

## SOLUBLE FRAGMENTS OF INFLUENZA VIRUS PB2 PROTEIN CAPABLE OF BINDING RNA-CAP

## TECHNICAL FIELD OF INVENTION

5 The present invention relates to soluble fragments of the Influenza virus RNA dependent RNA polymerase subunit PB2 and variants thereof, and crystallized complexes thereof comprising an RNA cap analog. This invention also relates to computational methods using the structural coordinates of said complex to screen for and design compounds that interact with the RNA cap binding pocket. In addition, this invention relates to methods identifying compounds that bind to  
10 PB2 polypeptide fragments comprising the RNA cap binding pocket, preferably inhibit the interaction with RNA caps or analogs thereof, by using said PB2 polypeptide fragments, preferably in a high-throughput setting. This invention also relates to compounds and pharmaceutical compositions comprising the identified compounds for the treatment of disease conditions due to viral infections caused by negative-sense single stranded RNA viruses.

15

## 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.  
20 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.

25

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) led directly to the development of neuraminidase inhibitors with anti-viral activity preventing the release of virus from the cells,  
30 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). In addition, more unspecific viral drugs, such as ribavirin, have been shown to work for treatment of Influenza infections (Eriksson et al., 1977). However, ribavirin is only approved in a few countries, probably due to severe side effects (Furuta et al., 2005). Clearly, new antiviral compounds are needed, preferably directed against different targets.

Influenza virus as well as Thogotovirus belong to the family of Orthomyxoviridae which, as well as the family of the Bunyaviridae, including the Hantavirus, Nairovirus, Orthobunyavirus, and Phlebovirus, 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; Kukkonen et al., 2005; Leahy et al., 1997; Noah and Krug, 2005). It binds to the 5' RNA cap of cellular mRNA molecules and cleaves the RNA cap together with a stretch of nucleotides. The capped RNA fragments serve as primers for the synthesis of viral mRNA. The polymerase is composed of three subunits: PB1 (polymerase basic protein), PB2, and PA. While PB1 harbors the endonuclease and polymerase activities, PB2 contains the RNA cap binding domain.

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., 2005). 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). 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). In addition, flutimide, a substituted 2,6-diketopiperazine, identified in extracts of *Delitschia confertaspera*, a fungal species, has been shown to inhibit the endonuclease of Influenza virus (Tomassini et al., 1996). Moreover, there have been attempts to interfere with viral transcription by nucleoside analogs, such as 2'-deoxy-2'-fluoroguanosine (Tisdale et al., 1995) 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.,

2005). Finally, by comparison studies between the binding mode of human cap binding protein eIF4E to RNA cap structures and Influenza virus RNP interaction with RNA cap structures Hooker et al. (2003) identified a novel cap analog that selectively interacts with Influenza virus, but not human cap binding protein. However, the major obstacle for identifying compounds that  
5 interact with the RNA cap binding pocket of PB2 and potentially interfere with RNA cap binding and thereby RNA polymerase activity was up to now that the structure and identity of said binding pocket was unknown.

Several attempts have been made to elucidate the RNA cap binding site, however, with  
10 controversial results. Cross-linking experiments indicated that two separate sequences, one N-(242-282) and one C-terminal (538-577) proximal segment of PB2, constitute the RNA cap-binding site of the Influenza virus RNA polymerase (Honda et al., 1999). Additional cross-linking experiments identified a sequence extending from amino acid 533 to amino acid 564 in the PB2 protein subunit, particularly amino acid residue Trp552, as potential interaction site for  
15 the RNA cap (Li et al., 2001). Furthermore, mutational analysis resulted in potential RNA cap binding amino acid residues Phe363 and Phe404 within PB2 (Fechter et al., 2003).

It is an object of the present invention to provide (i) high resolution structural data of the RNA cap binding pocket of PB2 by X-ray crystallography, (ii) computational as well as *in vitro*  
20 methods, preferably in a high-throughput setting, for identifying compounds that can bind to the RNA binding pocket of PB2, preferably competing with RNA cap binding and thereby interfering with RNA polymerase activity, and (iii) pharmacological compositions comprising such compounds for the treatment of infectious diseases caused by viruses using the cap snatching mechanism for synthesis of viral mRNA.

25 The present invention allows for the first time for the precise definition of the PB2 RNA cap-binding site within an independently folded domain. It has up to now been highly controversial where the site may be located. It was a common believe that a functional cap binding site requires all three polymerase subunits and possibly also viral RNA (Cianci et al., 1995; Li et al.,  
30 2001). The surprising achievement of the present inventors to recombinantly produce soluble PB2 polypeptide fragments comprising a functional RNA cap binding pocket allows to perform *in vitro* high-throughput screening for inhibitors of a functional site on Influenza virus polymerase using easily obtainable material from a straightforward expression system. Previous work on cap binding inhibitors has, for instance, used complete ribonucleoprotein particles

purified from Influenza virions (Hooker et al., 2003). Furthermore, by providing detailed structure coordinates of the RNA cap binding pocket within PB2, the present invention allows to use structure-based approaches to cap-binding inhibitor design, i.e., *in silico* screening and lead optimization.

5

#### SUMMARY OF THE INVENTION

In a first aspect the present invention relates to soluble polypeptide fragment, wherein said polypeptide fragment is (i) derived from the Influenza virus RNA-dependent RNA polymerase subunit PB2 or variant thereof and (ii) capable of binding to a RNA cap or analog thereof.

10

In a further aspect the present invention relates to a complex comprising the soluble polypeptide fragments of the present invention and a RNA cap or analog thereof.

15

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

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

20

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

25

In a further aspect the present invention relates to a method for identifying compounds which associate with all or part of the RNA cap binding pocket of PB2 or the binding pocket of a PB2 polypeptide variant, comprising the steps of

(a) constructing a computer model of said binding pocket defined by the structure coordinates of the complex of the present invention as shown in Figure 18;

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

30

(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 binding pocket in order to provide an energy-minimized configuration of the said compound in the binding pocket; and

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

5 In a further aspect the present invention relates to a compound identifiable by the in silico method of the present invention, under the provision that the compound is not m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, m<sup>7</sup>GpppUm, 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one or T-705 and is able to bind to the RNA cap binding pocket of PB2 or variant thereof.

10

In a further aspect the present invention relates to a compound identifiable by the in silico method of the present invention, under the provision that the compound is not m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, m<sup>7</sup>GpppUm, 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one or T-705 and is  
15 able to inhibit binding between the PB2 polypeptide, variant thereof or fragment thereof and the RNA cap or analog thereof.

In a further aspect the present invention relates to a method for identifying compounds which associate with the RNA cap binding pocket of PB2 or binding pockets of PB2 polypeptide  
20 variants, comprising the steps of (i) contacting the polypeptide fragment of the present invention or the recombinant host cell of the present invention with a test compound and (ii) analyzing the ability of said test compound to bind to PB2.

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

In a further aspect the present invention relates to a compound identifiable by the in vitro method of the present invention, under the provision that the compound is not m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, m<sup>7</sup>GpppUm, 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one or T-705 and is  
30 able to bind to the PB2 polypeptide, variant thereof or fragment thereof.

In a further aspect the present invention relates to a compound identifiable by the in vitro method of the present invention under the provision that the compound is not m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP,

m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, m<sup>7</sup>GpppUm, 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one or T-705 and is able to inhibit binding between the PB2 polypeptide, variant thereof or fragment thereof and the RNA cap or analog thereof.

5

In a further aspect the present invention relates to an antibody directed against the RNA cap binding domain of PB2.

10

In a further aspect the present invention relates to a use of a compound of the present invention, a pharmaceutical composition of the present invention or an antibody of the present invention for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with negative-sense ssRNA viruses.

#### BRIEF DESCRIPTION OF THE DRAWINGS

15 FIG. 1 illustrates full-length PB2 with potential binding sites for the PB1 subunit of the Influenza polymerase and potential RNA cap binding sites as predicted by the prior art (Cap-N/Cap-C see Honda et al. (1999), Cap-C (Trp552) see Li et al. (2001), Cap-M see Fechter et al. (2003). In addition, various PB2 fragments are shown that were generated for bacterial expression. All of these constructs are expressed at high levels and are soluble. From top to bottom the length of  
20 PB2 fragments encoded by 13 isolated clones are depicted, which encoded seven unique PB2 fragments. The isolated clones span nucleotides: Clone 68: 886 to 1449, Clone 16: 871 to 1452, Clone 15: 871 to 1452, Clone 57: 844 to 1449, Clone 13: 844 to 1449, Clone 10: 832 to 1443, Clone 28: 805 to 1452, Clone 23: 805 to 1452, Clone 12: 723 to 1425, Clone 4: 706 to 1491, Clone 30: 706 to 1491, Clone 07: 706 to 1491, and Clone 05: 706 to 1491 of SEQ ID NO: 25.

25

FIG. 2A shows a graph illustrating the purification of PB2 polypeptide fragment amino acids 318 to 483 on a Superdex 75 gel filtration column. The absorption at 280 nm in arbitrary units is plotted versus the elution time and fraction number.

30 FIG. 2B shows a coomassie stained SDS PAGE gel of fractions 24-34 of the PB2 polypeptide fragment (amino acids 318 to 483) after gel filtration. The molecular weight markers have sizes 116 - 66.2 - 45 - 35 - 25 - 18.4 and 14.4 kD.

FIG. 3A shows the native crystals of the PB2 polypeptide fragment (amino acids 318 to 483) in complex with m<sup>7</sup>GTP (5 mM) formed with precipitant solution 1.6 M sodium formate, 0.1 M citric acid pH 4.6.

- 5 FIG. 3B shows the crystals of partially seleno-methionylated PB2 polypeptide fragment (amino acids 318 to 483) in complex with m<sup>7</sup>GTP (5 mM) formed with precipitant solution 1.6 M sodium formate, 0.1 M citric acid pH 4.6.

- 10 FIG. 4 shows the X-ray diffraction pattern obtained on a native crystal on beamline BM14 at the European Synchrotron Radiation Facility. The diffraction pattern indexes in space-group C222<sub>1</sub> with cell dimensions a=92.2, b=94.4, c=220.4 Å. The top right inset shows the diffraction extending to 2.4 Å resolution. The bottom right inset shows the frozen crystal used for the data collection, which has an approximate size of 15 x 15 x 50 μm<sup>3</sup>.

- 15 FIG. 5 shows a ribbon diagram of the structure of the PB2 polypeptide fragment (amino acids 318 to 483) in complex with m<sup>7</sup>GTP. Secondary structure elements are labeled as calculated by DSSP (Holm, L. and Sander, C., 1993), with alpha helices in dark grey and beta-strands in light grey. The GTP is shown as a ball and stick model as are the side-chains of Phe323, Phe404 and His357. Significant loops, centered on residues 348 and 420 are labeled, as are the N- and C-  
20 terminal residues visible in the structure (respectively 320 and 483).

FIG. 6 shows an alternative view of the structure of the PB2 polypeptide fragment (amino acids 318 to 483) in complex with m<sup>7</sup>GTP revealing that the 420-loop projects into the solvent.

- 25 FIG. 7 depicts the conformation and the unbiased, experimental electron density of m<sup>7</sup>GTP surrounded by some of the amino acids that form the RNA cap binding pocket within PB2. The electron density corresponds to a map obtained by RESOLVE using experimental phases and non-crystallographic symmetry averaging contoured at 0.95 sigma.

- 30 FIG. 8 depicts the cap analog m<sup>7</sup>GTP embedded in the RNA cap binding pocket within PB2 with potential interacting atoms. Putative hydrogen bonds are denoted by dotted lines.

FIG. 9 shows an amino acid sequence alignment of the PB2 RNA cap-binding domains from Influenza A (strain A/Victoria/3/1975) and B (strain B/Lee/40). The secondary structure of

A/Victoria/3/1975 (calculated with DSSP (Holm and Sander, 1993)) is displayed over the sequence alignment. The structure based sequence alignments were plotted with ESPRIPT (<http://esprict.ibcp.fr/ESPrict/cgi-bin/ESPrict.cgi>). Residues on a solid black background are identical between the two sequences. The triangles indicate the principle residues interacting with the cap analog  $m^7GTP$ .

FIG. 10 shows an amino acid sequence alignment of the PB2 RNA cap-binding domains from influenza A (strain A/Victoria/3/1975), B (strain B/Lee/40) and C (strain Ann Arbor/1/50). Annotation is as for FIG. 9.

FIG. 11 schematically depicts the interactions between the RNA cap analog  $m^7GTP$  and the RNA cap binding pocket within PB2. The cap analog  $m^7GTP$  (labeled M7G) is drawn with dark grey bonds, protein residues with medium grey bonds and water molecules with light grey spheres. Hydrogen bonds are shown as dotted lines with the distance between donor and acceptor in Å. Hatched parts of spheres in dark grey indicate residues that are in van der Waals contact with the ligand. The diagram corresponds to the molecule with chain designation B in the coordinate file. The four other molecules in the asymmetric unit (chains A, B, D and F) give similar diagrams with minor differences in distances and water positions. The figure was made using LIGPLOT (Wallace et al., 1995).

FIG. 12 shows a coomassie stained SDS PAGE gel of the results of elution of indicated wild-type and single point mutants of the PB2 cap-binding domain, after binding to a  $m^7GTP$  Sepharose 4B resin at 4°C. Mutants E361A, F404A, H357A and K376A failed to bind to the  $m^7GTP$  Sepharose 4B resin under conditions where wild-type protein binds, whereas F323A had weak residual binding activity and F325A bound the resin nearly as well as wild-type, however, this binding activity is drastically reduced at 37°C (cf. Table 3). A H357W mutant was also made as influenzas B and C have this residue at this position. This mutant was solubly purified and found to have slightly enhanced binding to the  $m^7GTP$  Sepharose 4B resin compared to wild-type. Mutant  $\Delta VQ$  was generated by replacing residues Val421-Gln426 by three glycines (420-loop) and shows cap-binding activity.

FIG. 13 shows the cap-binding activity of wild-type or mutant recombinant polymerase complexes. Cultures of HEK293T cells were co-transfected with plasmids expressing PB1 and PA ( $\Phi$ ) or cotransfected with plasmids expressing PB1, PB2-His or mutants thereof and PA. Cell

extracts were analysed by pull-down with m<sup>7</sup>GTP-Sepharose at 4°C. The polymerase complexes retained and eluted with m<sup>7</sup>GTP were revealed by Western-blot with PA-specific antibodies. IN and E indicate input cell extract and eluted protein, respectively. The position of the PA-specific band is indicated to the right and the mobility of molecular weight markers is shown on the left.

5 Mutant ΔVQ was generated by replacing residues Val421-Gln426 by three glycines (420-loop).

FIG. 14 is a graph that shows replication activity of the wild-type or mutant polymerases. Wild-type and mutant mini-RNPs were reconstituted *in vivo* and purified by affinity chromatography with Ni<sup>2+</sup>-NTA-agarose. The accumulation of progeny RNPs was determined by Western-blot using anti-PA and anti-NP antibodies. The data are presented as percent of the wild-type value and are averages and ranges of two experiments. Mutant ΔVQ was generated by replacing residues Val421-Gln426 by three glycines (420-loop).

FIG. 15 is a graph that shows the *in vitro* transcription activity of purified wild-type or mutant RNPs using ApG as primer. The data presented are averages and ranges of two experiments. Mutant ΔVQ was generated by replacing residues Val421-Gln426 by three glycines (420-loop).

FIG. 16 is a graph that shows the *in vitro* transcription activity of purified wild-type or mutant RNPs using β-globin mRNA as primer. The data presented are averages and ranges of two experiments. Mutant ΔVQ was generated by replacing residues Val421-Gln426 by three glycines (420-loop).

