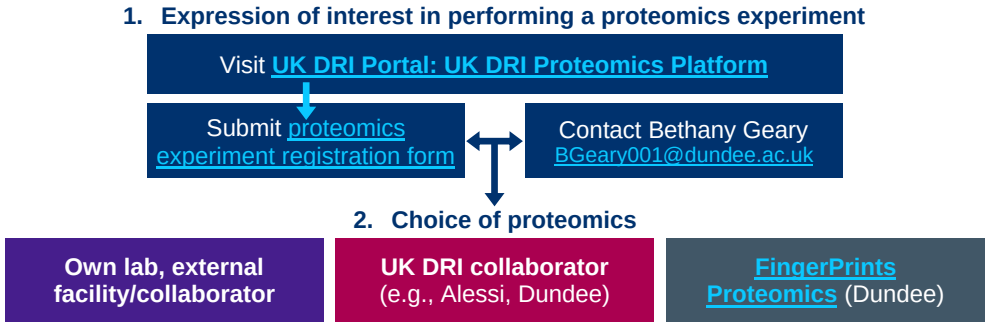
UK DRI Experimental and Reporting Guidelines for Proteomics

1 Overview

The use of proteomics mass spectrometry (MS) in neuroscience has bloomed in the past decade. The UK Dementia Research Institute (UK DRI) is committed to generating high-quality proteomics datasets that are findable, accessible, interoperable and reusable. To foster collaboration and strengthen experiments' reproducibility and replicability, the UK DRI has set up the following proteomics workflow:



The UK DRI Proteomics Data Working Group[1], with input from Dougie Lamont (Dundee), Prof Dario Alessi (Dundee) and Dr Harry Whitwell (Imperial), have also set out guidelines[2] to ensure that:

- datasets generated by UK DRI members meet community standards;
- the inherent variability of proteomics datasets is minimised;
- datasets contain the necessary information for others to replicate or reanalyse them.

The guidelines focus on five main stages of a proteomics experiment:

- **Experimental design**: Submitting experimental information (metadata) (e.g., origin/processing of biological material, numbers of replicates and experiment type) through the proteomics experiment registration form as soon as an experiment is planned. Early submission will greatly facilitate collaborations between UK DRI researchers and help avoid duplication of effort. [2.1]
- **Sample preparation**: Documenting the standard operating procedures (SOPs) used for sample processing and noting any deviations. [2.2]
- **Mass spectrometry**: Reporting the SOPs used (incl. software name and version, acquisition parameters) for liquid chromatography (LC) and MS; noting any deviations. [2.3] Please acquaint yourself with which files are mandatory/recommended/optional for a complete submission to a public data repository prior to the start of the experiment (see also [2.5]).
- **Data analysis**: Documenting the software and analysis pipeline (incl. version number, parameters and, if applicable, link to the code repository), statistical design and database(s) used. [2.4]
- **Data sharing**: Compiling and submitting all necessary data and metadata to a public repository. We recommend the EMBL-EBI PRoteomics IDEntifications (PRIDE) database. [2.5]

2 Detailed Guidelines

We strongly recommend performing and reporting the following steps and checks. As experimental design varies from study to study, if any step is not followed, please report the alternative step(s) taken to maintain the required standards. For further reading, please see the HUPO-PSI Minimum Information About a Proteomics Experiment (MIAPE) guidelines. Please contact Dr Bethany Geary (BGeary001@dundee.ac.uk) for any queries or concerns.

2.1 Experimental Design

Please submit information relevant to the proteomics experiment through the proteomics experiment registration form as soon as it is planned. Here are factors to consider while designing an experiment:

---

[1] Members include Dr Blanca Diaz-Castro (Lead, UK DRI at Edinburgh), Dr Johanna Jackson (UK DRI at Imperial), Sam Jackson (UK DRI HQ) and Dr Bethany Geary (Dundee).

[2] In alignment with guidelines from the HUPO Proteomics Standards Initiative (HUPO-PSI), the ProteomeXchange Consortium and journals (incl. Proteomics, Journal of Proteome Research, Molecular & Cellular Proteomics).

Founding funders: UKRI Medical Research Council, Alzheimer's Society, Alzheimer's Research UK

- One may need to perform pilot experiments to optimise the protocols before using key samples.
- One may need to consult a statistician for proper experimental design for projects with large cohorts.
- Randomisation of sample collection and data acquisition.
- For data-independent acquisition (DIA) analysis, consider whether a sample-specific library should be generated.
- Researchers must adhere to all relevant regulatory requirements, including those relating to information governance, ethical use of data and/or the Human Tissue Act 2004.
- An umbrella Research Collaboration Agreement is in place between the UK DRI centres and the University of Dundee to protect intellectual property and ensure that UK DRI polices are adhered to. In the event of carrying out work with Dundee, a standard Material Transfer Agreement (MTA) will be provided at the start of the collaboration, managed by Dr Bethany Geary.

