Journal of Chromatography A, 1218 (2011) 1715–1723

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Protein digestion optimization for characterization of drug–protein adducts using response surface modeling☆

L. Switzar, M. Giera*, H. Lingeman, H. Irth, W.M.A. Niessen

Division of BioMolecular Analysis, Faculty of Sciences, VU University Amsterdam, De Boelelaan 1083, 1081 HV, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 15 October 2010
Received in revised form 8 December 2010
Accepted 11 December 2010
Available online 21 December 2010

Keywords:
Drug–protein adducts
Design of Experiments
Response Surface Methodology
Enzymatic digestion
LC–MS/MS

ABSTRACT

The formation of drug–protein adducts *in vivo* may have important clinical and toxicological implications. Consequently, there is a great interest in the detection of these adducts and the elucidation of their role in the processes leading to adverse and idiosyncratic drug reactions. Enzymatic digestion is a crucial step in bottom-up proteomics strategies for the analysis of drug–protein adducts. The chosen proteolytic enzyme and digestion conditions have a large influence on the protein coverage of the modified protein and identification of its modification site. In this work, the enzymatic digestion conditions (pH, temperature and time) of trypsin and thermolysin were optimized specifically for the characterization of Human Serum Albumin (HSA) adducts. Using a Design of Experiments (DOE), it was found that of the three optimized parameters mainly pH and temperature showed strong effects on both responses. The optimized digestion conditions were different from those obtained from the suppliers or literature. Their application to HSA adducts resulted in improved protein coverage and signal intensity regarding the peptide containing the modification site, thereby highlighting the importance of a detailed optimization of digestion conditions.

1. Introduction

Drug–protein adducts are suggested to play a role as mediators of Adverse Drug Reactions (ADRs) and Idiosyncratic Drug Reactions (IDRs) [1]. Therefore, their detection and identification is crucial within the framework of drug safety [2]. In the last decade, several Liquid Chromatography–Mass Spectrometry (LC–MS) based strategies have been developed for the determination of drug–protein adducts [3], i.e., the screening of reactive drug intermediates trapped by small molecules such as glutathione (GSH) [4–6] and proteomics based methods analyzing the adduct formed by a drug and its protein target [7–11]. The latter strategies are mostly based on enzymatic digestion of the modified protein followed by LC–MS(MS) analysis of the resulting proteolytic peptides. These approaches allow for the detection of clinically relevant drug–protein adducts and their simultaneous identification thereby giving insight into the mechanisms underlying ADRs.

Two major factors influencing the success of such methods are protein coverage, linked to successful identification of the modified protein, and the detection of the specific peptides that contain the modification site. The latter defines the actual sensitivity of the method and, naturally, achieving high protein coverage increases the chance of detecting the modified peptides. The most delicate step in this respect is the digestion of drug–protein adducts. It is not only critical to choose the appropriate enzyme, but also to apply the right digestion conditions, such as buffer pH, digestion temperature and time. Enzyme suppliers usually provide optimal conditions for the delivered enzyme. In addition, a wide range of digestion conditions obtained with different substrates are available from literature and enzyme databases such as BRENDA (http://www.brenda-enzymes.org/). For example, the optimal digestion conditions of bovine trypsin (EC 3.4.21.4) according to several suppliers are 2–18 h digestion time (depending on the amount of protein) at a temperature of 37 °C in 50 mM ammonium hydrogencarbonate or 100 mM Tris–HCl, pH 8.5. However, other optima can be found in the literature, such as overnight digestion at 37 °C in 50 mM ammonium hydrogencarbonate buffer pH 7.8 [12] and 45 min digestion at 37 °C in 10 mM ammonium hydrogencarbonate buffer pH 8.5 [13], while BRENDA displays an optimal pH range of 7.0–8.7 and an optimal temperature range of 45–74 °C. The wide variety in published digestion optima complicates the selection of the correct digestion conditions based on literature data. Furthermore, digestion conditions often are optimized for specific protein targets, such as monoclonal antibodies [14], polyclonal ovine immunoglobulin G [15] and membrane proteins [16,17], or specific applications, such as on-line bioreactors

☆ Presented at the 30th International Symposium on Proteins, Peptides and Polynucleotides (ISPPP), Bologna, Italy, 5–8 September 2010.

* Corresponding author. Tel.: +31 20 5987538; fax: +31 20 5987543.
E-mail addresses: M.Giera@few.vu.nl, mgiera@few.vu.nl (M. Giera).

0021-9673/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2010.12.043

[18] and are mostly focused on improving the peptide yield and protein identification rate. Taken together, this underlines the need for a detailed and systematic optimization of enzymatic digestion conditions for drug–protein adducts.

Optimization of chemical processes is traditionally carried out using a One-Variable-At-a-Time (OVAT) approach. Commonly, a limited number of OVAT experiments are carried out in which the levels of one variable are changed while the others are kept constant [19]. A major disadvantage of OVAT approaches is the disregard of interactions between variables. Therefore, this methodology often does not lead to the true optimum and may even lead to different end results depending on the starting point [20]. In order to avoid the local optima, more experiments need to be performed, which makes this approach more costly in terms of analysis time and consumption of chemicals [19]. In contrast, DOE techniques, such as the Response Surface Methodology (RSM), change combinations of variables simultaneously which does allow for incorporation of the interaction effects [21]. Another advantage of this technique is the concurrent optimization of multiple responses in order to find the optimal compromise between them. Additionally, RSM only requires a small subset of experiments from all possible variable combinations to cover the design space, which significantly reduces the number of necessary experiments. These advantages allow for a more efficient and more accurate determination of the optimum conditions.

In this study, a RSM approach was applied to the optimization of the three above mentioned conditions (buffer pH, digestion temperature and time) for digestion of HSA adducts with trypsin and thermolysin. These enzymes were selected because of their varying specificities and efficiencies [22]. HSA is the most abundant serum protein and often a target for reactive intermediates of drugs because of the free thiol on cysteine-34 (Cys34) [23]. A wide range of drugs, or their metabolites, including the *N*-acetyl-*p*-benzoquinoneimine (NAPQI) intermediate from acetaminophen [7] and several intermediates of diclofenac [3], are known to covalently bind to this site *in vivo*, thereby causing severe ADRs [1]. For the RSM optimization experiments, a model adduct was prepared by modification of HSA with monochlorobimane (MCB), which was selected for the simplicity of the adduct formation [24]. The two responses used to evaluate the optimization were the protein coverage of HSA and the peak area of the modified Cys34 peptide. For comparison, the digestion optima obtained from the RSM and selected literature conditions were applied to the digestion of NAPQI–HSA adducts.

