Segment Self-Repulsion is the Major Driving Force of Influenza Genome Packaging

Sergey V. Venev and Konstantin B. Zeldovich

Phys. Rev. Lett. **110**, 098104 — Published 28 February 2013

DOI: 10.1103/PhysRevLett.110.098104

Segment Self-Repulsion is the Major Driving Force of Influenza Genome Packaging

Sergey V. Venev and Konstantin B. Zeldovich
Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA
(Dated: January 3, 2013)

The genome of influenza A virus consists of eight separate RNA segments, which are selectively packaged into virions prior to virus budding. The microscopic mechanism of highly selective packaging involves molecular interactions between packaging signals in the genome segments, and remains poorly understood. We propose that the condition of proper packaging can be formulated as a large gap between RNA-RNA interaction energies in the viable virion with 8 unique segments and in improperly packed assemblages lacking the complete genome. We then demonstrate that selective packaging of 8 unique segments into an infective influenza virion can be achieved by self-repulsion of identical segments at the virion assembly stage, rather than by previously hypothesized intricate molecular recognition of particular segments. Using Monte Carlo simulations to maximize the energy gap, without any other assumptions, we generated model 8-segment virions, which all display specific packaging, strong self-repulsion of the segments, and display reassortment patterns similar to natural influenza. The model provides a biophysical foundation of influenza genome packaging and reassortment, and serves as an important step towards robust sequence-driven prediction of reassortment patterns of the influenza virus.

PACS numbers: 87.19.xd, 87.16.A-, 87.14.gn

The influenza A virus is a pathogen of extreme significance for public health [1]. Among its unique characteristics is the segmented genome, which consists of 8 distinct RNA molecules of between 890 and 2341 nucleotides. All segments are essential for viral replication. The segmented genome allows influenza to mix and match segments when a host is co-infected by two or more strains at the same time, a process called reassortment. Although rare, reassortment events are believed to accelerate viral evolution, facilitate crossing host species barriers, create highly pathogenic strains [2] and are responsible for the major influenza epidemics [3]. A deeper understanding of the principles of influenza genome packaging and reassortment would greatly assist in influenza monitoring and potentially treatment.

The influenza virion lacks a rigid capsid, and consists of an 8-segment genome in the form of 8 RNA-protein complexes (vRNP), enveloped in a common lipid membrane. Electron microscopy and cryoelectron tomography studies demonstrate that vRNP are in direct contact in the virion, with 7 segments surrounding the central one in the 7+1 pattern [4, 5]. Packaging of the influenza genome is highly specific: while the probability to select 8 distinct segments at random is $8!/8^8 \approx 1/416$ [6], most of the observed virions contain exactly 8 unique segments forming the complete genome [7].

Packing of the viral genomes is an established problem of physical virology [8]. The genome is modeled by a polyelectrolyte molecule, constrained to the viral capsid. This makes it possible to calculate the nucleotide density profile within the capsid, and relate the genome size with capsid charge [9]. Unfortunately, these results are not directly applicable to the influenza virus, which lacks a rigid capsid, and assembles its genome from 8 independent segments. Here, instead of focusing on the structural properties of the virion, such as density profiles, we pose and address functional questions, which have so far eluded physical virology: what are the determinants of successful packaging of a segmented viral genome, and how viral fitness can be quantitatively related to the thermodynamic properties of the virions' constituents.

Both packaging and reassortment of the influenza virus rely on segment-segment interactions and molecular recognition during virion assembly. Since the 8 segments of the influenza genome encode 12 proteins, and each vRNP is made of an RNA molecule and exact same four proteins PA, PB-1, PB-2, NP [10], RNA-RNA and protein-RNA interactions must be involved in segment recognition. It has been demonstrated that influenza genome segments possess sequence features called packaging signals [11]. When the packaging signals are mutated, the virions fail to package properly; released virus progeny, if any, is not infective, although viral RNA is still being produced by the infected host cell. Bioinformatics and molecular biology studies have shown that the packaging signals are located towards the 3'- and 5'-ends of the segments, and partially overlap with protein-coding sequence [12, 13]. The current hypothesis of the highly selective genome packaging relies on highly evolved specific molecular recognition between vRNP, resulting in both packaging specificity and particular geometry of the influenza virion [14]. Below, we will present an alternative biophysical model of influenza packaging and reassortment.

