

MARTIN STEINEGGER

+82 · 2 · 880 · 4438 ◇ martin.steinegger@snu.ac.kr

502-423 1 Gwanak-ro, Gwanak-gu, ◇ Seoul, Korea

Born 29.01.1985 in Erding. German citizen.

CURRICULUM VITAE

Education

- 08/2014 - 08/2018 Ph.D. in Computer Science at the Technical University Munich (Passed with summa cum laude)
- 04/2013 - 08/2014 Master of Science in Computer Science at the Ludwig Maximilian University (Passed with merit)
- 09/2010 - 04/2013 Bachelor of Science in Bioinformatics at TU Munich / Ludwig Maximilian University
- 09/2006 - 07/2008 Business Informatics at the EDV-Schule Plattling (Passed as second best student)
- 10/2001 - 06/2005 Computer Engineering at the HTL Braunau (Technical college for electronics)

Research and Industry experience

- since 03/2024 Associate Professor at Seoul National University (tenured). *full-time*
- 03/2020 - 02/2024 Assistant Professor at Seoul National University. *full-time*
Laboratory of Machine Learning & Bioinformatics
- 10/2018 - 02/2020 Postdoctoral Fellow at the Salzberg Lab at the Johns Hopkins University School of Medicine. *full-time*
Pathogen detection in human metagenomic data.
- 08/2014 - 09/2018 PhD Student at the Quantitative and Computational Biology Laboratory at the Max-Planck Institute for Biophysical Chemistry. *full-time*
Ultrafast and sensitive sequence search methods in the era of next generation sequencing.
- 04/2016 - 07/2018 Collaboration with Seok Lab at the Seoul National University.
Large-scale de-novo structure prediction based on coevolution analysis of metagenomics-enriched multiple sequence alignments.
- 03/2015 - 04/2016 Collaboration with Notredame Lab at the Centre for Genomic Regulation in Barcelona.
Supporting the development of a large scale multiple sequence aligner.
- 08/2014 - 12/2014 Visiting Scientist at the Seok Lab, Seoul National University. *full-time*
Improving energy calculation for docking and protein structure prediction.
- 05/2012 - 07/2014 Research assistant at the Soeding Lab, Gene Center, LMU Munich. *part-time*
Improving HMM remote homologues protein search method.
- 08/2013 - 10/2013 Visiting Scientist at the Sali Lab, UCSF. *full-time*
Implementing a Bayesian inference framework to determine enzyme pathways.
- 07/2011 - 05/2012 Visiting Scientist at Rost Lab, Technical University Munich. *part-time*
Full In-Silico mutagenesis of the human proteome using the Cloud.
- 06/2011 - 05/2012 Technical Architect / Scrum Master at Medability. *part-time*
Developing a haptic surgery simulator
- 09/2008 - 06/2011 Software Engineer / Security Tester / Performance engineering at Accenture Technology Solutions. *full-time*
- 09/2007 - 01/2008 Software Engineer / Technical Architect at visionary people AG. *freelancer*
- 09/2005 - 06/2006 Bezirkskrankenhaus Haar. Military service (community service) *full-time*,

ACHIEVEMENTS AND QUALIFICATIONS

Distinctions

2025	Fellow , The Institute Invitation-only fellowship promoting interdisciplinary solutions for public benefit.
2024	Overton Prize , International Society for Computational Biology Annual award recognizing a computational biologist for outstanding accomplishments.
2024	Chair , RECOMB 2025 (Seoul) Organized flagship comp. bio. conference with 1,300+ attendees from 33 countries.
2024	Highly Cited Researcher , Clarivate Recognized among the top 1% most-cited researchers in biology and bioinformatics.
2024	Research Award , College of Natural Sciences, Seoul National University Awarded for innovative interdisciplinary research in structural bioinformatics.
2023	Remarkable Outputs Award , Swiss Institute of Bioinformatics Recognized our work to large-scale protein structure clustering.

Awards

2025	Microbes Mercator Fellow
2024	TUM Global Visiting Professor
2018	Poster award at the ECCB 2018
2016	Poster award at the Critical Assessment of Protein Structure Prediction 12 Conference
2015	Max Planck PhD fellowship
2013	Winner of the Twilio prize (~1000\$) at the Disrupt TechCrunch Hackathon (1200+ attendees)
2012	Excellence initiative research grant, Ludwig Maximilian University
2011	Finalist in the Big Data Challenge, CycleComputing
2008	Master prize of the Bavarian state government, EDV-Schule Plattling

Certificates

2011	Certified ScrumMaster (CSM)
2010	ASDA Application Developer (Massachusetts Institute of Technology / Accenture)
2010	Information Technology Infrastructure Library V3 Foundation
2010	SpringSource Certified Spring Professional
2010	ISTQB Certified Tester
2009	Sun Certified Java Programmer
2008	IBM Certified System Administrator

Languages

German	Native
English	Fluent
Korean	Beginner

TALKS, POSTERS, AND PUBLICATIONS

Talks

Delivered 80 talks across 17 countries; keynotes are highlighted in red.

05/2025	MPI-MR, Germany, Fast Search Methods to Organize the Structural Protein Universe
05/2025	xAlxBIO, Korea, Fast Search Methods to Organize the Structural Protein Universe
03/2025	Keystone, USA, Fast Search Methods to Organize the Structural Protein Universe
03/2025	Mosbach Symposia, Germany, Supercharged Protein Analysis in the era of AI
02/2025	KOGO, Korea, Supercharged Protein Analysis in the era of AI
11/2024	ETH, Switzerland, Methods related to the Nobel Prize of Chemistry 2024
11/2024	UNIL, Switzerland, Methods related to the Nobel Prize of Chemistry 2024
11/2024	ri.MED, Italy, Methods related to the Nobel Prize of Chemistry 2024
10/2024	LMU, Germany, Methods related to the Nobel Prize of Chemistry 2024
07/2024	ISMB 2024, Canada, Supercharged Protein Analysis in the era of AI
07/2024	ISCB Student Council, Canada, Metagenomic sequence analysis: from protein sequences to structures
07/2024	SMBE 2024, Mexico, Supercharged Protein Analysis in the era of AI
06/2024	EMBO Workshop, France, Supercharged Protein Analysis in the era of AI
05/2024	Wenner-Gren Symposia, Sweden, Supercharged Protein Analysis in the era of AI
05/2024	Quadram, UK, Metagenomic analysis: from sequence to structures
03/2024	UNIST, Korea, Supercharged Protein Analysis in the era of AI
02/2024	UCL, UK, Supercharged Protein Analysis in the era of AI
01/2024	UNIST, LMs and AI for Protein Analysis
12/2023	Tokyo University, Japan, Supercharged Protein Analysis in the era of AI
12/2023	DTMBIO, Japan, Supercharged Protein Analysis in the era of AI
12/2023	World Bio Innovation Forum, Online, Metagenome annotation in the era of next generation protein structure prediction
12/2023	LG, Korea, Supercharged Protein Analysis in the era of AI
11/2023	PSI seminar, China, Structure analysis in the era of next-generation structure prediction
11/2023	KSBI, Korea, Clustering predicted structures at the scale of the known Protein Universe
10/2023	University of Auckland, New Zealand, Structure analysis in the era of next-generation structure prediction
10/2023	KoSAIM, Korea, Clustering predicted structures at the scale of the known Protein Universe
10/2023	Sookmyeong University, Clustering predicted structures at the scale of the known Protein Universe
10/2023	Swedish/Korean metagenomics meeting, From sequence to structure
09/2023	ShanghaiTech, China, Clustering predicted structures at the scale of the known Protein Universe
08/2023	KRIBB, Korea, From protein sequence to structure
07/2023	CASP, USA, Clustering predicted structures at the scale of the known Protein Universe
06/2023	Korean In silico bioDesign and Discovery Society, Korea, Clustering predicted structures at the scale of the known Protein Universe
06/2023	Korean Society for Structural Biology, Korea, Clustering predicted structures at the scale of the known Protein Universe
06/2023	Joint Symposium of Hanyang Institute of Bioscience and Biotechnology, Korea, Clustering predicted structures at the scale of the known Protein Universe
03/2023	SAP, Germany, From protein sequence to structure

