

Daniela A. Gonzalez, PhD

Sheridan St, Jamaica Plain, MA 02130 | (857) 361-6551 danigon37@gmail.com - [linkedin.com/in/daniela-alejandra-gonzalez](https://www.linkedin.com/in/daniela-alejandra-gonzalez)

Professional Profile

PhD in Biological Sciences and MS in Data Science bridging AI/ML and wet-lab discovery in biologics and drug discovery. Proven track record in AI-wet lab integration, applying machine learning and deep learning models, including protein language models (ESM2), graph neural networks for drug-drug interactions, and multi-modal learning approaches, to accelerate biologics discovery. Experienced in assay development and optimization, stable cell line development, biotherapeutics characterization, and high-throughput screening within biopharma settings. Skilled in Python, PyTorch, TensorFlow, and data engineering (ETL pipelines, data ingestion, dashboards), with expertise in curating high-quality wet-lab data for computational modeling and ML/AI applications. Strong scientific leadership through 16+ publications, cross-functional collaboration across biology, chemistry, computational biology, and data science teams, and mentoring in academic and professional settings.

Skills

- **Biologics & AI-Wet Lab Integration:** 10+ years of experience in mammalian cell culture (CHO, HT-29, SK-MEL-28), functional and biochemical assay development, plate-reader-based assays (EC50/IC50, dose-response), and stable clone selection. Exposure to automated liquid handling (Hamilton) and biophysical characterization (Octet). Proven ability to bridge experimental biology and computational modeling for data-driven prioritization in biologics and oncology drug discovery.
- **Machine Learning & Deep Learning for Protein Science:** Hands-on experience with protein language models (ESM2) applying transfer learning with multi-modal inputs (sequence features, embeddings, biological annotations), and graph neural networks for drug-drug interaction prediction. Skilled in building predictive models for biological activity classification, assay outcome prediction, and data-driven compound/clone prioritization. Training in generative architectures (VAEs, diffusion models, transformers) for molecular representation and biologics design.
- **Programming & Data Engineering:** Proficient in Python (NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow), R (tidyverse, Bioconductor), SQL, and Bash. Experience building ETL (Extract-Transform-Load) and data ingestion pipelines for QC, normalization, and structuring of biological datasets into centralized storage (data lake), aligned with FAIR data principles. Familiar with ELN/LIMS documentation systems and data visualization/dashboards for reporting.
- **Computational Biology & Statistical Modeling:** Experience integrating in vitro experimentation with computational analysis for hypothesis generation and experimental design. Skilled in multi-assay data integration (bioassay + biophysics + omics), RNA-seq, transcriptomics workflows, Design of Experiments (DoE, RSM), multivariate analysis, pathway enrichment, and predictive modeling of complex biological datasets.
- **Scientific Leadership & Cross-Functional Collaboration:** 16+ publications in Q1/Q2 journals, Fulbright Fellowship, and mentoring experience through the American Statistical Association. Demonstrated success in cross-functional collaborations across biology, chemistry, computational biology, and data science teams, with strong communication through publications, conference presentations, and external scientific community engagement.

Work Experience

Research Scientist – The Institute for Experimental AI – Northeastern University

December 2025 – Present

Boston

- Building automated ETL and data retrieval pipelines from public biomedical databases via API integration, structuring large-scale chemical and biological data for downstream ML applications.
- Collaborating with cross-functional teams across biology, data science, and engineering to co-define data requirements, analytical frameworks, and evaluation strategies for complex biological datasets, ensuring alignment between experimental design and computational modeling needs.

Data Scientist intern - Cell Line Development - Takeda Pharmaceuticals

April – September 2025

Cambridge, USA

- Served as a scientific bridge between computational and experimental teams, leading training on wet-lab/dry-lab integration and translating analytical insights into actionable strategies for assay development and optimization of cell-based and biochemical assays across biologics campaigns.
- Developed end-to-end analytical workflows integrating cell line performance metrics with process metadata to enable structured, reproducible evaluation of therapeutic protein-producing clones, driving data-driven prioritization in biologics development.
- Built multivariate statistical models and predictive ML algorithms to assess fed-batch performance and early functional response across ~300 clones, reducing analysis turnaround time by ~40% and achieving ~91% R² in early identification of high-producing clones.
- Implemented standardized data preprocessing, QC pipelines (normalization, curve fitting, EC₅₀/IC₅₀ determination, reproducibility analysis), and ML-ready data curation practices to improve comparability of plate-reader and assay outputs across experiments, optimizing high-quality wet-lab data for computational modeling.
- Designed metadata schemas, traceability structures, and data integrity practices aligned with Quality-by-Design (QbD), GxP, and FAIR data principles, documenting all processes in ELN/LIMS to strengthen reproducibility and support scalable biologics development workflows.
- Supported automation of in-vitro assay workflows in cell line development, contributing to streamlined high-throughput screening and clone selection using automated liquid handling (Hamilton) and titer quantification platforms.

