# Supporting information for

## Mass spectrometric detection of targeting peptide bioconjugation to *Red clover necrotic mosaic virus*

Ruqi Wang[1], Dustin M. Lockney[1], Michael B. Goshe[2]*, Stefan Franzen[1]*

1. Department of Chemistry, North Carolina State University, Raleigh, NC 27695

2. Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695

### UPLC data and $MS^E$ chromatograms

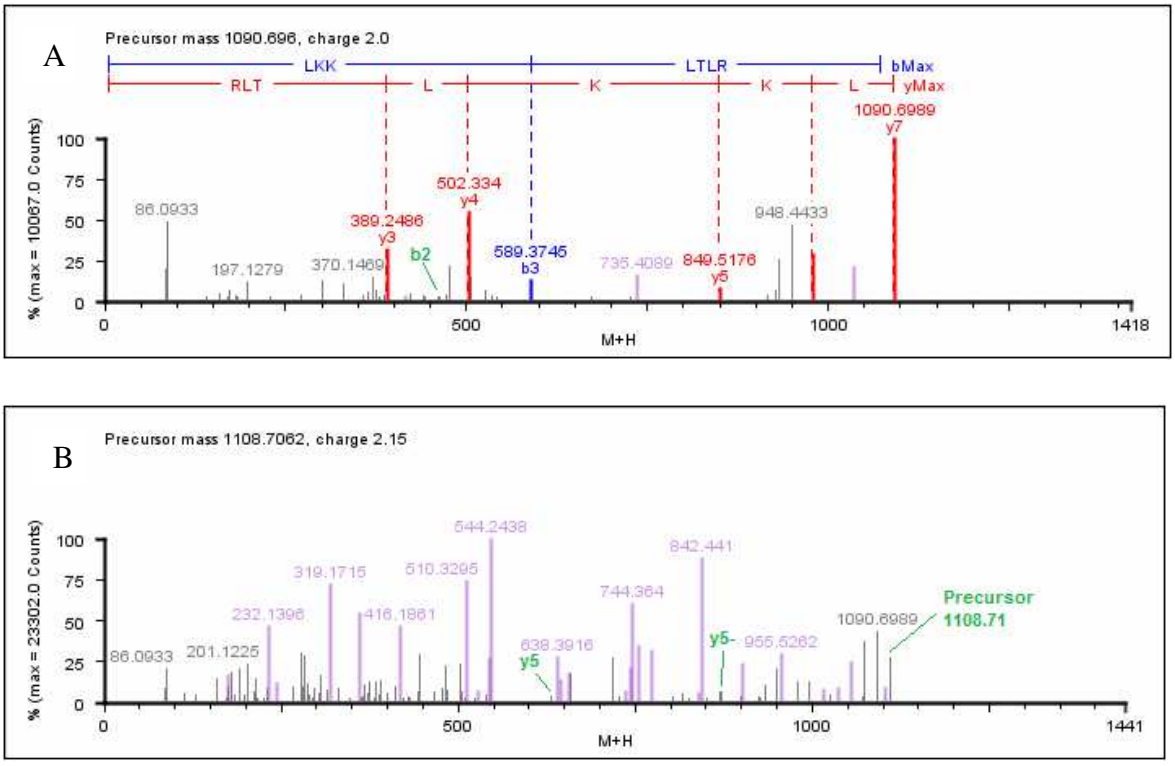


**Figure S1.** LC/$MS^E$ analysis of SMCC labeling of RCNMV. (A) $MS^E$ product ion spectrum of (101)LKK(SMCC)LTLR(107) with the $b_2$ ion containing the modification indicated in green. (B) $MS^E$ product ion spectrum of (101)LKK(SMCC-OH)LTLR(107) with the $y_5$ ion without modification (lower mass value) and the $y_5$ ion with modification (higher mass value) indicated in green.

