NAUTILUS™
BIOTECHNOLOGY

WHITE PAPER

The proteomics revolution - Accelerating basic research with proteomics

Biology researchers have learned much about the intricate inner workings of cells using classic molecular biology, biochemistry, and genetics techniques. Generally, these techniques minimize biological complexity by looking at only a small sub-component of a cell in a highly controlled experimental setting. For instance, researchers might break a small part of the cell (say a single protein) and see what effect that has on the behavior of the cell or organism. Although they have brought us far, such techniques are by their nature reductionist and make it difficult to assess the many interactions occuring in living things at any moment in a holistic way.

To truly understand how living things function in a mechanistic way that can be leveraged for applied purposes, researchers need to bridge the gap between small cellular components and broad organismal behaviors. They need to map out the intricate networks of interactions that connect cellular parts to organismal functions. Proteomics will enable them to do so.

Use this QR code to find all the white papers in this series.

Table of Contents

Introduction

Proteins carry out the vast majority of cellular processes and a cell's proteome, the dynamic composition of its full set of proteins and their associated distribution throughout the cell, determines cellular identity, form, and function. Proteomics technologies aim to reveal the abundances and locations of all proteins in cell samples right now. The dynamic patterns of these ever-changing "proteomes" comprise thousands to millions of fluctuations in protein levels, protein modifications, and protein distribution throughout the cell, and can help reveal the networks of protein interactions that are perturbed when cells experience various stimuli.

Below we discuss 3 types of research that can be accelerated by advanced proteomics technologies. This is far from an inclusive list and focuses on non-clinical research with a nod to clinical applications that can result from these research areas. As we get affordable, high-throughput proteomics technologies into the hands of researchers around the world, they'll develop myriad applications that will reveal much more about the inner-workings of living things than we ever thought possible.

Mapping the proteomic landscape of the cell surface

Cells are not homogenous bags of proteins. Instead, proteins are distributed to discrete locations throughout the cell where they act in coordination with other cellular components to carry out specialized processes. Proteins found at the cell surface are often inserted into the "plasma membrane," the fatty barrier separating the inside of the cell from its surroundings. These "membrane proteins" are particularly important because they make it possible to interact with and relay information about a cell's surroundings. While membrane proteins are traditionally difficult to study, new proteomics technologies will make it easier for scientists to learn more about these incredible molecular machines.

Membrane proteins have varied and astounding functions. For example, some proteins on the surface of immune cells detect parts of infectious bacteria. Once detected, these proteins trigger a cascade of events that can result in the consumption and later destruction of the bacteria. In another interesting example, proteins on the surface of nerve cells control the flow of ions in and out of the membrane. They enable our nervous systems to transmit highly controlled electrical signals. Still other proteins embedded in the surface of retinal cells detect specific wavelengths of light, a function that ultimately gives us color vision.

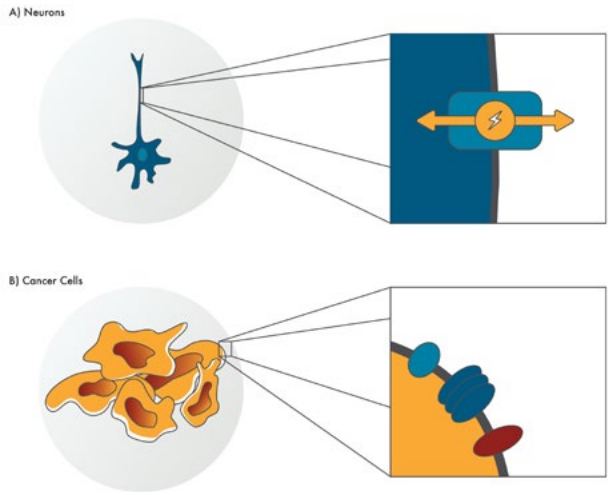


Figure 1. Cell surface proteins

Cell surface proteins or "membrane proteins" have a variety of fascinating functions such as (A) enabling neurons to send electrical signals. While they can be hard to isolate and study, if researchers can (B) find membrane proteins specific to cancer cells, these may make good drug targets.

