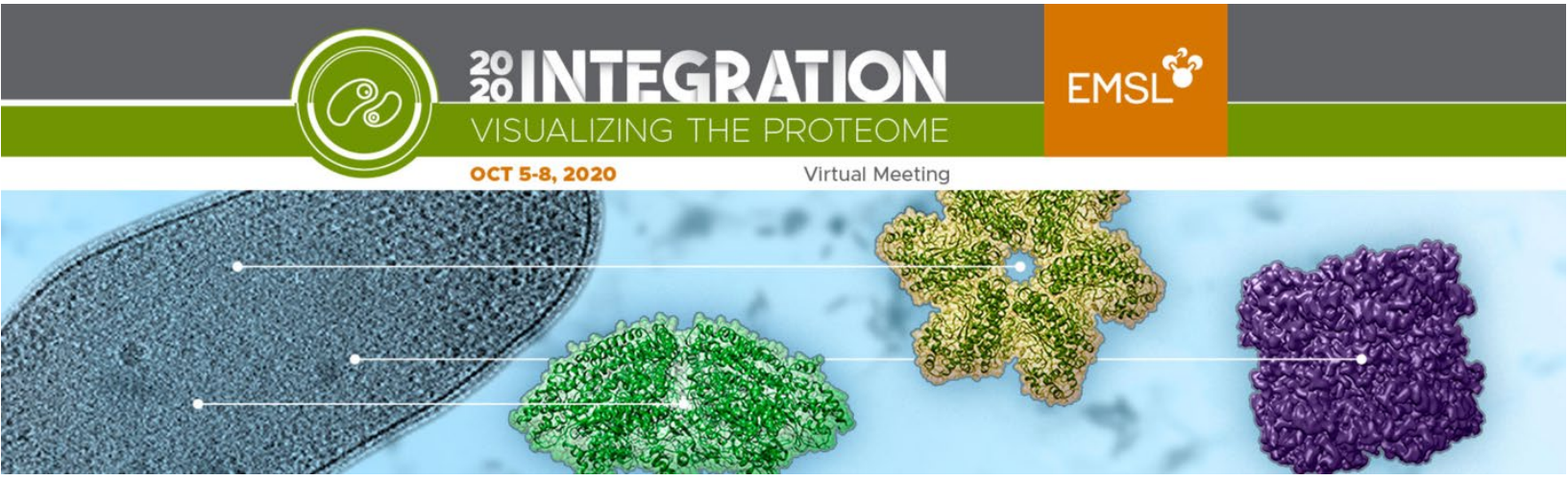
EMSL

2020 EMSL Integration Conference Report

Visualizing the Proteome

August 2021



PNNL-31802

U.S. DEPARTMENT OF ENERGY

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

2020 EMSL Integration Conference Report

Visualizing the Proteome

August 2021

Pacific Northwest National Laboratory
Richland, Washington 99354

Acronyms and Abbreviations

3-D	three-dimensional
APT	atom probe tomography
CCS	collisional cross-section
CE	capillary electrophoresis
cIEF	capillary isoelectric focusing
cryo-EM	cryogenic electron microscopy
cryo-ET	cryo-electron tomography
DP	dirigent protein
EMSL	Environmental Molecular Sciences Laboratory
HPC	high-performance computing
IM	ion mobility
MS	mass spectrometry
NanoPOTS	nanodroplet processing in one pot for trace samples
PIR	protein interaction reporter
PTM	post-translational modification
RSV	respiratory syncytial virus
SID	surface-induced dissociation
TEM	transmission electron microscopy

Contents

Figures

1.0 Introduction

The 2020 EMSL Integration Conference, titled “Visualizing the Proteome,” took place virtually, October 5–8, 2020. The switch from in-person to virtual conference was required because of the SARS-COV-2 pandemic. This was the first integration conference completely open to public and free, bringing together over 300 registered attendees from all over the world. EMSL Director Douglas Mans delivered a brief presentation on the first day to welcome everyone to the tenth annual integration conference while touching on how imaging and systems biology were part of the last year’s 201 active EMSL projects involving 685 users.

The format of the conference included morning science sessions, where people gave invited and contributed talks about state-of-the-art developments in 1) pushing the frontier of structural biology, 2) understanding subcellular compartmentalization and ultrastructure, 3) annotating protein function and modifications via native and top-down proteomics, 4) revealing cellular complexity and dynamics with single- to few-cell proteomics, and 5) advancing cellular models and protein design with high-performance computing (HPC). After the science sessions, EMSL scientists gave tutorials that provided insight into new capabilities and methods available via the EMSL user program. A virtual poster session, a meeting with the EMSL User Executive Committee, and an M.T. Thomas Award presentation also occurred during the four-day conference.

EMSL scientists and staff, along with support staff, organized the meeting and hosted the scientific community:

James Evans, Chemist, meeting co-organizer, session chair
Ljiljana Paša-Tolić, PNNL Laboratory Fellow, Chemist, meeting co-organizer
Mowei Zhou, Chemist, session chair
Trevor Moser, Chemist, session chair
Ying Zhu, Chemist, session chair
Aivett Bilbao Peña, Computational Scientist, session chair
Irina Novikova, Chemist, session chair
Eva Baroni, Project Coordinator
Mary McGown, Lead Administrator
Diane Showers, Administrator
Linda Isakson, Communications Partner
Corydon Ireland, Communications Professional
Beth Norris, Community Affairs Specialist

2.0 Science Sessions

2.1 Session 1: Pushing the frontier of structural biology with integrated approaches

Session Chair: James Evans

Day one started with a session on structural biology highlighting the use of cryogenic electron microscopy (cryo-EM), nuclear magnetic resonance, X-ray crystallography, and molecular dynamics simulations to solve the structure and mechanisms of macromolecular complexes. While this session included the most reductionist and highest-resolution interrogation of biological samples, it still demonstrated the power of multidisciplinary and multimodal approaches for addressing pressing biological challenges in bioimaging, health, and the environment.

"Gap junction channels visualized in a dynamic lipid environment"

In his talk, Steve Reichow of Portland State University described how gap junction channels allow neighboring cells to transmit metabolic and electrical signals through their coupled cytoplasms (Figure 1). This all happens within a dynamic local lipid environment that affects protein structure and stability. Reichow and coauthors described the underlying mechanism solved by combining cryo-EM with molecular dynamics simulations to study connexin-46/50 at a resolution just under two angstroms (Figure 1). For this class of gap junction channel, it was the highest-resolution imaging to date and showed both the protein structure as well as several bound lipids with enough resolution to be suitable for future structure-guided drug design.

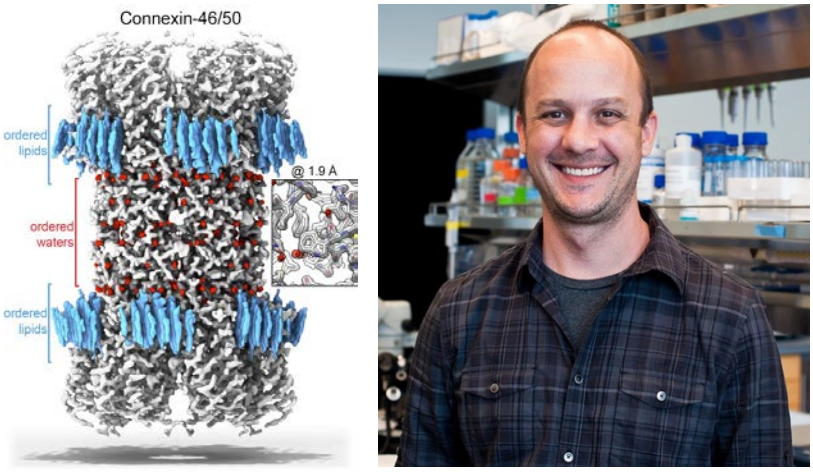


Figure 1. Molecular model of gap junction channel at 1.9Å resolution, showing protein in white, lipids in blue, and ordered water in red. Figure is courtesy of Steven Reichow, Portland State University.

"Structural basis for strand transfer inhibitor binding to HIV intasomes"

Dmitry Lyumkis of SALK Institute delivered a talk based on a recent paper in *Science* that continues his work on the structure of intasome—the viral machine that catalyzes the irreversible insertion of HIV DNA into target host chromatin (Figure 2). These macromolecular assemblies allow viruses, including HIV, to establish permanent infections in cells. Understanding intasomes will not only illuminate molecular pathways to infection, but also provide strategies of antiviral therapies for those infected with HIV.

Lyumkis and coauthors used high-resolution cryo-EM to decipher how a category of drugs, called integrase strand transfer inhibitors, bind to their drug target, HIV intasomes. To do so, they had to overcome challenges related to intasome biochemistry, which has been a hurdle for the past decade. Their work has provided mechanistic insights to improve drug binding and design. In particular, Lyumkis outlined progress on three developmental compounds that are candidate therapeutic agents and compared their biochemical, structural, and pharmacology features to the established drug Bictegravir.

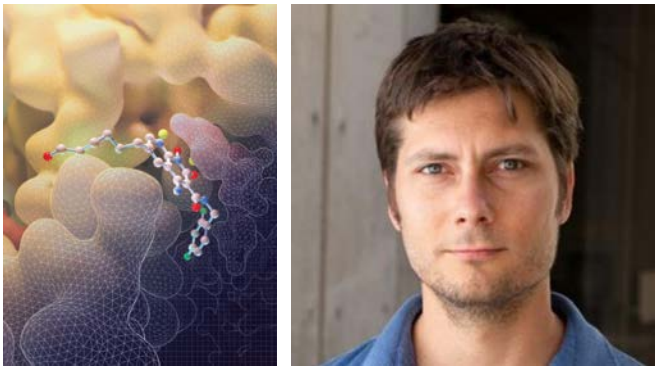


Figure 2. The image on the left depicts the molecular structure of an investigational compound that is one of the most potent Integrase Strand Transfer Inhibitors in development and how it binds to the HIV intasome. The structural biology data is being leveraged for structure-based drug design, in particular to combat the emergence of drug resistance. Figure is courtesy of Dmitry Lyumkis, SALK Institute.

"Pterocarpan synthase (PTS) structures suggest a common quinone methide-stabilizing function in dirigent proteins with dirigent-like domains"

Qingyan Meng of Washington State University brought plant proteins into the day-one conversation. She studies dirigent proteins (DPs), which eons ago emerged during the transition of plants from aquatic to terrestrial environments (Figure 3).

Importantly, DPs are entry points to distinct plant phenol classes, such as lignans, which may be useful in biofuel developments or other ways related to beneficial bioactivity. To date, however, only five to seven percent of the large multi-gene families related to DPs have identified biochemical functions.

Meng and her coauthors used biophysical technologies at EMSL and elsewhere to help unravel the architecture of plant protein complexes with DPs. Her work opens the door to systematically identify DPs and protein complexes that contain them, as well as decode their mechanism of action and explore how they can be harnessed to improve future bioproduction.

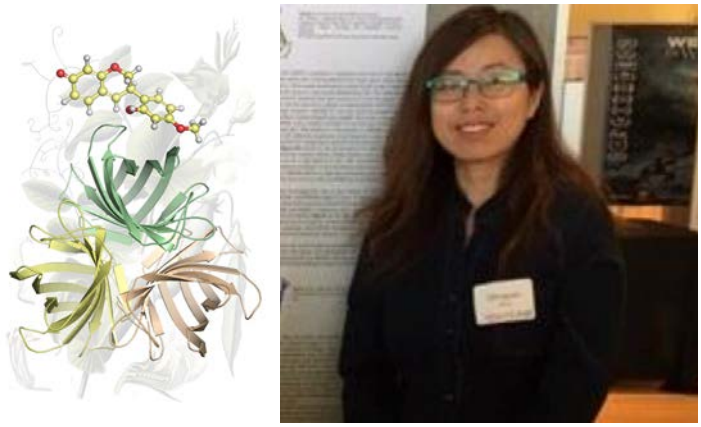


Figure 3. Biophysical technologies at EMSL and elsewhere helped unravel the architecture of plant dirigent protein complexes. Figure is courtesy of Qingyan Meng, a graduate research assistant at Washington State University.

"Vision for EMSL visual proteomics"

James Evans, an Integration 2020 co-organizer and scientist at EMSL, gave an overview talk that put EMSL in the context of all this important work. In particular, Evans pointed to the user facility's Structural Biology Integrated Research Platform, as well as expertise and instruments designed for imaging biological systems. Two new instruments this year got special attention: the Titan Krios G3i cryo-transmission electron microscopy (-TEM) and the Aquilos2 cryo-focused ion beam scanning electron microscopy. These two instruments can

provide atomic and nanoscale three-dimensional (3-D) structural imaging of individual proteins, protein complexes, organelles, whole cells, and tissue sections.

He addressed EMSL's efforts to overcome bottlenecks linking proteomics workflows to structural biology using a cell-free expression pipeline. Such integrated workflows can speed scientific discovery by minimizing the time needed to make suitable samples for structural analysis once a new target protein of interest is identified by proteomics approaches. He also highlighted the multimodal analysis of protein heterocomplexes using a combination of cryo-EM and native mass spectrometry (MS).

Evans ended his talk by giving an overview of the EMSL strategic objectives and decadal goals, including provision for enhanced visual proteomics, wherein researchers may combine many of the advanced capabilities housed within EMSL to create a unique and more holistic tool suite for mapping and identifying protein abundance, localization, and interactions within a single cell. The ability to characterize spatial organization of the proteome would expand future user science and start to refine and deepen our understanding of cellular organizations and molecular pathways to exploit new principles and enhance biodesign efforts for bioproduction or customized rhizosphere environments.

