Characterization of the Agent of ‘High Plains Disease’

MASS SPECTROMETRY DETERMINES THE SEQUENCE OF THE DISEASE-SPECIFIC PROTEIN*

Yi-Min She†ø, Dallas L. Seifers§, Steve Haber¶, Werner Ens†, and Kenneth G. Standing† ††

From †Department of Physics & Astronomy, University of Manitoba, Winnipeg, MB R3T 2N2, Canada, øStructural Biology and Biochemistry, Hospital for Sick Children, Toronto, ON M5G 1X8, Canada, §Agricultural Research Center, Kansas State University, Hays KS 67601-9228, USA, ¶Cereal Research Centre, Agriculture & Agri-food Canada, Winnipeg, MB R3T 2M9, Canada.

*The work at the University of Manitoba was supported by grants from the US National Institutes of Health (GM59240), and the Natural Sciences and Engineering Research Council of Canada (NSERC).

The high plains virus amino acid sequences reported in this paper have been deposited in the Swiss-Prot and TrEMBL knowledgebase under the accession numbers P83549 and P83550.
††To whom correspondence should be addressed: Department of Physics & Astronomy, University of Manitoba, Winnipeg, MB R3T 2N2, Canada.
Tel: 204-474-9358; Fax: 204-474-7622; E-mail: standin@cc.umanitoba.ca.

Running Title: Characterizing Proteins of the High Plains Virus

Footnote: The abbreviations used are: @A, acetylated alanine; BLAST, Basic Local Alignment Search Tool; CID, collision-induced dissociation; HPV, high plains virus; ELISA, enzyme-linked immunosorbent assay; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; MALDI, matrix-assisted laser desorption ionization; TOFMS, time-of-flight mass spectrometry; MS/MS, tandem mass spectrometry.

SUMMARY

The “32 kDa” protein specifically associated with High Plains disease was characterized by time-of-flight mass spectrometry, after the agent had been isolated in pure culture by “vascular puncture inoculation”, a novel mechanical means of transmission. Two isolates from different geographic locations each consisted of a mixture of sub-populations that were highly homologous to an amino acid sequence derived from a nucleotide sequence (U60141) deposited in GenBank by the Nebraska group as “the probable N-protein of High Plains Virus”. However, the U60141 sequence was found to be incomplete; de novo sequencing of peptides produced by proteolytic digestions of the “32 kDa” band from an SDS–PAGE separation showed that an additional 18 amino acid residues were present at
the N-terminus. BLAST examination of the sequence showed no significant homology with any protein in the databases, indicating that the infectious agent of High Plains disease is likely a member of a hitherto unclassified virus group.

INTRODUCTION.

In 1993, a disease affecting certain maize and winter wheat lines was observed in several High Plains states of the U.S. Great Plains region (1, 2). By the end of 1995 its presence had been confirmed over a wide area (3). This “High Plains Disease” has a relatively extensive host range, and it can cause severe or lethal symptoms on susceptible maize, wheat, barley, and several species of grasses (4). The disease is consistently associated with a characteristic cytopathology in which spherical or ovoid 150-200 nm double-membrane vesicles accumulate in the cytoplasm of affected cells (5). Analysis of the nucleic acid complement of infected plants revealed five characteristic species of double-stranded RNA, and a specific ribonucleoprotein, whose protein component has been estimated by SDS-PAGE to have a mass of ~32 kDa (3). The agent of High Plains Disease is commonly referred to as High Plains Virus (HPV), although its status as a plant virus has not been confirmed by the International Committee on the Taxonomy of Viruses.

The disease is characterized by the development of chlorotic spots that parallel the leaf veins. As the disease progresses, the chlorotic spots merge to give a generalized chlorosis. However, symptoms vary greatly with different host genotypes, environments, plant growth stage at time of initial infection, and combination with other infecting
viruses. HPV is transmitted by the same wheat curl mite (*Aceria tosichella* Keifer) vector as wheat streak mosaic virus (3, 6, 7). The similarities in: a) transmission by eriophyid mite vectors; b) induction of spot mosaic symptoms; and c) formation of 150-200 nm double membrane particles in the cytoplasm of affected cells; suggest that HPV is a member of a group of virus-like plant pathogens that cause wheat spot mosaic, fig mosaic, thistle mosaic, rose rosette, and redbud yellow ringspot diseases (5).

No one has yet succeeded in isolating and characterizing infectious HPV particles. Until recently, there was no other way to propagate the disease agent except to use live wheat curl mites. Using wheat curl mites for experimental transmission and propagation approximates what happens to plants grown in the field, but makes it very difficult to obtain HPV infections free of wheat streak mosaic virus, as the vector usually harbors both this virus and HPV (8).

We demonstrate here that the “32 kDa” protein specifically associated with HPV can be characterized by time-of-flight mass spectrometry (TOFMS), following HPV isolation in pure culture by “vascular puncture inoculation”(9, 10; 7), a novel mechanical means of transmission.

