

Review

# Misincorporation Proteomics Technologies: A Review

Joel R. Steele <sup>1,2</sup> , Carly J. Italiano <sup>2</sup> , Connor R. Phillips <sup>2</sup>, Jake P. Violi <sup>1,2</sup>, Lisa Pu <sup>2</sup> , Kenneth J. Rodgers <sup>2</sup>   
and Matthew P. Padula <sup>1,\*</sup> 

<sup>1</sup> 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.)

<sup>2</sup> 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



**Citation:** Steele, J.R.; Italiano, C.J.; Phillips, C.R.; Violi, J.P.; Pu, L.; Rodgers, K.J.; Padula, M.P. Misincorporation Proteomics Technologies: A Review. *Proteomes* **2021**, *9*, 2. <https://doi.org/10.3390/proteomes9010002>

Received: 16 December 2020

Accepted: 18 January 2021

Published: 21 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 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].

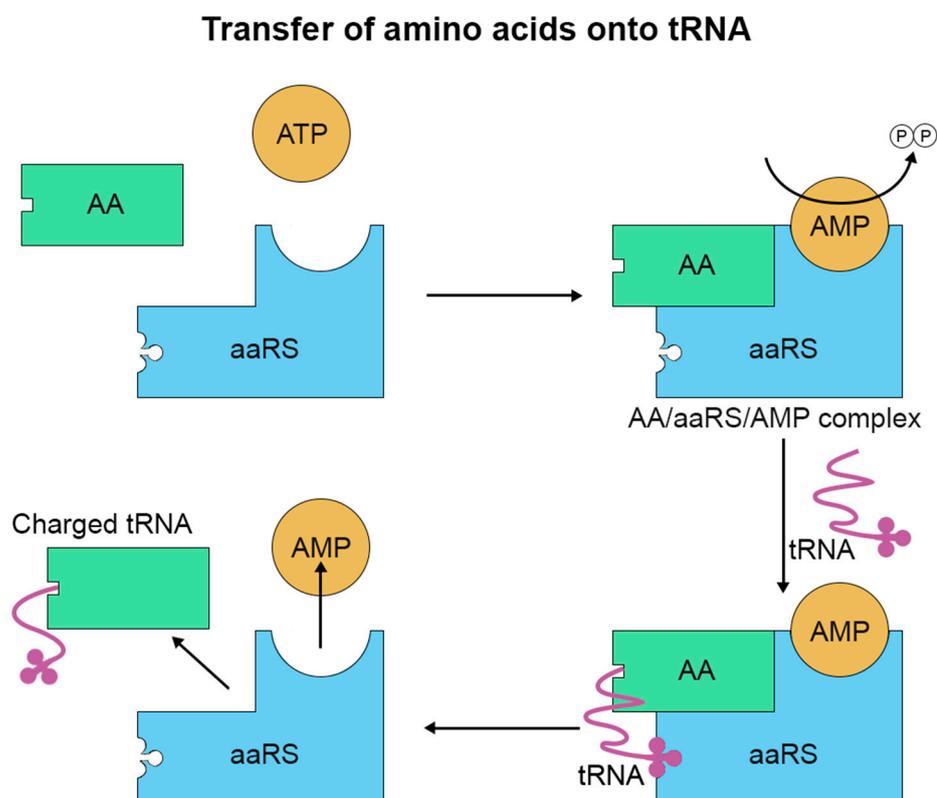
Additionally, and importantly for this review, variant proteoforms may be generated by mechanisms other than post-translational modification. During protein translation, an incorrect amino acid may be inserted into the growing peptide chain, resulting in a modification in the final proteoform. Such errors 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 NPAA 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

