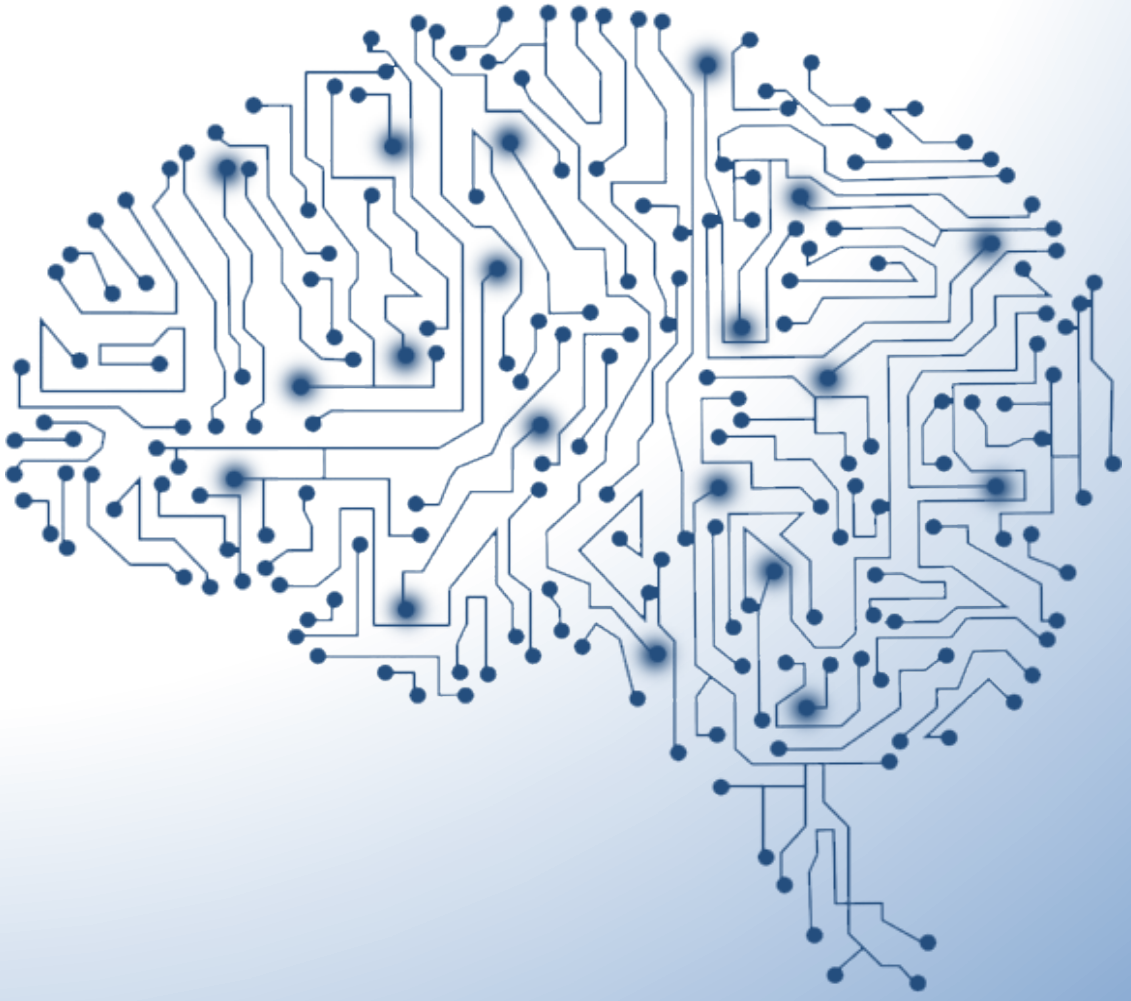
Metabolomics Association of North America Summer Symposium 2019

