Large fluctuations of the fusion intermediate help SARS-CoV-2 capture host cell membranes

Rui Su[1], Jin Zeng[1] and Ben O'Shaughnessy[1,*]

[1]Department of Chemical Engineering, Columbia University, New York, NY, USA

[*]Correspondence should be addressed to B.O'S (email: bo8@columbia.edu)

Abstract

Cell entry of SARS-CoV-2 is accomplished by the S2 subunit of the spike S protein on the virion surface by fusion of the viral envelope and host cell membrane. Fusion requires the prefusion S2 to transit to its potent, fusogenic form, the fusion intermediate (FI). However, the FI structure is unknown, detailed computational models of the FI are unavailable, and the mechanisms of fusion and entry remain unclear. Here, we constructed a full-length model of the CoV-2 FI by extrapolating from known CoV-2 pre- and postfusion structures. Atomistic and coarse-grained simulations showed the FI is a remarkably flexible mechanical assembly executing large orientational and extensional fluctuations due to three hinges in the C-terminal base. Fluctuations generate a large fusion peptide exploration volume and may aid capture of the host cell target membrane and define the clock for fluctuation-triggered refolding and membrane fusion. This work suggests several novel potential drug targets.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for the ongoing global pandemic COVID-19. In common with other coronaviruses, the lipid envelope of this ~100-nm diameter betacoronavirus[1] encloses a positive-sense single-stranded RNA genome complexed with the nucleocapsid (N) protein. The envelope houses 3 structural proteins: the spike (S) glycoprotein, and the envelope (E) and membrane (M) proteins[2]. COVID-19 is this century's third coronavirus epidemic, following SARS-CoV in 2002 and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012. Coronaviruses are likely to remain a global health threat for the foreseeable future.

Infection by SARS-CoV-2 begins with entry into lung, nasal mucosa or small intestine cells, catalyzed by the spike S protein, a trimeric class I fusion protein whose protomers have two subunits, S1 and S2[3,4]. Following S1-mediated binding to Angiotensin-Converting Enzyme 2 (ACE2) receptors on the host cell membrane, and proteolytic activation at locus S2′ by TMPRSS2 on the plasma membrane (PM)[5] or by cathepsins B and L in the endosome[6], S1 and S2 are thought to dissociate, releasing S2.

Entry is the job of the S2 subunit, by fusion of the viral envelope and host cell membrane. However, at this stage S2 is in its non-fusogenic prefusion state, with buried fusion peptides (FPs). Fusion requires S2 to undergo a major structural transition to its potent, fusogenic form, the fusion intermediate (FI)[7,8]. The S2 FI is the machine that accomplishes fusion and entry, bearing three N-terminal FPs that insert into the host cell membrane, the first breach. The FI then pulls the membranes together by refolding into the postfusion configuration, catalyzing membrane fusion[9]. The viral genomic material is delivered via the fusion pore[10], commandeering the host cell replication and translation machinery for progeny virion assembly and release by exocytosis[11].

The FI is the unsheathed weapon of CoV-2 entry, but little is known about this FI, or indeed the FI of any class I viral fusion protein, and the pathway remains conjectural despite indirect evidence[12,13]. While the prefusion and postfusion CoV-2 spike protein structures were determined by cryo-EM and x-ray crystallography[3,9,14,15], the FI structure, and the prefusion-to-FI transition and membrane fusion mechanisms are not established. However, the transition mechanism is suggested by that of another class I fusion protein, hemagglutinin (HA) of influenza, whose HA2 subunit may transit to the FI by a "spring-loaded" mechanism, in which the B-loop straightens into a helix[12]. The HR1 domain of CoV-2 and other coronaviruses has somewhat analogous loops[3,16], and proline substitutions in these loops stabilize the prefusion CoV-2 spike[3], suggesting CoV-2 may use a similar mechanism.

The prefusion spike protein and the entire virion were computationally studied [17-20], but detailed computational studies of the CoV-2 FI are unavailable, to our knowledge. Lack of experimental structural data has made simulation of class I fusion protein FIs difficult. A model structure of the Ebola GP2 FI was constructed[21] and local transitions of the HA and gp41 FIs were studied[22,23]. The FI has proved experimentally elusive, likely because FIs have sec-min lifetimes[7,24], far shorter than the pre- or postfusion lifetimes, and membrane fusion is thought a fast step in the virus life cycle[25]. Only very recently were FI structures of class I fusion proteins visualized for the first time, with cryo-EM: the HIV-1 gp41 intermediate, the parainfluenza F intermediate and the influenza HA2 intermediate[25-27].

The spike protein is an important target for vaccines, therapeutic antibodies and other antivirals. Most current recombinant neutralizing or vaccine-elicited antibodies to CoV-2 bind the S1 receptor binding domain[28-31], while class I viral fusion inhibitors are thought to bind the fusion-

executing subunit[15,32-34], including one FDA-approved drug against HIV[35]. Importantly, two recently emerged CoV-2 variants harboring spike protein mutations, B.1.351 (South African variant) and P.1 (Brazilian variant), escaped from two antibodies granted FDA emergency authorization, whereas the efficiency of S2-targeting antivirals was unaffected[36]. Moreover, S2 is conserved among coronaviruses[37]. Thus, unveiling the mechanisms governing the CoV-2 S2 fusion intermediate will play a vital role in the search for robust and pan-coronavirus antiviral drugs.

Here, we constructed a full-length structural model of the CoV-2 FI by extrapolating from known CoV-2 pre- and postfusion structures. In a multi-timescale approach, all-atom (AA) and MARTINI coarse-grained (CG) simulations tested the model and probed its long timescale dynamics. Simulations revealed the FI to be a remarkably flexible mechanical assembly subject to large orientational and extensional fluctuations due to three hinges in the C-terminal base. Fluctuations generate a large FP exploration volume and may aid capture of the host cell target membrane and define the clock for fluctuation-triggered refolding and membrane fusion. Our work suggests several novel potential drug targets.

Results

Model of the CoV-2 spike protein fusion intermediate

Entry of SARS-CoV-2 is catalyzed by its S protein, a trimeric class I fusion protein. Each protomer has two subunits, S1 and S2, with three S1 heads sitting upon a trimeric S2 stem[3]. Following binding of the S1 subunit to the host cell ACE2 receptor[5], S1 dissociates from S2[38] and the prefusion S2 trimer undergoes a major structural transition to its potent, fusogenic form, the fusion intermediate (FI)[7] (Fig. 1a).

Since the structure of the CoV-2 FI is not determined, we predicted it based on information from the partially solved pre- and postfusion cryo-EM structures[9] (Figs. 1b,c). A natural question is whether formation of the FI from its prefusion state follows a similar pathway to that proposed for influenza hemagglutinin (HA), another class I viral fusion protein. HA is proposed to use a loaded spring mechanism for the analogous transition[12]. Triggered by low endosomal PH, a loop region of the HA2 fusion subunit is thought to fold into a helix, straightening HA2 and thrusting the fusion peptides (FPs) toward the host cell membrane.

We hypothesized that, similarly, the prefusion-to-FI transition of the CoV-2 S2 subunit is triggered by a transition converting the heptad repeat 1 (HR1) domain into a continuous alpha helix (Figs. 1a,d). In this scenario all unstructured loop domains in the prefusion HR1 become helical and rotate like torsional springs, so that HR1 rotates and straightens into a continuous helix, itself continuous with the downstream central helix (CH). Occurring in all three S2 protomers, the result is a three-helix HR1-CH coiled coil in the FI trimer. This is the mechanical backbone of the FI. The S2′ site is assumed cleaved before this transition, exposing the FP N-terminus ready for host membrane capture, consistent with the finding that CoV-2 lung cell entry is blocked by inhibition of the serine protease TMPRSS2 that cleaves S2′ at the host cell plasma membrane[5]. This cleavage disconnects the golden peptide (GP), but GP remains physically attached to S2[9].

Consistent with this hypothesis, in the postfusion structure HR1 and CH are continuous helices in a trimeric coiled coil[9] (Fig. 1c). Thus, the hypothesis is tantamount to assuming that CH and HR1 adopt their postfusion structures in the FI.

We assume the β-hairpin (BH) domain adopts its postfusion configuration in the FI. Once the HR1 springs are released in the prefusion state, the HR1 and the upstream CR domains are pulled away from their prefusion locations. This would leave large destabilizing cavities in the BH domain, likely favoring its transition to the postfusion configuration where BH is raised to fill the cavities and assembles into a pyramidal base (Fig. 1e). For consistency the same is assumed of GP, since a GP β-strand interacts strongly with an antiparallel β-sheet in BH (Fig. 1a). GP also contributes two small helices, and a long helix in a CH coiled coil groove that completes a six-helix bundle just upstream of BH (Figs. 1a, e). This bundle provides structural support at the base of the CH-HR1 backbone (see later sections).

Downstream of BH the refolding hinge (RFH) domain, being remote from the loaded springs, was taken as the prefusion structure, while we used the HR2 structure of SARS-CoV from NMR[39] whose HR2 sequence is the same as that of CoV-2[15]. The unknown TM and CT structures were predicted by QUARK[40].

We integrated these components into the predicted CoV-2 FI structure shown in Fig. 1b (see "Methods"). The model implies the natural FI-to-postfusion refolding pathway that simultaneously pulls the membranes together[37], in which the unstructured N-terminal loop of the refolding hinge (RFH) domain folds into BH by contributing a β-strand to an antiparallel β-sheet (Figs. 1a, f). The remainder of RFH folds back as a leash packing a GP-CH groove in the six-helix bundle, ending with a small helical portion of RFH that attaches between the two small GP helices of the other two protomers, oriented almost perpendicular to the HR1-CH backbone. Refolding is completed when HR2 becomes partially unstructured to pack a second leash into a HR1-CH groove and supply one helix to a six-helix postfusion bundle with HR1, the fusion core[9].

Thus, the GP domain appears to chaperone refolding (Figs. 1a, f). (1) It initiates refolding of RFH, by providing the adjacent β-strand in BH onto which the RFH β-strand folds. (2) It guides refolding, by providing the GP-CH groove into which the RFH leash packs. (3) It may stabilize refolding, by providing two small helices that interact with the small RFH helix, pinning the RFH leash ends and hindering unzippering.

All-atom simulation of the fusion intermediate

Using complementary all-atom (AA) and coarse-grained (CG) MARTINI molecular dynamics (MD) approaches we tested the FI model of Fig. 1b and measured the FI dynamics (see "Methods").

During ~ 0.3 μs of AA simulation the basic secondary structure remained unaltered (Fig. S1), lending credibility to the model. The RFH-HR2 base region below the BH domain was highly flexible, allowing considerable tilt of the FI (Fig. 2 and Supplementary Movie 1). Interestingly, relative to the prefusion structure the three RFH helices became splayed with separated N-termini, in an inverted tripod appearing to offer structural support to the upstream BH and backbone. The HR2 helices became partially unstructured, a structural plasticity that helped the FI to tilt to greater angles at the membrane.

Three hinges endow the fusion intermediate with high flexibility

Next we measured long time FI dynamics using MARTINI CG simulations. This framework fixes the secondary structure, but accesses timescales two orders of magnitude greater. (Throughout, our model assumes S2′ is cleaved; we tested an uncleaved model with connected GP and FP, but the FPs were sequestered and unable to access the target membrane, Fig. S2.)

In 40 μs total running time over 5 runs, the FI exhibited large configurational fluctuations, bending and reorienting over a wide range of angles (Fig. 3a,c). To quantify the flexibility we

measured the curvature statistics along the FI (Figs. 3b, S3, Supplementary Movies 2,3 and "Methods"). This procedure identified three high flexibility hinge regions in the base. Mapping back to the atomistic structure located the hinges as unstructured loops at the BH/RFH, RFH/HR2 and HR2/TMD interfaces, respectively (Fig. 3b,c). The hinges had similar mean magnitudes of curvature, with large fluctuations (Fig. S4).

These hinges are closely related to and have roughly the same locations as three hinges identified in the prefusion CoV-2 S protein by a study combining cryoET and MD simulations[17]. Thus we adopt the "ankle, knee, hip" notation of that study. However, the hip hinge (RFH/BH interface) in our FI structure is more flexible than the prefusion hip due to the significantly altered RFH structure, with splayed helices (Fig. 2b).

We suggest that, by endowing the FI with high flexibility, the hinges let the FI search a large volume to capture target membrane efficiently. Further, the flexible hip hinge may function as the trigger for the refolding transition on the pathway to membrane fusion (see below).

Large fluctuations of the fusion intermediate lead to a large membrane capture volume

The first task of the unleashed FI is presumably to capture the host target membrane by insertion of the fusion peptides at the N-terminus of each protomer. Due to its flexibility the FI extension ranged from ~ 21-30 nm (mean ~ 26 nm), its orientation varied over angles ~ $\pm 60°$ to the membrane normal, and the FPs in consequence swept out volume at a rate ~ 750 $\mathrm{nm}^3\ \mu\mathrm{s}^{-1}$, for a total volume ~ 25,000 nm^3 (Figs. 3d,e, 95% of sampled locations).

Thus, due to the flexible base hinges combined with the large reach of the HR1-CH backbone, the FI accesses a substantial capture volume, equivalent to that of a ~ 36-nm-diameter sphere. This may help the virus to rapidly reconnect with the host cell following dissociation of S2 from the S1/ACE2 complex. Rapid host membrane capture may limit refolding back into the virion membrane in a postfusion state without host cell contact. Indeed, postfusion spike proteins were observed by cryoET on intact SARS-CoV-2 virions[41].

Structure and dynamics of the membrane-bound fusion peptide

We used a multiscale approach to study the secondary structure and spatiotemporal statistics of the membrane-bound FP removed from its host FI (Fig. 4a). The FI model assumes the prefusion FP, but its structure likely changes on binding to the membrane which provides a radically altered environment. Thus we used CG MARTINI MD to bind and equilibrate the bound FP (24 μs total), followed by 2 μs of AA simulation to realistically evolve the bound secondary structure (Fig. 4b).

During the AA simulation the secondary structure evolved somewhat (Figs. 4b, c and Supplementary Movies 4,5). The N-terminus helix barely changed, but the two C-terminal helices merged into one. (The mean total helix content was 35%, compared to ~20% from circular dichroism spectroscopy[42], a difference possibly due to the different simulated and experimental membrane compositions.) The equilibrated N-terminal and C-terminal helices were amphiphilic and anchored the FP to the membrane, with hydrophobic and hydrophilic residues oriented towards and away from the membrane, respectively (Figs. 4d,e).

In an 80 μs CG simulation we then measured the statistics of the bound, equilibrated FP (Supplementary Movies 6,7). The depth of residues decreased somewhat (Fig. S5a), and the C-terminal helix transiently unanchored for periods lasting ~0.3 μs (Fig. S6). Anchoring of this section required helical structure, since it became unanchored when a disordered secondary structure was imposed (Fig. S5b).

The bound FP had rms length ~ 1.5 nm and width ~ 0.9 nm, defined as the greater and smaller of the gyration tensor eigenvalues in the x-y plane, while the rms thickness was ~ 1.0 nm (Figs. 4f, S7 and "Methods"). The FP extended with the anchored helices at either end, roughly speaking, as the length was strongly correlated with the separation between the anchored helices. The radius of gyration autocorrelation function revealed a configurational memory time of ~ 270 ns (Fig. 4g), with similar times for the length, width and thickness (Fig. S8).

Measurement of fusion peptide-membrane binding rate constant

The target membrane is captured by insertion of the fusion peptide. To quantify the binding kinetics, we removed a FP from its FI host and measured the binding rate constant between two membranes separated by $h = 5.5$ nm (Fig. 5).

Prior to binding the FP collided ~ 27 times per μs with the membrane, defining a collision as an approach to within $R_{\mathrm{FP}} \sim 1.6$ nm, the rms FP end-to-end distance (Fig. S9). The FP then became irreversibly bound. Averaged over 10 MARTINI simulations the unbound probability decayed exponentially with time constant $\tau_{\mathrm{bind}} \sim 4.0 \pm 0.6$ μs (Fig. 5a). Thus, the long-time center of gravity binding rate was measured ($h > R_{\mathrm{FP}}$) and the initial condition was forgotten since diffusion-controlled effects were negligible[43].

The FP-membrane binding rate constant k_{bind} is defined in terms of an imagined situation with a solution of FPs at density c_{FP} contacting a membrane, such that $d\rho/dt = k_{\mathrm{bind}}\, c_{\mathrm{FP}}$ where ρ is the areal number density of bound FPs[43]. From the binding assay,

$$k_{\mathrm{bind}}^{\mathrm{FP}} = \frac{h}{2\,\tau_{\mathrm{bind}}} \approx 0.7 \text{ nm } \mu\text{s}^{-1}$$

where the factor of 2 reflects the two membranes.

Importantly, binding was mediated by the N-terminal helix of the FP, as this helix always provided first contact with the membrane during a binding event (Fig. 5b and Supplementary Movie 8). This is consistent with cleavage at the S2′ site being required for viral entry[5], as cleavage exposes the helix.

Interaction of the fusion intermediate with a target membrane

How rapidly will the FP bind the target membrane when attached to the FI? The estimated binding rate is ~ $k_{\mathrm{bind}}^{\mathrm{FP}}\, c_{\mathrm{FP}}^{1}(0)$ where $c_{\mathrm{FP}}^{1}(0) \sim 1/\Delta z$ is the 1D FP density evaluated at the membrane, $z = 0$, and $\Delta z \sim 10$ nm is the range of FP locations normal to the membrane (Fig. 3e). This gives an estimated binding time $\tau_{\mathrm{bind}} \sim \Delta z / k_{\mathrm{bind}}^{\mathrm{FP}} \sim 15$ μs.

