Available online at www.sciencedirect.com
ScienceDirect

Biochemical and Biophysical Research Communications 352 (2007) 177–180

BBRC

www.elsevier.com/locate/ybbrc

Influenza A H5N1 hemagglutinin cleavable signal sequence substitutions

Joel K. Weltman [a,*], Gail Skowron [a,b], George B. Loriot [c]

[a] *Department of Medicine, Brown University School of Medicine, 4 Wildacre lane, Barrington, RI 02806-2630, USA*
[b] *Division of Infectious Diseases, Roger Williams Medical Center, Boston University School of Medicine and Brown University School of Medicine, Providence, RI 02912, USA*
[c] *Center for Computation and Visualization, Brown University, Providence, RI 02912, USA*

Received 26 October 2006
Available online 10 November 2006

Abstract

Eleven influenza A H5N1 hemagglutinin N-terminal cleavable signal sequences, coded by single nucleotide substitutions relative to reference A/Viet Nam/1203/2004, were identified by BLASTN search of GenBank and were characterized by molecular modeling. The signal sequences statistically segregated into two classes of states. Members of one class were uncharged and conformationally compact while members of the second class each carried a 2+ electric charge and were conformationally extended. Virtual signal sequences, not found on GenBank and based upon hypothetical transversions in the third codon, had molecular characteristics intermediate to those of the two classes of actual signal sequences. The high incidence of non-synonymous substitutions (63.6%), the high transition/transversion ratio (10/1) and the results of molecular modeling all suggest that the N-terminal cleavable signal sequence is mutationally evolving more rapidly than proteins which must assume specific conformational states in the mature influenza virion.

Keywords: Influenza; H5N1; Avian; Hemagglutinin; Signal sequence; Mutations; Molecular mechanics; Evolution; Transitions; Transversions; Conformation

An analysis of single nucleotide substitutions in the influenza H5N1 hemagglutinin signal sequence based upon bioinformatics of the viral nucleic acid gene and molecular mechanics of the translated viral gene products is presented here. The cleavable, N-terminal 16-mer signal sequence of the hemagglutinin is essential for productive infection by the influenza A virus [1] and therefore, it is important to understand the fundamental principles that govern the structure, function, and evolution of the influenza hemagglutinin signal sequence.

Materials and methods

The signal sequence of the lethal, human isolate H5N1 influenza viral strain A/Viet Nam/1203/2004, was used as the reference for a BLASTN [2] search of the GenBank non-redundant database for influenza H5N1 strains with a hemagglutinin cleavable signal differing from that of the reference strain by a single nucleotide substitution. Signal sequences reported up to and including 2004, the year of isolation of the reference strain, were included in this study. The DNA sense representation of the antisense single stranded RNA signal sequence coding for the reference A/Viet Nam/1203/2004 viral signal sequence is: **atggagaaaatagtgcttctttttgcaatagtcagtcttgttaaaagt**. The amino acid sequence of the reference hemagglutinin signal sequence is: **MEKIVLLFAIVSLVKS**.

Molecular models of each of the hemagglutinin signal sequences were constructed with TINKER [3] using the amoebapro force field [4]. The hemagglutinin signal sequences were modeled with the carboxy termini *N*-methyl capped so as to mimic the covalent structure of the hemagglutinin precursor. The geometry of each 16-mer signal sequence peptide was optimized using TINKER MINIMIZE.EXE. Each geometry optimized signal sequence peptide was characterized with TINKER ANALYZE.EXE to obtain the total potential energy, the electric charge, the dipole moment and the radius of gyration. Non-parametric Mann–Whitney *U* statistical tests of significance [5] of the data were performed with MATLAB (Version 7.0.1.24704). Descriptive statistics, i.e., median, mode, range, mean, and skewness, were obtained with MS-Excel.

* Corresponding author. Fax: +1 401 863 9265.
E-mail address: joel_weltman@brown.edu (J.K. Weltman).

0006-291X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2006.10.184

Results and discussion

Identification of influenza A H5N1 hemagglutinin signal sequences

Eleven influenza A H5N1 hemagglutinin signal sequences with single nucleotide substitutions relative to the reference sequence were identified by the BLASTN search and are listed in Table 1 along with their nucleotide substitution and amino acid designation. Four synonymous substitutions (36.4%) and seven non-synonymous substitutions (63.6%) were reported by the BLASTN search. Ten single nucleotide substitutions in the signal sequence are transitions relative to the reference signal sequence. The six transitions in signal sequences K3E, K3R, K3K, A9A, K15E, and S16G involve purine bases while the four transitions in signal sequences F8L, F8F, V11A, and V14V involve pyrimidines. The purine transition that occurred at the third position of codon 3 was synonymous for LYS. The pyrimidine transition at the third position of codon 8 was synonymous for PHE. Synonymous transitions also occurred at the third positions of codons 9 and 14. One of the 11 nucleotide substitutions shown in Table 1, at the first nucleotide of the sixth codon, is a cytosine ⇒ adenine transversion that causes a conservative, non-synonymous LEU ⇒ ILE substitution at the sixth amino acid position of the signal sequence (designated L6I).

