Structures of the Foamy virus fusion protein reveal an unexpected link with the F protein of paramyxo- and pneumoviruses

Ignacio Fernández, François Bontems, Delphine Brun, Youna Coquin, Casper A Goverde, Bruno E Correia, Antoine Gessain, Florence Buseyne, Felix A Rey, Marija Backovic

STRUCTURAL BIOLOGY

Structures of the Foamy virus fusion protein reveal an unexpected link with the F protein of paramyxo- and pneumoviruses

Ignacio Fernández[1]*, François Bontems[1,2], Delphine Brun[1], Youna Coquin[3], Casper A. Goverde[4], Bruno E. Correia[4], Antoine Gessain[3], Florence Buseyne[3], Felix A. Rey[1], Marija Backovic[1]*

Foamy viruses (FVs) constitute a subfamily of retroviruses. Their envelope (Env) glycoprotein drives the merger of viral and cellular membranes during entry into cells. The only available structures of retroviral Envs are those from human and simian immunodeficiency viruses from the subfamily of orthoretroviruses, which are only distantly related to the FVs. We report the cryo–electron microscopy structures of the FV Env ectodomain in the pre- and post-fusion states, which unexpectedly demonstrate structural similarity with the fusion protein (F) of paramyxo- and pneumoviruses, implying an evolutionary link between the viral fusogens. We describe the structural features that are unique to the FV Env and propose a mechanistic model for its conformational change, highlighting how the interplay of its structural elements could drive membrane fusion and viral entry. The structural knowledge on the FV Env now provides a framework for functional investigations, which can benefit the design of FV Env variants with improved features for use as gene therapy vectors.

INTRODUCTION

Foamy viruses (FVs) belong to the *Spumaretrovirinae* subfamily of the *Retroviridae* family (*1*). They are unconventional retroviruses because their replication cycle displays several unique features that relate them to *Hepadnaviridae* and distinguish them from the better-characterized members of the *Orthoretrovirinae* subfamily, such as HIV and the human T-lymphotropic viruses (*2*, *3*). FVs are ancient viruses estimated to have existed ~450 million years ago (*4*–*6*). This long coevolution with hosts might explain their seemingly nonpathogenic nature, which has been exploited for the development of FVs as vectors for gene therapy (*7*, *8*). FVs are prone to cross-species transmission, and humans can be persistently infected with simian FVs (*2*, *9*).

As in all enveloped viruses, entry of FVs into host cells requires the fusion of viral and cellular membranes, a reaction catalyzed by the viral envelope (Env) protein. Retroviral Env proteins belong to the class I of viral fusion proteins, whose members are also present in influenza viruses, filoviruses, paramyxoviruses, pneumoviruses, arenaviruses, and coronaviruses. Class I fusion proteins share similar principles of oligomerization, proteolytic priming, and activation [reviewed in (*10*, *11*)]. The general paradigm is that an N-terminal signal peptide (SP) directs its cotranslational translocation into the endoplasmic reticulum, where it folds and oligomerizes into trimers that remain membrane anchored via a C-terminal transmembrane (TM) segment. The membrane-inserted SP is cleaved by a host signalase (*12*). The protomers are further proteolyzed by an additional host protease to generate two subunits which, based on the sequence positioning relative to the cleavage site, can be viewed as an N-terminal, peripheral subunit (N-SU) and a C-terminal, membrane-anchored subunit (C-SU) (*11*). The C-SU is maintained in a metastable, energy-loaded state within the primed trimer, and binding to a specific cellular receptor, either by the N-SU (*11*) or a separate viral glycoprotein (*13*), and/or protonation in the endosomal compartment (*10*, *14*) triggers a major conformational change allowing it to spring out to form an extended intermediate (*15*–*17*). In this state, the typically hydrophobic N-terminal end of the C-SU, termed "fusion peptide" (FP), inserts into the target membrane, while the protein remains anchored to the viral membrane via its TM segment (*18*). The C-SU quickly folds back and collapses into a low-energy, post-fusion trimeric state bringing target and viral membranes into apposition and coupling the released free energy from protein refolding to the required dehydration of the two outer leaflets of the membranes to drive their fusion (*19*).

The hallmark of all class I fusion proteins in the post-fusion state is a central trimeric coiled coil that begins at the C-terminal end of the FP and is formed by a long α helix that makes parallel interactions with its counterpart from the other protomers along the threefold molecular axis. The amino acid sequence in this region displays a seven-residue repeat pattern of nonpolar residues termed "heptad repeat A" (HRA, or HR1) (*20*, *21*). The C-terminal segment, near the TM anchor, which is often also α-helical and can be identified by HRs in the amino acid sequence [termed "heptad repeat B" (HRB, or HR2)], packs in an antiparallel fashion along the interhelical grooves of the trimeric coiled coil, placing FP and TM anchors at the same end of a trimeric rod-like structure. The resulting six-helix bundle (6HB) is present in many class I fusion proteins in their post-fusion state (*20*–*22*).

Heparan sulfate (HS) has been identified as an important attachment factor for FV entry (*23*, *24*). The existence of a protein receptor has been postulated (*24*), but its identity remains elusive. The segment containing residues 225 to 555 of FV Env was proposed to function as a receptor binding domain (RBD) based on the binding of recombinantly expressed FV Env fragments to cells (*25*). We reported an x-ray structure of the RBD, which revealed a previously

[1]Institut Pasteur, Université Paris Cité, CNRS UMR3569, Unité de Virologie Structurale, 75015 Paris, France. [2]Institut de Chimie des Substances Naturelles, CNRS UPR2301, Université Paris Saclay, 91190 Gif-sur-Yvette, France. [3]Institut Pasteur, Université Paris Cité, CNRS UMR3569, Unité d'Epidémiologie et Physiopathologie des Virus Oncogènes, 75015 Paris, France. [4]Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
*Corresponding author. Email: ignacio.fernandez@pasteur.fr (I.F.); marija.backovic@pasteur.fr (M.B.)

unidentified fold, and identified residues essential for HS binding (*26*). The RBD was shown to display dominant conformational epitopes targeted by human neutralizing antibodies (*27*).

Unlike the Env of most members of the *Orthoretrovirinae* subfamily, which are triggered by receptor binding at the plasma membrane, the FV Env–mediated fusion occurs in the acidic environment of the endosomes (*28*). The FV Env is composed of three subunits named leader peptide (LP), surface (SU), and TM (transmembrane) subunits (*2*) which correspond to the SP, N-SU, and C-SU, respectively, in the nomenclature that applies to all class I fusion proteins (*11*) and which will be used throughout the text for easier comparisons. Another distinguishing feature of FV Env is the presence of two cleavage sites for proteolytic maturation, an N-terminal furin-like site (residues 122-RRIAR-126) and a canonical furin site at position 567-RRKR-570 (fig. S1A). Cleavage at the first site, mediated by cellular furin (*29*), separates the SP from N-SU, while the second cleavage separates the N-SU from the C-SU. The SP remains associated with the mature FV Env trimer (*29*), which has two membrane-spanning segments, TM_1 and TM_2, within SP and C-SU, respectively (fig. S1A). Previous studies of intact FV Env on viral particles provided low-resolution cryo–electron microscopy (cryo-EM) maps that informed about the general architecture of the molecule and revealed that Env trimers form a lattice on the virion surface (*30*). The resolution of this structure was however insufficient to elucidate the atomic details of the protein or to understand its function.

Here, we describe the structures of the FV Env ectodomain determined by cryo-EM first in its post-fusion trimer of hairpin conformation, and then in the pre-fusion state. Our work shows that FV Env bears structural similarity to the fusion (F) protein of otherwise unrelated paramyxoviruses and pneumoviruses, with the notable insertion of an RBD upstream of the furin site that separates N-SU and C-SU. Comparison of FV Env in its two conformations, pre- and post-fusion, allows a description of the functional transition of the protein to drive membrane merger, revealing insights into the mechanism used by this ancient fusogen.

RESULTS AND DISCUSSION

The FV Env post-fusion state

We expressed the predicted ectodomain of the gorilla FV Env protein (described in detail in Material and Methods and fig. S1) and determined its structure by cryo-EM at 3.1-Å resolution (fig. S2 and table S1). The structure revealed a rocket-shaped Env trimer in the post-fusion conformation characteristic for class I proteins, with a 6HB projecting from its base (Fig. 1). The RBDs are arranged on the sides of a central shaft, with their long axes roughly parallel to the trimer axis (Fig. 1A). The low density in the regions occupied by the RBDs indicates their flexibility (fig. S2). The central shaft bears notable resemblance to the fusion (F) proteins of paramyxo- and pneumoviruses (Fig. 1B) (see below), displaying the same head, neck, and stalk regions, formed by domains I, II, and III as previously described for F (*31*, *32*). Domain I is a β sheet domain, domain II has an immunoglobulin superfamily fold (*33*), and domain III has an α/β topology, featuring a prominent "core β sheet" (fig. S3). When outlined on the primary structure of the protein, the domains of Env show a nested arrangement (Fig. 1C). The RBD, which constitutes a large portion of the N-SU, is the only domain made of a continuous stretch of amino acids (residues 217 to 543). The RBD is flanked by two linkers (L and L′), which connect it to domain III (Fig. 1D). Domain III is inserted between two β strands that belong to domain I, which is, in turn, flanked by domain II elements.

The FV Env coiled coil at the center of the 6HB is formed by a long α helix ("central helix") that packs in a parallel fashion against its counterparts from the other protomers along the molecular three-fold axis. The first resolved amino acid at the N terminus of the central helix, Asn^{573}, lies only two residues after the furin site (residues 567-RRKR-570), which has undergone cleavage in our construct (fig. S1B). The 21-turn long central helix encompasses the HRA segment, and its N-terminal portion includes the region predicted to be the FP (fig. S1, C and D) (*34*, *35*). The 16 N-terminal turns of the central helix (termed HRA^N helix) coil around each other along the trimer axis, while its 5 C-terminal turns (termed HRA^C helix) run parallel to HRA^C of the other protomers and also pack against helix α_2 of domain III (Fig. 1D). This helix α_2, which was designated as "HR**C**" in F homologs (Fig. 1B and fig. S3), connects to the RBD through a segment of random coil (linker L; residues 207 to 216) (Fig. 1D). The second linker, after the RBD (L′, residues 544 to 570), contains the furin cleavage site and is disordered in the structure.

At the C-terminal end of the construct, after domain II, the polypeptide chain adopts a random coil conformation, interrupted by α_5 helix, and follows the trimer interface along the central shaft, interacting with domain III of the adjacent protomers (Fig. 1, D and E). This segment is equivalent to the HRB linker and the HRB α helix in the F proteins (*36*) (for simplicity, we will apply the term HRB to designate the whole Env region spanning residues 845 to 884). The α_5 helix packs in the grooves of the HRA^N coiled coil, giving rise to the 6HB. The last resolved residue, Leu^{903}, belongs to a C-terminal helix termed α_C (Fig. 1D) that is juxtaposed to the FP, the element presumed to interact with the target membrane during fusion (Fig. 1A).