02/2023	University of Toronto, Canada, From protein sequence to structure
02/2023	Western, Canada, From protein sequence to structure
02/2023	Harvard, USA, Foldseek: fast and accurate protein structure search
02/2023	Stanford, USA, From protein sequence to structure
01/2023	MBU50, India, From protein sequence to structure
01/2023	Norwegian Biochemistry Society Meeting, Norway, From protein sequence to structure
01/2023	International Symposium on Structure and Folding of Disease Related Proteins, Korea, Foldseek: fast and accurate protein structure search
12/2022	ISCB-Asia/GIW, Taiwan, From protein sequence to structure
11/2022	Hanyang University, Korea, Next generation protein analysis tools in the ear of highly accurate protein structure prediction
11/2022	SNU Pharmaceutical department, Korea, Metagenomic sequence classification: from sequences to structures.
10/2022	Sungkyunkwan University, Korea, Next generation protein analysis tools in the ear of highly accurate protein structure prediction
09/2022	UNIST, Korea, Next generation protein analysis tools in the ear of highly accurate protein structure prediction
08/2022	Korea Brain Research Institute, Korea, Next generation protein structure analyze with ColabFold and Foldseek
06/2022	Korea Institute For Advanced Study, Korea, Fast structure prediction and search
05/2022	NWO Life, Nederland, Mega scale protein structure prediction and search
05/2022	Nobel Symposium, Sweden, Mega scale protein structure prediction and search
05/2022	Yonsei, Korea, Mega scale protein structure prediction and search
04/2022	Microbiome Forum Johns Hopkins, USA, Metagenomic sequence classification: from sequences to structures.
02/2022	BASF, Germany, Next generation protein analysis tools in the ear of highly accurate protein structure prediction
02/2022	KMB 2021, Korea, Mega scale protein structure prediction and search
11/2021	Swiss Institute of Bioinformatics, Switzerland, Next generation protein analysis tools in the ear of highly accurate protein structure prediction
11/2021	KSMCB 2021, Korea, Mega scale protein structure prediction and search
08/2021	Boston Protein Design and Modeling Club, USA, ColabFold - Making protein folding accessible to all via Google Colab!
07/2021	BiATA Conference, Russia, MMseqs2 profile/profile: fast and ultra sensitive searches beyond the twilight zone
06/2021	BVCN Conference, USA, Metagenomic pathogen detection using MMseqs2, Plass, and Linclust
12/2020	MicroEvo Meeting Informatics, Denmark, The unresolved dying of the Mariana crows
09/2020	Genome Informatics, UK, Protein-guided nucleotide viral genome assembly for huge metagenomic datasets
09/2019	University of Salzburg, Austria, New algorithms and tools for large-scale sequence analysis of metagenomic data
05/2019	University of Konstanz, Germany, New algorithms and tools for large-scale sequence analysis of metagenomic data
04/2019	RECOMB-SEQ 2019, USA, New algorithms and tools for large-scale sequence analysis of metagenomics data
01/2019	Seoul National University, Republic of Korea, Metagenomics data analysis on steroids

10/2018	Johns Hopkins University, USA, Metagenomics data analysis on steroids
09/2018	Max Planck Institute for Marine Microbiology, Germany, Metagenomics data analysis on steroids
07/2018	BiATA 2018, Russia, New algorithms and tools for large-scale sequence analysis of metagenomics data
07/2018	ISMB 2018, USA, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets
04/2018	European Bioinformatics Institute, England, Fast and sensitive protein sequence search, clustering and assembly tools for the analysis of massive metagenomics datasets
04/2018	NGS 2018, Spain, Fast and sensitive protein sequence search, clustering and assembly tools for the analysis of massive metagenomics datasets
01/2018	Johns Hopkins University, USA, Search, Clustering and Assembly tools for huge metagenomics datasets
01/2018	Rutgers University, USA, Search, Clustering and Assembly tools for huge metagenomics datasets
05/2017	Tokyo University, Japan, MMseqs2 / Linclust
05/2017	National Institute of Advanced Industrial Science, Japan, MMseqs2 / Linclust
06/2016	SocBIN2016, Russia, Sensitive protein sequence searching for the analysis of massive data sets
06/2015	Beijing Genomics Institute, China, HH-suite for sensitive protein sequence searching. / MMseqs for protein search
05/2015	Quest for Orthologs 4, Spain, MMseqs for clustering huge protein sets
03/2015	European Bioinformatics Institute, England, Sequence clustering and search in the era of NGS
06/2014	ISCB NGS14, Spain, MMseqs suite for fast and sensitive batch searching
06/2014	Hadoop User Group, Germany, In-Silico mutagenesis on Amazon EMR
09/2012	GMDS, Germany, Cloud architecture for In-Silico mutagenesis
12/2011	EDAM Meeting, Netherlands, Cloud architecture for PredictProtein
07/2010	University Cologne, Germany, Application Security
01/2010	Accenture community meeting, Germany, Web security

Poster

I have presented **40** poster as the Principal Investigator

11/2019	Genome Informatics 2019, USA, Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank
11/2019	Genome Informatics 2019, USA, New algorithms and tools for large-scale sequence analysis of metagenomic data
09/2018	ECCB18, USA, MMseqs2 desktop and local web server app for fast, interactive sequence searches
07/2018	ISMB 2018, USA, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets
04/2017	ISMB NGS 2017, Spain, Sensitive protein sequence searching for the analysis of massive data sets
12/2016	CASP12, Italy, Sensitive protein sequence searching for the analysis of massive data sets
04/2016	ISMB NGS 2016, Spain, Sensitive protein sequence searching for the analysis of massive data sets
03/2016	ABLS 2016, Belgium, Fast and sensitive searching of proteomic data
05/2015	Quest for Orthologs 4, Spain, MMseqs for clustering huge protein sets
09/2014	KIAS Conference on Protein Structure and Function, Republic of Korea, Accelerated pairwise HMM alignment using SIMD programming and improved secondary structure scoring

Research Grants

Money raised **PI 5.0 Mio., total 97.3 Mio. USD** (* indicates a joint grant)