2. Materials and methods

2.1. Reagents and materials

Human serum albumin (HSA), monochlorobimane (MCB), guanidine-HCl (G-HCl), ethanol, DL-dithiothreitol (DTT), iodoacetic acid (IHAc), thermolysin from *Bacillus thermoproteolyticus rokko* (EC 3.4.24.27), Tris–HCl, silver nitrate, sodium hydroxide, acetaminophen and the HPLC peptide standard mixture were purchased from Sigma Aldrich (Schnelldorf, Germany). Ammonium hydrogencarbonate, hydrochloric acid (HCl) 37% and diethyl ether were obtained from Riedel-de Haën (Seelze, Germany). Methanol, formic acid (FA), acetonitrile and chloroform came from Biosolve (Valkenswaard, The Netherlands). Trypsin from bovine pancreas (EC 3.4.21.4) was supplied by Roche (Almere, The Netherlands), acetone by Interchema (Oosterzee, The Netherlands), Bradford reagent by Biorad (Veenendaal, The Netherlands) and the synthetic peptide H-Pro-Pro-Pro-Pro-OH (Pro4) by Bachem (Weil am Rhein, Germany). Illustra NAP-25 gel-filtration columns with a bed volume of 2.5 mL, prepacked with G-25 DNA grade Sephadex, were obtained from GE Healthcare (Diegem, België). Water was purified by a Millipore (Amsterdam, The Netherlands) Milli-Q unit.

2.2. Design of Experiments

A RSM was applied for the optimization of digestion conditions of both enzymes with respect to the digestion of HSA adducts. A face-centered Central Composite Design (CCD) with uniform precision was created using JMP® 8.0.1 from SAS Institute Inc. (Cary, NC, USA). The CCD design was used to maximize two responses (protein coverage and peak area of the adducted Cys34 peptide) by optimization of three factors (buffer pH, digestion temperature and digestion time). The factor ranges were selected based on protease supplier's instructions and the BRENDA enzyme database. The applied factor ranges were pH 6–10, 24–50 °C and 1–12 h for trypsin and pH 5–9, 30–80 °C and 0.5–8 h for thermolysin. The complete DOE consisted of 40 randomized experiments per enzyme, including 6 center points and 1 replicate.

2.3. Sample preparation

2.3.1. Preparation of MCB–HSA adduct samples

For the RSM experiments, the MCB–HSA adduct was formed by adding a 50-fold molar excess of a 0.1 M solution of MCB in methanol to 5.5 mL of a 7.52 μM HSA solution in 50 mM ammonium hydrogencarbonate buffer pH 7.4. The reaction mixture was kept at 40 °C for 4 h after which the excess MCB was removed with a NAP-25 gel filtration column using 2 M G-HCl at pH 8.5 as the eluting buffer. The 35 cysteine residues of the denatured HSA were reduced by the addition of a 50-fold molar excess of 1 M DTT and alkylated using a 75-fold molar excess of 1 M IHAc. The reduced and alkylated MCB–HSA sample was split into three aliquots before being desalted using NAP-25 columns. As eluting buffers, three 50 mM ammonium hydrogencarbonate solutions were used with pH values corresponding to the three levels of the RSM design. A 100 μL aliquot of the desalted MCB–HSA was then digested with either trypsin or thermolysin (0.01 mg/mL in 0.1 mM HCl) using protein:enzyme ratios of 100:1 and 50:1, respectively. The enzymatic digestion was stopped with the addition of 10 μL of 10% FA. From a 12.3 μM internal standard (IS) solution of Pro 4 in water, 20 μL was added to the digested MCB–HSA samples to achieve a final concentration of 1.23 μM. The final volume of the samples was adjusted to 200 μL with water. The RSM experiments for trypsin and thermolysin were performed on different days and with different batches of the MCB–HSA adduct. A series of confirmation experiments were performed in triplicate to test whether the determined optimum digestion conditions lead to the predicted responses. These experiments were performed using the same batch of MCB–HSA for both enzymes, in order to guarantee comparability of the results.

2.3.2. Preparation of NAPQI–HSA adduct samples

The optimum digestion conditions obtained from the RSM of both enzymes were compared to literature conditions using the NAPQI–HSA adduct, which was prepared according to Hoos et al. [7]. This HSA adduct sample subsequently received the same treatment as described above for the MCB–HSA adducts applying either the optimum RSM digestion conditions or conditions obtained from literature. The selected literature values for trypsin digestion were taken from Aldini et al. [12] and consisted of overnight (13 h) digestion at 37 °C in 50 mM ammonium hydrogencarbonate buffer pH 7.8 and a protein:enzyme ratio of 20:1. For thermolysin, the reference digestion conditions were obtained from Bark et al. [25] and consisted of 15 min digestion at 65 °C in 100 mM ammonium hydrogencarbonate buffer pH 7.5 and a protein:enzyme ratio of 50:1.

2.4. LC–MS/MS analysis of digested HSA adduct samples

The digested HSA samples were analyzed with a 1200 series Rapid Resolution LC system coupled to a 6520 QTOF mass spectrometer (Agilent, Amstelveen, The Netherlands), that was controlled by the Agilent Masshunter Workstation Acquisition software (version B.02.00). The proteolytic peptides were separated on an Agilent XDB-C18 column (4.6 mm × 50 mm, 1.8 μm particles) that was protected by a guard column (4 mm × 2 mm) from Phenomenex (Utrecht, The Netherlands). Mobile phase A consisted of 5% acetonitrile and 0.1% FA in water, mobile phase B consisted of 95% acetonitrile, 5% water and 0.1% FA. The flow rate was set at 0.6 mL/min and the thermostated column compartment was maintained at 40 °C. Gradient elution was performed as follows: 0% B for the first 2 min, linearly increased to 40% B in 23 min, then set to 100% B and held constant for 4 min, followed by a re-equilibration at 0% B for 9 min. Using an internal switching valve, the LC flow from 2 to 25 min was directed to the mass spectrometer, which was operated in 2 GHz, extended dynamic range mode. The electrospray ionization source was operated in positive mode (ESI+), the capillary voltage was set to 3500 V and nitrogen (99.9990%) was used as the drying (350 °C) and nebulizer gas at a flow rate of 12 L/min and a pressure of 60 psig, respectively. Profile data were acquired in data-dependent mode where the most intense ion (m/z 200–3000) was selected for fragmentation and subsequently excluded from fragmentation for 0.2 min. MS/MS spectra were recorded from m/z 50 to 3000 at a rate of 1.02 spectra/s using a fixed collision energy voltage of 20 V and nitrogen was used as the collision gas. A blank sample and the HPLC peptide standard mixture were analyzed alternatively after every four sample runs to check the stability of the LC–MS system throughout the 45-h sequences.

2.5. Data analysis

Peak extraction (using a 20 ppm half-width m/z window) and integration was performed with Agilent Masshunter Qualitative Analysis software (version B.02.00). The peak areas were normalized to the IS. Mascot Distiller (version 2.3.2) and Mascot server (version 2.2) (Matrix Science Ltd., London, UK, www.matrixscience.com) were used for peak picking of the data, identification of the peptide sequences and calculation of the protein coverage. General peak picking settings, such as a minimum precursor mass of 500 Da, maximum precursor mass of 16,000 Da, a maximum intermediate scan count of 2, a signal-to-noise (S/N) of 10, were the same for all enzymes. The only enzyme-dependent setting was the default precursor charge range, which was set to 1–5 charges for tryptic digests and to 1–4 charges for thermolytic digests. The obtained peak lists were searched against the SwissProt database with carboxymethyl(cysteine) and chlorobimane(cysteine) (monoisotopic delta mass of 190.074 Da) or NAPQI(cysteine) (monoisotopic delta mass of 149.048 Da) selected as variable modifications. The peptide and MS/MS tolerance was set to 0.2 Da and the number of allowed missed cleavages was set to 3 for tryptic digests and to 4 for thermolytic digests. The cleavage definitions of thermolysin (*N*-terminal to A, F, I, L, and V) were determined experimentally (data not shown) and added to the enzyme database manually. Mascot peptide summary reports were formatted using an ion score cut-off value of 20 to remove random peptide matches and requiring a protein hit to include at least one top-ranking peptide match. Statistical evaluation of the RSM data was done with ANOVA functions embedded in JMP® 8.0.1. The optimum digestion conditions for maximization of both responses were determined using the "Maximize Desirability" function in the prediction profiler.