FIG. 17 is a graph that shows the ratio of ApG- versus β-globin mRNA-dependent *in vitro* transcription activity of purified wild-type or mutant RNPs. The data presented are averages and ranges of two experiments. Mutant ΔVQ was generated by replacing residues Val421-Gln426 by three glycines (420-loop).

FIG. 18 lists the refined atomic structure coordinates for PB2 polypeptide fragment amino acids 318 to 483 of SEQ ID NO.1 with Lys389. The fragment has an amino acid sequence according to SEQ ID NO: 11 in complex with the RNA cap analog 7-methyl-guanosine triphosphate (m<sup>7</sup>GTP). There are five molecules in the asymmetric unit with chains A, B, D, E, F. The RNA cap analog m<sup>7</sup>GTP is residue number 1 of each chain. There are 293 water molecules. The file header gives information about the structure refinement. "Atom" refers to the element whose coordinates are measured. The first letter in the column defines the element. The 3-letter code of

the respective amino acid is given and the amino acid sequence position. The first 3 values in the line "Atom" define the atomic position of the element as measured. The fourth value corresponds to the occupancy and the fifth (last) value is the temperature factor (B factor). The occupancy factor refers to the fraction of the molecules in which each atom occupies the position specified by the coordinates. A value of "1" indicates that each atom has the same conformation, i.e., the same position, in all molecules of the crystal. B is a thermal factor that measures movement of the atom around its atomic center. The anisotropic temperature factors are given in the lines marked "ANISOU". This nomenclature corresponds to the PDB file format.

## DETAILED DESCRIPTION OF THE INVENTION

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

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).

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.

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.

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](http://www.ebi.ac.uk/clustalw)) or Align (<http://www.ebi.ac.uk/emboss/align/index.html>) using standard settings, preferably for Align EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5. Those skilled in the art understand that it may be necessary to introduce gaps in either sequence to produce a satisfactory alignment. For example, residues 220 to 510 in the Influenza A virus PB2 subunit correspond to residues 222 to 511 and 227 to 528 in the Influenza B and C virus PB2 subunits, respectively. A thus generated alignment for three PB2 subunits is depicted in Fig. 10. Residues in two or more Influenza virus PB2 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, more preferably to less than 30% over a length of 10, 20 or 30 amino acids. A part of the best sequence alignment for the amino acid sequences of Influenza A (aa 315 to 498), B (aa 317 to 499), and C (aa 327 to 516) PB2 subunits is shown in Figures 9 and 10.

The present invention includes soluble Influenza virus RNA-dependent RNA polymerase PB2 subunit fragments, which are capable of binding to a RNA cap or analog thereof. The term "RNA-dependent RNA polymerase subunit PB2" preferably refers to the PB2 of Influenza A,

Influenza B and Influenza C virus, preferably having an amino acid sequence as set out in SEQ ID NO: 1, 2 or 3. "RNA-dependent RNA polymerase subunit PB2 variants" have at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g. Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with the amino acid sequence set forth in SEQ ID NOs: 1, 2, or 3. It is preferred that when a naturally occurring PB2 variant is aligned with a PB2 subunit according to SEQ ID NO:1, 2, or 3 that alignment will be over the entire length of the two proteins and, thus, that the alignment score will be determined on this basis. It is, however, possible that the natural variant may comprise C-terminal/N-terminal or internal deletions or additions, e.g. through N- or C-terminal fusions. In this case only the best aligned region is used for the assessment of the similarity and identity, respectively.

Preferably, and as set out in more detail below soluble fragments derived from these variants show the indicated similarity and identity, respectively, preferably within the region required for RNA cap binding. Accordingly, any alignment between SEQ ID NOs: 1, 2, or 3 and a PB2 variant should preferably comprise the RNA cap binding pocket. Thus, the above sequence similarity and identity, respectively, to SEQ ID NO: 1, 2, or 3 occurs at least over a length of 100, 110, 120, 130, 140, 150, 160, 165, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300 or more amino acids, preferably comprising the RNA cap binding pocket. Accordingly, in a preferred embodiment a RNA-dependent RNA polymerase subunit PB2 variant 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%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 1 over a length of 100 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 1 over a length of 110 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 1 over a length of 120 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 1 over a length of 130 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%,

13

least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 1 over a length of 300 amino acids.

- 5 In a preferred embodiment a RNA-dependent RNA polymerase subunit PB2 variant 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%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 100 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 110 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 120 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 130 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 140 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 150 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 160 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 165 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 170 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 180 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 190 amino acids, at

least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 200 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 210 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 220 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 230 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 240 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 250 amino acids or at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 2 over a length of 300 amino acids.

20 In a preferred embodiment a RNA-dependent RNA polymerase subunit PB2 variant 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%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3 over a length of 100 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3 over a length of 110 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3 over a length of 120 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3 over a length of 130 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3 over a length of 140

5 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3  
over a length of 160 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%,  
85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence  
similarity, preferably sequence identity to SEQ ID NO: 3 over a length of 165 amino acids, at  
least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%,  
10 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence  
identity to SEQ ID NO: 3 over a length of 170 amino acids, at least 60%, 65%, 70%, 80%, 81%,  
82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3 over a  
length of 180 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%,  
15 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence  
similarity, preferably sequence identity to SEQ ID NO: 3 over a length of 190 amino acids, at  
least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%,  
91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence  
identity to SEQ ID NO: 3 over a length of 200 amino acids, at least 60%, 65%, 70%, 80%, 81%,  
20 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3 over a  
length of 210 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%,  
86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence  
similarity, preferably sequence identity to SEQ ID NO: 3 over a length of 220 amino acids, at  
25 least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%,  
91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence  
identity to SEQ ID NO: 3 over a length of 230 amino acids, at least 60%, 65%, 70%, 80%, 81%,  
82%, 83%, 84%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3 over a  
30 length of 240 amino acids, at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 84%, 85%,

91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity to SEQ ID NO: 3 over a length of 300 amino acids.

A large number of natural PB2 variants of sequences according to SEQ ID NO: 1, 2, or 3 are known and have been described in the literature. All these PB2 variants are comprised and can be the basis for the soluble fragments of the present invention. Preferred examples for Influenza A, if SEQ ID NO: 1 is used as reference sequence, comprise mutations at one or more of Val227, Arg251, Ile255, Ser271, Gln288, Ile338, Val344, Arg355, Ile373, Leu374, Asn456, Val 461, and/or Ser497. In a preferred embodiment, said variants comprise one or more of the following mutations: Val227Met, Arg251Lys, Ile255Val, Ile255Thr, Ser271Ala, Gln288Leu, Ile338Val, Ile338Thr, Val344Leu, Arg355Lys, Ile373Thr, Leu374Ile, Asn456Ser, Val461Ile, and/or Ser497Asn. Preferred variants of the Influenza A virus PB2 subunit comprise a mutation at position 389 resulting in an amino acid exchange from arginine to lysine (e.g., SEQ ID NO: 11), optionally combined with one or more of the afore-mentioned mutations. Preferred variants of the Influenza B virus PB2 subunit, if SEQ ID NO: 2 is used as reference sequence, include mutations at one ore more of the following amino acid positions: Thr254, Met261, Ile269, Thr301, Ile303, Leu319, Ile346, Lys380, Arg382, Met383, Lys385, Lys397, Asn442, Ser456, Glu467, Leu468, and/or Thr494. In a preferred embodiment the PB2 subunit variant comprises one or more of the following mutations: Thr254Ala, Met261Thr, Ile269Val, Thr301Ala, Ile303Leu, Leu319Gln, Ile346Val, Lys380Gln, Arg382Lys, Met383Leu, Lys385Arg, Lys397Arg, Asn442Ser, Ser456Pro, Glu467Gly, Leu468Ser, and/or Thr494Ile. Preferred variants of the Influenza B virus PB2 subunit, if SEQ ID NO: 3 is used as reference sequence, include mutations at one ore more of the following amino acid positions: Leu311, Pro330, and/or Ser436. In a preferred embodiment, said mutations are as follows: Leu311Pro, Pro330Gln, and/or Ser436Thr.

The soluble fragments of the present invention are, thus, based on RNA-dependent RNA polymerase subunit PB2 or variants thereof as defined above. Accordingly, in the following specification the term “soluble polypeptide fragment(s)” and “PB2 polypeptide fragments” always comprises such fragments derived both from the PB2 proteins as set out in SEQ ID NO: 1, 2, or 3 and fragments derived from PB2 protein variants thereof, as set out above, which are capable of binding RNA cap or an analog thereof. However, the specification also uses the term “PB2 polypeptide fragment variants” or “PB2 fragment variants” to specifically refer to soluble PB” fragments, capable of binding to RNA cap or an analog thereof that are derived from RNA-

dependent RNA polymerase subunit PB2 variants. The soluble PB2 fragments of the present invention thus, preferably comprise, essentially consist or consist of sequences of naturally occurring Influenza virus subunit PB2. It is, however, also envisioned that the PB2 fragments variants further contain amino acid substitutions at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more amino acid positions, and have at least 60%, 65%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g. Align, using standard settings, preferably EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with the amino acid sequence set forth in SEQ ID NOs: 1, 2, or 3. It is understood that PB2 fragments of the present invention may comprise additional amino acids not derived from PB2, 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 NOs: 1, 2, or 3 at least over a length of 100, 110, 120, 130, 140, 150, 160, 165, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, or 270 amino acids, wherein the respective sequence of SEQ ID NO: 1, 2, or 3, preferably comprises the RNA cap binding pocket.

In a preferred embodiment, the soluble PB2 polypeptide fragment variants comprise at least or consist of the amino acid residues corresponding to amino acid residues 323 to 404 of Influenza A virus PB2 (SEQ ID NO: 13) 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 the amino acid sequence set forth in SEQ ID NO: 13, more preferably the PB2 polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 318 to 483 of Influenza A virus PB2 (SEQ ID NO: 11) 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 sequence set forth in SEQ ID NO: 11, more preferably the PB2 polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 235 to 496 of Influenza A virus PB2 (SEQ ID NO: 4) and have at least 60%,  
5 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,  
10 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: 4, more preferably the PB2 polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 220 to 510 of Influenza A virus PB2 and have at least 60%, more preferably 65%, more preferably 70%, more preferably 75%, more preferably 80%, 81%, 82%, 83%, 84%, 85%,  
15 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  
20 amino acid sequence set forth in SEQ ID NO: 1. In preferred embodiments, the Influenza A virus PB2 polypeptide fragment variants of the present invention comprise mutations, preferably naturally occurring mutations such as mutations in one or more of the following amino acid residues when compared to SEQ ID NO: 1: Val227, Arg251, Ile255, Ser271, Gln288, Ile338, Val344, Arg355, Ile373, Leu374, Lys389, Asn456, Val 461, and/or Ser497. In a preferred  
25 embodiment, said mutations are as follows: Val227Met, Arg251Lys, Ile255Val, Ile255Thr, Ser271Ala, Gln288Leu, Ile338Val, Ile338Thr, Val344Leu, Arg355Lys, Ile373Thr, Leu374Ile, Lys389, Asn456Ser, Val461Ile, and/or Ser497Asn. A preferred variant of the Influenza A virus PB2 subunit fragment comprises a mutation at position 389 resulting in an amino acid exchange from arginine to lysine (e.g., SEQ ID NO: 11), optionally combined with one or more of the  
30 afore-mentioned mutations.

In a preferred embodiment, the PB2 polypeptide fragment variants comprise at least or consist of the amino acid residues corresponding to amino acid residues 325 to 406 of Influenza B virus PB2 (derived from SEQ ID NO: 2) and have at least 80%, 81%, 82%, 83%, 84%, 85%, 86%,

87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best sequence alignment and/or over the region of the best sequence alignment, wherein the best sequence alignment is obtainable with art known tools, e.g. Align, using standard settings, preferably

5 EMBOSS::needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5, with the amino acid sequence set forth in SEQ ID NO: 2, more preferably the PB2 polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 320 to 484 of Influenza B virus PB2 (derived from SEQ ID NO: 2) and have at least 70%, more preferably 75%, more preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%,

10 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

15 NO: 2, more preferably the PB2 polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 237 to 497 of Influenza B virus PB2 (derived from SEQ ID NO: 2) and have at least 60%, more preferably 65%, more preferably 70%, more preferably 75%, more preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably

20 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, more preferably the PB2 polypeptide fragment variants

25 comprise at least the amino acid residues corresponding to amino acid residues 222 to 511 of Influenza B virus PB2 (derived from SEQ ID NO: 2) and have at least 60%, more preferably 65%, more preferably 70%, more preferably 75%, more preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence similarity, preferably sequence identity over the entire length of the fragment using the best

30 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. In preferred embodiments, the Influenza B virus PB2 polypeptide fragment variants of the present invention comprise mutations, preferably

naturally occurring mutations, at one ore more of the following amino acid positions compared to SEQ ID NO: 2: Thr254, Met261, Ile269, Thr301, Ile303, Leu319, Ile346, Lys380, Arg382, Met383, Lys385, Lys397, Asn442, Ser456, Glu467, Leu468, and/or Thr494. In a preferred embodiment, said fragment variants comprise one or more of the following mutations:

5 Thr254Ala, Met261Thr, Ile269Val, Thr301Ala, Ile303Leu, Leu319Gln, Ile346Val, Lys380Gln, Arg382Lys, Met383Leu, Lys385Arg, Lys397Arg, Asn442Ser, Ser456Pro, Glu467Gly, Leu468Ser, and/or Thr494Ile.

10 In a preferred embodiment, the PB2 polypeptide fragment variants comprise at least or consist of the amino acid residues corresponding to amino acid residues 335 to 416 of Influenza C virus PB2 (derived from SEQ ID NO: 3) and have at least 80%, more preferably 85%, more preferably 90%, most preferably 95% sequence similarity over the entire length of the fragment with the amino acid sequence set forth in SEQ ID NO: 3, more preferably the PB2 polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 330 to

15 501 of Influenza C virus PB2 (derived from SEQ ID NO: 3) and have at least 70%, more preferably 75%, more preferably 80%, more preferably 85%, most preferably 90% sequence similarity over the entire length of the fragment with the amino acid sequence set forth in SEQ ID NO: 3, more preferably the PB2 polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 242 to 514 of Influenza C virus PB2 (derived

20 from SEQ ID NO: 3) and have at least 60%, more preferably 65%, more preferably 70%, more preferably 75%, more preferably 80%, more preferably 85%, most preferably 90% sequence similarity over the entire length of the fragment with the amino acid sequence set forth in SEQ ID NO: 3, more preferably the PB2 polypeptide fragment variants comprise at least the amino acid residues corresponding to amino acid residues 227 to 528 of Influenza C virus PB2 (derived

25 from SEQ ID NO: 3) and have at least 60%, more preferably 65%, more preferably 70%, more preferably 75%, more preferably 80%, more preferably 85%, most preferably 90% sequence similarity over the entire length of the fragment with the amino acid sequence set forth in SEQ ID NO: 2. In preferred embodiments, the Influenza C virus PB2 polypeptide fragment variants of the present invention comprise mutations, preferably naturally occurring mutations such as

30 mutations in one or more of the following amino acid residues when compared to SEQ ID NO: 3: Leu311, Pro330, and/or Ser436. In a preferred embodiment, the fragment variants comprise one or more of the following mutations: Leu311Pro, Pro330Gln, and/or Ser436Thr.

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](http://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 as much as 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. It is particularly preferred that the expression and, preferably, purification of such a protein fragment is carried out as set out in more detail below in Example 2.

