



**HAL**  
open science

## **Species-Specific Functional Regions of the Green Alga Gamete Fusion Protein HAP2 Revealed by Structural Studies**

Eduard Baquero, Juliette Fedry, Pierre Legrand, Thomas Krey, Felix A. Rey

### ► **To cite this version:**

Eduard Baquero, Juliette Fedry, Pierre Legrand, Thomas Krey, Felix A. Rey. Species-Specific Functional Regions of the Green Alga Gamete Fusion Protein HAP2 Revealed by Structural Studies. *Structure*, 2019, 27 (1), pp.113 - 124.e4. <10.1016/j.str.2018.09.014>. <hal-03486165>

**HAL Id: hal-03486165**

**<https://hal.science/hal-03486165v1>**

Submitted on 20 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

## Species-specific functional regions of the green alga gamete fusion protein HAP2 revealed by structural studies

Eduard Baquero<sup>1,2</sup>, Juliette Fedry<sup>1,2,#</sup>, Pierre Legrand<sup>3</sup>, Thomas Krey<sup>1,2,#</sup> and Felix A. Rey<sup>1,2,¶</sup>

<sup>1</sup>Institut Pasteur, Unité de Virologie Structurale, 25-28 Rue du Docteur Roux, 75724 Paris Cedex 15, France

<sup>2</sup>CNRS UMR 3569, 25-28 Rue du Docteur Roux, 75724 Paris Cedex 15, France

<sup>3</sup>Synchrotron SOLEIL, L'Orme des Merisiers, 91192 Gif-sur-Yvette, France

<sup>¶</sup>Lead contact. Correspondence: Felix.Rey@pasteur.fr

<sup>#</sup>Present address: Bijvoet Center for Biomolecular Research, Utrecht University, 3584 CH Utrecht, the Netherlands (JF); Institute of Virology, Hannover Medical School and German Center for Infection Research, 30625 Hannover Germany (TK)

### Summary

The cellular fusion protein HAP2, which is structurally homologous to viral class II fusion proteins, drives gamete fusion across several eukaryotic kingdoms. Gamete fusion is a highly controlled process in eukaryotes, and is allowed only between same species gametes. In spite of a conserved architecture, HAP2 displays several species-specific functional regions that were not resolved in the available X-ray structure of the green alga *Chlamydomonas reinhardtii* HAP2 ectodomain. Here we present an X-ray structure resolving these regions, showing a target-membrane interaction surface made by three amphipathic helices in a horseshoe-shaped arrangement. HAP2 from green algae also features additional species-specific motifs inserted in regions that in viral class II proteins are critical for the fusogenic conformational change. Such insertions include a cystine-ladder-like module evocative of EGF-like motifs responsible for extracellular protein-protein interactions in animals, and a mucin-like region. These features suggest potential HAP2 interaction sites involved in gamete fusion control.

### Introduction

Cell-cell fusion is ubiquitous across eukaryotes. It is required for fertilization, where two gametes of opposite type merge together to form a zygote (Georgadaki et al., 2016). During the development of multicellular organisms, the formation of syncytial tissues also occurs via fusion of somatic cells (Zito et al., 2016). The molecular mechanism driving the merger of lipid bilayers has been studied extensively in the case of intracellular fusion events, such as the fusion of transport vesicles (Jahn and Scheller, 2006; Martens and McMahon, 2008; Sudhof and Rothman, 2009; Wickner and Schekman, 2008) or the homotypic fusion of intracellular organelles (Farmer et al., 2018; Hu and Rapoport, 2016). In contrast, relatively

