

in [linkedin.com/in/sacdallago](https://www.linkedin.com/in/sacdallago)

o github.com/sacdallago

Contact me in: English • Italiano • Deutsch • Español

Highlights

- Award winning AI in digital biology contributions and academic achievements featured on [Wired](#), [The Verge](#), and [the Guardian](#).
- Extensive record of interdisciplinary (biology, machine learning & software engineering) and collaborative work in academia and industry pioneering new academic fields (protein/genomic language models) & transferring to digital biology products
- Extensive record of Digital Biology ecosystem building, influencing research communities and supporting executives with product market fit and go-to-market
- 8+ years research impact in Computer Science, Biology, and Machine Learning with peer reviewed publications (30+), scholarships (4), individually awarded grants, editorial and chairing roles (4), and invited talks (20+).

Experience

Visiting Assistant Professor, Biostat & Bioinfo, and Cell Biology @ Duke |DURHAM, NC Feb'25 — Present
Principal Investigator (PI) of a small research group with focus on applied computational biology and machine learning approaches for biosequence learning.

Senior Applied Research Scientist, Digital Biology @ NVIDIA |BOSTON, MA Jul'22 — Present
Leading strategic research engagements and frontier applied research in computational biology, advisor to external small- to large techbio, pharma and research groups & internally serving as BioCS&BioML subject-matter-expert.

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, MA 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

Modeling aspects of the language of life through transfer-learning protein sequences The first comprehensive attempt at learning protein sequence representations from a large dataset using contextually aware NLP techniques (*SeqVec*). Subsequent work (*ProtTrans* in further publications) based on transformer architectures was backed by grants amounting to \$4.5 Million in compute, and kick-started many other initiatives in the community.

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

GenSLMs: Genome-scale language models reveal SARS-CoV-2 evolutionary dynamics With GenSLMs, we explored one approach to learning DNA sequences & genomes through large language- and diffusion models. The goal was to characterise SARS-CoV-2 in order to understand the dynamics of its evolution. This work won the Gordon Bell special prize. In parallel, I was working on a set of nucleotide transformer encoder models to study how model scale and data diversity affect downstream performance.

Articles: GenSLMs, *Int.J.HPC Apps.*, 10.1177/10943420231201154, NucTrans, *Nat. Methods*, 10.1038/s41592-024-02523-z,

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 pipeline for protein sequence generation and selection, and lastly explore an applied drug discovery use-case. The resulting software are useful for practitioners, and lay the foundation for reinforcement learning work.

Article: CSBJ, 10.1016/j.csbj.2022.11.014

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 [Machine Learning for Life and Material Sciences](#) workshop for ICML 2024; Co-editor of 2023/2024 Cold Spring Harbor Machine Learning for Protein Science and Engineering volume; Area chair for Machine Learning in Biology at International Super Computing 2022; Main organizer for the special session on [Representation Learning in Biology](#) at Intelligent Systems for Mol. Bio. 2021.

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](#)