**EXPERIMENTAL PROCEDURES**

*High Plains Virus (HPV) isolates* [] The HPV isolates to be characterized were obtained from symptomatic maize (*Zea mays* L.) collected in Kansas in 1996 (KS96) and Idaho in 1997 (ID97). Serological and infectivity assays were used to confirm the absence of other viruses (particularly wheat streak mosaic virus) in each HPV isolate. The identity of HPV was indicated by infectivity assay and SDS-PAGE. Confirmation
was obtained by ELISA and Western blotting, using antiserum developed and provided by S. G. Jensen (USDA/ARS, Lincoln, NE). Use of this antiserum permitted direct comparison of serological results from these investigations with those of previous studies of HPV (3, 4, 6-8). To ensure purity, the HPV isolates were maintained in culture in ‘Spirit’ corn by vascular puncture inoculation of kernels (9, 10; 7).

*Generation of HPV protein for analysis by TOFMS.* HPV-specific protein for mass spectrometric (MS) analysis was obtained after electrophoretic separation (SDS-PAGE), staining with Coomassie blue 250R, and excision of the ~32 kDa protein band. The excised HPV-specific protein band from each gel was digested with various endoproteinases as described by She *et al.* (11), using sequencing grade trypsin, Lys-C, Glu-C and Asp-N purchased from Roche Diagnostic Corporation (Indianapolis, IN). Enzymatic digestion was performed on the gel-separated proteins using either: 10 ng trypsin, 50 ng Lys-C, or Glu-C, in 25mM ammonium bicarbonate; or else 50 ng Asp-N in 10mM Tris-HCl (pH 7.6) solution. Peptide extractions with sonication usually yielded sufficient material for analyses covering most of the protein sequence.

*Derivatization.* To facilitate *de novo* peptide sequencing, we enhanced the production of y-ions from some of the selected peptides by derivatizing the in-gel tryptic digest with sulfonic acid. The peptides were dehydrated, then reacted with chlorosulfonylacetyl chloride, as described by Keough *et al.* (12). Improved yield of the sulfonated product was achieved by repeating the reaction three times.

*Mass Spectrometry and characterization of peptide amino acid sequences.* After digestion or chemical modification, the samples were analyzed by matrix-assisted laser
desorption ionization (MALDI) on a QqTOF mass spectrometer (13) as described previously (14, 15), using 2,5-dihydroxybenzoic acid (DHB) as matrix. MS/MS daughter ion spectra were obtained after collision-induced dissociation (CID) in the QqTOF collision cell. Protein or peptide sequences were then determined by manual interpretation of the MS and MS/MS measurements, and identified by database searching with the computer programs MS-Tag, BLAST, or FASTA.

RESULTS

Confirmation of the identity of HPV-specific protein

Protein extracts from infected tissue (propagated by vascular puncture inoculation) were shown by ELISA to react to an HPV antiserum (3). When examined by SDS-PAGE, the extracts yielded a ~32 kDa band (Fig. 1), which was not present in control extracts from uninfected tissue, and which reacted in Western blots to the specific antiserum.

Determination of the protein amino acid (aa) sequence

i) Deducing aa sequences from the initial tryptic digests

Figure 2 shows products from the tryptic digest of a gel slice containing the ‘32 kDa’ protein band extracted from isolate ID97. The spectrum is fairly complex, making analysis difficult. In addition, we believed initially that we were faced with a completely de novo sequencing problem, although this turned out to be only partially true. Thus it was clearly important to simplify the problem by choosing appropriate sequencing strategies.

Analysis of the prominent 2448/2476 Da “doublet” shown in Fig. 2 illustrates the difficulty. MS/MS measurements yielded similar daughter ion spectra in the lower m/z
range, from which we could extract putative $y_1$ to $y_{16}$ ions, indicating that the two parents were closely related. However, we were unable to find $y$ fragments between $y_{16}$ and the parent ion, or their b ion counterparts, so the most interesting parts of the peptides remained undefined.

Here we found sulfonation at the N-terminus (12) especially valuable. Fig. 3 shows the doublet before and after sulfonation. Although sulfonation has somewhat worsened the signal/noise ratio, the defining peaks are still well above background, and unit mass resolution is still obtained. Fig. 4 shows the corresponding MS/MS spectra. Thanks to sulfonation, these spectra can be interpreted directly as complete series of $y$ ions up to $y_{22}$ (as labeled in Fig. 4 and listed in Table II), permitting straightforward deduction of the corresponding aa sequences. Both peaks indeed correspond to the same peptide, apart from the two indicated single nucleotide mutations.

Similar analyses of a number of the peaks in Fig. 2 (with or without sulfonation) yielded the other peptide sequences exhibited in Table I; all were confirmed by MS/MS measurements. Searches of protein sequence databases revealed that some of these peptides corresponded to fragments of known ribosomal proteins, and some were homologous to $\beta$-glucosidase aggregating factor precursor, as indicated in the table. These proteins all have calculated masses in the range 27 to 32 kDa (Table I), so it is not surprising to find them in the rather large gel slice examined.

