B05 - PEPTIDE MAPPING（Revision 1）

INTRODUCTION

Proteins can exist as large complex structures, with some molecules in the population displaying heterogeneity in their primary sequence due to improper assembly, degradation or post-translational modification. The high molecular mass of proteins combined with their complexity makes it particularly challenging to chemically identify an intact protein product using a single analytical method. It is possible to cleave the test protein into smaller fragments which can be identified with sufficient mass resolution to determine the primary sequence of the protein. This process is the basis of the protein identification technique commonly known as peptide mapping. The peptide mapping technique involves a digestion step in which the protein is selectively cleaved at amide bonds between specific amino acid residues to yield a predictable set of peptides. Analytical chromatographic separation, detection, and identification of the peptide mixture reveal information on the amino acid sequence of the protein which can be used to identify the protein. Peptide mapping is a comparative procedure; the results from the test protein are contrasted with the results of the reference standard or material similarly treated to determine the identity of the test protein. This comparative identification confirms that the primary structure of the test protein matches that of the reference protein.

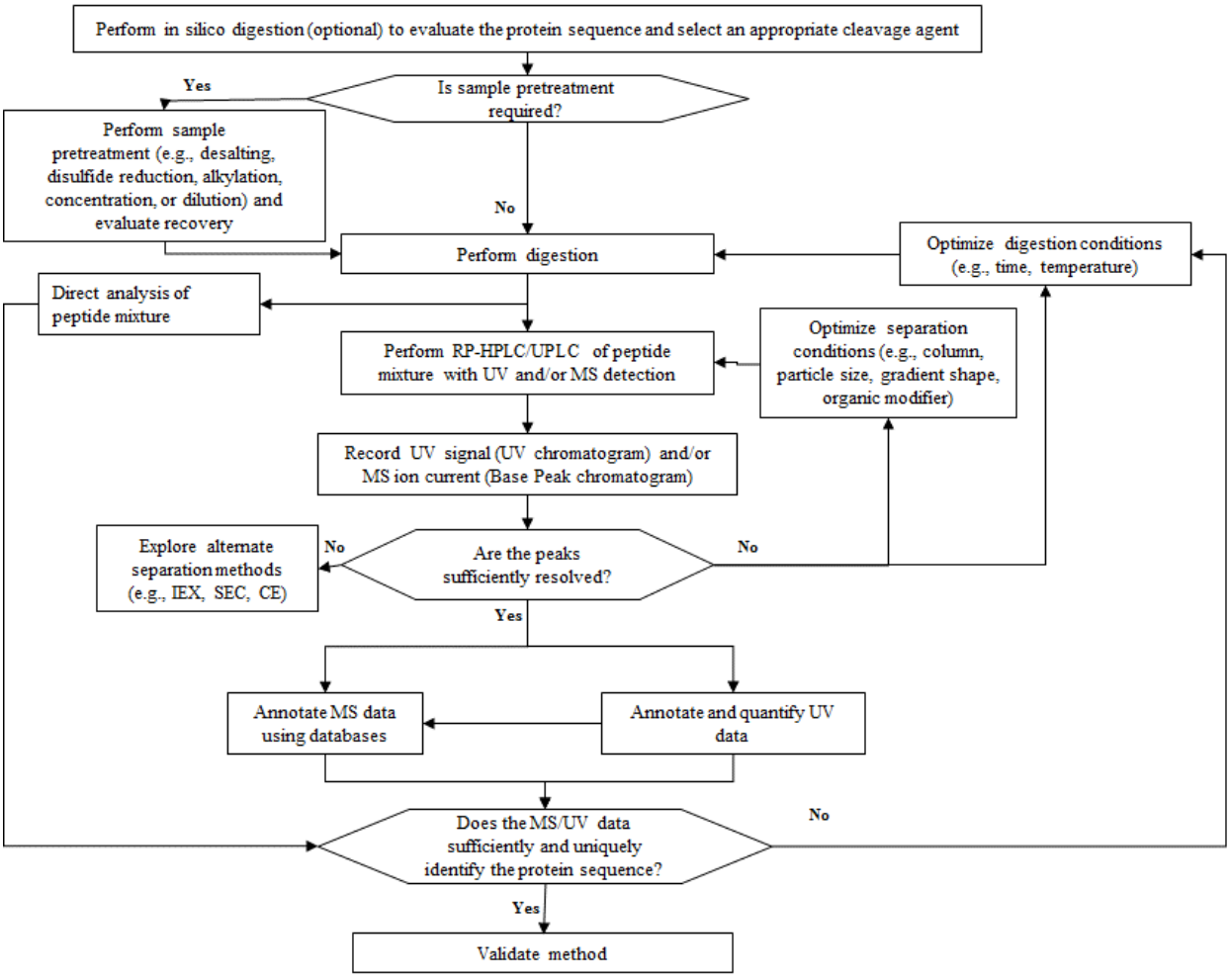
Peptide mapping’s ability to detect gross alterations in the primary sequence has resulted in many applications for the determination of protein quality which are outside the scope of this chapter. The purity of the test protein with regard to amino acid misincorporation or other misassembly such as disulfide bond scrambling, post-translational modifications, and degradation can be determined using a quantitative peptide map. Peptide mapping comparison during scale up or manufacturing changes can support studies of process consistency. Additionally, peptide mapping can be used to determine the degree and specific amino acid location of modifications such as glycosylation and conjugation (e.g., degree of pegylation).

