Available online at www.sciencedirect.com
ScienceDirect
Journal of Pharmaceutical and Biomedical Analysis 44 (2007) 388–395
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Structural characterization of PEGylated r*Hu*G-CSF and location of PEG attachment sites

Mario Cindrić [a,*], Tina Čepo [a], Nives Galić [b], Mirjana Bukvić-Krajačić [a], Nick Tomczyk [c], Johaness P.C. Vissers [c], Laura Bindila [d], Jasna Peter-Katalinić [d]

[a] *Pliva-Research & Development Ltd., Analytics, Prilaz baruna Filipovića 29, 10000 Zagreb, Croatia*
[b] *Department of Chemistry, Faculty of Science, Horvatovac 102A, 10000 Zagreb, Croatia*
[c] *Waters Corporation-MS Technologies Centre, Atlas Park Simonsway, Manchester M22 5PP, United Kingdom*
[d] *Institute for Medical Physics and Biophysics, Robert Koch Str. 31, University of Münster, 48149 Münster, Germany*

Received 31 October 2006; received in revised form 24 February 2007; accepted 27 February 2007
Available online 4 March 2007

Abstract

Mass spectrometry structural characterization is an essential tool in validating the quality of PEG-r*Hu*-proteins. However, in either case top-down or bottom-up fashion, the interference of high intensity PEG signals on MS detection or detrimental influence of PEG on protein structure, leads to incomplete structural characterization. We propose here a method that permits complete and reliable structural characterization of PEGylated recombinant human granulocyte-colony stimulating factor (r*Hu*G-CSF). The approach includes on-column 2-methoxy-4,5-dihydro-1*H*-imidazole derivatization of digested PEG r*Hu*G-CSF and subsequent LC/MS investigation. By comparing the LC/MS retention of derivatized and underivatized digested PEG r*Hu*G-CSF, location of the PEG attachment within r*Hu*G-CSF could be deduced. Besides, the protein sequence coverage and position of the disulfide bridges was fully deducible from the MS data interpretation. Additionally, ultra performance liquid chromatography–mass spectrometry-to-the-E (UPLC–MS^E) was introduced for analysis of label-free digested PEG r*Hu*G-CSF here to enable high resolution and mass accuracy of MS detection and facilitate deep structural insights of peptides.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Mono-PEGylated r*Hu*G-CSF; LC–MS/MS; MALDI-TOF MS; Lys derivatization; MS^E; N-terminal PEG-peptide

1. Introduction

Recombinant human proteins are widely distributed in clinical applications. However, the main drawback associated with the administration of most recombinant molecules is their rapid body clearance. One approach to improve the therapeutic efficacy by extended circulation time of protein *in vivo* and to decrease the dosing frequency is the covalent attachment of polyethylene glycol (PEG) to the recombinant protein [1–4]. PEG is usually attached to proteins at α-amino group of the N-terminal residue and ε-amino groups of lysine residues by alkylation or acylation [4]. The extent to which improvement of the PEG-recombinant proteins properties occurs is often dependent on the degree of PEG modification and the location of conjugated PEG on protein [3]. Recombinant human granulocyte-colony stimulating factor (r*Hu*G-CSF) is a hormone-like glycoprotein which regulates proliferation and differentiation of neutrophilic granulocytes. Recombinant human G-CSF is used for treatment of the myelosuppresion caused by chemotherapy [4]. The potential PEGylation sites in r*Hu*G-CSF are N-terminus and amino groups of the four lysine residues. PEGylated protein with single PEG group attached to the N-terminal residue of the r*Hu*G-CSF has the same activity as the non-PEGylated one, but essentially much longer circulation half-time, thus, reducing the necessity of daily administration [4]. To characterize PEGylated proteins different analytical techniques like gel electrophoresis [4,5], capillary electrophoresis [3,6,7], chromatography [8,9] and mass spectrometry [10,11] were so far employed.

* Corresponding author. Tel.: +385 1 372 31 19/91 250 23 71 (mobile); fax: +385 1 372 15 14.
E-mail address: mario.cindric@pliva.hr (M. Cindrić).

0731-7085/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2007.02.036

Matrix assisted laser desorption/ionization (MALDI)–mass spectrometry (MS) and MALDI–postsource decay(PSD) MS are viable and routinely applied tools for protein characterization *via* peptide mass fingerprint analysis (PMF) or peptide sequencing analysis, respectively [12]. However, for the characterization of PEGylated proteins this technique is better suited for the analysis of PEG dispersion and PEG-protein or PEG-peptide conjugation. The experimentally obtained masses of peptides are compared with the theoretical ones in sequence databases. This procedure can lead to ambiguous identification, because the obtained signals and their intensities in MS spectra depend on peptide intrinsic properties, amino acid composition, proline internal rearrangements or C- and N-terminus amino acid acidity and basicity, *etc*. Lys-terminated tryptic fragments are less accessible to MS-detection than Arg-terminated ones because of the lower pK_a of lysine than of arginine residue [13,14]. Beside that, MS/MS spectra of peptides carrying N-terminal modifications, conjugates and posttranslational modifications are complex and difficult to interpret. Different chemical modification of peptides were suggested to enhance peptide signals in MS or to obtain predominantly one type of fragments in PSD or MS/MS in order to achieve higher sequence coverage or even more, to directly allow *de novo* sequencing [15]. Chemical derivatization of the amino group of peptides usually includes guanidination [16–18] and sulfonation [19–23]. Guanidination of ε-amino group of lysine residue with *O*-methylisourea converts Lys to more basic homoarginine residue, yielding improved signal intensities in MS spectra. Moreover, it prevents sulfonation of ε-amino group, which can then be used for selective modification of N-terminus of peptides. Sulfonated peptides (by 2-sulfobenzoic acid, 4-sulfophenyl isothiocyanate, 3-sulfopropionic acid NHS-ester or chlorosulfonylacetyl chloride) produce in PSD MALDI experiments easily interpretable fragmentation patterns represented by series of y-type fragment ions. 2-methoxy-4,5-dihydro-1*H*-imidazole was introduced as a multifunctional labeling reagent which increases the ionization efficiencies of Lys-terminated fragments and simplifies tandem mass spectrum by almost exclusive generation of y-type fragment ions [24]. Beside improvement in ionization and fragmentation efficiency, derivatized lysine-containing peptides exhibit higher hydrophobicity and therefore, can be used for lysine determination. This feature is particularly beneficial for the analysis of chemically modified proteins (lysines are known to be one of the most reactive protein sites). Recently, we proposed a new on-column tagging method for the determination of disulfide bonds in recombinant human interferon α-2b and recombinant human erythropoietin based on the side reaction of the derivatization process with 2-methoxy-4,5-dihydro-1*H*-imidazole [25].

