^{the} Analytical Scientist

Developments in Peptide Mapping Technology for the Biopharmaceutical Industry

Developments in Peptide Mapping Technology for the Biopharmaceutical Industry

What is peptide mapping?

Peptide mapping is a critical workflow in biotherapeutic protein characterization and is essential for elucidating the primary amino acid structure of proteins. For recombinant protein pharmaceuticals, such as monoclonal antibodies (mAbs) and antibody-drug conjugates (ADCs), peptide mapping is used for proof of identity, primary structural characterization and quality assurance (QA).

Global regulatory agencies, including US Food and Drug Administration (US FDA) and European Medicines Agency (EMA), look to harmonized guidelines from the International Council for Harmonisation (ICH). ICH Q6B covers the test procedures and acceptance criteria for biologic drug products, and specifies the use of peptide mapping as a critical quality test procedure for drug characterization used to confirm desired product structure for lot release purposes.

In order to generate a peptide map, the therapeutic protein must first be digested into its constituent peptides via a chemical or enzymatic reaction. Robust separation and identification of the resultant peptides then provides insight into a protein's full sequence information; displaying each amino acid component and the surrounding amino acid microenvironment, including disulfide linkage information. Structural characterization at this level highlights post translational modifications (PTMs)



Figure 1: Overview of the typical structure and associated modifications found in a mAb.

such as site-specific glycosylation, amino acid substitutions (sequence variants) and/or truncations which may result from erroneous transcription of complementary DNA (Figure 1).

Within a bioproduction environment peptide mapping is necessary for manufacturing process monitoring and quality control (QC). It facilitates product comparability testing, which is necessary to identify any productrelated impurities, such as deamidation and/or oxidation following any formulation, manufacturing process or storage change.

Due to its complexity and inherent variability, peptide mapping is generally performed in a comparative manner; for example, biosimilars would be compared to a reference or control substance, such as the innovator biologic, in a side-byside experiment. An in-depth analysis is then required to identify minor and even isobaric differences in protein primary structure.

The modern biopharmaceutical and protein research laboratory is tasked with providing high quality analytical results, often in high-throughput, regulated environments. Some technologies currently employed for biopharmaceutical peptide mapping are subject to:

- high levels of irreproducibility
- poor sensitivity
- high levels of time-consuming manual work – with protracted methodologies that are not amenable to automation and often require 24 hours to achieve full protein digestion.

This variability impacts data confidence. Moreover, it increases potential for introduction of non-product related artefacts during manual sample handling. Outlined herein are the most recent advances in protein sample preparation chemistries, ultra-high performance liquid chromatography (UHPLC), mass spectrometry (MS) hardware and intuitive software for the generation of comprehensive, confident peptide maps.

Protein sample preparation

In-solution enzymatic digestion is the most commonly employed method of protein digestion, but there are several disadvantages to this approach. Before in-solution digestion can be performed, protein denaturation, disulfide bond reduction, and free sulphydryl alkylation are required. This is to unravel the protein's tertiary structure, break down its disulfide bonds, and prevent disulfide bond scrambling or shuffling. This results in a multistep preparation process using several reagents. These multiple steps provide the most open protein structure giving access for optimal enzymatic digestion. However, they can also introduce unwanted artefacts and amino acid modifications.

Protein denaturation

An assortment of denaturing methods can be used to unfold a protein (Figure 2), disrupt its non-covalent bonds, and reveal its primary structure. Denaturants can include strong acids or bases, inorganic salts or organic solvents, radiation, and heat. Commercially available 'optimised' denaturants can be employed, but it has also been found that heat alone can effectively denature proteins, avoiding the introduction of new chemicals and potential modifications/ side reactions into the sample, such as lysine carbamylation with the use of urea. Complete tryptic digestion of a mAb can be achieved in 60 minutes or less at 70°C, when the mAb is in an unfolded state and all the digestion sites are exposed to the protease (Figure 3).

Reduction of disulfide bonds

mAbs are typically comprised of four chains held together via disulfide bonding (Figure 4). Disulfide bonds between cysteine residues within a protein are often broken prior to digestion through reduction of their thiol functional group using chemicals which contain sulfhydryl or phosphine groups. Common reduction agents include dithiothreitol (DTT) or tris-2-



Figure 2. Unfolding or 'denaturing' of a native mAb protein.



