Structural Basis of the Influenza A Virus RNA Polymerase PB2 RNA-binding Domain Containing the Pathogenicity-determinant Lysine 627 Residue

Received for publication, December 8, 2008, and in revised form, December 29, 2008 Published, JBC Papers in Press, January 14, 2009, DOI 10.1074/jbc.C800224200

Takashi Kuzuhara,1 Daisuke Kise, Hiroko Yoshida1, Takahiro Horita, Yoshimi Murazaki, Akie Nishimura, Noriko Echigo, Hiroko Utsunomiya,2 and Hideaki Tsuge2

From the 1Laboratory of Biochemistry, Faculty of Pharmaceutical Sciences, and 2Institute for Health Sciences, Tokushima Bunri University, Tokushima 770-8514, Japan

Because the influenza A virus has an RNA genome, its RNA-dependent RNA polymerase, comprising the PA, PB1, and PB2 subunits, is essential for viral transcription and replication. The binding of RNA primers/promoters to the polymerases is an initiation step in viral transcription. In our current study, we reveal the 2.7 Å tertiary structure of the C-terminal RNA-binding domain of PB2 by x-ray crystallography. This domain incorporates lysine 627 of PB2, and this residue is associated with the high pathogenicity and host range restriction of influenza A virus. We found from our current analyses that this lysine is located in a unique “ϕ”-shaped structure consisting of a helix and an encircled loop within the PB2 domain. By electrostatic analysis, we identified a highly basic groove along with this ϕ loop and found that lysine 627 is located in the ϕ loop. A PB2 domain mutant in which glutamic acid is substituted at position 627 shows significantly lower RNA binding activity. This is the first report to show a relationship between RNA binding activity and the pathogenicity-determinant lysine 627. Using the Matras program for protein three-dimensional structural comparisons, we further found that the helix bundles in the PB2 domain are similar to that of activator 1, the 40-kDa subunit of DNA replication clamp loader (replication factor C), which is also an RNA-binding protein. This suggests a functional and structural relationship between the RNA-binding mechanisms underlying both influenza A viral transcription and cellular DNA replication. Our present results thus provide important new information for developing novel drugs that target the primer/promoter RNA binding of viral RNA polymerases.

In 1918, a pandemic expansion of influenza A virus resulted in ten million deaths worldwide (1), and strategies to prevent any future expansions of this virus are therefore an important endeavor (2, 3). Although inhibitors of neuraminidase and the M2 ion channel are widely used as anti-influenza A virus drugs (4, 5), some adverse effects of these agents and also the emergence of drug-resistant viruses have been reported (6, 7). Because the transcription and replication of influenza A virus requires the activity of its RNA-dependent RNA polymerase, this enzyme represents a very promising target for anti-viral drug development. Influenza A virus digests host mRNA via the PB1 subunit of its RNA polymerase (8). A 5′-RNA fragment of 10–15 nucleotides is used as a primer for transcription initiation (8) through primer and promoter RNA binding activities within the C-terminal region of its RNA polymerase PB2 domain (9–11). The PA subunit has protease activity and binds to PB (2). The determination of the tertiary structure of a protein is a valuable tool in the development of a targeting drug, and low resolution structures have previously been determined for the PA, PB1, and PB2 subunits of influenza A virus RNA polymerase by electron microscopy (12). In addition, the structure of PB2 nuclear localization signal domain (amino acid (aa)3 residues 687–759), the PB2 cap (7-methyl guanosine triphosphate)-binding domain (aa 318–483), and the C-terminal domain of PA with the N-terminal domain of PB1 have also been reported (13–17). The C-terminal domain of PB2 includes the Lys627 residue, which plays a role in the high pathogenicity and host range restriction of the influenza virus. The structure of PB2 has not been previously elucidated, and the structure-function relationships between this domain and Lys627 represent a significant advance in our understanding of the pathogenicity of influenza A.

EXPERIMENTAL PROCEDURES

Expression and Purification of the Influenza A Virus RNA Polymerase PB2 3/3 RNA-binding Domain—A cDNA fragment of the influenza A virus RNA polymerase PB2 3/3 domain (aa 535–759) was amplified by PCR from pBMSA-PB2 (RIKEN) using the following primers: PB2 Met 3/3, 5′-GCCGTTCATA-TGATGTGGGAGATTAATGGT-3′, and PB2 stop BamHI, 5′-GCCGTGGATCCCTTAATTTGATGCCATCCGAAAT-3′. The amplified DNA fragment was then subcloned into the pET28a(+) plasmid (Novagen, Madison, WI) at the NdeI and

1 This work was supported by a grant-in-aid for scientific research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. The atomic coordinates and structure factors (code 3CW4) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

2 The on-line version of this article (available at http://www.jbc.org/) contains two supplemental figures and a supplemental table.

