Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Crystallization and X-ray diffraction analysis of the RNA primer/promoter-binding domain of influenza A virus RNA-dependent RNA polymerase PB2

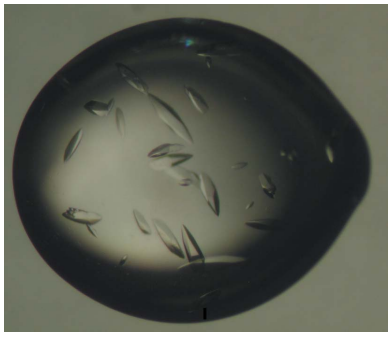
Takashi Kuzuhara,[a]*‡ Daisuke Kise,[a]‡ Hiroko Yoshida,[a]‡ Takahiro Horita,[a] Yoshimi Murazaki,[a] Hiroko Utsunomiya[b] and Hideaki Tsuge[b]*

[a]Laboratory of Biochemistry, Faculty of Pharmaceutical Sciences, Tokushima Bunri University, Yamashiro-cho, Tokushima 770-8514, Japan, and [b]Institute for Health Sciences, Tokushima Bunri University, Yamashiro-cho, Tokushima 770-8514, Japan

‡ The first three authors contributed equally.

Correspondence e-mail: kuzuhara@ph.bunri-u.ac.jp, tsuge@tokushima.bunri-u.ac.jp

Received 12 November 2008
Accepted 17 December 2008



The C-terminal domain protein (amino-acid residues 535–759) of the PB2 subunit of the RNA-dependent RNA polymerase from the highly pathogenic influenza A virus was expressed as a soluble protein in *Escherichia coli* and crystallized using sodium formate as a precipitant. Data sets were collected from crystals of native and selenomethionine-substituted protein on the KEK NW12 beamline at the Photon Factory and the crystals diffracted to a maximum resolution of 2.44 Å for the SeMet-derivative crystal. The native crystals were found to belong to space group $P3_221$, with unit-cell parameters $a = b = 52.5$, $c = 156.3$ Å. The Matthews value (V_M) was 2.7 Å^3 Da^{-1}, assuming the presence of one molecule in the asymmetric unit. The SeMet-derivative crystals were found to belong to the same space group, with unit-cell parameters $a = b = 52.6$, $c = 156.4$ Å. Attempts are being made to solve the structure by multi-wavelength anomalous dispersion phasing.

1. Introduction

Pandemic influenza A infections have resulted in ten million deaths worldwide (Taubenberger *et al.*, 2005). The transcription and replication of the influenza A virus requires the activity of its RNA-dependent RNA polymerase (Morse, 2007; Horimoto & Kawaoka, 2005). The influenza A virus digests host mRNA, uses a 5′-RNA fragment of 10–15 nucleotides as a primer for transcription initiation (Shi *et al.*, 1995) and possesses an RNA primer/promoter-binding activity within the C-terminal region of its RNA polymerase PB2 domain (Honda & Ishihama, 1997). Low-resolution structures have previously been determined of the PA, PB1 and PB2 subunits of influenza A virus RNA polymerase using electron microscopy (Torreira *et al.*, 2007). In addition, the structures of the PB2 nuclear localization signal domain (amino-acid residues 687–759; Tarendeau *et al.*, 2007), of the PB2 cap (7-methylguanosine triphosphate) binding domain (amino-acid residues 318–483; Guilligay *et al.*, 2008) and of the PA fragment with the PB1 fragment have also been reported (He *et al.*, 2008; Obayashi *et al.*, 2008), but no RNA primer/promoter-binding domain of this viral polymerase has yet been identified. It is important to determine the structure of this domain including Lys627 (the 627-region) as it may provide valuable insight for developing novel drugs against the influenza A virus. Recently, the structure of another 627-region of PB2 has been solved by Tarendeau *et al.* (2008).

In order to perform X-ray crystallography of the RNA polymerase of influenza A virus, it was necessary to obtain soluble quantities of the three core subunits of the enzyme. Each individual subunit was therefore divided into three portions and the C-terminal portion of the PB2 domain (PB2 3/3; amino-acid residues 535–759) was constructed. The protein was expressed in *Escherichia coli* as a His_6-tagged recombinant protein which was subsequently purified using Ni^{2+}–agarose. The PB2 3/3 protein was then further purified to near-homogeneity (98%) by cation-exchange chromatography.

