PosterREPRINT

Waters

# GLYCOSYLATION ANALYSIS OF RECOMBINANT E2 BVDV PROTEIN FROM A SEMI-PURIFIED FRACTION USING A WATERS® MICROMASS® CAPLC® Q-TOF™ SYSTEM

*Mark A Ritchie*[1]; Andrew C Gill[2]; Munir Iqbal[2]; Amit Pande[2]; John MacCauley[2]; Therese McKenna[1]; James I. Langridge[1]*

*[1]Waters Corp., Manchester, UK; [2]Institute for Animal Health, Compton Laboratories, Newbury, UK*

**Presented at ASMS, Montréal, Canada, 8th-12th June 2003**

## OVERVIEW

- Glycopeptide discovery by product ion directed MS/MS acquisition on a Waters® Micromass® Q-Tof™ Mass Spectrometer.
- Glycopeptide MS/MS spectra were interpreted using interactive software tools, to determine the composition and partial structure of the glycans displayed.
- Using this approach, three sites of N-linked glycosylation have been characterized on a recombinant E2 protein
- Data Directed Analysis (DDA™) in combination with databank searching allowed identification of contaminating, co-purified proteins.

## INTRODUCTION

Glycosylation can influence protein folding, affect the overall charge and solubility of a protein, sterically hinder protease digestion and act as a cellular recognition signal. The characterization of these structural features is therefore important in understanding biological function. Site-based glycosylation analysis provides invaluable information on the structure of a glycoprotein. The location of a carbohydrate moiety can be of as much importance as its structure. The analysis of glycopeptides is of greater value to the researcher than the `global' glycosylation mapping methods commonly used. The high mass accuracy, sensitivity and MS/MS function of a quadrupole time of flight (Q-Tof) mass spectrometer make it ideal for such an application. In conjunction with an on-line CapLC® system, the Q-Tof has the ability to specifically target eluting glycopeptides for MS/MS.

Bovine viral diarrhoea virus is a positive strand enveloped RNA virus in the genus pestivirus. The members of this genus cause infection in pigs and ruminants, and are unusual because they can be present in animals as a persistent infection in the absence of an immune response.

The virus has three glycoproteins;

- Erns, which is both a virus structural protein and secreted from infected cells, has ribonuclease activity, and binds to cel-surface glycosaminoglycans.
- E1, which is a structural component of the virion but no biochemical function has been associated with it.
- The third is E2, a structural component of the virion and is anchored to the virion through a C-terminal hydrophobic anchor. There is evidence to indicate that E2 binds to the cell surface and is the receptor binding protein. It is thought to be homologous at least in part to Tick Borne encephalitis E glycoprotein or Semliki-Forest E2 glycoprotein.

## METHOD

### Protein production and purification

BVDV E2 was expressed from a construct in Drosophila S2 cells. E2 was expressed to be secreted into the medium by adding an exogenous insect signal sequence (from the Bip protein) and had its C-terminal anchor removed; at its C-terminus it contained a His tag and a V5 epitope tag. It was purified from the medium by affinity chromatography on Ni-resin through its His tag. The sequence of the expressed protein is shown in Figure 1.

SGTSDMPLLLLLPLLWAGALAMDKLHLDCKPEYYYAIAKNDRIGPLGAE
GLTTVWKDYSPEMTLEDTMVIASCRDGKFMYLSRCTRETRYLAILHSRA
LPTSVVFKKLFEGQKQGDTVGMDDDFEFGLCPCDAKPIVRGKYNTTLLN
GPAFQMVCPIGWTGTVSCMLANRDTLDTAVVRTYRRSRPFPYRQGCITQ
KVLGEDLYDCILGGNWTCVTGDQLQYSGGSIESCKWCGFKFQRSEGLP
HYPIGKCRLKNETGYRLVDNTSCNREGVAIVPQGTVKCKIGDTTVQVIA
LDTKLGPMPCKPYEIISSEGPVEKTACTFNYTKTLKNKYFEPRDSYFQQY
MLKGEYQYWFDLEVTDHHRDYEFCRYPAQWRPLESRGPFEGKPIPNPLL
GLDSTRTGHHHHHH

*Figure 1. E2 BVDV fusion protein construct.*

## Protein digestion

The protein solution (100 µL of 600 µg/mL in 2M urea/$NH_4Ac$) was reduced for 3 hrs at room temperature by the addition of 10 µL 0.1M DTT, then alkylated for 3 hrs by the addition of 20 µL 0.1M IAA. The protein was then diluted 1/10 with 0.1M $NH_4Ac$ and digested with 0.5 µg of sequence grade trypsin (sigma) overnight. The digest was then diluted 1/10 with 0.1% formic acid, and 4 µL was used for each LC/MS/MS experiment.