The presence of the FP in the post-fusion form has not been seen in other class I fusion proteins, either because it was disordered or because it was not included in the expressed construct. The FP packs against the α_C helix (residues 886 to 902) at the tip of the coiled coil. The current paradigm is that an extended intermediate is formed during the fusion reaction, in which the FP interacts with the target membrane, while α_C remains in the proximity of the viral membrane. In the recombinant Env ectodomain, these two segments adopted a stabilizing conformation (fig. S1E), which we think is an artifact since the interaction could not occur in the context of the full-length Env embedded in the lipid bilayer. Analysis of hydrophobic clusters at the interface between the coiled coil and the zippering HRB supports the idea, indicating that a 6HB would likely form even in the absence of the FP and α_C as the hydrophobic cluster formed by helix α_5 is larger than the contacts keeping the α_C and FP together (fig. S1D).

The FV Env pre-fusion conformation

To stabilize the metastable pre-fusion state, we modified the expression construct as described in detail in Material and Methods and fig. S3. Briefly, a trimerization motif (foldon) was added to the C terminus, the furin site was replaced by a noncleavable GS linker, and the E630P mutation was introduced to impede refolding of the central helix. A cryo-EM dataset was collected and processed, resulting in a 3.8-Å resolution three-dimensional (3D) reconstruction (fig. S5 and table S1).The obtained atomic model was fitted in the 8.8-Å cryo-EM map reported for the full-length Env expressed on viral

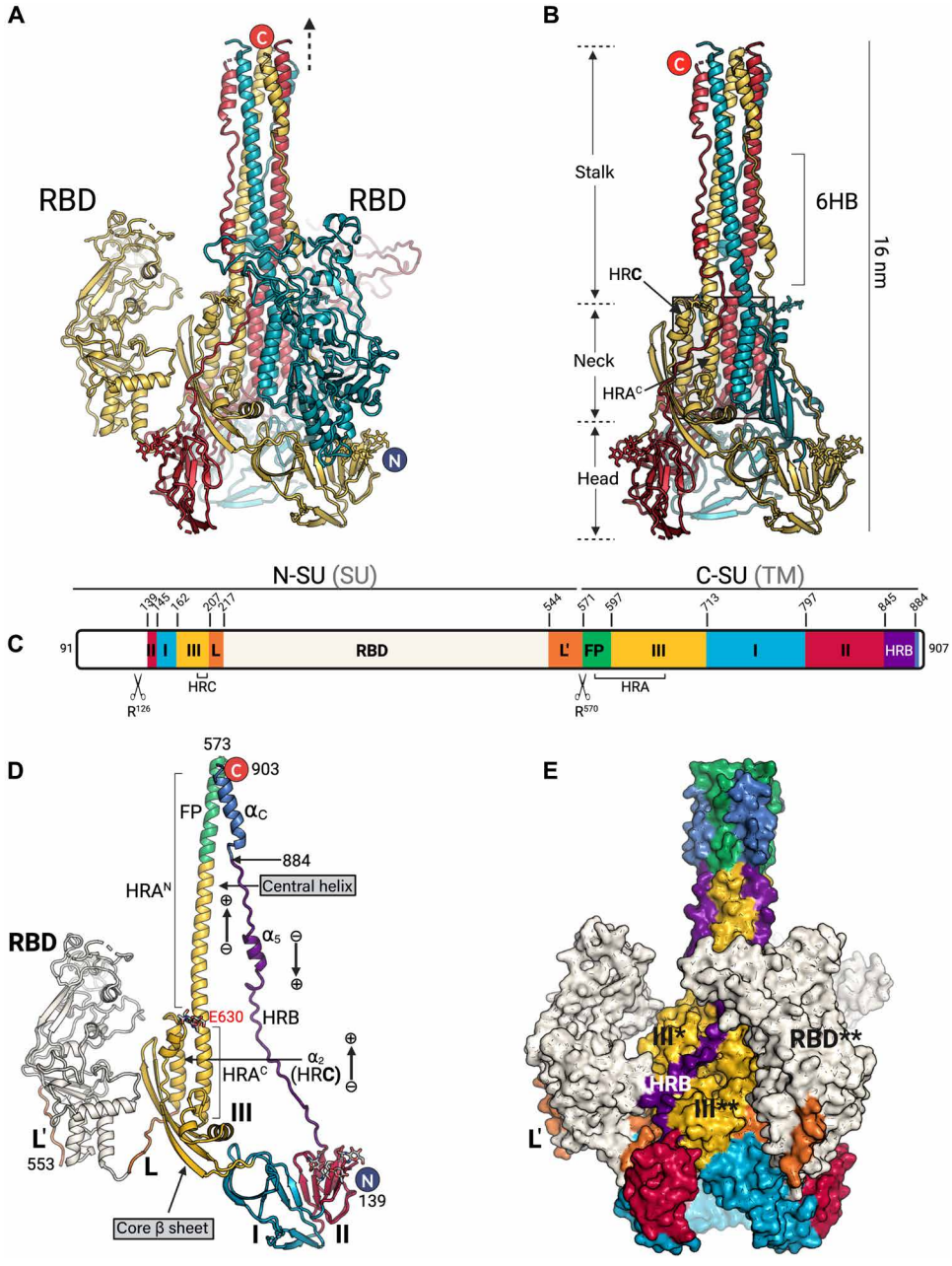

Fig. 1. Post-fusion FV Env structure. (**A**) Cartoon representation of the trimeric Env ectodomain, with each chain displayed in a different color. The first and last resolved residues (E139 and L903) are marked with "N" and "C," respectively. The dashed arrow indicates the membrane putative location. (**B**) The central shaft (the trimeric Env ectodomain without RBDs) is shown in cartoon representation. The helical core (three HRA^C and three HR**C** α helices, packing in parallel fashion) is indicated with the shaded box. (**C**) Schematic representation of the Env domains. The span of the N-SU and C-SU is shown on the top, with the conventional names used for subunits of the retroviral Envs, SU, and TM, respectively. The furin sites are indicated with the scissor icon. The domains are labeled with roman letters and colored as domain I (azure), II (cherry red), and III (bright gold). The numbers on the top correspond to the first residue of each domain. The FP (green), RBD linkers L and L′ (orange), RBD (beige), HRB (indigo), and α_C helix (navy blue) are highlighted. The HR**C** (residues 186 to 204) and the HRA segments (residues 576 to 651) of domain III are indicated below the scheme. The terminal regions not resolved in the structure are shown in white (residues 91 to 138 and 905 to 907). Other unresolved residues are indicated in fig. S6. (**D**) The Env protomer is shown in cartoon representation with domains colored as in (C). The central helix (HRA^N and HRA^C segments) and the HR**C** helix are labeled. The arrows indicate the helical dipolar moment. The core β sheet (domain III) is marked. (**E**) The trimeric Env ectodomain is colored by domain, as panel (C), and shown in molecular surface representation. The star symbols (* and **) designate the neighboring protomers.

vector particles (*30*) with a correlation coefficient of 0.91, validating that the Env ectodomain structure represents the fold adopted by the protein displayed on virions (fig. S4D). The map density was poor for N and C termini (fig. S6), including the foldon.

The pre-fusion Env trimer is a barrel-shaped molecule with two large cavities along the central threefold axis, one at the top (membrane-distal) and one at the bottom (membrane-proximal) side (Fig. 2A). The roof of the membrane-distal cavity is formed by the RBDs, which have a similar conformation to the one determined by x-ray crystallography (*26*). The local resolution of the map in the region corresponding to the RBD loops is >5.5 Å, highlighting their flexibility even in the context of trimeric Env and

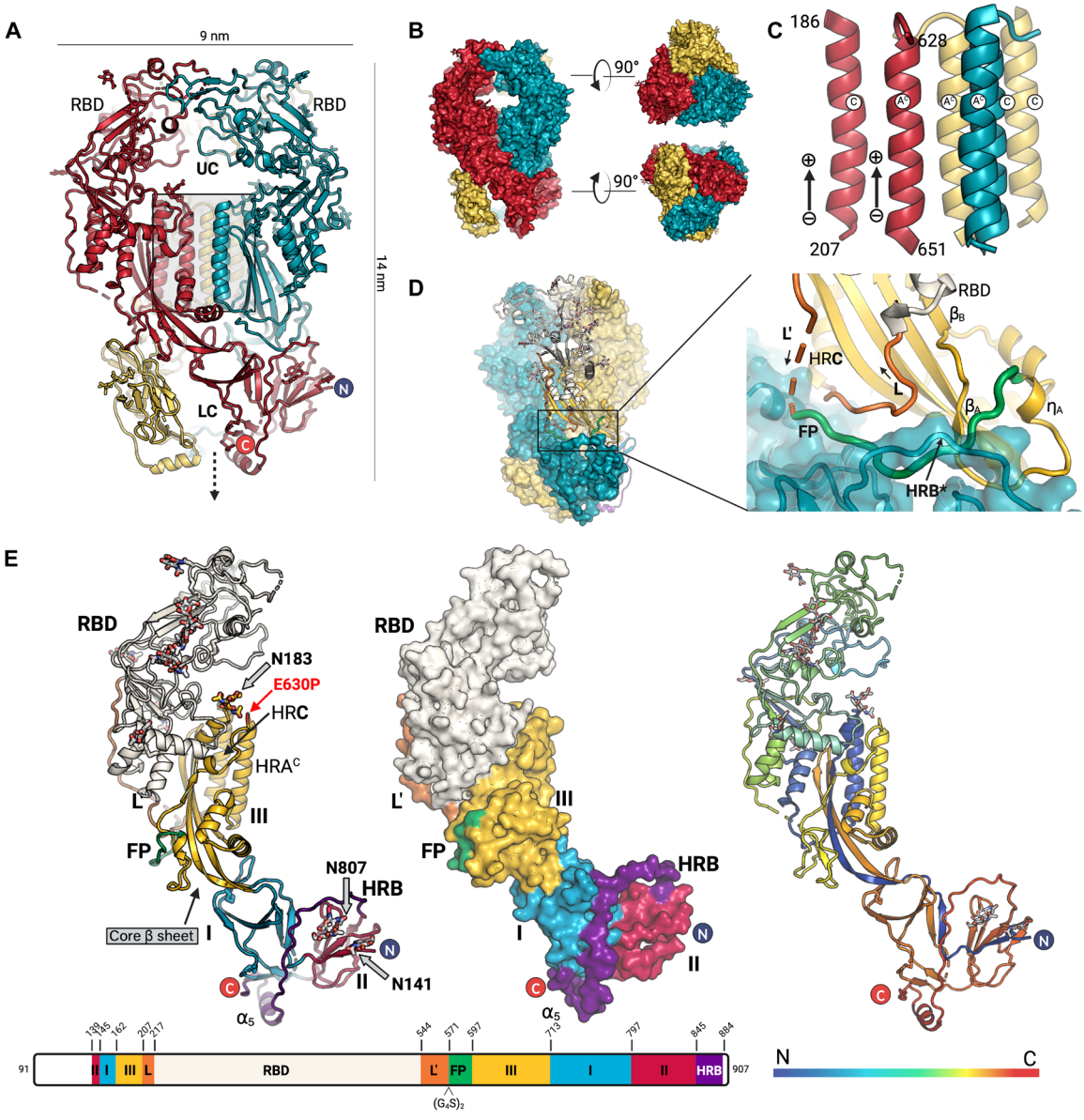

Fig. 2. Pre-fusion FV Env structure. (**A**) Cartoon representation of the trimeric Env ectodomain, with each chain displayed in a different color. The first and last resolved residues (E139 and S884) are marked with "N" and "C," respectively. The dashed arrow indicates the membrane putative location. UC and LC designate the upper and lower cavities, respectively. (**B**) Molecular surface representation of the trimer from (A). (**C**) Magnified region from the shaded box in (A), representing the helical core [three HRA^C ("A^C") and three HR**C** helices ("C")]. The arrows indicate the helical dipolar moments. (**D**) Magnified region (right) of the area indicated with the black line box on the trimeric Env (left). One protomer is shown in the cartoon, colored by domain, with two adjacent protomers shown as molecular surfaces (colored by chain). The FP inserts between the linker L and the HRB*. The structural elements in the HRA^N that follow the FP, the 3_{10} helix (η_A), and strands β_A and β_B are marked. The black arrows indicate the direction of the polypeptide chain. (**E**) Cartoon representation of Env ectodomain protomer (left) with domains colored as in the scheme below. The E630P mutation is marked with red letters. The HRA^C and HRC helices are labeled. The N-linked glycans are shown as stick models (indicated with gray arrows). The terminal regions not resolved in the structure are shown in white (residues 91 to 138 and 884 to 907). Other unresolved residues are indicated in fig. S6. The molecular surface of the same protomer (middle) illustrates the wrapping of the HRB around domains I and II, ending with the α_5 helix at the bottom of the molecule. The cartoon representation of the Env protomer is in blue to red, corresponding to the N to C termini (right).

implying a loose RBD-RBD interface (fig. S5). The membrane-proximal cavity is closed off by domains I and II, as well as the α_5 helix that packs underneath.

