Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The crystal structure of a simian Foamy Virus receptor binding domain provides clues about entry into host cells

Ignacio Fernández
Institut Pasteur
Lasse Toftdal Dynesen
https://orcid.org/0000-0003-3881-7131
Youna Coquin
Institut Pasteur
Riccardo Pederzoli
Institut Pasteur
Delphine Brun
Institut Pasteur
Ahmed Haouz
Plate-forme Cristallographie, Institut Pasteur https://orcid.org/0000-0003-1196-1635
Antoine Gessain
Institut Pasteur
Félix Rey
Institut Pasteur https://orcid.org/0000-0002-9953-7988
Florence Buseyne
Institut Pasteur https://orcid.org/0000-0002-9323-7837
Marija Backovic (marija.backovic@pasteur.fr)
Institut Pasteur https://orcid.org/0000-0001-8814-4428

Article

Keywords:

Posted Date: October 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2120175/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

The crystal structure of a simian Foamy Virus receptor binding domain provides clues about entry into host cells

Ignacio Fernández[1], Lasse Toftdal Dynesen[2], Youna Coquin[2], Riccardo Pederzoli[1], Delphine Brun[1], Ahmed Haouz[3], Antoine Gessain[2], Felix A. Rey[1], Florence Buseyne[2], Marija Backovic[1*]

[1]Institut Pasteur, Université Paris Cité, CNRS UMR3569, Unité de Virologie Structurale, 75015 Paris, France.

[2]Institut Pasteur, Université Paris Cité, CNRS UMR3569, Unité d'Epidémiologie et Physiopathologie des Virus Oncogènes, 75015 Paris, France.

[3]Institut Pasteur, Université Paris Cité, Plateforme de cristallographie-C2RT, CNRS UMR 3528, 75015 Paris, France.

[*]Corresponding author: Marija Backovic, marija@pasteur.fr

Abstract

The surface envelope glycoprotein (Env) of all retroviruses mediates virus binding to cells and fusion of the viral and cellular membranes. A structure-function relationship for the HIV Env that belongs to the Orthoretrovirus subfamily has been well established. Structural information is however largely missing for the Env of Foamy viruses (FVs), the second retroviral subfamily. FV Envs lack sequence similarity with their HIV counterpart. We present the X-ray structure of the receptor binding domain (RBD) of a simian FV Env at 2.6 Å resolution, revealing two subdomains and an unprecedented fold. We have generated a model for the organization of the RBDs within the trimeric Env which indicates that the upper subdomain is important for stabilization of the full-length Env, and have demonstrated that residues K342, R343, R359 and R369 in the lower subdomain play key roles in the interaction of the RBD and viral particles with heparan sulfate.

Introduction

Spumaretroviruses, also known as foamy viruses (FVs) are ancient retroviruses that have co-evolved with vertebrate hosts for over 400 million years [1,2]. FVs are prevalent in nonhuman primates, which can transmit them to humans, most often through bites [3]. Unlike their better-studied *Orthoretrovirinae* relatives (HIV being the most notable member) FVs have extremely slowly mutating genomes and do not induce severe pathologies despite integrating into the host genome and establishing lifelong persistent infections [4,5]. These features, along with broad tropism and host range [6], make FVs attractive vector candidates for gene therapy [7].

Viral fusion proteins drive membrane fusion by undergoing a conformational change, which can be triggered by acidification in an endosomal compartment and / or binding to a specific cellular receptor [8,9]. FVs enter cells by endocytosis, with fusion of the viral and cellular membranes occurring in the endosomal compartment in a pH-sensitive manner, leading to capsid release into the cytosol [10]. The exception is the prototype FV (PFV) which can also fuse at the plasma membrane [11]. The FV fusion protein, the envelope glycoprotein (Env), exhibits the organization of a class I fusogen [12], which are synthesized as single-chain precursors and fold into trimers, within which protomers are subsequently cleaved in secretory Golgi compartments. FV Env is cleaved twice by cellular furin during maturation, giving rise to 3 fragments: the leader peptide (LP), the surface (SU) subunit, which has the receptor binding domain (RBD), and the transmembrane subunit (TM), which harbors the fusion machinery. The structural information available for FV Env is limited to cryo-electron tomography (ET) of viral particles and a 9 Å cryo-electron microscopy (EM) reconstruction of PFV Env [13], which revealed LP-SU-TM trimers arranged in interlocked hexagonal assemblies [13,14], with an architecture that is different to that of HIV Env trimers [15].

Heparan sulfate (HS) is as an attachment factor for PFV and feline FV [16,17] but the requirements for a surface or intracellular receptor, which would trigger membrane fusion by FV Env, remain unclear. The search for a receptor has been complicated by the FV binding to HS, which is ubiquitously expressed on cells, masking potential candidates. A bipartite RBD, consisting of two discontinuous regions of the polypeptide chain, was identified within the FV SU by screening a panel of recombinant SU truncations for binding to cells [18]. FV Env is heavily glycosylated, with at least 13 predicted N-linked glycosylation sites. Mutational analysis has revealed that three of these N-sites are essential for the PFV infectivity – 2 located in the TM subunit, and one in the RBD. The latter

site, referred to as the glycosylation site 8 or N8 [19] is conserved across the FV subfamily, and has been suggested to play a direct role in binding to a receptor [18] (to distinguish the nomenclature of the predicted N-linked glycosylation sites (N1 to N15) from the single letter symbol for asparagine residues (N), the former will be underlined throughout the text). The remaining molecular determinants of the RBD interaction with host cells remain elusive, largely due to a lack of structural information, which has precluded rational approaches to mutagenesis and functional analyses. A high-resolution structure of the FV RBD, structural information regarding the RBDs organization within the Env trimer and how the RBDs contribute to the Env activation are not available. In this manuscript we present the first X-ray structure of the RBD from a zoonotic gorilla simian FV at 2.6 Å resolution, which reveals an entirely novel fold. We propose a model for the RBD assembly in the trimeric Env and report the identification of residues involved in HS binding, with functional and evolutionary implications discussed.

Results

The X-ray structure of the SFV RBD reveals a novel fold

Recombinant RBDs from several simian FV (SFV) strains were tested for production in *Drosophila* S2 insect cells, and only the RBD from gorilla SFV (strain SFVggo_huBAK74[20], genotype II; abbreviated as 'GII' herein) was both expressed in the quantities required for structural studies and also crystallized. The RBD was expected to be heavily glycosylated due to 8 predicted N-glycosylation sites (Fig. 1A). To increase the chances of generating well-diffracting crystals, a fraction of the purified protein was enzymatically deglycosylated (RBD^D). Crystals were obtained for the RBD^D as well as for the untreated protein (RBD^G). The RBD^D diffracted better (2.6 Å) than RBD^G and the structural analyses presented below were carried out using the RBD^D structure, unless otherwise noted. The data collection and structure determination statistics for both crystal forms are summarized in Table S1.

The SFV RBD folds into two subdomains, each with $\alpha+\beta$ topology, which we refer to as 'lower' (residues 218-245, 311-369 and 491-524) and 'upper' subdomains (residues 246-310 and 370-490) in reference to their positioning with respect to the viral membrane[13] (see below). The overall RBD fold approximates a ~65 Å long bean-shape, with the upper subdomain on the wider side (~45 Å diameter), and the N- and C-termini on the opposite, narrower side (~20 Å diameter)

(Fig. 1B). The lower subdomain is comprised of a three-helical bundle (α1, α7, α8) that packs against an anti-parallel, twisted four-stranded β-sheet (β14-β1-β5-β15) and against helix α2 (residues 333-346), which lays perpendicularly on the side of the bundle. Within the helical bundle and the β-sheet, the regions proximal to the N- and C- termini are tied together, and each structure is reinforced by disulfide bonds (DS) DS1 (C228-C503) and DS2 (C235-C318), respectively. Seventy- and 130-residue long segments, forming most of the upper subdomain, are inserted between β1 and β5, and between η4 and β14 of the lower subdomain, respectively (Fig. 2). The polypeptide chain extends upwards and back twice, finally yielding the outer strands of the β-sheet (β14 and β15). These secondary structural elements in the lower subdomain contain a prominent hydrophobic core that extends into the upper subdomain, which has lower secondary structure content (Table S2) and is stabilized by several networks of polar interactions (Fig. S1, Table S3). Four notable protrusions, designated loops 1 to 4 (L1-L4) emanate from the upper domain: loop 1 (L1, residues 253-270, connecting β2 and η1), loop 2 (L2, residues 276-281, connecting η1 and β3), loop 3 (L3, residues 414-436, connecting α6 and β9) and loop 4 (L4, residues 446-453, connecting β10 and β11). The loops L3 and L4 are particularly mobile in our structures as indicated by high B-factors (> 105 Å^2) for their Cα atoms (Fig. S2). Electron density was observed for 7 out of the 8 predicted N-glycosylation sites (Fig. 1C), allowing the modelling of at least one N-acetyl glucosamine (NAG) at each site (Fig. 2, Fig. S3).

Searching the PDB databank using the DALI algorithm[21], with the RBD and its substructures did not yield any meaningful results. Comparative analyses with the available structures of the RBDs from Orthoretroviruses (Fig. S4) did not reveal structural similarity, either at the level of the secondary structure topology or the three-dimensional fold. Therefore, the SFV RBD represents, to the best of our knowledge, an unprecedented fold.

The sugar attached to the strictly conserved 8th N-glycosylation site (N8) plays a structural role

There were no major differences between the X-ray structures of RBD^D and RBD^G (their superposition yielded a root mean square deviation (rmsd) below 1Å (Fig. S3B)), except for the different number of sugar units that we could built into the electron density maps (Fig. S3A). A prominent feature of the upper subdomain is the eighth N-linked sugar (N8)[18] attached to the α4 helix residue N390. The N390 and the first two NAG residues are buried in the RBD rendering the

EndoH/D cleavage site inaccessible (Fig. 3B), which allowed building of 10 sugars in the RBD^G and 8 in the deglycosylated protein crystals (Fig. 2, Fig. 3A). The N8 glycan emerges from a cavity that has N390 at its base and extends upwards, remaining in contact with the protein and preserving the same conformation in both crystal forms. Structural analyses revealed that the glycan establishes extensive van der Waal contacts with the residues underneath (buried surface area of 803 $Å^2$) and forms hydrogen bonds with main chain atoms from Y394 and I484 and the side chain of E361 (Fig. S6). The oligosaccharide covers a well-conserved and hydrophobic surface (Fig. 3C) and thus maintains the RBD fold and prevents aggregation, consistent with the reported misfolding and low levels of the secreted immunoadhesin carrying the RBD with a mutation in the N8 site[18]. The N8 is the only N-glycosylation site in the SU that is strictly conserved across the FV subfamily (Fig. S5), and the hydrophobic patch residues laying beneath it are conserved as well (Fig. 4C). Thus, N8 likely plays an important structural role in all FV RBDs.

