# Analysis of the Contrasting Pathogenicities Induced by the D222G Mutation in 1918 and 2009 Pandemic Influenza A Viruses

Cheng Shang,* Chris S. Whittleston,* Kyle H. Sutherland-Cash,* and David J. Wales*

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, U.K.

S Supporting Information

ABSTRACT: In 2009, the D222G mutation in the hemagglutinin (HA) glycoprotein of pandemic H1N1 influenza A virus was found to correlate with fatal and severe human infections. Previous static structural analysis suggested that, unlike the H1N1 viruses prevalent in 1918, the mutation did not compromise binding to human $\alpha2,6$-linked glycan receptors, allowing it to transmit efficiently. Here we investigate the interconversion mechanism between two predicted binding modes in both 2009 and 1918 HAs, introducing a highly parallel intermediate network search scheme to construct kinetically relevant pathways efficiently. Accumulated mutations at positions 183 and 224 that alter the size of the binding pocket are identified with the fitness of the 2009 pandemic virus carrying the D222G mutation. This result suggests that the pandemic H1N1 viruses could gain binding affinity to the $\alpha2,3$-linked glycan receptors in the lungs, usually associated with highly pathogenic avian influenza, without compromising viability.
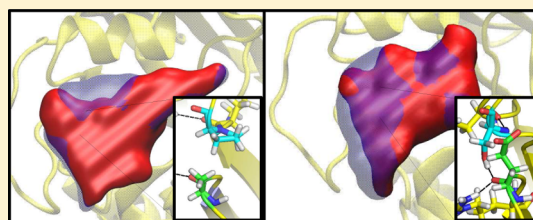
1. INTRODUCTION

2009 saw the first influenza pandemic of the 21st century, when an H1N1 influenza A virus spread to over 200 countries, resulting in more than 18 000 deaths globally.[1−3] The wide spread of the virus resulted from its efficient human-to-human transmission.[4−6] Transmission requires binding of the virus surface envelope protein, hemagglutinin (HA), to the host surface glycan receptors containing terminal sialic acid (SA).[7−9] The H1N1 human flu virus preferentially binds to receptors with a $\alpha2,6$ linkage between SA and galactose (Gal). These $\alpha2,6$-linked sialic acids ($\alpha2,6$-SA) are predominantly found in the epithelial cell surface of the human upper respiratory tract.[10−14] In contrast, the HAs of avian-adapted virus preferentially bind to $\alpha2,3$-SA, which are found in the digestive and respiratory tracts of birds, and on human lung alveolar cells.[9,10,12,15] Although the overall case fatality rate of 2009 H1N1 was not much different from seasonal flu, a specific mutation in HA at position 222 (H1 numbering) from aspartic acid to glycine (D222G) was more likely to be associated with severe or fatal cases.[16−20] The presence of this D222G mutation in both the HA from 1918 H1N1 and 2009 H1N1 viruses was reported to enhance HA binding to $\alpha2,3$-SAs, potentially causing the severe cases observed early in the 2009 pandemic.[21−26] The results also showed that the D222G mutation significantly weakened binding to $\alpha2,6$-SA for HA from 1918 H1N1, but only modestly affected the 2009 H1N1 virus.[21,22,24,26]

To explain this discrepancy, we previously performed computer simulations for the $\alpha2,6$-SA analog 6′SLN binding with HAs from the 1918 H1N1 (A/South Carolina/1/18, SC18) and 2009 H1N1 (A/Netherlands/602/2009, NL602) viruses with and without the D222G mutation.[21] For wild type SC18 (SC18-WT), two binding modes for $\alpha2,6$-SA were

identified. As shown in Figure 1b, these modes differ in the interactions formed between $\alpha2,6$-SA and HA, which alter the
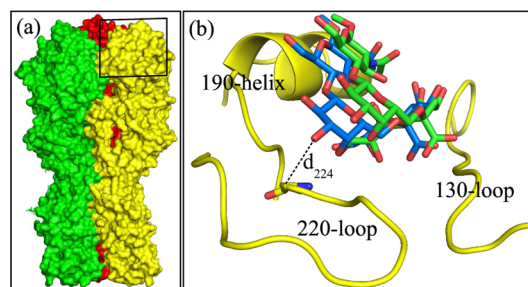

Figure 1. (a) Model of a HA trimer. The binding pocket region distal to the viral membrane is highlighted for one monomer. (b) Schematic illustration of the two binding modes; the 6′SLN glycan in mode 1 and mode 2 is colored blue and green, respectively. The distance between the α-C of 224A and the O3 atom of Gal (denoted as d_{224}) is used to distinguish the two binding modes in the present work.

position of the galactose sugar within the receptor binding pocket. Structures with shorter distances between the ligand and the residue at the 224 position are classified as mode 1, and those with longer distances are identified as mode 2. By definition, the ligand in mode 1 is buried deep inside the pocket, while in mode 2 it is situated higher up. NL602-WT was also observed to adopt two similar binding modes alongside many intermediate structures. For SC18 carrying the D222G mutation (SC18−D222G), almost no binding mode 2 was observed, while for NL602−D222G, although a

decrease in binding mode 2 was seen, it was not entirely absent. This contrasting stability is a result of the additional interaction between α2,6-SA and the side chains of 130K, 142K, and 224E, which can stabilize α2,6-SA binding to HA. We suggested that these interactions allow NL602−D222G to maintain binding to α2,6-SA, explaining the fitness of this mutant virus strain.

Other efforts have also been made to understand how the D222G mutation changes the binding preference, in both experiments and computer simulations. In 2013, Zhang et al.[22] reported the crystal structures of HA from both H1N1 A/California/04/2009 and SC18 flu virus bound to α2,6-SA and α2,3-SA analogs. They suggested that the D222G mutation results in a missing salt bridge between 222D and 219K, loosening the 220-loop and enabling the key residue 223Q to interact with the avian receptor, switching the binding preference. Using NMR and molecular dynamics (MD) simulation, Elli et al.[14,27] also reported stable structures of α2,6-SA binding to HA from 1918 flu before and after D222G mutation. The results suggested that the interaction between α2,6-SA and the 220-loop in wild type HA no longer exists after the mutation, which allows the galactose sugar to move away from the protein surface. Priyadarzini et al.[28] also performed MD simulations investigating the binding of Sialyldisaccharides to various HAs. Using Neu5Acα(2−6)Gal as the analog of α2,6-SA and HA from the A/swine/Iowa/30 virus, they also observed two binding modes. Due to the difference of the analog and HA model, these two binding modes are not exactly the same as those reported in 2010,[21] most likely due to differences in the HA model and glycan analogue employed.

