



Universiteit  
Leiden

The Netherlands

## Multi-level structural and functional characterization of therapeutic glycoproteins by mass spectrometry

Lippold, S.

### Citation

Lippold, S. (2021, November 30). *Multi-level structural and functional characterization of therapeutic glycoproteins by mass spectrometry*.

Retrieved from <https://hdl.handle.net/1887/3243984>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3243984>

**Note:** To cite this publication please use the final published version (if applicable).

# 3

## CYSTEINE AMINOETHYLATION ENABLES THE SITE-SPECIFIC GLYCOSYLATION ANALYSIS OF RECOMBINANT HUMAN ERYTHROPOIETIN USING TRYPSIN

**Steffen Lippold,<sup>1</sup> Alexander Büttner,<sup>2</sup> Matthew S.F. Choo,<sup>3</sup>  
Michaela Hook,<sup>2</sup> Coen J. de Jong,<sup>1</sup> Terry Nguyen-Khuong,<sup>3</sup>  
Markus Habeger,<sup>2</sup> Dietmar Reusch,<sup>2</sup> Manfred Wuhrer<sup>1</sup>  
and Noortje de Haan<sup>1</sup>**

<sup>1</sup>*Center for Proteomics and Metabolomics, Leiden University Medical Center,  
Albinusdreef 2, 2333 ZA Leiden, Netherlands*

<sup>2</sup>*Pharma Technical Development, Roche Diagnostics GmbH,  
82377 Penzberg, Germany*

<sup>3</sup>*Bioprocessing Technology Institute, Agency for Science Technology and Research,  
Singapore 138668*

*Reprinted and adapted from Anal. Chem. 2020, 92, 14, 9476–9481*

*DOI: 10.1021/acs.analchem.0c01794*

*Copyright © 2020 American Chemical Society*

Recombinant human erythropoietin (rhEPO) is an important biopharmaceutical for which glycosylation is a critical quality attribute. Therefore, robust analytical methods are needed for the in-depth characterization of rhEPO glycosylation. Currently, the protease GluC is widely established for the site-specific glycosylation analysis of rhEPO. However, this enzyme shows disadvantages, such as its specificity and the characteristics of the resulting (glyco)peptides. The use of trypsin, the gold standard protease in proteomics, as the sole protease for rhEPO is compromised, as no natural tryptic cleavage site is located between the glycosylation sites Asn24 and Asn38. Here, cysteine aminoethylation using 2-bromoethylamine was applied as an alternative alkylation strategy to introduce artificial tryptic cleavage sites at Cys29 and Cys33 in rhEPO. The (glyco)peptides resulting from a subsequent digestion using trypsin were analyzed by reverse-phase liquid chromatography-mass spectrometry. The new trypsin-based workflow was easily implemented by adapting the alkylation step in a conventional workflow and was directly compared to an established approach using GluC. The new method shows an improved specificity, a significantly reduced chromatogram complexity, allows for shorter analysis times, and simplifies data evaluation. Furthermore, the method allows for the monitoring of additional attributes, such as oxidation and deamidation at specific sites in parallel to the site-specific glycosylation analysis of rhEPO.

### 3.1. INTRODUCTION

RECOMBINANT human erythropoietin (rhEPO) is a successful therapeutic glycoprotein whose bioactivity and safety are highly affected by specific glycosylation features (e.g., the rhEPO half-life depends on the number of sialic acids) [11]. First-generation rhEPO exhibits three *N*-glycosylation sites (Asn24, Asn38, Asn83) and one *O*-glycosylation site (Ser126). *N*-Glycans of rhEPO show a wide structural diversity, including a varying number of antennae, *N*-acetylglucosamine (GlcNAc) repeats, terminating sialic acids (*N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc)), and glycan modifications, such as acetylation, phosphorylation, and sulfation [77]. The *O*-glycosylation site reportedly carries core 1-type structures with 0, 1, or 2 sialic acids, including acetylation [109]. Numerous mass spectrometry (MS)-based analytical methods have been described to characterize rhEPO glycosylation at the released glycan level [77, 110], by studying glycopeptides [59, 111–114], or, more recently, analyzing rhEPO in its intact form [59]. Between those methods, the bottom-up glycopeptide analysis is the only approach that allows the simultaneous characterization of *N*- and *O*-glycosylation heterogeneity in a site-specific manner [11].

