BOOK OF ABSTRACTS

https://caismd.indiayouth.info/

EDITORS:

Fidele Ntie-Kang (Buea, Cameroon)
Jutta Ludwig-Müller (Dresden, Germany)

Contents

FREE ONLINE WORKSHOP ON SECONDARY METABOLITE DISCOVERY

Theme : Computational Applications in Secondary Metabolite Discovery (CAiSMD), (08-10 March 2021)

You will find all the workshop related links under menu CAiSMD (https ://caismd.indiayouth.info/).

What to expect

In 3 days, the participants of this online workshop will take a survey of modern computer-based approaches for exploring natural products (or secondary metabolites) discovery in the omic age. Selected experts will give keynote lectures, hands on sessions, round table discussions, and oral presentations. Selected applicants (mostly early career scientists and students) will be invited to give oral presentations (15 minutes) and present posters in the form of (5 minutes) flash presentations.

Audience

M.Sc. and Ph.D. students, postdocs and early career researchers interested in bioinformatics, chemoinformatics, natural products chemistry, computational drug design, genomic analysis, with applications in drug discovery.

Registration

Participants should fill out the following registration form (https ://indiayouth.info/index.php/caismd/registration). Those who plan to do a presentation during the workshop should upload an abstract or poster accordingly using the prescribed abstract template. All applications for participation must be received by the deadline on 05 March 2021 by midnight CET.

Deadlines

Abstract for oral presentations (keynotes, standard oral presentations and flash presentations) must be received by midnight on 24 February 2021 CET. The final program of the workshop will be published by 05 March 2021.

Keynote presentations/Round table discussions

Selected experts will be directly invited by the organizers. Keynote lectures and round table discussions will last from 30 to 60 minutes.

Hands-on sessions

Selected experts will be directly invited by the organizers to give lectures and practical sessions on specific software tools or web servers. Each session will last for 90 minutes.

Oral presentations

This will last up to 15 minutes each and will be selected among applicants who submit an abstract using the provided template by the deadline of 24 February 2021 by midnight (CET).

Poster/flash presentations

Selected posters from M.Sc. and Ph.D. students will be presented in 5 minutes flash presentations. All submissions will be received by the expected deadline (24 February 2021 CET by midnight). Keynote presentations, round table discussions and hands on sessions are upon invitation by the workshop organizers. Presentations not accepted as standard (15 minutes) oral presentations will automatically be assigned as flash presentations or rejected.

Lecture options and web platform

All lecture slides will be uploaded by presenters 24 hours before the workshop and will be available for download by the workshop participants.

Certificates of participation

Participants who attend at least 60% of all lectures and take part in an online post-workshop survey will receive a signed certificate of participation.

Cost of participation

Free of charge

Language of workshop

English

Access to the Meeting

The general sessions (**keynote lectures**, **oral presentations** and **round table discussions**) are accessible through zoom. Information for access to the parallel hands-on sessions will be communicated by the presenters subsequently.

Meeting link : https ://tu-dresden.zoom.us/j/89855518329 ?pwd=TFVPTFFRT1lyVncyR2pXZlhXSTN3UT09

Meeting-ID : 898 5551 8329

Passcode : 33070891

Hands-on Sessions (Wednesday March 10, 10:00-11:30 am)

Session 1: Topic: Mining the Plant Specialized Metabolome with Mass Spectrometry: Library Matching and Molecular Networking with GNPS (**J.J.J. van der Hooft**)

Meeting link: https://us02web.zoom.us/j/83349814084?pwd=NWkrQnVtWW05NzVpd0hPNVJqSThPdz09

Meeting ID: 833 4981 4084

Passcode: 592037

Session 2: Topic: Virtual Identification of Bioactive Natural Products from African Flora (**T.M. Musyoka, D.M. Shaddrack, and F. Ntie-Kang**).

Meeting link: https://us02web.zoom.us/j/87895629764?pwd=V0VQM0dSME40Y1FOcEpUbUxDQU5aZz09

Meeting ID: 878 9562 9764

Passcode: 604210

Session 3: Topic: *In Silico* Prediction and Identification of Metabolites with BioTransformer : Enabling Secondary Metabolite Discovery (**Y. Djoumbou Feunang**).

Meeting link: https://corteva.zoom.com/j/98524404888?pwd=dHhhSnNTeGp0a2ZQZU9hRVRlY3FoUT09

Passcode: 401148

Session 4: Topic: New E-Resource for Drug Discovery (NERDD) (**C. Stork and N. Mathai**)

Meeting link: https://uib.zoom.us/j/61446200663?pwd=OElUMjRDaGJJQi9kbVZ6VU1CQTlpQT09

Meeting ID: 614 4620 0663

Passcode: z0rJxCK1

Session 5: Topic: Cheminformatics for Natural Products with the CDK (**M. Sorokina and C. Steinbeck**).

Meeting link: https://uni-jena-de.zoom.us/j/4830914532

Meeting ID: 483 091 4532

Speakers

Özlem Taştan Bishop

Özlem is a Full Professor in structural bioinformatics at Rhodes University. She received her B.Sc. degree in Physics from Boğaziçi University, Istanbul, Turkey. Then she moved to the Department of Molecular Biology and Genetics at the same University for her M.Sc. degree. She obtained her Ph.D. from Max-Planck Institute for Molecular Genetics and Free University, Berlin, Germany in 2003. While doing her Ph.D., Özlem became interested in structural biology, and during her postdoctoral positions (Texas University, USA ; University of Western Cape and University of Pretoria, South Africa) she gained experience in structural bioinformatics as well as structural biology. In 2009, Özlem took up an academic position at Rhodes University, South Africa. She established the Research Unit in Bioinformatics (RUBi) in 2013. She has graduated up to 13 Ph.D. and over 30 M.Sc. students since she joined Rhodes University. Özlem's broad research interest is structural bioinformatics and its applications to drug design and development. Her recent interest is in the allosteric mechanisms of proteins and understanding the effects of non-synonymous single nucleotide variations on protein structure and function, in order : to decode the underlying causes of many inherited diseases ; to uncover drug resistance mechanisms ; and to investigate drug sensitivity issues in certain populations for precision medicine purposes. She has published over 75 research articles.

Vanderlan da Silva Bolzani

Vanderlan da Silva Bolzani is a Full Professor at São Paulo State University (UNESP). Fellow of the Royal Society of Chemistry (UK), member of the Brazilian Academy of Science (ABC) and São Paulo Academy of Science (ACIESP), is also Science Productivity CNPq Fellow level 1A, and has received several awards; the most recently being Distinguished Woman in Science Chemistry and Chemical Engineering, conceived by ACS & IUPAC and Simão Mathias Medal and Elsevier-Capes in 2013. With a Ph.D. in Organic Chemistry, under the guidance of Professor Otto Richard Gottlieb, at University of São Paulo, in 1990 she was awarded a fellowship from the DAAD for a short training at the University of Hannover. After a post-doctorate at Virginia Polytechnic Institute (VPISU-USA) under the guidance of Professor David Kingston, she joined the São Paulo State University (UNESP), and since 2003, is a member of the Biota-FAPESP Program Coordination. Currently, she is the Director of the UNESP Technology Transfer Office (UNESP-AUIN). Her field of interest is plant science, and she has been involved in the isolation, bioactivity and function of secondary metabolites and peptides from plants. Also she has studied the biosynthesis of piperidine alkaloids, and recently she has been involved in metabolomics of medicinal plants. She has had strong involvement in human resource training, with over 50 Master's and Ph.D. students and several post-doctoral supervisions. She has been invited to give lectures and seminars worldwide. She has published more than 203 articles (h-index = 31, 3,565 citations), 5 book chapters and 7 patents. She also has strong work collaboration with National Pharmaceutical and Cosmetic Industries looking for new drugs from plant species. Dr. Bolzani was the President of the Brazilian Chemical Society from 2008-2010, and is currently a Counselor. She serves as a member of Editorial Boards of several scientific journals, e.g. *Journal of Natural Products*, *Natural Products Reports*, *Phytochemistry Letters*, and *Journal of Ethnopharmacology*. Since 2011 is a Visiting Professor at the Enzymologie Moléculaire et Fonctionnelle UR4-UPMC, Paris, and in 2012 she became a member of the L'Oreal Scientific Advisory Board. In October 2013 was elected to the World Academy of Science for the Advancement of Science in Developing Countries (TWAS).

Ana L. Chávez-Hernández

Ana Luisa Chávez-Hernández is a Ph.D. student in Chemical Science under the supervision of Dr. José Luis Medina-Franco, leader of DIFACQUIM Research Group, Department of Pharmacy of School of Chemistry at Universidad Nacional Autónoma de México. She has a Master of Chemical Science and a B.Sc. in Food Engineering. Currently, her research focuses on the development of libraries from natural products that can serve as building-blocks for the *de novo* design of natural products that are already reported in databases.

Miquel Duran-Frigola

Miquel's research interests lay at the intersection between drug discovery and large-scale biological data analysis. During his Ph.D. studies and early career, Miquel developed state-of-the-art *in silico* methods, producing over 30 scientific publications in a broad range of topics, from theoretical chemistry to phenotypic screening. Along this process, he has worked at IRB Barcelona, the Massachusetts Institute of Technology (MIT), Tel Aviv University, IS Global-CISM (Mozambique) and CIDRZ (Zambia). Recently, Miquel has founded the Ersilia Open Source Initiative. With Ersilia, Miquel aims to apply his expertise to the context of open science and global health, accelerating the discovery of treatments for neglected clinical needs. Miquel is the lead scientist behind the development of the Chemical Checker, a computational drug discovery tool based on the systematic integration of diverse bioactivity data.

Romuald Tematio Fouedjou

Romuald Tematio Fouedjou is currently a Visiting Researcher at the AgroEcoHealth Platform of the IITA (Benin) after his work at the Department of Chemistry (Research Unit of Environmental and Applied Chemistry, University of Dschang, Cameroon), where he received the Bachelor's, Master's and Ph.D. degrees in chemistry, specializing in organic chemistry/natural products chemistry in 2017. His research interests include bioactive secondary metabolites from medicinal plants and chemical transformations. He is also involved in various projects dealing with the *in silico* investigation of natural products from plants and pesticides in the environment.

