

---

---

# Identification of Amino Acid Substitutions in Avian Influenza Virus (H5N1) Matrix Protein 1 by Using Nanoelectrospray MS and MS/MS

Ning Liu,<sup>a</sup> Wenjun Song,<sup>b</sup> Kim-Chung Lee,<sup>a</sup> Pui Wang,<sup>b</sup> Honglin Chen,<sup>b</sup> and Zongwei Cai<sup>a</sup>

<sup>a</sup> Department of Chemistry, Hong Kong Baptist University, Hong Kong SAR, China

<sup>b</sup> State Key Laboratory for Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, Hong Kong SAR, China

---

Matrix protein 1 (M1), the major structural protein of the avian influenza virus, plays a critical role in regulation of viral RNA transcription via interaction with RNA and transportation of RNP cores. Mutations in M1 have been frequently observed in the highly virulent avian influenza H5N1 virus, which might be crucial to the pathogenic function. Here we report the characterization of mutated peptides in M1 purified from highly pathogenic avian influenza virus H5N1 by nanoelectrospray MS and MS/MS analyses on a quadrupole-time-of-flight mass spectrometer (Q-TOFMS). The specificity of tandem mass spectrometry allowed the identification of six amino acid (AA) substitutions in M1, including R95K, A166V, I168T, N207S, N224S, and R230K. Two commonly observed modifications such as oxidation and deamidation were accurately assigned in the protein. Bioinformatics analysis suggested some relationship between the amino acid substitution and structural property of M1 protein. Discussions on de novo sequencing of MS/MS spectra, especially in dealing with the AA substitutions, were provided. (J Am Soc Mass Spectrom 2009, 20, 312–320) © 2009 American Society for Mass Spectrometry

---

---

The influenza A virus causes human infection every year and continues to be a great threat against human health [1]. Among the various subtypes of influenza A virus, avian H5N1, one of the most highly pathogenic subtypes, has been found to cross species barrier and to result in human infection in recent years, including the 1997 and 2003 outbreaks in Hong Kong [2]. The avian influenza virus can acquire the ability for transmission to humans, mainly as a result of adaptive mutation or genetic reassortment in the strains. Matrix protein 1 (M1), the most abundant structural protein in influenza A virus, consists of 252 amino acids and plays vital roles in the virus life cycle, by regulating the transportation of ribonucleoprotein (RNP) cores into or out of the nucleus [3, 4] and controlling the termination of viral RNA synthesis through RNA binding [5, 6]. Although being the most conserved in the eight structural proteins, M1 has been frequently observed to exhibit some mutations at several sites, especially in avian H5N1 subtypes [7]. These mutations often resulted in AA substitutions in the primary structure of M1 and conferred some new properties by altering the hydrophobicity as well as the

secondary structures. It has also been indicated that some mutations in certain sites in M1 were crucial to pathogen function [8–10].

Analysis of antigens from influenza A virus by mass spectrometry-based approaches was reported [11–13]. MALDI-TOFMS has been applied to identify epitopes [11, 12] and post-translational modifications in influenza antigens [13]. In recent years, several studies reported mass spectrometric analysis of the mutation at the protein level [14–17]. The advantage of soft ionization techniques such as ESI and MALDI has allowed the characterization of large biomolecules such as proteins with high sensitivity and high accuracy. Using hemoglobin variants as model protein, a variety of mutations were successfully determined from MS analyses [14, 15]. A novel *cis* double mutation in transthyretin was determined by using liquid chromatography-electrospray ionization Fourier-transform ion-cyclotron-resonance mass spectrometry [16]. Structural studies of two allelic wheat glutenin subunits revealed a possible single amino acid substitution in subunit 1Bx20 [17].

We herein identified six AA substitutions in M1 protein by using nanoelectrospray MS and MS/MS. Two common modifications such as the oxidation at methionines and deamidation at asparagines were identified, and the modification sites were determined. The obtained results indicated that nanoelectrospray

---

Address reprint requests to Dr. H. Chen, State Key Laboratory for Emerging Infectious Diseases, Department of Microbiology, University of Hong Kong, Hong Kong SAR, China. E-mail: [hchen@hkucc.hku.hk](mailto:hchen@hkucc.hku.hk)  
Dr. Z. Cai, Department of Chemistry, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China. E-mail: [zwcai@hkbu.edu.hk](mailto:zwcai@hkbu.edu.hk)

ionization Q-TOFMS provided the capability to characterize the AA substitutions in M1 protein.

## Experimental

### *Chemicals and Materials*

HPLC grade ACN and methanol were from Fisher (Fairlawn, NY). Sequencing grade trypsin was obtained from Promega (Madison, WI). All other chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Avian influenza virus A/Dk/HN/303/2004(H5N1) [18] was propagated in a biosafety level 3 (BL-3) containment facility. The genome of this strain of duck-origin influenza virus was partially sequenced, which has been preserved in State Key Laboratory for Emerging Infectious Diseases, The University of Hong Kong.

### *Virus Cultivation*

Avian influenza virus strain A/Dk/HN/303/2004 (H5N1) was harvested from allantoic fluid of chicken embryos inoculated as 10-day old embryos and collected 48 h later. Several serial passages in chicken embryos were carried out to promote potential mutations in virus. Finally, due to the highly pathogenicity the virus possess, the virus-containing allantoic fluid was inactivated with 0.03% formalin at 4 °C for 72 h. All the experiments using the active virus were carried out in a biosafety level 3 laboratory. The resulting inactivated virus-containing allantoic fluid was then subjected to further experiments.