September 04, 2019



Quantum Chemistry and Computational Methods for Compound Identification

hosted by West Coast Metabolomics Center, UC Davis

Metabolomics Association of North America Summer Symposium 2019

hosted by West Coast Metabolomics Center, UC Davis

Quantum Chemistry and Computational Methods for Compound Identification

September 04, 2019

9 a.m. – 7 p.m.
UC Davis Genome Center Auditorium

Keynote speaker

Stefan Grimme,
University of Bonn, Germany

"Quantum chemistry prediction of NMR and EI mass spectra"

register & submit poster:
https://metabolomicsna.org

supported by: RCN 1743594

Speakers:

Yannick Djoumbou Feunang, Corteva Agriscience, USA
CFM-ID 3.0, ClassyFire, Biotransformer

Ryan Renslow, PNNL, USA
ISiCL ion mobility prediction

Stephen Stein, NIST, USA
MS and MS/MS libraries, hybrid searches

Jennifer Wei, Google Brain, USA
neural networks for EI mass spectra

Lloyd W. Sumner, University of Missouri, USA
combining MS/MS with NMR, and CCS

Xiuxia Du, University of North Carolina, USA
MS data processing, data deconvolution

Pieter Dorrestein, UC San Diego, USA
GNPS , MS/MS similarity propagation

Zheng-Jiang Zhu, CAS, Shanghai, China
MetDNA, ion mobility, DIA MS

David Grant, University of Connecticut, USA
Ecom50, predicting IR spectra, retention index

Ivana Blazenovic, DiscernDX, USA
compound ID confidence scoring

MANA Summer Symposium 2019

Quantum Chemistry and Computational Methods for Compound Identification

September 04, 2019
Program

8:00-9:00	breakfast and registration	
9:00 - 9:10	welcome - Fiehn, Oliver	US
9:10 - 10:00	Grimme, Stefan	Germany
10:00 - 10:30	Djoumbou Feunang, Yannick	US
	break	
11:00 -11:30	Renslow, Ryan	US
11:30-12:00	Stein, Stephen	US
12:00-12:30	Wei, Jennifer	US
	lunch break	
1:30-2:00	Sumner, Lloyd	US
2:00-2:30	Du, Xiuxia	US
2:30-3:00	Dorrestein, Pieter	US
	break	
3:30-4:00	Zhu, Zheng-Jiang	China
4:00-4:30	Grant, David	US
4:30-5:00	Ivana Blazenovic	US
5:00-7:00	reception, posters	

Speaker Abstracts

Keynote speaker
Stefan Grimme, University of Bonn, Germany

“Quantum chemistry prediction of NMR and EI mass spectra”

The GFN-xTB family of semi-empirical tight-binding methods is introduced[1]. The methods follow a global and element-specific parameters only strategy and are consistently parameterized for all elements through radon. Their original purpose and main target for the parameter optimization has been the computation of molecular geometries, vibrational frequencies, and non-covalent interactions. They are effectively used in the framework of meta-dynamics (MTD) to globally explore chemical compound, conformer, and reaction space.[2] For typical conformational search problems of organic drug molecules, the new MTD(RMSD) algorithm yields lower energy structures and more complete rotatmer/conformer ensembles at reduced computational effort. They are used in a fully automated procedure to compute high-resolution 1H-NMR spectra[3]. TB methods combined with the Fermi-smearing technique can also describe difficult electronic structures and covalent bond breaking at least qualitatively correct. This enables the 'first-principles' automated quantum chemistry computation of electron ionization mass spectra even for transition metal complexes[4]. The scope and limitations of various 'low-cost' quantum chemistry methods in typical chemistry applications like pKa or CD spectra calculation[5] are discussed.

[1] C. Bannwarth, S. Ehlert, S. Grimme, J. Chem. Theory Comput. (2019), 15, 1652-1671.
[2] S. Grimme, J. Chem. Theory Comput. (2019), DOI:10.1021/acs.jctc.9b00143
[3] S. Grimme, C. Bannwarth, S. Dohm, A. Hansen, J. Pisarek, P. Pracht, J. Seibert, F. Neese, Angew. Chem. Int. Ed. (2017), 56, 14763-14769.
[4] V. VÅsgeirsson, C. A. Bauer, S. Grimme, Chem. Sci. (2017), 8, 4879-4895.
[5] P. Pracht, R. Wilcken, A. Udvarhelyi, S. Rodde, S. Grimme, J. Comput. Aided Mol. Des. (2018), 32, 1139-1149.
J. Seibert, C. Bannwarth, S. Grimme, J. Am. Chem. Soc. (2017), 139, 11682-11685.