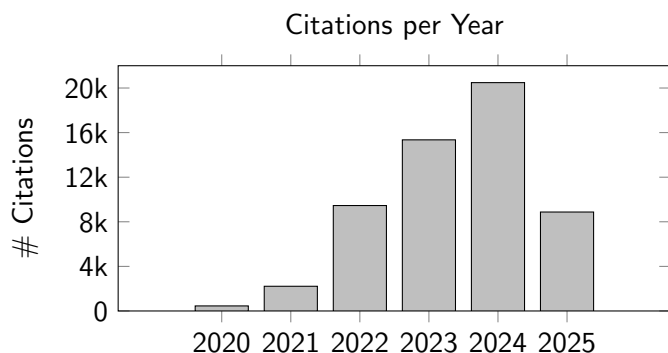
2024–2027	* NRF (Korea) and DFG (Germany) <i>Accurate and fast AI-based methods for predicting and classifying structurally resolved protein interactomes</i> 450 Mio KRW (317,617 USD); total 569,009 USD
2024–2031	* Novo Nordisk Foundation Ancient Environmental Genomics Initiative for Sustainability (AEGIS) PI 1,488 Mio KRW (974,026 USD); total 124,387.85 Mio KRW (87,768,576 USD)
2024–2029	National Research Foundation of Korea <i>De-novo enzyme design platform and its application in industrial enzyme development</i> PI 2,420 Mio KRW (1,707,833 USD); total 5,700 Mio KRW (4,022,582 USD)
2023–2025	Seoul National University <i>Accurate genomic annotation through a homology-aware AI model</i> PI 200 Mio KRW (141,153 USD)
2022–2025	Samsung <i>Rapid and precise diagnosis of infectious diseases using metagenomics</i> 90 Mio KRW (63,519 USD)
2021–2026	National Research Foundation of Korea <i>Folding the protein universe (FoldU): Metagenomics-scale protein structure prediction using machine learning</i> 728.581 Mio KRW (514,208 USD)
2021–2025	* National Research Foundation of Korea <i>Development of Cryo-EM/ET technology for 3D bio-imaging at molecular resolution</i> 388.3 Mio KRW (274,068 USD); total 2,975 Mio KRW (2,100,395 USD)
2020–2029	Seoul National University, Creative-Pioneering Researchers Program <i>Petasearch: Surveilling pathogens on a global scale</i> 568 Mio KRW (408,956 USD)
2020–2024	* National Research Foundation of Korea <i>In silico protein design by artificial intelligence and physical chemistry</i> 566.77 Mio KRW (400,120 USD); total 2,220 Mio KRW (1,567,243 USD)
2020–2023	National Research Foundation of Korea <i>Discovery of novel genomes through protein-guided assembly</i> 150 Mio KRW (105,865 USD)
2020–2021	Seoul National University, New Faculty Startup Fund <i>Capture probe design in the era of next-generation sequencing</i> 40 Mio KRW (28,230 USD)
2012	AMD Research Grant <i>~700 USD (one graphics card)</i>
2012	NVIDIA Research Grant <i>~3,000 USD (two graphics cards)</i>
2011	Amazon Research Grant <i>10,000 USD (AWS credits)</i>

Features of work or interviews

Interviews or coverages of our work.

- [1] Weijie Zhao (2025) New methods are revolutionizing biology: an interview with Martin Steinegger, *NSR*, doi: 10.1093/nsr/nwaf142
- [2] Andrea Lius (2025) Innovative platform empowers scientists to transform venoms into therapeutics, *American Society for Biochemistry and Molecular Biology*
- [3] Mallory L Wiper (2024) The 2024 ISCB Overton Prize Award—Dr Martin Steinegger, *Bioinformatics*, doi: 10.1093/bioinformatics/btae288
- [4] Andrea Lius (2025) Innovative platform empowers scientists to transform venoms into therapeutics, *American Society for Biochemistry and Molecular Biology*
- [5] Sara Reardon (2024) Five tasks that still challenge protein designer, *Nature*, doi: 10.1038/d41586-024-03595-9
- [6] Ewen Callaway (2024) Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures, *Nature*, doi: 10.1038/d41586-024-03214-7
- [7] Laurel Oldach (2024) Structure search suggests roles for thousands of viral proteins *c&en* ISSN 0009-2347
- [8] Raphael Rashid (2024) What will it take to open South Korean research to the world? *Nature Index* doi: 10.1038/d41586-024-02685-y
- [9] Laura Tran (2023) Lasker Award for Revolutionizing Protein Structure Predictions, *The Scientist*
- [10] Ewen Callaway (2023) The world's largest proteins? These mega-molecules turn bacteria into predators, *Nature*, doi: 10.1038/d41586-023-03937-z
- [11] Ewen Callaway (2023) 'A Pandora's box': map of protein-structure families delights scientists, *Nature*, doi: 10.1038/d41586-023-02892-z
- [12] Arunima Singh (2023) Speedier protein structure search, *Nature Methods*, doi: 10.1038/s41592-023-01953-5
- [13] Matthew Hutson (2023) Foldseek gives AlphaFold protein database a rapid search tool, *Nature*, doi: 10.1038/d41586-022-02083-2
- [14] Laurel Oldach (2022) Advancing structural biology to blazing speed, *ASBMB Today*
- [15] Ewen Callaway (2022) AlphaFold's new rival? Meta AI predicts shape of 600 million proteins *Nature*, doi: 10.1038/d41586-022-03539-1
- [16] Ewen Callaway (2022) 'The entire protein universe': AI predicts shape of nearly every known protein, *Nature*, doi: 10.1038/d41586-022-02083-2
- [17] Ewen Callaway (2022) What's next for AlphaFold and the AI protein-folding revolution, *Nature*, doi: 10.1038/d41586-022-00997-5
- [18] Henrik Müller (2022) Computerbasierte Proteinstruktur Vorhersage *Laborjournal* (in German)
- [19] Michael Eisenstein (2021) Artificial intelligence powers protein-folding predictions, *Nature*, doi: 10.1038/d41586-021-03499-y
- [20] Henrik Müller (2021) Interview mit Martin Steinegger über AlphaFold2 und ColabFold *Laborjournal* (in German)
- [21] Nikki Forrester, (2021) How new principal investigators tackled a tumultuous year, *Nature*, doi: 10.1038/d41586-021-01311-5
- [22] Lin Tang, (2020) Contamination in sequence databases, *Nature Methods*, doi: 10.1038/s41592-020-0895-8

Google Scholar - May 13 2025



	All	Since 2019
Citations	57,372	57,002
h-index	32	31
i10-index	54	53

Total 62 publications — Avg. Citation per Publication 927, Avg. Impact 189.1.

Listed below are papers with Impact > 8, ranked by Impact; “#” is rank, “*” marks Steinegger first/corresponding author.

#	TotalCitations	Impact	Year	First/Corresponding	Journal
1	35169	7033.80	2021		Nature
2	7060	1765.00	2022	*	Nat. Methods
3	3262	362.44	2017	*	Nat. Biotechnol.
4	2088	417.60	2021		IEEE TPAMI
5	1440	480.00	2023	*	Nat. Biotechnol.
6	1071	153.00	2019	*	BMC Bioinf.
7	838	104.75	2018	*	Nat. Commun.
8	776	258.67	2023		Nucleic Acids Res.
9	713	71.30	2016	*	Nucleic Acids Res.
10	697	116.17	2020		Curr. Protoc. Bioinf.
11	500	71.43	2019		Bioinformatics
12	445	63.57	2019	*	Nat. Methods
13	396	99.00	2022	*	Nat. Protocols
14	392	78.40	2021		Proteins
15	250	25.00	2016	*	Bioinformatics
16	246	49.20	2021		Nucleic Acids Res.
17	224	74.67	2023	*	Nature
18	212	35.33	2020		CASP14 abstract book
19	211	42.20	2021		Bioinformatics
20	209	34.83	2020	*	Genome Biol.
21	132	66.00	2024		NAR Genomics and Bioinformatics
22	101	25.25	2022		eLife
23	79	26.33	2023		Commun. Biol.
24	66	13.20	2021		Nucleic Acids Res.
25	57	19.00	2023		Trends in Biochemical Sciences
26	52	10.40	2021		Biodesign
27	42	21.00	2024	*	Nat. Protocols
28	40	13.33	2023		Biorxiv
29	35	8.75	2022		Front. Bioinf.
30	33	8.25	2022		eLife
31	29	9.67	2023		Genome Biol.
32	26	8.67	2023		Genome Biol.
33	18	9.00	2024	*	Nucleic Acids Res.
34	17	8.50	2024	*	bioRxiv
35	17	8.50	2024	*	Nat. Methods
36	14	14.00	2025	*	Nat. Methods
37	12	12.00	2025		bioRxiv

Publications

The most important articles are highlighted in red.