Unsurprisingly, membrane proteins are a fascinating subject for scientists wishing to understand how our bodies work, but they are also incredibly important in the clinic. Membrane proteins make useful drug targets because they're easier to reach than proteins buried inside the cell. They're also accessible to large, programmable biological therapeutics like antibodies and engineered cells.

Unfortunately, membrane proteins can be very difficult to study. They are not usually soluble in water and can be hard to purify from the fatty barriers where they reside. In addition, they sometimes stick to the materials scientists use to store and transport them. As a result, scientists often have to use strong chemicals to break apart the membranes housing these proteins and keep them from sticking to their containers. These chemicals are generally incompatible with standard protein analysis techniques.

These characteristics make membrane proteins difficult to study not only on an individual level but also at the level of the proteome. With standard proteomics techniques, scientists usually use enzymes to cut proteins into smaller, more manageable chunks. Membrane proteins contain few cut sites and the strong chemicals used to pull them out of the membrane disrupt the enzymes' cutting activity. Scientists have come up with a variety of clever ways to get around these issues, but they are usually complicated and time-consuming.

An additional problem for standard proteomics techniques is membrane protein abundance. Membrane proteins are generally low abundance and the signals they generate on standard proteomics platforms can be drown out by other proteins. This is because standard proteomics platforms don't look at individual proteins in isolation. Instead, they identify small protein chunks from many different proteins. Various kinds of analysis software then piece together the identities of the full proteins from the small chunks. This is much like putting together a very complicated puzzle. High abundance proteins generate many small chunks that can obscure the identities of the small abundance proteins.

Scientists can isolate membrane proteins and analyze them separately, but this can bias which proteins they ultimately see and complicate the experimental process.

Nautilus' technology gets around these issues in a number of ways. First off, scientists don't need to break apart their proteins before analyzing them on the Nautilus' platform. The platform is also compatible with the strong chemicals used to pull proteins out of membranes. Special isolation techniques are not required. Finally, full-length proteins are analyzed individually on the platform. This gives it higher sensitivity and makes it much less likely that signals from high abundance proteins will drown out membrane proteins.

Thus, the Nautilus platform opens the proteomic landscape of the cell surface to scientists. They can use just about any chemicals they wish before analyzing membrane proteins and they'll also get a full view of most of these proteins, regardless of their abundance.

With this landscape in view, scientists will be able to develop a clear understanding of how cells interact with their surroundings. For instance, they'll be able to determine what proteins stud different types of cells and, by associating particular membrane proteins with known cell-cell interactions, generate new hypotheses about how these interactions function at the molecular level. They'll also be able to probe how surface proteins change in various states of health and disease. For example, by studying the proteins found on the surface of cancer cells compared to healthy cells, they may find cancer-specific proteins. Thus, they'll be able to target these proteins with cancer-specific drugs - something that is only possible for a few types of cancer now.

Overall, advances in proteomics technology will enable researchers to answer a seemingly basic question: what proteins are found on the cell surface? The answers will advance our understanding of how the body works and improve health for many.

