Health & Medicine | J Lill, C Rose, R Mathews & M Schirle

Proteomic advances for the future of drug development

Proteomics has emerged as an essential tool in biotechnology's efforts to pioneer drug discovery. Techniques including comprehensive quantitative proteome analysis, sensitive single-cell proteomics, chemoproteomics, and artificial intelligence have helped the scientific community dig deep into the proteome and decipher intricate yet important pathways that could be of immense help in pathway and biomarker discovery and drug design. A US team at biotech company Genentech, consisting of Dr Jennie Lill, Dr Christopher Rose, and Dr Rodney Mathews, collaborating with Dr Markus Schirle of Novartis, is instrumental in driving the application of proteomics in the biotechnological and pharmaceutical sector. They examine the current landscape in proteomic research and how innovations are shaping the future of healthcare and medicine.

VectorMine/Shutterstock.com

Proteomics refers to the study of peptides and proteins and how they regulate cellular processes such as cell-signalling events. It has evolved to be one of biotechnology's essential tools to drive drug discovery over the years. Recently, novel proteomic techniques have been developed that allow researchers to dig even deeper into the proteome (protein composition) of a cell or tissue, increasing the granularity of research and sensitivity of the analysis. This has translated to discovering novel therapeutic targets, deciphering cellular processes and signalling pathways with more detail, and a better understanding of potential toxicity of drugs in cells.

The authors of a recent review paper are experts in proteomic research and bring decades of drug discovery experience with them. Dr Jennie R Lill, Dr Christopher M Rose and Dr W Rodney Mathews of Genentech (South San Francisco, California) collaborated with Dr Markus Schirle of Novartis

(Cambridge, Massachusetts) to review the current landscape in proteomic research and how innovations are contributing to the future of healthcare and medicine.

ADVANCES IN PROTEOMIC RESEARCH

Proteomic research has been most impactful primarily on pre-clinical studies wherein proteomic techniques have been used to identify drug candidates and specify their role in therapeutic applications. It has also found applications in biomarker discovery – another crucial aspect of therapeutic understanding and monitoring of diseases.

However, proteomics has historically suffered from a lack of sensitivity compared to its transcriptional profiling counterpart technologies. Thanks to the recent advances in sample collection, improved analytical processes like separation, improved ionisation, and better mass spectrometric instrumentation, followed by data collection and analysis, it is now possible to quantify more than 1,000 proteins from a single cell.

Enhancing the purity of samples introduced into the mass spectrometer at the beginning of an experiment can be one feasible solution to improve sensitivity in proteomic research. Reducing sample preparation time and minimal sample interaction with surrounding surfaces can ensure the purity of samples to a considerable extent and contribute to the sensitivity of the assay.

Tools like 'NanoPOTS' (nanodroplet processing in one pot for trace samples)
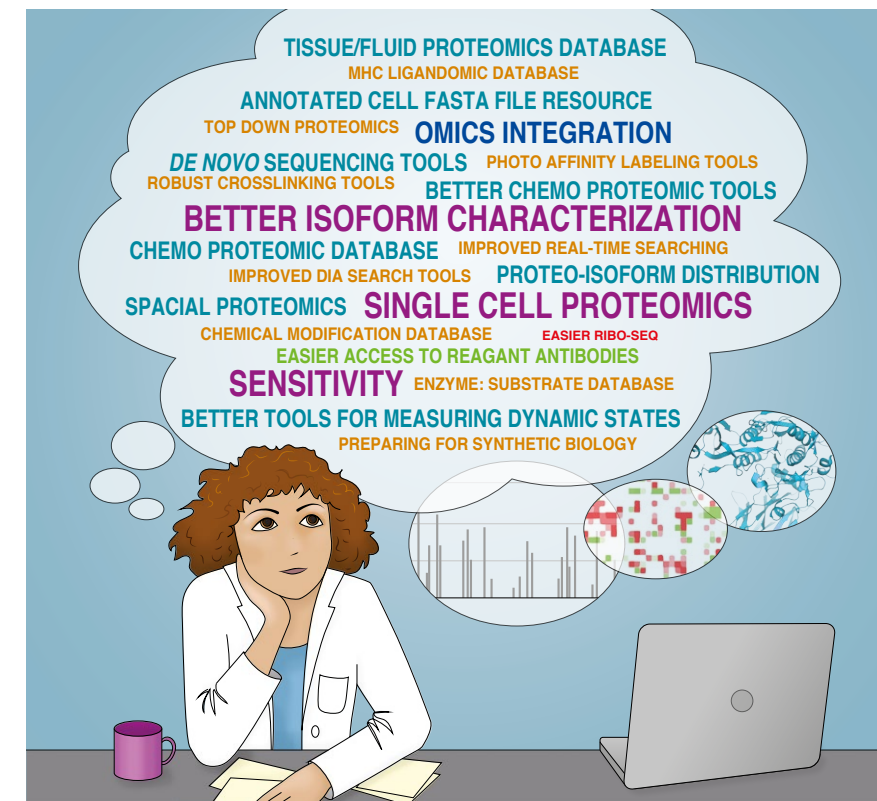
Novel proteomic techniques allow researchers to dig deeper into the proteome of a cell or tissue.

