Title	Identification of amino acid substitutions in mutated peptides of nucleoprotein from avian influenza virus
Author(s)	Liu, N; Lee, KC; Song, W; Wang, P; Cai, Z; Chen, H
Citation	Talanta, 2009, v. 78 n. 4-5, p. 1492-1496
Issued Date	2009
URL	http://hdl.handle.net/10722/59444
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Talanta xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Short communication

Identification of amino acid substitutions in mutated peptides of nucleoprotein from avian influenza virus

Ning Liu [a], Kim-Chung Lee [a], Wenjun Song [b], Pui Wang [b], Zongwei Cai [a,*], Honglin Chen [b,**]

[a] Department of Chemistry, Hong Kong Baptist University, Kowloon Tong, Hong Kong, SAR, China
[b] State Key Laboratory for Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, Hong Kong, SAR, China

ARTICLE INFO

Article history:
Received 15 December 2008
Received in revised form 27 January 2009
Accepted 30 January 2009
Available online xxx

Keywords:
Amino acid substitution
Mutated peptide
Nucleoprotein
H5N1 virus
Nanospray-MS/MS

ABSTRACT

Nucleoprotein (NP), the structural component of ribonucleoprotein complex of avian influenza virus, performs multiple essential functions in the regulation of viral RNA synthesis and in the control of nuclear traffic of viral proteins. Mutations have often been found in NP, some of which are relevant to viral survival strategies. In this study, we used nanospray-MS/MS to analyze tryptic digestion of nucleoprotein of avian influenza virus (H5N1) and to identify three mutated peptides. The MS/MS analyses allowed the confident determination of the three mutated amino acid residues F313Y, I194V and V408I/L in the mutated peptides of LLQNSQV**Y**SLIRPNENPAHK, GVGTMVMEL**V**R and ASAGQ**I/L**SVQPTFSVQR, respectively.

1. Introduction

The influenza A virus, especially the avian flu virus (H5N1), continues to be a global health threat. The avian influenza virus can mutate to acquire the ability for the transmission to humans and to facilitate the generation of pandemic and epidemic strains [1]. Nucleoprotein (NP), a polypeptide of 498 amino acids in length, is encoded by influenza A virus RNA segment 5. As the structural component of the virus transcription machinery, NP performs multiple essential functions throughout the virus life cycle, by regulating viral RNA synthesis through the interaction with other viral components [2,3] and by controlling the nuclear traffic of viral proteins and ribonucleoprotein complexes [4]. NP has been found to exhibit some mutations at several sites [5]. Some of the mutations often result in amino acids (AA) substitutions and thus may be relevant to viral survival strategies [6,7]. Thus, identification of the mutations may be important in the prevention and control of influenza pandemics.

Mass spectrometry with soft ionization techniques such as ESI and MALDI has been successfully applied to analyze peptides, proteins, and other large bio-molecules. Characterization of the mutations at the protein level by using mass spectrometry has been reported [8–11]. Previously, we reported the use of nanospray-MS and MS/MS to analyze the matrix protein 1 (M1) isolated and purified from the viral particles, in which AA substitutions were identified [12]. In this study, nanospray-MS/MS was applied to investigate another structural protein (NP) in H5N1 virus after the protein was isolated from SDS-PAGE. Three AA substitutions were identified through the *de novo* sequencing.

2. Experimental

2.1. Chemicals and materials

Avian influenza virus A/Chicken/Hong Kong/YU22/2002 (H5N1) [13] was kept and propagated in a biosafety level 3 (BL-3) containment facility. HPLC grade ACN and methanol were from Fisher (Fairlawn, NY, USA). Sequencing grade trypsin was obtained from Promega (Madison, WI, USA). All other chemicals were purchased from Sigma–Aldrich (St. Louis, MO, USA).

2.2. Virus cultivation

After passaged several times, avian influenza virus strain A/Chicken/Hong Kong/YU22/2002 (H5N1) was harvested from allantoic fluid of chick embryos inoculated as 10-day old embryos. The allantoic fluid was inactivated with 0.03% Formalin at 4 °C for 72 h to eliminate the highly pathogenicity the virus possessed before it was transferred to further experiments. All the

Abbreviations: AA, amino acid; NP, nucleoprotein.

