DR. MORGAN C. THOMAS

E-mail: morganthomas263@gmail.com ORCID: 0000-0002-1610-3499
GitHub • Scholar • LinkedIn • Blog

Education

· **PhD in Generative Structure-Based Drug Design** (in collaboration with Nxera Pharma, UK)
Supervisors: Prof. Andreas Bender & Dr. Chris de Graaf 2019 - 2024
University of Cambridge, Cambridge, UK

· **MChem Pharmaceutical Chemistry (1^{st})** (with a year in industry)
Supervisor: Prof. Emma Raven 2013 - 2017
University of Leicester, Leicester, UK

Experience

· **Postdoctoral Researcher** (in collaboration with Johnson & Johnson, Belgium)
Supervisor: Prof. Gianni De Fabritiis 2023 - Current
Universitat Pompeu Fabra, Barcelona, Spain

In my current role, I design, validate, and publish reinforcement learning (RL) algorithms applied to generative molecular design with chemical language models (CLMs). This includes state-of-the-art algorithms with respect to learning efficiency, multi-agent RL to enhance chemical exploration, fixing CLMs to substructure constrained generation only (e.g., scaffold-based design or fragment linking), and co-developing and maintaining ACEGEN (the first TorchRL-based library for generative molecular design with RL). I have also continued to develop MolScore and propose new benchmarks for the community.

· **External Consultant** 2021 - 2023
Nxera Pharma, Cambridge, UK

I worked alongside my PhD studies to develop a custom data curation pipeline for in-house and public GPCR structural data. I optimised protein structural alignment using tools such as Biopython and coded a custom interactive interface using PyQt and PyMol to facilitate visualisation and expert annotation.

· **PhD Student** (in collaboration with Nxera Pharma, UK) 2019 - 2024
University of Cambridge, Cambridge, UK

My PhD focused on the application of generative models to structure-based drug design. This culminated in identifying novel nanomolar A_{2A} receptor ligands with confirmed functional activity and x-ray crystal structures, achieving an 88% hit rate with methods developed during my PhD. These methods included 1) a more efficient RL algorithm to reduce compute time enabling the use of docking for structure-based scoring in a practical time frame, and 2) a python package to evaluate, benchmark, and help apply generative models to drug design called MolScore. Additionally I contributed to a COVID-19 project by prioritising radiolabelled ACE2 probes through free-energy simulations with GROMACS on AWS' cloud computing platform.

· **Graduate scientist** 2017 - 2019
AstraZeneca, Cambridge, UK

This cross-discipline graduate programme within the IMED Biotech Unit provided me with invaluable experience on the challenges and processes of drug design within a large pharmaceutical company. The graduate programme consisted of three 8-month rotations:

Bioinformatics Conducted hypothesis driven analysis of gene expression data in R using software such as DESeq2 and GSVA to assess phenotypic drug effects and aid project understanding. In addition, I developed a custom in-house gene signature used to inform regulation of STING-induced pathways used in pre-clinical studies.

Computational Chemistry Structure-based modelling of compounds for cross-project support by use of ligand docking, virtual screening, molecular dynamics and FEP techniques utilising an array of software including Schrodinger, MOE, Gaussian, OpenEye. Modelled covalent warhead reactivity using Gaussian and Jaguar QM software to support covalent warhead design. I also volunteered to prototype an ML-model for generative molecular design with reinforcement learning to enrich project chemistry by creating de novo libraries based on different scoring functions.

Structural Chemistry Learned the characterisation and importance of free ligand conformation in solution for project compounds, including macrocycles, PROTACS, and peptides by utilising 1D and 2D NMR experiments (including: 1H, 13C, HSQC, ROESY). This involved enumeration of possible conformations in silico and prediction of respective 1H chemical shifts by QM. These results were interpreted to aid drug design and improve ligand binding kinetics (k_{on}) by making design changes to increase the population of bioactive conformation in solution.

· **Master's student** 2016 - 2017
University of Leicester, Leicester, UK

Investigated protein-ligand binding of hemin to the PAS domain of hERG3, including solving the first x-ray crystallographic structure of the hERG3 PAS domain. This also included cell culture, protein expression and purification, UV-VIS spectroscopy, EPR, Raman, and CD spectroscopy. Additionally conducted site-directed mutagenesis confirming hypothesised hemin-binding cysteine.

