J Mol Model (2007) 13:1237–1244
DOI 10.1007/s00894-007-0245-0

ORIGINAL PAPER

Homology modeling and examination of the effect of the D92E mutation on the H5N1 nonstructural protein NS1 effector domain

Minyong Li · Binghe Wang

Received: 7 April 2007 / Accepted: 19 September 2007 / Published online: 5 October 2007

Abstract Virulent H5N1 strains of influenza virus often harbor a D92E point mutation in the nonstructural protein NS1. This crucial mutation has been correlated with increased virulence and/or cytokine resistance, but the structural implications of such a change are still unclear. Furthermore, NS1 protein could also be a potential target for the development of novel antiviral agents against H5N1 strains. Therefore, a reasonable 3D model of H5N1 NS1 is important for the understanding of the molecular basis of increased virulence and the design of novel antiviral agents. Based on the crystal structure of a non-H5N1 NS1 protein, a model of H5N1 NS1 was developed by homology modeling, molecular mechanics and molecular dynamics simulations. It was found that the D92E mutation could result in weakened interactions of the carboxylate side chain with other phosphorylated residues, thereby activating phosphorylation of NS1.

Keywords H5N1 · Avian influenza · Nonstructural protein NS1 · Homology modeling · Molecular mechanics · Molecular dynamics · Antiviral

Introduction

H5N1 strains of avian influenza A viruses are of increasing public health concern because of their occasional transmission to humans, causing highly contagious respiratory disease [1]. Furthermore, highly pathogenic H5N1 viruses have spread among poultry and wild aquatic birds from Asia to Europe and Africa since late 2003 [2]. The possibility of a global H5N1 flu pandemic, which could kill up to 140 million people and cost over 4 trillion US dollars in the worst case scenario [3], is not considered a remote possibility if left uncontrolled [4]. Therefore, there is an urgent need for an in-depth understanding of factors that contribute to the virulence of this type of influenza virus.

The genome of influenza virus consists of eight single-stranded negative sense RNA segments, segment 8 of which encodes a 26-kDa nonstructural protein NS1, which is critical to a viable infection [5]. NS1 protein has multiple functions, including inhibition of the cellular immune response during the viral life cycle and interference with splicing and polyadenylation as well as the resultant inhibition of nuclear export of cellular mRNAs in infected cells [6, 7]. For example, the NS1 effector domain functionally interacts with the cellular 30 kDa subunit of cleavage and polyadenylation specificity factor (CPSF), an essential component of the 3′-end processing machinery of cellular pre-mRNAs [8]. Consequently, NS1 inhibits 3′-end processing in vivo, resulting in uncleaved pre-mRNA remaining in the nucleus [8]. Substantial evidence suggests that NS1 also plays a role in virulence by abrogating the expression of antiviral genes, such as interferon (IFN), nuclear factor kappa B (NF-κB) and RNA-activated protein kinase PKR pathways, in host cells [9–11]. In addition, a recent large-scale sequence analysis also substantiates the idea that NS1 could contribute, at least in part, to the virulence of the H5N1 virus [12]. Therefore, a detailed understanding of the structural features of this protein becomes very important. The CPSF-30 binding site (residues 144–186) on H5N1 NS1 protein could be a potential target for the development of antiviral agents, and

M. Li · B. Wang (✉)
Department of Chemistry and Center for Biotechnology and Drug Design, Georgia State University,
Atlanta, GA 30302-4098, USA
e-mail: wang@gsu.edu

an in-depth understanding of its structural features will help drug design efforts [13, 14].

In virulent H5N1 strains, NS1 carries one of two mutations: a point mutation of aspartate 92 to glutamate (D92E) [15, 16], or a deletion of residues 80–84 [17]. These crucial mutations have been correlated with increased virulence, cytokine resistance or both in certain H5N1 strains. However, the structural implications of the D92E mutation in H5N1 NS1 are still unknown. Since there is no 3D structure available for the H5N1 NS1 protein, we undertook to establish a 3D structure of this NS1 protein using computational methods, and further explored the structural changes brought about by the D92E mutation. This work complements our research interest in understanding factors contributing to H5N1 virulence. In our previous work, we simulated how the respiratory duct cell surface carbohydrates SA-α-2, 3-Gal and SA-α-2, 6-Gal bind to H5N1 hemagglutinin [18]. The results of such computational work were later substantiated by crystallographic structure determination [19] and mutagenesis [20], respectively.

Materials and methods

Homology modeling

In this study, we were interested in gaining insight into the molecular structural features of the H5N1 NS1 protein using molecular modeling. In the absence of a well-defined, experimentally determined structure, homology modeling provides a rational alternative to develop a reasonable 3D model. It should be noted that homology modeling of proteins is currently the most accurate method for the prediction of 3D structure, yielding models suitable for a wide spectrum of applications, such as investigations into mechanism, structure-based drug development, and virtual screening [21–24]. This approach can produce a reasonable structural model for any given protein sequence that has related templates having more than 25% amino acid sequence identity [25].