2.2 Session 2: Peering inside cells to understand subcellular compartmentalization and ultrastructure

Session Chair: Trevor Moser

Intricate biological assemblies, such as cells and tissues, are marked by a complex spatial heterogeneity. In this session, speakers covered a range of modalities, including cryo-electron tomography (cryo-ET), super-resolution optical microscopy, and X-ray imaging, all of which were used to characterize the structure, localization, or distribution of subcellular structures to elucidate their function.

"Recent developments in whole cell cryo-electron tomography"

Elizabeth Wright (Figure 4) of the University of Wisconsin-Madison started day two of Integration 2020 with a detailed explanation of sample preparation of encapsulated viruses and demonstrated how cryo-ET could be used to support vaccine development by investigating the relationship between an engineered genotype and resulting phenotypic changes.

Figure 4. Wright focused her talk on the preparation of enveloped viruses for cryo-EM and characterization of respiratory syncytial virus (RSV) structure and assembly using cryo-ET.

She explained that enveloped viruses are difficult to study with cryo-EM because they bud from the host cell and are intrinsically pleomorphic. This makes them challenging for single-particle cryo-EM techniques, as the viral particle population is heterogeneous. Wright explained that cryo-ET is, however, ideal for studying pleomorphic structures, as increased volume information can be gained from a single sample rather than just a single projection.

Wright described the workflow and challenges related to purifying and preparing the isolated viral particles for cryo-ET. Because the viral particles must be isolated from the host cells, other impurities are often present on the grids, such as cellular debris and vesicles. She demonstrated how affinity-based capture of the viral particles improved particle distribution and reduced sample impurities across the grid. While this strategy improved the success of imaging individual particles, Wright noted uncertainty whether the observed viral morphology is representative of what is extruded from a host cell or a product of purification artifacts. She further explained that, because the viruses are purified from whole cells, there is no way to know

which particles are fully mature and extruded versus particles that are still developing internally within the host cell.

Wright then explained how whole-cell cryo-ET on infected host cells was performed to visualize the native structure of extruded RSV viral particles. Using a fluorescently tagged viral genome to label the virus, and immunolabeling the viral particles at the surface of the host cell, her group was able to perform correlative light and electron microscopy of RSV-infected cells. Using fluorescence to target cells that were replication-competent for the virus, cryo-ET confirmed the filamentous structure of released RSV particles.

Wright also showed evidence that codon deoptimization of the "G" glycoprotein resulted in a hexagonal packing at the viral surface that was not present in the native virus, while heat inactivation resulted in a change from filamentous to spherical morphology and conformational change of the "F" glycoprotein. Overall, her presentation illustrated how cryo-ET could complement vaccine development by showing how various strategies for live attenuation of viral particles affected the overall morphology of the virus.

"Dissecting a molecular machine by cryo-electron tomography"

Jun Liu (Figure 5) of the Yale University School of Medicine discussed using cryo-ET to characterize the flagellar motor machinery of the spirochete *Borrelia burgdoferi*. He explained that the motility of many microorganisms, including pathogens such as *B. burgdoferi* (the organism causing Lyme disease), is directly connected to the microorganisms' virulence. The relatively small cell size of *B. burgdoferi* makes it an ideal organism to study by cryo-ET.

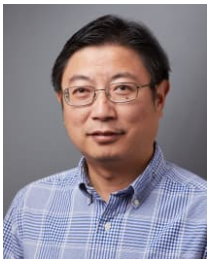


Figure 5. Liu told the audience how he and his team are using cryo-ET to gain structural insights into fundamental biological processes.

Liu described the unique structure and motility of spirochetes like *B. burgdoferi*, detailing their spiral cell shape and the periplasmic flagella running the length of the cell. These flagella rotate, causing translocation, which is unique in the bacterial world because their appendages are inside the cell, unlike many flagella with appendages outside the cell. The motor that drives these flagella is the type III secretion system, consisting of a membrane-bound export apparatus and a cytosolic ATPase complex functioning as a stator, and a C-ring that attaches to the stator and functions as the rotor. Using cryo-ET and subtomogram averaging the resulting volumes of the flagellar-specific type III secretion system (fT3SS), Liu and his team were able to resolve the structure of the 80-nm complex. While cryo-ET can provide some information, the technique cannot provide stochiometric, functional, or identity information as to which of the nine protein subunits make up a particular region of density in the final density map determined by cryo-ET and what their particular functions are in the complex.

To solve this problem, Liu described how his team performed knock-down mutations of stator proteins, MotA and MotB, resulting in the loss in motility in these cells. When comparing the sub-volume averages of the wild type fT3SS with the MotA and MotB mutants, a loss in density was detected in the transmembrane region at the edges of the stator, suggesting that these densities must correspond to the location of MotA and MotB in the assembled complex. Additionally, by creating mutants with altered CheY functionality, his team was able to generate structures of the fT3SS motors locked in either clockwise or counter-clockwise rotation structures and identify the mechanism of the conformational switching. Using this methodology, Liu and his team were able to infer the identities and locations of the other proteins in the C-ring complex, along with their stoichiometry.

"Leveraging super high optical resolution microscopy to probe the interaction zone between Clostridium thermocellum and biomass"

In his talk, John Yarbrough (Figure 6) of the National Renewable Energy Laboratory in Colorado described how he uses super-resolution optical microscopy to probe interfacial boundaries between *Clostridium, thermocellum,* and carbon substrate.

Figure 6. Yarbrough discussed an approach to study, close up, the deconstruction of biomass by *C. thermocellum* and better understand the critical region of enzyme-microbe-substrate interactions.

Yarbrough talked about the anaerobic organism *C. thermocellum* and its importance in biomass degradation. It currently has the highest known rate of cellulosic sugar solubilization, exceeding the performance of fungal systems. A key structure of *C. thermocellum* is the cellulosome, a multi-enzyme complex known to play a role in the digestion of cellulose in conjunction with *C. thermocellum*. Yarbrough noted that early EM images showed interfacial regions between *C. thermocellum* and cellulosic substrates, exposing density between the organism and the substate hypothesized to be a complex cellulosome matrix. By generating photoactivated green fluorescent protein and an Alexa Fluor tag that binds the cellosomes and cell wall with high specificity, Yarbrough demonstrated that the surface of *C. thermocellum* was coated with polycellulosomal protuberances.

By imaging the fluorescently tagged *C. thermocellum* growing on the cellulosic substrate Avicel, Yarbrough showed that the cellulosome complexes are associated with the substrate, adjacent and external to the organism that secreted them. He was able to visualize the spatial location of individual fluorescent events independent of collection parameters, such as intensity and brightness, using a density-based spatial clustering of applications with noise to filter out background noise from the fluorescent signal. This showed an increase in cellulosome clustering at the cell-substrate interface. Additionally, he confirmed the observation of changes in cellulosome clustering on the cell surface between different growth phases, indicating that the concentration and rate of cellulosome secretion are dependent on cell cycle.

"Endophyte-promoted phosphorous solubilization in Populus"

Tamas Varga (Figure 7) of EMSL described how his team, in close collaboration with Sharon Doty of the University of Washington, used X-ray tomography, X-ray microfluorescence, absorption spectroscopy, and multiomics to characterize endophytic phosphorus uptake in poplar.

Phosphorus is a critical nutrient for plants, but in soil, it is commonly bound in a metal complex, such as calcium phosphate, aluminum phosphate, or iron phosphate. These phosphorus metal complexes are largely unavailable to plants that require solubilized phosphate for uptake. Varga explained endophytes, which grow on the surface of plant roots, solubilize phosphates to make them available to the plants, but these inter-organismal relationships and communication channels remain largely unexplored.

Figure 7. Varga discussed the use of multiple techniques to study poplar plants and characterize how they absorb endophytic phosphorus.

Varga demonstrated that the endophyte strains used in their experiments could solubilize calcium phosphate and, to a lesser extent, aluminum phosphate and iron phosphate. When comparing the growth of poplar plants with or without endophytes, they noted that root length and leaf mass were greater in plants grown in soil inoculated with endophytes, although the changes were not significant. X-ray tomography, however, revealed that root volume was increased in endophyte inoculated plants, while root surface area was increased in non-inoculated plants, indicating increased finer root growth. By performing X-ray microfluorescence on thin sections of poplar roots to generate elemental maps of chemical distributions internal to the roots, they noted phosphorus hot spots within the endophyte inoculated roots that corresponded to calcium hotspots, indicating the uptake of calcium phosphate by the roots. In total, the chemical maps revealed phosphorus concentration in endophyte inoculated plants was consistently 20 to 30 percent higher than in the non-inoculated plants. The micro-X-ray absorption near-edge structure spectroscopy revealed that the phosphorus inside the roots was a mix of organic and inorganic phosphorus, whereas the inorganic phosphorus was mostly calcium phosphate. Finally, proteomics showed an up-regulation of phosphorus-related proteins linked to cell replication and phosphate transport in the endophyte inoculated root cells, while the non-inoculated roots showed increased levels of proteins involved in phosphate sensing and oxidative stress. In all, the results illustrated that microbes are available to plants to manipulate external soil chemistry.

2.3 Session 3: Revealing cellular complexity and dynamics with single- to few-cell proteomics

Session Chair: Ying Zhu

Biological systems are highly complex, containing diverse cell types and exhibiting tremendous cell-to-cell heterogeneity that dictates a multitude of functions in response to the physiological environment and external stimulations. Molecular measurements at the single-cell level can reveal mechanisms and unique features of cell populations, differentiation, impacts of microenvironments, and rare cells in complex cellular systems that are typically hidden in bulk cell measurements. While the rapid development and adoption of single-cell RNA sequencing have significantly advanced our knowledge on cellular heterogeneity, RNA transcripts are frequently recognized as unreliable indicators of cell function. The ultimate goal of the single-cell analysis is to directly measure the functional machinery of cells: the protein molecules and their native structures.

Speakers in this session covered key technology platforms for single-cell or near single-cell proteomics, including cryo-EM, MS, capillary electrophoresis (CE), and microfluidics. These talks also included the applications of single-cell proteomics to study protein-protein interactions, neuron diseases, poplar tolerance to abiotic stress, and embryonic development.

"Microfluidic sample preparation for EM Protein purification, single-cell analysis, and visual proteomics"

Thomas Braun (Figure 8) of the University of Basel, presented a couple of microfluidic technologies developed to enhance sample preparation for cryo-EM. He pointed out that the conventional methods are time-consuming, involve harsh treatments, and rely on a trial-and-error process. These methods are also largely inefficient and lead to 99.99 percent sample loss. To address these challenges, Braun's team developed a cryowriter, a robotic-controlled microfluidic capillary probe that can sample a very small amount of protein (nanoliter-scale), perform negative staining, and purify proteins for cryo-EM analysis. A unique advantage of cryowriter is its ability to significantly simplify the overall workflow for single-particle cryo-EM. To accomplish this, Braun's team developed a photo-cleavable affinity purification approach for protein enrichment, washing, and direct deposition to EM grids. The process was fully automated and generated reproducible results. Because of the improved sample recovery, the cryowriter enabled single-cell visual proteomics. The team applied a capillary probe to directly extract protein components from live single cells, perform negative staining, and deposit proteins on EM grids. By directly counting individual protein particles from single cells, they demonstrated that visual proteomics could differentiate single cells exposed to heat shock.

Figure 8. Braun illustrated how miniaturization using microfluidics during EM protein sample preparation promises to enable new applications for high-resolution structural analyses of proteins and for visual proteomics.

"Digging deeper into the proteome of single cells"

Ryan Kelly (Figure 9) from Brigham Young University gave a talk on how to dig deeper into the proteome of single cells by combining microfluidics with ultrasensitive liquid chromatography-MS. MS-based single-cell proteomics have the potential to transform biology and biomedical research. However, the measurement sensitivity is limited by the low sample recovery and low processing efficiency during sample preparation. He highlighted the great advantages of microfluidic nanodroplet technology for single-cell preparation, including increased trypsin digestion efficiency, reduced sample loss, and high robustness. When Kelly was a scientist in EMSL, he led the development of the nanoPOTS (nanodroplet processing in one pot for trace samples) platform for single-cell proteomics. The technology achieved approximately 500 times improvement in sensitivity compared with contemporary technologies. NanoPOTS also achieved the deepest proteome coverage of more than 1,000 proteins from single HeLa cells employing ion mobility filtering before MS analysis. To facilitate dissemination of nanoPOTS technology, Kelly's lab developed an autoPOTS system that uses microwell plates, an OT-2 liquid handler, and a modified liquid chromatography autosampler to enable fully automated proteomic sample preparation for single cells. The team further applied the platform to study single motor neurons and interneurons isolated from human brain tissues. They identified significant enrichment of high-affinity choline transporter proteins in motor neurons, illustrating the power of near single-cell measurements to track native spatial organization and map protein expression in tissues.

Figure 9. Kelly discussed the use of nanoPOTS for single-cell proteomics.

"Single-cell mass spectrometry goes embryonic"

Figure 10. Nemes discussed single-cell proteomics integrated with single-cell transcriptomics to better understand systems biology. Screenshot by Shalonne Luke, PNNL.

Peter Nemes (Figure 10) of the University of Maryland delivered a talk on the design and use of his lab's custom-built capillary electrophoresis mass spectrometry (CE-MS) technologies facilitating proteomic characterization of single differentiating cells in live *X. laevis* (frog) embryos. He emphasized that single-cell transcriptomics measurement cannot reflect the functional phenotype of the embryonic development process. Meanwhile, proteomics can give a more accurate spatiotemporally resolved picture of the fast biological process. Using their sensitive CE-MS system, they were able to identify 1,709 proteins from six single blastomeres at the eight-cell stage. Quantitative analysis revealed dorsal-ventral and animal-vegetal asymmetry. Next, to understand the dynamic proteome change during the development, the team developed a microprobe sampling technology. They can use the probe to directly sample nanoliter (10 to 15 nL) content from a live blastomere without killing it. After sampling, the blastomere can continue to differentiate. The new technology enabled them to study the neural tissue-fated dorsal lineage by repeating the sampling from the same embryo. Nemes concluded that single-cell proteomics integrated with single-cell transcriptomics raises new opportunities to better understand systems biology and the cellular mechanisms that coordinate normal and impaired development.