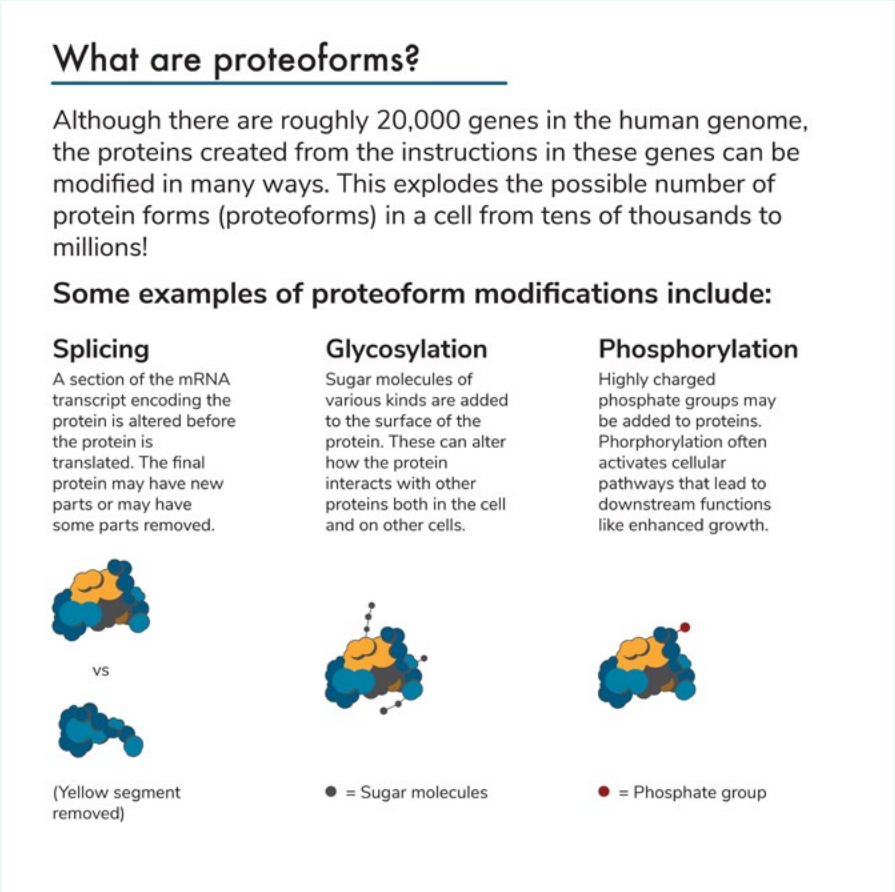
Enhancing our understanding of biological function and dysfunction through the identification of proteoforms

You may have heard that there are roughly 20,000 genes in the human genome. One might assume that, since genes encode proteins, there can be no more than 20,000 different proteins in any given cell. However, this is far from the case. In the process of decoding the information in a gene and using it to produce a protein, that information can be contorted in a variety of ways that leave out some parts of the original gene, start the decoding processes at different places in the gene, or even edit small portions of the information. In addition, once this information is used to create a protein, the protein itself can be modified with a variety of attachments such as chemicals and sugars. All this rearranging and modification explodes the number of possible "proteoforms" that could theoretically be produced in a cell from 20,000 to millions (Aebersold et al 2018).

The modifications found on proteoforms can change the underlying protein's function or even lead to its destruction. Indeed, the precise mix of proteoforms in a cell can have great impacts on cell processes, organ function, and total body health. Thus, knowing more about the proteoforms present in any given healthy or diseased cell may give scientists great insights into how these cells operate. Unfortunately, most proteomics technologies used today can only see a small fraction of proteoforms. Reasons why include:

LOW SENSITIVITY LIMITS THE PROTEOFORMS THAT CAN BE OBSERVED

There are many different ways proteins can be modified to create a wide array of proteoforms. However, for any given cell it's possible that only a few proteoforms will be present in high abundance. There may be low abundance proteoforms that have functional impacts on the cell, but they will be hard to see as the more abundant proteoforms drown out their signals on standard proteomics platforms. In addition, as most proteomics platforms look at groups of cells as opposed to individual cells, if one cell is dominated by a proteoform that is rare in other cells, that proteoform will be difficult to see.



LIMITATIONS DUE TO PROTEIN FRAGMENTATION

On standard proteomics platforms, full-length proteins are fragmented into small chunks. These small chunks are identified, and analysis software pieces together full proteins from the identified chunks. However, not all the chunks that make up the full protein will actually be identified. Some chunks will be left out of the analysis and those chunks could have modifications that are invisible to standard proteomics technologies.

LIMITED MEANS OF DETECTION

Standard proteomics platforms generally detect the mass of protein fragments in order to identify them. They look to see how many fragments of various weights are present and, by plotting abundance vs weight, they generate so-called "spectra" which are essentially signatures of different fragments. However, some fragments can generate very similar spectra and, without reference samples to distinguish between them, it can be difficult to identify fragments with never-before-seen modifications.