Ondrechen Research Group – Chemical Modeling Research Assistant

January 2024 – April 2025

Boston, USA

Northeastern University

- Conducted structure-based rational design workflows for drug discovery and hit identification, applying molecular docking, molecular dynamics simulation, and binding free-energy estimation to evaluate natural compounds against therapeutically relevant protein targets (LOX, COX, BRAF).
- Leveraged structural informatics and computational modeling to characterize protein-ligand binding stability, prioritize lead candidates based on predicted potency, and guide experimental validation strategies.
- Built QSAR predictive models using molecular fingerprints to relate chemical structure to biological activity, supporting data-driven compound prioritization.
- Identified top-ranked leads with improved affinity and interaction persistence vs. reference inhibitors, supporting advancement toward early hit validation.
- First-author manuscript under review; presented findings to multidisciplinary collaborators bridging computational and experimental biology.

Supervisor, Wet lab – EXP Makerspace

January 2024 – April 2025

Boston, USA

Northeastern University

- Managed operations of a multidisciplinary BSL-1/BSL-2 laboratory, overseeing experimental planning, resource allocation, equipment maintenance, and safety compliance.
- Mentored and trained students and researchers in mammalian and microbial cell culture, aseptic technique, and molecular biology workflows, fostering a collaborative and high-performing lab environment.
- Facilitated cross-functional collaboration between wet-lab and computational teams, improving experimental reproducibility, data traceability, and research workflow efficiency.
- Performed core analytical and molecular biology assays (DNA/RNA extraction, PCR, HPLC, fluorescence imaging, spectrophotometry) supporting biochemical characterization and cross-disciplinary research.

PhD grad student research -INBIOFIV-CONICET

Apr 2018 - Aug 2023

Tucumán, Argentina

- Integrated computational and experimental discovery approaches by collaborating on machine learning models for predicting docking scores and prioritizing compound libraries.
- Designed and optimized extraction and purification workflows for plant-derived bioactive compounds, applying ML-based response surface models to maximize yield and biological activity.
- Developed and executed in vitro cell-based and biochemical assays across 3 plant species (6 extracts) and 2 cancer cell lines to assess cytotoxic, chemopreventive, and anti-proliferative effects using plate-reader-based dose-response (EC50/IC50), and conducted functional validation studies in model organisms.
- Performed toxicological and dose-response profiling of lead compounds to guide early-stage therapeutic evaluation.

Education

Northeastern University

MS in Data Science (Khoury College of Computer Sciences)

Boston, Massachusetts

GPA: 3.77

National University of Tucumán

PhD in Biological Sciences (CONICET)

Tucumán, Argentina

GPA: 4.00

National University of Tucumán

BSc in Biotechnology

Tucumán, Argentina

GPA: 3.66

Awards, Mentorings and Publications

- American Statistical Association - SSDSE Mentoring Program for Professional Development – August 2025 - Present
- Scholarship AAUW for Graduate studies in the US to foster Women Graduate Education - October 2023
- Fulbright fellowship for pursuing a Master's in Data Science at Northeastern University– August 2023
- Full tuition waiver for MS in Data Science at Northeastern University - August 2023.
- Scholarship from CONICET (National Council for Scientific and Technical Research) Argentina to pursue a doctoral degree. Funded by CONICET - April 2018.
- Diploma of Distinction for outstanding participation in the Oral Presentations at the VII World Congress on Quinoa and Other Andean Grains. Chile - April 2018.
- Scholarship from the National Institute of Cancer from Argentina to promote cancer research in the interior of Argentina. Funded by the Ministry of Health of Argentina - August 2017.
- 10+ peer-reviewed publications, including 1 book chapter. [Google Scholar](#).