Simulations of an FI within striking range of a target host membrane revealed much slower binding than suggested by this estimate. We recorded no FP-mediated binding events in a total of ~ 300 μs of CG MARTINI simulations (over 20 independent runs) with a target membrane 20 nm from the viral membrane (Fig. 6a). In both the FP-only and the FI binding assays, the probability density for the FP N-terminal helix was severely depleted near the target membranes (Figs. 6b, c and S9, S10), so that the net probabilities to lie within 1 nm of the membrane were only ~ 0.5% and ~ 0.07% for the FP-only and FI cases, respectively. Accounting for the different spreads Δz of FP locations, the (suitably normalized) probability to lie within 1 nm of the membrane was only ~ 2-fold lower when attached to the FI.

Thus, when attached to the FI, insertion of the FP into the target membrane is considerably slower due to its local environment, such that FP-membrane binding requires ~ 300 μs or longer. Since this is unexplained by reduced membrane contact probability, we suggest orientational or energetic restrictions underlie the slower binding kinetics.

To verify that the FP is capable of maintaining a bound state, we enforced binding by pulling the FP of an FI with its head in a (1,2) configuration (see below) into the membrane (Fig. 6d). The FP remained stably bound for all of an 8 μs simulation.

Lateral organization of fusion peptides in the FI head optimizes membrane collisions

In the simulations above the FPs and upstream CR domains arranged themselves in a qualitatively universal spatial pattern at the N-terminal end of the HR1-CH backbone. The 3 FP-CR domains formed a blob we call the head, laterally extended relative to the backbone direction. The 3 FP N-terminal helices resided at either end of the head, two at one end, and one at the other (Fig. 6e and Supplementary Movie 9). Frequently the 2 helices at one end were almost merged. Accordingly, the N-terminal helix density distribution along the principal axis in the plane normal to the HR1 backbone typically had two peaks of unequal amplitude separated by ~5 nm. In other cases three equal amplitude peaks were present (Fig. 6f), indicating separated neighboring helices, or two peaks with smaller separation (Fig. S11).

The lateral orientation of the head, with FP N-terminal helices at either end, appears optimal for presenting the helices to the membrane for binding. With the host membrane ~ 20 nm away, likely imposed by the earlier S1-ACE2 binding episode, the FI is severely bent (Figs. 6e, f). In this configuration, when the head collides with the membrane it orients almost perpendicular to the membrane, forcing the helices at the end of the head into contact with the membrane.

The single unpaired N-terminal FP helix tended to be exposed, while the pair of helices tended to sequester one another. Thus, FP-membrane collisions were far more frequent for (1,2) than for (2,1) configurations, where a configuration is defined as (top, bottom) with the top of the head closest to the membrane. In runs with (2,1) configurations, despite the head approaching the membrane repeatedly the helices failed to collide with the membrane (Fig. S12). In ten other runs (224 μs total) with only (1,2) configurations, FP N-terminal helices were ~ 80% more frequently within 1 nm of the membrane compared to unbiased runs (Fig. 6c).

Discussion

The outbreak of Covid-19 saw rapid efforts to characterize the pre- and postfusion SARS-CoV-2 spike protein structures, both since determined by cryo-EM and x-ray crystallography[3,9,14,15]. However, the structure of the fusion intermediate, the fusogenic form of the spike protein that facilitates entry, has not been determined for CoV-2 or any coronavirus. Among class I fusion proteins, the fusion intermediates of HIV-1, parainfluenza and influenza were only recently visualized[25-27].

Lack of structural information has constrained computational study of fusion intermediates. Here we built an atomistic model of the CoV-2 FI, extrapolating from pre- and postfusion structures[9] (Fig. 1a). The model structure was stable in atomistic simulations (Fig. 2), and atomistic and coarse-grained simulations revealed large configurational fluctuations (Figs. 2, 3, 6).

The model implies a pathway from the prefusion to FI to postfusion states (Figs. 1a, 7). It suggests a loaded spring release mechanism generates the FI from the prefusion state, in which loop-to-helix transitions straighten and extend the HR1 domain. Given the postfusion HR1 is entirely helical[9], this inference is by far the most natural. Indeed, the recently solved cryo-EM structure of the HA extended intermediate revealed a long coiled coil in common with the

postfusion structure[26], confirming that influenza uses a loaded spring release mechanism in which the B-loop becomes helical[12]. The prefusion-to-FI transition for CoV-2 differs qualitatively from that of HA in two respects. First, four loaded springs release (all four loop regions in HR1, Figs. 1a, d). Second, straightening of HR1-CH would leave large cavities in the downstream BH domain, which we propose trigger a transition to the more stable postfusion BH configuration (Fig. 1e).

The release mechanism extends the FI to capture the host cell membrane. The present FI model implies the details of subsequent refolding, in which the refolding hinge (RFH, called the linker region in a study of SARS-CoV[44]) folds back, donating a β-strand and α-helix to the GP and packing a GP-CH six-helix bundle groove as a leash (Figs. 1a, f and 7d). Refolding delivers HR2 to the six-helix fusion core with HR1, while an unstructured portion of HR2 packs a HR1 groove as a leash. The RFH/HR2 leashes are conceptually similar to the leashes filling the hydrophobic grooves of the central coiled coil in influenza HA[45].

Simulations of the FP alone identified a 10 residue N-terminal helix that directs binding by making first contact with the membrane, and anchors the FP once bound (Fig. 5b). This accords with a recent AA study of the CoV-2 FP[46]. Moreover, in both AA and CG simulations this helix was the principal anchor stably attaching the FP to the membrane (Figs. 4e and S5a). The N-terminal helix, inherited from the prefusion structure, remained stable during a 2 μs AA simulation of a bound FP (Fig. 4c). A key role for the N-terminal FP helix is consistent with cleavage at the S2′ site being required for viral entry[5], as cleavage exposes the helix. The precise timing is unknown. Cleavage could occur before loaded spring release, as S2′ is not deeply buried within the prefusion structure[9] and may become further exposed when the structure near S2′ is destabilized following S1 binding to the host receptor[38] (Fig. 7b). Alternatively, cleavage could occur after release, protecting the viral membrane from self-penetration by sequestering the FP (Fig. S2).

Simulations showed the FI has high flexibility, which may aid membrane capture and trigger refolding. The flexibility is due primarily to three base hinges, termed knee, ankle and hip (Figs. 3b, c). These are very similar to the previously identified prefusion CoV-2 hinges, proposed to aid receptor binding[17]. The hinges may serve several vital roles for the FI. First, they greatly enlarge the region accessible to the FPs at the FI terminal, aiding capture of host cell membrane (Fig. 7c). Flexibility may also help multiple FIs simultaneously capture target membrane at different distances. Indeed, influenza, parainfluenza and HIV-1 appear to use several FIs for virus-host attachment[25,27,47]. Second, by allowing the ~25 nm long FI to bend significantly, the extreme FI flexibility may facilitate the prefusion-to-FI transition even in the confined circumstances of a nearby host cell membrane (Fig. 6a). Third, flexibility allows the FI to tilt its head and present the FPs directly to the target membrane, possibly using bending forces to assist FP insertion (Figs. 6e and 7c). Fourth, high FI flexibility implies large configurational fluctuations. Large hinge fluctuations occurred in AA and CG simulations (Figs. 2 and 3). Thus, after a microscopically long waiting time, a sufficient rotational hip hinge fluctuation may stochastically destabilize the RFH helix, triggering RFH refolding into BH and the GP-CH groove (Fig. 7d).

Secondary structures in the base of the FI were surprisingly dynamic, a structural plasticity that provided mechanical support and flexibility to the FI (Figs. 2b and S1). Relative to their prefusion configuration, the three RFH helices splayed into a flexible inverted tripod that acted like a suspension system buffering longitudinal fluctuations of the FI backbone (Figs. 2b and 3c). The splayed helices allowed the upstream RFH loops more freedom, so that the hip hinge was significantly more flexible than the prefusion[17] hip. The ease with which the helices splay suggests they interact weakly, consistent with their postfusion transformation to leashes.

The HR2 helices at the very base of the EI also showed structural plasticity. In the AA simulation, sections of HR2 made frequent transitions between helical and unstructured forms, provoked by large FI fluctuations (Figs. 2b and S1). When the FI tilted to a large angle, the most stretched outer HR2 helices became unstructured, presumably due to bending and extensional stresses. This appeared to increase the ankle hinge flexibility.

Target membrane capture by the FI was far slower than suggested by fusion peptide-only simulations. Interactions between membranes and the fusion peptides of CoV-2 and other viruses are often studied experimentally and computationally[42,46,48-50]. Most studies use isolated fusion peptides, unattached to the glycoprotein. We find that binding kinetics of the isolated Cov-2 FP are misleadingly rapid, with a binding rate constant $k_{\mathrm{bind}}^{\mathrm{FP}} \sim 0.7$ nm μs^{-1} (Fig. 5). Applied to the full FI-attached situation this rate constant predicted a ~ 15 μs delay time for the virus to capture the target membrane, whereas we recorded no membrane capture events in ~ 300 μs of coarse-grained simulation time (Fig. 6).

Thus, the binding rate of the FP N-terminal helix to the target membrane is much lower when attached to the FI. The (suitably normalized) probability for the FP N-terminal helix center of gravity to lie within 1 nm of the membrane was only ~ 2-fold lower when attached (Fig. 6c), suggesting orientational restrictions or an increased binding energy barrier due to neighboring CR domains may underlie the slower binding kinetics. Indeed, the FP configuration was entirely different when attached to the FI. With the CR domains, the FPs formed a laterally extended trimeric blob at the end of the HR1-CH backbone, the head. The N-terminal active helices located at either end of the head in a (1,2) or (2,1) configuration (Figs. 6e,f). Since most FI configurations were bent, this oriented the helices toward the membrane, and the unpaired helix made multiple membrane collisions. Spontaneous binding will presumably eventually occur, since the FP remained securely bound for all of a ~8 μs simulation when binding was enforced by pulling the FP into the membrane (Fig. 6d).

Our study suggests the golden peptide (GP) domain may chaperone refolding and is a potential therapeutic target. The GP is a disconnected fragment between the S1/S2 and S2′ sites, both of which must be cleaved for normal replication[51]. Since GP is found in similar locations in both pre- and postfusion states[9], it appears to interact strongly with components of S2. Its inclusion in the FI structure is thus obligatory. In the FI, GP participates in a long antiparallel β-sheet within the BH and wraps around the base of CH and may serve as a stabilizing socket for the CH-HR1 backbone (Fig. 7a).

GP appears to have a second critical role as a chaperone of refolding of the RFH domain (Fig. 7d). First, GP provides the β-strand that the RFH N-terminus loop folds onto as a parallel β-strand, helping to initiate RFH refolding. Second, GP/CH provides a groove for the RFH leash to pack, continuing refolding. Third, a small RFH helix is pinned by the GP catch, a U-shaped sequence including 2 small helices. The catch may rectify zippering of the leash, preventing its unravelling from the groove.

Interestingly, a recent study identified an 8-residue region in the prefusion RFH helix as the epitope of two cross-reactive monoclonal antibodies[52]. This region is the small RFH helix that engages the GP catch, and the four downstream residues. Thus, the antibodies may neutralize CoV-2 by binding RFH and interfering with the GP catch that rectifies RFH refolding.

GP is therefore a potential therapeutic target. HR2-derived peptides inhibit fusion by SARS-CoV-2 and MERS-CoV, presumably blocking formation of the HR1-HR2 six-helix fusion core[15,34,53]. Their efficiency as fusion inhibitors is insensitive to mutations in the spike protein[36], suggesting potential as robust antiviral drugs. We suggest similar strategies using GP- or RFH-

derived peptides could inhibit fusion by binding RFH or GP and disabling FI refolding. Such peptides might also stabilize FI for visualization of this critical intermediate. Indeed, the fusion intermediates of HIV gp41 and parainfluenza F were recently visualized by stabilizing the intermediate with HR2-derived fusion inhibitors[25,27].

What sets the timing of loaded spring release and refolding? We suggest refolding must occur after loaded spring release. In the prefusion state HR1 and CR reside in BH between RFH and the β-strands that help initiate RFH refolding, and HR1 helices fill the GP-CH grooves, preventing groove packing by the RFH leashes (Fig. 1a). Thus, refolding is unclamped by loaded spring release.

Following spring release, we propose the delay to refolding is the waiting time for a conformational FI fluctuation, a large hip rotation, sufficient to destabilize one of the RFH helices into an unstructured loop (Fig. 7d). The loop would be highly susceptible to GP-chaperoned refolding. Complete refolding of one protomer would draw the membranes together and rotate the FI, helping the remaining protomers to refold and position the trans-postfusion trimer ready for fusion and entry.

Acknowledgements

We gratefully acknowledge high-performance computing resources from Microsoft Azure, provided through the COVID-19 HPC Consortium from government, industry and academia that volunteers computing time and resources to support COVID-19 research. Particular thanks are due to Geralyn Miller, Jer-Ming Chia and John Sawyer at the Microsoft AI for Good Research Lab for building and maintaining the HPC system. Columbia University's Shared Research Computing Facility is gratefully acknowledged for the initial stages of this project.

Author contributions

B.O'S designed the research and performed mathematical analysis and analysis of the structure. R.S and J.Z built the structure and performed the simulations. R.S, J.Z and B.O'S analyzed the data. B.O'S and R.S wrote the paper with contributions from J. Z.

Competing interests

The authors declare no competing interests.

Methods

Building a complete structure for the fusion intermediate of the SARS-CoV-2 spike protein. The primary sequence of the SARS-CoV-2 S protein was obtained from the NCBI database (GenBank: MN908947). We first built the HR1-RFH portion of the FI (including the associated GP domain). We used Modeller[54] with two specified templates: the postfusion structure (PDB: 6XRA) for HR1-BH and GP, and the prefusion structure (PDB: 6XR8) for RFH. Several constraints including 3-fold symmetry and preservation of secondary structure were specified. The C-terminal domains (HR2, TMD and CT) were then appended to the HR1-RFH portion one by one using Modeller. The source for HR2 was a SARS-CoV HR2 NMR structure (PDB: 2FXP), while the TMD/CT structure was predicted by the QUARK server by providing the primary sequences. The FP and CR were then extracted as a whole from the prefusion structure (PDB: 6XR8) and appended to the HR1 domain in the FI with an arbitrary angle using Pymol. Finally, a complete GP structure was made in the FI using Modeller, by appending the N-terminal portion (residues 686-702) and C-terminal portion (residues 771-815) of the prefusion structure to the solved postfusion GP (residues 703-770).

Multiscale molecular dynamics simulations. We used two complementary molecular dynamics (MD) approaches, all-atom (AA) and coarse-grained (CG) MARTINI, to simulate the FI and its FP. For MARTINI simulations, all running times reported here are 4 times the raw MARTINI running time, to compensate for the faster sampling of the MARTINI model[55]. Time-dependent data in MARTINI simulations were analyzed after adjusting in this fashion. We used pure DPPC membranes to represent both the viral envelope and the host cell membrane for simplicity. All simulation systems were solvated with 150 mM NaCl at neutral pH. Simulations were run in the NPT ensemble at 1 bar and 310 K using GROMACS 2019.6[56,57]. For details of the simulations and the analysis please see *Supplementary Methods*.

All-atom simulation of the fusion intermediate. The FI model structure (Fig. 1b) was placed in a planar membrane in a $16 \times 16 \times 43$ nm^3 box using the CHARMM-GUI membrane builder[58]. The system was energy-minimized for 1,000 steps and equilibrated for 2.2 ns. The production simulation was run for 305 ns. Secondary structure was identified using the *dssp* algorithm[59,60].

Coarse-grained simulations of the fusion intermediate. The FI model structure (Fig. 1b) was first mapped into the MARTINI CG representation using the *martinize* utility[61,62] and then placed in a planar membrane in a $30 \times 30 \times 50$ nm^3 box using the *insane* utility[63]. The system was energy-minimized for 4,000 steps and equilibrated for 4 ns. Five production simulations starting from the same equilibrated system were run for 8 μs.

To fit a curve to represent the FI ectodomain backbone (residues 912-1237), we represented each residue by one point, calculated as the average position of the backbone (non-sidechain) beads of each residue among the three protomers. The FI backbone was then divided into two parts and each part separately smoothed using the LOWESS algorithm. The two parts were then joined and the final curve was fitted by the B-spline method. The location of a given residue was defined as the normalized arclength of the nearest point on the fitted curve to the position of this residue averaged over the three protomers, using only the backbone beads to locate the residues.

The isolated fusion peptide binding assay. The atomistic structure of the FP was extracted from the prefusion cryo-EM structure (PDB: 6XR8) and coarse-grained into MARTINI representation. Then the FP was placed ~1 nm above a planar bilayer in a $7 \times 7 \times 10$ nm^3 box using the *insane* utility[63]. By implementing periodic boundary condition, this is equivalent to using two planar membranes separated by ~5.5 nm. The C-terminal carboxyl group was neutralized as it connects to the CR in the full-length FI. Ten 24 μs parallel runs were performed after energy minimization and equilibration, all with the same initial conditions. A binding event was defined to be when the z coordinate of the FP center of gravity (COG) first had a value that positioned it below the upper membrane leaflet and above the lower membrane leaflet, where the leaflet locations were defined as the average locations of the PO4 beads in each leaflet.

All-atom simulation of a membrane-bound fusion peptide. In a given CG FP binding simulation the final configuration of the FP and the membrane to which it was bound was converted to atomic resolution in CHARMM36 force field[64,65] using the *backward* utility[66]. Then the system was re-solvated and equilibrated for 200 ns. The production simulation was run for 2 μs. The membrane insertion depth of each FP residue was defined as the vertical distance between the residue COG and the COG of PO_4 groups in the membrane leaflet to which the FP was bound.

