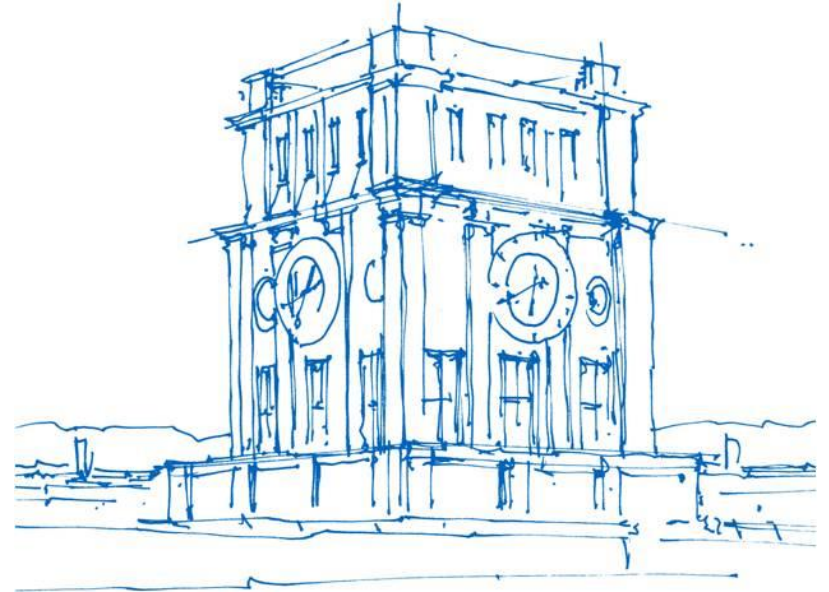# Computational Mass Spectrometry @ TUM/Freising

Prof. Dr. Mathias Wilhelm

TUM School of Life Sciences

Technische Universität München


Uhrenturm der TUM

About me and the group

Mathias Wilhelm

- Studied "Bioinformatics and Genome Research" and "Natural Sciences" @ Bielefeld
- Research assistant @ Boston Childrens Hospital/Havard Medical School
- PhD+PostDoc in Computational Mass Spectrometry @ TUM (Prof. Kuster)
- TT Professor in Computational Mass Spectrometry @ TUM/Freising since 2021
- Co-founder of OmicScouts and MSAID, two TUM "biotech" spin-offs
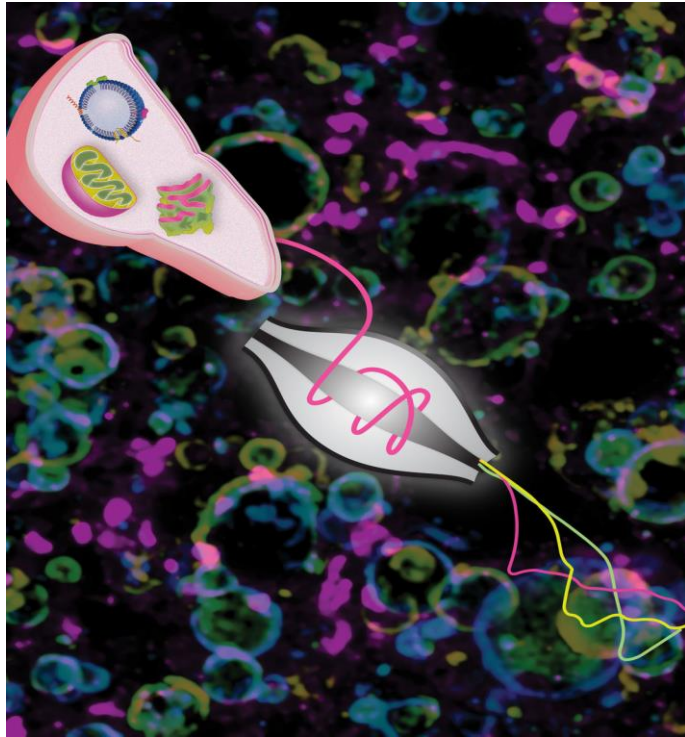
Computational Mass Spectrometry (MS)
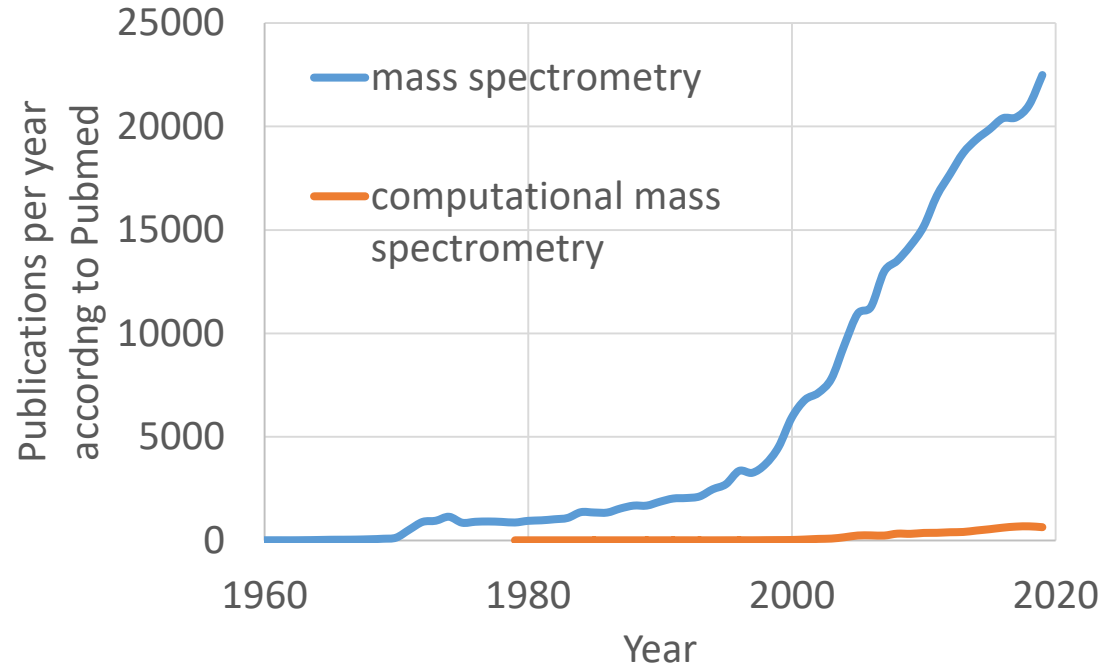https://www1.ls.tum.de/en/compms/home/
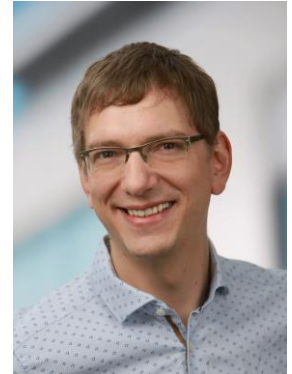
- Develop tools to evaluate, analyze, and interpret (MS) data in an integrative manner (multi omics).
- ~10 PhDs, 2 PostDocs
- ~10 parallel student projects (BSc, MSc, FP, …); >30 completed