Yannick Djoumbou Feunang

Yannick is a Data Scientist – Cheminformatician at Chemistry Data Science Research Group Corteva Agriscience, Indianapolis, Indiana. Dr. Yannick Djoumbou Feunang earned his Ph.D. in Microbiology and Biotechnology at the University of Alberta - Canada, in 2017, where his research focused on developing cheminformatics tools to enhance metabolomics. Some of his main contributions include software tools (e.g. ClassyFire, BioTransformer, and CFM-ID 3.0), with applications of ontology and linked data, as well as machine-learning, and knowledge-based artificial intelligence to biology and chemistry. Additionally, he has contributed to the development of databases such as DrugBank and HMDB. Since 2018, Dr. Djoumbou Feunang has worked as a Research Investigator for the Chemistry Data Science research group at Corteva Agriscience in Indianapolis, Indiana. His responsibilities include among others : (1) the development of machine learning models to support lead generation and optimization projects, and ; (2) the enhancement of Corteva's cheminformatics scientific computing platform. He also currently leads a project aiming at building a cutting-edge, adapted *in silico* metabolism platform at Corteva Agriscience.

Samuel A. Egieyeh

Dr. Samuel Ayodele Egieyeh is a seasoned and highly experienced pharmacist (qualified in 1997) with Bachelors and Masters degrees in Pharmacy, and a Ph.D. in Bioinformatics. He also has a post-graduate diploma in clinical research and drug development from the University of Basel (Switzerland). He is currently a senior lecturer in the discipline of Pharmacology and Clinical Pharmacy, School of Pharmacy, University of the Western Cape, Cape Town South Africa. He leads the Computational Pharmacology and Cheminformatics Group (CPCG). His research focuses on computational drug discovery and design ; data science (including cheminformatics, bioinformatics, machine learning and biostatistics techniques) for predictive drug development and precision medicine for infectious and non-communicable diseases. His presentation will show the use of data science in drug discovery and development, therapeutics, and precision medicine. The presentation will also introduce the "University of the Western Cape's Computational Pharmacology and Chemoinformatic Platform" that is used to analyze hit compounds from high-throughput assays (HTS) available in public bioactivity databases. The platform provides data analytics and data mining tools for researchers in drug discovery and design to analyze and mine useful data from their *in vitro* experiments in order to make rational and viable drug discovery and design decisions.

Justin J. J. van der Hooft

Justin van der Hooft is an Assistant Professor of Bioinformatics at Wageningen University, The Netherlands. He obtained a B.Sc. (2004) and M.Sc. (2007) in Molecular Sciences (Wageningen University). In 2012, he obtained his Ph.D. at the Biochemistry and Bioscience groups in Wageningen. He then moved to Glasgow, UK, for postdoctoral positions with Alan Crozier and Michael Barrett, subsequently. During his time in Glasgow, he developed mass spectrometry-based workflows to enhance metabolite annotation and also adopted molecular networking as a network-based annotation strategy. In particular, Justin set up a collaboration with Simon Rogers to develop MS2LDA for unsupervised substructure discovery in metabolomics data – based on topic modelling algorithms invented for text-mining approaches. In 2017, he took up a shared postdoctoral position between Marnix Medema and Pieter Dorrestein to explore the combination of genome and metabolome mining to accelerate natural product discovery. In January 2020, he started his own group in Wageningen that will develop computational metabolomics methodologies to decompose complex metabolite mixtures into their (sub)structures inspired by approaches in natural language processing and genomics. Justin will apply these workflows to study plant and microbiome-associated metabolites and the food metabolome.

Johannes Kirchmair

Johannes Kirchmair is an Assistant Professor in cheminformatics at the Department of Pharmaceutical Chemistry of the University of Vienna and head of the Computational Drug Discovery and Design Group (COMP3D). He also is a group leader at the Center for Bioinformatics (ZBH) of the University of Hamburg. After earning his Ph.D. from the University of Innsbruck (2007), Johannes started his career as an application scientist at Inte :Ligand GmbH (Vienna) and as a university assistant at his alma mater. In 2010 he joined BASF SE (Ludwigshafen) as a postdoctoral research fellow. Thereafter he worked as a research associate at the University of Cambridge (2010-2013) and ETH Zurich (2013-2014). Johannes held a Junior Professorship in applied bioinformatics at the University of Hamburg (2014 to 2018) and an Associate Professorship in bioinformatics at the University of Bergen (2018 to 2019). He has been a Visiting Professor or Lecturer at the National Institute of Warangal (2016), the University of Cagliari (2017) and the University of Vienna (2018). His main research interests include the development and application of computational methods for the prediction of the biological activities, metabolic fate and toxicity of small molecules (including natural products) in the context of drug discovery.

Jutta Ludwig-Müller

Prof. Jutta Ludwig-Müller is one of the organizers of the online workshop "Computational Applications in Secondary Metabolite Discovery (CAiSMD)". From 1999, she has been a Professor of Plant Physiology at the Technical University Dresden (Germany). After studying Biology at Frankfurt (Main) leading to a Diploma thesis on the topic "Auxin biosynthesis in Brassicaceae" (1981-1986), she completed her PhD dissertation in 1990 at the Botanical Institute of the Johann Wolfgang Goethe Universität (JWGU) in Frankfurt (Main) on the topic "Auxin biosynthesis and the role of auxins during the clubroot disease", with several visits to the laboratory of Dr. Ephraim Epstein, Volcani Center, Israel. After a research assistanship and postdoctoral stay in the USA, she completed her Habilitation in 1996. From 1995 to 1999, she was an Assistant Professor at the Botanical Institute of the JWGU. During this period she made several visits to the laboratory of Dr. Jerry D. Cohen, United States Department of Agriculture, Beltsville, Maryland (USA). Her research focuses on mechanisms leading to plant-pathogen (clubroot disease) interactions or the interaction with beneficial microorganisms (fungal endophytes). Her group is interested in the role of plant hormones and secondary plant products for development of pathogenic or symbiotic interactions of fungi with host plants. For better understanding of the events leading to pathogenesis, the team have begun to identify differentially expressed genes during host/pathogen interaction. She is also interested in the role of auxins and glucosinolates during plant development and therefore they study the biosynthesis, metabolism and regulation mechanisms controlling the endogenous concentrations of these compounds. In addition to glucosinolates, her team also investigates the bioactivity of other secondary plant metabolites such as flavonoids and phenylpropanoids. In 2019, she was awarded the Spiridion Brusina Medal of the Croatian Society of Natural Sciences. She is a member of several scientific organizations, including the American Society of Plant Biology (ASPB). the German Botanical Society (DBG), Deutscher Hochschulverband (DHV) and the International Plant Growth Substances Association. She is a member of the editorial boards of several journals, including being the Editor-in-Chief of the journal *Actualia*, (published by the German Botanical Society) and Editor-in-Chief of the journal *Journal of Plant Growth Regulation* (Springer).

External link :

https ://tu-dresden.de/mn/biologie/botanik/pflanzenphysiologie/die-professur

Neann Mathai

Neann Mathai is a Ph.D. student in Cheminformatics under the supervision of Assoc. Prof. Johannes Kirchmair at the Computational Biology Unit and the Department of Chemistry at the University of Bergen (Norway). She has a Master of Science in Computer Science (focused on computational glycobiology and vaccine development under Assoc. Prof. Michelle Kuttel) and a B.Sc. in Computer Science and Chemistry, both from the University of Cape Town, South Africa. She also spent an extended period of time in the United States where she worked on the statistical analysis of public health, epidemiological, and program management data and on the development of health informatics tools for large multinational organizations. Neann's main research focuses are on the development and validation of target predictions methods and the development of methods to curate small molecule libraries for screens.

Marnix Medema

Marnix Medema is an Assistant Professor of Bioinformatics at Wageningen University, The Netherlands. He obtained a Biology B.Sc. (Radboud University Nijmegen, 2006) and a Biomolecular Sciences M.Sc. (University of Groningen, 2008). In 2013, he completed his Ph.D. with Eriko Takano and Rainer Breitling in Groningen ; during this period, he was also a visiting fellow with Michael Fischbach at the University of California, San Francisco. Following a postdoc at the Max Planck Institute for Marine Microbiology in Bremen, Germany, he joined Wageningen University in 2015. There, his group develops computational methodologies to unravel natural product biosynthesis using omics data, and applies these methods to the study of molecular interactions in microbiomes. From 2020, his group is supported by an ERC Starting Grant.

José L. Medina-Franco

Dr. Medina-Franco holds a B.Sc. in Chemistry (1998, National Autonomous University of Mexico (UNAM), a M.Sc. and Ph.D. degree (2005, both from the UNAM). In 2005, Dr. Medina Franco joined the University of Arizona as a postdoctoral fellow under the supervision of Prof. Gerald Maggiora and he was named Assistant Member at the Torrey Pines Institute for Molecular Studies in Florida in August 2007. In 2013, he conducted research at the Mayo Clinic. In 2014 he joined UNAM and now is Full Time Research Professor. He leads the DIFACQUIM research group at UNAM. The research focus is on chemoinformatics and computer-aided drug design with applications on epigenetic targets and natural products. He is a member of the National Researcher System, National Council of Science and Technology in Mexico with the highest level, III. In 2016, he was appointed as Research Collaborator of the Mayo Clinic and in 2017 he was named Fellow of the Royal Society of Chemistry (UK) and was Visiting Professor of the University of Montreal at Quebec in 2019. José has published 225 peer-reviewed papers, 23 books chapters and issued one international patent. He has edited the books Epi-Informatics and Food Informatics. He serves as guest editor of the *Journal of Chemoinformatics*, *Frontiers in Chemistry* and *Biomolecules*, and is member of the Scientific Advisory Board of the company Epigenavir, LLC (USA). Since 2007 he has been PI in several research grants funded by government institutions and pharmaceutical companies. Dr. Medina Franco has supervised the research of 8 postdoctoral fellows, 5 Ph.D. students, 5 Masters, and 18 Undergraduate students. He teaches the course Chemoinformatics at UNAM.

Related links :

Research Group : https ://www.difacquim.com/

Research Gate : https ://www.researchgate.net/profile/Jose_Medina-Franco

LinkedIn : https ://www.linkedin.com/in/jose-l-medina-franco-0b653315/

Google Scholar : https ://scholar.google.com/citations ?user=xvyuVTYAAAAJ

Thommas M. Musyoka

Dr. Thommas Mutemi Musyoka has a Ph.D. in bioinformatics from Rhodes University, South Africa. He is involved in several computational studies for early stage identification of potential hits against infectious agents as well as deciphering their underlying mechanisms of drug resistance. He is also involved in the maintenance and upgrading of the South African Natural Compound Database (SANCDB). Currently, he is working as a postdoctoral research fellow in the Research Unit of Bioinformatics (RUBi), Rhodes University where he also mentors several postgraduate students working on different computational studies.