Figure 3. a) DSC thermogram of immunoglobulin G (IgG) (6mg/mL; mouse IgG2b in a 10mM phosphate buffer pH 8.1; 0.5°C/min)(1). b) Time course of digestion monitoring the appearance of a signature peptide. Complete digestion occurs at 60 min at 70°C.

carboxyethylphosphine hydrochloride (TCEP). This step can be omitted if the location of disulfide bonds is to be characterized as part of the peptide map.

Alkylation of cysteine residues

Cysteine alkylation is performed predigestion, following a reduction step. Alkylation of the free thiol group of cysteine residues to the stable S-carboxymethylcysteine ensures that disulfide bonds do not reform and disulfide scrambling is avoided. Often the alkylation reaction for peptides does not go to completion, resulting in a mixture of alkylated and nonalkylated peptides. This increases the number of chromatographic peaks detected, complicating data evaluation, and leading to inaccuracies in quantitative experiments.

Protein digestion

The most commonly employed protease for protein digestion is trypsin; a serine protease found in the digestive system of many vertebrates. Trypsin behaves in a highly predictable manner, cleaving proteins at the C-terminus of lysine (K) and arginine (R) amino acid residues, with the exception of when they are followed by a proline (P) residue. For most therapeutic proteins this pattern of cleavage provides peptides which are an amenable size for effective chromatographic retention (based on their pKa) and separation, and for optimal electrospray ionisation (ESI) and subsequent mass spectrometric analysis.

Peptides which are very small and polar are often difficult to retain using a standard C18 reversed-phase (RP) column, conversely, peptides which are very large and hydrophobic may be difficult to elute from these columns.

From a MS perspective, instruments often need to be optimized and calibrated for a specific m/z range. The ionization and ion transmission parameters will be tailored for the type and size of the peptides/proteins to be introduced. Most tryptic peptides are within the mass range of 200-2000 Da, thus facilitating easier instrument setup and optimization.

Of course not all protein sequences fall into this 'normal' distribution of K and R residues and will necessitate the use of proteases with different or even nonspecific selectivity for digestion. Other typical enzymes can be seen in Table 1.

Using proteases discretely and then mixing the two (or more) digests together prior to analysis provides redundancy in the sequence information obtained, but it can offer complementary and additional information where one enzyme alone proved insufficient. Through combined enzyme activity an increased number of sites can be cleaved and smaller peptides generated, which can be useful where the problem with using trypsin (for example) alone generated very large peptides, or where specific digestion sites were consistently missed.

A number of parameters require optimization with most protein digestion protocols:

• Protein:Protease ratio



Figure 4. Representation of mAb constituents after reduction of disulfide bonds, producing two heavy (H Chains) and two light (L Chains) chains.

Enzyme Name	Specificity	Cleaves after	Cleaves before	Except
Arg-C	C-terminus of Arginine residue	R		
Asp-N	N-terminal side of Aspartate		D	
Chymotrypsin	C-terminal side of Aryl Amino Acids	F, W, Y, or L		
Glu-C Protease	C-terminal side of Glutamate and Aspartate	E or D		
Lys-C Protease	C-terminal side of Lysine	К		
Pepsin	Non-specific			
Trypsin	C-terminal side of Basic Amino Acids	K or R		P is after K or R
Trypsin (Immobilized)	C-terminal side of Basic Amino Acids	K or R		P is after K or R

Table 1. Common proteases and their respective cleavage sites.

- Buffer conditions
- Temperature
- Time of digestion
- Cessation of digestion and cleanup or dilution

Once the protocol has been defined early in the development phase, it often remains with the drug product throughout its life cycle, in the development, manufacturing, and lot release phases.

Alternatives to traditional in-solution digestion processes include immobilized resin-based or magnetic bead digestion. These new techniques are dramatically



Figure 7. Relative abundance of 12 identified oxidations (a) and 7 deamidation (b) in different samples prepared using various digestion methods (2).

increasing in popularity. They often involve digestion at elevated temperatures with an excess of resinbased, heat-stable protease. This can provide highly reproducible (Figure 5) and efficient (Figure 6) digestion results thus eliminating the need for complex denaturation, reduction, and alkylation steps.

