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 03/2020	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

Awards, Fellowships and Achievements

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 price (∼1000$) at the Disrupt TechCrunch Hackathon (1200+ attendees)
2012	Excellence initiative research grant, Ludwig Maximilian University
2012	AMD Research grant (∼700$ one graphic card)
2012	NVIDA research grant (∼3000$ two graphic cards)
2011	Amazon research grant (10.000$ Amazon Web Services credits)
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

Technical Strengths

Programming	C++, C, Java, Shell scripting, Python
Cloud	Amazon Web Services
Testing	Unit, performance, functional, penetration test
Databases	BI, SQL, PL/SQL, Oracle 11g, MySQL, db4o

Languages

German	Native
English	Fluent
Korean	Beginner

Public source code

ColabFold	https://github.com/sokrypton/ColabFold
Foldseek	https://github.com/steineggerlab/foldseek
Conterminator	https://github.com/martin-steinegger/conterminator
Plass	github.com/soedinglab/plass
Linclust	github.com/soedinglab/mmseqs2
MMseqs2	github.com/soedinglab/mmseqs2
MMseqs	github.com/soedinglab/mmseqs
HH-suite	github.com/soedinglab/hh-suite

TALKS, POSTERS, AND PUBLICATIONS

Talks

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 ear of NGS
06/2014	ISCB NGS14, Spain, MMseqs suite for fast and sensitive batch searching
06/2014	Hadoop User Group, Germany, In-Silico mutagenises 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 **20** 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 programing and improved secondary structure scoring

Research Grants

PI is Martin Steinegger unless otherwise indicated

2023 - 2025	Seoul National University, "Accurate genomic annotation through a homology aware AI model" 200,000,000 KRW
2022 - 2025	Samsung, "Rapid and precise diagnosis of infectious diseases using metagenomics" 90,000,000 KRW
2020 - 2021	Seoul National University, the New Faculty Startup Fund, "Capture probe design in the era of next generation sequencing" 40,000,000 KRW
2020 - 2023	National Research Foundation of Korea, "Discovery of novel genomes through protein-guided assembly" NRF-2019R1A6A1A10073437, 150,000,000 KRW per year
2020 - 2024	Seoul National University, the Creative-Pioneering Researchers Program, "Petasearch: surveilling pathogens on a global scale", 320,000,000 KRW total
2020 - 2024	National Research Foundation of Korea, "In silico protein design by artificial intelligence and physical chemistry", NRF-2020M3A9G7103933, 450,000,000 KRW (PI : Chaok Seok)
2021 - 2026	National Research Foundation of Korea, "Folding the protein universe (FoldU): metagenomics scale protein structure prediction using maching learning", NRF-2021R1C1C102065, 743,450,000 KRW.
2021 - 2025	National Research Foundation of Korea, "Development of Cryo-EM/ET Technology for 3D Bio-imaging at Molecular resolution", NRF-2021M3A9I4021220, 500,000,000 KRW (PI : Roh, Soung-hun)

Features of work or interviews

Articles covering me or the work of our lab.

[1] Ewen Callaway (2023) 'A Pandora's box': map of protein-structure families delights scientists, *Nature* , doi: 10.1038/d41586-023-02892-z

[2] Arunima Singh (2023) Speedier protein structure search, *Nature Methods*, doi: 10.1038/s41592-023-01953-5

[3] Matthew Hutson (2023) Foldseek gives AlphaFold protein database a rapid search tool, *Nature*, doi: 10.1038/d41586-022-02083-2

[4] Ewen Callaway (2022) 'The entire protein universe': AI predicts shape of nearly every known protein, *Nature*, doi: 10.1038/d41586-022-02083-2

[5] Ewen Callaway (2022) What's next for AlphaFold and the AI protein-folding revolution, *Nature*, doi: 10.1038/d41586-022-00997-5

[6] Henrik Müller (2022) Computerbasierte Proteinstruktur Vorhersage *Laborjournal* (in German)

[7] Michael Eisenstein (2021) Artificial intelligence powers protein-folding predictions, *Nature*, doi: 10.1038/d41586-021-03499-y

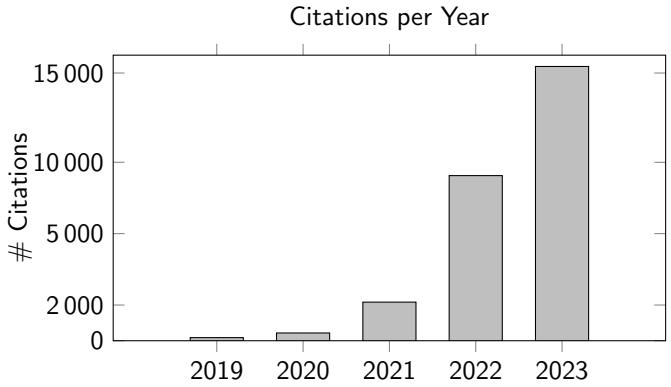
[8] Henrik Müller (2021) Interview mit Martin Steinegger über AlphaFold2 und ColabFold *Laborjournal* (in German)

[9] Nikki Forrester, (2021) How new principal investigators tackled a tumultuous year, *Nature*, doi: 10.1038/d41586-021-01311-5

[10] Lin Tang, (2020) Contamination in sequence databases, *Nature Methods*, doi: 10.1038/s41592-020-0895-8

Google Scholar

Google Scholar Profile. Data collected on November 6, 2023.



	All	Since 2018
Citations	27 875	27 790
h-index	25	25
i10-index	34	34

Publications

The most important articles are highlighted in red.

