Mass spectrometry analysis of non-protein amino acid misincorporation and proteomic changes in neurotoxicity-related cellular pathways

Joel Ricky Steele

Bachelors Medical Science (Hons)

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy from:

Faculty of Science,

University of Technology Sydney

2022

Certificate of Original Authorship

I, Joel Ricky Steele declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

Signature:

Production Note:
Signature removed prior to publication.

Date: 14th May 2022

Acknowledgements

The completion of a PhD thesis is a monumental task, and it takes many hands for this to be possible from those that came before us and those that assisted during the process, I would like to recognise all of those that made this possible. Jumbunna, I would like to acknowledge all of those that came before me, my brothers & sisters, my Ancestors, and the land we are on. I would like to thank the Aurora Education Foundation for providing me opportunities that have propelled my career down a path that I could never have dreamed possible. The many teachers that encouraged me to get where I am and fanned the flames for my love of science. Nicole Pitt, Margret St Hill and Beryl Chisholm.

My proteomics story began with a lecturer that called out some then shameful political figures and shone a light on the world of proteomics. Sitting in the lecture theatre I remember hearing about how amazing proteins were and that they were the machines of the cell and so much more, the shock that this field of research had not been covered in the previous two years of university or at high school was astounding. This field of research had the promise to change the world. I still think this rings true as we move towards research that characterises true phenotypes of disease. It was by happen chance that I saw Matthew Padula at the woollies at central, it was there that I asked him to allow me to perform research in his laboratory, this was the point where I can say my goals became clear on what sort of career I wanted to have and how I could make an impact on the world. I thank him for the opportunities and mentorship that he has provided me over the past eight years, the editing of my work, the many well worded references and everything else that goes along with mentoring a student. Thank you for exposing me to Lorne and welcoming me to the world of proteomics, I wish you the best in all things.

Beginning my research, I was taken under the wing of Dr Kate Harvey, who till this day has been one of the most kind-hearted, patient, and intelligent people I know. I would like to thank you for everything you have done for me, especially performing all those growth curves for me so I didn't have to manually figure out incubation times. Your friendship has been one of the best things in my life. Thank you also to Marina Harvey for the many vodka nights and the good life and academic advice.

I would like to thank all the members of the UTS Proteomics Facility over the past eight years, you have all been great to work beside and thank you all for your guidance/help. Furthermore, I would also like to thank Steven Djordjevic for his guidance as well as all the members of his lab over the years. This includes Ronnie, Jacqui, Iain, Mars, Cam, Max and many others.

Ken Rodgers, I would like to thank you for your mentorship and the opportunities you have afforded me, you took a chance on supervising me. You have been a great friend, helping me grow to the scientist I am and more importantly put up with my hairbrained ideas (and poor writing). Similarly, I would like to thank the many members of the Rodger laboratory over the years, beginning with Kate and Brendan. Thank you, Jake, for the help and friendship over the years, I wish you nothing but the best.

Thanks to all the undergrad friends who helped me get here, especially Dr. Angus Clulow for his Gus-Pass sessions, Dr. Fennel, Ante M and all those others for their care and friendship. Thank you to the close members of the office/labs, my fellow PhD-sufferers including Amber and Andrew.

Abstract

Neurodegenerative diseases cause significant morbidity and mortality globally, with the prevalence continuing to rise due to prolonged life expectancy. Many neurodegenerative disorders share a common pathology that involves protein misfolding, aggregation and deposition in the brain. Dietary intake of non-protein amino acids has previously been linked to such proteinopathies, with indirect evidence indicating potential misincorporation of non-protein amino acids into growing protein chains. Phenotypic and proteomic investigations could provide more direct evidence of misincorporation and further elucidate the role that non-protein amino acids may play in neurodegenerative disease. The aim of this work was to determine if non-protein amino acids incorporate into the human proteome at a level detectable by mass spectrometry, with a focus on the amino acids L-DOPA, BMAA, and azetidine 2-carboxylic acid. An enzymatic method for the conversion of tyrosine residues to L-DOPA was successfully developed, providing a basis for studying the incorporation of L-DOPA into proteins. L-DOPA incorporation into proteins was also detected following treatment of human neuronal cells *in vitro*, with quantitative proteomics revealing activation of the unfolded protein response, evidence of oxidative stress, and changes in pathways involved in neurodegenerative diseases. Meta-analysis of proteomics datasets revealed a significant effect of sample preparation on the oxidation of samples, which could potentially mask true *in vivo* oxidation. Labelling techniques and mass spectrometer resolution were also found to be important for the identification of unique peptides and modifications, including misincorporated amino acids. The treatment of human neuronal cells with BMAA *in vitro* induced proteomic changes indicating a profile of toxicity like that previously reported for glutamate-mediated excitotoxicity, but the incorporation of BMAA into proteins was not detected. Conversely, the incorporation of azetidine 2-carboxylic acid into proteins was readily detectable following *in vitro* treatment of cells, importantly in proteins involved in cell proteostasis. Azetidine 2-carboxylic acid also resulted in quantitative proteomic changes, including an increased abundance of protein folding machinery and a decreased abundance of translational machinery. The significant proteomic changes in neuronal cells following exposure to all three non-protein amino acids investigated indicated changes in pathways potentially related to neurodegeneration and neurotoxicity, indicating a potential role in such pathologies that should be further explored. This thesis also provided direct evidence that certain non-protein amino acids can be incorporated into human proteins at a level detectable by mass spectrometry, paving the way for future studies to further investigate the role of such amino acids in human disease.

Contents

List of Figures

List of Tables

List of publications

Publications associated with this thesis

Steele, J.R., Italiano, C.J., Phillips, C.R., Violi, J.P., Pu, L., Rodgers, K.J. & Padula, M.P. (2021), 'Misincorporation proteomics technologies: a review', *Proteomes*, vol. 9, no. 1, p. 2.

Steele, J.R., Strange, N., Rodgers, K.J. & Padula, M.P. (2021), 'A novel method for creating a synthetic L-DOPA proteome and in vitro evidence of incorporation', *Proteomes*, vol. 9, no. 2, p. 24.

Violi, J.P., Bishop, D.P., Padula, M.P., **Steele, J.R.** & Rodgers, K.J. (2020), 'Considerations for amino acid analysis by liquid chromatography-tandem mass spectrometry: A tutorial review', *TrAC Trends in Analytical Chemistry*, p. 116018.

Quinn, A.W., Phillips, C.R., Violi, J.P., **Steele, J.R.**, Johnson, M.S., Westerhausen, M.T. & Rodgers, K.J. (2021), 'β-Methylamino-L-alanine-induced protein aggregation in vitro and protection by L-serine', *Amino Acids*, vol. 53, no. 9, pp. 1351-9.

Samardzic, K., **Steele, J.R.**, Violi, J.P., Colville, A., Mitrovic, S.M. & Rodgers, K.J. (2021), 'Toxicity and bioaccumulation of two non-protein amino acids synthesised by cyanobacteria, β-N-Methylamino-L-alanine (BMAA) and 2, 4-diaminobutyric acid (DAB), on a crop plant', *Ecotoxicology and Environmental Safety*, vol. 208, p. 111515.

Other works published during PhD candidature

Widjaja, M., Harvey, K.L., Hagemann, L., Berry, I.J., Jarocki, V.M., Raymond, B.B.A., Tacchi, J.L., Gründel, A., **Steele, J.R.** & Padula, M.P. (2017), 'Elongation factor Tu is a multifunctional and processed moonlighting protein', *Scientific Reports*, vol. 7, no. 1, pp. 1-17.

Facey, J.A., **Steele, J.R.**, Violi, J.P., Mitrovic, S.M. & Cranfield, C. (2019), 'An examination of microcystin-LR accumulation and toxicity using tethered bilayer lipid membranes (tBLMs)', *Toxicon*, vol. 158, pp. 51-6.

Jarocki, V.M., **Steele, J.R.**, Widjaja, M., Tacchi, J.L., Padula, M.P. & Djordjevic, S.P. (2019), 'Formylated N-terminal methionine is absent from the Mycoplasma hyopneumoniae proteome: Implications for translation initiation', *International Journal of Medical Microbiology*, vol. 309, no. 5, pp. 288-98.

O'Rourke, M.B., Town, S.E., Dalla, P.V., Bicknell, F., Koh Belic, N., Violi, J.P., **Steele, J.R.** & Padula, M.P. (2019), 'What is normalization? The strategies employed in top-down and bottom-up proteome analysis workflows', *Proteomes*, vol. 7, no. 3, p. 29.

Berry, I.J., Widjaja, M., Jarocki, V.M., **Steele, J.R.**, Padula, M.P. & Djordjevic, S.P. (2021), 'Protein cleavage influences surface protein presentation in Mycoplasma pneumoniae', *Scientific Reports*, vol. 11, no. 1, pp. 1-15.

Chen, H., Wang, B., Li, G., **Steele, J.R.**, Stayte, S., Vissel, B., Chan, Y.L., Yi, C., Saad, S. & Machaalani, R. (2021), 'Brain health is independently impaired by E-vaping and high-fat diet', *Brain, Behavior, and Immunity*, vol. 92, pp. 57-66.

Prakash, A., Taylor, L., Varkey, M., Hoxie, N., Mohammed, Y., Goo, Y.A., Peterman, S., Moghekar, A., Yuan, Y., Glaros, T., **Steele, J.R.**, Faridi, P., Parihari, S., Srivastava, S., Otto, J.J., Nyalwidhe, J.O., Semmes, O.J., Moran, M.F., Madugundu, A., Mun, D.G., Pandey, A., Mahoney, K.E., Shabanowitz, J., Saxena, S. & Orsburn, B.C. (2021), 'Reinspection of a Clinical Proteomics Tumor Analysis Consortium (CPTAC) dataset with cloud computing reveals abundant post-translational modifications and protein sequence variants', *Cancers*, vol. 13, no. 20, p. 5034.

Conference Proceedings

Published abstracts

Rodgers, K., Chan, S. & **Steele, J.** (2017), 'Administration of L-tyrosine with levodopa could be neuroprotective in Parkinson's disease', Journal of Neurochemistry, vol. 142, Wiley 111 River St, Hoboken 07030-5774, NJ USA, pp. 245. **(Conference article)**

Rodgers, K., **Steele, J.** & Padula, M. (2017), 'A novel approach to detect the presence of levodopa (lDOPA) in the polypeptide chains of proteins', Journal of Neurochemistry, vol. 142, Wiley 111 River St, Hoboken 07030-5774, NJ USA, pp. 164. **(Conference article)**

Chen, H., **Steele, J.**, Li, G., Chan, Y., Oliver, B., Saad, S. & Machaalani, R. (2019), 'E-vapour inhalation–How does it affect memory?', IBRO Reports, vol. 6, pp. S208-S9. **(Conference article)**

Poster presentations

2017

22nd Annual Lorne Proteomics Symposium (Lorne, Australia)

Title: Using proteomic analysis to uncover the mechanisms of non-protein amino acids attributed to neurological diseases (#37). Authors: **Joel Steele**, Matt Padula, Kenneth Rodgers. Presented a lightning talk for this abstract, the poster also won an award.

2018

23rd Annual Lorne Proteomics Symposium (Lorne, Australia)

Title: Non-protein amino acids and neurological disease: their detection in human proteins and effects. Authors: **Joel Steele**, Matt Padula, Kenneth Rodgers.

2019

18th Human Proteome Organization World Congress – HUPO (Adelaide, Australia)

Three abstracts were awarded a poster presentation:

- Proteomic mapping of chemical warfare agent exposed plasma abs# 856.
- Mapping hydroxylated tyrosine in the human brain proteome: The formation and incorporation of L-DOPA abs# 857
- The neurotoxin β-Methylamino-L-alanine and its incorporation into proteins abs# 858

Abbreviations

α-2M	Alpha-2-macroglobulin
aaRS	Aminoacyl tRNA synthetase
AD	Alzheimer's disease
AEG	N-(2-aminoethyl) glycine
ALS	Amyotrophic lateral sclerosis
ALS-PDC	Amyotrophic lateral sclerosis-Parkinson's Dementia complex
AMBIC	Ammonium bi-carbonate
AQS	6-aminoquinolyl-N-hydroxysuccnimidyl-carbate
AZE	Azetidine-2-carboxylic acid
BCA	Bicinchonic acid
BMAA	β-methylamino-L-alanine
BOAA	β-N-oxalyl-α,β-L-diaminopropionic acid
BSA	Bovine serum albumin
CDC42	Cell division control protein 42 homolog
CHOP	CCAT-enhancer-binding protein homologous protein
CID	Collisional induced dissociation
CNS	Central nervous system
CSF	Cerebral spinal fluid
Da	Dalton
DAB	L-2,4-diaminobutyric acid
DDA	Data dependent analysis
D-DOPA	D-3,4-dihydroxyphenylalanine
DENR	Density-regulated protein
DIA	Data independent analysis
DMEM	Dulbecco's Modified Eagles' Medium
DTT	Dithiothreitol
EBT	1,1'-ethylidene-bis[L-tryptophan]
EDTA	Ethylenediaminetetraacetic acid
EMS	Eosinophilia Myalga Syndrome
ER	Endoplasmic reticulum
FDR	False discovery rate
FLD	Fluorescence detector
FL-HPLC	high-pressure liquid chromatography-fluorescence platform

GAPDH	glycerol-3-phosphate dehydrogenase
GC	Gas chromatography
GNB2	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2
HBB	Haemoglobin
HEPES	(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
HNRNPD	Heterogeneous nuclear ribonucleoprotein D0
IAA	iodoacetamide
IDA	Intelligent data Acquisition
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography tandem mass spectrometry
L-DOPA	L-3,4-dihydroxyphenylalanine
LFQ	Label free quantification
LOPIT	Localisation of organelle proteins by isotope tagging
m/z	Mass-to-charge ratio
MBP	Myelin basic protein
MEM	Minimum Essential Medium
MiP	Misincorporation proteomics
MND	Motor neuron disease
MRM	Multiple reaction monitoring
mRNA	Messenger RNA
MS	Multiple sclerosis
MS/MS	Tandem mass spectrometry
NBT	Nitroblue tetrazolium
NPAA	Non-protein amino acid
OST	Oligosaccharyl transferase complex
Ox-Met	Oxidised methionine
Ox-Phe	Oxidised phenylalanine
PB-DOPA	Protein bound L-DOPA
PBS	Phosphate-buffered saline
PD	Parkinson's disease
PMI	Post mortem interval
PPIA	Peptidyl-prolyl cis-trans isomerase A
PRM	Parallel reaction monitoring
PRMT1	Protein arginine N-methyltransferase

PSM	Peptide spectral match
PTM	Post translational modification
PVDF	Polyvinylidene fluoride
ROS	Reactive oxygen species
RTS-SPS-MS3	Real time search enabled SPS-MS3
SART1	U4/U6.U5 tri-snRNP-associated protein 1
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	standard error of the mean
SILAC	Stable isotope labelling by amino acids in cell culture
SLE	Systemic lupus erythematosus
SNRPG	Small nuclear ribonucleoprotein G
SNRPGP15	Putative small nuclear ribonucleoprotein G-like protein 15
SPRM1	Serine/arginine repetitive matrix protein 1
SPS-MS3	synchronous precursor selection based MS3
SRM	Single reaction monitoring
SRPR	Signal recognition particle receptor subunit alpha
TAILS	N-terminal isotopic labelling of substrates
TCA	Trichloroacetic acid
TCEP	tris(2-carboxyethyl)phosphine
TDP-43	TAR DNA-binding protein 43
TMT	Tandem Mass Tags
TOF	Time of flight
tRNA	Transfer RNA
UPR	Unfolded protein response
UTC-7	7M urea, 2M thiourea, 0.1% C7BzO

Thesis organisation

The organisation of this thesis is outlined below:

- Chapter One: This introduction frames the research questions for this thesis.

- Chapter Two: Published critical review of the literature concerned with NPAAs, the methods used to study their role and effect on an organism's proteome, and establishment of the formal pursuit of proteomic incorporation of NPAAs, with technologies and considerations outlined to advance the field of NPAA study.

- Chapter Three: A method for the enzymatic conversion of proteomes to contain L-DOPA to create reference mass spectra for analysis of samples that potentially contain proteoforms with L-DOPA incorporated, as well as providing evidence for L-DOPAs *in vitro* toxicity, and the parallels of the toxicity to a state of neurodegeneration highlighted.

- Chapter Four: Meta-analysis of publicly available data for the presence of proteoforms/peptidoform incorporated L-DOPA to establish a baseline of L-DOPA presence in the human proteome. The draft map of the human proteome (brain subset) was analysed as a baseline for control. A Parkinson's disease TMT labelling experiment on the substantia nigra was also analysed and finally a label free LC-MS/MS dataset of the proteome of the olfactory lobes of Parkinson's sufferers.

- Chapter Five: The effect of BMAA and Azetidine 2-carboxylic acid on the neuronal proteome of SH-SY5Y cells and their incorporation.

- Chapter Six: General discussion, future directions and concluding remarks.

- Appendices

- References

Chapter One: Introduction and Overview of Thesis

1.1 General Introduction

Due to an increase in life expectancy brought about by advances in medical science, neurodegenerative diseases are causing higher mortality and morbidity than ever as they are age-associated diseases [1-3]. Many of these disorders have a common pathology which involves protein misfolding, aggregation and deposition in the brain [1, 4-22]. In an attempt to increase quality of life for neurodegenerative disease sufferers, there has been a major research focus on identification of inherited or susceptibility genes [13, 14, 23-28]. The weakness of focusing on genetic mutations is that only 10% of patients across all neurodegenerative diseases have the familial form and 90% have a sporadic form, which implicates an environmental trigger as one of the causative agents [13, 23, 25-28]. Researchers have justified the study of genetic causes because the familial and sporadic forms share similar symptoms and prognosis, but this overlooks the important role that non-genetic factors play in the disease pathology [29]. Common features of these neurological disorders include aggregates of proteins both intra and extracellularly, generally referred to as amyloids although the composition of the aggregate is disease and protein specific. Several neurodegenerative diseases have been termed 'proteinopathies' due to their protein misfolding characteristics and include Alzheimer's, amyotrophic lateral sclerosis (ALS), Lewy body dementia, prion disease, Huntington's and Parkinson's disease [12, 29-37]. Other key features of these diseases are dysregulation in energy metabolism, mitochondrial dysfunction, and oxidative stress [33, 38-55]. Modified proteins resulting from these proteinopathies have also been investigated as a source of intra-neuron transmission via synaptic release of aggregates or toxic species referred to as 'prionoids' [32, 56-58].

The identification of a 'hot-spot' of ALS cases on the island of Guam in the 1950s stimulated interest amongst researchers as to possible environmental causes of neurodegenerative disease [59]. This particular disease did not seem to have a genetic link, disappearing over time within the indigenous population as they adopted an American lifestyle and diet. One factor that gained attention was consumption of food that contained the non-protein amino acid (NPAA) β-Methylamino-L-alanine (BMAA) in their traditional diet [60, 61]. A feature of this disorder was a long latency period and it was reported that migrants from Guam developed similar symptoms ~30 years after leaving. Within the literature, a mechanism was proposed for this and other proteinopathies that attempted to explain not only the latency of disease development but also the interplay between genetics and the environment. The basis of this hypothesis was that non-protein (non-canonical) amino acids could be associated or incorporated into proteins thereby promoting protein aggregation and deposition in the brain [62-75]. The alteration of a proteoform's structure by a single amino acid modification could

lead to misfolding and aggregation, in accordance with all the major pathways of protein folding including hydrophobic collapse and chaperone assisted stabilisation [76, 77]. Within the environment, organisms such as *fescue* grasses use this mechanism to inhibit the growth of neighbouring plants by excreting non-protein amino acids (NPAAs) in a process known as allelopathy. The uptake of these NPAAs results in protein aggregation and ribosomal stalling as the neighbouring plants do not have tRNAs that can distinguish the NPAA from its canonical analogue, with examples including meta-tyrosine and L-3,4, dihydrophenylalanine (L-DOPA) [66, 78]. This is an exploitation of the lack of fidelity of protein translation and within the environment there are over a thousand amino acids that have been identified, with some being so similar in structure to the canonical 20 proteogenic amino acids that they can theoretically be incorporated [64]. To understand how this is possible, protein translation needs to be outlined to demonstrate that NPAA incorporation is possible through its ability to fit within a tRNA binding pocket and escape the steps in proof reading employed by the cell [79].

Protein translation involves several steps starting with a free-floating amino acid and resulting in its addition to a growing peptide chain by the ribosome. A complete overview of the process can be found in chapter two (also *see* [80]). Briefly, in canonical synthesis the amino acid is charged to its corresponding aminoacyl tRNA synthetase forming a complex which subsequently will bind the charged amino acid to its appropriate tRNA and there are several editing steps to release incorrectly bound amino acids, with the resultant mistranslation rate general quoted to be 1/10,000 at any particular site [81]. Finally, the amino acid with the attached tRNA are used by the ribosome translating the mRNA transcript into a polypeptide chain. If not folded correctly, the subsequent chain will be targeted by degradation and the amino acid released to the free amino acid pool to potentially be used again [80].

1.1 Non-protein amino acids and incorporation

In the 1960s there was an interest in how, despite the protein synthesis machinery being programmed for the 20 canonical amino acids, certain NPAAs could be mistakenly incorporated into proteins in bacteria. Fowden *et al* (1967) utilised *Escherichia coli* as a model to demonstrate the allelopathic effects of these "toxic amino acid species". The studies investigated the localised incorporation of NPAAs into the active sites of enzymes inferred by loss of function in kinetic assays and the loss of protein solubility [82, 83]. The determination of the ability of an NPAA to be incorporated used a screening system model, whereby the NPAA being assessed is used as a replacement for each of the 20 'protein-coded' amino acids using the ATP-PPi exchange methodology. ATP-PPi exchange is where an amino acid tRNA Synthetase (aaRS) is combined with the NPAA and ATP to see if binding will occur, normally indicated by colourimetric stains or by determining the free concentration of amino acids at the end of incubation, which is used to infer incorporation [84]. For example, the incorporation of the

NPAA azetidine-2-carboxylic acid (AZE) has been investigated in a system in which human myelin basic protein was overexpressed in *E. coli* [74]. SDS-PAGE and mass spectrometry (Figure 1) showed a change in protein molecular weight which the authors claim demonstrates incorporation of AZE into the protein. However, there is no direct sequence evidence from a clear fragmentation (MS/MS) spectrum that shows actual protein identity or NPAA incorporation [74].

Fig. 5. Mass spectrometry of 18.5 kDa rmMBP (unmodified rmC1 variant) purified from 2xYT (yeast-tryptone) rich media supplemented with 1 mM Aze [1]. The main peak is at 19,421.50 Da (correct form).

Fig. 4. Mass spectrometry of 18.5 kDa rmMBP (unmodified rmC1 variant) purified from M9 minimal media supplemented with 1 mM Aze [1]. The calculated molecular mass of rmC1 is 19,421.59 Da (Bates et al., 2000). The expected difference for each Pro (2) → Aze (1) substitution is 14 Da. The main peak is at 19,422.00 Da (correct form), with successively smaller peaks at 19,408.60/19,408.00 Da (one Pro (2) → Aze (1) substitution), 19,393.70/19,392.00 Da (two Pro (2) → Aze (1) substitutions), and a barely detectable peak at roughly 19,380.00 Da (three Pro (2) → Aze (1) substitutions).

Figure 1 Published spectra of incorporation: AZE in place of proline [74]. Left spectra displays unmodified (Murine) recombinant myelin basic protein expressed in E. coli *synthesis model. The right spectra demonstrates mass subtraction indicative of NPAA replacement of proline. Peak 19,422.000 Da is the unmodified peak with smaller peaks of 19,408.0 Da indicative of a singular replacement of proline by AZE -14 Da, subsequently a double replacement of proline by AZE is evidenced by 19,393.6992 and a triple replacement with 19,380.00.*

There are a number of examples of NPAAs that could potentially play a role in neurodegenerative diseases. AZE has been the centre of a multiple sclerosis hypothesis proposed by Rubenstein *et al* [85], stating that the incorporation of AZE, which is produced by sugar beets (*Beta vulgaris*) and extracted products of which are found in food, into myelin in place of its analogue proline resulted in autoimmune attack [86, 87]. This research was performed in *E. coli* utilising an overexpression system of myelin basic protein and subsequent radioactive labelling [74, 85, 88, 89] with seminal studies by Zagari *et al* studying the effect of incorporation on bioinformatically modelled protein structures demonstrating dramatic perturbations [90] [91-95] . Published work in the field includes incubation of chick embryos with AZE resulting in incorporation of 10% with 48ug/ml (of embryo volume) and 61% at 194ug/ml, determined by a 'specific chemical procedure' utilising radio labelling and acid hydrolysis by Joel Rosenbloom and Dawin Prockop [96, 97]. This body of work, is consistent with, but does not definitively demonstrate incorporation of the NPAA into the protein.

Another example of a NPAA replacing a protein amino acid in protein synthesis is that of L-canavanine. L-canavanine is produced by the edible plants *Canvalia ensiformis* (common jack bean) and *Hedysarum alpinum* (a wild potato) and has been shown to compete with L-arginine for tRNA^{Arg} charging [98]. Experiments by Rosenthal *et al* (1987) utilised a radio isotope labelled canavanine to

investigate its incorporation into proteins in the tobacco hornworm (*Manduca sexta)* and tobacco budworm (*Heliothis virescens*), and demonstrated that the *H. virescens* tRNA[Arg] discriminates between L-canavanine and arginine, leading to a lower rate of incorporation in that species [98, 99]. Following this work, a study was performed on the anti-tumorigenic properties of L-canavanine in a murine model, with L-canavanine presence in the protein fraction found to be 1% at 4 and 24 hours post injection. The treatment resulted in a decline in animal health and histology of fibrosis in the pancreas with serum depletion of urea nitrogen and cholesterol [100]. The incorporation of L-canavanine into proteins has also been shown in SH-SY5Y neuroblastoma cells by radioactive labelling and confirmed by the stalling of mitochondrial ribosomal translation [101]. Direct evidence of misfolded proteins was not presented other than by inference from ribosomal stalling [101]. The conclusions of this experiment were that the use of L-canavanine causes stalling at the end of the ribosomal funnel at the interaction point with the aqueous cytoplasm. For the measurement of ribosomal stalling, pulsed *in organello* labelling was used which involves the addition of [^{35}S]methionine and [^{35}S]cysteine to a depleted media for 10 minutes and subsequent protein extraction, following which the proteins are resolved by SDS-PAGE and a phosphor-imaging scanner used to detect the amplified signal. This method relies on parallel Western blotting to identify mitochondrial proteins so they can be accurately quantified (method:[101, 102]). This method was used to infer that misfolded products were being produced, resulting in the observed toxicity and supported by further pulse-chase labelling experiments [101]. L-canavanine has been inferred to be incorporated into proteins but no studies of the larger translated proteins were performed, reportedly due to the toxicity of L-canavanine limiting long high dosage exposure. Proteomic evidence is lacking in the reported experiments which utilised Western blotting rather than mass spectrometry. Western blotting quantitation should not provide the sole identification of proteoforms as the usage of antibodies in an experiment that aims to produce non-native unmapped proteins could be leading to false binding due to alterations of the antibody's epitope.

The landmark paper by Dunlop *et al* (2013) provided some evidence for protein incorporation of the NPAA β-Methylamino-L-alanine (BMAA) and subsequent protein aggregation [103]. The utilisation of tritiated BMAA (^{3}H-BMAA) to treat the human cell lines SH-SY5Y (neuroblastoma), MRC-5 (lung fibroblast) and HUVEC (primary human umbilical vein endothelial cells) found radioactivity in isolated protein fractions, providing evidence of some association between BMAA and proteins [104]. It is worth noting that while the 10% tricholoroacetic acid (TCA) insoluble fraction is often referred to as the 'protein fraction', it will contain other molecules that have reduced solubility in TCA and can be isolated by centrifugation. Amino acid hydrolysis followed by derivatisation with 6-aminoquinolyl-N-hydroxysuccnimidyl-carbate (AQS), separation by liquid chromatography, and identification using

tandem-mass-spectrometry (LC-MS/MS) was used to measure the concentration of BMAA, and it was demonstrated to be present in the protein fraction in a dose-dependent manner with a linear correlation to the amount supplied [105]. To further support that the association was related to protein synthesis, the addition of cycloheximide (protein synthesis inhibitor) alongside BMAA significantly reduced the concentration of BMAA in the protein fraction [105]. In addition, fluorescence microscopy demonstrated an increase in levels of auto-fluorescent material in cells, a technique previously used to investigate incorporation of oxidised amino acids into proteins [106-108]. A further crucial experimental finding was that L-serine (250μM) supplementation decreased BMAA incorporation by ~70% when compared against the control condition whilst the enantiomer D-serine produced no statistically significant decrease [103].

Further studies by Glover *et al* (2014) built on Dunlop's work by utilising the commercial transcription kit PURExpress [109], a cell free system of purified *E. coli* transcriptional components. Utilising BMAA and two of its isomers, L-2,4-diaminobutyric acid (DAB) and N-(2-aminoethyl) glycine (AEG), as amino acid competitors in a screening of all 20 amino acids in the *E.coli* protein synthesis model showed that the kit has a lack of discrimination for NPAAs during protein synthesis alongside a BMAA incorporation rate of 0-80%. The incorporation of AEG should not be possible due to the distance between the primary amine and the carboxylic acid group being five carbons in length, as the normal formation of a peptide bond (dehydration reaction) requires the primary amine being one carbon separated from the COOH group. Therefore, canonically AEG should not form peptide bonds and incorporate into protein backbones, but within this de-novo synthesis model there was a measured incorporation of 7% for AEG and 3% for DAB which raises concerns over the validity of the peptide de-novo synthesis model applied [104].

The protein association of BMAA has also been shown by the depletion of its presence in the free amino acid pool, measured in treatment media by AQS and LC-MS/MS quantitation, demonstrating a 24% protein association to the protein fraction [110]. Within the literature, there is debate over whether the protein precipitation used during amino acid analysis of protein hydrosylates was incidentally precipitating BMAA alongside the protein (or not removing association) [110-112], so the denaturants SDS and dithiothreitol were used on the acetone precipitated pellet resulting in a release of half of the total BMAA that was previously assumed in studies to be protein associated. This paper displays that the amount of inferred incorporation in diseased brain tissue presented by authors in this field was inflated [110]. Taken together, Dunlop and Glover have determined that the incorporation of BMAA occurs in competition with serine, however these papers have not provided protein sequence information and thus actual incorporation of BMAA into proteins has still not been demonstrated. An attempt to investigate the incorporation of BMAA into proteins in place of serine

by utilising LC-MS/MS analysis of proteome digests from BMAA treated murine cell culture was performed by Beri *et al* (2017), but the authors were unable to provide evidence of incorporation [38]. The experimental design and methodologies used in previous studies do not allow for the detection of incorporation, and the reasons for this is discussed in a later section (chapter two).

For an NPAA to have an impact on human or animal health due its ability to be mistakenly incorporated into proteins there has to be a dietary or environmental source and a route by which it can enter the body. In the case of BMAA, a NPAA synthesised by cyanobacteria, there is the possibility of dietary consumption from contaminated seafood [113], inhalation following aerosolisation [114], and contamination of water sources by the cyanobacterial toxin [115-120]. Based on the concentrations of BMAA reported in the environment it is likely that the level of human exposure to BMAA is very low although chronic exposure might produce a pathological effect over a longer period of time if accumulation occurs. A prime example of high exposure to a NPAA is in the case of symptomatic treatment for Parkinson's disease (PD) by the NPAA L-3,4-dihydroxyphenylalanine (L-DOPA). L-DOPA is used in PD patients to restore the levels of dopamine, which restores motor function [57, 58, 121-131]. The use of L-DOPA in PD is inevitable in all cases as quality of life is increased and it is the most effective therapeutic in PD patients. However, the drug has been shown to be toxic by several studies by producing oxidative stress [57, 132-152].

Another group exposed to L-DOPA are students in higher education and the wider community who consume it as a "nootropic" drug [153]. Nootropics are cognitive performance enhancers mainly consisting of stimulants and, due to the legal grey area, L-DOPA has been found to be marketed as a nootropic which could be hazardous to human health [154, 155]. Despite the widespread prescribing of L-DOPA for over 50 years to treat PD, incorporation of L-DOPA into human proteins has not been directly analysed by proteomics. Inference by several radioactive labelling approaches and Western blotting for proteins containing L-DOPA has been applied but no sequence information has been reported in humans [9, 44, 45, 57, 58, 87, 108, 123-125, 156-167].

The charging of tRNA with L-DOPA for insertion into proteins was first observed to occur in 1966 using purified tRNAs from *E.coli* and *Bacillus subtilis*, specifically tRNA[Tyr] incubated with tritiated L-DOPA [168]. In the following decade, conflicting evidence was reported in a murine tRNA[Tyr] study by Hogenauer (1978) [169], resulting in the rejection of the possibility of incorporation of L-DOPA even at the therapeutic levels used in PD treatment. The issue of incorporation was not revisited until Rodgers *et al* (2004) provided evidence of incorporation *in vitro* utilising human and murine cell culture with radiolabelled L-DOPA, finding L-DOPA in the isolated protein fraction [108]. The body of work subsequently provided by Rodgers and colleagues has provided evidence not only of the

incorporation but the toxic effects of generated species *in vitro* utilising human cell culture [57, 58, 87, 106, 108, 124, 160, 164, 170, 171].

The detection method applied by Rodgers utilised the natural fluorescence of L-DOPA and tyrosine. A high-pressure liquid chromatography-fluorescence detection platform (FL-HPLC) was used for the detection of L-tyrosine, L-DOPA and meta-tyrosine which were quantitatively analysed post hydrolysis of the protein fractions and quantified against standards. This was orthogonally supported by scintillation counting of radio-labelled L-DOPA present in the TCA precipitated fractions [164]. The association of L-DOPA with proteins following exogenous supply was compared to the formation of L-DOPA due to hydroxyl radical attack on tyrosine, which would create the para-tyrosine (L-tyrosine) analogues, ortho-tyrosine and meta-tyrosine, from hydroxyl radical attack on phenylalanine. Experimental analysis showed low concentrations of L-DOPA (50μM) led to detectable levels (25-50 mMol DOPA per Mol tyrosine) in the protein fraction with a linear increase in concentration until 500μM treatment. To test the incorporation theory, cycloheximide was utilised to inhibit the synthesis of proteins; the results showed a 93% reduction in the amount of L-DOPA, which leaves the 7% attributable to protein association as protein synthesis had ceased or the proteins were synthesised prior to complete inhibition of synthesis by cycloheximide. Experimental labelling of proteins with radiolabelled L-leucine and L-DOPA allowed the degradation rates of proteins to be measured and demonstrated that radioactive L-DOPA turnover was slower than that of leucine at higher levels of DOPA; the data generally showed selective turnover of DOPA-containing proteins versus native proteins [164]. The resistance to degradation at higher levels of DOPA incorporation was interpreted as the generation of cross-linked protease-resistant proteins.

Further evidence for L-DOPA incorporation was provided by SDS-PAGE and utilised a redox staining method for catechols, using nitro blue tetrazolium (NBT). NBT-treatment of membranes following transfer from SDS-PAGE allowed the detection of proteins containing catechol groups such as DOPA in the cell lysate proteins. Positive staining was completely absent in proteins from untreated cells (indicating a lack of oxidation), but L-DOPA staining in proteins from DOPA-treated cells demonstrated the presence of catechol groups in many, if not all, proteins. The use of NBT-redox staining demonstrated a unique orthogonal way to detect L-DOPA in proteins, providing further evidence for its presence in proteins [87, 108]. Seen in Figure 2 (A) is the SDS-PAGE separated, NBT redox stained whole cell lysate from THP-1 human monocytes treated with cycloheximide, showing that only under protein synthetic conditions is L-DOPA protein-associated. Within Figure 2 (B) the same treatment of cells was performed at a 500μM L-DOPA concentration with 0.02 μM C^{14} L-DOPA tracer. A total protein stain was performed with subsequent redox staining and phosphor imaging demonstrating a clear pattern of L-DOPA containing proteins identifiable by both methodologies [108].
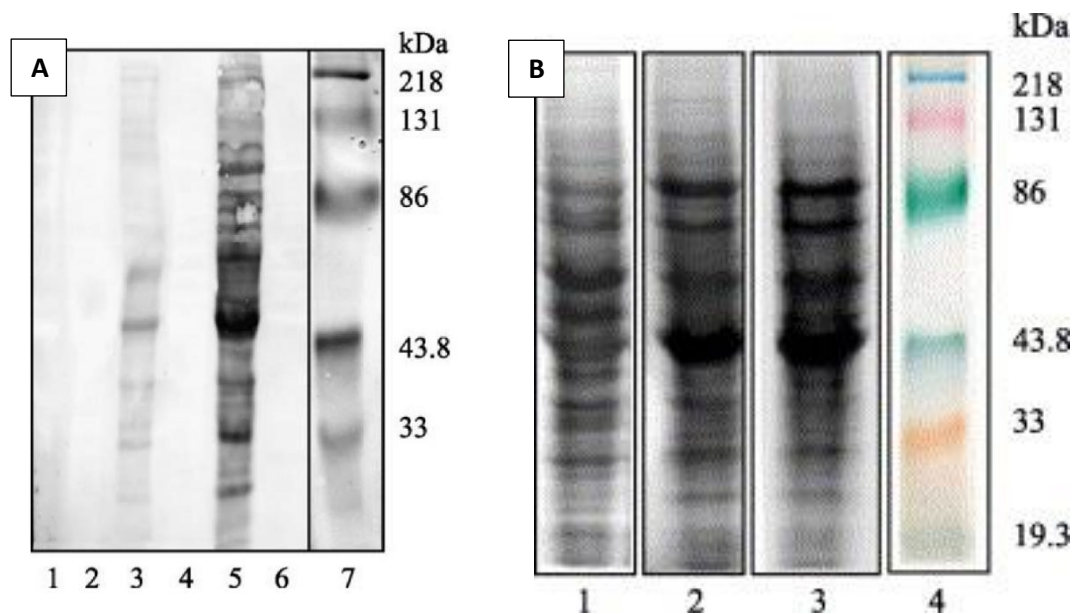
*Figure 2: Evidence of L-DOPA containing proteins. Images taken from Rodgers et al 2004 [108]. **A**: Human monocytes (THP-1) NBT stain of proteins indicating L-DOPA bound proteins. Lane 1-2: 0 µM L-DOPA; lanes 3-4: 150 µM L-DOPA; lanes 5-6: 500 µM L-DOPA. Lanes 2, 4, and 6 contained proteins from cycloheximide treated cells demonstrating inhibition of L-DOPA association. This redox stain demonstrates that L-DOPA is only found present in proteins during protein synthesis. **B**: Single-lane protein extract from THP-1 cells treated with 500µM L-DOPA with 0.02 µM [14C] L-DOPA, measured in series. Lane 1: amido black total protein stain. Lane 2: NBT redox stained proteins and lane 3 is radioactivity imaged by phosphor-imaging. This image further demonstrates that L-DOPA containing proteins are identified by NBT staining and confirmed by radioactive tracer presenting near identical banding of the same lane.*

Analysis of serum from L-DOPA treated PD patients found L-DOPA in circulation as its free form, with evidence of protein association/incorporation suggested by release of L-DOPA by hydrolytic digestion of the insoluble protein fraction and detection by FL-HPLC. The data presented in these studies provided evidence for L-DOPA incorporation into proteins *in vivo*, the authors arguing that it was unlikely to be due to oxidation since the oxidised forms of phenylalanine did not significantly differ in L-DOPA-treated and control patients [124]. Thompson *et al* also investigated protein associated L-DOPA and the effects on cellular homeostasis and protein turnover, reporting an upregulation in the expression of genes required to handle damaged or aggregated proteins [123]. Dunlop *et al* demonstrated, using microscopy, that cell-treatment with L-DOPA resulted in damage to lysosomes that eventually ruptured and this was interpreted as evidence for L-DOPA-containing protein accumulation in lysosomes [58]. The most recent studies published from Rodgers and colleagues produced evidence of incorporation in the neuroblastoma cell line SH-SY5Y and PD brain tissue. In an *in vitro* study using D-DOPA (the stereo isomer of L-DOPA) as a control for non-incorporation, it was demonstrated that the cell death was prominently due to the L-isomer [172]. Furthermore, fluorescent microscopy of SH-SY5Y cells and scanning electron microscopy show mitochondrial

dysfunction [57]. This research to date has provided evidence that L-DOPA can replace L-tyrosine in protein synthesis *in vitro* and *in vivo*. A study by Ozawa and colleagues using an *E. Coli* cell-free transcription/translation system demonstrated conclusively that L-DOPA can replace L-tyrosine in protein synthesis and results in the generation of proteins that are less soluble than the native forms [107, 124, 160]. The analysis of proteins containing L-DOPA by LC-MS/MS would greatly add to our knowledge of the biological mechanisms involved and consequences to cells and tissues from the generation of proteins containing incorporated L-DOPA. In addition, direct analysis of peptides containing peptide-bonded DOPA has not been performed but could provide data relevant to possible toxicity relating to this drug.

1.2 The absence of the smoking gun: Proof of NPAA Incorporation

A missing piece of experimental data to support the hypothesis of NPAA incorporation in proteins is MS/MS spectral evidence. The lack of this evidence arises because the most abundant molecules of a particular proteoform are translated 'normally' and the number of peptides containing the NPAA is below the limit of detection of methods that have been employed to date. It is also likely that the short exposure times used in *in vitro* and *in vivo* experimental models does not reflect the chronic exposure seen in clinical cases or post mortem. Thus, models of exposure need to be performed and investigated for NPAA protein incorporation and effects.

In most cases models of NPAA incorporation have not examined chronic exposure. Exposure of L-DOPA to neuronal cells for 7 days was performed by Chan *et al* but in most cell studies, exposure has been less than this [57]. Animal models of L-DOPA toxicity and human clinical trials have been extensive but even with proteomic analysis, the data sets have not been searched for or determined L-DOPA incorporation in models including zebra fish, murine models, non-human primates [163] and a plethora of human cells ranging from stem cells to neurons [140, 173-183]. Human clinical trials have indicated that cognitive decline due to L-DOPA is potentially masked by the positive activity relieving Parkinson's symptoms [177].

In the case of BMAA incorporation into proteins in the brains of individuals on Guam that was implicated in a complex disease that had features of ALS, PD, and dementia (termed ALS-PDC) and found in people exposed to high BMAA concentrations on the island of Guam, evidence shows a clear relationship between association to protein and neurodegeneration [60, 184-188]. Primate models have previously showed the symptoms [59] are caused by treatment with BMAA but it was not until Cox's 2016 paper was the histology examined and quantified. The author's noted that the low incorporation of 1/10,000 amino acids may beyond their capability to detect [63]. But given this evidence, there is still ambiguity around the mechanisms of pathogenesis that BMAA produces.

In all of the examples given so far, the evidence for incorporation is by inference but with little to no empirical phenotypic data at the proteomic level. Measurements need to be taken throughout exposure models to observe the changes in metabolism and protein abundance in order to observe the global effects *in vivo* [3, 56, 189-192]. The investigation of NPAA incorporation needs to performed along these lines of enquiry with proteomics to help elucidate their role in human disease.

1.3 Aims of this thesis

In the introduction, several NPAAs implicated in human neurological disease were discussed, namely: L-3,4-dihydroxyphenylalanine (L-DOPA), azetidine 2-carboxylic acid (AZE) and β-Methylamino-L-alanine (BMAA). While a range of *in vitro* and *in vivo* studies have provided evidence in support of their insertion into proteins in exchange for a protein amino acid, comprehensive proteomic analysis of the proteins has not been carried out but has the potential to provide insight into the impact of these NPAAs on human health.

- Overarching aim: To determine if non-protein amino acids incorporate into the proteome at a measurable level and if they cause neurodegeneration.
- Overarching hypothesis: NPAAs do incorporate into proteins at low levels resulting in neurodegenerative-like features via protein misfolding or effects upon cell status.

In extension to the overarching aims, there are several specific aims that will be addressed by this body of work.

1. To investigate whether L-DOPA is incorporated into human proteins *in vitro*. (Chapter 3)
2. To investigate whether L-DOPA is a naturally occurring protein constituent in normal human neuronal biology and Parkinson's disease. (Chapter 4)
3. To investigate the incorporation of BMAA and azetidine 2-carbocylic acid into the proteome and the effects on the proteome. (Chapter 5)

Chapter Two: Literature Review: "Misincorporation Proteomics Technologies: A Review"

Chapter Overview (published manuscript)

This chapter reviews the literature describing non-protein amino acid misincorporation into proteins and critically reviews proteomics-based approaches for the study of proteins containing NPAAs.

The current field of proteomics is sample and workflow centric. As the field matures many techniques are developed for specific applications to progress research into a particular biological material or protein analyte class of interest. Proteomics is a sample destructive and non-amplifiable technique and as such the major factors for consideration in proteomics are the amount of starting material, how to process to reduce loss and/or reduce complexity for deeper cataloguing or quantification, and finally how to best operate the mass spectrometry system to provide accurate identification and quantification.

This manuscript describes methods that have been developed for other avenues of proteome analysis that can be repurposed and applied to study misincorporation. Furthermore, this manuscript aimed to provide a unifying concept of misincorporation proteomics with the outlined work performed to date. Furthermore, the issues with performing misincorporation proteomics were outlined which include the major caveats of the field's tools which are linked to the concentration dependent analytical platforms (mass spectrometers) that are used for high throughput proteomics. The misincorporation events are of exceptionally low abundance and suffer from sequence isomer dilution, this being that positional isomers of NPAA incorporation leads to a dilution in signal falling below the limit of detection and in nearly every case do not have an established enrichment method or antibody that can be used for pull-downs to increase sensitivity and analytical depth.

The logical framework outlined within the manuscript can be ubiquitously applied across the entire field of proteomics, improving upon the depth and accuracy within experiments. Innovative methods for sample processing have been outlined as well as methods of operating mass spectrometry systems and the downstream bioinformatics packages required to handle these complex data.

Certificate of authorship and originality

This paper was published in *Proteomes* MDPI Ltd. I certify that the work presented in this chapter has not previously been submitted as part of the requirements for a degree. I also certify that I carried out the majority of the work presented in this paper.

- Joel Ricky Steele: Wrote the majority (~90%) of the manuscript and conceived the premise of the manuscript.

- Carly J. Italiano, Connor R. Philips, Jake P. Violi and Lisa Pu: contributed to the remaining ~10% of the manuscript, assisted in the construction of figures and proof read the final draft.

- Kenneth J. Rodgers and Matthew P. Padula: Proof-read and edited the manuscript and conceived project idea.

Primary Author   Co-author signatures

Joel Ricky Steele

Signature

Date: 10th October 2021.

Review

Misincorporation Proteomics Technologies: A Review

Joel R. Steele [1,2], **Carly J. Italiano** [2], **Connor R. Phillips** [2], **Jake P. Violi** [1,2], **Lisa Pu** [2], **Kenneth J. Rodgers** [2] and **Matthew P. Padula** [1,*]

[1]   Proteomics Core Facility and School of Life Sciences, The University of Technology Sydney, Ultimo, NSW 2007, Australia; Joel.Steele@uts.edu.au (J.R.S.); jake.violi@uts.edu.au (J.P.V.)
[2]   Neurotoxin Research Group, School of Life Sciences, The University of Technology Sydney, Ultimo, NSW 2007, Australia; carly.italiano@uts.edu.au (C.J.I.); connor.phillips@student.uts.edu.au (C.R.P.); Lisa.Pu@student.uts.edu.au (L.P.); kenneth.rodgers@uts.edu.au (K.J.R.)
*   Correspondence: Matthew.Padula@uts.edu.au

Abstract: Proteinopathies are diseases caused by factors that affect proteoform conformation. As such, a prevalent hypothesis is that the misincorporation of noncanonical amino acids into a proteoform results in detrimental structures. However, this hypothesis is missing proteomic evidence, specifically the detection of a noncanonical amino acid in a peptide sequence. This review aims to outline the current state of technology that can be used to investigate mistranslations and misincorporations whilst framing the pursuit as Misincorporation Proteomics (MiP). The current availability of technologies explored herein is mass spectrometry, sample enrichment/preparation, data analysis techniques, and the hyphenation of approaches. While many of these technologies show potential, our review reveals a need for further development and refinement of approaches is still required.

Keywords: misincorporation; non protein amino acids; post translational modifications

1. Introduction

The "central dogma" of molecular biology suggests that the translation of one gene results in the expression of a single protein [1]. However, translated proteins are known to exist as multiple biological variants or proteoforms [2]. These proteoforms are the result of modifications to the polypeptide chain, including the addition, subtraction, or alteration of chemical groups. Such modifications can endow proteoforms with biological activity or an altered function varying from the original proteoform [3]. Any modification that occurs to a proteoform once already translated is termed a post-translational modification (PTM), generating a new and distinct proteoform, adding the advantage of complexity to the proteome [4–6].

Additionally, and importantly for this review, variant proteoforms may be generated by mechanisms other than post-translational modification. During protein translation, an incorrect amino acid may be inserted into the growing peptide chain, resulting in a modification in the final proteoform. Such errors generate new, non-native proteoforms that have the potential to cause harm to the cell [7,8]. The mistranslational error rate reported in *Escherichia coli* (*E. coli*) is between 0.5–5% at any amino acid position [8]. This occurrence of errors in translation during synthesis can generate a new proteoform in the same way that cleavage or PTM can e.g., N-terminal methionine excision [8–10]. However, instead of providing an advantage to the cell, proteoforms produced through mistranslation often present a burden, as they are unpredictably generated with the resulting non-native proteoforms prone to misfolding [11,12]. To overcome this, the cell has sophisticated machinery to identify and degrade these proteoforms [13].

The "misincorporation" of incorrect amino acids into a proteoform need not be limited to the 20 canonical amino acids used in proteoform synthesis. Thousands of synthetic and naturally occurring nonprotein amino acids (NPAAs) exist, also referred to as "nonproteogenic", "noncanonical", "noncoded", or "non-natural" [14]. The infiltration of NPAAs

into the protein translation process adds further opportunity for mistakes to be made during translation, resulting in distinct and unintentionally produced proteoforms. Such misincorporation can cause aberrant modifications to proteoform structure, with biological ramifications for the cell or organism that can perturb normal cellular function. It has been shown that misincorporation of an NPAA can alter the 3D structure of a proteoform, resulting in aggregation [15]. Such misfolding and aggregation of proteoforms is known to be a hallmark of numerous degenerative neurological diseases [16] and for some diseases, an association already exists between exposure to NPAAs and disease development. This includes motor neuron disease (MND) [17,18], multiple sclerosis [19,20] and neuro-lathyrism [21,22]. As such, the exploration of NPAA misincorporation into proteoforms to date has largely focused on their underlying potential to trigger proteinopathies and cause neurodegeneration. To describe the study of NPAA misincorporation into the proteome, we introduce the term Misincorporation Proteomics (MiP), whereby during protein translation, a genetically encoded canonical amino acid is replaced by a NPAA.

2. Amino Acid Misincorporation

Protein translation involves the cognate amino acid (AA) being charged by its appropriate aminoacyl tRNA synthetase (aaRS; Figure 1). The charging is enabled by the hydrolysis of adenosine triphosphate (ATP) and allows the formation of an AA/aaRS/AMP complex. The cognate transfer RNA (tRNA) then binds to this aaRS/AA/AMP complex and an aminoacyl ester bond forms, transferring the aminoacyl group to the tRNA and releasing AMP [23,24].

Transfer of amino acids onto tRNA
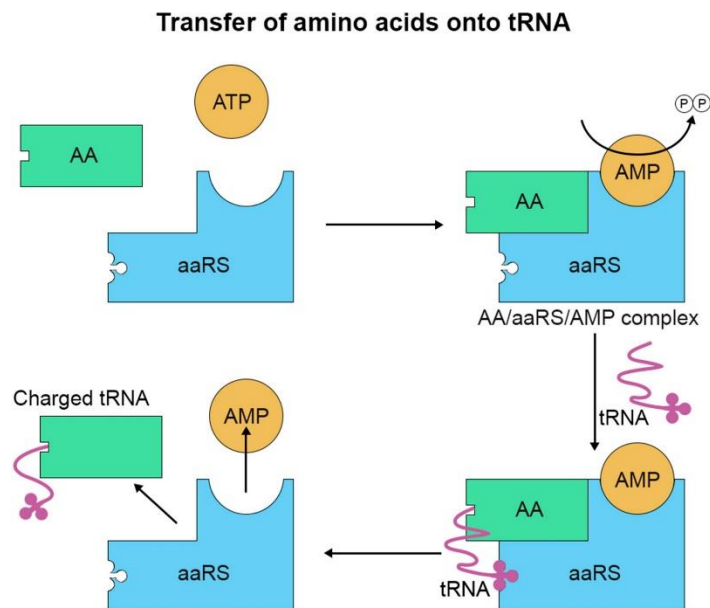
Figure 1. The transfer of amino acids (AA) onto tRNA through tRNA synthetase (aaRS) in the presence of ATP. ATP is converted to AMP activating the aaRS with the AA, an tRNA is then bound to the AA of the AA/aaRS/AMP complex subsequently the now charged tRNA dissociates.

Correct binding of amino acid to its aaRS is dependent on the fit between the amino acid and enzyme binding pocket [25]. If an incorrect amino acid binds the aaRS, this can be removed by a series of proofreading functions known as pre- and post-transfer

editing (Figure 2). Pre-transfer editing occurs before tRNA binds to the charged amino acid. Removal of the charged amino acid has been suggested to occur via various mechanisms, including selective release by the aaRS enzyme, translocation to a separate hydrolytic editing site on the enzyme, or hydrolysis at the primary active site of the aaRS [25,26]. Post-transfer editing occurs after the mischarged amino acid is attached to tRNA and involves the hydrolysis of the ester bond between the two. This can occur via cis-editing at a separate editing domain, by trans-editing factors that resample mischarged amino acid tRNA at the aaRS, or by free trans-editing factors [25,26]. The extent to which pre- and post-transfer editing occurs in human aaRS enzymes is poorly understood, as are the mechanisms by which pre- and post-transfer editing select for misincorporations of NPAAs in particular.
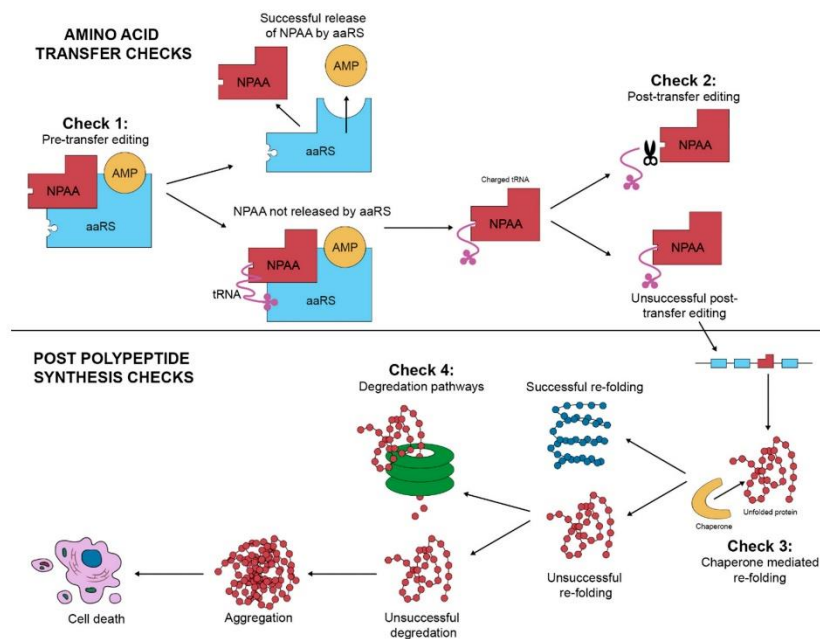
Figure 2. Checks of mistranslation within a cell. AAs are checked for correct charging pre- and/or post-transfer. Misfolded species can be refolded via chaperones or degraded by the various proteolytic pathways; aggregates that are not degraded can lead to cell death.

Despite the presence of checks to avoid incorrect amino acid incorporation, misincorporation can still occur. The misincorporation rate at any amino acid position is quoted as 1/10,000 when considering only the 20 canonical amino acids [9]. However, this rate may be higher when considering NPAAs could be present within the available pool of amino acids. The misincorporation of an NPAA into a growing polypeptide chain through its mischarging onto the aaRS, followed by the failure of the aaRS to remove the mischarged NPAA, allows for the non-native proteoform to be released. Additionally, an NPAA may outcompete the canonical amino acid for binding, especially when there is a higher concentration of NPAA present [15,18,27,28]. Depending on the position of NPAA misincorporation and the canonical amino acid that was replaced, there may be a conformational change in the proteoform, which predisposes it to misfolding and aggregation. Other issues may include replacement of amino acid residues that are essential sites for activity or post-translational modifications (PTM) in that proteoform. Such alterations to the amino

acid sequence can hinder a cells ability to operate, as the proteoforms may be nonfunctional, aggregate, or even acquire a toxic gain of function. An example of the impacts which NPAA misincorporation may cause is the neurological disease multiple sclerosis, the development of which has hypothesised links to misincorporation of the NPAA azetidine-2-carboxylic acid (azetidine), a proline analogue derived from the common agricultural crop sugar beet (beta vulgaris), into myelin basic protein [27]. The consequences posed by the misincorporation of NPAAs may be broad and unpredictable, making it important for the cell to have defences and solutions to this problem.

In addition to the checks during polypeptide synthesis that help to both initially avoid the mischarging of amino acids and to hydrolyse any mischarged amino acids from a tRNA, there are also checks in place following polypeptide synthesis that detect and remove misfolded proteoforms. Chaperones attempt to refold proteoforms to the correct conformational structure and if unsuccessful, degradation pathways such as the proteasome and autophagy are utilised [28]. The failure or bypassing of these checks altogether should trigger cell death. The toxic species of proteoform produced by the misincorporation process may aggregate and self-propagate, forming the basis for proteinopathies and NPAA pathologies [29–31]. As the induction of apoptosis/necrosis occurs when the burden of protein aggregates reaches a threshold, this presents a bigger problem for terminally differentiated cells such as neurons. Such cells cannot share the burden of aggregated proteins amongst daughter cells by dividing and must bear the burden fully, increasing their susceptibility to NPAA pathologies [31]. This may explain why diseases associated with NPAAs tend to be neurological in nature, and why it is of growing importance to investigate misincorporation and its consequences.

3. The Identification of NPAAs Misincorporation

The measurements of NPAA misincorporation to date have primarily relied on inference methodologies, indicating NPAA presence but not location, rather than the direct sequence localisation of a peptide or protein containing a NPAA (Table 1). Studies by Fowden and Richmond [32] in the 1960s were part of the foundational works for the exploration of NPAA misincorporation [33–35]. Here, NPAAs are recognised as structural analogues of canonical amino acids [36] with various toxic effects and as metabolites arising from their mimicry and competition with the canonical amino acids they resemble. This resulted in the term "antimetabolite" being used to describe these toxic NPAAs, with the arginine analogue Canavanine exemplifying this [37].

Various methods were employed during these early studies to investigate the misincorporation of NPAAs. This included the addition of NPAAs and canonical amino acids in different ratios to growing bacterial cultures to observe toxicity and infer competition between NPAA and canonical amino acids [32,38,39]. Such studies offered insight into the similarity of some NPAAs to specific canonical amino acids and therefore an understanding of what pathways would be most affected by the NPAA. Since these early studies, there has been a range of methods used to infer mistranslation and misincorporation of amino acids generally, including NPAAs. Some approaches included studying misincorporation at the tRNA level using radiolabelled amino acids and tRNA microarrays to detect misacylation [40]. Such methods have shown that the NPAA Beta-methyl-amino-L-alanine (BMAA) charges to both alanine and serine tRNAs and bypasses the proof-reading ability of the alanine aaRs, suggesting misincorporation at alanine positions [41,42]. An alternate approach that has proved popular is the use of radiolabelled amino acids, particularly in *E. coli*. Detection of a radioactive signal within isolated protein infers that the labelled amino acid has been incorporated, replacing the canonical amino acid, or is associated with the protein fraction. Examples include detection of radiolabelled cysteine in cysteine-free flagellin in *E. coli* [43]. Similar studies in *E. coli* [44,45] and rabbit reticulocytes [46] have also used this method.