Why computational mass spectrometry?

https://www.biochem.mpg.de/en/20181022-krahmer-mann

➤ Countless opportunities to develop computational tools to propel (protein-based) life science research

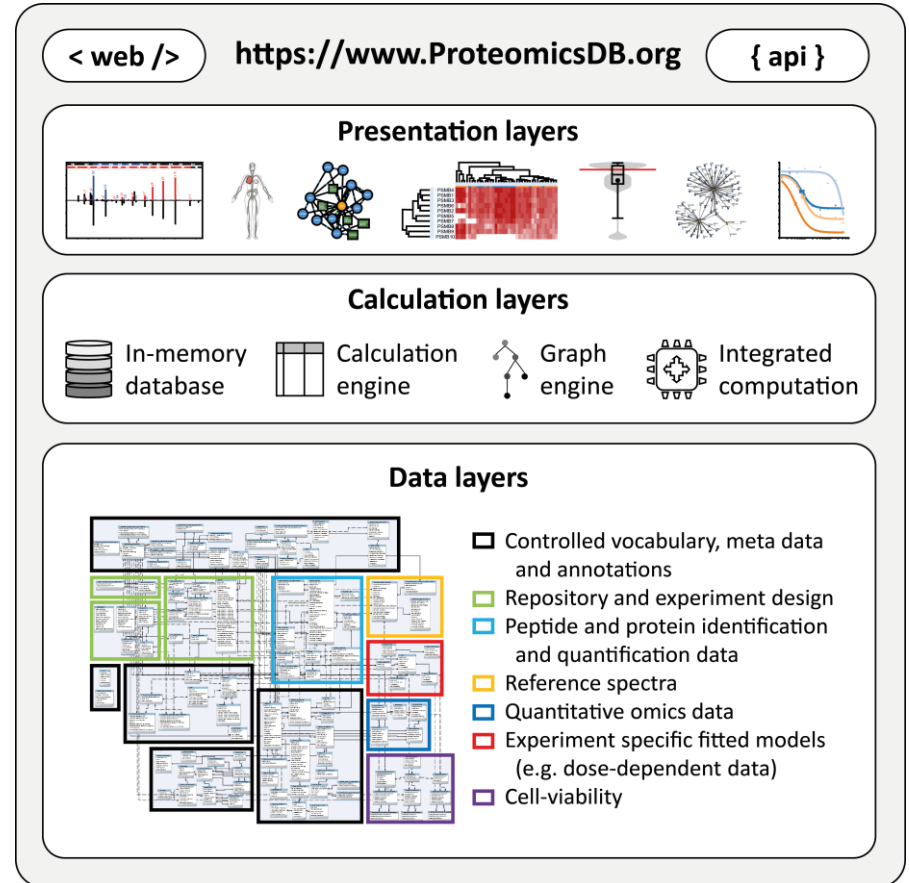# ProteomicsDB: A multi-omics database for life science