"Cell type specific proteomic responses to single and multiple abiotic stresses in poplar"

Figure 11. Balasubramanian shared how he and collaborators use EMSL's omics platforms to understand abiotic stress responses in various cell types of poplar trees. Photo by Andrea Starr, PNNL.

EMSL postdoc Vimal Kumar Balasubramanian (Figure 11), an expert in molecular plant phenotyping, focused his talk on understanding the plant responses to environmental stresses that occur at specific levels of single-cell-types in the leaf and root tissue of poplar trees. Abiotic stresses, such as high temperatures and a lack of fresh water, can cause reduced crop growth or even complete loss. Balasubramanian conducts research funded by the Department of Energy's (DOE) Office of Science Biological and Environmental Research program to identify stress-tolerant genes and promoters in poplar. Researchers can then use this knowledge to develop transgenic poplar trees that are tolerant to multiple stresses. Balasubramanian uses EMSL nanoPOTS technology to study the proteome regulations of poplar trees under drought, salinity, heat, and the combination of all three stresses. The team targeted distinct poplar tissues and cell types, including leaf mesophyll and xylem/phloem (vascular) cells, using cryo-sectioning and laser-capture microdissection techniques. EMSL's unique omics capabilities unraveled the abiotic stress-responsive genes and proteins in the different cell types of poplar leaf and root tissues. For example, cell-type-specific proteomic analysis revealed proteins involved in amino acid and carbohydrate metabolism, photosynthesis, sucrose biosynthesis/degradation, secondary metabolite biosynthesis, and antioxidant metabolism as stress-responsive proteins exclusively in palisade mesophyll or vascular cells during water-deficit, salt, heat, and combined stresses and in the recovery phase. These efforts contribute to better understanding the roles different poplar

cell types have in response to single and combined abiotic stresses, thus, facilitating the design of stress-tolerant poplar trees.

2.4 Session 4: Annotating protein function and modifications via native and top-down proteomics

Session Chair: Mowei Zhou

Most proteins are post-translationally modified and fold into specific conformations to carry out their function. They also interact with a variety of other molecules in biological systems. Genetic and conventional MS analysis cannot easily extract information regarding protein post-translational modifications (PTMs), higher-order structures, and molecular interactions. Significant technology development in the MS field has enabled new ways to study proteins. Top-down proteomics characterizes intact proteins, allowing the precise definition of "proteoforms"—functional forms of proteins bearing combinations of biologically relevant PTMs.

A single gene can make many different proteoforms because of versatile post-transcriptional and post-translational regulatory toolsets. These different proteoforms can assemble into protein complexes in specific ways to enable specific biological functions. Advanced MS instrumentation (including those at EMSL) can measure native protein complexes at significantly higher mass than standard mass spectrometers to enable native MS analysis revealing protein higher-order structures, protein-protein, protein-ligand, and protein-metal interactions. The presentations in this session featured five academic speakers who are experts in developing and applying novel MS techniques that could help the community understand protein structure and function.

"How unique is your plasma proteome? A personalized perspective"

The session's first plenary speaker, Albert Heck from Utrecht University and a well-known pioneer in the MS field, gave an overview of the recent advancements in MS instrumentation and its use for precision medicine (Figure 12). The ultra-high mass range Orbitrap technology from ThermoFisher Scientific plays a central role in enabling direct analysis of intact protein and protein complexes. Another exciting technology, charge detection MS, also implemented in Orbitrap systems, is a single-particle detection method that significantly improves sensitivity and resolution for top-down and native proteomics. This technique, combined with high mass resolution, has detected macromolecules as large as intact viruses and Immunoglobulin M.

Heck's recent work on human plasma proteome highlighted the complexity of PTMs and their largely underexplored roles in protein function. Using multiple separation and purification steps, his team isolated several important human plasma glycoproteins from 20 donors, including people in healthy and sepsis conditions. Deep characterization using native and top-down MS revealed tens to hundreds of proteoforms derived from a single gene, each with different combinations of glycosylation and/or phosphorylation. Interestingly, the proteoform profiles correlated well with genotypes, implying highly "personalized" proteoform profiles that may not be easily identified using conventional analytical techniques. The glycosylation profiles were also shown to respond to disease states and affect their binding characteristics to other biomolecules. The latest research demonstrated the unique advantage of native and top-down MS in deciphering heterogeneity at the molecular level and the potential for applications in personalized medicine.

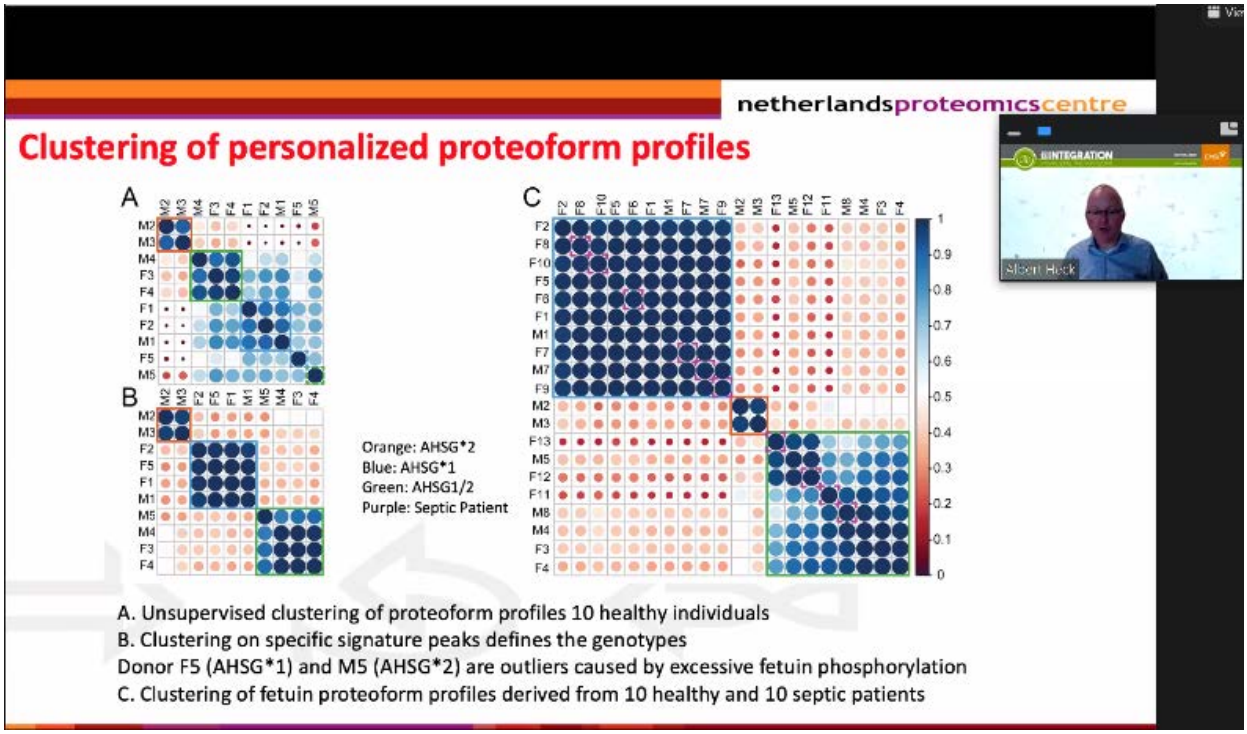


Figure 12. MS veteran Albert Heck discussed using direct analysis of intact proteins and protein complexes using advanced MS instrumentation. Screenshot by Shalonne Luke, PNNL.

"Towards protein structures from native mass spectrometry and single-particle imaging"

The second plenary speaker, Joseph Gault from the University of Oxford, was an Extraordinary Junior research fellow in chemistry at Queen's College in Dame Carol Robinson's group before transferring to an industry career at Vertex Pharmaceuticals in the United States (Figure 13). He discussed innovative studies at Oxford connecting native MS with single-particle imaging to push the frontiers of structural biology. The Nobel-Prize-winning technology, cryo-EM, like other single-particle imaging techniques, excels at taking structural snapshots of molecular machineries, facilitating a better understanding of biological processes. However, these high-resolution imaging techniques generally suffer from low-throughput, associated with sample heterogeneity resulting from dynamic structures, small molecule binding, etc. Native MS could fill many of the gaps and complement these analyses. For example, membrane protein complexes generally require very strict sample preparation conditions and are challenging targets for structural biology applications. The best conditions (e.g., choice of detergent) could be rapidly screened by native MS and transferred to cryo-EM analysis, while minimizing time and cost to maximize success.

Gault also proposed leveraging the power of chemical selectivity offered by native MS to improve the sample homogeneity, thereby improving the resolution attainable by single-particle imaging methods. The work included a novel strategy for preparing mass-selected proteins or complexes for imaging by depositing ions onto surfaces inside the mass spectrometer. This workflow opens new possibilities for using native MS to enrich and prepare low abundance proteins and protein assemblies from complex mixtures for cryo-TEM single-particle imaging. After meticulous improvements to the instrumentation, Gault and his colleagues were able to mass-isolate protein complexes, soft-land them on a surface, and image them using atomic force microscopy, transmission, and cryo-EM. Preliminary data suggested the soft-landed particles maintained the overall shape of their native structures. Future developments will allow MS-based selection and purification of

particles to significantly reduce sample heterogeneity for cryo-EM, which is a critical hurdle for high-resolution structural determination.

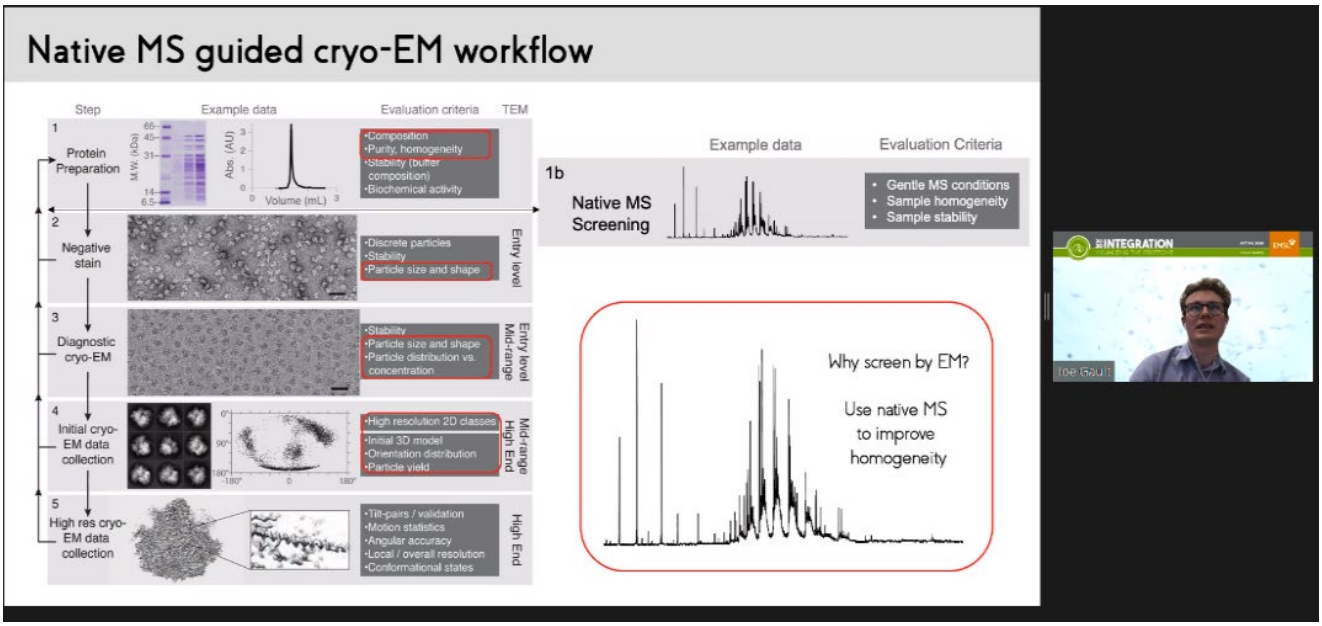


Figure 13. Gault spoke about combining native MS with single-particle imaging to understand biological processes better. Screenshot by Shalonne Luke, PNNL.

"Structural characterization of large and heterogenous B12 trafficking protein complexes by native ion mobility-mass spectrometry"

Varun Gadkari, a postdoctoral researcher at the University of Michigan, gave a contributed talk describing the use of ion mobility (IM)-MS to study native protein complexes (Figure 14). IM provides another dimension of separation and additional structural information to native MS measurements, separating heterogeneous species from one another, usually on the order of milliseconds (or possibly much longer).

Gadkari introduced large and heterogenous protein complexes related to the vitamin B12 (cobalamin) trafficking pathway. This pathway marks a complex intracellular route that delivers this water-soluble vitamin to target enzymes. B12 is needed for some vital human metabolic reactions and the prevention of certain medical complications, including neuropathies associated with the spinal cord. The proteins linked to the B12 trafficking pathway form a variety of complexes, which find equilibrium in solution. Using native MS techniques, Gadkari was able to resolve the individual protein complexes in the gas phase. He illustrated how this strategy could precisely assess protein composition, size, and stability of each species, instead of simply examining the whole ensemble of species. Gadkari's work further elucidated the unique power of native MS for defining the heterogeneity of proteins, greatly complementing other structural biology methods.