Fidele Ntie-Kang

Fidele Ntie-Kang is one of the organizers of the online workshop "Computational Applications in Secondary Metabolite Discovery (CAiSMD)". He currently heads the Molecular Simulations Laboratory, Chemistry Department, University of Buea (Cameroon). He studied Chemistry at the University of Douala (Cameroon) from 1999 to 2005, leading to Bachelor's and Master's degrees. His Ph.D., received from the University of Douala in 2014, was based on molecular modeling of anti-tubercular drug target to design novel inhibitors. This was followed by an Habilitation in Pharmaceutical Chemistry from Martin-Luther University Halle-Wittenberg, Halle (Saale) (Germany), under Prof. Wolfgang Sippl. He is concurrently a Guest Professor or Lecturer at the Technical University, Dresden (Germany), an Adjunct Professor at the Nelson-Mandela African Institution of Science and Technology, Arusha (Tanzania) and has previously taught a course on "Drug Design" in the Masters Program in Pharmaceutical Biology, University of Applied Sciences, Zittau/Görlitz (Germany). His current focus is the discovery of bioactive natural products from African flora by the use of virtual screening followed by *in vitro* assays. A major contribution of his research team has been the development of the African natural products database. He has authored or co-authored more than 80 publications and has served as Associate Editor in several journals, as Guest Editor in special issues in the journals *Molecules* and *Frontiers in Pharmacology*. He holds several awards, including the Georg Forster Fellowship from the Alexander von Humboldt Foundation (Germany) and is a member of several academic societies, including the American Chemical Society and the Cameroon Academy of Young Scientists. He is the editor of the book series "Chemoinformatics of Natural Products" published by Degruyter.

Victor C. Osamor

Victor C. Osamor is a Full Professor of Computer Science, former Head of Department, Computer and Information Sciences and Coordinator/ Director, Covenant University Center for Information Technology (CU CIT). Prof. Osamor is also the Applied Research Coordinator for World bank-6 million dollar funded Covenant Applied Information and Communication - African Center of Excellence (CApIC-ACE) research grant project with experience and previous involvement in NIH, NSF, H3Africa and H3Abionet research projects. He has experience in postgraduate supervision with graduated and current M.Sc. and Ph.D. students under his training. Prof. Osamor is an Adjunct Professor and External Examiner to other Universities and a Grant Reviewer for the Executive Government Agency of National Science Centre (Narodowe Centrum Nauki -NCN), Poland. He researches in Bioinformatics, Artificial Intelligence, Data analytics with other interest in Software techniques, Cyber security, Computer graphics and animation with certification in Artificial Intelligence, Predictive Analytics in IBM SPSS Modeller, Microsoft Technical Associate (MTA) in Database, Oracle Database and SQL programming. Prof. Osamor is a Marie Curie Fellow and belongs to Marie Curie Alumni Association (MCAA), having experience in European Commission/EU Co-funded projects with European Research Center for Informatics and Mathematics (ERCIM). Prof. Osamor is a book author, co-editor of a textbook, reviewer and has several publications in reputable journals to his credit especially in Bioinformatics. His Ph.D is in Computer Science and Bioinformatics and a Post-Doctorial fellowship is at University of Warsaw, Poland.

Vaishali M. Patil

Dr. Vaishali Manikrao Patil is an Associate Professor of Pharmaceutical Chemistry at KIET Group of Institutions, Ghaziabad. She obtained her undergraduate degree in pharmacy (2000) from North Maharashtra Technical University, Jalgaon and postgraduate in pharmaceutical chemistry (2007) from Dr. A. P. J. Abdul Kalam Technical University, Lucknow. Dr. Vaishali received her doctorate in Pharmaceutical Sciences from Birla Institute of Technology, Mesra in 2014. Dr. Vaishali has diversified experience of 15 years in academics, research and pharmaceutical industry. She is experienced in medicinal chemistry, computer aided drug design (CADD), QSAR and application of CADD tools for drug discovery and development. She is supervising three Ph.D. scholars and has supervised more than 20 M.Pharm. students. Her lab has received funds from various funding agencies like SERB-DST, BIRAB-DBT, and AKTU. Dr. Vaishali has made over 65 scholarly contributions, including nearly 56 peer-reviewed research papers (citations = 321 ; h-index = 10 ; i10-index = 11 ; Aug 2020). She has authored 03 books published including one by an international publisher. She has contributed 13 chapters in books published by Elsevier, Springer and Nova Science Publishers. She has presented her research work at various national and international platforms organized in the United States, Japan and Hungary. She is also serving as Editorial Board Member of scopus-indexed, peer-reviewed journal "Anti-Infective Agents" and "Coronaviruses".

Daniel M. Shaddrack

Daniel holds a Ph.D. in Computational Molecular Biophysics from NM-AIST, Tanzania and ICTP, Trieste, Italy. He is the Head of the Department of Chemistry, St John's University of Tanzania. Since 2018, Daniel has acted the role as a Director and co-organizer of the Biophysics /Biophysical Chemistry school/workshop in Tanzania/Africa. He is a member of various societies including ; Tanzania Chemical Society (TCS) and African Physical Society (AfPS). His areas of research focus on understanding the role of solvents in drug design, i.e. protein-ligand interaction, conformation of small molecules, protein conformations and drug delivery systems.

Maria Sorokina

Dr. Maria Sorokina is a Chem- and Bioinformatician in the Steinbeck group at The Friedrich-Schiller University in Jena, Germany. Her main research interests are natural products cheminformatics, challenges in storing complex biological data, deciphering enzyme functions and FAIR research data management. Dr. Sorokina studied at the University Paris-Saclay, where she obtained first her bachelor's degree in molecular and Cellular Biology and then her master's in bioinformatics and Biostatistics. She did her Ph.D. in bioinformatics at the Genoscope, the French National Sequencing Centre on new metabolic network representation for the discovery of metabolic pathways. After a short stay in Cologne, Germany, working on enzyme promiscuity in yeasts, she joined as a cheminformatician and data manager the collaborative research center ChemBioSys and the Steinbeck group at the Institute for Inorganic and Analytical Chemistry at the Friedrich-Schiller University in Jena.

Conrad Stork

Conrad Stork is a Ph.D. student in the Computational Drug Discovery and Design Group (COMP3D) group of Johannes Kirchmair at the Center for Bioinformatics (ZBH) at University of Hamburg. He completed a Master of Science in theoretical Chemistry at the University of Hamburg in the group of Prof. Carmen Herrmann where he investigated spin dynamics of surfaces using a first principles approach. He then changed research areas from theoretical Chemistry to cheminformatics and web development with a focus on understanding and predicting Frequent Hitters in biochemical assays. Conrad is also the lead developer of the New E-Resource for Drug Discovery (NERDD), a web server which hosts key cheminformatics tools developed for public use.

Kiran K. Telukunta

In 2018, Kiran completed his doctoral research in Pharmaceutical Bioinformatics at the University of Freiburg (Germany), while managing several Scientific IT infrastructure for Bioinformatic and Cheminformatics scientific projects, under Prof. Dr. Stefan Günther. Since then, he has worked on several projects, including as a Backend Developer and Software Engineer at Flowers-Software GmbH and Scientific Cloud Coordinator at VIB in Gent (Belgium). He has previously taught a course on "Drug Design" in the Master's Program in Pharmaceutical Biology, University of Applied Sciences, Zittau/Görlitz (Germany). He is Information System Administrator at ISES, a voluntary organisation in Freiburg (Germany). He is the One of the Directors of the TMS Foundation, which voluntarily prepared the submission platform for this online workshop. During the workshop, he plans to present some general guidelines and underlying principles of gene cluster analysis with applications using Galaxy.

Marilia Valli

Marilia is a postdoctoral researcher at the Sao Carlos Institute of Physics - University of Sao Paulo (IFSC-USP), working with the design of bioactive compounds for the treatment of neglected diseases and the development of NuBBE database, the first database of natural products from Brazilian biodiversity. Her main goal is to promote the use of the chemical diversity of Brazilian biodiversity in drug discovery. She obted her Ph.D. in Chemistry in 2014 with multidisciplinary experience on medicinal chemistry, natural products chemistry, Nuclear Magnetic Resonance (NMR) and organic chemistry. She has published 26 scientific papers and book chapters and has presented several of her results in national and international conferences. She is highly motivated in developing collaborative work and is currently volunteering as chair of the Younger Researchers committee of SBQ (Brazilian Chemical Society) and is an active member of the Younger Chemists committee (IYCN) at the IUPAC.

Tilmann Weber

Tilmann Weber is a Professor for Natural Products Genome Mining and Associate Scientific Director of the "Natural Products Genome Mining Group" at the Novo Nordisk Foundation Center for Biosustainability of the Technical University of Denmark. His main research interest is focused on deciphering the molecular pathways and engineering the biosynthesis of natural products by combining genetic, biochemical and bioinformatics methods. He is a pioneer in developing software for the automated genome mining (CLUSEAN, antiSMASH, antiSMASH-DB) and analysis of secondary metabolite biosynthetic pathways. His group was able to firstly elucidate the biosynthetic pathways of the elfamycin family of antibiotics and is deeply involved in developing CRISPR-based metabolic engineering tools for actinomycetes. Tilmann Weber is a member of the Editorial Board of Scientific Reports, Metabolic Engineering, and Associate Editor for Synthetic and Systems Biotechnology.

Ludger A. Wessjohann

Professor Wessjohann studied chemistry in Hamburg (Germany), Southampton (UK), and Oslo (Norway, Prof. Skattebøl). He earned his doctorate in 1990 with Prof. de Meijerein Hamburg. After a short period as lecturer in Brazil, he became a postdoctoral Feodor-Lynen fellow of the Alexander von Humboldt Foundation with Prof. Paul Wender at Stanford University (USA), working on the total synthesis of Taxol®. After an Assistant Professorship in Munich (LMU, 1992-1998), he was appointed to the Chair of Bioorganic Chemistry at the Vrije Universiteit Amsterdam (The Netherlands), working on organometallic chemistry and biocatalysis. Since 2001, he is director of the Department of Bioorganic Chemistry at the Leibniz Institute of Plant Biochemistry (IPB) in Halle (Germany), and in parallel holds the chair of natural product chemistry of the Martin Luther-University Halle-Wittenberg. From 2010-2017 he served as the Managing Director of the IPB (www.ipb-halle.de). Prof. Wessjohann focuses on the discovery, synthesis and application of natural products and bioactive derivatives thereof. He has over 400 publications, 30 patent applications, and is co-founder of six companies. He is a member of many boards and commissions, including recently the "mision de sabios" of the Colombian government. He received numerous scholarships, prizes and honors, e.g. Microsoft IT Founders Award, and is a foreign member of the Brazilian Academy of Science. Most recent chemoinformatics publication : "Evaluation of plant sources for antiinfective lead compound discovery by correlating phylogenetic, spatial, and bioactivity data (*Proc Natl Acad Sci*, **2020**, *117*, 12444-12451).