1 To whom correspondence may be addressed: Laboratory of Biochemistry, Faculty of Pharmaceutical Sciences, Tokushima Bunri University, Yamashiro-cho, Tokushima, 770-8514, Japan. Fax: 81-88-635-3051; E-mail: kuzuhara@ph.bunri-u.ac.jp.

2 To whom correspondence may be addressed: Institute for Health Sciences, Tokushima Bunri University, Yamashiro-cho, Tokushima, 770-8514, Japan. Fax: 81-88-622-2503; E-mail: tsuge@tokushima.bunri-u.ac.jp.

3 The abbreviations used are: aa, amino acids; SeMet, seleno-methionine; MAD, multiple anomalous dispersion; vRNA, viral RNA; CM, carbamoylmethyl.
Influenza A Virus RNA-binding Domain

BamHI restriction sites. This construct was then introduced into BL21-CodonPlus (Stratagene, La Jolla, CA) Escherichia coli. The induction of His$_6$-tagged recombinant protein expression was achieved by the addition of isopropyl β-d-thiogalactopyranoside in TBG-M9 medium, and this was followed by purification using Ni$^{2+}$-agarose. For further purification, the His-tagged proteins were cleaved by thrombin and purified using a HiTrap$^\text{TM}$ carboxymethyl (CM) FF column (GE Healthcare, Buckinghamshire, UK) with the AKTAPrime$^\text{TM}$ plus system (GE Healthcare). For phase determination, we expressed the recombinant proteins using seleno-methionine (SeMet)-containing M9 medium (18) in E. coli B836 (DE3) cells (Novagen) to enable multiple wavelength anomalous dispersion (MAD) analysis. Site-directed mutagenesis for lysine 627 was performed by PCR. We purified PB2 3/3 Glu627 and PB2 3/3 for reference using a HiTrap$^\text{TM}$ SP XL column (GE Healthcare).

**RNA Binding Analysis by Surface Plasmon Resonance (Biacore) Assay**—A Biacore 3000 instrument (Biacore AB, Upplands, Sweden) was used to perform surface plasmon resonance assays in accordance with the manufacturer’s instructions. The buffers used in these analyses were as follows: HBS-EP running buffer (10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% surfactant P20) and sensor chip streptavidin to which 100 μM of 20 μM biotinylated RNA oligomers were immobilized. A 5′-vRNA promoter (5′-AGUAGAAACAAGGCC-3′), capped primer (5′-m7GmAAUAUCUCAG-3′), or non-capped primer (5′-AmAAUACUCAG-3′) were synthesized by Hokkaido System Science Co. Ltd., (Hokkaido, Japan), and successful capping of the primers was confirmed by mass spectrometry (Voyager 2073 system). These molecules were then tested in the PB2 3/3 domain protein binding assays. Various concentrations (0.015, 0.044, 0.13, 0.40, 1.2, and 3.6 μM) of the purified influenza A virus RNA polymerase PB2 3/3 proteins were then applied to the RNA-immobilized sensor chip in HBS-EP buffer at a 20 μl/min flow rate. A blank control was used as a reference, and these values were subtracted from the raw data.

**Estimation of the K$_d$ (Dissociation Constant) Values of the PB2 3/3 Protein for Different RNA Substrates**—For affinity analyses, the association-dissociation curves for various concentrations of PB2 3/3 or PB2 3/3 Glu$^{627}$ protein were plotted using the BLAevaluation steady state affinity program (GE Healthcare) from which equilibrium dissociation constant (K$_d$) values (mol/liter) were calculated.

**Crystallization**—Purified samples were concentrated to 4–10 mg/ml by ultrafiltration. Optimal conditions for crystallization were determined initially using Crystal Screen 1$^\text{TM}$ and Crystal Screen 2$^\text{TM}$ (Hampton Research, Aliso Viejo, CA), which involved the hanging drop vapor diffusion method at 4°C. Native crystals were obtained using a reservoir solution containing 0.1 M sodium cacodylate hydrochloric acid, pH 6.2, and 2 M sodium formate. Crystals of selenomethionine-labeled PB2 3/3 protein were obtained using a reservoir solution containing 0.1 M sodium cacodylate hydrochloric acid, pH 6.0, and 3 M sodium formate.