In this work, the on-column 2-methoxy-4,5-dihydro-1*H*-imidazole derivatization method was introduced as an easy-to-use method for determination of the position of disulfide bridges and intact lysines in PEGylated protein. Additionally, the application of UPLC–MS^E for analysis of label-free peptides provided complementary data and greatly enhanced the peak resolution, mass accuracy and MS/MS abilities.

2. Experimental

2.1. Reagents and materials

For all solvent system preparations water of MILLI Q quality (resistivity less than 18.2 MΩ cm at 25 °C and total organic carbon less than 5 ppb) was used. Acetonitrile (ACN) of gradient grade chromatography and trifluoroacetic acid (TFA) were purchased from Merck (Darmstadt; Germany). For MALDI–MS analysis α-cyano-4-hydroxycinnamic acid (CHCA) and sinapic acid matrices were obtained from Sigma-Aldrich (St. Louis, USA) and Fluka (Buchs, Switzerland), repsectively. Dithiothreitol with a purity >99% and deuterium oxide of 99.9% atom were purchased from Sigma–Aldrich (St. Louis, MO, USA). Protein samples investigated in this study, recombinant human granulocyte-colony stimulating factor r*Hu*G-CSF (Neupogen®) and polyethylene glycol (PEG)-conjugated r*Hu*G-CSF (Neulasta®) were purchased from Amgen (USA) and further prepared according to the procedure described in Section 2.2. For protein digestion, Glu-C and Lys-C proteinases, respectively, were obtained from Roche (Mannheim, Germany); ammonium hydrogen carbonate (NH_4HCO_3) was bought from Riedel-deHaën (Hanover, Germany); Tris-buffer and β-mercaptoethanol were obtained from Merck (Darmstadt, Germany). For accurate calibration of peptide/protein spectra a ProteoMass Peptide & Protein MALDI–MS Calibration Kit MS-CAL1 from Sigma–Aldrich (St. Louis, USA) was used. On-column sample derivatization was carried out by using 2-methoxy-4,5-dihydro-1*H*-imidazole (LysTag 4H) derivatizing agent, provided by Agilent Technologies (Wilmington, USA).

Protein samples derivatization and purification were carried out by using Microcon YM-10 and ZipTip C_4 (tetrasilyl-silica gel) columns purchased from Millipore (Bedford, USA), PerfectPure C_{18} Tip (octadecylsilyl-silica gel), obtained from Eppendorf (Hamburg, Germany) and Peptide Cleanup C_{18} Spin Tubes provided by Agilent Technologies (Wilmington, USA).

2.2. Sample preparation

The recombinant PEG-r*Hu*G-CSF (Neulasta®, 10 mg/ml) was diluted at concentration 1.5 mg/ml with water in a total volume of 200 μl and desalted through Microcon YM-10. A volume of 200 μl was added and centrifuged at 2300 × *g* for 60 min. After centrifugation, the flow-through was discarded. The membrane was washed three times with 200 μl of water. After centrifugation (2300 × *g*, 40 min), flow-through was discarded each time. The sample was recovered by turning the tube upside down and centrifuging it at 300 × *g* for 5 min. The end volume of desalted sample was reconstituted up to the starting volume.

2.3. Digestion and purification

After desalting, PEG r*Hu*G-CSF was digested with endoproteinase Lys-C in a mixture containing 150 μl of protein (1.5 mg/ml), 10 μl of enzyme, 4.8 μl of Tris–acetic buffer (pH

8.5; 0.1 M) and 30 μl of water. The digestion mixture was incubated for 18 h at 37 °C. Protein digestion with endoproteinase Glu-C was performed by resuspending dried PEG r*Hu*G-CSF (SpeedVac) in 80 μl of NH_4HCO_3 (pH 7.8; 0.025 M) and 25 μl of endoproteinase Glu-C. The digestion mixture was incubated for 18 h at 37 °C.

After digestion, each peptide mixture of PEG r*Hu*G-CSF was dissolved 40-fold in water solution of 0.1% TFA in a total volume of 40 μl (~1.5 μg of protein digest) prior to MALDI-TOF analysis (underivatized r*Hu*G-CSF). The mixtures were purified and derivatized with 2-methoxy-4,5-dihydro-1*H*-imidazole on PerfectPure C_{18} Pipette Tip or Peptide Cleanup C_{18} Spin Tube according to procedure described previously by Cindrić et al. [25]. The purification and derivatization procedure are described in details for both, MALDI-TOF and LC/MS analyses.

2.4. MALDI-TOF analysis

Peptides resulted by protease treatment before and after derivatization were analyzed by an Applied Biosystems MALDI-TOF Voyager DE STR Biospectrometry Workstation (Foster City, CA, USA). The peptides were analyzed in reflector and linear ion mode under the following conditions: positive ion mode, accelerating voltage 20,000 V, grid voltage 66% and extraction delay time 150 ns for reflector ion mode; and positive ion mode, accelerating voltage 25,000 V, grid voltage 60% and extraction delay time 600 ns for linear ion mode. Intact r*Hu*G-CSF and conjugated PEG r*Hu*G-CSF were analyzed in linear mode under described conditions.

The matrices α-cyano-4-hydroxycinnamic and sinapic acid were prepared by dissolving 10 mg of matrix in 1 ml of acetonitrile–water (50:50, v/v).