2.2 Sample Preparation

Please document the SOPs used for sample processing and note any deviations. Here are steps that should be performed and reported (where applicable):

- **Protein purification**: As purification methods may use detergents that affect the LC performance, please report both the purification and subsequent sample detergent clean-up protocols. S-Trap or SP3 workflow can be used for the clean-up of strong detergents, e.g., 4% sodium dodecyl sulfate (SDS). A fraction of the protein sample can be kept for post-proteomics result validation.
- **Protein estimation**: This is to measure the protein amount in each sample before sample preparation to ensure enough starting material, equivalent input material for quantitative experiments, and to determine the correct amount of protease (protease/protein ratio) for efficient protein digestion. Please note some experimental designs (e.g., protein pull-down experiments, high-throughput experiments, low-yield samples) may not allow this. For tissue/cell lysate protein, quality control can be performed by doing sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) of the lysate.
- **Protein digestion**: For most experiments, protein will be digested into peptides. Please report both the protease used and the digestion protocol. For pull-down experiments, on-bead digestion protocol may be modified depending on the bait expression (Cho, et al. 2020).
- **Peptide quantitation**: This is to measure the sample amount after sample preparation to determine the efficiency of sample processing and allow normalisation of sample loading for quantitative experiments. For example, Pierce™ Quantitative Fluorometric Peptide Assay (catalog no. 23290).
- **Peptide desalting**: Peptides are desalted before MS. For pull-down experiments, it is recommended to use styrenedivinylbenzene-reverse phase sulfonate (SDB-RPS) stage-tips for desalting to avoid contamination from polymers and traces of detergents.
- **Peptide labelling**: Label-dependent quantification methods (e.g., tandem mass tagging, TMT) are often used.
- **Fractionation**: For label-dependent proteomics and phosphoproteomics experiments, it is recommended to do high-pH reversed-phase liquid chromatography (RPLC) pre-fractionation on a suitable high performance LC due to high peptide abundance (Yang, et al. 2012). For TMT-based pull-down experiments, stage-tip based mini fractionation is recommended as the peptide abundance is lower.

2.3 Mass Spectrometry

Please report the LC and MS methodology by documenting the SOPs used and noting any deviations during sample analysis by MS. Here are steps that should be performed and reported:

- **Instrument operation and maintenance**: Perform and document all maintenance, incl. cleaning, calibration, preventative maintenance and addition of new components (e.g., trapping/analytical columns). Perform system suitability test periodically to assess the instrument performance.
- **Instrument performance**: Run a protein digest standard as a quality control and performance check before and after each batch of samples run. Performance can be monitored by spiking a peptide retention time calibration mixture together with HeLa digest or pooled experimental samples to rescue technical batch effect. Use the same data analysis workflow and static FASTA database to document the numbers of peptides and proteins identified as a measure of performance.
- **No sample carryover test**: To assess if there is sample carryover, blank samples should be run between experimental samples. This should be done during method development at least and

throughout the experiments if needed to minimise sample carryover contamination. Blank runs should be saved to report no carryover.

- **Data acquisition**. Whenever possible, DIA is recommended to reduce variability and improve protein identification coverage. For TMT experiments, please state if MS2 or MS3 data acquisition is used (applicable only to Thermo Orbitrap Tribrid platforms such as Fusion, Lumos and Eclipse). MS3 is recommended to reduce isolation interference caused by co-eluting peptides.

2.4 Data Analysis

Please document the software and analysis pipeline used, including:

- **Software/pipeline information**: Such as version number, parameters, link to the code repository.
- **Search parameters**: Report all mass search parameters such as mass tolerance, false discovery rate, modifications, protein filtering, quantitation method.
- **Statistical design**: For example, *n* numbers, power calculations, normality, data transformation, statistical model used to determine differential expression.
- **Database**: Store and report the peptide search database and version used for protein identification. Recommended databases: Swiss-Prot (manually curated) and TrEMBL (includes alternative splicing, but not manually curated). Please also report and share any library generated for DIA analyses in, for example, PRIDE.

2.5 Data Sharing

In line with the UK DRI data sharing policy, (curated) proteomics data should be openly released in a public repository that is part of the ProteomeXchange Consortium no later than the publication of the main findings, the submission of the main findings to a preprint repository, or two years after generation, whichever is the earliest. We recommend that you deposit your (meta)data in PRIDE (how to submit)[3]. Please make note of the mandatory files (and standards) required for a complete submission prior to the start of the experiment, so you can compile the necessary (meta)data during the experiment and ensure your data is *born FAIR*.

3 Bibliography

- KF Cho, TC Branon, ND Udeshi, SA Myers, SA Carr, and AY Ting. 2020. "Proximity labeling in mammalian cells with TurboID and split-TurboID." *Nature Protocols* 15: 3971-3999.
- F Yang, Y Shen, DG Camp 2nd, and RD Smith. 2012. "High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis." *Expert Rev Proteomics* 9: 129-134.

Version	Date	Author(s)	Notes
1.0	13/09/2021	B. Diaz-Castro, B. Geary, J. Jackson, S. Jackson, I. Lin, R. Nirujogi	

[3] Submitted datasets are 'private' by default until made public. To share 'private' data, please create a reviewer account for your dataset during the submission process. The reviewer account, which you can share with your collaborator(s), will give access to all the files included in a dataset.