Top-Down Proteomics Workshop

Bridging the Gaps between Academia, Clinical Research, and Biopharmaceutical Industry

Presiders:

Yuri van der Burgt (Leiden University), Fanny Caroline Liu (Florida State University)

72nd ASMS Meeting

Anaheim, CA

June 3, 2024

Outline

- **Introduction** **5 min**
  - Consortium for Top-Down Proteomics (CTDP)
  - CTDP Early Career Researcher (ECR) Committees
- **Updates from interlaboratory research initiatives** **20 min**
  - Gold standard dataset (Kyowon Jeong, University of Tuebingen)
  - MS^2 Fragmentation (Luca Fornelli, University of Oklahoma)
- **Panel discussions** **50 min**
  - Top-down proteomics in the Clinical Research
  - Top-down proteomics in the Biopharmaceutical Industry

CTDP

Consortium for Top-Down Proteomics

Bringing Proteoforms to Life℠

Promoting collaboration, education, and innovative research to accelerate the comprehensive analysis of all human proteoforms

Currently Free to **JOIN** at **CTDP.org**

Building the Community

Proteoform Thursdays

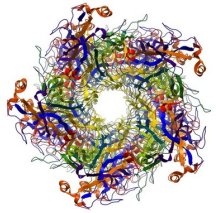
A monthly seminar series featuring the latest research from leading early-career scientists in top-down proteomics and proteoform biology

Resources for the Community

Software	17 publicly available software tools
Publications	Up to date bibliography of ~600 publications
Methods	Top-down standard methods
ProForma	Guide for standard proteoform notations
Proteoform Atlas	Database of proteoforms

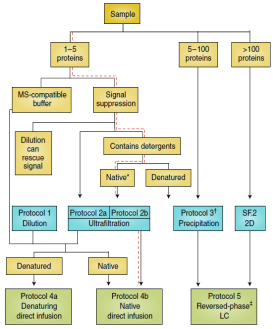
Recent Consortium Initiatives

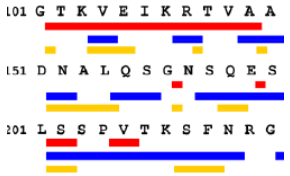
CTDP.org



Top-down mass spectrometry of native proteoforms and their complexes: a community study, Nature Methods, 2024

Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry, J. Am. Soc. Mass Spectrom., 2020





Best practices and benchmarks for intact protein analysis for top-down mass spectrometry, Nature Methods, 2019 [**>300 Citations**]

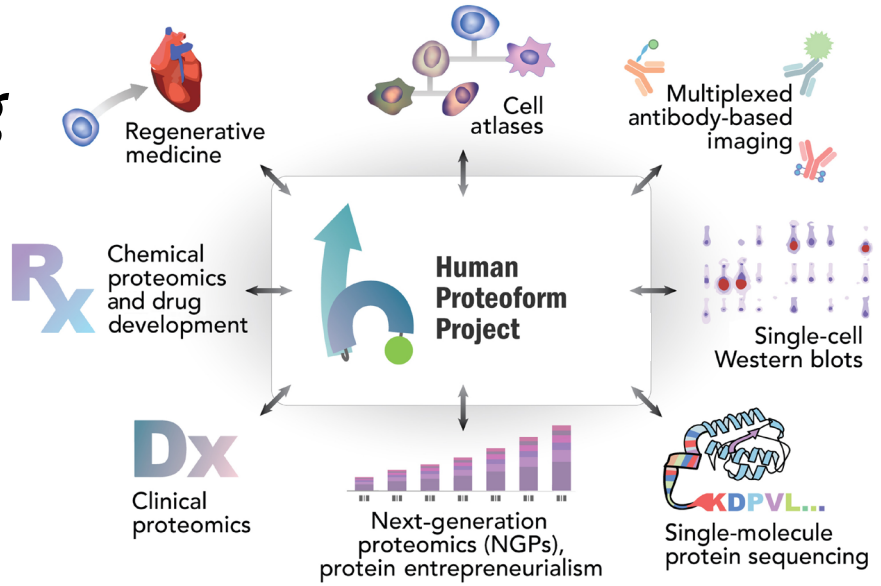
ProForma: A Standard Proteoform Notation, J. Proteome Res., 2018

Current Consortium Initiatives

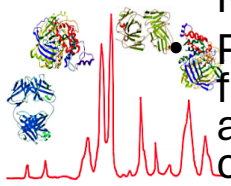
CTDP.org

The Human Proteoform Project: Defining the human proteome, Science Advances, 2021

- Ambitious initiative to provide the bridge from genotype to phenotype
- Drive the creation and development of new technologies and methodologies (similar to Human Genome Project)
- **>24,000 downloads**



CE-MS Initiative



- Engage a range of laboratories to assess the state of the art in CE-MS and CE-MS/MS
- Publish guidelines on the best practices for successful CE-MS and CE-MS/MS analysis of proteins and protein complexes

Human Proteoform Atlas

- Build infrastructure to host the human proteoforms, enabling the facile sharing of data and information
- Designed to interact with other important repositories and databases

CTDP Early Career Researcher (ECR) Committee

- **Main tasks**
  - Promote education and outreach
  - Propose and organize interlaboratory studies

- **2-Year term**
  - 1st ECR Committee, 2022-2024
  - 2nd ECR Committee, 2024-2026

1st CTDP ECR Committee, 2022-2024

Mowewi Zhou (Chair)

Pacific Northwest National Laboratory

Caroline Dehart (Co-Chair)

Frederick National Laboratory for Cancer Research

Serife Ayaz-Guner (Co-Chair Europe/Africa)

Izmir Institute of Technology

Kyowon Jeong (Co-Chair Europe/Africa)

University of Tuebingen

Frederik Lermyte (Co-Chair Europe/Africa)

Technical University of Darmstadt

Lissa Anderson (Co-Chair Americas)

National High Magnetic Field Laboratory

Luca Fornelli (Co-Chair Americas)

University of Oklahoma

Gloria Sheynkman (Co-Chair Americas)

University of Virginia

Huilin Li (Co-Chair Asia/Oceania)

Sun Yat-sen University

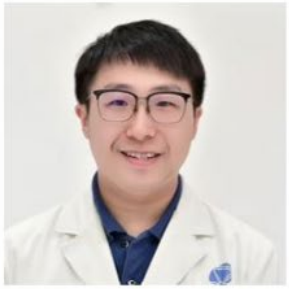
Guanbo Wang (Co-Chair Asia/Oceania)

Peking University

Interlab research initiatives

- **Gold-standard datasets**
  - Kyowon Jeong (University of Tuebingen)
  - Frederik Lermyte (Technical University of Darmstadt)
  - Serife Ayaz-Guner (Izmir Institute of Technology)
- **MS^2 Fragmentation**
  - Luca Fornelli (University of Oklahoma)
  - Lissa Anderson (National High Magnetic Field Laboratory)
  - Caroline Dehart (Frederick National Laboratory for Cancer Research)

2nd CTDP ECR Committee, 2024-2026



Cookson Chiu
(Shenzhen Bay Laboratory)

Bryon Drown
(Purdue University)

Elizabeth Duselis
(Genentech)

Fabio Gomes
(Virginia Commonwealth University)

Oliver Hale
(University of Birmingham)

Fanny Caroline Liu
(Florida State University)

Corinne Lutomski
(University of Oxford)

Rafael Melani
(Thermo Scientific)

Kellye Cupp-Sutton
(University of Alabama)

Sem Tamara
(Johnson & Johnson)

Currently planning new interlab initiatives.