2.5. LC–MS analysis

ESI-MS spectra were obtained on Q TRAP mass spectrometer (Applied Biosystems, Foster City, CA, USA) connected to Agilent 1100 LC instrument under the following conditions: positive ion mode, source temp 450 °C, scan rate 1000 amu/s, LIT fill time 40 ms, declustering potential 30 V, entrance potential 10 V, high usage of collision gas at collision energy values from 20 to 40 eV.

To separate Lys-C and Glu-C peptide fragments before and after 2-methoxy-4,5-dihydro-1*H*-imidazole derivatization, reversed phase chromatographic C_4 column, 4.6 mm × 250 mm (5 μm), pore size 30 nm (GRACE VYDAC, Hesperia, CA, USA) was used. The column was equilibrated for 10 min with channel A that contained 0.06% TFA (v/v). The peptides were eluted over 115 min with a linear gradient of 0–61% channel B that contained 0.06% TFA diluted in a solution of acetonitrile–water (90:10, v/v). The flow rate was 0.75 ml min^{-1}. Hydrophobic peptides were subsequently eluted over 10 min with 83% channel B at the flow rate 1.25 ml min^{-1}. After the elution of hydrophobic peptides, the column was rinsed with 100% channel B for 10 min at the same flow. Finally, after 5 min the flow was decreased to 0.75 ml min^{-1} and 0% channel B. The same conditions were kept for the last 5 min, thus, equilibrating the column for the following injection.

2.6. UPLC–MS analysis

ESI-MS spectra were obtained on Q-Tof Premier (v mode of operation) mass spectrometer (Waters Corporation, Manchester, UK) connected to an ACQUITY UPLC system (Waters Corporation) under the following conditions: positive ion mode, source temp 450 °C, capillary 3.2 kV, sample cone 25 V. LC–MS data was collected in an alternating, low energy and elevated energy (MS^E) mode of acquisition [26,27].

The spectral acquisition time in each mode was 0.5 s with a 0.1 s interscan delay. In low energy MS mode, data was collected at constant collision energy of 4 eV. In elevated energy MS mode, the collision energy was ramped from 15 to 30 eV during each 0.5 s data collection cycle.

To separate Glu-C peptide fragments, a reversed phase chromatographic, 2.1 mm × 150 mm C_{18} BEH (Bridged Ethylsiloxane/silica Hybrid), 1.7 μm, 130 Å pore size column (Waters Corporation) was used. The column was equilibrated for 0.5 min with 95% of channel A that contained 0.1% formic acid (v/v). The peptides were eluted over 120 min with a linear gradient of 0–55% channel B, that contained 0.1% formic acid diluted in acetonitrile (v/v). The flow rate was 0.3 ml min^{-1}. Hydrophobic peptides were eluted over the next 5 min with 80% mobile phase B. After that, elution conditions were set to initial conditions for 3 min.

3. Results and discussion

3.1. MALDI-TOF MS analysis of rHuG-CSF, N-terminally PEGylated rHuG-CSF and N-terminally PEGylated peptide of Lys-C digest

In order to precisely attest the conjunction of PEG with protein or N-terminal peptide, both the intact PEGylated r*Hu*G-CSF and the Lys-C digestion mixture (PEG N-terminal peptide) were analyzed on MALDI-TOF in positive linear ion mode. The average mass of the intact PEG-protein was confirmed by the ions at *m/z* 39,759 (Fig. 1A). The obtained value deviates from the theoretical one for about 1 kDa (theoretical mass is 38,800 Da, 18,800 Da for intact r*Hu*G-CSF and around 20,000 Da for PEG). The reason lies in the dispersion of the attached PEG polymer which causes the signal broadening (width at half height 2816 Da, Fig. 1A). Therefore, the mass error of approximately 950 Da is inevitable in this high molecular mass range. On the other hand, the intact r*Hu*G-CSF was analyzed in order to confirm the measurement accuracy in given conditions (Fig. 1B). The width of the r*Hu*G-CSF signal at half height is 30 Da which is significantly lower than that for the PEG-protein, thereby confirming the influence of polymer dispersion on signal broadening.

The purified Lys-C digest of PEGylated r*Hu*G-CSF was analyzed in both reflector (data not shown) and linear ion mode on MALDI-TOF. Signals obtained in reflector mode correspond to fragments L2 *m/z* 875.49, L3 *m/z* 1129.60 and L4 *m/z* 698.37

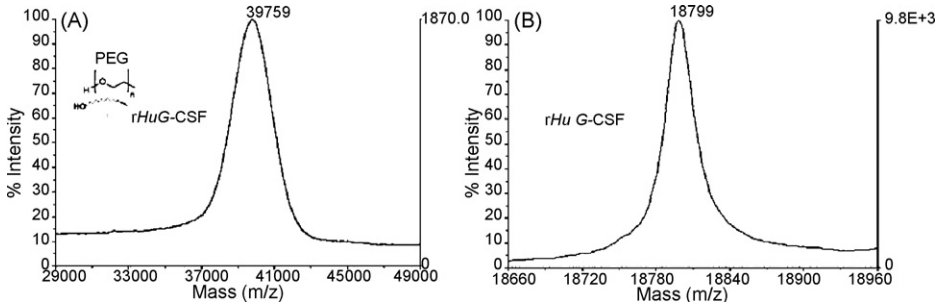

Fig. 1. MALDI-TOF mass spectra of intact PEGylated r*Hu*G-CSF with *m/z* 39,759 (A) and intact G-CSF with *m/z* 18,799 (B) analyzed in positive linear ion mode.

(data not shown). The N-terminal fragment L1 (*m/z* 1786.98) could not be detected in reflector mode, while the largest fragment, L5, with theoretical average *m/z* 14,386.67 does not contain any lysine and therefore no possible PEGylation site. On the contrary, the Lys-C digest of r*Hu*G-CSF yielded all four fragments; L1, L2, L3 and L4 (data not shown). The L1 N-terminal fragment from PEG-r*Hu*G-CSF Lys-C digest was analyzed in linear mode giving rise to ion at *m/z* 22,811 (Fig. 2). The obtained mass shows approximately the same measurement deviation of about 1 kDa and a signal width at half height of 2897 Da in comparison to the intact PEG-r*Hu*G-CSF mass measurement (see above).

