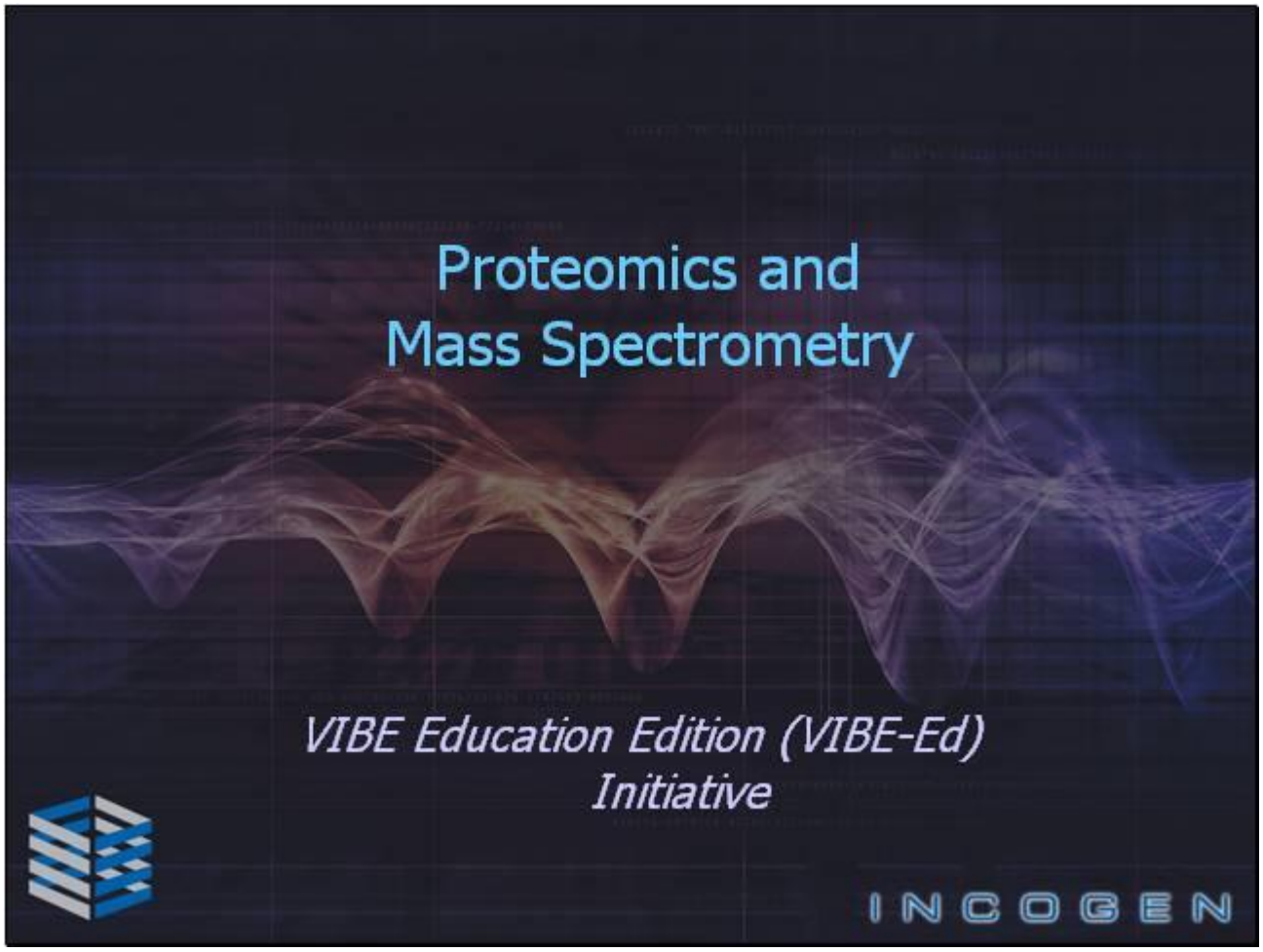
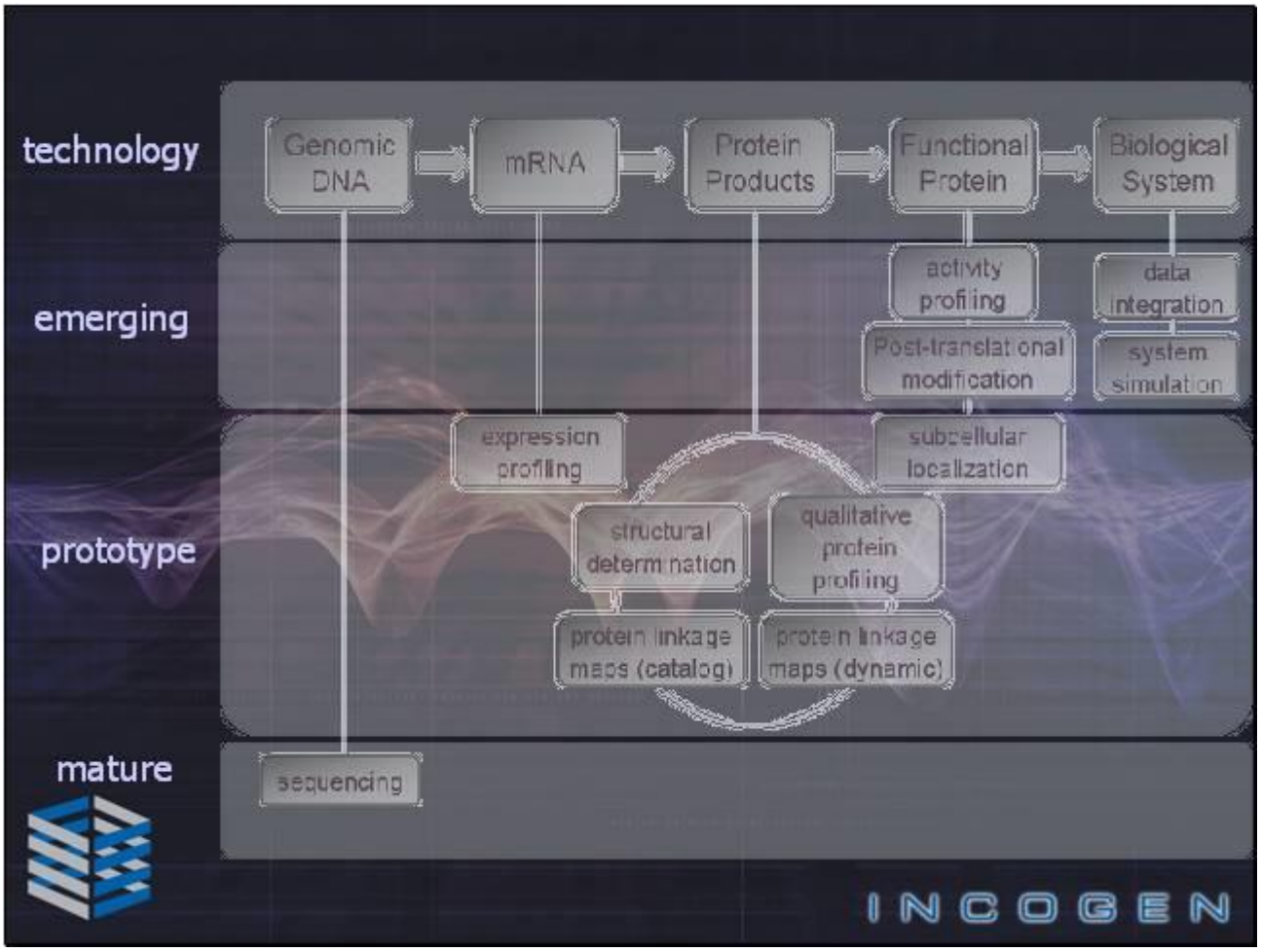
Slide 1 - Proteomics and Mass Spectrometry



Slide notes

This discussion is about proteomics and mass spectrometry.

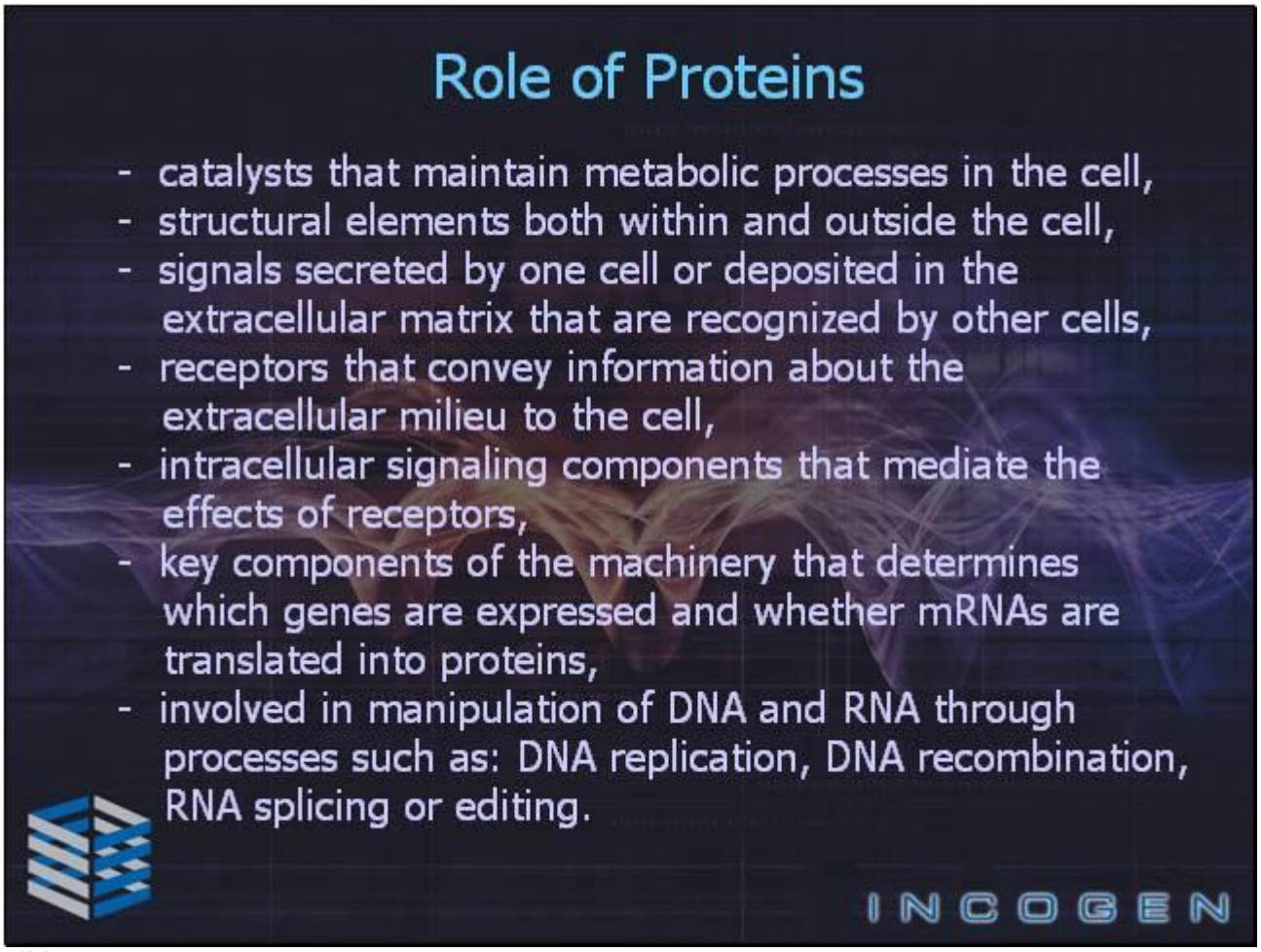
Slide 2 - Slide 2



Slide notes

Proteomics is the study of the structure, function, and interactions of the proteins produced by genes. As we move to these more complex systems, the technologies become less mature, in part because things become much more complicated and require the combination of expertise in different fields. For example, activity profiling has recently become of great interest; this involves measuring protein kinetics: where and when proteins move, how fast they move, the mechanisms of movement, degradation, and recycling. These are all important to understanding the life cycle of a gene's response to stimuli, and the study is just now getting underway.

