

# Sebastian Raschka

## Contact Information

---

**Email**            [mail@sebastianraschka.com](mailto:mail@sebastianraschka.com) (private)  
                      [raschkas@msu.edu](mailto:raschkas@msu.edu) (work)

**Website**        <https://sebastianraschka.com>

**GitHub**         <https://github.com/rasbt>

**Google Scholar**    Sebastian Raschka

## Education

---

**2012-2017**      Biochemistry and Molecular Biology and Quantitative Biology  
                      Ph.D. - [Michigan State University](#)

Thesis title: Uncovering Hidden Patterns of Molecular Recognition

Main research topics:

- 1) *Development and application of a new framework for high-throughput virtual screening of millions of molecules for antagonist and agonist discovery that lead to pioneering work in aquatic invasive species control.*
- 2) *Development of machine learning protocols for identifying chemical discriminants of biological activity.*
- 3) *Development of a novel algorithm for native binding mode prediction by rigidity analysis of protein-ligand complexes based on graph theory.*
- 4) *Discovery and characterization of novel patterns of preferential interactions between proteins and their ligands through statistical analyses of intermolecular interaction patterns and chemical preferences with implications for protein and ligand design and drug discovery.*
- 5) *Invention of 3D epitope-based virtual screening – a novel approach for inhibiting protein-protein interactions that enabled the discovery of inhibitors of cancer metastasis.*

Advisor: Prof. Dr. Leslie A. Kuhn

**2008-2012**      B.Sc. Biology Plus International - [Heinrich-Heine-Universitaet Duesseldorf](#)  
Thesis: *Investigation of potential CLAVATA2 interactors by Fluorescence Resonance Energy Transfer Analysis*

Advisor: Prof. Dr. Ruediger Simon

**2011-2012** Study Abroad - Michigan State University

Research during the 1st semester: *Acetylation-dependent SUMOylation of p53*

Advisor: Prof. Dr. Min-Hao Kuo

Research during the 2nd semester: *Phylogenetic studies of Rhagoletis flies and associated parasitoid Coptera wasps*

Advisor: Prof. Dr. Jim Smith

**1996-2007** High School Diploma - Gymnasium Rheinkamp Europaschule Moers

## Professional Experience and Employment History

---

**2012-2017** Protein Structural Analysis & Design Lab Prof. Kuhn  
**present** Graduate Research Assistant

Collaborated with experimental biologists (MSU Department of Fisheries and Wildlife) and designed machine learning models in time-sensitive projects to successfully predict receptor antagonists of sea lamprey mating pheromones. We discovered inhibitors that could block 93% of pheromone reception at  $10^{-9}$  M concentration in electro-olfactogram assays and at  $10^{-12}$  M concentration in natural water streams.

Developed a hypothesis-driven methodology and Python software framework, Screenlamp, for large-scale virtual screening allowing researchers to identify target molecules of protein receptors at an unprecedented scale in databases of more than 15 million compounds.

Developed algorithms and a Python/C software package, SiteInterlock, for predicting 3-dimensional protein-ligand binding poses on the time scale of seconds with a predictive accuracy equal to existing approaches. This novel approach is based on the hypothesis that interfacial rigidification of both the protein and ligand prove to be important characteristics of the native binding mode and are sensitive to the spatial coupling of interactions and bond-rotational degrees of freedom in the interface.

Discovered clear and strong patterns of chemical group matching preferences for intermolecular hydrogen bonding across a database of non-homologous protein-ligand complexes. The results from these statistical analyses lead to new insights that have wide implications for protein and ligand design, for instance, that the specificity of ligands is owed to a narrow geometry when forming cognate, intermolecular H-bonds. Further, a chemical preference key metric was designed based on the patterns of chemical groups participating in H-bonding that enables the prediction of native protein-ligand complexes.

Developed a novel approach, 3D epitope-based virtual screening, for identifying small molecule inhibitors of protein-protein interactions. This new method was applied to screen 10 million commercially available molecules with drug-like properties and identified two candidates that were shown to be successful as inhibitors of cancer metastasis in experimental assays performed by collaborators.

