POLYPEPTIDE FRAGMENTS COMPRISING ENDONUCLEASE ACTIVITY AND THEIR USE

TECHNICAL FIELD OF INVENTION

The present invention relates to polypeptide fragments comprising an amino-terminal fragment of the PA subunit of a viral RNA-dependent RNA polymerase or variants thereof possessing endonuclease activity, wherein said PA subunit is from a virus belonging to the Orthomyxoviridae family. This invention also relates to (i) crystals of the polypeptide fragments which are suitable for structure determination of said polypeptide fragments using X-ray crystallography and (ii) computational methods using the structural coordinates of said polypeptide to screen for and design compounds that modulate, preferably inhibit the endonucleolytically active site within the polypeptide fragment. In addition, this invention relates to methods identifying compounds that bind to the PA polypeptide fragments possessing endonuclease activity and preferably inhibit said endonucleolytic activity, preferably in a high throughput setting. This invention also relates to compounds and pharmaceutical compositions comprising the identified compounds for the treatment of disease conditions due to viral infections caused by viruses of the Orthomyxoviridae family.

BACKGROUND OF THE INVENTION

Influenza is responsible for much morbidity and mortality in the world and is considered by many as belonging to the most significant viral threats to humans. Annual Influenza epidemics swipe the globe and occasional new virulent strains cause pandemics of great destructive power. At present the primary means of controlling Influenza virus epidemics is vaccination. However, mutant Influenza viruses are rapidly generated which escape the effects of vaccination. In the light of the fact that it takes approximately 6 months to generate a new Influenza vaccine, alternative therapeutic means, i.e., antiviral medication, are required especially as the first line of defense against a rapidly spreading pandemic.

An excellent starting point for the development of antiviral medication is structural data of essential viral proteins. Thus, the crystal structure determination of the Influenza virus surface antigen neuraminidase (von Itzstein et al., 1993, Nature 363:418-423) led directly to the development of neuraminidase inhibitors with anti-viral activity preventing the release of virus from the cells, however, not the virus production. These and their derivatives have subsequently developed into the anti-Influenza drugs, zanamivir (Glaxo) and oseltamivir (Roche), which are currently being stockpiled by many countries as a first line of defense

against an eventual pandemic. However, these medicaments provide only a reduction in the duration of the clinical disease. Alternatively, other anti-Influenza compounds such as amantadine and rimantadine target an ion channel protein, i.e., the M2 protein, in the viral membrane interfering with the uncoating of the virus inside the cell. However, they have not been extensively used due to their side effects and the rapid development of resistant virus mutants (Magden et al., 2005, Appl. Microbiol. Biotechnol. 66:612-621). In addition, more unspecific viral drugs, such as ribavirin, have been shown to work for treatment of Influenza infections (Eriksson et al., 1977, Antimicrob. Agents Chemother. 11:946-951). However, ribavirin is only approved in a few countries, probably due to severe side effects (Furuta et al., 2005, Antimicrob. Agents Chemother. 49:981-986). Clearly, new antiviral compounds are needed, preferably directed against different targets.

Influenza virus A, B, C and Isavirus as well as Thogotovirus belong to the family of Orthomyxoviridae which, as well as the family of the Bunyaviridae, including the Hantavirus, Nairovirus, Orthobunyavirus, Phlebovirus, and Tospovirus, are negative stranded RNA viruses. Their genome is segmented and comes in ribonucleoprotein particles that include the RNA dependent RNA polymerase which carries out (i) the initial copying of the single-stranded virion RNA (vRNA) into viral mRNAs and (ii) the vRNA replication. For the generation of viral mRNA the polymerase makes use of the so called "cap-snatching" mechanism (Plotch et al., 1981, Cell 23:847-858; Kukkonen et al., 2005, Arch. Virol. 150:533-556; Leahy et al., 1997, J. Virol. 71:8347-8351; Noah and Krug, 2005, Adv. Virus Res. 65:121-145). The polymerase is composed of three subunits: PB1 (polymerase basic protein), PB2, and PA. For the cap-snatching mechanism, the viral polymerase binds via its PB2 subunit to the 5' RNA cap of cellular mRNA molecules which are cleaved at nucleotide 10 to 13 by the endonucleolytic activity of the polymerase. The capped RNA fragments serve as primers for the synthesis of viral mRNAs by the nucleotidyl-transferase center in the PB1 subunit (Li et al., 2001, EMBO J. 20:2078-2086). Finally, the viral mRNAs are 3'-end poly-adenylated by stuttering of the polymerase at an oligo-U motif at the 5'-end of the template. Recent studies have precisely defined the structural domain of PB2 responsible for cap-binding (Fechter et al., 2003, J. Biol. Chem. 278:20381-20388; Guilligay et al., 2008 Nat. Struct. Mol. Biol. 15:500-506). The endonucleolytic activity of the polymerase has hitherto been thought to reside in the PB1 subunit (Li et al, *supra*).

The polymerase complex seems to be an appropriate antiviral drug target since it is essential for synthesis of viral mRNA and viral replication and contains several functional active sites likely to be significantly different from those found in host cell proteins (Magden

et al., *supra*). Thus, for example, there have been attempts to interfere with the assembly of polymerase subunits by a 25-amino-acid peptide resembling the PA-binding domain within PB1 (Ghanem et al., 2007, J. Virol. 81:7801-7804). Moreover, there have been attempts to interfere with viral transcription by nucleoside analogs, such as 2'-deoxy-2'-fluoroguanosine (Tisdale et al., 1995, Antimicrob. Agents Chemother. 39:2454-2458) and it has been shown that T-705, a substituted pyrazine compound may function as a specific inhibitor of Influenza virus RNA polymerase (Furuta et al., *supra*). Furthermore, the endonuclease activity of the polymerase has been targeted and a series of 4-substituted 2,4-dioxobutanoic acid compounds has been identified as selective inhibitors of this activity in Influenza viruses (Tomassini et al., 1994, Antimicrob. Agents Chemother. 38:2827-2837). In addition, flutimide, a substituted 2,6-diketopiperazine, identified in extracts of *Delitschia confertaspora*, a fungal species, has been shown to inhibit the endonuclease of Influenza virus (Tomassini et al., 1996, Antimicrob. Agents Chemother. 40:1189-1193). However, the inhibitory action of compounds on the endonucleolytic activity of the viral polymerase was hitherto only studied in the context of the entire trimeric complex of the polymerase.

The PA subunit of the polymerase is functionally the least well-characterised, although it has been implicated in both cap-binding and endonuclease activity, vRNA replication, and a controversial protease activity. PA (716 residues in influenza A) is separable by trypsination at residue 213. The recently determined crystal structure of the C-terminal two-thirds of PA bound to a PB1 N-terminal peptide provided the first structural insight into both a large part of the PA subunit, whose function, however, still remains unclear, and the exact nature of one of the critical inter-subunit interactions (He et al., 2008, Nature 454:1123-1126; Obayashi et al., 2008, Nature 454:1127-1131). Systematic mutation of conserved residues in the PA amino-terminal domain have identified residues important for protein stability, promoter binding, cap-binding and endonuclease activity of the polymerase complex (Hara et al., 2006, J. Virol. 80:7789-7798). The enzymology of the endonuclease within the context of intact viral ribonucleoprotein particles (RNPs) has been extensively studied.

However, hitherto it was not possible to study the endonuclease activity of the PA subunit in the context of a polypeptide fragment possessing the endonucleolytic activity, since it was not known which domain is responsible for said activity. The present inventors surprisingly found that, contrary to the general opinion in the field, the endonucleolytic activity resides exclusively within the amino-terminal region of the PA subunit. The inventors have achieved to structurally characterize said domain by X-ray crystallography and identified the endonucleolytic active center within the amino-terminal PA polypeptide fragment.

Thus, the present invention provides the unique opportunity to study the endonucleolytic activity of the viral polymerase in the context of a polypeptide fragment which will considerably simplify the development of new anti-viral compounds targeting the endonuclease activity of the viral polymerase as well as the optimization of previously identified compounds. The surprising achievement of the present inventors to recombinantly produce PA polypeptide fragments possessing the endonucleolytic activity of the viral polymerase allows for performing *in vitro* high-throughput screening for inhibitors of a functional site on the viral polymerase using easily obtainable material from a straightforward expression system. Furthermore, the structural data of the endonucleolytic PA polypeptide fragment as well as of the enzymatically active center therein allows for directed design of inhibitors and *in silico* screening for potentially therapeutic compounds.

It is an object of the present invention to provide (i) high resolution structural data of the endonucleolytic amino-terminal domain of the viral polymerase PA subunit by X-ray crystallography, (ii) computational as well as *in vitro* methods, preferably in a high-throughput setting, for identifying compounds that can modulate, preferably inhibit, the endonuclease activity of the viral polymerase, preferably by blocking the endonucleolytic active site within the PA subunit, and (iii) pharmacological compositions comprising such compounds for the treatment of infectious diseases caused by viruses using the cap snatching mechanism for synthesis of viral mRNA.

SUMMARY OF THE INVENTION

In a first aspect, the present invention relates to a polypeptide fragment comprising an amino-terminal fragment of the PA subunit of a viral RNA-dependent RNA polymerase possessing endonuclease activity, wherein said PA subunit is from a virus belonging to the Orthomyxoviridae family.

In a further aspect, the present invention relates to an isolated polynucleotide encoding an isolated polypeptide fragment according to the present invention.

In a further aspect, the present invention relates to recombinant vector comprising the isolated polynucleotide according to the present invention.

In a further aspect, the present invention relates to a recombinant host cell comprising the isolated polynucleotide according to the invention or the recombinant vector according to the present invention.

In a further aspect, the present invention relates to a method for identifying compounds which modulate the endonuclease activity of the PA subunit of a viral RNA-dependent RNA polymerase from the Orthomyxoviridae family, comprising the steps of

(a) constructing a computer model of the active site defined by the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 18;

(b) selecting a potential modulating compound by a method selected from the group consisting of:

(i) assembling molecular fragments into said compound,

(ii) selecting a compound from a small molecule database, and

(iii) *de novo* ligand design of said compound;

(c) employing computational means to perform a fitting program operation between computer models of the said compound and the said active site in order to provide an energy-minimized configuration of the said compound in the active site; and

(d) evaluating the results of said fitting operation to quantify the association between the said compound and the active site model, whereby evaluating the ability of said compound to associate with the said active site.

In a further aspect, the present invention relates to a compound identifiable by the method according to the present invention, wherein said compound is able to modulate, preferably inhibit the endonuclease activity of the PA subunit or variant thereof.

In a further aspect, the present invention relates to a method for identifying compounds which modulate the endonuclease activity of the PA subunit or polypeptide variants thereof, comprising the steps of (i) contacting the polypeptide fragment according to the invention or the recombinant host cell according to the invention with a test compound and (ii) analyzing the ability of said test compound to modulate the endonuclease activity of said PA subunit polypeptide fragment.

In a further aspect, the present invention relates to a pharmaceutical composition producible according to the *in vitro* method of the present invention.

In a further aspect, the present invention relates to a compound identifiable by the *in vitro* method according to the invention, wherein said compound is able to modulate, preferably inhibit the endonuclease activity of the PA subunit or variant thereof

In a further aspect, the present invention relates to an antibody directed against the active site of the PA subunit or variant thereof

In a futher aspect, the present invention relates to the use of a compound according to the present invention, a pharmaceutical composition according to the present invention, or an

antibody according to the present invention for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with viruses of the Orthomyxoviridae family.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Assay of thermal stability of the PA-Nter structure using Thermofluor. The thermal shift assay was performed with different metal ions. For clarity, only the results obtained in absence of metal ion (full black line) or in presence of 1 mM $MnCl_2$ (dashed line) are shown. Arrows indicate the apparent melting temperature Tm.

Figure 2: Effects of different metal ions on thermal stability of PA-Nter. Summary of the different melting points (Tm) extracted from the thermal shift assay at pH 8.0 with different metal ions. The effect of $CoCl_2$ on protein stability at pH 7.0 was investigated but not interpretable due to quenching by the metal.

Figure 3: Effect of manganese on the structure of PA-Nter observed by far UV CD spectra. The secondary structure content of PA-Nter was monitored in absence (full line) or presence of 1 mM $MnCl_2$ (dashed line).

Figure 4: Assay of thermal stability with 2,4-Dioxo-4-phenylbutanoic acid (DPBA). Thermal shift assay with different concentrations of DPBA. DPBA further stabilizes PA-Nter in the presence of $MnCl_2$.

Figure 5: Time series of the endonuclease activity of PA-Nter. 10 µM purified panhandle RNA (ph-RNA) was incubated with 13 µM PA-Nter plus 1 mM $MnCl_2$. The incubation at 37°C was stopped by adding 20 mM EGTA after 5, 10, 20, 40, and 80 minutes (lanes 4 to 8, respectively). As controls, ph-RNA was incubated for 80 minutes at 37°C with only PA-Nter (lane 1) only $MnCl_2$ (lane 2) or PA-Nter and $MnCl_2$ plus 20 mM EGTA. The reaction products were loaded on an 8% acrylamide/8 M urea gel and stained with methylene blue.

Figure 6: Effect of divalent cations on PA-Nter endonuclease (RNase) activity. In the top panel (**A**), purified ph-RNA plus PA-Nter were incubated at pH 8 in the presence of β-mercaptoethanol and 1.5 mM $MnCl_2$, $CaCl_2$, $MgCl_2$, $ZnCl_2$, or $CoCl_2$. In the bottom panel (**B**), ph-RNA and PA-Nter were incubated at pH 7 with 1.5 mM $MnCl_2$, $CaCl_2$, $MgCl_2$, $NiCl_2$, or $CoCl_2$. After 30 minutes the reactions were stopped by adding 20 mM EGTA. Controls were performed using either salts or PA-Nter alone as indicated. The reaction products were loaded on 8% or 15% (for bottom panel) acrylamide/8 M urea and stained with methylene blue. Note

that at pH 7, $CoCl_2$ stimulated the endonuclease stronger than $MnCl_2$. At pH 8, $CoCl_2$ precipitates and, thus, does not activate the endonuclease activity.

Figure 7: PA-Nter endonuclease (RNase) activity on different RNA substrates. SRP *Alu*-RNA, tRNA, U-rich RNA, ph-RNA or short ph-RNA were incubated with PA-Nter plus 1 mM $MnCl_2$ (lanes 2, 4, 6, 8, and 10) or in the absence of PA-Nter (lanes 1, 3, 5, 7, and 9). The digestion was performed at 37°C. After 40 minutes the reaction was stopped by adding 20 mM EGTA. The reaction products were loaded on a 15% acrylamide/8 M urea gel and stained with methylene blue.

Figure 8: Endonuclease activity of PA-Nter on single stranded DNA. Single stranded DNA plasmid M13mp18 (100 ng/µl) (Fermentas) was incubated for 60 minutes at 37°C in the presence of PA-Nter plus $MnCl_2$ (lane 4). The reaction was stopped by adding 20 mM EGTA. As controls, M13mp18 was incubated with 1 mM $MnCl_2$ only (lane 2) or PA-Nter plus $MnCl_2$ and 20 mM EGTA (lane 3). The reaction products were loaded on a 0.8% agarose gel and stained with ethidium bromide.

Figure 9: Inhibition of PA-Nter endonuclease activity by 2,4-Dioxo-4-phenylbutanoic acid (DPBA). Cleavage of ph-RNA **(A)** or M13mp18 ssDNA **(B)** by PA-Nter was tested at 37°C during 40 minutes in the presence of 1 mM $MnCl_2$ and increasing concentrations of DPBA (0, 6.5, 13, 20, 26, 40, 65, 130, and 1000 µM). As a control, ph-RNA or ssDNA was incubated with 1 mM $MnCl_2$ alone (lanes 1). The reaction products were loaded on 8% acrylamide/8 M urea and stained with methylene blue **(A)** or on a 0.8% agarose gel and stained with ethidium bromide **(B)**.

Figure 10: Three-dimensional structure of PA-Nter. Ribbon diagram of the structure of influenza PA-Nter with α-helices (medium grey) and β-strands (light grey). The key active site residues are indicated in stick representation.

Figure 11: Sequence alignment of polypeptide fragments derived from the PA-subunit of representative influenza strains: A/Victoria/3/1975 (human H3N2; amino acid residues 1 to 209 of SEQ ID NO: 2), A/Duck/Vietman/1/2007 (avian H5N1; amino acid residues 1 to 209 of SEQ ID NO: 8), B/Ann Arbor/1/1966 (amino acid residues 1 to 206 of SEQ ID NO: 4) and C/Johannesburg/1/1966 (amino acid residues 1 to 189 of SEQ ID NO: 6). The secondary structure of A/Victoria/3/1975 is shown over the sequence alignment. The boxed sequences indicate sequence similarity between the four sequences. Residues in a solid black background are identical between the four sequences. The triangles indicate the key active site residues.

Figure 12: Representation of PA-Nter shaded according to residue conservation as based on the sequence alignment shown in Figure 11, with grey (not conserved), grey (equivalent residues) and black (100 % conserved).

Figure 13: Electrostatic surface potential of PA-Nter. The orientation is as in Figure 12. Electrostatic surface potential of PA-Nter in the absence of metal ions. The potential scale ranges from -10.0 kT/e (medium grey, acidic residues Asp(D) and Glu(E)) to 3.0 kT/e (dark grey, basic residues Lys(K) and Arg(R)).