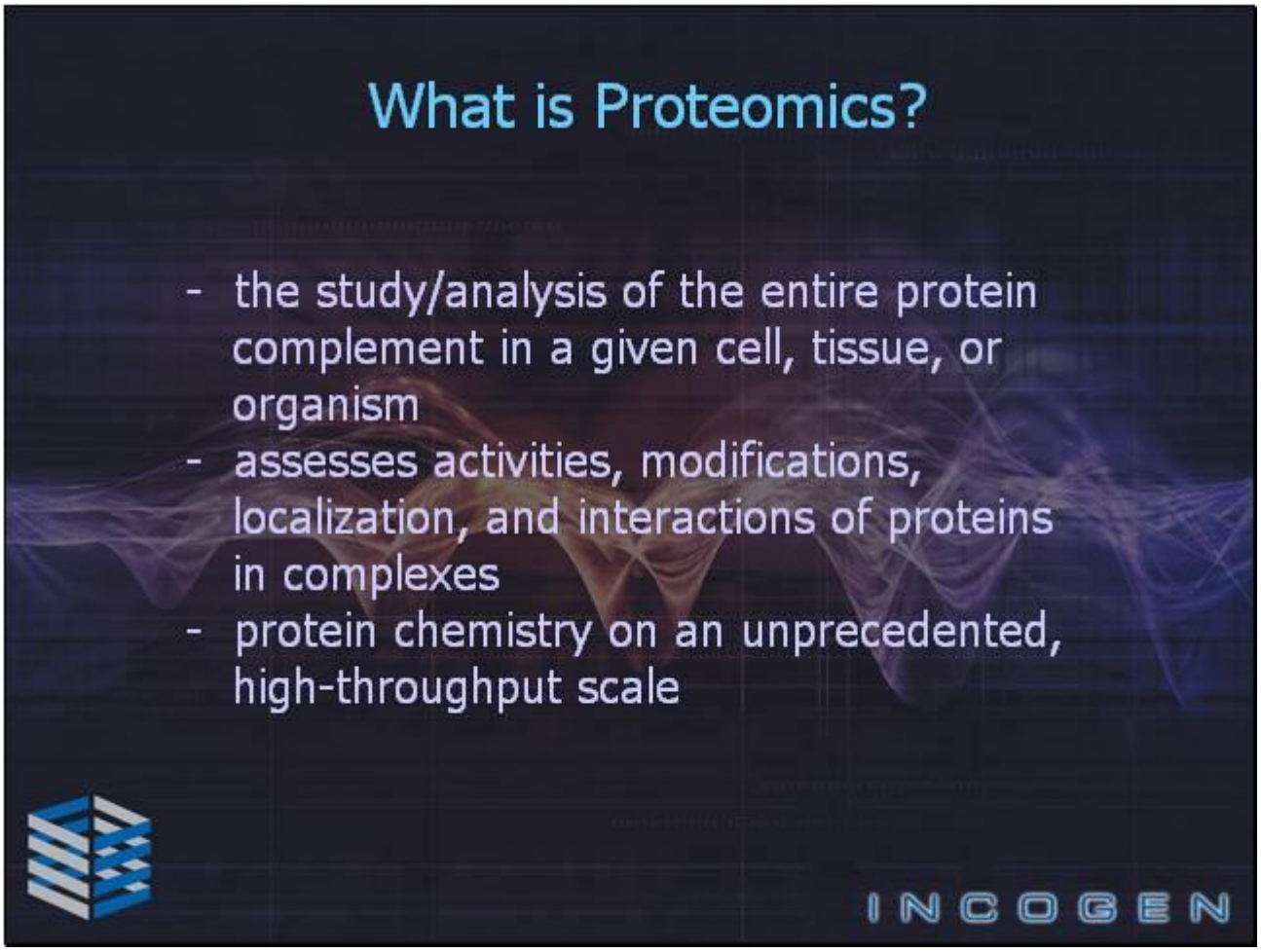
Slide 3 - Role of Proteins



Slide notes

What are some of the roles proteins play in our bodies? They act as catalysts that maintain metabolic processes in the cell. They are structural elements, both inside and outside the cell. They facilitate communication between cells and within the cell, and they act as sensors that cells use to remain in contact with the area around them. They are key components that help determine which genes are expressed, and they are involved in such processes as DNA replication and RNA splicing.

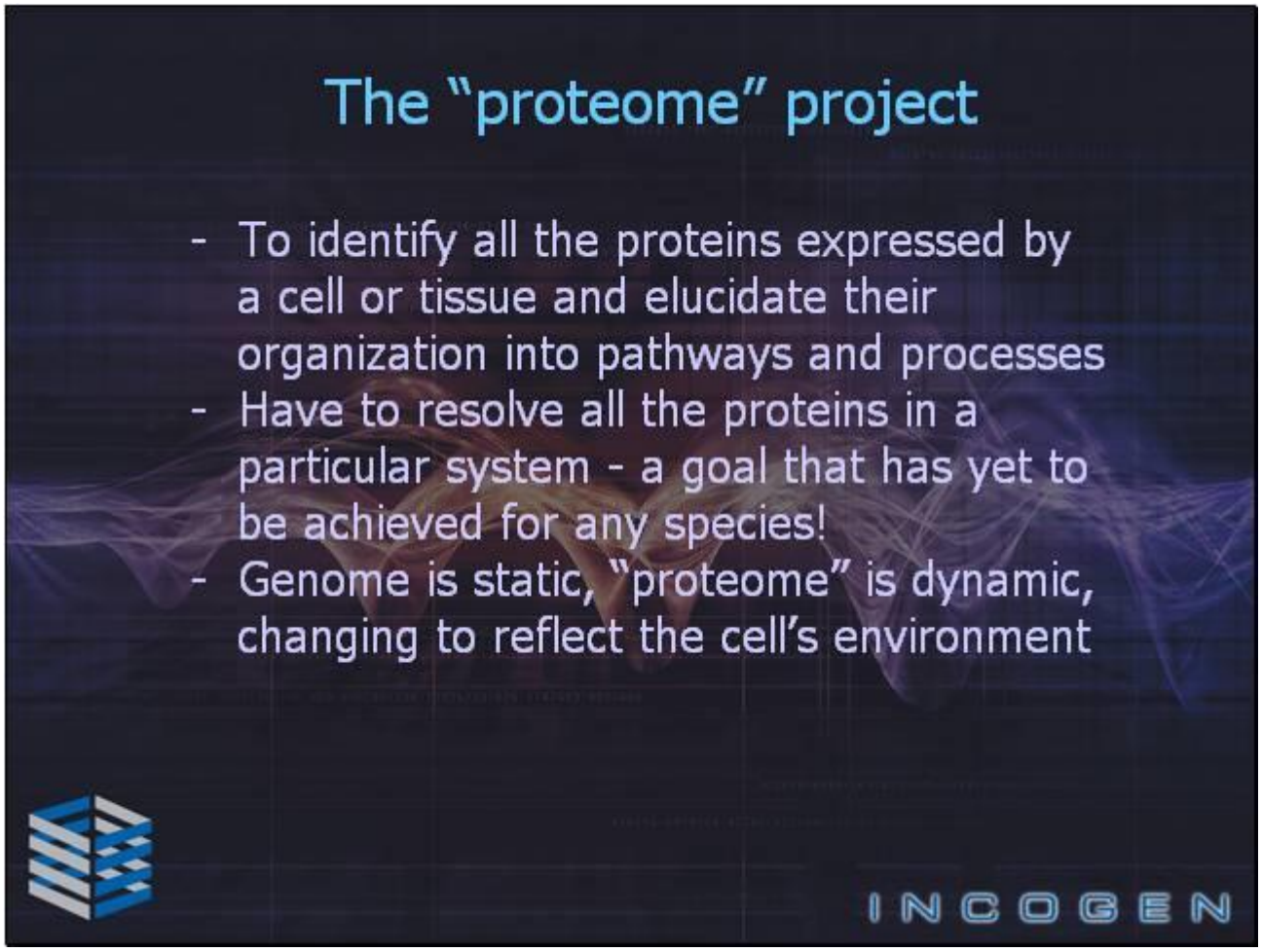
Slide 4 - What is Proteomics?



Slide notes

Proteomics is the study of the entire protein complement in a given cell, tissue, or organism. It involves the analysis of the activities, modifications, localization, and interactions of proteins in cells, tissues, etc. It is considered to be the next step in modern biology.

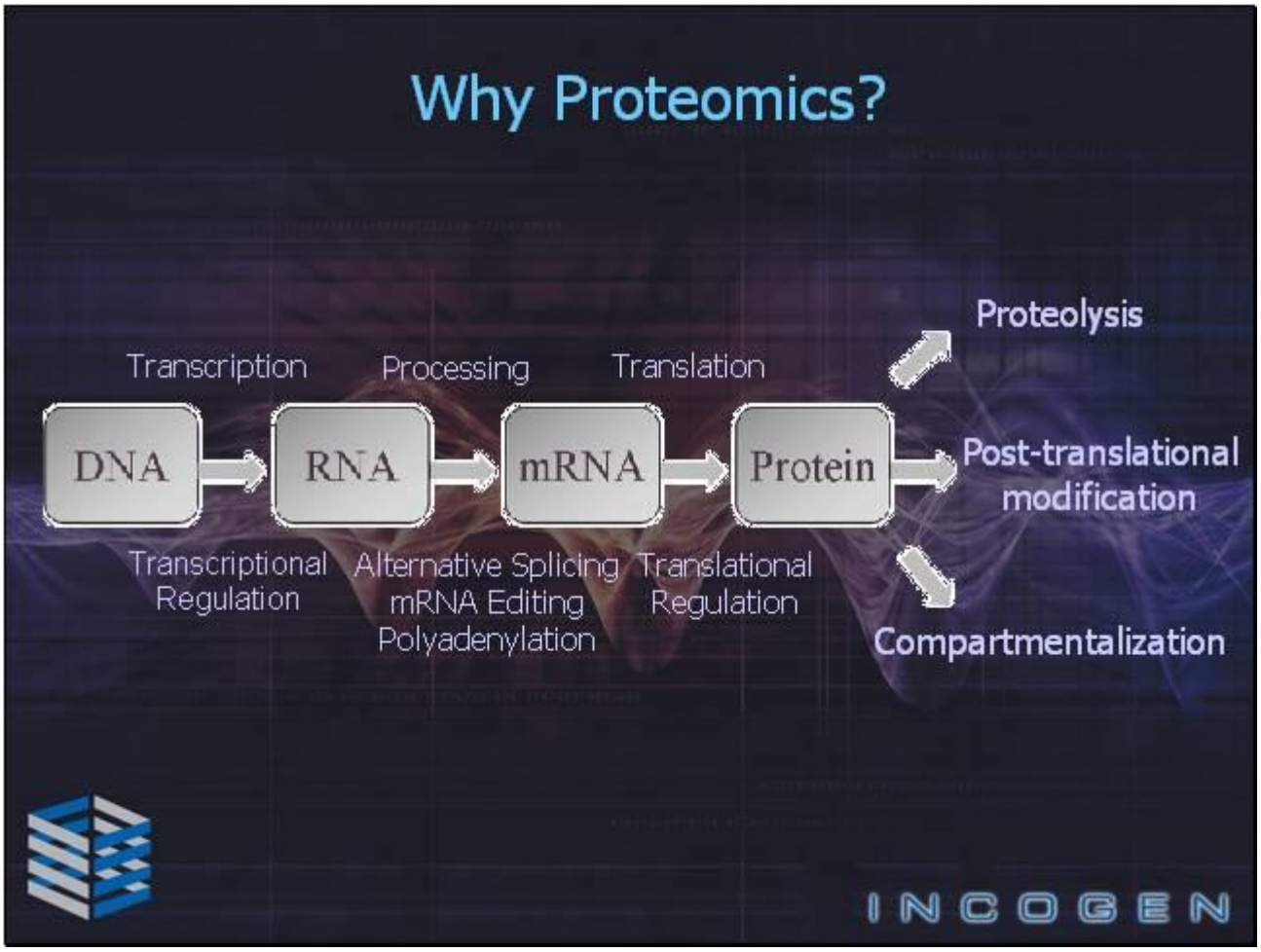
Slide 5 - The “proteome” project



Slide notes

The goal of the “proteome” project is to identify all of the proteins expressed by a cell or tissue and elucidate their organization into pathways and processes. This is quite a goal, considering that we have yet to resolve all of the proteins in any one species. In many ways, the genome is easier to characterize because it is static; the “proteome” is always changing to reflect the cell’s environment.

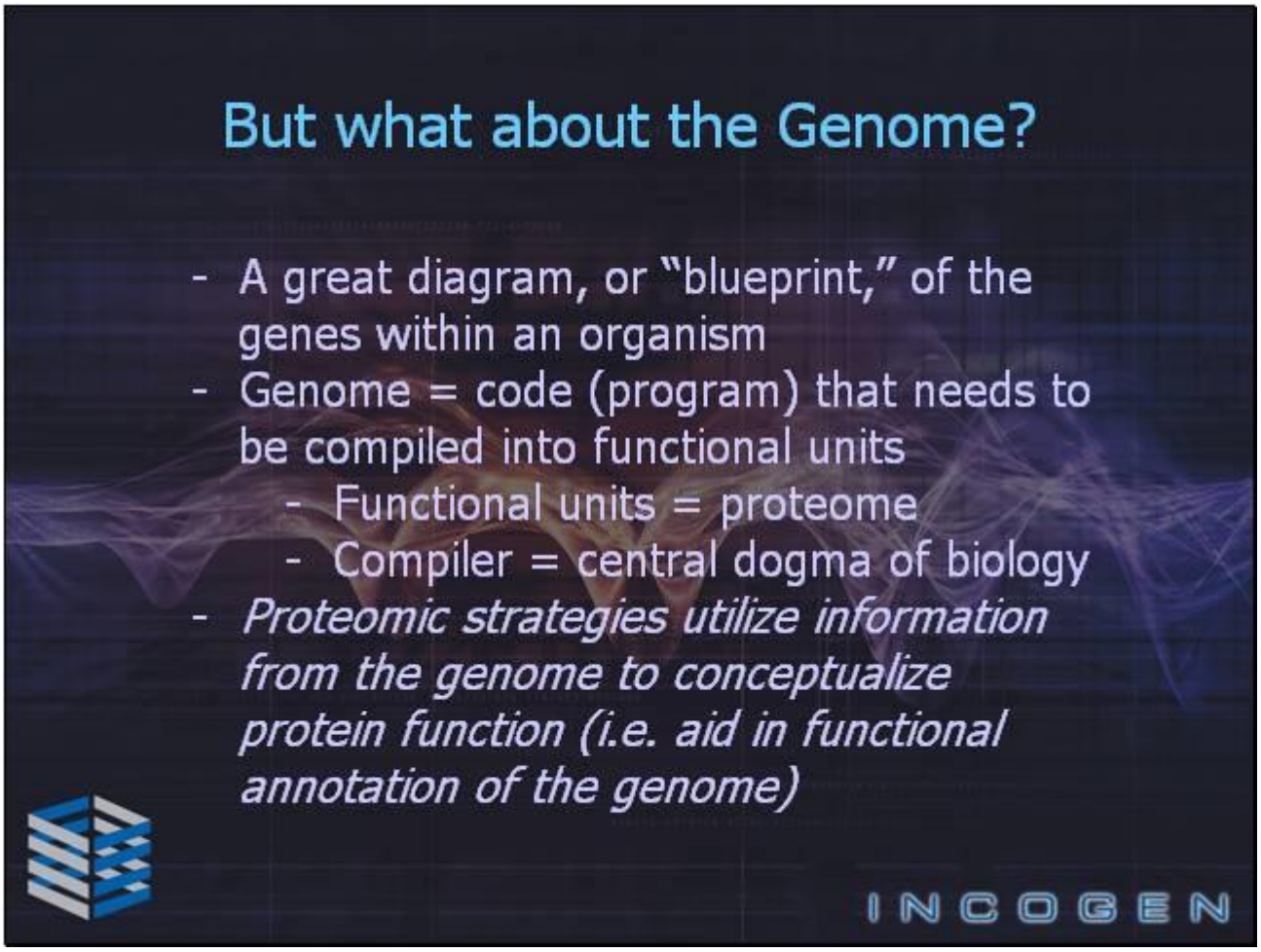
Slide 6 - Why Proteomics?