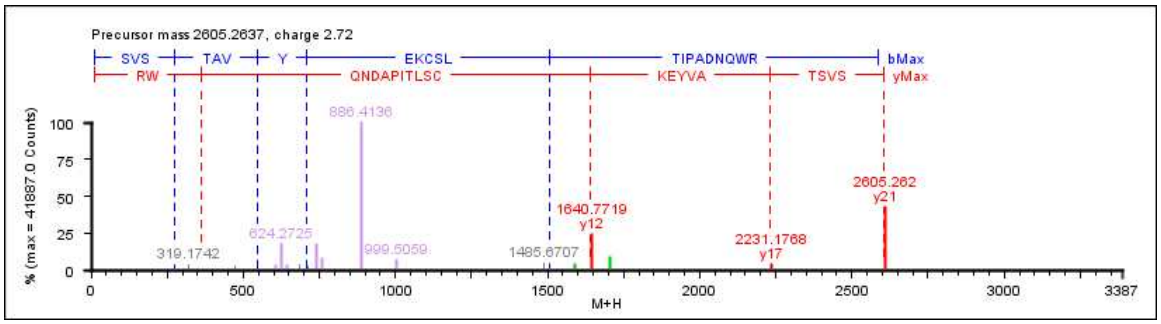


**Figure S2.** $MS^E$ product ion spectrum of (145)SVSTAVYEKC(SMCCOH)SLTIPADNQWR(165) generated from the sSMCC conjugated RCNMV sample.

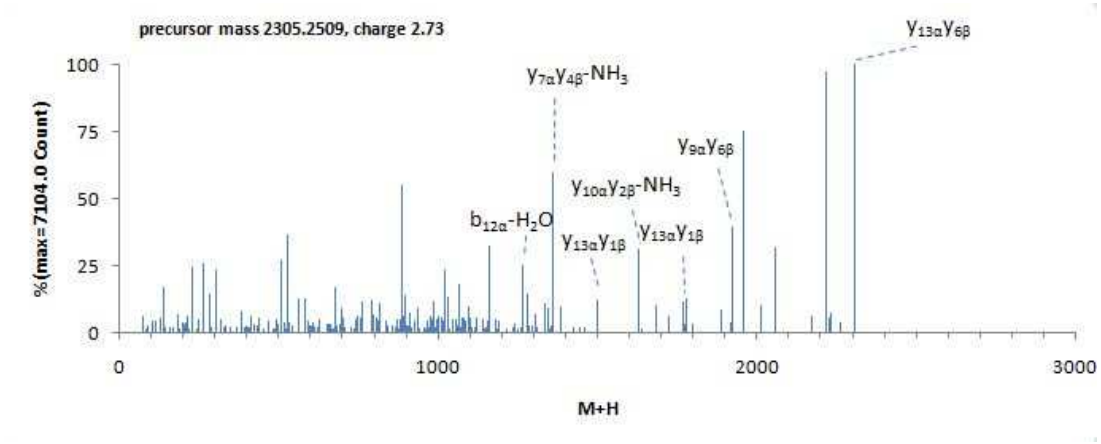


**Figure S3.** $MS^E$ product ion spectrum of TVAIPFAK(~LSTSFC)TQIIK generated from the RCNMV conjugated with the CD46 peptide. The most abundant ion $[M+H]^+$ at *m/z* 2305.25 corresponds to the intact peptide conjugate. Representative crosslinked peptide product ions are shown. Although not labeled, single peptide ions $y_{1\alpha}$, $y_{5\alpha}$, and $b_{7\alpha}$ were dectected for the capsid protein.

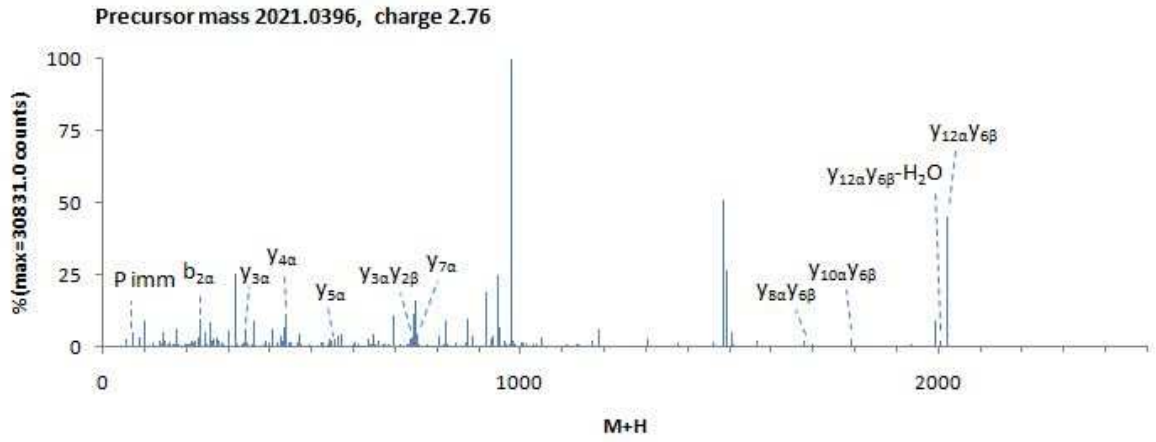


**Figure S4.** $MS^E$ product ion spectrum of (26)TVAIPFAK(~GSGSC)TQIIK(38) generated from the RCNMV conjugated with the GSGS peptide. The most abundant ion $[M+H]^+$ at *m/z* 2021.04 corresponds to the intact peptide conjugate. Representative crosslinked peptide product ions and single peptide product ions are shown. Although not completely labeled, all b and y ions for the capside peptide (α chain) were detected except $y_1$, $y_6$, and $b_4$; all b ions for GSGS peptide (β chain) were detected.