Coarse-grained simulations of an equilibrated fusion peptide bound to a membrane. Once the AA simulation was complete, the final configuration of the FP and the membrane to which it was bound was coarse-grained into MARTINI representation. The system was re-solvated, the energy minimized for 6,000 steps and the FP equilibrated for 0.8 ns. The production simulation was run for 80 μs. In a variation of this procedure, the FP C-terminal helix was transformed to an unstructured loop by changing the secondary structure file of the FP, an input to the coarse-graining utility.

The membrane insertion depth of each residue was defined in the same way as for the AA simulation (see above). The gyration tensor M of the FP was defined as

$$M = \frac{1}{N}\sum_{\mathrm{i}=1}^{N}(r_\mathrm{i} - r_\mathrm{c})\otimes(r_\mathrm{i} - r_\mathrm{c}),$$

where $\otimes$ represents the dyadic product, r_i is the coordinate of the ith bead in the FP, N is the total number of beads, and $\mathrm{r_c}$ is the COG of the FP, as $\mathrm{r_c} = \sum_{i=1}^{N} \mathrm{r}_i\,/N$. The radius of gyration R_g was computed as the square root of the trace of M. The length and the width of the FP were defined as, respectively, greater and smaller of the eigenvalues of M projected onto the x-y plane, and the thickness was defined as $\sqrt{M_{\mathrm{zz}}}$.

Simulations of the fusion intermediate interacting with a target membrane. In the ten unbiased simulations, ten snapshots from three FI simulations were used as initial conditions, in which the FI ectodomain protruded ~20 nm normal to the membrane. Another pre-equilibrated planar membrane (run for 4 μs) was placed 20 nm above the membrane anchoring the FI. Each configuration, containing an FI and two membranes, was re-solvated, equilibrated for 4 ns, and run for 8 μs. In the biased set of simulations, all ten runs (each lasting 22.4 μs) started from the same initial condition, in which the COG of the nearest N-terminal FP helix was within 1 nm of the membrane, with the FI head in the (1,2) configuration. Here, the membrane position was defined to be the mean location of all the PO_4 beads in the lower leaflet of the upper membrane.

Pulling a fusion peptide into the target membrane. An initial condition was chosen with the nearest N-terminal FP helix lying within 1 nm of the membrane and with the FI head in the (1,2) configuration. The helix was then pulled towards the target membrane at speed 2.5 nm/μs for ~1.2 μs, until the COGs of the helix and membrane were separated by ~0.1 nm in the *z* direction. The pulling force was then released, and the simulation was run for 8 μs.

Simulations of the fusion intermediate with uncleaved S2′ sites. To construct a FI with uncleaved S2' sites, the model FI structure (Fig. 1b) was fist coarse-grained into MARTINI representation. The FP N-terminus and the GP C-terminus were then pulled together at 10 nm/μs in a MARTINI simulation, with the residues in other domains fixed in space. The C-terminal portion of the postfusion GP domain that is unsolved in the postfusion structure[9] (residues 771-815) was forced to be a loop, to make pulling easier. Following the pulling episode, which lasted ~1.5 μs, the GP and the FP in one protomer were covalently connected. Using Pymol, this protomer was duplicated twice to make a trimer. Finally this FI models with uncleaved S2′ sites was placed in a planar membrane and simulated for 4 μs.

Data availability

All simulation data supporting the findings in this paper are available from the corresponding author (bo8@columbia.edu) upon reasonable request. The structure of the model of the SARS-CoV-2 spike fusion intermediate presented here is available at https://github.com/RuiSu11/cov-2_public

References

1 Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* **382**, 727-733, doi:10.1056/NEJMoa2001017 (2020).

2 Tu, Y.-F. *et al.* A Review of SARS-CoV-2 and the Ongoing Clinical Trials. *International Journal of Molecular Sciences* **21**, 2657 (2020).

3 Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263, doi:10.1126/science.abb2507 (2020).

4 Ziegler, C. G. K. *et al.* SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell* **181**, 1016-1035.e1019, doi:https://doi.org/10.1016/j.cell.2020.04.035 (2020).

5 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e278, doi:https://doi.org/10.1016/j.cell.2020.02.052 (2020).

6 Ou, X. *et al.* Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications* **11**, 1620, doi:10.1038/s41467-020-15562-9 (2020).

7 Harrison, S. C. Viral membrane fusion. *Nature Structural & Molecular Biology* **15**, 690-698, doi:10.1038/nsmb.1456 (2008).

8 Heald-Sargent, T. & Gallagher, T. Ready, Set, Fuse! The Coronavirus Spike Protein and Acquisition of Fusion Competence. *Viruses* **4**, 557-580 (2012).

9 Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586-1592, doi:10.1126/science.abd4251 (2020).

10 Cohen, F. S. How Viruses Invade Cells. *Biophysical Journal* **110**, 1028-1032, doi:10.1016/j.bpj.2016.02.006 (2016).

11 Jaafar, Z. A. & Kieft, J. S. Viral RNA structure-based strategies to manipulate translation. *Nature Reviews Microbiology* **17**, 110-123 (2019).

12 Carr, C. M. & Kim, P. S. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* **73**, 823-832, doi:https://doi.org/10.1016/0092-8674(93)90260-W (1993).

13 Kim, Y. H. *et al.* Capture and imaging of a prehairpin fusion intermediate of the paramyxovirus PIV5. *Proceedings of the National Academy of Sciences* **108**, 20992-20997, doi:10.1073/pnas.1116034108 (2011).

14 Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e286, doi:https://doi.org/10.1016/j.cell.2020.02.058 (2020).

15 Xia, S. *et al.* Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Research* **30**, 343-355, doi:10.1038/s41422-020-0305-x (2020).

16 Yuan, Y. *et al.* Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nature Communications* **8**, 15092, doi:10.1038/ncomms15092 (2017).

17 Turoňová, B. *et al.* In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203-208, doi:10.1126/science.abd5223 (2020).

18 Ali, A. & Vijayan, R. Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Scientific Reports* **10**, 14214, doi:10.1038/s41598-020-71188-3 (2020).

19 Casalino, L. *et al.* Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science* **6**, 1722-1734, doi:10.1021/acscentsci.0c01056 (2020).

20 Yu, A. *et al.* A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophysical Journal* **120**, 1097-1104, doi:https://doi.org/10.1016/j.bpj.2020.10.048 (2021).

21 Barfoot, S., Poger, D. & Mark, A. E. Understanding the Activated Form of a Class-I Fusion Protein: Modeling the Interaction of the Ebola Virus Glycoprotein 2 with a Lipid Bilayer. *Biochemistry* **59**, 4051-4058, doi:10.1021/acs.biochem.0c00527 (2020).

22 Lin, X., Noel, J. K., Wang, Q., Ma, J. & Onuchic, J. N. Atomistic simulations indicate the functional loop-to-coiled-coil transition in influenza hemagglutinin is not downhill. *Proceedings of the National Academy of Sciences* **115**, E7905-E7913, doi:10.1073/pnas.1805442115 (2018).

23 Lin, M. & Da, L.-T. Refolding Dynamics of gp41 from Pre-fusion to Pre-hairpin States during HIV-1 Entry. *Journal of Chemical Information and Modeling* **60**, 162-174, doi:10.1021/acs.jcim.9b00746 (2020).

24 Ivanovic, T., Choi, J. L., Whelan, S. P., van Oijen, A. M. & Harrison, S. C. Influenza-virus membrane fusion by cooperative fold-back of stochastically induced hemagglutinin intermediates. *eLife* **2**, e00333, doi:10.7554/eLife.00333 (2013).

25 Ladinsky, M. S. *et al.* Electron tomography visualization of HIV-1 fusion with target cells using fusion inhibitors to trap the pre-hairpin intermediate. *eLife* **9**, e58411, doi:10.7554/eLife.58411 (2020).

26 Benton, D. J., Gamblin, S. J., Rosenthal, P. B. & Skehel, J. J. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* **583**, 150-153, doi:10.1038/s41586-020-2333-6 (2020).

27 Marcink, T. C., Wang, T., des Georges, A., Porotto, M. & Moscona, A. Human parainfluenza virus fusion complex glycoproteins imaged in action on authentic viral surfaces. *PLOS Pathogens* **16**, e1008883, doi:10.1371/journal.ppat.1008883 (2020).

28 Chen, P. *et al.* SARS-CoV-2 neutralizing antibody LY-CoV555 in outpatients with Covid-19. *New England Journal of Medicine* **384**, 229-237 (2021).

29 Baum, A. *et al.* REGN-COV2 antibodies prevent and treat SARS-CoV-2 infection in rhesus macaques and hamsters. *Science* **370**, 1110-1115 (2020).

30 Jackson, L. A. *et al.* An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. *New England Journal of Medicine* **383**, 1920-1931, doi:10.1056/NEJMoa2022483 (2020).

31 Walsh, E. E. *et al.* Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates. *New England Journal of Medicine* **383**, 2439-2450, doi:10.1056/NEJMoa2027906 (2020).

32 Bosch, B. J. *et al.* Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8455-8460, doi:10.1073/pnas.0400576101 (2004).

33 Marcink, T. C. *et al.* Hijacking the Fusion Complex of Human Parainfluenza Virus as an Antiviral Strategy. *mBio* **11**, e03203-03219, doi:10.1128/mBio.03203-19 (2020).

34 de Vries, R. D. *et al.* Intranasal fusion inhibitory lipopeptide prevents direct-contact SARS-CoV-2 transmission in ferrets. *Science* **371**, 1379-1382, doi:10.1126/science.abf4896 (2021).

35 LaBonte, J., Lebbos, J. & Kirkpatrick, P. Enfuvirtide. *Nature Reviews Drug Discovery* **2**, 345 (2003).

36 Hoffmann, M. *et al.* SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell*, doi:https://doi.org/10.1016/j.cell.2021.03.036 (2021).

37 Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R. & Daniel, S. Coronavirus membrane fusion mechanism offers as a potential target for antiviral development. *Antiviral Research*, 104792 (2020).

38 Benton, D. J. *et al.* Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **588**, 327-330, doi:10.1038/s41586-020-2772-0 (2020).

39 Hakansson-McReynolds, S., Jiang, S., Rong, L. & Caffrey, M. Solution Structure of the Severe Acute Respiratory Syndrome-Coronavirus Heptad Repeat 2 Domain in the Prefusion State *. *Journal of Biological Chemistry* **281**, 11965-11971, doi:10.1074/jbc.M601174200 (2006).

40 Xu, D. & Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Structure, Function, and Bioinformatics* **80**, 1715-1735, doi:https://doi.org/10.1002/prot.24065 (2012).

41 Ke, Z. *et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498-502, doi:10.1038/s41586-020-2665-2 (2020).

42 Lai, A. L. & Freed, J. H. SARS-CoV-2 Fusion Peptide has a Greater Membrane Perturbating Effect than SARS-CoV with Highly Specific Dependence on Ca2+. *Journal of Molecular Biology* **433**, 166946, doi:https://doi.org/10.1016/j.jmb.2021.166946 (2021).

43 O'Shaughnessy, B. & Sawhney, U. Polymer Reaction Kinetics at Interfaces. *Physical Review Letters* **76**, 3444-3447, doi:10.1103/PhysRevLett.76.3444 (1996).

44 Fan, X., Cao, D., Kong, L. & Zhang, X. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nature Communications* **11**, 3618, doi:10.1038/s41467-020-17371-6 (2020).

45 Park, H. E., Gruenke, J. A. & White, J. M. Leash in the groove mechanism of membrane fusion. *Nature Structural & Molecular Biology* **10**, 1048-1053, doi:10.1038/nsb1012 (2003).

46 Khelashvili, G., Plante, A., Doktorova, M. & Weinstein, H. Ca²⁺-dependent mechanism of membrane insertion and destabilization by the SARS-CoV-2 fusion peptide. *Biophysical Journal* **120**, 1105-1119, doi:10.1016/j.bpj.2021.02.023 (2021).

47 Calder, L. J. & Rosenthal, P. B. Cryomicroscopy provides structural snapshots of influenza virus membrane fusion. *Nature Structural & Molecular Biology* **23**, 853-858, doi:10.1038/nsmb.3271 (2016).

48 Lai, A. L., Millet, J. K., Daniel, S., Freed, J. H. & Whittaker, G. R. The SARS-CoV Fusion Peptide Forms an Extended Bipartite Fusion Platform that Perturbs Membrane Order in a Calcium-Dependent Manner. *Journal of Molecular Biology* **429**, 3875-3892, doi:https://doi.org/10.1016/j.jmb.2017.10.017 (2017).

49 Baylon, J. L. & Tajkhorshid, E. Capturing Spontaneous Membrane Insertion of the Influenza Virus Hemagglutinin Fusion Peptide. *The Journal of Physical Chemistry B* **119**, 7882-7893, doi:10.1021/acs.jpcb.5b02135 (2015).

50 Marzinek, J. K. *et al.* A Funneled Conformational Landscape Governs Flavivirus Fusion Peptide Interaction with Lipid Membranes. *Journal of Chemical Theory and Computation* **14**, 3920-3932, doi:10.1021/acs.jctc.8b00438 (2018).

51 Bestle, D. *et al.* TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Science Alliance* **3**, e202000786, doi:10.26508/lsa.202000786 (2020).

52 Wang, C. *et al.* A conserved immunogenic and vulnerable site on the coronavirus spike protein delineated by cross-reactive monoclonal antibodies. *Nature Communications* **12**, 1715, doi:10.1038/s41467-021-21968-w (2021).

53 Lu, L. *et al.* Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nature Communications* **5**, 3067, doi:10.1038/ncomms4067 (2014).

54 Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **54**, 5.6.1-5.6.37, doi:https://doi.org/10.1002/cpbi.3 (2016).

55 Marrink, S. J., de Vries, A. H. & Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *The Journal of Physical Chemistry B* **108**, 750-760, doi:10.1021/jp036508g (2004).

56 Bekker, H. *et al.* in *Physics computing.* 252-256 (World Scientific Singapore).

57 Berendsen, H. J., van der Spoel, D. & van Drunen, R. J. C. p. c. GROMACS: a message-passing parallel molecular dynamics implementation. **91**, 43-56 (1995).

58 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859-1865, doi:https://doi.org/10.1002/jcc.20945 (2008).

59 Kabsch, W. & Sander, C. J. B. O. R. o. B. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **22**, 2577-2637 (1983).

60 Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. **39**, D411-D419 (2010).

61 de Jong, D. H. *et al.* Improved parameters for the martini coarse-grained protein force field. **9**, 687-697 (2013).

62 Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. J. T. j. o. p. c. B. The MARTINI force field: coarse grained model for biomolecular simulations. **111**, 7812-7824 (2007).

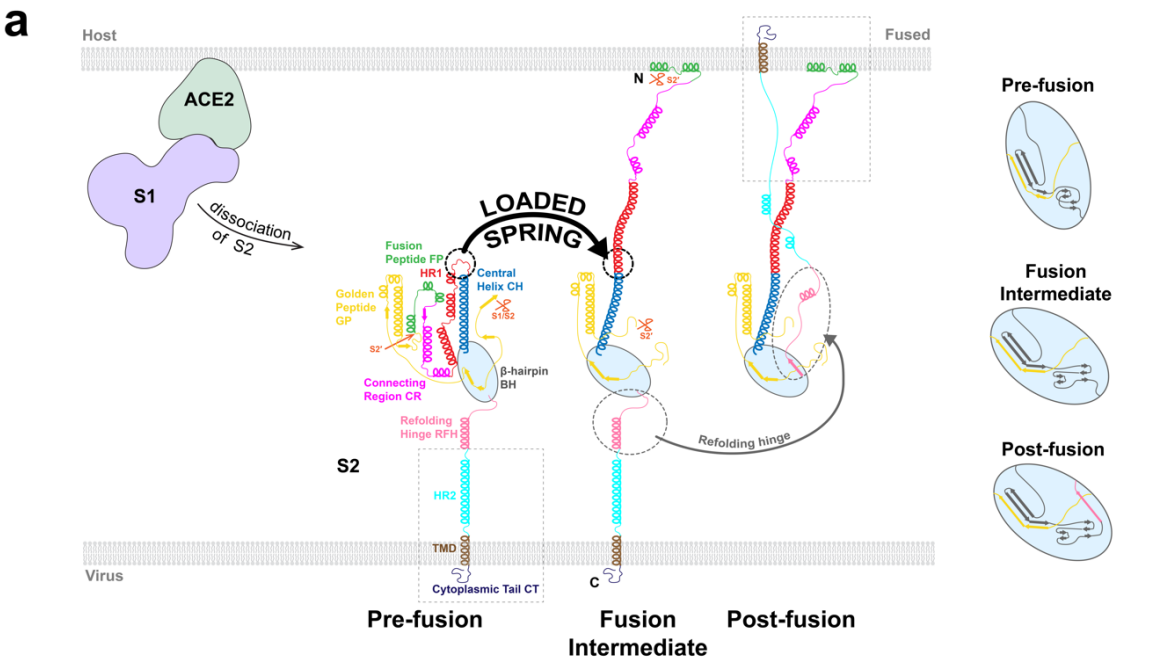
63 Wassenaar, T. A. *et al.* Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. **11**, 2144-2155 (2015).

