



**Christian Dallago** *November 11, 1992*

Last update on January 18, 2024

✉ [christian@dallago.us](mailto:christian@dallago.us) [in linkedin.com/in/sacdallago](https://www.linkedin.com/in/sacdallago)  
🐙 [github.com/sacdallago](https://github.com/sacdallago) ☎ (843) 410-9775

Contact me in: [English](#) • [Italiano](#) • [Deutsch](#) • [Español](#)

## Highlights

- Award winning scientific projects; teaching projects featured on [Wired](#), [The Verge](#), and [Guardian](#).
- Extensive record of interdisciplinary (biology, machine learning & software engineering) and collaborative work in academia and industry
- 7+ years research impact in Computer Science, Biology, and Machine Learning with publications (30+), scholarships (4), one individually awarded grant (100K EUR), editorial and chairing roles (3), and invited talks (20+).

---

## Experience

- Senior Solution Architect, Healthcare and Life Science @ NVIDIA** | MUNICH, DE *Jul'22 — Present*  
Helped small- to large biotech, pharma and research groups build applied machine learning solutions for biology & served as BioAI subject-matter-expert, helping cross-organizationally by shaping R&D and product directions, assisting on due diligence, and securing strategic collaborations.
- Senior Data Scientist, Group Data Analytics @ Allianz SE** | MUNICH, DE *Apr'21 — Jun'22*  
Owner of voice products to assist customers during their insurance journeys. Lead a diverse team of engineers and data scientists in building solutions leveraging AI & NLP.
- Scientific Researcher, Rostlab @ Technical University of Munich** | MUNICH, DE *Feb '19 — Feb '23*  
Contributed over 15 peer-reviewed articles in two years in the fields of biology, artificial intelligence and computer science to uncover biological mechanisms and how to use AI to improve human health.
- Software Engineering Researcher, Sanderlab @ Harvard Medical School** | BOSTON, USA *Jan'18 — Jan'19*  
Lead the development of two scientific software products to predict the 3 dimensional structures of proteins, and to digest biological knowledge from NLP in computable form to consume via AI algorithms.
- Scientific Collaborator, Rostlab @ Technical University of Munich** | MUNICH, DE *Oct '15 — Nov '17*  
Assisted the chair of computational biology in teaching activities for bachelor and master students in computer and data science, with a focus on DevOps and MLOps.
- Lead Engineer & Co-founder, Reebes** | BOLZANO, IT *Oct '14 — Jun '15*  
Co-founded a software company aimed at facilitating connections between local businesses and customers via a custom built online platform.
- Scientific Collaborator, Free University of Bolzano/Bozen** | BOLZANO, IT *Jul '13 — Jul '14*  
Contributed to the development of the first fully functional modular raspberry-pi cluster and to other creative uses of raspberry-pi's, for example to re-purpose broken electronics.
- System Administrator, Zucchitours S.N.C.** | BOLZANO, IT *Dec '10 — Dec '15*  
Responsible for the IT infrastructure of a travel agency. I implemented virtualization solutions to improve workflows and enable remote work.

More data on [LinkedIn](#).

---

## Education

- PhD in Informatics, TUM** | MUNICH, DE *Oct '17 — Dec '23*  
Grade: SUMMA CUM LAUDE (with highest honor), Major: BIOINFORMATICS
- MSc in Informatics, TUM** | MUNICH, DE *Apr '15 — Oct '17*  
GPA: 3.92/4, Major: BIOINFORMATICS, MINORS: SOFTWARE ENGINEERING, COMPUTER VISION
- BSc in Computer Science and Engineering, UniBZ** | BOLZANO, IT *Oct '11 — Oct '14*  
GPA: 3.63/4, Major: SOFTWARE ENGINEERING
- Visiting** [College of Charleston, SC, U.S.A.](#), 2012 (6mo) • [Universidad Pontificia de Salamanca, Campus de Madrid, Madrid, Spain](#), 2013 (6mo) • [Nanyang Technological University, Singapore](#), 2016 (6mo) • [Harvard Medical School, MA, U.S.A.](#), 2018 (1y)