Characterization of influenza A H5N1 hemagglutinin signal sequences

Molecular characterizations of the influenza A H5N1 hemagglutinin signal sequences that were identified by BLASTN search are given in Table 1. The median potential energy, electric charge, dipole moment, and radius of gyration of the hemagglutinin signal sequences shown in Table 1 were −269.1877 kcal/mol, +2 electron units, 41.249 D, and 14.951 Å, respectively. In all cases, the mode equaled the median. The mean potential energy, electric charge, dipole moment, and radius of gyration of the hemagglutinin signal sequences shown in Table 1 were −296.6158917 kcal/mol, +1.666666667 electron units, 36.52133 D, and 14.01383333 Å, respectively. In all cases, the parameter mean was less than the median, indicating non-normal distributions of molecular parameters. Accordingly, the hemagglutinin signal sequences were divided into two classes. Class 1 was defined as signal sequences K3E and K15E. Class 2 was defined as all of the other 10 signal sequences. The null hypothesis for the statistical identity of signal sequence classes 1 and 2 was rejected by the Mann–Whitney *U* test for all four molecular parameters with $p = 0.0303$ for each of the four. The probability that classes 1 and 2 do not represent statistically distinguishable molecular states is equal to 0.0303^4, or 8.4×10^{-7}. No other distribution of the signal sequences permitted such a definitive identification of two

Table 1
Influenza A H5N1 hemagglutinin signal sequence characterizations

Virus strain	GenBank Accession No.	Reference codon	Substituted codon	Signal sequence designation	Total potential energy (kcal/mol)	Total electric charge (electron units)	Dipole moment (D)	Radius of gyration (Å)
A/Viet Nam/ 1203/2004	AY651334			Reference	−269.1877	2.00	41.249	14.951
A/chicken/Viet Nam/159/2004	ABE97618	aaa	gaa	K3E	−425.2869	0.00	22.791	8.487
A/chicken/Viet Nam/c58/2004	AAW80718	aaa	aga	K3R	−299.5118	2.00	25.196	15.060
A/duck/Viet Nam/40/2004	DQ497673	aaa	aag	K3K	−269.1877	2.00	41.249	14.951
A/chicken/Sarburi/Thailand/ CU-27/2004	AAZ29968	ctt	att	L6I	−267.0616	2.00	43.931	14.825
A/chicken/Chachoengsao/ Thailand/ CU-11/2004	AAZ29955	ttt	ctt	F8L	−258.0824	2.00	43.457	15.113
A/goose/Hong Kong/739.2/ 2002	AY575871	ttt	ttc	F8F	−269.1877	2.00	41.249	14.951
A/chicken/Viet Nam/39/2004	AY651342	gca	gcg	A9A	−269.1877	2.00	41.249	14.951
A/mynas/Ranong/Thailand/ CU-209/2004	AAZ29981	gtc	gcc	V11A	−277.2001	2.00	30.006	16.382
A/duck/Hong Kong/821/2002	AAV97601	gtt	gtc	V14V	−269.1877	2.00	41.249	14.951
A/chicken/Suphanburi/ Thailand/ CU-9/2004	DQ083557	aaa	gaa	K15E	−418.1893	0.00	21.969	8.562
A/chicken/Hong Kong/NT93/ 2003	AAT73294	agt	ggt	S16G	−268.1201	2.00	44.661	14.982

Codons are listed in the DNA sense representation. Amino acid numbering is relative to the N-terminus.

Table 2
Virtual codon 3 transversion mutants in influenza A H5N1 hemagglutinin cleavable signal sequence

Virtual signal sequence	Reference codon	Hypothesized codon with transversion substitution	Total potential energy	Total electric charge	Dipole moment	Radius of gyration
K3Q	aaa	caa	−317.8533	1.00	33.303	9.498
K3T	aaa	aca	−283.2371	1.00	97.148	14.141
K3I	aaa	ata	−271.0886	1.00	32.127	9.726
K3N	aaa	aac	−276.8063	1.00	82.191	13.395
K3N	aaa	aat	−276.8063	1.00	82.191	13.395

Virtual mutants, not found on the GenBank, were constructed from hypothesized transversions at codon 3. Substitutions of amino acid residues at sequence residue position 3 were predicted from the hypothesized transversions according to the genetic code.

classes of molecular states. The members of class 1 are spatially compact and electrically uncharged. In contrast, the members of class 2 are in extended conformations and carry a 2+ charge. Segregation of the signal sequences into distinct compact and extended states demonstrates that functional influenza hemagglutinin signal sequences occur in nature over a wide range of molecular states.