The focus of this chapter will be on the use of peptide mapping for the chemical identification of a protein product where specificity is the primary attribute of the analytical method.

DEVELOPMENT OF A PEPTIDE MAPPING IDENTITY TEST PROCEDURE – POINTS TO CONSIDER

Prior to development of an identity test method procedure it is important to understand the application and level of specificity required to differentiate the identity of the test protein from other products processed in the same facility. In some instances orthogonal methods may be required to differentiate samples of structurally related proteins. Each protein presents unique characteristics that must be well understood so that the scientific approach used during development of the peptide map procedure will result in an analytical method that can be validated with sufficient specificity. The amino acid sequence of the test protein should be evaluated in order to select pretreatment and cleavage conditions resulting in optimal peptide length for analysis. Depending on application, complete or nearly complete sequence coverage is important, because there may be no prior knowledge of the alterations to the protein during development. The following points should be considered during development of a peptide mapping analytical technique. These elements are also presented graphically in *Figure 1*.

Figure 1. Identify Peptide Map Method and Target Performance Parameters



PRETREATMENT

Isolation and purification may be necessary for analysis of bulk drugs, dosage forms, or reference standards or materials containing interfering excipients or carrier proteins. Residual interfering substances may impact enzymatic cleavage efficiency and appearance of the peptide map. The impact of residual substances or the sample purification process on the final test peptide map should be assessed during the development process.

The tertiary structure of proteins may hinder full access of the cleavage enzyme to all cleavage sites resulting in unacceptable sequence coverage. The treatment of proteins

with chaotropic agents (e.g., guanidinium chloride, urea) and surfactants (e.g., sodium dodecyl sulfate) can be used to unfold the protein prior to digestion. Denaturing agents can affect enzyme activity and additional purification (e.g.diafiltration) or dilution steps may be needed prior to digestion. It may be necessary to reduce and alkylate the disulfide bonds prior to digestion in order to allow the enzyme to have full access to cleavage sites; however, the cysteine-to-cysteine linkage information is then lost. Common reagents for disulfide reduction include dithiothreitol and trialkylphosphine compounds such as tris(2-carboxyethyl)phosphine. Reagents for alkylating reduced cysteines include iodoacetamide, iodoacetic acid, and 4-vinylpyridine. The use of alkylating agents may create adducts which will impact the chromatographic separation and alter molecular weight of affected peptide.

Since peptide mapping is a comparative procedure; any purification or pretreatment steps performed on the test protein must also be performed on the product reference standard or material. The impact of residual substances, purification procedures, or pretreatment of the protein on method specificity and precision should be investigated during development and considered for inclusion in robustness studies conducted for method validation.