The three Env protomers wind about each other (Fig. 2B), making tight contacts in the central region through the HRA^C and the HR**C** (α_2) helices that run parallel to each other (Fig. 2C). The resulting helical core separates the membrane-distal and proximal cavities (Fig. 2B). The segment corresponding to the HRA^N helix in the post-fusion form, which includes the FP, adopts essentially a random coil conformation, with short segments of secondary structure (a 3_{10} helix η_A, followed by strands β_A and β_B and a short α_A helix; figs. S3 and S6). The FP runs below linker L of the same protomer, settling into a crevice formed by the HRB* segment, which partially secludes the FP from the outside (Fig. 2D) (the star sign "*" is used throughout the text to indicate belonging to the neighboring protomer). The stabilizing E630P mutation is in a loop connecting the HRA^N and HRA^C helices (Fig. 2E and fig. S6). Residues 557 to 581, comprising the second half of L′, are disordered, indicating that the furin site, which was substituted by a linker in our construct, would be exposed on the sides of the Env trimer where it is accessible for cleavage. The HRB region adopts an extended polypeptide conformation that wraps around domains II and I, ending with a short α_5 helix and strand β_{17}, which augments the domain I β sheet (fig. S3). The last amino acid with clear density, Ser^{884}, is immediately upstream the C-terminal helix α_C, which was observed only in the post-fusion form even though the expression constructs for the pre- and post-fusion Env ectodomains had the same boundaries (Fig. 1).

Several glycosylation sites were resolved in the FV ectodomain. The glycan attached to Asn^{183}, in a loop connecting helices α_1 and α_2 (HR**C**) of domain III (Fig. 2E), partially fills the membrane-distal cavity. The six glycans present in the RBD at Asn^{311}, Asn^{346}, Asn^{373}, Asn^{390}, Asn^{404}, and Asn^{524} project from the exposed surface of the trimer. Two additional glycans are visible in the structure, located on domain II at the base of the trimer, and attached to Asn^{141} and Asn^{807} (Fig. 2E).

Out of 20 cysteine residues, 18 were found to form nine intra-protomer disulfide bonds (DSB), and 2 cysteines formed one inter-protomer DSB. Residue C562 could not be modeled because of the poor quality of the cryo-EM map in this region, leaving C751 from the C-SU domain I unpaired. An unsharpened cryo-EM map provided the hint that C562 and C571* were in position to make an inter-protomer DSB (fig. S7A). SDS–polyacrylamide gel electrophoresis (SDS-PAGE) of cleaved and uncleaved constructs under non-reducing conditions showed that the Env trimer is covalently linked by DSBs between the N-SU from one protomer and the C-SU* (see Material and Methods; fig. S7, B and C). The inter-protomer DSB could be only attributed to C562-C751* since all the other cysteine residues are engaged in intra-protomer DSBs. Analysis of a C562A mutant confirmed this prediction (fig. S7C).

Additional insights into pre-fusion FV Env from AlphaFold multimer predictions

To obtain structural insights on regions that were not included in the expression construct or were not resolved in our stabilized Env ectodomain, we performed ab initio structural prediction of full-length FV Env using Alphafold multimer (AFM) (*37*), as described in Material and Methods. The AFM model for the full-length Env trimer has the predicted template modeling (pTM) score of 0.77 (the pTM reflects the expected similarity between the predicted and the real structure), and the interface pTM score of 0.76 (indicates the confidence of the docked protein interface domains). The global local distance difference test (LDDT) value for the AFM model is 66.75, and the "LDDT per residue" values are >70 for the majority of the ectodomain residues, except for the exposed loops containing the two cleavage sites and the FP segment (Fig. 3A and fig. S8A). Although the relative orientation of the domains I and II within each protomer varies in comparison to the experimental structure (Fig. 3A), the predicted trimeric model is notably similar to the experimentally determined structure, as indicated by the root mean square deviation (RMSD) of 1.7 Å for the superposition of 378 C_α atoms of the domain III in the context of the Env trimer (residues 162 to 206 and 631 to 712 in each protomer). The folding of the individual domains is preserved, as indicated by RMSD values < 1.4 Å (table S2). The regions corresponding to the FP, linker L, and HRB* have low confidence values and do not reveal the interaction observed between these elements in the cryo-EM structure.

The AFM prediction suggested that the extracellular LP residues 103 to 106 and 112 to 116 form two β strands, termed β_{-1} and β_0, respectively. Strand β_{-1} runs antiparallel to another predicted strand (β_{16}″, residues 862 to 865), which belongs to the HRB region and adopts a random coil conformation in our experimental structure (Fig. 3B). It is possible that the interactions between these two β strands further stabilize the HRB in the pre-fusion conformation by tying together the Env segments close to N and C termini. The β_0 strand (residues 112 to 116) forms a β hairpin with strand β_1, which was resolved in our structure, projecting the SP furin-like site to the side of the trimer and augmenting a domain II β sheet (Fig. 3 and fig. S3). When placed in the reported cryo-EM map of full-length Env (*30*), the β_0-β_1 hairpin fills the density that connects different trimers on the surface of viral particles, indicating that it could be an important structural element for the organization of the observed Env lattice (fig. S8B).

AFM further predicts that the C-terminal helix α_C, which was not observed in the experimental structure of pre-fusion Env, is part of an α-helical bundle, termed here domain IV, at the bottom of the pre-fusion trimer (Fig. 3B). This bundle contains four α helices and closes the lower cavity of the Env trimer, reinforced in part by the long TM_1, which crosses with the C-terminal TM_2 anchor at an angle (Fig. 3B). Two amphipathic α helices of the domain IV bundle (α_6 and α_7; fig. S6) are predicted to pack together roughly at the same height as the N terminus of TM_2 (residue Q100), in an orientation parallel to the viral membrane and facing the membrane with their hydrophobic sides (Fig. 3C). This observation suggests that the domain IV α helices may play the role of membrane-proximal regions (MPRs), which can be involved in membrane destabilization during fusion [reviewed in (*38*)]. Poor density was observed at the bottom of the maps of the pre-fusion Env ectodomain construct (ab initio models, classes 0 and 3; fig. S5), suggesting mobility of domains I and II, possibly due to the absence of the MPRs which were not included in our expression construct to avoid solubility problems. Last, AFM also predicts another α-helical bundle in the cytosolic side of the viral membrane (domain V) (Fig. 3C and fig. S3).

A common structural core in FV Env and paramyxo- and pneumovirus F

The topological distribution of secondary structure elements in FV Env domains I, II, and III (fig. S3C) shows only minor variations to

