



(51) International Patent Classification:

C40B 30/02 (2006.01) *C07K 14/145* (2006.01)
C40B 30/04 (2006.01) *G01N 23/00* (2006.01)
C12Q 1/70 (2006.01) *G01N 1/42* (2006.01)
C07K 14/08 (2006.01) *G01N 33/50* (2006.01)

(21) International Application Number:

PCT/US2015/065388

(22) International Filing Date:

11 December 2015 (11.12.2015)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/091,261 12 December 2014 (12.12.2014) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy St., Cambridge, Massachusetts 02138 (US).

(72) Inventors; and

(71) Applicants : **LIANG, Bo** [CN/US]; Apt. 2, 95 Boylston St., Brookline, Massachusetts 02445 (US). **LI, Zongli** [CN/US]; 90 Gary Road, Needham, Massachusetts 02494 (US). **JENNI, Simon** [CH/US]; Apt. 22F, Emerson Place, Boston, Massachusetts 02114 (US). **GRIGORIEFF, Nikolaus** [US/US]; 19070 Coton Farm Court, Leesburg, Virginia 20176 (US). **HARRISON, Stephen** [US/US]; 19 Sparhawk St., Brighton, Massachusetts 02135 (US).

(72) Inventor: **WHELAN, Sean P.J.**; 236 Broadway, Cambridge, Massachusetts 02139 (US).

(74) Agents: **RESNICK, David S.** et al.; Nixon Peabody LLP, 100 Summer St., Boston, Massachusetts 02110-2131 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

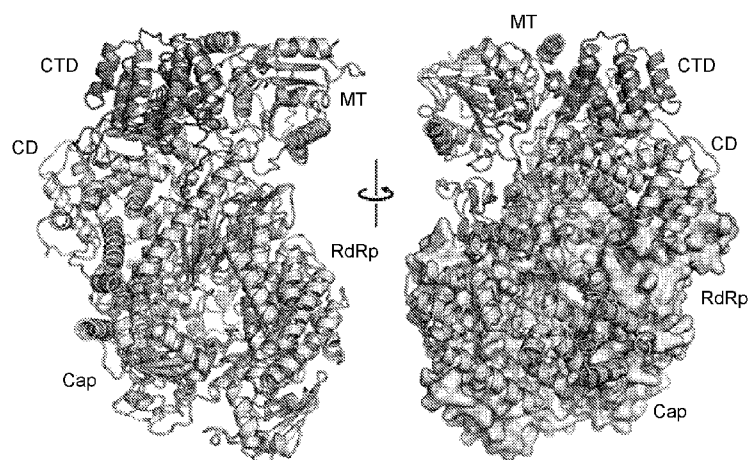
(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

[Continued on next page]

(54) Title: METHODS AND COMPOSITIONS RELATING TO HIGH RESOLUTION STRUCTURE OF NON-SEGMENTED NEGATIVE STRAND VIRUS L PROTEINS

**FIG. 3B**

(57) Abstract: Described herein are methods and compositions relating to the high resolution 3D structure determination for non-segmented negative strand (NNS) virus multifunction L proteins. High resolution 3D structure as described permits the application of a range of rational drug design and screening approaches to the identification of inhibitors of any of the varying functions of the NNS virus L proteins and thereby facilitates the identification of antiviral agents.



-
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) — before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

Published:

- with international search report (Art. 21(3))

METHODS AND COMPOSITIONS RELATING TO HIGH RESOLUTION STRUCTURE OF NON-SEGMENTED NEGATIVE STRAND VIRUS L PROTEINS

CROSS-REFERENCE TO RELATED APPLICATION

This application claims benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 62/091,261, filed December 12, 2014, the content of which is incorporated herein by reference in its entirety.

GOVERNMENT SUPPORT

This invention was made with government support under R01-AI059371 awarded by the National Institutes of Health (NIH). The government has certain rights in the invention.

FIELD OF THE INVENTION

The technology described herein relates to the identification of anti-viral therapeutics.

BACKGROUND

The non-segmented negative-strand (NNS) RNA viruses include some of the most lethal human and animal pathogens, including Ebola virus and rabies virus. Human respiratory syncytial virus infects 95% of infants before the age of 2 and can result in severe respiratory disease requiring hospitalization. According to the Center for disease Control, seasonal RSV infection accounts for 172,000 hospitalizations and 1.5 million outpatient visits among children younger than 5 yrs old, and 14,000 deaths among adults older than 65 years. Current therapeutics for RSV infection are limited in efficacy and high in cost. Therapeutics such as ribavirin have been used in the clinic, however evidence that it inhibits RS viral replication has been lacking.

For many members of the NNS family viruses there are no vaccines or efficacious antiviral drugs. Furthermore, rapid development of drug resistance to monotherapy has been observed for other RNA viruses, such as influenza (van der Vries et al., 2010; Zhu et al., 2012), suggesting that multiple antivirals will be required for long-term effective treatment of these diseases. Therefore, the development of new therapies is warranted, especially ones that could target multiple members of this human-pathogen laden virus order.

The multifunctional, large (L) polymerase proteins, carried within the virions (Baltimore et al., 1970), are produced by all NNS family members and have biochemical properties that distinguish them from most other RNA polymerases of viruses and polymerases of their hosts, which might make them

suitable as antiviral targets. In addition to their RNA-dependent RNA polymerase (RdRp) activity (Emerson and Wagner, 1973), NNS RNA virus L proteins catalyze an unusual sequence of mRNA capping reactions (Hercyk et al., 1988), and the RdRp itself polyadenylates the viral message (Hunt et al., 1984).

Relatively untargeted screens have identified small molecules that appear to interact with and/or inhibit certain NNS virus L proteins. Boehringer Ingelheim has reported the identification of a small molecule that appears to target L protein of RSV. An inhibitor of the polymerase from the measles virus identified by mapping a resistance mutation to the RdRp domain of L, has no activity against the polymerase of NNS family member RSV. An adenosine analog was reported that inhibits Ebola virus in a cell based assay, but interfered with replication of many viruses, raising the question of its impact on the host cell.

NNS virus L proteins have been difficult to study because of their size, low expression and dearth of robust biochemical assays of their functions.

SUMMARY OF THE INVENTION

The technology described herein relates to the identification of high resolution 3D structure for the L proteins of NNS viruses and the use of such structures to identify agents that inhibit one or more activities of the L protein. Molecules or compounds that inhibit L protein activity can provide antiviral therapies by disrupting viral replication, which depends upon the activities of the L protein. Where the multifunctional NNS virus L protein is essentially unique to the NNS family of viruses, therapies that selectively target L protein activity have a good likelihood of inhibiting viral replication with a reduced likelihood of unwanted effects on the host cell or organism. 3D structures can permit the rational design and/or refinement of candidate inhibitor molecules. However, to date, the size and dynamic structure of the NNS virus L proteins has precluded the application of standard approaches to determine their 3D structure. The technology described herein stems, in part, from the solution of an NNS virus L protein at high resolution, which permits the application of computer-modeling strategies to design or otherwise identify candidate inhibitor molecules. When used in combination with assays described herein for the various activities of the NNS virus, the high resolution structure also permits the confirmation and refinement of inhibitory molecules as antiviral agents.

The NNS virus L protein exemplified herein at high resolution is that of vesicular stomatitis virus (VSV). The approaches described herein to obtaining that high resolution structure are readily applied to the L proteins of other NNS viruses. However, not only does the solution of the high resolution structure for the VSV L protein described herein demonstrate the feasibility of generating such structures for any of the NNS viruses by, e.g., cryoelectron microscopy, but because the

structures of NNS virus L proteins are well-conserved between members of the NNS virus family, the solution of the VSV protein's structure also permits the use of that structure to predict, via computer modeling using amino acid sequence information for other NNS virus L proteins, the high resolution structures of those L proteins as well.

In one aspect, described herein is a method of identifying an inhibitor of a non-segmented negative strand (NNS) virus, the method comprising: (a) using a high resolution 3D structure of an NNS virus L protein to select a candidate inhibitor from a virtual library of 3D structures of potential candidates by performing structure based computational modeling; (b) obtaining the candidate inhibitor; and (c) measuring an activity of L protein when contacted with the candidate inhibitor obtained in step (b), wherein a decrease in the activity confirms that the candidate is an inhibitor of the NNS virus.

In another aspect, described herein is a method of identifying an inhibitor of a non-segmented negative strand (NNS) virus polymerase, the method comprising: (a) using a high resolution 3D structure of an NNS virus L protein to select a candidate inhibitor from a virtual library of 3D structures of potential candidates by performing structure based computational modeling; (b) obtaining the candidate inhibitor; and (c) measuring L protein polymerase activity when contacted with the candidate inhibitor obtained in step (b), wherein a decrease in the activity of the polymerase confirms that the candidate is an inhibitor of the NNS virus polymerase.

In another aspect, described herein is a method of identifying an inhibitor of an NNS virus, the method comprising: (a) screening a library of candidate molecules for inhibition of an NNS L protein in an *in vitro* biochemical assay for an L protein activity; (b) using a high resolution 3D structure of the L protein to determine the binding site of potential candidates obtained in step (a) by computer modeling; (c) identifying a potential refinement to the candidate inhibitor; (d) obtaining a candidate inhibitor molecule comprising the refinement identified in step (c); and (e) measuring the activity of the L protein when contacted with the candidate inhibitor obtained in step (d), wherein a decrease in the activity of the L protein confirms that the candidate is an inhibitor of the NNS virus.

In another aspect, described herein is a method of identifying an inhibitor of L protein of an NNS virus, the method comprising: (a) using a high resolution 3D structure of the L protein of vesicular stomatitis virus (VSV) to create an atomic model of the L protein of a second NNS virus by homology modeling the amino acid sequence of the L protein of the second NNS virus onto the high resolution 3D structure of the VSV L protein; (b) using the atomic model of the second NNS virus L protein to select a candidate inhibitor from a virtual library of 3D structures of potential candidates by performing structure based computational modeling; (c) obtaining the candidate inhibitor; and (d) measuring an activity of L protein of the second NNS virus when contacted with the candidate

inhibitor obtained in step (b), wherein a decrease in the activity confirms that the candidate is an inhibitor of the second NNS virus.

In one embodiment of each of the aspects noted above, the high resolution 3D structure has a resolution of at least 3.8 Å. In another embodiment, the high resolution 3D structure is an atomic structure determined via cryo-electron microscopy and single particle analysis.

In another embodiment of each of the aspects noted above, the NNS virus L protein is vesicular stomatitis virus L protein. In another embodiment, the VSV L protein high resolution structure has space group P1, so as to form a unit cell of dimensions of $a=112.0$ Å, $b=143.0$ Å, $c=106.0$ Å and $\alpha, \beta, \gamma=90^\circ$.

In one embodiment, the VSV L protein high resolution 3D structure has the coordinates as set out, for example, in the structure deposited with the RCSB Protein Data Bank with Reference Number 5A22 (this structure is accessible on the World Wide Web at rcsb.org/pdb/explore.do?structureId=5A22).

In another embodiment of each of the aspects noted above, the activity of the L protein measured is selected from the group consisting of RNA dependent RNA polymerase activity, GDP polyribonucleotidyltransferase (PRNTase) activity, methyltransferase activity, and the ability to structurally rearrange L protein domains during L protein functions. When the L protein activity measured is RNA dependent RNA polymerase activity, that activity can be further selected from the group consisting of transcript initiation, elongation and termination.

In another embodiment of certain aspects noted above, rather than selecting a candidate inhibitor from a virtual library of 3D structures of potential candidates, the high resolution 3D structure of the NNS virus L protein is used to design a potential inhibitor. In one embodiment, the potential inhibitor is designed to interact with an L protein domain selected from the group consisting of the RNA dependent RNA polymerase (RdRp) domain, the mRNA capping domain, the methyltransferase domain, the loop primer domain, the connector domain, and the C-terminal domain.

In another embodiment of each of the aspects noted above, the identified inhibitor binds a domain of the L protein at a region selected from the group consisting of the RNA dependent RNA polymerase (RdRp) domain, the mRNA capping domain, the methyltransferase domain, the loop primer domain, the connector domain, and the C-terminal domain of the L protein.

In another embodiment of each of the aspects noted above, the identified inhibitor attenuates the RNA dependent RNA polymerase activity, the GDP polyribonucleotidyltransferase (PRNTase) activity, the methyltransferase activity, the loop priming activity, or the ability of the L protein to structurally rearrange its domains during L one or more of its functions.

In another embodiment of each of the aspects noted above, the method further comprises the step of contacting the confirmed candidate inhibitor with a cell infected with an NNS virus.

In another embodiment of each of the aspects noted above, the NNS virus is selected from the viruses of the Filoviridae, Paramyxoviridae, Rhabdoviridae and Bornaviridae families. In one embodiment, the NNS virus is a Filovirus, selected from Ebola virus, Marburg virus and Cuevavirus. In one embodiment, the NNS virus is a Paramyxovirus selected from measles virus, Nipah virus, Hendra virus, respiratory syncytial virus and Newcastle disease virus. In one embodiment, the NNS virus is a Rhabdovirus selected from Lyssavirus, Vesiculovirus, Perhabdovirus, Sigmavirus, Ephemerovirus, Tibrovirus, Tupavirus and Sprivivirus.

In another aspect, described herein is a method of mapping the location of resistance to a known inhibitor of an NNS virus L protein on the 3D structure of the L protein, the method comprising homology-based structural mapping of the sequence comprising the resistance mutation onto the high resolution 3D structure of an NNS virus L protein. The high resolution 3D structure can be, for example, a 3.8 Å or greater resolution atomic structure of the chosen L protein.

In another aspect, described herein is a method of inhibiting an NNS virus, the method comprising contacting an NNS virus or a cell infected with an NNS virus with an inhibitor identified according to the method of any one of the aspects described herein above and/or in the Examples. The inhibitor can be administered to an individual in need of treatment or prophylaxis for infection with the NNS virus.

In another aspect, described herein is an NNS virus inhibitor identified using the method of any one of the aspects described herein above and/or in the Examples.

In another aspect, described herein is a method of designing a candidate inhibitor of an NNS virus, the method comprising: (a) identifying an NNS virus mutant that escapes the activity of a known inhibitor of the L protein of an that virus; (b) identifying the site of the L protein resistance mutation by sequence analysis; (c) using homology-based structural mapping of the sequence comprising the resistance mutation to map the mutation site onto a high resolution 3D structure of an NNS virus L protein; (d) using computer modeling with the high resolution 3D structure to selecting a candidate binding molecule that targets the site of the resistance mutation mapped in step (c); and (e) obtaining the candidate inhibitor molecule identified in step (d). In one embodiment, the method further comprises the step of contacting L protein of the NNS virus with the candidate inhibitor in an *in vitro* assay of L protein activity, wherein a decrease in L protein activity indicates the candidate inhibitor is an NNS virus inhibitor.

In another embodiment, the method further comprises the step of contacting the candidate inhibitor with the NNS virus and assaying viral function, wherein a decrease in viral function indicates the candidate inhibitor is an NNS virus inhibitor.

In another aspect, described herein is a method of designing a candidate inhibitor of an NNS virus, the method comprising: (a) identifying an NNS virus mutant that escapes the activity of a known inhibitor of the L protein of an that virus; (b) identifying the site of the L protein resistance mutation by sequence analysis; (c) using homology-based structural mapping of the sequence comprising the resistance mutation to map the mutation site onto a high resolution 3D structure of an NNS virus L protein; (d) identifying a potential refinement to the known inhibitor that permits interaction with the mutant L protein; (e) obtaining the inhibitor molecule comprising the refinement identified in step (d); and (f) assaying mutant virus replication or infection in the presence and absence of the refined inhibitor molecule, wherein a decrease in replication or infection indicates the refined inhibitor is effective against the mutant virus.

In another aspect, described herein is the use of an NNS virus inhibitor identified according to the methods of any one of the aspects or embodiments discussed above and/or described in the Examples, for the treatment of an individual infected with an NNS virus. In such use, the NNS virus can be selected, for example, from the group consisting of viruses of the Filoviridae, Paramyxoviridae, Rhabdoviridae and Bornaviridae families.

In another aspect, described herein is a computer-readable medium with high resolution 3D structure coordinate information for an NNS L protein stored thereupon.

DEFINITIONS:

As used herein, the term "non-segmented negative-strand (NNS) virus" refers to a virus that belongs to superfamily Mononegavirales and includes the families Rhabdoviridae (including, but not limited to vesicular stomatitis virus (VSV) and rabies virus), Paramyxoviridae (including, but not limited to measles and respiratory syncytial viruses (RSV), Filoviridae (including, but not limited to Ebola and Marburg viruses) and Bornaviridae (including, but not limited to Borna disease virus).

As used herein, the term "L protein of an NNS virus" refers to the large multifunctional proteins of non-segmented, negative strand RNA viruses that catalyzes RNA-dependent RNA polymerization with viral ribonucleoprotein as template, a non-canonical sequence of capping and methylation reactions, and polyadenylation of viral messages.

As used herein, the term "L protein activity" refers to any of the biological activities of an NNS virus L protein involved in and/or necessary for viral replication. Activities can include any of, for example,

RNA dependent RNA polymerase activity (including initiation, elongation and termination), GDP phosphoribonucleotidyltransferase activity, methyltransferase activity, viral mRNA capping activity, and the ability of the L protein to rearrange one or more of its domains during L protein function necessary for viral replication.

As used herein, the term “domain,” when used in reference to an NNS virus L protein or to a portion thereof, refers to a portion of an NNS virus L protein that performs a particular function, e.g., an enzymatic function including RNA-dependent RNA polymerase [RdRp], polyribonucleotidyl transferase [PRNTase], and methyltransferase functions and /or that defines a structural feature, e.g., a connector domain and/or C-terminal domain. The domains of an NNS virus L protein are shared across members of the family, such that alignment of, e.g., the VSV L protein with that of other family members permits the identification of the corresponding domains in the other family members. Sequence alignment in conjunction with a high resolution 3D structure permit the application of the 3D structure of one NNS virus L protein to make predictions regarding activities and inhibitors of other NNS virus L proteins.

As used herein, the term “RNA-dependent RNA polymerase domain” refers to that portion of an L protein that catalyzes the protein’s RNA polymerase activity. In the 2109 amino acid VSV L protein, the domain corresponds to amino acid residues 35-865 and encompasses a right hand, “fingers-palm-thumb” structure, with the catalytic site on the palm in a deep channel between the fingers and thumb subdomains.

As used herein, the term “capping domain” refers to that portion of the L protein that catalyzes viral mRNA capping and encompasses the GDP polyribonucleotidyl transferase (PRNTase) catalytic site. In the 2109 amino acid VSV L protein, the domain corresponds to amino acid residues 866-1334 and encompasses a GDP polyribonucleotidyl transferase (PRNTase) catalytic site of two conserved motifs—GxxT and HR, separated by ~ 70 residues, the largely α -helical, N-terminal half (residues 866–1100), which abuts the polymerase domain and the C-terminal half (1100–1334).

As used herein, the term “methyltransferase domain” refers to that portion of the L protein responsible for methyltransferase activity. In the 2109 amino acid VSV L protein, the domain corresponds to amino acid residues 1598-1892, comprises the methyltransferase catalytic site, and contacts both the connector and capping domains.

As used herein, the term “C-terminal domain” refers to that portion of the L protein occurring C-terminal to the methyltransferase domain. In the 2109 amino acid VSV L protein, the C-terminal domain corresponds to amino acid residues 1893-2109. The C-terminal domain is structurally an α -helical bundle, projecting a beak-like, β -hairpin supported by a second inter-helical loop. It extends

back against the RdRp, augmenting the β -hairpin that bears the catalytic Asp-Asn sequence at its tip, and terminates at the three-way junction of the capping, connector, and methyltransferase domains, where it has one or more contacts with each. The arm thus contributes to closing the multi-domain structure in the complex formed between the NNS viral L and P proteins.

As used herein, the term “connector domain” refers to that portion of the L protein occurring between the capping and methyltransferase domains. In the 2109 amino acid VSV L protein, the connector domain corresponds to amino acid residues 1358-1557, and is a bundle of eight helices.

As used herein, the term “high resolution 3D structure” refers to an atomic level protein or polypeptide structure of at least 3.8Å resolution, represented by a set of structure coordinates that can generate a 3D representation of the protein structure when applied to appropriate software. In some embodiments, the high resolution 3D structure has a resolution of at least 3.8Å, e.g., a resolution of at least 3.7Å, at least 3.6Å, at least 3.5Å, at least 3.4Å, at least 3.3Å, at least 3.2Å, at least 3.1Å, at least 3.0Å, at least 2.9Å, at least 2.8Å, at least 2.7Å, at least 2.6Å, at least 2.5Å, at least 2.4Å, at least 2.3Å, at least 2.2Å, at least 2.1Å, at least 2.0Å, or at least 1.9Å.

As used herein, the term “X Å atomic structure” refers to a three dimensional structure or set of structure coordinates that describes the 3D structure of a protein at a resolution of at least XÅ. The term “XÅ atomic structure of an NNS virus L protein” refers to the structure or set of coordinates that describes the structure of an entire NNS virus L protein including all of its domains, catalytic regions and structural elements at at least XÅ resolution. Thus, a 3.8 Å atomic structure of an NNS virus L protein is a three dimensional structure or set of structure coordinates that describes the 3D structure of an NNS L protein at a resolution of at least 3.8Å. High resolution atomic structures such as these can be obtained by methods including but not limited to cryo-electron microscopy, single particle analysis and x-ray crystallography. As an example, a 3.8 Å atomic structure of VSV L protein, solved by cryo-electron microscopy, exhibits the resolution of at least 3.8 Å with a space group P1, so as to form a unit cell of dimensions of a=112.0 Å, b=143.0 Å, c=106.0 Å and $\alpha, \beta, \gamma=90^\circ$.

As used herein, the term “atomic model” refers to a structural model of a protein at atomic resolution generated for example, by tracing the polypeptide chain or amino acid sequence of the target protein onto the three dimensional structure of target protein or that of another related homologous protein (e.g., an NNS virus L protein from another NNS viral species).

As used herein, the term “cryo-electron microscopy” or “cryo-EM” refers to a method of imaging specimens in a transmission electron microscope under cryogenic conditions. Cryo-EM can permit high resolution structure determination of proteins for which crystal preparation is difficult or impossible.

As used herein, the term “single-particle analysis” refers to a method of computerized analysis of images from negative staining and transmission electron microscopy by combining several digitized images of similar particles to build up a three-dimensional reconstruction of the particle.

As used herein, the term “virtual library of 3D structures” refers to a collection of data representing the 3 dimensional structures of potential candidate molecules, e.g., small organic (or, for that matter, inorganic) molecules, that, in concert with appropriate software permit the *in silico* docking of members of the library with an atomic structure model of, e.g., a target protein. A non-limiting example of such a library or database includes the ZINC database; see, e.g., wiki.bkslab.org/index.php/Category:ZINC15.

As used herein, the term “refinement” when used in relation to a molecule or compound refers to a modification of the chemical structure a given molecule or compound, e.g., a modification of a known or predicted inhibitor of a biological function. A “refinement” of a given molecule or compound can, for example, optimize the binding energy of the molecule comprising the refinement with a target site and/or reduce undesirable known or likely binding to another factor.

As used herein, the term “inhibitor” refers to a molecule or agent that attenuates, inhibits, decreases, reduces, or interferes with the structure, function and/or activity of an NNS virus L protein *in vitro*, *in situ*, and/or *in vivo*. Inhibitors include but are not limited to small peptide or peptide-like molecules, soluble peptides, and synthetic non-peptidyl organic or inorganic compounds.

As used herein, the term “structure-based computational modeling” refers to the use of the three-dimensional structure or structure coordinates of a drug target, e.g., a polypeptide, including, for example a large viral polypeptide, such as an NNS virus L protein, in conjunction with computer modeling to identify and investigate candidate drugs and their binding geometries and affinities. Structure-based computational modeling uses, for example, 3D structure coordinates obtained using methods including but not limited to cryo-electron microscopy, X-ray crystallography or NMR spectroscopy. Methods and software packages that permit structure-based computational modeling are described elsewhere herein and in the Examples.

As used herein, the term “homology modeling” refers to constructing an atomic-resolution model of a "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template"). The method relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. Methods and software packages that permit homology modeling are described elsewhere herein and in the Examples.

As used herein, “RNA dependent RNA polymerase (RdRp) activity” refers to the enzymatically catalyzed production of RNA from an RNA template. Assays for RdRp activity of an NNS virus L protein are described herein and demonstrated in the Examples.

As used herein, “GDP polyribonucleotidyl transferase (PRNTase) activity” refers to the transfer of a nascent RNA onto a GDP acceptor to form a GpppN cap structure, through a covalent L protein – P protein RNA intermediate. Assays for PRNTase activity are described herein and demonstrated in the Examples.

As used herein, “methyltransferase activity” refers to formation of the 7mGpppNpmNp cap structure upon methylation of the ribose O'2 and the guanosine N7. Assays for methyltransferase activity are described herein and demonstrated in the Examples.

The terms “attenuate”, “decrease”, “reduced”, “reduction”, or “inhibit” and grammatical variations thereof are all used herein to mean a decrease by a statistically significant amount. In some embodiments, “attenuate” “reduce,” “reduction” or “decrease” or “inhibit” typically mean a decrease by at least 10% as compared to a reference level (e.g., the absence of a given treatment, condition or agent) and can include, for example, a decrease by at least about 10%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99% , or more. As used herein, “reduction” or “inhibition” does not encompass a complete inhibition or reduction as compared to a reference level. “Complete inhibition” is a 100% inhibition as compared to a reference level.

The terms “increased”, “increase” or “enhance” or “activate” and grammatical variations thereof are all used herein to generally mean an increase by a statically significant amount; for the avoidance of doubt, the terms “increased”, “increase” or “enhance” or “activate” mean an increase of at least 10% as compared to a reference level, for example an increase of at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90% or up to and including a 100% increase or any increase between 10-100% as compared to a reference level, or at least about a 2-fold, or at least about a 3-fold, or at least about a 4-fold, or at least about a 5-fold or at least about a 10-fold increase, at least about a 20-fold increase, at least about a 50-fold increase, at least about a 100-fold increase, at least about a 1000-fold increase or more as compared to a reference level.

The term "location of resistance" refers to the structural site of the residues changed within a mutant L protein polypeptide chain that results in the virus being resistant to an inhibitor. The “location of

resistance” can be mapped onto the 3D structure of the wild-type protein to elucidate the particular activity the mutation changes and thereby permit the design of an inhibitor targeting that activity.

As used in this specification and the appended claims, the singular forms "a," "an," and "the" include singular and plural references unless the context clearly dictates otherwise. Thus for example, references to "the method" includes one or more methods, and/or steps of the type described herein and/or which will become apparent to those persons skilled in the art upon reading this disclosure and so forth. For example, the term "an NNS virus L protein inhibitor" includes reference to one or a plurality (*e.g.*, two or more) NNS virus L protein inhibitor(s) and the term " NNS virus L protein inhibitor " includes reference to one or a plurality (*e.g.*, two or more) of NNS virus L protein inhibitor(s) and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

Computer-readable storage media or computer readable media can be any available tangible media (*e.g.*, tangible storage media) that can be accessed by a computer, is typically of a non-transitory nature, and can include both volatile and nonvolatile media, removable and non-removable media. By way of example, and not limitation, computer-readable storage media can be implemented in connection with any method or technology for storage of information such as computer-readable instructions, program modules, structured data, or unstructured data. Computer-readable storage media can include, but are not limited to, RAM (random access memory), ROM (read only memory), EEPROM (erasable programmable read only memory), flash memory or other memory technology, CD-ROM (compact disc read only memory), DVD (digital versatile disk) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or other tangible and/or non-transitory media which can be used to store desired information. Computer-readable storage media can be accessed by one or more local or remote computing devices, *e.g.*, via access requests, queries or other data retrieval protocols, for a variety of operations with respect to the information stored by the medium.

Unless otherwise defined herein, scientific and technical terms used in connection with the present application shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. In one respect, the present invention relates to the herein described compositions, methods, and respective component(s) thereof, as essential to the invention, yet open to the inclusion of unspecified elements, essential or not ("comprising). In some embodiments, other elements to be included in the description of the composition, method or respective component thereof are limited to those that do not materially affect the basic and novel characteristic(s) of the

invention ("consisting essentially of"). This applies equally to steps within a described method as well as compositions and components therein. In other embodiments, the inventions, compositions, methods, and respective components thereof, described herein are intended to be exclusive of any element not deemed an essential element to the component, composition or method ("consisting of").

BRIEF DESCRIPTION OF THE FIGURES

FIGs. 1A-1E Electron Cryomicroscopic Reconstruction of VSV-L at High Resolution. **(FIG. 1A)** Raw image of VSV-L particles in vitreous ice recorded at 1.8 mm defocus. Scale bar, 10 nm. **(FIG. 1B)** Power spectrum of the image shown in (A), with plot of the rotationally averaged intensity versus resolution. Arrow indicates the spatial frequency corresponding to 3.8 Å resolution. **(FIG. 1C)** Representative class averages. Scale bar, 10 nm. **(FIG. 1D)** Fourier shell correlation analysis: FSC, correlation between the half-set three-dimensional reconstructions (solid blue line); C_{ref} , estimated correlation between the final map and a perfect reference map containing no errors, calculated from FSC (Rosenthal and Henderson, 2003); CC_{work} and CC_{free} , correlation between the final map and refined model for working and test set of structure factors, respectively. **(FIG. 1E)** Left: overview of VSV-L reconstruction. In the view shown, the particle (241 kDa) is ~110 Å long and 80 Å wide. Right: close-up view of a representative region in the polymerase domain (RdRp). The volume shown in close-up is from the protein interior, not on the RdRp surface. Density is shown as gray mesh; polypeptide-chain backbone of the refined model, as black ribbon; side-chain atoms (carbon, nitrogen, oxygen and sulfur, as sticks. Note the continuous backbone density, α -helical grooves and resolution of bulky side chains—features that allowed building and stereochemical refinement of the atomic model.

FIGs. 2A-2E Preparation of Structure Factors for Atomic Structure Refinement. **(FIG. 2A-2D)** Preparation of Fourier coefficients from the experimental reconstruction. **(FIG. 2A)** A mask is generated around the model. **(FIG. 2B)** Density outside the mask is flattened, and density inside the mask is put on absolute scale. **(FIG. 2C)** Amplitudes and phases are calculated from the flattened map by Fourier transformation (FFT). **(FIG. 2D)** Scaling of amplitudes before refinement. **(FIG. 2E)** Estimation of figures of merits from phase-angle differences between the two half-set reconstructions.

FIGs. 3A-3C Structure of VSV-L **(FIG. 3A)** Domain organization of VSV-L. The polymerase domain (RdRp), capping domain (Cap), connector domain (CD), methyltransferase (MT) domain, and C-terminal domain (CTD) are indicated. Amino-acid residue numbers indicate functional domain boundaries. Flexible linkers 1 and 2 connect Cap to CD and CD to MT domain, respectively. Conserved regions within L proteins of non-segmented negative-strand (NNS) RNA viruses are labeled CR I–VI. Asterisks indicate the position of active site residues. **(FIG. 3B)** Ribbon diagram of

VSV-L polypeptide chain; domains indicated as in (A). **(FIG. 3C)** Substrate channels and internal cavities of VSV-L, depicted as white surface enclosed by the structure in ribbon representation. In this orientation, the entrance to the template channel leading to the active site faces down; the channel runs between the RdRp and capping domains. Nucleotides can access the RdRp active site through the channel in the foreground.

FIG. 4 Secondary structure diagram of VSV-L. Secondary structure elements along the VSV-L sequence are shown as cylinders and arrows for α helices and β strands, respectively. Domains are indicated as in FIG 3, with the exception of the polymerase domain (RdRp), which is indicated as in FIG 5, with the palm domain, the fingers and the thumb. Domain boundaries are indicated by the corresponding residue numbers.

FIGs. 5A-5C Polymerase RdRp Domain **(FIG. 5A)** Structure of the RdRp domain. Residues 35–865 are shown in ribbon representation in conventional orientation (viewed from inside the surrounding “cage” and upside down with respect to the view in Figure 2). The palm subdomain, the fingers, and the thumb are indicated. The N-terminal region is gray. **(FIG. 5B)** Close-up view of the active site. Palm, fingers and thumb are indicated as in (A). The GDN active site motif is at a β -hairpin in the palm domain. A model for the positions of the template RNA strand and two nucleotides is derived from the reovirus I3 initiation complex (PDB: 1n1h) after superposition on VSV-L RdRp. The priming loop (residues 1157–1173) intruding from the capping domain (gray) positions the initial nucleotide of the transcript. **(FIG. 5C)** Similarity of the VSV-L RdRp domain to those of other viral polymerases. Structures of influenza virus B polymerase (PDB: 4wrt; PA residues 248–716 and PB1 residues 1–616), reovirus I3 (PDB: 1muk; residues 2–890) and rotavirus VP1 (PDB: 2r7q; residues 2–778) are shown with the same orientation and coloring scheme as the VSV-L RdRp in (A).

FIGs. 6A-6D Capping Domain **(FIG. 6A)** Structure of the capping domain. Residues 866–1334 are in ribbon representation. Motifs GxxT and HR are sites of guanosine nucleotide binding and of covalent RNA attachment, respectively. Residues corresponding to positions of inhibitor-resistance mutations in human RSV polymerase (Liuzzi et al., 2005) are shown as spheres. **(FIG. 6B)** Close-up of the active site. **(FIG. 6C)** Configuration of the priming loop in VSV-L. Only the RdRp and capping domains are shown. The priming loop (residues 1157–1173) protrudes from the capping domain into the active site of the RdRp domain. **(FIG. 6D)** Proposed domain shifts to allow transcript elongation and eventual template release.

FIG. 7 Density around the Priming Loop, Contours from a 5 Å resolution density map outline the flexible priming loop that projects from the capping domain. Heavy black lines show the C α trace for the loop and adjacent polypeptide chain.

FIGs. 8A-8D Domain Reorganization (FIG. 8A) Projection angle matching between class averages of negatively stained complexes of VSV-L and P protein (top row) (Rahmeh et al., 2010, 2012) and projections calculated from the model (middle row). The bottom row shows the model in the same orientation with the individual domains colored as in FIG. 3. Numbers are correlation coefficients between model and negative-stain class averages. VSV-L:P (1–106) corresponds to the structure determined here. In the panel for VSV-L:P, an arrow indicates additional density observed in some class averages that was attributed to the bound P dimer. Scale bar, 10 nm. **(FIG. 8B)** Difference density map ($\text{map}_{\text{observed}} - \text{map}_{\text{model}}$) calculated to 5 Å resolution and shown together with the model. The map shows density present in the image reconstruction that could not be fit with a molecular model. Strong density—presumably from linkers 1 and 2, which enter and leave this density at defined points, and with potential contribution from P (tentative assignment indicated by “?”)—lines the groove between the capping and connector domains. Projection angle matching of VSV-L fragments. For VSV-L(1–1557), the negative-stain class averages suggest a conformationally variable connection between the connector domain (CD) and the polymerase (RdRp) and capping domain (Cap). The “doughnut” part of the image was selected and aligned residues 35–1334. **(FIG. 8C, FIG. 8D)** Full-length VSV-L without P. CD, MT, and CTD extend in variable orientation from the RdRp-Cap doughnut.

FIGs. 9A-9D Correlation of Images of Negatively Stained VSV-L and Fragments with Projected Views of Model-Based Density, For each panel, an image from negative-stain electron microscopy (Rahmeh et al., 2012) was correlated with projections of a density map calculated from the molecular model of VSV-L, using routines in SPIDER (Shaikh et al., 2008). The images for (A)–(D) are those in the first, third, fifth and sixth panels, respectively, in the top row of FIG. 8A. For (A) and (C), which have a subsidiary maximum in the correlation plot, the projected views for both peaks have been shown. Angular coordinates as defined in SPIDER.

FIGs. 10A-10C Methyltransferase Domain (FIG. 10A) Structure of the methyltransferase: residues 1598–1892 in ribbon representation. The consensus fold of the S-adenosyl methionine-dependent methyltransferase subdomain is highlighted. The N-terminal and C-terminal regions are in gray. **(FIG. 10B)** Close-up of the active site. The SAM/SAH binding-site motif, GxGxG, is between b1 and aA. An SAH molecule's location is derived from a superposition of its complex with dengue virus NS5 MT (PDB: 119k). Residues that participate in the methyltransferase activity are in stick detail. **(FIG. 10C)** Comparison of VSV-L MT domain with other viral AdoMet-dependent methyltransferases. Structures of dengue virus NS5 MT (PDB: 119k; residues 7–267) and vaccinia virus VP39 MT (PDB: 1av6; residues 3–297) are shown in the same orientation and highlight scheme as in (A).

FIGs. 11A-11B Secondary Structure Diagram of the Methyltransferase Domain, (FIG. 11A) Diagram of the consensus fold for AdoMet-dependent methyl transferases; α helices, β strands, and termini are

represented by circles, triangles and rectangles, respectively. The consensus AdoMet-dependent methyl transferase fold is highlighted, other regions in gray (as in FIG. 10). The positions of SAM/SAH and the active site are indicated. **(FIG. 11B)** Diagram of VSV-L MTase, vaccinia virus VP39 MTase, and flavivirus NS5 MTase, in the same highlight/color scheme as (A). Insertions or deletions of α helices or β strands are indicated as empty circles.

FIG. 12 Sequence alignment. Homologous L proteins include those of Rabies, Ebola, Measles and RSV. Alignment of their sequences shows the overall arrangement of various domains, identifies the active site residues of the protein for RdRp, PRNTase and methyltransferase activities and suggests domain boundaries for expressing various fragments of the proteins from VSV and related viruses.

FIGs. 13A-13C Production and activity of MV polymerase complex. **(FIG. 13A)** Purification of MV polymerase complex. L and P were co-expressed in insect cells and purified as described in Material and Methods and analyzed by SDS-PAGE. Position of L, P and co-purified Hsp90 are indicated. **(FIG. 13B)** De novo RNA synthesis by MV L-P (lane 1) and MV L-P/Hsp90 on a non-encapsidated 19 nt long RNA template. Reactions were set up in presence of [α^{32} P]-GTP, quenched by the addition EDTA/formamide and analyzed on a 20% polyacrylamide/7 M urea gel. De novo initiation product sizes are indicated on the left. **(FIG. 13C)** De novo RNA synthesis by MV L-P/Hsp90 in presence of increase concentration of Geldanamycin.

FIGs. 14A-14B Inhibition of MV L-P by As136a. **(FIG. 14A)** Structure of the As136a inhibitor is indicated on top. Inhibitory effect of increasing As136a concentrations on RNA synthesis by MV L-P. Total amount of products observed in (B) were quantified and the percentage of MV L-P RNA synthesis inhibition by the different As136a concentrations were plotted. Error bars represent the standard deviation from the mean of independent experiments. **(FIG. 14B)** 20% polyacrylamide/7 M urea gel showing the inhibitory effect of increasing As136a concentrations on RNA synthesis by MV L-P.

FIGs. 15A-15B Inhibition of MV L-P by ERDRP0519. **(FIG. 15A)** Structure of the ERDRP0519 inhibitor is indicated on top. Inhibitory effect of increasing ERDRP0519 concentrations on RNA synthesis by MV L-P. Total amount of products observed in (B) were quantified and the percentage of MV L-P RNA synthesis inhibition by the different inhibitor concentrations were plotted. Error bars represent the standard deviation from the mean of independent experiments. **(FIG. 15B)** 20% polyacrylamide/7 M urea gel showing the inhibitory effect of increasing As136a concentrations on RNA synthesis by MV L-P.

FIGs. 16A-16D Effect of ERDRP0519 on MV L resistant-mutants in complex with P. Effect of increasing ERDRP0519 concentrations on RNA synthesis by MV **(FIG. 16A)** L_{T776A/T751I}-P, **(FIG.**

16B) L_{T776A-P} and **(FIG. 16C)** L_{T751I-P} **(C)**. **(FIG. 16D)** Total amount of products observed in **(B)** and **(C)** were quantified and the percentage of RNA synthesis inhibition by the different inhibitor concentrations were plotted. Error bars represent the standard deviation from the mean of independent experiments.

FIGs. 17A-17C. Effect of inhibitor on recombinant measles containing substitutions in L To demonstrate that the two mutations lead to complete resistance of the measles virus to the non-nucleoside inhibitor, a recombinant measles virus KS strain expressing eGFP was engineered. The kinetics of replication of the engineered variant MV-L_{T776A/T751I} were reduced compared with the wild type parent as evidenced by the decreased viral yield at 24h post inoculation of cells (**FIG. 17A**) consistent with a fitness cost to the resistance mutations. The recombinant MV-L_{T776A/T751I} variant was resistant to the non-nucleoside inhibitor on viral replication as judged by eGFP expression in the presence of 25 μ M of the bioavailable inhibitor (**FIG. 17B, 17C**). This result demonstrates that substitutions in measles L lead to complete resistance of the virus to the non-nucleoside inhibitor in vitro and in cells but uncover a fitness cost to viral replication associated with those substitutions.

FIG. 18 MV L Amino acid sequence alignment with CDV and VSV. Amino acid sequence alignment of L from MV, CDV and VSV has been realized using T-Coffee and analyzed using Esript. The sequences are shown from MV L residue 266 to 1257. Resistant mutations that have been identified for MV with As136a and CDV with ERDRP0519 are noted with dots and squares, respectively.

FIGs. 19A-19C Localization of resistant mutation in L 3D structure. **(FIG. 19A)** Overall structure of VSV L showing the resistant mutations obtained for MV and CDV in side chain detail. **(FIG. 19B)** RdRp domain of VSV L shown as the right hand orientation. The GDN residues responsible for RNA synthesis are shown in yellow and the resistant mutants in side chain detail. **(FIG. 19C)** Model of MV L ERDRP0519 binding domain build using VSV L map (PDB) and coot. In yellow are shown the GDN residue, in side chain detail the resistant mutants and in separate stick figure the ERDRP0519 inhibitor.

FIGs. 20A-20C Production and activity of RABV polymerase complex. **(FIG. 20A)** Purification of RABV polymerase complex elements. L (left) and P (right) were individually expressed in insect cells and bacteria, respectively. Both proteins were purified as described in Material and Methods analyzed by SDS-PAGE. **(FIG. 20B)** De novo RNA synthesis by RABV L on a RABV Le19 RNA template in absence or presence of either RABV P or VSV P. Reactions were set up in presence of [α 32P]-GTP, quenched by the addition EDTA/formamide and analyzed on a 20% polyacrylamide/7 M urea gel. De novo initiation product sizes are indicated on the left. **(FIG. 20C)**. De novo RNA synthesis by VSV L on a VSV Le19 RNA template in absence or presence of either VSV P or RABV P.

FIG. 21 Identification of the minimal P L-enhancing domain. All P proteins were individually expressed in bacteria and purified as described in Examples. **(FIG. 21)** De novo RNA synthesis by RABV L on a RABV Le19 RNA template in absence or presence of either RABV P or VSV P, P₁₋₉₁, P₂₁₋₉₁, P₁₁₋₉₁, P₁₋₄₀, P₁₋₅₀ or P₁₁₋₅₀.

FIG. 22A-22C Template specificity for RNA synthesis by RABV L-P. **(FIG. 22A)** De novo RNA synthesis by RABV L-P on RABV Le19 and VSV Le19 RNA templates. Sequence of both RNAs are indicated on top, stars indicating non-conserved nucleotide between the two templates. **(FIG. 22B)** Comparison of total RNA synthesis by RABV L-P on RABV Le19 and VSV Le19. Total RNA synthesis was quantified, normalized to levels of RNA synthesis on RABV Le19 and graphed. Error bars represent the standard deviation from the mean of three independent experiments. **(FIG. 22C)** Individual reaction products of different sizes synthesized in **(FIG. 22A)** for RABV Le19 (top) and VSV Le19 (bottom) were quantified and normalized to the size of the 11 mer. Error bars represent the standard deviation from the mean of three independent experiments.

FIG. 23 Template specificity for RNA synthesis by VSV L-P. **(FIG. 23A)** De novo RNA synthesis by VSV L-P on VSV Le19 and RABV Le19 RNA templates. **(FIG. 23B)** Comparison of total RNA synthesis by VSV L-P on VSV Le19 and RABV Le19. Total RNA synthesis was quantified, normalized to levels of RNA synthesis on VSV Le19 and graphed. Error bars represent the standard deviation from the mean of three independent experiments. **(FIG. 23C)** Individual reaction products of different sizes synthesized in **(FIG. 23A)** for RABV Le19 (top) and VSV Le19 (bottom) were quantified and normalized to the size of the 11 mer. Error bars represent the standard deviation from the mean of three independent experiments.

FIG. 24A-24B Activity of RABV L on VSV L N-RNA. **(FIG. 24A)** Model showing domain structure of P highlighting the CTD which was exchanged. **(FIG. 24B)** Activities of L-P complexes on VSV N-RNA templates.

FIG. 25 Model for Transcription of an RNP Complex by VSV-L Left: initiation complex, with domains organized as in the structure described here. Arrow shows direction of capping domain displacement required for the transition to an elongation complex. Right: elongation requires both displacement of the capping domain (with likely accompanying reorganization of the CD, MT, and CTD) and displacement of two to three N subunits from the template residues looped into the polymerase. The N subunits are shown linked as a continuous chain, as suggested by their structure (Green et al., 2006).

FIG. 26 Correlation of Model and Density, Related to Experimental Procedures. Residue-by-residue correlation, with each panel corresponding to a domain. Residue numbers above the plots for the

capping and connector domains indicate segments of poor density. The gap in the methyltransferase plot corresponds to a short loop omitted from the model. Calculated with CNS.

FIG. 27 Start-stop model of sequential transcription. A schematic of the VSV genome is depicted 3'-5' showing the leader (le) and trailer (tr) regions that contain essential cis-acting signals necessary to promote RNA synthesis. At each gene junction a highly conserved sequence 3'-AUACUUUUUUU GA UUGUCNNUAG-5' provides critical signals for the termination and polyadenylation of the upstream mRNA and initiation and capping of the downstream mRNA. The viral N=nucleocapsid, P=phospho, M=matrix, G=glyco- and L=large protein genes are shown along with the mRNA that are sequentially transcribed from them.

FIG. 28A-28D CryoEM reconstruction of VSV L stabilized by a fragment of P. (FIG. 28A) Surface view of density map. Oval indicates approximate region of the detailed map in (FIG. 28C). The polymerase domain is the large, roughly spherical lobe in the lower half of the figure; the capping and methyltransferase domains are at the top. (FIG. 28B) Surface view as in (FIG. 28A), but superposed on rotavirus VP1 ribbon diagram (pale magenta). The superposition illustrates the similarity of the polymerase domain and VP1 (1086 residues). Individual secondary structural elements such as alpha helices may have different lengths and somewhat different local packings. The overall topology of the chains are the same over most of the polymerase domain. The colored ovals at the top are rough outlines of the three smaller domains packed against one surface of the polymerase domain. Dashed oval shows the outline of a domain to the rear of the other two. The channel through which nascent transcript emerges (by analogy with VP1) leads toward this domain. (FIG. 28C) Detail of density for part of an α -helix, as outlined in (FIG. 28A). (FIG. 28D) Detail of density for the polymerase catalytic site.

FIG. 29A-29B (FIG. 29A) Isolation of an N-P-L complex. SDS-PAGE gel of a purified complex comprising VSV L together with a fragment of P, and an N-terminal truncation of N is shown. The N-protein was truncated by the deletion of the N-terminal arm which favors the formation of the RNA-template associated form of N. The viral N and P proteins were co-expressed in E. coli and L from a recombinant baculovirus vector in insect cells. The complex purified by nickel affinity chromatography and size exclusion. **(FIG. 29B)** Enzymatic properties of the N-P-L complex. RNA synthesis activity of L (1) L-P1-106 (2) and L-P1-106-N22-423 (3) on a naked 19 mer RNA.

FIG. 30A-30C Structures of N in complex with P. (FIG. 30A) Structure of three molecules of N-RNA in complex with the P_{CD} PDB = 3HHZ (FIG 30B) Structure of N in complex with the N-terminal domain of P. P binds in between the N-terminal and the C-terminal of N, PDB = 3PMK^{t451}. (FIG. 30C) Location of the R146 residue that is a histidine in the poIR1 mutant. Shown on the structure of the N-RNA, PDB = 2GIC.