3. Results and discussion

3.1. Method improvements for RSM experiments

A DOE was used for efficient optimization of a multi-parameter analytical method. A RSM significantly reduces the number of experiments, which is advantageous in terms of instrumental analysis time and sample handling. Depending on the number of variables to be optimized, a predefined number of experiments must be performed for each enzyme and this number increases rapidly with each added variable. For a replicated RSM design with three variables, the number of experiments amounts to 40 and with four variables to 62. Critical and time-consuming steps in such a procedure are the preparation of the samples, the time needed to analyze them, and the time needed for data interpretation and processing. Robotic sample preparation would simplify the first step, whereas the sample analysis can be performed in unattended overnight experiments. With this in mind, it is imperative that the general analytical procedure is sufficiently well developed and optimized with respect to the parameters that are not evaluated in the DOE. Otherwise, no reliable data may be obtained within the series of experiments. Therefore, prior to starting with the RSM experiments, a number of parameters were optimized and considered, such as the chromatography, the use of an internal standard, sample handling steps prior to the actual digestion (denaturation, reduction, alkylation and removal of reagents involved in these steps).

3.1.1. LC–MS separation and detection

Protein digests are complex samples containing a multitude of proteolytic peptides. Successful identification of the protein under investigation relies on the acquisition of high-quality MS/MS spectra, which in turn depends on the resolution of the preceding LC separation step. Separation of protein digest samples with conventional LC systems requires long gradient runs, often exceeding 90 min [14,26], in order to achieve a satisfactory chromatographic resolution. However, it is preferable to complete the analysis of all samples from a DOE within a reasonable time in order to avoid effects of extraneous factors. The use of Ultra High Performance Liquid Chromatography (UHPLC) or Rapid Resolution Liquid Chromatography (RRLC) systems with analytical columns containing sub-2 μm particles offer substantial benefits over conventional LC separations in terms of speed and resolution. Using this technology, a fast and high-resolution separation method with a total run time of 38 min was developed for the separation of HSA digests, which is significantly shorter than achieved with lower-resolution instruments. The increased separation speed resulted in narrow peaks with peak widths at half height ≤0.2 min. The MS settings were, therefore, adjusted to ensure the collection of sufficient data points to accurately detect each peak to enable appropriate peak area determination for the Cys34 adducted peptide, and to obtain sufficient MS/MS data for Mascot database searches. This was achieved by selecting only the highest-intensity ion from the full scan spectrum as the precursor ion for fragmentation. Subsequent exclusion of this ion for 0.2 min allowed for selection of lower-intensity precursor ions.

3.1.2. Selection of the internal standard

Since one of the measured responses of the DOE was the peak area of the MCB–Cys34 peptide, it is essential to normalize the obtained peak areas. For correct mapping of the behavior of the studied analyte, the most appropriate IS for application in proteomics experiments is a synthetic peptide with an amino acid sequence that cannot be attributed to the protein that was digested. Considering the nature of this study, a synthetic peptide had to be selected taking into account the cleavage sites of the proteolytic

enzymes for which the digestion conditions were optimized. Even though the IS was added after stopping the enzymatic reaction, residual enzyme activity could cleave the IS and thereby negatively influence peak area normalization. The best option for an IS was found to be a peptide consisting solely of proline residues, as this is not a cleavage site for the investigated enzymes.

3.1.3. Sample preparation

With the fast LC–MS/MS method in place, reduction of sample handling steps was investigated next. From a series of preliminary experiments (data not shown), it was found that denaturation, reduction and alkylation of the protein were crucial for obtaining a high digestion efficiency. Deletion of any of these steps from the sample preparation protocol resulted in a significant decrease in the protein coverage. Moreover, failure to remove the applied denaturation, reduction and alkylating agents prior to enzymatic digestion negatively influenced the enzyme activity. Removal of the excess modifying agent after adduct formation is also necessary to prevent the formation of other cysteine adducts during the reduction of the protein S–S bonds. Gel-filtration columns were used for these sample clean-up steps because of their compatibility with a wide range of common buffers, including G-HCl. Therefore, denaturation and removal of the excess modifying agent could be combined in a single step by using 2.0 M G-HCl as the eluting buffer. A 6.0 M G-HCl buffer (pH ≈ 6) is commonly used for denaturation of proteins, but the high salt concentration was found to block the gel-filtration columns and reduced their lifetime. Using a lower concentration of the G-HCl buffer did not noticeably affect the denaturation of the adducted protein. Increasing the pH value of the G-HCl buffer to pH 8.5 facilitated the subsequent reduction (DTT is active at pH > 7) and alkylation reactions (IHAc is selective for cysteines at pH 8.5). The alkylation process was further improved by adding a larger excess of alkylating agent over reducing agent. Previously, a 50-fold molar excess of both reagents was added, in sequence, to the denatured protein, but the presence of reducing agents during the alkylation step quenches the alkylating agent. It was observed that under these conditions all S–S bridges were reduced, but less than 35% of the detected cysteines were alkylated. By increasing the molar excess of alkylating agent to 75-fold (1.5 × the excess of reducing agent), the number of detected alkylated cysteine residues increased to >90%. The denaturation, reduction and alkylation were followed by a desalting and simultaneous solvent exchange step to the appropriate digestion buffer using the gel-filtration columns.

3.2. RSM

The above described analytical procedure was developed to facilitate the optimization of the enzymatic digestion conditions of trypsin and thermolysin. The varying specificities and efficiencies of these enzymes result in characteristic peptide profiles and corresponding protein coverages. While trypsin generally displays high protein coverage, it has only two cleavage sites and, therefore, produces larger peptide fragments. When using less than optimum digestion conditions, the increased occurrence of missed cleavages may prevent detection of these large peptides, resulting in decreased coverage. On the other hand, thermolysin has a broader specificity, having five main cleavage sites, and produces peptides with smaller chain lengths that are easily detectable. However, when the chain lengths become too small (less than 5 amino acids) elucidation of the amino acid sequence becomes more difficult, which, in turn, also leads to decreased protein coverage. Therefore, the digestion conditions of the above mentioned enzymes were optimized to maximize the coverage. The three factors to be optimized, buffer pH, digestion temperature and time, were chosen based on their influence on the digestion efficiency. Another advantage of applying a DOE is the possibility to optimize multiple responses simultaneously. Generally, protein coverage is the only response that is optimized in protein digestion experiments, but in this particular case detection of a specific part of the protein is essential, namely the site of adduct formation. In this respect, sensitivity for detection of the modification site is paramount since drug–protein adducts are often low-abundant. Therefore, the peak areas of the various modified Cys34 peptides generated by the two enzymes were chosen as the second response for which the digestion process was optimized. After optimization of the digestion conditions using the MCB–HSA adduct, the obtained optimum digestion protocols of the two enzymes were compared by application to a clinically relevant drug–protein adduct to determine the best candidate for identification of HSA adducts.