The previous work described above mainly focused on the differences between static structures; the kinetics of the binding process were not addressed. Having identified alternative binding modes, one can investigate not only their kinetic stability, both in terms of on/off rates and the free energy of binding, but also how readily they interconvert. This information, especially the transition mechanism, the barriers, and the key intermediates, could be helpful in identifying additional therapeutic targets and provide a more complete understanding of the effect of a particular mutation on viral fitness. In the present work, we investigate interconversion pathways between two binding modes to elucidate the contrasting affinity of the two different H1N1 virus mutants for α2,6-SA. A new intermediate network scheme is introduced to accelerate the search for kinetically relevant paths in the potential energy landscape. By comparing the distribution of energy barriers and energy changes of key rearrangements, we find that the D222G mutation exchanges the energetic preference of the two binding modes in both SC18 and NL602 HAs and that the effect is more dramatic in SC18. By analyzing the shape of the binding pocket, we have identified differences between the two binding modes for the 1918 and 2009 strains. Our results indicate that changes at positions 183 and 224 generate a smaller pocket in NL602 than in SC18, permitting greater flexibility in the α2,6-SA transformation between the two modes in NL602.

2. COMPUTATIONAL DETAILS

2.1. Simulation Setup. The AMBER ff99SB[29−31] molecular mechanics force field in conjunction with the GB^OBC Generalized Born implicit solvent method ($igb = 2$)[32] was used to represent the protein. The potential was modified to remove a discontinuity at the nonbonded interaction cutoff, which was set to 16 Å. The trisaccharide, NeuAcα2,6-Galβ1−

4GlcNAcβ1 (6′SLN), with a methyl cap attached to the O1 atom of the terminal sugar, was used to represent α2,6-SA in our simulations, since it has been seen experimentally to bind to a range of H1N1 flu viruses.[33,34] Parameters for the three sugars that comprise this analog (NeuAc, Gal, and GlcNAc) were obtained from the GLYCAM06g[35] force field, as in our previous study, where the binding modes were first identified.[21]

2.2. Structure Generation. The initial structure of the HA in NL602-WT was based upon the crystal structure in the SC18 H1N1.[21] The order parameter used as the criterion to choose initial structures was the separation of the center of mass of the backbone atoms of residue 224 and the center of mass of the O2 and O3 atoms of the α2,6-SA galactose sugar. This order parameter was chosen because it was found to demarcate two distinct binding modes, which had been observed in MD snapshots from earlier work.[21] Once these snapshots had been sorted into one of these binding modes (or intermediate structures), we selected representatives based on the following criteria: (i) structures temporally separated from an observed transition between modes and (ii) structures clearly assigned as mode 1 or mode 2. These choices ensure that the structures are indeed representative, prior to minimization.

For SC18-WT and NL602-WT, the selected structures were first locally minimized. We then manually interpolated between the two end points, removing those differences in binding mode 2 far away from the pocket, which do not affect interconversion behavior. For SC18−D222G, binding mode 2 was entirely absent from the MD trajectory reported previously.[21] Hence, in the current work, a structure for mode 2 was generated using the ligand from SC18-WT, superimposing it into the pocket of SC18−D222G, incorporating residue moves around the pocket corresponding to binding mode 2, and finally performing a local minimization. For NL602−D222G, although some of the structures seen in previous work exhibit features of binding mode 2,[21] the total number is very small. In the present work, binding mode 2 was also generated using the ligand from NL602-WT mode 2, superimposing it into the binding pocket, then minimizing. Figure 2 shows the structure of all the end points after minimization. A detailed description can be found on our website[36] and in the Supporting Information.

2.3. Identifying the Fastest Path. The doubly nudged elastic band (DNEB) method,[37−39] as implemented in the OPTIM program,[40] was used to generate likely candidate structures for transition states along the pathway between a pair of end point structures. These stationary points were then refined accurately using hybrid eigenvector-following.[41,42] The two minima that each transition state connects were identified by calculating approximate steepest-descent paths. All the minima and transition states characterized during the connection of each pair of end points were collected in the initial database generated by the intermediate network scheme, as described in the following section. The databases were then expanded using the discrete path sampling (DPS) framework,[43−45] with several SHORTCUT, SHORTCUT BARRIER, and UNTRAP runs,[46,47] as implemented in the PATHSAMPLE program.[48] Dijkstra's shortest-path algorithm[49] with appropriate edge weights[43,45] was used to determine the discrete path that makes the largest contribution to the rate constant between the end points of interest (the fastest path).

2.4. Intermediate Network Scheme. Efficiently finding a connected pathway between two significantly different end points can be very difficult if attempted in a single step due to
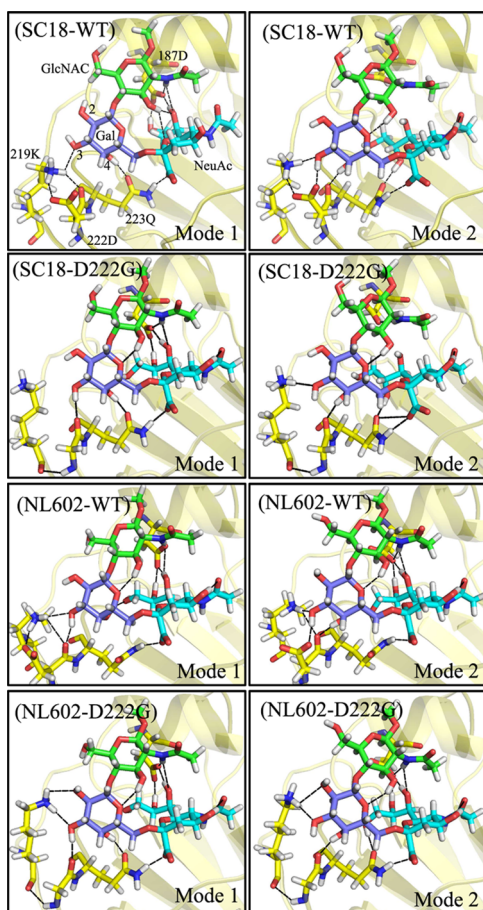
Figure 2. Structures of all initial end points. The names of key residues in SC18-WT binding mode 1 are labeled.
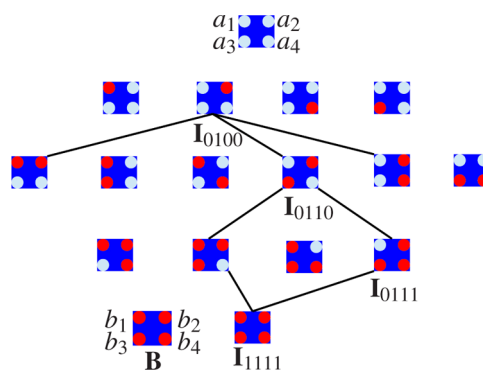