Figure 14: Comparison of PA-Nter with other nucleases of the PD-(D/E)XK superfamily. Comparison of PA-Nter (left, **A**), *P. furiosus* Holliday junction resolvase Hjc (PDB entry 1GEF) (middle, **B**) and *E. coli Eco*RV restriction enzyme (PDB entry 1STX, product complex with DNA and manganese) (right, **C**) after superposition of the conserved core active site structural motif. The rootmean-square-deviations are 2.9 Å for 77 aligned Cα atoms of Hjc and 2.46 (3.1) Å for 55 (72) aligned Cα atoms of *Eco*RV. Secondary-structure elements are as in Figure 10 with key active sites residues in stick representation.

Figure 15: Details of the manganese ion interactions with the active sites of influenza PA-Nter (molecule A) (left, **A**) and *E. coli EcoR*V restriction enzyme (product complex) (right, **B**). The active site elements and residues are shown respectively in leight grey and dark grey (left) and dark grey (right). Manganese ions and water molecules are respectively medium grey and dark grey spheres. The anomalous difference map contoured at 3 σ, calculated using manganese K edge (wavelength 1.89) diffraction data and model phases, is in dark grey. Peak heights are 14.1, 10.1, and 5.0 σ for Mn1, Mn2 and the sulphur of Cys45 respectively. Note that in metal dependent nucleases, the exact configuration of the metal ions and acidic side chains subtly depends on the reaction co-ordinate.

Figure 16: Superposition of the active sites of influenza PA-Nter and *E. coli Eco*RV restriction enzyme. PA-Nter secondary structure elements and active sites residues (indicated with PA) are shown in light grey with the manganese ions in medim grey. Superposed are the equivalent elements of *Eco*RV (PDB entry 1STX) (Horton and Perona, 2004, Biochemistry 43:6841-6857) in dark grey (indicated with E) for the protein and dark grey for the manganese ions. Key active site metal binding and catalytic functional groups of the two proteins align.

Figure 17: Comparison of *Eco*RV product complex (**B**) and Pa-Nter with Glu66 from a neighbouring molecule (**A**). The active site elements and residues of PA-Nter (molecule A) are shown in light grey with manganese ions in medium grey and the Glu66 containing loop of the adjacent molecule in light grey. In the same orientation, after superposition of the two

structures, *E. coli Eco*RV restriction enzyme (PDB entry 1STX) (Horton and Perona, *supra*) is shown in dark grey with the DNA bases in light grey and the manganese ions in medium grey. The carboxyl function of Glu59 superimposes on the scissile phosphate of dA7 whereas the well-ordered sulphate ion found in the active site of PA-Nter occupies the position of the phosphate part of dT8.

Figure 18: Refined atomic structure coordinates for PA polypeptide fragment amino acids 1 to 209 according to amino acids 1 to 209 of the amino acid sequence set forth in SEQ ID NO: 2. There are three molecules in the asymmetric unit denoted A, B, and C. The file header gives information about the structure refinement. "Atom" refers to the element whose coordinates are measured. The first letter in the column defines the element. The 3-letter code of the respective amino acid is given and the amino acid sequence position. The first 3 values in the line "Atom" define the atomic position of the element as measured. The fourth value corresponds to the occupancy and the fifth (last) value is the temperature factor (B factor). The occupancy factor refers to the fraction of the molecules in which each atom occupies the position specified by the coordinates. A value of "1" indicates that each atom has the same conformation, i.e., the same position, in all molecules of the crystal. B is a thermal factor that measures movement of the atom around its atomic center. The anisotropic temperature factors are given in the lines marked "ANISOU". This nomenclature corresponds to the PDB file format.

DETAILED DESCRIPTION OF THE INVENTION

Before the present invention is described in detail below, it is to be understood that this invention is not limited to the particular methodology, protocols and reagents described herein as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art.

In the following, the elements of the present invention will be described. These elements are listed with specific embodiments, however, it should be understood that they may be combined in any manner and in any number to create additional embodiments. The variously described examples and preferred embodiments should not be construed to limit the present invention to only the explicitly described embodiments. This description should be understood to support and encompass embodiments which combine the explicitly described

embodiments with any number of the disclosed and/or preferred elements. Furthermore, any permutations and combinations of all described elements in this application should be considered disclosed by the description of the present application unless the context indicates otherwise. For example, if in a preferred embodiment the polypeptide fragment of the present invention corresponds to amino acids 1 to 209 of the amino acid sequence set forth in SEQ ID NO: 2 and in another preferred embodiment the PA polypeptide fragment according to the present invention may ba tagged with a peptide-tag that is preferably cleavable from the PA polypeptide fragment, preferably using a TEV protease, it is a preferred embodiment of the invention that the polypeptide fragment corresponding to amino acids 1 to 209 of the amino acid sequence set forth in SEQ ID NO: 2 is tagged with a peptide-tag that is cleavable from the PA polypeptide using a TEV protease.

Preferably, the terms used herein are defined as described in "A multilingual glossary of biotechnological terms: (IUPAC Recommendations)", H.G.W. Leuenberger, B. Nagel, and H. Kölbl, Eds., Helvetica Chimica Acta, CH-4010 Basel, Switzerland, (1995).

To practice the present invention, unless otherwise indicated, conventional methods of chemistry, biochemistry, and recombinant DNA techniques are employed which are explained in the literature in the field (cf., e.g., *Molecular Cloning: A Laboratory Manual*, 2nd Edition, J. Sambrook et al. eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps. As used in this specification and the appended claims, the singular forms “a”, “an”, and “the” include plural referents, unless the content clearly dictates otherwise.

Several documents are cited throughout the text of this specification. Each of the documents cited herein (including all patents, patent applications, scientific publications, manufacturer's specifications, instructions, etc.), whether *supra* or *infra*, are hereby incorporated by reference in their entirety. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

The term “polypeptide fragment” refers to a part of a protein which is composed of a single amino acid chain. The term “protein” comprises polypeptide fragments that resume a secondary and tertiary structure and additionally refers to proteins that are made up of several

amino acid chains, i.e., several subunits, forming quartenary structures. The term "peptide" refers to short amino acid chains of up to 50 amino acids that do not necessarily assume secondary or tertiary structures. A "peptoid" is a peptidomimetic that results from the oligomeric assembly of N-substituted glycines.

Residues in two or more polypeptides are said to "correspond" to each other if the residues occupy an analogous position in the polypeptide structures. As is well known in the art, analogous positions in two or more polypeptides can be determined by aligning the polypeptide sequences based on amino acid sequence or structural similarities. Such alignment tools are well known to the person skilled in the art and can be, for example, obtained on the World Wide Web, e.g., ClustalW (www.ebi.ac.uk/clustalw) or Align (http://www.ebi.ac.uk/emboss/align/index.html) using standard settings, preferably for Align EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5. Those skilled in the art understand that it may be necessary to introduce gaps in either sequence to produce a satisfactory alignment. For example, residues 1 to 196 in the Influenza A virus PA subunit correspond to residues 1 to 195 and 1 to 178 in the Influenza B and C virus PA subunits, respectively. Residues in two or more PA subunits are said to "correspond" if the residues are aligned in the best sequence alignment. The "best sequence alignment" between two polypeptides is defined as the alignment that produces the largest number of aligned identical residues. The "region of best sequence alignment" ends and, thus, determines the metes and bounds of the length of the comparison sequence for the purpose of the determination of the similarity score, if the sequence similarity, preferably identity, between two aligned sequences drops to less than 30%, preferably less than 20%, more preferably less than 10% over a length of 10, 20 or 30 amino acids. A part of the best sequence alignment for the amino acid sequences of Influenza A (aa 1 to 209), B (aa 1 to 206), and C (aa 1 to 189) PA subunits is shown in Figure 11.

For example, amino acids Tyr24, His41, Glu80, Arg84, Leu106, Asp108, Glu119, Ile120, Tyr130, Glu133, Lys134, and Lys137 of the amino acid sequence set forth in SEQ ID NO: 2 (Influenza A virus PA subunit) correspond to amino acids Phe24, His41, Glu81, Arg85, Leu107, Asp109, Glu120, Val121, Tyr131, Lys134, Lys135, and Lys138 of the amino acid sequence set forth in SEQ ID NO: 4 (Influenza B virus PA subunit) and amino acids Ala24, His41, Glu65, Arg69, Leu91, Asp93, Glu104, Ile105, Tyr115, Ser118, Lys119, and Lys122 of the amino acid sequence set forth in SEQ ID NO: 6 (Influenza C virus PA subunit), respectively.

The present invention includes Influenza virus RNA-dependent RNA polymerase PA subunit fragments possessing endonuclease activity. The term "RNA-dependent RNA polymerase subunit PA" preferably refers to the PA subunit of Influenza A, Influenza B, or Influenza C virus, preferably having an amino acid sequence as set out in SEQ ID NO: 2, 4, or 6. "RNA-dependent RNA polymerase subunit PA variants" have at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g., Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with the amino acid sequence set forth in SEQ ID NO: 2, 4, or 6. It is preferred that when a naturally occurring PA variant is aligned with a PA subunit according to SEQ ID NO: 2, 4, or 6 that the alignment will be over the entire length of the two proteins and, thus, that the alignment score will be determined on this basis. It is, however, possible that the natural variant may comprise C-terminal/N-terminal or internal deletions or additions, e.g., through N- or C-terminal fusions. In this case, only the best aligned region is used for the assessment of similarity and identity, respectively. Preferably and as set out in more detail below, fragments derived from these variants show the indicated similarity and identity, respectively, preferably within the region required for endonuclease activity. Accordingly, any alignment between SEQ ID NO: 2, 4, or 6 and a PA variant should preferably comprise the endonuclease active site. Thus, the above sequence similarity and identity, respectively, to SEQ ID NO: 2, 4, or 6 occurs at least over a length of 100, 110, 120, 130, 140, 150, 160, 165, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300 or more amino acids, preferably comprising the endonuclease active site. A large number of natural PA variants of sequences according to SEQ ID NO: 2, 4, or 6 are known and have been described in the literature. All these PA variants are comprised and can be the basis for the polypeptide fragments of the present invention. Preferred examples of the Influenza A PA subunit, if SEQ ID NO: 2 is used as reference sequence, comprise mutations at one or more of positions Phe4, Ala20, Leu28, Glu31, Val44, Tyr48, Asn55, Gln57, Gly58, Val62, Leu65, Asp66, Thr85, Gly99, Ala100, Glu101, Ile118, Ile129, Asn142, Ile145, Glu154, Lys158, Asp164, Ile171, Lys172, Ile178, Asn184, and/or Arg204. In a preferred embodiment, said variant comprises one or more of the following mutations: Phe4Leu, Ala20Thr, Leu28Pro, Glu31Lys, Val44Ala, Tyr48His, Asn55Asp, Gln57Arg, Gly58Ser, Val62Ile, Leu65Ser, Asp66Gly, Thr85Ala, Gly99Lys, Ala100Val, Glu101Asp, Ile118Thr, Ile129Thr, Asn142Lys, Ile145Leu,

Glu154Gly, Lys158Gln, Asp164Val, Ile171Val, Lys172Arg, Ile178Val, Asn184Ser, Asn184Arg, and/or Arg204Lys. Preferred variants of the Influenza B virus PA subunit, if SEQ ID NO: 4 is used as reference sequence, include mutations at one ore more of the following amino acid positions: Thr60, Asn86, Arg105, Asn158, His160, and/or Ile196. In a preferred embodiment the Influenza B virus PA subunit variant comprises one or more of the following mutations: Thr60Ala, Asn86Thr, Arg105Lys, Asn158Asp, His160Ser, and/or Ile196Val. Preferred variants of the Influenza C virus PA subunit, if SEQ ID NO: 6 is used as reference sequence, include mutations at one or more of the following amino acid positions: Thr11, Leu53, Ser58, Gly70, and/or Ala111. In a preferred embodiment, said mutations are as follows: Thr11Ala, Leu53Met, Ser58Asn, Gly70Arg, and/or Ala111Thr.

The polypeptide fragments of the present invention are, thus, based on RNA-dependent RNA polymerase subunit PA or variants thereof as defined above. Accordingly, in the following specification the terms "polypeptide fragment(s)" and "PA polypeptide fragments" always comprise such fragments derived both from the PA proteins as set out in SEQ ID NO: 2, 4, or 6 and fragments derived from PA protein variants thereof, as set out above, possessing endonuclease activity. However, the specification also uses the term "PA polypeptide fragment variants" or "PA fragment variants" to specifically refer to PA fragments possessing endonuclease activity that are derived from RNA-dependent RNA polymerase subunit PA variants. The PA polypeptide fragments of the present invention thus preferably comprise, essentially consist or consist of sequences of naturally occurring viral PA subunits, preferably Influenza virus PA subunit. It is, however, also envisioned that the PA fragment variants further contain amino acid substitutions at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more amino acid positions, and have at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g., Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with the amino acid sequence set forth in SEQ ID NO: 2, 4, or 6. It is understood that PA fragments of the present invention may comprise additional amino acids not derived from PA, like, e.g., tags, enzymes etc., such additional amino acids will not be considered in such an alignment, i.e., are excluded from the calculation of the alignment score. In a preferred embodiment, the above indicated alignment score is obtained when aligning the sequence of the fragment with SEQ ID NO: 2, 4, or 6 at least over a length of 70,

80, 90, 100, 110, 120, 130, 140, 150, 160, 165, 170, 180, or 190 amino acids, wherein the respective sequence of SEQ ID NO: 2, 4, or 6, preferably comprises the endonuclease active site.

In a preferred embodiment, the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 196 of Influenza A virus PA or consist of amino acid residues 1 to 196 (derived from SEQ ID NO: 2) and have at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g., Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with amino acid residues 1 to 196 of the sequence set forth in SEQ ID NO: 2, more preferably the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 209 of Influenza A virus PA or consist of amino acid residues 1 to 209 (derived from SEQ ID NO: 2) and have at least 70%, more preferably 75%, more preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g., Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with the amino acid residues 1 to 209 of the amino acid sequence set forth in SEQ ID NO: 2, more preferably the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 213 of Influenza A virus PA or consist of amino acid residues 1 to 213 (derived from SEQ ID NO: 2) and have at least 60%, more preferably 65%, more preferably 70%, more preferably 75%, more preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g., Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with amino acid residues 1 to 213 of the amino acid sequence set forth in SEQ ID NO: 2. In preferred embodiments, the Influenza A virus PA polypeptide fragment variants of the present invention comprise mutations, preferably naturally occurring mutations such as mutations in

one or more of the following amino acid residues when compared to SEQ ID NO: 2: Phe4, Ala20, Leu28, Glu31, Val44, Tyr48, Asn55, Gln57, Gly58, Val62, Leu65, Asp66, Thr85, Gly99, Ala100, Glu101, Ile118, Ile129, Asn142, Ile145, Glu154, Lys158, Asp164, Ile171, Lys172, Ile178, Asn184, and/or Arg204. In a preferred embodiment, said variant comprises one or more of the following mutations: Phe4Leu, Ala20Thr, Leu28Pro, Glu31Lys, Val44Ala, Tyr48His, Asn55Asp, Gln57Arg, Gly58Ser, Val62Ile, Leu65Ser, Asp66Gly, Thr85Ala, Gly99Lys, Ala100Val, Glu101Asp, Ile118Thr, Ile129Thr, Asn142Lys, Ile145Leu, Glu154Gly, Lys158Gln, Asp164Val, Ile171Val, Lys172Arg, Ile178Val, Asn184Ser, Asn184Arg, and/or Arg204Lys.

In a preferred embodiment, the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 195 of Influenza B virus PA or consist of amino acid residues 1 to 195 (derived from SEQ ID NO: 4) and have at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g., Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with amino acid residues 1 to 195 of the amino acid sequence set forth in SEQ ID NO: 4, more preferably the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 206 of Influenza B virus PA or consist of amino acid residues 1 to 206 (derived from SEQ ID NO: 4) and have at least 70%, more preferably 75%, more preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g., Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with the amino acid residues 1 to 206 of the sequence set forth in SEQ ID NO: 4, more preferably the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 210 of Influenza B virus PA or consist of amino acid residues 1 to 210 (derived from SEQ ID NO: 4) and have at least 60%, more preferably 65%, more preferably 70%, more preferably 75%, more preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best

sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g., Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with amino acid residues 1 to 210 of the amino acid sequence set forth in SEQ ID NO: 4. In preferred embodiments, the Influenza B virus PA polypeptide fragment variants of the present invention comprise mutations, preferably naturally occurring mutations, at one ore more of the following amino acid positions compared to SEQ ID NO: 4: Thr60, Asn86, Arg105, Asn158, His160, and/or Ile196. In a preferred embodiment the Influenza B virus PA subunit variant comprises one or more of the following mutations: Thr60Ala, Asn86Thr, Arg105Lys, Asn158Asp, His160Ser, and/or Ile196Val.

