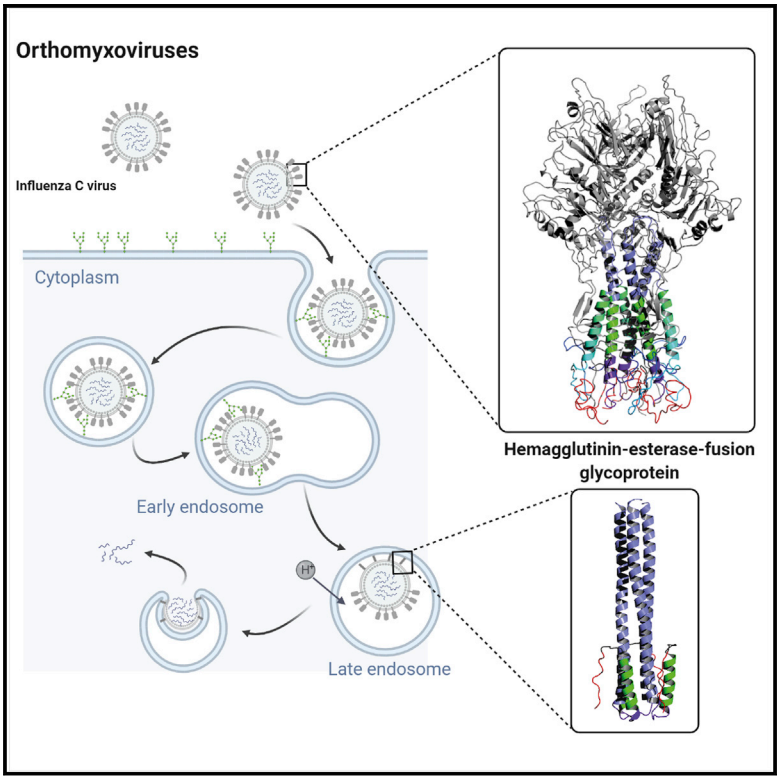
Cell Reports

Article

# Snapshot of an influenza virus glycoprotein fusion intermediate

## Graphical abstract




## Authors

Vitor Hugo B. Serrão, Jonathan D. Cook, Jeffrey E. Lee

## Correspondence

jeff.lee@utoronto.ca

## In brief

Structural snapshots of pre-and post-fusion glycoproteins have revealed the dynamic conformational rearrangements necessary for fusion, but the intermediate states have resisted capture. Serrão et al. describe a late-stage extended intermediate structure from influenza C virus. This structure extends our understanding of the mechanisms of virus cell fusion.

## Highlights

- Trapped crystal structure of a class I viral glycoprotein extended intermediate
- ICV $HEF_2$ intermediate displays a second chain reversal region at its C terminus
- Fluorescence studies reveal the role of tryptophans in forming the fusion intermediate
- The trimeric central and outer helices form a stable core early in the fusion process

Serrão et al., 2021, Cell Reports *35*, 109152
May 18, 2021 © 2021 The Author(s).
https://doi.org/10.1016/j.celrep.2021.109152

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

# Snapshot of an influenza virus glycoprotein fusion intermediate

Vitor Hugo B. Serrão,[1] Jonathan D. Cook,[1,2] and Jeffrey E. Lee[1,3,*]
[1]Department of Laboratory Medicine and Pathobiology, Temerty Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada
[2]Present address: School of Medicine, Faculty of Health Sciences, Queen's University, Kingston, ON K7L 3N6, Canada
[3]Lead contact
*Correspondence: jeff.lee@utoronto.ca
https://doi.org/10.1016/j.celrep.2021.109152

## SUMMARY

Enveloped virus entry requires the fusion of cellular and viral membranes, a process directed by their viral fusion glycoproteins. Our current knowledge of this process has been shaped by structural studies of the pre- and post-fusion conformations of these viral fusogens. These structural snapshots have revealed the start and end states necessary for fusion, but the dynamics of the intermediate conformations have remained unclear. Using the influenza C virus hemagglutinin-esterase-fusion glycoprotein as a model, we report the structural and biophysical characterization of a trapped intermediate. Crystallographic studies revealed a structural reorganization of the C terminus to create a second chain reversal region, resulting in the N and C termini being positioned in opposing directions. Intrinsic tryptophan fluorescence and bimane-induced quenching measurements suggest intermediate formation is mediated by conserved hydrophobic residues. Our study reveals a late-stage extended intermediate structural event. This work adds to our understanding of virus cell fusion.

## INTRODUCTION

The dynamic fusion of lipid membranes is essential for productive infection of mammalian cells by enveloped viruses. The mixing of lipid bilayers is kinetically disfavored, presumably due to the large repulsive force of hydration that maintains physical separation between the outer leaflets of cell and virus membranes (Cohen and Melikyan, 2004). To overcome the kinetic barrier, viruses encode a surface glycoprotein (GP) to catalyze entry into the host cell. For human viral pathogens such as influenza, HIV-1, Ebola, coronaviruses, and many others, the envelope GP mediates binding to host cell receptor(s) and fusion of the host cell and virus membrane bilayers (Podbilewicz, 2014; White et al., 2008).

The prototypical viral GP, and one of the most thoroughly characterized, is the influenza A (IAV) hemagglutinin (HA) (Skehel and Wiley, 2000; Wilson et al., 1981). HA is synthesized as a single polypeptide ($HA_0$) that is then proteolyzed to form disulfide-linked heterodimers ($HA_1$ and $HA_2$). The $HA_1$ mediates attachment of the virus to the host receptor, whereas the $HA_2$ subunit catalyzes membrane fusion. The $HA_2$ subunit contains an N-terminal hydrophobic fusion peptide and C-terminal transmembrane helix that anchors the HA to the virus membrane. Each HA decorates the surface of the virus particle as a trimer in a metastable, fusion-competent state (Carr and Kim, 1993). Influenza infection starts when the HA binds sialic acids on host cell receptors leading to receptor-mediated endocytosis. The low pH environment of the endosome triggers a concerted series of irreversible conformational rearrangements. The first crucial transition is into the extended pre-hairpin intermediate. In this intermediate, the $HA_2$ fusion subunit embeds a hydrophobic fusion peptide into the cell membrane to form a physical tether between the viral and host lipid bilayers. This intermediate is hypothesized to be short lived as $HA_2$ then folds back onto itself. The bilayer fusion proceeds through a hemifusion stalk intermediate, which allows mixing of the outer membrane leaflets. The final formation of the post-fusion, six-helix bundle (6HB) leads to the expansion of the fusion pore and transfer of the viral genetic material into the cell cytoplasm. It is thought that the refolding of HA into the 6HB conformation provides the energetics necessary to overcome the activation barrier (Baquero et al., 2013; Podbilewicz, 2014; White et al., 2008).

Structural characterization of IAV HA at neutral and low pH states have provided intimate snapshots of the start and end states of fusion and have been instrumental in defining the steps in the reaction coordinate (Bullough et al., 1994; Carr and Kim, 1993; Wilson et al., 1981). The conformational changes of IAV HA, like other class I viral GPs, have historically been thought to occur through a "spring-loaded jackknife" mechanism, where low pH triggers an irreversible conformational change to facilitate membrane merger (Carr and Kim, 1993). Recently, the spring-loaded jackknife mechanism has been reassessed using single-molecule Förster resonance energy transfer (smFRET), cryo-electron microscopy (cryo-EM), and hydrogen-deuterium exchange mass spectrometry (HDX-MS) techniques (Calder and Rosenthal, 2016; Chlanda et al., 2016; Gui et al., 2016; Lee, 2010). These studies have revealed that the mechanism of HA activation is much more than a two-state mechanism; rather,


This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

HA forms dynamic transient ensembles of intermediate states during fusion. The structural identities of transition intermediates between the pre- and post-fusion states have been poorly characterized due to their short-lived transient nature. Recently, Benton et al. discovered three early intermediate states of IAV virus HA by single-particle cryo-EM (Benton et al., 2020). The cryo-EM intermediate structures reveal the early HA transitions involved in low pH activation. Moreover, they describe the formation of the extended pre-hairpin intermediate. However, what happens after the formation of the early-stage ensemble of extended intermediates leading up to the formation of the post-fusion 6HB remains unknown.

The influenza C virus (ICV) presented us with an opportunity to capture a viral GP intermediate. ICV contains only one virally encoded GP on the surface of the virus, a hemagglutinin-esterase fusion (HEF) that has receptor binding, receptor destroying, and membrane fusion activities (Wang and Veit, 2016). The ICV HEF combines the functions of the HA and neuraminidase proteins encoded by IAV and influenza B viruses (IBVs). Like IAV and IBV HAs, ICV HEF is post-translationally cleaved by host proteases to form the attachment ($HEF_1$) and fusion ($HEF_2$) subunits. The structure of the pre-fusion ICV HEF is structurally similar to IAV and IBV HAs (Rosenthal et al., 1998; Taylor, 1949). Although, ICV is related to IAV and IBV, major differences exist in the fusion kinetics between these orthomyxoviruses (Formanowski et al., 1990). There is a lag before the onset of fusion by ICV that is not observed in fusion by IAV or IBV. Specifically, fusion activity was reported between a pH value of 5.6–6.1 with hemolysis occurring at slightly lower pH values of 5.1–5.7. Based on virus morphological changes and endogenous tryptophan fluorescence experiments, Formanowski et al. (1990) suggested that a conformational change in ICV HEF is the rate-limiting step in fusion. The precise mechanism of the ICV kinetic lag is poorly defined. It is possible that the close packing of HEF in a hexagonal array on the virus surface hinders conformational changes or that the fusion event requires the cooperation of multiple HEF molecules (Formanowski et al., 1990). Alternatively, the kinetic lag may be due to a slow buildup of a stable ICV HEF intermediate.