Furthermore, the use of simplified digestion protocols has proven, under rigorous testing, that in comparison to traditional in-solution digestion methods, similar levels for all modifications are detected, and no significant trends of increased PTMs

In-Solution, Heat

400 µg rituximab were denatured in 50 mM tris hydrochloride (HCl) at pH 8.0 and 70 °C for 75 min, followed by a reduction step using 5 mM DTT for 30 min at 70 °C. Alkylation was performed with 15 mM iodoacetamide (IAA) for 30 min at room temperature, and the reaction was quenched by addition of 9 mM DTT. The sample was then diluted 1:10 (v/v) with 50 mM tris HCl pH 8.0. Trypsin was added with a protein/protease ratio of 40:1 (w/w) and digestion was allowed to proceed overnight at 37 °C. Digestion was stopped by addition of trifluoroacetic acid (TFA) to a final concentration of 0.5%.

In-Solution, Urea

400 μ g rituximab were denatured for 75 min in 7 M urea and 50 mM tris HCl at pH 8.0, followed by a reduction step using 5 mM DTT for 30 min at 37 °C. Alkylation was performed with 15 mM IAA for 30 min at room temperature, and the reaction was quenched by addition of 9 mM DTT. The sample was then diluted 1:10 (v/v) with 50 mM tris HCl pH 8.0 to reach a final urea concentration below 1 M. Trypsin was added with a protein/protease ratio of 40:1 (w/w) and digestion was allowed to proceed overnight at 37 °C. Digestion was stopped by the addition of TFA to a final concentration of 0.5%.

SMART Digest

50 μ L rituximab sample, adjusted to 2 mg/mL with water, was diluted 1:4 (v/v) with the SMART Digest kit buffer provided with the kit. It was then transferred to a reaction tube containing 15 μ L of the SMART Digest kit resin slurry, corresponding to 14 μ g of heat-stabile, immobilized trypsin. A time course experiment was performed and tryptic digestion was allowed to proceed at 70 °C for 15, 30, 45, and 75 min at 1400 rpm; a digestion time of 45–60 min was found to be sufficient to achieve digestion completeness for mAb samples. After the digestion, the reaction tube was centrifuged at 7000 rpm for 2 min, the supernatant was transferred to a new tube, and the centrifugation step was repeated. Disulfide bonds were reduced by incubation for 30 min at 37 °C with 5 mM DTT.

(Sample names: SMART Digest, 15, 30, 45, 75 min). All samples were diluted with 0.1% formic acid (FA) in water to a final protein concentration of 100 ng/ μ L, and 2.5 μ g were loaded on the column for all runs.

in any of the conditions are observed. Noteworthy, for many modification sites, e.g. deamidation of asparagine residues, the level of certain modifications detected in resin-based digest samples was actually lower compared to insolution digest samples (Figure 7).

The use of pre-optimized digestion



Figure 5. UV chromatographic overlay of three seperate Thermo Scientific[™] SMART Digest[™] Kit digestions of the same mAb conducted by three individual operators. 15 marked peptides in each sample were used for inter-user/inter-day RSD calculations. Retention time RSD = 0.02% and peak area RSD = 4.5%.



Figure 6. Schematic of a traditional in-solution digestion protocol (a) vs. the Thermo Scientific[™] SMART Digest[™] Kit digestion protocol (b).

kits eliminates the necessity to optimize the ratio of protein to protease, buffer conditions, and digestion temperature. The time of digestion is easily optimized by taking time-points for analysis until no intact protein or very large peptides remain. Clean-up is usually a simple centrifugation or solid phase extraction (SPE). For a range of proteins the optimal digestion times are often provided by the kit manufacturers.

Separation of peptides

The UHPLC system employed for a peptide mapping experiment must be capable of delivering highly stable and precise flow rates and gradients; thus

providing highly reproducible retention times (RTs). RT stability increases confidence in peak assignments. This is of particular importance where ultraviolet (UV) detection alone is employed, and assignments are made solely based on analyte elution time.