Figure 3. Schematic illustration of the intermediate network scheme. The blue square represents end points **A** and **B**. Four small circles represent the variable fragments in **A** and **B**, which are white ($a_1 \cdots a_4$) in **A** and red ($b_1 \cdots b_4$) in **B** (**B** $\cong$ **I**$_{1111}$). Six intermediates and possible connections among them are highlighted.

limitations in how intelligently we can interpolate between them. We found that the main differences between the two end points are mainly rotations of 6′SLN and the side chains of residues on the edge of the binding pocket. The main body of the complex, including the position of the center of mass of 6′SLN in the pocket, is almost the same. This observation inspired a new intermediate network interpolation technique to optimize refinement of the kinetic transition network and ensure that we have included as many paths as possible in the initial database. This scheme is illustrated in Figure 3 and consists of the following steps:

(1) For two end points **A** and **B**, identify all the differences of local conformations and label these differences as a_i and b_i, ($i = 1, n$). The end point **A** can then be denoted as the set **A** = $\{a_1, a_2, ..., a_n\}$. For example, in Figure 3, the molecular fragment in question is represented by the blue square, which should be similar for **A** and **B**. Four differences are identified and labeled with small circles, which are white in **A** and red in **B**.

(2) New configurations are generated for end point **A** by transplanting the different local **B** conformations into the original **A** minimum. We adopt the notation **I**$_\epsilon$ to denote the minimum obtained by relaxing the structure after the local conformation replacement, where ϵ is a binary string, identifying the origin of each local configuration. Those originally from **A** and **B** are denoted as "0" and "1", respectively. The leftmost digit represents a_1 or b_1. For example, in Figure 3, **I**$_{0110}$ = $\{a_1, b_2, b_3, a_4\}$

corresponds to the minimum obtained by relaxing the structure with the b_2 and b_3 conformations replacing a_2 and a_3. Hence **I**$_{0\cdots0}$ = **A**. We note that **I**$_{1\cdots1}$ can differ in detail from **B** because the local relaxations after transplanting conformations can produce minor changes in nearby side chains. In principle the total number of **I**$_a$ is 2^n, including **A**. However, some minima obtained after relaxation may be the same, because certain substitutions may have to follow a particular sequence to be favorable.

(3) Double-ended connection attempts are then run for all the possible **I**$_\epsilon$ and **I**$_\gamma$ corresponding to a single local conformation change between ϵ and γ. Here we require the pathway search algorithm to locate multiple transition states, ensuring that any given intermediate pairs will eventually be connected, as implemented in the "missing connection" approach,[50] which is coded in OPTIM.[40] In Figure 3, we have shown several possible connections. In principle, the total number of connections covered by this network could be $n \times 2^{n-1} + 1$, where the extra +1 corresponds to the case when **I**$_{1\cdots1}$ $\neq$ **B**. The number will be less than this upper bound if some of the minima generated in step 2 are the same.

We consider this procedure to be converged when pathways have been found between all of the states **I**. One advantage of the scheme is that all the minimizations in step 2 and the connection attempts in step 3 can be run in parallel independently. The corresponding pathways can have multiple transition states with additional local minima stabilized by nonlocal interactions. However, they are precisely the paths that are likely to have the lowest barriers and make the largest contribution to the overall rate constant. For the ideal case that **I**$_{1\cdots1}$ = **B**, and any pair of intervening minima have only one possible connection, corresponding to a single transition state, then in principle, this network will cover all the possible paths that connect the two end points. In practice, we may have to group more than one residue into a single fragment so that one single connection in step 3 still contains several elementary steps and alternative possibilities. Hence, it is still worthwhile to further refine the initial database that includes all the minima and transition states from the intermediate network connection searches.

3. RESULTS AND DISCUSSION

3.1. Fastest Paths for the Four Strains. The glycan binding pocket is located on the top of each monomer of the HA trimer. As shown in Figure 1b, it consists mainly of three secondary structure features: the 190-helix (residues 187 to 195), 220-loop (residues 218 to 225), and 130-loop (residues 132 to 135).[2,22,51] The aspartic acid residues located at the 222 and 187 positions contribute most to the binding of $\alpha2,6$-SA. By definition, the fastest path makes the largest contribution to the rate constant between the end points of interest. Analyzing this path helps us to understand how displacements of individual residues contribute to the overall interconversion mechanism between the two binding modes. The fastest paths in each of the four strains all involve loss of the interaction between NeuAc and 187D and formation of a new hydrogen-bond between Gal and the 220-loop. Several motions are shared by the fastest paths of all four strains, namely (1) the rotation of the 4-hydroxyl of Gal toward the backbone of 222D/G (G_4 movement), (2) the 3-hydroxyl of Gal binding to the side chain of 222D of wild type HAs, or moving to where it would have bound to the side chain of 222D in mutant HAs with 222G (G_3 movement), and (3) the rotation of the side chain of 187D toward the outside of the binding pocket (R_{187} movement).

For the fastest path in SC18-WT, we can distinguish four sets of movements. NeuAc first loses interactions with 187D by rotating the 7-hydroxyl and forms new intramolecular hydrogen-bonds, after which the G_4 and G_3 movements occur in succession. The 3-hydroxyl of Gal points toward the 2-hydroxyl of Gal in the binding mode 1, which is unique to the SC18-WT strain. These two steps together involve a highest barrier of 7.8 kcal/mol and make the largest contribution to the distance change between the α-C of 224A and the O3 atom of Gal (denoted as d_{224}), which distinguishes the two binding modes. The final change is the R_{187} motion, where 187D loses most of its interactions with the ligand, leaving only one hydrogen-bond with GlcNAc. Further details of the fastest path can be found in the Supporting Information.

In SC18−D222G, the position of the ligand is first changed by the rotation of the 8-hydroxyl of NeuAc, which breaks its hydrogen-bond with 187D and forms new hydrogen-bonds within NeuAc itself. This step has a high barrier of 14.1 kcal/mol and is endothermic by 11.7 kcal/mol. Next are the G_4 and G_3 movements, which are similar to those in SC18-WT, except that the 3-hydroxyl of Gal interacts with the backbone of 222G in binding mode 1. The G_3 movement still results in the largest change in d_{224} but the corresponding energy changes are much smaller than those in SC18-WT. The final steps again correspond to motion of R_{187}.

For NL602-WT, the first steps involve breaking the hydrogen-bond between 224E and the 2-hydroxyl of Gal, followed by conformational changes of the 224E side chain. This step does not exist in SC18 since the 224A side chain in SC18 is too short to interact with the ligand. The 9-hydroxyl of Gal then rotates, forming a hydrogen-bond with 224E, which shifts the whole 220-loop toward the ligand. The G_4, G_3, and R_{187} movements that follow have two interesting differences from previous pathways: (1) For both SC18-WT and SC18−D222G strains, we observed translational motion of the ligand toward the 220-loop when the G_4 and G_3 movements happen, which is not seen in NL602-WT. (2) After the R_{187} motion, the ligand in SC18-WT and SC18−D222G moves away from the

base of the pocket, losing interactions with 187D, while in NL602-WT, most of the hydrogen-bonds remain after the R_{187} displacement. These differences are discussed further below.