Site-specific glycopeptide analysis requires the generation of proteolytic peptide moieties covering a single glycosylation site [115]. Therefore, protease selection is a crucial step in analytical method development. Trypsin is the gold standard in proteomics due to its high specificity and robustness [116]. However, the use of this protease alone is not sufficient for site-specific glycosylation analysis specifically for rhEPO, as no natural tryptic cleavage sites (Lys and Arg) are located between the glycosylation sites Asn24 and Asn38 [11, 111]. GluC is currently the protease of choice in most rhEPO glycopeptide studies [111–114]. However, the use of this enzyme is generally compromised by its pH dependent specificity, its low digestion efficiency, and the generation of relatively large peptide portions [68]. Hence, numerous missed cleaved products and inconsistent results for the obtained peptide moieties are reported, hampering the data analysis of rhEPO glycopeptides [59, 111, 113, 114]. To overcome the disadvantages of the generation of rhEPO glycopeptides by GluC, different approaches were established. For example, trypsin and GluC were used in a double digestion of rhEPO [117] or information from two independent complementary digestions was combined [59, 114]. However, these strategies are less consistent and more time-consuming than the use of a single specific protease.

In the current study, site-specific glycosylation analysis of rhEPO using trypsin as the sole proteolytic enzyme was enabled via the aminoethylation of cysteines. Using this approach, cysteines are transformed into pseudolysines and recognized as substrates by trypsin [118]. Aminoethylation can be performed using 2-bromoethylamine, as was first described in 1956 [119]. This approach was applied by several others to facilitate proteolytic cleavage using trypsin [120, 121]. For example, cysteine aminoethylation and trypsin digestion were recently combined in a biopharmaceutical application, increasing the sequence coverage of the complementarity determining region in a

monoclonal antibody [121]. In the current study on rhEPO, cysteine aminoethylation created two tryptic cleavage sites between Asn24 and Asn38 at Cys29 and Cys33. This approach was readily integrated into a conventional reverse-phase liquid chromatography (RP-LC)-MS workflow for rhEPO multiple-attribute monitoring at the (glyco)peptide level. As compared to conventional methods, the current approach resulted in a lower diversity in the cleavage products, shorter peptide portions for the glycopeptides, and shorter analysis times. Furthermore, it prevented the interference of oxidation (Met54) and deamidation (Asn47) sites with the assessment of the Asn38 glycosylation.

## 3.2. EXPERIMENTAL SECTION

### 3.2.1. CHEMICALS AND SAMPLES

All chemicals had at least analytical grade quality. Further information about the chemicals used can be found in the **Information S1**. A reference standard of rhEPO produced in a Chinese hamster ovary (CHO) cell line was provided by Roche Diagnostics (Penzberg, Germany).

### 3.2.2. SAMPLE PREPARATION

A detailed description of the sample preparation can be found in the **Information S1**. Briefly, 250 µg of rhEPO was denatured, reduced (20 mM DTT), and then alkylated by either 60 mM iodoacetic acid (for GluC digestion as reported [122]) or 60 mM 2-bromoethylamine (for 1 h at 60 °C for trypsin digestion) in a final volume of 320 µL. Upon buffer exchange, samples were digested for 16-18 h by GluC (25 °C, enzyme/protein 1:25) or trypsin (37 °C, enzyme/protein 1:100) in 450 µL of 50 mM ammonium bicarbonate, pH 7.8. Sample preparation was performed in triplicate per digestion protocol.