**X-ray Diffraction Analysis**—Diffraction data sets for Native and SeMet crystals were collected at 100 K, using a mixture of paratone-N (Hampton Research) and paraffin oil as a cryoprotectant, on a beamline PF-AR NW12 instrument at the Photon Factory (Tsukuba, Japan). The image data were processed using HKL2000 (19) and SCALA (20). The phase was obtained by MAD with three wavelength SeMet data sets using SOLVE and RESOLVE (21). The initial models were constructed using the XtalView program (22). Iterative refinement and model building were subsequently performed using the REFMAC5 module of the CCP4 suite (22). There were no residues found in the disallowed region of the Ramachandran plot as determined by PROCHECK (23).

**RESULTS AND DISCUSSION**

To perform x-ray crystallographic analysis of the RNA polymerase PB2 of influenza A virus (H1N1, A/PR/8/34), it was necessary to obtain soluble quantities of the enzyme. The PB2 subunit was therefore divided into three portions. We expressed PB2 3/3 (aa 535–759) in E. coli as a His$_6$-tagged recombinant protein that was subsequently purified from soluble fractions using Ni$^{2+}$-agarose. We further purified the PB2 3/3 domain protein by cation-exchange chromatography to near homogeneity (98%).

We then subjected the recombinant PB2 3/3 protein to crystallization as described under “Experimental Procedures.” Native crystals were found to belong to the space group P3$_1$2$_1$2$_1$ with the properties a = b = 52.5 Å, c = 156.3 Å. The Matthews volume (V$_{\text{m}}$) was 2.7 Å$^3$ per molecule in an asymmetric unit. SeMet crystals were found in the same space group with the properties a = b = 52.6 Å, c = 156.4 Å. The crystal structure of PB2 3/3 was then determined by MAD phasing guided by the six SeMet positions out of the 10 methionine sites using SeMet crystals (24). The model structure for PB2 3/3 was constructed from 208 residues (aa 535–742), except for the C-terminal 18 residues, and 86 water molecules (R value, 23.7%, and R$_{free}$, 29.0%) using reflection data of up to 2.7 Å. There were no residues found in the disallowed region of the Ramachandran plot.

The determined structure for PB2 3/3 is shown in Fig. 1 and consists of eight helices and seven strands (Fig. 1, a and b). The
The N-terminal region (aa 535–684) is rich in α-helices, whereas the C-terminal portion is predominantly composed of β-strands (Fig. 1, a and b). The structure of the PB2 3/3 C terminus (aa 685–742) consists of three strands and one helix, which is quite similar to the already reported structure of the importin-binding domain of PB2 (Fig. 1b) (13). The N-terminal region, which has not previously been reported, consists of seven helix bundles and four β-strands (Fig. 1a). The C-terminal importin-binding region and the N terminus are linked by a "tethering loop" (aa 676–692) (Fig. 1a). Because the N terminus of PB2 3/3 is located in front of the importin-binding surface of its C-terminal region in the tertiary structure, the binding of the C-terminal region of PB2 to importin requires its N-terminal portion to move or rotate away (domain-domain movement).

Recently, the other structure of the domain was reported by Tarendeau et al. (25). There are eight amino acid differences among the two structures outside of the N-terminal regions (former: this structure (Protein Data Bank accession code 3cw4); latter: Protein Data Bank accession code 2vy6) A613T, F636L, A661T, V667I, T676I, G682S, A684S, and K702R. A661T, G682S, A684S, and K702R are host range mutations from (avian to avian) to (human to human). Unless these changes occur, the tertiary structure is basically identical to the root mean square deviation 1.103 Å. These host range mutations never affect the tertiary structure, suggesting that they directly interact with other host or virus factors. Their structures also do not include the residues Asp678–Gly685 in the tethering loop.

We identified a characteristic structure in PB2 3/3 in which one helix (aa 589–605) is encircled by a loop (aa 623–635) (Fig. 1b–d). We denoted this the "ϕ structure" because its shape resembles this Greek letter (Fig. 1, b and c). In the encircled "ϕ loop," two prolines and one phenylalanine (PFXXXP) are conserved among the influenza A, B, and C virus PB2 subunits (supplemental Fig. 1), suggesting important roles, possibly including structural maintenance, for this region. The residue at position 627 in PB2 has been shown to be a determinant of pathogenicity and host range restriction for the influenza A virus (26–28). Interestingly, lysine 627 is located in the "ϕ loop" (Fig. 1, b and c). In the "ϕ helix," there is also a conserved sequence GXXYXhhR ("h" denotes a hydrophobic amino acid), which would be expected to contribute to the essential functions of the overall ϕ structure in the life cycle of the influenza A virus. Because the side chains of Lys (627: red) QSRMQ (purple) are toward the outside on the flexible ϕ loop, they would be expected to function in interactions with other molecules. The inherent flexibility of this structure further suggests important roles for the PB2 domain, which include the pathogenicity of this virus, as discussed in the next section.

