NEW TOOLS FOR PROTEOME-WIDE CROSS-LINKING MASS SPECTROMETRY

PhD Pre-Defense – Micha J. Birklbauer

27th February 2025

BIOINFORMATICS
RESEARCH GROUP
HAGENBERG

Supervised by FH-Prof. PD DI Dr. Stephan Winkler

in close cooperation with

JKU JOHANNES KEPLER UNIVERSITÄT LINZ

FH OBERÖSTERREICH UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

IMP Research Institute of Molecular Pathology

PROTEOMICS

❑ Proteins are the core components of life

❑ They are integral to controlling cellular function and state

❑ Ultimately they determine the characteristics of any organism

❑ Proteomics is the study of all proteins

MOTIVATION

❑ Research of native protein structures and their interactions is crucial for understanding diseases and pathological processes (vital for drug development)

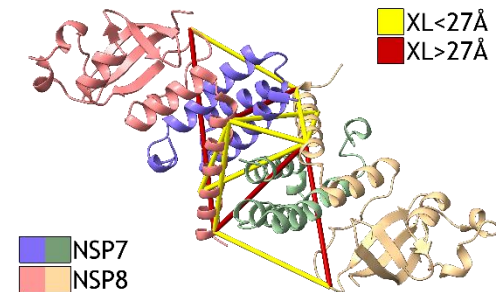❑ In this regard Cross-Linking Mass Spectrometry (XLMS) has become the tool of choice