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

The term "associate" as used in the context of identifying compounds with the methods of the present invention refers to a condition of proximity between a moiety (i.e., chemical entity or compound or portions or fragments thereof), and a binding pocket of PB2. 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 "RNA cap" refers to a cap structure found on the 5' end of an mRNA molecule and consists of a guanine nucleotide connected to the mRNA via an unusual 5' to 5' triphosphate linkage. This guanosine is methylated on the 7 position. Further modifications include the possible methylation of the 2' hydroxy-groups of the first 3 ribose sugars of the 5' end of the mRNA. "RNA cap analogs" refer to structures that resemble the RNA cap structure. Examples of RNA cap analogs include 7-methyl-guanosine ( $m^7G$ ), 7-methyl-guanosine monophosphate ( $m^7GMP$ ), 7-methyl-guanosine triphosphate ( $m^7GTP$ ), 7-methyl-guanosine linked via a 5' to 5' triphosphate bridge to guanosine ( $m^7GpppG$ ), 7-methyl-guanosine linked via a 5' to 5' triphosphate bridge to guanosine methylated at the 2' OH position of the ribose ( $m^7GpppGm$ ), 7-methyl-guanosine linked via a 5' to 5' triphosphate bridge to adenosine ( $m^7GpppA$ ), 7-methyl-guanosine linked via a 5' to 5' triphosphate bridge to adenosine methylated at the 2' OH position of the ribose ( $m^7GpppAm$ ), 7-methyl-guanosine linked via a 5' to 5' triphosphate bridge to cytidine ( $m^7GpppC$ ), 7-methyl-guanosine linked via a 5' to 5' triphosphate bridge to cytidine methylated at the 2' OH position of the ribose ( $m^7GpppCm$ ), 7-methyl-guanosine linked via a 5' to 5' triphosphate bridge to uridine ( $m^7GpppU$ ), 7-methyl-guanosine linked via a 5' to 5' triphosphate bridge to uridine methylated at the 2' OH position of the ribose ( $m^7GpppUm$ ). Thus, in a preferred embodiment of the present invention the cap analog is selected from the group consisting of  $m^7G$ ,  $m^7GMP$ ,  $m^7GTP$ ,  $m^7GpppN$ ,  $m^7GpppNm$ , where N is a nucleotide, preferably G, A, U, or C. In another preferred embodiment, the RNA cap analogs may comprise additional, e.g., 1 to 15 nucleotides, i.e., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15, wherein the nucleotides are preferably independently selected from the group consisting of A, C, G, U, and/or T. These further nucleotides are, preferably linked via a phosphoester, phosphodiester or phosphotriester bond or a non-hydrolyzable analog thereof to the first nucleotide N in  $m^7GpppN$  or  $m^7GpppNm$ .

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).

5 The term "binding pocket" refers to a three-dimensional structure formed by the polypeptide fragments of the invention, i.e., the RNA cap binding domain of PB2, that is lined with amino acids that directly contact the RNA cap or amino acid residues that position the amino acid residues that are in direct contact with the RNA cap (second layer amino acid residues), e.g., Arg332 and Ser337. As a result of its shape the binding pocket accommodates the RNA cap or analog thereof. The RNA cap binding pocket of PB2 is defined by structure coordinates  
10 originating from the analysis of the crystal structure data of the complex between a PB2 polypeptide fragment comprising the RNA cap binding site and a RNA cap analog. The term "binding pocket" also includes binding pockets of PB2 polypeptide fragment variants.

15 The term "RNA cap binding domain of PB2" refers to the minimal polypeptide fragment of PB2 that comprises the RNA binding pocket in its native three-dimensional structure.

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  
20 ribonucleotide bases and includes DNA and RNA molecules, both sense and anti-sense strands. The term comprises cDNA, genomic DNA, and recombinant DNA. A polynucleotide may consist of an entire gene, or a portion thereof.

25 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  
30 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 PB2 polypeptide fragment or variants thereof according to the invention. This polynucleotide may be found inside the host cell (i) freely dispersed as such, (ii) incorporated in a recombinant vector, or (iii) integrated into the host cell genome or mitochondrial DNA. The recombinant cell can be used for expression of a polynucleotide of interest or for amplification of the polynucleotide or the recombinant vector of the invention. The term "recombinant host cell" includes the progeny of the original cell which has been transformed, transfected, or infected with the polynucleotide or the recombinant vector of the invention. A recombinant host cell may be a bacterial cell such as an *E. coli* cell, a yeast cell such as *Saccharomyces cerevisiae* or *Pichia pastoris*, a plant cell, an insect cell such as SF9 or Hi5 cells, or a mammalian cell. Preferred examples of mammalian cells are Chinese hamster ovary (CHO) cells, green African monkey kidney (COS) cells, human embryonic kidney (HEK293) cells, HELA cells, and the like.

As used herein, the term "crystal" or "crystalline" means a structure (such as a three-dimensional solid aggregate) in which the plane faces intersect at definite angles and in which there is a regular structure (such as internal structure) of the constituent chemical species. The term "crystal" can include any one of: a solid physical crystal form such as an experimentally prepared crystal, a crystal structure derivable from the crystal (including secondary and/or tertiary and/or quaternary structural elements), a 2D and/or 3D model based on the crystal structure, a representation thereof such as a schematic representation thereof or a diagrammatic representation thereof, or a data set thereof for a computer. In one aspect, the crystal is usable in X-ray crystallography techniques. Here, the crystals used can withstand exposure to X-ray beams and are used to produce diffraction pattern data necessary to solve the X-ray crystallographic structure. A crystal may be characterized as being capable of diffracting X-rays in a pattern defined by one of the crystal forms depicted in T. L. Blundell and L. N. Johnson, "Protein Crystallography", Academic Press, New York (1976).

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

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

5

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 a binding pocket) 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.

15

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 PB2 polypeptide fragment or RNA cap binding pocket therein from the backbone of PB2 or the RNA cap binding pocket therein as defined by the structure coordinates of the PB2-m<sup>7</sup>GTP complex according to Figure 18.

20

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.

25

The term "fitting program operation" refers to an operation that utilizes the structure coordinates of a chemical entity, binding pocket, molecule or molecular complex, or portion thereof, to associate the chemical entity with the binding pocket, molecule or molecular complex, or portion thereof. This may be achieved by positioning, rotating or translating the chemical entity in the binding pocket to match the shape and electrostatic complementarity of the binding pocket. Covalent interactions, non-covalent interactions such as hydrogen bond, electrostatic,

30

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 binding pocket.

5

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 bind to the polypeptide fragment of interest, i.e., the PB2 polypeptide fragment of the invention or variants thereof comprising the RNA cap binding pocket. 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 binds to 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.

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 bind to 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 RNA cap binding domain of PB2 or a peptide 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')<sub>2</sub>, 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, pp. 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.

The present inventors have found that there are Influenza virus RNA-dependent RNA polymerase subunit PB2 derived fragments, which are soluble and capable of binding to a RNA cap or analog thereof and, thus, suitable for a crystallization to obtain structural information but also to carry out binding studies with recombinant proteins. It is one aspect of the invention to provide a soluble polypeptide fragment, wherein said polypeptide fragment is (i) derived from the Influenza virus RNA-dependent RNA polymerase subunit PB2 and (ii) capable of binding to a RNA cap or analog thereof. The RNA-dependent RNA polymerase subunit PB2 from which the soluble fragments of the invention are derived are preferably derived from an Influenza A, B or C virus. The minimal length of the soluble fragment of the present invention is determined by its ability to bind RNA cap or an analog thereof. Accordingly, it is preferred that the PB2 polypeptide fragment of the invention comprises at least or consists of amino acid residues 323 to 404 or 320 to 483 of SEQ ID NO 1 or amino acids corresponding thereto, e.g. in PB2 polypeptide fragment variant; at least or consists of amino acid residues 325 to 406 or 320 to 484 of SEQ ID NO 2 or amino acids corresponding thereto, e.g. in PB2 polypeptide fragment variant; or at least or consists of amino acid residues 335 to 416 or 330 to 501 of SEQ ID NO 3 or amino acids corresponding thereto, e.g. in PB2 polypeptide fragment variant. On the other hand it has been observed that the solubility of the fragments decreases if their length extends beyond certain C- and/or N-terminal boundaries. For Influenza A type derived soluble fragments these

boundaries are with reference to SEQ ID NO: 1 preferably for the N-terminus amino acid 220 or an amino acid corresponding thereto in a variant and for the C-terminus amino acid 510 or an amino acid corresponding thereto in a variant. For Influenza B type derived soluble fragments these boundaries are with reference to SEQ ID NO: 2 preferably for the N-terminus amino acid 222 or an amino acid corresponding thereto in a variant and for the C-terminus amino acid 511 or an amino acid corresponding thereto in a variant. For Influenza C type derived soluble fragments these boundaries are with reference to SEQ ID NO: 3 preferably for the N-terminus amino acid 227 or an amino acid corresponding thereto in a variant and for the C-terminus amino acid 528 or an amino acid corresponding thereto in a variant. The term "soluble" in this context refers to polypeptide fragments, which are dissolvable at a concentration of at least 0.1 mg/ml solvent, more preferably at least 0.5 mg/ml, more preferably at least 1 mg/ml, more preferably at 5 mg/ml or more. Suitable solvents are buffer systems comprising Tris·HCl at concentrations ranging from 0.01 M to 3 M, preferably 0.05 M to 2 M, more preferably 0.1 M to 1 M, at pH 3 to pH 9, preferably pH 4 to pH 9, more preferably pH 7 to pH 9 and optionally a reducing agent such as dithiothreitol (DTT) or TCEP·HCl (Tris(2-carboxyethyl) phosphine hydrochloride) at a concentration of 1 mM to 20 mM.

In a preferred embodiment, said polypeptide fragment is purified to an extent to be suitable for crystallization. Preferably has a purity of at least 95%, 96%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%. Purity of a protein can be assessed by standard methods including SDS polyacrylamide gel electrophoresis and staining by silver staining or HPLC and detection at 280 nm.

In another preferred embodiment, the

(i) N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 220 or higher, e.g. 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290, 295, 300, 305, 310, 315, 320, or 323, and the C-terminus is identical to or corresponds to amino acid position 510 or lower, e.g. 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483, of the amino acid sequence of PB2 according to SEQ ID NO: 1, more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 220 and the C-terminus is identical to or corresponds to amino acid position 510, 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 225 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488,

31

- identical to or corresponds to amino acid position 280 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 285 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 290 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 295 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 300 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 305 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; ; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 310 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; ; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 315 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; ; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 320 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483; or ; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 323 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower, in particular 505, 500, 495, 490, 489, 488, 487, 486, 485, 484 or 483;
- (ii) N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 222 or higher, e.g. 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290, 295, 300, 305, 310, 315, 320, 325, and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490,

489, 488, 487, 486, 485, or 484, of the amino acid sequence of PB2 according to SEQ ID NO: 2, more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 222 or higher and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490,

5 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 225 or higher and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 230 or higher

10 and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 235 or higher and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490,

15 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 240 or higher and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 245 or higher

20 and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 250 or higher and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490,

25 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 255 or higher and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 260 or higher

30 and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 265 or higher and the C-terminus is identical to or corresponds to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490,

489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 270 or higher and the C-terminus is identical to or corresponds to to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of  
5 said polypeptide fragment is identical to or corresponds to amino acid position 275 or higher and the C-terminus is identical to or corresponds to to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 280 or higher and the C-terminus is identical to or corresponds to to amino  
10 acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 285 or higher and the C-terminus is identical to or corresponds to to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of  
15 said polypeptide fragment is identical to or corresponds to amino acid position 290 or higher and the C-terminus is identical to or corresponds to to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 295 or higher and the C-terminus is identical to or corresponds to to amino  
20 acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 300 or higher and the C-terminus is identical to or corresponds to to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of  
25 said polypeptide fragment is identical to or corresponds to amino acid position 305 or higher and the C-terminus is identical to or corresponds to to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 310 or higher and the C-terminus is identical to or corresponds to to amino  
30 acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 315 or higher and the C-terminus is identical to or corresponds to to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; more preferably the N-terminus of

said polypeptide fragment is identical to or corresponds to amino acid position 320 or higher and the C-terminus is identical to or corresponds to to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484; or more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 325 or higher and the C-terminus is identical to or corresponds to to amino acid position 511 or lower, e.g. 510, 505, 500, 499, 498, 497, 496, 495, 494, 493, 492, 491, 490, 489, 488, 487, 486, 485, or 484;

or

(iii) N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 227 or higher, e.g. 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290, 295, 300, 305, 310, 315, 320, 325, 330, or 335, and the C-terminus is identical to or corresponds to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501, of the amino acid sequence of PB2 according to SEQ ID NO: 3, more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 227 or higher and the C-terminus is identical to or corresponds to to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 230 or higher and the C-terminus is identical to or corresponds to to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 235 or higher and the C-terminus is identical to or corresponds to to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 240 or higher and the C-terminus is identical to or corresponds to to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 245 or higher and the C-terminus is identical to or corresponds to to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 250 or higher and the C-terminus is identical to or corresponds to to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505,

36

lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 305 or higher and the C-terminus is identical to or corresponds to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 227 or higher and the C-terminus is identical to or corresponds to amino acid position 310 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 315 or higher and the C-terminus is identical to or corresponds to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 320 or higher and the C-terminus is identical to or corresponds to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 325 or higher and the C-terminus is identical to or corresponds to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 330 or higher and the C-terminus is identical to or corresponds to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501; or more preferably the N-terminus of said polypeptide fragment is identical to or corresponds to amino acid position 335 or higher and the C-terminus is identical to or corresponds to amino acid position 528 or lower, e.g. 525, 520, 519, 518, 517, 516, 515, 514, 513, 512, 511, 510, 509, 508, 507, 506, 505, 504, 503, 502, or 501.

In another embodiment said polypeptide fragment consist, essentially consists or corresponds to an amino acid sequence selected from the group of amino acid sequences according to SEQ ID NO: 4 to 13 and variants thereof, which retain the ability to associate with an RNA cap or analog thereof and are soluble. In preferred embodiments, said polypeptide fragments comprise amino acid substitutions, insertions, or deletions, preferably naturally occurring mutations as set forth above. Preferably the PB2 polypeptide fragment of the invention has or corresponds to the amino acid residues 235 to 496 (SEQ ID NO: 4), 241 to 483 (SEQ ID NO: 5), 268 to 483 (SEQ ID NO:

6), 277 to 480 (SEQ ID NO: 7), 281 to 482 (SEQ ID NO: 8), 290 to 483 (SEQ ID NO: 9), 295 to 482 (SEQ ID NO: 10), 318 to 483 (SEQ ID NO: 11), 320 to 483 (SEQ ID NO: 12), or 323 to 404 (SEQ ID NO: 13).

- 5 In another aspect, the invention provides a complex comprising the PB2 polypeptide fragment as described above and a RNA cap or analog thereof. Preferably, the cap analog is selected from the group consisting of m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, and m<sup>7</sup>GpppUm. In one embodiment of the present invention the polypeptide fragment in the complex consists of an amino acid sequence according to SEQ
- 10 ID NO: 11 and the cap analog is m<sup>7</sup>GTP, the complex having the structure defined by the structural coordinates as shown in Figure 18. In a preferred embodiment, any of the complexes of the invention comprise a crystalline form, preferably with space group C222<sub>1</sub> and unit cell dimensions of a = 9.2 nm, b = 9.4 nm; c = 22.0 nm ± 0.3 nm. Preferably said crystal diffracts X-rays to a resolution of 3.0 Å or higher, preferably 2.8 Å or higher, more preferably 2.6 Å or
- 15 higher, most preferably 2.4 Å or higher.

