Atomistic simulations indicate the functional loop-to-coiled-coil transition in influenza hemagglutinin is not downhill

Xingcheng Lin[a,b], Jeffrey K. Noel[c], Qinghua Wang[d], Jianpeng Ma[a,d,e], and José N. Onuchic[a,b,f,g,1]

[a]Center for Theoretical Biological Physics, Rice University, Houston, TX 77030; [b]Department of Physics and Astronomy, Rice University, Houston, TX 77005; [c]Crystallography, Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany; [d]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030; [e]Department of Bioengineering, Rice University, Houston, TX 77030; [f]Department of Chemistry, Rice University, Houston, TX 77005; and [g]Department of BioSciences, Rice University, Houston, TX 77005

Contributed by José N. Onuchic, June 13, 2018 (sent for review March 29, 2018; reviewed by Brian Dyer, Angel E. Garcia, and Zaida A. Luthey-Schulten)

Influenza hemagglutinin (HA) mediates viral entry into host cells through a large-scale conformational rearrangement at low pH that leads to fusion of the viral and endosomal membranes. Crystallographic and biochemical data suggest that a loop-to-coiled-coil transition of the B-loop region of HA is important for driving this structural rearrangement. However, the microscopic picture for this proposed "spring-loaded" movement is missing. In this study, we focus on understanding the transition of the B loop and perform a set of all-atom molecular dynamics simulations of the full B-loop trimeric structure with the CHARMM36 force field. The free-energy profile constructed from our simulations describes a B loop that stably folds half of the postfusion coiled coil in tens of microseconds, but the full coiled coil is unfavorable. A buried hydrophilic residue, Thr59, is implicated in destabilizing the coiled coil. Interestingly, this conserved threonine is the only residue in the B loop that strictly differentiates between the group 1 and 2 HA molecules. Microsecond-scale constant temperature simulations revealed that kinetic traps in the structural switch of the B loop can be caused by nonnative, intramonomer, or intermonomer β-sheets. The addition of the A helix stabilized the postfusion state of the B loop, but introduced the possibility for further β-sheet structures. Overall, our results do not support a description of the B loop in group 2 HAs as a stiff spring, but, rather, it allows for more structural heterogeneity in the placement of the fusion peptides during the fusion process.

all-atom molecular dynamics | B-loop transition | structural heterogeneity | buried water | sequence divergence

Influenza hemagglutinin (HA) is the membrane surface glycoprotein responsible for guiding the invasion of flu viruses into their host cells. It is a homotrimer with each monomer composed of two domains, a globular head domain (HA_1) that wraps around a stem domain (HA_2 together with HA_1 termini) (Fig. 1). HA_2 contains a C-terminal transmembrane peptide that anchors it to the viral membrane and a buried N-terminal fusion peptide (FP). The overall picture that has emerged from many years of study (1) is that the low-pH conditions in the endosome trigger a large conformational change in HA_2 that (*i*) allows the burial of the FPs into the endosomal membrane and (*ii*) guides the subsequent fusion of the viral and endosomal membranes. While the endpoints of the HA_2 structural rearrangement (2–4) are known from X-ray crystallography, how the intermediate steps of the conformational rearrangement are coordinated to perform these two crucial functions remains unclear. Thus, an understanding of the pH-induced structural transition of HA_2 will shed light on the physical mechanism of the invasion of influenza viruses.

A seminal discovery was that the postfusion structure of HA_2 is significantly more stable than the prefusion structure of HA_2 in the absence of HA_1 (3, 5). This means that the free energy released during refolding can be used to help drive the fusion process. At the end of the conformational rearrangement, the postfusion structure (Fig. 1) shows that the transmembrane region and the FPs are located in the same membrane. Completing the rearrangement before inserting FPs into the host membrane leads to a nonfusogenic, dead-end state, where the FPs are stuck at the viral membrane (6, 7). Thus, the conformational rearrangement must take place in at least two stages: a membrane insertion stage, followed by a final rearrangement, where the final rearrangement is sufficiently delayed to allow for host membrane insertion. How the kinetics of the HA rearrangement are controlled remains unclear, but likely involves differential kinetic barriers and stabilities of the various prefusion to postfusion refolding substeps. HA_2 divides into at least five subdomains, S1–S5, that undergo distinct structural rearrangements during the refolding (Fig. 1). An attractive model, first described in terms of a "spring-loaded" model (5, 8), suggests that S1 (A helix) and S2 (B loop) refold at an early stage, with the resulting coiled coil extending the FPs away from HA_2 and the viral membrane and toward the host membrane. The C-terminal portion of the protein (S4 and S5) then refolds at a later stage. Although direct observation of the HA rearrangement is complicated, this general model of an extended HA_2 intermediate is supported by several studies (5, 9, 10).

It is now well recognized that, similarly to HA, some protein sequences can fold into multiple folded structures (11–13). While the global folds of these proteins can stably switch based on

Significance

Influenza hemagglutinin (HA) is the viral membrane protein that guides entry of flu viruses into host cells. A global refolding of the stem domain of HA, HA_2, is crucial to successful viral membrane fusion, but the molecular mechanism relating this conformational change of HA_2 to its fusogenic properties remains unclear. We use molecular dynamics simulations to quantify the thermodynamic importance of the B loop domain for driving the overall HA_2 rearrangement. We find that the B loop's loop-to-coiled-coil refolding is not strongly downhill and thus is a possible target for therapeutics. A buried hydrophilic amino acid is implicated in destabilizing the coiled coil, and the sequence divergence at this position may indicate functional differences between group 1 and 2 HAs.

Author contributions: X.L., J.K.N., Q.W., J.M., and J.N.O. designed research, performed research, analyzed data, and wrote the paper.

Reviewers: B.D., Emory University; A.E.G., Los Alamos National Laboratory; and Z.A.L.-S., University of Illinois at Urbana-Champaign.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: jonuchic@rice.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1805442115/-/DCSupplemental.

Published online July 16, 2018.