FIGS. 31A-31B The P oligomerization domain is dispensable. (FIG. 31A) Reconstitution of RNA synthesis on the N-RNA template. (FIG 31B) Plaque morphology of wild type recombinant virus (Pwt) or that lacking the P oligomerization domain (Pmin0D).

FIGs. 32A-32C Inhibition by M of RNA synthesis on a naked RNA template by L and L-P. (FIG 32A) Structure of the C-terminal domain of VSV M, note the N-terminal domain is not resolved⁽²⁴⁾. (FIG 32B) Inhibitory effect of increasing M concentrations on RNA synthesis by VSV L-P. Reactions were performed in presence of [³²P]-GTP and analyzed on a 20% polyacrylamide/7M urea gel. (FIG 32C) Inhibitory effect of increasing M concentrations on RNA synthesis by VSV L and L-P. Total amount of products was quantified and the percentage of VSV L RNA synthesis inhibition by the different M concentrations was plotted. Error bars represent the standard deviation from the mean of independent experiments.

FIG. 33 Non nucleoside inhibitors of Measles virus RNA-dependent RNA polymerase,

FIG. 34 Effect of ERDRP0519 on measles virus L/P complex. The inhibitor inhibits with an IC50 of 100nM

FIG. 35 Compounds AS-136a and ERDRP-00519 inhibit measles virus RdRp activity. Values are expressed relative to vehicle-treated samples and represent average of three experiments.

FIG. 36 Effect of inhibitors on MV L resistant-mutants in complex with P.

FIG. 37 Measles virus L/P inhibitor mutant sequence modeled onto the VSV polymerase. The mutation site identified close to the active site.

FIG. 38 Localization of resistant mutation in L 3D structure.

FIG. 39 RNA synthesis by MV L-P complex.

FIG. 40 Ribbon diagram of protein structure.

DETAILED DESCRIPTION OF THE INVENTION

As discussed in the Summary above, the technology described herein relates to the identification of high resolution 3D structure for the L proteins of NNS viruses and the use of such structures to identify agents that inhibit one or more activities of the L protein. The high resolution 3D structures of NNS virus L proteins have been elusive, due in large part to the large size of the protein and the dynamic nature of the various protein domains. The size and flexibility of the protein has precluded,

for example, the use of the standard X-ray crystallography approach, as these properties make the preparation of the necessary protein crystals difficult, if not impossible to achieve. Despite these issues, described herein is the solution of an NNS virus L protein at high resolution. The solution of the VSV L protein's structure at this resolution not only demonstrates the ability to apply the same or very similar methods to the solution of other NNS virus L proteins, but, using the VSV L protein high resolution 3D structure and homology modeling based upon the amino acid sequences of other L proteins, one can prepare useful 3D structures or models for other members of the family. Further, the high resolution 3D structure of the VSV L protein permits the mapping of mutations that result in resistance to known L protein inhibitors onto the high resolution structure to permit the rational design of inhibitors with improved activity or even, for example, broader activity against more than one NNS virus.

The following describes the various methods and considerations necessary to use the technology disclosed herein.

NNS viruses

The non-segmented negative-strand (NNS) RNA viruses include some of the most lethal human and animal pathogens, including Ebola virus and rabies virus. Their multifunctional, large (L) polymerase proteins, carried within the virions (Baltimore et al., 1970), have biochemical properties that distinguish them from most other RNA polymerases of viruses or of their hosts. In addition to their RdRp activity (Emerson and Wagner, 1973), NNS RNA virus L proteins catalyze an unusual sequence of mRNA capping reactions (Hercyk et al., 1988), and the RdRp itself polyadenylates the viral message (Hunt et al., 1984). A nucleocapsid (N) protein sheath coats the genomic RNA, and the viral polymerase uses this N-RNA complex as template, rather than uncoated RNA.

Understanding of RNA synthesis in NNS RNA viruses comes principally from studies of vesicular stomatitis virus (VSV)—an enveloped, bullet-shaped, rhabdovirus, closely related to rabies virus. VSV causes an acute disease of livestock. The VSV L protein does not bind the N-RNA template directly but requires a cofactor, the viral phosphoprotein (P), as a bridge (Green and Luo, 2009). Delivery of the N-RNA-L-P complex into the cytoplasm when the virus enters a cell initiates infection. Transcription starts at the 3' end of the genome and produces a triphosphate 47-nucleotide leader RNA, followed by sequential transcription of five capped and polyadenylated mRNAs (Abraham and Banerjee, 1976; Ball and White, 1976; Whelan and Wertz, 2002). At each gene junction, L terminates synthesis of the upstream gene, adding a polyadenylate (polyA) tail by iterative transcription of a U7 tract (Barr and Wertz, 2001; Barr et al., 1997; Stillman and Whitt, 1997). It then transcribes the downstream gene. Approximately 30% attenuation occurs at each successive gene, with dissociation of the template (Iverson and Rose, 1981). Replication also starts at the 3' end of the

genome, but encapsidation by newly synthesized N accompanies synthesis of the nascent RNA strand, and in this mode, L ignores all the cis-acting signals that dictate sequential transcription of mRNAs (La Ferla and Peluso, 1989; Patton et al., 1984; Peluso and Moyer, 1983).

The various enzymatic activities of L are tightly linked. A GDP polyribonucleotidyl transferase (PRNTase) adds the cap structure (Ogino and Banerjee, 2007) when the nascent RNA chain length has reached 31 nucleotides, as shown by artificially stalling transcription at various chain lengths (Tekes et al., 2011). The unconventional mechanism of cap addition proceeds through a covalent adduct between a histidine residue on L (H1227) and the monophosphate nascent RNA, which is transferred onto a GTP-derived GDP acceptor (Li et al., 2008; Ogino and Banerjee, 2007). Failure to cap results in the premature termination of transcription, which links cap addition to RdRp processivity (Li et al., 2008; Stillman and Whitt, 1999; Wang et al., 2007). Subsequent methylation is first at the 2'O position on the ribose of the first nucleotide and then on the N7 of the capping guanylate—opposite to the typical order of modification (Rahmeh et al., 2009). Failure to methylate, either by exogenously manipulating the concentration of S-adenosyl homocysteine (SAH) to compete with the S-adenosyl methionine methyl donor or by mutating critical catalytic residues in the methylase domain, can result in hyper-polyadenylation of mRNA transcripts, linking the methylation activity to the RdRp (Galloway and Wertz, 2008; Li et al., 2009; Rose et al., 1977). Both cap addition and subsequent methylation require specific sequence elements within the first 10 nt of the mRNA (Wang et al., 2007). The uncapped leader lacks those signals (Li et al., 2008; Ogino and Banerjee, 2007).

During RNA synthesis, a few subunits of N dissociate from the template RNA for access to the catalytic site of RdRp and then reassociate as the process continues (Albertini et al., 2006; Green et al., 2006). P can help coordinate these events (Green and Luo, 2009). There are nine nucleotides associated with each N subunit in the ribonucleoprotein (RNP). Full polymerase processivity and correct recognition of the cis-acting signals in the viral genome both require N (Morin et al., 2012).

Images from negative-stain electron microscopy (EM) of purified VSV-L show a ring-like “core” decorated by a set of three variably oriented globular appendages (Rahmeh et al., 2010). Truncations of L map the RdRp to the ring-like domain. Formation of a complex with P causes the globular appendages to reorganize into a compact tail, wrapped onto one side of the ring (Rahmeh et al., 2010). Complex formation with P enhances RdRp initiation and processivity, but P itself has no enzymatic activity. A fragment of P comprising residues 35–106 is sufficient to induce the conformational rearrangement (Rahmeh et al., 2012). Segments in the center and at the C-terminal end of the P polypeptide chain mediate dimerization and N-RNA binding, respectively (Ding et al., 2006; Green and Luo, 2009). Images from negative-stain EM of purified L-P complexes are a mixture of single

and dimeric L species, in which the two L molecules have variable relative orientation (Rahmeh et al., 2010). However these studies do not define what form of complex prevails during RNA synthesis.

Herein, the structure of a complex of L with P(35–106) by electron cryomicroscopy (cryo-EM) is determined. Into a density map at 3.8 Å resolution, a complete model of the 2109-residue polypeptide chain is built. The result shows that even for a fully asymmetric structure with molecular mass <250 kDa, the features in a density map from single-particle cryo-EM can be well enough defined for a complete de novo chain trace. Five domains have been distinguished: three—an RNA-dependent RNA polymerase (RdRp), an mRNA capping domain, and a methyltransferase domain—with assigned enzymatic activity and two—a connector between the capping and methyltransferase domains and a C-terminal domain—that appear to have largely organizational roles. It is believed that the conformation of the protein in the complex examined in the examples described herein corresponds to an initiation state, ready to accept the 3' end of a template. Elongation beyond one or two nucleotides will require large-scale domain reorganization. As such, relative positions of the domains and the ability to block movements then provides another level of activity to target for potential antivirals.

The NNS viruses can be divided into two main groups based on their genomic RNA: non-segmented NS (NNS) RNA viruses and segmented NS (SNS) RNA viruses. The NNS RNA viruses comprise four families, the Rhabdoviridae (vesicular stomatitis virus (VSV) and rabies virus), Paramyxoviridae (measles and respiratory syncytial viruses (RSV)), Filoviridae (Ebola and Marburg viruses) and the Bornaviridae (Borna disease virus). The SNS RNA viruses comprise three families, the Arenaviridae (lymphocytic choriomeningitis virus and Machupo virus (MACV)), Bunyaviridae (Rift Valley fever virus) and Orthomyxoviridae (influenza A virus).

The NS RNA viruses share common replicative machinery comprising a protein–RNA complex. It is composed of the genomic RNA which is wrapped over its entire length by the nucleotide protein (N) to form the ribonucleocapsid (N-RNA), and the polymerase complex. One of the features of Mononegavirales is that this polymerase complex consists of two main elements: the protein L and its co-factor, the phosphoprotein (P) which allows interaction between L and N-RNA complex. Those NP-RNA templates are copied by the virally encoded RNA dependent RNA polymerase (RdRP) in two synthetic modes: mRNA transcription and genome replication. For the NNS RNA viruses that single genome contains a tandem array of 5–10 genes that are sequentially copied by the viral polymerase, whereas the SNS RNA virus polymerases copy each template into a single mRNA.

The enzymatic activities necessary for copying of the NP-RNA templates include an RNA-dependent RNA polymerase (RdRp), as well as the enzymes required for mRNA cap formation namely a GDP:polyribonucleotidyltransferase (PRNTase) and a dual specificity mRNA cap methyltransferase

that are only utilized during mRNA transcription. All of the necessary enzymatic activities reside within a 250 kDa large (L) polymerase protein. Bioinformatics analyses have identified six conserved domains (CR I to CR VI) in NNSV L proteins that are connected by variable linker regions (Figure ; Poch et al., 1988, 1990; Svenda et al., 1997). However, the precise roles for each of these L domains in RdRp function are still largely unclear. CR I has been implicated in L oligomerization (Cevik et al., 2003, 2004; Smallwood and Moyer, 2004) and L-P interactions (Horikami et al., 1994; Holmes and Moyer, 2002; Cevik et al., 2003, 2004; Chattopadhyay and Shaila, 2004), CR III is involved in phosphodiester bond formation for RNA polymerization (Malur et al., 2002b), and CR VI contains methyltransferase activity (Poch et al., 1990; Ferron et al., 2002). A conserved GXXTnHR motif in CR V of VSV L is thought to mediate unusual capping of the viral mRNAs through transfer of 5'-monophosphate-mRNA onto GDP.

L protein as a target

An inherent challenge of all pathogen-directed drug discovery campaigns is a narrow indication spectrum of the therapeutic candidate, limiting inhibitory activity to a specific member or, at best, a single genus within the virus family. The L protein represents a rich target for drug discovery campaigns, due to its multidomain organization and the concentration of several essential enzymatic activities in a single protein. The precedence established by the development of inhibitors of, for instance, HIV reverse transcriptase and Hepatitis C virus polymerase underscores the value of high-resolution structural information for the identification and optimization of hit structures, the molecular understanding of the mechanism of inhibitory activity, and, potentially, the proactive design of analogs with increased resilience against viral escape from inhibition (Nijhuis et al., 2009; Adams et al., 2010; Das et al., 2011; Halfon and Locarnini, 2011; Canhou, 2012; Lloyd et al., 2014). However, the drug development field is hampered by the current lack of high-resolution structural information for any mononegavirales L protein. The high resolution NNS virus L protein structure described herein is a major step forward in this regard.

An envisioned drug application including post-exposure prophylactic use affects the drug profile requested of a desirable anti-NNS therapeutic; a successful candidate must be safe and efficacious, amenable to cost-effective manufacture, ideally be shelf-stable at ambient temperature, and must be orally bioavailable. Of small-molecule chemical compounds, large molecule biologics, and peptidic biopharmaceutical as candidate drug classes, small-molecules are most suitable to fulfill these divergent demands (Ganellin et al., 2013). Two main classes of polymerase-targeted drugs are currently in clinical use for, among others, antiretroviral therapy, human cytomegalovirus therapy, and HCV therapy, competitive nucleotide/nucleoside substrate analogs and non-nucleoside allosteric inhibitors (Sun et al., 2007, 2008; Andrei et al., 2008; Brown, 2009; Krumm et al., 2011; Mercorelli et al., 2011).

*Assays for L protein**RdRp assay*

The structural insights obtained from NS RNA virus templates indicate that the RNA must be displaced from NP to facilitate its copying by polymerase. Those observations catalyzed the development of an improved system to study polymerase biology *in vitro* in which polymerase alone could use naked RNA as template. These *in vitro* models are powerful tools to validate and study how selected inhibitors of L protein and/or NNS virus polymerase function. In agreement with a model of N dissociation, it has been demonstrated via an *in vitro* system that NNS L protein, in the absence of NP and P, initiates RNA synthesis on a naked RNA corresponding to the first 19 nucleotides of the genomic RNA. The *in vitro* initiation assay uses essentially only the enzyme and substrate components: L protein and naked RNA. RNA corresponding to the first 19 nt of the genome (Le19) can be chemically synthesized and used as template for RNA synthesis *in vitro*. Reaction products can be visualized following 30 min of incubation, separation on a polyacrylamide/ urea gel and analyzed using autoradiography. The *in vitro* assay has been shown to be adapted to study the activity of L protein of other viruses of the NNS family, e.g. measles virus L protein and rabies virus L protein. Detail for these assays are described in the examples herein.

In vitro capping assay

to test for capped products, *in vitro* transcription reactions are carried out in the presence of [α -32P]GTP; RNA is purified and digested with TAP, and the products resolved by thin layer chromatography. Product that comigrated with a 7mGp marker, confirms that the RNA is capped and methylated.

In vitro methylation assay

triphosphorylated RNA transcripts are capped in the presence of α -32P GTP (3,000 Ci/mmol; Perkin-Elmer) and vaccinia virus capping enzyme (Epicentre) in 15- μ l reaction mixtures. After 2 h at 37°C, the reaction mixtures are treated with 10 U of calf intestinal alkaline phosphatase (NEB) for 1 h, followed by purification. Capped RNA is quantified by measuring the incorporation of [α -32P]GTP and extrapolation of molar quantities from a standard curve of moles versus counts per minute. For methylation at the G-N-7 position, 100 μ M S-adenosyl methionine is included in the capping reaction mixture. For methylation at the ribose 2'-O position or at both the G-N-7 and ribose 2'-O positions, the capped RNA or the G-N-7-methylated RNA is incubated with 100 μ M SAM and 7.5 U of vaccinia virus 2'-O MTase (Epicentre). The products of this reaction are purified, digested with nuclease P1, and resolved by TLC on PEI cellulose F sheets. Nuclease P1 digestion cleaves the 3'-5' phosphodiester bonds in single-stranded RNA but does not cleave the 5'-5' bond of the cap structure,

resulting in the release of GpppA from unmethylated RNA. Following incubation of the RNA with purified L protein, the products of nuclease P1 digestion comigrated with a m7GpppAm marker, demonstrating that L possesses both G-N-7 and 2'-O MTase activities.

Rational drug design

In another embodiment, the technology devised herein provides a method of using an atomic structure of this invention in an inhibitor screening assay, the method comprising selecting a potential inhibitor by performing rational drug design with the three-dimensional structure determined for the L protein. The selecting is performed using with computer modeling. The rationally designed or improved candidate inhibitor is then contacted with an L protein and the ability of the potential inhibitor to inhibiting L protein activity *in vitro*, infection or replication of an NNS virus among others in measured.

Numerous computer programs are available and suitable for rational drug design and the processes of computer modeling, model building, and computationally identifying, selecting and evaluating potential inhibitors of L protein. These include, for example, GRID (available from Oxford University, UK), MCSS (available from Molecular Simulations Inc., Burlington, Mass.), AUTODOCK (available from Oxford Molecular Group), FLEX X (available from Tripos, St. Louis, Mo.), DOCK (available from University of California, San Francisco), CAVEAT (available from University of California, Berkeley), HOOK (available from Molecular Simulations Inc., Burlington, Mass.), and 3D database systems such as MACCS-3D (available from MDL Information Systems, San

Leandro, Calif.), and UNITY (available from Tripos, St. Louis, Mo.). Potential agents can also be computationally designed "de novo" using such software packages as LUDI (available from Biosym Technologies, San Diego, Calif.), LEGEND (available from Molecular Simulations Inc., Burlington, Mass.), and LEAPFROG (Tripos Associates, St. Louis, Mo.). Compound deformation energy and electrostatic repulsion, can be evaluated using programs such as GAUSSIAN 92, AMBER, QUANTA/CHARMM, AND INSIGHT II/DISCOVER. These computer evaluation and modeling techniques can be performed on any suitable hardware including for example, workstations available from Silicon Graphics, Sun Microsystems, and the like. Any such use, as pertains to the present invention is to be considered an embodiment thereof. These techniques, methods, hardware and software packages are representative and are not intended to be a comprehensive listing, and any such use represents an embodiment of the invention. Other modeling techniques known in the art can also be employed in accordance with this invention. See for example, N. C. Cohen, *Molecular Modeling in Drug Design*, Academic Press (1996) (and references therein), and software identified at internet sites including the CMBI Cheminformatics Suite at cheminf.cmbi.ru.nl/news/chemnews_his.shtml (Feb.

27, 2006), and the NIH Molecular Modeling Home Page at World Wide Web at [?muni.cz/usr/mejzlik/mirrors/molbio.info.nih.gov/modeling/softWare list/](http://www.muni.cz/usr/mejzlik/mirrors/molbio.info.nih.gov/modeling/softWare list/) (Feb. 27, 2006).

The agent can be selected by performing rational drug design with the three-dimensional high resolution structure (or structures) determined as herein, especially in conjunction with computer modeling and methods described above. The agent is then obtained from commercial sources or is synthesized from readily available starting materials using standard synthetic techniques and methodologies known to those of ordinary skill in the art. The agent is then assayed, in one embodiment, to determine its ability to inhibit structure, function and/or activity of an NNS virus L protein or a domain thereof.

The potential inhibitory activity of a chemical compound on an NNS L protein can be analyzed prior to its actual synthesis and testing by the use of computer modeling techniques, as is known to those of ordinary skill in the art. One of ordinary skill in the art can use, in other embodiments of this invention, any one of several methods to screen chemical entities or fragments for their ability to associate with L protein or a domain thereof. Selected fragments or chemical entities can then be modeled in a variety of orientations, or docked, onto a specific target domain. Docking can be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics with standard molecular mechanics forcefields, such as CHARMM and AMBER. Specialized computer programs can also assist in the process of selecting fragments or chemical entities. These include, in one embodiment, the programs GRID, MCSS, AUTODOCK and DOCK.

In another embodiment, the agent can be designed as a whole or “de novo” design using an empty binding site. These methods can include the use of programs such as LUDI, LEGEND and LeapFrog, each of which represents an embodiment of the technology described herein.

Embodiments of various aspects described herein can be defined in any of the following numbered paragraphs:

1. A method of identifying an inhibitor of a non-segmented negative strand (NNS) virus, the method comprising: (a) using a high resolution 3D structure of an NNS virus L protein to select a candidate inhibitor from a virtual library of 3D structures of potential candidates by performing structure based computational modeling; (b) obtaining the candidate inhibitor; and (c) measuring an activity of L protein when contacted with the candidate inhibitor obtained in step (b), wherein a decrease in the activity confirms that the candidate is an inhibitor of the NNS virus.

2. A method of identifying an inhibitor of a non-segmented negative strand (NNS) virus polymerase, the method comprising: (a) using a high resolution 3D structure of an NNS virus L protein to select a

candidate inhibitor from a virtual library of 3D structures of potential candidates by performing structure based computational modeling; (b) obtaining the candidate inhibitor; and (c) measuring L protein polymerase activity when contacted with the candidate inhibitor obtained in step (b), wherein a decrease in the activity of the polymerase confirms that the candidate is an inhibitor of the NNS virus polymerase.

3. A method of identifying an inhibitor of an NNS virus the method comprising: (a) screening a library of candidate molecules for inhibition of an NNS L protein in an *in vitro* biochemical assay for an L protein activity; (b) using a high resolution 3D structure of the L protein to determine the binding site of potential candidates obtained in step (a) by computer modeling; (c) identifying a potential refinement to the candidate inhibitor; (d) obtaining a candidate inhibitor molecule comprising the refinement identified in step (c); and (e) measuring the activity of the L protein when contacted with the candidate inhibitor obtained in step (d), wherein a decrease in the activity of the L protein confirms that the candidate is an inhibitor of the NNS virus.

4. A method of identifying an inhibitor of L protein of an NNS virus, the method comprising: (a) using a high resolution 3D structure of the L protein of vesicular stomatitis virus (VSV) to create an atomic model of the L protein of a second NNS virus by homology modeling the amino acid sequence of the L protein of the second NNS virus onto the high resolution 3D structure of the VSV L protein ; (b) using the atomic model of the second NNS virus L protein to select a candidate inhibitor from a virtual library of 3D structures of potential candidates by performing structure based computational modeling;

(c) obtaining the candidate inhibitor; and (d) measuring an activity of L protein of the second NNS virus when contacted with the candidate inhibitor obtained in step (b), wherein a decrease in the activity confirms that the candidate is an inhibitor of the second NNS virus.

5. The method of any one of paragraphs 1-4 wherein the high resolution 3D structure has a resolution of at least 3.8 Å.

6. The method of paragraph 4 wherein the high resolution 3D structure is an atomic structure determined via cryo-electron microscopy and single particle analysis.

7. The method of any one of paragraphs 1-4 wherein the NNS virus L protein is vesicular stomatitis virus L protein.

8. The method of paragraph 7 wherein the VSV L protein high resolution structure has space group P1, so as to form a unit cell of dimensions of $a=112.0$ Å, $b=143.0$ Å, $c=106.0$ Å and $\alpha, \beta, \gamma=90^\circ$.

9. The method of paragraph 7 wherein the VSV L protein high resolution 3D structure has the coordinates as set out, for example, in the structure deposited with the RCSB Protein Data Bank with Reference Number 5A22, and is accessible on the World Wide Web at rcsb.org/pdb/explore.do?structureId=5A22.
10. The method of any of paragraphs 1, 3 or 4 wherein the activity of the L protein measured is selected from the group consisting of RNA dependent RNA polymerase activity, GDP polyribonucleotidyltransferase (PRNTase) activity, methyltransferase activity, and the ability to structurally rearrange L protein domains during L protein functions.
11. The method of paragraph 2 or paragraph 10 wherein the RNA dependent RNA polymerase activity measured is selected from the group consisting of transcript initiation, elongation and termination.
12. The method of paragraph 1, 2 or 4 wherein, rather than selecting a candidate inhibitor from a virtual library of 3D structures of potential candidates, the high resolution 3D structure of the NNS virus L protein is used to design a potential inhibitor.
13. The method of paragraph 12 wherein the potential inhibitor is designed to interact with an L protein domain selected from the group consisting of the RNA dependent RNA polymerase (RdRp) domain, the mRNA capping domain, the methyltransferase domain, the loop primer domain, the connector domain, and the C-terminal domain.
14. A method of mapping the location of resistance to a known inhibitor of an NNS virus L protein on the 3D structure of the L protein, the method comprising: (a) homology-based structural mapping of the sequence comprising the resistance mutation onto the high resolution 3D structure of an NNS virus L protein.
15. The method of paragraph 14 wherein the high resolution 3D structure is a 3.8Å resolution structure.
16. A method of inhibiting an NNS virus, the method comprising contacting an NNS virus or a cell infected with an NNS virus with an inhibitor identified according to the method of any one of paragraphs 1-4.
17. An NNS virus inhibitor identified using the method of any one of paragraphs 1-4.

18. The method of any one of paragraphs 1-4 wherein the identified inhibitor binds a domain of the L protein at a region selected from the group consisting of the RNA dependent RNA polymerase (RdRp) domain, the mRNA capping domain, the methyltransferase domain, the loop primer domain, the connector domain, and the C-terminal domain of the L protein.
19. The method of any one of paragraphs 1-4 wherein the identified inhibitor attenuates the RNA dependent RNA polymerase activity, the GDP polyribonucleotidyltransferase (PRNTase) activity, the methyltransferase activity, the loop priming activity, or the ability of the L protein to structurally rearrange its domains during L one or more of its functions.
20. The method of any one of paragraphs 1-4 which further comprises the step of contacting the confirmed candidate inhibitor with a cell infected with an NNS virus.
21. The method of any of paragraphs 1-4 wherein the NNS virus is selected from the viruses of the Filoviridae, Paramyxoviridae, Rhabdoviridae and Bornaviridae families.
22. The method of paragraph 21 wherein the NNS virus is a Filovirus, selected from Ebola virus, Marburg virus and Cuevavirus.
23. The method of paragraph 21 wherein the NNS virus is a Paramyxovirus selected from measles virus, Nipah virus, Hendra virus, respiratory syncytial virus and Newcastle disease virus.
24. The method of paragraph 21 wherein the NNS virus is a Rhabdovirus selected from Lyssavirus, Vesiculovirus, Perhabdovirus, Sigmavirus, Ephemero virus, Tibrovirus, Tupavirus and Sprivivirus.
25. A method of designing a candidate inhibitor of an NNS virus, the method comprising: (a) identifying an NNS virus mutant that escapes the activity of a known inhibitor of the L protein of an that virus; (b) identifying the site of the L protein resistance mutation by sequence analysis; (c) using homology-based structural mapping of the sequence comprising the resistance mutation to map the mutation site onto a high resolution 3D structure of an NNS virus L protein; (d) using computer modeling with the high resolution 3D structure to selecting a candidate binding molecule that targets the site of the resistance mutation mapped in step (c); and (e) obtaining the candidate inhibitor molecule identified in step (d).
26. The method of paragraph 25 further comprising the step of contacting L protein of the NNS virus with the candidate inhibitor in an *in vitro* assay of L protein activity, wherein a decrease in L protein activity indicates the candidate inhibitor is an NNS virus inhibitor.

27. The method of paragraph 25 further comprising the step of contacting the candidate inhibitor with the NNS virus and assaying viral function, wherein a decrease in viral function indicates the candidate inhibitor is an NNS virus inhibitor.

28. A method of designing a candidate inhibitor of an NNS virus, the method comprising: (a) identifying an NNS virus mutant that escapes the activity of a known inhibitor of the L protein of an that virus; (b) identifying the site of the L protein resistance mutation by sequence analysis; (c) using homology-based structural mapping of the sequence comprising the resistance mutation to map the mutation site onto a high resolution 3D structure of an NNS virus L protein; (d) identifying a potential refinement to the known inhibitor that permits interaction with the mutant L protein; (e) obtaining the inhibitor molecule comprising the refinement identified in step (d); and (f) assaying mutant virus replication or infection in the presence and absence of the refined inhibitor molecule, wherein a decrease in replication or infection indicates the refined inhibitor is effective against the mutant virus.

29. Use of an NNS virus inhibitor identified according to the methods of any one of paragraphs 1-4, 25 and 28 for the treatment of an individual infected with an NNS virus.

30. The use of paragraph 29 in which the NNS virus is selected from the group consisting of viruses of the Filoviridae, Paramyxoviridae, Rhabdoviridae and Bornaviridae families.

31. A computer-readable medium with high resolution 3D structure coordinate information for an NNS L protein stored thereupon.

EXAMPLES

Example 1

Protein expression and purification

Protein expression and purification were carried out by insect expression of VSV-L, using baculovirus vector created with pFastBac Dual (Invitrogen), placing L under the control of the polyhedrin promoter and green fluorescent protein (GFP) under control of the P10 promoter (to visualize expression). Sf21 cells were infected, incubated at 27 °C for 60-72 hours, and harvested as cell pellets by centrifugation. Following lysis by sonication and removal of cell debris by centrifugation, the L protein was purified by Ni-nitrilotriacetic acid (NTA) chromatography followed by Hi-Trap S

and size exclusion chromatography. VSV P, residues 35-106 with an N-terminal 6xHis-tag followed by a tobacco-etch virus (TEV) protease recognition motif, was expressed in Rosetta BL21 (DE3) E. coli cells grown in LB medium containing 100 µg/mL ampicillin. Protein expression with 0.8 mM IPTG at optical density 0.8, and incubated overnight at 18 °C. The P(35-106) fragment was first purified by Ni-NTA agarose chromatography followed by tag removal by incubating with TEV protease overnight at 4 °C. A second round of Ni-NTA chromatography separated cleaved P(35-106) from uncleaved product. The cleaved proteins were dialyzed against 25 mM Tris pH7.4, 250 mM NaCl, 1 mM DTT. Purified VSV-L and P(35-106) were incubated overnight at 4 °C in a molar ratio of 1:4, and the complex was isolated on a Superdex 200 gel filtration column in 25 mM HEPES pH 7.4, 250 mM NaCl, 6 mM MgSO₄, 0.5 mM TCEP.

These methods can be applied to any NNS virus L protein.

Electron microscopy

The purified VSV-L:P complex was examined negatively-stained samples on a Philips CM10 electron microscope (EM). For cryo preparation, 3.5 µL of protein was applied at ~0.35 mg/mL to a Quantifoil R1.2/1.3 Cu grid (400 mesh) (Quantifoil, Germany) that had been glow discharged at 40 mA for 30 s. Grids were plunge-frozen with an FEI Vitrobot Mark I, with the following settings: 65 % humidity, offset -3, blot time 2 s, drain time 1 s. Images were recorded with liquid-nitrogen cooling on a Tecnai F20 EM (FEI) with a CT3500 cryo-specimen holder (Gatan); the microscope was operated at 200 kV; the defocus range was 0.9-2.3 µm. A semi-automated acquisition program, UCSFImage4 to record movies with a K2 Summit direct detector (Gatan), operated in superresolution mode with dose fractionation. The nominal magnification was 29,000x, corresponding to a calibrated magnification of 40,410x on the sensor plane of the camera. The beam intensity was set to 8 e/pixel/s. During a 6 s exposure, 30 frames of 200 ms each were collected for a total electron dose of 31 e/Å². Frames were binned over 2x2 pixels, yielding a pixel size of 1.24 Å, and aligned to each other using the program dosefgpu_driftcorr.

Image processing

A total of 356,611 particles were picked by hand from 6x binned images from 1272 movies. Two-dimensional classification with multivariate statistical analysis (MSA) was carried out in IMAGIC and K-means classification in TIGRIS (<http://tigris.sourceforge.net>). Image defocus was determined with CTFFIND3 and class averages were calculated with full contrast transfer function (CTF) correction. Good class sums were selected as references for particle alignment and subsequent re-classification, iterating twice. An initial model, calculated with EMAN2 (e2initialmodel.py) used 292 class averages. Refinement and three-dimensional classification (3 classes) in FREALIGN of the 292 CTF-corrected class averages resulted in one class with about 10 Å resolution, which was used as an initial reference for refinement and classification of the full particle stack.

FREALIGN was also used for refinement and three-dimensional classification (3 classes) of the full-dose particle stack, initially using 3x binned images (high resolution limit 10 Å). After 11 cycles of refinement and classification, the computation switched to unbinned particles. After 160 cycles, the resolution had gradually extended to 7 Å, at which point the best set of 155,443 particles were extracted. A further 100 cycles of refinement and classification (3 classes) extended the resolution to 6 Å. 74,940 particles were then extracted with the best scores from the two best classes. A final 7 cycles of refinement of angles and shifts (using a 6 Å reference model) alternated between the full-dose (31 e/Å²) stack and a low-dose (12 e/Å²) stack. The final map (3.8 Å resolution at FSC=0.143 criterion) was calculated from the low-dose images. A similar approach can permit yet further refinement in resolution.

Model building

The polypeptide chain of the RdRp, ring-like domain was traced using the programs O and Coot. Coot was used to place standard poly-alanine α -helices into evident helical density features, which were usually well enough defined to determine polarity, and connected the helices with poly-alanine loops, following strong density. The reovirus and rotavirus polymerases (λ 3 and VP1, respectively) were used as connectivity guides, having established correspondence of helical segments over a span of about 600 total residues. The methyltransferase domain was built following a similar strategy, guided by the consensus fold of S-adenosylmethionine (SAM)-dependent transferases. The capping domain, connector domain, and C-terminal domain have no known homologs; the initial models therefore were built depending on density, confirming and adjusting connectivity subsequently by reference to the amino-acid sequence. Side-chain density was strong enough in secondary structure elements of each domain to establish the sequence register. Secondary structure prediction (www.predictprotein.org) helped locate principal helices and strands. The entire structure was checked and corrected with O, using the lego-loop provision to rebuild many of intersecondary-structure loops and adjusting side-chain torsion angles to fit density. For the following segments of the capping and connector domains, the density did not allow confident assignment of backbone stereochemistry, and C α positions will have larger errors than in the rest of the model: 1159-1171; 1210-1226; 1308-1334; 1512-1518; 1534-1541.

The structure coordinate information for the VSV L protein determined as described herein is deposited with the RCSB Protein Data Bank with Reference Number 5A22, and is accessible on the World Wide Web at rcsb.org/pdb/explore.do?structureId=5A22. See also Liang et al., Cell 162: 314-327, 2015, which provides structure figures with details in color; the reference is incorporated herein by reference. The 2109 amino acid VSV L protein sequence is available at GenBank Accession No. P03523 (Indiana strain) and P16379 (New Jersey strain). L protein sequences of other NNS viruses

are readily available. Examples include, but are not limited to Rabies virus L protein (Accession No. P16289), Synochus yellow net virus (Accession No. P31332), Measles virus (Accession No. P12576), Sendai virus (Accession No. Q06996), human Parainfluenza virus 2 (Accession No. P26676), human Parainfluenza virus 3 (Accession No. P12577), human Parainfluenza virus 5 (Accession No. Q03396), Newcastle disease virus (Accession No. P11205), human Respiratory syncytial virus (Accession No. P28887), Ebola virus Zaire (Q05318), Marburg virus (Accession No. AAR85466.1), Cuevavirus (Accession No. YP_004928143), Nipah virus (Accession No. AEZ01398.1), Hendra virus (Accession No. NP_047113.3), Lyssavirus (Accession No. YP_001285397.1), Perhabdovirus (Accession No. YP_007641367.1), Eel virus (Accession No. AHD46130.1), Sigmavirus (Accession No. ACU65445.1), Ephemerovirus (Accession No. YP_009094434.1), Tibrovirus (Accession No. YP_007641376.1), Tupavirus (Accession No. YP_238534.1) and Sprivivirus (Accession No. ABW24037.1). The sequences for the L proteins of other strains of the various NNS viruses are also readily available in, e.g., GenBank.

Structure refinement

The density map was fine sampled on a grid with 0.72 Å spacing and transferred the density (and model) into a P1 cell ($a=112$ Å, $b=143$ Å, $c=106$ Å, with 90° angles) using MAPROT from the CCP4 suite. This map was further solvent flattened (FIG. 2) by calculating a mask around the model with a probe radius of 3.9 Å and setting grid points outside the masked region to a constant value corresponding to 0.33 e/Å³. Density within the mask was the set to an absolute scale by determining a scale factor assuming 0.33 e/Å³ and 0.43 e/Å³, respectively for solvent and protein within the mask (determining the "dry" protein volume from its mass and partial specific volume; the "hydration" calculated this way is about 0.3 w/w). Map and mask operations were carried out with MAPMAN. The amplitudes (FP) and phases (PHIO) were calculated from the solvent-flattened map. Although amplitude standard deviations (SIGFP) was not used in any of our calculations, dummy values were supplied (SIGFP = 0.1 FP) to satisfy input requirements of certain programs. 4 % of the structure factors was flagged as a cross validation set for calculating R_{free}. Figures of merit (FOM) were estimated from the phase angle difference between the two half-set reconstructions and calculated Hendrickson-Lattman coefficients from PHIO and FOM. The structure was refined against amplitudes and phases, including data to a minimum Bragg spacing of 3.8 Å. Procedures in PHENIX were applied, using a protocol with several rounds of individually restrained positional and B-factor refinement, including one round of torsion-angle simulated annealing and real-space refinement. Appropriate weights were determined for the experimental terms in the target function by monitoring R_{free}, model geometry, and B-factor statistics. Secondary structure restraints were applied throughout the refinement and Ramachandran restraints in the final round. The final model was analyzed with MolProbity. Refinement and model statistics are shown in **Table 1** below.

Table 1.

Refinement	
Space group	P1
Cell dimensions $\text{\AA}$	
a, b, c ($\text{\AA}$)	112.0, 143.0, 106.0
α, β, γ ($^\circ$)	90.0, 90.0, 90.0
Asymmetric unit composition	
Number of residues	2004
Non-hydrogen atoms	16077
Resolution ($\text{\AA}$)	143.1 – 3.80 (3.87 – 3.80)
Number of reflections (work / free)	62328 (2490) / 3570 (140)
$R_{\text{work}} / R_{\text{free}}$ (%)	28.2 / 29.6 (85.8 / 78.3)
$CC_{\text{work}} / CC_{\text{free}}$	0.88 / 0.86 (0.18 / 0.05)
Phase angle difference ($^\circ$)	33.2 (73.0)
Wilson B factor ($\text{\AA}^2$)	60.2
Model statistics	
B factors	
Average $\langle B \rangle$ ($\text{\AA}^2$)	112.4
$\langle B_i - B_j \rangle$ ($\text{\AA}^2$)	25.3
R.m.s deviations	
Bond lengths ($\text{\AA}$)	0.009
Bond angles ($^\circ$)	1.547
Ramachandran angles	84.1 % / 13.9 % / 2.0 % (favored / allowed / outliers)
MolProbity clash score $^\circ$	13.05, 97 th percentile (N = 37, 3.00 – 999.00 $\text{\AA}$)
MolProbity score $^\circ$	2.79, 93 rd percentile (N = 342, 3.25 – 4.20 $\text{\AA}$)

^a Highest resolution shell is shown in parenthesis.

^b $CC = \Sigma(F_{\text{map}} F_{\text{model}}) / (\Sigma(F_{\text{map}}^2) \Sigma(F_{\text{model}}^2))^{1/2}$, correlation between experimental and model structure factors for working and test set, respectively.

^c 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

Example 2

High resolution structure determination

Images from VSV-L bound with P(35–106) were recorded from grids prepared for cryo-EM as described in Experimental Procedures. All data were taken on an FEI F20 microscope with a Gatan K2 Summit detector. Following two-dimensional classification with IMAGIC (van Heel et al., 1996) and TIGRIS (<http://tigris.sourceforge.net>) and calculation of an initial three dimensional reference density with EMAN2 (Tang et al., 2007), FREALIGN was used for refinement and three-dimensional classification. FIG.1 shows images, summarizes stages of the analysis, and illustrates the final 3.8 $\text{\AA}^\circ$ resolution density map.

Domain Organization of VSV-L

The polypeptide chain from residue 35 to the carboxy terminus, residue 2109 was traced, leaving out poorly ordered linker segments 1335–1357 and 1558–1597 and a short, poorly ordered loop 1840–1849. Some segments of ~10–15 residues between 1100 and 1334 and between 1358 and 1557 are in poor density, and the chain trace in those regions is approximate. Because the beginning and end of the segment and “puddles” of density leave little uncertainty about the overall course of the segment in question, those residues are kept in the model.

Considerable attention was paid to correct stereochemistry during model building, but to adjust both the fit and the stereochemistry beyond the capacity of visual inspection, one round of refinement was carried out, by calculating structure factors from the map and using both amplitudes and phases in the target function, with secondary-structure hydrogen bonds imposed as restraints. After some minor adjustments and one more cycle, *R*_{free} and *R*_{work} were 29.6 and 26.2, respectively (FIG 2; Table 1). The residues assigned to domains and linkers as follows: RdRp, 35–865; capping domain, 866–1334; linker 1 1335–1357; connector domain, 1358–1557; linker 2, 1558–1597; methyltransferase, 1598–1892; C-terminal domain, 1893–2109 (FIG 2, FIG 3, FIG4 ; Table 1). The boundaries between the capping domain and the connector domain and between the latter and the methyltransferase are evident from the disordered linkers that intervene. The boundary between the RdRp and the capping domain corresponds to a previously identified tryptic cleavage site. Although the interface between the two domains is relatively extensive and well packed, fragment 1–860 can be expressed independently (Rahmeh et al., 2010). The boundary between the connector domain and the methyltransferase domain also corresponds to two independently stable fragments: 1–1593 and 1594–2109. Between the methyltransferase domain and the C-terminal domain is an extended but relatively polar interface. Negative stain electron microscopy of a fragment that includes residues 1594–2109 shows that stain can penetrate between the two domains and that their connection might be flexible (Rahmeh et al., 2010). There are well defined secondary-structure features in the density for each of the domains, leading us to believe that poor density for certain segments reflects local disorder, rather than failure of three-dimensional classification algorithms to detect overall variation in position or orientation of any particular domain with respect to the others.

RdRp

The RdRp has at its core a right-hand, “fingers-palm-thumb” structure (residues 360–865) common to a very large group of RNA and DNA polymerases (FIG 5). The catalytic site on the palm is in a deep channel between the fingers and thumb subdomains (extended from the palm as if in a loose hand grip). Appended to the core on the N-terminal side is a globular region (residues 1–359) that closes the channel on one end and reinforces the relatively slender thumb subdomain. From the appearance of VSV-L in negative-stain electron microscopy, and in particular from the size and staining of a “doughnut-like” part, the RdRp can be similar in cage-like structure to the dsRNA virus polymerases (Rahmeh et al., 2010). Those enzymes have their catalytic sites at the center of an enclosed cavity, connected to the exterior by four channels, for template entrance, template exit, transcript exit, and NTP access (Lu et al., 2008; Tao et al., 2002). Comparison of the chain trace with their structures shows that this suggestion was correct, with one modification. The dsRNA virus RdRps have a C-terminal “bracelet” domain that encircles the exit path for the template and includes a site for binding the methyl G cap on the non-template, plus-sense strand (Lu et al., 2008; Tao et al., 2002). In VSV L, the capping domain, which has no structural similarity to the bracelet domain of the

dsRNA virus RdRps, occupies the corresponding space. That is, residue 865, taken as the end of the RdRp, is at the C terminus of the thumb.

The positions of secondary structural elements were compared within VSV L, reovirus $\lambda 3$ (Tao et al., 2002), rotavirus VP1 (Lu et al., 2008), and the heterotrimeric influenza virus polymerase (Pflug et al., 2014; Reich et al., 2014). The secondary structural elements with correspondences in the three other polymerases extend from about residue 107 in VSV-L to the end of the RdRp domain (FIG 5). The analogous parts of reovirus $\lambda 3$ encompass residues 150–860 (approximately); those of rotavirus VP1, residues 135–750; those of human influenza virus B polymerase, residues 415 to the C terminus (714) of the PA subunit, and residues 8–586 of the PB1 subunit. The homology thus extends from the middle of PA into PB1. The region in common between VSV-L and influenza virus PB1 corresponds to the fingers-palm-thumb core structure, and the region shared with PA is a large part of the RdRp N-terminal domain.

Capping Domain

Unlike the corresponding host-cell process, the capping reaction of NNS RNA viruses proceeds from a covalent linkage between the 5' end of the RNA and a histidine residue, with attack on that linkage by a guanosine nucleotide. The enzyme is thus a polyribonucleotidyl transferase (PRNTase) rather than a guanylyl transferase. Two conserved motifs—GxxT and HR, separated by ~70 residues—mark the catalytic site (FIG 6A and 6B). The former participates in guanosine nucleotide binding; the latter is the site of covalent RNA attachment. The domain has no structural homologs that could be detected with standard search methods. The largely α -helical, N-terminal half (residues 866–1100), which abuts the polymerase domain, is well ordered. The C-terminal half (1100–1334) has several poorly defined segments, including the loop that bears the HR sequence. Despite uncertain definition of side chains, the separation of the two conserved sites is 10 Å, appropriate if a GTP bound at the former is to attack the histidine-liganded 5' phosphate of the nascent RNA at the latter (FIG 6B). Positions corresponding to sites in human respiratory syncytial virus (hRSV) of resistance mutations to a small-molecule capping inhibitor (Liuzzi et al., 2005) impinge on the active site from three sides (FIG 6A); their locations, and the relatively poor definition of the active site in the map, suggest that activation of the domain, perhaps by binding the 5' end of the nascent message, induces a conformational rearrangement, similar to the domain closures seen in many enzymes when they bind their substrate. Two candidate Zn sites, one with clear density where two Cys (residues 1120 and 1123) and two His (1294 and 1296) ligand the likely Zn ion, and one with three Cys (residues 1081, 1299, and 1302) and a Glu (1108), contribute structural integrity to the capping domain. The sites are close to each other and well outside the catalytic center. In both cases, the liganding residues are present as a conserved set in most NNS virus L proteins and absent as a set in the others.

A loop between residues 1157 (the threonine of the GxxT motif) and 1173 projects back into the cavity of the polymerase domain (FIG 5B, 6C, and 7). The poorly ordered tip of this loop occupies the same position as the priming loop in the reovirus polymerase (Tao et al., 2002). The loop in VSV-L polymerase domain that corresponds to the I3 priming loop is shorter than its reovirus homolog, and the capping-domain loop projects over it. Neither polymerase requires a polynucleotide primer to initiate, and the priming loop in the reovirus polymerase supports the initiating nucleoside triphosphate. As elongation proceeds, the tip of the loop recedes to make room for the dsRNA replication product or for the short doublestranded region just upstream of the newly added nucleotide during transcription (Tao et al., 2002). This loop, which contacts the minor groove of the nascent product, can also enhance fidelity, by retarding elongation of mismatches detected by poor minor-groove geometry and allowing more time for ATPbased pyrophosphorolysis of the mismatch. The position of the likely priming loop of VSV-L on the capping domain, adjacent to the GxxT residues, suggests coupling of capping to initiation of polymerization.

Connector Domain

The connector domain is a bundle of eight helices (FIG 3B); it appears to have largely an organizational role in positioning or spacing the catalytic domains. Disordered linkers, 23 and 40 residues long, respectively, lead into and out of the connector domain. The endpoints of these linkers in well-defined density show that they must occupy an extended groove between the capping and connector domains; the groove also extends into the interface between the capping and methyltransferase domains

(FIG 8). Strong, low resolution density features fill this groove, but they are not sharp enough to suggest particular linker conformations (FIG 8B). The location of P indicated by negative-stain electron microscopy (FIG 8A and 9; Table 2) leads us to suggest that the groove also holds some or all of the P fragment present in the L-P complex imaged. Because P(35–106) locks the smaller domains of L into a fixed configuration, it is plausible that it might do so by stabilizing folded structures for the two linker segments—gluing them down, so to speak, alongside the connector domain (FIG. 8B).