- In one embodiment, the protein solution suitable for crystallization may include in aqueous solution the PB2 polypeptide fragment or variant thereof at a concentration of 5 mg/ml to 20 mg/ml, preferably at 8 mg/ml to 18 mg/ml, more preferably at 11 mg/ml to 15 mg/ml, a RNA
- 20 cap analog at a concentration between 2 mM and 10 mM, preferably 3 mM and 8 mM, more preferably 4 mM and 6 mM, optionally a buffer system such as Tris·HCl at concentrations ranging from 0.01 M to 3 M, preferably 0.05 M to 2 M, more preferably 0.1 M to 1 M, at pH 3 to pH 9, preferably pH 4 to pH 9, more preferably pH 7 to pH 9 and optionally a reducing agent such as dithiothreitol (DTT) or TCEP·HCl (Tris(2-carboxyethyl) phosphine hydroxychloride) at
- 25 a concentration of 1 mM to 20 mM. The PB2 polypeptide fragment or variant thereof or the complex comprising the PB2 polypeptide fragment or variant thereof and an RNA cap or analog 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 is mixed with an equal volume of a precipitant solution such
- 30 as sodium formate, ammonium sulphate, polyethylene glycol of various sizes. 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 1.5 to 2 M sodium formate, and 0.05 to 0.15 M citric acid at pH

4 to pH 5. In another embodiment, the precipitant solution comprises 7% PEG6000, 1 M LiCl, 0.1 M citric acid pH 5.0, 10 mM TCEP.HCl.

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 PB2 polypeptide fragments of the inventions or variants thereof alone or in complex with a compound. Such variations include, but are not limited to, adjusting pH, protein concentration and/or crystallization temperature, changing the identity or concentration of salt and/or precipitant used, using a different method for crystallization, or introducing additives such as detergents (e.g., TWEEN 20 (monolaurate), LDOA, Brij 30 (4 lauryl ether)), sugars (e.g., glucose, maltose), organic compounds (e.g., dioxane, dimethylformamide), lanthanide ions, or poly-ionic compounds that aid in crystallizations. High throughput crystallization assays may also be used to assist in finding or optimizing the crystallization condition.

Microseeding may be used to increase the size and quality of crystals. In brief, micro-crystals are crushed to yield a stock seed solution. The stock seed solution is diluted in series. Using a needle, glass rod or strand of hair, a small sample from each diluted solution is added to a set of equilibrated drops containing a protein concentration equal to or less than a concentration needed to create crystals without the presence of seeds. The aim is to end up with a single seed crystal that will act to nucleate crystal growth in the drop.

The manner of obtaining the structure coordinates as shown in Figure 18, interpretation of the coordinates and their utility in understanding the protein structure, as described herein, are commonly understood by the skilled person and by reference to standard texts such as J. Drenth, "Principles of protein X-ray crystallography", 2<sup>nd</sup> 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 a 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), phases calculated with SHARP (Vonrhein et al., 2006), and improved with solvent flattening and non-crystallographic symmetry averaging (e.g., with RESOLVE (Terwilliger, 2000). Model autobuilding can be done, e.g., with ARP/wARP (Perrakis et al., 1999) and refinement with, e.g., REFMAC (Murshudov, 1997). Furthermore, the structure coordinates (Figure 18) of the PB2 fragment provided by the present invention are useful for the structure determination of PB2 polypeptides of other Orthomyxoviridae genera, or PB2 polypeptide variants that have amino acid substitutions, deletions, and/or insertions using the method of molecular replacement.

It is another aspect of the present invention to provide an isolated polynucleotide coding for the above-mentioned PB2 polypeptide fragments and variants thereof. The molecular biology methods applied for obtaining such isolated nucleotide fragments are generally known to the person skilled in the art (for standard molecular biology methods see Sambrook et al., Eds., "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), which is incorporated herein by reference). For example, RNA can be isolated from Influenza virus infected cells and cDNA generated applying reverse transcription polymerase chain reaction (RT-PCR) using either random primers (e.g., random hexamers of decamers) or primers specific for the generation of the fragments of interest. The fragments of interest can then be amplified by standard PCR using fragment specific primers.

In a preferred embodiment the isolated polynucleotide coding for the preferred embodiments of the soluble PB2 polypeptide fragments are derived from SEQ ID NOs: 23 (Influenza A), 26 (Influenza B), or 27 (Influenza C). In a preferred embodiment, the isolated polynucleotide coding for the soluble Influenza A virus PB2 polypeptide fragment or variant thereof is derived from SEQ ID NO: 25 which comprises an amino acid exchange from arginine to lysine at position 389 when compared to SEQ ID NO: 1. In an even more preferred embodiment, the isolated polynucleotide coding for the Influenza A virus PB2 polypeptide fragment or variants thereof is derived from SEQ ID NO: 26 which is a DNA sequence optimized for *E. coli* codon usage. In that context, derived refers to the fact that SEQ ID NOs: 23, 24, 25, 26, or 27 encode the full-length PB2 polypeptides and, thus, polynucleotides coding for preferred PB2 polypeptide fragments comprise deletions at the 5' and 3' ends of the polynucleotide as required by the respectively encoded PB2 polypeptide fragment.

In one embodiment, the present invention relates to a recombinant vector comprising said isolated polynucleotide. The person skilled in the art is well aware of techniques used for the incorporation of polynucleotide sequences of interest into vectors (also see Sambrook et al., 1989). Such vectors include any vectors known to the skilled person including plasmid vectors, 5 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 10 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.

15 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 20 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, 25 Factor Xa, PreScission, TEV protease, and the like. The recognition sites for such proteases are well known to the person skilled in the art. 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 30 which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

In another aspect, the present invention provides a recombinant host cell comprising said isolated polynucleotide or said recombinant vector. The recombinant host cells may be prokaryotic cells such as archaea 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<sup>TM</sup> (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 can 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 associate with all or part of the RNA cap binding pocket of PB2 or a binding pocket of PB2 polypeptide variant, comprising the steps of (a) constructing a computer model of said binding pocket defined by the structure coordinates of the complex as shown in Figure 18; (b) selecting a potential binding 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 binding pocket in order to provide an energy-minimized configuration of the said compound in the binding pocket; and (d) evaluating the results of said fitting operation to quantify the association between the said compound and the binding pocket model, whereby evaluating the ability of said compound to associate with the said binding pocket.

For the first time, the present invention permits the use of molecular design techniques to identify, select, or design potential binding partners for the RNA cap binding pocket of PB2 or RNA cap binding pockets of PB2 polypeptide variants, preferably inhibitors of RNA cap binding, based on the structure coordinates of said binding pocket according to Figure 18. Such a predictive model is valuable in light of the higher costs associated with the preparation and testing of the many diverse compounds that may possibly bind to said binding pocket. In order to use the structure coordinates generated for the PB2 polypeptide fragment in complex with the RNA cap analog 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 (A. Sali and T. L. Blundell, *J. Mol. Biol.*, 234, pp. 779-815 (1993) 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 bind to the RNA cap binding pocket of PB2 or PB2 polypeptide variants. This process may begin by a visual inspection of, for example, a three-dimensional computer model of the RNA cap binding pocket of PB2 based on the structural coordinates according to Figure 18. Selected fragments or chemical compounds may then be positioned in a variety of orientations or docked within the binding pocket. 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 (D. S. Goodsell et al., "Automated Docking of Substrates to Proteins by Simulated Annealing", *Proteins: Struct., Funct., Genet.*, 8, pp. 195-202 (1990); AUTODOCK is available from The Scripps Research Institute, La Jolla, CA) and (ii) DOCK (I. D. Kuntz et al., "A Geometric Approach to Macromolecule-Ligand Interactions", *J. Mol. Biol.*, 161, pp. 269-288 (1982); 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 (P. A. Bartlett et al., "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules", in Molecular Recognition in Chemical and Biological Problems", Special Publication, Royal Chem.

- 5 Soc., 78, pp. 182-196 (1989); G. Lauri and P. A. Bartlett, "CAVEAT: A Program to Facilitate the Design of Organic Molecules", J. Comp. Aid. Mol. Des., 8, pp. 51-66 (1994); 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 Y. C. Martin, "3D Database Searching in Drug Design", J. Med. Chem., 35, pp. 2145-2154 (1992)), and (iii) HOOK (M. B.  
10 Eisen et al., "HOOK: A Program for Finding Novel Molecular Architectures that Satisfy the Chemical and Steric Requirements of a Macromolecule Binding Site", Proteins: Struct., Funct., Genet., 19, pp. 199-221 (1994); HOOK is available from Molecular Simulations Incorporated, San Diego, CA).

- 15 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 RNA cap binding pocket of PB2 or binding pockets of PB2 polypeptide variants. In this screening, the quality of fit of such compounds to the binding pocket may be judged either by shape complementarity or by estimated interaction energy (E. C. Meng et al., J. Comp. Chem., 13, pp. 505-524 (1992)).

- 20 Alternatively, a potential binding partner for the RNA cap binding pocket of PB2, preferably an inhibitor of RNA cap binding, may be designed *de novo* on the basis of the 3D structure of the PB2 polypeptide fragment in complex with the RNA cap analog according to Figure 18. There are various *de novo* ligand design methods available to the person skilled in the art. Such  
25 methods include (i) LUDI (H.-J. Bohm, "The Computer Program LUDI: A New Method for the *De Novo* Design of Enzyme Inhibitors", J. Comp. Aid. Mol. Des., 6, pp. 61-78 (1992); LUDI is available from Molecular Simulations Incorporated, San Diego, CA), (ii) LEGEND (Y. Nishibata and A. Itai, Tetrahedron, 47, pp. 8985-8990 (1991); LEGEND is available from Molecular Simulations Incorporated, San Diego, CA), (iii) LeapFrog (available from Tripos  
30 Associates, St. Louis, MO), (iv) SPROUT (V. Gillet et al., "SPROUT: A Program for Structure Generation", J. Comp. Aid. Mol. Des., 7, pp. 127-153 (1993); SPROUT is available from the University of Leeds, UK), (v) GROUPBUILD (S. H. Rotstein and M. A. Murcko "GroupBuild: A Fragment-Based Method for *De Novo* Drug Design", J. Med. Chem., 36, pp. 1700-1710 (1993)), and (vi) GROW (J. B. Moon and W. J. Howe, "Computer Design of Bioactive

Molecules: A Method for Receptor-Based *De Novo* Ligand Design", *Proteins*, 11, pp 314-328 (1991)).

- 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 RNA cap binding pocket interacting compounds have been described and include, for example, N. C. Cohen et al., "Molecular Modeling Software and Methods for Medicinal Chemistry", *J. Med. Chem.*, 33, pp. 883-894 (1990); M. A. Navia and M. A. Murcko, "The Use of Structural Information in Drug Design", *Curr. Opin. Struct. Biol.*, 2, pp. 202-210 (1992); L. M. Balbes et al., "A Perspective in Modern Methods in Computer-Aided Drug Design", *Reviews in Computational Chemistry*, Vol. 5, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, pp. 37-380 (1994); W. C. Guida, "Software for Structure-Based Drug Design", *Curr. Opin. Struct. Biol.*, 4, pp. 777-781 (1994).
- 15 A molecule designed or selected as binding to the RNA cap binding pocket of PB2 or a binding pocket of a PB2 variant 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 (M. J. Frisch, Gaussian, Incorporated, Pittsburgh, PA), (ii) AMBER, version 4.0 (P. A. Kollman, University of California, San Francisco, CA), (iii) QUANTA/CHARMM (Molecular Simulations Incorporated, San Diego, CA), (iv) OPLS-AA (W. L. Jorgensen, "OPLS Force Fields", *Encyclopedia of Computational Chemistry*, Schleyer, Ed., Wiley, New York (1998) Vol. 3, pp. 1986-1989), and (v) Insight II/Discover (Biosym 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 RNA cap binding pocket of PB2 or a binding pocket of a PB2 variant by the same computer methods described in detail above.

In one embodiment of the above-described method of the invention, the RNA cap binding pocket of PB2 or a binding pocket of a PB2 polypeptide variant comprises amino acids Phe323, His357, and Phe404 of PB2 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, Phe404, Phe325, Phe330, and Phe363 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, and Phe363 according to SEQ ID NO: 1 or amino acids corresponding thereto. In yet another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Glu361, and Lys376 according to SEQ ID NO: 1 or amino acids corresponding thereto. In yet another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Ser320, Arg332, Ser337, and Gln406 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Lys339, Arg355, Asn429, and His432 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, and Lys376 according to SEQ ID NO: 1 or amino acids corresponding thereto. In yet another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Ser320, Arg332, Ser337, and Gln406 according to SEQ ID NO: 1 or amino acids corresponding thereto. In yet another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Lys339, Arg355, Asn429, and His432 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Glu361, Lys376, Ser320, Arg332, Ser337, and Gln406 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Glu361, Lys376, Lys339, Arg355, Asn429, and His432 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432 according to SEQ ID

NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, and Gln406 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Lys339, Arg355, Asn429, and His432 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, and Met431 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, and Met431 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, and Met431 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, and Met431 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, His432, and Met431 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Glu361, and Lys376 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe363, Glu361, and Lys376 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Glu361, and Lys376 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, and Ser337 according to SEQ ID

NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids His357, Phe404, Phe325, Glu361, Lys376, Arg332, Ser337, and Met431 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids His357, Phe404, Phe325, Glu361, Lys376, Arg332, Ser337, Met431, Lys339, Arg355, Asn429, and His432 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, and Phe325 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, and Glu361 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, and Lys376 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Glu361, and Lys376 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325 and Glu361 according to SEQ ID NO: 1 or amino acids corresponding thereto. In another embodiment, said binding pocket comprises amino acids Phe323, His357, Phe404, Phe325 and Lys376 according to SEQ ID NO: 1 or amino acids corresponding thereto. Furthermore, in other embodiments, the above defined binding pockets may optionally comprise an amino acid corresponding to amino acid Met431 according to SEQ ID NO: 1 or amino acids corresponding thereto.

In a further aspect of the above-described method of the invention, the RNA cap binding pocket of PB2 is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, and Phe404 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, Phe404, Phe325, Phe330, and Phe363 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, and Phe363 according to Figure 18. In yet another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Glu361, and Lys376 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Ser 320, Arg332, Ser337, and Gln406 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Lys339, Arg355, Asn429, and His432

according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, and Lys376 according to Figure 18. In yet another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Ser320, Arg332, Ser337, and Gln406 according to Figure 18. In yet another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Lys339, Arg355, Asn429, and His432 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Glu361, Lys376, Ser320, Arg332, Ser337, and Gln406 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Glu361, Lys376, Lys339, Arg355, Asn429, and His432 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, and Gln406 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Lys339, Arg355, Asn429, and His432 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, and Met431 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, and Met431 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2

SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, and Met431 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, and Met431 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, His432, and Met431 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Glu361, and Lys376 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Glu361, and Lys376 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Glu361, and Lys376 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Glu361, and Lys376 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, and Ser337 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids His357, Phe404, Phe325, Glu361, Lys376, Arg332, Ser337, and Met431 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids His357, Phe404, Phe325, Glu361, Lys376, Arg332, Ser337, Met431, Lys339, Arg355, Asn429, and His432 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, and Phe325 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, and Glu361 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, and Lys376 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Glu361, and Lys376 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, and Glu361 according to Figure 18. In another embodiment, said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, Phe404, Phe325, and Lys376 according to Figure 18. Furthermore, in other embodiments, the binding pockets as defined above may optionally be additionally

defined by the structure coordinates of PB2 SEQ ID NO:1 amino acid Met431 according to Figure 18.