In a preferred embodiment, the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 178 of Influenza C virus PA or consist of amino acid residues 1 to 178 (derived from SEQ ID NO: 6) and have at least 80%, more preferably 85%, more preferably 90%, most preferably 95% sequence similarity over the entire length of the fragment with amino acid residues 1 to 178 of the amino acid sequence set forth in SEQ ID NO: 6, more preferably the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 189 of Influenza C virus PA or consist of amino acid residues 1 to 189 (derived from SEQ ID NO: 6) and have at least 70%, more preferably 75%, more preferably 80%, more preferably 85%, most preferably 90% sequence similarity over the entire length of the fragment with amino acid residues 1 to 189 of the amino acid sequence set forth in SEQ ID NO: 6, more preferably the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 193 of Influenza C virus PA or consist of amino acid residues 1 to 193 (derived from SEQ ID NO: 6) and have at least 60%, more preferably 65%, more preferably 70%, more preferably 75%, more preferably 80%, more preferably 85%, most preferably 90% sequence similarity over the entire length of the fragment with amino acid residues 1 to 193 of the amino acid sequence set forth in SEQ ID NO: 6. In preferred embodiments, the Influenza C virus PA polypeptide fragment variants of the present invention comprise mutations, preferably naturally occurring mutations such as mutations in one or more of the following amino acid residues when compared to SEQ ID NO: 6: Thr11, Leu53, Ser58, Gly70, and/or Ala111. In a preferred embodiment, said mutations are as follows: Thr11Ala, Leu53Met, Ser58Asn, Gly70Arg, and/or Ala111Thr.

In the context of the present invention, the term "PA-Nter" refers to a polypeptide fragment which consists of amino acid residues 1 to 209 of the amino acid sequence as set forth in SEQ ID NO: 2 with an additional amino-terminal linker, i.e., GMGSGMA (SEQ ID NO: 19).

If a PA polypeptide fragment of the present invention comprises one of the above outlined amino acid residues, it is preferred that the other amino acid residues are not derived from the respective Influenza A, B, or C virus PA protein.

The term "sequence similarity" means that amino acids at the same position of the best sequence alignment are identical or similar, preferably identical. "Similar amino acids" possess similar characteristics, such as polarity, solubility, hydrophilicity, hydrophobicity, charge, or size. Similar amino acids are preferably leucine, isoleucine, and valine; phenylalanine, tryptophan, and tyrosine; lysine, arginine, and histidine; glutamic acid and aspartic acid; glycine, alanine, and serine; threonine, asparagine, glutamine, and methionine. The skilled person is well aware of sequence similarity searching tools, e.g., available on the World Wide Web (e.g., www.ebi.ac.uk/Tools/similarity.html).

The term "soluble", as used herein, refers to a polypeptide fragment which remains in the supernatant after centrifugation for 30 min at 100,000 x g in an aqueous buffer under physiologically isotonic conditions, for example, 0.14 M sodium chloride or sucrose, at a protein concentration of at least 200 µg/ml, preferably of at least 500 µg/ml, preferably of at least 1 mg/ml, more preferably of at least 2 mg/ml, even more preferably of at least 3 mg/ml, even more preferably of at least 4 mg/ml, most preferably of at least 5 mg/ml in the absence of denaturants such as guanidine or urea in effective concentrations. A protein fragment that is tested for its solubility is preferably expressed in one of the cellular expression systems indicated below.

The term "purified" in reference to a polypeptide, does not require absolute purity such as a homogenous preparation, rather it represents an indication that the polypeptide is relatively purer than in the natural environment. Generally, a purified polypeptide is substantially free of other proteins, lipids, carbohydrates, or other materials with which it is naturally associated, preferably at a functionally significant level, for example, at least 85% pure, more preferably at least 95% pure, most preferably at least 99% pure. The expression "purified to an extent to be suitable for crystallization" refers to a protein that is 85% to 100%, preferably 90% to 100%, more preferably 95% to 100% pure and can be concentrated to higher than 3 mg/ml, preferably higher than 10 mg/ml, more preferably higher than 18 mg/ml without precipitation. A skilled artisan can purify a polypeptide using standard

techniques for protein purification. A substantially pure polypeptide will yield a single major band on a non-reducing polyacrylamide gel.

The term "associate" as used in the context of identifying compounds with the methods of the present invention refers to a condition of proximity between a moiety (i.e., chemical entity or compound or portions or fragments thereof), and an endonuclease active site of the PA subunit. The association may be non-covalent, i.e., where the juxtaposition is energetically favored by, for example, hydrogen-bonding, van der Waals, electrostatic, or hydrophobic interactions, or it may be covalent.

The term "endonuclease activity" or "endonucleolytic activity" refers to an enzymatic activity which results in the cleavage of the phosphodiester bond within a polynucleotide chain. In the context of the present invention, the polypeptide fragments possess an endonucleolytic activity, which is preferably not selective for the polynucleotide type, i.e., the polypeptide fragments according to the present invention preferably exhibit endonucleolytic activity for DNA and RNA, preferably for single stranded DNA (ssDNA) or single stranded RNA (ssRNA). In this context, "Single stranded" means that a stretch of preferably at least 3 nucleotides, preferably at least 5 nucleotides, more preferably at least 10 nucleotides within the polynucleotide chain are single stranded, i.e., not base paired to another nucleotide. Preferably, the endonucleolytic activity of the polypeptide fragments according to the present invention is not dependent on recognition sites, i.e., specific nucleotide sequences, but results in unspecific cleavage of polynucleotide chains. For example, the skilled person may test for endonucleolytic activity of polypeptide fragments according to the present invention by incubating RNA or DNA substrates such as panhandle RNA or a linear or circular single stranded DNA, e.g., the circular M13mp18 DNA (MBI Fermentas), with or without the respective polypeptide fragment, for example, at 37°C for a certain period of time such as for 5, 10, 20, 40, 60, or 80 minutes, and test for the integrity of the polynucleotides, for example, by gel electrophoresis.

The term "nucleotide" as used herein refers to a compound consisting of a purine, deazapurine, or pyrimidine nucleoside base, e.g., adenine, guanine, cytosine, uracil, thymine, deazaadenine, deazaguanosine, and the like, linked to a pentose at the 1' position, including 2'-deoxy and 2'-hydroxyl forms, e.g., as described in Kornberg and Baker, DNA Replication, 2nd Ed. (Freeman, San Francisco, 1992) and further include, but are not limited to, synthetic nucleosides having modified base moieties and/or modified sugar moieties, e.g., described generally by Scheit, Nucleotide Analogs (John Wiley, N.Y., 1980).

The term "isolated polynucleotide" refers to polynucleotides that were (i) isolated from their natural environment, (ii) amplified by polymerase chain reaction, or (iii) wholly or partially synthesized, and means a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases and includes DNA and RNA molecules, both sense and anti-sense strands. The term comprises cDNA, genomic DNA, and recombinant DNA. A polynucleotide may consist of an entire gene, or a portion thereof.

The term "recombinant vector" as used herein includes any vectors known to the skilled person including plasmid vectors, cosmid vectors, phage vectors such as lambda phage, viral vectors such as adenoviral or baculoviral vectors, or artificial chromosome vectors such as bacterial artificial chromosomes (BAC), yeast artificial chromosomes (YAC), or P1 artificial chromosomes (PAC). Said vectors include expression as well as cloning vectors. Expression vectors comprise plasmids as well as viral vectors and generally contain a desired coding sequence and appropriate DNA sequences necessary for the expression of the operably linked coding sequence in a particular host organism (e.g., bacteria, yeast, plant, insect, or mammal) or in *in vitro* expression systems. Cloning vectors are generally used to engineer and amplify a certain desired DNA fragment and may lack functional sequences needed for expression of the desired DNA fragments.

"Recombinant host cell", as used herein, refers to a host cell that comprises a polynucleotide that codes for a polypeptide fragment of interest, i.e., the PA polypeptide fragment or variants thereof according to the invention. This polynucleotide may be found inside the host cell (i) freely dispersed as such, (ii) incorporated in a recombinant vector, or (iii) integrated into the host cell genome or mitochondrial DNA. The recombinant cell can be used for expression of a polynucleotide of interest or for amplification of the polynucleotide or the recombinant vector of the invention. The term "recombinant host cell" includes the progeny of the original cell which has been transformed, transfected, or infected with the polynucleotide or the recombinant vector of the invention. A recombinant host cell may be a bacterial cell such as an *E. coli* cell, a yeast cell such as *Saccharomyces cerevisiae* or *Pichia pastoris*, a plant cell, an insect cell such as SF9 or Hi5 cells, or a mammalian cell. Preferred examples of mammalian cells are Chinese hamster ovary (CHO) cells, green African monkey kidney (COS) cells, human embryonic kidney (HEK293) cells, HELA cells, and the like.

As used herein, the term "crystal" or "crystalline" means a structure (such as a three-dimensional solid aggregate) in which the plane faces intersect at definite angles and in which there is a regular structure (such as internal structure) of the constituent chemical species. The term "crystal" can include any one of: a solid physical crystal form such as an experimentally

prepared crystal, a crystal structure derivable from the crystal (including secondary and/or tertiary and/or quaternary structural elements), a 2D and/or 3D model based on the crystal structure, a representation thereof such as a schematic representation thereof or a diagrammatic representation thereof, or a data set thereof for a computer. In one aspect, the crystal is usable in X-ray crystallography techniques. Here, the crystals used can withstand exposure to X-ray beams and are used to produce diffraction pattern data necessary to solve the X-ray crystallographic structure. A crystal may be characterized as being capable of diffracting X-rays in a pattern defined by one of the crystal forms depicted in T. L. Blundell and L. N. Johnson, "Protein Crystallography", Academic Press, New York (1976).

The term "unit cell" refers to a basic parallelepiped shaped block. The entire volume of a crystal may be constructed by regular assembly of such blocks. Each unit cell comprises a complete representation of the unit of pattern, the repetition of which builds up the crystal.

The term "space group" refers to the arrangement of symmetry elements of a crystal. In a space group designation the capital letter indicates the lattice type and the other symbols represent symmetry operations that can be carried out on the contents of the asymmetric unit without changing its appearance.

The term "structure coordinates" refers to a set of values that define the position of one or more amino acid residues with reference to a system of axes. The term refers to a data set that defines the three-dimensional structure of a molecule or molecules (e.g., Cartesian coordinates, temperature factors, and occupancies). Structural coordinates can be slightly modified and still render nearly identical three-dimensional structures. A measure of a unique set of structural coordinates is the root mean square deviation of the resulting structure. Structural coordinates that render three-dimensional structures (in particular, a three-dimensional structure of an enzymatically active center) that deviate from one another by a root mean square deviation of less than 3 Å, 2 Å, 1.5 Å, 1.0 Å, or 0.5 Å may be viewed by a person of ordinary skill in the art as very similar.

The term "root mean square deviation" means the square root of the arithmetic mean of the squares of the deviations from the mean. It is a way to express the deviation or variation from a trend or object. For purposes of this invention, the "root mean square deviation" defines the variation in the backbone of a variant of the PA polypeptide fragment or the enzymatically active center therein from the backbone of the PA polypeptide fragment or the enzymatically active center therein as defined by the structure coordinates of the PA polypeptide fragment PA-Nter according to Figure 18.

As used herein, the term "constructing a computer model" includes the quantitative and qualitative analysis of molecular structure and/or function based on atomic structural information and interaction models. The term "modeling" includes conventional numeric-based molecular dynamic and energy minimization models, interactive computer graphic models, modified molecular mechanics models, distance geometry, and other structure-based constraint models.

The term "fitting program operation" refers to an operation that utilizes the structure coordinates of a chemical entity, an enzymatically active center, a binding pocket, molecule or molecular complex, or portion thereof, to associate the chemical entity with the enzymatically active center, the binding pocket, molecule or molecular complex, or portion thereof. This may be achieved by positioning, rotating or translating the chemical entity in the enzymatically active center to match the shape and electrostatic complementarity of the enzymatically active center. Covalent interactions, non-covalent interactions such as hydrogen bond, electrostatic, hydrophobic, van der Waals interactions, and non-complementary electrostatic interactions such as repulsive charge-charge, dipole-dipole and charge-dipole interactions may be optimized. Alternatively, one may minimize the deformation energy of binding of the chemical entity to the enzymatically active center.

As used herein, the term "test compound" refers to an agent comprising a compound, molecule, or complex that is being tested for its ability to inhibit the endonucleolytic activity of the polypeptide fragment of interest, i.e., the PA polypeptide fragment of the invention or variants thereof possessing endonucleolytic acitvity. Test compounds can be any agents including, but not restricted to, peptides, peptoids, polypeptides, proteins (including antibodies), lipids, metals, nucleotides, nucleotide analogs, nucleosides, nucleic acids, small organic or inorganic molecules, chemical compounds, elements, saccharides, isotopes, carbohydrates, imaging agents, lipoproteins, glycoproteins, enzymes, analytical probes, polyamines, and combinations and derivatives thereof. The term “small molecules” refers to molecules that have a molecular weight between 50 and about 2,500 Daltons, preferably in the range of 200-800 Daltons. In addition, a test compound according to the present invention may optionally comprise a detectable label. Such labels include, but are not limited to, enzymatic labels, radioisotope or radioactive compounds or elements, fluorescent compounds or metals, chemiluminescent compounds and bioluminescent compounds. Well known methods may be used for attaching such a detectable label to a test compound. The test compound of the invention may also comprise complex mixtures of substances, such as extracts containing natural products, or the products of mixed combinatorial syntheses. These

can also be tested and the component that inhibits the endonucleolytic activity of the target polypeptide fragment can be purified from the mixture in a subsequent step. Test compounds can be derived or selected from libraries of synthetic or natural compounds. For instance, synthetic compound libraries are commercially available from Maybridge Chemical Co. (Trevillet, Cornwall, UK), ChemBridge Corporation (San Diego, CA), or Aldrich (Milwaukee, WI). A natural compound library is, for example, available from TimTec LLC (Newark, DE). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal cell and tissue extracts can be used. Additionally, test compounds can be synthetically produced using combinatorial chemistry either as individual compounds or as mixtures. A collection of compounds made using combinatorial chemistry is referred to herein as a combinatorial library.

In the context of the present invention, "a compound which modulates the endonucleolytic activity" may increase or decrease, preferably inhibit the endonucleolytic activity of the PA subunit or the viral RNA-dependent RNA polymerase or a variant thereof. Preferably, such a compound is specific for the endonucleolytic activity of the viral PA subunit or variant thereof and does not modulate, preferably decrease the endonucleolytic activity of other endonucleases, in particular mammalian endonucleases.

The term "a compound which decreases the endonucleolytic activity" means a compound which decreases the endonucleolytic activity of the PA subunit of the viral RNA-dependent RNA polymerase from the Orthomyxoviridae family or a variant thereof by 50%, more preferably by 60%, even more preferably by 70%, even more preferably by 80%, even more preferably by 90%, and most preferably by 100% compared to the endonucleolytic activity of the PA subunit or a variant thereof without said compound but with otherwise the same reaction conditions, i.e., buffer conditions, reaction time and temperature. It is most preferred that the compound which decreases the endonucleolytic activity of the PA subunit or a variant thereof inhibits said activity, i.e., decreases said activity by at least 95%, preferably by 100% compared to the activity without the compound. It is particularly preferred that the compound that decreases or inhibits the endonucleolytic activity of the PA subunit or a variant thereof specifically decreases or inhibits the endonucleolytic activity of the PA subunit or a variant thereof but does not inhibit the endonucleolytic activity of other endonucleases such as RNase H or restriction endonucleases to the same extent, preferably not at all. For example, the skilled person may set up the following samples with the same buffer and reaction conditions as well as substrate and endonuclease concentrations: (1) substrate such as panhandle RNA, endonucleolytically active PA polypeptide fragment or

variant thereof, (2) substrate such as panhandle RNA, endonucleolytically active PA polypeptide fragment or variant thereof, test compound, (3) substrate such as panhandle RNA, reference endonuclease such as RNAse H, (4) substrate such as panhandle RNA, reference nucleotide such as RNAse H, test compound. After incubation of the samples, the skilled person may analyze the substrate, for example, by gel electrophoresis. Test compounds which result in cleaved substrate in sample (2) and intact substrate in sample (4) are preferred.

The term "in a high-throughput setting" refers to high-throughput screening assays and techniques of various types which are used to screen libraries of test compounds for their ability to inhibit the endonuclease activity of the polypeptide fragment of interest. Typically, the high-throughput assays are performed in a multi-well format and include cell-free as well as cell-based assays.

The term "antibody" refers to both monoclonal and polyclonal antibodies, i.e., any immunoglobulin protein or portion thereof which is capable of recognizing an antigen or hapten, i.e., the PA polypeptide fragment possessing endonucleolytic activity or a peptide thereof. In a preferred embodiment, the antibody is capable of binding to the enzymatically (endonucleolytically) active center within the PA polypeptide fragment or variant thereof. Antigen-binding portions of the antibody may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. In some embodiments, antigen-binding portions include Fab, Fab', $F(ab')_2$, Fd, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies such as humanized antibodies, diabodies, and polypeptides that contain at least a portion of an antibody that is sufficient to confer specific antigen binding to the polypeptide.