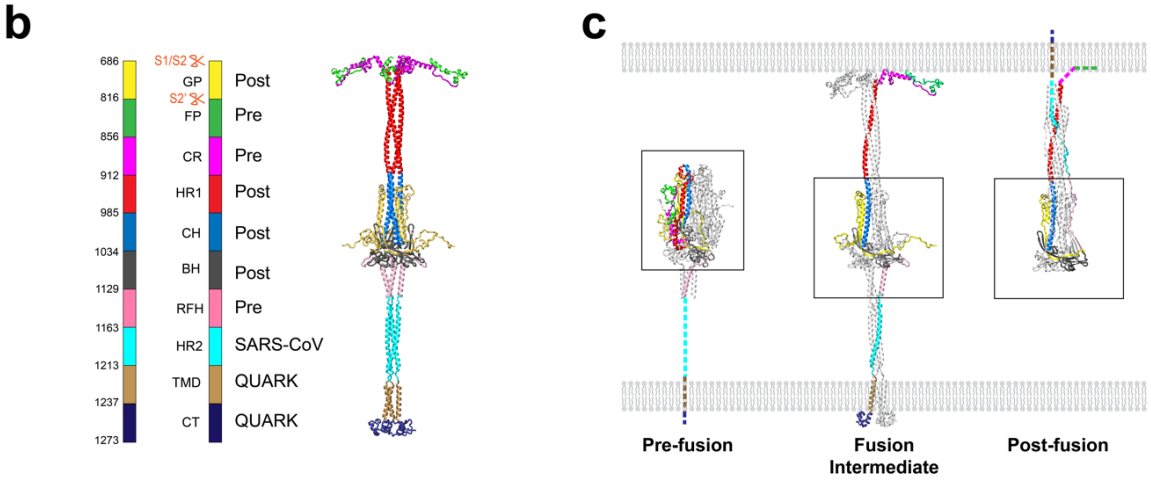
64 Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. **8**, 3257-3273 (2012).

65 Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. **114**, 7830-7843 (2010).

66 Wassenaar, T. A. *et al.* Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. **10**, 676-690 (2014).

Figures





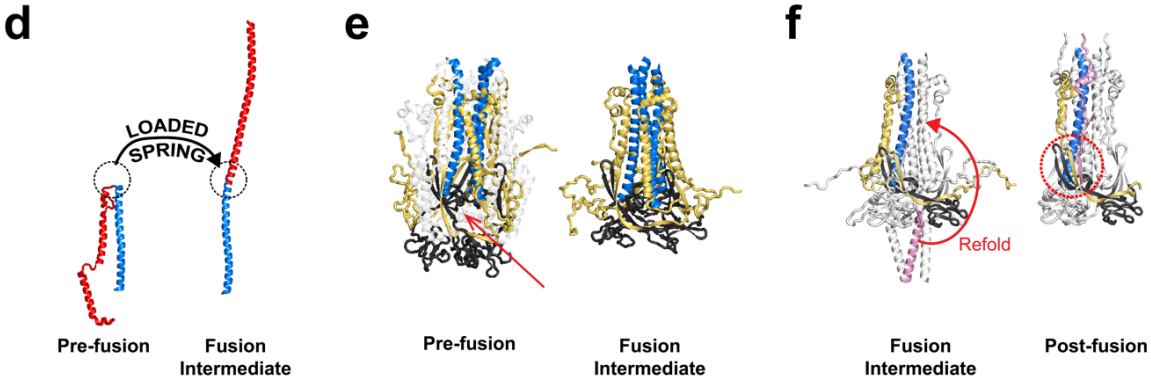


Figure 1. Model of the SARS-CoV-2 spike protein fusion intermediate

(a) Model of the fusion intermediate (FI) of the CoV-2 spike (S) protein, schematic. One protomer of the S trimer is shown. Schematics of the pre- and postfusion S2 states from the known structures, other than unknown regions (boxed). Details of BH and adjacent domains, at right. Following dissociation of S2 from S1, the unstructured pre-fusion HR1 loops become helical (loaded spring release), giving the HR1-CH backbone that thrusts the FPs toward the host cell membrane. The FI subsequently refolds into the postfusion structure, driven by structural changes in RFH and chaperoned by GP, with RFH and HR2 providing leashes that pack the HR1-CH backbone grooves.

(b) Model of the FI of the CoV-2 spike (S) protein, exact structure. Source structures for each S2 subunit domain are indicated, either CoV-2 prefusion (PDB: 6XR8), CoV-2 postfusion (PDB: 6XRA) or HR2 of CoV (PDB: 2FXP). Transmembrane domain (TMD) and cytoplasmic tail (CT) structures predicted by QUARK.

(c) Comparison between predicted FI structure and known crystal structures of the prefusion (PDB: 6XR8) and postfusion (PDB: 6XRA) CoV-2 S2 subunit. One protomer highlighted in color. Dashed lines: missing domains from partially solved crystal structures.

(d) Details of loaded spring transition.

(e) β-hairpin (BH) domains in the known prefusion and predicted FI structures (boxed regions of (c)). The RFH domain is omitted from the FI BH for clarity. Following the loaded spring transition, HR1, CR and FP (shown faint in the prefusion BH) vacate their prefusion locations in BH. The resultant cavity (arrow) would presumably be unstable. We assume the FI adopts the more compact postfusion BH structure (right).

(f) The golden peptide (GP) chaperones refolding of the fusion intermediate (FI) into the postfusion structure. Blowups of boxed regions in (c) are shown. Refolding of the refolding hinge (RFH) domain is guided by GP. RFH forms a parallel β-strand with GP (red circle), the RFH unstructured portion packs the CH-GP groove, and RFH helices interact with two GP helices. Colored BH and CH belong to one protomer; colored RFH belongs to a different protomer.

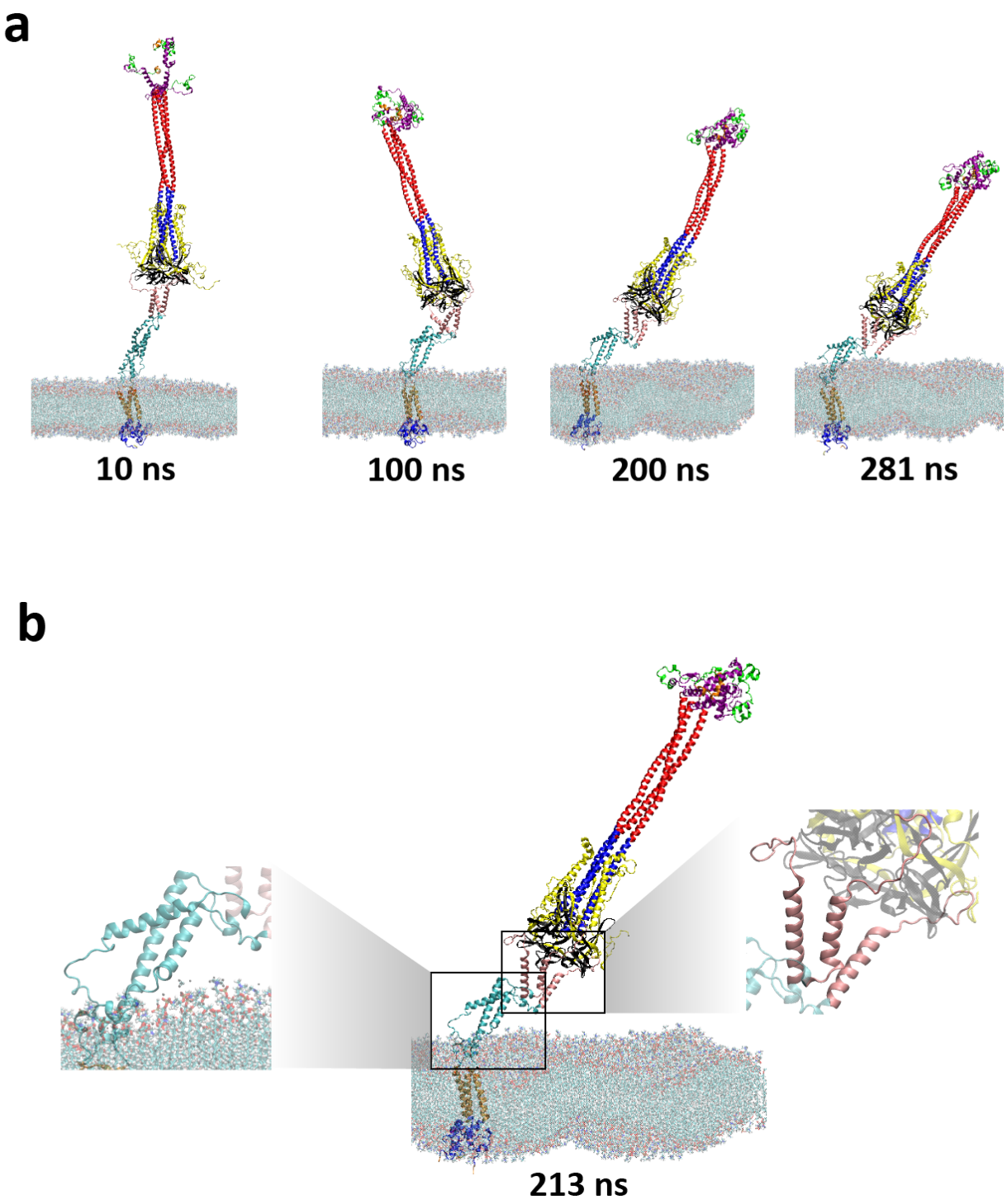


Figure 2. All-atom simulation of the SARS-CoV-2 fusion intermediate
Color code for this and all subsequent figures, as for Fig. 1. In addition, the N-terminal helices of the fusion peptides are shown orange.
(a) Snapshots of the fusion intermediate (FI) during the ~ 0.3 μs all-atom simulation of the model of Fig. 1b. The secondary structure was stable. The FI undergoes large bending and extensional fluctuations throughout the simulation.

(b) Snapshot of the FI after 213 ns of the AA simulation. The RFH and HR2 domains (highlighted) have secondary structural plasticity. Relative to the prefusion structure, the RFH helices splay into an inverted tripod, a mechanical suspension system for the upstream HR1-CH backbone, BH and GP domains. HR2 had dynamic secondary structure, with helices becoming partially unstructured. Bending of the FI stretched the outermost HR2 helices, partially converting the helices into unstructured sections.

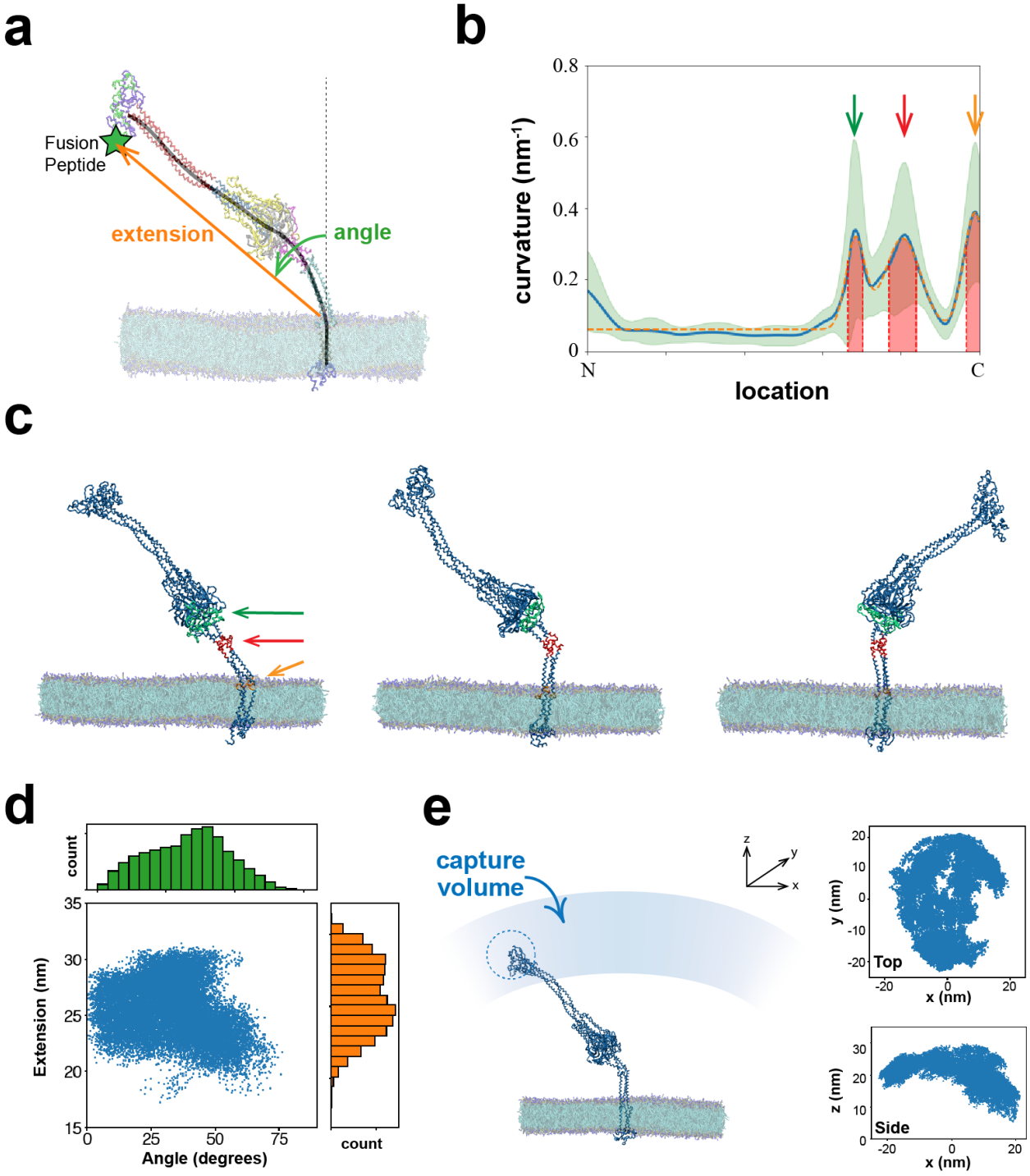


Figure 3. The fusion intermediate is highly flexible and visits a large capture volume
(a) In coarse-grained MARTINI simulations the FI had large configurational fluctuations, measured by the extension and angle of orientation of the FI backbone (black curve).
(b) Time averaged backbone curvature versus normalized backbone arclength. Three high curvature hinges are apparent (arrows). Each hinge region (red) was defined as the quarter width of a fitted Gaussian (orange). Green envelope indicates SD.

(c) Simulation snapshots with the three hinges highlighted, identified as residues 1084-1138, 1156-1178 and 1204-1213.
(d) Distributions of FI extensions and angles.
(e) The FI has a large capture volume. Top and side views of FP locations visited. The FI extension and orientation ranges are ~21-30 nm and of $\pm 56°$, respectively (95% of sampled values) so that a large capture volume is swept out over time, shown schematically (left). Dashed circle: approximate region explored by the FP in 1 μs.
(b), (d), (e) Statistics are averages over the last 4 μs of five 8 μs runs, for a total of 20 μs simulation time.

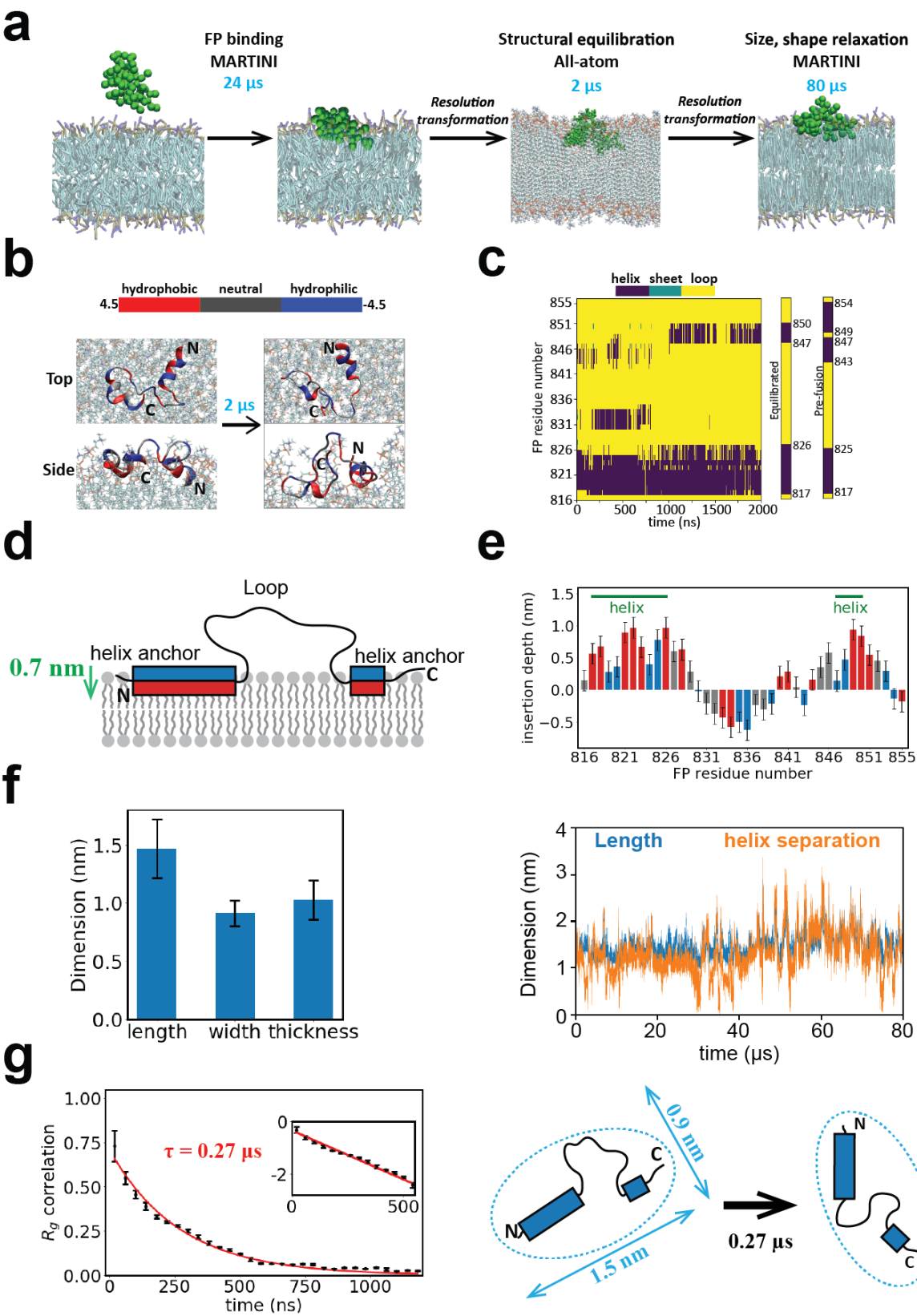


Figure 4. Multiscale simulations of the membrane-bound fusion peptide.
(a) Multiscale simulation strategy to measure secondary structure and spatiotemporal statistics of the membrane-bound FP. The prefusion structure is coarse-grained to MARTINI representation and bound to and relaxed within the membrane in a 24 μs simulation (binding required ~ 4 μs). Following backmapping to atomistic resolution, the secondary structure of the membrane-bound FP is relaxed in a 2 μs AA simulation. Assigning each residue its most frequently visited secondary structure during the final 0.8 μs of the AA simulation, the FP is again coarse-grained and its spatial dimensions and relaxation time measured in an 80 μs CG simulation.
(b) Evolution of FP structure during the 2 μs AA simulation of (a). Initial and final states are shown. FP resides in one of three colors depending on the hydrophobicity.
(c) Evolution of bound FP secondary structure during the 2μs AA simulation of (a). The initial (prefusion) and final (equilibrated) structures are compared. For each residue the equilibrated structure shows the most frequently adopted in the final 0.8 μs.
(d) Equilibrated bound FP following the AA equilibration of (a), schematic. The principal anchor is the amphiphilic N-terminal helix, with a secondary amphiphilic C-terminal helix anchor. Hydrophobicity color scheme as for (b).
(e) Mean membrane insertion depth profile along the bound FP in the equilibrated structure represented in (c) (see "Methods"). Mean values over 0.8 μs.
(f) Length and helix separation of the bound FP during the 80 μs MARTINI simulation of (a). Mean dimensions averaged over the final 78 μs (left).
(g) Temporal correlation function of the radius of gyration of the bound FP yields shape memory time $\tau = 269 \pm 1$ ns . (Bin size, 40 ns. 100 data points per bin.) Inset: log-lin representation. Dashed lines: exponential fit. Top view, schematic (right). All error bars: SD.