little is known about extracellular fusion (Hernandez and Podbilewicz, 2017; Sampath et al., 2018). The only well-characterized extracellular fusion process is that of enveloped viruses, which fuse the viral lipid bilayer either with the plasma membrane or with an endosome from the luminal side (Harrison, 2015). HAP2/GCS1 is the first *bona fide* gamete fusion protein identified. Discovered in flowering plants (Johnson et al., 2004; Mori et al., 2006), subsequent studies in the green alga *Chlamydomonas reinhardtii* and the malaria organism *Plasmodium* showed that it was essential for bilayer merger (Liu et al., 2008). The presence of HAP2 in organisms across the main branches of eukaryotes (Cole et al., 2014; Ebchuqin et al., 2014; Kawai-Toyooka et al., 2014; Steele and Dana, 2009) suggests that it is likely an ancestral gamete fusogen already present at the origin of eukaryotic life (Wong and Johnson, 2010). HAP2 is present exclusively in male gametes in most of the organisms where it was identified - or in minus type gametes in the case of isogamous organisms such as algae. The gamete fusion process has been studied in highest detail in *C. reinhardtii*, where HAP2 was shown to localize to the “mating structure”, a projecting organelle at the plasma membrane located in between the two flagella of the *minus* gametes (Pan et al., 2003).

We recently reported the x-ray structure of the *C. reinhardtii* HAP2 ectodomain. Although this structure lacked information about important functional regions, it revealed clear homology to class II viral fusion proteins (Fedry et al., 2017), a notion supported by parallel functional studies on HAP2 from other eukaryotic phyla (Angrisano et al., 2017; Pinello et al., 2017; Valansi et al., 2017). The structural study also showed that the HAP2 ectodomain trimerizes upon insertion to membranes, adopting a hairpin arrangement typical of the post-fusion form of the viral proteins, with the transmembrane (TM) proximal region and the fusion loops together at one end of the trimer (Figure 1).

In spite of being structurally unrelated, the three characterized structural classes of viral fusion proteins, termed class I, II and III (Igonet and Rey, 2012), were shown to function via a similar mechanism (Harrison, 2008). Anchored in the viral lipid envelope by a C-terminal TM segment, they have a bulky N-terminal ectodomain folded in a metastable conformation at the viral surface. Interactions with the target membrane trigger an exergonic conformational change during which the protein exposes a polypeptide segment, termed “fusion peptide” when it is N-terminal or “fusion loop” when it is internal. This segment inserts firmly into the outer leaflet of the target cell membrane while the protein adopts a transient, elongated trimeric conformation (Figure 1A). The trimer subunits then further re-organize into a hairpin that brings both membranes into close apposition (within 1 nm), thereby overcoming the

repulsive force between the two membranes generated by the required dehydration of the outer leaflets to allow lipid contact, helping catalyze the membrane fusion reaction.

The reported HAP2 structure displayed the three characteristic  $\beta$ -sheet-rich domains of viral class II fusion proteins, organized around the central domain I. In the post-fusion form, domain I is a 10-stranded  $\beta$ -sandwich with the inner  $A_0B_0I_0H_0G_0$   $\beta$ -sheet packing against the outer  $J_0C_0D_0E_0F_0$   $\beta$ -sheet (Figure 1B). At the domain I “front” end, two of the connections between consecutive  $\beta$ -strands make very long excursions to form the elongated domain II: the  $D_0E_0$  connection spans domain II  $\beta$ -strands  $a'$  through  $e'$ , while the  $H_0I_0$  connection makes  $\beta$ -strands  $f$  through  $l$  (Figure 1B and C). The fusion loops are located in the segment connecting  $\beta$  strands  $c$  and  $d$ , at the very tip of domain II. At the opposite end of domain I - the “bottom” end - a flexible linker connects to domain III, which has an immunoglobulin superfamily fold and is in turn connected to the C-terminal viral TM anchor via a flexible segment called “stem”. Although the structure of HAP2 in the pre-fusion form is not known, structural studies on pre- and post-fusion forms of viral class II fusion proteins (Bressanelli et al., 2004; Gibbons et al., 2004; Guardado-Calvo et al., 2017; Halldorsson et al., 2016; Modis et al., 2004) have shown that the transition mostly involves a rearrangement of the three domains with respect to each other, with domain III undergoing an important translation to relocate to the sides of the trimer (diagrammed in Figure 1A). Importantly, in all cases studied, the central domain I undergoes the most significant internal rearrangement, with  $\beta$ -strand  $A_0$  switching from the inner to the outer  $\beta$ -sheet. This change is accompanied by formation of  $\beta$ -strand  $J_0$ , which in the pre-fusion form is part of the linker between domains I and III, and by a significant rearrangement of the  $A_0B_0$ ,  $C_0D_0$  and  $E_0F_0$  inter-strand connections (Figure 1). These studies suggest that these same regions are also likely to undergo structural changes during the HAP2 trimerization to adopt its post-fusion form.