In one aspect, the present invention provides a method for computational screening according to the above-described method for compounds able to associate with a binding pocket that is a variant to the RNA cap binding pocket of PB2 according to Figure 18. In one embodiment said variant of said binding pocket has a root mean square deviation from the backbone atoms of amino acids Phe323, His357, and Phe404; of amino acids Phe323, Phe404, Phe325, Phe330, and Phe363; of amino acids Phe323, His357, Phe404, Phe325, Phe330, and Phe363; of amino acids Phe323, His357, Phe404, Glu361, and Lys376; of amino acids Phe323, His357, Phe404, Ser 320, Arg332, Ser337, and Gln406; of amino acids Phe323, His357, Phe404, Lys339, Arg355, Asn429, and His432; of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, and Lys376; of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Ser320, Arg332, Ser337, and Gln406; of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Lys339, Arg355, Asn429, and His432; of amino acids Phe323, His357, Phe404, Glu361, Lys376, Ser320, Arg332, Ser337, and Gln406; of amino acids Phe323, His357, Phe404, Glu361, Lys376, Lys339, Arg355, Asn429, and His432; of amino acids Phe323, His357, Phe404, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432; of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, and Gln406; of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Lys339, Arg355, Asn429, and His432; of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432; of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432; of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, and His432, of amino acids Phe323, His357, Phe404, and Met431, of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, and Met431, of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, and Met431, of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, and Met431, of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, Lys376, Ser320, Arg332, Ser337, Gln406, Lys339, Arg355, Asn429, His432, and Met431, of amino acids Phe323, His357, Phe404, Phe325, Glu361, and Lys376, of amino acids Phe323, His357, Phe404, Phe325, Phe363, Glu361, and Lys376, of amino acids Phe323, His357, Phe404, Glu361, and Lys376, of amino acids Phe323, His357, Phe404, Phe325, Phe330, Phe363, Glu361, and Ser337,

of amino acids His357, Phe404, Phe325, Glu361, Lys376, Arg332, Ser337, and Met431, of amino acids His357, Phe404, Phe325, Glu361, Lys376, Arg332, Ser337, Met431, Lys339, Arg355, Asn429, and His432, of amino acids Phe323, His357, Phe404, and Phe325, of amino acids Phe323, His357, Phe404, and Glu361, of amino acids Phe323, His357, Phe404, and Lys376, of amino acids Phe323, His357, Phe404, Glu361, and Lys376, of amino acids Phe323, His357, Phe404, Phe325, and Glu361, of amino acids Phe323, His357, Phe404, Phe325, and Lys376, the above amino acid combinations optionally including amino acid Met431 according to Figure 18 of not more than 3 Å. In another embodiment, the said root mean square deviation is not more than 2.5 Å. In another embodiment, the said root mean square deviation is not more than 2 Å. In another embodiment, the said root mean square deviation is not more than 1.5 Å. In another embodiment, the said root mean square deviation is not more than 1 Å. In another embodiment, the said root mean square deviation is not more than 0.5 Å..

In a preferred embodiment, m<sup>7</sup>GTP is sandwiched between His357 on the solvent side and a cluster of five phenylalanines on the protein side, principally Phe323 and Phe404, but also Phe325, Phe330, and Phe363. In this preferred embodiment, specific recognition of the guanosine base is principally achieved by Glu361 hydrogen bonding to N2 and N1 positions of the guanine, which also helps to neutralize the delocalized positive charge of the m<sup>7</sup>GTP. Also Lys376 makes a long hydrogen bond to the O6. There are two well-ordered, buried water molecules in the ligand pocket which interact with Glu361, Lys376, and Gln406 but not directly with the ligand. Important second layer residues are preferably Arg332 and Ser337 which hydrogen bond to His357. Within hydrogen bonding distance of the N2 of the base is either a water molecule or Ser320. The N7 methyl group is in van der Waals contact with the side-chain of Gln406 (3.4 Å) and the carbonyl oxygen of Phe404 (3.4 Å) is in contact with the side chain of Met431 slightly further away. In this embodiment, the triphosphate is bent round towards the base. The alpha-phosphate interacts with His432 and Asn429 and the gamma-phosphate interacts with basic residues His357, Lys339, and Arg355.

If computer modeling according to the methods described hereinabove indicates binding of a compound to the RNA binding pocket of PB2 or the binding pocket of a PB2 polypeptide variant, said compound may be synthesized and optionally the ability of said compound to bind to said binding pocket may be tested *in vitro* or *in vivo* comprising the further step of (e) synthesizing said compound, and optionally (f) contacting said compound with the PB2 polypeptide fragment or variant thereof or the recombinant host cell of the invention and a RNA

cap or analog thereof to determine the ability of said compound to inhibit binding between said PB2 polypeptide fragment and said RNA cap or analog thereof. The quality of fit of such compounds to the binding pocket may be judged either by shape complementarity or by estimated interaction energy (E. C. Meng et al., J. Comp. Chem., 13, pp. 505-524 (1992).

- 5 Methods for synthesizing said compounds are well known to the person skilled in the art or such compounds may be commercially available. Examples for methods for determining said RNA cap binding inhibitory effect of the identified compounds are described hereinafter.

- 10 It is another aspect of the invention to provide a compound identifiable by the above-described method, under the provision that said compound is not any of the RNA cap analogs selected from the group consisting of m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, and m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, and m<sup>7</sup>GpppUm or the small molecule inhibitors of RNA cap binding 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one (RO0794238, Hooker et al., 2003) or T-705 (Furuta et al., 2005) and is able to bind to the RNA
- 15 cap binding pocket of PB2. In another aspect, the present invention refers to a compound identifiable by the above-described method, under the provision that said compound is not any of the RNA cap analogs selected from the group consisting of m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, and m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, and m<sup>7</sup>GpppUm or the small molecule inhibitors of RNA cap binding 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-
- 20 dihydro-purin-6-one or T-705 and is able to inhibit binding between the PB2 polypeptide fragment and the RNA cap or analog thereof. 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,
- 25 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
- 30 compounds or elements, fluorescent compounds or metals, chemiluminescent compounds and bioluminescent compounds.

In a further aspect, the present invention provides a method for identifying compounds which associate with the RNA cap binding pocket of PB2 or a binding pocket of a PB2 polypeptide

variant, comprising the steps of (i) contacting the PB2 polypeptide fragment or the recombinant host cell of the invention with a test compound and (ii) analyzing the ability of said test compound to bind to the PB2 polypeptide fragment or variant thereof.

- 5 In one embodiment, the interaction between the PB2 polypeptide fragment or variant thereof and a test compound may be analyzed in form of a pull down assay. For example, the PB2 polypeptide fragment may be purified and may be immobilized on beads. In one embodiment, the PB2 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
- 10 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.
- 15 In another embodiment, the interaction between the PB2 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 PB2 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,
- 20 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
- 25 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 soluble PB2 polypeptide fragment or variants thereof
- 30 according to the invention. Binding of said polypeptide may be verified by a PB2 polypeptide fragment specific antibody and chemiluminescence or colorimetric reactions as described above.

In a further embodiment, purified soluble PB2 polypeptide fragments may be incubated with a peptide array and binding of the PB2 polypeptide fragments to specific peptide spots

corresponding to a specific peptide sequence may be analyzed, for example, by PB2 polypeptide specific antibodies, antibodies that are directed against an epitope-tag fused to the PB2 polypeptide fragment, or by a fluorescence signal emitted by a fluorescent tag coupled to the PB2 polypeptide fragment.

5

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 PB2 polypeptide fragment may be, for example, verified by immunoprecipitation of said polypeptide and verification of the presence of the label.

10

In a preferred embodiment, the above-described method for identifying compounds which associate with the RNA cap binding pocket of PB2 or a binding pocket of a PB2 polypeptide variant comprises the further step of adding a RNA cap or analog thereof. In another embodiment of said method, the ability of said test compound to bind to PB2 in presence of said RNA cap or analog thereof or the ability of said test compound to inhibit binding of said RNA cap or analog thereof to PB2 is analyzed. A compound is considered to inhibit RNA cap or RNA cap analog binding if binding is reduced by the compound at the same molar concentration as the RNA cap or RNA cap analog by more than 20%, by more than 30%, by more than 40%, by more than 50%, preferably by more than 60%, preferably by more than 70%, preferably by more than 80%, preferably by more than 90%. In preferred embodiments, the above-described pull down, ELISA, peptide array, FRET, and co-immunoprecipitation experiments may be carried out in presence of an RNA cap structure or an analog thereof, preferably a labeled RNA cap structure or analog thereof, in presence or in absence of a test compound. In one embodiment, the RNA cap or analog thereof is added prior to addition of said test compound. In a further embodiment, the RNA cap or analog thereof is added concomitantly with addition of said test compound. In yet another embodiment, the RNA cap or analog thereof is added after addition of said test compound.

20

25

30

In a preferred embodiment, the ability of the identifiable test compound to interfere with the interaction of a PB2 polypeptide fragment of the invention to a RNA cap or analog thereof may be tested by incubating said polypeptide fragment with 7-methyl-GTP Sepharose 4B resin (GE

Healthcare) in presence or absence of said compound and comparing, preferably quantifying, the amount of bound PB2 polypeptide fragment with and without said compound, e.g., on a coomassie stained SDS PAGE gel or using Western blot analysis.

5 The ability of the RNA cap or analog thereof to associate with the PB2 fragment or variant thereof according to the present invention in presence or absence of the test compound is evaluated. In one embodiment, this may be achieved using a fluorescein-labeled RNA cap analog, fluorescein-labeled 7-methyl-guanosine monophosphate (m<sup>7</sup>GMP), in a fluorescence polarization assay as described by Natarajan et al. (2004). Alternatively, in another embodiment  
10 a ribose diol-modified fluorescent cap analog, anthraniloyl-m<sup>7</sup>GTP, may be used in a fluorescence spectroscopic assay as set forth by Ren et al. (1996). In a further embodiment, radioactively labeled RNA caps may be incubated with the PB2 polypeptide fragment or variant thereof in presence or absence of a test compound, wherein the RNA cap may be added prior to, concomitantly with, or after addition of the test compound. The cap binding reaction is UV-  
15 crosslinked, denatured, and analyzed by gel electrophoresis as described by Hooker et al. (2003). In another embodiment, the PB2 polypeptide fragments of the invention may be immobilized on a microtiter plate, incubated with a labeled RNA cap or analog thereof in presence or absence of a test compound, wherein the RNA cap or analog thereof may be added prior to, concomitantly with, or after addition of the test compound, and cap binding is analyzed by verifying the  
20 presence of the label after thorough washing of the plate. Signal intensities of the RNA cap or RNA cap analog label in wells with test compound to signal intensities of said label without test compound are compared and a compound is considered to inhibit RNA cap or RNA cap analog binding if binding is reduced by more than 50%, preferably by more than 60%, preferably by more than 70%, preferably by more than 80%, preferably by more than 90% as described above.

25 In a preferred embodiment, the above-described method for identifying compounds which associate with the RNA cap binding pocket of PB2 or a binding pocket of a PB2 polypeptide variant, preferably inhibit binding of RNA caps or analogs thereof, is performed in a high-throughput setting. In a preferred embodiment, said method is carried out in a multi-well  
30 microtiter plate as described above using immobilized PB2 polypeptide fragments or variants thereof according to the present invention and labeled RNA caps or analogs thereof. 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. The  
5 RNA caps or analogs thereof may be added prior to, concomitantly with, or after the addition of the library compound. The ability of the compound to inhibit RNA cap binding may be assessed as described above.

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

In a further embodiment of the invention, the test compound applied in any of the above described methods is a small molecule. In a preferred embodiment, said small molecule is  
20 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*  
25 identification of compounds that associate with the RNA cap binding pocket of the PB2 polypeptide fragment or variant according to the invention and/or inhibit RNA cap binding to said binding pocket, said methods further comprise the step of formulating the identifiable compound or a pharmaceutically acceptable salt thereof with one or more pharmaceutically acceptable excipient(s) and/or carrier(s). In another aspect the present invention provides a  
30 pharmaceutical composition producible according to the afore-mentioned method. A compound according to the present invention can be administered alone but, in human therapy, will generally be administered in admixture with a suitable pharmaceutical excipient, diluent, or carrier selected with regard to the intended route of administration and standard pharmaceutical practice (see hereinafter).

In the aspect of computational modeling or screening of a binding partner for the RNA cap binding pocket of PB2 or a binding pocket of a PB2 variant, 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", 8<sup>th</sup> Edition, L. S. Goodman, A. Gilman, T. W. Rall, A. S. Nies, & P. Taylor, Eds., Pergamon Press (1985); W. L. Jorgensen & E. M. Duffy, *Bioorg. Med. Chem. Lett.*, 10, pp. 1155-1158 (2000). 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 (C. A. Lipinski et al., *Adv. Drug Deliv. Rev.*, 23, pp. 3-25 (1997)). 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 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 embodiment the present invention provides a compound identifiable by the methods described above, comprising the step of contacting a PB2 polypeptide fragment or PB2 polypeptide fragment variant of the present invention or a recombinant host cell of the present invention with a test compound, which is different from the RNA cap analogs selected from the group consisting of m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, and m<sup>7</sup>GpppUm, and the small molecule inhibitors of cap binding 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one (RO0794238, Hooker et al., 2003) and T-705 (Furuta et al., 2005) and is able to bind to PB2. In a preferred aspect, the present invention relates to a compound identifiable by the methods described above, comprising the step of contacting a PB2 polypeptide fragment or variant thereof or a recombinant host cell of the present invention with a test compound, which is different from the RNA cap analogs selected from the group consisting of m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, and m<sup>7</sup>GpppUm, and the small molecule inhibitors of cap binding 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one and T-705 and is able to inhibit binding between PB2 and the RNA cap or analog thereof.

In another aspect, the present invention provides an antibody directed against the RNA cap binding domain of PB2. In a preferred embodiment, said antibody recognizes the RNA cap binding domain of a polypeptide fragment or variant of a fragment indicated above, in particular selected from the group of polypeptides defined by SEQ ID NO: 14 to 22. In particular, said antibody specifically binds to an epitope comprising one or more of above indicated amino acids, which define the binding pocket. In that context, the term epitope has its art recognized meaning and preferably refers to stretches of 4 to 20 amino acids, preferably 5 to 18, 6 to 16, or 7 to 14 amino acids. Accordingly, preferred epitopes have a length of 4 to 20, 5 to 18, preferably 6 to 16 or 7 to 14 amino acids and comprise one or more of Ser320, Phe323, Phe325, Phe330, Arg332, Ser337, Lys339, Arg355, His357, Glu361, Phe363, Lys376, Phe404, Gln406, Met431 and/or

His432 of SEQ ID NO: 1 or a corresponding amino acid. 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')<sub>2</sub>, 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 RNA cap binding domain of PB2 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 RNA cap binding domain of PB2 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 RNA cap binding domain of PB2 or a fragment thereof. Cells that produce antibodies directed against the RNA cap binding domain of PB2 or fragments 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", E. Harlow and D. Lane, Eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1990), which is incorporated herein by reference, for support regarding generation of antibodies.

In another aspect, the present invention relates to the use of a compound identifiable by the above-described methods that is able to bind to the RNA cap binding pocket of PB2 or a binding pocket of a PB2 polypeptide variant and/or is able to inhibit binding of an RNA cap or analog thereof to said binding pocket, the pharmaceutical composition described above, or the antibody of the present invention for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with negative-sense single stranded RNA viruses. In a preferred embodiment, said disease conditions are caused by viral infections with non-segmented negative-sense single stranded RNA viruses, i.e., the order of Mononegavirales, comprising the Bornaviridae, Filoviridae, Paramyxoviridae, and Rhabdoviridae families. In a more preferred embodiment, said disease condition is caused by segmented negative-sense single stranded RNA viruses comprising the family of Orthomyxoviridae, including the genera Influenza A virus, Influenza B virus, Influenza C virus, Thogotovirus, and Isavirus, the families of Arenaviridae and Bunyaviridae, including the genera Hantavirus, Nairovirus, Orthobunyavirus, Phlebovirus, and Tospovirus. In an even more preferred embodiment said disease condition is caused by an infection with a virus species selected from the group consisting of Borna disease virus, Marburg virus, Ebola virus, Sendai virus, Mumps virus, Measles virus, Human respiratory syncytial virus, Turkey rhinotracheitis virus, Vesicular stomatitis Indiana virus, Nipah virus, Henda virus, Rabies virus, Bovine ephemeral fever virus, Infectious hematopoietic necrosis virus, Thogoto virus, Influenza A virus, Influenza B virus, Influenza C virus, Hantaan virus, Crimean-congo hemorrhagic fever virus, Rift Valley fever virus, La Crosse virus, preferably Influenza A virus, Influenza B virus, Influenza C virus, Thogoto virus, or Hantaan virus, more preferably Influenza A virus, Influenza B virus, Influenza C virus, most preferably Influenza A virus.

