Data Analysis and Statistics

## TMT quantitative mass spec pipeline

To ensure consistency across all publicly available mass spectrometry datasets that used TMT isobaric mass tagging, we applied the same standardized analysis pipeline previously published. This approach allows for better comparability between datasets. Several of the datasets included in this repository were generated and analyzed at the European Molecular Biosciences Laboratories (EMBL) in Heidelberg, Germany, where Dr. Frank Stein developed and implemented the data analysis pipeline. In summary, we processed the mass spectrometry (MS) data using FragPipe1, following established methodologies outlined in Farley et al. (2024, A), Farley et al. (2024, B), and Thomas et al. (2025)2–4. To ensure accurate protein identification, we used a UniProt FASTA database that included common contaminants and reversed sequences for validation. Proteins were included in the final dataset only if they contained at least two unique peptides, each at least seven amino acids long.

To account for potential technical variations across experiments, we applied a normalization step using the ‘removeBatchEffect’ function from the limma package, stabilizing variance in log2-transformed TMT reporter ion intensities5,6. Further normalization was carried out using the ‘normalizeVSN’ function of the same package. We identified differentially expressed proteins using the moderated t-test within limma, incorporating replicate information into the statistical model through the ‘lmFit’ function.

For the protein-lipid probe interaction experiments (+/- UV exposure), proteins were categorized based on their enrichment in the +UV condition. Proteins showing a log2 fold-change of at least 1 and a p-value below 0.05 were classified as “enriched hits,” while those showing similar enrichment trends but with p-values above 0.05 were labeled as “enriched candidates.”

To better understand the biological significance of the detected proteins, we conducted gene ontology (GO) enrichment analysis using the ‘compareCluster’ function from the ‘clusterProfiler’ package7. This method evaluates whether specific GO terms are overrepresented in the dataset compared to a background gene set. We performed enrichment analysis for three GO categories: Cellular Component (CC), Molecular Function (MF), and Biological Process (BP). The reference database used was ‘org.Pf.plasmo.db.’ To quantify enrichment, we calculated the odds ratio by comparing the proportion of genes associated with each GO term in our dataset (‘GeneRatio’) with the proportion in the reference background set (‘BgRatio’). A value greater than 1 suggests that a given GO term is more prevalent in our dataset than expected by chance.

1. Yu, F. *et al.* [Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform](https://doi.org/10.1038/s41467-023-39869-5). *Nat Commun* **14**, 4154 (2023).

2. Farley, S., Stein, F., Haberkant, P., Tafesse, F. G. & Schultz, C. [Trifunctional Sphinganine: A New Tool to Dissect Sphingolipid Function](https://doi.org/10.1021/acschembio.3c00554). *ACS Chem. Biol.* **19**, 336–347 (2024).

3. Farley, S. E. *et al.* [Trifunctional fatty acid derivatives: The impact of diazirine placement](https://doi.org/10.1039/D4CC00974F). *Chem. Commun.* **60**, 6651–6654 (2024).

4. Thomas, A. *et al.* Trifunctional lipid derivatives: PE’s mitochondrial interactome. *Chem. Commun.* 10.1039.D4CC03599B (2025) doi:[10.1039/D4CC03599B](https://doi.org/10.1039/D4CC03599B).

5. Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. [Variance stabilization applied to microarray data calibration and to the quantification of differential expression](https://doi.org/10.1093/bioinformatics/18.suppl_1.S96). *Bioinformatics* **18**, S96–S104 (2002).

6. Ritchie, M. E. *et al.* [Limma powers differential expression analyses for RNA-sequencing and microarray studies](https://doi.org/10.1093/nar/gkv007). *Nucleic Acids Research* **43**, e47–e47 (2015).

7. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. [clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters](https://doi.org/10.1089/omi.2011.0118). *OMICS: A Journal of Integrative Biology* **16**, 284–287 (2012).