Table 2

	Cross correlation coefficient				Z score ^b	3D-plot
	Max	Mean	Min	Sigma		
VSV-L : P(41-106)						
Class 1	0.88	0.78	0.65	0.05	1.89	Figure S4A
Class 2	0.93	0.88	0.85	0.02	2.32	
Class 3	0.88	0.80	0.67	0.06	1.67	
Class 4	0.88	0.83	0.78	0.02	2.50	
VSV-L : P						
Class 1	0.78	0.70	0.59	0.04	1.81	Figure S4C
Class 2	0.75	0.61	0.47	0.07	1.88	Figure S4D
VSV-L(35-860)						
Class 1	0.87	0.80	0.73	0.03	2.59	Figure S4B
Class 2	0.92	0.84	0.74	0.04	1.84	
VSV-L(35-1114)						
Class 1	0.90	0.86	0.82	0.02	2.59	Figure S4B
Class 2	0.90	0.88	0.80	0.02	2.05	
VSV-L(35-1557)						
Class 1	0.73	0.67	0.60	0.02	2.46	Figure S4B
Class 2	0.73	0.64	0.57	0.04	2.37	
VSV-L(1598-2109)						
Class 1	0.83	0.72	0.58	0.07	1.42	Figure S4B
Class 2	0.72	0.62	0.52	0.05	1.95	

^a Cross correlation coefficients were calculated for all the class averages from negative stain data shown in Figure 5, pairing each to 799 reference images (all possible orientations sampled at a 5° angular grid) of 2D projections from the model 3D volume. Then the mean, maximum, minimum and standard deviation were calculated.

^b Z score is defined as number of standard deviations (Sigma) of Max above Mean.

Methyltransferase Domain

The methyltransferase domain has the structure characteristic of many other domains that catalyze transfer of a methyl group from S-adenosyl methionine (FIG. 10 and FIG. 11). It methylates both the ribose O20 and the guanosine N7 (Rahmeh et al., 2009). Most of the domain superposes extremely well on the flavivirus methyltransferases, also dual specificity enzymes (Egloff et al., 2002; Ray et al., 2006; Zhou et al., 2007). Evidence for functional feedback from the VSV-L methyltransferase to the RdRp comes from the observation that addition of S-adenosyl homocysteine, which inhibits methylation, leads to hyperpolyadenylation; mutations that prevent methyl transfer have a similar effect (Galloway and Wertz, 2008; Li et al., 2009; Rose et al., 1977). The methyltransferase domain contacts both the connector and the capping domains, but it has no direct contact with the RdRp. Moreover, there is no obvious “tunnel” that would allow the 5' end of the transcript to move from the catalytic site of the capping domain to the catalytic site of the methyltransferase domain. Therefore, the L protein undergoes a substantial conformation change following initiation of polymerization and that the inter-domain communication shown in this structure is relevant to formation of the first one or two phosphodiester bonds, but not to subsequent elongation and 5'-end modification.

C-Terminal Domain

Like the connector domain, the C-terminal domain, which terminates in a ~25-residue long C-terminal “arm,” appears to have an essentially organizational role (FIG 3B). It is largely an α -helical bundle, but a projecting, almost beak-like, β -hairpin supported by a second interhelical loop, imparts a

noticeable asymmetry. The C-terminal arm, a feature that appears from sequence alignments to be conserved among NNS viral polymerases, but variable in length, extends back against the RdRp, augmenting the b-hairpin that bears the catalytic Asp-Asn sequence at its tip, and terminates at the three-way junction of the capping, connector, and methyltransferase domains, where it has one or more contacts with each. The arm thus contributes to closing the multi-domain structure seen in the L-P complex, further stabilized by the phosphoprotein, P.

Template Channel

Superpositions of related positions in reovirus $\lambda 3$ (Tao et al., 2002) and rotavirus VP1 (Lu et al., 2008) have allowed us to model a bound template and a template-primer-NTP complex, because I3 was catalytically active in the crystals studied used to determine the structure and VP1 in its crystals incorporated template in a sequence-specific register (Tao et al., 2002). The template entrance channel is at the interface between the RdRp and the capping domain (FIG 3C). Polar and especially basic residues project into the groove from both sides. As in all polymerases of this family, the template runs across a “fingers loop” (residues 523–545 in VSV-L) and twists sharply to present the templating base to the catalytic center. A hydrophobic residue in the loop (Phe541 in VSV-L) bears on the templating base to enforce correct base pairing with the incoming nucleoside triphosphate. For initiation at the 3' end of the viral RNA (either for replication or for transcription of leader RNA), the priming nucleoside triphosphate will rest against the loop from the capping domain described above (FIG 5B). Any further elongation, after forming the initial phosphodiester bond, will require this loop to move, and substantial elongation will almost certainly require displacement of the entire capping domain (FIG 6D). To accommodate transcriptional elongation, the entire array of smaller domains can reorganize.

Domain Reorganization

The configurations of VSV-L characterized herein by negative-stain electron microscopy illustrate the potential for large-scale domain reorganization (Rahmeh et al., 2010). Images of L alone show a core “doughnut,” which admits stain at its center, three globular appendages, in apparently variable positions and orientations with respect to each other and to the core. Addition of P, or of the peptide, residues 35–106, used to stabilize the complex studied here, locks the appendages in place (Rahmeh et al., 2010, 2012). Many of the projections of this locked structure resemble a FIG “10.” Class averages from these images agree extremely well with projections of the structure described (FIG 8A), as do class averages from images of four different L fragments (FIG 8C). One of these (1–860) corresponds precisely to the RdRp. Another, 1–1121, includes the RdRp and the largely helical, N-terminal half of the capping domain. The tryptic cleavage that initially generated that fragment is in a surface loop. The last fragment previously imaged by negative staining comprises the methyltransferase and C-terminal domains (FIG 8C). These comparisons show that previous

assignment of the three globular appendages to the capping domain, the methyltransferase domain, and one unassigned domain (Rahmeh et al., 2010) should be modified. The capping domain is part of the doughnut, and the appendages correspond to the connector, methyltransferase, and C-terminal domains, respectively (FIG 8D). The linkers between the capping domain and the connector and between the connector and the methyltransferase clearly allow the latter two to move away from the rest of the molecule; good definition in negative stain for the third globular appendage suggests that in the unlocked structure, the C-terminal arm also pulls away from the RdRp. Many of its interactions, as it inserts back against the rest of the molecule in the structure determined herein, are indeed with the connector and methyltransferase domains. Images of negatively stained complexes of L with intact, dimeric P often show two, linked, FIG-“10” L molecules, but occasionally the P dimer does not recruit a second L and appears as a surface feature on the hook of the “10” (FIG 8A). Comparison of the L structure with these projections is consistent with our proposal that the interaction with P(35–106) that stabilizes the “10” conformation is with the linker segments at either end of the connector domain (FIG 8B).

Cryo-electron microscopy

High-resolution cryo-EM structure determination has until recently relied on either high symmetry or large size—for example, icosahedral viruses, which have both, or ribosomes, which are large enough to produce reasonable contrast for getting started with iterative determination of particle orientations and centers (Grigorieff and Harrison, 2011). The molecular structure of an asymmetric protein of total mass <250 kDa. Dose fractionation (“movies”) has been determined herein, permitted by use of a direct electron detector, and refinement and maximum-likelihood classification procedures (Lyumkis et al., 2013), implemented in FREALIGN, facilitated a resolution adequate to build an atomic model.

Sequential Transcription

A de novo initiation event with ATP as initiating nucleotide appears to start synthesis of each mRNA transcript. The structure is interpreted as that of an early initiation state, representing an L-P complex ready for loading onto the end of the template to synthesize leader RNA. During the transition to elongation the priming loop—contributed by the capping domain—must shift out of the way to accommodate the product. Inspection further suggests that after addition of only a few more nucleotides, the capping domain as a whole must withdraw from tight contact with the RdRp to allow further elongation, as there do not appear to be clear exit channels for transcript and template. Upon termination of a transcript, the polymerase reinitiates on the next gene, but the efficiency of producing the succeeding transcript is only ~70% (Iverson and Rose, 1981). The template entrance channel in VSV-L is at the interface of the capping domain and the RdRp, and dissociation of the template will be straightforward (when initiation or early elongation aborts) if that interface opens as suggested. Otherwise, the entire template would have to thread through the active site and emerge through

another channel. Transcription of the downstream gene probably requires reestablishing the inter-domain contact seen in our structure, so that the priming loop can reinsert into the active site of the RdRp domain for subsequent de novo initiation. Inhibition at any one of these rearrangements could be effective to inhibit protein activity. That is, the identification of these necessary domain movements identifies new targets for inhibition. Such targets have the benefit that they are likely to be unique to the NNS viruses or even to a specific virus examined, preventing likelihood of side effects.

Coupling of Capping, Polymerase, and Methyltransferase Activities

Displacement of the capping domain from the RdRp as elongation proceeds might have two consequences. First, the active site of the PRNTase can reorganize (e.g., by “domain closure”) into a better ordered configuration than the one seen in the present structure. Second, because the capping domain faces both the connector and methyltransferase domains, its displacement might also induce rearrangement of the rest of the capping machinery. A large-scale reorganization of this kind could account for some of the observed functional crosstalk between the capping and polymerase activities.

A cap is added only when the length of a transcript has reached 31 nucleotides (Tekes et al., 2011). *In vitro*, very short (up to 5 nt) transcripts can be capped in trans by L, but this process is inefficient and fails completely with longer transcripts (Li et al., 2008; Ogino and Banerjee, 2007). Conversely, mutations in L that disrupt cap addition cause premature termination (Li et al., 2008, 2009). Mutations in the specific, cis-acting signals at the 5' end of the nascent strand, which are absent from the leader RNA, also block cap addition and result in premature termination of that transcript (Li et al., 2008; Ogino and Banerjee, 2007; Stillman and Whitt, 1997; Wang et al., 2007). The precision of the 31-nt requirement suggests that the reorganized structure that allows elongation is a well-defined state, rather than a loosely ordered one.

Reorganization of the capping machinery can also account for why mRNA cap methylation requires no additional chain length (Tekes et al., 2011). In the configuration represented by the structure determined herein, the catalytic sites for capping and methylation are distant from each other. If the smaller domains move away from the polymerase core, the capped, nascent RNA could probably release from the capping enzyme and gain immediate access to the methylase domain. Methylation in trans can occur under some circumstances, but previous work has shown that transcripts stalled at a chain length of 31 nt are fully methylated—presumably in cis—by the stalled L (Tekes et al., 2011).

Inhibition of mRNA cap methylation by high concentrations of S-adenosyl homocysteine can result in hyper-polyadenylation, demonstrating a linkage between the methylase and RdRp domains (Galloway and Wertz, 2008; Li et al., 2009; Rose et al., 1977). The RdRp domain of L carries out polyadenylation by iterative transcription of a gene-end U tract element. Some, but not all, mutations

that inhibit methylation result in the hyper- polyadenylation phenotype (Galloway and Wertz, 2008), indicating that the crosstalk mechanism is not a readout of cap modification but probably a consequence of interactions between domains and between the protein and the nascent transcript.

The cap methylase of VSV participates in both ribose 2'O and guanine-N7 methylation reactions. The preferred substrate for all other ribose 2'O methyltransferases is 7mGpppN and like other proteins that recognize the mRNA cap structure—such as eIF4E-2'O methylases—those enzymes position the ribose in the active site by p-p stacking interactions with the 7mG RNA. The order of cap methylation in VSV is reversed. Methylation of 2'O precedes and facilitates subsequent methylation of guanine-N7. The absence of aromatic residues that could participate in such interactions with a 7mGpppN RNA in the VSV methyltransferase is consistent with this altered reaction sequence.

The N Protein

The template for polymerase is not naked RNA, but a complex in which the template RNA is encased within the nucleocapsid protein sheath. In that complex the RNA bases are not accessible to the RdRp of L, and the N protein must transiently dissociate from the RNA for the RdRp to proceed (Green et al., 2006). The structure of L allows us to estimate that 20–25 nt of the template strand are threaded through the polymerase domain. Accordingly, because each molecule of N covers 9 nt of RNA, two or three molecules of N must be displaced from the template strand at any one time. Adjacent N subunits in the RNP interact stably, embracing each other through N- and C-terminal extensions (Green et al., 2006). Thus, looping out of template RNA need not entail dissociation of N from the RNP coil. Indeed, if the linked chain of N subunits is considered as the analog of a cRNA strand, then the displaced N is the counterpart of a looped-out plus-sense strand during transcription by the related polymerases of dsRNA viruses. This N-protein bridge could account for the precision of the 31-nt length of nascent transcript required for cap addition, perhaps by creating a defined spacing between the RdRp and the popped-out capping domain. The N protein influences L activity, as recognition of the cis-acting signals in the genome requires it and as its presence influences incorporation by L of substituted nucleotide analogs (Morin and Whelan, 2014). It can also be necessary for capping.

The P Protein

P is an adaptor that engages both the N-RNA template complex and the L protein. A small, globular domain at the C-terminal end of P (residues 195–265) interacts with the N-RNA complex (Green and Luo, 2009). This domain could in principle move from one subunit to the next as polymerization proceeds. The structure described herein contains only part of the N-terminal region of P. Although it is poorly ordered in the density map, the P(35–106) can occupy some of the strong, low resolution density features between the capping and connector domains, locking in the linker segments at both

ends of the latter. Depending on the chain polarity and on the flexibility of intervening segments, the C-terminal domain of P could lie near the opening through which RNA enters the active site of the RdRp and thus, through its interaction with N, be part of the process that feeds template rapidly through the polymerase channel.

Homologies and Comparisons

Homologous L proteins include those of rabies, Ebola, measles, and respiratory syncytial viruses. Alignment of their sequences (FIG. 2) shows the same overall arrangement of the various domains, identifies the active site residues of the protein for RdRp, PRNTase, and methyltransferase activities and suggests domain boundaries for expressing various fragments of the proteins from VSV and related viruses. All NNS RNA viruses have a polymerase complex that comprises the enzymatic subunit, L, and an equivalent of the VSV phosphoprotein, P. In some cases, additional viral proteins (VP24 in the case of the filoviruses, M2-1 in the case of respiratory syncytial virus) are necessary for full polymerase processivity. The three-dimensional interconnections among domains in the VSV polymerase suggest that binding of these accessory proteins to any of the smaller domains might influence large-scale rearrangements and affect enzyme processivity.

Parts of L are structurally similar to the heterotrimeric polymerase of influenza virus. The major differences are the capping machineries, reflecting the distinction between influenza-virus cap-snatching and NNS RNA-virus cap synthesis (Reich et al., 2014). A structural feature of the influenza virus polymerase that is also absent from the VSV polymerase is the long PB1 arm, which is probably required for a step of mRNA transcription (Pflug et al., 2014; Reich et al., 2014).

The VSV-L structure allows preliminary interpretation of mechanisms for inhibitors of its homologs. For measles virus, resistance to a non-nucleoside analog inhibitor that blocks gene expression maps to the polymerase domain (Krumm et al., 2014). The positions of those mutations, mapped onto VSV-L, flank the GDNQ motif at the RdRp catalytic site, suggesting an allosteric mechanism, like that of the non-nucleoside analog reverse transcriptase inhibitors for HIV. A compound active against RSV targets the capping domain. The locations of positions in VSV-L that correspond to sites of resistance mutations in RSV-L are consistent with our proposal that a domain closure accompanies activation of the capping activity after polymerization has commenced (Liuzzi et al., 2005).

Example 3

Extension of in vitro L protein assay to Measles virus

Recombinant Measles virus L-P complex was expressed as described above for VSV L protein with some modifications. Briefly, 7xHis tagged L and P were co-expressed in *Spodoptera fugiperda* 21 (Sf21) cells and affinity purified with Ni-NTA agarose (Qiagen) followed by size exclusion

chromatography (Superdex 200 HR 10/30, GE healthcare) for the WT protein, in a buffer containing 50 mM Hepes pH 7.4, 400 mM NaCl, 1 mM DTT and 10% glycerol.

Development of an in vitro RNA synthesis assay for MV

To develop an in vitro assay for measles virus RNA dependent RNA polymerase the L and P proteins of the Khartoum Sudan (KS) strain of measles virus were first expressed in insect cells from a single recombinant baculovirus. The resulting L-P complex was purified by affinity for Ni-NTA agarose and subsequent size exclusion chromatography (FIG 13A). A portion of the L-P complex co-purified with an additional protein which identified by mass-spectrometry as heat shock protein (Hsp) 90 (FIG 13, lane 2). To determine whether the purified polymerase retains RNA dependent RNA polymerase activity a simple *in vitro* assay using as template a chemically synthesized naked RNA corresponding to the first 19-nt of the measles virus genome was developed. The L-P, and the L-P-Hsp90 complexes exhibit RdRP activity in this assay (FIG 13B), and inclusion of the Hsp90 inhibitor geldanamycin had no effect on this activity (FIG 13C). These results demonstrate that the purified polymerase complexes are equally active in vitro and that Hsp90 is not required for the RdRP activity of the complex.

In vitro Polymerase assays for measles virus L protein were carried by using 1.4 μ M of template with 0.1 μ M of MV L-P in a reaction mixture containing 20 mM Tris-base pH 8, 10 mM NaCl, 6 mM MgCl₂, 200 μ M UTP, 1.5 mM ATP, 1.5 mM CTP and 165 nM of [α 32P]-GTP (3000 Ci/mmol). Reactions were incubated at 30°C for 3 hours, and stopped by addition of EDTA/formamide. Reactions products were resolved on a denaturing polyacrylamide gel electrophoresis (20% polyacrylamide, 7 M urea) in TBE buffer, and analyzed by autoradiography. The sizes of the products were determined by comparison with a 19 nt marker RNA labeled by T4 polynucleotide kinase (New England Biolabs) using [γ 32P]-ATP (3000 Ci/mmol). Assays studying the inhibitory effects of As136a and ERDRP0519 were carried out using the indicated concentrations of the inhibitor. All radioisotopes were purchased from PerkinElmer.

Modeling resistance mutations

Full-length L proteins of MV (GenBank: ADO17332.1), CDV (GenBank: ACZ56434.1) and VSV (UniProtKB/Swiss-Prot: P03523.2) have been aligned using T-Coffee and shown using Esript and location of mutations were identified and position on the VSV L structure (PDB: 5A22) using Pymol. For the modeling of ERDRP0519 into the binding domain of MV L, ERDRP0519 has been drawn using Chemdraw and modeled using PHENIX ref). The polypeptide chain of MV binding domain has been traced with the MV L sequence (GenBank: ADO17332.1) and the VSV L map (PDB: 5A22) using Coot.

Effects of a non-nucleoside inhibitor and a more bioavailable derivative on RNA synthesis activity by MV L-P

Using cell-based luciferase replicon reporter assays for MV and canine distemper virus (CDV) the non-nucleoside inhibitor, 1-Methyl-N-[4-(1-piperidinylsulfonyl)phenyl]-3-(trifluoromethyl)-1H-pyrazole-5-carboxamide (As136a) and a more bioavailable derivative (S)-1-methyl-N-(4-((2-(2-morpholinoethyl)piperidin-1-yl)sulfonyl)phenyl)-3-(trifluoromethyl)-1H-pyrazole-5-carboxamide (ERDRP0519), were shown to inhibit viral gene expression. Selection of viral mutants resistant to those inhibitors identified mutations in the L gene have been described {Krumm, 2014 #281; Ndungu, 2012 #280}. To directly demonstrate that the non-nucleoside inhibits RNA synthesis their effect was tested on RdRP activity *in vitro*. The inhibitors block RNA synthesis initiation by L-P with an IC₅₀ of 70 nM for As136a and 100 nM for ERDRP0519 (FIG. 14) as demonstrated by the uniform inhibition of all of the products of RNA synthesis. This result shows that the non-nucleoside inhibitors inhibit the MV RdRP.

Effect of L gene mutations that lead to resistance to the non-nucleoside inhibitor

Resistance mutations were previously shown to map to the RdRp domain of L {Krumm, 2014 #281}. As a first step to understand which mutations provide resistance we expressed and purified polymerase complexes carrying either the mutation L_{T776A} (L_{T776A}-P) or L_{T751I} (L_{T751I}-P). When compared to wild type L-P, each of the substitutions led to a significant difference in IC₅₀ of 1.3 μM for L_{T776A}-P and 110 μM for L_{T751I}-P (FIG. 16B, C, D)- a 10 and 1000 fold change. The effect of combined mutations on MV L by generating an L protein that carries both T776A and T751I (L_{T776A/T751I}-P) was studied. The combination of these two mutations on L confer a complete resistance to L_{T776A/T751I}-P as no inhibition is detected even in the presence of 2.5 mM of inhibitor (FIG. 16A). These results suggest that the inhibitors bind directly MV L in an area of L located close to its RNA synthesis active site.

Recombinant measles containing substitutions in L exhibits complete resistance to non-nucleoside inhibitor

To demonstrate that the two mutations lead to complete resistance of the measles virus to the non-nucleoside inhibitor a recombinant measles virus KS strain expressing eGFP was engineered (FIG. 17). The kinetics of replication of the engineered variant MV-L_{T776A/T751I} were reduced compared with the wild type parent as evidenced by the decreased viral yield at 24h post inoculation of cells (FIG. 17A) consistent with a fitness cost to the resistance mutations. The recombinant MV-L_{T776A/T751I} variant was resistant to the non-nucleoside inhibitor on viral replication as judged by eGFP expression in the presence of 25 μM of the bioavailable inhibitor. This result demonstrates that substitutions in measles L lead to complete resistance of the virus to the non-nucleoside inhibitor *in vitro* and in cells but uncover a fitness cost to viral replication associated with those substitutions.

Example 5*Application of findings for VSV L protein to Rabies virus*

Rabies virus (RABV) is a non-segmented negative-strand (NNS) RNA virus from the Rhabdoviridae family that generates lethal infection in the brain of humans and mammals causing an estimated 55000 deaths per year with 4 out of 10 deaths occurring in children under the age of 15. Although an efficient vaccine-based treatment exists, its poor access in rural regions of Asia and Africa suggests that the development of antiviral could be a possible alternative.

NNS RNA viruses share common replicative machinery that consists of a ribonucleoprotein (RNP) complex composed by the genomic RNA completely coated by the viral nucleoprotein (N) to form the N-RNA template for transcription and replication by the polymerase complex. This latter consists of the RNA-dependent RNA polymerase (RdRP) that resides within the 242-kDa large (L) protein and the essential 33-kDa co-factor phosphoprotein (P) that bridges the interaction between N-RNA and the RdRP. RABV P is a multi-functional protein composed of three domains, a C-terminal domain that binds the genomic N-RNA, a central domain responsible for auto-oligomerization and a N-terminal domain with a dual function of L-binding and N0-binding domain (with N0 being the free-RNA monomeric soluble form of N). Extensive studies of the vesicular stomatitis virus (VSV), another member of the rhabdovirus, have shown that L also carries two more enzymatic activities, a GDP:polyribonucleotidyltransferase that caps the newly synthesized mRNA, and a dual specificity mRNA cap methyltransferase. Moreover, electron microscopy revealed that L undergoes a conformational change via binding with P and renders the RdRP more active and processive.

The NNS RNA viruses minimal RNA synthesis machinery consists of L and a naked RNA has been demonstrated. Using purified recombinant VSV L and a synthetic RNA corresponding to the first 19 nt of the VSV genome, a new *in vitro* assay to study RNA synthesis for VSV has been developed. This assay became a powerful method for multiple studies of the replicative machinery: investigate the RNA synthesis initiation mechanism and requirements, study the effects of viral or other or cellular proteins on RNA synthesis and processivity, discovery and characterization of the mechanism of action of small molecule inhibitors. We sought an RNA synthesis *in vitro* assay for RABV for which no RNA synthesis assay exists.

Materials and methods*Protein expression and purification*

Recombinant L and P were expressed as described in (Rahmeh et al, 2010). Briefly, 6xHis tagged L was expressed in *Spodoptera fugiperda* 21 (Sf21) cells and affinity purified with Ni-NTA agarose (Qiagen) followed by ion exchange chromatography. 10xHis tagged P was expressed in *Escherichia*

coli BL21 (DE3) cells or Sf21 cells then affinity purified with Ni-NTA agarose (Qiagen) followed by gel filtration (Superdex 200 HR 10/30, GE healthcare).

In vitro VSV L polymerase assay. RNA templates were chemically synthesized and purified (Integrated DNA Technologies). Standard polymerase assays were carried out using 0.2 μ M of template with 0.2 μ M of VSV L in a reaction mixture containing 20 mM Tris-base, pH 8, 50 mM NaCl, 2 mM DTT and 0.5% (vol/vol) Triton X-100, 6 mM MgCl₂, 200 μ M UTP, 1.5 mM ATP, 1.5 mM CTP and 165 nM of [α 32P]-GTP (3000 Ci/mmol). Reactions were incubated at 30°C for 3 hours, and stopped by the addition of EDTA/formamide. Reaction products were resolved using denaturing polyacrylamide gel electrophoresis (20% polyacrylamide, 7 M urea) in TBE buffer, and analyzed by autoradiography. The sizes of the products were determined by comparison to a 19 nt marker RNA labeled by T4 polynucleotide kinase (New England Biolabs) using [γ 32P]-ATP (3000 Ci/mmol). Assays studying the effects of NTP concentration were carried out using the indicated concentrations. For monitoring the effect of varying GTP concentration reactions were supplemented with 165 nM of [α 32P]-ATP (3000 Ci/mmol). To determine the first nucleotide incorporated during initiation, standard reactions were carried out with 165 nM [γ 32P]-GTP (3000 Ci/mmol) or [γ 32P]-ATP (3000 Ci/mmol). On random RNA templates, AC RNA and ACG RNA, 200 μ M of GTP, 1.5 mM of each, ATP and CTP and 165 nM of [α 32P]-UTP (3000 Ci/mmol) were used. Internal initiation reactions were performed using 3'-AUGCUUCUGUUUGUUUGGUA-5' (Le+1), 3'-GAUGCUUCUGUUUGUUUGGUA-5' (Le+2) or 3'-GAAUGCUUCUGUUUGUUUGGUA-5' (Le+3) as a template. The effect of P was examined by the addition of 0.2 μ M of P to the standard reaction. The effect of ribavirin triphosphate (RTP) was evaluated on the C3A template with 165 nM of [α 32P]-UTP (3000 Ci/mmol), 1.5 mM of each, ATP and CTP and 200 μ M of either GTP or RTP. Assays studying the effects of RTP concentrations were carried out by adding the indicating concentrations of RTP in the standard polymerase assay. All radioisotopes were purchased from PerkinElmer.

Results

Establishment and characterization of a minimal in vitro RNA synthesis assay for RABV

To establish a minimal *in vitro* RNA synthesis assay for RABV, the RABV L and P proteins were expressed independently in insect cells and bacteria respectively, and purified them by Ni-NTA affinity followed by a size exclusion chromatography (FIG 20A). Using a system developed for VSV {Morin, 2012 #96}, another member of the rhabdovirus family, it is shown that RABV L can use a non-encapsidated RNA corresponding to the first 19 nt of the 3' end of RABV genome (RABV Le19) to synthesize RNA (FIG. 20B). This result demonstrates that the *in vitro* system previously developed for the vesiculovirus can be adapted for the lyssavirus. In the presence of RABV P, L RNA synthesis activity and processivity is enhanced (FIG. 20B). In order to observe if P of different rhabdoviruses

could enhance the activity of any other L proteins, the RABV L has been incubated with either RABV or VSV P, and VSV L with either VSV P or RABV P. RABV P can only enhance RABV L activity, while VSV P can only enhance VSV L (FIG. 20B,C) This result shows that only when L and P are coming from the same virus L has an activity enhanced by P, demonstrating that the interaction L-P is specific of each virus species.

Identification of the minimal P L-enhancing domain

A previous study have demonstrated by co- immunoprecipitation that the region from aa 1 to 19 of P is sufficient for L-binding but could be stabilized by extending this region to aa 52 {Chenik, 1998 #262}. Interestingly, the VSV L-binding domain is located just upstream the oligomerization domain (aa 92-131) {Gerard, 2009 #267;Rahmeh, 2012 #165}. Based on these data six P-deletions mutants were generated within the region 1-91 to identify the minimal P region that could still enhance both RNA synthesis and processivity of L, P1-91, P21-91, P11-91, P1-40, P1-50, and P11-50. While P21-91 does not enhance L RNA synthesis activity, both P1-91 and P11-91 enhance L activity that demonstrates that the region 11-21 of P is essential for the binding of L (FIG. 21). In the same way P1-40 does not modify the activity of L while P1-50, enhances both RNA synthesis and processivity showing that the region 40-50 is also critical for L enhancement. Based on these results, P11-50 was generated and observed that this domain enhanced L polymerase activity. Thus the minimal L-binding domain of P from aa 11 to 50 have been defined herein. The results confirms the previous results showing that the L-binding region of L was encompassed within residues 1-52 {Chenik, 1998 #262} and show that the first 10 aa are not essential for the enhancement of L.

Template specificity of the RABV polymerase complex

The sequence properties of the template on the RNA synthesis activity of L were investigated. The activity of RABV L on RABV Le19 and a synthetic RNA corresponding to the first 19 nt of the 3' end of VSV genome (VSV Le19) were compared. The RABV L-P can initiate RNA synthesis on VSV Le19 (FIG 22A). We have previously demonstrated that the first three nt of the template are essential for VSV L to initiate RNA synthesis {Morin, 2012 #96}. RABV Le19 and VSV Le19 share the same first three nucleotides with the first different nucleotide located at position 4, and 10 others located at position 5, 6, 8, 9, 13, 15, 16, 17, 18 and 19 (FIG. 22A). Although RABV L initiates on VSV Le19 its efficiency of initiation is lower than on RABV Le19 (FIG. 22B), suggesting a role of the nt located further the first three positions in stabilizing the initiation complex. Looking at the products size level, each RNA template is observed to generate a specific pattern (FIG. 22C). When RABV L-P use RABV Le19 as a template the products-size that are synthesized the most are about from 9 to 13 and 19 nt long with the 12 mer synthesized at a larger quantity than the other one. Interestingly very few RNA products are synthesized from 14 to 18 nt long indicating a good processivity of the polymerase around this length. On VSV Le19, RABV L-P present a products distribution different

than the one observe with RV Le19. While the 5 mer is produced at a larger level, L-P synthesized larger quantity of the long size product (>10 nt) with a slight reduction around the size of 15 nt. This result suggests that interactions between the template and the entry tunnel might influence the RNA synthesis by the polymerase.

To compare if these results would be similar with another rhabovirus L-P complex, a similar experiment was conducted using VSV L-P instead of RABV L-P (FIG. 22). As observed with RABV L-P, VSV L-P can initiate on both VSV Le19 and RABV Le19 templates, however initiation with RABV Le19 is less efficient than with VSV Le19, with around 50% less synthesis (FIG 23A, 23B). Surprisingly the profile of level of synthesis size by size with VSV L-P on RABV Le 19 is very similar as the one observed with RABV L-P showing the 12 mer synthesized at a larger quantity and few RNA synthesis for the sizes from 15 to 18 nt long (FIG. 23C). This suggests that the RNA sequence itself can drive the RNA synthesis property of both polymerases. However when VSV Le19 is used as a template the 2 polymerases present some slight differences.

RABV L cannot copy the VSV N-RNA template

We have demonstrated herein that RABV L initiate RNA synthesis on a synthetic RNA corresponding to the 3' end on the VSV genome (FIG. 22). However the genomic RNA of the NNS RNA viruses is completely enveloped by N to form the N-RNA meaning that during the copy of the template by the polymerase complex a small number of N have to dissociate locally to let access the RNA into L active site. Based on these two facts, the RABV L can copy VSV N-RNA by using a chimer P protein that could bind both RABV L and VSV N-RNA. During transcription, P binds L via its N-terminal end and N-RNA via its C-terminal end to bridge the interaction between the polymerase and its template. A chimer P (cRABV P) composed of the first 91 aa of the N-terminal end of RABV P and the oligomerization and C-terminal domains of VSV P, residues 107 to 265 was designed (FIG. 24A). While VSV L-P can use VSV N-RNA as a template and synthesized the 47 nt long leader RNA, the RABV L/cRABVP complex cannot copy the template (FIG. 29B). Moreover the lack of any bands below the size of 47 nt indicates that initiation does not seem to occur using the two complex RABV L/cRABVP and VSV-N-RNA. These results suggest that during RNA synthesis initiation of the polymerase complex on the N-RNA another interaction than L/P/N-RNA might be involved for a correct initiation of RNA synthesis. Perhaps, a direct binding L-N occurs at early stage of the RNA synthesis initiation as well as during elongation.

*In vitro assays**RNA synthesis and initiation in vitro assay*

Standard polymerase assays were carried out using 0.2 μ M of template with 0.2 μ M of VSV L in a reaction mixture containing 20 mM Tris-base, pH 8, 50 mM NaCl, 2 mM DTT and 0.5% (vol/vol) Triton X-100, 6 mM MgCl₂, 200 μ M UTP, 1.5 mM ATP, 1.5 mM CTP and 165 nM of [α 32P]-GTP (3000 Ci/mmol). Reactions were incubated at 30°C for 3 h, and stopped by the addition of EDTA/formamide. Reactions products were resolved using denaturing polyacrylamide gel electrophoresis (20% polyacrylamide, 7 M urea) in TBE buffer, and analysed by autoradiography. The sizes of the products were determined by comparison with a 19-nt marker RNA labelled by T4 polynucleotide kinase (PNK) (New England Biolabs) using [γ 32P]-ATP (3000 Ci/mmol). Assays studying the effects of NTP concentration were carried out using the indicated concentrations. For monitoring the effect of varying GTP concentration, reactions were supplemented with 165 nM of [α 32P]-ATP (3000 Ci/mmol). To determine the first nucleotide incorporated during initiation, standard reactions were carried out with 165 nM [γ 32P]-GTP (3000 Ci/mmol) or [γ 32P]-ATP (3000 Ci/mmol). On random RNA templates, AC RNA and ACG RNA, 200 μ M of GTP, 1.5 mM of each, ATP and CTP and 165 nM of [α 32P]-UTP (3000 Ci/mmol) were used. Internal initiation reactions were performed using 3'-AUGCUUCUGUUUGUUUGGUA-5' (Le+1), 3'-GAUGCUUCUGUUUGUUUGGUA-5' (Le+2), or 3'-GAAUGCUUCUGUUUGUUUGGUA-5' (Le+3) as a template. The effect of P was examined by the addition of 0.2 μ M of P to the standard reaction. The standard polymerase assay was supplemented with 250 μ M GTP and 0.2 μ M of P when N-RNA was used as a template. The effect of RTP was evaluated on the C3A template with 165 nM of [α 32P]-UTP (3000 Ci/mmol), 1.5 mM of each, ATP and CTP and 200 μ M of either GTP or RTP. Assays studying the effects of RTP concentrations were carried out by adding the indicating concentrations of RTP in the standard polymerase assay. All radioisotopes were purchased from Perkin-Elmer.

in vitro capping assay

To test for capped products, *in vitro* transcription reactions are carried out in the presence of [α -32P]GTP; RNA was purified and digested with TAP, and the products resolved by thin layer chromatography. Product that comigrated with a 7mGp marker, which confirmed that the RNA is capped and methylated.

in vitro methylation assay

Triphosphorylated RNA transcripts are capped in the presence of 20 μ Ci of [α -32P]GTP (3,000 Ci/mmol; Perkin-Elmer) and 7.5 U of the vaccinia virus capping enzyme (Epicentre) in 15- μ l reaction mixtures. After 2 h at 37°C, the reaction mixtures were treated with 10 U of calf intestinal alkaline

phosphatase (NEB) for 1 h, followed by purification. Capped RNA was quantified by measuring the incorporation of [α - 32 P]GTP and extrapolation of molar quantities from a standard curve of moles versus counts per minute. For methylation at the G-N-7 position, 100 μ M SAM is included in the capping reaction mixture. For methylation at the ribose 2'-O position or at both the G-N-7 and ribose 2'-O positions, the capped RNA or the G-N-7-methylated RNA was incubated with 100 μ M SAM and 7.5 U of vaccinia virus 2'-O MTase (Epicentre). The products of this reaction were purified, digested with nuclease P1, and resolved by TLC on PEI cellulose F sheets. Nuclease P1 digestion cleaves the 3'-5' phosphodiester bonds in single-stranded RNA but does not cleave the 5'-5' bond of the cap structure, resulting in the release of GpppA from unmethylated RNA. Following incubation of the RNA with purified L protein, the products of nuclease P1 digestion comigrated with a m⁷GpppAm marker, demonstrating that L possesses both G-N-7 and 2'-O MTase activities.

All patents, patent applications, and publications identified in this document are expressly incorporated herein by reference for the purpose of describing and disclosing, for example, the methodologies described in such publications that might be used in connection with the present invention. These publications are provided solely for their disclosure prior to the filing date of the present application. Nothing in this regard should be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention or for any other reason. All statements as to the date or representation as to the contents of these documents is based on the information available to the applicants and does not constitute any admission as to the correctness of the dates or contents of these documents.

It is understood that the foregoing detailed description and the following examples are illustrative only and are not to be taken as limitations upon the scope of the invention. Various changes and modifications to the disclosed embodiments, which will be apparent to those of skill in the art, may be made without departing from the spirit and scope of the present invention. Further, all patents, patent applications, and publications identified are expressly incorporated herein by reference for the purpose of describing and disclosing, for example, the methodologies described in such publications that might be used in connection with the present invention. These publications are provided solely for their disclosure prior to the filing date of the present application. Nothing in this regard should be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention or for any other reason. All statements as to the date or representation as to the contents of these documents are based on the information available to the applicants and do not constitute any admission as to the correctness of the dates or contents of these documents.

It should be understood that this invention is not limited to the particular methodology, protocols, and reagents, etc., described herein and as such may vary. 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 is defined solely by the claims.

REFERENCES

- Abraham, G., and Banerjee, A.K. (1976). Sequential transcription of the genes of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA* 73, 1504–1508.
- Adams, P.D., Afonine, P.V., Bunko' czi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., apral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- Albertini, A.A., Wernimont, A.K., Muziol, T., Ravelli, R.B., Clapier, C.R., Schoehn, G., Weissenhorn, W., and Ruigrok, R.W. (2006). Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 313, 360–363.
- Ball, L.A., and White, C.N. (1976). Order of transcription of genes of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA* 73, 442–446.
- Baltimore, D., Huang, A.S., and Stampfer, M. (1970). Ribonucleic acid synthesis of vesicular stomatitis virus, II. An RNA polymerase in the virion. *Proc. Natl. Acad. Sci. USA* 66, 572–576.
- Barr, J.N., and Wertz, G.W. (2001). Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 30-AUAC-50 tetranucleotide: implications for the mechanism of transcription termination. *J. Virol.* 75, 6901–6913.
- Barr, J.N., Whelan, S.P., and Wertz, G.W. (1997). cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *J. Virol.* 71, 8718–8725.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21.
- Ding, H., Green, T.J., Lu, S., and Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* 80, 2808–2814.
- Egloff, M.P., Benarroch, D., Selisko, B., Romette, J.L., and Canard, B. (2002). An RNA cap (nucleoside-20-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.* 21, 2757–2768.
- Emerson, S.U., and Wagner, R.R. (1973). L protein requirement for in vitro RNAsynthesis by vesicular stomatitis virus. *J. Virol.* 12, 1325–1335.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.
- Galloway, S.E., and Wertz, G.W. (2008). S-adenosyl homocysteine-induced hyperpolyadenylation of vesicular stomatitis virus mRNA requires the methyltransferase activity of L protein. *J. Virol.* 82, 12280–12290.

- Green, T.J., and Luo, M. (2009). Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc. Natl. Acad. Sci. USA* 106, 11713–11718.
- Green, T.J., Zhang, X., Wertz, G.W., and Luo, M. (2006). Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 313, 357–360.
- Grigorieff, N., and Harrison, S.C. (2011). Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Curr. Opin. Struct. Biol.* 21, 265–273.
- Hereyk, N., Horikami, S.M., and Moyer, S.A. (1988). The vesicular stomatitis virus L protein possesses the mRNA methyltransferase activities. *Virology* 163, 222–225.
- Hunt, D.M., Smith, E.F., and Buckley, D.W. (1984). Aberrant polyadenylation by a vesicular stomatitis virus mutant is due to an altered L protein. *J. Virol.* 52, 515–521.
- Iverson, L.E., and Rose, J.K. (1981). Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. *Cell* 23, 477–484.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110–119.
- Krumm, S.A., Yan, D., Hovingh, E.S., Evers, T.J., Enkirch, T., Reddy, G.P., Sun, A., Saindane, M.T., Arrendale, R.F., Painter, G., et al. (2014). An orally available, small-molecule polymerase inhibitor shows efficacy against a lethal morbillivirus infection in a large animal model. *Sci. Transl. Med.* 6, 232ra252.
- La Ferla, F.M., and Peluso, R.W. (1989). The 1:1 N-NS protein complex of vesicular stomatitis virus is essential for efficient genome replication. *J. Virol.* 63, 3852–3857.
- Li, J., Rahmeh, A., Morelli, M., and Whelan, S.P. (2008). A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. *J. Virol.* 82, 775–784.
- Li, J., Rahmeh, A., Brusic, V., and Whelan, S.P. (2009). Opposing effects of inhibiting cap addition and cap methylation on polyadenylation during vesicular stomatitis virus mRNA synthesis. *J. Virol.* 83, 1930–1940.
- Liuzzi, M., Mason, S.W., Cartier, M., Lawetz, C., McCollum, R.S., Dansereau, N., Bolger, G., Lapeyre, N., Gaudette, Y., Lagace, L., et al. (2005). Inhibitors of respiratory syncytial virus replication target cotranscriptional mRNA guanylation by viral RNA-dependent RNA polymerase. *J. Virol.* 79, 13105–13115.
- Lu, X., McDonald, S.M., Tortorici, M.A., Tao, Y.J., Vasquez-Del Carpio, R., Nibert, M.L., Patton, J.T., and Harrison, S.C. (2008). Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. *Structure* 16, 1678–1688.
- Lyumkis, D., Brilot, A.F., Theobald, D.L., and Grigorieff, N. (2013). Likelihoodbased classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* 183, 377–388.

- Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* 142, 334–347.
- Morin, B., and Whelan, S.P. (2014). Sensitivity of the polymerase of vesicular stomatitis virus to 20 substitutions in the template and nucleotide triphosphate during initiation and elongation. *J. Biol. Chem.* 289, 9961–9969.
- Morin, B., Rahmeh, A.A., and Whelan, S.P. (2012). Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. *EMBO J.* 31, 1320–1329.
- Ogino, T., and Banerjee, A.K. (2007). Unconventional mechanism of Mrna capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol. Cell* 25, 85–97.
- Patton, J.T., Davis, N.L., and Wertz, G.W. (1984). N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitisvirus. *J. Virol.* 49, 303–309.
- Peluso, R.W., and Moyer, S.A. (1983). Initiation and replication of vesicular stomatitis virus genome RNA in a cell-free system. *Proc. Natl. Acad. Sci. USA* 80, 3198–3202.
- Pflug, A., Guilligay, D., Reich, S., and Cusack, S. (2014). Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* 516, 355–360.
- Rahmeh, A.A., Li, J., Kranzusch, P.J., and Whelan, S.P. (2009). Ribose 20-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein. *J. Virol.* 83, 11043–11050.

CLAIMS

1. A method of identifying an inhibitor of a non-segmented negative strand (NNS) virus, the method comprising:
 - (a) using a high resolution 3D structure of an NNS virus L protein to select a candidate inhibitor from a virtual library of 3D structures of potential candidates by performing structure based computational modeling;
 - (b) obtaining the candidate inhibitor; and
 - (c) measuring an activity of L protein when contacted with the candidate inhibitor obtained in step (b), wherein a decrease in the activity confirms that the candidate is an inhibitor of the NNS virus.
2. A method of identifying an inhibitor of a non-segmented negative strand (NNS) virus polymerase, the method comprising:
 - (a) using a high resolution 3D structure of an NNS virus L protein to select a candidate inhibitor from a virtual library of 3D structures of potential candidates by performing structure based computational modeling;
 - (b) obtaining the candidate inhibitor; and
 - (c) measuring L protein polymerase activity when contacted with the candidate inhibitor obtained in step (b), wherein a decrease in the activity of the polymerase confirms that the candidate is an inhibitor of the NNS virus polymerase.
3. A method of identifying an inhibitor of an NNS virus the method comprising:
 - (a) screening a library of candidate molecules for inhibition of an NNS L protein in an *in vitro* biochemical assay for an L protein activity;
 - (b) using a high resolution 3D structure of the L protein to determine the binding site of potential candidates obtained in step (a) by computer modeling;
 - (c) identifying a potential refinement to the candidate inhibitor;
 - (d) obtaining a candidate inhibitor molecule comprising the refinement identified in step (c); and
 - (e) measuring the activity of the L protein when contacted with the candidate inhibitor obtained in step (d), wherein a decrease in the activity of the L protein confirms that the candidate is an inhibitor of the NNS virus.
4. A method of identifying an inhibitor of L protein of an NNS virus, the method comprising:
 - (a) using a high resolution 3D structure of the L protein of vesicular stomatitis virus (VSV) to create an atomic model of the L protein of a second NNS virus by homology modeling the amino acid sequence of the L protein of the second NNS virus onto the high resolution 3D structure of the VSV L protein ;

- (b) using the atomic model of the second NNS virus L protein to select a candidate inhibitor from a virtual library of 3D structures of potential candidates by performing structure based computational modeling;
- (c) obtaining the candidate inhibitor; and
- (d) measuring an activity of L protein of the second NNS virus when contacted with the candidate inhibitor obtained in step (b), wherein a decrease in the activity confirms that the candidate is an inhibitor of the second NNS virus.
5. The method of any one of claims 1-4 wherein the high resolution 3D structure has a resolution of at least 3.8 Å.
6. The method of claim 4 wherein the high resolution 3D structure is an atomic structure determined via cryo-electron microscopy and single particle analysis.
7. The method of any one of claims 1-4 wherein the NNS virus L protein is vesicular stomatitis virus L protein.
8. The method of claim 7 wherein the VSV L protein high resolution structure has space group P1, so as to form a unit cell of dimensions of $a=112.0$ Å, $b=143.0$ Å, $c=106.0$ Å and $\alpha, \beta, \gamma=90^\circ$.
9. The method of claim 7, wherein the VSV L protein high resolution 3D structure has the coordinates as set out in the structure deposited with the RCSB Protein Data Bank with Reference Number 5A22.
10. The method of any of claims 1, 3 or 4 wherein the activity of the L protein measured is selected from the group consisting of RNA dependent RNA polymerase activity, GDP polyribonucleotidyltransferase (PRNTase) activity, methyltransferase activity, and the ability to structurally rearrange L protein domains during L protein functions.
11. The method of claim 2 or claim 10 wherein the RNA dependent RNA polymerase activity measured is selected from the group consisting of transcript initiation, elongation and termination.
12. The method of claim 1, 2 or 4 wherein, rather than selecting a candidate inhibitor from a virtual library of 3D structures of potential candidates, the high resolution 3D structure of the NNS virus L protein is used to design a potential inhibitor.
13. The method of claim 12 wherein the potential inhibitor is designed to interact with an L protein domain selected from the group consisting of the RNA dependent RNA polymerase (RdRp) domain, the mRNA capping domain, the methyltransferase domain, the loop primer domain, the connector domain, and the C-terminal domain.

14. A method of mapping the location of resistance to a known inhibitor of an NNS virus L protein on the 3D structure of the L protein, the method comprising:
 - (a) homology-based structural mapping of the sequence comprising the resistance mutation onto the 3.8 Å atomic structure of an NNS virus L protein.
15. The method of claim 14 wherein the high high resolution 3D structure is a 3.8Å resolution structure.
16. A method of inhibiting an NNS virus, the method comprising contacting an NNS virus or a cell infected with an NNS virus with an inhibitor identified according to the method of any one of claims 1-4.
17. An NNS virus inhibitor identified using the method of any one of claims 1-4.
18. The method of any one of claims 1-4 wherein the identified inhibitor binds a domain of the L protein at a region selected from the group consisting of the RNA dependent RNA polymerase (RdRp) domain, the mRNA capping domain, the methyltransferase domain, the loop primer domain, the connector domain, and the C-terminal domain of the L protein.
19. The method of any one of claims 1-4 wherein the identified inhibitor attenuates the RNA dependent RNA polymerase activity, the GDP polyribonucleotidyltransferase (PRNTase) activity, the methyltransferase activity, the loop priming activity, or the ability of the L protein to structurally rearrange its domains during L one or more of its functions.
20. The method of any one of claims 1-4 which further comprises the step of contacting the confirmed candidate inhibitor with a cell infected with an NNS virus.
21. The method of any of claims 1-4 wherein the NNS virus is selected from the viruses of the Filoviridae, Paramyxoviridae, Rhabdoviridae and Bornaviridae families.
22. The method of claim 21 wherein the NNS virus is a Filovirus, selected from Ebola virus, Marburg virus and Cuevavirus.
23. The method of claim 21 wherein the NNS virus is a Paramyxovirus selected from measles virus, Nipah virus, Hendra virus, respiratory syncytial virus and Newcastle disease virus.
24. The method of claim 21 wherein the NNS virus is a Rhabdovirus selected from Lyssavirus, Vesiculovirus, Perhabdovirus, Sigmavirus, Ephemerovirus, Tibrovirus, Tupavirus and Sprivivirus.
25. A method of designing a candidate inhibitor of an NNS virus, the method comprising:

- (a) identifying an NNS virus mutant that escapes the activity of a known inhibitor of the L protein of an that virus;
- (b) identifying the site of the L protein resistance mutation by sequence analysis;
- (c) using homology-based structural mapping of the sequence comprising the resistance mutation to map the mutation site onto a high resolution 3D structure of an NNS virus L protein;
- (d) using computer modeling with the high resolution 3D structure to selecting a candidate binding molecule that targets the site of the resistance mutation mapped in step (c);
- (e) obtaining the candidate inhibitor molecule identified in step (d).