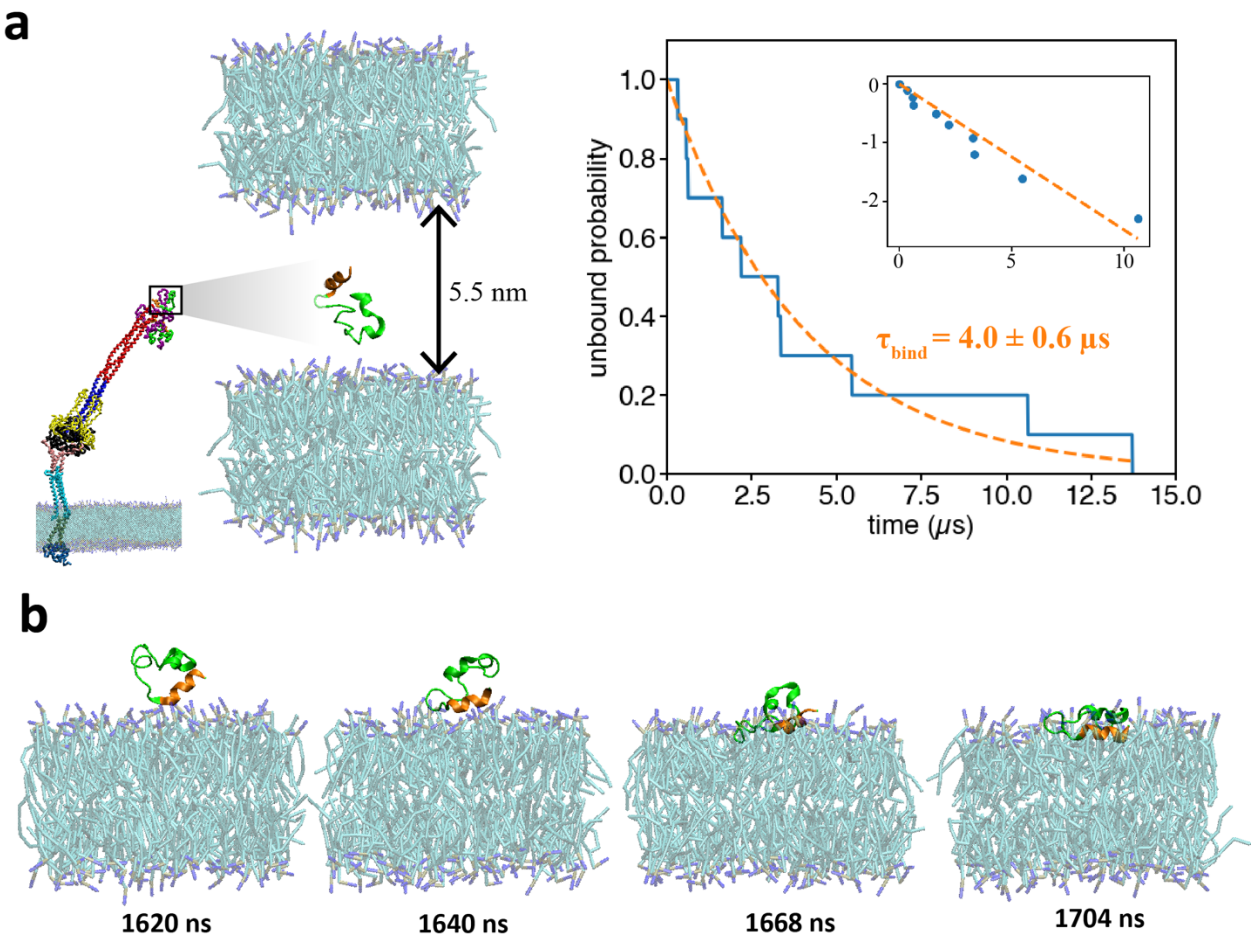


Figure 5. Membrane binding kinetics of an isolated fusion peptide

(a) Binding assay to measure the membrane binding rate constant, $k_{\mathrm{bind}}^{\mathrm{FP}}$, of a FP removed from its host FI. Initially the FP is positioned between two membranes separated by 5.5nm (left). FP dynamics are simulated using the CG MARTINI force field and the time to irreversibly bind the membrane measured. The unbound fraction (blue, right) among ten simulated FPs decays exponentially with time constant $\tau_{\mathrm{bind}} = 4.0 \pm 0.6$ μs (dashed orange curves). Inset: log-lin representation.

(b) Typical binding event. The N-terminal helix (orange) is the first binding contact. To show secondary structure, the FP was back-mapped to all-atom representation.

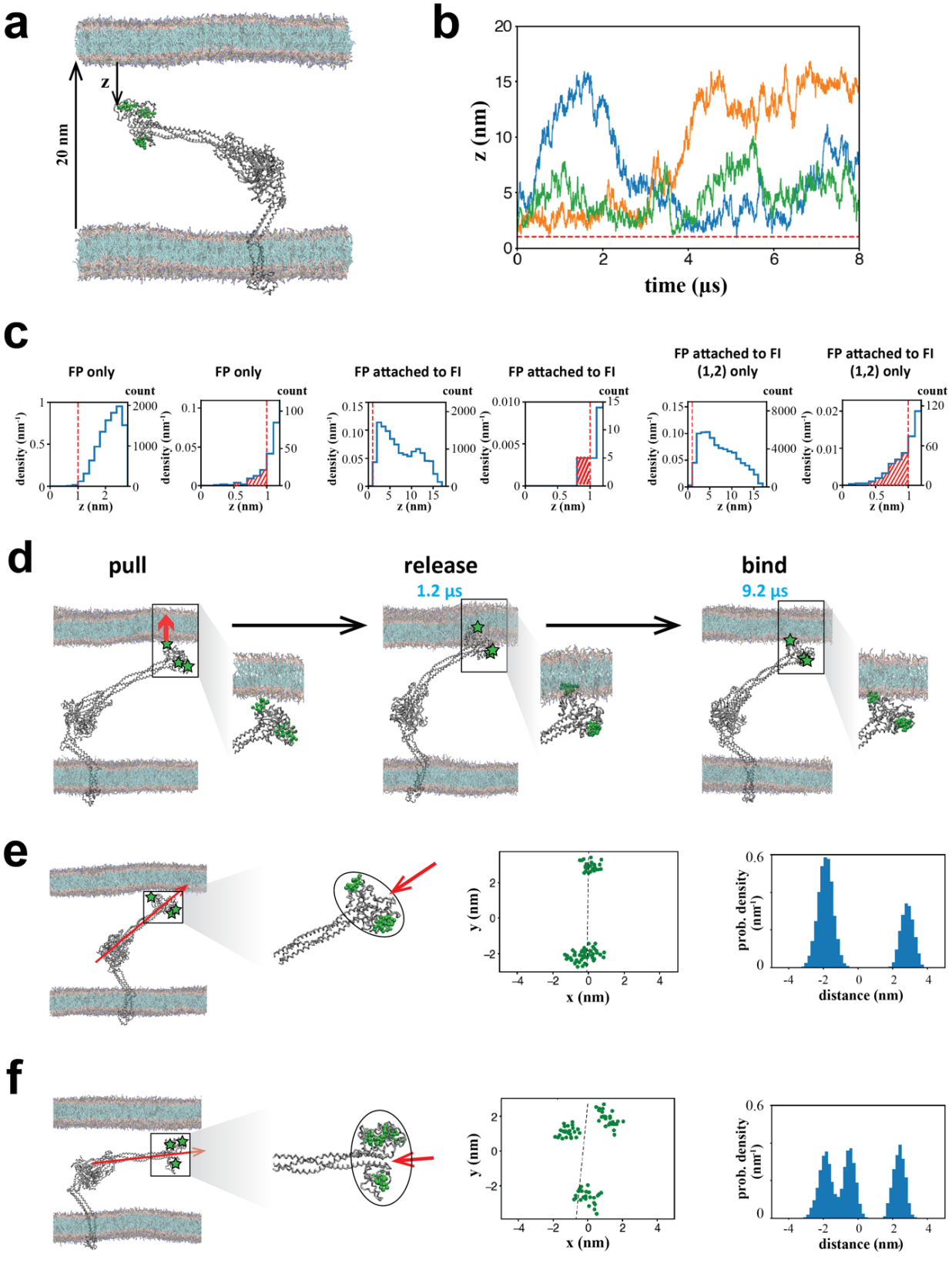


Figure 6. Interaction of the fusion intermediate with a target membrane.
(a) Snapshot from the CG MARTINI simulation. The FI interacts with a target membrane 20 nm from the viral membrane.
(b) Distance of the nearest N-terminal FP helix from the membrane versus time in three runs.
(c) Probability density versus distance *z* of the N-terminal FP helix from the membrane, during a binding assay for an isolated FP (Fig. 5) or a binding assay for a FP attached to the FI (nearest FP). Both FP-only and FI binding assays reveal a density depletion zone at the membrane. The net probability to lie within 1nm of the membrane is 0.51% (isolated FP) or 0.07% (FP attached to FI), shown hatched. In biased FI runs using only (1,2) configurations, the probability was 0.12%.
(d) Enforced binding of a FP. The FP N-terminal helix was pulled into the membrane over a period of 1.2 μs, and then released. The FP remained bound to the membrane for all of a 8 μs CG MD simulation.
(e) At the N-terminal end of the FI backbone the three FP-CR domains organize into a laterally extended blob, the FI head. The FPs reside at either end of the head (green stars locate N-terminal helices). A (1,2) configuration is shown (one FP at the membrane proximal end, two FPs at the distal end, high membrane collision frequency). End view (perspective of red arrow) of N-terminal helix beads and density distribution of the helix beads along the principal axis (dashed black line) in the plane normal to the backbone. One FP is isolated while the FP pair is merged.
(f) Same as (d), for a (2,1) head configuration (low membrane collision frequency).

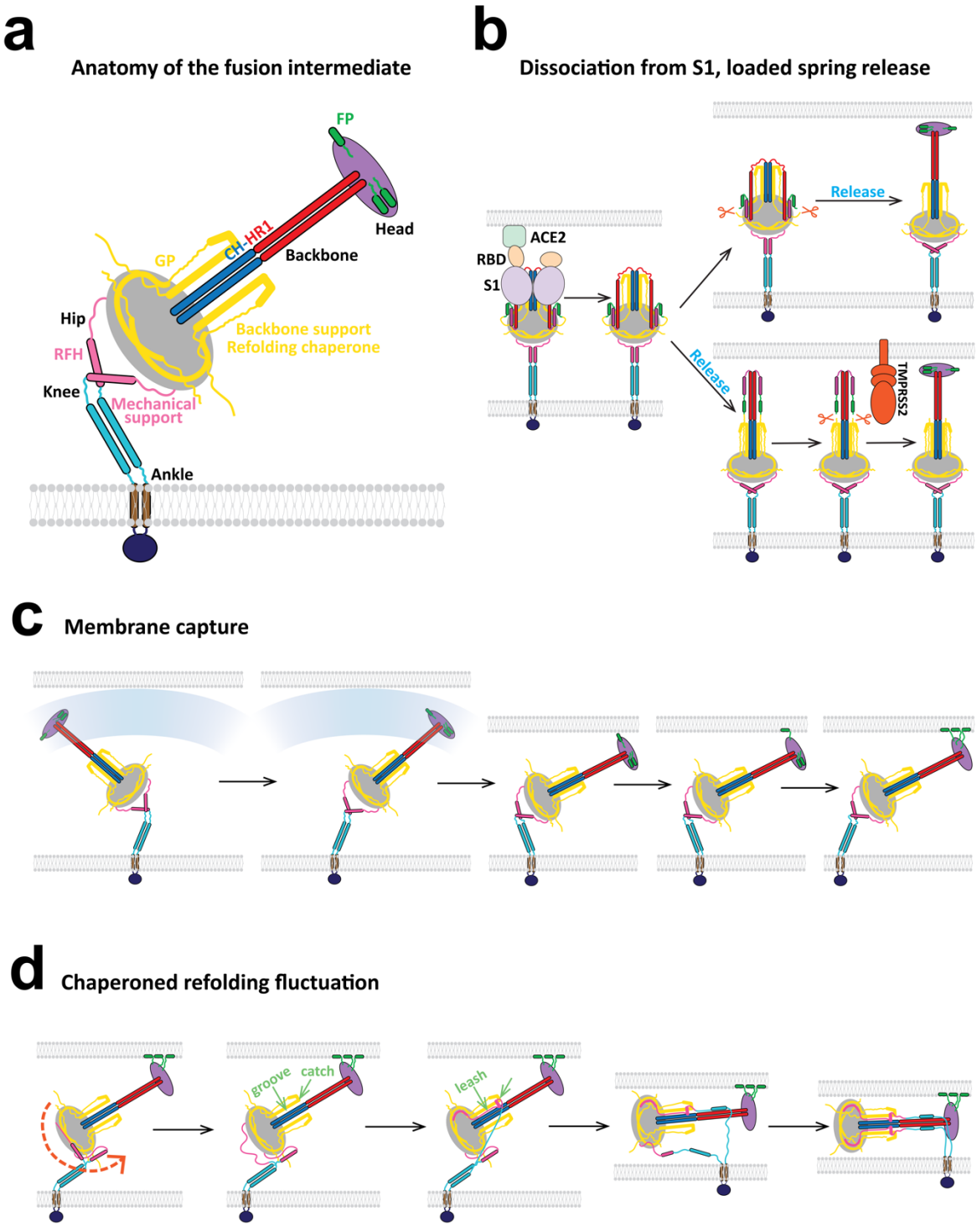


Figure 7. Model of the SARS-CoV-2 fusion intermediate and the pathway to fusion
(a) Schematic of the fusion intermediate. The ankle, knee and hip hinges impart high flexibility to the FI. RFH is an inverted tripod suspension system buffering longitudinal backbone fluctuations. GP supports the backbone and chaperones refolding. The CH-HR1 backbone

provides mechanical strength and reach. The FP-CR head houses the fusion peptides for host membrane capture.

(b) Pathway to the fusion intermediate. Following dissociation of S2 from the S1/ACE2 complex, a loaded spring release mechanism generates the fusion intermediate after proteolytic cleavage at the S2′ site (upper pathway) or before cleavage (lower pathway). RBD, receptor binding domain of S1. TMPRSS2, transmembrane protease serine 2.

(c) Schematic of host cell membrane capture by the fusion intermediate. Three base hinges endow the fusion intermediate with high flexibility and large configurational fluctuations, such that the N-terminal fusion peptides sweep out a large volume for membrane capture.

(d) Model of fluctuation-triggered, GP-chaperoned refolding. A sufficiently large rotational fluctuation at the RFH/BH hip joint unfolds a RFH helix into an unstructured loop. The loop is grabbed by a GP β-strand in BH, initiating RFH refolding, and guided into a GP-CH groove which it packs as a leash. Leash zippering into the groove is stabilized by the GP catch, preventing unzippering. Refolding of the HR2 leash completes the refolding process for one protomer, pulling the membranes together and helping the other protomers refold. The trans postfusion structure catalyzes membrane fusion in cooperation with other refolded fusion proteins.