For the fastest path in NL602−D222G, the R_{187} displacement comes first, together with hydrogen-bond breaking between 224E and the 2-hydroxyl of Gal. As discussed for NL602-WT, no interactions between 187D and the ligand are lost. Then, following the G_4 and G_3 movements, d_{224} increases from 5.2 to 6.4 Å. The final two steps correspond to motion of residues 180 H and 191L toward the base of the pocket.

3.2. Kinetics of the Key Elementary Step. By comparing the fastest path that connects the two binding modes in each case, we find that the shift of the 3-hydroxyl of Gal, either binding to the side chain of 222D of wild type HAs, or moving to where it would have bound to the side chain of 222D in mutant HAs with 222G (the G_3 movement), is the key elementary step differentiating the two binding modes. This step is directly affected by the D222G mutation.

Although the G_3 displacement is common to all four cases, the fastest transition paths exhibit systematic differences. In SC18-WT, the 3-hydroxyl does not bind to the backbone first before binding to the side chain of 222D. In NL602-WT this change, and the binding of the 4-hydroxyl of Gal to the backbone of 222D, occur simultaneously. To compare further we collect all the transition states corresponding to steps where the 3-hydroxyl of Gal either binds to the side chain of 222D in wild type HAs, or moves to the place where it would have bound to the side chain of 222D in mutant HAs with 222G. The barrier and energy change distributions for the four cases are illustrated in the contour plots shown in Figure 4. In SC18-WT, 127 out of 2742 transition states in the database correspond to this step. The peak value of the distribution is around (0.7, −1.7), corresponding to an energy barrier of
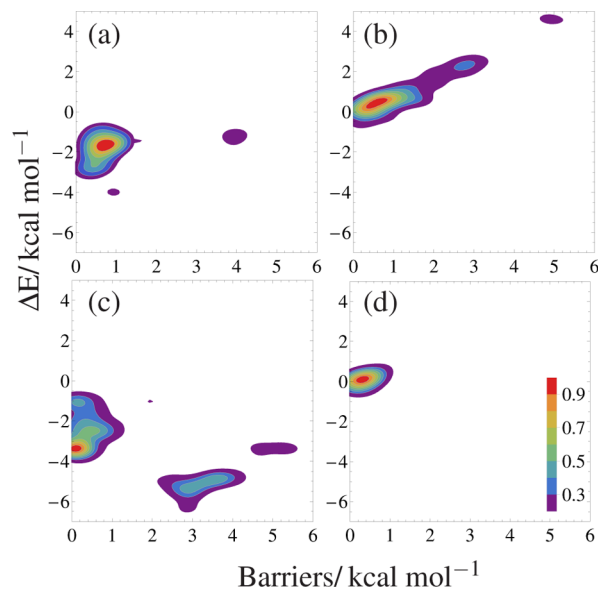

Figure 4. Contour plots illustrating the distribution of the barriers and the energy changes corresponding to the 3-hydroxyl of Gal, either binding to the side chain of 222D in (a) SC18-WT and (c) NL602-WT, or moving to where it would have bound to the side chain of 222D in (b) SC18−D222G and (d) NL602−D222G with 222G. The height of the maximum peak in the distributions is scaled to unity. The x axis corresponds to the height of the barrier and the y axis corresponds to the energy difference.

0.7 kcal/mol and an energy difference of −1.7 kcal/mol. The pathways that contribute to this peak all involve breaking the hydrogen-bond between the 3-hydroxyl of Gal and the backbone of 222D. The transition states that contribute to the shoulder peak around (0.5, −2.5) are broadly similar, with insignificant differences in nearby side chains, for example, the position of the 7-hydroxyl of NeuAc. The step in which the 3-hydroxyl of Gal breaks hydrogen-bonds with non-222D residues mainly contributes to the small peak at (3.9, −1.2), as we observed in the fastest path.

In SC18−D222G, 126 out of 3475 transition states in the database correspond to the key elementary step. The main peak shifts to (0.6, 0.5), which still mainly correlates with interactions of the 3-hydroxyl of Gal and the backbone of 222G in the initial state. The energy barrier remains almost the same as in the wild type HA, while the overall energy difference increases by at least 2.2 kcal/mol, as a result of the D222G mutation. The small peak located at around (2.8, 2.3) corresponds to cases where the 3-hydroxyl of Gal is not bound to the backbone of 222G in the initial state. The structure of the corresponding initial state is very similar to the one reported by Zhang et al.,[22] where the 3-hydroxyl binds to 219K and the 4-hydroxyl of Gal binds to the backbone of 222G.

In NL602-WT, 332 out of 3719 transition states in the database correspond to the 3-hydroxyl motion. The main peak in Figure 4c is around (0.1, −3.4) and represents paths where the 3-hydroxyl breaks the hydrogen-bond to the backbone of 222D, while simultaneously binding to the side chain of 222D. Compared to SC18-WT, this step is significantly faster and more exothermic in NL602-WT. The same mechanism also contributes to the broad peak ranging from (2.7, −5.5) to (3.6, −4.9), and in these paths, the 4-hydroxyl of Gal is already bound to the backbone of 222D in the initial state. The 4-hydroxyl subsequently binds to the backbone of 222D, as in the fastest path. Another small peak located at around (0.1, −1.0) corresponds to paths where the 3-hydroxyl of Gal does not interact with the backbone of 222D in the initial state.

In NL602−D222G, 128 out of 3524 transition states in the database correspond to 3-hydroxyl motion. Only one peak located at (0.3, 0.1) can be seen in Figure 4d, corresponding to paths with the 3-hydroxyl of Gal bound to the backbone of 222G in the initial state.

