

# Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit

Franck Tarendeau<sup>1,5</sup>, Julien Boudet<sup>2,5</sup>, Delphine Guilligay<sup>1</sup>, Philippe J Mas<sup>1</sup>, Catherine M Bougault<sup>2</sup>, Sébastien Boulo<sup>3</sup>, Florence Baudin<sup>3</sup>, Rob W H Ruigrok<sup>3</sup>, Nathalie Daigle<sup>4</sup>, Jan Ellenberg<sup>4</sup>, Stephen Cusack<sup>1</sup>, Jean-Pierre Simorre<sup>2</sup> & Darren J Hart<sup>1</sup>

**The trimeric influenza virus polymerase, comprising subunits PA, PB1 and PB2, is responsible for transcription and replication of the segmented viral RNA genome. Using a novel library-based screening technique called expression of soluble proteins by random incremental truncation (ESPRIT), we identified an independently folded C-terminal domain from PB2 and determined its solution structure by NMR. Using green fluorescent protein fusions, we show that both the domain and the full-length PB2 subunit are efficiently imported into the nucleus dependent on a previously overlooked bipartite nuclear localization sequence (NLS). The crystal structure of the domain complexed with human importin  $\alpha 5$  shows how the last 20 residues unfold to permit binding to the import factor. The domain contains three surface residues implicated in adaptation from avian to mammalian hosts. One of these tethers the NLS-containing peptide to the core of the domain in the unbound state.**

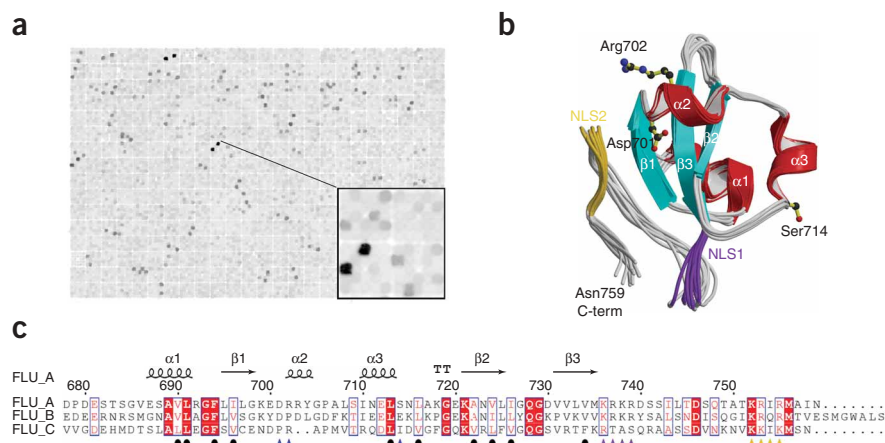
The eight segments of the influenza virus RNA genome are packaged into ribonucleoprotein particles (RNPs) containing the nucleoprotein (NP) and the trimeric RNA-dependent RNA polymerase complex, which comprises subunits PA, PB1 and PB2. The polymerase operates in two distinct modes: in the first, it transcribes virally encoded genes, using a cap-snatching mechanism to prime transcription and ensure proper 5' capping of viral messenger RNA; in the second, it replicates full-length viral RNA to produce first positive-strand complementary RNA and then progeny viral RNA<sup>1</sup>. The PB1 subunit binds the conserved 5' and 3' ends of the viral RNA<sup>2–4</sup> and carries both the polymerase active site and the endonuclease activity that cleaves host-cell mRNA bound by the cap-binding PB2 subunit<sup>5</sup>. PB1 residues implicated in the endonuclease and polymerase active sites have been identified<sup>5</sup>, although the location of the cap-binding site of PB2 remains controversial<sup>5–7</sup>. The polymerase is active in the nucleus and NLSs have been identified on PB1 (ref. 8), PB2 (ref. 9) and PA<sup>10</sup>. Recent results suggest that cytoplasmically expressed PB1 and PA may be imported by the importin RanBP5 as a subcomplex, which then assembles with separately imported PB2 (refs. 11,12). The host proteins mediating PB2 nuclear import have not been characterized, although it seems to be chaperoned by Hsp90 during transport<sup>13</sup>.