* Corresponding author. Fax: +852 34117348.
** Corresponding author.
E-mail addresses: zwcai@hkbu.edu.hk (Z. Cai), hlchen@hkucc.hku.hk (H. Chen).

0039-9140/$ – see front matter © 2009 Published by Elsevier B.V.
doi:10.1016/j.talanta.2009.01.057

experiments using the active virus were carried out in a bio-safety level three laboratory.

2.3. Isolation of virus by ultracentrifugation

The inactivated allantoic fluid was cleared from large debris by low-speed centrifugation. The virus was isolated and purified from the supernatant by ultracentrifugation in gradient sucrose cushion as described [12]. The virus band was carefully collected and stored at −80 °C until use.

2.4. SDS-PAGE

The purified virus particles were lysed with equal volume of reducing sample loading buffer (2% SDS, 20% glycerol, 10% 2-mercaptoethanol, 20 mM Tris–Cl and 0.001% Bromophenol Blue, pH 6.8) and kept at 95 °C for 5 min. The protein concentration was determined by using the Micro BCA (bicinchoninic acid) protein assay kit (Pierce, Rockford, IL, USA) with BSA as a standard protein. The sample (approximate 1.0 μg protein) was then diluted with appropriate amount of reducing sample loading buffer just before SDS-PAGE analysis. Electrophoretic analyses were made in a Mini-Cell system (Bio-Rad, Hercules, CA, USA), and run in 12% Tris–glycine–SDS polyacrylamide gels with a 5% stacking gel at a constant voltage setting of 200 V. After electrophoretic separation, the gels were stained with colloidal Coomassie G250 and scanned with a calibrated densitometer (GS800, Bio-Rad) (Fig. 1).In-gel digestionThe gel bands of interest were cut off and transferred into 0.6 ml Eppendorf vials. After being washed with Milli-Q water several times, the gel bands were cut into pieces of about 1.0 mm^3. The gel pieces were destained, reduced, alkylated and then in-gel digested as described [14]. The resulting tryptic peptides were extracted by a solution of 5% TFA in 50% ACN. The extract was dried in a vacuum centrifuge and then re-dissolved in 0.5% TFA and 5% ACN prior to the mass spectrometric analysis.Mass spectrometric analysis

Samples were loaded into a PicoTip emitter (New Objective, USA) and analyzed on a quadrupole orthogonal acceleration time-of-flight mass spectrometer (QSTAR, Applied Biosystems, CA) equipped with an external nanospray ion source (Protana A/S, Odense, Denmark) as described previously [15]. After the full-scan mass spectra of tryptic peptides were obtained in TOFMS mode, the parent ions of interest were subject to sequence analysis in product ion mode where the resolution of Q1 was typically set at unit mass as long as the fragment ion intensity was high enough. For database searching in MS/MS mode, Mascot generic files were created by using a script embedded in the Analyst QS 1.1 software (MDS Sciex). The obtained peak lists were searched against the SwissProt database in the entry of *other viruses* on an in-house Mascot server (Matrix Science, London, UK) with the following parameters: peptide tolerance, 0.2 Da; MS/MS tolerance, 0.2 Da; one missed cleavage; fixed cysteine carbamidomethylation and variable modifications such as asparagine/glutamine deamidation and methionine oxidation. 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). For the use of sequence tags for database searching in the sequence query mode, the obtained sequence tags were searched against the SwissProt database in the entry of *other viruses* with the parameters similar to those used in MS/MS search mode. The peptide charge was set as the charge state of the individual peptide being analyzed.