Although they drive fusion in an overall similar way, the viral fusion proteins of the various classes have different ways of ensuring a stable enough insertion into the target membrane to withstand the molecular gymnastics diagrammed in Figure 1A. In the case of the influenza virus hemagglutinin (HA) (the prototype class I fusion protein), the fusion peptide displays a random coil conformation in the pre-fusion form (Wilson et al., 1981), but restructures upon exposure to a membrane into an  $\alpha$ -helical hairpin projecting nine bulky non-polar side chains (Harter et al., 1989; Lorieau et al., 2010). In contrast, for certain class II proteins such as that of the Rift Valley fever virus (RVFV, a bunyavirus), the fusion loop maintains the same conformation it has in the pre-fusion form. In this case, a pre-formed pocket accommodates the head groups of glycerophospholipids of the membrane’s outer leaflet by making multiple

hydrogen bonds (Guardado-Calvo et al., 2017). Combined with the insertion of only one or two bulky aromatic side chains of the fusion loop, these polar interactions provide stable enough anchoring to drive the fusion process.

Here, we report the X-ray structure of the *C. reinhardtii* HAP2 ectodomain derived from a different crystal form that resolves most of the regions that were missing in the previous structure. It shows that the fusion loops are folded as three amphipathic helices in a horse-shoe-shaped arrangement. Mutagenesis altering the non-polar character of the exposed residues in the fusion helices disrupted the ability of the HAP2 ectodomain to insert into liposomes. This structure also showed the presence of an unanticipated cystine ladder-like motif reminiscent of the epidermal-growth factor (EGF)-like domains inserted in loop C<sub>0</sub>D<sub>0</sub>, packing against a mucin-like region in loop E<sub>0</sub>F<sub>0</sub> at the bottom end of domain I.

## Results

**Crystallization of the intact HAP2 ectodomain.** The crystals used to determine the previously reported HAP2 X-ray structure grew only after limited proteolysis of the HAP2 ectodomain (Fedry et al., 2017). Here we screened for crystallization conditions of the unproteolysed HAP2 ectodomain in complex with HAP2-specific monoclonal antibodies (Mabs) obtained previously (Fedry et al., 2017). The complex with the single chain variable fragment (scFv) of one of these antibodies resulted in crystals in the hexagonal space group P6<sub>3</sub>22 diffracting to 2.6Å resolution (Figure S1). Structure determination by molecular replacement using the atomic model of HAP2, PDB code 5MF1, resulted in an electron density map of HAP2 displaying 1 protomer per asymmetric unit (i.e., with the trimer axis coincident with a crystallographic 3-fold axis) and 80% of the volume occupied by solvent (Table 1). Unexpectedly, this map did not reveal density for a bound scFv. It is possible that formation of the crystal lattice displaced the bound antibody, or that it is present but disordered in the large solvent volumes in between HAP2 molecules (Figure S1). In any event, the electron density allowed tracing the polypeptide chain for most of the HAP2 ectodomain, including loops that were not resolved in the available structure. In total, the new model contains 551 out of 573 residues (Figure 1B), with a break of about 15 residues in the *fg* loop in domain II – a loop projecting laterally, perpendicular to the trimer axis and roughly midway between top and bottom. Potential antibody binding to this loop could explain the absence of scFv density in the crystal, and will be the object of future experiments. Also missing was part of loop E<sub>0</sub>F<sub>0</sub> at the trimer bottom (Figure S1).

