NAUTILUS™
BIOTECHNOLOGY

WHITE PAPER

The proteomics revolution - Unlocking the complexity of proteins to fuel precision medicine

When the first draft of the human genome was completed in the early 2000s, many hoped the ensuing genomics revolution would usher in a new era in drug development. Researchers would use their understanding of the genome to create personalized “precision medicines” that target diseases at their root genetic causes and have a high likelihood of clinical success. However, this has not been possible for most diseases and only around 13% of all drugs entering clinical trials make it to market (Wong et al 2019). Billions of dollars still go into developing drugs that ultimately fail (Wouters et al 2020). Here we describe how proteomics can radically improve the drug development process.

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

Table of Contents

The need for a proteomics revolution

Instead of being the hoped-for panacea, the genomics revolution has highlighted just how complex biology is. Genes generally encode proteins, the molecular machines that carry out most cellular processes. A cell's full set of genes, its "genome," is generally static. By contrast, a cell's full set of proteins, its "proteome," is highly dynamic. Proteins move around within cells, and the composition of the proteome depends on a cell's position in the body, age, stimuli, and more. We cannot analyze the proteome using genomics, but the vast majority of approved drugs target proteins (Santos et al 2017). Thus, being blind to the proteome and its complexity makes it difficult to develop effective drugs.

To accelerate drug development, we need to unlock the complexity of proteins and harness the power of proteomics.

Unlike its DNA-based counterpart (the "genome"), if you start talking to people about the "proteome" you'll get blank stares. Most people probably only think of proteins as parts of their diets - mysterious things listed on cereal boxes. But, if the genome is the blueprint for your cells, the proteome is the live tour of your home, complete with your children running from place to place and painting the walls. It provides us with much more information than the genome and changes from day to day, hour to hour, minute to minute.

Given that the proteome is composed of millions of protein isoforms, only recently have researchers been able to capture, store, and analyze large amounts of proteomic data. Advances in nano-scale fabrication and imaging methodologies now make it possible to rapidly characterize nearly all of the proteins in a cell sample with single-molecule precision. With these advancements bolstering us, we're ready for a proteomics revolution that will finally fuel precision medicine and lead to safer, more effective drugs.

Proteomics and drug development - new solutions to well-known problems

As our ability to quantify, localize, and study the individual proteins making up the proteome advances, biological research will change in ways we can't imagine. Nonetheless, we expect the expanded use of proteomics will positively impact drug development in some essential and predictable ways. These include:

IMPROVING OUR UNDERSTANDING OF THE MECHANISMS OF DISEASE

Genomics can identify genetic variants associated with disease. However, simply knowing that a variant is associated with a disease tells us little about if or how the variant causes disease.

Instead, we need to know how genetic variants impact proteins - the molecular machines behind cellular functions. It is the activities of proteins that usually cause cells to be healthy or diseased. Genetic variants can alter proteins in many ways (Figure 1). For example, they might:

- Alter a protein's structure and function
- Keep the protein from functioning at all
- Cause the protein to interact with other parts of the cell in non-productive ways
- Change where the protein localizes in the cell
- Change where the protein is produced in the body
- Change the protein's abundance

In the worst case, a genetic variant might just be a bystander. It might have no mechanistic relevance at all.

Thus, researchers need more than genetic information to understand the mechanisms of disease. Using proteomics, researchers can probe how a genetic variant changes an encoded protein and the cell at large.

For example, let's say a genetic variant encodes a broken protein that normally increases the abundance of other proteins. Breaking this protein decreases the production of "downstream" proteins. Researchers can observe this decrease using proteomics. If they know the functions of the downstream proteins, they can hypothesize reasons why changing their abundance leads to disease. Further experiments will solidify the causal links to disease.

In addition, many diseases have no genetic origins. Their causes instead lie in interactions between cells and the environment. These include interactions with:

- Disease-causing organisms (pathogens)
- Toxic chemicals
- Allergens
- Lifestyle factors

Genomics cannot give us clues about the causes of these diseases, but proteomics can. For example, in some diseases proteins glom onto one another and form "protein aggregates." These aggregates can disrupt cellular functions and lead to such things as neurological disorders. Often, the forces causing protein aggregation aren't genetically determined. Thus we cannot observe protein aggregation with genomics but we can see it with proteomics. When developing drugs to treat aggregates, we can also use proteomics to quickly determine whether the aggregates disappear.

