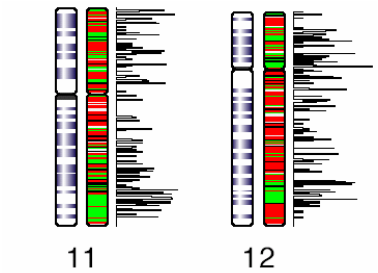
Mass Spectroscopy in Health and Environmental Science

S.E. Stein, P. Neta, and L. Kilpatrick (838)

The health of a cell and of an organism is reflected by the proteins that it contains. The recent delineation of the human genome coupled with radically new, sensitive methods for biomolecule detection by mass spectrometry has made it possible to measure a large fraction of these proteins, opening up a range of new, targeted methods for disease detection and prevention. To do this requires the ability to identify massive numbers of proteins, often at low concentrations and with poor separation so that any analysis experiment must deal with mixtures of hundreds or thousands of different proteins.

CSTL launches the first large-scale attempt to determine the variability of peptide spectra.



http://www.peptideatlas.org/

The use of mass spectra for protein identification is illustrated with recent work from the Peptide Atlas project. In this figure, the usual reproduction of the gene (in black and white) is coupled with a stick diagram representing the number of peptides identified by their mass spectra. The colored bands are an attempt to show whether there are more (green) or fewer (red) peptides than expected. The project has cataloged 225,000 spectra for more than 26,000 proteins. The number of spectra and proteins analyzed has grown dramatically in the last few years, but little attention has been paid to the basic reproducibility of peptide mass spectra and to the development of effective means of establishing their accuracy. There has not been any significant measure of the reproducibility of the spectra from the same individual, the reproducibility of the spectra from different individuals, and the variations that are mapped to disease states. This information is required in order to establish reliable collections of peptide spectra for use in research and clinical studies. The mass spectra for these proteins are being generated by liquid chromatography (LC), mass spectrometry (MS), and specifically by LC/MS/MS. In these experiments the proteins are chemically broken apart by specific enzymes in well-understood and predictable ways. The resulting mixtures are separated by liquid chromatography, the ions are produced by an electrospray apparatus, and then the molecular ions are fragmented to give spectra that are characteristic of the specific peptide that has been eluted. The small components, peptides, can be identified and their identity tied back to the genome. In this way, peptide spectra become fingerprints of the protein and can even establish its particular chemical state (such as post translational modifications). This mechanism has been refined by researchers around the world and the ease with which the experiments can be done is a major part of the explosion in information on the proteome.

The first large-scale attempt to determine the variability of these peptide spectra is being done at NIST. In this project we are using thousands of openly available spectra along with some specific measurements made at NIST. By looking at the variability created by changing the apparatus and measurement conditions in well-understood ways, we are beginning to understand the best way to establish a measure of confidence in a peptide spectrum and to define new algorithms to compensate for this variability in identification of the proteins by electrospray LC/MS/MS. An example of the data can be seen in the figure. Here we use a very slightly modified version of the NIST MS search software to display and search a given protein mass spectrum produced from instruments at NIST (both in the NIST Gaithersburg labs and at the NIST facilities at the Hollings Marine Laboratory in Charleston, SC) with the data for the same protein from other laboratories.

Plans are being developed to build high-quality libraries of commonly observed peptides from both directed measurements of selected proteins at NIST and from the vast and growing information available in repositories being developed in the health science community. This will enable a significant increase in reliability of protein identification as well as enable the development of refined methods for using sequence information for matching experimentally determined peptide spectra. While the scale of this work is very large, the tools developed for use with EI mass spectrometry here at NIST along with tools for the specific problems raised with mass spectra of proteins in other institutions make the effort possible.