However, other peaks were more directly related to our investigation. We had hoped only that some of the observed peptides might serve to characterize HPV, and that perhaps there would be some homology between these putative HPV peptides and the recorded sequences of other viruses. Instead, we were pleasantly surprised to find that
most of the peptides corresponded almost exactly to portions of a specific amino acid sequence listed as AAB03575, and designated as the “The probable N-protein of High Plains Virus” (16), of which we had not previously been aware. AAB03575 had been deduced from the cDNA sequence (U60141) of a reverse-transcribed RNA component of an infectious isolate prepared by the Nebraska group, and deposited in GenBank, but not otherwise published. Thus our mass spectrometric measurements served to confirm the previously tentative identification of U60141 with High Plains Disease (see discussion below).

Unexpectedly, however, the tryptic peptide fragment at m/z = 3522.653 that included the N-terminal sequence deduced from U60141, had an additional single glutamic acid (D) residue at its N-terminus (Fig. 5), showing that the database-listed sequence was not complete. Moreover, several additional peaks had been observed and their corresponding peptide fragments sequenced de novo (Table I–part 4). While these sequences all corresponded to the same protein, its peptides had no closely matching counterparts in any protein database. Was this part of a separate protein, or was its sequence part of the ‘complete’ HPV N-protein sequence? None of the observed tryptic fragments overlapped the U60141-deduced aa sequence (Table I), leaving this question open.

**ii Additional proteolytic digests** —To decide between the two possibilities, we generated new sets of fragments using different endoproteinases (Asp-N and Glu-C). Fortunately this succeeded in generating peptides that also started from the N-terminus of the “unknown” protein but were considerably larger. Sequencing these peptides by MS and MS/MS measurements gave overlaps of as many as 13 residues with the sequence
predicted by U60141; see Table III. Thus an additional 18 residues were found to be part of the actual HPV aa sequence, so the sequence predicted by the nucleic acid data (U60141) is indeed incomplete (see discussion below).

**iii Correction of the N-terminal residue assignment** — One discrepancy remained. The assignment of N-terminal residues for the “new” peptides shown in Tables I and III was based on a nominal interpretation of the MS data that allowed an error tolerance up to ~50 mDa., but closer examination reveals inconsistencies in this interpretation. Note that all values of Δm for the “new” N-terminal peptides are negative, with an average value of about ~32 mDa, (Table III), whereas Δm for the collection of peptides listed in the upper part of Table I has an rms value ~7.8 mDa, and an average value <0.8 mDa, (similar to the values obtained in previous measurements on the same instrument (14, 15)). Such a discrepancy might easily be overlooked if less accurate MS techniques had been used, but it is clearly unacceptable for measurements with our QqTOF instrument.

MS/MS measurements on the “new” peptides yielded additional information. Fig. 6 shows the simplest example— the daughter ion spectrum from breakup of the 607 Da parent ion. Table IVa lists the difference between observed and calculated masses for the daughter ions on the basis of the nominal sequence assignment. All b ion masses differ from the calculated values by more than 30 mDa. By contrast, the masses of the y₁ to y₄ daughter ions all differ from calculation by 5 mDa or less, whereas the mass of y₅ differs by 37 mDa.. Results of MS/MS measurements on the other “new” peptides were consistent with this pattern (see supplemental figures S1-S5), unambiguously pointing to an error in the assignment of the first residue.
We then realized that the mass (113.048 Da) of acetylated alanine (@A) is almost identical to that of leucine/isoleucine (113.084 Da), yielding a plausible alternative assignment for the N-terminal residue. Table 4b shows the masses calculated assuming @A to be at the N-terminus. These all agree with the observed masses of the b and y ions within 5 mDa., and similar improvements were obtained in the Δm’s of Tables I and III — convincing evidence that acetylated alanine rather than leucine/isoleucine constitutes the actual N-terminus of the HPV protein. (We note that hydroxylation of proline yields a compound (hydroxyproline) with the same elemental composition as @A, and thus the same mass, but this is an unusual modification that would not be expected to occur at the N-terminus).

iv) Amino acid variations in the KS96 and ID97 HPV isolates

The sequence deduced from U60141 consists of 270 aa residues, whereas the sequences determined for both KS96 and ID97 contain, as noted above, an extra 18 residues at the N-terminus. An additional three isolates from different geographic locations (not described here) were also found to contain the same extra residues, so it appears that HPV normally contains 288 aa residues.

The overall results are shown in Table V. For both isolates, more than 85% of the residues were determined. The isolates each consist of a mixture of sub-populations, one of which contains an aa sequence corresponding directly to the one derived from U60141. Sub-populations with E, D or N at position 186 are present in both. Other variations that define isolate subpopulations are listed in Tables I and V.

In all cases, the C-terminal 270 aa residues of the HPV isolates are highly homologous to the U60141 sequence, and from Table V it can be seen that a consensus
sequence accounting for about 88% of the observed residues can be assembled.