THE BENEFIT OF DETECTING NEW PROTEOFORMS

There are many unknowns when it comes to the world of proteoforms. Scientists have made theoretical predictions about the number of proteoforms that could possibly exist, but it is not at all clear what fraction of these proteoforms are actually made, how they might be distributed across cells, and what functional consequences they have. Nonetheless, we do know that some protein modifications are highly consequential. For example, we know that an overabundance of certain types of modifications to proteins that scaffold DNA are associated with poor outcomes in cancer (Nebbioso et al 2018). If scientists can identify more proteoforms in particular cells, they can more confidently associate them with various states of health and disease. Some may even become biomarkers that indicate when a person has a particular disease and others may be targeted by novel drugs.

ENHANCING PROTEOFORM DETECTION WITH THE NAUTILUS PLATFORM

At Nautilus, we're developing a proteomics platform that makes it easier to identify a much wider variety of proteoforms. Some of the ways the platform enhances proteoform identification include:

- *The platform analyzes individual proteins in isolation.* This gives the platform higher sensitivity than standard proteomics platforms and thus more potential to detect relatively rare proteoforms.
- *The platform analyzes full-length proteins.* Thus, it is theoretically possible to identify modifications across the entire length of a protein.
- *The platform looks at many characteristics of proteins other than their mass.* The platform uses "probes" that detect particular sequences of protein building blocks in full-length proteins. These probes can be designed to detect building blocks that have been modified in a variety of ways and thereby detect various proteoforms.

By getting a more in-depth view of the proteoforms that exist across cells, we'll take the first step toward understanding how these proteoforms impact health and disease. Once scientists observe differences in proteoform abundance across cells, they can investigate whether those differences are functionally significant, but the first step is seeing what proteoforms actually exist. Nautilus' platform can bring us a long way toward accomplishing this key step.

Proteomics and developmental biology - how do cells and tissues get their identities?

Developmental biology is the study of how organisms grow from single cells. Research in this field has inspired advances in evolution, stem cell biology, cloning, and much more, but there are many open questions. In particular, scientists don't have a full understanding of what defines the different types of cells that can be found in an organism and they are beginning to see that many cellular identities are fluid. Some cells can take on new identities and even replace others after tissue damage (Spatz et al 2021). So what really makes a skin cell a skin cell or an immune cell an immune cell? It's the proteins, the molecular machines that give these cells their functions, that matter.

Scientists can track cell lineages through development, know many of the key proteins that drive cells to become one type or another, and have analyzed many cell types through Herculean efforts like the Human Protein Atlas. Nevertheless, the older techniques used in many of these projects don't always provide a full view of the many proteins and protein isoforms produced in the human body. Recent efforts using advanced proteomics technologies dig deeper and find new proteins associated with particular tissues (e.g. skin). Using novel, high-throughput proteomics platforms like those being developed at Nautilus, researchers can probe the millions of proteins and protein isoforms produced in the human body and get a much fuller picture of cellular identity across all tissues.

With such comprehensive data, it may become easier to understand how to do things like coax one type of cell to become another. This would be useful not only to prove that certain proteins are sufficient to define cellular identities but also because a variety of diseases manifest when people lack particular cells. For example, the liver detoxifies many compounds that enter our bodies but can itself be damaged by these compounds. This leads to impaired liver function and can ultimately result in death. Researchers could perhaps develop new means of treating liver disease by getting other cells to replace diseased or dead liver cells. Similarly, diabetes can arise when patients lack the pancreatic beta cells that produce insulin. Researchers are making efforts to treat diabetes by getting more pancreas cells to transform into insulin-producing beta cells (Zhong and Jiang 2019). Proteomics technologies could accelerate these and similar efforts by precisely highlighting what proteins need to be altered in order to make replacement cells fully functional.

