Development of an enrichment method for azide-containing peptides to study proteome dynamics by mass spectrometry

Nessen, M.A.

Citation for published version (APA):
Chapter 1

General Introduction
To understand the basic principles of life, knowledge of the behavior of cells at the molecular level is needed. One key aspect is the dynamics of the cellular proteome, the proteins present in the cell at a given moment. Cellular proteome dynamics comprises changes in protein synthesis rates, post-translational modifications and protein-protein interactions on a proteome-wide scale. Information on these different levels will reveal regulatory circuits underlying cellular adaptation and thereby give insight on cellular behavior. In this thesis the development of enrichment methods directed towards azide-containing peptides aimed at mass spectrometric analysis of the different aspects of proteome dynamics is described.

Mass spectrometry of peptides and proteins

Mass spectrometry is a very fast and highly sensitive method that since the advent of soft ionization techniques such as electrospray ionization (ESI)\(^1\) and matrix assisted laser desorption ionization (MALDI)\(^2\) can be used to study a wide range of questions at the protein level.

A common approach is to subject isolated proteins or proteomes first to digestion, generating a mixture of peptides. Subsequent mass spectrometric analysis of this peptide mixture in combination with database searching allows identification and quantification of the proteins. Many proteomic studies are focused on identification and quantification of as many proteins as possible that are expressed under specific conditions. However, the enormous complexity of peptide mixtures from complete proteomes implies that even with the most advanced fractionation methods and mass spectrometric equipment, identification and quantification is often based on only a few or even single unique peptides. Especially, in studies in which the interest lays in a subset of peptides, such as in phosphoproteomics\(^3\),\(^4\), this might hamper identification of peptides of interest.

Complexity of peptide samples

The complexity of the derived peptide mixtures is given by the number of proteins present in the cell at a certain time point, combined with the differential modifications of proteins and the dynamic range of expression. For example, the genome of the bacterium *Escherichia coli* comprises around 4200 (potential) genes, from which it can be expected that around 2000 proteins are expressed at the same time. Taking into account that the average sized protein after tryptic digestion yields 30 peptides, mixtures of at least 60,000 peptides are obtained. These numbers do not take into account post-translational modifications and the
differential level of expression of proteins. When the interest lies in a subset of peptides, in a specific modification or in proteins that are low abundant, mass spectrometric analysis will not be able to give a complete or even representative coverage. Reduction of the complexity is needed to improve the analysis, which can be accomplished by two approaches.

**Reduction of sample complexity on protein level**

First, complexity of the (isolated) protein sample can be reduced before generating the peptide mixture. Most commonly, one- or two-dimensional gel electrophoresis is used, separating proteins based on molecular mass (one dimensional) or combined with net charge in a second dimension (two dimensional). From the gel, proteins are digested and the tryptic peptide mixtures are analyzed by mass spectrometry. In this way, two-dimensional gel electrophoresis yielded identification of more than 1000 proteins from *E. coli*.\(^5\) Furthermore, gel electrophoresis can be used to obtain protein expression profiles, allowing quantification based on the intensity of the spots on the gel.

However, there are some drawbacks: often proteins are not completely resolved, yielding mixtures in which low abundant proteins cannot be detected. As well, hydrophobic proteins, or proteins with extreme basicity or acidity are underrepresented in the analysis. Furthermore, the method is laborious and time consuming. This has lead to the development of automated methods based on the separation of peptides, as will be discussed in the next section.

Alternative methods to reduce complexity are based on depletion of high abundant proteins or selective isolation of proteins of interest. For the first, immuno affinity separation has been shown to selectively remove highly abundant proteins from serum, improving the detection of lower abundant proteins in human serum.\(^5\) Using the affinity for lectin, glycoproteins can be selectively isolated and analyzed by mass spectrometry.\(^7\)

**Reduction of sample complexity on peptide level**

A second approach to reduce sample complexity is based on the separation after tryptic digestion on the peptide level.\(^8\) Prior to ESI-MS analysis, peptides are separated on a reversed phase column directly coupled to the mass spectrometer. To improve separation further, two- or three-dimensional liquid chromatography methods have been used prior to mass spectrometric analysis, combining e.g. strong cation exchange (SCX) or hydrophilic interaction chromatography (HILIC) with reversed phase chromatography.\(^9,11\) This has lead to identification and quantification of up to hundreds of proteins for *E. coli*.\(^12\)
Additional fractionation of peptides can be accomplished by selective separation of peptides of interest. For phosphorylated peptides separation by immobilized metal affinity chromatography (IMAC) and more recently TiO$_2$ has been proven to be very successful, identifying numerous phosphorylated peptides.