Current concern about the adaptation of highly pathogenic H5N1 avian influenza strains to humans has highlighted the need to understand the genetic and molecular determinants of virulence and

transmissibility of influenza viruses<sup>14</sup>. Recent studies have focused attention on polymerase subunit and NP point mutations that accompany natural or laboratory-selected adaptations from avian to mammalian hosts, notably in the case of the 1918 pandemic strain<sup>15–18</sup>. However, the functional implications of these mutations—for instance, whether they affect polymerase activity, assembly, stability or interactions with host factors—are poorly understood. This is partly due to the lack of detailed structural information on the polymerase, although low-resolution negative-stain electron microscopic reconstructions of a mini-RNP<sup>19</sup> and of the polymerase trimer<sup>20</sup> are available. The intractability of the polymerase to structure determination results partly from the fact that bacterially expressed subunits are insoluble and yields of active enzyme produced in insect cells or mammalian cells are prohibitively low. Additionally, biochemical or bioinformatic knowledge of the domain structure of the polymerase subunits is limited. As a new approach to overcome these problems, we devised a directed evolution-type screen called ESPRIT to isolate bacterial clones expressing soluble fragments of polymerase subunits from an essentially complete library of terminal truncation variants. Here we present the structural and functional characterization of a PB2 C-terminal domain identified by this screen. This is the first successful application of a new soluble expression screening strategy that should be of general use in finding soluble domains in otherwise insoluble proteins for both structural and functional studies.

<sup>1</sup>European Molecular Biology Laboratory (EMBL) Grenoble Outstation, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France. <sup>2</sup>Institut de Biologie Structurale Jean-Pierre Ebel UMR 5075 CNRS-CEA-UJF, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France. <sup>3</sup>Unit of Virus Host Cell Interactions UMR 5233 UJF-EMBL-CNRS, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France. <sup>4</sup>EMBL Gene Expression Programme, Meyerhofstrasse 1, D-69117 Heidelberg, Germany. <sup>5</sup>These authors contributed equally to this work. Correspondence should be addressed to D.J.H. (hart@embl.fr) or S.C. (cusack@embl.fr).

Received 13 December 2006; accepted 31 January 2007; published online 25 February 2007; doi:10.1038/nsmb1212



**Figure 1** Identification and solution NMR structure of PB2 C-terminal domain. **(a)** Protein expression screen of 26,880 random deletion constructs of the *pb2* gene. Stable expression of soluble protein results in efficient *in vivo* labeling of a C-terminal biotin acceptor peptide. Detection is by fluorescent streptavidin and fluorimaging. **(b)** Ribbon diagram of the ten lowest-energy NMR structures superimposed using backbone heavy atoms (r.m.s. deviation 0.94 Å for 224 atoms). Indicated are Asp701, Arg702 and Ser714, which are implicated in cross-species transmission, and basic regions corresponding to the minor (purple) and major (gold) sites of the bipartite NLS. **(c)** Primary sequence alignment comparing influenza (FLU) A, B and C strains. Indicated are residues implicated in cross-species transmission (blue triangles), the minor site (purple triangles) and major site (gold triangles) of the bipartite NLS, conserved buried hydrophobic residues (black ovals) and secondary structure elements.

## RESULTS

### Isolation of a soluble domain by random construct screening

An exonuclease protocol<sup>21</sup> was used to generate a comprehensive, unbiased set of 5' *pb2* gene deletions, to which N-terminal methionine and lysine were genetically fused. A library of approximately 26,880 constructs with a seven-fold oversample of each clone was then arrayed robotically onto nitrocellulose membranes and screened for soluble protein expression in a colony format using a C-terminal biotin acceptor peptide<sup>22</sup>. This short peptide is biotinylated *in vivo* by endogenous BirA and has been extensively exploited as a tool for affinity purification and immobilization using avidin. Here we used efficient biotinylation of expressed variants as a quantitative readout for expression of stable, soluble protein. Colony array blots were prepared by *in situ* lysis and expression of biotinylated proteins was detected with a fluorescent streptavidin conjugate and fluorimeter (Fig. 1a). Intense positive clones were sequenced to identify the truncation boundaries, revealing that many were predicted to express C-terminal protein fragments of 80–110 amino acid residues. Two clones in particular, corresponding to residues 678–759 (designated DPDE, from the first four residues) and 661–759 (TTKR), were