are now available for more efficient inputs of low-level materials for mass spectrometric analyses. The device consists of glass chips made using UV-based lithographic techniques with hydrophilic pedestals. The chips also have a glass spacer sealed to a glass slide coated with membranes. These pedestals are surrounded by hydrophobic surfaces, giving the sample preparation space a nanowell-like appearance.

In these nano-chambers, the different phases of proteomic sample preparation, like reduction, alkylation, and proteolytic digestion, can be performed at a miniaturised scale in a humidified environment. The membranous coating of the nano-chambers minimises sample exposure time and thereby reduces evaporation. The glass substrates allow sample manipulation in smaller volumes (as little as < 200 nanolitres (nL)) and can also minimise chances of sample loss by reducing sample interaction with adherent surfaces or adsorption, thanks to the substrates' hydrophobic chemical nature. Thus, the overall chemical nature

of the NanoPOTS prevents undue sample loss and ensures the purity of samples for analysis.

Ion mobility spectroscopy (IMS) is another novel technique on the rise. It allows for the separation of ions in the gas phase based on their mobility in a carrier buffer gas. Performing IMS prior to mass spectrometric analysis separates the noise from the sample on the basis of charge, enabling more purified samples to be fed into the spectrometer. This increases the likelihood of a more specific analysis.

Recent advances have made it possible to quantify more than 1,000 proteins from a single cell.

Multiplexing technologies have not only increased the number of proteomes to be analysed in one go, but they have also dramatically improved the ability to assay various genotypes, treatments, or time points in a single discovery proteomics experiment.

Single Cell ProtEomics by Mass Spectrometry (SCoPE-MS) has garnered

praise among the research community as a novel method to use labelled proteomes from a single cell for more sensitive analysis. The SCoPE-MS platform uses isobaric chemical labels to differentiate a carrier proteome with approximately 25–500x the amount of a single cell proteome from the experimental samples under investigation. The carrier proteome is added to the experimental single-cell proteomes to yield a single analytical sample, and the carrier proteome facilitates identification and quantitation of peptides and proteins in the experimental samples.

Particularly in the context of chemical labelling techniques, improved MS data acquisition techniques like intelligent data acquisition (IDA) conducted with the help of real-time analysis of MS data have enabled improvements in data collection. The generated datasets show an increased depth and sensitivity of proteomic analysis, thus increasing the level of detail with which a cellular proteome can be studied.

A new entrant in the field of targeted proteomics is triggered by offset, multiplexed, accurate mass, high resolution, and absolute quantification (TOMAHAQ). This technique uses a combination of isobaric labels and synthetic peptides to multiplex protein probes within a sensitive targeted assay. As the TOMAHAQ data acquisition progresses, the synthetic peptide, also known as the trigger peptide, is identified first. This further initiates an offset analysis on the endogenous target peptide. Once this endogenous target peptide is sequenced, the corresponding fragment ions are isolated for final quantitative analysis.