### 3.2.3. REVERSE-PHASE LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY

Ten µL of each sample (around 5 µg) was analyzed by C8 RP-LC-MS/MS. A C8 BEH column (2.1 mm × 150 mm, 1.7 µm, 130 Å, Waters) was used on a UPLC system (Vanquish Horizon, Thermo Scientific). The flow rate was kept at 300 µL/min, and the column temperature was kept at 65 °C. A 65 min gradient of 1% to 80% B (mobile phase A, 0.1% formic acid in water; mobile phase B, 0.1% formic acid in acetonitrile) was applied (**Table S1**). The UPLC was hyphenated to an LTQ Orbitrap Velos (Thermo Scientific). MS1 information was obtained for  $m/z$  200-2,000 in the positive ion mode. MS2 scans were acquired by CID with a normalized collision energy of 35% for the five most intense parent ions. Detailed information about the MS settings can be found in **Information S2** [122].

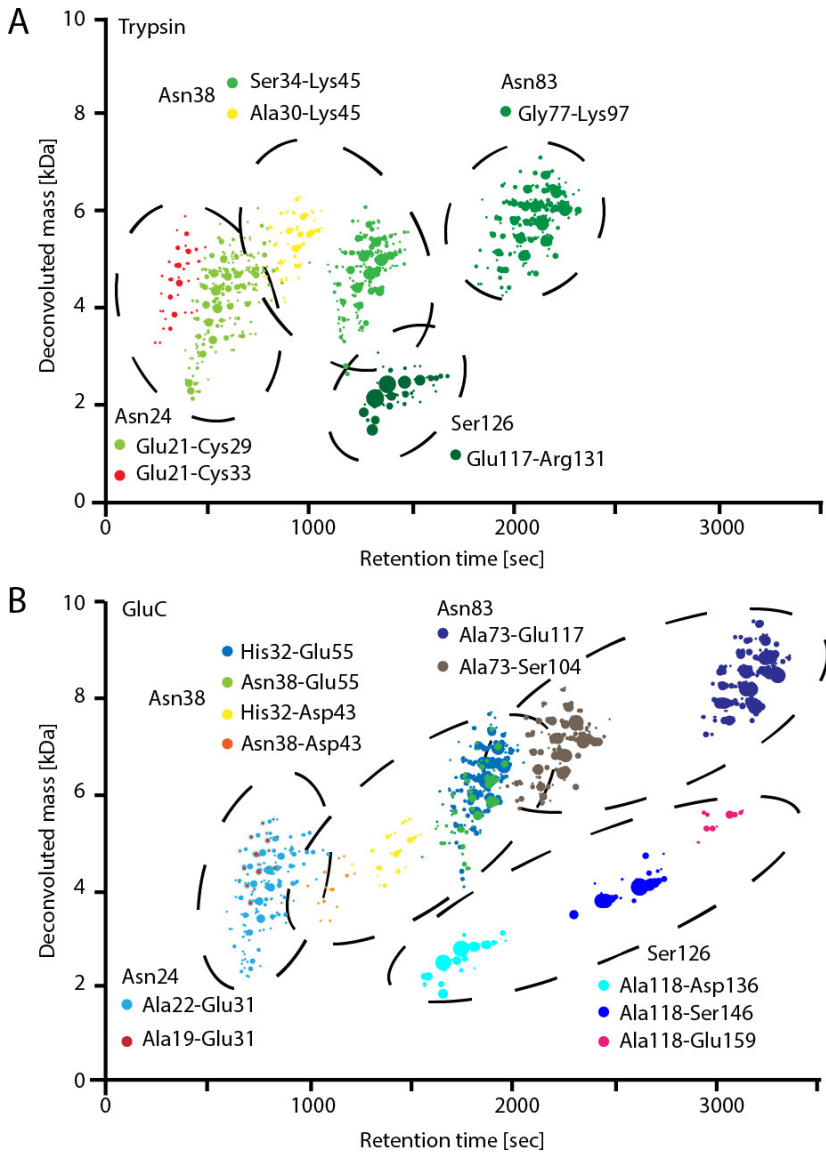
### 3.2.4. DATA ANALYSIS

A detailed overview of the data analysis workflow can be found in **Information S3** and **Figure S1**. In short, GlycopeptideGraphMS (ver. 2.04) [89], in combination with the MS2 confirmation (manual and using Byonic) of selected glycoforms (**Figure S2**) and the literature [122], was used for assignment of the glycopeptides. An initial analyte quality control and automatic quantification of the main cleavage products was performed in LacyTools (ver. 1.0) [90]. Skyline (ver. 19.1.0.193) was used for a manual integration and the quality control of the main glycopeptides of interest for all relevant cleavage products [123].