The ideal characteristics of an UHPLC system for biopharmaceutical workflows have been extensively reviewed elsewhere (3) but one choice that should be considered is whether a binary (high pressure gradient) or quaternary (low pressure gradient) system is most appropriate. For highthroughput environments, requiring peptide map run times < 20 minutes, a binary pumping system can offer lower gradient delay volumes (GDV) and therefore faster re-equilibration and reinjection times. For separations where throughput is not paramount and system flexibility for other applications or method development is required, a quaternary system also provides highly reliable flow delivery.

Traditionally peptide mapping was a 'long' gradient separation process, with hour-long runs being commonplace. Peptide mapping is such a fundamental workflow, used right the way through the biotherapeutic pipeline, that it can be advantageous to speed things up. If rapid peptide maps could be performed during the cell line development/clone selection phase, this would potentially expedite the transfer of 'hot' candidates into drug process development. In process development, the ability to run faster peptide maps, but maintain the quality and information gained would then propagate into a QA/QC environment; the move to continuous bioprocessing is also an area where this could be of great benefit for rapid at-line process monitoring.

By varying the pressure and flow rate, gradients can be reduced from >30 minutes to just 5 minutes. A 1000 bar pressure system would be capable of achieving a 13 minute gradient, and a 1500 bar system would have the pressure capabilities to achieve a 5 minute separation (Figure 8).

For all five gradient times tested, from 30 minutes down to 5 minutes, a very good separation was achieved and sequence coverages of 100% were obtained from all separation times, both for the light and heavy chain of rituximab.

Column chemistries & conditions for separation of peptides

The column dimension and stationary phase employed is fundamental to



Figure 8. Total ion chromatograms obtained from peptide-mapping experiments of rituximab applying gradient lengths 30, 20, 13, 8, and 5 min (4). Flow rates and resulting pressures are indicated in the individual traces.

The commercially available mAb rituximab (F. Hoffmann-La Roche) was digested using the SMART Digest kit. The sample was diluted 1:4 with the SMART Digest kit digestion buffer included in the kit, and enzymatic digestion was allowed to proceed at 70°C for 45 min at 1400 rpm on a shaker. Disulfide bonds were reduced after the digestion by incubation for 30 minutes at room temperature with 5 mM TCEP. A Thermo Scientific[™] Vanguish[™] UHPLC system with a 2.1 × 250 mm Thermo Scientific[™] Acclaim[™] Vanquish[™] C18, 2.2 µm column and gradients of water and acetonitrile (ACN) with 0.1% FA each were used to separate the peptide mixtures. Five different separation times were applied and compared: 5, 8, 13, 20, and 30 min for the gradient ramping from 4% to 55% eluent B (0.1 % FA in 8:2 ACN/water (v/v)).1.1 (5 min), 1.0 (8 min), 0.6 (13 min), 0.4 (20 min), and 0.4 mL/min (30 min). The Thermo Scientific[™] O Exactive[™] HF hybrid Quadrupole-Orbitrap[™] MS equipped with a HESI-II probe was used for mass spectrometric detection using a full MS/dd-MS2 (Top 5) experiment. The data were acquired with the Thermo Scientific[™] Chromeleon[™] Chromatography Data System, version 7.2 SR4, and Thermo Scientific[™] BioPharma Finder[™] software, version 1.0, was used for subsequent data analysis.

the success of any peptide mapping experiment, with RP chemistries widely used in the biopharmaceutical industry. Silica-based C18 stationary phases with virtually zero silanophilic activities result in superior separation of peptides with minimal band broadening. The need for the commonly used ion-pairing agent, trifluoroacetic acid (TFA), to improve chromatographic peak shapes is no longer a necessity in LC-MS peptide mapping workflows, which allows the use of solvent additives such as FA, which substantially lowers signal suppression and consequently provides a boost in sensitivity with MS detection.

Protein digests are often very complex samples, containing tens or



Figure 9. Example peptide MS/MS mass spectrum showing amino acid fragment ions.



Figure 10. Rapid instrument scan speed for fast peptide mapping (9).

even hundreds of peptides, meaning gradient elution is mandatory. Long columns packed with small particles can be used to achieve larger peak capacities. Additionally, longer gradient times, involving shallow gradient slopes, can be used to achieve higher peak capacities. This has been extensively proven in bottom-up proteomics, where long gradients, up to 12 hours long, have been used to fully exploit the potential resolving power of UHPLC capillary columns (5).