**DISCUSSION**

Previous attempts to characterize the HPV agent have been hindered by the need to transmit the infectious agent with the wheat curl mite vector, which naturally transmits wheat streak mosaic virus along with HPV. Vascular puncture inoculation eliminates the problem, making it possible to establish pure cultures of HPV-infected plants, and thus ensuring that protein unique to infected plants is indeed disease-specific.

Our investigation demonstrates the power of combining this method of sample preparation with protein characterization by mass spectrometry. In the present case, the analysis was aided by the unanticipated discovery that our MS-determined sequence had extensive homology with an aa sequence that had been deduced from a partial cDNA of dsRNA extracted from ribonucleoproteins of diseased plants (U60141). Nevertheless, our MS analysis revealed that the U60141 sequence was incomplete, and indeed this might have been suspected from the absence of an open reading frame in U60141, likely arising from the inherent difficulties in generating full-length cDNA clones from dsRNA.

Although the MS-based approach to analyzing HPV-specific protein avoids these difficulties entirely, determining the complete protein sequence nonetheless required a thorough MS analysis. Database searching followed simply by protein identification would of course have yielded no information about the additional aa residues at the N-terminus, illustrating the limitations of the commonly used “high-throughput” measurements. Moreover, it was essential in this case to use an array of different proteinases to obtain digest fragments that would define the additional residues. This
emphasizes the need to adapt one’s strategy to observations as they are made. Another feature of the measurement was the ability to distinguish between leucine and acetylated alanine ($\Delta m=36$ mDa) as the N-terminal residue, which could only be accomplished with high mass accuracy MS. As an obvious consequence of this observation, we predict that a complete cDNA transcript of RNA from the HPV-specific ribonucleoprotein will contain codons for Ala-Leu/Ile-Ser… rather than Leu/Ile-Leu/Ile-Ser… corresponding to the protein’s N-terminal.

Several naturally-occurring sequence variants were readily characterized in these analyses as components of the overall population of HPV-specific protein molecules. This is a critical advantage of mass spectrometry over nucleic acid sequencing of viral nucleic acid cDNA, as the latter necessarily involves the analysis of solely a clonal sub-population.

Examination of the sequence by BLAST (17, 18) failed to find any significant homologies (except for U60141) in the protein- or translated nucleic acid sequence databases. Nevertheless, as mentioned above, biological evidence suggests that HPV is related to a group of unusual virus-like plant pathogens that cause wheat spot mosaic, fig mosaic, thistle mosaic, rose rosette, and redbud yellow ringspot diseases (5), none of which have yet been sequenced. On the basis of the present results, we believe that a good method of investigating such a relationship at the molecular level, is to extract protein fractions from plants affected by these diseases, and subject them to MS sequence analysis; this task is now much easier with a ‘type’ sequence (HPV) in hand. Such evidence will be important in addressing the outstanding questions of the taxonomic affiliations of this group of unusual plant pathogens.
Acknowledgements

We thank Jeff Ackerman, James McNabb, and Victor Spicer for excellent technical support.

References

1. Jardine, D. J., Bowden, R. L., and Jensen, S. G. (1994) *Phytopathology* **84**, 1117-1118 (Abstract)

2. Jensen, S. G., and Lane, L. C. (1994) *Phytopathology* **84**, 1158 (Abstract)

3. Jensen, S. G., Lane, L. C., and Seifers, D. L. (1996) *Plant Dis.* **80**, 1387-1390

4. Seifers, D. L., Harvey, T. L., Martin, T. J., and Jensen, S. G. (1998) *Plant Dis.* **82**, 875-879

5. Ahn, K. K., Kim, K. S., Gergerich, R. C., and Jensen, S. G. (1998) *J. Submicrosc. Cytol. Pathol.* **30**, 563-571

6. Seifers, D. L., Harvey, T. L., Martin, T. J., and Jensen, S. G. (1997) *Plant Dis.* **81**, 1161-1166

7. Seifers, D. L., Harvey, T. L., Louie, R., Gordon, D.T., and Martin, T. J., (2002) *Plant Dis.* **86**,138-142

8. Marçon, A., Kaeppler, S. M., Jensen, S. G., Senior, L. and Stuber, C. (1999) *Crop Sci.* **39**, 1171-1177

9. Louie, R., (1995) *Phytopathology*, **85**, 139-143
10. Louie, R. and Seifers, D. L. (1996) *Phytopathology*, **86**, S1. (Abstract)

11. She, Y. M., Wang, G. Q., Loboda, A., Ens W., Standing, K. G. and Burczynski, F. J. (2002) *Anal. Biochem*. **310**, 137-147

12. Keough, T., Youngquist, R. S. and Lacey, M. P. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 7131-7136

13. Loboda, A. V., Krutchinsky, A. N., Bromirski, M., Ens, W. and Standing, K. G. (2000) *Rapid Commun. Mass Spectrom.* **14**, 1047-1057

14. Seifers, D. L., Salomon, R., Marie-Jeanne, V., Alliot, B., Signoret, P., Haber, S., Loboda, A., Ens, W., She, Y. M. and Standing, K. G. (2000) *Phytopathology* **90**, 505-513.