We postulate that viral fitness depends on the packaging success of the virus: given similar viral RNA production rates, a strain that often produces virions with incomplete genomes will have a lower rate of produc-

tion of infective, complete-genome virions, and so will be at a fitness disadvantage compared to a strain that always packs 8 unique segments into every virion released. To make this concept quantitative, let us describe influenza virions by their packing lists, an 8-component vector of the copy number of each of the 8 segment types present in the virion. A viable virion has the packing list $V = (1,1,1,1,1,1,1,1)$, where each segment is present in the virion only once. The simplest improperly packaged, non-infective virion has the packing list of the $U = (0,2,1,1,1,1,1,1)$ type, where one of the segments is missing, and another is present in duplicate, Figure 1A. For influenza, there are 7*8=56 different packing lists of this type. The genome packaging problem can then be formulated by requiring the viable packing list to have the lowest possible energy of segment-segment interactions. In this case, V will be the most probable state compared to improperly packaged states $\{U\}$ during virion assembly. It is convenient to introduce the virion packaging efficiency Π as

$$\Pi = \frac{\sum_k \exp(-\beta E(V,k))}{\sum_k \exp(-\beta E(V,k)) + \sum_k \sum_{i=1}^{56} \exp(-\beta E(U_i,k))}, \quad (1)$$

where $E(X,k)$ is the energy of segment-segment interactions in the virion described by the packing list X, the state index k denotes the relative positional shifts of the segments present in X, and $\beta = 1/k_BT$. For simplicity, we consider only the 56 simplest mispackaged states in the partition function. The sum in the numerator reflects our assumption that viral fitness is determined by mere presence of the complete genome V, but not by its specific configuration. The larger the energy difference between the viable packaging and mispackaged states, the higher is the probability Π to assemble a virion containing the complete genome, and a strain with the higher Π will produce more virions containing the complete genome. Therefore, viral fitness is an increasing function of Π, and evolution of influenza towards proper packaging must increase Π, or the energy gap between V and the ensemble $\{U\}$ of mispackaged states. In this aspect the model is similar to the concept of the energy gap in protein folding, where the native folded conformation is separated from misfolded decoy states by a significant energy gap ΔG [15]. The folding free energy ΔG, equivalent to viral Π, has been shown to be a important physical constraint in protein evolution [16].

What is the physical mechanism of segment interactions, ultimately responsible for setting the energies $E(V), E(U_i)$? The current view is that selective packaging (thus, large gap) is achieved through intricate molecular recognition between the vRNP. Here, we propose an alternative mechanism of highly selective influenza genome packaging: instead of the specific attractive interactions between particular segments, successful packaging can be achieved due to repulsion between identical segments at the virion assembly stage, Figure 1B. In this way, a segment assembly with two or more copies of the same segment becomes thermodynamically unfavorable and therefore is unlikely to form a virion. Unlike the specific interactions model, the self-repulsion model does not require coevolution of multiple highly specific interfaces between separate RNA molecules. The specific

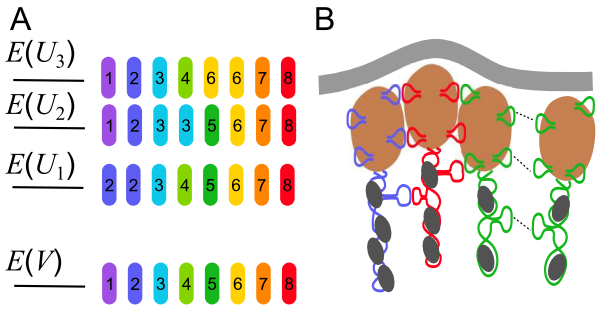

FIG. 1. (Color online) (A) The fitness/packaging hypothesis: the virion consisting of 8 unique segments has the lowest energy of segment-segment interactions. The gap between $E(V)$ and $E(U_i)$ defines the selective advantage of properly packed virions V with the complete set of genes. (B) Each vRNP consists of an RNA molecule condensed on NP proteins (gray) and loaded to the polymerase complex (brown). Prior to budding, the assemblage of vRNP in the nascent virion is in thermodynamic equilibrium with free vRNP in the cytoplasm. We hypothesize that RNA-RNA interactions between packaging signals (protruding loops on the vRNP) are responsible for specific packaging; protein-protein interactions provide nonspecific attraction between the vRNP.