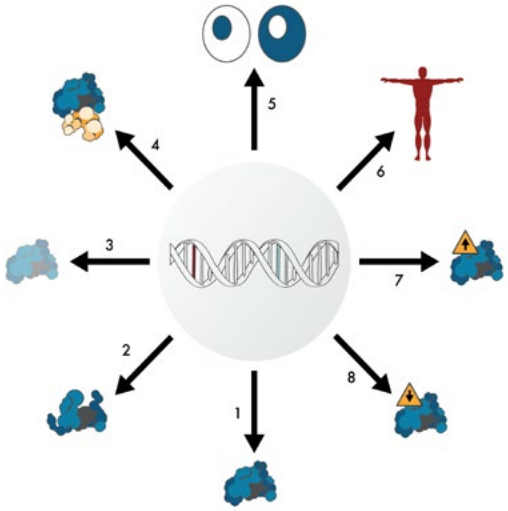

Figure 1. Using proteomics to understand the mechanisms of diease.

Although genomics data and genetic variants (red and blue base pairs) can point researchers to proteins that may be involved in disease, they provide little information about how those variants lead to disease. Genetic variants might (1) not have any impact on the encoded proteins, (2) encode proteins with altered structure, (3) with loss of function, (4) with non-productive interactions, (5) with different sub-cellular localization, (6) with expression in a different part of the body, (7) with increased expression, or (8) with decreased expression.

Proteomics data can help researchers understand how genetic variants lead to disease by indicating which of the mechanisms indicated here are most likely. With this knowledge, they can later create drugs that counteract the mechanisms of disease more directly.

IDENTIFYING POTENTIAL DRUG TARGETS

Even if we have some insights into the mechanisms of disease, it can be difficult to establish possible drug targets. For that, we need an empirical evaluation of protein abundance in diseased and healthy cells. Proteomics technologies can identify proteins with low abundance in healthy cells and high abundance in diseased cells. Proteomics can also tell us whether these proteins are at cellular locations that are easy to reach with drugs.

For example, with proteomics, researchers can identify proteins found at high levels on the surface of diseased cells but not healthy ones. These proteins may be easy to target with antibodies and other kinds of drugs.

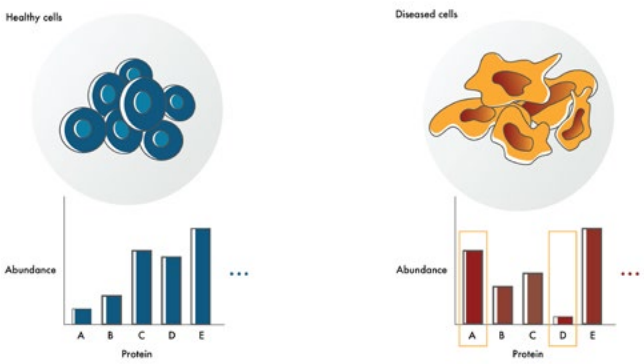

Figure 2. Identifying possible drug targets

Researchers can use proteomics to identify proteins that are at elevated or decreased levels in diseased cells compared to healthy cells (like proteins A and D). The proteins that are elevated in diseased cells may make good drug targets. Those with decreased levels may explain disease mechanism.

MAKING EFFECTIVE DRUGS WITH FEW SIDE EFFECTS

Drug developers cannot go straight from a list of protein abundances in healthy and diseased cells to the clinic. First, they must create drugs that interact with their protein targets in ways that actually impact disease. In addition, they must show that these drugs don't cause side effects. Proteomics helps them do both.

Differences in the proteomes of healthy and diseased cells provide objective measures of disease. If a drug makes a diseased cell's proteome look like that of a healthy cell, it is likely to be effective. Here the proteome, and potentially a select set of protein abundances, acts as a "biomarker." That is, an objective, biological measure of drug effectiveness and disease severity.

