A primer on methods in proteomics and metabolomics

Michael Schubert

19th December 2010

1 Introduction

This document aims to provide a summary of methods in proteomics as well as metabolomics as demanded by section A of the „Proteome and Metabolome Informatics“ module of the Applied Bioinformatics MSc at Cranfield University. It was generally tried to answer the questions in order, but this concept was sacrificed for the sake of clarity where needed.

2 Proteomics

This section summarises the different methods empleyed oin proteomics, although the general principles apply to similar fields, such as metabolomics, as well. The experimental methods described here in detail are to be understood as valid for the metabolomics section as well, which is why it is a little more thorough than the second part.

2.1 Separation of compounds

2.1.1 Extraction

Classical (liquid-liquid) extraction was one of the first methods to try and separate different components in a solution. It is based on components being differently soluble in two immiscible fluids where each component will be enriched in the fluid it is more following Nernst's Law of distribution

$$K = \frac{c_i(phase\ A)}{c_i(phase\ B)}$$

which states that the relative concentrations will be constant provided the two solutions are in equilibrium with each other. Compounds can also be extracted from a solid matrix, giving rise to the technique of solid phase extraction.

2.1.2 Chromatography

The principle of chromatography is having a stationary phase (a matrix) where a mobile phase including the mixture of compounds to be separated migrates through. Each compound can either interact with the mobile or the stationary phase, the level of which is resulting in migration speed. The simplest

method of chromatography is performed with a thin layer as stationary phase (TLC), which in most cases consists of silica gel (polar). The mobile phase in that case is some or a mixture of organic solvents (less polar), that migrates through the stationary phase due to capillary action. The compound is held back each time it interacts with the stationary phase, resulting in less polar compounds travelling further on the layer.

Other methods of chromatography include:

- Size exclusion chromatography (SEC): In this method, the macromolecular compounds do not interact with the stationary phase surface which is comprised of porous particles. Instead, depending on the compound's volume, it can either migrate between particles or in the pores within individual ones. The smaller it is and thus the more pores it migrates through, the longer it takes for it to elute.
- Ion exchange chromatography (IEX): The stationary phase consists of specific ions, which can bind compounds of the opposite charge. Usually, this method is employed so that the interaction is strong enough not only to delay migration, but to completely prevent it under given conditions (defined as „loading" of a column). The bound compounds are subsequently eluted under different conditions (e.g., a change a change of pH value or salt concentration).
- Affinity chromatography: The interaction partner of a compound of interest is specifically binding to it. This can either be achieved by using: a column with affixed antibodies to it, or affinity tags such as a nickle column that captures Histidin hexamers (HIS-Tag).
- Gas chromatography (GC): The compound of interest must be transferred into the mobile, non-reacitve gasious phase without being degraded. The gas is then transferred through a very long but narrow column, being separated by the usual principle. Since this process is very dependent on temperature, it must be kept constant for its duration.
- High performance liquid chromatography (HPLC): The main difference between HPLC and other chromatography methods is that the mobile phase is pumped through the column at a high pressure instead of just gravity-induced flow. In most cases, the column is more polar than the mobile phase. The reverse (termed reverse phase chromatography) can also be used, especially for protein compounds. Also, the column is loaded with the sample and then eluted by application of a salt- or pH-, or temperature gradient.

2.1.3 Electrophoresis

The basis of electrophoresis is that compounds move in an electric field according to their charge. Usually, they are are embedded in a matrix that slows their migration so that an effective separation can be achieved. This matrix is in most cases a polymer made of either Agarose or Acrylamide (poly-Acrylamide gel electrophoresis, PAGE) where the concentration of Agarose and the ratio of Acrylamide to Bisacrylamide are respectively responsible for the degree of cross-linking and thus the amount of migration slowing. However, this only allows for separation by compound charge and not by size. This is no issue in the case where charge corresponds to compound size (e.g., a DNA molecule), but not for proteins. There, size-based separation can be accomplished by coating the protein with a compound as sodium dodecyl sulfate (SDS), resulting in a charge that corresponds to size.

Additionally, it is possible to apply gradients to the matrix (gel), either by pore size or pH value. In the first case, the resistance to compound movement is increased with distance (i.e., the pore size is decreased), which leads to sharper separation. In the latter case, the pH gradient in the gel causes protein charges to vary according to their isoelectric point, termed isoelectric focussing (IEF). Once they reach a position in the gel where the pH is equal to their isoelectric point, they exhibit no net charge and thus no longer move in the electric field.

