<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Identification of amino acid substitutions in mutated peptides of nucleoprotein from avian influenza virus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Liu, N; Lee, KC; Song, W; Wang, P; Cai, Z; Chen, H</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Talanta, 2009, v. 78 n. 4-5, p. 1492-1496</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2009</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/59444">http://hdl.handle.net/10722/59444</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; Talanta. Copyright © Elsevier BV.</td>
</tr>
</tbody>
</table>
Short communication

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

Ning Liu, Kim-Chung Lee, Wenjun Song, Pui Wang, Zongwei Cai, Honglin Chen

A R T I C L E   I N F O

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

A B S T R A C T

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 LQNSQYVSLIPENPAHK, GVGTMVMELVR and ASAGQI/LSVQPTSVQR, respectively.

© 2009 Published by Elsevier B.V.

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 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 [2,3]. Thus, identification of the muta-

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
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 digestion The 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 mm3. 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 for 1 h 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 time-of-flight mass spectrometer (QSTAR, Applied Biosystems, CA) equipped with an external nanospray ion source (Protaga 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.

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 LLQNFSLIRPNENPAHK (Fig. 2). The mutated peptide could be easily located with the native NP peptide sequence in the residues from 306 to 325 or LLQKFSLIRPNENPAHK 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 F313 → Y through the following sequence analyses and illustrations. Theoretically, four other AA substitutions with the nominal mass shift of 16 Da, namely V312 → D, S310 → C, A323 → S, P318,322 → L, could exist.
Table 1
Summary of proteolytic peptides identified in nucleoprotein of avian influenza H5N1 virus (A/Chicken/Hong Kong/YU22/2002).

<table>
<thead>
<tr>
<th>Peak ID</th>
<th>Peptide sequence</th>
<th>Charge status</th>
<th>Calculated m/z</th>
<th>Measured m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>GVFELSDEK</td>
<td>2</td>
<td>512.25</td>
<td>512.26</td>
</tr>
<tr>
<td>T2</td>
<td>LIQNSITIER</td>
<td>2</td>
<td>593.84</td>
<td>593.85</td>
</tr>
<tr>
<td>T3</td>
<td>MVLSAFDER</td>
<td>2</td>
<td>534.26</td>
<td>534.27</td>
</tr>
<tr>
<td>T4</td>
<td>M*VLSAFDER</td>
<td>2</td>
<td>542.26</td>
<td>542.27</td>
</tr>
<tr>
<td>T5</td>
<td>EGYSLVGIDPFR</td>
<td>2</td>
<td>676.85</td>
<td>676.84</td>
</tr>
<tr>
<td>T6</td>
<td>MM*ESARPEDVSFQGR</td>
<td>3</td>
<td>585.93</td>
<td>585.94</td>
</tr>
<tr>
<td>T7</td>
<td>M<em>M</em>ESARPEDVSFQGR</td>
<td>3</td>
<td>591.26</td>
<td>591.27</td>
</tr>
<tr>
<td>T8</td>
<td>LLQNSQVYSLIRPNENPAHK</td>
<td>3</td>
<td>774.42</td>
<td>774.41</td>
</tr>
<tr>
<td>T9</td>
<td>GVGTM<em>VM</em>ELVR</td>
<td>2</td>
<td>612.31</td>
<td>612.31</td>
</tr>
<tr>
<td>T10</td>
<td>TTIM<em>AAFN</em>GNTEGR</td>
<td>2</td>
<td>750.34</td>
<td>750.35</td>
</tr>
<tr>
<td>T11</td>
<td>NPGN*AEFEDLTFLAR</td>
<td>2</td>
<td>847.90</td>
<td>847.91</td>
</tr>
<tr>
<td>T12</td>
<td>GVQIASNENM<em>EDAM</em>DSNTLELR</td>
<td>3</td>
<td>785.36</td>
<td>785.37</td>
</tr>
<tr>
<td>T13</td>
<td>ASAGQIS/IQPTFSVQR</td>
<td>2</td>
<td>845.45</td>
<td>845.44</td>
</tr>
</tbody>
</table>

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

3.2. Identification of I194V substitution

The obtained product ion spectrum of the doubly charged ion (Fig. 2) of 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 1165.59, 1067.44, 969.36, and 871.25 at the high end of the spectrum were observed for the R-terminated tryptic peptide. Thus, the de novo sequencing of this spectrum suggested GVT as its N-terminal sequence of the corresponding peptide. It was easy to locate GVT in an expected tryptic sequence GVGTM*VM*ELVR (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 theoretical value of 615.32 for the doubly charged ion of the expected sequence GVT.