3.2.1. Trypsin RSM results

Trypsin is a widely applied proteolytic enzyme because of its high efficiency and specificity, cleaving predominantly at the *C*-terminal side of lysine and arginine residues. Under optimal conditions, tryptic digestions are characterized by a low number of missed cleavages and high protein coverage, which was also observed in the RSM experiments. In general, trypsin performed well under a wide range of conditions producing a protein coverage ranging from 81 to 93%, except when a combination of high pH and high temperature was used. These extreme conditions significantly reduced the digestion efficiency, regardless of the length of digestion time (1–12 h), which decreased the protein coverage to ≤48%. This is further exemplified by the deviating chromatographic peptide profile as compared to profiles obtained using less extreme digestion conditions (Fig. 1).

Due to the reduced digestion efficiency, the MCB–Cys34 peptide was not detected in these incomplete tryptic digestion samples. In the remaining 36 experiments, the MCB–Cys34 peptide was identified predominantly without missed cleavages, while the other peptides identified by Mascot exhibited 0, 1 or 2 missed cleavages. The amino acid sequences of the identified tryptic MCB–Cys34 peptides can be found in Table 1. The sequence coverage of the MCB–Cys34 peptide without missed cleavages was 71% allowing for accurate identification of the modification site.

Evaluation of the RSM results showed that pH and temperature influenced both responses. Fig. 2a shows a surface plot of the factors pH and temperature for the response protein coverage. It reveals that the digestion pH should be decreased accordingly when a higher digestion temperature is applied and vice versa, in order to achieve similar protein coverage. The highest predicted protein coverage (96%) is obtained when performing the digestion at pH 6 and a temperature of 47 °C. The factor time by itself did not seem to have an influence on either of the responses. However, the surface plot of the effect of time and temperature on the peak area of the MCB–Cys34 peptide shows that a combination of a low digestion temperature and long digestion time leads to the highest peak area of this peptide. Protein coverage, on the other hand, would benefit from a higher digestion temperature, but digestion at elevated temperatures significantly decreases the peak area of the MCB–Cys34 peptide, which would mean a loss in sensitivity of the method. Therefore, a compromise has to be made to obtain the optimal digestion conditions for both responses.

Using the "Maximize Desirability" function of the prediction profiler, the most desirable factor settings to maximize both responses simultaneously were obtained through an iterative process. The prediction profiler in Fig. 3 shows that for this specific optimization experiment the optimum tryptic digestion conditions are digestion at pH 8.4 and 24 °C for 11.6 h. The predicted maximum response values for these conditions are a protein coverage of 91% and a peak area of the ALVLIAFAQYLQQC(MCB)PFEDHVK peptide of 7.72×10^5.

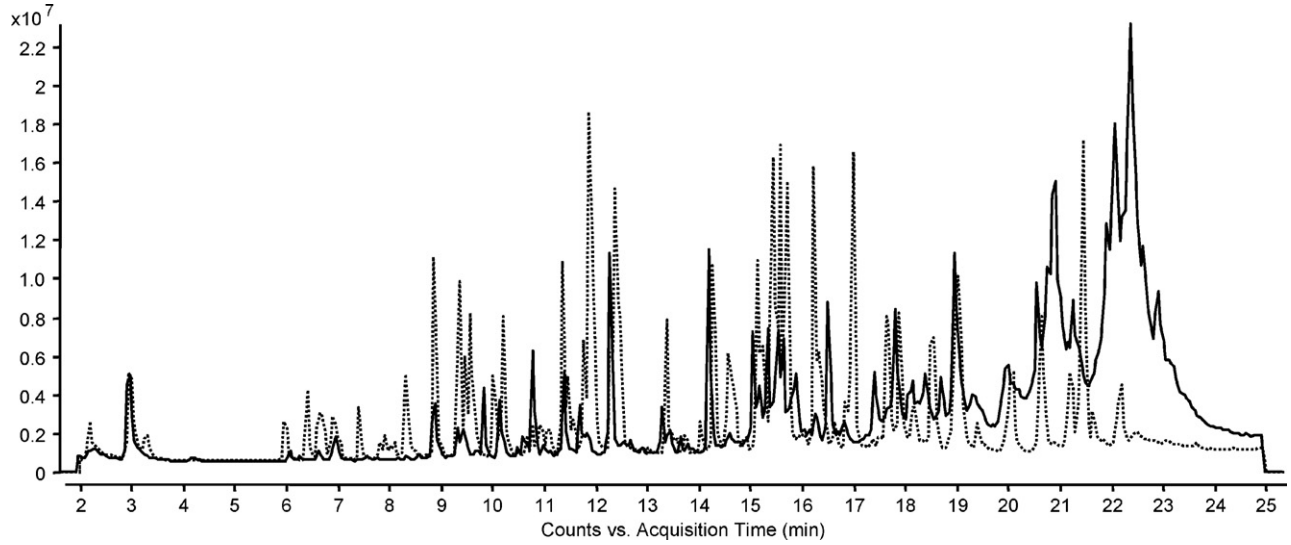

Fig. 1. LC–MS chromatograms showing the peptide profiles that are obtained through tryptic digestion of MCB–HSA under different conditions. Solid line: digestion at pH 10, 50 °C for 12 h. Dotted line: digestion at pH 8, 37 °C for 6 h.

Table 1
Sequences of the detected MCB–Cys34 peptides obtained through tryptic or thermolytic digestion of MCB–HSA.

Enzyme	MCB–Cys34 peptide	# Missed cleavages	m/z $[M+nH]^{n+}$	Sequence coverage (%)
Trypsin	ALVLIAFAQYLQQC[a]PFEDHVK	0	875.127^{3+}	71
	ALVLIAFAQYLQQC[a]PFEDHV*K*LVNEVTEFAK	1	939.276^{4+}	26
Thermolysin	LQQC[a]P	0	*nd*	–
	LQQC[a]PFEDHV	1	653.797^{3+}	89
	LQQC[a]PFEDHV*K*	2	511.926^{4+}	100
	LQQC[a]P*F*EDHV*KL*	3	549.618^{4+}	50

nd = not detected
[a] Indicates the MCB–Cys34 modification site. Amino acids in italics represent missed cleavages.

The optimum pH and time values are comparable with the manufacturer's instructions, but the found optimum temperature is lower than expected. The discrepancy between the RSM optima and those indicated by the manufacturers could be explained by the fact that the conditions for optimal enzyme activity are often determined using relatively small substrates. Judging by the wide range of trypsin digestion optima applied to different substrates published in the literature, the enzyme activity optima cannot be directly extrapolated to every substrate or analytical system. In addition, protein digestion is generally evaluated by protein coverage and the use of a second response in this study may also lead to different optimum conditions.