26. The method of claim 25 further comprising the step of contacting L protein of the NNS virus with the candidate inhibitor in an *in vitro* assay of L protein activity, wherein a decrease in L protein activity indicates the candidate inhibitor is an NNS virus inhibitor.

27. The method of claim 25 further comprising the step of contacting the candidate inhibitor with the NNS virus and assaying viral function, wherein a decrease in viral function indicates the candidate inhibitor is an NNS virus inhibitor.

28. A method of designing a candidate inhibitor of an NNS virus, the method comprising:

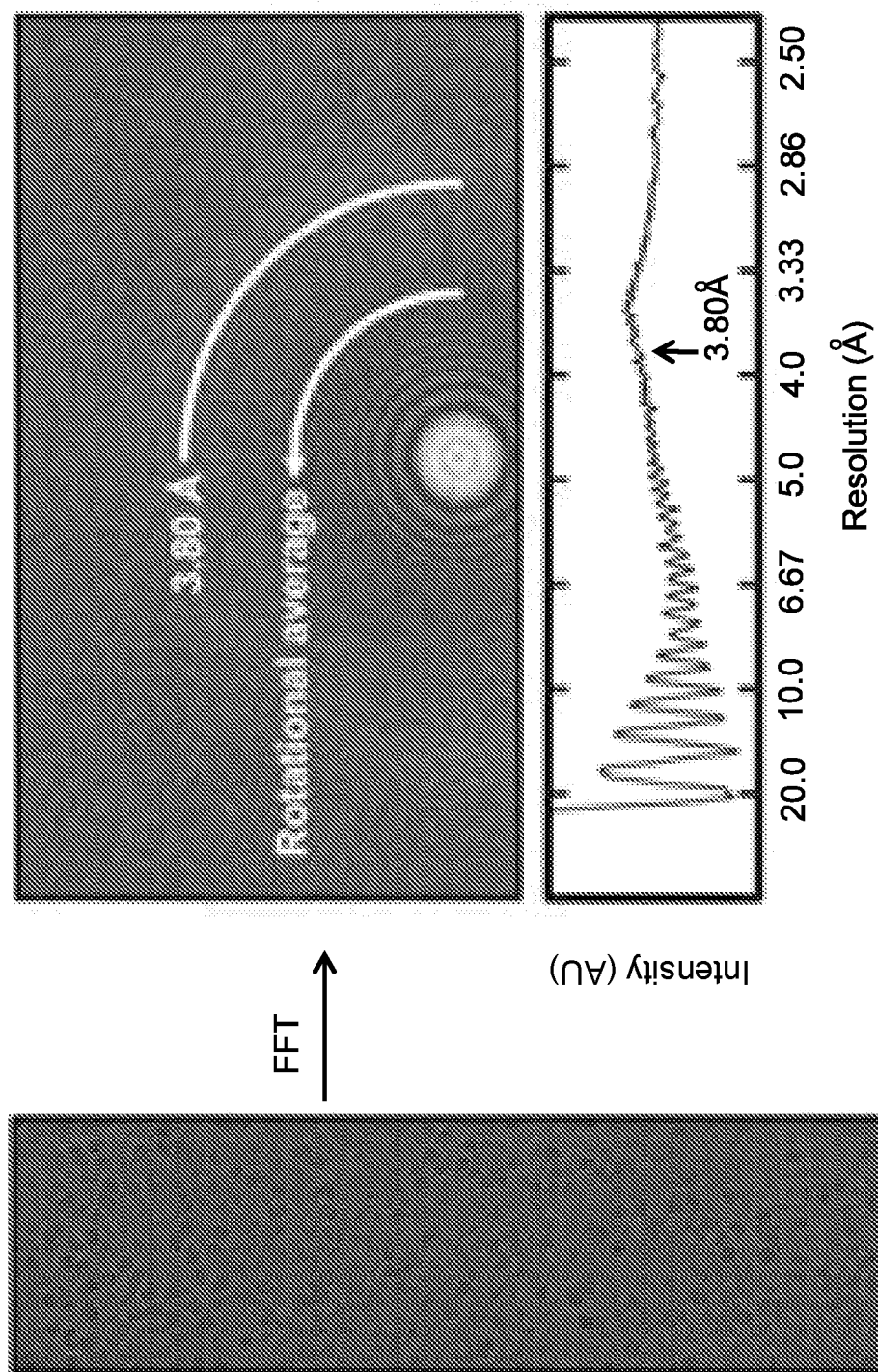
- (a) identifying an NNS virus mutant that escapes the activity of a known inhibitor of the L protein of an that virus;
- (b) identifying the site of the L protein resistance mutation by sequence analysis;
- (c) using homology-based structural mapping of the sequence comprising the resistance mutation to map the mutation site onto a high resolution 3D structure of an NNS virus L protein;
- (d) identifying a potential refinement to the known inhibitor that permits interaction with the mutant L protein;
- (e) obtaining the inhibitor molecule comprising the refinement identified in step (d); and
- (f) assaying mutant virus replication or infection in the presence and absence of the refined inhibitor molecule, wherein a decrease in replication or infection indicates the refined inhibitor is effective against the mutant virus.

29. Use of an NNS virus inhibitor identified according to the methods of any one of claims 1-4, 25 and 28 for the treatment of an individual infected with an NNS virus.

30. The use of claim 29 in which the NNS virus is selected from the group consisting of viruses of the Filoviridae, Paramyxoviridae, Rhabdoviridae and Bornaviridae families.

31. A computer-readable medium with high resolution 3D structure coordinate information for an NNS L protein stored thereupon.

1/69

*FIG. 1B**FIG. 1A*

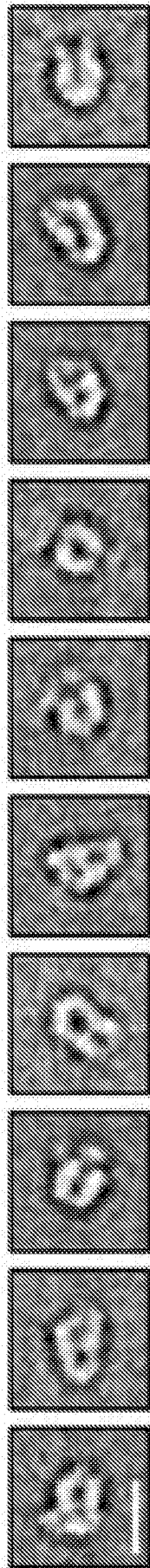
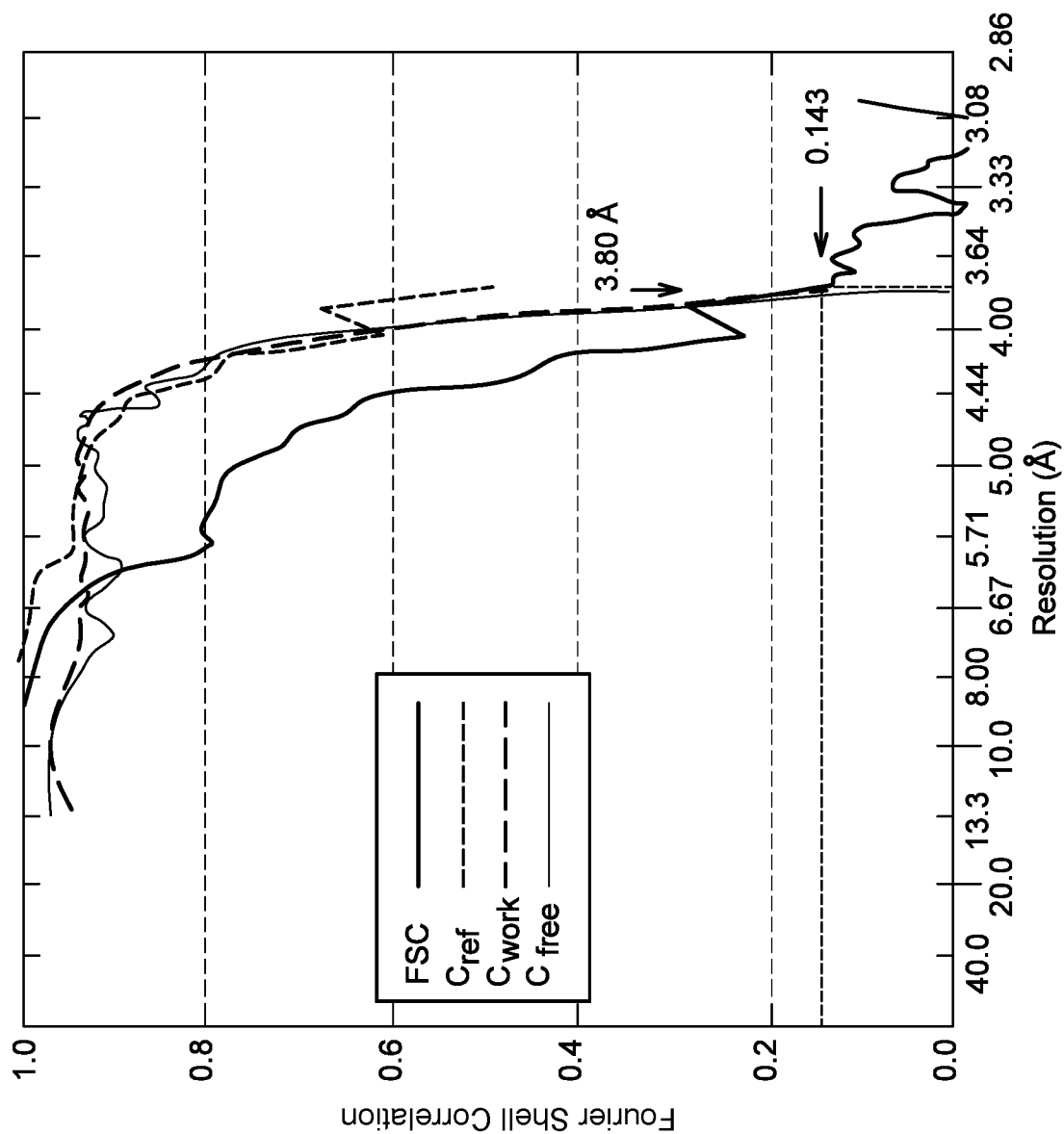


FIG. 1C

3/69

**FIG. 1D**

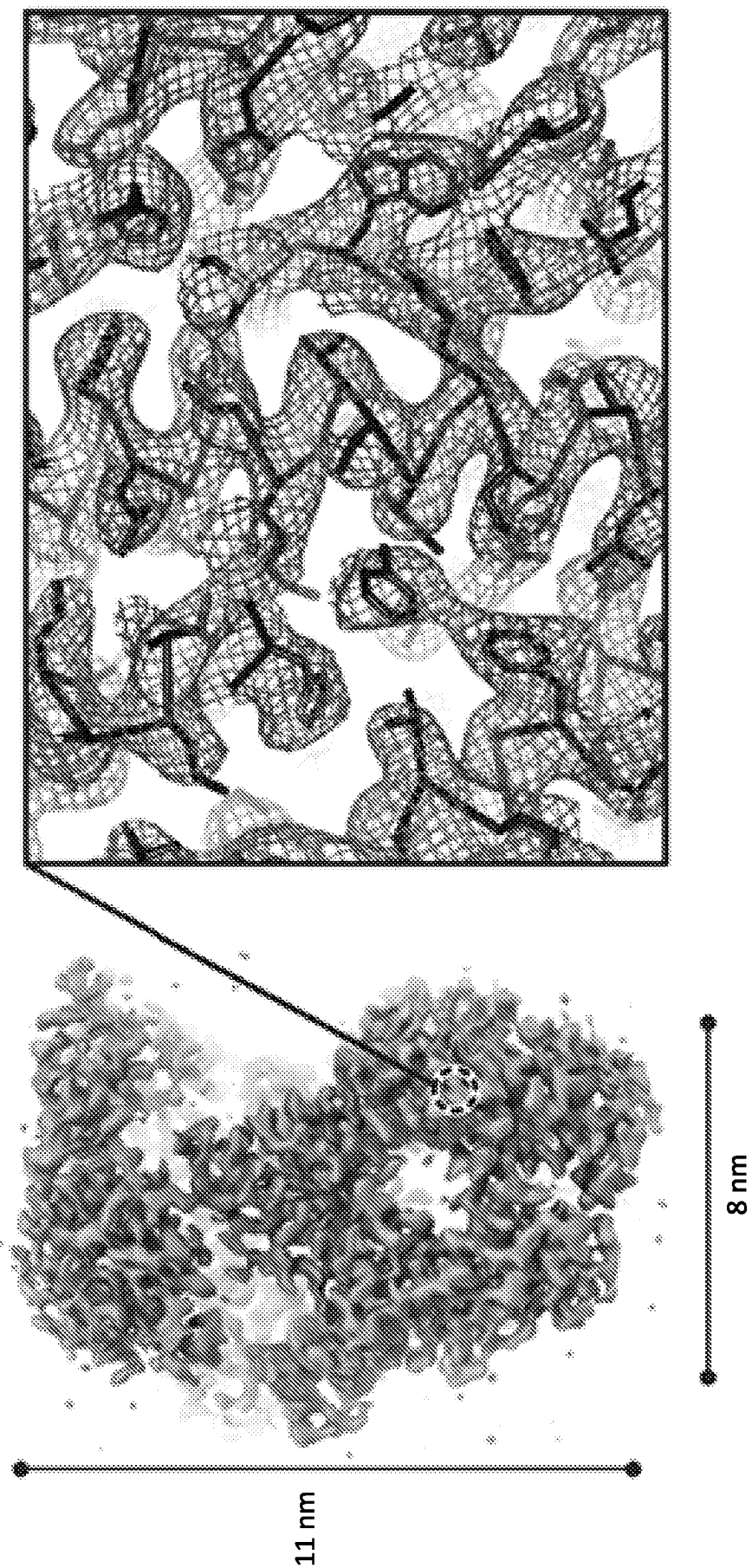


FIG. 1E

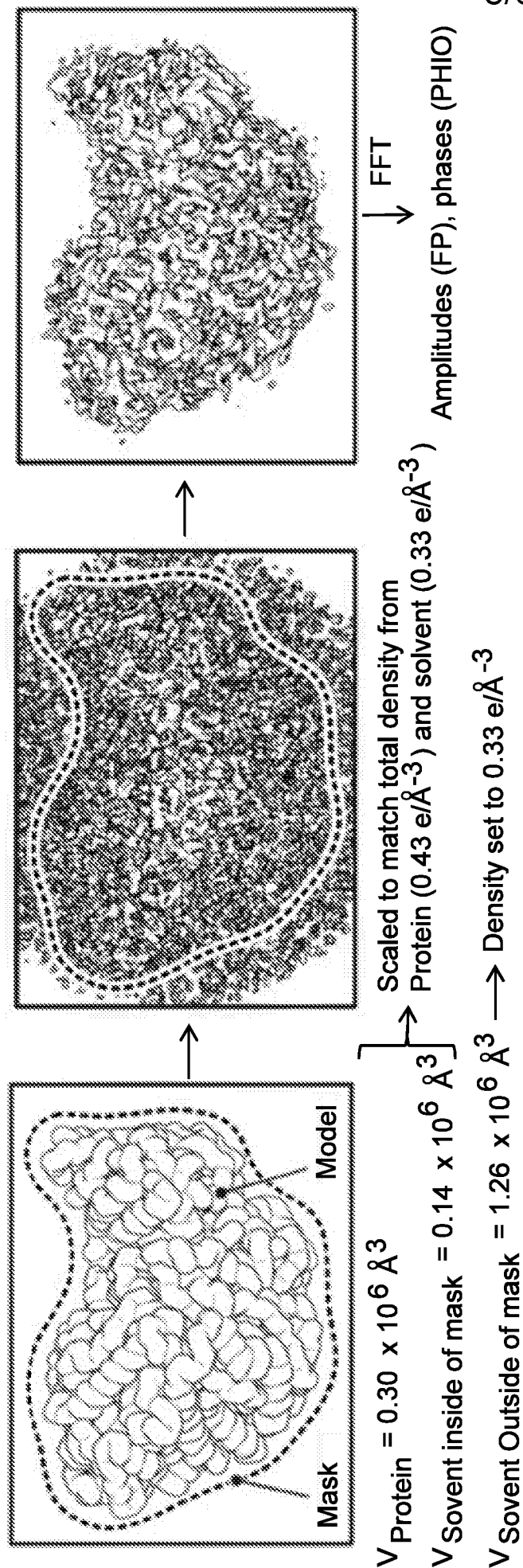


FIG. 2A

FIG. 2B

FIG. 2C

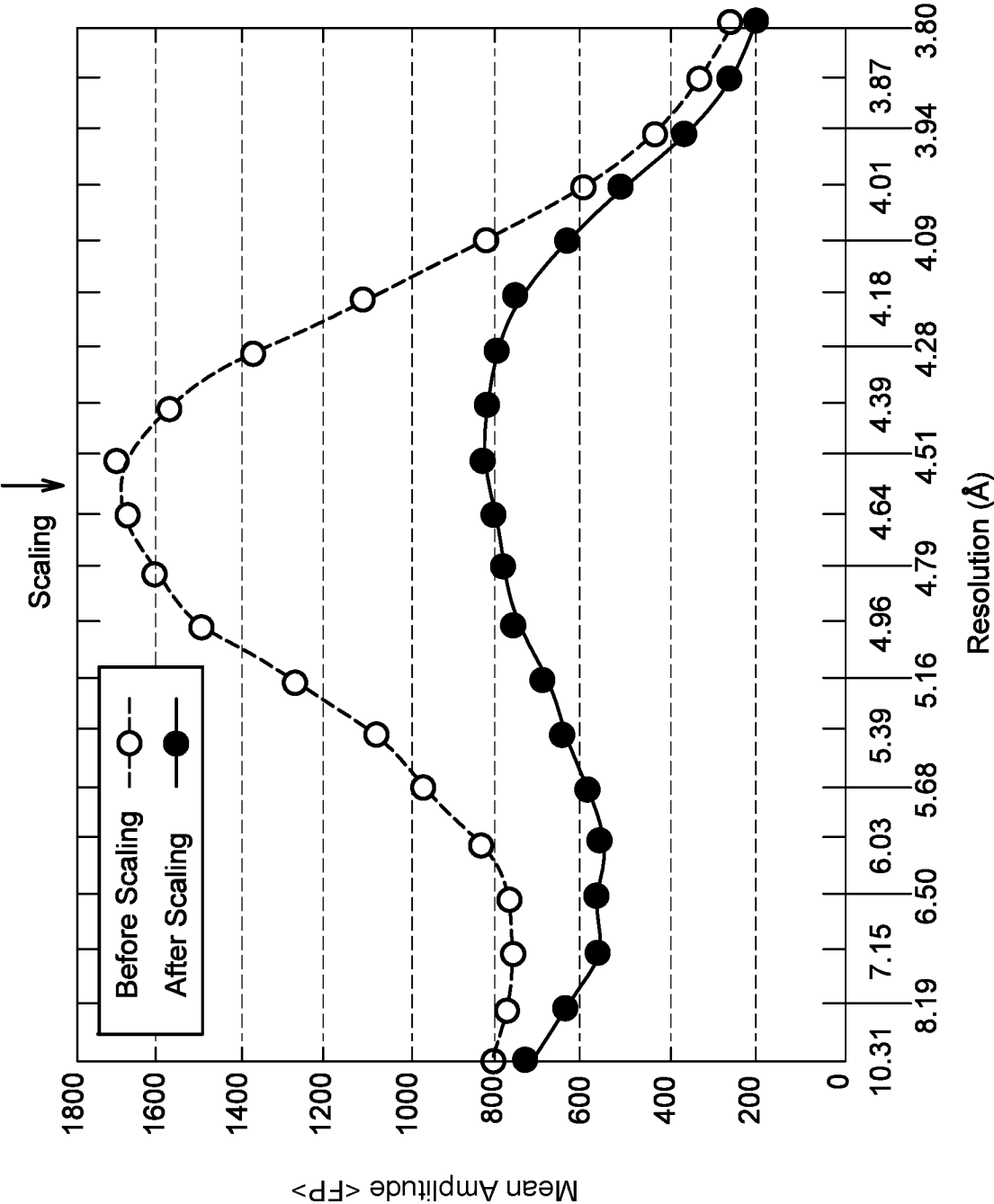


FIG. 2D

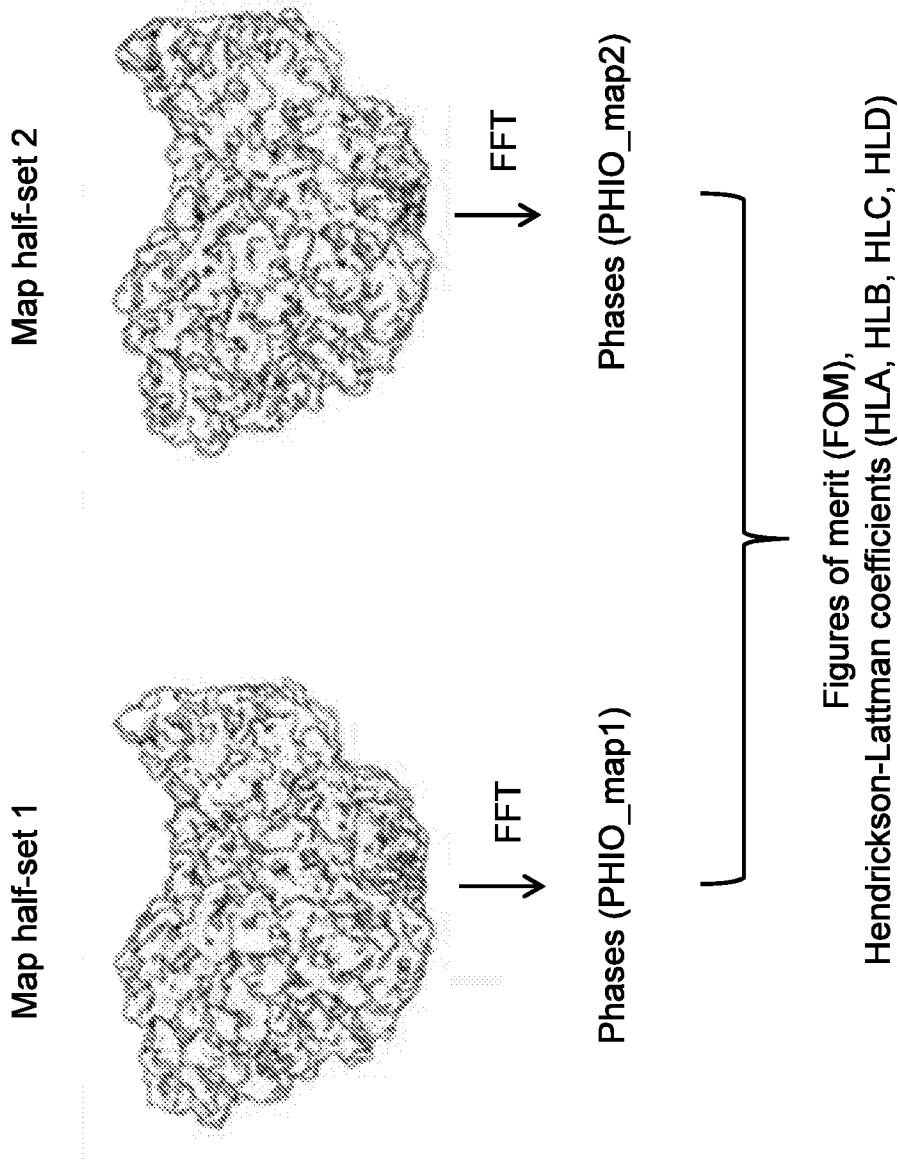


FIG. 2E

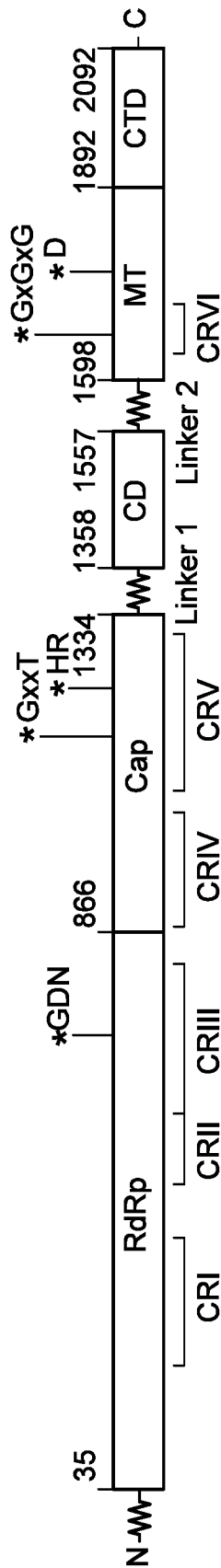
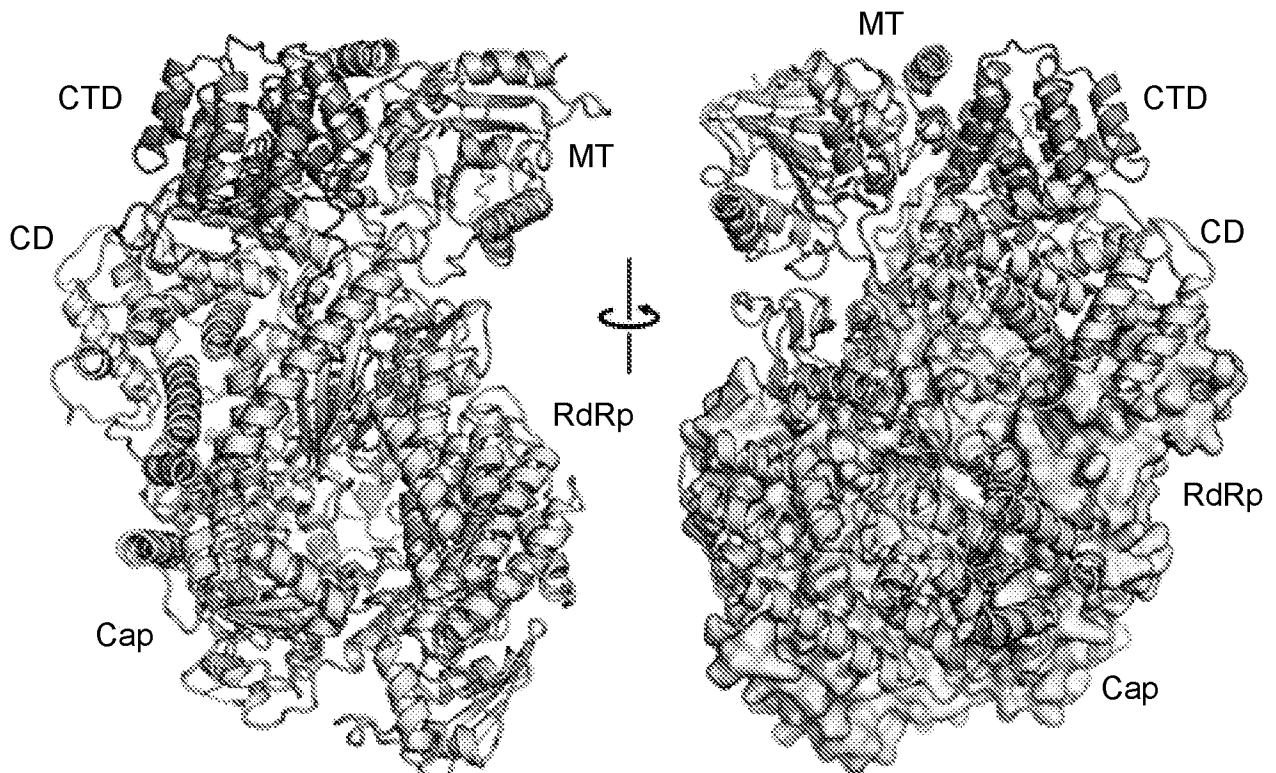
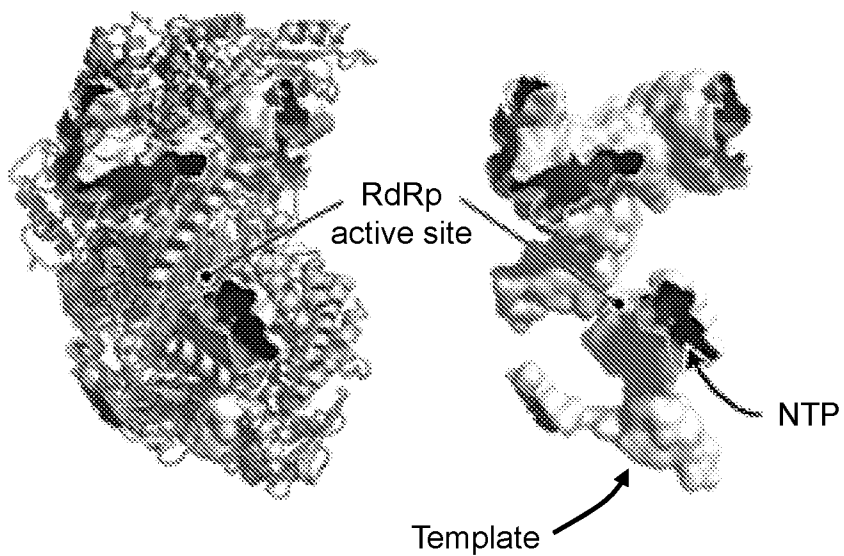
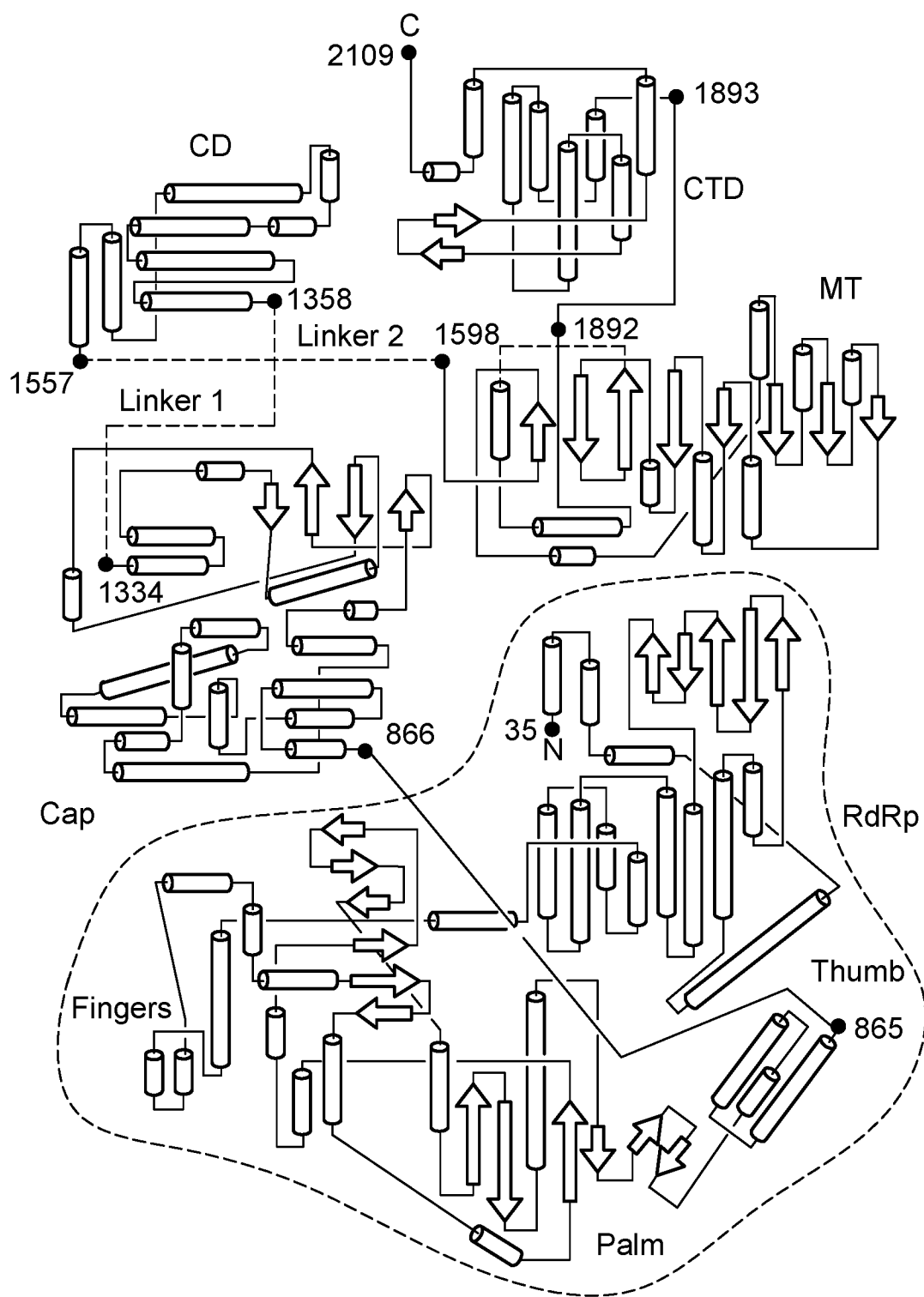


FIG. 3A

9/16

*FIG. 3B**FIG. 3C*

10/69

**FIG. 4**

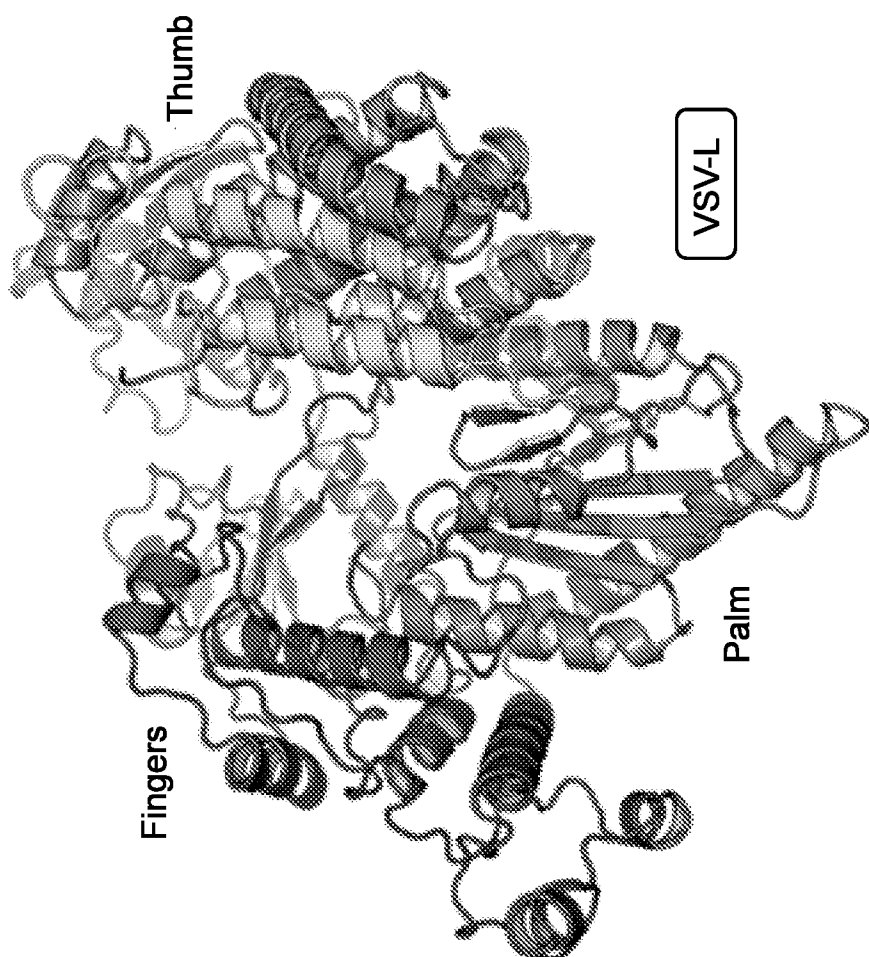


FIG. 5A

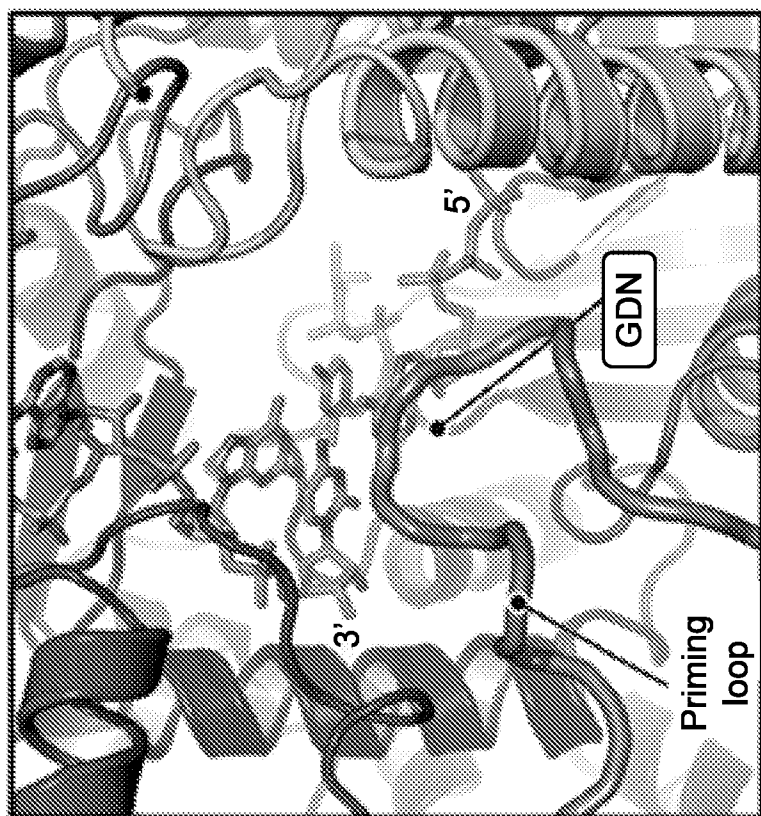


FIG. 5B

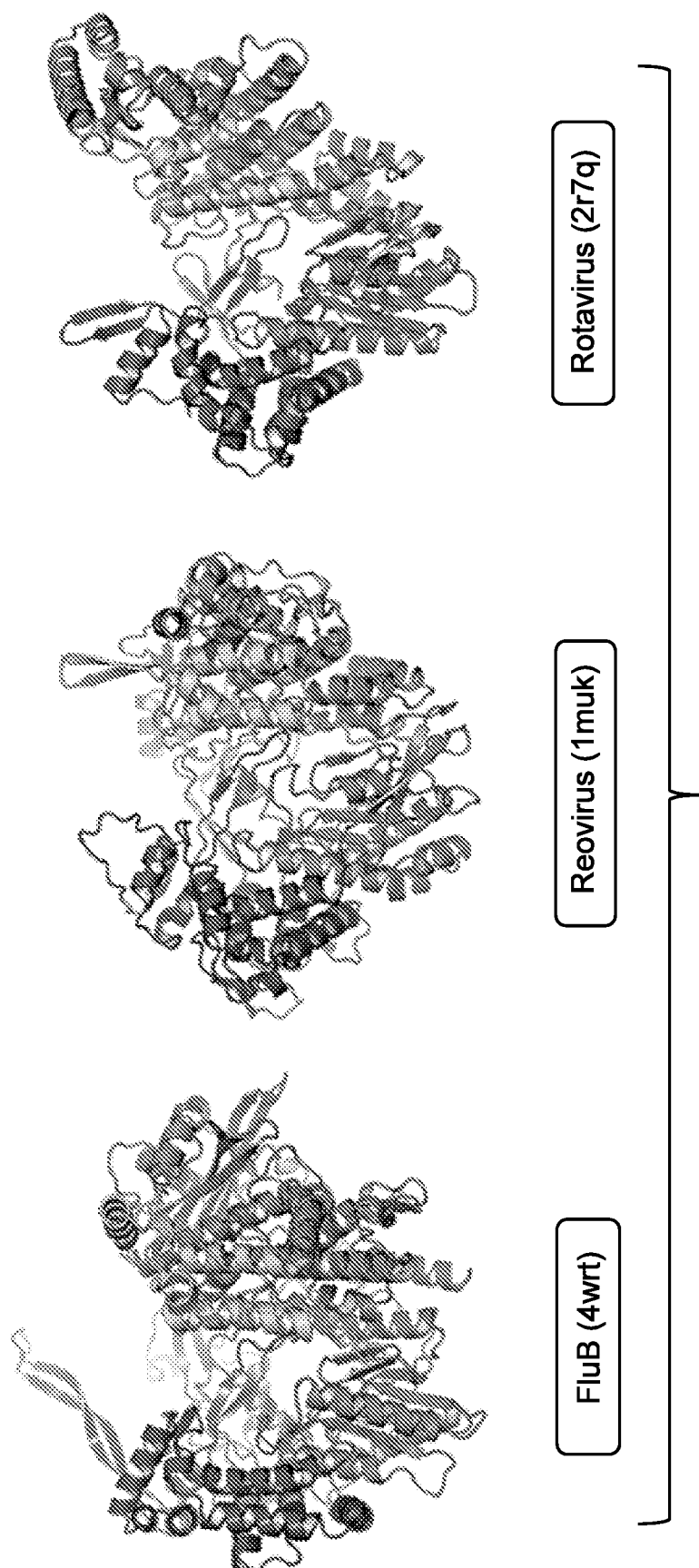


FIG. 5C

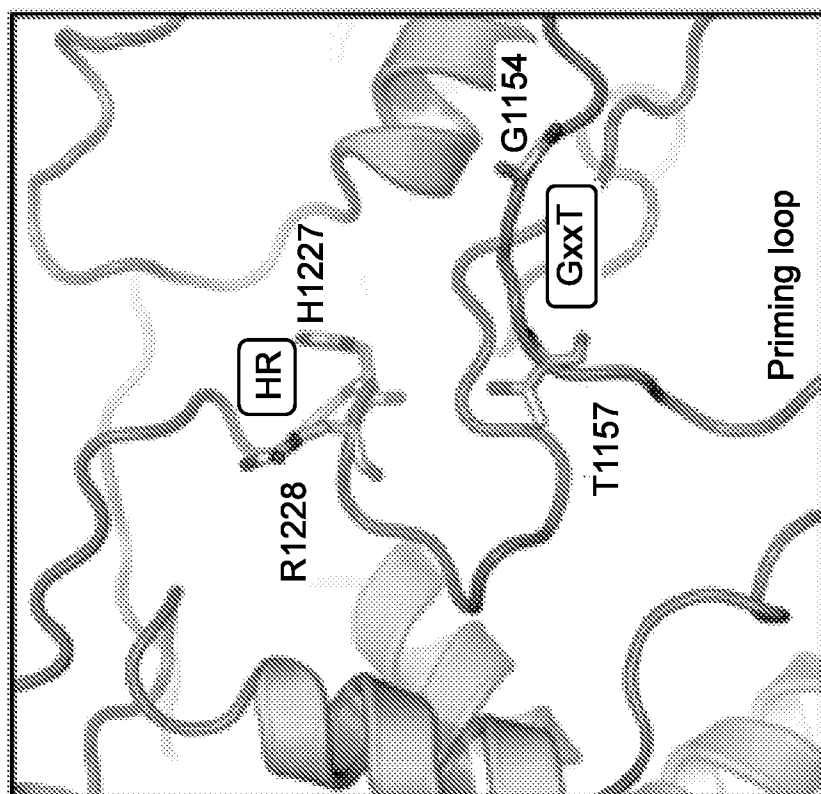


FIG. 6B

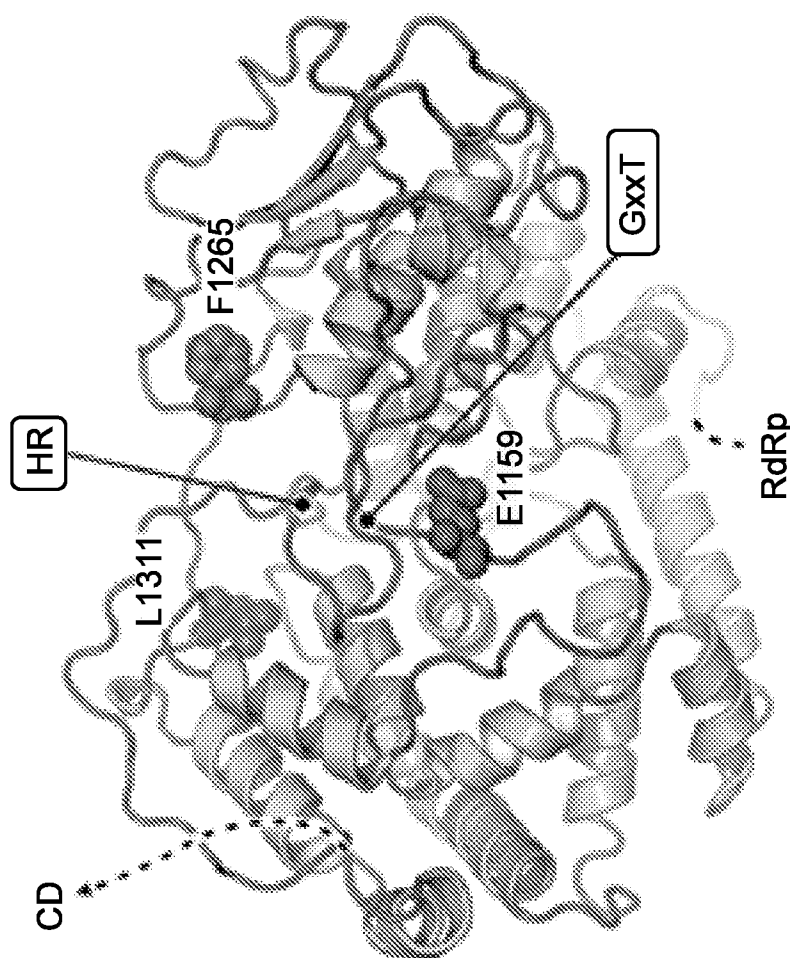


FIG. 6A

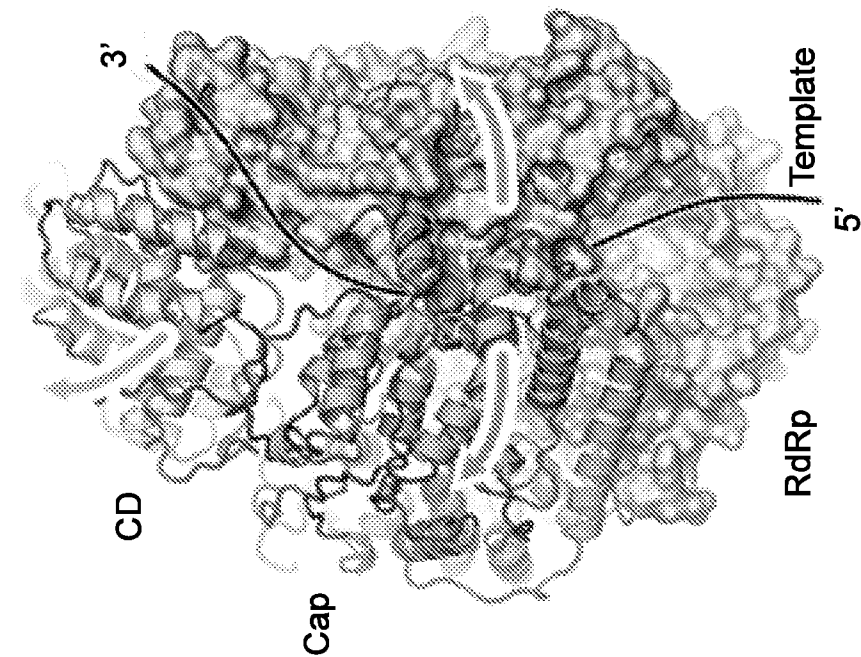


FIG. 6D

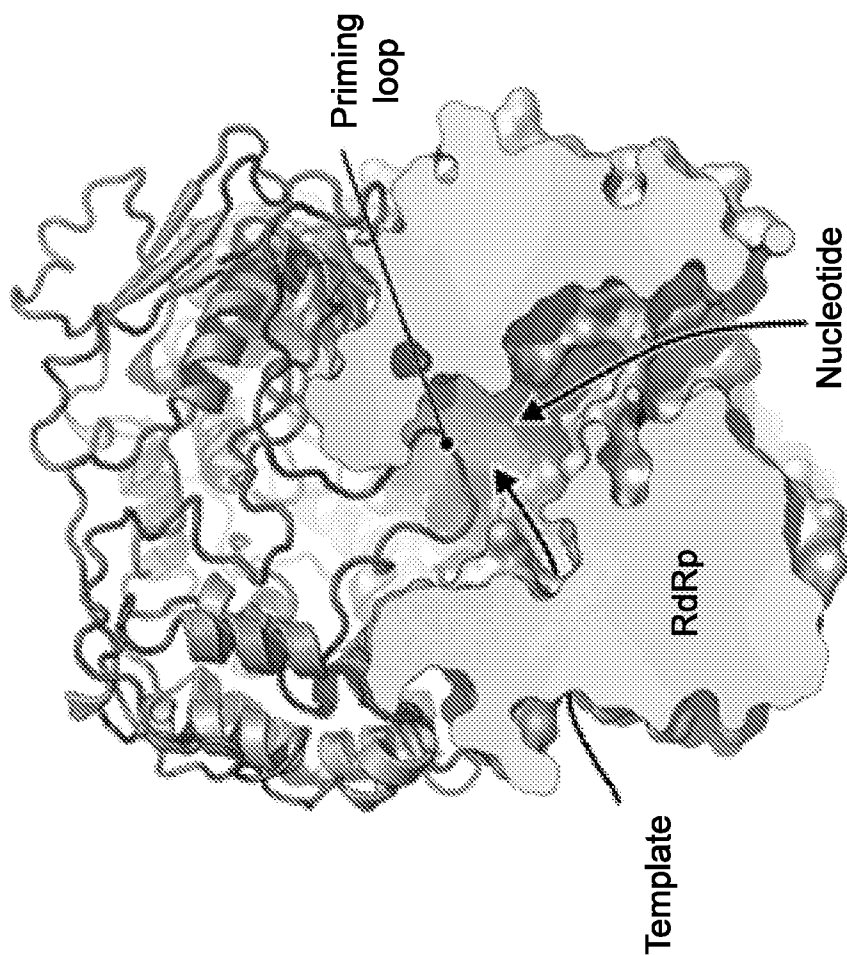
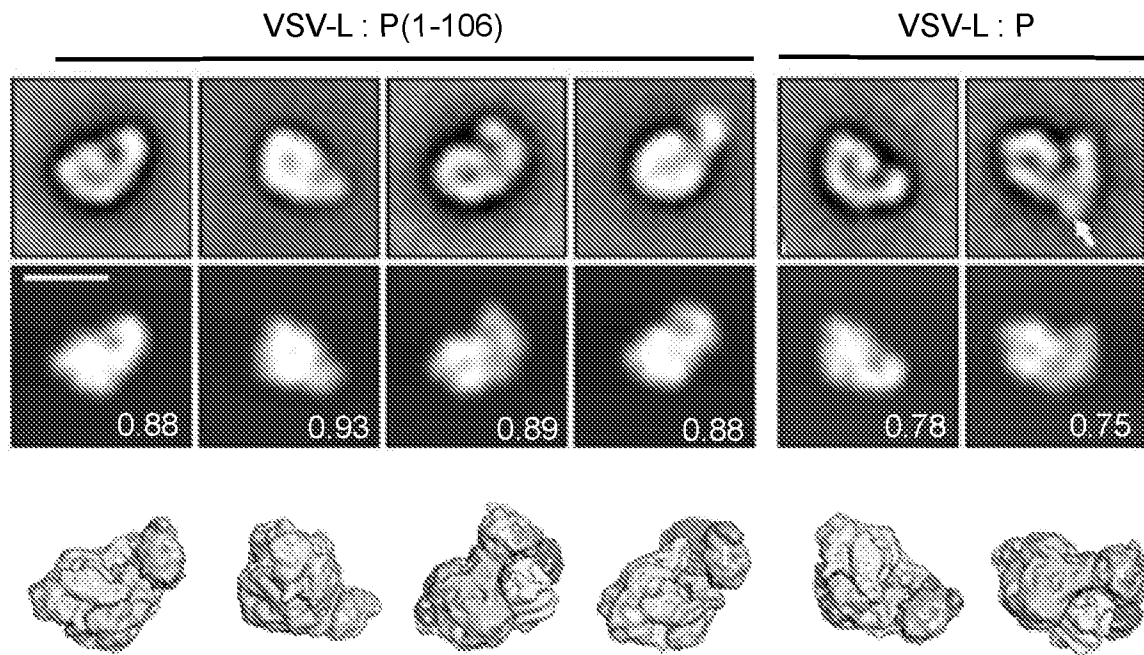
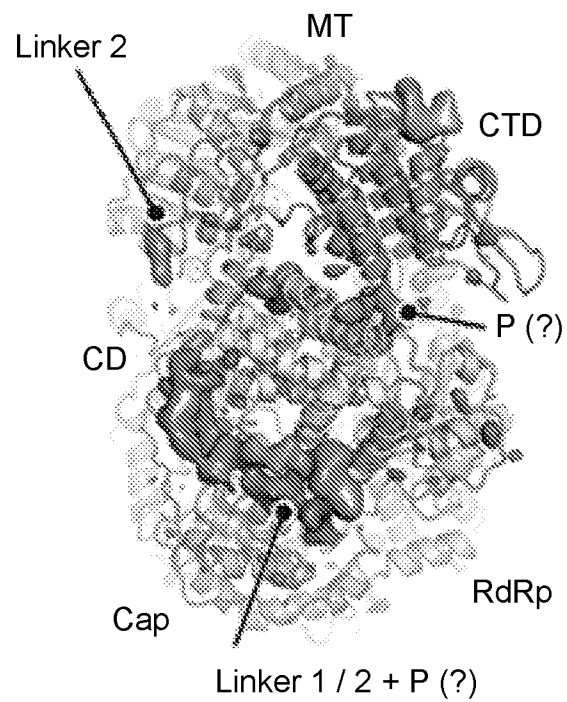