It is also possible to apply two different gradients one after another, resulting in separation by two different compound properties. For instance, a step of IEF could be followed by SDS-PAGE. This is particularly relevant in complex samples where more than one compounds share a specific size or a specific isoelectric point.

2.2 Mass spectrometry methods

One of the most widely employed methods for compound identification in proteomics, as well as many other fields, is mass spectrometry. Therein, samples are ionised and the resulting ions are separated on basis of their mass to charge ratio (m/z), which is in turn detected. This procedure is very sensitive and does only require a minimal amount of the sample. For protein samples, some kind of digestion (mostly Trypsin) usually precedes the actual analysis since peptides are easier to analyse compared to intact proteins and a more uniform fragmentation is achieved compared to using ionisation fragmentation directly.

2.2.1 Ionisation

In principle, there are two kinds of ionisation methods: the so-called hard and soft ones. The difference between these is if and how much a compound gets fragmented when it is ionised. Regarding protein identification, one does want only fragmentation along the peptide backbone and have intact residues. Thus, generally only soft methods are used in proteomics:

- Electrospray ionisation (ESI): An aerosol is created by spraying the sample through a thin needle, which, when the solvent evaporates, is divided into even smaller droplets because of the electrostatic repulsion between sample molecules.
- Matrix-assisted laser desorption ionisation (MALDI): The sample is dried in a matrix of carrier molecules, e.g. Sinipinic acid or α-CHCA, which both are desorbed by a short laser pulse, thereby forming ions of the sample. This method is not used as often because it is not easy to combine it with tandem methods or LC.

The corresponding N- or C-terminal ions are called a/b/c and x/y/z respectively, in accordance to where the peptide bond is broken (cf. figure 1).

2.2.2 Mass analyser

The idea of a mass analyser is to separate ions bei their mass-to-charge ratio so a specific ion can be detected. The most widely used ones are:

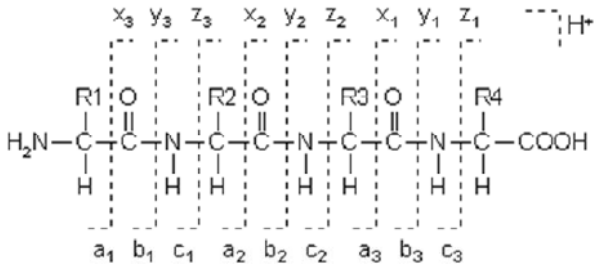

Figure 1: Naming convention of different fragment ions

- Time of Flight (TOF): Ions are accelerated by the short application of an electromagnetic field where ions with a higher m/z suffer more from inertia than the light ones. The time which each ion needs to reach the detector is measured. This technique allows for rapid recording of spectra but has limited resolution.
- Quadrupole: This mass analyser consists of two times two rods with an applied DC and RF, respectively. At any given time, only ions with a specific m/z can travel along the rods until the detector, the others are filtered out.
- Other methods include an Ion Trap or Ion Cyclotron Resonance.

2.2.3 Detection and classification

For detection, usually an electron multiplier is used. When the sample is relatively pure or has been subjected to a suitable chromatography method before, the peak chromatogram can be searched against a library of known compounds. This approach is very efficient when (a) the sample is not too complex and (b) the compound is known.

2.2.4 Tandem MS

To overcome some limitations of classic mass spectrometers, they can also be used one after another (tandem-MS). For instance, three Quadrupoles can be set up serially. The mode of operation of each one can either be a full scan (all ions are let through, each m/z at a specific time), a targeted scan (only the ions with a specific m/z are let through), or it can also act as a collision cell (Helium is pumped into the otherwise evacuated chamber and ions are thereby broken down to smaller pieces). One obvious application of these possibilities would be to use the first Quadrupole to always only let through a specific m/z at any given time, use a second one to break the precurser-ion down further, and analyse the fragment ions with a third Quadrupole. An other possibility would be to couple a Quadrupole with a TOF analyser. The result of both is in essence a 2D-MS spectrum.

2.3 Quantification

When talking about quantification, it should be noted that there are two different types, absolute and relative. While absolute quantification tries to determine how much e.g. picomoles of a certain compound are present in the sample, relative quantification is satisfied with information on how much one compound is more abundant than the other (e.g. 2 times).