- Please reach out if you are interested in being involved.
- Please use our survey to inform us of challenges when using the TDP approach.

Progress report on CTDP ECR Gold standard dataset project

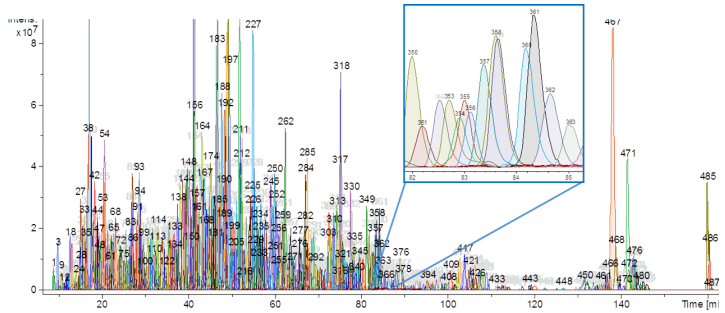
Kyowon Jeong

Kohlbacher lab, University of Tübingen in Germany

CTDP ECR committee

Gold standard TDP dataset

- Gold standard TDP dataset: a set of high-quality TDP MS1 and MS2 spectra, with thorough assignments.



- Ideally, all peaks should be completely annotated leading to the perfect annotation of all precursor and fragment ion masses.

- Standard to measure the goodness of your own datasets and analysis pipelines

- A step stone for method development!

Implication of gold standard dataset

- Validation and evaluation of existing TDP workflows
- Develop computational method for data analysis.
- Setting up TDP protocols for research objectives.

Gold standard TDP dataset project

CTDP
Consortium for Top-Down Proteomics
Bringing Proteoforms to Life℠

Mowewi Zhou (Chair)

Pacific Northwest National Laboratory

Caroline Dehart (Co-Chair)

Frederick National Laboratory for Cancer Research

Serife Ayaz-Guner (Co-Chair Europe/Africa)

Izmir Institute of Technology

Kyowon Jeong (Co-Chair Europe/Africa)

University of Tuebingen

Frederik Lermyte (Co-Chair Europe/Africa)

Technical University of Darmstadt

+ many contributors!!

Lissa Anderson (Co-Chair Americas)

National High Magnetic Field Laboratory

Luca Fornelli (Co-Chair Americas)

University of Oklahoma

Gloria Sheynkman (Co-Chair Americas)

University of Virginia

Huilin Li (Co-Chair Asia/Oceania)

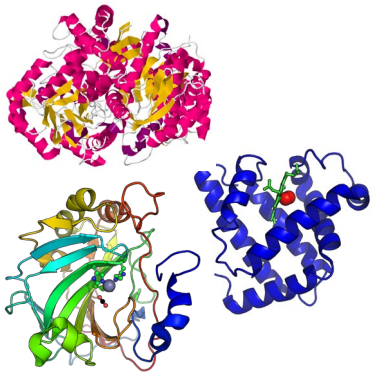
Sun Yat-sen University

Guanbo Wang (Co-Chair Asia/Oceania)

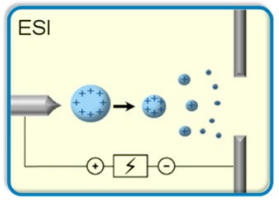
Peking University

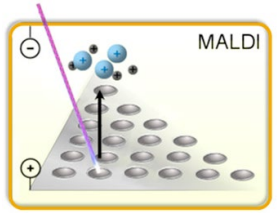
Design & objectives

Selected well known proteins:
Myoglobin, carbonic anhydrase ...

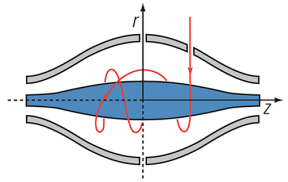


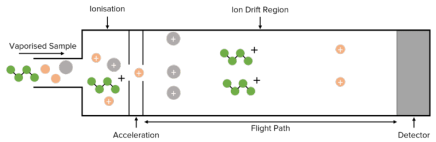
Ion source

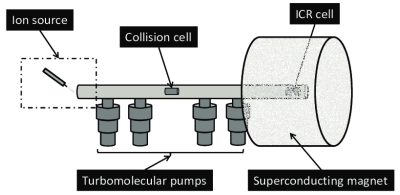




Mass analyzer



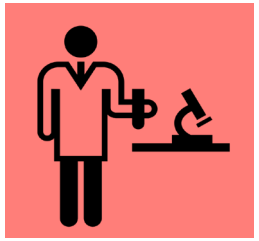


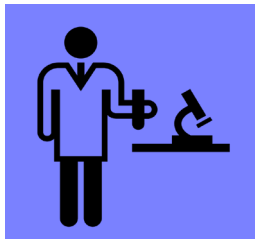


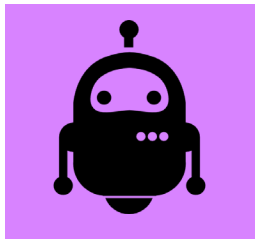
Fragmentation

- MS 1 only
- HCD
- ETD

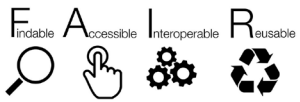
Annotation





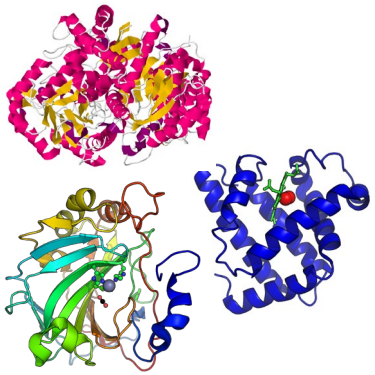


Distribution

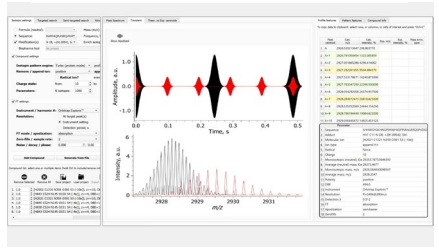


Design & objectives

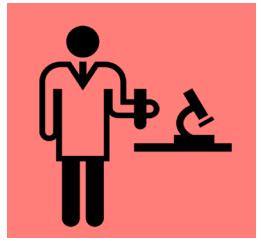
Selected well known proteins:
Myoglobin, carbonic anhydrase ...