Yet, even if a drug appears to mitigate disease, it may have unbearable side effects for patients. These can be caused by unanticipated drug-protein interactions. Proteomics can help drug developers avoid side effects by giving them a more holistic understanding of such drug-protein interactions.

Indeed, undesirable interactions like these manifest when drug developers focus too much on interactions between drugs and their intended targets. They may make drugs that bind really well to their targets. Yet, if they don't know how a drug impacts other proteins, they won't get a full understanding of its "mechanism of action." The drug might bind to other "off-target" proteins causing "off-target toxicity." It might even cause the target to interact with other cellular components in unexpected ways (Roberts 2018). If they choose a different target, they may be able to create drugs that interact with it more specifically. However, if they don't start with a holistic understanding of drug-protein interactions, they won't realize this is a problem.

Researchers don't usually see the effects of unexpected drug-protein interactions until the drug enters the clinic. As a result, researchers create drugs that have small "therapeutic windows." At a low dose, these drugs selectively interact with the appropriate targets and have positive effects. However, as the dose increases, unintended interactions begin and toxicity dominates.

We anticipate advances in analyzing and quantifying proteins will give researchers a more holistic understanding of drug-protein interactions. Before dosing patients and discovering all sorts of unexpected toxicity, drug developers can instead look for signs that the drug and/or interactions with its target cause unwanted changes to the proteome. They can eliminate potential targets before developing drugs against them and stop studies before entering expensive clinical trials.

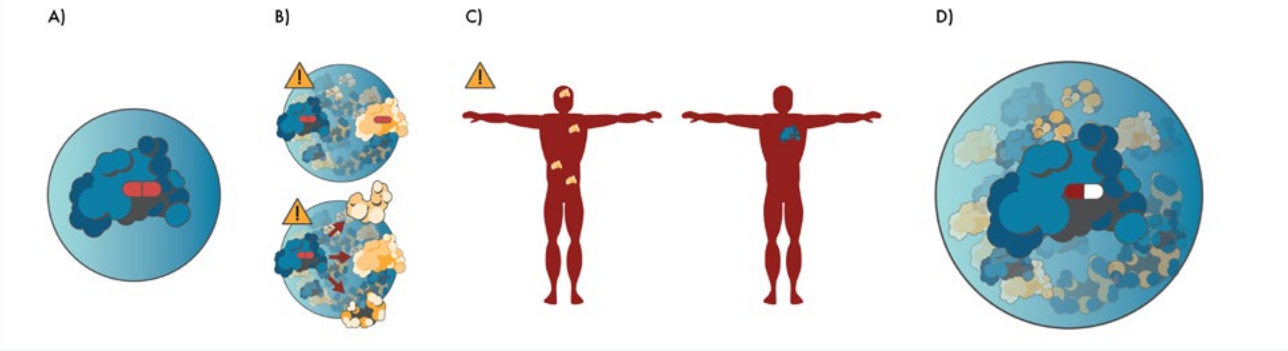

Figure 3. Preventing toxicity through proteomics

A) Researchers often create drugs (red pill) that target proteins in isolation.

B) However, this can lead to many unexpected interactions between the drug and other proteins (like the yellow protein above) once the drug enters cells. Alternatively, the drug may cause the target protein to interact with other proteins. These unintended interactions can lead to cellular malfunctions and toxicity.

C) Drugs created this way may also target proteins that are expressed at many different places in the body (left, yellow protein). These may be bad drug targets if the proteins only cause disease associated with a particular part of the body. Better drug targets (right, blue protein) may only be expressed in the part of the body where they cause disease.

D) Proteomics gives researchers a more holistic view of protein interactions during the drug development process. This enables them to create drugs (red/white pill) that are unlikely to have have "off-target" interactions and which target proteins that are only expressed at the disease site.

