INFLUENZA A 2009 PANDEMIC H1N1 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 possessing endonuclease activity, wherein said PA subunit is from Influenza A 2009 pandemic H1N1 virus or is a variant thereof. 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 which are able to modulate, preferably to inhibit, the endonuclease activity of the PA subunit polypeptide fragment or variant thereof of the present invention and pharmaceutical compositions comprising said compounds for the treatment of disease conditions caused by viral infections with viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, preferably caused by viral infections with Influenza A 2009 pandemic H1N1 virus. Preferably, said compounds are identifiable by the methods disclosed herein or said pharmaceutical compositions are producible by the methods disclosed herein.

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.

It has been found recently that, contrary to the general opinion in the field, the endonucleolytic activity resides exclusively within the PA subunit of the RNA-dependent RNA polymerase of the Influenza A H3N2 virus (Dias et al., 2009).

The present inventors have now achieved to structurally characterize the PA domain of the Influenza A 2009 pandemic H1N1 virus by X-ray crystallography and identified the endonucleolytic active center within said domain. The present inventors surprisingly found that polypeptide fragments of the PA subunit of said virus readily crystallized and that, thus, said polypeptide fragments are very suitable to study the endonucleolytic activity of the RNA-dependent RNA polymerase of the Influenza A 2009 pandemic H1N1 virus in the context of said polypeptide fragments in order to simplify the development of new anti-viral compounds targeting the endonuclease activity of said viral polymerase as well as to optimize previously identified compounds.

The achievement of the present inventors to recombinantly produce PA polypeptide fragments possessing the endonucleolytic activity of the RNA-dependent RNA polymerase of the Influenza A 2009 pandemic H1N1 virus allows for performing *in vitro* high-throughput screening for inhibitors of a functional site on said viral polymerase using easily obtainable material from a straightforward expression system. Furthermore, the structural data of the endonucleolytic PA H1N1 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.

The present inventors further managed, for the first time, the co-crystallization of a PA polypeptide fragment and of a variant thereof of the PA subunit of Influenza A 2009 pandemic H1N1 virus with a bound inhibitor and found that, thus, the development of new anti-viral compounds targeting the endonuclease activity of the RNA-dependent RNA polymerase of the Influenza A 2009 pandemic H1N1 virus can be improved. Particularly, the co-crystallization data show, for the first time, in detail which amino acids comprised in the active site of a PA polypeptide fragment of the PA subunit of Influenza A 2009 pandemic H1N1 virus are especially involved in compound binding. This new knowledge allows the optimized design of modifications to existing inhibitors in order to improve their potency or the design and optimization of novel inhibitors that effectively block endonuclease activity.

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 H1N1 PA subunit by X-ray crystallography, (ii) high resolution structural data of the endonucleolytic amino-terminal domain of the viral polymerase H1N1 PA subunit co-crystallized with a known inhibitor by

X-ray crystallography, (iii) 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 of the Influenza A 2009 pandemic H1N1 virus, preferably by blocking the endonucleolytic active site within the H1N1 PA subunit, and (iv) 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 Influenza A pandemic 2009 H1N1 virus according to SEQ ID NO: 2 or is a variant thereof, wherein said variant comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto.

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

In a further aspect, the present invention relates to a 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 present 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 RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or a variant thereof comprising the steps of:

(a) constructing a computer model of the active site defined by (i) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 1, (ii) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 2, (iii) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 3, (iv) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 4, (v) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 5, (vi) the structure coordinates of the

polypeptide fragment according to the present invention as shown in Figure 15, or (vii) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 16,

(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 method for identifying compounds, which modulate the endonuclease activity of the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or a variant thereof comprising the steps of:

(i) contacting the polypeptide fragment or variant thereof according to the present invention or the recombinant host cell according to the present 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 or variant thereof.

In a further aspect, the present invention relates to a compound which is able to modulate, preferably to inhibit, the endonuclease activity of the PA subunit polypeptide fragment or variant thereof according to the present invention. Preferably, said compound is identifiable by the (*in vitro*) methods according to the present inventions.

In a further aspect, the present invention relates to a pharmaceutical composition comprising the compound of the present invention or a pharmaceutically acceptable salt thereof and one or more pharmaceutically acceptable excipient(s) and/or carrier(s). Preferably, said pharmaceutical composition is producible according to the *(in vitro)* methods of the present invention.

In a further aspect, the present invention relates to an antibody directed against the active site of the PA subunit of the Influenza A 2009 pandemic H1N1 virus according to SEQ

ID NO: 2 or a variant thereof, wherein said variant comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto.

In a further 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, Bunyaviridae family and/or Arenviridae family, preferably caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

In a further aspect, the present invention relates to the use of 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3), 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2), 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1), or [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate (EGCG) for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

In a further aspect, the present invention relates to a compound according to the present invention, a pharmaceutical composition according to the present invention, or an antibody according to the present invention for treating, ameliorating, or preventing disease conditions caused by viral infections with viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, preferably caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

In a further aspect, the present invention relates to 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3), 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2), 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1), or [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate (EGCG) for treating, ameliorating, or preventing disease conditions caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

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 198 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 be 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 198 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 quaternary 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. 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.

The present invention relates to Influenza A 2009 pandemic H1N1 virus RNA-dependent RNA polymerase PA subunit fragments possessing endonuclease activity. The term "RNA-dependent RNA polymerase PA subunit" refers to the PA subunit of Influenza A 2009 pandemic H1N1 virus having an amino acid sequence as set forth in SEQ ID NO: 2. The term "RNA-dependent RNA polymerase PA subunit variant" refers to a PA subunit variant of Influenza A 2009 pandemic H1N1 virus which comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto and preferably has 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%, or 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. It is preferred that when a PA subunit variant is aligned with a PA subunit according to SEQ ID NO: 2 that the alignment will be over the entire length of the two polypeptides and, thus, that the alignment score will be determined on this basis. It is, however, possible that the PA subunit 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, 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 and a PA subunit variant should preferably comprise the endonuclease active site. Thus, the above sequence similarity and identity to SEQ ID NO: 2 occurs at least over a length of 80, 90, 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.

The polypeptide fragments of the present invention are, thus, based on Influenza A 2009 pandemic H1N1 virus RNA-dependent RNA polymerase PA subunit or variants thereof as defined above. Accordingly, in the following specification, the terms "polypeptide fragment(s)" and "PA polypeptide fragment(s)" always comprise fragments derived both from

the PA protein as set out in SEQ ID NO: 2 and from PA protein variants thereof, as set out above, possessing endonuclease activity.

However, the specification also uses the terms "PA polypeptide fragment variants" or "PA fragment variants" to specifically refer to PA polypeptide fragments or PA fragments possessing endonuclease activity that are derived from Influenza A 2009 pandemic H1N1 virus RNA-dependent RNA polymerase PA subunit variants. The PA polypeptide fragments of the present invention, thus, preferably comprise, essentially consist or consist of sequences of the naturally occurring Influenza A 2009 pandemic H1N1 virus PA subunit. It is, however, also envisioned that the PA polypeptide fragment variants comprise the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto and preferably 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/or 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. 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 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 sequence of SEQ ID NO: 2 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 190 of Influenza A 2009 pandemic H1N1 virus PA subunit according to SEQ ID NO: 2 or consist of amino acid residues 1 to 190 of Influenza A 2009 pandemic H1N1 virus PA subunit according to 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 190 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 198 of Influenza A 2009 pandemic H1N1 virus PA subunit according to SEQ ID NO: 2 or consist of amino acid residues 1 to 198 of Influenza A 2009 pandemic H1N1 virus PA subunit according to 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 198 of the amino acid sequence set forth in SEQ ID NO: 2. It is also preferred that the PA polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 1 to 200 of Influenza A 2009 pandemic H1N1 virus PA subunit according to SEQ ID NO: 2 or consist of amino acid residues 1 to 200 of Influenza A 2009 pandemic H1N1 virus PA subunit according to 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 200 of the amino acid sequence set forth in SEQ ID NO: 2. It should be noted that all of the above mentioned preferred PA polypeptide fragment variants comprise the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto.

In the context of the present invention, the term "PA-Nter" refers to a polypeptide fragment which consists of amino acid residues 1 to 198 of the amino acid sequence as set forth in SEQ ID NO: 2 (PA H1N1 1 to 198) and optionally of an additional amino-terminal linker, i.e. MGSGMA (SEQ ID NO: 3). In the context of the present invention, the term "PA-Nter mutant" refers to a polypeptide fragment (variant) which consists of amino acid residues 1 to 198 of the amino acid sequence as set forth in SEQ ID NO: 2 with amino acids 52 to 64

replaced by the amino acid glycine (PA H1N1 1 to 198 Δ52-64: Gly) and optionally of an additional amino-terminal linker, i.e. MGSGMA (SEQ ID NO: 3).

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 90% or 95% pure, most preferably at least 99% pure. The expression "purified to an extent to be suitable for crystallization" refers to a polypeptide that is 85% to 100%, preferably 90% to 100%, more preferably 95% to 100% or 98% 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, 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, cRNA, 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 Influenz A pandemic H1N1 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 High Five 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 cubic or 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 1.

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 of the Influenza A 2009 pandemic H1N1 virus or the Influenza A 2009 pandemic H1N1 virus RNA-dependent RNA polymerase or a variant thereof. Preferably, such a compound is specific for the endonucleolytic activity of the Influenza A 2009 pandemic H1N1 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 Influenza A 2009 pandemic H1N1 virus or the Influenza A 2009 pandemic H1N1 virus RNA-dependent RNA polymerase 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 said 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 of the Influenza A 2009 pandemic H1N1 virus 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 of the Influenza A 2009 pandemic H1N1 virus or a variant thereof specifically decreases or inhibits the endonucleolytic activity of said 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 H1N1 polypeptide

fragment or variant thereof, (2) substrate such as panhandle RNA, endonucleolytically active PA H1N1 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 (4) and intact substrate in sample (2) 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 Influenza A 2009 pandemic H1N1 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 Influenza A 2009 pandemic H1N1 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 inventors surprisingly found that a small independently folded domain derived from the N-terminus of the PA subunit of Influenza A 2009 pandemic H1N1 virus RNA-dependent RNA polymerase 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. The present inventors surprisingly found that the domain derived from the N-terminus of the PA subunit of Influenza A 2009 pandemic H1N1 virus RNA-dependent RNA polymerase readily crystallized and that, thus, said domain is very dedicated for the identification, selection and design of new anti-viral compounds targeting the endonuclease activity of the Influenza A 2009 pandemic H1N1 polymerase as well as for the optimization of previously identified compounds by computer methods.

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 Influenza A 2009 pandemic H1N1 virus according to SEQ ID NO: 2 or is a variant thereof, wherein said variant comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto.

In a preferred embodiment of the present invention, said variant comprises the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 64 replaced by the amino acid glycine. In a further preferred embodiment of the present invention, said variant comprises the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 72 replaced by the amino acid glycine. In another further preferred embodiment of the present invention, said variant comprises the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 69 replaced by the amino acid glycine.