For treating, ameliorating, or preventing said disease conditions the medicament of the present invention can be administered to an animal patient, preferably a mammalian patient, preferably a human patient, orally, buccally, sublingually, intranasally, via pulmonary routes such as by inhalation, via rectal routes, or parenterally, for example, intracavernosally, intravenously, intra-arterially, intraperitoneally, intrathecally, intraventricularly, intraurethrally 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. 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 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  
5 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  
10 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.,  
15 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.,  
20 sorbitan trioleate.

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  
25 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  
30 present invention may be varied or adjusted from about 1 mg to about 1000 mg per m<sup>2</sup>, preferably about 5 mg to about 150 mg/m<sup>2</sup> according to the particular application and the potency of the active component.

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

## EXAMPLES

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

### Example 1: Generation of PB2 expression constructs

An *E. coli* codon optimized PB2 gene (Geneart) SEQ ID NO: 25, based on the amino acid sequence of Influenza A/Victoria/3/1975(H3N2) PB2 (de la Luna et al., 1989), was cloned into vector pET9a (Novagen) modified to introduce 5' AatII/AscI and 3' NsiI/NotI restriction site pairs for directional exonuclease III truncation, and provide a 3' sequence encoding the C-terminal biotin acceptor peptide GLNDIFEAQKIEWHE. A first unidirectional truncation plasmid library was generated from the 3' end (Tarendeau et al., 2007), pooled and used as the substrate for a 5' deletion reaction. Linearised plasmids encoding inserts of 150 to 250 amino acids were isolated from an agarose gel, religated and used to transform *E. coli* BL21 AI containing the RIL plasmid (Stratagene) for expression screening using hybridization of Alexa488 streptavidin (Invitrogen) to colony blots (Tarendeau et al., 2007). Clones were ranked upon their fluorescence signal and the first 36 tested for expression of purifiable protein by Ni<sup>2+</sup> affinity chromatography. Thirteen clones were identified and sequenced revealing 7 unique constructs in the region of interest (see Figure 1):

nt 703-1488 (aa 235-496),

nt 721 -1449 (aa 241-483),

nt 802-1449 (aa 268-483),  
nt 829-1440 (aa 277-480),  
nt 841-1446 (aa 281-482),  
nt 868-1449 (aa 290-483),  
5 nt 883-1446 (aa 295-482)

These PB2 polynucleotide fragments were generated by polymerase chain reaction (PCR) using the synthetic gene as template and cloned into the expression vector pETM11 plasmid (EMBL) between NcoI and XhoI sites. The pETM11 expression vector allows for expression of His<sub>6</sub>-tagged fusion protein. The His<sub>6</sub>-tag can be cleaved off the fusion protein using TEV protease.

10

#### Example 2: Identification of a minimal cap-binding fragment

PB2 polypeptide fragments 235-496, 241-483, 268-483, and 290-483 were expressed in *E. coli* using the plasmid constructs described in Example 1 and the general expression and purification  
15 protocol described in Example 3. It was found that purified PB2 fragment 235-496 is only soluble at low concentration, whereas fragments 241-483, 268-483 and 290-483 are more soluble and could bind specifically to a m<sup>7</sup>GTP sepharose column, indicating cap-binding activity. For these assays 0.2 mg of protein was loaded onto 50µL of 7-methyl-GTP Sepharose 4B resin (GE Healthcare) and incubated for 3 h at 4°C for binding. After washing with buffer containing 50  
20 mM Tris•HCl (pH 8.0), 200 mM NaCl, 2 mM DTT, the proteins were eluted by centrifugation with the same buffer containing 1 mM m<sup>7</sup>GTP and analysed on SDS-PAGE. During the course of these experiments it was noted that a smaller degradation fragment could also bind m<sup>7</sup>GTP and this was identified by N-terminal sequencing and mass-spectroscopy to be residues 318-483 and furthermore this fragment was proteolytically stable. The corresponding polynucleotide  
25 fragment (nt 952-1449) was cloned into pETM11 as described in Example 1.

#### Example 3: PB2 fragment expression and purification

Said plasmids were transformed in *E. coli* cells using chemical transformation. The protein was  
30 expressed in *E. coli* strain BL21-CodonPlus-RIL (Stratagene) in LB medium. After a 16 to 20 h hours induction at 25°C with 0.2 mM isopropyl-β-thiogalactopyranoside (IPTG), the cells were harvested and re-suspended in a lysis buffer (50 mM Tris•HCl (pH 8.0), 300 mM NaCl, 5 mM 2-mercaptoethanol) and sonicated. After centrifugation, the cleared lysate was directly loaded on a nickel affinity column (Chelating sepharose from GE Healthcare, loaded with Ni<sup>2+</sup> ions). The Ni-

sepharose resin was extensively washed with a 50 mM Tris•HCl (pH 8.0), 1 M NaCl, 15 mM imidazole, 5 mM 2-mercaptoethanol buffer, and then with a 50 mM Tris•HCl (pH 8.0), 200 mM NaCl, 50 mM imidazole, 5 mM 2-mercaptoethanol buffer. The protein was then eluted with a 50 mM Tris•HCl (pH 8.0), 200 mM NaCl, 0.5 M imidazole, 5 mM 2-mercaptoethanol buffer. The purified protein was incubated for TEV protease cleavage overnight at 10°C. After dialysis against 50 mM Tris•HCl (pH 8.0), 200 mM NaCl, 15 mM imidazole, 5 mM 2-mercaptoethanol buffer, a second nickel affinity step was performed to remove His-tag labeled TEV protease. The unbound protein was recovered and concentrated at 5-8 mg/ml for gel filtration on Superdex 75 column (GE Healthcare) in order to improve purity. The result of gel filtration for PB2 polypeptide fragment (amino acids 318 to 483) is shown in FIG 2A. This fragment (amino acids 318 to 483) could be concentrated up to 24 mg/ml. The purity of the protein was evaluated by SDS PAGE and coomassie staining (see FIG. 2B). A typical yield is 120 mg of pure protein per liter of bacterial culture.

#### 15 Example 4: Co-crystallization of PB2 and m<sup>7</sup>GTP

For crystallization trials a protein solution of PB2 fragment 318-483 at 11-15 mg/ml in 50 mM Tris•HCl (pH 8.0), 200 mM NaCl, 2 mM dithiothreitol (DTT), 5 mM m<sup>7</sup>GTP (Sigma) was used. A robot (Cartesian) was used to screen 576 crystallization conditions using the sitting drop vapor diffusion method, by mixing 100 µl of protein solution with 0.1 µl of different precipitant solutions. Crystals were identified in two distinct conditions. These conditions were manually optimized using the hanging drop vapor diffusion method with 1 µl of protein solution mixed with 1 µl of precipitant solution. The best crystals were obtained with a precipitant solution comprising 0.1 M citric acid pH 4.6, 1.6-1.8 M sodium formate. These crystals were of space group C222<sub>1</sub>, with unit cell dimensions a = 9.2 nm, b = 9.4 nm; c = 22.0 nm ± 0.3 nm Å. Some of the smaller crystals resulted in X-ray diffraction with a resolution of at least 2.4 Å. Larger crystals were found to be unusable in that they were not single crystals. In the second crystallization condition found, the precipitant solution comprises, 7% PEG6000, 1 M LiCl, 0.1 M citric acid pH 5.0, 10 mM TCEP HCl. This condition gives thin plate-like crystals, of undetermined space-group, with evidence of diffraction to at least 3.3 Å resolution, but the diffraction quality is much inferior and these crystals were not pursued.

#### Example 5: Preparation of seleno-methionine labeled protein.

Seleno-methionated protein was produced in minimum M9 medium as described (Van Duyne, 1993). Since fully seleno-methionated protein did not crystallize, partially labeled protein was produced using 50 mg/L of selenomethionine and 5 mg/L of methionine, instead of 60 mg/l of selenomethionine. Expression, lysis, and purification conditions were as for native protein (Example 3). Electrospray mass-spectroscopy showed that, whereas the fully seleno-methionated protein had a MW of 19654 Daltons, corresponding to 11/11 methionines being substituted, the partially labeled protein had an average molecular weight of 19578 (9.4/11, i.e., 85% substituted) with individual peaks being obtained for protein species with 7-11 seleniums.

#### 10 Example 6: Structure determination and refinement

The structure was determined with the single wavelength anomalous diffraction (SAD) method using partially seleno-methionylated protein as described in Example 5. Partially seleno-methionylated protein actually gave rise to larger single crystals than wild-type. Native data was collected at 2.4 Å resolution on a small crystal (size 15 x 15 x 50  $\mu\text{m}^3$ ) of space-group C222<sub>1</sub> and cell dimensions a=92.2, b=94.4, c=220.4 Å on beamline BM14 at the ESRF (Figure 4). Partially seleno-methionine labeled protein crystals were measured on ID29 at 2.9 Å resolution for structure determination and later at 2.3 Å resolution on ID14-EH1 on a larger crystal for structure refinement. The structure was solved by the SAD method using AUTOSHARP (Vonrhein et al., 2006). 50/55 selenium atom positions were found. Phases and the map were improved with RESOLVE (Terwilliger, 2000) making use of 4-fold non-crystallographic symmetry determined from the selenium positions, although it turned out that there are actually five molecules in the asymmetric unit (solvent content 55%). These are arranged in an unusual way. Four well-ordered molecules (average B-factor about 30Å<sup>2</sup>), denoted A-D are arranged with 4-fold point symmetry. The fifth molecule mediates the packing of the tetramers in the crystallographic c-direction but is partially disordered (average B-factor about 51Å<sup>2</sup>). After transferring phases to the native data, ARP/wARP (Perrakis et al., 1999) was able to automatically build most of the structure. Clear residual electron density was observed for the m<sup>7</sup>GTP bound to each molecule. Refinement was performed with REFMAC (Murshudov, 1997) using tight NCS restraints for the majority of each molecule and TLS parameters for each molecule as a whole. The final R-factor (R-free) is 0.186 (0.235) with 293 water molecules. According to MOLPROBITY (Lovell et al., 2003), the quality of the structure is excellent, as indicated by this server output:

REMARK 40

REMARK 40 MOLPROBITY STRUCTURE VALIDATION

REMARK 40 MOLPROBITY OUTPUT SCORES:

REMARK 40 ALL-ATOM CLASHSCORE : 8.40

5 REMARK 40 BAD ROTAMERS : 3.4% 24/708 (TARGET 0-1%)

REMARK 40 RAMACHANDRAN OUTLIERS : 0.0% 0/795 (TARGET 0.2%)

REMARK 40 RAMACHANDRAN FAVORED : 97.7% 777/795 (TARGET 98.0%)

10 The cap-binding domain of PB2 has a compact, well-ordered mixed  $\alpha$ - $\beta$  fold with a prominent anti-parallel beta-sheet comprising strands  $\beta$ 1,  $\beta$ 2,  $\beta$ 5,  $\beta$ 6 and  $\beta$ 7 packed onto a helical bundle comprising  $\alpha$ 1,  $\alpha$ 2,  $\alpha$ 3 and  $\alpha$ 4, with the longest helices  $\alpha$ 1 and  $\alpha$ 3 being aligned roughly anti-parallel. The connection between  $\beta$ 2 and  $\beta$ 5 is extended into a long beta hairpin, comprising strands  $\beta$ 3 and  $\beta$ 4 and 348-loop, which crosses over to the helical side of the molecule. Between  $\alpha$ 2 and  $\alpha$ 3, a poorly ordered loop, designated the 420-loop, juts into the solvent Figures 4 and 5).

15 In the Influenza A PB2 cap-binding domain, the m<sup>7</sup>G sits on a hydrophobic platform formed principally by Phe-323 (from  $\beta$ 1, more stacked against the ribose) and Phe-404 (from the C-terminal of helix  $\alpha$ 1, more stacked against the base), but neither are exactly parallel to the base. These two aromatic residues are part of a remarkable cluster of five phenylalanines including also Phe-325, Phe-330 and Phe-363. On the solvent side of the ligand, the sandwich is completed

20 by His-357 (from the C-terminal of  $\beta$ 4), a residue type not previously observed in this role. The key acidic residue is Glu-361 which hydrogen bonds to the N1 and N2 positions of the guanine and Lys-376 is also involved in base specific recognition via an interaction with O6 (Figures 6, 7, and 10). Unusually for cap-binding proteins, there are no direct interactions with either ribose hydroxyl. The triphosphate is bent round towards the base. The alpha-phosphate interacts with

25 His-432 and Asn-429 and the gamma-phosphate interacts with basic residues His-357, Lys-339 and Arg-355.

#### Example 7: Mutational analysis of the cap-binding site

30 To verify that the structurally observed protein-ligand interactions are important for *in vitro* cap-binding, we made alanine mutations of seven m<sup>7</sup>GTP-contacting residues (F323, F325, E361, F363, F404, H357, K376) using the QuickChange site directed mutagenesis kit (Stratagene). These single point mutants were expected to show reduced cap-binding activity, which was assessed by their ability to bind to m<sup>7</sup>GTP-Sepharose resin at 4°C (Table 4, Figure 12). The

mutants were purified as described in Example 3 for wild-type PB2. Mutant F363A expressed poorly and with low solubility in *E. coli*, presumably due to misfolding, and was not further pursued. Mutants H357A, E361A, K376A and F404A expressed well in soluble form but failed to bind to m<sup>7</sup>GTP-Sepharose under the assay conditions, whereas F323A had weak binding activity (Figure 12). Surprisingly, mutant F325A, though poorly expressed, bound the resin nearly as well as the wild-type. We show that this mutant's activity is strongly temperature-dependent (Table 3). A H357W substitution was also made and consistent with the expected improved stacking with the base, this mutant domain showed enhanced m<sup>7</sup>GTP binding. Results are shown in Figure 12 and Table 4.

#### Example 8: Quantitative analysis of cap-binding affinity using surface Plasmon resonance measurements

To provide a more quantitative comparison of the binding affinity of the various mutants we performed surface plasmon resonance (SPR) measurements with immobilised cap-binding domain and m<sup>7</sup>GTP as the analyte. SPR analysis was performed on a Biacore T100 machine equipped with CM5 sensor chips (Biacore AB, Uppsala, Sweden). Wild-type and mutant PB2 cap-binding domains were diluted to 50 µg/mL in 10 mM sodium acetate, pH 6.0 and immobilized to the sensor surface using standard amine-coupling chemistry. Activated/deactivated sensor surface served as reference. Two-fold serial dilutions of m<sup>7</sup>GTP, m<sup>7</sup>GpppG and GTP ranging from 0.47-1 mM were injected over the immobilized PB2 proteins and reference surfaces at a flow rate of 30 µL/min. Running buffer was 10 mM Hepes, pH 7.4 containing 150 mM NaCl and 1 mM DTT. To test the effects of temperature upon binding, the experiments were performed at 5, 15, 25 and 37°C. All acquired sensorgrams were processed using double referencing and the equilibrium dissociation constants were determined by steady state affinity evaluation with the Biacore T100 Evaluation Software. These are referred to as 'apparent' dissociation constants as the surface chemistry and molecular orientation associated with immobilisation could affect the values. The errors are those associated with the fitting as reported by the software. Variability between experiments with different immobilizations of proteins was < 25% (data not shown). Results for the apparent equilibrium disassociation constant K<sub>D</sub> obtained at 25°C (Table 1) show a trend consistent with the m<sup>7</sup>GTP-Sepharose binding results, with the highest affinity being for the H357W mutant (K<sub>D</sub> = 24 ± 3 µM) followed by the wild-type (K<sub>D</sub> = 177 ± 34 µM), and the E361A and K376A mutants having the lowest affinity (K<sub>D</sub>>1500 µM).