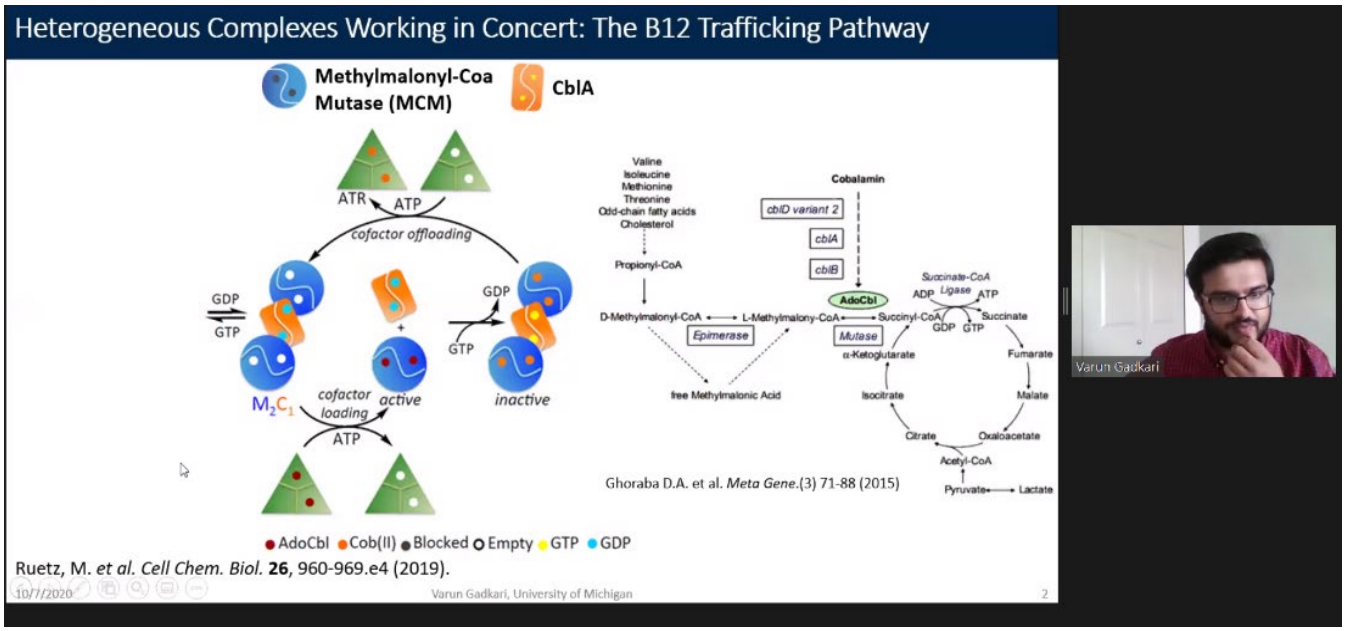


Figure 14. Varun Gadkari described advantages of ion mobility separations coupled with native MS for characterization of molecular machines. Screenshot by Andrea Starr, PNNL.

"Automated capillary isoelectric focusing-tandem mass spectrometry for qualitative and quantitative top-down proteomics"

Despite the novel research discussed by the previous speakers, native and top-down proteomics methods still suffer from lower sensitivity and coverage for complex samples when compared to the more conventional, bottom-up proteomics approach (i.e., digesting proteins into small peptides for MS detection). Better separation techniques will be essential for making these advanced MS technologies compatible with complex biological samples. Liangliang Sun (Figure 15) from Michigan State University gave a contributed a talk describing the advantages of capillary isoelectric focusing (cIEF) for top-down proteomics. cIEF was first used for protein MS in the 1990s but is now undergoing a research revival because of improved instrumentation. Sun explained how he uses top-down proteomics to identify and quantify proteoforms from complex biological matrices (e.g., cell lysates).

Figure 15. Sun shared the benefits of using high-resolution MS to characterize proteoforms in zebrafish brains. Screenshot by Andrea Star, PNNL.

Sun and his coworkers used high-resolution MS instrumentation to identify nearly 800 proteoforms in a single cIEF-MS analysis of *E. coli* lysate. He also quantitatively compared the proteoforms in male and female zebrafish brains and noted drastic differences that were hidden when using the more conventional bottom-up approach. Sun has done the first such comparison studies characterizing proteomes in a proteoform-specific manner. His group continues to collaborate with developmental biologists and bioinformatians to further the practicality of cIEF-MS technology for better characterization of functional proteoforms.

"Interactome studies in aging and heart failure"

Jim Bruce at the University of Washington gave the last presentation in the session. He uses a different approach from the other speakers to study protein-protein interactions related to aging and heart failure. Although conventional bottom-up proteomics workflows cannot capture higher-order structure information, chemical labeling strategies can introduce footprints on proteins to capture interactions. One common method is known as chemical crosslinking, whereby a specific crosslinker is introduced to link reactive amino acid residues on proteins (e.g., lysine, aspartic acid, glutamic acid, cysteine) in proximity *in vivo* or under native conditions *in vitro*. The crosslinker produces unique mass shifts that can be captured by MS and used to identify interacting partners.

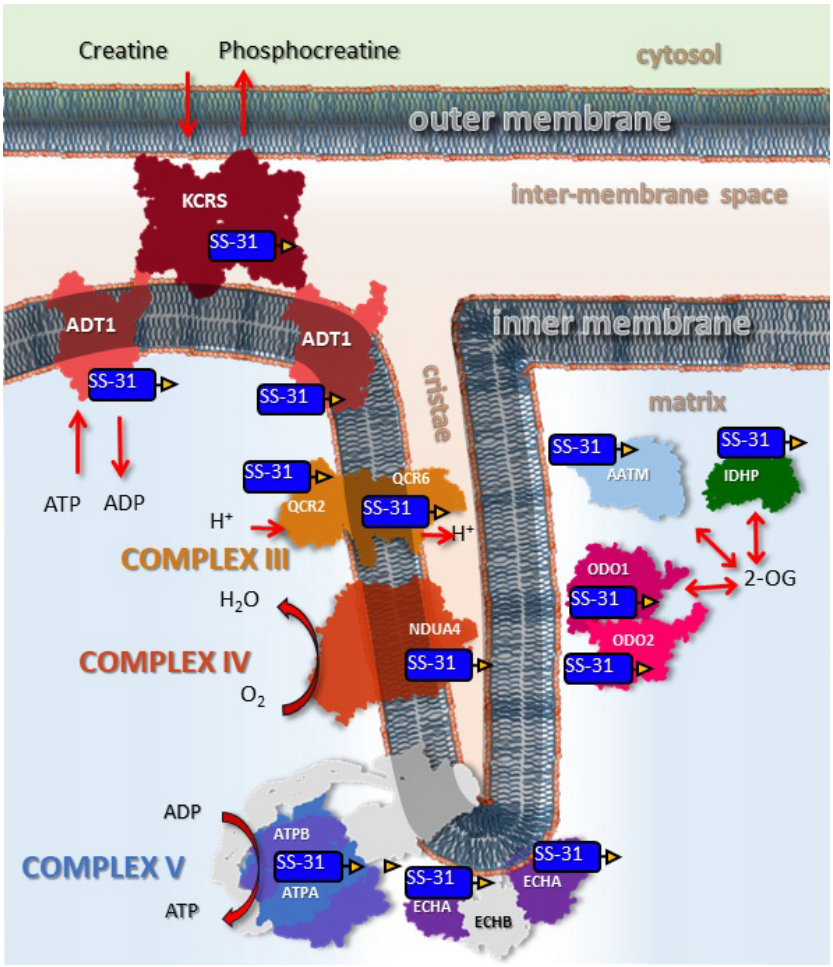


Figure 16. A graphical abstract of the talk by Bruce on PIR technologies developed in his lab. Image is courtesy of Jim Bruce, University of Washington.

Bruce's research focuses on developing *in vivo* crosslinking methods, which can capture protein-protein interactions in a cellular environment. However, such analysis has many technical challenges. Crosslinked peptides are usually at low intensity, and non-crosslinked peptides can easily mask their signal. His team developed the Protein Interaction Reporter (PIR) technology to address this challenge. A reporter moiety is incorporated into the crosslinker design to generate signature fragments in mass spectra to help fish out the crosslinked peptides (Figure 16). He applied PIR to provide insight on the mitochondrial interactome of heart tissue and helped guide the science to emerging mitochondrial therapies for heart disease and aging. The crosslinkers are cell-permeable and can capture the "interactome" in living cells without protein isolation, allowing insights into transient interaction that may not survive lysis conditions or extensive purification. Bruce also applied the PIR crosslinking technology to study structural changes in the proteome in response to small molecule binding (e.g., pharmaceuticals). The work with the synthetic tetrapeptide, elamipretide (SS-31), which improves mitochondrial function and undergoing clinical trials for heart failure treatment, has so far identified 12 mitochondrial protein targets that are helping explain some of the observed beneficial effects of this therapy, including reduced protein leak and improved ATP production.

2.5 Session 5: Advancing cellular models and protein design with scientific computing

Session Chair: Aivett Bilbao Peña, EMSL

Presentations in session five showcased a variety of research techniques for protein visualization, spatial organization, and cellular models supported by scientific computing. The advancements and multidisciplinary nature of the scientific computing field were discussed, along with their essential role in visual proteomics. Specific topics included studying protein compaction using IM-MS and molecular dynamics simulations, integrating electron microscopy and X-ray techniques with bioinformatics to characterize and ultimately engineer bacterial microcompartments, developing structural pattern mining algorithms in cryo-electron tomography, and applying machine learning to identify protein signatures in atom probe tomography (APT).

"Studying quantitative protein compaction, unfolding, and dissociation with ion mobility-mass spectrometry"

University of Oregon assistant professor James Prell (Figure 17) explained how IM-MS could help determine the biomolecular structure and energetics of native proteins and nanoscale macromolecular assemblies. IM-MS separates protein ions in the gas phase by their orientationally averaged collisional cross-section (CCS), reflecting their overall size and shape.

Figure 17. Prell described IM-MS approaches for characterizing proteins and nanoscale macromolecular assemblies.

Prell presented results from a systematic study of structural changes in well-characterized proteins induced by adjuvant gas-phase methods using molecular dynamics simulations, kinetic modeling, and other experiments. Experimental CCS with helium and nitrogen as a buffer gas showed that the proteins are compacted up to 22 percent in the gas phase compared to their crystal structures. Protein compaction is among the structural alterations that occur when proteins are transferred to a gas phase. Molecular dynamics simulations performed with the GROMACS software comparing five common force fields all resulted in compaction, although to varying extents. Comparing the average percentage difference in experimental versus simulated CCS, the GROMOS force fields outperformed the others evaluated, achieving average deviations from experiments that were less than five percent.

His group feels they have "a handle on compaction" and now investigates native ion unfolding and dissociation, including utilizing collision-induced unfolding, to measure multiple transitions or mobility jumps and provide quantitative information about the folding domains and unfolding energetics.

This study led to several important conclusions. First, molecular dynamics simulations can accurately model protein compaction, which occurs both at the surface and in the interior of proteins, but the choice of a force field is critical. Second, compaction increases slightly with charge but is compensated by increased ion-dipole interactions. Third, unfolding barrier thermodynamics of native proteins can be accurately measured with IM-MS, under gentle nano-electrospray ionization conditions, thus, facilitating analysis of many samples that do not easily crystallize or exhibit extreme heterogeneity. Fourth, Coulomb repulsion plays a large role in gas-phase unfolding.

"Elucidating gas-phase unfolding of protein complexes through steered molecular dynamics simulations"

Researcher Chae Kyung Jeon (Figure 18), a graduate student from the University of Michigan, delivered a talk on molecular dynamics simulations that shed light on the gas-phase unfolding of protein complexes. It is part of a larger research pursuit developing integrative models of gas-phase protein structures.

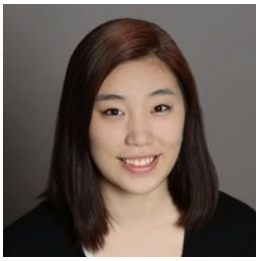


Figure 18. Jeon discussed molecular dynamics simulations in the pursuit of developing integrative models of gas-phase protein structures.

She discussed IM-MS instrumentation and the collision-induced unfolding technique. Her team aims to combine molecular dynamics simulations with IM-MS measurements of domain-deleted protein assemblies to ascertain the level of domain-correlated unfolding within multi-protein complexes. To learn more about unfolding pathways, they performed steered molecular dynamics, a modeling approach that allows the unfolding of specific protein regions.

So far, the researchers uncovered that steered molecular dynamics strongly support sub-domain unfolding; that with more experimental data, they can refine their models; and that different molecular dynamics approaches may be needed.

"Integrated Methods for Characterizing and Engineering Bacterial Microcompartments"

Cheryl Kerfeld (Figure 19), a professor of structural bioengineering at Michigan State University, highlighted the use of electron microscopy and X-ray techniques, such as small angle X-ray scattering and X-ray foot printing, combined with bioinformatics tools to understand organelle-like compartments in bacterial cells. The ultimate goal of these studies is to characterize and potentially engineer these functionally diverse, selectively permeable metabolic modules found across the bacterial kingdom.

Figure 19. Kerfeld discussed using integrative techniques to understand the organelles in bacterial cells.

This research has a unifying concept—modularity—a hierarchical nesting of function from protein domains to organelles to ecosystems, enabling a higher-order complexity and function to emerge from separate but combined evolutionary units. To conclude her presentation, she emphasized that, in part, breaking through to new frontiers will require integrating imaging—visualizations across scales, from the atomic nanoscale to the cellular mesoscale.

