LETTERS

Crystal structure of an avian influenza polymerase PA_N reveals an endonuclease active site

Puwei Yuan[1]*, Mark Bartlam[2]*, Zhiyong Lou[3]*, Shoudeng Chen[1], Jie Zhou[1], Xiaojing He[1], Zongyang Lv[1], Ruowen Ge[4], Xuemei Li[1,3], Tao Deng[2,5], Ervin Fodor[5], Zihe Rao[1,2,3] & Yingfang Liu[1]

The heterotrimeric influenza virus polymerase, containing the PA, PB1 and PB2 proteins, catalyses viral RNA replication and transcription in the nucleus of infected cells. PB1 holds the polymerase active site[1] and reportedly harbours endonuclease activity[2], whereas PB2 is responsible for cap binding[2–4]. The PA amino terminus is understood to be the major functional part of the PA protein and has been implicated in several roles, including endonuclease[5] and protease activities[6] as well as viral RNA/complementary RNA promoter binding[7]. Here we report the 2.2 ångström (Å) crystal structure of the N-terminal 197 residues of PA, termed PA_N, from an avian influenza H5N1 virus. The PA_N structure has an α/β architecture and reveals a bound magnesium ion coordinated by a motif similar to the (P)DX_N(D/E)XK motif characteristic of many endonucleases. Structural comparisons and mutagenesis analysis of the motif identified in PA_N provide further evidence that PA_N holds an endonuclease active site. Furthermore, functional analysis with *in vivo* ribonucleoprotein reconstitution and direct *in vitro* endonuclease assays strongly suggest that PA_N holds the endonuclease active site and has critical roles in endonuclease activity of the influenza virus polymerase, rather than PB1. The high conservation of this endonuclease active site among influenza strains indicates that PA_N is an important target for the design of new anti-influenza therapeutics.

Highly pathogenic avian influenza A virus strains with H5N1 subtype are entrenched in poultry worldwide and pose a growing threat to human health. Of the 387 reported human cases of avian influenza since 2003, 245 have been fatal (World Health Organisation, September 2008). Understanding the viral replication mechanism mediated by the viral RNA polymerase is crucial for the development of new anti-influenza therapeutics to increase preparedness against a global influenza pandemic.

The influenza virus RNA-dependent RNA polymerase is a heterotrimeric complex (PA, PB1 and PB2) harbouring several enzymatic activities for catalysing both viral RNA transcription (vRNA → messenger RNA) and replication (vRNA → cRNA → vRNA). The functions of PB1 and PB2 are well defined[2], whereas PA has been implicated in a diverse range of functions[5,8–15]. Mutational analysis of PA suggests that PA_N is involved in several functions of the polymerase complex, including protein stability, endonuclease activity, cap binding and promoter binding[5,7].

The PA protein can be cleaved by limited tryptic digestion into two domains: a smaller N-terminal ~25 kDa domain, and a larger carboxy terminal ~55 kDa domain[5,16]. The crystal structure of the large C-terminal domain of PA (residues 257–716, termed PA_C), understood to bind to PB1 for complex formation and nuclear transport, was recently determined by our group and others[17,18]. The structure of PA_C in complex with an inhibitory N-terminal peptide from PB1 (termed $PB1_N$) provides a starting point for further investigation into the structure and function of the influenza virus polymerase and a target for the discovery of anti-influenza therapeutics. In particular, the high conservation of the $PB1_N$ binding site points towards a viable target to design alternatives to the influenza drugs oseltamivir (Tamiflu) and zanamivir (Relenza), which are at present beset by problems of resistance. Here we extend our previous study on the C-terminal region of PA by investigating the structure and function of the smaller PA N-terminal domain.

The N-terminal domain of PA covering residues 1–256 and termed PA_N, was cloned from an avian type A virus isolate (A/goose/Guangdong/1/96 (H5N1)) and expressed in *Escherichia coli*. Recombinant PA_N was purified and crystallized, and the subsequent crystal structure of PA_N was determined by multi-wavelength anomalous dispersion using 2.9 Å data from a seleno-methionyl derivative protein and 2.2 Å native data (Fig. 1a). From the resulting experimental electron density map, PA_N was traced from residues 1–197, with the exception of a flexible loop region from residues 49–75 and the entire C-terminal region from residues 198–256. Data collection, phasing and refinement statistics are listed in Supplementary Table 1.

The PA_N structure has an α/β architecture with five mixed β-strands forming a twisted plane surrounded by seven α-helices (Fig. 1a, b). A negatively charged cavity surrounded by helices α2–α5 and strand β3 houses a bound metal ion (Fig. 1c), which we identified as magnesium for the following reasons: a high concentration of magnesium (0.1 M) was added to the crystallization solution, and atomic absorption results showed a significant amount of magnesium in the washed and resolved PA_N crystals in an approximately 13:9 molar ratio (data not shown). The metal is directly coordinated by five ligands: the acidic residues E80 and D108, three water molecules stabilized by residue H41, E119 and the carbonyl oxygens of L106 and P107 (Fig. 1D). All six amino acids involved in coordinating Mg^{2+} are conserved among influenza A, B and C viruses, with the exception of P107 which is replaced with alanine and cysteine in influenza B and C viruses, respectively (Supplementary Fig. 1).