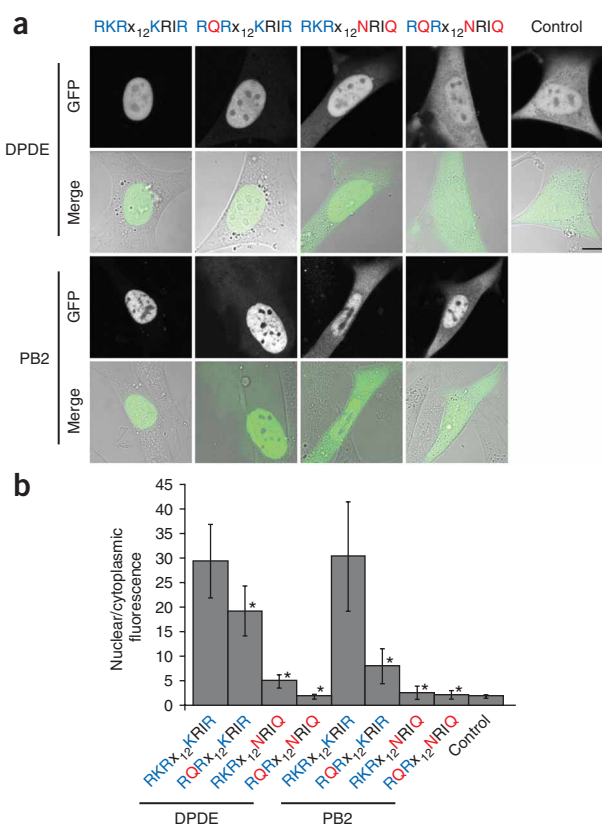
**Figure 2** Nuclear import of PB2 C-terminal domain and full-length PB2 subunit is directed by a bipartite NLS. **(a)** Steady-state subcellular localization of PB2 C-terminal domain (DPDE) or full-length PB2 fused to enhanced GFP in NIH 3T3 cells. Constructs have either the wild-type NLS (737-RKR<sub>x12</sub>KRIR-755) or the point mutations shown in red. Images show 2- $\mu$ m single confocal sections of GFP fluorescence (GFP) and merged DIC and fluorescence (Merge). Contrast and brightness have been adjusted for display purposes. Scale bar, 10  $\mu$ m. **(b)** Ratio of nuclear to cytoplasmic concentration of GFP-fusion proteins, as a measure of nuclear import. Nuclear and cytoplasmic mean fluorescence intensities were background-subtracted and normalized to differences in total cell fluorescence intensity for each mutant. Mean nuclear/cytoplasmic fluorescence intensities are shown with s.d. ( $n = 12$  cells for each mutant). Asterisks mark significant differences from wild-type construct ( $P < 0.05$ , Student's *t*-test).

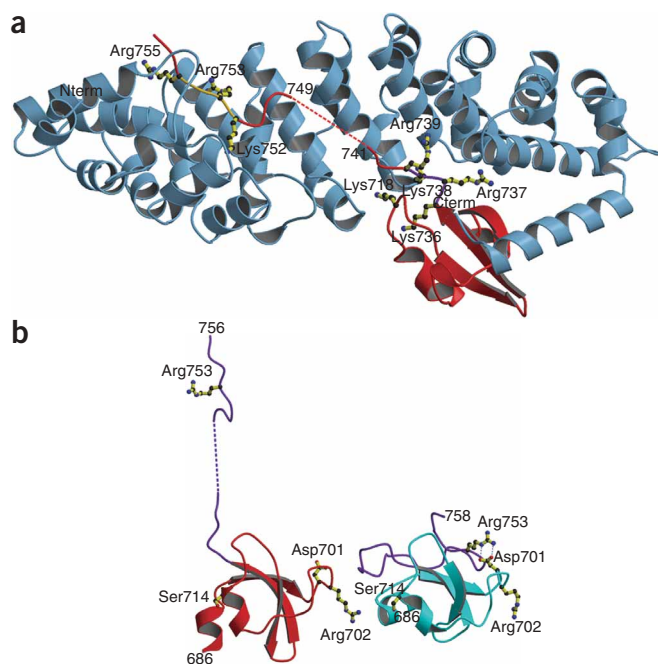
expressed in soluble form at high levels. Multimilligram quantities of each protein were purified and characterized as monodisperse by gel filtration.