"Structural pattern mining in cryo-electron tomography"

Carnegie Mellon University computational biologist Min Xu (Figure 20) uses machine learning strategies linked to Cryo-ET to look for patterns of protein structures in macromolecules. The tomograms captured by Cryo-ET can yield 3-D density maps that represent the structure and spatial organization of single cells in near-native states and at sub-molecular resolutions. Samples are frozen quickly to avoid forming ice crystals, and two-dimensional projection images are acquired through rotational tilting. Data are reconstructed to generate 3-D volumes representative of the original structure.

Pattern mining algorithms can be applied to systematically detect macromolecular structures in cellular tomograms. Challenges with this kind of data analysis are related to missing data, low signal-to-noise ratio, highly diverse structural content, and the sheer volume of data. A single run can capture hundreds of tomographs, which contain millions of macromolecules.

Xu and his team developed and applied various deep learning techniques for addressing large-scale sub-tomogram classification and segmentation challenges. These techniques include domain adaptation and randomization, open-set learning, and few-shot learning. Another algorithm presented was based on autoencoders and K-means clustering for mining structural patterns, such as ribosome macromolecules and vesicular membranes. Xu also presented their unsupervised deep learning algorithms for subtomogram alignment.

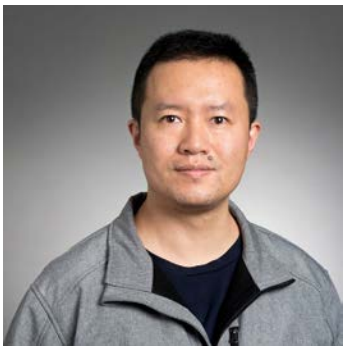


Figure 20. Xu presented research using cryo-ET to capture cells' 3-D electron density distribution at sub-molecular resolution and close-to-native state.

To facilitate large-scale analysis, this team developed AITom, an open-source artificial intelligence platform for cryo-ET. The framework includes new and previously published algorithms implemented in Python and C++ for performing cryo-ET data analysis through both the traditional template-based or template-free approaches, as well as deep learning.

He concluded his presentation by remarking that artificial intelligence facilitates understanding and analysis of tomograms and mentioned that AITom is an active software development project looking for collaborators.

"The application of machine learning for the identification of protein signatures from atom probe tomography data"

EMSL materials scientist Mark Wirth (Figure 21) described how the APT technique generates 3-D compositional maps of various materials. Samples are first sharpened to a needle-like tip less than 100 nanometers across and evaporated from the surface using a thermally assisted electric field. Data in the ATP 3-D maps contain the position of atoms and small molecular ions whose chemical identity is determined using time-of-flight mass spectrometry. Recent advances in the cryogenic preparation, handling, and APT analysis of hydrated materials have opened the possibility of APT analysis of hydrated soft biological samples. Challenges with the current implementation include molecular fragmentation and possible field evaporation aberrations during APT analysis, which complicate the identification and determination of protein structure.

Figure 21. Wirth discussed using APT with machine learning to determine protein structures from APT data.

The team is exploring and developing a machine-learning-based approach to determine protein structures from APT data. They developed and trained a neural network to combine Protein Databank Data and experimental APT data. The Protein Databank Data online database of protein structures experimentally determined by TEM, X-ray diffraction, or nuclear magnetic resonance techniques, was designed to simulate APT data for training data for the neural network. Given the input features represented as the distances to the seven nearest neighbors and their chemical identities for each ion, this initial model classified ions as belonging or not to the protein backbone. Application to APT experimental data of hematite (Fe_3O_4) nanoparticles embedded in resin showed that in 5.7 million ions, the model labeled fewer than 100 ions as protein backbone false positives.

Wirth concluded that their neural network could successfully identify protein fragments in resin. Future work includes using semantic segmentation and 3-D convolutional neural networks, boosting detector efficiency, and analyzing proteins embedded in non-crystalline water ice, and ultimately, progressing from APT data to protein structures.

3.0 Additional Activities

Beyond the science sessions, guests had additional opportunities to learn and participate. A brief account of some of their experiences follows.

3.1 M.T. Thomas Award

After the science session on Monday, October 5, 2020, EMSL Director Douglas Mans introduced Dušan Veličković, recipient of the 2019 M.T. Thomas Award for Outstanding Postdoctoral Achievement. This award recognized outstanding accomplishments by a postdoctoral research associate who performed at least a portion of their research at EMSL. EMSL's User Executive Committee served as the selection board. Veličković was the twenty-fourth scientist to receive this award since its inception in 1996.

The M.T. Thomas Award recognizes a postdoctoral researcher who has used EMSL resources within the last three years to develop a conceptual advance that has influenced or has the potential to influence their field. To be eligible for this award, the candidate must demonstrate a high level of scientific accomplishment, creativity, and problem-solving relevant to EMSL's mission, and Veličković exceeded these criteria. A native of Belgrade, Serbia, he earned a doctorate in analytical chemistry from the University of Belgrade in 2011, as the youngest PhD candidate in the Biochemistry Department. He started his outstanding postdoctoral career at EMSL in 2016 and became an EMSL scientist in 2020.

Veličković's research interest is focused on biological mass spectrometry, specifically matrix-assisted laser desorption/ionization or MALDI MS imaging. As a postdoc, he transformed EMSL's MS imaging capabilities and expanded the number of EMSL users interested in using MS imaging from a few to more than a half-dozen yearly. His efforts also led to an increase in external funding for MS imaging to support important biological and environmental research.

One of Veličković's many achievements was his research with EMSL user Gary Stacey at the University of Missouri to understand metabolic transactions occurring in the plant-microbe interaction zone. This work was highlighted in DOE's budget request to Congress for fiscal year 2020. Some of his other accomplishments as an EMSL postdoc included co-authoring 12 scientific publications in peer-reviewed journals (eight as first author), delivering eight conference presentations, supporting 10 user projects resulting in seven publications, winning two travel awards, and co-authoring three external and two internal proposals.

As the 2019 M.T. Thomas Award recipient, Veličković received a plaque recognizing him for his outstanding postdoctoral achievement. He also received a stipend of up to $2,000 to attend a domestic conference of his choice to serve as an EMSL ambassador and present his research using EMSL science capabilities.

3.2 User Executive Committee Forum

The EMSL User Executive Committee hosted a one-hour forum on Zoom with current, past, and possible future users of EMSL. The forum's purpose was intended to introduce the committee members to the users they represent and foster constructive dialog about the EMSL user program.

3.3 Poster Session

Twenty research posters, Appendix A, were hosted on the virtual web platform Discord that offered scientists the opportunity to learn more about their colleagues' research and spark conversations for future collaborations and networking.

3.4 Increased Opportunities for Discussions and Networking

Throughout the conference, the Discord web platform was open, and a channel for each science session let attendees ask presenters questions during and after each talk. These channels permitted longer and more detailed discussions than would have been possible with the short Q&A period between the end of one presentation and the start of the next. With attendees and presenters connected remotely over more than 12 time zones, the Discord channels made it possible for participants to be read the posters and engage with others any time throughout the conference.

3.5 How to Write a Successful Proposal Panel Discussion

In addition to the technical and science tutorials, a panel discussion was held on day four to guide future users on accessing the EMSL user program and the type of science problems addressed in proposals that lead to successfully awarded projects. EMSL's Deputy of User Services, Terry Law, presented an introductory overview; the panelists also included Nancy Hess (Environmental Transformations and Interactions Lead), Scott Baker (Functional and Systems Biology Lead), Scott Lea (Molecular Bioimaging [name later changed to Structural Biology] IRP Lead), and Lee Ann McCue (Chief Information Officer [title later changed to Chief Data and Analytics Officer]).

3.6 Tutorial: Single-Particle Cryo-Electron Microscopy

To complement the day one talks on single-particle imaging, EMSL's Trevor Moser gave a tutorial on using single-particle cryo-EM to determine protein structures. He discussed the steps in sample preparation, data collection, and data processing required to generate a high-resolution electron density map of a protein using cryo-EM.

Moser explained that sample preparation begins with the expression and purification of a homogenous protein sample, which may be accomplished with cell-free expression, *in vivo*, or purification from a native system but stressed that the final product must have high purity and be structurally homogenous. He detailed the factors involved in preparing this protein sample for cryo-EM, including considerations for substrate preparation, such as glow discharging, surface functionalization, or surfactant additives, to obtain a sample with protein particles evenly distributed over the substrate surface in a thin layer of amorphous ice. To freeze these samples, he gave a generic set of freezing conditions and shared that ideal sample preparation conditions often vary between different protein samples and even vary from sample to sample, noting that exact conditions must be determined empirically.

After demonstrating a successful sample preparation, Moser described the considerations for collecting a high-resolution cryo-EM data set for single-particle analysis, including microscope alignment; image acquisition parameters, such as pixel size and dose rate; and an acquisition strategy for improving imaging throughput. He described the various software for collecting data sets and suggested that, generally, depending on the density and distribution of protein particles, a data set of several thousand images may be adequate for structural generation, although larger data sets could further improve resolution.

Moser then discussed the steps required to process data sets to produce a final structural map. Motion correction of individual subframes is performed first to remove vibrational motion artifacts, followed by correction of the contract transfer function to preserve high-frequency phase information. Particle positions could then be picked and extracted from each micrograph, and bad picks or particles could be cleaned up with two-dimensional classification. Moser then described how the final library of extracted particles was used to reconstruct the 3-D volume of the protein, which could be used for atomic modeling. Finally, he gave a brief demonstration of on-the-fly data processing, where images could be processed and an initial structure generated as images came off the microscope during a data collection run.

3.7 Tutorial: NanoPOTS

On day two, EMSL scientist Ying Zhu gave a technical tutorial on the basics of near single-cell proteomics. Zhu introduced near single-cell proteomics and different technologies developed to date. He pointed out cell-type-specific proteomics that can be used for bacteria and yeast, while single-cell proteomics can be used on mammalian and plant systems. Zhu also introduced cell isolation technologies available in EMSL and general considerations during cell isolation. Unlike conventional proteomics, specific care must be taken during the protein extraction and trypsin digestion process. EMSL provides the most sensitive MS platform to analyze the single-cell samples based on the nanoPOTS technology; however, MS parameters must be carefully optimized to maximize the performance. Finally, Zhu provided several study cases using EMSL's single-cell proteomics capability, including the study of hair cell development, single islet heterogeneity, and a cell-type-specific proteomic study of tomato fruit. A video at the end of the tutorial demonstrated the general processes of using nanoPOTS for sample preparation of single-cell proteomics.

3.8 Tutorial: Native MS

After the poster session on day three, EMSL staff scientist Mowei Zhou gave a tutorial on native MS. As demonstrated by several speakers in the morning session, native MS is an emerging technology that complements other structural biology tools such as cryo-EM. In his tutorial, Zhou reviewed recent advances in native MS, including several publications in the *Science* and *Nature* journals from the invited speakers and their co-workers. Native MS has been used to characterize stoichiometry of complexes, higher-order structure, and dynamics (e.g., conformational change, allostery). With proper conditions, membrane protein complexes could also be studied. He also described the basic workflow, EMSL instrumentation, and various sample preparation factors to consider for potential users who want to apply native MS to their research.

In the end, Zhou described two recent EMSL projects utilizing the native MS capability to solve difficult structural biology problems. Both projects were in collaboration with Vicki Wysocki's group at the National Resource for Native MS-Guided Structural Biology at the Ohio State University. During his PhD tenure in Wysocki's lab, Zhou developed surface-induced dissociation (SID) to dissect protein structures. He covered an ongoing EMSL user project with Bradley Tebo (Oregon Health and Science University), and Thomas Spiro and Alexandra Soldatova (both at the University of Washington). They studied a biomineralization enzyme: Mnx from a marine bacterium. Mnx oxidizes Mn(II) and (III) into Mn(IV) oxide and are important in element cycles in biogeochemistry. Mnx is the first enzyme in its family that forms multimeric protein complexes. The structure was unknown, without a homology model to reference. Mnx is also recalcitrant to crystallization and cryo-EM analysis. Zhou worked closely with the Tebo and Wysocki teams and applied native MS and SID to determine the quaternary structure. Based on the MS data, a structural model was built. Intriguingly, the model was well-represented by recent images of Mnx using low-resolution cryo-EM. The team is continuing to obtain high-resolution cryo-EM structures and studying metal-binding properties using native MS.

The other example was an EMSL intramural project in collaboration with Hanjo Hellmann at Washington State University in Pullman. EMSL staff James Evans, Irina Novikova, and Zhou led the project. Hellmann studies plant pesudoenzyme PDX1.2, which is involved in vitamin B6 synthesis. PDX1.2 is catalytically inactive and forms heterocomplexes with its homolog PDX1.3—an active canonical enzyme. Evans and Novikova obtained high-resolution cryo-EM structures for the heterocomplexes between PDX1.2 and PDX1.3, but the two proteins could not be distinguished in the structure due to high similarity and heterogeneity in the assembly. Zhou again used native MS with SID to define the heterogeneity of the assembly. His data revealed some level of symmetry of assembly, shedding light on the mechanism of the hetero association. Both examples from recent EMSL projects highlighted how complementary native MS is to other structural biology techniques. They also echoed the message from other speakers in the session—native MS could help define the heterogeneity of protein complexes.