The RBD fold is predicted to be conserved within the *Spumaretrovirinae* subfamily

To investigate potential conformational differences between RBDs from different species, we used AlphaFold (AF) [22] software to predict the RBD structures from members of each of the 5 FV genera, some of which exist as two genotypes due to the modular nature of FV Env [23]. Within each FV Env, a ~250-residue long region within the RBD, termed the variable or 'SU^{var}' region, defines two co-circulating genotypes , I and II, which have been found in gorillas [24], chimpanzees[25] and mandrills[23], among others. The SU^{var} regions share less than 70% identity (Fig. S5), while the rest of Env residues are highly conserved (>95% sequence identity). The SU^{var} is located within the upper subdomain and encompasses loops L1-L4 (residues 282-487 in GII RBD; Fig. S7).

All the generated models have high confidence metrics (Fig. S8) and display a conserved fold in agreement with a sequence identity >30%. Significant deviations were found only in the loops within the SU^{var}. The Template Modelling score (TM-score), which is, unlike rmsd, a length-independent measure of structural similarity [26] has the average value of 0.891 for the 11 compared structures. The AF model of the GII RBD and our experimentally determined structure of the same strain superimpose with a TM-score of 0.96 and rmsd of 1.5 Å for 320 out of 328 Cα atoms aligned, confirming the high accuracy of the AF model. A common core' (CC), which includes the ensemble of residues with Cα rmsd values smaller than 4Å for all the pairwise superpositions, was calculated by the mTM-align webserver [27]. The CC of the FV RBD contains 239 out of 308 aligned residues (Fig.

S9), with most CC residues belonging to the secondary structure elements forming the lower subdomain. The loops in the upper subdomain are largely not a part of the CC (Fig. S9).

Fitting of the RBD atomic model into Env cryo-EM density maps reveals the trimeric RBD arrangement

To investigate the RBD arrangement within trimeric Env, we fitted the RBD atomic model into the 9Å cryo-EM map reported for trimeric PFV (a chimpanzee genotype I FV) Env expressed on foamy viral vector (FVV) particles [13]. The fitting was justified by the high structural conservation between gorilla and chimpanzee RBDs, indicated by a TM-score of 0.88 for the superposition of the GII RBD structure and the predicted PFV RBD model (Fig. S8).

The RBD fitting was performed with the fit-in-map function in Chimera suite [28]. The correlation coefficient of 0.96 strongly suggests that the recombinantly expressed RBD represents its biologically relevant conformation as observed at the surface of virus particles. The three RBDs are arranged around a central cavity at the apex (membrane-distal region) of Env (Fig. 4A). The analyses of the macromolecular surfaces of the trimeric RBD model, carried out in PDBePISA [29], revealed a limited inter-protomer interface (<10% of the entire RBD solvent accessible surface) established solely by loops L1, L3 and L4 that form a ring-like structure at the RBD apex, leaving most of the RBD exposed (Fig. 4B). The three L1 loops engage in homotypic interactions at the center of the RBD, forming an inner ring, while each L3 loop contacts the L4 of a neighboring protomer, further stabilizing the interface. The sequences of the interacting loops are poorly conserved within the FV subfamily (Fig. S10).

The 7 N-linked glycans that we could resolve in the RBD structure are all fully solvent-exposed (Fig. 4C). The RBD N- and C- termini point towards the membrane, indicating that the lower half of the Env density is occupied by the TM subunit and the remaining SU residues, as previously suggested [13].

Positively charged residues in the lower subdomain form a heparan sulfate (HS) binding site

To locate potential HS binding regions with RBD, we investigated electrostatic potential surface distribution and identified a large, continuous region in the lower subdomain with strong positive potential (Fig. 5A). We next analyzed the RBD structure with the ClusPro server that predicts putative HS binding sites on protein surfaces [30]. K342 and R343 in α2 helix, R359 in the proceeding

184 η4 helix and R369 in an extended chain region were among the residues that had the highest
185 number of contacts with HS models that were docked onto the surface (Fig. 5B, 5C). The four
186 residues also mapped within the positively-charged region in the lower subdomain.

187 Based on the ClusPro predictions, we produced two GII RBD variants (K342/R343, termed
188 'mut1', and R359/R369, termed 'mut2') and tested their binding to HS immobilized on a Sepharose
189 matrix (Fig. 6A). The two RBD variants eluted at the same volume on size exclusion chromatography
190 consistent with the expected size of a monomer (Fig. S11), indicating that the introduced mutations
191 did not cause protein misfolding. The WT RBD was retained on the column and eluted at 300 mM
192 sodium chloride concentration, while mut1 and mut2 variants were not retained and eluted in the
193 flow-through fraction. The observed loss of heparin binding capacity strongly suggests that residues
194 K342, K343, R359, and R369 are directly involved in interactions with HS.

195 We used flow cytometry to investigate the interaction between the GII RBD and HS on cells
196 (Fig. S12A). We found that the monomeric RBD did not bind to HT1080 cells even at high protein
197 concentrations (Fig. 6B). We therefore tested a longer construct, the GII Env ectodomain, which
198 spontaneously form trimers, hypothesizing that an oligomer would yield higher signal due to avidity
199 effects. The trimeric ectodomain bound to HT1080 cells (Fig. 6B), and the mutations K342A/R343A
200 and R356A/R369A (mut1 and mut2, respectively) were introduced into the ectodomain background
201 to render them suitable for flow cytometry experiments. We assayed the binding of the Env
202 ectodomains to HT1080 and to BHK-21 cells (Fig. S12B), which are susceptible to infection by gorilla
203 FVs [31]. We quantified HS expression levels by flow cytometry concomitantly with the binding
204 experiments, and verified that BHK-21 cells expressed lower HS levels than HT1080 cells, as had
205 been reported [17] (Fig. S12C). The HS expression levels were 10 to 30-fold lower on BHK-21 cells
206 compared to HT1080 cells (Fig. 6C). The binding of the WT ectodomain was lower on BHK-21 than
207 to HT1080 cells at the highest protein concentrations tested. The binding signal was dose-
208 dependent and one log lower for mut1 and mut2 ectodomain variants in comparison with the WT
209 protein on both cell lines (Fig. 6C).

210 To prove that the designed mutations specifically affected the interaction with cellular HS,
211 we measured binding to HT1080 cells that were pre-treated with heparinase, which removed more
212 than 90% of HS from the cells (Fig. S12D). Binding of the WT ectodomain to heparinase-treated cells
213 was diminished about 100-fold when compared to buffer-treated cells, while the binding of the
214 mut1 and mut2 variants was not affected by heparinase treatment (Fig. 6D).

The importance of residues K342, R343, R356, R369 for viral was tested using FVVs that express either GII WT, mut1 or mut2 Env on their surface. The total number of FVV particles released by the transfected cells, measured by RT-qPCR, was ~6-fold lower for the mut1 compared to WT FVVs, while mut2 had the same particle production as WT (Fig. S13A). The infectious titers were 34- and 65-fold lower for mut1 and mut2, respectively, compared to WT (Fig. S13B). The proportion of infectious particles (the infectious titer, Fig. S13B) divided by the total number of FVVs (Fig. S13A) was 0.7% for WT Env FVVs, while the values for mut1 and mut2 FVVs were 3- and 22-fold lower, respectively (Fig. 6E). We measured the binding of FVVs to cells by RT-qPCR, and found that the binding was also reduced 3- and 23-fold for FVVs carrying mut1 and mut2 Envs, respectively, compared to the WT Env FVVs (Fig. 6F). Thus, binding to cells and entry levels were decreased to the same extent for the FVVs carrying Env proteins with mutations in the HS binding site.

The results described for the recombinant Env proteins and FVVs carrying full-length Env agree with the biochemical data (Fig. 6A) and demonstrate that residues K342, R343, R356, R369 play a crucial role in virus interaction with HS.

Discussion

FV RBD adopts a novel fold and is composed of two subdomains. We determined the X-ray structure of the RBD from a gorilla FV, revealing a fold (Fig. 1, Fig. 2) distinct from the available orthoretroviral RBD structures i.e. from the Friend murine leukemia virus, feline leukemia virus, human endogenous retrovirus EnvP(b)1 (gammaretrovirus genus) [32-34], and gp120 from HIV[35] (lentivirus genus) (Fig. S4). This finding expends the repertoire of unique FV features (assembly, particle release [36], replication [37]) that are not shared with Orthoretroviruses, and is consistent with the lack of Env sequence conservation between Orthoretroviruses and FVs.

The gammaretroviral RBDs are relatively small (~200 residues) and fold into an antiparallel β-sandwich with two extended loops that give rise to a helical subdomain that sits on top of the sandwich [32,33] (Fig. S4). The helical subdomain defines the tropism for cellular receptors [38] and shows high sequence variability within the genus. HIV interacts with its cognate receptor CD4 through gp120, its SU, which is larger (~450 residues) and has two subdomains, inner and outer. The receptor binding surface of gp120 is formed by secondary structure elements from both subdomains [35]. Variable loops project out from the gp120 core, and participate in receptor binding and immune

invasion [39]. It is possible to argue that the FV RBD global organization into two subdomains - the lower, which is better conserved, and upper, which contains the protruding loops and is variable in sequence - is reminiscent of the characteristics described above for the Orthoretrovirus RBDs. Whether the presence of similar features implies similar function remains to be investigated.

