PEARLS

RNA 3-dimensional structural motifs as a critical constraint of viroid RNA evolution

Ying Wang[1]*, Craig L. Zirbel[2]*, Neocles B. Leontis[3]*, Biao Ding[4,5]†

1 Department of Biological Sciences, Mississippi State University, Starkville, Mississippi, United States of America, **2** Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio, United States of America, **3** Department of Chemistry and Center for Biomolecular Sciences, Bowling Green State University, Bowling Green, Ohio, United States of America, **4** Department of Molecular Genetics, The Ohio State University, Columbus, Ohio, United States of America, **5** The Center for RNA Biology, The Ohio State University, Columbus, Ohio, United States of America

† Deceased.
* wang@biology.msstate.edu (YW); zirbel@bgsu.edu (CLZ); leontis@bgsu.edu (NBL)
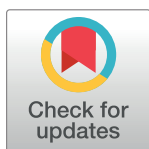
Introduction

Viroids are circular noncoding RNAs infecting plants [1, 2]. During infection, viroids, like RNA viruses, generate swarms of sequence variants called quasispecies [3, 4]. Viroids in Avsunviroidae family replicate in chloroplasts and display the highest mutation rates among all living entities [5]. Viroids in Pospiviroidae family replicate in the nucleus with a relatively lower mutation rate resembling some RNA viruses [6]. Those sequence variants generated during replication are described by the concept of sequence space, which harnesses a geometric representation to illustrate genetic similarities via physical distances. Given the high mutation rate and fast propagation, viroid RNAs have a potentially large sequence space for the evolution of new variants. However, in reality, they use only a small fraction of this space. Constraints of viral sequence space may include genome size, replication fidelity, error thresholds, host or tissue tropism, etc. These factors have been nicely reviewed elsewhere [3, 7, 8] and are not the focus of this Pearl. In addition, RNA secondary structures have been considered, though not adequately, as a constraint factor [8]. Viroids, in contrast to viruses, entirely rely on their RNA structural motifs for function due to their noncoding nature, which offers insights into their capacity to explore regions of sequence space influenced by RNA structures.

Here, we describe that 3-dimensional (3D) structural motifs formed by non–Watson-Crick (non-WC) base pairs in viroid RNAs act as a critical constraint for the sequence space of viroid genome evolution. This constraint operates because RNA 3D motifs can play crucial roles by mediating (1) RNA–RNA interactions for the folding of a part or a whole of RNA into a distinct tertiary conformation and (2) RNA–protein interactions. Therefore, mutations in a 3D motif that do not disrupt the structure and function will be retained in the population, whereas mutations that disrupt the 3D structures of motifs, and consequently the function, will be lost.

Question 1: What are the features of local RNA 3D structural motifs?

RNA 3D structures, to a first approximation, are composed of helices (formed by contiguous WC base pairs such as adenine [A]–uridine [U], guanine [G]–cytosine [C], and GU base pairs) and loops, both of which are shown in RNA 2D structures (Fig 1A). The loops are usually structured by additional interactions, including non-WC base pairs, base–backbone interactions, and base stacking (Fig 1A). In larger RNAs, these "local" loops can bind to helices or other loops distant in the 2D structure, stabilizing a larger-scale 3D structure. The loops have

Fig 1. RNA structure basis. (A) Simple illustration of RNA primary, 2D, and 3D structures. (B) Three edges of adenine nucleotides. (C) Isosteric AG tHS and CU tHS base pairs. Glycosidic bond orientations are highlighted with magenta arrows. C1'-C1' distances are highlighted with dashed magenta lines. (D) tHS IsoDiscrepancy Index heat map from the RNA Basepair Catalog. Any base combinations in the tHS family are listed, and AG vs CU is marked in magenta dot. Lower numeric value (less than 2.2 in blue color) dictates isosteric base pairs. Values between 2.2 and 3.5, colored in yellow, show nearly isosteric base pairs. Values above 3.5, colored in orange or red, dictate the nonisosteric base pairs. A, adenine nucleotides; C, cytocine nucleotides; G, guanine nucleotides; tHS, Trans Hoogsteen/Sugar edge; U, uridine nucleotides.

https://doi.org/10.1371/journal.ppat.1006801.g001

been described in detail by atomic-resolution crystallography and NMR spectroscopy studies [9]. Loop geometries and interaction details are typically conserved in homologous positions across species. Those RNA loop geometries that recur in nonhomologous positions of unrelated RNA molecules, with at most minor variations, are referred to as recurrent RNA 3D motifs [10, 11].

Question 2: What are non-WC base pairs?