A Dali (http://www.ebi.ac.uk/dali) search identified several similar structures to PA_N, including: the *Tt*1808 hypothetical protein from *Thermus thermophilus* HB8 (Protein Data Bank (PDB) accession 1WDJ, *Z*-score 4.8, root mean squared deviation (r.m.s.d.) 3.4 Å for 87 amino acids), the restriction endonuclease SdaI (PDB accession 2IXS, *Z*-score 3.9, r.m.s.d. 4.0 Å for 95 amino acids) and the Holliday junction resolvase Hjc (PDB accession 1GEF, *Z*-score 3.8, r.m.s.d. 3.0 Å for 76 amino acids). SdaI belongs to the type IIP DNA restriction endonucleases and its C-terminal domain exhibits a classic nuclease motif containing a seven-stranded planar β-sheet surrounded by four

[1]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. [2]College of Life Sciences and Tianjin Key Laboratory of Protein Science, Nankai University, Tianjin 300071, China. [3]Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China. [4]Department of Biological Sciences, National University of Singapore, Singapore 117543. [5]Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK.
*These authors contributed equally to this work.

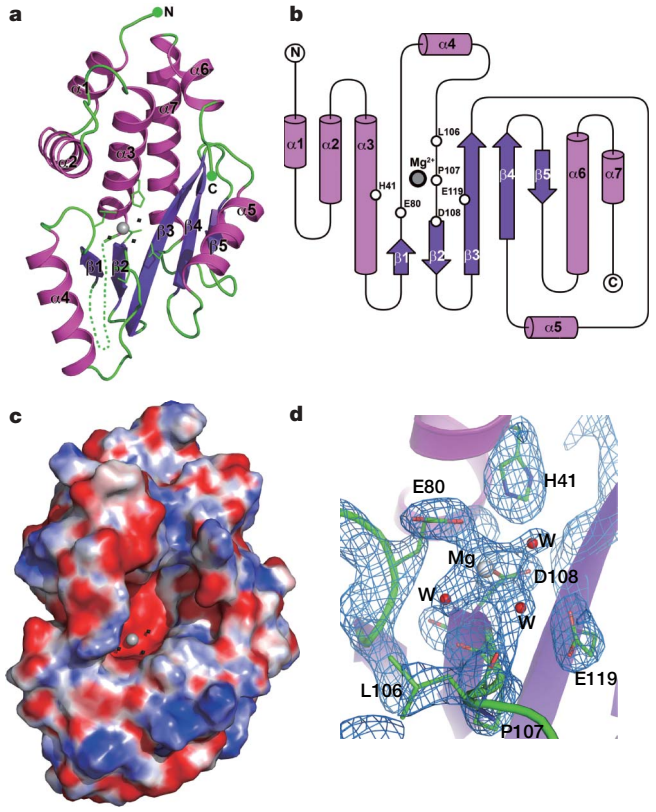

Figure 1 | The PA_N structure. **a**, Ribbon representation showing the PA_N structure. The structure is coloured according to secondary structure elements: α-helices are pink, β-strands are magenta, and loops are green. Individual secondary structure elements are labelled. The Mg^{2+} ion is shown by a silver sphere and the three water molecules are indicated by black dots. **b**, Topology figure of the PA_N structure coloured according to the scheme in **a**. **c**, Surface representation showing the same view of PA_N as in **a**, coloured by electrostatic charge from red ($-10\,k_BT/e_c$, in which k_B is the Boltzmann constant, T is temperature and e_c is the electron charge) to blue ($+10\,k_BT/e_c$). The Mg^{2+} ion is shown as a silver sphere and water molecules are shown by black spheres. **d**, Close-up view of the Mg^{2+} binding site covered by a $2F_o - F_c$ electron density map (contoured at 1.5σ). Residues coordinating the Mg^{2+} ion are shown in stick representation and labelled. The Mg^{2+} ion is shown by a silver sphere and water molecules are shown by red spheres. The PA_N structure is in the same orientation and coloured according to the scheme in **a**.