It is preferred that the polypeptide fragment according to the present invention is soluble, preferably 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 according to the present invention 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 20 mg/ml, e.g. of 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 15.5, 16, 16.5, 17, 17.5, 18, 18.5, 19, 19.5, or 20 mg/ml, preferably of 8 to 15 mg/ml, most preferably of 10 to 15 mg/ml in a buffer system such as HEPES or 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 substances such as sodium formate, ammonium sulphate, lithium sulphate, magnesium acetate, manganese acetate, sodium chloride, glycerol 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 $MgCl_2$ and $MnCl_2$, at concentrations ranging from 0.1 to 50 mM, preferably 0.5 to 25 mM, more preferably 1 to 10 mM or 1 to 5 mM.

Preferably, the precipitant/reservoir solution comprises sodium formate at concentrations ranging from 0.5 to 2 M, preferably 1 to 1.8 M, a buffer system such as HEPES at concentrations ranging from 10 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, and/or ethylene glycol at concentrations ranging from 1% to 20%, preferably 2% to 8%, more preferably 2 to 5%.

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 2.0 M sodium formate, 80 to 120 mM HEPES pH 6.5 to pH 7.5, and 2 to 5% glycol or glycerol.

In another embodiment, the precipitant solution comprises, preferably essentially consists of or consists of 1.6 M sodium formate, 0.1 M HEPES pH 7.0, and 5% glycol or

glycerol, and the crystallization/protein solution comprises, preferably essentially consists or consists of 10 to 15 mg/ml in 20 mM HEPES pH 7.5, 150 mM NaCl, 2.0 mM $MnCl_2$, and 2.0 mM $MgCl_2$.

In another embodiment, the PA polypeptide fragment is co-crystallizable with a compound, preferably with a compound that modulates, preferably inhibits, the endonuclease activity of the PA polypeptide fragment, preferably with a compound according to Tables 1 or 2, using (i) an aqueous protein solution with a concentration of the PA polypeptide fragment of 5 to 20 mg/ml, e.g., 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 15.5, 16, 16.5, 17, 17.5, 18, 18.5, 19, 19.5, or 20 mg/ml, preferably of 8 to 15 mg/ml, most preferably of 10 to 15 mg/ml in a buffer system such as HEPES or 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 substances such as sodium formate, ammonium sulphate, lithium sulphate, magnesium acetate, manganese acetate, sodium chloride, ethylene glycol, glycerol, or PEG. For co-crystallization, a compound, preferably a compound that modulates, e.g. inhibits, the endonuclease activity of the PA polypeptide fragment, preferably a compound according to Tables 1 or 2, is added to the aqueous protein solution to a final concentration of between 0.5 and 5 mM, preferably of between 1.5 and 5 mM, i.e. 0.5, 1, 1.5, 2, 2.5, 3, 4.5 or 5 mM.

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 $MgCl_2$ and $MnCl_2$, at concentrations ranging from 0.1 to 50 mM, preferably 0.5 to 25 mM, more preferably 1 to 10 mM or 1 to 5 mM.

Preferably, the precipitant/reservoir solution comprises ammonium sulphate at concentrations ranging from 0.1 to 2.5 M, preferably 0.1 to 2.0 M, a buffer system such as Bis-Tris at concentrations ranging from 10 mM to 1 M, preferably 50 mM to 500 mM, more preferably 75 to 150 mM, at preferably pH 4 to 7, more preferably pH 5 to 6, and/or PEG such as PEG 3350 at concentrations ranging from 1% to 30%, preferably 15% to 30%, more preferably 20 to 25%.

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 protein solution. For co-crystallization, the aqueous protein solution comprising the PA polypeptide

fragment or variant thereof and the compound may be mixed with an equal volume of the precipitant solution.

In a preferred embodiment, the protein solution comprises, preferably essentially consists or consists of 10 to 15 mg/ml PA polypeptide fragment in 20 mM HEPES pH 7.5, 150 mM NaCl, 2.0 mM $MnCl_2$, and 2.0 mM $MgCl_2$ and 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3) in a final concentration of 1.5 mM, and the reservoir solution/precipitant solution comprises, preferably essentially consists of or consists of 2.0 M ammonium sulphate and 0.1M Bis-Tris pH5.5 (see also Table 1).

In another preferred embodiment, the protein solution comprises, preferably essentially consists or consists of 10 to 15 mg/ml PA polypeptide fragment in 20 mM HEPES pH 7.5, 150 mM NaCl, 2.0 mM $MnCl_2$, and 2.0 mM $MgCl_2$ and 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2) in a final concentration of 1.5 mM, preferably soaked into crystals initially grown with ribo-Uridine Monophosphate (rUMP), and the reservoir solution/precipitant solution comprises, preferably essentially consists of or consists of 0.1 M ammonium sulphate, 0.1 M Bis-Tris pH5.5 and 25 % (w/v) PEG 3350 (see also Table 1).

In a further preferred embodiment, the protein solution comprises, preferably essentially consists or consists of 10 to 15 mg/ml PA polypeptide fragment in 20 mM HEPES pH 7.5, 150 mM NaCl, 2.0 mM $MnCl_2$, and 2.0 mM $MgCl_2$ and ribo-Uridine Monophosphate (rUMP) in a final concentration of 5 mM, and the reservoir solution/precipitant solution comprises, preferably essentially consists of or consists of 0.1 M ammonium sulphate, 0.1 M Bis-Tris pH5.5 and 25 % (w/v) PEG 3350 (see also Table 1).

Other preferred reservoir solutions/precipitant solutions for the co-crystallization of the PA polypeptide fragment or variant thereof, e.g. with 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1) or with [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate (EGCG), can further be taken from Table 1.

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 Figures 1 to 5, 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).

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 190, preferably amino acids 1 to 198, preferably amino

acids 1 to 200, of the PA subunit of the RNA-dependent RNA polymerase of Influenza A 2009 pandemic H1N1 virus according to SEQ ID NO: 2 or variants thereof.

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 preferred embodiment, the polypeptide fragment according to the present invention is purified to an extent to be suitable for co-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 present 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 and/or magnesium. Thus, in a preferred embodiment, the polypeptide fragment of the present invention is present in complex with (i) two manganese cations, (ii) two magnesium cations, or (iii) one manganese and one magnesium cation. In a preferred embodiment, the divalent cations are coordinated by amino acids corresponding to amino acids Glu80 and Asp108 (second cation) and amino acids corresponding to amino acids His41, Asp108, Ile120 and Glu119 (first cation) as set forth in SEQ ID NO: 2. In preferred embodiment, the divalent cation is manganese for site 1 and manganese or magnesium for site 2, i.e. manganese for site 1 and magnesium for site 2 (see Figure 6) or manganese for site 1 and manganese for site 2 (see Figures 7 to 10, 17 and 18)

In a further embodiment, the polypeptide fragment according to the present invention is a polypeptide fragment, wherein 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 200, e.g., 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: 2, preferably the C-terminus is identical to or corresponds to an amino acid at a position selected from 190 to 200, more preferably 190 to 198 of the amino acid sequence of the PA subunit according to SEQ ID NO: 2, or is a variant thereof, wherein said variant comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto and retains the endonuclease activity.

In a preferred embodiment of the present invention, said variant comprises the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 64 replaced by the amino acid glycine. In a more preferred embodiment of the present invention, said variant consists of the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 64 replaced by the amino acid glycine and optionally of an amino-terminal linker having the amino acid sequence MGSGMA (SEQ ID NO: 3).

In a further preferred embodiment of the present invention, said variant comprises the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 72 replaced by the amino acid glycine. In a more preferred embodiment of the present invention, said variant consists of the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 72 replaced by the amino acid glycine and optionally of an amino-terminal linker having the amino acid sequence MGSGMA (SEQ ID NO: 3).

In another further preferred embodiment of the present invention, said variant comprises the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 69 replaced by the amino acid glycine. In a more preferred embodiment of the present invention, said variant consists of the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 69 replaced by the amino acid glycine and optionally of an amino-terminal linker having the amino acid sequence MGSGMA (SEQ ID NO: 3).

Preferably, 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 190, 10 to 190, 15 to 190, 20 to 190, 5 to 198, 10 to 198, 15 to 198, 20 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 and variants thereof, wherein said variants comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto and retain the endonuclease activity.

In a further embodiment, the polypeptide fragment (variant) according to the present invention

(a) consists of amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 and optionally of an amino-terminal linker having the amino acid sequence MGSGMA (SEQ ID NO: 3) and has the structure defined by (i) the structure coordinates as shown in Figure 1, (ii) the structure coordinates as shown in Figure 2,

(iii) the structure coordinates as shown in Figure 3, (iv) the structure coordinates as shown in Figure 4, or (v) the structure coordinates as shown in Figure 5, or

(b) consists of amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 64 replaced by the amino acid glycine and optionally of an amino-terminal linker having the amino acid sequence MGSGMA (SEQ ID NO: 3) and has the structure defined by (vi) the structure coordinates as shown in Figure 15, or (vii) the structure coordinates as shown in Figure 16.

It is preferred that said polypeptide fragment according to the present invention has the crystal structure defined by the structure coordinates as shown in Figure 1 without co-crystallization with a compound (native structure).

It is further preferred that said polypeptide fragment (variant) according to the present invention has the crystal structure defined by the structure coordinates as shown in Figures 2 to 5, 15 or 16 after co-crystallization with a compound, preferably with a compound that modulates, preferably inhibits, the endonuclease activity of said polypeptide fragment.

It is particularly preferred that said polypeptide fragment (variant) has the structure defined by the structure coordinates as shown in Figures 2 to 5, 15 or 16 after co-crystallization with a compound according to Tables 1 or 2, i.e. the structure defined by the structure coordinates as shown in Figures 2 and 3 after co-crystallization with 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3), the structure defined by the structure coordinates as shown in Figure 4 after co-crystallization with 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid, (EMBL-R05-2), the structure defined by the structure coordinates as shown in Figure 5 after co-crystallization with ribo-uridine monophosphate (rUMP), the structure defined by the structure coordinates as shown in Figure 15 after co-crystallization with 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1), or the structure defined by the structure coordinates as shown in Figure 16 after co-crystallization with [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate (EGCG).

Considering the above, the skilled person will readily understand that said polypeptide fragment (variant) has a crystal structure which is only defined by the structure coordinates of its amino acids and optionally of its bound divalent cations comprised in Figures 1, 2 to 5, 15 or 16 and that it is not defined by the structure coordinates of the respective compound used for co-crystallization which are also comprised in Figures 2 to 5, 15 or 16, i.e. the compound EMBL-R05-3 has the residue descriptor *ci3* in Figures 2 and 3, the compound EMBL-R05-2

has the residue descriptor *ci2* in Figure 4, the compound rUMP has the residue descriptor *U* in Figure 5, the compound EMBL-R05-1 has the residue descriptor *ci1* in Figure 15 and the compound EGCG has the residue descriptor *tte* in Figure 16.