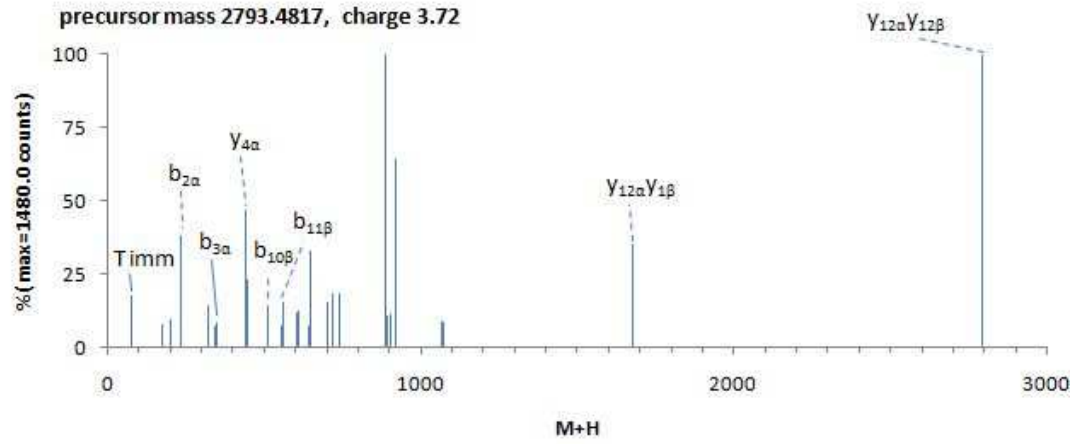


**Figure S5.** $MS^E$ product ion spectrum of (34)TVAIPFAK(~AHAVDINGNQVC)TQIIK(45) generated from the RCNMV conjugated with the ADH304 peptide. The most abundant ion $[M+H]^+$ at *m/z* 2793.48 corresponds to the intact peptide conjugate. Representative crosslinked peptide product ions and single peptide product ions are shown. Although not labeled, product ions $b_1$, $b_2$, $y_1$, $y_3$-$y_5$ for capsid (α chain) were detected, and $b_{10}$ and $b_{11}$ ions for the ADH peptide (β chain) were detected.

**Use of various proteases for digestion of RCNMV to extend peptide sequence coverage**

In order to dectect capsid peptide sequencesnot covered by trypsin and chymotrypsin digestion, other enzymes were used. Pepsin has a primary digestion site at F, W, Y, E, and I residues, but further digestion can be random. Digestion was performed in 10 mM HCl, and after certain time, the reaction was quenched by adding 7uL of 100mM sodium phosphate buffer (pH~8.0) to increase the pH above 7. Then the sample

was frozen at -20℃ until further processing could be performed. Urea (8 M) or heating in

a boiling water bath (5 min) was used to unfold/denature the protein. The method was applied on bovine serum albumin (BSA) as a reference. Ratios of pepsin and RCNMV from 1:5 to 1:500 (wt: wt) were investigated.

The results of the pepsin digestion using a 1:100 ratio of pepsin:RCNMV (wt:wt) are shown in Figure 5. For the first few minutes, the virus was digested very quickly, but appeared to be arrested within 5 min as indicated by the large protein/peptide pieces that were left undigested. The reason for this result cannot be self-digestion or inactivation of the protease since pepsin was able to be detected in the control sample (Lane P). To attempt to circumvent this problem, additional pepsin was added at time points 15, 30 and 45 min after the initiation of the reaction to ensure that active pepsin was present in the solution during the course of the digestion. This approach promoted further digestion, but incomplete RCNMV digestion was still evident.