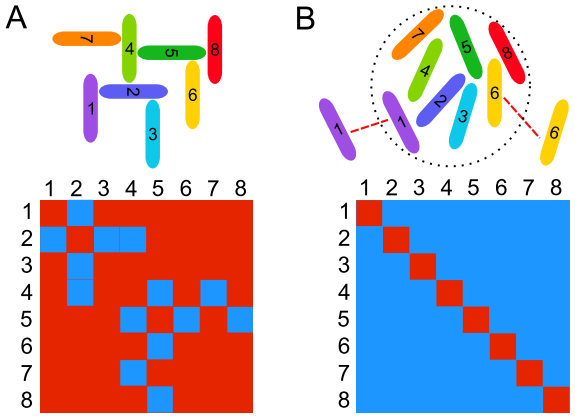

FIG. 2. (Color online) (A) Selective packaging of a segmented genome via specific segment-segment interactions. The segment interaction matrix E_{ij} is overall repulsive (red) with an intricate pattern of attraction (blue) between particular segments. (B) Alternatively, selective packaging can be achieved by self-repulsion of identical segments (dashed lines) combined with weak non-specific attraction between the different segments; the E_{ij} matrix exhibits a diagonal pattern. The diagrams represent the possible patterns of segment interactions rather than actual geometry of the virion.

interaction and self-repulsion models make qualitatively different predictions for the pattern of segment-segment interaction matrix. In the specific interaction model, one expects to observe distinct energy minima for the interactions of a segment with its cognate partner(s), Figure 2A. In the self-repulsion model, the interactions between most of the segments amount to weak, nearly nonspecific attraction, while the interaction of a segment with a copy of itself is strongly repulsive (Figure 2B). The physical interactions between the vRNP consist of RNA-RNA, protein-protein, and RNA-protein interactions. RNA-RNA interactions can be represented by Watson-Crick pairing between the packaging signals. Since the NP protein carries a large positive charge, it compensates most of the electric charge of the RNA. Therefore, a weak attraction between the vRNP due to van der Waals protein-protein interactions can be expected. The proposed model is compatible with the observed 7+1 nematic geometry of vRNP in the virion. It is believed [17] that virion assembly proceeds at the surface of the cell membrane (Figure 1B) and there is no conclusive data suggesting that positions of the specific segments within the 7+1 pattern are fixed. Therefore, lateral 2D diffusion of the segments already attached to the cell membrane can provide the necessary mobility and ensure thermodynamic sampling of all possible pairwise interactions between the segments in the pre-virions.

To model RNA-RNA interactions, we assume that each segment carries a packaging signal of length L_n, which is accessible for base-pairing with packaging signals on the other segments, Figure 1B. To account for the possible differences in length of the packaging signals, we chose 8 different values of L_n, $30 \le L_n \le 44$. The nucleotide interaction energies are $\epsilon_{AU} = -2, \epsilon_{GC} = -3$ (in arbitrary units) and zero otherwise, reflecting Watson-Crick base pairing. While, in reality, packaging signals are likely not continuous and may consist of disjoint RNA loops protruding from the vRNP, detailed consideration of these effects is beyond the scope of our model, and does not affect its general conclusions. The interaction energy of the two viral segments i and j is then approximated by hybridizing two linear RNA sequences,

$$E_{ij}(\{p,q\}) = \sum_{m=1}^{\min(L_i,L_j)} \epsilon(n^i_{m+p}, n^j_{m+q}), \qquad (2)$$

where n^i_m is the nucleotide type at the m-th position of segment i, and p and q are the relative shifts of the packaging segments due to possible geometric misalignment in the virion. We assumed $-3 \le p, q \le 3$, i.e. up to 6 nucleotide relative shift between the interacting sequences. The set $\{p,q\}$ is symbolically denoted as k in the partition sum, Eq. 1. The energy $E(X,k)$ of the segment assemblage (pre-virion) with the packing list $X = \{x_i\}$ and shifts k is then the sum of all pairwise interactions