### *Isolation of Virus by Ultracentrifugation*

The virus-containing allantoic fluid was cleared from large debris by centrifugation at 5000 *g* for 30 min. The virus was collected from the supernatant by pelleting it through a 5-step discontinuous gradient cushion consisting of 20, 30, 40, 50, and 60% (wt/vol) sucrose, in a Beckman Rotor SW40 Ti at 155,000 *g* (35,000 rpm) at 4 °C for 60 min. The resulting virus band between 40% and 50% sucrose was carefully aspirated out from the SW40 tubes, followed by dilution with four volumes of buffer (100 mM Tris, 10 mM MgCl<sub>2</sub>, pH 7.4) to form a suspension. The virus was then applied to a 2-step discontinuous gradient of 30% and 60% (wt/vol) sucrose in Ultra clear SW60 tubes. Centrifugation was run in a Beckman Rotor SW60 Ti at 38,800 rpm at 4 °C for 60 min. The virus band was carefully collected, suspended in and then freeze at –80 °C until use. The purity of the virus was checked by SDS-PAGE.

### *Purification of M1 Protein*

Matrix protein 1 was purified from A/Dk/HN/303/2004(H5N1) by extraction with acidic chloroform-methanol, as described by Gregoriades [19]. Protein concentration was assayed by using the DC RC

protein assay kit (Bio-Rad, Hercules, CA, USA) with BSA as a standard protein.

### *SDS-PAGE*

Electrophoretic analyses were made in a Mini-Cell system (Bio-Rad, Hercules, CA), and run in 12% tris-glycine-SDS polyacrylamide gels with a 5% stacking gel at a constant voltage setting of 100 V. Samples were prepared by dilution with reducing sample loading buffer and boiling at 95 °C for 5 min. After electrophoretic separation, the gels were stained with colloidal Coomassie G250 and scanned with a calibrated densitometer (GS800, Bio-Rad).

### *In-Solution Digestion*

The protein sample (approximate 1.0 ug protein) was prepared by dilution with equal volume of 25 mM ammonium bicarbonate. After cysteines in M1 were reduced and then alkylated, the sample was digested by addition of sequencing grade modified porcine trypsin (Promega, Madison, WI) in a trypsin to protein weight ratio of 1:20. The mixture was left for 16 h or overnight at 37 °C. The resulting tryptic peptide mixture was stored at –20 °C until analysis.

### *Nanoelectrospray Mass Spectrometric Analysis*

After desalted and purified by using ZipTip C<sub>18</sub> pipette tips (Millipore, Bedford, MA), the tryptic peptides were analyzed on a quadrupole orthogonal acceleration time-of-flight mass spectrometer (QSTAR; Applied Biosystems, CA) equipped with an external nanoelectrospray ion source (Protana A/S, Odense, Denmark). Renin substrate was used to calibrate the instrument in the both MS and MS/MS mode. Approximately 2 μL of sample was loaded into a PicoTip emitter (New Objective, Woburn, MA). After the full mass spectra (parent ions) of tryptic peptides were obtained in TOFMS mode, the charge states of the peptides were assigned from the observed stable isotope pattern. The parent ions of interest were then subject to sequence analysis in product ion mode. In the product ion mode acquisition, the resolution of Q1 was typically set at a unit to insure the “purity” of the selected ions as long as their intensity was high enough. Manual de novo sequencing of peptide tandem mass spectra was performed with the aid of Pepsea (1.1) in Analyst QS 1.1 software (MDS SCIEX, Concord, Canada).

### *Bioinformatics Analysis of the AA Substitutions Identified in M1 Protein*

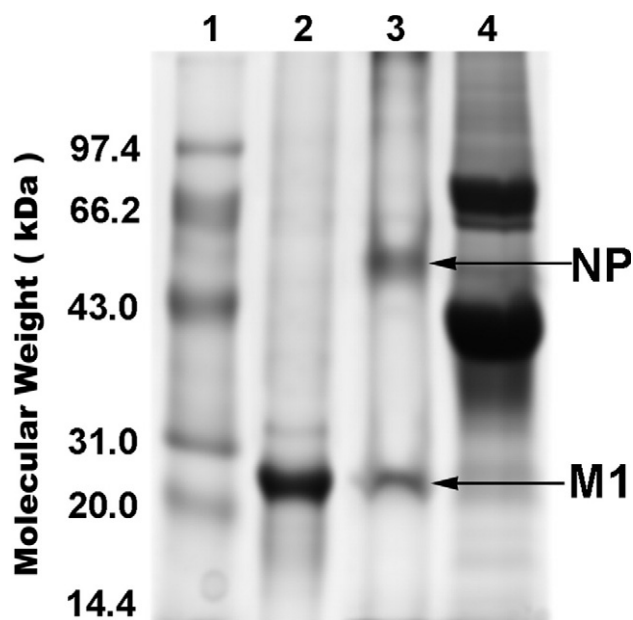
Hydrophobicity values of the six AA substitutions were calculated using ProtScale at ExPASy [20] with the Kyte-Doolittle hydrophobicity scale. Secondary structures were predicted by using nnPREDICT [21]. Protein stability changes upon mutations were predicted by using I-Mutant 2.0 [22].

## Results and Discussion

### Purification and Identification of M1 Protein

The cultivation of influenza virus in chicken embryos was performed in several serial passages to improve the chance of multigenic mutations [23]. The purification of virus particles from allantoic fluid was carried out by using ultracentrifugation in discontinuous sucrose cushion. After pelleted in SW40 rotor, the virus particles were subjected to a further purification in a small SW60 rotor to eliminate the possible contaminations from allantoic fluid. Finally, M1 protein was extracted from the purified virus by using acidic chloroform-methanol. The purity of M1 protein was checked by SDS-PAGE (Figure 1). The purified M1 protein was then subjected to tryptic digestion and mass spectrometric analysis.