It is also preferred that said polypeptide fragment (variant) having the structure defined by the structure coordinates as shown in Figure 1 has a crystalline form with space group *C*2 and unit cell dimensions of a = 26.36 nm ± 0.5 nm, b = 6.62 nm ± 0.3 nm, c = 6.63 nm ± 0.3 nm, α = 90 deg, β = 96 ± 2 deg, γ = 90 deg, having the structure defined by the structure coordinates as shown in Figures 2 to 5 has a crystalline form with space group $P2_12_12_1$ and unit cell dimensions of a = 5.46 ± 0.3 nm, b= 12.25 ± 0.4 nm, c=13.0 ± 0.3 nm, α = 90 deg, β = 90 deg, γ = 90 deg, having the structure defined by the structure coordinates as shown in Figure 15 has a crystalline form with space group $P\,6_222$ and unit cell dimensions of a = 7.50 nm ± 0.3 nm, b = 7.50 nm ± 0.3 nm, c = 12.00 nm ± 0.5 nm, α = 90 deg, β = 90 deg, γ = 120 deg, or having the structure defined by the structure coordinates as shown in Figure 16 has a crystalline form with space group $P6_422$ and unit cell dimensions of a = 9.99 nm ± 0.5 nm, b = 9.99 nm ± 0.5 nm, c = 8.27 nm ± 0.3 nm, α = 90 deg, β = 90 deg, γ = 120 deg.

Preferably, the crystal of the polypeptide fragment (variant) 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 or 1.9 Å or higher. For example, the crystal of the polypeptide fragment (variant) diffracts X-rays to a resolution of 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8 Å or higher.

It is another aspect of the present invention to provide an isolated polynucleotide coding for the above-mentioned (isolated) PA polypeptide fragments and variants thereof according to the present invention.

In a preferred embodiment of the present invention, the isolated polynucleotide codes for the PA polypeptide fragment which comprises the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2.

In another preferred embodiment of the present invention, the isolated polynucleotide codes for the PA polypeptide fragment variant which comprises the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 64 replaced by the amino acid glycine.

In a further preferred embodiment of the present invention, the isolated polynucleotide codes for the PA polypeptide fragment variant which comprises the amino acids 1 to 198 of

the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 72 replaced by the amino acid glycine.

In another further preferred embodiment of the present invention, the isolated polynucleotide codes for the PA polypeptide fragment variant which comprises the amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 69 replaced by the amino acid glycine.

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 A 2009 pandemic H1N1 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.

The isolated polynucleotides coding for the Influenza A 2009 pandemic H1N1 virus RNA-dependent RNA polymerase PA subunit fragments are derived from SEQ ID NO: 1. In this context, "derived" refers to the fact that SEQ ID NO: 1 encodes the full-length PA polypeptide and, thus, polynucleotides coding for preferred PA polypeptide fragments may comprise deletions at the 3'- and/or 5'-ends of the polynucleotides as required by the respectively encoded PA polypeptide fragments.

In another aspect, the present invention relates to a recombinant vector comprising the isolated polynucleotide according to the present invention. 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: 12). 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 the isolated polynucleotide according to the present invention or the recombinant vector according to the present invention. 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 RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or a variant thereof comprising the steps of:

(a) constructing a computer model of the active site defined by (i) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 1, (ii) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 2, (iii) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 3, (iv) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 4, (v) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 5, (vi) the structure coordinates of the polypeptide fragment (variant) according to the present invention as shown in Figure 15, and/or (vii) the structure coordinates of the polypeptide fragment (variant) according to the present invention as shown in Figure 16,

(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 preferred embodiment, a computer model of the active site defined by (i) the structure coordinates of the polypeptide fragment according to the present invention (i.e. amino acid structure coordinates) as shown in Figure 1 and the structure coordinates of the divalent cations as shown in Figure 1, (ii) the structure coordinates of the polypeptide fragment according to the present invention (i.e. amino acid structure coordinates) as shown in Figure 2 and the structure coordinates of the divalent cations as shown in Figure 2, (iii) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 3 and the structure coordinates of the divalent cations as shown in Figure 3, (iv) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 4 and the structure coordinates of the divalent cations as shown in Figure 4, (v) the structure coordinates of the polypeptide fragment according to the present invention as shown in Figure 5 and the structure coordinates of the divalent cations as shown in Figure 5, (vi) the structure coordinates of the polypeptide fragment (variant) according to the present invention (i.e. amino acid structure coordinates) as shown in Figure 15 and the structure coordinates of the divalent cations as shown in Figure 15, and/or (vii) the structure coordinates of the polypeptide fragment (variant) according to the present invention (i.e. amino acid structure coordinates) as shown in Figure 16 and the structure coordinates of the divalent cations as shown in Figure 16 is constructed in step a) of the method of the present invention.

Preferably, the modulating compound associates with, preferably binds to, the endonucleolytically active site within the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus 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 of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or variant thereof decreases said activity, more preferably said compound inhibits said activity. Preferably, the compound decreases the endonucleolytic activity of the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or 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 said 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 of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or 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.

The present invention permits the use of molecular design techniques to identify, select, or design compounds that potentially modulate the endonucleolytic activity of the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or variant thereof, based on the structure coordinates of the (native) endonucleolytically active site according to Figures 1. For the first time, the present invention further permits the optimized identification, selection and design of compounds that potentially modulate the endonucleolytic activity of the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or variant thereof, based on the structure coordinates of the endonucleolytically active site according to Figures 2 to 5, 15 or 16. Said structure coordinates have been achieved from polypeptide fragments (polypeptide fragment variants) according to the present invention which have been co-crystallized with a modulating compound, preferably with 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3) (see Figures 2 and 3), with 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2) (see Figure 4), with ribo-Uridine monophosphate (rUMP) (see Figure 5), with 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1) (see Figure 15), or with [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate (EGCG) (see Figure 16) (see also Tables 1 or 2). The present inventors have namely surprisingly found that the structure coordinates of the polypeptide fragment (variant) of the present invention without bound compound (i.e. native polypeptide fragment (variant)) (see, for example, Figure 1) differ from the structure coordinates of the polypeptide fragment (variant) of the present invention with bound compound (see, for example, Figures 2 to 5, 15 or 16) as said polypeptide fragment (variant) changes its conformation after compound binding. Thus, this new three-dimensional knowledge allows the optimized design of modifications to existing inhibitors in order to improve their potency or the design and optimization of novel inhibitors that effectively block endonuclease activity.

Such predictive models are 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 Influenza A 2009 pandemic H1N1 PA polypeptide fragment (variant) 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 Influenza A 2009 pandemic H1N1 PA subunit or Influenza A 2009 pandemic H1N1 PA subunit 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 Figures 1 to 5, 15 or 16. 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 Influenza A 2009 pandemic H1N1 PA subunit or active sites of Influenza A 2009 pandemic H1N1 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 Influenza A 2009 pandemic H1N1 PA subunit or Influenza A 2009 pandemic H1N1 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 Figures 1 to 5. 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 Influenza A 2009 pandemic H1N1 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 Influenza A 2009 pandemic H1N1 PA subunit or variant thereof by the same computer methods described in detail above.

In one embodiment of the above-described method of the present invention, the endonucleolytically active site of the Influenza A 2009 pandemic H1N1 PA subunit or variant thereof comprises amino acids Glu80, Glu119, Asp108, Ile120, and His41 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41 and Lys34 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41 and Tyr24 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41 and Arg84 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41 and Phe105 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, and Tyr130 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, and Ile38 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another

embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, and Arg124 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, and Tyr24 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, and Arg84 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, and Phe105 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, and Tyr130 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, and Ile38 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38 and Arg124 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Glu26 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124 and Glu26 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Lys134 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, and Lys134 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38,

and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Gly26 and Lys134 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Gly26 and Lys134 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Leu106, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Leu106, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Glu26, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Glu26, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Lys134, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Lys134, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Lys134, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Lys134, and Lys137 of the PA subunit

according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, Lys134, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Glu26, Lys134, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In yet another embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, Lys134, Leu106, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Glu26, Lys134, Leu106, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

In a further embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, and Glu26 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In a further embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, and Lys134 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In a further embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In a further embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In a further embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, and Lys134 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In a further embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, Lys134, and Leu106 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto. In a further embodiment, said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, Lys134, Leu106, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

In one embodiment of the above-described method of the invention, the endonucleolytically active site of the Influenza A 2009 pandemic H1N1 PA subunit or variant thereof is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, and His41 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, and His corresponding to amino acids Glu80, Glu119, Asp108, Ile120, and His41 of SEQ ID NO: 2, respectively, according to

Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41 and Lys34 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, and His41 and Lys34 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41 and Tyr24 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, and Tyr corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41 and Tyr24 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41 and Arg84 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His and Arg corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41 and Arg84 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41 and Phe105 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His and Phe corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41 and Phe105 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41 and Tyr130 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His and Tyr corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41 and Tyr130 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, and Ile38 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His and Ile corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41 and Ile38 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, and Arg124 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, and Arg corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, and Arg124 of SEQ ID NO: 2, respectively, according to Figures 15 or 16.

In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, and Tyr24 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, and Tyr corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, and Tyr24 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, and Arg84 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, and Arg corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, and Arg84 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, and Phe105 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, and Phe corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, and Phe105 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, and Tyr130 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, and Tyr corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, and Tyr130 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, and Ile38 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, and Ile corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, and Ile38 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Arg124 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, and Arg corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Arg124 of SEQ ID NO: 2, respectively, according to Figures 15 or 16.

In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Glu26 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, and Glu corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Glu26 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Lys134 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Lys134 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Leu106 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, and Leu corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Leu106, of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Lys137 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Lys137 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Gly26 and Lys134 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, Glu, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, and Lys134 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Leu106, and

Lys137 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, Leu, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Leu106, and Lys137 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, and Lys137 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, Glu, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, and Lys137 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, and Leu106 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, Glu, and Leu, corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, and Leu106 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Lys134, and Leu106 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, Lys, and Leu corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Lys134, and Leu106 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Lys134, and Lys137 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, Lys, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Lys134, and Lys137 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, Lys134, and Leu106 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg,

Phe, Tyr, Ile, Glu, Lys, and Leu corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, Lys134, and Leu106 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, Lys134, Leu106, and Lys137 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, Glu, Lys, Leu, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, Lys134, Leu106, and Lys137 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In yet another embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Glu26, Lys134, Leu106, and Lys137 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Lys, Tyr, Arg, Phe, Tyr, Ile, Arg, Glu, Lys, Leu, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Arg124, Glu26, Lys134, Leu106, and Lys137 of SEQ ID NO: 2, respectively, according to Figures 15 or 16.