15. She, Y. M., Haber, S., Seifers, D.L., Loboda, A., Chernushevich, I., Perreault, H., Ens, W. and Standing, K. G. (2001) *J. Biol. Chem.* **276**, 20039-20047

16. Hall, J. S., Jensen, S. G. and French, R. C. (1996) “The Nucleotide Sequence of RNA3 - the probable N-protein of High Plains Virus” Genbank Accession No. U60141.1.

17. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389-3402

18. Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A., Bork, P., Ens, W., and Standing, K. G. (2001) *Anal Chem.* **73**, 1917-1926.
Figure Captions

Fig. 1  SDS-PAGE profile of the ID97 and KS96 high plains virus (HPV) isolates. Proteins were concentrated from ‘Spirit’ corn by minipurification, and separated by SDS-PAGE. Lanes 1 and 2 contain corn infected by the KS96 and ID97 HPV isolates respectively, lane 3 is healthy corn, and lane 4 shows molecular mass standards; (numbers to the right of the standards are in kDa). Arrows in lanes 1 and 2 mark the location of the ~32-kDa protein associated with HPV.

Fig. 2 Peptide mapping of the in-gel tryptic digest of the HPV-ID97 isolate. Peaks are labeled according to the categories subsequently identified, shown in Table I: 1. Ribosomal Proteins; 2. Beta-glucosidase aggregating factor precursor; 3. Putative High Plains Virus Protein U60141; 4. Unknown protein (not found in database).

Fig. 3 Expanded MALDI MS spectra of the in-gel tryptic digest of the HPV-ID97 isolate: A– before derivatization; B– after derivatization by sulfonic acid.

Fig. 4 MS/MS spectra of the Fig. 3B doublet, i.e., daughter ion spectra after sulfonic acid derivatization and CID: A– from the 2570 Da parent; B– from the 2598 Da parent. The
sequence predicted by U60141 (aa residues 32-53) is shown above. These correspond to residues 50-71 of the revised sequence shown in Table V.

Fig. 5 MS/MS spectrum of the peptide ion at m/z 3522.653. obtained from a tryptic digestion. Analysis shows that the peptide contains the sequence DGFVTSSDIETTVHDFSYEKPDLSVDGFSKL, i.e. there is an additional aspartic acid residue (D) at the N-terminus of the HPV protein sequence predicted by U60141.

Fig. 6 MS/MS spectrum of the 607 Da tryptic N-terminal peptide after CID. The daughter ions include a complete set of both N-terminal b ions and C-terminal y ions, defining the peptide sequence as acetylated ALSFK (see Table III and text). Daughter ions from other N-terminal parents are shown in supplemental figures S1-S5.
FIG 3
FIG 4
### Table I
Identification of Peptides from the in-gel tryptic digest of the HPV-ID97 isolate