For wild type HAs, we see that binding the 3-hydroxyl of Gal to the side chain is generally an energetically favorable process. When the hydroxyl shifts from interacting with the backbone to the side chain, the barrier is always lower than 1 kcal/mol. As a result, mode 2 corresponds to an energy minimum in this transformation, increasing the equilibrium population and making it accessible during MD simulations, as observed in previous work.[21]

For mutant HAs, mode 2 is generally energetically unfavorable compared to mode 1. In particular, for SC18−D222G mode 2 is so unfavorable that the reverse process (mode 2 to mode 1) has a barrier of only 0.1 kcal/mol. According to the energy profile, other processes, such as the rotation of the 187D residue, are also more endothermic in SC18−D222G than in NL602−D222G. As a result, the equilibrium occupation probability of mode 2 in SC18−D222G is rather low, which is consistent with the previous observation that mode 2 is rarely observed during MD simulations.[21] For NL602−D222G, the paths are similar, but the energy difference is lower. Hence, binding mode 2 can still be observed in the NL602−D222G mutant HA, albeit with

decreased probability. A more detailed analysis of this key elementary step will be addressed in future work

3.3. Shape of the Binding Pocket. One reason why the binding of the 3-hydroxyl of Gal to the side chain of 222D is more exothermic and has a lower barrier in NL602 may be the differences in the shape of the binding pocket. When the ligand binds to 222D/G or unbinds from 187D in SC18, we always observe a significant shift of the Gal and GlcNAc sugars toward the 222 position. However, in NL602, there is no such translation. We examined the shape of the pocket for the four strains using the POVME program[52,53] and illustrate the differences observed in Figure 5. Structures were aligned based
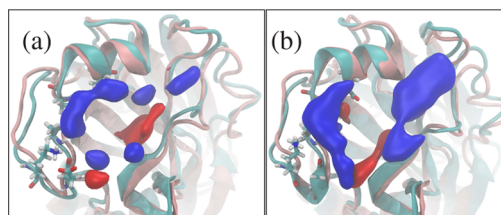

Figure 5. Pocket shape differences between (a) SC18-WT (cyan) and NL602-WT (pink) and (b) SC18−D222G (cyan) and NL602−D222G (pink). The blue and red regions correspond to areas where the pocket in SC18 is larger and than that in NL602, respectively.

on the 190-helix, 130-loop, and 220-loop using VMD.[54,55] Two major differences between SC18-WT and NL602-WT can be observed in Figure 5a: (1) The pocket in SC18-WT is larger than that in NL602-WT on the side of 220-loop and 190-helix, where the majority of the interactions between ligand and pocket are. (2) The pocket in NL602-WT is larger on the side of the 130-loop and also between the 220-loop and 130-loop. Comparing D222G mutants in Figure 5b, the blue region is larger than that in Figure 5a. However, the extra volume is located mainly at the top of the pocket, which may arise due to measurement error, since the pockets are open at the top.

To analyze the pocket shape differences further, in Figure 6 we show a triangular region defined by 187D, 219K, and 222D/G on the edge of the binding pocket.
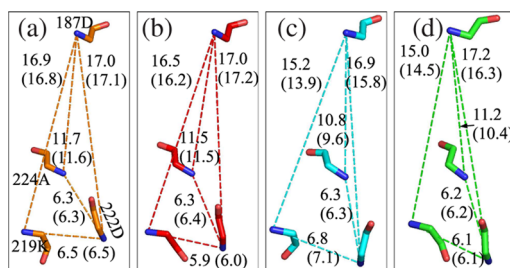

Figure 6. (a−d) Triangle formed by 187D, 219K, and 222D/G in mode 1 of (a) SC18-WT, (b) SC18−D222G, (c) NL602-WT, and (d) NL602−D222G, respectively. The distances between residues are calculated between the backbone nitrogen atoms, and the values in brackets are the corresponding distances in mode 2.

The shape of this region provides a simple representation of the 220-loop side of the binding pocket, where the majority of the rearrangements between the two binding modes are localized. We find that the 219K residue in SC18 is about 1−2 Å further away from 187D but slightly closer to 222D/G compared to NL602. As a result, 219K is nearer to the center of the pocket in NL602.

Another difference is the position of residue 224. In SC18, 224A interacts with the 4-hydroxyl of Gal in binding mode 1 and 224E in NL602 interacts with the 3- and 4-hydroxyl of Gal. Hence, 224A/E mainly determines how far to the left (from the viewpoint of Figure 6) and how deep α2,6-SA binds within the pocket. In SC18, the position of 224A is close to the center of the triangle. The distance between 224A and the line connecting 187D and 222D/G is 2.8 Å in the wild type HA and 2.5 Å in the 222G mutant HA. In NL602, the position of 224E is close to the line connecting 187D and 222D/G. The distance between 224E and the line connecting 187D and 222D/G is 1.1 Å in the wild type HA and 1.4 Å in the mutant HA, and the distance between 224A and 187D in SC18 is also longer than that in NL602. Since 222D protrudes slightly from the edge of the pocket, the position of 224A in SC18 allows the ligand to bind deeper and closer to the left side. The variations in the distance between the 220-loop and 190-helix confirm the shape differences observed from the POVME measurement: the pockets in SC18-WT and SC18−D222G are larger than for NL602-WT and NL602−D222G on the side of the 220-loop.

We also used POVME to identify changes in pocket shape during the interconversion. For SC18-WT, the first nontrivial shape change of the pocket results from the rotation of the 4-hydroxyl of Gal (Figure 7a), which shrinks the volume between the 220-loop and the 130-loop. This reduction in size is caused by both the conformational change of the 223Q side chain and the movement of the 220-loop toward the 130-loop. Another significant shape change results from the rotation of 187D (Figure 7b), which reduces the size of the region in front of 187D. The observations in the case of SC18−D222G are similar to those for SC18-WT. The first few movements, including the rotation of the 8-hydroxyl of Gal and the movement of the 4- and 3-hydroxyl of Gal, contribute to the shrinking of the pocket between the 220- and 130-loops, as shown in Figure 7c. The side chain rotation of 187D then removes the region in front of 187D (Figure 7d).

The changes in pocket shape for NL602-WT differ greatly from those for SC18-WT. The rotation of the 9-hydroxyl of Gal shrinks the pocket between the 220-loop and 190-helix and enlarges the volume between the 190-helix and 130-loop, as shown in Figure 7e. The rotation of 187D, however, contributes less to the shape change than it does for SC18-WT, which is consistent with the previous observation that the rotation of 187D does not remove many interactions with the ligand. The observations for the NL602−D222G simulations are similar to those for NL602-WT. The rotation of 187D slightly shrinks the region in front of 187D, and enlargement of the pocket next to 224E occurs via movement of the 224E side chain (Figure 7g). The next few steps, including the movement of the 4- and 3-hydroxyl and residues 180 H and 191L, slightly shrink the pocket between the 220-loop and 190-helix, as for NL602-WT.