The tryptic digests were analyzed by using nanoelectrospray MS and MS/MS. After the MS spectrum of the tryptic digests was recorded, the MS/MS analysis of each peak was performed. With the aid of MASCOT and Pepsea software, the peptides that had not undergone mutation were identified. A MASCOT score of 418 was obtained after the database searching against the SwissProt database in the entries of other viruses on a local Mascot server with the following parameters: one missed cleavage, fixed cysteine carbamidomethylation, and variable modifications such as asparagine/glutamine deamidation and methionine oxidation. Careful evaluation of the available data also allowed the identification of several mutated peptides. Accordingly, a



**Figure 1.** Purification of M1 from allantoic fluid. The virus particles were subject to ultracentrifugation in sucrose cushion to obtain purified virus. M1 protein was extracted by acidic chloroform-methanol from the purified virus. Samples were examined on a 12% SDS-PAGE, followed by colloidal Coomassie G250 staining. Lane 1: marker, lane 2: purified M1, lane 3: purified virus, lane 4: allantoic fluid.

total of 19 tryptic sequences (T1–T19) were identified, which allowed the assignment of 14 unique peptide sequences (Table 1). Sequence coverage of 70% was obtained. By performing external calibration using rein substrate, the Q-TOFMS analysis of the peptides was achieved with mass accuracy within 27 ppm for the  $m/z$  values of monoisotopic peaks.

### Identification of R95K Substitution

Through the careful inspection of the available MS/MS spectra, some spectra with very similar fragmentation pattern and/or ion series were found and sorted into individual groups. Two doubly charged ion peaks at  $m/z$  933.40 (T5) and 925.41 (T6), as well as three triply charged ion peaks at  $m/z$  622.61 (T5), 629.93 (T5), and 617.28 (T6) displayed similar MS/MS spectra. A comprehensive analysis of their MS/MS spectra allowed determination of the primary sequence of the peaks at  $m/z$  617.28 and 925.41 as FVQNALDGDGDPNNMDK (Figure 2). The peaks at  $m/z$  622.61 and 933.40 were identified as the peptide similar to that from  $m/z$  617.28 and 925.41 except with the methionine oxidized. The ion at  $m/z$  629.93 was the sodium adduct of the ion at  $m/z$  622.61. Among these five signals observed, the ion at  $m/z$  622.61 displayed the highest intensity and thus was selected for the subsequent investigation.

Through comparison with the expected sequence in M1 protein, it was easy to locate the peptide sequence in the residues from 79 to 95 in M1 as FVQNALNGNGDPNNMDR. In Figure 2, the obtained  $y$ -series ions enabled the determination of two deamidation modifications at N<sub>85</sub> and N<sub>87</sub> in the corresponding peptide. Similar deamidation modifications were not observed in the remaining three asparagines in this peptide. It was noticed that the neighboring residues next to the deamidated asparagines (N<sub>85</sub> and N<sub>87</sub>) were glycines. The observed pattern of deamidation was consistent with the findings that glycine was observed as the most common ( $N + 1$ ) neighboring residue in proteins in which deamidation was established and characterized [24, 25]. Accordingly, theoretical values of the deamidated peptide FVQNALNGNGDPNNMDR are  $m/z$  631.93 and 947.41 for triply charged and doubly charged species, respectively, considering that the methionine is oxidized. The measured mass difference between the expected and measured peptides was  $-28.02$  and  $-27.96$  Da, which was calculated from the data of T5 with double charge and triple charge, respectively. Two possible amino acid residues, namely V<sub>80</sub>, R<sub>95</sub>, in T5 peptide may have a nominal mass shift of  $-28$  Da upon AA substitution: R→K ( $-28.01$  Da), V→A ( $-28.03$  Da) and R→Q ( $-28.04$  Da). Since Q was not a typical C-terminal residue of tryptic peptides, AA substitution of R→Q ( $-28.04$  Da) was first eliminated. So the site upon mutation could be suggested at R<sub>95</sub> (R→K) or V<sub>80</sub> (V→A).

The analysis of MS/MS spectrum of the ion peaks was then performed to discriminate which one of the

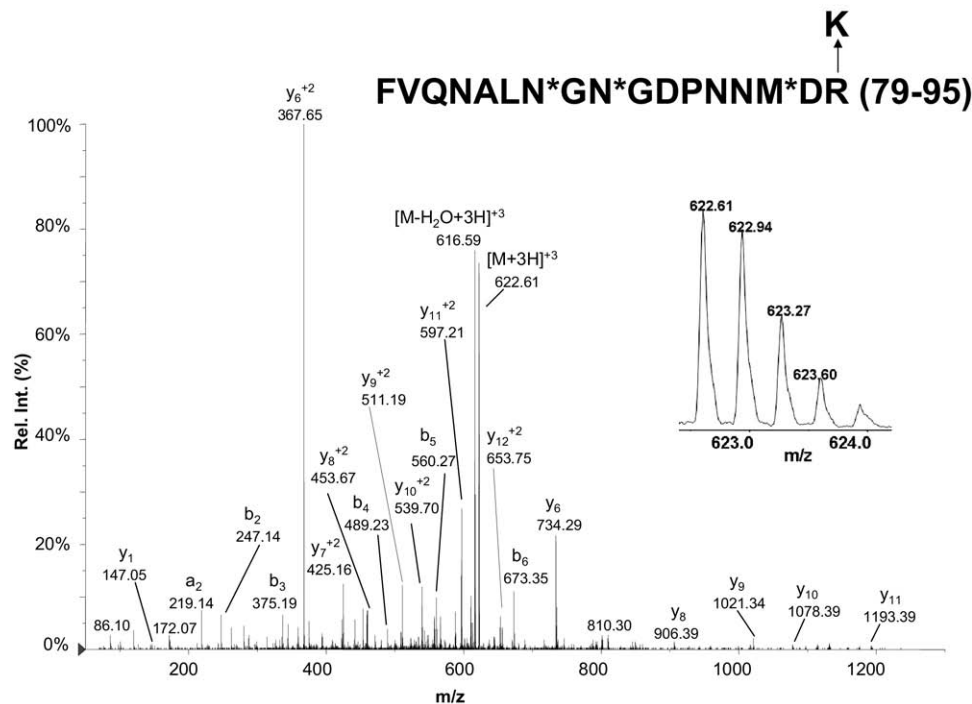
**Table 1.** Summary of tryptic peptides identified in M1 protein from the avian influenza virus [A/Dk/HN/303/2004(H5N1)]