3.2. LC–MS analysis of rHuG-CSF and N-terminally PEGylated rHuG-CSF, Lys-C and Glu-C digests

PEGylation reaction of r*Hu*G-CSF can generate not only the desired PEGylation of N-terminus, but also side products like PEGylated lysines or altered PEGylated r*Hu*G-CSF variants that cannot be further separated in purification procedure (*e.g.*, gel filtration, ion exchange and hydrophobic interaction chromatography). Even though the N-terminal reductive alkylation is expected to be a highly selective reaction in acidic environment (α-amino pK_a is 7.8–8.0 and lysine ε-amino pK_a is 10.0–10.2), only a clear cut evidence of N-terminal mono-PEGylation would be the final proof of the product quality. For MS-analytical purposes the PEG macromolecule (of approximate size 20 kDa) incorporation into the protein macromolecule

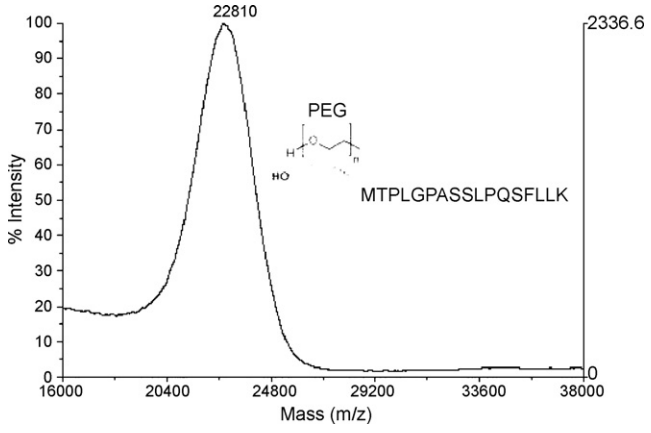

Fig. 2. MALDI-TOF mass spectra of Lys-C N-terminal fragment PEG-L1 with m/z 22 810 analyzed in positive linear ion mode.

is detrimental to structural intact protein (top down) and peptide digest (bottom up) analyses. One of the possible approaches to circumvent this problem was previously described in literature by Kinstler et al. and includes dual digestion with Lys-C and CNBr [4]. By their method N-terminal PEGylation and the amino acid composition of Lys-C/CNBr-derived N-terminal fragment can be confirmed after the fragment isolation and N-terminal sequencing. As an alternative to the off-line approach, derivatization of all lysines and methylation of cysteines (in the same reaction with 2-methoxy-4,5-dihydro-1*H*-imidazole derivatization or *O*-methylisourea) can reveal exact position of cysteines (disulfide bridges) and lysines (possible side reaction groups) in only one LC–MS experiment, confirming this way the preservation of r*Hu*G-CSF protein structure. After Lys-C or Glu-C digests derivatization with 2-methoxy-4,5-dihydro-1*H*-imidazole, lysines and cysteines derivatives produced after the tagging reaction will show a mass increment shift of 68.04 Da for each lysine and 14.02 Da per methylated cysteine [25], taking into account reduced disulfide bridges (–SH HS–). Another benefit of the direct LC–MS approach is that the amino acid composition of each protein fragment can be easily deduced in additional MS/MS experiment after peptide mass fingerprinting (PMF) analysis of Glu-C or Lys-C digests (except for Lys-C fragment L5 with average *m/z* 14,387). To validate this experimental concept for PEGylated r*Hu*G-CSF in the first step *in silico* cleavage with proteinases Lys-C and Glu-C was carried out to derive the masses of peptide fragments necessary for total ion chromatogram (TIC) PMF elucidation (Fig. 3).

After derivatization with Lys-C, all lysine-containing peptides exhibited higher hydrophobicity (in interaction with TFA and stationary phase) and delayed retention time with respect to their underivatized counterparts (Fig. 4). The shift in the retention time can, thus, reveal whether Lys amino group is free or blocked. Fragments L2 and L3 were shifted for 9 and 6 min, respectively, and fragment L4 connected by disulfide bond to fragment L5 before derivatization (L4-L5) appeared in chromatogram after –S–S– bridge methylation as a Lys-derivatized and cysteine methylated fragment $L4^{m}*$. The other part of the L4-L5 fragment, terminal fragment L5 does not contain any lysine but after derivatization all three cysteines were methylated in either case of intermolecular (Cys37-Cys43, fragments L4-L5) or intramolecular (Cys65-Cys75, fragment L5) disulfide bond. Fragment $L4^{m}*$ at *m/z* 780.3 with mass increment of 14 Da for methylated Cys and 68 Da for derivatized Lys relative to underivatized fragment L4 (*in silico m/z* 698.3) and fragment

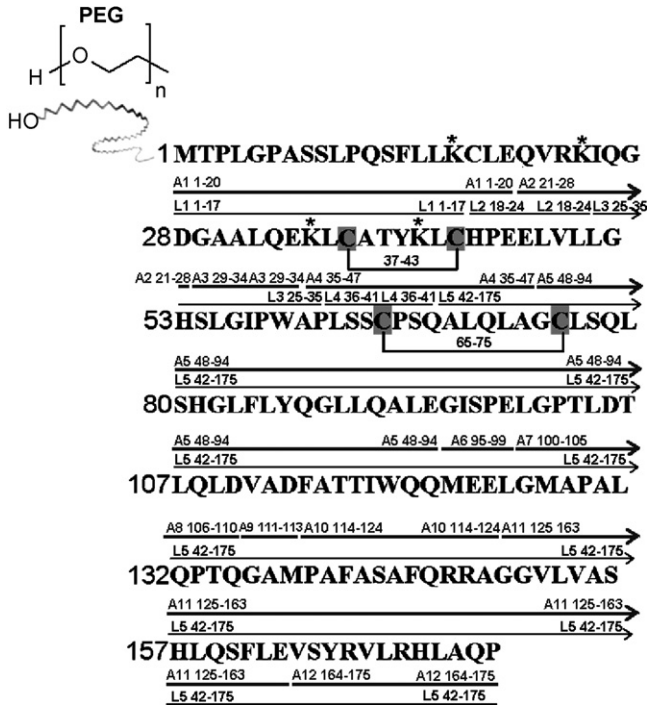

Fig. 3. Primary amino acid sequence of PEGylated r*Hu*G-CSF and Lys-C and (L) Glu-C (A) peptide fragments (black and gray, respectively) obtained after *in silico* digestion. Lysines as potential reactive sites are marked with asterisks. Disulfide bridges are highlighted at positions Cys37-Cys43 and Cys65-Cys75.