### NMR solution structure of the DPDE domain

The <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectrum of DPDE shows a good dispersion of the amide resonances at 10 °C, characteristic of a folded domain (Supplementary Fig. 1 online). TTKR differs only in the extent of disordered N-terminal peptide. Measurements on a sample of uniformly <sup>13</sup>C,<sup>15</sup>N-labeled DPDE led to the assignment of backbone chemical shifts for all residues except Asp678, Gly693 and Ser741. At 25 °C, the intensities of the resonances of the N-terminal (678–685) and C-terminal (744–759) regions were reduced, as were those of Arg702, Gln728, Gly729 and Lys738. Measurements of <sup>15</sup>N-backbone relaxation at 10 °C and analysis of data showed that, with the exception of Lys738, all temperature-sensitive residues belong to flexible parts of the molecule (data not shown). The lack of <sup>1</sup>H-<sup>1</sup>H NOEs for residues 678–684 confirms the absence of a stable conformation for the N-terminal extremity.

Structure calculations were therefore performed for residues 685–759 using NMR restraints measured at 10 °C. Corresponding structural statistics are presented in Supplementary Table 1 online.





**Figure 3** X-ray structure of the PB2 C-terminal domain complexed with importin  $\alpha 5$ . **(a)** Ribbon diagram showing DPDE (red) bound to human importin  $\alpha 5$  (blue), comprising ten armadillo repeats. The C-terminal helix of the importin is unpacked and mediates domain-swap dimer formation in the crystal. The bipartite NLS at the C terminus of DPDE binds classically within the superhelical groove of importin  $\alpha 5$ . Basic residues Arg737, Lys738 and Arg739 from the minor site (purple) interact with the C-terminal armadillo repeats; Lys752, Arg753 and Arg755 from the major site (gold) interact with the N-terminal armadillo repeats. Lys736 does not interact with importin  $\alpha 5$  but makes intramolecular hydrogen bonds in DPDE, perhaps preventing further unfolding of the C terminus. Lys718 makes three hydrogen bonds with importin  $\alpha 5$ . **(b)** Comparison of the PB2 domain structure in complexed (red) and free solution state (cyan) demonstrates unfolding of residues 736–759 (purple) upon binding to importin  $\alpha 5$ . Residue Asp701, important in host specificity and virulence, forms a salt bridge with Arg753 of the major NLS motif and tethers the C terminus to the core of the domain in the unbound state. Residues Arg702 and Ser714 are also implicated in interspecies transmission. Note different orientations of the N-terminal helix of DPDE in the two structures.

The solution structure of the C-terminal domain of PB2 reveals a compact  $\alpha$ - $\beta$  domain with a hydrophobic core formed by a three-stranded antiparallel  $\beta$ -sheet, onto which a short amphipathic  $\alpha$ -helix packs (**Fig. 1b**). Except for the flexible loop 726–730 connecting the  $\beta 2$  and  $\beta 3$  strands and the last 20 residues at the C terminus, the structure was determined with a high resolution (r.m.s. deviation of  $0.42 \pm 0.12$  Å, calculated for the 81 backbone atoms of the secondary structure elements). Regions of lower structural resolution are correlated with enhanced dynamics (**Supplementary Fig. 2** online). The PB2 C-terminal domain seems to be a novel structural motif, as no similar protein with a Z-score  $> 2$  was found using Dali<sup>23</sup>. It is likely that the equivalent domains of PB2 from influenza B and C strains have similar structures because of their relatively high sequence conservation, particularly of buried hydrophobic residues (**Fig. 1c**).