For our homology modeling, we thus required an experimentally determined structure with high homology to the

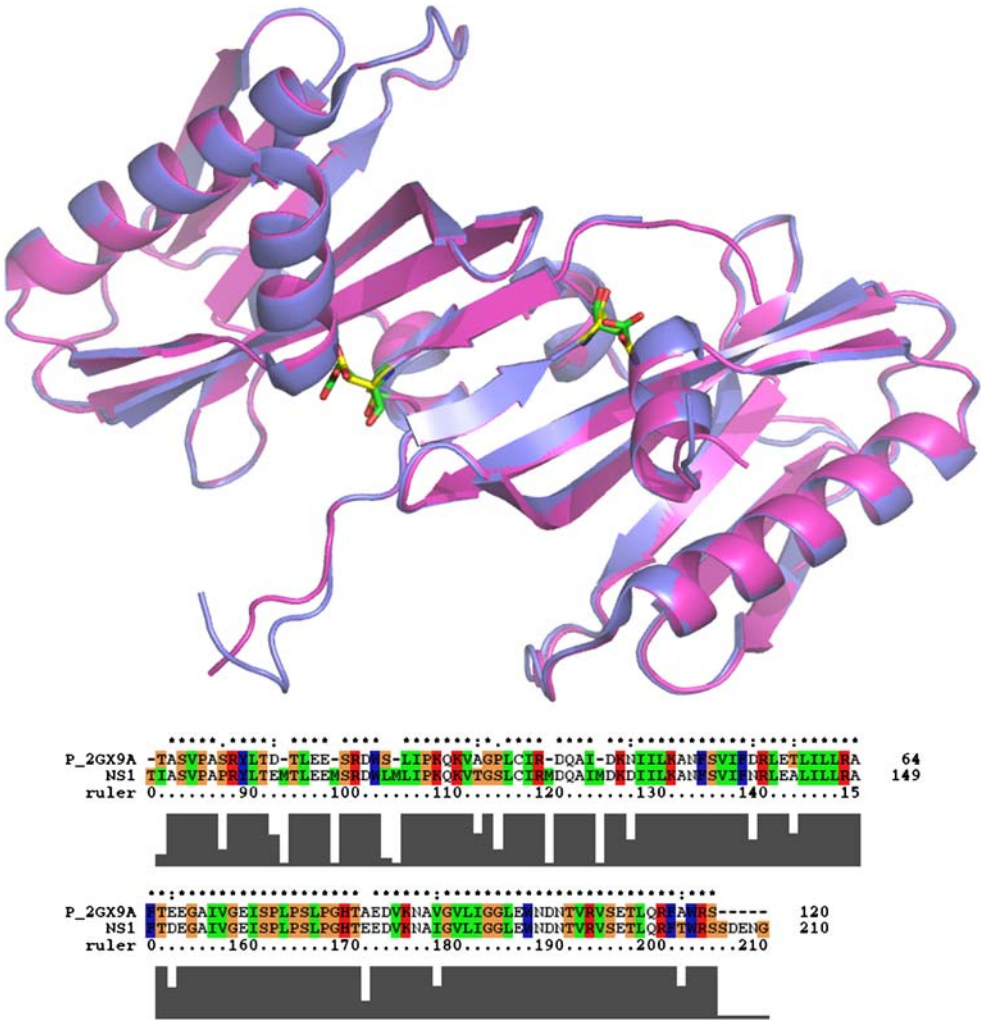

Fig. 1 The sequence and structural alignment of H5N1 NS1 to influenza virus A NS1. *Pink ribbon* H5N1 NS1 homology model, *blue ribbon* influenza virus A NS1, *green sticks* Glu 92, *yellow sticks* Asp 92

H5N1 NS1 protein. The crystal structure of the non-H5N1 influenza virus A NS1 protein is known [26]. The construction of protein models by homology modeling normally proceeds along a series of well-defined and commonly accepted steps: (1) sequence alignment between the target and the template; (2) building an initial model; (3) refining the model; and (4) evaluating the quality of the model [23].

The amino acid sequence of strain A/Hong Kong/156/97 H5N1 NS1 was obtained from the SWISS-PROT database (SWISS-PROT entry NS1-IAHO3) [27]. The X-ray structure of non-H5N1 influenza virus A NS1 was obtained from the RCSB Protein Database Bank (PDB entry 2GX9) [26]. Sequence alignment was conducted using the CLUSTAL X 1.83 program [28], and default parameters were applied. Aligned results were inspected and adjusted manually to minimize the number of gaps and insertions. The final sequence and structural alignment of H5N1 NS1 to non-H5N1 NS1 (Fig. 1) shows about 90% sequence identity of the effector domain, which suggests that the most important part of the sequence is conserved. Therefore, we conclude that this alignment is suitable for the construction of a reliable 3D model for H5N1 NS1.