Drawbacks to such long gradient times are twofold:

- 1. Low sample throughput
- 2. Decreased detection sensitivity due to in-column peak dilution



Figure 12. Sequence coverage map of the heavy (right) and light chain (left) of rituximab, showing 100% sequence coverage. Processed using Thermo Scientific™ BioPharma Finder™ software.

Änalytical Scientist

White Paper

The sensitivity issue is particularly important when peptide mapping is performed with UV detection alone. There is a risk that the low-level components will fail to be detected. It is therefore important to achieve high peak capacity and at the same time keep the total run time reasonably short.

It is well documented that temperature also has a major impact on the peak capacity (6,7) of peptides separated by RP. Peptides have increased diffusion rates at higher temperatures, ultimately producing narrower peaks, and thus delivering higher peak capacity peptide maps. Temperature increases from 40 - 80 °C can lead to > 20 % increases in peak capacity for 30 minute gradients (8).

The use of mobile phase pre-heating can avoid thermal mismatch which results in retention and diffusion rate variations inside the column, and ultimately band broadening.

Peptide mass spectrometry

Peptide separation must be combined with some form of detection. Characterization methods based on UHPLC and MS are among the most powerful protein characterization techniques. UHPLC coupled to high resolution, accurate mass (HRAM), Orbitrap-based MS provides precise mass measurements for each peptide from even the most complex samples. Hybrid quadrupole-Orbitrap based LC-MS/MS instruments allow each peptide to be isolated and further fragmented, using higher-energy collisional dissociation (HCD) in the gas phase, which induces fragmentation of peptides along their backbone producing signature fragment ions - the masses of which correspond to the amino acid sequence of the peptide (Figure 9).

Taking the sequence of your protein and performing *in silico* digestion provides a reference dataset in which



Figure 11. Stacked chromatograms of the total ion current (TIC) chromatogram (a) and the UV trace at 214 nm (b) of a SMART Digest Kit digested rituximab sample with subtracted blank baseline. Peak assignment of the tryptic peptides from rituximab (c). Peak labels with 1 correspond to the light chain, and those with 2 correspond to the heavy chain of the mAb. The number after the colon indicates the amino acid region of this particular tryptic peptide.

to compare. An MS generated mass list containing the m/z of both the peptide (MS precursor ion) and fragment (MS/ MS product ions) can be searched against the in silico reference list and thus the protein can be characterized via mapping of its constituent peptides. Modifications can be identified by appending static or variable accurate mass shifts to the in silico reference data. Both MS and MS/MS data is then used for the identification of peptides, facilitating identification of very small peptides and disulfide linked peptides, which could be missed by MS/MS alone.

HRAM peptide mapping analysis

provides fast, precise, and reproducible results which facilitates primary level biotherapeutic characterization with confidence. In a HRAM peptide mapping experiment the MS parent ion spectra reveal the accurate precursor ion masses of the peptides, and MS/ MS product ion spectra reveal the band y-ion amino acid fragments of each peptide, and thus information on the modifications that are present.

Fast UHPLC separations require high MS instrument scan speeds to deliver high quality MS and MS/MS spectra within the time frame of a typical chromatographic peak. The scan speeds achievable using Orbitrap-based



Figure 13. Chromatographic peak shading showing relative quantitation of co-eluting peptides using Thermo Scientific™ BioPharma Finder™ software.

instrumentation have been demonstrated to exceed the demands of even the most rapid UHPLC separation protocols. A data dependent acquisition analysis incorporating MS/MS analysis of the top 5 most intense MS peaks during a rapid 5 minute UHPLC gradient with typical chromatographic peak width of 2.4s results in a sum of 31 scans per peak (Figure 10).

Peptide mapping methods are often developed using combined UV and MS detection. This simplifies the transfer between research and routine environments where UV detection is frequently used alone. In highthroughput routine workflows, peptide mapping experiments are performed for antibody identity confirmation, PTM characterization, and stability studies. Figure 11 shows an example of an overlay of a UV trace at 214 nm and the total ion current (TIC) chromatogram obtained from MS analysis of the same molecule (rituximab) under the same separation conditions. Prior LC-UV-MS setup combined with the high RT stability of the UHPLC system allows confident UV peak assignments (Figure 11 (b)).