grimme@thch.uni-bonn.de

Stefan Grimme, PhD studied Chemistry and finished his Ph.D. in 1991 in Physical Chemistry on laser spectroscopy. He habilitated in Theoretical Chemistry in the group of Sigrid Peyerimhoff. In 2000, he got the C4 chair for Theoretical Organic Chemistry at the University of Muenster. In 2011, he accepted an offer as the head of the founded Mulliken Center for Theoretical Chemistry at the University of Bonn. He has published more than 490 research articles and is the recipient of the 2013 SchrVðdinger medal of the World Organization of Theoretically Oriented Chemists (WATOC). In 2015, he was further awarded the ‚Gottfried Wilhelm Leibniz-Preis, of the DFG. His main research interests are the development and application of quantum chemical methods for large molecules, density functional theory, noncovalent interactions, and their impact in chemistry.

Speaker 02

Yannick Djoumbou Feunang, Corteva Agriscience, USA

"Cheminformatics Tools for Enabling Metabolomics"

Endogenous (metabolic) and exogenous (environmental) transformations are mainly responsible for an estimated 95% of the human chemical exposome (~3 million compounds) that remains largely uncharacterized or unknown. The identification and characterization of these compounds is crucial in understanding the short- and long-term effects that they may have on human and animal health, as well as the environment. Yet, it is a daunting task. In this presentation I will describe several computational tools that I contributed developing to enable Metabolomics. In particular, I will present: 1) ClassyFire and ChemOnt, freely available resources for the automated hierarchical structure-based classification of chemicals; 2) BioTransformer, a software tool for predicting metabolic biotransformation products arising from human metabolism and environmental microbial degradation; and 3) CFMID 3.0, the latest release of CFM-ID, a software tool designed to accurately predict mass spectra for rapid compound identification. These tools were developed as part of my research at the University of Alberta, Canada.

yannick.djoumbou@gmail.com

Yannick Djoumbou Feunang joined Corteva Agriscience in 2018 as a Data scientist/Cheminformatician, after a PhD and Post-Doc journey at the University of Alberta, Canada. Over the last 10 years, his research interests have focused on exploring the interface between Biology and Chemistry, using Computing Science. In 2018, he earned the Scholarship for Excellence from the Chemical Information division of the American Chemical Society, for his work on in silico metabolism prediction. Moreover, he has worked on and in silico MS-spectra prediction and compound identification, Cheminformatics database development, and Biochemical/Biomedical ontology development projects.

Currently, Yannick is a member of the Chemistry Data Science and Informatics (ChemDSI) group at Corteva Agriscience, where he contributes to the development of predictive models, as well as the development of the Cheminformatics workbench, cutting-edge Cheminformatics scientific computing platform.

Speaker 03

Ryan Renslow, PNNL, USA

“ISiCL ion mobility prediction, deep learning, quantum chemistry / DFT”

We are advancing the use of computationally calculated chemical properties to enable the identification of molecules with reduced reliance on the use of data from analysis of authentic reference materials. Our molecular identification pipeline, ISiCLE (In Silico Chemical Library Engine), uses a large-scale computational chemistry platform that exploits PNNL’s high-performance computational quantum chemistry software, NWChem, to calculate small molecule chemical properties, such as collision cross section (CCS) and nuclear magnetic resonance (NMR) chemical shift. In this symposium I will discuss the details of the ISiCLE workflow, as well as report on recent changes on how adduct ionization sites are predicted. I will present several successful application examples, including identifying environmental degradation products, separating molecular isomers, and correcting mislabeled isobaric isosteres. Finally, the challenges and future plans of this approach will be discussed, including a comparisons to recent deep learning approaches.