Following alignment, the backbone coordinates of the residues in the H5N1 NS1 monomer were generated with the MODELLER 7v7 program using default parameters [29]. The monomer coordinates were then aligned with template dimer respectively to obtain the dimer homology model. In homology modeling, it is very important that appropriate steps are built into the process to assess the quality of the model. For this, we used PROCHECK [30], ERRAT [31], VERIFY-3D [32], WHAT-IF [33] and PROSA2003 [34] to assess the quality of the H5N1 NS1 protein model constructed.

Molecular dynamics simulation

The rough dimer model was soaked using the TIP3P water model [35], subjected to 500 steps of steepest descent, 1,500 steps of conjugate gradient while restraining all other atoms, and molecular dynamics (MD) simulations at 300 K for 10 ns using the SANDER module in AMBER 8 [36]. The whole system was heated from 0 to 300 K over 30 ps with harmonic constraints. This was followed by equilibration for 100 ps at 300 K with harmonic constraints gradually reduced to zero and an additional 200 ps equilibration of the system without constraints. The equilibrated systems were used to perform 10 ns of MD simulations with periodic boundary conditions. As a control, the non-H5N1 NS1 protein was also simulated using the same protocol.

Hardware and software

The homology modeling (MODELLER 7v7), binding analysis (HBPLUS 3.06 and Ligplot 4.22), and visualization of models (PyMOL 0.99 [37]) were carried out on a Linux workstation. The molecular mechanics and MD simulations (AMBER 8) were performed on a Linux-based 40-node cluster. The PROCHECK, ERRAT, VERIFY-3D (http://nihserver.mbi.ucla.edu/SAVS/), WHAT-IF (http://swift.cmbi.kun.nl/WIWWWI/) validation and NetPhos phosphorylation site prediction (http://www.cbs.dtu.dk/services/NetPhos/) were executed on-line.

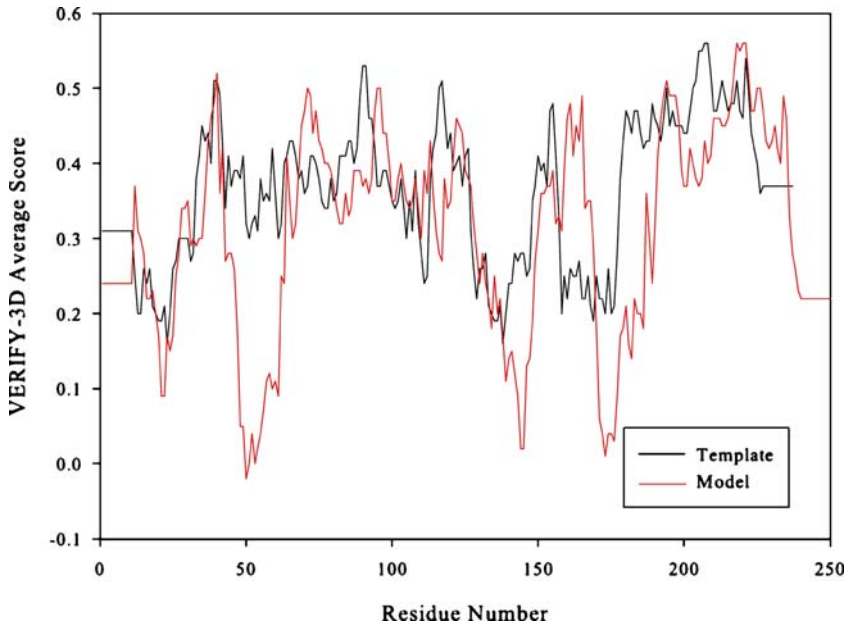

Fig. 2 VERIFY-3D score profiles calculated for the template and the homology model. The percentage of residues with a score > 0.2 should be more than 80% for a reliable model

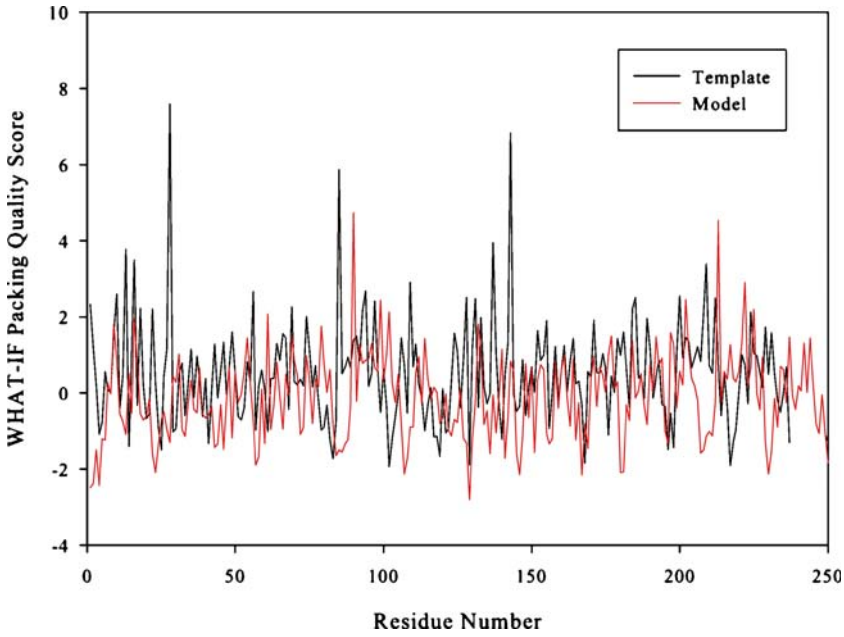

Fig. 3 WHAT-IF packing quality scores calculated for the template and the homology model. The score should be above −5 for a reliable model