In order to increase the coverage of proteins identified in proteomic studies, methods have been developed to enrich peptides containing rare amino acids. By isolation of only a few peptides per protein, complexity of the peptide mixture will be reduced enormously. The most common amino acid used for this purpose is cysteine, representing around 85% of all proteins in E. coli. Several reagents have been developed that are directed towards the thiol-group of the cysteine residue. One example is the isotope coded affinity tag (ICAT), a thiol reactive reagent, containing an affinity tag for enrichment by biotin-avidin affinity purification and a linker that can be used for incorporation of stable isotopes for quantification.

Another approach that has been developed by Gevaerts et al. is based on a change in chromatographic behavior of peptides upon a selective chemical modification of one of the consisting amino acids and is called combined fractional diagonal chromatography (COFRADIC). Selective modification of
peptides after separation on a reversed phase column, and running the modified samples again under the same conditions, allowed enrichment of methionine-containing peptides\textsuperscript{23, 24} and N-terminal peptides.\textsuperscript{25, 26} Methionine is a relatively rare amino acid, with an abundance of around 2.5%, but it does represent 99.7% of all proteins in \textit{E. coli}. When N-terminal methionine residues are not taken into account, 95.8% of the proteins are represented.\textsuperscript{23}

Other amino acids being used for selective enrichment, either via covalent solid phase or affinity purification, include histidine,\textsuperscript{27} tryptophan\textsuperscript{28} and tyrosine.\textsuperscript{29} These examples of enrichment methods for peptides bearing a specific amino acid illustrate the interest in improvement of sample preparation to enhance the sensitivity of mass spectrometric analysis. However, with the techniques described above only partial elucidation of proteome dynamics can be obtained. Studies on (rare) post-translational modifications, newly synthesized proteins or the interactions between proteins (in a protein complex) by mass spectrometry, give rise to samples in which peptides of interest are represented by the minority of a mixture that is in addition very complex.

**Enrichment of peptides labeled with a bio-orthogonal label**

A way to facilitate the study of proteome dynamics is by the introduction of a bio-orthogonal label in or to a peptide or protein. This label should allow selective sequestration and analysis of the peptides of interest. Further requirements of bio-orthogonal labels are:

i) compatibility with the natural environment of the organism or cell, not modifying or be modified by molecules present

ii) the possibility to selectively introduce the label in or to the protein (or peptide) of interest, without disturbance of the molecular structure

iii) after labeling the introduced moiety should be the handle for a selective reaction to enrich the desired peptide(s).

In Figure 1, the approach for selective enrichment of peptides of interest is given. First, a label (in this case an azide) is introduced in or to the protein, either \textit{in vivo} or \textit{in vitro}, after which digestion by trypsin yields a peptide mixture of which the minority of the peptides will contain the label. Selective enrichment via covalent capturing on a solid phase will allow isolation of the peptides of interest and analysis of the simplified peptide mixture by mass spectrometry.
Bio-orthogonal reactions directed towards azides

Bio-orthogonal labels that have been used in living systems include ketones, alkynes and azides, of which the last is most popular.\textsuperscript{30-33} The azide is a small functional group and inert towards the biological system. It does, however, provide a chemical handle that can be used for selective modification.

The first bio-orthogonal reaction performed in living systems was the reaction of the azide with a phosphine, known as the Staudinger ligation\textsuperscript{34-36} (Figure 2a). In this way, azide-containing sugars have been selectively modified on cell surfaces\textsuperscript{34} and a purified azide-containing protein was labeled by a fluorophore.\textsuperscript{37}

A second orthogonal reaction of azides is the reaction with alkynes. The cycloaddition of the azide with alkynes proved to be very selective, but for a good and selective reaction elevated temperatures were needed.\textsuperscript{38} To increase the reaction rate and selectivity, Cu(I) was used as a catalyst for the reaction with terminal alkynes (Figure 2b).\textsuperscript{39, 40} However, because of its toxicity, Cu(I) is not suitable to be used for labeling in living systems.