Peak ID	Peptide sequence	Charge status	Calculated $m/z$ (monoisotopic)	Measured $m/z$ (monoisotopic)	Error (ppm)	Residues
T1	LEDVFAGK	2	439.73	439.74	23	28–35
T2	NTDLEVLN*EWLK	2	739.86	739.87	14	36–47
T3	TRPILSPLTK	3	375.90	375.91	27	48–57
T3	TRPILSPLTK	2	563.35	563.36	18	48–57
T4	GILGFVFTLTPSER	3	545.98	545.98	0	58–72
T4	GILGFVFTLTPSER	2	818.46	818.45	-12	58–72
T5	FVQNALN*GN*GDPNNM*DK	3	622.60	622.61	16	79–95
T5	FVQNALN*GN*GDPNNM*DK	2	933.40	933.40	0	79–95
T5	FVQNALN*GN*GDPNNM*DK+Na	3	629.93	629.93	0	79–95
T6	FVQNALN*GN*GDPNNMDK	3	617.27	617.28	16	79–95
T6	FVQNALN*GN*GDPNNMDK	2	925.40	925.41	11	79–95
T7	EITFHGAK	2	451.74	451.74	0	106–113
T8	M*GTVTTEVAFGLVCATCEQIADSQHR	3	928.43	928.44	11	135–160
T9	QMVTTTNPLIR	2	637.35	637.35	0	164–174
T10	QM*VTTTNPLIR	2	645.35	645.35	0	164–174
T11	MVLASTTAK	2	461.26	461.26	0	179–187
T12	M*VLASTTAK	2	469.25	469.26	21	179–187
T13	AM*EQM*AGSSEQAAEAM*EVASQAR	3	811.01	811.01	0	188–210
T14	AM*EQMAGSSEQAAEAM*EVASQAR	3	805.68	805.68	0	188–210
T15	AM*EQMAGSSEQAAEAMEVASQAR	3	800.35	800.36	12	188–210
T16	QMVOAMR	2	432.21	432.22	23	211–217
T17	TIGTHPSSSAGLK	2	628.34	628.34	0	218–230
T17	TIGTHPSSSAGLK	3	419.23	419.23	0	218–230
T18	DNLLENLQAYQK	2	724.87	724.87	0	231–242
T18	DNLLENLQAYQK	3	483.58	483.59	21	231–242
T19	M*GVQMQR	2	433.20	433.21	23	244–250

M\*: mono-oxidized methionine; N\*: deamidated asparagine; the identified mutated amino acid residues were highlighted.

two residues was substituted. The observation of diagnostic  $y_1$  ion of  $m/z$  147 for K as well as absence of  $y_1$  ion of 175 for R in the MS/MS spectra of T5 and T6

suggested that the tryptic peptide was very likely to have a K residue at the C-terminus. Additionally, the detection of  $b_2$  ion at  $m/z$  247.14 and the  $a_2$  ion at  $m/z$



**Figure 2.** ESI-MS/MS spectrum of triply charged T5 peak at  $m/z$  622.61. It was identified as the mutated peptide (FVQNALN\*GN\*GDPNNM\*DK) of an expected sequence (residues 79–95) from tryptic digestion of M1 (M\*: mono-oxidized methionine; N\*: deamidated asparagine).

219.14 indicated with great certainty that the first two residues in the peptide were FV, of which the theoretical mass values were 247.14 and 219.15, respectively. Therefore, the site upon mutation was confidently determined at R<sub>95</sub> (R→K).

#### Identification of A166V and I168T Substitutions

Two doubly charged ion peaks T9 and T10 with  $m/z$  637.35 and 645.35, respectively, were detected with a very similar pattern of fragment ion series (Figure 3). The produced fragment ion at  $m/z$  175 indicated that the corresponding peptides possibly ended with an R at the C terminus. De novo sequencing of the spectra suggested both peptides shared the same internal sequence of VTTTN, which was deduced from the  $y$ -ion series with  $m/z$  1014.54, 915.48, 814.46, 713.41, 612.37, and 498.32. The inverse sequence NTTTV deduced from the same ion series, which were considered as  $b$ -ions, was less possible because the  $y$ -ion intensity of R-terminated tryptic peptides would often be most prominent in the spectra, particularly at the high end of the spectra. The nominal mass difference between T9 and T10 was 16 Da, indicating that an oxidation, one of the most commonly observed modifications in peptides (usually at methionine), might take place in T10. It should be mentioned that the mass shift of 16 Da between T9 and T10 might not be incurred by AA substitution due to the empirical assumption that certain AA substitution