L5^{mmm} at average *m/z* 14,427.6 with mass increment of 42 Da for three methylated Cys relative to underivatized fragment L5 (*in silico* average *m/z* 14,386.7) were compared to the fragment L4-L5 at average *m/z* 15,079.3 (Fig. 4). After molecular mass comparison of fragment L4-L5 obtained before derivatization and fragments L4^{m}* and L5^{mmm} after derivatization, position and oxidation state of disulfide bonds could be estimated. Defining the number and position of cysteines in a complex conjugated protein represents a key step in analysis of the protein primary structure, especially at chemically modified proteins. By this relatively simple approach (after designation of Lys-C fragments retention time) we unambiguously confirmed that PEGylation of r*Hu*G-CSF did not alter or reduce disulfide bonds in comparison to r*Hu*G-CSF.

Further, all Lys-C peptide fragments were identified before and after derivatization by mass or retention time increment. The most critical fragment, PEG-L1, for the MS analysis, which contains both PEGylated N-terminus (20 kDa) and free Lys 17, was shifted for 0.7 min after derivatization, thus, pointing out that Lys 17 was free of PEG before derivatization. Worthwhile to mention here is that in the case of potential partial Lys 17 PEGylation, it would be very difficult without applied derivatization to chromatographically discriminate the fragment PEG-L1 with PEGylated lysine from the same fragment with PEGylated N-terminus. Reproducibility of the method was evaluated by five separate sample preparations and consecutive analyses of five underivatized and five derivatized Lys-C digests. The average retention time shifting between fragments PEG-L1 and PEG-L1* was 0.73 ± 0.11 min.

The proposed concept was additionally confirmed by comparison of PEG- r*Hu*G-CSF Glu-C digests, before and after derivatization (Fig. 5). Lysine-containing Glu-C fragments (A1^{m}*, A2* and A4^{mm}**) were completely derivatized as in the case of the lysine-containing Lys-C fragments, even though lysines were not positioned at the C-termini. All other peaks eluted at the same retention time as the ones before derivatization (A3, A5, A6, A7, A8, A10 and A11), except for the doubly methylated fragment A5^{mm} (Cys65-Cys75) which was shifted for 6.6 min and tripeptide fragment A9 that was not observed. After 10 consecutive injections, the average shift of retention time between fragments PEG-A1 and PEG-A1^{m}* was 1.10 ± 0.11 min.

Comparison and analysis of fragments PEG-L1, PEG-A1, PEG-L1* and PEG-A1^{m}* mass spectra (Fig. 6) revealed a truncation of N-terminal fragment invariably at the position Pro11.

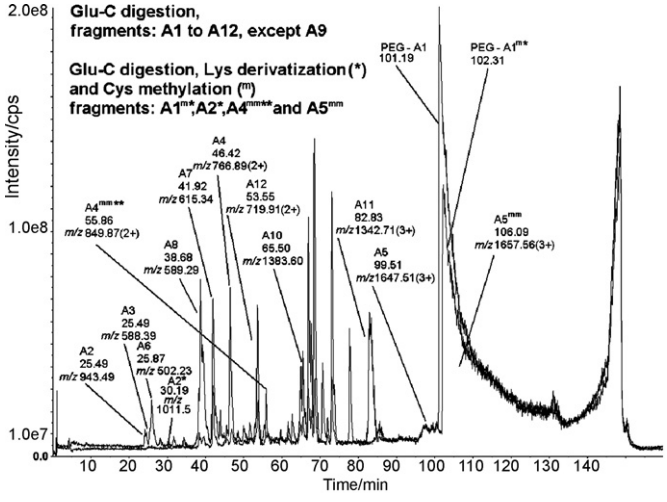

Fig. 4. Overlaid total ion chromatograms of Glu-C PEGylated r*Hu*G-CSF digests before and after lysine derivatization and cysteine methylation. Derivatized lysine-containing fragments are marked with asterisk and methylated cysteine-containing fragments with letter m.

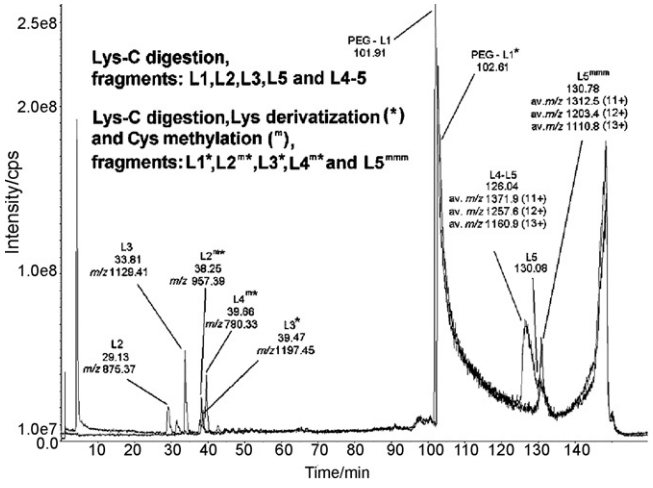

Fig. 5. Overlay of total ion chromatograms of Lys-C PEGylated r*Hu*G-CSF digests before and after lysine derivatization and cysteine methylation. Derivatized lysine-containing fragments are marked with asterisk and methylated cysteine-containing fragments with letter m.