The term "pharmaceutically acceptable salt" refers to a salt of a compound identifiable by the methods of the present invention or a compound of the present invention. Suitable pharmaceutically acceptable salts include acid addition salts which may, for example, be formed by mixing a solution of compounds of the present invention with a solution of a pharmaceutically acceptable acid such as hydrochloric acid, sulfuric acid, fumaric acid, maleic acid, succinic acid, acetic acid, benzoic acid, citric acid, tartaric acid, carbonic acid or phosphoric acid. Furthermore, where the compound carries an acidic moiety, suitable pharmaceutically acceptable salts thereof may include alkali metal salts (e.g., sodium or potassium salts); alkaline earth metal salts (e.g., calcium or magnesium salts); and salts formed with suitable organic ligands (e.g., ammonium, quaternary ammonium and amine cations formed using counteranions such as halide, hydroxide, carboxylate, sulfate, phosphate, nitrate, alkyl sulfonate and aryl sulfonate). Illustrative examples of

pharmaceutically acceptable salts include, but are not limited to, acetate, adipate, alginate, ascorbate, aspartate, benzenesulfonate, benzoate, bicarbonate, bisulfate, bitartrate, borate, bromide, butyrate, calcium edetate, camphorate, camphorsulfonate, camsylate, carbonate, chloride, citrate, clavulanate, cyclopentanepropionate, digluconate, dihydrochloride, dodecylsulfate, edetate, edisylate, estolate, esylate, ethanesulfonate, formate, fumarate, gluceptate, glucoheptonate, gluconate, glutamate, glycerophosphate, glycolylarsanilate, hemisulfate, heptanoate, hexanoate, hexylresorcinate, hydrabamine, hydrobromide, hydrochloride, hydroiodide, 2-hydroxy-ethanesulfonate, hydroxynaphthoate, iodide, isothionate, lactate, lactobionate, laurate, lauryl sulfate, malate, maleate, malonate, mandelate, mesylate, methanesulfonate, methylsulfate, mucate, 2-naphthalenesulfonate, napsylate, nicotinate, nitrate, N-methylglucamine ammonium salt, oleate, oxalate, pamoate (embonate), palmitate, pantothenate, pectinate, persulfate, 3-phenylpropionate, phosphate/diphosphate, picrate, pivalate, polygalacturonate, propionate, salicylate, stearate, sulfate, subacetate, succinate, tannate, tartrate, teoclate, tosylate, triethiodide, undecanoate, valerate, and the like (see, for example, S. M. Berge et al., "Pharmaceutical Salts", J. Pharm. Sci. 66:1-19 (1977)).

The term "excipient" when used herein is intended to indicate all substances in a pharmaceutical formulation which are not active ingredients such as, e.g., carriers, binders, lubricants, thickeners, surface active agents, preservatives, emulsifiers, buffers, flavoring agents, or colorants.

The term "pharmaceutically acceptable carrier" includes, for example, magnesium carbonate, magnesium stearate, talc, sugar, lactose, pectin, dextrin, starch, gelatin, tragacanth, methylcellulose, sodium carboxymethylcellulose, a low melting wax, cocoa butter, and the like.

DETAILED DESCRIPTION

The present invention establishes for the first time a unique role for the PA subunit of influenza virus polymerase and contradicts the widely held view that the endonuclease active site is located within the PB1 subunit. The present inventors surprisingly found that a small independently folded domain derived from the N-terminus of the PA subunit exhibits the functional properties of the endonuclease reported for the trimeric complex, although this activity was thought to be detectable only in the trimeric complex. Moreover, the inventors found that this PA polypeptide fragment can easily be produced by recombinant means and thus is suitable for *in vitro* studies on the endonucleolytic activity and and its modulation as well as for crystallization to obtain structural information in particular on the active site.

It is one aspect of the present invention to provide a polypeptide fragment comprising an amino-terminal fragment of the PA subunit of a viral RNA-dependent RNA polymerase possessing endonuclease activity, wherein said PA subunit is from a virus belonging to the Orthomyxoviridae family. Preferably, the polypeptide fragment is soluble in an aqueous solution. The minimal length of the polypeptide fragment of the present invention is determined by its ability to cleave polynucleotide chains such as panhandle RNA or single stranded DNA, i.e., the minimal length of the polypeptide is determined by its endonucleolytic activity. Preferably, the endonuclease activity is not dependent on the polynucleotide type, and thus, may be exerted on DNA and RNA, preferably on single stranded DNA and RNA. Preferably, the endonuclease activity is not dependent on specific recognition sites within the substrate polynucleotide.

In a preferred embodiment, the polypeptide fragment is suitable for crystallization, i.e., preferably the polypeptide fragment is crystallizable. Preferably, the crystals obtainable from the polypeptide fragment according to the invention are suitable for structure determination of the polypeptide fragment using X-ray crystallography. Preferably, said crystals are greater than 25 micron cubes and preferably are radiation stable enough to permit more than 85% diffraction data completeness at resolution of preferably 3.5 Å or better to be collected upon exposure to monochromatic X-rays.

In one embodiment, the polypeptide fragment is crystallizable using (i) an aqueous protein solution, i.e., the crystallization solution, with a protein concentration of 5 to 10 mg/ml, e.g., 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, or 10 mg/ml, preferably of 8 to 10 mg/ml in a buffer system such as Tris-HCl at concentrations ranging from 10 mM to 3 M, preferably 10 mM to 2 M, more preferably 20 mM to 1 M, at pH 3 to pH 9, preferably pH 4 to pH 9, more preferably pH 7 to pH 9 and (ii) a precipitant/reservoir solution comprising one or more compounds such as sodium formate, ammonium sulphate, lithium sulphate, magnesium acetate, manganese acetate, or ethylene glycol. Optionally, the protein solution may contain one or more salts such as monovalent salts, e.g., NaCl, KCl, or LiCl, preferably NaCl, at concentrations ranging from 10 mM to 1 M, preferably 20 mM to 500 mM, more preferably 50 mM to 200 mM, and/or divalent salts, e.g., $MnCl_2$, $CaCl_2$, $MgCl_2$, $ZnCl_2$, or $CoCl_2$, preferably $MnCl_2$, at concentrations ranging from 0.1 to 50 mM, preferably 0.5 to 25 mM, more preferably 1 to 10 mM. Preferably, the precipitant/reservoir solution comprises Li_2SO_4 at concentrations ranging from 0.5 to 2 M, preferably 1 to 1.5 M, a buffer system such as MES at concentrations ranging from 20 mM to 1 M, preferably 50 mM to 500 mM, more preferably 75 to 150 mM, at preferably pH 4 to 8, more preferably pH 5 to 7, magnesium

acetate and/or manganese acetate at concentrations ranging from 1 to 100 mM, preferably from 5 to 20 mM, and/or ethylene glycol at concentrations ranging from 1% to 20%, preferably 2% to 8%, more preferably 2 to 4%. The PA polypeptide fragment or variant thereof is preferably 85% to 100% pure, more preferably 90% to 100% pure, even more preferably 95% to 100% pure in the crystallization solution. To produce crystals, the protein solution suitable for crystallization may be mixed with an equal volume of the precipitant solution. In a preferred embodiment, the crystallization medium comprises 0.05 to 2 µl, preferably 0.8 to 1.2 µl, of protein solution suitable for crystallization mixed with a similar, preferably equal volume of precipitant solution comprising 1.0 to 1.4 M Li_2SO_4, 80 to 120 mM MES pH 5.5 to pH 6.5, 5 to 15 mM magnesium acetate and/or manganese acetate, and 2 to 4% ethylene glycol. In another embodiment, the precipitant solution comprises, preferably essentially consists of or consists of 1.2 M Li_2SO_4, 100 mM MES pH 6.0, 10 mM magnesium acetate and/or 10 mM manganese acetate, preferably 10 mM magnesium acetate, and 3% ethylene glycol, and the crystallization/protein solution comprises, preferably essentially consists or consists of 5 to 10 mg/ml protein, 20 mM Tris pH 8.0, 100 mM NaCl, and 2.5 mM $MnCl_2$.

Crystals can be grown by any method known to the person skilled in the art including, but not limited to, hanging and sitting drop techniques, sandwich-drop, dialysis, and microbatch or microtube batch devices. It would be readily apparent to one of skill in the art to vary the crystallization conditions disclosed above to identify other crystallization conditions that would produce crystals of PA polypeptide fragments of the inventions or variants thereof alone or in complex with a compound. Such variations include, but are not limited to, adjusting pH, protein concentration and/or crystallization temperature, changing the identity or concentration of salt and/or precipitant used, using a different method for crystallization, or introducing additives such as detergents (e.g., TWEEN 20 (monolaurate), LDOA, Brij 30 (4 lauryl ether)), sugars (e.g., glucose, maltose), organic compounds (e.g., dioxane, dimethylformamide), lanthanide ions, or poly-ionic compounds that aid in crystallizations. High throughput crystallization assays may also be used to assist in finding or optimizing the crystallization condition.

Microseeding may be used to increase the size and quality of crystals. In brief, micro-crystals are crushed to yield a stock seed solution. The stock seed solution is diluted in series. Using a needle, glass rod or strand of hair, a small sample from each diluted solution is added to a set of equilibrated drops containing a protein concentration equal to or less than a

concentration needed to create crystals without the presence of seeds. The aim is to end up with a single seed crystal that will act to nucleate crystal growth in the drop.

The manner of obtaining the structure coordinates as shown in Figure 18, interpretation of the coordinates and their utility in understanding the protein structure, as described herein, are commonly understood by the skilled person and by reference to standard texts such as J. Drenth, "Principles of protein X-ray crystallography", 2nd Ed., Springer Advanced Texts in Chemistry, New York (1999); and G. E. Schulz and R. H. Schirmer, "Principles of Protein Structure", Springer Verlag, New York (1985). For example, X-ray diffraction data is first acquired, often using cryoprotected (e.g., with 20% to 30% glycerol) crystals frozen to 100 K, e.g., using a beamline at a synchrotron facility or a rotating anode as an X-ray source. Then, the phase problem is solved by a generally known method, e.g., multiwavelength anomalous diffraction (MAD), multiple isomorphous replacement (MIR), single wavelength anomalous diffraction (SAD), or molecular replacement (MR). The sub-structure may be solved using SHELXD (Schneider and Sheldrick, 2002, Acta Crystallogr. D. Biol. Crystallogr. (Pt 10 Pt 2), 1772-1779), phases calculated with SHARP (Vonrhein et al., 2006, Methods Mol. Biol. 364:215-30), and improved with solvent flattening and non-crystallographic symmetry averaging, e.g., with RESOLVE (Terwilliger, 2000, Acta Cryst. D. Biol. Crystallogr. 56:965-972). Model autobuilding can be done, e.g., with ARP/wARP (Perrakis et al., 1999, Nat. Struct. Biol. 6:458-63) and refinement with, e.g., REFMAC (Murshudov, 1997, Acta Crystallogr. D. Biol. Crystallogr. 53: 240-255). The skilled person can use the structure coordinates (Figure 18) as input for secondary analysis, including the determination of electrostatic surface potential (see Figure 13), which aids in the determination of side groups in test compounds, which are likely to interact with a surface area of the PA of a given electrostatic potential, preferably in the active site. In order to use the structure coordinates generated for the PA polypeptide fragment it is necessary to convert the structure coordinates into a three-dimensional shape. This is achieved through the use of commercially available software that is capable of generating three-dimensional graphical representations of molecules or portions thereof from a set of structure coordinates. An example for such a computer program is MODELER (Sali and Blundell, 1993, J. Mol. Biol. 234:779-815 as implemented in the Insight II Homology software package (Insight II (97.0), Molecular Simulations Incorporated, San Diego, CA)). Such a three-dimensional graphical representations can be use with suitable programs including (i) Gaussian 92, revision C (Frisch, Gaussian, Incorporated, Pittsburgh, PA), (ii) AMBER, version 4.0 (Kollman, University of California, San Francisco, CA), (iii) QUANTA/CHARMM (Molecular

Simulations Incorporated, San Diego, CA), (iv) OPLS-AA (Jorgensen, 1998, Encyclopedia of Computational Chemistry, Schleyer, Ed., Wiley, New York, Vol. 3, pp. 1986-1989), and (v) Insight II/Discover (Biosysm Technologies Incorporated, San Diego, CA) to generate graphic representations of, e.g. electrostatic potential. Similarly, the structural information can be combined with information on the conservation of residues as depicted in Figure 11 at the various amino acid positions (see Figure 12) to highlight those residues at the surface of the PA and/or in the active site, which are particularly conserved between different virus isolates and, consequently, are likely to be also present in mutants of thoses viruses or other isolates. This suitable in the the skilled person is able to derive information on the relevance of the residues Furthermore, the structure coordinates (Figure 18) of the Influenza A virus PA fragment PA-Nter provided by the present invention are useful for the structure determination of PA polypeptides of other viruses from the Orthomyxoviridae family, or PA polypeptide variants that have amino acid substitutions, deletions, and/or insertions using the method of molecular replacement.

In a preferred embodiment of the polypeptide fragment according to the invention, the PA subunit is from Influenza A, B, or C virus or is a variant thereof, preferably from Influenza A virus or a variant thereof. Preferably, the amino terminal PA fragment comprised within the polypeptide fragment according to the present invention corresponds to, preferably essentially consists or consists of at least amino acids 1 to 196, preferably amino acids 1 to 209, preferably amino acids 1 to 213 of the PA subunit of the RNA-dependent RNA polymerase of Influenza A virus or variants thereof, i.e., amino acid residues 1 to 196, 1 to 209, or 1 to 213 of the amino acid sequence as set forth in SEQ ID NO: 2.

In a preferred embodiment, the polypeptide fragment according to the present invention is purified to an extent to be suitable for crystallization, preferably it is 85% to 100%, more preferably 90% to 100%, most preferably 95% to 100% pure.

In another embodiment, the polypeptide fragment according to the invention is capable of binding to divalent cations. Preferably, the polypeptide fragment according to the present invention is bound to one or more divalent cation(s), preferably it is bound to two divalent cations. In this context, the divalent cation is preferably selected form the group consisting of manganese, cobalt, calcium, magnesium, and zinc, and is more preferably manganese or cobalt, most preferably manganese. Thus, in a preferred embodiment, the polypeptide of the present invention is present in complex with two manganese cations. In a preferred embodiment, the divalent cations are coordinated by amino acids corresponding to amino

acids Glu80 and Asp108 (first cation) and amino acids corresponding to amino acids His41, Asp108, and Glu119 (second cation) as set forth in SEQ ID NO: 2.

In a preferred embodiment of the polypeptide fragment according to the present invention, (i) the N-terminus is identical to or corresponds to amino acid position 15 or lower, e.g., at position 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1, and the C-terminus is identical to or corresponds to an amino acid at a position selected from positions 186 to 220, e.g., 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, or 220 of the amino acid sequence of the PA subunit according to SEQ ID NO: 2; preferably the N-terminus is identical to or corresponds to amino acid position 15 or lower, e.g., at position 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 and the C-terminus is identical to or corresponds to an amino acid at a position selected from 196 to 220 of the amino acid sequence of the PA subunit according to SEQ ID NO: 2; more preferably the N-terminus is identical to or corresponds to amino acid position 15 or lower, e.g., at position 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1, and the C-terminus is identical to or corresponds to an amino acid at a position selected from 196 to 209 of the amino acid sequence of the PA subunit according to SEQ ID NO: 2, (ii) the N-terminus is identical to or corresponds to amino acid position 15 or lower, e.g., amino acid position 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1, and the C-terminus is identical to or corresponds to an amino acid at a position selected from positions 185 to 217, e.g., 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, or 217 of the amino acid sequence of the PA subunit according to SEQ ID NO: 4; preferably the N-terminus is identical to or corresponds to amino acid position 15 or lower, e.g., amino acid position 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1, and the C-terminus is identical to or corresponds to an amino acid at a position selected from positions 195 to 217 of the amino acid sequence of the PA subunit according to SEQ ID NO: 4; more preferably the N-terminus is identical to or corresponds to amino acid position 15 or lower, e.g., amino acid position 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1, and the C-terminus is identical to or corresponds to an amino acid at a position 195 to 206 of the amino acid sequence according to SEQ ID NO: 4, or (iii) the N-terminus is identical to or corresponds to amino acid position 15 or lower, e.g., amino acid position 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1, and the C-terminus is identical to or corresponds to an amino acid at a position selected from positions 168 to 200, e.g., amino acid position 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, or 200 of the

amino acid sequence of the PA subunit according to SEQ ID NO: 6; preferably the N-terminus is identical to or corresponds to amino acid position 15 or lower, e.g., amino acid position 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1, and the C-terminus is identical to or corresponds to an amino acid at a position selected from positions 178 to 200 of the amino acid sequence according to SEQ ID NO: 6, and variants thereof, which retain the endonuclease activity; more preferably the N-terminus is identical to or corresponds to amino acid position 15 or lower, e.g., amino acid position 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1, and the C-terminus is identical to or corresponds to an amino acid at a position selected from positions 178 to 189 of the amino acid sequence according to SEQ ID NO: 6; and in each case variants of the amino acid sequence according to SEQ ID NO: 2, 4 or 6, which retain endonuclease activity.

In another embodiment said polypeptide fragment has or corresponds to an amino acid sequence selected from the group of amino acid sequences consisting of amino acids 5 to 196, 10 to 196, 15 to 196, 20 to 196, 5 to 209, 10 to 209, 15 to 209, 20 to 209 of the amino acid sequence set forth in SEQ ID NO: 2 and variants thereof, which retain the endonucleolytic activity. In another embodiment said PA polypeptide fragment has or corresponds to amino acids selected from the group of amino acid sequences consisting of amino acids 5 to 195, 10 to 195, 15 to 195, 20 to 195, 5 to 206, 10 to 206, 15 to 206, 20 to 206 of the amino acid sequence set forth in SEQ ID NO: 4 and variants thereof, which retain the endonucleolytic activity. In another embodiment said PA polypeptide fragment has or corresponds to amino acids selected from the group of amino acid sequences consisting of amino acids 5 to 178, 10 to 178, 15 to 178, 20 to 178, 5 to 189, 10 to 189, 15 to 189, 20 to 189 of the amino acid sequence set forth in SEQ ID NO: 6 and variants thereof, which retain the endonucleolytic activity. In preferred embodiments, said polypeptide fragments comprise amino acid substitutions, insertions, or deletions, preferably naturally occurring mutations as set forth above.