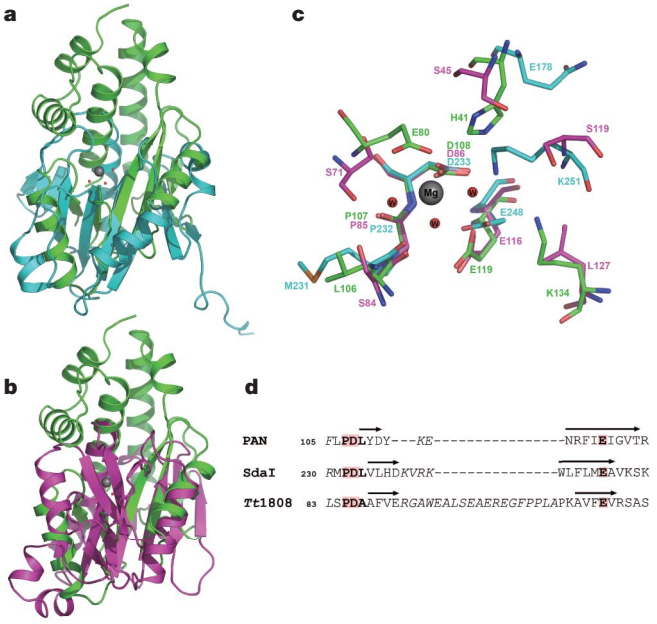

Figure 2 | Structural comparisons suggest PA_N holds an endonuclease active site. **a**, Superposition of the PA_N (green) and SdaI (cyan) structures, viewed in the same orientation as Fig. 1a. **b**, Superposition of the PA_N (green) and *Tt*1808 (magenta) structures, viewed in the same orientation as Fig. 1a. **c**, Superposition of the putative active site residues of PA_N (green) with the active sites of SdaI (cyan) and *Tt*1808 (magenta). Residues are shown in stick representation and labelled. The Mg^{2+} ion and coordinating water molecules from the PA_N structure are shown as silver and red spheres, respectively. **d**, Structure-based sequence alignment of PA_N with SdaI (PDB accession 2IXS) and *Tt*1808 (PDB accession 1WDJ), showing the structurally-conserved endonuclease motif (highlighted in red). Amino acids shown in italic are not aligned.

α-helices. The hypothetical *Tt*1808 is also predicted to be an endonuclease on the basis of structural similarity, and both proteins contain a conserved (P)DX_N(D/E)XK active site motif characteristic of type II endonucleases, with only the lysine position in the motifs varying. The overall structure of PA_N is comparable to both SdaI and *Tt*1808 (Fig. 2a, b). In the case of SdaI, PA_N shares a similar core structure and strands β1–β5 match closely with the central β-sheet (strands β8 and β10–β13) of SdaI (Fig. 2a). There is particular similarity between residues coordinating Mg^{2+} in PA_N and residues in the SdaI active site motif. P107, D108 and E119 of PA_N align well to P232, D233 and E248 of SdaI, respectively, suggesting that PA_N might contain endonuclease activity with a motif similar to the (P)DX_N(D/E)XK motif characteristic of many endonucleases (Fig. 2c, d).

The endonuclease activity of the influenza virus polymerase is critical for snatching capped primers from host mRNA to initiate mRNA transcription. PA is reported to be involved in endonuclease activity, either via its N- or C-terminal domain[5,11]; in particular, several mutations in the N-terminal of PA inhibited this activity[5]. However, it was previously reported that the endonuclease activity of the influenza polymerase is located in PB1, and the acidic residues E508, E519 and D522 are essential for this activity[2], whereas earlier studies suggested that the PB2 subunit might contain the endonuclease active site[19–21].

To resolve this issue, we mutated residues in PA_N involved in metal coordination, including H41, E80, L106, P107, D108 and E119, and performed *in vivo* ribonucleoprotein (RNP) reconstitution assays followed by primer extension analysis of the steady-state levels of all three species of viral RNAs synthesized by the virus polymerase (Fig. 3a)[11]. We also included residue K134, which is located close to the proposed endonuclease active site and has been shown to specifically inhibit viral RNA transcription but not replication in previous studies[5]. Polymerases with E80A, D108, E119A and K134A point mutations in PA showed background levels of mRNA synthesis, while retaining significant cRNA and vRNA synthesis activity, in comparison with the wild-type polymerase (Fig. 3a, b). In contrast, L106A and P107A retained mRNA synthesis activity (L106A at near wild-type levels), probably because they interact with the Mg^{2+} ion via their carbonyl oxygens (Figs 1d and 3a). However, they both showed a reduction in cRNA and vRNA synthesis, whereas H41A showed no detectable synthesis of any of the three viral RNAs. Notably, point mutations of the PB1 residues, E508, E519 and D522 (Fig. 3a), which have previously been claimed to be the polymerase endonuclease active centre[2], resulted in significant levels of activity. Sequence similarity searches do not identify a possible endonuclease activity motif around residues E508, E519 and D522 in PB1 (results not shown). These observations strongly suggest that PA provides a centre for polymerase endonuclease activity, rather than PB1.

