

Crystal structure of the polymerase PA_C-PB1_N complex from an avian influenza H5N1 virus

Xiaojing He¹, Jie Zhou¹, Mark Bartlam², Rongguang Zhang³, Jianyuan Ma¹, Zhiyong Lou⁴, Xuemei Li^{1,4}, Jingjing Li¹, Andrzej Joachimiak³, Zonghao Zeng¹, Ruowen Ge⁵, Zihao Rao^{1,2,4} & Yingfang Liu¹

The recent emergence of highly pathogenic avian influenza A virus strains with subtype H5N1 pose a global threat to human health¹. Elucidation of the underlying mechanisms of viral replication is critical for development of anti-influenza virus drugs². The influenza RNA-dependent RNA polymerase (RdRp) heterotrimer has crucial roles in viral RNA replication and transcription. It contains three proteins: PA, PB1 and PB2. PB1 harbours polymerase and endonuclease activities and PB2 is responsible for cap binding^{3,4}; PA is implicated in RNA replication^{5–10} and proteolytic activity^{11–14}, although its function is less clearly defined. Here we report the 2.9 ångström structure of avian H5N1 influenza A virus PA (PA_C, residues 257–716) in complex with the PA-binding region of PB1 (PB1_N, residues 1–25). PA_C has a fold resembling a dragon's head with PB1_N clamped into its open 'jaws'. PB1_N is a known inhibitor that blocks assembly of the polymerase heterotrimer and abolishes viral replication. Our structure provides details for the binding of PB1_N to PA_C at the atomic level, demonstrating a potential target for novel anti-influenza therapeutics. We also discuss a potential nucleotide binding site and the roles of some known residues involved in polymerase activity. Furthermore, to explore the role of PA in viral replication and transcription, we propose a model for the influenza RdRp heterotrimer by comparing PA_C with the λ 3 reovirus polymerase structure, and docking the PA_C structure into an available low resolution electron microscopy map.

PA is an important protein in the polymerase heterotrimer and may be required for replication and transcription of viral RNA (vRNA) and endonuclease cleavage of the cap RNA primer^{5–9,15}. It reportedly induces proteolysis of viral and host proteins^{11–14} and may also be involved in virus assembly¹⁶. Recombinant PA can be cleaved into two main fragments by trypsin digestion: a ~25-kilodalton (kDa) fragment containing the amino-terminal region of PA, and a ~55-kDa fragment containing the remainder of the protein¹⁵. The N-terminal fragment is understood to be sufficient for proteolytic activity and important for RNA synthesis activity of the polymerase complex^{11,14,15}. The carboxy-terminal region of PA is thought to bind to PB1 for complex formation and nuclear transport^{17–19}. Previous reports have shown that the PB1 N-terminal 25 residues (PB1_N) specifically bind to the C-terminal region of PA^{2,19,20}.

The structure of an N-terminally truncated PA covering residues 257–716 (termed PA_C) was determined in a complex with a known interaction peptide from PB1 (PB1_N, residues 1–25, Supplementary Fig. 2B) by X-ray crystallography (see Supplementary Table 1 for statistics). PA_C consists of 13 α -helices, one short 3_{10} helix (η 1), nine β -strands and several loops/turns (Fig. 1a, b). PA_C resembles the head of a dragon and can be subdivided into two parts: domain I, the 'brain', and domain II, the 'mouth' (Fig. 1a). The strands β 1 to β 7

of the brain form a twisted plane surrounded by five α -helices (α 1, α 2, α 3, α 6 and α 7) and the η 1 helix. Anti-parallel strands β 6 and β 7 are linked by a four-residue turn and extend from the left side of the PA_C brain. The mouth consists of strands β 8 and β 9, and helices α 4, α 5 and α 8– α 12; helices α 11 and α 13 form the lower jaws, whereas helices α 8 and α 10 form the upper jaws. Structural comparison using Dali (<http://www.ebi.ac.uk/dali>) indicated that PA_C has a new fold.