In another preferred embodiment, the polypeptide fragment according to the present invention consists of amino acids 1 to 209 of the amino acid sequence set forth in SEQ ID NO: 2 and has the structure defined by the structure coordinates as shown in Figure 18.

In another embodiment, the polypeptide fragment according to the present invention has a crystalline form, preferably with space group $P4_32_12$, with unit cell dimensions of preferably a = b = 6.71 ± 0.2 nm, c = 30.29 nm ± 0.4 nm. In another embodiment, the crystals according to the invention are hexagonal plates with preferred unit cell dimensions of a = b = 6.79 nm, c = 49.4 nm, $\alpha = \beta = 90°$, and $\gamma = 120°$ having preferably a trigonal or hexagonal

space group. Preferably, the crystal of the polypeptide fragment diffracts X-rays to a resolution of 2.8 Å or higher, preferably 2.6 Å or higher, more preferably 2.5 Å or higher, even more preferably 2.4 Å or higher, most preferably 2.1 Å or higher.

It is another aspect of the present invention to provide an isolated polynucleotide coding for the above-mentioned PA polypeptide fragments and variants thereof. The molecular biology methods applied for obtaining such isolated nucleotide fragments are generally known to the person skilled in the art (for standard molecular biology methods see Sambrook et al., Eds., "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), which is incorporated herein by reference). For example, RNA can be isolated from Influenza virus infected cells and cDNA generated applying reverse transcription polymerase chain reaction (RT-PCR) using either random primers (e.g., random hexamers of decamers) or primers specific for the generation of the fragments of interest. The fragments of interest can then be amplified by standard PCR using fragment specific primers.

In a preferred embodiment the isolated polynucleotide coding for the preferred embodiments of the PA polypeptide fragments are derived from SEQ ID NO: 1 (Influenza A), 3 (Influenza B), or 6 (Influenza C). In this context, "derived" refers to the fact that SEQ ID NO: 1, 2, and 3 encode the full-length PA polypeptides and, thus, polynucleotides coding for preferred PA polypeptide fragments may comprise deletions at the 3'- and/or 5'-ends of the polynucleotide as required by the respectively encoded PA polypeptide fragment.

In one embodiment, the present invention relates to a recombinant vector comprising said isolated polynucleotide. The person skilled in the art is well aware of techniques used for the incorporation of polynucleotide sequences of interest into vectors (also see Sambrook et al., 1989, *supra*). Such vectors include any vectors known to the skilled person including plasmid vectors, cosmid vectors, phage vectors such as lambda phage, viral vectors such as adenoviral or baculoviral vectors, or artificial chromosome vectors such as bacterial artificial chromosomes (BAC), yeast artificial chromosomes (YAC), or P1 artificial chromosomes (PAC). Said vectors may be expression vectors suitable for prokaryotic or eukaryotic expression. Said plasmids may include an origin of replication (ori), a multiple cloning site, and regulatory sequences such as promoter (constitutive or inducible), transcription initiation site, ribosomal binding site, transcription termination site, polyadenylation signal, and selection marker such as antibiotic resistance or auxotrophic marker based on complementation of a mutation or deletion. In one embodiment the polynucleotide sequence of interest is operably linked to the regulatory sequences.

In another embodiment, said vector includes nucleotide sequences coding for epitope-, peptide-, or protein-tags that facilitate purification of polypeptide fragments of interest. Such epitope-, peptide-, or protein-tags include, but are not limited to, hemagglutinin- (HA-), FLAG-, myc-tag, poly-His-tag, glutathione-S-transferase- (GST-), maltose-binding-protein- (MBP-), NusA-, and thioredoxin-tag, or fluorescent protein-tags such as (enhanced) green fluorescent protein ((E)GFP), (enhanced) yellow fluorescent protein ((E)YFP), red fluorescent protein (RFP) derived from Discosoma species (DsRed) or monomeric (mRFP), cyan fluorescence protein (CFP), and the like. In a preferred embodiment, the epitope-, peptide-, or protein-tags can be cleaved off the polypeptide fragment of interest, for example, using a protease such as thrombin, Factor Xa, PreScission, TEV protease, and the like. Preferably, the tag can be cleaved of with a TEV protease. The recognition sites for such proteases are well known to the person skilled in the art. For example, the seven amino acid consensus sequence of the TEV protease recognition site is Glu-X-X-Tyr-X-Gln-Gly/Ser, wherein X may be any amino acid and is in the context of the present invention preferably Glu-Asn-Leu-Tyr-Phe-Gln-Gly (SEQ ID NO: 21). In another embodiment, the vector includes functional sequences that lead to secretion of the polypeptide fragment of interest into the culture medium of the recombinant host cells or into the periplasmic space of bacteria. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

In another aspect, the present invention provides a recombinant host cell comprising said isolated polynucleotide or said recombinant vector. The recombinant host cells may be prokaryotic cells such as archea and bacterial cells or eukaryotic cells such as yeast, plant, insect, or mammalian cells. In a preferred embodiment the host cell is a bacterial cell such as an *E. coli* cell. The person skilled in the art is well aware of methods for introducing said isolated polynucleotide or said recombinant vector into said host cell. For example, bacterial cells can be readily transformed using, for example, chemical transformation, e.g., the calcium chloride method, or electroporation. Yeast cells may be transformed, for example, using the lithium acetate transformation method or electroporation. Other eukaryotic cells can be transfected, for example, using commercially available liposome-based transfection kits such as Lipofectamine™ (Invitrogen), commercially available lipid-based transfection kits

such as Fugene (Roche Diagnostics), polyethylene glycol-based transfection, calcium phosphate precipitation, gene gun (biolistic), electroporation, or viral infection. In a preferred embodiment of the invention, the recombinant host cell expresses the polynucleotide fragment of interest. In an even more preferred embodiment, said expression leads to soluble polypeptide fragments of the invention. These polypeptide fragments may be purified using protein purification methods well known to the person skilled in the art, optionally taking advantage of the above-mentioned epitope-, peptide-, or protein-tags.

In another aspect, the present invention relates to a method for identifying compounds which modulate the endonuclease activity of the PA subunit of a viral RNA-dependent RNA polymerase from the Orthomyxoviridae family or a variant thereof, comprising the steps of

(a) constructing a computer model of the active site defined by the structure coordinates of the polypeptide fragment according to the present invention shown in Figure 18;

(b) selecting a potential activity modulating compound by a method selected from the group consisting of:

(i) assembling molecular fragments into said compound,

(ii) selecting a compound from a small molecule database, and

(iii) *de novo* ligand design of said compound;

(c) employing computational means to perform a fitting program operation between computer models of the said compound and the said active site in order to provide an energy-minimized configuration of the said compound in the active site; and

(d) evaluating the results of said fitting operation to quantify the association between the said compound and the active site model, whereby evaluating the ability of said compound to associate with the said active site.

Preferably, the modulating compound binds to the endonucleolytically active site within the PA subunit or variant thereof. The modulating compound may increase or decrease, preferably decrease said endonucleolytic activity.

In a preferred embodiment of this aspect of the present invention, the compound that modulates the endonuclease activity of the PA subunit or a variant thereof decreases said activity, more preferably said compound inhibits said activity. Preferably, the compound decreases the endonucleolytic activity of the PA subunit or a variant thereof by 50%, more preferably by 60%, even more preferably by 70%, even more preferably by 80%, even more preferably by 90%, and most preferably by 100% compared to the endonucleolytic activity of the PA subunit or a variant thereof without said compound but with otherwise the same reaction conditions, i.e., buffer conditions, reaction time and temperature. It is particularly

preferred that the compound specifically decreases or inhibits the endonucleolytic activity of the PA subunit or a variant thereof but does not decrease or inhibit the endonucleolytic activity of other endonucleases, in particular of mammalian endonucleases, to the same extent, preferably not at all.

For the first time, the present invention permits the use of molecular design techniques to identify, select, or design of compounds potentially modulating the endonucleolytic activity of the PA subunit or variants thereof, based on the structure coordinates of the endonucleolytically active site according to Figure 18. Such a predictive model is valuable in light of the higher costs associated with the preparation and testing of the many diverse compounds that may possibly modulate the endonucleolytic activity. In order to use the structure coordinates generated for the PA polypeptide fragment it is necessary to convert the structure coordinates into a three-dimensional shape. This is achieved through the use of commercially available software that is capable of generating three-dimensional graphical representations of molecules or portions thereof from a set of structure coordinates. An example for such a computer program is MODELER (Sali and Blundell, 1993, J. Mol. Biol. 234:779-815 as implemented in the Insight II Homology software package (Insight II (97.0), Molecular Simulations Incorporated, San Diego, CA)).

One skilled in the art may use several methods to screen chemical entities or fragments for their ability to modulate the endonucleolytic activity of the PA subunit or PA polypeptide variants. This process may begin by a visual inspection of, for example, a three-dimensional computer model of the endonucleolytically active site of PA based on the structural coordinates according to Figure 18. Selected fragments or chemical compounds may then be positioned in a variety of orientations or docked within the active site. Docking may be accomplished using software such as Cerius, Quanta, and Sybyl (Tripos Associates, St. Louis, MO), followed by energy minimization and molecular dynamics with standard molecular dynamics force fields such as OPLS-AA, CHARMM, and AMBER. Additional specialized computer programs that may assist the person skilled in the art in the process of selecting suitable compounds or fragments include, for example, (i) AUTODOCK (Goodsell et al., 1990, Proteins: Struct., Funct., Genet. 8: 195-202; AUTODOCK is available from The Scripps Research Institute, La Jolla, CA) and (ii) DOCK (Kuntz et al., 1982, J. Mol. Biol. 161:269-288; DOCK is available from the University of California, San Francisco, CA).

Once suitable compounds or fragments have been selected, they can be designed or assembled into a single compound or complex. This manual model building is performed using software such as Quanta or Sybyl. Useful programs aiding the skilled person in

connecting individual compounds or fragments include, for example, (i) CAVEAT (Bartlett et al., 1989, in Molecular Recognition in Chemical and Biological Problems, Special Publication, Royal Chem. Soc. 78:182-196; Lauri and Bartlett, 1994, J. Comp. Aid. Mol. Des. 8:51-66; CAVEAT is available from the University of California, Berkley, CA), (ii) 3D Database systems such as ISIS (MDL Information Systems, San Leandro, CA; reviewed in Martin, 1992, J. Med. Chem. 35:2145-2154), and (iii) HOOK (Eisen et al., 1994, Proteins: Struct., Funct., Genet. 19:199-221; HOOK is available from Molecular Simulations Incorporated, San Diego, CA).

Another approach enabled by this invention, is the computational screening of small molecule databases for compounds that can bind in whole or part to the endonucleolytically active site of the PA subunit or active sites of PA polypeptide variants. In this screening, the quality of fit of such compounds to the active site may be judged either by shape complementarity or by estimated interaction energy (Meng et al., 1992, J. Comp. Chem. 13:505-524).

Alternatively, a potential modulator for the endonucleolytic activity of the PA subunit or polypeptide variant thereof, preferably an inhibitor of the endonucleolytic activity, may be designed *de novo* on the basis of the 3D structure of the PA polypeptide fragment according to Figure 18. There are various *de novo* ligand design methods available to the person skilled in the art. Such methods include (i) LUDI (Bohm, 1992, J. Comp. Aid. Mol. Des. 6:61-78; LUDI is available from Molecular Simulations Incorporated, San Diego, CA), (ii) LEGEND (Nishibata and Itai, Tetrahedron 47:8985-8990; LEGEND is available from Molecular Simulations Incorporated, San Diego, CA), (iii) LeapFrog (available from Tripos Associates, St. Louis, MO), (iv) SPROUT (Gillet et al., 1993, J. Comp. Aid. Mol. Des. 7:127-153; SPROUT is available from the University of Leeds, UK), (v) GROUPBUILD (Rotstein and Murcko, 1993, J. Med. Chem. 36:1700-1710), and (vi) GROW (Moon and Howe, 1991, Proteins 11:314-328).

In addition, several molecular modeling techniques (hereby incorporated by reference) that may support the person skilled in the art in *de novo* design and modeling of potential modulators and/or inhibitors of the endonucleolytically active site, preferably binding partners of the endonucleolytically active site, have been described and include, for example, Cohen et al., 1990, J. Med. Chem. 33:883-894; Navia and Murcko, 1992, Curr. Opin. Struct. Biol. 2:202-210; Balbes et al., 1994, Reviews in Computational Chemistry, Vol. 5, Lipkowitz and Boyd, Eds., VCH, New York, pp. 37-380; Guida, 1994, Curr. Opin. Struct. Biol. 4:777-781.

A molecule designed or selected as binding to the endonucleolytically active site of the PA subunit or variants thereof may be further computationally optimized so that in its bound state it preferably lacks repulsive electrostatic interaction with the target region. Such non-complementary (e.g., electrostatic) interactions include repulsive charge-charge, dipole-dipole and charge-dipole interactions. Specifically, the sum of all electrostatic interactions between the binding compound and the binding pocket in a bound state, preferably make a neutral or favorable contribution to the enthalpy of binding. Specific computer programs that can evaluate a compound deformation energy and electrostatic interaction are available in the art. Examples of suitable programs include (i) Gaussian 92, revision C (Frisch, Gaussian, Incorporated, Pittsburgh, PA), (ii) AMBER, version 4.0 (Kollman, University of California, San Francisco, CA), (iii) QUANTA/CHARMM (Molecular Simulations Incorporated, San Diego, CA), (iv) OPLS-AA (Jorgensen, 1998, Encyclopedia of Computational Chemistry, Schleyer, Ed., Wiley, New York, Vol. 3, pp. 1986-1989), and (v) Insight II/Discover (Biosysm Technologies Incorporated, San Diego, CA). These programs may be implemented, for instance, using a Silicon Graphics workstation, IRIS 4D/35 or IBM RISC/6000 workstation model 550. Other hardware systems and software packages are known to those skilled in the art.

Once a molecule of interest has been selected or designed, as described above, substitutions may then be made in some of its atoms or side groups in order to improve or modify its binding properties. Generally, initial substitutions are conservative, i.e., the replacement group will approximate the same size, shape, hydrophobicity and charge as the original group. It should, of course, be understood that components known in the art to alter conformation should be avoided. Such substituted chemical compounds may then be analyzed for efficiency of fit to the endonucleolytically active site of the PA subunit or variant thereof by the same computer methods described in detail above.

In one embodiment of the above-described method of the invention, the endonucleolytically active site of the PA subunit or variant thereof comprises amino acids corresponding to amino acids Asp108, Ile120, and Lys134 of the PA subunit according to SEQ ID NO: 2. In another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile120, Lys134, and His41 according to SEQ ID NO: 2. In another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile120, Lys134, and Glu80 according to SEQ ID NO: 2. In another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile120, Lys134, and Glu119 according to SEQ ID NO: 2. In another embodiment, said active site comprises amino

acids corresponding to amino acids Asp108, Ile120, Lys134, His41, Glu80, and Glu119 according to SEQ ID NO: 2. In yet another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Tyr24 according to SEQ ID NO: 2. In yet another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Arg84 according to SEQ ID NO: 2. In yet another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Leu106 according to SEQ ID NO: 2. In yet another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Tyr130 according to SEQ ID NO: 2. In yet another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Glu133 according to SEQ ID NO: 2. In yet another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Lys137 according to SEQ ID NO: 2. In another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile 120, Lys134, His41, Glu80, Glu119, Tyr24, Arg84, and Leu106 according to SEQ ID NO: 2. In another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile 120, Lys134, His41, Glu80, Glu119, Tyr130, Glu133, and Lys137 according to SEQ ID NO: 2. In another embodiment, said active site comprises amino acids corresponding to amino acids Asp108, Ile 120, Lys134, His41, Glu80, Glu119, Tyr24, Arg84, Leu106, Tyr130, Glu133, and Lys137 according to SEQ ID NO: 2.

In a further aspect of the above-described method of the invention, the endonucleolytically active site of the PA subunit or a variant thereof is defined by the structure coordinates of the PA SEQ ID NO: 2 amino acids Asp108, Ile120, and Lys134 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, and His41 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, and Glu80 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, and Glu119 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, and Glu119 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Tyr24 according to

Figure 18. In yet another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Arg84 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Leu106 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Tyr130 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Glu133 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Lys137 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, Tyr24, Arg84, and Leu106 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, Tyr130, Glu133, and Lys137 according to Figure 18. In another embodiment, said active site is defined by the structure coordinates of PA SEQ ID NO: 2 amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, Tyr24, Arg84, Leu106, Tyr130, Glu133, and Lys137 according to Figure 18.

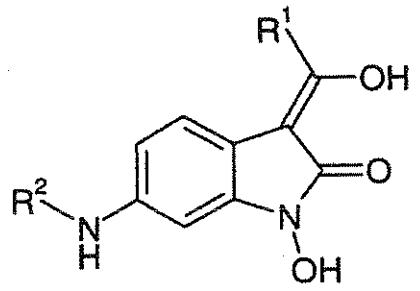
In one aspect, the present invention provides a method for computational screening according to the above-described method for compounds able to modulate and/or associate with an endonucleolytically active site that is a variant to the endonucleolytically active site of the PA subunit according to Figure 18. In one embodiment, said variant of said active site has a root mean square deviation from the backbone atoms of amino acids Asp108, Ile120, and Lys134, of amino acids Asp108, Ile120, Lys134, and His41, of amino acids Asp108, Ile120, Lys134, and Glu80, of amino acids Asp108, Ile120, Lys134, and Glu119, of amino acids Asp108, Ile120, Lys134, His41, Glu80, and Glu119, of amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Tyr24, of amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Arg84, of amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Leu106, of amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Tyr130, of amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Glu133, of amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, and Lys137, of amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, Tyr24, Arg84, and Leu106, of amino acids Asp108, Ile120, Lys134, His41, Glu80, Glu119, Tyr130, Glu133, and Lys137, of amino acids Asp108, Ile120,

Lys134, His41, Glu80, Glu119, Tyr24, Arg84, Leu106, Tyr130, Glu133, and Lys137 according to Figure 18 of not more than 3 Å. In another embodiment, the said root mean square deviation is not more than 2.5 Å. In another embodiment, the said root mean square deviation is not more than 2 Å. In another embodiment, the said root mean square deviation is not more than 1.5 Å. In another embodiment, the said root mean square deviation is not more than 1 Å. In another embodiment, the said root mean square deviation is not more than 0.5 Å.