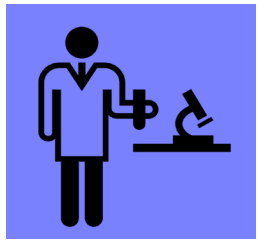


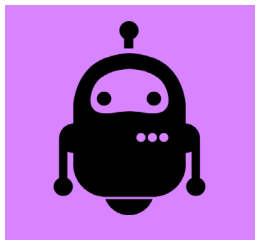
FTMS Simulator



Annotation







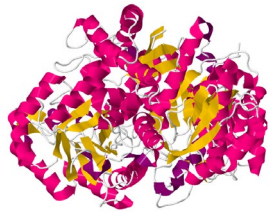
Distribution

Findable Accessible Interoperable Reusable

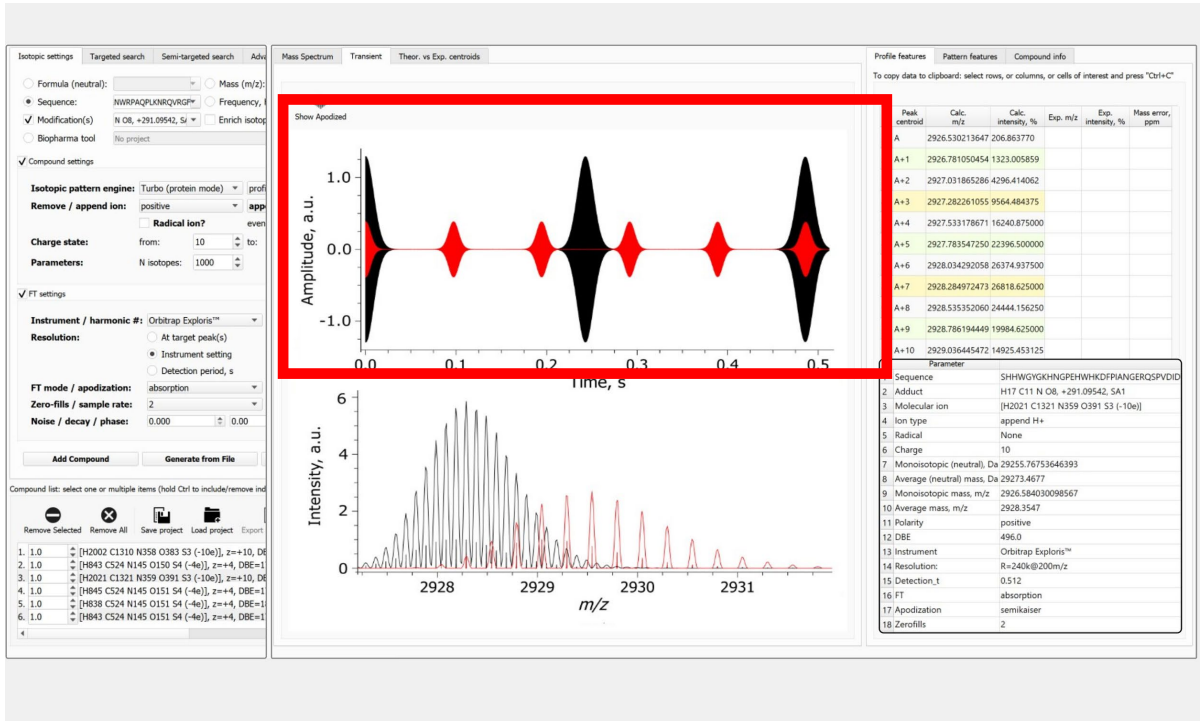
in silico spectra by FT simulator

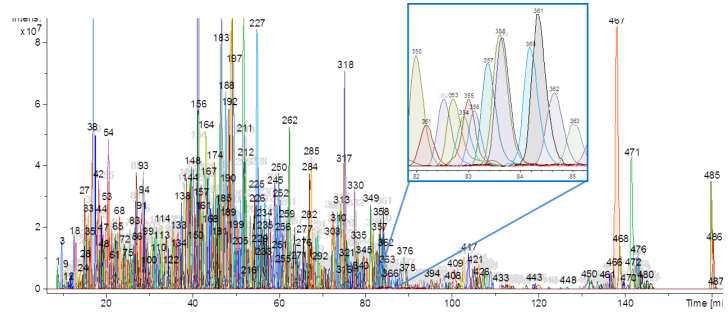
Spectroswiss

FTMS Simulator



Resolution
Noise level
Ion type
Intensities
Instrument type
...



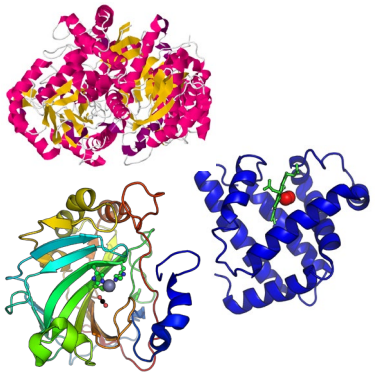


FTMS Data Simulation via Time-Domain Transients

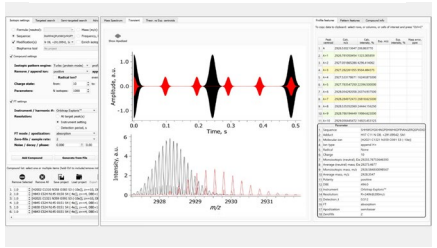
All true fragment masses are known!!

Design & objectives

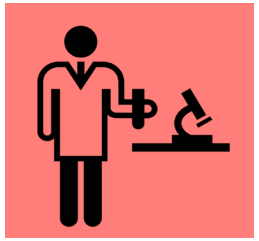
Selected well known proteins:
Myoglobin, carbonic anhydrase …

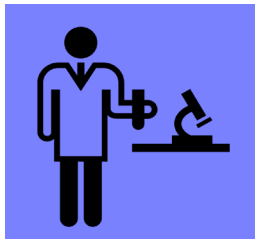


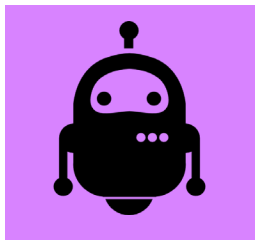
FTMS Simulator



Annotation ***w/ FPR xx% and FNR xx%***







Distribution

Annotation team

Dr. Philipp Kaulich in Kiel University

Kiel University
Christian-Albrechts-Universität zu Kiel

Dr. Wonhyeuk Jung in Yale University

Yale University
School of Medicine

Dr. Serife Ayaz-Guner in Abdullah Gül University

ABDULLAH GÜL
ÜNİVERSİTESİ

Entrapping manual annotators

- The annotators annotate the spectrum given target sequences based on our guidelines.
- Target sequences define the masses to look at.
- For those target masses, FNR is expected to be ~0.0
- But in reality, other masses would be present – modified proteoforms, internal ions, etc.
- If we provide the *in silico* datasets only containing correct masses, FNR cannot be measured for such unexpected masses.
- Thus, we provide the datasets with slightly modified sequences that contain some portion of incorrect masses.
- For those masses (only known for data generation team) we measure FNR.