### Importin $\alpha 5$ binding and live-cell localization studies

Efficient nuclear localization of PB2 is believed to depend on an internal region (residues 449–495) and a classical monopartite basic motif, 736-KRKR-739 (ref. 9), which is present in DPDE. Mutation<sup>9</sup> of the basic motif, deletion<sup>24</sup> of a C-terminal region containing it or deletion of the internal domain<sup>9</sup> all abrogate nuclear localization. In the NMR structure, 736-KRKR-739 immediately follows strand  $\beta 3$ , and all the basic side chains are surface exposed, except that of Lys736, which is partly buried. To investigate the nuclear localization properties of the C-terminal domain, we studied its interaction with human importin  $\alpha 5$  ( $\alpha$  importin), the classical eukaryotic nuclear import adaptor. We first showed that DPDE forms a stable complex with human importin  $\alpha 5$  *in vitro* (**Supplementary Fig. 3** online), suggesting that the domain may be capable of directing nuclear import. We then performed live-cell localization studies using a green fluorescent protein (GFP)-DPDE fusion, which was found almost exclusively in the nucleus ( $29 \pm 8$  times more nuclear than cytoplasmic), whereas a GFP control remained equilibrated between the cytoplasm and nucleus (**Fig. 2**). As previous results showed that the monopartite motif was itself insufficient for nuclear import<sup>9</sup>, we examined

sequence alignments for other residues that might be implicated (**Fig. 1c**). These show that although the monopartite NLS motif is conserved in influenza B, only two basic residues occur in influenza C (KRTA). However, a second basic motif, 752-KRIR-755, conserved in influenza B and C strains, suggested the presence of a classical bipartite NLS (KR $x_{12-15}$ KR $x$ R, where  $x$  is any residue and  $x_{12-15}$  is any sequence 12–15 residues long)<sup>25</sup>. To test this, the GFP-DPDE fusion was mutated in the putative bipartite NLS (**Fig. 2**). Mutating the first ('minor') part of the NLS (K738Q) mildly but significantly diminished nuclear import ( $19 \pm 5$  times more nuclear;  $P = 0.0004$  compared with wild-type), in agreement with studies on the whole PB2 protein<sup>9</sup>. Mutating the second ('major') part (K752N R755Q) greatly diminished nuclear import ( $5 \pm 1$  times more nuclear;  $P = 1.6 \times 10^{-8}$  compared with wild-type). Combining minor (K738Q) and major (K752N R755Q) NLS mutations abolished import to levels of GFP alone ( $2 \pm 1$  times more nuclear;  $P = 9.5 \times 10^{-8}$  compared with wild-type and  $P = 0.86$  compared with GFP). Similar data were obtained with a second mutation in the minor region (R737Q; data not shown) and with full-length PB2-GFP fusions (**Fig. 2**). Consistent results were also obtained when DPDE domains bearing the same NLS mutants were analyzed for importin  $\alpha 5$  complex formation *in vitro* (**Supplementary Fig. 3**). Thus, the C-terminal PB2 sequence 738-KR $x_{12}$ KRIR-755 functions as an efficient, classical, bipartite nuclear import signal, with a similar response to mutations as that of the retinoblastoma protein (KR $x_{11}$ KKLR) or phosphoprotein NIN2 (KR $x_{12}$ KKSK)<sup>25,26</sup>. Although this NLS is functional in the isolated domain as well as the full-length PB2 subunit, we cannot exclude the possibility that other sequences present in PB2 are responsible for the very low residual nuclear localization seen with the double NLS mutant of the whole protein.

### Cocrystal structure of DPDE with human importin $\alpha 5$

To provide further evidence for this revised mechanism of PB2 nuclear import, we cocrystallized DPDE with residues 66–512 of human importin  $\alpha 5$  (lacking the autoinhibitory N-terminal region) and determined the X-ray structure at 2.2-Å resolution (see Methods and **Supplementary Fig. 4** online). The structure of importin  $\alpha 5$  comprises ten armadillo repeats and is similar to that of yeast karyopherin  $\alpha$  (PDB 1BK5), with an r.m.s. deviation of 2 Å (for 400 C $\alpha$  positions of 422 residues aligned) and 56% sequence identity. This is the first structure of a human importin  $\alpha$  family member and the first of a complex of importin with a complete folded domain, rather than simply an NLS-containing peptide. DPDE interacts with