We set out to understand whether it is possible to trap a longer-lived ICV $HEF_2$ intermediate by carefully controlling the pH in ICV $HEF_2$. Using biophysical techniques, we detected a distinct stable conformational state in ICV $HEF_2$ at pH 6.0 or higher that was not observed in analyses of IAV or IBV HAs. This led to rationally designed crystallization experiments that resulted in the determination of a 2.4 Å resolution crystal structure of a late-stage extended ICV $HEF_2$ intermediate. This $HEF_2$ intermediate is different from the previously described early fusion intermediates and the post-fusion 6HB. The intermediate is characterized by a structurally flexible C-terminal region and a second chain reversal region. We hypothesize that this structure represent a late-stage extended intermediate conformation before the foldback to form the 6HB. Our data present a detailed atomic-level dissection of the intermediate states of viral fusion, specifically addressing the ensemble of extended intermediate conformations present leading into the final post-fusion state. This contributes now to a more complete understanding of the conformational gymnastics required for fusion.

## RESULTS

### Distinct conformational states of ICV $HEF_2$ detected as a function of pH

In order to determine whether ICV fusion intermediates can be trapped, we focused on the $HEF_2$ fusion subunit. ICV $HEF_2$ (residues 481–621) was designed based on the construct boundaries of $HA_2$ post-fusion ectodomain subunits from IAV and IBV. ICV $HEF_2$ and IAV and IBV $HA_2$ fusion subunits were heterologously expressed and purified from *E. coli* to homogeneity (Figure S1A). Although these three proteins have similar biological functions, their sequences are not conserved (Figure S1B). Initially, we performed circular dichroism (CD) analysis on ICV $HEF_2$ as a function of pH. The CD spectra of ICV $HEF_2$ clearly show the presence of two distinct, stable conformational states between pH 4.5 and 7.5 (Figure 1A; Figure S2A). At pH values higher than 5.5, ICV $HEF_2$ adopts a major $\alpha$-helical conformation, whereas at pH values below 5.5, both $\alpha$-helical and $\beta$ sheet content are present (Table S1). Previous crystallographic analyses indicate that IAV and IBV HA post-fusion 6HB states have a central coiled-coil $HA_2$ core with a small antiparallel $\beta$ sheet outer layer (Figure S3) (Bullough et al., 1994; Ni et al., 2014). Analysis of thermal denaturation curves of ICV $HEF_2$ revealed a reversible and two-state equilibrium between pH 5.5 and 7.5 (Figure 2). At 37°C and pH 6.0, the Gibbs free energy profile of ICV $HEF_2$ was clearly indicative of an energetically stable intermediate (Figure 1B).

IAV and IBV HA were not observed to undergo a lag before the onset of fusion (Formanowski et al., 1990). The secondary structural composition of IAV and IBV $HA_2$ are invariant as a function of pH, suggesting that the maintenance of the canonical post-fusion 6HB structure is maintained over the pH range tested (Table S1). The lack of any detectable stable fusion intermediates of IAV or IBV $HA_2$ is consistent with previous studies that showed IAV HA undergoes spontaneous and reversible transitions between multiple conformations (Das et al., 2018; Ivanovic and Harrison, 2015). The ICV $HEF_2$ conformation observed at pH 6.0 (Figure S2A) is consistent with functional studies that demonstrated ICV fusion activity at pH values between 5.6 and 6.1; therefore, the conformation at pH 6.0 may represent a fusion intermediate state (Formanowski et al., 1990).

### Trapping an ICV $HEF_2$ intermediate structure

Understanding the free energy profile of ICV $HEF_2$ allowed us to rationally design structural studies to trap the ICV intermediate. A slightly shorter ICV $HEF_2$ construct (residues 495–586) was required to obtain well-diffracting crystals at pH ~6.0 (Figures S1C–S1E; Table S2). This shorter ICV $HEF_{495–586}$ variant had similar CD spectral characteristics and structural pH dependency as the longer ICV $HEF_{481–621}$ construct used for biophysical studies (Figure S2A). The pH of the crystallization drop was measured at 6.1, thus consistent with a pH that can capture the primarily $\alpha$-helical conformation observed in our CD studies and with functional experiments that showed ICV fusion activity in the pH range from 5.6 to 6.1 (Formanowski et al., 1990).

The ICV $HEF_2$ structure is trimeric with each of its protomers consisting of a 15-turn N-terminal helix (residues 496–550; segments A-B-C) and a second, smaller 5-turn C-terminal $\alpha$ helix

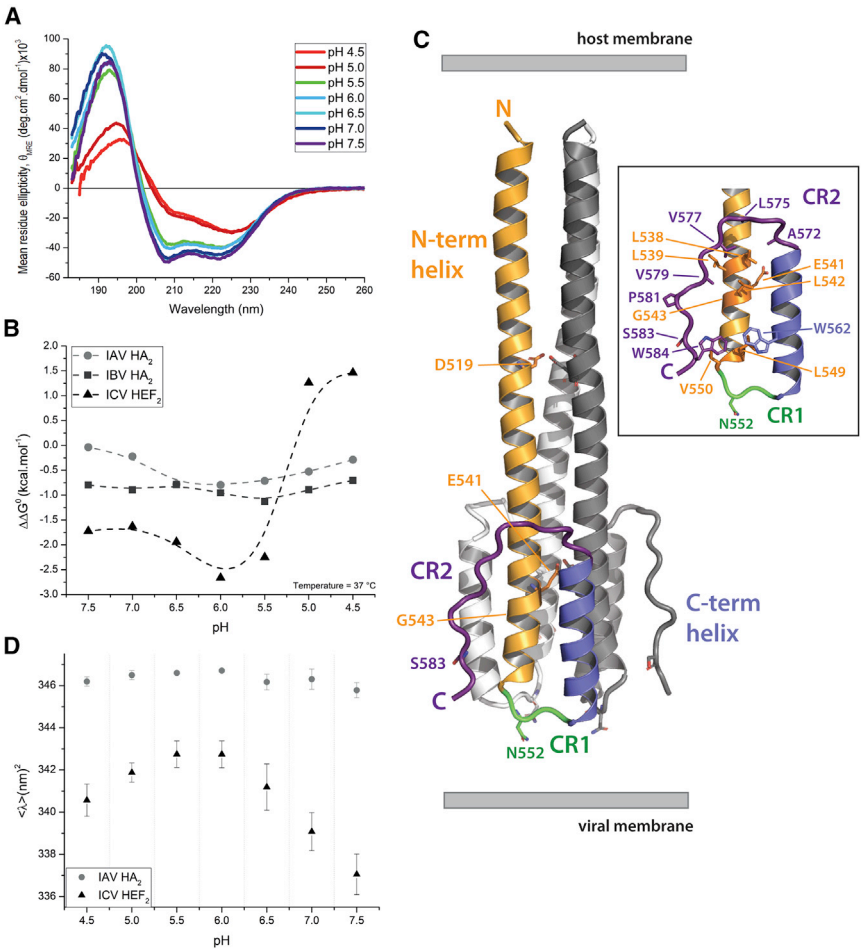



**Figure 1. ICV $HEF_2$ structural and biophysical analysis**

(A) CD spectra of ICV $HEF_2$ as a function of pH.

(B) Gibbs free energy landscape as a function of pH for ICV $HEF_2$ (black triangles), IAV $HA_2$ (gray circles), and IBV $HA_2$ (dark-gray squares).

(C) Structure of the trimeric ICV $HEF_2$ intermediate. Three protomers of the trimer are shown as multicolor, white, and gray chains. Inset shows the hydrophobic interactions made between CR2 and the central helical core. A predicted N-linked glycan site (N552) is found in CR1.

(D) Wavelength at the center of spectral mass calculated from the intrinsic tryptophan fluorescence emission spectra of ICV $HEF_2$ and IAV $HA_2$ as a function of pH. Spectroscopic and thermodynamic analysis are represented as the mean ± SEM from biological triplicates.

See also Figures S1, S2, S4, and S5 and Tables S1 and S2.

(residues 555–570; segment D), separated by a 4-residue chain reversal region (CR1) (Figures 1C and 3). The N- and C-terminal α-helical coils are generated by heptad-repeat motifs. Each central helical core is lined primarily with leucine and isoleucine residues, with the exceptions of two layers of aspartic and glutamic acid residues (Asp519 and Glu541), which form hydrophobic interactions between protomers. The outer helix packs against the central helical core in an anti-parallel orientation to form the trimeric helical bundle. The coil accommodates stabilizing salt-bridge interactions between adjacent protomers without interrupting the heptad-repeat motif. At the base of ICV $HEF_2$, the heptad-repeat stutters at Gly543, and two turns later the helix breaks to form the short CR1 region that contains Asn552, which is glycosylated in the pre-fusion structure of ICV HEF (Rosenthal et al., 1998). This positions three N-linked glycans at the base of the trimer pointing outward toward the bulk solvent.