In a further embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, and Glu26 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, and Glu corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, and Glu26 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In a further embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, and Lys134 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, and Lys134 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In a further embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, and Leu106 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, and Leu corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, and Leu106 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In a further embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, and Lys137 according to Figures 1 to 5 or by the structure

coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, and Lys137 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In a further embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, and Lys134 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Glu, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, and Lys134 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In a further embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, Lys134, and Leu106 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Glu, Lys, and Leu corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, Lys134, and Leu106 of SEQ ID NO: 2, respectively, according to Figures 15 or 16. In a further embodiment, said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, Lys134, Leu106, and Lys137 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, His, Glu, Lys, Leu, and Lys corresponding to amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, Lys134, Leu106, and Lys137 of SEQ ID NO: 2, respectively, according to Figures 15 or 16..

The inventors of the present invention have surprisingly found that the amino acids His 41, Glu80, Asp108, Glu119, and Ile120 according to SEQ ID NO: 2 within the active site of Influenza A 2009 pandemic H1N1 PA subunit are required for divalent cation binding (see for example Figures 1 and 6), while the amino acids Tyr24, Lys34, Ile38, Arg84, Phe105, and Tyr130 according to SEQ ID NO: 2 within the active site of Influenza A 2009 pandemic H1N1 PA subunit are required for compound (i.e. inhibitor) binding in the structures shown in Figures 2 to 5 (see also Figures 7 to 10 and Tables 1 and 2) and are those amino acids which conformation changes from the unligated structure (see Figures 1 and 6). In addition, the inventors of the present invention determined that the amino acid residues Leu106, Lys134, Lys137 and Glue26 according to SEQ ID NO: 2 within the active site of Influenza A 2009 pandemic H1N1 PA subunit are also required to make contacts with the tested compounds (see Tables 1 and 2, Figures 2 to 5 and Figures 7 to 10). See also Figures 15 to 19. In addition, the inventors of the present invention have found that the amino acid Arg124 within the active site of Influenza A 2009 pandemic H1N1 PA subunit is required for compound (i.e. inhibitor) interaction in the structures shown in Figures 16 and 18. This three-dimensional knowledge of

the compound/ligand interacting residues and the plasticity of the active site is important for the optimised design of modifications to existing inhibitors in order to improve their potency. In addition, the design, identification and selection of new anti viral compounds that interact with these amino acids is, thus, highly preferable.

If computer modeling according to the methods described hereinabove indicates binding of a compound to the active site of the Influenza A 2009 pandemic H1N1 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, more preferably inhibit, the endonucleolytic activity of the Influenza A 2009 pandemic H1N1 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 according to the present invention or the recombinant host cell according to the present invention and determine the ability of said compound to (i) bind to the active site and/or (ii) modulate, preferably decrease, more preferably inhibit, the endonucleolytic activity of the Influenza A 2009 pandemic H1N1 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 present invention to provide a compound which is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit or variant thereof according to the present invention. It is preferred that the compound is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit or variant thereof according to the present invention by associating with, preferably by binding to, the endonucleolytically active site of said PA subunit or variant thereof.

It is also an aspect of the present invention to provide a compound which is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof according to the present invention. It is preferred that the compound is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit polypeptide fragment or variant thereof according to the present invention by associating with, preferably by binding to, the endonucleolytically active site of said PA subunit polypeptide fragment or variant thereof.

It is further an aspect of the present invention to provide a compound which is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit of other viruses, preferably of viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family. It is preferred that the compound is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit of other viruses, preferably of viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, by associating with, preferably by binding to, the endonucleolytically active site of said PA subunit.

Preferably, the endonucleolytically active site of the above-mentioned PA subunits or variants thereof, or PA subunit polypeptide fragments or variants thereof has the structure as defined above.

Thus, it is preferred that said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, and His41 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

It is more preferred that said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Lys34, Tyr24, Arg84, Phe105, Tyr130, and Ile38 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or comprises amino acids Glu80, Glu119, Asp108, Ile120, His41 Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Arg124 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

It is also more preferred that said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, His41, Glu26, Lys134, Leu106, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

It is most preferred that said active side comprises amino acids Glu80, Glu119, Asp108, Ile120, His41 Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, Glu26, Lys134, Leu106, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto, or comprises amino acids Glu80, Glu119, Asp108, Ile120, His41 Lys34, Tyr24,

Arg84, Phe105, Tyr130, Ile38, Arg124, Glu26, Lys134, Leu106, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

Preferably, the endonucleolytically active site of the above-mentioned PA subunits or variants thereof, or PA subunit polypeptide fragments or variants thereof is defined by the structure coordinates of the PA subunit as mentioned above.

The ability of a compound to associate with, preferably to bind to, the endonucleolytically active site of the afore-mentioned PA subunit or variant thereof, or PA subunit polypeptide fragment or variant thereof, and/or to modulate, preferably to decrease, more preferably to inhibit the endonucleolytic activity of the PA subunit or variant thereof, or PA subunit polypeptide fragment or variant thereof can easily be assessed. For example, the purified PA subunit or 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 or PA subunit polypeptide fragment 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. Further suitable methods are described below.

Preferably, said compound is identifiable by the above-described method. More preferably, said compound is identified by the above-described method.

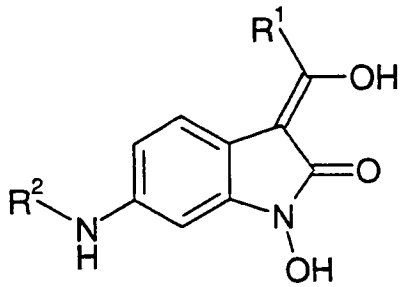
The compound identifiable by the above-described method may be able to modulate the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit or variant thereof. The compound identifiable by the above-described method may be able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof according to the present invention. It is also possible that the compound identifiable by the above-mentioned method may be able to modulate, preferably to decrease, most preferably to inhibit, the

endonuclease activity of the PA subunit of other viruses, preferably of viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family.

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 2,6-diketopiperazine or a derivative thereof, a substituted 2,6-diketopiperazine or a derivative thereof, 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 preferred embodiment of the compound according to the present invention, the compound is not 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid, 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid, 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid, ribo-Uridine

Monophosphate (rUMP), desoxy-thymidinmonophosphat (dTMP), or 2.4-dioxo-4-phenylbutanoic acid (DPBA).

In a further preferred embodiment of the compound according to the present invention, the compound is not a catechin, preferably not a (-)-epigallocatechin gallate (EGCG), such as [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate, a (-)-epicatechin gallate (ECG), (-)-epigallocatechin (EGC), (-)-epicatechin (EC), or gallic acid (GA). In another preferred embodiment of the compound of the present invention, the compound is not a phenethylphenylphthalimide, a phenethylphenylphthalimide analog derived from thalidomide or raltegravir (also a diketobutanoic acid derivative).

In a further aspect, the present invention provides a (an *in vitro*) method for identifying compounds, which modulate, preferably decrease, more preferably inhibit, the endonuclease activity of the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or a variant thereof comprising the steps of:

(i) contacting said polypeptide fragment or variant thereof according to the present invention or said recombinant host cell according to the present invention with a test compound, and

(ii) analyzing the ability of said test compound to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of said PA subunit polypeptide fragment or variant thereof.

It is preferred that said compounds modulate, preferably decrease, more preferably inhibit, the endonuclease activity of the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or a variant thereof by associating with, preferably by binding to, the endonucleolytically active site of said PA subunit.

In another further aspect, the present invention provides a (an *in vitro*) method for identifying compounds, which bind to the endonucleolytically active site, (and) preferably modulate, more preferably decrease, most preferably inhibit, the endonuclease activity of the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or a variant thereof comprising the steps of:

(i) contacting said polypeptide fragment or variant thereof according to the present invention or said 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 site, (and) preferably to modulate, more preferably to decrease, most preferably to inhibit, the endonuclease activity of said PA subunit polypeptide fragment or variant thereof.

In one embodiment, the interaction between the Influenza A 2009 pandemic H1N1 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 or variant thereof according to the invention 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 Influenza A 2009 pandemic H1N1 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 Influenza A 2009 pandemic H1N1 PA polypeptide fragments or variants 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 Influenza A 2009 pandemic H1N1 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 Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variants thereof.

Preferably, the ability of the test compound to inhibit the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof is analyzed in the above-described methods.

In a preferred embodiment, the above-described methods for identifying compounds which associate with, preferably bind to, the endonucleolytically active site, modulate, decrease, and/or inhibit the endonucleolytic activity of the Influenza A 2009 pandemic H1N1

PA subunit polypeptide fragment or variant thereof are performed in a high-throughput setting. In a preferred embodiment, said methods are 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 A 2009 pandemic H1N1 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 A 2009 pandemic H1N1 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 A 2009 pandemic H1N1 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, preferably bind to, the endonucleolytically active site, modulate, decrease, and/or inhibit the endonucleolytic activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof according to the present invention, said methods further comprise the step of formulating said compound or a pharmaceutically acceptable salt thereof with one or more pharmaceutically acceptable excipient(s) and/or carrier(s).

As already mentioned above, it is an aspect of the present invention to provide a compound which is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit or variant thereof according to the present invention. It is preferred that the compound is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit or variant thereof according to the present invention by associating with, preferably by binding to, the endonucleolytically active site of said PA subunit or variant thereof.

It is also an aspect of the present invention to provide a compound which is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof according to the present invention. It is preferred that the compound is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit polypeptide fragment or variant thereof according to the present invention by associating with, preferably by binding to, the endonucleolytically active site of said PA subunit polypeptide fragment or variant thereof.

It is further an aspect of the present invention to provide a compound which is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit of other viruses, preferably of viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family. It is preferred that the compound is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit of other viruses, preferably of viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, by associating with, preferably by binding to, the endonucleolytically active site of said PA subunit.

The ability of a compound to associate with, preferably to bind to, the endonucleolytically active site of the afore-mentioned PA subunit or variant thereof, or PA subunit polypeptide fragment or variant thereof, and/or to modulate, preferably to decrease, more preferably to inhibit the endonucleolytic activity of the PA subunit or variant thereof, or PA subunit polypeptide fragment or variant thereof can easily be assessed. For example, the purified PA subunit or 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 or PA subunit polypeptide fragment 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. Further suitable methods are described below.

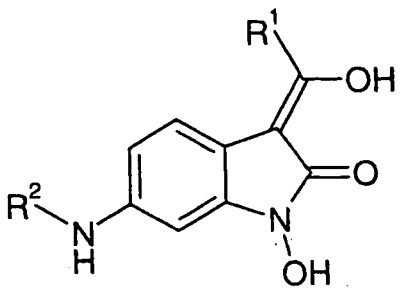
Preferably, said compound is identifiable by the above-described (*in vitro*) methods. More preferably, said compound is identified by the above-described (*in vitro*) methods.

The compound identifiable by the above described (*in vitro*) methods may be able to modulate the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit or variant thereof. The compound identifiable by the above-described (*in vitro*) methods may be able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof according to the present invention.

It is also possible that the compound identifiable by the above-described (*in vitro*) methods may be able to modulate, preferably to decrease, most preferably to inhibit, the endonuclease activity of the PA subunit of other viruses, preferably of viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family.