FIG. 6C



FIG. 7

16/69

*FIG. 8A**FIG. 8B*

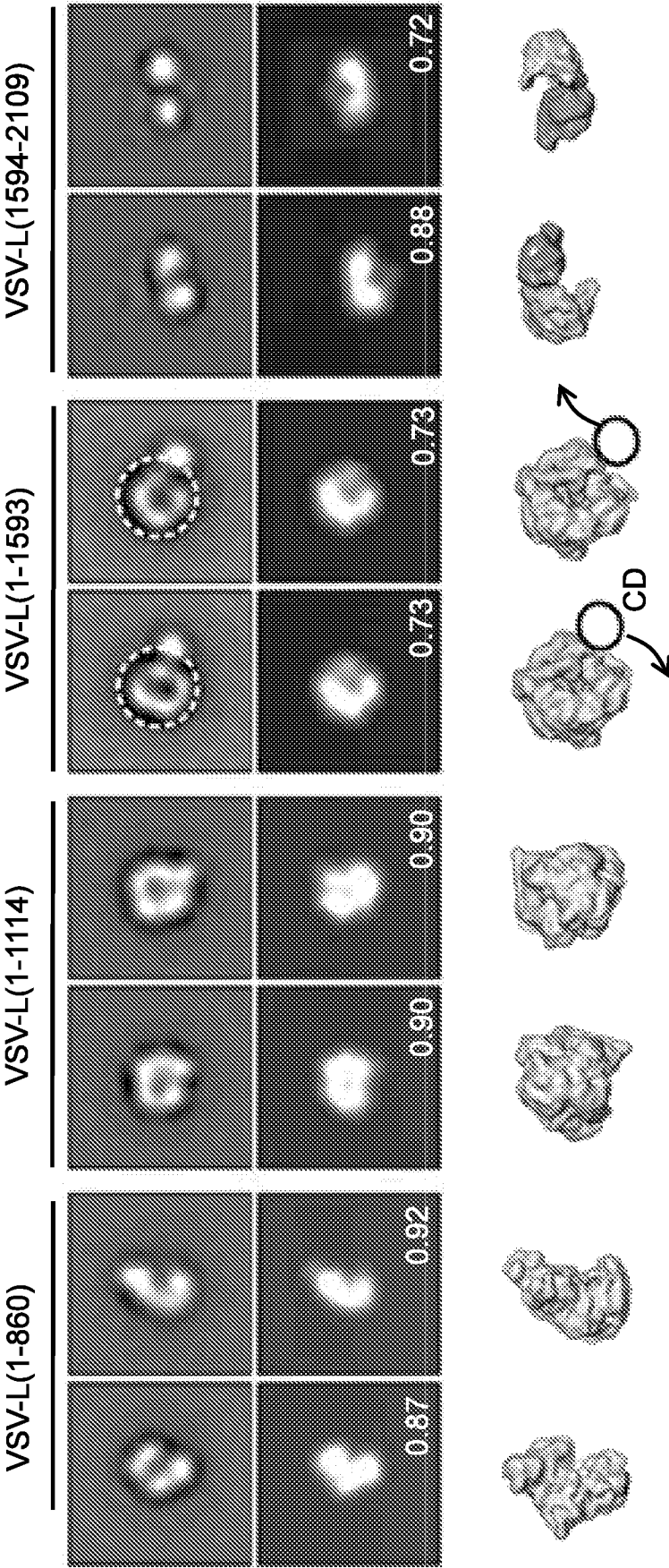


FIG. 8C

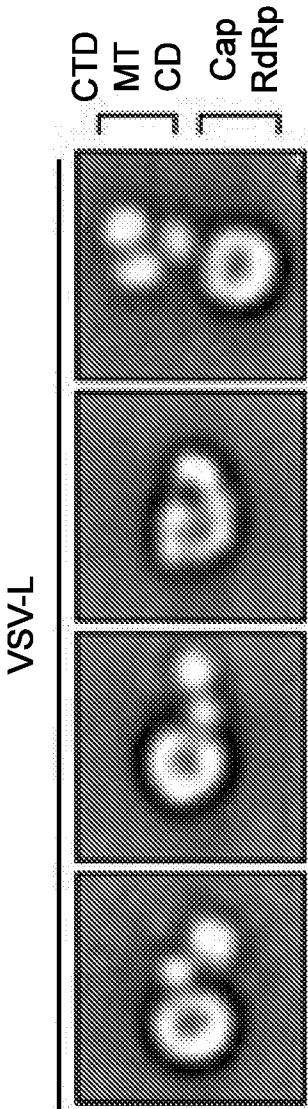


FIG. 8D

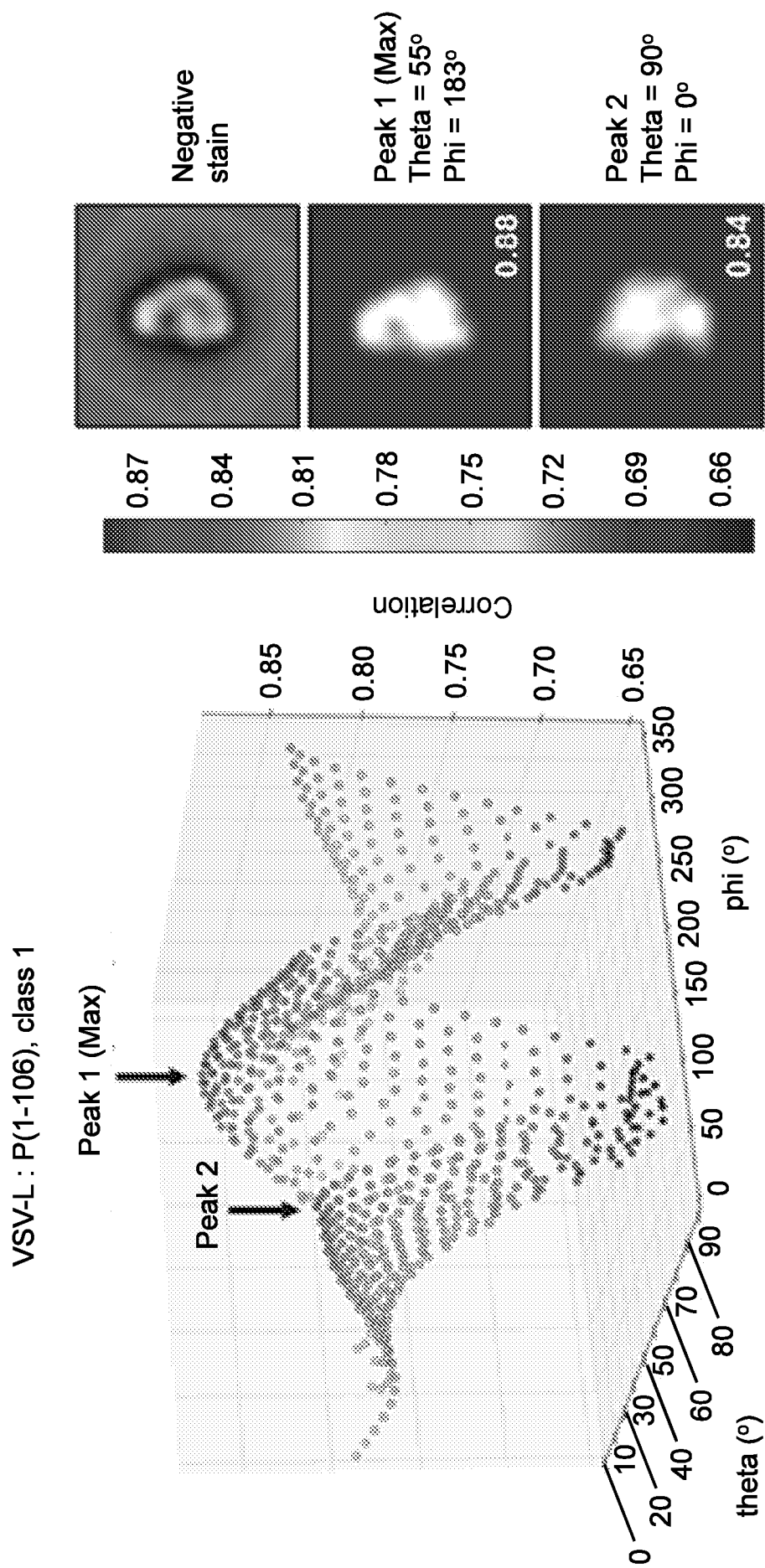


FIG. 9A

VSV-L: P(1-106), class 3

Peak 1 (Max)

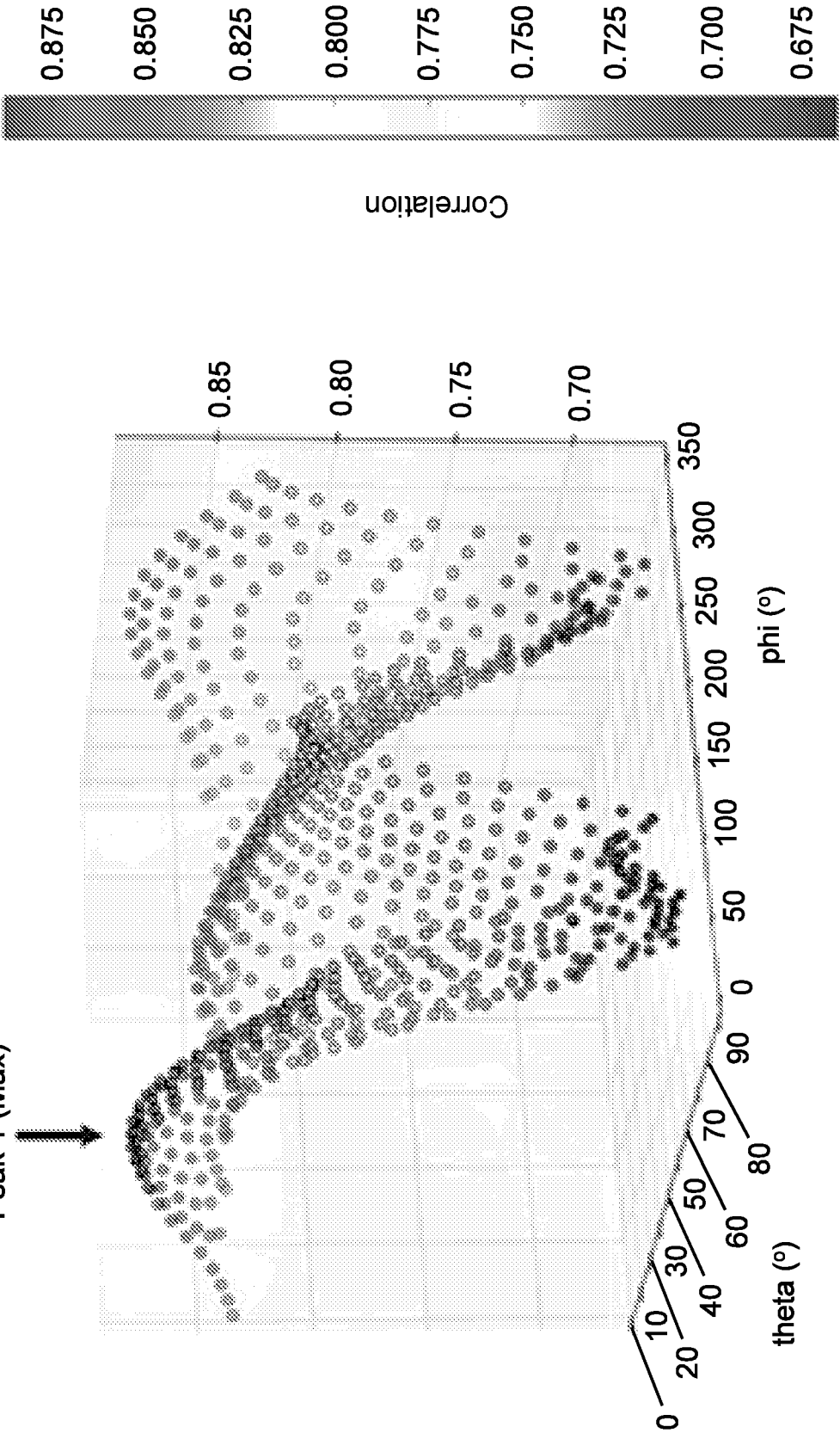


FIG. 9B

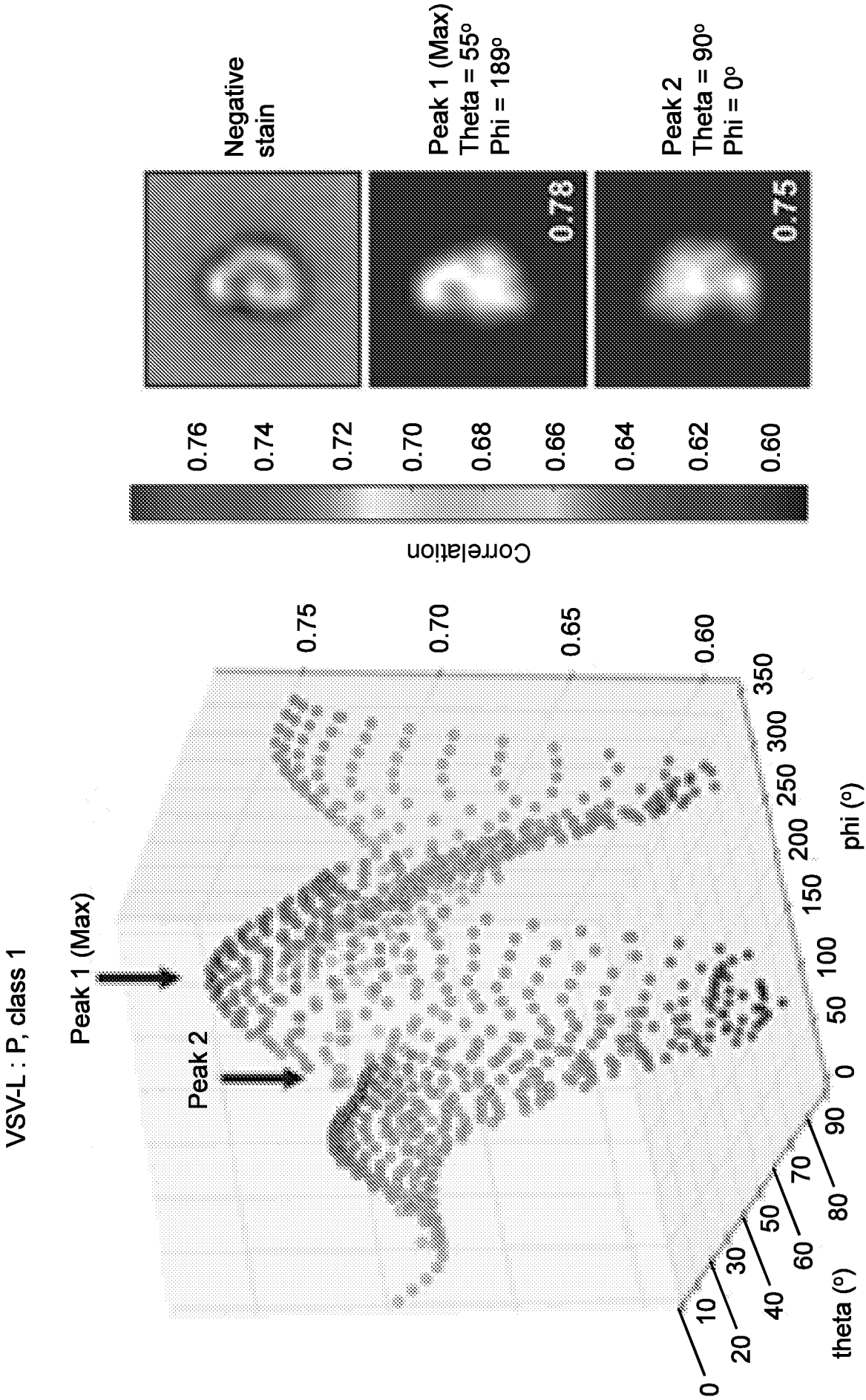


FIG. 9C

VSV-L : P, class 2

Peak 1 (Max)

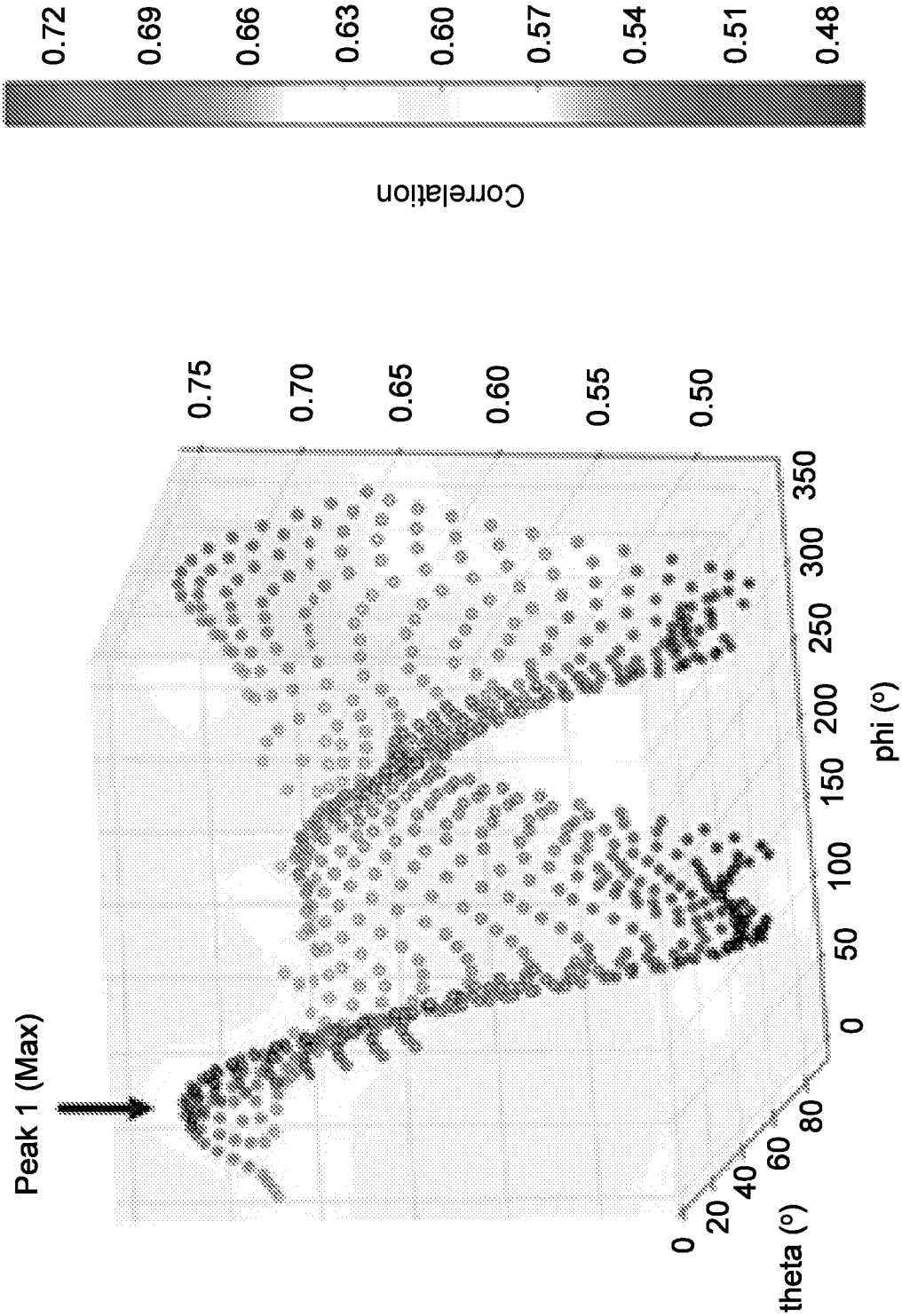


FIG. 9D

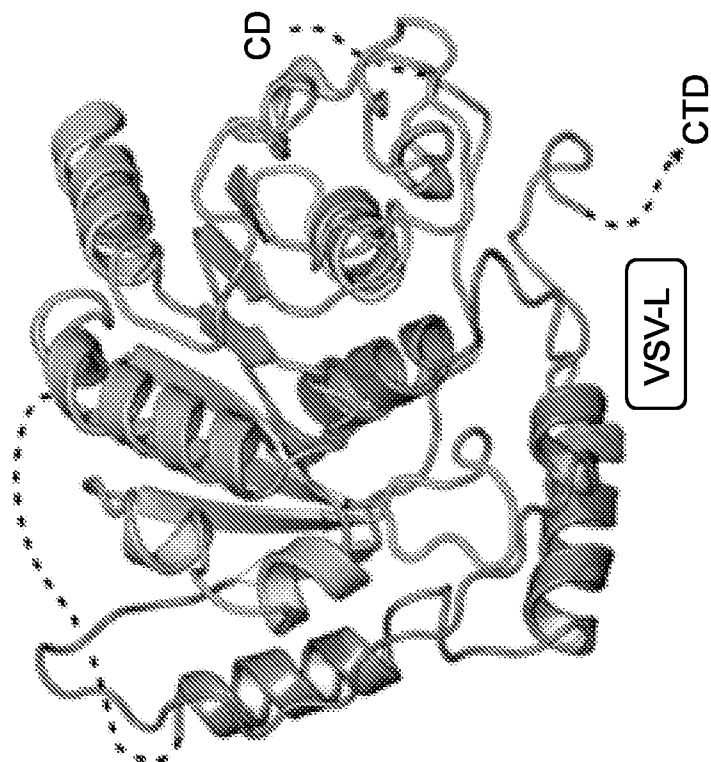


FIG. 10A

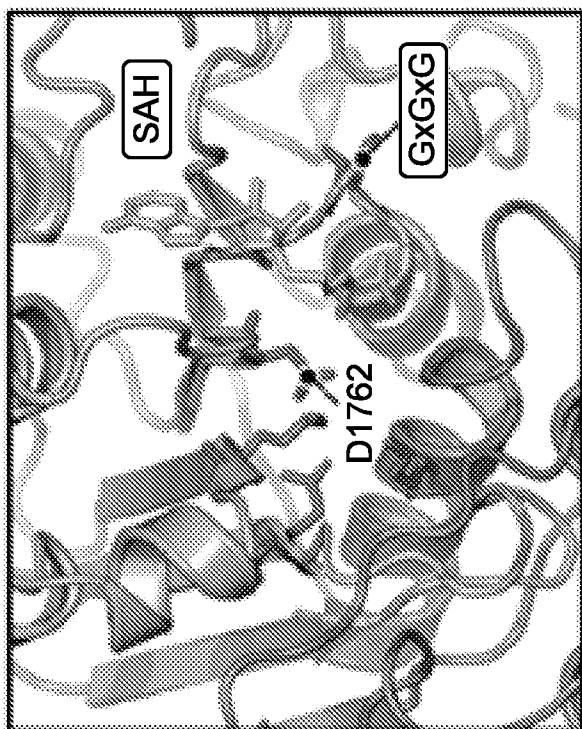
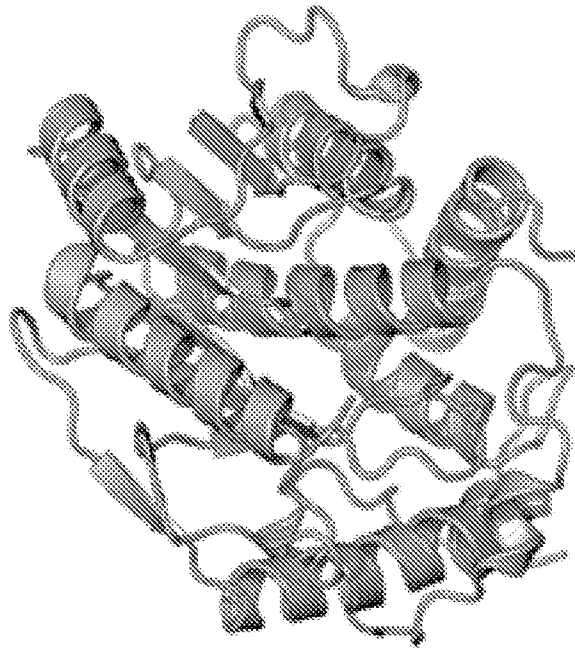
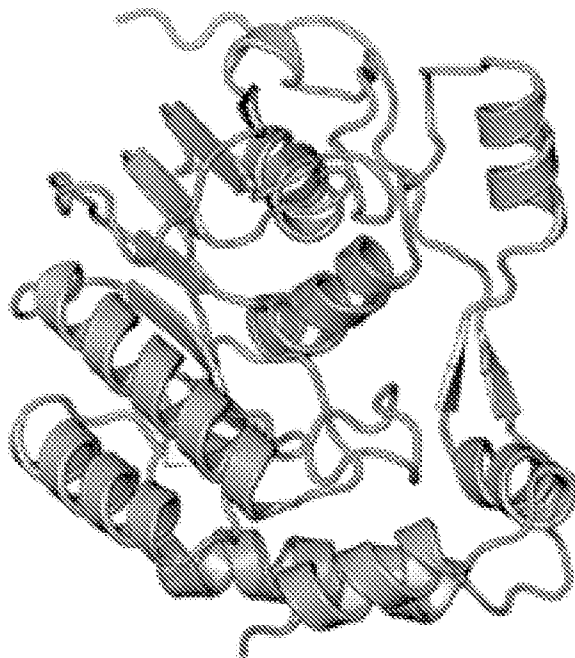


FIG. 10B

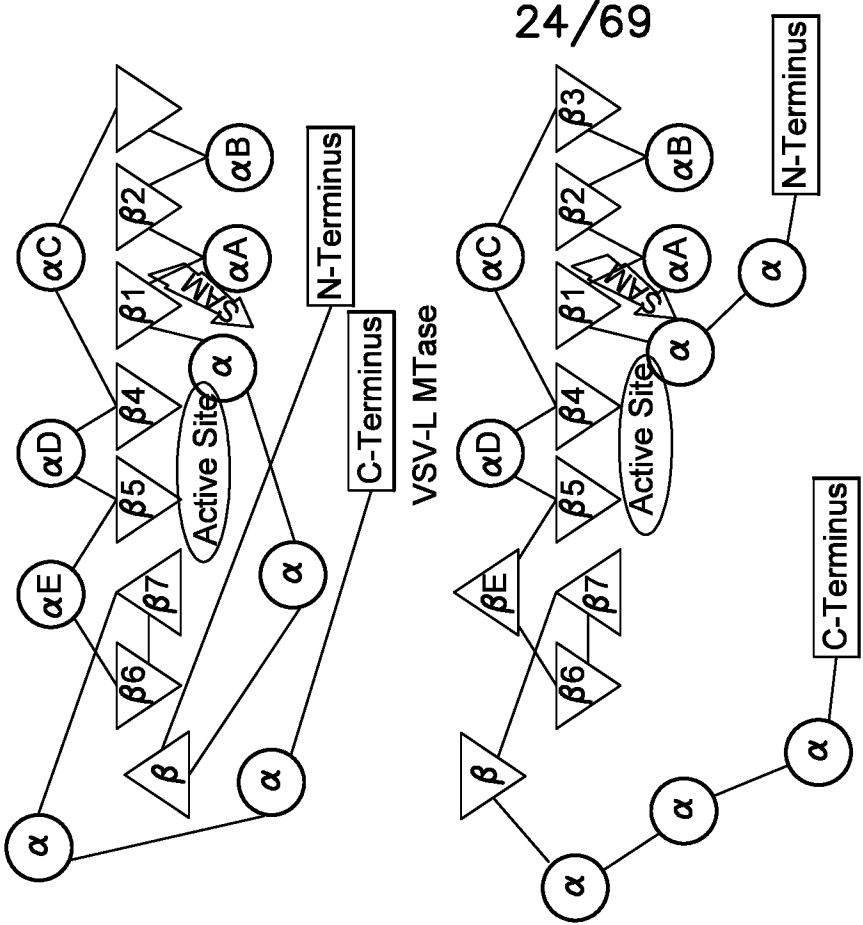


Vaccinia virus (1av6)



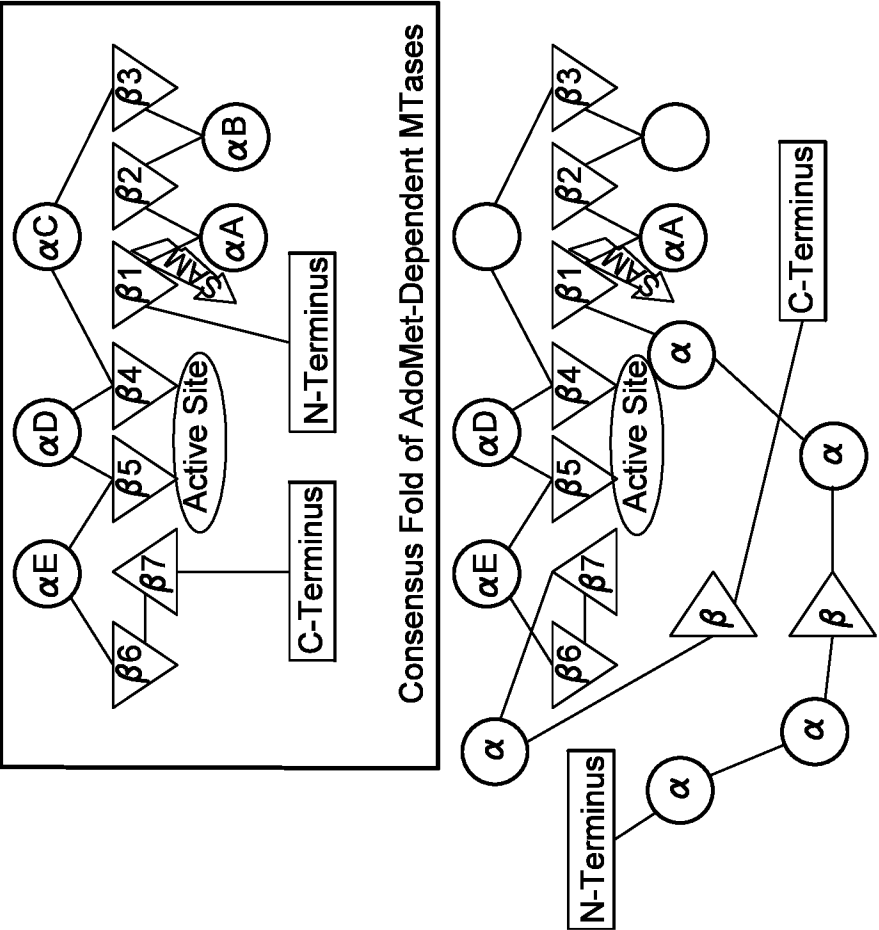
Dengue virus (1l9k)

FIG. 10C



Vaccinia Virus VP39 MTase

FIG. 11B



Flavivirus NS5 MTase

FIG. 11A

25/69

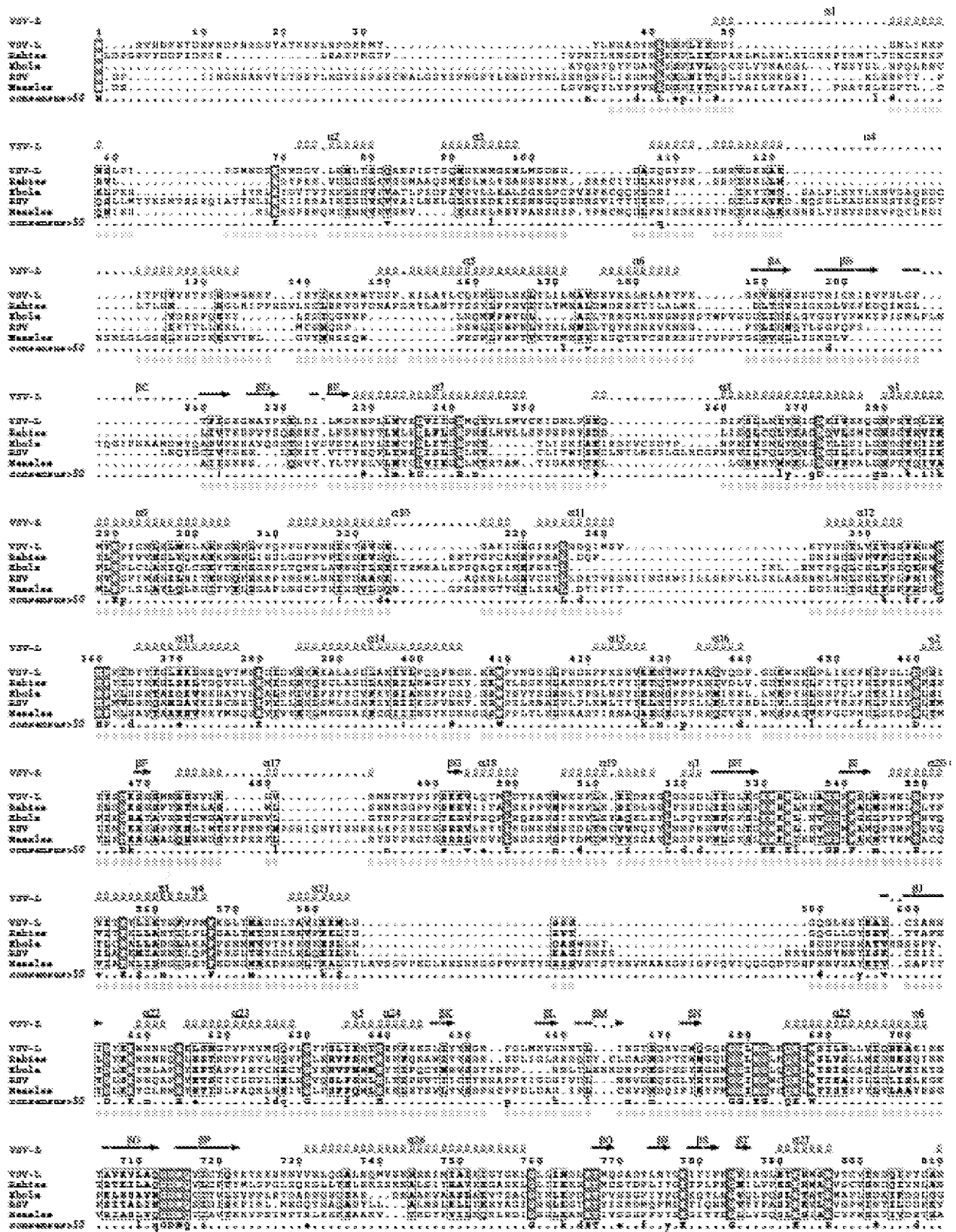
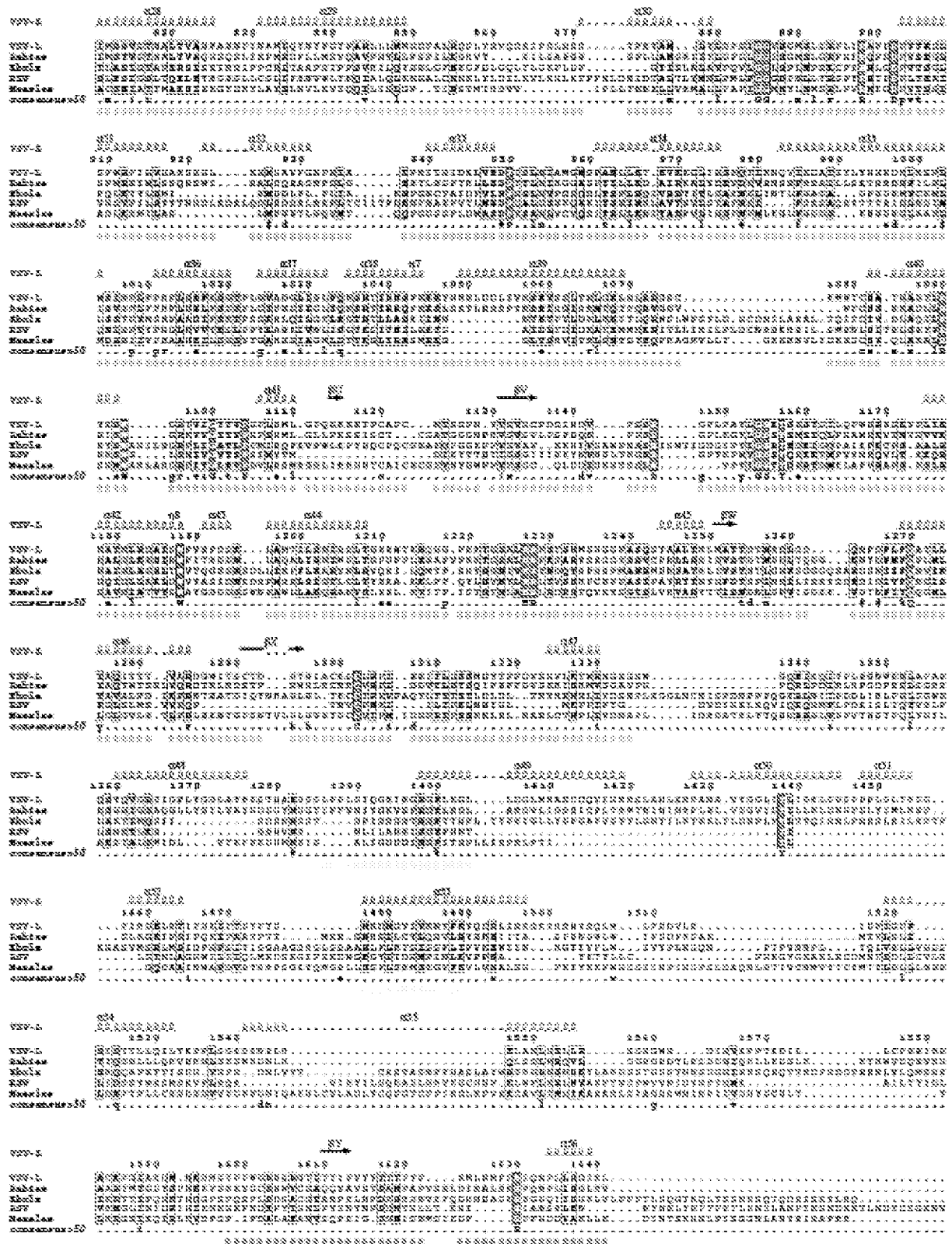