A second chain reversal region (CR2; residues 571–585; segments E-F) is observed on the C-terminal side of the second outer helix in the ICV $HEF_2$ intermediate structure. In the post-fusion 6HB conformation, the N and C termini point in the same direction toward the host cell membrane (Figure 3). In contrast, in the ICV $HEF_2$ intermediate CR2 causes the C terminus to point toward the viral membrane, allowing a connection between the viral and host membranes. Ser583 in the CR2 is positioned on the outer $HEF_2$ surface. In the wild-type protein, this residue is a cysteine, and its location would allow the formation of the inter-subunit $HEF_1$-$HEF_2$ disulfide linkage.

## Second chain reversal region involves hydrophobic interactions and a tryptophan pair

CR2 forms an unstructured coil that packs toward a hydrophobic interface formed between its central and outer helices (segments C-D). Leu538, Leu539, and Leu542 in the central helix (segment C) form a hydrophobic pocket that interacts with Ala572, Leu575, Val577, and Val579 from the CR2 region (segment E). In addition, Trp584 located at the C terminus of CR2 (segment F) forms an edge-to-edge aromatic interaction with Trp562 (~3.5 Å apart) located on the outer helix (segment D) and hydrophobic residues Leu549 and Val550 on the central helix (segment C) (Figure 1C, inset). All residues in the hydrophobic pocket and the CR2 region, including the tryptophan residues, are strictly conserved in all ICV isolates (Figure S4).

CR2 (segment D) does not form any crystal contacts in our structure. However, the last five residues of the C terminus of ICV $HEF_2$ form 3 H-bond contacts to a crystallographic symmetry-related molecule (Figure S5). In contrast, there are 7 hydrophobic interactions formed by the CR2 region packing into the central helical core. Thus, we predict the formation of the CR2 region likely outweighs the C-terminal crystal contact interactions. Regardless, it was important to ensure that the formation of CR2 and the Trp562-Trp584 interaction are not artifacts of crystallization. The Trp562-Trp584 interaction is the only tryptophan pair in ICV $HEF_2$; thus, the motions and interaction can be specifically monitored in solution by fluorescence. Intrinsic tryptophan fluorescence analysis revealed a shift in the peak center of mass

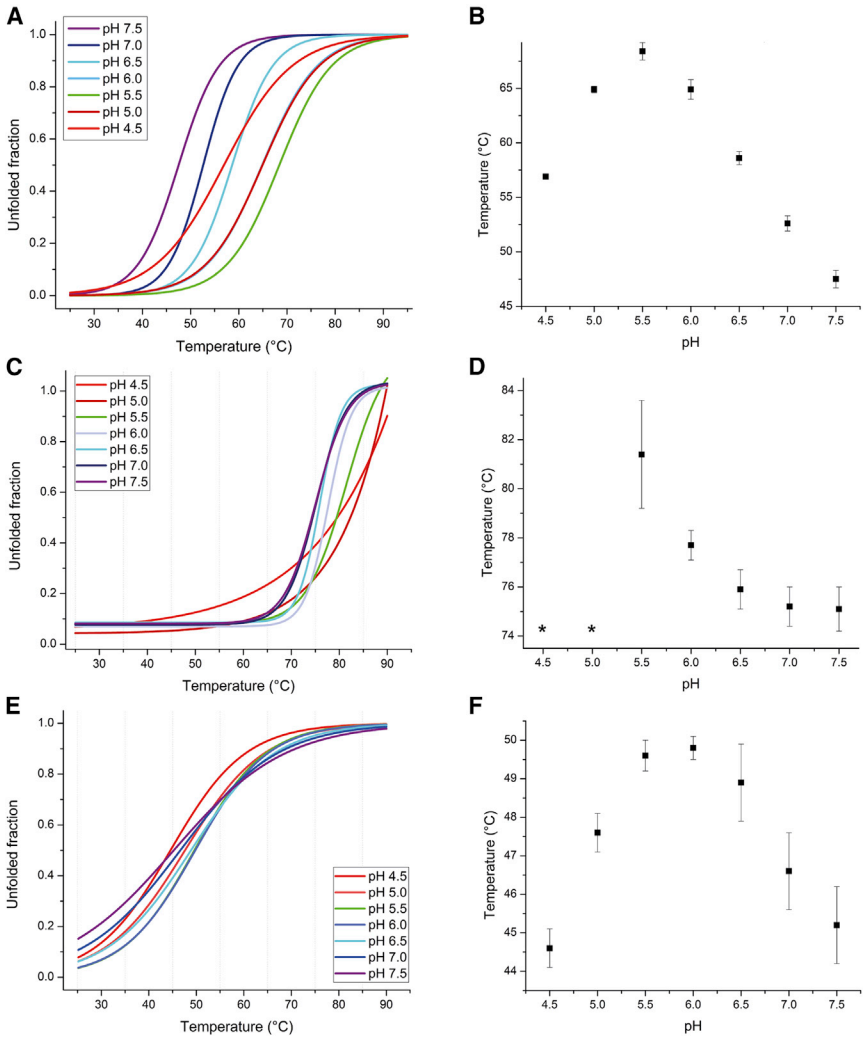



**Figure 2. Thermal melts for IAV $HA_2$, IBV $HA_2$, and ICV $HEF_2$**

Non-linear biphasic sigmoidal curve to determine the apparent melting temperature(s) ($T_m$ values) as a function of pH for (A and B) ICV $HEF_2$, (C and D) IAV $HA_2$, and (E and F) IBV $HA_2$. The asterisks indicate that $T_m$ values could not be experimentally determined. All CD spectra and thermal denaturation assays were performed in biological triplicates. See also Figure S2 and Table S1.

from the emission maximum at pH values lower than 6.0 (Figure 1D). A pH-dependent change in tryptophan fluorescence was also observed in the native infectious ICV particle (Formanowski et al., 1990). In contrast, for IAV $HA_2$, which contains one tryptophan, there was no appreciable shift in intrinsic tryptophan fluorescence as a function of pH (Figure 1D).

To specifically probe the Trp562-Trp584 interaction, we site-specifically labeled ICV $HEF_2$ with the fluorescent probe bimane and used tryptophan-induced quenching (TrIQ) to determine the specific orientation of the ICV $HEF_2$ C-terminal region as a function of pH in solution. The TrIQ approach exploits the strong quenching effect of tryptophan on the bimane fluorescence emission intensity. Only a single fluorescent probe was incorporated into the protein at specific locations to provide direct measurements of protein conformation. Further, this quenching occurs only over short distances when the $C_\alpha$-$C_\alpha$ distance between the bimane-cysteine attachment point and tryptophan quencher is between 5 and 15 Å. Single cysteine point mutations were introduced into strategic positions in both short ICV $HEF_{495–586}$ and long ICV $HEF_{481–621}$ constructs to allow for specific labeling with bimane. The cysteine mutation Q558C is adjacent to Trp562 and was designed to directly probe the formation of the Trp562-Trp584 interaction (Figure S6A). We expected increased fluorescence quenching upon CR2 formation and the approach of Trp584 toward Trp562. Based on the post-fusion IAV and IBV $HA_2$ structures, the C terminus folds back into HR1 helix at low pH. A second cysteine mutant, H518C, was introduced in the middle of the central HR1 helix to probe the foldback of the C-terminal region and Trp584 into the HR1 helix at low pH (Figure S6A). Each of the cysteine mutants behaved similarly to wild-type ICV $HEF_2$ based on size-exclusion chromatography (SEC) and CD profiles, suggesting that no structural changes resulted from the cysteine substitutions (Figures S6B and S6C).

Fluorescence measurements revealed tryptophan-induced quenching of the bimane at Q558C in ICV $HEF_{495–586}$ and ICV $HEF_{481–621}$ at pH values ranging from 5.5 to 6.0 but not at pH values above 6.0 or below 5.5 (Figure 4; Figure S7; Table S3). As expected, the H518C bimane-labeled mutants showed strong quenching only at pH values lower than 5.5, suggesting a foldback of the ICV $HEF_2$ C-terminal region and Trp584 against the central HR1 helix. The TrIQ profiles of both the short ICV $HEF_{495–586}$ and longer ICV $HEF_{481–621}$ constructs were similar (Figure 4). Taken together with the intrinsic tryptophan fluorescence data, the TrIQ studies support a conformational change that positions Trp584 in the vicinity of Trp562 when the pH is in the range of 5.5–6.0. Importantly, the short ICV $HEF_{495–586}$ construct behaves similarly to the longer but more complete ICV $HEF_{481–621}$ fusion ectodomain, which is similar in length to the post-fusion IAV and IBV $HA_2$ previously used for crystallization. Finally, this analysis confirmed that CR2, observed in the ICV $HEF_2$ crystal structure, exists in solution.