## LC/MS/MS Experiment

- The column was equilibrated with 5% ACN for 10 mins. Samples were injected using an autosampler onto a Waters Symmetry300™ $C_{18}$ Opti-Pak™ and on-line separation was carried out on a Waters Atlantis™ $C_{18}$ 75 µm column. A gradient from 0-40% B (B=95% ACN 0.1% FA) over 40 minutes was provided using a Waters CapLC.
- All data was acquired on a Micromass Q-Tof Ultima™ API mass spectrometer fitted with a nanoflow ESI ZSpray™ source. Mass accuracy was enhanced by the use of a NanoLockSpray™ source, which switches every 10 s to a reference spray (flow rate of 500 nL/min) containing 100 fmol/µL GFP.

## Precursor Ion Discovery

- The Q-Tof was run in Precursor Ion Discovery (PID) mode (Figure 2), where the voltage on the gas collision cell is switched alternately between high (30 v) and low (8 v) every second. This provides both a standard low energy MS spectrum and a spectrum containing all of the product ions derived from the intact molecular species observed in the standard scan. The MassLynx™ Software was modified to specifically allow the selection for MS/MS of ions with 4 + to 7+ charges, thus increasing the likelihood of selecting the larger, highly-charged glycopeptide species over the smaller, non-glycosylated tryptic peptides.

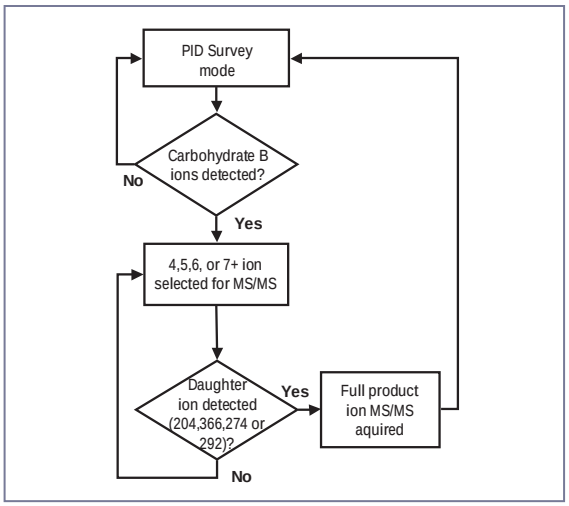



*Figure 2.*

- The MS functions are acquired in centroid mode, providing increased mass accuracy allowing real time product ion detection. Upon detection of carbohydrate B ions ($[Hex]^1$+ m/z 204, $[HexHexNac]^1$+ m/z 366, $[NeuAc - H_20]^1$+ m/z 274; $[NeuAc]^1$+ m/z 292) the instrument switches to the product ion MS/MS mode and selects the most intense multiply charged ion for fragmentation. Once this precursor ion has been selected, MS/MS will only continue if it yields the expected carbohydrate b type product ion(s) (Figure 2). During the MS/MS fragmentation step a collision energy ramp of 20-40 V was applied to obtain a diverse range of fragment ions, in order to provide as much structural information as possible.
- For comparative purposes, the experiment was repeated but the instrument was prevented from switching into the MS/MS mode. The low energy survey allowed chromatography to be monitored, while reconstructed mass chromatograms for the glycopeptide B ions in the high-energy survey reveals the elution profile of the glycopeptides.

**Data Acquisition - DDA**

- The sample was also analyzed in Data Directed Acquisition (DDA) mode, where only the charge and intensity of an ion is considered when selecting species for fragmentation by MS/MS.
- This approach resulted in the analysis of non-glycosylated peptides by MS/MS, providing peptide sequence information on the sample.