ryan.renslow@pnnl.gov

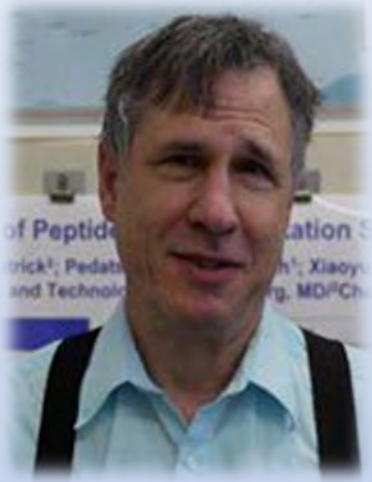
Ryan Renslow is an expert in mathematical modeling, with a focus in biological systems and molecular modeling and quantum chemical calculations. Dr. Renslow has over 50 publications, and expertise in quantitative biology, density functional theory, and biological data analysis automation. A chemical engineer by training, his research has been routinely multi-disciplinary, working alongside microbiologists, ecologists, physicians, dentists, human health researchers, and researchers specializing in metabolomics and exposomics.
Ryan Renslow received his B.S. and M.S. and his PhD in Chemical Engineering at the Washington State University , Pullmam, WA.. In 2010 he was the NIH Protein Biotechnology Fellow, Washington State University, Pullman, WA . Ryan Renslow received 2015 the Linus Pauling Distinguished Postdoctoral Fellowship at PNNL, Richland, WA .

Speaker 04

Stephen E. Stein, NIST, USA

"Spectral-library based methods for identifying compounds not in the library"

There is wide agreement that the inability to reliably identify a large fraction of the ions generated in mass-spectral-based metabolomics studies greatly hinders progress in this field. While identification through libraries can, in principle, provide a solution, libraries are, and are likely to remain too limited in scope to provide a solution to this problem. Currently, perhaps 10% of the ions observed in urine or plasma can be identified by libraries. However, current libraries contain an enormous amount of information linking chemical structures to spectra that can be exploited identify compounds absent from the library. This is done by, in effect, making use of fragmentation pathways of for spectra present in the library to tentatively compounds not present in the library. We discuss two such methods: 1) the 'hybrid search' which finds compounds that differ from a library compound by an 'inert' chemical group and 2) our MS Interpreter program which validates the correctness of a spectrum/structure pair using basic ion fragmentation principles and thermochemistry, both derived from library spectra.



stephen.stein@nist.gov

Stephen E. Stein, PhD is NIST Fellow, and Director of the NIST Mass Spectrometry Data Center in the Biomolecular Measurement Division, Material Measurement Laboratory, National Institute of Standards and Technology. Dr. Stein conceives, plans, directs and personally conducts advanced research and development in chemical data analysis, with emphasis on mass spectral data and search algorithms, deconvolution algorithms, quality measurement and new approaches for chemical identification in proteomics and metabolomics, methods for automated chemical data analysis, algorithms for spectra and property prediction from chemical structural data.

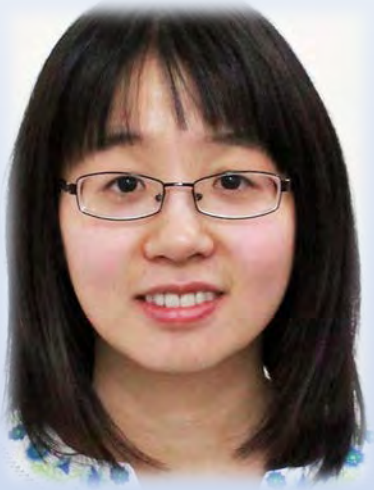
Dr. Stein received his BS in Chemistry from U Rochester, NY, in 1969, and his PhD in Physical Chemistry from U Washington in 1974. He was Associate Professor in Chemistry at West Virginia U until 1982 and serves as director of the NIST MS data center since 1988.