## Selected Academic Work

**Learned embeddings from deep learning to visualize and predict protein sets** A python framework that leverages learned representations (embeddings) from trained protein language models (ELMo, BERT, and more) to provide computable representations for downstream protein predictions (e.g., subcellular location and secondary structure). The first framework of its kind: it quickly grew a community and dedicated users. It now features a hosted API and two front-ends, and is in use by large projects with over 3000 monthly users.

**Role:** Initiator & Lead author

**Find it online:** [bioembeddings.com](https://bioembeddings.com)

**Article/Proceeding:** ISMB&LMRL&Current Protocols, [10.1002/cpz1.113](https://doi.org/10.1002/cpz1.113)

**Modeling aspects of the language of life through transfer-learning protein sequences** The first comprehensive attempt at learning protein sequence representations using contextually aware NLP techniques (*SeqVec*). This transformative piece led to companies like Facebook, Google and Salesforce following in the steps of our initial findings. Subsequent work (*ProtTrans* in further publications) was backed by grants amounting to \$4.5 Million in compute and revolutionised the field.

**Articles:** SeqVec, *BMC Bioinformatics*, [10.1186/s12859-019-3220-8](https://doi.org/10.1186/s12859-019-3220-8); ProtTrans, *IEEE Trans. on Pat. Analysis and Mach. Intel.*, [10.1109/TPAMI.2021.3095381](https://doi.org/10.1109/TPAMI.2021.3095381)

**Languages, frameworks and technologies:** Python (pytorch, AllenNLP), JavaScript (react), Transformers (T5, BERT, XLNet)

**Capturing scientific knowledge in computable form** Biofactoid: an online workflow for biological scientists that leverages computational agents to assist humans in data curation, e.g. by reading existing literature via NLP and providing users recommendations and Q/A bots. This innovative system is being integrated in the publishing workflow of major companies such as Elsevier and Springer Nature.

**Role:** Project Manager & Developer

**Languages and frameworks:** Node.js

**Find it online:** [biofactoid.org](https://biofactoid.org)

**Article/Proceeding:** OSF Preprints, [10.31219/osf.io/zep3x](https://doi.org/10.31219/osf.io/zep3x)

**Embeddings from deep learning transfer GO annotations beyond homology** A first-principles approach to using protein-embeddings from protein LMs (see above) to annotate proteins with unknown Gene Ontology (GO) annotations based on proteins with known GO annotation. This piece of work highlighted the fundamental difference between purely sequence-based methods (like homology) and learned sequence models (like protein LMs), highlighting the larger information content of the latter in its most naive setting.

**Role:** Co-initiator, Co-author, & Developer.

**Languages and frameworks:** Python **Find it online:** [embed.protein.properties](https://embed.protein.properties)

**Article/Proceeding:** Scientific Reports, [10.1038/s41598-020-80786-0](https://doi.org/10.1038/s41598-020-80786-0)

**SARS-CoV-2 structural coverage map reveals viral protein assembly, mimicry, and hijacking mechanisms** An extensive analysis of the mechanisms of SARS-CoV-2 obtained by combining a wealth of predictions and experimental data into a unique visual resource.

**Role:** Co-author, & Developer.

**Languages and frameworks:** Python, JavaScript **Find it online:** [aquaria.ws/covid](https://aquaria.ws/covid)

**Article/Proceeding:** Molecular Systems Biology, [10.15252/msb.202010079](https://doi.org/10.15252/msb.202010079)

**End-to-end multitask learning, from protein language to protein features without alignments** Elnaggar A., Heinzinger M., Dallago C., Rost B.