· **Industrial student** 2015 - 2016
Cambridge Research Biochemicals, Stockton-upon-tees, UK

Synthesis, purification, and quality control of short peptide chains (1-40 residues) via solid phase peptide synthesis. Gained hands-on experience in a production-driven SME environment.

Skills & Activities

Technical skills: Git, Unix/Shell, Python (RDKit, PyTorch, TorchRL, PyG, Scikit-learn, Biopython, Streamlit, PyQt), SQL, PyMol, Schrodinger (Glide, FEP), Gromacs, ChemDraw, MOE, ROCS, rDock, Vina

Collaboration: Consistently worked in industry-collaborative environments throughout Postdoctoral and PhD research, learning to effectively mediate sometimes conflicting objectives and feedback.

Mentoring: Successfully defined projects for, and supervised three visitors, one resulted in a publication and promotion with the other securing an MSCA funded doctoral program. Recently joined as a PhD committee board member to continue supervision duties.

Leadership: Took organisational ownership of events such as the Global Graduate Presentation Sessions. Led projects as an external consultant by defining and actioning project direction while cross-referencing with team members. Initiated and led grant applications to develop strategic, long-term research programs beyond individual projects.

Communication: Disseminated research at varying levels through peer-reviewed journals, book chapters, conferences presentations, and informal blog posts.

Peer-review: Frequent peer-reviewer of top scientific journals, including Nature Communications, Journal of Chemical Information and Modelling, Journal of Cheminformatics, and Journal of Computer-Aided Drug Design.

Awards & Funding

· **Postdoctoral scholarship** 2023
One of two successful candidates for the Flanders innovation and entrepreneurship (VLAIO) project.

· **PhD Studentship** 2019
Established a studentship wholly funded by Nxera Pharma, UK

· **Celltech Award** 2017
Highest grade in MChem Pharmaceutical chemistry

· **RSC Year in Industry Fellowship** 2016

Conference Presentations

· Talk - Cambridge Cheminformatics Meeting 20/02/2025
· Talk - Drug Discovery Chemistry (Europe) 04/12/2024
· Talk - UKQSAR 17/10/2024
· Poster - 9th Join Sheffield Conference on Chemoinformatics 21/06/2023
· Talk - ELLIS ML4Molecules workshop 28/11/2022
· Talk - MGMS Young Modellers' Forum **(Best Talk)** 25/11/2022
· Poster - UKQSAR **(Best Poster)** 08/11/2022
· Poster - 5th RSC/CICAG Artificial Intelligence in Chemistry 01/09/2022
· Talk - International Conference on Chemical Structures 13/06/2022
· Talk - 5th European Research Network on Signal Transduction 05/10/2021
· Talk - 4th RSC/CICAG Artificial Intelligence in Chemistry 27/09/2021
· Poster - 4th European Research Network on Signal Transduction 12/04/2021
· Poster - ELLIS ML4Molecules workshop 12/12/2020
· Talk - German Conference on Cheminformatics 02/11/2020
· Poster - 3rd RSC/CICAG Artificial Intelligence in Chemistry 28/09/2020
· Talk - Cambridge Cheminformatics Meeting 02/09/2020

References

Available upon request.

Publications

· **Thomas M**, Bou A, De Fabritiis G. Test-Time Training Scaling for Chemical Exploration in Drug Design. *submitted to Forty-Second International Conference on Machine Learning, available on arXiv*

· **Thomas M**, Bou A, De Fabritiis G. REINFORCE-ING Chemical Language Models in Drug Design. arXiv preprint arXiv:2501.15971. 2025

· Peter S, Siragusa L, **Thomas M**, Palomba T, Cross S, O'Boyle NM, Bajusz D, Ferenczy GG, Keseru GM, Bottegoni G, Bender B. Comparative Study of Allosteric GPCR Binding Sites and Their Ligandability Potential. Journal of Chemical Information and Modeling. 2024;64(21):8176-92.