3.9 Tutorial: Scientific Computing

Aivett Bilbao presented a tutorial about scientific computing on day four. She is a computational scientist at EMSL with a computer engineering background specialized in MS algorithms. She discussed visual proteomics and scientific computing definitions, and described how scientific computing is applied at EMSL to better understand complex systems. She also shared the history of the five HPC systems that EMSL has provided to the scientific community since its inception in 1998. She provided details about the new EMSL HPC, named Tahoma, including hardware specifications, software that will be available, and the process to access and use it. She also compared and illustrated with analogies the advantages and disadvantages of hardware architectures using a central processing unit versus the more recent graphics processing unit, both now provided within Tahoma.

Bilbao described examples of visual-proteomics-related scientific computing research using HPC and mass spectrometry. The first example was Joon-Yong Lee's and colleagues' (https://www.biorxiv.org/content/10.1101/428334v1), "Proteomics of Natural Bacterial Isolates Powered by Deep Learning-based *de novo* Identification." This work takes advantage of deep learning and HPC using graphics processing units. The second example was a study published by Parisa Hosseinzadeh and collaborators (DOI: 10.1126/science.aap7577) performing modeling and design of peptide macrocycles using the Rosetta software. The third example was a work in progress by Joseph Laureanti, Bojana Ginovska, Mowei Zhou, Dennis Thomas, and John Cort in collaboration with Norman Lewis' group at Washington State University. The team is performing molecular dynamics simulations to understand plant dirigent proteins using the GROMACS software. The fourth example was a work performed by Paul Piehowski and Ying Zhu et al. (DOI 10.1038/s41467-019-13858-z), showcasing an MS imaging workflow for spatially resolved proteomics utilizing the Trelliscope software platform for visualization.

Lastly, she described a top-down MS study performed by Mowei Zhou et al. (DOI 10.1016/j.ymeth.2019.10.007) characterizing post-translational modifications of intact histones from sorghum leaves in drought. She introduced PSpecter, a new interactive proteomics software she developed with David Degnan, Mowei Zhou, Amanda White, Lisa Bramer, and Lee Ann McCue. PSpecter is open-source R software integrating bottom-up and top-down database search engines using Docker and HDF5 technologies.

Bilbao concluded her tutorial with three remarks: software is as important as the instruments and techniques that generate the data, scientific computing is a multidisciplinary field that supports a wide range of research and plays a pivotal role in advancing visual proteomics, and finally, more computer science and software engineering are needed to implement robust, reproducible, scalable, and user-friendly visual proteomics software tools.

4.0 Conclusion

EMSL sponsored the 2020 Integration virtual conference, October 8–10, to share what is new and possible in visual proteomics. Intricate biological assemblies, such as macromolecular machines, cells, and tissues, are marked by complex spatial heterogeneity. To understand that, visualizing and quantifying cellular components are necessary—and modern microscopy and MS could help. One leitmotif of Integration 2020 was the range of advanced mass spectrometers, microscopes, and computational tools that EMSL offers to users under one roof to visualize proteome's spatial organization. Scientific sessions and tutorials touched on themes of structural biology, subcellular compartmentalization, annotating protein function, cellular dynamics, and advanced protein modeling and design. The conference had strong international representation, with 346 registrants from 36 countries, from Cameroon and Spain to India and Myanmar, a virtual United Nations of visual proteomics—about three-quarters of them early in their careers. The guests from the scientific community and representatives from the Office of Science Biological and Environmental Research program and the EMSL User Executive Committee shared four days of great science, with many opportunities to learn, connect, and collaborate.

4.1 Looking Ahead to 2021

Join us for the 2021 EMSL Integration Conference, October 4–7. This year's conference will focus on Environmental Sensors. Scientists worldwide from academia, industry, and government research laboratories are welcome to attend free of charge. During the virtual, four-day conference attendees will discuss:

- Plant Biosensors: discovering the molecular dynamics of plant responses to environmental changes
- Microbial Sensors: monitoring nitrogen, carbon, and water cycling in soil
- Rhizosphere Sensors: decoding plant-microbe communication
- Emerging Technologies: development and implementation
- Detecting biological traits across scales.

Appendix A – Agenda

EMSL

www.emsl.pnl.gov

Environmental Molecular Sciences Laboratory
902 Battelle Boulevard • P.O. Box 999 • Richland, WA 99352

Environmental Molecular Sciences Laboratory (EMSL) Integration 2020
Visualizing the Proteome
October 5-8, 2020

AGENDA

Main Session Zoom Link *(this link is the same for each day)*
https://pnnl.zoomgov.com/webinar/register/WN_JIFD9RiQRMGnEAyOyEceeg

Discord Link: https://discord.gg/Snq7MuC

Monday, October 5, 2020				**Location**
Session 1: **Pushing the frontier of structural biology with integrated approaches** Session Chair: James Evans				
8:00 am	Instructions & Intro EMSL Director	James Evans	EMSL	Zoom
8:10 am	Welcome	Douglas Mans, EMSL Director	EMSL	Zoom
8:20 am	Plenary #1	Steve Reichow – "Gap junction channels visualized in a dynamic lipid environment"	Portland State University	Zoom
9:00 am	Session Speaker #1	Dmitry Lyumkis – "Structural Basis for Strand Transfer Inhibitor Binding to HIV Intasomes"	Salk Institute	Zoom
9:20 am	Contributed Talk #1	Qingyan Meng – "Pterocarpan synthase (PTS) structures suggest a common quinone methide-stabilizing function in dirigent proteins with dirigent-like domains"	Washington State University	Zoom
9:40 am	James Evans – "Vision for EMSL Visual Proteomics"		EMSL	Zoom
10:00 am	Break			
10:15 am	MT Thomas Award Introduction by Douglas Mans			Zoom
10:30 am	MT Thomas Award Speaker	Dusan Velickovic – "Molecular snapshots of complex environmental biochemistry"	EMSL	Zoom

EMSL is located at PNNL

Pacific Northwest NATIONAL LABORATORY

11:15 am	Transition from ZOOM webinar to ZOOM meeting			
11:25 am	Introduction to the User Executive Committee and Discussion **Zoom Link:** https://pnnl.zoomgov.com/j/1610350860?pwd=U0lRU2JiY3VmdDMyVnlUcHFUNlBydz09 Meeting ID: 161 035 0860 | Passcode: 175082 | Dial in: +1 669 254 5252 US (San Jose) +1 646 828 7666 US (New York)			Zoom
12:30 pm – 1:30 pm	**Tutorial:** Cryo-EM **Zoom Link:** https://pnnl.zoomgov.com/j/1610350860?pwd=U0lRU2JiY3VmdDMyVnlUcHFUNlBydz09 Meeting ID: 161 035 0860 | Passcode: 175082 | Dial in: +1 669 254 5252 US (San Jose) +1 646 828 7666 US (New York)			Zoom
Tuesday, October 6, 2020				**Location**
Session 2: **Peering inside cells to understand subcellular compartmentalization and ultrastructure** Session Chair: Trevor Moser				
8:00 am	Plenary #2	Elizabeth Wright – "Recent developments in whole cell cryo-electron tomography"	University of Wisconsin-Madison	Zoom
8:40 am	Session Speaker #2	Jun Liu – "Dissecting a molecular machine by cryo-electron tomography"	Yale University	Zoom
9:00 am	Contributed Talk #2	John Yarbrough – "Leveraging Super High Optical Resolution Microscopy to Probe the Interaction Zone Between Clostritium thermocellum and Biomass"	National Renewable Energy Laboratory	Zoom
9:20 am	Contributed Talk #3	Tamas Varga – "Endophyte-Promoted Phosphorous Solubilization in Populus"	EMSL	Zoom
9:40 am	Break			
Session 3: **Revealing cellular complexity and dynamics with single to few cell proteomics** Session Chair: Ying Zhu				
10:00 am	Plenary #3	Thomas Braun – "Microfluidic sample preparation for EM Protein purification, single cell analysis, and visual proteomics"	University of Basel	Zoom

10:40 am	Session Speaker #3	Ryan Kelly – "Digging deeper into the proteome of single cells"	Brigham Young University	Zoom
11:00 am	Session Speaker #4	Peter Nemes – "Single-cell mass spectrometry goes embryonic"	University of Maryland	Zoom
11:20 am	Contributed Talk #4	Vimal Kumar Balasubramanian –"Cell-type Specific Proteomic Responses to Single and Multiple Abiotic Stresses in Poplar"	EMSL	Zoom
12:30 pm – 1:30 pm	**Tutorial:** Few-cell proteomics Zoom Link: https://pnnl.zoomgov.com/j/1610350860?pwd=U0lRU2JiY3VmdDMyVnlUcHFUNlByd z09 Meeting ID: 161 035 0860 | Passcode: 175082 | Dial in: +1 669 254 5252 US (San Jose) +1 646 828 7666 US (New York)			Zoom
Wednesday, October 7, 2020				**Location**
Session 4: **Annotating protein function and modifications via native and top-down proteomics** Session Chair: Mowei Zhou				
8:00 am	Plenary #4	Albert Heck – "How unique is your plasma proteome? A personalized perspective."	Utrecht University	Zoom
8:40 am	Plenary #5	Joseph Gault – "Towards Protein Structures from Native Mass Spectrometry & Single Particle Imaging"	Vertex Pharmaceuticals/ University of Oxford	Zoom
9:20 am	Contributed Talk #5	Varun Gadkari – "Structural Characterization of Large and Heterogenous B12 Trafficking Protein Complexes by Native Ion Mobility-Mass Spectrometry"	University of Michigan	Zoom
9:40 am	Contributed Talk #6	Liangliang Sun – "Automated capillary isoelectric focusing-tandem mass spectrometry for qualitative and quantitative top-down proteomics"	Michigan State University	Zoom
10:00 am	Break			
10:15 am	Session Speaker #5	Jim Bruce – "Interactome studies in aging and heart failure"	University of Washington	Zoom

10:35 am	Virtual Poster Session and Networking **Discord Link**: https://discord.gg/Snq7MuC			Discord
12:30 pm – 1:30 pm	**Tutorial:** Native Mass Spectrometry Zoom Link: https://pnnl.zoomgov.com/j/1610350860?pwd=U0lRU2JiY3VmdDMyVnlUcHFUNlBydz09 Meeting ID: 161 035 0860 | Passcode: 175082 | Dial in: +1 669 254 5252 US (San Jose) +1 646 828 7666 US (New York)			Zoom
Thursday, October 8, 2020				**Location**
Session 5: **Advancing cellular models and protein design with scientific computing** Session Chair: Aivett Bilbao				
8:00 am	Session Speaker #6	Jim Prell – "Studying Quantitative Protein Compaction, Unfolding, and Dissociation with Ion Mobility-Mass Spectrometry"	University of Oregon	Zoom
8:20 am	Session Speaker #7	Chae Kyung Jeon – "Elucidating Gas-phase Unfolding of Protein Complexes through Steered Molecular Dynamics Simulations"	University of Michigan	Zoom
8:40 am	Plenary #6	Cheryl Kerfeld – "Integrated Methods for Characterizing and Engineering Bacterial Microcompartments"	Michigan State University	Zoom
9:20 am	Contributed Talk #7	Min Xu – "Structural pattern mining in cryo-electron tomography"	Carnegie Mellon University	Zoom
9:40 am	Contributed Talk #8	Mark Wirth – "The application of machine learning for the identification of protein signatures from APT data"	EMSL	Zoom
10:00 am	Break			
10:30 am – 11:30 am	**Tutorial:** Scientific Computing Zoom Link: https://pnnl.zoomgov.com/j/1610350860?pwd=U0lRU2JiY3VmdDMyVnlUcHFUNlBydz09 Meeting ID: 161 035 0860 | Passcode: 175082 | Dial in: +1 669 254 5252 US (San Jose) +1 646 828 7666 US (New York)			Zoom

11:30 am – 12:30 pm	**Tutorial:** How to write a successful user proposal Zoom Link: https://pnnl.zoomgov.com/j/1610350860?pwd=U0lRU2JiY3VmdDMyVnlUcHFUNlBydz09 Meeting ID: 161 035 0860 | Passcode: 175082 | Dial in: +1 669 254 5252 US (San Jose) +1 646 828 7666 US (New York)	Zoom

Poster Session – Wednesday 10/7/20

Discord Link: https://discord.gg/Snq7MuC

Poster Title	Name
Niche differentiation of bacteria and eukaryotes in carbon and nitrogen cycling of different habitats in a temperate coniferous forest: A soil metaproteomic approach	Baldrian, Petr
Pushing structure analysis by cryo-EM beyond 1.5A resolution	Brink, Jacob
Copper availability controls cell structure, cell energies and MxaF/XoxF expression in Methylomicrobium alacaliphilum 20Z, a model methanotroph possessing only particulate methane-monooxygenase	Collins, David
Pterocarpan synthase (PTS) structures suggest a common quinone methide-stabilizing function in dirigent proteins with dirigent-like domains	Cort, John
Seasonal differentiation of the nitrogen- and carbon cycle in a coniferous forest using metaproteomics	Human, Zander
Integrative omic and transgenic analyses reveal the positive effect of ultraviolet-B irradiation on salvianolic acid biosynthesis through up-regulation of SmNAC1	Lu, Xu
The National Center for X-ray Tomography: What can it do for you do?	McDermott, Gerard
Objective crystallographic symmetry classifications for electron crystallography based structural biology	Moeck, Peter
In liquid infrared scattering scanning near-field optical microscopy for chemical and biological nano-imaging	O'Callahan, Brian
Deletion of the ntrYX two component system in Thidobacter sphaeroides causes the generation of diverse extracelluar membrane structures	Parrell, Daniel
Towards the development of a bio-APT capability at EMSL	Perea, Daniel
Investigating protein complex subunit composition, conformation, and topology with combined native ion mobility-mass spectrometry and computation approaches	Rolland, Amber

Endophyte-Promoted Phosphorous Solubilization in Populus	Varga, Tamas
Advanced techniques to Study Protein Structure and Function at SSRL	Wierman, Jennifer
Investigation of Sorghum Histone Post-Translational Modifications as Epigenetic Markers for Drought Acclimation using Top-Down Mass Spectrometry	Wilson, Jesse
The application of machine learning for the identification of protein signatures from APT data	Wirth, Mark
Ion mobility fractionation enhanced mass spectrometry acquistion method for single-cell proteomics	Woo, Jongmin
Automated capillary isoelectric focusing-tandem mass spectrometry for qualitative and quantitative top-down proteomics	Xu, Tian
CorRelator: an interactive and flexible tool for high-precision cryo-correlative light and electron microscopy	Yang, Jae
Leveraging Super High Optical Resolution Microscopy to Probe the Interaction Zone Between Clostritium thermocellum and Biomass	Yarbrough, John

Appendix B – Attendance Statistics and Feedback

EMSL 2020 INTEGRATION CAMPAIGN

EMSL

Facebook	3,789 Impressions	190 Engagements
LinkedIn	431 Impressions	39 Engagements
Twitter	18,956 Impressions	569 Engagements
Email	6,800 Sent to	225 Clicked thru
Website	520 Page views	460 Unique views

Impressions = number of times the post has shown up on Facebook, Twitter, LinkedIn.