Supplementary Information

Large fluctuations of the fusion intermediate help SARS-CoV-2 capture host cell membranes

Rui Su[1], Jin Zeng[1], Ben O'Shaughnessy[1,*]

[1]Department of Chemical Engineering, Columbia University, New York, NY, USA

[*]Correspondence should be addressed to B.O'S (email: bo8@columbia.edu)

Supplementary Methods

This section provides more detailed information for each subsection in "Methods".

All-atom simulation of the fusion intermediate.
The full-length fusion intermediate (FI, Fig. 1b) with its TMD inserted in the membrane was placed in a simulation box of $16 \times 16 \times 43\ \mathrm{nm}^3$. The membrane-protein system was built using the CHARMM-GUI membrane builder[1], consisting of 786 lipids for the pure DPPC membrane. The termini and ionizable residues were treated in their charged states assuming neutral pH. The disulfide bond in FP was added according to the prefusion structure (PDB: 6XR8), while the disulfide bonds in the other domains, BH and GP, were added based on the postfusion structure (PDB: 6XRA, the disulfide bonds were conserved in the pre- and postfusion structures). The resulting simulation box contained approximately 300,000 water molecules and was neutralized with 150 mM NaCl ions. The TIP3P model was used for water[2].

The system was first energy-minimized for 1,000 steps. Then, 2 equilibrations in the NVT ensemble were each performed for 0.1 ns with position restraints on all protein atoms. Subsequently, 4 equilibrations in the NPT ensembles were each performed for 0.5 ns with position restraints on protein heavy atoms. The production simulation was run for 305 ns in the NPT ensemble at 1 bar and 310 K and with a time step of 2 fs. The temperature and pressure were maintained using the Nose´-Hoover thermostat[3,4] and Parinello-Rahman barostat[5], respectively. All the energy-minimization, equilibration, and production simulations were performed using GROMACS 2019.6[6,7]. The secondary structure of each residue in the FI was analyzed using the *dssp* algorithm[8,9] for every 0.1 ns.

Coarse-grained simulations of the fusion intermediate.
Atomistic coordinates of the modeled full length fusion intermediate (Fig. 2a) were converted onto the secondary structure based coarse-grained MARTINI 2.2 topology[10,11] using the *martinize* utility. The termini and ionizable residues were treated in their charged states assuming neutral pH. Disulfide bonds were added to the same residues as in the all-atom simulation. The simulation box of $30 \times 30 \times 50\ \mathrm{nm}^3$ was generated using the *insane* utility[12], in which the coarse-grained (CG) FI with its TMD inserted in a crystalline DPPC bilayer consisting of 3,024 coarse-grained lipids that represents the viral envelope.

The system was first energy-minimized for 2,000 steps in vacuum. Subsequently, the box was solvated with approximately 300,000 coarse-grained water particles and neutralized with 150 mM NaCl ions. The system was then energy-minimized for 2,000 steps and equilibrated for 4 ns in NPT ensemble sampling. Then the system was subjected to 5 independent production simulations, each lasting 8 μs in the NPT ensemble. Equations of motion were integrated using the Verlet leapfrog algorithm with a 80 fs time step. Bonds were constrained with the LINCS algorithm. The system temperature (310K) was maintained by the velocity rescale thermostat[13]. The membrane tension (0.05 pN/nm) and the system pressure (1 bar) were maintained by the Berendsen barostat with surface-tension coupling[14].

Fitting a curve to the fusion intermediate ectodomain
To fit a curve to represent the FI ectodomain backbone (residues 912-1237) in the MARTINI simulations of the FI (Figs. 3 and S2), we first represented the backbone by points. Each point represented a residue and its position was calculated by the averaged position of the MARTINI

backbone bead (the non-sidechain bead) of each residue among the three protomers. The coordinates of the beads were extracted from the simulation trajectory using the *mdtraj*[15] python package. Then the FI backbone was divided into an upper part (residues 912-1191) and a lower part (1152-1237) with an overlapping region of 40 residues. Each part was separately smoothed by the following steps. (1) All points (x_i, y_i, z_i) were calibrated so that the center of gravity (COG) was at the origin, (2) All points were rotated so that the new x,y,z axis are aligned with the eigenvectors of the gyration tensor of the rotated points (x_i', y_i', z_i'). This rotation maximized the root mean square projected length onto the z axis. (3) A smoothed curve $(\overline{x_i'}, \overline{y_i'}, z_i')$ was generated by the LOWESS algorithm in python, in which the smoothed $\overline{x_i'}$ and $\overline{y_i'}$ value for each z_i' was set by its neighboring points spanning one tenth of the entire z range. Then the smoothed upper/lower part was rotated back to the original orientation and the overlapping region was averaged between the upper/lower parts. Finally the reconnected points were smoothed by the B-spline method with 4^{th} order polynomial functions. To find the location (from the N- to the C-terminus) of each domain on the fitted the curve, the curve was first reparametrized by its normalized arclength. The location of each residue was determined as the normalized arclength of the nearest point on the fitted curve to the position of this residue averaged over the three protomers, using only the backbone bead to locate the residues.

The isolated fusion peptide binding assay.
The atomistic structure of the fusion peptide (FP, residue 816-855) was extracted from the crystal structure of the prefusion structure (PDB: 6XR8). Then the atomistic structure of the FP was coarse-grained into the MARTINI representation. The ionizable residues were treated in their charged states assuming neutral pH. The C-terminal carboxyl group at residue 855 was neutralized by changing the type of the backbone bead from Qa to Na and changing the backbone bead charge from -1 to 0 in the itp topology file, as the residue 855 connects CR in the full-length FI. The disulfide bond was added according to the solved prefusion structure (PDB: 6XR8). The coarse-grained FP was placed approximately 1 nm above a crystalline DPPC bilayer consisting of 162 coarse-grained lipids in a $7 \times 7 \times 10\ \mathrm{nm}^3$ box using the *insane* utility[12]. By implementing periodic boundary condition, this is equivalent to placing a FP between two planar membranes separated by ~5.5 nm.

The system was first energy-minimized for 500 steps in vacuum. Subsequently, the box was solvated with approximately 2,000 coarse-grained water particles and neutralized with 150 mM NaCl ions. The system was then energy-minimized for 500 steps and equilibrated for 4 ns in the NPT ensemble with a 80 fs time step. Then the system was subjected to 10 independent production simulations, each lasting 24 μs in the NPT ensemble. The system temperature (310K) was maintained by the velocity rescale thermostat[13]. The membrane tension (0.05 pN/nm) and the system pressure (1 bar) were maintained by the Berendsen barostat with surface-tension coupling[14].

All-atom simulation of a membrane-bound fusion peptide.
In one of the CG simulations in the FP-only binding assay, the final configuration of the FP and the membrane to which it was bound is converted to atomic resolution in CHARMM36 force field[16,17] using the *backward* utility[18]. The ionizable residues are treated in their charged states assuming neutral pH. The C-terminal carboxyl group is neutralized. The Disulfide bond is added to the same residues as in the FP-only binding assay. The resulting simulation box contained approximately 9,000 water molecules and was neutralized with 150 mM NaCl ions. The TIP3P model was used for water[2].

4 equilibrations in the NVT and NPT ensembles were each performed for 50 ns with position restraints on peptide heavy atoms. The production simulation was run for 2 μs in NPT ensembles at 1 bar and 310 K and with a time step of 2 fs. The temperature and pressure were maintained using the Nose´-Hoover thermostat[3,4] and Parinello-Rahman barostat with isotropic coupling[5], respectively. The secondary structure of each residue in the FP was analyzed using the *dssp* algorithm[8,9] for every 0.1 ns.

Coarse-grained simulations of an equilibrated fusion peptide bound to a membrane.
In the AA simulation of a membrane-bounded FP simulation, the final configuration of the FP and the membrane to which it was bound was converted to MARTINI CG representation using the *martinize* utility. The ionizable residues were treated in their charged state. The C-terminal carboxyl group was neutralized.

The system was first energy-minimized for 1,000 steps in the vacuum. Subsequently, the box was solvated with water particles and ions to attain a salt concentration of 0.15 M. The system was then energy-minimized for 5,000 steps and equilibrated for 0.8 ns in the NPT ensemble. Then the system was subjected to the production simulation for 80 μs in the NPT ensemble. The system temperature (310K) and pressure (1 bar) were maintained by the velocity rescale thermostat and Parinello-Rahman barostat, respectively.

Coarse-grained simulations of the fusion intermediate interacting with a target membrane.
In the unbiased set of simulations, ten snapshots from three FI simulations (one membrane) were chosen as the initial condition in which the FI ectodomain protruded ~20 nm normal to the membrane. Another pre-equilibrated planar membrane (run for 4 μs) was placed 20 nm above the membrane anchoring the FI in the selected configurations. Each configuration containing an FI and two membranes was re-solvated by approximately 200,000 coarse-grained water particles and neutralized by 150 mM NaCl ions. Then, each of the 10 systems was equilibrated for 4 ns, and subject to a production simulation for 8 μs in the NPT ensemble. The system temperature (310K) was maintained by the velocity rescale thermostat. The membrane tension (0.05 pN/nm) of the two membranes and the system pressure (1 bar) was maintained by the Berendsen barostat with surface-tension coupling.
In the biased set of simulations, all ten runs started from the same initial condition, in which the COG of the nearest N-terminal FP helix was within 1 nm of the adjacent membrane and the FI head was in the (1,2) configuration. Here, the membrane position was defined to be the mean location of all the PO4 beads in the lower leaflet of the upper membrane. Each production simulation lasted 22.4 μs with the same system temperature, pressure and the membrane tension.

Pulling the fusion peptide into the target membrane.
An initial condition was chosen with the nearest N-terminal FP helix lying within 1 nm of the membrane and with the FI head in the (1,2) configuration. The FP N-terminal helix was pulled vertically towards the upper membrane at a constant rate of 2.5 nm/μs by a harmonic potential with a force constant of 1,000 kJ mol^{-1} nm^{-2}. The system temperature, membrane tension, and system pressure were maintained at the same value using the same thermostat and barostat as mentioned above. The pulling process took ~1.2 μs. In the final configuration the distance between the COG of the FP N-terminal helix and the COG of the entire upper membrane reached ~0.1 nm, and the FP N-terminal helix was pulled into the adjacent membrane patch by ~1 nm. Then the force was released and a subsequent 8 μs CG simulation was run, in which the system temperature,

membrane tension, and system pressure were maintained at the same value using the same thermostat and barostat.

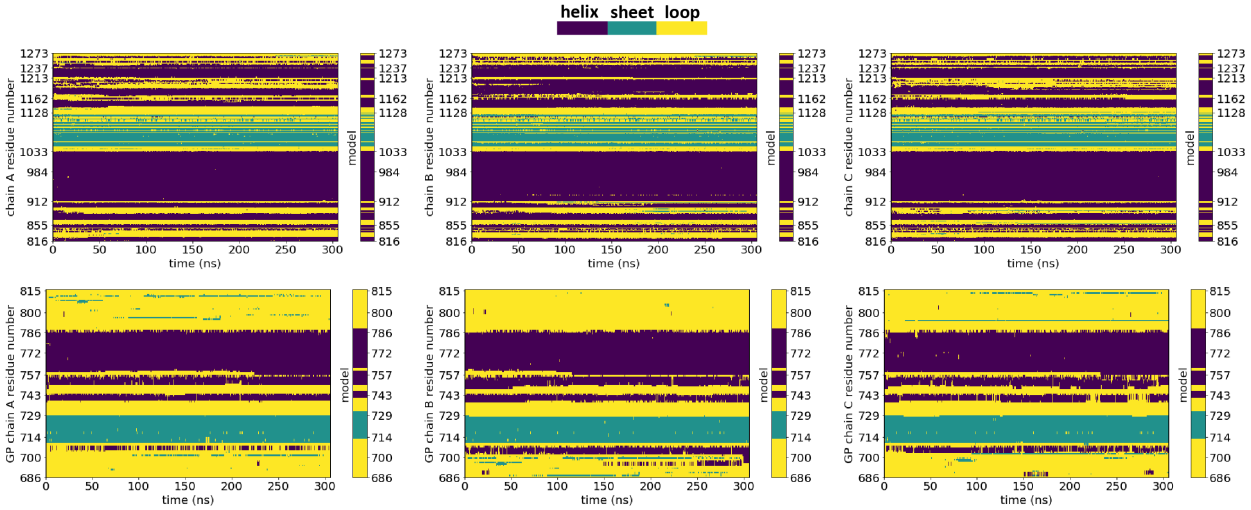
Coarse-grained simulation of a fusion intermediate with uncleaved S2′ sites.
Atomistic coordinates of the full-length fusion intermediate (Fig. 1b) were converted onto the MARTINI 2.2 topology using the *martinize* utility and placed in a simulation box of $40 \times 40 \times 50$ nm^3. The unsolved C-terminal part of the GP domain in the postfusion structure[19] (residues 771-815) was forced to be a loop, by changing the input secondary structural file to the *martinize* utility. The terminal and ionizable residues were treated in their charged states assuming neutral pH. The box was then solvated by approximately 600,000 coarse-grained water particles and was neutralized by 150 mM NaCl ions.

In each protomer, the FP N-terminus residue (residue 816) and the GP C-terminus residue (residue 815) were pulled together at a constant rate of 10 nm/μs by a harmonic potential with a force constant of 500 kJ mol^{-1} nm^{-2}. Position constrains by a harmonic potential with force constant of $1{,}000$ kJ mol^{-1} nm^{-2} were applied to the beads in the domains other than GP, FP and CR. The pulling simulation took ~1.5 μs so that in the final configuration, the COG distance between residues 815 and 816 in one of the three protomers reaches ~0.5 nm.

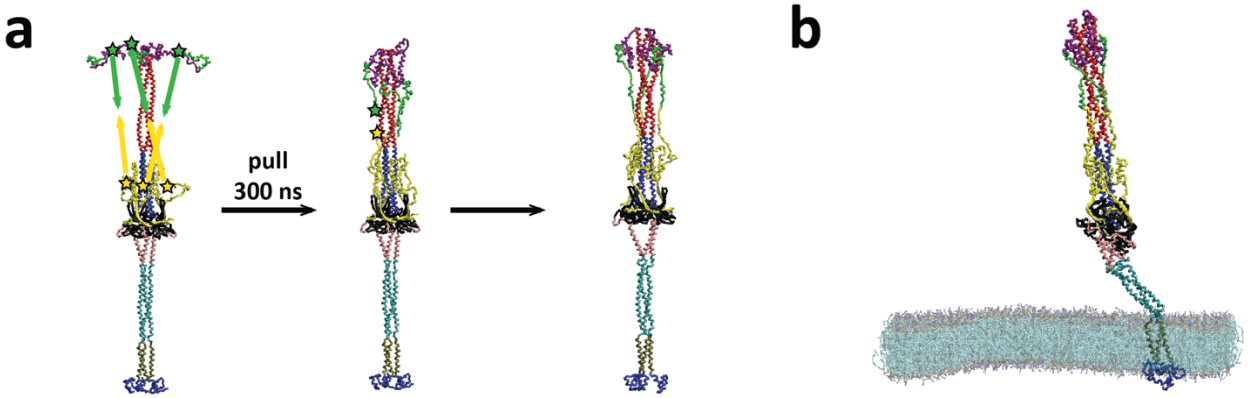
The final coordinates of the coarse-grained FI were then backmapped into atomistic resolution in CHARMM36 force field[16,17]. The residues 815 and 816 were covalently connected in the protomer with the smallest separation between the two residues, and this protomer was duplicated twice to make a trimer using Pymol. Now a structure of the FI with its S2′ sites uncleaved was created. Then, the all-atom structure of the FI with its S2′ sites uncleaved was converted into MARTINI coarse-grained representation using the *martinize* utility. A simulation box of $30 \times 30 \times 50$ nm^3 was generated using the *insane* utility, in which the coarse-grained FI with its S2′ sites uncleaved was inserted in a crystalline DPPC bilayer consisting of 2,831 coarse-grained lipids that represented the viral envelope.

The system containing the coarse-grained FI with its S2′ sites uncleaved on the viral envelope was first energy-minimized for 2,000 steps in the vacuum. Subsequently, the box was solvated with approximately 300,000 coarse-grained water particles and neutralized with 150 mM NaCl ions. The system was then energy-minimized for 2,000 steps and equilibrated for 4 ns in NPT ensemble sampling. Then the system was subjected to a production simulation lasting 4 μs in the NPT ensemble. The system temperature, membrane tension, and system pressure were maintained at the same value using the same thermostat and barostat as mentioned above.