Hardware

- 2x 6 TB RAM in-memory database
- ~300 TB storage
- Multiple CPU and GPU servers

➤ **Database, backend, and frontend dev**

Quantitative data for

- Human, Arabidopsis, mouse and rice
- Soon to be extended to ~100 most important crops and ~300 bacteria
- Drug sensitivity measurements, …

➤ **Multi-omics in life science research**

Prosit: An artificial neural network for peptide properties
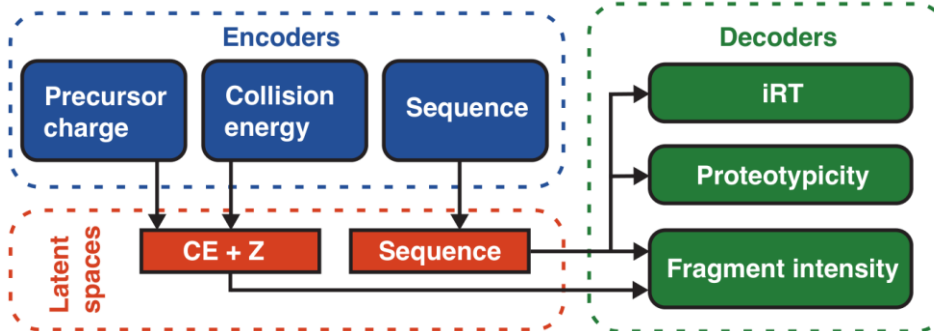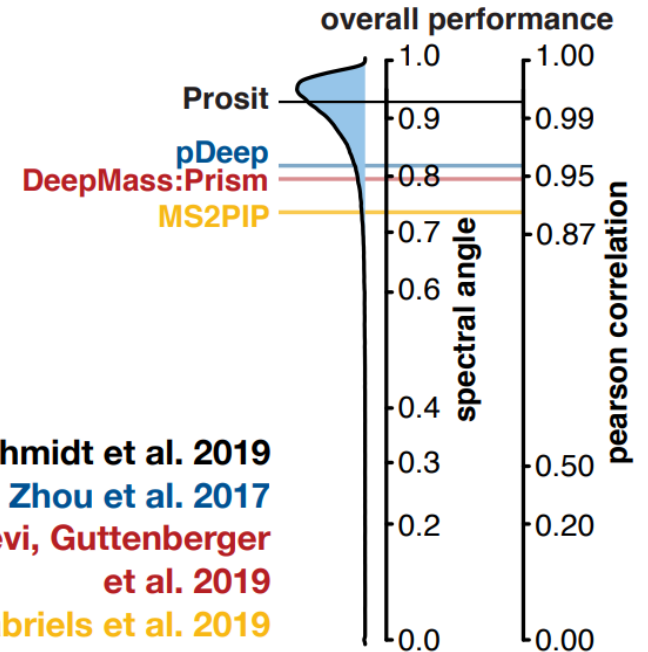
Deep learning peptide properties

- Training data of >100M reference samples
- CNN, LSTM, and Transformer models
- Served >15 billion predictions over the past 4 years

➤ **ML and DL model development and utilization**

Prosit Deep Learning architecture

Gessulat, Schmidt et al. 2019
Zhou et al. 2017
Shivani, Levi, Guttenberger et al. 2019
Gabriels et al. 2019

Teaching and projects

WE WANT YOU

https://www1.ls.tum.de/en/compms/teaching/   APPLY HERE => https://www.mls.ls.tum.de/en/compms/research/theses/