Results and discussion

Homology modeling

The root-mean-square deviation (RMSD) between the backbone atoms of the template and the homology model was 0.67 Å, again indicating a close homology (Fig. 1). Therefore, we felt that we had a reasonable and reliable conformation for further simulation.

ERRAT is a so-called "overall quality factor" for non-bonded atomic interactions, and higher scores mean higher quality [31]. The normally accepted range is >50 for a high quality model [31]. In the current case, the ERRAT score is 83.2, well within the range of a high quality model; the ERRAT score for the template is 97.5. Thus, the backbone conformation and non-bonded interactions of the homology model are all within a normal range.

VERIFY-3D uses energetic and empirical methods to produce averaged data points for each residue to evaluate the quality of protein structures [32]. Using this scoring function, if more than 80% of the residue has a score of >0.2 then the protein structure is considered high quality [32]. For our H5N1 NS homology model, 82% of the residues have a score of >0.2 (Fig. 2).

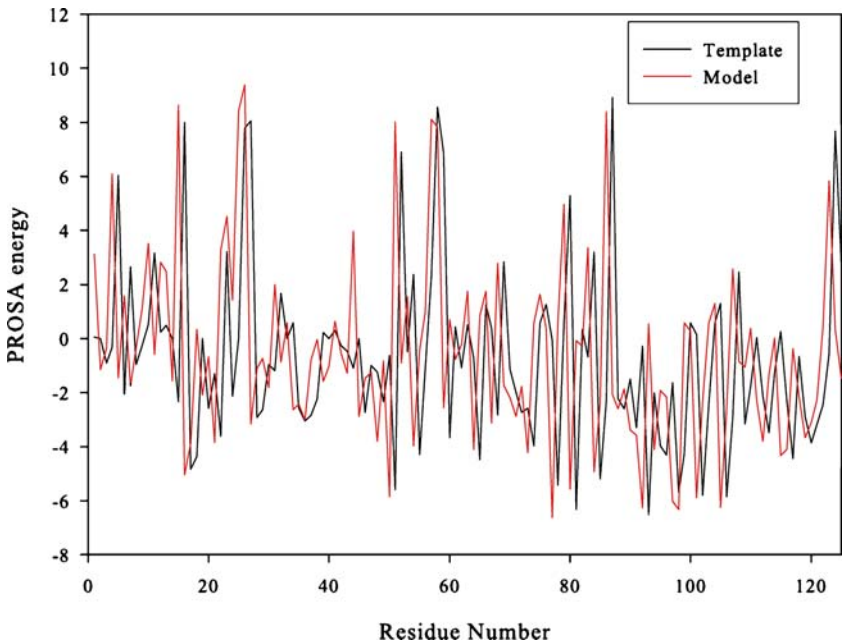

Fig. 4 PROSA2003 energy profiles calculated for the template and the homology model

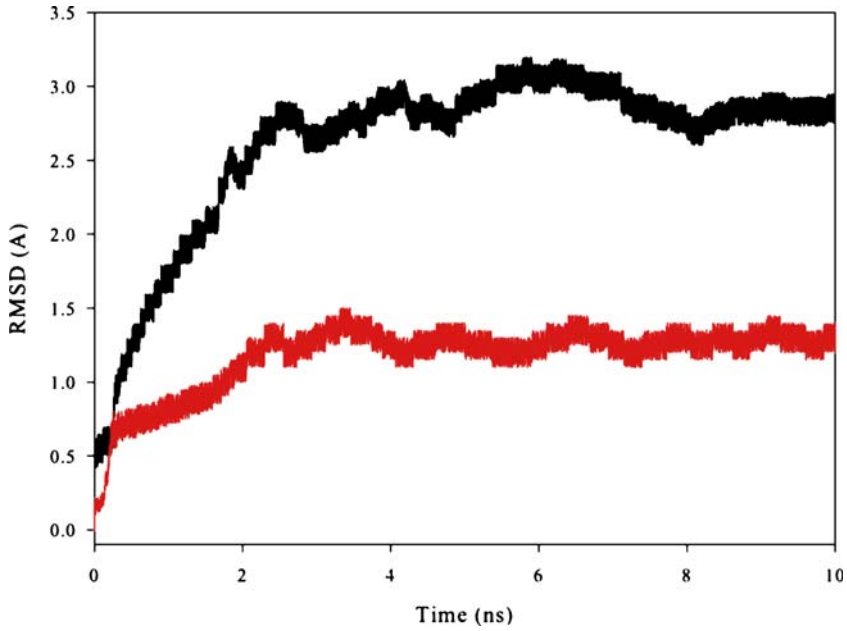

Fig. 5 Time dependence of the root-mean-square deviation (RMSD; Å) from the homology model of H5N1 NS1 and the crystal structures of non-H5N1 NS1 for the Cα atoms in the 10 ns molecular dynamics (MD) simulations