This has led to the development of a group of alkynes that have improved reactivity because of ring-strain (Figure 2c).\textsuperscript{31, 32, 41} Bertozzi et al. have invested in the development of a range of different cyclooctynes that can be used for the strain-promoted (3+2) cycloaddition with azides \textit{in vivo} (Figure 3). The first generation of cyclooctynes (1-3) had a much slower reactivity compared to the Cu(I) catalyzed cycloaddition.\textsuperscript{42, 43} Inclusion of one fluorine in the octyne ring improved reactivity modestly (4).\textsuperscript{43} A difluorinated cycloctyne (DIFO, 5) showed similar reactivity as the Cu(I) mediated cycloadditions.\textsuperscript{44} Other cyclooctynes have been developed, obtained via a simpler synthetic route (6-7)\textsuperscript{45} or with improved solubility (8).\textsuperscript{46} Furthermore, a dibenzocyclooctyne (9) has been described.
by Boons et al., with a comparable reactivity to DIFO, but synthetically more easily accessible.\textsuperscript{47} In general, this group of molecules gives a fast and specific reaction with azides, which is of great importance for selective analysis by mass spectrometry.

Several studies have shown the applicability of azides in biological systems\textsuperscript{35, 42-45, 48-53}: to visualize glycan trafficking in the cell, azide-reactive fluorophores can be used\textsuperscript{44, 51} and global \textit{de novo} protein synthesis has been monitored by using azide-containing amino acids.\textsuperscript{50, 54-56} The use of azide-containing glycans,\textsuperscript{57} lipids\textsuperscript{30} or azide-containing reagents directed towards posttranslational modifications in combination with a sequestration method for azides will allow the mapping of posttranslational modification sites. In the following sections, an overview of the use of azides to study proteome dynamics is given, summarizing what has been done and discussing what potential this bio-orthogonal label has.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{cyclooctyne_reagents.png}
\caption{Overview of cyclooctyne reagents for strain-promoted (3+2) cycloaddition with azides.}
\end{figure}
The application of azides in biological systems for the study of proteome dynamics

The use of azide-glycans to study glycosylation of proteins and trafficking in the cell membrane

Visualization of glycan trafficking in the cell has been accomplished by azide-reactive fluorophores. Cells treated with different types of azide-containing sugars (example in Figure 4a) incorporate the labeled sugars via the biosynthetic machinery. After incorporation, the azide moiety is used for selective labeling by fluorophores via the Staudinger ligation or via the more successful strain-promoted (3+2) cycloaddition. This allowed metabolic labeling of glycans in an auxotroph strain of Saccharomyces cerevisiae, visualization of glycan trafficking dynamics in Jurkat cells, tagging of glycans in living mice and the imaging of glycans in developing zebrafish.

To study the glycoproteome, these azide-sugars can also be used for the enrichment of glycosylated peptides, allowing identification of post-translational modified proteins. After incorporation of the sugars, proteins are isolated and digested to peptides. A mixture of peptides will be obtained from which glycosylated peptides can be isolated by selective capturing via the azide moiety. Mass spectrometric analysis of the isolated (and deglycosylated) peptides will allow identification of the glycosylation sites and mapping of the glycoproteome.
The use of azide-containing amino acids to study protein synthesis in E. coli

The incorporation of non-natural amino acids in proteins is of great use in protein engineering and functional studies. Chemical (reactive) groups can be introduced site- or residue specifically, giving possibilities for manipulation of protein structure and function.\textsuperscript{49,62,63} In addition, it allows tagging and visualization of proteins \textit{in vivo}.\textsuperscript{30}

Azide-containing amino acids that have been described to be successfully incorporated in proteins in \textit{E. coli} are azidophenylalanine (11)\textsuperscript{64} and the methionine analogue azidohomoalanine (azhal, 12)\textsuperscript{35} (Figure 4b). A new orthogonal aminoacyl-tRNA synthetase/tRNA pair with specificity for azidophenylalanine was developed for site specific incorporation of the photocross-linking amino acid. Azhal is incorporated into proteins by the cell’s translational machinery in methionine auxotroph \textit{E. coli} cells grown in methionine depleted, azhal rich medium.\textsuperscript{35,49,55}

The aminoacyl-tRNA synthetase for methionine (MetRS) tolerates the non-natural amino acid azhal and readily incorporates it into proteins, with a $k_{\text{cat}}/K_m$ that is 390 times lower for azhal\textsuperscript{35} and to ensure complete labeling of proteins absence of methionine is required. Labeling of proteins can be achieved by using a methionine auxotroph \textit{E. coli} strain, unable to generate methionine itself and dependent on medium supplements, and methionine depleted growth medium.