Wintersemester 2023/24

Titel	LV-Nr.	Art	Dauer	Termine	Vortragende/r (Mitwirkende/r)
Examenskolloquium - Oberseminar Bioinformatik (IN2122) ⬚	0000004257	KO	2	🕐⬚	M. Wilhelm ⬚ [L]
Forschungspraktium Bioinformatik ⬚	0000004253	FO	12	🕐⬚	M. Wilhelm ⬚
Fortgeschrittenen-Praktikum Bioinformatik ⬚	0000004256	PR	8	🕐⬚	M. Wilhelm ⬚ [L]
Problem-basiertes Lernen ⬚	0000004252	SE	2	🕐⬚	M. Wilhelm ⬚ [L]
Systems BioMedicine	0000001385	VO	2	🕐⬚	M. List ⬚ [L] , J. Pauling ⬚ , M. Schirmer ⬚ , M. Wilhelm ⬚
Systems BioMedicine ⬚	0000002445	UE	3	🕐⬚	M. List ⬚ [L] , Q. Manz ⬚ , M. Schirmer ⬚ , M. Wilhelm ⬚

Sommersemester 2024

Titel	LV-Nr.	Art	Dauer	Termine	Vortragende/r (Mitwirkende/r)
Angewandte Data Science in den Biowissenschaften ⬚	0000000867	VO	2	🕐⬚	M. Wilhelm ⬚ , M. Schirmer ⬚ , M. List ⬚ , M. Witting ⬚
Examenskolloquium - Oberseminar Bioinformatik (IN2122) ⬚	0000000442	KO	2	🕐⬚	M. Wilhelm ⬚ [L]
Forschungspraktium Bioinformatik ⬚	0000002257	FO	12	🕐⬚	M. Wilhelm ⬚
Fortgeschrittenen-Praktikum Bioinformatik ⬚	0000001278	PR	8	🕐⬚	M. Wilhelm ⬚
Masterpraktikum Bioinformatik ⬚	0000001275	PR	10	🕐⬚	M. Wilhelm ⬚
Problem-basiertes Lernen ⬚	0000001277	SE	2	🕐⬚	M. Wilhelm ⬚
Weiterführende Bioinformatik (IN2399) ⬚	0000000177	VO	3	🕐⬚	N. Kamal ⬚ , B. Rost ⬚ , L. Richter ⬚ , M. Wilhelm ⬚

Parent Project	Project title	Supervisor	[BSc|MSc]	Skills
e.g. ORIGINs/Prosit/	Or key words of work, e.g. add xx features		BSc/MSc/bc	e.g. Software development,
ProteomicsDB	Add BioSchemas cards for proteins/drugs/cell lines	Armin	both	JS, webdev
Prosit	Astral fine-tuning	Joel	both	Python, DL, MS
Prosit	How to de-duplicate peptides/spectra in training for improved learning	Wassim	MSc?	Python, DL, MS
ProteomicsDB	Systematic rescoring of all data	Armin	both	Python, SQL
Prosit	SHAP value background distribution assessment	Joel	both?	Python, DL
Prosit	SHAP values for intensity model	Joel	both?	Python, DL
DLOmix	Automatic publishing to Koina	Omar	both?	Python
ORIGIN	Encoding unknown experimental conditions using LS input	Zixuan (Cory) or Ayla		Python, DL
Oktoberfest	Orbitrap vs Ion trap rescoring gains comparison	sensitivty vs specificity	Mario	master	a bit of python for data analy
Prosit	ABSciex model	Joel		Python, DL, a bit of MS
Prosit	Rolling collision energy model	Joel		Python, DL, a bit of MS
Prosit	ETD,EThcD, ... support	Joel		Python, DL, a bit of MS
Prosit	Combined intensity/RT model	Armin		Python, DL, a bit of MS
Prosit	MSA phospho model	Joel		Python, DL, a bit of MS
ProteomicsDB	Convert to USI	Armin		
ORIGIN	Elution width predictor, iRT to RT (range) alignment	Zixuan (Cory)	MSc	Python, data preprocessing,
Prosit	Evaluate training data size <-> transformer model size dependency on various datasets of increasing complexity (tryptic, HLA, TMT, CID)	Joel		Python, DL, a bit of MS
Prosit	Evaluate Dynamic length Prosit model for rescoring	Joel		data analysis, a bit of MS
general / Review	review of existing search engines, their features, use cases & outputs	Ayla	BSc	literature research, data ana
ENB	Evaluation and implementation of QC metrics for "The Proteomes that Feed the Wo	Armin	both	literature research, MS, data
DROP2AI	Refitting and integration of drug response curves into proteomicsDB	Mario	both	Python, webdev depending
DROP2AI	Systematic outlier detection, refitting and comparison of cell viability data	Mario	both	Python, a bit of machine lea
Oktoberfest	Systematic rescoring and comparison of a large-scale Metaproteomics dataset	Ayla & Mario	BSc	Python, data analysis, data