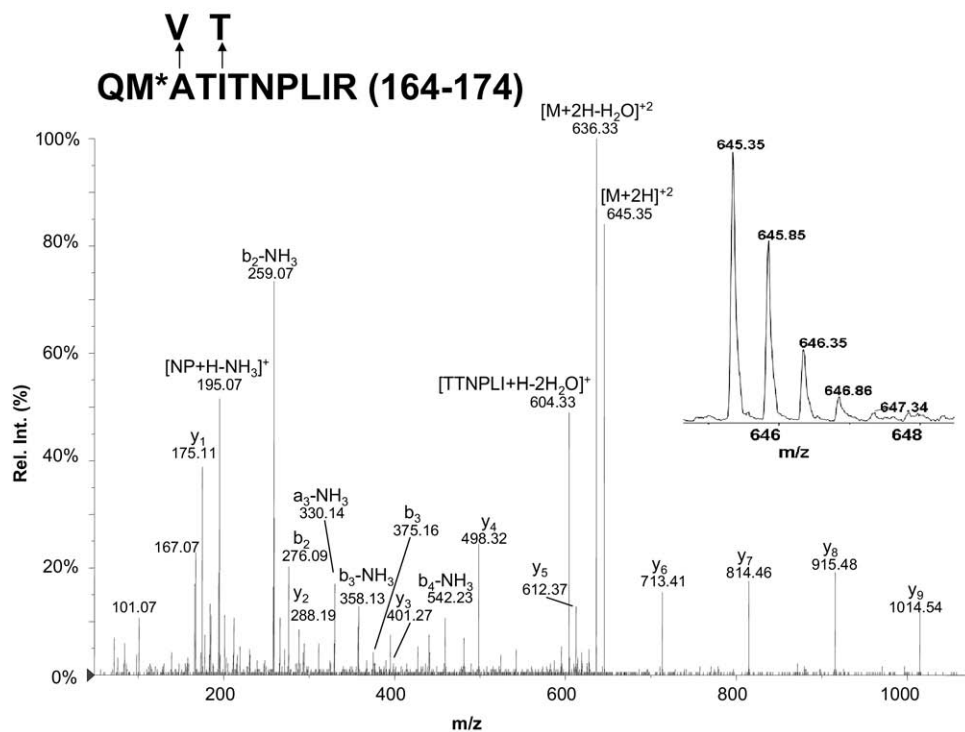
resulted from gene mutation often exhibited in a homogeneous way.

The expected sequence of the tryptic peptide ranging from 164 to 174 in M1 protein was QMATITNPLIR. The theoretical  $m/z$  values of the doubly charged ions were 637.35 and 629.35 for the methionine-oxidized and native peptides, respectively. Although the theoretical value of 637.35 was the same as the measured value of T9 (637.35), the MS/MS spectrum of T9 exhibited totally different fragment ion series from those theoretically deduced from QMATITNPLIR. Therefore, T9 did not have the sequence of QMATITNPLIR.

Finally, the doubly charged ion peak T10 with  $m/z$  645.35 was identified as the mutated peptide from the expected sequence QMATITNPLIR (164–174) with two mutated sites: A<sub>166</sub> (A→V) and I<sub>168</sub> (I→T). The combination of the AA substitution: A→V (28.03 Da) and I→T (−12.04 Da) resulted in a difference of nominal mass of 16 Da between the expected and the mutated peptides. In addition to the  $y$ -ion series from  $y_4$  to  $y_9$ , other characteristic ions such as  $b_2$ ,  $b_2$ -NH<sub>3</sub>,  $b_3$ ,  $b_3$ -NH<sub>3</sub>, and  $b_4$ -NH<sub>3</sub> were also readily assigned, confirming the identification. T9 was then identified as the same sequence with T10, but with the notoxidized methionine.

#### Identification of N207S Substitution

The expected sequence of the peptide ranging from 188 to 210 in M1 protein was AMEQMAGSSEQAAEAMEVAN-



**Figure 3.** ESI-MS/MS spectrum of doubly charged T10 peak at  $m/z$  645.35. It was identified as the mutated peptide (QM\*VTTTNPLIR) of an expected sequence (residues 164–174) from tryptic digestion of M1 (M\*: mono-oxidized methionine).

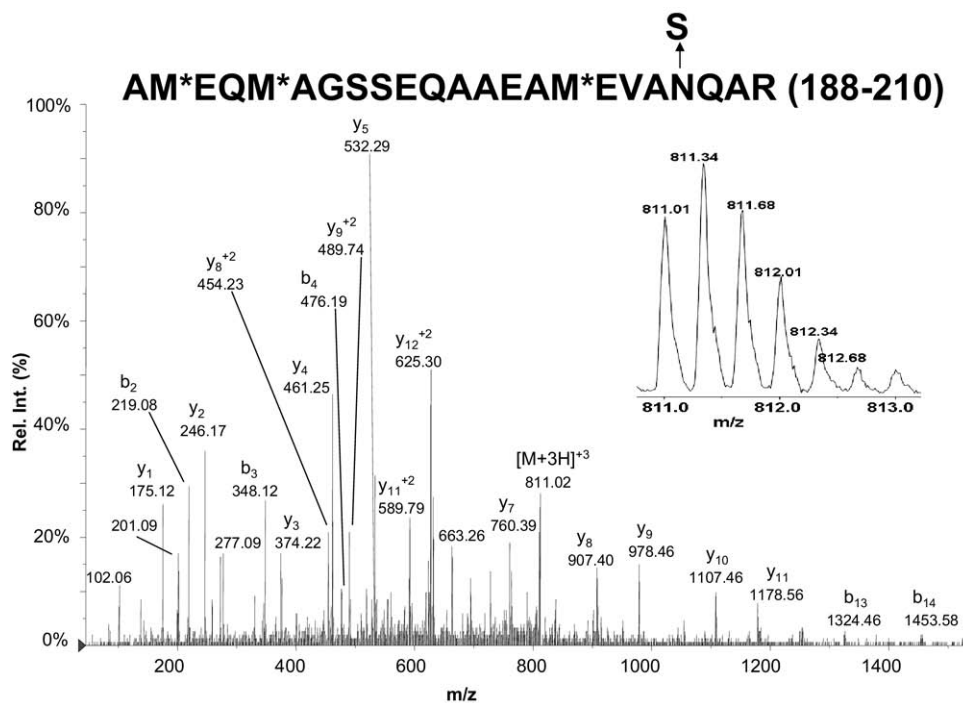
QAR. The theoretical  $m/z$  values of the triply charged ions were 809.35, 814.68, and 820.01 for the one, two, and three oxidized methionines, respectively. The peaks of T15, T14, and T13 with  $m/z$  800.36, 805.68 and 811.01, respectively, were identified as the sequence of AMEQMAGSSEQAAEAMEVANQAR with various oxidative states of methionines. However, a difference of nominal mass of  $-27$  Da was observed for T15, T14, and T13, compared with the same sequence peptide with one, two, and three oxidized methionines. Two different types of AA substitutions with the nominal mass shift of  $-27$  Da upon AA substitution were possible, namely  $N \rightarrow S$  ( $-27.01$  Da),  $K \rightarrow T$  ( $-27.05$  Da). Since one N and no K residue was presented in the peptide sequence, the mutation site could be determined at the N residue ( $N \rightarrow S$ ). The clearly assigned  $y$ -ion series from  $y_1$  to  $y_{13}$  confirmed the mutation at the position (Figure 4).