Another method of inferring amino acid misincorporation is to study the effects on the proteoforms themselves if misincorporation occurs. This includes measuring changes in

the isoelectric point of a proteoform due to different amino acid residues being present [47], or restoration of enzymatic activity, including fluorescence, of the resulting proteoform species [48–53]. Other methods include inference of incorporation by an inability to detect free COOH or NH2 groups of the NPAA unless first hydrolysed from protein (dinitrophenyl assay). Alternatively, the use of detectors coupled to chromatographic and ionophoric separation have also been used for the identification of NPAAs from a protein hydrolysate [36]. While these various techniques have provided useful information, many also rely on the use of bacterial systems, limiting their application in mammalian research. Most importantly, all these methods offer only an indirect measurement of misincorporation and cannot definitively characterise the misincorporations of NPAAs. For a review of these techniques, refer to the work of Ribas de Pouplana et al. [54].

There are also direct methods of detecting misincorporation that involve mass spectrometry. This provides the opportunity to localise a NPAA in a peptide sequence and identify misincorporation of incorrect amino acids based on side-chain modifications or the use of a modified database algorithm [55,56]. Identification of NPAAs in the hydrolysate of protein fractions via high-pressure liquid chromatography (HPLC) coupled to either mass spectrometric (MS) (including tandem (MS/MS)) [41], or spectrophotometric detectors (such as ultraviolet-visible spectroscopy (UV-VIS)), is a routine technique. Additionally, there has been an analytical method developed for assaying misincorporation in overexpressed proteins in *E. coli* and yeast called MS-READ, which utilises a genetically modified overexpression model combined with affinity purification [25]. In a similar manner, ESI-MS has been used to study the incorporation of L-3,4-dihydroxyphenylalanine (L-DOPA) in proteins expressed in *E. coli*. However, although a mass shift corresponding to L-DOPA was observed in the protein and tryptic peptides, fragmentation of the trypsin-generated peptides was not performed, preventing localisation of the L-DOPA in the peptide sequence [57]. The use of hydrolysis and detection methods indicating NPAA presence do not provide proteoform sequence information which is essential in studying the effects of NPAAs on biology.

Table 1. Direct and indirect methods used to determine NPAA misincorporation.

Analysis Type	Method of Analysis	Ref.
Direct	-MS mass shift analysis	[57]
	-MS/MS peptide spectral analysis	NPD
	-MS-READ	[25]
Indirect	-Hydrolysed amino acid analysis	[41]
	-Radio-labelled amino acid analysis	[40]
	-Amino acid competition studies	[32]
	-tRNA micro-assay	[40]
	-Proteoform isoelectric point analysis	[47]
	-Enzymatic activity assays	[48]
	-Dinitrophenyl assay	[36]

NPD stands for no published data.

The sensitivity of the methods that can be employed to explore misincorporation varies with the more sensitive methods sacrificing direct site localisation for a decreased limit of detection. These highly sensitive methods include radiolabelling experiments, amino acid analysis of protein hydrolysate, ELISAs and antibody-based microscopy. These methods can be employed alongside exploratory proteomic methods and the usage of targeted mass spectrometry for quantification can also be employed following identification of misincorporation species. When applied to the quantification of the misincorporation species, indirect methods like amino acid hydrolysis have sensitivity in the parts-per-billion (ppb) or pictogram range. Antibody-based detection methods using coupled enzyme-based reporters, such as horseradish peroxidase (HRP), could theoretically detect a single antigen

molecule if enough substrate is converted to a detectable product but detection limits will vary with antibodies and variation of molecular structure around the NPAA site.

As the advancement of proteomics has led to an increase in the ability to detect and characterise canonical amino acid mistranslation, the applications of the techniques have diversified. This review will firstly outline key considerations in the exploratory analysis of MiP, including experimental considerations and biological assumptions that should be understood before utilising certain methods and models. Secondly, a summary of the mass spectrometric technologies that have become available in the pursuit of investigating these misincorporations will be discussed, including targeted global proteomics, data-independent acquisition technologies, precursor ion scanning, and unique instrument system configurations. Considerations for sample processing and potential enrichment strategies that can be employed or require development are introduced before a discussion of the diverse range of pipelines, processes, and data analysis suites originally designed for PTM identification that can be repurposed for MiP. While the bulk of this paper will focus on the ramifications for human samples, many of the techniques discussed are still applicable to a large array of different sample types. From the outset of a study, it is essential to understand that for the generation of high-quality data in MiP, there needs to be enough of the NPAA misincorporation in a proteoform at a specific location prior to protein extraction as analysis methods in proteomics, such as gel electrophoresis and mass spectrometry, are concentration sensitive technologies.

4. Key Considerations in Mistranslation Proteomics

The nature of NPAA misincorporation is random or stochastic and when considering a single site of incorporation, and excluding mechanistic interference, the proportion of proteoforms containing a misincorporation will be several fold less abundant given the highest observed error rates of 10% in canonical amino acid substitution at a single site substitution (Asn->Asp) in *E. coli* [55]. Additionally, detection becomes even more difficult when considering that misincorporation may randomly occur at more than one site in a peptide sequence, resulting in a wider variety of proteoforms, and thus peptidoforms, all at a lower abundance. For example, an incorporation rate of 1 in 1000 with three possible sites would have a 1×10^{-12} abundance in comparison to the unmodified molecule, which would be beyond the limits of detection for the most advanced mass spectrometers. This makes the detection and subsequent investigation of the mistranslated species unlikely without utilising specific extraction and pre-concentration/sample enrichment techniques.

Another important consideration for the study of misincorporation is the suitability of the models (yeast/*E. coli*/mammalian cell lines) and the time frame which is appropriate for sufficient incorporation for detection versus the transient or acute effects on the proteome. The use of animal models could be employed but resource limitations and time constraints are often beyond that of most laboratories. While this review is focused on the study of misincorporation, the investigation into misincorporational effects on the proteomic landscape would also be important, as this could determine the underlying pathology necessary to classify systemic disease pathways (NPAA-pathology). This may help elaborate biological cases arising from exposure to NPAAs and help build upon biomarker classification for individual NPAAs. When considering the complexity of NPAA pathology, studies in cellular models may be limited in their application to human disease processes, especially when perturbation of PTMs that have biological function occur Table 2.

19

Table 2. Amino acids and their biological post translational modifications.

Site of Modification	Letter Symbol	Modification
Alanine	A	Acetylation, methylation, o-linked glycosylation
Arginine	R	Acetylation, ADP-ribosylation, Citrullination, Hydroxylation, Methylation, N linked-glycosylation, Phosphorylation
Asparagine	N	Acetylation, ADP-ribosylation, Citrullination, Hydroxylation, methylation, N-linked-Glycosylation
Aspartic acid	D	Acetylation, Caspase cleavage, Hydroxylation, N-linked glycosylation, Phosphorylation
Cysteine	C	Acetylation, ADP-ribosylation, methylation, Nitration, Oxidation, Palmitoylation, Prenylation, S-nitroylation, Ubiquitination
Glutamic acid	E	Acetylation, Caspase cleavage, Hydroxylation, N-linked glycosylation, Phosphorylation
Glutamine	Q	Methylation
Glycine	G	Acetylation, Caspase cleavage, Hydroxylation, N-linked glycosylation, Phosphorylation
Histidine	H	Acetylation, Caspase cleavage, Hydroxylation, N-linked glycosylation, Phosphorylation
Isolecine	I	Acetylation, Caspase cleavage, Hydroxylation, N-linked glycosylation, Phosphorylation
Leucine	L	Hydroxylation, methylation
Lysine	K	Acetylation interruption(mass)
Methionine	M	Acetylation, Fromylation, Methylation, oxidation, Ubiquitylation
Phenylalanine	F	Acetylation, Caspase cleavage, Hydroxylation, N-linked glycosylation, Phosphorylation
Proline	P	Acetylation, Hydroxylation, Methylation, O-linked Glycosylation
Serine	S	Acetylation, Methylation, N-linked & O linked Glycosylation, Palmitoylation, Phosphorylation, Prenylation, Sulfation
Threonine	T	Acetylation, Methylation, O-linked glycosylation, Palmitoylation, Phosphorylation, sulfation
Tryptophan	W	Hydroxylation, Methylation, Prenylation, Ubiquitylation
Tyrosine	Y	Acetylation, Hydroxylation, Nitration, O-Linked Glycosylation, Phosphorylation, Sulfation
Valine	V	Acetylation, Hydroxylation

The table provides context for how a misincorporation at a particular site could have large biological impacts in cell function [58].

Detection of NPAA incorporation in humans may only be possible in individuals exposed to a subtoxic intake of a NPAA over a prolonged time of many years i.e., supplementary intake of norvaline in bodybuilders [59], azetidine (from beta vulgaris) in meat and milk consumers [60], and Parkinson's disease patients undergoing L-DOPA therapeutic treatment [15]. However, when compared to cell lines and bacterial models, samples are not as readily available. Choosing overexpression models is not a biologically relevant methodology but it will give an approximate representation of the incorporation rate of NPAAs, and researchers should use caution when drawing inferences between organisms and human disease. The impact on cellular systems in multicellular organisms will be different from that of single cellular models and will not consider the larger repertoire of proteostasis mechanisms.

20

5. Sample Processing and Enrichment

Sample processing and enrichment techniques are critical to ensuring that downstream analytical methods have the optimal chance of detecting low abundance proteoforms, such as those containing NPAAs. All the techniques presented here can increase the chance of identification, however, the success of all MiP experimental approaches will be limited by the amount of the NPAA containing species in the sample. Therefore, most explorations of NPAA's have relied upon inference methods showing their presence but not sequence location via methods including radiolabelling and amino acid assaying of hydrolysed proteins, rather than direct localisation [40–46,48–55,57]. The use of fractionation, enrichment, and depletion are essential, but the caveat is always the potential loss of target proteoform due to the sample processing. As an example, the starting amount of protein used to examine the acetylated proteome was reported as 15mg to assay several thousand peptides [61], whilst typical shotgun proteomic experiments are performed on several micrograms. To this end, quantifying the amount of NPAA using radiolabels or hydrolysate amino acid analysis can infer the amount of starting protein that is required and if the NPAA is indeed present in a sample prior to the MiP workflow employed.

The enrichment of proteoforms containing a NPAA can employ various strategies including; antibodies (pull-downs), metal and chemical affinities (chromatography), or fractionation that concentrates targeted species. Antibody-based affinity enrichment is often the first choice of enrichment for PTM studies [62] and requires a highly specific antibody. Antibodies are time-consuming and expensive to create, especially for small discrete chemical groups such as an amino acid side chain [63,64], therefore manufacture requires considerable supporting evidence of specificity. Enrichment strategies such as metal affinity as used in phosphopeptide enrichment may enrich an NPAA containing species, although its application will be limited to those NPAAs whose chemistry is amenable to binding.

Fractionation methods can help detect low abundance proteoforms or peptides and the methodologies generally applied are based on molecular weight, charge, and hydrophobicity. The aforementioned methods do not specifically target peptides containing NPAAs but may give a better chance at identification by dynamic range reduction and increase in the amount of a specific proteoform or peptidoform. As NPAA incorporation may be implicated in proteinopathies and neurodegeneration, enrichment of NPAA containing proteoforms by isolation of insoluble proteoforms from cell lysates is an attractive selection method that has been employed for such protein aggregates as β-amyloid and Tau [65].

Approaches to create sufficient levels of NPAA-containing proteins for detection include the use of in vitro systems whereby protein production can be controlled, and high levels of NPAAs introduced. The limitation with such approaches is that the organism level disease pathology is not seen. An example of these methods is MS-READ [25] to study NPAA incorporation into over-expressed GFP-elastin protein in *E. coli* or yeast. This methodology is quoted as providing quantitation down to the 1/10,000 mistranslational rate for canonical and NPAA species, applying the use of affinity columns to the GFP-elastin protein and then chromatographic separation of all the captured species. This model can establish a baseline of NPAA incorporation, but a mammalian cell line used in the same manner will provide a closer estimate of the rate of incorporation specific to human tRNA binding/proofreading. The NPAA BMAA has been investigated by this method alongside tRNA binding studies that implicated incorporation in place of alanine (contrary to past research identifying the site as serine [29]), with the product being un-quantifiable [42]. Further complications arise when model systems such as *E. coli* or yeast have alternate biological responses to an NPAA compared to mammalian systems. For example, the response and misincorporation dynamics of BMAA in *E. coli* appears markedly different than in mammalian cells [41], and *E. coli* are reported to have yet unidentified mechanisms of avoiding misincorporation of this NPAA [66]. Quantitative analysis of NPAA-modified versus native proteoforms is not currently possible as the ionisation and chromatographic profiles (retention times) between the two peptide or

proteoform species will differ. The amount of the modified proteoform can be inferred from a loss of intensity in the unmodified species relative to the abundance of the other proteotypic peptides assigned to the open reading frame, as shown in Figure 3. The synthetic modification of all species is possible in certain instances where the chemistry is simple, such as the addition of a hydroxyl group to Tyrosine to form the NPAA L-DOPA with the subsequent use of heavy labelled oxygen supplied via peroxide to allow relative comparison. This method has been applied for methionine oxidation [67]. In the same manner, the synthesis of NPAA-containing peptides could be used to quantify the amount if the target species is known. This approach has been applied to the study of the entire human proteome by the Kuster lab and JPT peptides, who have synthesised all human peptide sequences and acquired LC-MS/MS orbitrap data to form a pan-human spectral library and the tool PROSIT [68].
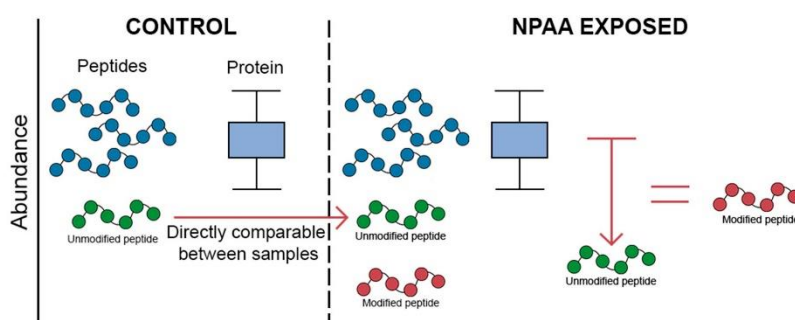
Figure 3. How to compare mistranslated sequence abundances. Within a "Control" vs. "Treated" sample it is only possible to estimate the relative amount of the modified peptide species by a concurrent decrease in the amount of the unmodified peptide. This in essence is a decrease in the proteoform-typic peptide in comparison to the native proteoform-peptides.

Abundance is only directly comparable between identical species. Conversely, the ionisation efficiency of the modified sequence will not reflect the missing amount of unmodified peptide. Thus, only a decline in the abundance of the unmodified peptide can be quantified accurately rather than the increased abundance of the NPAA-containing peptide.

The reader is cautioned against the use of peptide labelling procedures used for relative quantification by multiplexing samples, i.e., Tandem Mass Tags (TMT) [69] and Isobaric tags for relative and absolute quantitation (iTRAQ), in the identification of NPAA containing species [70]. The NPAA containing peptides do not exist in a control sample and by combining multiple proteomes the signal of these unique species will be diluted. This may be why Beri et al. [71] were unable to identify BMAA-containing peptides when searched as a serine modification, however recent data suggests that the modification may be in place of alanine [72].

It is important to note that if starting material is scarce, that sample processing techniques requiring higher starting material should not be used. As such, the development of specific sample processing techniques for studying MiP is equally as important as the mass spectrometric technologies employed.

6. Mass Spectrometric Technologies

The current state of the art in proteomic technologies has lowered the amount of sample required to less than a microgram for a comprehensive shotgun proteomics analysis, however the linear dynamic range has stretched marginally to approximately four-five orders of magnitude in orbitrap mass spectrometers [72]. There are many ways to increase sensitivity to identify and selectively sequence these lower abundance ions. For confident

spectral identifications and resolving of parent mass ions, a High-Resolution instrument like that of an FT-ICR/orbitrap or high-resolution Quadrupole-Time of Flight (Q-TOF) should be used.

Generally, the peptides containing NPAAs are orders of magnitude lower in abundance in comparison to the total proteome. As such without sample enrichment strategies, novel instrument operation techniques, correct ion dissociation techniques, and the hardware being able to intelligently operate, identification is simply not feasible. Herein, the technologies concerning the mass spectrometer that can increase the identification of the low abundance misincorporations will be covered: Mass spectrometer base requirements; Untargeted (either Data-Dependent Analysis (DDA) or Data Independent Acquisition (DIA)), Targeted single reaction monitoring (SRM) and multiple or parallel reaction monitoring (MRM/PRM), hybrid methods, and dynamic range optimisations.

7. Mass Spectrometer Base Requirements and Desirable Features

The advent of high-resolution accurate mass (HRAM) instruments has enabled the discrimination of low abundance ions from background noise especially in the instances of FT-ICR/FT-orbitrap. The higher resolution settings on the orbitrap of 240,000–1,000,000 FWHM can determine the exact elemental composition of an ion. When applying this to peptide sequences, it becomes feasible to accurately identify not only the sequence but also other chemical properties of the peptides. For example, the characteristic fragmentation patterns under various energies and fragmentation types, revealing the presence of PTMs.

In Orbitrap instruments, this high resolution is usually sacrificed in MS2 acquisition for speed as 17,500 resolution is used in standard DDA methods, in conjunction with the high-resolution MS1 scan, allowing an accurate assignment of fragments to amino acid sequence. Within the isolation window for precursor transmission accumulation and fragmentation (1.4 Da in an orbitrap ($\pm$ 0.7 Da)) there is a potential that the fragmented species may be derived from more than one single ion population given the likelihood of similar m/z ion species being found within the isolation window selected for fragmentation. This produces chimeric fragmentation spectra that may confuse the correct assignment of the precursor's structure, although modern peak-picking algorithms somewhat overcome this. This is critical when an investigation of MiP requires accurate assignment. A way to further overcome this issue is BOXCAR which is explored below [73].

8. Data Dependent Analysis

Current untargeted exploration of the proteome by LC-MS/MS involves the use of Data-Dependent Acquisition (DDA), a method of operating the mass spectrometer to sequentially select from an initial precursor scan of the most abundant precursor species for fragmentation. This mode of operation is an important foundation for data generation as each MS/MS spectrum should be derived from a single precursor (excluding the aforementioned chimeric spectra). However, DDA favours sequencing of the high abundance peptides in the sample, precluding the detection of low abundance species without prior offline enrichment, depletion, or fractionation. However, there are several ways to increase the depth of observed ions by the addition of inclusion/exclusion lists to detect lower abundance ions, but the generation of these lists requires prior knowledge of the sample and NPAA containing species.

Due to the nature of ion accumulation and space charging effects, only a certain number of ions can be measured in the MS1 scan and lower abundance ions are masked by the presence of higher abundance ions [74,75]. The way in which one can identify low abundance species, particularly of a NPAA containing peptide, is to limit the precursor mass range using gas-phase fractionation to break up the ion current into smaller ranges [76]. This would turn a traditional acquisition using a 300–1500 m/z range into 2–7 mass ranges, which are subsequently combined bioinformatically.

A more robust approach to sampling the ion current and limiting the dynamic range transmitted is the usage of BOXCAR. Developed by the Mann lab, this can increase the

23

dynamic range ten-fold for the detection of ions [77]. This methodology uses a precursor ion scan to survey the ion species distribution across the mass range, with the mass range subsequently divided into windows with each having the same ion auto gain control (AGC) target. The number of these BOXCAR scans can be increased to further increase the number of lower abundance ions that can be accumulated by narrowing the window while keeping the same AGC target (which in our experience is three). This method is excellent for finding low abundant species but does not increase the ion accumulation of low abundance species for a MS/MS scan, increasing the probability of their identification. Furthermore, the BOXCAR scans occupy a significant amount of duty cycle time, limiting the number of precursor ions that can be fragmented. Jenkins and Orsburn's preprint on BOXCAR-assisted mass fragmentation (BAMF) offers key insight into how partial ion stream sampling can increase the depth of identification when combined with fragmentation [75]. The benefit of low abundance ion identification is that it allows larger spectral libraries to be created and queried. A realised version of this methodology was also developed by the Gygi lab and similarly limits transmission by using parallel-notched waveform isolation [78].

The incorporation of targeting can be used concurrently in a DDA experiment and added to the mass spectrometer's cycle time. The usage of targeting lists requires prior knowledge of the peptide species that could contain an NPAA, and this can be generated in silico by taking currently identified species of highly abundant and confidently matched peptides, then adding the mass difference that incorporation would theoretically produce to create a theoretical precursor mass for targeting. However, without knowledge of the chromatographic elution profile, the target list cannot be scheduled, and thus the length of the list must be restricted to manage duty cycle with expected chromatographic peak width to ensure sufficient measurements across the peak, as well as differentiating isobaric peptides being detected at differing retention times over the duration of chromatographic separation. This issue of ions similar to the targeted precursors being detected during the chromatographic separation can be overcome by using the targeting mode within MaxQuant Live [79] that can dynamically warp global targeting lists. Dynamic warping of targeting lists combats retention time drift that is apparent in intra sample injections as the gradient conditions may have been slightly altered causing drift and a potential loss of the targets during a separation. There is also a function built-in called BatMode enabling fragmentation of the targeted precursor mass during every duty cycle for the predicted elution window. This method utilises mass inclusion lists for fragmentation which can be implemented on any Q-Exactive mass spectrometer that can perform DDA and handle targeting schemes, which in turn decreases the limits of detection.

9. Data Independent Acquisition

Data independent proteomic approaches offer an increased depth of quantitative analysis with the ability to generate a permanent digital ion record of the entire sample being analysed [80]. The drawbacks of this technology are the reliance on generated spectral libraries from DDA sample analysis, which are necessary for searching, and the computational overhead required to generate them. For human samples, a synthetic library of peptides that have been fragmented in an orbitrap instrument has been used to create the PROSIT tool [68]. This allows the researcher to download a spectral library based on a FASTA file provided and align it to the retention times of the instrument. This increases the identification rate by using "match between runs" in various software and increases the richness of the data acquired as exemplified by studies on the PTM phosphorylation [81].

Unfortunately, NPAA containing peptides are not present in the various databases of PROSIT and SRM/human peptide atlas. Acquiring data in a DIA manner will allow retrospective interrogation for these peptides, with a caveat that the amount of NPAA-containing peptides present for analysis must be above the lower limits of detection. The ways to overcome the lack of theoretical data available to predict the effect of NPAA incorporations on produced MS/MS spectra is through the use of tools (such as DeepLC [82]) that enable the prediction of retention times and elution windows for these unknown

species. This allows smarter scheduling and the ability to potentially target tens of thousands of precursors in a single injection. In the data analyses section of this manuscript, the tools that can be used for in silico mining will be covered.

10. Immonium Ion and Precursor Ion Scanning

The use of precursor ion scanning in combination with stable isotope labelled amino acids was developed by Purcell and Williamson [83]. This technique relies upon a diagnostic immonium fragment ion that is unique to an amino acid. In the classical method, stable isotope labelled analogues are used in cellular treatment and produce mass shifted product ions. In this instance, it relies upon the NPAA to produce a unique product ion such as an immonium ion (a fragment resulting from the amino acid side chain). If an immonium ion is formed from the NPAA of interest, it is then feasible to use a stable isotope containing NPAA, such as N-15, during treatment to produce this "diagnostic ion". The size of certain immonium ions may be below the scanning range in most general shotgun proteomic experiments, which may require specific operational parameters and precludes the use of certain instruments.

11. Ion Mobility Mass Spectrometry

A highly desirable feature for the analysis of MiPs and even PTMs is ion mobility separation devices (IMS) available on several state-of-the-art mass spectrometric platforms, with most major vendors offering a variation. For a review on the different types of IMS please see: [84]. For further reading, see [85–87] and for example of novel usage, refer to [88]. In theory, NPAA containing sequence isomers could have the same m/z and potentially same elution time in chromatography producing chimeric fragmentation spectra and being indistinguishable in traditional proteomic LC-MS/MS methods. IMS offers the ability to further separate ions using the third dimension of Collisional Cross Section (CCS) to separate isobaric, co-eluting peptides. While there are several variants of IMS, there are three variants commonly used with MS/MS that would be best suited: traveling wave ion mobility spectrometry (TWIMS), drift-time ion mobility spectrometry (DTIMS), and trapped ion mobility spectrometry (TIMS). TWIMS and DTMS determine mobility via how the ions drift in a cell filled with gas, with the only difference being the application of voltage in DTMS. TIMS determines mobility by trapping the ions in place and having gas pass through the cell. IMS may overcome chimeric spectra production and distinguishing precursors and products [74] but its immediate usage is to increase the depth of analysis of a sample by fractionating the ion current. Implementing any of the approaches mentioned within this manuscript in combination with IMS could provide high-quality empirical data for the incorporation of NPAAs.

12. When Is an Incorporation Real?

When performing PTM or MiP analysis, it is important to establish parameters for correct positive identification of a modified spectra that are sensible. The requirements are a statistically significant peptide sequence identification, based on robust and established statistical analysis and false discovery rate determination [11,89–94]. Positional fragments for the NPAA in an MS/MS spectrum are mandatory for site localisation, similar to the Ascore developed to account for sequence isomers in phosphoproteomic experiments [95]. On the rare occasion that sequence isomers elute together and form a chimeric spectrum, it may not be possible to assign a localisation for the site of misincorporation. The next level of identification should be that of a spectrum matching the unmodified peptide existing within the sample, showing that the peptide belongs to the parent protein or open reading frame. The prime consideration should be the peptide's presence exclusively in the conditions where misincorporation can occur otherwise, an identification is likely a false positive. As the search space is increased to allow multiple variable modifications on a single peptide, the chances of a forced match/false positive occurring become statistically likely.

25

13. Data Analysis Techniques

Once the acquisition of data has been performed, the most time-intensive part of a proteomic experiment begins, that of data analysis. A typical proteomics experiment will employ packages or pipelines that will perform the general required processing, precursor/fragment extraction, centroiding, and mass recalibration. The spectra will then either be searched directly against an in-silico generated tryptic digest database created from the FASTA files for the relevant organism/s and Peptide-Spectrum Matches (PSMs) validated through statistical models or through de novo sequencing of the spectra performed with or without reference to a database.

The repurposing of proteomic tools to specifically explore MiP involves the mass shift of the NPAA incorporation being known at each position of incorporation. Tools that can consider the NPAA as a "PTM" by indicating it as a variable modification in traditional database searching is possible if incorporation is at a high enough abundance to produce an MS/MS spectrum of sufficient quality and the substitution position is correctly characterised. These mass shifts for literarily relevant NPAA incorporations are listed in Table 3, and for further reading on NPAAs, see [96–98]. Traditional database searching is heavily reliant on the database using a known variable modification to identify NPAA species and as such, a method known as "open" searching is recommended for MiP exploration as it allows an exploration of the NPAA's presence at multiple sites and positions in an unbiased manner. The underlying flaw of traditional database searching is still not resolved using spectral libraries in DIA. For identification of a NPAA containing species to be possible in DIA the reference database still must have identified the peptide, it is foreseeable that bioinformatically this problem could be addressed by the assignment of unassigned, co-eluting set of fragments that perfectly align in retention time in relation to the identified unmodified species.

Table 3. Toxic NPAAs, their canonical amino acid homologue, disease, and mass shift for incorporation.

NPAA	Homologue	Mass Shift	Theoretical Immonium Ion (*m/z*)	Disease Associated with Toxicity	Source
BMAA	Serine	+13.0316	73.0766	MND [99] AD/PD [100]	Cycad palms and cyanobacteria [101,102]
	Alanine	+29.0266			
L-DOPA *	Tyrosine	+15.9949	152.0712	PD [15]	Velvet bean plant (*Mucuna pruriens*) [103]
	Phenylalanine	+31.9898			
Meta-tyrosine *	Tyrosine	NMS	136.0762	AD/PD [104]	Fescue grass (*Festuca* spp.) [105]
	Phenylalanine	+15.9949			
Ortho-tyrosine *	Tyrosine	NMS	136.0762	AD/PD, marker of Asctheleroisis [104]	Oxidation product of phenylalanine
	Phenylalanine	+15.9949			
Norvaline	Leucine	−14.0157	72.08132	Cytotoxic [61]	Nutritional supplement [61]
	Isoleucine				
	Valine	NMS			
Azetidine 2 carboxylic acid	Proline	−14.0157	56.05002	MS [19]	Lily of the valley (Convallaria majalis); Garden beet (Beta vulgaris) [9]
Canavanine	Arginine	+1.9793	131.0933	MS/Systemic lupus erythematosus [106]	Jack bean plant (*Canavalia ensiformis*) [107]
Norleucine	Methionine	−17.9564	86.09697	NNAD	Bacteria
Mimosine	Tyrosine	+16.9902	153.0664	NNAD	Leucaena (*Leucaena* spp.) and some *Mimosa* species [108]

NMS stands for no mass shift. NNAD stands for no named associated disease and * denotes NPAAs that are formed via oxidation of a proteogenic AA. Calculated theoretical immonium ions were assumed to be in positive ionisation mode.

26

Several programs that have been developed that can perform open searching on datasets. These programs have developed ways to restrict the computational search space considered to localise PTMs in unmatched spectra. The main examples include; Fragpipe [109], MetaMorpheus [91], Open-pfind [110], tagGraph [111], Peaks Studio [112], Proteome Discoverer [113], Byonic [114], MaxQuant [92], for more examples see Table 4. These programs are essential to the exploration of NPAA localisation within the proteome. As an example, we will cover a method that has been developed for systematic detection of amino acid substitutions that does not rely on the use of a genetically altered or modified cellular expression system [8,25,42] and does not require the use of stable isotope labelling [115].

Table 4. Tools for proteomic analysis that can be used in the pursuit of MiP.

Program	GUI	Cost	Open Searches	Accessibility	DIA Searching	Paper
Byonic	Yes	Licensed	Yes	Easy	No	[114]
EncyclopeDIA	Yes	Free	No	Easy	Exclusively	[116]
Fragpipe *Msfragger, Philosopher, PTMShepard*	Yes	Free to academics	Yes	Easy	Yes	[109]
Galaxy P	Yes	Free	No	Intermediary	Being implemented	[117]
Mascot	Yes	Licensed	No	Easy	No	[118]
MassIVE	Yes	Free	Yes	Easy	Yes	[119]
Maxquant *Andromeda*	Yes	Free	Yes, Dependent peptides	Intermediary	No	[92]
MetaMorpheus	Yes	Free	Yes	Intermediary	No	[91]
OpenMS	Yes	Free	Yes *	Advanced	Yes	[120]
Open-pFind	Yes	Free Licensed	Yes	Easy	No	[110]
Peaks Studio	Yes	Licensed	Yes	Easy	Yes	[121]
Protein Pilot, PeakView	Yes	Licensed	Yes: Protein Pilot	Easy	Yes: PeakView	[122]
Proteome Discoverer	Yes	Licensed	Through Nodes	Easy	Yes	[123]
R workflows *	Yes	Free	Yes	Advanced	Yes	[124]
Skyline Signal quantification (DIA, MRM, PRM)	Yes	Free	N/A	Intermediary	No	[125]
SpectroMine *PTM Shepard*	Yes	Licensed	Yes	Easy	No	[126]
Spectronaut	Yes	Licensed	No	Easy	Yes	[127]
TagGraph	No	Free	Yes	Advanced	No	[111]
Trans-Proteomic Pipeline *PTMProphet*	Yes	Free	Yes	Advanced	No	[128]

GUI: Graphical User Interface. * Consultation of Bioconductor resources is recommended for a grounding in R implemented workflows. Engine/algorithm names are italicised.

The methodology developed by Mordret et al. [11] relies upon a "blind modification" search and repurposes the MaxQuant "dependent peptides" search to explore amino acid substitutions [129]. This methodology could be repurposed in a NPAA exposure model with subsequent identification of incorporation, given a large enough starting sample with known NPAA protein association. This methodology employed two forms of fractionation, solubility and strong cation exchange, which can be routinely performed by most laboratories. The peptide search relies on the principle that the modified peptide sequence will be of lower abundance compared to the identified unmodified counterpart, the peptides that are assigned to known modifications from the Unimod database are

filtered out and peptides remaining unexplained are considered for NPAA incorporation. Within Mordret's paper they were also able to align chromatographic elution profiles for substituted amino acids which could be used to determine the elution profiles of NPAA species. This paper provides the best pipeline for analysing NPAA containing proteomes and is highly recommended for the investigation of the MiPome.

14. Conclusions, a Future Direction and a Best Practice for MiP

Outlined throughout this paper are ways to increase the identification of NPAA containing species and have framed this as mistranslational proteomics or MiP. Key consideration should be given to the sample selection and model creation that offer the highest level of misincorporation, and it is advised that the investigating lab employs amino acid analysis on the proteome prior to committing resources to proteome analysis. The next step in a workflow is to employ an enrichment method or fractionation to further reduce the dynamic range of protein abundance. It is noted here that development of a specific enrichment method may be required, either chemical or an antibody-affinity based. Investigators are also cautioned against the use of peptide labelling tags for use in NPAA peptide identification.

It was outlined that operation of a mass spectrometer that can transmit a smaller mass range of ions to increase the dynamic range, such as that of BOXCAR or parallel-notched wave form isolation, will enhance the depth of identification and that IMS can be used to further separate out NPAA containing species from higher abundance peptides and separate chimera producing sequence isomers. The data analysis workflow is the key to whether a NPAA containing species will be identified, and several programs are listed that can perform open searches that can be used to explore the MiPome. Furthermore, a complete pipeline was identified that can be employed for the investigation of NPAA proteoforms systematically without the requirement of genetically modified models or use of expensive stable isotopic reagents.

The further progression of the entire field of proteomics towards the study of proteoforms will continue to increase the ability to detect and quantify the MiPome, which will provide insight into how disease related to NPAA incorporation progresses and infer potential treatments.

References

1. Crick, F. Central dogma of molecular biology. *Nature* **1970**, *227*, 561–563. [CrossRef] [PubMed]
2. Smith, L.M.; Kelleher, N.L.; Consortium for Top Down, P. Proteoform: A single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186–187. [CrossRef] [PubMed]
3. Aebersold, R.; Agar, J.N.; Amster, I.J.; Baker, M.S.; Bertozzi, C.R.; Boja, E.S.; Costello, C.E.; Cravatt, B.F.; Fenselau, C.; Garcia, B.A.; et al. How many human proteoforms are there? *Nat. Chem. Biol.* **2018**, *14*, 206–214. [CrossRef] [PubMed]
4. Shliaha, P.V.; Gorshkov, V.; Kovalchuk, S.I.; Schwammle, V.; Baird, M.A.; Shvartsburg, A.A.; Jensen, O.N. Middle-Down Proteomic Analyses with Ion Mobility Separations of Endogenous Isomeric Proteoforms. *Anal. Chem.* **2020**, *92*, 2364–2368. [CrossRef] [PubMed]

5. Thygesen, C.; Boll, I.; Finsen, B.; Modzel, M.; Larsen, M.R. Characterizing disease-associated changes in post-translational modifications by mass spectrometry. *Expert Rev. Proteom.* **2018**, *15*, 245–258. [CrossRef]
6. Berry, I.J.; Steele, J.R.; Padula, M.P.; Djordjevic, S.P. The application of terminomics for the identification of protein start sites and proteoforms in bacteria. *Proteomics* **2016**, *16*, 257–272. [CrossRef]
7. Mohler, K.; Ibba, M. Translational fidelity and mistranslation in the cellular response to stress. *Nat. Microbiol.* **2017**, *2*, 17117. [CrossRef]
8. Garofalo, R.; Wohlgemuth, I.; Pearson, M.; Lenz, C.; Urlaub, H.; Rodnina, M.V. Broad range of missense error frequencies in cellular proteins. *Nucleic. Acids Res.* **2019**, *47*, 2932–2945. [CrossRef]
9. Drummond, D.A.; Wilke, C.O. The evolutionary consequences of erroneous protein synthesis. *Nat. Rev. Genet.* **2009**, *10*, 715–724. [CrossRef]
10. Jarocki, V.M.; Steele, J.R.; Widjaja, M.; Tacchi, J.L.; Padula, M.P.; Djordjevic, S.P. Formylated N-terminal methionine is absent from the Mycoplasma hyopneumoniae proteome: Implications for translation initiation. *Int. J. Med. Microbiol.* **2019**, *309*, 288–298. [CrossRef]
11. Mordret, E.; Dahan, O.; Asraf, O.; Rak, R.; Yehonadav, A.; Barnabas, G.D.; Cox, J.; Geiger, T.; Lindner, A.B.; Pilpel, Y. Systematic Detection of Amino Acid Substitutions in Proteomes Reveals Mechanistic Basis of Ribosome Errors and Selection for Translation Fidelity. *Mol. Cell* **2019**, *75*, 427–441e425. [CrossRef] [PubMed]
12. Drummond, D.A.; Wilke, C.O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **2008**, *134*, 341–352. [CrossRef] [PubMed]
13. Metcalf, D.J.; Garcia-Arencibia, M.; Hochfeld, W.E.; Rubinsztein, D.C. Autophagy and misfolded proteins in neurodegeneration. *Exp. Neurol.* **2012**, *238*, 22–28. [CrossRef] [PubMed]
14. Bell, E.A. Nonprotein amino acids of plants: Significance in medicine, nutrition, and agriculture. *J. Agric. Food Chem.* **2003**, *51*, 2854–2865. [CrossRef]
15. Chan, S.W.; Dunlop, R.A.; Rowe, A.; Double, K.L.; Rodgers, K.J. l-DOPA is incorporated into brain proteins of patients treated for Parkinson's disease, inducing toxicity in human neuroblastoma cells in vitro. *Exp. Neurol.* **2012**, *238*, 29–37. [CrossRef]
16. Ross, C.A.; Poirier, M.A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **2004**, *10*, S10–S17. [CrossRef]
17. Spencer, P.S.; Nunn, P.B.; Hugon, J.; Ludolph, A.C.; Ross, S.M.; Roy, D.N.; Robertson, R.C. Guam Amyotrophic-Lateral-Sclerosis Parkinsonism Dementia Linked to a Plant Excitant Neurotoxin. *Science* **1987**, *237*, 517–522. [CrossRef]
18. Cox, P.A.; Banack, S.A.; Murch, S.J. Biomagnification of cyanobacterial neurotoxins and neurodegenerative disease among the Chamorro people of Guam. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13380–13383. [CrossRef]
19. Rubenstein, E. Misincorporation of the Proline Analog Azetidine-2-Carboxylic Acid in the Pathogenesis of Multiple Sclerosis: A Hypothesis. *J. Neuropathol. Exp. Neurol.* **2008**, *67*, 1035–1040. [CrossRef]
20. Bessonov, K.; Bamm, V.V.; Harauz, G. Misincorporation of the proline homologue Aze (azetidine-2-carboxylic acid) into recombinant myelin basic protein. *Phytochemistry* **2010**, *71*, 502–507. [CrossRef]
21. Ravindranath, V. Neurolathyrism: Mitochondrial dysfunction in excitotoxicity mediated by L-beta-oxalyl aminoalanine. *Neurochem. Int.* **2002**, *40*, 505–509. [CrossRef]
22. Yan, Z.Y.; Spencer, P.S.; Li, Z.X.; Liang, Y.M.; Wang, Y.F.; Wang, C.Y.; Li, F.M. Lathyrus sativus (grass pea) and its neurotoxin ODAP. *Phytochemistry* **2006**, *67*, 107–121. [CrossRef] [PubMed]
23. Rajendran, V.; Kalita, P.; Shukla, H.; Kumar, A.; Tripathi, T. Aminoacyl-tRNA synthetases: Structure, function, and drug discovery. *Int. J. Biol. Macromol.* **2018**, *111*, 400–414. [CrossRef] [PubMed]
24. Bullwinkle, T.; Lazazzera, B.; Ibba, M. Quality Control and Infiltration of Translation by Amino Acids Outside of the Genetic Code. *Annu. Rev. Genet.* **2014**, *48*, 149–166. [CrossRef] [PubMed]
25. Mohler, K.; Aerni, H.R.; Gassaway, B.; Ling, J.; Ibba, M.; Rinehart, J. MS-READ: Quantitative measurement of amino acid incorporation. *Biochim. Biophys. Acta Gen. Subj.* **2017**, *1861*, 3081–3088. [CrossRef]
26. Ling, J.; Reynolds, N.; Ibba, M. Aminoacyl-tRNA synthesis and translational quality control. *Annu. Rev. Microbiol.* **2009**, *63*, 61–78. [CrossRef]
27. Giannopoulos, S.; Samardzic, K.; Raymond, B.B.A.; Djordjevic, S.P.; Rodgers, K.J. L-DOPA causes mitochondrial dysfunction in vitro: A novel mechanism of L-DOPA toxicity uncovered. *Int. J. Biochem. Cell Biol.* **2019**, *117*, 105624. [CrossRef]
28. Dunlop, R.A.; Cox, P.A.; Banack, S.A.; Rodgers, K.J. The non-protein amino acid BMAA is misincorporated into human proteins in place of L-serine causing protein misfolding and aggregation. *PLoS ONE* **2013**, *8*, e75376. [CrossRef]
29. Rubenstein, E.; Zhou, H.L.; Krasinska, K.M.; Chien, A.; Becker, C.H. Azetidine-2-carboxylic acid in garden beets (Beta vulgaris). *Phytochemistry* **2006**, *67*, 898–903. [CrossRef]
30. Cox, P.A.; Davis, D.A.; Mash, D.C.; Metcalf, J.S.; Banack, S.A. Dietary exposure to an environmental toxin triggers neurofibrillary tangles and amyloid deposits in the brain. *Proc. Biol. Sci.* **2016**, *283*. [CrossRef]
31. Rodgers, K.J. Non-protein amino acids and neurodegeneration: The enemy within. *Exp. Neurol.* **2014**, *253*, 192–196. [CrossRef] [PubMed]
32. Richmond, M.H. The effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol. Rev.* **1962**, *26*, 398–420. [CrossRef] [PubMed]
33. Fowden, L.; Lewis, D.; Tristram, H. Toxic amino acids: Their action as antimetabolites. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1967**, *29*, 89–163. [CrossRef] [PubMed]

29

34. Fowden, L.; Richmond, M.H. Replacement of proline by azetidine-2-carboxylic acid during biosynthesis of protein. *Biochim. Biophys. Acta* **1963**, *71*, 459–461. [CrossRef]

35. Fowden, L. Nitrogenous compounds and nitrogen metabolism in the Liliaceae. 6. Changes in nitrogenous composition during the growth of Convallaria and Polygonatum. *Biochem. J.* **1959**, *71*, 643–648. [CrossRef]

36. Richmond, M.H. Incorporation of canavanine by Staphylococcus aureus 524 SC. *Biochem. J.* **1959**, *73*, 261–264. [CrossRef]

37. Walker, J.B. Canavanine and homoarginine as antimetabolites of arginine and lysine in yeast and algae. *J. Biol. Chem.* **1955**, *212*, 207–215. [CrossRef]

38. Schwartz, J.H.; Maas, W.K. Analysis of the inhibition of growth produced by canavanine in Escherichia coli. *J. Bacteriol.* **1960**, *79*, 794–799. [CrossRef]

39. Aronson, J.N.; Wermus, G.R. Effects of M-Tyrosine on Growth and Sporulation of Bacillus Species. *J. Bacteriol.* **1965**, *90*, 38–46. [CrossRef]

40. Netzer, N.; Goodenbour, J.M.; David, A.; Dittmar, K.A.; Jones, R.B.; Schneider, J.R.; Boone, D.; Eves, E.M.; Rosner, M.R.; Gibbs, J.S.; et al. Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature* **2009**, *462*, 522–526. [CrossRef]

41. Main, B.J.; Bowling, L.C.; Padula, M.P.; Bishop, D.P.; Mitrovic, S.M.; Guillemin, G.J.; Rodgers, K.J. Detection of the suspected neurotoxin beta-methylamino-l-alanine (BMAA) in cyanobacterial blooms from multiple water bodies in Eastern Australia. *Harmful Algae* **2018**, *74*, 10–18. [CrossRef] [PubMed]

42. Han, N.C.; Bullwinkle, T.J.; Loeb, K.F.; Faull, K.F.; Mohler, K.; Rinehart, J.; Ibba, M. The mechanism of beta-N-methylamino-l-alanine inhibition of tRNA aminoacylation and its impact on misincorporation. *J. Biol. Chem.* **2020**, *295*, 1402–1410. [CrossRef]

43. Edelmann, P.; Gallant, J. Mistranslation in *E. coli*. *Cell* **1977**, *10*, 131–137. [CrossRef]

44. Bouadloun, F.; Donner, D.; Kurland, C.G. Codon-specific missense errors in vivo. *EMBO J.* **1983**, *2*, 1351–1356. [CrossRef]

45. Khazaie, K.; Buchanan, J.H.; Rosenberger, R.F. The accuracy of Qβ RNA translation. *Eur. J. Biochem.* **1984**, *144*, 491–495. [CrossRef]

46. Loftfield, R.B.; Vanderjagt, D. The frequency of errors in protein biosynthesis. *Biochem. J.* **1972**, *128*, 1353–1356. [CrossRef]

47. Parker, J.; Friesen, J.D. "Two out of three" codon reading leading to mistranslation in vivo. *MGG Mol. Gen. Genet.* **1980**, *177*, 439–445. [CrossRef]

48. Toth, M.J.; Murgola, E.J.; Schimmel, P. Evidence for a unique first position codon-anticodon mismatch in vivo. *J. Mol. Biol.* **1988**, *201*, 451–454. [CrossRef]

49. Stansfield, I.; Jones, K.M.; Herbert, P.; Lewendon, A.; Shaw, W.V.; Tuite, M.F. Missense translation errors in Saccharomyces cerevisiae11Edited by K. Nagai. *J. Mol. Biol.* **1998**, *282*, 13–24. [CrossRef]

50. Curran, J.F.; Yarus, M. Base substitutions in the tRNA anticodon arm do not degrade the accuracy of reading frame maintenance. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 6538. [CrossRef]

51. Meyerovich, M.; Mamou, G.; Ben-Yehuda, S. Visualizing high error levels during gene expression in living bacterial cells. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11543–11548. [CrossRef] [PubMed]

52. Kramer, E.B.; Farabaugh, P.J. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* **2007**, *13*, 87–96. [CrossRef] [PubMed]

53. Javid, B.; Sorrentino, F.; Toosky, M.; Zheng, W.; Pinkham, J.T.; Jain, N.; Pan, M.; Deighan, P.; Rubin, E.J. Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1132–1137. [CrossRef] [PubMed]

54. Ribas de Pouplana, L.; Santos, M.A.S.; Zhu, J.-H.; Farabaugh, P.J.; Javid, B. Protein mistranslation: Friend or foe? *Trends Biochem. Sci.* **2014**, *39*, 355–362. [CrossRef]

55. Ruan, B.; Palioura, S.; Sabina, J.; Marvin-Guy, L.; Kochhar, S.; LaRossa, R.A.; Söll, D. Quality control despite mistranslation caused by an ambiguous genetic code. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16502–16507. [CrossRef]

56. Yu, X.C.; Borisov, O.V.; Alvarez, M.; Michels, D.A.; Wang, Y.J.; Ling, V. Identification of Codon-Specific Serine to Asparagine Mistranslation in Recombinant Monoclonal Antibodies by High-Resolution Mass Spectrometry. *Anal. Chem.* **2009**, *81*, 9282–9290. [CrossRef]

57. Ozawa, K.; Headlam, M.J.; Mouradov, D.; Watt, S.J.; Beck, J.L.; Rodgers, K.J.; Dean, R.T.; Huber, T.; Otting, G.; Dixon, N.E. Translational incorporation of L-3,4-dihydroxyphenylalanine into proteins. *FEBS J.* **2005**, *272*, 3162–3171. [CrossRef]

58. Creasy, D.M.; Cottrell, J.S. Unimod: Protein modifications for mass spectrometry. *Proteomics* **2004**, *4*, 1534–1536. [CrossRef]

59. Samardzic, K.; Rodgers, K.J. Cytotoxicity and mitochondrial dysfunction caused by the dietary supplement l-norvaline. *Toxicol. In Vitro* **2019**, *56*, 163–171. [CrossRef]

60. Samardzic, K.; Rodgers, K.J. Cell death and mitochondrial dysfunction induced by the dietary non-proteinogenic amino acid L-azetidine-2-carboxylic acid (Aze). *Amino Acids* **2019**, *51*, 1221–1232. [CrossRef]

61. Svinkina, T.; Gu, H.; Silva, J.C.; Mertins, P.; Qiao, J.; Fereshetian, S.; Jaffe, J.D.; Kuhn, E.; Udeshi, N.D.; Carr, S.A. Deep, Quantitative Coverage of the Lysine Acetylome Using Novel Anti-acetyl-lysine Antibodies and an Optimized Proteomic Workflow. *Mol. Cell. Proteom. MCP* **2015**, *14*, 2429–2440. [CrossRef] [PubMed]

62. Larsen, M.R.; Trelle, M.B.; Thingholm, T.E.; Jensen, O.N. Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques* **2006**, *40*, 790–798. [CrossRef] [PubMed]

63. Hattori, T.; Koide, S. Next-generation antibodies for post-translational modifications. *Curr. Opin. Struct. Biol.* **2018**, *51*, 141–148. [CrossRef] [PubMed]

64. Zhao, Y.; Jensen, O.N. Modification-specific proteomics: Strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* **2009**, *9*, 4632–4641. [CrossRef] [PubMed]

65. Gozal, Y.M.; Duong, D.M.; Gearing, M.; Cheng, D.; Hanfelt, J.J.; Funderburk, C.; Peng, J.; Lah, J.J.; Levey, A.I. Proteomics analysis reveals novel components in the detergent-insoluble subproteome in Alzheimer's disease. *J. Proteome Res.* **2009**, *8*, 5069–5079. [CrossRef] [PubMed]

66. van Onselen, R.; Cook, N.A.; Phelan, R.R.; Downing, T.G. Bacteria do not incorporate β-N-methylamino-l-alanine into their proteins. *Toxicon* **2015**, *102*, 55–61. [CrossRef] [PubMed]

67. Bettinger, J.Q.; Welle, K.A.; Hryhorenko, J.R.; Ghaemmaghami, S. Quantitative Analysis of in Vivo Methionine Oxidation of the Human Proteome. *J. Proteome Res.* **2020**, *19*, 624–633. [CrossRef]

68. Gessulat, S.; Schmidt, T.; Zolg, D.P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; et al. Prosit: Proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **2019**, *16*, 509–518. [CrossRef]

69. Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895–1904. [CrossRef]

70. Aggarwal, K.; Choe, L.H.; Lee, K.H. Shotgun proteomics using the iTRAQ isobaric tags. *Brief. Funct. Genom.* **2006**, *5*, 112–120. [CrossRef]

71. Beri, J.; Nash, T.; Martin, R.M.; Bereman, M.S. Exposure to BMAA mirrors molecular processes linked to neurodegenerative disease. *Proteomics* **2017**, *17*, 10. [CrossRef]

72. Hecht, E.S.; Scigelova, M.; Eliuk, S.; Makarov, A. Fundamentals and Advances of Orbitrap Mass Spectrometry. In *Encyclopedia of Analytical Chemistry*; Wiley: Hoboken, NJ, USA, 2019; pp. 1–40. [CrossRef]

73. Meier, F.; Geyer, P.E.; Virreira Winter, S.; Cox, J.; Mann, M.J.N.M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat. Methods* **2018**, *15*, 440–448. [CrossRef] [PubMed]

74. Tarasova, I.A.; Surin, A.K.; Fornelli, L.; Pridatchenko, M.L.; Suvorina, M.Y.; Gorshkov, M.V. Ion Coalescence in Fourier Transform Mass Spectrometry: Should We Worry about This in Shotgun Proteomics? *Eur. J. Mass Spectrom.* **2015**, *21*, 459–470. [CrossRef] [PubMed]

75. Ledford, E.B.; Rempel, D.L.; Gross, M.L. Space charge effects in Fourier transform mass spectrometry. II. Mass calibration. *Anal. Chem.* **1984**, *56*, 2744–2748. [CrossRef] [PubMed]

76. Pino, L.K.; Just, S.C.; MacCoss, M.J.; Searle, B.C. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *J. Mol. Cell. Proteom.* **2020**. [CrossRef] [PubMed]

77. Jenkins, C.; Orsburn, B. BoxCar Assisted MS Fragmentation (BAMF). *BioRxiv* **2019**, 860858. [CrossRef]

78. Erickson, B.K.; Schweppe, D.K.; Yu, Q.; Rad, R.; Haas, W.; McAlister, G.C.; Gygi, S.P. Parallel Notched Gas-Phase Enrichment for Improved Proteome Identification and Quantification with Fast Spectral Acquisition Rates. *J. Proteome Res.* **2020**, *19*, 2750–2757. [CrossRef]

79. Wichmann, C.; Meier, F.; Virreira Winter, S.; Brunner, A.-D.; Cox, J.; Mann, M. MaxQuant.Live Enables Global Targeting of More Than 25,000 Peptides. *J. Mol. Cell. Proteom.* **2019**, *18*, 982–994. [CrossRef]

80. Searle, B.C.; Swearingen, K.E.; Barnes, C.A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **2020**, *11*, 1548. [CrossRef]

81. Hogrebe, A.; von Stechow, L.; Bekker-Jensen, D.B.; Weinert, B.T.; Kelstrup, C.D.; Olsen, J.V. Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nat. Commun.* **2018**, *9*, 1045. [CrossRef]

82. Bouwmeester, R.; Gabriels, R.; Hulstaert, N.; Martens, L.; Degroeve, S. DeepLC can predict retention times for peptides that carry as-yet unseen modifications. *BioRxiv* **2020**. [CrossRef]

83. Williamson, N.A.; Reilly, C.; Tan, C.T.; Ramarathinam, S.H.; Jones, A.; Hunter, C.L.; Rooney, F.R.; Purcell, A.W. A novel strategy for the targeted analysis of protein and peptide metabolites. *Proteomics* **2011**, *11*, 183–192. [CrossRef]

84. Cumeras, R.; Figueras, E.; Davis, C.E.; Baumbach, J.I.; Gracia, I. Review on ion mobility spectrometry. Part 1: Current instrumentation. *Analyst* **2015**, *140*, 1376–1390. [CrossRef] [PubMed]

85. Ridgeway, M.E.; Lubeck, M.; Jordens, J.; Mann, M.; Park, M.A. Trapped ion mobility spectrometry: A short review. *Int. J. Mass Spectrom.* **2018**, *425*, 22–35. [CrossRef]

86. Winter, D.L.; Wilkins, M.R.; Donald, W.A. Differential Ion Mobility–Mass Spectrometry for Detailed Analysis of the Proteome. *Trends Biotechnol.* **2019**, *37*, 198–213. [CrossRef]

87. Campuzano, I.D.G.; Giles, K. Historical, current and future developments of travelling wave ion mobility mass spectrometry: A personal perspective. *Trac. Trends Anal. Chem.* **2019**, *120*, 115620. [CrossRef]

88. Bekker-Jensen, D.B.; Martínez-Val, A.; Steigerwald, S.; Rüther, P.; Fort, K.L.; Arrey, T.N.; Harder, A.; Makarov, A.; Olsen, J.V. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Mol. Cell. Proteom.* **2020**, *19*, 716–729. [CrossRef] [PubMed]

89. Kong, A.T.; Leprevost, F.V.; Avtonomov, D.M.; Mellacheruvu, D.; Nesvizhskii, A.I. MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **2017**, *14*, 513–520. [CrossRef]

90. Burger, T. Gentle Introduction to the Statistical Foundations of False Discovery Rate in Quantitative Proteomics. *J. Proteome Res.* **2018**, *17*, 12–22. [CrossRef]

91. Solntsev, S.K.; Shortreed, M.R.; Frey, B.L.; Smith, L.M. Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *J. Proteome Res.* **2018**, *17*, 1844–1851. [CrossRef] [PubMed]
92. Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301–2319. [CrossRef] [PubMed]
93. Nesvizhskii, A.I.; Vitek, O.; Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* **2007**, *4*, 787–797. [CrossRef] [PubMed]
94. Ma, K.; Vitek, O.; Nesvizhskii, A.I. A statistical model-building perspective to identification of MS/MS spectra with Peptide-Prophet. *BMC Bioinform.* **2012**, *13*, S1. [CrossRef] [PubMed]
95. Beausoleil, S.A.; Villen, J.; Gerber, S.A.; Rush, J.; Gygi, S.P. A probability-based approach for high-throughput protein phosphory-lation analysis and site localization. *Nat. Biotechnol.* **2006**, *24*, 1285–1292. [CrossRef] [PubMed]
96. Yoganathan, S.; Vederas, J.C. 5.02—Nonprotein l-Amino Acids. In *Comprehensive Natural Products II*; Liu, H.-W., Mander, L., Eds.; Elsevier: Oxford, UK, 2010; pp. 5–70. [CrossRef]
97. Rodgers, K.J.; Samardzic, K.; Main, B.J.J.P.T. Toxic nonprotein amino acids. *Plant Toxins* **2015**, *1*, 1–20.
98. Rodgers, K.J.; Shiozawa, N. Misincorporation of amino acid analogues into proteins by biosynthesis. *Int. J. Biochem. Cell Biol.* **2008**, *40*, 1452–1466. [CrossRef]
99. Torbick, N.; Ziniti, B.; Stommel, E.; Linder, E.; Andrew, A.; Caller, T.; Haney, J.; Bradley, W.; Henegan, P.L.; Shi, X. Assessing Cyanobacterial Harmful Algal Blooms as Risk Factors for Amyotrophic Lateral Sclerosis. *Neurotox. Res.* **2018**, *33*, 199–212. [CrossRef]
100. Silva, D.F.; Candeias, E.; Esteves, A.R.; Magalhães, J.D.; Ferreira, I.L.; Nunes-Costa, D.; Rego, A.C.; Empadinhas, N.; Cardoso, S.M. Microbial BMAA elicits mitochondrial dysfunction, innate immunity activation, and Alzheimer's disease features in cortical neurons. *J. Neuroinflamm.* **2020**, *17*, 332. [CrossRef]
101. Vega, A. α-Amino-β-methylaminopropionic acid, a new amino acid from seeds of Cycas circinalis. *Phytochemistry* **1967**, *6*, 759–762. [CrossRef]
102. Cox, P.A.; Banack, S.A.; Murch, S.J.; Rasmussen, U.; Tien, G.; Bidigare, R.R.; Metcalf, J.S.; Morrison, L.F.; Codd, G.A.; Bergman, B. Diverse taxa of cyanobacteria produce beta-N-methylamino-L-alanine, a neurotoxic amino acid. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5074–5078. [CrossRef]
103. Tomita-Yokotani, K.; Hashimoto, H.; Fujii, Y.; Nakamura, T.; Yamashita, M. Distribution of L-DOPA in the root of velvet bean plant (*Mucuna pruriens* L.) and gravity. *Uchu Seibutsu Kagaku* **2004**, *18*, 165–166. [PubMed]
104. Ipson, B.R.; Fisher, A.L. Roles of the tyrosine isomers meta-tyrosine and ortho-tyrosine in oxidative stress. *Ageing Res. Rev.* **2016**, *27*, 93–107. [CrossRef] [PubMed]
105. Huang, T.; Rehak, L.; Jander, G. meta-Tyrosine in Festuca rubra ssp. commutata (Chewings fescue) is synthesized by hydroxylation of phenylalanine. *Phytochemistry* **2012**, *75*, 60–66. [CrossRef] [PubMed]
106. Constantin, M.-M.; Nita, I.E.; Olteanu, R.; Constantin, T.; Bucur, S.; Matei, C.; Raducan, A. Significance and impact of dietary factors on systemic lupus erythematosus pathogenesis. *Exp. Med.* **2019**, *17*, 1085–1090. [CrossRef]
107. Kitagawa, M.; Tomiyama, T. A new amino-compound in the jack bean and a corresponding new ferment.(I). *J. Biochem.* **1929**, *11*, 265–271. [CrossRef]
108. Rosenthal, G.A. *Plant Nonprotein Amino and Imino Acids: Biological, Biochemical, and Toxicological Properties*; Academic Press: New York, NY, USA, 1982.
109. Yu, F.; Teo, G.C.; Kong, A.T.; Haynes, S.E.; Avtonomov, D.M.; Geiszler, D.J.; Nesvizhskii, A.I. Identification of modified peptides using localization-aware open search. *Nat. Commun.* **2020**, *11*, 4065. [CrossRef]
110. Chi, H.; Liu, C.; Yang, H.; Zeng, W.F.; Wu, L.; Zhou, W.J.; Wang, R.M.; Niu, X.N.; Ding, Y.H.; Zhang, Y.; et al. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat. Biotechnol.* **2018**. [CrossRef]
111. Devabhaktuni, A.; Lin, S.; Zhang, L.; Swaminathan, K.; Gonzalez, C.G.; Olsson, N.; Pearlman, S.M.; Rawson, K.; Elias, J.E. TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. *Nat. Biotechnol.* **2019**, *37*, 469–479. [CrossRef]
112. Tran, N.H.; Rahman, M.Z.; He, L.; Xin, L.; Shan, B.; Li, M. Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* **2016**, *6*, 31730. [CrossRef]
113. Li, Q.; Shortreed, M.R.; Wenger, C.D.; Frey, B.L.; Schaffer, L.V.; Scalf, M.; Smith, L.M. Global Post-Translational Modification Discovery. *J. Proteome Res.* **2017**, *16*, 1383–1390. [CrossRef]
114. Bern, M.; Kil, Y.J.; Becker, C. Byonic: Advanced peptide and protein identification software. *Curr. Protoc. Bioinform.* **2012**. [CrossRef] [PubMed]
115. Cvetesic, N.; Semanjski, M.; Soufi, B.; Krug, K.; Gruic-Sovulj, I.; Macek, B. Proteome-wide measurement of non-canonical bacterial mistranslation by quantitative mass spectrometry of protein modifications. *Sci. Rep.* **2016**, *6*, 28631. [CrossRef] [PubMed]
116. Searle, B.C.; Pino, L.K.; Egertson, J.D.; Ting, Y.S.; Lawrence, R.T.; MacLean, B.X.; Villen, J.; MacCoss, M.J. Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat. Commun.* **2018**, *9*, 5128. [CrossRef] [PubMed]
117. Boekel, J.; Chilton, J.M.; Cooke, I.R.; Horvatovich, P.L.; Jagtap, P.D.; Käll, L.; Lehtiö, J.; Lukasse, P.; Moerland, P.D.; Griffin, T.J. Multi-omic data analysis using Galaxy. *Nat. Biotechnol.* **2015**, *33*, 137–139. [CrossRef]

118. Koenig, T.; Menze, B.H.; Kirchner, M.; Monigatti, F.; Parker, K.C.; Patterson, T.; Steen, J.J.; Hamprecht, F.A.; Steen, H. Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. *J. Proteome Res.* **2008**, *7*, 3708–3717. [CrossRef]

119. Choi, M.; Carver, J.; Chiva, C.; Tzouros, M.; Huang, T.; Tsai, T.H.; Pullman, B.; Bernhardt, O.M.; Huttenhain, R.; Teo, G.C.; et al. MassIVE.quant: A community resource of quantitative mass spectrometry-based proteomics datasets. *Nat. Methods* **2020**. [CrossRef]

120. Röst, H.L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.C.; Gutenbrunner, P.; Kenar, E.; et al. OpenMS: A flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods* **2016**, *13*, 741–748. [CrossRef]

121. Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337–2342. [CrossRef]

122. Sew, Y.S.; Aizat, W.M.; Razak, M.S.F.A.; Zainal-Abidin, R.-A.; Simoh, S.; Abu-Bakar, N. Comprehensive proteomics data on whole rice grain of selected pigmented and non-pigmented rice varieties using SWATH-MS approach. *Data Brief* **2020**, *31*, 105927. [CrossRef]

123. Rinas, A.; Espino, J.A.; Jones, L.M. An efficient quantitation strategy for hydroxyl radical-mediated protein footprinting using Proteome Discoverer. *Anal. Bioanal. Chem.* **2016**, *408*, 3021–3031. [CrossRef]

124. Gatto, L.; Christoforou, A. Using R and Bioconductor for proteomics data analysis. *Biochim. Biophys. Acta Proteins Proteom.* **2014**, *1844*, 42–51. [CrossRef] [PubMed]

125. MacLean, B.; Tomazela, D.M.; Shulman, N.; Chambers, M.; Finney, G.L.; Frewen, B.; Kern, R.; Tabb, D.L.; Liebler, D.C.; MacCoss, M.J. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966–968. [CrossRef] [PubMed]

126. Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O.M.; Treiber, T.; Bruderer, R.; Reiter, L. Surpassing 10000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol. Omics* **2019**, *15*, 348–360. [CrossRef] [PubMed]

127. Bernhardt, O.; Selevsek, N.; Gillet, L.; Rinner, O.; Picotti, P.; Aebersold, R.; Reiter, L. Spectronaut: A Fast and Efficient Algorithm for MRM-Like Processing of Data Independent Acquisition (SWATH-MS) Data. Available online: https://biognosys.com/media.ashx/spectronaut-a-fast-and-efficient-algorithm-for-mrm-like-swath-processing.pdf (accessed on 21 January 2021).