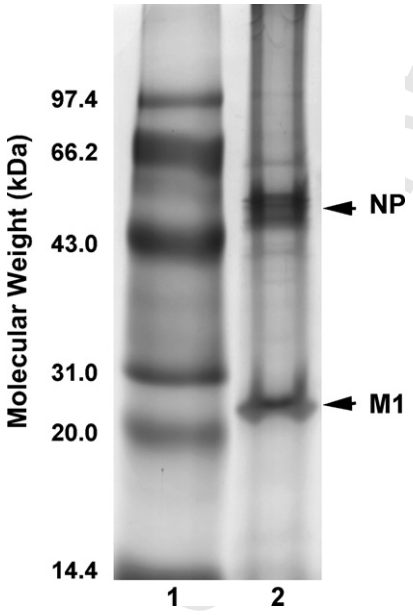

Fig. 1. SDS-PAGE separation of viral proteins from virus lysate. After being purified by ultracentrifugation in sucrose cushion, the virus particles were lysed and the viral proteins were separated on a 12% SDS-PAGE, followed by colloidal Coomassie G250 staining (Lane 1: marker, lane 2: virus lysate).

3. Results and discussion

3.1. Analysis of nucleoprotein

The virus particles were purified from allantoic fluid by using ultracentrifugation in discontinuous sucrose cushion. The purified virus was lysed and separated on SDS-PAGE. The protein bands (around 56 kDa) were cut off and subject to in-gel digestion. While the successful separation of viral proteins was performed by using 1-D SDS-PAGE, it should be noted that multidimensional separations such as 2-D electrophoresis and 2-D chromatography might significantly improve the analytical capability for more complicated protein mixtures [16]. However, in our efforts to separate the major antigens from influenza virus, the full-length nucleoprotein was difficult to be detected on 2D gels, even with pH range of 3–11.

The tryptic digests were analyzed by using nanospray-MS/MS. A Mascot score of 345 which was the sum of the unique ion scores was obtained for the identification of the nucleoprotein from avian influenza H5N1 virus (A/Chicken/Hong Kong/YU22/2002) after the database searching in the MS/MS search mode. The Mascot searches identified ten expected sequences in the nucleoprotein. The *de novo* sequencing of the MS/MS data allowed the identification of three mutated peptides. Accordingly, a total of 13 tryptic sequences (T1–T13) were identified, which allowed the assignment of 11 unique peptide sequences (Table 1).

3.2. Identification of F313Y substitution

With the interpretation of obtained MS/MS spectrum of the T8 peak and the aid of Pepsea software, the peak with the triply charged ion at *m/z* 774.41 was identified as LLQNSQVYSLIRPNENPAHK (Fig. 2). The mutated peptide could be easily located with the native NP peptide sequence in the residues from 306 to 325 or LLQNSQVFSLIRPNENPAHK from the comparison with the expected peptide sequence in NP. Compared to the theoretical *m/z* value of 769.09 for the triply charged ion of the native peptide, a difference of nominal mass of 16 Da was observed for T8. The mass difference of 16 Da was identified as the substitution of $F_{313} \rightarrow Y$ through the following sequence analyses and illustrations. Theoretically, four other AA substitutions with the nominal mass shift of 16 Da, namely $V_{312} \rightarrow D$, $S_{310} \rightarrow C$, $A_{323} \rightarrow S$, $P_{318,322} \rightarrow L$, could exist

Table 1
Summary of proteolytic peptides identified in nucleoprotein of avian influenza H5N1 virus (A/Chicken/Hong Kong/YU22/2002).

Peak ID	Peptide sequence	Charge status	Calculated *m/z*	Measured *m/z*	Residues
T1	GVFELSDEK	2	512.25	512.26	462–470
T2	LIQNSITIER	2	593.84	593.85	56–65
T3	MVLSAFDER	2	534.26	534.27	66–74
T4	M*VLSAFDER	2	542.26	542.27	66–74
T5	EGYSLVGIDPFR	2	676.85	676.84	294–305
T6	MM*ESARPEDVSFQGR	3	585.93	585.94	447–461
T7	M*M*ESARPEDVSFQGR	3	591.26	591.27	447–461
T8	LLQNSQV**Y**SLIRPNENPAHK	3	774.42	774.41	306–325
T9	GVGTM*VM*EL**V**R	2	612.31	612.31	185–195
T10	TTIM*AAFN*GNTEGR	2	750.34	750.35	423–436
T11	NPGN*AEFEDLTFLAR	2	847.90	847.91	247–261
T12	GVQIASNENM*EAM*DSNTLELR	3	785.36	785.37	362–382
T13	ASAGQIS**I**/**L**QPTFSVQR	2	845.45	845.44	401–416