To address this further, we partially purified a recombinant viral RNA polymerase complex with mutations in the putative endonuclease active site from transfected 293T cells (Fig. 3c). Approximately equal amounts of RNA polymerase were used in transcription assays *in vitro* using an ApG primer that circumvents the need for cap-binding and endonuclease cleavage for transcription initiation. Wild-type or near wild-type levels of activity were observed for the E80A, D108A,

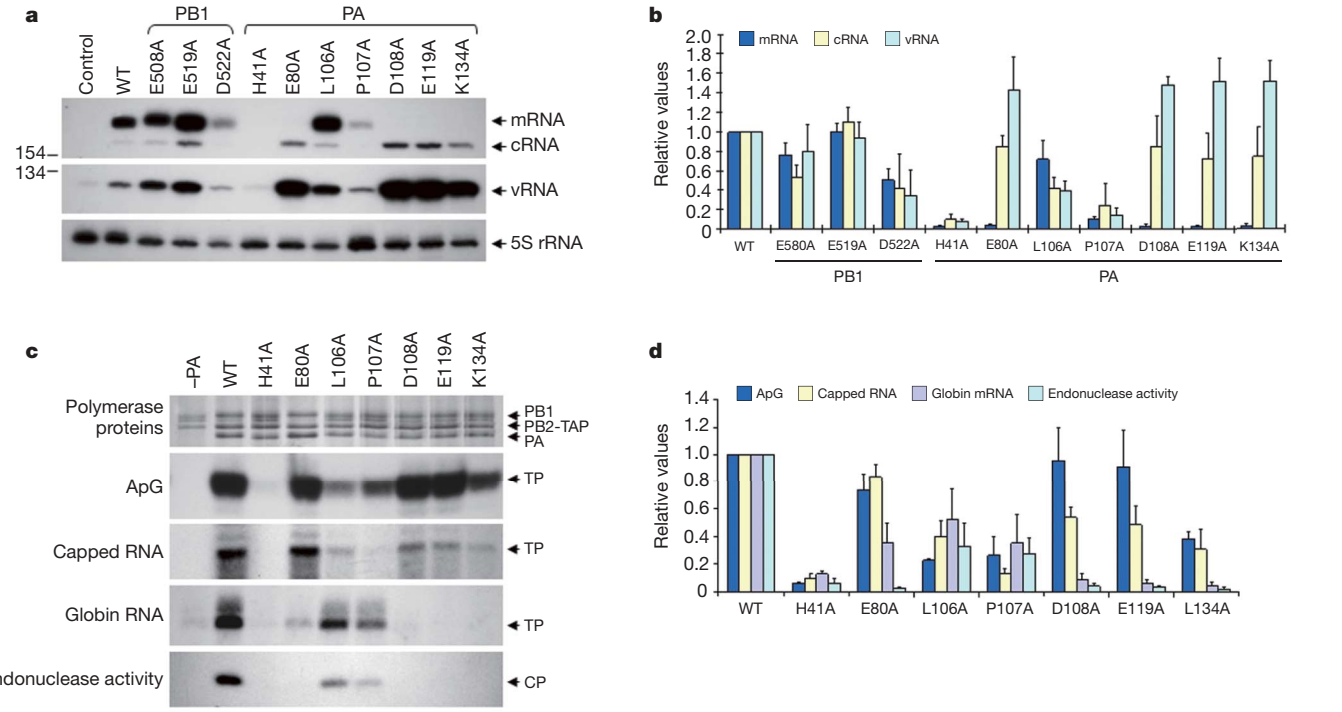

Figure 3 | Effects of mutations in the PA and PB1 subunits on RNA polymerase function. **a**, The effect of polymerase mutations on viral RNA transcription and replication. Wild-type or mutant ribonucleoprotein complexes were reconstituted in 293T cells and viral RNA levels were analysed using a primer extension assay. **b**, Quantification of viral RNA levels from **a** by phosphorimaging. **c**, Effects of mutations on polymerase complex formation, ApG, capped RNA and globin mRNA primed transcription initiation, and endonuclease activity. The position of the three polymerase subunits, PB1, PB2-TAP and PA is indicated. The position of the transcription products (TP) and endonuclease cleavage products (CP) is also indicated. The major, 7-nucleotide capped cleavage product is shown. **d**, Quantification of the results from **c** obtained by phosphorimaging analysis. Data presented in **b** and **d** are an average of three independent experiments, and standard deviations are shown.

E119A and K134A mutants, whereas the L106A and P107A mutants showed reduced but detectable activity (Fig. 3c, d). No activity was detected for the H41A mutant. Similar results were observed if a capped RNA primer, requiring no endonuclease cleavage before elongation, was provided. In contrast, if globin mRNA was provided as a source of capped RNA primer, requiring endonuclease cleavage, most mutations with the exception of L106A and P107A resulted in diminished activity, strongly suggesting that the ability of the polymerase to endonucleolytically cleave capped RNA is primarily affected by the mutations. Indeed, a direct endonuclease cleavage assay using the mutant polymerases showed that all mutants had no detectable activity, with the exception of the L106A and P107A mutations that are proposed to be involved via their carbonyl oxygens (see earlier). It should be noted that although the E80A mutant exhibited some activity in globin mRNA primed transcription, no activity was detected in the endonuclease assay; this may be owing to the different nature of the capped RNA used in the different assays (globin mRNA versus poly(A)-tailed capped RNA). These results obtained *in vitro* are in agreement with the results obtained in the cell-based RNP reconstitution assay and fully support the involvement of this region of PA in endonuclease cleavage activity. Nuclease contamination of PA_N purified from *E. coli* made it impossible to determine whether the isolated PA_N domain has endonuclease activity *in vitro* (data not shown).