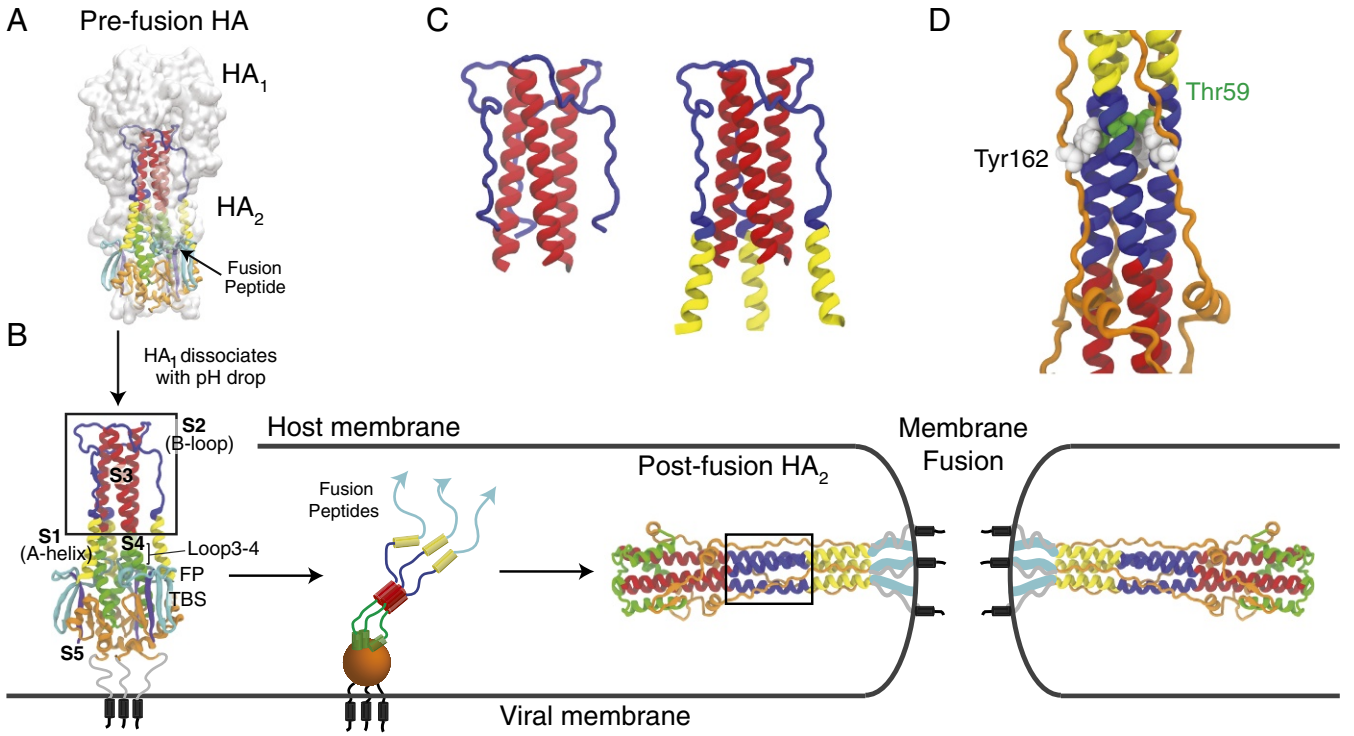

Fig. 1. The pH-induced HA transition. (*A*) The prefusion configuration of HA [Protein Data Bank (PDB) ID code 2HMG (2)] including the head domain HA_1 (white surface representation) and stem domain HA_2 (colored cartoon representation). (*B*) The reduction of pH triggers the dissociation of HA_1 from HA_2, which starts an extensive structural rearrangement ending in the postfusion configuration [PDB ID code 1QU1 (4)]. Successful membrane fusion presumably places the FPs and transmembrane helices in the same membrane. Note that HA_1 remains localized near HA_2 because of a disulfide bond between regions S5 and TBS (two β-strands). The B-loop domain (blue) changes from a random loop to a coiled-coil structure (highlighted by boxes). The central stalk, S3, maintains its coiled-coil structure during the HA_2 transition. The A-helix domain (yellow) shifts >100 Å from the preconfiguration to postconfiguration. Little is known about intermediate configurations during this transition. (*C*) The simulation constructs: the entire B loop and S3 (*Left*); and A helix, B loop, and S3 (*Right*). (*D*) The long coiled coil in the postfusion structure is wrapped by S5 (orange domain), which may provide additional stabilization to the coiled-coil configuration. In the crystal structure, Tyr162 is the only residue in S5 that buries deeply into the B loop. A buried water is coordinated between Thr59 and Tyr162.

solvent conditions or binding partners, the various subdomains or foldons (14) of these proteins are likely to be frustrated on their own, preferring one local fold only marginally more than the alternative. The postfusion B loop has a coiled-coil structure, but in the context of the crystal, it is wrapped by loops coming from S5 (Fig. 1*D*). Here, our central question is whether the B loop is able to quickly and stably refold from its random loop structure attached to S3 in the prefusion structure to its trimeric coiled-coil postfusion structure in the absence of the S5 interactions. We study the B loop's ability to act as a structural switch with detailed molecular simulations. Some indications for the pH sensitivity of the B loop itself have been seen in vitro by using a 36-residue B-loop construct, showing monomers with little helical structure at pH 7 and trimers with significant helical content at pH 4.8 (5).

The energy landscape theory of protein folding (15–17) describes folding as a diffusive process across a high-dimension energy landscape, where the competition between configurational entropy and energetics can lead to free-energy barriers to folding. Fast refolding of the B loop implies a refolding barrier small enough that the B loop can reconfigure much faster than the time scale of the full HA_2 transition. Additionally, involvement of the B loop with the removal of the buried FPs would imply a large free-energy bias toward the postfusion coiled-coil structure. By studying the folding landscape of the B loop, the ability of the B loop to shape the overall HA refolding process can be tested. The postfusion B loop has a coiled-coil fold (18, 19), a motif where two or more α-helices associate together with complimentary amphipathic sequence patterns. The simplest description of the folding of a coiled coil is as a hydrophobic "zipper," where each turn of the coiled coil quickly nucleates the folding of the next turn because of the low effective loop length (20). Thus, the folding speed for stable coiled coils can be as fast as several microseconds (21, 22). Real sequences though, with deviations from ideal hydrophobic packing, can show more complicated folding kinetics. For example, studies of the leucine zipper protein GCN4 revealed a double-exponential fit to the unfolding rate, indicating the existence of intermediate states (23, 24). GCN4 folding, recently observed with pH-jump experiments (25), can be as fast as 30 μs when tethered. An additional complicating factor is that the symmetry of a homotrimer can lead to frustration on the folding landscape (26, 27). Even considering these factors able to slow the folding, it is very likely that the refolding time of a stable B loop is much shorter than the long time scale of fusion [tens of seconds (28)]. Thus, the question of the thermodynamic stability of the final coiled-coil fold is of greatest functional importance.

In this work, we use all-atom explicit-solvent simulations using the CHARMM36 forcefield to characterize the refolding of the B loop from the group 2 H3 HA variant. Simulations follow the B loop from the prefusion loop to the postfusion trimeric coiled coil. The pH-sensitive residues in the B loop are protonated to model a low-pH environment. The free-energy profiles derived from these simulations indicate that the B loop can rapidly and stably fold half of the coiled coil, but further folding is not favorable. The burial of Thr59 in group 2 HAs is implicated as the source of the destabilization of the N-terminal region of the coiled coil by creating a water pocket in the hydrophobic

core. In sharp contrast, group 1 HAs have a hydrophobic Met59. Kinetic unfolding simulations confirm that a Thr59Met mutant has a dehydrated hydrophobic core and is more stable, providing evidence that Thr59 is destabilizing the N-terminal half of the B-loop coiled coil. Additional long unbiased kinetic trajectories were consistent with the free-energy profiles and gave information on possible kinetic traps to B-loop refolding, consisting of nonnative β-sheet formation.

Results

The B-loop domain of influenza HA subtype H3 folds into a trimeric coiled coil in the absence of the HA_1 head domain. The B-loop sequence features a heptad repeat pattern creating amphipathic helices able to form a continuous hydrophobic core buried among the three helices (Fig. 1*B*). From the A helix (S1), the B loop (S2), and the stable central coiled coil (S3), there is a 45-residue stretch where the residues at positions "a" and "d" are hydrophobic, interrupted only by Thr59 in the B loop, and flanked by Thr87 and Thr41. Additionally, coiled coils can be stabilized by salt-bridge interactions at positions "e" and "g," of which the B loop has one, Lys62 and Glu67. The B loop has several charged residues. Including the flanking Arg54 and Arg76, there are 12 charged residues per monomer. The coiled-coil structure is known to be sensitive to pH, with little helicity and probable loss of the trimer at pH 7 and strong helicity and trimer formation at pH 4 (5), where the charge is between +4 and +5 per monomer. Secondary structure prediction algorithms predict the B loop to be helical (29), although recent experiments on B-loop monomers have suggested that the B loop is only 12% helical at pH 4.8 (30).

The formation of the B-loop coiled coil is inconsistent with the prefusion HA structure. Thus, a highly stable and downhill refolding B loop has been suggested as the force that drives the initial stages of the HA conformational change (5, 8). To test this hypothesis, we characterized the folding and free-energy landscape of an S2–S3 trimer, where the B-loop trimer is folding from a prenucleated S3 coiled coil, as it does in nature. As simulating the full HA molecule is prohibitively expensive, our choice of this subset of HA_2 deserves a brief discussion. The prefusion loop structure was stabilized by interaction with HA_1 (Fig. 1*A*), which is believed to be destabilized and dissociated from the B loop through electrostatic repulsion caused by the lowered pH in the endosome (31). Here, we assumed that any remaining interaction between HA_1 and the B loop was weak and nonspecific, and, thus, we were able to remove it from the simulation. S1 (the A helix) is a second structural feature that is missing from the simulation. S1 makes extensive contacts with the anchored portion of HA_1 and, additionally, is attached to the FPs buried within S4. In the initial stages of the HA conformational rearrangement, if B-loop refolding is to be a dominant force for driving the dissociation of S1 and the FPs, the B loop will be refolding without interaction with S1 apart from its N-terminal attachment. Thus, to focus on the B loop's early role in driving subsequent refolding, we removed S1 from the simulation. We explore the consequences of including S1 later (Microsecond-Scale Kinetic Simulations).