M*: mono-oxidized methionine; N*: deamidated asparagine; the identified mutated amino acid residues were labeled with bold.

in the expected peptide sequence of LLQNSQVFSLIRPNENPAHK. The possibility for AA substitution of $S_{310} \rightarrow C$ was first eliminated because the C residue, if existed, should have been alkylated by iodoacetamide in the sample preparation process. The possibilities of $V_{312} \rightarrow D$, $A_{323} \rightarrow S$ and $P_{318,322} \rightarrow L$ were removed because the corresponding characteristic fragment ions were not detected in the MS/MS spectrum of the mutated peptide. For example, the AA substitution of $V_{312} \rightarrow D$ should have produced the y_{13}^{+2} ion at *m/z* 761.91, which, however, was not observed (Fig. 2). For the possibility of $A_{323} \rightarrow S$, y_3 ion at *m/z* 371.20, y_4 ion at *m/z* 468.26 and y_5 ion at *m/z* 582.30 should have been produced. The AA substitution of $P_{318} \rightarrow L$ should have produced y_8 ion at *m/z* 922.47, y_8^{+2} ion at *m/z* 461.74, y_{10} ion at *m/z* 1191.66 and y_{10}^{+2} ion at *m/z* 596.33, etc., from MS/MS analysis of the mutated peptide. Similarly, the $P_{322} \rightarrow L$ should have produced y series ions from y_4 to y_8 at *m/z* 468.29, 582.34, 711.34, 825.42 and 922.4, respectively. Therefore, the elimination of other impossible AA substitutions suggested the possible mutation at F_{313} residue (F → Y). The clearly detected both y and b ion series confirmed the identification of the mutation position (Fig. 2).

3.3. Identification of I194V substitution

The obtained product ion spectrum of the doubly charged ion peak T9 at *m/z* 612.31 showed a characteristic fragment ion at *m/z* 175, indicating that the corresponding peptide might possibly be ended with an R at C terminus (Fig. 3). Prominent y ion series of *m/z* 1166.59, 1067.44, 1010.56, 909.45 at the high end of the spectrum were observed for the R-terminated tryptic peptide. Thus, the *de novo* sequencing of this spectrum suggested GVGT as its N-terminal sequence of the corresponding peptide. It was easy to locate GVGT in an expected tryptic sequence GVGTMVMELIR (185–195) in NP with the theoretical *m/z* value of 619.32 for the doubly charged ion, considering that two methionines were oxidized (Table 1). However, a mass difference of 14 Da was observed for the detected doubly charged ion of T9 when compared to the