Characterization of hypothesized influenza A H5N1 hemagglutinin signal sequences

Hemagglutinin signal sequences were hypothesized based upon transversions at signal sequence codon 3 (aaa). One transversion, aaa ⇒ taa, from the reference codon leads to the taa termination codon and is not considered further. The remaining five hypothesized transversions were used to predict amino acid substitutions at position 3 of signal sequences which were then modeled and analyzed statistically, as described above. As shown in Table 2, transversion substitutions in codon 3 relative to the reference codon lead to virtual signal sequences with either GLN, THR, ILE, or ASN substituting for LYS in amino acid position 3 (K3Q, K3T, K3I, K3N (from aac), and K3N (from aat)). The median potential energy, charge, dipole moment, and radius of gyration of the virtual signal sequences were −276.8063 kcal/mol, +1 electron unit, 82.191 D, and 13.395 Å units. The results in Table 2, obtained for the virtual signal sequences with hypothesized transversion substitutions in codon 3, were compared with those for the reference signal sequence and signal sequences K3E, K3R, and K3K produced by transition substitutions at codon 3 (Table 1). The potential energy, charge, dipole moment, and radius of gyration of the virtual signal sequences all were indistinguishable by the non-parametric Mann–Whitney *U* test from those obtained for the actual signal sequences that had been identified on the GenBank ($0.095 < p < 0.857$). In all cases the results for the virtual signal sequences either overlapped with, or were contained within the bounds of the results for the actual signal sequences. In all cases, the results for at least two members of the four virtual signal sequences were within the bounds of the results of the four actual signal sequences. Thus transitions and transversions at codon 3 of the N-terminal cleavable signal sequence region of the influenza A H5/N1 hemagglutinin gene appear to play opposing roles: transition mutations driving the limits of evolution, tranversion mutations potentially playing a moderating evolutionary role. The ratio of synonymous to non-synonymous substitution of 4/7 is lower than the range of 7–16 reported for the influenza virus overall [6], suggesting that the hemagglutinin cleavable signal sequence undergoes particularly rapid mutational evolution, in keeping with the high error rate of the viral RNA polymerase complex [7–9].

Hemagglutinin signal sequence substitutions and molecular evolution of influenza A H5N1

Although the LYS ⇒ GLU substitution at the third amino acid position of the N-terminal hemagglutinin signal sequence (Table 1) has a drastic effect on signal sequence conformation, this substitution, based on a nucleotide transition, nevertheless yields biologically active molecules in nature. In contrast, virtual signal sequences, resulting from hypothetical nucleotide transversions at codon 3, possess overall molecular parameters lying within the boundaries of those of the natural structures (Table 2) but have not been observed in nature. These results suggest that molecular evolution at codon 3 of the influenza hemagglutinin N-terminal cleavable signal sequence has been regulated by the nucleotide transition/transversion substitution ratio and not by the molecular characteristics of the translated viral gene product. The transition/transversion ratio over the entire genome of the mature varies with a range from 2 to 4 [6], considerably lower than the value of 10 reported here for the cleavable hemagglutinin signal sequence, which is removed from the precursor hemagglutinin [10] before the mature virion is secreted. The high transition/transversion ratio and the wide variation of charge and spatial conformation of the hemagglutinin signal sequence suggest that the cleavable signal sequence is mutating more rapidly than proteins which must assume specific conformational states in the mature influenza virion.

Acknowledgments

The authors thank William Thompson, PhD and Robert Pressman, PhD for their critical reviews of the manuscript.

References

[1] K. Sekikawa, C.J. Lai, Defects in functional expression of an influenza virus hemagglutinin lacking the signal peptide sequences, Proc. Natl. Acad. Sci. USA 80 (1983) 3563–3567.

[2] S.F. Altschul et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.

[3] J.W. Ponder, F.M. Richards, An efficient Newton-like method for molecular mechanics energy minimization of large molecules, J. Comput. Chem. 8 (1987) 1016–1024.

[4] J.W. Ponder, D.A. Case, Force fields for protein simulations, Adv. Protein Chem. 66 (2003) 27–85.

[5] H.B. Mann, D.R. Whitney, On a test of whether one of 2 random variables is stochastically larger than the other, Ann. Math. Stat. 18 (1947) 50–60.

[6] J. K Taubenberger et al., Characterization of the 1918 influenza virus polymerase genes, Nature. 437 (2005) 889–892.

[7] J.D. Parvin et al., Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1, J. Virol. 59 (1986) 377–383.

[8] E. Domingo, J.J. Holland, RNA mutations and fitness for survival, Annu. Rev. Microbiol. 51 (1997) 151–178.

[9] E. Fodor, G.G. Brownlee, Influenza Virus Replication, in: C.W. Potter (Ed.), Influenza, Elsevier, 2002, pp. 1–29.

[10] M.J. Gething, J. Sambrook, Construction of influenza haemagglutinin genes that code for intracellular and secreted forms of the protein, Nature. 300 (1982) 598–603.