Director Dept. Bioorganic Chemistry & Full Professor of Chemistry Leibniz Institute of Plant Biochemistry (IPB)

Address : Weinberg 3, 06120 Halle (Saale), Germany

Tel : +49 (345)5582-1301, Fax : +49 (345) 5582-1309

E-mail : wessjohann@ipb-halle.de

http ://www.ipb-halle.de/en/research/bioorganic-chemistry/

Orcid : 0000-0003-2060-8235

Paul Zierep

Dr. Paul Zierep completed his bachelor's degree in chemistry as part of the binational French-German "Regio Chimica" program at the Universities of Mulhouse and Freiburg. He continued his studies in the master program Biochemistry and Biophysics. He just finished his Ph.D. in the bioinformatics group of Prof. Stefan Günther at the University of Freiburg. The main focus of his doctoral studies was the prediction of gene cluster metabolites based on computational approaches. The findings of the Ph.D. were compiled into a web server called SeMPI. SeMPI v1 focused on the combination of PKS prediction approaches with a database screening of the metabolites in the StreptomeDB v2. The second version was extended to NRPS products and included a large range of publicly available natural product databases. At the end of his Ph.D. he visited the group of Prof. Bjorn Peters at the La Jolla Institute for Immunology in California. There he applied his knowledge of cheminformatics to implement the first general approach for the prediction of non-peptidic epitopes. He is currently conducting a short term Post-Doc position at Prof. Stefan Günther's group, where he is further investigating the immunogenic potential of small molecules.

Schedule

VIRTUAL WORKSHOP ON PLANT-BASED SECONDARY METABOLITES

Theme : Computational Applications in Secondary Metabolite Discovery (CAiSMD)

Planned period

08-10 March 2021

Organizers

Jutta Ludwig-Müller, Technical University Dresden, Dresden, Germany

Fidele Ntie-Kang, University of Buea, Buea, Cameroon

Invited speakers

Vanderlan da Silva Bolzani, São Paulo, Brazil

Yannick Djoumbou, Indianapolis, IN, USA

Samuel A. Egieyeh, Cape Town, South Africa

Miquel Duran-Frigola, Cambridge, United Kingdom

Justin van der Hooft, Wageningen, The Netherlands

Johannes Kirchmair, Vienna, Austria

Marnix Medema, Wageningen, The Netherlands

José L. Medina-Franco, Mexico City, Mexico

Victor C. Osamor, Ota, Nigeria

Maria Sorokina, Jena, Germany

Özlem Tastan Bishop, Grahamstown, South Africa

Tilmann Weber, Lyngby, Denmark

Ludger A. Wessjohann, Halle (Saale), Germany

Many more ...

NB :

Draft Program Program follows Central European Time (CET)

HS : Hands-on Session, *KL* : Keynote Lecture, *OP* : Oral Presentation, *RTD :* Round Table Discussion.

Day 1

Chairs : *F. Ntie-Kang and J. Ludwig-Müller*

15 :30-16 :00

Registration finalization (for those still to register), welcome speech, general announcements by organizers, announcing the keynote speaker.

16 :00-16 :45

KL01 : Chemoinformatics in Bioactive Natural Product Discovery and Metabolomics Research (**L.A. Wessjohann**).

16 :45-17 :00

Announcements and closing remarks of day 1, closing.

Day 2

08 :00-08 :45

KL02 : Computational Tools for Biosynthetic Pathway Discovery in Plants (**M. Medema**).

08 :45-09 :00

Introducing Day 2 (Organizers).

Chair Morning Session : Bioinformatics Applications (***F. Ntie-Kang***).

09 :00-09 :15

OP01 : Application of Clustering on Secondary Metabolites (**K.K. Telukunta**).

09 :15-09 :30

OP02 : SeMPI 2.0 – A Web Server for PKS and NRPS Predictions combined with Metabolite Screening in Natural Product Databases (**P. Zierep**).

09 :30-09 :45

Coffee break

09 :45-10 :00

OP03 : OsamorSoft : A Tool for Clustering Genomic Data (**V.C. Osamor**).

10 :00-10 :15

OP04 : Natural Products Database from Brazilian Biodiversity, a Powerful Tool for Science, Technology and Innovation (**M. Valli and V. da Silva Bolzani**).

10 :15-10 :30

OP05 : Bio-/Chemoinformatics for Drug Discovery : UWC's Open Access Platform (**S.A. Egieyeh**).

10 :30-12 :00

RTD01 : (P. Zierep, M. Medema, T. Weber, and all speakers of morning session).

12 :00-13 :30

Lunch break

Chair Afternoon Session : Chemoinformatics Applications (***J. Ludwig-Müller***).

13 :30-13 :45

OP06 : Data-driven Prediction of Bioactivity Profiles for Uncharacterized Natural Compounds (**M. Duran-Frigola**).

13 :45-14 :00

OP07 : Artificial Intelligence /Machine Learning for Secondary Metabolite Prediction (**Y. Djoumbou Feunang**).

14 :00-14 :15

OP8 : Fragment Library of Natural Products and Compound Databases for Drug Discovery (**A.L. Chávez-Hernández and J.L. Medina-Franco**).

14 :15-14 :30

OP9 : *In Silico* Prediction of the Macromolecular Targets of Natural Products (**J. Kirchmair**).

14 :30-15 :00

Coffee break

Chair Evening Session : Outlook into the Future (***Ö. Tastan Bishop***).

15 :00-15 :15

OP10 : Challenges in Building NoSQL Databases for Natural Products Research (**M. Sorokina**).

15 :15-16 :00

KL 03 : Natural Products Genome Mining : Where Do We Go from Here ? (**T. Weber**).

16 :00-16 :15

Closing remarks, announcements.

Day 3

08 :00-08 :45

KL 04 : Bioinformatics Approaches for Early Drug Discovery : Tapping into the African Experience (**Ö. Tastan Bishop**).

08 :45-09 :00

Introducing Day 3 (organizers).

Chair Early Morning Session : Young Investigators Session (***M. Medema***).

09 :00-09 :15

OP11 : QSAR Studies on Some Anti-Cancer Kinase Inhibitors (**V.M. Patil**).

09 :15-09 :30

OP12 : Cameroonian Medicinal plants as Potential candidates of SARS-CoV-2 Inhibitors : *In Silico* Investigation (**R.T. Fouedjou**).

09 :30-10 :00

Coffee break

Chair Late Morning Session : Hands-on Sessions (***J. Ludwig-Müller***).

10 :00-11 :30

Parallel Hands-on Sessions (Open Source Tools to Enhance the Bio-/Chemoinformatic Assisted Natural Product-Based Drug Discovery).

Session 1

HS01 : Mining the Plant Specialized Metabolome with Mass Spectrometry : Library Matching and Molecular Networking with GNPS (**J.J.J. van der Hooft**).

Session 2

HS02 : Virtual Identification of Bioactive Natural Products from African Flora (**T.M. Musyoka, D.M. Shaddrack, and F. Ntie-Kang**).

Session 3

HS03 : *In Silico* Prediction and Identification of Metabolites with BioTransformer : Enabling Secondary Metabolite Discovery (**Y. Djoumbou Feunang**).

Session 4

HS04 : New E-Resource for Drug Discovery (NERDD) (**C. Stork and N. Mathai**).

Session 5

HS05 : Cheminformatics for Natural Products with the CDK (**M. Sorokina and C. Steinbeck**).

11 :30-13 :00

RTD2 : General Discussion on the Entire Workshop (Interactive Session Organized by Early Career Researchers Focused on the Essential Lessons to be Drawn from the Workshop). Chair (***S.A.T. Fobofou***)

13 :00-13 :15

Closing Remarks on the Workshop.

Abstracts of Keynote Lectures

KL01: Chemoinformatics in Bioactive Natural Product Discovery and Metabolomics Research

Ludger A. Wessjohann,[1,2*]
Wolfgang Brandt,[1] Mohamad A.A. Farag,[1,3] Lutz Weber,[4] Anne-Kathrin Hartig,[1] Katrin Franke,[1] Pauline Stark,[1] Laura Holzmeyer,[5] Alexandra Müllner-Riehl,[2,5] Jan Schnitzler[2,5] et al.

[1]Department of Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry (IPB), Weinberg 3, D-06120 Halle (Saale), Germany
[2]co-affiliation: German Centre for Integrative Biodiversity Research (iDiv)
[3]American University Cairo, Cairo, Egypt
[4]Ontochem GmbH, Halle (Saale) Germany.
[5]Institute of Biology, Leipzig University, Leipzig, Germany

[*]Corresponding author (Email: wessjohann@ipb-halle.de)

Background: Informatic methods are of increasing importance to extract knowledge from large data sets in chemical and life sciences, be it literature data, physicochemical and spectroscopic or biological data.

Content: I will present an overview, how various informatic methods and toolsdeveloped by us and with partners can support the selection of sources and the identification of natural products of relevance [1-3]. A focus will be on the chemoinformatic analysis of (1.) database and text large data analyses, and (2.) the role metabolic profiling will play in future bioactive compound discovery, going beyond classical isolation processes. Examples will include: A survey of the flora of Java, metabolic profiling of various medicinal plants, with an emphasis on *Hypericum* sp. (St. Johns worts).

Keywords: metabolomics, chemoinformatics, data mining, structure-activity-relationship, natural products, structure elucidation; NMR, MS, *Hypericum*, anticancer, antibiotics, neuroactives.

References:
[1] A. Feiner, N. Pitra, P. Matthews, et al. *Plant Cell Environ*, **2021**, *44*, 323-338.
[2] B. Michels, K. Franke, A. Weiglein, et al. *J Exp Biol*, **2020**, *223*, jeb223982.
[3] L. Holzmeyer, A.K. Hartig, K. Franke, et al. *Proc Natl Acad Sci*, **2020**, *117*, 12444-12451.