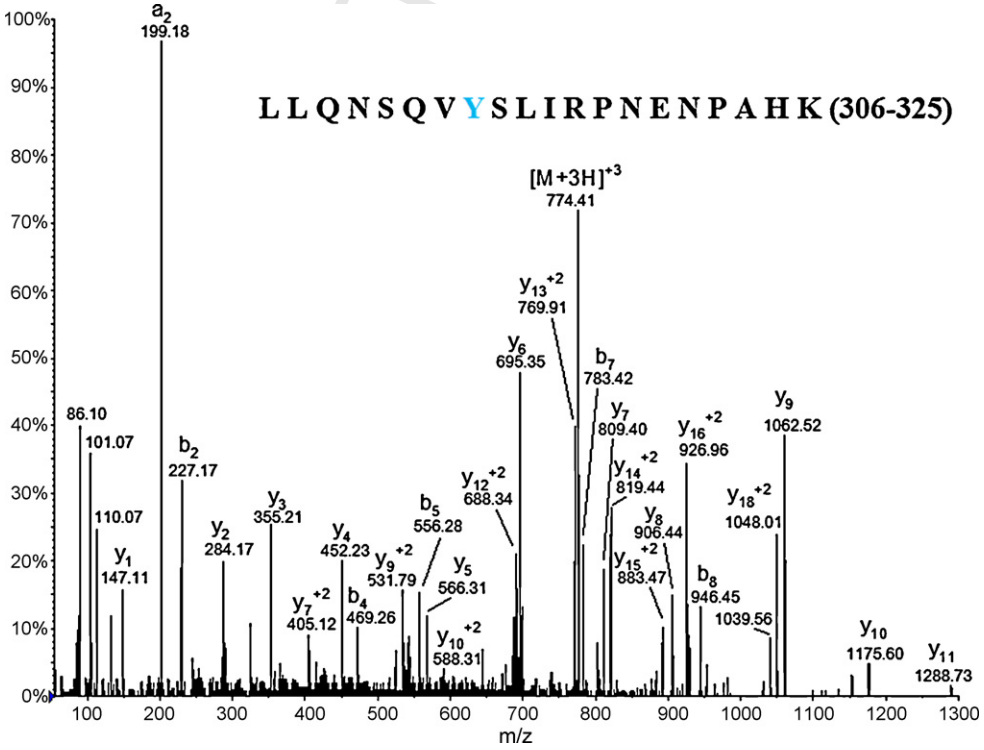

Fig. 2. ESI-MS/MS spectrum of triply charged T8 peak at *m/z* 774.41 that was identified as the mutated peptide LLQNSQV**Y**SLIRPNENPAHK with the sequence of residues 306–325 from tryptic digestion of nucleoprotein.

Please cite this article in press as: N. Liu, et al., Talanta (2009), doi:10.1016/j.talanta.2009.01.057

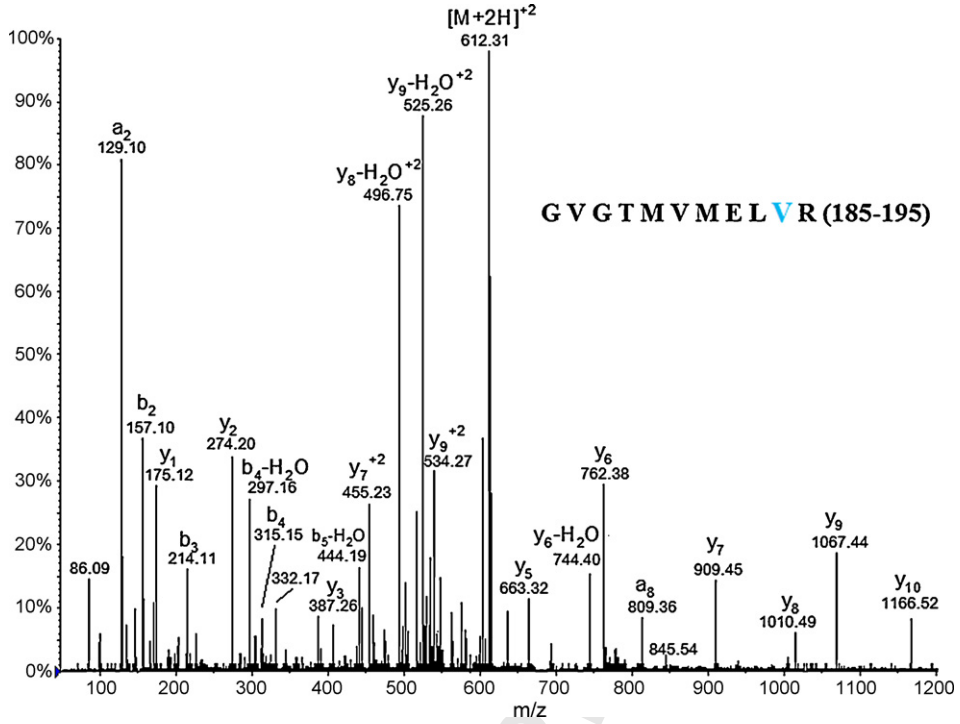

Fig. 3. ESI-MS/MS spectrum of doubly charged T9 peak at *m/z* 612.31 that was identified as the mutated peptide GVGTM**V**MELVR with the sequence of residues 185–194 from tryptic digestion of nucleoprotein.