XL<27Å
XL>27Å

NSP7
NSP8

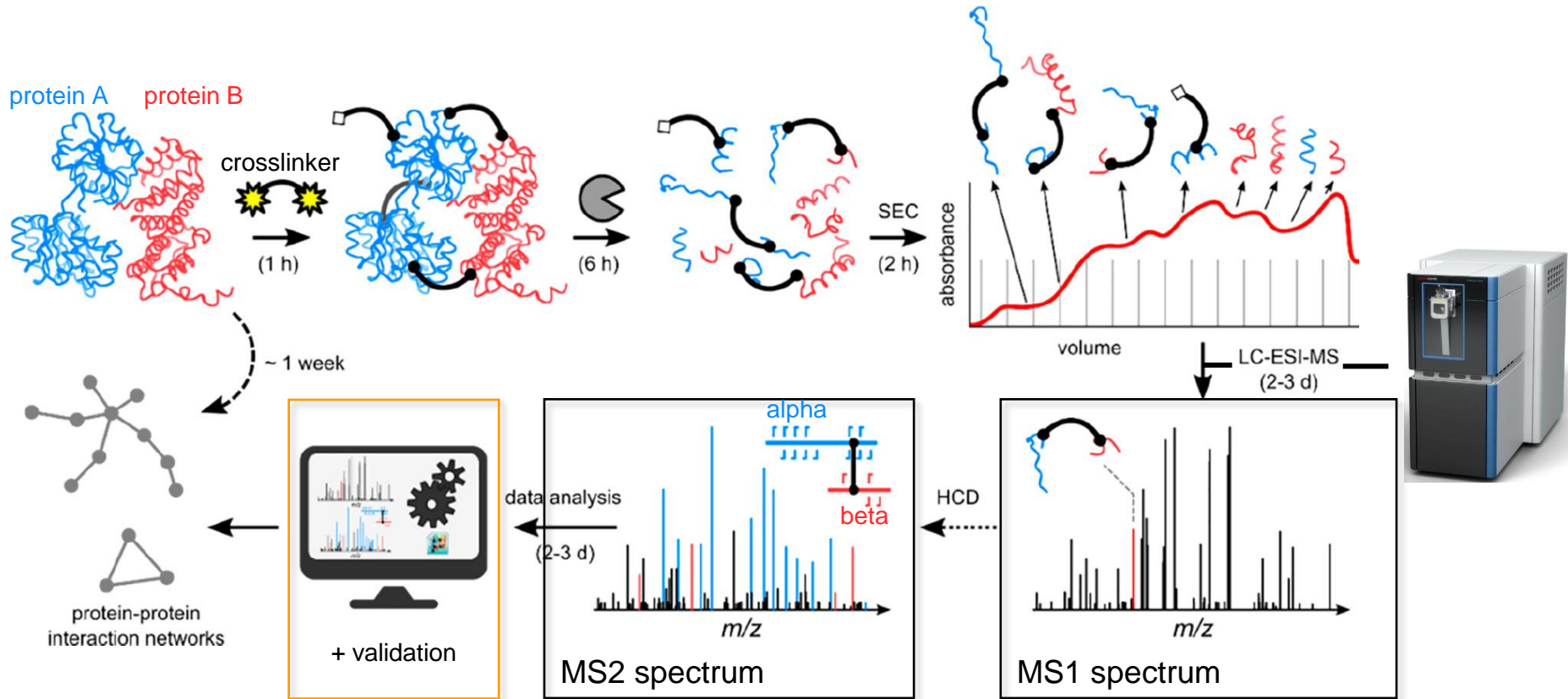**Figure 1**

CROSS-LINKING MASS SPECTROMETRY

Figure 2*

MY RESEARCH:

- Identification of crosslinks from MS2-MS3 acquisition workflows
- Identification of crosslinks from non-cleavable reagents

*Figure adopted from Piersimoni et al., *Chemical Reviews*, 2021

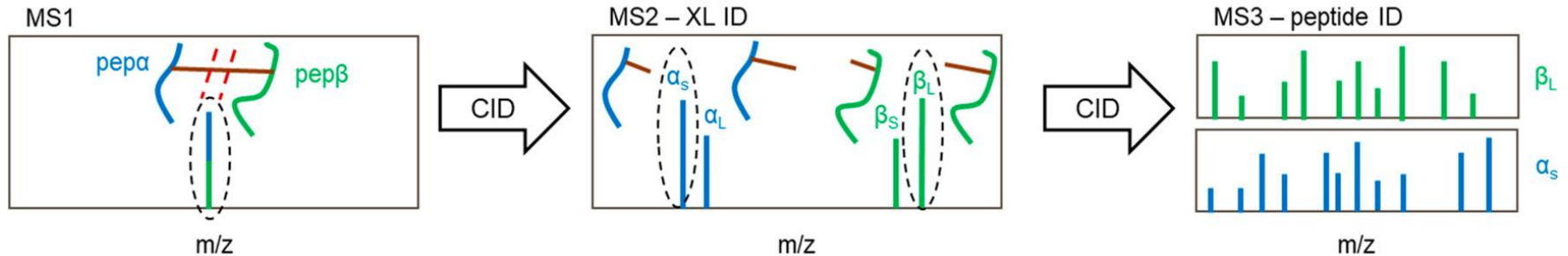# MS2-MS3-BASED CROSSLINK IDENTIFICATION

Figure 3*

MOTIVATION
❑ One of the most common workflows
❑ Software exist, but limited in sensitivity and specificity (crucial for biological application)

PROBLEM
❑ Algorithm of high sensitivity and specificity needed

CHALLENGES
❑ Incorrect precursor isolation (1)
❑ Incorrect C12 assignment from isotope distribution (2)
❑ Matching and combining MS2 and MS3 information

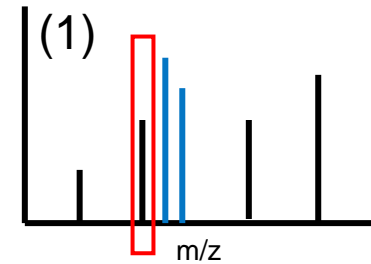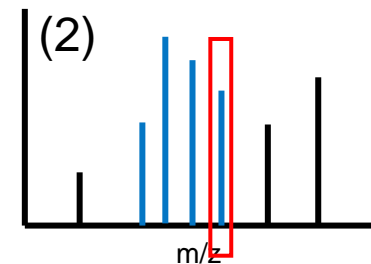**Figure 4**