Managed IT systems in the Kuhn lab, including hardware, software, servers, computing clusters, and Linux workstations.

**2017-  
present**

Joint Research Collaboration with the MSU i-PRoBe Lab  
*Research Collaborator*

Designed and implemented a novel artificial neural network architecture for deep learning based on ideas from the field of general adversarial networks with applications to facial recognition and biometrics. This novel architecture, *semi-adversarial neural networks*, showed unprecedented performance in perturbing gender information in large image databases while retaining the utility of face matching software.

**May 2012-  
Aug 2012**

Computational Pharmaceutical Chemistry Lab Dr. Holger Gohlke  
*Research Internship*

Research project: *Analysis of protein thermostability using a constraint network approach based on graph theory.*

Conducted molecular dynamics simulations and analyzed the resulting trajectories to predict the effect of binding-site mutations on the thermostability of various proteins.

**May 2010-  
Oct 2010**

Immulaab Duesseldorf  
*Internship*

Performed computational and statistical analyses of real-time qPCR data from human saliva samples to diagnose paradontose causing bacteria.

**Oct 2007-  
Mar 2008**

St. Thekla home for the elderly, Rheinberg, Germany  
*Community Service*

Assisted the elderly with household tasks and the participation in social activities.

## Teaching Experience

---

- Aug 2012-2017**      Project Design  
*Co-Supervisor*  
Composed projects for undergraduate and graduate students in the Protein Structure and Analysis Laboratory and trained students in computational and statistical subjects.
- Aug 2016**      Machine Learning with scikit-learn at [SciPy 2016](#), Austin, Texas  
*Instructor*  
Taught an 8-hour workshop on machine learning and scikit-learn (approx. 100 participants).
- May 2013**      Software Carpentry Instructor Training, Toronto, Canada  
Participated in a one-week Software Carpentry instructor training course.
- May 2014-  
Jul 2014**      BMB 401: Comprehensive Biochemistry, Michigan State University  
*Graduate Teaching Assistant*  
Held weekly office hours and organized weekly study groups to help students with the course material. Co-supervised examinations (approx. 300 students).
- Jun 2010-  
Aug 2010**      Bio 250: Genetics, Heinrich-Heine Universitaet Duesseldorf  
*Student Teaching Assistant*  
Co-supervised laboratory sessions and designed evaluated weekly written tests to help students with the course material and preparation for the exam (approx. 200 students).

## Honors and Awards

---

- *MSU BMB Outstanding Graduate Student Award (2016-2017)*
- *ACM Computing Reviews' Best of 2016* (Python Machine Learning book)

## Clubs/Extracurricular Activities

---

- Oct 2016-present**      **MSU Data Science Club**  
*Member and speaker*  
Advised MSU students in applying data science techniques to research problems.
- Jan 2017-June 2017**      **Udacity Deep Learning Nanodegree**  
*Continued education*  
Completed a 16-week course implementing various deep neural networks for text and image recognition and generation.
- Oct 2013-Apr 2014**      **Council of Graduate Students, Michigan State University**  
*Departmental representative*  
Participated in council meetings and discussions, representing the Department of Biochemistry and Molecular Biology.

## Technical Skills

---

- *Computational biology and structural modeling*  
Expert in using protein-ligand docking tools and protein-ligand scoring functions for pose-prediction and conducting large-scale ligand-based virtual screenings for agonist/antagonist discovery. Other skills include the use of molecular dynamics simulations, energy minimization, protonation and charge assignment, conformer analysis, 3-dimensional structure alignments and pharmacophore detection, homology modeling, and multiple sequence alignments (amino acid sequences and DNA) in research projects. Trained and helped supervise undergraduate research assistants and graduate rotation students.
- *Machine learning and deep learning*  
Author of the best-selling book *Python Machine Learning, Packt Publishing Ltd., 2015* and instructor of various machine learning workshops. Expert knowledge in parametric and non-parametric modeling for supervised learning. Expert knowledge in unsupervised learning, including clustering, dimensionality reduction, and manifold learning. Good familiarity with TensorFlow and capable of implementing deep artificial neural networks, including convolutional neural networks and recurrent neural networks with long short term memory, from scratch for large-scale image classification and speech recognition on distributed systems.
- *Statistics and probability theory*  
Experienced in using various statistical analyses, including frequentist methods for hypothesis testing and pattern recognition based on Bayesian decision theory.