Slide notes

Why do we study proteomics? Humans have about 30000 genes, fewer than some “lower” organisms, but there are 8 different regulatory mechanisms at four different times that can influence which proteins will be created from these genes. Proteomics allows us to compare control and experimental samples to look for differences in the amount and type of proteins present. Additionally, we can look at the proteins directly instead of inferring the information from the genome.

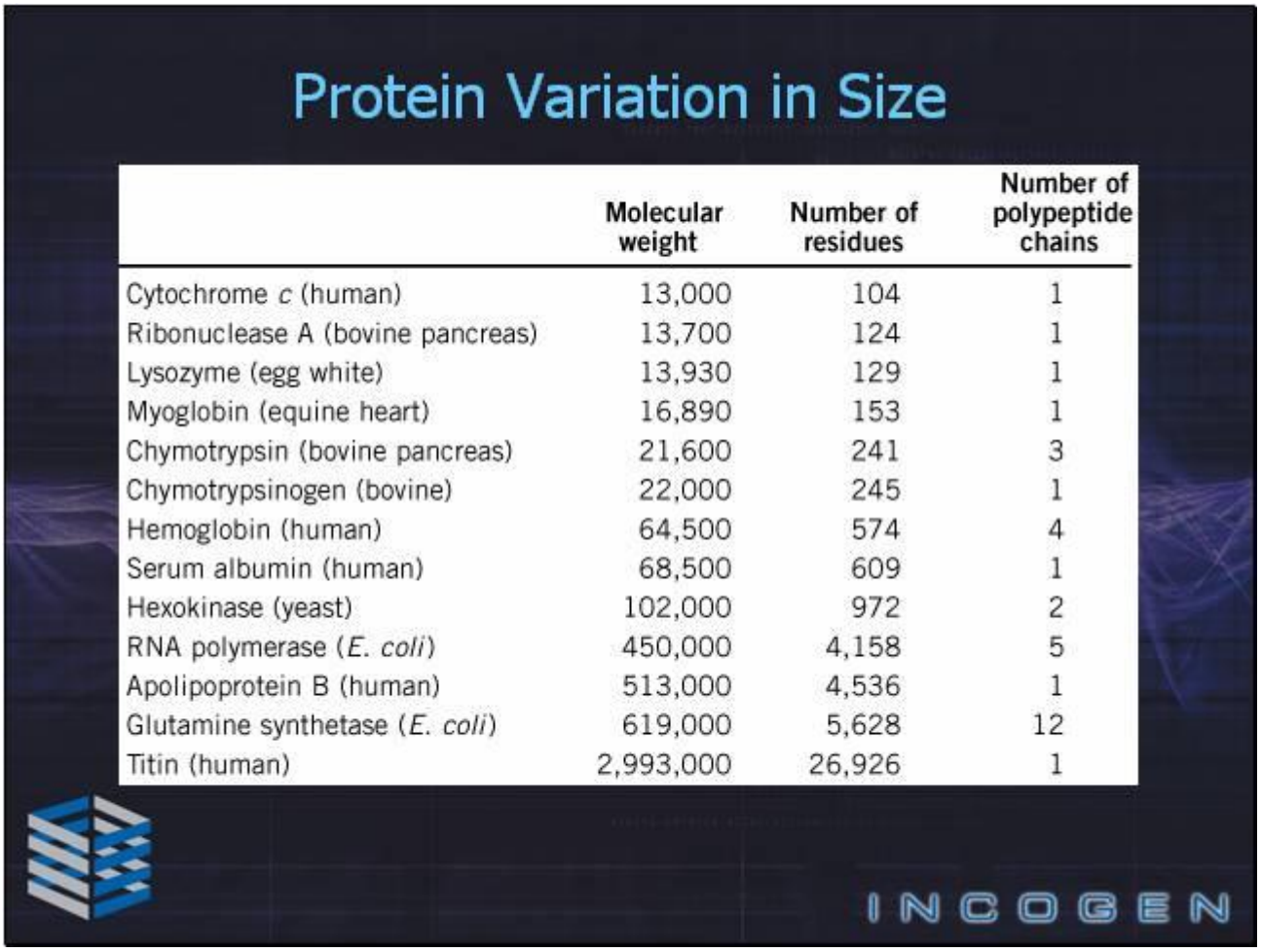
Slide 7 - But what about the Genome?



Slide notes

If proteomics is so important, why bother studying genomics? The genome is an excellent blueprint of the genes within an organism. Using a computer science analogy, the genome is the code or program that needs to be compiled into a proteome before it can be used. Proteomic strategies use information about the genome to conceptualize protein function.

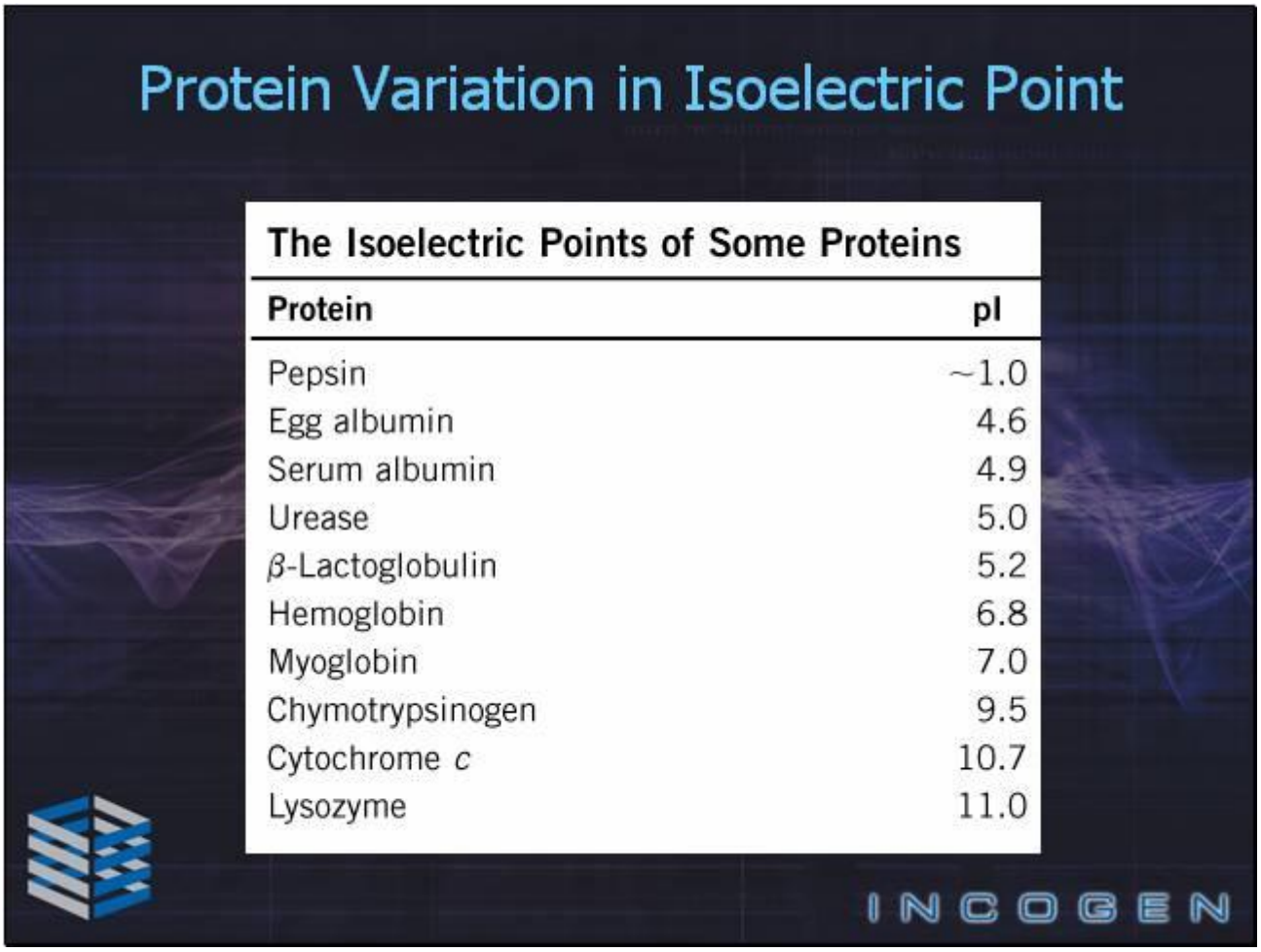
Slide 8 - Protein Variation in Size



Slide notes

There is a large variation in the physical characteristics of proteins. This table lists the weight and length of a representative sampling of proteins. Note that they both can vary by several orders of magnitude from smallest to largest.

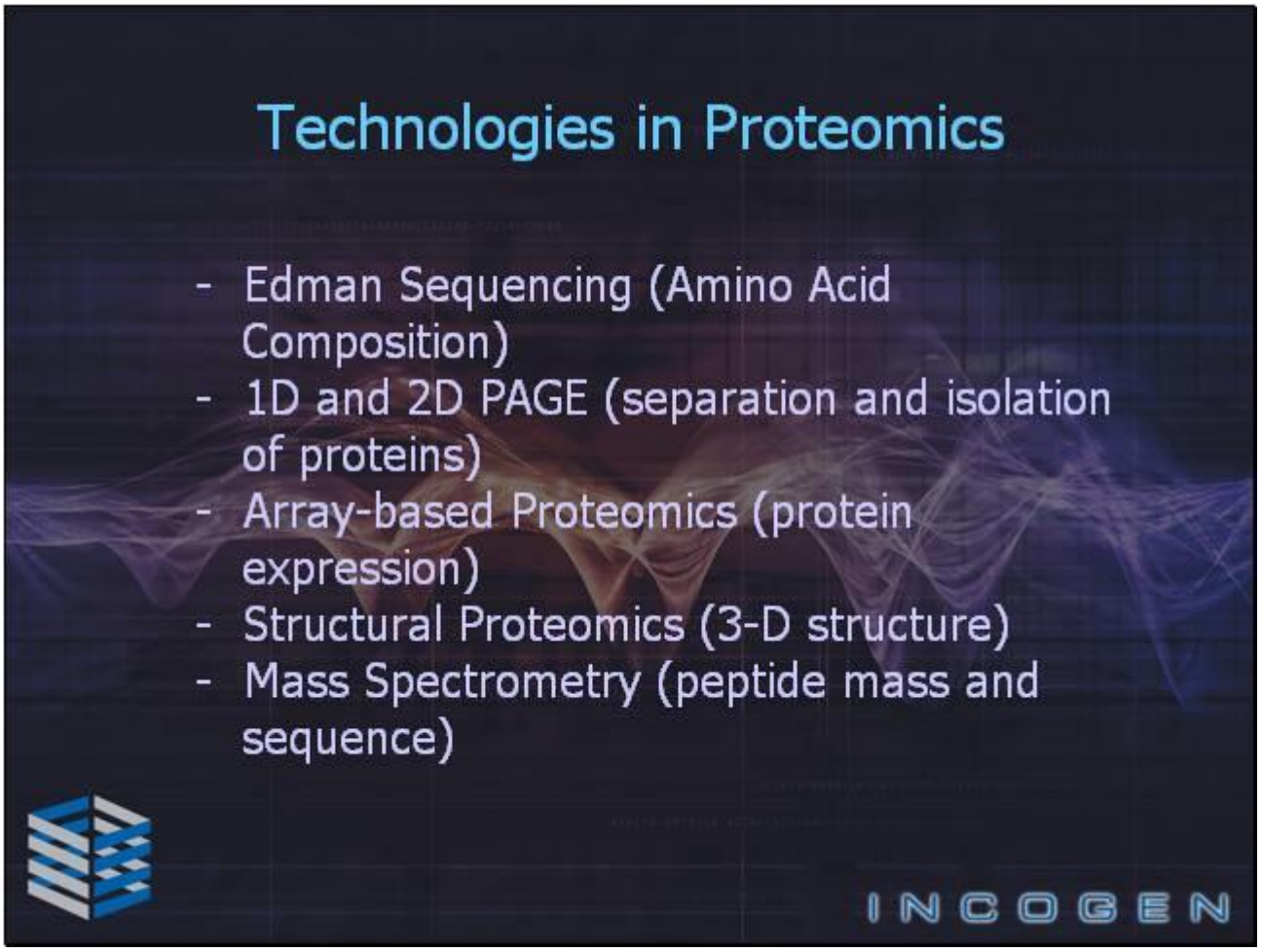
Slide 9 - Proteins Variation in Isoelectric Point



Slide notes

This table shows the variation in isoelectric point for a sampling of proteins. The isoelectric point is the pH at which the protein has an equal number of positive and negative charges. Just from these few proteins, we can tell that the isoelectric point varies from extremely acidic to neutral to fairly basic.