A deep semi-circular groove (Figs 1a and 2a, b) extending down to the core of the brain is ~25 Å in diameter and highly basic, which is favourable for RNA binding. The extended α 3– α 4 loop consists largely of negatively charged residues and is situated below this deep groove, where it completes the circular structure and forms a 'neck' for the dragon head (Fig. 1a). Residues lining the putative RNA binding groove, including K328, K539, R566 and K574, are highly conserved in all three influenza virus types (Supplementary Fig. 1A, blue stars). The K539A mutation significantly disrupts complementary RNA (cRNA) and vRNA synthesis, suggesting that the groove is important for RNA binding⁵.

An elliptical ~14 Å diameter channel, formed by the concave side of the twisted β -sheet and the inside ends of several helices from domain I, lies between the two domains and connects to the putative RNA binding groove (Fig. 2a, b). It is loosely occupied by the N terminus of the PA_C polypeptide, which is too far from either side to form strong interactions with residues lining the channel and may be a crystallographic artefact (Fig. 2b, c). Conserved residues K539, R566, K574 and N696 lie on the surface connecting the groove to the entrance of the channel. Residues E410, K461, E524 and K536 lie on the inside surface of the channel (Supplementary Fig. 1A, blue stars; and Fig. 2a). Mutations of several of these residues are known to influence polymerase activity⁵: E524A impairs RNA synthesis activity and inhibits virus production; E410A decreases polymerase activity; and the K536A/W537A double mutation impairs RNA synthesis activity of the heterotrimer (Fig. 2c)¹⁶. The precise role of this channel remains to be established, although the highly conserved nature points towards its functional importance.

The R638A mutation promotes generation of interfering RNA, thus giving rise to transcription defects that can be overcome by the C453R mutation (Fig. 2c)⁹. The long side chain of R638 (α 10) forms a hydrogen bond and an ion bridge with the carbonyl oxygen and side chain of residue E449 (α 5), respectively (Supplementary Fig. 3B). C453 (α 6) is about 5 Å from R638 and E449. The R638A mutation would break the interaction between R638 and E449 and cause α 5 to move into the proposed RNA binding surface between PA and PB1, thus interfering with RNA synthesis activity of the heterotrimer. The C453R mutation would enable a new interaction with E449 to be formed that would restrain α 5, thus at least partially recovering PA

¹National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. ²College of Life Sciences and Tianjin State Laboratory of Protein Sciences, Nankai University, Tianjin 300071, China. ³Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. ⁴Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China. ⁵Department of Biological Sciences, National University of Singapore, 117543 Singapore.

from the defect caused by the R638A mutation. The long side chain of C453R may interfere with the normal function of R638, which would explain why a virus containing a single C453R mutation could not be produced. Other mutations known to affect polymerase activity are listed in Supplementary Table 2.

PB1_N binds obliquely between the jaws of PA_C with its N terminus pointing towards the back of the mouth and its C terminus extending outwards (Figs 1a and 2d). Residues from 5 to 11 of PB1_N fold into a short helix (η 2). The interaction with PB1_N is largely mediated by α 8, α 10, α 11 and α 13 of PA_C; the four helices form a hydrophobic core at the tip of the mouth and tightly interact with PB1_N by means of hydrophobic interactions, hydrogen bonds and van der Waals forces (Fig. 2d). Other minor interactions are provided by α 4 and the α 9– α 10 loop. A short LLFL motif from residues 7–10 of PB1_N, known to be important for the interaction with PA¹⁹, interacts with the PA_C hydrophobic core formed by F411 (α 4), M595 (α 8), L666 (α 11), W706 and F710 (α 13), and V636 and L640 (α 10; Fig. 1a). W706 also interacts with residues V3 and N4; Q408 and N412 (α 4) interact with

V3 and D2; and Q670 (α 11) interacts with PB1_N residues F9, V12, P13 and A14. Residues 620 and 621 on β 8 are also located in the PB1_N interaction surface (Supplementary Fig. 1A, green arrows). The W706A/Q670A double mutation disrupts the binding of PB1_N to PA_C (Fig. 2e), as do the L666G/F710E, L666G/F710G and W706A/F710Q double mutations (Supplementary Fig. 4).