Coiled-Coil Formation Is Not Downhill for the B Loop. We characterized the folding of the B-loop trimer with fully atomistic molecular dynamics simulations, in which the protein was protonated as if at pH 4 (see Materials and Methods for details). The folding was monitored with two reaction coordinates. $N_{\rm hb}$ counted the number of α-helical hydrogen bonds formed in the B-loop trimer and indicated the degree of helicity. $Q_{\rm post}$ counted the number of postfusion contacts that were formed and was a good monitor of the formation of tertiary interactions existing in the crystallographic coiled-coil structure of the B loop. Fig. 2 shows the free energy as a function of these two reaction coordinates. Also shown are the centroid structures from eight

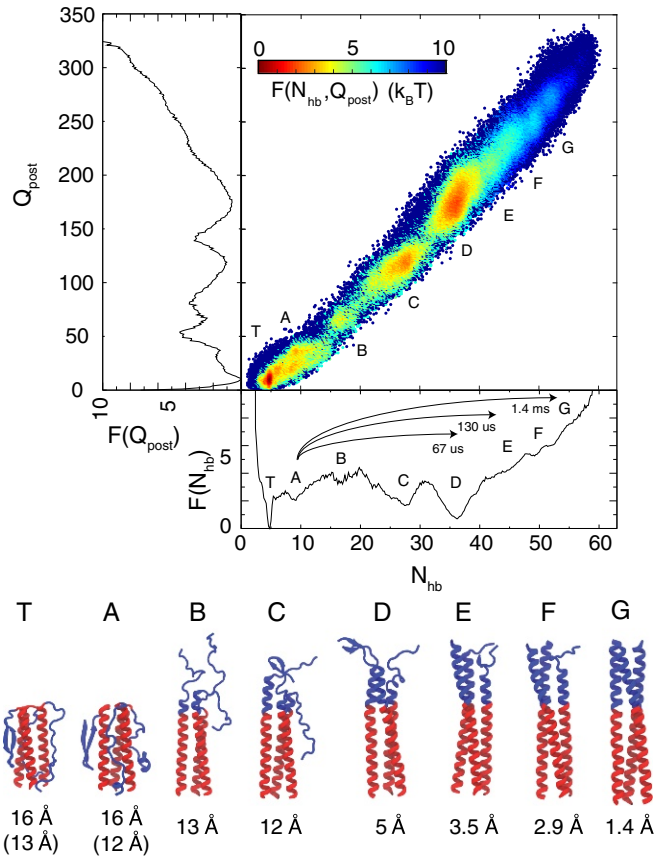

Fig. 2. Free-energy profiles of the B-loop transition and the centroid structures from eight representative $N_{\rm hb}$ ensembles at physiological temperature (310 K). (*Upper*) The two-dimensional free-energy profile as a function of $N_{\rm hb}$ and $Q_{\rm post}$, colored by free energy. The one-dimensional plots to the left and below show the projection of the free energy onto $Q_{\rm post}$ and $N_{\rm hb}$, respectively. The prefusion crystal structure has $N_{\rm hb} = 6$ and $Q_{\rm post} = 7$. The postfusion crystal structure has $N_{\rm hb} = 60$ and $Q_{\rm post} = 441$. Also indicated are the mean first passage times for the B-loop transitions from basin A to basin D, E, and G. (*Lower*) The backbone rmsd of the B loop in the representative ensembles relative to the postfusion crystal structure is shown. If the rmsd to the prefusion structure is smaller, its value is shown in parentheses.

different $N_{\rm hb}$ ensembles, a long-lived trapped structure (E_T) and structures with roughly zero to six formed α-helical turns per B-loop monomer ($E_A - E_G$). The overall folding was seen to be a zipping-like mechanism, where progressive helical structure was formed from C- to N-terminal and in lock-step with tertiary interactions.

The most striking result was that the B loop was thermodynamically most stable near E_D, where only half of the coiled-coil structure is formed, as opposed to the post-fusion-like E_G. Unbiased simulations only sampled between $E_A - E_D$ (*SI Appendix*, Fig. S3*A*) and, thus, umbrella simulations were necessary to connect the unbiased sampling to the postfusion ensemble (E_G). A simple lack of conformational sampling was not a good explanation for the absence of E_G in the unbiased simulations because the conformational entropy of $E_E - E_G$ is more restricted than E_D due to the forming coiled coil. Thus, the absence of E_G in the unbiased simulations suggested a free-energy penalty preventing sampling of E_G. This hypothesis was confirmed by performing two umbrella simulations of increasing helical bias, which connected E_D with E_E (*SI Appendix*, Fig. S3*B*) and E_E with E_G (*SI Appendix*, Fig. S3*C*), respectively.

The resulting free-energy profile was computed by using the combined data. The folding progressed from a disordered B-loop

ensemble 1–2 k_BT above E_D over a ≈ 2 k_BT barrier to E_C and then over a ≈ 2 k_BT barrier to E_D. E_G was seen to be ≈ 8 k_BT above E_D. The existence of an intermediate at E_D was corroborated by an additional set of melting simulations starting from the full postfusion coiled-coil structure of S1–S3 (*SI Appendix*, Fig. S10).

The free-energy profiles shown in Fig. 2 describe a B loop that spends most of its time in E_D. Significant time was also spent in a particular configuration E_T where each monomer of the B loop made extensive and unique contacts with S3 (*SI Appendix*, Fig. S9). In E_T, the N termini of the B loop are interacting with the C termini of S3, which is likely an artifact of introducing the artificial termini for the simulation. Nevertheless, this trap along with E_A highlights that there is residual tendency for the B loop to interact with S3. Despite the high α-helical secondary structure propensity, the B loop spent $\approx 20\%$ of the time as a monomer in β-strand structures (*SI Appendix*, Fig. S8). As can be seen in the centroid structures of E_T and E_A, these β-strand structures can associate with S3. This behavior was consistent with the frustrated and slow folding B loop observed with dual-basin structure-based models (9).

Thr59 Interrupts the Coiled-Coil Hydrophobic Core. The stability of coiled-coil structures stems from a heptad repeat pattern of buried hydrophobic residues (positions a and d) flanked by highly polar residues locking the coiled-coil configuration (18, 19). The most stable ensemble, E_D, buried the C-terminal heptad repeat pattern of Ile77–Val73–Phe70–Ile66–Phe63 until its interruption at the small polar Thr59, which was surrounded by the polar Asn60 and the charged Lys58 (Fig. 3*A*). A heptad repeat stutter (19, 32) had a 3–4–4–3–4 hydrophobic pattern instead of the canonical 3–4–3–4. There was a heptad stutter at positions 62–59, and the heptad pattern then continued from Lys58 into S1. Conspicuously, the configurations zipping up past Phe63 ($N_{hb} > 40$) were destabilized relative to E_D.

The formation of the hydrophobic core was quantified by the radial distribution function (RDF) of water molecules around residues forming the hydrophobic core of the coiled coil (Fig. 3*B*). The burial of the C-terminal hydrophobic core was quantified by the RDF of Ile66 (the rest of the RDFs can be found in *SI Appendix*, Fig. S6), which was solvated in E_A, but dry in E_D and E_G. In contrast, even in E_G, Thr59 had a peak at the radius of a water, showing that it was hydrated even when folded into the core. Note that the neighbors of Thr59, Asn60 or Lys58, were not observed to fold into the core in E_G and displaced Thr59. Seemingly, Thr59 reduced the free-energy gain of the next zipping step after E_D by maintaining water in contact with the hydrophobic residues. Put another way, there is a likely free-energy penalty in confining waters to the empty space left by Thr59 and surrounded above and below by hydrophobic residues. The presence of a polar residue inside the coiled-coil hydrophobic core is not a unique trait of the B loop. A similar cavity is seen in the GCN4 coiled coil, where two ordered water molecules cling to a buried asparagine (33).

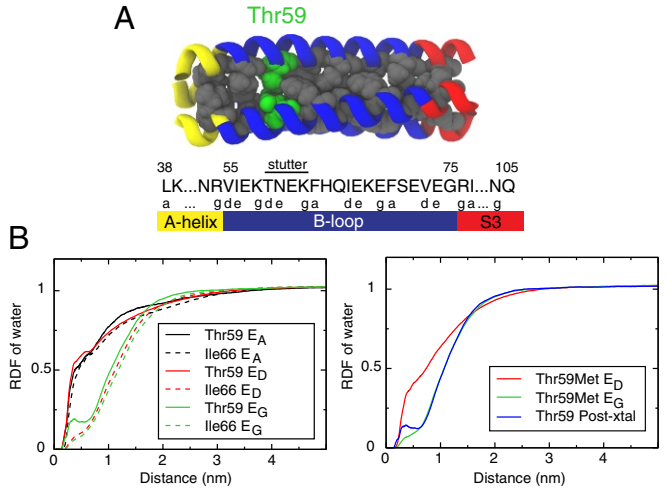

Fig. 3. The radial distribution function (RDF) of water around side-chain atoms of core residues. (*A*) The sequence and location of Thr59. (*B*, *Left*) RDFs around Thr59 and Ile66 in E_A, E_D, and E_G. The RDF of Thr59 has a peak at the radius of water, indicating solvation even in the fully coiled-coil E_G. As a comparison, Ile66 is dehydrated in states D and G. (*B*, *Right*) RDFs around Met59 in a single-residue mutant B loop at E_D and E_G. Also, as a control, a simulation starting from the postfusion coiled-coil B loop was performed, and the RDF around Thr59 was calculated.