- Stable Isotope Labelling (realtive approach): When comparing the relative abundances of a certain compound within two different samples, one could label one with a stable isotope such as ^{15}N, mix them, and then compare the peak size with the known mass offset due to the labelling. This method is expensive and difficult to perform, since it is hard to get a 100% pure isotope. Its advantages on the other hand are that pooling of samples is possible and there thus is no machine variance between runs, and the compuational analysis is simple. One application of this technique is Stable Isotope Labelling with Amino Acids in Cell culture (SILAC), which aims to label all of a certain amino acid. This approach faces the before mentioned problems, as well as amino acid conversions and the need for auxotrophic strains. One attempt to alleviate these issues is the development if iTRAQ, which appends isochore tags to compounds of interest that can be distinguished by MSMS.

- Calibration with a known amount of internal standard (absolute approach): In this approach, some kind of labelled compound that is similar to the one of interest is added to the sample. The question thereby is how similar the internal standard needs to be to the compound of interest, since variability can be introduced by different tryptic digestion, ionisation, or other efficiencies. Based on different assumptions, methods as AQUA or QconCAT were developed. An alternative approach is to add a completely unrelated protein, which performs surprisingly well.

3 Metabolomics

In contrast to proteomics, metabolomics takes care of small molecule analysis. The methods employed are generally the same as in proteomics, but there are suble differences.

3.1 Methods employed

The three main methods that are employed in metabolomics are HPLC-MS, GC-MS, and NMR. The details of HPLC, GC, and MS were already extensively discussed in section 2 including their advantages and disadvantages, which is why they are not featured in full detail here. The reader is referred to the corresponding sections when not specifically discussing the combinations of methods. All limitations of the single methods are of course valid for the combined approaches.

- HPLC-MS: In HPLC-MS, a complex sample is first unraveld by means of HPLC and then subjected to MS analysis. The HPLC step thereby produces a number of peaks, which are analysed independently by MS. For details on the basic methods, see the description in the proteomics section. An advantage of HPLC over GC is that a sample can be analysed in its native state (as solution) and does not need to be evaporated. Furthermore, polar as well as organic molecules can be analysed using normal and reverse phase HPLC, respectively. Disadvantages include the potential degradation of modified silica residues.

- GC-MS: In GC-MS, the separation step is carried out by gas instead of liquid chromatography. Its advantages are a high resolution and sensitivity, as well as high speed and general accuracy. Disadvantages include sample preparation, e.g. for evaporation or derivatisation in order to cope with reactive or otherwise unstable compounds.

- NMR: This method is based on a property of certain atomic nuclei, that if their spin is not equal to zero, absorb energy in a magnetic field and radiate this energy back out. The frequency with which this occurs is specific to the nucleus and its surrounding atoms, and is termed resonance frequency. Exciting all the nuclei and recording their resonance frequencies leads to spectra similar to the other methods. The most important nuclei for this kind of analysis are ^1H and ^{13}C. An advantage of this method, in contrast to MS methods, that it does not destroy the sample. Disadvantages compared to MS methods are, for instance, lower sensitivity and a higher price.

3.2 Peak resolution

When a peak in GC corresponds to two distinct metabolites, there are various methods to distinguish between those. One choice would be to choose a different separation technique in which the metabolites that previously eluted at the same time now elute at different times. In order to achieve this, their interaction with the column needs to be at least slightly different, so one could use a different column for which this is true or try another chromatography method, such as HPLC. Also, an additional pre-separation step could be taken, with e.g. other chromatography or electrophoresis methods.

If these intrumental techniques could not be used or were unsuccessful for whatever reason, the peaks can also be resolved when analysing the MS spectrum. In general, there are two possibilites: one is to separate the two compounds by using tandem-MS (since even if two compounds elute at the same time, they are unlikely to have the same m/z ratio) and analyse each one afterwards, which should generally be the preferred approach. The other method is to identify the compounds within one MS spectrum by their individual peaks. This works well when they have few fragment sizes in common, which can be but does not have to be the case. If if is, the ratio of each compound in the GC-peak can be calculated by using the MS peaks as a basis. This can be done by, for instance, library seraches of MS peak data.

References

All information was compiled from the lecture slides of the Proteomics and Metabolomics module at Cranfield University, as well as notes taken during these sessions.