Current preprints

- [1] Lee S., Kim J., Mirdita M., **Steinegger M.** (2025) Easy and interactive taxonomic profiling with Metabuli App *bioRxiv*, doi: 10.1101/2025.03.10.642298 (under review Bioinformatics)
- [2] Santus L., Espinosa-Carrasco J., Rauschnig L., Mir-Pedrol J., Trujnara I., Vignoli A., Mansouri L., Baltzis A., Floden E.W., Di Tommaso P. and others (2025) An nf-core framework for the systematic comparison of alternative modeling tools: the multiple sequence alignment case study *bioRxiv*, doi: 10.1101/2025.03.14.642603
- [3] Yeo J., Han Y., Bordin N., Lau A.M., Kandathil S.M., Kim H., Levy Karin E., Mirdita M., Jones D.T., Orengo C., **Steinegger M.** (2025) Metagenomic-scale analysis of the predicted protein structure universe *bioRxiv*, doi: 10.1101/2025.04.23.650224
- [4] Fernandez-Guerra A., Wörmer L., Borrel G., Delmont T.O., Elberling B., Elvert M., Eren A.M., Gribaldo S., Henriksen R.A., Hinrichs K-U. and others (2025) Two-million-year-old microbial communities from the Kap København Formation in North Greenland *bioRxiv*, doi: 10.1101/2023.06.10.544454 (under review Nature)
- [5] Kallenborn F., Chacon A., Hundt C., Sirelkhatim H., Didi K., Dallago C., Mirdita M., Schmidt B., **Steinegger M.** (2025) GPU-accelerated homology search with MMseqs2 *bioRxiv*, doi: 10.1101/2024.11.13.623350 (under review Nature Methods)
- [6] Gilchrist C.L.M., Mirdita M., **Steinegger M.** (2024) Multiple protein structure alignment at scale with FoldMason *bioRxiv*, doi: 10.1101/2024.08.01.606130 (under review Science)
- [7] St. John P., Lin D., Binder P., Greaves M., Shah V., St. John J., Lange A., Hsu P., Illango R., Ramanathan A. and others (2024) BioNeMo Framework: A Modular, High-Performance Library for AI Model Development in Drug Discovery *arXiv*, doi: 10.48550/arXiv.2411.10548
- [8] Caroline Puente-Lelievre, Ashar J. Malik, Jordan Douglas, David Ascher and others (2023) Tertiary-interaction characters enable fast, model-based structural phylogenetics beyond the twilight zone *bioRxiv*, doi: 10.1101/2023.12.12.571181
- [9] Tymor Hamamsy, Meet Barot, James T. Morton, **Martin Steinegger**, Richard Bonneau, Kyunghyun Cho (2023) Learning sequence, structure, and function representations of proteins with language models *bioRxiv*, doi: 10.1101/2023.11.26.568742
- [10] Weissenow, K., Heinzinger, M., **Steinegger, M.**, and Rost, B. (2022) Ultra-fast protein structure prediction to capture effects of sequence variation in mutation movies, *bioRxiv*, doi: 10.1101/2022.11.16.471726
- [11] Vanni, C., Schechter, M., Delmont, T., and others (2021), AGNOSTOS-DB: a resource to unlock the uncharted regions of the coding sequence space *bioRxiv*, doi: 10.1101/2021.06.07.447314

Peer-reviewed manuscripts

2025 (9)

- [1] Kim D., Park S., **Steinegger M.** (2025) Unicore enables scalable and accurate phylogenetic reconstruction with structural core genes *GBE*, doi: 10.1093/gbe/evaf109
- [2] Park J.B., Lee G., Han Y-Y., Kim D., Heo K., Kim J., Park J., Yun H., Lee C.W., Cho H-S. and others (2024) Structural basis of the catalytic and allosteric mechanism of bacterial acetyltransferase PatZ *PNAS*, doi: 10.1101/2024.11.12.623305 (accepted)

- [3] Moi D., Bernard C., **Steinegger M.**, Nevers Y., Langleib M., Dessimoz, C., (2025) Structural phylogenetics unravels the evolutionary diversification of communication systems in gram-positive bacteria and their viruses, *Nature Structural & Molecular Biology* (accepted)
- [4] Hsiao, M-H., Miao, Y., Liu, Z., Schütze, K., Limjunyawong, N., Chien, D. C-C., Monteiro, W. D., Chu, L-S., Morgenlander, W. and others (2025) Molecular Display of the Animal Meta-Venome for Discovery of Novel Therapeutic Peptides *Molecular & Cellular Proteomics*, doi: 10.1016/j.mcpro.2024.100901
- [5] Karin, E. L., **Steinegger, M.** (2025) Cutting Edge Deep-Learning Based Tools for Metagenomic Research *National Science Review*, doi: 10.1093/nsr/nwaf056
- [6] Fleming, J., Magaña, P., Nair, S., Tsenkov, M., Bertoni, D., Pidruchna, I., Querino Lima Afonso, M., Midlik, A., Paramval, U., Židek, A. and others (2025) AlphaFold Protein Structure Database and 3D-Beacons: New Data and Capabilities *Journal of Molecular Biology*, doi: 10.1016/j.jmb.2025.168967
- [7] Seo, C. W., Yoo, S., Cho, Y., Kim, J. S., **Steinegger, M.**, Lim, Y. W. (2025) FunVIP: Fungal Validation and Identification Pipeline Based on Phylogenetic Analysis *Journal of Microbiology*, 63(4):e2411017, doi: 10.71150/jm.2411017
- [8] Hyunuk Eom, Sukhwan Park, Kye Soo Cho, Jihyeon Lee, Hyunbin Kim, Stephanie Kim, and others (2025) Discovery of Highly Active Kynureninases for Cancer Immunotherapy through Protein Language Model *Nucleic Acids Research*, doi: 10.1101/2024.01.16.575968
- [9] Kim, W. and Mirdita, M. and Levy K. L. Gilchrist., L. M. C., Schweke, H., Söding, J., Levy, Emmanuel D., **Steinegger, M.** (2025) Rapid and Sensitive Protein Complex Alignment with Foldseek-Multimer, *Nature Methods*, doi: 10.1038/s41592-025-02593-7

2024 (6)

- [1] Heinzinger, M., Weissenow, K., Sanchez, JG., Henkel, A., Mirdita, M., **Steinegger, M.**, Rost, Burkhard (2024) ProstT5: Bilingual Language Model for Protein Sequence and Structure *NAR Genomics and Bioinformatics*, doi: 10.1101/2023.07.23.550085
- [2] Kim R. S., Levy K. E., Mirdita M., Chikhi R., **Steinegger, M.** (2024) BFVD-a large repository of predicted viral protein structures, *Nucleic Acids Research* doi: 10.1186/s40168-024-01904-y
- [3] Kim G., Lee S., Levy K. E., Kim H., Moriwaki Y., Ovchinnikov S., **Steinegger, M.**, Mirdita M. (2024) Easy and accurate protein structure prediction using ColabFold *Nature Protocols*, doi: 10.21203/rs.3.pex-2490/v1
- [4] Kim, J., **Steinegger, M.** (2024) Metabuli: sensitive and specific metagenomic classification via joint analysis of amino-acid and DNA, *Nature Methods*, doi: 10.1038/s41592-024-02273-y
- [5] Jochheim, A., Jochheim F.A., Kolodyazhnaya A., Morice E., Koitka F., Mirzaiebadizi A., **Steinegger, M.** and Söding J.. (2024) Strain-Resolved de-Novo Metagenomic Assembly of Viral Genomes and Microbial 16S rRNAs. *Microbiome* doi: 10.1186/s40168-024-01904-y
- [6] Busley, A. V., Gutierrez-Gutierrez, O., Hammer, E., **Steinegger, M.**, and others (2024) LZTR1 polymerization provokes cardiac pathology in recessive Noonan syndrome, *Cell Reports*, doi: 10.1016/j.celrep.2024.114448