Thr59 Is the Hallmark of Group 2 HA Molecules. Bioinformatic studies of influenza HA sequence evolution (34, 35) classify HA into two groups (36). These two groups are believed to have separated ~1,500 y ago (37). Thr59 is the only residue of the B loop strictly conserved within group 2 HA sequences, despite its disruption of the heptad repeat pattern and its apparent disruption of the hydrophobic core. Group 1 HAs have instead a conserved hydrophobic Met59. Position 59 is the strongest differentiator between the two groups (*SI Appendix*, Fig. S1).

Although a full thermodynamic characterization of a Thr59Met mutant or a group 1 B loop is beyond the scope of the current study, we performed a 130-ns constant temperature simulation on Thr59Met mutant. The mutation was introduced into the centroid structure of E_G and led to a dry core (Fig. 3*B*). As a control, to show that the time scale of the simulation was long enough to calculate the water dynamics, a simulation of the same duration showed water penetrating into the core at position 59 for wild type started from the postfusion crystal structure. Additionally, thermal unfolding simulations indicated that hydrophobic mutations at position 59 were stabilizing. When started from the postfusion structure and subjected to a temperature ramp, wild-type B loop unfolded at ~450 K, whereas Thr59Met and Thr59Val were shifted up by ~25 K (Fig. 4*A*).

As evidenced by the postfusion crystal structure, any residual disruption of the B-loop coiled coil was eventually overcome by the refolding of the S1 coiled coil and/or S5 wrapping around the B loop. Tyr162 may be important: It is positioned to coordinate a buried water in the hydrophilic pocket created by Thr59 (Fig. 1*D*). Bioinformatic analysis does not indicate any significant sequence deviations between group 1 and 2 in the S5 region interacting with the B loop. In the case of position 162, tyrosine was completely conserved for group 2, but in group 1 it was phenylalanine in 9% of sequences. In the future, a group 1 postfusion crystal structure would provide insight on any side-chain packing differences in the B-loop region between group 1 and 2 HAs.

Estimating the Kinetics of the B-Loop Transition from Free-Energy Profiles. To calculate rates from free-energy profiles, we describe the folding of the B loop in terms of diffusive motion on a free-energy landscape (see Materials and Methods for details). While one-dimensional reaction coordinates can provide intuitive descriptions of the dynamics, the full dynamics are multidimensional. In the case of protein folding, collective coordinates based on the degree of nativeness (e.g., Q_{post}) have been shown to behave diffusively and capture transition-state ensembles (38, 39). Here, for the case of B-loop folding, the high degree of similarity between free-energy profiles and diffusion coefficients calculated for two different collective coordinates gives additional evidence that they are appropriate for estimating rates. The rates calculated here for B-loop folding are likely an upper bound since the additional energetic roughness from nonspecific

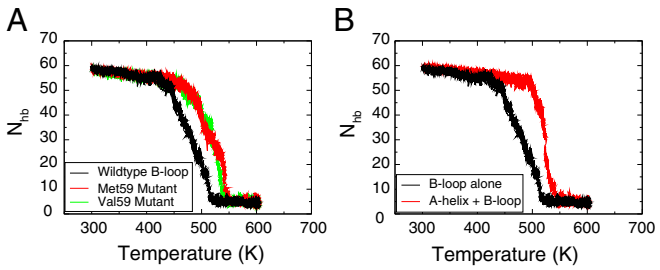

Fig. 4. Stability to thermal unfolding. N_{hb} of the B loop is used to monitor the unfolding during a 500-ns simulation with a temperature ramping from 300 K to 600 K. (*A*) Comparison of melting curves between the wild-type H3 B loop (black) with two mutants where Thr59 of the sequence was mutated to Met59 (red) and Val59 (green). The melting temperatures of these two mutants have been raised by ∼25 K. (*B*) Comparison of melting curves between the wild-type S2 + S3 and with A helix added. With the stability provided by the A helix (red), the B loop has a ∼50 K higher melting temperature than without the A helix (black). Also, the A-helix-included system melts more cooperatively, with a steeper slope downward than the S2 + S3 system. All melting curves are averages over three independent simulations. The individual curves are shown in *SI Appendix*, Fig. S10.

interactions with HA_1 or the drag of S1 should further slow the folding.

Mean first passage times obtained from our free-energy profiles are indicated in Fig. 2. It took over a millisecond for the B loop to reach a post-fusion-like coiled coil (E_G) from the prefusion ensemble (E_A). In comparison, the folding time for a fast-folding helical bundle protein is ≈10 μs (40, 41) and for a tethered leucine zipper is ≈30 μs (25). The dimeric, pH-sensitive leucine zipper was thermodynamically two-state, but showed multiphase folding kinetics on a microsecond time scale similar to our current B-loop folding landscape. For the B loop, the large free-energy penalty to fully fold was the main factor in slowing down the folding. Reaching the most stable intermediate E_D took ≈60 μs, a bit slower than a tethered leucine zipper.

Microsecond-Scale Kinetic Simulations. To connect the thermodynamic sampling data more directly to the kinetic behavior of the B-loop transition, we further performed unbiased, microsecond-scale, constant-temperature simulations (42). Three runs with S2 + S3 were performed for 1, 3, and 7.2 μs. The shorter runs formed long-lived β-structures and were terminated for this reason. The final snapshots from the 1- and 3-μs runs are shown in Fig. 5*B*.

In the 7-μs trajectory, the B loop quickly reached E_D (∼3 μs) but did not form the full coiled coil during the remaining 4 μs (Fig. 5*A*). The C-terminal half of the B loop folded into a helical bundle, followed by the complete folding of two helical monomers. However, progress stopped at this stage, as the third monomer had difficulty zipping up or docking to create the hydrophobic core. In line with the thermodynamic sampling, this kinetic run suggests that a barrier exists to forming the hydrophobic core and that the core is not seen to form around Thr59. Also consistent with the thermodynamics, the kinetic run relaxed back into E_D at 7.0 μs rather than remaining stably in E_E before continuing toward the full coiled coil as downhill folding would suggest.

In the full HA_2 molecule, the N termini of the B loop were attached to a 16-residue A helix (S1, residues 38–53; Fig. 1). Experimentally, the B loop was found to be more stable in a coiled-coil structure when the A helix was present (5). To investigate how the A helix may affect the kinetics of the B-loop transition, similarly to the S2 + S3 simulations, two runs were performed for 6 and 7 μs for S1 + S2 + S3 (Fig. 6). These simulations showed some structural similarity to the S2 + S3 simulations. Specifically, we found several intermediate structures of the B loop, where one or two monomers went almost completely helical, while the third stayed as a random loop, unable to form the hydrophobic core. While there was no indication that a free A-helix helps speed the folding of the B loop, the short time scale of the kinetic simulations makes such a determination difficult.

The most interesting feature of the kinetic simulations was the prevalence of β secondary structure in both the B loop and A helix. Both segments showed almost no helical secondary structure when surrounded by solvent and showed affinity for forming β-strands. As an isolated monomer, the B loop had on average 20% β secondary structure (*SI Appendix*, Fig. S8). Existence of the other two monomers and the A helices can stabilize these β secondary structures. Intramonomer β-hairpins were seen both between A helix and B loop (Fig. 6*A*, I1) and within the B loop alone (Fig. 5*B*). Intermonomer β-strands were also seen both between the A helix and B loop (Fig. 6*A*, I2) and within the B loop (Fig. 5*B*). β structures were also observed in the thermodynamic simulations, including the centroids of E_T, E_A, and E_D (Fig. 2). This β propensity was not observed in the bioinformatic prediction by Psipred (*SI Appendix*, Fig. S5) (43). This could be because Psipred is based on crystal structures, and the A helix and B loop have clear coiled-coil-like sequences that indeed form coiled coils in their folded end states. These simulations suggest that even these highly predicted helical sequences can form a multitude of secondary structures and nonnative trapped β structures en route to their stable folded structures.

A Helix Stabilizes the Postfusion Structure. Melting simulations starting from the postfusion structure corroborated the experimental finding that addition of the A helix stabilizes the coiled-coil structure (5). Two melting simulations were started from the postfusion structure, one with S2 + S3 and the other S1 + S2 + S3. Addition of the A helix shifted the melting temperature up by ∼50 K (Fig. 4*B*). Interestingly, the melting curves showed an intermediate that had the C-terminal half of the B loop in a coiled-coil structure that resembled the E_D structural ensemble from the thermodynamic simulations (*SI Appendix*, Fig. S10).