128. Shteynberg, D.D.; Deutsch, E.W.; Campbell, D.S.; Hoopmann, M.R.; Kusebauch, U.; Lee, D.; Mendoza, L.; Midha, M.K.; Sun, Z.; Whetton, A.D.; et al. PTMProphet: Fast and Accurate Mass Modification Localization for the Trans-Proteomic Pipeline. *J. Proteome Res.* **2019**, *18*, 4262–4272. [CrossRef] [PubMed]

129. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [CrossRef] [PubMed]

Chapter Three: A Novel Method for Creating a Synthetic L-DOPA Proteome and *In Vitro* Evidence for Incorporation of L-DOPA into Proteins

Chapter Overview (published manuscript)

The NPAA L-DOPA has been used ubiquitously in Parkinson's treatment and has been highlighted by previous work [57, 193] to be an agent of oxidative stress and cell damage [58, 79, 106-108, 123, 124, 127, 149, 172, 194-197]. Although DOPA is an inherently reactive molecule, the ability of L-DOPA to cause oxidative stress to cells in culture is largely an artefact due to the high oxygen concentrations (~20%) to which the cells are exposed, as well as the presence of metal ions in the culture medium [198]. The ability of L-DOPA to produce a comparable level of oxidative stress *in vivo* to what has been reported in cell studies is unlikely and it is possible that L-DOPA toxicity *in vivo* is not due to its ability to cause oxidative stress due to the high levels of antioxidants and detoxification systems that exist *in vivo*.

Studying the incorporation of L-DOPA into the proteome of PD patients might help us understand the functional ramifications of the presence of DOPA in a cell's proteome. A method was therefore created for the modification of a proteome through conversion of tyrosine residues to L-DOPA. This method offers the potential to artificially create L-DOPA containing proteins in control samples which can then help with the identification of DOPA-containing peptides across patient sample cohorts, allowing an exploratory proteomics approach with a sample specific synthetic peptide collection for targeting and identification transference between runs.

The method described employs the use of the enzyme tyrosinase to convert tyrosine residues in proteins to L-DOPA. In these studies, the method was first trailed on a synthetic peptide (Synthesised proteotypic peptide for the protein UCP-5 from the Payne laboratory, University of Sydney), proving the procedure viable before being used to create whole proteome samples. The initial peptide method did not scale up to full proteome conversion, with the final method adapted from an industrial conversion proof of concept paper [199]. The tyrosine conversion efficiencies achieved were approximately 10% regardless of the enzyme ratio used, but off target oxidation of phenylalanine was greatly reduced when using a higher ratio of tyrosinase to protein.

Furthermore, an *in vitro* L-DOPA treatment experiment using the human neuronal cell line SH-SY5Y was performed demonstrating an increase in the amount of L-DOPA within proteins as well diverse proteome changes. Quantitative analysis demonstrated cellular responses to toxic effects including the activation of the unfolded protein response, oxidative stress and changes in pathways involved in

neurodegenerative diseases. This work addressed the aim of detecting incorporation of L-DOPA and providing evidence for its possible role in contributing to neurodegeneration.

Certificate of authorship and originality

This paper was published in *Proteomes* (MDPI Ltd). I certify that the work presented in this chapter has not previously been submitted as part of the requirements for a degree. I also certify that I carried out the majority of the work presented in this paper.

- Joel Ricky Steele: Wrote the majority (~95%) of the manuscript, performed all experimental work and performed the majority of the data analysis
- Natalie Strange: Provided intellectual input into the experimental approaches, assisted in data analysis, proof reading and editing the manuscript including final proofs.
- Kenneth J. Rodgers and Matthew P. Padula: Proof-read the manuscript and oversaw the project.

Primary and corresponding author

Joel Ricky Steele

Signature

Date: 10th October 2021.

Co-author signatures

MDPI

Article

A Novel Method for Creating a Synthetic L-DOPA Proteome and In Vitro Evidence of Incorporation

Joel Ricky Steele [1,2,*], Natalie Strange [3], Kenneth J. Rodgers [2] and Matthew P. Padula [1]

[1] Proteomics Core Facility and School of Life Sciences, The University of Technology Sydney, Ultimo, NSW 2007, Australia; matthew.padula@uts.edu.au

[2] Neurotoxin Research Group, School of Life Sciences, The University of Technology Sydney, Ultimo, NSW 2007, Australia; kenneth.rodgers@uts.edu.au

[3] School of Life Sciences, The University of Technology Sydney, Ultimo, NSW 2007, Australia; natalie.strange@student.uts.edu.au

[*] Correspondence: joel.steele@uts.edu.au

Abstract: Proteinopathies are protein misfolding diseases that have an underlying factor that affects the conformation of proteoforms. A factor hypothesised to play a role in these diseases is the incorporation of non-protein amino acids into proteins, with a key example being the therapeutic drug levodopa. The presence of levodopa as a protein constituent has been explored in several studies, but it has not been examined in a global proteomic manner. This paper provides a proof-of-concept method for enzymatically creating levodopa-containing proteins using the enzyme tyrosinase and provides spectral evidence of in vitro incorporation in addition to the induction of the unfolded protein response due to levodopa.

1. Introduction

The non-protein amino acid (NPAA) L-3,4-dihydroxyphenylalanine (L-DOPA), commercially known as 'levodopa', is used to restore dopamine levels in the damaged substantia nigra of Parkinson's disease (PD) patients [1] although it has also been hypothesised to accelerate PD pathology and neurodegeneration [2] by causing neuronal toxicity [2–16]. The mechanisms of L-DOPA-induced neurodegeneration include induction of oxidative stress and the misincorporation of L-DOPA into proteins in place of L-tyrosine, resulting in protein misfolding and the formation of proteolysis-resistant aggregates [3,5,7,17–19]. To date, in vitro studies have not examined or identified the proteins containing incorporated L-DOPA and their potential role in disease pathology, with Parkinson's disease datasets being a prime candidate [20].

Protein-bound DOPA (PB-DOPA) can form via several mechanisms, including direct incorporation of free L-DOPA in place of tyrosine into a growing polypeptide chain [21]. Alternatively, L-DOPA can be formed via hydroxyl attack on tyrosine residues by a reactive oxygen species (ROS) [21–25], independent of free L-DOPA. Similarly, hydroxyl radical attack on the phenyl ring of phenylalanine residues forms isomers of tyrosine [22,26], the position of the hydroxyl group of these isomers (meta and ortho) differing from that of the protein amino acid L-tyrosine (para-tyrosine) (Figure 1). It is known that free L-DOPA can induce oxidative stress, resulting in ROS generation via interaction with iron and copper to produce Fenton reaction products [27], which leads to further hydroxylation of free or protein incorporated tyrosine, thereby producing free and PB-DOPA [5,22].
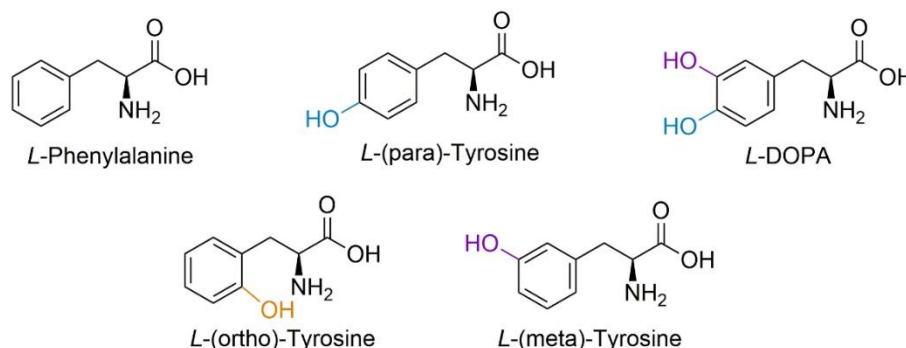
Figure 1. Structural forms of phenylalanine, para-tyrosine, meta-tyrosine, ortho-tyrosine, and L-DOPA. Hydroxyl radical attack on the phenyl ring of phenylalanine residues can form isomers of tyrosine, the canonical form being para-tyrosine, but can also form meta-tyrosine and ortho-tyrosine; hydroxyl attack of tyrosine can also form *L*-DOPA.

The misincorporation rate of canonical amino acids in general is reported to be approximately one in 10,000; this low level of misincorporation makes it difficult to detect DOPA incorporation or other misincorporated amino acids [18,21,28,29]. However, it is possible that misincorporation levels of NPAAs could be higher since they can escape proofreading mechanisms that have evolved to detect the mistaken incorporation of a canonical amino acid. As proteomics utilises mass spectrometers, which are concentration-dependent detectors, methodologies should be applied to enhance the ability to detect these low-abundance species (for a comprehensive review, see Steele et al., 2021 [28]). The most sensitive way to detect low-abundance species in exploratory proteomics is the use of data-independent acquisition (DIA), but this requires prior identification of the peptide species and its fragments for a searchable spectral library, especially in instances of modified peptides [29]. By producing synthetic peptides for subsequent proteomic analysis and library inclusion, it becomes possible to identify very low levels of the modified proteoform. However, the production of such a synthetic library is prohibitively expensive and the libraries currently available do not contain NPAA modifications [30].

The best measure of the global oxidative state of a proteome is the level of methionine oxidation, which occurs via the addition of a single oxygen atom, creating methionine sulfoxide with a mass shift of +15.99 Da, homologous to that of hydroxylation [31,32]. However, protein extraction and preparation can increase the likelihood of forming oxidised methionine and even electrospray ionisation can result in this modification [29,33,34]. Due to the multiple possible end-points of hydroxyl attack, the identification of oxidised phenylalanine may help in determining if PB-DOPA has been formed from incorporation or from hydroxyl attack. Within plant and animal tissues, the conversion of tyrosine to L-DOPA can be carried out by the enzyme tyrosinase, which has also been used in industrial synthesis of L-DOPA and is essential to the production of dopamine in the human brain [33,34]. Unlike oxidation reactions, the enzyme tyrosinase has the capacity to generate peptide-bonded DOPA (from PB-tyrosine conversion) without generating other oxidised amino acid residues such as ortho- and meta-tyrosine.

To overcome the poor detection of low-ion-abundance peptides from PB-DOPA proteoforms, a method for creating a positive control spectral library needs to be produced to enable future application to clinical samples by enabling DIA analyses. A method for generating a positive control spectral library and avoiding the high cost of chemically synthesising a DOPA-containing peptide library is to use the enzyme tyrosinase to specifically modify tyrosine residues to L-DOPA [34,35]. The aim of this work is to establish the use of tyrosinase to convert proteomes, enabling subsequent analyses. Furthermore, we

explore the effects of DOPA on a neuronal cell line and investigate the proteins that contain PB-DOPA following treatment with L-DOPA.

2. Materials and Methods

2.1. Materials

The following reagents were sourced from Sigma-Aldrich (Saint Louis, MI, USA): L-DOPA (Cat #D9628); urea (Cat #U5378); thiourea (Cat #T7875); C7BzO (Cat #C0856); Dulbecco's Modified Eagle Medium (DMEM; Cat#D6429) (New York, NY, USA); cOmplete™ Protease Inhibitor Cocktail (Cat #11697498001) (Roche, USA); tris(2-carboxyethyl)phosphine (TCEP; Cat #C4706); ammonium bicarbonate (Cat #A6141); Empore™ Extraction Disk Cartridge C18 (Cat #66873-U) (Supelco, Bellefonte, PA, USA); acetonitrile hypergrade for LC–MS LiChrosolv® (Cat #1.00029) (Truganina, Australia); triflouroacetic acid (Cat #T6508); tyrosinase (Cat #T3824); boric acid (Cat #1.00165); sodium hyrdoxide (Cat # 221465); ascorbic acid (Cat # A4544); hydroxylamine solution (Cat # 467804). The following reagents were sourced from Thermo Fisher: TrypLE™ Express Enzyme (Cat #12604013) (Gibco, New York, NY, USA); Pierce™ BCA Protein Assay Kit (Cat #23225) (Thermo Scientific, Rockford, IL, USA). The UCP-5 peptide was synthesised by the Payne Laboratory with sequence: EEGVLALYSGIAPALLR.

2.2. Tyrosinase Method Optimisation

The neuroblastoma cell line SH-SY5Y (ATCC #CRL-2266) was utilised to optimise a microscale proteome tyrosinase conversion. Six T175 flasks were treated in Dulbecco's Modified Eagle Medium with 10% v/v fetal bovine serum (FBS) at 37 °C with 5% CO_2. Three flasks were treated with 500 μM L-DOPA-containing media for 24 h; the DOPA addition was from a 3 mM stock in 10 mM HCl with control flasks receiving an equal amount of sterilised HCl solution. Cells were washed three times with PBS, de-adhered using TrypLE (Gibco, New York, NY, USA), pelleted at 1500 rcf and snap frozen in liquid nitrogen.

2.3. Protein Extraction and Processing

Cell pellets were solubilised in UTC7 (8 M urea, 2 M thiourea, 0.1% w/v C7BzO) with cOmplete protease inhibitors (Roche). Pellets were kept on ice when adding solubilisation buffer, then subjected to probe sonication with 80–100% intensity for 3 × 30 s rounds with resting on ice between rounds. Samples were centrifuged at 10,000× g to pellet insoluble material and the supernatant transferred to fresh tubes. Protein yields were then assayed via gel-based densitometry [36] and BCA assay (Thermo Fisher, Rockford, MA, USA). Protein samples were then reduced using 5 mM tris(2-carboxyethyl)phosphine (TCEP) to break disulfide linkages and linearise proteins; the cysteines were then blocked with 20 mM acrylamide monomers for 90 min. Protein digestion was performed on 100 μg of sample with the UTC7 solution diluted by 8-fold with 100 mM ammonium bicarbonate solution to enable the activity of trypsin by ensuring the pH is above 8.0 and the concentration of urea is below 1 M. Peptides were cleaned of contaminating buffer components using C18 solid-phase extraction columns (Empore™ Extraction Disk Cartridge). Briefly, the columns were activated with 100% acetonitrile (ACN) then equilibrated with 2% v/v ACN with 0.2% v/v trifluoroacetic acid (TFA). Samples were then applied to the equilibrated column and then washed with 2% v/v ACN; elution was achieved with 50% v/v ACN, 0.1% v/v TFA. The organic solvent was then removed by rotary vacuum-assisted evaporation under non-heating conditions. These samples were then analysed utilising a LC–MS/MS on a QExactive Plus (Thermo Fisher Scientific, Bremen, Germany) [37].

2.4. Control Peptide Lysate Preparation

For the creation of a L-DOPA-positive control proteome, the three non-treated replicate digests were pooled for processing, allowing comparison of the controls against the L-DOPA-treated sample in terms of percentage conversion. All samples underwent the same

methodology until the point before injection onto the mass spectrometer. The peptide solutions were aliquoted and frozen at -80 °C, then lyophilised in a freeze drier (Christ). These peptides were then stored at -20 °C until required.

2.4.1. Tyrosinase Conversion Proof of Concept

Previous work investigating tyrosinase conversion used much larger amounts of material than would normally be available for proteomic studies. Thus, a proof-of-concept microscale experiment was performed utilising a synthetic peptide sequence from UCP-5, a mitochondrial protein implicated in Parkinson's disease [38]. Following the method by Ito et al. 1984 [35], conversion efficiencies were so low that detectable conversion on an entire proteome was not possible.

2.4.2. Microscale Whole Proteome Tyrosinase Conversion

An optimised method for tyrosinase conversion adapted from Ceinska et al. 2016 [34] was applied to the control peptide lysate. Molar concentrations for the reaction were based on an average peptide weight of 1 kDa (https://www.bioline.com/media/calculator/01_0 4.html) (accessed on 1 April 2018). The reaction buffer final constitution was of 0.5 mM borate buffer adjusted to pH 7.0 with NaOH, 2 mM ascorbic acid, 6.7 mM hydroxylamine and varying ratios of tyrosinase to peptide digest. For concentration-dependent scaling down, 1 mM (100 μg) of peptide was reacted with differing ratios of tyrosinase (10:1, 2:1, 1:1) in 100 μL reaction volumes. The most detrimental factor reported for the industrialised conversion method is the lack of aeration in the reaction vessel and the exhaustion of free oxygen during back conversions of deoxy-tyrosine and creation of the hydroxylated residues. To prevent this, the micro-scaling of the reaction inside of a microtube combined with low-speed desktop vortexing ensures aeration of the reaction. The reaction was performed at 30 °C for 240 min. To remove the tyrosinase from the peptide proteome to be analysed, a 3 kDa molecular weight cut off spin filter (Pall) was used as the intact 112 kDa tyrosinase enzyme is retained and the peptides freely pass through. This sample was then diluted to the correct concentration for injection onto the QExactive Plus mass spectrometry platform with interfering buffer components washed away during peptide trapping before analysis.

2.5. MALDI Analysis of Synthetic UCP-5 Conversion

Native and converted UCP-5 peptides were spotted at a volume of 1 μL combined with 1 μL of 5 mg/mL CHCA dissolved in 50% *v/v* acetonitrile, 0.1% *v/v* TFA, 10 mM $NH_4H_2PO_4$ onto a clean 386-well OptiTOF target plate (AB Sciex) and allowed to dry. Peptide samples were then analysed using a 5800 MALDI-TOF/TOF (AB Sciex) mass spectrometer in positive ion reflector mode. Laser intensity was set to 2600 for MS parent ion scans and 3000 for MSMS fragmentation ion scans; 400 laser shots were averaged for MS scans and up to 1250 shots were averaged for MSMS scan with the Dynamic Exit algorithm selected. MS parent ion scans were calibrated using the TOF/TOF standards while MS/MS fragmentation ion scans were calibrated using the fragments of Glu-Fibrinopeptide B present in the TOF/TOF standards mixture. The resulting MS and MS/MS spectral data were then converted to Mascot Generic Format using T2D Extractor and the data searched with Mascot (provided by the Australian Proteomics Computational Facility, hosted by the Walter and Eliza Hall Institute for Medical Research Systems Biology Mascot Server) against the Human Proteome with variable modification Y(+15.99). Theoretical fragments were generated utilising the fragment ion calculator hosted by http://db.systemsbiology.net (accessed on 20 August 2017).

2.6. Q-Exactive Plus LC–MS/MS

Peptides were analysed using a QExactive Plus mass spectrometer (Thermo Fisher Scientific) and a M-Class Acquity liquid chromatograph (Waters). Peptides were subject to electrospray ionisation (ESI) at 2.6 kV using a PicoFrit column (75 mm ID × 300 mm;

New Objective, Woburn, MA, USA) packed with Daisogel SP-ODS-BIO C18 resin (3 μm) (Osaka Soda Co, Osaka, Japan). Data-dependent acquisition was performed using the top 12 highest-abundance ions with dynamic exclusion of precursors for 30 s post-fragmentation, resolution was set to 70,000 for a survey scan range of $350_{m/z}$–$1500_{m/z}$, with an AGC target of 1×10^6 and a maximum injection time (I.T) of 100 ms, fragmentation scans were acquired at a resolution of 17,500, with an I.T of 50 ms, an isolation window of $1.4_{m/z}$, a mass range beginning at $50_{m/z}$ and a normalized collision energy (NCE) of 28. Prior to ESI, 1 μg of peptide samples was loaded on to a nanoEase Symmetry C18 trapping column (180 μm × 20 mm), analytical chromatographic separation was performed utilising buffer A containing 99.8% ultra-pure water with 0.2% formic acid and buffer B containing 99.8% acetonitrile and 0.2% formic acid. Separation occurred over a 140 min gradient at a constant flow rate of 300 nL/min, and gradient steps were linear with buffer A (95, 95, 60, 20, 99, 99) at minute (0, 1, 121, 123, 125, 127, 150).

2.7. Data Analysis

Raw files were loaded into PEAKS Studio (v10.6, build 2020 1015) [39]. Searches were performed using the following parameters: a parent mass tolerance of 20 ppm; a fragment tolerance of 0.02 Da; semi-specific trypsin with three missed cleavages; de novo sequencing with a maximum of modifications three per peptide—fixed: propionamide (C), variable: carbamylation (K and N-term, identified on a first pass PTM search) and oxidation (M, Y, F). PEAKS database searching was performed against the human proteome containing isoforms indexed on the 3 September 2017 (71,540 entries combined from TREMBL and SwissProt). Proteins containing greater than 10,000 amino acids were removed, and standard contaminants were also included. PEAKS PTM searching was performed on the previous database search results using the inbuilt 312 modifications. Results were filtered to include de novo scores ≥50, peptides -10lgP score of ≥16.1, and a 1% FDR, reporting 40,752 identified PSMs. Proteins at 1% FDR with one unique peptide resulted in a total of 9462 identified proteins. PTMs with an Ascore above 1 (/1000) were retained.

The control and treatment technical injections were also searched using MaxQuant (v1.6.14.0) against the same database and common contaminants. All parameters can be found within the MaxQuant summary text file (PRIDE: PXD025759). Briefly, parameters were set to a minimum peptide length of 7, fixed modification propionamide (C), variable modifications of oxidation (M, Y, F) and N-terminal acetylation. MaxQuant LFQ was performed with a minimum of two unique peptides used for protein identification and one unique plus razor peptide used for quantitation with a minimum ratio of two. An FDR of 0.01 was used for PSM protein and site, with match between runs used with a matching window of 0.7 min and an alignment window of 20 min.

MaxQuant output files were loaded into the Mass Dynamics webserver ([40], https://app.massdynamics.com/) (accessed on 1 March 2021) for analysis, a public link to this project has been generated https://app.massdynamics.com/p/dfbb55e9-0a6d-450e-8e8e-3f0c070e5bf9 (accessed on 1 March 2021). A total of 2047 protein groups were identified, and 1846 protein groups quantified. Proteins with an adjusted FDR of less than 5% and 2-fold up or down in abundance were used for further pathway analysis (Supplementary File S3). Pathway analysis was performed using the pathways tool that queries Reactome repository [41]. Proteins significantly changing in abundance (adjusted *p*-value of <0.05 and a $\log_2$-fold change >1) were exported from Mass Dynamics project and annotated using UniProt's ID mapping function. Further analysis of protein interactions between up and down proteins were also performed within STRING (v11.0) [42]; medium confidence was used with all interaction data sources enabled.

3. Results

3.1. Tyrosinase Conversion of Synthetic Peptide UCP5

To demonstrate that tyrosinase could be used to convert tyrosine residues in a peptide to DOPA, the synthetic peptide UCP5 (SLC25A14) (mitochondrial uncoupling protein) was

chosen for conversion due to its implication in the maintenance of mitochondrial stability and its exploration in models of oxidative stress in SH-SY5Y cells [20]. The tryptic peptide sequence used—EEGVLALYSGIAPALLR—was successfully converted to contain DOPA by the method of Ito et al. [35]. Figure 2 shows the theoretical fragments generated by the native peptide and the L-DOPA-containing peptide with spectral evidence confirming this pattern shown underneath.
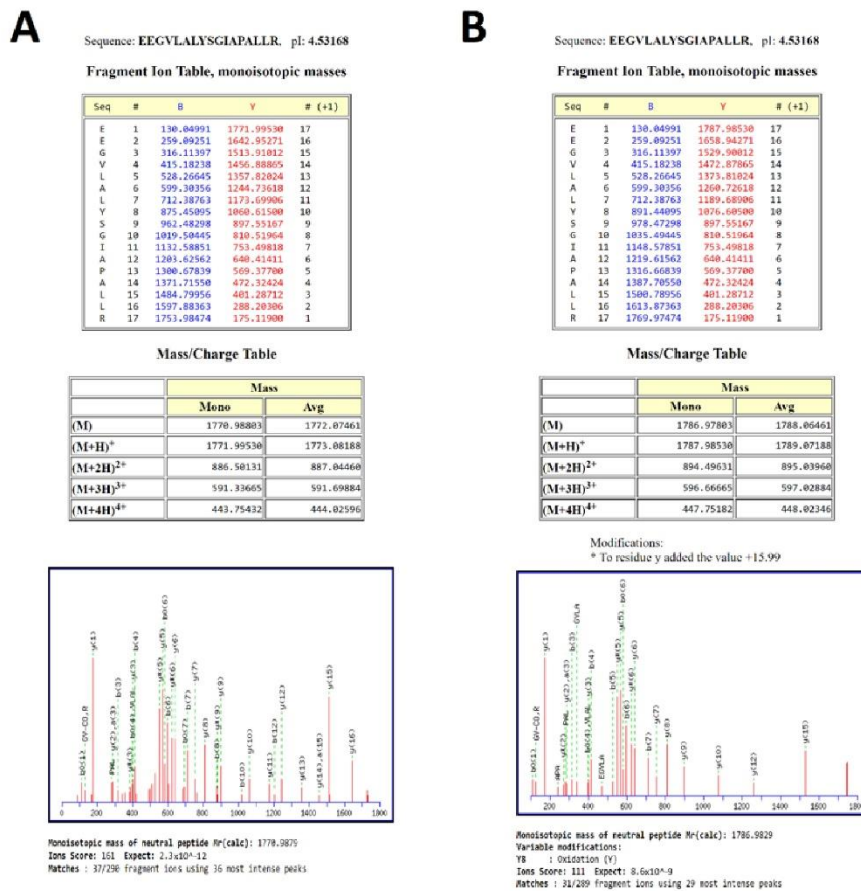
Figure 2. Spectral evidence of a positive control for PB-DOPA. (**A**) The native peptide sequence. (**B**) Addition of oxygen to tyrosine indicated by the increase in peptide mass and altered masses of the b8 and y10 ions.

3.2. Proteome Analysis

The method of Ito et al. [35] was not successful when applied on a proteome level (data not shown). Thus, the tyrosinase conversion method outlined by Cieńska et al. [34] was scaled down and applied to the SH-SY5Y control proteome. To assess the off-target effects, the percentages of tyrosine (Y), methionine (M) and phenylalanine (F) residues displaying oxidation are presented in Figure 3. Regardless of residue or peptide sequence, the percentage oxidation displays a similar pattern for each sample. The DOPA-treated biological samples display a significant increase in methionine oxidation (~26%) versus

the control (~1%), with a non-statistically significant increase in the amount of DOPA and oxidised phenylalanine. All three tyrosinase amounts used for treatment resulted in a >10% increase in DOPA-containing peptides, with the 1:1 conversion having a higher specific L-DOPA formation noted by the decreased amount of phenylalanine oxidation. The conversion process led to >75% methionine oxidation across all three enzyme:substrate ratios and, regardless of which ratio was used, DOPA formation is significantly lower than the 95% conversion achieved on free tyrosine reported in the unscaled method [34]. However, a significantly lower number of spectra were detected following treatment, and even less following the conversion process, despite an equal concentration of protein being processed for each sample. A label-free quantitative analysis was performed in PEAKS Studio to generate the total ion count normalization factors displayed in Figure 4, which indicates a notable decrease in the amount of sample ionised in treatments and conversions relative to control.
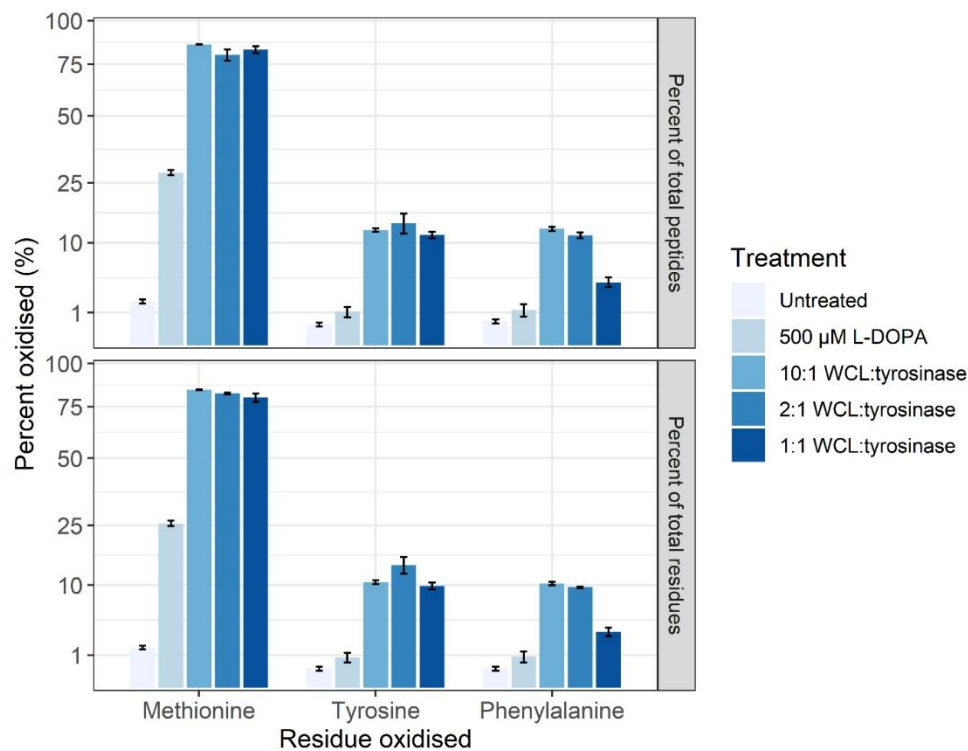
Figure 3. The overall oxidation profile of untreated, DOPA-treated and converted SH-SY5Y proteomes. The oxidation profile of methionine, tyrosine (L-DOPA), and phenylalanine is shown for each treatment group. The level of oxidation is presented as a percentage of the total number of identified peptides containing M, Y or F (**top**) and individual residues (**bottom**). The three biological control samples were pooled for the tyrosinase conversion, using tyrosinase to peptide digest ratios of 10:1, 2:1, and 1:1. Regardless of residue or peptide sequence, the percentage of oxidation follows a similar pattern for each sample. Error bars represent the standard error of the mean (SEM).
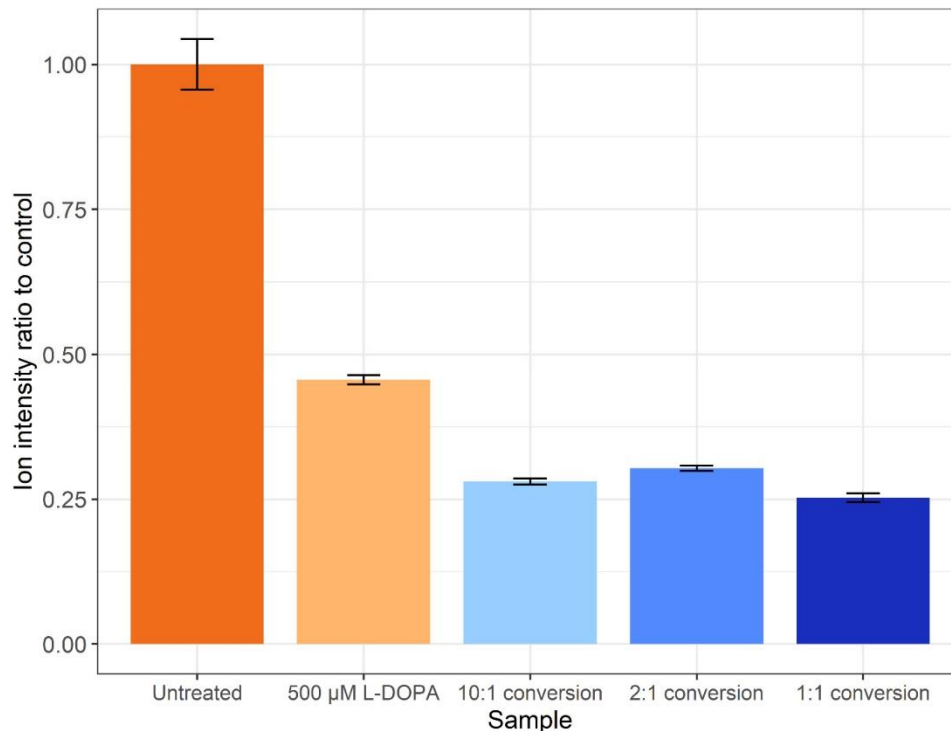
Figure 4. The ratio of total ion intensities compared to the control. Error bars represent the SEM.

3.3. Proteoform Analysis

Proteoforms containing DOPA were generated by incubation of cells with L-DOPA in cell culture or by conversion of tyrosine present in tryptic peptides generated from control cell lysates to DOPA using the enzyme tyrosinase. From the output data, DOPA-containing peptide sequences were identified in untreated control, DOPA-treated or tyrosinase-converted samples (Table 1). The identification was based on feature detection and 'match between runs' or peptide identification transference rather than peptide identification in each injection, and sequences that could not be confidently attributed to a condition were discarded. From the control samples, only 34 peptide sequences could be accurately identified as containing DOPA whilst for the DOPA-treated samples there is spectral evidence for 101 sequences containing DOPA. The tyrosinase conversion produced 532 DOPA-containing peptide sequences. The sequences and associated protein identifications can be found in Supplementary File S2. For the DOPA-containing peptide sequences, protein identifications were extracted and compiled, and 75 active UniProt entries were found in the DOPA-treated cells and 37 in the control. Overlap of proteins containing DOPA between control and DOPA treatment was performed (Supplementary File S2), resulting in 30 proteins being unique to control and 68 unique to treatment. Four proteins were identified in both groups: valine–tRNA ligase; 26S proteasome regulatory subunit 6A; calreticulin; tubulin beta-4A chain; clathrin heavy chain 1; clathrin heavy chain; and heat shock protein HSP 90-beta. Analysis of the biological process GO terms unique to control identified four key terms that covered the identifications; response to stress; organelle organization; post-transcriptional regulation of gene expression; and sulfur compound

biosynthetic process. Analysis of the biological process GO terms of the unique to DOPA with strongest FDR identified; protein folding; establishment of localisation; transport; and regulation of biological quality.

Table 1. Peptide sequences containing DOPA within each and proteins identities matched (using accessions and UniProt ID matching).

Identification	Control	Treatment	Conversion
peptides	34	101	532
proteins	37	75	317

When analysing the tyrosinase-converted samples to identify DOPA-containing peptides, four peptides were identified in the biological samples from quality fragment data generated in the converted samples (Table 2), which would have otherwise not been detected with confidence.

Table 2. Peptide identifications due to conversion samples containing quality spectra. The four peptide sequences were only able to be identified due to the converted sample having a high-quality spectrum.

Localisation	Sequence	Protein Accession	PTM Site Ascore
Control	AAGGDGDDSLY(+15.99)PIAVLIDELR	P30154 ǀ 2AAB_HUMAN	Y11:L-DOPA:1000.00
DOPA	DLYANTVLSGGTTMY(+15.99)PGIADR	P60709 ǀ ACTB_HUMAN:P63261 ǀ ACTG_HUMAN	Y15:L-DOPA:27.62
DOPA	GINPDEAVAY(+15.99) GAAVQAGVLSGDQ(+0.98)DTGDLVLLDVC (+71.04)PLTLGIETVGGVMTK	P11021 ǀ GRP78_HUMAN	Y10:L-DOPA:1000.00;Q23:Deamidation (NQ):0.00;C34:Propionamide:1000.00
DOPA	DLYAN(+.98)TVLSGGTTMY(+15.99) PGIADR	P60709 ǀ ACTB_HUMAN:P63261 ǀ ACTG_HUMAN	N5:Deamidation (NQ):1000.00;Y15:L-DOPA:10.83

3.4. Biological Insights from Pathway Analysis

Quantitative analysis was performed on the control vs. treated samples using MaxQuant. Pathway analysis was subsequently performed using the Mass Dynamics webserver, with the volcano plot shown in Figure 5. All proteins with a $\log_2$-fold change ≥ 1 and adjusted *p*-value < 0.05 were used for subsequent pathway analyses. The project results can be accessed and interacted with by the public link provided. The statistically significantly enriched pathways (<5% FDR) can be found in the mass dynamic project, which shows proteins involved in the metabolism of RNA and its processing to be decreased in abundance.

Quantitative analysis of the proteome quantified 1846 protein groups with 390 significantly changing in abundance (adjusted *p*-value <0.05 and a 2-fold change in abundance). Within the treatment, 50 proteins were increased in abundance, and 340 decreased (Supplementary File S3).

Proteins found to be significantly decreased in abundance in the treatment group include Parkinson's disease protein 7 (PARK7), elongation factor G, mitochondrial (GFM1), superoxide dismutase (SOD1), high mobility group protein HMG-I/HMG-Y, transcription factor A, mitochondrial (TFAM). The proteins significantly increased in abundance include dihydrofolate reductase (DHFR), sorting nexin-3 (SNX3) and s-formylglutathione hydrolase. Proteins related to neurodegenerative disease were also found to be significantly changed in abundance, including proteins involved in Parkinson's disease (PARK7) and amyotrophic lateral sclerosis (ubiquilin-4, superoxide dismutase, and RNA-binding protein FUS).
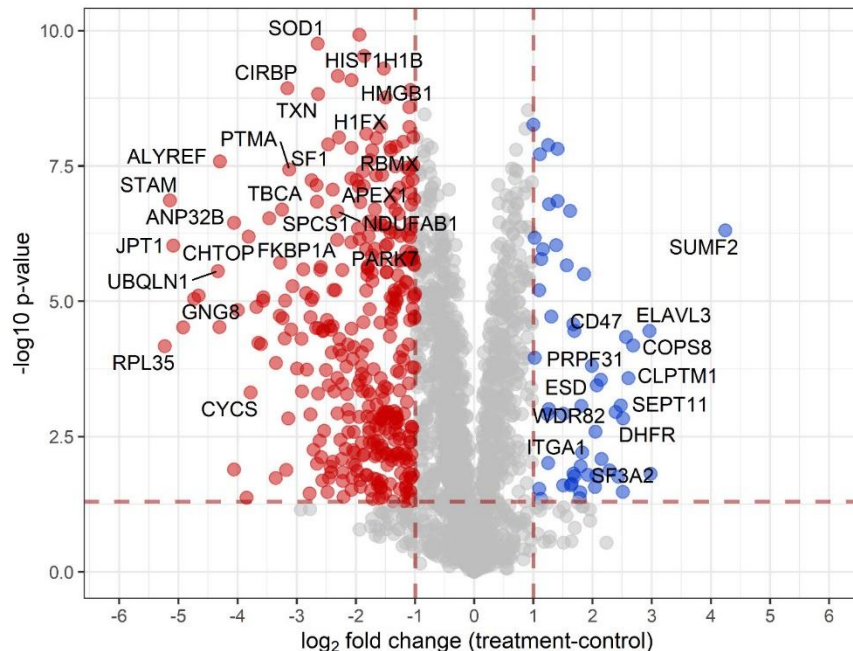
Figure 5. Volcano plot of proteins from label-free quantitative analysis. Coloured proteins have *p*-value <0.05 and a log$_2$ fold change of ≥1. Blue proteins are up in treatment, and red proteins are down.

Analysis of interactions between proteins changing in abundance was performed using STRING. Of the 50 proteins of increased abundance, 49 were able to be mapped within STRING. Statistical analysis identified an enrichment of pathways involved in the unfolded protein response and components of the proteostatic mechanisms of the cell. The STRING analysis of the proteins significantly increased in abundance can be found at the following permalink: https://version-11-0b.string-db.org/cgi/network?networkId=bXEIWsTDm0Bp (accessed on 4 March 2021). Of the 340 proteins decreased in abundance within treated cells, 316 were mapped within STRING; the analysis identified a decrease in proteins involved in mRNA metabolic processes and gene expression. Further KEGG pathway analysis of proteins decreasing in abundance showed proteins involved in thermogenesis and oxidative phosphorylation, and disease pathways including Alzheimer's disease, Parkinson's disease, and Huntington's disease. Disease-annotated proteins are mainly mitochondrial proteins essential for mitochondrial homeostasis (NDUFA4, SOD1, TFAM, COX5A) or the prevention of apoptosis (BID and CYCS). The STRING analysis of the proteins significantly decreased in abundance can be found at the following permalink: https://version-11-0b.string-db.org/cgi/network?networkId=bjdJP4IT0l36 (accessed on 4 March 2021).

4. Discussion

In this study, we demonstrate the first use of the tyrosinase reaction to convert whole-sample peptide digests for subsequent proteomic analysis, providing spectral evidence for the incorporation of DOPA into the proteome of a neuronal human cell line. To the best of our knowledge, this is the first presentation of shotgun proteomic data analysed for DOPA as a protein constituent on a proteome scale and the first to provide spectral data generated

45

by DDA to a public repository. These data also provide an example of where matching between runs (peptide identification transference) from a tyrosinase-converted proteome allowed identification of DOPA-containing sequences in biological samples (Table 2).

The initial trialing of tyrosinase as a conversion method utilised the method presented by Ito et al. [35], with successful conversion of a synthetic peptide from the protein UCP5 (SLC25A14) shown in Figure 2. This method was not successful in a whole proteome conversion (data not shown), so an alternative method was employed that in theory presented an efficiency of conversion of >95% on free tyrosine [34]. The Cieńska et al. (2016) [34] protocol needed to be scaled down to be economically feasible and accommodate the smaller sample material present in typical proteomic samples. The conversion of peptide residues to DOPA was achieved at a ~10% efficiency (Figure 3) regardless of the ratio of tyrosinase to peptide; these data indicate that factors other than the amount of tyrosinase are affecting the efficiency of conversion. The converted samples displayed an increase from 1% methionine oxidation within the control sample to ~75% across all three conversion samples indicating a high level of oxidation has occurred during conversion.

The conversion of proteins to contain L-DOPA is a pursuit within the field of synthetic biomaterials and has potential as a biological glue. Several pursuits of conversion of marine proteins from mussels have reported <15% yield on peptide conversions or 8% on specific residue conversion with tyrosinase [43]. Furthermore, the efficiency may be lower due to a loss of peptide solubility as it has been reported that use of strong acid conditions such as acetic acid can enhance the recovery of L-DOPA-containing proteins when separating from insoluble material [43]. Analysis of the amount of ionised peptide (Figure 4) agrees with the loss of peptides as amounts loaded were assumed on the amount of control peptide used for each conversion reaction. The decrease in ionisable peptides could be explained by several factors, including a loss of solubility, or some proteins possibly remained bound to the tyrosinase enzyme. It has been shown previously that the incorporation of L-DOPA into a non-solvent-exposed region of a protein can lead to protein insolubility, whilst external residues do not affect protein solubility [21]. Future improvements to the conversion experiment could include the use of a modified version of tyrosinase as the mtyr-CNK tyrosinase that is active at below 30 °C and optimally active at pH 7.0 may result in a marked increase in the specificity and reduction in oxidations found in this study [44].

There are several factors listed in the literature that could be leading to an increase in the amount of oxidation products within the sample. Tyrosinase isolates have been known to be contaminated with other oxidases [44], which could be leading to the off-target effects. L-DOPA may not have been the only product, there may have been 3,4,5-trihydroxyphenylalanine (TOPA) as has been reported in a 2002 study [45]. Any free tyrosine remaining in the lyophilised peptide sample may have been converted to L-DOPA, which can in turn increase the amount of ROS species. Similarly, the formation of L-DOPA in the peptide backbone could also be increasing the amount of ROS generated, especially if the tyrosinase powder contained impurities such as metals to allow Fenton chemistry to progress [25,46]. One of the then stated objectives of this work was the creation of a spectral library for use in DIA based experiments. The low conversion efficiency of tyrosine to DOPA prevents the use of these data in the construction of a comprehensive library and the optimisations listed above need to be performed before tyrosinase conversion becomes a viable method for the generation of spectral libraries for DIA experiments.

Treatment of the SH-SY5Y cells with 500 μM L-DOPA resulted in a significant increase in methionine oxidation (25%) (Figure 3) which is used in a proteome as a measure of oxidative stress [47]. However, L-DOPA treatment did not result in a statistically significant increase in the amount of DOPA- or oxidised phenylalanine-containing peptides which may be reflective of the neuronal cell line's ability to turn over oxidised and damaged proteins [48] or the acute nature of the treatment.

Analysis of L-DOPA-containing sequences resulted in the identification of 37 (30 unique) proteins in the control and 75 (68 unique) in L-DOPA-treated cells. Analysis of the proteins containing L-DOPA unique to control via GO terms identified general terms including

response to stress; organelle organization; post-transcriptional regulation of gene expression; and sulfur compound biosynthetic process. These do not highlight a specific group of proteins in the cell that would be ascribed to detrimental pathology but can be taken forward in literature as naturally occurring. Analysis of the proteins containing L-DOPA unique to treatment revealed four GO terms including protein folding; establishment of localisation; transport; and regulation of biological quality. These terms taken together show proteins that are responsible for the quality control of proteins are being either damaged by hydroxyl attack or are aggregating within the cell due to high levels of L-DOPA incorporation. From these experimental results, it is not possible to determine whether the L-DOPA proteins are indeed aggregating, which is a function of the protein half-lives, or whether it is a totally random pattern. Further experiments involving the use of the D-DOPA isomer may allow the distinction of ROS species generation and the effects of incorporation as this would still result in oxidative stress but not incorporation. The use of degradomics with stable isotope labels in pulse chase experiments during treatment may help identify the proteostatic influence due to L-DOPA treatment and the functions behind incorporation.

Pathway analysis within Mass Dynamics on the L-DOPA-treated samples showed a statistical significantly enrichment for RNA metabolism and processing pathways within proteins decreasing in abundance (see external project). A decrease in RNA expression is supported by the oxidative stress observed in Parkinson's disease studies [49,50]. Within the treated sample, superoxide dismutase (SOD1) was decreased in abundance; this protein is involved in the scavenging of free radicals within the cell and acts to protect proteins from damage. A decrease in the amount of SOD1 would lead to a decreased capacity to cope with oxidative stress, and SOD1 has been found oxidatively damaged in Alzheimer's disease and Parkinson's disease patients [51].

Analysis of changes in protein abundance in STRING revealed that proteins involved in the unfolded protein response were increased due to treatment. Concurrently, proteins involved in transcription were down-regulated, suggesting that L-DOPA treatment results in cell stress related to protein misfolding. Several proteins involved in neurodegenerative diseases were significantly changed by L-DOPA treatment, namely those involved in mitochondrial homeostasis providing further evidence that L-DOPA may induce proteome changes that have previously been linked to neurodegenerative pathology [52–55]. As this study used an acute treatment, future proteomic analyses may benefit from chronic L-DOPA exposure. Exploring longer treatments may help to further understand the underlying mechanisms related to L-DOPA incorporation in proteins [7].

The experimental data demonstrate cell stress under L-DOPA treatment, which agrees with in vitro studies that have shown that L-DOPA can be toxic to dopaminergic neurons [16,56] and raises the possibility that L-DOPA could impair neuronal function in vivo. Progressive peripheral nerve deficits have been reported with increasing exposure to L-DOPA [57]; in a subset of L-DOPA-treated patients, a decrease in striatal DA transport binding was found, indicating a reduction in dopaminergic neuron function or number [58]. The toxic effects of L-DOPA in vitro can be divided into those due to oxidative stress, which is likely to be an in vitro phenomenon, and those that are due to an effect on protein, which is independent of oxidative stress [5,7]. These have been hypothesised to be a result of the mistaken incorporation of L-DOPA into proteins [17] which is further evidenced by the prevalence of misfolded protein responses identified within these experimental data.

This proteomic experiment was performed without fractionation or enrichment. It is foreseeable that highly fractionating the samples could increase the identification of L-DOPA-containing proteins. Furthermore, online/inline fractionations such as that of gas-phase fractionation can be performed on a pooled sample to enable matching between runs, which in essence is a manner for reproducing highly empirical DIA libraries [59]. Enrichment methodologies are routinely used in the PTM analysis field of phospho-proteomics [60,61] and the creation of either a chemical or antibody-based enrichment method for L-DOPA could then be applied to biological or tyrosinase-converted samples.

The use of an antibody affinity method would require high specificity that would be expensive to raise; a chemical-based affinity method may offer more feasibility due to the unique catechol chemical structure of DOPA.

5. Conclusions

This work developed a microscale tyrosinase conversion method for the creation of DOPA-containing samples to aid in the exploration of its presence in the human proteome. This work also identified proteins containing DOPA generated in vitro via treatment with L-DOPA and an increase in the unfolded protein response with significantly changing proteins related to neurodegenerative diseases. Further optimisation experiments listed in the discussion need to be performed to create a sufficiently comprehensive spectral library for future DIA experiments.

References

1. Rascol, O.; Payoux, P.; Ory, F.; Ferreira, J.J.; Brefel-Courbon, C.; Montastruc, J.-L. Limitations of current Parkinson's disease therapy. *Ann. Neurol.* **2003**, *53*, S3–S12. [CrossRef]
2. Parkkinen, L.; O'Sullivan, S.S.; Kuoppamäki, M.; Collins, C.; Kallis, C.; Holton, J.L.; Williams, D.R.; Revesz, T.; Lees, A.J. Does levodopa accelerate the pathologic process in Parkinson disease brain? *Neurology* **2011**, *77*, 1420–1426. [CrossRef]
3. Alexander, T.; Sortwell, C.E.; Sladek, C.D.; Roth, R.H.; Steece-Collier, K. Comparison of neurotoxicity following repeated administration of L-dopa, D-dopa, and dopamine to embryonic mesencephalic dopamine neurons in cultures derived from Fisher 344 and Sprague-Dawley donors. *Cell Transplant.* **1997**, *6*, 309–315. [CrossRef]
4. Asanuma, M.; Miyazaki, I.; Ogawa, N. Dopamine- or L-DOPA-induced neurotoxicity: The role of dopamine quinone formation and tyrosinase in a model of Parkinson's disease. *Neurotox. Res.* **2003**, *5*, 165–176. [CrossRef]
5. Chan, S.W.; Dunlop, R.A.; Rowe, A.; Double, K.L.; Rodgers, K.J. L-DOPA is incorporated into brain proteins of patients treated for Parkinson's disease, inducing toxicity in human neuroblastoma cells in vitro. *Exp. Neurol.* **2012**, *238*, 29–37. [CrossRef] [PubMed]
6. Colamartino, M.; Santoro, M.; Duranti, G.; Sabatini, S.; Ceci, R.; Testa, A.; Padua, L.; Cozzi, R. Evaluation of Levodopa and Carbidopa Antioxidant Activity in Normal Human Lymphocytes In Vitro: Implication for Oxidative Stress in Parkinson's Disease. *Neurotox. Res.* **2014**, *27*, 106–117. [CrossRef] [PubMed]
7. Giannopoulos, S.; Samardzic, K.; Raymond, B.B.; Djordjevic, S.P.; Rodgers, K.J. L-DOPA causes mitochondrial dysfunction in vitro: A novel mechanism of L-DOPA toxicity uncovered. *Int. J. Biochem. Cell Biol.* **2019**, *117*, 105624. [CrossRef] [PubMed]
8. Koh, S.-H.; Park, H.-H.; Choi, N.-Y.; Lee, K.-Y.; Kim, S.; Lee, Y.J.; Kim, H.-T. Protective effects of statins on l-DOPA neurotoxicity due to the activation of phosphatidylinositol 3-kinase and free radical scavenging in PC12 cell culture. *Brain Res.* **2011**, *1370*, 53–63. [CrossRef]

9.　Kostrzewa, R.M.; Kostrzewa, J.P.; Brus, R. Neuroprotective and neurotoxic roles of levodopa (L-DOPA) in neurodegenerative disorders relating to Parkinson's disease. *Amino Acids* **2002**, *23*, 57–63. [CrossRef]

10.　Lipski, J.; Nistico, R.; Berretta, N.; Guatteo, E.; Bernardi, G.; Mercuri, N.B. L-DOPA: A scapegoat for accelerated neurodegeneration in Parkinson's disease? *Prog. Neurobiol.* **2011**, *94*, 389–407. [CrossRef] [PubMed]

11.　Mytilineou, C.; Han, S.-K.; Cohen, G. Toxic and Protective Effects of l-DOPA on Mesencephalic Cell Cultures. *J. Neurochem.* **1993**, *61*, 1470–1478. [CrossRef]

12.　Park, K.H.; Choi, N.-Y.; Koh, S.-H.; Park, H.-H.; Kim, Y.S.; Kim, M.-J.; Lee, S.-J.; Yu, H.-J.; Lee, K.-Y.; Lee, Y.J.; et al. L-DOPA neurotoxicity is prevented by neuroprotective effects of erythropoietin. *NeuroToxicology* **2011**, *32*, 879–887. [CrossRef]

13.　Pedrosa, R.; Soares-Da-Silva, P. Oxidative and non-oxidative mechanisms of neuronal cell death and apoptosis by L-3,4-dihydroxyphenylalanine (L -DOPA) and dopamine. *Br. J. Pharmacol.* **2002**, *137*, 1305–1313. [CrossRef] [PubMed]

14.　Perveen, A.; Khan, H.Y.; Hadi, S.M.; Damanhouri, G.A.; Alharrasi, A.; Tabrez, S. Pro-oxidant DNA Breakage Induced by the Interaction of L-DOPA with Cu(II): A Putative Mechanism of Neurotoxicity. In *Genedis 2014: Neurodegeneration*; Vlamos, P., Alexiou, A., Eds.; Springer Int Publishing Ag: Cham, Switzerland, 2015; pp. 37–51.