importin  $\alpha 5$  with its C-terminal residues beyond Lys736 in an extended conformation, permitting binding of the bipartite NLS to two distinct regions within the superhelical groove of importin  $\alpha 5$  (Fig. 3a). The mode of NLS binding is similar to that previously observed for bipartite NLS peptides binding yeast and mouse  $\alpha$  importins<sup>25,27</sup>. Strong interactions of the side chains of three basic residues in both the minor (737-RKR-739) and major (752-KRIR-755) sites, within discrete pockets of the importin, are supplemented by additional hydrogen bonds, notably from the NE1 atoms of importin tryptophans Trp149, Trp191, Trp234 and Trp360 to the DPDE NLS peptide main chain carbonyl oxygens (Supplementary Fig. 5 online). Residues 742–747 of the long (12-residue) linker are not visible in the electron density, nor are the C-terminal residues 758–759, presumably owing to flexibility. The side chain of Lys736 of DPDE makes strong intradomain hydrogen bonds with the main chains of Ser714 and Leu716 in both the free and bound states of the DPDE domain (Supplementary Fig. 5), consistent with it not being crucial for nuclear import<sup>9</sup>. The folded region of the DPDE domain packs against importin  $\alpha 5$  with a total buried surface area of about 916 Å<sup>2</sup>; in particular, Lys718 makes three hydrogen bonds with Gly284, Asn286 and Thr325 of importin  $\alpha 5$  (Fig. 3a).

## DISCUSSION

The structure of the core of the DPDE domain (residues 693–736) observed in solution by NMR is very similar to the structure in the crystalline state in complex with importin  $\alpha 5$  (r.m.s. deviation of C $\alpha$  positions of 1.87 Å for 41 residues aligned), although the X-ray structure shows more secondary structure (Fig. 3b). The comparison reveals how the region 736–759 unfolds to allow binding of the bipartite NLS to the import factor (Fig. 3b). In the unbound NMR structure, the carboxyl group of Asp701 forms a salt bridge with Arg753, an important basic residue of the major site of the bipartite NLS motif, thus tethering the C-terminal extremity to the core of the domain. Binding to importin  $\alpha 5$  requires that this interaction be broken, as shown by the structure of the complex (Fig. 3a). A number of recent studies have shown that PB2 residue 701 is important for interspecies transmission. The D701N substitution has been found to contribute considerably to the acquisition of pathogenicity to mice of both H5N1 (ref. 15) and H7N7 (ref. 17) avian strains and has been observed in both fatal and nonfatal cases of H5N1 infecting humans in Vietnam<sup>28</sup>. We note that the mutation D701N would disrupt the salt bridge with Arg753 of the NLS, potentially altering the energetics of importin binding and, in turn, the efficiency of trimeric polymerase assembly. The adjacent residue 702 is also an important determinant of host specificity and, with only rare exceptions, is an arginine in human isolates, including the 1918 H1N1 pandemic virus, and a lysine in avian strains<sup>16</sup>. This residue is completely solvent exposed in both structures, and its function is not yet known. Mutation of a third residue in this domain, Ser714, in common with position 701, affects levels of polymerase activity of reconstituted ribonucleoprotein complexes in mammalian 293T cells; mutants with Asn701 and Arg714 have the highest activity<sup>17</sup>. The fact that these three cross-species-transfer residues<sup>15–17</sup> are exposed to solvent suggests a possible role of this domain in intermolecular interactions with other viral or host proteins; the interaction with importin  $\alpha 5$  and the regulation of nuclear transport of PB2 both could be important in this respect.

## METHODS

**Identification of soluble constructs.** DNA encoding PB2 from strain A/Victoria/3/75(H3N2)<sup>29</sup> fused to a C-terminal biotin acceptor peptide was

truncated from the 5' end with exonuclease III. An *Escherichia coli* BL21 strain was transformed with a pool of 24,036 plasmids, and clones were screened for protein expression by robotic arraying onto nitrocellulose membranes over agar. Colonies were lysed *in situ* and streptavidin conjugates used to detect biotinylated proteins (Supplementary Methods online).

**NMR spectroscopy and structure calculation.** Heteronuclear NMR data for assignment and extraction of structural restraints were collected at 10 °C on a 0.67 mM uniformly <sup>13</sup>C,<sup>15</sup>N-labeled DPDE sample in 50 mM Tris buffer (90% (v/v) H<sub>2</sub>O, 10% (v/v) D<sub>2</sub>O) at pH 7.5 with 200 mM NaCl, 1 mM EDTA and 0.01% (w/v) NaN<sub>3</sub>. Structure calculations including residues 685–759 were performed using CNS<sup>30</sup> and refined in explicit water (Supplementary Methods).