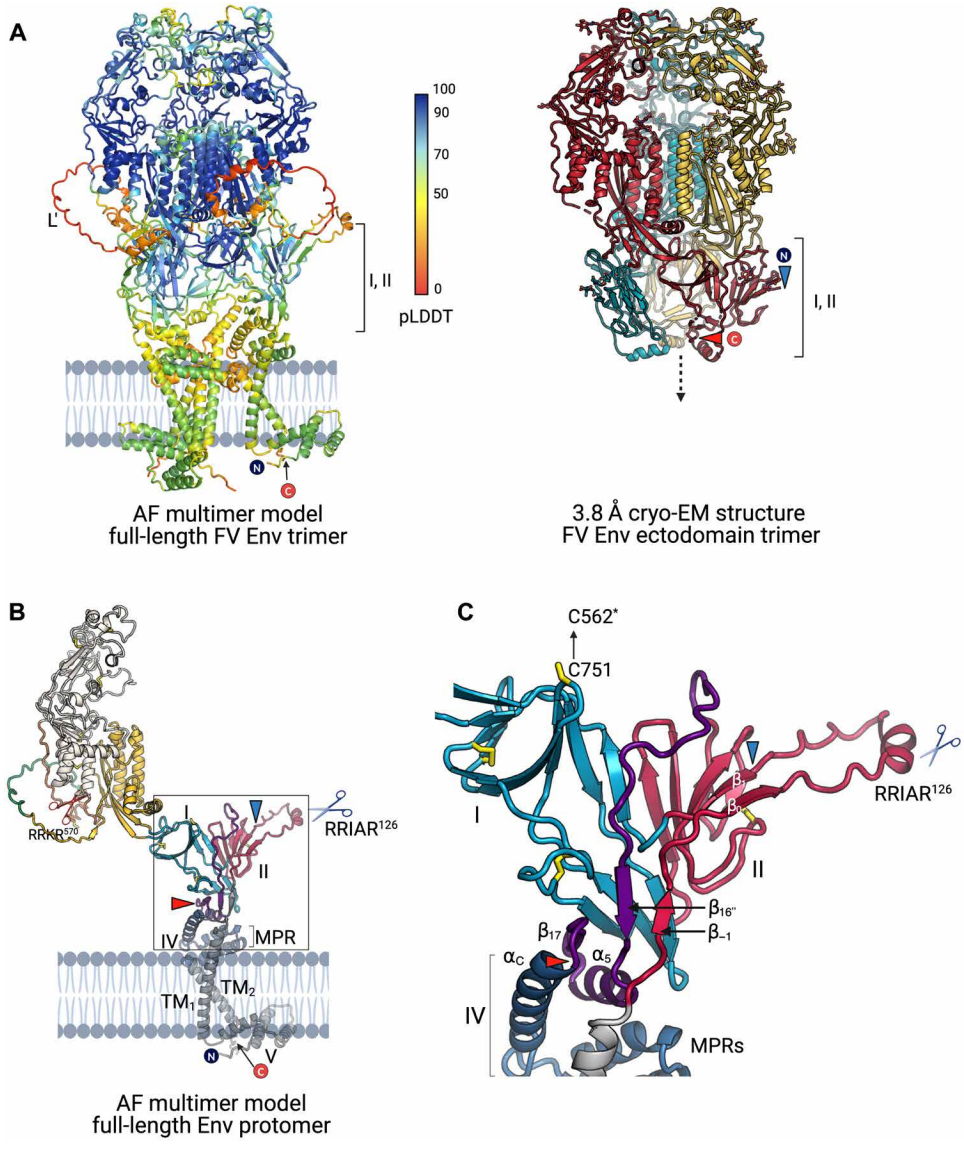

Fig. 3. AF Multimer predicted model of the full-length Env trimer in the pre-fusion state. (**A**) The AFM model of the full-length trimer is colored according to the per-residue confidence metric called the "predicted local distance difference test" (pLDDT). The pLDDT can have a value between 0 and 100: the higher the pLDDT number, the higher the confidence for the residue placement in the model. The predicted location of domains I and II, upward compared to the experimental structure, is indicated, as well as the N and C termini. The experimental cryo-EM structure of the pre-fusion Env ectodomain is shown for comparison on the right. The protomers are each shown in a different color. The blue and red triangles designate the first and last resolved residue, Glu^{139} and Ser^{884}, respectively. (**B**) A protomer of the AFM model shown in (A) is displayed and colored by domain using the color scheme defined in Fig. 1C. The domains are marked with roman letters I to V. The helical bundle that comprises α_C and the putative membrane-proximal region (MPR) helices is colored in navy blue. The TM helices belonging to the SP and C-SU, the TM_1 and TM_2, respectively, are indicated and colored in gray together with the helical bundle of the luminal side that forms domain V. The blue and red triangles designate the first and last resolved residue of the experimental structure shown in (A). (**C**) Domains I and II and the MPR of the Env pre-fusion model boxed in (B) are magnified to highlight α_C as a part of domain IV, as well as β strands preceding the furin-like cleavage site. DSBs are shown in yellow stick representation, and the side chain of C571, which forms an inter-protomer DSB with C562* in the AFM model, is indicated.

those observed in paramyxovirus and pneumovirus F (*39*, *40*). The RMSD calculated for the superposition of 147 C_α, corresponding to domains I, II, and III of FV Env and parainfluenza virus type 3 F is 3.6 Å, indicating comparable 3D organization of the three domains (Fig. 4). The intertwined pattern of nested domains shown in Fig. 1C was first described for paramyxovirus and pneumovirus F (fig. S9) (*40*, *41*) and was also observed, along the structural similarity, for the fusion subunit of the coronavirus spike protein extending the homology (spike does not have domain II and domain I is smaller) (*42*, *43*). The conservation of the structural core, composed of the core β sheet and the helical bundle, now provides an additional evolutionary link between paramyxo-, pneumo-, and coronaviruses, with the *Spumaretrovirinae* subfamily of the retroviruses.

The sugar attached to the highly conserved Asn^{183} of FV Env (fig. S6), which projects into the membrane-distal cavity and partially covers

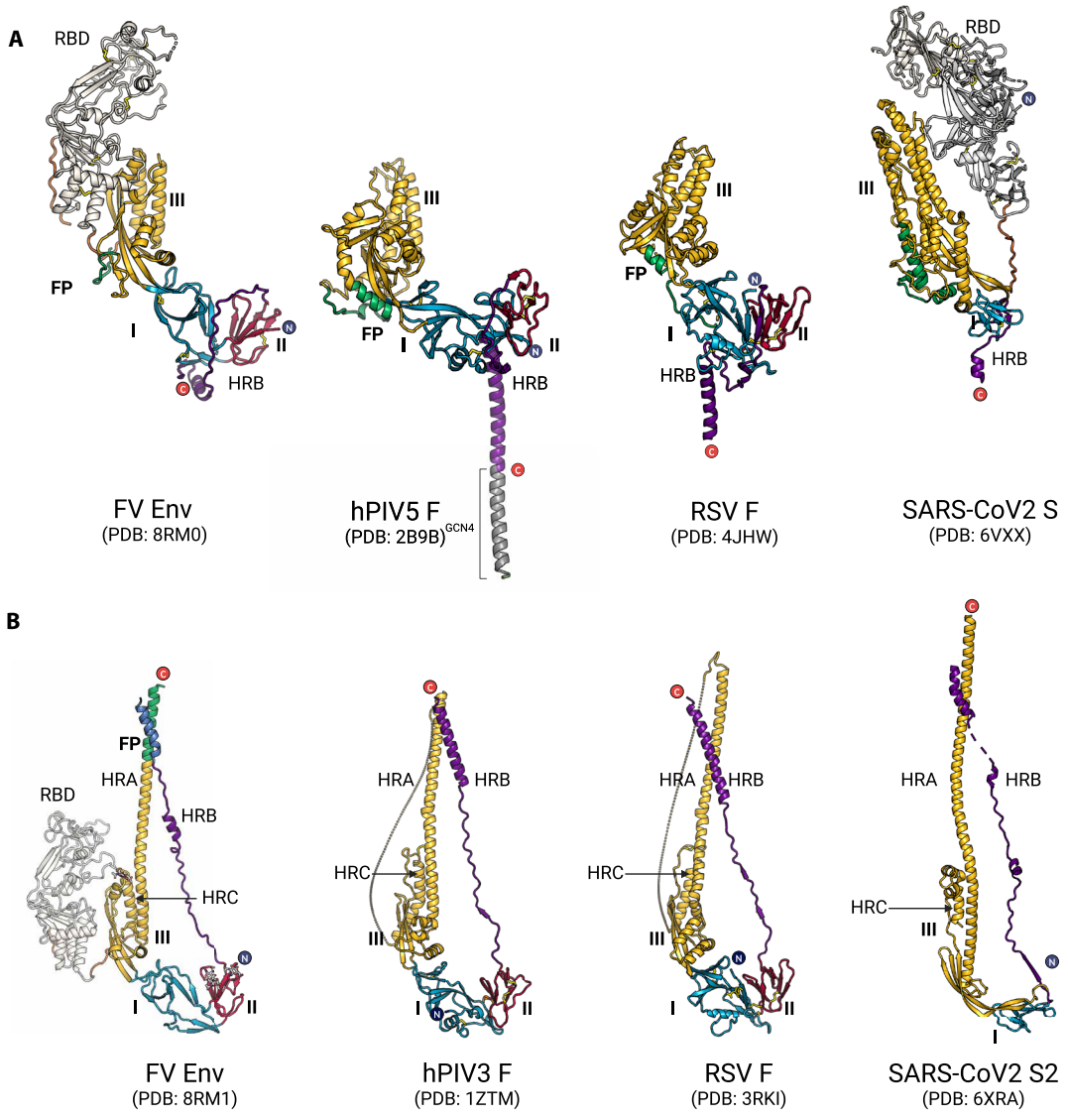

Fig. 4. Structural similarity between fusion proteins of FV, paramyxo-, pneumo-, and coronaviruses. Pre-fusion and post-fusion structures are shown in (**A**) and (**B**), respectively. Only protomers are displayed for clarity. Domains are colored according to the scheme in Fig. 1C. hPIV stands for human parainfluenza virus (now called human respirovirus, member of *Paramyxoviridae*), RSV for respiratory syncytial virus (now called human orthopneumovirus, member of *Pneumoviridae*), and SARS-CoV-2 for severe acute respiratory syndrome coronavirus 2 (member of *Coronaviridae*). PDB codes for each structure are indicated. The terminology used for coronavirus spike domains (*75*) is different from one established for F proteins; for simplicity, we applied the former to all the proteins belonging to this subset.

the HRA^N segment in the pre-fusion conformation, does not form polar contacts with the protein and does not have an extensive buried surface area (Fig. 2E), suggesting a role other than a structural one, which we had attributed to the oligosaccharide linked to the RBD Asn^{390} (*26*). A prototype FV vector (FVV) carrying the Env N183Q variant had the same infectivity as the wild type (WT) (*44*). The F proteins of some pneumo- and paramyxoviruses contain a glycan at the equivalent location, i.e., just upstream of the HR**C** helix, which is fully exposed at the apex and was shown to be important for immune evasion (*45*, *46*). The glycan at Asn^{183} in FV Env does not mask the RBD subdomain targeted by neutralizing antibodies in vivo (*27*, *47*), but it projects hovering on top of a refolding region and could mask potentially neutralizing epitopes located there.

Similar to the structure of the pre-fusion uncleaved paramyxovirus F protein, the FP of FV Env wraps mostly around the exterior of the trimer at the inter-protomer interface (*41*). In contrast, in the pre-fusion structure of pneumovirus F, the FP was found packing against an adjacent protomer but buried within the interior of the F trimer (*40*). A unique feature that distinguishes respiratory syncytial virus (a pneumovirus) F from its paramyxovirus and FV counterparts is the presence of two consecutive furin sites preceding the FP (Fig. 4), both of which must be cleaved for the pre-fusion trimer to assemble, as the intervening p27 segment interferes with trimerization (*48*).

The most prominent feature that distinguishes Env from F in the pre-fusion form is the RBD insertion in Env, in line with FVs not having a dedicated, additional protein for receptor binding, as

paramyxo- and pneumoviruses do (*49*). In this context, the N-SU and C-SU of FV Env would correspond to the equivalent subunits of the F protein (Fig. 5 and fig. S9), separated by the amino acidic sequence of the RBD that is accommodated at the periphery of the F scaffold in the tertiary structure (Fig. 4).

Another important difference between the pre-fusion conformations of F and FV Env is observed at the C-terminal region. The HRB segment connecting domain II to the α_5 helix of Env initially follows a path similar to that of its counterpart in the F proteins, interacting along the sides of domain I to reach the "bottom" of the trimer. However, the HRB helix α_5 packs underneath domain II in an orientation parallel to the viral membrane in Env (Figs. 2E and 6A), rather than forming the coiled-coil "stalk" structure observed in paramyxovirus F (*41*) and coronavirus spike (*43*), or the inverted pyramid three-helix bundle observed for pneumovirus F (*40*). This difference may arise from constraints imposed by two membrane-spanning segments in the FV Env. TM_1 is required because the preceding N-terminal region interacts with the Gag protein and is essential for FV budding (*50*, *51*), in contrast to orthoretrovirus Env, which binds to Gag via the cytosolic C-terminal segment downstream of TM_2.

The proposed model for the FV Env conformational change

During the conformational transition, the RBDs undergo a major repositioning (their centers of mass shift ~30 Å; table S3) while maintaining their fold (table S4). As was observed for the F proteins, the FV Env domains I and II also preserve their 3D folds and relocate as rigid bodies (Fig. 6 and table S4). The core β sheet and HR**C** helices from domain III remain unchanged in the Env, F, and spike and serve as a hub around which the HRA segment refolds. The helical core however collapses inward during the pre- to post-fusion transition in the F proteins, compacting the head region (*40*, *41*). Conversely, the conformation of the helical core is maintained in FV Env (the superposition of C_α atoms has the RMSD of 1.2 Å), conferring a larger core around which the structural rearrangement takes place.