Mutation of several residues in the region from D108–F117 of PA_N, including W88, P103, Y110 and F117, results in impaired polymerase activity affecting both transcription and replication[5], whereas the D111A/Y112A double mutation reportedly interferes with transcription and replication[9]. W88, P103, Y110 and F117 form part of a hydrophobic region linking helix α4 to the rest of the PA_N structure (Supplementary Fig. 2a), whereas D111 and Y112 interact with R75 on the loop between α3 and β1 (Supplementary Fig. 2b). Their respective mutations would be expected to destabilize the protein structure and therefore impair its function. Hence the region covering D108 to F117, which overlaps with the endonuclease active site motif, should also be important for the correct folding and stability of the protein and therefore essential for its function. Furthermore, our results indicate that the H41A, L106A and P107A mutations affected the replicative activity of the viral RNA polymerase, suggesting that Mg^{2+} binding might also contribute to the structural integrity of the polymerase in addition to its required role for endonuclease cleavage of capped RNA.

Taken together, the structure and functional assays demonstrate that PA_N contains an endonuclease active site with a putative $P_{107}D_{108}X_{10}E_{119}K_{134}$ active site motif. The first step in the general mechanism of phosphodiester hydrolysis is the preparation of the attacking nucleophile by deprotonation, usually involving a base to deprotonate a water molecule. Lysine is often considered as a general base candidate in endonucleases but it is not strictly conserved[22,23]. The prime candidate for this lysine in PA_N is K134, which is oriented towards the active site; a K134A mutant in functional assays shows a phenotype fully compatible with being involved in endonucleolytic activity. However, from the structure, K134 is located about 7 Å from the Mg^{2+} ion and is therefore too far from the active site to be directly involved. We speculate that in the context of the polymerase trimer, the relative position of α5 harbouring K134 might be slightly altered allowing its direct involvement. A second possibility is that an alternative residue, that is, H41, might serve as the base and K134 has an indirect role.

Our structural data provide no information on how the capped RNA substrate could access the endonuclease active site as it reveals no obvious sites for RNA binding. One possible RNA binding site could be formed by a cluster of four arginines on the protein surface:

two arginine residues (R124 and R125) on the β3–α5 loop, and two arginines (R192 and R196) on helix α7 (Supplementary Fig. 2c). Capped RNA has to be accessed by PB2 and PB1, with PB2 providing the cap-binding domain and PB1 being involved in the elongation of cleaved capped RNA. Previous studies have shown that capped RNA can be cross-linked to PB2 and PB1 (refs 2, 4), but no direct interaction has been reported for PA and capped RNA. We speculate that the capped RNA binding site might be formed by residues contributing from PB1 and PB2 and a co-crystal structure of PA with capped RNA, and possibly including PB1 and PB2, might be required to resolve this issue. PA_N is also associated with the nuclear transport of PB1 (refs 24, 25), and the arginine cluster mentioned above coincides with two putative nuclear transport motifs that are proposed to lie within the N-terminal domain of PA: residues 124–139 (NLS1) and 186–247 (NLS2)[26]. This arginine cluster exposed on the protein surface is consistent with the previous proposal[26] that PA_N contains a bipartite nuclear localization signal.

PA has been linked to proteolysis of viral and host proteins. Residues 1–247 are claimed to be sufficient to induce protein degradation *in vivo* with T157 being involved in both protease and polymerase replication activities[6,27], whereas a conflicting report places S624 of PA_C at the protease active site[12]. Inspection of the environment surrounding T157 in the PA_N structure reveals no obvious candidates for a protease active site (Supplementary Fig. 2d), and our own *in vitro* protease assays following the protocol previously described[12] indicate that PA_N has no detectable proteolytic activity (Supplementary Fig. 3). Further studies are therefore required to clarify the role of PA in protease activity and to determine the location of the active site. Nevertheless, several residues surrounding T157, including E154, K158, D160, E165, E166, R168 and R170, are highly conserved across influenza species (Supplementary Fig. 1), suggesting that this region is an important part of the polymerase complex. For example, region 163–178 of PA has been implicated in the regulation of the promoter binding activities of the RNA polymerase[7].