| m/z(Obs.) | MH+(Calc.) | [m (mDa)] | Peptide sequence | Residues |
|----------|------------|-----------|-----------------|----------|
| 1        |            |           |                 |          |
| 1. Ribosomal proteins:               |            |           |                 |          |
| a. 40S S4. Accession number: 7440143, MW: 30171.5 Da | 1268.633   | 1268.628 +5 | [R]GIPYLNTYDGR[T] | 135-145  |
|         | 2043.042   | 2043.030 +12 | [R]SVQFGQKIGIPYLNTYDGR[T] | 128-145  |
|         | 2261.169   | 2261.165 +4 | [R]HWMLDKLGGAFAPKPGPHK[S] | 17-37    |
| b. 40S S6. Accession number: 22296425, MW: 28472.3 Da | 1040.647   | 1040.647 0 | [R]LVTPLTLQR[K] | 183-191  |
| c. 60S L2. Accession number: 266944, MW: 28270.8 Da | 1059.576   | 1059.585 -9 | [R]VTFRHPFR[F] | 61-68    |
|         | 1079.535   | 1079.538 -3 | [K]AGNAHYKSYR[V] | 182-190  |
|         | 2696.295   | 2696.298 -3 | [R]GVAMNPVEHPHGGNGNHQHASTVR[R] | 201-226  |
|         | 2852.410   | 2852.399 +11 | [R]GVAMNPVEHPHGGNGNHQHASTVR[R] | 201-227  |
| d. 60S L7A. Accession number: 542158, MW: 29304.0 Da | 1354.643   | 1354.639 +4 | [K]ANFNDKDDEVV[K] | 216-226  |
|         | 1383.816   | 1383.811 +5 | [R]LKVPPALNQFT[R] | 67-78    |
|         | 1733.979   | 1733.981 -2 | [K]QFQGGGALPPKKDLHR[F] | 31-46    |
| e. 60S L10A. Accession number: 20521241, MW: 27101.4 Da | 1562.910   | 1562.917 -7 | [R]FSGSVKLPHIPRPK[M] | 48-61    |
|         | 1966.976   | 1966.962 +14 | [K]VGPGWGGSGGSPVDITAEPQR[L] | 163-182  |
|         | 2208.126   | 2208.141 -15 | [K]VGPGWGGSGGSPVDITAEPQR[L] | 163-184  |
|         | 2411.231   | 2411.236 -5 | [K]TPIGGTVVNNWAVYDGPQMN[L] | 44-67    |
|         | 2586.314   | 2586.322 -8 | [R]AQGLHIQAGNWVNSFSLFVDQR[F] | 72-94    |
|         | 2858.452   | 2858.454 -2 | [K]SITVATGIAVTSEIFASYVDSAGQTQSAGR[W] | 185-213  |
| 2. Beta-glucosidase aggregating factor precursor. Accession number: 9313027, MW: 31792.0 Da | 1179.638   | 1179.637 +1 | [K]KYEEIVKDR[S] | 122-135  |
|         | 1716.720   | 1716.718 +2 | [K]NTINFNEDDFMNK[L] | 122-135  |
|         | 1771.935   | 1771.917 +18 | [K]IKYNHESDLVAPMVR[Q] | 183-197  |
|         | 1772.918   | 1772.901 +17 | [K]IKYNHESDLVAPMVR[Q] | 183-197  |
| 3. High plains virus. Accession number: 1407699, MW: 31374.7 Da | 585.290    | 585.290 0 | [R]WSHK[V] | 175-178  |
|         | 780.406    | 780.414 -8 | [K]YEEIVK[D] | 221-226  |
|         | 851.438    | 851.438 0 | [K]NVVSYLN[F] | 86-92    |
|         | 908.504    | 908.509 -5 | [K]YEEIVK[D] | 220-226  |
|         | 1050.544   | 1050.548 -4 | [K]VFNSWSKR[Y] | 109-116  |
|         | 1054.573   | 1054.580 -7 | [K]WHRVELK[I] | 175-182  |
|         | 1179.638   | 1179.637 +1 | [K]YEEIVKDR[S] | 220-228  |
|         | 1716.720   | 1716.718 +2 | [K]NTINFNEDDFMNK[L] | 122-135  |
|         | 1771.935   | 1771.917 +18 | [K]IKYNHESDLVAPMVR[Q] | 183-197  |
|         | 1772.918   | 1772.901 +17 | [K]IKYNHESDLVAPMVR[Q] | 183-197  |
|         | 1786.927   | 1786.916 +11 | [K]IKYEHESDLVAPMVR[Q] | 183-197  |
|         | 1815.086   | 1815.078 +8 | [K]LTFFDIYVGKLKN[K] | 68-83    |
|         | 1944.063   | 1944.059 +4 | [K]RNISDVMIDIVGKDVIAK[K] | 202-219  |
|         | 2448.218   | 2448.216 +2 | [K]SLLSDDGWHVVAQSVTNSER[L] | 32-53    |
|         | 2476.239   | 2476.247 -8 | [K]SLLTSEGWHVVAQSVTNSER[L] | 32-53    |
|         | 2920.442   | 2920.447 -5 | [K]YNILEDKDFKINIDSSLSTIK[N] | 237-261  |
|         | 2933.476   | 2933.478 -2 | [K]YNILEDKDFKINIDSSLSTIK[N] | 237-261  |
|         | 2948.570   | 2948.577 -7 | [R]GNISDVMIDIVGKDIIAKKYYEIVKDR[S] | 203-228  |
|         | 3522.653   | 3522.644 +9 | DGFVTSDDIETTVHDFSYEKPDLSVDGFSKL[S] | (-1)-31  |
| Mass (m/z) | Expected Mass | Monoisotopic Mass | Peptide Sequence | Status |
|-----------|---------------|-------------------|-----------------|--------|
| 607.350   | 607.382       | LLSFK^k           | ?               |        |
| 1292.723  | 1292.758      | LLSFKNSSGVLK^k    | ?               |        |
| 1491.856  | 1491.890      | LLSFKNSSGVLKAK^k  | ?               |        |

a. Measured by MALDI QqTOF mass spectrometry.
b. Protonated monoisotopic mass calculated for the peptide sequence shown.
c. Numbered according to the protein sequence in the NCBI database.
d. #174 M V
e. #174 M V
f. #44 A T; #57 T A; #59 C Y
g. #186 E N
h. #186 E D
i. #35 S T; #37 D E
j. #252 D N; #254 S T
k. L means Leu or Ile, and the N-terminal L is actually acetylated Ala (@A)--see text.
m. Masses (and Δm values) were calculated for the incorrect nominal assignment just noted.
Table II
Daughter ions from CID fragmentation of parent ions:
m/z= 2570.189 (SLLSSDGWHIVVAYQVTSNER)
m/z= 2598.218 (SLLTSEGWHIVVAYQVTSNER)