5

Table 1: Apparent equilibrium dissociation constants ( $K_D$ ) for interaction between  $m^7GTP$  and different point mutants of the PB2 cap-binding domain determined by SPR at 25°C.

10

| protein          | $K_D$ ( $\mu M$ ) |
|------------------|-------------------|
| <b>Wild-type</b> | $177 \pm 34$      |
| <b>H357W</b>     | $24 \pm 3$        |
| <b>F323A</b>     | $285 \pm 45$      |
| <b>F325A</b>     | $418 \pm 99$      |
| <b>H357A</b>     | $831 \pm 45$      |
| <b>F404A</b>     | $1277 \pm 120$    |
| <b>E361A</b>     | $>1500$           |
| <b>K376A</b>     | $>1500$           |

15

Taken together, the above data confirm the functional importance of these contact residues to  $m^7GTP$  by the PB2 domain *in vitro*. Additional results for the dinucleotide cap analogue,  $m^7GpppG$ , and GTP binding to the wild-type cap-binding domain are shown in Table 2.

20

Table 2: Apparent equilibrium dissociation constants ( $K_D$ ) for interaction between wild-type cap-binding domain of PB2 and different cap analogues determined by SPR at 25°C, compared to values for  $IC_{50}$  measured by inhibition of capped RNA cross-linking by cap analogues (Hooker *et al.* 2003)

25

| Analyte                      | $K_D$ ( $\mu M$ ) | $IC_{50}$ ( $\mu M$ ) |
|------------------------------|-------------------|-----------------------|
| <b><math>m^7GpppG</math></b> | $171 \pm 19$      | 127                   |
| <b><math>m^7GTP</math></b>   | $177 \pm 34$      | 113                   |
| <b>GTP</b>                   | $1019 \pm 100$    | 550                   |

30

To investigate the temperature dependence of binding between the cap-binding domain and  $m^7GTP$  we made SPR measurements of  $K_D$  on wild-type and selected mutant domains at various temperatures up to 37°C (Table 3). The cap-binding affinity of all proteins tested decreased significantly with increasing temperature. However, whereas the apparent  $K_D$  of the wild-type increased by a factor of 2.7 between 5 and 37°C, that of the F325A mutant increased by more

than 20 times. Moreover, unlike the wild-type domain, the binding activity of the F325A mutant is irreversibly lost after incubation at 37°C (data not shown), presumably due to thermal denaturation. This temperature sensitivity therefore provides a consistent rationale for the observed near normal cap-binding activity of the F325A mutant at 4°C, both in the domain and the trimer, and the functional impairment of the F325A mutant polymerase at 37°C (cf. Example 10).

Table 3: Temperature dependence of dissociation constants ( $K_D$ ) for interaction between  $m^7$ GTP and different point mutants of the cap-binding domain.

| Temperature | $K_D$ ( $\mu$ M) |        |           |            |             |
|-------------|------------------|--------|-----------|------------|-------------|
|             | Wild-type        | H357W  | F323A     | H357A      | F325A       |
| 5°C         | 100 ± 20         | 12 ± 2 | 168 ± 29  | 394 ± 26   | 170 ± 44    |
| 15°C        | 133 ± 26         | 16 ± 3 | 215 ± 38  | 551 ± 42   | 223 ± 55    |
| 25°C        | 177 ± 34         | 24 ± 3 | 285 ± 45  | 831 ± 44   | 419 ± 99    |
| 37°C        | 268 ± 45         | 45 ± 6 | 884 ± 140 | 1123 ± 130 | 3787 ± 1700 |

#### Example 9: Cap-binding activity of wild-type and mutant polymerase complexes

To investigate the relevance of the cap-binding site for the biological activity of influenza RNPs we introduced the same mutations into full-length PB2 of influenza A/Victoria/3/75 and tested the activity of the resultant trimeric polymerase or recombinant RNPs. None of the mutations altered the accumulation of PB2 when expressed in the context of the polymerase complex, nor changed the capacity of PB2 to form trimeric polymerase complexes (Table 4).

For cap-binding of trimeric polymerase, HEK293T cells were co-transfected with plasmids pCMVPB1 encoding the PB1 subunit, pCMVPB2His encoding the PB2 wild-type subunit or mutants thereof, and pCMVPA encoding the PA subunit of the influenza virus RNA polymerase and cell extracts were prepared 24 h post-transfection. The extracts were prepared in a buffer containing 50 mM Tris-HCl, 100 mM NaCl, 5 mM EDTA, 1 mM DTT, 0.2% NP40, protein inhibitors (Roche, Complete cocktail) at pH 7.5 and incubated with  $m^7$ GTP-Sepharose (GE Healthcare) for 3 h at 4°C. After washing with 100 volumes of the same buffer, the protein retained was eluted with the same buffer containing 1 mM  $m^7$ GTP and was analysed by Western blot using PA-specific antibodies (Figure 13). Under these conditions wild-type polymerase was not retained on a control Sepharose resin. Mutant H357W behaved as wild-type, whereas all

other mutations abolished binding, except F325A that still bound, mirroring the results the results obtained for the isolated domain.

#### Example 10: Replication and transcription activities of wild-type and mutant recombinant mini-RNPs

The replication activity of the wild-type or mutant polymerases was tested by measuring the accumulation of RNPs in a recombinant system *in vivo* (Ortega et al., 2000; Martin-Benito et al., 2001). For RNP reconstitution, HEK293T cells were co-transfected with plasmids pCMVPB1, pCMVPB2His, pCMVPA, pCMVNP and pHHCclone23 using a calcium-phosphate precipitation protocol. At 24 h post-transfection, extracts were prepared and RNPs were purified by chromatography over Ni<sup>2+</sup>-NTA-agarose resin as described (Area et al., 2004). Western-blotting was carried out using antibodies specific for PA and NP. Progeny RNPs were extracted, purified and their accumulation monitored using anti-PA and anti-NP antibodies. Most mutations did not alter substantially the replication activity of the RNPs (Figure 14 and Table 4). Only mutations F325A and K376A reduced partially the accumulation of RNPs.

The transcription activity of purified wild-type or mutant RNPs was tested *in vitro* using either ApG or  $\beta$ -globin mRNA as primers, to reveal cap-independent or cap-dependent transcription, respectively. For *in vitro* RNA synthesis, equal amounts of purified RNPs, as determined by Western blot were incubated in 20  $\mu$ l reaction mixtures containing 50 mM Tris-HCl (pH 8.0), 2 mM MgCl<sub>2</sub>, 100 mM KCl, 1 mM dithiothreitol, 0.5 mM each ATP, CTP, and UTP, 10  $\mu$ M  $\alpha$ -<sup>32</sup>P-GTP (0.5  $\mu$ Ci/nmol), 10  $\mu$ g/ml actinomycin D, 1 U/ml of human placental RNase inhibitor and either 100  $\mu$ M ApG or 10  $\mu$ g/ml  $\beta$ -globin mRNA, for 60 min at 30°C. The synthesized RNA was TCA-precipitated and recovered by filtration in a dot-blot apparatus. Incorporation was quantified with a phosphorimager. Most mutant RNPs were fully active when primed with ApG (Figure 15 and Table 4): Mutations H357A and F404A led to RNPs transcriptionally more efficient than wild-type and only mutant F325A showed a strong deficiency.

In contrast,  $\beta$ -globin mRNA dependent transcription was severely affected for most of the mutants, except mutations H357W and H357A that were fully and partially active, respectively (Figure 16 and Table 4). The ratio of  $\beta$ -globin versus ApG-primed transcription activity (Figure 17 and Table 4) shows that all mutants except H357W were strongly affected in their ability to use  $\beta$ -globin mRNA as primer for transcription.

All but one of the mutations designed to reduce cap-binding activity abolished interaction with m<sup>7</sup>GTP-Sepharose, both as protein domains and as mutant polymerase complexes (Table 4, Figures 12 and 13). The exception, F325A, was poorly expressed in the domain context and severely affected in most of the functional assays at the polymerase or RNP level at 37°C (Table 4). However it retained reasonable cap-binding affinity at 4°C (Figures 12 and 13) suggesting a possible temperature sensitive phenotype.

Table 4: Summary of biochemical and biological activity of PB2 cap-binding site mutants

|           | Solubility of domain | Cap-binding to domain at 4°C | Cap-binding to polymerase at 4°C | RNP accumulation | ApG-dependent transcription | β-globin-dependent transcription | Ratio β-globin/ApG transcription |
|-----------|----------------------|------------------------------|----------------------------------|------------------|-----------------------------|----------------------------------|----------------------------------|
| Wild-type | +++                  | +++                          | +++                              | +++              | +++                         | +++                              | +++                              |
| F323A     | ++                   | +                            | -                                | +++              | +                           | -                                | -                                |
| F325A     | +/-                  | ++                           | ++                               | ++               | -                           | -                                | -                                |
| H357A     | +++                  | -                            | -                                | ++++             | ++++                        | +                                | -                                |
| H357W     | +++                  | ++++                         | +++                              | ++               | +++                         | ++++                             | +++                              |
| E361A     | ++                   | -                            | -                                | +++              | +++                         | -                                | -                                |
| F363A     | -                    | nd                           | nd                               | nd               | nd                          | nd                               | nd                               |
| K376A     | +++                  | -                            | -                                | +                | +++                         | -                                | -                                |
| F404A     | +++                  | -                            | -                                | ++++             | ++++                        | -                                | -                                |
| ΔVQ       | +++                  | +++                          | +                                | +++              | ++                          | -                                | -                                |

nd: not determined

#### Example 11: Cap-dependent transcription requires integrity of the 424-loop

We were intrigued by the conspicuously exposed 420-loop (Figures 5 and 6), whose sequence is relatively well conserved between influenza A and B (Figure 9). To investigate the possible function of this loop, we shortened it by replacing residues Val421-Gln426 by three glycines (mutant ΔVQ). The isolated cap-binding domain bearing this mutation behaved as wild-type with respect to m<sup>7</sup>GTP-Sepharose binding activity (Figure 12) and apparent K<sub>D</sub> (data not shown). In the context of the trimeric polymerase and recombinant RNPs, this mutant retained the ability to bind to cap-analogue resins, albeit to a reduced extent, and yielded wild-type levels of replication and ApG primed transcription activity (Figures 14 and 15). In contrast, the mutant

was unable to perform  $\beta$ -globin-dependent transcription *in vitro* (Figures 16, 17, and Table 4). The  $\Delta VQ$  mutant thus is defective in cap-dependent transcription but not cap-binding. We speculate that this loop may play an allosteric role in regulating the activity of the PB1 subunit.

## REFERENCES CITED

- Area, E. et al., (2004), "3D structure of influenza virus polymerase complex: localization of subunit domains", PNAS USA, 101, pp. 308-313
- 5
- Balbes, L. M. et al., (1994), "A Perspective in Modern Methods in Computer-Aided Drug Design", Reviews in Computational Chemistry, Vol. 5, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, pp. 37-380
- 10
- Bartlett, P. A. et al., (1989), "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules", in Molecular Recognition in Chemical and Biological Problems", Special Publication, Royal Chem. Soc., 78, pp. 182-196
- Berge, S. M. et al., (1977), "Pharmaceutical Salts", J. Pharm. Sci., 66, pp. 1-19
- 15
- Blundell, T. L. and Johnson, L. N., (1976), "Protein Crystallography", Academic Press, New York
- Bohm, H.-J., (1992), "The Computer Program LUDI: A New Method for the *De Novo* Design of Enzyme Inhibitors", J. Comp. Aid. Mol. Des., 6, pp. 61-78
- 20
- Cianci, C. et al., (1995), "Differential activation of the influenza virus polymerase via template RNA binding", J. Virol., 69, pp. 3995-3999.
- 25
- Cohen, N. C. et al., (1990), "Molecular Modeling Software and Methods for Medicinal Chemistry", J. Med. Chem., 33, pp. 883-894
- Drenth, J., "Principles of protein X-ray crystallography", 2<sup>nd</sup> Ed., Springer Advanced Texts in Chemistry, New York (1999)
- 30
- Eisen, M. B. et al., (1994), "HOOK: A Program for Finding Novel Molecular Architectures that Satisfy the Chemical and Steric Requirements of a Macromolecule Binding Site", Proteins: Struct., Funct., Genet., 19, pp. 199-221

- Eriksson, B. et al., (1977), Inhibition of Influenza virus ribonucleic acid polymerase by ribavirin triphosphate. *Antimicrob. Agents Chemother.*, 11, pp. 946-951
- 5 Fechter, P. et al., (2003), "Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding", *J. Biol. Chem.*, 278, pp. 20381-20388
- Furuta, Y. et al., (2005), "Mechanism of action of T-705 against influenza virus", *Antimicrob. Agents Chemother.*, 49, pp. 981-986
- 10 Ghanem, A. et al., (2007), "Peptide-mediated interference with influenza A virus polymerase". *J. Virol.*, 81, pp. 7801-7804
- Gillet, V. et al., (1993), "SPROUT: A Program for Structure Generation", *J. Comp. Aid. Mol. Des.*, 7, pp. 127-153
- 15 Goodman, L. S. et al., Eds., (1985), "Goodman and Gilman's The Pharmacological Basis of Therapeutics", 8<sup>th</sup> Edition, Pergamon Press
- Goodsell, D. S. et al., (1990), "Automated Docking of Substrates to Proteins by Simulated
- 20 Annealing", *Proteins: Struct., Funct., Genet.*, 8, pp. 195-202
- Guida, W. C., (1994), "Software for Structure-Based Drug Design", *Curr. Opin. Struct. Biol.*, 4, pp. 777-781
- 25 Harlow, E. and Lane, D., Eds., (1990), "Antibodies: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Holm, L. and Sander, C., (1993), "Protein structure comparison by alignment of distance matrices", *J. Mol. Biol.*, 233, pp. 123-138.
- 30 Honda, A. et al., (1999), "Two separate sequences of PB2 subunit constitute the RNA cap-binding site of influenza virus RNA polymerase", *Genes to Cells*, 4, pp. 475-485

Hooker, L. et al., (2003), "Quantitative analysis of the influenza virus RNP interaction with RNA cap structures and comparison to human cap binding protein eIF4E", *Biochemistry*, 42, pp. 6234-6240

- 5 Jorgensen, W. L., (1998), "OPLS Force Fields", *Encyclopedia of Computational Chemistry*, Schleyer, Ed., Wiley, New York, Vol. 3, pp. 1986-1989

Kukkonen, S. K. et al (2005), "L protein, the RNA-dependent RNA polymerase of hantaviruses", *Arch. Virol.*, 150, pp. 533-556

10

Kuntz, I. D. et al., (1982), "A Geometric Approach to Macromolecule-Ligand Interactions", *J. Mol. Biol.*, 161, pp. 269-288

15

Lauri, G. and Bartlett, P. A., (1994), "CAVEAT: A Program to Facilitate the Design of Organic Molecules", *J. Comp. Aid. Mol. Des.*, 8, pp. 51-66

Leahy, M. B. et al, (2005), "In Vitro Polymerase Activity of Thogoto Virus: Evidence for a Unique Cap-Snatching Mechanism in a Tick-Borne Orthomyxovirus", *J. Virol.*, 71, pp. 8347-8351

20

Leuenberger, H.G.W. et al., Eds., (1995), "A Multilingual Glossary of Biotechnological Terms: IUPAC Recommendations", *Helvetica Chimica Acta*, CH-4010 Basel, Switzerland,

25

Li, M.-L. et al., (2001), "The active site of the influenza cap-dependent endonuclease are on different polymerase subunits", *EMBO J.*, 20, pp. 2078-2086

Lipinski, C. A. et al., (1997), "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings", *Adv. Drug Deliv. Rev.*, 23, pp. 3-25

30

Lovell, S.C. et al., (2003), "Structure validation by Calpha geometry: phi,psi and Cbeta deviation", *Proteins*, 50, pp. 437-450.