- *Large-scale data analysis and visualization*  
Analysis of small and large datasets, primarily using Python libraries ([matplotlib](#) and [seaborn](#)) and R.
- *High performance computing*  
User of the High Performance Computing Cluster (HPCC) at Michigan State University and Amazon EC2 for high-throughput virtual screening and machine learning applications.
- *Git version control*  
Multi-year experience in reproducible research and the development and maintenance of open source projects under version control via GitHub and GitLab.
- *Programming Languages*  
Python (very strong knowledge, preferred language for scientific programming)  
R, C, Bash, SQL, HDF5 (strong knowledge)  
C++, JavaScript, HTML, CSS (good knowledge)  
Java, Scala (basic knowledge)

## Publications

---

### Books

- Raschka, Sebastian *Introduction to Artificial Neural Networks and Deep Learning (In prep.)* [<https://github.com/rasbt/deep-learning-book>].
- Raschka, Sebastian, Vahid Mirjalili. *Python Machine Learning. 2nd Edition*. Birmingham, UK: Packt Publishing, 2017. ISBN: 978-3958457331.
- Raschka, Sebastian, Vahid Mirjalili. *Python Machine Learning. 2nd Edition (German Transl.)*. Frechen, Germany: mitp Verlag, 2017. ISBN: 978-3958457331.
- Raschka, Sebastian, David Julian, and John Hearty. *Python: Deeper Insights into Machine Learning*. Birmingham, UK: Packt Publishing, 2016. ASIN: B01LD8K994.
- Raschka, Sebastian. *Python Machine Learning*. Birmingham, UK: Packt Publishing, 2015. ISBN: 978-1783555130. (Amazon bestseller in Data Mining 2015 & 2016, Packt bestselling title 2015 & 2016, more than 45,000 copies sold; ACM Computing Reviews' [Notable Computing Books and Articles of 2016](#))
- Raschka, Sebastian. *Python Machine Learning (Russian Transl.)*. Russia: DMK Press, 2017. ISBN: 978-5970604090.
- Raschka, Sebastian. *Python Machine Learning (Russian Transl.)*. Poland: Helios, 2017. ISBN: 978-8328336131.
- Raschka, Sebastian. *Python Machine Learning (Korean Transl.)*. Korea: Ji & Son, 2017. ISBN: 979-1187497035.

- Raschka, Sebastian. *Machine Learning mit Python (German Transl.)*. Frechen, Germany: mitp Verlag, 2016. ISBN: 978-3958454224.
- Raschka, Sebastian. *Machine learning con Python. Costruire algoritmi per generare conoscenza (Italian Transl.)*. Italy: Guida Completa, 2016. ISBN: 978-8850333974.
- Raschka, Sebastian. *Python Machine Learning (Japanese Transl.)*. Japan: Impress Top Gear, 2016. Japanese: 978-4844380603.
- Raschka, Sebastian. *Python Machine Learning (Chinese Transl.)*. New Taipei City, Taiwan: DrMaster Press Co., Ltd., 2016. ISBN: 978-9864341405.
- Raschka, Sebastian. *Heat Maps in R: How-To*. Birmingham, UK: Packt Publishing, 2013. ISBN: 978-1782165644. (more than 500 copies sold)

## Book Chapters

- Raschka, Sebastian, Leslie A. Kuhn, Anne M. Scott, and Weiming Li. *Computational Drug Discovery and Design: Automated Inference of Chemical Group Discriminants of Biological Activity from Virtual Screening Data*. Springer, 2017. [In press]