| y ion  | m/z(Obs.) | MH+(Calc.) | △m(mDa) | y ion  | m/z(Obs.) | MH+(Calc.) | △m(mDa) |
|--------|-----------|------------|---------|--------|-----------|------------|---------|
| y1     | 175.108   | 175.119    | -11     | y1     | 175.107   | 175.119    | -12     |
| y2     | -         | 304.162    | -       | y2     | -         | 304.162    | -       |
| y3     | 391.189   | 391.194    | -5      | y3     | 391.194   | 391.194    | 0       |
| y4     | 505.242   | 505.237    | +5      | y4     | 505.230   | 505.237    | -7      |
| y5     | 606.289   | 606.284    | +5      | y5     | 606.294   | 606.284    | +10     |
| y6     | 705.349   | 705.353    | -4      | y6     | 705.352   | 705.353    | -1      |
| y7     | 792.380   | 792.385    | -5      | y7     | 792.384   | 792.385    | -1      |
| y8     | 920.432   | 920.444    | -12     | y8     | 920.449   | 920.444    | +5      |
| y9     | 1083.496  | 1083.507   | -11     | y9     | 1083.509  | 1083.507   | +2      |
| y10    | 1154.545  | 1154.544   | +1      | y10    | 1154.543  | 1154.544   | -1      |
| y11    | 1253.620  | 1253.612   | +8      | y11    | 1253.607  | 1253.612   | -5      |
| y12    | 1352.681  | 1352.682   | 0       | y12    | 1352.682  | 1352.682   | +1      |
| y13    | 1465.760  | 1465.765   | -5      | y13    | 1465.769  | 1465.765   | +4      |
| y14    | 1602.832  | 1602.824   | +8      | y14    | 1602.836  | 1602.824   | +12     |
| y15    | 1788.906  | 1788.903   | +3      | y15    | 1788.907  | 1788.903   | +4      |
| y16    | 1845.931  | 1845.925   | +6      | y16    | 1845.937  | 1845.925   | +12     |
| y17    | 1960.960  | 1960.952   | +8      | y17    | 1974.971  | 1974.967   | +4      |
| y18    | 2047.998  | 2047.984   | +14     | y18    | 2061.998  | 2061.999   | -1      |
| y19    | 2135.006  | 2135.016   | -10     | y19    | 2163.035  | 2163.047   | -12     |
| y20    | 2248.085  | 2248.100   | -15     | y20    | 2276.123  | 2276.131   | -8      |
| y21    | 2361.181  | 2361.184   | -3      | y21    | 2389.216  | 2389.215   | +1      |
| y22    | 2448.201  | 2448.216   | -15     | y22    | 2476.242  | 2476.247   | -5      |
Table III
Proteolytic fragments from the HPV protein
LLSFKNSSGVLKAKTLKDGFVTSDDIETTVHDFSYEKPDLSVDGFSLKS...

| Enzyme | n/z (Obs.) | m/z (Calc.) | △m (mDa) | Fragment sequence |
|--------|------------|-------------|----------|------------------|
| Trypsin | 3522.653 | 3522.644 | + 9 | DGFVTSDDIETTVHDFSYEKPDLSVDGFSLKV |
| Asp-N  | 3322.769 | 3322.789 | - 20 | LLSFKNSSGVLKAKTLKDGFVTSDDIETTVH |
| Glu-C  | 2884.534 | 2884.567 | - 33 | LLSFKNSSGVLKAKTLKDGFVTSDDIE |
| Asp-N  | 1834.080 | 1834.117 | - 37 | LLSFKNSSGVLKAKTLK |
| Trypsin | 1491.856 | 1491.890 | - 34 | LLSFKNSSGVLKAK |
| Trypsin | 1292.723 | 1292.758 | - 35 | LLSFKNSSGVKL |
| Trypsin | 607.350  | 607.382  | - 32 | LLSFK |

The N-terminal portion of the original U60141 sequence is shown shaded.
Table IV

*Daughter ions from the N-terminal tryptic peptide parent ion at m/z 607. 345*

a Comparison with mass values calculated from the sequence LLSFK

| b ion | m/z (Obs.) | MH+ (Calc.) | △m(mDa) | y ion | m/z (Obs.) | MH+ (Calc.) | △m(mDa) |
|-------|------------|-------------|---------|-------|------------|-------------|---------|
| b1    | -          | 114.092     | -       | y1    | 147.109    | 147.113     | -0.004  |
| b2    | 227.138    | 227.176     | -0.038  | y2    | 294.184    | 294.182     | +0.002  |
| b3    | 314.176    | 314.208     | -0.032  | y3    | 381.217    | 381.214     | +0.003  |
| b4    | 461.244    | 461.276     | -0.032  | y4    | 494.303    | 494.298     | +0.005  |
| b5    | 589.339    | 589.371     | -0.032  | y5    | 607.345    | 607.382     | -0.037  |