#### Identification of N224S and R230 Substitutions

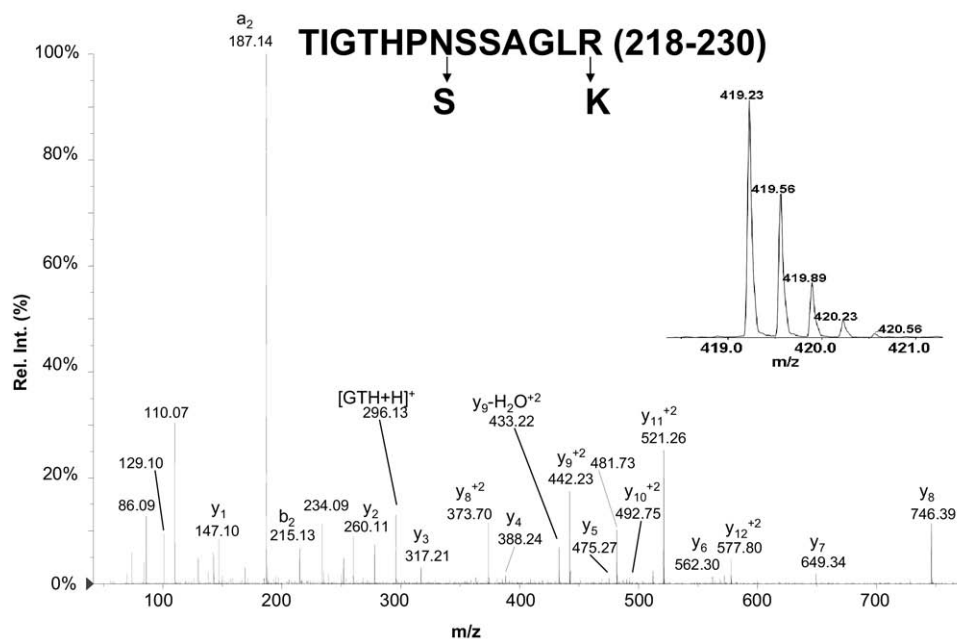
The isotope pattern in MS spectra and the fragmentation pattern in MS/MS spectra of the both ions at  $m/z$  628.34 and  $m/z$  419.23 demonstrated that they were doubly charged and triply charged ions of a same peptide (T17), respectively. De novo sequencing of the MS/MS spectra of T17 suggested a partial sequence of PSSAG(L/I)K (Figure 5), which was very similar to the expected sequence of PNSSAGLR (223–230) in M1 protein. The expected sequence of the tryptic peptide containing PNSSAGLR was TIGTHPNSSAGLR (218–

230) with a theoretical  $m/z$  values at 655.84 and 437.56 for doubly charged and triply charged ions, respectively. Therefore, it could readily assign the sequence of T17 as a mutated sequence of TIGTHPNSSAGLR. The nominal mass shift between the mutated and expected peptide was  $-55$  Da. The only type of AA substitution exhibiting the nominal mass shift of  $-55$  Da was  $R \rightarrow T$  ( $-55.05$  Da), which was unlikely because T was not a typical cleavage site of trypsin. Therefore, combination of multiple AA substitutions was investigated for this peptide. Similar to the MS/MS spectrum of T5 and T6, a fragment ion at  $m/z$  147 was observed and no signal at  $m/z$  175 was found, suggesting that the AA substitution of  $R \rightarrow K$  ( $-28.01$  Da) took place at the C terminus. The additional mass difference of 27 Da might result from one of the two potential AA substitutions, namely  $N \rightarrow S$  ( $-27.01$  Da) or  $K \rightarrow T$  ( $-27.05$  Da). Because only one N was present in the expected peptide sequence, the AA substitution  $N \rightarrow S$  ( $-27.01$  Da) at N<sub>224</sub> could be readily deduced. The suggestion was supported by the detection of  $y$  ion series in MS/MS spectra of the corresponding peptide. Similar to the identification of T9 and T10, T17 was identified as a tryptic peptide TIGTHPSSAGLK with two AA substitutions in M1 protein.

In summary, six AA substitutions were identified by nano-electrospray MS and MS/MS in four tryptic peptides of M1 protein of avian influenza H5N1 virus (Figure 6). Noticeably, none of the peaks of the corresponding expected peptides of these four mutated peptides were detected, indicating that the AA substi-



**Figure 4.** ESI-MS/MS spectrum of triply charged T13 peak at  $m/z$  811.01. It was identified as the mutated peptide (AM\*EQM\*AGSSEQAAEAM\*EVASQAR) of an expected sequence (residues 188–210) from tryptic digestion of M1 (M\*: mono-oxidized methionine).