## Peer Reviewed Journals

- Mirjalili, Vahid, Sebastian Raschka, Anoop Namboodiri, and Arun Ross. *Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images*. [Accepted for publication] [[arXiv preprint](#)]
- Raschka, Sebastian, Alex Wolf, and Leslie A. Kuhn. *Protein-ligand interfaces are polarized: Discovery of a strong trend for intermolecular hydrogen bonds to favor donors on the protein side with implications for predicting and designing ligand complexes*. [Submitted]
- Raschka, Sebastian, Nan Liu, Santosh Gunturu, Anne M. Scott, Mar Huertas, Weiming Li, and Leslie A. Kuhn. *Facilitating the hypothesis-driven prioritization of small molecules in large databases: Screenlamp and its application to GPCR inhibitor discovery*. [In revision]
- Raschka, Sebastian. *BioPandas: Working with molecular structures in pandas DataFrames*. The Journal of Open Source Software 2.14 (2017).
- Raschka, Sebastian, Joseph Bemister-Buffington, and Leslie A. Kuhn. *Detecting the native ligand orientation by interfacial rigidity: SiteInterlock*. Proteins: Structure, Function and Bioinformatics 84.12 (2016): 1888-1901.

## Preprints

- Raschka, Sebastian. *MusicMood: Predicting the mood of music from song lyrics using machine learning*. arXiv preprint arXiv:1611.00138 (2016).

- Raschka, Sebastian. *Naive Bayes and Text Classification I-Introduction and Theory*. arXiv preprint arXiv:1410.5329 (2014).



## Selected Oral Presentations

---

- Raschka, Sebastian. *Building hypothesis-driven virtual screening pipelines for millions of molecules*. ODSC West. Nov 2017. San Francisco, CA.
- Raschka, Sebastian. *Introduction to Deep Learning with TensorFlow*. PyData Ann Arbor. Aug 2017. Ann Arbor, MI.
- Raschka, Sebastian. *Efficient large-scale virtual screening with molecular data frames*. SciPy 2017 Lightning Talk. Jul 2017. Austin, TX.
- Raschka, Sebastian, J. Bemister-Buffington, and L.A. Kuhn. *Binding pose prediction via rigidity analysis and large-scale pheromone antagonist discovery for invasive species control*. MSU Computational and Systems Biology Club. Mar 2017. East Lansing, MI.
- Raschka, Sebastian. *Machine Learning and Performance Evaluation – Overcoming the Selection Bias*. DataPhilly. Nov 2016. Philadelphia, PA.
- Raschka, Sebastian, J. Bemister-Buffington, and L.A. Kuhn. *Detecting the Native Ligand Orientation by Interfacial Rigidity*. BMB Departmental Retreat. Oct 2016. Okemos, MI.
- Raschka, Sebastian. *Getting Started with Data Science*. MSU DataScience. Sep 2016. East Lansing, MI.
- Raschka, Sebastian. *Learning Scikit-Learn*. PyData 2016. Aug 2016. Chicago, IL.
- Raschka, Sebastian, and A. C. Mueller. *Machine Learning with Scikit-Learn*. SciPy 2016. Jul 2016. Austin, TX.
- Raschka, Sebastian, J. Bemister-Buffington, and L.A. Kuhn. *A Novel Approach to Protein-Ligand Binding Mode Prediction by Rigidity Analysis Using Graph Theory*. Biomolecular Science Gateway. Feb 2016. East Lansing, MI.
- Raschka, Sebastian. *Practical Data Science. An Introduction to Supervised Machine Learning and Pattern Classification*. MSU NextGen Bionformatics. Feb 2015. East Lansing, MI.
- L. A. Kuhn, N. Liu, A. Scott, S. Gunturu, S. Raschka, M. Huertas, and W. Li, *Screening and Identifying Pheromone Receptor Antagonists for Invasive Species Control*, ACS Division of Environmental Chemistry Session: Great Lakes Restoration Initiative: An Environmental Chemistry Challenge”, 248th ACS National Meeting and Exposition San Francisco, CA, August 10-14, 2014