LC/MS$^{E}$ analysis of RCNMV digested by pepsin did not provide reproducible results. Some samples had higher coverage but questionable peptide identications. By examining the masses of these peptides, we found that the majority of them are below 1000 Da which were difficult to identify due to the low number of product ions. This agrees with the SDS-PAGE results, in that pepsin seemed to digest only the solvent accessible part of the virus, but was not able further access the virus. In contrast, trypsin

(Lane Try) was able to more completely digest the virus except for the 6-10 kDa segments were no cleavage sites were available. In an attempt to obtain sequence information regarding this 6-10 kDa segment, in-gel digestion of this region was performed using pepsin but the small peptides generated precluded peptide detection.

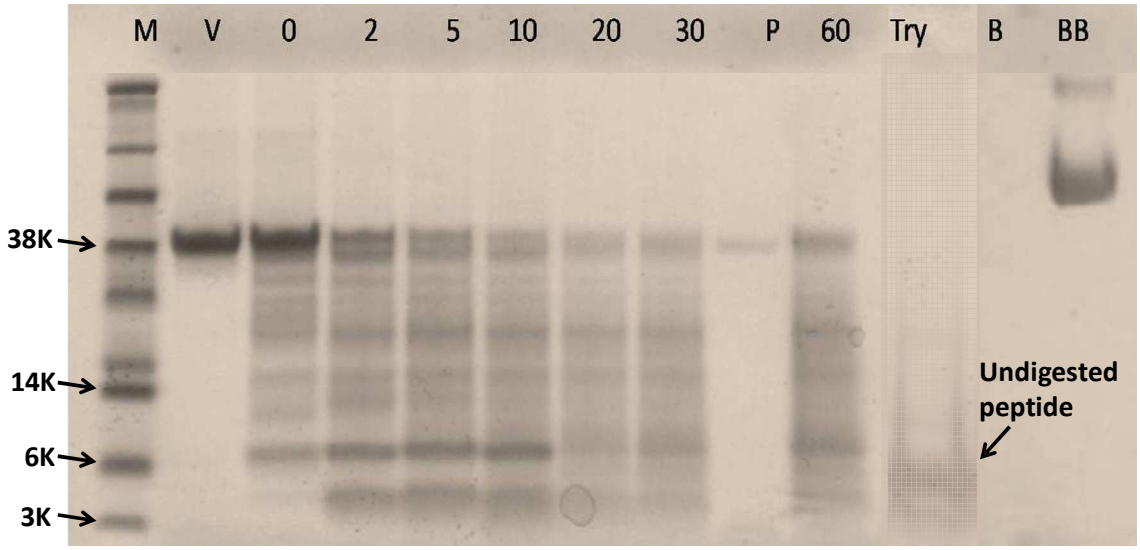


**Figure S5.** Combined SDS-PAGE analysis of RCNMV digestion using pepsin. For digestion, a ratio of 1:100 (pepsin:virus, wt:wt) was used. M=marker, V=virus, P=pepsin, Try= Trypsin digestion overnight, B=BSA, BB=BSA blank, and numbers=#minutes of digestion . Proteins were stained with coomassie blue.

GluC hydrolyzes proteins at glutamic acids and aspartic acids (much slower) in ammonium bicarbonate aspartic acids buffer at pH 4.0 or phosphate buffer at pH 7.8. GluC was explored as a possible avenue to increase capsid sequence coverage because there are several D and E residues in the region 176-286 where trypsin does not give adequate coverage. Unfortunately, GluC digestion did not meet our expectations. As indicated by SDS-PAGE analysis in Figure S6, GluC did not digest the capsid protein significantly in either buffer despite digesting overnight and using a 1:20 ratio of GluC:virus (wt:wt).In addition, no useful data were obtained by LC/MS$^{E}$ analysis.

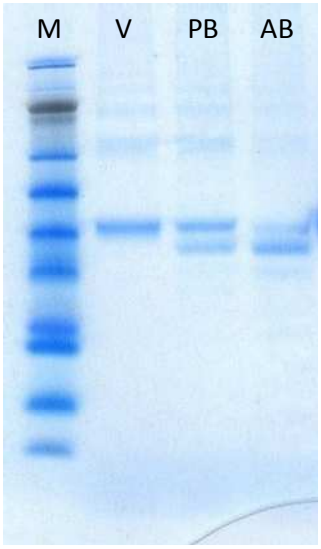


**Figure S6.** SDS-PAGE analysis of RCNMV digestion using GluC. For digestion, a ratio of 1:20 (GluC:virus, wt:wt) was used. M=marker, V=virus, PB=phosphate buffer, and AB= ammonium bicarbonate buffer. . Proteins were stained with coosmassie blue.