**Data Processing**

Glycosylation analysis

Deconvoluted fragment ion MS/MS data can then be interpreted using Carbotools, a specialized carbohydrate interpretation tool (MassLynx 4.0 Software, Waters Corp. Manchester, UK.) Neutral loss of carbohydrate masses can be calculated and annotated onto the MS/MS data, allowing characterization of the carbohydrate moiety. (See Figure 3)

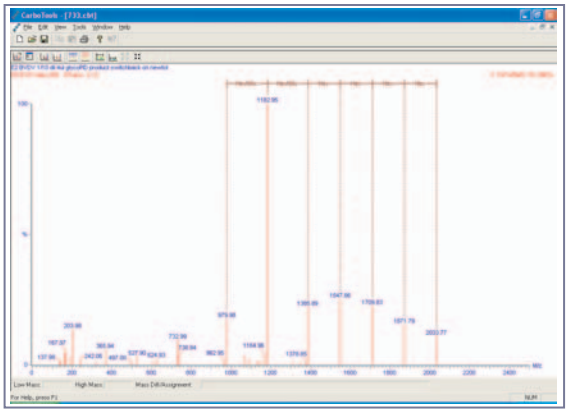


*Figure 3.*

The smallest y-type fragment ion (nomenclature as Domon and Costello) detected is usually the glycopeptide y1 ion, (the peptide with the reducing terminus N-acetylhexosamine attached). Working up from this intense y1 ion, a structure for the carbohydrate moiety present on the glycopeptide can be established. Working down in mass from the Y1 ion, the mass of the peptide can be determined by subtracting the monoisotopic mass of HexNAc (203.07Da). The mass of the peptide can then be used to identify the corresponding amino acid sequence and to potentially locate the site of glycosylation. This is

facilitated in the case of N-linked glycans, by the glycosylation consensus sequence of N-X-S/T/C (where X is any amino acid other than Proline). In an experiment where the protein sequence of interest is known, theoretical prediction of the glycopeptides present and their associated m/z values can be calculated prior to analysis.

### Protein Identification

Data from the LC/MS/MS DDA experiment was processed using ProteinLynx™ Global SERVER 2.0 (PLGS2). The raw data files were deconvoluted and deisotoped using the MaxEnt™ lite algorithm, a lockmass is applied and then the processed data is searched against nrdb (NCBI) using a peptide mass tolerance of 20 ppm and a fragment tolerance of 0.1 Da. Modifications were taken into consideration through the use of the Automod™ program.

## Results

### Protein identification

A results summary of the data from PLGS2 is shown in table 1 below

| Protein | Mw (Da) | N° tryptic peptides | Coverage (%) |
|---|---|---|---|
| SD02860p, Tequila gene product | 267,451 | 16 | 11.4 |
| E2 glycoprotein [pestivirus type 1] | 25,951 | 8 | 15.7 |
| Modified Beta Trypsin | 23,293 | 7 | 32.8 |
| Keratin 1 | 65,977 | 7 | 13.5 |

Table1: PLGS2 results summarized:

- The sample appears to be heavily contaminated with SD02860p from Drosophila, a protease with tryptic, chymotryptic and non-specific protease activity. Inspection of the sequence shows clusters of Histidine, which presumably allowed it to bind to the IMAC column during purification. This contaminant is 10 times the MW of the E2 protein.
- Also trypsin autolysis products and tryptic peptides of keratin were found, which were presumably post-purification contaminants.
- The E2 construct sequence does not appear in the nrdb, but peptides from the E2 portion of this protein have been identified from searching against the nrdb databank, including one peptide (Lys253-Arg260) containing the glycosylation consensus sequence.

### Glycopeptide PID

- A chromatogram from the PID experiment with the MS/MS switching capability disabled is presented in Figure 4 (a, b). Reconstructed mass chromatograms for the carbohydrate B ions shows that there is a wide range of glycosylated peptides, more than is to be expected for a relatively small protein like E2.

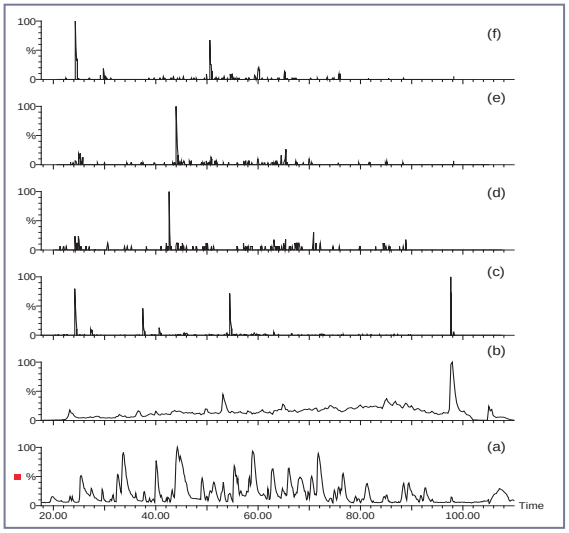



*Figure 4.*

- Similar mass chromatograms for the MS/MS data (Figure 4 c-f) show the most intense glycopeptide ions elute close to the start and end of the chromatographic gradient, with numerous weak ions eluting throughout the experiment.