Figure 5

*Figure adopted from Matzinger et al., *Journal of Proteome Research*, 2021

MS2-MS3-BASED CROSSLINK IDENTIFICATION

INCORRECT PRECURSOR ISOLATION

❑ Verify precursor ion by comparison to MS2 spectrum
❑ If MS3 spectrum originates from non-crosslinked
   precursor → discard spectrum

Figure 6

INCORRECT C12 ASSIGNMENT

❑ Re-calculate C12 mass from MS2 spectrum based on
   known rules for isotope distributions

Figure 7

MATCHING AND COMBINING MS2 / MS3

❑ Match via precursor and product ion mass
   considering mass tolerance and retention time frame
❑ Combination via novel scoring function

$$Score(peptide) = \max\left(scores_{peptide}\right) * (1 + \frac{p}{100})^{n-1}$$

Where:

  $scores_{peptide}$ is a list of all scores of that peptide
  p is a user-defined boost parameter (default 20)
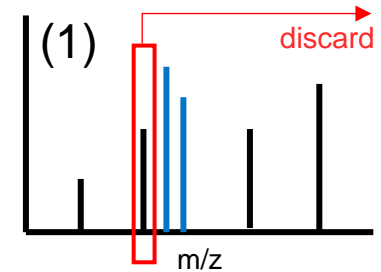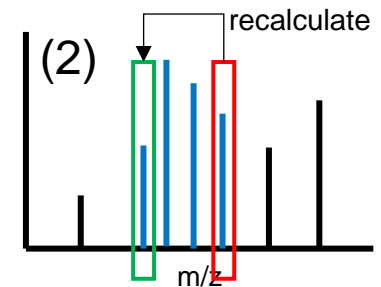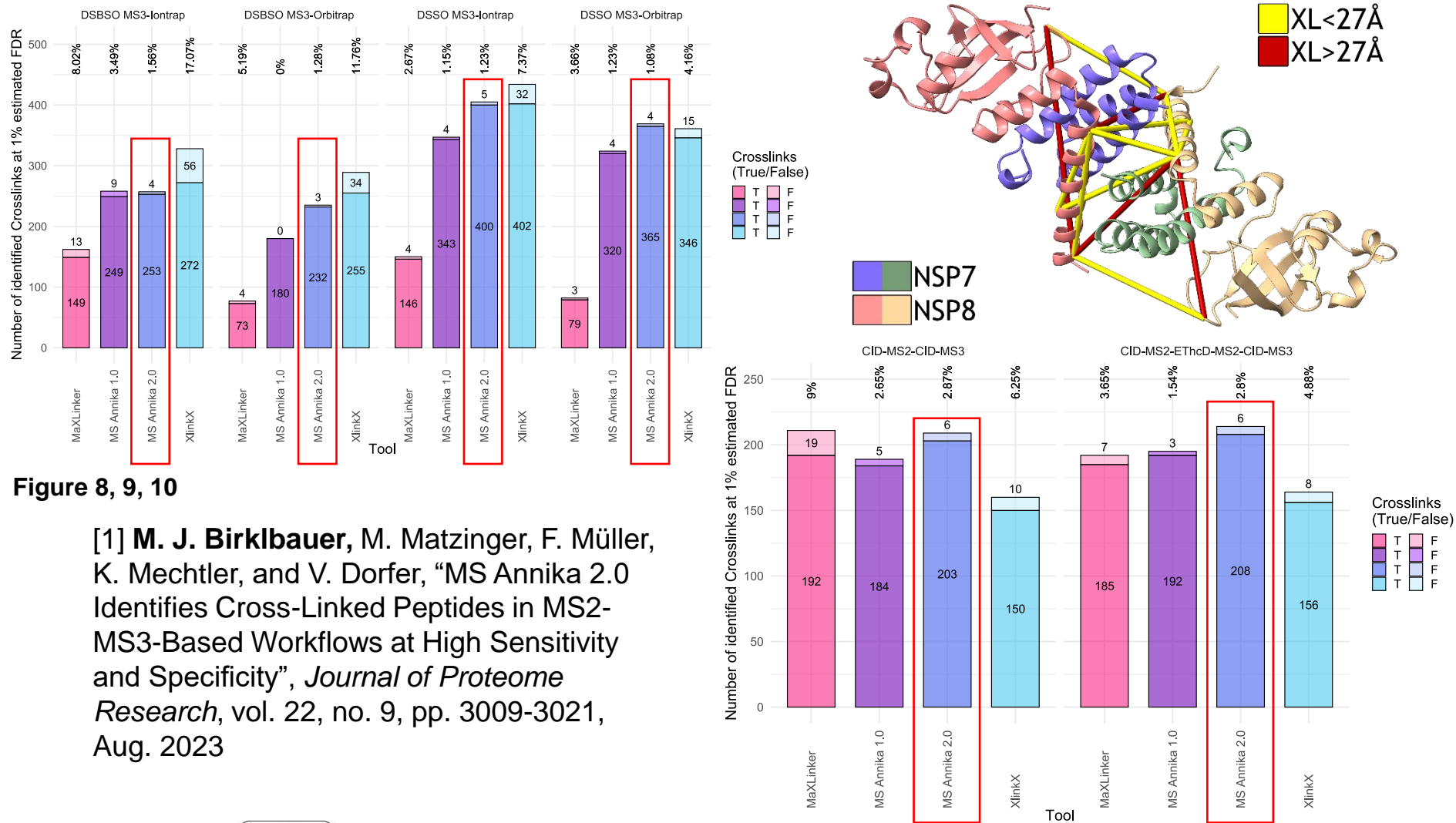 n is the number of unique scan numbers that peptide has been identified in