Speaker 05

Jennifer Wei, Google Brain, USA

“neural networks for predicting EI mass spectra”

We propose to improve the library’s coverage by augmenting it with synthetic spectra that are predicted from candidate molecules using machine learning. We contribute a lightweight neural network model that quickly predicts mass spectra for small molecules, averaging 5 ms per molecule with a recall-at-10 accuracy of 91.8%. Achieving high-accuracy predictions requires a novel neural network architecture that is designed to capture typical fragmentation patterns from electron ionization. We analyze the effects of our modeling innovations on library matching performance and compare our models to prior machine-learning-based work on spectrum prediction.



weijennifer@google.com

Jennifer Wei is a software engineer with the Brain Research team in Cambridge, MA. She received her PhD in Chemical Physics at Harvard University. Her primary research interests are the applications of machine learning for small molecules towards predicting molecular properties and chemical phenomena. She has published research on applications of machine learning to reaction prediction, inverse design of molecules, and mass spectrometry prediction.

Speaker 06

Lloyd W. Sumner, U Missouri, USA

"Development of Integrated Computational and Empirical Tools to Address the Metabolomics Grand Challenges of Confident Metabolite Identification and Increased Depth-of-Coverage"

The vast utility of metabolomics is well documented in the literature; however, its full scientific promise has not yet been realized due to multiple technical challenges. These grand challenges include large-scale confident metabolite identifications and greater depth of coverage. We have developed sophisticated spectral, computational and integrated experimental metabolomics tools for the systematic and biologically directed annotation of plant metabolomes and for greater metabolome depth of coverage. This presentation will describe a UHPLC-QTOF-MS/MS mass spectral library, custom software entitled Plant Metabolite Annotation Toolbox (PlantMAT), software for CCS prediction, sophisticated UHPLC-timsTOF-MS/MS and UHPLC-QTOF-MS-SPE-NMR instrumental ensembles that are being used for 'sequencing' the first plant metabolomes.

sumnerlw@missouri.edu

Lloyd W. Sumner acquired his Ph.D. in analytical chemistry from Oklahoma State University in 1993. He joined the Samuel Roberts Noble Foundation in 1999 and rose to the rank of Professor within the Plant Biology Division. He then moved to the University of Missouri, Columbia in January 2016 as a Professor in the Biochemistry Department and Director of the MU Metabolomics Center. Dr. Sumner's research program focuses upon the development, integration, and application of large-scale biochemical profiling technologies to better understand plant specialized metabolism. Dr. Sumner's research is or has been graciously supported by the University of Missouri, Noble Foundation, NSF 2010, NSF MCB, NSF MRI, NSF-JST, NSF-IOS, NSF-PGRP and The Oklahoma Commission for the Advancement of Science and Technology. Dr. Sumner is currently an AAAS Fellow; Former Treasurer and President of the Metabolomics Society, Lifetime Honorary Member of the Metabolomics Society, Distinguished Alumni of Cameron University, and President Emeritus of the Phytochemical Society of North America.

Speaker 07

Xiuxia Du, UNC Charlotte, USA

"Spectral deconvolution for constructing pure mass spectra for compound identification"

Spectral deconvolution is an essential step in preprocessing untargeted GC-MS metabolomics data. It computationally separates ions that are in the same mass spectrum but belong to coeluting compounds that are not resolved completely by chromatography. As a result of this computational separation, spectral deconvolution produces pure fragmentation mass spectra.

Each mass spectrum consists of ions from a single compound, allowing identification or annotation of compounds through searching spectral libraries.

Traditionally, spectral deconvolution has been achieved by using a two-step model peak approach. The first step detects the presence of components and selects a model peak for each perceived component. At this stage, a component generally consists of multiple chromatographic peaks that have a very similar peak shape. From these peaks, a model peak is selected that can best represent the elution profile of the component. The second step decomposes each detected peak into a linear combination of the model peaks and constructs a pure fragmentation spectrum for each perceived component. This two-step approach is computationally efficient and provides model peaks that are similar to real chromatographic peaks in shape. However, there is always the risk that a selected model peak has actually been produced by two or more co-eluting components and therefore is inappropriate to serve as a model peak. This inappropriate model peak selection would cause incorrect pure fragmentation spectra for all the involved co-eluting compounds and eventually errors in library matching. To address the limitations of the traditional approach, we have developed a multivariate curve resolution (MCR)-based method. At this symposium, I will describe the fundamental differences between the two approaches and how we have implemented the MCR-based method while overcoming the inherent computational complexity.

xiuxia.du@uncc.edu

Xiuxia Du is an expert in metabolomics informatics, statistics, and machine learning. In particular, she is focused in developing computational algorithms and software tools for preprocessing mass spectrometry-based metabolomics data.