## DISCUSSION

Obtaining high-resolution structural information on viral fusion intermediates are rare. Initially, evidence for viral fusion GP

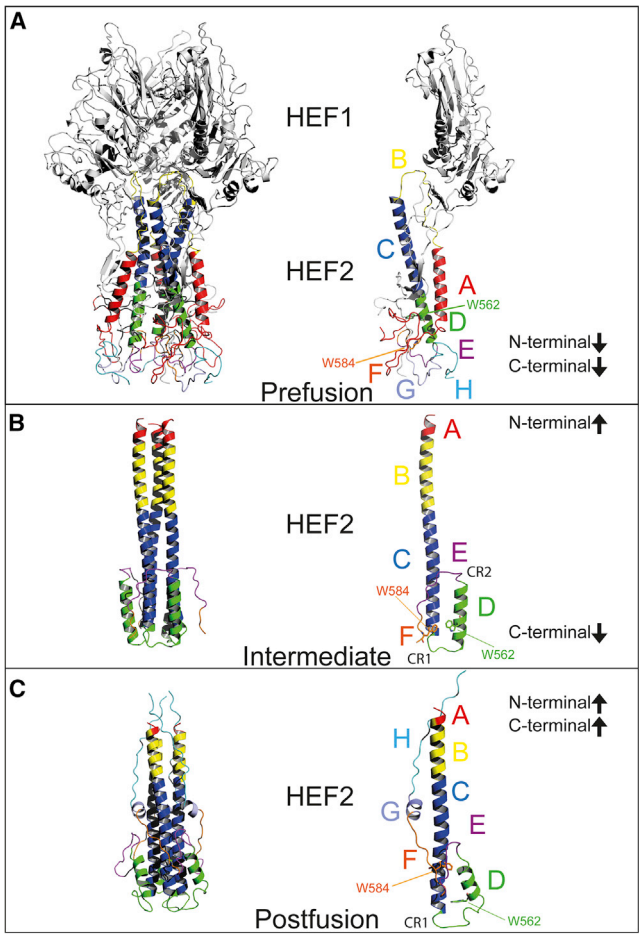



**Figure 3. Comparison of pre-fusion, extended intermediate, and post-fusion states of ICV $HEF_2$**

(A) The pre-fusion metastable state crystal structure presents parallel N and C termini pointed in the direction of the viral membrane.

(B) The ICV $HEF_2$ intermediate crystal structure reveals that segments A-B-C-D forms a core that is structurally invariant and maintained to the final post-fusion 6HB. Segments E-F form a second chain reversal region that positions the N and C termini in an antiparallel orientation, bridging viral and host membranes.

(C) Homology model of the post-fusion ICV $HEF_2$ structure reveals that at low pH foldback in segments E-F results in formation of the stable 6HB. All structures are shown as cartoons with the eight segments of $HEF_2$ that undergo conformational changes highlighted in different colors from A to H. The structures used to generate this figure were as follows: prefusion, PDB: 1FLC; intermediate, structure presented in this manuscript, PDB: 6WKO; and post-fusion, homology model generated using the ITASSER server (Roy et al., 2010) and the IBV $HA_2$ post-fusion structure as a template (PDB: 4NKJ).

See also Figure S3.

intermediates was obtained through indirect experimental observations. For example, the transiently extended intermediate in human parainfluenza virus 5 F was detected using electron microscopy based on the inter-bilayer spacing of the virion trapped in the process of fusing to the target membrane (Kim et al., 2011). The bilayer spacing is consistent with the expected distance of a computationally determined extended intermediate model, where the N-terminal region has sprung into a trimeric coiled coil toward the host membrane and the C terminus is anchored in the viral membrane. Further indirect evidence of viral fusion intermediates come from studies of HIV-1 gp41. Studies of peptides such as T-20/enfuvirtide (Wild et al., 1994) or C34 (Chan et al., 1998), which mimic the binding of the outer heptad repeat-2 (HR2) helix, indicate that the trimeric central helical core must be formed prior to the foldback of the HR2 helix to form the final post-fusion 6HB. A number of cryo-electron tomographic studies have imaged conformational changes in HA on the virus surface and captured its interaction with the membrane (Calder and Rosenthal, 2016; Chlanda et al., 2016; Fontana et al., 2012; Gui et al., 2016; Lee, 2010). Analyses of the low-resolution tomograms led to the hypothesis that the fusion peptide is released prior to the dissociation of the $HA_1$ attachment subunits. More definitive studies came recently where smFRET measurements showed that low pH activation of IAV HA pseudotyped onto lentivirus particle resulted in reversible sampling of multiple intermediate HA conformations prior to the formation of the extended intermediate (Das et al., 2018). The interaction of $HA_2$ with the host membrane then promotes the irreversible conformational change to the post-fusion state.

A major breakthrough was made when cryo-EM reconstructions captured conformational intermediates of IAV HA after low pH treatment (Benton et al., 2020). Three structural transitions were observed directly after pH triggering. After 10 s of low pH exposure, the envelope-distal domain of $HA_1$ dilates and tilts away from the central helical axis followed by subunit opening and the disordering of the membrane-proximal region (Figure 5). At 60 s, the $HA_2$ propels a 150 Å triple helix (corresponding to segments A-B-C-D) at the center of HA to form an extended pre-hairpin intermediate. This rearrangement inserts the fusion peptide into the host membrane, analogous to the action of a spear. Pulse-labeled hydrogen-deuterium exchange mass spectrometry (HDX-MS) studies further showed that intact virions upon low pH activation produced an intermediate ensemble that displayed concurrent reorganization of the $HA_1$ subunit (Benhaim et al., 2020). While these results clearly show the presence of early transient viral fusion intermediate states, there are still major questions that remain unresolved: What are the transitions that follow the early extended intermediate states to result in the final post-fusion conformation? What are the rearrangements that need to occur at the membrane-proximal region of the glycoprotein? What are the residues that mediate the formation of other fusion intermediates?

From smFRET, cryo-EM, and HDX-MS studies, the extended intermediate was shown to be an ensemble of conformations (Calder and Rosenthal, 2016; Chlanda et al., 2016; Gui et al., 2016; Lee, 2010). The ICV $HEF_2$ crystal structure presented here likely belongs to one of the extended intermediate states, as it is capable of forming a physical protein bridge between the viral and host membranes (Figure 3). This conformation captured in the current structure is likely further along the pathways toward the post-fusion state than the previously described IAV early-stage extended intermediates, which clearly displayed a central long helix composed of segments A-B-C-D (Benton et al., 2020). This is in contrast to the post-fusion state where segments D-E-F-G-H fold back onto a central trimeric helix composed of segments A-B-C to juxtapose the N and C termini

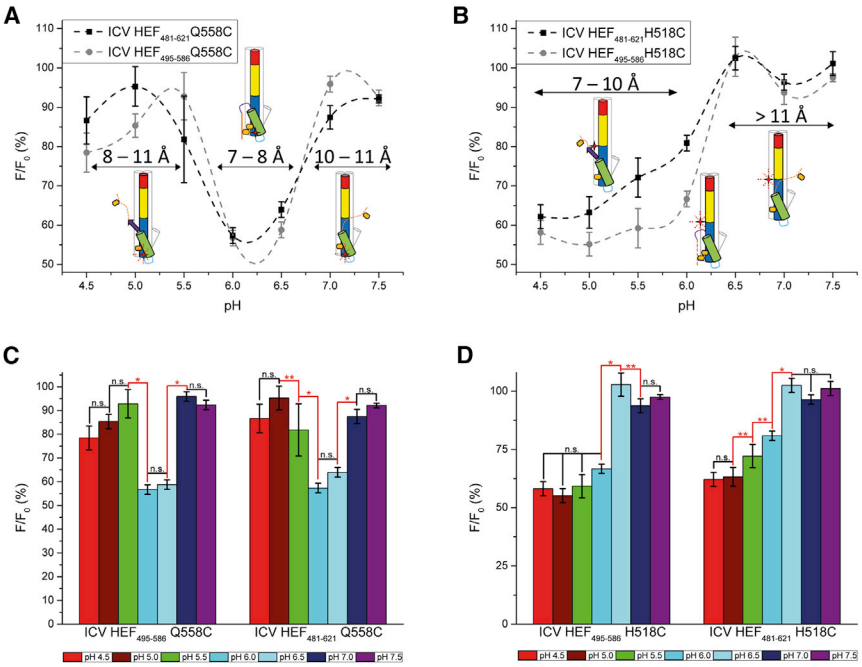



**Figure 4. Tryptophan-induced quenching**

(A and B) Site-directed bimane labeling at positions (A) Q558C and (B) H518C for ICV $HEF_{495–586}$ and ICV $HEF_{481–621}$ allowed the measurement of ICV $HEF_2$ motions as a function of pH. Tryptophan residues that come into close proximity to the bimane excited state will quench its fluorescence signal. The Q558C bimane-labeled mutants were generated to probe the interaction of W562–W584. At pH values >6.0, no quenching was observed, thus suggesting the lack of any interactions between W562 and W584. At pH 5.5–6.0, fluorescence quenching was detected, indicating the formation of the W562–W584 interaction and CR2, as observed in the crystal structure of the trapped fusion intermediate state. The H518C bimane-labeled mutants were generated to probe interactions with the HR1 helix. Fluorescence quenching was only observed at pH values lower than 5.5, consistent with the packing of W584 into the HR1 central helix, as observed in many low pH influenza structures. The inset schematic cartoons of ICV $HEF_2$ show the location of the probe (red star) and tryptophan residues (orange rectangles). While there are no structures of ICV $HEF_2$ at low pH, we hypothesize that the ICV $HEF_2$ C terminus refolds toward the central HR1 trimeric helix based on the structural changes observed at low pH in IAV HA. The Cα-Cα distances, shown in Å, are estimated from the measured ICF $HEF_2$ quenching fractions ($F/F_0$) and quenching-based distance measurements reported by Jones Brunette and Farrens, 2014. (C and D) Statistical analysis of the change in TrIQ $F/F_0$ values as a function of pH for (C) Q558C and (D) H518C ICV $HEF_{495–586}$ and ICV $HEF_{481–621}$. Data are presented as the (mean ± SEM) from biological quadruplets. One-way ANOVA followed by a Tukey post hoc comparison test was calculated. n.s. is not significant ($p > 0.05$); *$p < 0.0001$; **$p < 0.001$.