PB1_N inhibits influenza A viral replication by interfering with polymerase activity, presumably by blocking assembly of the polymerase heterotrimer². Our structure identifies a very important PB1 binding region of PA, and can thus be used as a basis to design new anti-influenza compounds. Protein–protein interfaces often involve a large surface area, which can present problems for drug discovery. In this case, however, relatively few residues drive the binding of PB1_N to PA_C, indicating that designing small molecule inhibitors

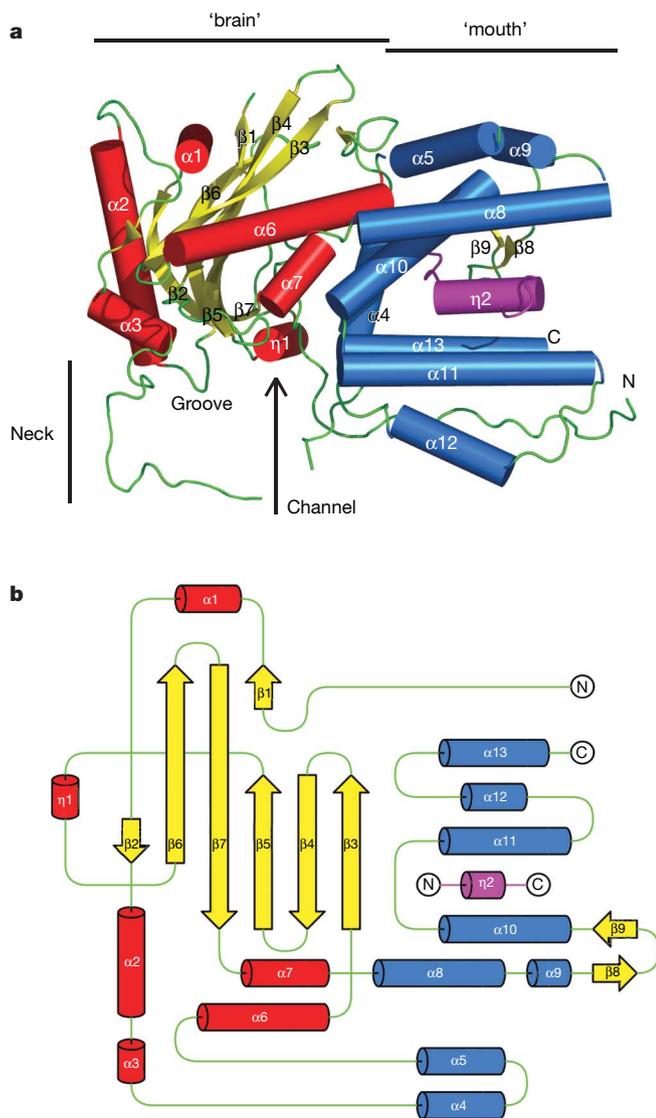


Figure 1 | The PA_C-PB1_N complex structure. **a**, The overall structure of the PA_C-PB1_N complex. The structure is coloured according to secondary structure and elements are labelled. Helices are shown as cylinders and are red in the brain domain and blue in the mouth domain; strands are yellow and loops are green. The PB1_N peptide is magenta. **b**, Topology diagram for the PA_C-PB1_N complex structure. The colouring is consistent with **a**.