The suite of ADAP algorithms and the associated software tool that her team have developed provide metabolomics researchers an alternative informatics approach to preprocess their untargeted LC-MS and GC-MS metabolomics data. The source code of all of the algorithms and the related software tools is publicly available on Github for close scrutiny by the metabolomics community.

Prof Xiuxia Du received her B.S. and M.S. in Electrical Engineering at the Hefei University of Technology , China. In 2005 she finished her PhD at Washington University in St. Louis and started her postdoc at PNNL in Computational Proteomics. Professor Du is since 2008 a Professor at the Department of Bioinformatics & Genomics University of North Carolina at Charlotte.

Speaker 08

Pieter Dorrestein, UC San Diego, USA

“Mass Spectrometry Annotation and Analysis at the Repository Scale”

This lecture will describe the emerging approaches that are being developed by the GNPS user community to annotate molecules not just at the individual compound level but also using different levels of information at the repository scale. It will describe the logistical challenges and informatics solutions to enable comparison of data from different labs, often collected on different instruments, but also computational strategies for discovery and annotation of biologically relevant molecular ion partners of small molecules, another level of annotation. Finally it will describe the importance of FAIR principles to improve the amount of information that can be obtained using mass spectrometry.

pdorrestein@ucsd.edu

Pieter Dorrestein is an Associate Professor at the University of California - San Diego. He is the Director of the Therapeutic Discovery Mass Spectrometry Center and a Co-Director, Institute for Metabolomics Medicine in the Skaggs School of Pharmacy & Pharmaceutical Sciences, Departments of Pharmacology, Chemistry and Biochemistry. Dr. Dorrestein was trained by Tadgh Begley in the chemical biology of enzymes involved in vitamin biosynthesis and by Neil Kelleher and Christopher Walsh, whom were co-sponsors of his NRSA postdoctoral fellowship, in Top and Middle down mass spectrometry on proteins that made small molecules of therapeutic value. Since his arrival to UCSD in 2006, Dr. Dorrestein has been pioneering the development of mass spectrometry methods to study the chemical and ecological crosstalk between population of organisms for agricultural, diagnostic and therapeutic applications.

Speaker 09

Zheng-Jiang Zhu, CAS, Shanghai, China

Interdisciplinary Research Center on Biology and Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, 200032 P. R. China

"MetDNA: Metabolic Reaction Network-based Recursive Metabolite Annotation for Untargeted Metabolomics"

Large-scale metabolite annotation is a challenge in liquid chromatogram-mass spectrometry (LC-MS)-based untargeted metabolomics. Here, we develop a metabolic reaction network (MRN)-based recursive algorithm (MetDNA) that substantially expands metabolite annotations without the need for a comprehensive standard spectral library. Metabolites and their reaction-paired neighbor metabolites tend to share similar MS2 spectra due to structural similarity. Based on this rationale, MetDNA characterizes initial seed metabolites using a small tandem spectral library, and utilize their experimental MS2 spectra as surrogate spectra to annotate their reaction-paired neighbor metabolites which are subsequently served as the basis for recursive analysis. We further showcase the utility and versatility of MetDNA using different LC-MS instrumentations, data acquisition methods, and biological sample types, and demonstrate that about 2,000 metabolites can cumulatively be annotated from one experiment. MetDNA largely expands the annotation of metabolites, thereby allowing quantitative assessment for not just metabolic pathways but also multi-omic studies, such as integrative analysis between metabolomics and transcriptomics.

jiangzhu@sioc.ac.cn

Zheng-Jian Zhu received his B.S. degree in Chemistry from Nanjing U in 2006, and his Ph.D. degree in Chemistry from U Massachusetts at Amherst in 2011. He worked as postdoctoral researcher with Gary Siuzdak at Scripps from 2011 to 2013. From 2013, he was appointed as Professor in Shanghai Institute of Organic Chemistry (SIOC) and Interdisciplinary Research Center on Biology and Chemistry (IRCBC), Chinese Academy of Sciences (CAS).