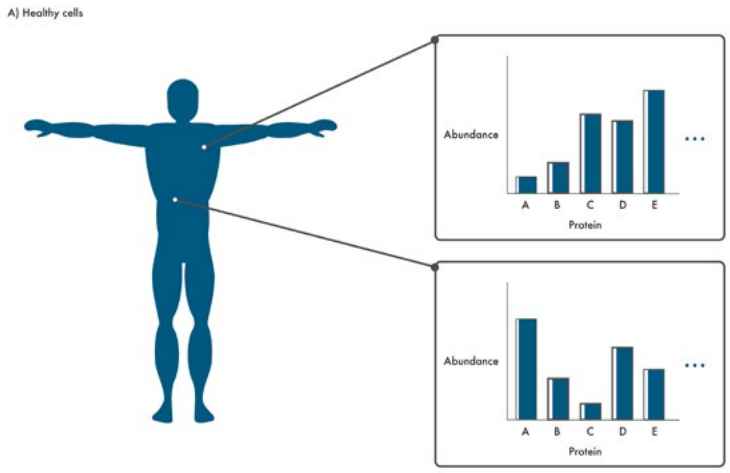


Figure 2. What gives cells their unique identities?

Cells in different parts of the body play vastly different roles in maintaining human health. As the molecular machines that give cells their functions, proteins largely dictate cellular identity. With widely accessible proteomics technologies, researchers will be able to get a better understanding of how protein composition determines cellular function and identity.

Looking forward to future discoveries enabled by proteomics

Proteomics gives us the ability to quickly grasp the full array of changes in protein abundance, form, and distribution that result from whatever biological interaction or stimulus one would like to study. This burgeoning field makes it possible to gather a ton of information on how cells react to one another and with their surroundings. By using proteomics technologies to dissect any biological process they're interested in, researchers will gain a much fuller understanding of how these processes work and they should be able to alter such processes for a variety of goals. This will make it possible for scientists to better understand the complexities of biology and leverage that understanding to create new metrics of human health, therapeutics, and biotechnologies.

References

Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA, Ge Y, Gunawardena J, Hendrickson RC, Hergenrother PJ, Huber CG, Ivanov AR, Jensen ON, Jewett MC, Kelleher NL, Kiessling LL, Krogan NJ, Larsen MR, Loo JA, Ogorzalek Loo RR, Lundberg E, MacCoss MJ, Mallick P, Mootha VK, Mrksich M, Muir TW, Patrie SM, Pesavento JJ, Pitteri SJ, Rodriguez H, Saghatelian A, Sandoval W, Schlüter H, Sechi S, Slavoff SA, Smith LM, Snyder MP, Thomas PM, Uhlén M, Van Eyk JE, Vidal M, Walt DR, White FM, Williams ER, Wohlschlager T, Wysocki VH, Yates NA, Young NL, Zhang B. How many human proteoforms are there? Nat Chem Biol. 2018 Feb 14;14(3):206-214. doi: 10.1038/nchembio.2576. PMID: 29443976; PMCID: PMC5837046.

Nebbioso A, Tambaro FP, Dell'Aversana C, Altucci L. Cancer epigenetics: Moving forward. PLoS Genet. 2018 Jun 7;14(6):e1007362. doi: 10.1371/journal.pgen.1007362. PMID: 29879107; PMCID: PMC5991666.

Spatz LB, Jin RU, Mills JC. Cellular plasticity at the nexus of development and disease. Development. 2021 Feb 5;148(3):dev197392. doi: 10.1242/dev.197392. PMID: 33547203.

Dyring-Andersen B, Løvendorf MB, Coscia F, Santos A, Møller LBP, Colaço AR, Niu L, Bzorek M, Doll S, Andersen JL, Clark RA, Skov L, Teunissen MBM, Mann M. Spatially and cell-type resolved quantitative proteomic atlas of healthy human skin. Nat Commun. 2020 Nov 5;11(1):5587. doi: 10.1038/s41467-020-19383-8. PMID: 33154365; PMCID: PMC7645789.

Zhong F, Jiang Y. Endogenous Pancreatic Beta Cell Regeneration: A Potential Strategy for the Recovery of Beta Cell Deficiency in Diabetes. Front Endocrinol (Lausanne). 2019 Feb 20;10:101. doi: 10.3389/fendo.2019.00101. PMID: 30842756; PMCID: PMC6391341.