15.　Song, J.; Kim, B.C.; Nguyen, D.-T.T.; Samidurai, M.; Choi, S.-M. Levodopa (L-DOPA) attenuates endoplasmic reticulum stress response and cell death signaling through DRD2 in SH-SY5Y neuronal cells under α-synuclein-induced toxicity. *Neuroscience* **2017**, *358*, 336–348. [CrossRef]

16.　Basma, A.N.; Morris, E.J.; Nicklas, W.J.; Geller, H. l-DOPA Cytotoxicity to PC12 Cells in Culture Is via Its Autoxidation. *J. Neurochem.* **2002**, *64*, 825–832. [CrossRef] [PubMed]

17.　Rodgers, K.J.; Hume, P.M.; Morris, J.G.L.; Dean, R.T. Evidence for L-dopa incorporation into cell proteins in patients treated with levodopa. *J. Neurochem.* **2006**, *98*, 1061–1067. [CrossRef] [PubMed]

18.　Rodgers, K.J.; Shiozawa, N. Misincorporation of amino acid analogues into proteins by biosynthesis. *Int. J. Biochem. Cell Biol.* **2008**, *40*, 1452–1466. [CrossRef] [PubMed]

19.　Koh, S.-H.; Kim, S.H.; Kim, H.-T. Role of glycogen synthase kinase-3 in l -DOPA-induced neurotoxicity. *Expert Opin. Drug Metab. Toxicol.* **2009**, *5*, 1359–1368. [CrossRef]

20.　Jami, M.-S.; Pal, R.; Hoedt, E.; Neubert, T.A.; Larsen, J.P.; Møller, S.G. Proteome analysis reveals roles of L-DOPA in response to oxidative stress in neurons. *BMC Neurosci.* **2014**, *15*, 1–11. [CrossRef]

21.　Ozawa, K.; Headlam, M.J.; Mouradov, D.; Watt, S.J.; Beck, J.L.; Rodgers, K.J.; Dean, R.T.; Huber, T.; Otting, G.; Dixon, N.E. Translational incorporation of L-3,4-dihydroxyphenylalanine into proteins. *FEBS J.* **2005**, *272*, 3162–3171. [CrossRef]

22.　Rodgers, K.J.; Dean, R.T. Metabolism of protein-bound DOPA in mammals. *Int. J. Biochem. Cell Biol.* **2000**, *32*, 945–955. [CrossRef]

23.　Dunlop, R.A.; Rodgers, K.J.; Dean, R.T. Recent developments in the intracellular degradation of oxidized proteins1,2. *Free. Radic. Biol. Med.* **2002**, *33*, 894–906. [CrossRef]

24.　Rodgers, K.J.; Hume, P.M.; Dunlop, R.A.; Dean, R.T. Biosynthesis and turnover of DOPA-containing proteins by human cells. *Free. Radic. Biol. Med.* **2004**, *37*, 1756–1764. [CrossRef] [PubMed]

25.　Dunlop, R.A.; Dean, R.T.; Rodgers, K.J. The impact of specific oxidized amino acids on protein turnover in J774 cells. *Biochem. J.* **2008**, *410*, 131–140. [CrossRef] [PubMed]

26.　Matayatsuk, C.; Poljak, A.; Bustamante, S.; Smythe, G.A.; Kalpravidh, R.W.; Sirankapracha, P.; Fucharoen, S.; Wilairat, P. Quantitative determination of ortho- and meta-tyrosine as biomarkers of protein oxidative damage in beta-thalassemia. *Redox Rep.* **2007**, *12*, 219–228. [CrossRef]

27.　Pattison, D.I.; Dean, R.T.; Davies, M.J. Oxidation of DNA, proteins and lipids by DOPA, protein-bound DOPA, and related catechol(amine)s. *Toxicology* **2002**, *177*, 23–37. [CrossRef]

28.　Steele, J.; Italiano, C.; Phillips, C.; Violi, J.; Pu, L.; Rodgers, K.; Padula, M. Misincorporation Proteomics Technologies: A Review. *Proteomes* **2021**, *9*, 2. [CrossRef]

29.　Searle, B.C.; Pino, L.K.; Egertson, J.D.; Ting, Y.S.; Lawrence, R.T.; MacLean, B.X.; Villén, J.; MacCoss, M.J. Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat. Commun.* **2018**, *9*, 1–12. [CrossRef]

30.　Zolg, D.P.; Wilhelm, M.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Delanghe, B.; Bailey, D.J.; Gessulat, S.; Ehrlich, H.-C.; Weininger, M.; et al. Building ProteomeTools based on a complete synthetic human proteome. *Nat. Methods* **2017**, *14*, 259–262. [CrossRef]

31.　Walker, E.J.; Bettinger, J.Q.; Welle, K.A.; Hryhorenko, J.R.; Ghaemmaghami, S. Global analysis of methionine oxidation provides a census of folding stabilities for the human proteome. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6081–6090. [CrossRef]

32.　Bettinger, J.Q.; Welle, K.A.; Hryhorenko, J.R.; Ghaemmaghami, S. Quantitative Analysis of in Vivo Methionine Oxidation of the Human Proteome. *J. Proteome Res.* **2020**, *19*, 624–633. [CrossRef]

33.　Carballo-Carbajal, I.; Laguna, A.; Romero-Giménez, J.; Cuadros, T.; Bové, J.; Martinez-Vicente, M.; Parent, A.; Gonzalez-Sepulveda, M.; Peñuelas, N.; Torra, A.; et al. Brain tyrosinase overexpression implicates age-dependent neuromelanin production in Parkinson's disease pathogenesis. *Nat. Commun.* **2019**, *10*, 1–19. [CrossRef]

34.　Cieńska, M.; Labus, K.; Lewańczuk, M.; Koźlecki, T.; Liesiene, J.; Bryjak, J. Effective L-Tyrosine Hydroxylation by Native and Immobilized Tyrosinase. *PLoS ONE* **2016**, *11*, e0164213. [CrossRef]

35.　Ito, S.; Kato, T.; Shinpo, K.; Fujita, K. Oxidation of tyrosine residues in proteins by tyrosinase. Formation of protein-bonded 3,4-dihydroxyphenylalanine and 5-S-cysteinyl-3,4-dihydroxyphenylalanine. *Biochem. J.* **1984**, *222*, 407–411. [CrossRef]

36. Vincent, S.G.; Cunningham, P.R.; Stephens, N.L.; Halayko, A.J.; Fisher, J.T. Quantitative densitometry of proteins stained with Coomassie Blue using a Hewlett Packard scanjet scanner and Scanplot software. *Electrophoresis* **1997**, *18*, 67–71. [CrossRef]

37. Roediger, B.; Lee, Q.; Tikoo, S.; Cobbin, J.C.; Henderson, J.M.; Jormakka, M.; O'Rourke, M.B.; Padula, M.P.; Pinello, N.; Henry, M.; et al. An Atypical Parvovirus Drives Chronic Tubulointerstitial Nephropathy and Kidney Fibrosis. *Cell* **2018**, *175*, 530–543. [CrossRef]

38. Ho, P.W.; Ho, J.W.; Liu, H.-F.; So, D.H.; Tse, Z.H.; Chan, K.-H.; Ramsden, D.B.; Ho, S.-L. Mitochondrial neuronal uncoupling proteins: A target for potential disease-modification in Parkinson's disease. *Transl. Neurodegener.* **2012**, *1*, 3. [CrossRef]

39. Tran, N.H.; Qiao, R.; Xin, L.; Chen, X.; Liu, C.; Zhang, X.; Shan, B.; Ghodsi, A.; Li, M. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat. Methods* **2019**, *16*, 63–66. [CrossRef]

40. Bloom, J.I.; Triantafyllidis, A.; Burton, P.; Infusini, G.; Webb, A.I. Mass Dynamics 1.0: A streamlined, web-based environment for analyzing, sharing and integrating Label-Free Data. *bioRxiv* **2021**. [CrossRef]

41. Jassal, B.; Matthews, L.; Viteri, G.; Gong, C.; Lorente, P.; Fabregat, A.; Sidiropoulos, K.; Cook, J.; Gillespie, M.; Haw, R.; et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **2020**, *48*, D498–D503. [CrossRef]

42. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef]

43. Hwang, D.S.; Gim, Y.; Yoo, H.J.; Cha, H.J. Practical recombinant hybrid mussel bioadhesive fp-151. *Biomaterials* **2007**, *28*, 3560–3568. [CrossRef]

44. Do, H.; Kang, E.; Yang, B.; Cha, H.J.; Choi, Y.S. A tyrosinase, mTyr-CNK, that is functionally available as a monophenol monooxygenase. *Sci. Rep.* **2017**, *7*, 17267. [CrossRef]

45. Burzio, L.A.; Waite, J. The Other Topa: Formation of 3,4,5-Trihydroxyphenylalanine in Peptides. *Anal. Biochem.* **2002**, *306*, 108–114. [CrossRef]

46. Rodgers, K.J.; Wang, H.; Fu, S.; Dean, R.T. Biosynthetic incorporation of oxidized amino acids into proteins and their cellular proteolysis. *Free Radic. Biol. Med.* **2002**, *32*, 766–775. [CrossRef]

47. Ghesquière, B.; Jonckheere, V.; Colaert, N.; Van Durme, J.; Timmerman, E.; Goethals, M.; Schymkowitz, J.; Rousseau, F.; Vandekerckhove, J.; Gevaert, K. Redox Proteomics of Protein-bound Methionine Oxidation. *Mol. Cell. Proteom.* **2011**, *10*. [CrossRef]

48. Dunlop, R.A.; Brunk, U.T.; Rodgers, K.J. Proteins containing oxidized amino acids induce apoptosis in human monocytes. *Biochem. J.* **2011**, *435*, 207–216. [CrossRef]

49. Alam, Z.I.; Jenner, A.; Daniel, S.E.; Lees, A.J.; Cairns, N.; Marsden, C.D.; Jenner, P.; Halliwell, B. Oxidative DNA Damage in the Parkinsonian Brain: An Apparent Selective Increase in 8-Hydroxyguanine Levels in Substantia Nigra. *J. Neurochem.* **2002**, *69*, 1196–1203. [CrossRef]

50. Cenini, G.; Lloret, A.; Cascella, R. Oxidative Stress in Neurodegenerative Diseases: From a Mitochondrial Point of View. *Oxidative Med. Cell. Longev.* **2019**, *2019*, 1–18. [CrossRef]

51. Choi, J.; Rees, H.D.; Weintraub, S.T.; Levey, A.I.; Chin, L.-S.; Li, L. Oxidative Modifications and Aggregation of Cu,Zn-Superoxide Dismutase Associated with Alzheimer and Parkinson Diseases. *J. Biol. Chem.* **2005**, *280*, 11648–11655. [CrossRef]

52. Corona, J.C.; Duchen, M.R. Impaired mitochondrial homeostasis and neurodegeneration: Towards new therapeutic targets? *J. Bioenerg. Biomembr.* **2015**, *47*, 89–99. [CrossRef]

53. Kim, H.; Perentis, R.J.; Caldwell, G.; Caldwell, K.A. Gene-by-environment interactions that disrupt mitochondrial homeostasis cause neurodegeneration in C. elegans Parkinson's models. *Cell Death Dis.* **2018**, *9*, 555. [CrossRef]

54. Kang, I.; Chu, C.T.; Kaufman, B.A. The mitochondrial transcription factor TFAM in neurodegeneration: Emerging evidence and mechanisms. *FEBS Lett.* **2018**, *592*, 793–811. [CrossRef]

55. Markaki, M.; Tavernarakis, N. Mitochondrial turnover and homeostasis in ageing and neurodegeneration. *FEBS Lett.* **2020**, *594*, 2370–2379. [CrossRef] [PubMed]

56. Pardo, B.; Mena, M.; Casarejos, M.; Paíno, C.; De Yébenes, J. Toxic effects of L-DOPA on mesencephalic cell cultures: Protection with antioxidants. *Brain Res.* **1995**, *682*, 133–143. [CrossRef]

57. Toth, C.; Breithaupt, K.; Ge, S.; Duan, Y.; Terris, J.M.; Thiessen, A.; Wiebe, S.; Zochodne, D.W.; Suchowersky, O. Levodopa, methylmalonic acid, and neuropathy in idiopathic Parkinson disease. *Ann. Neurol.* **2010**, *68*, 28–36. [CrossRef] [PubMed]

58. Fahn, S.; Oakes, D.; Shoulson, I.; Kieburtz, K.; Rudolph, A.; Lang, A.; Olanow, C.W.; Tanner, C.; Marek, K. Levodopa and the progression of Parkinson's disease. *N. Engl. J. Med.* **2004**, *351*, 2498–2508.

59. Searle, B.C.; Swearingen, K.E.; Barnes, C.A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **2020**, *11*, 1548. [CrossRef]

60. Ping, L.; Kundinger, S.R.; Duong, D.M.; Yin, L.; Gearing, M.; Lah, J.J.; Levey, A.I.; Seyfried, N.T. Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci. Data* **2020**, *7*, 315. [CrossRef]

61. Su, Z.; Burchfield, J.G.; Yang, P.; Humphrey, S.J.; Yang, G.; Francis, D.; Yasmin, S.; Shin, S.-Y.; Norris, D.M.; Kearney, A.L.; et al. Global redox proteome and phosphoproteome analysis reveals redox switch in Akt. *Nat. Commun.* **2019**, *10*, 1–18. [CrossRef]

Chapter Four: Meta-analysis of publicly available data to identify the presence of L-DOPA in proteins

4.1 Introduction

Oxidised amino acids can be formed in cells by reaction with free radical species formed through Fenton chemical reactions or from the escape of radicals from mitochondria or lysosomes [172]. The oxidative damage to proteins can cause loss of function affecting a protein's degradation with potentially wide-reaching effects on metabolic pathways within cells. In dividing cells, the burden of damaged proteins that cannot be degraded by the proteolytic machinery of the cell can be distributed into daughter cells, but within the central nervous system, neuronal cells are predominantly a terminally differentiated population making them acutely susceptible to oxidised protein aggregate-based toxicity [58, 160, 171, 200-202]. Oxidative stress and the resultant increased levels of damaged or non-native proteins may overwhelm the proteostasis mechanisms of the cell and cause a decline in cell function. Elevated levels of oxidised and damaged proteins are linked to many neurological diseases and are also a characteristic of aging neuronal tissues [203].

The formation of oxidised amino acids can occur while they are part of a protein or within the pool of free amino acids in a cell or tissue. There is a possibility that oxidised amino acids released by proteolysis could then re-cycle through proteins; this is something that to our knowledge has not been explored. A prime example of an oxidised amino acid that is produced intentionally is L-DOPA. Within the cell the amino acid tyrosine becomes hydroxylated to form L-DOPA, an intermediary in the synthesis of the signalling molecule dopamine [107, 108, 160, 164, 170, 201]. L-DOPA is a highly reactive molecule, with the capability to participate in fenton reactions [204]. Upon entry to dopaminergic neurons, L-DOPA is converted to dopamine by Aromatic Amino Acid De-carboxylase and the resultant dopamine stored by vesicular monoamine transporter 2 in vesicles [205, 206].

In this chapter, the term protein bound L-DOPA (PB-DOPA) will be used to denote the presence of L-DOPA in a protein without empirical determination of its origin which could either be from oxidation of a tyrosine residue or from the incorporation of L-DOPA into the protein by the synthetic machinery of the cell in place of a tyrosine residue [160].

L-DOPA is used as a therapeutic drug to restore the reduced ability of PD patients to produce dopamine. Depletion of dopamine in PD is caused by loss of dopaminergic neurons in the substantia nigra and results in a progressive loss of the ability to initiate movement [33]. The presence of L-DOPA itself has been shown to result in oxidative stress in several studies [133, 136, 176, 194, 197, 207, 208] and has been demonstrated to be a component of brain proteins in Parkinson's disease by redox staining of extracted brain protein and quantification of DOPA by FL-HPLC [57]. Several studies have

been performed looking at oxidative damage and its role in other neurodegenerative diseases such as AD. Examples of this research are within the field of redox proteomics, with primary works by Allan Butterfield [40, 56, 190, 209-211], namely the triangle of death which implicates mitochondrial dysfunction (Reported as a novel toxicity of L-DOPA [172]), oxidative stress and altered protein homeostasis as causes of neurodegenerative diseases [1].

The presence of PB-DOPA in neurological proteomes could offer insight into disease states and identify previously uncharacterised proteoforms that may have pathogenic roles [33, 57, 133, 134, 136, 173, 174, 177, 180]. Oxidative stress increases over the course of a neuronal cell's life, accompanied by a decreased ability to deal with misfolded proteins [212]. It is also known that L-DOPA-containing proteins can form proteolytic resistant aggregates and possibly oligomers [12, 29, 32, 34-36, 58, 80, 213-217], a prime example being alpha-synuclein containing L-DOPA [193].

The previous chapter identified L-DOPA incorporation in a cell model and described a method for creating L-DOPA containing peptides from any proteomic sample. The present chapter aims to look for evidence for the incorporation of L-DOPA into proteins in healthy neuronal tissue while simultaneously looking at the formation of oxidised forms of phenylalanine (Ox-Phe) and methionine (Ox-Met). We also compare the level of amino acid modification in tissues from L-DOPA-treated PD patients against controls and identify which proteins contain these specific oxidative modifications. These data are essential to further assess if L-DOPA treatment results in an increase in the levels of PB-DOPA and if the modified proteins could have functional ramifications that could lead to increased neuronal toxicity. These results are also applicable to the study of oxidative stress in healthy neuronal ageing. To achieve these aims we reinterpreted publicly available data. The proteomics field has a long history of repository deposition of experimental data, making it feasible to explore other hypotheses without the need to obtain patient brain samples and generate data that would not be feasible in the time frame of a PhD candidature.

4.2 Materials and methods

4.2.1 Meta-analyses of online datasets for L-DOPA incorporation

The proteomic data repository PRIDE DB was accessed, and complete projects were downloaded and analysed as per the original authors' parameters through PEAKS Studio but with the addition of variable modifications of oxidation of tyrosine, methionine and phenylalanine. The datasets used were PXD000561 "draft map of the human proteome" (Brain tissue subsection), PXD000427 "substantia nigra in Parkinson's disease" and PXD008036 "Parkinson's Olfactory Lobe Proteome" which also included the variable search of oxidation on aspartic acid, asparagine, proline, arginine, lysine, and the C-terminal of a protein. These datasets were analysed subsequently using excel, R and Graphpad Prism

9 to produce oxidation levels of amino acids and allow statistical analysis. STRING DB (outlined in 4.2.3) was used for biological network generation and analysis inferences with stringent cut-offs applied. All networks generated have an associated permanent link that can be accessed indefinitely.

4.2.2 PEAKS software methodology

To standardise the analysis of all datasets, a singular human UniProt SwissProt/KB proteome downloaded on 03/09/2017 was used (Appendix A_1). The database consists of all Uniprot accessions associated to the human proteome; this includes the trEMBL KB datasets and all validated isoforms with a valid protein count at 71,540. Utilising a large reference proteome gives a higher chance of identifying unique PTM's by being able to identify isoforms that may contain unique PTMs. Proteomic analysis was performed in PEAKS Studio X Plus version 10.5. *De novo* sequencing was performed with confident identifications (>50% localisation score) used for a database search with the same parameters presented by each dataset with the addition of oxidised tyrosine (PB-DOPA) and oxidised phenylalanine (Ox-Phe). The output of this search was then subjected to an open PTM search using the 313 in-built modifications to identify other co-occurring PTMs which was used for the complete analysis. All peptide spectrum matches were filtered at a minimum FDR of 1% at the peptide level and a -10logP at the protein level (99% confidence equivalent for PEAKS protein outputs).

4.2.3 STRING DB analysis methodology

STRING DB is an analysis tool for functional enrichment of protein-protein interactions by network generation and contains information in the form of experimental interactions, *in silico* predictions and text mined interactions. The use of STRING results in functional enrichment analysis by way of accessing and summarising biological information and ontologies as well as being able to summarise lists of proteins and presents them in a more visually informative diagram. Protein accession outputs from PEAKS Studio were cleaned to produce a compatible format for STRING multiple accession inputs. The lists of exclusive proteins were parsed into individual protein networks with the mapping to STRING items retained. Due to the redundancy and lack of experimental evidence for a large proportion of the proteins identified, not every protein identified can be analysed in the subsequent networks. Physical interaction networks were generated excluding text mining as a source for this analysis as the text mining feature does not provide a constructive network when exploring known interactions. Minimum confidence required for an interaction was set to "high confidence (0.700)" and the edges displayed were strength based. K-means clustering was applied to each network generated and was supervised for cluster number input per network. String ontology and biological mapping was created in an unsupervised manner with all results being statistically significant (p<0.05). Venn diagrams were generated using Bioinformatics & Evolutionary Genomics website[218]

4.3 Results

4.3.1 Human draft proteome reanalysis (Brain)

With the advent of public repositories for the sharing of proteomics data, it has become possible to mine and reanalyse data acquired by other researchers whose experiments had a different focus, but samples analysed could potentially have of broader interest. One such dataset is the draft map of the human proteome [219], and this dataset was reanalysed for markers of oxidative damage and for the presence of PB-DOPA. This dataset is described as healthy control tissue, and it was used to set a threshold for PB-DOPA in a normal biological context. All associated project files and outputs can be found in Appendix 4_1.

4.3.1.1 Overview of data

The raw data files of the draft human proteome were downloaded and analysed using the settings listed in the publication [219], with the addition of oxidation of tyrosine (PB-DOPA) and the oxidation of phenylalanine (Ox-Phe). The analysis produced a total of 18,770 proteins identified with 7,148 protein groups, from 196,403 peptides matching to 126,637 sequences (Table 1).

Table 1: Summary statistics of the reanalysed human brain proteome.

	#Features	Identified			#Peptides	#Seq	#Proteins		
		#PSMs	#Scans	#Features			Groups	All	Top
Total	3861299	918775	918775	613094	196403	126637	7148	18770	8454
Adult Frontal Cortex 1	938290	187212	187212	131875	68052	47906	6664	17074	7817
Adult Frontal Cortex 2	917040	245106	245106	163109	83832	71346	6924	18096	8165
Adult Frontal cortex 3	822134	190447	190447	127283	62211	48016	6449	16535	7562
Foetal brain one	850334	154343	154343	99229	65423	46720	6803	17463	8030
Foetal Brain two	333501	141667	141667	91598	38634	30587	5471	13940	6364

*Breakdown per original tissue sample is displayed. Columns: **Scans** refers to MS/MS scans and features here refers to features with an associated identification. **Features** are precursors identified with a signal. **Identified** have associated matched information: peptide spectral matches **PSMs**, **Peptide**s refers to the number of peptides whilst **#seq (sequences)** are the unique sequences identified. Protein **Groups** refers to matches of proteins that are indistinguishable, **All** proteins refers to matches regardless of distinguishing peptides and **Top** refers to the high confidence matches.*

4.3.1.2 Proteome oxidation level analysis

Oxidative analysis was performed at the residue level, which was produced by identifying individual peptides containing the modified form of the amino acid relative to the total number of residues of that unmodified amino acid expressed as a percentage (Figure 3) (Appendix 4_1_2-A Supplementary file 1). Analysis of the oxidated amino acid levels of the human brain proteome displayed a range of values: PB-DOPA from 0.4% to 1.5% whilst the phenylalanine oxidation ranged from 0.6% to 1.5% and the methionine oxidation ranged from 14% to 70%.

There was an experiment-wide high level of methionine oxidation but PB-DOPA and hydroxylated phenylalanine remained at less than two percent, while the high level of methionine oxidation is contrasted against chapter three where only ~25% of methionine's were oxidised in the L-DOPA treated condition. The ratio of PB-DOPA to Phe-Ox within individual peptides is only noticeably shifted towards PB-DOPA in the foetal brain 2 and equal in adult brain 1. Two different fractionation techniques were used in the generation of this data with adult brain 3 and foetal brain 2 subject to gel-based fractionation while the others used high pH reverse phase chromatography for sample fractionation. All samples besides adult brain 2 had greater than 40% oxidation with the adult brain 1 having over 70% methionine oxidation. There is no clear pattern of increased methionine oxidation correlating with the formation of PB-DOPA or Ox-Phe.
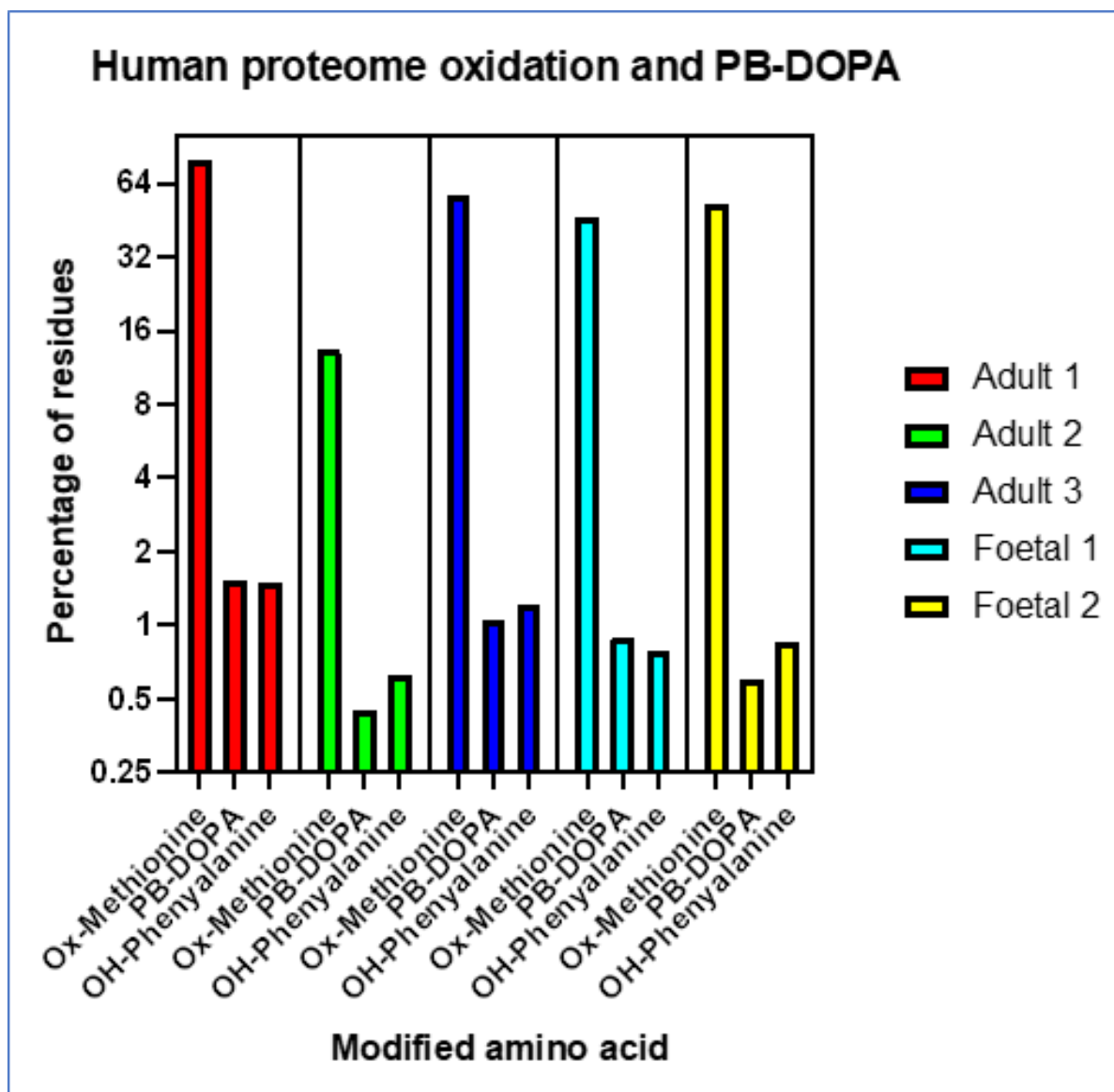
Figure 3: Human proteome oxidation and PB-DOPA (Brain tissue subset). The modified residues as a percentage of total specific amino acid containing residues. This data set is a foundation for the human proteome analysis and its data is used across the field of proteomics. Evident here is a high percentage of methionine oxidation with varying levels of previously unidentified PB-DOPA and hydroxylated phenylalanine. Noted here is that the foetal brains that have not been exposed to long term environmental factors and have similar ratios to the adult brains.

4.3.1.3.2 Identification of peptides containing PB-DOPA and Ox-Phe

The peptide results file was filtered for protein sequences containing PB-DOPA and Ox-Phe. At the 1% FDR and Ascore >500, 650 peptides were identified as containing PB-DOPA (196,403 peptides were identified in total, see Table 1), and of these 631 were assigned to a protein's accession number. The accessions were filtered for unique entries resulting in 577 matches mapped to the current UniProt repository, and all accessions were matched resulting in 570 proteins (due to merges in the human proteome), the complete annotated list of these can be found in Appendix 4_1_2-A (Supplementary

file 1 and intermediary files in 4_1_2-B) with associated gene ontology mappings. There were 824 peptides containing Oxidised Phe, with 805 having accessions assigned, 671 unique accessions remained after filtering, matching to 662 active entries. These data were downloaded and added to the aforementioned Appendix 4_1_2-A (Supplementary file 1 and intermediary files in 4_1_2-B).

4.3.1.3.3 Analysis of peptides containing PB-DOPA and Ox-Phe

Accession lists for the stringently filtered peptides were used to generate a Venn diagram to demonstrate the overlapping protein accessions and to investigate if there are differences in the formation of each modification (Figure 4). This analysis has left multiple accessions from a single peptide intact as this is likely important information as selection of razor assignment or another abbreviation technique results in potential proteoform loss. Given the high redundancy of the FASTA database used, there is a potential for protein isoforms to possibly be in both unique sections (See section 4.2.2).
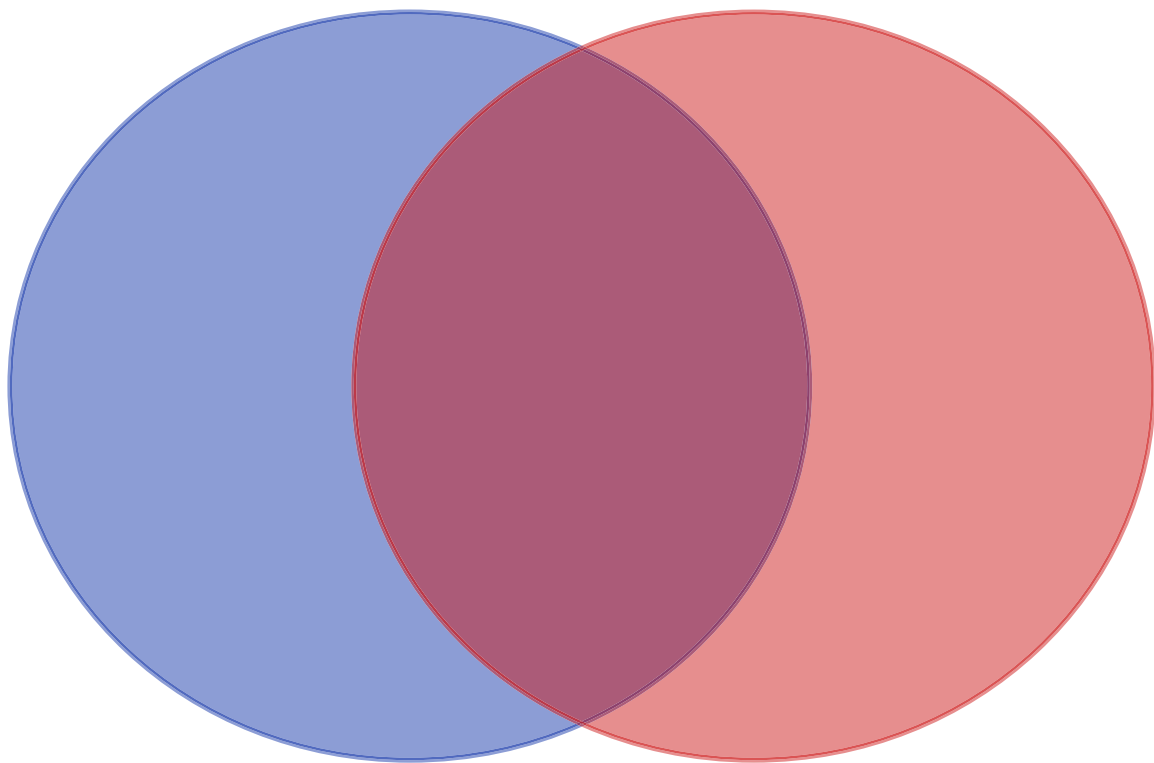
Figure 4: Venn diagram of PB-DOPA and Ox-Phe accessions (Protein identities) attached to peptide sequences containing each modification. PB-DOPA unique of 344, Ox-Phe of 438 and an overlap of 233. Total of 1,015 accessions.

4.3.1.3.4 Proteins containing both PB-DOPA and Ox-Phe

Network analysis was performed for the three Venn diagram sections. The overlapping 233 accessions mapped to 159 identities in STRING as seen in Figure 5 (permalink: https://version-11-5.string-db.org/cgi/network?networkId=btpkvD2aRk5C). Based on a physical component network analysis, the tubulins and actins are central to the network of overlapping oxidised proteins. A large number of these proteins were annotated to be either extracellular by way of vesicle or exosome (123). Mitochondrial annotations made up 35 (22%) of the identities with several of these implicated in Parkinson's and other neurodegenerative diseases which is likely a function of the tissue type itself. Proteins of interest include myelin basic protein (MBP), myelin proteolipid protein (PLP1), thioredoxin-dependent peroxide reductase (PRDX2) and three solute carriers (SLC1A2, SLC25A4, SLC25A6) which are responsible for the transport of excitatory amino acids and ATP/ADP exchange. Three heat shock proteins were also found to be common to both forms of oxidation (HSPA8, HSP90AA1, HSP90AB1).
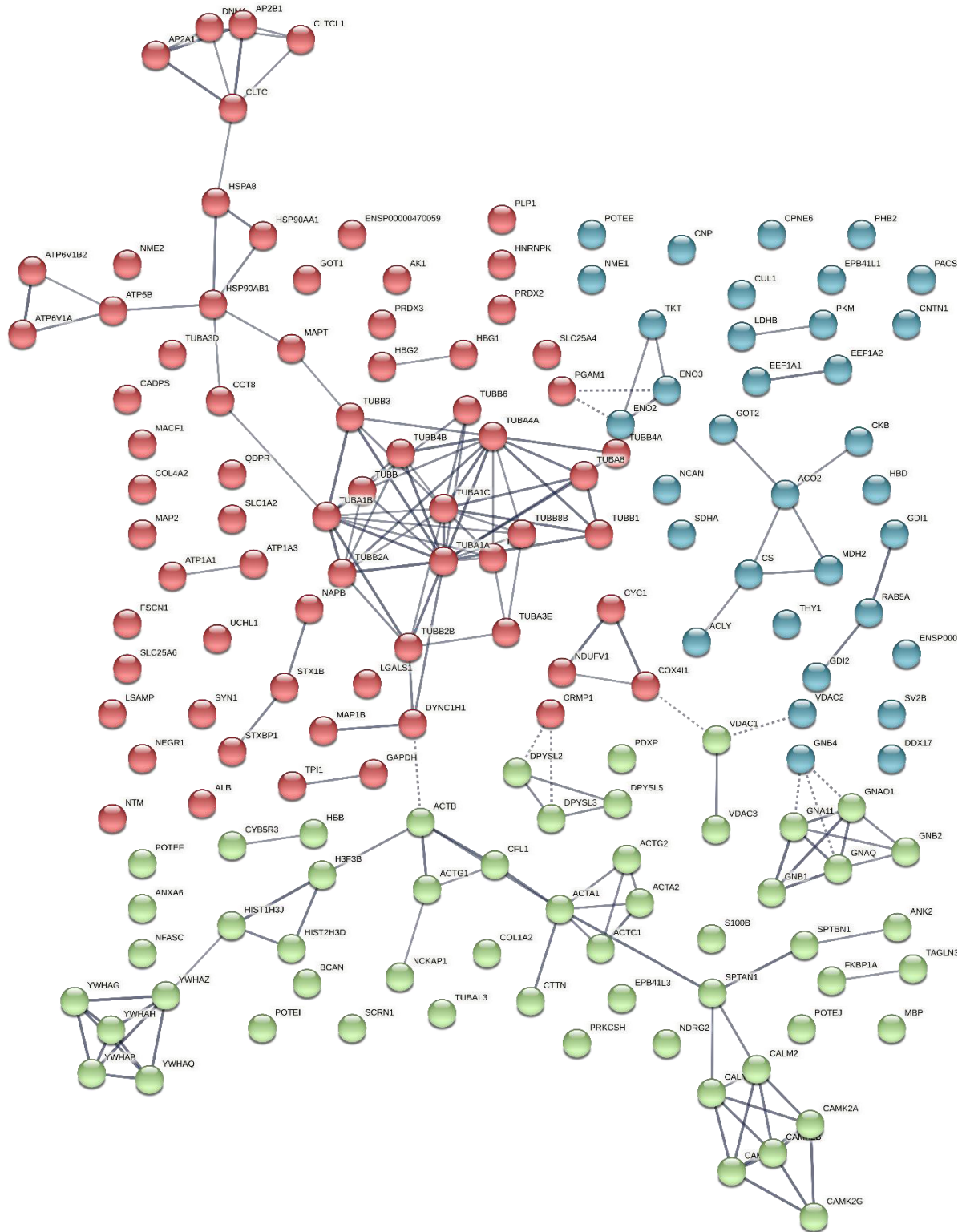
Figure 5: Proteins containing both PB-DOPA and oxidised phenylalanine. Red: 70 identity cluster containing tubulin proteins, solute carriers, heat shock proteins with other chaperones, Thioredoxin-dependent peroxide reductase and other mitochondrial proteins involved in ATP synthesis. Green: Cluster with 55 identities containing actin proteins, calmodulins and enriched for vesicles at the cellular location (42). Blue: Cluster with 34 identities containing mostly extracellular exosome annotation (21), mitochondrial proteins involved in metabolism including glycolysis, gluconeogenesis and the TCA cycle.

4.3.1.3.5 Proteins exclusively containing PB-DOPA

Network analysis of the proteins mapped from the PB-DOPA exclusive accessions was performed, resulting in 229 matches with the central clustering being histones and the ubiquitin-proteasomal system (Figure 6). (permalink: https://version-11-5.string-db.org/cgi/network?networkId=birlUaFdZV9U). Enrichment analysis identified the majority of proteins being involved in cellular component organisation and/or biogenesis, with a fraction in ATP metabolic processes and cellular protein complex assembly. The majority of the proteins identified had two major roles with either being in protein binding (139) or the RNA binding (192). Proteins were annotated to mainly extracellular exosome/space/region or vesicle, with several KEGG pathways identified including: systemic lupus erythematosus, Parkinson's disease, amyotrophic lateral sclerosis and prion disease. Of the proteins identified, a pubmedID (27448508) was enriched for with 14/79 identities found, and this publication identified aggregated proteins that can be used to distinguish AD hippocampus samples from normal controls [220]. Myelin basic protein was identified within this analysis from an alternative identifier to the overlapping analysis above. The interacting groups of proteins forming the majority of identities in the three clusters (43) are annotated to be part of the mitochondrion.
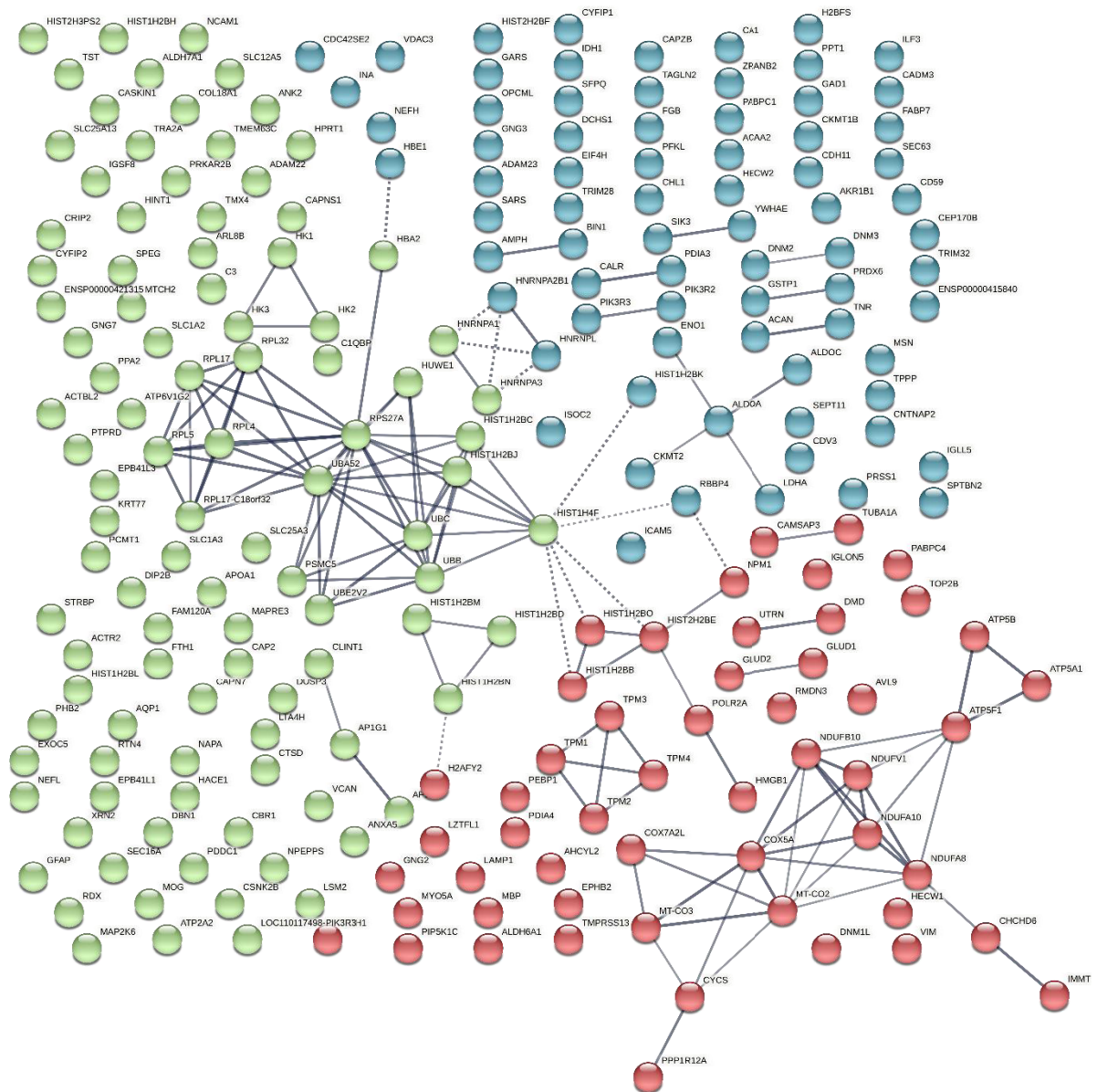
Figure 6: Protein's containing PB-DOPA based on unique identifiers. A total of 433 identifiers matched to 229 proteins upon STRING analysis. Green: 102 identities, mapping to central nervous system disease (21), extracellular exosome (49), nine identities mapping (9/79) to proteins that can be used to distinguish AD aggregates from healthy control. Blue: 74 identities, mapping to extracellular exosome and synapse/ cell junction, immune system proteins and binding proteins. Red: 53 identities mapping to mitochondrial key proteins involved in oxidative phosphorylation and NADH respiratory chain.

4.3.1.3.6 Proteins exclusively containing Ox-Phe

Protein accessions from Ox-Phe only peptides resulted in 313 matches in STRING (Figure 7). The String network has central clustering of heterogeneous nuclear ribonucleoproteins, ribosomal and T-complex proteins (permalink: https://version-11-5.string-db.org/cgi/network?networkId=bi4L6LuMRs0N). The majority of the protein's biological process categories are cellular component organisation or biogenesis (177) with an extensive list of processes that compose the majority of cell functions. KEGG pathways with the lowest false discovery rate included dopaminergic synapse proteins (20/128) and a series of synapse specific proteins. Disease associated gene annotations included disease (148) with sub-terms of anatomical identity (114), nervous system disease (64) and neurodegenerative disease (23). Cellular component ontology enrichment included extracellular vesicle (123), cytosol (193), vesicle (162), mitochondrial (37) and exosome (121) with the majority being cytoplasmic (276).
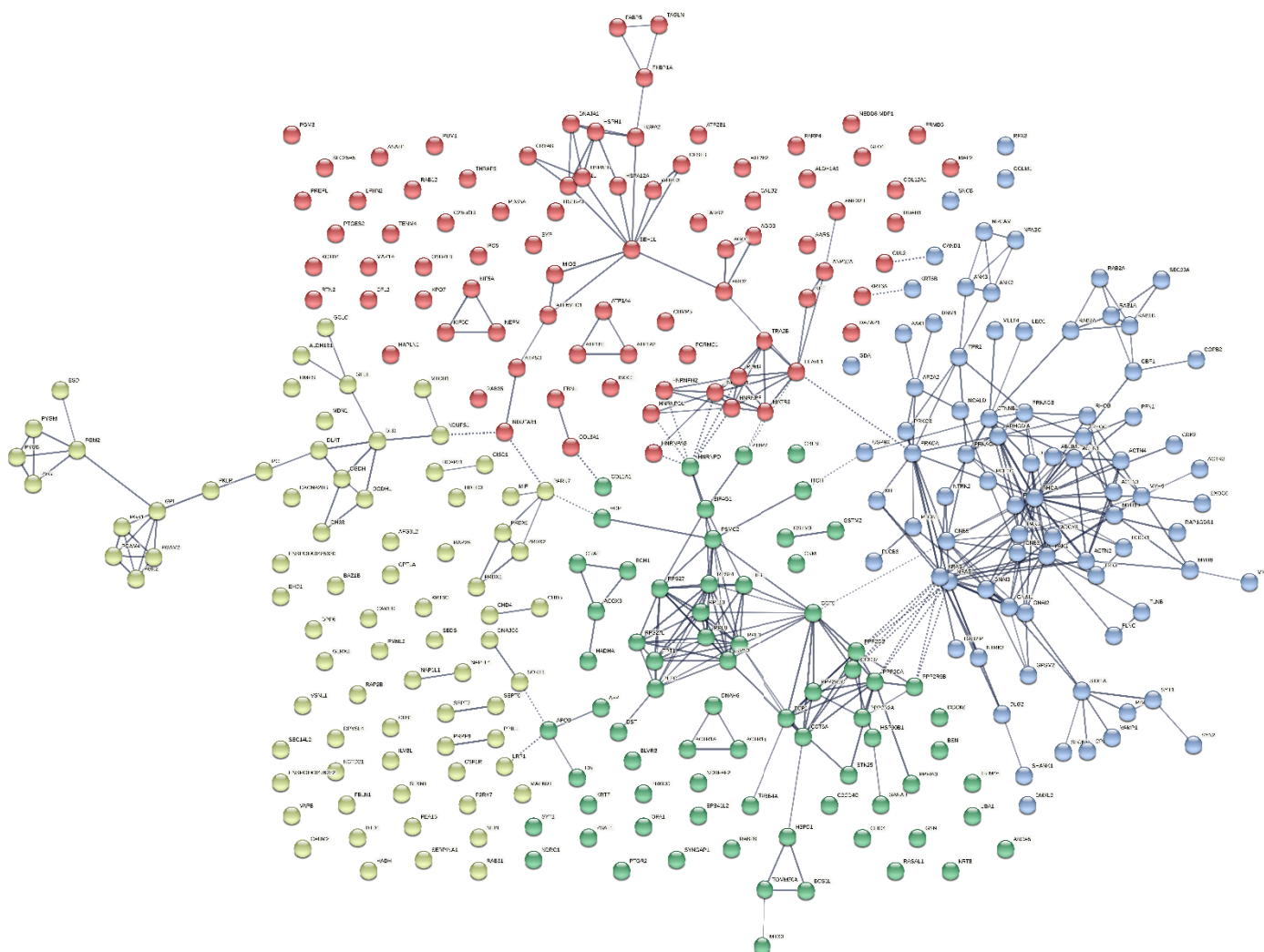
Figure 7: Network analysis of proteins containing only Ox-Phe containing peptides, 313 identities matched out of 438 supplied. Red: 82 identities mapping to extracellular vesicle (27) ribonucleoprotein complex (14) and organelle (75). Proteins involved in Prion disease including Heat shock proteins were also within this cluster. Yellow: 77 identities mapping to several oxidoreductase complex members, Peroxiredoxin isoforms 1 2 5 and linked interactor PARK7. Green: 72 identities including ribosomal proteins, chaperones and enriched for extracellular vesicle (36). Blue: 82 identities of a highly interconnected cluster. Biological processes enriched for include signalling (62) and vesicle mediated transport (39). Mapping KEGG pathways included synaptic pathways for Serotonergic, cholinergic, glutamatergic, dopaminergic synapses and several other neuronal signalling-based pathways.

4.3.2 Analysis of substantia nigra from Parkinson's patients' and control patients' (TMT dataset: PXD000427)

Analyses of a publicly available dataset (PXD000427) was performed utilising PEAKS software. The dataset consisted of substantia nigra samples from three healthy controls (70M, 61F, 87F) and from three Parkinson's disease patients (73M, 79M, 85F). All associated project files can be found under Appendix 4_2, including entire Peaks project with subsequent analyses. The analysis resulted in the identification of 17,247 peptides and 6,609 proteins with 2,579 indistinguishable protein groups (Table 2), whereas the original publication alongside this dataset reported the identification of 1,795 protein groups. The increase in protein identification is likely a function of the Peaks studios *de novo* sequencing power and is not in dispute of the initial publication's findings. The average age for the control patient group was 72.7 ± 13.2 (standard deviation [SD]) and for the Parkinson's group 79± 6 years, who had a disease duration range of 2-16 years. Post mortem intervals were 24 ± 11 hours (SD) for control and 18.7 ± 10.1 for Parkinson's disease.

Table 2: Overview of dataset produced by reanalysis of PXD000427.

		Total
#Scans	MS1	94645
	MS/MS	175638
#Features		507952
Identified	#PSMs	61049
	#Scans	61049
	#Features[**]	33519
#Peptides		17247
#Sequences		16097
#Proteins[*]	Groups	2579
	All	6609
	Top	3139

Columns: **Features** are precursors identified with a trace. **Scans** refers to spectra generated at the MS1 and MS/MS levels and **Features** here refers to features with an associated identification. **Identified** have associated matched information: peptide spectral matches **PSMs**, **Peptides** refers to the number of peptides whilst sequences are the unique sequences identified. **Sequences** are stripped sequences supported by peptides regardless of modification. **Protein Groups** refers to matches of proteins that are indistinguishable, **all** proteins refer to matches regardless of distinguishing peptides and **Top** refers to the high confidence matches.

Quantitative analysis of this dataset was hampered by the data curator not providing the lot number of the TMT labelling kit to allow isotope purity correction. Applying a 1% adjusted FDR and stringent filtering resulted in only one protein being significantly differentially abundant and it was alpha-1-antitrypsin at a ratio of 0.46 to the control expression. The quoted publication set the analysis as using a log2 fold change of 1.35 as the threshold for differentially expressed proteins, and further to this normalisation was performed based on the mean with CVs assessed based on an internal standard, which allowed the assessment of sample variability. The original paper lists 204 proteins as statistically significantly changing. The analysis performed used similar cut offs as presented by the initial paper, with a -10logp of the p-value of 20 was equalling a p-value of 0.01, two unique peptides used for quantification and a peptide level FDR of 1%. Manual interpretation of the LacB spike protein intensity used as a normalisation candidate confirmed no dramatic variation as reported by the original authors and as such the standard TIC method was used for normalisation in Peaks Studio. Quantitative analysis using the authors defined cut-off identified the housekeeping protein glycerol-3-phosphate dehydrogenase as being upregulated in Parkinson's as well as a subunit of ferritin and aquaporin (Table 3). Proteins decreased in abundance included the alpha anti-trypsin protein and, of Parkinson's biological significance, tyrosine 3-monooxygenase, the enzyme responsible for the conversion of tyrosine to L-DOPA. Several of the proteins found to be differentially abundant are found in circulating plasma and are highlighted in red font, this is important as blood contamination is a known issue in proteomic preparation as the protein content is greater than that of tissues. The primary component human albumin was not found to be differentially abundant, if contamination by blood had occurred it is likely that protein albumin would also vary in the same manner as the identified blood components. The lack of statistical difference in the amount of quantified albumin must be interpreted with caution as proteins that dominate the ion population are difficult to quantify, due to ion suppression and dynamic range factors.

Table 3: Proteins of differential abundance in the substantia nigra of Parkinson's patients versus healthy age-matched control patients.

Accession	Significance	-10lgP	Ratio Parkinson's	#Unique	Description
P21695	20.63	141.1	1.62	8	Glycerol-3-phosphate dehydrogenase [NAD(+)] cytoplasmic
P02792	23.11	143.3	1.61	7	Ferritin light chain*
V9PBN7	22.19	173	1.49	16	Aquaporin 4 isoform delta4
P55087	22.19	173	1.49	16	Aquaporin-4
P02675	20.57	143.3	0.66	5	Fibrinogen beta chain
D6REL8	20.57	104.2	0.66	3	Fibrinogen beta chain
C9JEU5	24.58	164.2	0.6	9	Fibrinogen gamma chain
P02679	24.58	164.2	0.6	9	Fibrinogen gamma chain
C9JC84	24.58	164.2	0.6	9	Fibrinogen gamma chain
C9JPQ9	24.58	83.19	0.58	2	Fibrinogen gamma chain (Fragment)
C9JU00	24.58	83.19	0.58	2	Fibrinogen gamma chain (Fragment)
P07101	39.45	165	0.55	9	Tyrosine 3-monooxygenase
P01834	40.24	166.6	0.54	6	Immunoglobulin kappa constant
P02671	33.05	154.2	0.53	5	Fibrinogen alpha chain
P02763	23.57	126.8	0.49	2	Alpha-1-acid glycoprotein 1
P01009	68.84	201.6	0.46	18	Alpha-1-antitrypsin
A0A024R6I7	68.84	201.6	0.46	18	Alpha-1-antitrypsin
P00738	25.88	192.1	0.45	2	Haptoglobin
J3QLC9	25.88	187	0.45	2	Haptoglobin (Fragment)
H3BS21	25.88	139.7	0.45	2	Haptoglobin (Fragment)
J3KRH2	25.88	126.8	0.45	2	Haptoglobin (Fragment)
J3KSV1	25.88	98.86	0.45	2	Haptoglobin (Fragment)
J3QQI8	25.88	98.86	0.45	2	Haptoglobin
H3BMJ7	25.88	98.86	0.45	2	Haptoglobin
A0A0G2JRN3	68.84	191.4	0.44	16	Alpha-1-antitrypsin
G3V2B9	68.84	115.1	0.41	5	Alpha-1-antitrypsin (Fragment)

Descriptions in red are circulating blood components, a small non-statistically significant change in abundance of human albumin (ratio of 0.95 in PD relative to control). *Ferritin was indicated as being expressed by glial cell populations within the original manuscript

Analysis of the quantitative PSMs (Peptide spectral matches with reporter ion measurements) from the dataset was performed for peptides containing an Ascore of 1000 for each modification generating a ratio of area in the Parkinson's samples to control (Figure 8). The number of PSMs identified for Ox-Phe was 91, 65 for PB-DOPA and 509 for oxidised methionine. Analysis of the means revealed no statistical difference across all three modifications in comparison to control (ratio of reporter ions).
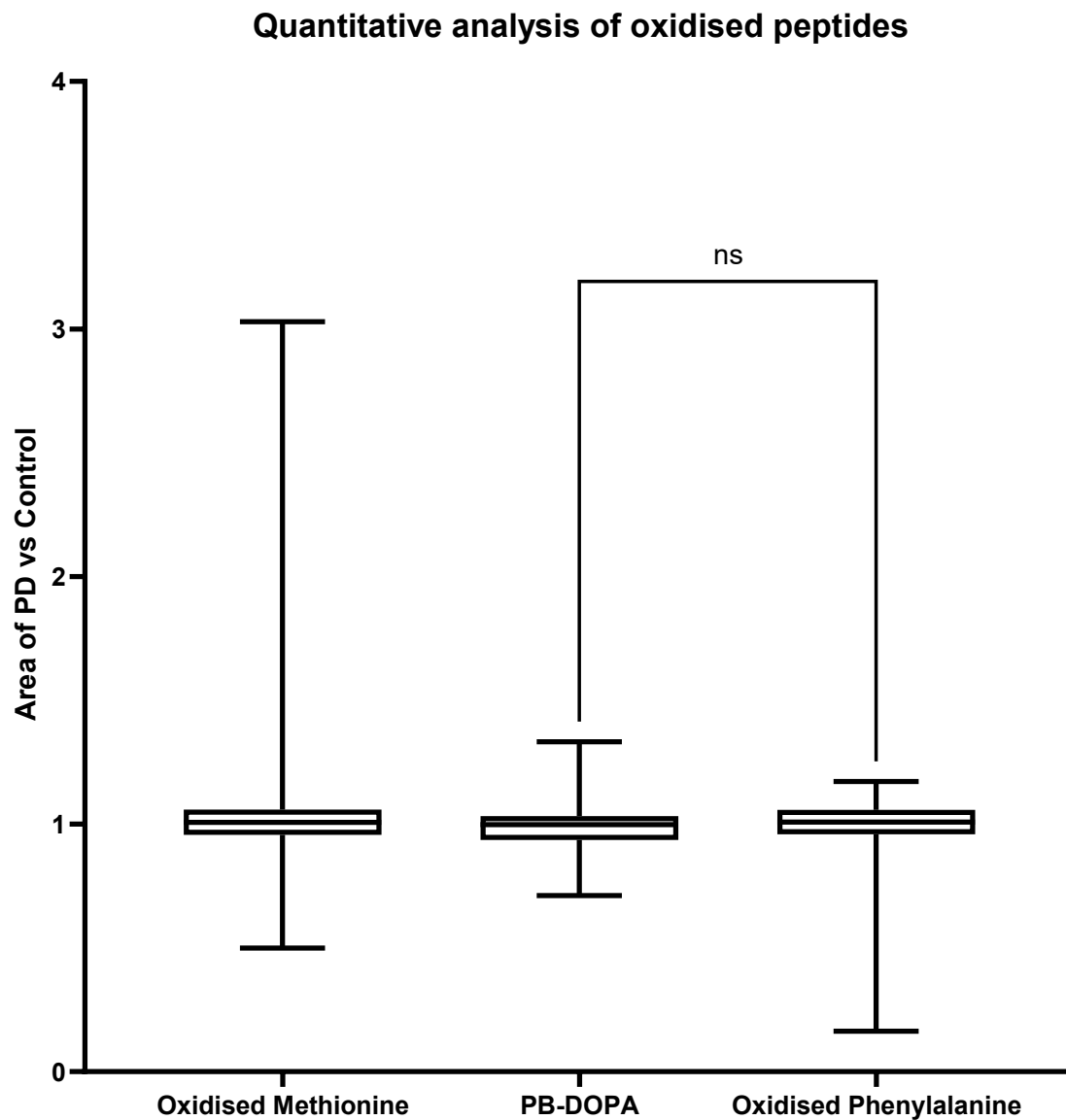
Figure 8: Tukey boxplot representing the quantitative amount of each oxidation site in PD relative to the control. No statistical difference in the means were found (students t-test). The number of PSMs used for each analysis were 509 (Oxidised Methionine), 65 (PB-DOPA) and 91 (Oxidised Phe).

Analysis of peptides containing confident Ascores of 1000 was performed (Ascore is an ambiguity score developed for phosphorylation site localisation, it is a -10logp, with 1000 being the highest assignable certainty [221]) for PB-DOPA and Ox-Phe, resulting in the identification of 46 peptides containing PB-DOPA, 47 containing oxidised phenylalanine and a further overlapping 10 peptides containing confident assignment of both modifications. Analysis of the peptide's protein assignments resulted in 111 PB-DOPA, 126 Ox-Phe and 31 dual mapped accessions which were then used to generate a Venn diagram (Figure 9). All data for this section can be found in (Appendix 4_2_2-C: Supplementary file 1).
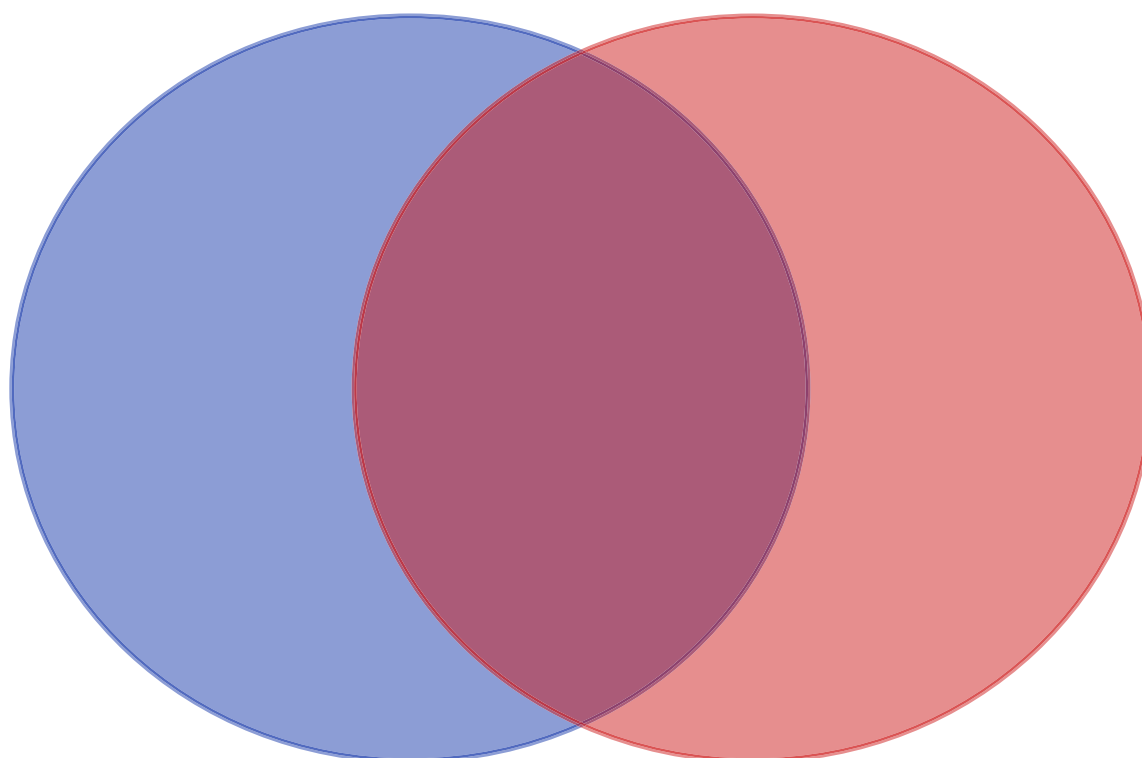
Figure 9: Venn diagram of proteins containing PB-DOPA and Ox-Phe from the PXD000427 PD-TMT dataset. This was created using all accessions mapping to all peptides for each respective modification.