## 3.3. RESULTS AND DISCUSSION

### 3.3.1. PROTEOLYTIC PEPTIDES

All expected peptides, with at least three amino acids, were assigned for the tryptic rhEPO digest (**Figure S3** and **Table S2**). The efficiencies of the Cys alkylation with 2-bromoethylamine (97.3%) and the subsequent tryptic digestion (99.3%) were evaluated based on one Cys-containing peptide with an adjacent Arg (**Figure S4** and **Table S3**). The alkylation selectivity was evaluated by considering the top seven most abundant peptides, and showed negligible rates (<1%) of Met, Glu, Asp, and Tyr aminoethylation (**Figures S5-S10** and **Table S4**). The non-selectively alkylated sites are in line with the reported (*O*-)alkylation byproducts in *S*-alkylation reactions [124]. Trace amounts of peptides were observed of which the presence is likely explained by the tryptic cleavage of alkylated Met, Glu, or Tyr (**Figures S7** and **S9**). For the glycopeptides obtained in the trypsin workflow, cleavages were observed C-terminally to Lys (Asn38 and Asn83), Arg (Ser126), and aminoethylated Cys (Asn24) (**Figure 3.1** and **Table S5**). Aminoethylation of the cysteines and their cleavage by trypsin was confirmed by fragmentation of the non-glycosylated Asn24 (theoretical  $m/z$  980.4359, 1+) and Asn38 (theoretical  $m/z$  665.8466, 2+) peptides (**Figure S11**). For both Asn24 and Asn38, an additional missed cleaved peptide portion was detected, which was overalkylated. The location of the extra aminoethyl group was assigned to Cys29 in the MS2 analysis of the non-glycosylated Asn24 peptide Glu21–Cys31 (**Figure S12**), likely preventing trypsin cleavage. The estimated relative abundance of the overalkylated and missed cleaved versions of the Asn24 and Asn38 glycopeptides was 9.1% and 10.5%, respectively (**Table S6**). Considering the theoretically increased ionization efficiency of the missed cleaved version, due to the additional amines in the elongated sequence [125], the true relative abundance may be even lower and consequently negligible. For the other glycosylation sites (Asn83 and Ser126), specific peptide portions were obtained using trypsin (**Figure 3.1**).



**Figure 3.1:** Visualization of detected rhEPO glycopeptides in representative replicates by RP-LC-MS/MS analysis upon (A) trypsin digestion after aminoethylation and (B) GluC digestion after carboxymethylation. Glycopeptide assignment and visualization was performed using GlycopeptideGraphMS [89]. The diameter of the data points indicate the relative abundance of the glycopeptides (logarithmic scaling between intensities from  $1 \times 10^6$  to  $1 \times 10^{12}$ ). Data points with the same color have the same peptide backbone but different glycan compositions.