WHAT-IF is used to check the normality of the local environment of amino acids [33]. For the WHAT-IF evaluation, the quality of the distribution of atom types is determined around amino fragments. For a reliable structure, the WHAT-IF packing scores should be above −5.0. In this case, none of the scores for each residue is lower than −5.0 as depicted in Fig. 3. Therefore, the WHAT-IF evaluation also indicates that the homology model structure is very reasonable.

The interaction energy per residue was also calculated using the PROSA2003 program [34]. In this analysis, the interaction energy of each residue with the remainder of a protein is computed to judge whether or not it fulfills certain energy criteria. Figure 4 displays the PROSA2003 energy profiles calculated for the homology model along with the template. The energy profile of the homology model is consistent with a reliable conformation based on its similarity with that of the template.

In summary, the geometric quality of the backbone conformation, the residue interaction, the residue contact and the energy profile of the structure are all well within the limits established for reliable structures. All evaluation suggests that a model of high quality for the H5N1 NS1 protein has been obtained that will allow examination of protein–substrate interactions. The resultant structures were then analyzed using the HBPLUS 3.06 [38] and Ligplot 4.22 [39] programs to identify specific interactions.

Molecular dynamics simulations

During the MD simulation, The RMSD (Å) of the protein (H5N1 NS1 and non-H5N1 NS1) backbone atoms relative to the crystal structure of non-H5N1 NS1 studied here are plotted as a function of time (Fig. 5). It should be noted that after 3 ns, the RMSD of each system tends to converge, indicating that these two systems are stable and have reached equilibrium.

However, it should be emphasized that the protein backbone in the simulation deviates more significantly from H5N1-NS1 than in the non-H5N1 NS1 simulation, as

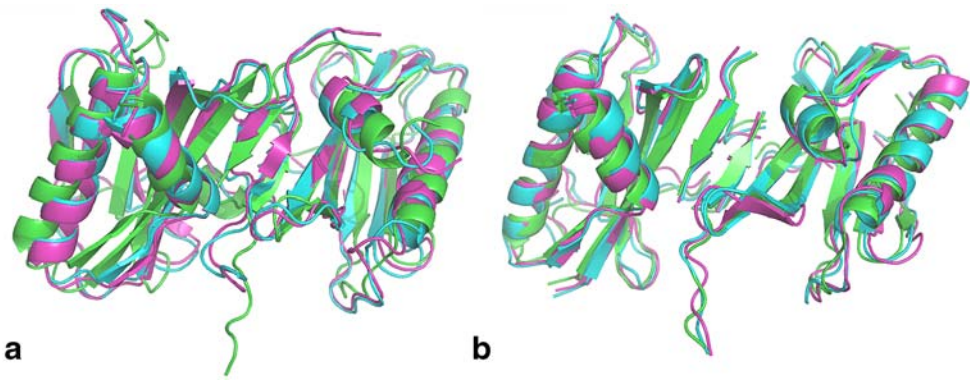

Fig. 6 Superposition of snapshots picked from the two MD trajectories: **a** H5N1 NS1 homology model, **b** non-H5N1 NS1 crystal structure after 0 (*green ribbon*), 5 (*blue ribbon*) and 10 ns (*pink ribbon*) MD simulation

Table 1 Quality of structures checked by PROCHECK during molecular dynamics (MD) simulation

Protein	Time	Ramachandran plot quality (%)[a]				Goodness factor[b]		
		Core	Allowed	General	Disallowed	Dihedrals	Covalent	Overall
H5N1 NS1	0 ns	93.6	5.5	0.0	0.9	0.07	−0.19	−0.03
	5 ns	91.4	7.7	0.0	0.9	−0.37	−0.34	−0.09
	10 ns	90.5	8.6	0.0	0.9	−0.43	−0.42	−0.05
non-H5N1 NS1	0 ns	90.9	8.7	0.5	0.0	−0.16	0.50	0.10
	5 ns	92.7	6.8	0.5	0.0	−0.25	0.54	0.05
	10 ns	93.2	6.4	0.5	0.0	−0.31	0.30	0.09

[a] Ramachandran plot qualities show the percentage (%) of residues belonging to the core, allowed, generally allowed and disallowed region of the plot
[b] Goodness factors show the quality of covalent and overall bond/angle distances; these scores should be above -0.5 for a reliable model

shown in Fig. 6. These results indicate that non-H5N1 NS1 has less structural fluctuation than H5N1 NS1, and thus, non-H5N1 NS1 should have more stability than H5N1 NS1 in the MD simulation.