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, 2,6-diketopiperazine or a derivative thereof, a substituted 2,6-diketopiperazine or a derivative thereof, 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 preferred embodiment of the compound according to the present invention, the compound is not 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid, 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid, 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid, ribo-Uridine Monophosphate (rUMP), desoxy-thymidinmonophosphat (dTMP), or 2.4-dioxo-4-phenylbutanoic acid (DPBA).

In a further preferred embodiment of the compound according to the present invention, the compound is not a catechin, preferably not a (-)-epigallocatechin gallate (EGCG), such as [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate, a (-)-epicatechin gallate (ECG), (-)-epigallocatechin (EGC), (-)-epicatechin (EC), or gallic acid (GA). In another preferred embodiment of the compound of the present invention, the compound is not a phenethylphenylphthalimide, a phenethylphenylphthalimide analog derived from thalidomide or raltegravir (also a diketobutanoic acid derivative).

It is an aspect of the present invention to provide a pharmaceutical composition comprising the afore-mentioned compound of the present invention or a pharmaceutically acceptable salt thereof. It is also an aspect of the present invention to provide a pharmaceutical composition comprising the afore-mentioned compound of the present invention or a pharmaceutically acceptable salt thereof and one or more pharmaceutically acceptable excipient(s) and/or carrier(s).

As already mentioned above, the compound comprised in the pharmaceutical compositions of the present invention (i) is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit or variant thereof according to the present invention, e.g. by binding to the endonucleolytically active site of said subunit or variant thereof, (ii) is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof according to the

present invention, e.g. by binding to the endonucleolytically active site of said polypeptide fragment or variant thereof, or (iii) is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit of other viruses, preferably of viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, e.g. by binding to the endonucleolytically active site of said PA subunit. Preferably, the compound comprised in the pharmaceutical compositions of the present invention is identifiable by the above-described methods. More preferably, the compound comprised in the pharmaceutical compositions of the present invention is identified by the above-described methods.

Preferably, said pharmaceutical composition(s) is (are) producible according to the afore-mentioned methods.

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 Influenza A 2009 pandemic H1N1 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.

In another aspect, the present invention provides an antibody directed against the active site of the PA subunit of the Influenza A 2009 pandemic H1N1 virus according to SEQ ID NO: 2 or a variant thereof, wherein said variant comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto.

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: 4 to 11, i.e., amino acids 20 to 30 (SEQ ID NO: 4), 32 to 40 (SEQ ID NO: 5), 41 to 51 (SEQ ID NO: 6), 80 to 90 (SEQ ID NO: 7), 100 to 110 (SEQ ID NO: 8), 115 to 125 (SEQ ID NO: 9), 130 to 140 (SEQ ID NO: 10), and 176 to 186 (SEQ ID NO: 11) of the amino acid sequence as set forth in SEQ ID NO: 2.

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, e.g. stretches of 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 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, e.g. have a length of 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids, and/or comprise one or more of Lys34, Glu26, Ile38, Tyr24, His41, Glu80, Arg84, Leu106, Asp108, Glu119, Ile120, Arg124, Tyr130, Lys134, Phe105, and Lys137 of SEQ ID NO: 2 or one or more corresponding amino acid(s). Thus, in a preferred embodiment, said antibody recognizes a polypeptide fragment of a length between 5 and 15 amino acids of the amino acid sequence as set forth in SEQ ID NO: 2, wherein the polypeptide fragment comprises one or more amino acid residues selected from the group consisting of Lys34, Glu26, Ile38, Tyr24, His41, Glu80, Arg84, Leu106, Asp108, Glu119, Ile120, Tyr130, Lys134, Phe105, Lys137, and Arg124.

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 Influenza A 2009 pandemic H1N1 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 Influenza A 2009 pandemic H1N1 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 H1N1 PA endonuclease domain or a fragment thereof. Cells that produce antibodies directed against the H1N1 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 the compound(s) according the present invention, the pharmaceutical composition(s) according to the present invention, or the 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, Bunyaviridae family and/or Arenviridae family, preferably caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

Said compound (i) is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit or variant thereof according to the present invention, e.g. by binding to the endonucleolytically active site of said PA subunit or variant thereof, (ii) is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof according to the present invention, e.g. by binding to the endonucleolytically active site of said PA subunit polypeptide fragment or variant thereof, or (iii) is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit of other viruses, preferably of viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, e.g. by binding to the endonucleolytically active site of said PA subunit.

It is preferred that said compound(s) is (are) identifiable by the above-described methods. It is further preferred that said pharmaceutical composition(s) is (are) producible according to the afore-mentioned methods. Thus, preferably, 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 Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof, and/or is able to modulate, preferably to decrease, more preferably to inhibit the endonucleolytic activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof, or the pharmaceutical composition producible according to the afore-mentioned methods for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with viruses of the Orthomyxoviridae family Bunyaviridae family and/or Arenviridae family, preferably caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

In a further aspect, the present invention relates to the compound(s) according the present invention, the pharmaceutical composition(s) according to the present invention, or the antibody according to the present invention for treating, ameliorating, or preventing disease conditions caused by viral infections with viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, preferably caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

Said compound (i) is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit or variant thereof according to the present invention, e.g. by binding to the endonucleolytically active site of said PA subunit or variant thereof, (ii) is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof according to the present invention, e.g. by binding to the endonucleolytically active site of said PA subunit polypeptide fragment or variant thereof, or (iii) is able to modulate, preferably to decrease, more preferably to inhibit, the endonuclease activity of the PA subunit of other viruses, preferably of viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, e.g. by binding to the endonucleolytically active site of said PA subunit.

It is preferred that said compound(s) is (are) identifiable by the above-described methods. It is further preferred that said pharmaceutical composition(s) is (are) producible according to the afore-mentioned methods. Thus, preferably, the present invention relates to a compound identifiable by the above-described methods that is able to bind to the endonucleolytically active site of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof, and/or is able to modulate, preferably to decrease, more preferably to inhibit the endonucleolytic activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof, or the pharmaceutical composition producible according to the afore-mentioned methods for treating, ameliorating, or preventing

disease conditions caused by viral infections with viruses of the Orthomyxoviridae family Bunyaviridae family and/or Arenviridae family, preferably caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

In another aspect, the present invention relates to the use of the compound 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3), of the compound 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2), of the compound 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1), or of the compound [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate (EGCG) for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with Influenza A 2009 pandemic H1N1 virus, as the inventors of the present invention have determined that said compounds are able to bind to the endonucleolytically active site of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof, and are able to modulate, i.e. to inhibit, the endonucleolytic activity of the Influenza A 2009 pandemic H1N1 PA subunit polypeptide fragment or variant thereof.

In an further aspect, the present invention relates to 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3), 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2), 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1), or [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate (EGCG) for treating, ameliorating, or preventing disease conditions caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

In another aspect, the present invention relates to the use of a 4-substituted 2,4-dioxobutanoic acid compound or a green tea catechin for treating, ameliorating, or preventing disease conditions caused by viral infections with Influenza A 2009 pandemic H1N1 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.

Various modifications and variations of the invention will be apparent to those skilled in the art without departing from the scope of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in the relevant fields are intended to be covered by the present invention.

The following figures and examples are merely illustrative of the present invention and should not be construed to limit the scope of the invention as indicated by the appended claims in any way.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-5 (common annotation): Refined atomic structure coordinates for PA polypeptide fragment amino acids 1 to 198 according to amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 (PA H1N1 1 to 198), with or without bound compounds. For each structure there are generally four molecules in the crystallographic asymmetric unit (ASU) denoted A, B, C and D. In Figures 1 to 5, however, only one selected molecule is shown (see below). 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 equivalent molecules of the crystal. B is a thermal factor that measures movement of the atom around its atomic center. This nomenclature corresponds to the Protein Data Bank (PDB) format.

Figure 1: Structural co-ordinates of the native (i.e. without compound/ligand) H1N1 PA endonuclease domain in standard Protein Data Bank (PDB) format. Only the chain A (with associated divalent cations, i.e. one magnesium and one manganese ion, and water molecules) from the asymmetric unit is included. Chains B, C and D are very similar.

Figure 2: Structural co-ordinates of the H1N1 PA endonuclease domain with bound 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3) in standard Protein Data Bank (PDB) format. Only the chain A (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. The compound EMBL-R05-3 has residue descriptor *ci3*.

Figure 3: Structural co-ordinates of the H1N1 PA endonuclease domain with bound 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3) in standard Protein Data Bank (PDB) format. Only the chain D (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. The compound EMBL-05-03 has a different configuration than in chain A (Figure 2). Said compound has residue descriptor *ci3*.

Figure 4: Structural co-ordinates of the H1N1 PA endonuclease domain with bound 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2) in standard Protein Data Bank (PDB) format. Only the chain A (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. Chains B, C and D are very similar. The compound EMBL-R05-2 has residue descriptor *ci2*.

Figure 5: Structural co-ordinates of the H1N1 PA endonuclease domain with bound ribo-Uridine Monophosphate (rUMP) in standard Protein Data Bank (PDB) format. Only the chain A (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. Chains B, C and D are very similar. The compound rUMP has residue descriptor *U*.

Figure 6: Diagram using the structural co-ordinates of Figure 1 illustrating the endonuclease active site (PA polypeptide fragment (chain A)), showing the divalent cations (one manganese and one magnesium) and key active site residues.

Figure 7: (A) Diagram using the structural co-ordinates of Figure 2 illustrating the endonuclease active site (PA polypeptide fragment (chain A)), showing the bound compound EMBL-R05-3, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Arg84) or are close to it and **(B)** Diagram comparing the co-ordinates of Figure 1 (chain subjacent on left hand side) and Figure 2 showing change in conformation of the loop in the vicinity of Tyr24 upon binding of EMBL-R05-3 (chain A). Tyr24 side-chain moves to partially stack with the chlorobenzene of EMBL-R05-3.

Figure 8: Diagram using the structural co-ordinates of Figure 3 illustrating the endonuclease active site (PA polypeptide fragment (chain D)), showing the bound compound EMBL-R05-3, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Tyr24 and Arg84) or are close to it.

Figure 9: Diagram using the structural co-ordinates of Figure 4 illustrating the endonuclease active site (PA polypeptide fragment (chain A)), showing the bound compound EMBL-R05-2, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Tyr24 and Phe105) or are close to it.

Figure 10: Diagram using the structural co-ordinates of Figure 5 illustrating the endonuclease active site (PA polypeptide fragment (chain A)), showing the bound compound rUMP, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Tyr24 and Lys34) or are close to it.

Figure 11: 15% PAGE gels and gel filtration profile from a typical purification of H1N1 PA1-198. The arrow indicates faint traces of residual MBP.

Figure 12: Frozen crystal of H1N1 PA-Nter co-crystallised with rUMP in the P212121 space-group.

Figure 13: Divalent ion co-ordination in the native endonuclease structure.