Crystallographic refinement led to a final model containing 4136 atoms in total, with free R factor of 24.3% at 2.6Å. Comparison of the new model with the previously determined structure resulted in a root mean square deviation (RMSD) of 0.528 Å for 528 aligned C $\alpha$  atoms, indicating that they are very close, with minor changes at the tip of domain II, where the proteolytic treatment to obtain the crystals may have locally affected the protein conformation.

### **Potential protein-protein interaction modules in domain I**

**Mucin-like region (MLR).** The new crystals revealed density for loops C<sub>0</sub>D<sub>0</sub> and E<sub>0</sub>F<sub>0</sub> (Figure 2). These two segments are highly variable in sequence and length in HAP2 orthologs across eukaryotes (Figures S2 and S3), and are particularly long in HAP2 from green algae. In *C. reinhardtii* HAP2, the longest is loop E<sub>0</sub>F<sub>0</sub> (residues 239-284), a 45-residue long segment rich in proline and serine/threonine residues (Figure S2A), which are typical of heavily O-glycosylated proteins such as the mucins (Shogren et al., 1989) (Figure 2A, left panel and Figure S2A). MLRs, which are known to confer unique rheological properties on proteins (Tabak, 1995), are present on viral proteins involved in attachment to cells, for instance in glycoprotein C of herpes simplex viruses (Rajcani and Vojvodova, 1998), the respiratory syncytial virus glycoprotein G (Satake et al., 1985; Wertz et al., 1985), or in the surface glycoprotein of Ebola virus (Kirchdoerfer et al., 2017). These viral proteins have in common the recognition and binding of glycosaminoglycans (GAGs) present at the surface of target cells, and the MLR was shown to modulate GAG binding (Altgarde et al., 2015). In animals, MLRs are found on many proteins anchored at the cell surface, as well in secreted proteins, and were found to modulate cell recognition and adhesion via sugar/lectin interactions (Tran and Ten Hagen, 2013). Although the new crystals did not entirely resolve the E<sub>0</sub>F<sub>0</sub> loop, they showed density for its N-terminal six amino acids (sequence PTAPSP, downstream of strand E<sub>0</sub>) as well as its C-terminal 14 residues (271-APQFLSPPAPSTRE-285) preceding  $\beta$ -strand F<sub>0</sub> (Figure 2A), accounting in total for 40% of the loop. The latter the proline rich segment packs against the bottom of domain I, and also contacts loops A<sub>0</sub>B<sub>0</sub> and C<sub>0</sub>D<sub>0</sub> of the neighboring subunit. Of note, the downstream F<sub>0</sub> and G<sub>0</sub>  $\beta$ -strands in the previously reported structure were miss-assigned and were out of register by two residues, due to poor electron density in this region and the more limited resolution of those crystals. In addition, the density that had been interpreted as a potential O-glycosylation on residue Thr577 at the end of domain III, is clear in the new structure to correspond to an N-linked glycan attached to Asn578. These errors were corrected in the present structure.