Experimental datasets

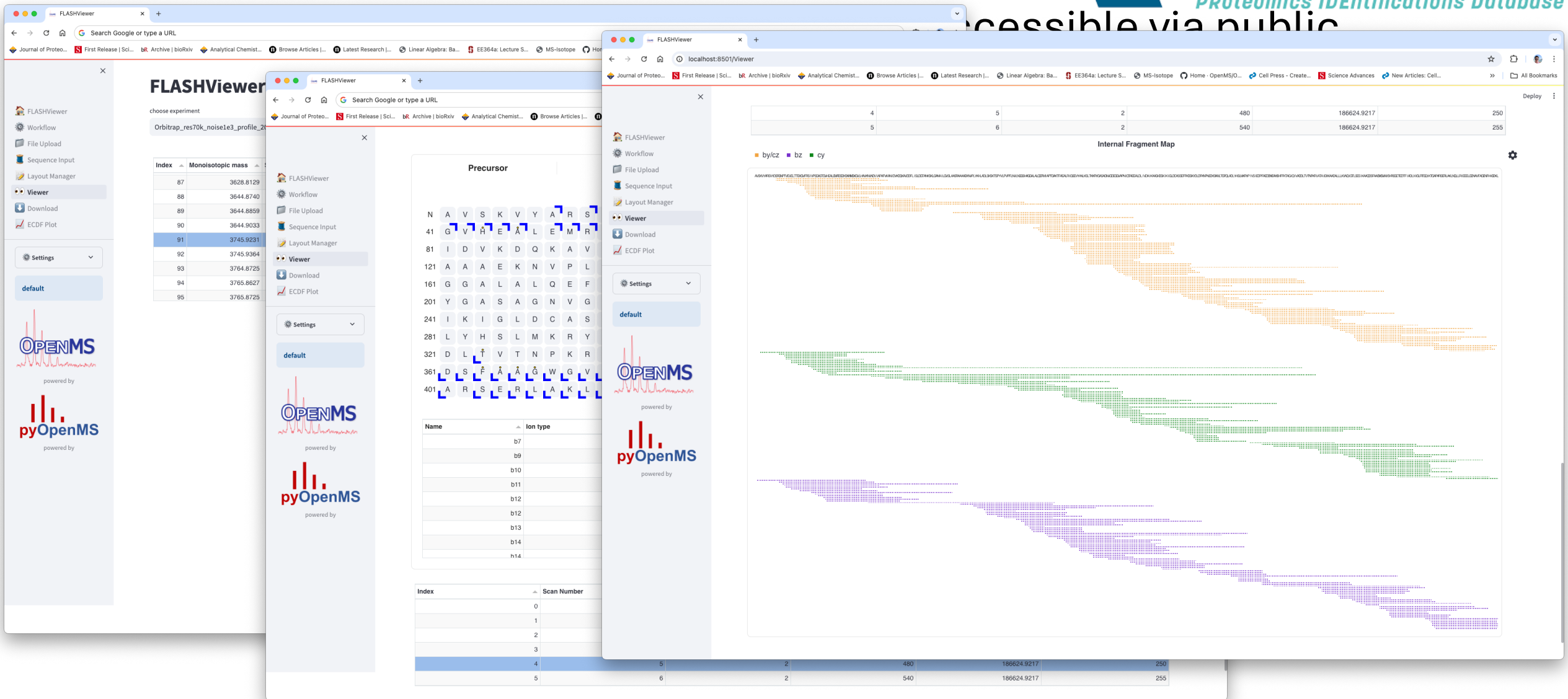
- Initiative launched in 2023 (currently ~10 participants)
- Participants cover the major instrument types for top-down protein MS (data acquisition will start in summer 2024)
- More annotators would be recruited who are willing to help process Orbitrap, Synapt (Waters), and solariX (Bruker) data

Distribution of the datasets

Ayesha Feroz

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

PRIDE
PRoteomics IDEntifications Database



One example use case of *in silico* spectra

- Knowing ALL masses that generated spectrum can be handy when analyzing internal fragment ions.

- To confirm an internal fragmentation by a mass, we need confirm two things:
  - The mass is correct.
  - The assignment is correct

Assignment ambiguity

Employing a 1 ppm mass error tolerance as the only search constraint allowed ca. 80% of internal fragments to be unambiguously assigned within the mass range of 9–29 kDa and ca. 70% at 66 kDa (Table S1). The 1 ppm error

analytical chemistry

pubs.acs.org/ac Article

Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment

Nicholas D. Schmitt, Joshua M. Berger, Jeremy B. Conway, and Jeffrey N. Agar*

Cite This: *Anal. Chem.* 2021, 93, 6355–6362 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A major limitation of intact protein fragmentation is the lack of sequence coverage within proteins' interiors. We show that collisionally activated dissociation (CAD) produces extensive

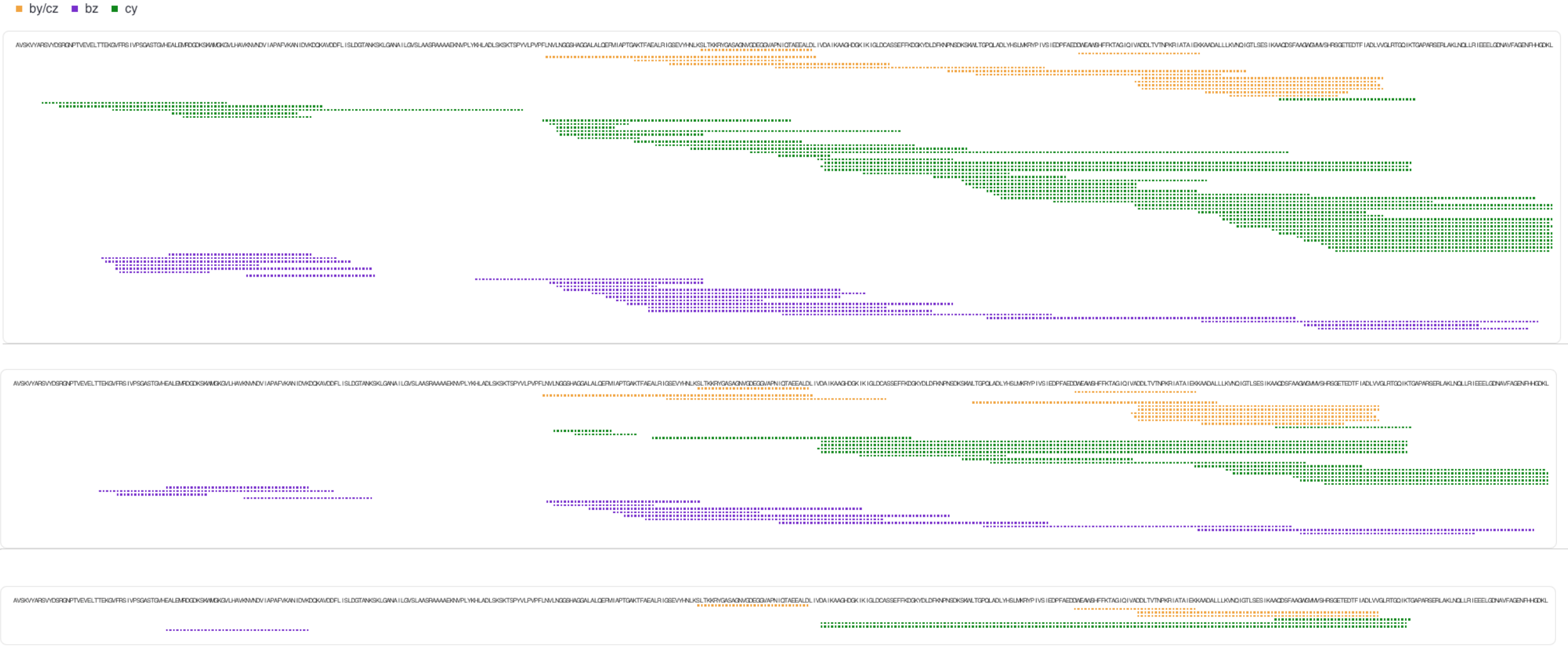
Assignment ambiguity + mass error?

- What if masses are incorrect but we do not know?
- What is the expected erroneous internal ion annotation rate w.r.t. mass error rate?
- This is a tricky question to answer with experimental datasets because we do not have golden incorrect masses!!
- But we can have 100% incorrect deconvolved masses from *in silico* datasets!

Procedure

- We generated *in silico* spectrum from Yeast enolase protein (46.5 kDa, 436 amino acids) only with b and y ions but ***without*** any internal ion.
- Deconvolved using FLASHDeconv yielding 250 masses : 189 correctly matched.
- What about 61 remaining obvious ***false positive but non-internal*** ion masses?

61 incorrect masses vs. internal ion fragments



2, 1, 0 ppm tolerance for internal ion matching

Conclusion

- Gold standard dataset project is on its way
- Trying to collect high quality fully annotated MS and MS2 spectra
- Multiple labs will be involved to generate high quality datasets from various protocol and instruments
- *in silico* spectra will serve as
  - Gold standard dataset itself
  - A method for manual validation quality evaluation
  - Resource for other implications (e.g., internal fragment ion study)
- The datasets will be openly accessible through a public repository.
- We will host a webpage for the interactive visualization of the datasets.