The recombinant PB2 3/3 protein was subjected to crystallization. Native crystals were found to belong to space group $P3_221$, with unit-cell parameters $a = b = 52.5$, $c = 156.3$ Å. The Matthews value (V_M) was 2.7 Å^3 Da^{-1}, assuming the presence of one molecule in the asymmetric unit. Selenomethionine-derivative crystals were obtained in the same space group, with unit-cell parameters $a = b = 52.6$, $c = 156.4$ Å. The crystal structure of PB2 3/3 is currently under analysis using multiwavelength anomalous dispersion (MAD) phasing (Hendrickson & Ogata, 1997).

2. Methods and results

2.1. Plasmid construction, expression and purification of the influenza A RNA polymerase PB2 RNA primer/promoter-binding domain

Influenza (A/PR/8/34) RNA polymerase PB2 plasmid, pBMSA-PB2, as deposited by Dr Susumu Nakada, was provided by the DNA Bank, RIKEN BioResource Center (Tsukuba, Japan) with the support of the National Bio-Resources Project of the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT). cDNA fragments of the RNA primer/promoter-binding domain (PB2 3/3: amino-acid residues 535–759) of the influenza A RNA-dependent RNA polymerase PB2 subunit were amplified by PCR. The following primers were used to amplify the fragment for subcloning into a protein-expression vector: PB2 3/3 Met, GCCGTTCATATGATGT-GGGAGATTAATGGT; PB2 3/3 stop, GCCGTTGGATCCTTAAT-TGATGGCCATCCGAAT. The amplified DNA fragment was subcloned into the pET28a(+) plasmid (Novagen, Madison, Wisconsin, USA) at the *Nde*I and *Bam*HI restriction sites. The construct was introduced into BL21-CodonPlus (Stratagene, La Jolla, California, USA) *E. coli* cells. The induction of His_6-tagged recombinant protein expression from these constructs was achieved by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) and this was followed by purification using Ni^{2+}–agarose. To prepare the cell lysate, buffer consisting of 20 m*M* Tris–HCl pH 7.9, 10% glycerol, 500 m*M* NaCl, 0.1% NP-40, 0.35% 2-mercaptoethanol, 0.1 m*M* phenylmethylsulfonyl fluoride (PMSF), 10 µg ml^{-1} leupeptin, 10 µg ml^{-1} pepstatin was used. For Ni^{2+}–agarose chromatography, elution buffer consisting of 20 m*M* Tris–HCl pH 7.9, 10% glycerol, 250 m*M* NaCl, 0.1% NP-40, 200 m*M* imidazole, 0.35% 2-mercaptoethanol was used. For further purification, the His-tagged proteins were cleaved by thrombin and purified using a HiTrap CM FF column (GE Healthcare, Buckinghamshire, UK) with the Akta Prime Plus system (GE Healthcare, Buckinghamshire, UK). For CM column chromatography, buffer *A* (20 m*M* Tris–HCl pH 6.5, 10% glycerol, 50 m*M* NaCl) and buffer *B* (20 m*M* Tris–HCl pH 6.5, 10% glycerol, 500 m*M* NaCl) were used.

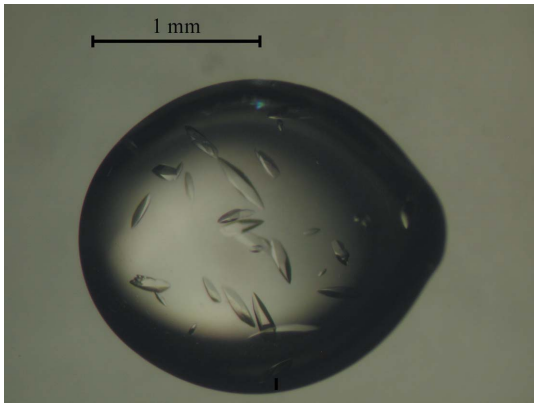


Figure 1
Photograph of a native crystal of the C-terminal domain of influenza A virus RNA-dependent RNA polymerase PB2. The dimensions of the largest native crystal were approximately 0.3 × 0.1 × 0.1 mm.