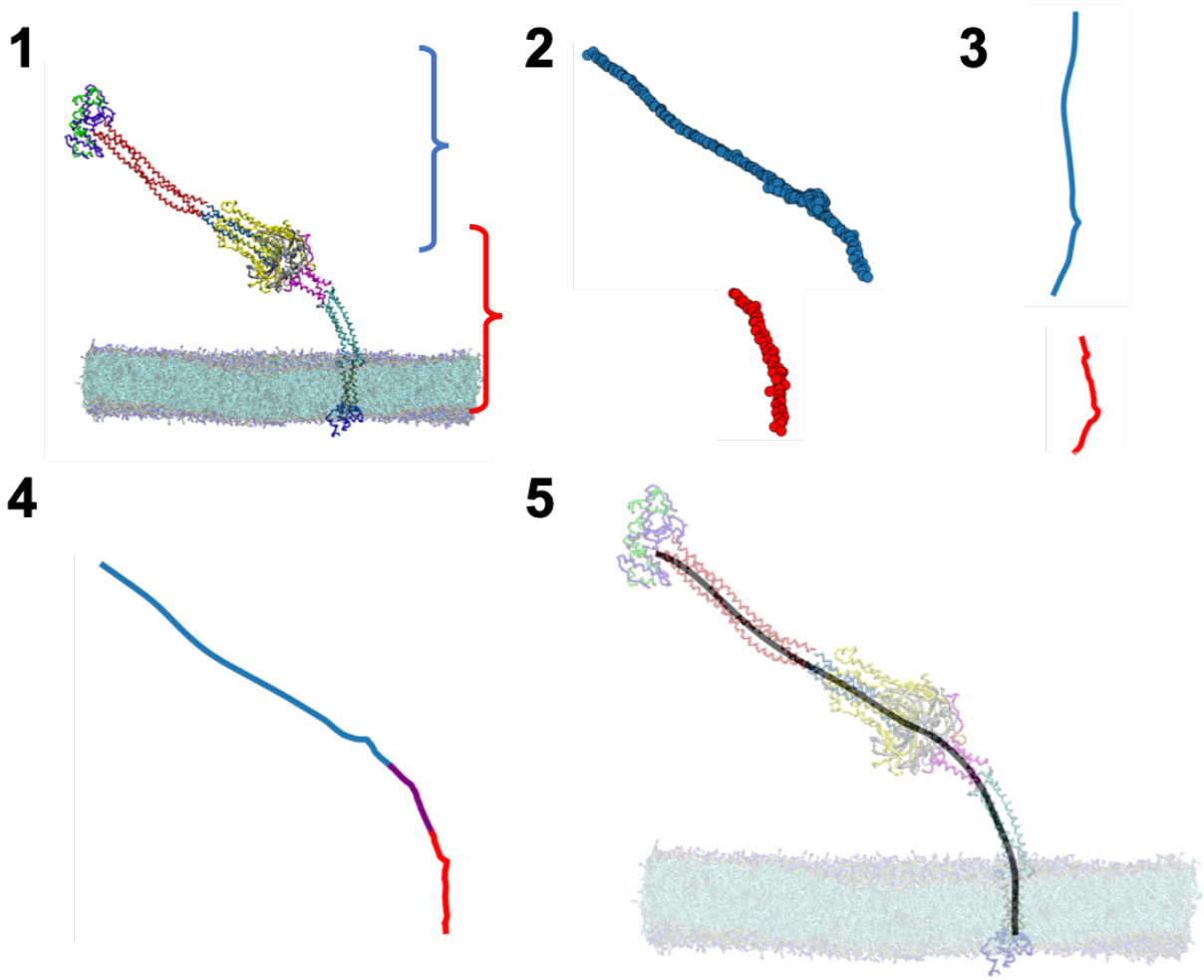
Supplementary Figures



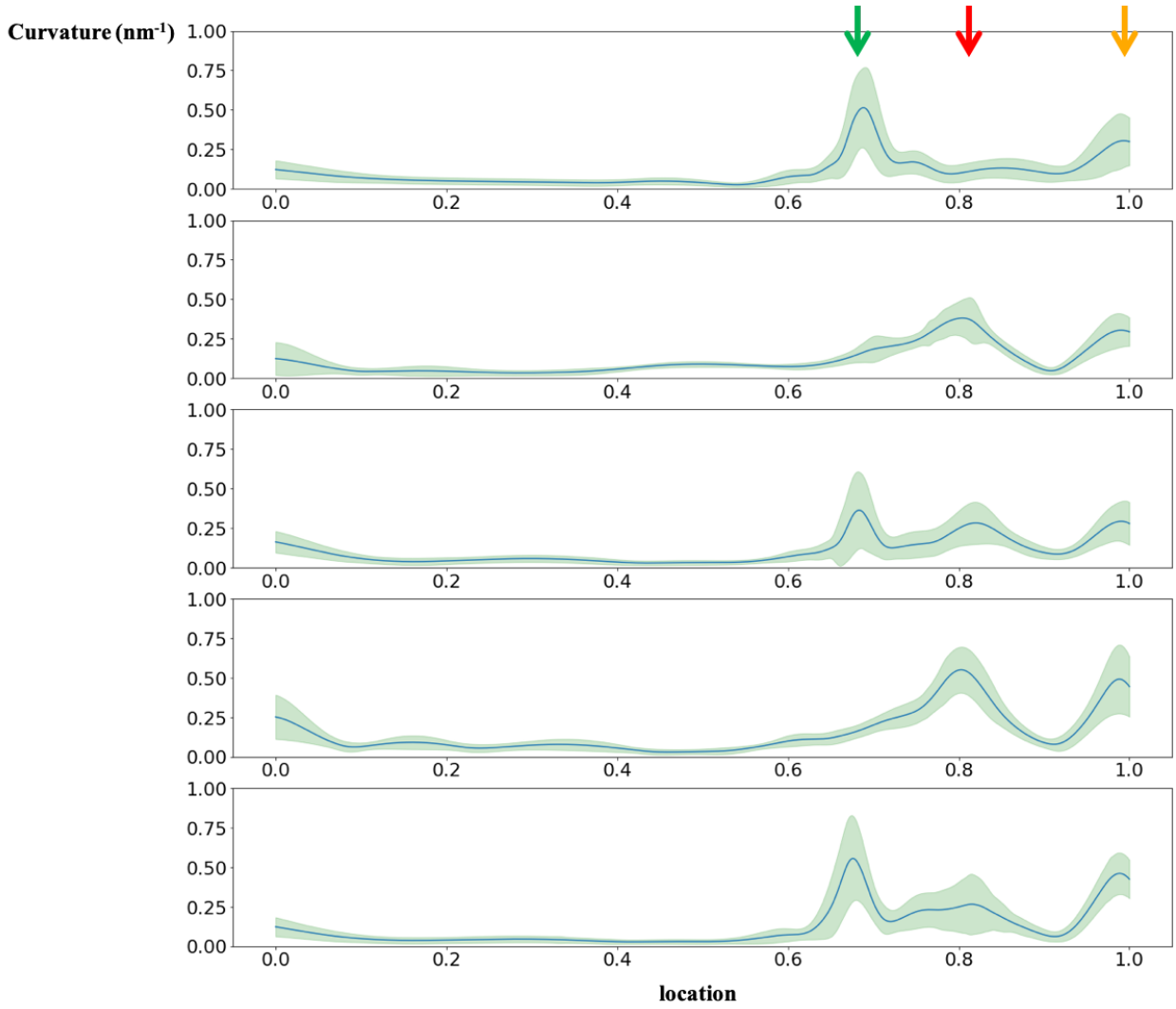
Supplementary Figure 1. Evolution of the secondary structure of the fusion intermediate during 305 ns of all-atom simulation.
The secondary structure of each residue was measured every 0.1 ns during the AA simulation (Fig. 2). Each panel refers to one protomer of either the main body of the S2 subunit (top row) or the cleaved GP (bottom row). The secondary structure of the fusion intermediate model of Figure 1b is shown to the right of each panel for comparison.



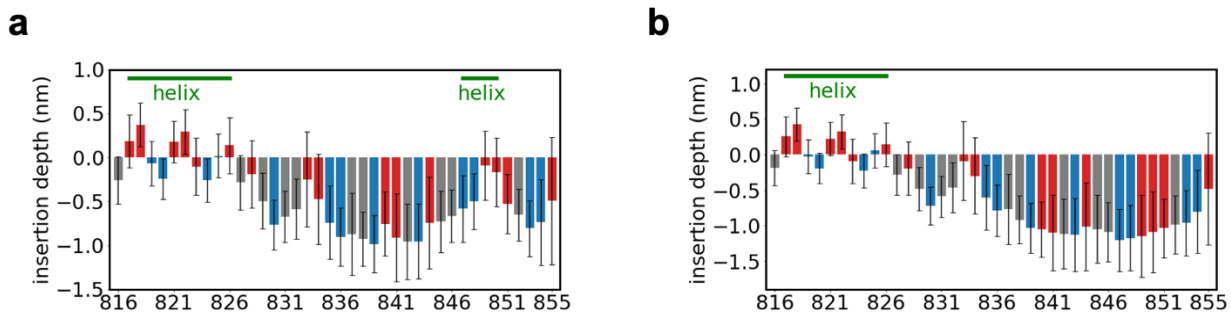
Supplementary Figure 2. Simulation of the fusion intermediate with uncleaved S2′ sites.
(a) Construction procedure for a fusion intermediate with uncleaved S2′ sites. Starting from the model structure in Fig. 2b, the C-terminus of three GPs and the N-terminus of three FPs were pulled together in ~1.5 μs in a MARTINI CG simulation. Then the C-terminus of GP and the N-terminus of FP in one protomer were connected covalently (stars). This protomer was duplicated twice to generate a FI homotrimer with uncleaved S2′ sites.
(b) Snapshot from a 4 μs simulation. The fusion peptides (green) are sequestered. The FI with uncleaved S2′ sites exhibits similar flexibility to the normally cleaved FI.



Supplementary Figure 3. Fitting a curve to represent the fusion intermediate ectodomain. (1) The coordinate of each coarse-grained bead is extracted by Python package *mdtraj*. (2) Each residue is represented by one point, averaged over three protomers. The upper part of the FI (blue, residues 912-1191) and the lower part (red, residues 1152-1237) are separated. (3) The upper and the lower parts are rotated so that the three principal axes are aligned with three eigenvectors of the gyration tensor of the rotated points. The rotated points are smoothed by the LOWESS algorithm. (4) The upper and lower parts are rotated back to their original orientations and reconnected. The overlapping region (purple, residues 1152-1191) is averaged over the upper and lower parts. (5) The final curve is fitted by the B-spline method.

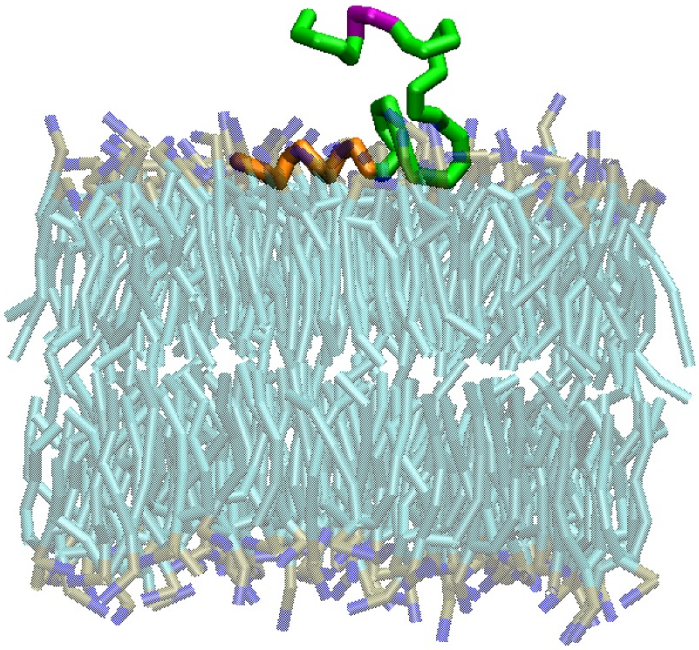


Supplementary Figure 4. Time averaged backbone curvature versus normalized backbone arclength for five parallel runs. The three hinges (Fig. 3b, arrows) show different magnitudes of curvature in the five parallel runs. Green envelope indicates SD.



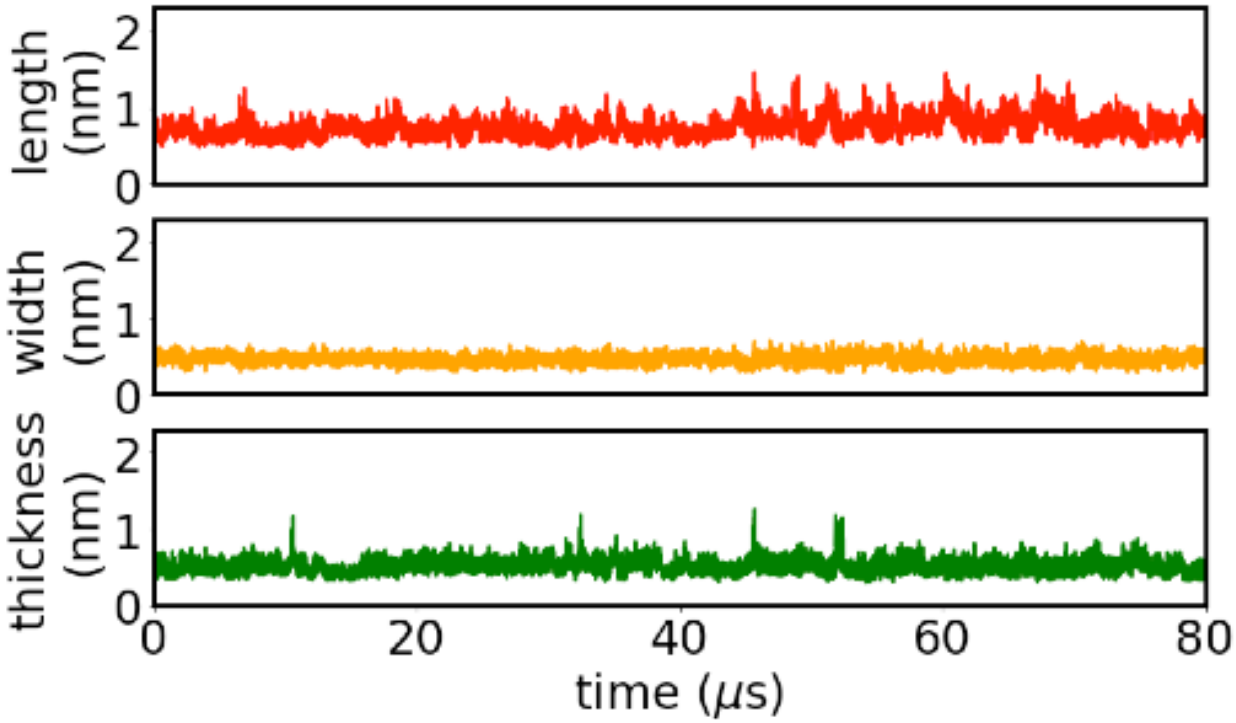
Supplementary Figure 5. Fusion peptide residue insertion depth measured in MARTINI coarse-grained simulations.

The insertion depth is defined as the vertical distance between the center of gravity (COG) of each fusion peptide residue and the COG of the PO_4 beads in the leaflet to which the fusion peptide is bound (same definition as for the all-atom simulation). The secondary structure used is **(a)** the equilibrated secondary structure from the all-atom simulation, and **(b)** the equilibrated secondary structure but imposing the C-terminal helix to be a loop. The insertion depths are averaged over the last 78 μs. Error bars: SD over the same time frame. Bars were colored by the hydrophobicity of the corresponding residues (red: hydrophobic, grey: neutral, blue: hydrophilic), using the same color scheme as in Fig. 4b.

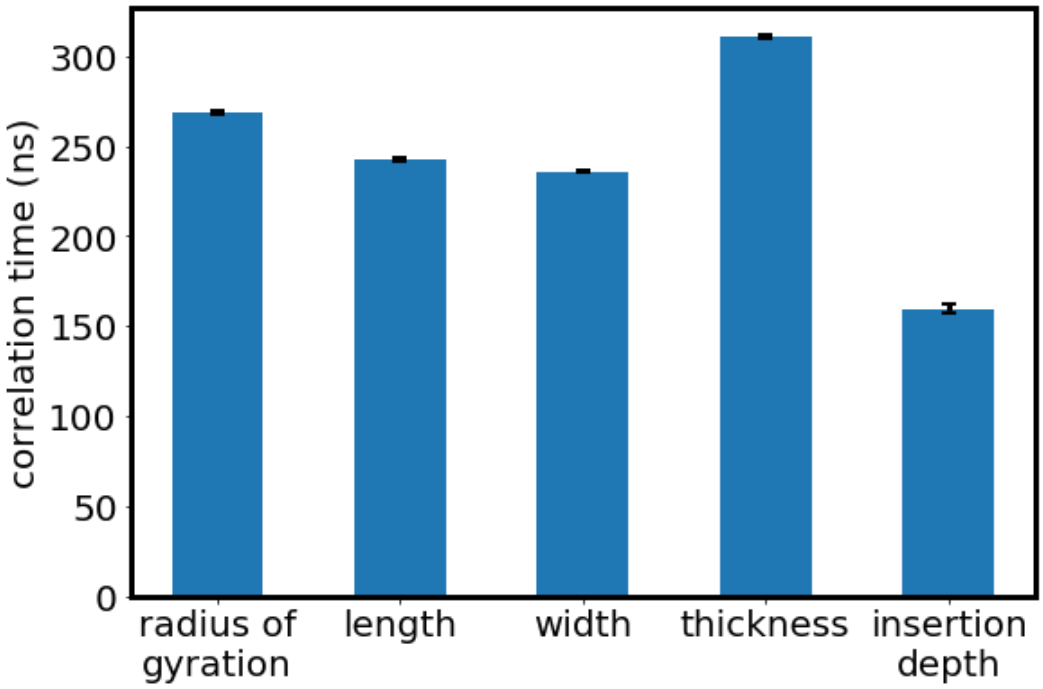


Supplementary Figure 6. MARTINI simulation snapshot showing a transient detachment of the fusion peptide C-terminal helix

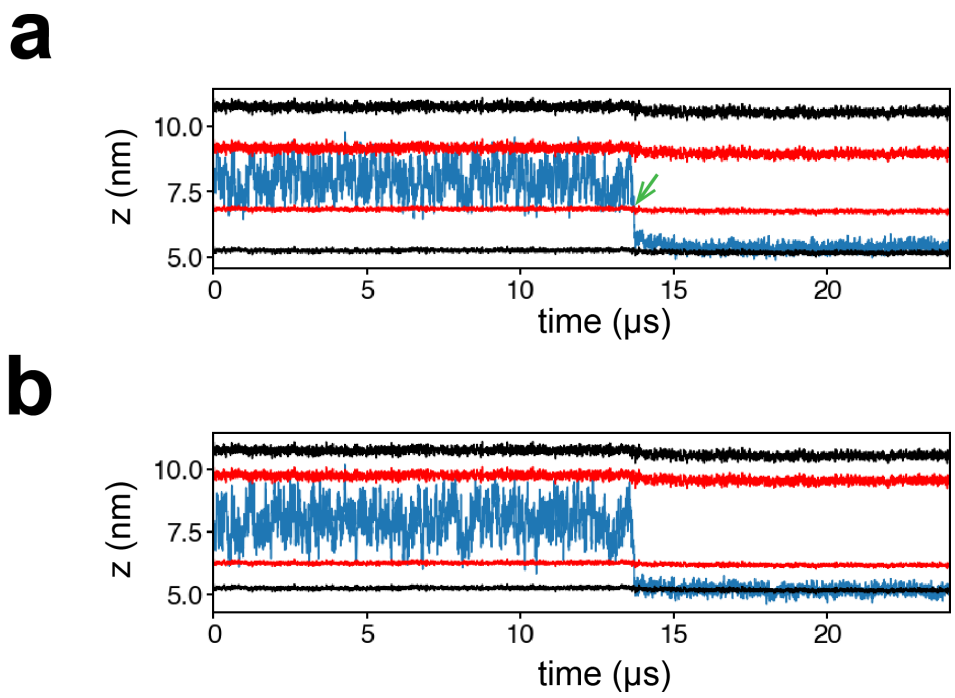
The fusion peptide C-terminal helix (purple) unanchored transiently from the membrane for ~0.3 μs (the instance depicted here occurred after ~52 μs of the simulation). In contrast, the N-terminal helix (orange) always remained buried in the membrane.