Network analysis was performed for the proteins containing both Ox-Phe and PB-DOPA, with 48 accessions mapped to 24 STRING identities (Figure 10). The majority of the proteins were structural proteins including tubulins, actin and the myelin proteins (PLP1 and MBP). The strongest enriched

term for location was the internode region of the axon (2), dense body (2) and the myelin sheath (4). Enriched biological processes included; substantia nigra development (3) with the identities being PLP1, MBP and actin cytoplasmic 1 (ACTB). (Permalink: https://version-11-5.string-db.org/cgi/network?networkId=bIsj3ujegYw3)
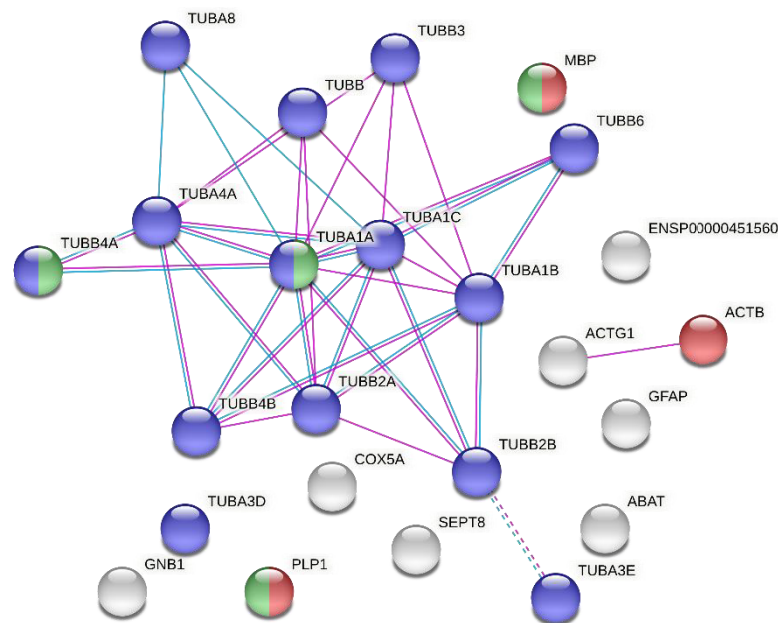
Figure 10: STRING network of proteins annotated to contain both PB-DOPA and Ox-Phe. The 48 unique accessions were mapped to 24 string identities with the majority being tubulins. Blue: Parkin-ubiquitin proteasomal system pathway (WP2359). Red: Substantia nigra development (GO:0021762). Green: Myelin sheath (GO:0043209).

Network analysis was performed using STRING DB with the 63 unique accessions mapping to 27 identities (Figure 11). (Permalink: https://version-11-5.string-db.org/cgi/network?networkId=bjyBq6o9PJlJ). This network contained a central node of heat shock protein 90 alpha with an interaction cluster involving haemoglobin (HBB), and this core cluster is annotated to be chaperone mediated autophagy (HAS-9613829) which indicates that these proteins containing both modifications are exposed to oxidative damaging conditions. Several proteins involved in hydrogen peroxide catabolism and KEGG disease pathways (Parkinson's, Prion, Amyotrophic lateral sclerosis) were also identified. Myelin proteolipid protein was also identified as containing PB-DOPA by a unique accession and is a protein involved centrally in multiple sclerosis.
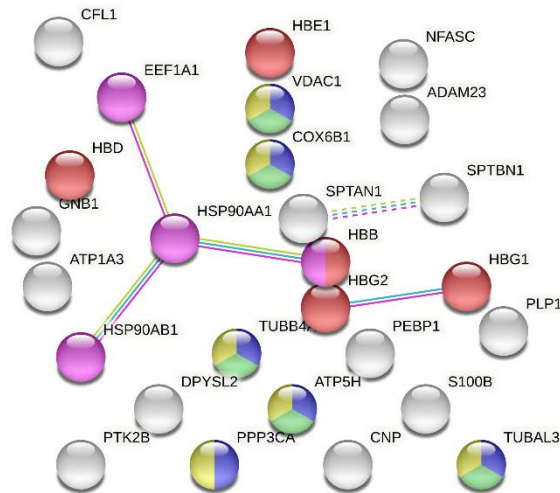
Figure 11: STRING network generated from PB-DOPA accessions that were not mapped to contain Ox-Phe. Network contains 27 mapped identities. Pink: Chaperone mediated autophagy (4). Red: Hydrogen peroxide catabolic process (5). Yellow/Blue/Green: Amyotropic lateral sclerosis, prion and Parkinson's disease.

Network analysis was performed using the 78 oxidised phenylalanine exclusive accessions matching to 31 identities within STRING (Figure 12). (Permalink: https://version-11-5.string-db.org/cgi/network?networkId=bypPaRm1p31v) A central tubulin interaction network was identified and the structural myelin basic protein (MBP) identified by a unique identifier, which is the protein involved in multiple sclerosis. The majority of the network proteins were annotated to be exosomal proteins and several neurodegenerative disease pathways were also enriched due to the tubulin core.
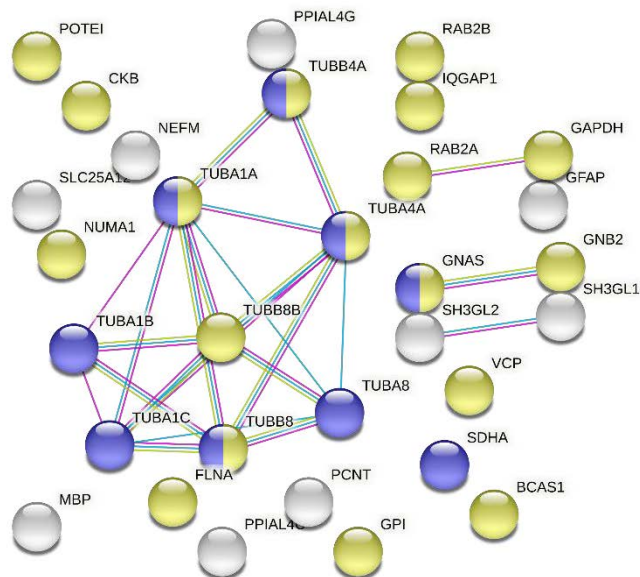
Figure 12: STRING network of proteins annotated to containing Ox-Phe by unique accession, 31 identities mapped from 78 accessions. Yellow: Extracellular exosome (#18). Blue: Parkinson's disease (#9, hsa05012).

4.3.3 Olfactory proteome: analysis of label-free quantitative dataset

Olfactory dysfunction precede the more pronounced symptoms of movement issues in Parkinson's by several years, as such it is a primary candidate for identification of biomarkers that could allow earlier diagnosis. Analysis of the Olfactory LC-MS/MS dataset resulted in the identification of 3,818 protein groups (12,143 proteins) with 34,191 sequences, 47,049 peptides and 913,928 PSMs. Individual samples reproducibly identified similar numbers of peptides and proteins across the sample set (Appendix 4_3). The analysis of PXD008036 for oxidation levels on peptides and at Met, DOPA and Phe residues produced no significant differences for any modification (Figure 13). The small number of control samples included in this dataset provides a small statistical power. The range of oxidation for methionine and phenylalanine are wider than that of the control, and further investigation of underlying factors that could be affecting these levels are explored below.
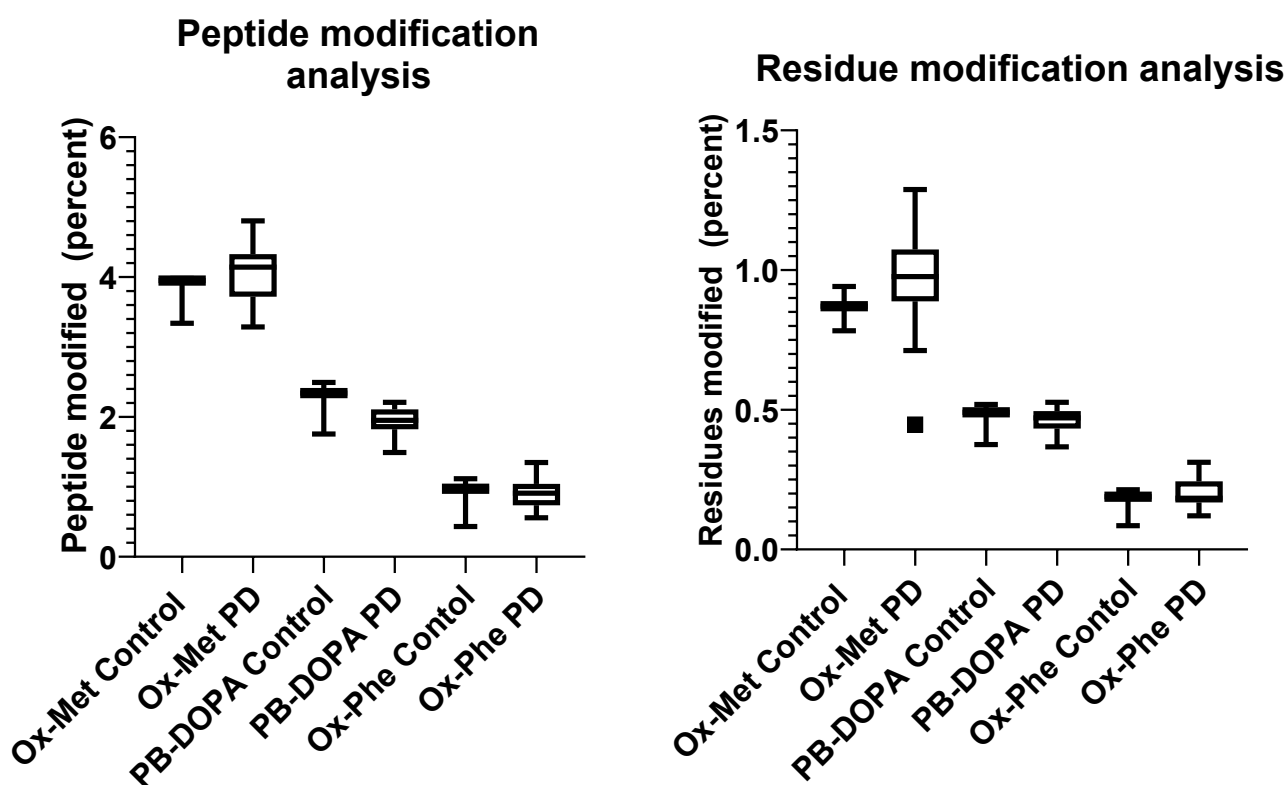
Figure 13: Boxplots of percentage modification compared at the peptide level (left) and the residue level (Right). PB-DOPA is compared against tyrosine residues or peptides. No statistical difference was determined between the controls and Parkinson's samples at any of the modification sites or at either the peptide or residue level. Three controls and 21 Parkinsons samples were included in this dataset.

Investigation of the relationship between post-mortem interval and amino acid oxidation produced no significant correlations for any of the three amino acid modifications. The three control brains used in this dataset contained pathologies that have influenced the levels of protein damage in group

comparisons; C008 was a female aged 93 that had "age-related changes", C048 had "microvascular pathology", and C064 had "cerebellar infarcts". The high level of C048/46 hydroxylation may be incompatible with use as a negative control.

Site specific modification is displayed for peptide (Figure 14) and residue (Figure 15), and the three controls displayed a distribution across the entire range of samples recorded for the Parkinson's samples. Sample C008, C048 and C064 had an increase in post mortem interval and demonstrate an increase in Ox-Phe and PB-DOPA relative to the PMI, which called for correlation analysis of the underlying clinical variables (see following page). Assessing the modification at the peptide and residue level identifies a similar pattern in the Ox-Phe/PB-DOPA levels across all samples, albiet at a lesser level indicating that singular peptide modifications are more likely than multi-site modificaitons. There was an observered variation in the Ox-Met for PD-295/PD-423 suggesting that a peptide may have more (PD423) or less oxidation (PD295) at multiple methionine positions.

Percentage of site containing peptides modified per sample
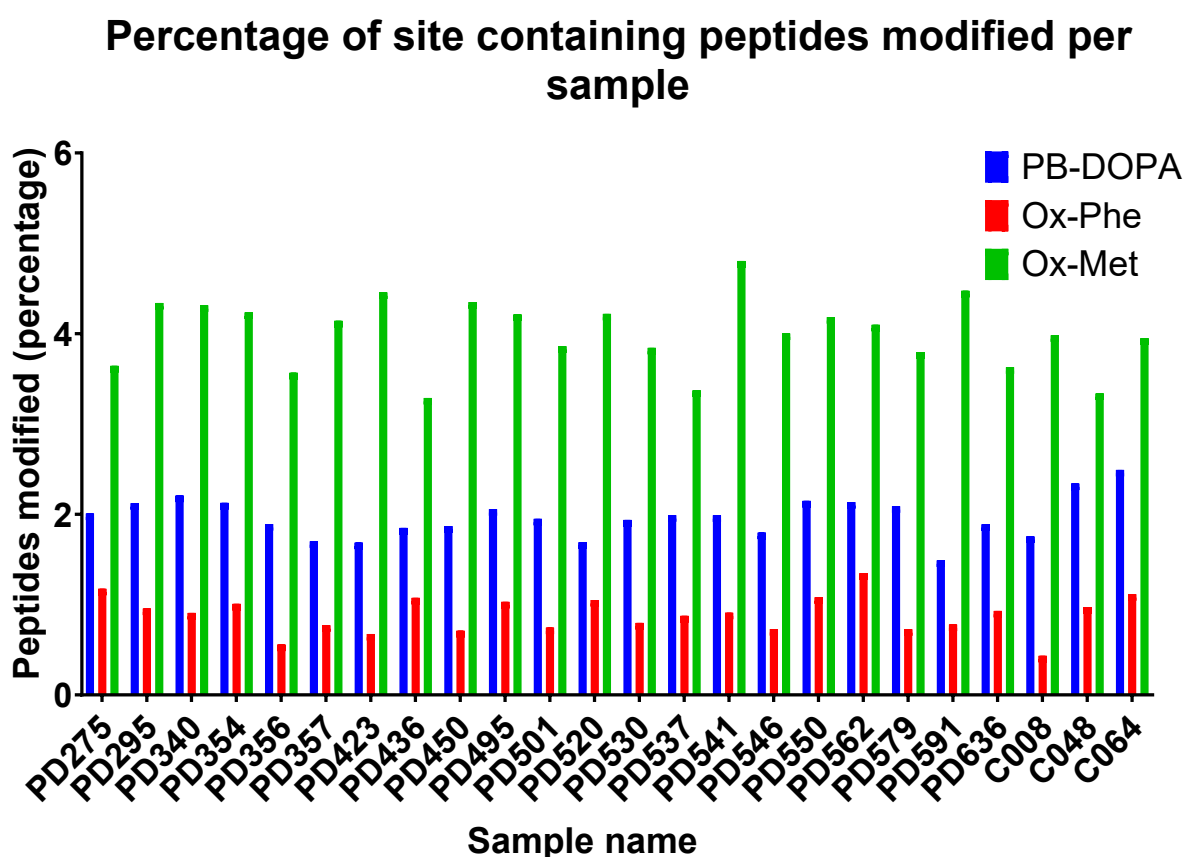
Figure 14: Percentage of modified peptides relative to peptides containing the amino acid site for modification. A high variability is observed, the control samples (C008, C048, C064) range from lowest to highest percentage for both Ox-Phe and PB-DOPA. Their appears to be no relationship of PD and PD-DOPA percentage, but low sample size may play a factor.
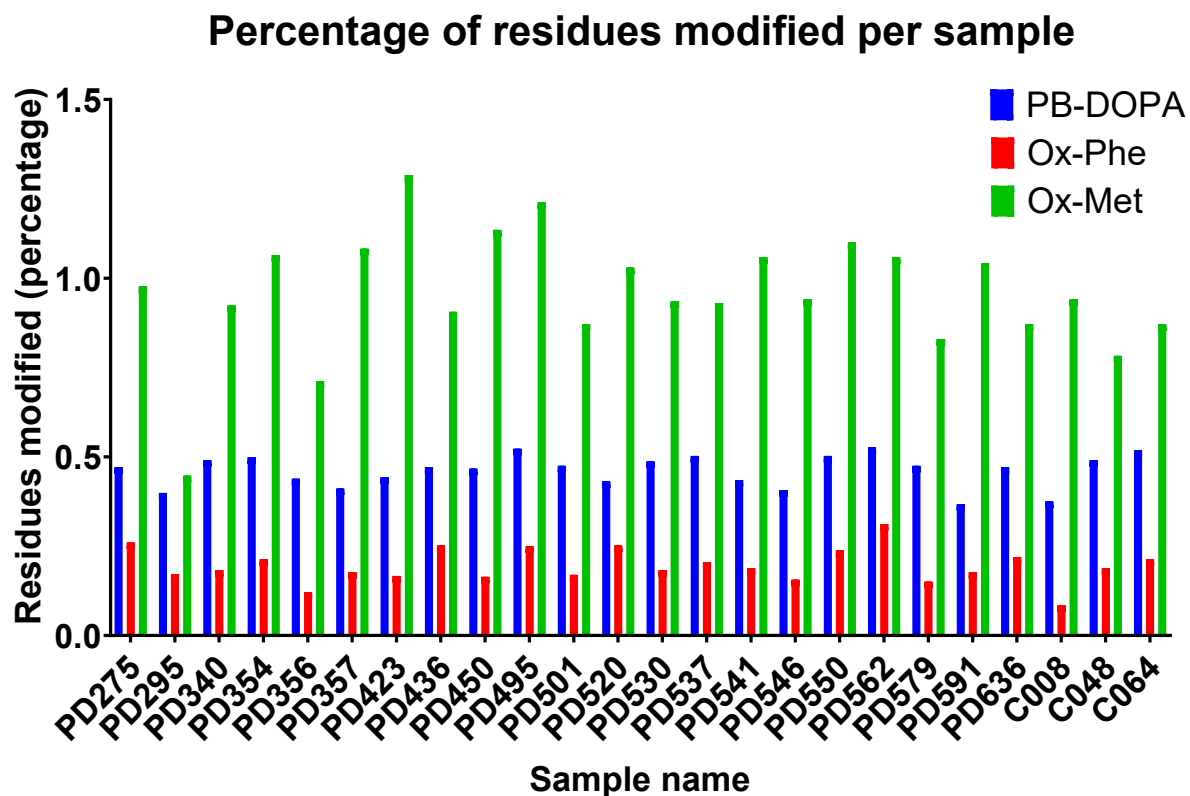
Figure 15: Percentage of modified residues containing the amino acid site modification compared to all sites of that amino acid. Residue level analysis displays a similar profile to the analysis performed at the peptide level, two measurements do not follow a similar trend which is samples PD-295 and PD-423 which both display a different Ox-Met status.

Correlation analysis was performed for the clinical information supplied alongside the dataset, including post mortem interval (PMI), disease onset age and disease duration and these were compared against levels of PB-DOPA, Ox-Phe and Ox-Met. The two groups were analysed separately. The Pearson's correlation analysis (Figure 16) and two tailed t-tests were performed at a 95% confidence interval with some correlations having statistical significance including: Y_by_residue and F_by_residue, age and onset and duration and onset.

Pearsons correlation of Parkinson's variables
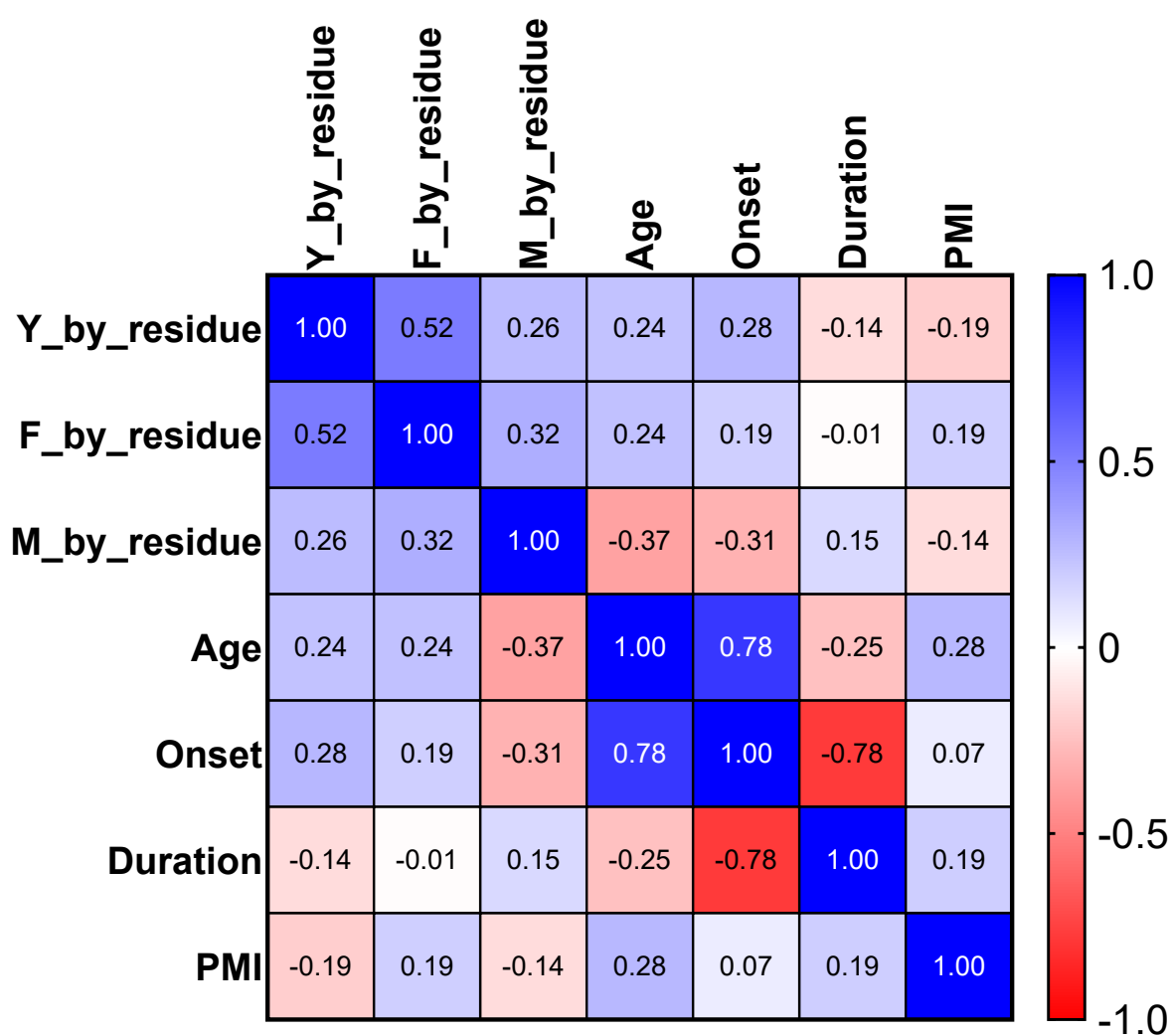
Figure 16: Pearson's correlation analysis of Parkinson's sample attributes against oxidation sites by residue (Y= PB-DOPA, M= Ox-Met, F= Ox=Phe). Oxidation of Phe and Tyr were positive correlated with a p value of 0.017. Onset and age were positively correlated with a p-value of 2.87E^-05 and duration and onset negatively correlated with a p-value of 3.62E^-05. Duration = length of time diagnosed with PD. Onset = age of disease onset.

Correlation analysis was performed on the controls separately from the Parkinson's samples to prevent interference of disease state in the underlying analysis (Figure 17). Strong positive correlation was observed for oxidation on tyrosine by residue (PB-DOPA) and the oxidation of phenylalanine by residue, with a p-value of 0.002. Strong negative correlation was observed between Age:PB-DOPA (p-value:0.018) and Age:Ox-Phe (p-value:0.016).
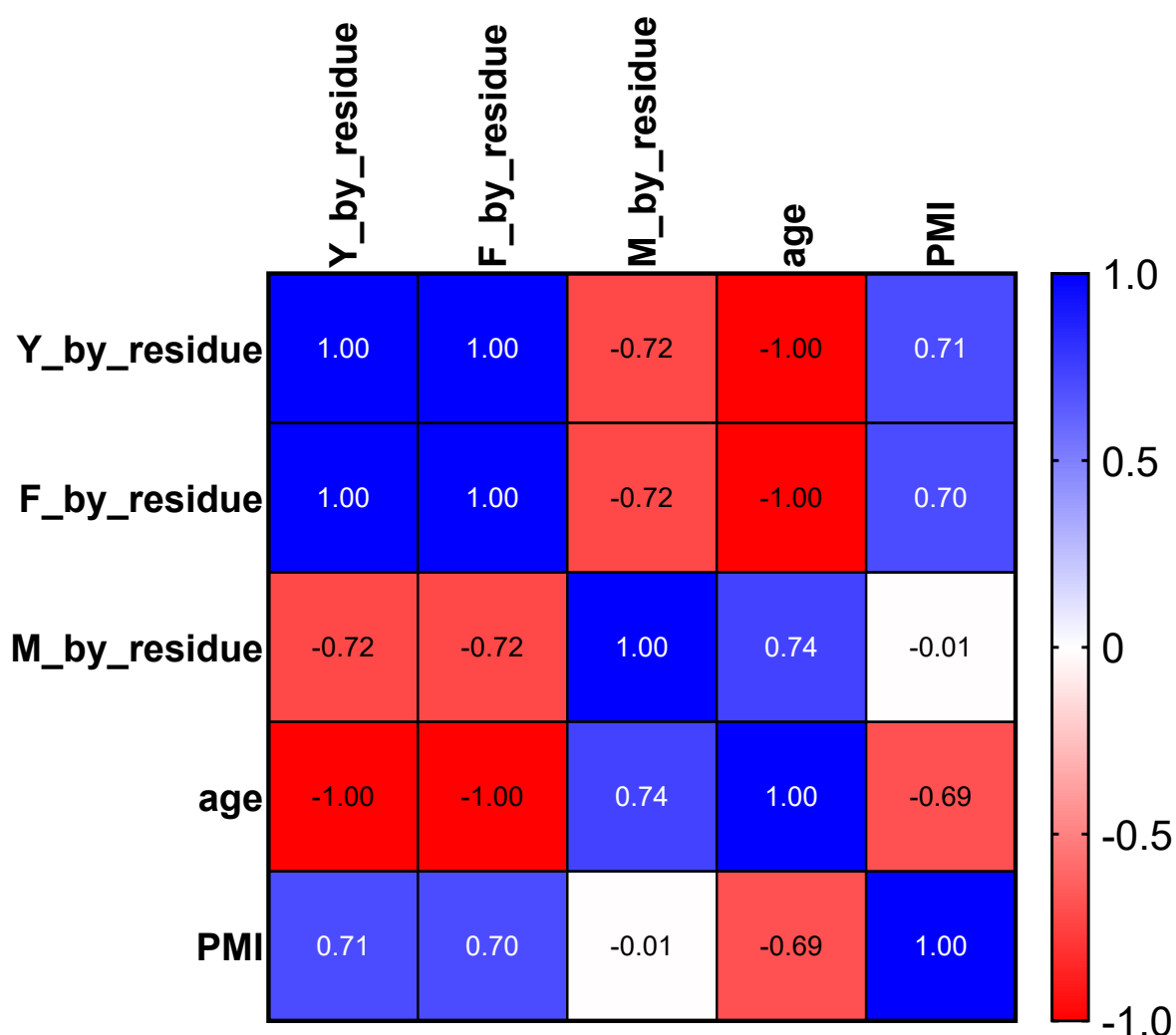
Figure 17: Pearson's correlation analysis of control sample attributes against oxidation sites by residue (Y= PB-DOPA, M= Ox-Met, F= Ox=Phe). PB-DOPA and Ox-Phe were positively correlated with each other with statistical significance and age negatively correlated against both of these.

Peptide centric analysis was performed to determine the presence of PB-DOPA and Ox-Phe, and there was only a single peptide sequence meeting the cut-off criteria that contained a single site of PB-DOPA on a sequence in the control group, K.TY(+15.99)SWDNAQVLLVGNK.C, mapping to Ras-related protein Rab-3A (P20336), which plays a role in regulated exocytosis and secretion. Given the decreased spectral quality across the entire dataset (manual interpretation) and the lack of modifications assigned as exclusive to a condition, further analyses were not performed.

4.4 Discussion

In chapter three, the presence of L-DOPA as a protein constituent was established following treatment of a human neuronal cell line with L-DOPA, and changes to the proteome were identified consistent with oxidative damage to proteins. The present chapter aimed to assess the presence of L-DOPA and Ox-Phe in human samples as a baseline occurrence and to investigate if there is a shift towards increased L-DOPA over Ox-Phe in PD tissues that would be consistent with incorporation of L-DOPA into proteins. Data open access requirements alongside publications has allowed researchers to re-analyse and interpret datasets further, extending the usability of high-quality experiments and investigation of alternative hypothesis, referred to as making data access Findable, Accessible, Interoperable, and Reusable (FAIR) [222]. Herein we have reanalysed three datasets for their oxidation profile, the first being the human draft proteome's brain samples [219], then a TMT substantia nigra Parkinson's vs control dataset [44], and finally a label free proteomic analysis of the olfactory lobes of healthy controls and Parkinson's patients [223]. The choice of these three datasets was made on a biological and technical basis, the initial human dataset providing a baseline for studies of human proteome research. The substantia nigra TMT dataset was used as it encompassed clinically relevant samples and utilised what is normally the most quantitatively accurate method of untargeted shotgun proteomics [224]. Finally, the olfactory lobe analysis was chosen after the TMT dataset produced no unique peptides containing L-DOPA and was a label free methodology. All three datasets were from fresh frozen brains rather than formalin fixed tissues.

4.4.1 Human draft proteome reanalysis (Brain)

The draft map of the human proteome utilised extensive fractionation to delve further into the proteome than had ever been attempted before, providing evidence for the existence of a large amount of the predicted proteome, and this data was the cumulation and foundation for proteomic studies resulting in advances in tools and is still an ongoing pursuit today [225]. The brain subsection of the dataset was analysed, containing three adult frontal cortices and two foetal brains. Reanalysis of this dataset was performed using open PTM searching in the PEAKS Studio software which provides a depth of information that has not been published to date on this dataset. This is also the first interrogation of aspects of the oxidative state of the samples that has been performed. It is known

that post mortem interval can affect tissues and result in oxidative damage [226], but given there is no reported PMI for these samples, it is not possible to derive this impact beyond a statement in the paper referring to these samples as being a part of a rapid sample acquisition program that aimed to decrease these effects [219]. Furthermore, oxidation can occur during sample preparation further hampering analysis of oxidative modifications [227].

4.4.1.1 The draft human proteome as a model dataset

The proteomic depth achieved in this landmark study was revolutionary at the time and still to this day has achieved a depth rarely seen. Given that these samples were chosen because of their histologically normal status it was intended that this dataset would provide a baseline measurement of what could be expected for the basal levels of PB-DOPA in neuronal tissues as well as identify general changes to the oxidative status of proteins. Our results indicated extensive oxidation had occurred in all samples. With the exception of adult brain 2, the methionine oxidation level was over 40%, with adult brain one having 70% of methionine residues oxidised, a high level of oxidation similar to that observed in chapter three (Steele *et al*, 2021. Figure 3 [228]) upon the enzymatic conversion of samples rather than in the biological treatments. Methionine oxidation levels can be used as a measure of the general oxidation status of a sample and can be attributed to both *in situ* biological sources [210] and also sample preparation artefacts [229]. The range of PB-DOPA was between 0.4% to 1.5% of tyrosine residues and similarly the range of oxidised phenylalanine was between 0.6% and 1.5% of phenylalanine residues, which are equivalent to the ranges observed in Chapter Three within the biological experiments (Chapter Three: Figure three). These results suggest that these modifications are maintained at a low level and are less susceptible to oxidative modification during proteomic analyses. The relatively low percentages of total PB-DOPA does indicate the need for an enrichment method to better identify these species in future analyses. As discussed in Chapter Three, low abundance modifications are not readily detected and the level seen here is less than phosphorylated protein concentrations which are routinely analysed using enrichment methods.

The ratio of PB-DOPA to oxidised phenylalanine may be an important metric in understanding whether incorporation or oxidative stress has resulted in the formation of PB-DOPA. The present data indicates that hydroxyl attack on phenylalanine occurs at a rate similar to that on tyrosine, with the methionine oxidation level inferring the general oxidative state. The results indicate a varying degree of PB-DOPA presence compared to Ox-Phe, however there was a general trend showing that increased levels of methionine oxidation correlated with an increase in both Phe and Tyr oxidative modifications. Given that these are histologically normal samples, the shifts in the PB-DOPA ratio may not reflect a pathology and inform a possibility that this level of variability between individuals is biologically normal. Also, given the extreme level of methionine oxidation observed (as compared to the biological

treatments in chapter three, Figure 3), there appears to not be a similar dramatic increase in the oxidation state of tyrosine or phenylalanine suggesting these modifications may be a more stable representation of what is a biological profile of oxidation. Furthermore, as methionine is a key free radical scavenger in the proteome it is more susceptible to modification [230] due to it containing sulphur, similarly cysteine also contains sulphur and its di-sulphide bridge formation can be interrupted causing protein structural changes [231, 232].

The absence of well-defined sample history also affects the ability to draw any information about whether these samples are healthy or affected to different degrees by ageing as it has been noted previously that age related declines result in oxidative stress in the brain [203]. This is a drawback of re-interpretation of public datasets as a large amount of sample information is missing, not just the history of clinical samples but also omitted steps of sample processing and acquisition.

The possible reasons for the large degree of methionine oxidation and the variability of these levels could be the manner in which the samples were fractionated, with two of the samples fractionated by gel-based electrophoresis showing a higher level of oxidation of phenylalanine then PB-DOPA while having a similar amount of oxidised methionine. The electrophoretic buffers may have been responsible for the high levels of oxidation as they contained 100mM DTT and were run for a long electrophoretic time. A way to ensure this does not happen in gel-based separation is to use thioglycolate as a free radical scavenger in the cathode solution [233]. Other sources of non-biological oxidation may be the amount of acid contained in the loading buffers and the time that samples were waiting to be acquired in an autosampler of the LC-MS/MS system or time spent digesting with trypsin. In this particular dataset no steps were taken to reduce or prevent sample oxidation, which brings into question the validity of re-interpretation of certain datasets and sets a benchmark that samples should be processed in a universal manner when comparing oxidative status.

Given the possible dual source of oxidation within a sample, either by biology or sample preparation, there has been recent work in which the proteome was artificially oxidised using heavy hydrogen peroxide [229, 234], similar to the objectives of chapter 3. This may prove a viable way to analyse the proteome by artificially converting the sample to have the heavy oxidation in unmodified peptides, allowing relative quantitation against the biologically modified species within a single scan, as the two isobaric species will have the same retention time properties [229]. The application of comparing isobarically labelled peptide species produces highly accurate information and is the foundation for TMT based quantification [235]. In the same manner it may be possible to employ the tyrosinase conversion procedure developed in the previous chapter under heavy oxygen atmosphere, as the conversion procedure consumes oxygen dissolved within the buffer and performing the procedure by

bubbling through heavy oxygen will result in isobaric species that now have PB-DOPA that can be compared to natively occurring PB-DOPA containing 'light' oxygen to study a relative species ratio (see chapter three for further explanation).

4.4.1.2 Network analysis of proteins containing PB-DOPA and Ox-Phe

In order to understand the pathways that could be affected through oxidation of specific proteins, network analysis was performed. Proteins localised to parts of the cell that produce ROS are more likely to be modified; these include proteins in and around the mitochondria as well as those associated with peroxisomes/lysosomes [236]. As observed in Figure 4 was a similar number of proteins identified as exclusively containing either PB-DOPA (344) or Ox-Phe (438), the overlap being large at 233 protein accessions which suggests basal levels of modification and the potential that these proteins can be considered as housekeeping in subsequent study analyses.

4.4.1.2.1 Network analysis of proteins containing both PB-DOPA and Ox-Phe

While protein oxidation during sample processing was possible in the samples analysed, it is likely that some protein oxidation would also have occurred *in vivo*. In these control samples the proteins found to contain both PB-DOPA and Ox-Phe are likely to be a product of oxidative stress since exogenous DOPA has not been supplied. The central network consists of the tubulin and actin proteins of the cell, with most of the proteins identified being annotated as vesicle or extracellular vesicles in location. The localisation of proteins to be packaged or extracellular is likely a reflection of proteins that are found exported from the cell and are more likely to contain damaged residues by ROS attack [237] especially in the extracellular space and mitochondria [238] or in association with the lysosome [106]. Some of the proteins found to be common are proteins related to disease including the myelin basic protein and proteolipid protein (MBP, PLP1), both of which are involved in multiple sclerosis, an autoimmune disease characterised by the destruction of the myelin sheathing of the CNS, and modification of the protein structure by such PTMs may form antigenic species [85, 86, 239, 240]. Proteins found to have both PB-DOPA and Ox-Phe included thioredoxin-dependent peroxide reductase which is primarily responsible for the reduction of hydrogen peroxide, the primary oxidative radical [241]. Cellular structural proteins such as actin and tubulins are modified in both ways and this could be due to the long-lived nature of the protein structures accumulating the ROS damage over time. These structures may be considered markers of oxidative attack and are possibly affected by post-mortem decay of the cell and the release of mitochondrial oxidative potential upon the cell death. It is also known that tubulins participate in thiol based redox reactions with the protein glyceraldehyde-3-phosphate dehydrogenase (GAPDH). In experiments by Landion *et al* (2014) [242] it was found that tubulin reduced oxidised cysteines by 50% on GAPDH.

4.4.1.2.2 Network analysis of proteins exclusively containing PB-DOPA

Analysis of the DOPA containing proteins (Figure 6) identified histones as being the centre of the network and linked to the ubiquitin-proteasomal system (UPS). The proteins involved in the UPS are inherently more likely to be modified by ROS attack, as such future analyses should discount these for misincorporation based PB-DOPA unless stable isotope L-DOPA is supplied *in vitro* [57, 66, 87, 108, 124, 164, 172, 208]. The majority of proteins identified are found to be to be binders of either protein or RNA, suggesting that these proteins are more likely to be coming into contact with other biomolecules across the cell, leading to more chance of interaction with ROS species. The likelihood of mitochondria-related proteins containing more oxidised or PB-DOPA proteins has been reported previously by our laboratory upon treatment of neuroblastoma cells with exogenous L-DOPA, resulting in up regulation of the endosomal-lysosomal degradation and a decline in mitochondrial function [105, 107, 108, 123, 164, 170, 172], which may become exacerbated upon death. As reported above, two of the adult brains (Adult 1 and 3) had a large amount of PB-DOPA and Ox-Phe in comparison to the other samples and this may be a function of the donor's age. The annotation to extracellular space for many of the proteins further reinforces that some of these proteins may be undergoing oxidative attack in the extracellular space or conversely are exported to the extracellular space after damage.

Myelin basic protein was identified again in the PB-DOPA exclusive dataset by a unique accession, as multiple accessions represent the protein (splice variants/isoforms). The inclusion of these variants is justified as unique sequences that may define a particular isoform may contain a modification that otherwise would have been missed, and in this case, we can consider the different identities to be representative peptidoforms. Utilising the abbreviated and non-redundant SwissProt database may produce a more simplistic identification result for use in quantitative studies, but by doing this a large number of proteoforms are missed and this further hampers the identification of PTMs/MiPs as sequences that contain these modifications may not have been included. This information is utterly essential in moving forward as a field and when performing peptide centric MiP/PTM analyses as the specific proteoforms may be the disease trigger in proteinopathies [228, 243, 244].

4.4.1.2.3 Network analysis of proteins exclusively containing Ox-Phe

The proteins containing Ox-Phe are formed by hydroxyl attack on phenylalanine post protein synthesis [208], and these proteins are likely to be found in regions of the cell that are exposed to high levels of oxidative stress or free radicals [208], with a third option being partial degradation inside of lysosomes or faulty accumulation [245]. Of the proteins identified to contain Ox-Phe, an over representation of proteins involved in the movement of cargo through the cell was identified. This, along with the statistical enrichment for proteins localised to synapses of different types including dopaminergic and exosomes, suggests that oxidised proteins are likely originating in cells involved in neurological

signalling and are linked to the signalling processes themselves. The identification of oxidised dopaminergic synaptic proteins, also suggests that these regions of the brain are producing more oxidative potential during the neuron's life and post mortem which is possibly due to higher intracellular quinone synthesis [246, 247]. The release of catecholamines into synaptic clefts is a possible source for oxidation of the exosome associated proteins, making proteins separate from this environment resulting in more profound changes to cell status and should be considered a significant perturbation. The secreted proteins containing oxidative damage provides evidence that these proteins are being distributed throughout the brain, and if these are indeed toxic proteoforms the propagation could be leading to neurodegenerative state [32, 36, 215-217, 234, 248-256].

4.4.1.3 Summary of findings on the human proteome

Despite these samples being considered healthy controls by the study's authors, several hundred proteins were found to contain PB-DOPA and Ox-Phe. Due to the unknown underlying nature of post mortem intervals and the age of the brain tissue, it is hard to unravel what factors contribute to the oxidation at these sites. Within this dataset an exceptionally high level of oxidised methionine was observed when compared to chapter three [228] and the olfactory dataset (section 4.3.1, Figure 13). This high level of oxidation suggests that the sample processing methods contributed to the sample's oxidative status, and controlling for this is required in further PB-DOPA analysis especially when assessing the validity of other published data.

Given the confounding nature of the high levels of oxidation of the samples, it is hard to place a fixed basal value on the oxidative/PB-DOPA state of the human neuronal proteome which was the aim of the analysis. However, given that the ranges in the amount of PB-DOPA/Ox-Phe does not scale proportionally with the methionine level, it is possible that these results are valid but do call for further analyses of other datasets. Analysis of a dataset that have a fixed processing method for all samples is needed to understand if Parkinson's disease and L-DOPA supplementation have an effect. Regardless of the aforementioned increase in oxidised methionine levels, the amount of PB-DOPA and Ox-Phe are reinforced to be a small percentage of the proteome, regardless of variability in sample preparation.

4.4.2 Parkinson's substantia nigra TMT dataset: PXD000427

The analysis performed on the previous dataset failed to produce a robust measurement for the oxidation status of the human neuronal proteome, whereby only suggesting a low amount of Ox-Met and PB-DOPA. As a result, the dataset PXD000427 was chosen as it did not suffer from variable sample preparation and also contained both healthy control tissues and PD tissue. Furthermore, this dataset presents a viable relative-comparison for what oxidation levels and PB-DOPA therapeutic intervention can be expected in Parkinson's disease versus control tissues. This experiment also provides the opportunity assess the usage of TMT methodology for analysis of PB-DOPA.

TMT analysis is the most accurate untargeted shotgun proteomic method due to the relative abundances of isobaric tags being directly compared in a single scan resulting in fewer missing values when compared to label free quantification [257]. TMT also has the benefit of not requiring retention time correction based on liquid chromatographic performance, this issue is apparent in inter-run sample comparisons as a precursor is considered a feature which can have other precursors of similar mass eluting close by (feature in this context refers to the peak formed from precursor ion clusters that map to a sequence without fragmentation spectra being generated).

Analysis of the dataset produced 6,609 proteins representing 2,579 groups, which is likely a function of the verbose isoform database employed (high sequence redundancy). Peaks Studio identified several hundred more proteins at the group level than the original publication, which is likely a function of the superior performance offered by the *de novo* sequencing combined with neural network approach used by the software [258, 259]. Only one complete injection set was used to perform subsequent analysis and this likely explains the disparity in quantitively differentially abundant proteins identified, which was further compounded by the lack of TMT lot number being provided as isotopic purity correction cannot be applied for the TMT reporter channels. Furthermore, this dataset was acquired using MS2 based reporter tag analysis which is not the most quantitatively accurate version of TMT analysis due to ion ratio suppression, although this method does result in a higher scan speed in tribrid instruments (this experiment used a LTQ-orbitrap which is a hybrid based instrument) prior to the advent of RTS based SPS-MS3 [224, 260-262]. Notably if using a smaller sample size, a faster and more quantitatively accurate version of TMT analysis can be performed by taking half of the reporter tags from the 16plex kit and using the unique balancers as an extra quantitative ion, resulting in two points of quantification for the precursor. Furthermore, this version of TMT does not suffer from isotopic impurities and results in a high scan speed while also resulting in better quantitative accuracy. Given the small number of samples used in this initial study, if a small cohort was to be used again this would be the best experimental approach [262].

4.4.2.1 PXD000427: Peptide modification levels for PB-DOPA, oxidised phenylalanine and methionine

Analysis of the peptide modifications of interest found no statistical difference in the ratio between the control group and the PD samples, indicating that the pathology and/or the supplementation of L-DOPA does not appear to have an effect on the amount of PB-DOPA within the substantia nigra, at least for those peptides that are common as no unique peptides were identified in either group. This result is in contrast to work performed by our laboratory [57, 193], where it was found that tissue of the substantia nigra, cingulate gyrus and corpus coliseum had an increased level of PB-DOPA (Increases of ~50%, ~25% and 65% respectively). Also reported alongside our laboratories studies was that therapeutic L-DOPA treatment of greater than 20 years in duration resulted in an increased PB-DOPA content in plasma proteins, but these results also utilised FL-HPLC and amino acid hydrolysis which provides a more sensitive level of quantification of the released PTM-containing AA than proteome level analysis (Chapter two). One further benefit for the FL-HPLC analysis detecting more PB-DOPA, is that the hydrolysis of proteins captures species that cannot be solubilised by proteomic buffer systems and are not amendable to LC-MS/MS analysis.

The lack of unique identification and differential abundance (modified peptides) may be a function of the lack of detectable unique peptides and further supports that enrichment is needed to identify proteins containing PB-DOPA due to the low level of PB-DOPA identified previously, which is akin to the study of phosphorylated peptides. Signalling studies in phospho-proteomics is not on a whole proteome without specific enrichment [263], with typical workflows beginning with greater than 500µg of total peptide to enrich for enough total phosphopeptide for a single injection of 100-1000ng. It is foreseeable that similarly large peptide mass amounts may be required for enrichment analysis of PB-DOPA containing peptides and that future studies should include a non-mass limited sample, for example human plasma which has a high protein content.

4.4.2.2 PXD000427: Quantitative analysis of the differences between control and Parkinson's disease samples

Quantitative analysis of all identified peptides, regardless of modification, resulted in 11 protein groups found to be changing in abundance, with the majority of these proteins annotated as being constituents of human plasma, potentially indicating a systematic change in the blood or contamination of the tissue during dissection. The majority of the plasma associated proteins were decreased in abundance in the Parkinson's samples which could indicate less blood contamination or else an actual decrease in the levels of these proteins. The protein ferritin was identified as being localised to oligodendrocytes and microglial cells in the initial publication which does suggest changing cellular populations confirmed by immunohistochemistry. For a more robust analysis, inclusion of blood analysis alongside such datasets should be considered to rule out these underlying

contaminants, and possibly mass spectrometric imaging of intact sections. The use of mass spectrometric imaging of sections allows the separation of peptide signal based on structural components of the dissected tissue, enabling differences in cell populations and possibly the vasculature [264]. The changes observed in these plasma-associated proteins are likely to be related to changes in the cell population since, if blood was a major contaminant, proteins such as human serum albumin and haemoglobin would have been present at higher levels in the control [265]. When assessing the raw quantitative data there was a statistically non-significant decrease in the human albumin levels of 0.95 to control which does not align with the ratios observed in the other blood component proteins, suggesting these are systemic changes and not blood contamination.

The proteins increased in abundance included glycerol-3-phosphate dehydrogenase (GAPDH) which produces dihydroxyacetone phosphate as well as NADH, and this change in abundance may be representative of an increased metabolic rate. Furthermore, it is known that GAPDH itself can suffer oxidative attack and may be a marker of metabolic disturbance [242], and other studies have used it as a housekeeping gene when studying the PD protein PARKIN which this current work indicating it to be a poor choice [266]. Aquaporin-4 was also found to be of increased abundance and is responsible for the exchange of water across the blood brain barrier and is required for the clearance of beta amyloid fibrils, based on data generated by studies on the rodent equivalent aquaporin-4 [267]. The increased need for clearance of cellular debris from the Parkinson's brain may be the reason for aquaporin-4 increasing in abundance as a compensatory mechanism, and exploration of other Parkinson's brain tissue proteomes would be required to confirm this. The protein tyrosine 3-monooxygenase (tyrosine 3-hydroxylase), also known as tyrosinase, was found to be significantly decreased in the Parkinson's substantia nigra and this is pathologically relevant, as this is one of the two enzymes responsible for the conversion of tyrosine into dopamine and was also the enzyme employed in the previous chapter for enzymatic creation of proteins containing PB-DOPA [208]. This loss of tyrosinase would likely be the direct result of dopaminergic neuron loss and could be considered in full agreement with biological analysis performed by the original authors of this data [44]. The decrease in tyrosinase with no detectable difference in PB-DOPA is likely a function of the depth of proteome analysis achieved, further fractionation or enrichment may assist in the quantification of these low-abundant peptides.

4.4.2.3 PXD000427: Proteins containing both PB-DOPA and Ox-Phe

Network analysis of the proteins containing both PB-DOPA and Ox-Phe resulted in the identification of a tubulin cluster and structural proteins of the myelin sheath (Figure 10). Furthermore, proteins of the parkin-proteosomal system were found to be modified alongside the identification of the term Parkinson's disease from KEGG enrichment, suggesting that these proteins are indeed mapped to PD

and that their modification could be at a basal level in these tissues. In the previous analysis of the human draft proteome dataset these proteins were similarly identified, lending further evidence that these cellular components are more prone to oxidation, especially when coming into contact with damaged protein species [107]. The identification of proteins related to the sub cellular location of the internode region of the axon was found, and this aligns with substantia nigra tissue development specifically as these tissues have more contact with dopamine and potentially L-DOPA. Furthermore, given that these are substantia nigra sections it is likely that there is a large amount of free L-DOPA within the cells and a highly oxidative environment is being produced [57, 193].

4.4.2.4 PXD000427: PB-DOPA only containing proteins

Analysis of the proteins containing PB-DOPA resulted in a network containing hydrogen peroxide catabolic proteins and heat shock protein 90 alpha (Figure 11). The presence of proteins involved in misfolding is more likely due to oxidative attack as these proteins are going to be present in regions of the cell where oxidised species are produced, such as around the mitochondria. The identification of proteins coming into contact with ROS sources in the cell is supported by the identification of hydrogen peroxide catabolic proteins and is in agreement with the misfolding observed in neurodegeneration and aging [1, 203]. Myelin proteolipid protein was once again identified here as containing PB-DOPA via a unique accession that did not contain Ox-Phe, meaning that this protein isoform is undergoing oxidative attack rather than L-DOPA incorporation. Several disease related proteins were identified, with the majority being tubulins that were identified to contain Ox-Phe. These data further display that identification is directed towards proteins that contain both forms of modification which further reinforces that oxidative attack of these species is likely occurring naturally in both healthy tissue and Parkinson's.

4.4.2.5 PXD000427: Ox-Phe only containing proteins

Network analysis of the Ox-Phe exclusive accessions displayed another tubulin cluster and the structural MBP (Figure 12). There were minute abundance differences including the central housekeeping protein GAPDH and several other vesicle related proteins, this again suggests that proteins are of similar function and location as to those discussed in the previous two sections. GAPDH itself has been implicated in neurogenerative diseases by being a known interactor with toxic proteoforms and can itself aggregate upon high levels of dopamine content within the cell which may be an exacerbated state in PD therapeutic treatment [268]. GAPDH being oxidatively attacked may result in a loss of function, as there is a decrease in effective amounts, the cell may increase the overall abundance of this protein to combat possible aggregate and oligomer formation [268]. The identification of oxidation on GAPDH indicates it as a possible target for further study by affinity pulldown and comprehensive characterisation of oxidation sites, this would include the use of

orthogonal proteases alongside sample processing methods that protects the native oxidation state of the protein sample [269, 270].

4.4.3 PXD000427: Summary of dataset analysis

Upon network-based analysis of the proteins possessing PB-DOPA and Ox-Phe modifications, similar terms (GO-pathway/location, Disease, KEGG) were identified in both networks suggesting that these modification sites do not correspond to a pathology but are more symptomatic of normal cell functioning. The substantia nigra is the most catechol-rich region of the brain and has a high dopamine and L-DOPA concentration relative to the other tissues of the brain [271], making it a location where incorporation could potentially occur alongside oxidative stress. The oxidative levels of both groups were the same and not statistically significant for any of the three modifications examined and given that the data is TMT labelled it is not possible to perform a direct residue or peptide analysis as per the last dataset as there are no missing values in the matrix. The lack of exclusive assignment of peptide sequences to each dataset means that performing a sample centric oxidation level analysis will not produce any difference. These data suggest that even given extensive fractionation and the use of a six-channel TMT, it is not possible to **identify** unique peptides with the oxidation or misincorporation of interest as no peptide containing a modification was unique to the Parkinson's group or control. This is likely because of the peptides containing these modifications are well below the detection limit of the TMT labelled samples.

Future studies of such tissues may benefit from the usage of a method for creation of a synthetic PB-DOPA positive proteome (as outlined in chapter three), which would be akin to the usage of booster channels employed in single cell proteomics to overcome the detection issues inherent in low abundance samples [272]. The conversion efficiency for the synthetic PB-DOPA proteome presented in the previous chapter was lower than would be needed for this proposed study but further manipulation of the supply of oxygen and use of a higher purity tyrosinase could potentially result in a higher conversion. Once a high purity synthetically converted sample is achieved utilising a pooled sample of the entire cohort, it would be possible to stack the channels of the TMT plex to contain one or more channels of this high purity converted proteome by either using multiple channels to bias sequencing towards these species or by using a single booster channel of several fold higher amount, making the sequencing based on this channel's composition specifically [273]. To further build on this logical framework, the use of booster channels that occupies higher ion content than the actual biological samples results in the ion population reaching the intensity required for triggering, this is akin to how the TOMAHAQ ("triggered by offset, multiplexed, accurate mass, high resolution, and absolute quantitation") method uses high concentration TMT-zero labelled synthetic peptides to trigger acquisition of unseen TMT ions further up the spectra by using mass-offset acquisition [274].

TOMAHAQ methodology can be applied without that absolute quant making this TOMAH, which may indeed be a method producing superior data as there is known issues in single cell proteomics whereby the use of a booster channel of >20x that of the samples results in inaccurate quantitative measurements [275]. Further to this, the use of a pooled sample exposed to hydrogen peroxide as another booster channel would further allow the detection and possible quantification of ROS affected peptides.

Other potential future directions for overcoming this lack of misincorporation detection would be to stack TMTplexes with more PD samples to bias the sequencing towards the proteome of the diseased tissue, which has the same effect by biasing the analysis to sequence proteins of higher abundance relative to the PD samples. This in essence is the dilution of control biological signal from the overall mass of the plex, resulting in low abundance peptides present in the PD samples being more likely to reach the threshold required to trigger a data dependent MS/MS scan.

Using a higher plexing capacity labelling kit will also result in further depth of proteome analysis, through combining this with current synchronous precursor selection based MS3 (SPS-MS3) or even real time search enabled SPS-MS3 (RTS-SPS-MS3) which may result in a better depth of identification and be helpful in targeting species of interest [276]. Current SPS-MS3 uses the isolation of the peptide precursors for acquisition in the ion trap for identification spectra in tribrid instruments whilst the orbitrap is used to measure the ratio of the isolated reporter fragments released from the precursor, leading to a reduction in what is termed ratio compression in MS2 based TMT experiments [277]. Current SPS-MS3 mass spectrometers including the Lumos and Eclipse also use advanced peak determination and theoretical charge envelope tolerances to ensure the species isolated is a peptide species [278] which further enhances selection of meaningful peptides. The RTS-SPS-MS3 approach utilises real time live searching of acquired Ion trap MS2 spectra to ensure that the peptide species fragmentation spectra result in an identification, and those that do not pass this step are then not acquired for quantification which is the rate limiting step in MS3 based experiments, as the orbitrap has significantly higher overhead scan time than the associated ion trap fragmentation [276]. By employing RTS-SPS-MS3 there is also the option to use a peptide close out approach whereby if a MS/MS spectrum has already been identified with statistical significance and a pre-determined number of peptides, the mass spectrometer will instead choose other more unique peptides for identification of proteins not yet identified or closed out. Furthermore, in enabling RTS-SPS-MS3, variable modifications can be set to include PTMs of interest in the database search, enabling Ox-Phe/PB-DOPA as variable modifications will increase their acquisition as these species will be identified and prioritised at acquisition. This technology is still fledgling with improvements to be made, and a foreseeable method development would be the inclusion of peptides for acquisition

based on in-silico generation of targets of interest, which could be considered global targeting lists based on peptide fingerprint generation. The reason behind the employment in global targeting schemes is primarily due to the computational overhead required to employ such methods, especially for the generation of retention time-based models for dynamic warping. The use of a program framework like that of Superhirn developed by the Aebersold lab [279] and implemented in the AcquireX [280, 281] framework would be the realised version of this technology and could lead to a revolution in "untargeted" proteomics specifically for PTMs of interest.

The samples used in this study were not well annotated in the terms of the regime of L-DOPA consumption so direct comparison of incorporation rates cannot be performed. As such it is highly recommended if performing analysis of such clinical datasets that as much information surrounding the pharmaceutical interventions and their durations is included. The lack of unique modified peptides to either condition indicated that exploration of a label free quantitative dataset may result in the more robust identification of L-DOPA misincorporation, and this would be due to unique peptides of low abundance having the chance to be acquired within individual samples runs as they would otherwise be diluted too far below the limits of detection in a TMT experiment. Biological signal dilution should be acknowledged as a drawback of labelling experiments especially moving forward in MiP/PTM analyses in the same way that sequence isomer dilutions are.

4.5.3 Olfactory Proteome dataset

Analysis of a label free quantitative dataset (PXD008036) was performed as the analysis of the previous dataset suggested that the dilution of unique biological signal in TMT prevents unique peptide identification, furthermore this dataset had extensive annotation of post mortem interval, duration and onset of disease, however no information on L-DOPA treatment was contained within this dataset. Upon manual analysis and assessment of spectral quality it was identified that the TOF based mass spectrometer did not provide clear, noise free spectra when manually inspected. As such, an increase in spectral quality filtration was used and setting the FDR at 0.1% at the peptide level and 0.01% at the protein level produced data approximate to that an orbitrap instrument produces as the matching of noised fragment ions can confuse modification assignment and leads to false positives within this dataset. This is an important point for PTM/MiP-based analysis and it is likely an overlooked point that spectral quality from a TOF should be subject to more stringent filtering. It is uncertain as to how an ion mobility cell separation may impact on the spectral quality as this may lead to less noisy spectra by removing contaminating ions, but these analyses support the use of orbitrap resolution for MIP and PTM's in future analyses. When analysing this dataset, access to a high-performance computing facility was gained and, with the Peaks Studio software running in a Linux environment,

the variable oxidation was left open to all described amino acid sites as per the UNIMOD resource to understand the impact of possible sites of modification.