corresponding theoretical value of the sequence of GVGTMVMELR (185-195) in NP. Four possible AA substitutions, namely $E_{192} \rightarrow D$, $T_{188} \rightarrow S$, $L_{193} \rightarrow V$ and $I_{194} \rightarrow V$ in the peptide, might have a nominal mass shift of −14 Da. Because the four AA substitutions had the same exact value of mass shift, they could not be distinguished from each other without the MS/MS analysis. The interpretation of the MS/MS spectrum of T9 was therefore performed to illustrate possibility of the AA substitution and the site of peptide mutation. The detection of y_2 ion at *m/z* 274 suggested that the tryptic peptide likely contained a V residue next to the C-terminus. Additionally, the absence of ions at *m/z* 288 and *m/z* 401 eliminated the possibilities of $E_{192} \rightarrow D$, $T_{188} \rightarrow S$ and $L_{193} \rightarrow V$. In addition to the y_2

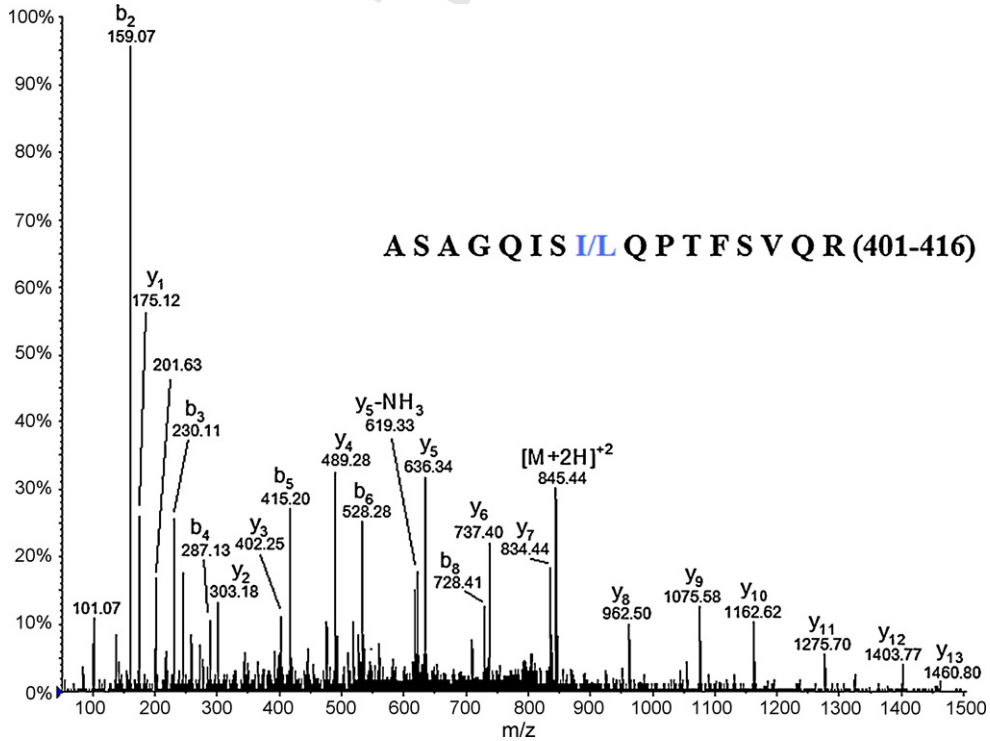

Fig. 4. ESI-MS/MS spectrum of triple-charged T13 peak at *m/z* 845.44 that was identified as the mutated peptide ASAGQIS**I/L**QPTFSVQR with the sequence of residues 401–416 from tryptic digestion of nucleoprotein.