KL02: Computational Tools for Biosynthetic Pathway Discovery in Plants

Marnix H. Medema[1,*]
[1]Bioinformatics Group, Wageningen University, Wageningen, The Netherlands
[*]Corresponding author (Email: marnix.medema@wur.nl)

Plants, fungi and bacteria produce a wealth of specialized metabolites, which are of great importance from both ecological and clinical perspectives. Due to the accelerated accumulation of omics data, computational methods have become more and more important to identify these molecules and to assess their biological activities. Here, I will highlight the work performed in my research group on using these approaches to accelerate natural product discovery, as well as to study microbe-microbe and host-microbe interactions in human, plant and animal microbiomes. Specifically, I will discuss the use of computational approaches to investigate biosynthetic diversity across large numbers of genomes, and integrative genome/metabolome mining to link gene clusters to molecules.

KL03: Natural Products Genome Mining: Where do we go from here?

Tilmann Weber[1,*]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark
[*]Corresponding author (Email: tiwe@biosustain.dtu.dk)

Genome analyses of many microorganisms but also higher organisms indicate that the genetic potential to synthesize specialized metabolites is far beyond the number of molecules observed in traditional screenings. With the availability of cheap and easy-to-obtain whole genome sequences (at least for microorganisms), *in silico* genome mining has become an indispensable tool to complement the classical chemistry-centred approach to identify and characterize novel secondary / specialized metabolites. Since the initial release in 2011, the open source genome mining pipeline antiSMASH [1] (https://antismash.secondarymetabolites.org), which is developed in collaboration with the group of Marnix Medema (University of Wageningen) and many international contributors, has become one of the most widely used tools. We recently released version 6 of antiSMASH, including an improved user interface, new detection modules, a new cluster comparison tool, and many internal optimizations. Users can easily analyse genomic sequences for the presence of secondary metabolite biosynthetic gene clusters with antiSMASH. To provide extensive analysis options of the data generated with antiSMASH, we have extended the framework with the antiSMASH database [2], (https://antismash-db.secondarymetabolites.org/), a user-friendly application allowing to browse and query antiSMASH v5 annotations. In version 3, the database contains 147 517 high quality BGC regions from 388 archaeal, 25 236 bacterial and 177 fungal genomes. These basic genome mining technologies build the foundations of further in silico studies towards a more comprehensive “Genome Analytics” platform, which we use to streamline our natural product discovery and characterization efforts.

Keywords: antiSMASH; genome analytics; genome mining

References:

[1] K. Blin, S. Shaw, K. Steinke, et al. *Nucleic Acids Res*, **2019**, *47*, W81-W87.

[2] K. Blin, S. Shaw, S.A. Kautsar, et al. *Nucleic Acids Res*, **2021**, *49*, D639–D643.

KL04: Bioinformatics Approaches for Early Drug Discovery: Tapping into the African Experience

ÖzlemTastan Bishop[1,*]

[1]Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda, 6140, South Africa
[*]Corresponding author (Email: o.tastanbishop@ru.ac.za)

In the context of early drug discovery research, my group's main interests include the identification of novel and alternative drug targeting sites (i.e. allosteric sites) and of hit compounds for communicable and non-communicable diseases. We are also interested to understand the effects of nonsynonymous single nucleotide variations (snSNVs) on protein structure and function, in order: to assess the reasons behind many inherited diseases; to uncover the association to drug resistance mechanisms; to link to drug sensitivity issues in certain populations for precision medicine development; and many other applications. An understanding of the underlying resistance mechanism due to variations at molecular level is essential and can lead either to modifications of current approved drugs to get more effective ones or to the design of new inhibitors that overcome resistance mutations. This presentation will include examples of our recent work on understanding the underlying drug resistance mechanisms and identification of allosteric modulators as an alternative to orthosteric drugs.

Keywords: Allostery; communicable and non-communicable diseases; Nonsynonymous single nucleotide variations (nsSNV).

References:
[1] O. Sheik Amamuddy, G.M. Verkhivker, Ö. Tastan Bishop. *J Chem Inf Model*, **2020**, *60*, 5080-5102.
[2] O. Sheik Amamuddy, T.M. Musyoka, R.A. Boateng, et al. *Comput Struct Biotechnol J*, **2020**,*18*, 1103-1120.
[3] O. Sheik Amamuddy, W. Veldman, C. Manyumwa, et al. *Int J Mol Sci*, **2020**, *21*, 847.
[4] A. Amusengeri, L. Astl, K. Lobb, et al. *Int J Mol Sci*, **2019**, *20*, 5574.
[5] D.L. Penkler, Ö. Tastan Bishop. *Sci Rep*, **2019**, *9*, 1600.
[6] A. Amusengeri, Ö. Tastan Bishop. *Molecules*, **2019**, *24*, 188.

Abstracts of Oral Presentations

OP01: Application of Clustering on Secondary Metabolites

Kiran K Telukunta[1]

[1]Tarunavadaanenasaha Muktbharatonnayana Samstha Foundation, Hyderabad, India
[*]Corresponding author (Email: kiran.telukunta@indiayouth.info)

Background: Introducing analysis of clustering methods for secondary metabolites using Galaxy framework

Methods: One of the key approaches in analysis of data is recognizing the patterns of data. This becomes an arduously significant laborious process in unlabeled data. An approach to solve this kind of problem, which is also an unsupervised learning of data, is called clustering. Clustering of data sets containing DNA sequence data, gene expression data, and small molecules which are part of drug-like compounds of the chemical universe for drug discovery into groups of data points having similar patterns helps us to recognize more relevant information from the raw data [1].

Results: The evolution of clustering has spread widely and has given rise to many algorithms in reorganizing data into similar patterns. For each research task it becomes significant to understand the requirement of the data and to make similar points of the data to fall in the same cluster along with dissimilar data to fall in different clusters and choose accordingly suitable algorithms.

Conclusions: The presentation discusses clustering and various methods [2] and algorithms available along with few supporting cases [3]. These can be general guidelines in adopting clustering techniques in solving research problems.

Keywords: algorithms; clustering; DNA sequence data; Galaxy; gene expression data; secondary metabolites.

References:

[1] J. MacCuish, N. MacCuish, ***in*** Clustering in Bioinformatics and Drug Discovery, 1-3, **2011**
[2] A. Khanteymoori, A. Kumar, **2021** Clustering in Machine Learning (Galaxy Training Materials). /training-material/topics/statistics/tutorials/clustering_machinelearning/tutorial.html Online; accessed Wed Feb 17 2021
[3] K.K. Telukunta *Development and application of ligand-based cheminformatics tools for drug discovery from natural products*, **in**, PhD Thesis, Freiburg, Germany, **2018**.

OP02: SeMPI 2.0 – A Web Server for PKS and NRPS Predictions combined with Metabolite Screening in Natural Product Databases

Paul F. Zierep,[1] Adriana T. Ceci,[2] Ilia Dobrusin,[1] Sinclair Rockwell-Kollmann,[1] and Stefan Guenther[1,*]

[1]Pharmaceutical Bioinformatics, Albert-Ludwigs-University Freiburg, Germany
[2]Department of Cellular, University of Trento, Italy
[*]Corresponding author (Email: stefan.guenther@pharmazie.uni-freiburg.de)

Background: The constantly increasing amount of published genomic data provides the opportunity for an efficient identification of gene clusters by genome mining. SeMPI 2.0 [1] provides a comprehensive prediction pipeline, which includes the screening of the scaffold in publicly available natural compound databases.

Methods: The pipeline was completely redesigned, including improved profile hidden Markov models (pHMMs) and a novel algorithm for substrate specific prediction. Furthermore, we developed a novel natural products screening algorithm, that allows for the detection of homologous structures even for partial incomplete clusters.

Results: In a benchmark based on 559 gene clusters, SeMPI v2 achieved comparable or better results than antiSMASH v5.

Conclusion: The pipeline allows to link gene clusters to known natural products and therefore also provides a metric to estimate the novelty of the cluster if a matching scaffold cannot be found. Additionally, the SeMPI v2 web server provides features that can help to further investigate a submitted gene cluster, such as the incorporation of a genome browser, and the possibility to modify a predicted scaffold in a workbench before the database screening.

Keywords: Genome-Mining; Machine Learning; Nonribosomal Peptide Synthetases; Polyketide Synthases; Secondary Metabolites.

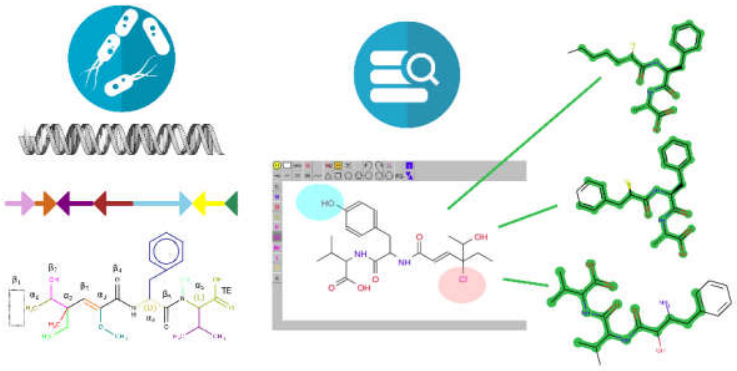


Figure 1: Graphical abstract taken form [1].

References:
[1] P.F. Zierep, A.T. Ceci, I. Dobrusin, et al. *Metabolites*, **2021**,*11*,13.

OP03: OsamorSoft: A Tool for Clustering Genomic Data

Victor Chukwudi Osamor[1*], and Theresa Okediya[1]

[1]Department of Computer and Information Sciences, Covenant University, Ota, Nigeria

[*]Corresponding author (Email: victor.osamor@covenantuniversity.edu.ng; vcosamor@gmail.com)

Background: Clustering of genomic datasets is often a way of processing through-put data to allow for intuitive unveiling of hidden patterns that connotes some status and functionalities in a system. Often, there are challenges of different clustering algorithms giving different results on the same dataset due to differences in the techniques in algorithmic development and assumptions. Our aim is to describe OsamorSpreadsheet [1] and employ different clustering techniques to cluster molecules from natural products databases [2-3] and investigate the level of cluster quality using our newly developed cluster validation tool called OsamorSoft [1]. This is to enhance the search protocol for a potential lead compounds for drug discovery.