Peptide mass fingerprint interpretation Sophisticated software solutions are available which can automate data processing and facilitate in-depth data interpretation. Intuitive workflows provide a powerful, yet user-friendly approach to data processing with interactive plotting features facilitating thorough data interpretation. MS generated spectra can be searched against predicted spectra and/or gold standard data. Peptide identification is achieved by comparing the experimental fragmentation spectrum to the predicted spectrum of each native or modified peptide. Peak areas of related peptide ions under their extracted ion chromatograms (XIC) can be used for relative quantification of modified peptides

Utilizing experimental data, powerful software solutions can seamlessly provide comprehensive results; including amino acid sequence confirmation with mass tolerance, modification, identification, RT, and confidence information.

In order for the analyst to efficiently mine their data, effective visualization tools are required. Routinely, peptide



Where could the future of peptide mapping take us?

Establishing QC for biotherapeutics involves the need to measure numerous critical quality attributes (CQAs).The ability to use HRAM MS in peptide mapping workflows allows the direct measurement of multiple CQAs, which can eliminate traditional lot release testing and increase product knowledge.

CQA measurements via a HRAM MS peptide mapping workflow can both reduce the number of parallel analyses required and increase the product quality profile. This is being referred to as a multi-attribute method (MAM).

Read more about this exciting new direction for peptide mapping within the QC environment.

Figure 14. Schematic showing an LC-UV/MS optimized peptide mapping workflow.

maps containing the entire protein sequence annotated with identified peptides are used (Figure 12). Coverage maps help to quickly assess the success of a peptide mapping experiment and determine the sequence coverage achieved. Additional information can be provided by intensity colour coding.

Biotherapeutic molecules are extremely complex with an abundance of opportunities for PTMs and sequence variants. For this reason it is often necessary to delve deeper into data to identify these modifications, which are often present at low levels. Visualization tools such a chromatographic peak shading (Figure 13) can highlight co-eluting peptides ensuring that all variants are identified.

Conclusion

Peptide mapping will remain an essential step in biotherapeutic characterization. It is a powerful technique which provides important information at numerous stages within the biotherapeutic developmental process. The most fundamental requirement in any peptide mapping workflow is reproducibility, which enables users to confidently assign data differences to the sample and not the methodological conditions employed. Reproducibility is influenced by protein digestion, chromatographic separation, detector performance and linearity, and consistency in data handling. A comprehensive peptide mapping workflow with standardization of the various steps of the process improves peptide mapping reproducibility and increases analytical confidence.

References

- A. W. Vermeer and W. Norde. The thermal stability of immunoglobulin: unfolding and aggregation of a multi-domain protein. Biophysical Journal (78) 2000.394–404
- M. Samonig et al. SMART Digest compared to classic in-solution digestion of rituximab for in-depth peptide mapping characterization. 2016 AN72141-EN 0916S.
- 3. R.E. Moore 'Ultra' Biotherapeutic Characterisation. Innovations in

Pharmaceutical Technology. 2016. 32-37.

- R. Moore and M. Samonig. High-Throughput, High-Resolution Peptide Maps. GEN. Bioprocessing Tech Notes. (36) 2016. 12
- T. Köcher, R. Swart, and K. Mechtler. Ultra-High-Pressure RPLC Hyphenated to an LTQ-Orbitrap Velos Reveals a Linear Relation between Peak Capacity and Number of Identified Peptides. 2011, 83 (7). 2699–2704.
- U.D Neue and J.R Mazzeo. A theoretical study of the optimization of gradients at elevated temperature. J. Sep. Sci. (24). 2001. 921.
- 7. X. Wang, W.E. Barber and P.W Carr. A practical approach to maximizing peak capacity by using long columns packed with pellicular stationary phases for proteomic research. J. Chromatogr. (1107) 2006. 139.
- M. De Pra. High Resolution Peptide Mapping for Biopharmaceutical Analysis. 2014 TN71301-EN 0914S.
- M. Samonig et al. High-Throughput Peptide Mapping with the Vanquish UHPLC System and the Q Exactive HF Mass Spectrometer. 2015. AN71709-EN 0815S.

For more information on peptide mapping of biotherapeutic proteins please visit www.thermofisher.com/upgradeyourmaps