Figure 14: Electron density for rUMP and divalent cations (manganese, Mn1 and Mn2) in co-crystals with H1N1 PA-Nter. Refined 2Fo-Fc electron density contoured at 1.1 σ. Unbiased Fo-Fc electron density contoured at 2.8 σ. Anomalous difference map contoured at 4 σ.

Figures 15-16 (common annotation): Refined atomic structure coordinates for PA polypeptide fragment amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 64 replaced by the amino acid glycine (PA H1N1 1 to 198 Δ52-64: Gly) with bound compounds. For each structure there is one molecule in the crystallographic asymmetric unit (ASU). 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 equivalent molecules of the crystal. B is a thermal factor that measures movement of the atom around its atomic center. This nomenclature corresponds to the Protein Data Bank (PDB) format.

Figure 15: Structural co-ordinates of the H1N1 PA endonuclease domain with bound 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1) in standard Protein Data Bank (PDB) format. The chain A (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. The compound EMBL-R05-1 has residue descriptor *ci1*.

Figure 16: Structural co-ordinates of the H1N1 PA endonuclease domain with bound epigallocatechin 3-gallate (EGCG) in standard Protein Data Bank (PDB) format. The chain A (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. The compound EGCG has residue descriptor *tte*.

Figure 17: Diagram using the structural co-ordinates of Figure 15 illustrating the endonuclease active site (PA polypeptide fragment), showing the bound compound EMBL-R05-1, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Tyr24 and Arg84) or are close to it.

Figure 18: (A) Diagram using the structural co-ordinates of Figure 16 illustrating the endonuclease active site (PA polypeptide fragment), showing the bound compound EGCG, the divalent cations (two manganese ions) and key active site residues that interact with the compound or are close to it. **(B)** Bound EGCG, the divalent cations (two manganese ions) and key active site residues that interact with the compound or are close to it. Interactions less than 3.3 Å (grey dotted lines), additional possible interactions less than 4 Å (dark grey dotted lines).

Figure 19: Superposition of all diketo inhibitor compounds (EMBL-R05-1, EMBL-R05-2, and EMBL-R05-3A and D) and EGCG bound in PA active site. As shown in Figure 19, the mode of binding of the three diketo inhibitors to the metals is conserved (although there is some variability in exact position) but the two 'arms' of each compound are inserted into different combinations of the pockets 1 to 4. EMBL-R05-1 has a similar configuration to EMBL-R05-3D, with the two arms occupying pockets 2 and 3. EMBL-R05-3A occupies pockets 2 and 4. EMBL-R05-2, which differs notably from R05-1 and R05-3 in the point of substitution on the piperidinyl ring (Table 2) occupies pocket 3 and uniquely pocket 1 (Figure 7B). The green tea compound EGCG occupies pockets 3 and 4. More potent and specific compounds could be perhaps designed that occupy more than two of the pockets.

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.

1. Methods

1.1 Cloning, expression and purification of PA-Nter (PA H1N1 1 to 198) and PA-Nter mutant (PA H1N1 1 to 198 Δ52-64: Gly) from influenza strain A/California/04/2009-H1N1

The DNA coding for PA-Nter (residues 1-198) (see SEQ ID NO: 1 and 2) from influenza strain A/California/04/2009-H1N1 was synthesized (PA H1N1 1 to 198) and sub-cloned in the expression vector pESPRIT002 (EMBL) by GeneArt, Regensburg, Germany. The sequence was designed to contain a MGSGMA (SEQ ID NO: 3) polypeptide linker between the tobacco etch virus (TEV) cleavage site at the N-terminus to obtain 100% cleavage by TEV protease.

To further improve crystallisation properties, a deletion of part of the flexible loop (52-73) was engineered by site directed mutagenesis. To this end, a PCR amplification of the whole vector containing the wild-type gene was performed using two primers flanking the mutation site, one of them phosphorylated, and TurboPfu polymerase (Stratagene). Subsequently template vector was digested with DpnI (New England Biolabs) and the mutated vector was re-ligated. In the PA-Nter mutant, the amino acid sequence encompassing amino acids 52-64 was replaced by a single glycine (PA H1N1 1 to 198 Δ52-64: Gly).

The wild-type and mutant plasmids were transformed to *E. coli* BL21(DE3) (Stratagene) and the protein was expressed in LB medium overnight at 20°C after induction at an OD 0.8-1.0 with 0.2 mM isopropyl-β-thiogalactopyranoside (IPTG). The protein was purified by an immobilized metal affinity column (IMAC). A second IMAC step was performed after cleavage by the His-tagged TEV protease, followed by gel filtration on a Superdex 75 column (GE Healthcare). Finally, the protein was concentrated to 10-15 mg/ml. See **Figure 11**.

1.2 Compounds

Compounds used for co-crystallisation are given in **Table 2**. Compound rUMP was purchased from Sigma. Compounds EMBL-R05-1, EMBL-R05-2, and EMBL-R05-3 (first

described in {Tomassini, 1994 #397}) were custom synthesised by Shanghai ChemPartner. EGCG was purchased from Sigma (E4143).

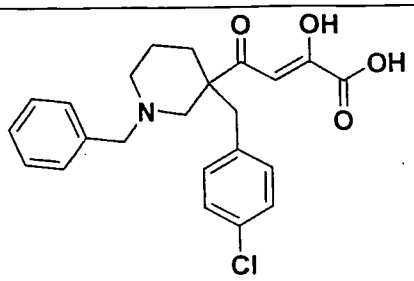
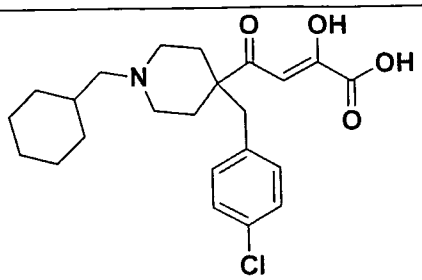
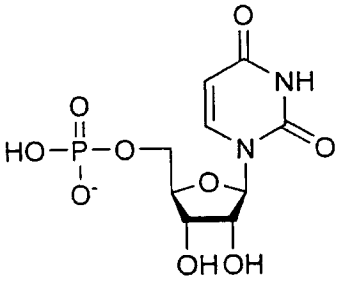
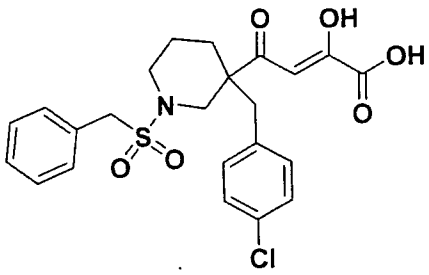
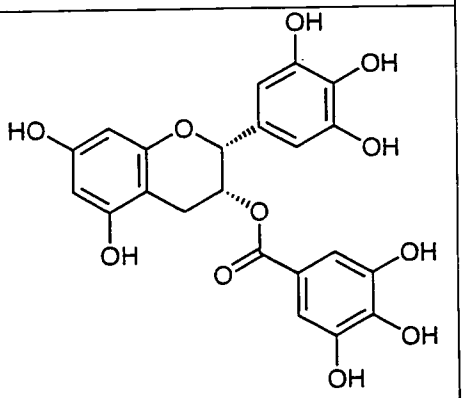
1.3 Crystallization

Initial sitting drop screening was carried out at 20°C mixing 100 nL of protein solution (15 mg/mL) with 100 nL of reservoir solution using a Cartesian robot at the EMBL Grenoble crystallization platform. Around 600 conditions were screened. Subsequently, larger crystals were obtained at 20°C by the hanging drop method mixing protein and reservoir solutions in a ratio of 1:1. The protein solution contained 10-15 mg/mL of PA-N-ter in 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM $MnCl_2$, 2 mM $MgCl_2$. The refined reservoir compositions for native crystals and co-crystallization with different compounds/ligands are listed in **Table 1**. For co-crystallisations, compounds/ligands rUMP, EMBL-R05-3 and EGCG were added to the protein solution to final concentrations of 5mM, 1.5mM and 5mM, respectively. Native crystals and those co-crystallized with EMBL-R05-3 and EGCG were flash frozen in liquid nitrogen in their reservoir solution with additional 25% glycerol as cryo-protectant. Co-crystals with rUMP were frozen in their reservoir solution with additonal rUMP at 10mM concentration and additional 20% glycerol as cryoprotectant. The structure with EMBL-R05-2 was obtained by soaking co-crystals of PA-N-ter and rUMP for 2h with reservoir solution containing 1.5mM EMBL-R05-2 and no rUMP (the inhibitor displaces the rUMP) followed by cryo protection in reservoir solution containing 20% glycerol and 1.5mM EMBL-R05-2. **Figure 12** shows a typical co-crystal with rUMP. The structure with EMBL-R05-1 was obtained by soaking co-crystals of PA-N-ter mutant and dTMP for 2h with reservoir solution containing the inhibitor followed by cryo protection in reservoir solution containing 20% glycerol and the inhibitor. See also **Table 2** for the compounds/ligands.

Table 1: Summary of crystallisation conditions and crystallographic parameters for various compounds

Fragment/ Figure(s)	Compound/ Ligand (final concentration)	Crystallisation reservoir condition	Space group and unit cell parameters	Resolution Refinement R-factor/R-free
PA H1N1 1 to 198 fragment (chain A), see **Figures 1** and **6**	No compound (native)	1.6 M sodium formate 0.1 M HEPES pH7 5% Glycerol	*C*2 4 Molecules/ASU 263.630 66.240 66.32 90.00 95.98 90.00	2.1 Å 0.222 /0.268
PA H1N1 1 to 198 fragment (chain A), see **Figures 2** and **7** PA H1N1 fragment (chain D), see **Figures 3** and **8**	EMBL-R05-3 (1.5 mM)	2.0M ammonium sulphate 0.1M BisTris pH5.5	*P* 2_1 2_1 2_1 4 Molecules/ASU 54.57 122.54 129.78 90.00 90.00 90.00	2.5 Å 0.205 /0.276
PA H1N1 1 to 198 fragment (chain A), see **Figures 4** and **9**	EMBL-R05-2 (1.5 mM, soaked into crystals initially grown with rUMP)	0.1 M ammonium sulphate 0.1 M Bis-Tris pH5.5 25 % (w/v) PEG 3350	*P* 2_1 2_1 2_1 4 Molecules/ASU 56.59 120.81 128.20 90.00 90.00 90.00	2.07 Å 0.205 /0.259
PA H1N1 1 to 198 fragment (chain A), see **Figures 5** and **10**	rUMP (5.0 mM)	0.1 M ammonium sulphate 0.1 M Bis-Tris pH5.5 25 % (w/v) PEG 3350	*P* 2_1 2_1 2_1 4 Molecules/ASU 54.94 120.11 128.05 90.00 90.00 90.00	2.05 Å 0.206 /0.246
PA H1N1 1-198 Δ52-64: Gly Fragment (chain A) see **Figures 15** and **17**	EMBL-R05-1 (1.5 mM, soaked into crystals initially grown with dTMP)	25-30% PEG4K 0.1M Tris pH8.5 0.2M NaCl	*P* $6_2$22 (180) 1 molecule/ASU a=b= 75.06 c=120.05 Å	1.9 Å 0.201/0.251
PA H1N1 1-198 Δ52-64: Gly Fragment (chain A) see **Figures 16** and **18**	EGCG (5mM)	10% peg 3350, 0.1M NaCl, 0.1M Hepes, pH 7.0	*P6₄22* (181) 1 molecule/ASU a=b= 99.9 c=82.7 Å	2.65 Å 0.250 /0.304