Another targeted proteomic approach employs nanopore technology for selective detection of amino acids when a protein is passed through the pore. A number of different techniques have been implemented to feed the protein through the pore, such as attaching a DNA tag, using an unfoldase (enzymes responsible for the catalytic unfolding of proteins) or adherent negative ionic detergents. However, it remains challenging to

precisely distinguish between the 20 proteinogenic amino acids, owing to the small size of amino acids in comparison to a monophosphate nucleotide.

USING A MULTI-OMICS APPROACH FOR BETTER ANALYSIS AND IDENTIFICATION

Scientists have also attempted to gain a holistic sense of an organism, cell, or biological pathway by analysing its proteomic, genomic, and lipidomic data components together. Simply put, these datasets form a comprehensive molecular footprint of a given biological pathway. However, this requires in-depth knowledge and understanding of the complexity and nuances of each -omic approach. And while multi-omic integration is still evolving, intriguing examples are emerging which will pave the way for more adaptation among the scientific community.

Amidst the ongoing COVID-19 pandemic, combining multi-omics results generated analyte clusters enriched in severe COVID-19 cases, including the protein gelsolin (GSN) and the metabolite citrate. This association is interesting because GNS is a Ca^{2+}-activated actin-severing protein and citrate is a calcium chelator, pointing towards a potential role of cytoskeleton modifications during virus–host interactions in COVID-19. This example shows how an understanding of the biological roles of molecules is vital to revealing the importance of enriched clusters. Advances in proteomic research thus far have led to the application of machine learning (ML)-based approaches to develop algorithms that can predict the severity of COVID-19. Multi-omics-based analysis has also helped researchers identify two key metabolites associated with COVID-19 severity – kynurenine and quinolinic acid, both of which are crucial for the immune system to recruit inflammatory cells.

MACHINE LEARNING AND LARGE-SCALE DATA AVAILABILITY

There has been a machine-learning boom in proteomics is as well. ML algorithms such as linear discriminant analysis (LDA) or support vector machines (SVM) have been used in the past to distinguish between true and false peptide identifications. Recently, neural networks have also emerged as useful proteomic tools.

One of the first MS spectrum prediction algorithms, MS2PIP, demonstrated the possibility of predicting spectral patterns. Recently, two deep learning algorithms, Prosit and DeepMass:Prism, have been increasingly used to accurately predict MS spectra. So far, proper functioning of the algorithms typically requires training data with target peptide sequence, possible modifications, and fragmentation modes. The result is a match score between the spectrum predicted by the algorithm to that of an experimental (protein) spectrum. The match score can then be used to help discriminate true from false identifications. This form of cross-checking can increase identification accuracy by as much as 50% for searches involving larger databases (eg, MHC-associated peptide searches).

While ML algorithms are slowly gaining popularity, the availability of large-scale genetic and transcriptomic data has helped researchers refine their understanding of common cancer mutations. This is particularly important as new therapeutic modalities, such as cellular therapies, aim to target up-regulated proteins in the tumour tissues (tumour associated antigens, TAA) or the mutated

Multi-omics-based analysis has helped researchers identify two key metabolites associated with COVID-19 severity.

cancer proteins, differentiating them from normal cells.

CHEMOPROTEOMICS AND DRUG DEVELOPMENT

Chemoproteomics refers to techniques that aim to identify and characterise the interaction of proteins with chemical matter. These workflows are specifically important in identifying drug–target interactions in cells or cell-derived samples such as cell lysates or enriched subcellular fractions to aid in the identification of the efficacy target of a drug or target occupancy studies.

Chemoproteomics workflows typically share four general steps: 1) selection of disease-relevant input material and adequate preparation; 2) treating samples with a probe to allow for target binding; 3) separating the compound-interacting proteins from the rest of the proteome by a variety of means, including affinity enrichment or detection of changes in protein stability upon binding to a compound; and 4) detecting the interacted and quantified proteins against an untreated control, typically using quantitative mass spectrometry.