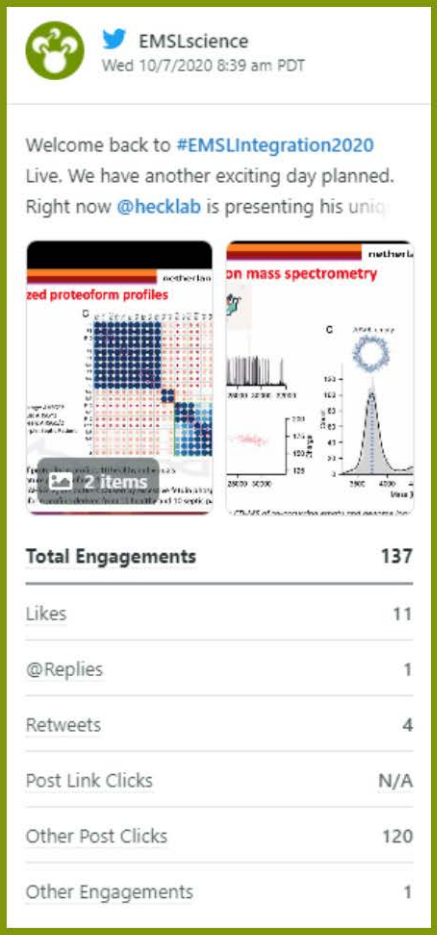
Engagements = number of times people "liked" or commented on the post.

Clicked thru = number of people who clicked on the website link embedded in the email.

Page views = number of times the web page was viewed.

Unique views = the number of unique individuals who viewed the webpage.

Top Post

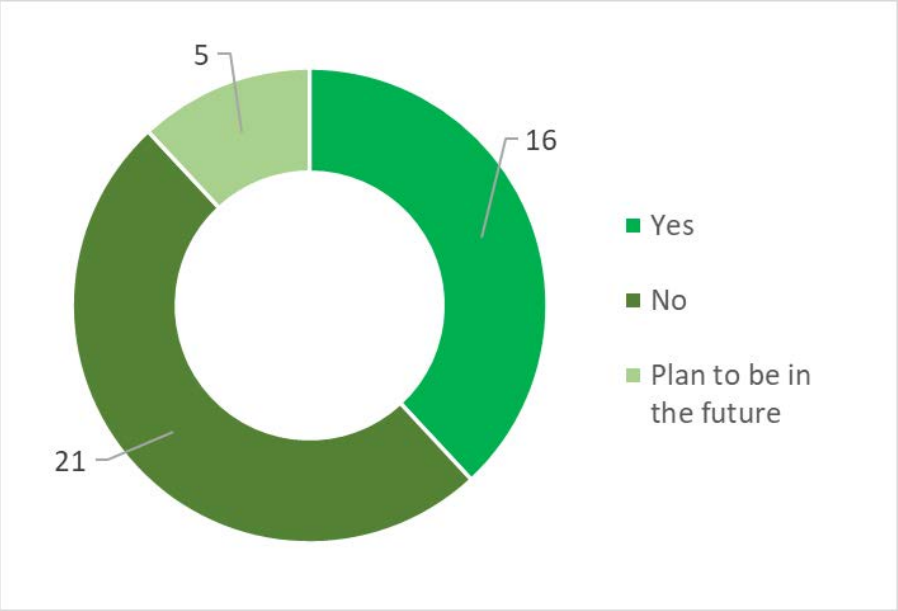


Impressions: 1,752
# of people reached: 5,277

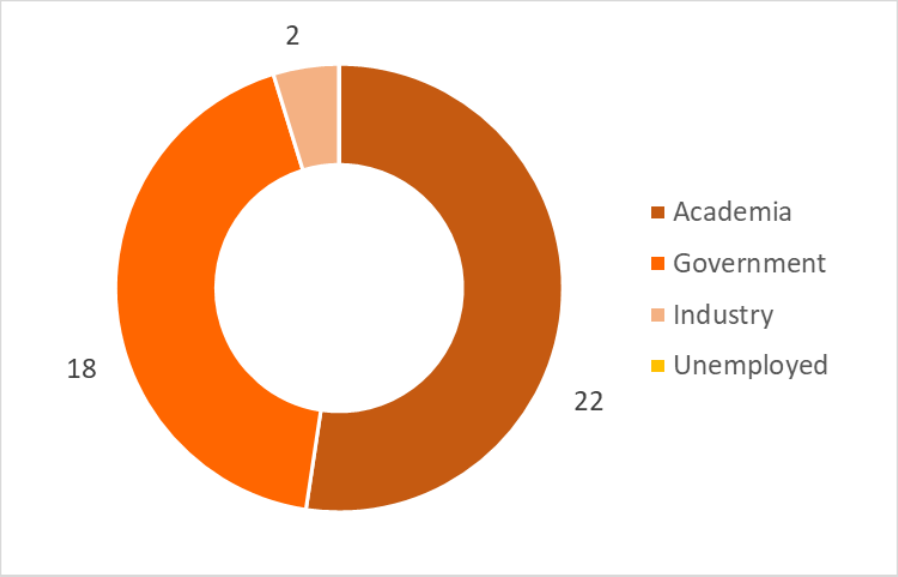
Appendix C – Feedback Survey

On Friday, October 9, EMSL Communications sent an email to 2020 Integration Conference registrants requesting they complete a feedback survey. The survey was open for three weeks and closed on October 30. Of the 330+ registrants, there were 42 respondents. Here is a summary of their responses.

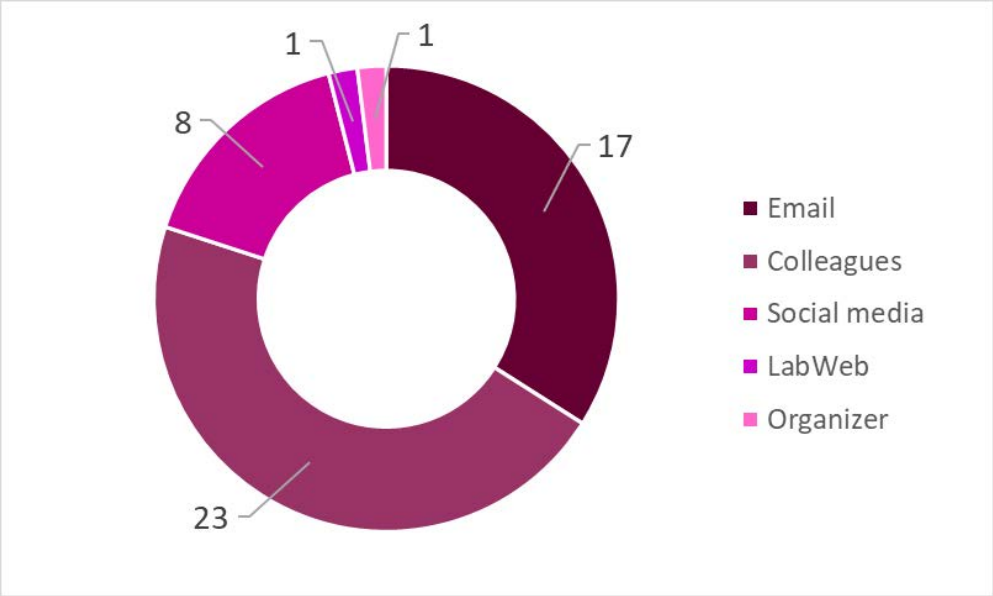
1. **Are you an EMSL user?** Total respondents – 42



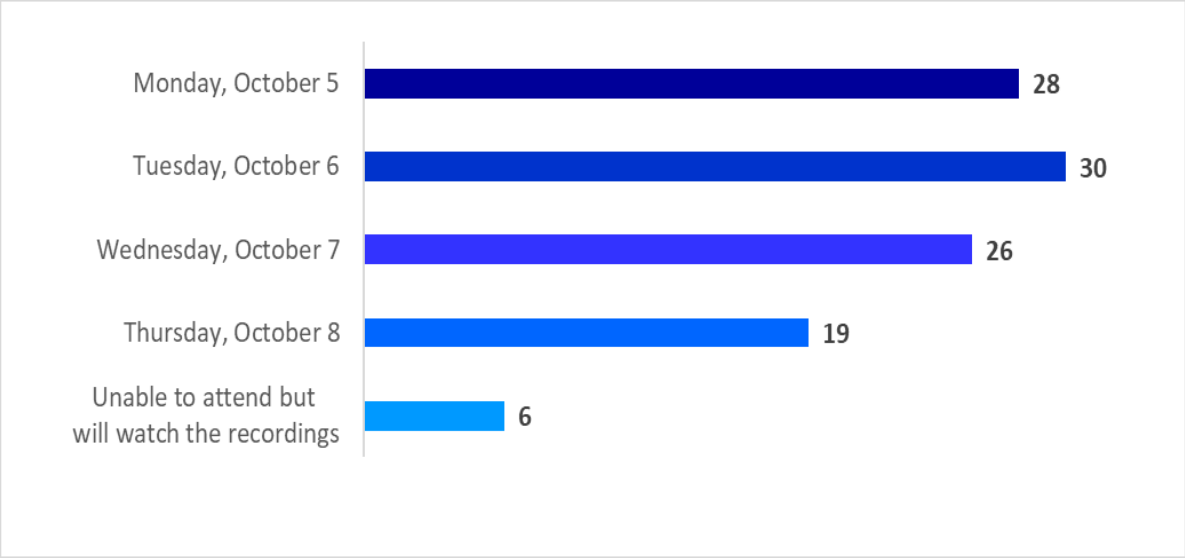
2. **Which sector are you currently employed with?** Total respondents – 42



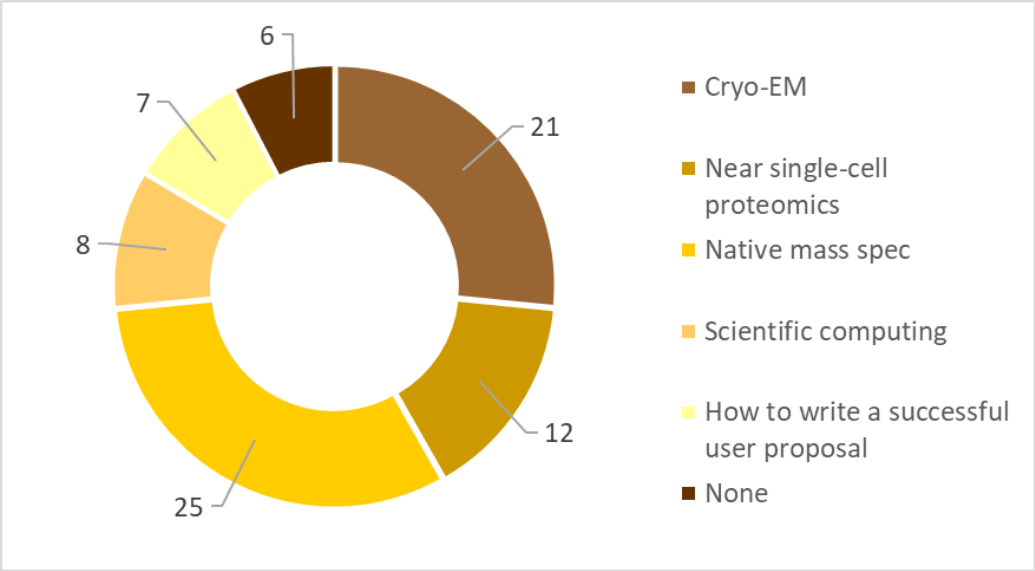
3. **How did you hear about the 2020 EMSL Integration Meeting?** (Check all that apply) Total respondents – 41