The availability of structures for the N- and C-terminal domains of PA enables us to build up a more complete picture of the PA protein (Supplementary Fig. 4). PA_N was expressed from residues 1–256, although residues 198–256 could not be observed from the structure, and PA_C expressed from residues 257–716 lacks any regular secondary structure in its N-terminal 17. This leads us to suspect that the region linking PA_N and PA_C is flexible. Our attempts to express the central linker region of PA from residues 197–256 alone proved unsuccessful, and secondary structure predictions indicate that this region should possess more than 60% random coil (data not shown). A recent study confirmed our observations and showed that neither the N- nor the C-terminal domains of PA could ensure a stable interaction with PB1 without the presence of the linker[16]. Furthermore, protease treatment of the PA–PB1 complex showed that its PA protein is markedly more stable than free PA, indicating that the linker is protected from digestion by PB1 and forms an essential part of the subunit interface[16]. This interface spans much of the PA sequence, with the exception of the N-terminal 154 amino acids, and is consistent with previous reports mapping the interaction domain of PA and PB1 to the C-terminal two-thirds of PA[13,28]. The presence of the linker between the N- and C-terminal domains of PA provides a degree of conformational flexibility that may enable it to have a role in regulating polymerase functions through conformational changes of polymerase complexes[15]. In fact, mutations in the linker region are known to affect polymerase function: L226P results in accumulation of the polymerase complex in an inactive form, whereas F205A and L214A mutants diminish RNA synthesis activity[5].

We have determined the crystal structure of the N-terminal domain of PA from residues 1–197 and identified an endonuclease active motif in this domain. Mutational analysis of the identified motif confirmed that PA_N, rather than PB1, harbours an endonuclease active site and suggests that the activity should be Mg^{2+}-dependent. Furthermore, the high conservation of residues in the site indicates that it is general to all PA_N of different influenza viruses. This provides a feasible target for the discovery of new anti-influenza therapeutics, which are urgently needed owing to the increasing problems of resistance to available influenza drugs such as oseltamivir (Tamiflu) and zanamivir (Relenza). The PA_N structure reported here, combined with our previous structure of the C-terminal domain of PA, offers a more complete view of the PA protein and its functional roles in the influenza polymerase complex. In particular, our work provides a much-needed structural basis for researchers to understand the role of PA_N in protein stability, endonuclease activity, cap binding and vRNA promoter binding.

METHODS SUMMARY

Residues 1–256 of the avian H5N1 influenza A virus (A/goose/Guangdong/1/96) PA gene were cloned into the pGEX-6p vector (GE Healthcare) and overexpressed in *E. coli* strain BL21. The recombinant protein was purified with a glutathione affinity column (GE Healthcare). Glutathione *S*-transferase (GST) was cleaved with PreScission protease (GE Healthcare). The protein complex was further purified by Q ion exchange and Superdex-200 gel filtration chromatography (GE Healthcare).

Native PA_N was crystallized in the space group $P1$ using 25% PEG8000 at pH 6.5. A selenomethionyl derivative was crystallized in $P6_422$ using 20% PEG3350 at pH 6.5. The structure was phased to 2.9 Å by multiple-wavelength anomalous dispersion from a selenomethionyl derivative, and traced using 2.2 Å native data. The final model, with an R-factor of 23% and R_{free} of 25%, contains residues 1–197 of PA and is missing residues 49–75.

Plasmids pcDNA-PA, pcDNA-PB1, pcDNA-PB2, pcDNA-NP and pPOLI-NA-RT used in ribonucleoprotein reconstitution assays were identical to those in previous reports[11,29]. Primer extension assays of viral RNAs were performed as previously described[29].

Point mutations were introduced using PCR by designing mutated residues in primers. Mutated genes were cloned into the pcDNA3.1 vector. The pcDNA-PA, pcDNA-PB1 and pcDNA-PB2-TAP plasmids were co-transfected into 293T cells, which were collected 2 days later. Recombinant proteins were purified by tandem affinity purification (TAP) and stored at −80 °C[5]. *In vitro* transcription assays using ApG or capped RNA primers or globin mRNA as a source of capped RNA primer were performed as described[5,11]. Capped RNA (m^7GpppACACUUGCUUUUG) for the capped RNA primed transcription assay has been described[30]. The capped ^{32}P-labelled RNA was poly(A)-tailed for endonuclease cleavage assays[11]. A 7-nucleotide major and a 13-nucleotide minor cleavage product were identified in endonuclease cleavage reactions, in agreement with the finding that endonucleolytic cleavage occurs preferentially at purine residues.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 November; accepted 12 December 2008.
Published online 4 February 2009.

1. Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8,** 3867–3874 (1989).
2. Li, M. L., Rao, P. & Krug, R. M. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *EMBO J.* **20,** 2078–2086 (2001).
3. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Mol. Biol.* **15,** 500–506 (2008).
4. Fechter, P. *et al.* Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. *J. Biol. Chem.* **278,** 20381–20388 (2003).
5. Hara, K., Schmidt, F. I., Crow, M. & Brownlee, G. G. Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *J. Virol.* **80,** 7789–7798 (2006).
6. Sanz-Ezquerro, J. J., Zurcher, T., de la Luna, S., Ortin, J. & Nieto, A. The amino-terminal one-third of the influenza virus PA protein is responsible for the induction of proteolysis. *J. Virol.* **70,** 1905–1911 (1996).
7. Maier, H. J., Kashiwagi, T., Hara, K. & Brownlee, G. G. Differential role of the influenza A virus polymerase PA subunit for vRNA and cRNA promoter binding. *Virology* **370,** 194–204 (2008).
8. Lee, M. T. *et al.* Definition of the minimal viral components required for the initiation of unprimed RNA synthesis by influenza virus RNA polymerase. *Nucleic Acids Res.* **30,** 429–438 (2002).
9. Regan, J. F., Liang, Y. & Parslow, T. G. Defective assembly of influenza A virus due to a mutation in the polymerase subunit PA. *J. Virol.* **80,** 252–261 (2006).

10. Naffakh, N., Massin, P. & van der Werf, S. The transcription/replication activity of the polymerase of influenza A viruses is not correlated with the level of proteolysis induced by the PA subunit. *Virology* **285,** 244–252 (2001).
11. Fodor, E. *et al.* A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J. Virol.* **76,** 8989–9001 (2002).
12. Hara, K. *et al.* Influenza virus RNA polymerase PA subunit is a novel serine protease with Ser624 at the active site. *Genes Cells* **6,** 87–97 (2001).
13. Zurcher, T., de la Luna, S., Sanz-Ezquerro, J. J., Nieto, A. & Ortin, J. Mutational analysis of the influenza virus A/Victoria/3/75 PA protein: studies of interaction with PB1 protein and identification of a dominant negative mutant. *J. Gen. Virol.* **77,** 1745–1749 (1996).
14. Huarte, M. *et al.* Threonine 157 of influenza virus PA polymerase subunit modulates RNA replication in infectious viruses. *J. Virol.* **77,** 6007–6013 (2003).
15. Kawaguchi, A., Naito, T. & Nagata, K. Involvement of influenza virus PA subunit in assembly of functional RNA polymerase complexes. *J. Virol.* **79,** 732–744 (2005).
16. Guu, T. S., Dong, L., Wittung-Stafshede, P. & Tao, Y. J. Mapping the domain structure of the influenza A virus polymerase acidic protein (PA) and its interaction with the basic protein 1 (PB1) subunit. *Virology* **379,** 135–142 (2008).
17. He, X. *et al.* Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* **454,** 1123–1126 (2008).
18. Obayashi, E. *et al.* The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454,** 1127–1131 (2008).
19. Blok, V. *et al.* Inhibition of the influenza virus RNA-dependent RNA polymerase by antisera directed against the carboxy-terminal region of the PB2 subunit. *J. Gen. Virol.* **77,** 1025–1033 (1996).
20. Honda, A., Mizumoto, K. & Ishihama, A. Minimum molecular architectures for transcription and replication of the influenza virus. *Proc. Natl Acad. Sci. USA* **99,** 13166–13171 (2002).
21. Shi, L., Summers, D. F., Peng, Q. & Galarz, J. M. Influenza A virus RNA polymerase subunit PB2 is the endonuclease which cleaves host cell mRNA and functions only as the trimeric enzyme. *Virology* **208,** 38–47 (1995).
22. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. Structure of restriction endonuclease BamHI and its relationship to EcoRI. *Nature* **368,** 660–664 (1994).
23. Lukacs, C. M., Kucera, R., Schildkraut, I. & Aggarwal, A. K. Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5 A resolution. *Nature Struct. Biol.* **7,** 134–140 (2000).
24. Fodor, E. & Smith, M. The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *J. Virol.* **78,** 9144–9153 (2004).
25. Nieto, A. *et al.* Nuclear transport of influenza virus polymerase PA protein. *Virus Res.* **24,** 65–75 (1992).
26. Nieto, A., de la Luna, S., Barcena, J., Portela, A. & Ortin, J. Complex structure of the nuclear translocation signal of influenza virus polymerase PA subunit. *J. Gen. Virol.* **75,** 29–36 (1994).
27. Perales, B. *et al.* The replication activity of influenza virus polymerase is linked to the capacity of the PA subunit to induce proteolysis. *J. Virol.* **74,** 1307–1312 (2000).
28. Toyoda, T., Adyshev, D. M., Kobayashi, M., Iwata, A. & Ishihama, A. Molecular assembly of the influenza virus RNA polymerase: determination of the subunit-subunit contact sites. *J. Gen. Virol.* **77,** 2149–2157 (1996).
29. Vreede, F. T., Jung, T. E. & Brownlee, G. G. Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates. *J. Virol.* **78,** 9568–9572 (2004).
30. Brownlee, G. G., Fodor, E., Pritlove, D. C., Gould, K. G. & Dalluge, J. J. Solid phase synthesis of 5′-diphosphorylated oligoribonucleotides and their conversion to capped m7Gppp-oligoribonucleotides for use as primers for influenza A virus RNA polymerase *in vitro*. *Nucleic Acids Res.* **23,** 2641–2647 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Chen and K. Yu for providing the A/goose/Guangdong/1/96 influenza PA gene, R. Zhang and A. Joachimiak for assistance with data collection, and G. G. Brownlee for providing materials and for valuable advice. This work was supported by the National Natural Science Foundation of China (grant numbers 30599432, 30221003 and 30721003), the Ministry of Science and Technology International Cooperation Project (2006DFB32420), the Ministry of Science and Technology 863 Project (2006AA02A314 and 2006AA02A322), the Ministry of Science and Technology 973 Project (2006CB504300 and 2007CB914300) and the Medical Research Council (G0700848).