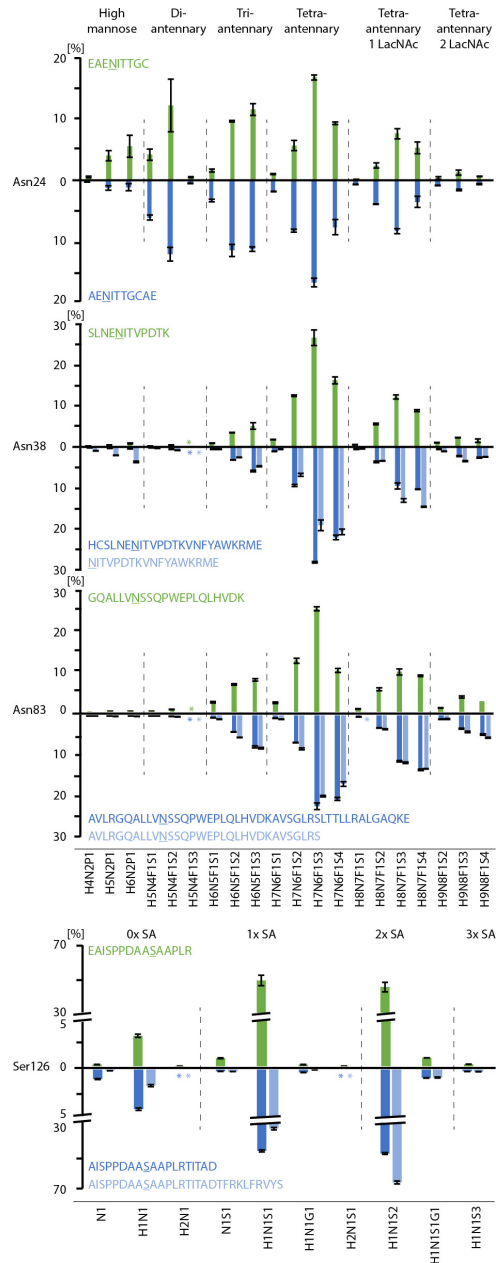
GluC-derived rhEPO glycopeptides showed C-terminal cleavages to Glu (Asn24, Asn38, Asn83, and Ser126), Asp (Asn38 and Ser126), and Ser (Asn83 and Ser126), which were reported earlier for rhEPO [111]. The unspecific cleavage of GluC resulted in several digestion products per glycosylation site. Missed cleaved products of Asn24 (Ala19–Glu37) and Asn38 (Asn38–Glu55) resulted in coeluting glycopeptides in the same cluster as the respective main peptide moieties (**Figure 3.1**), hampering the data analysis [59, 114]. The highest heterogeneity in the obtained GluC peptide portions was observed for Asn38, which showed four cleavage products (**Figure 3.1** and **Table S5**), as reported previously [114]. Furthermore, Ser126 was detected on three different peptide portions, while the GluC digestion resulted in two products for both Asn24 and Asn83. The abundance of additional peptides ranged from 0.3% for Asn38 (Asn38–Asp43) to 29.2% for Ser126 (Ala118–Ser104) (**Table S6**). For further data analysis, peptides of interest were selected based on the relative abundance (>20%). In addition, a GluC cleavage product of Asn38 (Asn38–Glu55) was included, as its glycosylation site is located next to the cleavage site. The rhEPO cleavage products of GluC are not consistently reported in different studies [59, 111, 113, 114]. This may be due to different protease sources, sample preparations, and data analysis procedures.

Using the novel trypsin workflow, all glycopeptide peptide moieties, including the missed cleaved products for Asn24 and Asn38, were well separated by RP-LC, which allowed for straightforward data integration. Furthermore, the trypsin workflow enabled a shorter analysis time as compared to the GluC workflow using the same RP-LC gradient. The maximum retention time for the trypsin workflow was 37 min (Asn83-H7N6F1S4), while that of GluC was 55 min (Asn83-H7N6F1S4). In summary, the trypsin workflow reduced the heterogeneity of proteolytic glycopeptides as compared to the GluC workflow (6 vs 11 glycopeptides for four glycosylation sites) because of the higher protease specificity of trypsin.

### 3.3.2. SITE-SPECIFIC GLYCOSYLATION ANALYSIS

In line with previous reports on rhEPO glycosylation, the trypsin-based workflow showed mainly fucosylated complex-type *N*-glycans with varying numbers of antennae (2-4), LacNAc repeats (1-3), and sialylation (0-5) (**Table S7**) [77, 109, 110]. Moreover, site-specific glycosylation differences were characterized, indicating a relatively high content of high mannose-type and diantennary complex-type structures on Asn24; Asn38 and Asn83 predominantly showed tetra-antennary complex-type *N*-glycans (**Figure 3.2**) [77]. For the *O*-glycosylation site, mainly core 1-type structures with one and two sialic acids were detected, which was in line with those reported in the literature [109]. Low abundances of Neu5Gc-containing glycoforms were assigned for both *N*- and *O*-glycopeptides (**Table S7**). Interestingly, glycopeptide compositional isomers were separated in the RP-LC (**Figure S13**). However, the isomers were not individually considered for further data processing due to uncertainties in their assignment. A detailed exploration of the isomeric structure assignment was considered out of the scope of this report.