The overall results of the PROCHECK analyses in MD simulation are shown in Table 1. PROCHECK is used to check the detailed residue-by-residue stereochemical quality of a protein structure. The Ramachandran plots for non-H5N1 NS1 show high similarity during the whole MD simulation. Compared with the non-H5N1 NS1, the H5N1 NS1 homology model has a similar Ramachandran plot, with a relatively low percentage of residues having disallowed torsional angles (0.9%) after 0, 5 and 10 ns MD simulation, respectively. These results again indicate the high quality of our homology model and reasonableness of the MD simulations.

A comparison of the amino acid sequences of the NS1 proteins from various influenza viruses showed that the H5N1/97 NS1 is unique in that it has a Glu residue at position 92, while all of the other known influenza virus strains have an Asp residue at this position [15, 16]. This mutation is very important for the resistance of H5N1/97 to antiviral cytokines. To understand the structural implications of this D92E nutation, three snapshots, after 0, 5 and 10 ns simulation, were taken from each MD trajectory for this Asp/Glu residue structural analysis.

Figure 7 compares the interactions of Asp 92 with other residues in non-H5N1 NS1 after 0, 5 and 10 ns MD simulation. At the beginning of MD simulation, Asp 92 is located at the bottom of a structurally dynamic cleft and is involved in strong hydrogen-bond interactions with Ala 132, Ser 195, and Thr 197, and hydrophobic interactions with Lys 131 (Fig. 7a). After 5 and 10 ns of MD simulation, Asp 92 still has multiple hydrogen-bond interactions with Ala 132, Ser 195 and Thr 197 but lacks hydrophobic contacts with Lys 131.

However, as depicted in Fig. 8, in the case of H5N1 NS1 with the D92E mutation, the interactions of Glu 92 with other residues are different during the whole MD simulation. In the initial conformation, the Glu side chain could engage in only one hydrogen-bond interaction with Ala 132, and hydrophobic interactions with Thr 91, Met 93, Lys 131, Ser 195 and Thr 197. After 5 ns of MD simulation, Glu 92 is positioned to have a weaker hydrophobic interaction with Thr 91, Met 93, Lys 131, Glu 196, and Thr 197, and no hydrogen bond with Ala 132. In brief, the

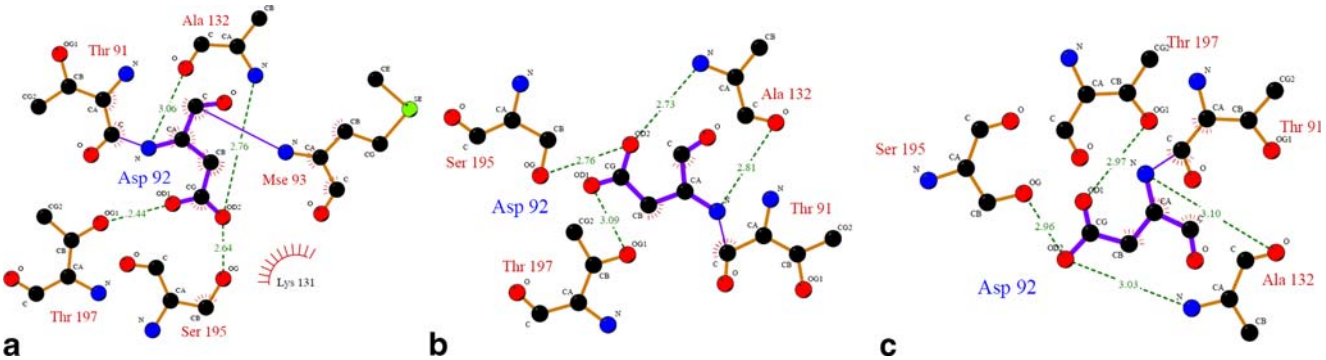

Fig. 7 A schematic illustration of the interactions of Asp 92 with other residues in non-H5N1 NS1 after 0 (**a**), 5 (**b**) and 10 ns (**c**) MD simulation. *Dotted lines* Hydrogen bonds (distances in Å)