DIGESTION

The choice of a cleavage technique is protein-dependent. Some of the more common cleavage agents, both enzymatic and chemical, and their specificity are shown in *Table 1*. There may be specific reasons for using other cleavage agents or combinations of methods.

Table 1. Examples of Cleavage Agents

Type	Agent	Specificity
Enzymatic	Trypsin, EC 3.4.21.4	*C*-terminal side of Arg and Lys
	Chymotrypsin, EC 3.4.21.1	*C*-terminal side of hydrophobic residues (e.g., Leu, Met, Ala, aromatics)

Type	Agent	Specificity
	Pepsin A (Pepsin), EC 3.4.23.1	Low-specificity digest
	Lysyl endopeptidase (Lys-C endopeptidase), EC 3.4.21.50	*C*-terminal side of Lys
	Glutamyl endopeptidase (Glu-C endoproteinase; V8 protease); (from *S. aureus* strain V8), EC 3.4.21.19	*C*-terminal side of Glu and Asp
	Peptidyl-Asp metalloendopeptidase (Asp-N endoproteinase), EC 3.4.24.33	*N*-terminal side of Asp
	Clostripain (Arg-C endopeptidase), EC 3.4.22.8	*C*-terminal side of Arg
Chemical	Cyanogen bromide	*C*-terminal side of Met
	2-Nitro-5-thiocyanobenzoic acid	*N*-terminal side of Cys
	O-Iodosobenzoic acid	*C*-terminal side of Trp and Tyr
	Dilute acid	Asp and Pro
	3-Bromo-3-methyl-2-(2-nitrophenylthio-3*H*-indole (BNPS-skatole)	Trp

Factors that impact the effectiveness and reproducibility of protein digestion include pH, digestion buffer, temperature, time, and ratio of digest enzyme/reagent to protein. The optimal digestion mixture pH is generally determined by the enzyme or reagent, not the protein being analyzed. For example, a highly acidic environment (e.g., pH 2, formic acid) is necessary when using cyanogen bromide as a cleavage agent; however, a slightly alkaline environment (pH 8) is optimal when using trypsin as a cleavage agent.

The optimal temperature is dependent on the cleavage reagent; for example, most enzymes have optimum activity in a range of 25°–37°. The temperature can define the specificity of the enzyme to some extent. In these cases the adjustment of the temperature can be used to optimize the digestion conditions for certain proteins. Ideally, the digestion temperature will minimize sample-related chemical side reactions, such as deamidation, and protein aggregation while maximizing the susceptibility of the test protein to digestion while maintaining the activity of the cleavage agent.

It is necessary to ensure the digestion time is sufficient for intended use to avoid variable digests. A simple time-course study should be performed to ensure sufficient digestion with minimal peptide fragments resulting from partial digestion. Time of digestion varies from minutes to days and aliquots of a single reaction may be appropriately stabilized for analysis to determine the time required for complete digestion of the protein.

Sufficient cleavage agent should be used to attain the desired level of digestion within a practical time period (i.e., 2–20 h), while the amount of cleavage agent is minimized to avoid its contribution to the peptide map. For an enzymatic digest, the protein-to-protease mass ratio between 20:1 and 200:1 is generally used. In cases where the cleavage agent is unstable, the cleavage efficiency may be improved by making multiple additions of cleavage agent. Enzymes may be bound to a solid support to allow the use of higher relative amounts of protease while avoiding enzyme autolysis contamination and contribution of enzyme fragments to the peptide map. Chemical cleavage reagents are usually used in significant molar excess, and may need to be removed at the end of the digestion.

The optimal concentration of the test protein in the digestion should be empirically determined. The concentration should be low enough to minimize the potential aggregation of intact and partially digested proteins but must be sufficient to result in detection of all peptides following chromatographic separation with the selected detection method. Sample dilution or sample concentration by techniques such as centrifugal filtration may be required. Any dilution or concentration steps performed on the test protein must also be performed on the product reference standard or material. Protein recovery should be evaluated for any concentration step and the impact of dilution or concentration on method specificity and precision should be investigated during development and considered for inclusion in robustness studies conducted for method validation.

