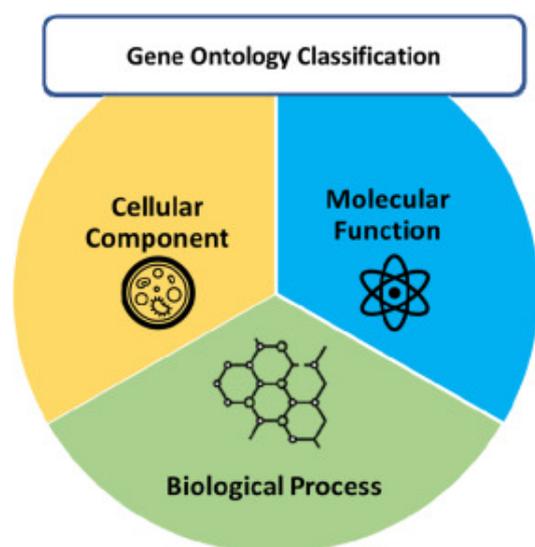


## Thesis

# Knowledge Graph Assisted Protein Models

Deep neural networks have been widely applied in the field of bioinformatics. The methodology applied in the field of natural language processing (NLP) and computer vision (CV) has been transferred to effectively obtain protein representation. Taking the task of protein-protein interaction as example, one existing work [1] designs an end-to-end Siamese residual framework to encode the acid amino sequence. Another work [2] transfers graphical representation learning method to encode contextual protein representation.

However, the works mentioned above only focus on protein biological feature, no matter whether the protein model encodes the amino acid sequence or the protein interaction information. Only few works utilized the structural information from the knowledge graphs (KGs) to enhance the protein representation.



In this work, you will work on modeling protein representation learning with external textual knowledge graph [3]. You will be encouraged to investigate different methods in the related fields, ideally aiming at a publication in relevant venues. You are further encouraged to realize your own ideas. Note that this project does not require expertise in biology, even though it involves processing KGs that contain biological data.

Some of the following may be of use in this project:

- Python programming skills
- Deep learning framework (e.g. PyTorch)
- Basic knowledge of knowledge graphs

For further information, please contact Fengyu Cai.

Fachbereich 18  
Elektrotechnik und  
Informationstechnik  
Selbstorganisierende Systeme

Department 18  
Electrical Engineering and  
Information Technology  
Self-Organizing Systems

Prof. Dr. Heinz Koepl  
Head of lab

Fengyu Cai  
Project supervisor

Rundeturmstraße 12  
64283 Darmstadt

Phone: +49 615116 - 57243  
fengyu.cai@tu-darmstadt.de  
<https://www.bcs.tu-darmstadt.de>

Date  
June 29, 2022

## References

- [1] M. Chen, C. J. Ju, G. Zhou, X. Chen, T. Zhang, K. Chang, C. Zaniolo, and W. Wang, “Multifaceted protein-protein interaction prediction based on siamese residual RCNN,” *Bioinform.*, vol. 35, no. 14, pp. i305–i314, 2019.
- [2] G. Lv, Z. Hu, Y. Bi, and S. Zhang, “Learning unknown from correlations: Graph neural network for inter-novel-protein interaction prediction,” in *IJCAI*. ijcai.org, 2021, pp. 3677–3683.
- [3] D. Yu, Y. Yang, R. Zhang, and Y. Wu, “Knowledge embedding based graph convolutional network,” in *WWW. ACM / IW3C2*, 2021, pp. 1619–1628.