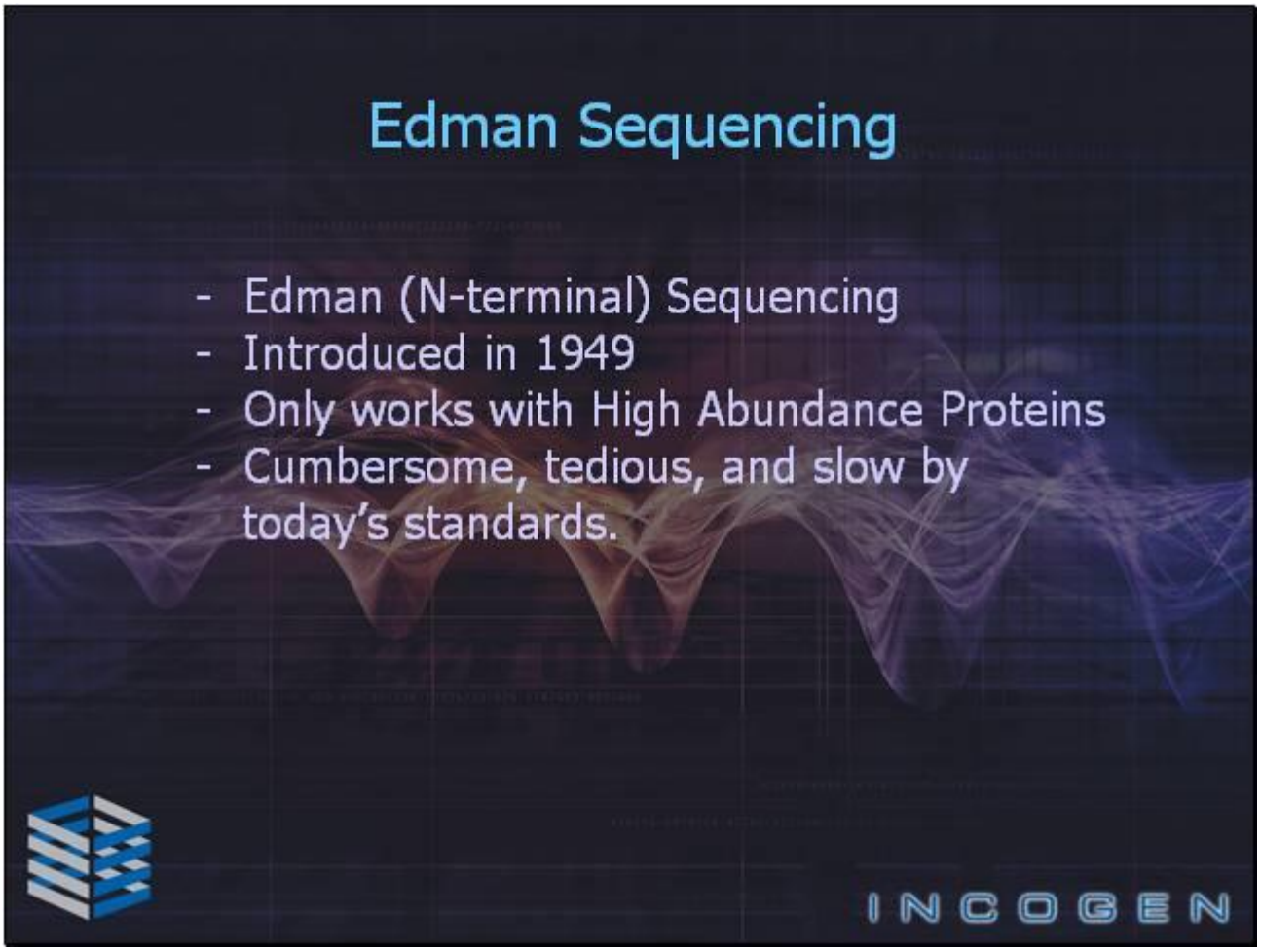
Slide 10 - Technologies in Proteomics



Slide notes

During the remainder of this lecture, we will briefly discuss some common technologies used to study proteins.

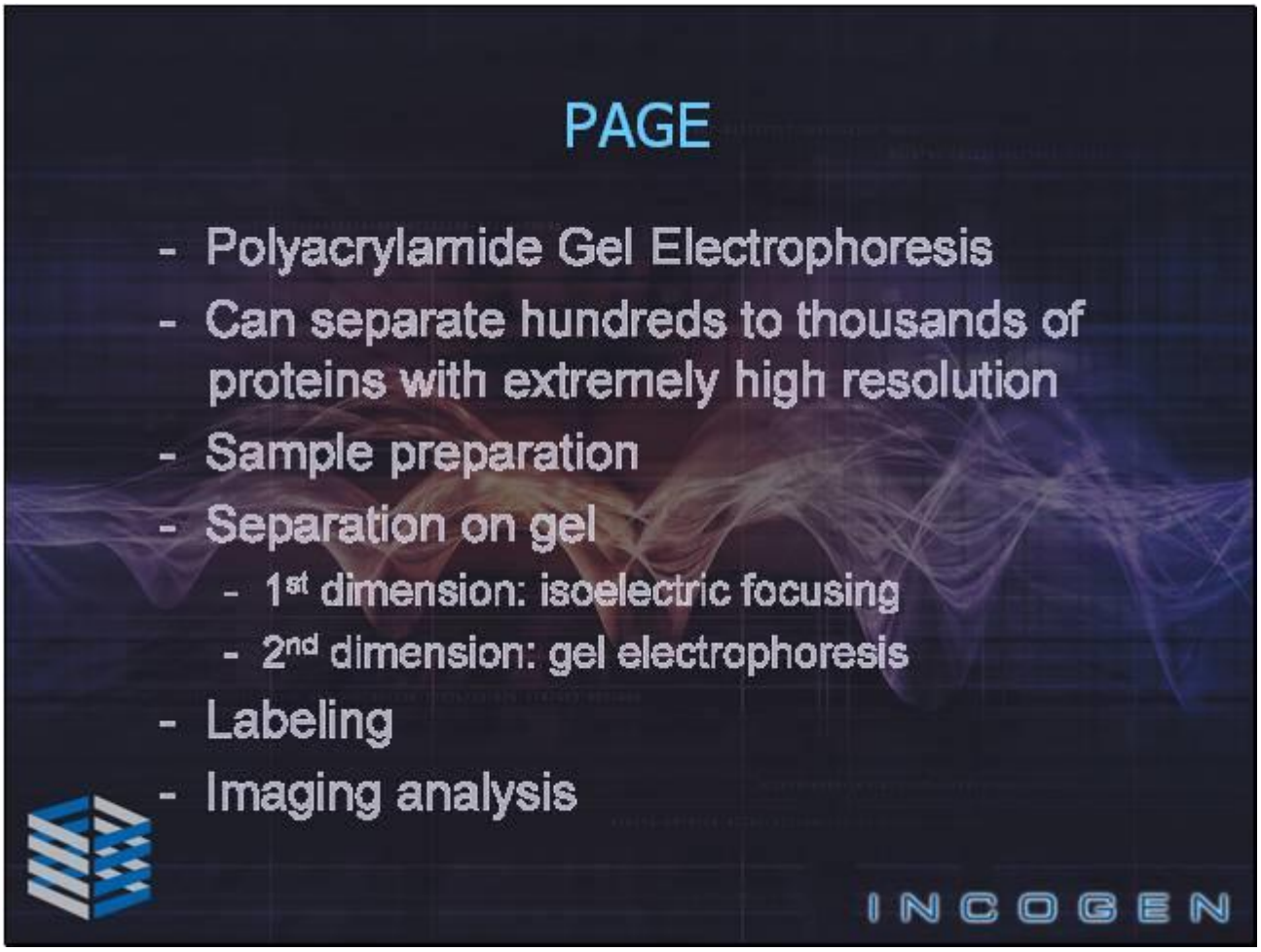
Slide 11 - Edman Sequencing



Slide notes

Edman, or N-terminal, sequencing was one of the first methods used to study proteins. It was introduced in 1949 and only works with high abundance proteins. By today's standards, it is cumbersome, slow, and tedious.

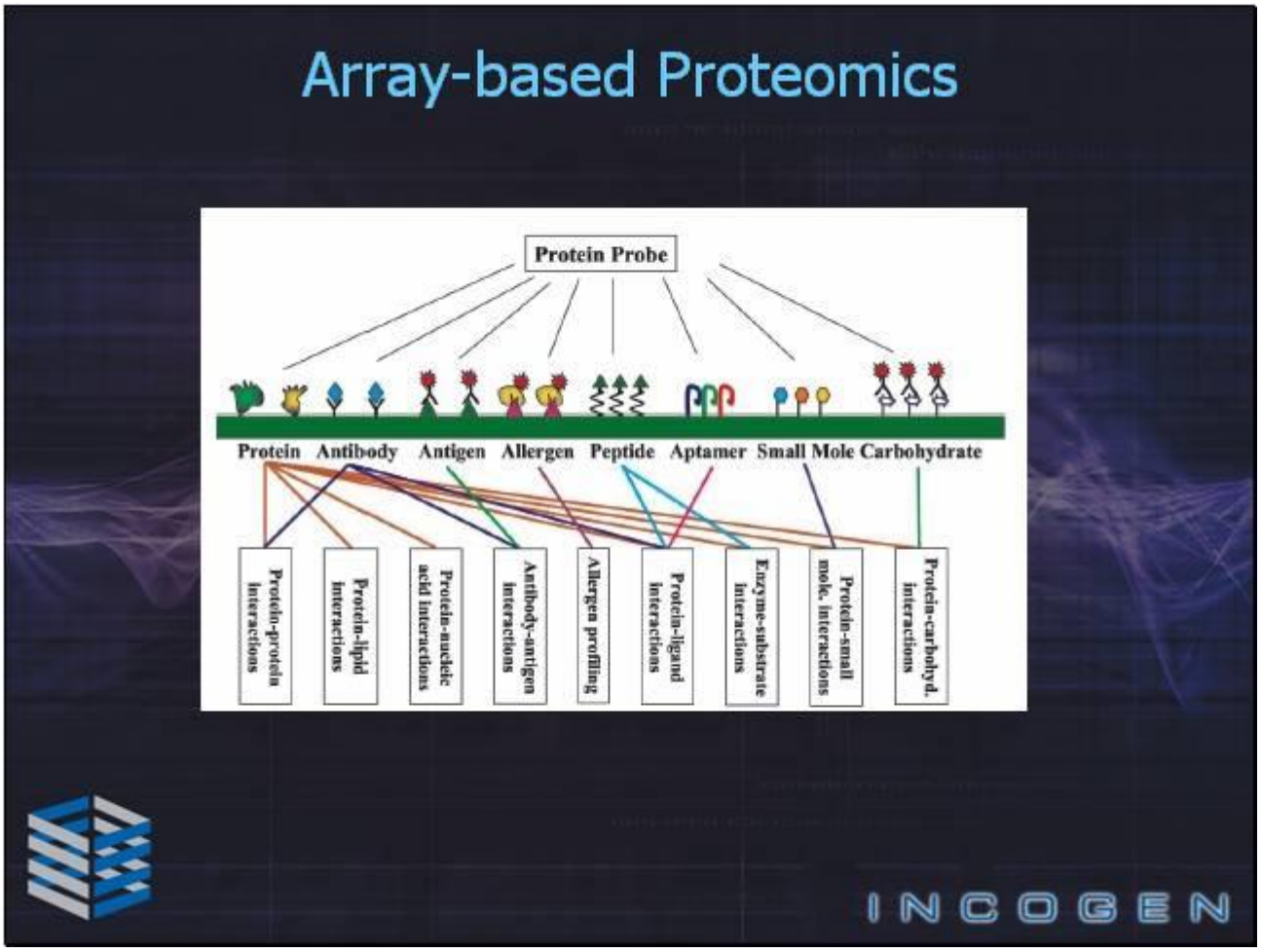
Slide 12 - PAGE



Slide notes

PAGE, or polyacrylamide gel electrophoresis, can separate hundreds to thousands of proteins with extremely high resolution. Since its inception more than 30 years ago, the process has not changed; it works by separating proteins using isoelectric focusing in one dimension and their molecular weights using gel electrophoresis in the other. Although this technique is very precise, it has drawbacks. Spots often overlap, making identifications difficult. The process is slow and tedious and contains “open” phases where contamination can occur. It is seen by many to be an art rather than a science.