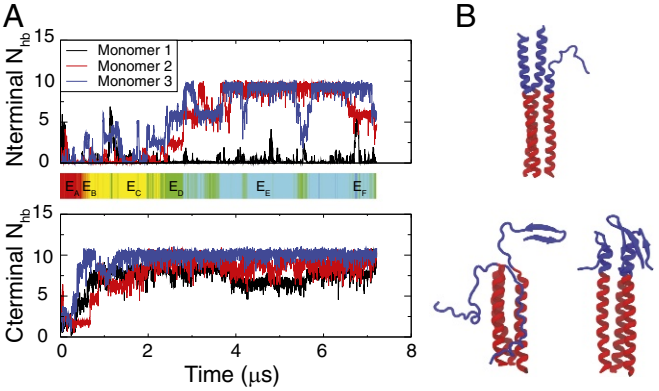

Fig. 5. Constant-temperature (T = 350 K) kinetic simulations of S2 + S3 system. (*A*) N_{hb} of each B-loop monomer changes with simulation time in one simulation trajectory. (*A*, *Upper*) N_{hb} of the N-terminal half of the B loop. (*A*, *Lower*) N_{hb} of the C-terminal half of the B loop. The C-terminal half of the B loop is almost fully folded in a stable state after 2 μs. The color bar shows the corresponding states reached by the B loop during this simulation. Color code: red, E_A; orange, E_B; yellow, E_C; green, E_D; cyan, E_E; blue, E_F. The B loop reaches E_E after 3 μs while eventually returns to E_D. (*B*, *Upper*) One snapshot from the latter half of the simulated trajectory showing a typical structure of the B loop kept from forming the hydrophobic core around Thr59. (*B*, *Lower*) Snapshots of the B loop showing nonnative β structures from the end of the two other independent runs.

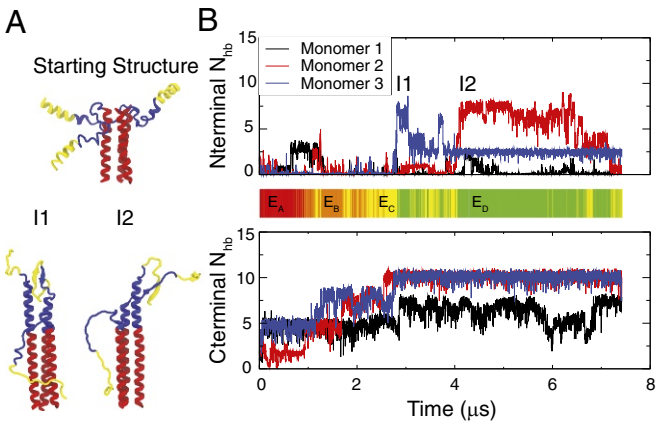

Fig. 6. Constant temperature kinetic simulations of the S1 + S2 + S3 system. The simulations show similar behavior to the simulations of S2 + S3 system. (*A*, *Upper*) Simulations start from a prefusion-like structure, where the A helix has been lifted slightly away from S3 to reduce their initial interactions. (*A*, *Lower*) Two intermediate structures, I1 and I2, taken from the simulations shows the existence of nonnative β structures formed between the B loop and the A helix. (*B*) N_{hb} of each B-loop monomer changes with simulation time. (*B*, *Upper*) N_{hb} of the N-terminal half of the B loop. (*B*, *Lower*) N_{hb} of the C-terminal half of the B loop. The color bar shows the corresponding state reached by the B loop during this simulation. Color bar code same as Fig. 5.

This provides additional evidence that the C-terminal core of the B loop is more stable than the N-terminal core.

Discussion and Conclusion

The transition of the B loop is of particular interest in the study of HA because the full coiled-coil transition of the B loop and A helix would extend the FPs of HA_2 toward the endosomal membrane. This spring-loaded process is attractive because it provides a simple explanation for one of the necessary functions of HA: facilitating the burial of FPs in the host membrane (5, 8). Direct experimental observation of the dynamical features of the B-loop transition and HA intermediates in general are difficult due to the complicated bimembrane system. Thus, molecular simulations offer another avenue to understanding the conformational transitions of HA. Here, the free-energy profile generated for the B loop calls into question the description of the B loop as a simple loaded spring. The result that the post-fusion-like ensemble E_G is significantly less stable than a half-formed coiled coil may be surprising, but it is not at odds with the full HA_2 postfusion structure being the lowest free-energy configuration. In the context of the postfusion structure, the S1 coiled coil and the contacts with S5 can provide additional stability to the B-loop coiled coil.

The simulation results we have presented are compatible with previous in vitro experiments (5, 30) that focused on a 36-residue B-loop construct. An early experiment concluded from circular dichroism (CD) that below pH 5, the B loop formed a stable trimeric coiled coil that showed 90% helicity (5). More recent experiments indicated that the helicity of the Loop36 monomer was very low and insensitive to pH (30), and, thus, suggest that forming the trimer is important for the observation of a highly helical Loop36. This is in line with our simulations of a B-loop monomer that showed 3% helicity and ∼20% β-character (*SI Appendix*, Fig. S8). Our simulations of the B-loop trimer show that an ensemble E_D, with roughly half of the B-loop coiled coil formed, is most thermodynamically stable. Analyzing the helicity for the same residues as in Loop36 (Arg54 → Ile89), which additionally contains 16 residues of S3, show E_D contains ≈80% helicity and E_E contains ≈90% helicity. Although the simulated ensemble E_D appears to be slightly less helical than the Loop36 trimer, absolute quantification of secondary structure with CD is difficult and typically involves uncertainties of 10% (44). Thus, the behavior of our simulated B loop is a viable interpretation of the available experimental data.

The sequence patterns of coiled coils are able to encode a wide range of stabilities. Interestingly, hydrophilic residues at buried positions a and d occur with low frequency [15–20% (45–47)], and mutational studies have indicated that the positions and identities of these residues can alter properties such as parallel vs. antiparallel arrangement and the degree of multimerization (48–50). Buried hydrophilic residues destabilize coiled coils relative to an ideal hydrophobic heptad pattern (47), but by doing so can promote structural specificity through encouraging the hydrophilic residues to localize together. Thus, their placement can be important for protein regions undergoing functional rearrangements. The group 2 H3 HA studied here contains one buried hydrophilic residue in its B loop, Thr59 at heptad position d. There is a stutter in the B loop of HA, and its position is conserved in most class I (α-helical) and class III (mixed α/β) viral fusion proteins (51). The functional significance of heptad repeat stutters remains unclear, but it has been shown that removal of the stutter in the B loop leads to supercoiling by one third turn in the B loop (52). The sequence Thr59–Asn60–Glu61–Lys62 in the B loop contains the two abnormalities, the buried hydrophilic Thr59 and the heptad stutter.

It is important to consider that our results are for H3, a group 2 HA, and that the sequence differences in other HA subtypes may lead to different B-loop behavior. The conservation of position 59, which strictly contains Thr59 in group 2 HA and Met59 in group 1 HA, suggests functional importance. We showed that Thr59 disrupts the hydrophobic core of the H3 B loop and that mutation to hydrophobic residues, Thr59Met or Thr59Val, created a dry hydrophobic core (Fig. 3*B*) and stabilized the B loop to thermal unfolding (Fig. 4*A*). It remains to be investigated, with a methionine at position 59, whether group 1 HAs have a more stable B-loop postfusion coiled coil and behave more closely to spring-loaded refolding or a B loop with a comparable stability due to destabilization at other positions.

In conclusion, free-energy profiles of the loop-to-coiled-coil transition describe a B loop that stably folds half of the postfusion coiled coil in tens of microseconds, but the full coiled coil is unstable. These results are compatible with previous experiments on the B loop, and the time scale is consistent with folding of a tethered leucine zipper. Thermal unfolding simulations provide indirect evidence that adding the A helix stabilizes the full B loop. Thus, it may be that the B loop is only fully stabilized by the eventual docking of the A-helix trimer or the final packing of S5. The burial of an N-terminal threonine triplet is shown to be the likely cause of the unstable full B-loop coiled-coil structure, and, thus, the sequence divergence at this position may signal a difference in the behavior of group 1 and 2 HAs, at least in the B-loop transition. Since the B-loop behavior can be influenced by the whole B-loop sequence, B-loop studies of several HA2 subtypes will be enlightening. Time-resolved pH-jump experiments on membrane-bound proteins show great promise for capturing these conformational changes in HA (53). A slower transition of the B loop and a long-lived intermediate may provide functional benefits by providing additional free-energy drive for membrane fusion and contributing extra time to the association among multiple HAs (54). We can speculate that this intermediate state is followed by the final rearrangement of S4 and S5, providing sufficient free energy to stabilize the final structure and possibly bringing the viral and cell membranes together.

