**1 Peptide Identification**

The currency in proteomics are the peptides. In proteomics, proteins are digested to so-called peptides since they are much easier to handle biochemically than proteins. Proteins are very different some are very sticky while others are soluble in aqueous solutions while again are only soluble in membranes. Therefore, proteins are chopped up into peptides because it is far to assume, that for each protein, there will be some peptides behaving well so that they can be measured with the mass spectrometer. This step introduces another problem, the so-called protein inference problem. In this package, we do not look at all touch upon the protein inference.

### 1.1 Computing the Parent Ion Mass

The function `parentIonMass` relates the mass of an amino acid sequence while the function `nchar` returns a hydrophobicity value for a given amino acid which can be used to predict the retention time.

```r
# library(protviz)
# + pch = 16,
# # fragment Ion('HTLNQIDSVK')[[1]]
# # parentIonMass('HTLNQIDSVK')
# defaultIon
# + protViz::iRTpeptides$peptide)
# list of 2
# - title : chr “178: (rt=22.3807) [20080816_23_fetuin_160.RAW]
# - samples : chr “178: (rt=22.3807) [20080816_23_fetuin_160.RAW]

# parentIonMass(irt.peptide)
# defaultIon
# + protViz::iRTpeptides$peptide)
# list of 2
# - title : chr “178: (rt=22.3807) [20080816_23_fetuin_160.RAW]
# - samples : chr “178: (rt=22.3807) [20080816_23_fetuin_160.RAW]
```

### 1.2 In-silico Peptide Fragmentation

Given a peptide sequence and a tandem mass spectrum for the assignment of a candidate peptide an in-silico fragment ions is computed. The function `defn` does perform this for each fragment in the closed peak in the MS2. If the difference between the in-silico mass and the measured mass is inside the ‘accuracy’ window of the mass spec device the in-silico fragment ion is considered a potential hit.

```r
# peptide.sequence <- “HTLNQIDSVK”
# + defn(“HTLNQIDSVK”),
# + defn(“HTLNQIDSVK”)
```

### 1.3 Fragment Ion Matching

In almost all cases there are not more than three to six repetitions. For the moment there are limited options defining other forms of fragment ions for ETD (c and z ions).

```r
c <- b + (Nitrogen + (3* Hydrogen))
# Nitrogen <- 14.003074
# Hydrogen <- 1.007825
```

### 2 Quantification

For an overview of Quantitative Proteomics read e.g. [12] The authors are aware that meaningful statistics usually require a much higher number of biological replicates. In almost all cases there are not more than three to six repetitions. For the moment there are limited options due to the availability of mass spectrometers and the limit of the technologies.

#### 2.1 Absolute Label-Free

The data set `retinol2` contains a subset of our real data labeled in [9]. The example below shows a visualization using box plots. It graphs the abundance of four protein dependency from the fedcon concentration spiked into the sample.

```r
# par(mfrow=c(1,2)
# + qqline(asinh(iTRAQ[,i]), col='grey')
# + asinh(iTRAQ[,i])
# + asinh(iTRAQ[,i])
# + asinh(iTRAQ[,i])
```

#### 2.2 Relative Label-Free

LCMS based label-free quantification is a very popular method to extract relative quantitative information from mass spectrometry experiments. At the AGC we use the software ProgressCMS for this work flow.

```r
# par(mfrow=c(1,2)
# + protViz::iRTpeptides$peptide)
# + protViz::iRTpeptides$peptide)
# + protViz::iRTpeptides$peptide)
# + protViz::iRTpeptides$peptide)
```

This right figure above shows the correlation between runs on protein level (values are asinh transformed). While the perfect correlation while black indicates a poor correlation. Sticking is the fact that the six biological replicates for each condition group very well.

### 2.3 ITRAQ – Two Group Analysis

The data for the next section is an ITRAQ-8-plex experiment where two conditions are compared (each condition has four biological replicates)

#### 2.3.1 Sanity Check

```r
# par(mfrow=c(1,2)
# + main = "boxplot(artifact, c(0:3))",
# + main = "boxplot(artifact, c(0:3))"
```

This figure shows the five proteins which are tested using the t-test function if they differ across conditions using the four biological replicates.

### References