Each RNA base has 3 edges, the WC, Hoogsteen, and Sugar edges, that can potentially hydrogen bond (H-bond) with other base edges in loop motifs (Fig 1B) [12]. According to the relative positions of glycosidic bonds, for each pair of interacting edges, there are 2 possible orientations, called "cis" (together) and "trans" (opposed). In total, there are 12 base-pairing geometries. Sequence variations observed for paired positions in RNA motifs are typically isosteric, meaning that base substitutions occupying similar space are potentially interchangeable without disrupting 3D structures [13]. To qualify, those base interactions should (1) use the same edges for interaction, (2) share the same orientations (cis or trans) of glycosidic bonds, and (3) occupy the same C1'-C1' distance in space. Base pair isostericity reduces the range of base substitution in 3D motifs (Fig 1C). Features of all possible RNA base pairings, including edge interactions, glycosidic bond orientations, and C1'-C1' distances, are displayed in the RNA Basepair Catalog (http://ndbserver.rutgers.edu/ndbmodule/services/BPCatalog/bpCatalog.html). The Catalog provides a numerical measure of the degree of isostericity among different base combinations for each of the 12 base-pairing geometries, displayed in interactive heat maps, illustrated by AG versus CU trans Hoogsteen/Sugar Edge base pairs in Fig 1D.

Question 3: Why are 3D structures of RNA loop motifs critical for function?

In a regular RNA helix, only the minor groove is easily accessible to proteins, while the major groove is too narrow for inserting alpha helixes, as occurs in DNA–protein complexes. The minor groove (sugar) edges of the nucleotides display a smaller difference between AU and GC base pairs than the major groove, but some amino groups, such as GN2 in guanine nucleotide, can sometimes constrain RNA sequence variations when H-bonding with proteins. More common functional sites are the loop regions of an RNA that provide specific binding locations for proteins or other molecules. Non-WC base pairs in RNA loops expose WC edges and widen the major groove. The WC edges are more distinct across the 4 bases, which allows for specific interactions critical for function.

Question 4: What is the evidence that RNA 3D motifs are critical for viroid infection?

RNA secondary structures of potato spindle tuber viroid (PSTVd), the type species of Pospiviroidae family, have been well characterized through chemical mapping (Fig 2A) [14, 15]. Noteworthy is that both studies, including the recently developed Selective 2'-hydroxyl acylation analyzed by primer extension experiments, support the existence of base pairs within loop motifs [14, 15], and 18 out of 27 RNA loops in the PSTVd genome are critical for either replication or systemic spreading [16], both of which are commonly used for assessing the fitness of viruses as surrogates [3].