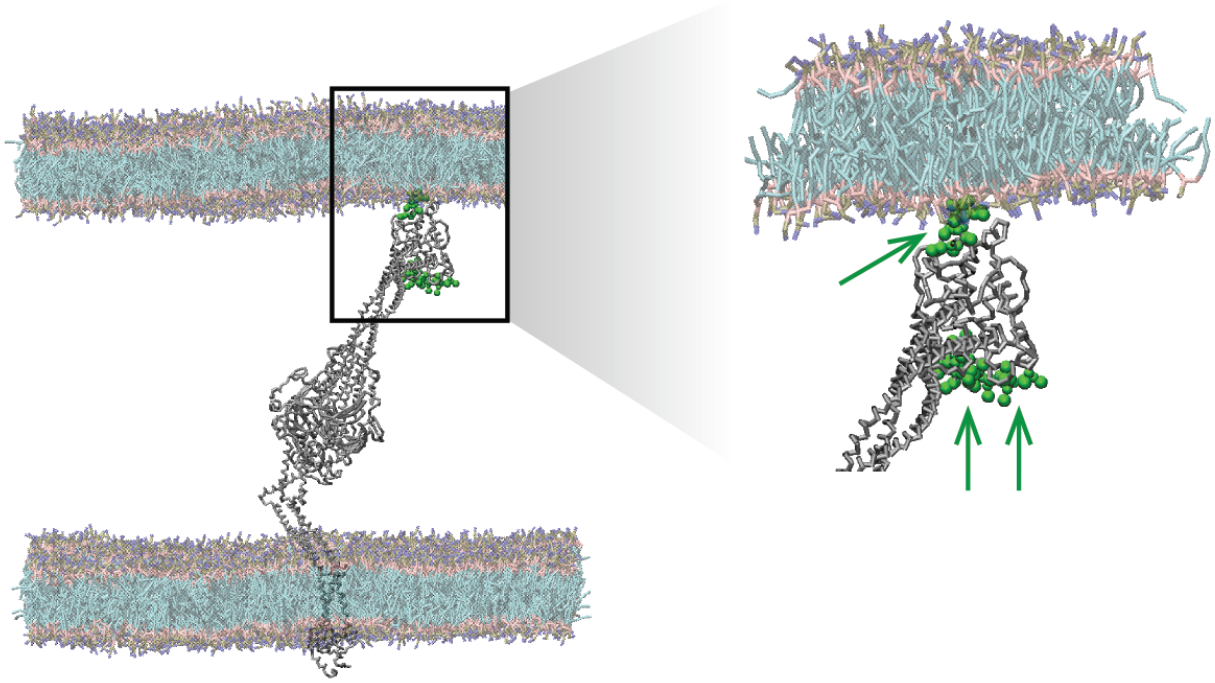
Supplementary Figure 7. Measured length, width and thickness versus time of the fusion peptide during the MARTINI coarse-grained simulation of an equilibrated fusion peptide bound to a membrane.



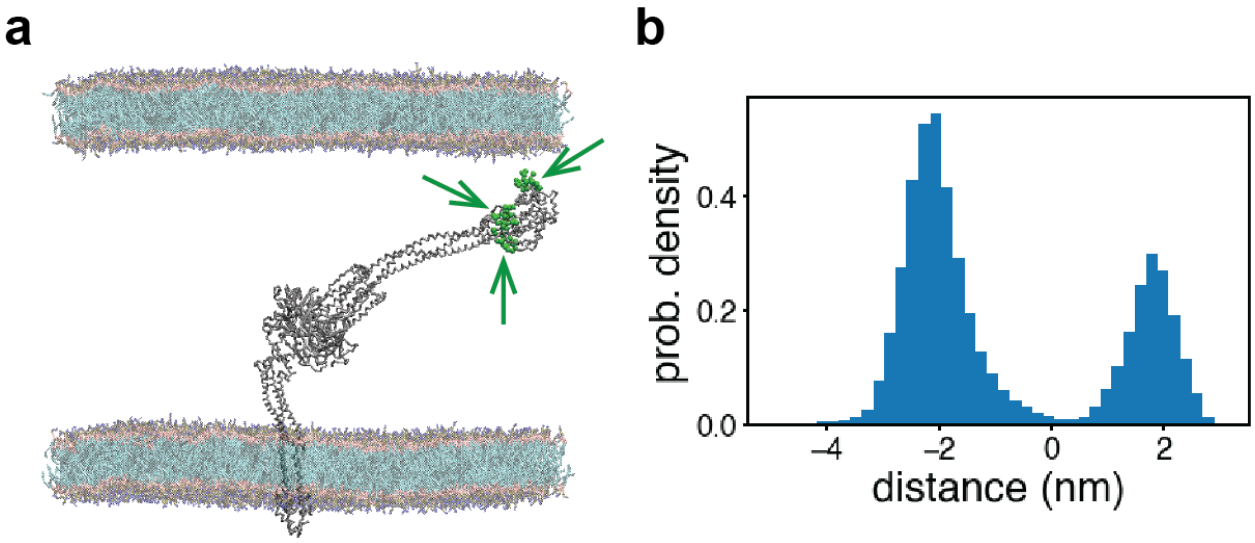
Supplementary Figure 8. Correlation times of the fusion peptide shape properties during MARTINI coarse-grained simulation of an equilibrated fusion peptide bound to a membrane.
The radius of gyration, length, width, thickness and insertion depth of the fusion peptide are calculated for the last 78 µs of the MARTINI simulation. The radius of gyration, length, width and thickness are calculated from the gyration tensor (see "Methods"). The insertion depth of the entire fusion peptide is defined as the vertical distance between the center of gravity of the fusion peptide and the center of gravity of the PO_4 beads in the leaflet to which the FP is bound. The correlation time is calculated in the same way as in Fig. 4g. Error bars: 95% confidence interval.



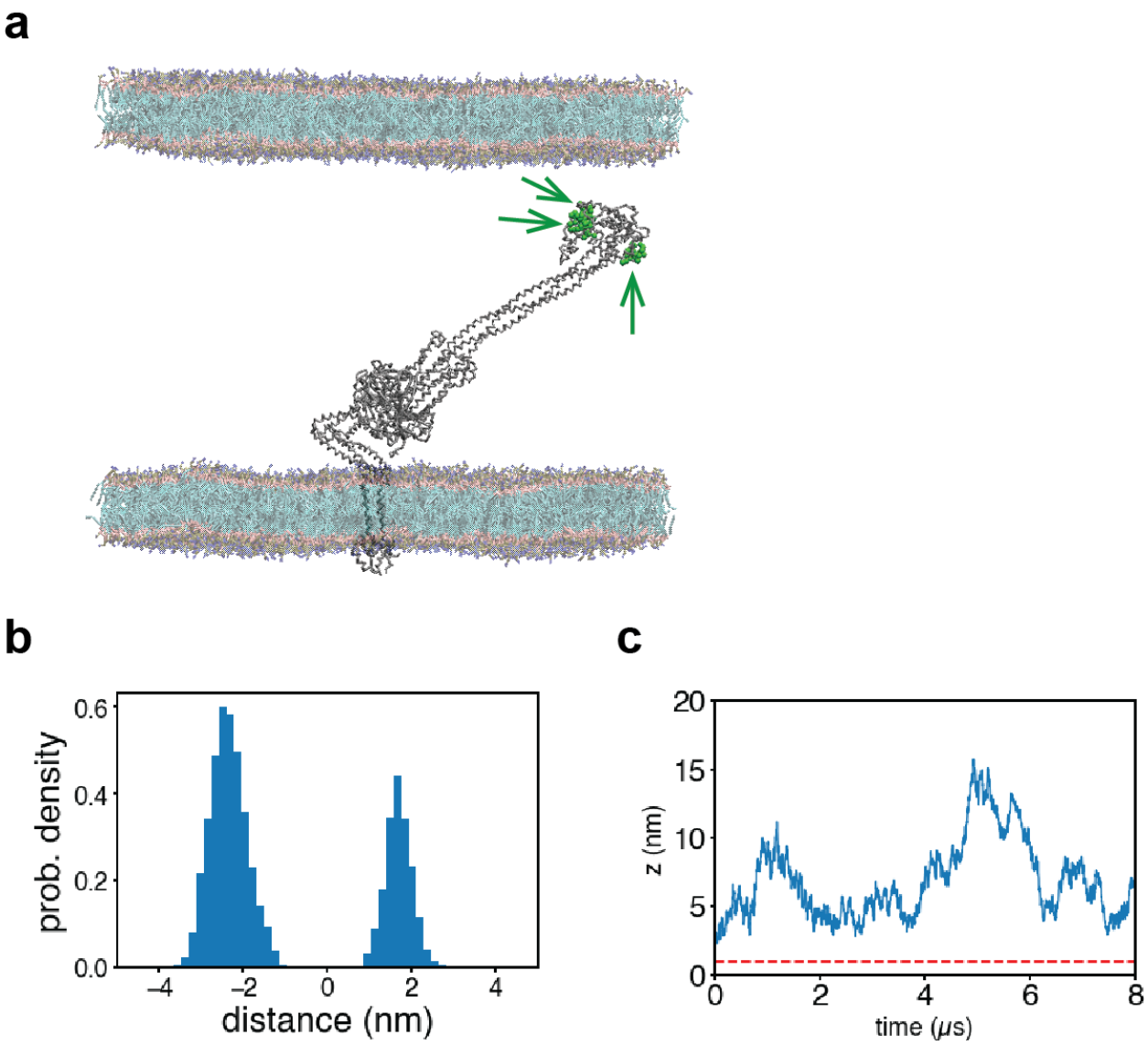
Supplementary Figure 9. Evolution of fusion peptide vertical position during a binding assay of an isolated fusion peptide.
(a) Evolution of the center of gravity of the fusion peptide (blue). The vertical positions of the two leaflets to which the FP is likely to bind are calculated by averaging the PO_4 bead positions in each leaflet (black). A collision event is defined as an approach to the membrane to within $R_{\mathrm{FP}} \sim 1.6$ nm of either leaflet (black), where R_{FP} is the rms FP end-to-end distance. A binding event is defined to be when the FP center of gravity first has a value that positioned it below the upper membrane leaflet and above the lower membrane leaflet (green arrow).
(b) Evolution of the center of gravity of the fusion peptide N-terminal helix (blue). The positions of the two leaflets are defined in the same way as for (a). Before the binding event, the N-terminal helix approached several times to within 1 nm of either leaflet (red).



Supplementary Figure 10. Snapshot of a fusion intermediate interacting with the target membrane with one fusion peptide approaching the target membrane to within 1 nm.
The fusion peptide N-terminal helix beads are shown as green spheres. A zoom-in of the three fusion peptides (green arrows) and the membrane is shown at right. The center of gravity of the closest fusion peptide N-terminal helix is within 1 nm of the membrane, whose position is defined as the average position of the PO_4 beads in the lower leaflet.



Supplementary Figure 11. Fusion intermediate simulation with a small separation between the fusion peptides in the head.
(a) A simulation snapshot with the fusion peptide N-terminal helix beads labeled as green spheres. **(b)** The density distribution of the helix beads (defined in the same way as for Figs. 6e,f) reveals a smaller separation between the two peaks (~4 nm) than for Fig. 6e.



Supplementary Figure 12. Fusion intermediate simulation with head in the (2,1) configuration.
(a) Simulation snapshot with the fusion peptide N-terminal helix beads shown as green spheres. A merged fusion peptide N-terminal helix pair lies close to the membrane, while the unpaired N-terminal helix is distal from the membrane.
(b) Density distribution of the helix beads.
(c) Time evolution of the distance between the closest fusion peptide N-terminal helix and the target membrane. Distance is defined as for Fig. 6b. Although the fusion peptides frequently lie close to the membrane, the N-terminal helix fails to approach the target membrane to within 1 nm (red dashed line).

Legends for Supplementary Movies

Supplementary Movie 1. All-atom simulation of the SARS-CoV-2 fusion intermediate. The color code is as for Fig. 2.

Supplementary Movie 2. Coarse-grained MARTINI simulation of the SARS-CoV-2 fusion intermediate. The three hinges are colored according to the code of Fig. 3c.

Supplementary Movie 3. Simulation of Movie 2, but now including a fitted curve (red) to the ectodomain backbone of the fusion intermediate (blue) whose TMD is anchored in the viral envelope (green).

Supplementary Movie 4. All-atom simulation of a membrane-bound fusion peptide (side view). The residues are colored according to their hydrophobicity, using the color scheme of Fig. 4b.

Supplementary Movie 5. Simulation of Movie 4, top view.

File name: Supplementary Movie 6
Description: Coarse-grained MARTINI simulation of an equilibrated fusion peptide bound to a membrane (side view). The N-terminal helix (orange) was always buried in the membrane, but after ~52 μs of the simulation the C-terminal helix (purple) transiently unanchored from the membrane for ~0.3 μs.

Supplementary Movie 7. Simulation of Movie 6, top view.

Supplementary Movie 8. Binding assay for an isolated fusion peptide (coarse-grained MARTINI simulation). The fusion peptide, removed from the FI, becomes bound to the membrane after ~1.6 μs, with the N-terminal helix (orange) providing first stable contact. The fusion peptide remains bound for the remaining ~22.4 μs of the simulation. The binding events of Fig. 5b are snapshots from this movie.

Supplementary Movie 9. Simulation of the fusion intermediate interacting with a target membrane, 8 μs duration. One of the fusion peptide N-terminal helices (green spheres) repeatedly approaches the membrane to within 1 nm, but fails to bind. Fig. 6e shows a snapshot from this movie.

Supplementary References

1 Jo, S., Kim, T., Iyer, V. G. & Im, W. J. J. o. c. c. CHARMM-GUI: a web-based graphical user interface for CHARMM. **29**, 1859-1865 (2008).

2 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. J. T. J. o. c. p. Comparison of simple potential functions for simulating liquid water. **79**, 926-935 (1983).

3 Hoover, W. G. J. P. r. A. Canonical dynamics: Equilibrium phase-space distributions. **31**, 1695 (1985).

4 Nosé, S. J. T. J. o. c. p. A unified formulation of the constant temperature molecular dynamics methods. **81**, 511-519 (1984).

5 Parrinello, M. & Rahman, A. J. J. o. A. p. Polymorphic transitions in single crystals: A new molecular dynamics method. **52**, 7182-7190 (1981).

6 Bekker, H. *et al.* in *Physics computing.* 252-256 (World Scientific Singapore).

7 Berendsen, H. J., van der Spoel, D. & van Drunen, R. J. C. p. c. GROMACS: a message-passing parallel molecular dynamics implementation. **91**, 43-56 (1995).

8 Kabsch, W. & Sander, C. J. B. O. R. o. B. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **22**, 2577-2637 (1983).

9 Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. **39**, D411-D419 (2010).

10 de Jong, D. H. *et al.* Improved parameters for the martini coarse-grained protein force field. **9**, 687-697 (2013).

11 Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. J. T. j. o. p. c. B. The MARTINI force field: coarse grained model for biomolecular simulations. **111**, 7812-7824 (2007).

12 Wassenaar, T. A. *et al.* Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. **11**, 2144-2155 (2015).

13 Bussi, G., Donadio, D. & Parrinello, M. J. T. J. o. c. p. Canonical sampling through velocity rescaling. **126**, 014101 (2007).

14 Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A. & Haak, J. R. J. T. J. o. c. p. Molecular dynamics with coupling to an external bath. **81**, 3684-3690 (1984).

15 McGibbon, R. T. *et al.* MDTraj: a modern open library for the analysis of molecular dynamics trajectories. **109**, 1528-1532 (2015).

16 Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. **8**, 3257-3273 (2012).

17 Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. **114**, 7830-7843 (2010).

18 Wassenaar, T. A. *et al.* Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. **10**, 676-690 (2014).

19 Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. **369**, 1586-1592 (2020).