Dr. Zhu develops mass spectrometry based metabolomics technology with applications in health and disease related research. Dr. Zhu has published >50 peer-reviewed papers on the prestigious journals including Nature, Nature Chemistry, Nature Biotechnology, Nature Communications, JACS, eLife, Analytical Chemistry and Bioinformatics with >3000 times in citation. Dr. Zhu is supported by the Thousand Youth Talents Program. See http://www.zhulab.cn for more information.

Speaker 10

David Grant, U Connecticut, USA

"ISiCL ion mobility prediction, deep learning, quantum chemistry / DFT"

The long-term objective of the research in my lab is to develop analytical, computational and database tools that can be used to rapidly identify the chemical structures of metabolites in human biofluids. Our computational algorithms predict physical/chemical properties of compounds contained in large chemical databases. The physical/chemical properties chosen are those that can be experimentally determined for any unknown compound by HPLC-mass spectrometry. These include retention indices, collision induced dissociation fragmentation spectra, Ecom50, collision-cross section and infrared spectra. Compounds in databases (for example PubChem) whose predicted properties most closely match experimental properties are prioritized as likely candidates for the unknown. By facilitating the structural identification of unknown chemical compounds these tools will enhance the ability of metabolomics to compliment and synergize other areas of biomedical research.

david.grant@uconn.edu

David Grant received BS degrees in Biology and Chemistry from Idaho State University in 1979 and a PhD in Environmental Toxicology from Michigan State University in 1987. After postdoctoral training in molecular toxicology at The University of California-Davis, he began a tenure-track position in 1995 in the Department of Pharmacology and Toxicology at the University of Arkansas for Medical Sciences. He moved to the University of Connecticut in 2001 and is currently a Professor in the Department of Pharmaceutical Sciences. His research is focused on developing analytical and computational approaches to determine the structure of unknown metabolites in human biofluids. Dr. Grant has served as a member of The American Society for Pharmacology and Experimental Therapeutics, The International Society for the Study of Xenobiotics, The American Society for Mass Spectrometry and The Metabolomics Society. He serves on NIH study sections in the areas of metabolomics and analytical biochemistry.

Speaker 11

Ivana Blazenovic, DiscernDX, USA

“Confidence in compound ID: software, databases and tools.”

The annotation of small molecules remains a major challenge in untargeted mass spectrometry-based metabolomics. Critical discussion is needed on structured elucidation approaches and software that are designed to help during the annotation of unknown compounds. Only by elucidating unknown metabolites first is it possible to biologically interpret complex systems, to map compounds to pathways and to create reliable predictive metabolic models for translational and clinical research. We present in silico fragmentation tools such as MS-FINDER, CFM-ID, MetFrag, ChemDistiller and CSI:FingerID that can annotate compounds from existing structure databases and that have been used in the CASMI (critical assessment of small molecule identification) contests. The use of retention time models from liquid chromatography and the utility of collision cross-section modelling from ion mobility experiments are discussed. Furthermore, we present annotation of all spectra in a sample using variety of freely available m/z-RT databases, mass spectral libraries (MoNA, NIST17, LipidBlast, CarniBlast etc).Text

Ivana Blazenovic earned her phD at TU Braunschweig, Germany in the field of Metabolomics and Microbiology and her MSc at University of Zagreb, Croatia in the field of Biotechnology. Dr. Ivana Blazenovic is currently a senior scientist working in a clinical start up company DiscernDx in Palo Alto. Prior to working in industry she was a post doctoral scholar in the Fiehn lab at UC Davis working on multiple pilot projects with a focus on up-to-date mass spectrometry based bioinformatics and cheminformatics to discover, identify and validate small molecules of interest such as urinary tract infection biomarkers, as foundation for clinical microbiology, natural products etc. Her other work experience includes international research in industrial metabolomics, Agilent Technologies, Metabolomic Discoveries (now Metabolon) and waste water analytics in a pharmaceutical company Pliva.

ivana.blazenovic@gmail.com