Acknowledgement

CTDP
Consortium for Top-Down Proteomics
Bringing Proteoforms to Life℠

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Ayesha Feroz
Dr. Oliver Kohlbacher

Spectroswiss

Dr. Yury Tsybin
Dr. Konstantin Nagornov

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Dr. Frederik Lermyte

C A U
Kiel University
Christian-Albrechts-Universität zu Kiel

Dr. Philipp Kaulich

ABDULLAH GÜL
ÜNİVERSİTESİ

Dr. Serife Ayaz-Guner

Yale University
School of Medicine

Dr. Wonhyeuk Jung

ECR CTDP Protocol on MS²: an update

Luca Fornelli, Lissa Anderson, Caroline DeHart, and ECR Committee

Top-down workshop

72nd ASMS Conference

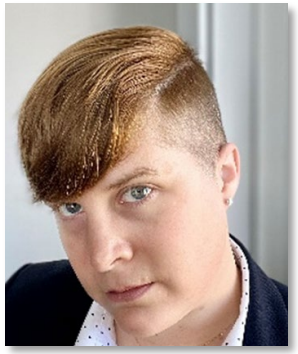
Anaheim, CA, June 3, 2024

Protocol on MS^2: **Organizers**

CTDP
Consortium for Top-Down Proteomics

- Project design & implementation:

Lissa Anderson
National High Magnetic Field Laboratory



Caroline DeHart
Frederick National Laboratory for Cancer Research

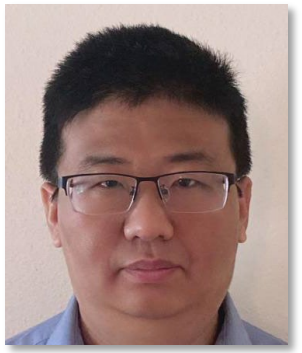
Luca Fornelli
University of Oklahoma

- Data analysis and bioinformatics:

Kyowon Jeong
University of Tuebingen

Ryan Fellers
Northwestern University

Joseph Greer
Northwestern University



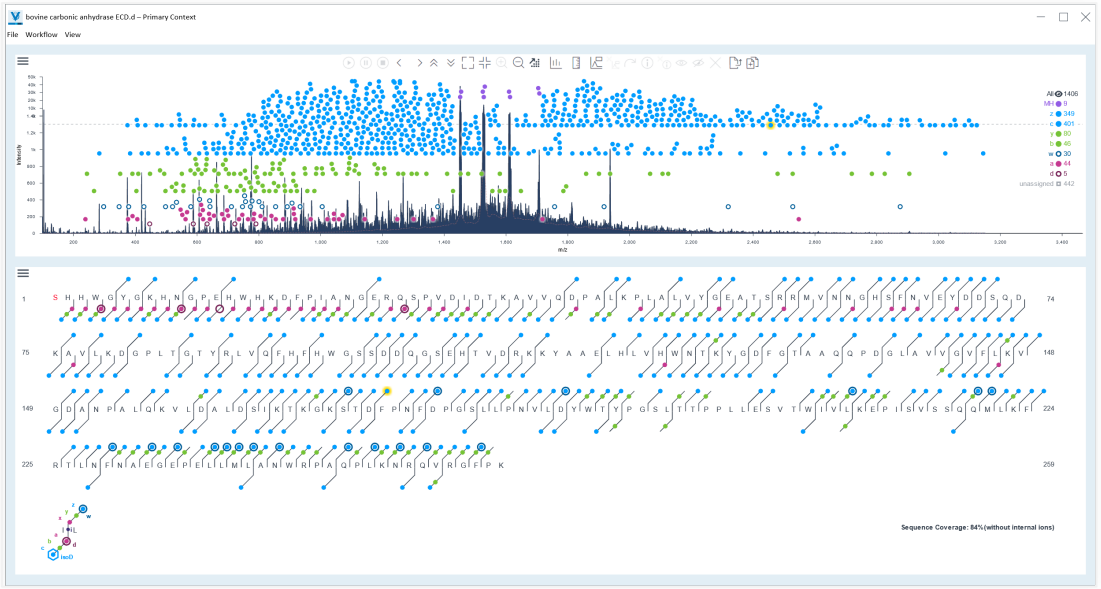
Xiaowen Liu
Tulane University

Adrian Guthals
Agilent

What is ExDViewer?

Agilent freeware for analysis and visualization of top- and middle-down protein mass spectrometry data.

- For protein sequence confirmation



Key Features

- In addition to Agilent .d, accepts other vendor and open-source file formats as input.
- One **deconvolution** workflow with beginner-friendly **presets**. Match to a target sequence or use untargeted deconvolution.
- Annotate **CID, ECD**, and other fragmentation modes.
- Build custom target sequences in the **Target Editor**.
- Use **interactive visualizations** for data review.
- Use the **variable modification search** to help find evidence for sequence variants.
- **Share** a URL link to view results online.

What does freeware mean? As freeware, ExDViewer is Agilent proprietary software that is distributed at no monetary cost to users under specific terms of use defined in the End User License Agreement. ExDViewer is not free software or open-source software.

Previous CTPD work on MS[1]