If computer modeling according to the methods described hereinabove indicates binding of a compound to the active site of the PA subunit or a variant thereof, said compound may be synthesized and optionally said compound or a pharmaceutically acceptable salt thereof may be formulated with one or more pharmaceutically acceptable excipient(s) and/or carrier(s). Thus, the above-described method may comprise the further step of (e) synthesizing said compound and optionally formulating said compound or a pharmaceutically acceptable salt thereof with one or more pharmaceutically acceptable excipient(s) and/or carrier(s). Optionally, the ability of said compound or of a pharmaceutically acceptable salt thereof or of a formulation thereof to modulate, preferably decrease, preferably inhibit the endonucleolytic activity of the PA subunit or variant thereof may be tested *in vitro* or *in vivo* comprising the further step of (f) contacting said compound with the PA polypeptide fragment or variant thereof or the recombinant host cell of the invention and to determine the ability of said compound to (i) bind to the active site and/or (ii) to modulate, decrease, or inhibit the endonucleolytic activity of the PA subunit polypeptide fragment or variant thereof. The quality of fit of such compounds to the active site may be judged either by shape complementarity or by estimated interaction energy (Meng et al., 1992, J. Comp. Chem. 13:505-524). Methods for synthesizing said compounds are well known to the person skilled in the art or such compounds may be commercially available.

It is another aspect of the invention to provide a compound identifiable by the above-described method, wherein said compound is able to modulate the endonuclease activity of the PA subunit or variant thereof. In another aspect, the present invention refers to a compound identifiable by the above-described method, wherein said compound is able to decrease, preferably inhibit the endonuclease activity of the PA subunit or variant thereof, e.g., the PA subunit polypeptide or variant thereof according to the present invention. Compounds of the present invention can be any agents including, but not restricted to, peptides, peptoids, polypeptides, proteins (including antibodies), lipids, metals, nucleotides, nucleosides, nucleic acids, small organic or inorganic molecules, chemical compounds, elements, saccharides, isotopes, carbohydrates, imaging agents, lipoproteins, glycoproteins,

enzymes, analytical probes, polyamines, and combinations and derivatives thereof. The term "small molecules" refers to molecules that have a molecular weight between 50 and about 2,500 Daltons, preferably in the range of 200-800 Daltons. In addition, a test compound according to the present invention may optionally comprise a detectable label. Such labels include, but are not limited to, enzymatic labels, radioisotope or radioactive compounds or elements, fluorescent compounds or metals, chemiluminescent compounds and bioluminescent compounds. In a preferred embodiment of the compound according to the present invention, the compound is not a 4-substituted 2-dioxobutanoic acid, a 4-substituted 4-dioxobutanoic acid, a 4-substituted 2,4-dioxobutanoic acid, a pyrazine-2,6-dione or a substituted pyrazine-2,6-dione such as flutimide, an N-hydroxamic acid, or an N-hydroxymide. In particular, the compound according to the present invention is not a compound according to Formula I:



R^1 = Me, Et, Ph
R^2 = acyl, carbamoyl, sulfonyl.

Formula I

In a further aspect, the present invention provides a method for identifying compounds which bind to the endonucleolytically active site, preferably modulate, more preferably decrease, most preferably inhibit the endonuclease activity of the PA subunit or polypeptide variants thereof, comprising the steps of (i) contacting the PA polypeptide fragment according to the present invention or a recombinant host cell according to the present invention with a test compound and (ii) analyzing the ability of said test compound to bind to the endonucleolytically active site, to modulate, to decrease, or to inhibit the endonuclease activity of said PA subunit polypeptide fragment.

In one embodiment, the interaction between the PA polypeptide fragment or variant thereof and a test compound may be analyzed in form of a pull down assay. For example, the PA polypeptide fragment may be purified and may be immobilized on beads. In one embodiment, the PA polypeptide fragment immobilized on beads may be contacted, for example, with (i) another purified protein, polypeptide fragment, or peptide, (ii) a mixture of proteins, polypeptide fragments, or peptides, or (iii) a cell or tissue extract, and binding of proteins, polypeptide fragments, or peptides may be verified by polyacrylamide gel

electrophoresis in combination with coomassie staining or Western blotting. Unknown binding partners may be identified by mass spectrometric analysis.

In another embodiment, the interaction between the PA polypeptide fragment or variant thereof and a test compound may be analyzed in form of an enzyme-linked immunosorbent assay (ELISA)-based experiment. In one embodiment, the PA polypeptide fragment or variant thereof according to the invention may be immobilized on the surface of an ELISA plate and contacted with the test compound. Binding of the test compound may be verified, for example, for proteins, polypeptides, peptides, and epitope-tagged compounds by antibodies specific for the test compound or the epitope-tag. These antibodies might be directly coupled to an enzyme or detected with a secondary antibody coupled to said enzyme that – in combination with the appropriate substrates – carries out chemiluminescent reactions (e.g., horseradish peroxidase) or colorimetric reactions (e.g., alkaline phosphatase). In another embodiment, binding of compounds that cannot be detected by antibodies might be verified by labels directly coupled to the test compounds. Such labels may include enzymatic labels, radioisotope or radioactive compounds or elements, fluorescent compounds or metals, chemiluminescent compounds and bioluminescent compounds. In another embodiment, the test compounds might be immobilized on the ELISA plate and contacted with the PA polypeptide fragment or variants thereof according to the invention. Binding of said polypeptide may be verified by a PA polypeptide fragment specific antibody and chemiluminescence or colorimetric reactions as described above.

In a further embodiment, purified PA polypeptide fragments may be incubated with a peptide array and binding of the PA polypeptide fragments to specific peptide spots corresponding to a specific peptide sequence may be analyzed, for example, by PA polypeptide specific antibodies, antibodies that are directed against an epitope-tag fused to the PA polypeptide fragment, or by a fluorescence signal emitted by a fluorescent tag coupled to the PA polypeptide fragment.

In another embodiment, the recombinant host cell according to the present invention is contacted with a test compound. This may be achieved by co-expression of test proteins or polypeptides and verification of interaction, for example, by fluorescence resonance energy transfer (FRET) or co-immunoprecipitation. In another embodiment, directly labeled test compounds may be added to the medium of the recombinant host cells. The potential of the test compound to penetrate membranes and bind to the PA polypeptide fragment may be, for example, verified by immunoprecipitation of said polypeptide and verification of the presence of the label.

In another embodiment, the ability of the test compound to modulate, preferably decrease, more preferably inhibit the endonucleolytic activity of the PA subunit polypeptide fragment or variant thereof is assessed. For example, the purified PA subunit polypeptide fragment and a substrate thereof such as panhandle RNA or single stranded DNA are contacted in presence or absence of varying amounts of the test compound and incubated for a certain period of time, for example, for 5, 10, 15, 20, 30, 40, 60, or 90 minutes. The reaction conditions are chosen such that the PA subunit polypeptide is endonucleolytically active without the test compound. The substrate is then analyzed for degradation/endonucleolytic cleavage, for example, by gel electrophoresis. Alternatively, such a test may comprise a labeled substrate molecule which provides a signal when the substrate molecule is endonucleolytically cleaved but does not provide a signal if it is intact. For example, the substrate polynucleotide chain may be labeled with fluorescent reporter molecule and a fluorescence quencher such that the fluorescent reporter is quenched as long as the substrate polynucleotide chain is intact. In case the substrate polynucleotide chain is cleaved, the fluorescent reporter and the quencher are separated, thus, the fluorescent reporter emits a signal which may be detected, for example, by an ELISA reader. This experimental setting may be applied in a multi-well plate format and is suitable for high throughput screening of compounds regarding their ability to modulate, decrease, or inhibit the endonuclease activity of the PA subunit polypeptide fragment or variants thereof.

In a preferred embodiment, the above-described method for identifying compounds which associate with the endonucleolytically active site, modulate, decrease, or inhibit the endonucleolytic activity of the PA subunit polypeptide fragment or variant thereof is performed in a high-throughput setting. In a preferred embodiment, said method is carried out in a multi-well microtiter plate as described above using PA polypeptide fragments or variants thereof according to the present invention and labeled test compounds.

In a preferred embodiment, the test compounds are derived from libraries of synthetic or natural compounds. For instance, synthetic compound libraries are commercially available from Maybridge Chemical Co. (Trevillet, Cornwall, UK), ChemBridge Corporation (San Diego, CA), or Aldrich (Milwaukee, WI). A natural compound library is, for example, available from TimTec LLC (Newark, DE). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts can be used. Additionally, test compounds can be synthetically produced using combinatorial chemistry either as individual compounds or as mixtures.

In another embodiment, the inhibitory effect of the identified compound on the Influenza virus life cycle may be tested in an *in vivo* setting. A cell line that is susceptible for Influenza virus infection such as 293T human embryonic kidney cells, Madin-Darby canine kidney cells, or chicken embryo fibroblasts may be infected with Influenza virus in presence or absence of the identified compound. In a preferred embodiment, the identified compound may be added to the culture medium of the cells in various concentrations. Viral plaque formation may be used as read out for the infectious capacity of the Influenza virus and may be compared between cells that have been treated with the identified compound and cells that have not been treated.

In a further embodiment of the invention, the test compound applied in any of the above described methods is a small molecule. In a preferred embodiment, said small molecule is derived from a library, e.g., a small molecule inhibitor library. In another embodiment, said test compound is a peptide or protein. In a preferred embodiment, said peptide or protein is derived from a peptide or protein library.

In another embodiment of the above-described methods for computational as well as *in vitro* identification of compounds that associate with the endonucleolytically active site, modulate, decrease, or inhibit the endonucleolytic activity of the PA subunit polypeptide fragment or variant thereof according to the present invention, said methods further comprise the step of formulating the identifiable compound or a pharmaceutically acceptable salt thereof with one or more pharmaceutically acceptable excipient(s) and/or carrier(s). In another aspect the present invention provides a pharmaceutical composition producible according to the afore-mentioned method. A compound according to the present invention can be administered alone but, in human therapy, will generally be administered in admixture with a suitable pharmaceutical excipient, diluent, or carrier selected with regard to the intended route of administration and standard pharmaceutical practice (see hereinafter).

In the aspect of computational modeling or screening of a binding partner for the endonucleolytically active site, a modulator, and/or inhibitor of the endonucleolytic activity of the PA subunit polypeptide fragment or variant thereof according to the present invention, it may be possible to introduce into the molecule of interest, chemical moieties that may be beneficial for a molecule that is to be administered as a pharmaceutical. For example, it may be possible to introduce into or omit from the molecule of interest, chemical moieties that may not directly affect binding of the molecule to the target area but which contribute, for example, to the overall solubility of the molecule in a pharmaceutically acceptable carrier, the bioavailability of the molecule and/or the toxicity of the molecule. Considerations and

methods for optimizing the pharmacology of the molecules of interest can be found, for example, in "Goodman and Gilman's The Pharmacological Basis of Therapeutics", 8th Edition, Goodman, Gilman, Rall, Nies, & Taylor, Eds., Pergamon Press (1985); Jorgensen & Duffy, 2000, Bioorg. Med. Chem. Lett. 10:1155-1158. Furthermore, the computer program "Qik Prop" can be used to provide rapid predictions for physically significant descriptions and pharmaceutically-relevant properties of an organic molecule of interest. A 'Rule of Five' probability scheme can be used to estimate oral absorption of the newly synthesized compounds (Lipinski et al., 1997, Adv. Drug Deliv. Rev. 23:3-25). Programs suitable for pharmacophore selection and design include (i) DISCO (Abbot Laboratories, Abbot Park, IL), (ii) Catalyst (Bio-CAD Corp., Mountain View, CA), and (iii) Chem DBS-3D (Chemical Design Ltd., Oxford, UK).

The pharmaceutical composition contemplated by the present invention may be formulated in various ways well known to one of skill in the art. For example, the pharmaceutical composition of the present invention may be in solid form such as in the form of tablets, pills, capsules (including soft gel capsules), cachets, lozenges, ovules, powder, granules, or suppositories, or in liquid form such as in the form of elixirs, solutions, emulsions, or suspensions.

Solid administration forms may contain excipients such as microcrystalline cellulose, lactose, sodium citrate, calcium carbonate, dibasic calcium phosphate, glycine, and starch (preferably corn, potato, or tapioca starch), disintegrants such as sodium starch glycolate, croscarmellose sodium, and certain complex silicates, and granulation binders such as polyvinylpyrrolidone, hydroxypropylmethyl cellulose (HPMC), hydroxypropylcellulose (HPC), sucrose, gelatin, and acacia. Additionally, lubricating agents such as magnesium stearate, stearic acid, glyceryl behenate, and talc may be included. Solid compositions of a similar type may also be employed as fillers in gelatin capsules. Preferred excipients in this regard include lactose, starch, a cellulose, milk sugar, or high molecular weight polyethylene glycols.

For aqueous suspensions, solutions, elixirs, and emulsions suitable for oral administration the compound may be combined with various sweetening or flavoring agents, coloring matter or dyes, with emulsifying and/or suspending agents and with diluents such as water, ethanol, propylene glycol, and glycerin, and combinations thereof.

The pharmaceutical composition of the present invention may contain release rate modifiers including, for example, hydroxypropylmethyl cellulose, methyl cellulose, sodium carboxymethylcellulose, ethyl cellulose, cellulose acetate, polyethylene oxide, Xanthan gum,

Carbomer, ammonio methacrylate copolymer, hydrogenated castor oil, carnauba wax, paraffin wax, cellulose acetate phthalate, hydroxypropylmethyl cellulose phthalate, methacrylic acid copolymer, and mixtures thereof.

The pharmaceutical composition of the present invention may be in the form of fast dispersing or dissolving dosage formulations (FDDFs) and may contain the following ingredients: aspartame, acesulfame potassium, citric acid, croscarmellose sodium, crospovidone, diascorbic acid, ethyl acrylate, ethyl cellulose, gelatin, hydroxypropylmethyl cellulose, magnesium stearate, mannitol, methyl methacrylate, mint flavoring, polyethylene glycol, fumed silica, silicon dioxide, sodium starch glycolate, sodium stearyl fumarate, sorbitol, xylitol.

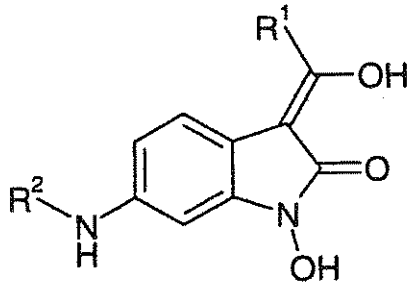
For preparing suppositories, a low melting wax, such as a mixture of fatty acid glycerides or cocoa butter, is first melted and the active component is dispersed homogeneously therein, as by stirring. The molten homogeneous mixture is then poured into convenient sized molds, allowed to cool, and thereby to solidify.

The pharmaceutical composition of the present invention suitable for parenteral administration is best used in the form of a sterile aqueous solution which may contain other substances, for example, enough salts or glucose to make the solution isotonic with blood. The aqueous solutions should be suitably buffered (preferably to a pH of from 3 to 9), if necessary.

The pharmaceutical composition suitable for intranasal administration and administration by inhalation is best delivered in the form of a dry powder inhaler or an aerosol spray from a pressurized container, pump, spray or nebulizer with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, a hydrofluoroalkane such as 1,1,1,2-tetrafluoroethane (HFA 134A.TM.) or 1,1,1,2,3,3,3-heptafluoropropane (HFA 227EA.TM.), carbon dioxide, or another suitable gas. The pressurized container, pump, spray or nebulizer may contain a solution or suspension of the active compound, e.g., using a mixture of ethanol and the propellant as the solvent, which may additionally contain a lubricant, e.g., sorbitan trioleate.

It is another aspect of the invention to provide a compound identifiable by the above-described method, wherein the compound is able to modulate the endonuclease activity of the PA subunit or variant thereof. In another aspect, the present invention refers to a compound identifiable by the above-described method, wherein the compound is able to decrease, preferably inhibit the endonuclease activity of the PA subunit or variant thereof, e.g., the PA subunit polypeptide or variant thereof according to the present invention. Compounds of the

present invention can be any agents as described above for the *in silico* screening methods. In a preferred embodiment of the compound according to the present invention, the compound is not a 4-substituted 2-dioxobutanoic acid, a 4-substituted 4-dioxobutanoic acid, a 4-substituted 2,4-dioxobutanoic acid, a pyrazine-2,6-dione or a substituted pyrazine-2,6-dione such as flutimide, an N-hydroxamic acid, or an N-hydroxymide. In particular, the compound according to the present invention is not a compound according to Formula I:



R^1 = Me, Et, Ph
R^2 = acyl, carbamoyl, sulfonyl.