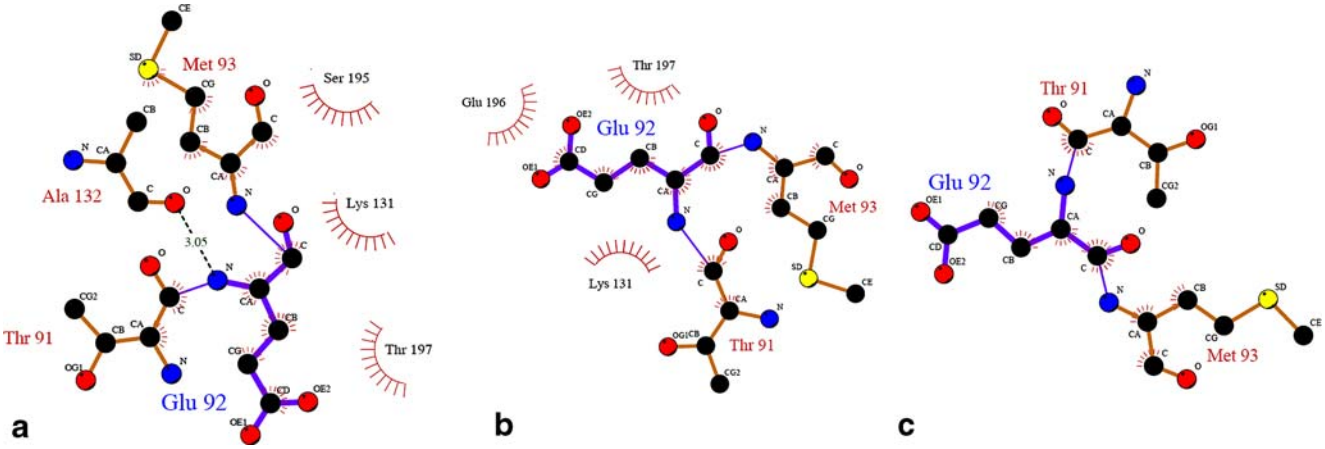

Fig. 8 A schematic illustration of the interactions of Asp 92 with other residues in non-H5N1 NS1 after 0 (**a**), 5 (**b**) and 10 ns (**c**) MD simulation. *Dotted lines* Hydrogen bonds (distances in Å)

interactions of Glu 92 with other residues become weaker and weaker during the MD timeline. Thus, this critical mutation could cause the local destabilization of the region around Glu 92 and also "free up" Ser 195 and Thr 197. Although there has been little study involving the H5N1 virus in this regard, for many other viruses it is well known that phosphorylation of the NS1 protein is very important during viral DNA amplification [40, 41]. Therefore, it is conceivable that the D92E mutation could "free up" Ser 195 and Thr 197 for stronger interactions with host targets or make them more readily available for phosphorylation.

In order to examine whether Ser 195 and Thr 197 could be reasonable phosphorylation sites, the whole H5N1 NS1 sequence was scanned and possible phosphorylation sites were analyzed by NetPhos 2.0, a widely accepted program for phosphorylation site prediction [42]. The results indicate that Ser 195 and Thr 197 are very reasonable phosphorylation sites (Fig. 9), thus further substantiating the possibility that phosphorylation is a factor. In addition, influenza A virus NS1 can be activated by PKR kinase [43], further implying that NS1 phosphorylation could have significant biological consequences.

All of the above could help explain the structural basis for the increased virulence resulting from the D92E mutation of the NS1 protein in H5N1. These data also suggest a possible role for phosphorylation-mediated regulation of H5N1 NS1.

Conclusions

In the present investigation, we have simulated the structure of the H5N1 NS1 protein using homology modeling, molecular mechanics, and MD calculations. The results indicate that the D92E mutation could cause weakened interactions of the carboxylate side chain of Glu with other residues such as Ser 195 and Thr 197. The weakened interactions could "free up" these two residues for stronger interactions with host targets or for phosphorylation, which could in turn lead to other more profound changes of the protein structure and function. Computational analysis using NetPhos 2.0 indicates that both Ser 195 and Thr 197 are very reasonable phosphorylation sites, thus suggesting phosphorylation as one possible scenario for modifying the function of H5N1 NS1 upon D92E mutation. The detailed structural understanding achieved may also help structure-based design of novel antiviral agents against the H5N1 avian influenza virus.

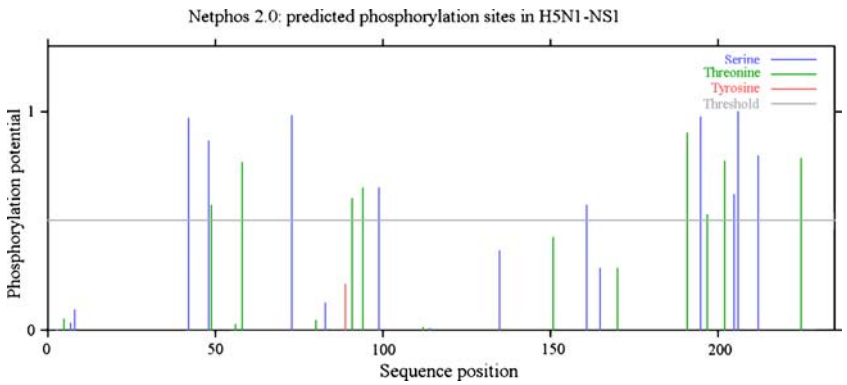

Fig. 9 Kinase-specific NS1 phosphorylation sites as predicted by NetPhos 2.0

Acknowledgments Financial support from the Georgia Cancer Coalition, Georgia Research Alliance, and the National Institutes of Health (CA123329, CA113917) is gratefully acknowledged. We also thank Mr. Victor Bolet for his technical support on computation.