In the pre-fusion FV Env state, the RBDs are held in place by loose inter-protomer interactions at the top of the trimer and by contacts between linker L, which precedes the RBD, with the FP and HRB* segment (Figs. 2D and 6A). On the basis of the comparison of the pre- and post-fusion structures of FV Env, we propose the following sequence of events for the conformational transition. Binding to HS and/or a putative receptor pulls the RBDs away from each other, resulting at least in a partial opening of the Env trimer. The movement of the RBDs, and consequently the linker L, destabilizes the interface between L, FP, and HRB* (Fig. 1D), liberating the two latter segments to establish new contacts. As a result, the HRB region unwraps from domain I and swings out to extend upward, pulling along domains II and I. Complete opening of the RBD trimer apex is necessary to create space for the refolding of the HRA^N segment into the central helix to form the coiled coil and project the FP into the target membrane to make an elongated intermediate. The acidic environment of the endosomes and residues sensitive to pH could play a role in the destabilization of the MPRs and TM anchors, which also need to separate for membrane fusion to complete.

Unlike class II fusogens (*52*), the class I fusion proteins do not adopt a single fold and cannot be considered to imply homology because their hallmark—the coiled coil motif, accommodating the C-terminal segments of the protein in the interhelical grooves—is simple and could have evolved independently during evolution more than once. We propose that the class I proteins are divided into at least two subclasses of likely different origins. One subclass, which we term class Ia, includes orthoretrovirus Env, influenza virus HA, filovirus GP, and arenavirus GPC, where the membrane fusion function is confined to the C-SU that results from the maturation cleavage. The N-SU contacts with C-SU in the post-fusion state are absent or limited to a few secondary structure elements (*53*). The paramyxo- and pneumovirus F, coronavirus spike, and FV Env have a more complex topology, with N-SU and C-SU contributing to all three domains, making N-SU a structural component of the C-SU refolding fusion machinery. Because of the large topological differences between both subclasses, and the lack of fusogens showing

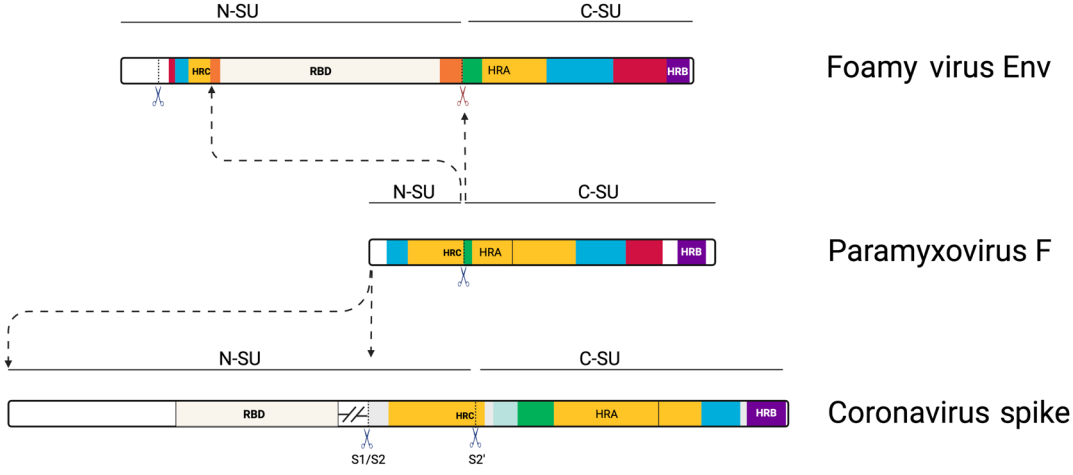

Fig. 5. Modular nature of class Ib fusion proteins. Schematic representations of the primary structure of the fusion proteins from FVs, paramyxoviruses, and coronaviruses. The segments constituting domains I, II, and III are colored according to the scheme from Fig. 1C. The dashed arrows indicate the incorporation sites of the RBD modules. The furin cleavage sites are marked with the scissors symbol. The light green box in the coronavirus spike, just upstream of the FP, corresponds to the FP proximal region, a segment observed only in the spike (see fig. S9 for details) (*76*).

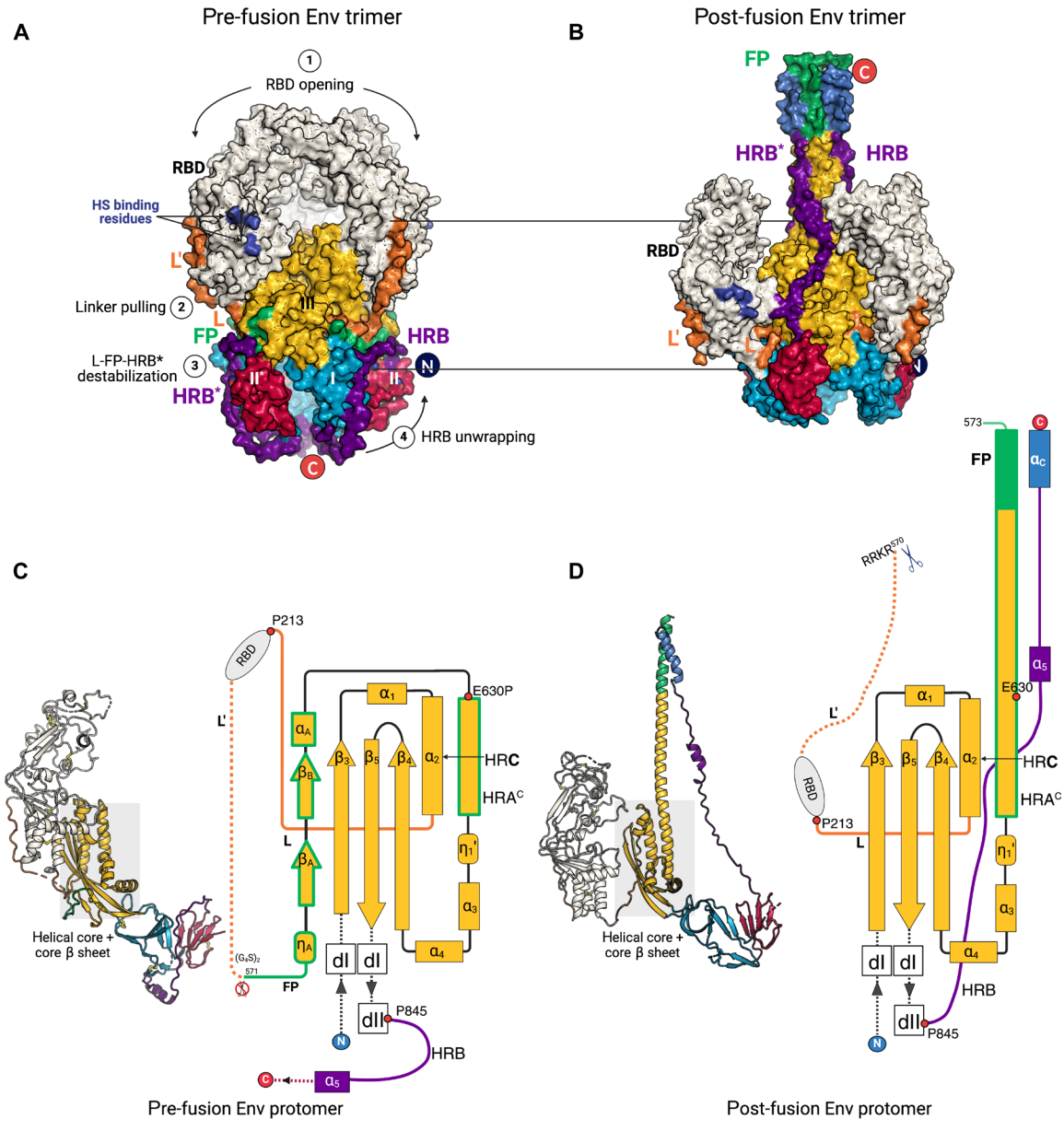

Fig. 6. Insights into the conformational change mechanisms from the comparisons of pre- and post-fusion Env structures. (**A** and **B**) Molecular surface representations of the pre-fusion and post-fusion Env trimer, respectively, colored by domain and according to the scheme from Fig. 1C. The pre- and post-fusion structures were aligned on the domain III helical core (HRA^C and HR**C** helices). The HS binding residues in the RBD domain, K342, R343, R356, and R369, are highlighted in dark blue. Numbers 1 to 4 illustrate a possible order of events during the conformational change: 1 - lateral movements of the RBDs, 2 - RBD pulling of the linker L, 3 - destabilization of the L-FP-HRB* interface and liberation of the FP, and 4 - unwrapping of the HRB, which flips and zippers down the post-fusion coiled coil. The star symbol "*" designates the structural elements belonging to the adjacent protomer. (**C** and **D**) The cartoon representations of the Env protomers in the pre- and post-fusion states, respectively. The inlets show the secondary structure topology of the domain III elements core β sheet and the HRA region that undergoes refolding, which is highlighted with green borders. The locations of the E630P mutation and the strictly conserved prolines, residues 213 and 845, which represent the points where the polypeptide chain changes direction, are indicated with red circles.

sequential gain/loss of the domains, we hypothesize that subclass Ia and subclass Ib proteins have evolved independently from different origins. Conversely, and despite the lack of amino acid sequence conservation, the subclass Ib fusion proteins from paramyxoviruses, pneumoviruses, coronaviruses, and FVs share a common structural core indicating that they derive from a distant common ancestor (Fig. 4). Within subclass Ib, the F protein is the smallest, with a fold that represents the minimal structural scaffold that can accommodate receptor binding modules, either at its N terminus, as in coronavirus spike, or inserted between the two subunits, as in the FV Env (Fig. 5 and fig. S9). The haemagglutinin-esterase-fusion protein of influenza C virus presents a similar modular nature, where the esterase and the RBD might have been incorporated into an ancestral membrane-fusion protein (*54*). Phylogenetic analyses revealed that

FVs originated >450 million years ago making them one of the most ancient groups of viruses known so far (*6*). One could speculate that the FV *env* could have provided a template for excisions or insertion of fragments, such as the RBD. Consistent with this idea are the findings highlighting the modular nature of simian FV genomes and the existence of two *env* variants that differ in a region that is flanked by two recombination hotspots within the N-SU (*55*).

The lack of structural similarity between the FV Env and the Env of orthoretroviruses, along with the reported novel fold for the FV RBD (*26*), further supports the classification of FVs as a separate retroviral subfamily, which had originally been done based on the unique features of their replication cycle (*1*). Many questions, however, remain to be answered, primarily the requirement and nature of the cellular receptor and the fusion triggering mechanism that may rely on the presence of the receptor as well as acidification. The atomic-resolution structures of the FV Env in the pre- and post-fusion states now provide a framework for further functional investigations such as the FV cell tropism and molecular features controlling the Env fusogenicity. The structural information can also be exploited for the design of FV Env variants with improved features for their use as gene therapy vectors.

MATERIALS AND METHODS

Construct design

We predicted the boundaries of the gorilla simian FV Env ectodomain (GII-K74 strain, accession number JQ867464) using Phyre2 (*56*), which predicted two TM helices: the first between residues 60 and 90 and the second between 948 and 978. We tested constructs bearing C-terminal truncations at residues 907, 936, or 945 and selected 91 to 907 as the best ectodomain candidate based on its higher expression.