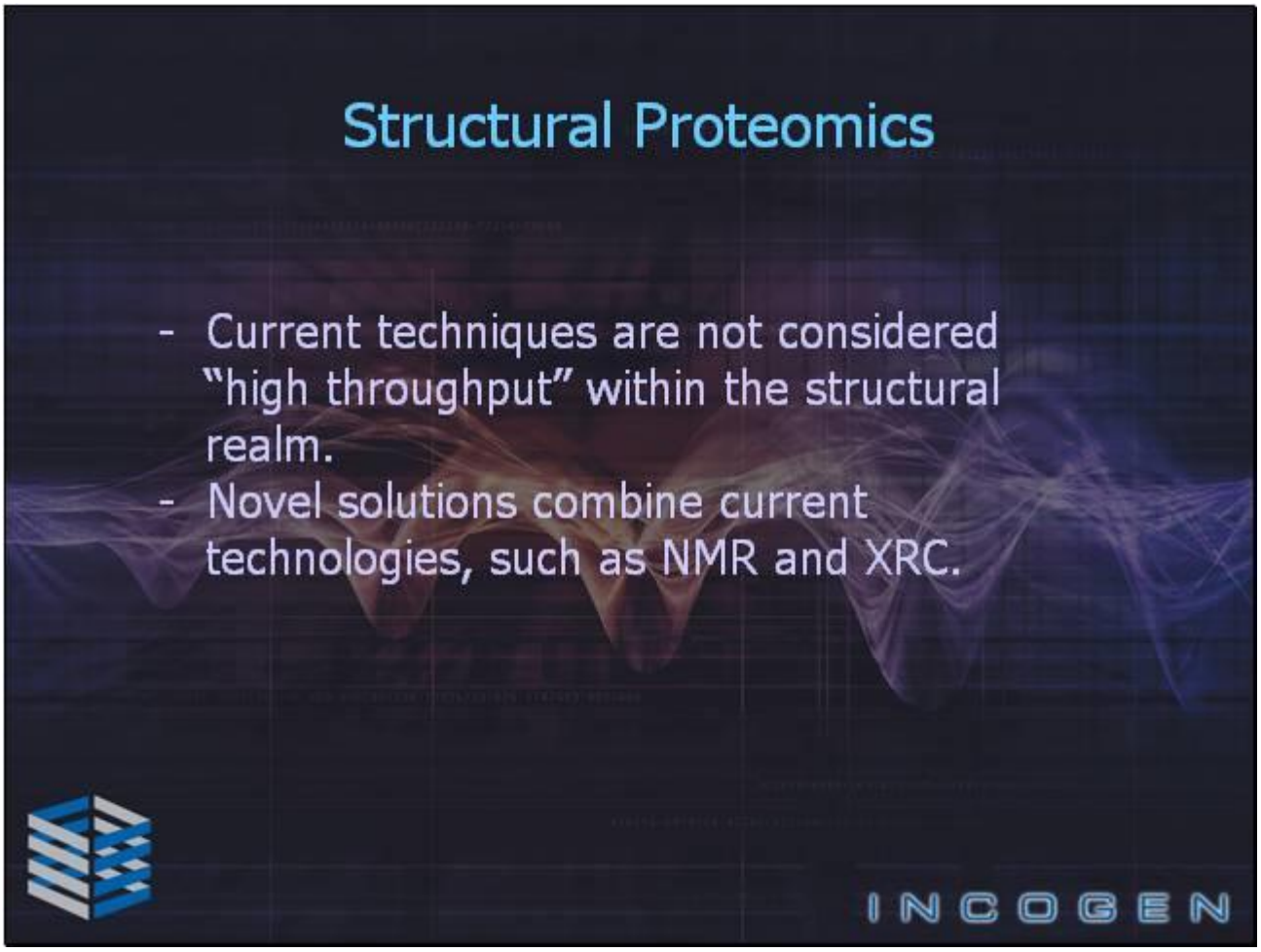
Slide 13 - Array-based Proteomics



Slide notes

We also use experiments similar to microarray experiments to determine under which conditions proteins are expressed and with what other proteins they interact.

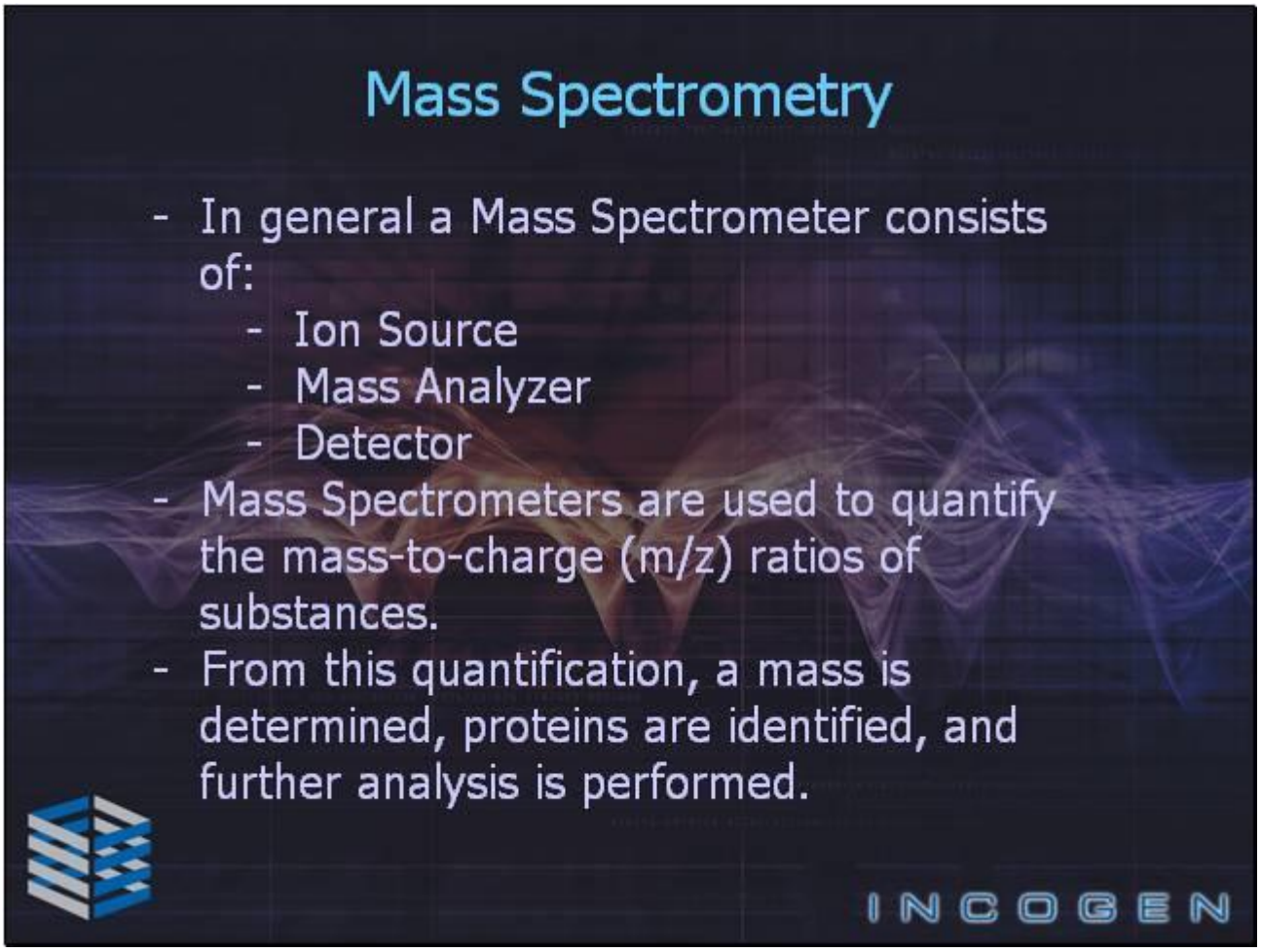
Slide 14 - Structural Proteomics



Slide notes

Structural proteomics studies the physical structure of proteins, both in their natural states and crystallized. Current techniques are slow and tedious, and so are not considered “high throughput” within the structural realm.

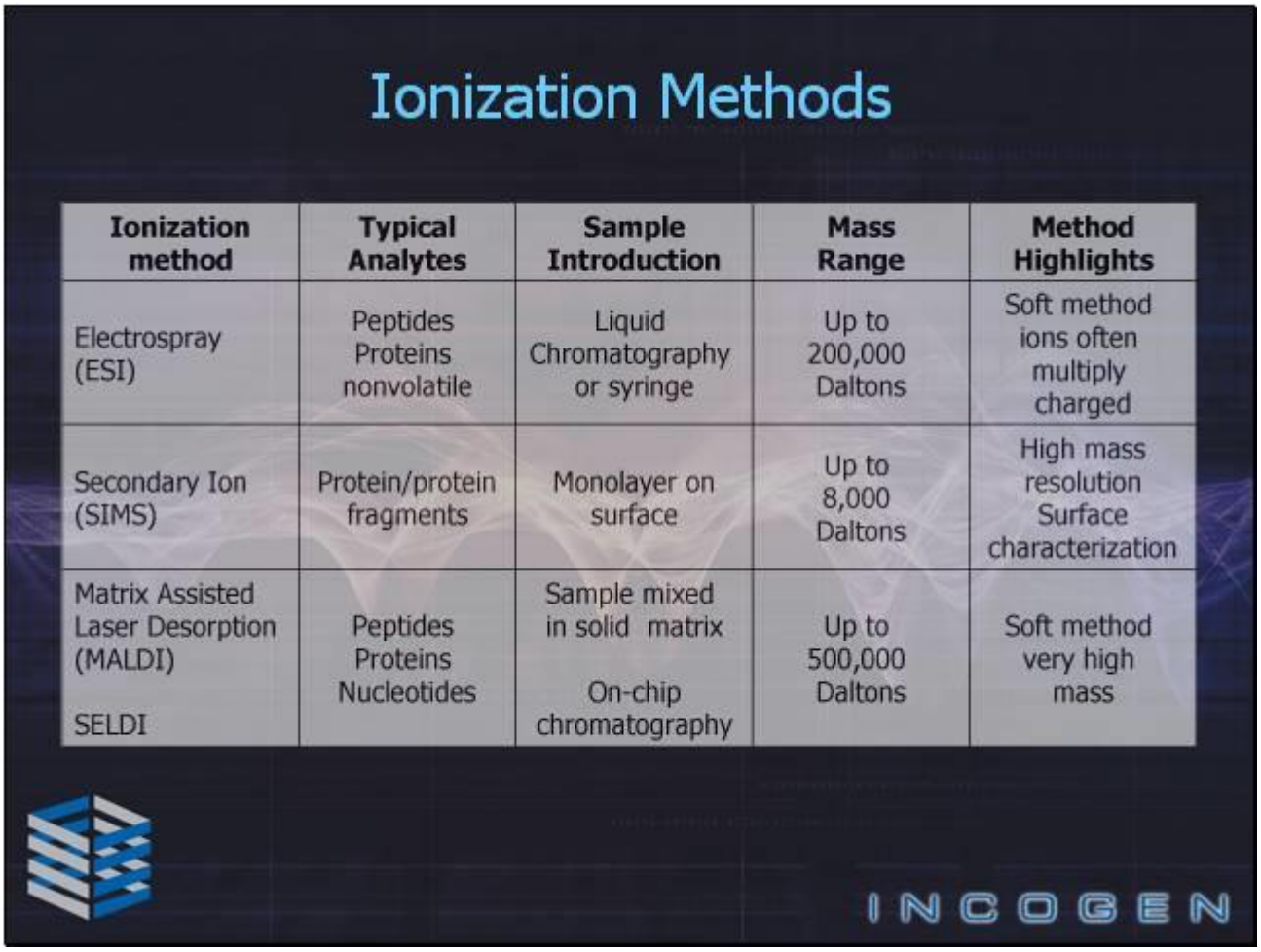
Slide 15 - Mass Spectrometry



Slide notes

Mass spectrometry is also used to identify the composition and structure of proteins. A mass spectrometer consists of an ion source, a mass analyzer, and a detector. They are used to quantify mass-to-charge ratios. From this quantification, a mass can be determined and the proteins can be identified.