## Analysis of Glycopeptide MS/MS Data

- A number of good quality glycopeptide MS/MS spectra were acquired, clearly showing the presence of carbohydrate moieties.
- The structures were generally small, N-linked core structures, with some core fucosylation.
- Due to the extent of contamination and the likelihood of non-specific protease activity identification of the glycosylation site is problematic.
- Table 2 summarizes only tryptic or semi-tryptic peptides matching to the E2 or SD02860p proteins. Some structures could not be assigned to a specific peptide. Even with a mass accuracy of sub 20 ppm, it is not possible to determine these peptides by their mass alone due to the large number of candidate peptide sequences and potential glycosylation sites on the Tequila gene product.

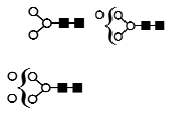
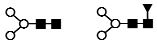
| Glycan structures | Peptide Mass (Da) | Peptide ID |
|---|---|---|
|  | 979.56 | E2<br>$_{253}$LKNETGYR$_{260}$ |
|  | 3442.66 | E2<br>$_{139}$GKYNTTLLNGPAFQMVCPI GWTGTVSCMLANR$_{170}$ |
|  | 3260.54 | E2<br>$_{298}$LGPMPCKPYEIISSEGPVE KTACTFNYTK$_{326}$ |
|  | 3492.78 | E2<br>$_{307}$EIISSEGPVEKTACFNYTKT LKNKYFEPR$_{336}$<br>(chymotryptic/tryptic) |
|  | 1539.75 | Tequila<br>$_{436}$LHTFPIYPPVDNK(T)$_{488}$ |
|  | 3294.52 | Not assigned |
|  | 3835.90 | Not assigned |
|  | 2610.23 | Not assigned |
|  | 1876.82 | Not assigned |
|  | 2265.18 | Not assigned |

*Table 2: Glycopeptide analysis results*

- Even with these challenging circumstances, where there is a high level of contamination, it is possible to identify and characterize three out of the five theoretical glycosylation sites on the E2 protein.
- The 253-260 peptide appears to be the most heterogeneous site, with three glycoforms: one core and two extended high mannose type structures, but it appears that there are no core fucosylated structures present.
- The single site identified on the contaminating Tequila gene product appears to have only one structure, a fucosylated trimannosyl core.

PosterREPRINT

Waters

- It is not known, from this study, if the other two uncharacterized sites on E2 are glycosylated. This may be easier to assess on a purified sample, without the presence of the contaminating protein and the high levels of non-specific cleavage.

## CONCLUSION

- Glycopeptide-PID coupled with the Q-Tof LC/MS/MS system can successfully characterize glycosylation sites even in a complex mixture of peptides produce by a mixture of protease activity.
- In cases where the identity of the peptide is elusive, strategies are available for glycosylation site determination.
- One such strategy would involve PNG`ase digestion of the mixture to release the glycans. The mass of the deglycosylated peptide would shift by 1Da as the Asn previously glycosylated is converted into Asp. Taking this into account, an include list could be generated in a DDA-type experiment to target these peptides for MS/MS-based identification.
- Alternatively an MS3 type experiment utilizing cone/ion tunnel based fragmentation could be used to target the intense Y1 ion for MS/MS and obtain a sequence for peptide portion for identification purposes.
- Future studies will concentrate on a highly purified form of the E2 protein to determine the full microheterogeneity present upon each glycosylation site.

**To whom all correspondence should be addressed:**
**e-mail: Mark_Ritchie@waters.com**

WATERS CORPORATION
34 Maple St.
Milford, MA 01757 U.S.A.
T: 508 478 2000
F: 508 872 1990
www.waters.com

Waters

MICROMASS MS TECHNOLOGIES

Opt-In! My Profile www.waters.com



LLOYD'S REGISTER QUALITY ASSURANCE ISO9001

UKAS QUALITY MANAGEMENT 001

IT