**Crystallography.** Cocrystals of DPDE and human importin  $\alpha 5$  (NSMB Protein AAP35605; residues 66–512) were grown by vapor diffusion with 2-methyl-2,4-pentanediol (MPD) as precipitant. Data to 2.2-Å resolution were measured on beamline ID23-1 at the European Synchrotron Radiation Facility, using X-rays of wavelength 1.072 Å and a temperature of 100 K, and integrated with XDS<sup>31</sup>. Two complexes were located in the asymmetric unit by molecular replacement using PDB 1Q1S (mouse importin  $\alpha$  with NLS peptide) as search model. The map was greatly improved using RESOLVE<sup>32</sup> with two-fold noncrystallographic symmetry averaging, allowing the complete structure (residues 84–508 of importin  $\alpha 5$ , and 686–741 and 748–757 of DPDE) to be constructed and refined using REFMAC<sup>33</sup>. The final *R*-factor was 20.6% (*R*<sub>free</sub> = 24.7%). According to Molprobity (http://molprobity.biochem.duke.edu/), 98.4% and 99.6% of the total of 966 residues are within the favored or allowed regions of the Ramachandran plot, respectively, with four outliers (see Table 1 and Supplementary Methods).

**Nuclear localization.** Mouse Swiss National Institutes of Health embryonic fibroblasts (NIH 3T3) were transfected with plasmids encoding the PB2

**Table 1 Crystallographic data collection and refinement statistics**

	Human importin $\alpha 5$ with influenza A PB2 C-terminal domain
<b>Data collection</b>	
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	94.97, 100.55, 151.84
Resolution (Å)	50–2.2 (2.2–2.3)
<i>R</i> <sub>merge</sub>	0.093 (0.694)
<i>I</i> / $\sigma$ <i>I</i>	9.40 (2.41)
Completeness (%)	99.5 (99.9)
Redundancy	4.00 (4.06)
<b>Refinement</b>	
Resolution (Å)	30–2.2 (2.257–2.200)
No. reflections (total / free)	74,089 / 2,969
<i>R</i> <sub>work</sub>	0.206 (0.284)
<i>R</i> <sub>free</sub>	0.247 (0.340)
No. atoms	7,863
Protein	7,597 <sup>a</sup>
Water	266
<i>B</i> -factors (Å <sup>2</sup> )	41.15
Protein (chain A, chain B)	41.9, 48.3
DPDE (chain D, chain E)	62.7, 51.4
Water	33.2
R.m.s. deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.273

Values in parentheses are for highest-resolution shell. <sup>a</sup>Two molecules in the asymmetric unit.

domain or full-length PB2 fused C-terminally to enhanced GFP. Cells were imaged by quantitative confocal microscopy (**Supplementary Methods**).

**Accession codes.** BioMagResBank (<http://www.bmrb.wisc.edu>): The  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts for the C-terminal domain of PB2 with Pro679 in the *trans* (major) and *cis* (minor) conformations have been deposited with accession number 7056 and linked to the atomic coordinates of the ten structures of lowest energy in the Protein Data Bank (PDB 2GMO and RCSB 037302). Protein Data Bank: coordinates and structure factors of the PB2 domain-importin  $\alpha 5$  complex have been deposited with accession codes 2JDQ and R2JDQSF, respectively.

*Note: Supplementary information is available on the Nature Structural & Molecular Biology website.*

#### ACKNOWLEDGMENTS

We thank J. Ortin (Centro Nacional de Biotecnología, CSIC, Madrid) for the *pb2* gene, A. Favier (Institut de Biologie Structurale, Grenoble) for NMR scripts and the EMBL Centre for Molecular and Cellular Imaging for suggestions. Screening for crystals was done by the Partnership for Structural Biology high-throughput crystallization facility. We thank the European Synchrotron Radiation Facility and EMBL Joint Structural Biology group for assistance with the synchrotron beamtime and T. Crepin (EMBL, Grenoble) for help with data collection. Partial funding was provided by the European Commission Framework 5 Integrated Project 'Structural Proteomics in Europe' (SPINE, contract QL-G-CT-2002-00988).