**Cystine ladder-like (CLL) motif.** The 37-residue long C<sub>0</sub>D<sub>0</sub> loop is completely resolved in the new structure, and displays a CLL motif with three parallel disulfide bonds (Figure 2A, right panel). This motif is present only in HAP2 from green algae (Figure S3), across which it is highly conserved (Figure 2B). Authentic cystine ladder motifs have been observed in  $\theta$ -defensins (Conibear et al., 2012), which are 18-residue long cyclic peptides with antifungal, antibacterial, and antiviral activities first isolated from primates (Tang et al., 1999). The CLL motif in the C<sub>0</sub>D<sub>0</sub> loop is however more reminiscent of EGF-like domains (Fig. 2D), which are extra-cellular modules identified in animals, found as single or as multiple copies connected sequentially (Bork et al., 1996). Protein-protein contacts involving EGF-like modules are associated with blood coagulation, complement activation, cell adhesion, or determination of embryonic cell fate during animal development (Bork et al., 1996; Carpenter and Cohen, 1990; Stenflo, 1991). These domains are also around 40 aa long and contain six conserved cysteine residues forming three disulfide bonds. The arrangement of these bonds is intermediate between the CLL and the “cystine knot motif” observed in many bioactive peptides (Daly and Craik, 2011). In HAP2, the CLL motif is inserted at the turn of a  $\beta$ -hairpin formed by domain I consecutive  $\beta$ -strands C<sub>0</sub> and D<sub>0</sub>, and so the N- and C-termini are next to each other (Figure 2D, left panel). In contrast, the “chain of beads” organization of sequential EGF-like domains directs their termini to opposite ends (Figure 2D). This different topology results in a change in the disulfide connectivity, with a 1-3, 2-4 and 5-6 bonding pattern in EGF-like modules and 1-2, 3-6 and 4-5 in the HAP2 CLL (Figure 2B). The HAP2 C<sub>0</sub>D<sub>0</sub> loop displays an electrostatically charged surface potential (Figure 2C) – a hallmark of EGF-like modules - although the CLL does not display the Ca<sup>2+</sup> binding motif present in a subset of the EGF-like domains (Rao et al., 1995).

### **The stem region.**

Another region that had not been resolved previously is the stem segment that connects domain III to the C-terminal TM helix. The new structure resolves the N-terminal 12 residues of the stem, showing that it is directed toward the tip of domain II of the adjacent “left” subunit in the trimer (Figure 3). Compared to the two viral class II proteins for which the stem is resolved, RVFV Gc and rubella virus (RV) E1, the path of the stem to reach the fusion loops in the HAP2 post-fusion trimer is different. In spite of a similar positioning of domain III, the stem does not turn back to interact with domain II from the same protomer (in yellow in Figure 3) as in RVFV. HAP2 is also different to RV E1, as in this case not only the stem but also domain III interacts with the alternative adjacent subunit (to the “right” in the trimer,

Figure 3, right panel). The HAP2 stem fills a surface groove in the adjacent protomer, formed between the *bc* loop and the *ae fg*  $\beta$ -sheet of domain II. The polar contacts include a short segment of antiparallel main-chain  $\beta$ -interactions (strand *m*, residues 587-589, Figures 1B and 1C) with the *bc* loop, inducing formation of strand *s* (residues 149-151) (Figure 3, inset). The construct used for crystallization ended at residue G592, followed by the affinity tag. The last residue ordered in the structure is N591 (Figure 1C), with the stem ending roughly at the level of the *bdc*  $\beta$ -sheet in the domain II tip. In full-length *C. reinhardtii* HAP2, there are 42 aa between this point and the TM helix, including an amphipathic membrane proximal external region (MPER). Comparison with the viral proteins (Figure 3) suggests that the distance remaining to reach the fusion loops would be spanned by about 10-12 residues, until about position 604. The sequence of the remaining C-terminal part (residues 605-633, Figure S4) is compatible with the MPER present in a number of viral fusion proteins (Cai et al., 2011; Guardado-Calvo and Rey, 2017; Shogren et al., 1989). The MPER has so far eluded structural characterization for any fusion protein in the context of the post-fusion form (the MPER is not represented, for clarity, in the diagram of Fig. 1A).