Methods: The molecules under study were obtained from the African natural compounds database [2-3]. In the quest to identify a potential lead compound, we used Atom pair fingerprint to compute the molecular distances and the Tanimoto coefficient for the similarity measure with a threshold cutoff starting from 0.4. K-means clustering and hierarchical clustering tests were carried out for the purpose of comparison, resulting in the generation of hitmap for the purpose of visualization. We also employ OsamorSoft [1] to validate the cluster quality obtained during the experiment.

Results: As the molecular distances and ward's method are applied, the molecules were grouped into two clusters. The resulting compounds were reduced to 22 and were then clustered into two groups. We found that cpm1 and cmp4 were grouped while the other 17 compounds were included in the same cluster as cmp3 indicating consistency with the earlier clustering techniques used.

Conclusion: From the available preliminary result obtained in addition to the similarly property principle, molecules with similar structures are likely to have the same properties. This likely portends that they have similar properties and suggests that further test as a potential lead compounds may be necessary.

Keywords: clustering; lead molecule; natural products; OsamorSpreadsheet; OsamorSoft

References:

[1] I.P. Osamor, V.C. Osamor. *J Big Data*, **2020**, *7*, 48.

[2] C.V. Simoben, A.F.A. Moumbock, A. Qaseem, et al. *Mol Inf*, **2020**, *39*, 2000163.

[3] F. Ntie-Kang, K.K. Telukunta, K. Doering, et al. *J Nat Prod*, **2017**, *80*, 2067-2076.

OP04: Natural Products Database from Brazilian Biodiversity, a Powerful Tool for Science, Technology and Innovation

Marilia Valli,[1] and Vanderlan da Silva Bolzani[1,*]

[1]Nuclei of Bioassays, Biosynthesis and Ecophysiology of Natural Products (NuBBE), Department of Organic Chemistry, Institute of Chemistry, Sao Paulo State University – UNESP, Araraquara, Brazil
[*]Corresponding author (Email: vanderlan.bolzani@unesp.br)

The vast and biologically rich Brazilian biodiversity represents an outstanding opportunity for the development on the state-of-art of biology, chemistry sciences, medicinal chemistry, and in many other research areas. This vast biodiversity distributed in important endemic biomes and ecosystems, represents a powerful source for the creation of a database of natural chemical diversity useful for science, technology and innovation. Thus, secondary metabolites may contribute enormously to studies in several scientific areas and are important tools for data repositories, including drug discovery. Several strategies are used for the design of bioactive compounds and the use of databases has played an important role in this area. Brazilian biodiversity stands out in this context and the first library of natural products from Brazilian biodiversity (NuBBE Database) [1], provides valuable and integrative data for medicinal chemistry. Our research efforts have been engaged in developing NuBBE Database ($NuBBE_{DB}$, http://nubbe.iq.unesp.br/portal/nubbedb.html) [2], the first database of natural products from Brazilian biodiversity. A chemoinformatic analysis of $NuBBE_{DB}$ shows that $NuBBE_{DB}$ has a large percentage of drug-like compounds and support that this is a promising source of molecules for drug discovery and medicinal chemistry [3]. Aimed at developing this database, an innovative and unique collaborative research work was established with the Chemical Abstracts (CAS-ACS), the most important chemical database in the world to set the first comprehensive database of natural products in Brazil. Promoting the successful use of Brazilian biodiversity [4,5], would be a demonstration that technological innovation is achievable, supporting economic growth without a negative impact in the environment.

Keywords: Brazilian Biodiversity; Database; Natural Products.

References:
[1] M. Valli, R.N. dos Santos, L.D. Figueira, et al. *J Nat Prod*, **2013**, *76*, 439-444.
[2] A.C. Pilon, M. Valli, A.C. Dametto, et al. *Sci Rep*, **2017**, *7*, 7215.
[3] F.I. Saldívar-González, M. Valli, A.D. Andricopulo, et al. *J Chem Inf Model*, **2019**, *59*, 74-85.
[4] J.M. Galúcio, E.F. Monteiro, D.A. de Jesus, et al. *Comput Biol Chem*, **2019**, *83*, 107102.
[5] A.F. Vásquez, A. Reyes Muñoz, J. Duitama, et al. *J Biomol Struct Dyn*, **2020**, *13*, 1-15.

OP05: Bio-/Chemoinformatics for Drug Discovery: UWC's Open Access Platform

Samuel Egieyeh,[1,2,*] and Alan Christoffels[2]

[1]School of Pharmacy, University of the Western Cape, Cape Town, 7535 South Africa
[2]South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Cape Town 7535, South Africa
[*]Corresponding author (Email: segieyeh@uwc.ac.za)

In drug discovery, the journey from “hits” to “drug candidates” may be tedious, long and expensive. A high quality drug candidate must exhibit a balance of many properties, including potency, ADME (Absorption, Distribution, Metabolism and Excretion) and safety/toxicity. Hence a multi-parameter optimization strategy is required. However, the “hits” to “leads” optimization and “leads” to “drug candidates” optimization requires rigorous chemical-bioactivity data analytics, data mining and transformation of data into knowledge that will aid multi-parameter optimization as well as the rational selection of compounds with the highest chance of success. The field of cheminformatics provided the right tools and techniques to tackle this challenge to drug discovery. Here we present the role of cheminformatics in modern drug discovery. This presentation will define cheminformatics, justify the need for it in drug discovery, outline the cheminformatic resources and toolkits as well as delve into specific case studies that show the immense contributions of cheminformatics to drug discovery and design. The presentation will also introduce the “University of the Western Cape’s Computational Pharmacology and Chemoinformatics Platform” that is currently used to analyze hit compounds from high-throughput assays (HTS) available in public bioactivity databases. The platform provides an opportunity for researchers in drug discovery to analyze and mine useful data from their *in-vitro* experiments in order to make rational and viable drug discovery and design decisions.

Keywords: bioinformatics; chemoinformatics tools; drug discovery; natural products.

OP06: Data-driven Prediction of Bioactivity Profiles for Non-Characterised Natural Products

Miquel Duran-Frigola,[1,2,*] Martino Bertoni,[2] Pau Badia-i-Mompel,[2] Gemma Turon,[1] Edoardo Gaude,[1] and Patrick Aloy[2,3]

[1]Ersilia Open Source Initiative, Cambridge, UK
[2]Joint IRB-BSC-CRG Programme in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Catalonia, Spain
[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain
[*]Corresponding author (Email: miquel@ersilia.io)

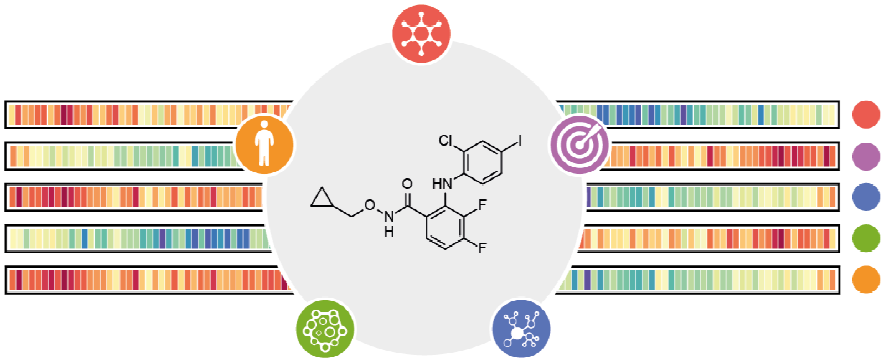
Background: The broad release of small molecule bioactivity data has prompted enriched representations of compounds, reaching beyond chemical structures and capturing their known biological properties. Unfortunately, 'bioactivity descriptors' are not available for most small molecules, which limits their applicability to a few thousand well characterized compounds.

Methods: We present a collection of deep neural networks able to infer bioactivity signatures for any compound of interest, even when little or no experimental information is available for them. Our 'signaturizers' relate to bioactivities of 25 different types (including target profiles, cellular response and clinical outcomes) and can be used as drop-in replacements for chemical descriptors in day-to-day chemoinformatics tasks.

Results: We illustrate how inferred bioactivity signatures are useful to navigate the chemical space in a biologically relevant manner, unveiling higher-order organization in natural product collections. Moreover, we implement a battery of signature-activity relationship models and show a substantial improvement in performance, with respect to chemistry-based classifiers, across a series of biophysics and physiology activity prediction benchmarks.

Conclusion: Large-scale inference of bioactivity profiles can set the basis for automated annotation of compound collections, including drugs, metabolites and natural products.

Keywords: Bioactivity signatures; Data-driven drug discovery.



OP07: Artificial Intelligence /Machine Learning for Secondary Metabolite Prediction

Yannick Djoumbou Feunang[1*]

[1]Corteva Agriscience, Indianapolis, Indiana, USA

*Corresponding author: (Email: yannick.djoumboufeunang@corteva.com)

In silico metabolism prediction tools provide a unique perspective to studying the chemical exposome, and how its changes affect the environment. Classical applications of such tools include, but are not limited to metabolite discovery, environmental fate prediction, ADMET profiling, and molecular design. Several approaches and methods to address the prediction of secondary metabolites have been described, and implemented in a comprehensive list of tools that include expert-, machine learning-, and QM-based systems, or hybrids thereof. In spite of the numerous reported successes, many limitations still hamper the wide adoption of those tools. In this presentation, we will describe the impact of artificial intelligence in the development of secondary metabolite prediction systems, along with the most commonly implemented approaches. Moreover, we will provide examples of the application of *in silico* metabolism prediction tools, such as BioTransformer, in the identification of secondary metabolites. Furthermore, we will discuss some of the prevalent limitations that hamper the widespread adoption of such tools, and propose solutions.

Keywords: expert system; deep learning; enzyme-substrate specificity; machine learning; metabolism prediction; metabolite identification; metabolic pathway; metabolomics; precision; recall; smarts; smirks.

References:

[1] Y. Djoumbou-Feunang, J. Fiamoncini, A. Gil-de-la-Fuente, et al. *J Cheminform*, **2019***, 11,* 2.

OP08: Fragment Library of Natural Products and Compound Databases for Drug Discovery

Ana L. Chávez-Hernández,[1,*] Norberto Sánchez-Cruz,[1] and José L. Medina-Franco[1]

[1]DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, México City, México
[*]Corresponding author (Email: anachavez3026@gmail.com)

Background: Natural Products (NP) possess unique functional groups [1]. However, few amounts from them were obtained during extraction and purification procedures. So, to maximize the use of NP, we propose to generate fragments libraries from NP that can be used such as building blocks for the synthesis of the so-called "pseudo NP"[2].