2023 (12)

- [1] Lee, S., Kim, G., Karin, EL., Mirdita, M., Park, S., Chikhi, R.; Babaian, A., Kryshtafovych, A., **Steinegger, M.** (2023) Petabase-scale Homology Search for Structure Prediction, *CSH Perspective Biology*, doi: 10.1101/cshperspect.a041465

- [2] Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., and Oleg K., Tunyasuvunakool, K., Laydon, A., Židek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., Birney, E., **Steinegger M.**, Hassabis, D., Velankar S. (2023), AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences *Nucleic Acids Research* doi: 10.1093/nar/gkad1011
- [3] Basu S., Zhao B., Biró B., Faraggi E., Gsponer J., Hu G., Kloczkowski A., Malhis N., Mirdita M., Söding J. and others (2023) DescribePROT in 2023: more, higher-quality and experimental annotations and improved data download options *Nucleic Acids Research*, 52(D1): D426–D433, doi: 10.1093/nar/gkad985
- [4] Varabyou, A., Sommer, M. J., Erdogdu, B., Shinder, I., Minkin, L., Chao, K.-H., Park, S., Heinz, J., Pockrandt, C., Shumate, A., Rincon, N., Puiu, D., **Steinegger, M.**, Salzberg, S. L., and Pertea, M. (2023) CHES3: an improved, comprehensive catalog of human genes and transcripts based on large-scale expression data, phylogenetic analysis, and protein structure, *Genome Biology*, doi: 10.1186/s13059-023-03088-4
- [5] Barrio-Hernandez, Inigo, Yeo, J., Jänes, J., Mirdita M., Gilchrist C. L.M., Wein, T., Varadi, M.; Velankar, S., Beltrao, P., **Steinegger, M.** (2023) Clustering predicted structures at the scale of the known protein universe, *Nature*, doi: 10.1038/s41586-023-06510-w
- [6] Liu D., and **Steinegger M.** (2023), Block aligner: fast and flexible pairwise sequence alignment with SIMD-accelerated adaptive blocks *Bioinformatics*, doi: 10.1093/bioinformatics/btad487
- [7] Jeong, E., Kim, W., Son S., Yang, S., Gwon, D., Hong, J., Cho Y., Jang, C., **Steinegger, M.**, Lim, Y. and Kang, K. (2023) Qualitative metabolomics-based characterization of a phenolic UDP-xylosyltransferase with a broad substrate spectrum from *Lentinus brumalis* *Proceedings of the National Academy of Sciences* doi: 10.1073/pnas.2301007120
- [8] van Kempen, M., Kim, S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist C. L.M., Söding, J. and **Steinegger, M.** (2023) Fast and accurate protein structure search with Foldseek, *Nature Biotechnology*, doi: 10.1101/2022.02.07.479398
- [9] Ruperti, F., Papadopoulos, N., Musser, JM., Mirdita M., **Steinegger, M.** and Arendt D. Cross-phyla protein annotation by structural prediction and alignment *Genome Biology* doi: 10.1186/s13059-023-02942-9
- [10] Kim H., Mirdita M., **Steinegger, M.** (2023) Foldcomp: a library and format for compressing and indexing large protein structure sets, *Bioinformatics*, doi: 10.1093/bioinformatics/btad153
- [11] Bordin N., Dallago C., Heinzinger M., Kim S., Littmann M., Rauer C., Steinegger M., Rost B., Orengo C. (2023) Novel machine learning approaches revolutionize protein knowledge, *Trends in Biochemical Sciences*, doi: 10.1016/j.tibs.2022.11.004
- [12] Bordin N., Sillitoe I., Nallapareddy V. M., Rauer C., Lam D. S., Waman P. V., Sen N., Heinzinger M., Littmann M., Kim S., Velankar S., Steinegger M., Rost B., Orengo C. (2023) AlphaFold2 reveals commonalities and novelties in protein structure space for 21 model organisms, *communications biology* , doi: 10.1038/s42003-023-04488-9

2022 (11)

- [1] Olenyi T., Marquet C., Heinzinger M., Kröger B., Nikolova T., Bernhofer M., Sändig P., Schütze K., Littmann M., Mirdita M., Steinegger M., Dallago C., Rost B. (2022) LambdaPP: Fast and accessible protein-specific phenotype predictions, *Protein Science*, doi: 10.1016/j.tibs.2022.11.001
- [2] Sommer, M., Cha S., Varabyou, A., Rincon M., Park S., Minkin I., Pertea M., **Steinegger, M.**, Salzberg L. S., (2022) Structure-guided isoform identification for the human transcriptome. *elfie*, doi: 10.7554/eLife.82556 (#corresponding)

- [3] Trinquier J., Petti S., Feng S., Söding J., **Steinegger M.**, Ovchinnikov S. (2022) SWAMPNN: End-to-end protein structures alignment *Machine Learning in Structural Biology Workshop, NeurIPS 2022*
- [4] Schütze K., Heinzinger M., **Steinegger M.**, Rost B. (2022) Nearest neighbor search on embeddings rapidly identifies distant protein relations *Frontiers in Bioinformatics*, 2:1033775, doi: 10.3389/fbinf.2022.1033775 :contentReference[oaicite:0]index=0
- [5] Varadi, M., Nair, S., Sillitoe, I., Tauriello, G., and others (2022) 3D-Beacons: decreasing the gap between protein sequences and structures through a federated network of protein structure data resources, *GigaScience*, vol. 11, doi: 10.1093/gigascience/giac211
- [6] Kim, D., Gilchrist C. L.M., Chun J., **Steinegger M.** (2022) UFCG: database of universal fungal core genes and pipeline for genome-wide phylogenetic analysis of fungi *Nucleic Acids Research*, *accepted*, doi: TBA
- [7] Lu J., Rincon N., Wood E D., Breitwieser F., Pockrandt C., Langmead B., Salzberg L S. and **Steinegger M.** (2022), Metagenome analysis using the Kraken software suite *Nature Protocols*, doi: 10.1038/s41596-022-00738-y
- [8] Varadi M., Nair S., Sillitoe I., Tauriello G., Anyango S., Bienert S., Borges C., Deshpande M., Green T., Hassabis D. and others (2022) 3D-Beacons: decreasing the gap between protein sequences and structures through a federated network of protein structure data resources *GigaScience*, doi: 10.1093/gigascience/giac118
- [9] Mirdita M., Schütze K., Moriwaki Y., Heo L., Ovchinnikov S. and **Steinegger M.** (2022), ColabFold: Making protein folding accessible to all *Nature Methods*, doi: 10.1038/s41592-022-01488-1
- [10] Choi, Hyun-Kyu, Hyunook Kang, Chanwoo Lee, Hyun Gyu Kim, Ben P. Phillips, Soohyung Park, Charlotte Tumescheit, and others (2022). Evolutionary Balance between Foldability and Functionality of a Glucose Transporter *Nature Chemical Biology*, doi:10.1038/s41589-022-01002-w.
- [11] Vanni, C., Schechter, M., Silvia G., Barberán, A., Buttigieg, P., Casamayor, E., Delmont, T., Duarte, C., Eren, A. and Finn, R. and others (2022), Unifying the global coding sequence space enables the study of genes with unknown function across biomes *elife*, doi: 10.7554/eLife.67667