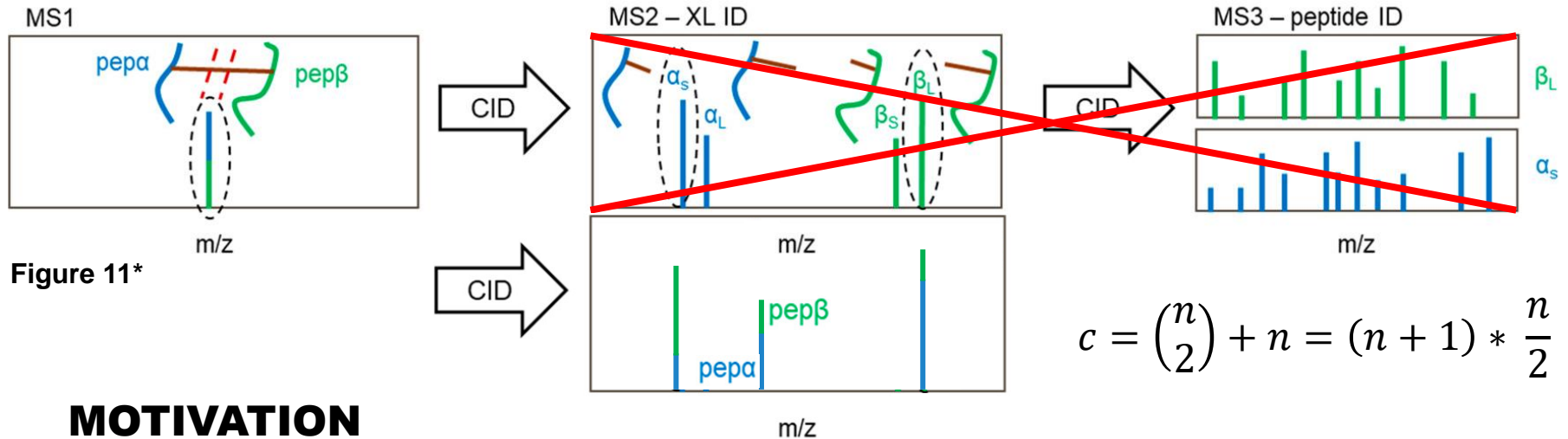# MS2-MS3-BASED CROSSLINK IDENTIFICATION

❑ Algorithms implemented in **MS Annika 2.0** [1]