Peptide modification analysis was performed at the residue and peptide level, and the distribution of modifications at both levels were comparable (Figures 13). There was no statistical significance in the percentage of modified peptides at either level, suggesting that indeed there is no significant difference in the amount of oxidised peptides. In addition, it is also possible that the small number of control samples makes it difficult to reach statistical significance especially with the variability within the three control samples being high. The range of methionine oxidation level was significantly lower than that of the chapter three dataset, with the range being 0.5% to 1.3%, and the draft map of the human proteome dataset, being from 40-70%, suggesting that these samples were not adversely affected by sample preparation to the same degree. The only other proteomic paper to give a percentage to the level of oxidised methionine produced a number of 6% in human epidermal fibroblast cells in normal tissue culture conditions [229]. The range of PB-DOPA in this dataset was from 0.4-0.5% whilst the range in the draft map of the human proteome dataset was 0.4-1.5% and the Ox-Phe levels were 0.08-0.25% in the LFQ dataset and 0.7-1.5% suggesting that the oxidative state of the draft map of the human proteome dataset is partially attributed to sample preparation.

Correlation analyses were performed for the controls and PD samples separately to investigate if factors such as age and PMI affect the oxidation state of the samples. Within the PD samples, PB-DOPA and Ox-Phe levels were positively correlated with statistical significance, suggesting a co-formation of both modifications. Samples that do not contain a positive correlation between PB-DOPA and Ox-Phe would be of interest in future analyses. The PD onset was also positively correlated with age, which agrees that increased age is a risk factor for Parkinson's disease [282] as such the age of onset and duration is negatively correlated suggesting that the older the individual the lower the survival duration with Parkinson's. Annotation of L-DOPA therapeutic dosage course alongside each sample will be essential in order to determine if L-DOPA is a factor for disease acceleration, notably a control group should be used of Parkinson's samples that had not undergone therapeutic intervention with L-DOPA rather than a healthy control to investigate effects and possible incorporation.

Within the control samples, the level of PB-DOPA and Ox-Phe were positively correlated with statistical significance, again suggesting that both modifications are equally likely and subject to a similar rate of development. The amount of PB-DOPA and Ox-Phe negatively correlated with age which runs counter to the literature that oxidative stress increases with age [203]. These data suggest that the formation of PB-DOPA and Ox-Phe correlate strongly in both the PD and in healthy control samples. Given the small sample size and the large variability in post mortem

interval and age, further studies need to be performed with larger sample sizes to overcome these confounding factors. It has been demonstrated previously that the oxidation state of tissues is dependent on PMI [283], reflected at the protein level through carbonylation and HNE species which were not identified upon PTM analysis. This further suggests that an enrichment method should be developed and employed, not only for PB-DOPA but also for the various oxidative protein modifications.

Analysis to determine the unique peptides between conditions was performed and only a single exclusive unique PB-DOPA containing peptide was identified which was within the control group. Further analysis was not performed as this, combined with poor spectral quality and low control sample size, would not offer meaningful data beyond that already outlined by the previous two analyses. Analysis of a similarly defined clinical dataset alongside a large number of controls is needed to pull apart the distribution of PB-DOPA as a component of control and diseased tissue. It is also foreseeable that PB-DOPA is not increasing in abundance within the tissues and that the cell may indeed remove these proteins efficiently.

4.5 Conclusions and Future Directions

Analysis of the three datasets has identified important considerations for the assessment of PB-DOPA as a proteome constituent, namely that; sample preparation has a dramatic effect on the oxidative status of samples, potentially masking *in vivo* oxidation; TMT may prevent the identification of unique peptide species that could arise in individuals; TOF based instruments may not offer the resolution required to perform PTM and MIP analysis.

The oxidative status of the samples attributed to sample preparation may have an effect on the status of both Ox-Phe and PB-DOPA, and a larger sample size with better designed workflow that accounts for oxidation is needed. The paramount requirements for a better designed experiment are more apt clinical descriptions of individuals and sample preparation that is not variable. This chapter outlines a need for the development of enrichment techniques specifically for PB-DOPA containing peptides alongside the utilization of artificial oxidation techniques to label with ^{18}O in order to isolate sample preparation artefacts from endogenous biological oxidation [229].

Future directions for proteomic bioinformatics analyses of datasets similar to this call for more oxidative site-specific modifications to be considered, namely investigation of the carbonyl and HNE produced species as they are stable markers that have been explored in other works [56, 190, 210, 284, 285]. These were included in these analyses by PTM open searching but may require analysis at the *de novo* sequencing level. The computationally heavy analysis via Peaks Studio searching required months of time per dataset and the current advent of faster searching programs such as that of

FRAGPIPE would speed up the re-analysis of these datasets and allow fragmentation spectra-based searching which can help in identifying species containing PB-DOPA by the identification of the immonium ion [286-288]. Further re-interpretation of various tissue datasets would help to build a more conclusive picture of what proteins contain PB-DOPA and Ox-Phe in non-diseased proteomes. As more datasets are re-analysed an understanding of how individual sample preparation methods affect the modifications can be characterised and acted upon, which will enable public datasets that analyse Parkinson's disease samples to be benchmarked in the context of oxidation.

The analysis of PB-DOPA is difficult as the generation of *in-silico* databases for the modification suffers from the possibility that the actual oxidation site is not assignable by being confounded by the other possible sites of oxidation (modification isomers). Further to this, the possibility that the PB-DOPA has undergone quinone formation (the loss of double hydrogen causing the two oxygen atoms to become double bonded to their respective carbon) was not included in these analyses as the search times already spanned several years and its inclusion would have prevented completion of this thesis. Quinone analysis should be considered in the future and added to base search as a variable modification. In Peaks Studio this can be implemented by the iterative searching of a dataset and application of new *de novo* sequencing parameters [258, 259]. The confounding issue with iteratively open-searching datasets is that as more variable modifications are considered at more sites an inflation of false positives does occur. This is partially overcome in Peaks analyses as only the *de novo* peptides that failed to be confidently identified in a previous database search are used in subsequent analyses. This type of iterative searching is not true open searching and the distinction between types of open searches can be found in chapter three. Briefly, all open searching methods restrict the possible search space as computationally considering every possible PTM in every possible site in every combination border on being a mathematically hard problem. Restrictions on number of variables reduce possible search space, such that it becomes feasible to explore PTM occurrence. These include limiting multiple occupation sites considered or by using different metrics such as delta mass searches for identification rather than exploring every possible combination of PTM in parallel (Fragpipe [289, 290] and Andromeda [dependent peptide search] [291, 292]) [228].

In summary L-DOPA was considered a prime candidate for NPAA study as it has a known route of entry into the human body unlike other NPAAs that have disputed routes of exposure alongside dosage level. The confounding factor in the analysis of L-DOPA incorporation is that endogenous levels exist as it is an intermediary synthesised in both the gut and brain [293]. Furthermore, L-DOPA can form from oxidation of the free amino acid tyrosine and the modification of interest which is PB-DOPA can be formed in a similar manner. The data analyses presented within this chapter demonstrate that the identifiable L-DOPA containing proteins are vastly linked to proteins undergoing oxidative damage,

alongside its presence being in the less than 6% range. Data analysis was hampered by poor descriptions in publicly available data, highlighting a need for well annotated datasets and methodologies that consider oxidation during processing. Furthermore, oxidation of the neuronal proteome occurs during aging and is found ubiquitously across neurodegenerative diseases, which further hampered the determination of PB-DOPAs origin and role in the disease process. The subsequent chapter will investigate the incorporation of two NPAAs not known to exist endogenously within the human body and can be readily attributed to dietary exposure which have been suggested in playing a role in neurodegenerative disease.

Chapter 5: Investigation of β-Methylamino-L-alanine (BMAA) and azetidine 2-carboxylic acid (AZE) incorporation and the effects on neuronal cells *in vitro*

5.1 Introduction

Chapters three and four investigated the incorporation of L-DOPA into the human proteome, however analysis of incorporation of L-DOPA into proteins was complicated by the possible endogenous formation of PB-DOPA by oxidation of tyrosine residues. The presence of DOPA at a site encoded for tyrosine could therefore be due to either hydroxyl attack on the tyrosine residue *in situ* or the mistaken incorporation of L-DOPA [57]. In the studies presented in this chapter we investigate the incorporation of two NPAAs whose presence cannot be due to the chemical modification of a canonical amino acid. These two NPAAs have been linked to neurodegeneration and their effects have been hypothesised to be the result of their misincorporation into proteins. The NPAAs of interest are β-Methylamino-L-alanine (BMAA) and azetidine-2-carboxylic acid (AZE), which are linked to motor neurone disease and multiple sclerosis respectively. Both NPAAs have a unique mass shift that is distinct from the masses of the canonical amino acids they are hypothesised to replace and are unlikely to be formed by other biological means within the cell, making them ideal candidates for MiP investigation (see Chapter Two for mass shifts).

Motor Neurone Disease (MND) (also known as amyotrophic lateral sclerosis [ALS]) is characterised by a progressive loss of motor neurones. After diagnosis, patients on average live for four years and throughout this time experience a progressive loss of function [294]. MND cases are 90% sporadic, so for most patients there is no known genetic cause which has hampered the development of biomarkers and therapeutic interventions [295]. Recent modelling studies have suggested that six steps, or insults, could be required for disease precipitation [296]. One known risk factor is age and is considered be one of the steps. Further to this, some known genetic mutations decrease the number of insults required to precipitate the disease down to possibly two insults [296, 297]. This modelling approach has been shown to give consistent results when applied to many patient data sets [296, 297] and the interplay between genetic and environmental factors it is now generally considered linked in MND cases [296].

The hypothesis for BMAA involvement in MND arose from the identification of a cluster of cases of a disease-complex that had features of amyotrophic lateral sclerosis, Parkinson's and dementia (ALS-PDC) on the island of Guam in the 1950s [298]. The incidence of this complex disorder was 50 to 100 times that of the disease worldwide and, since it affected immigrants as well as the local Guamanians, it was likely that environmental factors were involved [59, 186, 299, 300]. One unique feature of the

Guamanian lifestyle was the dietary consumption of cycads, due to extensive cycad forests on the island. Analysis of cycad seeds by Armando Vega in a laboratory in London identified the amino acid BMAA for the first time, and this compound then attracted interest as a possible cause [301]. The source of BMAA was later identified as being cyanobacteria (blue-green algae) growing on the roots of cycad trees. In addition to using cycad seeds to make flour, the seeds were eaten by fruit bats which were then consumed by humans providing a biomagnified source of BMAA. [184]

A study in primates showed that BMAA administered by gavage resulted in ALS like motor dysfunction [63] and the formation of protein aggregate pathology. The aggregate pathology was also partially rescued by co-administration of serine [63]. Furthermore, BMAA has been shown to be associated with the protein fraction of ALS-PDC brains and quantified in several other models, although these analyses utilise amino acid analysis on protein hydrosylates which is the most quantitatively sensitive method for detection of amino acids (as described in chapter two) but at the loss of information about the location at the protein level [185, 187, 298, 302-305]. *In vitro* studies using both neuronal cell lines and primary cells have described a range of toxic effects such as protein aggregation and apoptosis (karayam and speth), reinforced by studies from our laboratory. Our laboratory has also demonstrated cell protection from supplementation of culture medium with the protein amino acid L-Serine [105].

Incorporation of BMAA into polypeptide sequences is hotly debated in the literature [63, 65, 67, 104, 110, 306-310] and proteomic experiments have not been able to definitively prove incorporation [38]. Given the debate in the literature and the lack of evidence for actual protein sequence level incorporation, BMAA exposed samples need further analysis to determine if it is actually incorporating into proteins or is just associated with cellular components that can be isolated by protein precipitation methods [311]. There is further ambiguity around the incorporation site of BMAA, previously reported as in place of serine [103] but more recent studies by the Ibba lab have demonstrated the site to likely be in place of alanine [312]. At a minimum, BMAA can be considered a neurotoxic amino acid and its effects are similar to the pathology exhibited by several neurodegenerative diseases including MND and warrants further investigation.

AZE has been the centre of a multiple sclerosis hypothesis proposed by Rubenstein *et al* [85]. MS is an autoimmune disorder characterised by the demyelination of the central nervous system [313]. AZE is produced ubiquitously across vegetables but in particularly high concentrations in bulbous plants such as sugar beets (*B. vulgaris*) which are used as a source of sucrose and as a gavage for livestock (concentrations quoted at 3.4 mg/kg in molasses or 0.58 mg/kg in shredded beet pulp), namely dairy producing cattle [86]. The incorporation of AZE into proteins has been demonstrated using an *E. coli* overexpression system of myelin basic protein with subsequent radioactive labelling [74, 85]. Studies

by Zagari *et al* bioinformatically modelled the effect of incorporation on predicted protein structures demonstrating alteration, which further supports the incorporation of AZE as being a potential trigger for detrimental myelin structures [88-90, 93]. More recent work by Song *et al.* (2017) [314] utilised a His-tagged myelin basic protein over expression system in HEK-293 cells and used affinity purification with subsequent shotgun-based proteomic analysis to demonstrate incorporation across the majority of proline sites, also reporting that proline supplementation lowers the toxic effects of AZE by competition for incorporation. Despite these experimental approaches, no proteome wide analysis has been published identifying whether specific proteins or sites are susceptible to incorporation or if the proteotype induced is akin to other known neurodegenerative diseases, which is of key importance in the proteinopathies that have proline containing trigger proteins, such as Tau, TDP-43 and α-synuclein. Competition of NPAAs with the canonical amino acid for incorporation is a possible therapeutic strategy for preventing and/or reducing toxicity of NPAAs and this includes serine for BMAA, tyrosine for L-DOPA and proline for AZE [228].

Both the NPAAs of interest for this chapter have data suggesting that they are incorporated into the proteome resulting in toxicity via indirect methods. The BMAA hypothesis has not generated any proteomic data of incorporation thus far but given its potential link to MND, its exploration is important. AZE has been demonstrated to be incorporated into proteins and has had mass spectral data generated [314], however this data is not available publicly and the experimental approach did not allow for assessment of its penetration into the whole proteome of a cell. Furthermore, it was reported that proline can lower the toxic effects of AZE on cell viability and as such will be supplied to neurons alongside AZE to understand the proteomic changes associated with supplementation [314].

This experimental study will utilise the neuroblastoma cell line SH-SY5Y to study the human neuronal proteome, the treatment conditions including a non-treated control alongside 500 µM AZE, 500 µM BMAA and a 500 µM AZE + Proline treatment in order to investigate the potential for proline to reduce misincorporation. Given the difficulty in detecting aggregating proteins in the wider field of proteomics, a harsh solubilisation buffer will be used in an attempt to identify species that may be escaping standard proteomic methods of solubilisation through the use of in gel digestion of biological replicates. The use of in-gel digestion allows the removal of the harsh buffer and surfactant components from the sample and allows digestion of otherwise inaccessible proteoforms [315].

5.2 Methods

5.2.1 Cell culture:

The neuroblastoma cell line SH-SY5Y (passage 26) was serum starved overnight prior to treatment in 175cm^2 culture flasks (80% confluent). Treatment was performed using EMEM containing 1% v/v FBS, and biological triplicates were treated with: unmodified EMEM as a control, 500 μM BMAA which is a treatment reported to result in decreased cell viability [105], 500μM AZE which is a dose that does not result in complete cell death, and 500 μM AZE and 500μM Proline for a period of 24 hours. Cells were washed with PBS warmed to 37°C and subsequently detached using TrypLE express (Gibco) before cells were pelleted at 900 rcf for 5 minutes with the pellet being washed with warm PBS thrice and snap-frozen using liquid nitrogen for storage prior to protein extraction.

5.2.2 Sample processing

When analysing protein aggregates, the selection of sample buffer is important as the inability to solubilise these components in a proteomic workflow prevents their analysis. Earlier work in proteomics focused heavily on the determination of optimal buffer systems that allow broad capture of the proteome at the cost of losing classes of proteins that otherwise would require tailored buffer systems. Given this, a paper by Shevchenko *et al* trailed sixteen extraction buffers showing that a Triton X-114 solution gave the highest protein counts and namely enriched for trans-membrane proteins which are difficult to solubilise due to their hydrophobic nature [316]. This buffer was used in this experiment in order to enhance recovery of insoluble neuronal proteins and protein aggregates that could be under-represented in traditional analysis.

Cells were solubilised using 500 μL 1% v/v Triton X-114, 10 mM Tris-HCl pH 7.4, 100mM PBS, 0.15 M NaCl, 1 mM Ethylenediaminetetraacetic acid (EDTA), and 1x complete ultra-protease inhibitors (Roche). Samples were probe sonicated 3x30 seconds using a Vibra-cell VC50 (Sonics & Materials Inc, USA) ultra-probe sonicator at 30% intensity with resting on wet ice between rounds. Samples were reduced with 5 mM Tris(2-carboxyethyl)phosphine (TCEP) and alkylated with 20 mM acrylamide monomers at room temperature for 90 minutes. Protein content was quantified using gel-based densitometry [317]. Sample were loaded onto hand-cast gels of 12% acrylamide. An eight-point standard curve was constructed for all assays consisting of serial dilution of bovine serum albumin (BSA) (Sigma-Aldrich Cat) resuspended in Laemmli buffer that was boiled and clarified. Gels were electrophoresed at 100 V until samples were clear of the well with a visible gap between the upper limit of sample and the base of the lane well [318].

5.2.3 Gel based protein quantification

Gels were extracted from casting plates, rinsed quickly with ultra-pure water and placed into fixing solution (40% methanol with 10% glacial acetic acid) for a minimum of one hour. Fixing solution was changed to a Coomassie Brilliant Blue stain (Two stock solutions are mixed in a ratio of 4:1, solution 1 [45% methanol, 3 g/L Coomassie Brilliant Blue G-250, 45% ultrapure water with 10% phosphoric acid], solution 2 [50% $(NH4)_2SO_4$] made fresh to increase quality of staining and incubated for 4 hours on gel rocker. Gels were de-stained utilising 1% acetic acid and imaged utilising a Typhoon FLA 9500 (GE Healthcare Life Sciences) in infrared (675/710 nm) [319]. Density analysis was performed utilising QuantityOne® (Bio-Rad USA) with the known BSA dilution series used to construct a standard curve. Equal sample mass (100 µg) was then loaded and separated using hand cast 12% acrylamide gels, fixed and stained with Coomassie blue, imaged as described above, and subsequently subjected to in gel digestion.

5.2.4 In gel digestion

Whole lanes from Coomassie stained gels were excised and diced into 1mm cubes using clean scalpels. Samples were then de-stained of Coomassie using a 50:50 of Acetonitrile (ACN) to 100mM Ammonium bicarbonate (AMBIC) solution. Gel pieces were dehydrated using 100% ACN and rehydrated with 100 mM AMBIC containing 12.5 ng/ul Trypsin Gold (Promega) and incubated for 18 hours at 37°C. Tubes were then bath sonicated for 10 minutes, supernatant removed and 50:50 AMBIC:ACN added to cover the gel pieces, bath sonication repeated, and supernatant combined in a fresh tube. Peptide fractions were then vacuum evaporated to near dryness, and mass spec sample buffer A added (0.1% Formic acid, 2% ACN) to a final concentration of 2 µg/ 5 µl (based on the amount loaded into the gels) and subsequently analysed by LC-MS/MS utilising the same methods found in chapter three in technical triplicate.

5.2.5 Bioinformatics

The raw MS/MS data files were searched using PEAKS Studio X-Pro version 10.6 (Bioinformatics Solutions, Canada). Samples were searched against the human SwissProt database downloaded June 2020 and common contaminants, using three missed cleavages and semi-specific trypsin cleavage, propionamide on cysteine as a fixed modification, variable modifications of; norvaline substitution for isoleucine (for comparison with an external data set), oxidation of methionine, N-terminal acetylation, AZE substitution for proline and BMAA substitution for alanine or serine. Ascore was set to >500 for strong localisation of a modification and further criteria of only peptides matching to the treatment condition considered for further analysis (false positives were identified by presence of a matched peptide in BMAA or control samples), spectra produced in only single sample were manually inspected for spectral quality, chimeric spectra were only considered if well annotated, and sequences

supported by overlapping confident peptides were also retained. All biological information relating to protein annotations was taken from UniProt by the use of protein accessions unless otherwise specified. Label free quantitative analysis was performed using the PEAKS Studio feature based LFQ method with a 6-minute retention time shift tolerance and a 1% FDR threshold. GraphPad Prism 9 software was used for the generation of boxplots and statistical analysis. STRING DB was used to generate protein interaction networks and to perform biological enrichment analyses as outlined in section 4.2.3, briefly, physical interaction networks were generated with cluster analysis performed, permanent links are retained for future access to the interaction networks.

5.3 Results

5.3.1 Overview of dataset for all conditions

From the PEAKS Studio PTM analysis there was a total of 3,941 proteins (2,985 groups) identified from this experiment with 27,893 sequences, 32,700 peptides and 654,778 PSMs (Appendix 5_1). The filtered results resulted in a 0.2% FDR on the PSM level, 1% for peptide sequences and 0.5% for protein groups. From the label free quantitative output there was less than 1.2% missing values on all samples once matching between runs was applied (between 7-15% prior to matching). Based on the metrics reported inside of PEAKS, peptide quality was set ≥ 10, average area ≥ 1E4, charge state between two and six, Peptide ID count of at least one and detected in at least one per group, a 1% FDR was applied and a fold change ≥2 with at least 2 peptides using the PEAKSQ algorithm, which is a repurposing of the MaxLFQ method [258, 259, 291, 292]. This resulted in 375 protein groups found to be differentially abundant across the entire dataset with over 27,471 feature vectors with associated IDs supporting these proteins.

5.3.2 BMAA treatment results

Inspection of the data at the peptide level showed no spectra definitively indicating the presence of BMAA incorporation and no discernible bias towards those samples treated with BMAA containing matches. However, examining the PSMs prior to filtration showed that there were 4,626 alanine and 1,913 serine BMAA substitutions which were all deemed false through extensive manual validation and interpretation of fragmentation spectra. Further criteria that were used to filter the PSMs containing BMAA were that the spectra were identified in control samples or AZE samples, and that no PTM combinations were also found to better explain the BMAA peptides. The framework for assignment of true NPAA incorporation can be found in chapter two. All identification results can be found in appendix 5_2-A. Quantitative analysis using Peaks Studio resulted in fourteen proteins identified as differentially abundant with seven proteins increasing and seven decreasing (Table 4), with two keratins being present in these filtered results and ignored during further analysis. It is important to note here that the significantly changing proteins identified for BMAA treatment were also evident in treatment with AZE.

Table 4: Proteins significantly changing upon BMAA treatment.

Accession	BMAA fold change	Aze fold change	Aze & Pro fold change	Description
Q99873	2.47	1.82	2.21	Protein arginine N-methyltransferase 1
P68871	2.4	2.98	1.27	Hemoglobin subunit beta *
P02042	2.4	2.98	1.27	Hemoglobin subunit delta *
P62879	2.2	2.41	2.29	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2
P35908	**2.19**	**1.57**	**3.65**	**Keratin type II cytoskeletal 2 epidermal**
P60953	2.08	2.56	1.75	Cell division control protein 42 homolog
P01023	2.06	2.61	2.13	Alpha-2-macroglobulin
P62308	0.5	0.48	0.71	Small nuclear ribonucleoprotein G **
A8MWD9	0.5	0.48	0.71	Putative small nuclear ribonucleoprotein G-like protein 15 **
O43583	0.49	0.5	0.44	Density-regulated protein
O43290	0.49	0.79	0.53	U4/U6.U5 tri-snRNP-associated protein 1
Q14103	0.44	0.53	0.4	Heterogeneous nuclear ribonucleoprotein D0
P62937	0.44	0.49	0.5	Peptidyl-prolyl cis-trans isomerase A
P35527	**0.39**	**0.35**	**0.28**	**Keratin type I cytoskeletal 9**

*Raw fold-change presented, two contamination proteins of Keratin have been bolded and excluded from further analysis. Red proteins are increased in abundance within BMAA vs control and green is down. * Undistinguished protein isoforms.*

Of the proteins found to be increased in abundance, Protein arginine N-methyltransferase (PRMT1) is an RNA binder and histone methylator. The indistinguishable haemoglobin subunits (beta and delta) identified as increased in abundance are involved in hydrogen peroxide catabolic processes and the positive regulation of cell death. Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2 (GNB2) is involved in calcium channel regulation, G protein-coupled receptor signalling and protein folding. Cell division control protein 42 homolog (CDC42) is involved in a plethora of biological processes and molecular functions including the positive regulation of neuron apoptotic processes and the coordination of physical cellular components. Alpha-2-macroglobulin is responsible for inhibition of proteinases by trapping and reducing the activity against higher molecular weight proteins.

The proteins decreasing in abundance included several proteins involved in the translation machinery of the cell. Small nuclear ribonucleoprotein G (SNRPG) is involved in RNA binding, histone mRNA metabolic processes and spliceosome complex assembly, and this protein was indistinguishable from a protein Putative small nuclear ribonucleoprotein G-like protein 15 (SNRPGP15) which has similarly

annotated function. Density-regulated protein (DENR) is also a mRNA binding protein and is involved in translation initiation and ribosome disassembly, and this protein was indistinguishable from U4/U6.U5 tri-snRNP-associated protein 1 (SART1) which also has the role of RNA binding and splicesomal assembly. Heterogeneous nuclear ribonucleoprotein D0 (HNRNPD) is a chromatin and RNA binding protein, involved in RNA catabolism and processing as well as positive regulation of translation, telomere maintenance and RNA based spliceosome splicing, this protein was indistinguishable from Peptidyl-prolyl cis-trans isomerase A (PPIA). PPIA is also an RNA binding protein which is a negative regulator of oxidative stress induced apoptosis.

5.3.3 Azetidine treatment results

5.3.4.1 Incorporation of Azetidine into proteins

Analysis using Peaks Studio resulted in 10,068 PSMs containing AZE with 1,045 peptides remaining with an Ascore above 500 and applying criteria of only being identified in AZE treated conditions (AZE and AZE + Proline) resulted in the identification of 662 AZE containing peptides with 648 having accessions assigned. Manual inspection of peptide spectra resulted in a final definitive list of 467 peptides containing AZE matching to 270 unique proteins (Appendix 5_2). Upon manual inspection, the AZE replacement in a peptide sequence results in the masking of the next y-ion in the series and is similar to the effect when proline itself fragments preferentially. Furthermore, the immonium ion for AZE was routinely seen (56.05 Da) which provided an extra piece of supporting evidence in many spectra (diagnostic ion). Immonium ions are unique ion fragments that are from the side chain of interest and broken free from the peptide backbone and as such are used as a complementary ion in peptide sequencing. When performing manual sequencing of fragmentation spectra, the detection of these ions further validates correct sequence assignment.

The number of peptides manually validated per treatment group are 463 peptides for AZE only and 33 in the AZE + proline condition, the majority of proteins identified in the proline supplemented group were also identified in the AZE only group (28) with only 5 being unique (Figure 18). Intensity based analysis of the manually validated AZE containing peptides demonstrated a 6.8-fold difference across all AZE peptides in each sample. Whilst a comparison of the intensities of AZE peptides identified in both conditions demonstrated a 12.7-fold reduction in intensity. Within this experiment proline supplementation was not optimised to produce a reduction in AZE but regardless has caused a reduction in number and amount of AZE incorporation.
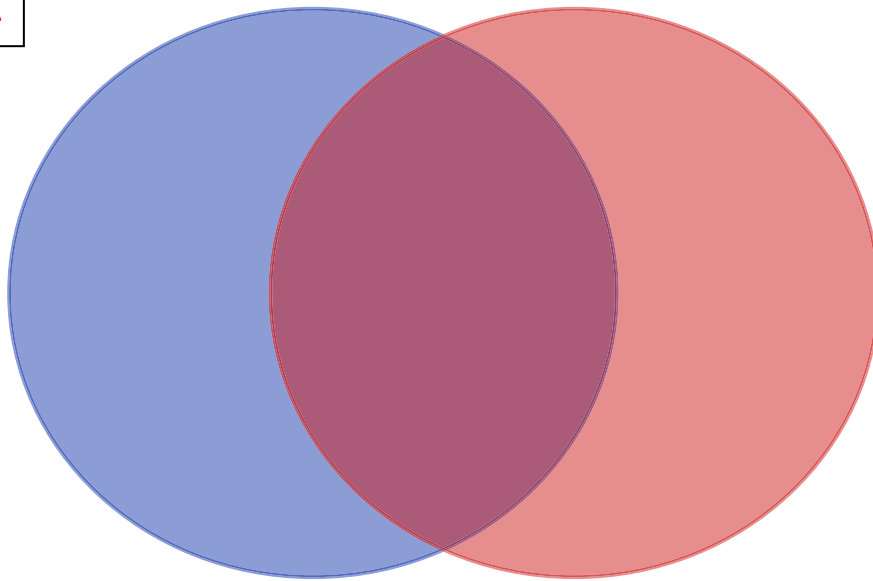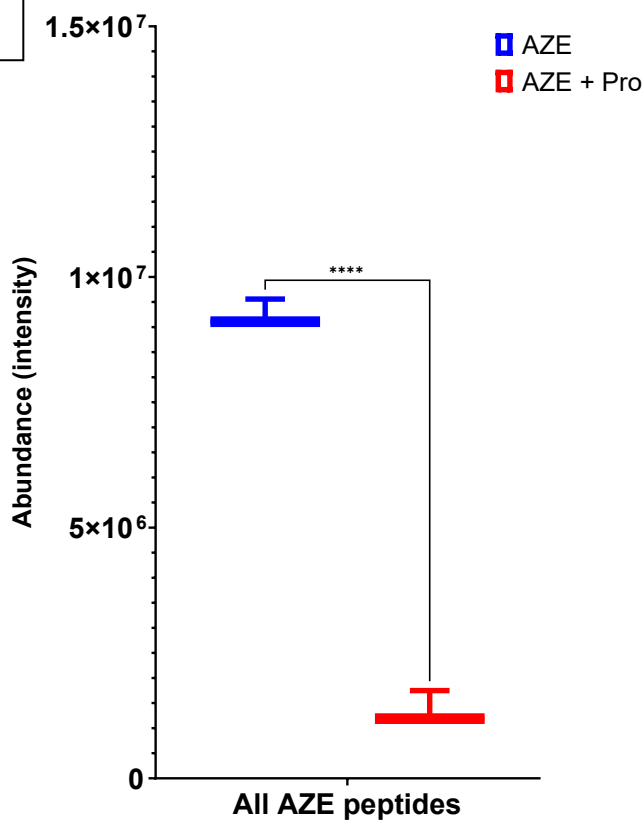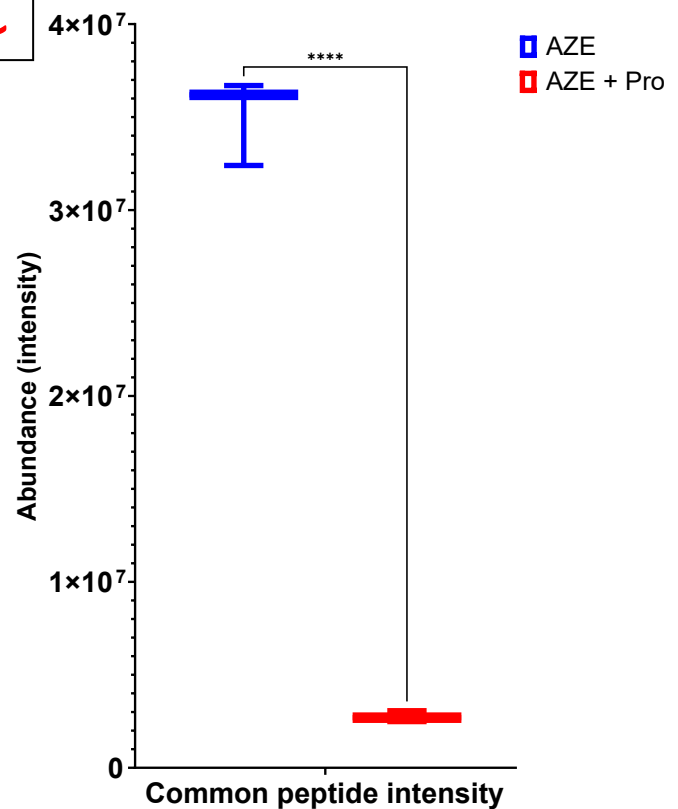
Figure 18 Analysis of AZE containing peptides across both conditions. A) Venn diagram displaying the overlap and distinct number of AZE containing peptides with 14-fold less peptides identified containing AZE upon proline supplementation. B) Boxplot of all manually validated AZE containing peptide intensities averaged per condition, displaying a 6.8-fold difference in means with statistical significance (p <0.0001 N=393). C) Boxplot of AZE containing peptides identified in both conditions, there is a 12.7-fold reduction in peptide intensities upon proline supplementation (p <0.0001 N= 28).

5.3.4.2 Network and enrichment analysis of proteins containing misincorporated azetidine-2-carboxylic acid

The AZE containing proteins mapped to 252 STRING proteins. A physical subnetwork was generated using only experimental data and databases as sources with kmeans clustering used set to six clusters (Figure 19) (Appendix 5_2-C). (permalink: https://version-11-5.string-db.org/cgi/network?networkId=biEQHyGEkTOu). The network generated by physical interaction, removed sources from text mining and used only empirical experimental derived protein-protein interactions, however the size of this network was large in spite of these more stringent parameters. From this network, kmeans clustering was used to divide the network into six clusters and these individually were investigated for enrichment. The enrichment results on the entire network resulted in major terms in all the functional enrichment modules that STRING analyses. Biological processes found to be statistically enriched in STRING DB include protein folding (44/213, FDR: 1.19e-32), organelle organization (117/3450, FDR: 1.56e-21), and response to unfolded protein (25/166, FDR:1.47e-14). The enriched cellular component terms included; extracellular exosome (175/2099, FDR:3.37e-102) and membrane-bounded organelle (246/12427, FDR: 2.08e-37). The enriched KEGG pathway terms included; systemic lupus erythematosus (29/93, FDR: 3.20e-26), prion disease (32/265, FDR: 2.18e-18) and Parkinson's disease (29/240, FDR: 1.13e-16).
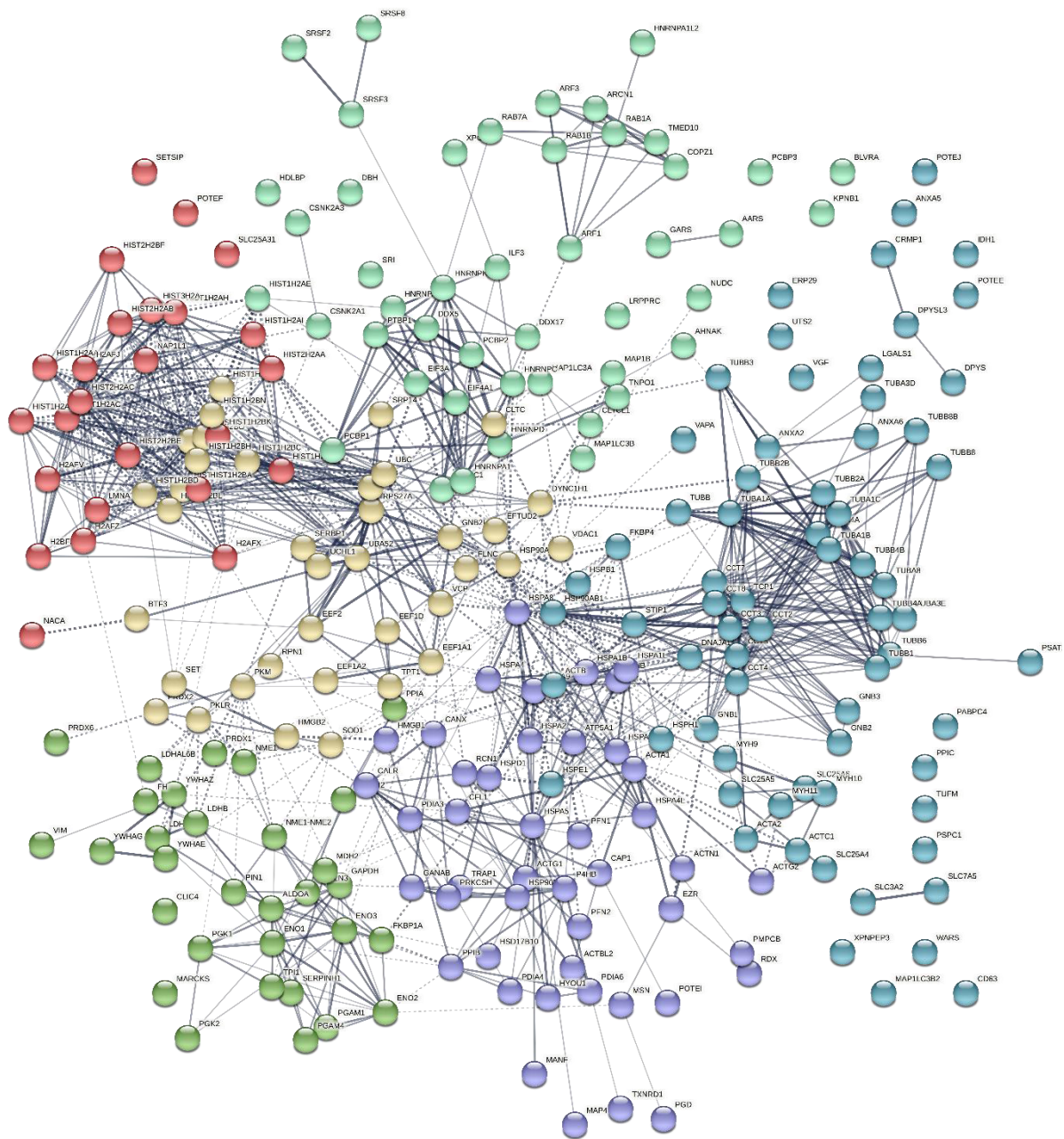
Figure 19: STRING physical network of proteins containing AZE. Six clusters generated by Kmeans clustering. Red: nucleosome and histones. Yellow: Extracellular exosome nucleosome and involved in protein ubiquitination and amyloid fibre formation, notably eight of which have been identified in the mediation of protein aggregation and cytotoxicity in AD. Lime-Green: Extracellular exosome and space, identical protein binding, Nucleoside diphosphate phosphorylation. Green: mRNA splicing by spliceosome, mRNA binding, annotated as extracellular exosome and vesicle components. Blue: Protein folding, protein localisation and microtubule cytoskeleton organisation, annotated to be extracellular exosome, supramolecular fibre and microtubules. Purple: Protein folding, response to unfolded protein, RNA binding, chaperone mediated protein folding, annotated to be extra cellular exosome, extracellular space and focal adhesion.

Enrichment analysis of disease-gene associations identified proteins associated with neurodegenerative diseases characterised by neuronal death and matching proteinopathies (Table 5). Given the diverse array of proteins identified including those identified including the unfolded protein response (Figure 19; purple cluster), these results are likely to be biologically relevant rather than statistical inflation related to cell lines proteome.

Table 5: Statistically enriched disease-gene associations from proteins containing AZE. Extended table matches can be found in Appendix 5_1 supplementary file 1 and sheet "AZE_enrichment.DISEASES"

Term ID	Description	Observed	Background	FDR
DOID:1289	Neurodegenerative disease	21	462	0.0047
DOID:0050539	Charcot-Marie-Tooth disease type 2	6	34	0.0232
DOID:231	Motor neuron disease	8	78	0.0232
DOID:070355	Multisystem proteinopathy	3	3	0.0431
DOID:331	Central nervous system disease	31	1107	0.0431
DOID:870	Neuropathy	14	297	0.0431

5.3.4.3 PEAKS quantitative analysis of protein abundance following azetidine-2-carboxylic acid treatment

Quantitative analysis of protein abundance changes upon AZE treatment was performed using the Peaks Studio software. Proteins having a two-fold change in abundance were retained during this analysis (Appendix 5_2-B) resulting in the identification of 239 proteins changing in abundance, with 19 found to be increased and 220 found to be decreased upon AZE treatment (1% adjusted FDR, two unique peptides).

The 19 proteins of increased abundance were analysed using STRING DB for physical interaction resulting in 20 matching proteins due to protein isoform matches, and kmeans clustering was used with four clusters set as visually four distinct regions were present (Figure 20 Permalink (https://version-11-5.string-db.org/cgi/network?networkId=bvMOTbAUctzn). The biological processes terms enriched included; hydrogen peroxide catabolic processes, protein folding and stress response proteins. The enriched cellular component terms from this network included; extracellular space, extracellular exosome, vesicle and oligosaccharyl transferase complex (OST). Several of these proteins play a role in collagen binding and other structural protein biosynthesis such as actin. Taken together these proteins indicate a response to oxidative stress by the increase in the HBB cluster and protein misfolding, the OST being intertwined with the response to protein misfolding as it assists in the stabilisation of misfolded proteins.
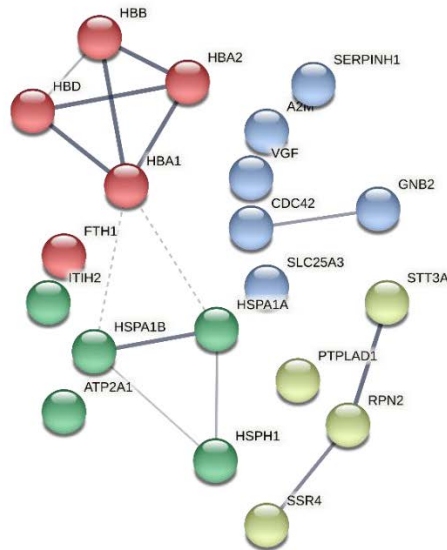
Figure 20: Physical protein network of proteins greater than two-fold increased upon AZE treatment. Red: Hydrogen peroxide catabolic processes and receptor mediated endocytosis. Yellow: Oligosaccharyltransferase complex and endoplasmic reticulum protein processing. Green: Heat shock proteins responsible for misfolded protein refolding, annotated to be components of the aggresome and endoplasmic reticulum. Blue: Proteins annotated to be located in the extracellular space, with functions related to collagen binding, mitochondrial phosphate import, cell-cell communication and proteinase inhibition.

The 220 proteins of decreased abundance were analysed by STRING DB generating a physical network of 211 identities with text-mined data excluded from the analysisand the physical network clustered to seven modules using kmeans clustering (Figure 21 Permalink (https://version-11-5.string-db.org/cgi/network?networkId=bJQt1OHlJVt7). Analysis of the overall enrichment identifies proteins involved in mRNA metabolic processes as decreased and translation initiation with a large cluster of ribosomal proteins identified by the purple cluster (36 proteins annotated as ribosomal). Notably the protein pyrroline-5-carboxylate reductase 1 (PYCR1), which is the house keeping enzyme responsible for the last step of proline biosynthesis, was found to be decreased in abundance. Using this as a highlighting enzyme when interpreting the network, another eight other proteins involved in amino acid biosynthesis were identified with several being mitochondrial (ASNS, ACO2, IDH3A, SHMT2, PHGDH, PSAT1, CBSL, PYCR1, BCAT1). Several proteins annotated as being involved in amyotrophic lateral sclerosis were found within this network (TUBB6, TUBB3, SRSF7, HNRNPA1, UBQLN2, HNRNPA2B1, HNRNPA1L2, MATR3, UBQLN4, SRSF3, HNRNPA3, KLC2, ALYREF, ADRM1) as well as proteins that are responsible for the negative regulation of apoptosis and immune signalling.
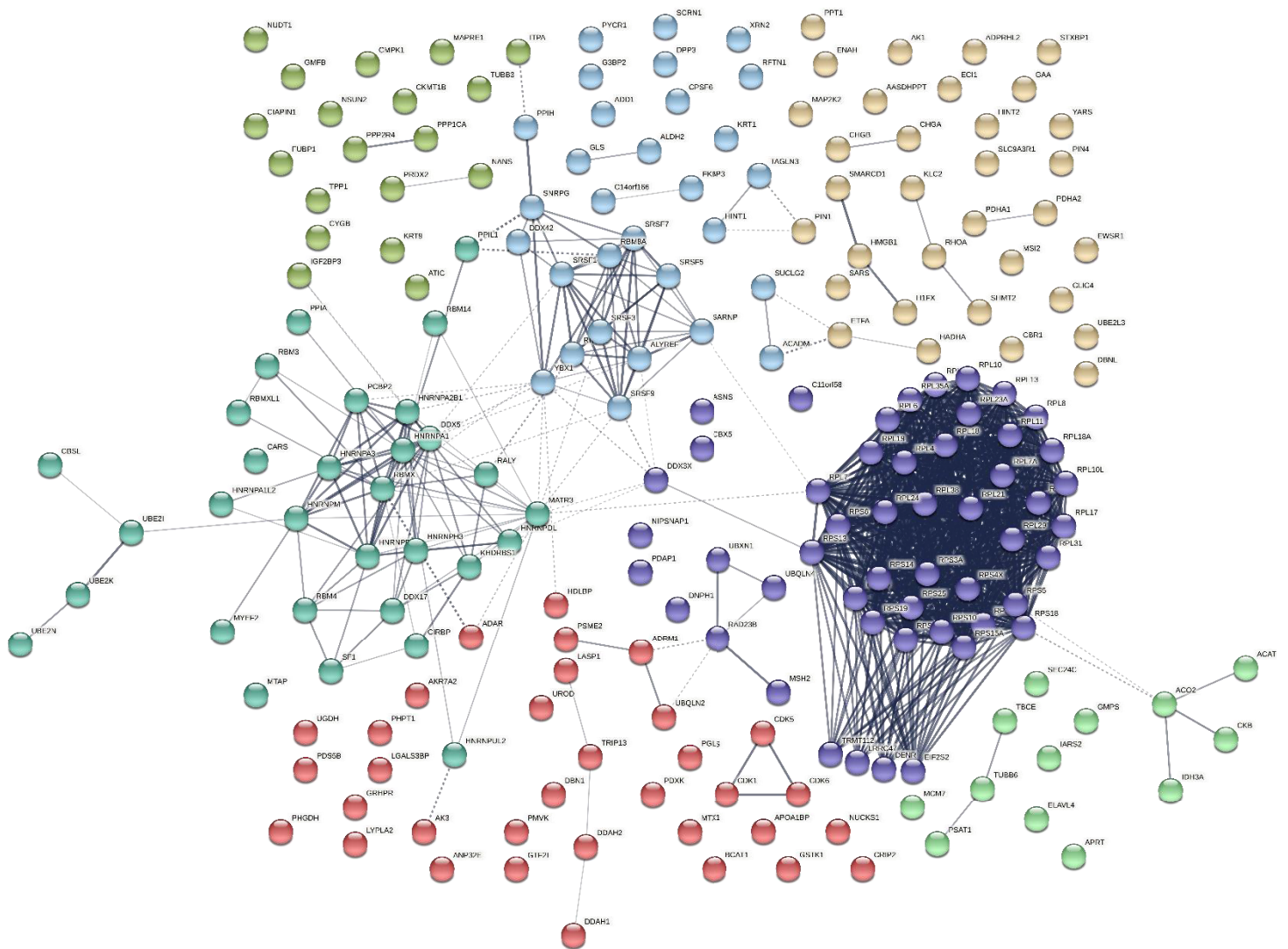
Figure 21: STRING generated physical network on the 220 proteins of decreased abundance due to AZE treatment. Red: Small molecule metabolic processes (nitrogen pathway matched), extracellular exosome, vesicle and cytosol. Brown: Carboxylic acid metabolic process, small molecule metabolic processes with proteins mapped to mitochondrial components and extracellular exosomes. Olive: Proteins involved in the suppression of cell death and cell growth, these proteins were statistically enriched to be exosomes and exist in extracellular space. Green: Proteins involved in mitochondrial small molecule processing, proteins involved in tubulin folding, neuron specific RNA processing and proteins involved in protein transport. Blue: Proteins involved in mRNA/RNA splicing and processing, myelin expression factor 2 the transcriptional repressor is also present in this cluster. Light-sky-blue: Proteins involved in mRNA/RNA processing and transport as part of the spliceosomal complex with proteins involved in protein localisation and intracellular transport. (Purple): Ribosomal proteins as the core cluster with proteins present that negatively regulate cell apoptosis and are responsible for ubiquitin based proteasomal degradation.

Quantitative analysis was performed on the cellular proteins following co-treatment with proline and AZE vs AZE alone to assess the potential of proline to negate the toxic effects and incorporation of AZE into proteins. Analysis identified 54 proteins found to be differentially abundant, with six proteins increasing and 48 proteins decreasing in abundance (Appendix 5_2-B). Supplementation with proline reduced the number of differentially regulated proteins to 22% of that with AZE alone. With AZE-treatment alone 8% of proteins differentially regulated increased in abundance and in the presence of proline this was similar at 7%. Four non-contaminate proteins were found be increased in abundance (Table 6). Asparagine synthetase (ASNS) is responsible for the synthesis of asparagine, and involved in the negative regulation of apoptosis and positive regulation of mitotic cell cycle. Chromogranin-A (CHGA) is a protein that has several roles in biological processes including; being a free radical scavenging molecule, innate immune response mast cell degranulation and the negative regulation of neuron death [320]. General transcription factor II-I (GTF2I) is involved in the negative regulation of angiogenesis and the positive regulation of DNA repair. Cyclin-dependent kinase 6 is involved in cell cycle control and cell differentiation.

Table 6: Proteins found to be increased in abundance upon the cotreatment of proline and azetidine versus azetidine alone.

Accession	Control	BMAA	Aze & Pro	Description
P08243	3.53	4.62	3.91	Asparagine synthetase [glutamine-hydrolyzing]
P10645	5.57	4.03	2.25	Chromogranin-A
P78347	4.7	3.91	2.11	General transcription factor II-I
Q00534	3.5	3.01	2.01	Cyclin-dependent kinase 6

The 48 proteins of decreased abundance were analysed using STRING with the generation of a physical network with 43 matches. The STRING network was then clustered using kmeans clustering with the number of clusters set to four as visually four sets could be seen (Figure 22) (Permalink: https://version-11-5.string-db.org/cgi/network?networkId=bJmp9sPov1wV). The network generated identified several biological processes enriched for including; Protein folding in the endoplasmic reticulum, positive regulation of tau-protein kinase activity, response to unfolded protein and cell redox homeostasis. Several KEGG pathways were enriched including protein processing in the endoplasmic reticulum and antigen processing and presentation.
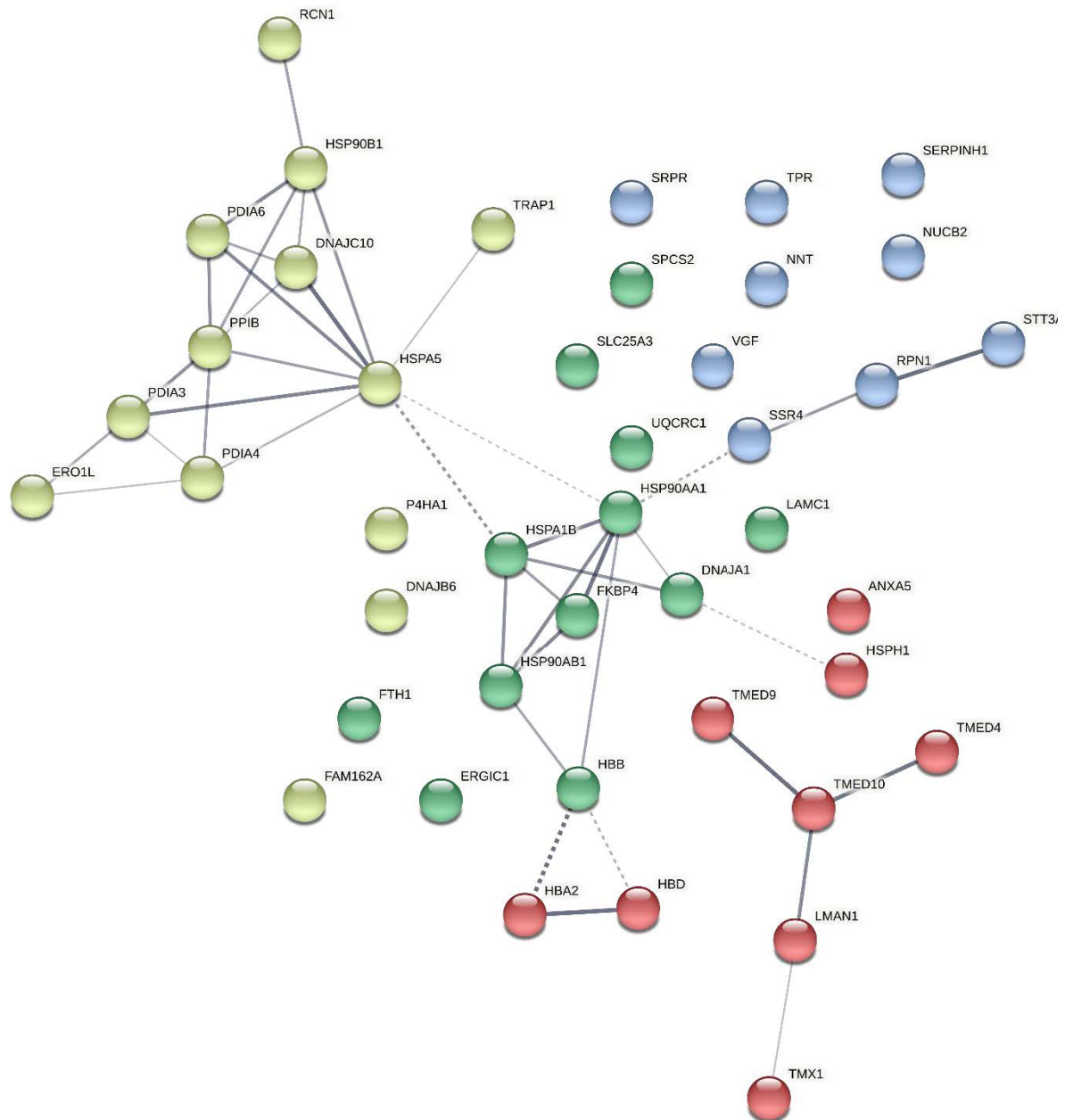
Figure 22: STRING physical network of proteins decreased in abundance due to the addition of proline versus azetidine alone. Red: Golgi organisation, vesicle mediated transport and oxygen transport. Yellow: Protein folding in the endoplasmic reticulum, ATF6-mediated unfolded protein response, cell redox homeostasis, proteins involved in the prion disease pathway (HSPA5, HSP90B1, PDIA3) with mapping as chaperones and endoplasmic specific proteins. Green: Protein folding/refolding, rneutrophil degranulation, mitochondrian, necropoptosis, antigen processing and presentation, prion disease proteins (HSPA1B, UQCRC1, LAMC1). Blue: Proteins located in the endoplasmic reticulum.

5.4 Discussion

5.4.1 BMAA *in vitro* treatment does not result in detectable misincorporation

BMAA has been investigated previously for evidence of protein misincorporation using shotgun proteomics [38] and it was reported that no incorporation was detectable when analysed alongside orthogonal methods of amino acid hydrolysis with targeted SRM analysis and using *in vitro* synthesis kits. Similarly, through manual interrogation of peptide spectrum data of BMAA treated SH-SY5Y cells in the present studies, there was no BMAA misincorporation detected with no indicative bias even in location unassignable modifications. Beri *et al* examined only BMAA incorporation in place of serine, whilst recently the possibility of alanine replacement has been proposed [312]. The analysis presented in this chapter included consideration of this secondary site for incorporation and failed to find evidence of incorporation at either site. Regardless of incorporation, the exposure to BMAA was shown by Beri *et al* to result in changes in protein abundance consistent with neurodegenerative damage [38].

In order to validate the presence of NPAA incorporation in a manner that overcomes detection sensitivity, utilisation of a human recombinant protein expression system alongside affinity purification may prove viable as it increases the ability to detect low abundance proteins. Even though MS-READ failed to identify the BMAA containing species, it may be a function of the non-mammalian expression system utilised [312] (previously detailed in Chapter Two for further information). The generation of *in silico* targeting schemes for the mass spectrometer based on previously acquired samples may prove useful [228] alongside the acquisition of either amino acid analysis of the proteome or the use of radiolabelled BMAA to establish incorporation rates into the proteome alongside the BMAA proteomic experiment.

The method applied in the present studies aimed to increase the solubilisation, and thus detection, of peptides from potentially insoluble protein species produced via incorporation. Since we did not detect BMAA-containing peptides, it is hard to evaluate the success of this approach. Excretion of the modified protein species by vesicle-mediated cargo transport to the extracellular environment may have occurred and this would not have been detected by our experimental approach. Given that both experiments utilised non-differentiated cells, it is also possible that any BMAA-containing proteins could be spread to daughter cells preventing accumulation at levels toxic to the cell and more easily detectable by MS. These cells are also highly metabolically active being an immortalised cell line with a higher protein turnover rate than a primary neuronal cell, potentially neutralising any protein accumulation [321]. Furthermore, given the interval length from BMAA exposure to ALS-PDC reported in humans [61], the short treatment time used in this experiment might not reflect the type of insult that would occur *in vivo*, and experiments on differentiated mixed cell populations like organoids over

very long periods of time at low dosages may provide a more accurate model to assess BMAA incorporation. The data presented in this chapter does not reject the hypothesis for BMAA incorporation outright but does provide more evidence for it not being at a level detectable with current, state of the art, proteomic techniques without an enrichment method. Use of non-conventional methods of protein solubilisation, such as with ionic liquids which are themselves able to enhance the solubility of hydrophobic molecules through their ability to interact in both cationic anionic manners, could be key for recovering lost modified proteins [322].

5.4.2 BMAA treatment results cellular stress responses and perturbs the spliceosome

Label free quantitative analysis of BMAA treated cells versus control cells identified fourteen differentially abundant proteins (Table 4). The proteins of increased abundance were involved in cell signalling and response to stress. The latter is consistent with previous cell studies [105] as well as a primate study which reported protein aggregates in the primate brain [63]. Protein arginine N-methyltransferase (PRMT1) is an RNA binding protein, histone methylator and is responsible for the methylation of RBM15 leading to its degradation and preventing terminal differentiation of megakaryocytes [323]. There was no evidence of methylation on RBM15 (via manual interpretation of the PTM search results) found in this dataset but given the increase of PRMT1, it is likely to play a role in RNA binding and protein degradation. PRMT1 has been shown to be a highly expressed protein involved in the proliferation of breast cancers, which does not agree with other studies where by BMAA treatment results in slowing of cell cycle progression [324, 325]. Furthermore, the reported inhibition of cell cycle progression with BMAA was within cells without glutamate receptors, suggesting that the mechanism was independent of excitoxicity [324].

Hemoglobin subunits (either beta or delta) were identified as having an increased abundance and were indistinguishable due to no proteotypic peptides being detected to differentiate the proteoforms, with both variants involved in oxygen transport. Both hemoglobulin subunits play a role in hydrogen peroxide catabolism and cellular oxidant detoxification, which may be a response to possible oxidative stress [326]. Hemoglobulin subunit beta also has a role in the positive regulation of cell death [327] and positive regulation of nitric oxide biosynthesis [328]. The potential role in cells responding to BMAA treatment may be that oxidative stress is increasing and cell death is occurring in the cell population which may be part of the mechanism underpinning the reported neuronal toxicity [38, 62, 63, 110, 111, 307, 329]. Another protein significantly increased was Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2 (GNB2) which functions as a calcium channel regulator and acts in the G protein-coupled receptor signalling pathway. A known mechanism of BMAA toxicity is excitotoxicity through its interaction with bicarbonate resulting in the formation of carbamate adducts which are capable of triggering glutamate receptors [306], leading to dysregulated

calcium homeostasis which can result in apoptosis. SH-SY5Y cells are routinely used for modelling oxidative stress and upon differentiation can serve as a model of Parkinson's disease since they express a dopaminergic phenotype. It is unclear whether they can respond to glutamate since the presence of ionotropic glutamate receptors (iGluR) or metabotropic glutamate receptors (mGluR) has not been shown consistently [330]. Regardless of the absence of the NDMA receptors in SH-SY5Y cells, glutamate induced oxidative stress has been reported independent of receptors due to increased intracellular Ca^{2+} [330]. The manner in which this could be occurring with BMAA treatment is through the formation of a carbamate adduct which triggers the glutamate and cystine antiporter, resulting in the depletion of glutathione (GSH) and down regulation of SOD resulting in apoptosis which is also a characteristic change seen in MND [331].

