



Tutorial 4: Targeted Proteomics

1 Introduction to targeted proteomics

Targeted proteomics, where a HPLC-MS/MS instrument is pre-programmed to monitor specific peptides as surrogates for proteins of interest remains popular due to its perceived high sensitivity compared to data dependent acquisition (DDA) and the fact that it can be conducted using relatively inexpensive mass spectrometers with low mass accuracy.

A targeted proteomics experiment may involve selected reaction monitoring (SRM) whereby the mass spectrometer is programmed to monitor masses associated with specific parent ions and specific product ions derived from these, or utilise a simple inclusion list of peptide ions to monitor. In both cases, a target mass may be monitored over a full chromatography run, or scheduled within a defined retention time (RT) window during which the target peptide is expected to elute.

Selection of surrogate peptides for targeted proteomics experiments is a laborious process, as many factors define what makes a good surrogate: mass spectrometry detectability of the peptide, likely quality of its fragmentation spectrum, matrix effects, and cleavage efficiency. GIO contains a number of tools designed to predict these factors, thereby assisting in the selection of peptides for targeted proteomics.

2 Targeted proteomics tools in GIO

The following tools are provided in the *GIO:Targeted proteomics* section of tools:

Digest: Taking as input a FASTA file of protein sequences, this tool digests the proteins into their tryptic peptides.

Peptide MW and hydrophobicity: Calculates the molecular weight of peptides from sequences provided in FASTA format or in the rows of a tabular file. Also provides an estimate of the peptide hydrophobicity using the SSRCalc algorithm¹.

MC:Pred²: Taking as input a FASTA file of protein sequences, digest proteins into their tryptic peptides and predicts the likelihood of missed cleavage occurring for both termini.

CONSeQuence2³: Predicts peptide detectability from sequence.

PREGO⁴: Predicts suitability of peptide for SRM from sequence.

Intercalc: Calculates the likelihood of matrix effects for a given peptide due to the presence of other peptides with similar m/z and retention time.

With the exception of *Digest*, these tools have generally been implemented such that they take as input a tabular file – in which peptides are arranged in rows – and append one or more columns of new information, e.g. possible level of interference (Intercalc) or detectability score (CONSeQuence2). They can therefore be chained together into workflows that gradually build a picture of each peptide's suitability as a surrogate.

¹ <http://www.ncbi.nlm.nih.gov/pubmed/15238601>

² <http://www.ncbi.nlm.nih.gov/pubmed/22804685>

³ <http://www.ncbi.nlm.nih.gov/pubmed/21813416>

⁴ <http://www.ncbi.nlm.nih.gov/pubmed/26100116>

3 Sample workflow to inform surrogate peptide selection

A sample workflow is provided in GIO under the name *Surrogate peptide selection*, and is reproduced here in Figure 1. This workflow begins with two FASTA files, one containing sequences of target proteins and the other containing background proteins. The background proteins comprise proteins that could feasibly be expected to be present within the sample – typically the proteome of the species under study plus common contaminants.

In the first step of the workflow, both sets of proteins are digested into tryptic peptides using the *Digest* tool. The molecular weight and hydrophobicity of each peptide from both the targets and background is then calculated. This information is used by the Intercalc tool to provide an indication of the potential background interference for each peptide. CONSeQuence2 is then applied (twice in this case, once using random forests and once with adaptive boosting) to generate detectability scores for each peptide. PREGO is then used to calculate peptide suitability scores, and the accumulated information augmented with missed cleavage scores from MC:pred using Galaxy's built-in *Join* tool.

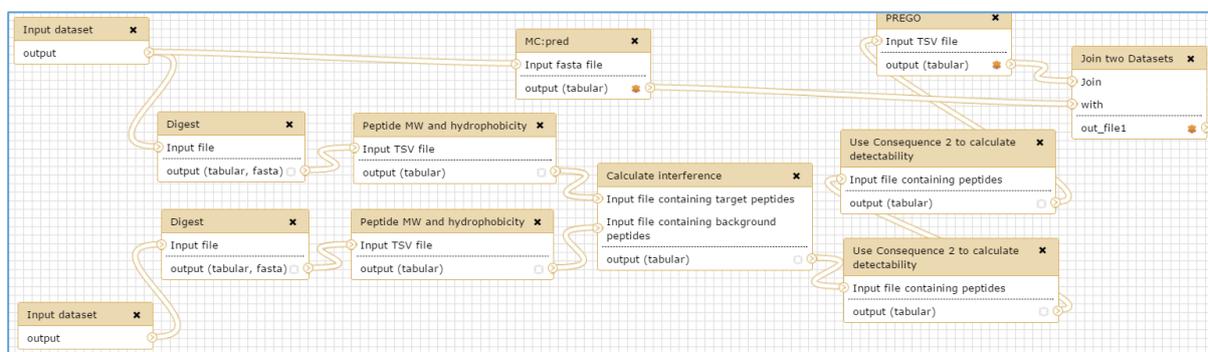


Figure 1: Workflow to aid surrogate peptide selection.

To summarise, the workflow takes as input a list of proteins that may be in the sample, a list of target proteins that we want to monitor, and ultimately produces a table in which each row contains a peptide annotated with various numerical indicators of its suitability as a surrogate peptide for targeted proteomics.

4 Further workflow steps

The workflow described above results in a tabular file containing all potential target peptide sequences, together with the parent protein name, peptide sequence, mass, hydrophobicity, potential background interference and multiple estimates of detectability. While this contains useful information for selecting surrogate peptides, it cannot in itself be used to conduct a targeted proteomics experiment – further work is needed to extract a list of peptides to monitor from the available information. The most obvious way to achieve this is by combining the various scores into a single value for ranking peptides, using Galaxy's built-in *Compute* tool (found in the *Text manipulation* section of tools), and using *Filter and sort* tools to produce a shortlist of peptides to monitor. Once this list has been produced the *Compute* tool can be used to determine approximate RT from the calculated hydrophobicity for the specific type of chromatography to be employed, using simple linear calibration, thereby providing a scheduled list of masses to monitor.

5 Summary and next steps

This tutorial has introduced the targeted proteomics tools that are currently available within GIO, and suggested how they can be combined into workflows for designing targeted proteomics experiments. We do not endorse a particular tool or workflow because of the variety of targeted

proteomics protocols that exist and the diverse range of biological questions to which they are applied. Rather, GIO provides immense flexibility in which tools can be combined, parameters set and bespoke workflows developed for selecting surrogate peptides for specific studies. Once created, these workflows can be reused and shared using standard Galaxy functionality, helping to ensure repeatable research.

Last updated 16 November 2015.