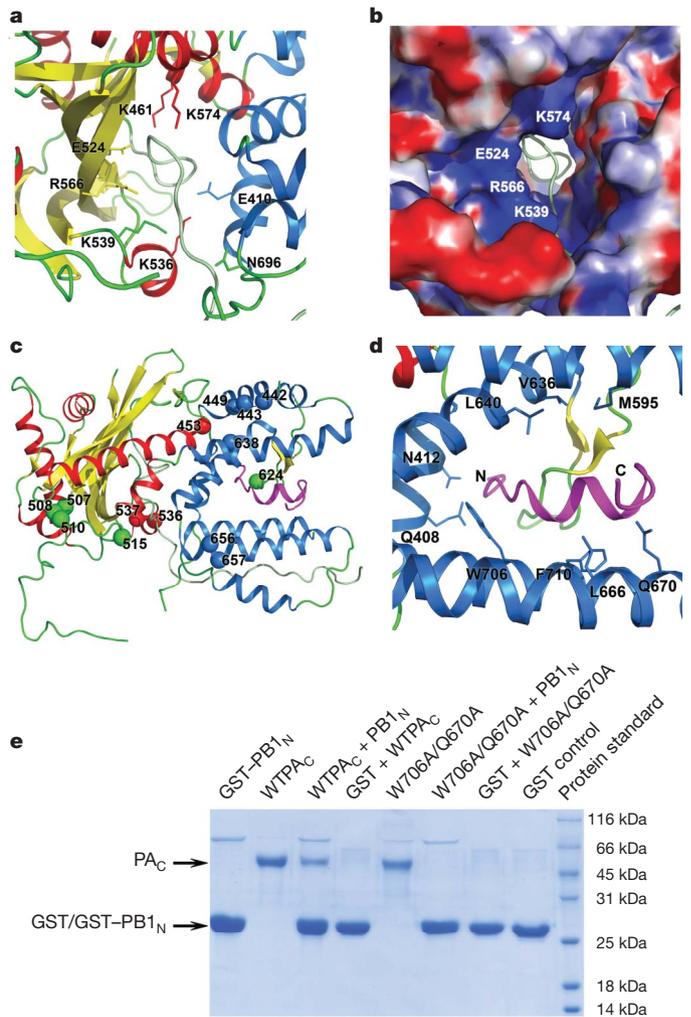


Figure 2 | Representations of the PA_C groove and channel. **a**, An enlarged view of the PA_C groove and channel, rotated by 90° about the *x* axis relative to Fig. 1a and with the same colouring scheme. The extended N-terminal loop from residues 257–288 is coloured pale green. Conserved residues lining the groove and channel are shown in stick representation and labelled. **b**, A surface representation of the same region of PA_C coloured by electrostatic charge from red ($-10 k_b T/e_c$) to blue ($+10 k_b T/e_c$), where k_b is the Boltzmann constant, T is temperature and e_c is the electron charge. Residues lining the highly basic groove are labelled. **c**, Location of known mutations in the PA_C structure. PA_C and PB1_N are shown in ribbon representation and coloured according to the scheme in Fig. 1. **d**, Interaction between PA_C and PB1_N. PA_C and PB1_N are shown in ribbon representation and coloured according to the scheme in Fig. 1. Residues interacting with the PB1_N peptide are shown in stick representation and labelled. **e**, *In vitro* binding assay for PB1_N to the W706A/Q670A PA_C double mutant. WTPA_C, wild-type PA_C.

of this interaction is feasible. Mutation of specific residues in PB1_N, including V3, N4, P5, L7, L8, F9 and L10, causes the loss of more than two-thirds of the binding affinity, with significant reductions in polymerase activity and virus production¹⁹. D2V and A14D mutations in PB1_N do not significantly influence the binding affinity to PA, but they abolish polymerase activity and virus production, suggesting that D2 and A14 have other crucial functions¹⁹. An L13P mutation in PB1 from a H7N7 virus is associated with increased virulence²¹; in our PB1_N structure from a H5N1 virus, P13 breaks the η 2 helix and the associated structural changes may result in increased RNA synthesis by PB1. PB1_N residues interacting with PA are conserved across type A, B and C influenza viruses, and PA residues shown to interact with PB1_N are similarly conserved (Supplementary Fig. 1A, B). Anti-virals designed on the basis of PB1_N, therefore, may be effective against most influenza A strains. Furthermore, the high conservation of the PB1_N binding site on PA suggests that anti-virals targeting this site may be less susceptible to problems of resistance associated with drugs targeting the neuraminidase and haemagglutinin.