Although the protein coverage remained reasonably stable and the MCB–Cys34 peptide could be identified under most digestion conditions, the peak area of this peptide changed significantly with varying conditions. Therefore, increasing the temperature to 37 °C

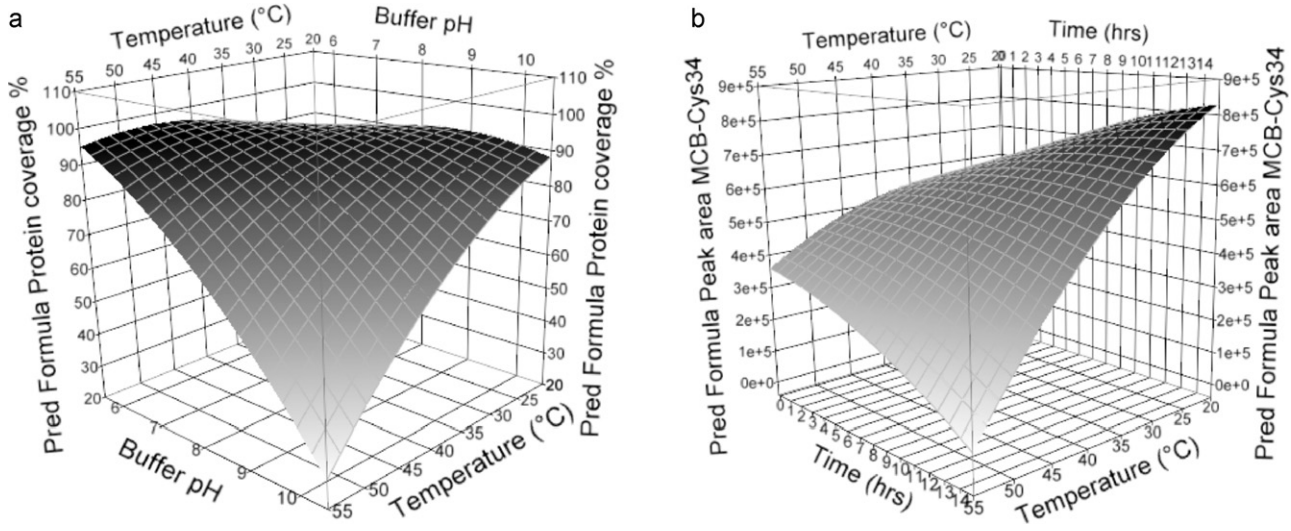

Fig. 2. Surface plots showing the effects of buffer pH and digestion temperature on protein coverage (a) and temperature and time on the peak area of the MCB–Cys34 peptide (b).

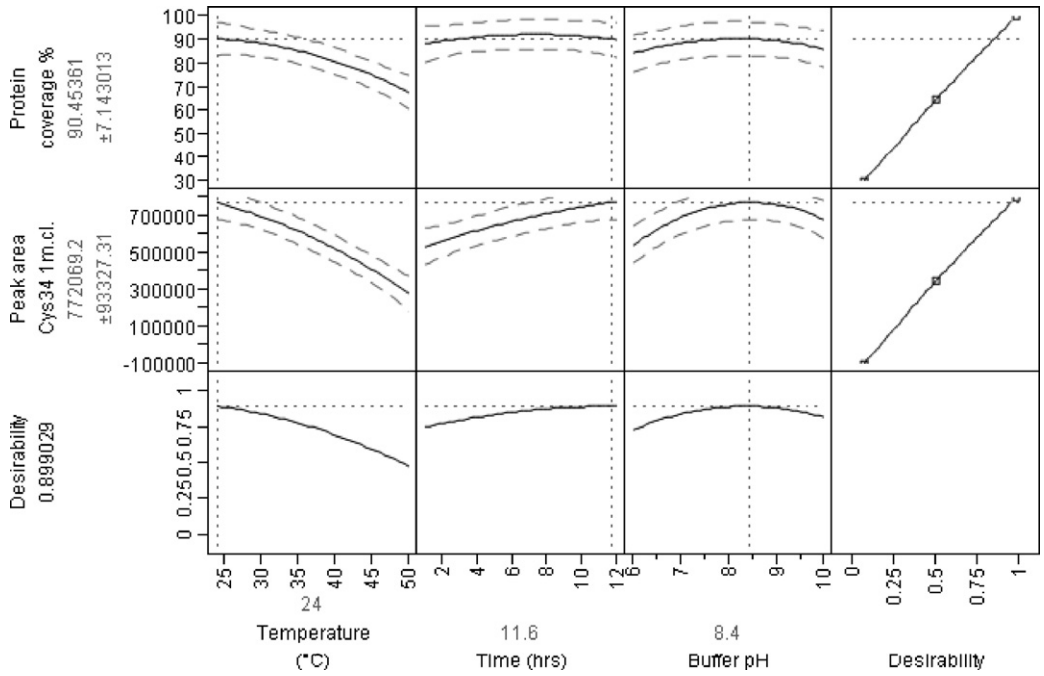

Fig. 3. Prediction profiler plot showing the optimum tryptic digestion conditions and the predicted responses for the complete model under these conditions. The dashed lines indicate the 95% confidence intervals, the dotted vertical lines indicate the optimum level for each of the factors and the dotted horizontal lines show the predicted value of both responses ± their standard deviations.

would lead to a significant loss in peak area of the MCB–Cys34 peptide. Additionally, the prediction profiler also showed that a shorter digestion time would lead to a slight increase in protein coverage. However, this would negatively affect the peak area of the MCB–Cys34 and, thus, is not advantageous for the whole model. Furthermore, a significant decrease in MS signal was observed during the long LC–MS sequence of the 40 trypsin RSM samples. A possible explanation for this observation could be the analysis of incomplete tryptic digests obtained under less than optimal conditions. In general, trypsin produces larger peptide fragments and with an incomplete digestion these even larger fragments may pollute the ion source.

Statistical evaluation of the RSM by Analysis of Variance (ANOVA) showed that a well-fitted model ($R^2 > 0.8$ [27]) was obtained for both responses. The R^2 values for protein coverage and peak area of the MCB–Cys34 peptide were 0.859 and 0.894, respectively, indicating that the model could explain 85.9% and 89.4% of the variation in the respective responses. Additionally, *F*-tests revealed that the regression for both responses (*F*-values of 20.34 and 28.04) was statistically significant at a confidence level of >99.9% and that an insignificant proportion of the pure error is explained by variation due to lack of fit (*F*-values of 30.41 and 5.03 with *P*-values of 0.0001 and 0.0025). The centerpoint experiments showed limited variation for protein coverage with a %RSD of 2.4% and slightly more for the peak area of the MCB–Cys34 peptide with a %RSD of 8.1%. The predictive quality of the model was exemplified by the confirmation experiments performed at the optimum tryptic digestion conditions. An average ($n = 3$) protein coverage of 86 ± 1% and a peptide peak area of $7.12 \times 10^5 \pm 15\%$ were obtained, which is in good agreement with the predicted responses.