△m (average) = -33 mDa

Δm (average y1-y4) = +1.5 mDa

b Comparison with mass values calculated from the sequence @ALSFK

| b ion | m/z (Obs.) | MH+ (Calc.) | △m(mDa) | y ion | m/z (Obs.) | MH+ (Calc.) | △m(mDa) |
|-------|------------|-------------|---------|-------|------------|-------------|---------|
| b1    | -          | 114.056     | -       | y1    | 147.109    | 147.113     | -0.004  |
| b2    | 227.138    | 227.140     | -0.002  | y2    | 294.184    | 294.182     | +0.002  |
| b3    | 314.176    | 314.172     | +0.004  | y3    | 381.217    | 381.214     | +0.003  |
| b4    | 461.244    | 461.240     | +0.004  | y4    | 494.303    | 494.298     | +0.005  |
| b5    | 589.339    | 589.335     | +0.004  | y5    | 607.345    | 607.346     | -0.001  |

△m (average) = 2.5 mDa

Δm (average) = 1 mDa

@ indicates N-terminal acetylation.
Table V

Comparison of the measured amino acid sequences of the HPV protein isolates ID97 and KS96 with the sequence predicted by U60141

| Position | ID97 | KS96 |
|----------|------|------|
| 001      | @ALSFKNSSGV LKAHLKDGDF VTSSDIETTV HDFSYEKPDLS SSVDFSLKS | ----- | ----- |----- |----- |
| 051      | LLLGDWFHIV VAYQSTNSEE RLNNKKKNKK TQRKFLFTFD IIVIPGLKPN | ----- | ----- |----- |----- |
| 101      | KSKNVSYSNR FMALCIGMIC YHKWKVFNW SNKRYEDNKN TINFNEDDF | ----- | ----- |----- |----- |
| 151      | MNKLAMSAGF SKEHKYHWFY STGFETYFDI FPAEVIAMSL FRWHSRVELK | ----- | ----- |----- |----- |
| 201      | IKEYEHESDLV APMVRQVTKR GNISDVMDI VGDJIAKKYE EIVKDRSSIG | ----- | ----- |----- |----- |
| 251      | IGTKYNILD EFKDIIFKID SSLSSTIKN CFNKLIDGE | ----- | ----- |----- |----- |

The top line shows the amino acid sequence deduced from U60141 (NCBI database accession number 1407699), with the addition of the 18 residues determined by de novo sequencing in the present measurements (shaded region); @ indicates N-terminal acetylation. Dashes indicate agreement with the sequence predicted by U60141. Observed sub-populations are indicated as follows: §§§ for SSD or TSE; ‡ for E, N or D; ∑ for I or T; ¥¥¥ for DSS or NST; †† for ID or T.
Figures submitted as supplemental materials: further mass spectrometric evidence for the N-terminal addition of 18 amino acid residues to the protein sequence predicted by U60141.

Supplemental Figure Captions

Fig. S1 MS/MS daughter ion spectrum from CID of the parent ion at m/z 1292.723, obtained from a tryptic digestion. Analysis shows that the peptide contains the sequence @ALSFKNSSGVLK, where @ indicates N-terminal acetylation, and L indicates Leucine or Isoleucine.

Fig. S2 MS/MS spectrum of the peptide ion at m/z 1491.856, obtained from a tryptic digestion. Analysis shows that the peptide contains the sequence @ALSFKNSSGVLKAK, where @ indicates N-terminal acetylation, and L indicates Leucine or Isoleucine.

Fig. S3 MS/MS spectrum of the peptide ion at m/z 1834.080, obtained from an Asp-N digestion. Analysis shows that the peptide contains the sequence @ALSFKNSSGVLKAKTLK, where @ indicates N-terminal acetylation, and L indicates Leucine or Isoleucine.

Fig. S4 MS/MS spectrum of the peptide ion at m/z 2884.534, obtained from a Glu-C digestion. Analysis shows that the peptide contains the sequence
@ALSFKNSSGVLKAKTLKDGFVTSSDIE, where @ indicates N-terminal acetylation, and L indicates Leucine or Isoleucine.

Fig. S5 MS/MS spectrum of the peptide ion at m/z 3322.769, obtained from an Asp-N digestion. Analysis shows that the peptide contains the sequence @ALSFKNSSGVLKAKTLKDGFVTSSDIETTVH, where @ indicates N-terminal acetylation, and L indicates Leucine or Isoleucine.
FIG S1
FIG S2
FIG S3
FIG S4
FIG S5

Counts/bin

m / z

b_3

y_4

b_5-H_2O

b_6

y_6

b_7

y_7

b_8

y_8

b_9

y_9

b_10

y_10

b_11

y_11

b_12

y_12

b_13

y_13

b_14

y_14

b_15

y_15

b_16

y_16

b_17

y_17

b_18

y_18

b_19

y_19

b_20

y_20

b_21

y_21

b_22

y_22

b_23

y_23

b_24

y_24

b_25

y_25

b_26

y_26

b_27

y_27

b_28

y_28

b_29

y_29

b_30

y_30

b_31

y_31

@ALSFKNSSCVLKAKTLKDGFVTSSDIETTVH
Characterization of the agent of 'High Plains Disease': Mass spectrometry determines the sequence of the disease-specific protein
Yi-Min She, Dallas L Seifers, Steve Haber, Werner Ens and Kenneth G Standing

J. Biol. Chem. published online October 14, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M308506200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2003/10/30/M308506200.DC1