These techniques can also be used to identify mechanisms of action and biomarkers specific to cells to enhance the drug-development process.

REVOLUTIONISING DRUG DEVELOPMENT

The authors of this study believe that proteomics can contribute significantly to drug discovery and development. Breakthrough innovations like enhanced sensitivity, multi-omic data integration, chemoproteomic technologies, and advances in biomarker discovery, including improved analytics, have the potential to revolutionise the drug-discovery and development process.

Behind the Research

Jennie Lill

Christopher M Rose

W Rodney 'Rod' Mathews

Markus Schirle

W: www.gene.com/scientists/our-scientists/jennie-lill   **W:** www.gene.com/scientists/our-scientists/chris-rose
W: www.gene.com/scientists/our-scientists/rod-mathews   **E:** markus.schirle@novartis.com

Detail

Bio

Jennie Lill is Executive Director of the Proteomics and Next Generation Sequencing department at Genentech in South San Francisco, California. She has worked on numerous drug discovery efforts over the past 20 years and led the Engineered T cell programme for Research at Genentech, as well as being heavily involved in other cancer immunology modalities. Her research focuses on using proteomics to study cell death mechanisms and associated proteolysis events and the MHC Ligandome in the context of disease.

Christopher M Rose is Director of Discovery Proteomics within Genentech's Microchemistry, Proteomics, and Lipidomics department. He currently leads efforts in developing and applying cutting-edge quantitative proteomic technologies aimed at exploring biological pathways related to potential therapeutic targets and/or molecules. His lab is currently focused on developing proteomics technologies to enable single-cell proteomic analyses and measuring MHC associated peptidomes to understand the processing and presentation of important cancer neoantigens as it relates to therapeutic intervention.

W Rodney 'Rod' Mathews is Director of the Technology Group in the OMNI Biomarker Development department at Genentech. The OMNI Biomarker Development department is responsible for the development and execution of biomarkers for all non-oncology programmes at Genentech. Dr Mathews' group is focused on the use of specialised technology, including mass spectrometry, mass cytometry, and flow cytometry, to discover, develop and implement clinical biomarkers. Dr Mathews has over 35 years of drug discovery and development experience, with an emphasis on leveraging the analytical power of mass spectrometry for target identification, lead optimisation, understanding mechanism of action, clinical biomarkers and development of novel therapeutics.

Markus Schirle is Research Investigator at Novartis Institutes for BioMedical Research. He established an independent platform for chemical and affinity proteomics for target identification (ID) and mechanism of action (MoA) studies in what is now the Chemical Biology and Therapeutics department. Currently, Dr Schirle is leading a group dedicated to the identification and characterisation of compound–protein interactions using proteomics, chemical biology, in vitro biophysics and biochemistry, as well as structural biology. His team is responsible for affinity-based approaches to target ID/MoA as well as applications of covalent chemoproteomics and complementary approaches to the identification of ligandable pockets and chemical starting points for Novartis projects globally, as well as the exploration of new modalities.

Research Objectives

The researchers review areas of innovation in the field of proteomics and their effects on biotechnological and pharmacological research.

References

Lill, JR, Mathews, WR, Rose, CM, Schirle, M, (2021) Proteomics in the pharmaceutical and biotechnology industry: a look to the next decade. *Expert Review of Proteomics*, 18(7), 503–526. doi.org/10.1080/147894 50.2021.1962300

Personal Response

Which is the most promising proteomic innovation that you have come across?

We are very excited about the single molecule sequencing and the possibilities that it will open up to be able to fully characterise a single copy of a protein in its cellular context.

Do you think that combining ML algorithms to biotechnology research can replace human efforts?

I believe that these will be complementary to one another. ML is still in its infancy, but has already shown great power in developing in silico based prediction algorithms, data search pipelines and other interesting tools. However, humans have a unique way of thinking and I believe that human and ML efforts will work best when hand in hand.