See also Figures S6 and S7 and Table S3.

and form the 6HB (Figure S3). In our ICV $HEF_2$ structure, we observed a structure that is in transition between the early extended intermediate and final post-fusion conformation. There is a foldback of segment D onto the central helix (segments A-B-C), but segments E-F have not folded back into a conformation as observed in the post-fusion 6HB structure. Instead, segments E-F form the second chain reversal region that points the $HEF_2$ C terminus back toward the viral membrane (Figure 3). Regardless of the different folds in the helical segments, both the cryo-EM IAV HA and our ICV $HEF_2$ structures likely belong to an ensemble of extended intermediate conformations.

Our data suggest the order in which conformational changes of the HA occur (Figure 5). After the formation of the extended intermediate and the long central trimeric helix (segments A-B-C-D), segment D breaks from the central trimeric helix and folds back to form a small outer helix. This creates the base of the 6HB. Therefore, our structure clearly shows that the small outer helix forms before the C-terminal segments G-H are packed antiparallel against it. Segments A-B-C-D form a structurally invariant core that is maintained in the final post-fusion conformation (Figures 3 and 5). Concomitant with the foldback of segment D to form the outer helix, segments E-F form a second chain reversal region immediately after the outer helix (segment D). The random coiled nature of the CR2 region combined with the predicted unstructured membrane-proximal region provides flexibility. After the formation of the extended intermediate, the fusion subunit needs to tilt and may break its 3-fold symmetry as it transitions to the final post-fusion 6HB conformation. The flexibility of the CR2 region provides the slack to allow the tilting of the fusion subunit. The tilting of the fusion subunit was indirectly shown to be important for orthomyxovirus membrane fusion as the antibody FISW84 binds the C-terminal membrane-proximal region of HA and appears to neutralize the virus by sterically restricting the range of HA orientations required for fusion (Benton et al., 2018).

The formation of the intermediate in ICV $HEF_2$ is likely driven by hydrophobic forces from conserved residues located on both the CR2 region (segments E-F) and the segment C-D helices. The use of hydrophobic interactions to form the late-stage intermediate is consistent with the biology of orthomyxoviruses. For example, influenza viruses are endocytosed and enter early endosomes (pH ~6), followed by subsequent trafficking to late endosomes (pH ~5) where fusion occurs (Lakadamyali et al., 2004; Matlin et al., 1981). The use of non-ionizable residues to form the intermediate would make biochemical sense, as the strength of hydrophobic and aromatic interactions will not be affected as the virus migrates through the various endosomal stages and pH changes. In other viruses that undergo fusion at low pH, such as avian sarcoma leukosis virus, hydrophobic residues rather than electrostatics are proposed to stabilize viral fusogen conformations (Aydin et al., 2013, 2014).

In summary, our results provide experimental evidence for the existence of a distinct viral fusion intermediate. The atomic-resolution model of this intermediate state has potential utility in

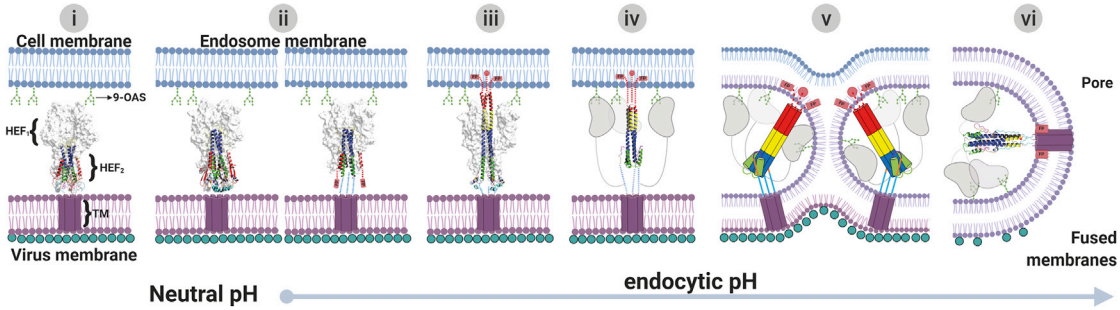



**Figure 5. Membrane fusion reaction coordinate of ICV HEF**

(i) Prefusion conformation: ICV HEF sits on the surface of the virus in a pre-fusion metastable conformation. Initial attachment of the virus to the host cell is mediated through the attachment of the pre-fusion HEF to N-acetyl-9-O-acetylneuraminic acid (9-OAS; shown as green branches) located on the host membrane. More than one ICV HEF protein is likely required at the site of fusion. The transmembrane domain (TM) is represented in purple. The HEF cytoplasmic tail interacts with the matrix protein (shown as blue circles) located inside the viral membrane. (ii) Early intermediate: after endocytosis and trafficking to the endosome, the low pH triggers large concerted conformational changes at the $HEF_1$-$HEF_2$ interface. The three membrane-distal $HEF_1$ domains rotate as a rigid body away from the central $HEF_2$ helix. Moreover, the membrane-proximal domain becomes disordered and rearranges, with the fusion peptide being released from its pre-fusion state. (iii) Extended intermediate: the constraints of $HEF_1$ (gray) on $HEF_2$ are released to allow the hydrophobic fusion peptide (FP) from $HEF_2$ to insert into the host cell lipid bilayer. The $HEF_2$ helix straightens into a single helical intermediate to form a physical anchor between the viral and host cell membranes. This is the extended pre-hairpin intermediate. (iv) Late extended intermediate: the segment D (green) helix rearranges and folds back to form an outer helix forming the intermediate conformation. The second chain reversal region is flexible thus providing slack so that ICV $HEF_2$ can tilt during the subsequent foldback of the C-terminal region. The trimeric central and outer segment D helix in $HEF_2$ remains intact till the final post-fusion conformation. (v) Collapse of intermediate: the fold back of the extended intermediate induces contact and mixing of the outer viral and host cell membranes to form the membrane hemifusion intermediate. (vi) Post-fusion conformation: final formation of the post-fusion $HEF_2$ 6HB leads to the expansion of the fusion pore. The matrix layer is disrupted at the location of the fusion pore. The structures used in this illustration are as follows: panel (i) pre-fusion ICV HEF, PDB: 1FLC; panel (ii) IAV HA dilated early intermediate, PDB: 6Y5I and 6Y5J; panel (iii) IAV HA extended intermediate, PDB: 6Y5K; ICV $HEF_2$ late extended intermediate, PDB: 6WKO; and post-fusion ICV $HEF_2$ based on a homology model using IBV $HA_2$ (PDB: 4NKJ) as a template. Structures are shown as ribbons with the eight segments of $HEF_2$ that undergo conformational changes highlighted in different colors. The figure was generated with BioRender.com

numerous applications. For example, this model can be used to identify sites for the incorporation of molecular probes to detect or trap other orthomyxoviral fusion GP intermediates. Given that influenza HA/HEF is the prototypical class I viral GP, the model principles developed here can also be applied toward other viral fusogens. Moreover, understanding the complete repertoire of structural intermediates will guide the design of inhibitors that block fusion and development of stabilized fusion intermediates that can be used as immunogens in vaccines. The use of trapped fusion intermediates as immunogens has the advantage of exposing transiently available or normally inaccessible epitopes to the humoral immune system. Our results demonstrate that ICV HEF is an ideal system to evaluate intermediate states along the viral fusion reaction coordinate, thus helping to identify the missing pieces in our understanding of the dynamics of virus membrane fusion.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial strains
- METHOD DETAILS
  - Influenza viral fusion glycoprotein expression and purification
  - Secondary structural and melting temperature determination by circular dichroism
  - Free energy profile determination
  - ICV $HEF_2$ crystallization and structure determination
  - Intrinsic tryptophan fluorescence assays
  - Tryptophan-induced fluorescence quenching
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Biophysical analysis
  - Crystallographic data analysis

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109152.