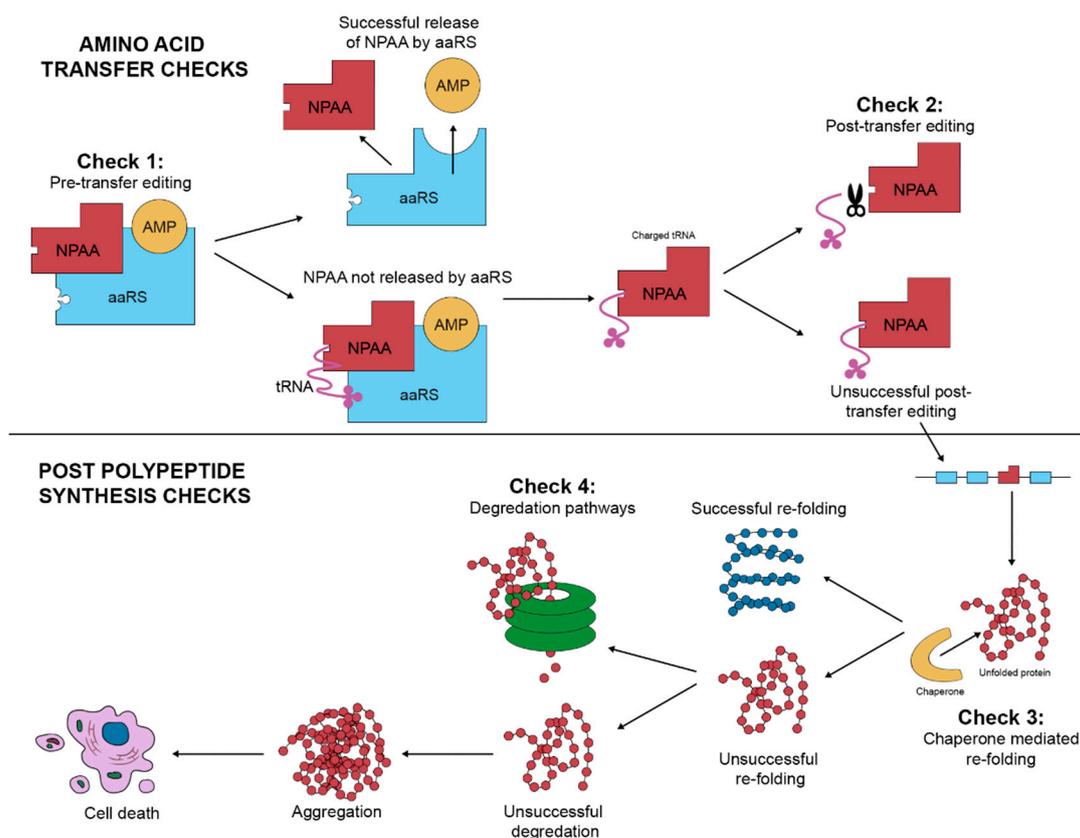
Protein translation involves the cognate amino acid (AA) being charged by its appropriate aminoacyl tRNA synthetase (aaRS; Figure 1). The charging is enabled by the hydrolysis of adenosine triphosphate (ATP) and allows the formation of an AA/aaRS/AMP complex. The cognate transfer RNA (tRNA) then binds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



**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 NPAA 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 NPAA 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 cell's 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 NPAA 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 NPAA 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 NH<sub>2</sub> 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  $1 \times 10^{-12}$  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.

**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
Isoleucine	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.