PA_C shares similarity with part of the N-terminal domain of reovirus polymerase λ 3, and we used this similarity to propose a model for the RdRp heterotrimer (Fig. 3 and Supplementary Fig. 5). The λ 3 polymerase has three domains that form a hollow 'cage': the N-terminal domain I forms one side of the polymerase; the central catalytic domain II (residues 380–890) contains 'finger', 'palm' and 'thumb' subdomains for catalysis of RNA synthesis; and the C-terminal domain III covers the catalytic cleft and completes the polymerase²². Superimposing and substituting PA_C for the λ 3 polymerase N-terminal domain shows that it is close to the λ 3 catalytic domain, which has an equivalent function to PB1, with very few conflicts between the two polypeptides (Fig. 3a). PA is oriented with the mouth directed towards the λ 3 catalytic domain. PA_C is predicted

from the λ 3 model to have few interactions with the C-terminal domain, consistent with observations that PA and PB2 both form stable complexes with PB1 but have no direct interaction with each other¹⁷.

The spatial arrangement of PA, PB1 and PB2 within the heterotrimer is only partially understood. To put PA_C into the context of the larger RdRp heterotrimer, we docked the crystal structure into an available 26 Å negative staining electron microscopy map of the isolated polymerase heterotrimer (Fig. 3c)^{23,24}. The isolated polymerase map shows a hollow, globular conformation and PA_C fits reasonably well into the front face of the heterotrimer, away from the site of interaction with NP monomers and in good agreement with immunolabelling using an antibody raised against residues 400–716 of PA^{23,24}. Domains II and III of λ 3 were loosely docked into the EM map to highlight the approximate locations of PB1 and PB2 in the heterotrimer (Fig. 3c). The EM polymerase model bears an overall resemblance to the λ 3 polymerase: both form a cage-like structure with four channels providing access to the catalytic site²².

Positioning PA_C into the polymerase heterotrimer enabled us to infer its potential roles in influenza virus replication/transcription. The conserved groove and β 3– β 4 loop of PA_C are located on the inner surface of the polymerase cage. The polymerase is proposed to use different binding strategies for vRNA and cRNA promoters²⁵, although it is not clear if it uses a single promoter binding pocket or separate binding pockets. PA is required for vRNA synthesis but preferentially binds to the cRNA promoter over the vRNA promoter, mainly by the N terminus of the protein²⁶. We propose that the groove/channel in PA_C assists in recruiting cRNA; however, another study suggests that PB1 binds the cRNA promoter²⁷, and further work is needed to resolve this controversy.

H510 lies on the β 4– β 5 loop and is a potential site for endonuclease activity mediated by PB1; a H510A mutation in PA impairs nuclease activity⁵. The area including H510, which should form an endonuclease centre with PA N-terminal residues D108 and K134¹⁵, is directed into the inner cavity of our polymerase model, where it should cleave RNA to produce primers for transport to the catalytic centre of PB1. A proposed nucleotide binding motif from residues 502–509 resembles the GXXXXGKT/S motif²⁸ and lies on the same loop as H510 (Supplementary Fig. 1A, yellow box). The conserved D529 site described previously²⁸ is too far to co-operate with this proposed NTP-binding motif, although D478 on the α 4– β 4 turn is close to the motif and in a region that interacts with R508 and H510. Another possible role for the RNA groove in PA_C may be for host messenger RNA binding before endonuclease cleavage, which is consistent with the endonuclease defect caused by mutation of H510, although this remains to be confirmed by further studies. Because the N-terminal one-third of PA is also reported to be required for polymerase endonuclease activity¹⁵, it should be located adjacent to the region including H510 of PA_C. Further studies are required to determine the precise location and function of the PA N terminus in the heterotrimer.