The protein CDC42 was increased upon BMAA treatment, and this protein is involved in a wide range of functions including the extension and maintenance of filopodia, positive regulation of stress fibre formation and positive regulation of neuronal apoptosis. The CDC42 increase does not point to one change in cell function but may be a marker for the diverse changes that are occurring due to BMAA treatment, and the various underlying processes affected. Functional assays alongside microscopy techniques could prove useful for untangling this result as well as the use of more advanced proteomic techniques to characterise other interaction partners.

The final protein found to be increased upon BMAA treatment was alpha-2-macroglobulin (α-2M), a large protein which is responsible for inhibition of all four classes of proteinases that target high molecular weight proteins. This protein is known to target metalloproteinases, playing a role in age related disease and cancer progression by mechanisms involving zinc binding, as evidenced in cervical cancer [332]. The literature implication of a α-2M role within cancer and age-related diseases does not provide a clear answer, but its implication in age related disease progression is salient as one known factor for MND is age, warranting further exploration to characterise the phenotype parallels to MND pathogenesis.

The proteins of decreased abundance are primarily involved in translation machinery of the cell suggesting a decrease in protein synthesis in response to BMAA exposure. There was a repeated association with proteins of the splicesomal complex including SNRPG, SART1, and HNRNPD. The splicosome is a complex that is responsible for the splicing of pre-mRNA and its dysfunction has been linked to MND [333, 334]. Decrease in the protein abundance of the components of the spliceosome itself has been found to be caused by neurofibril tangle formation resulting from tau protein aggregation [335], which has also been reported as a result of BMAA treatment in a primate model [63]. It is possible that disruption of the spliceosome may result in frontotemporal dementia-MND

[336], further supporting the hypothesis that BMAA can act as a factor in the pathogenesis of MND [38].

Certain proteins that changed significantly in abundance in the BMAA treated samples were also found to be similarly changing in abundance upon AZE treatment and slightly less so in the AZE-proline treated samples. This may result from a general toxicity induced by these molecules to the cell but could also provide further evidence that proteinogenic NPAAs may induce a similar cellular response, which is characterised by changes homologous to neurodegenerative states. BMAA treatment results in a cell state that has underlying features of neurodegenerative diseases, namely the splicesomal nature of MND [333]. BMAA was found to not be incorporated into proteins which has been reported by one other proteomic experiment to date [38].

5.4.3 Azetidine 2-carboxylic acid incorporates readily into the neuronal proteome.

Analysis of AZE treated cell proteins for AZE misincorporation resulted in over 10,000 AZE peptide spectral matches which with manual inspection and filtration resulted in 467 peptides with high confidence spectra demonstrating incorporation. The position of proline within a peptide fragmentation spectrum has been known to result in the diminished signal of the N-terminal fragment ions or the next y-ion of increasing m/z in a peptide fragmentation series [337], and it was similarly found to occur with AZE incorporation in place of proline. This thesis is the first to report this effect. The immonium ion of 56.05 Da was found in many spectra also showing AZE incorporation, and the immonium ion was used as further evidence that a sequence assignment was correct. This diagnostic immonium ion also makes it possible to perform precursor ion scanning proteomic workflows and assists with sequence assignment in DIA experiments [338, 339]. Upon filtration for unique protein accessions, 270 proteins containing AZE remained, with a large number of proteins containing multiple peptides and sites of incorporation (Appendix 5_2-C).

Misincorporation of AZE has been previously shown by Song et al [314] in a HEK293 myelin basic protein expression system with a wide population of peptides species detected. This led the authors to suggest that antigenic peptide species could have been generated. Furthermore, the exposure of lambs to AZE-containing sugar beet silage resulted in an autoimmune disease phenotype similar to Multiple Sclerosis [340]. The data presented herein provides additional evidence for the incorporation of AZE into proteins and highlights the need to perform animal-based experiments analysing not only the proteome of the central nervous system and various fluids, but also to carry out immunological assays to detect possible antigenic peptides [341].

The proteins containing AZE were used to generate STRING networks which were analysed for functional enrichment in order to understand which proteins could be modified and subject to misfolding, in addition to those proteins already described in the literature such as the collagens, haemoglobin and keratins [239]. Analysis of the AZE protein network as a whole resulted in 252 matched identifiers with the majority of the proteins being annotated as components of extracellular exosomes (175) and membrane-bound organelles (246). The incorporation of AZE into proteins suggests that proteins containing AZE are readily exported, which could prove a mechanism of dissemination across the central nervous system and would suggest they could potentially be used as biomarkers for AZE exposure [342]. It has been reported that exosomes containing myelin proteins could possibly perpetuate and maintain immunogenicity in T-cells which highlights a possibility that the generation of exosomes containing proteins with incorporated AZE may be a factor in triggering or enhancing multiple sclerosis [343].

The biological process terms enriched across the entire network include protein folding, organelle organisation and response to unfolded proteins. These enriched terms suggest that the proteostasis systems are infiltrated by AZE and that this could be hampering efficient protein refolding, resulting in a toxic gain in function by loss of tertiary structure stability with subsequent protein aggregation. It is also important to note here that this may be a runaway effect (or positive feedback loop) of proteins responsible for dealing with misfolding being misfolded and subsequently have loss of function or toxic gain of function, which may also be a function of the increase in these specific proteins due to wide spread misincorporation. Within the cluster-based analysis of the STRING network, there was also splicesomal proteins found to contain incorporated AZE and they are annotated as being involved in amyloid fibre formation which is indicated in several neurodegenerative diseases [335, 344]. In addition, there were several KEGG disease terms statistically enriched including prion disease, systemic lupus erythematosus and Parkinson's disease. In support of this disease association, enrichment also identified neurodegenerative disease, motor neurone disease, multisystem proteinopathy, central nervous system disease and neuropathy. The identification of these ontologies in our *in vitro* studies does not show that incorporation of AZE causes MS but it does demonstrate a potential role for AZE in causing neurological proteinopathy and potentially immune effects which supports Rubenstein's AZE-MS hypothesis [86]. It is important to emphasise that proteins rich in proline may be more heavily impacted by the incorporation of AZE resulting in the formation of aberrant structures or conversion of disordered regions to prion like structures. Proteins linked to neurodegeneration that are known to be proteinopathic are prime candidates for potential gain of

toxic function by misincorporation. For example TDP-43 is implicated in MND and has 16 proline residues, and within the region of known pathogenic mutation (residue 169-393 for natural disease-causing variants) there are nine prolines spread across known disordered regions and within the ubiquilin interaction region of the protein which could also prevent efficient degradation [345, 346].

5.4.3.2 Azetidine 2-carboxylic acid results in protein misfolding and a decrease in protein synthesis machinery

Quantitative analysis was performed on the AZE treated cells versus control cells resulting in 239 proteins found to be of differential abundance with a bias to more proteins decreasing in abundance (220). This suggests that the proteome is geared towards the degradation of proteins upon treatment and is likely a reflection of the increased incorporation which would be leading to misfolded proteins. Proteins with increased abundance were analysed using STRING with the resultant biological process terms enriched including hydrogen peroxide catabolic processes, protein folding and stress response proteins. Since we have demonstrated that some proteins contained AZE, these data would suggest that proteostatic mechanisms in the cell have recognised the aberrant proteins. In addition, it is possible that the insertion of AZE alters the tertiary structure of certain proteins exposing previously buried regions of the protein's that are hydrophobic, increasing their propensity for aggregation [201, 347, 348]. The increase in proteins responsible for hydrogen peroxide catabolism indicates an oxidative stress state which has been shown to lead to protein misfolding as well as being a symptom of protein misfolding including aberrant proteoforms [209], which is highly linked to neurological diseases [349]. Furthermore, the potential damage to mitochondrial integrity by misfolded proteins could also be resulting in the production of hydrogen peroxide within the cell [51]. Proteins of increased abundance were found to be part of the OST, one of the best characterised glycosylation pathways which in turn could be another mechanism the cell is utilising to decrease the burden of misfolded proteins as glycosylation is known to destabilise unfolded proteins, which enables proteostasis tools to access the protein [345].

Network analysis of the 220 proteins with decreased abundance revealed 36 ribosomal proteins alongside proteins involved in proline biosynthesis, including mitochondrial specific proteins as well as several proteins to have known pathogenic roles in ALS. Myelin expression factor 2, which is a transcriptional repressor, was found to be decreased suggesting that AZE treatment results in the increased expression of myelin basic protein, suggesting that myelin basic protein is specifically adversely affected by incorporation, possibly producing misfolded forms of the protein which coincides with its proline rich region at residues 90-102 which, if altered, would dramatically effect protein folding [85]. Myelin basic protein turnover has not been described in relation to AZE treatment of cells, and this data indicates it could be a candidate for degradomics

based analysis using pulse chase experiments similar to those performed in a recent Alzheimer's disease model (*see* [350]). Briefly; the supply of a stable isotope during the treatment of cells (known as: stable isotope labelling by amino acids in cell culture [SILAC]) with AZE labels the newly synthesised proteins with the introduced isotope, allowing comparison against the existing protein population prior to treatment which can then be measured over time [351]. Furthermore, the use of N-terminal isotopic labelling of substrates (TAILS) can be employed to study what cleavages events are occurring in a temporal fashion within the proteome, a method recently applied to study mitochondrial protein degradation in Parkinson's disease [352] (further reading *see* [353-356]. Alternatively, the use of radioactive tracers during treatment would result in newly synthesised proteins being detectable by scintillation or phosphor-imaging [108]. This method is more sensitive for an entire proteome turnover rate and should be employed alongside a proteomic based degradomics approach.

The proteins of increased abundance also included two heat shock proteins (70kDa and 105kDa) which are both strong regulators of proteostasis by acting as chaperones and preventing denatured protein aggregation, the 105kDa variant being also responsible for positive regulation of MHC class I biosynthesis implicating further potential immunological effects of the AZE treatment. The increase of heat shock protein 70kDa has been identified as a key feature of MS pathology, and this increase has also been highly linked to immunological response [357]. These proteins and their direction of abundance change further indicate a substantial basis for AZE producing MS like cellular responses.

5.4.3.3 Proline supplementation during Azetidine 2-carboxylic acid treatment reduces toxic effects

It has been shown in the literature that supplementation of proline alongside AZE treatment can prevent cellular toxicity and as such, proline was included in this experiment in order understand the proteome level changes that occur during this treatment. Analysis of the number of peptides identified upon proline supplementation resulted in a decrease in the number of peptides containing AZE, and it was also found that the average intensity of AZE peptides were decreased by 6.8-fold when compared to the AZE alone treatment. Furthermore, direct comparison of the AZE-peptides found in all biological replicates produced a statistically significant 12.7-fold reduction in ion intensity when compared to the AZE alone treatment. These metrics demonstrate a dramatic decrease in the number/amount of AZE-containing peptides upon proline supplementation, and it is foreseeable that further time course analysis alongside varying combinations would reveal an accurate concentration to outcompete AZE misincorporation. Quantitative analysis resulted in the identification of 54 proteins as differentially abundant with two keratins of increased abundance discarded from analysis. Two of the proteins found to be increased play a role in the prevention of cell death (ASNS, CHGA) indicating a pro-survival state in comparison to the AZE treatment alone which was largely shifted towards apoptosis, which agrees with previous literature [314].

The proteins of increased abundance are annotated as being involved in cell survival whilst the proteins of decreased abundance are involved in misfolding and cell redox homeostasis, and these data suggests that the supplementation of proline reduces the toxic effects of AZE by reducing protein misfolding and oxidative stress. This agrees with Song *et al* who demonstrated proline's rescue effect in Hela cell culture and zebrafish embryos [314]. The rescue of AZE toxicity exemplifies a competition for incorporation mechanism but does not provide useful information about the possible pathogenesis caused by *in vivo* consumption of AZE in the broader community. Amino acid analysis of food stocks which have AZE sources may prove useful in identifying presence but given the possible inducement of Multiple Sclerosis there should be a ban on the use of *Beta vulgaris* or at minimum a toxicity threshold established.

5.5 Conclusions

BMAA-treatment induced a pattern of toxicity similar to that previously reported for glutamate-mediated excitotoxicity. We were unable to detect any incorporation into proteins. The toxic changes identified could therefore be independent of incorporation and result from perturbations to other cellular processes. It has been consistently shown that L-serine protects against BMAA toxicity but this was not examined in the present study and is the subject of other work by the author. These data suggest that molecules other than proteins are impacted by BMAA, given that serine is protective upon BMAA treatment pathways that use this amino acid include sphingolipids that take serine as a head group. As such prime candidates for a metabolomic/lipidomic extraction and analysis would be sphingolipids and molecules involved in their metabolism as they contain L-serine and their synthesis could be impaired in the presence of BMAA.

AZE-treatment of cells resulted in incorporation that was readily detectable at a proteomic level, and the proteins containing AZE are annotated as being part of systems within the cell involved in proteostasis, with a large number of these found to be exosome related proteins suggesting proteins involved in export from the cell may be less likely to be degraded than other proteins. Quantitative analysis of the proteome resulted in the identification of proteins involved in protein folding being increased in abundance and the ribosomal machinery being decreased in abundance. The use of proline to reduce the toxic effects of AZE was shown to be effective and provides evidence that AZE induced misfolding could potentially be rescued. Analysis of the number and abundance of AZE containing peptides further revealed a significant and large decrease upon proline supplementation alongside AZE treatment. It is possible that the incorporation of AZE into proteins is of neurological significance and may be a contributing factor in the pathogenesis of MS, through outright toxic effects on the cell and possible immunological effects. There are a number of points where AZE can enter the human food chain but at this point none are under surveillance, as such levels of human exposure are not known.

Chapter 6: General discussion and future directions

6.1 Introduction

With an ageing population, neurodegenerative diseases are having an increased impact on human health and society. Extensive research in this space has employed genome centric approaches and this, to some extent, has resulted in the impact of environmental factors being overlooked [13, 23, 25-28]. Exposure to environmental factors is complicated by the possibility that multiple exposures could be required to trigger a disease state [297]. In addition, there are unknowns, such as the importance of exposure at the time when the nervous system is developing and the complexity of genetic susceptibility to specific environmental factors. [13, 23, 25-28]. The hallmarks of neurodegenerative diseases generally include protein aggregates in which the aggregation of particular proteins are associated strongly with specific diseases, termed as proteinopathies due to misfolding and deposition in cells, and include Alzheimer's, amyotrophic lateral sclerosis, prion disease, Huntington's and Parkinson's disease [12, 29-37]. Further characteristics of these diseases include metabolic changes, mitochondrial dysfunction and oxidative stress with, in some instances, a role for immunologically driven inflammation [33, 38-55].

Given the latency of neurodegenerative disease alongside a complex interplay of genetic factors as well as ageing; a proposed contributor to the pathogenesis is the mis-incorporation of NPAAs into proteins resulting in an acceleration in the formation of aberrant protein structures promoting neurodegeneration [79]. Research into NPAAs as possibly "toxic amino acid species" began with the work of Fowden *et al* in the 1960's in bacteria, and work has since been performed utilising various methods to determine toxicity and incorporation with the majority of studies using indirect methods to assess incorporation with only a few studies attempting to utilise proteomic methods [38, 82, 83, 87, 208, 314]. The absence of absolute evidence for incorporation within a single protein or at a global proteomic level has resulted in an incomplete understanding of the role of these NPAAs in human neurological disease [228].

This thesis aimed to study NPAAs, their localisation in proteoforms, and their effects on the cellular proteome using proteomic approaches. The primary aim was to determine if NPAAs are detectable within tryptic peptides derived from proteoforms and this primary aim was achieved for the NPAAs L-DOPA and AZE. Secondary to this, was to understand the effects of these NPAAs on the cellular proteome and investigate any connections to the phenotypes displayed during neuronal degeneration in humans. The misincorporation of the NPAA L-DOPA was demonstrated to result in protein misfolding and oxidative stress in chapter three. Similarly, AZE treatment resulted in both oxidative stress and protein misfolding. Furthermore, a significant number of AZE-peptides was confirmed,

suggesting mis-incorporation may play a substantial role in creating toxic neuronal stress. Proline supplementation alongside AZE treatment reduced detectable misincorporation alongside a decrease in the strength of the unfolded protein response.

6.2 Misincorporation proteomics is an emerging field

Researchers have made headway in the study of misincorporation and the mistranslation rate into proteoforms in general [312, 358, 359]. The use of affinity-based methods with recombinant protein expression systems has offered insight into mistranslation rates and has similarly been applied to some instances of misincorporation, namely BMAA and AZE [312, 314]. These methods have provided important insight into the mechanistic rates of incorporation, but further whole-proteome analyses need to be performed to understand if NPAAs offer a route of pathogenicity that could explain the causes of many sporadic neurological diseases that, despite many detailed studies, show no genetic link [79].

The published MiP review (chapter two) provided a guide to the best practice approaches and can be ubiquitously applied to analyse low abundance peptides present in numerous types of proteomic samples [228]. The use of analytical chemistry methods that target specific amino acids for increased sensitivity and quantification alongside proteomic experiments is emphasised in the review, in conjunction with radiolabelling treatments to trace the penetration of the NPAA into the proteome. The use of classic indirect methods alongside current proteomic approaches provides additional evidence for incorporation and determines if quantification of the NPAA incorporation is possible within a model system with regard to the dynamic range of a mass spectrometer. The application of DIA is highlighted to allow the future re-interrogation of datasets that may contain NPAAs where their analysis was not the original primary objective of the experiment, along with the use of other intra-dynamic range increasing methods such as BOXCAR. Furthermore, the use of targeted methods such as PRM/SRM is suggested in tandem with inclusion and exclusion lists, to acquire more meaningful proteomic data. The use of bioinformatic tools that incorporate open searching as part of the core functionality of the software by either *de novo* sequencing or fragment indexation is outlined as essential to the exploration of NPAAs, especially in the majority of instances where the site of incorporation is not well characterised or unknown. The use of tools that enable the prediction of retention times is also outlined, and the use of retention times as a metric for the presence of an NPAA in a peptide sequence adds an extra dimension of confidence once known positives are characterised along with the retention time shift. This review provided an overview of the pursuits so far in NPAA study and frames the pursuit as a formal direction to be taken in proteomics. Furthermore, it provides foundation to this thesis and outlines key considerations in analysing NPAA MiP data.

6.2.1 Future directions in field of misincorporation proteomics

The application of a combination of open searching and more intelligent mass-spectrometric data acquisition could provide key insight into the NPAA misincorporation into proteins and provide an understanding of how NPAAs impact upon structures and functions as they become increasingly more prevalent in the human environment by climate change [360] in the case of BMAA and dietary infiltration in the case of AZE [85, 361, 362]. Further exploration and development of novel methods is required, such as the hyphenation of approaches including; affinity-based methods developed for NPAAs of interest, *in silico* retention time prediction using neural networks, targeted mass spectrometry alongside DDA/DIA (referred to as hybrid acquisition strategies) and the use of open searching with large computational resources for analysis of all permutations of NPAA containing peptides. The application of analytical chemistry approaches used for small molecule quantification should be integrated into this framework.

Implementation of targeted MiP analysis would use stable isotope labelled spike in peptides that represent proteoforms that are important in the pathways of specific diseases, these are purchasable from several companies and should be used in futures analyses of any experiments i.e. JPT SpikeTides™ for cytokine signalling (JPT Peptide Technologies GmbH, Berlin, Germany). A multifaceted approach applying analytical techniques to the quantification of the NPAA in the amino acid synthetic pool alongside protein content should be performed, as this allows connection between wider analytical observations of presence alongside penetration into the proteome. The use of pulse chase labelling experiments and other radioactive tracer experiments alongside the study of *in vitro* incorporation should be performed to understand protein turnover and rates of degradation. Longitudinal proteomics studies should be performed on populations that are exposed to the NPAAs that are potential importance to human health alongside genome, transcriptome and metabolomic profiling; this could be by routine collection of sera alongside possible cerebral spinal fluid cumulating in brain biopsies once deceased. The generation of larger and more multi-omics datasets is essential to further build this fledging field and determine the role of NPAAs in neurodegeneration, this should be initiated by consultation with experts in sample acqusiton, sample processing, data acquisition, bioinformatics, clinical research and statistics. omics data generatation Comprehensive study of dietary intake of NPAAs alongside clusters of patients suffering neurodegeneration should be performed, to provide further evidence of exposure that might not be directly quantifiable from complex human samples. Immunological studies and cytokine signalling analyses should be performed on further models exploring NPAAs, exploring the potential of NPAAs to produce autoimmune responses and if these states are reflected in the identification of antigenic peptides.

When considering misincorporated NPAAs as the mechanism underpinning proteinopathies several aspects come to the fore including time, proteoforms, cell type, immune system and protein turnover. The misincorporated proteoforms produced may only be toxic under certain conditions and within certain populations of cells. Given enough time the benign miscincorporation containing proteoforms may become toxic i.e. molarity changes or exposure to other molecules in the cell. Furthermore, aging is known to lead to a decrease in proteostasis as cell health declines and metabolism slows, this may allow the correct conditions for where individual protein species overwhelm the ability of the cell to deal with these pathogenic aggregates[39, 40, 203, 211, 363]. Neuronal cell populations in the brain are long lived and do not divide, which would otherwise shed aggregate load into daughter cells, this makes these populations more susceptible to damage by protein aggregation[79, 200]. Given the sheer complexity of the brain itself and complex lifestyle factors in the population, large scale studies that are performed in a longitudinal manner may help to pull apart this tangle of factors.

6.3 A method for synthesising L-DOPA positive controls and investigation of L-DOPA as a possible source of oxidative stress

The possibility of L-DOPA incorporating into proteoforms could have ramifications for the treatment of Parkinson's patients and warranted an initial investigation *in vitro*. However, with a lack of available and reliable methods for the enrichment of L-DOPA containing peptides, a method for creating positive control peptides containing DOPA was needed to create reference fragmentation spectra for library searching of experimental data. In chapter three it was demonstrated that conversion of tryptic peptides to contain L-DOPA was possible using the enzyme tyrosinase, offering a new tool for the study of L-DOPA incorporation into peptides; it was demonstrated to allow identification transfer (match between runs in MaxQuant [291]) and revealed treatment-based incorporation of L-DOPA into peptides that would have otherwise been undetectable. This method resulted in the conversion of 10% of all tyrosine residues, and further optimisation needs to be performed prior to complete application on clinically relevant samples. In spite of this, the method offers an economically viable alternative to analyse thousands of peptide targets without the need for costly synthesis of all L-DOPA containing peptides in the human proteome.

It was also demonstrated that L-DOPA produces oxidative stress *in vitro* and protein misfolding via the effects on the protein expression and modification profile. These results may not be applicable *in vivo* due to the lower oxygen content, but this does suggest that the therapeutic supply of L-DOPA could be leading to an increase in oxidative burden in the patient's neuronal proteome. Even if this is not directly measurable it is probable that supplementation is leading to cellular stress as redox homeostasis is under pressure. Further exploration of L-DOPA incorporation was indicated and was the basis for chapter four's investigation of publicly available datasets of brain samples from Parkinson's patients likely exposed to L-DOPA.

6.3.1 Future directions in proteomic studies investigating PB-DOPA

The results of chapter 3 indicated that further optimisation of the enzymatic conversion procedure is required through the use of higher purity tyrosinase and heavy oxygen which could result in more accurate analysis of samples for their PB-DOPA content. Usage of artificial oxidation of the sample by heavy peroxide is also an avenue that should be explored alongside these conversions that could enable more robust quantification [229], as the incorporation of heavy oxygen allows relative comparison of the modified and unmodified species which would otherwise be impossible as different peptide species are not directly comparable due to having different chromatographic retention times and electrospray ionisation efficiencies (see chapter two). The use of synthetic peptides for targeted analysis in large cohorts is routinely used in large targeted quantitative studies and should be performed once targets are identified, while the application of tyrosinase to a pooled sample will

create a sample specific L-DOPA containing positive control which allows the rapid exploration of diverse samples without prior identification of possible targets.

In vivo analysis of animal models utilising L-DOPA at high dose, low dose, and chronic treatment may prove helpful in understanding the extent of L-DOPA incorporation into proteins *in vivo* as it is not impacted by a concurrent disease state in patients. Parkinson's longevity studies based on plasma analysis over the course of years or decades are necessary and would allow a pair wise analysis, performing proteomics alongside amino acid analysis may prove a more accurate experimental approach. Furthermore, application of complimentary techniques should be performed; redox staining, antibody analysis with western blotting or dot blotting for oxidative modifications, and targeted global proteomics for markers of oxidative stress or pathogenic proteoforms.

6.4 Meta-analyses enable re-interrogation of *in vivo* experiments, establishing the presence of L-DOPA in human proteomes

The ability to re-interrogate public data made it possible to explore experiments that would usually be inaccessible to researchers due to restricted funding and time constraints. However, the analyses performed in chapter four were hampered by poor annotation of sample attributes and experimental designs which were not focused on analysing NPAA incorporation. It is suggested that a stronger focus on more comprehensive annotation of datasets in repositories is performed to enable the selective reinterpretation of samples, this would be particularly valuable in the pursuit of oxidation and NPAA incorporation.

The re-analysis of these *in vivo* data identified ranges of DOPA in the proteome of between 0.2-1.5% of tyrosine residues with no statistical difference in abundance identified between the DOPA-containing peptides in Parkinson's samples versus healthy controls. Furthermore, no significant trend towards an increase in PB-DOPA compared to Ox-Phe was observed which might have supported incorporation of L-DOPA *in vivo*. These findings were similar to those found *in vitro* following the treatment of the SH-SY5Y cell line with L-DOPA in chapter three, which identified a base-line oxidation level is maintained after treatment with exogenous L-DOPA. The exploration of these peptides containing L-DOPA by an enrichment method is called for in future analyses.

6.4.1 Future directions

The reanalysis of datasets should be performed as more become available on the PRIDE repository. Several datasets identified for analysis were still under embargo and not publicly available, and as these enter the public domain it becomes possible to further explore these modifications of interest from clinically relevant *in vivo* samples. These meta-analyses were hampered by access to computational time that would make it possible to analyse large datasets in an acceptable time frame,

and since the undertaking of these analyses, faster search software has become available, such as that of FRAGPIPE, to perform open searching [228, 286-288]. Also, to this end, the advent of deploying proteomics analysis software on large computing clouds has enabled the reanalysis of much larger datasets, exemplified by a co-authored article with Prakash [364]. Given the low level of L-DOPA presence in proteomes, an enrichment method similar to that used in phospho-proteomics is required to identify the low abundance species that are below detectable levels. An enrichment method for this modification is being developed in house and was to be included with this thesis but clinical samples analysed had sample storage issues due to a change in location of the laboratory. Briefly, this method used alumina resin to bind DOPA peptides, a series of washing steps were then employed before the DOPA-containing peptides were released from the resin at a low pH. Method development using radiolabelled DOPA-containing peptides could affinity purify 95% of DOPA-containing peptides from a cell lysate. Preliminary proteomic analysis of treated cell line samples resulted in a 900% increase in the detectable amount of L-DOPA containing peptides (data not shown). In future, this approach will be applied to plasma and tissues from DOPA-treated mice and if successful could be applied to human tissues. Furthermore, the reanalysis of the clinical plasma samples which suffered sample storage issues is possible as aliquoted stocks were kept in reserve in case sample processing failed.

Analysis of separate regions of the brain in parallel to using a singular method of identification may prove more useful in untangling this oxidative pattern, such as a repurposing of the Localisation of Organelle Proteins by Isotope Tagging (LOPIT) method from the Lilley lab [365]. Briefly, this would require labelling of specific entire anatomical structures of the brain with individual TMT labels in order to understand the distribution of oxidation patterns across tissues and regions of the brain. Statistical methods for characterising organelle-based identification and localisation are already employed in the LOPIT method and if applied to a well sectioned brain, it would provide a map of the oxidative profile and, given enough points of reference, an accurate representation of oxidation within each structure would be created [365]. This method may prove useful but with the caveat that these data have identified a bias towards homology and mask unique species in TMT workflows. As such, application of an enrichment method on already labelled peptide pools may further enhance the identification of oxidation sites.

6.5 The NPAA azetidine 2-carboxylic acid is readily incorporated into proteomes triggering a cellular response to aberrant proteins.

The work of chapter five identified that AZE is able to incorporate into proteins in place of proline with the highest content protein species being those annotated to be a part of the unfolded protein response, which were also found to be quantitatively increasing in abundance. The co-supplementation of proline alongside AZE significantly lowers the number and amount of AZE mis-incorporation, coinciding with a decrease in the strength of the unfolded protein response.

6.5.1.1 Role in multiple sclerosis

Rubenstein's hypothesis that AZE is a factor in Multiple Sclerosis is ground breaking and seminal [85]; his hypothesis is well constructed but has been supported only by limited experimental analysis by himself and others. The principle underlying the hypothesis is further supported by the data presented in chapter 5, which found that AZE is readily incorporated into proteins and induces a strong unfolded protein response indicated by the increased abundance of the proteins involved the response. The diversity of proteins identified that contain AZE reinforces the likelihood that given sufficient exposure time, immunogenic species from structural proteins such as myelin will form, which could be a trigger to Multiple Sclerosis. It was further shown that the use of proline co-supplementation lowers the toxicity that results from AZE treatment and reduces the unfolded protein response within cells. A recent publication from our laboratory also demonstrated that AZE results in inflammation and apoptosis in microglial cells, further reinforcing its role in immunological changes [366].

6.5.1.2 Potential role for AZE in wider diseases of neurological nature

AZE incorporation may induce protein misfolding and subsequently oxidative stress requiring the autophagic and lysosomal systems of the cell to be upregulated [85, 87, 107, 209]. Rubenstein's work stops short on proposing that this NPAA may be a trigger for other neurological diseases through promoting the synthesis of non-native proteins. The work of Al-Chalabi identified that six insults are required to result in the precipitation of MND and it is a possibility that exposure to enough AZE at a time when proline is low in dietary intake may form an insult on the terminally differentiated neurons within the CNS [296]. MS has a much higher incidence in parts of the world where sugar beet (rather than cane) are the main source of sugar. It is important to note that the extracted sugar beet pulp is widely used as a stock feed in places such as Tasmania which have a very high incidence of MS [367].

6.5.1.3 Future directions for analysis and implication of AZE incorporation

This work indicated that AZE is indeed incorporated into the proteome at detectable levels resulting in an increased abundance of proteins involved misfolding, indicating a response to damaged proteins in neuronal cells. Experiments that should be performed to validate and build on this work include treatment of cells that produce myelin such as oligodendrocytes, as this would provide a model close to what would occur *in vivo* upon treatment as they are the targets of MS based neuronal degradation [368]. Furthermore, the analysis of AZE treatment of primary motor neurones may provide useful information in comparison to familial primary cells of MND as this may demonstrate an insult that could result in MND. The study of AZE treated primates might provide insight into potential links between MS and AZE exposure. Post-mortem analysis of tissues from MS patients for the presence of AZE could provide evidence of AZE exposure.

6.5.2 BMAA is not incorporated into the proteome at a detectable level

In chapter five, BMAA incorporation was investigated via proteomics and no positive spectral identification was obtained to support its incorporation into proteins during protein synthesis. While BMAA is consistently found in the 'protein fraction' in extracts of a wide range of material [63, 65, 67, 103, 104, 110, 112, 113, 115, 116, 361, 369-372] we did not obtain any evidence for incorporation into the peptide chain. This was also reported elsewhere and the link between incorporation and release from protein is ambiguous, as there are other small molecules that may be attached to BMAA. However, BMAA's role in inducing neurodegenerative disease is supported by epidemiological studies [70, 373-376] but the lack of actual evidence of incorporation could suggest another mechanism of toxicity and other avenues of systems biology study are required to untangle the actual role of BMAA. Alternative mechanisms of toxicity besides incorporation have been reported within the literature on BMAA, these include its ability to be a glutamate receptor agonist leading to overstimulation, interactions with neuromelanin leading to oxidative stress [63, 104, 111, 185, 302, 305, 306, 324]

Data not displayed within this thesis includes more than seven complete BMAA treatment proteomic experiments whereby no identification of incorporation was found, including the bioinformatic analysis of BMAA replacing every possible amino acid and possibly as a conjugate or forming a novel PTM by other structural bindings. Furthermore, experiments using a neuronal differentiated cell line along with a wide array of BMAA dosages and treatment times also resulted in no detectable incorporation. Research around BMAA's role in MND should focus on other potential avenues of toxicity rather than misincorporation, and there does appear to be evidence that proteinopathic states are resultant of BMAA treatment and this may be due to a functional interaction or incorporation into other molecular species.

6.6 General future directions

This thesis has investigated the incorporation effects of three NPAAs by proteomic approaches. Besides the three studied herein, there are several other NPAAs of potential significance to human health including the NPAAs norvaline, canavanine, homoarginine, homolysine and several other modified amino acids that have been shown to be toxic [228]. An analysis of a norvaline treated dataset was performed and while the replacement of valine results in no mass shift and was not assessed, the incorporation in place of leucine/isoleucine was investigated yielding no definitive spectral evidence [228]. A potential method to explore its incorporation in a proteomic manner would be to analyse peptides by mass spectrometry while using ion mobility separation, with synthesised peptides to understand the isomeric effect on collisional cross section/drift time. The use of amino acid analysis with derivatisation may also allow the separation of the norvaline isomer from valine, which would justify the costs involved in synthesising norvaline containing standards.

6.6.1 Technologies that can be applied non-protein amino acid proteomics studies

The use of targeting methods using large *in silico* generated inclusion lists is a powerful method that could be used for identifying possible low abundant species peptides containing a misincorporated NPAA, working by only accumulating ions of interest from the ion beam and selectively monitoring reactions for diagnostic ions of interest. A program created by the Aebersold lab called SuperHirn [279] was able to generate iterative inclusion and exclusion lists on an LTQ orbitrap mass spectrometer and this program was trialled in this work but was not successful due to incompatibilities in software versioning. A similar line of enquiry was undertaken through the utilisation of MaxQuant Live [377] generated inclusion-exclusion-lists which were readily able to identify several thousand new peptides from sample injections, but none were positive for BMAA incorporation. Modifying the Xcalibur based AcquireX program used in Tribrid based Orbitrap mass spectrometers (also included in the Exploris series) may be feasible, as this program is utilised for metabolic experiments in the exact way that Superhirn was employed whereby iterative analyses generate inclusion and exclusion lists. By modifying AcquireX and forcing the program to dynamically load inclusion and exclusion lists for peptides, the depth of acquisition on NPAA containing samples would be unparalleled and would likely offer an advance to ultra-depth proteomics in itself. Proteomic analysis is quoted to be incompatible with the AcquireX software due to the large amount of precursor ions that elute during chromatographic runs, but this does not consider the use of fine-tuned FAIMS to include inclusion lists of only the charge state of interest. As such, previously identifying peptides of high abundance and subsequently creating retention time aligned targeting windows for modified ions based on this will decrease the limits of detection possibly leading to identification.

Given that low abundance of modified peptides is an inherent problem in any mis-translational/misincorporation analysis, use of larger mass spec ion beam capture (fill times/accumulation times) and higher resolutions may provide a better chance of detection. With the sensitivity of the latest generation of both the Bruker systems (TIMS-TOF series) and the Thermo Fisher Scientific systems (Exploris 480 Orbitrap and Eclipse Tribrid), it becomes possible to analyse proteomes with less than 200 pg of starting material [272]. Strategies to modify methods, and utilization of ion mobility separations, can increase the ability to resolve out novel species, as is exemplified by the use of FAIMS in phosphoproteomics experiments [378].

The use of antibody-based affinity pull downs for targets is routinely performed across the field of proteomics as it enables identification of target species that would otherwise be undetectable and as such this should be performed on the NPAAs that are the subject of this thesis [379]. The use of antibody-based affinity methods is possible, but raising highly specific antibodies to the side chain of the NPAA would be cost-prohibitive and time consuming for routine use. The exploratory generation of a metal-chemical affinity method for the NPAA L-DOPA was performed using radioactively labelled L-DOPA resulting in good initial data (data not shown). Enrichment resulted in the increase of detectable L-DOPA by several hundred-fold, and subsequently this was applied to a Parkinson's plasma sample set and analysed. Unfortunately, issues with sample acquisition and storage led to extreme sample degradation and confounding results. The results of this investigation were originally slated to be a published chapter of this thesis. As such, further repetition of this experiment is desirable, and would likely result in novel data about the incorporation of L-DOPA into the proteins present in plasma of Parkinson's patients. There is also the possibility to automate the enrichment method utilising liquid handling robots capable of performing bead-based preparation like that of Agilent's Bravo system [380] or the Thermo-Fisher platform, the KingFisher [381], which will enable processing steps to be carried out in a highly efficient manner, limiting storage time during intermediary processing steps.

6.6.2 Biological experiments that should be performed to investigate each NPAA in order to identify misincorporation

Each NPAA comes with its own unique challenges in the exploration of its role in a wider biological context and as such, experiments that should be performed that would better capture the misincorporation status of samples are outlined herein. The NPAA L-DOPA should be explored in well annotated *in vivo* experiments involving patients with and without L-DOPA supplementation and age matched controls, the specific goals being to quantify the amount of PB-DOPA and other oxidative markers. Ideally, sample numbers should be in the hundreds and consist of post-mortem tissues, plasma obtained over a time course of years and properly stored, and possibly CSF samples to adequately investigate the propagation of PB-DOPA through the brain. The analysis of such a

comprehensive dataset via proteomics would allow for not only the investigation of the MiPome but also metrics that may either reinforce or refute its involvement in neurodegeneration. For the NPAA AZE, experiments covering several avenues should be explored in parallel, including immunopeptidomic analysis performed using proteins created *in vitro*. Ideally the use of high myelin expressing cell lines should be used and subsequently, the treatment of animal models. The ideal models for AZE treatment would be either murine or primate, and a chronic model of AZE exposure is of paramount importance. All biological fluids should be investigated for AZE incorporation by amino acid analysis and proteomic methods, while nerve potential studies alongside these data may also offer insight into the breakdown of the myelin sheath.

6.7 Concluding Remarks

The results presented in this thesis provide evidence that certain NPAAs can incorporate into the human proteome *in vitro* at detectable levels. The exploration into the effects of these NPAAs via proteomics is a feasible undertaking, with many tools already established that can help explore their role in the wider context of human disease [228]. The application of proteomics to exploring biological problems is accelerating and has become ubiquitous with most forms of system-wide analysis endeavours, revolutionising research, with its tools developed by the day. As neurodegeneration continues to increase alongside longevity of humans, the study of these diseases/proteinopathies will continue to advance with the exploration by proteomics at the foreseeable forefront. NPAAs are pervasive and distributed through our diet and their links to neurological and even immunological disease has been established but not fully explored. With the advent of more advanced protein sequencing technologies coinciding with the quantum leap of protein structure prediction offered by Alpha Fold, a dramatic increase in the capacity to accurately predict proteoform structures is possible [382]. The pieces are beginning to fall into place to understand how the integration of NPAAs into the human proteome may result toxic proteoforms and ultimately negatively impact on human health, and the promise of proteomic enabled precision medicine may offer a therapeutic horizon.

Appendices

Alongside this thesis there are several large datasets and supplementary files, the organisation of these appendices are numbered according to the chapters of the thesis.

A_1: General files and installers

Software installers are archived within this repository to enable reproduction of analyses; https://drive.google.com/drive/folders/1D9wk41QFNpV8SWIyiU-6k62_mNs8V1Q2?usp=sharing. Within this folder the indexed human proteome can be found, which was used for all analyses named "Human Proteome PhD indexed 3.9.17.fasta".

A_2: No associated data

A_3: Chapter 3

All data generated by this chapter can be accessed at PRIDE DB under the identifier: PXD02575. Furthermore, biological analyses can be accessed at: https://app.massdynamics.com/p/dfbb55e9-0a6d-450e-8e8e-3f0c070e5bf9

A_4: Chapter 4

Hyperlink to chapter files:
https://drive.google.com/drive/folders/1RBV9QAC7sbvi5OUJjl7VBS89vaT1UfH0?usp=sharing

Appendix 4_1_1 Human brain proteome Peaks project

The complete Peaks Studio project file can be accessed by this hyperlink: https://drive.google.com/drive/folders/1RC09GlSVnwVFBKpRt2GOjGT_7hWLbrzn?usp=sharing. This project file can be opened freely by Peaks Studio by activating the viewer mode, versions past 10.5 will be able to open this project. Files size is 30GB uncompressed.

Appendix 4_1_2 Human brain proteome outputs and supplementary files

A) Peaks outputs: Outputs for protein and peptide Level

"[PXD000561 Human Brain Proteome_Complete PTM Search_134pep25pro20_1](#)"

The files within this folder are:

peptide features.csv

peptide.csv

Supplementary file 1; used to perform peptide modification analysis.

protein-peptides.csv

proteins.csv

DB search psm.csv

de novo only peptides.csv

B) Analysis files for each modification (L-DOPA and Ox-Phe). "A" denotes Ascore threshold applied to each file, "wAcsn" are peptides with an accession. The "Worked" excel documents are the final steps in analysis.

DOPA_peptides_500A.csv

DOPA_Peptides_A500_wAcsn.csv

DOPA_Peptides_A500_wAcsn_Worked.xlsx

OXPhe_peptide_A500.csv

OXPhe_peptide_A500_wAcsn.csv

OXPhe_peptide_A500_wAcsn_worked.xlsx

Venn_Diagram_IDs.xlsx

Appendix 4_2_1 Substantia nigra TMT (PXD000427) Peaks project file

The complete Peaks Studio project file can be accessed by this hyperlink: https://drive.google.com/drive/folders/1owelINsLvgxdqTO5tYjibF31qjjpPRh5?usp=sharing. This project file can be opened freely by Peaks Studio by activating the viewer mode, versions past 10.5 will be able to open this project. Files size is 9GB uncompressed.

Appendix 4_2_2 Substantia nigra TMT (PXD000427) outputs and supplementary files

A) Peaks outputs and associate analyses: Outputs for PTM search and Quantitation. "PXD000427 PD Substantia Nigra_QUANTITATION_28". The key files are "Quant_PSM_Analysis.xlsx" which was the end product of PSM level quantitation and the Graphpah prism file: "Graphpad_PDsn_PSM_AQUant.pzfx" used to generate Figure 8.

B) "PXD000427 PD Substantia Nigra_QUANTITATION_28_non_sig_for_albumin" Used to determine the presence of Albumin and its quantitative ratio found in "proteins.csv".

C) "PXD000427 PD Substantia Nigra_PEAKS PTM_27_1FDR", this output-set was used to perform peptide centric analysis for the modifications of interest. The key file within this analysis is "Supplementary file 1".

Appendix 4_3_1 Olfactory proteome (PXD008036) Peaks project file

The complete Peaks Studio project file can be accessed by this hyperlink: https://drive.google.com/drive/folders/1JgV-TPRluur9RxJ9jmhVSw8Azz3lODLr?usp=sharing. This project file can be opened freely by Peaks Studio by activating the viewer mode, versions past 10.5 will be able to open this project. Files size is 22.8GB uncompressed.

Appendix 4_3_2 Olfactory proteome (PXD008036) outputs and supplementary files

A) Peaks outputs and associated analyses: "PXD008036_190714_PEAKS PTM_28_01FDR". The key files include; Graphpad prism file used for the generation of Figures 13-17 named "Data_analysis_Olfactory.pzfx" along with the excel file used in the analysis of modification sites named "Supplementary file 1".

B) An indexed version of the R-analysis code used to generate data, this file has been created using the zipping program 7zip. "PXD008036_190714_PEAKS PTM_28_01FDR_Ranalyses.7z"

A_5:Chapter 5

The hyperlink to chapter appendix:
https://drive.google.com/drive/folders/1o9fMqDh545A82SgS_mKInkZZw4bCXwLa?usp=sharing.

Appendix 5_1 Peaks project file

The complete Peaks Studio project file can be accessed by this hyperlink:
https://drive.google.com/drive/folders/1pF0PJD0vbkKgGV65qNyqg0VoXFxGBC3h?usp=sharing. This
project file can be opened freely by Peaks Studio by activating the viewer mode, versions past 10.5
will be able to open this project. Files size is 28.5gb uncompressed.

Appendix 5_2 Outputs and supplementary files

A) Peaks PTM search outputs containing the complete set of files: "NPAA Experiment Ctrl BMAA
AZE AZ+PRO_1_PEAKS PTM_83_fullCSV"

B) Peaks LFQ search outputs containing the complete set of files: "NPAA Experiment Ctrl BMAA
AZE AZ+PRO_1_LABEL FREE_85csvs"

C) Peaks PTM search analysis files for peptide centric analyses: "NPAA Experiment Ctrl BMAA
AZE AZ+PRO_1_PEAKS PTM_83_AZE_csv", the two worked files include an intensity analysis
used to generate Figure 18 ("Intensity analysis of AZE peptides.xlsx") and the peptide centric
analysis with manually verified spectra ("Peptide_centric_analysis.xlsx", sheet: Final filtered
contains all spectra validated). The GraphPad prism project file can also be found within this
appendix.

References

1. Di Domenico, F., et al., *The triangle of death in Alzheimer's disease brain: the aberrant cross-talk among energy metabolism, mammalian target of rapamycin signaling, and protein homeostasis revealed by redox proteomics.* Antioxidants & redox signaling, 2017. **26**(8): p. 364-387.

2. *Neurological disorders : public health challenges.* 2006, World Health Organisation: Switzerland. p. 39-51.

3. Demartini, D.R., et al., *Alzheimer's and Parkinson's diseases: an environmental proteomic point of view.* J Proteomics, 2014. **104**(Supplement C): p. 24-36.

4. Xicoy, H., Wieringa, B., and Martens, G.J., *The SH-SY5Y cell line in Parkinson's disease research: a systematic review.* Mol Neurodegener, 2017. **12**(1): p. 10.

5. Zhang, H., Duan, C., and Yang, H., *Defective autophagy in Parkinson's disease: lessons from genetics.* Mol Neurobiol, 2015. **51**(1): p. 89-104.

6. Xin, W., et al., *Toxic Oligomeric Alpha-Synuclein Variants Present in Human Parkinson's Disease Brains Are Differentially Generated in Mammalian Cell Models.* Biomolecules, 2015. **5**(3): p. 1634-51.

7. Tofaris, G.K., Layfield, R., and Spillantini, M.G., *α-Synuclein metabolism and aggregation is linked to ubiquitin-independent degradation by the proteasome.* FEBS Letters, 2001. **509**(1): p. 22-26.

8. McNaught, K.S. and Olanow, C.W., *Protein aggregation in the pathogenesis of familial and sporadic Parkinson's disease.* Neurobiol Aging, 2006. **27**(4): p. 530-45.

9. Blesa, J. and Przedborski, S., *Parkinson's disease: animal models and dopaminergic cell vulnerability.* Front Neuroanat, 2014. **8**: p. 155.

10. Bennett, M.C., et al., *Degradation of alpha-synuclein by proteasome.* J Biol Chem, 1999. **274**(48): p. 33855-8.

11. Grace, E.A. and Busciglio, J., *Aberrant activation of focal adhesion proteins mediates fibrillar amyloid beta-induced neuronal dystrophy.* J Neurosci, 2003. **23**(2): p. 493-502.

12. Hipp, M.S., Park, S.H., and Hartl, F.U., *Proteostasis impairment in protein-misfolding and -aggregation diseases.* Trends Cell Biol, 2014. **24**(9): p. 506-14.

13. Therrien, M., Dion, P.A., and Rouleau, G.A., *ALS: Recent Developments from Genetics Studies.* Curr Neurol Neurosci Rep, 2016. **16**(6): p. 59.

14. Van Cauwenberghe, C., Van Broeckhoven, C., and Sleegers, K., *The genetic landscape of Alzheimer disease: clinical implications and perspectives.* Genet Med, 2016. **18**(5): p. 421-30.

15. Vogelgesang, S., et al., *Deposition of Alzheimer's beta-amyloid is inversely correlated with P-glycoprotein expression in the brains of elderly non-demented humans.* Pharmacogenetics, 2002. **12**(7): p. 535-41.

16. Valente, E.M., et al., *Hereditary early-onset Parkinson's disease caused by mutations in PINK1.* Science, 2004. **304**(5674): p. 1158-60.

17. Tsuji, T., et al., *Proteomic profiling and neurodegeneration in Alzheimer's disease.* Neurochem Res, 2002. **27**(10): p. 1245-53.

18. Suh, Y.H. and Checler, F., *Amyloid precursor protein, presenilins, and alpha-synuclein: molecular pathogenesis and pharmacological applications in Alzheimer's disease.* Pharmacol Rev, 2002. **54**(3): p. 469-525.

19. Spires-Jones, T.L. and Hyman, B.T., *The intersection of amyloid beta and tau at synapses in Alzheimer's disease.* Neuron, 2014. **82**(4): p. 756-71.

20. Hallengren, J., Chen, P.C., and Wilson, S.M., *Neuronal ubiquitin homeostasis.* Cell Biochem Biophys, 2013. **67**(1): p. 67-73.

21. Nivon, M., et al., *NFkappaB is a central regulator of protein quality control in response to protein aggregation stresses via autophagy modulation.* Mol Biol Cell, 2016. **27**(11): p. 1712-27.

22.   Kopito, R.R., *Aggresomes, inclusion bodies and protein aggregation.* Trends Cell Biol, 2000. **10**(12): p. 524-30.

23.   Ghasemi, M. and Brown, R.H., Jr., *Genetics of Amyotrophic Lateral Sclerosis.* Cold Spring Harb Perspect Med, 2018. **8**(5).

24.   Chiesa, P.A., et al., *Revolution of Resting-State Functional Neuroimaging Genetics in Alzheimer's Disease.* Trends Neurosci, 2017. **40**(8): p. 469-480.

25.   de la Torre, J.C., *Genetics of Alzheimer's Disease*, in *Alzheimer's Turning Point*. 2016, Springer International Publishing: Cham. p. 75-83.

26.   Feany, M.B., *ASIP Outstanding Investigator Award Lecture. New approaches to the pathology and genetics of neurodegeneration.* Am J Pathol, 2010. **176**(5): p. 2058-66.

27.   Renton, A.E., Chio, A., and Traynor, B.J., *State of play in amyotrophic lateral sclerosis genetics.* Nat Neurosci, 2014. **17**(1): p. 17-23.

28.   Spatola, M. and Wider, C., *Genetics of Parkinson's disease: the yield.* Parkinsonism Relat Disord, 2014. **20 Suppl 1**: p. S35-8.

29.   Lim, J. and Yue, Z., *Neuronal aggregates: formation, clearance, and spreading.* Dev Cell, 2015. **32**(4): p. 491-501.

30.   Yang, H. and Hu, H.Y., *Sequestration of cellular interacting partners by protein aggregates: implication in a loss-of-function pathology.* FEBS J, 2016. **283**(20): p. 3705-3717.

31.   Golde, T.E., et al., *Thinking laterally about neurodegenerative proteinopathies.* J Clin Invest, 2013. **123**(5): p. 1847-55.

32.   Aguzzi, A. and Rajendran, L., *The transcellular spread of cytosolic amyloids, prions, and prionoids.* Neuron, 2009. **64**(6): p. 783-90.

33.   Michel, P.P., Hirsch, E.C., and Hunot, S., *Understanding Dopaminergic Cell Death Pathways in Parkinson Disease.* Neuron, 2016. **90**(4): p. 675-91.

34.   Labbadia, J. and Morimoto, R.I., *The biology of proteostasis in aging and disease.* Annu Rev Biochem, 2015. **84**(1): p. 435-64.

35.   Gidalevitz, T., Prahlad, V., and Morimoto, R.I., *The stress of protein misfolding: from single cells to multicellular organisms.* Cold Spring Harb Perspect Biol, 2011. **3**(6).

36.   Frost, B., Jacks, R.L., and Diamond, M.I., *Propagation of tau misfolding from the outside to the inside of a cell.* J Biol Chem, 2009. **284**(19): p. 12845-52.

37.   Soto, C. and Estrada, L.D., *Protein misfolding and neurodegeneration.* Arch Neurol, 2008. **65**(2): p. 184-9.

38.   Beri, J., et al., *Exposure to BMAA mirrors molecular processes linked to neurodegenerative disease.* Proteomics, 2017. **17**(17-18): p. 10.

39.   Kaufman, D.M., et al., *Ageing and hypoxia cause protein aggregation in mitochondria.* Cell Death Differ, 2017. **24**(10): p. 1730-1738.

40.   Butterfield, D.A., Palmieri, E.M., and Castegna, A., *Clinical implications from proteomic studies in neurodegenerative diseases: lessons from mitochondrial proteins.* Expert Rev Proteomics, 2016. **13**(3): p. 259-74.

41.   Gatt, A.P., et al., *Dementia in Parkinson's disease is associated with enhanced mitochondrial complex I deficiency.* Mov Disord, 2016. **31**(3): p. 352-9.

42.   Nisar, R., et al., *Diquat causes caspase-independent cell death in SH-SY5Y cells by production of ROS independently of mitochondria.* Arch Toxicol, 2015. **89**(10): p. 1811-25.

43.   Ryan, B.J., et al., *Mitochondrial dysfunction and mitophagy in Parkinson's: from familial to sporadic disease.* Trends Biochem Sci, 2015. **40**(4): p. 200-10.

44.   Licker, V., et al., *Proteomic analysis of human substantia nigra identifies novel candidates involved in Parkinson's disease pathogenesis.* Proteomics, 2014. **14**(6): p. 784-94.

45.   Dexter, D.T. and Jenner, P., *Parkinson disease: from pathology to molecular disease mechanisms.* Free Radic Biol Med, 2013. **62**: p. 132-144.

46.   Greene, A.W., et al., *Mitochondrial processing peptidase regulates PINK1 processing, import and Parkin recruitment.* EMBO Rep, 2012. **13**(4): p. 378-85.

47. Exner, N., et al., *Mitochondrial dysfunction in Parkinson's disease: molecular mechanisms and pathophysiological consequences.* EMBO J, 2012. **31**(14): p. 3038-62.

48. Chou, J.L., et al., *Early dysregulation of the mitochondrial proteome in a mouse model of Alzheimer's disease.* J Proteomics, 2011. **74**(4): p. 466-79.

49. Schneider, L., et al., *Differentiation of SH-SY5Y cells to a neuronal phenotype changes cellular bioenergetics and the response to oxidative stress.* Free Radic Biol Med, 2011. **51**(11): p. 2007-17.

50. Geisler, S., et al., *PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1.* Nat Cell Biol, 2010. **12**(2): p. 119-31.

51. Lin, M.T. and Beal, M.F., *Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases.* Nature, 2006. **443**(7113): p. 787-95.

52. Hinerfeld, D., et al., *Endogenous mitochondrial oxidative stress: neurodegeneration, proteomic analysis, specific respiratory chain defects, and efficacious antioxidant therapy in superoxide dismutase 2 null mice.* J Neurochem, 2004. **88**(3): p. 657-67.

53. Reimertz, C., et al., *Gene expression during ER stress-induced apoptosis in neurons: induction of the BH3-only protein Bbc3/PUMA and activation of the mitochondrial apoptosis pathway.* J Cell Biol, 2003. **162**(4): p. 587-97.

54. Ravindranath, V., *Neurolathyrism: mitochondrial dysfunction in excitotoxicity mediated by L-beta-oxalyl aminoalanine.* Neurochem Int, 2002. **40**(6): p. 505-9.

55. Hirai, K., et al., *Mitochondrial abnormalities in Alzheimer's disease.* J Neurosci, 2001. **21**(9): p. 3017-23.

56. Butterfield, D.A. and Perluigi, M., *Redox Proteomics: A Key Tool for New Insights into Protein Modification with Relevance to Disease.* Antioxid Redox Signal, 2017. **26**(7): p. 277-279.

57. Chan, S.W., et al., *L-DOPA is incorporated into brain proteins of patients treated for Parkinson's disease, inducing toxicity in human neuroblastoma cells in vitro.* Exp Neurol, 2012. **238**(1): p. 29-37.

58. Dunlop, R.A., Brunk, U.T., and Rodgers, K.J., *Proteins containing oxidized amino acids induce apoptosis in human monocytes.* Biochem J, 2011. **435**(1): p. 207-16.

59. Spencer, P.S., et al., *Guam amyotrophic lateral sclerosis-parkinsonism-dementia linked to a plant excitant neurotoxin.* Science, 1987. **237**(4814): p. 517-22.

60. Banack, S.A. and Cox, P.A., *Biomagnification of cycad neurotoxins in flying foxes: implications for ALS-PDC in Guam.* Neurology, 2003. **61**(3): p. 387-9.

61. Cox, P.A. and Sacks, O.W., *Cycad neurotoxins, consumption of flying foxes, and ALS-PDC disease in Guam.* Neurology, 2002. **58**(6): p. 956-9.

62. Main, B.J., Dunlop, R.A., and Rodgers, K.J., *The use of L-serine to prevent beta-methylamino-L-alanine (BMAA)-induced proteotoxic stress in vitro.* Toxicon, 2016. **109**: p. 7-12.

63. Cox, P.A., et al., *Dietary exposure to an environmental toxin triggers neurofibrillary tangles and amyloid deposits in the brain.* Proc Biol Sci, 2016. **283**(1823).

64. Rodgers, K.J., Samardzic, K., and Main, B.J., *Toxic Nonprotein Amino Acids*, in *Plant Toxins*, P. Gopalakrishnakone, R.C. Carlini, and R. Ligabue-Braun, Editors. 2015, Springer Netherlands: Dordrecht. p. 1-20.

65. Jiang, L., *Mass Spectrometry of Non-protein Amino Acids: BMAA and Neurodegenerative Diseases.* 2015.

66. Dunlop, R.A., Main, B.J., and Rodgers, K.J., *The deleterious effects of non-protein amino acids from desert plants on human and animal health.* Journal of Arid Environments, 2015. **112**: p. 152-158.

67. Berntzon, L., et al., *Detection of BMAA in the human central nervous system.* Neuroscience, 2015. **292**: p. 137-47.

68. Juntas-Morales, R., et al., *[Environmental factors in ALS].* Presse Med, 2014. **43**(5): p. 549-54.

69. Stommel, E.W., Field, N.C., and Caller, T.A., *Aerosolization of cyanobacteria as a risk factor for amyotrophic lateral sclerosis.* Med Hypotheses, 2013. **80**(2): p. 142-5.

70.   Bradley, W.G., et al., *Is exposure to cyanobacteria an environmental risk factor for amyotrophic lateral sclerosis and other neurodegenerative diseases?* Amyotroph Lateral Scler Frontotemporal Degener, 2013. **14**(5-6): p. 325-33.

71.   Huang, T., Rehak, L., and Jander, G., *meta-Tyrosine in Festuca rubra ssp. commutata (Chewings fescue) is synthesized by hydroxylation of phenylalanine.* Phytochemistry, 2012. **75**: p. 60-6.

72.   Jiao, C.J., et al., *Factors affecting beta-ODAP content in Lathyrus sativus and their possible physiological mechanisms.* Food Chem Toxicol, 2011. **49**(3): p. 543-9.

73.   Nunn, P.B., et al., *Toxicity of non-protein amino acids to humans and domestic animals.* Nat Prod Commun, 2010. **5**(3): p. 485-504.

74.   Bessonov, K., Bamm, V.V., and Harauz, G., *Misincorporation of the proline homologue Aze (azetidine-2-carboxylic acid) into recombinant myelin basic protein.* Phytochemistry, 2010. **71**(5-6): p. 502-7.

75.   Pablo, J., et al., *Cyanobacterial neurotoxin BMAA in ALS and Alzheimer's disease.* Acta Neurol Scand, 2009. **120**(4): p. 216-25.