We have shown that for the first 30 minutes after inoculation the growth of the methionine auxotroph MTD123 on azhal is similar to growth on methionine, with a doubling time of approximately 1 hour.\textsuperscript{55} Toxic side effects of azhal not related to incorporation in proteins were excluded by growth on mixtures of azhal and methionine. And even though many proteins retain their function after incorporation of azhal, it has an effect on cell functioning and cell growth on the longer term. Nevertheless, to study the dynamics of the proteome labeling periods of only minutes will be used, a time scale on which cell functioning is expected not to be affected yet.

To follow protein synthesis the methionine-analogue azhal (12), containing an azide moiety, has been incorporated into proteins in mammalian (HEK 293) cells\textsuperscript{50,54} and in methionine auxotroph \textit{E. coli} cells.\textsuperscript{55,56} This allows discrimination between newly synthesized proteins, containing azhal, and the proteins already present, containing methionine. In addition, it provides a handle for the selective enrichment of the newly synthesized proteins or peptides derived from the labeled proteins, enhancing mass spectrometric analysis.
Identification of newly synthesized proteins will allow us to follow changes in protein composition upon a switch in environmental conditions of *E. coli* and by pulse-labeling experiments with azhal protein synthesis rates can be determined. Furthermore, comparison of protein synthesis rates on a proteome wide scale with mRNA levels at the time of the pulse allows to determine whether changes of protein synthesis rates are achieved by adjusting mRNA levels (regulation at transcription level) or by adjusting the frequency of translation of the mRNA (regulation at translation level). Such data will reveal regulatory circuits underlying cellular behavior during adaptation.

**Other applications of azide-containing amino acids in biological systems**

Other applications of azide-containing amino acids in biological systems include the incorporation of azhal in virus particles to construct polyvalent particles with the possibility to introduce a wide variety of functional groups. And most recently, azhal has been used to visualize *in situ* newly synthesized proteins in rat hippocampal neurons, by coupling of an alkyne-fluorophore via the copper catalyzed cycloaddition. In this study, diffusion of newly synthesized proteins in neurons has been studied as well. The newly synthesized azhal-containing proteins were first coupled to DIFO (5) linked to a biotin affinity group. Then, labeling with streptavidin quantum dot particles allowed visualization of the proteins in living neurons.

**The use of azide-containing cross-linkers to study protein-protein interactions**

A last application of azide-containing molecules to study proteome dynamics discussed here is the chemical cross-linking of proteins. Protein cross-linking can provide information on spatial distance constraints to reveal protein-protein interactions and to enable validation of 3-D structure models. The mapping of linked amino acids will be tremendously facilitated by sequestration of the rare and low abundant cross-linked peptides in protein digests, using azide-containing cross-linkers, such as the recently described bis(succinimidyl)-3-azidomethyl glutarate (BAMG (13); figure 4c). This amine-reactive bis(succinimidyl) ester is used to covalently link primary amines in a protein or protein complex, which are found in the lysine residues and the N-terminus. From the linked amines, distances can be derived by the length of the spacer of the cross-linker used. Hereby models of proteins can be validated and refined and interaction sites of protein complexes can be determined.

However, the combination of the low-abundance of cross-links and the high number of possible combinations of linked peptides makes the identification of
cross-links very difficult. The enrichment of structure informative cross-links will improve the analysis enormously, which in the future will allow analysis of protein complexes.

**Outline of the thesis**

In this thesis the development and application of enrichment methods directed towards azide-containing peptides to aid in mass spectrometric analysis are described. In *Chapter 2* the development of an enrichment method based on covalent capturing on solid phase is described. *Chapter 3* shows the development of an alternative enrichment method based on affinity purification using fluororous solid phase extraction (FSPE). *Chapter 4* describes the study on the use of the azide reactive cyclooctyne (ARCO)-resin to study different aspects of proteome dynamics. In *Chapter 5* the use of the ARCO-resin has been optimized to study protein structures in combination with chemical cross-linking and mass spectrometry. Finally, *Chapter 6* discusses the advantages and disadvantages of the developed method(s), presents improvements to the method and looks ahead to application of the method for further study.
Chapter 1

References

General Introduction

Chapter 1