FIG. 12



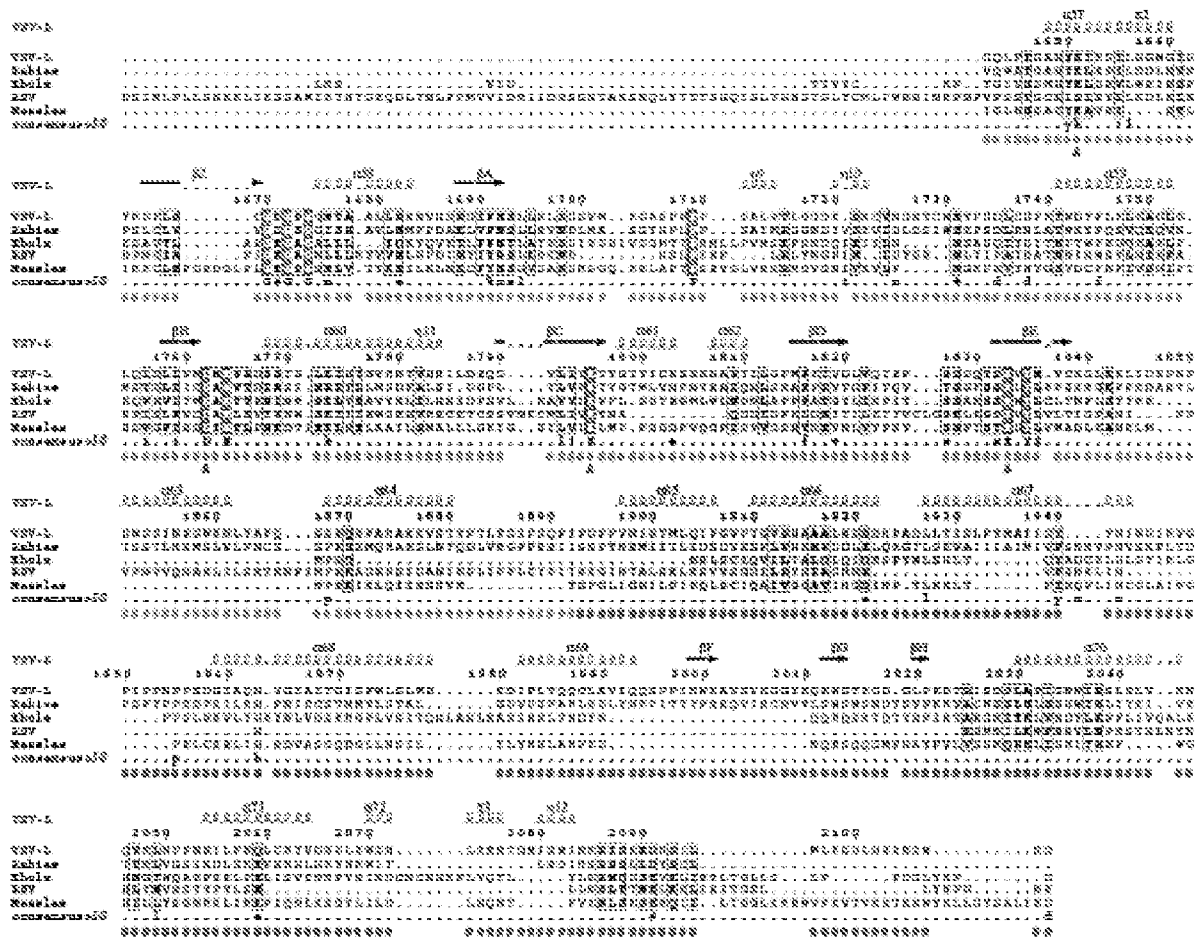


FIG. 12 Cont'd

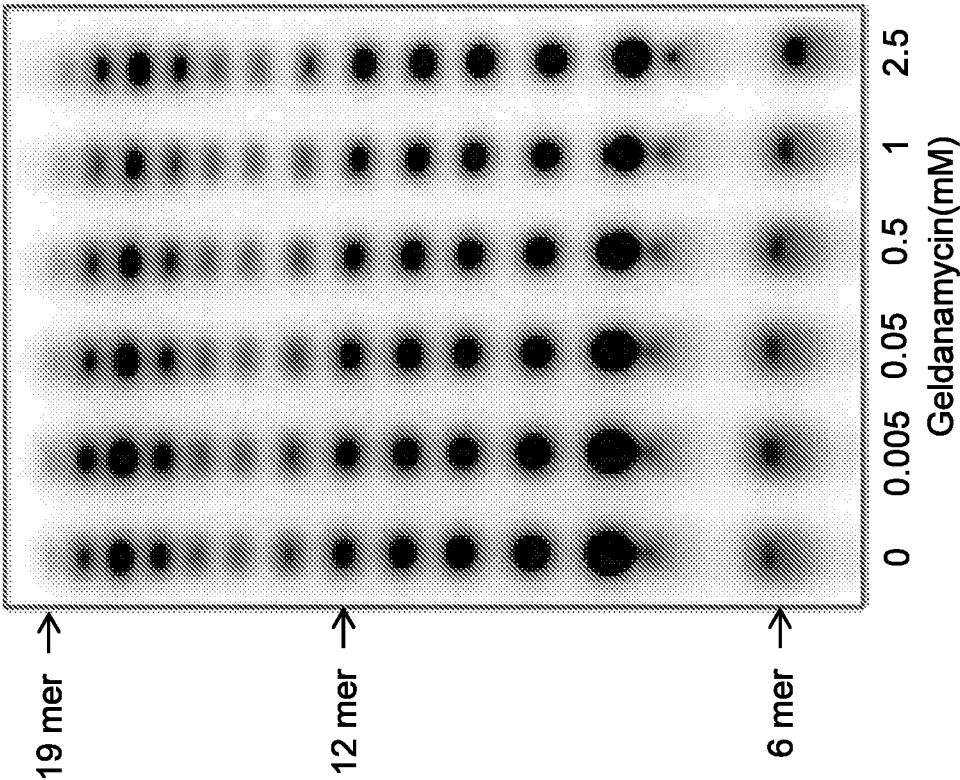


FIG. 13C

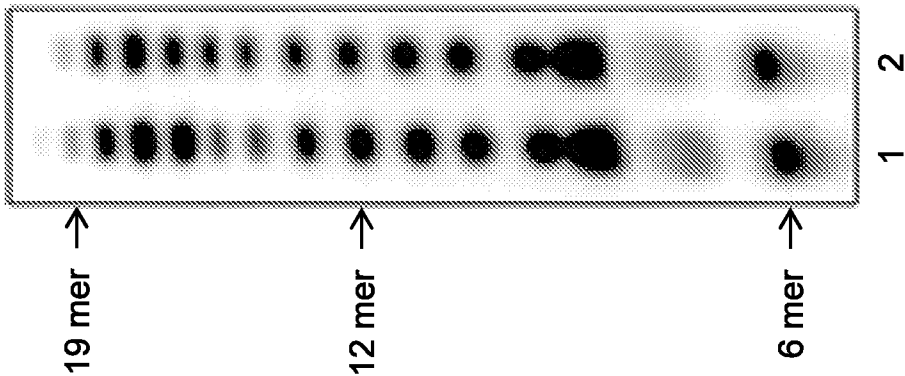


FIG. 13B

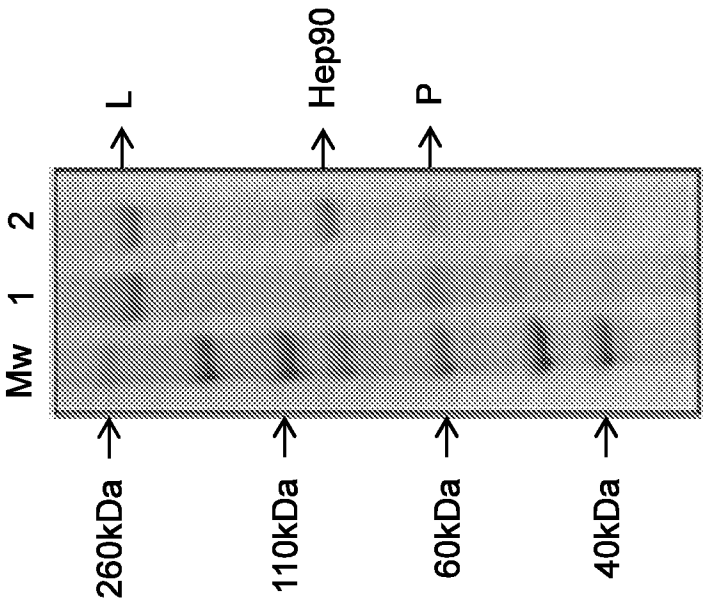
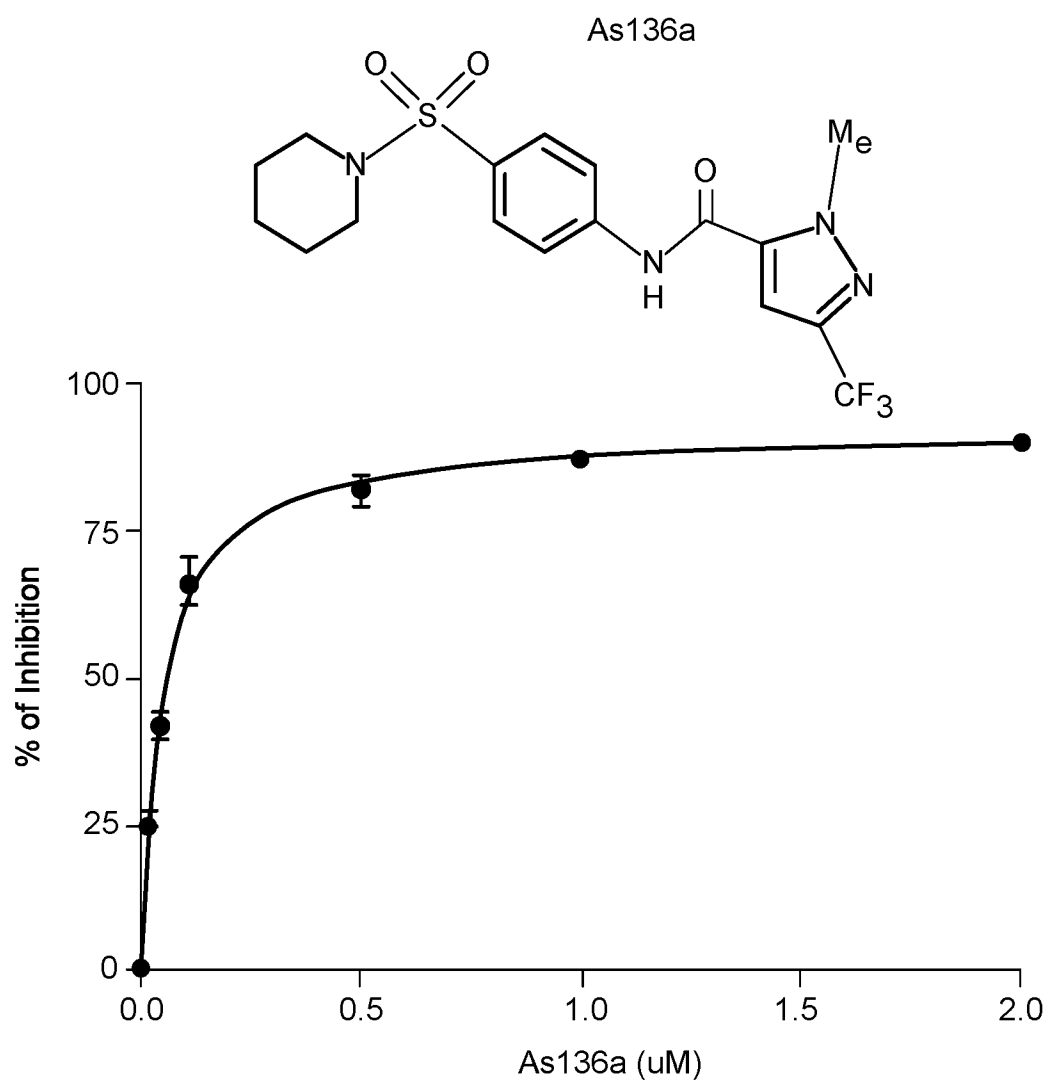
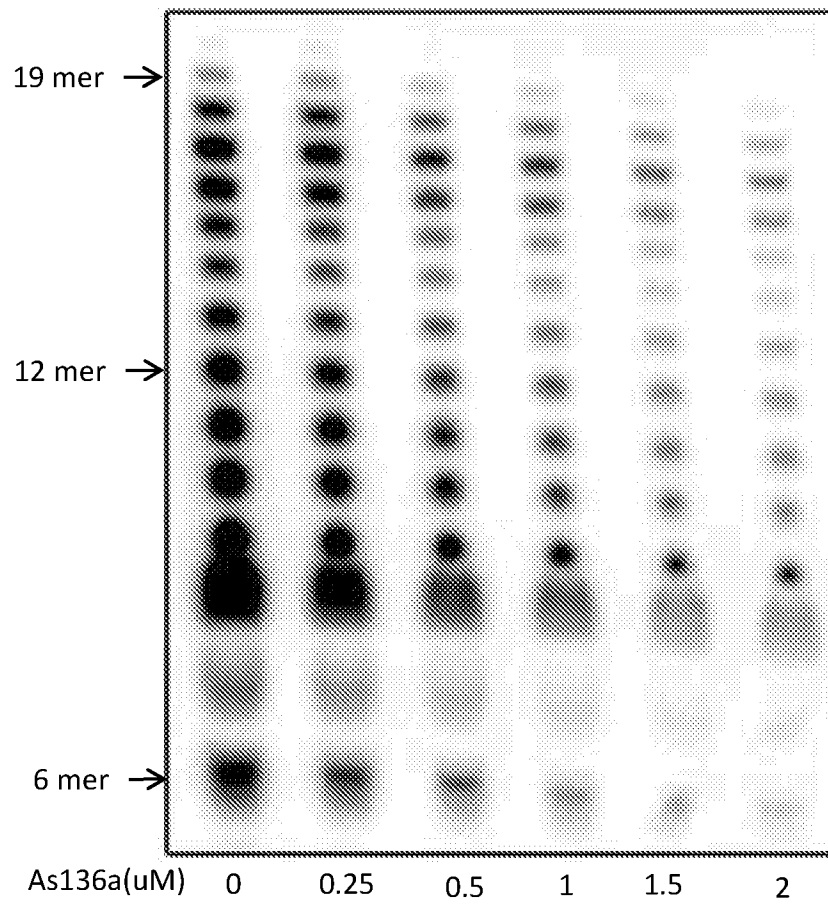


FIG. 13A

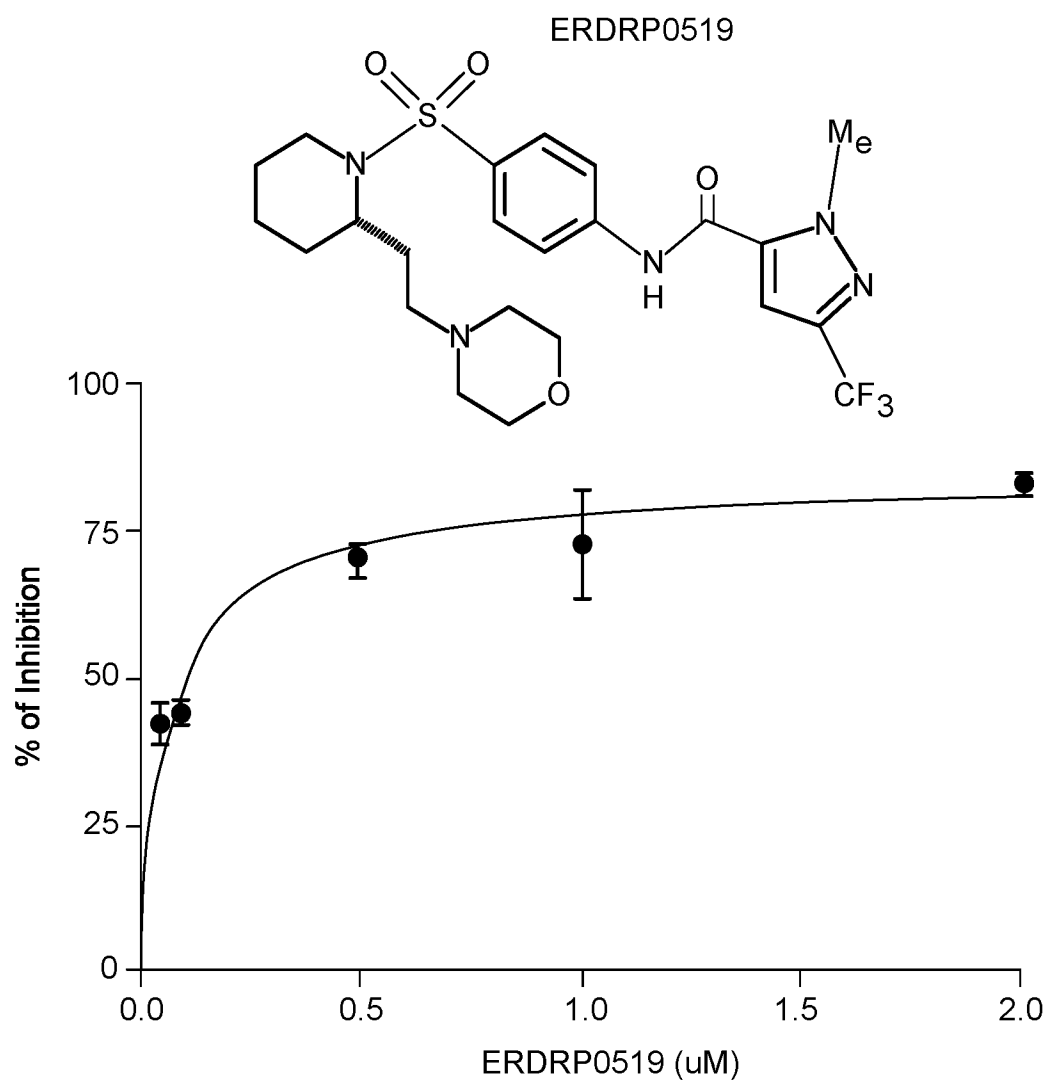
29/69

**FIG. 14A**

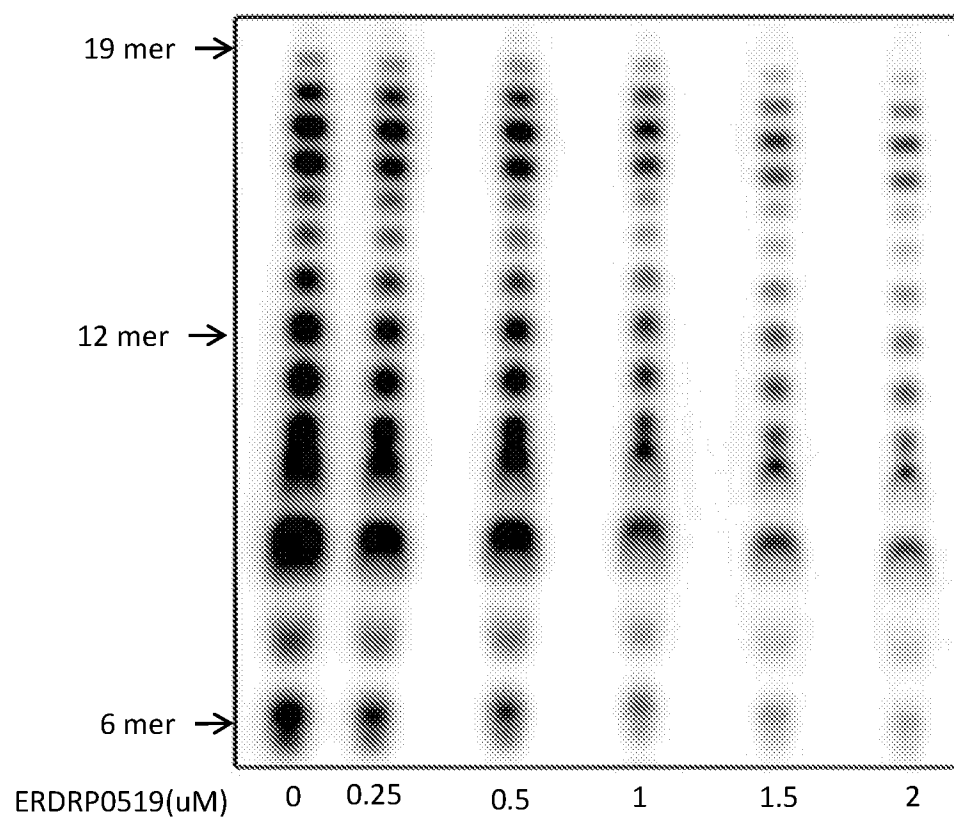
30/69

*FIG. 14B*

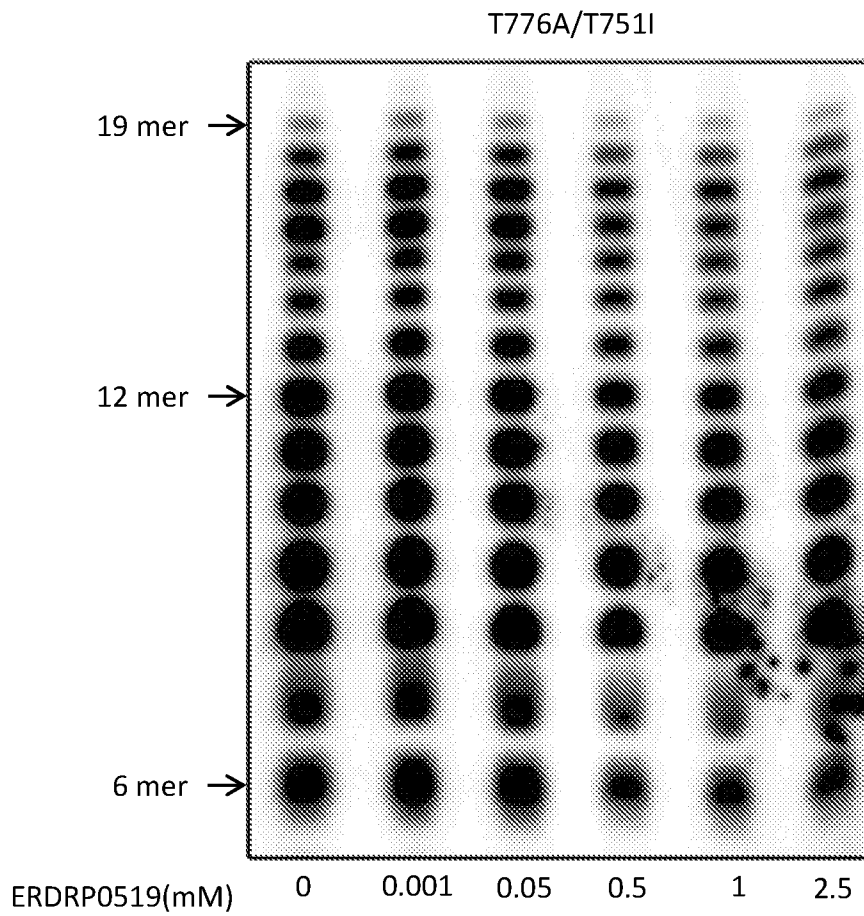
31/69

**FIG. 15A**

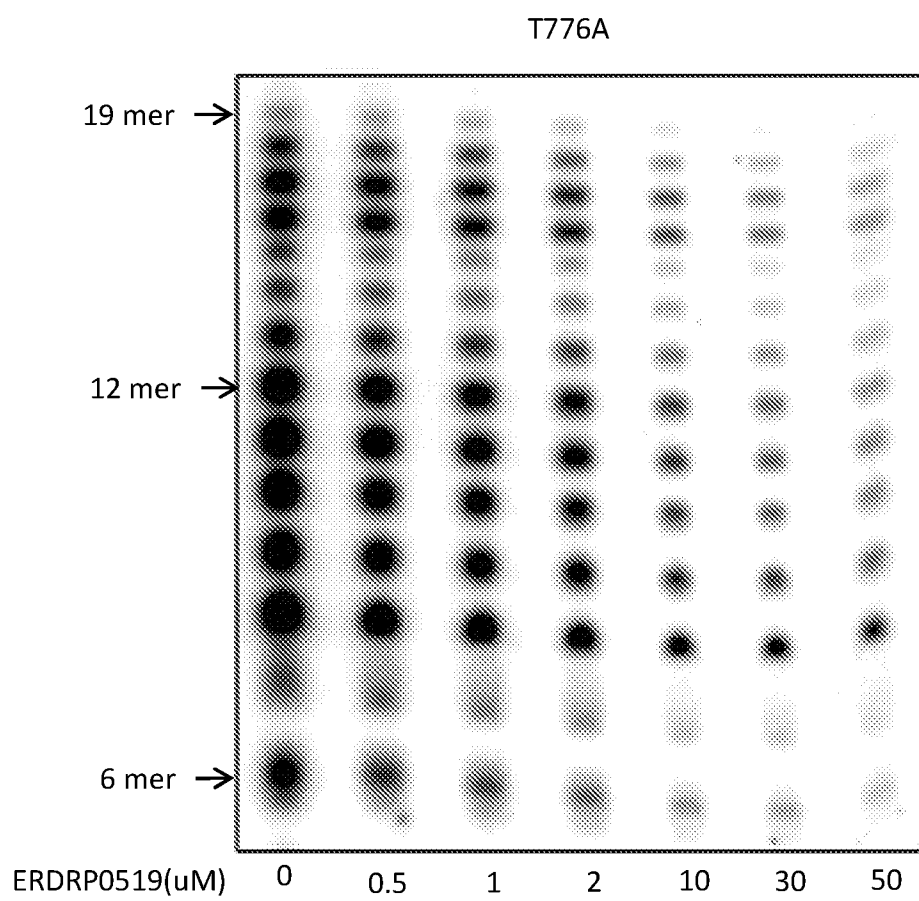
32/69

*FIG. 15B*

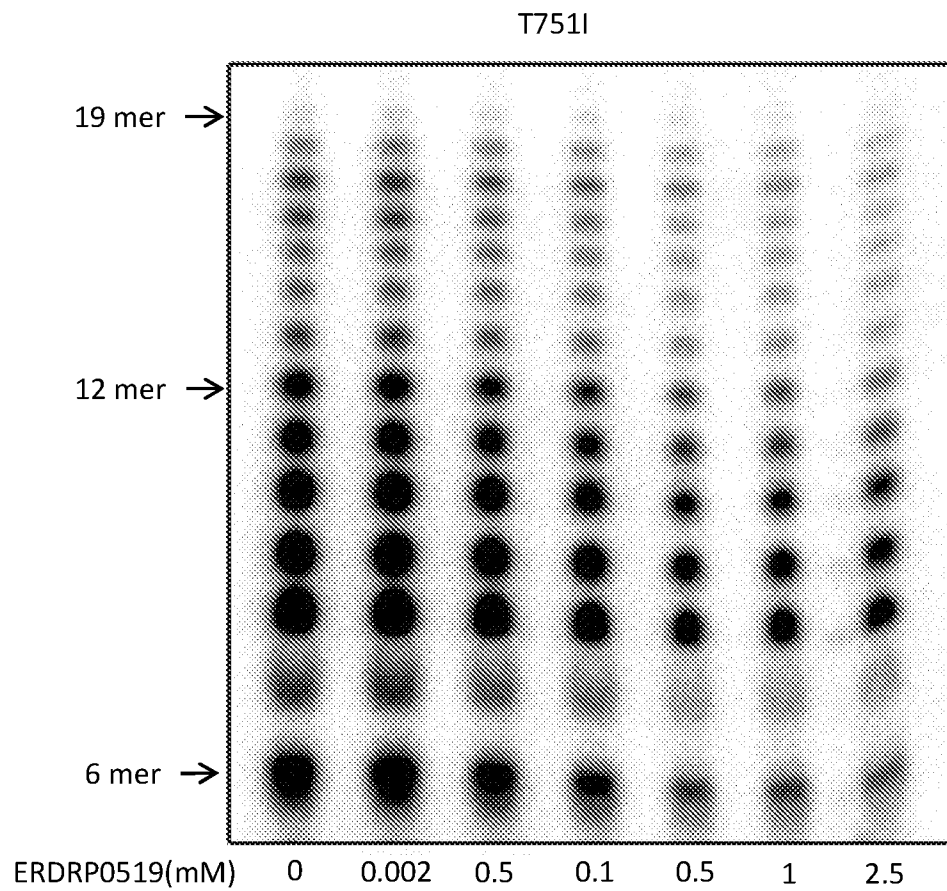
33/69

*FIG. 16A*

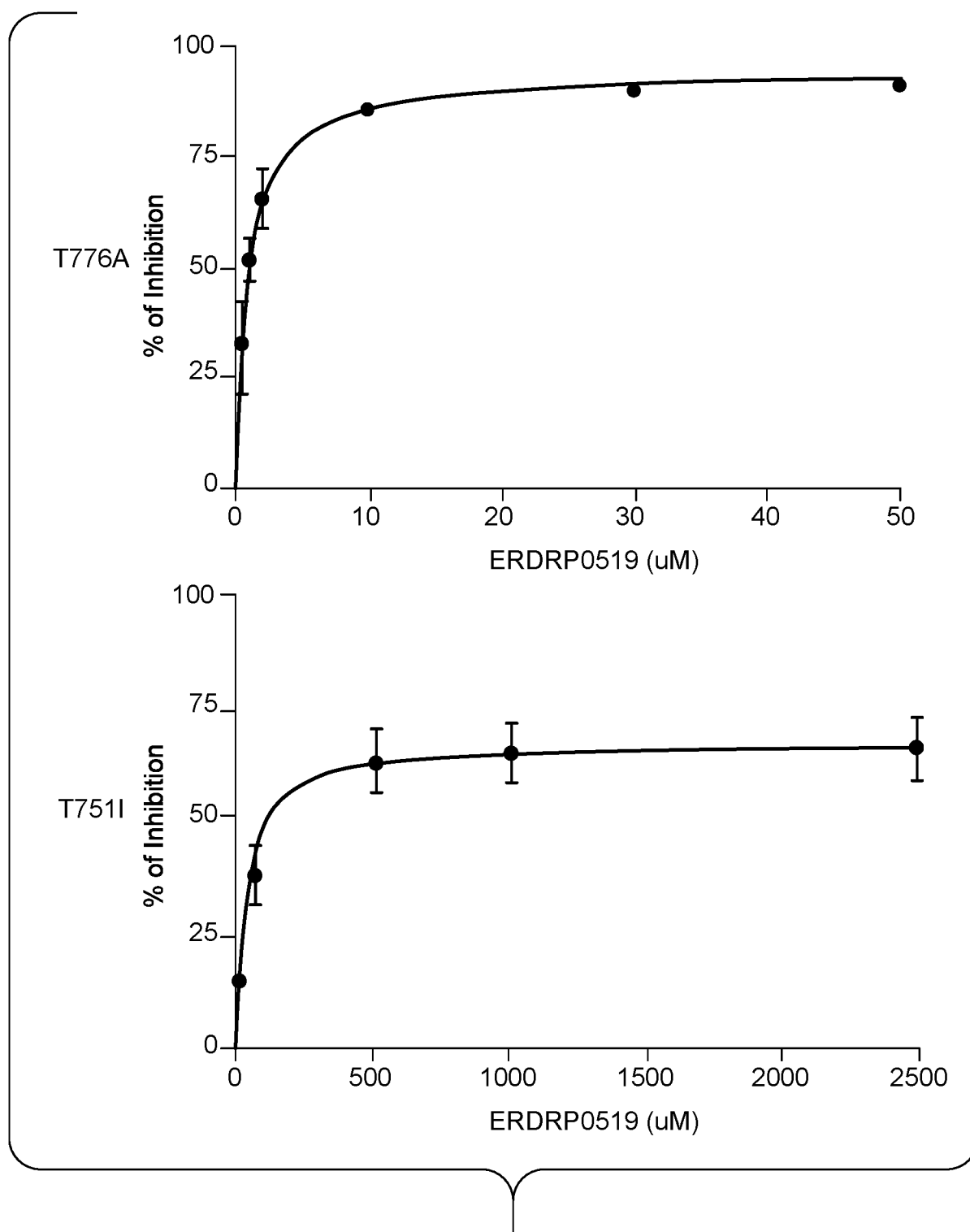
34/69

*FIG. 16B*

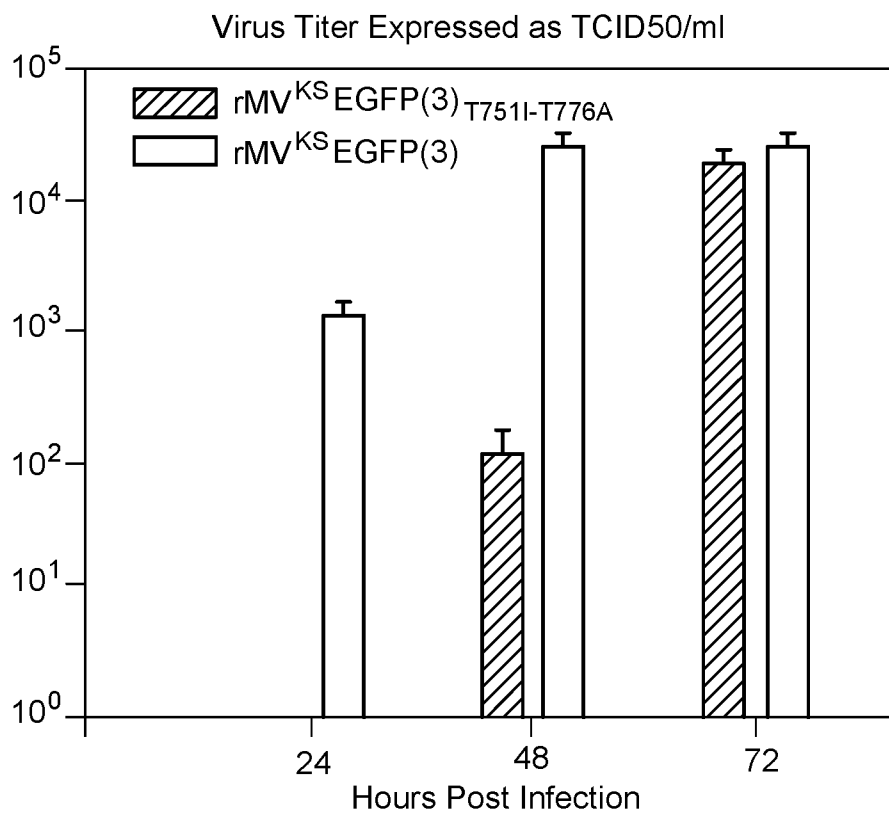
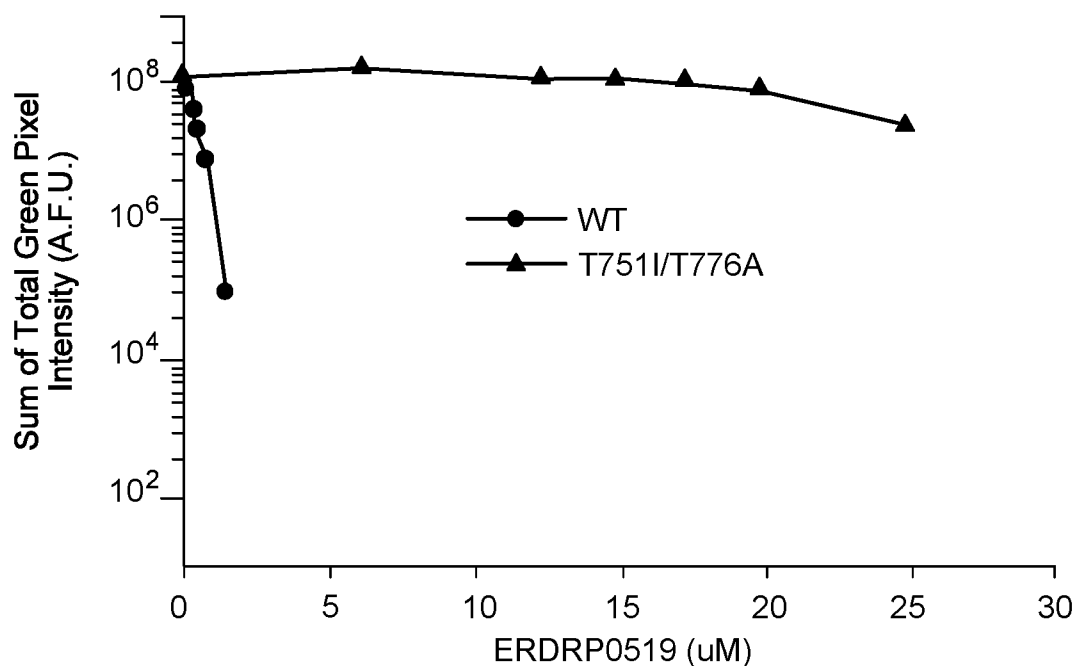
35/69

*FIG. 16C*

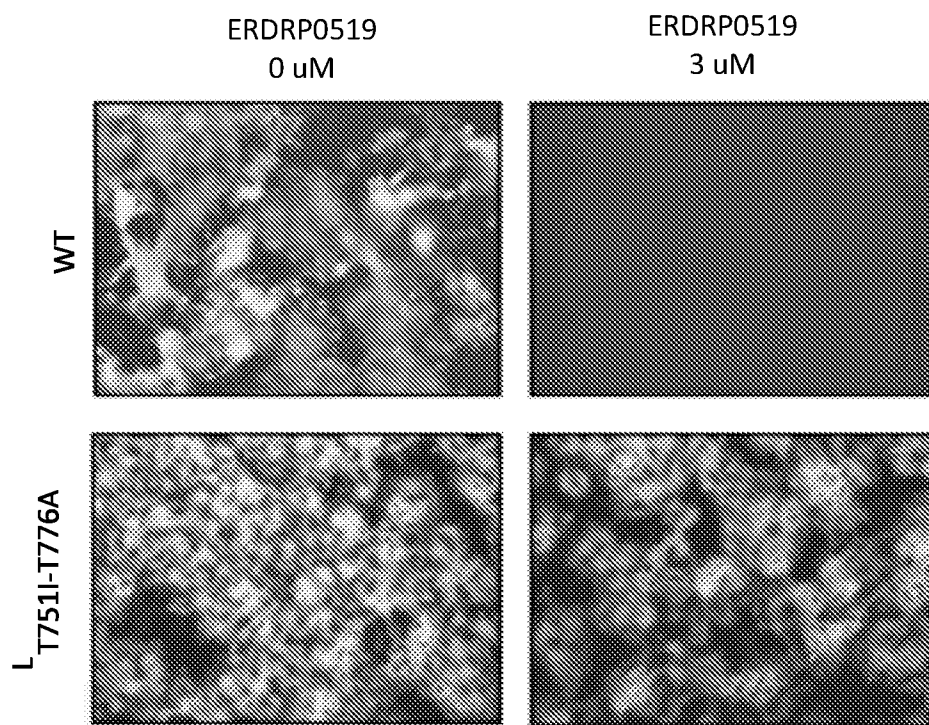
36/69

**FIG. 16D**

37/69

**FIG. 17A****FIG. 17C**

38/69

*FIG. 17B*

39/69

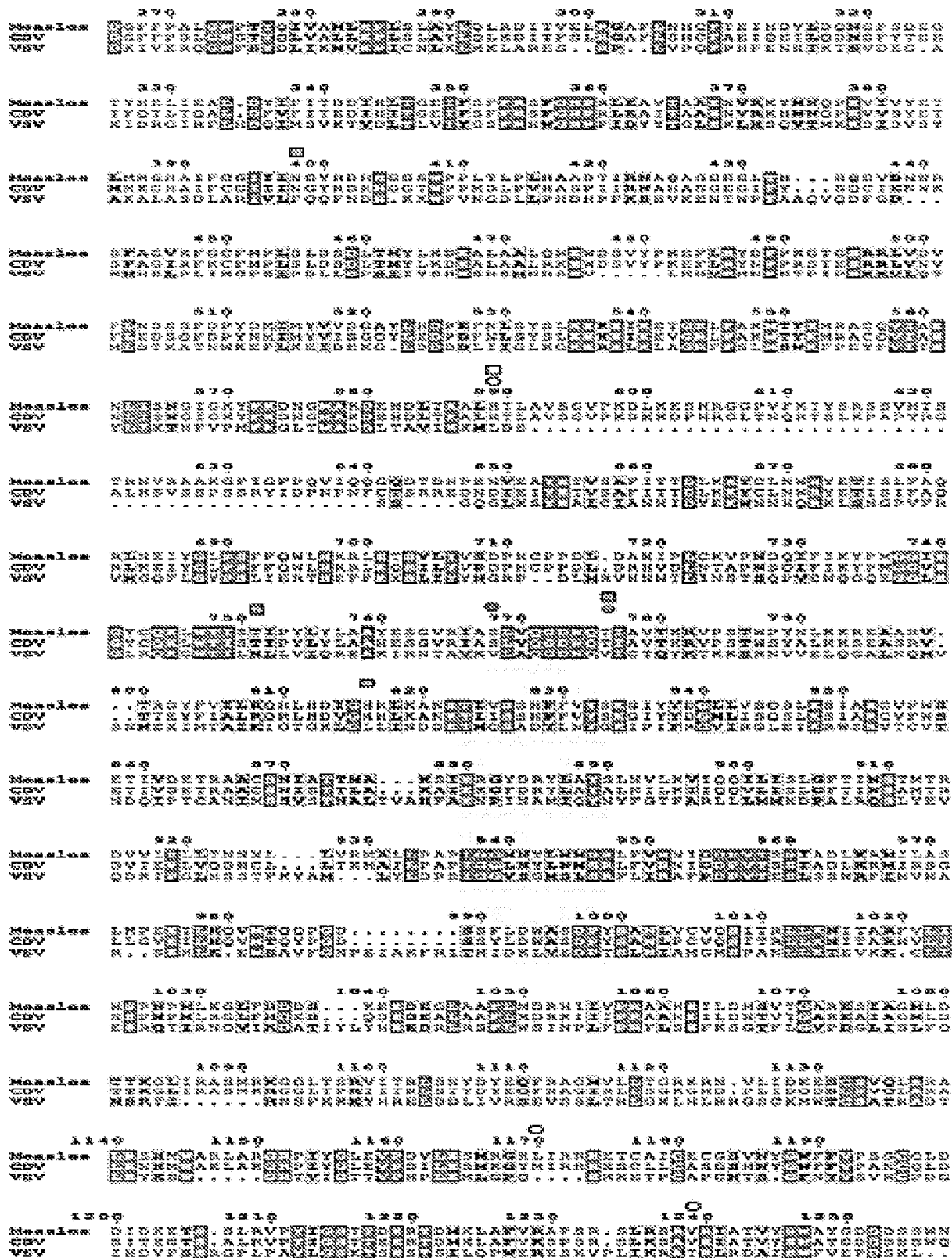


FIG. 18

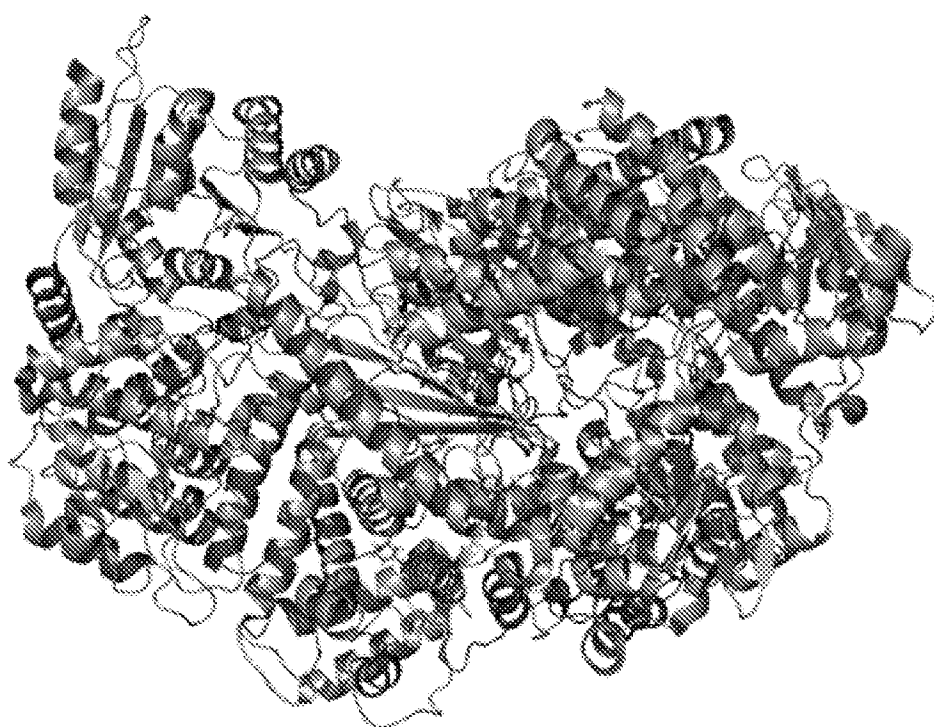


FIG. 19A

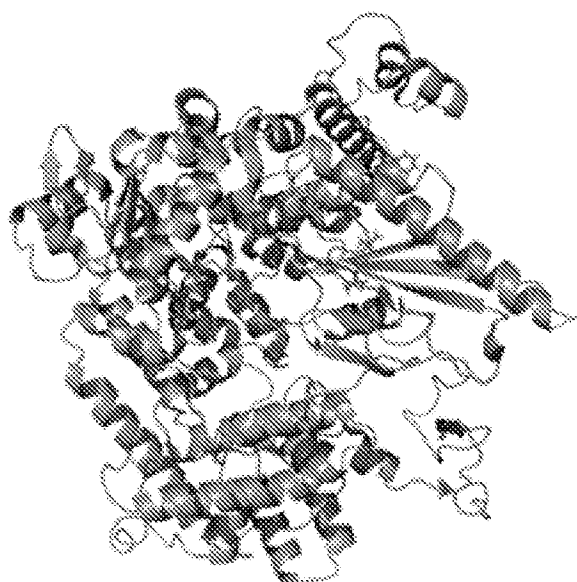


FIG. 19B

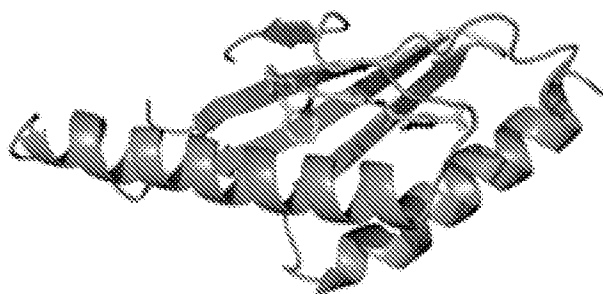
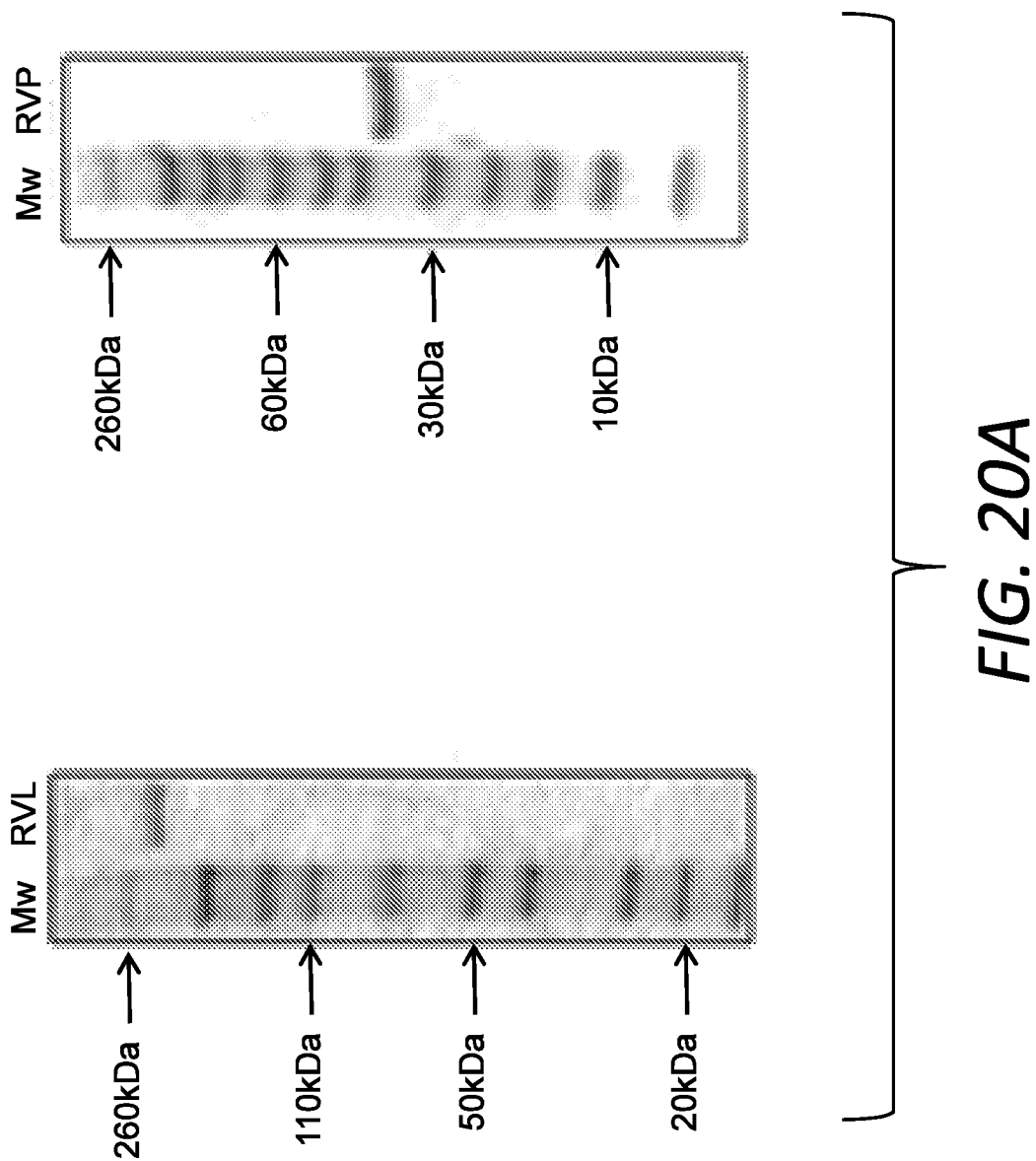