The digestion step can introduce ambiguities in the peptide map as a result of side reactions, such as nonspecific cleavage, deamidation, disulfide isomerization, oxidation of methionine residues, carbamylation of lysine residues, or formation of pyroglutamic

groups created from the deamidation of glutamine at the *N*-terminus of a peptide. Autolysis may introduce extraneous peaks produced by the proteolytic enzyme digesting itself. The intensities of autolysis peptide peaks are dependent on the enzyme to substrate ratio and the modifications and quality of the enzyme used. To avoid autolysis reagent solutions of proteolytic enzymes should be prepared at a pH which inhibits enzyme activity or the reagent solutions should be prepared immediately before use. Modified enzymes, where changes are made to the protease to prevent autolysis, may be used. Commercial preparations of trypsin (often called “proteomics grade”) are available in which the lysine residues of the enzyme have been methylated or acetylated to reduce the number of autolytic cleavage sites. To identify digestion artifacts, a blank determination is performed using a digestion control with all the reagents except the test protein.

SEPARATION

Chromatographic separation of the peptide mixture resulting from the digestion step is meant to resolve its complexity so that a valid interpretation of the data is meaningful and reproducible. The complexity of the peptide map will ultimately dictate the optimal set of chromatography conditions, column, and mobile phases. Method optimization experiments will be required to obtain the highest quality reproducible chromatogram. The molecular weight of the test protein will also influence the complexity of the map and the optimal separation.

Many techniques (e.g., ion-exchange high performance liquid chromatography [HPLC; ultra-high pressure liquid chromatography may also be suitable and as a subset of HPLC should be considered interchangeable with HPLC throughout this chapter], hydrophobic interaction HPLC, and capillary electrophoresis) have been used to separate peptides for peptide map analysis. However, reversed phase HPLC (RP-HPLC) is the most common method for the peptide mapping separation step and will be the focus of this chapter.

The selection of a chromatographic column is empirically determined for each protein. Columns with different pore sizes (80–1000 Å) or nonporous based on silica, polymeric,

or hybrid supports have been shown to give adequate separation. Columns with particle sizes <2 µm are available and are typically more efficient than those with 3–5 µm particle sizes. Generally, octyl or octadecylsilyl bonded phases are ideal for peptides. Octadecylsilane (C18) with 300 Å or smaller pores is the most commonly employed bonded phase for the peptide mapping separation step.

The most common mobile phase for the RP-HPLC separation of peptides is water with acetonitrile as the organic modifier; however other organic modifiers such as methanol, isopropyl alcohol, or *n*-propyl alcohol can be employed. Solvents such as the propyl alcohols in the mobile phase may be useful for separating samples that contain many highly hydrophobic peptides; however, it should be noted that hydrophilic or small peptides may possibly elute in a column void volume. Mobile phase additives such as acids, bases, buffer salts, and ion-pairing reagents are generally needed to produce high quality chromatographic separations of peptides. The most common mobile phase additive has been trifluoroacetic acid (TFA) with typical concentrations of 0.05%–0.2% being employed. The use of phosphate as an additive is less common but can be useful. Volatile acids and salts can be used in the mobile phase to improve compatibility with mass spectrometer detection. While TFA has a significant positive impact on the quality of peptide separation, sensitivity with mass spectrometer detection can suffer with TFA due to ion suppression. Formic acid, acetic acid, or combinations of these with TFA increase mass spectrometer sensitivity by reducing ion suppression. Temperature control of the chromatographic column is necessary to achieve good reproducibility. The column temperature may be used to optimize peptide separation or improve the retention or elution of certain peptides since the resolution typically increases with temperature for a reversed-phase column.