To predict the function of the PB2 3/3 domain, we calculated the electrostatic potentials on its surface using the GRASP pro-
gram (Fig. 2). One surface of PB2 3/3 (Fig. 2a) is highly basic when compared with the opposite side, suggesting that it may play an important role in the association with RNA primers and/or promoters as described in a later section (Fig. 2a). The basic surface of PB2 consists of three parts: (I) a small portion separated by the φ loop (Lys^{586}, Arg^{589}), (II) a boundary of the φ loop (Lys^{627}, Arg^{630}), and (III) a basic groove separated by the φ loop (Fig. 2a). This basic surface is also separated from non-basic surfaces by helix 4 (Fig. 2a).

Based upon our prediction that the basic surface would function in RNA binding, we tested whether the recombinant PB2 3/3 product had RNA primer and/or promoter binding activities using surface plasmon resonance analysis. The interactions of various concentrations (0.015, 0.044, 0.13, 0.40, 1.2, and 3.6 μM) of recombinant PB2 3/3 and PB2 3/3 Glu^{627} proteins were incubated with the RNA oligonucleotides that had been immobilized on a sensor chip. The 5’-vRNA promoter (5’-AGUAAACAAAGGCC-3’), a capped primer (5’-m7GmpAAUACUAAG-3’), or a non-capped primer (5’-AmAAUACUAAG-3’) was tested for PB2 3/3 domain protein binding. The vertical axis in each panel indicates the resonance units (RU), which reflect the number of attached molecules observed in real time (horizontal axis, full scale is 250 s). The red, magenta, green, cyan, blue, and brown lines indicate the binding curves for the 0.015, 0.044, 0.15, 0.44, 1.2, and 3.6 μM protein concentrations, respectively. The data shown indicate the resonance units after subtraction of the background values (no RNA measurements). The labels association and dissociation represent the buffer in the presence or absence of the input molecules, respectively. Affinity analysis of the binding of oligomer RNA to recombinant PB2 3/3 protein was performed, and the calculated $K_d$ and $R_{max}$ values are shown in supplemental Table 1.
To determine whether any preferences exist for the binding of RNA to the PB2 3/3 domain protein, the $K_d$ values were determined for various RNA molecules using the BIAevaluation program (Fig. 3). The $K_d$ value for 5′-vRNA promoter was calculated as 3.2 μM, whereas those for capped and non-capped primers were found to be 12 and 8.8 μM, respectively. This suggests that PB2 3/3 has 2–3-fold higher binding affinity for vRNA promoter in comparison with capped and non-capped RNA oligomers. Honda and Ishihama’s review (9) has suggested that this domain has primer binding activity but did not show any data in support of this. Jung and Brownlee (11) have shown that the PB2 subunit binds to the vRNA promoter but did not show which domain contributes to this activity. We show in our current study that the PB2 3/3 domain has RNA binding activity with a $K_d$ in the μM range. This affinity was found not to be dependent on capping, but our current data do not exclude the possibility that the whole PB2 subunit or trimer RNA polymerase complex binds to capped primers through this domain. The affinity of binding to the vRNA promoter is higher than that found for capped primer RNA, suggesting that this domain facilitates promoter binding with the PB1 subunit (11).