## Poster Presentations

---

- S. Raschka, N. Liu, S. Gunturu, A.M. Scott, M. Huertas, W. Li, and L.A. Kuhn *Screenlamp: A Software Framework for Hypothesis-Driven Ligand Discovery Based on Virtual Screening and Machine Learning*. Great Lakes Bioinformatics and Canadian Computational Biology Conference 2016, University of Toronto, May 16-19, 2016
- A. Scott, M. Huertas, S. Raschka, N. Liu, L.A. Kuhn, and W. Li, *Regulation of Pheromone Induced Responses with Behavioral Antagonists for Invasive Sea Lamprey (Petromyzon marinus L.) Control*. Behaviour 2015 conference, Cairns, Australia, August 8, 2015
- S. Raschka, N. Liu, S. Gunturu, A.M. Scott, M. Huertas, W. Li, and L.A. Kuhn, *A Hypothesis-Driven Virtual Screening Methodology for Structure-Based Inhibitor Discovery*. Quantitative Biology Symposium, Michigan State University, East Lansing, August 18, 2014
- A. Scott, M. Huertas, N. Liu, S. Gunturu, S. Raschka, L. A. Kuhn, and W. Li, *Identification of pheromone receptor antagonists for invasive sea lamprey control*. International Congress on the Biology of Fish, Heriot-Watt University, UK, 3-7 August, 2014

## Peer Reviewing

---

- Journal of Machine Learning Research (JMLR)
- American Chemical Society (ACS) Publications
- The Journal of Open Source Software (JOSS)
- Scientific Reports, Nature Publishing Group

## Open Source Projects and Contributions

---

- Contributor to the Python machine learning library [scikit-learn](#)
- Contributor to the deep learning library [TensorFlow](#)
- Creator and maintainer of molecular data frame library [BioPandas](#)
- Creator and maintainer of the Python data science library [mlxtend](#)
- Creator and maintainer of [Screenlamp](#), A Python toolkit for facilitating hypothesis-driven, ligand-based screening of large molecule libraries
- Journal of Open Source Software reviewer

- Member of the [conda-forge](#) organization, which provides recipes, build infrastructure, and distributions for the conda package manager
- Member of the Python Software Foundation
- An extended open source portfolio is available on GitHub: <https://github.com/rasbt>

## Selected Journalism Contributions

---

- Computer Business Reviews, 2017: *The Future of Machine Learning and Data Science*
- KDnuggets, 2016: *Top 2016 KDnuggets Stories*
- KDnuggets, 2016: *Data Preparation Tips, Tricks, and Tools: An Interview with the Insiders*
- Huffington Post, 2016: *The Impact of Machine Learning on Healthcare*
- Forbes, 2016: *What Do Computational Biologists Do?*
- Hacker Bits print magazine, 2016: *Everyone is talking about deep learning these days, but what the heck is it?*
- Open Data Science, 2016: *Trifecta: Python, Machine Learning, and Dueling Languages A.I.*
- Open Data Science, 2016: *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*
- Partially Derivative Podcast Interview, 2017: *Model Evaluation with Sebastian Raschka*
- Data Science at Home Podcast Interview, 2016: *30 min with data scientist Sebastian Raschka*
- Becoming a Data Scientist Podcast, 2016: *30 min with data scientist Sebastian Raschka*

## Selected Press Coverage

---

- A.I. & Optimization, 2016: *Top 22 Python Programming Books*
- KDnuggets, 2016: *Top 10 Essential Books for the Data Enthusiast*
- KDnuggets, 2016: *5 Machine Learning Projects You Can No Longer Overlook*
- Techopedia, 2016: *Artificial Intelligence: Experts To Follow on Twitter*
- InformationWeek, 2016: *10 Big Data Books To Boost Your Career*
- OpenSource.org, 2015: *Publisher's picks: 29 open source books for 2015*
- Datawerq, 2015: *50 Data leaders to follow on Twitter*
- Onanalytica, 2015: *Artificial Intelligence & Machine Learning: Top 100 Influencers and Brands*
- Analytics Vidhya, 2015: *Top Data Scientists to follow on GitHub*