Table 1
Data-collection and processing statistics for native and selenomethionine-substituted PB2.

(*a*) Statistical analysis of the reflection data. Values in parentheses are for the highest resolution shells.

	Native	SeMet	SeMet	SeMet
Beamline	KEK AR NW12			
Wavelength (Å)	1.00000	0.97920	0.97936	0.96413
Resolution (Å)	2.70	2.44	2.44	2.44
Total No. of observations	41868	96928	96000	96049
No. of unique reflections	7319	18021	17959	17976
Redundancy	5.7	5.4	5.4	5.4
Completeness (%)	98.8 (100)	99.6 (96.9)	99.2 (93.2)	99.0 (90.8)
R_{merge}† (%)	3.9 (6.3)	6.3 (27.0)	6.4 (28.2)	6.4 (28.7)
$\langle I/\sigma(I)\rangle$	45.8	20.9	20.2	19.5

(*b*) Statistics for phase calculation.

Resolution (Å)	34.3–2.7
Figure of merit	0.70
Score	16.41

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

2.2. Crystallization

The purified samples were concentrated to 12.7 mg ml^{-1} (native protein) or 4.3 mg ml^{-1} (selenomethionine-labelled protein) by ultrafiltration. The optimal conditions for crystallization were initially determined using Crystal Screens 1 and 2 (Hampton Research, Aliso Viejo, California, USA). The searches were performed using the hanging-drop vapour-diffusion method at 277 K. The protein sample was in 10 m*M* Tris–HCl pH 7.5, 10% glycerol, 50 m*M* NaCl. 1 µl protein solution and 1 µl precipitant solution were mixed in each condition and used for crystallization. Native crystals were obtained using reservoir solutions containing 0.1 *M* sodium cacodylate–HCl pH 6.2, 2.2 *M* sodium formate (Fig. 1) and 0.1 *M* sodium citrate tribasic dihydrate pH 5.2, 1.8 *M* sodium formate. Crystals of selenomethionine-labelled PB2 3/3 protein were obtained using a reservoir solution containing 0.1 *M* sodium cacodylate–HCl pH 6.0, 3 *M* sodium formate. For the native crystal, we collected complete data to 2.7 Å resolution, although the crystal diffracted beyond 2.7 Å resolution. The SeMet-derivativized crystals were grown at a higher precipitant concentration. However, there was no great difference in diffraction quality between the native and SeMet-derivatized crystals when the same-sized crystals were used.

2.3. Preliminary X-ray diffraction analysis

Diffraction data sets for native and SeMet crystals were collected at 100 K using a mixture of Paratone-N (Hampton Research, Aliso Viejo, California, USA) and paraffin oil as a cryoprotectant and using synchrotron facility AR NW12 at the Photon Factory, Tsukuba, Japan. The image data were processed using *HKL*-2000 (Otwinowski & Minor, 1997) and *SCALA* (Collaborative Computational Project, Number 4, 1994). The data and statistics are summarized in Table 1. The initial phase was obtained by MAD with three-wavelength SeMet data sets using *SOLVE* and *RESOLVE* (Terwilliger, 2003). The electron-density map showed that the structure of the PB2 3/3 protein (amino-acid residues 535–759) consists of two subdomains. The second domain is the nuclear localization signal domain