Table 2: Compounds used

Compound	CA Index Name	Formula	Chemical structure
EMBL-R05-3 (Tomassini et al. Antimicrob Agents Chemother 1994, 38:2827-2837)	4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid	C23 H24 Cl N O4	
EMBL-R05-2 (Tomassini et al. Antimicrob Agents Chemother 1994, 38:2827-2837)	4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid	C23 H30 Cl N O4	
rUMP (ribo-uridine monophosphate)	5'-Uridylic acid	C9 H13 N2 O9 P	
EMBL-R05-1 (Tomassini et al. Antimicrob Agents Chemother 1994, 38:2827-2837)	4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid	C23 H24 Cl N S O6	
(-)-Epigallocatechin gallate (EGCG)	[(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate	C22 H18 O11	

1.4 Crystal structure determination

Diffraction data were collected on various beamlines at the European Synchrotron Radiation Facility. Data sets were integrated with XDS and scaled with XSCALE. Subsequent data analysis was performed with the CCP4i programme suite. The initial H1N1 structure was solved with molecular replacement with PHASER using the previously determined H3N2 PA N-ter structure (Dias et al., Nature 2009). Subsequent co-crystal structures were determined with PHASER using the H1N1 structure. Refinement was carried out with REFMAC and/or model building with COOT or O. In the *C*2 and $P2_12_12_1$ crystal forms there are four molecules per asymmetric unit. However due to structural variations between the molecules due to plasticity (in particular the 53-73 region) and the generally good resolution, NCS restraints were not applied.

2. Results

2.1 PA H1N1 polypeptide fragment or PA H1N1 polypeptide fragment variant generation and crystallisation

The inventors of the present invention found that a polypeptide fragment comprising amino acids 1 to 198 of SEQ ID NO: 2 of influenza strain A/California/04/2009-H1N1 (2009 pandemic strain) (PA H1N1 1 to 198) readily crystallised with and without relevant compounds/ligands (see **Figures 1 to 10**). Crystallisation properties could further be improved with a polypeptide fragment variant comprising amino acids 1 to 198 of SEQ ID NO: 2 of influenza strain A/California/04/2009-H1N1 (2009 pandemic strain) with amino acids 52 to 64 replaced by the amino acid glycine (PA H1N1 1 to 198 Δ52-64: Gly) (see **Figures 15 to 18**).

Thus, in contrast to the numerous unsuccessful attempts undertaken in the prior art, the inventors of the present invention were able to obtain structures of PA H1N1 polypeptide fragments or PA H1N1 polypeptide fragment variants with and without compounds/ligands.

The structures of PA H1N1 polypeptide fragments or PA H1N1 polypeptide fragment variants with and without compounds/ligands are described below. All these structures show in detail how these compounds/ligands bind directly to the metal ions as well as interacting with a number of residues in the active site. Furthermore several of the interacting residues change conformation upon ligand binding, information which was unavailable before. This three-dimensional knowledge of the ligand interacting residues and the regions of plasticity of the active site is critical for the optimised design of modifications to existing inhibitors to

improve their potency or for structure based design and optimisation of novel inhibitors that effectively block endonuclease activity.

2.2 H1N1 PA native structure

Structural co-ordinates of the native (i.e. without compound/ligand) H1N1 PA endonuclease domain (PA H1N1 1 to 198) (chain A) in standard Protein Data Bank (PDB) format are shown in **Figure 1**. Only the chain A (with associated divalent cations, i.e. one magnesium and one manganese ion, and water molecules) from the asymmetric unit is included. Chains B, C and D are very similar. A diagram using the structural co-ordinates of **Figure 1** illustrating the endonuclease active site (PA polypeptide fragment (chain A)) is shown in **Figure 6**. It shows the divalent cations (one manganese and one magnesium) and key active site residues.

2.3 EMBL-R05-3-bound structure

Structural co-ordinates of the H1N1 PA endonuclease domain (PA H1N1 1 to 198) (chain A) with bound 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3) in standard Protein Data Bank (PDB) format are shown in **Figure 2**. Only the chain A (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. The compound EMBL-R05-3 has residue descriptor *ci3*. A diagram using the structural co-ordinates of **Figure 2** illustrating the endonuclease active site (PA polypeptide fragment (chain A)) is shown in **Figure 7A**. It shows the bound compound EMBL-R05-3, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Arg84) or are close to it.

A diagram comparing the co-ordinates of **Figure 1** (chain subjacent on left hand side) and **Figure 2** showing change in conformation of the loop in the vicinity of Tyr24 upon binding of EMBL-R05-3 (chain A) is shown in **Figure 7B**. The loop around Tyr24 is poorly ordered in the native structure. Tyr24 side-chain moves to partially stack with the chlorobenzene of EMBL-R05-3. This indicates a plasticity of the active site and an induced fit mode of ligand binding. An important conclusion for designing more potent inhibitors is to ensure that the extensions ('arms') to any ion-binding scaffold optimise interactions in one or more pocket(s). Imperfect matching will lead to residual flexibility and sub-optimal potency, as seems to be the case for the current compounds, none of which exhibit very well defined, full occupancy binding modes.

Further, structural co-ordinates of the H1N1 PA endonuclease domain (PA H1N1 1 to 198) (chain D) with bound 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3) in standard Protein Data Bank (PDB) format are shown in **Figure 3**. Only the chain D (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. The compound EMBL-05-3 has a different configuration than in chain A (**Figure 2**). Said compound has residue descriptor *ci3*. A diagram using the structural co-ordinates of **Figure 3** illustrating the endonuclease active site (PA polypeptide fragment (chain D)) is shown in **Figure 8**. It shows the bound compound EMBL-R05-3, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Tyr24 and Arg84) or are close to it.

2.4 EMBL-R05-2-bound structure

Structural co-ordinates of the H1N1 PA endonuclease domain (PA H1N1 1 to 198) (chain A) with bound 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2) in standard Protein Data Bank (PDB) format are shown in **Figure 4**. Only the chain A (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. Chains B, C and D are very similar. The compound EMBL-R05-2 has residue descriptor *ci2*. A diagram using the structural co-ordinates of **Figure 4** illustrating the endonuclease active site (PA polypeptide fragment (chain A)) is shown in **Figure 9**. It shows the bound compound EMBL-R05-2, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Tyr24 and Phe105) or are close to it.

2.5 Ribo-Uridine Monosphosphate (rUMP)-bound structure

Structural co-ordinates of the H1N1 PA endonuclease domain (PA H1N1 1 to 198) (chain A) with bound ribo-Uridine Monophosphate (rUMP) in standard Protein Data Bank (PDB) format are shown in **Figure 5**. Only the chain A (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. Chains B, C and D are very similar. The compound rUMP has residue descriptor *U*.

Co-crystallisation trials with rUMP gave large, well-ordered crystals in a new orthorhombic space-group (**Figures 12** and **14**).

A diagram using the structural co-ordinates of **Figure 5** illustrating the endonuclease active site (PA polypeptide fragment (chain A)) is shown in **Figure 10**. It shows the bound compound rUMP, the divalent cations (two manganese) and key active site residues that

interact with the compound (notably Tyr24 and Lys34) or are close to it. The rUMP binds with two oxygens of the phosphate completing the co-ordination sphere of Mn1, one of them also co-ordinating Mn2. The base is well stacked on Tyr24 and Lys34 makes a hydrogen bond to the O2 position. The ribose hydroxyl groups do not make hydrogen bonds to the protein, consistent with the fact that deoxy ribose binds equally well and the protein is a DNAase as much as an RNAase {Dias, 2009 #448}.

The conformation observed for rUMP is quite different from that previously published (PDB entry 3hw3 {Zhao, 2009 #444}). The latter structure was obtained by soaking nucleotides into existing crystals of the endonuclease in the absence of manganese and the electron density is very poor. In this structure, a water molecule replaces Mn1 and a magnesium ion replaces Mn2. This difference in metal ligation is reflected in the altered conformation of Glu119. The ribose and base positions are quite different from the positions in the structure provided herein and unable to interact with Lys34 or Tyr24. The differences between the two structures may reflect firstly the lack of manganese and secondly the fact that soaking pre-grown crystals does not allow the active site to adapt to the ligand as is more likely the case for co-crystallisation.

2.6 EMBL-R05-1-bound structure

Structural co-ordinates of the H1N1 PA endonuclease domain (PA H1N1 1 to 198 Δ52-64: Gly) with bound 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1) in standard Protein Data Bank (PDB) format are shown in **Figure 15**. The chain (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. The compound EMBL-R05-1 has residue descriptor *cil*. A diagram using the structural co-ordinates of **Figure 15** illustrating the endonuclease active site (PA polypeptide fragment) is shown in **Figure 17**. It shows the bound compound EMBL-R05-1, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Tyr24 and Arg84) or are close to it.

2.7 EGCG-bound structure

Epigallocatechin 3-gallate (EGCG), is the ester of epigallocatechin and gallic acid and is the most abundant catechin in green tea. It has recently been reported that EGCG inhibits the influenza endonuclease{Kuzuhara, 2009 #629}. Co-crystallisation of PA H1N1 1 to 198 Δ52-64: Gly with (-)-Epigallocatechin gallate gave a new crystal form (**Table 1**) diffracting to

2.65 Å resolution. The compound was clearly observed in the active site. Strong extra density also exists around a 2-fold crystallographic axis and most likely represents other EGCG molecules trapped by crystal packing.

Structural co-ordinates of the H1N1 PA endonuclease domain (PA H1N1 1 to 198 Δ52-64: Gly) with bound epigallocatechin 3-gallate (EGCG) in standard Protein Data Bank (PDB) format is shown in **Figure 16**. The chain (with associated divalent cations, i.e. two manganese ions, and water molecules) from the asymmetric unit is included. The compound EGCG has residue descriptor *tte*. A diagram using the structural co-ordinates of **Figure 16** illustrating the endonuclease active site (PA polypeptide fragment) is shown in **Figure 18A**. It shows the bound compound EGCG, the divalent cations (two manganese ions) and key active site residues that interact with the compound or are close to it. **Figure 18B** shows the bound EGCG, the divalent cations (two manganese ions) and key active site residues that interact with the compound or are close to it. Interactions less than 3.3 Å (grew dotted lines), additional possible interactions less than 4 Å (dark grey dotted lines) are shown.