2021 (10)

- [1] Seok, C. and Baek, M. and *Steinegger, M.* and Park, H. and Lee, G. and Won, J. (2021) Accurate protein structure prediction: what comes next? *Biodesign* doi: 10.34184/kssb.2021.9.3.47
- [2] Pockrandt, C., **Steinegger M.**, Salzberg L S. (2022), PhyloCSF++: A fast and user-friendly implementation of PhyloCSF with annotation tools. *Bioinformatics*, doi: 10.1093/bioinformatics/btab756
- [3] Jumper, J., Evans, R., Pritzel, A., Green, T. and others (2021), Applying and improving AlphaFold at CASP14 *Proteins: Structure, Function, and Bioinformatics*, doi: 10.1002/prot.26257
- [4] Jumper J., Evans R., Pritzel A., Green T. and others (2021), Highly accurate protein structure prediction with AlphaFold. *Nature*, doi: 10.1038/s41586-021-03819-2
- [5] Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., **Steinegger, M.** and others (2021), ProtTrans: Towards Cracking the Language of Life's Code Through Self-Supervised Deep Learning and High Performance Computing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2021.3095381
- [6] Aevansson, A., Kaczorowska, A., Adalsteinsson, A., Ahlqvist, J., Al-Karadaghi, S., and others (2021), Going to extremes – a metagenomic journey into the dark matter of life *FEMS Microbiology Letters* 10.1093/femsle/fnab067
- [7] Michael Bernhofer, Christian Dallago, Tim Karl, and others (2021), PredictProtein-Predicting

Protein Structure and Function for 29 Years, *Nucleic Acids Research*, doi: 10.1093/nar/gkab354

- [8] Mirdita, M. and **Steingger, M.**, and Breitwieser, F. and Soeding, J. and Karin, E. L. (2021), Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics*, doi: 10.1101/2020.11.27.401018
- [9] Credle, J. J., Robinson, M., Gunn, J., Monaco, D., Sie, B., Tchir, A. L., Hardick, J., Zheng, X., Shaw-Saliba, K., and Rothman, Richard and others (2021), Highly multiplexed oligonucleotide probe-ligation testing enables efficient extraction-free SARS-CoV-2 detection and viral genotyping. *Modern Pathology*, doi: 10.1038/s41379-020-00730-5
- [10] Zhao, Bi, Katuwawala, Akila, Oldfield, Christopher J, Dunker, A Keith and others (2021), DescribePROT: database of amino acid-level protein structure and function predictions. *Nucleic Acids Research*, doi: 10.1093/nar/gkaa931

2020 (3)

- [1] Gabler F., Nam S., Till S., Mirdita M., **Steinegger M.**, Söding J., Lupas A, Alva V., (2020), Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics*, doi: 10.1002/cpbi.108
- [2] Park S., **Steinegger, M.**, Cho H. and Chun J. (2020) Metagenomic Association Analysis of Gut Symbiont *Limosilactobacillus reuteri* Without Host-Specific Genome Isolation. *Frontiers in Microbiology*, doi: 10.3389/fmicb.2020.585622
- [3] **Steinegger, M.**, Salzberg L S. (2020) Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biology*, doi: 10.1186/s13059-020-02023-1

2019 (3)

- [1] **Steinegger, M.**, Markus Meier, Milot Mirdita, Harald Vöhringer, Stephan J. Haunsberger, and Söding, J. (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*, doi: 10.1186/s12859-019-3019-7
- [2] **Steinegger, M.**, Milot Mirdita, and Söding, J. (2019) Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. *Nature Methods*, **16**, 603–606, doi: 10.1038/s41592-019-0437-4
- [3] Milot Mirdita, **Steinegger, M.** and Söding, J. (2019) MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics*. doi: 10.1093/bioinformatics/bty1057

2018 (3)

- [1] **Steinegger, M.**, and Söding, J. (2018) Clustering huge protein sequence sets in linear time. *Nature Communications* doi: 10.1038/s41467-018-04964-5
- [2] Forslund K., Pereira C., Capella-Gutierrez S. and others (2018) Gearing up to handle the mosaic nature of life in the quest for orthologs *Bioinformatics*, bf 34, i323–i329, doi: 10.1093/bioinformatics/btx542
- [3] Mahlich Y., **Steinegger, M.**, Rost, B. and Bromberg Y. (2018) HFSP: High speed homology-driven function annotation of proteins. *Bioinformatics*, bf 34, i304–i312, doi: 10.1093/bioinformatics/bty262

2017 (2)

- [1] **Steinegger, M.**, and Söding, J. (2017) MMseqs2: Sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, **35**, 1026–1028, doi: 10.1038/nbt.3988
- [2] Mirdita, M.[#], von den Driesch[#], L., Galiez, G., Martin, M., Söding, J.^{*}, and **Steinegger, M.**^{*} (2017) Uniclust databases of clustered and deeply annotated protein sequences and alignments.

2016 (1)

- [1] Hauser M.[#], **Steinegger, M.[#]**, and Söding, J. (2016) MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics*, **32**, 1323–1330. doi: 10.1093/bioinformatics/btw006. (#Equal contributions.)

2014 (1)

- [1] Kajan L., Yachdav G., Vicedo E., **Steinegger M.**, Mirdita M., Angermüller C., Böhm A., Domke S., Ertl J., Mertes C., Reisinger E., Staniewski C., B. Rost (2014) Cloud prediction of protein structure and function with PredictProtein for Debian. *BioMed research international*, doi: 10.1155/2013/398968

Non peer-reviewed articles

- [1] **Steinegger M.** and Goiss, H. (2011) Introducing a Model-based Automated Test Script Generator. *Testing Experience*, 70–76