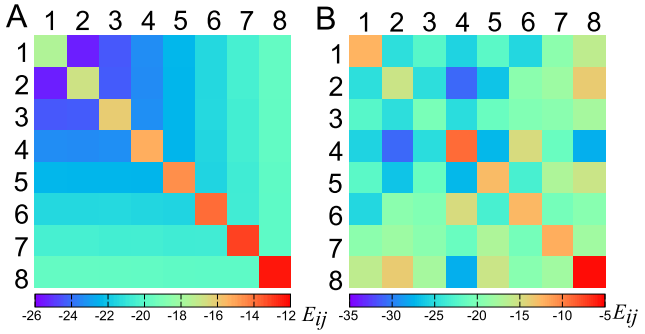

FIG. 3. (Color online) Segment interaction energies E_{ij} for the average over 10^4 virions, average $\Pi = 0.348$ (A), and a representative evolved virion with $\Pi = 0.294$ (B). Simulations to maximize Π, without any additional assumptions, lead to the emergence of segment self-repulsion. The color key below the plots encodes for interaction energy, from attractive (purple) to repulsive (red). Overall change of the energies with segment number is a result of differences in the length of packaging signals.

between the segments,

$$E(X,k) = \sum_{i=1}^{8} \frac{x_i(x_i-1)}{2} E_{ii}(k) + \sum_{i=1,j>i}^{8} x_i x_j E_{ij}(k). \quad (3)$$

To simulate evolution of proper packaging, we used a Monte Carlo approach by starting each virion with 8 random packaging signals, introducing random mutations, and accepting or rejecting the mutations according to the Metropolis criterion so as to increase Π at the temperature $T = 8$, and 10^4 virions were evolved. As expected, initial random sequences produced virions with a very low packaging efficiency, $\Pi = 1.3 \cdot 10^{-4} \pm 0.0011$ (mean $\pm$ SD). After 5000 rounds of mutation and selection, the average Π increased 2600-fold to 0.348 ± 0.040.

The pattern of segment-segment interactions evolved in this model clearly supports the proposed hypothesis of segment self-repulsion, and rules out the model of specific segment "docking" interactions. Figure 3A shows the 8x8 matrix of segment interactions E_{ij}, averaged over 10^4 independently evolved virions. Segment self-repulsion is clearly manifested, as the diagonal elements E_{ii} are significantly higher than off-diagonal ones ($P < 10^{-2}$). A typical E_{ij} matrix of an individual virion is shown in Figure 3B and does not match the specific interaction model. Variations of the microscopic model of segment interactions, e.g. the ϵ matrix did not affect these qualitative conclusions (data not shown). Evolved RNA sequences did not show extreme GC content, low complexity, or other obvious artifacts.

The predictions of our simplistic model are in good agreement with experiment. Recently, Chou *et al* [7] used single molecule fluorescence to quantify co-packaging of the segments in immobilized virions. They determined that the probabilities to find a PB2 segment together

with each of the other 7 segments exceeded 88%, and estimated that over 50% of the virions that contain PB2 segment also contain all other 7 segments. These figures are in line with our thermodynamic estimate of packaging efficiency, $\Pi \approx 0.35$. Chou *et al* have also found that over 90% of the observed virions contained only one copy of each segment, which justifies our definition of V and the choice of the 56 double-occupancy states $\{U\}$ as mispackaged decoys. Similarly, Noda *et al* [17] observed that all 30 virions surveyed contained 8 different vRNP. Prediction of the most probable number of segments in the virion is beyond the scope our model, as it requires precise knowledge of the binding constants between the vRNP and cell surface, as well as the absolute energies of vRNP interactions and proper accounting for all entropic effects. The mean field treatment of segment interactions in Eq. 3 leads to the prediction that relative positions of the segments within the virion are random, apart from the trivially stronger attraction between longer segments. This prediction does not contradict the current experimental data.

One of the challenges to proper virion assembly is to prevent accidental packaging of the host RNA; influenza evolved multiple RNA recognition strategies, which include the use of negative-strand genome and specific docking of the polymerase complex at the conserved ends of the genome segments [10, 18]. We determined packaging efficiency Π for 10^4 evolved model virions with one of the packaging signals replaced by a random sequence of equal length, mimicking a host RNA molecule which did not coevolve with the virus. On average, inclusion of a single random segment into a virion decreased its packaging efficiency 14-fold to $\Pi = 0.024 \pm 0.044$. Therefore, we find that model viruses resist incorporation of host RNA via the segment self-repulsion, and highly specific segment docking is not necessary to prevent incorporation of exogenous RNA into the virions.