We report the crystal structure of a principal part of the influenza polymerase PA protein in complex with an inhibitory peptide of PB1. The availability of the PA_C–PB1_N structure offers a starting point for further investigation into the structure and function of the influenza virus polymerase. Together with recent structures of PB2 domains for nuclear import²⁹ and cap binding⁴, the PA_C–PB1_N structure reported here will increase our understanding of the molecular basis for influenza virus replication and transcription. Moreover, there is an urgent need for improved influenza therapeutics; the PA_C–PB1_N complex structure provides a key conserved target for the design of a new generation of compounds that inhibit polymerase assembly and activity.

METHODS SUMMARY

Residues 257–716 of the avian H5N1 influenza A virus (A/goose/Guangdong/1/96) PA gene were cloned into the pGEX-6p vector (GE Healthcare) and

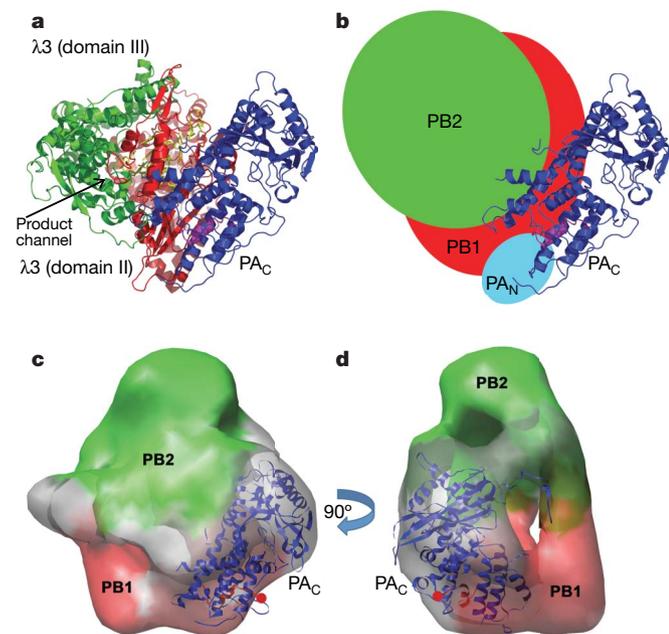


Figure 3 | Model of the influenza polymerase heterotrimer. **a**, Model of the influenza polymerase heterotrimer constructed using the λ 3 reovirus polymerase structure²². PA_C is coloured blue, PB1_N is magenta, λ 3 domain II is red and the λ 3 domain III is in green. **b**, Putative arrangement of PA, PB1 and PB2 in the influenza polymerase heterotrimer. Colouring is the same as in **a**, with PB1 in red and PB2 in green. **c**, **d**, PA_C docked into the electron microscopy map of the isolated heterotrimer²⁴. PA_C is shown in ribbon representation and in blue. Putative locations of the PB1 and PB2 proteins are shown in red and green, respectively.

overexpressed in *Escherichia coli* strain BL21. A gene fragment encoding residues 1–25 of PB1 was also cloned into the same vector and overexpressed in BL21 cells. Cell cultures expressing each of these two proteins were mixed in approximately 1:1 molar ratios of expressed proteins. Recombinant proteins were co-purified with a glutathione affinity column (GE Healthcare). Glutathione S-transferase (GST) was cleaved with PreScission protease (GE Healthcare), and the protein complex was further purified by Q ion exchange chromatography and Superdex-200 gel filtration chromatography (GE Healthcare).

The PA_C–PB1_N complex was crystallized in the space group *P*₄₁₂₁ using 1–1.5 M sodium acetate as the precipitant at a pH of 7.9. The structure was phased to 3.6 Å by multiple-wavelength anomalous dispersion from a selenomethionyl derivative, and traced using 2.9 Å native data. The final refined model, with an *R* factor of 22% and an *R*_{free} of 26%, contains residues 257–716 of PA (with some missing internal fragments) and residues 1–15 of PB1.