Please cite this article in press as: N. Liu, et al., Talanta (2009), doi:10.1016/j.talanta.2009.01.057

ion, other y series ions as well as the b series ions such as b_2, b_3, b_4-H_2O and b_4, were also readily assigned, confirming the identification of the mutation site at I_{194} (I → V) in the mutated peptide GVGTMVMEL**V**R.

The partial sequence VGT was deduced from the initial interpretation of the mass spectrum shown in Fig. 3, which could be included into a qualifier in an alternative database searching algorithm in Mascot called sequence query. The Mascot result layout containing the parameters and conditions in database searching was provided in supplementary materials (Fig. S1). The use of the sequence tag qualifier [612.31 etag(1166.52, VGT, 909.45)] in the database searching against SwissProt in the entry of *other virus* in the sequence query mode of Mascot software had identified the GVGTMVMELIR (185–195) in the nucleoprotein of avian influenza virus (A/Chicken/Hong Kong/YU22/2002), but with M_{191} oxidized and the unsuspected M_{189} modification with a mass shift of 1.98 Da. However, the fragmentation pattern of the ions in the MS/MS spectrum of the peak T9 did not support the sequence query result regarding the unsuspected modification on M_{189}. Careful interpretation of the MS/MS data indicated that the M_{189} was oxidized (with a mass shift of +16 Da) and the I_{194} was substituted by V residue (with a mass shift of −14 Da). The unsuspected modification (with a mass shift of 1.98 Da) obtained from the sequence query database searching resulted from the combination of oxidation at M_{189} and substitution at I_{194}.

3.4. Identification of V408I/L substitution

Similar to the interpretation of the product ion spectrum of peak T9, a fragment ion at *m/z* 175 from the MS/MS analysis of peak T13 with the doubly charged ion at *m/z* 845.44 suggested that the corresponding peptide might end with an R at C terminus (Fig. 4). An internal sequence tag GQIS was deduced from the detected ion series at *m/z* 1460.80, 1403.77, 1275.70, 1162.62 and 1075.58. The possible tryptic peptide including this sequence tag (GQIS) in NP was ASAGQISVQPTFSVQR (401–416) with the theoretical *m/z* value of 838.44 for the doubly charged ion. The observed mass difference of 14 Da for T13 might be resulted from AA substitution at one of six possible sites, namely G_{404} → A, V_{408} → I/L, V_{414} → I/L, S_{402} → T, S_{407} → T and S_{413} → T. Given the fact that the sequence tag GQIS has already been deduced from the detected high-end y ion series, possibilities for G_{404} → A and S_{407} → T were eliminated. Similar to the interpretation of the product ion spectrum of peak T8, evaluation on the fragment ions of peak T13 indicated that the residue V_{408} (V → I/L) was the site of the AA substitution in ASAGQIS**I/L**QPTFSVQR. The clearly assigned y and b ion series eliminated the mutation possibilities at other residues in this peptide.

The sequence qualifier [845.44 etag(1460.80, GQIS, 1075.58)] was searched against the SwissProt database in sequence query mode of Mascot software. The corresponding Mascot result layout was provided in supplementary materials (Fig. S2). As a result, the sequence ASAGQISVQPTFSVQR (401–416) in nucleoprotein was identified, but with a mass shift of 14 Da which occurred at V_{408}. Therefore, the substitution of V408I/L was readily determined without the need of interpreting other fragment ions such as y and b series ions at the low-end of the spectrum.

4. Conclusion

Three AA substitutions, namely F313Y, I194V and V408I/L, were identified from the nanospray-MS/MS analysis of mutated peptides in nucleoprotein of avian influenza H5N1 virus. With the interpretation of the obtained MS/MS data, the database searching on a local Mascot server with MS/MS ion search option allowed the identification of ten expected sequences in nucleoprotein from avian influenza H5N1 virus (A/Chicken/Hong Kong/YU22/2002). Three mutated peptides were identified by *de novo* interpretation of the available data. An alternative database searching algorithm [17] called sequence query search option in Mascot was then used to confirm the results of the mutated peptides, whose specific sequence tags were readily obtained from the interpretation of MS/MS data. The results indicated that the sequence tag search algorithm was effective to directly identify mutated peptides when the sequence tags could be accurately deduced. This approach might serve as complementary to *de novo* sequencing in interpretation of MS/MS data from mutated peptides and provided useful sequence information that would facilitate *de novo* sequencing of mutated peptides. The determination of the mutated residues in peptides involved in NP provided an actual example of *de novo* sequencing of the mutated peptides, which might be useful for better understanding the mutability and structure–function relationship of the key protein in influenza virus.