#### AUTHOR CONTRIBUTIONS

D.J.H. conceived the ESPRIT method. D.J.H., F.T. and P.J.M. implemented ESPRIT. D.G. purified wild-type and mutant DPDE for *in vitro* binding studies and crystallization and made double-labeled protein for NMR. C.M.B., J.B. and J.-P.S. performed the NMR measurements and structural analysis. S.B. purified importin  $\alpha 5$  under the supervision of F.B. and cocrystallized it with DPDE. R.W.H.R. and S.C. initiated the influenza polymerase project, and S.C. determined the crystallographic structure. F.T. and N.D. performed nuclear import assays with instrumentation and methodology established by J.E. S.C. and D.J.H. compiled the text, with contributions from all authors.

#### COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Structural & Molecular Biology* website for details).

Published online at <http://www.nature.com/nsmb/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

- Noah, D.L. & Krug, R.M. Influenza virus virulence and its molecular determinants. *Adv. Virus Res.* **65**, 121–145 (2005).
- 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).
- Li, M.L., Ramirez, B.C. & Krug, R.M. RNA-dependent activation of primer RNA production by influenza virus polymerase: different regions of the same protein subunit constitute the two required RNA-binding sites. *EMBO J.* **17**, 5844–5852 (1998).
- Gonzalez, S. & Ortin, J. Characterization of influenza virus PB1 protein binding to viral RNA: two separate regions of the protein contribute to the interaction domain. *J. Virol.* **73**, 631–637 (1999).
- 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).
- Honda, A., Mizumoto, K. & Ishihama, A. Two separate sequences of PB2 subunit constitute the RNA cap-binding site of influenza virus RNA polymerase. *Genes Cells* **4**, 475–485 (1999).
- 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).
- Nath, S.T. & Nayak, D.P. Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). *Mol. Cell. Biol.* **10**, 4139–4145 (1990).
- Mukaigawa, J. & Nayak, D.P. Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2. *J. Virol.* **65**, 245–253 (1991).
- Nieto, A. *et al.* Nuclear transport of influenza virus polymerase PA protein. *Virus Res.* **24**, 65–75 (1992).
- 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).
- Deng, T. *et al.* Role of ran binding protein 5 in nuclear import and assembly of the influenza virus RNA polymerase complex. *J. Virol.* **80**, 11911–11919 (2006).
- Naito, T., Momose, F., Kawaguchi, A. & Nagata, K. Involvement of hsp90 in assembly and nuclear import of influenza virus RNA polymerase subunits. *J. Virol.* **81**, 1339–1349 (2007).
- Russell, C.J. & Webster, R.G. The genesis of a pandemic influenza virus. *Cell* **123**, 368–371 (2005).
- Li, Z. *et al.* Molecular basis of replication of duck H5N1 influenza viruses in a mammalian mouse model. *J. Virol.* **79**, 12058–12064 (2005).
- Taubenberger, J.K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
- Gabriel, G. *et al.* The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl. Acad. Sci. USA* **102**, 18590–18595 (2005).
- Salomon, R. *et al.* The polymerase complex genes contribute to the high virulence of the human H5N1 influenza virus isolate A/Vietnam/1203/04. *J. Exp. Med.* **203**, 689–697 (2006).
- Martin-Benito, J. *et al.* Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO Rep.* **2**, 313–317 (2001).
- 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).
- Ostermeier, M. & Lutz, S. The creation of ITCHY hybrid protein libraries. *Methods Mol. Biol.* **231**, 129–141 (2003).
- Beckett, D., Kovaleva, E. & Schatz, P.J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921–929 (1999).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- Peralles, B., de la Luna, S., Palacios, I. & Ortin, J. Mutational analysis identifies functional domains in the influenza A virus PB2 polymerase subunit. *J. Virol.* **70**, 1678–1686 (1996).
- Fontes, M.R., Teh, T., Jans, D., Brinkworth, R.I. & Kobe, B. Structural basis for the specificity of bipartite nuclear localization sequence binding by importin- $\alpha$ . *J. Biol. Chem.* **278**, 27981–27987 (2003).
- Zacksenhaus, E., Bremner, R., Phillips, R.A. & Gallie, B.L. A bipartite nuclear localization signal in the retinoblastoma gene product and its importance for biological activity. *Mol. Cell. Biol.* **13**, 4588–4599 (1993).
- Conti, E. & Kuriyan, J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin  $\alpha$ . *Structure* **8**, 329–338 (2000).
- de Jong, M.D. *et al.* Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nat. Med.* **12**, 1203–1207 (2006).
- 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).
- Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
- Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
- Terwilliger, T.C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).
- Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).