Figure 8, 9, 10

[1] **M. J. Birklbauer,** M. Matzinger, F. Müller, K. Mechtler, and V. Dorfer, "MS Annika 2.0 Identifies Cross-Linked Peptides in MS2-MS3-Based Workflows at High Sensitivity and Specificity", *Journal of Proteome Research*, vol. 22, no. 9, pp. 3009-3021, Aug. 2023

NON-CLEAVABLE CROSSLINK IDENTIFICATION

Figure 11*

$$c = \binom{n}{2} + n = (n+1) * \frac{n}{2}$$

MOTIVATION
- ❑ Non-cleavable crosslinkers are the most common crosslinker reagent
- ❑ Limited software support for data analysis, especially for complex samples

PROBLEM
- ❑ An efficient algorithm for proteome-wide searches is needed

CHALLENGES
- ❑ Reduction of search space
- ❑ Efficient search algorithm for considering potentially millions of peptide candidates and tackling the *n*-squared problem

*Figure adopted from Matzinger et al., *Journal of Proteome Research*, 2021

NON-CLEAVABLE CROSSLINK IDENTIFICATION

REDUCTION OF SEARCH SPACE

❑ Divide and conquer approach: identify one of the two
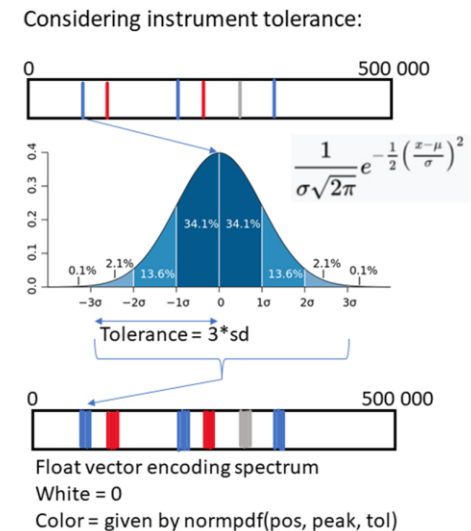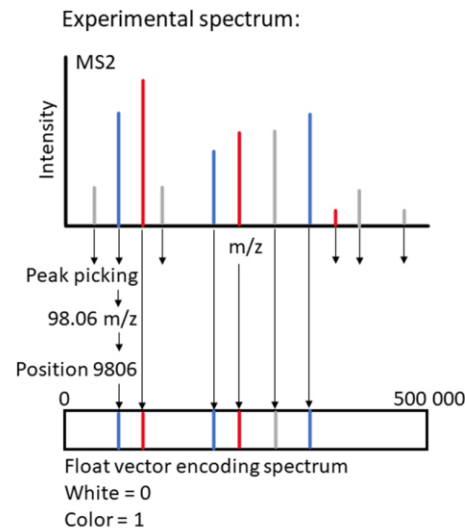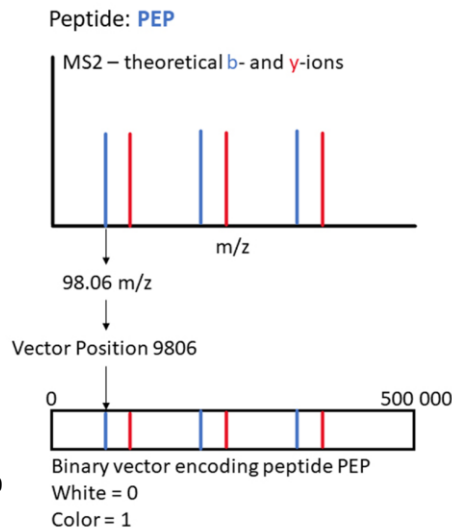   cross-linked peptides first, the second peptide can
   then be more easily inferred