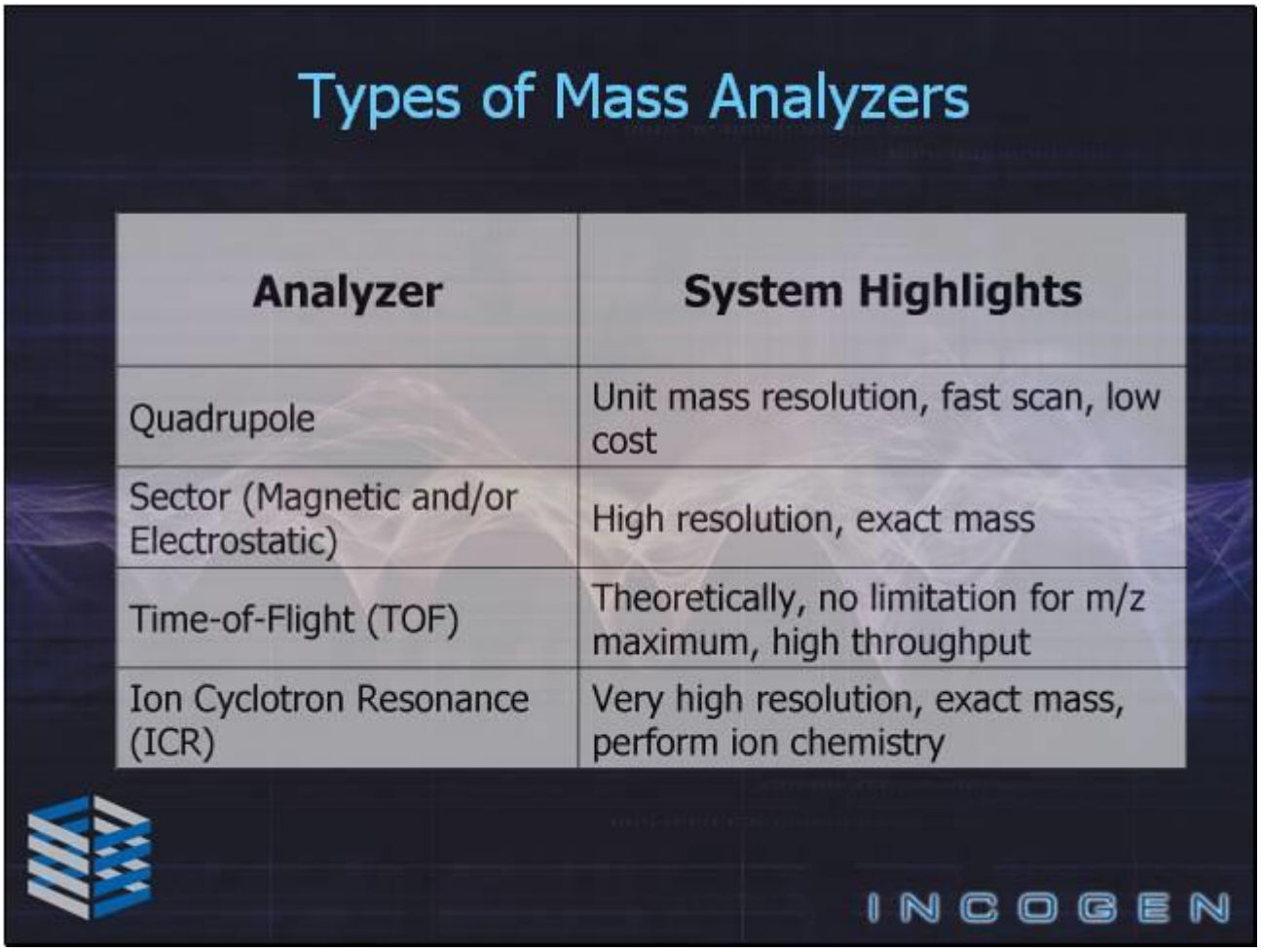
Slide 16 - Ionization Methods



Slide notes

The first step in mass spectrometry is to ionize the substance that is being analyzed. There are a variety of methods that are used to ionize samples, depending on their state and approximate molecular weight. Electrospray is typically used for post-source decay termination of amino acid sequences. Matrix assisted laser desorption ionization time of flight, or MALDI-TOF, is used mainly for peptide mass mapping, although some newer machines are sensitive enough for protein sequencing. SELDI is used for a wide range of protein profiling and is considered a fairly new technology.

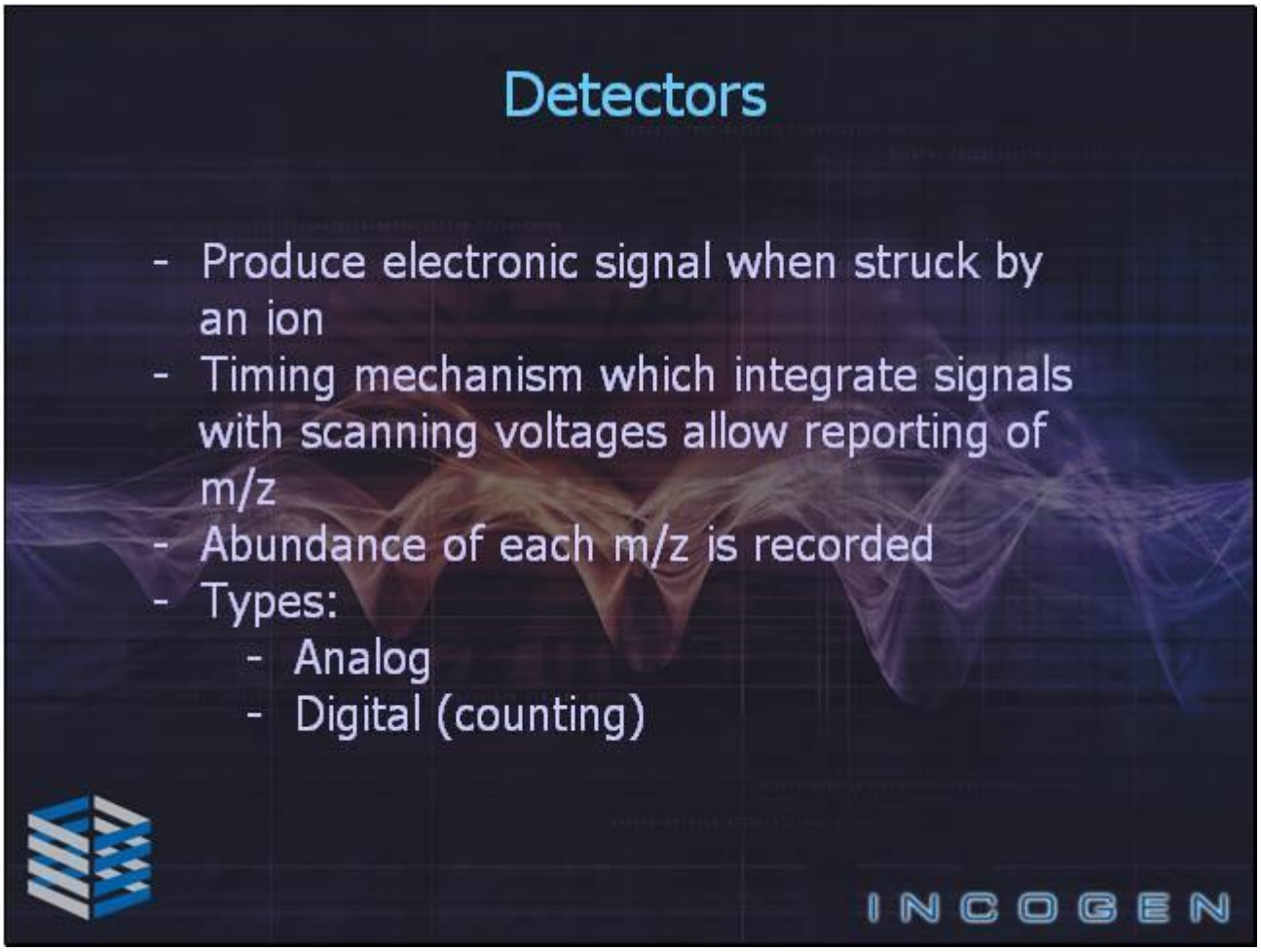
Slide 17 - Types of Mass Analyzers



Slide notes

There are also a variety of mass analyzers that can be used. They each have their strengths. For example, the quadripole is fast and relatively cheap, while the sector gives more accurate masses.

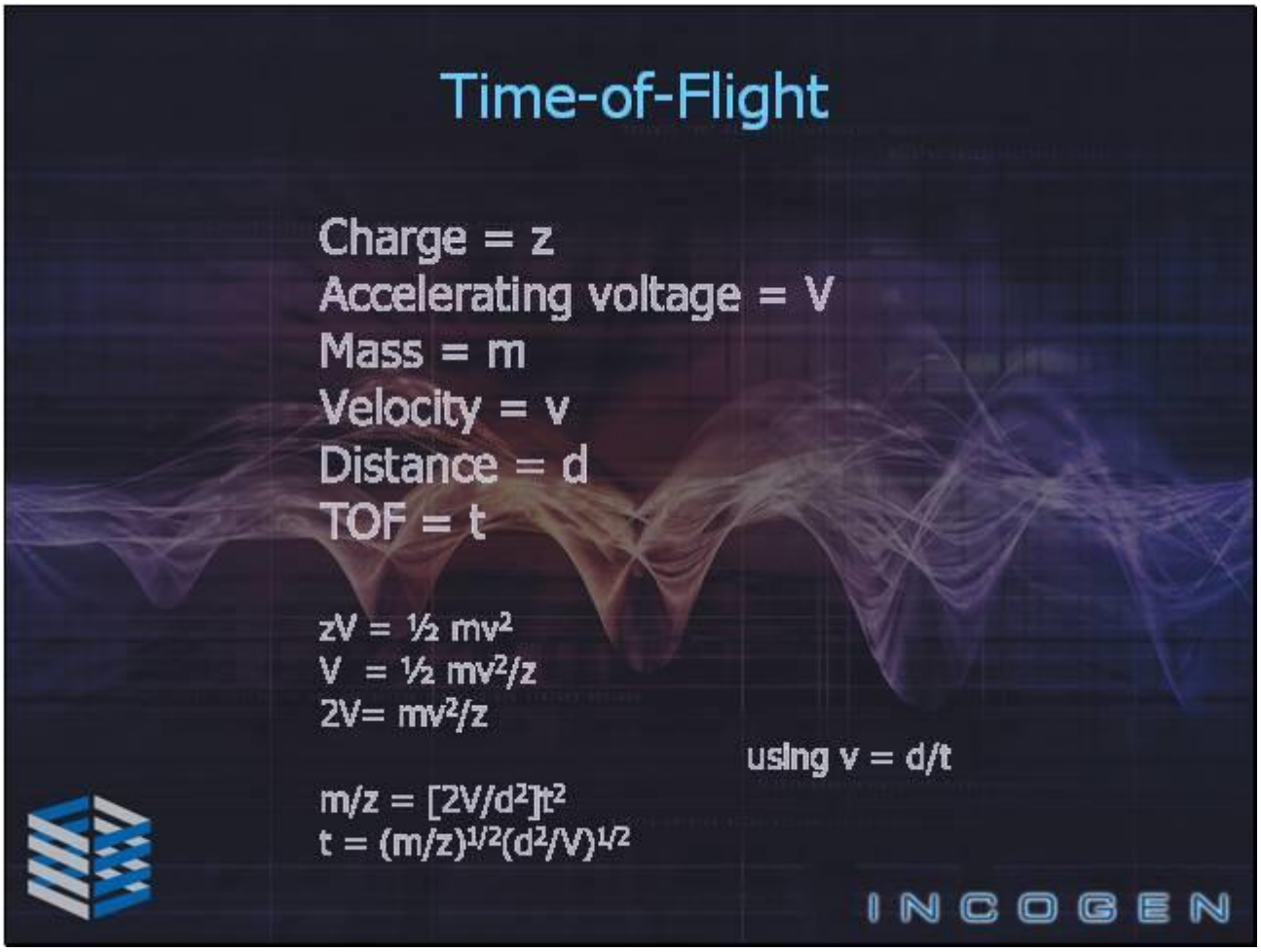
Slide 18 - Detectors



Slide notes

There are many types of detectors, but they all work by producing an electronic signal when struck by an ion. Timing mechanisms which integrate those signals with the scanning voltages allow the instrument to report which m/z strikes the detector. The mass analyzer sorts the ions according to m/z and the detector records the abundance of each m/z. Regular calibration of the m/z scale is necessary to maintain accuracy in the instrument. Calibration is performed by introducing a well-known compound into the instrument and "tweaking" the circuits so that the compound's molecular ion and fragment ions are reported accurately.

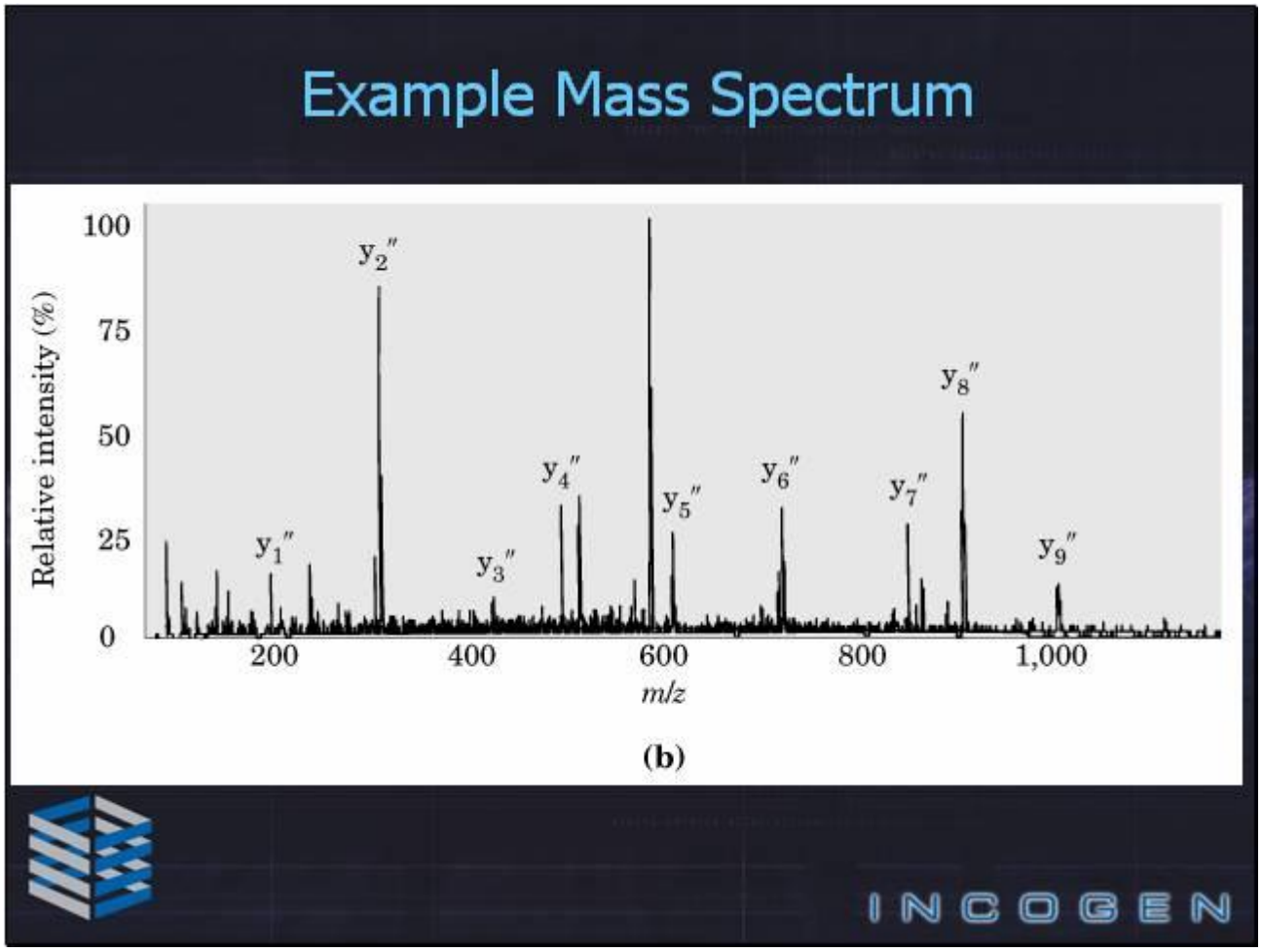
Slide 19 - Time-of-Flight



Slide notes

The Mass Spectrometer measures time-of-flight (TOF) for the ions, which is the amount of time required by the ions to travel from the source to the detector. From this, we can obtain the mass-to-charge ratios, or "m-over-z" (m/z) values of the ions.