3.2.2. Thermolysin RSM results

Thermolysin is a thermostable enzyme that is able to withstand elevated temperatures. The optimum digestion temperature, according to the supplier, is 70 °C at a pH value of 8.0, but an optimal digestion time is not supplied and BRENDA does not contain information about the digestion time either. However, thermolysin is suggested to exhibit an accelerated rate of reaction at elevated temperatures [25]. Therefore, a wider temperature range of 30–80 °C and shorter digestion times of 0.5–8 h were chosen for this design. Table 2 shows the complete design and the obtained response values. From this table, it can already be seen that the results can be clustered into three groups consisting of low, intermediate and high response values. Especially for the low values, the response can be correlated to the application of a specific combination of conditions, in this case high pH and high digestion temperature.

As anticipated, the peptide profile resulting from thermolytic digestion of MCB–HSA is different from that obtained with tryptic digestion. The LC–MS chromatograms shown in Fig. 4 contain a higher number of peptides due to the lower specificity of this enzyme and a larger number of missed cleavages.

Due to the higher number of missed cleavages, several MCB–Cys34 peptides with different chain lengths were detected (see Table 1). Although, for this enzyme, a proline residue in this position should not interfere with the cleavage, the shortest possible peptide with a sequence of LQQCP was not detected. This could be due to the fact that the MCB adduct is located on the cysteine next to the proline or because of the digestion efficiency of this enzyme. Three other MCB–Cys34 peptides with an increasing number of missed cleavages, LQQCPFEDHV, LQQCPFEDHVK and LQQCPFEDHVKL, were detected. Of these three, the MCB–Cys34 peptide with two missed cleavages (LQQCPFEDHVK) had the highest intensity and sequence coverage of 100% allowing for unambiguous identification of the modification site. Therefore, this peptide was used as the second response of the RSM, besides the protein coverage.

Similar to trypsin, the application of extreme conditions in experiments 6, 12, 37 and 40 led to a low protein coverage of <15% and failure to detect the MCB–Cys34 peptide. Although the protein coverage in the remaining experiments, ranging from 44 to 74%, was lower than that obtained with trypsin, still high-confidence identification of the target protein could be achieved. Again, the main parameters temperature and pH proved to have an effect on the response values, as shown in Fig. 5a for the peak area of the MCB–Cys34 peptide. This surface plot also shows that there is a

Table 2
RSM design and experimental data of thermolysin digestion optimization.

Exp. #	pH	°C	Hours	Protein coverage (%)	Peak area MCB–Cys34 ($\times 10^6$)
1	7[a]	55	4.25	64	1.12
2	7	30	4.25	62	1.18
3	7[a]	55	4.25	69	1.18
4	5	80	0.5	47	0.17
5	7	80	4.25	67	0.54
6	9	80	8	9	0.00
7	7[a]	55	4.25	62	1.04
8	9	30	8	67	1.14
9	5	30	0.5	62	0.43
10	5	80	0.5	52	0.20
11	5	30	8	46	0.44
12	9	80	0.5	15	0.00
13	5	55	4.25	59	0.73
14	7[a]	55	4.25	64	1.00
15	7	30	4.25	68	1.03
16	9	30	0.5	54	0.50
17	7[a]	55	4.25	61	1.19
18	7	55	0.5	74	1.22
19	7[a]	55	4.25	64	1.07
20	9	30	8	73	1.25
21	5	80	8	51	0.21
22	7	55	8	66	1.04
23	9	30	0.5	59	0.48
24	7[a]	55	4.25	64	1.01
25	5	30	8	44	0.44
26	9	55	4.25	51	0.42
27	7[a]	55	4.25	65	1.14
28	7	55	0.5	66	1.26
29	7[a]	55	4.25	70	1.07
30	9	55	4.25	51	0.40
31	7[a]	55	4.25	61	1.13
32	7[a]	55	4.25	64	1.16
33	7	80	4.25	52	0.38
34	5	80	8	51	0.19
35	5	55	4.25	62	0.80
36	7[a]	55	4.25	57	1.01
37	9	80	0.5	11	0.00
38	7	55	8	65	1.05
39	5	30	0.5	62	0.63
40	9	80	8	8	0.00

[a] Center point.

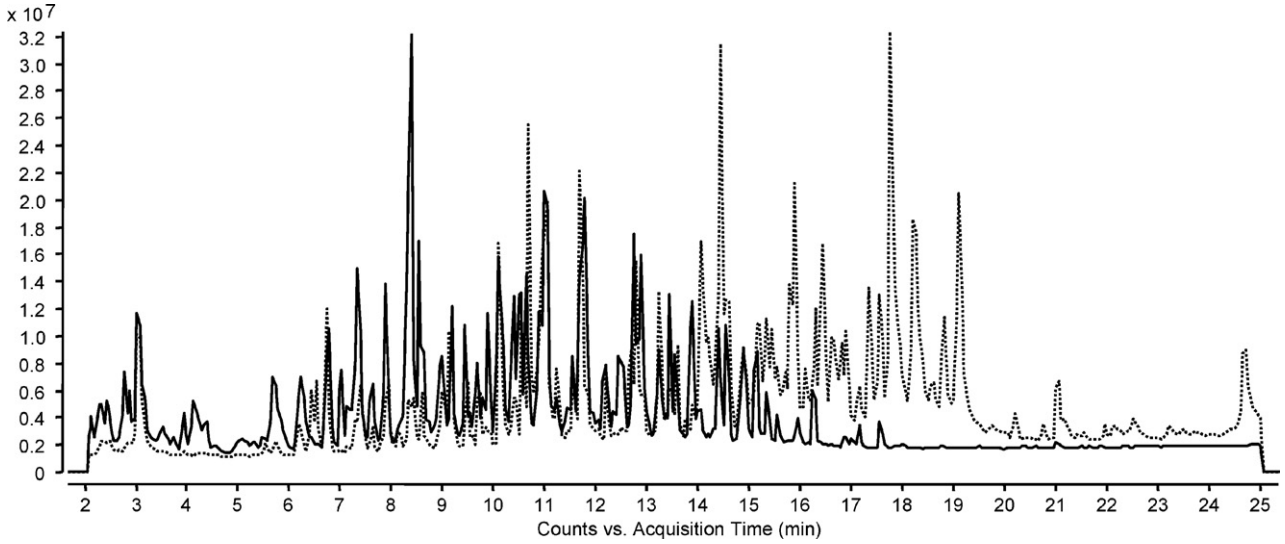

Fig. 4. LC–MS chromatograms showing the peptide profiles that are obtained through thermolytic digestion of MCB–HSA under different conditions. Solid line: digestion at pH 7, 55 °C for 4.25 h. Dotted line: digestion at pH 9, 30 °C for 0.5 h.

true optimum temperature and buffer pH for achieving maximum peak area, which is at pH 7.4 and 38 °C. Somewhat surprisingly, as shown in Fig. 5b, there was little effect of time on protein coverage for this specific system. Moreover, lower, rather than elevated temperatures resulted in the highest obtained protein coverage.