Co-authors over the past 5 years (bold: \geq two papers), Total 508 co-authors

J. Abrahamson, S. G. Acinas, T. B. Adalsteinsson, **J. Adler**, A. Aevansson, M. Q. L. Afonso, J. Ahlqvist, M. R. Ahmadian, S. Al-Karadaghi, J. Allison, J. Altenbuchner, J. Altmüller, A. Anandkumar, C. Angerer, S. Anyango, **D. Arendt**, G. Armstrong, H. Arsin, D. Ascher, H. Ashkenazy, A. Babaian, **T. Back**, M. Baek, M. Baker, **A. J. Ballard**, A. Baltzis, A. Barberán, M. Barot, I. Barrio-Hernandez, S. Basu, **R. Bates**, P. Beltrao, N. Ben-Tal, **T. Berghammer**, C. Bernard, **M. Bernhofer**, **D. Bertoni**, D. Bhowmik, S. Bienert, P. Binder, E. Birney, B. Biró, S. Bodenstern, R. Bonneau, **N. Bordin**, C. Borges, G. Borrel, R. Bouckaert, K. Boyd, A. Brace, D. Brandt, F. Breitwieser, **F. P. Breitwieser**, **A. Bridgland**, Y. Bromberg, D. H. Brookes, A. Busia, A. V. Busley, P. L. Buttigieg, L. Böhmer, Z. Cao, E. O. Casamayor, S. Cha, A. Chacon, X. Chang, K.-H. Chao, D. C.-C. Chien, **R. Chikhi**, **H.-S. Cho**, K. Cho, K. S. Cho, **Y. Cho**, H.-J. Choi, H.-K. Choi, H.-Y. Chou, L.-S. Chu, S. Chu, J. Chun, M. Cichowicz-Cieślak, I. C. Cirstea, **E. Clancy**, N.-C. community, S. A. K. Cornish, A. D. S. Costa, J. Courtin, **A. Cowie**, J. J. Credle, L. Cyganek, S. Dabrowski, H. Dahle, **C. Dallago**, S. Darabi, E. Dawson, **T. O. Delmont**, S. Deorowicz, M. Deshpande, C. Dessimoz, **K. Didi**, S. Djefane, X. Dong, S. Dorawa, J. Douglas, C. M. Duarte, A. K. Dunker, J. Dusaucy, L. Edwards, B. Elberling, A. Elnaggar, M. Elvert, F. Enault, M. Engler, H. Eom, B. Erdogdu, **A. M. Eren**, S. H. Eshleman, J. Espinosa-Carrasco, **R. Evans**, **E. Faraggi**, A.-E. Fedøy, T. Feher, S. Feng, **A. Fernandez-Guerra**, A. Fernández-Guerra, **M. Figurnov**, R. D. Finn, J. Fleming, E. W. Floden, S. Freitag-Pohl, H. O. Fridjonsson, C. Fu, M. Fullmer, **C. Galiez**, E. Garriga, T. Gaudet, P. Gawron, M. Geiger, T. Gibbs, C. Gilchrist, **C. L. Gilchrist**, C. L. M. Gilchrist, M. Gill, F. O. Gloeckner, E. Glomsaker, F. O. Glöckner, T. Goldberg, J. J. Gray, M. Greaves, **T. Green**, L. Gremer, S. Gribaldo, **J. Gsponer**, W. Gu, E. E. Gudmundsdóttir, H. Gudmundsson, A. Gudyś, E. S. Gundesø, J. Gunn, Ó. Gutiérrez-Gutiérrez, M. Guérin, D. Gwon, T. Hamamsy, E. Hammer, C. Han, Y. Han, Y.-H. Han, **Y.-Y. Han**, K. Hansen, J. Hardick, D. Hariharan, G. Hasenfuss, **D. Hassabis**, A. Hatos, Y. He, T. Hegedus, J. Heinz, **M. Heinzinger**, M. L. Hekkelman, A. Helleux, C. Henke, A. Henkel, R. A. Henriksen, R. J. Henriksen, **K. Heo**, L. Heo, K.-U. Hinrichs, K. Hippe, S. Hjörleifsdóttir, H. Hong,

J. Hong, O. G. Hreggvidsson, M.-H. Hsiao, D. J. Hsu, P. Hsu, G. Hu, C. Hundt, M. Håkansson, R. Illango, W. Im, **R. Jain**, C.-Y. Jang, S.-E. Jang, Y. Jarosz, A. Jasilionis, S. Jayaraman, E. Jeong, H. Ji, S. Jiang, **A. Jochheim**, **F. A. Jochheim**, J. S. John, P. S. John, D. T. Jones, L. Jones, R. Joosten, **J. Jumper**, A. Jurczak-Kurek, J. Jänes, B. L. Jónsdóttir, I. Jónsdóttir, A.-K. Kaczorowska, T. Kaczorowski, L. Kajan, J. Kalinowski, F. Kallenborn, K. Kamata, S. M. Kandathil, H. Kang, K. B. Kang, **E. L. Karin**, T. Karl, A. Katuwawala, G. Kaushik, **K. Kavukcuoglu**, M. V. Kempen, **D. Kim**, **G. Kim**, H. G. Kim, **H. Kim**, **J. Kim**, J. S. Kim, **J.-S. Kim**, R. S. Kim, S. A. Kim, **S. Kim**, S. S. Kim, W. Kim, K. H. Kjær, M. Kleinsorge, **A. Kloczkowski**, S. A. A. Kohl, S. A. Kohl, **P. Kohli**, F. Koitka, **A. Kolodyazhnaya**, T. S. Korneliussen, M. Korshunova, S. Kothen-Hill, R. Kottmann, O. Kovalevskiy, P. L. Kozlowski, L. Kraft, **M. Krupovic**, A. Kryshtafovych, B. Kröger, **L. Kurgan**, K. Kwiatkowska-Semrau, S. D. Lam, O. Lanes, A. Lange, J. Lange, M. Langleib, **B. Langmead**, H. B. Larman, N. K. Larsen, A. M. Lau, **A. Laydon**, J. Lebrat, C. Lee, **C. W. Lee**, **G. Lee**, G. R. Lee, H. Lee, J. Lee, M. S. Lee, S. Lee, Y. Lee, E. D. Levy, Z. Li, **Y. W. Lim**, N. Limjunyawong, D. Lin, J. Linares-Pastén, **M. Littmann**, D. Liu, M. Liu, Y. Liu, Z. Liu, M. Livne, A. S. Lorentsen, **J. Lu**, D. Lundin, T. Lutterman, D. Ma, P. Magana, P. Magaña, A. Mahajan, **N. Malhis**, A. J. Malik, S. Malina, L. Mansouri, F. Marbach, C. Marquet, N. J. Matzke, Z. McClure, **C. Meyer**, Y. Miao, A. Midlik, E. A. Miller, I. Minkin, L. Minkin, J. Mir-Pedrol, **M. Mirdita**, A. Mirzaiebadizi, A. Mitchell, J. Mitchell, D. Moi, D. Molodenskiy, D. Monaco, W. D. Monteiro, A. Moradzadeh, W. Morgenlander, E. Morice, É. Morice, **Y. Moriwaki**, J. T. Morton, O. Mosafi, H. Mostafa, **J. M. Musser**, L. Naef, **S. Nair**, V. Nallapareddy, Y. Nashed, Y. Nevers, **S. Nikolov**, T. Nikolova, C. Notredame, S. O'Donoghue, Z. Obradovic, C. J. Oldfield, **T. Olenyi**, OPMC, **C. Orengo**, **S. Ovchinnikov**, **M. Pacholska**, S. Paliwal, **N. Papadopoulos**, C. Pape, **U. Paramval**, H. Park, **J. B. Park**, **J. Park**, S. Park, M. W. Pedersen, V. K. Pedersen, A. Pekosz, Y. Peng, R. Perez-Laso, **M. Perte**, **S. Petersen**, S. Petti, B. P. Phillips, **I. Pidruchna**, D. Piovesan, **C. Pockrandt**, A. M. Poole, **A. Potapenko**, N. Prasad, **A. Pritzel**, C. Puente-Lelievre, D. Puiu, J. Qiu, S. Rabhi, M. Radhakrishnan, A. Ramanathan,

F. Ramezanghorbani, **C. Rauer**, L. Rauschning, C. Regep, G. Rehawi, D. Reidenbach, **D. Reiman**, G. Renaud, C. Ricketts, **N. Rincon**, M. L. Robinson, S. H. Roh, **S.-H. Roh**, B. C. Roland, B. Romera-Paredes, B. Romera-Paredes, **O. Ronneberger**, **B. Rost**, R. E. Rothman, **F. Ruperti**, A. H. Ruter, P. W. Sackett, E. Salladini, **S. L. Salzberg**, J. G. Sanchez, K. K. Sand, C. Sander, L. Santus, V. Satagopam, A. Schafferhans, **M. S. Schechter**, A. Schlessinger, B. Schmidt, R. Schneider, H. Schroeder, T. Schwede, H. Schweke, **K. Schütze**, N. Sen, **A. W. Senior**, C. W. Seo, **C. Seok**, **Y.-J. Seok**, K. Shah, V. Shah, J. Shan, K. Shaw-Saliba, T. Shimko, I. Shinder, A. Shumate, B. Sie, M. Sikora, **I. Sillitoe**, **D. Silver**, S. Sinai, **H. Sirelkhatim**, K. Siren, J. Soeding, H.-G. Sommer, M. J. Sommer, S. Son, W. J. Song, S.