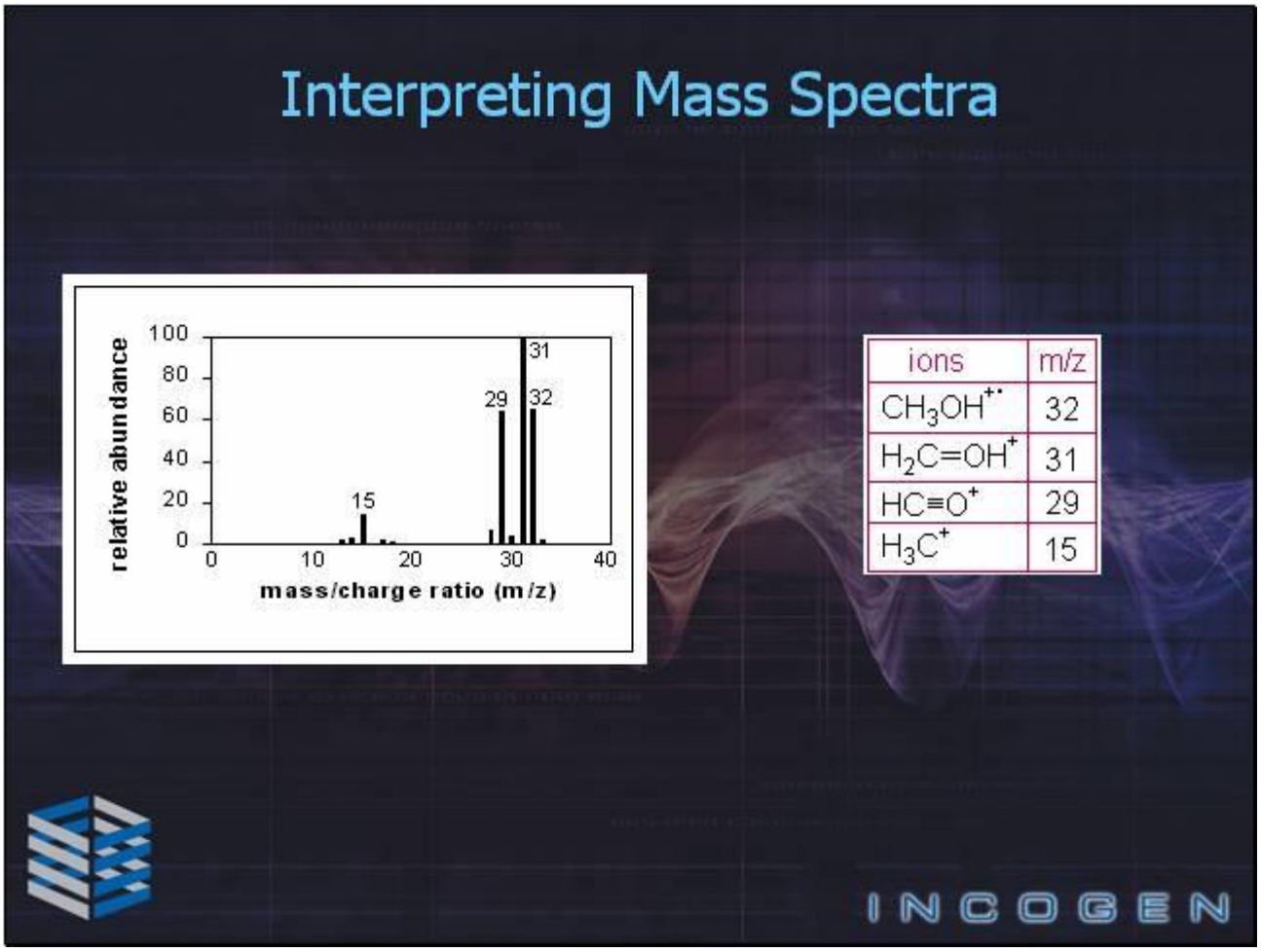
Slide 20 - Slide 20



Slide notes

Here is an example of a spectrum.

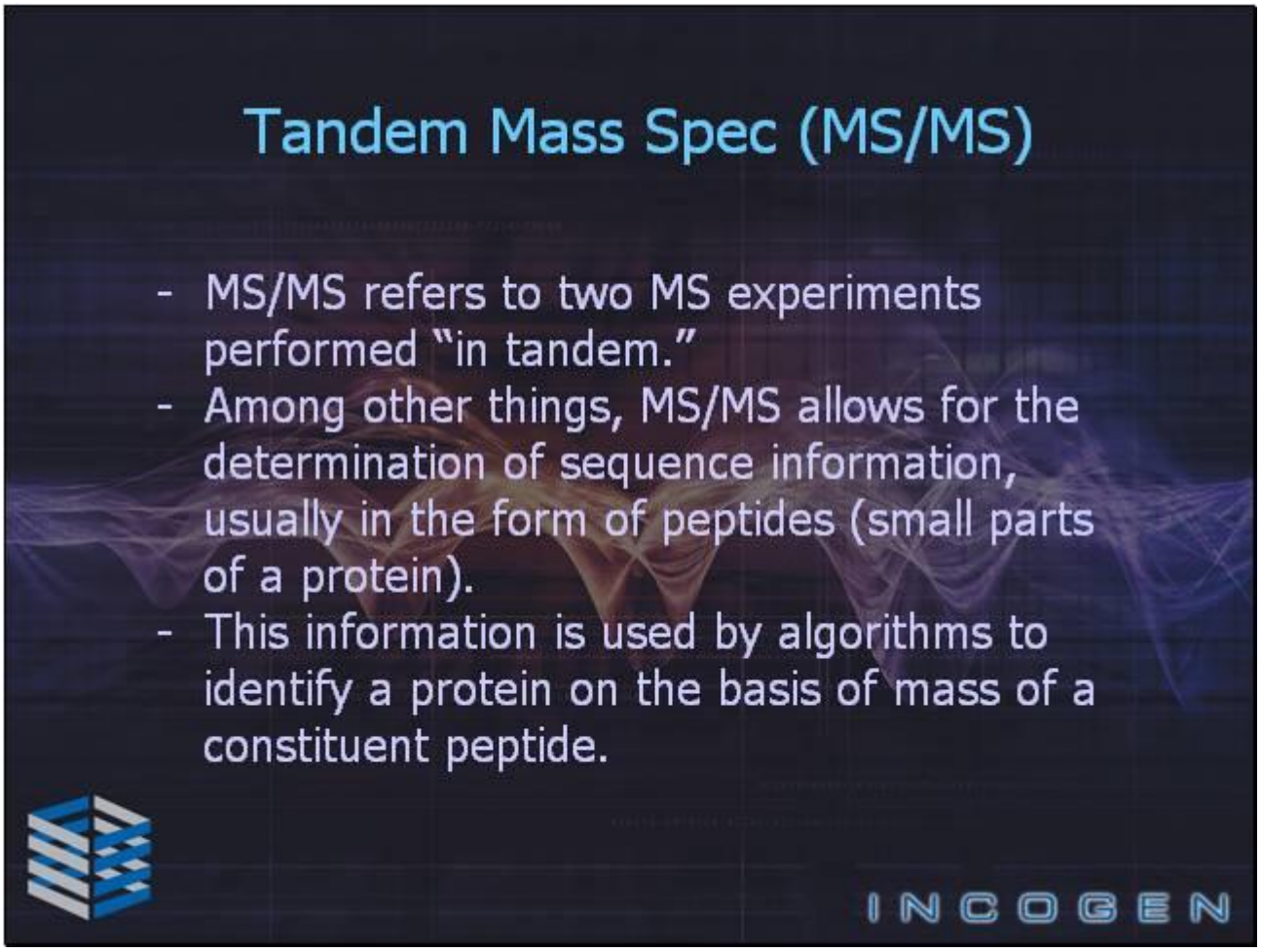
Slide 21 - Interpreting Mass Spectra



Slide notes

In general, proteins break up into fairly standard and common parts that have well defined m/z, so interpreting a mass spectrum is a matter of identifying which m/z values show a positive number of counts and then matching that m/z with an ion. It seems straightforward; however, the interpretation of results obtained from complex samples can be difficult.

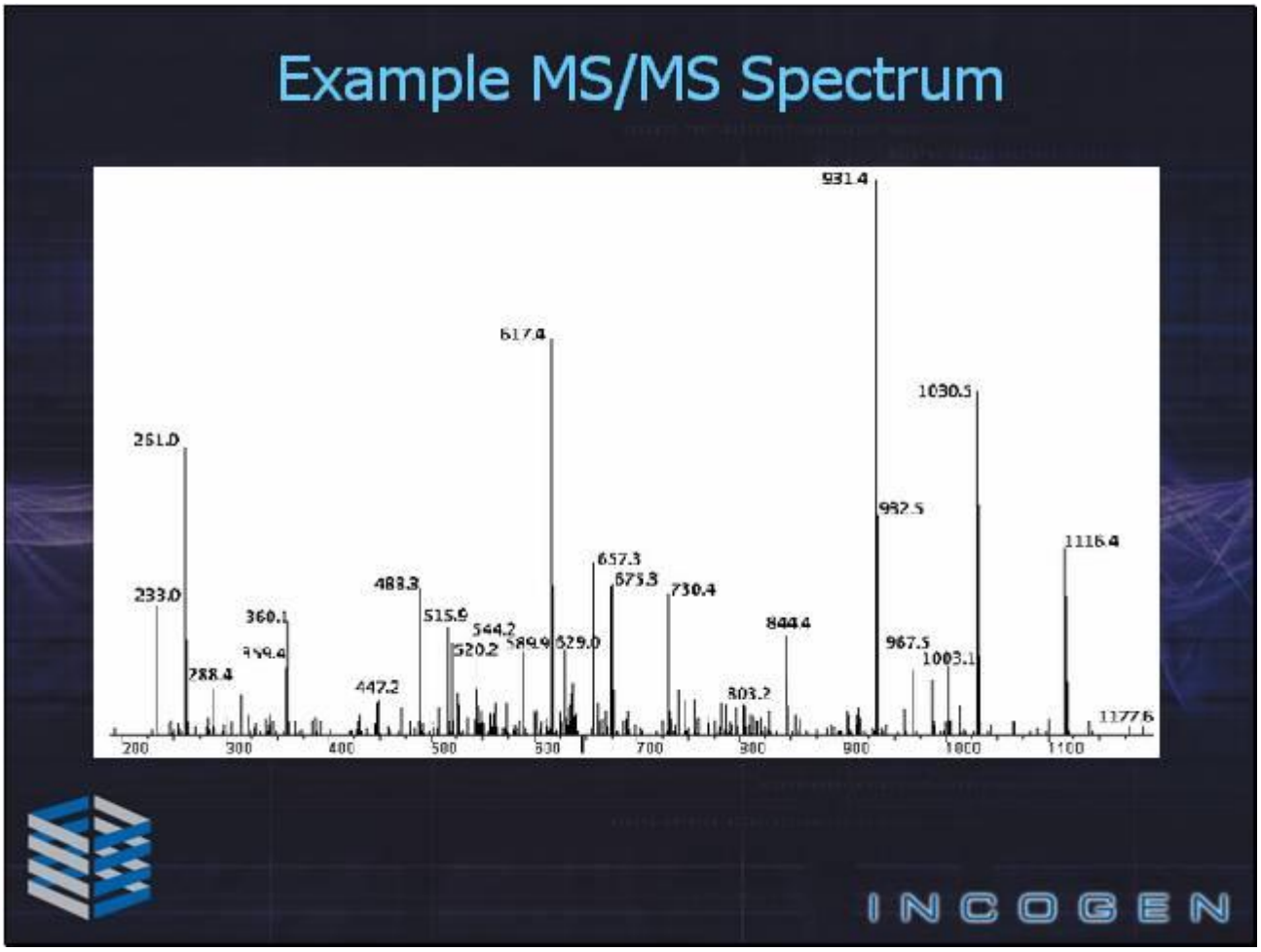
Slide 22 - Tandem Mass Spec (MS/MS)



Slide notes

It is also possible to set up a mass spectrometry experiment "in tandem", or one right after the other. Tandem mass spec experiments are used to identify proteins based on the masses of their constituent parts.