FIG. 19C



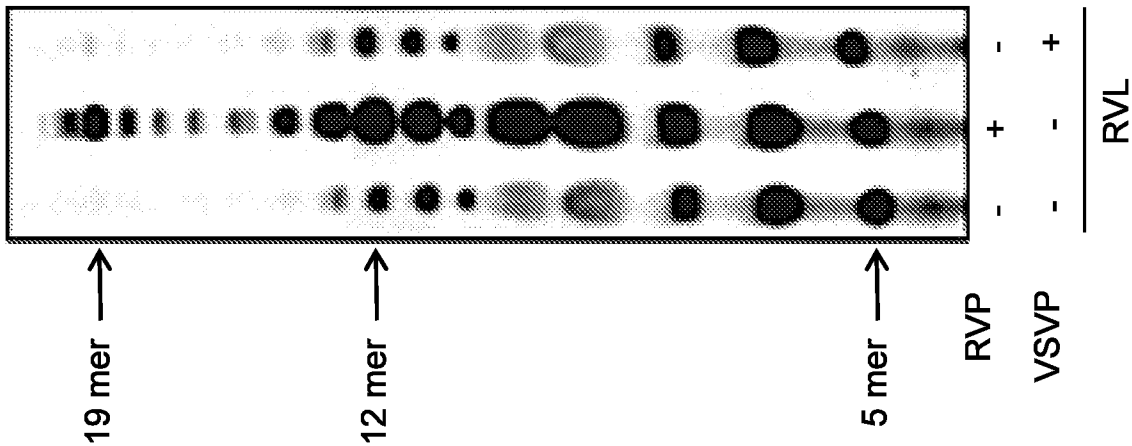


FIG. 20B

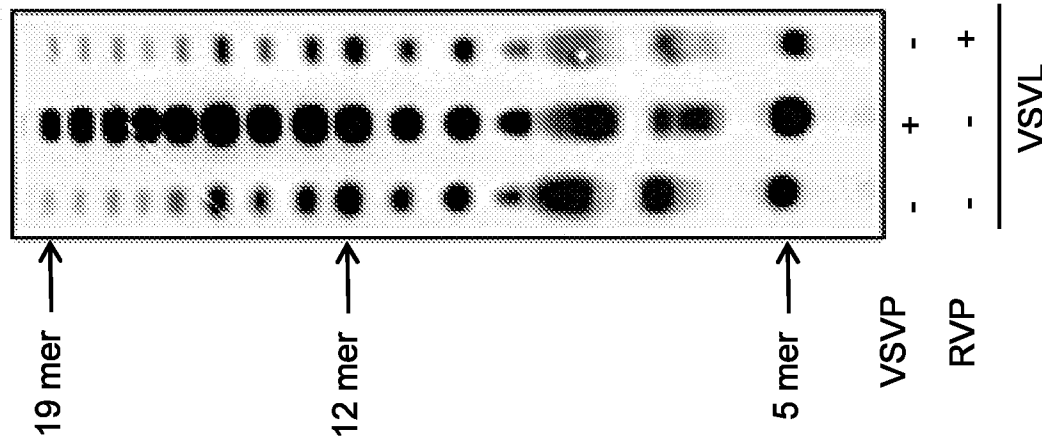
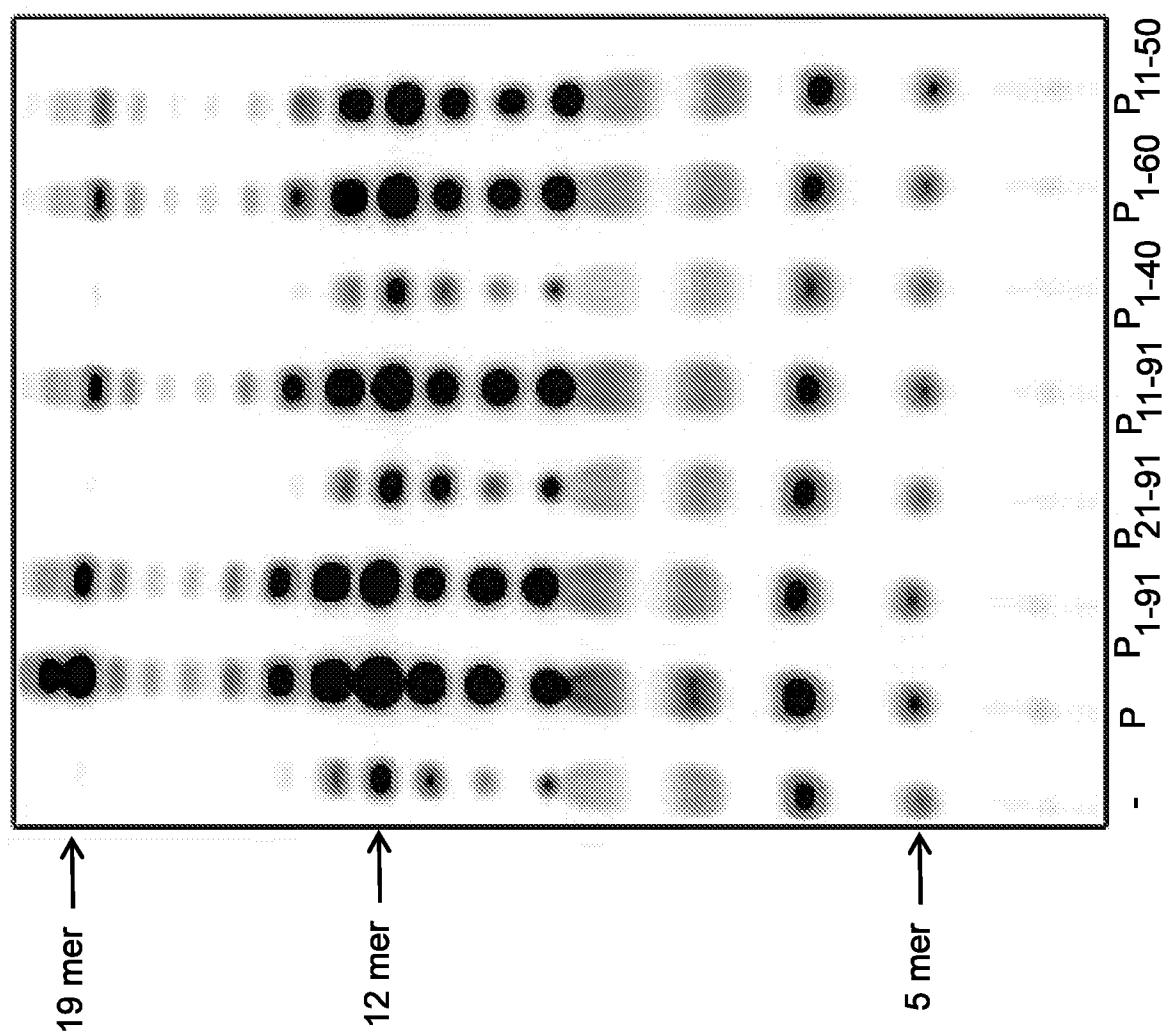


FIG. 20C

*FIG. 21*

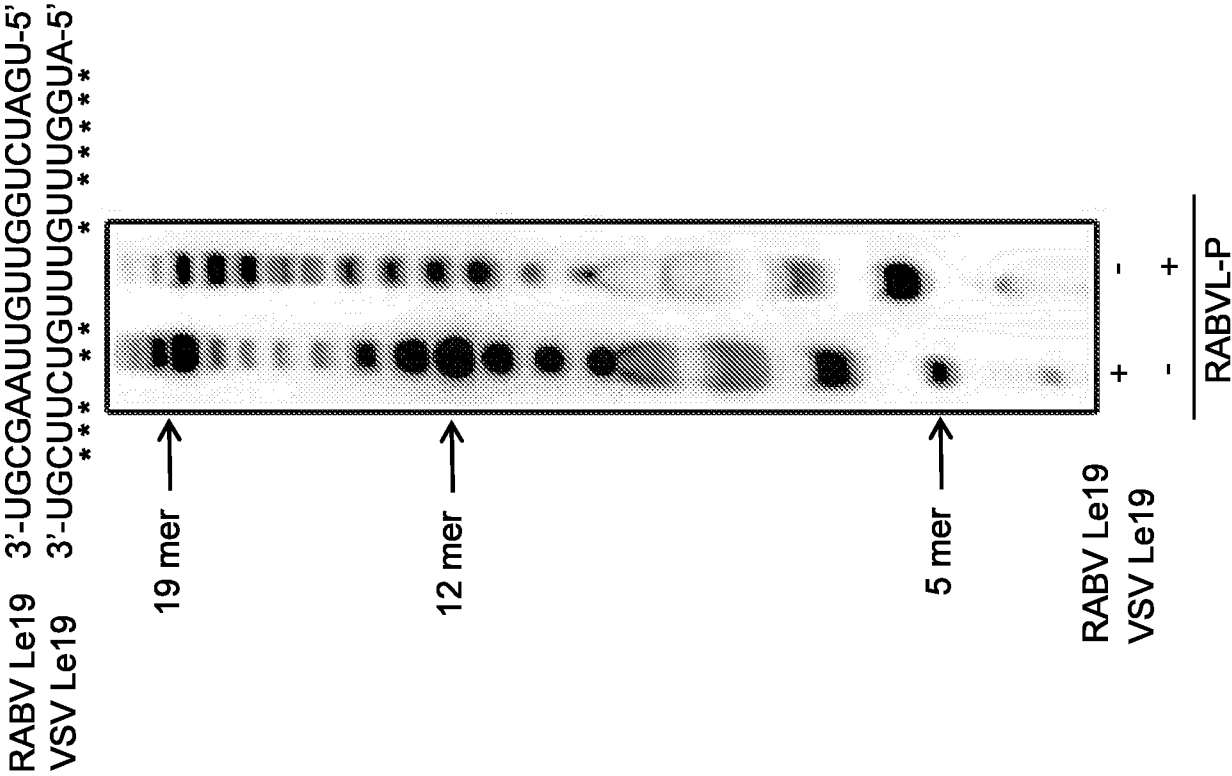


FIG. 22A

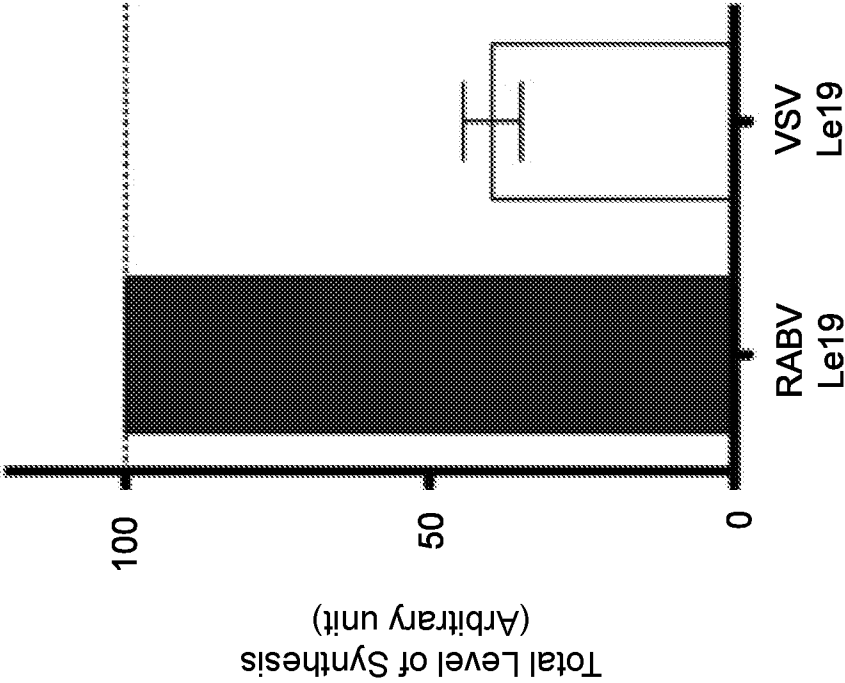
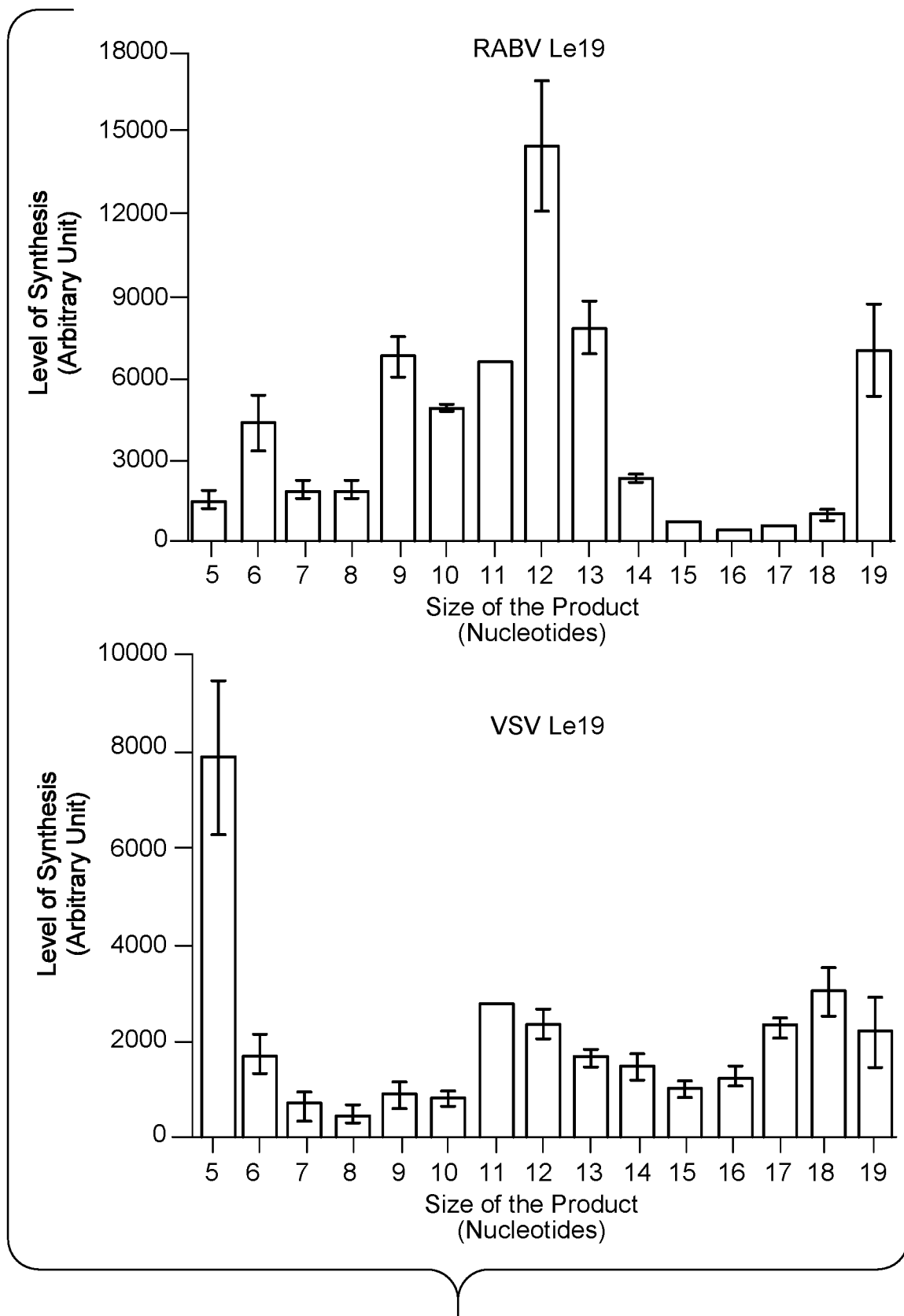
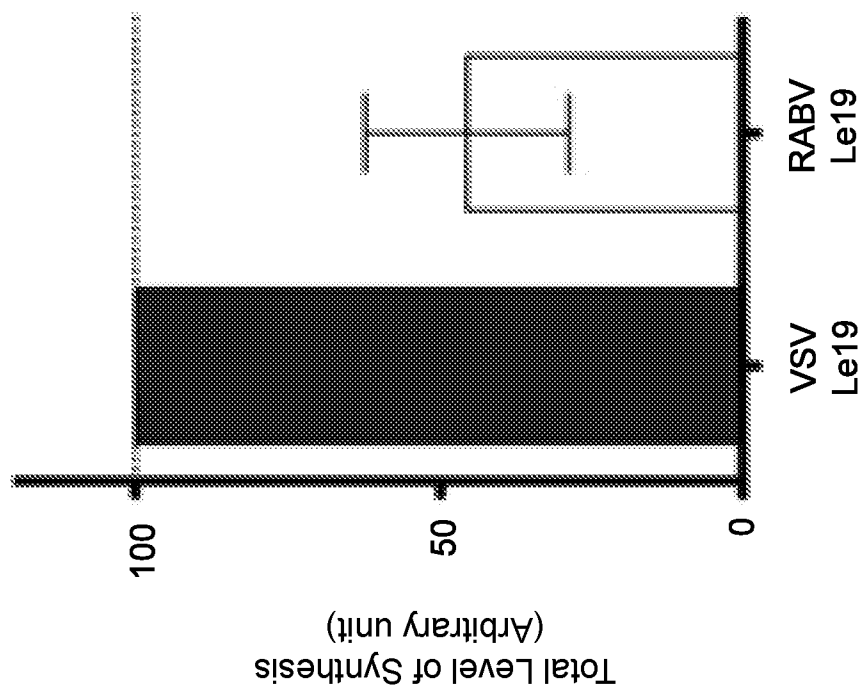
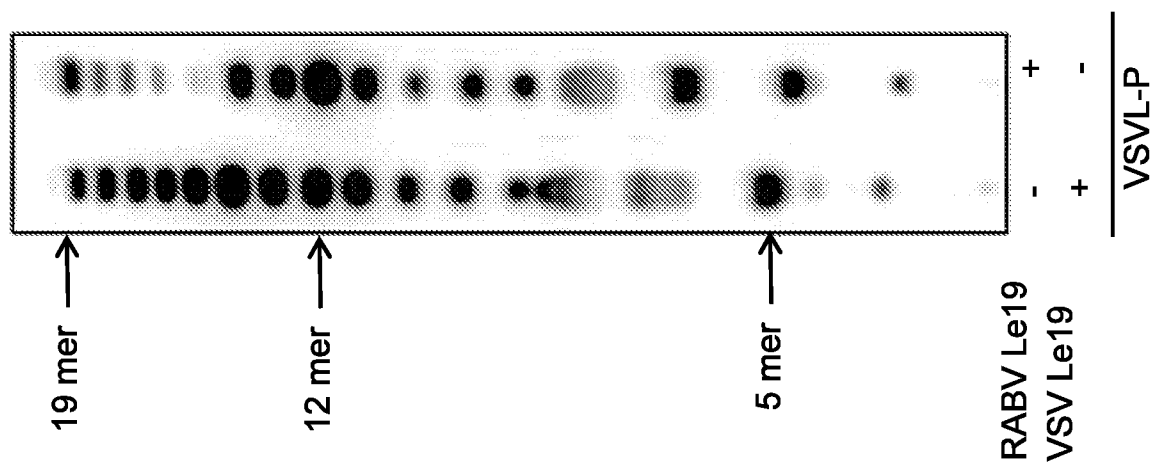
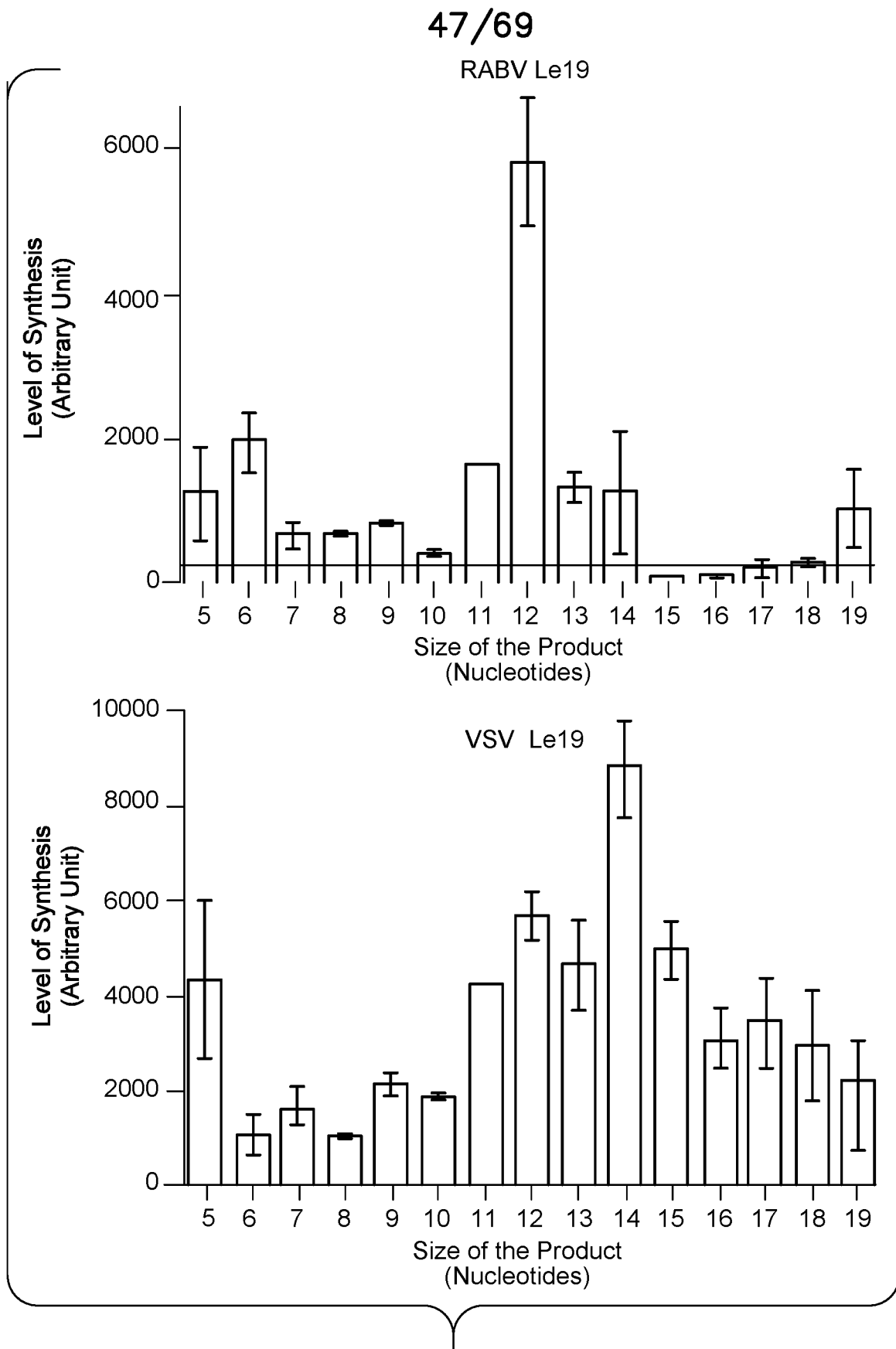


FIG. 22B

45/69

**FIG. 22C**



**FIG. 23C**

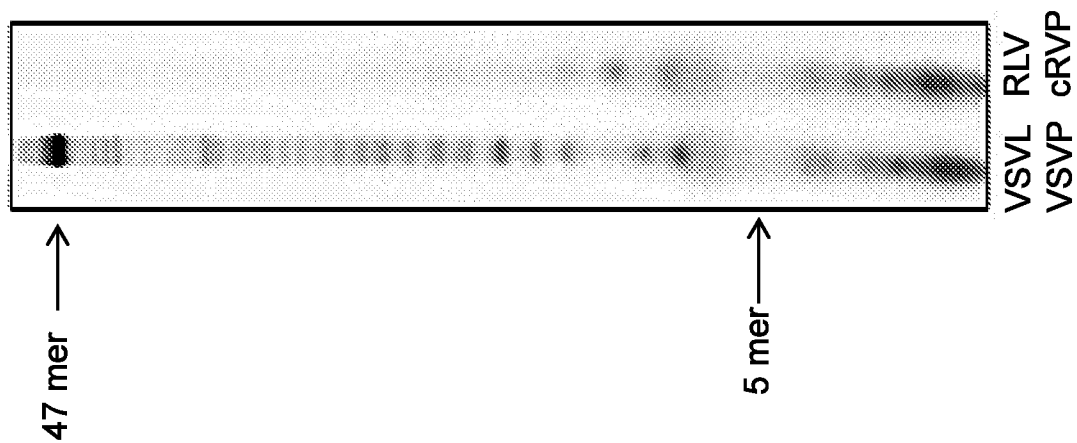


FIG. 24B

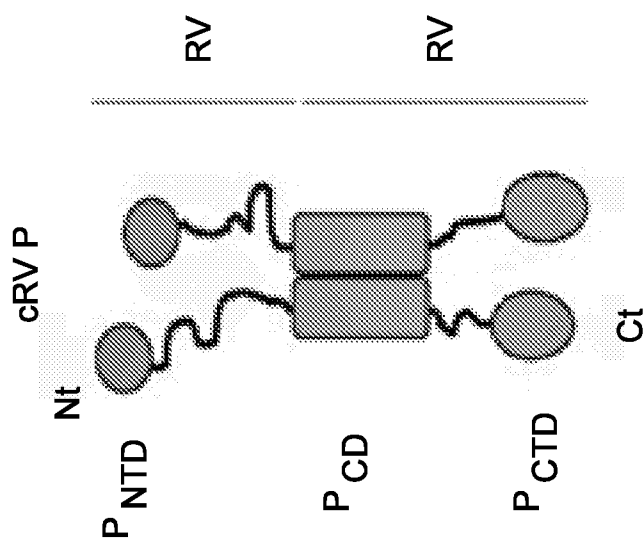


FIG. 24A

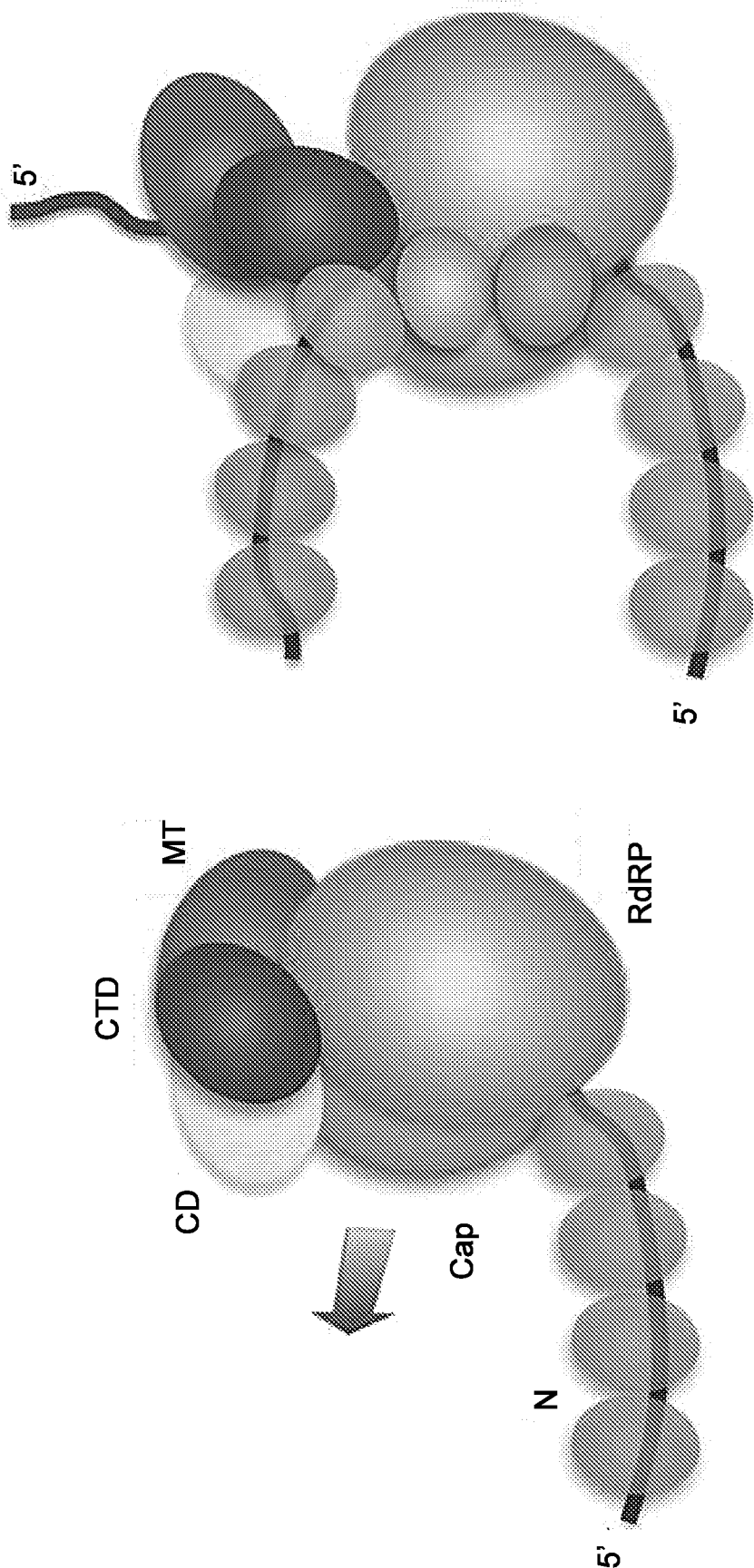


FIG. 25

50/69

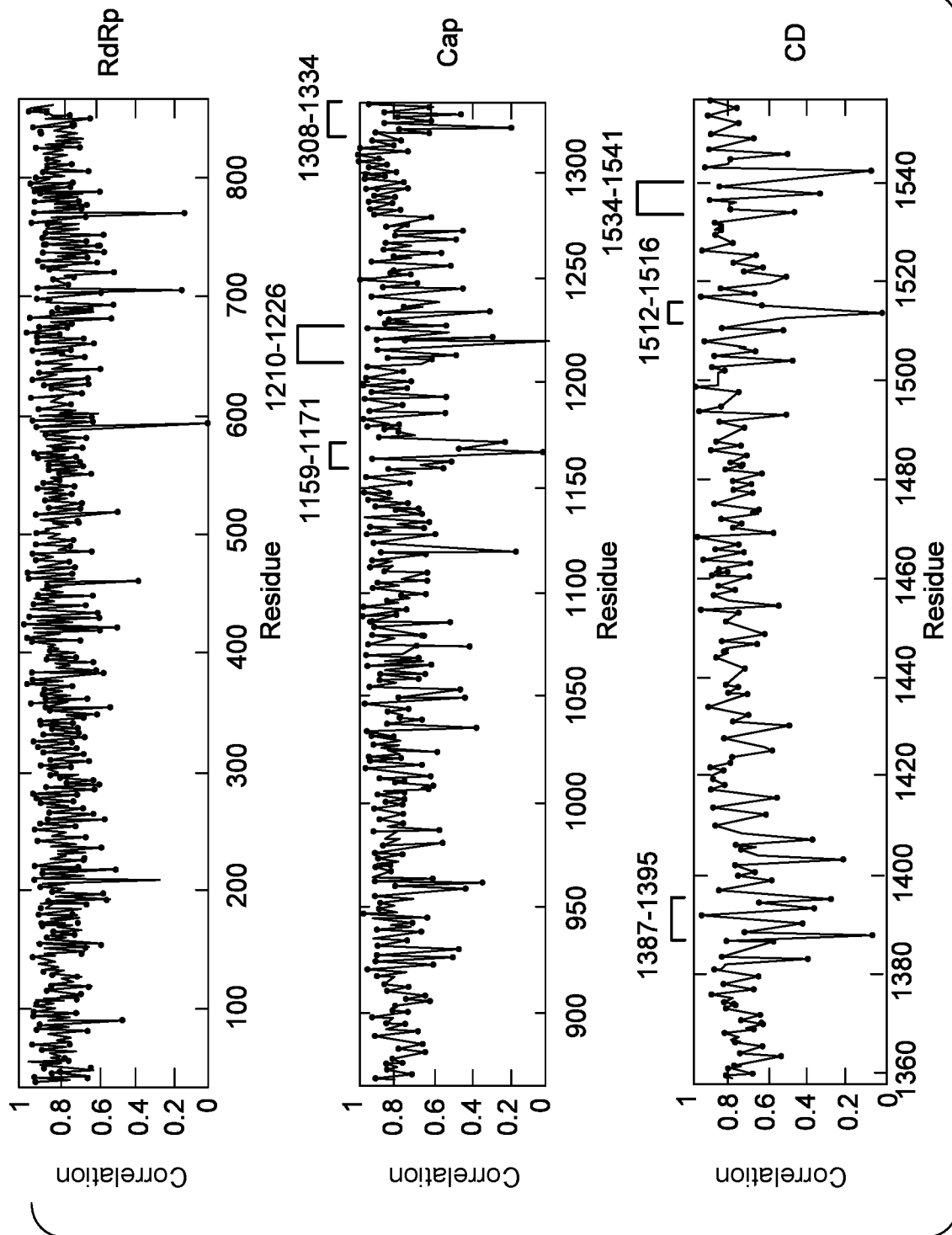


FIG. 26

51/69

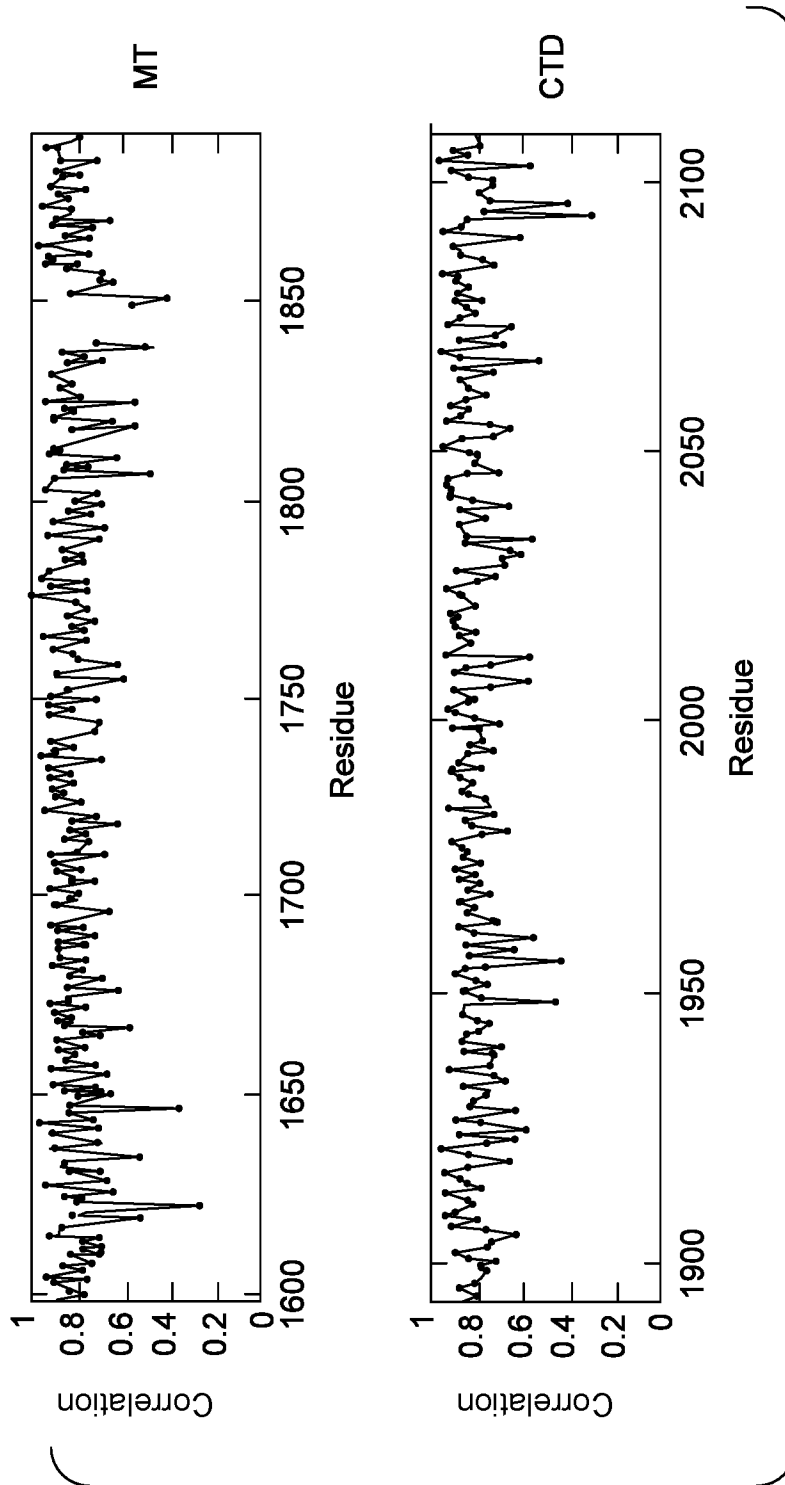


FIG. 26 Cont'd

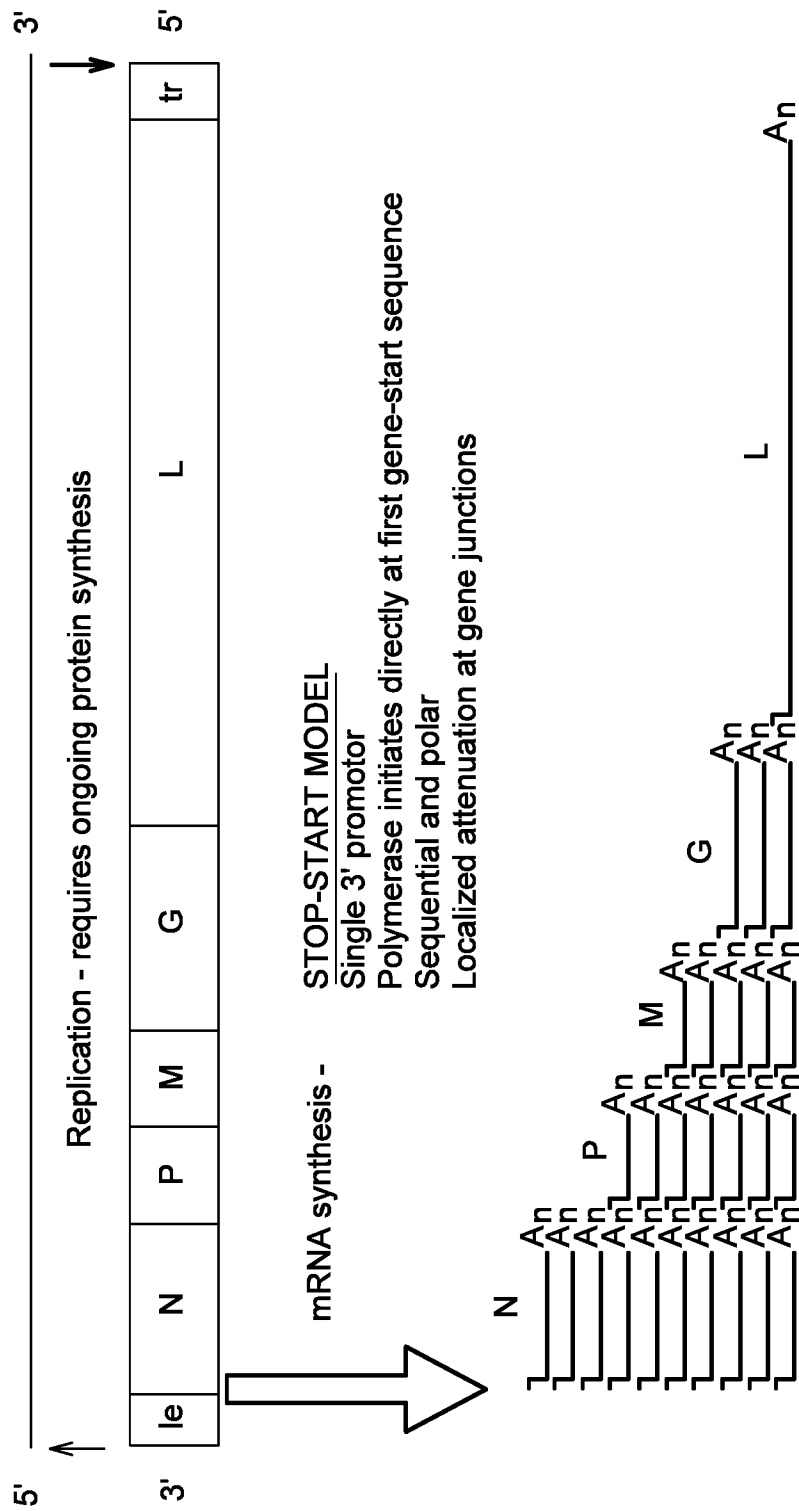


FIG. 27

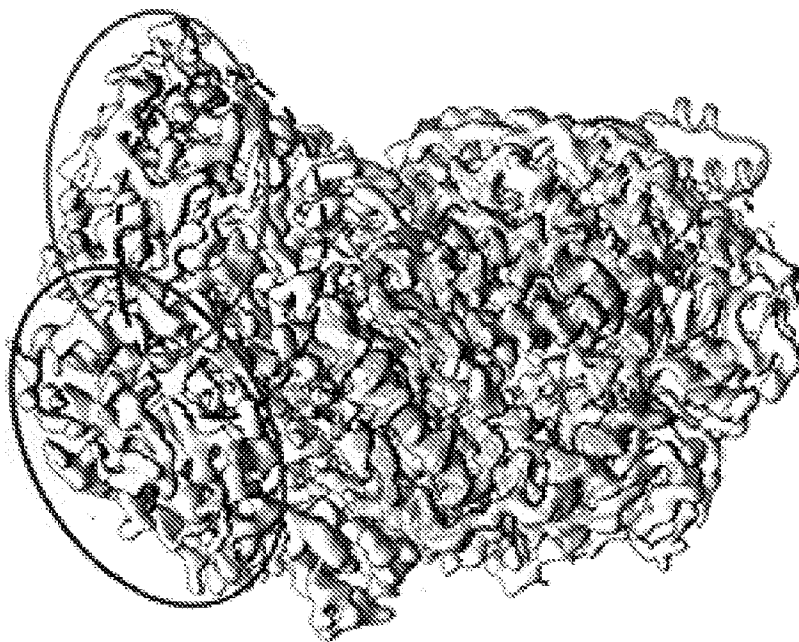


FIG. 28B

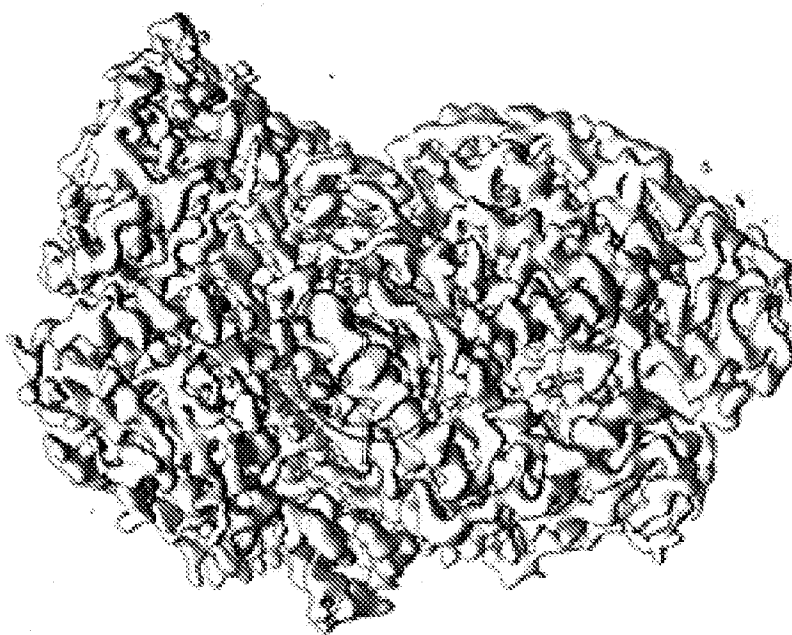


FIG. 28A

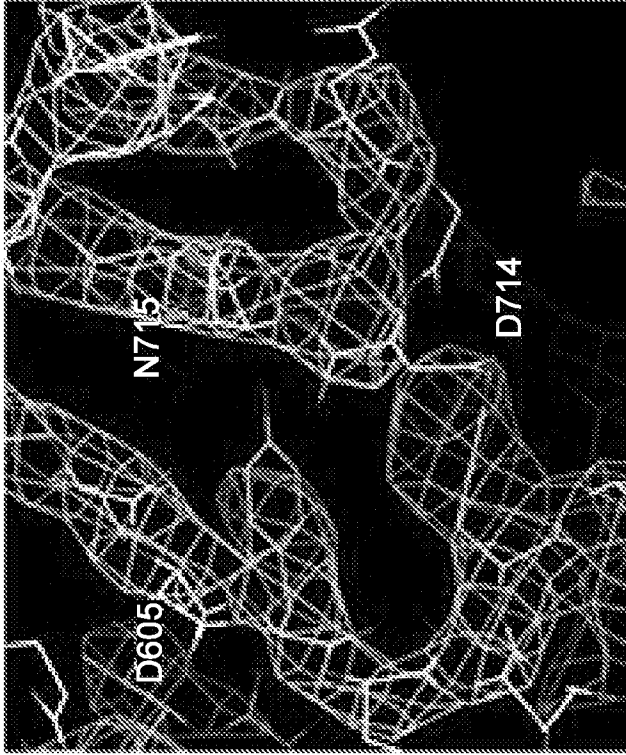


FIG. 28D

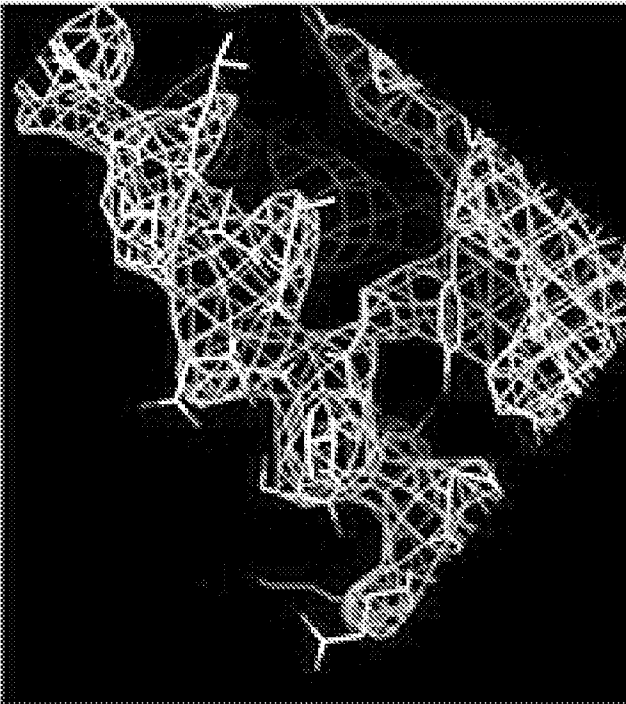


FIG. 28C

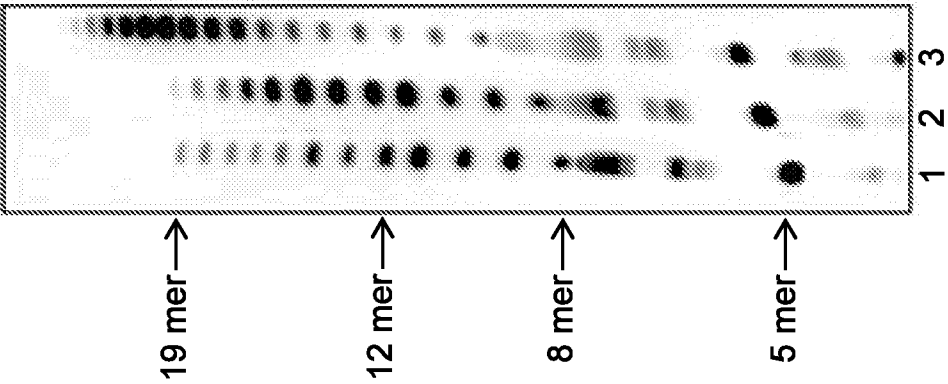


FIG. 29B

VSIV L + P1-106 + N22-423

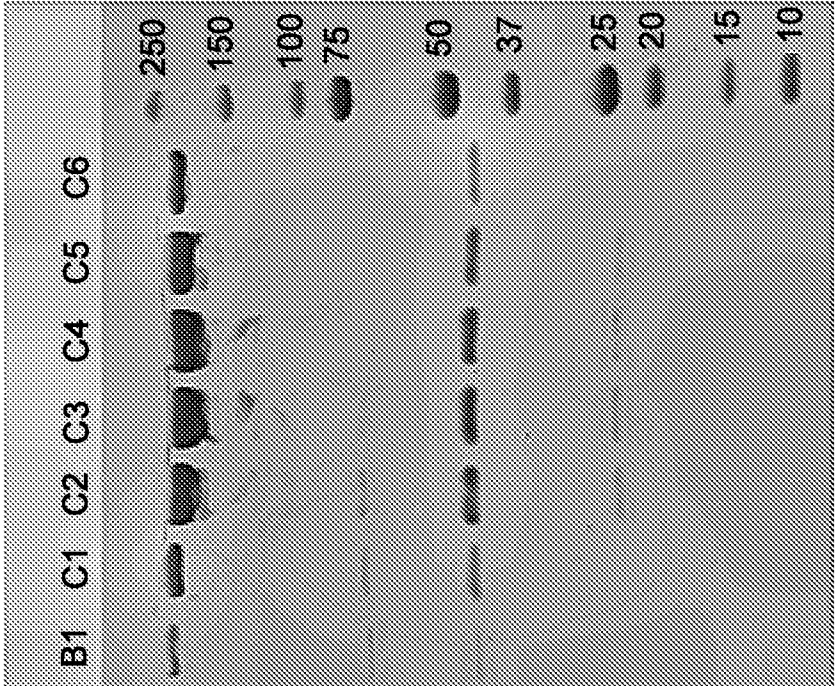


FIG. 29A

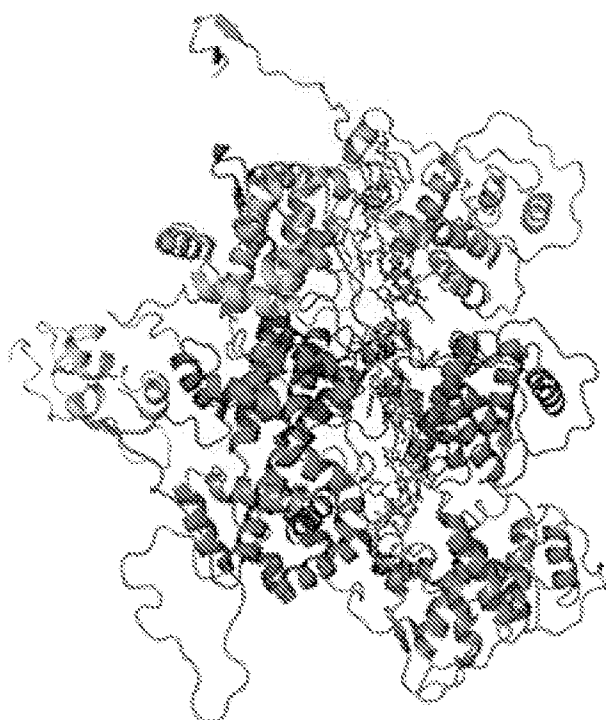


FIG. 30A

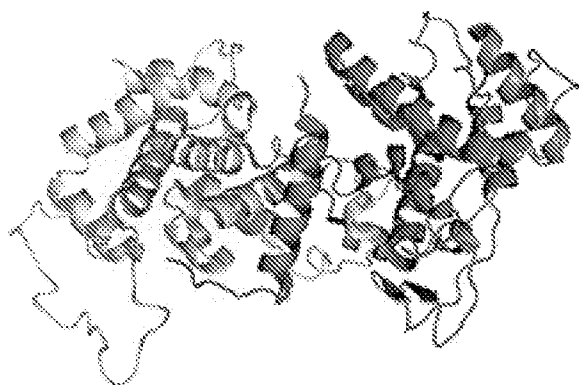


FIG. 30B

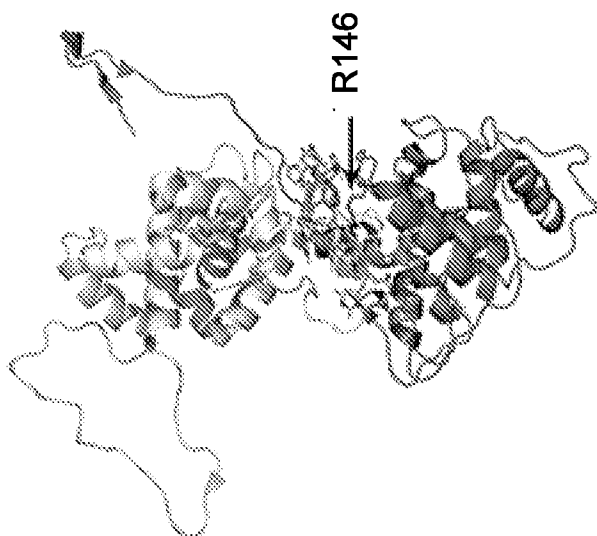
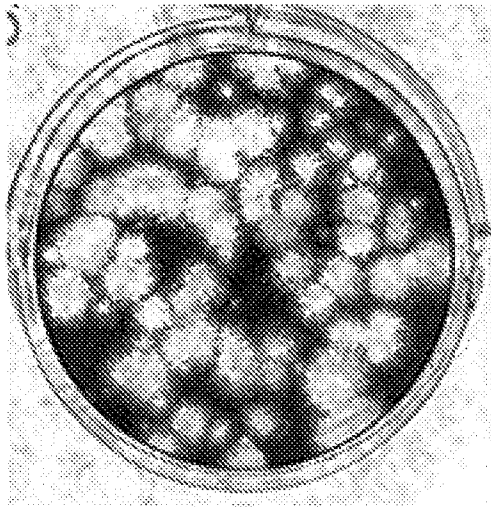
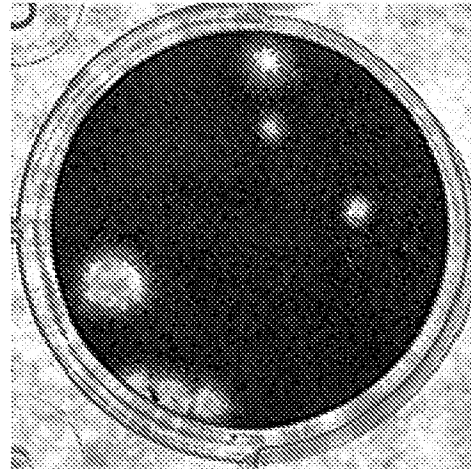