Acknowledgement

The authors would like to acknowledge financial support from the Faculty Research Grant (FRG/07-08/11-21) from Hong Kong Baptist University.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.talanta.2009.01.057.

References

[1] G. Brankston, L. Gitterman, Z. Hirji, C. Lemieux, M. Gardam, Lancet Infect. Dis. 7 (2007) 257.
[2] M.A. Rameix-Welti, A. Tomoiu, E. Dos Santos Afonso, S. van der Werf, N. Naffakh, J. Virol. 83 (2008) 1320.
[3] S.L. Noton, E. Medcalf, D. Fisher, A.E. Mullin, D. Elton, P. Digard, J. Gen. Virol. 88 (2007) 2280.
[4] S. Boulo, H. Akarsu, R.W. Ruigrok, F. Baudin, Virus Res. 124 (2007) 12.
[5] K. Bragstad, L.P. Nielsen, A. Fomsgaard, Virol. J. 5 (2008) 40.
[6] K. Ohba, S. Yoshida, M. Zahidunnabi Dewan, H. Shimura, N. Sakamaki, F. Takeshita, N. Yamamoto, K. Okuda, Vaccine 25 (2007) 4291.
[7] G.F. Rimmelzwaan, A.C. Boon, J.T. Voeten, E.G. Berkhoff, R.A. Fouchier, A.D. Osterhaus, Virus Res. 103 (2004) 97.
[8] K. Tanaka, S. Takenaka, S. Tsuyama, Y. Wada, J. Am. Soc. Mass Spectrom. 17 (2006) 508.
[9] A.K. Mandal, S. Bisht, V.S. Bhat, P.R. Krishnaswamy, P. Balaram, Clin. Biochem. 41 (2008) 75.
[10] A.I. Nepomuceno, C.J. Mason, D.C. Muddiman, H.R. Bergen 3rd, S.R. Zeldenrust, Clin. Chem. 50 (2004) 1535.
[11] V. Consolo, S. Foti, R. Saletti, S. Gilbert, A.S. Tatham, P.R. Shewry, J. Mass Spectrom. 39 (2004) 66.
[12] N. Liu, W. Song, K.C. Lee, P. Wang, H. Chen, Z. Cai, J. Am. Soc. Mass Spectrom. 20 (2009) 312.
[13] K.S. Li, Y. Guan, J. Wang, G.J. Smith, K.M. Xu, L. Duan, A.P. Rahardjo, P. Puthavathana, C. Buranathai, T.D. Nguyen, A.T. Estoepangestie, A. Chaisingh, P. Auewarakul, H.T. Long, N.T. Hanh, R.J. Webby, L.L. Poon, H. Chen, K.F. Shortridge, K.Y. Yuen, R.G. Webster, J.S. Peiris, Nature 430 (2004) 209.
[14] W. Song, Q. Lin, S.B. Joshi, T.K. Lim, C.L. Hew, Mol. Cell Proteom. 5 (2006) 256.
[15] N. Liu, W. Song, P. Wang, K. Lee, W. Chan, H. Chen, Z. Cai, Proteomics 8 (2008) 1851.
[16] D.A. Lubman, M.T. Kachman, H.X. Wang, S.Y. Gong, F. Yan, R.L. Hamler, K.A. O'Neil, K. Zhu, N.S. Buchanan, T.J. Barder, J. Chromatogr. B 782 (2002) 183.
[17] M. Mann, M. Wilm, Anal. Chem. 66 (1994) 4390.