The RBDs form a cage-like structure at the membrane-distal side of Env. We fitted the experimentally determined RBD structure into the low-resolution density map (Fig. 4A) obtained by cryo-EM single particle reconstruction of trimeric PFV Env expressed on FVV particles [13]. The resulting model of the RBD trimer arrangement is consistent with the biochemical and functional data presented here; as expected, the HS binding residues (K342, R343, R356, R369) and 7 N-linked carbohydrates map to the Env surface that is exposed to the solvent (Fig. 4C). According to our model, the L1-L4 loops, located at the top of the upper subdomain of each protomer, form the inter-protomer interface (Fig. S10) leaving, just below, a cavity that was clearly observable in the cryo-EM maps [13]. The loop sequences are poorly conserved across the FV family (Fig. S10), and the superpositions of the AF models of 11 FV RBDs reveal only slight structural differences, which are mostly limited to this variable region containing the loops (Fig. S8, Fig. S9). Based on these observations we speculate that the inter-protomer interactions mediated by the loops maintain a closed Env conformation, stabilizing the Env protein in its metastable pre-fusion form that is displayed on the virion surface. It is possible to envisage that different sets of interacting residues in the loops of different FVs can stabilize the RBDs within the trimeric Env sufficiently well to maintain its native state. We hypothesize that the RBDs are loosely bound within native Env to enable them to readily dissociate upon a fusion trigger, which could be delivered in the endosome (acidic pH) and / or by a specific cellular receptor. In that respect, the FV upper domain loops could play a role equivalent to V1/V2/V3 loops in HIV Env [40]. It will be also important to discern the RBD molecular determinants, if any, that drive the membrane fusion at the plasma membrane, as used by PFV, in comparison to all the other FVs that fuse in the endosomes [11].

The structure explains why upper RBD domain is tolerant to deletions or substitutions. Based on the ability of Env truncated variants to bind to cells, Duda *et al.* defined the RBD of PFV Env as a region spanning residues 225-555[18] (residues 226-552 in gorilla GII Env (Fig. S14A)). Within the

proposed region, the terminal segments were found to be essential for the RBD to retain its cell binding activity, while the central region was dispensable. This RBD region, which is not essential for binding to cells [18], termed also RBDjoin (Dynesen *et al.* accompanying manuscript), maps to the top of the RBD, is clamped by two intra-region DS bonds, and encompasses L3 and L4 (Fig. S5). Its location, away from the HS binding residues, is consistent with the ability of the PFV SU truncation lacking the "non-essential" region to bind to cells at the levels measured for WT protein[18]. The AF model of the PFV RBD lacking this RBDjoin region reveals a 3D fold very similar to that of the complete RBD (Fig. S14B). Considering the RBD arrangement within trimeric Env, which we show is maintained by the loops in the RBDjoin region (Fig. 3), we hypothesized that the loop deletions, while tolerated by isolated, monomeric RBD, would likely be detrimental for the integrity of the trimeric Env and its function in fusion and entry. Dynesen *et al.* indeed show that the viruses with Env carrying deletions in loop regions lose infectivity (accompanying paper).

The lower RBD subdomain carries the residues involved in HS binding. Our data demonstrate that K342/R343 and R356/R369 are the key residues for the RBD interaction with HS immobilized on an inert matrix or expressed on cells (Fig. 6) and that HS is an attachment factor for gorilla FV, expanding upon previous reports for PFV[17]. In SFV Envs, the residue at position equivalent to 343 in gorilla GII Env is always an arginine or lysine, while arginine is strictly conserved at position 356 (Fig. S5). Residues at positions 342 and 369 are less conserved among FV Envs, although they are usually surrounded by positive or polar residues. This suggests that the R343 and R356 may be important for HS binding in all FVs, while other positively charged residues, specific to each virus, could be dispersed in the patch with high positive electrostatic potential (Fig. 5A) contributing to the HS binding in a virus-specific context. Existence of an FV receptor had been proposed by Plochmann *et al.* since a total lack of HS did not abolish FV infection, but HS has also been proposed to function as a true FV receptor [16]. The residual Env binding to cells devoid of HS, which we observed both for the WT and the HS-binding impaired variants (Fig. 6D), is consistent with the existence of another cell receptor(s) in FV entry. The HS-binding defective Env variants we generated will be useful tools in the research of the potential proteinaceous receptor, as they eliminate binding to HS, which is a widely expressed attachment factor.

Concluding remarks. In this manuscript we have described the first X-ray structure of a FV RBD and validated that the novel fold is the one adopted in the native Env. We identified, within the RBD, two subdomains in terms of their structure, conservation, and function: the upper subdomain, which encompasses the majority of the genotype-specific region, and is likely involved in maintaining the closed prefusion Env conformation, and a more conserved, lower subdomain, important for binding to the attachment factor HS. We generated AF models for 11 additional FV RBDs, highlighting its conserved three-dimensional conformation. This information is critical for understanding virus-cell interactions and provides a framework for structure-driven mutagenesis studies necessary for establishing the molecular basis of FV entry and recognition by neutralizing antibodies as described in Dynesen *et al.* (accompanying manuscript). The AlphaFold [22] algorithm cannot predict the arrangement of oligosaccharides at the surface of glycoproteins. The previously reported functional observations on FV Envs, along with the role of N8 can now be understood in light of the experimentally derived structure, underscoring the necessity for structure determination by experimental means. Identification of HS binding residues will aid the search for additional putative FV receptor(s). Insights into the structure-function relationship of the metastable, multimeric and heavily glycosylated FV Env, as well as unraveling the molecular basis of receptor activation and membrane fusion, will require integrated biology efforts and experimental structural methods.

Methods

Expression construct design (SFV RBD and ectodomains for HS binding studies)

A flow-cytometry assay was developed by Duda *et al.* to detect binding of recombinantly expressed foamy virus Env variants to cells [18]. By using a panel of SU truncations fused to the Fc region of murine IgG (immunoadhesins) the authors showed that the RBD – defined as the minimal region of the PFV Env sufficient for binding to cells - encompassed residues 225 to 555 (corresponding to residues 226 to 552 in gorilla FV RBD (GII-K74 strain, accession number JQ867464)[41] (Fig. S5). When designing the expression construct for SFV RBD, we also considered the secondary prediction generated by the Phyre2 webserver [42]. Residue I225 was in the middle of a putative helix (residues 220-230), leading us to choose an upstream residue R218 as the N-terminus of the construct (Fig. 1A, Fig. S5).

The Phyre2 webserver was also used to design the ectodomain construct, which starts after the first predicted transmembrane helix (S91) and encompasses residues up to I905.

Recombinant SFV RBD and ectodomain production and purification

For structural studies the RBD (residues 218-552, GII-K74 strain, Env accession number JQ867464)[43] was cloned into a modified pMT/BiP insect cell expression plasmid (Invitrogen) designated pT350, which contains a divalent-cation inducible metallothionein promoter, the BiP signal peptide at the N-terminus (MKLCILLAVVAFVGLSLG), and a double strep tag (DST) (AGWSHPQFEKGGGSGGGSGGGSWSHPQFEK) at the C-terminus, as previously described [43]. This plasmid was co-transfected to *Drosophila* Schneider line 2 cells (S2) with the pCoPuro plasmid for puromycin selection[44]. The cell line has undergone selection in serum-free insect cell medium (HyClone, GE Healthcare) containing 7 μg/ml puromycin and 1% penicillin/streptomycin. For the protein production stage, the cells were grown in spinner flasks until the density reached approximately 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 supernatant was concentrated and used for affinity purification using a Streptactin column (IBA). Approximately 20 milligrams of recombinant RBD were obtained per liter of S2 cell culture. The DST was removed by incubating the protein with 64 units of Enterokinase light chain (BioLabs) in 10 mM Tris-HCl, 100 mM NaCl, 2 mM

354 $CaCl_2$, pH 8.0, at room temperature, overnight. The proteolysis reaction was buffer exchanged into
355 10 mM Tris-HCl, 100mM NaCl, pH 8.0, and subjected to another affinity purification, recovering the
356 flow-through fraction containing the untagged RBD. The protein was concentrated and its enzymatic
357 deglycosylation with EndoD and EndoH was set up at room-temperature following overnight
358 incubation with 1000 units of each glycosidase in 50 mM Na-acetate, 200 mM NaCl, pH 5.5. The
359 protein was further purified on a size exclusion chromatography (SEC) column Superdex 200 16/60
360 (Cytiva) in 10 mM Tris-HCl, 100 mM NaCl, pH 8.0, concentrated in VivaSpin concentrators to 8.2
361 mg/ml and used as such for crystallization trials.

362 For cell binding experiments the RBD construct was cloned in a pcDNA3.1(+) derived plasmid,
363 for expression in mammalian cells. The expression plasmid was modified by inserting a CMV exon-
364 intron-exon sequence that increases the expression of recombinant proteins. The RBD was cloned
365 downstream of the CD5 signal peptide (MPMGSLQPLATLYLLGMLVASCLG) with an enterokinase
366 cleavage site and a DST tag in the C-terminus. The HS mutants were generated by site-directed
367 mutagenesis. The plasmids coding for the recombinant proteins were transiently transfected in
368 Expi293F™ cells (Thermo Fischer) using FectroPRO® DNA transfection reagent (Polyplus), according
369 to the manufacturer's instructions. The cells were incubated at 37 °C for 5 days after which the
370 cultures were centrifuged. The protein was purified from the supernatants by affinity
371 chromatography using a StrepTactin column (IBA), followed by SEC on a Superdex 200 10/300
372 column (Cytiva) equilibrated in 10 mM Tris-HCl, 100 mM NaCl, pH 8.0. The peak corresponding to
373 the monomeric protein was concentrated and stored at -80 °C until used.

374 The wild-type gorilla GII FV ectodomain was cloned into the pT350 vector and used as a
375 template for generating the heparan-sulphate binding mutants by site-directed mutagenesis.
376 *Drosophila* S2 cells were stably transfected with all the vectors, as previously mentioned. The
377 ectodomains expression followed the same steps reported for the RBD production and after 6 days
378 they were purified from the cell supernatants by affinity chromatography using a StrepTactin
379 column (IBA) and SEC on a Superose 6 10/300 column (Cytiva) in 10mM Tris-HCl, 100 mM NaCl, pH
380 8.0. The fractions within the peak corresponding to the trimeric ectodomain were concentrated in
381 VivaSpin concentrators and stored at -80 °C until used.

