego, E., Makkapati, N., Parish, S. V., Medellin, B., & Berrondo, M. (2020). A Review of Deep Learning Methods for Antibodies. *Antibodies*, *9*(2), 12. Retrieved from www.mdpi.com/journal/antibodies doi: 10.3390/antib9020012

List of Figures

List of Tables

A. Appendix

A.1. Link to the Code

The link to the entire code of this thesis can be found at:
`https://github.com/ibmm-unibe-ch/OAS_paired_sequences_cls.git`

A.2. Declaration of Independence