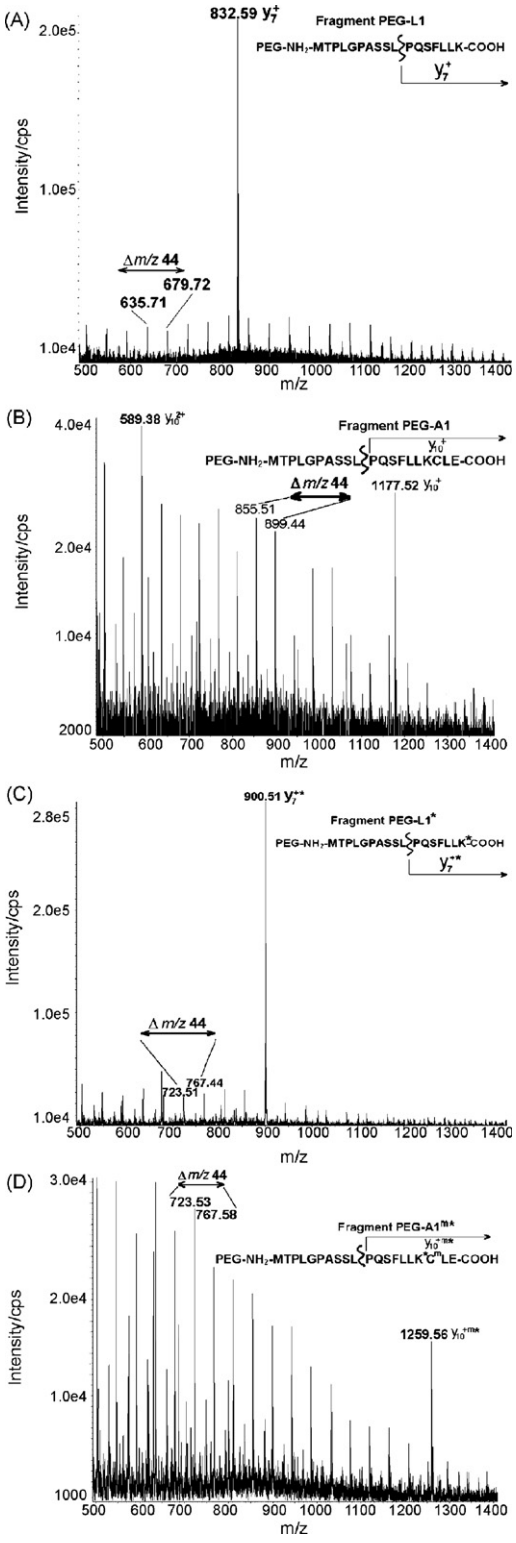

Fig. 6. Mass spectra of Lys-C and Glu-C N-terminal fragments PEG-L1 (A) and PEG-A1 (B) with the singly charged ions at *m/z* 832.59 and 1177.52, respectively. Both fragments are truncated at the same position Pro11 revealing that Lys17 is not PEGylated. Mass spectra of Lys-C and Glu-C N-terminal fragments PEG-L1* (C) and PEG-A1^{m}* (D) with the singly charged monoisotopic ions at *m/z* 900.51 and 1259.56 were obtained after derivatization. Mass increment of 68 Da for derivatized lysine and 14 Da for cysteine methylation additionally confirmed N-terminus mono-PEGylation.

Besides shifting of the fragment retention time after derivatization, the N-terminal fragment cleavage is an additional data to prove the Lys17 non-PEGylation. Moreover, the size of the PEGylated fragments, about 22 kDa and the high level of noise arising from permanent in-source PEG fragmentation (Δ*m/z* 44, visible in Fig. 6) would most probably make impossible the peptide mass analysis without the PEGylated N-terminal fragment truncation. Combining the data obtained by PMF extracted from TICs, the determination of the shifts in retention time of derivatized PEGylated peptide fragments (Lys-C and Glu-C) and mass spectra analysis of the same fragments before and after derivatization, it can be concluded that the developed method allows for the exact determination of the PEG position in the molecule, as well as r*Hu*G-CSF primary amino acid sequence preservation.

3.3. UPLC/MS^E analysis of N-terminally PEGylated rHuG-CSF and Glu-C digest

Peptide mapping as one of the most versatile and widely used means for primary structure analysis of proteins can be achieved with a large palette of analytical methods [28]. Simple comparison of UV–vis chromatograms obtained for a reference substance and a test substance is not sufficient for protein structural elucidation and finally sequence analysis. Mass spectrometry can upgrade the UV–vis chromatography by PMF analysis. Yet, some of the modifications like oxidation, partial oxidation, N-terminal modification, truncation, deamidation and partial deamidation may frequently remain undetectable by this approach and/or by LC–MS/MS peptide mapping. MS^E concept introduced for the first time in 2005 [29–31] resides in parallel alternating scans for acquisition at either low collision energy to obtain precursor ion information (PMF), or ramping of high collision energy to obtain full-scan accurate mass fragment, precursor ion and neutral loss information. By using ultra performance liquid chromatography (UPLC), the resolution of chromatographic peaks corresponding to digested PEGylated proteins was here improved up to 10 times (Fig. 7) relative to classical HPLC peptide mapping (Fig. 5). A full width at half

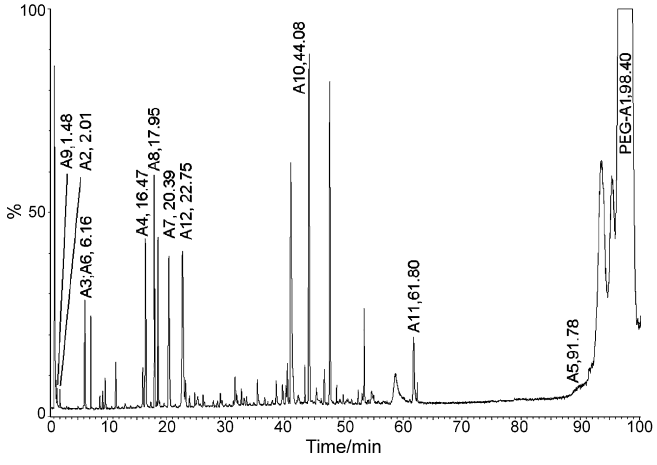

Fig. 7. Total ion chromatogram of Glu-C PEGylated r*Hu*G-CSF digest performed on UPLC–MS system. Precursor ions were annotated after low collision energy (4 eV) TIC chromatogram examination.