Three-dimensional non-WC base pair arrangements in several PSTVd RNA motifs were annotated recently. Zhong et al. [17] analyzed the PSTVd loop E motif and validated the 3D
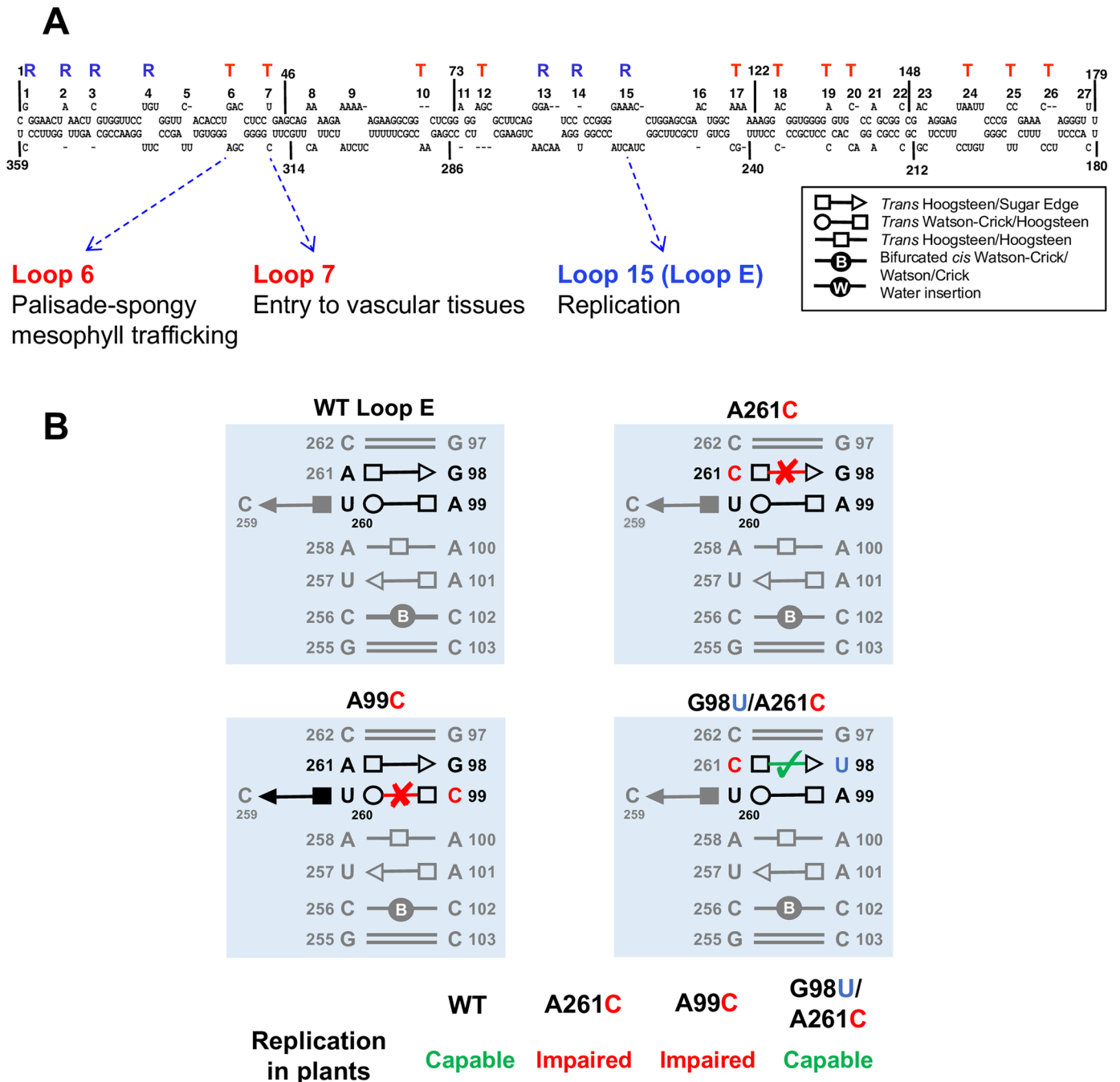
Fig 2. PSTVd RNA structures. (A) The 2D organization of PSTVd RNA genome. 3D structural arrangements and the function of loop 6, loop 7, and loop E are listed [17–19]. "T" and "R" depict the functions in "trafficking" and "replication," respectively [16]. (B) Disruptive and compensatory PSTVd loop E mutants predicted by isostericity [17]. Illustration for the replication of PSTVd variants in tomato plants, verified by northern blots [17], is shown in the lower panel. PSTVd, potato spindle tuber viroid; WT, wild-type.

https://doi.org/10.1371/journal.ppat.1006801.g002

structural arrangements therein. Interestingly, variants predicted to form nonisosteric base pairs (A99C and A261C) impair the replication capacity, while compensatory mutants (G98U/A261C) predicted to recover the original non-WC base pair restore the replication capacity (Fig 2B), demonstrating that isostericity dictates the function of RNA motifs.

Following this study, 3D structural arrangements of 2 additional PSTVd motifs have been shown to play critical roles [18, 19]. U43/C318 forms a single base pair motif (cis WC/WC) with a water insertion, termed loop 7, that regulates the entry of PSTVd to vascular tissues for spreading [19]. In addition, the neighboring loop 6 governs trafficking from palisade mesophyll to sponge mesophyll in plant leaves by forming specific non-WC base pairs [18]. Noteworthy is that saturated mutational analyses showed that the functional variants in each loop share isosteric structures.

Question 5: What is the evidence for RNA 3D structural motifs constraining viroid evolution?

Because some RNA 3D motifs control viral infection, strong selective pressures exist for maintaining the 3D motif structures that constrain the variation in sequence space. Mutational analyses on loop E, loop 6, and loop 7 all support this [17–19]. Taking loop 6 as an example, the 3D structure of this 3×3 loop was predicted using sequence-based homology search against RNA structure database [20], and the predicted model was consistently supported by data from functional mutagenesis analyses and chemical probing [19]. PSTVd loop 6 has a total of 4^6 possible sequence combinations, but there are only 8 functional variants out of 49 possible isosteric combinations [19]. Therefore, isostericity in RNA 3D motifs significantly reduced the sequence variations in PSTVd loop 6 by 84-fold (= $4^6/49$) and testing for function by an additional factor of 6 (= 49/8), indicating that RNA 3D structural motifs serve as a critical constraining factor.