## 5. Sample Processing and Enrichment

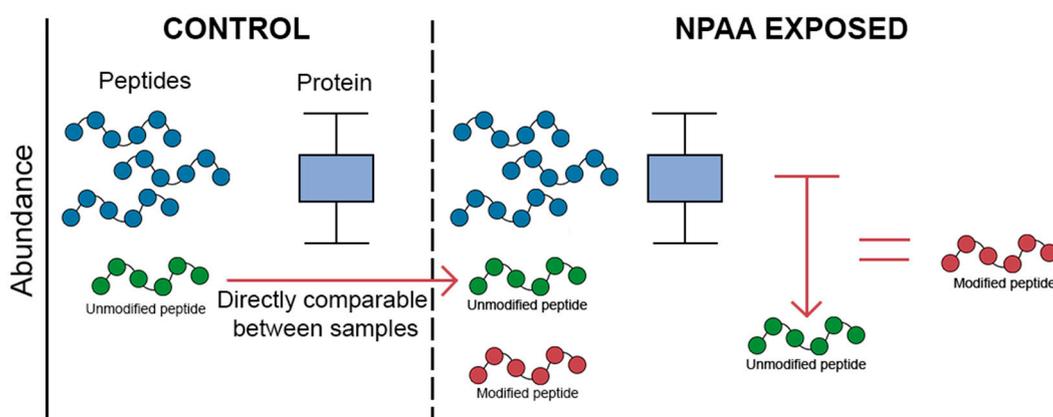
Sample processing and enrichment techniques are critical to ensuring that downstream analytical methods have the optimal chance of detecting low abundance proteoforms, such as those containing NPAA. All the techniques presented here can increase the chance of identification, however, the success of all 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 NPAA 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 NPAA but may give a better chance at identification by dynamic range reduction and increase in the amount of a specific proteoform or peptidiform. 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  $\beta$ -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 NPAA 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-elastic 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-elastic 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 NPAA 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 MS<sub>2</sub> acquisition for speed as 17,500 resolution is used in standard DDA methods, in conjunction with the high-resolution MS<sub>1</sub> 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 MS<sub>1</sub> 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

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 DTIMS determine mobility via how the ions drift in a cell filled with gas, with the only difference being the application of voltage in DTIMS. 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.

### 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 ( <i>Mucuna pruriens</i> ) [103]
	Phenylalanine	+31.9898			
Meta-tyrosine *	Tyrosine	NMS	136.0762	AD/PD [104]	Fescue grass ( <i>Festuca</i> 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				
Azetidine 2 carboxylic acid	Proline	−14.0157	56.05002	MS [19]	Lily of the valley ( <i>Convallaria majalis</i> ); Garden beet ( <i>Beta vulgaris</i> ) [9]
Canavanine	Arginine	+1.9793	131.0933	MS/Systemic lupus erythematosus [106]	Jack bean plant ( <i>Canavalia ensiformis</i> ) [107]
Norleucine	Methionine	−17.9564	86.09697	NNAD	Bacteria
Mimosine	Tyrosine	+16.9902	153.0664	NNAD	Leucaena ( <i>Leucaena</i> spp.) and some <i>Mimosa</i> 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.

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						
<i>Msfragger, Philosopher, PTMShepard</i>	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						
<i>Andromeda</i>	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						
<i>PTM Shepard</i>	Yes	Licensed	Yes	Easy	No	[126]
Spectronaut	Yes	Licensed	No	Easy	Yes	[127]
TagGraph	No	Free	Yes	Advanced	No	[111]
Trans-Proteomic Pipeline						
<i>PTMPProphet</i>	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.

**Author Contributions:** Conceptualization, J.R.S. and M.P.P., J.R.S., C.J.I., C.R.P., J.P.V., L.P., K.J.R. and M.P.P. contributed to the writing, drafting and editing of the manuscript. J.R.S., L.P., C.J.I. and J.P.V. contributed the construction of figures. All authors have read and agreed to the published version of the manuscript

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** J.R.S.: J.P.V., C.J.I. and L.P. are recipients of the Australian Government Research Training Program Stipend. J.R.S. is also the recipient of the Jumbunna Postgraduate Research Scholarship and the University of Technology Sydney's Research Excellence Scholarship.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### 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](#)]

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.; Ibbas, 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 Q $\beta$  RNA translation. *Eur. J. Biochem.* **1984**, *144*, 491–495. [[CrossRef](#)]
46. Lofffield, 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 cerevisiae* 11 Edited 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  $\beta$ -N-methylamino-l-alanine into their proteins. *Toxicon* **2015**, *102*, 55–61. [[CrossRef](#)] [[PubMed](#)]
67. Bettinger, J.Q.; Welle, K.A.; Hryhorenko, J.R.; Ghaemmaghani, 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.; Rühner, 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 PeptideProphet. *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 phosphorylation 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.  $\alpha$ -Amino- $\beta$ -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. Cveticic, 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](#)]