Crystallization

Crystallization trials were performed in 200 nanoliter sitting drops formed by mixing equal volumes of the protein and reservoir solution in the format of 96 Greiner plates, using a Mosquito robot, and monitored by a Rock-Imager at the Core Facility for Protein Crystallization at Institut Pasteur in Paris, France [45]. The native RBD^D crystal used for data collection was grown in 0.1M Tris pH 8.5, 3.5M sodium formate (NaCOOH). For the derivative data, the RBD^D crystal, grown in 0.1M Tris pH 8.5, 3.25M sodium formate, was soaked overnight in the same crystallization solution supplemented with 0.5M sodium iodide and directly frozen using the mother liquor containing 33% ethylene glycol as cryo-buffer. The RBD^G crystals were obtained from a solution containing 0.2M ammonium tartarate ($(NH_4)_2\,C_4H_4O_6$) and 20% w/v PEG 3350.

X-ray diffraction data collection and SFV RBD structure determination

The native, the derivative (iodine-soaked) and the 'glycosylated' data were all collected at 100K on the Proxima-1 [46] or Proxima-2A beamlines at the SOLEIL synchrotron source (Saint Aubin, France), using the Pilatus Eiger X 16M or Eiger X 9M detectors (Dectris), respectively.

We obtained trigonal crystals, space group 322_1 for the RBD^D (2.6 Å), $P3_121$ (later found to be $P3_221$) for the derivative RBD^D (3.2 Å), and hexagonal crystals for the RBD^G protein (2.8 Å, space group $P6_1$). Diffraction data were processed using XDS [47] and scaled and merged with AIMLESS [48]. The high-resolution cut-off was based on the statistical indicator CC1/2 [49]. Several applications from the CCP4 suite were used throughout processing [50]. The statistics are given in Table S1.

To solve the structure of the RBD^D, the AutoSol pipeline from the Phenix suite[51,52] was employed, using the anomalous data set, searching for 20 iodine sites and specifying two NCS copies in the asymmetric unit (ASU). AutoSol reliably determined the substructure, composed of 20 iodine sites. The refined anomalous phases were internally used to phase the entire protein with the aid of density modification. The result of the process was a structure with a low R-factor; moreover, the density modified map showed a good contrast between the protein and the solvent and helical features clearly discernible. The initial assignment of the space group of the anomalous data was tentative, as the screw axis that is present in the cell allows for two alternatives ($P3_121$ or $P3_221$). The enantiomorph ambiguity was resolved after density modification with the anomalous phases

and model building by looking at the map and its quality. AutoSol unambiguously selected the correct space group, which is $P3_221$. The structure was further improved in Buccaneer [53] in 'experimental phases' mode, using the density modified map from AutoSol and the refined substructure from AutoSol. Finally, the BUCCANEER model was refined against the native data at 2.6 Å by iterative rounds of phenix.refine [51], BUSTER [54,55] and Coot [56], which was used throughout all model building and refinement to inspect and manually correct the model.

To solve the structure of the RBD^G, the RBD^D was used as a search-model in Molecular Replacement in Phaser [57] from the Phenix suite. In this case, the ASU was found to contain two molecules, which were again refined using a combination of BUSTER and phenix.refine.

For both models, the 2|F|o-|F|c and |F|o-|F|c electron density maps were used to unambiguously identify the carbohydrate moieties and built them. For both models, the final stereochemistry was assessed by MolProbity (http://molprobity.biochem.duke.edu/)[58].

The final maps showed clear, interpretable electron density, except for a region comprising residues 419-427 precluding building on these 9 amino acids and indicating inherent flexibility of the region. The atomic models were refined to R_{work}/R_{free} of 0.21/0.25 and 0.19/0.23, for the RBD^D and RBD^G crystals, respectively.

Cells, viral sequences and production of foamy virus viral vectors

Baby Hamster Kidney (BHK)-21 cells (ATCC-CLL-10) were cultured in DMEM-glutamax-5% fetal bovine serum (FBS) (PAA Laboratories). HT1080 cells (ECACC 85111505) were cultured in EMEM-10% FBS supplemented with 1x L-glutamine and 1x non-essential amino acids (NEAA). Human embryonic kidney 293T cells (CRL-3216) were cultured in DMEM-glutamax-10% FBS.

Foamy virus isolates were named according to the revised taxonomy [20]and short names were used for gorilla and chimpanzee strains [25]. The four-component FVV system (plasmids pcoPG, pcoPP, pcoPE, pcu2MD9-BGAL (a transfer plasmid encoding for β-galactosidase)) and the gorilla Env construct containing sequences from the zoonotic and GI-D468 (JQ867465) and GII-K74 (JQ867464) *env* genes (EnvGI-SUGII) have been described [25,59], also in Dynesen *et al.* (accompanying manuscript). Briefly, the genotype II Env construct we used (EnvGI-SUGII[25]) is comprised of the SU is from the GII-BAK74 genotype, and the LP and TM from the GI strain BAD468, the latter two being very conserved between GI and GII.

Mutations in the RBD predicted heparan sulfate binding site (K342A/R343A and R356A/R369A) were introduced to this gorilla Env plasmid containing full-length GII SU. FVVs were produced by co-transfection of four plasmids (gag:env:pol:transgene (β-galactosidase) ratio of 8:2:3:32). Three μg total DNA and eight μl polyethyleneimine (JetPEI, #101-10N, Polyplus, Ozyme) were added to 0.5 x 10^6 HEK 293T cells seeded in 6-well plates. Supernatants were collected 48 hours post transfection, clarified at 1,500 x g for 10 min, and stored as single-use aliquots at -80°C. Vector infectivity was determined by transducing BHK-21 cells with serial five-fold dilutions of vectors and detecting β-galactosidase expression after 72 hours of culture at 37°C. Plates were fixed with 0.5% glutaraldehyde in PBS for 10 min at room temperature (RT), washed with PBS and stained with 150 μl X-gal solution containing 2 mM $MgCl_2$, 10 mM potassium ferricyanide, 10 mM potassium ferrocyanide and 0.8 mg/ml 5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside in PBS for 3 hours at 37°C. Plates were counted on a S6 Ultimate Image UV analyzer (CTL Europe, Bonn, Germany). One blue cell was defined as one infectious unit. Cell transduction by FVV is a surrogate for viral infectivity and FVV titers were expressed as infectious units/ml.

The yield of FVV particles was estimated by the quantification of particle-associated transgene RNA. FVVs RNAs were extracted from raw cell supernatants with QIAamp Viral RNA Extraction Kit (Qiagen). RNAs were treated with DNA free kit (Life Technologies), retro-transcribed with Maxima H Minus Reverse Transcriptase (Thermo Fischer Scientific) using random primers (Thermo Fischer Scientific), according to manufacturer's instructions. qPCR was performed on cDNA using BGAL primers (BGAL_F 5' AAACTCGCAAGCCGACTGAT 3' and BGAL_R 5' ATATCGCGGCTCAGTTCGAG 3') with a 10-min-long denaturation step at 95°C and 40 amplification cycles (15s at 95°C, 20s at 60°C and 30s at 72°C) carried out with an Eppendorf realplex2 Mastercycler (Eppendorf). 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/ml, considering that each particle carries 2 copies of the transgene.

Prediction of RBD heparan-binding site and mutant design

The server ClusPro (https://cluspro.org/login.php) was used for identifying a potential heparin-binding site [30,60-62]. The server generated 13 models of a fully sulfated tetra-saccharide heparin fragment docked to the FV RBD and a list of atom-atom contacts between the heparin chain and the protein residues that was used to generate the plots on Figure 5B.

Env interactions with heparan-sulfate assayed by binding to heparan-sulfate Sepharose

100 µg of recombinant FV RBDs (wild-type, R356A/R369A, K342A/R343A) were injected at 1 ml/min onto a Heparin-Sepharose column (Cytiva) previously equilibrated with running buffer (10 mM Tris-HCl, 100 mM NaCl, pH 8.0). After washing, a linear gradient (50% in 30 minutes) of elution buffer (10 mM Tris-HCl, 2M NaCl, pH 8.0) was applied.

Env interactions with heparan-sulfate on cells (in vitro): Env protein binding assay

HT1080 and BHK-21 adherent cells were detached with Trypsin-EDTA and 5 x 10^5 cells were used per condition. Cell washing and staining steps were performed in PBS-0.1% BSA at 4°C. SFV Env ectodomains were added to the cell pellet for 1 hour. Cells were washed twice, incubated with anti-StrepMAB-Classic-HRP antibody that recognizes the strep tag at the C-terminus of the SFV Env ectodomain (7.5 µg/ml, IBA Lifesciences #2-1509-001) for 1 hour, washed twice and incubated with the secondary antibody coupled to fluorophore AF488, anti-HRP-AF488 (0.75 µg/ml, Jackson ImmunoResearch, #123-545-021) for 30 min. Cells were washed and fixed in PBS-2% PFA at RT for 10 min and kept at 4°C until acquisition. A minimum of 25,000 cells were acquired on a CytoFLEX cytometer (Beckman Coulter). Data were analyzed using Kaluza software (Beckman Coulter). Viable single cells were selected by the sequential application of gates on FSC-A/SSC-A and SSC-A/SSC-H dot-plots (Fig. S12A). Cells labelled with the two secondary antibodies only were used as a reference. SFV Env binding was expressed as the ratio of mean fluorescence intensity (MFI) from the cells that were incubated with the recombinant ectodomains vs. untreated cells (Fig. S12B).

Heparan sulfate removal and detection

Cells were treated with Trypsin-EDTA and 5 x 10^5 cells were labelled per condition. Cells were washed once with PBS-0.1% BSA prior to incubation with 0.1 mIU/ml heparinase III from *Flavobacterium heparinum* (Sigma-Aldrich, #H8891) in 20 mM Tris-HCl, 0.1 mg/mL BSA and 4 mM $CaCl_2$, pH 7.45 for 15 min. at 37°C. Heparan sulfate was detected by staining with F58-10E4 antibody (5 µg/ml, AmsBio, UK #370255-S) and anti-mouse IgM-AF488 antibodies (2 µg/ml, Invitrogen #A-21042). The neoantigen generated by HS removal (ΔHS) was detected with the F69-3G10 antibody (10 µg/ml, AmsBio #370260-S) and anti-mIgG-AF647 antibodies (4 µg/ml, Invitrogen #A-31571). Cell staining and washing were performed in PBS-0.1% BSA at 4°C. Incubation times were 60 and 30 min

for primary and secondary antibodies, respectively. Cytometer acquisition, and data analysis were performed as described for Env binding (Fig. S12). Cells labelled with secondary antibodies only were used as a reference. Levels of HS and ΔHS staining were expressed as the ratio of MFI from labelled to unlabeled cells (Fig. S12C).