In summary, the two strains from 1918 South Carolina are more flexible between the 220-loop and 130-loop but more rigid between the 220-loop and 190-helix when compared to the 2009 Netherlands strains. This observation is consistent with the distance changes among key residues (Figure 6). The maximum distance variation in SC18-WT is only 0.3 Å for 187D and 222D. However, in NL602, the distance from 187D to residues 219K, 222D, and 224A is reduced by roughly 1 Å in binding mode 2. Since the ligand mainly interacts with the 190-helix and the 220-loop, the flexibility of this region plays an important role in the two binding modes. When the
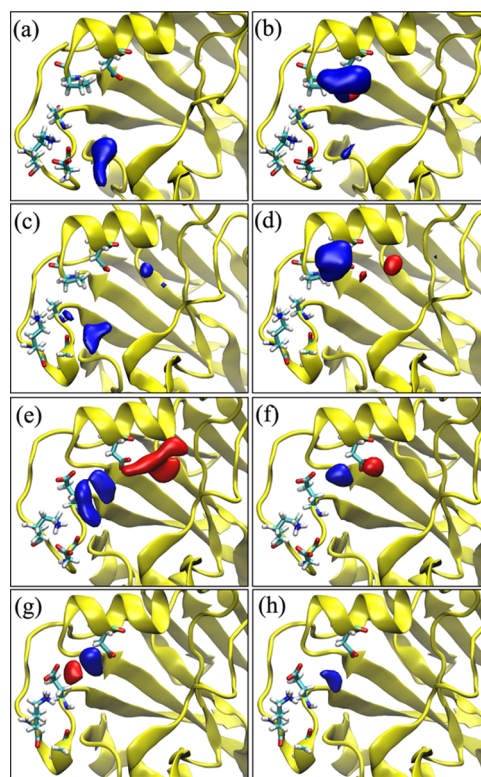

Figure 7. Binding pocket changes during the interconversion between the two modes in four different strains. The blue and red regions correspond to volumes lost and gained after a particular displacement, respectively. (a) G_4 movement in SC18-WT (m3−m4 in Figure S1). (b) R_{187} movement in SC18-WT (m5−m9 in Figure S1). (c) Rotation of the 8-hydroxyl of Gal and G_4 and G_3 shifts in SC18−D222G (m1−m9 in Figure S3). (d) R_{187} movement in SC18−D222G (m9−m11 in Figure S3). (e) Rotation of the 9-hydroxyl of Gal and G_4 shifts in NL602-WT (m4−m6 in Figure S5). (f) R_{187} movement in NL602-WT (m7−m9 in Figure S5). (g) R_{187} movement in NL602−D222G (m5−m7 in Figure S7). (h) G_4, G_3, 180H, and 191L movement in NL602−D222G (m7−m11 in Figure S7).

transformation between the two modes occurs, due to the shape and the size of the pocket in NL602, the 220-loop can adapt its position to reflect the position of the ligand, which allows most interactions with the edge of the pocket to be conserved. However, for SC18, because the pocket is larger, the ligand must move from one side to the other during the transformation to maintain these interactions, and the 220-loop cannot adapt to the ligand position as flexibly as for NL602. As a result, the binding of α2,6-SA to SC18 relies more heavily on the side chain of 222D than in NL602, since the ligand has limited access to the edge of the pocket. The additional interactions between the ligand and the pocket in NL602 that were reported previously can also been explained in the light of this observation.[21]

Interestingly, as shown in Figure 8a, the junction structures between the 190-helix and 220-loop, which determine the relative position of the two regimes, are different in the two HAs. In SC18 183P is in front of the 190-helix and does not interact with other residues nearby. The hydrogen-bonds between 215A and 182P, together with 217R and 224A, determine the relative positions of the 190-helix and the 220-loop. The residues at the junction between the 190-helix and 220-loop change from 183P and 224A in SC18, to 183S and 224E in NL602. Due to these differences, the shape of the
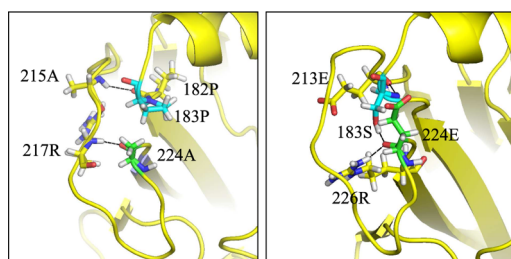
Figure 8. Key residues in (a) SC18 and (b) NL602 that determine the relative positions of the 190-helix and 220-loop.

pocket in NL602 is affected by the interaction between 183S and 213E, as well as the interaction between 183S, 224E and 226R, as shown in Figure 8b.

There have been several experimental results reported since 2010 that the S183P mutation[56−61] and the E224A mutation[62−64] with or without other mutations, are correlated with the pathogenesis and transmission changes observed for H1N1. Since the majority of the shape difference between SC18 and NL602 during the binding mode interconversion occurs between the 190-helix and 220-loop, the accumulated mutations at positions 183 and 224 may explain the differences in pocket shape flexibility during binding mode interconversion observed here.

4. CONCLUSION

In the present work, we have investigated the transformation mechanism between two binding modes for α2,6-SA in both the wild type and mutant H1N1 hemagglutinin proteins from 1918 and 2009 pandemic flu using the 6′SLN ligand. We introduced an intermediate network scheme to accelerate the search process of pathways connecting the two binding modes. The advantage of this procedure is that searches can be systematically conducted in parallel for all the single changes in identifiable local conformational states. By comparing the paths predicted to dominate the kinetics, we have identified the key elementary step that distinguishes the two binding modes, namely the 3-hydroxyl of Gal either binding to the side chain of 222D in wild type HAs or moving to the place where it would have bound to the side chain of 222D in mutant HAs with 222G. We find that the different residues at the 183 and 224 positions in the two HAs change the connecting pair between the 190-helix and 220-loop, and hence the shape of the pocket. The smaller size of the binding pocket in NL602 allows the ligand to interact with more residues in NL602 than in SC18 and also makes the 220-loop more flexible.

The D222G mutation exchanges the energy preference between the two binding modes in both SC18 and NL602 hemagglutinins. However, the transition between mode 1 and mode 2 raises the energy more in SC18 than in NL602, probably because of the different shape of the binding pocket.

The different shape, the flexibility of the pocket during binding mode interconversion, and the resulting energy profile changes provide insight into the contrasting pathogenicities induced by the D222G mutation in the SC18 and NL602 viruses. These results may help to explain the difference between the impact of the 1918 and 2009 H1N1 strains. A change in preference from α2,6- to α2,3-linked sialic acid is associated with a shift in pathogenicity. However, in the 2009 D222G mutant, the 2,6 affinity is not weakened as much as for the corresponding 1918 mutant. Understanding such dis-

tinctions at an atomic level of detail should be helpful in predicting the potential emergence of strains that may pose serious threats to human health.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The website for the structures of end points and the example input for PATHSAMPLE. Structural details of each end point. The energy profile of the fastest paths. Detailed descriptions of each path. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: cs778@cam.ac.uk (C.S.).
*E-mail: csw34@cam.ac.uk (C.S.W.).
*E-mail: khs26@cam.ac.uk (K.H.S.-C.).
*E-mail: dw34@cam.ac.uk (D.J.W.).

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Resende, P. C.; Motta, F. C.; Oliveira, M. d. L. A.; Gregianini, T. S.; Fernandes, S. B.; Cury, A. L. F.; do Carmo D.Rosa, M.; Souza, T. M. L.; Siqueira, M. M. *PLoS One* **2014**, *9*, e92789.