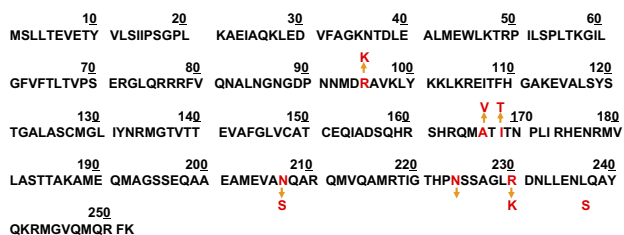


**Figure 5.** ESI-MS/MS spectrum of triply charged T17 peak at  $m/z$  419.23. It was identified as the mutated peptide (TIGTHPNSSAGLK) of an expected sequence (residues 218–230) from tryptic digestion of M1.

tutions resulted from the gene mutations in M1 might take place in a homogenous way. This phenomenon was different from the modification on amino residues such as deamidation or oxidation, which showed several variants of the same peptide with different status of modification. The MS/MS spectra of T9 and T10 exhibited similar fragmentation patterns, suggesting they might share the same or very similar sequence. Therefore, the mass difference of 16 Da between these two peptides was subjectively determined as the modification occurring at certain residue (oxidation at methionine), rather than AA substitution. Similar fragmentation was also found in the analysis of T13, T14, and T15. If a group of peaks with very similar fragmentation patterns were detected in tryptic peptides, the difference in nominal mass should not be attributed to AA substitution, but to probable modifications. When the AA-substituted peptide was detected, the corresponding native peptide was usually not found in the same sample. AA substitutions could be usually detected in variants of the same protein

originating from individual samples with different physiological status. This observation might be useful in de novo sequencing of MS/MS spectra, especially in the analysis of real samples underlying possible AA substitution.

Analysis of antigens from influenza A virus by mass spectrometry-based approaches was reported [11–13]. An immunoaffinity-coupled mass spectrometric method was developed and used to identify an antigenic epitope to a monoclonal antibody against hemagglutinin (HA), one of the envelope proteins in the viral particles [11, 12]. By using MALDI-TOFMS, acylation modification in the C-terminal HA<sub>2</sub> was investigated [13]. In the present study, nanoelectrospray MS and MS/MS were employed to analyze the primary structure of the M1 from a highly pathogenic avian influenza A (H5N1) strain through the de novo sequencing of peptides. During nanospray ESI-MS analysis of mutated peptides, mass accuracy lower than 27 ppm was achieved with external calibration. At this level of mass accuracy, some AA substitutions could not be confidently assigned without MS/MS analysis. For the identification of R95K in the peptide T5 and T6, for example, the theoretical mass shift of the two possible AA substitutions was  $-28.01$  and  $-28.03$  Da for the mutation of R→K and V→A, respectively. Because the measured mass shift was  $-28.02$  Da between the expected and mutated T5 peptide, it was impossible to determine whether the mutation was R→K or V→A in the peptide. However, if mass accuracy of 1 to 2 ppm could be achieved, for example by FT-MS, assignment of the different AA substitution could be possible.



**Figure 6.** Summary of the mutation sites in M1 protein of avian influenza virus [A/Dk/HN/303/2004(H5N1)] identified by using nanoelectrospray MS and MS/MS.

**Table 2.** Summary of changes in hydrophobicity, secondary structure and protein stability upon the six AA substitutions identified in M1 from the avian influenza virus [A/Dk/HN/303/2004(H5N1)]

Residue position of substitution	Change of residue	Change in hydrophobicity (Kyte-Doolittle)	Change in secondary structure	$\Delta\Delta G^*$	Change in protein stability, RI*
95	R→K	−0.800→−0.733	H	−1.05	Decrease, 9
166	A→V	−0.878→−1.189	C→E	−0.22	Decrease, 1
168	I→T	0.222→−0.089	C	−1.97	Decrease, 8
207	N→S	−0.978→−0.678	H	−0.64	Decrease, 4
224	N→S	−1.067→−0.767	C	−0.09	Decrease, 6
230	R→K	0.056→0.122	C	−1.43	Decrease, 7

H: alpha helix; E: extended strand; C: random coil;  $\Delta\Delta G^*$ :  $\Delta G$  (free energy value of mutated sequence) –  $\Delta G$  (free energy value of expected sequence) in kcal/mol;  $\Delta\Delta G < 0$ : decrease stability;  $\Delta\Delta G > 0$ : increase stability; RI\*: reliability index for the predicted sign.

### Changes in Hydrophobicity, Secondary Structure, and Protein Stability

Among the six mutations identified, I168T was found to be a non-synonymous mutation, resulting in the change from hydrophobic to hydrophilic at the mutation site. The other five mutations, namely R95K, A166V, N207S, N224S, and R230K, were synonymous and would not have the change. Prediction for the secondary structure by nnPREDICT software indicated that there was a change from random coil to extended strand in the secondary structure at residue of 166 upon the substitution of A to V. The other five AA substitutions did not alter the secondary structure at the respective site. Changes of protein stability upon the individual mutations were also evaluated by using I-Mutant 2.0 software. Estimation of protein stability was performed by calculating the free-energy change ( $\Delta\Delta G$ ) upon mutation. The result showed that each of the six AA substitutions exhibited a decrease in free-energy and decrease in the stability of M1 protein (Table 2).

### Conclusions

Six AA substitutions in avian influenza H5N1 virus matrix protein 1 located in four tryptic peptides of the protein were identified from the nano-electrospray Q-TOFMS and MS/MS analyses, of which some relevance between the changes in protein structures and the AA substitutions were revealed from the bioinformatics analyses. Mass spectrometric analyses also allowed the site assignment of some commonly observed modifications such as oxidation (M) and deamidation (N) in M1 protein. Among the four mutated peptides identified in the tryptic digest of M1, two peptides bearing only one AA substitution were initially identified. A great benefit can be derived from significant improvement of measurement accuracy (~1 to 2 ppm) in analyzing peptides bearing only one AA substitution when a high-resolution mass spectrometer such as FT-MS is used. But for the determination of the two peptides bearing two AA substitutions, MS analysis with high-resolution mass measurement alone is insufficient, whereas the interpretation of MS/MS fragmentations will play a pivotal role. The specificity of MS/MS

enables the de novo sequencing of the mutated peptides, which may be useful for investigating various virus species with high mutability, such as the influenza virus.