Materials and Methods

The behavior of the B loop is the focus of this work, and the simulations were performed with truncated subdomains of HA_2. We were specifically

Table 1. Summary of all simulations in this paper

Set	Mutant	System	Length (μs)	Temperature, K	MD simulation type
1	WT*	S2† + S3‡	23.0	300–600	CST
1	WT	S2 + S3	8.0	300–550	CST + Umbrella $N_{hb} = 60$
1	WT	S2 + S3	7.0	300–550	CST + Umbrella $N_{hb} = 55$
2	WT	S2 + S3	1.0	350	Unbiased
2	WT	S2 + S3	3.0	350	Unbiased
2	WT	S2 + S3	7.2	350	Unbiased
3	WT	S1§ + S2 + S3	6.0	350	Unbiased
3	WT	S1 + S2 + S3	7.5	350	Unbiased
4	WT	S2 + S3	0.5	300 → 600	Thermal unfolding
4	WT	S1 + S2 + S3	0.5	300 → 600	Thermal unfolding
4	Thr59Met	S2 + S3	0.5	300 → 600	Thermal unfolding
4	Thr59Val	S2 + S3	0.5	300 → 600	Thermal unfolding

*Group 2 H3 HA, sequence from ref. 2.
†B loop (residues 54–75 of HA_2).
‡Unchanged region between prefusion and postfusion HA structures (residues 76–105 of HA_2).
§A helix (residue 38–53 of HA_2).

exploring to what extent the structural switch of the B loop can drive the overall conformational rearrangement of HA_2. Four sets of simulations were performed (Table 1). The first set of simulations used an S2 + S3 construct with a combination of unbiased and umbrella-biased simulations that used continuous simulated tempering [CST (55, 56)] for an increased rate of thermodynamic convergence. This first set was used to generate the free-energy profile (Fig. 2) and consumed the majority of the computing cost. The second set of simulations was unbiased constant temperature kinetic simulations on the same S2 + S3 system started from a prefusion-like configurating. The third set of simulations was the same as the second set, except that the A helix was included (S1 + S2 + S3). The fourth set of simulations was melting simulations used to roughly determine relative stability between different B-loop constructs.

We first discuss parameters common to all of the simulations. The sequence of group 2 H3 HA was used for simulation because it is the only sequence for which both prefusion (2) and postfusion (4) B-loop structures exist (sequence: LKSTQAAIDQINGKLNRVIEKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALENQ). All simulations were performed with the CHARMM36 force field (57) and TIP3P water, which gives a balanced treatment of protein secondary structures (58). The N termini were capped by ACE (acetyl group) and C termini by NME (methylamine). The simulations used a time step of 2.0 fs, were neutralized by 24 Cl^- ions and contained an ionic concentration of NaCl similar to endosomes of 0.15 M. The simulations were performed in the NVT ensemble because of the CST enhanced sampling protocol.

An important consideration for the HA_2 rearrangement is its pH sensitivity. Several regions of HA are suggested to respond to lowered pH, including dissociation of HA_1 (59), destabilization of the fusion peptide burial pocket (54, 60), and stability of the B-loop trimeric coiled coil (5). Interestingly, isolated HA_2 folds into the postfusion structure at both neutral and low pH (61), suggesting a limited role for pH in HA_2. Nonetheless, since there is experimental evidence for pH sensitivity of the isolated B-loop fragment considered here [most stable below pH 5 (5)], we modeled the effect of pH in the simulations. To create the most stable environment for the B-loop coiled coil, we protonated the titratable residues of the B loop as though at pH 4. This had the beneficial side-effect of simplifying the computation because this pH is low enough that all titratable residues will be protonated with high probability for the entire simulation. Although methods for simulating a system in a constant-pH environment have been developed, both for coarse-grained systems (62, 63) and explicit solvent simulations (64), these methods for explicit solvent simulations still prove to be too computationally expensive for the size and time scales of interest for the B-loop transition. There were a total of three Asp residues (zero in the B loop), 10 Glu residues (six in the B loop), and one His residue (one in the B loop) in each monomer of the simulation system. Calculation of the pK_a values for titratable residues was done with PROPKA (65) for the prefusion and postfusion structures of the isolated S2 + S3 construct (*SI Appendix*, Table S2). The results show that any residues with a $pK_a < 4.0$ were located between the C-terminal and Glu74, while in the B loop, residues should be protonated at pH 4.0. We additionally protonated even the few residues with $pK_a < 4.0$ in S3, because we reasoned that these negative charges would lead to trapping of the now positively charged B loop near prefusion-like configurations. Thus, this change to the system should only stabilize the postfusion structure.

The pK_a values of titratable residues in the postfusion B-loop structure shared the same pattern as the prefusion structure, except for one residue, Glu67, which had a $pK_a < 4.0$ in the final state due to the formation of a salt-bridge interaction with Lys62. Glu67 is located in the middle of the B loop, so the recovery of its charge state could either stabilize the basin D or G in our free-energy plot (Fig. 2). Because it is intractable to change the protonation state of this residue during the simulation, we performed a free-energy perturbation (FEP) calculation (66) after the simulation was completed. Our FEP analysis (*SI Appendix*, Fig. S4) suggested that adding a negative charge to Glu67 did not change the free energy of basin D relative to basin G, but instead would likely stabilize the prefusion state even further. Therefore, the protonation of this residue does not affect the main conclusion from our thermodynamic sampling. In sum, our protonation scheme, while below the range of pHs probed by experiments, is set up in favor of stabilizing the postfusion B-loop structure. Fig. 2 demonstrates that even in this situation, the B loop is unstable by itself in the postfusion state.

CST Simulations. Simulation set 1 used CST (55, 56) as a method to accelerate the sampling of protein conformations. CST continuously changes the temperature of a single simulation copy while maintaining a generalized canonical ensemble. This method has been used to study the folding of proteins (56). A brief description of the method can be found in *SI Appendix*. The simulated construct is very similar to the Loop36 protein studied experimentally (5), but was extended to include the entire S3. The simulated system (truncated system of S2 + S3; Fig. 1*C*) was placed in a dodecahedral box of length 102 Å with 74,587 atoms. To fix the orientation of the molecule in the box, the C_α atoms of S3 were restrained to their crystallographic positions by using isotropic harmonic position restraints. In this way, we were only investigating the conditional thermodynamic landscape of the B loop in the presence of S3 near its crystallographic position. Since the S3 coiled coil is conserved during the overall rearrangement of HA_2 (the preconfigurations and postconfigurations have a C_α rmsd < 1.3 Å), restraining S3 should not affect the results. To reduce the possibility for the N termini of S2 to interact with the C termini of S3, an unphysical configuration in a complete HA, a 1-nm radius excluded volume interaction was applied between the atoms of the terminal residues. The simulations were performed with an in-house version of Gromacs (Version 4.6.3) (67, 68) edited to include CST and PLUMED (69) for umbrella sampling.

During 23 μs of unbiased CST sampling, the postfusion ensemble was not reached, so additional umbrella sampling based on the number of α-helical hydrogen bonds in the B loop, N_{hb}, was used to ensure that configurations resembling the postfusion state were sampled. The umbrella potential was

$$V_{\text{umbrella}} = \frac{1}{2}k(N_{hb} - N_{hb}^0)^2, \quad [1]$$

where N_{hb}^0 was either 55 or 60 and $k = 0.09$ kJ/mol. N_{hb} is defined in Eq. **2**. The umbrella simulations were started from a basin D structure, and only the portion of the trajectory after the protein left basin D for the first time was used in the free-energy calculation. This should have eliminated any

bias on the starting structure (70). The unbiased sampling (i.e., $k = 0$) and the umbrella sampling were combined with the weighted histogram analysis method (71). The reweighted free-energy profiles were consistent and explained why umbrella sampling was necessary, because the postfusion ensemble was less stable by ~8 k_BT, and thus unlikely to be sampled.

Kinetic Simulations. In simulation set 2, we performed conventional NPT simulations for the same system (system S2 + S3) as set 1. The simulations were initialized from the prefusion configuration of HA_2. These simulations were performed on the Anton1 supercomputer (42) with a rectangular box of 114 × 99 × 99 Å containing 110,141 atoms. Since we found in the CST simulations that the free-energy landscape of the B-loop transition at 350 K was similar to that at 310 K (*SI Appendix*, Fig. S2*A*), we ran the kinetic simulations at 350 K for the enhanced diffusion rate. Harmonic position restraints were applied to the C_α atom of residue 80 in S3 to maintain the orientation of the simulated system.