A simple generalization of our model permits analysis of influenza reassortment. In a reassortant virion, several segments out of 8 arise from one of the parent strains, while the second parent provides the remaining segments. To describe the possible reassortant virions, let us represent them by binary numbers between 0 (all 8 segments are from strain 1) and $2^8-1 = 255$ (all segments are from strain 2); the reassortment index r is $r = \sum_{i=1}^{8} 2^{i-1} a_i$, where $a_i = 0$ if segment i comes from strain 1, and $a_i = 1$ otherwise. We can then apply the concept of packaging efficiency Π to characterize the reassortment: if a virion with a certain mix of unique segments (from either parent strain) has a high probability of packaging, comparable to that of the parent strains, such a reassortant is a candidate for high fitness. Although influenza reassortment is a complex phenomenon, and functional incompatibilities between the segments can render a well-packaged virion non-infective [19], experiments show that proper packaging is a necessary condition for the emergence of

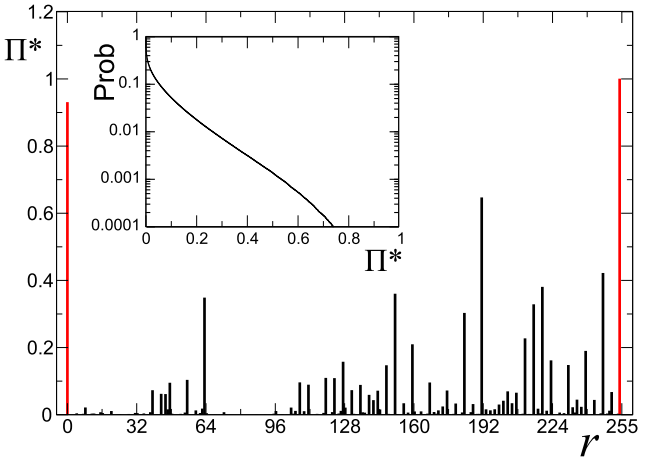

FIG. 4. (Color online) Normalized packaging efficiency Π^* for two parental strains (red) and the 254 non-trivial reassortants (black) as function of the reassortment index r. Most of the reassortants have $\Pi^* \ll 1$, well below the one of either parental strain. Inset: Probability to find a reassortant virion with a given Π^* or above. Virions with $\Pi^* \approx 1$ have the same packaging efficiency as the parent strains, and are exceedingly rare.

reassortant strains [20].

The packaging efficiency Π of a reassortant strain can be expressed by a straightforward generalization of Eqs. (1,3), where each segment can be taken from either of the parent strains according to the reassortment index r, and there are always 8 segments in a virion.

To simulate reassortment, we chose 1000 model virions evolved using the procedure described above, and calculated Π_r for all $254 = 2^8 - 2$ reassortants for every pair of parental virions. Figure 4 shows a typical bar chart of normalized packaging efficiency $\Pi^* = \Pi_r / \max(\Pi_0, \Pi_{255})$ for the 254 reassortants emerging from two parental virions. Remarkably, most reassortants fail to package the complete genome ($\Pi^* \ll 1$) and only a select few have Π approaching the packaging efficiency of the pure strains. Statistical analysis of $1.25 \cdot 10^5$ pairs of virions showed that on average only 0.12% of nontrivial reassortants had $\Pi \geq 0.5 \max(\Pi_0, \Pi_{255})$, i.e. at least one-half of the packaging efficiency of the better-packaging pure strain (Figure 4, inset). These findings support the notion of reassortment as a rare event, and suggest that if reassortment happens, it proceeds in a highly specific manner, depending on the packaging signals of the co-infecting strains.

To summarize, we considered influenza packaging and reassortment as thermodynamic competition between properly and improperly packed assemblages of vRNP prior to virion budding. Viral fitness is then proportional to the probability Π to find a virion containing all 8 unique, essential genome segments. Optimizing RNA-RNA interactions between packaging signals to maximize Π, we have shown that thermodynamic evolution of specific packaging results in self-repulsion of identical

segments, rather than in intricate segment recognition. Evolved model viruses have a high packaging efficiency Π compared to random packaging signals, do not package exogenous RNA, and are capable of reassortment remarkably similar to natural patterns. Obviously, packaging is just one of the many constraints required for an infective influenza virion, together with stable and functional proteins and functional regulatory RNA sequences. Since the packaging signals in real influenza partially overlap with the coding sequence through the use of synonymous codons [12], it is clear that evolution of packaging is compatible with other functional constraints on viral RNA.