FVVs binding assay

HT1080 cells were incubated with FVV particles (1, 10 and 100 particles/cell) on ice for 1h. Cells were washed 3 times with PBS to eliminate unbound FVVs and RNAs were extracted using RNeasy plus mini Kit (Qiagen) according to manufacturer's protocol. RT was performed as described for FVVs RNA quantification. Bound FVV 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 $2^{-\Delta\Delta Ct}$.

Statistics

The infectious titers, particle concentration, percentages of infectious particles and quantity of bound FVVs carrying wildtype and mutant SU were compared using the paired t-test.

Data availability

The data related to the X-ray structures determined for the SFV GII RBD^{D} and RBD^{G} have been deposited to the RCSB protein databank under PDB accession codes 8AEZ and 8AIZ, respectively.

References

1. Aiewsakun, P. & Katzourakis, A. Marine origin of retroviruses in the early Palaeozoic Era. *Nat Commun* **8**, 13954 (2017).
2. Rethwilm, A. & Bodem, J. Evolution of foamy viruses: the most ancient of all retroviruses. *Viruses* **5**, 2349-74 (2013).
3. Pinto-Santini, D.M., Stenbak, C.R. & Linial, M.L. Foamy virus zoonotic infections. *Retrovirology* **14**, 55 (2017).
4. Buseyne, F. et al. Clinical signs and blood test results among humans infected with zoonotic simian foamy virus: a case-control study. *Journal of Infectious Diseases* **218**, 144-151 (2018).
5. Ledesma-Feliciano, C. et al. Feline Foamy Virus Infection: Characterization of Experimental Infection and Prevalence of Natural Infection in Domestic Cats with and without Chronic Kidney Disease. *Viruses* **11**(2019).
6. Meiering, C.D. & Linial, M.L. Historical perspective of foamy virus epidemiology and infection. *Clinical Microbiology Reviews* **14**, 165-+ (2001).
7. Rajawat, Y.S., Humbert, O. & Kiem, H.-P. In-Vivo Gene Therapy with Foamy Virus Vectors. *Viruses* **11**, 1091 (2019).
8. White, J.M. & Whittaker, G.R. Fusion of Enveloped Viruses in Endosomes. *Traffic* **17**, 593-614 (2016).
9. Harrison, S.C. Viral membrane fusion. *Virology* **479-480**, 498-507 (2015).
10. Picard-Maureau, M., Jarmy, G., Berg, A., Rethwilm, A. & Lindemann, D. Foamy virus envelope glycoprotein-mediated entry involves a pH-dependent fusion process. *J Virol* **77**, 4722-30 (2003).
11. Dupont, A. et al. Identification of an Intermediate Step in Foamy Virus Fusion. *Viruses* **12**(2020).
12. Rey, F.A. & Lok, S.M. Common Features of Enveloped Viruses and Implications for Immunogen Design for Next-Generation Vaccines. *Cell* **172**, 1319-1334 (2018).
13. Effantin, G. et al. Cryo-electron microscopy structure of the native prototype foamy virus glycoprotein and virus architecture. *PLoS Pathog* **12**, e1005721 (2016).
14. Wilk, T. et al. The intact retroviral Env glycoprotein of human foamy virus is a trimer. *J Virol* **74**, 2885-7 (2000).
15. Pancera, M. et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455-61 (2014).
16. Nasimuzzaman, M. & Persons, D.A. Cell Membrane-associated heparan sulfate is a receptor for prototype foamy virus in human, monkey, and rodent cells. *Mol Ther* **20**, 1158-66 (2012).
17. Plochmann, K. et al. Heparan sulfate is an attachment factor for foamy virus entry. *J Virol* **86**, 10028-35 (2012).
18. Duda, A., Luftenegger, D., Pietschmann, T. & Lindemann, D. Characterization of the prototype foamy virus envelope glycoprotein receptor-binding domain. *J Virol* **80**, 8158-67 (2006).
19. Luftenegger, D., Picard-Maureau, M., Stanke, N., Rethwilm, A. & Lindemann, D. Analysis and function of prototype foamy virus envelope N glycosylation. *J Virol* **79**, 7664-72 (2005).
20. Khan, A.S. et al. Spumaretroviruses: Updated taxonomy and nomenclature. *Virology* **516**, 158-164 (2018).
21. Holm, L. DALI and the persistence of protein shape. *Protein Sci* **29**, 128-140 (2020).
22. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
23. Aiewsakun, P. et al. Modular nature of simian foamy virus genomes and their evolutionary history. *Virus Evol* **5**, vez032 (2019).
24. Richard, L. et al. Cocirculation of Two env Molecular Variants, of Possible Recombinant Origin, in Gorilla and Chimpanzee Simian Foamy Virus Strains from Central Africa. *J Virol* **89**, 12480-91 (2015).
25. Lambert, C. et al. 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).

26. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702-10 (2004).
27. Dong, R., Pan, S., Peng, Z., Zhang, Y. & Yang, J. mTM-align: a server for fast protein structure database search and multiple protein structure alignment. *Nucleic Acids Res* **46**, W380-W386 (2018).
28. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
29. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-97 (2007).
30. Kozakov, D. et al. The ClusPro web server for protein-protein docking. *Nat Protoc* **12**, 255-278 (2017).
31. Couteaudier, M., Calzada-Fraile, D., Montange, T., Gessain , A. & Buseyne, F. Inhibitors of the interferon response increase the replication of gorilla simian foamy viruses. *Virology* **541**, 25-31 (2020).
32. Fass, D., Harrison, S.C. & Kim, P.S. Retrovirus envelope domain at 1.7 angstrom resolution. *Nat Struct Biol* **3**, 465-9 (1996).
33. Fass, D. et al. Structure of a murine leukemia virus receptor-binding glycoprotein at 2.0 angstrom resolution. *Science* **277**, 1662-6 (1997).
34. McCarthy, K.R. et al. Structure of the Receptor Binding Domain of EnvP(b)1, an Endogenous Retroviral Envelope Protein Expressed in Human Tissues. *mBio* **11**(2020).
35. Kwong, P.D. et al. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648-59 (1998).
36. Lindemann, D., Hutter, S., Wei, G. & Lochelt, M. The Unique, the Known, and the Unknown of Spumaretrovirus Assembly. *Viruses* **13**(2021).
37. Rethwilm, A. The replication strategy of foamy viruses. *Curr Top Microbiol Immunol* **277**, 1-26 (2003).
38. Battini, J.L., Heard, J.M. & Danos, O. Receptor choice determinants in the envelope glycoproteins of amphotropic, xenotropic, and polytropic murine leukemia viruses. *J Virol* **66**, 1468-75 (1992).
39. Chen, B. Molecular Mechanism of HIV-1 Entry. *Trends Microbiol* **27**, 878-891 (2019).
40. Wang, H. et al. Cryo-EM structure of a CD4-bound open HIV-1 envelope trimer reveals structural rearrangements of the gp120 V1V2 loop. *Proc Natl Acad Sci U S A* **113**, E7151-E7158 (2016).
41. Rua, R., Betsem, E., Calattini, S., Saib, A. & Gessain, A. Genetic characterization of simian foamy viruses infecting humans. *Journal of Virology* **86**, 13350-13359 (2012).
42. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-58 (2015).
43. Krey, T. et al. The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. *PLoS Pathog* **6**, e1000762 (2010).
44. Backovic, M. & Krey, T. Stable Drosophila Cell Lines: An Alternative Approach to Exogenous Protein Expression. *Methods Mol Biol* **1350**, 349-58 (2016).
45. Weber, P. et al. High-Throughput Crystallization Pipeline at the Crystallography Core Facility of the Institut Pasteur. *Molecules* **24**(2019).
46. Chavas, L.M.G. et al. PROXIMA-1 beamline for macromolecular crystallography measurements at Synchrotron SOLEIL. *J Synchrotron Radiat* **28**, 970-976 (2021).
47. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
48. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-14 (2013).
49. Karplus, P.A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-3 (2012).
50. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-42 (2011).

51. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
52. Terwilliger, T.C. et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* **65**, 582-601 (2009).
53. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002-11 (2006).
54. Blanc, E. et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* **60**, 2210-21 (2004).
55. Bricogne, G. et al. BUSTER. 2.8.0 edn (Global Phasing Ltd., Cambridge, United Kingdom, 2009).
56. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
57. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
58. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
59. Hutter, S., Mullers, E., Stanke, N., Reh, J. & Lindemann, D. Prototype foamy virus protease activity is essential for intraparticle reverse transcription initiation but not absolutely required for uncoating upon host cell entry. *Journal of Virology* **87**, 3163-3176 (2013).
60. Mottarella, S.E. et al. Docking server for the identification of heparin binding sites on proteins. *J Chem Inf Model* **54**, 2068-78 (2014).
61. Vajda, S. et al. New additions to the ClusPro server motivated by CAPRI. *Proteins* **85**, 435-444 (2017).
62. Desta, I.T., Porter, K.A., Xia, B., Kozakov, D. & Vajda, S. Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure* **28**, 1071-1081 e3 (2020).
63. DeLano, W.L. The PyMOL Molecular Graphics System. (DeLano Scientific, San Carlos, CA, USA, 2002).
64. Kyte, J. & Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-32 (1982).
65. Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112-128 (2018).