Double mutations in PA_C were introduced using the PCR method by designing mutated residues in primers. The mutated genes were also cloned into the pGEX-6p vector. For *in vitro* binding experiments, the wild-type or double mutant PA_C proteins were separately purified from *E. coli* strain BL21. Purified GST–PB1_N fusion peptide or GST alone (native control) were first immobilized on glutathione affinity-column resins (GSH), followed by addition of approximately fivefold molar excess of purified wild-type/double mutant PA_C. GSH resins were extensively washed with PBS (pH 7.4) to remove unbound proteins. The bound protein was eluted from GSH resins with GST elution buffer containing 10 mM reduced glutathione. The eluted sample was analysed by SDS–PAGE electrophoresis and stained by Coomassie blue.

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

Received 25 March; accepted 23 May 2008.

Published online 9 July 2008.

- Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
- Ghanem, A. *et al.* Peptide-mediated interference with influenza A virus polymerase. *J. Virol.* **81**, 7801–7804 (2007).
- 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).
- Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Biol.* **15**, 500–506 (2008).
- 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).
- Jung, T. E. & Brownlee, G. G. A new promoter-binding site in the PB1 subunit of the influenza A virus polymerase. *J. Gen. Virol.* **87**, 679–688 (2006).
- 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).
- Huarte, M. *et al.* Threonine 157 of influenza virus PA polymerase subunit modulates RNA replication in infectious viruses. *J. Virol.* **77**, 6007–6013 (2003).
- Fodor, E., Mingay, L. J., Crow, M., Deng, T. & Brownlee, G. G. A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase promotes the generation of defective interfering RNAs. *J. Virol.* **77**, 5017–5020 (2003).
- Sugiura, A., Ueda, M., Tobita, K. & Enomoto, C. Further isolation and characterization of temperature-sensitive mutants of influenza virus. *Virology* **65**, 363–373 (1975).
- 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).
- 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).
- Rodríguez, A., Perez-Gonzalez, A. & Nieto, A. Influenza virus infection causes specific degradation of the largest subunit of cellular RNA polymerase II. *J. Virol.* **81**, 5315–5324 (2007).
- 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).
- 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).
- 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).
- 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).
- Ohtsu, Y., Honda, Y., Sakata, Y., Kato, H. & Toyoda, T. Fine mapping of the subunit binding sites of influenza virus RNA polymerase. *Microbiol. Immunol.* **46**, 167–175 (2002).
- Perez, D. R. & Donis, R. O. Functional analysis of PA binding by influenza A virus PB1: effects on polymerase activity and viral infectivity. *J. Virol.* **75**, 8127–8136 (2001).
- Gonzalez, S., Zurcher, T. & Ortin, J. Identification of two separate domains in the influenza virus PB1 protein involved in the interaction with the PB2 and PA subunits: a model for the viral RNA polymerase structure. *Nucleic Acids Res.* **24**, 4456–4463 (1996).
- Gabriel, G. *et al.* Differential polymerase activity in avian and mammalian cells determines host range of influenza virus. *J. Virol.* **81**, 9601–9604 (2007).
- Tao, Y., Farsetta, D. L., Nibert, M. L. & Harrison, S. C. RNA synthesis in a cage—structural studies of reovirus polymerase λ 3. *Cell* **111**, 733–745 (2002).
- Area, E. *et al.* 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc. Natl Acad. Sci. USA* **101**, 308–313 (2004).
- Torreira, E. *et al.* Three-dimensional model for the isolated recombinant influenza virus polymerase heterotrimer. *Nucleic Acids Res.* **35**, 3774–3783 (2007).
- Deng, T., Vreede, F. T. & Brownlee, G. G. Different de novo initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. *J. Virol.* **80**, 2337–2348 (2006).
- 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).
- Gonzalez, S. & Ortin, J. Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *EMBO J.* **18**, 3767–3775 (1999).
- de la Luna, S., Martinez, C. & Ortin, J. Molecular cloning and sequencing of influenza virus A/Victoria/3/75 polymerase genes: sequence evolution and prediction of possible functional domains. *Virus Res.* **13**, 143–155 (1989).
- Tarendeau, F. *et al.* Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nature Struct. Biol.* **14**, 229–233 (2007).