The self-repulsion mechanism we observed is clearly related to the statistical enhancement of self-interactions of random patterns [21], which follows from the symmetry considerations; interactions between identical RNA molecules are statistically different from interactions between dissimilar sequences. The self-repulsion mechanism is not specific to influenza, and should have been in place at the earliest stages of evolution, enabling efficient assembly of multiple RNA molecules into functional multi-segmented genomes. Although our model is simplistic, and would greatly benefit from high-resolution structures of vRNP, we believe it is useful as a baseline framework for biophysical analysis of packaging and reassortment of segmented viral genomes. Further extension of the physics-driven approach to influenza genome assembly will improve identification of packaging signals in real influenza sequences.

The authors wish to acknowledge the support of DARPA (Prophecy Program, Contract No. HR0011-11-C-0095) and the contributions of all the members of the ALiVE (Algorithms to Limit Viral Epidemics) working group.

[1] R. J. Webby and R. G. Webster, Science **302**, 1519 (2003).
[2] C. Li, M. Hatta, C. A. Nidom, Y. Muramoto, S. Watanabe, G. Neumann, and Y. Kawaoka, Proceedings of the National Academy of Sciences (2010), 10.1073/pnas.0912807107.
[3] E. D. Kilbourne, Emerging infectious diseases **12**, 9 (2006).
[4] T. Noda, H. Sagara, A. Yen, A. Takada, H. Kida, R. H. Cheng, and Y. Kawaoka, Nature **439**, 490 (2006).
[5] E. Fournier, V. Moules, B. Essere, J.-C. C. Paillart, J.-D. D. Sirbat, C. Isel, A. Cavalier, J.-P. P. Rolland, D. Thomas, B. Lina, and R. Marquet, Nucleic acids research **40**, 2197 (2012).
[6] M. Enami, G. Sharma, C. Benham, and P. Palese, Virology **185**, 291 (1991).
[7] Y.-y. Chou, R. Vafabakhsh, S. Doğanay, Q. Gao, T. Ha, and P. Palese, Proceedings of the National Academy of Sciences **109**, 9101 (2012).
[8] A. Siber, A. L. Božič, and R. Podgornik, Physical chemistry chemical physics : PCCP **14**, 3746 (2012).
[9] V. A. Belyi and M. Muthukumar, Proceedings of the National Academy of Sciences **103**, 17174 (2006), http://www.pnas.org/content/103/46/17174.full.pdf+html.
[10] E. C. Hutchinson, J. C. von Kirchbach, J. R. Gog, and P. Digard, The Journal of general virology **91**, 313 (2010).
[11] Y. Fujii, H. Goto, T. Watanabe, T. Yoshida, and Y. Kawaoka, Proceedings of the National Academy of Sciences **100**, 2002 (2003).
[12] J. R. Gog, E. D. S. Afonso, R. M. Dalton, I. Leclercq, L. Tiley, D. Elton, J. C. von Kirchbach, N. Naffakh, N. Escriou, and P. Digard, Nucleic Acids Research **35**, 1897 (2007).
[13] K. Fujii, Y. Fujii, T. Noda, Y. Muramoto, T. Watanabe, A. Takada, H. Goto, T. Horimoto, and Y. Kawaoka, Journal of virology **79**, 3766 (2005).
[14] T. Noda and Y. Kawaoka, Proceedings of the National Academy of Sciences **109**, 8797 (2012).
[15] E. Shakhnovich, *Chemical Reviews*, Chem. Rev. **106**, 1559 (2006).
[16] K. B. Zeldovich, P. Chen, and E. I. Shakhnovich, Proceedings of the National Academy of Sciences **104**, 16152 (2007).
[17] T. Noda, Y. Sugita, K. Aoyama, A. Hirase, E. Kawakami, A. Miyazawa, H. Sagara, and Y. Kawaoka, Nature Communications **3**, 639+ (2012).
[18] P. Resa-Infante, N. Jorba, R. Coloma, and J. Ortin, RNA biology **8**, 207 (2011).
[19] C. Li, M. Hatta, S. Watanabe, G. Neumann, and Y. Kawaoka, Journal of Virology **82**, 11880 (December 1, 2008), http://jvi.asm.org/content/82/23/11880.full.pdf+html.
[20] Q. Gao and P. Palese, Proceedings of the National Academy of Sciences **106**, 15891 (2009).
[21] D. B. Lukatsky, K. B. Zeldovich, and E. I. Shakhnovich, Physical Review Letters **97**, 178101+ (2006).