DETECTION

While RP-HPLC is the most common separation method employed with peptide mapping for identity testing, the most common detection method is ultraviolet (UV) light absorption at 214 nm. The peptides resulting from protein digestion may not contain amino acids with aromatic side chains that absorb light at higher wavelengths (e.g., 280

nm) so detection at 214 nm (i.e., wavelength where peptide bonds absorb light) is essential to ensure sequence coverage of the protein while minimizing background due to the mobile phase. Other detection methods may also be suitable.

The limitation of UV detection is that it provides no peptide structural information. Mass spectrometry is a useful detection method which provides mass information to aid in identification of peptides, as well as selectivity in cases when peptides co-elute. In most applications, the RP-HPLC effluent can be directly introduced into the mass spectrometer, provided that the mobile phase is compatible. Specific mobile phase considerations are dependent on the ionization method selected. Electrospray ionization (ESI) is the most common method for the introduction of proteins and peptides into the mass analyzer, and volatile, water-solvent mixtures provide the greatest ionization efficiency. Because ionization by ESI in the presence of solvent is limited to molecules more basic than the solvent, formic acid or acetic acid are commonly added to the mobile phase. Buffers and salts should be minimized since they can reduce signal, and nonvolatile salts can deposit in the source. As mentioned previously, TFA should be avoided because it can result in ion suppression, a type of matrix interference, which may reduce the signal of some peptides, particularly when ESI is used. Ion suppression may also reduce the ionization efficiency of glycosylated peptides, resulting in reduced sensitivity. It is thus important to test optimal combination to achieve optimal results for both UV and MS detection.

DATA ANALYSIS

Peptide mapping is a comparative procedure. To determine if the test protein is the desired protein of interest, the test protein's peptide map must be compared to the peptide map of the reference standard or material generated using identical pre-treatment, separation and detection procedures. Visual comparison of the retention times, the peak responses (the peak area or the peak height), the number of peaks, and the overall elution pattern is the first step of the procedure. It is a best practice to conduct a further non-subjective analysis of the peak response ratios of the critical peaks and the peak retention times. If all critical peaks in the test protein digest and in the reference standard or material digest have the same retention times and peak

response ratios, then the identity of the test protein is confirmed. For example, peptide mapping tests for monoclonal antibody samples often include a common Fc peptide that is used as a reference peak. This reference peptide can be spiked into the sample digest and then peak response ratios and retention times can be examined in comparison with the reference peak's predefined acceptance criteria. The method of comparison selected should depend on the complexity of the resulting peptide map and the specificity required for the particular identity test application (e.g. differentiation between different protein products manufactured at the same facility or differentiation of variants of the same protein product).

When high specificity is required, a mass spectrometer can be used for routine analyses to provide insight into peptide modifications, truncations, missed cleavages, impurities, and unresolved co-eluting peak(s) under a single peak.

POINTS TO CONSIDER PRIOR TO VALIDATION

During the development of the peptide mapping procedure, knowledge and experience are gained that lead to selection of system suitability criteria and analytical method validation acceptance criteria. A final review of the procedure prior to validation can ensure that the procedure is ready for validation, reducing risk of failure to meet criteria. As a general procedure, peptide mapping may encompass a significant range of experimental designs, applications, and requirements for performance. As a consequence, in a general text, it is not possible to set out specific system suitability or validation criteria. The following elements are suggested for evaluation prior to starting the validation.

It should be noted that the scope of this document does not include routine application of mass spectroscopic (MS)–based peptide mapping applications; however, the application of mass spectrometry for structural identification of peptides during the development of peptide mapping methods is a best practice. Mass spectrometric detection can be utilized to evaluate the following performance parameters.

Coverage

Coverage refers to the percentage of the target protein sequence which is recovered in the peptide map and identified as discretely resolved peaks. Although no specific figure can be identified for all applications, coverage approaching 95% has been found to be an acceptable performance target for a peptide mapping procedure.

Specific Bond Cleavages

The specific bonds cleaved by the chosen enzyme or chemical digestion procedure should be identified and listed.

