Alexey Strokach

PhD Student • University of Toronto • Toronto, ON

alex.strokach@utoronto.ca | (647) 226-4521 | ostrokach | ostrokach.gitlab.io

Education

Dec. 2023 (Expected) — PhD in Computer Science — **University of Toronto**, Toronto, ON
Supervisor: Dr. Philip M. Kim
GPA: 3.93 / 4.0
Awards: • NSERC Postgraduate Scholarship (2018-2021, $63000)

Dec. 2016 — MSc in Computer Science — **University of Toronto**, Toronto, ON
Supervisor: Dr. Philip M. Kim
GPA: 4.0 / 4.0

Aug. 2013 — MSc in Pharmaceutical Sciences — **University of Toronto**, Toronto, ON
Supervisor: Dr. James W. Wells
GPA: 4.0 / 4.0
Awards: • Ontario Graduate Scholarship (2011-2012, $15000)

Aug. 2010 — BMSc in Medical Sciences (Honors) — **University of Western Ontario**, London, ON
GPA: 88.7 / 100
Awards:
- Dean's Honor List (2006-2010)
- CIS Academic All-Canadian in Cross Country and Track and Field (2009)
- Queen Elisabeth II Scholarship (2006-2009, $3500 x 4)
- Reinhard Konrad Memorial Awards in Science (2008, $1000)
- Laurene Paterson Estate Scholarship (2007, $2000)
- The Western Scholarship of Distinction (2006, $1000)

Work Experience

Jan. 2022–Present — Software Developer — **Google**, Toronto, ON
Working on machine learning models for detecting fraud in advertising.

May–Aug. 2021 — Software Engineering Intern — **Google**, Remote, ON
Implemented differentially private mechanisms and trained ML models using anonymized data.

May–Aug. 2019 — Software Engineering Intern — **Google**, Sunnyvale, CA
Designed and evaluated APIs for interacting with Apache Beam batch and streaming pipelines.

Sep. 2010–Present — Teaching Assistant — **University of Toronto**, Toronto, ON
Developed content, held tutorials, graded assignments and exams for the following courses:

• CSC413H – Neural Networks and Deep Learning	1 term
• CSC411H – Introduction to Machine Learning	1 term
• CSC384H – Introduction to Artificial Intelligence	3 terms
• CSC343H – Introduction to Databases	3 terms
• CSC311H – Introduction to Machine Learning	1 term
• CSC209H – Software Tools and Systems Programming	1 term
• CSC148H – Introduction to Computer Science	6 terms
• CSC128H – Content Developer for PythonTA	1 term
• CSC120H – Computer Science for the Sciences	1 term
• CSC108H – Introduction to Computer Programming	2 terms
• C4M – Computing for Medicine	1 term
• CS Help Center	1 term
• PHM140H / PHC301H – Molecular Pharmacology	2 terms
• PHM225H – Introduction to Biochemistry and Molecular Biology	1 term

May–Aug. 2009 — Summer Research Assistant — **University of Western Ontario**, London, ON
Supervisors: Dr. David W. Shoesmith and Dr. J. Clara Wren
Performed computational modeling of gas-phase radiolysis and carbon steel corrosion.

Miscellaneous

- Reviewer for the *Machine Learning in Computational Biology (MLCB)* workshop and the *Computational and Structural Biotechnology Journal*, *Biotechnology Letters* and *BMC Supplements* journals.
- Competed on Varsity Cross Country and Track and Field teams for five years.

Select Research Projects

2018–Present
Paper
Web server
Code
UofT News
The Varsity

Fast and Flexible Design of Novel Proteins (ProteinSolver)
Supervisor: Dr. Philip M. Kim
Technologies: Python, PyTorch, PyTorch Geometric, Voila, Docker, Kubernetes
- Pre-trained a **graph convolutional network** to reconstruct masked amino acids in proteins.
- Evaluated the quality of the learned embeddings using a variety of downstream applications.
- Generated new protein sequences which underwent successful experimental validation.
- Currently extending this approach to generate proteins with entirely "hand-drawn" topologies.

2017–2018
Rankings
Report
Code
Code AZ

Predicting Synergistic and Beneficial Drug Combinations
Supervisor: Dr. Philip M. Kim
Technologies: Python, Scikit-learn, XGBoost, Tensorflow, MySQL
- Created a **data warehouse** with information about drugs (rdkit), drug targets (STITCH), protein interactions (Mentha), protein functional associations (STRING), and cell line- and drug-dependent gene expression levels (LINCS), mutation profiles (CCLE), and essentiality scores (Achiles).
- Compared **decision tree**- and **neural network-based** models for predicting drug combinations.
- Participated in the AstraZeneca-Sanger Drug Combination Prediction DREAM Challenge.