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited with the Protein Data Bank under the accession number 3EBJ. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Z.R. (raozh@xtal.tsinghua.edu.cn) and Y.L. (liuy@ibp.ac.cn).

METHODS

Protein expression and purification. Residues 1–256 of the avian H5N1 influenza A virus (A/goose/Guangdong/1/96 (H5N1)) PA gene were cloned into the pGEX-6p vector (GE Healthcare) and overexpressed in *E. coli* strain BL21. The recombinant proteins were then purified with a glutathione affinity column (GE Healthcare). After cleavage of the GST tag with PreScission protease (GE Healthcare), the protein complex was further purified by Q ion exchange chromatography and then by Superdex-200 gel filtration chromatography (GE Healthcare).

Data collection and structure determination. Crystals were obtained by the vapour diffusion method and grown using with 25% PEG8000 (native crystals) or 20% PEG3350 (selenomethionyl derivative crystals) as precipitant at pH 6.5. The native crystals belong to the space group $P1$ with cell parameters $a = 51.1$ Å, $b = 59.8$ Å, $c = 67.2$ Å, $\alpha = 96.6^\circ$, $\beta = 96.8^\circ$, $\gamma = 109.5^\circ$ and four molecules per asymmetric unit. The selenomethionyl derivative crystals belong to the space group $P6_422$ with cell parameters $a = 73.8$ Å, $b = 73.8$ Å, $c = 123.4$ Å with one molecule in the asymmetric unit.

X-ray diffraction data were collected from a single selenomethionyl derivative PA_N crystal at three wavelengths around the selenium K-edge at the Advanced Photon Source (APS), Chicago, and processed using HKL2000 (ref. 31) to 2.9 Å resolution. An additional high resolution native data set was collected in-house to 2.2 Å resolution using a Rigaku FR-E Cu-Kα rotating anode generator and R-AXIS IV++ detector. The positions of six selenium atoms in the asymmetric unit were determined from the selenomethionyl derivative data sets using SHELX[32]. Initial phase estimates were obtained by MAD phasing with the program MLPHARE and improved by solvent flattening using DM at a nominal solvent content of 41%[33]. The initial electron density map showed clear solvent boundaries, and it was possible to trace 80 residues of the four major helices. Following limited positional refinement using CNS 34, the working R-factor for this model for all amplitudes to 2.8 Å resolution had converged to 49%. After more than 50 cycles of manual main chain building and phase combination, the working R-factor for this model had decreased to 37%. This initial model was used for molecular replacement using PHASER[34] to obtain four clear solutions in the asymmetric unit of the native data set. Further refinement and rebuilding of this model was performed with the guidance of $2F_o - F_c$ and $F_o - F_c$ density maps using CNS[35]. In the final stages, the refinement had converged to yield an R-factor of 23.1% (R_{free} of 25.2%) for all data in the resolution range 30.0–2.2 Å, with refinement of individual isotropic temperature factors and use of bulk solvent correction. Finally, water molecules were added by inspection of sA-weighted ($F_o - F_c$) and $2F_o - F_c$ electron density maps. A total of 107 water molecules were added to the model, and refinement and adjustment of these was performed until no further difference density could reasonably be ascribed to water. Inspection of difference Fourier maps in the later stages of refinement showed electron density (at $>4\sigma$) near D108, and the bound Mg^{2+} ion with three coordinated solvent molecules were carefully added into the model. When analysed using PROCHECK, the final structure has 89% of residues in the most favoured regions of the Ramachandran plot, and only seven residues are in disallowed regions as the result of reasonable hydrogen bonds.

Protease activity assay. The protease activity of PA_N was measured using a previously described protocol[12]. In brief, Suc-LLVY-MCA (Sigma-Aldrich) was used as a substrate, and the fluorescence of 7-amino-4-methylcoumarin from peptide hydrolysis was measured by a Hitachi F-4500 fluorescence spectrophotometer at wavelength 405 nm. Chymotrypsin (Sigma-Aldrich) was used as a positive control.

31. Otwinowski, Z. & Minor, W. in *Macromolecular Crystallography* Vol. 276 (eds Carter, C. W. Jr & Sweet, R. M.) Part A 307–326 (Academic, 1997).
32. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
33. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
34. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D* **57**, 1373–1382 (2001).
35. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).