The WT Env ectodomain served as the template for introduction of mutations commonly used to stabilize viral fusion proteins in their pre-fusion conformation (*57*): (i) the R126A mutation in the furin-like site on the LP that prevented the SP/N-SU cleavage (*29*), (ii) the canonical furin site (567 to 570) was replaced with a flexible and noncleavable linker (GGGGSGGGGS) abrogating the cleavage, and (iii) a foldon trimerization motif was added at the C terminus (SAIGGYIPEAPRDGQAYVRKDGEWVLLSTFLG). The resulting expression construct was termed variant 1 (V.1 on fig. S4A). Cryo-EM analysis and the 3D ab initio models revealed that about 40% of the particles appeared to have a conformation reminiscent of the one obtained by docking the RBDs into a low-resolution reconstruction of the pre-fusion full-length Env trimer on virions (*26*, *30*). This conformation is characterized by having the RBDs—which are readily identified by their characteristic bean shape—assembled near to each other in a closed conformation (fig. S4B). To further stabilize the pre-fusion form, we performed a proline scan around a region of the central helix where we anticipated a helix break in the pre-fusion form, based on similar work on other class I fusion proteins (*58*, *59*). We thus identified mutant E630P (construct V.2 on fig. S4A), which resulted in ~70% of the particles appearing in the closed, presumably pre-fusion form, and which allowed us to generate maps and build an atomic model. The E630P did not influence vector particle production, infectivity, and cell binding (fig. S4D), indicating that the mutation itself was not sufficient to lock the Env in the pre-fusion state, and suggesting larger importance of the furin site substitution and foldon for Env stabilization.

Protein expression and purification

The codon-optimized DNA sequence coding for the WT gorilla FV Env ectodomain from strain SFVggo_huBAK74 [accession code JQ867464.1 (*60*)] was synthesized by GenScript and cloned into a modified pMT/BiP insect cell expression plasmid (Invitrogen) designated pT350, which contains a divalent-cation inducible metallothionein promoter, the BiP SP at the N terminus (MKLCILLAVVAFVGLSLG), and a double strep tag (AGWSHPQFEKGGGSGGGSGGGSWSHPQFEK) at the C terminus (*61*). Mutagenesis to obtain the stabilized ectodomain construct, as well as the R570T and C562A mutants, was performed by GenScript.

Each plasmid was cotransfected in Drosophila Schneider line 2 cells (S2) with the pCoPuro plasmid for puromycin selection (*62*). Both cell lines underwent selection in serum-free insect cell medium (HyClone, GE Healthcare) containing puromycin (7 μg/ml) and 1% penicillin/streptomycin. For the protein production stage, the cells were grown in spinner flasks until the density reached ~1×10^7 cells/ml, at which point the protein expression was induced with 4 μM $CdCl_2$. After 6 days, the cells were separated by centrifugation, and the supernatants were concentrated and used for affinity purification using a StrepTactin column (IBA). The proteins were further purified by size exclusion chromatography (SEC) on a Superose6 16/60 (Cytiva) column (WT ectodomain) or using a Superdex 200 10/300 (Cytiva) column (stabilized ectodomain) in 10 mM tris and 100 mM NaCl (pH 8.0). Eluted fractions were analyzed by SDS-PAGE under reducing and nonreducing conditions, and some of them were pooled and concentrated (WT ectodomain) or stored separately (stabilized ectodomain).

Sample preparation for cryo-EM

To find the best construct stabilized in the pre-fusion conformation, a screening of candidates was performed by preparing grids with 3 μl of protein (1.5 μM trimer in SEC buffer), which were applied on Quantifoil R1.2/1.3 200 mesh copper grids that had been glow-discharged twice using a Pelco glow discharge system at 15 mA for 25 s. Samples were immediately vitrified in 100% liquid ethane using a Mark IV Vitrobot (Thermo Fisher Scientific) by blotting for 3.5 s with Whatman no. 1 filter paper at 9°C and 100% relative humidity. Three microliters of the selected stabilized ectodomain was added to Quantifoil R2/2 200 mesh copper grids with a 2-nm continuous carbon film on top (Delta Microscopies), which were glow-discharged twice beforehand using a Pelco glow discharge system at 15 mA for 25 s. Samples were immediately vitrified in 100% liquid ethane using a Mark IV Vitrobot (Thermo Fisher Scientific) by blotting for 3.5 s with Whatman no. 1 filter paper at 9°C and 100% relative humidity, after 10-s waiting time. Three microliters of the WT ectodomain (1.0 μM trimer in SEC buffer) was added to Quantifoil R1.2/1.3 200 mesh copper grids (Delta Microscopies), which had been glow-discharged as indicated before. Samples were immediately vitrified by blotting for 4 s with Whatman no. 1 filter paper at 9°C and 100% relative humidity.

Cryo-EM data collection, processing, refinement, and modeling (post-fusion form)

Data of the WT ectodomain were acquired on a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 kV, using EPU automated image acquisition software (Thermo Fisher Scientific). Movies were collected on a Gatan K3 direct electron detector operating in counting mode at a nominal

magnification of ×105,000 (0.85 Å/pixel) using a defocus range of −1.0 to −3.0 μm. Movies were collected over a 2-s exposure and a total dose of ~50 e−/Å^2.

Data were processed by using RELION 3.0 (*63*) following the workflow shown in fig. S2. Briefly, row movies were processed using MotionCor2, with a five-by-five patch-based alignment, keeping all frame and dose-weighting up to the total exposure. The CTFs of the dose-weighted images were determined using CTFFIND4.1 (*64*). Images having a CTF maximal resolution worse than 4 Å, as well as those with important ice contamination or large carbon area, were discarded. Particles were picked using the RELION Laplacian of Gaussian; they were extracted and submitted to three rounds of 2D classification. The particles from the selected classes were used to generate an initial 3D model, which, in turn, was used to perform 3D classification. The position of particles from one class was refined, and the resulting map (4-Å resolution) was filtered to 20 Å and used to perform a round of auto-picking with a 3D reference. The new set of particles was extracted and used for two cycles of 2D classification, and those from the best classes were submitted to a 3D classification job with 50 iterations and the initial 3D model as a reference. Particles belonging to the best 3D class were re-extracted (without binning) and used to perform per-particle CTF refinement, particle polishing, and another per-particle CTF refinement (shiny particles). A final round of 3D classification rendered four classes, two of which were selected to combine their particles in a last 3D refinement round, resulting in a 3.1-Å resolution map.

Building the N-SU atomic model started by placing the crystal structure of the RBD [PDB: 8AIC (*26*)] into the sharpened EM map using Chimera and manually extending the N terminus in Coot. The first residues of the TM atomic model were built by the map-to-model tool from the Phenix suite, which generated a segment of the HR1 helix that was manually extended in N- and C-terminal directions in Coot. Following this strategy, an initial model was obtained and it was refined with one round of real-space refinement including simulated annealing in Phenix, followed by iterative rounds of manual building and real-space and B-factor refinement in Coot (*65*) and Phenix, with secondary structure restraints. Symmetry operators were obtained from the EM map, and they were used to place three copies of the protomer within the map using tools from Phenix. Validation of model coordinates was performed using MolProbity (*66*).

Cryo-EM data collection, processing, refinement, and modeling (pre-fusion form)

Screening of the stabilized ectodomain candidates was performed by collecting single-particle cryo-EM data on a Glacios transmission electron microscope (Thermo Fisher Scientific) operating at 200 kV, using the EPU automated image acquisition software (Thermo Fisher Scientific). Movies were collected on a Falcon 4i direct electron detector operating in counting mode at a nominal magnification of ×150,000 (0.95 Å/pixel) using a defocus range of −1.0 to −2.75 μm. Movies were collected over a 4.3-s exposure and a total dose of ~40 e−/Å^2.

Single-particle cryo-EM data of the stabilized ectodomain were acquired on a Glacios transmission electron microscope (Thermo Fisher Scientific) operating at 200 kV, using the EPU automated image acquisition software (Thermo Fisher Scientific). Movies were collected on a Falcon 4i direct electron detector operating in counting mode at a nominal magnification of ×150,000 (0.95 Å/pixel) using a defocus range of −1.0 to −2.75 μm. Movies were collected over a 4.3-s exposure and a total dose of ~40 e−/Å^2. All movies were motion-corrected and dose-weighted with MotionCorr2 (*67*), and the aligned micrographs were used to estimate the defocus values with patchCTF within cryoSPARC (*68*). CryoSPARC blob picker was used for automated particle picking, and the resulting particles were used to obtain initial 2D references. Five initial ab initio 3D models were generated in cryoSPARC and the one with the most particles was selected to perform a heterogeneous refinement into three classes without imposing any symmetry. The particles from the best class (which generated a map with more density) were subjected to a 3D classification into five classes using cryoSPARC without initial volumes or masks. The result with the best density across all protomers was chosen for a nonuniform refinement with C3 symmetry in cryoSPARC (*68*). The final map was sharpened with DeepEMhancer (*69*).

Model building started with the domains I, II, and III (missing the HRA refolding region) from the pre-fusion model and later from the AFM model of the trimeric full-length Env. The individual domains were fitted into the sharpened EM map using UCSF Chimera (*70*). The regions connecting the individual domains were deleted and manually rebuilt. This initial model was refined with one round of real-space refinement including simulated annealing in Phenix (*71*) followed by iterative rounds of manual building and real-space and B-factor refinement in Coot (*65*) and Phenix, with secondary structure restraints. Symmetry operators were obtained from the EM map with the map symmetry tool in Phenix, and they were used to place the three copies of the protomer within the trimeric map with the apply-NCS-operators program in Phenix. Validation of model coordinates was performed using MolProbity (*66*).

Cells, sequences, and production of Foamy virus viral vectors

The four-component FVV system [plasmids pcoPG, pcoPP, EnvGI-SUGII, and pcu2MD9-BGAL (a transfer plasmid encoding for β-galactosidase)] has been described (*47*). The EnvGI-SUGII plasmid encodes a full-length gorilla Env that comprises the LP and TM sequences from the zoonotic SFVggo_huBAD468 (JQ867465) and the SU from the SFVggo_huBAKK74 (JQ867464). We refer to this Env as WT. The mutation E630P was introduced in WT Env. FVVs were produced by cotransfection of the four plasmids (gag:env:-pol:transgene β-galactosidase) at a ratio of 7:2:3:26. Three micrograms of total DNA and 8 μl of polyethyleneimine (JetPEI, Polyplus, Ozyme) were added to 0.5 × 10^6 human embryonic kidney 293 T cells seeded in six-well plates. Supernatants were collected 48 hours after transfection, clarified at 1500*g* for 10 min, and stored as single-use aliquots at −80°C. Vector infectivity was determined by transducing BHK-21 cells with serial fivefold dilutions of vectors and detecting β-galactosidase expression after 72 hours of culture at 37°C. Plates were fixed with 0.5% glutaraldehyde in phosphate-buffered saline (PBS) for 10 min at room temperature, washed with PBS, and stained with 150 μl of X-Gal solution containing 2 mM $MgCl_2$, 10 mM potassium ferricyanide, 10 mM potassium ferrocyanide, and 5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside (0.8 mg/ml) in PBS for 3 hours at 37°C. Counting was done on an S6 Ultimate Image UV analyzer (CTL Europe), with one blue cell defined as one infectious unit. Cell transduction by FVV is a surrogate for viral infectivity and FVV titers were expressed as infectious units per milliliter.