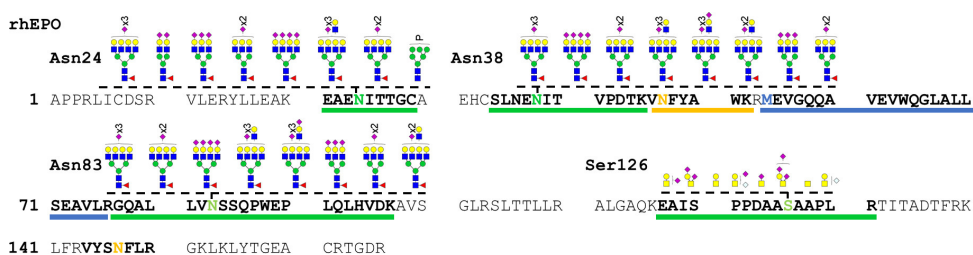
**Figure 3.2:** Site-specific quantitative comparison of the main N- and O-glycoforms (>0.5% relative abundance excluding acetylated variants, **Table S7**) of the peptide moieties of interest obtained by trypsin (green) or GluC (blue) digestion of rhEPO. Bars and error bars represent the mean values and standard deviations of triplicate measurements, respectively. Asterisks (\*) indicate nondetected glycoforms. Compounds are represented as H, hexose; N, N-acetylhexosamine; F, fucose; S, N-acetylneuraminic acid; G, N-glycolylneuraminic acid; and P, phosphate. SA: sialic acid (S or G).

A total number of 39 (74), 32 (56), 28 (58), and 10 (20) compositions for the main trypsin-derived Asn24, Asn38, Asn83, and Ser126 glycopeptide clusters, respectively, were assigned, excluding (including) acetylation variants (**Table S7**). This amounts to 55 (100 including acetylation) different *N*-glycoform compositions. Both *N*- and *O*-glycopeptides revealed low abundant glycoforms, which were previously not reported for rhEPO. For example, hybrid-type structures were detected on Asn24, and a disialyl motif (H1N1S3) was detected on Ser126 (for MS2, see **Table S7** and **Figure S14**). Of note, the glycopeptide feature assignment was based on the retention time clustering using GlycopeptideGraphMS and prior knowledge [89, 122]. To limit the assignments of false-positive glycoforms, MS2 validation was performed for at least one glycoform per cluster (**Figure S2**), and all MS1 signals were subjected to accurate mass and isotopic pattern matching. Glycoforms that have not been reported in CHO cell-produced rhEPO were in part validated by MS2 (**Figure S14**). The new trypsin method resulted in similar site-specific glycosylation profiles as compared to the reference GluC method (**Figure 3.2**). Interestingly, the relative abundance of high mannose glycans at Asn38 was higher for GluC peptide moieties, which had the cleavage site adjacent to the glycosylation site (Asn38–Glu55). Furthermore, differences in the relative abundances of GluC *O*-glycopeptides carrying 46% or 33% H1N1S1 and 47% or 64% H1N1S2 were observed on the shorter or longer peptide, respectively. The findings for Asn38 suggest a glycoform-dependent preference of the GluC digestion when the glycan is located subsequent to the cleavage site. Such effects were shown before for trypsin, which cleaved less efficiently when a fucosylated glycan was present next to the targeted Arg [96]. For the *O*-glycopeptides, the effects might be attributed to differences in MS responses between different cleavage products. To our knowledge, no reports have as of yet investigated the differences in the relative quantification of glycosylation for differently cleaved peptide moieties in rhEPO. Here we show that tryptic peptides result in a simplified data interpretation as compared to GluC, as less peptide heterogeneity is introduced.