Acknowledgements

This work was funded by the 'Agence Nationale de la Recherche' [ANR-10-LABX62-IBEID, Intra-Labex Grant, M.B] and the 'Programme de recherche transversal from Institut Pasteur' [PTR2020-353 ZOOFOAMENV, F.B.]. The funding agencies had no role in the study design, generation of results, or writing of the manuscript. We thank the staff from the Utechs Cytometry & Biomarkers and Crystallography platform at the Institut Pasteur, the synchrotron source Soleil (Saint-Aubin, France) for granting access to the facility, and to the staff of Proxima 1 and Proxima 2A beamlines for their kindness and assistance during X-ray data collection. We are grateful to Jan Hellert, Pablo Guardado-Calvo and Philippe Afonso for the discussions and advice, with special thanks to Max Baker for reading the manuscript and English language corrections.

Author contributions

Conceptualization: MB, FB

Realization of the assays: IF, LTD, YC, DB, RP, AH

Resources: MB, FB, FAR, AG

Supervision: MB, FB

Writing—original draft: MB

Writing—review & editing: IF, LTD, FB, FAR

Funding acquisition: MB, FB

Competing interests

The authors have no competing interests to declare.

Figures

Figure 1: Overview of the novel fold adopted by the SFV RBD

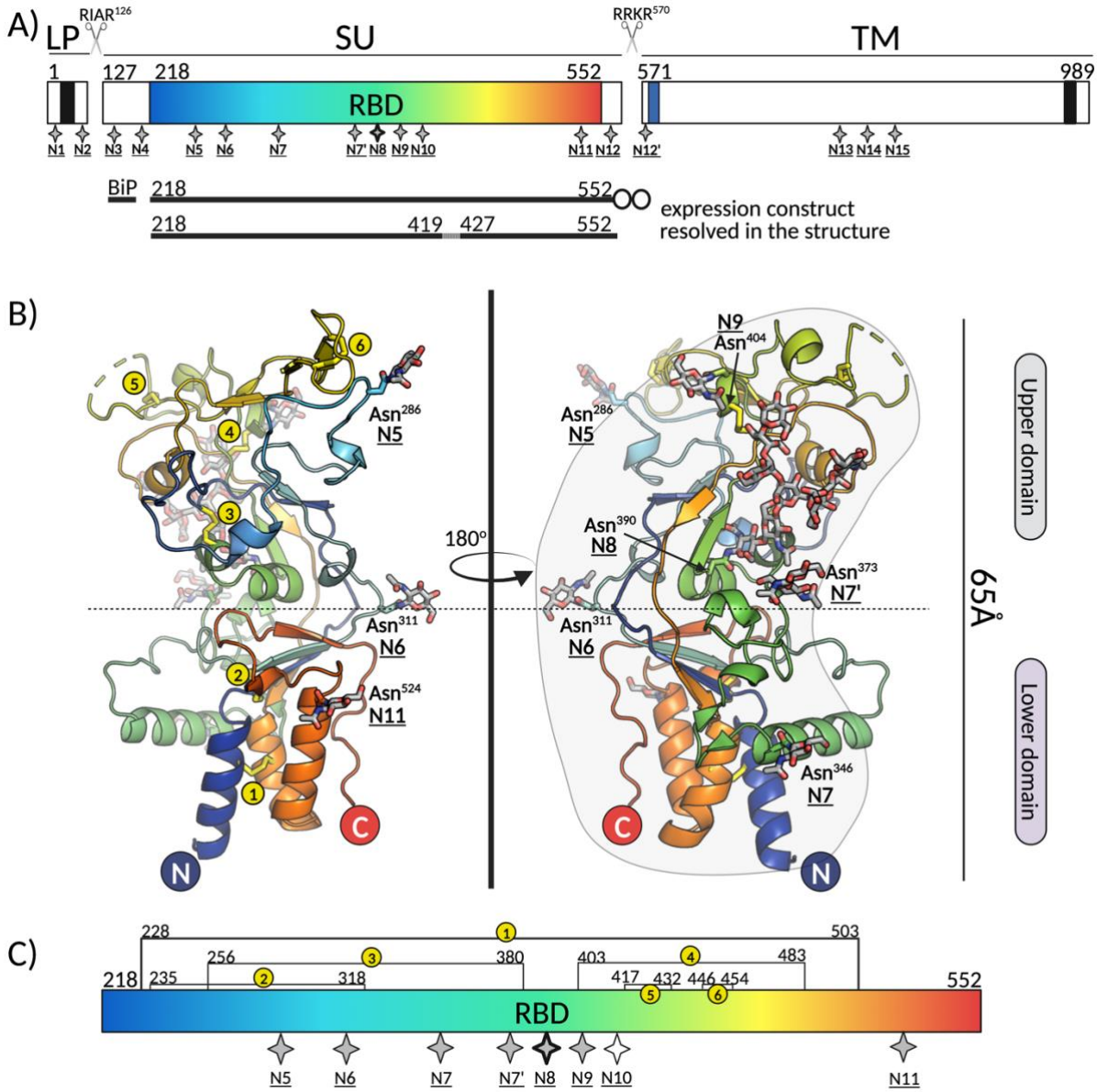


Figure 1 legend. A) Schematic representation of SFV Env protein organization indicating the three constituent chains: leader peptide (LP), surface subunit (SU), and transmembrane subunit (TM). The transmembrane domains anchoring the LP and TM in the membrane are represented as black boxes; the receptor binding domain (RBD) within SU is highlighted in blue-red spectrum; the fusion peptide at the N-terminus of the TM is shown in blue. The furin sites between the LP and SU (RIAR^{126}), and SU and TM (RRKR^{570}) are indicated with scissors symbols. The expression construct contained the exogenous BiP signal at the N-terminus, residues 218 to 552 of the SFV gorilla GII Env and a double strep tag at the C-terminus (shown as two circles). The region comprising residues 420-426 is drawn as a dashed line because it was not seen in the electron density map. The 17 putative N-glycosylation sites for gorilla SFV Env are indicated with star symbols and labels N1 to N15, following the previously established nomenclature [19].

693 **B) The X-ray structure of the RBD^{D}** is shown in ribbon model colored from N- to C-terminus in blue
694 to red spectrum, respectively. The dashed line indicates the separation between the upper and
695 lower subdomains. The N-glycosylation sites are indicated with N, and the sugars as well as the
696 asparagine side chains carrying them are displayed as sticks.

697 **C) The schematic representation of the RBD**. The potential N-glycosylation sites are indicated with
698 star symbols. The sites with sugars built in the RBD^{D} structure are shown as grey stars, and the site
699 N10 (N411) that showed no electron density for the carbohydrate in neither RBD^{D} nor RBD^{G} is
700 labeled with an empty star symbol. The site N8 (N390) that contains the long, partially buried sugar
701 is highlighted with a thicker border. The locations of six disulfides are indicated with yellow circles
702 and numbers 1-6. The figure was created with Pymol [63] and BioRender.com.

703

Figure 2: SFV RBD secondary structure topology diagram

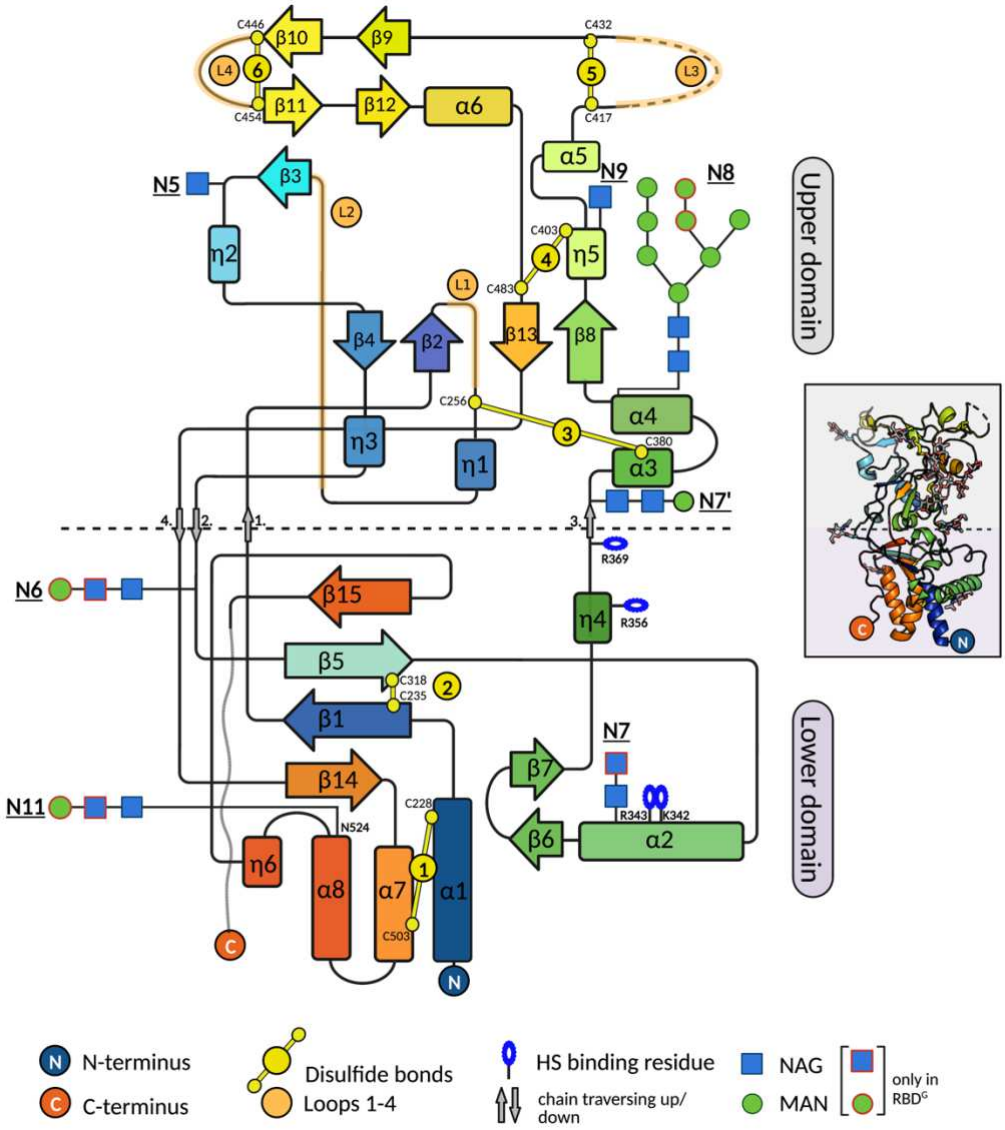


Figure 2 legend: The horizontal dashed line designates the boundary between the lower and upper subdomains. The NAG and MAN units built only into the RBD^{G} (and not RBD^{D}) structure are indicated with red frames. The figure was created with BioRender.com.