Although the factor 'time' did not have a large effect on either of the two responses (Fig. 6), the highest peak area of the MCB–Cys34

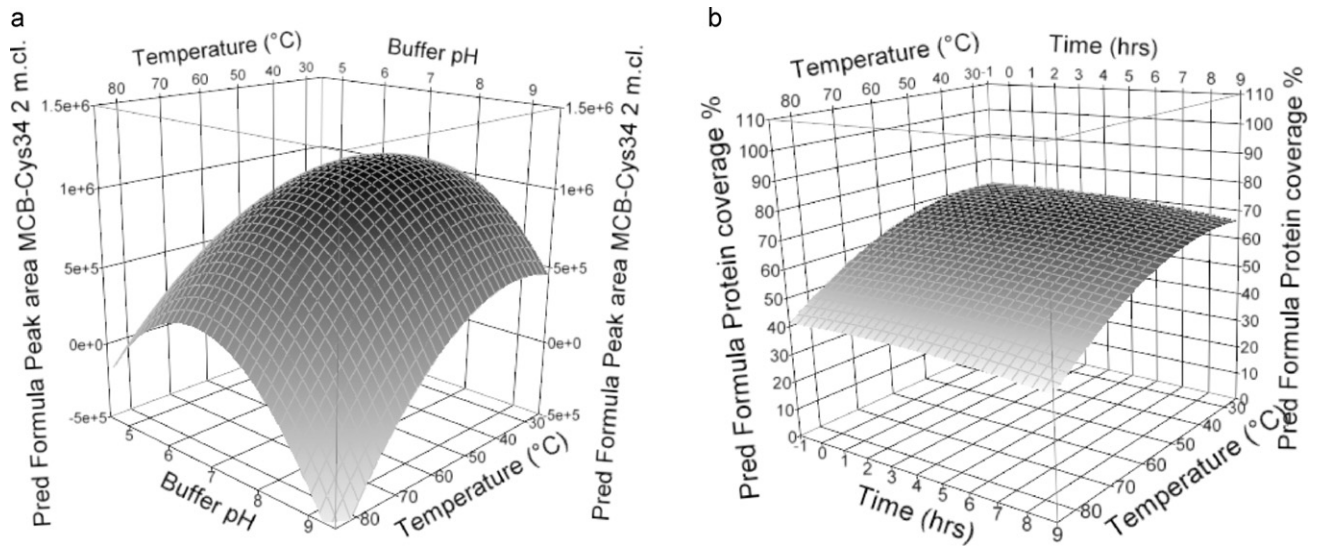

Fig. 5. Surface plots showing the effects of buffer pH and digestion temperature on the peak area of the MCB–Cys34 peptide (a) and temperature and time on the protein coverage (b).

peptide was achieved after 8 h of digestion. Slightly different optimum conditions were obtained for each of the responses separately, but the optimum conditions for the whole model were digestion at pH 7.5 and 34 °C for 8 h. The predicted protein coverage and peak area of the MCB–Cys34 peptide at these conditions were 69% and 1.32×10^6, respectively. As mentioned above, the optimum digestion temperature is much lower than expected based on supplier's instructions and literature, and a longer digestion time is preferred. The optimum pH also does not concur with the provided instructions, but was an exact match with the literature pH value.

Statistical analysis was also performed on the thermolysin data revealing that the RSM model fitted well, represented by R^2 values of 0.885 and 0.926 for protein coverage and peak area of the MCB–Cys34 peptide, respectively. Similar to the trypsin RSM model, *F*-tests at a >99.9% confidence level indicated that the regression for both responses was statistically significant (*F*-values of 25.60 and 41.48) and that an insignificant proportion of the pure error is explained by variation due to lack of fit (*F*-values of 13.908 and 19.273). The centerpoint experiments showed limited variation for both responses with %RSD values of 5.4% for protein coverage and 6.3% for peak area of the MCB–Cys34 peptide. The confirmation experiments resulted in a protein coverage of $62 \pm 2\%$ and a peak area of $1.00 \times 10^6 \pm 3\%$, which is in good agreement with the RSM optima as well.

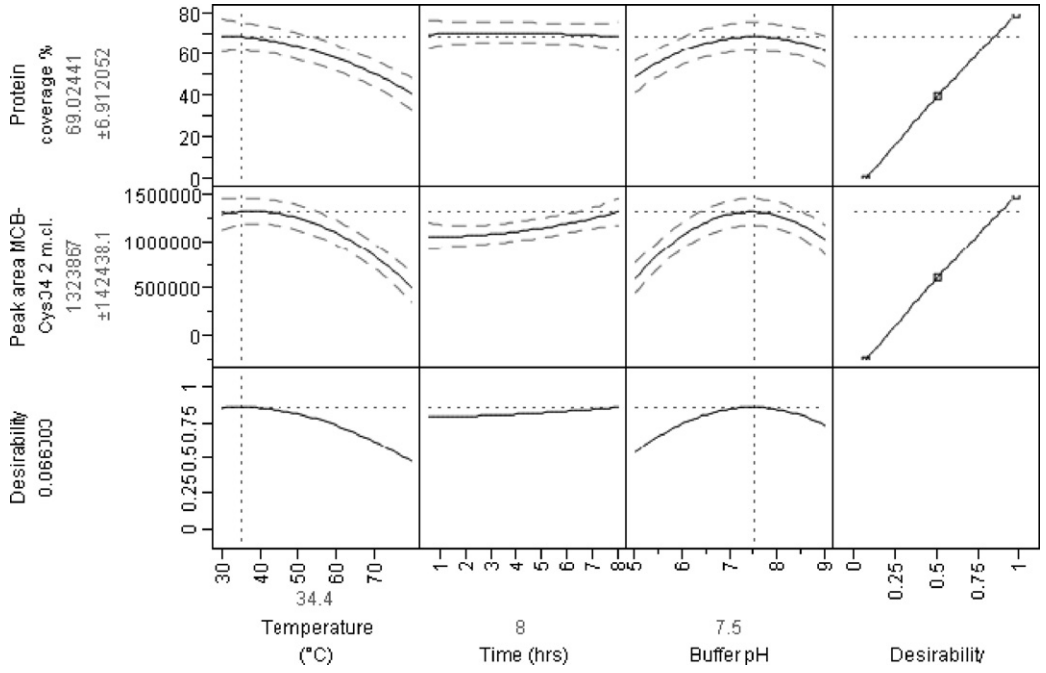

Fig. 6. Prediction profiler plot showing the optimum thermolytic digestion conditions and the predicted responses for the complete model under these conditions. The dashed lines indicate the 95% confidence intervals, the dotted vertical lines indicate the optimum level for each of the factors and the dotted horizontal lines show the predicted value of both responses ± their standard deviations.

Table 3
Digestion conditions and results from the comparison experiment using NAPQI–HSA.

Digestion conditions		pH	Temperature (°C)	Time (h)	Protein cov. (% ± %RSD)	Peak area NAPQI–Cys34 ($\times 10^5$ ± %RSD)
Trypsin	RSM optima	8.4	24	11.6	85 ± 3	6.49 ± 2
	Literature	7.8	37	13	82 ± 0	4.38 ± 5
Thermolysin	RSM optima	7.5	34	8	58 ± 6	2.34 ± 1
	Literature	7.5	65	0.25	36 ± 12	0.33 ± 30

3.3. Comparison experiments using NAPQI–HSA

Since the RSM optimization was performed using a model HSA adduct, the optimized conditions for tryptic and thermolytic digestion were applied to another, clinically relevant HSA adduct. Additionally, a comparison was made with conditions obtained from literature to assess whether the optimized protocols lead to better results. These values, as well as the RSM optima and the responses are shown in Table 3.