· **Thomas M**, Matricon PG, Gillespie RJ, Napiórkowska M, Neale H, Mason JS, Brown J, Fieldhouse C, Swain NA, Geng T, O'Boyle NM, Deflorian F, Bender A, de Graaf C. Modern hit-finding with structure-guided de novo design: identification of novel nanomolar adenosine A2A receptor ligands using reinforcement learning. *accepted to Nature Communications, available on ChemRxiv*

· Bou A, **Thomas M**, Dittert S, Navarro C, Majewski M, Wang Y, Patel S, Tresadern G, Ahmad M, Moens V, Sherman W. ACEGEN: Reinforcement learning of generative chemical agents for drug discovery. Journal of Chemical Information and Modeling. 2024;64(15):5900-11.

· **Thomas M**, Ahmad M, Tresadern G, De Fabritiis G. PromptSMILES: prompting for scaffold decoration and fragment linking in chemical language models. Journal of Cheminformatics. 2024;16(1):77.

· **Thomas M**, O'Boyle NM, Bender A, De Graaf C. MolScore: a scoring, evaluation and benchmarking framework for generative models in de novo drug design. Journal of Cheminformatics. 2024;16(1):64.

Handa K, **Thomas MC**, Kageyama M, Iijima T, Bender A. On the difficulty of validating molecular generative models realistically: a case study on public and proprietary data. Journal of Cheminformatics. 2023;15(1):112.

· **Thomas M**, Bender A, de Graaf C. Integrating structure-based approaches in generative molecular design. Current Opinion in Structural Biology. 2023;79:102559.

· Staniszewska AD, Armenia J, King M, Michaloglou C, Reddy A, Singh M, San Martin M, Prickett L, Wilson Z, Proia T, Russell D, **Thomas M**, Delpuech O, O'Connor MJ, Leo E, Angell H, Valge-Archer V. PARP inhibition is a modulator of anti-tumor immune response in BRCA-deficient tumors. Oncoimmunology. 2022;11(1):2083755.

· **Thomas M**, O'Boyle NM, Bender A, De Graaf C. Re-evaluating sample efficiency in de novo molecule generation. arXiv preprint arXiv:2212.01385. 2022

· **Thomas M**, O'Boyle NM, Bender A, De Graaf C. Augmented Hill-Climb increases reinforcement learning efficiency for language-based de novo molecule generation. Journal of cheminformatics. 2022;14(1):68.

· **Thomas M**, Boardman A, Garcia-Ortegon M, Yang H, de Graaf C, Bender A. Applications of artificial intelligence in drug design: opportunities and challenges. Artificial Intelligence in Drug Design. 2022:1-59.

· **Thomas M**, Smith RT, O'Boyle NM, de Graaf C, Bender A. Comparison of structure-and ligand-based scoring functions for deep generative models: a GPCR case study. Journal of cheminformatics. 2021;13(1):39.

· Raubo P, Carbajo RJ, McCoull W, Raubo J, **Thomas M**. Diversity-orientated synthesis of macrocyclic heterocycles using a double SN Ar approach. Organic & Biomolecular Chemistry. 2021;19(28):6274-90.

· Burton MJ, Cresser-Brown J, **Thomas M**, Portolano N, Basran J, Freeman SL, Kwon H, Bottrill AR, Llansola-Portoles MJ, Pascal AA, Jukes-Jones R, Chernova T, Schmid R, Davies NW, Storey NM, Dorlet P, Moody PCE, Mitcheson JS, Raven EL. Discovery of a heme-binding domain in a neuronal voltage-gated potassium channel. Journal of Biological Chemistry. 2020;295(38):13277-86.

· McAulay K, Hoyt EA, **Thomas M**, Schimpl M, Bodnarchuk MS, Lewis HJ, Barratt D, Bhavsar D, Robinson DM, Deery MJ, Ogg DJ, Bernardes GJL, Ward RA, Waring MJ, Kettle JG. Alkynyl benzoxazines and dihydroquinazolines as cysteine targeting covalent warheads and their application in identification of selective irreversible kinase inhibitors. Journal of the American Chemical Society. 2020 May 15;142(23):10358-72.