3.3. Identification of I194V substitution

In the expected peptide sequence of I194V, C was first eliminated because the C residue, if existed, should have been alkylated by iodoacetamide in the sample preparation process. The possibilities of V312 → D, A323 → S and P318,322 → 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 V312 → D should have produced the y13+2 ion at m/z 761.91, which, however, was not observed (Fig. 2). For the possibility of A323 → S, y3 ion at m/z 371.20, y4 ion at m/z 468.26, and y5 ion at m/z 582.30 should have been produced. The AA substitution of P318 → L should have produced y8 ion at m/z 922.47, y8+2 ion at m/z 461.74, and y10 ion at m/z 1191.66 and y10+2 ion at m/z 596.33, etc., from MS/MS analysis of the mutated peptide. Similarly, the 1535–1595 ion series from Y4 to Y8 at m/z 468.29, 582.34, 711.34, 825.42, and 922.40, respectively, were not detected, suggesting the possibility of other impossible AA substitutions. Therefore, the elimination of other impossible AA substitutions suggested the possible mutation at F313 residue (F → Y). The clearly detected both y and b ion series confirmed the identification of the mutation position (Fig. 2).
Corresponding theoretical value of the sequence of GVGMVMLIR (185–195) in NP. Four possible AA substitutions, namely E192 → D, T188 → S, L193 → V and I194 → 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 y2 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 E192 → D, T188 → S and L193 → V. In addition to the y2
ion, other y series ions as well as the b series ions such as \textit{b}_{2}, \textit{b}_{3}, \textit{b}_{4}^{-2} \textit{H}_{2}\textit{O} and \textit{b}_{4}^{-1} \textit{H}_{2}\textit{O} were also readily assigned, confirming the identification of the mutation site at \textit{I}_{194} (1 \rightarrow \textit{V}) in the mutated peptide \textit{GVTGMVMELVIR}.

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 \textit{GVTGMVMELVIR} (185 \rightarrow 195) in the nucleoprotein of avian influenza virus (A/Chicken/Hong Kong/YU22/2002), but with \textit{M}_{191} oxidized and the unsuspected \textit{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 \textit{M}_{189}. Careful interpretation of the MS/MS data indicated that the \textit{M}_{189} was oxidized (with a mass shift of +16 Da) and the \textit{I}_{194} was substituted by \textit{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 \textit{M}_{189} and substitution at \textit{I}_{194}.

3.4. Identification of \textit{V408I/L} substitution

Similar to the interpretation of the product ion spectrum of peak T9, a fragment ion at \textit{m/z} 175 from the MS/MS analysis of peak T13 with the doubly charged ion at \textit{m/z} 854.44 suggested that the corresponding peptide might end with an \textit{R} at C terminus (Fig. 4). An internal sequence tag \textit{GQIS} was deduced from the detected ion series at \textit{m/z} 1460.80, 1403.77, 1374.74, 1361.70, 1162.62 and 1075.58. The possible tryptic peptide including this sequence tag (\textit{GQIS}) in NP sequence was \textit{ASAGQISVQPTFSVQR} (401 \rightarrow 416) with the theoretical \textit{m/z} value of 838.44 for the \textit{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 \textit{G}_{404} \rightarrow \textit{A}, \textit{V}_{408} \rightarrow \text{IL}, \textit{V}_{404} \rightarrow \text{IL}, \textit{S}_{402} \rightarrow \textit{T}, \textit{S}_{407} \rightarrow \textit{T} and \textit{S}_{413} \rightarrow \textit{T}. Given the fact that the sequence tag \textit{GQIS} has already been deduced from the detected high-end \textit{y} ion series, possibilities for \textit{G}_{404} \rightarrow \textit{A} and \textit{S}_{407} \rightarrow \textit{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 \textit{V}_{408} (\textit{V} \rightarrow \text{IL}) was the site of the AA substitution in \textit{ASAGQISVQPTFSVQR}. The clearly assigned \textit{y} and \textit{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 \textit{ASAGQISVQPTFSVQR} (401 \rightarrow 416) in nucleoprotein was identified, but with a mass shift of 14 Da, which occurred at \textit{V}_{408}. Therefore, the substitution of \textit{V408I} was readily determined without the need of interpreting other fragment ions such as \textit{y} and \textit{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 \textit{de novo} interpretation of the available data. An alternative database searching algorithm \cite{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 \textit{de novo} sequencing in interpretation of MS/MS data from mutated peptides and provided useful sequence information that would facilitate \textit{de novo} sequencing of mutated peptides. The determination of the mutated residues in peptides involved in NP provided an actual example of \textit{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

\begin{enumerate}
\end{enumerate}