Methods: We generated fragment libraries from the COlleCtion of Open NatUral producTs (COCONUT), and other reference-data sets such as food chemical compounds (FooDB); compounds that not showed activity although that have been thoroughly tested (DCM), and two data sets related to COVID-19 research: a focused library from the Chemical Abstract Service (CAS) and inhibitors of the main protease of SARS-CoV-2 (3CLP). The fragments were generated using the algorithm of Retrosynthetic Combinatorial Analysis Procedure (RECAP), and then we calculated structural features such as the fraction of chiral carbons; fraction of sp^3 carbons; atoms of carbon, oxygen, nitrogen. Finally, we generated a visual representation of the chemical space for both compounds and fragments implementing the recently developed algorithm TMAP (Tree MAP) [3]. This method allows the visual representation of many molecules that are difficult to visualize using other standard methods such as principal component analysis (PCA).

Results: Molecular fragments retained the structural characteristics from original compounds which was consistent with the results of our previous work [4]. Both fragments and compounds from COCONUT were structurally more complex than reference data sets. Also, it was observed that the chemical space defined by the structural diversity of each data set is mostly defined by NP and DCM, both for compounds and fragments.

Conclusion: In general, molecular fragments retained their structural characteristics from original compounds. Also fragments would be used as building blocks in the *de novo* design of bioactive molecules despite that the original compounds have weak biological activity such as DCM [5].

Keywords: Chemoinformatics; COVID-19; drug discovery; natural products fragments; natural products informatics; SARS-CoV-2.

References:
[1] A. Pahl, H. Waldmann, K. Kumar. *Chim Int J Chem*, **2017**, *71*, 653–660.
[2] A. Christoforow, J. Wilke, A. Binici, et al. *Angew Chemie Int Ed*, **2019**, *58*, 14715–14723.
[3] D. Probst, J. L. Reymond. *J Cheminform*, **2020**, *12*, 12.
[4] A.L. Chávez-Hernández, N. Sánchez-Cruz, J.L. Medina-Franco. *Mol Inf*, **2020**, *39*, 2000050.
[5] A.L. Chávez-Hernández, N. Sánchez-Cruz, J.L. Medina-Franco. *Biomolecules*, **2020**, *10*, 1518.

OP09: Computational Prediction of the Macromolecular Targets of Natural Products

Johannes Kirchmair,[1*] Neann Mathai,[2] Ya Chen,[3] Christina de Bruyn Kops,[3] and Conrad Stork[3]

[1]Department of Pharmaceutical Chemistry, University of Vienna, 1090 Vienna, Austria
[2]Department of Chemistry and Computational Biology Unit (CBU), University of Bergen, N-5020 Bergen, Norway
[3]Center for Bioinformatics, Universität Hamburg, 20146 Hamburg, Germany
[*]Corresponding author (Email: johannes.kirchmair@univie.ac.at)

A key question in the research of natural products is the identificationof their macromolecular target. Pinpointing these interactions usually entails substantial experimental efforts, with uncertain outcomes. Computational methods have the potential to make a significant contribution towards target identification but most existing methods are focused on synthetic small molecules [1].In this contribution we will provide a timely overview of the state-of-the-art in computational target prediction, with focus on natural products research. We will discuss the scope and limitations of the available methods and models, and provide guidance on how to make best use of these methods and understand the reliability of predictions.The approaches covered by this contribution will include similarity-based methods, network-based approaches, machine learning methods and docking approaches.

Keywords: natural products; target prediction; cheminformatics.

Reference:
[1] Y. Chen, J. Kirchmair. *Mol Inf*, **2020**, *39*, 2000171.

OP10: Challenges in Building NoSQL Databases for Natural Products Research

Maria Sorokina,[1,*] and Christoph Steinbeck[1]

[1]Institute for Inorganic and Analytical Chemistry, Friedrich Schiller University, Jena, Germany
[*]Corresponding author (Email: maria.sorokina@uni-jena.de)

Well organized data facilitates scientific discoveries, and this is particularly important when it comes to the massive chemical datasets. In recent years, the NoSQL (standing for “not only SQL”) databases gained popularity over the classical SQL databases due to their flexible schemas for data storage and retrieval, enhancing the traditional table structures found in relational databases. MongoDB [1] is a document-based noSQL database particularly suitable for storing the very diverse and sparse data on natural products, and easy data querying and crossing. This database type is rapidly gaining popularity in the cheminformatics community, as more and more chemical interfaces for it are developed to enable similarity and substructure searches. The example of the COCONUT database [2] for natural products will be discussed as an illustration of the topic.

References:

[1] mongoDB: the database for modern applications (https://www.mongodb.com/de).
[2] M. Sorokina, P. Merseburger, K. Rajan, M. A. Yirik, C. Steinbeck. Coconut Online: Collection of Open Natural Products Database. ResearchSquare. DOI: 10.21203/rs.3.rs-75600/v1

OP11: QSAR Studies on Some Anti-Cancer Kinase Inhibitors

Vaishali M. Patil,[1,*] and Neeraj Masand[1]

[1]Computer Aided Drug Design Lab, KIET Group of Institutions, Delhi-NCR, Ghaziabad-201206, India

[*]Corresponding author (Email: vaishuwise@gmail.com)

The field of anti-cancer drug discovery has witnessed role of AXL i.e. a receptor tyrosine kinase as a potential target. It has demonstrated role in promoting tumor growth and metastasis [1]. A series of substituted pyrimidines has been selected to derive the structure-activity relationship towards inhibition of the receptor tyrosine kinase AXL. These small molecule inhibitors were selected followed by calculation of molecular descriptors i.e. hydrophobic, steric and electronic. The identified contributing molecular descriptors can be further explored to design novel analogs as potential anti-cancer agents. The newly designed analogs were screened for binding interactions and pharmacokinetic profile.

Reference:

[1] A. Mollard, S.L. Warner, L.T. Call, et al. *ACS Med Chem Lett*, **2011**, *2,* 907-912.

OP12: Cameroonian Medicinal Plants as Potential Candidates of SARS-CoV-2 Inhibitors: *In Silico* Investigation

Romuald Tematio Fouedjou,[1,*] Samir Chtita,[2] and Loris Alvine Djoumbissie[3]

[1]Department of Chemistry, University of Dschang, Dschang, Cameroon
[2]Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, Casablanca, Morocco
[3]Department of Chemistry, University of Dschang, Dschang, Cameroon
[*]Corresponding author (Email: tematioromy@yahoo.fr)

Background: Since December 2019, the pandemic of coronavirus has emerged in Wuhan, and was spreading throughout the rest of the world [1]. The International Committee on Taxonomy of Viruses has renamed the provisionally named 2019-nCoV as Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) [1].Since prehistoric, different parts of plants are used to treat several diseases among which respiratory infections [2]. For this purpose, ethnobotanical surveys were carried out within healers, and the plants used to prepare recipes against respiratory tract diseases were collected from the Cameroonian flora.
Methods: As part of the promotion of Cameroonian medicinal plants, various bioactive compounds have been isolated for their biological activities. These results show that plants could be an important reservoir of bioactive molecules. In order to identify inhibitors that could facilitate the development of potential anti COVID-19 drug candidates.Some compounds isolated from Cameroonian medicinal plants used to cure respiratory infections were screened, for their affinity against SARS CoV-2.
Results: The structures of 74 plants metabolites were converted into a single database format using *Discovery Studio* software. After the optimization of the top candidates, molecular docking analysis were furthermore conducted of the top candidates with SARS-CoV-2 main protease and spike protein, which represents potential therapeutic targets for the inhibition of SARS-CoV-2 replication.
Conclusion: Five compounds give the best preliminary results with lowest docking scores and were found to have significant affinity with the selected target. Molecular modeling was performed, and the results also illustrate thatH-bonding and hydrophobic interactions are crucial for the stability of docked complexes. Investigation of ADMET parameters helps to filter among the top hits and identify one compound, as a suitable lead candidate.
Keywords: ADMET, *Asteraceae*, Cameroonian medicinal plants, Molecular modeling, SARS-CoV-2.

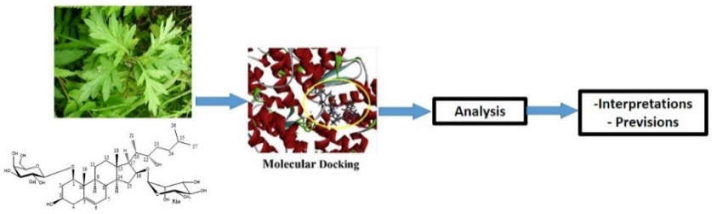


Figure 1: Graphical abstract.

References:
[1] A.E. Gorbalenya, S.C. Baker, R.S. Baric, et al. *Nat Microbiol*, **2020**, *5*, 536–544.
[2] V. Kumar, S. Suri, R. Prasad, et al. *Agric & Food Secur*, **2019**, *8*, 6.

Abstracts of Hands-on Sessions

HS01: Mining the Plant Specialized Metabolome with Mass Spectrometry: Library Matching and Molecular Networking with *GNPS*

Justin J.J. van der Hooft[1*]

[1]Bioinformatics Group, Wageningen University, Wageningen, the Netherlands
[*]Corresponding author (Email: justin.vanderhooft@wur.nl)

Plants produce a myriad of specialized metabolites that support various important physiological and defensive functions. To fully appreciate the plant specialized metabolome, mass spectrometry techniques are currently widely employed to measure hundreds to thousands to metabolites per plant extract. Whilst such wide-screen metabolomics approaches yield an unprecedented amount of metabolic information on plant species, it also comes with a new challenge: most plant metabolites have not been fully characterized yet, making their identification from mass spectra in metabolomics methods very difficult. To overcome this challenge, computational metabolomics workflows have recently been introduced that aid in the processing, analysis, and annotation of plant metabolomics profiles. In particular, the use of mass spectrometry fragmentation (MS/MS) has been advocated as this supports metabolite annotation by providing characteristic mass spectral patterns for particular metabolite classes or scaffolds. The *GNPS* platform was initially built to facilitate the analysis of microbial natural products data and extended to other sources including plants [1,2,3]. In this hands-on tutorial, we will - within the web based *GNPS* environment - study a recently published plant metabolomics dataset that investigates the chemical diversity across 71 Rhamnaceae plants [4]. Various families and genera of two main Rhamnaceae clades were profiled using liquid chromatography coupled to MS/MS. You will explore the dataset and find out how spectral library matching and molecular networking assist in metabolite annotation. Each of these concepts will be shortly introduced before diving into the plant metabolite data. Finally, the *GNPS* platform now also offers tools to study the original plant metabolomics profiles; for example, to check the LC-MS peak shape of metabolites of interest to validate their presence,or to study the occurrence of particular metabolites across the dataset when quantification information is available; for example, across genera or clade, as enabled through the use of consistent curated metadata [5]. You will get the chance to touch upon this for the Rhamnaceae dataset. I will finalize with an outlook on exciting on-going developments [6] and how this will affect plant metabolomics and natural product research in general.