Figure 3: The oligosaccharide linked to N390 plays a structural role in the RBD

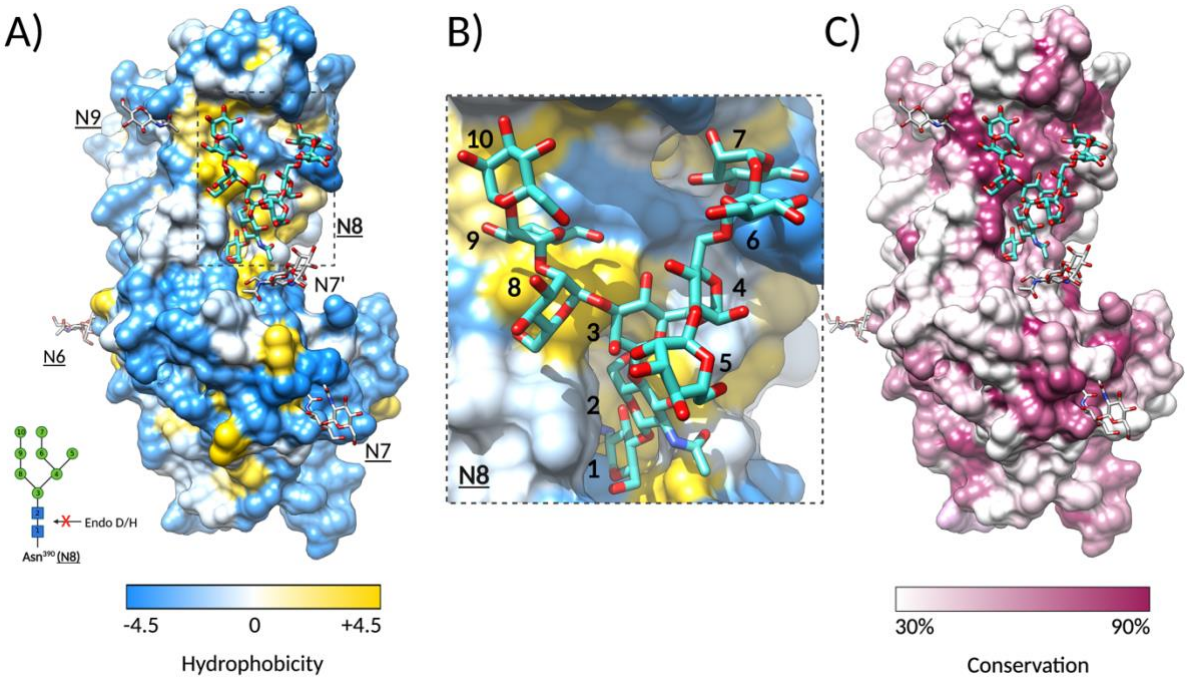


Figure 3 legend: A) Molecular surface representation of the SFV RBD^{G} colored by residue hydrophobicity. Hydrophobicity for each residue was calculated according to the Kyte and Doolittle scale [64] in Chimera [28], with the gradient color key indicating the lowest hydrophobicity in blue, to the highest hydrophobicity in yellow. The sugars at sites N6, N7, N7', and N9 are displayed as white sticks, and the sugar attached to N8 as cyan sticks. The inlet shows the bond cleaved by glycosydases Endo D/H, which is protected in N8.

B) The N8 sugar attached to N390 covers a hydrophobic region. Magnified region within the dashed-line rectangle in panel A) Is shown. The NAG and MAN residues are labeled with numbers that correspond to the N-oligosaccharide drawn in the inlet of panel A).

C) The hydrophobic patch covered by N8 is well conserved. The SFV RBD surface is rendered by residue conservation in Chimera [28], according to the % of the identical residues in the 11 FV Env sequences (alignment shown in Fig. S5). Residues conserved in less than 30% and more than 90% of sequences are colored in white and purple, respectively, and residues in between with a white-purple gradient, as indicated on the color key below the surface representation.

Figure 4: The RBDs form a trimeric assembly at the apex of the full-length Env

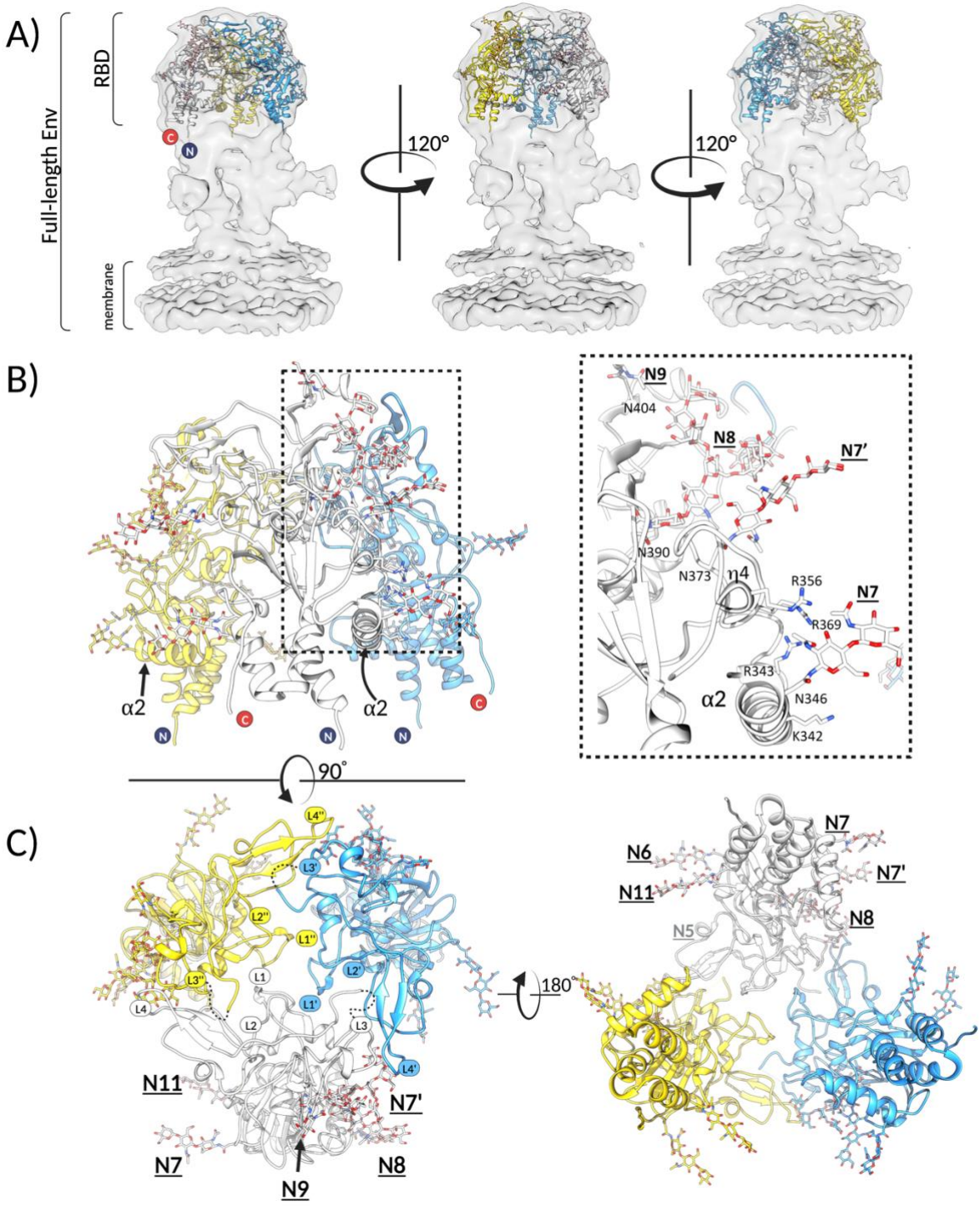


Figure 4 legend: A) Three SFV RBD protomers were fitted in the 9Å cryo-EM map (EMBD: 4013) obtained by cryo-EM 3D reconstruction of the full-length PFV Env expressed on viral vector particles [13]. The map is shown in light grey surface, and RBDs in cartoon mode, with each protomer colored differently (yellow, white, light blue).

B) The three RBDs fitted as explained in panel A) are shown to illustrate that the α2 and η4 helices, which carry the HS binding residues (K342, R343, R359, R369), as well as the N-linked glycosylations

738 (N6, N7, N7', N8, N9 and N11) point outward and are solvent accessible. The boxed region on the
739 left panel is magnified and displayed on the right panel (only one protomer, colored in white, is
740 represented for clarity purposes).

741 **C)** The views at the trimeric RBD arrangement from the top i.e. looking at the membrane (left) and
742 bottom i.e. looking from the membrane (right) are shown. The RBDs form interprotomer contacts
743 via the L1-L4 in the upper domain. The loops belonging to each protomer are designated as L, L' and
744 L''. Images in all three panels were generated in Chimera [28].