Question 6: How do viroids adapt to new environments while under constraints to form RNA 3D motifs?

While maintaining the 3D structure of RNA loop motifs is pivotal, isosteric base substitutions may allow infection of new tissues or hosts. Previously, no infectious PSTVd strain for *Nicotiana tabacum* (tobacco) was observed in nature. However, in planta selection assays identified the C259U substitution in PSTVd loop E that led to the emergence of a new infectious strain for tobacco [21]. A subsequent study in transgenic tobacco also showed substitutions in loop E (C259U or U257A) enabling PSTVd infection of tobacco [22]. Both substitutions are predicted to be isosteric with the original wild-type (WT) sequences [17]. Therefore, isosteric base substitutions in loop E can both maintain the local 3D structure and allow for the emergence of new infectious PSTVd variants.

Conclusions and perspectives

Maintaining structures of RNA 3D motifs serves as a critical constraint of viroid evolution. In RNA 3D motifs, isosteric base substitutions in noncanonical base pairs are required to maintain 3D motif structure, greatly reducing the range of possible base substitutions. Maintaining functional interactions with proteins reliant on specific nucleotide–residue combinations further reduces the space of possible base changes.

RNA 3D motifs may be a constraint for viruses as well. Despite differences in their genetic makeups and unique infection and evolution pathways, different viral and viroid RNAs should all share one common property: RNA 3D motif–based RNA–RNA, RNA–protein, and RNA–small ligand interactions necessary for completing life cycles [23–27]. Therefore, understanding how RNA 3D structural motifs play a role in viral infection and their exploration for regions of sequence space may potentially improve the prediction of outbreaks of new viruses.

Acknowledgments

References

1. Ding B. The biology of viroid-host interactions. Annu Rev Phytopathol. 2009; 47:105–31. Epub 2009/04/30. https://doi.org/10.1146/annurev-phyto-080508-081927 PMID: 19400635.

2. Flores R, Gago-Zachert S, Serra P, Sanjuan R, Elena SF. Viroids: survivors from the RNA world? Annu Rev Microbiol. 2014; 68:395–414. Epub 2014/07/09. https://doi.org/10.1146/annurev-micro-091313-103416 PMID: 25002087.

3. Lauring AS, Andino R. Quasispecies theory and the behavior of RNA viruses. PLoS Pathog. 2010; 6(7): e1001005. Epub 2010/07/28. https://doi.org/10.1371/journal.ppat.1001005 PMID: 20661479.

4. Brass JR, Owens RA, Matousek J, Steger G. Viroid quasispecies revealed by deep sequencing. RNA Biol. 2017; 14(3):317–25. Epub 2016/12/28. https://doi.org/10.1080/15476286.2016.1272745 PMID: 28027000.

5. Gago S, Elena SF, Flores R, Sanjuan R. Extremely high mutation rate of a hammerhead viroid. Science. 2009; 323(5919):1308. Epub 2009/03/07. https://doi.org/10.1126/science.1169202 PMID: 19265013.

6. Lopez-Carrasco A, Ballesteros C, Sentandreu V, Delgado S, Gago-Zachert S, Flores R, et al. Different rates of spontaneous mutation of chloroplastic and nuclear viroids as determined by high-fidelity ultra-deep sequencing. PLoS Pathog. 2017; 13(9):e1006547. Epub 2017/09/15. https://doi.org/10.1371/journal.ppat.1006547 PMID: 28910391.

7. Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de la Iglesia F, Lafforgue G, et al. The evolutionary genetics of emerging plant RNA viruses. Mol Plant Microbe Interact. 2011; 24(3):287–93. https://doi.org/10.1094/MPMI-09-10-0214 PMID: 21294624.

8. Holmes EC. Error thresholds and the constraints to RNA virus evolution. Trends Microbiol. 2003; 11 (12):543–6. Epub 2003/12/09. PMID: 14659685.

9. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. RNA. 2001; 7 (4):499–512. PMID: 11345429.

10. Lescoute A, Leontis NB, Massire C, Westhof E. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. Nucleic Acids Res. 2005; 33(8):2395–409. https://doi.org/10.1093/nar/gki535 PMID: 15860776.

11. Petrov AI, Zirbel CL, Leontis NB. Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. RNA. 2013; 19(10):1327–40. Epub 2013/08/24. https://doi.org/10.1261/rna.039438.113 PMID: 23970545.

12. Leontis NB, Westhof E. The annotation of RNA motifs. Comp Funct Genomics. 2002; 3(6):518–24. Epub 2008/07/17. https://doi.org/10.1002/cfg.213 PMID: 18629252.