EFFICIENT SEARCH ALGORITHM

❑ Encode peptides and spectra as sparse vectors
❑ Fast scoring by matrix multiplication

Figure 12

Micha J. Birklbauer

NON-CLEAVABLE CROSSLINK IDENTIFICATION

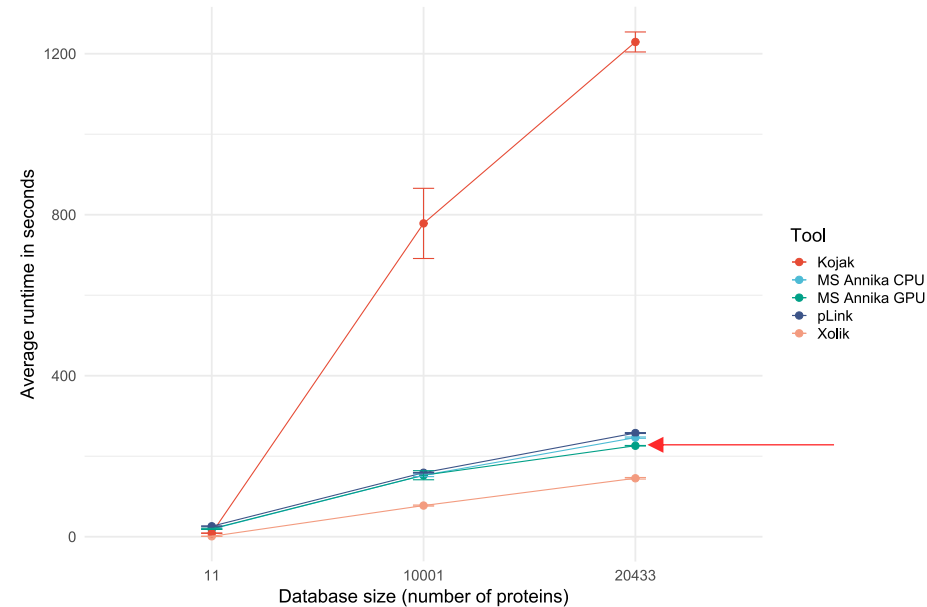❏ Algorithms implemented in **MS Annika 3.0** [2]

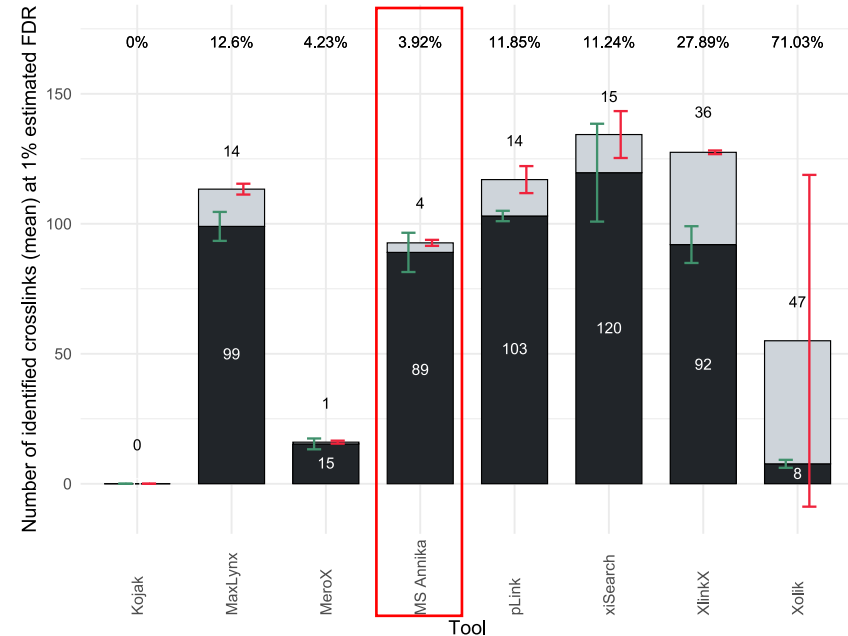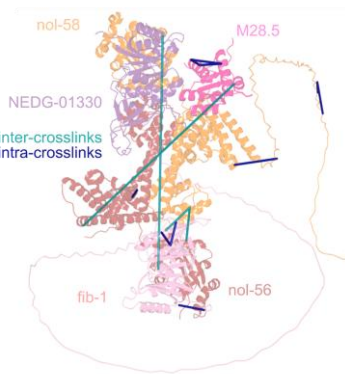**Figure 13, 14, 15**