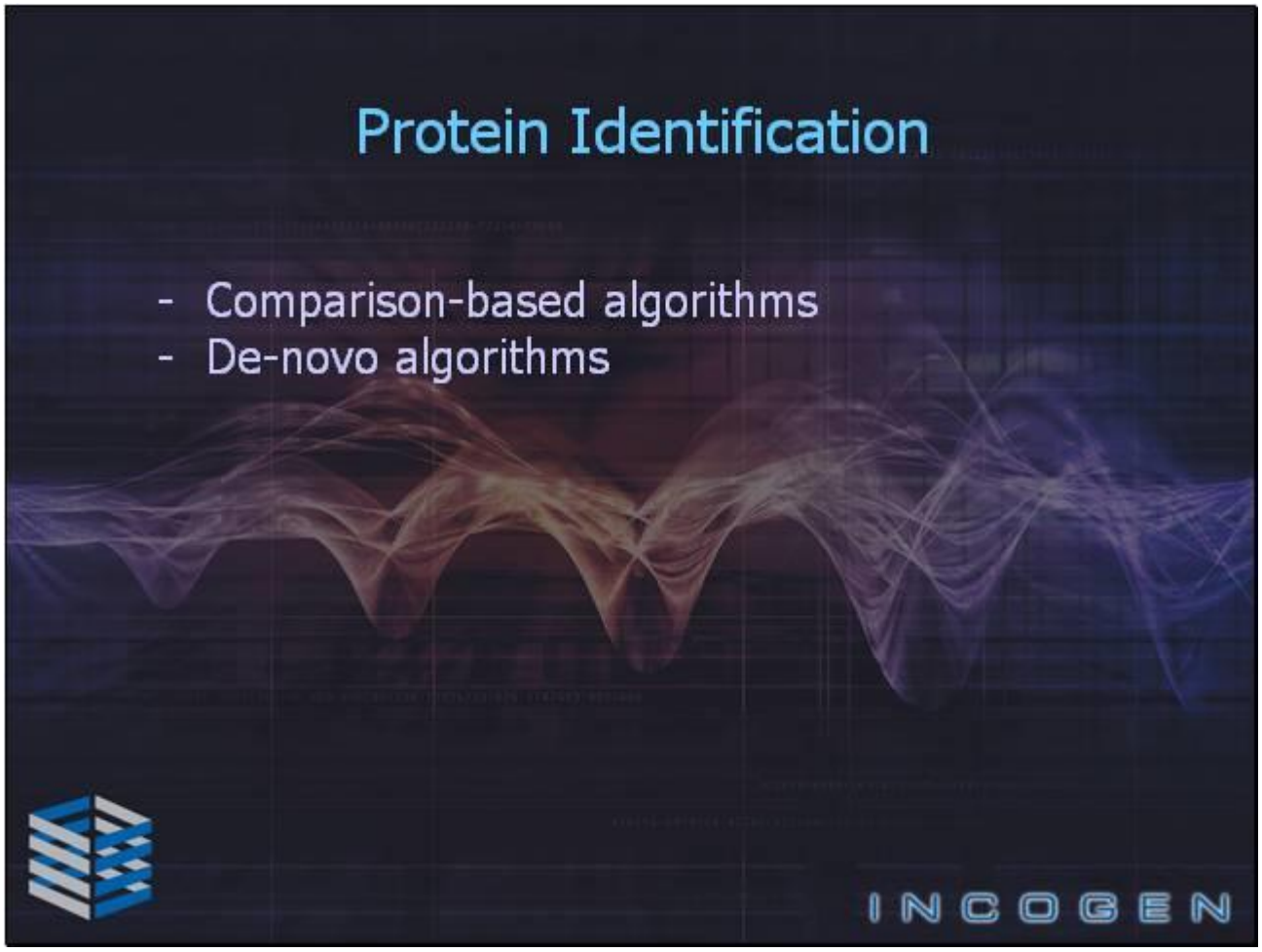
Slide 23 - Example MS/MS Spectrum



Slide notes

This is an example of a tandem mass spec experiment.

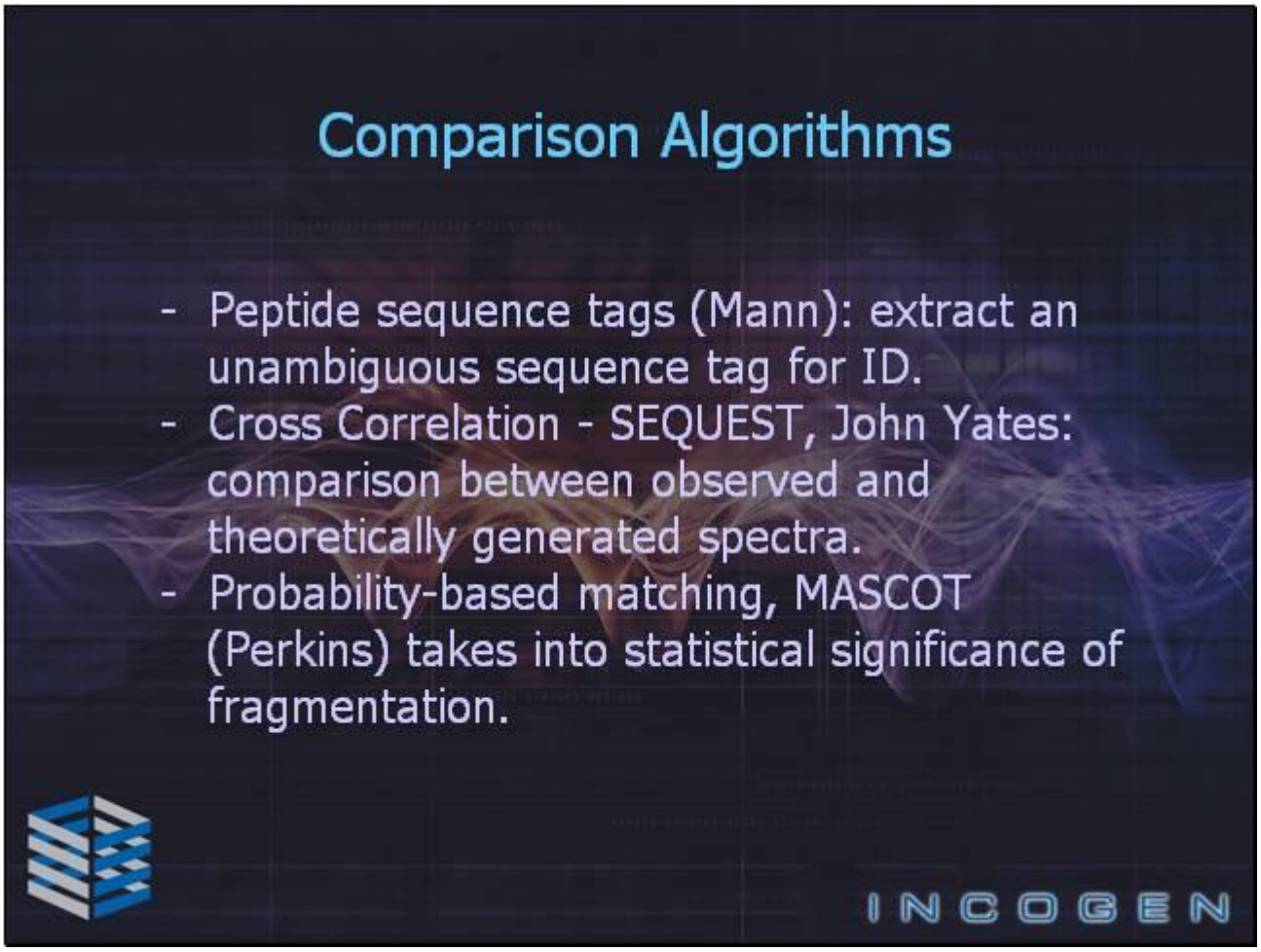
Slide 24 - Protein Identification



Slide notes

Protein identification from tandem mass spec experiments is accomplished using comparison-based algorithms or de novo algorithms.

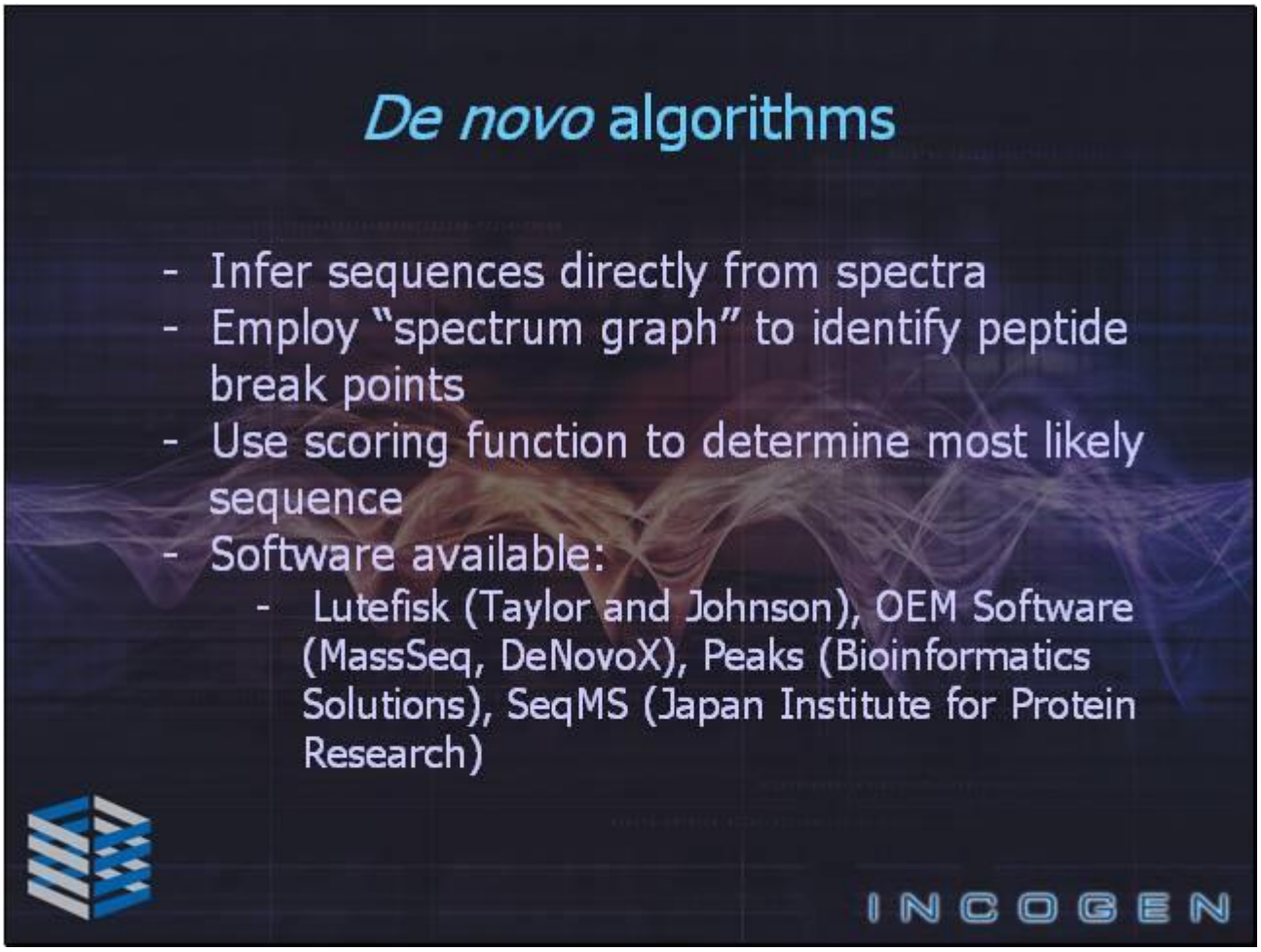
Slide 25 - Comparison Algorithms



Slide notes

A few comparison algorithms have been developed. Protein sequence tags, developed by Mann, extract an unambiguous sequence tag as identification. A sequence tag is a short stretch of sequence read out from a fragment spectrum of a peptide. Sequest, developed by John Yates, uses a cross correlation technique which converts the character-based representation of amino acid sequences in a protein database to fragmentation patterns which are compared against the MS/MS spectrum generated on the target peptide. Another commonly used tool, MASCOT, developed by Perkins, computes the probability that the observed match between the experimental data and mass values calculated from a candidate peptide or protein sequence is a random event.

Slide 26 - De novo algorithms



Slide notes

In the de novo algorithms, the sequence is inferred directly from the spectra. Then a "spectra graph" is employed to identify the peptide break points. A scoring function is then used to determine the most likely sequence. Several software programs have been written utilizing the de novo algorithms, but this remains an unsolved problem and an area of active research.

Slide 27 - Acknowledgements

Slide notes

Finally, we would like to acknowledge the following contributors.