In simulation set 3, the A helix was added (Fig. 1*C*). These simulations were designed to explore possible interactions between the A helix and the B loop during the conformational rearrangement. The 6-μs simulation started from the prefusion configuration, and the 7-μs simulation started with the A helix slightly extended and making no contacts with S3. As above, these simulations were performed on the Anton1 supercomputer (42) with a rectangular box of 119 × 119 × 138 Å (a total of 191,717 atoms). Again, residue 80 of S3 was restrained.

Melting Simulations. In simulation set 4, we used thermal unfolding trajectories to estimate the relative stability of various B-loop configurations. Each simulation was run for 500 ns and repeated three times. The setup of the simulation system was the same as in set 1.

Order Parameters for Free-Energy Calculation and Umbrella Sampling. Two order parameters were used throughout our analysis. N_{hb} counts the number of hydrogen bonds formed in the B-loop helices, and Q_{post} measures the similarity of the B loop to the postfusion state. Since we wanted to avoid any precise structural bias from the existing postfusion crystal structure in our thermodynamic simulations, we used N_{hb} for umbrella sampling. In the postsimulation analyses, we used both N_{hb} and Q_{post}. The equation for N_{hb} was taken from PLUMED (69): $N_{hb} = \sum b_{i,i+4}$, where

$$b_{ij}(t) = \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m}, \quad [2]$$

and $n = 8$, $m = 12$, and $r_0 = 2.5$ Å (72, 73). i and $i + 4$ enumerated the hydrogen and oxygen atoms forming the backbone hydrogen bonds of an α-helix. r_0 defined the length of a well-formed hydrogen bond. This choice for n and m made b_{ij} close to a step function and differentiable. Q_{post} quantified the similarity of the B loop to the postfusion crystal structure by counting the number of native atomic contacts formed in a protein configuration. The native contacts were defined as those atom pairs that were close together in the postfusion structure of the B loop, as calculated by the Shadow contact map (74). Small asymmetries in the homotrimeric contact map created by small asymmetries in the crystal structure were removed. In total, 441 native atomic contacts were obtained.

First Passage Times and Diffusion Coefficients. The kinetics of the B-loop transition can be estimated from a free-energy profile along a collective coordinate by using the following formula (38):

$$\frac{1}{k} = \langle \tau \rangle = \int_{N_{hb}^{pre}}^{N_{hb}^{post}} dN_{hb} \int_0^{N_{hb}} dN'_{hb} \frac{\exp[(F(N_{hb}) - F(N'_{hb}))/k_BT]}{D(N_{hb})}, \quad [3]$$

where N_{hb}^{pre} and N_{hb}^{post} refer to the number of α-helical hydrogen bonds at the starting and ending states of the B-loop transition. $F(N_{hb})$ is the free energy and $D(N_{hb})$ is the diffusion coefficient.

With CST, the kinetic information of the protein dynamics at a specific temperature is lost, due to the changing simulation temperatures. Thus, to estimate $D(N_{hb})$, we performed additional constant temperature simulations at each basin identified from the free-energy profile. Three trajectories of 300 ns each were computed for each basin. The three initial structures for each basin's simulations were determined by clustering the CST trajectories in each basin and taking the central structure from the top three clusters in each basin. The mean square deviation of the N_{hb} value scaled approximately linearly with time when τ is large, which is one indicator of diffusive behavior (*Appendix*, Fig. S7). The following formula relates the mean squared deviation with the diffusion coefficient,

$$D(N_{hb}) = \lim_{\Delta\tau \to \infty} \frac{1}{2} \frac{\partial}{\partial \tau} \langle \Delta N_{hb}(\Delta\tau)^2 \rangle, \quad [4]$$

and in practice means determining the slope of a linear fit to $\langle \Delta N_{hb}(\Delta\tau)^2 \rangle$ vs. $\Delta\tau$ at sufficiently large $\Delta\tau$. The calculated diffusion coefficients are listed in *SI Appendix*, Table S1. There was a large variation in the three computed diffusion coefficients for each basin, and, thus, it did not seem meaningful to use these coefficients in a basin-dependent manner in Eq. **3**. Thus, a constant diffusion coefficient for the whole profile was estimated by averaging $\Delta N_{hb}(\Delta\tau)$ at each $\Delta\tau$ over all of the 300-ns simulations. A constant value of 16.5 $N_{hb}^2/\mu s$ was used in Eq. **3** to determine the first passage times in Fig. 2.

Folding rates are dependent on temperature through changes in either the free-energy profile or the diffusion coefficient. We also calculated the free-energy and diffusion coefficients at 350 K. The free-energy profile at higher temperature was very similar (*SI Appendix*, Fig. S2, *Left*), while the diffusion coefficients were larger (*SI Appendix*, Table S1).

Analysis of HA Sequences. HA sequences were retrieved from the Influenza Research Database (75). We used all HA sequences up to now (a total of 103,091 sequences including partial sequences). Multiple sequence alignment was conducted by using the program MAFFT (76). To classify HA into two groups according to their subtypes, we used results from previous studies (77). Hellinger distance was used to capture the difference in sequence position probability distributions between group 1 and 2 HAs:

$$H_j(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{20} (\sqrt{p_i} - \sqrt{q_i})^2}, \quad [5]$$

where $P(p_1, p_2, \ldots p_{20})$ and $Q(q_1, q_2, \ldots q_{20})$ are the probability of each of the 20 amino acids in group 1 and 2 HAs. To compare the distribution of residues of the B loop in these two groups, only positions j with $1\text{-}H_j^2(P, Q) < 0.5$ were selected for visualization (Fig. S1).

ACKNOWLEDGMENTS. This work was supported by the Center for Theoretical Biological Physics sponsored by NSF Grant PHY-1427654 and also by NSF Division of Chemistry Grant 1614101 and Welch Foundation Grant C-1792. J.K.N. is a Humboldt Postdoctoral Fellow. X.L. was partially supported by National Institute of Health Grant R01-GM110310 at the early states of this work. J.M. is supported by National Institutes of Health Grants R01-GM067801, R01-GM116280, and R01-GM127628, National Science Foundation Grant MCB-0818353, and Welch Foundation Grant Q-1512. Q.W. is supported by the National Institutes of Health Grants R01-AI067839, R01-GM116280, and R01-GM127628, and Welch Foundation Grant Q-1826. Anton computer time (Grant PSCA14037P) was provided by the Pittsburgh Supercomputing Center (PSC) through National Institutes of Health Grant R01-GM116961. The Anton machine at PSC was generously made available by D. E. Shaw Research. This work used the Extreme Science and Engineering Discovery Environment through allocation TG-MCB140274, which is supported by National Science Foundation Grant ACI-1548562.

1. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin. *Annu Rev Biochem* 69:531–569.
2. Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* 289:366–373.
3. Bullough PA, Hughson FM, Skehel JJ, Wiley DC (1994) Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* 371:37–43.
4. Chen J, Skehel JJ, Wiley DC (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc Natl Acad Sci USA* 96:8967–8972.
5. Carr CM, Kim PS (1993) A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* 73:823–832.
6. Weber T, et al. (1994) Evidence for H(+)-induced insertion of influenza hemagglutinin HA2 N-terminal segment into viral membrane. *J Biol Chem* 269:18353–18358.
7. Wharton SA, et al. (1995) Electron microscopy of antibody complexes of influenza virus haemagglutinin in the fusion pH conformation. *EMBO J* 14:240–246.
8. Carr CM, Chaudhry C, Kim PS (1997) Influenza hemagglutinin is spring-loaded by a metastable native conformation. *Proc Natl Acad Sci USA* 94:14306–14313.
9. Lin X, et al. (2014) Order and disorder control the functional rearrangement of influenza hemagglutinin. *Proc Natl Acad Sci USA* 111:12049–12054.
10. Ivanovic T, Harrison SC (2015) Distinct functional determinants of influenza hemagglutinin-mediated membrane fusion. *eLife* 4:e11009.
11. Murzin AG (2008) BIOCHEMISTRY: Metamorphic proteins. *Science* 320:1725–1726.