### References

1. Brankston, G.; Gitterman, L.; Hirji, Z.; Lemieux, C.; Gardam, M. Transmission of Influenza A in Human Beings. *Lancet Infect. Dis.* **2007**, *7*, 257–265.
2. Nicholson, K. G.; Wood, J. M.; Zambon, M. Influenza. *Lancet.* **2003**, *362*, 1733–1745.
3. Martin, K.; Helenius, A. Nuclear Transport of Influenza Virus Ribonucleoproteins: The Viral Matrix Protein (M1) Promotes Export and Inhibits Import. *Cell.* **1991**, *67*, 117–130.
4. Whittaker, G.; Bui, M.; Helenius, A. Nuclear Trafficking of Influenza Virus Ribonucleoproteins in Heterokaryons. *J. Virol.* **1996**, *70*, 2743–2756.
5. Elster, C.; Larsen, K.; Gagnon, J.; Ruigrok, R. W.; Baudin, F. Influenza virus M1 protein binds to RNA through its nuclear localization signal. *J. Gen. Virol.* **1997**, *78*, 1589–1596.
6. Garcia-Robles, I.; Akarsu, H.; Müller, C. W.; Ruigrok, R. W.; Baudin, F. Interaction of Influenza Virus Proteins with Nucleosomes. *Virology.* **2005**, *332*, 329–336.
7. Anwar, T.; Lal, S. K.; Khan, A. U. Matrix Pprotein 1: A Comparative in Silico Study on Different Strains of Influenza A H5N1 Virus. *Bioinformatics.* **2006**, *1*, 253–256.
8. Hui, E. K.; Smee, D. F.; Wong, M. H.; Nayak, D. P. Mutations in Influenza Virus M1 CCHH, the Putative Zinc Finger Motif, Cause Attenuation in Mice and Protect Mice Against Lethal Influenza Virus Infection. *J. Virol.* **2006**, *80*, 5697–5707.
9. Govorkova, E. A.; Gambaryan, A. S.; Claas, E. C.; Smirnov, Y. A. Amino Acid Changes in the Hemagglutinin and Matrix Proteins of Influenza A (H2) Viruses Adapted to Mice. *Acta Virol.* **2000**, *4*, 241–248.
10. Liu, T.; Ye, Z. Restriction of Viral Replication by Mutation of the Influenza Virus Matrix Protein. *J. Virol.* **2002**, *76*, 13055–13061.
11. Kiselar, J. G.; Downard, K. M. Antigenic Surveillance of the Influenza Virus by Mass Spectrometry. *Biochemistry.* **1999**, *38*, 14185–14191.
12. Morrissey, B.; Downard, K. M. A Proteomics Approach to Survey the Antigenicity of the Influenza Virus by Mass Spectrometry. *Proteomics.* **2006**, *6*, 2034–2041.
13. Serebryakova, M. V.; Kordyukova, L. V.; Baratova, L. A.; Markushin, S. G. Mass Spectrometric Sequencing and Acylation Character Analysis of C-Terminal Anchoring Segment from Influenza A Hemagglutinin. *Eur. J. Mass Spectrom.* (Chichester, Eng). **2006**, *12*, 51–62.
14. Tanaka, K.; Takenaka, S.; Tsuyama, S.; Wada, Y. Determination of Unique Amino Acid Substitutions in Protein Variants by Peptide Mass Mapping with FT-ICR MS. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 508–513.
15. Mandal, A. K.; Bisht, S.; Bhat, V. S.; Krishnaswamy, P. R.; Balaram, P. Electrospray Mass Spectrometric Characterization of Hemoglobin Q (Hb Q-India) and a Double Mutant Hemoglobin S/D in Clinical Samples. *Clin. Biochem.* **2008**, *41*, 75–81.
16. Nepomuceno, A. I.; Mason, C. J.; Muddiman, D. C.; Bergen, H. R. III; Zeldenrust, S. R. Detection of Genetic Variants of Transthyretin by Liquid Chromatography-Dual Electrospray Ionization Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry. *Clin. Chem.* **2004**, *50*, 1535–1543.
17. Cunsolo, V.; Foti, S.; Saletti, R.; Gilbert, S.; Tatham, A. S.; Shewry, P. R. Structural Studies of the Allelic Wheat Glutenin Subunits 1Bx7 and 1Bx20 by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry and High-Performance Liquid Chromatography/Electrospray Ionization Mass Spectrometry. *J. Mass Spectrom.* **2004**, *39*, 66–78.
18. Li, K. S.; Guan, Y.; Wang, J.; Smith, G. J.; Xu, K. M.; Duan, L.; Rahardjo, A. P.; Puthavathana, P.; Buranathai, C.; Nguyen, T. D.; Estoepangestie, A. T.; Chaisingh, A.; Auewarakul, P.; Long, H. T.; Hanh, N. T.; Webby,



- R. J.; Poon, L. L.; Chen, H.; Shortridge, K. F.; Yuen, K. Y.; Webster, R. G.; Peiris, J. S. Genesis of a Highly Pathogenic and Potentially Pandemic H5N1 Influenza Virus in Eastern Asia. *Nature*. **2004**, *430*, 209–213.
19. Gregoriades, A. The Membrane Protein of Influenza Virus: Extraction from Virus and Infected Cell with Acidic Chloroform-Methanol. *Virology*. **1973**, *54*, 369–383.
  20. <http://ca.expasy.org/tools/protscale.html>.
  21. <http://www.cmpfarm.ucsf.edu/~nomi/nnpredict.html>.
  22. <http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi>.
  23. Widjaja, L.; Ilyushina, N.; Webster, R. G.; Webby, R. J. Molecular Changes Associated with Adaptation of Human Influenza A Virus in Embryonated Chicken Eggs. *Virology*. **2006**, *350*, 137–145.
  24. Robinson, N. E.; Robinson, A. B. Deamidation of Human Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12409–12413.
  25. Robinson, N. E. Protein Deamidation. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5283–5288.