The proteomics revolution is upon us

The promises of proteomics are many. By knowing exactly what proteins are in a given cell, it will be easier to:

- Understand the mechanisms of disease
- Identify possible drug targets
- Make drugs that effectively treat disease with fewer side effects

Until now, the technologies available for measuring the proteome have fallen behind the promise. These technologies classically rely on mass spectrometry. They have pushed the field forward and made the importance of the proteome clear, but they are not accessible to many researchers and generate data that can be difficult to analyze.

This is perhaps unsurprising given the complexity of the proteome and the challenges involved in quantifying it. Large abundance proteins can easily drown out small abundance proteins in individual cells and individual proteins themselves consist of 20 different amino acids. These interact with one another in intricate ways in 3D space.

In addition, single genes can encode many different forms (isoforms) of proteins. All of these isoforms can be further modified in various ways. From thousands of genes come millions of proteins.

On top of all that, the proteome is in constant flux. Proteins move around the cell and their abundances change over time. While the genome you're born with is the genome you die with, the proteome changes from day-to-day, cell-to-cell.

At Nautilus, we've developed a platform of technologies that demystify this complexity using single-molecule analysis and machine learning. Our platform leverages massive chips composed of billions of wells that isolate single proteins (one protein per well). These are similar to the ever-improving microchips in your computer. We repeatedly flow fluorescent affinity probes over each protein in a chip and these probes selectively bind to motifs found in some fraction of the proteins. Using optical imaging techniques, we can look at each well and determine if binding occurs.

Our platform can measure many proteins at once because our affinity probes bind to multiple protein species, but this also makes binding measurements noisy. This is where machine learning comes in. Sometimes an affinity probe will fail to bind a protein even if it has a target motif. Other times affinity probes will bind to proteins that don't have their target motifs. Machine learning algorithms assess these stochastic binding events and determine which of a known set of proteins is most likely in a given well. In doing so, our technology effectively extracts signal from noise - something researchers would have a hard time doing on their own.

The simultaneous, repeated analysis of billions of protein molecules provides us with unprecedented sensitivity and dynamic range. By tallying up the proteins in each well, we build a full picture of the proteome from data collected at the scale of single protein molecules and deliver the user a simple list of proteins and their abundances. We can even look at very low-abundance proteins whose signals would be drowned out by high-abundance proteins on other analysis platforms. This is the biological equivalent of finding a needle in a haystack!

Of course, this requires our machine learning tools to build their predictions from known proteins. Toward this end, we are actively working with researchers to train our platform. We believe we'll be able to accurately identify 95% of the human proteome soon.

Democratizing proteomics and fueling precision medicine

With our platform and simple readout, we will democratize proteomics. Researchers in biomedicine, regardless of their background, will be able to use our technologies. In fact, we hope to make it possible for anyone who wants a proteome to get one. The platform will store researchers' proteomic data and give them access to standardized analysis tools.

This accessible proteomics platform will catalyze drug development. Using it, researchers will be able to establish proteomes for any cells they wish. With these in hand, it will be easier to figure out exactly what's gone awry in diseased cells and create drugs that target diseases at their molecular causes with few side effects.

The wave of precision and personalized medicine is coming soon, fueled by a proteomics revolution. Individuals will feel the benefit of this proteomics revolution in the form of more effective therapies. They'll think of proteins as drivers of health and not just items listed on cereal boxes.

References

Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics. 2019 Apr 1;20(2):273-286. doi: 10.1093/biostatistics/kxx069. Erratum in: Biostatistics. 2019 Apr 1;20(2):366. PMID: 29394327; PMCID: PMC6409418.

Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. JAMA. 2020 Mar 3;323(9):844-853. doi: 10.1001/jama.2020.1166. PMID: 32125404; PMCID: PMC7054832.

Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, Overington JP. A comprehensive map of molecular drug targets. Nat Rev Drug Discov. 2017 Jan;16(1):19-34. doi: 10.1038/nrd.2016.230. Epub 2016 Dec 2. PMID: 27910877; PMCID: PMC6314433.

Roberts RA. Understanding drug targets: no such thing as bad news. Drug Discov Today. 2018 Dec;23(12):1925-1928. doi: 10.1016/j.drudis.2018.05.028. Epub 2018 May 24. PMID: 29803936.