## ACKNOWLEDGMENTS

We thank Dr. Donald Benton (Francis Crick Institute, London UK) for providing the coordinates to the cryo-EM IAV HA low pH intermediate structures prior to official PDB release. In addition, we thank Jean Jakoncic, Alexei Soares, Martin Fuchs, and Vivian Stojanoff on Beamline 17ID-1 (AMX) and 17ID-2 (FMX) at the National Synchrotron Light Source II (NSLS-II) for synchrotron access and support. Funding from a NSERC Discovery Grant (RGPIN-2019-057), CIHR Operating Grant (MOP-115066; PJT-173301), an Ontario Early Researcher Award (ER-13-09-116), and a Canada Research Chair (Tier II) to J.E.L. are gratefully acknowledged. Biophysics and structural biology infrastructure were supported by funding from the Canada Foundation for Innovation John R. Evans Leaders Fund (CFI-JELF). V.H.B.S. is the recipient of a Banting

Postdoctoral Fellowship, and J.D.C. received stipend support from Vanier-CGS and Ontario Graduate Scholarships. Support for work performed at the Center for Biomolecular Structure beamline LIX (16ID) | AMX (17ID-1) | FMX (17ID-2) at NSLS-II is provided by NIGMS P30GM133893 and DOE BER-KP1607011. NSLS-II is supported by DOE BES-FWP-PS001. We also thank Patrick Yip, Dr. Karen Siu, and other members of the Lee lab for experimental assistance and useful discussions.

## AUTHOR CONTRIBUTIONS

Conceptualization, V.H.B.S., J.D.C., and J.E.L.; methodology, V.H.B.S., J.D.C., and J.E.L.; investigation, V.H.B.S. and J.E.L.; data curation, V.H.B.S and J.E.L.; writing – original draft, V.H.B.S., J.D.C., and J.E.L.; writing – review & editing, V.H.B.S. and J.E.L.; funding acquisition, J.E.L.; resources, J.E.L.; supervision, J.E.L.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 13, 2021
Revised: March 31, 2021
Accepted: April 27, 2021
Published: May 18, 2021

## SUPPORTING CITATIONS

The following references appear in the supplemental information: Chen et al. (1999); Mansoor et al. (2010); Matsuzaki et al. (2016).

## REFERENCES

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367.

Aydin, H., Smrke, B.M., and Lee, J.E. (2013). Structural characterization of a fusion glycoprotein from a retrovirus that undergoes a hybrid 2-step entry mechanism. FASEB J. *27*, 5059–5071.

Aydin, H., Cook, J.D., and Lee, J.E. (2014). Crystal structures of beta- and gammaretrovirus fusion proteins reveal a role for electrostatic stapling in viral entry. J. Virol. *88*, 143–153.

Baquero, E., Albertini, A.A., Vachette, P., Lepault, J., Bressanelli, S., and Gaudin, Y. (2013). Intermediate conformations during viral fusion glycoprotein structural transition. Curr. Opin. Virol. *3*, 143–150.

Benhaim, M.A., Mangala Prasad, V., Garcia, N.K., Guttman, M., and Lee, K.K. (2020). Structural monitoring of a transient intermediate in the hemagglutinin fusion machinery on influenza virions. Sci. Adv. *6*, eaaz8822.

Benton, D.J., Nans, A., Calder, L.J., Turner, J., Neu, U., Lin, Y.P., Ketelaars, E., Kallewaard, N.L., Corti, D., Lanzavecchia, A., et al. (2018). Influenza hemagglutinin membrane anchor. Proc. Natl. Acad. Sci. USA *115*, 10112–10117.

Benton, D.J., Gamblin, S.J., Rosenthal, P.B., and Skehel, J.J. (2020). Structural transitions in influenza haemagglutinin at membrane fusion pH. Nature *583*, 150–153.

Bullough, P.A., Hughson, F.M., Skehel, J.J., and Wiley, D.C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. Nature *371*, 37–43.

Calder, L.J., and Rosenthal, P.B. (2016). Cryomicroscopy provides structural snapshots of influenza virus membrane fusion. Nat. Struct. Mol. Biol. *23*, 853–858.

Carr, C.M., and Kim, P.S. (1993). A spring-loaded mechanism for the conformational change of influenza hemagglutinin. Cell *73*, 823–832.

Chan, D.C., Chutkowski, C.T., and Kim, P.S. (1998). Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. Proc. Natl. Acad. Sci. USA *95*, 15613–15617.

Chen, J., Skehel, J.J., and Wiley, D.C. (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.

Chlanda, P., Mekhedov, E., Waters, H., Schwartz, C.L., Fischer, E.R., Ryham, R.J., Cohen, F.S., Blank, P.S., and Zimmerberg, J. (2016). The hemifusion structure induced by influenza virus haemagglutinin is determined by physical properties of the target membranes. Nat. Microbiol. *1*, 16050.

Cohen, F.S., and Melikyan, G.B. (2004). The energetics of membrane fusion from binding, through hemifusion, pore formation, and pore enlargement. J. Membr. Biol. *199*, 1–14.

Das, D.K., Govindan, R., Nikić-Spiegel, I., Krammer, F., Lemke, E.A., and Munro, J.B. (2018). Direct Visualization of the Conformational Dynamics of Single Influenza Hemagglutinin Trimers. Cell *174*, 926–937.e12.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr. *69*, 1204–1214.

Fontana, J., Cardone, G., Heymann, J.B., Winkler, D.C., and Steven, A.C. (2012). Structural changes in Influenza virus at low pH characterized by cryo-electron tomography. J. Virol. *86*, 2919–2929.

Formanowski, F., Wharton, S.A., Calder, L.J., Hofbauer, C., and Meier-Ewert, H. (1990). Fusion characteristics of influenza C viruses. J. Gen. Virol. *71*, 1181–1188.

Greenfield, N.J. (2006). Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat. Protoc. *1*, 2527–2535.

Gui, L., Ebner, J.L., Mileant, A., Williams, J.A., and Lee, K.K. (2016). Visualization and Sequencing of Membrane Remodeling Leading to Influenza Virus Fusion. J. Virol. *90*, 6948–6962.

Ivanovic, T., and Harrison, S.C. (2015). Distinct functional determinants of influenza hemagglutinin-mediated membrane fusion. eLife *4*, e11009.

Jones Brunette, A.M., and Farrens, D.L. (2014). Distance mapping in proteins using fluorescence spectroscopy: tyrosine, like tryptophan, quenches bimane fluorescence in a distance-dependent manner. Biochemistry *53*, 6290–6301.

Kim, Y.H., Donald, J.E., Grigoryan, G., Leser, G.P., Fadeev, A.Y., Lamb, R.A., and DeGrado, W.F. (2011). Capture and imaging of a prehairpin fusion intermediate of the paramyxovirus PIV5. Proc. Natl. Acad. Sci. USA *108*, 20992–20997.

Lakadamyali, M., Rust, M.J., and Zhuang, X. (2004). Endocytosis of influenza viruses. Microbes Infect. *6*, 929–936.

Lee, K.K. (2010). Architecture of a nascent viral fusion pore. EMBO J. *29*, 1299–1311.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D Struct. Biol. *75*, 861–877.

Mansoor, S.E., Dewitt, M.A., and Farrens, D.L. (2010). Distance mapping in proteins using fluorescence spectroscopy: the tryptophan-induced quenching (TrIQ) method. Biochemistry *49*, 9722–9731.

Matlin, K.S., Reggio, H., Helenius, A., and Simons, K. (1981). Infectious entry pathway of influenza virus in a canine kidney cell line. J. Cell Biol. *91*, 601–613.

Matsuzaki, Y., Sugawara, K., Furuse, Y., Shimotai, Y., Hongo, S., Oshitani, H., Mizuta, K., and Nishimura, H. (2016). Genetic Lineage and Reassortment of Influenza C Viruses Circulating between 1947 and 2014. J. Virol. *90*, 8251–8265.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Cryst. *40*, 658–674.

Ni, F., Chen, X., Shen, J., and Wang, Q. (2014). Structural insights into the membrane fusion mechanism mediated by influenza virus hemagglutinin. Biochemistry *53*, 846–854.

Podbilewicz, B. (2014). Virus and cell fusion mechanisms. Annu. Rev. Cell Dev. Biol. *30*, 111–139.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320–W324.

Rosenthal, P.B., Zhang, X., Formanowski, F., Fitz, W., Wong, C.H., Meier-Ewert, H., Skehel, J.J., and Wiley, D.C. (1998). Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. Nature *396*, 92–96.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. *5*, 725–738.

Skehel, J.J., and Wiley, D.C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. *69*, 531–569.

Taraska, J.W., Puljung, M.C., and Zagotta, W.N. (2009). Short-distance probes for protein backbone structure based on energy transfer between bimane and transition metal ions. Proc. Natl. Acad. Sci. USA *106*, 16227–16232.

Taylor, R.M. (1949). Studies on survival of influenza virus between epidemics and antigenic variants of the virus. Am. J. Public Health Nations Health *39*, 171–178.

Wang, M., and Veit, M. (2016). Hemagglutinin-esterase-fusion (HEF) protein of influenza C virus. Protein Cell *7*, 28–45.

White, J.M., Delos, S.E., Brecher, M., and Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. Crit. Rev. Biochem. Mol. Biol. *43*, 189–219.

Whitmore, L., and Wallace, B.A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. *32*, W668–W613.

Wild, C.T., Shugars, D.C., Greenwell, T.K., McDanal, C.B., and Matthews, T.J. (1994). Peptides corresponding to a predictive α-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. Proc. Natl. Acad. Sci. USA *91*, 9770–9774.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315.

Wilson, I.A., Skehel, J.J., and Wiley, D.C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature *289*, 366–373.