Formula I

In another aspect, the present invention provides an antibody directed against the endonuclease domain of the PA subunit. In a preferred embodiment, said antibody recognizes the endonuclease domain by recognition of a polypeptide fragment selected from the group of polypeptides defined by SEQ ID NO: 9 to 17, i.e., amino acids 20 to 30 (SEQ ID NO: 9), 35 to 45 (SEQ ID NO: 10), 75 to 85 (SEQ ID NO: 11), 80 to 90 (SEQ ID NO: 12), 100 to 110 (SEQ ID NO: 13), 107 to 112 (SEQ ID NO: 20), 115 to 125 (SEQ ID NO: 14), 125 to 135 (SEQ ID NO: 15), 130 to 140 (SEQ ID NO: 16), and 135 to 145 (SEQ ID NO: 17) of the amino acid sequence as set forth in SEQ ID NO: 2. Preferably said antibody recognizes the amino sequence PDLYDYK (SEQ ID NO: 20). In particular, said antibody specifically binds to an epitope comprising one or more of above indicated amino acids, which define the active site. In this context, the term epitope has its art recognized meaning and preferably refers to stretches of 4 to 20 amino acids, preferably 5 to 18, 5 to 15, or 7 to 14 amino acids. Accordingly, preferred epitopes have a length of 4 to 20, 5 to 18, preferably 5 to 15, or 7 to 14 amino acids and comprise one or more of Asp108, Ile120, Lys134, His41, Glu80, Glu119, Tyr24, Arg84, Leu106, Tyr130, Glu133, and/or Lys137 of SEQ ID NO: 2 or one or more corresponding amino acid(s).

The antibody of the present invention may be a monoclonal or polyclonal antibody or portions thereof. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. In some embodiments, antigen-binding portions include Fab, Fab', $F(ab')_2$, Fd, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies such as

humanized antibodies, diabodies, and polypeptides that contain at least a portion of an antibody that is sufficient to confer specific antigen binding to the polypeptide. The antibody of the present invention is generated according to standard protocols. For example, a polyclonal antibody may be generated by immunizing an animal such as mouse, rat, rabbit, goat, sheep, pig, cattle, or horse with the antigen of interest optionally in combination with an adjuvant such as Freund's complete or incomplete adjuvant, RIBI (muramyl dipeptides), or ISCOM (immunostimulating complexes) according to standard methods well known to the person skilled in the art. The polyclonal antiserum directed against the endonuclease domain of PA or fragments thereof is obtained from the animal by bleeding or sacrificing the immunized animal. The serum (i) may be used as it is obtained from the animal, (ii) an immunoglobulin fraction may be obtained from the serum, or (iii) the antibodies specific for the endonuclease domain of PA or fragments thereof may be purified from the serum. Monoclonal antibodies may be generated by methods well known to the person skilled in the art. In brief, the animal is sacrificed after immunization and lymph node and/or splenic B cells are immortalized by any means known in the art. Methods of immortalizing cells include, but are not limited to, transfecting them with oncogenes, infecting them with an oncogenic virus and cultivating them under conditions that select for immortalized cells, subjecting them to carcinogenic or mutating compounds, fusing them with an immortalized cell, e.g., a myeloma cell, and inactivating a tumor suppressor gene. Immortalized cells are screened using the PA endonuclease domain or a fragment thereof. Cells that produce antibodies directed against the PA endonuclease domain or a fragment thereof, e.g., hybridomas, are selected, cloned, and further screened for desirable characteristics including robust growth, high antibody production, and desirable antibody characteristics. Hybridomas can be expanded (i) *in vivo* in syngeneic animals, (ii) in animals that lack an immune system, e.g., nude mice, or (iii) in cell culture *in vitro*. Methods of selecting, cloning, and expanding hybridomas are well known to those of ordinary skill in the art. The skilled person may refer to standard texts such as "Antibodies: A Laboratory Manual", Harlow and Lane, Eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1990), which is incorporated herein by reference, for support regarding generation of antibodies.

In another aspect, the present invention relates to the use of a compound identifiable by the above-described methods that is able to bind to the endonucleolytically active site of the PA subunit polypeptide fragment or variant thereof, and/or is able to modulate, preferably decrease, more preferably inhibit the endonucleolytic activity of the PA subunit polypeptide fragment or variant thereof, the pharmaceutical composition described above, or the antibody

of the present invention for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with negative-sense single stranded RNA viruses of the family of Orthomyxoviridae. In a preferred embodiment, said disease conditions are caused by viral infections with Influenza A virus, Influenza B virus, Influenza C virus, Isavirus, or Thogotovirus. In an even more preferred embodiment, said disease condition is caused by an infection with a virus species selected from the group consisting of Influenza A virus, Influenza B virus, Influenza C virus, most preferably Influenza A virus.

For treating, ameliorating, or preventing said disease conditions the medicament of the present invention can be administered to an animal patient, preferably a mammalian patient, preferably a human patient, orally, buccally, sublingually, intranasally, via pulmonary routes such as by inhalation, via rectal routes, or parenterally, for example, intracavernosally, intravenously, intra-arterially, intraperitoneally, intrathecally, intraventricularly, intra-urethrally intrasternally, intracranially, intramuscularly, or subcutaneously, they may be administered by infusion or needleless injection techniques.

The pharmaceutical compositions of the present invention may be formulated in various ways well known to one of skill in the art and as described above.

The pharmaceutical preparation is preferably in unit dosage form. In such form the preparation is subdivided into unit doses containing appropriate quantities of the active component. The unit dosage form can be a packaged preparation, the package containing discrete quantities of preparation, such as packeted tablets, capsules, and powders in vials or ampoules. Also, the unit dosage form can be a capsule, tablet, cachet, or lozenge itself, or it can be the appropriate number of any of these in packaged form.

The quantity of active component in a unit dose preparation administered in the use of the present invention may be varied or adjusted from about 1 mg to about 1000 mg per m^2, preferably about 5 mg to about 150 mg/m^2 according to the particular application and the potency of the active component.

The compounds employed in the medical use of the invention are administered at an initial dosage of about 0.05 mg/kg to about 20 mg/kg daily. A daily dose range of about 0.05 mg/kg to about 2 mg/kg is preferred, with a daily dose range of about 0.05 mg/kg to about 1 mg/kg being most preferred. The dosages, however, may be varied depending upon the requirements of the patient, the severity of the condition being treated, and the compound being employed. Determination of the proper dosage for a particular situation is within the skill of the practitioner. Generally, treatment is initiated with smaller dosages, which are less than the optimum dose of the compound. Thereafter, the dosage is increased by small

increments until the optimum effect under circumstances is reached. For convenience, the total daily dosage may be divided and administered in portions during the day, if desired.

EXAMPLES

The Examples are designed in order to further illustrate the present invention and serve a better understanding. They are not to be construed as limiting the scope of the invention in any way.

Summary of the Examples

PA-Nter, residues 1-209 of the amino acid sequence set forth in SEQ ID NO: 2 (A/Victoria/3/1975 (H3N2)) was expressed in *E. coli* and purified by affinity and gel filtration chromatography. The influence of metal ions on thermal stability was tested by thermofluor assays (Ericsson et al., 2006, Anal. Biochem. 357:289-298). The endonuclease activity was tested by incubation at 37°C of 13 μM PA-Nter with 10 μM of various RNA substrates: Alu-RNA; 110 nucleotides of the Alu-domain of *P. horikoshii*, SRP RNA, *C. albicans* tRNA Asn, U-rich RNA (5'-GGCCAUCCUGU$_7$CCCU$_{11}$CU$_{19}$-3'; SEQ ID NO: 18, Saito et al., 2008, Nature 454:523-527), panhandle RNA (ph-RNA) of 81 nucleotides (Baudin et al., 1994, EMBO J. 13:3158-3165), short ph-RNA of 36 nucleotides comprising just the conserved 3'- and 5'-ends with a short linker, and circular single stranded DNA (M13mp18) (Fermentas). Crystals diffracting to 2 Å resolution were obtained at 20°C by the hanging drop method using a protein solution of 5-10 mg/ml in 20 mM Tris pH 8.0, 100 mM NaCl, and 2.5 mM $MnCl_2$ and a reservoir composition of 1.2 M Li_2SO_4, 100 mM MES pH 6.0, 10 mM magnesium acetate and 3% ethylene glycol. Diffraction data were collected on beamlines ID14-4 and ID23-1 at the European Synchrotron Radiation Facility (ESRF). The structure was solved by the single-wavelength anomalous dispersion (SAD) method using a gadolinium chloride soaked crystal. Nine sites were found by SHELXD (Schneider and Sheldrick, 2002, Acta Crastallogr. D. Biol. Crystallogr. 58:1772-1779) and refined with SHARP (de La Fortelle et al., 1997, Methods in Enzymology 276:472-494). After three-fold NCS averaging with RESOLVE (Terwilliger, 2002, Acta Crystallogr. D. Biol. Crystallogr. 58:2213-2215) an interpretable map was obtained and much of the model could be built with ARP/wARP (Perrakis et al., 1999, Nat. Struct. Biol. 6:458-463). Additionally, data were measured on a native crystal at the manganese K edge (X-ray wavelength 1.89 Å) to reveal the location and identity of bound manganese ions through anomalous difference Fourier synthesis. There are

three molecules in the asymmetric unit denoted A, B, and C. The metal ion structure is best defined in molecule A. The crystallographic statistics are summarized in Table 1 and more details available in the experimental Examples below.

Table 1: Data collection and refinement statistics of PA-Nter

	PA-Nter native	PA-Nter Mn K-edge	PA-Nter Gd derivative
Data collection			
Beamline (ESRF)	ID14-4	ID23-1	ID14-4
Wavelength (Å)	0.976	1.892	1.008
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions			
a, *b*, *c* (Å)	67.1, 67.1, 302. 9	67.9, 67.9, 300.8	67.8, 67.8, 300.4
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 106.24, 90.0
Resolution (Å)	50-2.05 (2.05-2.10)*	30-2.60 (2.6-2.7)*	30-2.5 (2.5-2.6)*
R_{merge}	0.056 (0.690)	0.055 (0.484)	0.058 (0.539)
$I / \sigma I$	17.6 (2.2)	17.8 (2.5)	14.5 (2.1)
Completeness (%)	93.2 (99.4)	99.7 (99.8)	97.9 (98.0)
Redundancy	4.84 (5.64)	3.66 (3.44)	3.63 (3.15)
Refinement			
Resolution (Å)	30-2.05 (2.05-2.10)*		
Total No. reflections / free	39715/2118		
R_{work}	0.217 (0.278)		
R_{free}	0.268 (0.320)		
No. atoms			
Protein	4742		
Water/sulphate/Mn ions	152/8/5		
Average B-factors (Å^2) All atoms Chains A, B, D	 45.8 41.5 , 40.0, 57.0		
R.m.s. deviations			
Bond lengths (Å)	0.014		
Bond angles (°)	1.363		
Ramachandran Plot**			
Favoured (%)	98.1		
Allowed (%)	99.8		

Example 1: Cloning, Expression and Purification

The DNA coding for PA residues 1-209 of the amino acid sequence set forth in SEQ ID NO: 2 (A/Victoria/3/1975 (H3N2)) was cloned into a pET-M11 expression vector (EMBL) between the NcoI and XhoI sites. A polypeptide linker having the amino acid sequence GMGSGMA (SEQ ID NO: 19) was engineered after the tobacco etch virus (TEV) cleavage site to obtain a 100% cleavage by TEV protease. This vector was used to transform the BL21(DE3)-RIL-CodonPlus *E. coli* strain (Stratagene). The protein was expressed in LB

medium overnight at 15°C after induction with 0.1 mM isopropyl-β-thiogalactopyranoside (ITPG). The protein was purified by an immobilised metal affinity column (IMAC). A second IMAC step was performed after cleavage using a His-tagged TEV protease, followed by gel filtration on a Superdex 200 column (GE Healthcare). Finally, the protein was concentrated to 5 to 10 mg/ml.

Example 2: Endonuclease Assay

All ribonucleic acid substrates for endonuclease assays were obtained by *in vitro* T7 transcription as described previously (Price et al., 1995, J. Mol. Biol. 249:398-408). Two structured RNAs were used: Alu-RNA; 110 nucleotides comprising the *Alu*-domain of *Pyrococcus horikoshii* signal recognition particle (SRP) RNA (unpublished construct) and *Candida albicans* $tRNA^{Asn}$ composed of 76 nucleotides (unpublished construct). We also used a uridine-rich unstructured RNA of 51 nucleotides (U-rich RNA; 5'-$GGCCAUCCUGU_7CCCU_{11}CU_{19}$-3'; SEQ ID NO: 18) (Saito et al., 2008, Nature 454:523-527) and two partially folded RNAs derived from influenza A virus genomic RNA segment 5: a panhandle RNA (ph-RNA) of 81 nucleotides (Baudin et al., 1994, EMBO J. 13:3158-3165) and a shorter panhandle RNA (short ph-RNA) of 36 nucleotides comprising just the conserved 3'- and 5'-ends with a short linker (unpublished construct). The endonuclease activity was also tested using a circular single stranded DNA (M13mp18) (Fermentas).

RNA cleavage was performed by incubating 13 µM PA-Nter with various RNA substrates (all at 10 µM) at 37°C in a final volume of 50 µL. The reaction buffer was 20 mM Tris-HCl pH 8, 100 mM NaCl, 10 mM β-mercaptoethanol, and 1 mM metal salts. Incubations were stopped by addition of EGTA at a final concentration of 20 mM. The reaction products were loaded on 8 M urea polyacrylamide gels (8% or 15%) and stained with methylene blue. The effect of divalent cations on the RNAse activity of PA-Nter was tested at pH 8 (with β-mercaptoethanol) and pH 7 (without β-mercaptoethanol) by incubating ph-RNA with PA-Nter in the presence of different metal salts: $MnCl_2$, $CaCl_2$, $MgCl_2$, $ZnCl_2$ (or $NiCl_2$ at pH 7) and $CoCl_2$. For DNA cleavage, circular single stranded M13mp18 DNA was used. In the 10 µL reaction volume (same buffer as for RNA), 100 ng/µL of purified plasmid M13mp18 was incubated for 60 minutes in the presence of PA-Nter and 1 mM $MnCl_2$. The reaction products were loaded on a 0.8% agarose gel and stained with ethidium bromide. For endonuclease inhibition by 2,4-Dioxo-4-phenylbutanoic acid (DPBA), PA-Nter and ph-RNA or single stranded M13mp18 DNA were incubated in the presence of 1 mM $MnCl_2$ and increasing concentrations of DPBA. Because DPBA is poorly soluble in water, a stock solution of

65 mM DPBA was prepared in 50% ethanol that was further diluted so that only 1 µL of DPBA solution had to be added to each reaction mix to obtain the required final concentration. Addition of the inhibitor in ethanol did not change the pH of the reaction mixture and the addition of the same concentration of ethanol alone had no effect on nuclease activity (not shown).

Using a partially structured 81nt ph-RNA it could be demonstrated that PA-Nter has intrinsic RNase activity that is divalent cation dependent (Figure 5). Consistent with the results on RNPs (Doan et al., 1999, Biochemistry 38:5612-5619, strong activity was observed at pH 8 with manganese and weaker activity with magnesium ions. At pH 7, the PA-Nter endonuclease activity was also observed with cobalt (Figure 6). After 40 minutes incubation highly structured RNAs such as tRNA and SRP *Alu*-RNA were relatively resistant to degradation, partially structured ph- and short-ph-RNAs were partially degraded and unstructured U-rich RNA was completely degraded, suggesting that the enzyme is single-strand specific (Figure 7). The enzyme also completely degraded circular ssDNA showing that it is a nonspecific endonuclease (Figure 8). The endonuclease activity on both RNA and DNA was inhibited in a dose dependent manner by the compound 2,4-dioxo-4-phenylbutanoic acid, a known inhibitor of influenza endonuclease (Figure 9). The K_i for this compound is estimated at 26 µM, in excellent agreement with the IC_{50} reported for the same compound inhibiting cleavage of capped RNA by the intact influenza virus polymerase (Tomassini et al., 1994, Antimicrob. Agents Chemother 38:2827-2837).

Example 3: Thermal Shift Assay

Thermal shift assays were performed with 10 µM of PA-Nter in 20 mM Tris-HCl pH 7.0 or 8.0, 100 mM NaCl and a 5X dilution of SYPRO Orange dye (Invitrogen) as described (Ericsson et al., 2006, Anal. Biochem. 357:289-298). The dye was excited at 490 nm and the emission light was recorded at 575 nm while the temperature was increased by increments of 1°C per minute from 25 to 75°C. Control assays were carried out in the absence of protein or dye to check that no fluorescence signal was recorded.

The thermal shift assay was performed to investigate the thermal stability of PA-Nter in presence and absence of divalent cations. The experiments revealed a significant increase in thermal stability (apparent melting temperature shifts from 44°C to 57°C) upon addition of manganese ions and to a lesser extent upon addition of calcium and magnesium ions (Figures 1 and 2).Titrating the compound 2,4-dioxo-4-phenylbutanoic acid, a known inhibitor of influenza endonuclease, to manganese bound PA-Nter increases the thermal stability even

further (apparent melting temperature shifts form 59°C to 65°C) (Figure 4), whereas the inhibitor has no effect on metal-free enzyme (data not shown).