Table 1
In silico calculated and monoisotopic singly charged m/z values of Glu-C peptides calculated from obtained multiply charged ions after PEGylated r*Hu*G-CSF digestion

$[M+H]^+$ Glu-C fragment *m/z in silica*	Molecular formula	$[M+H]^+$ observed *m/z*	Error (ppm)
PEG-A1 truncated fragment 1177.6291	C54H88N12O15S	Truncated fragment 1177.6321	2.5
A2 943.5325	C39H70N14O13	943.5242	8.8
A3 588.2993	C24H41N7O10	588.2938	9.3
A4 1532.7241	C67H105N17O20S2	1532.7257	1.0
A5 4940.6211	C227H359N57O62S22	4940.5837	7.6
A6 502.2513	C21H35N5O9	502.2477	7.2
A7 615.3353	C27H46N6O10	615.3274	12.8
A8 589.3197	C25H44N6O10	589.3149	8.14
A9 304.1509	C12H21N3O6	304.1479	9.8
A10 1383.6255	C62H90N14O20S	1383.6223	2.3
A11 4026.0846	C180H285N51O50S2	4025.9960	22.0
A12 1438.8283	C65H107N21O16	1438.8247	2.5

maximum of peptide peaks obtained by UPLC/MS^E analysis was in average 0.17 min relative to 0.6 min obtained by LC/MS analysis. Improvement of peak width and resolution obtained on BEH columns raised overall MS sensitivity by a factor of 10. However, in comparison to the triple quadrupole instrument, Q-TOF instrument can provide precise MS/MS data with accurately measured masses. Thus, position of disulfide bridges, PEG attachment sites and protein sequence coverage can be determined from high reliable experimental data in only one analysis. For routine analyses and quality control of the PEGylated products "classical" liquid chromatography system (*e.g.*, C_4 column, 4.6 mm × 250 mm) coupled to the triple quadrupole instrument should be sufficient but we strongly suggest development of the appropriate method on the UPLC/Q-TOF instrument because of implementation of data/information dependant analysis or MS^E concept that reduce sample consumption and time needed for development phase.

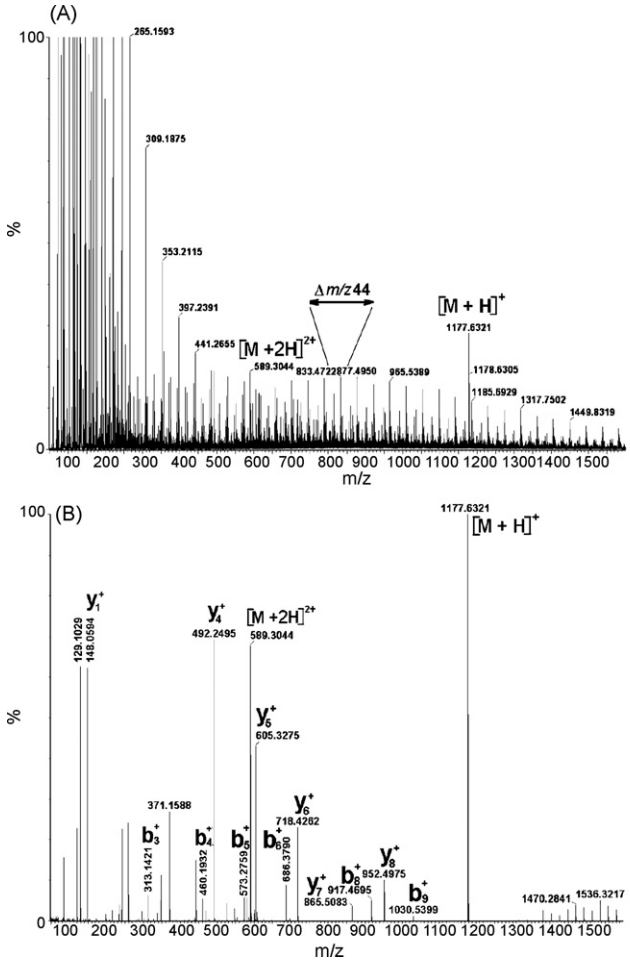

Fig. 8. MS^E spectra (high energy ramp 15–30 eV) of truncated fragment PQSFLLKCLE, PEG-A1 extracted from high collision energy TIC chromatogram before (A) and after (B) PEG signals subtraction (96–98 min).

Mass accuracy with mass error up to 25 ppm presented in Table 1 can be further improved using lock-spray (without ion suppression effect) or some other internal calibrant introduced in the ion source (not applied in this study).

Nevertheless, obtained mass error less than 50 ppm provided high reliable data for PMF analysis and less then 25 ppm for non-lockmass corrected MS^E sequencing of the Glu-C peptides. Fragmentation data of the truncated fragment PEG-A1 extracted from the spectrum that was highly polluted with PEG signals demonstrated the validity of the MS^E concept (Fig. 8).

4. Conclusions

To verify the PEGylation status of recombinant proteins with respect to the localization of PEG attachment sites and degree of PEGylation, and ultimately to assert the amenability of PEGylated recombinant protein for therapeutic purposes, mass spectrometry is an indispensable tool. However, in MS analysis of PEGylated recombinant proteins, the high intensity of PEG related signals hinder the detection of PEGylated peptides and their fragmentation analysis.

In this study, it was demonstrated that on-column 2-methoxy-4,5-dihydro-1*H*-imidazole derivatization of the Lys containing peptides and its side reaction, *i.e.* methylation of Cys, significantly contributed to enhancement of the MS characterization of PEGylated recombinant proteins. The study was exemplified on recombinant human granulocyte colony stimulating factor *rHu*G-CSF. Comparative LC/MS assessment of derivatized and underivatized Lys-C and Glu-C digested PEGylated r*Hu*G-CSF

revealed through different retention times, PMF and MS data inspection: localization of PEG attachment sites, Cys position and lysine-derivatization versus its PEGylation and primary protein structure. These data supplemented by the MS-information obtained by use of ULPC-MSE at greatly improved resolution and sensitivity permitted an unambiguous and complete structural determination of PEGylated r*Hu*G-CSF. Thus, the results presented here assert the viability of the method for characterization of PEGylated r*Hu*G-CSF. We consider this approach of broad applicability for validation of PEGylated recombinant proteins quality as therapeutic agents in clinical applications.