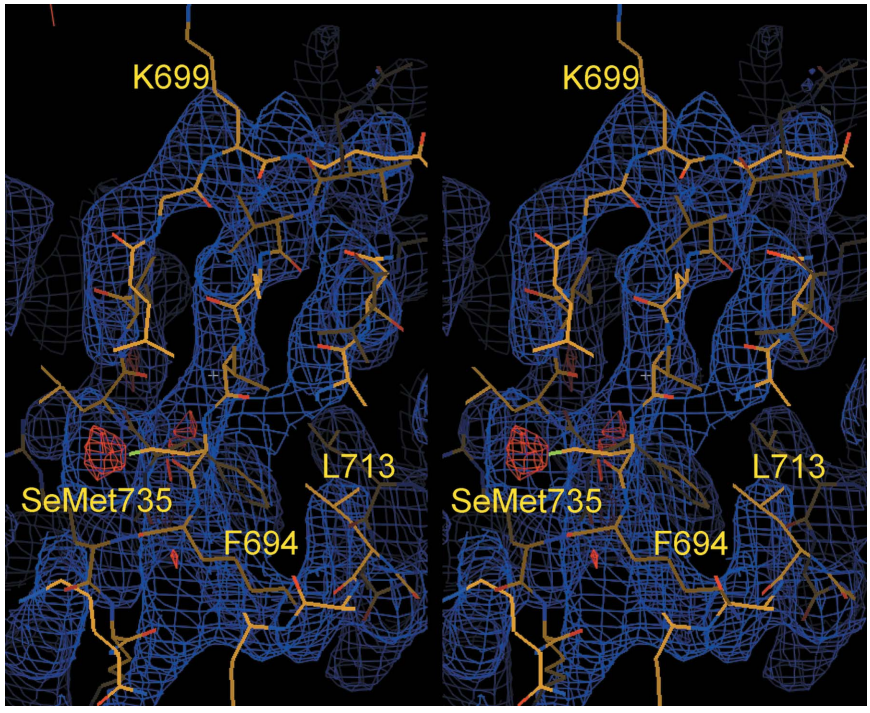


Figure 2
The structure of the nuclear localization signal domain (from PDB entry 2jdq; Tarendeau *et al.*, 2007) was fitted into the electron-density map (*RESOLVE* experimental map). The colours of the contour levels are blue and red for 1σ and 5σ, respectively.

described by Tarendeau *et al.* (2007). The structure of this domain was fitted into the electron-density map (Fig. 2). The first domain includes Lys627. This residue is associated with high pathogenicity and is a host-range determinant in human-to-human infection. Currently, both model building and crystallographic refinement are being carried out against a native data set.

Recently, the structure of another 627-domain was reported (Tarendeau *et al.*, 2008). There are eight amino-acid differences between the two structures outside the N-terminal regions: A613T, F636L, A661T, V667I, T676I, G682S, A684S and K702R. A661T, G682S, A684S and K702R are host-range mutations from A2A (avian to avian) to H2H (human to human). A comparable study will reveal whether these host-range mutations affect the tertiary structure. The fact that two structures will be solved in different crystal forms may shed some light on artifacts arising from crystal packing.

This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan. We thank N. Igarashi, N. Matsugaki, Y. Yamada and S. Wakatsuki for data collection at PF AR NW12.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R. W., Ortin, J., Hart, D. J. & Cusack, S. (2008). *Nature Struct. Mol. Biol.* **15**, 500–506.
He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., Ge, R., Rao, Z. & Liu, Y. (2008). *Nature (London)*, **454**, 1123–1126.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Honda, A. & Ishihama, A. (1997). *Biol. Chem.* **378**, 483–488.
Horimoto, T. & Kawaoka, Y. (2005). *Nature Rev. Microbiol.* **3**, 591–600.
Morse, S. S. (2007). *Nature Med.* **13**, 681–684.
Obayashi, E., Yoshida, H., Kawai, F., Shibayama, N., Kawaguchi, A., Nagata, K., Tame, J. R. & Park, S. Y. (2008). *Nature (London)*, **454**, 1127–1131.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Shi, L., Summers, D. F., Peng, Q. & Galarz, J. M. (1995). *Virology*, **208**, 38–47.
Tarendeau, F., Boudet, J., Guilligay, D., Mas, P. J., Bougault, C. M., Boulo, S., Baudin, F., Ruigrok, R. W., Daigle, N., Ellenberg, J., Cusack, S., Simorre, J. P. & Hart, D. J. (2007). *Nature Struct. Mol. Biol.* **14**, 229–233.
Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R. W., Cusack, S. & Hart, D. J. (2008). *PLoS Pathog.* **4**, e1000136.
Taubenberger, J. K., Reid, A. H., Lourens, R. M., Wang, R., Jin, G. & Fanning, T. G. (2005). *Nature (London)*, **437**, 889–893.
Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.
Torreira, E., Schoehn, G., Fernández, Y., Jorba, N., Ruigrok, R. W., Cusack, S., Ortín, J. & Llorca, O. (2007). *Nucleic Acids Res.* **35**, 3774–3783.