2.8 Superposition of all diketo inhibitor compounds

A superposition of all diketo inhibitor compounds (EMBL-R05-1, EMBL-R05-2, and EMBL-R05-3A and D) and EGCG bound in PA active site is shown in **Figure 19**. As shown in **Figure 19**, the mode of binding of the three diketo inhibitors to the metals is conserved (although there is some variability in exact position) but the two 'arms' of each compound are inserted into different combinations of the pockets 1 to 4. EMBL-R05-1 has a similar configuration to EMBL-R05-3D, with the two arms occupying pockets 2 and 3. EMBL-R05-3A occupies pockets 2 and 4. EMBL-R05-2, which differs notably from R05-1 and R05-3 in the point of substitution on the piperidinyl ring (**Table 2**) occupies pocket 3 and uniquely pocket 1 (**Figure 7B**). The green tea compound EGCG occupies pockets 3 and 4. More potent and specific compounds could be perhaps designed that occupy more than two of the pockets.

2.9 Divalent cation binding

In the native structure (**Figure 1** and **Figure 6**), two divalent cations are observed in the active site. These are identified (using magnitude of electron density and anomalous scattering, data not shown) to be a manganese atom in site 1 (Mn1 in **Figure 13**), ligated by His41, Asp108, Glu119, Ile120 and two water molecules (W4 and W5) and a magnesium ion in site 2 (Mg2 in **Figure 13**), ligated by Glu80 and Asp108 and four water molecules (W1, W2, W3 and W4). In the other structures (**Figures 2-5** and **Figures 7-10** as well as **Figures**

15-18) the ions in both site 1 and site 2 are manganese ions (Mn1 and Mn2), as identified by magnitude of electron density and anomalous scattering (e.g. for rUMP, see **Figure 14**).

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 Influenza A 2009 pandemic H1N1 virus according to SEQ ID NO: 2 or is a variant thereof, wherein said variant comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto.

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 any one of claims 1 to 3, wherein the amino terminal fragment corresponds to at least amino acids 1 to 190 of the PA subunit of the RNA-dependent RNA polymerase of Influenza A 2009 pandemic H1N1 virus according to SEQ ID NO: 2.

5. The polypeptide fragment of any one of claims 1 to 4, wherein said polypeptide fragment is purified to an extent to be suitable for crystallization.

6. The polypeptide fragment of any one of claims 1 to 5 to which two divalent cations are bound, wherein the divalent cation is preferably manganese and/or magnesium.

7. The polypeptide fragment of any one of claims 1 to 6, 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 190 to 198 of the amino acid sequence of the PA subunit according to SEQ ID NO: 2, or is a variant thereof, wherein said variant comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto and retains the endonuclease activity.

8. The polypeptide fragment of any one of claims 1 to 7, which

(a) consists of amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 and optionally of an amino-terminal linker having the amino acid sequence MGSGMA (SEQ ID NO: 3) and which has the structure defined by (i) the structure coordinates as shown in Figure 1, (ii) the structure coordinates as shown in Figure 2, (iii) the structure coordinates as shown in Figure 3, (iv) the structure coordinates as shown in Figure 4, or (v) the structure coordinates as shown in Figure 5, or

(b) consists of amino acids 1 to 198 of the amino acid sequence set forth in SEQ ID NO: 2 with amino acids 52 to 64 replaced by the amino acid glycine and optionally of an amino-terminal linker having the amino acid sequence MGSGMA (SEQ ID NO: 3) and which has the structure defined by (vi) the structure coordinates as shown in Figure 15, or (vii) the structure coordinates as shown in Figure 16.

9. The polypeptide fragment of claim 8, wherein the polypeptide fragment having the structure defined by (i) has a crystalline form with space group *C*2 and unit cell dimensions of a = 26.36 nm ± 0.5 nm, b = 6.62 nm ± 0.3 nm, c = 6.63 nm ± 0.3 nm, α = 90 deg, β = 96 ± 2 deg, γ = 90 deg, (ii) to (v) has a crystalline form with space group $P2_12_12_1$ and unit cell dimensions of a = 5.46 ± 0.3 nm, b= 12.25 ± 0.4 nm, c=13.0 ± 0.3 nm, α = 90 deg, β = 90 deg, γ = 90 deg, (vi) has a crystalline form with space group $P\,6_222$ and unit cell dimensions of a = 7.50 nm ± 0.3 nm, b = 7.50 nm ± 0.3 nm, c = 12.00 nm ± 0.5 nm, α = 90 deg, β = 90 deg, γ = 120 deg, or (vii) has a crystalline form with space group $P6_422$ and unit cell dimensions of a = 9.99 nm ± 0.5 nm, b = 9.99 nm ± 0.5 nm, c = 8.27 nm ± 0.3 nm, α = 90 deg, β = 90 deg, γ = 120 deg.

10. The polypeptide fragment of claims 8 or 9, wherein the crystal diffracts X-rays to a resolution of 2.6 Å or higher, preferably 2.1 Å or higher or 1.9 Å or higher.

11. An isolated polynucleotide coding for an isolated polypeptide fragment of any one of claims 1 to 10.

12. A recombinant vector comprising said isolated polynucleotide of claim 11.

13. A recombinant host cell comprising said isolated polynucleotide of claim 11 or said recombinant vector of claim 12.

14. A method for identifying compounds, which modulate the endonuclease activity of the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or a variant thereof comprising the steps of:

(a) constructing a computer model of the active site defined by (i) the structure coordinates of the polypeptide fragment of claim 8 as shown in Figure 1, (ii) the structure coordinates of the polypeptide fragment of claim 8 as shown in Figure 2, (iii) the structure coordinates of the polypeptide fragment of claim 8 as shown in Figure 3, (iv) the structure coordinates of the polypeptide fragment of claim 8 as shown in Figure 4, (v) the structure coordinates of the polypeptide fragment of claim 8 as shown in Figure 5, (vi) the structure coordinates of the polypeptide fragment of claim 8 as shown in Figure 15, or (vii) the structure coordinates of the polypeptide fragment of claim 8 as shown in Figure 16,

(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.

15. The method of claim 14, wherein said active site comprises amino acids Glu80, Glu119, Asp108, Ile120, and His41 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

16. The method of claim 15, wherein said active site further comprises amino acids Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Arg124 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

17. The method of claims 15 or 16, wherein said active site further comprises amino acids Glu26, Lys134, Leu106, and Lys137 of the PA subunit according to SEQ ID NO: 2 or amino acids corresponding thereto.

18. The method of claim 14, wherein said active site is defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu80, Glu119, Asp108, Ile120, and His41 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Glu, Asp, Ile, and His corresponding to amino acids Glu80, Glu119, Asp108, Ile120, and His41 of SEQ ID NO: 2, respectively, according to Figures 15 or 16.

19. The method of claim 18, wherein said active site is further defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Arg124 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Lys, Tyr, Arg, Phe, Tyr, Ile, and Arg corresponding to amino acids Lys34, Tyr24, Arg84, Phe105, Tyr130, Ile38, and Arg124 of SEQ ID NO: 2, respectively, according to Figures 15 or 16.

20. The method of claims 18 or 19, wherein said active site is further defined by the structure coordinates of the PA subunit SEQ ID NO: 2 amino acids Glu26, Lys134, Leu106, and Lys137 according to Figures 1 to 5 or by the structure coordinates of the PA subunit amino acids Glu, Lys, Leu, and Lys corresponding to amino acids Glu26, Lys134, Leu106, and Lys137 of SEQ ID NO: 2, respectively, according to Figures 15 or 16.

21. The method of any one of claims 14 to 20 further comprising the 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).

22. The method of claim 21 further comprising the step of:

(f) contacting said compound and said polypeptide fragment or variant thereof of any one of claims 1 to 10 or said recombinant host cell according to claim 13 and determine the ability of said compound to modulate the endonuclease activity of said PA subunit polypeptide fragment or variant thereof.

23. A method for identifying compounds, which modulate the endonuclease activity of the PA subunit of a RNA-dependent RNA polymerase from Influenza A 2009 pandemic H1N1 virus or a variant thereof comprising the steps of:

(i) contacting said polypeptide fragment or variant thereof of any one of claims 1 to 10 or said recombinant host cell of claim 13 with a test compound, and

(ii) analyzing the ability of said test compound to modulate the endonuclease activity of said PA subunit polypeptide fragment or variant thereof.

24. The method of claim 23, wherein the ability of said test compound to inhibit the endonuclease activity of said PA subunit polypeptide fragment or variant thereof is analyzed.

25. The method of claims 23 or 24, wherein said method is performed in a high-throughput setting.

26. The method of any one of claims 23 to 25, wherein said test compound is a small molecule.

27. The method of any of one claims 23 to 26, wherein said test compound is a peptide or protein.

28. The method of any one of claims 23 to 27, 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).

29. A compound which is able to modulate, preferably to inhibit, the endonuclease activity of the PA subunit polypeptide fragment or variant thereof of any one of claims 1 to 10.

30. The compound of claim 29, wherein said compound is identifiable by the method of any one of claims 14 to 22 or by the method of any one of claims 23 to 28.

31. A pharmaceutical composition comprising the compound of claims 29 or 30 or a pharmaceutically acceptable salt thereof and one or more pharmaceutically acceptable excipient(s) and/or carrier(s).

32. The pharmaceutical composition of claim 31, wherein said pharmaceutical composition is producible according to the method of claims 21 or 28.

33. An antibody directed against the active site of the PA subunit of the Influenza A 2009 pandemic H1N1 virus according to SEQ ID NO: 2 or a variant thereof, wherein said variant comprises the amino acid serine at an amino acid position 186 according to SEQ ID NO: 2 or at an amino acid position corresponding thereto.

34. The antibody of claim 33, wherein said antibody recognizes a polypeptide fragment of a length between 5 and 15 amino acids of the amino acid sequence as set forth in SEQ ID NO: 2, wherein the polypeptide fragment comprises one or more amino acid residues selected from the group consisting of Lys34, Glu26, Ile38, Tyr24, His41, Glu80, Arg84, Leu106, Asp108, Glu119, Ile120, Tyr130, Lys134, Phe105, Lys137, and Arg124.

35. Use of a compound according to claims 29 or 30, a pharmaceutical composition according to claims 31 or 32, or an antibody according to claims 33 or 34 for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, preferably caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

36. Use of 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3), 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2), 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-

oxo-2-butenoic acid (EMBL-R05-1), or [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate (EGCG) for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

37. A compound according to claims 29 or 30, a pharmaceutical composition according to claims 31 or 32, or an antibody according to claims 33 or 34 for treating, ameliorating, or preventing disease conditions caused by viral infections with viruses of the Orthomyxoviridae family, Bunyaviridae family and/or Arenviridae family, preferably caused by viral infections with Influenza A 2009 pandemic H1N1 virus.

38. 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-3), 4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-2), 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (EMBL-R05-1), or [(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl]3,4,5-trihydroxybenzoate (EGCG) for treating, ameliorating, or preventing disease conditions caused by viral infections with Influenza A 2009 pandemic H1N1 virus.