- Magden, J. et al., (2005), "Inhibitors of virus replication: recent developments and prospects", Appl. Microbiol. Biotechnol., 66, pp. 612-621
- 5 Martin, Y. C., (1992), "3D Database Searching in Drug Design", J. Med. Chem., 35, pp. 2145-2154
- Martin-Benito, J et al., (2001), "Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle", EMBO Rep., 2, pp. 313-317
- 10 Meng, E. C. et al., (1992), "Automated Docking with Grid-Based Energy Evaluation", J. Comp. Chem., 13, pp. 505-524
- Moon, J. B. and Howe, W. J., (1991), "Computer Design of Bioactive Molecules: A Method for Receptor-Based *De Novo* Ligand Design", Proteins, 11, pp. 314-328
- 15 Murshudov, G. N., (1997), "Refinement of macromolecular structures by the maximum-likelihood method", Acta Crystallogr. D. Biol. Crystallogr. 53, pp. 240-255.
- Natarajan, A. et al., (2004), Synthesis of fluorescein labeled 7-methylguanosinemono-phosphate. Bioorg. Med. Chem. Lett., 14, pp. 2657-2660
- 20 Navia, M. A. and Murcko, M. A., (1992), "The Use of Structural Information in Drug Design", Curr. Opin. Struct. Biol., 2, pp. 202-210
- 25 Nishibata, Y. and Itai, A., (1991), "Automatic creation of drug candidate structures based on receptor structure. Starting point for artificial lead generation", Tetrahedron, 47, pp. 8985-8990
- Noah, D. L. and Krug, R. M., (2005), "Influenza virus virulence and its molecular determinants", Adv. Virus Res., 65, pp. 121-145
- 30 Ortega, J. et al., (2000), "Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification", J. Virol., 74, pp. 156-163

- Perrakis A. et al., (1999), "Automated protein model building combined with iterative structure refinement", *Nat. Struct. Biol.*, 6, pp. 458-63
- 5 Plotch, S. J. et al., (1981), "A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription", *Cell*, 23, pp. 847-858
- Ren, J. et al., (1996), "Synthesis of a fluorescent 7-methylguanosine analog and a  
10 fluorescence spectroscopic study of its reaction with wheat germ cap binding proteins", *Nucleic Acids Res.*, 15, pp. 3629-3634
- Rotstein, S. H. and Murcko, M. A., (1993), "GroupBuild: A Fragment-Based Method for *De Novo* Drug Design", *J. Med. Chem.*, 36, pp. 1700-1710
- 15 Ryder, S. P. and Strobel, S. A., (1999), "Nucleotide analog interference mapping", *Methods*, 18, pp. 38-50
- Sali, A. and Blundell, T. L., (1993), "Comparative protein modelling by satisfaction of spatial  
20 restraints", *J. Mol. Biol.*, 234, pp. 779-815
- Sambrook, J. et al., Eds., (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- 25 Schneider T. R. and Sheldrick G. M., (2002), "Substructure solution with SHELXD", *Acta Crystallogr. D. Biol. Crystallogr.* (Pt 10 Pt 2), pp. 1772-1779
- Schulz, G. E. and Schirmer, R. H., (1985), "Principles of Protein Structure", Springer Verlag, New York
- 30 Tarendeau, F. et al., (2007), "Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit", *Nat. Struct. Mol. Biol.*, 14, pp. 229-233

Terwilliger, T. C., (2000), "Maximum likelihood density modification," *Acta Cryst. D. Biol. Crystallogr.*, 56, pp. 965-972

5 Tisdale, M. et al., (1995), "Inhibition of influenza virus transcription by 2'-deoxy-2'-fluoroguanosine", *Antimicrob. Agents Chemother.*, 39, pp. 2454-2458

10 Tomassini, J. et al., (1994), "Inhibition of cap (m<sup>7</sup>GpppXm)-dependent endonuclease of influenza virus by 4-substituted 2,4-diocobutanic acid compounds", *Antimicrob. Agents Chemother.*, 38, pp. 2827-2837

Tomassini, J. et al., (1996), "A novel antiviral agent which inhibits the endonuclease of influenza viruses", *Antimicrob. Agents Chemother.*, 40, pp. 1189-1193

15 Van Duyne, G. D. et al., (1993), "Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin", *J. Mol. Biol.*, 229, pp. 105-124

Von Itzstein, M. et al., (1993), "Rational design of potent sialidase-based inhibitors of influenza virus replication", *Nature*, 363, pp. 418-423

20 Vonnrhein C. et al., (2006), "Automated Structure Solution With autoSHARP", *Methods Mol. Biol.*, 364, pp. 215-30

25 Wallace, A. C. et al., (1995). "LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions", *Prot. Eng.*, 8, pp. 127-134.

## Claims

1. A soluble polypeptide fragment, wherein said polypeptide fragment is (i) derived from the Influenza virus RNA-dependent RNA polymerase subunit PB2 or variant thereof and (ii)  
5 capable of binding to a RNA cap or analog thereof.
2. The polypeptide fragment of claim 1, wherein said polypeptide fragment is purified to an extent to be suitable for crystallization.
- 10 3. The polypeptide fragment of claims 1 or 2, wherein the Influenza virus RNA-dependent RNA polymerase subunit PB2 is from an Influenza A, B, or C virus or variant thereof.
4. The polypeptide fragment of any of claims 1 to 3, wherein  
15 (i) the N-terminus is identical to or corresponds to amino acid position 220 or higher and the C-terminus is identical to or corresponds to amino acid position 510 or lower of the amino acid sequence of PB2 according to SEQ ID NO: 1,  
(ii) the N-terminus is identical to or corresponds to amino acid position 222 or higher and the C-terminus is identical to or corresponds to amino acid position 511 or  
20 lower of the amino acid sequence of PB2 according to SEQ ID NO: 2, or  
(iii) wherein the N-terminus is identical to or corresponds to amino acid position 227 or higher and the C-terminus is identical to or corresponds to amino acid position 528 or lower of the amino acid sequence of PB2 according to SEQ ID NO: 3  
and variants thereof, which retain the ability to associate with an RNA cap or analog thereof.  
25
5. The polypeptide fragment of claim 4, wherein said polypeptide fragment has or corresponds to an amino acid sequence selected from the group of amino acid sequences according to SEQ ID NO: 4 to 13 and variants thereof, which retain the ability to associate with an RNA cap or analog thereof.  
30
6. A complex comprising the polypeptide fragment of any of claims 1 to 5 and a RNA cap or analog thereof.

7. The complex of claim 6, wherein the cap analog is selected from the group consisting of m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, and m<sup>7</sup>GpppUm.
- 5 8. The complex of claim 6, wherein said polypeptide fragment consists of an amino acid sequence according to SEQ ID NO: 11 and said cap analog is m<sup>7</sup>GTP, having the structure defined by the structure coordinates as shown in Figure 18.
9. The complex of claim 8, wherein said complex has a crystalline form with space group  
10 C222<sub>1</sub> and unit cell dimensions of a = 9.2 nm, b = 9.4 nm; c = 22.0 nm (± 0.3 nm)
10. The complex of claims 8 and 9, wherein the crystal diffracts X-rays to a resolution of 3.0 Å or higher, preferably 2.4 Å or higher.
- 15 11. An isolated polynucleotide coding for an isolated polypeptide of any of claims 1 to 5.
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  
20 recombinant vector of claim 12.
14. A method for identifying compounds which associate with all or part of the RNA cap binding pocket of PB2 or the binding pocket of a PB2 polypeptide variant, comprising the steps of
- 25 (a) constructing a computer model of said binding pocket defined by the structure coordinates of the complex of claim 8 as shown in Figure 18;
- (b) selecting a potential binding 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  
30 (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 binding pocket in order to provide an energy-minimized configuration of the said compound in the binding pocket; and

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

- 5 15. The method of claim 14, wherein said binding pocket comprises amino acids Phe323, His357, and Phe404 of PB2 according to SEQ ID NO: 1 or amino acids corresponding thereto.
- 10 16. The method of claim 15, wherein said binding pocket further comprises amino acids Phe325, Phe330, and Phe363 of PB2 according to SEQ ID NO: 1 or amino acids corresponding thereto.
- 15 17. The method of claims 15 or 16, wherein said binding pocket further comprises amino acids Glu361, and Lys376 of PB2 according to SEQ ID NO: 1 or amino acids corresponding thereto.
18. The method of any of claims 15 to 17, wherein said binding pocket further comprises amino acids Ser320, Arg332, Ser337, and Gln406
- 20 19. The method of any of claims 15 to 18, wherein said binding pocket further comprises amino acids Lys339, Arg355, Asn429, and His432 of PB2 according to SEQ ID NO: 1 or amino acids corresponding thereto.
- 25 20. The method of claim 14, wherein said binding pocket is defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe323, His357, and Phe404 according to Figure 18.
- 30 21. The method of claim 20, wherein said binding pocket is further defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Phe325, Phe330, and Phe363 according to Figure 18.
22. The method of claims 20 or 21, wherein said binding pocket is further defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Glu361, and Lys376 according to Figure 18.

23. The method of any of claims 20 or 22, wherein said binding pocket is further defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Ser320, Arg332, Ser337, and Gln406 according to Figure 18.

5

24. The method of any of claims 20 to 23, wherein said binding pocket is further defined by the structure coordinates of PB2 SEQ ID NO: 1 amino acids Lys339, Arg355, Asn429, and His432 according to Figure 18.

10 25. The method of claim 20, wherein the binding pocket of a PB2 polypeptide variant has a root mean square deviation from the backbone atoms of the amino acids Phe323, His357, and Phe404 of said binding pocket of not more than 2.5 Å.

26. The method of any of claims 14 to 25 comprising the further step of

15 (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).

27. The method of claim 26 comprising the further step of

20 (f) contacting said compound, said polypeptide fragment of any of claims 1 to 4 or said recombinant host cell according to claim 12 and a RNA cap or analog thereof to determine the ability of said compound to inhibit binding between said PB2 polypeptide fragment and said RNA cap or analog thereof.

25 28. A compound identifiable by the method of claims 14 to 27, under the provision that the compound is not  $m^7G$ ,  $m^7GMP$ ,  $m^7GTP$ ,  $m^7GpppG$ ,  $m^7GpppGm$ ,  $m^7GpppA$ ,  $m^7GpppAm$ ,  $m^7GpppC$ ,  $m^7GpppCm$ ,  $m^7GpppU$ ,  $m^7GpppUm$ , 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one, T-705 or  $m^7Gppp(N)_{1-15}$ , wherein N is A, Am, G, Gm, C, Cm, U or Um and is able to bind to the RNA cap binding pocket of PB2 or variant thereof.

30

29. A compound identifiable by the method of claims 14 to 27, under the provision that the compound is not  $m^7G$ ,  $m^7GMP$ ,  $m^7GTP$ ,  $m^7GpppG$ ,  $m^7GpppGm$ ,  $m^7GpppA$ ,  $m^7GpppAm$ ,  $m^7GpppC$ ,  $m^7GpppCm$ ,  $m^7GpppU$ ,  $m^7GpppUm$ , 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one, T-705 or  $m^7Gppp(N)_{1-15}$ , wherein N is A, Am, G, Gm, C, Cm, U or Um

and is able to inhibit binding between the PB2 polypeptide, variant thereof or fragment thereof and the RNA cap or analog thereof.

30. A method for identifying compounds which associate with the RNA cap binding  
5 pocket of PB2 or binding pockets of PB2 polypeptide variants, comprising the steps of (i)  
contacting said polypeptide fragment of claims 1 to 4 or said recombinant host cell according  
to claim 13 with a test compound and (ii) analyzing the ability of said test compound to bind  
to PB2.
- 10 31. The method of claim 30, comprising the further step of adding a RNA cap or analog  
thereof.
32. The method of claim 31, wherein the ability of said test compound to bind to PB2 or a  
variant thereof in presence of said RNA cap or analog thereof or the ability of said test  
15 compound to inhibit binding of said RNA cap or analog thereof to PB2 or a variant thereof is  
analyzed.
33. The method of claims 31 or 32, wherein said RNA cap or analog thereof is added  
prior, concomitantly, or after addition of said test compound.
- 20 34. The method of any of claims 30 to 33 performed in a high-throughput setting.
35. The method of any of claims 30 to 34, wherein said test compound is a small molecule.
- 25 36. The method of any of claims 30 to 34, wherein said test compound is a peptide or  
protein.
37. The method of any of claims 26, 27, or 30 to 36, wherein said method further  
comprises the step of formulating said compound or a pharmaceutically acceptable salt  
30 thereof with one or more pharmaceutically acceptable excipient(s) and/or carrier(s).
38. A pharmaceutical composition producible according to the method of claims 27 or 37.

39. A compound identifiable by the method of any of claims 30 to 37, under the provision that the compound is not m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, m<sup>7</sup>GpppUm, 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one, T-705 or m<sup>7</sup>Gppp(N)<sub>1-15</sub>, wherein N is A, Am, G, Gm, C, Cm, U or Um and is able to bind to the PB2 polypeptide, variant thereof or fragment thereof.
40. A compound identifiable by the method of any of claims 29 to 36, under the provision that the compound is not m<sup>7</sup>G, m<sup>7</sup>GMP, m<sup>7</sup>GTP, m<sup>7</sup>GpppG, m<sup>7</sup>GpppGm, m<sup>7</sup>GpppA, m<sup>7</sup>GpppAm, m<sup>7</sup>GpppC, m<sup>7</sup>GpppCm, m<sup>7</sup>GpppU, m<sup>7</sup>GpppUm, 2-Amino-7-benzyl-9-(4-hydroxy-butyl)-1,9-dihydro-purin-6-one, T-705 or m<sup>7</sup>Gppp(N)<sub>1-15</sub>, wherein N is A, Am, G, Gm, C, Cm, U or Um, and is able to inhibit binding between the PB2 polypeptide, variant thereof or fragment thereof and the RNA cap or analog thereof.
41. An antibody directed against the RNA cap binding domain of PB2.
42. The antibody of claim 41, wherein said antibody recognizes a polypeptide fragment selected from the group of polypeptides defined by SEQ ID NO: 14 to 22.
43. Use of a compound according to claims 28, 29, 39, or 40, a pharmaceutical composition according to claim 36, or an antibody according to claims 41 or 42 for the manufacture of a medicament for treating, ameliorating, or preventing disease conditions caused by viral infections with negative-sense ssRNA viruses.
44. The use of claim 43, wherein said disease condition is caused by viral infections of the Mononegavirales order comprising the Bornaviridae, Filoviridae, Paramyxoviridae, and Rhabdoviridae families.
45. The use of claim 43, wherein said disease condition is caused by the Orthomyxoviridae, Arenaviridae, or Bunyaviridae families.
46. The use of claim 43, wherein said disease condition is caused by a virus selected from the group consisting of Borna disease virus, Marburg virus, Ebola virus, Sendai virus, Mumps virus, Measles virus, Human respiratory syncytial virus, Turkey rhinotracheitis virus,

Vesicular stomatitis Indiana virus, Nipah virus, Henda virus, Rabies virus, Bovine ephemeral fever virus, Infectious hematopoietic necrosis virus, Thogoto virus, Influenza A virus, Influenza B virus, Influenza C virus, Hantaan virus, Crimean-congo hemorrhagic fever virus, Rift Valley fever virus, and La Crosse virus.

5

47. The use of claim 43, wherein said disease condition is caused by a virus selected from the group consisting of Influenza A virus, Influenza B virus, Influenza C virus, Thogoto virus, and Hantaan virus.