Srinivasan, Ste, J. Su, E. Suhajda, D. Svergun, P. Sánchez, P. Sändig, **J. Söding**, G. Tauriello, A. Tchir, L. Tenorio-Ku, H. Tomlinson, P. D. Tommaso, S. Tosatto, C. Trefois, J. Trinquier, I. Trujnara, **M. Tsenkov**, **C. Tumescheit**, **K. Tunyasuvunakool**, **C. Vanni**, **A. Varabyou**, **M. Varadi**, **S. Velankar**, I. Veseli, A. Vignoli, **O. Vinyals**, G. Vriend, V. P. Waman, D. Wang, K. Wang, Y. Wang, A. M. Waterhouse, T. Wein, **K. Weissenow**, D. Willbold, E. Willerslev, B. Wollnik, J. Won, **D. E. Wood**, L. Wörmer, D. Xu, G. Yachdav, J. Yang, S. Yang, **J. Yeo**, S. Yoo, T.-Y. Yoon, F. Yuan, **H. Yun**, J. Zhang, **B. Zhao**, L. Zhao, X. Zheng, X. Zhou, Y. Zhou, H. Zhu, **M. Zielinski**, W.-H. Zimmermann, Á. Ú. Átlason, M. Žure, **A. Židek**.

Web servers and public resources

Tool	Type	Description
COLABFOLD	M, W	ColabFold makes fast and highly accurate protein folding accessible to all (AlphaFold2/ESMfold/BioEMU/RoseTTAFold2) >57M web uses, >3.3k installs per day . https://colabfold.com
FOLDSEEK	M, W	Ultra-fast, sensitive protein structure search using our 3D structural alphabet (3Di). >1M web uses, >300k times . https://foldseek.com
MMSEQS2	M	Fast, sensitive sequence search and clustering, 10,000-fold faster than BLAST, comparable sensitivity to PSI-BLAST. >1.2M installs . https://mmseqs.com
FOLDSEEK-MULTIMER	M, W	Fast method to compare protein structure complexes. https://search.foldseek.com/multimer
AFDB-CLUSTER	M, R, W	Clusters ~214M AlphaFold DB structures into ~2.3M structural groups, revealing novel families and organizing global protein structure space. https://afdb-cluster.foldseek.com
FOLDMASON	M, W	Rapid multiple protein structure alignment leveraging Foldseek's 3Di alphabet. https://foldmason.foldseek.com
PLASS	M	Protein-level assembler reconstructing proteins directly from short reads; recovers up to 10× more proteins from metagenomes compared to nucleotide-based assembly. https://plass.mmseqs.com
BFVD	R, W	Comprehensive repository with >350,000 viral protein structure predictions (via ColabFold), facilitating large-scale structural virology analyses. https://bfvd.foldseek.com
METACLUST	R	Database of 424M clustered metagenomic proteins (50% and 95% identity); enhances diversity for sensitive profile-based searches and structure prediction. https://metaclust.mmseqs.com
BFD	R	Database of ~300M clusters (30% identity) from ~2.5B sequences; foundational dataset for training AlphaFold2 and improving protein annotation. https://bfd.mmseqs.com
AFESM-CLUSTER	R, W	Integration of AlphaFold DB and ESMAtlas (~821M total models); identified 11,000 novel domain combinations and biome-specific structural groups. https://afesm.foldseek.com
UNICLUST	R	Hierarchical clustering of UniProt sequences (30/50/90% identity); provides consistent annotations and precomputed MSAs for sensitive homology detection. https://uniclust.mmseqs.com

Legend: M: Method, R: Resource, W: Webserver

TEACHING

Academic Service

2025	Chair RECOMB in Seoul
2024	Program committee RECOMB
2023	Program committee RECOMB-Seq
2021	Organizer "Symposium on Bioinformatics for Metagenomic analysis"

Lectures, seminars, and lab classes

2025	Introduction to bioinformatics (undergraduate course). Seoul National University
2025	Integrative biology (graduate course). Seoul National University
2024	Introduction to bioinformatics (undergraduate course). Seoul National University
2023	Advanced topics in bioinformatics (graduate course). Seoul National University
2023	Integrative biology (graduate course). Seoul National University
2023	Introduction to bioinformatics (undergraduate course). Seoul National University
2022	Integrative biology (graduate course). Seoul National University
2022	Introduction to bioinformatics (undergraduate course). Seoul National University
2021	Advanced topics in bioinformatics (graduate course). Seoul National University
2021	Integrative biology (graduate course). Seoul National University
2021	Introduction to bioinformatics (undergraduate course). Seoul National University
2020	Deep dive into metagenomic data using metagenome-atlas and MMseqs2 at ECCB 2020 in Spain.
2018	Modern and scalable tools for efficient analysis of very large metagenomic at ECCB18 in Greece.
2012	Bioinformatics tutorial for bachelor students: Development of tutorial material and teaching at the Ludwig Maximilian University.
2009 - 2011	Database faculty at Accenture. Regularly held Oracle database seminars and reworked the course material. Full-time 2 day seminars for Accenture consultants
2010 - 2011	Security training at Accenture. Helped create a security curriculum and held seminars.
2010	Java architecture seminars at Accenture, Full-time 5 days workshop for Java consultants

(Co-)Supervised theses

02/2025: Dongwook Kim, Ph.D. *Scalable methods for species comparison and phylogeny using protein sequences and structures*

02/2025: Jihyeon Kim, Ph.D. *Integrative Bioinformatics Approaches for Bacterial Species Delineation Using Genomic and Proteomic Data*

02/2025: Yeong Ouk, Ph.D. *Development and Application of a System for Large-Scale Microbiome Data Analysis*

09/2024: Sein Park, Ph.D. *Computational Frameworks for Bacterial Genomic Analysis at Multiple Taxonomic Levels*

09/2023: Seongin Na, Ph.D., Bioinformatics, Seoul National University (co-advisor)
Discovery of Core Genes in Prokaryotes and Phylogenomics-based Application to Taxonomy

09/2023: Jaebeom Kim, B.Sc., Bioinformatics, Seoul National University
Sensitive and specific metagenomic classification by joint analysis of DNA and amino acid sequences

09/2023: Sewon Lee, B.Sc., Biology, Seoul National University (best thesis award)

Improving protein structure prediction using petascale sequence search

Michael Heinziger, Ph.D., Technical University of Munich (co-advisor)

How to speak protein? Representation learning for protein prediction

09/2023: SooHyun Kim, B.Sc., Biology, Seoul National University

New Methods for Ribozyme Discovery

05/2022: HyeonSeok Oh, Ph.D., Bioinformatics, Seoul National University (co-advisor)

Understanding human gut microbiota and its application for human health using computational methods

09/2021: Minghang Lee, B.Sc., Biology, Seoul National University (best thesis award)

Petasearch: Fast, approximate comparison of huge sequence datasets

09/2021: Sukhwan Park, B.Sc., Biology, Seoul National University

Methodology of building Empirical Codon Substitution Model using XRate

09/2021: Doyoung Kim, B.Sc., Biology, Seoul National University

Fast homology detection neural network based profile prediction

09/2016: Milot Mirdita, M.Sc., Computer Science, LMU Munich

Uniclust - clustered and deeply annotated protein sequence databases

10/2014: Lars von der Driesch, M.Sc., Bioinformatics, LMU Munich / TU Munich

Deep clustering and annotation of the Uniprot database

05/2014: Stefan Haunsberger, B.Sc., Bioinformatics, Hochschule Weihenstephan-Triesdorf

Fast AVX-based Forward-Backward and Maximum Accuracy algorithms for pairwise alignment of profile hidden Markov models