The yield of FVV particles was estimated by the quantification of particle-associated transgene RNA as described in (*26*). FVV RNAs were extracted from raw cell supernatants with QIAamp Viral RNA Extraction Kit (Qiagen). RNAs were treated with a DNA-free kit (Life Technologies) and retro-transcribed with Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific) using random primers (Thermo Fisher Scientific) according to the manufacturer's instructions. Quantitative polymerase chain reaction (qPCR) was performed on cDNA using BGAL primers (BGAL_F 5′ AAACTCGC AAGCCGACTGAT 3′ and BGAL_R 5′ ATATCGCGGCTCAGTTCGAG 3′) with a 10-min denaturation step at 95°C and 40 amplification cycles (15 s at 95°C, 20 s at 60°C, and 30 s at 72°C) carried out with an Eppendorf Realplex 2 Mastercycler. A standard curve prepared with serial dilutions of pcu2MD9-BGAL plasmid was used to determine the copy number of FVVs. Results were expressed as vector particles per milliliter, considering that each particle carries two copies of the transgene.

The capacity of FVVs to bind to HT1080 cells was tested by incubating cells with FVV particles (1, 10, and 100 particles per cell) on ice for 1 hour. Cells were washed three times with PBS to eliminate unbound FVVs, and RNAs were extracted using RNeasy Plus Mini Kit (Qiagen) according to the manufacturer's protocol. Reverse transcription (RT) was performed as described for FVV RNA quantification. Bound FVVs were quantified by qPCR of *bgal* gene as described for vector titration; cells were quantified by a qPCR amplifying the *hgapdh* gene with the following primers: hGAPDH_F 5′ GGAGCGAGATCCCTCCAAAAT 3′ and hGAPDH_R 5′ GGCTGTTGTCATACTTCTCATGG 3′. The qPCR reaction conditions were the same as those used to amplify the *bgal* gene. Relative mRNA expression of *bgal* versus *hgapdh* was calculated using the ΔΔCt method, and relative binding as 2−ΔΔCt.

Alphafold multimer prediction of full-length FV Env trimer

For the generation of the FV Env model, we used ColabFold v1.5.1 (*72*) with the weights of AlphaFold multimer v3 (*37*). The multiple sequence alignment (MSA) was generated using the mmseqs2 (*73*) and the uniref30 database (*74*), resulting in an MSA of 79 sequences. The template search did not return any matches, so the model was generated from only MSAs as input. All the models were predicted using 20 recycles on an NVIDIA V100 GPU (taking about 5 hours per prediction).

Statistics

The infectious titers, particle concentration, percentages of infectious particles, and a number of bound FVVs carrying WT and E630P Env variants were compared using a two-way paired *t* test, with *P* values indicated on the graph.

Supplementary Materials

This PDF file includes:

Figs. S1 to S9
Tables S1 to S4
References

REFERENCES AND NOTES

1. A. S. Khan, J. Bodem, F. Buseyne, A. Gessain, W. Johnson, J. H. Kuhn, J. Kuzmak, D. Lindemann, M. L. Linial, M. Lochelt, M. Materniak-Kornas, M. A. Soares, W. M. Switzer, Spumaretroviruses: Updated taxonomy and nomenclature. *Virology* **516**, 158–164 (2018).
2. D. M. Pinto-Santini, C. R. Stenbak, M. L. Linial, Foamy virus zoonotic infections. *Retrovirology* **14**, 55 (2017).
3. D. Lindemann, O. Herchenroder, in *Field's Virology*, P. M. Howley, D. M. Knipe, Eds. (Wolters Kluwer, ed. 7, 2022), vol. 3, chap. 20, pp. 679–705.
4. P. Aiewsakun, Avian and serpentine endogenous foamy viruses, and new insights into the macroevolutionary history of foamy viruses. *Virus Evol.* **6**, vez057 (2020).
5. X. Wang, Y. Chen, J. Cui, Identification of cartilaginous fish endogenous foamy virus rooting to vertebrate counterparts. *J. Virol.* **97**, e01816–e01822 (2023).
6. P. Aiewsakun, A. Katzourakis, Marine origin of retroviruses in the early Palaeozoic Era. *Nat. Commun.* **8**, 13954 (2017).
7. E. Simantirakis, I. Tsironis, G. Vassilopoulos, FV vectors as alternative gene vehicles for gene transfer in HSCs. *Viruses* **12**, 332 (2020).
8. J. Kim, G. E. Lee, C. G. Shin, Foamy virus integrase in development of viral vector for gene therapy. *J. Microbiol. Biotechnol.* **30**, 1273–1281 (2020).
9. F. Buseyne, E. Betsem, T. Montange, R. Njouom, C. Bilounga Ndongo, O. Hermine, A. Gessain, Clinical signs and blood test results among humans infected with zoonotic simian foamy virus: A case-control study. *J. Infect. Dis.* **218**, 144–151 (2018).
10. S. C. Harrison, Viral membrane fusion. *Virology* **479–480**, 498–507 (2015).
11. F. A. Rey, S. M. Lok, Common Features of Enveloped Viruses and Implications for Immunogen Design for Next-Generation Vaccines. *Cell* **172**, 1319–1334 (2018).
12. N. McCaul, M. Quandte, I. Bontjer, G. van Zadelhoff, A. Land, E. T. Crooks, J. M. Binley, R. W. Sanders, I. Braakman, Intramolecular quality control: HIV-1 envelope gp160 signal-peptide cleavage as a functional folding checkpoint. *Cell Rep.* **36**, 109646 (2021).
13. T. S. Jardetzky, R. A. Lamb, Activation of paramyxovirus membrane fusion and virus entry. *Curr. Opin. Virol.* **5**, 24–33 (2014).
14. J. M. White, G. R. Whittaker, Fusion of Enveloped Viruses in Endosomes. *Traffic* **17**, 593–614 (2016).
15. M. S. Ladinsky, P. N. Gnanapragasam, Z. Yang, A. P. West, M. S. Kay, P. J. Bjorkman, Electron tomography visualization of HIV-1 fusion with target cells using fusion inhibitors to trap the pre-hairpin intermediate. *eLife* **9**, e58411 (2020).
16. Y. H. Kim, J. E. Donald, G. Grigoryan, G. P. Leser, A. Y. Fadeev, R. A. Lamb, W. F. DeGrado, Capture and imaging of a prehairpin fusion intermediate of the paramyxovirus PIV5. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20992–20997 (2011).
17. D. J. Benton, S. J. Gamblin, P. B. Rosenthal, J. J. Skehel, Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* **583**, 150–153 (2020).
18. B. Apellaniz, N. Huarte, E. Largo, J. L. Nieva, The three lives of viral fusion peptides. *Chem. Phys. Lipids* **181**, 40–55 (2014).
19. S. C. Harrison, Mechanism of membrane fusion by viral envelope proteins. *Adv. Virus Res.* **64**, 231–261 (2005).
20. D. C. Chan, D. Fass, J. M. Berger, P. S. Kim, Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273 (1997).
21. W. Weissenhorn, A. Dessen, S. C. Harrison, J. J. Skehel, D. C. Wiley, Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426–430 (1997).
22. W. Weissenhorn, A. Carfi, K. H. Lee, J. J. Skehel, D. C. Wiley, Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. *Mol. Cell* **2**, 605–616 (1998).
23. M. Nasimuzzaman, D. A. Persons, Cell Membrane-associated heparan sulfate is a receptor for prototype foamy virus in human, monkey, and rodent cells. *Mol. Ther.* **20**, 1158–1166 (2012).
24. K. Plochmann, A. Horn, E. Gschmack, N. Armbruster, J. Krieg, T. Wiktorowicz, C. Weber, K. Stirnnagel, D. Lindemann, A. Rethwilm, C. Scheller, Heparan sulfate is an attachment factor for foamy virus entry. *J. Virol.* **86**, 10028–10035 (2012).
25. A. Duda, D. Luftenegger, T. Pietschmann, D. Lindemann, Characterization of the prototype foamy virus envelope glycoprotein receptor-binding domain. *J. Virol.* **80**, 8158–8167 (2006).
26. I. Fernandez, L. T. Dynesen, Y. Coquin, R. Pederzoli, D. Brun, A. Haouz, A. Gessain, F. A. Rey, F. Buseyne, M. Backovic, The crystal structure of a simian Foamy Virus receptor binding domain provides clues about entry into host cells. *Nat. Commun.* **14**, 1262 (2023).
27. L. T. Dynesen, I. Fernandez, Y. Coquin, M. Delaplace, T. Montange, R. Njouom, C. Bilounga-Ndongo, F. R. Rey, A. Gessain, M. Backovic, F. Buseyne, Neutralization of zoonotic simian foamy viruses: Genotype-specific epitopes within the receptor binding domain. *PLOS Pathog.* **19**, e1011339 (2023).
28. M. Picard-Maureau, G. Jarmy, A. Berg, A. Rethwilm, D. Lindemann, Foamy virus envelope glycoprotein-mediated entry involves a pH-dependent fusion process. *J. Virol.* **77**, 4722–4730 (2003).
29. A. Duda, A. Stange, D. Luftenegger, N. Stanke, D. Westphal, T. Pietschmann, S. W. Eastman, M. L. Linial, A. Rethwilm, D. Lindemann, Prototype foamy virus envelope glycoprotein leader peptide processing is mediated by a furin-like cellular protease, but cleavage is not essential for viral infectivity. *J. Virol.* **78**, 13865–13870 (2004).
30. G. Effantin, L. F. Estrozi, N. Aschman, P. Renesto, N. Stanke, D. Lindemann, G. Schoehn, W. Weissenhorn, Cryo-electron microscopy structure of the native prototype foamy virus glycoprotein and virus architecture. *PLOS Pathog.* **12**, e1005721 (2016).