Example 4: Far UV Circular Dichroism (CD) Spectroscopy

Far-UV CD spectra were recorded with 1 mM path length at 20°C on a JASCO model J-810 CD spectro-polarimeter equipped with a Peltier thermostat. The PA-Nter concentration was 10 µM in 10 mM Tris-HCl, pH 8.0, 10 mM NaCl in the presence or absence of 1 mM $MnCl_2$. Mean residue ellipticity was calculated using the number of residues (PA-Nter is 209 residues long plus 7 additional residues before the starting methionine). Wavelength scans were recorded from 200 to 260 nm and averaged over eight consecutive scans (0.5 nm increment, 1 s response, 1 nm bandwidth and 50 nm/min scanning speed).

The structural effect of manganese binding to PA-Nter, investigated by CD spectroscopy, revealed a significant increase in helical content (estimated 8 to 9 residues) upon addition of 1 mM Mn^{2+} (Figure 3).

Example 5: Crystallization and Crystallography

Initial sitting drop screening was carried out at 20°C mixing 100 nL of protein solution (6 mg/ml) with 100 nL of well solution using a Cartesian robot. Subsequently, larger crystals were obtained at 20°C by the hanging drop method following a ratio of 1:1 well:protein solutions. The protein solution was at 5-10 mg/ml in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 2.5 mM $MnCl_2$. The reservoir composition was 100 mM MES pH 6.0, 1.2 M Li_2SO_4, 10 mM magnesium acetate, 3% ethylene glycol after refinement of the crystallisation condition. Crystals appeared after 1-2 weeks and were typically of a volume of 50x50x15 μm^3.

Crystals were frozen in liquid nitrogen in the presence of 22% ethylene glycol for cryoprotection. Diffraction data were collected at 100 K on beamlines ID14-4 and ID23-1 at the European Synchrotron Radiation Facility (ESRF) and all data were integrated and scaled in the space group $P4_32_12$ using the XDS suite (Kabsch, 1993, J. Appl. Cryst. 26:795-800). The best native data were collected to 2.05 Å resolution at a wavelength of 0.976 Å, after soaking with additional 10 mM $MnCl_2$ for 2 minutes. Additionally, data was measured on native crystals at a wavelength of 1.89 Å (close to the manganese K edge) to reveal the location and identity of any bound manganese ions. The structure was solved with a highly redundant data set to 2.5 Å resolution collected at a wavelength of 1.008 Å from a crystal soaked for 6 h in mother liquor containing 5 mM $GdCl_3$. Three initial Gd sites were located on the basis of their anomalous differences using SHELXD (Schneider and Sheldrick, 2002,

Acta Crystallogr. D. Biol. Crystallogr. 58:1772-1779) as implemented in HKL2MAP (Pape and Schneider, 2004, J. Appl. Cryst. 37:843-844). These initial sites were refined and experimental phases to 3.5 Å were calculated using the single anomalous dispersion (SAD) procedure in SHARP (de La Fortelle et al., 1997, Methods in Enzymology 276:472-494). After several iterative cycles a further 6 sites were identified in the residual maps and the phases were refined to 2.5 Å. These initial phases were improved with the density modification package SOLOMON in SHARP. Finally, a clearly interpretable map was obtained by using 3-fold NCS operators identified from the 9 Gd sites by RESOLVE (Terwilliger, 2002, Acta Crystallogr. D. Biol. Crystallogr. 58:2213-2215) for averaging with DM (Cowtan, 1994, Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31:34-38) as implemented in CCP4 (Collaborative Computational Project, 1994, Acta Crystallogr. D. Biol. Crystallogr. 50:760-763). This averaged map was of sufficient quality for RESOLVE (Terwilliger, 2003, Acta Crystallogr. D. Biol. Crystallogr. 59:45-49) to build 396 out of 648 possible amino acids, of which 85 could be sequence assigned. A manually modified model and a subsequent high resolution data set to 2.05 Å were then put into ARP/wARP (Perrakis et al., 1999, Nat. Struct. Biol. 6:458-463) resulting in a more complete model. This model was refined with Refmac (Murshudov, 1997, Acta Crystallogr. D. Biol. Crystallogr. 53:240-255) iterated with manual rebuilding cycles in O (Jones et al., 1991, Acta Crystallogr. A 47:110-119). Using TLS refinement and tight NCS restraints on parts of the structure, the final R-factor (R-free) is 0.233 (0.291). According to MOLPROBITY (Lovell et al., 2003, Proteins 50:437-450), 97.5%, 99.8% are respectively in the favoured and allowed region of the Ramachandran plot. The crystallographic details are summarized in Table 1. There are three molecules in the asymmetric unit denoted A, B, and D. The metal ion structure is best defined in molecule A. Different molecules have regions 69-74 and 134-143 more or less well ordered. 6 residues of the N-terminal tag and residues 204-209 are not visible. Molecule D is the least well ordered overall (Table 1). In the described structure the crystal contact between two of the molecules (B and D) exhibits multiple conformations perhaps accounting for the relatively high R-factor of the native data for the resolution. Structure figures were drawn with PyMOL (DeLano, 2002, available online at http://www.pymol.sourceforge.net). The sequence alignment in Figure 11 was drawn with ESPript (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) (Gouet et al., 1999, Bioinformatics 15:305-308). The electrostatic surface (Figure 13) was calculated using DelPhi (Rocchia et al., 2002, J. Comput. Chem. 23:128-137). Structural similarity searches were performed with

MSDFOLD (http://www.ebi.ac.uk/msdsrv/ssm/cgi-bin/ssmserver) and Dalilite (http://www.ebi.ac.uk/Tools/dalilite/index.html).

We grew small square-plate crystals of PA-Nter in the presence of both manganese and magnesium that diffracted to about 2 Å resolution, with three independent molecules in the asymmetric unit. The crystal structure reveals a single, folded domain with residues 1-196 visible, comprising seven α-helices and a mixed, five-stranded β-sheet (Figure 10). The structure based sequence alignment amongst influenza A, B and C viruses (Figure 11) projected onto a surface representation reveals a very highly conserved depression that is strongly negatively charged due to a concentration of acidic residues (Figures 12 and 13), suggestive of an active site. A structure similarity search gave no high scoring hits indicating that the global fold is novel. The most similar protein found is the archaeal Holliday junction resolvase Hjc from *Pyrococcus furiosus* (Nishino et al., 2001, Structure 9:197-204). The structural alignment of PA-Nter with Hjc superposes helix α3 and strands β1-5 (Figure 14, left and middle panel) encompassing a structural motif characteristic of many nucleases including resolvases and type II restriction enzymes. The motif includes catalytically important divalent metal ion binding acidic residues Asp33 and Glu46 of Hjc upon which Asp108 and Glu119 of PA-Nter exactly superpose. Structural alignment of PA-Nter with type II restriction endonucleases such as *Bam*HI or *Eco*RV reveals a similar superposition of active site elements (Figure 14, right panel). Catalytically important Glu45, Asp74, Asp90 and Lys92 of *Eco*RV align with His41, Asp108, Glu119 and Lys134 of PA-Nter, respectively, although the lysines are positioned differently in the primary sequence (Figure 16). The conserved lysine is implicated in stabilizing the attacking hydroxide nucleophile during catalysis. Thus PA-Nter is a new member of the PD-(D/E)XK nuclease superfamily which encompasses a diversity of enzymes involved in various aspects of DNA metabolism. In PA-Nter, the characteristic motif occurs at 107-PDLYDYK (SEQ ID NO: 20), although the separation between the two acidic residues is unusually short and the putative catalytically important lysine (Lys134) has 'migrated' to an alternative position, as in some other members of the superfamily. Within this family, PA-Nter is unusual in that it is biologically functional as an RNase and has a histidine in the active site.

To confirm that the conserved acidic residues of PA-Nter are metal binding residues we calculated an anomalous difference map using data collected at the manganese K absorption edge. Two manganese ions were identified in each active site as adjacent anomalous peaks separated by about 3.8 Å (Figure 15, left panel). The stronger peak (Mn1) is co-ordinated by Glu80, Asp108 and two water molecules; the weaker site (Mn2) by His41,

Asp108, Glu119 and the carbonyl oxygen of Ile120. The cited residues are absolutely conserved in all influenza virus PA sequences (except for Ile120 which is conservatively substituted) (Figure 11). The two metal sites correspond closely with those observed in restriction enzymes such as *Eco*RV (Figure 15, right panel). His41 (positioned as Glu45 in *Eco*RV) from helix α3 could be important in conferring manganese specificity, since magnesium and calcium bind less readily to histidine. Manganese binding by His41 and the resulting stabilization of helix α3 could account for the additional helical content (estimated as 8-9 residues) detected upon incubating PA-Nter with manganese (Figure 3). In the crystal, Mn1 is also co-ordinated by Glu59 from a loop of an adjacent molecule. Superposition of DNA complexes of *Bam*HI or *Eco*RV on PA-Nter shows that the Glu59 carboxylate group corresponds closely to the position of the scissile phosphate group (Figure 17). Thus our structure mimics a substrate or product complex.

Our structural and biochemical results combined with previous observations on the trimeric polymerase provide compelling evidence that PA-Nter is the endonuclease that cleaves host mRNAs during cap-snatching. First, the domain has intrinsic RNA and DNA endonuclease activity which is preferentially activated by manganese, in accordance with observations reported for the viral RNPs (Figure 6). Second, this activity is inhibited by a compound known to inhibit influenza endonuclease activity with a nearly identical K_i (Figure 9). Third, the domain contains a structural motif characteristic of the catalytic core of a broad family of nucleases, including type II endonucleases. The active site features a cluster of three acidic residues (Glu80, Asp108 and Glu119) and a putative catalytic lysine (Lys134) (Figures 14 to 16). Fourth, these acidic residues, together with His41, are all absolutely conserved in influenza viruses, co-ordinate two manganese ions in a configuration consistent with a two-metal dependent reaction mechanism as proposed for many nucleases (Figure 15, left panel).

Claims

1. A polypeptide fragment comprising an amino-terminal fragment of the PA subunit of a viral RNA-dependent RNA polymerase possessing endonuclease activity, wherein said PA subunit is from a virus belonging to the Orthomyxoviridae family.

2. The polypeptide fragment of claim 1, which is soluble.

3. The polypeptide fragment of claim 1 or 2, which is crystallizable.

4. The polypeptide fragment of claim 3 which is crystallizable using a protein solution of 5 to 10 mg/ml in 20 mM Tris pH 8.0, 100 mM NaCl and 2.5 mM $MnCl_2$ and a reservoir solution consisting of 1.2 M Li_2SO_4, 100 mM MES pH 6.0, 10 mM magnesium acetate, and 3% ethylene glycol.

5. The polypeptide fragment of any one of claims 1 to 4, wherein the PA subunit is from Influenza A, B, or C virus or is a variant thereof.

6. The polypeptide fragment of any one of claims 1 to 5, wherein the amino terminal fragment corresponds to at least amino acids 1 to 196 of the PA subunit of the RNA-dependent RNA polymerase of Influenza A virus (SEQ ID NO: 2).

7. The polypeptide fragment of any one of claims 1 to 6, wherein said polypeptide fragment is purified to an extent to be suitable for crystallization.

8. The polypeptide fragment of any one of claims 1 to 7 to which two divalent cations are bound, wherein the divalent cation is preferably manganese.

9. The polypeptide fragment of any of claims 1 to 8, wherein

(i) the N-terminus is identical to or corresponds to amino acid position 1 and the C-terminus is identical to or corresponds to an amino acid at a position selected from positions 196 to 209 of the amino acid sequence of the PA subunit according to SEQ ID NO: 2,

(ii) the N-terminus is identical to or corresponds to amino acid position 1 and the C-terminus is identical to or corresponds to an amino acid at a position selected from positions 195 to 206 of the amino acid sequence of the PA subunit according to SEQ ID NO: 4, or

(iii) wherein the N-terminus is identical to or corresponds to amino acid position 1 and the C-terminus is identical to or corresponds to an amino acid at a position selected from positions 178 to 189 of the amino acid sequence of the PA subunit according to SEQ ID NO: 6,

and variants thereof, which retain the endonuclease activity.

10. The polypeptide fragment of any one of claims 1 to 9 which consists of amino acids 1 to 209 of the amino acid sequence set forth in SEQ ID NO: 2 and optionally an amino-terminal linker having the amino acid sequence GMGSGMA (SEQ ID NO: 19) and which has the structure defined by the structure coordinates as shown in Figure 18.

11. The polypeptide fragment of any one of claims 1 to 10, wherein said polypeptide fragment has a crystalline form with space group $P4_32_12$ and unit cell dimensions of a = b = 6.71 ± 0.2 nm, c = 30.29 nm ± 0.4 nm.

12. The polypeptide fragment of claim 10 or 11, wherein the crystal diffracts X-rays to a resolution of 2.5 Å or higher, preferably 2.1 Å or higher.

13. An isolated polynucleotide coding for an isolated polypeptide of any one of claims 1 to 12.

14. A recombinant vector comprising said isolated polynucleotide of claim 13.

15. A recombinant host cell comprising said isolated polynucleotide of claim 13 or said recombinant vector of claim 14.

16. A method for identifying compounds which modulate the endonuclease activity of the PA subunit of a viral RNA-dependent RNA polymerase from the Orthomyxoviridae family, comprising the steps of

(a) constructing a computer model of the active site defined by the structure coordinates of the polypeptide fragment of claim 10 as shown in Figure 18;

(b) selecting a potential modulating compound by a method selected from the group consisting of:

(i) assembling molecular fragments into said compound,

(ii) selecting a compound from a small molecule database, and

(iii) *de novo* ligand design of said compound;

(c) employing computational means to perform a fitting program operation between computer models of the said compound and the said active site in order to provide an energy-minimized configuration of the said compound in the active site; and

(d) evaluating the results of said fitting operation to quantify the association between the said compound and the active site model, whereby evaluating the ability of said compound to associate with the said active site.

17. The method of claim 16, wherein said active site comprises amino acids corresponding to amino acids Asp108, Ile120, and Lys134 of the PA subunit according to SEQ ID NO: 2.

18. The method of claim 17, wherein said active site further comprises amino acids corresponding to amino acids His41, Glu80, and Glu119 of the PA subunit according to SEQ ID NO: 2.

19. The method of claims 17 or 18, wherein said active site further comprises the amino acids corresponding to amino acids Tyr24, Arg84, Leu106, Tyr130, Glu133, and Lys137 of the PA subunit according to SEQ ID NO: 2.

20. The method of claim 16, wherein said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Asp108, Ile120, and Lys134 according to Figure 18.

21. The method of claim 20, wherein said active site is further defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids His41, Glu80, and Glu119 according to Figure 18.

22. The method of claims 20 or 21, wherein said active site is further defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Tyr24, Arg84, Leu106, Tyr130, Glu133, and Lys137 according to Figure 18.

23. The method of claim 21, wherein the active site of a PA subunit polypeptide fragment variant has a root mean square deviation from the backbone atoms of the amino acids Asp108, Ile120, and Lys134 of said active site of not more than 2.5 Å.

24. The method of any of claims 16 to 23 comprising the further step of
(e) synthesizing said compound and optionally formulating said compound or a pharmaceutically acceptable salt thereof with one or more pharmaceutically acceptable excipient(s) and/or carrier(s).

25. The method of claim 24 comprising the further step of
(f) contacting said compound and said polypeptide fragment of any of claims 1 to 12 or said recombinant host cell according to claim 15 and determine the ability of said compound to modulate the endonuclease activity of said PA subunit polypeptide fragment.

26. A compound identifiable by the method of any one of claims 16 to 25, wherein said compound is able to modulate the endonuclease activity of the PA subunit or variant thereof.

27. A compound identifiable by the method of claims 16 to 25, wherein said compound is able to inhibit the endonuclease activity of the PA subunit polypeptide fragment of any one of claims 1 to 12 or variant thereof.

28. A method for identifying compounds which modulate the endonuclease activity of the PA subunit or polypeptide variants thereof, comprising the steps of (i) contacting said polypeptide fragment of claims 1 to 12 or said recombinant host cell of claim 15 with a test compound and (ii) analyzing the ability of said test compound to modulate the endonuclease activity of said PA subunit polypeptide fragment.

29. The method of claim 28, wherein the ability of said test compound to inhibit the endonuclease activity of said PA subunit polypeptide fragment is analyzed.

30. The method of claims 28 or 29 performed in a high-throughput setting.

31. The method of any of claims 28 to 30, wherein said test compound is a small molecule, wherein a small molecule has a molecular weight between 50 and about 2,500 Daltons.

32. The method of any of claims 28 to 31, wherein said test compound is a peptide or protein.

33. The method of any one of claims 16 to 25 and 28 to 32, wherein said method further comprises the step of formulating said compound or a pharmaceutically acceptable salt thereof with one or more pharmaceutically acceptable excipient(s) and/or carrier(s).

34. A pharmaceutical composition producible according to the method of claims 24 or 33.

35. A compound identifiable by the method of any of claims 28 to 33, wherein said compound is able to modulate the endonuclease activity of the PA subunit or variant thereof.

36. A compound identifiable by the method of any of claims 28 to 33, wherein said compound is able to inhibit the endonuclease activity of the PA subunit polypeptide fragment of any one of claims 1 to 12 or variant thereof.

37. Use of a compound according to claims 26, 27, 35, or 36, a pharmaceutical composition according to claim 34 for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with viruses of the Orthomyxoviridae family.

38. The use of claim 37, wherein said disease condition is caused by a virus selected from the group consisting of Influenza A virus, Influenza B virus, and Influenza C virus.