We next focused on the lysine residue at position 627, which is located in the $\phi$ loop (Figs. 1, b and c, and 2, a and b), because this residue has been shown to be a determinant of the pathogenicity and host range restriction of the influenza A virus (4–6). When the residue at position 627 is glutamic acid, influenza A virus shows low temperature sensitivity in mammalian cells (30). When this residue is lysine, the virus shows high pathogenicity in mammals (26, 28). Lysine is almost exclusively found in human influenza virus isolates, whereas glutamic acid is almost exclusively found in avian influenza virus isolates at this position. This amino acid was also previously reported to be involved in the interaction with NP (nucleoprotein) protein (30, 31). Based on our data regarding the basic surface of this domain, we predicted that the residue would also contribute to the binding of acidic molecules such as RNA. Thus, to investigate whether lysine 627 contributes to RNA binding activity, we replaced it with glutamic acid at this position in the PB2 3/3 protein (PB2 3/3 Glu$^{627}$). During the purification of the PB2 3/3 and PB2 3/3 Glu$^{627}$ proteins using CM column chromatography, we found that they had quite different properties. PB2 3/3 bound to CM resin, whereas PB2 3/3 Glu$^{627}$ did not (Fig. 2, c–e). This suggests that the substitution of only one lysine at position 627 with glutamic acid has profound effects upon the surface electric charges of this domain, although there are 13 basic residues on the basic surface of the PB2 3/3 domain. Because the lysine at position 627 is located in the center of the basic surface and within the $\phi$ loop (Fig. 2a), it may have strong effects upon surface electric charges.

To investigate whether lysine 627 contributes to RNA binding, we examined the PB2 3/3 and PB2 3/3 Glu$^{627}$ proteins for this activity using a surface plasmon resonance assay for four types of oligonucleotide RNAs, such as primer/promoter RNA (Fig. 3). The RNA binding activity of PB2 3/3 Glu$^{627}$ was significantly decreased for all RNA oligomers tested when compared with that of PB2 3/3 (Fig. 3 and supplemental Table 1). We thus concluded that residue 627 is indeed involved in RNA binding. The differences in the RNA binding activity of PB2 might therefore contribute to the pathogenicity of influenza in humans. Our current results are also not necessarily inconsistent with previous reports for NP as it is not unusual for a domain to have multiple functions and bind to several molecules. Taken together, our present findings thus indicate that the PB2 surface is a potential candidate for novel anti-influenza A virus drugs from the viewpoint of pathogenicity and host range restriction levels.

To further investigate the functional and structural relationships of the recombinant PB2 3/3 domain, we screened for proteins with a similar tertiary structure using the Matras program (32). The results of these searches indicated that the helix bundles in the N-terminal portion of PB2 3/3 have similarities with that of subunit C from DNA clamp loader (replication factor C). The tertiary structure of DNA clamp loader was reported previously by Kuriyan and colleagues (33) (supplemental Fig. 2a, blue line). The red and blue lines overlap at the helix as shown by the stereo structures (supplemental Fig. 2a). Kuriyan’s model illustrates that DNA clamp loader binds to RNA primers, proliferating cell nuclear autoantigen (PCNA), DNA clamp, and DNA polymerase at the initiation of DNA replication (supplemental Fig. 2c). Taken together, our analysis outlines the structural and functional relationship between influenza A virus transcription and cellular DNA replication from the viewpoint of RNA primer binding. Because each of the two domains, influenza RNA polymerase PB2 3/3 and DNA clamp loader, have RNA binding activity, it is probable that these helix bundles are involved with an RNA binding function.

In conclusion, the unique $\phi$-shaped structure within the influenza A virus RNA-dependent RNA polymerase and the associated structure-function relationships between RNA binding and pathogenicity for this enzyme have the potential to lead to novel anti-influenza A virus drugs. The development of such agents may be achieved by performing docking simulations in silico in which the surfaces of these unique domains can be comprehensively tested using small molecule databases.

Acknowledgments—We thank N. Igarashi, N. Matsugaki, Y. Yamada, and S. Wakatsuki for data collection using the PF AR-NW12. The influenza (A/PR/8/34) RNA polymerase PB2 plasmid, pbMSA-PB2, was provided by the DNA Bank, RIKEN BioResource Center (Tsukuba, Japan; originally deposited by Dr. Susumu Nakada) with the support of National Bio-Resources Project of the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT).

REFERENCES
Influenza A Virus RNA-binding Domain

Structural Basis of the Influenza A Virus RNA Polymerase PB2 RNA-binding Domain Containing the Pathogenicity-determinant Lysine 627 Residue
Takashi Kuzuhara, Daisuke Kise, Hiroko Yoshida, Takahiro Horita, Yoshimi Murazaki, Akie Nishimura, Noriko Echigo, Hiroko Utsunomiya and Hideaki Tsuge

doi: 10.1074/jbc.C800224200 originally published online January 14, 2009

Access the most updated version of this article at doi: 10.1074/jbc.C800224200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2009/01/14/C800224200.DC1

This article cites 33 references, 3 of which can be accessed free at
http://www.jbc.org/content/284/11/6855.full.html#ref-list-1