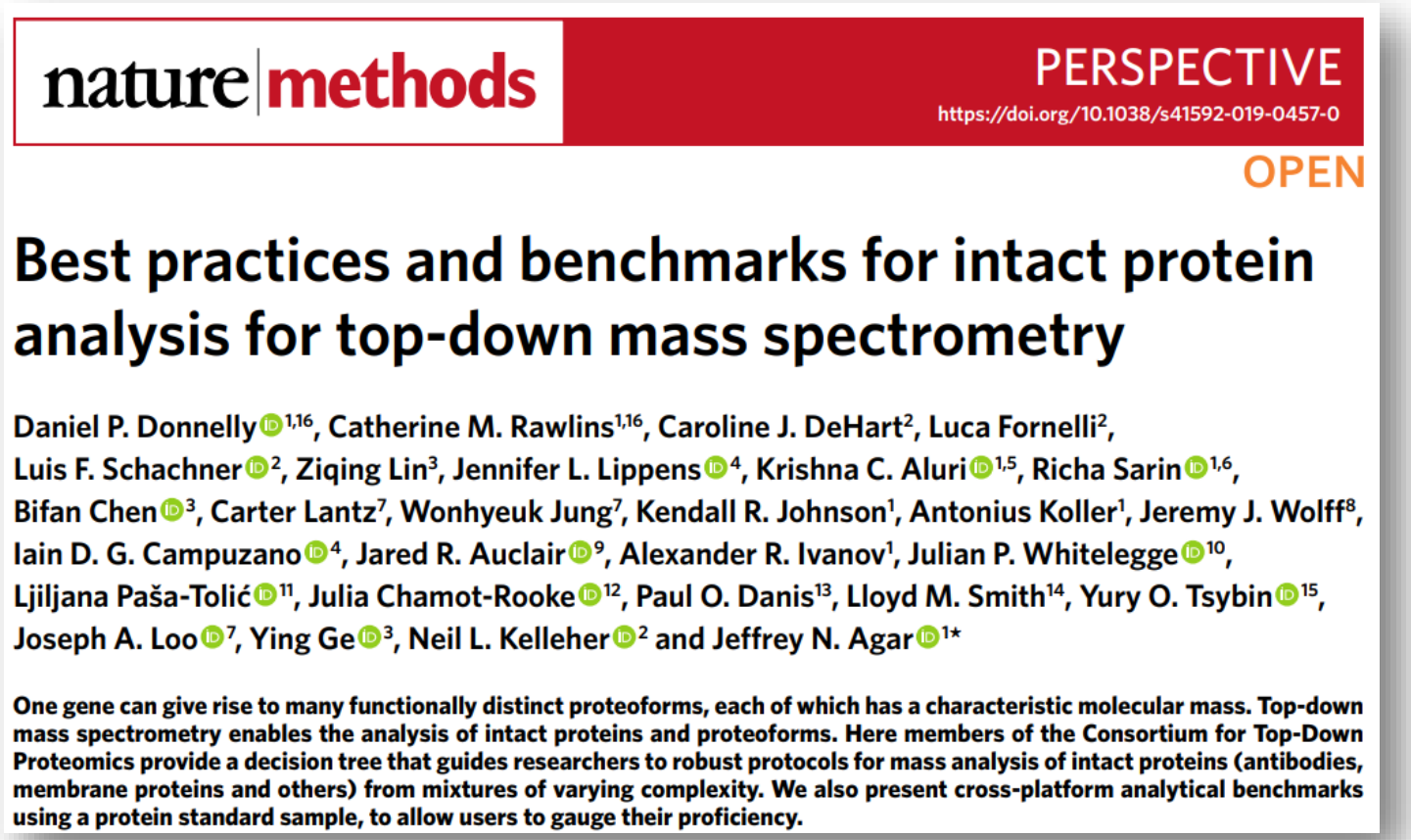
CTDP
Consortium for Top-Down Proteomics

- **Blueprint** for the new study:

- Interlaboratory study led by Prof. Agar in 2019, published on *Nature Methods*.

- This work included useful information for scientists approaching top-down MS for the first time.

- Intact mass analysis via MS1 is usually the entry point to the top-down universe.



nature methods
PERSPECTIVE
https://doi.org/10.1038/s41592-019-0457-0
OPEN

Best practices and benchmarks for intact protein analysis for top-down mass spectrometry

Daniel P. Donnelly[1,16], Catherine M. Rawlins[1,16], Caroline J. DeHart[2], Luca Fornelli[2], Luis F. Schachner[2], Ziqing Lin[3], Jennifer L. Lippens[4], Krishna C. Aluri[1,5], Richa Sarin[1,6], Bifan Chen[3], Carter Lantz[7], Wonhyeuk Jung[7], Kendall R. Johnson[1], Antonius Koller[1], Jeremy J. Wolff[8], Iain D. G. Campuzano[4], Jared R. Auclair[9], Alexander R. Ivanov[1], Julian P. Whitelegge[10], Ljiljana Paša-Tolić[11], Julia Chamot-Rooke[12], Paul O. Danis[13], Lloyd M. Smith[14], Yury O. Tsybin[15], Joseph A. Loo[7], Ying Ge[3], Neil L. Kelleher[2] and Jeffrey N. Agar[1*]

One gene can give rise to many functionally distinct proteoforms, each of which has a characteristic molecular mass. Top-down mass spectrometry enables the analysis of intact proteins and proteoforms. Here members of the Consortium for Top-Down Proteomics provide a decision tree that guides researchers to robust protocols for mass analysis of intact proteins (antibodies, membrane proteins and others) from mixtures of varying complexity. We also present cross-platform analytical benchmarks using a protein standard sample, to allow users to gauge their proficiency.

Protocol on MS^2: **Goals**

CTDP
Consortium for Top-Down Proteomics

- Description of main **data acquisition parameters** and **their impact** on MS^2 spectra:
  - For various instruments
  - Using different ion activations
- Distinction between **two common scenarios** in data acquisition parameters for top-down MS:
  - "targeted" (e.g., analysis of recombinant proteins, analysis of biotherapeutics)
  - "high-throughput" (i.e., large scale top-down proteomics)
- Definition of **realistic expectations** for MS^2 results:
  - What is a good sequence coverage?
  - What metrics matter?

Protocol on MS^2: **Study structure**

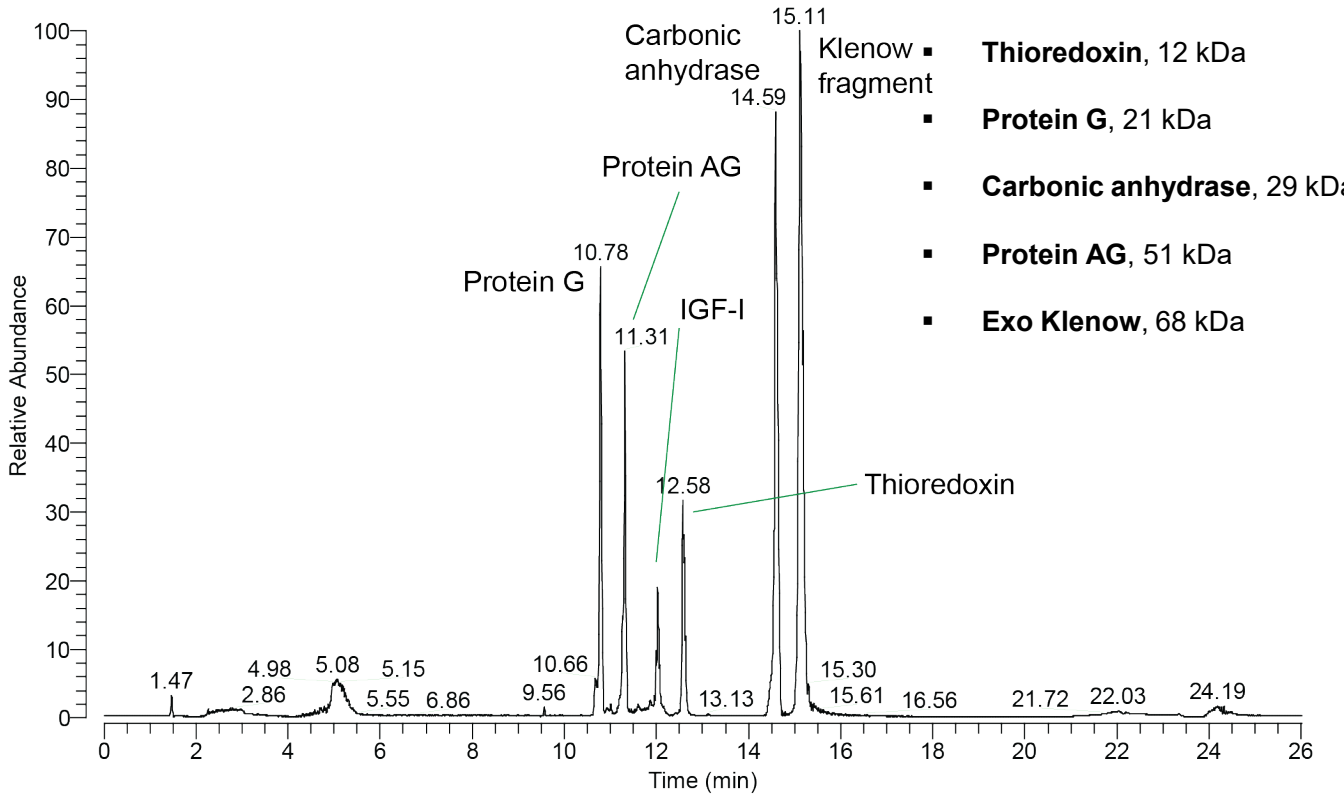
- The study will be **divided into two Phases**:

- **Phase 1**: targeted MS^2 analysis of a commercial six protein standard

  - *analysis to be performed as LC-MS^2*
  - *various ion activation techniques*

- **Phase 2**: high-throughput analysis of a complex proteoform mixture (*E. coli*) using:

  - *set of parameters for "optimal proteoform characterization"*
  - *set of parameters for "speed"*



- Pierce Intact Protein Standard (Thermo Scientific)
  - **IGF-I**, 9 kDa
  - **Thioredoxin**, 12 kDa
  - **Protein G**, 21 kDa
  - **Carbonic anhydrase**, 29 kDa
  - **Protein AG**, 51 kDa
  - **Exo Klenow**, 68 kDa

Phase 1: **Data analysis**

CTDP
Consortium for Top-Down Proteomics

- Data analysis will be **centralized** (files uploaded online):

1. **Conversion:** from proprietary to mzML (or mzXML) format
2. **Deconvolution**: FLASHDeconv (Kyowon Jeong)
3. **Fragment matching**: ProSight Lite (Ryan Fellers + Joseph Greer)

- Deconv parameters will be optimized for each instrument platform (using preliminary data)

Phase 1: **Results & Output**

CTDP
Consortium for Top-Down Proteomics

- **Sequence coverage** (fraction of identified backbone cleavages over the total)
- **p-score**
- **Number of matched fragments** and **percentage of matched fragments** over total

- Definition of **two sets of parameters**:

- Parameters that, while still compatible with LC time scale, provide the best protein characterization at the expense of speed – "targeted top-down MS" parameters.

- Parameters that, sacrificing sequencing performance, optimize data acquisition cycle and keep p-score good enough – "high-throughput top-down proteomics" parameters.

Participants, vendors & instruments

CTDP
Consortium for Top-Down Proteomics

- 21 T FT-ICR

- Project participant: **Lissa Anderson**, High Magnetic Field National Laboratory

CID, ETD

NATIONAL HIGH MAGNETIC FIELD LABORATORY

Participants, vendors & instruments

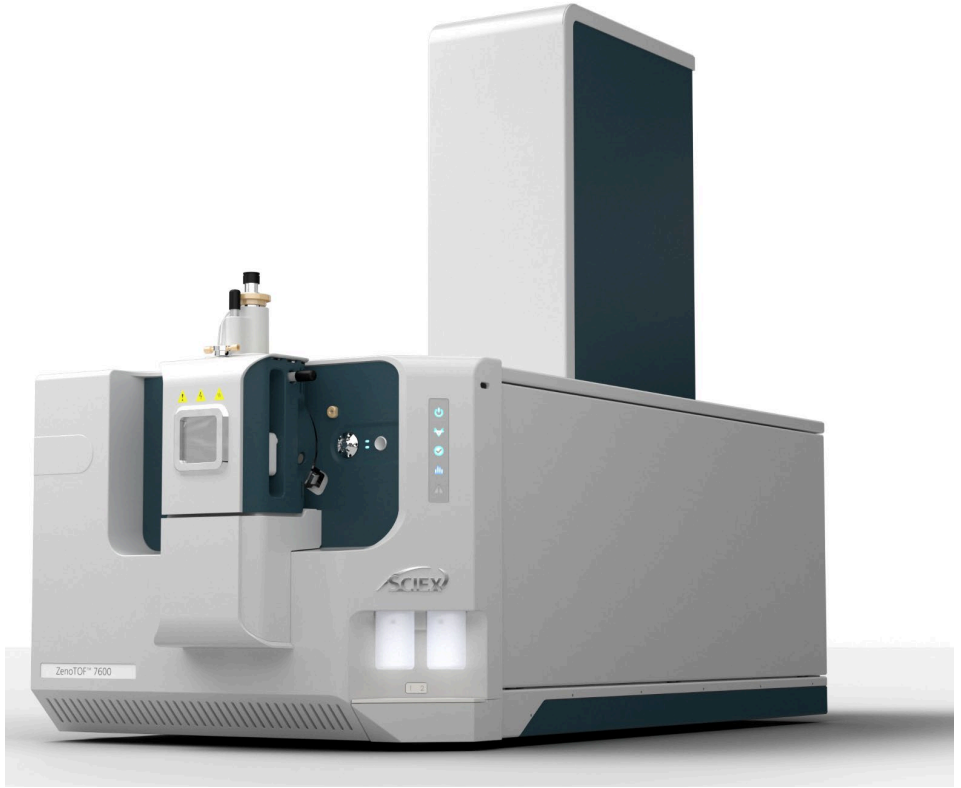
CTDP
Consortium for Top-Down Proteomics

- ZenoTOF 7600

- Vendor contact: Sahana Mollah

- Project participant: **Rick Searfoss**, Washington University St. Louis

CAD, EAD

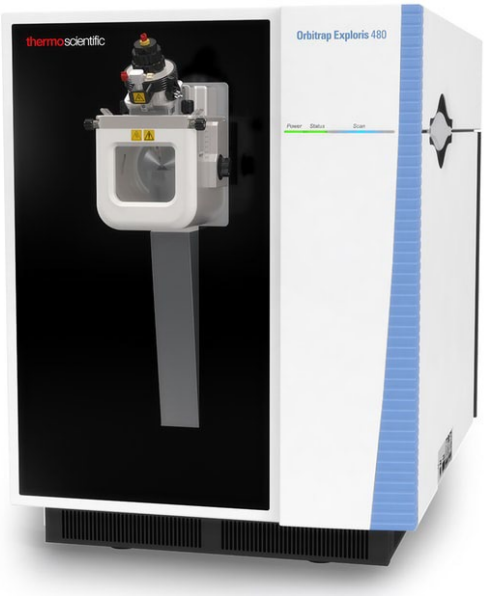


SCIEX

Participants, vendors & instruments

CTDP
Consortium for Top-Down Proteomics

- Exploris 480



- Vendor contact: Rafael Melani

- Project participant: **Grace Scheidemantle**,
  Frederick National Laboratory for Cancer Research

HCD

thermo scientific

Participants, vendors & instruments

CTDP
Consortium for Top-Down Proteomics

- Orbitrap Fusion Lumos Tribrid

- Vendor contact: Rafael Melani

- Project participant: **Robert D'Ippolito**, Frederick National Laboratory for Cancer Research

HCD, CID, ETD

thermo scientific

Participants, vendors & instruments

CTDP
Consortium for Top-Down Proteomics

- Orbitrap Fusion Lumos Tribrid

- Vendor contact: Rafael Melani

- Project participant: **Philipp Kaulich**, Christian-Albrechts-Universität zu Kiel

HCD, ETD, UVPD, EThcD

thermo scientific

Participants, vendors & instruments

CTDP
Consortium for Top-Down Proteomics

- Orbitrap EclipseTribrid

- Vendor contact: Rafael Melani
- Project participant: **Amal Eltobshi**, University of Oklahoma