Acknowledgements

The authors would like to thank PLIVA-Research and Development Ltd. and Waters for their comprehensive help and support. The work was partly supported by the Ministry of Science, Education and Sports of the Republic of Croatia through grant no. 0119641.

References

[1] B.B. Yang, P.K. Lum, M.M. Hayashi, L.K. Roskos, J. Pharm. Sci. 93 (2004) 1367–1373.
[2] M.C. Heller, J.F. Carpenter, T.W. Randolph, J. Pharm. Sci. 88 (1999) 58–64.
[3] M.J. Roberts, J.M. Harris, J. Pharm. Sci. 87 (1998) 1440–1445.
[4] O.B. Kinsler, D.N. Brems, S.L. Lauren, A.G. Paige, J.B. Hamburger, M.J. Treuheit, Pharm. Res. 13 (1996) 996–1002.
[5] D.H. Na, E.J. Park, Y.S. Youn, B.W. Moon, Y.W. Jo, S.H. Lee, W.B. Kim, Y. Sohn, K.C. Lee, Electrophoresis 25 (2004) 476–479.
[6] D.H. Na, M.O. Park, S.Y. Choi, Y.S. Kim, S.S. Lee, S.D. Yoo, H.S. Lee, K.C. Lee, J. Chromatogr. B 754 (2001) 259–263.
[7] D.H. Na, K.C. Lee, Anal. Biochem. 331 (2004) 322–328.
[8] S.P. Monkarsh, Y. Ma, A. Aglione, P. Bailon, D. Ciolek, B. DeBarbieri, M.C. Graves, K. Hollferder, H. Michel, A. Palleroni, J.E. Porter, E. Russoman, S. Roy, Y.C.E. Pan, Anal. Biochem. 247 (1997) 434–440.
[9] P.I. Guerra, C. Acklin, A.A. Kosky, J.M. Davis, M.J. Treuheit, D.N. Brems, Pharm. Res. 15 (1998) 1822–1827.
[10] E. Watson, B. Shah, R. DePrince, R.W. Hendren, R. Nelson, BioTechniques 16 (1994) 278–281.
[11] H. Lee, T.G. Park, J. Pharm. Sci. 92 (2003) 97–103.
[12] K. Gevaert, J. Vandekerckhove, Electrophoresis 21 (2000) 1145–1154.
[13] T. Keough, M.P. Lacey, R.S. Youngquist, Rapid Commun. Mass Spectrom. 16 (2002) 1003–1015.
[14] E. Krause, H. Wenschuh, P.R. Jungblut, Anal. Chem. 71 (1999) 4160–4165.
[15] K.G. Standing, Curr. Opin. Struct. Biol. 13 (2003) 595–601.
[16] S.L. Cockrill, K.L. Foster, J. Wildsmith, A.R. Goodrich, J.G. Dapron, T.C. Hassell, W.K. Kappel, G.B.I. Scott, BioTechniques 38 (2005) 1–4.
[17] R.L. Beardsley, J.P. Reilly, Anal. Chem. 74 (2002) 1884–1890.
[18] F.L. Brancia, S.G. Oliver, S.J. Gaskell, Rapid Commun. Mass Spectrom. 14 (2000) 2070–2073.
[19] B. Samyn, G. Debyser, K. Sergeant, B. Devreese, J. Van Beeumen, J. Am. Soc. Mass Spectrom. 15 (2004) 1838–1852.
[20] P. Chen, S. Nie, W. Mi, X.C. Wang, S.P. Liang, Rapid Commun. Mass Spectrom. 18 (2004) 191–198.
[21] T. Keough, M.P. Lacey, R.J. Strife, Rapid Commun. Mass Spectrom. 15 (2001) 2227–2239.
[22] T. Keough, M.P. Lacey, R.S. Youngquist, Rapid Commun. Mass Spectrom. 14 (2000) 2348–2356.
[23] T. Keough, R.S. Youngquist, M.P. Lacey, Anal. Chem. (2003) 156A–165A.
[24] E.C. Peters, D.M. Horn, D.C. Tully, A. Brock, Rapid Commun. Mass Spectrom. 15 (2001) 2387–2392.
[25] M. Cindrić, T. Čepo, A. Škrlin, M. Vuletić, L. Bindila, Rapid Commun. Mass Spectrom. 20 (2006) 694–702.
[26] R.H. Bateman, J.I. Langridge, T. McKenna, K. Richardson, Methods and Apparatus for Mass Spectrometry, UK Patents GB 2,363,249/GB 2,364,168 and US Patent 2,385,918A (2003).
[27] R.H. Bateman, R. Carruthers, J.B. Hoyes, C. Jones, J.I. Langridge, A. Millar, J.P.C. Vissers, J. Am. Soc. Mass Spectrom. 13 (2002) 792–803.
[28] J. Bongers, J.J. Cummings, M.B. Ebert, M.M. Federici, L. Gledhill, D. Gulati, G.M. Hilliard, B.H. Jones, K.R. Lee, J. Mozdzanowski, M. Naimoli, S. Burman, J. Pharm. Biomed. Anal. 21 (2000) 1099–1128.
[29] R.S. Plumb, K.A. Johnson, P. Rainville, B.W. Smith, I.D. Wilson, J.M. Castro-Perez, J.K. Nicholson, Rapid Commun. Mass Spectrom. 20 (2006) 1989–1994.
[30] J.C. Silva, R. Denny, C.A. Dorschel, M.V. Gorenstein, G.Z. Li, K. Richardson, D. Wall, S.J. Geromanos, Mol. Cell. Proteomics 5 (2006) 589–607.
[31] J.C. Silva, R. Denny, C.A. Dorschel, M.V. Gorenstein, I.J. Kass, G.Z. Li, T. McKenna, M.J. Nold, K. Richardson, P. Young, S. Geromanos, Anal. Chem. 77 (2005) 2187–2200.