Winter, G., Waterman, D.G., Parkhurst, J.M., Brewster, A.S., Gildea, R.J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I.D., et al. (2018). DIALS: implementation and evaluation of a new integration package. Acta Crystallogr. D Struct. Biol. *74*, 85–97.

## STAR★METHODS

### KEY RESOURCES TABLE

| Reagent or resource | Source | Identifier |
|---|---|---|
| **Bacterial and virus strains** | | |
| *Escherichia coli* BL21 (DE3) competent cells | Novagen | Cat#69450 |
| **Chemicals, peptides, and recombinant proteins** | | |
| IAV $HA_2$ (378-515, C481S) recombinant protein | This paper | Uniprot: Q9WFX3 - Influenza A/Brevig Mission/1/1918 H1N1 |
| IBV $HA_2$ (390-532, C489S) recombinant protein | This paper | Uniprot: P03462 - Influenza B/HK/8/1973 |
| ICV $HEF_2$ (495-586, C583S) recombinant protein | This paper | Uniprot: A8E060 - C/Johannesburg/1/1966 |
| ICV $HEF_2$ (481-621, C583S) recombinant protein | This paper | Uniprot: A8E060 - C/Johannesburg/1/1966 |
| ICV $HEF_2$ (495-586, Q558C, C583S) recombinant protein | This paper | Uniprot: A8E060 - C/Johannesburg/1/1966 |
| ICV $HEF_2$ (495-586, H518C, C583S) recombinant protein | This paper | Uniprot: A8E060 - C/Johannesburg/1/1966 |
| ICV $HEF_2$ (481-621, Q558C, C583S, C591S, C595S) recombinant protein | This paper | Uniprot: A8E060 - C/Johannesburg/1/1966 |
| ICV $HEF_2$ (481-621, H518C, C583S, C591S, C595S) recombinant protein | This paper | Uniprot: A8E060 - C/Johannesburg/1/1966 |
| Isopropyl β-D-1-thiogalactopyranoside | Bioshop | Cat#IPT001.50 |
| Ampicillin, Sodium salt | Bioshop | Cat#AMP201.100 |
| Tris base | Bioshop | Cat#TRS003.10 |
| Sodium Chloride | Bioshop | Cat#SOD002.10 |
| Imidazole | Bioshop | Cat#IMD510.1 |
| Ni-NTA agarose affinity resin | QIAGEN | Cat#30230 |
| Thrombin, restriction grade | Millipore Sigma | Cat#69671 |
| Potassium phosphate monobasic | Bioshop | Cat#PPM666.1 |
| Potassium phosphate dibasic | Bioshop | Cat#PPD303.1 |
| Sodium fluoride | Bioshop | Cat#SFL001.500 |
| Monobromobimane | Millipore Sigma | Cat#B4380 |
| NeXtal Tubes Cryos Suite crystallization screen (condition 24) | QIAGEN | Cat#130703 (condition 24) |
| **Deposited data** | | |
| IAV $HA_2$ 6HB postfusion crystal structure | Protein Data Bank | PDB: 1QU1 |
| IBV $HA_2$ 6HB postfusion crystal structure | Protein Data Bank | PDB: 4NKJ |
| ICV $HEF_2$ intermediate crystal structure | This paper | PDB: 6WKO |
| **Software and algorithms** | | |
| Boltzmann biphasic sigmoidal fit | Origin 2017 | https://www.originlab.com/ |
| Standard deviation | Origin 2017 | https://www.originlab.com/ |
| Dichroweb | http://dichroweb.cryst.bbk.ac.uk/html/home.shtml | http://dichroweb.cryst.bbk.ac.uk/html/home.shtml |
| Gibbs free energy determination | Origin 2017 | https://www.originlab.com/ |
| DIALS | https://dials.github.io/ | v3.3.0 |
| Aimless | CCP4 | v7.1.009 |
| Coot | CCP4 | Version 0.8.9.2 |
| Phenix.phaser | Phenix | Version 1.17 |

(*Continued on next page*)

*Continued*

| Reagent or resource | Source | Identifier |
|---|---|---|
| Phenix.refine | Phenix | Version 1.17 |
| MolProbity | MolProbity | http://molprobity.biochem.duke.edu/ |
| PHENIX.polygon | Phenix | Version 1.17 |
| Fluorescence center of mass determination | Origin 2017 | https://www.originlab.com/ |
| One-way ANOVA with a Tukey's posthoc test | GraphPad Prism software, L.L.C | v9.0.0 (121) |
| Pymol | The PyMOL Molecular Graphics System (Schrödinger, LLC.). | https://pymol.org/2/ |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jeffrey E. Lee (jeff.lee@utoronto.ca).

### Materials availability

Reagents generated in this study are available from the Lead Contact upon request.

### Data and code availability

The atomic coordinates and structure factors for the ICV $HEF_{495-586}$ pH 6.1 structure described in this manuscript have been deposited in the Protein Data Bank - PDB under accession code: 6WKO.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Bacterial strains

*Escherichia coli* BL21 (DE3) competent cells were obtained from Novagen/Millipore Sigma (Cat#69450).

## METHOD DETAILS

### Influenza viral fusion glycoprotein expression and purification

DNA sequences corresponding to the IAV $HA_2$ (residues 378-515; A/Brevig Mission/1/1918 H1N1; Uniprot Q9WFX3), IBV $HA_2$ (residues 390-532; B/HK/8/1973; Uniprot P03462), and ICV $HEF_2$ (residues 495-586 and 481-621; C/Johannesburg/1/1966; Uniprot A8E060) core fusion domain were codon-optimized, commercially synthesized, and subcloned into pET46 Ek/LIC. A thrombin protease site was inserted downstream of the vector encoding the N-terminal 6-histidine tag. In order to avoid non-specific intermolecular disulfide-mediated aggregation, cysteine to serine mutations (IAV $HA_2$: C481S, IBV $HA_2$: C489S; ICV $HEF_2$: C583S) were generated. The constructs are denoted as IAV $HA_2$, IBV $HA_2$, ICV $HEF_{495-586}$, and $HEF_{481-621}$ (Robert and Gouet, 2014).

All influenza viral fusion subunits were expressed in *E. coli* BL21 (DE3) (Novagen). Cultures were grown to an $OD_{600}$ of 0.6 at 37°C and recombinant protein expression was induced with a final concentration of 0.5 mM isopropyl β-*D*-1-thiogalactopyranoside (IPTG; Bioshop) for 20 hours at 18°C. Pelleted cells were resuspended in 50 mM Tris-HCl pH 8.0, 300 mM NaCl, and 20 mM imidazole. Bacterial cells were lysed using a hydraulic cell disruption system (Constant Systems), and protein was purified by standard Ni-NTA affinity chromatography (QIAGEN). Proteins were then dialyzed against 10 mM Tris-HCl pH 7.5, and 150 mM NaCl and digested with restriction-grade thrombin (Millipore Sigma) at 4°C for 16 hours. Prior to biophysical assays or crystallization trials, ICV $HEF_2$ was further purified by size exclusion chromatography (SEC) using a custom Superdex 75 prep grade XK 16/40 column (GE) equilibrated in 10 mM Tris-HCl pH 7.5, and 150 mM NaCl. Protein concentration was determined by absorbance at $\lambda_{280nm}$, and purity was monitored by 16% SDS-polyacrylamide gel.

### Secondary structural and melting temperature determination by circular dichroism

Minor modifications were made for the purification of the influenza virus fusion subunits for CD spectroscopy. Cell pellets were resuspended in lysis buffer (10 mM potassium phosphate pH 8.0, 150 mM sodium fluoride, and 20 mM imidazole), and cells were lysed as described above. All fusion subunits were bound onto Ni-NTA resin, washed with 10 column volumes (CV) lysis buffer, and eluted in 5 CV elution buffer (10 mM potassium phosphate pH 8.0, 150 mM NaF, and 500 mM imidazole). The purified proteins were dialyzed overnight at 4°C against 10 mM potassium phosphate pH 7.5, 150 mM NaF prior to SEC on a Superdex 75 prep grade XK 16/40 column.

All CD spectral scans and thermal melting curves of ICV $HEF_{495-586}$, ICV $HEF_{481-621}$, IAV $HA_2$, and IBV $HA_2$ were acquired on a Jasco J-1500 spectropolarimeter in a 1-mm pathlength quartz cuvette (Helma) at a protein concentration of 0.2 $mg.ml^{-1}$. Spectra were recorded over a pH range from 4.5 to 7.5 (0.5 pH unit intervals) using 100 mM potassium phosphate buffer adjusted to the desired pH. CD wavelength scans were collected at 20°C between 183-260 nm at a rate of 50 $nm \cdot min^{-1}$. Raw CD ellipticity values were baseline subtracted, averaged over eight accumulations, and converted to mean residue ellipticity ($\theta_{MRE}$) in units of degree $cm^2$ $dmol^{-1}$ as presented in Equation (1):

$$\theta_{MRE} = \frac{MW}{(N-1)} x \frac{\theta}{(10.d.c)} \quad \text{(Equation 1)}$$

where *MW* is the molecular weight for each sample in Daltons; *N* is the number of amino acids; θ is the ellipticity in millidegree; *d* corresponds to the optical pathlength in cm; *c* is the protein concentration in $mg.mL^{-1}$. Secondary structural compositions (unstructured coils, turns, β strands, and helices) of all influenza viral fusion subunits were calculated using the CDSSTR method and the Set7 (190-240 nm) library with the *Dichroweb* server (Whitmore and Wallace, 2004).