12. Ramírez-Sarmiento CA, Noel JK, Valenzuela SL, Artsimovitch I (2015) Interdomain contacts control native state switching of RfaH on a dual-funneled landscape. *PLoS Comput Biol* 11:e1004379.
13. Giri Rao VVH, Desikan R, Ayappa KG, Gosavi S (2016) Capturing the membrane-triggered conformational transition of an α-helical pore-forming toxin. *J Phys Chem B* 120:12064–12078.
14. Panchenko AR, Luthey-Schulten Z, Wolynes PG (1996) Foldons, protein structural modules, and exons. *Proc Natl Acad Sci USA* 93:2008–2013.
15. Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci USA* 84:7524–7528.
16. Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci USA* 89:8721–8725.
17. Whitford PC, Sanbonmatsu KY, Onuchic JN (2012) Biomolecular dynamics: Order—disorder transitions and energy landscapes. *Rep Prog Phys* 75:076601.
18. Lupas AN, Gruber M (2005) The structure of α-helical coiled coils. *Advances in Protein Chemistry* (Elsevier), Vol 70, pp 37–38.
19. Mason JM, Arndt KM (2004) Coiled coil domains: Stability, specificity, and biological implications. *ChemBioChem* 5:170–176.
20. Plotkin SS, Onuchic JN (2000) Investigation of routes and funnels in protein folding by free energy functional methods. *Proc Natl Acad Sci USA* 97:6509–6514.
21. Yang WY, Gruebele M (2003) Folding at the speed limit. *Nature* 423:193–197.
22. Zhu Y, et al. (2003) Ultrafast folding of 3d: A de novo designed three-helix bundle protein. *Proc Natl Acad Sci USA* 100:15486–15491.
23. Dragan AI, Privalov PL (2002) Unfolding of a leucine zipper is not a simple two-state transition. *J Mol Biol* 321:891–908.
24. Wang T, Lau WL, DeGrado WF, Gai F (2005) T-jump infrared study of the folding mechanism of coiled-coil GCN4-p1. *Biophysical J* 89:4180–4187.
25. Donten ML, et al. (2015) pH-jump induced leucine zipper folding beyond the diffusion limit. *J Phys Chem B* 119:1425–1432.
26. Noel JK, et al. (2012) Mirror images as naturally competing conformations in protein folding. *J Phys Chem B* 116:6880–6888.
27. Levy Y, Cho SS, Shen T, Onuchic JN, Wolynes PG (2005) Symmetry and frustration in protein energy landscapes: A near degeneracy resolves the Rop dimer-folding mystery. *Proc Natl Acad Sci USA* 102:2373–2378.
28. Otterstrom JJ, et al. (2014) Relating influenza virus membrane fusion kinetics to stoichiometry of neutralizing antibodies at the single-particle level. *Proc Natl Acad Sci USA* 111:E5143–E5148.
29. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res* 41:W349–W357.
30. Casali M, Banta S, Zambonelli C, Megeed Z, Yarmush ML (2008) Site-directed mutagenesis of the hinge peptide from the hemagglutinin protein: Enhancement of the pH-responsive conformational change. *Protein Eng Des Selection* 21:395–404.
31. Huang Q, Opitz R, Knapp EW, Herrmann A (2002) Protonation and stability of the globular domain of influenza virus hemagglutinin. *Biophysical J* 82:1050–1058.
32. Hicks MR, Holberton DV, Kowalczyk C, Woolfson DN (1997) Coiled-coil assembly by peptides with non-heptad sequence motifs. *Folding Des* 2:149–158.
33. Oshaben KM, Salari R, McCaslin DR, Chong LT, Horne WS (2012) The native GCN4 leucine-zipper domain does not uniquely specify a dimeric oligomerization state. *Biochemistry* 51:9581–9591.
34. Medina RA, García-Sastre A (2011) Influenza S viruses: New research developments. *Nat Rev Microbiol* 9:590–603.
35. Burke DF, Smith DJ (2014) A recommended numbering scheme for influenza A HA subtypes. *PLoS ONE* 9:e112302.
36. Kovácová A, Ruttkay-Nedecký G, Haverlík IK, Janecek S (2002) Sequence similarities and evolutionary relationships of influenza virus A hemagglutinins. *Virus Genes* 24:57–63.
37. Suzuki Y, Nei M (2002) Origin and evolution of influenza virus hemagglutinin genes. *Mol Biol Evol* 19:501–509.
38. Socci ND, Onuchic JN, Wolynes PG (1996) Diffusive dynamics of the reaction coordinate for protein folding funnels. *J Chem Phys* 104:5860–5868.
39. Best RB, Hummer G, Eaton WA (2013) Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci USA* 110:17874–17879.
40. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334:517–520.
41. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit'. *Curr Opin Struct Biol* 14:76–88.
42. Shaw DE, et al. (2009) *Millisecond-Scale Molecular Dynamics Simulations on Anton* (ACM, New York), p 1.
43. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195–202.
44. Wiedemann C, Bellstedt P, Görlach M (2013) CAPITO—A web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics* 29:1750–1757.
45. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* 252:1162–1164.
46. Woolfson DN, Alber T (1995) Predicting oligomerization states of coiled coils. *Protein Sci* 4:1596–1607.
47. Akey DL, Malashkevich VN, Kim PS (2001) Buried polar residues in coiled-coil interfaces. *Biochemistry* 40:6352–6360.
48. Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262:1401–1407.
49. Lumb KJ, Kim PS (1995) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* 34:8642–8648.
50. Oakley MG, Kim PS (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil †. *Biochemistry* 37:12603–12610.
51. Igonet S, et al. (2011) X-ray structure of the arenavirus glycoprotein GP2 in its postfusion hairpin conformation. *Proc Natl Acad Sci USA* 108:19967–19972.
52. Higgins CD, Malashkevich VN, Almo SC, Lai JR (2014) Influence of a heptad repeat stutter on the pH-dependent conformational behavior of the central coiled-coil from influenza hemagglutinin HA2: The heptad repeat stutter in influenza HA2. *Proteins Struct Funct Bioinf* 82:2220–2228.
53. Jeong BS, Dyer RB (2017) Proton transport mechanism of M2 proton channel studied by laser-induced pH jump. *J Am Chem Soc* 139:6621–6628.
54. Lin X, Noel JK, Wang Q, Ma J, Onuchic JN (2016) Lowered pH leads to fusion peptide release and a highly dynamic intermediate of influenza hemagglutinin. *J Phys Chem B* 120:9654–9660.
55. Zhang C, Ma J (2010) Enhanced sampling and applications in protein folding in explicit solvent. *J Chem Phys* 132:244101.
56. Zhang C, Ma J (2012) Folding helical proteins in explicit solvent using dihedral-biased tempering. *Proc Natl Acad Sci USA* 109:8139–8144.
57. Best RB, et al. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J Chem Theor Comput* 8:3257–3273.
58. Best RB, Buchete NV, Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophys J* 95:L07–L09.
59. Zhou Y, Wu C, Zhao L, Huang N (2014) Exploring the early stages of the pH-induced conformational change of influenza hemagglutinin: pH sensing in influenza hemagglutinin. *Proteins Struct Funct Bioinf* 82:2412–2428.
60. Ivanovic T, Choi JL, Whelan SP, van Oijen AM, Harrison SC (2013) Influenza-virus membrane fusion by cooperative fold-back of stochastically induced hemagglutinin intermediates. *eLife* 2:e00333.
61. Chen J, et al. (1995) A soluble domain of the membrane-anchoring chain of influenza virus hemagglutinin (HA2) folds in *Escherichia coli* into the low-pH-induced conformation. *Proc Natl Acad Sci USA* 92:12205–12209.
62. Lee MS, Salsbury FR, Brooks CL (2004) Constant-pH molecular dynamics using continuous titration coordinates. *Proteins* 56:738–752.
63. O Brien EP, Brooks BR, Thirumalai D (2012) Effects of pH on proteins: Predictions for ensemble and single-molecule pulling experiments. *J Am Chem Soc* 134:979–987.
64. Goh GB, Knight JL, Brooks CL (2012) Constant pH molecular dynamics simulations of nucleic acids in explicit solvent. *J Chem Theor Comput* 8:36–46.
65. Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct Funct Bioinf* 61:704–721.
66. Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys* 22:1420–1426.
67. Zang T, Yu L, Zhang C, Ma J (2014) Parallel continuous simulated tempering and its applications in large-scale molecular simulations. *J Chem Phys* 141:044113.
68. Páll S, Abraham MJ, Kutzner C, Hess B, Lindahl E (2015) Tackling Exascale software challenges in molecular dynamics simulations with GROMACS. *Solving Software Challenges for Exascale*, eds Markidis S, Laure E (Springer International, Cham, Switzerland), Vol 8759, pp. 3–27.
69. Bonomi M, et al. (2009) Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* 180:1961–1972.
70. Rosta E, Hummer G (2015) Free energies from dynamic weighted histogram analysis using unbiased Markov state model. *J Chem Theor Comput* 11:276–285.
71. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13:1011–1021.
72. Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* 44:97–179.
73. McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol* 238:777–793.
74. Noel JK, Whitford PC, Onuchic JN (2012) The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *J Phys Chem B* 116:8692–8702.
75. Zhang Y, et al. (2017) Influenza research database: An integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res* 45:D466–D474.
76. Katoh K (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066.
77. Air GM (1981) Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A virus. *Proc Natl Acad Sci USA* 78:7639–7643.