[2] **M. J. Birklbauer**, F. Müller, S. S. Geetha, M. Matzinger, K. Mechtler, and V. Dorfer, "Proteome-wide Non-Cleavable Crosslink Identification with MS Annika 3.0 Reveals the Structure of the C. elegans Box C/D Complex", *Communications Chemistry*, vol. 7, no. 1, Dec. 2024

Improved structure of the *C. elegans* Box C/D RNP complex
Database size: ~26 000 proteins

RESEARCH PROGRESS

Cross-linking Publications (Published or in Preparation)

- **M. J. Birklbauer,** M. Matzinger, F. Müller, K. Mechtler, and V. Dorfer, "MS Annika 2.0 Identifies Cross-Linked Peptides in MS2-MS3-Based Workflows at High Sensitivity and Specificity", *Journal of Proteome Research*, vol. 22, no. 9, pp. 3009-3021, Aug. 2023

- **M. J. Birklbauer**, F. Müller, S. S. Geetha, M. Matzinger, K. Mechtler, and V. Dorfer, "Proteome-wide Non-Cleavable Crosslink Identification with MS Annika 3.0 Reveals the Structure of the C. elegans Box C/D Complex", *Communications Chemistry*, vol. 7, no. 1, Dec. 2024

- F. Müller, B. R. Brutiu, I. Saridakis, T. Leischner, **M. Birklbauer**, M. Matzinger, T. Lendl, S. Shaaban, V. Dorfer, N. Maulide, and K. Mechtler, "A Journey Towards Developing a New Cleavable Crosslinker Reagent for In-Cell Crosslinking", in review at *Communications Chemistry*

- **M. J. Birklbauer**, F. Müller, M. Matzinger, K. Mechtler, and V. Dorfer, "Unified Down-Stream Analysis of Crosslink Results with pyXLMS", is a planned journal article to be submitted in 2025 to *Bioinformatics*

- F. Müller, **M. J. Birklbauer**, D. Hollenstein, V. Dorfer, and K. Mechtler, "Optimized Crosslink Quantification using Data-Independent Acquisition", is a planned journal article to be submitted in 2025 to *Nature Communications*

Other Publications (Published or in Preparation)

- A. Grimaud, **M. J. Birklbauer**, L. Levitsky, L. Buur, V. Gorshkov, Z. Udvardy, C. Lennartsson, and V. Schwämmle, "Making Sense of Internal Ions" is a planned journal article to be submitted in 2025 to the *Journal of Proteome Research*

Software

- ***MS Annika*** is a cross-linking search engine capable of identifying crosslinks from both cleavable and non-cleavable reagents, and with support for a variety of acquisition workflows including multi-stage tandem acquisition. MS Annika is available at https://github.com/hgb-bin-proteomics/MSAnnika

Conference Contributions

21/03/2022	Poster at the EuBIC-MS Winter School
13/09/2022	Talk at APRMS 2022
15/01/2023	Talk and poster at the EuBIC-MS Developers Meeting
27/09/2023	Poster at APRMS 2023
15/01/2024	Co-Organizer and poster at the EuBIC-MS Winter School
25/09/2024	Talk and „Best Presentation" winner at APRMS 2024