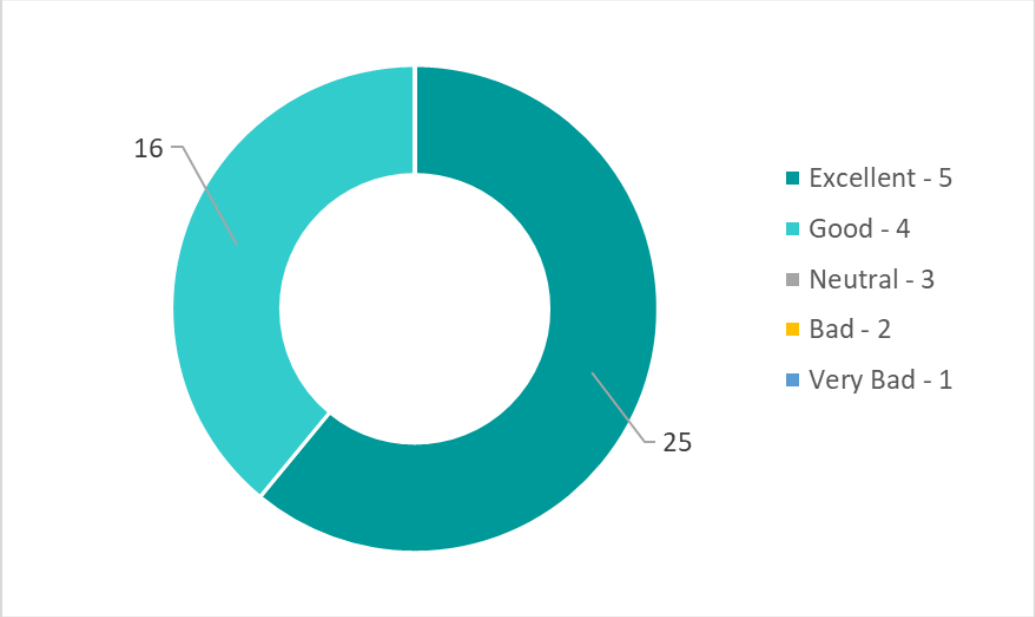
4. **What days did you attend**? (Check all that apply) Total respondents – 42



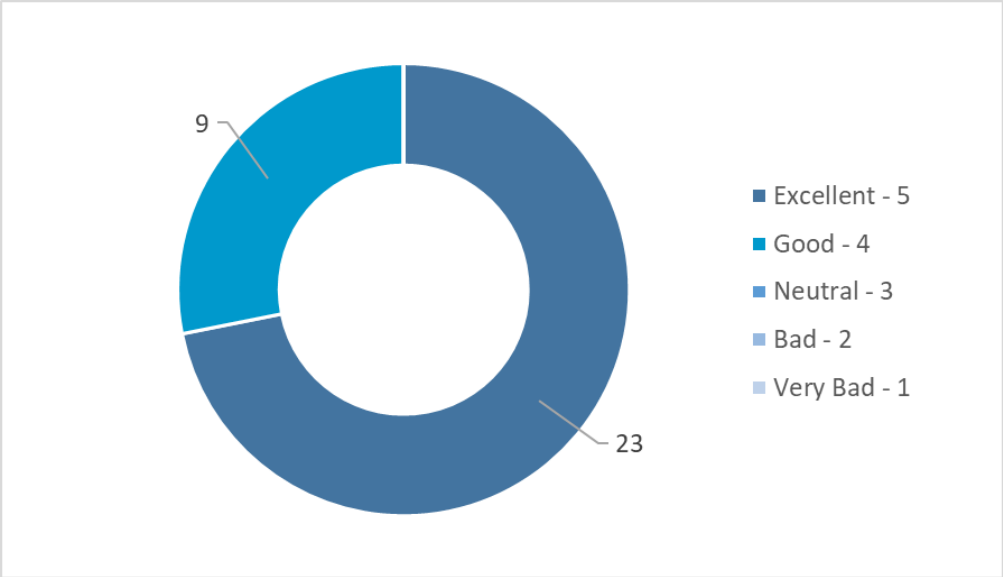
5. **Which tutorial did you attend?** (Check all that apply) Total respondents – 39



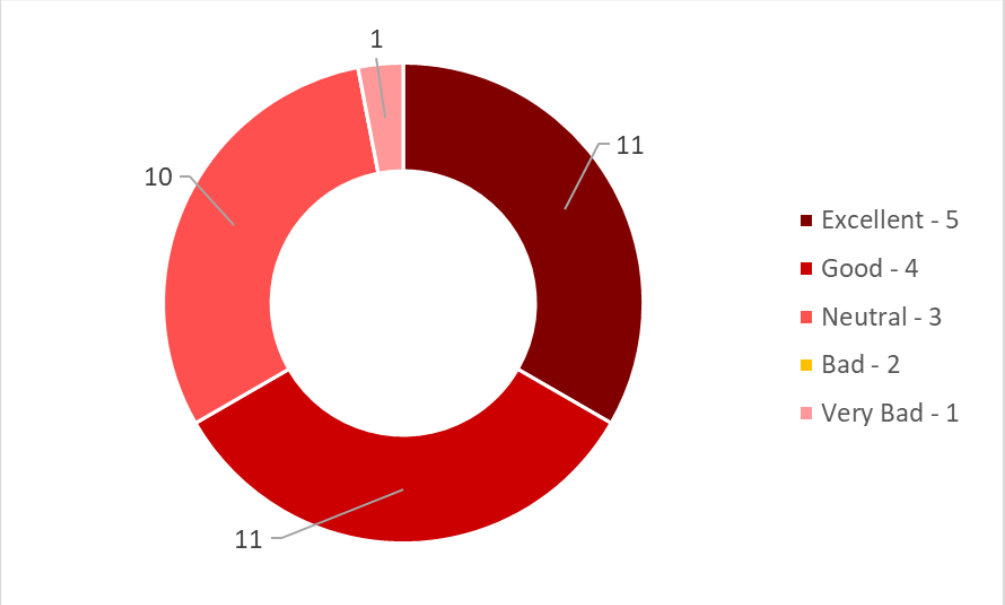
6. **How do you rate the quality of presentations given at 2020 Integration?** Total respondents – 41



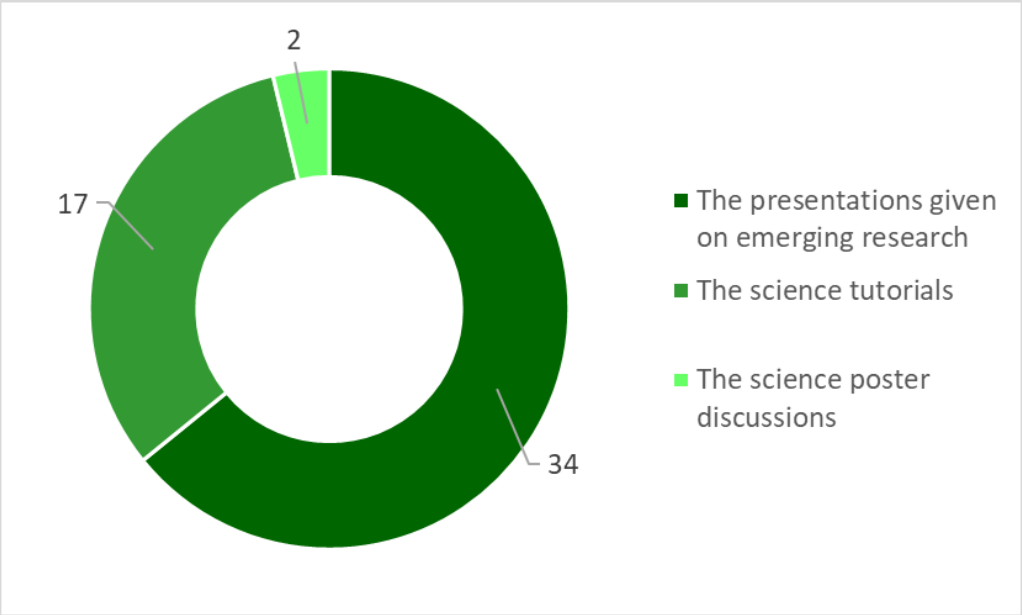
7. **How do you rate the quality of the science tutorials given at 2020 Integration?** Total respondents – 32



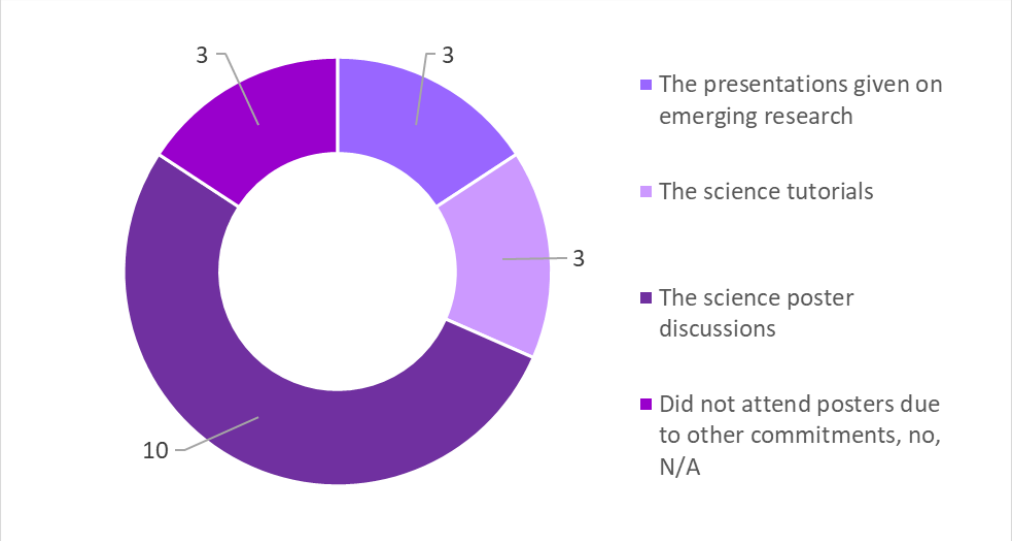
8. **How do you rate Discord for ease of use in chatting and the poster session?** Total respondents – 33



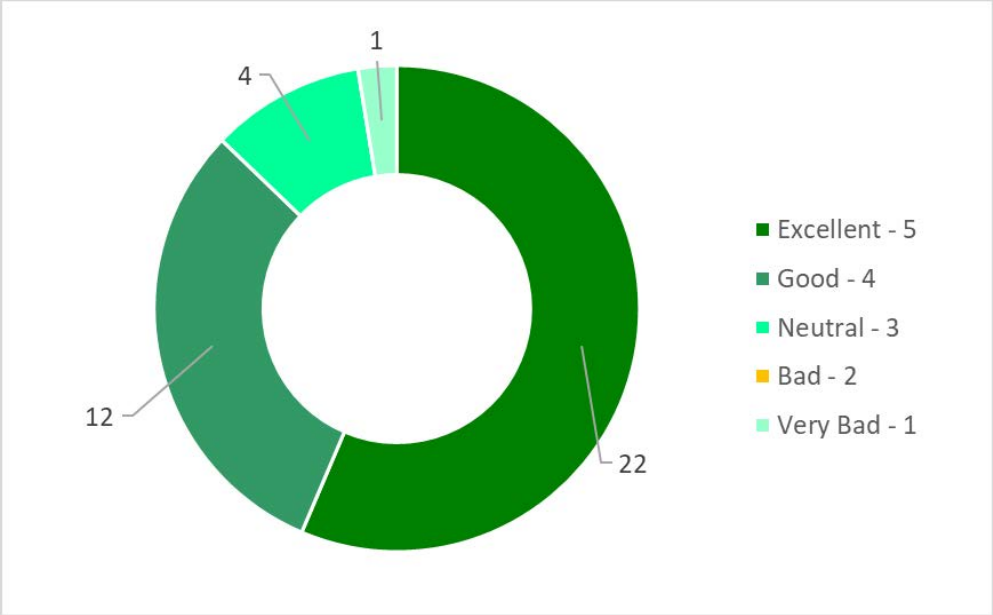
9. **What aspects of the meeting did you find most useful?** (Check all that apply) Total respondents – 41



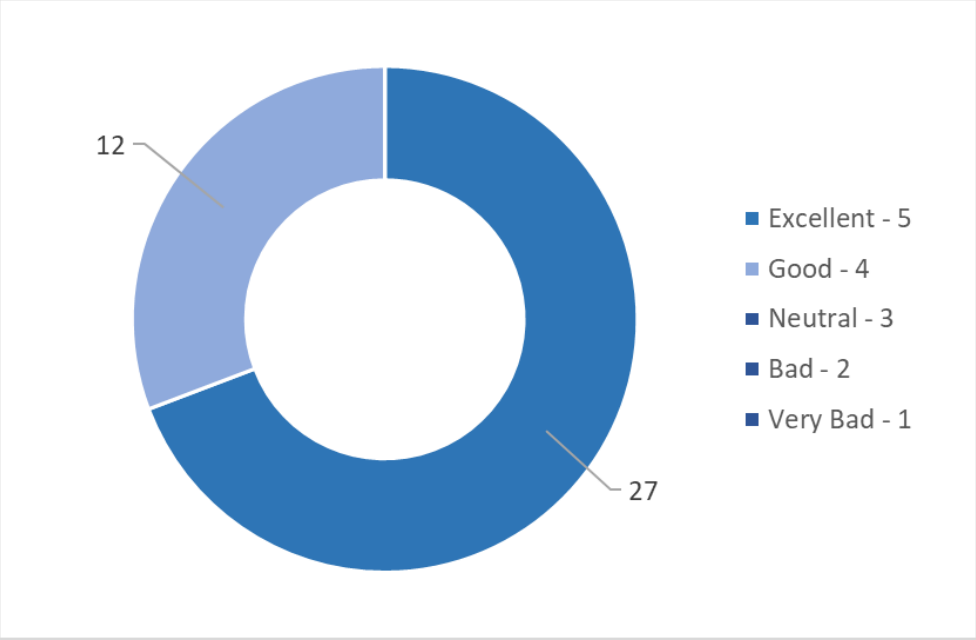
10. **What aspects of the meeting did you find least useful?** (Check all that apply) Total respondents – 19



11. **Rate the logistics support provided for Discord and Zoom before and during the meeting.** Total respondents – 39



12. **Overall, how do you rate the 2020 EMSL Integration Meeting?** Total respondents – 39




EMSL