(2) Xu, R.; McBride, R.; Nycholat, C. M.; Paulson, J. C.; Wilson, I. A. *J. Virol.* **2012**, *86*, 982−990.

(3) Medina, R. A.; Garcia-Sastre, A. *Nat. Rev. Microbiol.* **2011**, *9*, 590−603.

(4) Fraser, C.; Donnelly, C. A.; Cauchemez, S.; Hanage, W. P.; van Kerkhove, M. D.; Hollingsworth, T. D.; Griffin, J.; Baggaley, R. F.; Jenkins, H. E.; Lyons, E. J.; Jombart, T.; Hinsley, W. R.; Grassly, N. C.; Balloux, F.; Ghani, A. C.; Ferguson, N. M.; Rambaut, A.; Pybus, O. G.; Lopez-Gatell, H.; Alpuche-Aranda, C. M.; Chapela, I. B.; Zavala, E. P.; Guevara, D. M. E.; Checchi, F.; Garcia, E.; Hugonnet, S.; Roth, C.; Collaboration, T. W. R. P. A. *Science* **2009**, *324*, 1557−1561.

(5) Munayco, C. V.; Gomez, J.; Laguna-Torres, V. A.; Arrasco, J.; Kochel, T. J.; Fiestas, V.; Garcia, J.; Perez, J.; Torres, I.; Condori, F.; Nishiura, H.; Chowell, G. *Euro Surveill* **2009**, *14*, 6−10.

(6) Nishiura, H.; Castillo-Chavez, C.; Safan, M.; Chowell, G. *Euro Surveill* **2009**, *14*, 2−5.

(7) Sauter, N. K.; Bednarski, M. D.; Wurzburg, B. A.; Hanson, J. E.; Whitesides, G. M.; Skehel, J. J.; Wiley, D. C. *Biochemistry* **1989**, *28*, 8388−8396.

(8) Takemoto, D.; Skehel, J.; Wiley, D. *Virology* **1996**, *217*, 452−458.

(9) Gambaryan, A.; Tuzikov, A.; Piskarev, V.; Yamnikova, S.; Lvov, D.; Robertson, J.; Bovin, N.; Matrosovich, M. *Virology* **1997**, *232*, 345−350.

(10) de Graaf, M.; Fouchier, R. A. M. *EMBO J.* **2014**, *33*, 823−841.

(11) Ibricevic, A.; Pekosz, A.; Walter, M. J.; Newby, C.; Battaile, J. T.; Brown, E. G.; Holtzman, M. J.; Brody, S. L. *J. Virol.* **2006**, *80*, 7469−7480.

(12) Shinya, K.; Ebina, M.; Yamada, S.; Ono, M.; Kasai, N.; Kawaoka, Y. *Nature* **2006**, *440*, 435−436.

(13) van Riel, D.; Munster, V. J.; de Wit, E.; Rimmelzwaan, G. F.; Fouchier, R. A.; Osterhaus, A. D.; Kuiken, T. *Am. J. Pathol.* **2007**, *171*, 1215−1223.

(14) Elli, S.; Macchi, E.; Rudd, T. R.; Raman, R.; Sassaki, G.; Viswanathan, K.; Yates, E. A.; Shriver, Z.; Naggi, A.; Torri, G.; Sasisekharan, R.; Guerrini, M. *Biochemistry* **2014**, *53*, 4122−4135.

(15) Russell, R.; Stevens, D.; Haire, L.; Gamblin, S.; Skehel, J. *Glycoconjugate J.* **2006**, *23*, 85−92.

(16) Kilander, A.; Rykkvin, R.; Dudman, S. G.; Hungnes, O. *Euro Surveill* **2010**, *15*, 6−8.

(17) Chen, G.-W.; Shih, S.-R. *Emerg. Infect. Dis.* **2009**, *15*, 1897−1903.

(18) Puzelli, S.; Facchini, M.; Spagnolo, D.; de Marco, M. A.; Calzoletti, L.; Zanetti, A.; Fumagalli, R.; Tanzi, M. L.; Cassone, A.; Rezza, G.; Donatelli, I.; H1N1, S. G. P. *Emerg. Infect. Dis.* **2010**, *16*, 863−865.

(19) Mak, G. C.; Au, K. W.; Tai, L. S.; Chuang, K. C.; Cheng, K. C.; Shiu, T. C.; Lim, W. *Euro Surveill* **2010**, *15*, 15−16.

(20) Wedde, M.; Wählisch, S.; Wolff, T.; Schweiger, B. *PLoS One* **2013**, *8*, e57059.

(21) Chutinimitkul, S.; Herfst, S.; Steel, J.; Lowen, A. C.; Ye, J.; van Riel, D.; Schrauwen, E. J. A.; Bestebroer, T. M.; Koel, B.; Burke, D. F.; Sutherland-Cash, K. H.; Whittleston, C. S.; Russell, C. A.; Wales, D. J.; Smith, D. J.; Jonges, M.; Meijer, A.; Koopmans, M.; Rimmelzwaan, G. F.; Kuiken, T.; Osterhaus, A. D. M. E.; Garca-Sastre, A.; Perez, D. R.; Fouchier, R. A. M. *J. Virol.* **2010**, *84*, 11802−11813.

(22) Zhang, W.; Shi, Y.; Qi, J.; Gao, F.; Li, Q.; Fan, Z.; Yan, J.; Gao, G. F. *J. Virol.* **2013**, *87*, 5949−5958.

(23) Stevens, J.; Blixt, O.; Glaser, L.; Taubenberger, J. K.; Palese, P.; Paulson, J. C.; Wilson, I. A. *J. Mol. Biol.* **2006**, *355*, 1143−1155.

(24) Liu, Y.; Childs, R. A.; Matrosovich, T.; Wharton, S.; Palma, A. S.; Chai, W.; Daniels, R.; Gregory, V.; Uhlendorff, J.; Kiso, M.; Klenk, H.-D.; Hay, A.; Feizi, T.; Matrosovich, M. *J. Virol.* **2010**, *84*, 12069−12074.

(25) Tumpey, T. M.; Maines, T. R.; van Hoeven, N.; Glaser, L.; Solórzano, A.; Pappas, C.; Cox, N. J.; Swayne, D. E.; Palese, P.; Katz, J. M.; García-Sastre, A. *Science* **2007**, *315*, 655−659.

(26) Belser, J. A.; Jayaraman, A.; Raman, R.; Pappas, C.; Zeng, H.; Cox, N. J.; Katz, J. M.; Sasisekharan, R.; Tumpey, T. M. *PLoS One* **2011**, *6*, e25091.

(27) Sassaki, G. L.; Elli, S.; Rudd, T. R.; Macchi, E.; Yates, E. A.; Naggi, A.; Shriver, Z.; Raman, R.; Sasisekharan, R.; Torri, G.; Guerrini, M. *Biochemistry* **2013**, *52*, 7217−7230 PMID:24015903.