HCD, CID, ETD, UVPD

thermo scientific

Participants, vendors & instruments

CTDP
Consortium for Top-Down Proteomics

- Modified TimsTOF Pro

- Vendor contact: Melvin Park

- Project participant: **Fanny Caroline Liu**, Florida State University

CAD

BRUKER

Participants, vendors & instruments

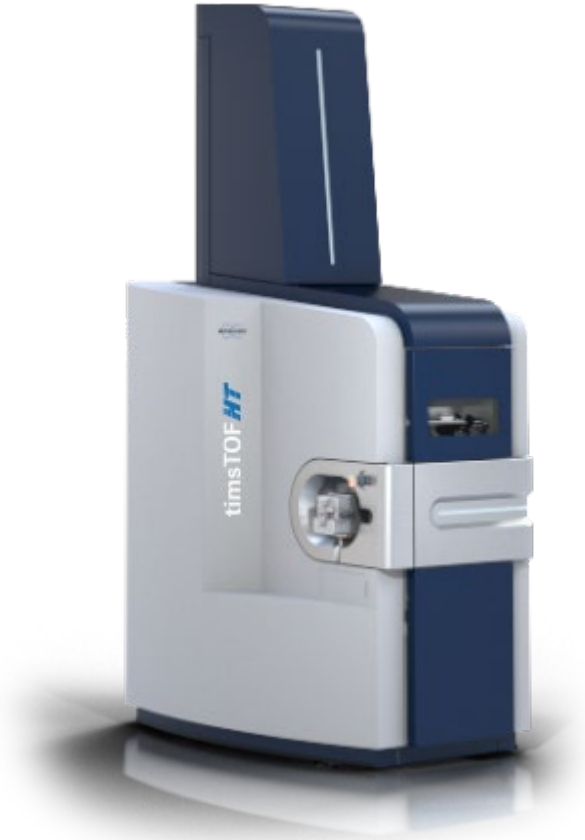
CTDP
Consortium for Top-Down Proteomics

- TimsTOF HT

- Vendor contact: Melvin Park

- Project participant: **Bryon Drown**, Purdue University

CAD



BRUKER

Participants, vendors & instruments

CTDP
Consortium for Top-Down Proteomics

- 6545XT AdvanceBio LC/Q-TOF

- Vendor contact: Rachel Franklin

- Project participant: **Carter Lantz**, Texas A&M University

CAD, ECD

Agilent

Participants, vendors & instruments

CTDP
Consortium for Top-Down Proteomics

- SELECT SERIES MRT

- Vendor contact: Brad Williams

- Project participant: **TBD**

Waters™

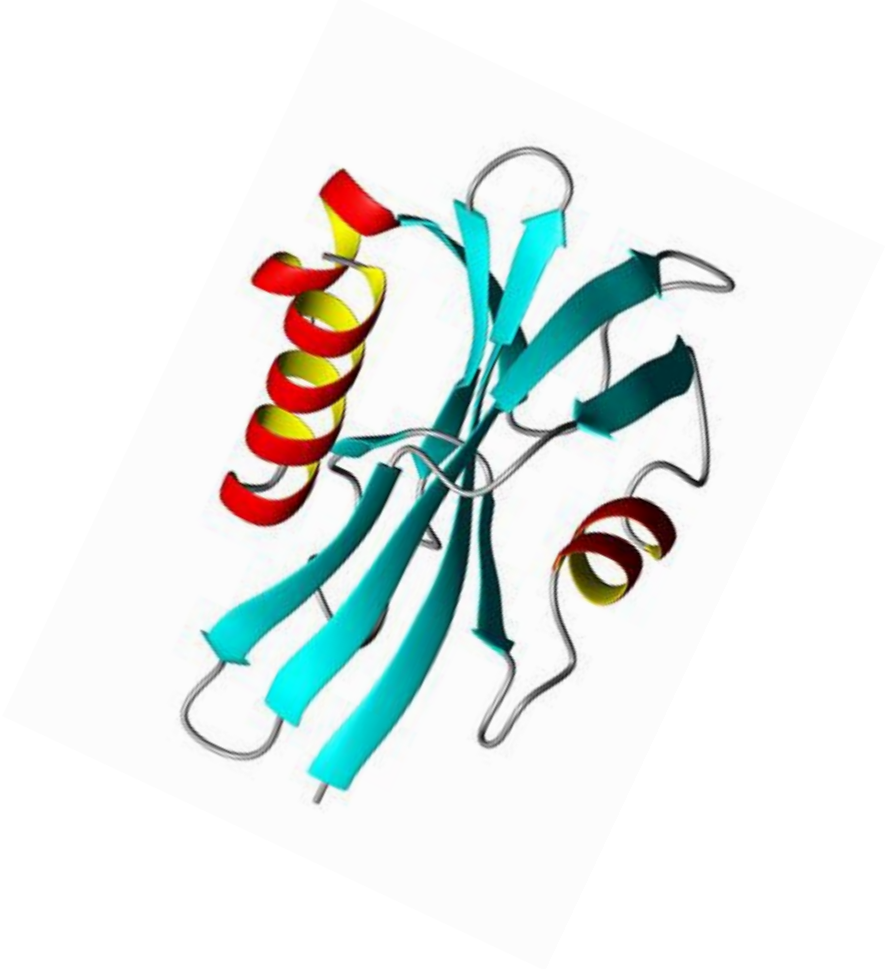
Updated timeline and milestones

- Update to community at Top-Down workshop @ ASMS 2024 → about 50% data collected, remaining data: post-ASMS to first 2 weeks of July.
- **Phase 1** to be completed by August 2024 (plus some time for data analysis/polishing).
- **Meeting** to discuss Phase 1 results by October 2024 (to be organized by CTDP).
- **Phase 2** starting in July/August 2024.

Contact information

CTDP
Consortium for Top-Down Proteomics

- Lissa: anderson@magnet.fsu.edu
- Caroline: caroline.dehart@nih.gov
- Luca: luca.fornelli@ou.edu
- Kyowon: kyowon.jeong@uni-tuebingen.de
- Ryan: ryan.fellers@northwestern.edu
- Joe: joe.greer@northwestern.edu
- Adrian: adrian.guthals@agilent.com

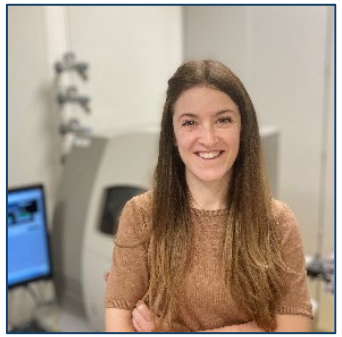


Panel Discussions

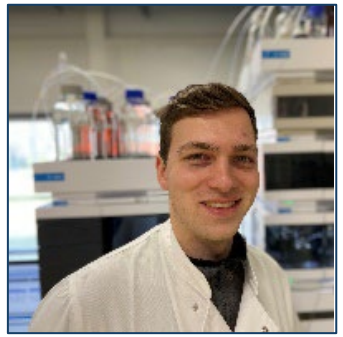
Clinical Research

Julia Chamot-Rooke
(Institut Pasteur)

Bryon Drown
(Purdue University)



Elena Domínguez-Vega
(Leiden University)



Christoph Gstöttner
(Leiden University)

Biopharmaceutical Industry

Pavel Bondarenko
(Amgen)

Lauren Adams
(Merck)

Jake Melby
(AstraZeneca)

Weijing Liu
(Thermo Fisher Scientific)

17th Uppsala conference on Electron Capture and Transfer Dissociation

- July 15-17, 2024
- Technical University of Darmstadt
- organized by Frederik Lermyte.
- early-bird registration (only €80 for academics, later going up to €100)
- abstract submission are still open.

Confirmed speakers

UppCon 2024 will feature an excellent lineup of confirmed speakers, including

- Professor → **Jon Amster** (University of Georgia, USA)
- Professor → **Joseph Beckman** (Oregon State University, USA)
- Professor → **Kristina Håkansson** (University of Michigan, USA)
- Professor → **Albert Heck** (Utrecht University, NL)
- Professor → **Joseph Loo** (UCLA, USA)
- Professor → **Peter O'Connor** (University of Warwick, UK)
- Professor → **Brandon Ruotolo** (University of Michigan, USA)
- Professor → **Richard Scheltema** (University of Liverpool, UK)
- Professor → **Frank Sobott** (University of Leeds, UK)
- Professor → **Roman Zubarev** (Karolinska Institutet, SE)

Thank you for participating in this work

See you in the Top-Down Proteomics Workshop at ASMS 2025