31. L. Chen, J. J. Gorman, J. McKimm-Breschkin, L. J. Lawrence, P. A. Tulloch, B. J. Smith, P. M. Colman, M. C. Lawrence, The structure of the fusion glycoprotein of Newcastle disease virus suggests a novel paradigm for the molecular mechanism of membrane fusion. *Structure* **9**, 255–266 (2001).
32. K. A. Baker, R. E. Dutch, R. A. Lamb, T. S. Jardetzky, Structural basis for paramyxovirus-mediated membrane fusion. *Mol. Cell* **3**, 309–319 (1999).
33. P. Bork, L. Holm, C. Sander, The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320 (1994).
34. M. Wang, H. Zhang, Q. M. Liu, Y. Sun, Z. Li, W. H. Liu, X. H. He, J. Song, Y. X. Wang, Structure of transmembrane subunits gp47 of the foamy virus envelope glycoproteins. *Acta Virol.* **60**, 181–189 (2016).
35. R. M. Epand, R. F. Epand, Factors contributing to the fusogenic potency of foamy virus. *Biochem. Biophys. Res. Commun.* **284**, 870–874 (2001).
36. T. A. Sergel, L. W. McGinnes, T. G. Morrison, Mutations in the fusion peptide and adjacent heptad repeat inhibit folding or activity of the Newcastle disease virus fusion protein. *J. Virol.* **75**, 7934–7943 (2001).
37. R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, O. Ronneberger, S. Bodenstein, M. Zielinski, A. Bridgland, A. Potapenko, A. Cowie, K. Tunyasuvunakool, R. Jain, E. Clancy, P. Kohli, J. Jumper, D. Hassabis, Protein complex prediction with AlphaFold-Multimer. bioRxiv 2021.10.04.463034 (2022). https://doi.org/10.1101/2021.10.04.463034.
38. J. M. White, A. E. Ward, L. Odongo, L. K. Tamm, Viral membrane fusion: A Dance between proteins and lipids. *Annu. Rev. Virol.* **10**, 139–161 (2023).
39. H. S. Yin, R. G. Paterson, X. Wen, R. A. Lamb, T. S. Jardetzky, Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9288–9293 (2005).
40. J. S. McLellan, M. Chen, S. Leung, K. W. Graepel, X. Du, Y. Yang, T. Zhou, U. Baxa, E. Yasuda, T. Beaumont, A. Kumar, K. Modjarrad, Z. Zheng, M. Zhao, N. Xia, P. D. Kwong, B. S. Graham, Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113–1117 (2013).
41. H. S. Yin, X. Wen, R. G. Paterson, R. A. Lamb, T. S. Jardetzky, Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. *Nature* **439**, 38–44 (2006).
42. A. C. Walls, M. A. Tortorici, J. Snijder, X. Xiong, B. J. Bosch, F. A. Rey, D. Veesler, Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11157–11162 (2017).
43. A. C. Walls, M. A. Tortorici, B. J. Bosch, B. Frenz, P. J. M. Rottier, F. DiMaio, F. A. Rey, D. Veesler, Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **531**, 114–117 (2016).
44. D. Luftenegger, M. Picard-Maureau, N. Stanke, A. Rethwilm, D. Lindemann, Analysis and function of prototype foamy virus envelope N glycosylation. *J. Virol.* **79**, 7664–7672 (2005).
45. M. B. Battles, V. Mas, E. Olmedillas, O. Cano, M. Vazquez, L. Rodriguez, J. A. Melero, J. S. McLellan, Structure and immunogenicity of pre-fusion-stabilized human metapneumovirus F glycoprotein. *Nat. Commun.* **8**, 1528 (2017).
46. H. C. Aguilar, K. A. Matreyek, C. M. Filone, S. T. Hashimi, E. L. Levroney, O. A. Negrete, A. Bertolotti-Ciarlet, D. Y. Choi, I. McHardy, J. A. Fulcher, S. V. Su, M. C. Wolf, L. Kohatsu, L. G. Baum, B. Lee, N-glycans on Nipah virus fusion protein protect against neutralization but reduce membrane fusion and viral entry. *J. Virol.* **80**, 4878–4889 (2006).
47. C. Lambert, M. Couteaudier, J. Gouzil, L. Richard, T. Montange, E. Betsem, R. Rua, J. Tobaly-Tapiero, D. Lindemann, R. Njouom, A. Mouinga-Ondeme, A. Gessain, F. Buseyne, Potent neutralizing antibodies in humans infected with zoonotic simian foamy viruses target conserved epitopes located in the dimorphic domain of the surface envelope protein. *PLOS Pathog.* **14**, e1007293 (2018).
48. A. Krarup, D. Truan, P. Furmanova-Hollenstein, L. Bogaert, P. Bouchier, I. J. Bisschop, M. N. Widjojoatmodjo, R. Zahn, H. Schuitemaker, J. S. McLellan, J. P. Langedijk, A highly stable prefusion RSV F vaccine derived from structural analysis of the fusion mechanism. *Nat. Commun.* **6**, 8143 (2015).
49. M. Porotto, I. Devito, S. G. Palmer, E. M. Jurgens, J. L. Yee, C. C. Yokoyama, A. Pessi, A. Moscona, Spring-loaded model revisited: Paramyxovirus fusion requires engagement of a receptor binding protein beyond initial triggering of the fusion protein. *J. Virol.* **85**, 12867–12880 (2011).
50. T. Wilk, V. Geiselhart, M. Frech, S. D. Fuller, R. M. Flugel, M. Lochelt, Specific interaction of a novel foamy virus Env leader protein with the N-terminal Gag domain. *J. Virol.* **75**, 7995–8007 (2001).
51. J. Tobaly-Tapiero, P. Bittoun, M. L. Giron, M. Neves, M. Koken, A. Saib, H. de The, Human foamy virus capsid formation requires an interaction domain in the N terminus of Gag. *J. Virol.* **75**, 4367–4375 (2001).
52. P. Guardado-Calvo, F. A. Rey, The viral class II membrane fusion machinery: Divergent evolution from an ancestral heterodimer. *Viruses* **13**, 2368 (2021).
53. L. Rutten, Y. T. Lai, S. Blokland, D. Truan, I. J. M. Bisschop, N. M. Strokappe, A. Koornneef, D. van Manen, G. Y. Chuang, S. K. Farney, H. Schuitemaker, P. D. Kwong, J. P. M. Langedijk, A universal approach to optimize the folding and stability of prefusion-closed HIV-1 envelope trimers. *Cell Rep.* **23**, 584–595 (2018).
54. P. B. Rosenthal, X. Zhang, F. Formanowski, W. Fitz, C. H. Wong, H. Meier-Ewert, J. J. Skehel, D. C. Wiley, Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* **396**, 92–96 (1998).
55. P. Aiewsakun, L. Richard, A. Gessain, A. Mouinga-Ondémé, P. V. Afonso, A. Katzourakis, Modular nature of simian foamy virus genomes and their evolutionary history. *Virus Evol.* **5**, vez032 (2019).
56. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
57. H. Ebel, T. Benecke, B. Vollmer, Stabilisation of viral membrane fusion proteins in prefusion conformation by structure-based design for structure determination and vaccine development. *Viruses* **14**, 1816 (2022).
58. H. Qiao, S. L. Pelletier, L. Hoffman, J. Hacker, R. T. Armstrong, J. M. White, Specific single or double proline substitutions in the "spring-loaded" coiled-coil region of the influenza hemagglutinin impair or abolish membrane fusion activity. *J. Cell Biol.* **141**, 1335–1347 (1998).
59. R. W. Sanders, J. P. Moore, Virus vaccines: Proteins prefer prolines. *Cell Host Microbe* **29**, 327–333 (2021).
60. R. Rua, E. Betsem, S. Calattini, A. Saib, A. Gessain, Genetic characterization of simian foamy viruses infecting humans. *J. Virol.* **86**, 13350–13359 (2012).
61. T. Krey, J. d'Alayer, C. M. Kikuti, A. Saulnier, L. Damier-Piolle, I. Petitpas, D. X. Johansson, R. G. Tawar, B. Baron, B. Robert, P. England, M. A. Persson, A. Martin, F. A. Rey, The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. *PLOS Pathog.* **6**, e1000762 (2010).
62. M. Backovic, T. Krey, Stable drosophila cell lines: An alternative approach to exogenous protein expression. *Methods Mol. Biol.* **1350**, 349–358 (2016).
63. T. Nakane, D. Kimanius, E. Lindahl, S. H. Scheres, Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).
64. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
65. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
66. V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
67. S. Q. Zheng, E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
68. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
69. R. Sanchez-Garcia, J. Gomez-Blanco, A. Cuervo, J. M. Carazo, C. O. S. Sorzano, J. Vargas, DeepEMhancer: A deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
70. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
71. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkoczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
72. M. Mirdita, K. Schutze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
73. M. Steinegger, J. Soding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
74. B. E. Suzek, Y. Wang, H. Huang, P. B. McGarvey, C. H. Wu, C. UniProt, UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).
75. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
76. W. Shi, Y. Cai, H. Zhu, H. Peng, J. Voyer, S. Rits-Volloch, H. Cao, M. L. Mayer, K. Song, C. Xu, J. Lu, J. Zhang, B. Chen, Cryo-EM structure of SARS-CoV-2 postfusion spike in membrane. *Nature* **619**, 403–409 (2023).
77. S. V. Strelkov, P. Burkhard, Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J. Struct. Biol.* **137**, 54–64 (2002).
78. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

79. P. Gouet, E. Courcelle, D. I. Stuart, F. Metoz, ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
80. L. Richard, R. Rua, E. Betsem, A. Mouinga-Ondeme, M. Kazanji, E. Leroy, R. Njouom, F. Buseyne, P. V. Afonso, A. Gessain, Cocirculation of two *env* molecular variants, of possible recombinant origin, in gorilla and chimpanzee simian foamy virus strains from Central Africa. *J. Virol.* **89**, 12480–12491 (2015).
81. T. A. Galvin, I. A. Ahmed, M. Shahabuddin, T. Bryan, A. S. Khan, Identification of recombination in the envelope gene of simian foamy virus serotype 2 isolated from *Macaca cyclopis*. *J. Virol.* **87**, 8792–8797 (2013).
82. S. S. Guo, J. Liu, X. G. Zhou, G. J. Zhang, DeepUMQA: Ultrafast shape recognition-based protein model quality assessment using deep learning. *Bioinformatics* **38**, 1895–1903 (2022).
83. W. L. DeLano. The PyMOL Molecular Graphics System. (DeLano Scientific, 2002).

Acknowledgments: We thank the staff from the Nanoimaging Core Facility at Institut Pasteur for technical assistance during EM data collection. We are grateful to J. Hellert and P. G. Calvo for the discussions and advice and for the critical reading of the manuscript. M.B. dedicates this manuscript to the memory of R. Lamb for advice over the years and for his guidance while M.B. served as a postdoc in the Jardetzky laboratory. **Funding:** This work was supported by the Integrative Biology of Emerging Infectious Diseases Laboratory of Excellence (ANR-10-LABX62-IBEID, Intra-Labex Grant, to M.B.) and "Programme de recherche transversal from Institut Pasteur" (PTR2020-353 ZOOFOAMENV, Florence Buseyne). The funding agencies had no role in the study design, generation of results, or writing of the manuscript. **Author contributions:** Conceptualization: M.B., I.F., F.A.R., A.G., and C.A.G. Investigation: I.F., Y.C., F.Bu., D.B., and C.A.G. Methodology: I.F., B.E.C., and C.A.G. Validation: I.F., M.B., F.A.R., F.Bu., and C.A.G. Visualization: M.B., I.F., F.Bu., and C.A.G. Formal analysis: I.F., F.Bo., and C.A.G. Data curation: M.B. and C.A.G. Resources: M.B., F.A.R., D.B., B.E.C., and F.Bu. Supervision: M.B., I.F., F.A.R., and B.E.C. Software: F.Bo., B.E.C., and C.A.G. Writing—original draft: I.F. and M.B. Writing—review and editing: M.B., I.F., A.G., and F.Bu. Project administration: M.B., I.F., and F.A.R. Funding acquisition: M.B., F.Bu., F.A.R., and B.E.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The coordinates for the cryo-EM structural models have been deposited to the RCSB protein databank under PDB accession codes 8RM0 (pre-fusion Env) and 8RM1 (post-fusion Env). The EM maps have been deposited to the EMDB under accession codes EMD-19347 (C3) and EMD-50530 (C1) for the pre-fusion Env, and EMD-19348 (C3) and EMD-50529 (C1) for the post-fusion Env. The Alphafold multimer model of the FV Env trimer in the pre-fusion state is available in ModelArchive (modelarchive.org) with the accession code ma-q5jmc.

Submitted 20 February 2024
Accepted 6 September 2024
Published 11 October 2024
10.1126/sciadv.ado7035