(28) Priyadarzini, T. R. K.; Selvin, J. F. A.; Gromiha, M. M.; Fukui, K.; Veluraja, K. *J. Biol. Chem.* **2012**, *287*, 34547−34557.

(29) Case, D.; Darden, T.; Cheatham, T.; Simmerling, C.; Luo, R.; Merz, K.; Pearlman, D.; Crowley, M.; Walker, R.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, S.; Gohlke, H.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Ross, W. *AMBER 9*; University of California, 2006.

(30) Pearlman, D.; Case, D.; Caldwell, J.; Ross, W.; Cheatham III, T.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

(31) Case, D.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Computat. Chem.* **2005**, *26*, 1668−1688.

(32) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins: Structure, Function, and Bioinformatics* **2004**, *55*, 383−394.

(33) Liao, H.-Y.; Hsu, C.-H.; Wang, S.-C.; Liang, C.-H.; Yen, H.-Y.; Su, C.-Y.; Chen, C.-H.; Jan, J.-T.; Ren, C.-T.; Chen, C.-H.; Cheng, T.-J. R.; Wu, C.-Y.; Wong, C.-H. *J. Am. Chem. Soc.* **2010**, *132*, 14849−14856.

(34) Blixt, O.; Head, S.; Mondala, T.; Scanlan, C.; Huflejt, M. E.; Alvarez, R.; Bryan, M. C.; Fazio, F.; Calarese, D.; Stevens, J.; Razi, N.; Stevens, D. J.; Skehel, J. J.; van Die, I.; Burton, D. R.; Wilson, I. A.; Cummings, R.; Bovin, N.; Wong, C.-H.; Paulson, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17033−17038.

(35) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Computat. Chem.* **2008**, *29*, 622−655.

(36) http://www-wales.ch.cam.ac.uk/examples/PATHSAMPLE/ (accessed March 31, 2015).

(37) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082−2094.

(38) Sheppard, D.; Terrell, R.; Henkelman, G. *J. Chem. Phys.* **2008**, *128*, 134106.

(39) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9978−9985.

(40) Wales, D. J. OPTIM: A Program for Optimising Geometries and Calculating Pathways. http://www-wales.ch.cam.ac.uk/OPTIM/ (accessed March 31, 2015).

(41) Munro, L. J.; Wales, D. J. *Phys. Rev. B* **1999**, *59*, 3969−3980.

(42) Kumeda, Y.; Wales, D. J.; Munro, L. J. *Chem. Phys. Lett.* **2001**, *341*, 185−194.

(43) Wales, D. J. *Mol. Phys.* **2002**, *100*, 3285−3305.

(44) Wales, D. J. *Mol. Phys.* **2004**, *102*, 891−908.

(45) Wales, D. J. *Int. Rev. Phys. Chem.* **2006**, *25*, 237−282.

(46) Carr, J. M.; Wales, D. J. *J. Chem. Phys.* **2005**, *123*, 234901.

(47) Strodel, B.; Whittleston, C. S.; Wales, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 16005−16014.

(48) Wales, D. J. PATHSAMPLE: A Program for Generating Connected Stationary Point Databases and Extracting Global Kinetics. http://www-wales.ch.cam.ac.uk/PATHSAMPLE/ (accessed March 31, 2015).

(49) Dijkstra, E. *Numer Math* **1959**, *1*, 269−271.

(50) Carr, J. M.; Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2005**, *122*, 234903.

(51) Stevens, J.; Corper, A. L.; Basler, C. F.; Taubenberger, J. K.; Palese, P.; Wilson, I. A. *Science* **2004**, *303*, 1866−1870.

(52) Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A. *J. Mol. Graphics Modell.* **2011**, *29*, 773−776.

(53) Durrant, J. D.; Votapka, L.; Sørensen, J.; Amaro, R. E. *J. Chem. Theory Comput.* **2014**, *10*, 5047−5056.

(54) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33−38.

(55) Eargle, J.; Wright, D.; Luthey-Schulten, Z. *Bioinformatics* **2006**, *22*, 504−506.

(56) Ilyushina, N. A.; Khalenkov, A. M.; Seiler, J. P.; Forrest, H. L.; Bovin, N. V.; Marjuki, H.; Barman, S.; Webster, R. G.; Webby, R. J. *J. Virol.* **2010**, *84*, 8607−8616.

(57) Ye, J.; Sorrell, E. M.; Cai, Y.; Shao, H.; Xu, K.; Pena, L.; Hickman, D.; Song, H.; Angel, M.; Medina, R. A.; Manicassamy, B.; Garcia-Sastre, A.; Perez, D. R. *PLoS Pathog.* **2010**, *6*, e1001145.

(58) O'Donnell, C. D.; Vogel, L.; Wright, A.; Das, S. R.; Wrammert, J.; Li, G.-M.; Zheng, N.-Y.; Song, H.; Yewdell, J. W.; Ahmed, R.; Wilson, P. C.; Subbarao, K. *mBio* **2012**, *3*, e00120-12.

(59) Marzano, A.; Marengo, A.; Ruggiero, T.; Allice, T.; Sanna, C.; Alessandria, C.; Morgando, A.; Sciandrello, M. C.; Franzin, A. M.; Rizzetto, M.; Ghisetti, V. *J. Med. Virol.* **2013**, *85*, 1−7.

(60) Ruggiero, T.; de Rosa, F.; Cerutti, F.; Pagani, N.; Allice, T.; Stella, M. L.; Milia, M. G.; Calcagno, A.; Burdino, E.; Gregori, G.; Urbino, R.; di Perri, G.; Ranieri, M. V.; Ghisetti, V. *Influenza Other Respir Viruses* **2013**, *7*, 1416−1426.

(61) de Vries, R. P.; de Vries, E.; Martínez-Romero, C.; McBride, R.; van Kuppeveld, F. J.; Rottier, P. J. M.; García-Sastre, A.; Paulson, J. C.; de Haan, C. A. M. *J. Virol.* **2013**, *87*, 13868−13877.

(62) van Doremalen, N.; Shelton, H.; Roberts, K. L.; Jones, I. M.; Pickles, R. J.; Thompson, C. I.; Barclay, W. S. *PLoS One* **2011**, *6*, e25755.

(63) Martínez-Romero, C.; de Vries, E.; Belicha-Villanueva, A.; Mena, I.; Tscherne, D. M.; Gillespie, V. L.; Albrecht, R. A.; de Haan, C. A. M.; García-Sastre, A. *J. Virol.* **2013**, *87*, 6507−6511.

(64) de Vries, R. P.; de Vries, E.; Moore, K. S.; Rigter, A.; Rottier, P. J. M.; de Haan, C. A. M. *J. Biol. Chem.* **2011**, *286*, 5868−5875.