Major Peaks

The major peptides recovered from the specific bond cleavages should be identified and listed.

Partial Cleavages

Peptide bonds susceptible to partial or incomplete cleavage and their associated chromatographic peaks or signals should be identified.

Minor/Non-specific Cleavages

The extent of cleavage at non-specific bonds should be identified and limited.

Protease-derived Peaks

If a protease is used for the test protein digestion then any peaks above background derived from the protease should be identified and, where appropriate, limited.

Undigested “Core” Protein

Undigested or partially digested protein (often called “core”) should be identified and limited.

Mean Peptide Length

Describes the peptide set produced by the combination of the chosen protease and/or chemical cleavage reagent and the test protein. This is a trade-off between smaller peptides, which show a higher level of structural selectivity with peptide mapping but produce a more complex map with more peaks, and longer peptides which produce simpler maps but with less resolving capacity for structural variants. No specific peptide length is suitable for all applications, but a mean peptide length of 10-20 residues is often considered appropriate.

Resolution Capacity

Resolution capacity refers to the capacity of the separation system to resolve the peptide set generated by the protease or chemical cleavage reagent. For example, a digest may produce 30 peptides but only 20 peaks due to co-elutions or nonrecoveries. Problematic separations should be identified and resolved by appropriate chromatographic procedures and, if necessary, controlled by the use of peptide reference standard or material or system performance criteria.

System Suitability Criteria Selection

System suitability criteria should be developed to ensure that the elements of the procedure for protein digestion, separation, and detection have successfully provided a structural identification of the test protein at the level of unambiguity required for the application. System suitability criteria evaluated during routine analysis for identity tests will typically include an assessment of the reference protein digest chromatogram and may include such performance characteristics as:

- Qualitative similarity to reference chromatogram
- Extent of digestion

- Partial cleavages
- Non-specific cleavages
- Peak heights/signal to noise ratio
- Peak shape
- Peak retention time
- Resolution of specific peaks

For test method procedures that require sample isolation, purification, or concentration a sample recovery criteria should be determined and included as part of the system suitability assessment. In cases where digestion artifacts may be present, assessment of a blank digestion control may be needed to demonstrate a lack of interference.

VALIDATION

Before validating a peptide mapping procedure, the procedure should have been developed to its final form and documented with system suitability criteria. Each time the procedure is performed the results are evaluated against the system suitability criteria to determine if the procedure has successfully provided reproducible results consistent with previous testing instances. Pre-approved acceptance criteria often evolve based on the system suitability criteria of the procedure. The elements of the analytical validation protocol are as follows:

Specificity

Method performance requirements will vary depending on the application of the identity test method and may require a risk assessment to understand what degree of specificity is needed to differentiate the identity of the test protein from other products processed in the same facility. Peptide mapping is a comparative technique confirming that the primary structure of the test protein matches that of the reference protein. Specificity is established by the comparison of the peptide maps of a suitable reference standard or material and samples of structurally related proteins. The selection of comparator samples should be selected based on a risk assessment of other products processed in the same facility and should be documented in the validation protocol. In order to minimize the inherent variability of the test, the procedure is executed on reference standard or material and test protein during the same testing instance. A peptide

mapping test design that analyzes the test protein digest, reference standard or material digest, and a 1:1 (v/v) comixture of the test protein and reference standard or material after digestion is a useful specificity validation experiment. Occasionally a peak can appear in a test protein's peptide map that elutes at a slightly different retention time then the corresponding peak in the reference standard or material peptide map, leading the analyst to judge the peaks as nonidentical. Testing a co-mixture sample during the specificity validation experiment can demonstrate that two peaks are identical if they co-elute in the co-mixture peptide map and confirm the identity. Chemically modified forms of the reference standard or material can be produced by exposure to conditions of pH, temperature, or chemical agents known to cause alteration of the primary structure. These alterations typically include deamidation of asparagine and glutamine residues, oxidation of methionine, histidine, or tryptophan residues, and acid catalyzed cleavage of peptide bonds. Peptide maps of a chemically modified reference standard or material and the reference standard or material can be compared based on predetermined acceptance criteria to demonstrate if the specificity of the peptide mapping procedure is affected by amino acid side chain modifications.