All thermal denaturation assays were carried out at a wavelength (222 nm) selected to maximize the CD signal by increasing the temperature from 20°C to 95°C in 5°C intervals with 120 s equilibration between temperature points. The resultant change in ellipticity was buffer baseline subtracted, normalized between 0 (folded) and 1 (unfolded), and fit using the program Origin2017 to a non-linear biphasic sigmoidal curve to determine the apparent melting temperatures ($T_m$). All CD spectra and thermal denaturation assays were performed in biological triplicates.

## Free energy profile determination

The analyses of thermal denaturation curves of IAV $HA_2$, IBV $HA_2$ and ICV $HEF_2$ revealed a reversible and two-state equilibriums between pH 5.5 and 7.5. The apparent change in *Gibbs* free energy was calculated for all protein samples using Equations (2) and (3) (Greenfield, 2006):

$$\Delta G_i^0 = -RTlnK_{eq},\ K_{eq} = \frac{\theta - \theta_{unfolded}}{\theta_{folded} - \theta} \quad \text{(Equation 2)}$$

$$\Delta\left(\Delta G^0\right) = \Delta G^0_{unfolded} - \Delta G^0_{folded} \quad \text{(Equation 3)}$$

where $\Delta G_i^0$ is the *Gibbs* free energy for each *i* apparent state, R is the gas constant, T temperature in *Kelvin*, θ is the ellipticity in degrees measured at each wavelength. The $\Delta(\Delta G^0)$ is the difference in $\Delta G^0$ between the unfolded and folded microstates. The free energy landscape profiles were determined at 37°C, a temperature below the melting temperature, from biological triplicate experiments.

## ICV $HEF_2$ crystallization and structure determination

Extensive sparse-matrix crystallization screenings of ICV $HEF_{495-586}$ and $HEF_{481-621}$ were performed by sitting drop vapor diffusion in 96-well 2-drop Art Robbins Intelliplates using the Douglas Instrument Oryx 8 liquid handling system. Crystallization of the ICV $HEF_{495-586}$ pH 6.1 intermediate was obtained by concentrating the protein to 60 $mg.mL^{-1}$ and fast diluting to a final concentration of 10 $mg.mL^{-1}$ using 50 mM MES pH 6.0 directly in the sitting drop prior to sparse-matrix screening. ICV $HEF_{495-586}$ crystals were obtained in 85 mM tri-sodium citrate pH 5.6, 29.75% (v/v) tert-butanol, and 15% (v/v) glycerol. The measured pH of the actual crystallization condition (protein buffer plus crystallization reagent) was pH 6.1. Prior to diffraction experiments, crystals were transferred to a drop of cryo-protectant consisting of the mother liquor supplemented with 20% (v/v) glycerol and immediately plunged into liquid nitrogen. Diffraction data for ICV $HEF_{495-586}$ crystals were remotely collected on beamline 17ID-1 (AMX) at the National Synchrotron Light Source-II (NSLS-II; Upton, NY). Diffraction data were reduced using DIALS (Winter et al., 2018), and scaling was performed using Aimless (Evans and Murshudov, 2013). The structure of ICV $HEF_{495-586}$ was determined by molecular replacement, using the program PHENIX.phaser (McCoy et al., 2007) and a polyalanine model of the trimeric IBV $HA_2$ inner HR1 helices (residues 18-77; PDB code: 4NKJ:A) as an initial search model. Iterative rounds of model rebuilding and refinement were performed using the program Coot and PHENIX.refine (Afonine et al., 2012), respectively. Clear electron density was observed for the main chain and side chain residues that allowed for manual model building **(**Figure S1E**)**. Model validation was performed using MolProbity (Williams et al., 2018), Coot (Emsley and Cowtan, 2004), and PHENIX.polygon (Liebschner et al., 2019). Data collection and refinement statistics are presented in Table S2. The ICV $HEF_{495-586}$ structural model was deposited under PDB accession code 6WKO.

## Intrinsic tryptophan fluorescence assays

The relative tryptophan environment was analyzed by monitoring intrinsic fluorescence on a Synergy Neo2 multimode plate reader (BioTek Instruments) with monochromatic excitation at (295 ± 3) nm. ICV $HEF_{495-586}$, ICV $HEF_{481-621}$, and IAV $HA_2$ were prepared at 1 $mg.mL^{-1}$ in the same buffers as used for CD analysis. A fluorescence emission spectral scan was acquired between 303-500 nm. Center of mass for each analyzed spectrum was determined in Origin2017 as shown in Equation (4):

$$\langle\lambda\rangle = \frac{\sum \lambda_i . F_i}{\sum F_i} \quad \text{(Equation 4)}$$

where $< \lambda >$ is the center of spectral mass, and $F_i$ is the fluorescence intensity for each discrete wavelength $\lambda_i$. Mean values and standard deviations were calculated from biological duplicate and technical triplicate experiments.

### Tryptophan-induced fluorescence quenching

The conformational changes of ICV $HEF_2$ as a function of pH were evaluated by tryptophan-induced quenching (TrIQ). Cysteine residues were introduced into ICV $HEF_{481-621}$ and $HEF_{495-586}$ at positions Q558 and H518 to allow for bimane labeling. To ensure that there was a specific, single label on each construct, native cysteine residues at C583, C591, and C595 in ICV $HEF_{481-621}$ and C583 in ICV $HEF_{495-586}$ were substituted with serine. Four constructs were codon-optimized, gene synthesized, and cloned into a pET46 Ek/LIC vector: ICV $HEF_{495-586}$ Q558C, C583S; ICV $HEF_{495-586}$ H518C, C583S; ICV $HEF_{481-621}$ Q558C, C583S, C591S, C595S; ICV $HEF_{481-621}$ H518C, C583S, C591S, C595S. All cysteine mutants were expressed and purified using the protocol for the wild-type ICV $HEF_2$ constructs. SEC-purified samples in 10 mM potassium phosphate pH 7.5, and 150 mM NaCl were concentrated to 1.8 $mg.mL^{-1}$. To assess whether the cysteine substitutions affected the folding of ICV $HEF_2$, a CD spectrum at pH 7.5 was recorded under the same conditions as previously described for wild-type ICV $HEF_2$. All ICV $HEF_2$ cysteine mutants were treated with a final concentration of 5 mM monobromobimane (mBBr, Sigma-Aldrich) for 2 hours at 22°C (Jones Brunette and Farrens, 2014; Taraska et al., 2009). Excess mBBr label was removed using a desalting HiTrap column (GE). Success of protein labeling was assessed by analyses of UV absorbance and fluorescence excitation at 295 nm and 390 nm, respectively.

TrIQ fluorescence signal from the bimane was recorded at 22°C using a Synergy Neo2 multimode plate reader (BioTek Instruments) with monochromatic excitation at (395 ± 5) nm (Taraska et al., 2009). All bimane-labeled ICV $HEF_2$ mutants were prepared at 0.5 $mg.mL^{-1}$ in 100 mM potassium phosphate buffer adjusted to the desired pH (between 4.5 to 7.5 in 0.5 pH unit intervals). Fluorescence emission spectral scans were acquired between 425-600 nm and averaged from technical triplicates, buffer subtracted, and then normalized from 0 to 1. Normalized data from four independently expressed and purified batches of protein were averaged and used to calculate the standard error of the mean (SEM). Free mBBr at a final concentration of 850 μM was used to measure the fluorescence emission reference ($F_0$). The quenching factor ($F/F_0$) was calculated for each pH and mutant construct as described in Equation (5):

$$\left(\frac{F}{F_0}\right)_i = \left(\frac{I_{max,\lambda}}{I_{0,\lambda}}\right)_i \quad \text{(Equation 5)}$$

where $I_{max,\lambda}$ is the normalized maximum fluorescence emission value at a particular wavelength (λ), while $I_{0,\lambda}$ is the normalized maximum fluorescence emission value of a reference free mBBr at the same wavelength. One-way ANOVA was used to determine whether there are any statistically significant differences between $F/F_0$ values as a function of pH, and a *Tukey's post hoc* test was performed to determine the groups that were statistically different (GraphPad Prism 9.0). The average $C_\alpha$-$C_\alpha$ distance between the probe and tryptophan residues was estimated using the calculated ICV $HEF_2$ quenching fraction ($F/F_0$) and comparing to quenching-based distance estimates reported by Jones Brunette and Farrens (2014).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Biophysical analysis

The statistical analyses of biophysical experiments shown in Figures 1, 2, 4, S2, S6, and S7 were performed using Origin2017. One-way ANOVA used in combination with a *Tukey's post hoc* test for the TrIQ assays shown in Figure 4 were performed using GraphPad Prism software v9.0. All quantitative experimental data are presented as mean ± SEM. The number of biological or technical replicates are stated in the figure legends.

### Crystallographic data analysis

Crystallographic data were processed and analyzed using DIALS, Aimless, PHENIX and MolProbity. Data and refinement statistics and analyses are described in Method details and Table S2.