Keywords: mass spectrometry; molecular networking; natural products; substructure discovery; specialized metabolites.

References:
[1] M. Wang, J.J. Carver, V.V. Phelan, et al. *Nat Biotechnol*, **2016**, *34*, 828-837.
[2] A. Aron, E.C. Gentry, K.L McPhail, et al. *Nat Protoc*, **2020**, *15*, 1954–1991.
[3] L.F. Nothias, D. Petras, R. Schmid, et al. *Nat Methods*, **2020**, *17*, 905-908.
[4] K.B. Kang, M. Ernst, J.J.J. van der Hooft, et al. *Plant J*, **2019**, *98*, 1134-1144.
[5] A.K. Jarmusch, M. Wang, C.M. Aceves, et al. *Nat Methods*, **2020**, *17*, 901-904.
[6] F. Huber, L. Ridder, S. Verhoeven, et al. bioRxiv, **2020**, *DOI*:10.1101/2020.08.11.245928v2.

HS02: Virtual Identification of Bioactive Natural Products from African Flora

Daniel M Shadrack,[1] Thommas Musyoka,[2] and Fidele Ntie-Kang[3,*]

[1]Department of Chemistry, St John's University of Tanzania, P.O.Box 47 Dodoma, Tanzania
[2]Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda, 6140, South Africa
[3]Department of Chemistry, University of Buea, P. O. Box 63 Buea, Cameroon

[*]Corresponding author (Email: fidele.ntie-kang@ubuea.cm)

The South African Natural Compounds Database (SANCDB - https://sancdb.rubi.ru.ac.za/) [1] is a collection of 1,012 compounds derived from South African natural sources. Since its inception in 2015, the database has been used for various machine learning and *in silico* virtual drug screening studies with a recent study identifying several potential hits against severe acute respiratory syndrome coronavirus 2 (SARS-COV-2). As part of a recent update, a unique feature incorporating the compound dataset analogs from two leading commercial databases (Molport and Mcule) was included. The feature will not only allow users to explore a larger chemical space during screening but also allow them to seamlessly purchase compounds for their biological studies. Participants will be introduced to the database with emphasis on how they can obtain compounds for both their virtual screening and biological studies. The second part of the session (approximately 20 minutes) will focus on natural products databases originating from the regions of Northern [2] and East Africa [3] (http://african-compounds.org/anpdb/). The participants will explore the web tools that permit to search the databases, including similarity and sub-structure searching for privileged scaffolds. After you have been introduced to the natural product databases from African sources, their contents, compound classes and potential for lead compound discovery, the last session (about 50 minutes) will introduce state-of-the-art computational techniques used in lead compound identification from electronic databases, e.g. molecular docking and pharmacophore-based searching. In this section you will be introduced to the approaches used to perform *in silico* screening of libraries containing natural products against a main protease for the COVID-19. You will also be briefly introduced to other sophisticated tools like molecular dynamics and metadynamics, just on the fly. You will, therefore, learn how to perform virtual screening from large libraries (focusing on natural products libraries from African sources). A similar approach could be used in your own project of interest, involving another drug target.

Keywords: lead identification; natural products; molecular docking; virtual screening; drug discovery.

References:
[1] R. Hatherley, D.K. Brown, T.M. Musyoka, et al. *J Cheminform*, **2015**, *7*, 29.
[2] F. Ntie-Kang, K.K. Telukunta, K. Döring, et al. *J Nat Prod*, **2017**, *80*, 2067–2076.
[3] C.V. Simoben, A. Qaseem, A.F.A. Moumbock, et al. *Mol Inf*, **2020**, *39*, 2000163.

HS03: *In Silico* Prediction and Identification of Metabolites with BioTransformer: Enabling Secondary Metabolite Discovery

Yannick Djoumbou Feunang,[1*] Jesse Balcer,[1] and Dirk Tomandl[1]

[1]Corteva Agriscience, Indianapolis, Indiana, US

*Corresponding author: (Email: yannick.djoumboufeunang@corteva.com)

In silico metabolism prediction tools provide a unique perspective to studying the chemical exposome, and how its changes affect the environment. Classical applications of such tools include, but are not limited to metabolite discovery, environmental fate prediction, ADMET profiling, and molecular design. In spite of their reported successes and enormous potential, the widespread adoption of metabolism prediction tools is still hampered by several factors such as the moderate prediction accuracy, the limited domains of applicability, as well the accessibility and shareability, among others. In the first part of this presentation, we will describe BioTransformer, an open source software tool, and freely accessible server for the prediction of human CYP450-catalyzed metabolism, human gut microbial degradation, human phase-II metabolism, human promiscuous metabolism, and environmental microbial degradation. Additionally, BioTransformer assists in metabolite identification, and metabolic pathway prediction. In the second part, we will present an assessment of BioTransformer's performance in the prediction of metabolism for diverse sets of molecules, including but not limited to pharmaceuticals, pesticides, and phytochemicals. Overall, BioTransformer was shown to achieve moderately high precision (0>.46), and higher recall (>0.84) when predicting human metabolism of drugs, lipids, and phytochemicals, compared to two commercially available tools. On the other hand, the overall precision (~0.3) and recall (~0.6) achieved for the metabolism prediction of agrochemical suggest that improvements are needed to cover more Ag-relevant chemical as well as biological species. In the third part of this presentation, we will illustrate a few examples of its application as demonstrated by various published scientific studies. Finally, we will share future perspectives for this open source project, and describe how it could significantly benefit the exposure science and regulatory communities.

Keywords: expert system; machine learning; metabolism prediction; metabolite identification; metabolic pathway; metabolomics; precision; recall.

References:

[1] Y. Djoumbou-Feunang, J. Fiamoncini, A. Gil-de-la-Fuente, et al. *J Cheminform*, **2019***, 11,* 2.

HS04: New E-Resource for Drug Discovery (NERDD)

Conrad Stork,[1*] and Neann Mathai[2]

[1]Center for Bioinformatics, Universität Hamburg, 20146 Hamburg, Germany
[2]Department of Chemistry and Computational Biology Unit (CBU), University of Bergen, N-5020 Bergen, Norway
[*]Corresponding author (Email: stork@zbh.uni-hamburg.de)

NERDD (New E-Resource for Drug Discovery) [1] is an easy-to-use web service that gives researchers access to state-of-the-art computational tools that can provide guidance for drug discovery projects: (i) FAME3 [2], a machine learning model to predict the sites of metabolism (SoM) of compounds for phase I and phase II metabolism, (ii) GLORY [3] and GLORYx [4], which utilize predicted SoMs and a set of manually curated reaction rules to predict the likely metabolites of xenobiotics, (iii) Hit Dexter 2.0 [5], which predicts the likelihood of a compound being a frequent hitter in biochemical assays and flags compounds for which extra caution should be exercised when interpreting experimental measurements, (iv) NP-Scout [6], which is a tool to elucidate the natural product likeness of compounds, and (v) Skin Doctor CP [7], a model for the prediction of the skin sensitization potential of small molecules. NERDD is actively being maintained and expanded as more tools for its catalog are being developed. All of the tools are based on state-of-the-art machine learning models with high accuracies on holdout data. In this session, we will give detailed insights into NERDD, the science behind the individual tools and how they work, as well as the scope and limitations of the individual models. We will walk participants through the interpretation of the results, the different modes under which the models can be run, and the applicability domains of the tools. The session will be the perfect venue to discuss the tools with the developers of NERDD and to give valuable feedback and suggestions that you would like to see incorporated.

Keywords: natural products; metabolism; cheminformatics; web server; tutorial

References:

[1] C. Stork, G. Embruch, M. Šícho, et al. *Bioinformatics*, **2020**, 36, 1291.
[2] M. Šícho, C. Stork, A. Mazzolari, et al. *J Chem Inf Model*, **2019**, 59, 3400.
[3] C. de Bruyn Kops, C. Stork, M. Šícho, et al. *Front Chem*, **2019**, 7, 402.
[4] C. de Bruyn Kops, M. Šícho, A. Mazzolari, et al. *Chem Res Toxicol*, **2020**, doi: 10.1021/acs.chemrestox.0c00224.
[5] C. Stork, Y. Chen, M. Šícho, et al. *J Chem Inf Model*, **2019**, 59, 1030.
[6] Y. Chen, C. Stork, S. Hirte, et al. *Biomolecules*, **2019**, 9, 43.
[7] A. Wilm, U. Norinder, M.I. Agea, et al. *Chem Res Toxicol*, **2020**, doi: 10.1021/acs.chemrestox.0c00253.

HS05: Cheminformatics for Natural Products with the CDK

Maria Sorokina,[1,*] Aziz M. Yirik,[1] Jonas Schaub,[1] and Christoph Steinbeck[1]

[1]Institute for Inorganic and Analytical Chemistry, Friedrich Schiller University, Jena, Germany
[*]Corresponding author (Email: maria.sorokina@uni-jena.de)

The Chemistry Development Kit (CDK) [1] is one of the main programming toolkits for processing and analyzing chemical information. It is available as modular Java libraries, easy to use, open and free, and available to download at https://cdk.github.io/. It can also be easily integrated with Maven and Gradle. During this hands-on, we are going to present the key CDK concepts, such as molecule representation and manipulation, diverse fingerprints, and molecular descriptors. Examples of the usage of CDK for natural products discovery and analysis, like the NP-likeness scorer [2] and the Sugar Removal Utility [3], will also be presented. All code examples presented during this workshop will be provided. The only pre-requisite for attending this workshop is some coding experience.

[1] E. L. Willighagen, J. W. Mayfield, J. Alvarsson, et al. *J Cheminform*, **2017**, *9*, 33.
[2] M. Sorokina, C. Steinbeck. *J Cheminform*, **2019**, *11*, 55.
[3] J. Schaub, A. Zielesny, C. Steinbeck, M. Sorokina. *J Cheminform*, **2020**,*12*, 67.