76.   Daggett, V. and Fersht, A.R., *Is there a unifying mechanism for protein folding?* Trends Biochem Sci, 2003. **28**(1): p. 18-25.

77.   Jahn, T.R. and Radford, S.E., *The Yin and Yang of protein folding.* FEBS J, 2005. **272**(23): p. 5962-70.

78.   Matsumoto, H., *The mechanisms of phytotoxic action and selectivity of non-protein aromatic amino acids L-DOPA and m-tyrosine.* Journal of Pesticide Science, 2011. **36**(1): p. 1-8.

79.   Rodgers, K.J., *Non-protein amino acids and neurodegeneration: the enemy within.* Exp Neurol, 2014. **253**: p. 192-6.

80.   Bergmann, T.J., Pisoni, G.B., and Molinari, M., *Quality control mechanisms of protein biogenesis: proteostasis dies hard.* 2016.

81.   Drummond, D.A. and Wilke, C.O., *The evolutionary consequences of erroneous protein synthesis.* Nat Rev Genet, 2009. **10**(10): p. 715-24.

82.   Fowden, L., Lewis, D., and Tristram, H., *Toxic amino acids: their action as antimetabolites.* Adv Enzymol Relat Areas Mol Biol, 1967. **29**: p. 89-163.

83.   Fowden, L. and Richmond, M.H., *Replacement of proline by azetidine-2-carboxylic acid during biosynthesis of protein.* Biochimica et Biophysica Acta, 1963. **71**: p. 459-461.

84.   Kiick, K.L. and Tirrell, D.A., *Protein Engineering by In Vivo Incorporation of Non-Natural Amino Acids: Control of Incorporation of Methionine Analogues by Methionyl-tRNA Synthetase.* Tetrahedron, 2000. **56**(48): p. 9487-9493.

85.   Rubenstein, E., *Misincorporation of the proline analog azetidine-2-carboxylic acid in the pathogenesis of multiple sclerosis: a hypothesis.* J Neuropathol Exp Neurol, 2008. **67**(11): p. 1035-40.

86.   Rubenstein, E., et al., *Azetidine-2-carboxylic acid in the food chain.* Phytochemistry, 2009. **70**(1): p. 100-4.

87.   Rodgers, K.J. and Shiozawa, N., *Misincorporation of amino acid analogues into proteins by biosynthesis.* Int J Biochem Cell Biol, 2008. **40**(8): p. 1452-66.

88.   Zagari, A., et al., *The effect of the l-azetidine-2-carboxylic acid residue on protein conformation. IV. Local substitutions in the collagen triple helix.* Biopolymers, 1994. **34**(1): p. 51-60.

89.   Zagari, A., Nemethy, G., and Scheraga, H.A., *The effect of the L-azetidine-2-carboxylic acid residue on protein conformation. I. Conformations of the residue and of dipeptides.* Biopolymers, 1990. **30**(9-10): p. 951-9.

90.   Kao, W.W. and Prockop, D.J., *Proline analogue removes fibroblasts from cultured mixed cell populations.* Nature, 1977. **266**(5597): p. 63-4.

91.   Uitto, J. and Prockop, D.J., Biochim. biophys. Acta, 1974. **336**: p. 234-251.

92.   Uitto, J., Dehm, P., and Prockop, D.J., Biochim. biophys. Acta, 1972. **278**: p. 601-605.

93.   Takeuchi, T. and Prockop, D.J., Biochim. biophys. Acta, 1969. **175**: p. 142-155.

94.   Jimenez, S. and Rosenbloom, J., Arch. Biochem. Biophys., 1974. **163**: p. 459-465.

95.   Inouye, K., Sakakibara, S., and Prockop, D.J., Biochim. biophys. Acta, 1976. **420**: p. 133-141.

96.   Rosenbloom, J. and Prockop, D.J., J. biol. Chem., 1971. **246**: p. 1549-1555.

97.   Rosenbloom, J. and Prockop, D.J., J. biol. Chem., 1970. **245**: p. 3361-3368.

98.   Rosenthal, G.A., *L-Canavanine: a higher plant insecticidal allelochemical.* Amino Acids, 2001. **21**(3): p. 319-30.

99.   Rosenthal, G.A. and Dahlman, D.L., *L-Canavanine and protein synthesis in the tobacco hornworm Manduca sexta.* Proc Natl Acad Sci U S A, 1986. **83**(1): p. 14-8.

100.   Thomas, D.A. and Rosenthal, G.A., *Toxicity and pharmacokinetics of the nonprotein amino acid L-canavanine in the rat.* Toxicol Appl Pharmacol, 1987. **91**(3): p. 395-405.

101.   Konovalova, S., et al., *Exposure to arginine analog canavanine induces aberrant mitochondrial translation products, mitoribosome stalling, and instability of the mitochondrial proteome.* Int J Biochem Cell Biol, 2015. **65**: p. 268-74.

102.   Fernandez-Silva, P., et al., *In vivo and in organello analyses of mitochondrial translation.* Methods Cell Biol, 2007. **80**: p. 571-88.

103.   Dunlop, R.A., et al., *The non-protein amino acid BMAA is misincorporated into human proteins in place of L-serine causing protein misfolding and aggregation.* PLoS One, 2013. **8**(9): p. e75376.

104.   Chernoff, N., et al., *A critical review of the postulated role of the non-essential amino acid, beta-N-methylamino-L-alanine, in neurodegenerative disease in humans.* J Toxicol Environ Health B Crit Rev, 2017. **20**(4): p. 1-47.

105.   Main, B.J. and Rodgers, K.J., *Assessing the Combined Toxicity of BMAA and Its Isomers 2,4-DAB and AEG In Vitro Using Human Neuroblastoma Cells.* Neurotox Res, 2018. **33**(1): p. 33-42.

106.   Dunlop, R.A., Brunk, U.T., and Rodgers, K.J., *Oxidized proteins: mechanisms of removal and consequences of accumulation.* IUBMB Life, 2009. **61**(5): p. 522-7.

107.   Dunlop, R.A., Dean, R.T., and Rodgers, K.J., *The impact of specific oxidized amino acids on protein turnover in J774 cells.* Biochem J, 2008. **410**(1): p. 131-40.

108.   Rodgers, K.J., et al., *Biosynthesis and turnover of DOPA-containing proteins by human cells.* Free Radic Biol Med, 2004. **37**(11): p. 1756-64.

109.   Shimizu, Y., et al., *Cell-free translation reconstituted with purified components.* Nat Biotechnol, 2001. **19**(8): p. 751-5.

110.   Glover, W.B., Mash, D.C., and Murch, S.J., *The natural non-protein amino acid N-beta-methylamino-L-alanine (BMAA) is incorporated into protein during synthesis.* Amino Acids, 2014. **46**(11): p. 2553-9.

111.   Munoz-Saez, E., et al., *Analysis of beta-N-methylamino-L-alanine (L-BMAA) neurotoxicity in rat cerebellum.* Neurotoxicology, 2015. **48**: p. 192-205.

112.   Jiang, L., et al., *Selective LC-MS/MS method for the identification of BMAA from its isomers in biological samples.* Anal Bioanal Chem, 2012. **403**(6): p. 1719-30.

113.   Jiang, L., et al., *Quantification of neurotoxin BMAA (beta-N-methylamino-L-alanine) in seafood from Swedish markets.* Sci Rep, 2014. **4**: p. 6931.

114.   Cox, P.A., et al., *Cyanobacteria and BMAA exposure from desert dust: a possible link to sporadic ALS among Gulf War veterans.* Amyotroph Lateral Scler, 2009. **10 Suppl 2**: p. 109-17.

115.   Violi, J.P., et al., *Prevalence of beta-methylamino-L-alanine (BMAA) and its isomers in freshwater cyanobacteria isolated from eastern Australia.* Ecotoxicol Environ Saf, 2019. **172**: p. 72-81.

116.   Esterhuizen-Londt, M. and Pflugmacher, S., *Vegetables cultivated with exposure to pure and naturally occurring beta-N-methylamino-L-alanine (BMAA) via irrigation.* Environ Res, 2019. **169**: p. 357-361.

117.   Main, B.J., et al., *Detection of the suspected neurotoxin beta-methylamino-l-alanine (BMAA) in cyanobacterial blooms from multiple water bodies in Eastern Australia.* Harmful Algae, 2018. **74**: p. 10-18.

118. Lage, S., et al., *BMAA extraction of cyanobacteria samples: which method to choose?* Environ Sci Pollut Res Int, 2016. **23**(1): p. 338-50.

119. Lage, S., et al., *BMAA in shellfish from two Portuguese transitional water bodies suggests the marine dinoflagellate Gymnodinium catenatum as a potential BMAA source.* Aquat Toxicol, 2014. **152**: p. 131-8.

120. Al-Sammak, M.A., et al., *Co-occurrence of the cyanotoxins BMAA, DABA and anatoxin-a in Nebraska reservoirs, fish, and aquatic plants.* Toxins (Basel), 2014. **6**(2): p. 488-508.

121. Kartvelishvili, E., et al., *Chimeric human mitochondrial PheRS exhibits editing activity to discriminate nonprotein amino acids.* Protein Sci, 2016. **25**(3): p. 618-26.

122. Chan, S., et al., *L-DOPA is incorporated into brain proteins of patients treated for Parkinson's disease, inducing toxicity in human neuroblastoma cells in vitro*. Vol. 238. 2011. 29-37.

123. Thompson, A.M., et al., *Evidence that DOPA-Derivatives are Generated After L-DOPA Incorporation into Proteins by Mammalian Cells.* The Journal of Adhesion, 2009. **85**(9): p. 561-575.

124. Rodgers, K.J., et al., *Evidence for L-dopa incorporation into cell proteins in patients treated with levodopa.* J Neurochem, 2006. **98**(4): p. 1061-7.

125. Asanuma, M., Miyazaki, I., and Ogawa, N., *Dopamine- or L-DOPA-induced neurotoxicity: the role of dopamine quinone formation and tyrosinase in a model of Parkinson's disease.* Neurotox Res, 2003. **5**(3): p. 165-76.

126. Alexander, T., et al., *Comparison of neurotoxicity following repeated administration of l-dopa, d-dopa and dopamine to embryonic mesencephalic dopamine neurons in cultures derived from Fisher 344 and Sprague-Dawley donors.* Cell Transplant, 1997. **6**(3): p. 309-15.

127. Spencer, J.P., et al., *Evaluation of the pro-oxidant and antioxidant actions of L-DOPA and dopamine in vitro: implications for Parkinson's disease.* Free Radic Res, 1996. **24**(2): p. 95-105.

128. Li, C.L., Werner, P., and Cohen, G., *Lipid peroxidation in brain: interactions of L-DOPA/dopamine with ascorbate and iron.* Neurodegeneration, 1995. **4**(2): p. 147-53.

129. Pardo, B., et al., *Toxic effects of L-DOPA on mesencephalic cell cultures: protection with antioxidants.* Brain Res, 1995. **682**(1-2): p. 133-43.

130. Walkinshaw, G. and Waters, C.M., *Induction of apoptosis in catecholaminergic PC12 cells by L-DOPA. Implications for the treatment of Parkinson's disease.* J Clin Invest, 1995. **95**(6): p. 2458-64.

131. Basma, A.N., et al., *L-DOPA CYTOTOXICITY TO PC12 CELLS IN CULTURE IS VIA ITS AUTOXIDATION.* Journal of Neurochemistry, 1995. **64**(2): p. 825-832.

132. Asanuma, M. and Miyazaki, I., *3-O-Methyldopa inhibits astrocyte-mediated dopaminergic neuroprotective effects of L-DOPA.* BMC Neurosci, 2016. **17**(1): p. 52.

133. Perveen, A., et al., *Pro-oxidant DNA Breakage Induced by the Interaction of L-DOPA with Cu(II): A Putative Mechanism of Neurotoxicity*, in *GeNeDis 2014*, P. Vlamos and A. Alexiou, Editors. 2015, Springer Int Publishing Ag: Cham. p. 37-51.

134. Colamartino, M., et al., *Evaluation of levodopa and carbidopa antioxidant activity in normal human lymphocytes in vitro: implication for oxidative stress in Parkinson's disease.* Neurotox Res, 2015. **27**(2): p. 106-17.

135. Bizzarri, B.M., et al., *Current Advances in L-DOPA and DOPA-Peptidomimetics: Chemistry, Applications and Biological Activity.* Curr Med Chem, 2015. **22**(36): p. 4138-65.

136. Jami, M.S., et al., *Proteome analysis reveals roles of L-DOPA in response to oxidative stress in neurons.* BMC Neurosci, 2014. **15**: p. 93.

137. Mushtaq, M.N., Sunohara, Y., and Matsumoto, H., *l-DOPA inhibited the root growth of lettuce by inducing reactive oxygen species generation.* Weed Biology and Management, 2013. **13**(4): p. 129-134.

138. Park, K.H., et al., *L-DOPA neurotoxicity is prevented by neuroprotective effects of erythropoietin.* Neurotoxicology, 2011. **32**(6): p. 879-87.

139. Koh, S.H., et al., *Protective effects of statins on L-DOPA neurotoxicity due to the activation of phosphatidylinositol 3-kinase and free radical scavenging in PC12 cell culture.* Brain Res, 2011. **1370**: p. 53-63.

140. Koh, S.H., Kim, S.H., and Kim, H.T., *Role of glycogen synthase kinase-3 in l-DOPA-induced neurotoxicity.* Expert Opin Drug Metab Toxicol, 2009. **5**(11): p. 1359-68.

141. Aluf, Y., Vaya, J., and Finberg, J.P.M., *Characterization of oxidative stress in amphetamine-and L-dopa-treated rats using novel multifunctional marker molecules.* Journal of Neural Transmission, 2008. **115**(10): p. 1478-1478.

142. Hachinohe, M. and Matsumoto, H., *Involvement of melanin synthesis and reactive oxygen species in phytotoxic action of l-DOPA in carrot cells.* Crop Protection, 2007. **26**(3): p. 294-298.

143. Rocchitta, G., et al., *Endogenous melatonin protects L-DOPA from autoxidation in the striatal extracellular compartment of the freely moving rat: potential implication for long-term L-DOPA therapy in Parkinson's disease.* J Pineal Res, 2006. **40**(3): p. 204-13.

144. Isobe, C., et al., *Cabergoline scavenges peroxynitrite enhanced by L-DOPA therapy in patients with Parkinson's disease.* Eur J Neurol, 2006. **13**(4): p. 346-50.

145. Hachinohe, M. and Matsumoto, H., *Involvement of reactive oxygen species generated from melanin synthesis pathway in phytotoxicty of L-DOPA.* J Chem Ecol, 2005. **31**(2): p. 237-46.

146. Shi, Y.L., Benzie, I.F., and Buswell, J.A., *L-DOPA oxidation products prevent H2O2-induced oxidative damage to cellular DNA.* Life Sci, 2002. **71**(26): p. 3047-57.

147. Kostrzewa, R.M., Kostrzewa, J.P., and Brus, R., *Neuroprotective and neurotoxic roles of levodopa (L-DOPA) in neurodegenerative disorders relating to Parkinson's disease.* Amino Acids, 2002. **23**(1-3): p. 57-63.

148. Koshimura, K., et al., *Effects of dopamine and L-DOPA on survival of PC12 cells.* J Neurosci Res, 2000. **62**(1): p. 112-9.

149. Migheli, R., et al., *Enhancing effect of manganese on L-DOPA-induced apoptosis in PC12 cells: role of oxidative stress.* J Neurochem, 1999. **73**(3): p. 1155-63.

150. Kim-Han, J.S. and Sun, A.Y., *Protection of PC12 cells glutathione peroxidase in L-DOPA induced cytotoxicity.* Free Radic Biol Med, 1998. **25**(4-5): p. 512-8.

151. Nakao, N., Nakai, K., and Itakura, T., *Metabolic inhibition enhances selective toxicity of L-DOPA toward mesencephalic dopamine neurons in vitro.* Brain Res, 1997. **777**(1-2): p. 202-9.

152. Mena, M.A., Davila, V., and Sulzer, D., *Neurotrophic effects of L-DOPA in postnatal midbrain dopamine neuron/cortical astrocyte cocultures.* J Neurochem, 1997. **69**(4): p. 1398-408.

153. Saiz Garcia, H., et al., *Nootropics: Emergents drugs associated with new clinical challenges.* European Psychiatry, 2020. **41**(S1): p. s877-s878.

154. Tomen, D. *L-DOPA*. 2016 [cited 2017 28/10/17]; Available from: https://nootropicsexpert.com/l-dopa/.

155. *L-DOPA*. 28/10/2017]; Available from: http://www.smarternootropics.com/table-of-contents/l-dopa/.

156. Badger, J.L., et al., *Parkinson's disease in a dish - Using stem cells as a molecular tool.* Neuropharmacology, 2014. **76 Pt A**: p. 88-96.

157. First, E.A., *L-DOPA ropes in tRNA(Phe).* Chem Biol, 2011. **18**(10): p. 1201-2.

158. Moor, N., Klipcan, L., and Safro, M.G., *Bacterial and eukaryotic phenylalanyl-tRNA synthetases catalyze misaminoacylation of tRNA(Phe) with 3,4-dihydroxy-L-phenylalanine.* Chem Biol, 2011. **18**(10): p. 1221-9.

159. Gundersen, V., *Protein aggregation in Parkinson's disease.* Acta Neurol Scand Suppl, 2010(190): p. 82-7.

160. Ozawa, K., et al., *Translational incorporation of L-3,4-dihydroxyphenylalanine into proteins.* FEBS J, 2005. **272**(12): p. 3162-71.

161. Muller, T., et al., *Is levodopa toxic?* J Neurol, 2004. **251 Suppl 6**(S6): p. VI/44-6.

162. Simon, N., et al., *The effects of a normal protein diet on levodopa plasma kinetics in advanced Parkinson's disease.* Parkinsonism Relat Disord, 2004. **10**(3): p. 137-42.

163.   Mytilineou, C., et al., *Levodopa is toxic to dopamine neurons in an in vitro but not an in vivo model of oxidative stress.* J Pharmacol Exp Ther, 2003. **304**(2): p. 792-800.

164.   Rodgers, K.J., et al., *Biosynthetic incorporation of oxidized amino acids into proteins and their cellular proteolysis.* Free Radical Biology and Medicine, 2002. **32**(8): p. 766-775.

165.   Melamed, E., et al., *Levodopa--an exotoxin or a therapeutic drug?* J Neurol, 2000. **247 Suppl 2**: p. II135-9.

166.   Dizdar, N., et al., *L-dopa pharmacokinetics studied with microdialysis in patients with Parkinson's disease and a history of malignant melanoma.* Acta Neurol Scand, 1999. **100**(4): p. 231-7.

167.   Riley, P.A., *The great DOPA mystery: the source and significance of DOPA in phase I melanogenesis.* Cell Mol Biol (Noisy-le-grand), 1999. **45**(7): p. 951-60.

168.   Calendar, R. and Berg, P., *The catalytic properties of tyrosyl ribonucleic acid synthetases from Escherichia coli and Bacillus subtilis.* Biochemistry, 1966. **5**(5): p. 1690-5.

169.   Hogenauer, G., Kreil, G., and Bernheimer, H., *Studies on the binding of DOPA (3,4-dihydroxyphenylalanine) to tRNA.* FEBS Lett, 1978. **88**(1): p. 101-4.

170.   Rodgers, K.J. and Dean, R.T., *Metabolism of protein-bound DOPA in mammals.* Int J Biochem Cell Biol, 2000. **32**(9): p. 945-55.

171.   Shiozawa-West, N., Dunlop, R.A., and Rodgers, K.J., *Using an in vitro model to study oxidised protein accumulation in ageing fibroblasts.* Biochim Biophys Acta, 2015. **1850**(11): p. 2177-84.

172.   Giannopoulos, S., et al., *L-DOPA causes mitochondrial dysfunction in vitro: A novel mechanism of L-DOPA toxicity uncovered.* Int J Biochem Cell Biol, 2019. **117**: p. 105624.

173.   Song, J., et al., *Levodopa (L-DOPA) attenuates endoplasmic reticulum stress response and cell death signaling through DRD2 in SH-SY5Y neuronal cells under alpha-synuclein-induced toxicity.* Neuroscience, 2017. **358**: p. 336-348.

174.   Szadejko, K., et al., *Polyneuropathy in levodopa-treated Parkinson's patients.* J Neurol Sci, 2016. **371**: p. 36-41.

175.   Nikolova, G., et al., *Protective effect of two essential oils isolated from Rosa damascena Mill. and Lavandula angustifolia Mill, and two classic antioxidants against L-dopa oxidative toxicity induced in healthy mice.* Regul Toxicol Pharmacol, 2016. **81**: p. 1-7.

176.   Stednitz, S.J., et al., *Selective toxicity of L-DOPA to dopamine transporter-expressing neurons and locomotor behavior in zebrafish larvae.* Neurotoxicol Teratol, 2015. **52**(Pt A): p. 51-6.

177.   Olanow, C.W., *Levodopa: effect on cell death and the natural history of Parkinson's disease.* Mov Disord, 2015. **30**(1): p. 37-44.

178.   Jang, W., et al., *1,25-dyhydroxyvitamin D3 attenuates L-DOPA-induced neurotoxicity in neural stem cells.* Mol Neurobiol, 2015. **51**(2): p. 558-70.

179.   Alirezaei, M., et al., *Beneficial antioxidant properties of betaine against oxidative stress mediated by levodopa/benserazide in the brain of rats.* J Physiol Sci, 2015. **65**(3): p. 243-52.

180.   Stansley, B.J. and Yamamoto, B.K., *L-dopa-induced dopamine synthesis and oxidative stress in serotonergic cells.* Neuropharmacology, 2013. **67**: p. 243-51.

181.   Yuan, H., et al., *Treatment strategies for Parkinson's disease.* Neurosci Bull, 2010. **26**(1): p. 66-76.

182.   Lee, E.S., et al., *The role of 3-O-methyldopa in the side effects of L-dopa.* Neurochem Res, 2008. **33**(3): p. 401-11.

183.   Chen, L., et al., *Unregulated cytosolic dopamine causes neurodegeneration associated with oxidative stress in mice.* J Neurosci, 2008. **28**(2): p. 425-33.

184.   Cox, P.A., Banack, S.A., and Murch, S.J., *Biomagnification of cyanobacterial neurotoxins and neurodegenerative disease among the Chamorro people of Guam.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13380-3.

185.   Karamyan, V.T. and Speth, R.C., *Animal models of BMAA neurotoxicity: a critical review.* Life Sci, 2008. **82**(5-6): p. 233-46.

186. Zhang, Z.X., et al., *Motor neuron disease on Guam: geographic and familial occurrence, 1956-85.* Acta Neurol Scand, 1996. **94**(1): p. 51-9.

187. Banack, S.A. and Murch, S.J., *Multiple neurotoxic items in the Chamorro diet link BMAA with ALS/PDC.* Amyotroph Lateral Scler, 2009. **10 Suppl 2**: p. 34-40.

188. Bell, E.A., *The discovery of BMAA, and examples of biomagnification and protein incorporation involving other non-protein amino acids.* Amyotroph Lateral Scler, 2009. **10 Suppl 2**: p. 21-5.

189. Moya-Alvarado, G., et al., *Neurodegeneration and Alzheimer's disease (AD). What Can Proteomics Tell Us About the Alzheimer's Brain?* Mol Cell Proteomics, 2016. **15**(2): p. 409-25.

190. Butterfield, D.A., et al., *Redox proteomics analysis to decipher the neurobiology of Alzheimer-like neurodegeneration: overlaps in Down's syndrome and Alzheimer's disease brain.* Biochem J, 2014. **463**(2): p. 177-89.

191. Garcia-Garcia, A., et al., *Biomarkers of protein oxidation in human disease.* Curr Mol Med, 2012. **12**(6): p. 681-97.

192. Sultana, R., et al., *Do proteomics analyses provide insights into reduced oxidative stress in the brain of an Alzheimer disease transgenic mouse model with an M631L amyloid precursor protein substitution and thereby the importance of amyloid-beta-resident methionine 35 in Alzheimer disease pathogenesis?* Antioxid Redox Signal, 2012. **17**(11): p. 1507-14.

193. Chan, S., *Studies investigating Levodopa incorporation into proteins: Implications for its use in Parkonson's Disease*, in *Medicine Faculty*. 2011, University of Sydney: N/A. p. 224.

194. Mytilineou, C., Han, S.K., and Cohen, G., *Toxic and protective effects of L-dopa on mesencephalic cell cultures.* J Neurochem, 1993. **61**(4): p. 1470-8.

195. Spencer, J.P., et al., *Intense oxidative DNA damage promoted by L-dopa and its metabolites. Implications for neurodegenerative disease.* FEBS Lett, 1994. **353**(3): p. 246-50.

196. Pattison, D.I., Dean, R.T., and Davies, M.J., *Oxidation of DNA, proteins and lipids by DOPA, protein-bound DOPA, and related catechol(amine)s.* Toxicology, 2002. **177**(1): p. 23-37.

197. Pedrosa, R. and Soares-da-Silva, P., *Oxidative and non-oxidative mechanisms of neuronal cell death and apoptosis by L-3,4-dihydroxyphenylalanine (L-DOPA) and dopamine.* Br J Pharmacol, 2002. **137**(8): p. 1305-13.

198. Clement, M.V., et al., *The cytotoxicity of dopamine may be an artefact of cell culture.* J Neurochem, 2002. **81**(3): p. 414-21.

199. Cienska, M., et al., *Effective L-Tyrosine Hydroxylation by Native and Immobilized Tyrosinase.* PLoS One, 2016. **11**(10): p. e0164213.

200. Bufalino, M.R. and van der Kooy, D., *The aggregation and inheritance of damaged proteins determines cell fate during mitosis.* Cell Cycle, 2014. **13**(7): p. 1201-7.

201. Dunlop, R.A., Rodgers, K.J., and Dean, R.T., *Recent developments in the intracellular degradation of oxidized proteins.* Free Radic Biol Med, 2002. **33**(7): p. 894-906.

202. Bullwinkle, T., Lazazzera, B., and Ibba, M., *Quality Control and Infiltration of Translation by Amino Acids Outside of the Genetic Code*, in *Annual Review of Genetics, Vol 48*, B.L. Bassler, Editor. 2014. p. 149-166.

203. Castelli, V., et al., *Neuronal Cells Rearrangement During Aging and Neurodegenerative Disease: Metabolism, Oxidative Stress and Organelles Dynamic.* Front Mol Neurosci, 2019. **12**(132): p. 132.

204. Rehmani, N., et al., *Copper-mediated DNA damage by the neurotransmitter dopamine and L-DOPA: A pro-oxidant mechanism.* Toxicol In Vitro, 2017. **40**: p. 336-346.

205. Lohr, K.M. and Miller, G.W., *VMAT2 and Parkinson's disease: harnessing the dopamine vesicle.* Expert Rev Neurother, 2014. **14**(10): p. 1115-7.

206. Chagraoui, A., et al., *L-DOPA in Parkinson's Disease: Looking at the "False" Neurotransmitters and Their Meaning.* 2020. **21**(1): p. 294.

207. Ptolemy, A.S., Lee, R., and Britz-McKibbin, P., *Strategies for comprehensive analysis of amino acid biomarkers of oxidative stress.* Amino Acids, 2007. **33**(1): p. 3-18.

208. Steele, J.R., et al., *A Novel Method for Creating a Synthetic L-DOPA Proteome and In Vitro Evidence of Incorporation.* Proteomes, 2021. **9**(2): p. 24.

209. Butterfield, D.A., et al., *Oxidative Stress and the Triangle of Death in Alzheimer Disease Brain: The Aberrant Crosstalk among Energy Metabolism, MTOR Signaling and Protein Homeostasis Revealed by Redox Proteomics.* Free Radical Biology and Medicine, 2016. **100**(Supplement): p. S158.

210. Butterfield, D.A. and Dalle-Donne, I., *Redox proteomics: from protein modifications to cellular dysfunction and disease.* Mass Spectrom Rev, 2014. **33**(1): p. 1-6.

211. Butterfield, D.A., et al., *Mass spectrometry and redox proteomics: applications in disease.* Mass Spectrom Rev, 2014. **33**(4): p. 277-301.

212. Hipp, M.S., Kasturi, P., and Hartl, F.U., *The proteostasis network and its decline in ageing.* Nat Rev Mol Cell Biol, 2019. **20**(7): p. 421-435.

213. Carballo-Carbajal, I., et al., *Brain tyrosinase overexpression implicates age-dependent neuromelanin production in Parkinson's disease pathogenesis.* Nat Commun, 2019. **10**(1): p. 973.

214. Hetz, C. and Mollereau, B., *Disturbance of endoplasmic reticulum proteostasis in neurodegenerative diseases.* Nat Rev Neurosci, 2014. **15**(4): p. 233-49.

215. Prusiner, S.B., *Cell biology. A unifying role for prions in neurodegenerative diseases.* Science, 2012. **336**(6088): p. 1511-3.

216. Holmes, B.B. and Diamond, M.I., *Cellular mechanisms of protein aggregate propagation.* Curr Opin Neurol, 2012. **25**(6): p. 721-6.

217. Murphy, R.M., *Peptide aggregation in neurodegenerative disease.* Annu Rev Biomed Eng, 2002. **4**: p. 155-74.

218. *Bioinformatics & Evolutionary Genomics website, https://bioinformatics.psb.ugent.be/webtools/Venn/. Accessed December 2021.*

219. Kim, M.S., et al., *A draft map of the human proteome.* Nature, 2014. **509**(7502): p. 575-81.

220. Ayyadevara, S., et al., *Proteins that mediate protein aggregation and cytotoxicity distinguish Alzheimer's hippocampus from normal controls.* Aging Cell, 2016. **15**(5): p. 924-39.

221. Beausoleil, S.A., et al., *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nat Biotechnol, 2006. **24**(10): p. 1285-92.

222. Caufield, J.H., et al., *A Second Look at FAIR in Proteomic Investigations.* J Proteome Res, 2021. **20**(5): p. 2182-2186.

223. Lachen-Montes, M., et al., *Unveiling the olfactory proteostatic disarrangement in Parkinson's disease by proteome-wide profiling.* Neurobiol Aging, 2019. **73**: p. 123-134.

224. McAlister, G.C., et al., *MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes.* Anal Chem, 2014. **86**(14): p. 7150-8.

225. Omenn, G.S., et al., *Progress Identifying and Analyzing the Human Proteome: 2021 Metrics from the HUPO Human Proteome Project.* J Proteome Res, 2021. **20**(12): p. 5227-5240.

226. Eckman, J., et al., *Oxidative Stress Levels in the Brain Are Determined by Post-Mortem Interval and Ante-Mortem Vitamin C State but Not Alzheimer's Disease Status.* Nutrients, 2018. **10**(7).

227. Hoofnagle, A.N., et al., *Recommendations for the Generation, Quantification, Storage, and Handling of Peptides Used for Mass Spectrometry-Based Assays.* Clin Chem, 2016. **62**(1): p. 48-69.

228. Steele, J.R., et al., *Misincorporation Proteomics Technologies: A Review.* Proteomes, 2021. **9**(1): p. 2.

229. Bettinger, J.Q., et al., *Quantitative Analysis of in Vivo Methionine Oxidation of the Human Proteome.* J Proteome Res, 2020. **19**(2): p. 624-633.

230. Aledo, J.C., *Methionine in proteins: The Cinderella of the proteinogenic amino acids.* Protein Sci, 2019. **28**(10): p. 1785-1796.

231. Paulsen, C.E. and Carroll, K.S., *Cysteine-mediated redox signaling: chemistry, biology, and tools for discovery.* Chem Rev, 2013. **113**(7): p. 4633-79.

232. Brenig, K., et al., *The Proteomic Landscape of Cysteine Oxidation That Underpins Retinoic Acid-Induced Neuronal Differentiation.* J Proteome Res, 2020. **19**(5): p. 1923-1940.

233. Sun, G. and Anderson, V.E., *Prevention of artifactual protein oxidation generated during sodium dodecyl sulfate-gel electrophoresis.* Electrophoresis, 2004. **25**(7-8): p. 959-65.

234. Serpa, J.J., et al., *Using isotopically-coded hydrogen peroxide as a surface modification reagent for the structural characterization of prion protein aggregates.* J Proteomics, 2014. **100**: p. 160-6.

235. Pappireddi, N., Martin, L., and Wuhr, M., *A Review on Quantitative Multiplexed Proteomics.* Chembiochem, 2019. **20**(10): p. 1210-1224.

236. Snezhkina, A.V., et al., *ROS Generation and Antioxidant Defense Systems in Normal and Malignant Cells.* Oxid Med Cell Longev, 2019. **2019**: p. 6175804.

237. Ottaviano, F.G., Handy, D.E., and Loscalzo, J., *Redox regulation in the extracellular environment.* Circ J, 2008. **72**(1): p. 1-16.

238. Stolzing, A., Wengner, A., and Grune, T., *Degradation of oxidized extracellular proteins by microglia.* Arch Biochem Biophys, 2002. **400**(2): p. 171-9.

239. Rubenstein, E., et al., *Azetidine-2-carboxylic acid in garden beets (Beta vulgaris).* Phytochemistry, 2006. **67**(9): p. 898-903.

240. Rubenstein, E., *Biologic effects of and clinical disorders caused by nonprotein amino acids.* Medicine (Baltimore), 2000. **79**(2): p. 80-9.

241. Cao, Z., Bhella, D., and Lindsay, J.G., *Reconstitution of the mitochondrial PrxIII antioxidant defence pathway: general properties and factors affecting PrxIII activity and oligomeric state.* J Mol Biol, 2007. **372**(4): p. 1022-1033.

242. Landino, L.M., Hagedorn, T.D., and Kennett, K.L., *Evidence for thiol/disulfide exchange reactions between tubulin and glyceraldehyde-3-phosphate dehydrogenase.* Cytoskeleton (Hoboken), 2014. **71**(12): p. 707-18.

243. Pauwels, J. and Gevaert, K., *Mass spectrometry-based clinical proteomics - a revival.* Expert Rev Proteomics, 2021. **18**(6): p. 411-414.

244. Lisitsa, A., et al., *Profiling proteoforms: promising follow-up of proteomics for biomarker discovery.* Expert Rev Proteomics, 2014. **11**(1): p. 121-9.

245. Settembre, C., et al., *Lysosomal storage diseases as disorders of autophagy.* Autophagy, 2008. **4**(1): p. 113-4.

246. Jana, S., et al., *Mitochondrial dysfunction mediated by quinone oxidation products of dopamine: Implications in dopamine cytotoxicity and pathogenesis of Parkinson's disease.* Biochim Biophys Acta, 2011. **1812**(6): p. 663-73.

247. Sulzer, D. and Zecca, L., *Intraneuronal dopamine-quinone synthesis: a review.* Neurotox Res, 2000. **1**(3): p. 181-95.

248. Jakobson, C.M. and Jarosz, D.F., *Organizing biochemistry in space and time using prion-like self-assembly.* Curr Opin Syst Biol, 2018. **8**: p. 16-24.

249. Brundin, P. and Melki, R., *Prying into the Prion Hypothesis for Parkinson's Disease.* J Neurosci, 2017. **37**(41): p. 9808-9818.

250. Manecka, D.L., et al., *The Neuroprotective Role of Protein Quality Control in Halting the Development of Alpha-Synuclein Pathology.* Front Mol Neurosci, 2017. **10**: p. 311.

251. Sudhakaran, I.P. and Ramaswami, M., *Long-term memory consolidation: The role of RNA-binding proteins with prion-like domains.* RNA Biol, 2017. **14**(5): p. 568-586.

252. Goedert, M., *NEURODEGENERATION. Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled Abeta, tau, and alpha-synuclein.* Science, 2015. **349**(6248): p. 1255555.

253. Poehler, A.M., et al., *Autophagy modulates SNCA/alpha-synuclein release, thereby generating a hostile microenvironment.* Autophagy, 2014. **10**(12): p. 2171-92.

254. Kfoury, N., et al., *Trans-cellular propagation of Tau aggregation by fibrillar species.* J Biol Chem, 2012. **287**(23): p. 19440-51.

255. Goedert, M., Clavaguera, F., and Tolnay, M., *The propagation of prion-like protein inclusions in neurodegenerative diseases.* Trends Neurosci, 2010. **33**(7): p. 317-25.

256. Prusiner, S.B., *Prions.* Proc Natl Acad Sci U S A, 1998. **95**(23): p. 13363-83.

257. O'Connell, J.D., et al., *Proteome-Wide Evaluation of Two Common Protein Quantification Methods.* J Proteome Res, 2018. **17**(5): p. 1934-1942.

258. Tran, N.H., et al., *Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry.* Nat Methods, 2019. **16**(1): p. 63-66.

259. Tran, N.H., et al., *De novo peptide sequencing by deep learning.* Proc Natl Acad Sci U S A, 2017. **114**(31): p. 8247-8252.

260. Erickson, B.K., et al., *Parallel Notched Gas-Phase Enrichment for Improved Proteome Identification and Quantification with Fast Spectral Acquisition Rates.* J Proteome Res, 2020. **19**(7): p. 2750-2757.

261. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics.* Nat Methods, 2011. **8**(11): p. 937-40.

262. Johnson, A., Stadlmeier, M., and Wuhr, M., *TMTpro Complementary Ion Quantification Increases Plexing and Sensitivity for Accurate Multiplexed Proteomics at the MS2 Level.* J Proteome Res, 2021. **20**(6): p. 3043-3052.

263. Humphrey, S.J., et al., *High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform.* Nat Protoc, 2018. **13**(9): p. 1897-1916.

264. Heijs, B., et al., *Brain Region-Specific Dynamics of On-Tissue Protein Digestion Using MALDI Mass Spectrometry Imaging.* J Proteome Res, 2015. **14**(12): p. 5348-54.

265. Tomascova, A., et al., *A comparison of albumin removal procedures for proteomic analysis of blood plasma.* Gen Physiol Biophys, 2019. **38**(4): p. 305-314.

266. Wang, H.Q., et al., *Cell type-specific upregulation of Parkin in response to ER stress.* Antioxid Redox Signal, 2007. **9**(5): p. 533-42.

267. Iliff, J.J., et al., *A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid beta.* Sci Transl Med, 2012. **4**(147): p. 147ra111.

268. Nakajima, H., et al., *Glyceraldehyde-3-phosphate dehydrogenase aggregate formation participates in oxidative stress-induced cell death.* J Biol Chem, 2009. **284**(49): p. 34331-41.

269. Virág, D., et al., *Current Trends in the Analysis of Post-translational Modifications.* Chromatographia, 2019. **83**(1): p. 1-10.

270. Leutert, M., Entwisle, S.W., and Villen, J., *Decoding Post-Translational Modification Crosstalk With Proteomics.* Mol Cell Proteomics, 2021. **20**: p. 100129.

271. Sarre, S., et al., *Biotransformation of L-DOPA to dopamine in the substantia nigra of freely moving rats: effect of dopamine receptor agonists and antagonists.* J Neurochem, 1998. **70**(4): p. 1730-9.

272. Budnik, B., et al., *SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation.* Genome Biol, 2018. **19**(1): p. 161.

273. Tsai, C.F., et al., *Motif-centric phosphoproteomics to target kinase-mediated signaling pathways.* Cell Rep Methods, 2022. **2**(1): p. 100138.

274. Liu, F., et al., *Enhancing Data Reliability in TOMAHAQ for Large-Scale Protein Quantification.* Proteomics, 2020. **20**(11): p. e1900105.

275. Ctortecka, C., et al., *Quantitative Accuracy and Precision in Multiplexed Single-Cell Proteomics.* Anal Chem, 2022. **94**(5): p. 2434-2443.

276. Schweppe, D.K., et al., *Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics.* J Proteome Res, 2020. **19**(5): p. 2026-2034.

277. Hogrebe, A., et al., *Benchmarking common quantification strategies for large-scale phosphoproteomics.* Nat Commun, 2018. **9**(1): p. 1045.

278. Myers, S.A., et al., *Evaluation of Advanced Precursor Determination for Tandem Mass Tag (TMT)-Based Quantitative Proteomics across Instrument Platforms.* J Proteome Res, 2019. **18**(1): p. 542-547.

279. Mueller, L.N., et al., *SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling.* Proteomics, 2007. **7**(19): p. 3470-80.

280. David, L., Kang, J., and Chen, S., *Untargeted Metabolomics of Arabidopsis Stomatal Immunity.* Methods Mol Biol, 2021. **2200**: p. 413-424.

281. Kang, J., et al., *Three-in-One Simultaneous Extraction of Proteins, Metabolites and Lipids for Multi-Omics.* Front Genet, 2021. **12**: p. 635971.

282. Hindle, J.V., *Ageing, neurodegeneration and Parkinson's disease.* Age Ageing, 2010. **39**(2): p. 156-61.

283. Eckman, J., et al., *Oxidative Stress Levels in the Brain Are Determined by Post-Mortem Interval and Ante-Mortem Vitamin C State but Not Alzheimer's Disease Status.* Nutrients, 2018. **10**(7): p. 883.

284. Butterfield, D.A., *Proteomics: a new approach to investigate oxidative stress in Alzheimer's disease brain.* Brain Res, 2004. **1000**(1-2): p. 1-7.

285. Butterfield, D.A., Boyd-Kimball, D., and Castegna, A., *Proteomics in Alzheimer's disease: insights into potential mechanisms of neurodegeneration.* J Neurochem, 2003. **86**(6): p. 1313-27.

286. Yu, F., et al., *Identification of modified peptides using localization-aware open search.* Nat Commun, 2020. **11**(1): p. 4065.

287. Chang, H.Y., et al., *Crystal-C: A Computational Tool for Refinement of Open Search Results.* J Proteome Res, 2020. **19**(6): p. 2511-2515.

288. Kong, A.T., et al., *MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics.* Nat Methods, 2017. **14**(5): p. 513-520.

289. da Veiga Leprevost, F., et al., *Philosopher: a versatile toolkit for shotgun proteomics data analysis.* Nat Methods, 2020. **17**(9): p. 869-870.

290. Avtonomov, D.M., Kong, A., and Nesvizhskii, A.I., *DeltaMass: Automated Detection and Visualization of Mass Shifts in Proteomic Open-Search Results.* J Proteome Res, 2019. **18**(2): p. 715-720.

291. Tyanova, S., Temu, T., and Cox, J., *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.* Nat Protoc, 2016. **11**(12): p. 2301-2319.

292. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.

293. Xue, R., et al., *Peripheral Dopamine Controlled by Gut Microbes Inhibits Invariant Natural Killer T Cell-Mediated Hepatitis.* Front Immunol, 2018. **9**: p. 2398.

294. Turner, M.R., et al., *Pattern of spread and prognosis in lower limb-onset ALS.* Amyotroph Lateral Scler, 2010. **11**(4): p. 369-73.

295. Byrne, S., et al., *Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis.* J Neurol Neurosurg Psychiatry, 2011. **82**(6): p. 623-7.

296. Al-Chalabi, A., et al., *Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study.* Lancet Neurol, 2014. **13**(11): p. 1108-1113.

297. Chio, A., et al., *The multistep hypothesis of ALS revisited: The role of genetic mutations.* Neurology, 2018. **91**(7): p. e635-e642.

298. Liu, X., et al., *β-N-methylamino-L-alanine induces oxidative stress and glutamate release through action on system Xc–.* Experimental neurology, 2009. **217**(2): p. 429-433.

299. Murch, S.J., et al., *Occurrence of beta-methylamino-l-alanine (BMAA) in ALS/PDC patients from Guam.* Acta Neurol Scand, 2004. **110**(4): p. 267-9.

300. Murch, S.J., Cox, P.A., and Banack, S.A., *A mechanism for slow release of biomagnified cyanobacterial neurotoxins and neurodegenerative disease in Guam.* Proc Natl Acad Sci U S A, 2004. **101**(33): p. 12228-31.

301. Vega, A., *α-Amino-β-methylaminopropionic acid, a new amino acid from seeds of Cycas circinalis.* Phytochemistry, 1967. **6**(5): p. 759-762.

302. Bradley, W.G., *Possible therapy for ALS based on the cyanobacteria/BMAA hypothesis.* Amyotroph Lateral Scler, 2009. **10 Suppl 2**: p. 118-23.

303. Brand, L.E., *Human exposure to cyanobacteria and BMAA.* Amyotroph Lateral Scler, 2009. **10 Suppl 2**: p. 85-95.

304. Cox, P.A., *Conclusion to the Symposium: the seven pillars of the cyanobacteria/BMAA hypothesis.* Amyotroph Lateral Scler, 2009. **10 Suppl 2**: p. 124-6.

305. Karlsson, O., et al., *Selective brain uptake and behavioral effects of the cyanobacterial toxin BMAA (beta-N-methylamino-L-alanine) following neonatal administration to rodents.* Toxicol Sci, 2009. **109**(2): p. 286-95.

306. Delcourt, N., et al., *Cellular and Molecular Aspects of the beta-N-Methylamino-l-alanine (BMAA) Mode of Action within the Neurodegenerative Pathway: Facts and Controversy.* Toxins (Basel), 2017. **10**(1): p. 6.

307. Popova, A.A. and Koksharova, O.A., *Neurotoxic Non-proteinogenic Amino Acid beta-N-Methylamino-L-alanine and Its Role in Biological Systems.* Biochemistry (Mosc), 2016. **81**(8): p. 794-805.

308. Glover, W.B., Cohen, S.A., and Murch, S.J., *Liquid chromatography and mass spectrometry for the analysis of N-beta-methylamino-L-alanine with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate.* Methods Mol Biol, 2015. **1208**: p. 379-91.

309. Merel, S., et al., *State of knowledge and concerns on cyanobacterial blooms and cyanotoxins.* Environ Int, 2013. **59**: p. 303-27.

310. Masseret, E., et al., *Dietary BMAA exposure in an amyotrophic lateral sclerosis cluster from southern France.* PLoS One, 2013. **8**(12): p. e83406.

311. Nunn, P.B., *50 years of research on alpha-amino-beta-methylaminopropionic acid (beta-methylaminoalanine).* Phytochemistry, 2017. **144**: p. 271-281.

312. Han, N.C., et al., *The mechanism of beta-N-methylamino-l-alanine inhibition of tRNA aminoacylation and its impact on misincorporation.* J Biol Chem, 2020. **295**(5): p. 1402-1410.

313. Olek, M.J., *Multiple Sclerosis.* Ann Intern Med, 2021. **174**(6): p. ITC81-ITC96.

314. Song, Y., et al., *Double mimicry evades tRNA synthetase editing by toxic vegetable-sourced non-proteinogenic amino acid.* Nat Commun, 2017. **8**(1): p. 2281.

315. Lu, X. and Zhu, H., *Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins.* Mol Cell Proteomics, 2005. **4**(12): p. 1948-58.

316. Shevchenko, G., et al., *Comparison of extraction methods for the comprehensive analysis of mouse brain proteome using shotgun-based mass spectrometry.* 2012. **11**(4): p. 2441-2451.

317. Vincent, S.G., et al., *Quantitative densitometry of proteins stained with coomassie blue using a Hewlett Packard scanjet scanner and Scanplot software.* Electrophoresis, 1997. **18**(1): p. 67-71.

318. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.

319. Gauci, V.J., Padula, M.P., and Coorssen, J.R., *Coomassie blue staining for high sensitivity gel-based proteomics.* J Proteomics, 2013. **90**: p. 96-106.

320. Mohseni, S., et al., *Antioxidant properties of a human neuropeptide and its protective effect on free radical-induced DNA damage.* J Pept Sci, 2014. **20**(6): p. 429-37.

321. Brown, J.L., et al., *Tumor burden negatively impacts protein turnover as a proteostatic process in noncancerous liver, heart, and muscle, but not brain.* J Appl Physiol (1985), 2021. **131**(1): p. 72-82.

322. Taoka, M., et al., *An Ionic Liquid-Based Sample Preparation Method for Next-Stage Aggregate Proteomic Analysis.* Anal Chem, 2019. **91**(21): p. 13494-13500.

323. Zhang, L., et al., *Cross-talk between PRMT1-mediated methylation and ubiquitylation on RBM15 controls RNA splicing.* Elife, 2015. **4**: p. e07938.

324. Okamoto, S., et al., *beta-N-methylamino-L-alanine (BMAA) suppresses cell cycle progression of non-neuronal cells.* Sci Rep, 2018. **8**(1): p. 17995.

325. Lee, Y.J., et al., *Downregulation of PRMT1 promotes the senescence and migration of a non-MYCN amplified neuroblastoma SK-N-SH cells.* Sci Rep, 2019. **9**(1): p. 1771.

326. Cooper, C.E., et al., *Engineering tyrosine residues into hemoglobin enhances heme reduction, decreases oxidative stress and increases vascular retention of a hemoglobin based blood substitute.* Free Radic Biol Med, 2019. **134**: p. 106-118.

327. Kapralov, A., et al., *Peroxidase activity of hemoglobin-haptoglobin complexes: covalent aggregation and oxidative stress in plasma and macrophages.* J Biol Chem, 2009. **284**(44): p. 30395-407.

328. Suzuki, S., et al., *Hemoglobin augmentation of interleukin-1 beta-induced production of nitric oxide in smooth-muscle cells.* J Neurosurg, 1994. **81**(6): p. 895-901.

329. Lurling, M., Faassen, E.J., and Van Eenennaam, J.S., *Effects of the cyanobacterial neurotoxin -N-methylamino-L-alanine (BMAA) on the survival, mobility and reproduction of Daphnia magna.* Journal of Plankton Research, 2010. **33**(2): p. 333-342.

330. Kritis, A.A., et al., *Researching glutamate - induced cytotoxicity in different cell lines: a comparative/collective analysis/study.* Front Cell Neurosci, 2015. **9**(91): p. 91.

331. Baillet, A., et al., *The role of oxidative stress in amyotrophic lateral sclerosis and Parkinson's disease.* Neurochem Res, 2010. **35**(10): p. 1530-7.

332. Mocchegiani, E., et al., *Role of zinc and alpha2 macroglobulin on thymic endocrine activity and on peripheral immune efficiency (natural killer activity and interleukin 2) in cervical carcinoma.* Br J Cancer, 1999. **79**(2): p. 244-50.

333. Tsuiji, H., et al., *Spliceosome integrity is defective in the motor neuron diseases ALS and SMA.* EMBO Mol Med, 2013. **5**(2): p. 221-34.

334. Matera, A.G. and Wang, Z., *A day in the life of the spliceosome.* Nat Rev Mol Cell Biol, 2014. **15**(2): p. 108-21.

335. Hsieh, Y.C., et al., *Tau-Mediated Disruption of the Spliceosome Triggers Cryptic RNA Splicing and Neurodegeneration in Alzheimer's Disease.* Cell Rep, 2019. **29**(2): p. 301-316 e10.

336. Ito, D., Hatano, M., and Suzuki, N., *RNA binding proteins and the pathological cascade in ALS/FTD neurodegeneration.* Sci Transl Med, 2017. **9**(415).

337. Dong, N.-p., Zhang, L.-x., and Liang, Y.-z., *A comprehensive investigation of proline fragmentation behavior in low-energy collision-induced dissociation peptide mass spectra.* International Journal of Mass Spectrometry, 2011. **308**(1): p. 89-97.

338. Larsen, M.R., et al., *Analysis of posttranslational modifications of proteins by tandem mass spectrometry: Mass Spectrometry For Proteomics Analysis.* 2006. **40**(6): p. 790-798.

339. Bateman, R.H., et al., *A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation.* J Am Soc Mass Spectrom, 2002. **13**(7): p. 792-803.

340. Chalmers, G.A., *Swayback (enzootic ataxia) in Alberta lambs.* Can J Comp Med, 1974. **38**(2): p. 111-7.

341. Purcell, A.W., Ramarathinam, S.H., and Ternette, N., *Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics.* Nat Protoc, 2019. **14**(6): p. 1687-1707.

342. Zhou, X., et al., *The function and clinical application of extracellular vesicles in innate immune regulation.* Cell Mol Immunol, 2020. **17**(4): p. 323-334.

343. Galazka, G., et al., *Multiple sclerosis: Serum-derived exosomes express myelin proteins.* Mult Scler, 2018. **24**(4): p. 449-458.

344. Li, D., et al., *Neurodegenerative diseases: a hotbed for splicing defects and the potential therapies.* Transl Neurodegener, 2021. **10**(1): p. 16.

345. Shental-Bechor, D. and Levy, Y., *Effect of glycosylation on protein folding: a close look at thermodynamic stabilization.* Proc Natl Acad Sci U S A, 2008. **105**(24): p. 8256-61.

346.   Xu, C. and Ng, D.T., *Glycosylation-directed quality control of protein folding.* Nat Rev Mol Cell Biol, 2015. **16**(12): p. 742-52.

347.   Laurence, J.S. and Middaugh, C.R., *Fundamental Structures and Behaviors of Proteins*, in *Aggregation of Therapeutic Proteins*. 2010. p. 1-61.

348.   Mills, B.J. and Laurence Chadwick, J.S., *Effects of localized interactions and surface properties on stability of protein-based therapeutics.* J Pharm Pharmacol, 2018. **70**(5): p. 609-624.

349.   Finelli, M.J., *Redox Post-translational Modifications of Protein Thiols in Brain Aging and Neurodegenerative Conditions-Focus on S-Nitrosation.* Front Aging Neurosci, 2020. **12**(254): p. 254.

350.   Hark, T.J., et al., *Pulse-Chase Proteomics of the App Knockin Mouse Models of Alzheimer's Disease Reveals that Synaptic Dysfunction Originates in Presynaptic Terminals.* Cell Syst, 2021. **12**(2): p. 141-158 e9.

351.   Kristensen, L.P., et al., *Temporal profiling and pulsed SILAC labeling identify novel secreted proteins during ex vivo osteoblast differentiation of human stromal stem cells.* Mol Cell Proteomics, 2012. **11**(10): p. 989-1007.

352.   Lualdi, M., et al., *Exploring the Mitochondrial Degradome by the TAILS Proteomics Approach in a Cellular Model of Parkinson's Disease.* Front Aging Neurosci, 2019. **11**: p. 195.

353.   Berry, I.J., et al., *The application of terminomics for the identification of protein start sites and proteoforms in bacteria.* Proteomics, 2016. **16**(2): p. 257-72.

354.   Prudova, A., et al., *Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics.* Mol Cell Proteomics, 2010. **9**(5): p. 894-911.

355.   Lange, P.F. and Overall, C.M., *Protein TAILS: when termini tell tales of proteolysis and function.* Curr Opin Chem Biol, 2013. **17**(1): p. 73-82.

356.   Kleifeld, O., et al., *Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products.* Nat Biotechnol, 2010. **28**(3): p. 281-8.

357.   Mansilla, M.J., Montalban, X., and Espejo, C., *Heat shock protein 70: roles in multiple sclerosis.* Mol Med, 2012. **18**(1): p. 1018-28.

358.   Mohler, K., et al., *MS-READ: Quantitative measurement of amino acid incorporation.* Biochim Biophys Acta Gen Subj, 2017. **1861**(11 Pt B): p. 3081-3088.

359.   Mohler, K. and Ibba, M., *Translational fidelity and mistranslation in the cellular response to stress.* Nat Microbiol, 2017. **2**: p. 17117.

360.   Abeysiriwardena, N.M., Gascoigne, S.J.L., and Anandappa, A., *Algal Bloom Expansion Increases Cyanotoxin Risk in Food.* Yale J Biol Med, 2018. **91**(2): p. 129-142.

361.   Samardzic, K., et al., *Toxicity and bioaccumulation of two non-protein amino acids synthesised by cyanobacteria, beta-N-Methylamino-L-alanine (BMAA) and 2,4-diaminobutyric acid (DAB), on a crop plant.* Ecotoxicol Environ Saf, 2021. **208**: p. 111515.

362.   Samardzic, K. and Rodgers, K.J., *Cell death and mitochondrial dysfunction induced by the dietary non-proteinogenic amino acid L-azetidine-2-carboxylic acid (Aze).* Amino Acids, 2019. **51**(8): p. 1221-1232.

363.   Labbadia, J. and Morimoto, R.I., *The Biology of Proteostasis in Aging and Disease.* Annual Review of Biochemistry, 2015. **84**(1): p. 435-464.

364.   Prakash, A., et al., *Reinspection of a Clinical Proteomics Tumor Analysis Consortium (CPTAC) Dataset with Cloud Computing Reveals Abundant Post-Translational Modifications and Protein Sequence Variants.* Cancers (Basel), 2021. **13**(20): p. 5034.

365.   Geladaki, A., et al., *Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics.* Nat Commun, 2019. **10**(1): p. 331.

366.   Piper, J.A., et al., *Pro-Inflammatory and Pro-Apoptotic Effects of the Non-Protein Amino Acid L-Azetidine-2-Carboxylic Acid in BV2 Microglial Cells.* Curr Issues Mol Biol, 2022. **44**(10): p. 4500-4516.

367. Simpson, S., Jr., et al., *Trends in the epidemiology of multiple sclerosis in Greater Hobart, Tasmania: 1951 to 2009.* J Neurol Neurosurg Psychiatry, 2011. **82**(2): p. 180-7.

368. Oluich, L.J., et al., *Targeted ablation of oligodendrocytes induces axonal pathology independent of overt demyelination.* J Neurosci, 2012. **32**(24): p. 8317-30.

369. Cianca, R.C., et al., *Reversed-phase HPLC/FD method for the quantitative analysis of the neurotoxin BMAA (beta-N-methylamino-L-alanine) in cyanobacteria.* Toxicon, 2012. **59**(3): p. 379-84.

370. Andrys, R., et al., *Improved detection of beta-N-methylamino-L-alanine using N-hydroxysuccinimide ester of N-butylnicotinic acid for the localization of BMAA in blue mussels (Mytilus edulis).* Anal Bioanal Chem, 2015. **407**(13): p. 3743-50.

371. Banack, S.A., et al., *Distinguishing the cyanobacterial neurotoxin beta-N-methylamino-L-alanine (BMAA) from other diamino acids.* Toxicon, 2011. **57**(5): p. 730-8.

372. Banack, S.A., et al., *Detection of cyanobacterial neurotoxin beta-N-methylamino-l-alanine within shellfish in the diet of an ALS patient in Florida.* Toxicon, 2014. **90**: p. 167-73.

373. Caller, T.A., et al., *Spatial clustering of amyotrophic lateral sclerosis and the potential role of BMAA.* Amyotroph Lateral Scler, 2012. **13**(1): p. 25-32.

374. Banack, S.A., et al., *Detection of cyanotoxins, beta-N-methylamino-L-alanine and microcystins, from a lake surrounded by cases of amyotrophic lateral sclerosis.* Toxins (Basel), 2015. **7**(2): p. 322-36.

375. Banack, S.A., Caller, T.A., and Stommel, E.W., *The cyanobacteria derived toxin Beta-N-methylamino-L-alanine and amyotrophic lateral sclerosis.* Toxins (Basel), 2010. **2**(12): p. 2837-50.

376. Caller, T.A., et al., *A cluster of amyotrophic lateral sclerosis in New Hampshire: a possible role for toxic cyanobacteria blooms.* Amyotroph Lateral Scler, 2009. **10 Suppl 2**: p. 101-8.

377. Wichmann, C., et al., *MaxQuant.Live Enables Global Targeting of More Than 25,000 Peptides.* 2019. **18**(5): p. 982-994.

378. Muehlbauer, L.K., et al., *Global Phosphoproteome Analysis Using High-Field Asymmetric Waveform Ion Mobility Spectrometry on a Hybrid Orbitrap Mass Spectrometer.* Anal Chem, 2020. **92**(24): p. 15959-15967.

379. Zhao, Y. and Jensen, O.N., *Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques.* Proteomics, 2009. **9**(20): p. 4632-41.

380. Leutert, M., et al., *R2-P2 rapid-robotic phosphoproteomics enables multidimensional cell signaling studies.* Mol Syst Biol, 2019. **15**(12): p. e9021.

381. Tape, C.J., et al., *Reproducible automated phosphopeptide enrichment using magnetic TiO2 and Ti-IMAC.* Anal Chem, 2014. **86**(20): p. 10296-302.

382. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.