2013–2016
Thesis
Web server
Slides (CAGI5)
Poster
ELASPIC
ELASPIC2

Predicting the Effect of Mutations on a Genome-Wide Scale
Supervisor: Dr. Philip M. Kim
Technologies: Python, Scikit-learn, XGBoost, Django, MySQL, AIOHTTP
- Created › 100k homology models of proteins and protein-protein interactions.
- Trained a **gradient boosted decision tree** model to predict the $\Delta\Delta G$ associated with mutations.
- Performed feature elimination while optimizing performance on external validation tasks.
- Implemented a web server backend which executes user-submitted jobs on a SLURM cluster.

2010–2013
Thesis
Slides
Poster (FRET)
Poster (SPT)

Probing the Oligomeric Status of G Protein-Coupled Receptors using Förster Resonance Energy Transfer (FRET) and Single-Particle Tracking (SPT)
Supervisor: Dr. James W. Wells
Technologies: SPT microscopy, FRET microscopy, MatLab, cell culture, molecular cloning, PCR
- Created DNA constructs for proteins tagged with fluorescent proteins or other labels.
- Expressed fluorescently-tagged proteins in cells and collected thousands of microscopy images.
- Derived **4 mathematical models** describing expected patterns of FRET for different interactions.
- Fit single particle fluorescence to Poisson distributions using custom CUDA kernels.

Publications

Strokach A, Kim PM. *Deep generative modeling for protein design*. **Current Opinion in Structural Biology**. 2021 (submitted).

Strokach A, Becerra D, Corbi-Verge C, Perez-Riba A, Kim PM. *Computational generation of proteins with predetermined three-dimensional shapes using ProteinSolver*. **STAR Protocols**. 2021; doi: 10.1016/j.xpro.2021.100505.

Strokach A, Lu TY, Kim PM. *ELASPIC2 (EL2): Combining contextualized language models and graph neural networks to predict effects of mutations*. **Journal of Molecular Biology**. 2021; doi: 10.1016/j.jmb.2021.166810.

Strokach A, Becerra D, Corbi-Verge C, Perez-Riba A, Kim PM. *Fast and flexible protein design using deep graph neural networks*. **Cell Systems**. 2020; 11: 1–10. doi: 10.1016/j.cels.2020.08.016.

Strokach A, Corbi-Verge C, Kim PM. *Predicting changes in protein stability caused by mutations using sequence- and structure-based methods in a CAGI5 blind challenge*. **Human Mutation**. 2019; 40: 1414–1423. doi: 10.1002/humu.23852.

Savojardo C, Petrosino M, Babbi G, Bovo S, Corbi-Verge C, **et al.**. *Evaluating the predictions of the protein stability change upon single amino acid substitutions for the FXN CAGI5 challenge*. **Human Mutation**. 2019; 0. doi: 10.1002/humu.23843.

Strokach A, Corbi-Verge C, Teyra J, Kim PM. *Predicting the Effect of Mutations on Protein Folding and Protein-Protein Interactions*. **Methods in Molecular Biology**. 2019; 1851: 1–17. doi: 10.1007/978-1-4939-8736-8_1.

Witvliet D*, **Strokach A***, Giraldo-Forero AF, Teyra J, Colak R, and Kim PM. *ELASPIC web-server: proteome-wide structure based prediction of mutation effects on protein stability and binding affinity*. **Bioinformatics** 32 (10): 1589–1591. doi: 10.1093/bioinformatics/btw031.

Conference Posters

Strokach A, Becerra D, Corbi-Verge C, Perez-Riba A, Kim PM *Designing real novel proteins using deep graph neural networks*. Machine Learning in Computational Biology (MLCB); **NeurIPS Workshop on Learning Meaningful Representation of Life (LMRL)**. Vancouver, Canada, 2019.

Strokach A, Wells JW. *Using Förster resonance energy transfer to distinguish between proteins that exist as monomers, transient oligomers and stable oligomers in the plasma membrane of living cells*. **Chemical Biophysics Symposium**. Toronto, Canada, 2013.

Strokach A, Wells JW. *Probing the Oligomeric Status of G Protein-Coupled Receptors by Single-Molecule Fluorescence*. **Great Lakes GPCR Retreat**. London, Canada, 2012.