[1] Heinzinger, M., Weissenow, K., Sanchez, JG.;,Henkel, D., **Steinegger, M.**, Rost, Burkhard (2023) ProstT5: Bilingual Language Model for Protein Sequence and Structure *bioRxiv*, doi: 10.1101/2023.07.23.550085

[2] 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, *bioRxiv*, doi: 10.1101/2023.07.10.548308

[3] Kim, J., **Steinegger, M.** (2023) Metabuli: sensitive and specific metagenomic classification via joint analysis of amino-acid and DNA, *bioRxiv*, doi: 10.1101/2023.05.31.543018

[4] Fernandez-Guerra, A. Borrel, G., Delmont, OT., and others (2023) A 2-million-year-old microbial and viral communities from the Kap København Formation in North Greenland *bioRxiv*, doi: 10.1101/2023.06.10.544454

[5] 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

[6] Busley, A. V., Gutierrez-Gutierrez, O., Hammer, E., **Steinegger, M.**, and others (2023) LZTR1 polymerization provokes cardiac pathology in recessive Noonan syndrome, *bioRxiv*, doi: 10.1101/2023.01.10.523203

[7] 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

[1] 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., Žídek, 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

[2] 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) CHESS 3: 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

[3] Barrio-Hernandez, Inigo, Yeo, J., Jänes, J., Mirdita M., Gilchrist C. L.M., Wein, T., Varadi, M.; Velankar, S., Beltrao, P., **Steinegger, M.** Clustering predicted structures at the scale of the known protein universe, *Nature*, doi: 10.1038/s41586-023-06510-w

[4] 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

[5] 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

[6] 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

[7] 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

[8] 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

[9] 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

[10] 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

[11] Bordin N., Dallago C., Heinzinger M., Kim S., Littmann M., Rauer C., Steinegger M., Rost B., Orengo C. (2022) Novel machine learning approaches revolutionize protein knowledge, *Trends in Biochemical Sciences*, doi: 10.1016/j.tibs.2022.11.004

[12] Sommer, M.,, Cha S., Varabyou, A., Rincon M., Park S., Minkin I., Pertea M., **Steinegger, M.**[#],, Salzberg L. S.[#], (2022) Highly accurate isoform identification for the human transcriptome, *elfie*, doi: 10.7554/eLife.82556 ([#]correspoding)

[13] 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

[14] 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

[15] 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

[16] 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

[17] 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.

[18] 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

[19] 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

[20] Pockrandt, C., **Steinegger M.**, Salzberg L S. (2021), PhyloCSF++: A fast and user-friendly implementation of PhyloCSF with annotation tools. *Bioinformatics*, doi: 10.1093/bioinformatics/btab756

[21] 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

[22] 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

[23] 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 Lifes Code Through Self-Supervised Deep Learning and High Performance Computing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2021.3095381

[24] Aevarsson, 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

[25] 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

[26] 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

[27] 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

[28] Zhao, Bi and Katuwawala, Akila and Oldfield, Christopher J and Dunker, A Keith and Faraggi, Eshel and Gsponer, Jörg and Kloczkowski, Andrzej and Malhis, Nawar and Mirdita, Milot and Obradovic, Zoran and others (2021), DescribePROT: database of amino acid-level protein structure and function predictions. *Nucleic Acids Research*, doi: 10.1093/nar/gkaa931

[29] 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

[30] 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

[31] **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

[32] **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

[33] **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

[34] 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

[35] **Steinegger, M.**, and Söding, J. (2018) Clustering huge protein sequence sets in linear time. *Nature Communications* doi: 10.1038/s41467-018-04964-5

[36] 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

[37] 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

[38] **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

[39] 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. *Nucleic Acids Research*, **45**, D170–D176, doi: 10.1093/nar/gkw1081. . ([#]Equal contributions.) ([*]Corresponding authors.)

[40] 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.)

[41] 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

TEACHING

Academic Service

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

Lectures, seminars, and lab classes

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

03/2017 - 09/2023: Seongin Na, Ph.D., Bioinformatics, Seoul National University (co-advisor)
Discovery of Core Genes in Prokaryotes and Phylogenomics-based Application to Taxonomy

03/2021 - 09/2023: Jaebeom Kim, B.Sc., Bioinformatics, Seoul National University
Sensitive and specific metagenomic classification by joint analysis of DNA and amino acid sequences

03/2023 - 09/2023: Sewon Lee, B.Sc., Biology, Seoul National University (best thesis award)
Improving protein structure prediction using petascale sequence search

07/2023: Michael Heinziger, Ph.D., Technical University of Munich (co-advisor)
How to speak protein? Representation learning for protein prediction

03/2023 - 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

03/2021 - 09/2021: Minghang Lee, B.Sc., Biology, Seoul National University (best thesis award)
Petasearch: Fast, approximate comparison of huge sequence datasets

03/2021 - 09/2021: Sukhwan Park, B.Sc., Biology, Seoul National University
Methodology of building Empirical Codon Substitution Model using XRate

03/2021 - 09/2021: Doyoung Kim, B.Sc., Biology, Seoul National University
Fast homology detection neural network based profile prediction

03/2016 - 09/2016: Milot Mirdita, M.Sc., Computer Science, LMU Munich
Uniclust - clustered and deeply annotated protein sequence databases

04/2014 - 10/2014: Lars von der Driesch, M.Sc., Bioinformatics, LMU Munich / TU Munich
Deep clustering and annotation of the Uniprot database

11/2013 - 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