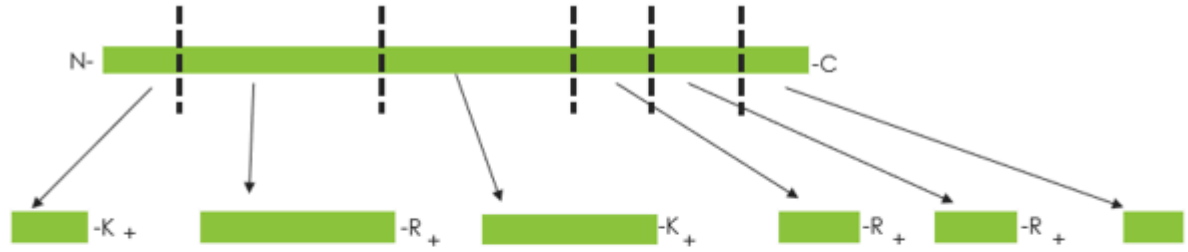
Peptide Mapping Sample Preparation - Good Practices for Tryptic Digests

In the world of biotherapeutics and biosimilars, regulatory agencies have characterization requirements that are primarily focused around the physicochemical properties of the therapeutics that affect efficacy, safety, and immunogenicity. Chromatography plays an important role in three specific areas of the physicochemical property analysis, being higher order structure, primary structure, and post translational modifications. One key technique to determine primary structure is peptide mapping, which typically involves a serine protease digestion of a purified or isolated protein followed by analysis using HPLC. The most common protease for these types of digestions is trypsin, however even within the digestion step there are many important practices we can apply to improve the result, and these will be discussed in the tech tip below.

Most protein digests are done using trypsin due to its well defined specificity. Trypsin will only hydrolyze peptide bonds in which the carbonyl group is followed by either an arginine (Arg) or lysine (Lys) residue, with the following notable exceptions: when Lys and Arg are N-linked to negatively charged amino acids, or if proline (Pro) is positioned on the C terminal side Lys or Arg.

Trypsin is also favored as the resulting peptides have basic residues at the C-terminus, making them easier to ionize by mass spec, which is typically the preferred method of detection for these applications. One final benefit of using trypsin is the mass range of the resulting peptides, which is typically between 600 and 4000 Da which is ideal for subsequent analysis when using either LC/UV LC MS/MS.



N – ala – ser – pro – lys – gly – gly – arg – trp – asp – glu – lys – cys - **C**

N – asp – lys – glu – lys – glu –arg – trp – gly – arg – pro – cys - **C**

A basic tryptic digest comprises of a number of steps, for which a few key points of the process are summarized below:

Denaturation and reduction: A chaotrope is necessary to completely unfold the protein such that it accessible for trypsin, for which urea or guanidine are commonly used. Urea is known to reduce sequence coverage as it is can break down and subsequently carbamylate both lysines and cysteines, and so will be avoided in some laboratories. The easiest reagent to use for denaturing is guanidine hydrochloride (GnHCl), however it must be removed with buffer exchange prior to digestion, as it inhibits trypsin.

Alkylation: There are two common alkylating agents, Iodoacetamide (IAM) and Iodoacetic acid (IAA). IAM is the most common as it reacts quickly, however IAA is also used, as it shows less non-specific alkylation. With both reagents the addition of dithiothreitol (DTT) will also help prevent further non-specific alkylation.

Buffer exchange: Once alkylation has taken place, it is critical to then remove residual alkylating reagents as well as any denaturing components prior to the digest. There are three main approaches to this type of sample clean up: dialysis, molecular weight cutoff spin filters, and desalting columns. Dialysis leads to long clean up times, but there is very little loss of sample. Spin filters and desalting columns are susceptible to loss, but do offer a more rapid method for sample clean up over dialysis.

Digestion: Here we are looking at the considerations for digestion using trypsin. Tryptic peptides can be relatively short, and there is also the possibility missed cleavages will occur. Porcine trypsin is most commonly used for digestion, and this enzyme has a preference for Arg, with a lower efficiency for Lys. Other missed cleavages include Asp or Glu next to Lys or Arg on the C terminus, as well as Proline (Pro) on the N terminus. Trypsin alone can only cover a restricted portion of the proteome, and parallel proteases also need to be used, including AspN or GluC in many cases for complete sequence coverage.

When sequence coverage is critical orthogonal proteases can be used, and are summarized below.

Enzyme	Cleavage Site	Advantage	Limitations
ArgC	C-terminal of R	Increase proteome coverage qualitatively	Larger peptide fragments than trypsin
AspN	N-terminal of D	Wide range pH	Larger peptide fragments than trypsin. Missed cleavages are possible
Chymotrypsin	C-terminal of F, Y, L, W	Orthogonal proteome space to trypsin	Non-specific for hydrophobic peptides resulting in missed cleavages
GluC	C-terminal of D	Increase proteome coverage qualitatively	Larger peptide fragments than trypsin
LysC	C-terminal of K	Complements trypsin	Overlaps with trypsin so not a great orthogonal protease
Pepsin	C-terminal of Y F W	High activity at low pH used for HDX-MS	Complex resulting spectra

Below is a summary of the type of parameter restrictions you should look to follow when digesting proteins with trypsin.

Parameter	Range	Example	Considerations
Enzyme to substrate ratio	1:5 – 1:50	1:20	Optimize digestion length and temperature
Temperature (°C)	37-90	56	Workflow throughput Deamidation above 37 °C
Digestion Length (hours)	2-16	4	Digestion completion Non-trypsin activity Trypsin autocleavage
pH	7.5-8.2	7.6	Deamidation at pH > 7.5 Optimal trypsin activity
Protease Supplementation	n/a	LysC	Digestion completion

In conclusion, it is not always the analytical method or HPLC analysis which fully dictates sequence coverage when looking at primary structure of a protein. The steps and precautions taken during sample preparation can also play a vital role, and some of these are summarized above. Where appropriate steps are taken in the digestion of the protein, the subsequent HPLC or LCMS analysis becomes a much easier process to ensure adequate coverage is obtained.