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; J. Ortin and O. Llorca for supplying electron microscopy maps; C. Yang, X. Su, F. Sun, L. Wang and R.-M. Xu for advice and technical assistance; and S. Harrison, P. Kuhn, X. Chen and T. Toyoda for discussion. This work was supported by the National Natural Science Foundation of China (grant numbers 30599432 and 30221003), the Ministry of Science and Technology International Cooperation Project (grant number 2006DFB32420), the Ministry of Science and Technology 863 Project (grant numbers 2006AA02A314 and 2006AA02A322) and the Ministry of Science and Technology 973 Project (grant numbers 2006CB504300 and 2007CB914300).

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank under accession number 3CM8. 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 257–716 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. A gene fragment that encodes the N-terminal 25 residues of PB1 was also cloned into the same vector and overexpressed in BL21 cells. Cell cultures expressing each of these two proteins were mixed in approximately 1:1 molar ratios of expressed proteins. The recombinant proteins were then co-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).

In vitro binding experiments. Double point mutations in PA_C were introduced using the PCR method by designing mutated residues in primers. The mutated genes were also cloned into the pGEX-6p vector. For *in vitro* binding experiments, the wild-type or double mutant PA_C proteins were separately purified from *E. coli* strain BL21 as described in the main text. The purified GST–PB1_N fusion peptide or GST alone (native control) was first immobilized on GSH resins, followed by the addition of approximately fivefold molar excess of purified wild-type/double mutant PA_C. GSH resins were extensively washed with PBS (pH 7.4) to remove unbound proteins, and the bound protein was eluted from the GSH resins with GST elution buffer containing 10 mM reduced glutathione. The eluted sample was then analysed by SDS–PAGE electrophoresis and stained by Coomassie blue.

Crystallization and structure determination. Crystals were obtained by the vapour diffusion method with 1–1.5 M sodium acetate as the precipitant at pH 7.9. Native 2.9 Å data were collected in-house ($\lambda = 1.5418$ Å) using an FR-E SuperBright rotating anode X-ray source equipped with an R-Axis IV++ image plate detector (Rigaku). Anomalous X-ray diffraction data were collected to 3.6 Å at peak ($\lambda = 0.9783$ Å) and inflection ($\lambda = 0.9785$ Å) wavelengths from a selenomethionyl-derivative protein at the Advanced Photon Source, Chicago. Diffraction data were processed by HKL-2000 (ref. 30). Peak and inflection data sets were used for phasing by the multi-wavelength anomalous dispersion method³¹. A total of 17 expected heavy atoms were located by SHELXD³² and initial phases were calculated using SHARP³³. Density modification was performed by DM³⁴ and Solomon³⁵. Some secondary structure elements were clearly observed in the experimental electron density map. Approximately 60% of residues were traced automatically by ARP/wARP³⁶ after phase extension by Phenix³⁷ using 2.9 Å native data and phases to 3.6 Å. The remainder of the model was built manually using the program Coot³⁸. The model was refined against native data at 2.9 Å resolution using the programs CNS³⁹ and Refmac⁴⁰. The final refined model, with an *R* factor of 22% and an *R*_{free} of 26%, contains residues 257–716 of PA with some missing internal fragments, and residues 1–15 of PB1. All diagrams were prepared using PyMOL (<http://www.pymol.org>) and electrostatic surface charges were calculated by APBS. Docking of the PA_C crystal structure into the low-resolution electron microscopy map and visualization of the electron microscopy model were performed with UCSF Chimera (<http://www.cgl.ucsf.edu/chimera/>).

30. Otwinowski, Z. & Minor, W. in *Macromolecular Crystallography*, part A (eds Carter C. W. Jr & Sweet, R. M.) 307–326 (Academic, New York, 1997).
31. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
32. Sheldrick, G. M. *Direct Methods for Solving Macromolecular Structures* (ed. Fortier, S.) 401–411 (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998).
33. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
34. Cowtan, K. D. & Zhang, K. Y. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72**, 245–270 (1999).
35. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
36. Perakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
37. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
39. Brunger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).