Tryptic digestion of NAPQI–HSA using the RSM optimum conditions leads to similar protein coverage and peak area of the modified Cys34 peptide, as compared to those obtained with MCB–HSA. The same can be concluded for the protein coverage obtained with thermolytic digestion of either MCB–HSA or NAPQI–HSA. However, the peak area of the NAPQI–Cys34 peptide is 4-fold lower than the peak area of the MCB–Cys34 peptide. This may most likely be attributed to different ionization efficiencies of the adducted peptides. These results show that the optimized digestion protocol can be successfully applied to different HSA Cys34 adducts, as long as the modification itself does not interfere with cleavage of the protein due to sterical hindrance, for instance.

For both enzymes, the optimized digestion protocols lead to higher protein coverage and peak area of the NAPQI–Cys34 peptide than obtained with values chosen from literature. Trypsin is a very versatile enzyme and will lead to sufficiently high protein coverage even at less than optimal conditions, which was also evident from the RSM results. However, when investigating a certain part of the protein or modification site, it is worthwhile to optimize the digestion conditions and test different enzymes to improve the limit of detection for the peptide containing this site. With the optimized conditions from the RSM, the peak area of the tryptic NAPQI–Cys34 peptide was 50% higher than that obtained with the conditions taken from literature. In the case of thermolysin, the applied digestion conditions had a large effect on both responses. Using the RSM optima for thermolytic digestion, the protein coverage improved by 1.5-fold and the NAPQI–Cys34 peak area by 7-fold, which may become crucial when analyzing lower concentrations of the adducted protein.

4. Conclusion

The results presented in this study highlight the importance of detailed optimization of protein digestion conditions for each analytical system. In particular, when specific target peptides have to be detected, such as in proteomics-based drug–protein adduct studies where detection and identification of the modification site is paramount, much stands to be gained. By applying a DOE, interactions between the different factors were taken into account and a suitable compromise could be made to obtain optimum conditions for both responses, thereby rendering OVAT approaches inadequate. The optimum digestion conditions found for trypsin and thermolysin showed discrepancies with the optima given by the supplier and the values found in the literature. Especially for thermolysin, these differences were significant, rendering a thorough optimization of the digestion conditions for this enzyme mandatory, as shown here in the field of drug–protein adduct research. In both cases, the optimized digestion protocols lead to improved digestion of the studied HSA adducts.

Acknowledgements

We thank Filipe Santos and Johan van Heerden from the Molecular Cell Physiology department at the Faculty of Earth and Life Sciences of the VU University Amsterdam for their help with the statistical evaluation of the data. This work was performed within the framework of the Dutch Top Institute Pharma project "Towards novel translational safety biomarkers for adverse drug toxicity [Project no. D3-201-1]".

References

[1] S. Zhou, E. Chan, W. Duan, M. Huang, Y.Z. Chen, Drug Metab. Rev. 37 (2005) 41.
[2] D.C. Evans, A.P. Watt, D.A. Nicoll-Griffith, T.A. Baillie, Chem. Res. Toxicol. 17 (2004) 3.
[3] S. Zhou, J. Chromatogr, B. Analyt, Technol. Biomed. Life Sci. 797 (2003) 63.
[4] J. Zheng, L. Ma, B. Xin, T. Olah, W.G. Humphreys, M. Zhu, Chem. Res. Toxicol. 20 (2007) 757.
[5] S. Ma, M. Zhu, Chem. Biol. Interact. 179 (2009) 25.
[6] S. Ma, R. Subramanian, J. Mass Spectrom. 41 (2006) 1121.
[7] J.S. Hoos, M.C. Damsten, J.S.B. de Vlieger, J.N.M. Commandeur, N.P.E. Vermeulen, W.M.A. Niessen, H. Lingeman, H. Irth, J. Chromatogr, B. Analyt, Technol. Biomed. Life Sci. 859 (2007) 147.
[8] M.C. Damsten, J.N.M. Commandeur, A. Fidder, A.G. Hulst, D. Touw, D. Noort, N.P.E. Vermeulen, Drug Met. Dispos. 35 (2007) 1408.
[9] J. Carol-Visser, M. van der Schans, A. Fidder, A.G. Huist, B.L.M. van Baar, H. Irth, D. Noort, J. Chromatogr, B. Analyt, Technol. Biomed. Life Sci. 870 (2008) 91.
[10] H. Badghisi, D.C. Liebler, Chem. Res. Toxicol. 15 (2002) 799.
[11] X.X. Yang, Z.P. Hu, S.Y. Chan, S.F. Zhou, Clin. Chim. Acta 365 (2006) 9.
[12] G. Aldini, G. Vistoli, L. Regazzoni, L. Gamberoni, R.M. Facino, S. Yamaguchi, K. Uchida, M. Carini, Chem. Res. Toxicol. 21 (2008) 824.
[13] C.S. Allardyce, P.J. Dyson, J. Coffey, N. Johnson, Rapid Commun. Mass Spectrom. 16 (2002) 933.
[14] L.W. Dick Jr., D. Mahon, D. Qiu, K.C. Cheng, J. Chromatogr, B. Analyt, Technol. Biomed. Life Sci. 877 (2009) 230.
[15] C. Cresswell, A.R. Newcombe, S. Davies, I. Macpherson, P. Nelson, K. O'Donovan, R. Francis, Biotechnol. Appl. Biochem. 42 (2005) 163.
[16] B. Rietschel, S. Bornemann, T.N. Arrey, D. Baeumlisberger, M. Karas, B. Meyer, Proteomics 9 (2009) 5553.
[17] F. Fischer, A. Poetsch, Proteome Sci. 4 (2006) 2.
[18] E. Calleri, C. Temporini, E. Perani, C. Stella, S. Rudaz, D. Lubda, G. Mellerio, J.L. Veuthey, G. Caccialanza, G. Massolini, J. Chromatogr. A 1045 (2004) 99.
[19] M.A. Bezerra, R.E. Santelli, E.P. Oliveira, L.S. Villar, L.A. Escaleira, Talanta 76 (2008) 965.
[20] H. Tye, Drug Discov. Today 9 (2004) 485.
[21] O.W. Gooding, Curr. Opin. Chem. Biol. 8 (2004) 297.
[22] R. Chen, X. Jiang, D. Sun, G. Han, F. Wang, M. Ye, L. Wang, H. Zou, J. Proteome Res. 8 (2009) 651.
[23] W.E. Funk, H. Li, A.T. Lavarone, E.R. Williams, J. Riby, S.M. Rappaport, Anal. Biochem. 400 (2010) 61.
[24] J.C. Fernandez-Checa, N. Kaplowitz, Anal. Biochem. 190 (1990) 212.
[25] S.J. Bark, N. Muster, J.R. Yates, 3rd, G. Siuzdak, J. Am. Chem. Soc. 123 (2001) 1774.
[26] Y. Qiu, L.Z. Benet, A.L. Burlingame, J. Biol. Chem. 273 (1998) 17940.
[27] T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, A. Nystrom, J. Pettersen, R. Bergman, Chemometr. Intell. Lab. Syst. 42 (1998) 3.