BioRxiv - 2019 - [10.1101/864405](https://doi.org/10.1101/864405)

**Posters:** CellMap visualizes protein-protein interactions and subcellular localization, ISMB, 2017 • EVcouplings.org: web service for analysis of protein coevolution and prediction of 3D structures and mutation effects, ISMB, 2019 • Factoid: Making Biological Pathways in Research Articles Easy to Find and Access, ISMB, 2019 • Encoding Evolution Information through Multi-Task Alignment Free Protein Prediction (MT-ALF), NeurIPS, 2019

**GenSLMs: Genome-scale language models reveal SARS-CoV-2 evolutionary dynamics** With GenSLMs, we explored one approach to learning DNA sequences through large language and diffusion models. The goal was to obtain a genome-wide view of an organism, and in particular SARS-CoV-2, in order to understand the dynamics of its evolution. This work won the Gordon Bell special prize, and exemplified a large-scale experimentation workflow completely driven in-silico. In parallel, I was working on a set of nucleotide transformer models to study how model scale and data diversity affect downstream performance.

**Article:** BioRxiv, [10.1101/2022.10.10.511571](https://doi.org/10.1101/2022.10.10.511571)

**Languages, frameworks and technologies:** Python (pytorch, pytorch lightning), Transformers (GPT2), Diffusion

**From sequence to function through structure: Deep learning for protein design** Invited to contribute to Computational and Structural Biotechnology Journal, I reached out to colleagues internationally to pursue three objectives: create an up-to-date collection of tooling for protein design and engineering, develop a non end-to-end pipeline for protein sequence generation and selection, and lastly demonstrate a use-case of this technology for industrial applications. The result of this effort are useful for practitioners and researchers, and lay the foundation for reinforcement learning work.

**Article:** CSBJ, [10.1016/j.csbj.2022.11.014](https://doi.org/10.1016/j.csbj.2022.11.014)

**Languages, frameworks and technologies:** Python (pytorch), Transformers (T5, GPT2)

Other contributions on [Google Scholar](#) and [GitHub](#).

---

## Grants & Scholarships

**Bundesministerium für Bildung und Forschung (BMBF) - 01IS17049** Grant for Junior Scientists, Software Campus 2.0 (TU München, 2020-2022; Individual; Total: 100.000 EUR)

**Bundesministerium für Bildung und Forschung (BMBF) - 031L0168** Shared

**Deutsche Forschungsgemeinschaft (DFG) - RO 1320/4-1** Shared

**Deutschlandstipendium & IBM** Scholarship for extraordinary students, 2015/2016; 2016/2017; EUR 7.200

**Provincia Autonoma di Bolzano** Scholarship for post-graduate students, 2018; 2019; EUR 18.000

---

## Other academic contributions

**Reviewer** Nature Biotechnology, Nature Machine Intelligence, Nature Communications, OUP Bioinformatics, PLoS Computational Biology, IEEE/ACM transactions on computational biology and bioinformatics, and more.

**Organization** Main organizer for the special session on [Representation Learning in Biology](#) at Intelligent Systems for Mol. Bio. 2021; Area chair for "Machine Learning in Biology" at International Super Computing 2022; Co-editor of 2023 Cold Spring Harbor volume titled "Machine Learning for Protein Science and Engineering" (in-production).

---

## Teaching

**Javascript technology** Technical University of Munich - Fall & Spring 2016-2017

**Web App Development with the Power of Node.js** MOOC via edX & Technical University of Munich

**Protein Prediction I & II for Computer Scientists** Technical University of Munich - Spring & Fall 2019-2023

---

## Selected Media

2016, [This Computer Algorithm Predicted Who Will Die Next on Game of Thrones](#) TIME

2019, [Machine learning algorithm predicts who will be left standing in 'Game of Thrones'](#) phys.org

2021, [3D analysis of SARS-CoV-2 reveals clues on virus tactics](#) phys.org

2022, [Gordon Bell Special Prize Goes to LLM-Based Covid Variant Prediction](#) HPCwire