13. Stombaugh J, Zirbel CL, Westhof E, Leontis NB. Frequency and isostericity of RNA base pairs. Nucleic Acids Res. 2009; 37(7):2294–312. Epub 2009/02/26. https://doi.org/10.1093/nar/gkp011 PMID: 19240142.

14. Gast FU, Kempe D, Spieker RL, Sanger HL. Secondary structure probing of potato spindle tuber viroid (PSTVd) and sequence comparison with other small pathogenic RNA replicons provides evidence for central non-canonical base-pairs, large A-rich loops, and a terminal branch. J Mol Biol. 1996; 262 (5):652–70. Epub 1996/10/11. https://doi.org/10.1006/jmbi.1996.0543 PMID: 8876645.

15. Giguere T, Adkar-Purushothama CR, Perreault JP. Comprehensive secondary structure elucidation of four genera of the family Pospiviroidae. PLoS ONE. 2014; 9(6):e98655. Epub 2014/06/05. https://doi.org/10.1371/journal.pone.0098655 PMID: 24897295.

16. Zhong X, Archual AJ, Amin AA, Ding B. A genomic map of viroid RNA motifs critical for replication and systemic trafficking. Plant Cell. 2008; 20(1):35–47. Epub 2008/01/08. https://doi.org/10.1105/tpc.107.056606 PMID: 18178767.

17. Zhong X, Leontis N, Qian S, Itaya A, Qi Y, Boris-Lawrie K, et al. Tertiary structural and functional analyses of a viroid RNA motif by isostericity matrix and mutagenesis reveal its essential role in replication. J Virol. 2006; 80(17):8566–81. Epub 2006/08/17. https://doi.org/10.1128/JVI.00837-06 PMID: 16912306.

18. Zhong X, Tao X, Stombaugh J, Leontis N, Ding B. Tertiary structure and function of an RNA motif required for plant vascular entry to initiate systemic trafficking. Embo J. 2007; 26(16):3836–46. Epub 2007/07/31. https://doi.org/10.1038/sj.emboj.7601812 PMID: 17660743.

19. Takeda R, Petrov AI, Leontis NB, Ding B. A three-dimensional RNA motif in Potato spindle tuber viroid mediates trafficking from palisade mesophyll to spongy mesophyll in Nicotiana benthamiana. Plant Cell. 2011; 23(1):258–72. Epub 2011/01/25. https://doi.org/10.1105/tpc.110.081414 PMID: 21258006.

20. Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. J Math Biol. 2008; 56(1–2):215–52. Epub 2007/08/19. https://doi.org/10.1007/s00285-007-0110-x PMID: 17694311.

21. Wassenegger M, Spieker RL, Thalmeir S, Gast FU, Riedel L, Sanger HL. A single nucleotide substitution converts potato spindle tuber viroid (PSTVd) from a noninfectious to an infectious RNA for nicotiana tabacum. Virology. 1996; 226(2):191–7. Epub 1996/12/15. https://doi.org/10.1006/viro.1996.0646 PMID: 8955038.

22. Zhu Y, Qi Y, Xun Y, Owens R, Ding B. Movement of potato spindle tuber viroid reveals regulatory points of phloem-mediated RNA traffic. Plant Physiol. 2002; 130(1):138–46. https://doi.org/10.1104/pp.006403 PMID: 12226494.

23. Lee N, Moss WN, Yario TA, Steitz JA. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. Cell. 2015; 160(4):607–18. https://doi.org/10.1016/j.cell.2015.01.015 PMID: 25662012.

24. Fok V, Mitton-Fry RM, Grech A, Steitz JA. Multiple domains of EBER 1, an Epstein-Barr virus noncoding RNA, recruit human ribosomal protein L22. RNA. 2006; 12(5):872–82. Epub 2006/03/25. https://doi.org/10.1261/rna.2339606 PMID: 16556938.

25. Tycowski KT, Guo YE, Lee N, Moss WN, Vallery TK, Xie M, et al. Viral noncoding RNAs: more surprises. Genes Dev. 2015; 29(6):567–84. Epub 2015/03/21. https://doi.org/10.1101/gad.259077.115 PMID: 25792595.

26. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr., Swanstrom R, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009; 460(7256):711–6. Epub 2009/08/08. https://doi.org/10.1038/nature08237 PMID: 19661910.

27. Ooms M, Huthoff H, Russell R, Liang C, Berkhout B. A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. J Virol. 2004; 78(19):10814–9. Epub 2004/09/16. https://doi.org/10.1128/JVI.78.19.10814-10819.2004 PMID: 15367648.