References

1. Claas EC, Osterhaus AD, van Beek R, De Jong JC, Rimmelzwaan GF, Senne DA, Krauss S, Shortridge KF, Webster RG (1998) Lancet 351:472–477
2. Webby RJ, Webster RG (2003) Science 302:1519–1522
3. Gostin LO (2006) JAMA 295:554–556
4. Longini IM Jr, Nizam A, Xu S, Ungchusak K, Hanshaoworakul W, Cummings DA, Halloran ME (2005) Science 309:1083–1087
5. Krug RM, Yuan W, Noah DL, Latham AG (2003) Virology 309:181–189
6. Fortes P, Lamond AI, Ortin J (1995) J Gen Virol 76:1001–1007
7. Chen Z, Li Y, Krug RM (1999) EMBO J 18:2273–2283
8. Nemeroff ME, Barabino SM, Li Y, Keller W, Krug RM (1998) Mol Cell 1:991–1000
9. Talon J, Horvath CM, Polley R, Basler CF, Muster T, Palese P, Garcia-Sastre A (2000) J Virol 74:7989–7996
10. Wang X, Li M, Zheng H, Muster T, Palese P, Beg AA, Garcia-Sastre A (2000) J Virol 74:11566–11573
11. Zurcher T, Marion RM, Ortin J (2000) J Virol 74:8781–8784
12. Obenauer JC, Denson J, Mehta PK, Su X, Mukatira S, Finkelstein DB, Xu X, Wang J, Ma J, Fan Y, Rakestraw KM, Webster RG, Hoffmann E, Krauss S, Zheng J, Zhang Z, Naeve CW (2006) Science 311:1576–1580
13. Twu KY, Noah DL, Rao P, Kuo RL, Krug RM (2006) J Virol 80:3957–3965
14. De Clercq E (2006) Nat Rev Drug Discov 5:1015–1025
15. Seo SH, Hoffmann E, Webster RG (2002) Nat Med 8:950–954
16. Seo SH, Hoffmann E, Webster RG (2004) Virus Res 103:107–113
17. Lipatov AS, Andreansky S, Webby RJ, Hulse DJ, Rehg JE, Krauss S, Perez DR, Doherty PC, Webster RG, Sangster MY (2005) J Gen Virol 86:1121–1130
18. Li M, Wang B (2006) Biochem Biophys Res Commun 347:662–668
19. Yamada S, Suzuki Y, Suzuki T, Le MQ, Nidom CA, Sakai-Tagawa Y, Muramoto Y, Ito M, Kiso M, Horimoto T, Shinya K, Sawada T, Kiso M, Usui T, Murata T, Lin Y, Hay A, Haire LF, Stevens DJ, Russell RJ, Gamblin SJ, Skehel JJ, Kawaoka Y (2006) Nature 444:378–382
20. Yang ZY, Wei CJ, Kong WP, Wu L, Xu L, Smith DF, Nabel GJ (2007) Science 317:825–828
21. Li M-Y, Lu J-F, Xia L (2005) Acta Chim Sin 63:1875–1883
22. Du LP, Li MY, Tsai KC, You QD, Xia L (2005) Biochem Biophys Res Commun 332:677–687
23. Ginalski K (2006) Curr Opin Struct Biol 16:172–177
24. Du L, Li M, You Q, Xia L (2007) Biochem Biophys Res Commun 355:889–894
25. Tramontano A (1998) Methods 14:293–300
26. Bornholdt ZA, Venkataram Prasad BV (2006) Nat Struc Mol Biol 13:559–560
27. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M (2003) Nucleic Acids Res 31:365–370
28. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) Nucleic Acids Res 25:4876–4882
29. Fiser A, Sali A (2003) Methods Enzymol 374:461–491
30. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) J Biomol NMR 8:477–486
31. Colovos C, Yeates TO (1993) Protein Sci 2:1511–1519
32. Eisenberg D, Luthy R, Bowie JU (1997) Methods Enzymol 277:396–404
33. Vriend G, Sander C (1993) J Appl Crystallogr 26:47–60
34. Tomii K, Hirokawa T, Motono C (2005) Proteins 61(Suppl 7):114–121
35. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
36. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz Jr KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) J Comput Chem 26:1668–1688
37. DeLano WL (2006) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, http://www.pymol.org
38. McDonald IK, Thornton JM (1994) J Mol Biol 238:777–793
39. Wallace AC, Laskowski RA, Thornton JM (1995) Protein Eng 8:127–134
40. Corbau R, Salom N, Rommelaere J, Nuesch JP (1999) Virology 259:402–415
41. Nuesch JP, Lachmann S, Corbau R, Rommelaere J (2003) J Virol 77:433–442
42. Blom N, Gammeltoft S, Brunak S (1999) J Mol Biol 294:1351–1362
43. Li S, Min JY, Krug RM, Sen GC (2006) Virology 349:13–21