745

Figure 5: Prediction of HS binding residues and design of the variants impaired in binding

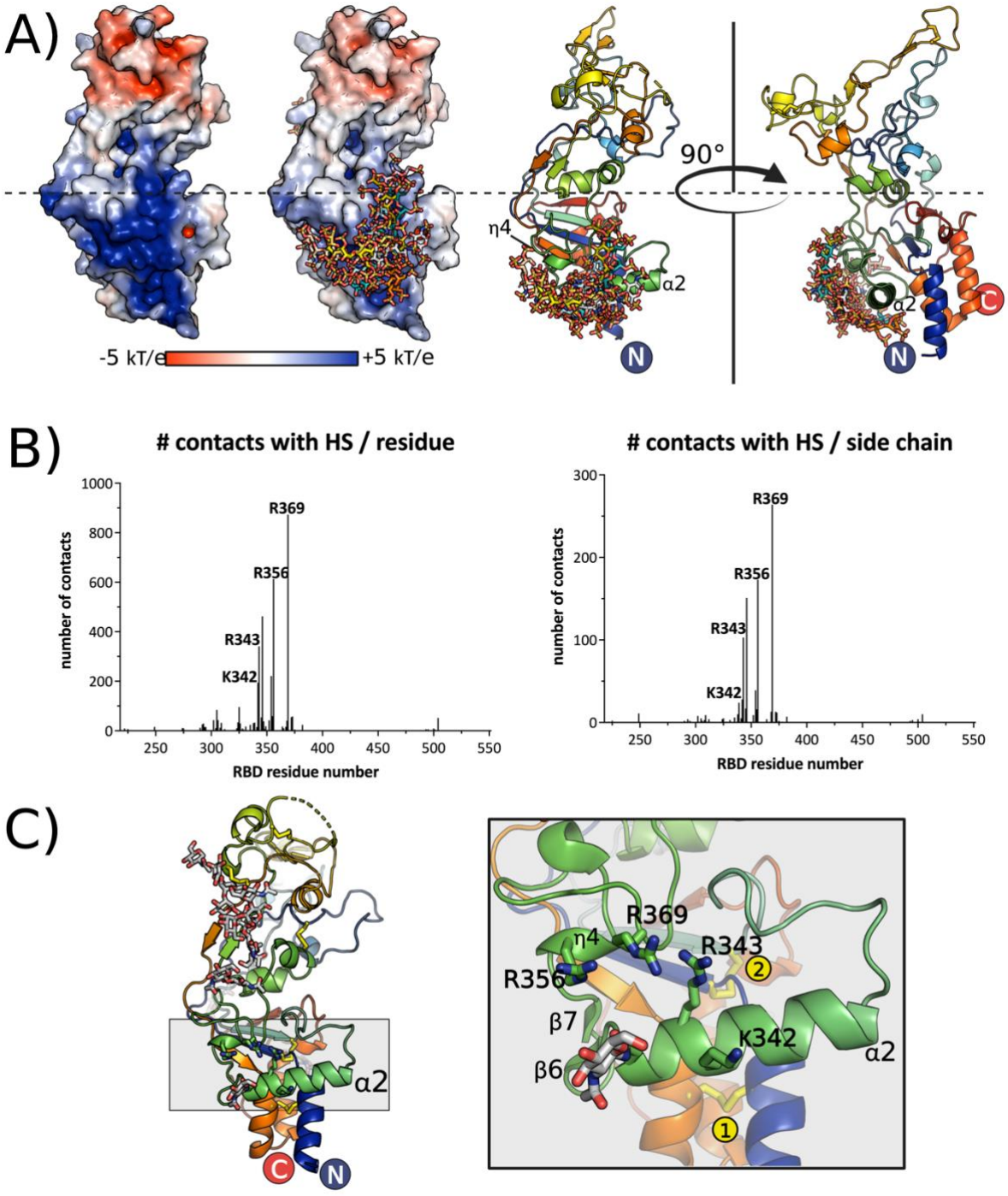


Figure 5 legend: A) Electrostatic potential distribution was calculated using Adaptive Poisson-Boltzmann Solver [65] module in Pymol [63] and plotted on the solvent accessible surface of the RBD, with red corresponding to the negative, and blue to positive potentials (2 left panels). The ensemble of HS molecules modeled by ClusPro [30] map to the lower subdomain and are displayed in sticks on the two right panels, which show the RBD in cartoon model and in two orientations to illustrate the location of predicted HS binding secondary structure elements.

B) Predicted number of contacts per residue and per side chain atoms calculated by ClusPro and plotted for each RBD residue, revealing the most likely candidates to be engaged in HS binding.

C) Structure of RBD is shown in cartoon, with the region containing α2 and η4 helices highlighted in grey. Magnification of the grey boxed region is shown on the right panel, with the relevant secondary structure elements and predicted HS binding residues shown in sticks. Two disulfide bonds are indicated with yellow circles. The figure was created with Pymol [63] and BioRender.com.

Figure 6: The SFV RBD residues K342, R343, R356 and R369 mediate Env binding to HS

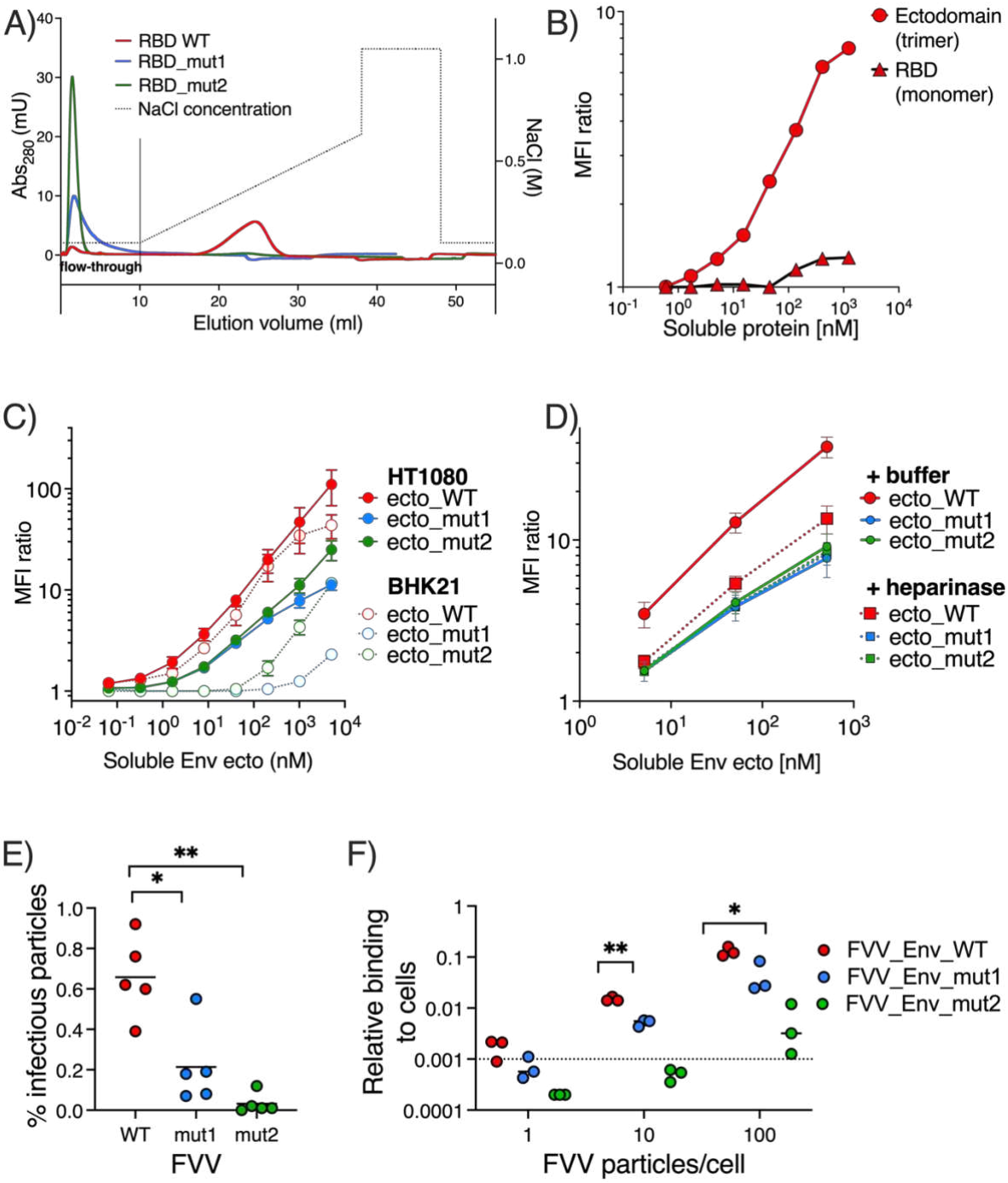


Figure 6 legend. A) Heparin-Sepharose chromatogram of the recombinant SFV RBD, WT (red) and variants with mutations in HS binding residues, mut1 (K342A/R343A) in blue, and mut2 (R356A/R369A) in green. Dotted line indicates salt concentration, plotted on the right y-axis.

B) Binding of recombinant WT RBD and ectodomains to HT1080 cells. To be comparable with the RBD, the concentration for the ectodomain is calculated and plotted for the monomeric protein.

C) Binding of WT, mut1 (K342A/R343A), and mut2 (R356A/R369A) recombinant Env ectodomains to HT1080 cells and BHK-21 cells. SFV Env binding level was expressed as the ratio of MFI from protein

771 treated to untreated cells (Fig. S12B). Mean and SD from two independent experiments are shown.
772 Cell HS expression levels were monitored on the day of each of the 2 experiments (the Env binding
773 levels were 85.4 and 61.6 for HT1080 cells, 8.40 and 2.68 for BHK-21 cells (Fig. S12C)).

774 **D)** Binding of WT, mut1 (K342A/R343A), and mut2 (R356A/R369A) recombinant Env ectodomain
775 variants to HT1080 cells treated with heparinase or buffer was quantified at increasing ectodomain
776 concentrations. Ectodomain binding level was expressed as the ratio of MFI from protein treated to
777 untreated cells (Fig. S12D). Mean and SD from two independent experiments are shown.

778 **E)** The percentage of infectious FVV particles carrying WT, mut1 or mut2 Env was calculated as the
779 ratio between the number of infectious particles (determined by titration of susceptible cells) (Fig.
780 S13A) and the amount of vector particles obtained by RT-qPCR (Fig. S13B). The different FVVs were
781 compared using the paired t test, * $p<0.05$, ** $p < 0.01$.

782 **F)** Binding of FVVs carrying the WT, mut1 or mut2 Env to HT1080 cells. The FVVs were incubated
783 with HT1080 cells on ice for 1 hour at different particle/cell ratios, before washing and quantification
784 of *the remaining* vector particles by RT-PCR. Lines represent mean values from tested FVV batches.
785 The dotted line represents the quantification threshold. The different FVVs were compared using
786 the paired t test, * $p<0.05$, ** $p < 0$.

787

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FernandezFoamyvirusRBDSI.pdf
- D1292124233valreportfullP1.pdf
- D1292124316valreportfullP1.pdf
- fernandeznrreportingsummary.pdf