FIG. 30C



P_{wt}



P_{minOD}

FIG. 31B

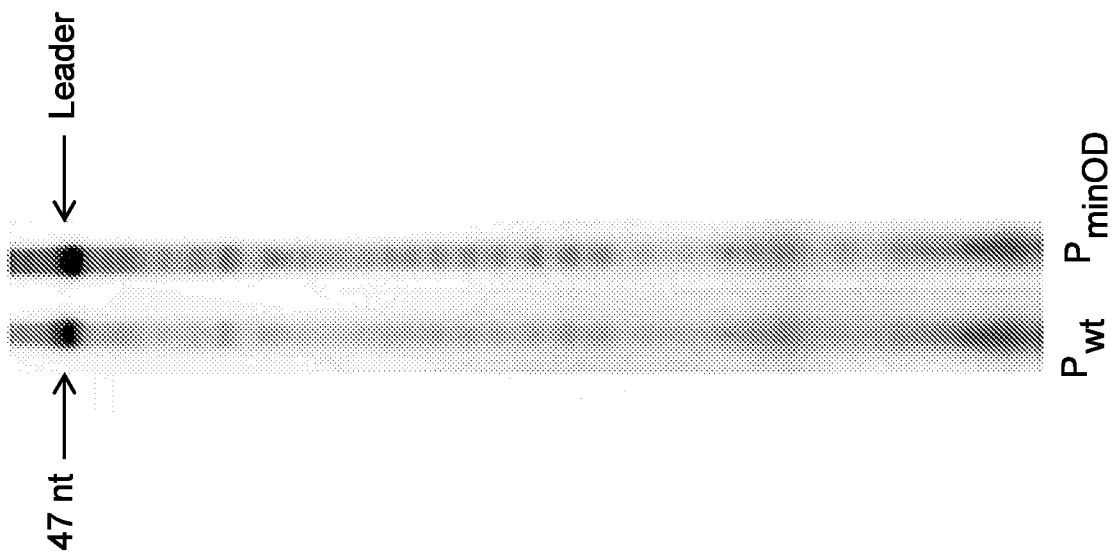


FIG. 31A

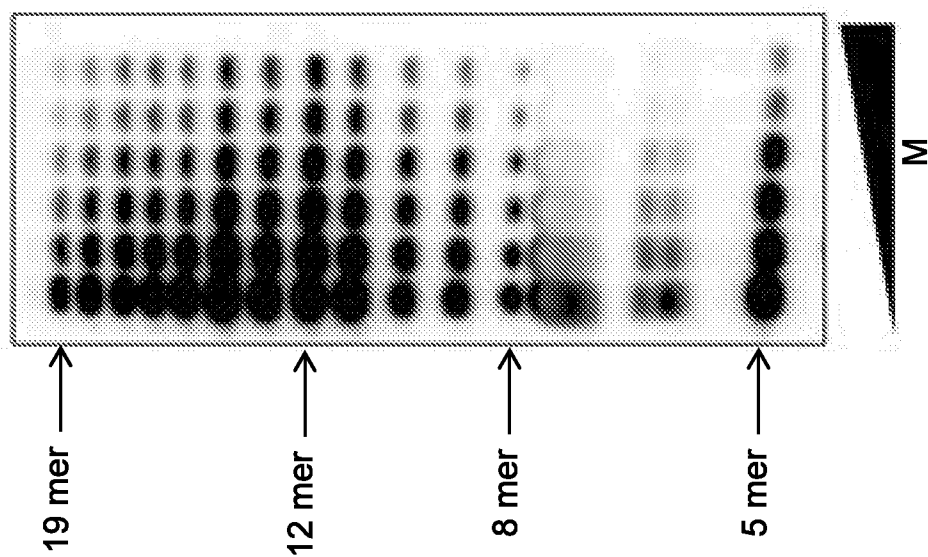


FIG. 32B

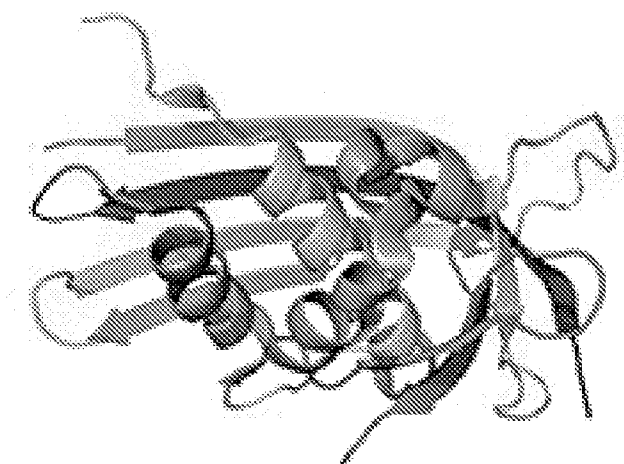
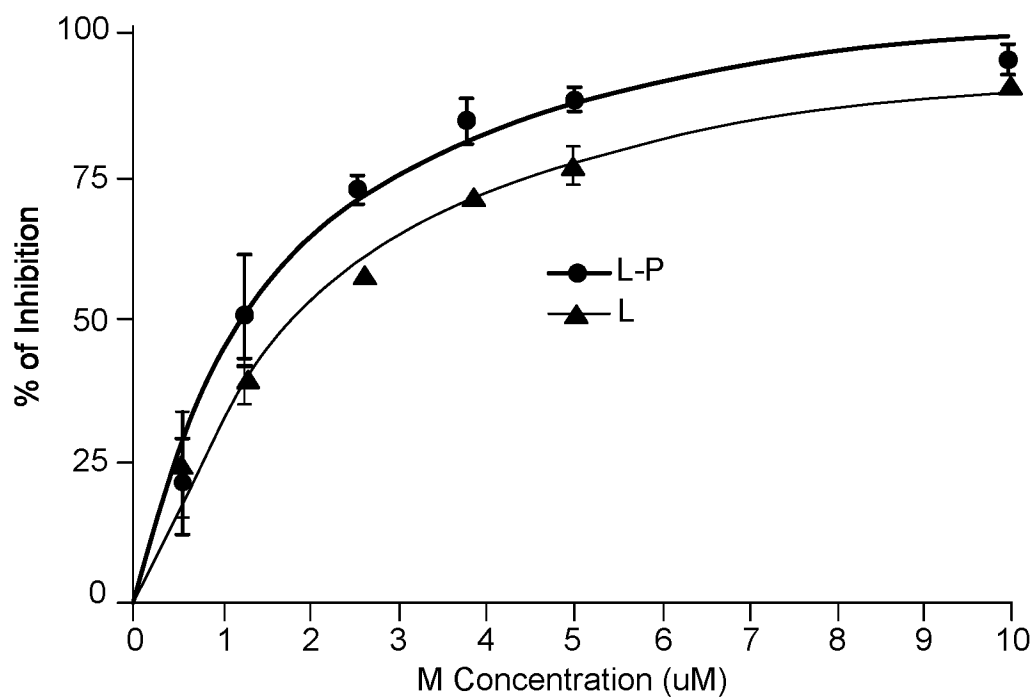
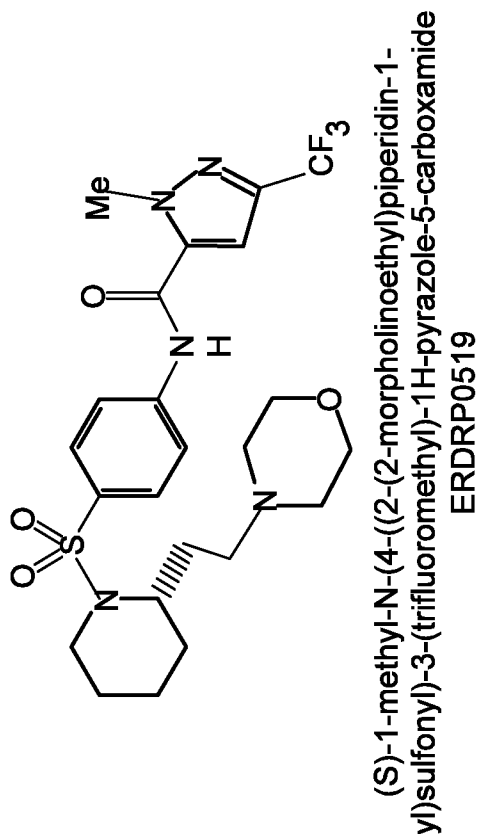
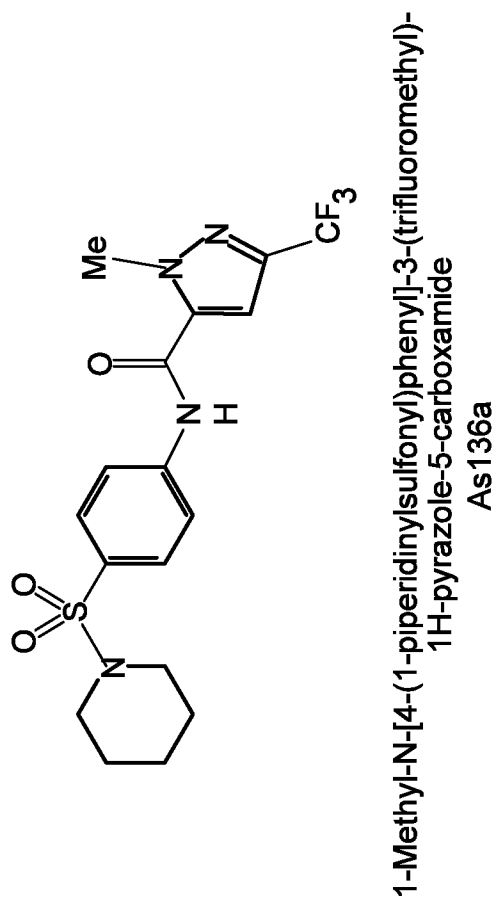


FIG. 32A

59/69

**FIG. 32C**

Non-nucleoside inhibitors of MV polymerase



60/69

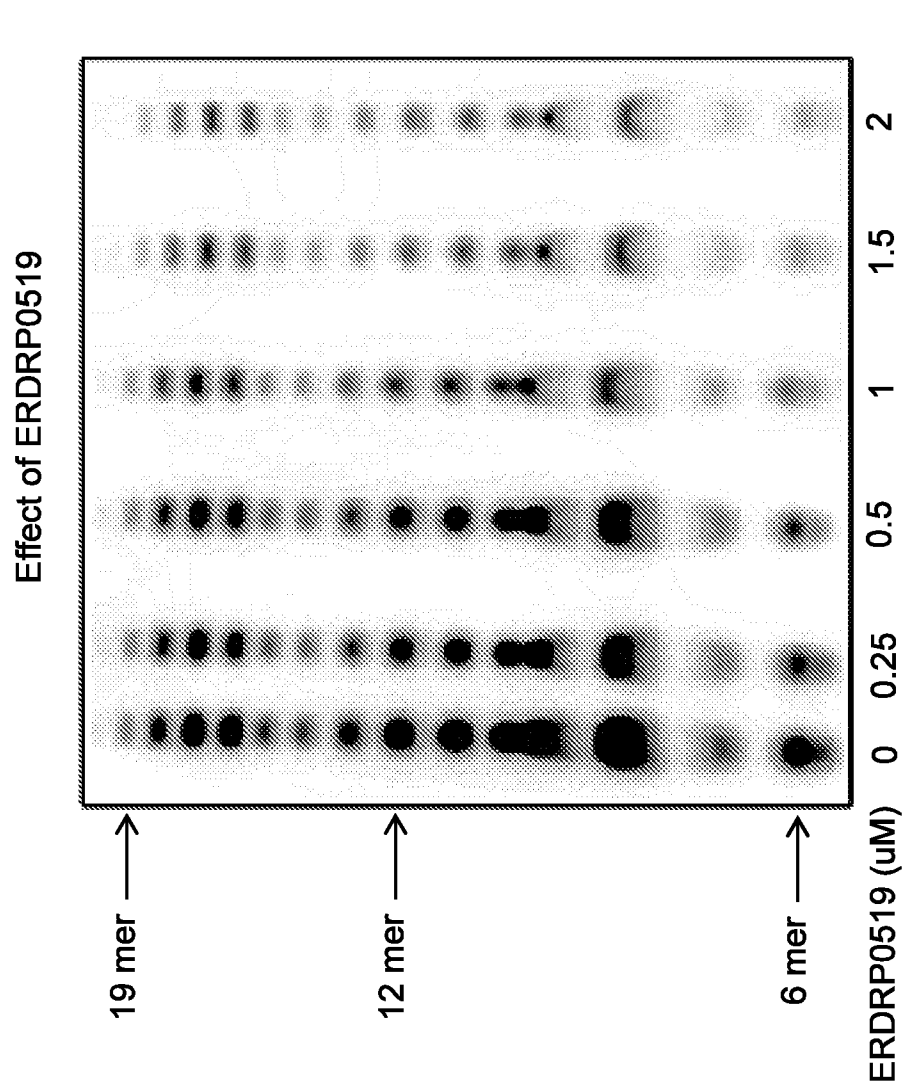
Published in final edited form as:
J Med Chem. 2012 May 10; 55(9): 4420-4230. doi:10.1021/jm201699w.

Non-nucleoside inhibitors of the Measles Virus RNA-dependent RNA Polymerase: Synthesis, Structure-Activity Relationships and Pharmacokinetics

J. Mmaina Ndungu¹, Stefanie A. Krumm², Dan Yan², Richard F. Arrendale¹, G.Prabhakar Reddy¹, Taylor Evers¹, Randy Howard¹, Michael G. Natchus¹, Manohar T. Saindane¹, Dennis C. Liotta^{1,3}, Richard K Plemper^{2,4}, James P. Snyder^{1,3}, and Aiming Sun^{1,1}

What is the mechanism of action?

FIG. 33



*ERDRP0519 inhibits MV L/P with an IC50 of 100nM

FIG. 34

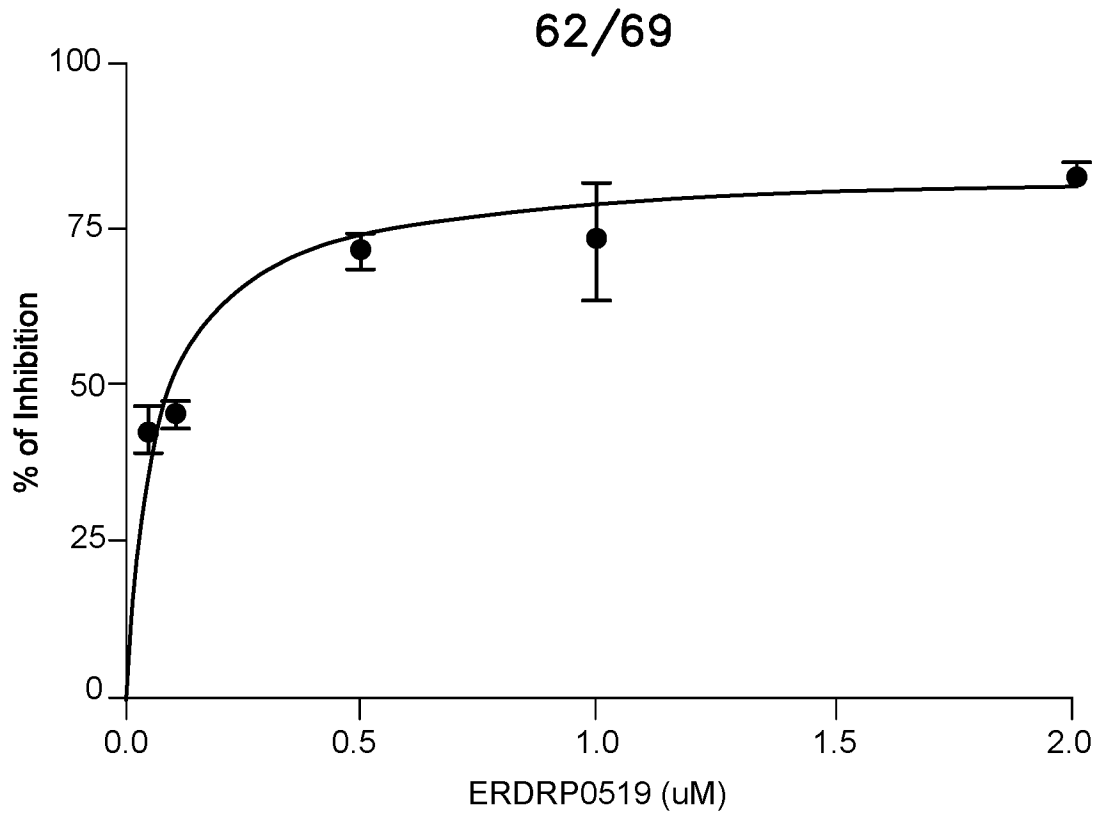


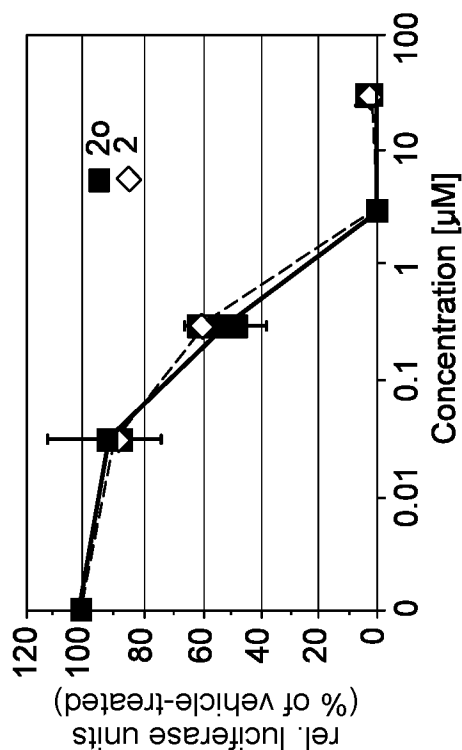
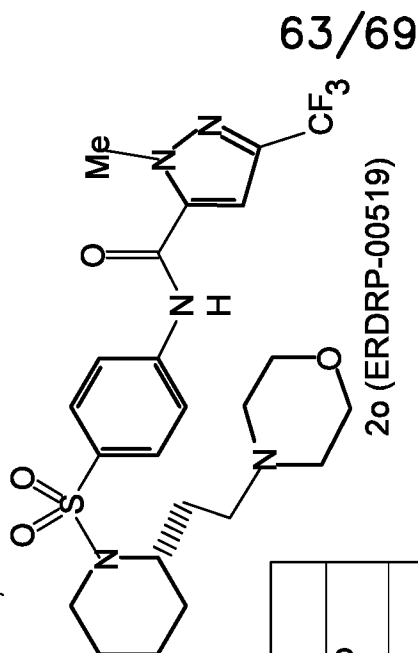
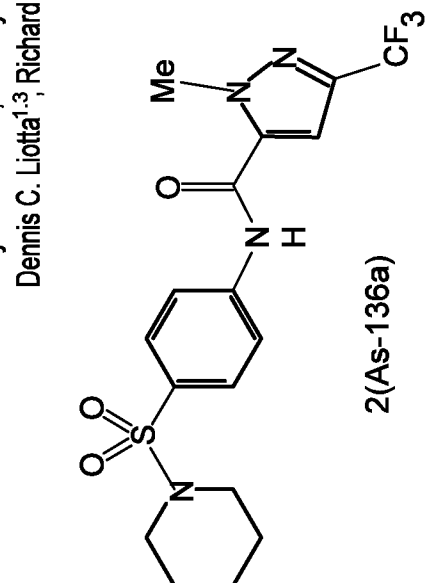
FIG. 34 Cont'd

Published in final edited form as:

J Med Chem. 2012 May 10; 55(9): 4420-4230. doi:10.1021/jm201699w.

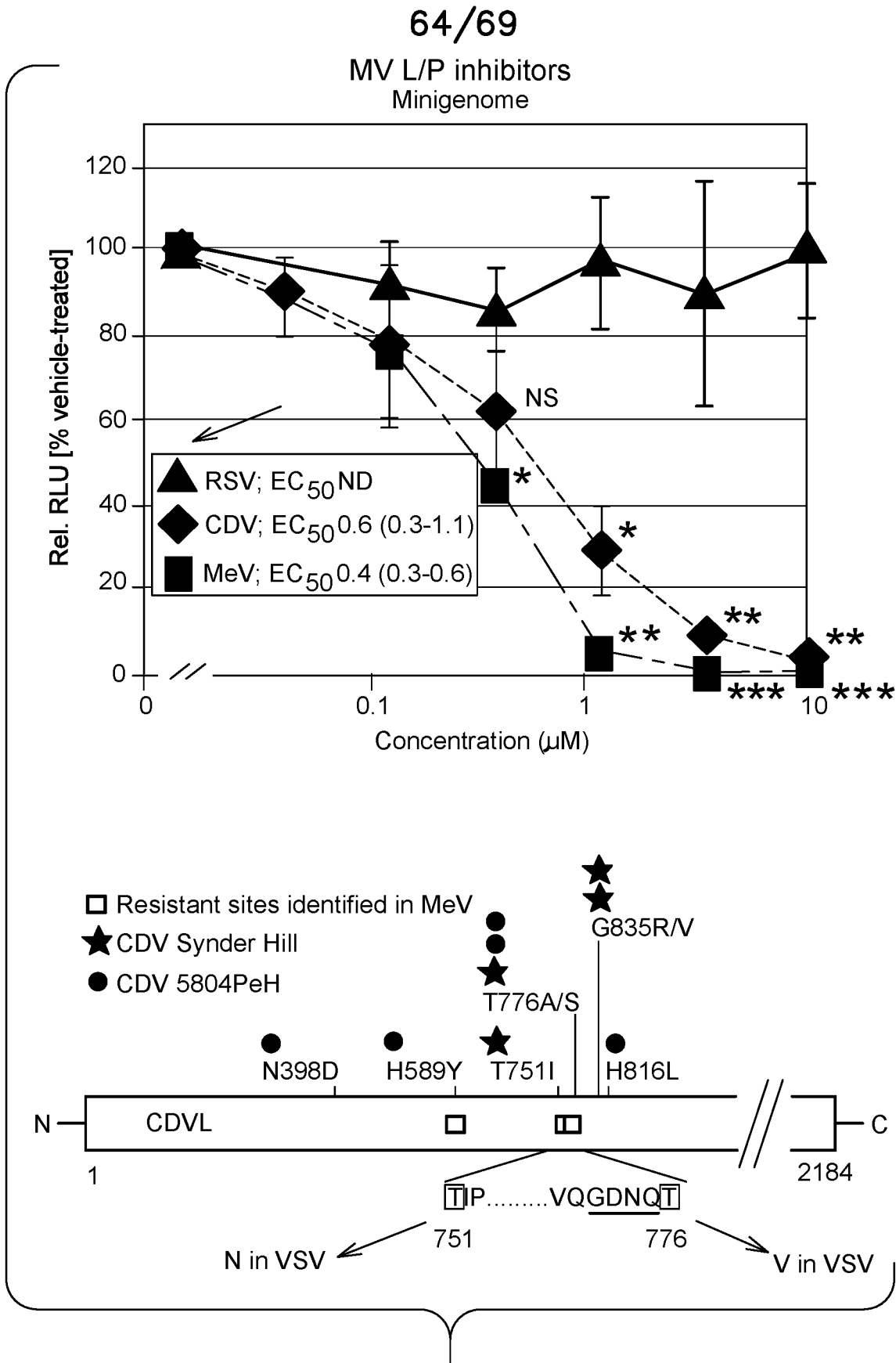
Non-nucleoside inhibitors of the Measles Virus RNA-dependent RNA Polymerase: Synthesis, Structure-Activity Relationships and Pharmacokinetics

J. J. Mmaina Ndungu¹, Stefanie A. Krumm², Dan Yan², Richard F. Arrendale¹, G. Prabhakar Reddy¹, Taylor Evers¹, Randy Howard¹, Michael G. Natchus¹, Manohar T. Saindane¹, Dennis C. Liotta^{1,3}, Richard K Plemper^{2,4}, James P. Snyder^{1,3}, and Aiming Sun¹

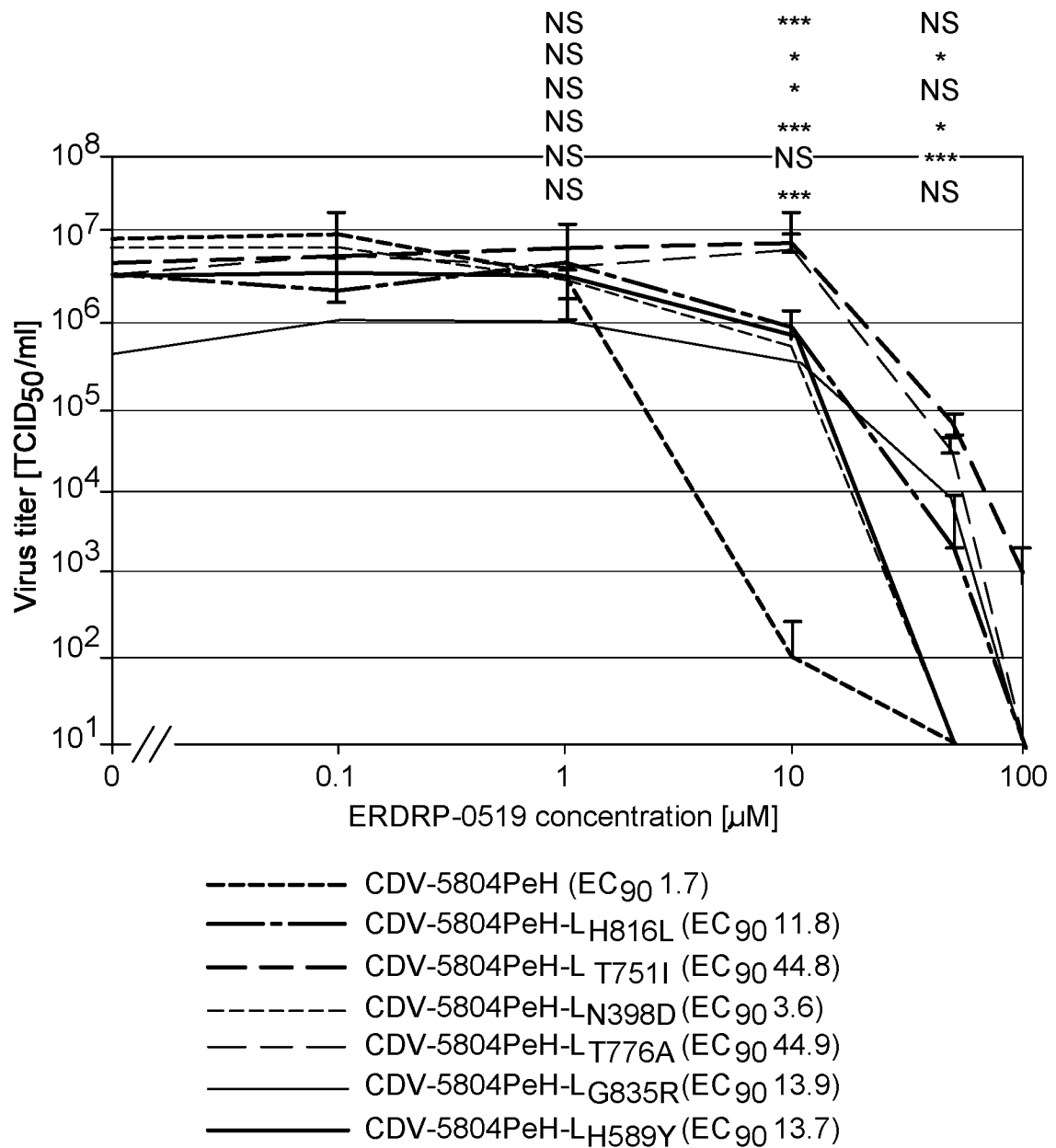


Compounds 2o and 2 inhibit viral RdRp activity with equal potency. Values are expressed relative to vehicle-treated samples and represent averages of three experiments $\pm$ SD.

FIG. 35

**FIG. 36**

65/69

**FIG. 36 Cont'd**

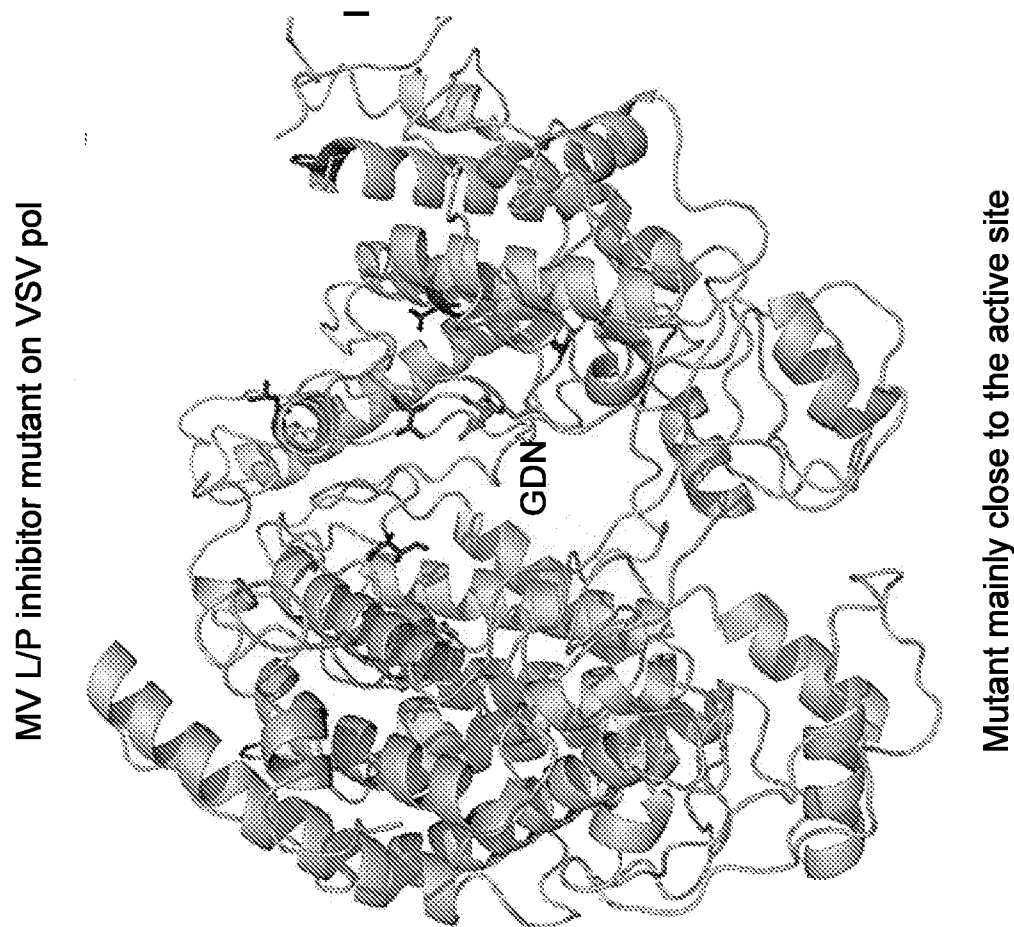
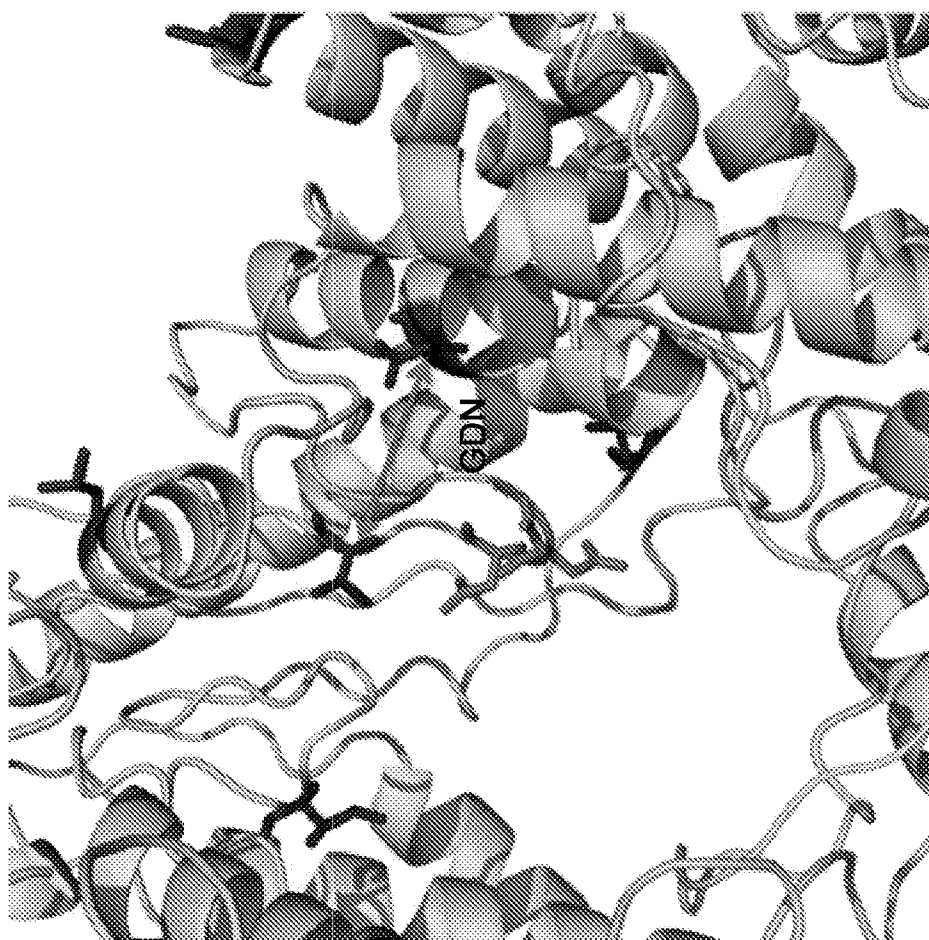


FIG. 37

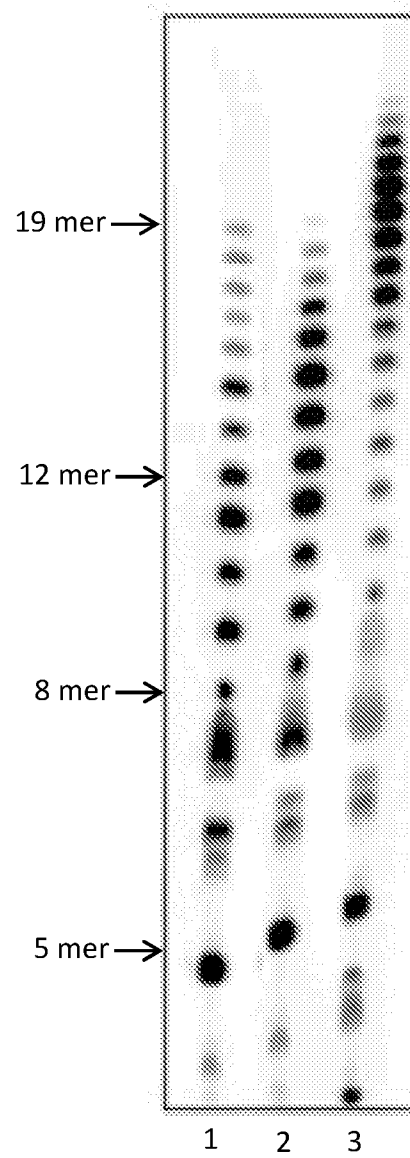
MV L/P inhibitor mutant on VSV pol

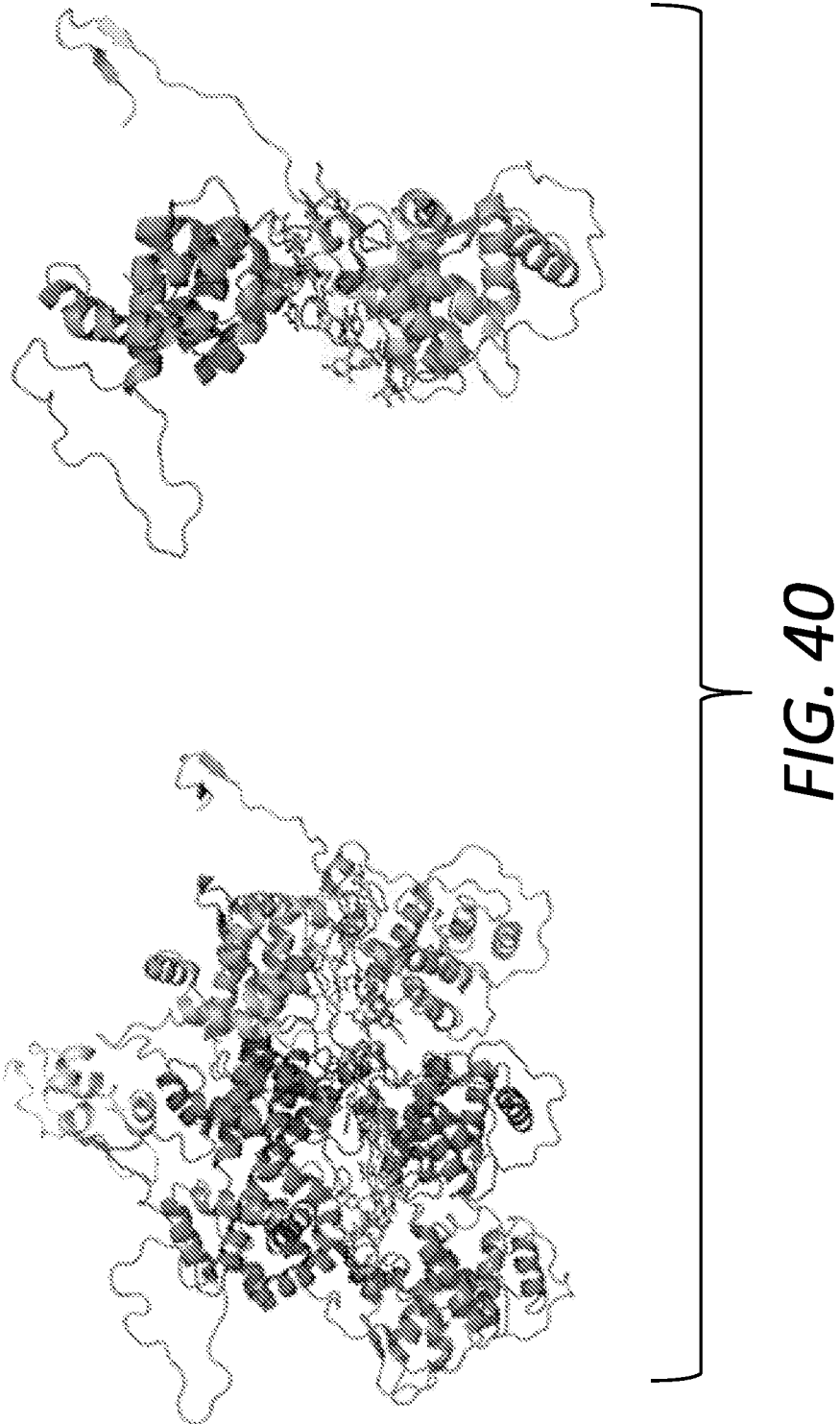


Interact with the 2 threonines
to block the flexibility of the
catalytic loop

FIG. 38

68/69

*FIG. 39*



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 15/65388

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C40B 30/02, 30/04; C12Q 1/70; C07K 14/08, 14/145; G01N 23/00, 1/42, 33/50 (2016.01)

CPC - G01N 2333/08, 2333/145, 23/311; H01J 37/26; G06F 19/16; C40B 30/02, 30/04; C07K 14/08, 14/145

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8): C40B 30/02, 30/04; C12Q 1/70; C07K 14/08, 14/145; G01N 23/00, 1/42, 33/50 (2016.01)

CPC: G01N 2333/08, 2333/145, 23/311; H01J 37/26; G06F 19/16; C40B 30/02, 30/04; C07K 14/08, 14/145

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

CPC: G01N 2333/08, 2333/145, 23/311; H01J 37/26; G06F 19/16; C40B 30/02, 30/04; C07K 14/08, 14/145 (text search)

USPC: USPC: 250/311; 702/19; 436/501; 506/8, 9; 435/5, 235.1 (text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Electronic data bases: PatBase; Google Patents; Google Scholar

Search terms: high resolution viral protein structure, virtual chemical library screening, inhibitor, L protein, RNA-dependent RNA polymerase (RdRp), Methyltransferase (MTase), Non-segmented negative strand (NNS) virus, Vesicular stomatitis virus, filovirus, rabdo

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VISWANATHAN et al. Identification of a novel inhibitor of dengue virus protease through use of a virtual screening drug discovery Web portal. J Chem Int Model 27 Oct 2014 Vol 54 No 10 Pages 2816-2825. Especially abstract.	1-10, 12-31
A	GALLOWAY et al. Analysis of a structural homology model of the 2'-O-ribose methyltransferase domain within the vesicular stomatitis virus L protein. Virology 5 December 2009 Vol 382 No 1 Pages 69-82. Especially abstract, pg 70 fig 1	1-10, 12-31
A	RAHMEH et al. Molecular architecture of the vesicular stomatitis virus RNA polymerase. Proc Nat Acad Sci 16 November 2010 Vol 107 No 46 Pages 20075-20080. Especially abstract; pg 20076 fig 1E, 2C.	1-10, 12-31
Y,P	LIANG et al. Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. CELL 16 July 2015 Vol 162 No 2 Pages 314-327. Especially entire article.	1-10, 12-31
A	GALLOWAY et al. S-adenosyl homocysteine-induced hyperpolyadenylation of vesicular stomatitis virus mRNA requires the methyltransferase activity of L protein. J Virol December 2008 Vol 82 No 24 pages 12280-12290. Especially abstract, pg 12281 fig 1, pg 12287 col 1 para 1.	1-10, 12-31
A	BUJNICKI et al. In silico identification, structure prediction and phylogenetic analysis of the 2'-O-ribose (cap 1) methyltransferase domain in the large structural protein of ssRNA negative-strand viruses. Protein Eng February 2002 Vol 15 No 2 Pages 101-108. Especially abstract.	1-10, 12-31

☒ Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

30 March 2016

Date of mailing of the international search report

21 APR 2016

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 15/65388

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 11
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 15/65388

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2011/263554 A1 (KAWAOKA et al.) 27 October 2011 (27.10.2011). Especially para [0005]	1-10, 12-31
A	US 2011/160188 A1 (PLEMPER et al.) 30 June 2011 (30.06.2011). Especially para [0011], [0104]	1-10, 12-31