Precision

To facilitate the determination of the precision (repeatability and intermediate precision) of the peptide mapping procedure, an empirical method of quantifying peak responses (peak areas or peak heights) and peak retention factor should be part of the procedure. One approach is to make peak response and peak retention time comparisons that are expressed relative to a highly reproducible reference peak within the same chromatogram. The precision results obtained during the analytical procedure validation are reported and should meet the acceptance criteria of the validation. Failure of the precision results to meet the acceptance criteria can lead the analyst to reassess the digestion and/or separation steps in the procedure.

Robustness

Factors such as composition of the mobile phase, protease quality or chemical reagent purity, column variation and age, digestion temperature, and digest stability are likely to

affect the overall performance of the test and its reproducibility. Tolerances for each of the key parameters are evaluated and baseline limits established in case the test is used for routine lot release purposes.

The impact of small variations in purification, pretreatment, dilution, or concentration procedures of the protein sample on recovery should be identified during the development process and controlled. Impact of residual substances remaining after sample preparation on method specificity and precision should be considered. Critical parameters identified during development should be included in robustness studies conducted for method validation.

Many protein fragmentation strategies employ the use of proteolytic enzymes. As a result, the digestion portion of the peptide mapping procedure is inherently more sensitive to minor variation of test parameters. These parameters may include all or a subset of the following: digestion pH, buffer, buffer concentration, ionic strength, digestion temperature, digestion kinetics, test protein concentration, protease quantity, protease quality, and the stability of the digest. Using a design of experiments approach, the identified critical parameters are systematically studied to understand their impact on method variability. Those digestion parameters where small variations have been shown to impact the precision of the peptide mapping procedure should be carefully controlled within the test procedure using operating ranges established and validated by these studies.

To evaluate the protease quality or chemical reagent purity a sample of the reference standard or material is prepared and digested with different lots of cleavage agent. The chromatograms for each digest are compared in terms of peak areas, peak shape, and number. The same procedure can be applied to other critical chemicals or pretreatment procedures used during sample preparation, such as reducing and *S*-carboxymethylation reagents.

The length of time a digest can be held before proceeding to the separation step of the procedure, as well as the conditions under which the digest is stored before separation,

are assessed. Several aliquots from a single digest are stored under different storage conditions and resolved by the chromatographic method. These maps are then evaluated for significant differences.

During the separation step column-to-column variability, even within a single column lot, can affect the performance of the peptide mapping procedure. To evaluate column lot differences, reference standard or material of the protein of interest is digested and the digest is subjected to separation using different column lots from a single manufacturer. The resulting peptide maps are then evaluated in terms of the overall elution profile, retention times, and resolution according to predetermined acceptance criteria.

To evaluate the lifetime of a column in terms of robustness, a single digest of the reference standard or material can be analyzed using the peptide mapping procedure with columns that vary by the injection number history (e.g., 10 injections per column to 250 injections per column). The resulting peptide maps can then be compared for significant differences in peak broadening and overall resolution. As a column ages, an increase in back pressure might be observed that can affect the peptide map. System suitability or assay validity criteria can be designed to be diagnostic of column aging or other events that may affect the peptide mapping results.

SUMMARY

The peptide mapping procedure consists of multiple steps possibly including protein isolation, denaturation, chemical modification (e.g., blocking sulfhydryl groups) if necessary, protein digestion, peptide separation and detection, and data analysis. Each step should be optimized during development to result in a well qualified analytical procedure for the peptide mapping identity test. In combination with the use of a suitable reference standard or material, system suitability criteria should be chosen that evaluate if all the steps in the procedure worked together properly to produce a successful peptide map of that reference standard or material that is consistent with the validation of the analytical procedure. When properly developed, validated, and

performed the analytical peptide mapping procedure can be used to verify the identity of the test protein which is a critical quality attribute of the product.