In addition to the different glycoforms, the non-glycosylated variants of each of the tryptic cleavage products of Asn24 (2.3%), Asn38 (0.001%), and Ser126 (5.1%) were also detected (**Table S8**). It should be noted that the estimation of the abundance of the non-glycosylated variants would be more accurate after deglycosylation, as the ionization efficiency of the peptide with and without glycan may be vastly different, as previously reported [13]. The GluC-derived peptides showed inconsistent results for non-glycosylated variants on different cleavage products, in particular when the cleavage site was adjacent to the glycosylation site (Asn24 and Asn38). For example, non-glycosylated variants were only detected for Asn38 peptide moieties with the glycosylation site adjacent to the cleavage site (**Table S8**). In addition, the non-glycosylated variant of NITGCAE was observed for the Asn24 site in the GluC digest (**Figure S15**). However, no glycopeptides were observed for this peptide moiety. Overall, the trypsin workflow facilitated a consistent site occupancy determination.

### 3.3.3. OXIDATION AND DEAMIDATION

In addition to glycosylation, the post-translational oxidation (Met54) and deamidation (Asn47 and Asn147) of rhEPO were previously reported to impact its structure and biological activity [126]. Trypsin-based RP-LC-MS peptide mapping to assess the oxidation and deamidation of rhEPO is preferred over GluC-based methods, as trypsin allows the analysis of Asn47 deamidation and Met54 oxidation on separate peptides (**Figure 3.3, S16-S18, and Table S9**) [126]. GluC cleavage faces the issue of Met54 and Asn47 being present on the same peptide moiety as Asn38. This complicates the quantification of the individual post-translational modifications, as both the oxidation and the deamidation result in additional peaks per glycoform. Additionally, it results in ambiguities as the mass difference between a Neu5Gc and a Neu5Ac is exactly the same as the mass increment upon Met54 oxidation (15.9949 Da). Thus, besides site-specific glycosylation analysis, the new trypsin-based workflow offers an approach for multiple-attribute monitoring of rhEPO, including its deamidation and oxidation.



**Figure 3.3:** Sequence of rhEPO. Tryptic peptides covering glycosylation sites (green) and relevant deamidation (orange) and oxidation sites (blue) are highlighted. For the glycosylation sites, the eight most abundant glycoforms are presented. Structural ambiguities of the glycoforms were reduced by information from the literature [77].

## 3.4. CONCLUSIONS

We demonstrated the feasibility and advantages of cysteine aminoethylation with subsequent trypsin digestion for the site-specific glycosylation analysis of rhEPO. To our knowledge, this is the first report that describes trypsin as the sole protease for the cleavage between the glycosylation sites Asn24 and Asn38 of rhEPO. Trypsin showed a high specificity, resulting in six peptide portions covering the four glycosylation sites. In contrast, GluC showed a total of 11 peptide portions for the four glycosylation sites. In the direct comparison between carboxymethylated GluC-derived and aminoethylated trypsin-derived glycopeptides, we found comparable site-specific relative quantification results for the glycosylation of rhEPO. It should be noted that the current study is of exploratory nature, and further validation is needed to assess the quantification consistency of low-abundant glycoforms. The decreased heterogeneity in the proteolytic cleavage products using trypsin resulted in a significantly reduced

sample complexity, causing less ambiguities and facilitating straightforward data analysis. Moreover, the tryptic digestion did not show glycosylation-dependent cleavage and appeared highly suitable for the parallel monitoring of deamidation and oxidation in a multiple-attribute monitoring approach. Finally, the described sample preparation can be integrated in current rhEPO glycopeptide mapping workflows with minor adaptations, making it an attractive method for the biopharmaceutical sector.

## SUPPORTING INFORMATION

Supporting information is available free of charge via <https://pubs.acs.org/doi/10.1021/acs.analchem.0c01794?goto=supporting-info>.

## ACKNOWLEDGEMENTS

The authors would like to thank Peter van Veelen for his input on cysteine aminoethylation. Peter van Veelen and Arnoud de Ru are acknowledged for their support with the Byonic searches. This research was funded by the European Commission H2020 (Analytics for Biologics project, Grant 765502).