### **The HAP2 membrane insertion region**

The packing between HAP2 trimers in the new crystals involves 2-fold symmetric head-to-head interactions along the 3-fold crystallographic axis (Figure S1). This “dimer of trimers” is likely to correspond to the “hexamer” observed in the size-exclusion chromatography (SEC) profile of the purified recombinant HAP2 ectodomain (Figure S1A, left panel). In this assembly, the non-polar residues at the tip of domain II are masked from solvent. This interaction between trimers results in the fusion loops being entirely resolved in the new structure. The segment between  $\beta$ -strands *c* and *d* of domain II, which is 40 residues long, projects two loops (*cd1* and *cd2* in Figure 3B) exposing non-polar side chains. In place of the single  $\alpha$ F fusion helix observed in *A. thaliana* HAP2 and the three potential short fusion loops of *T. cruzi* HAP2 (Figure 4A), in *C. reinhardtii* the HAP2 membrane insertion surface is composed of helix  $\alpha$ F1 in loop *cd1* and  $\alpha$ F2 in *cd2*, running roughly antiparallel to each other (Figure 5, left panels), and a small  $3_{10}$  helix ( $\eta$ F) also in loop *cd2* directly preceding  $\alpha$ F2. The three “fusion helices” are arranged into a horseshoe shape, exposing the non-polar residues toward the target membrane (Figures 4 and 5).  $\alpha$ F1 is 9 aa long and exposes residues W173 and F177;  $\eta$ F is only one turn and exposes F192 and W193, and  $\alpha$ F2 is 7 aa long and exposes P196, L197, L200 and I201 (Figure 4). In addition to providing a firm hold of the protein onto the target membrane, the substantial membrane insertion surface is likely to generate a

significant stress upon insertion into the outer leaflet of the opposing gamete membrane. The resulting lipid bilayer destabilization can potentially also play a role in facilitating the downstream membrane fusion process.

The HAP2 *cd* segment is long in comparison with the *cd* loop of the fusion protein of the well-studied flaviviruses (10 residues) (Modis et al., 2004) or bunyaviruses like RVFV (15 residues) (Guardado-Calvo et al., 2017), but not when comparing to other class II viral fusion proteins. For instance, RV E1 has a 48 aa long *cd* inter-strand connection (DuBois et al., 2013), featuring an additional  $\beta$  strand (*c'*) that augments the *bdc*  $\beta$ -sheet at the domain II tip into *bdcc'* and two fusion loops, *cc'* and *c'd* (Figure 3B). In HAP2, *cd1* and *cd2* are separated by the arginine residue (Figure 4A) making a strictly conserved salt bridge (R185-E126 in *C. reinhardtii* HAP2) with a glutamic acid in  $\beta$ -strand *b* (Fedry et al., 2017). This salt-bridge is a distinctive HAP2 feature among all class II proteins, linking the variable membrane insertion region to the highly conserved core of the protein (Fedry et al., 2018). The R185 side chain not only makes a salt bridge but also donates multiple hydrogen bonds to backbone carbonyls of the highly conserved “*ij* loop” (Figure 4A), so named because it takes the place of the *ij*  $\beta$ -hairpin of viral class II proteins. The HAP2 *ij* loop is the central part of the pfam 10699 motif (Finn et al., 2014), a ~50 aa long stretch of conserved sequence that allowed the identification of HAP2 in numerous eukaryotic species.

Although the aa 184-190 stretch (i.e., including the conserved R185) in between *cd1* and *cd2* was ordered in the previous structure and formed an  $\alpha$ -helix termed  $\alpha_0$ , it displays an extended conformation in the new structure (Figures 1B and 4A, left panel). This segment is also extended in *Arabidopsis thaliana* HAP2 (Figure 4A, middle panel) (Fedry et al., 2018). The difference between the two *C. reinhardtii* HAP2 structures in this stretch could have been induced by the *in situ* proteolytic treatment used to grow the initial crystals.  $\alpha_0$  helix folding/unfolding could also be a mechanism required to fully project the *cd* loop toward the target membrane during the fusogenic conformational change of HAP2. Indeed, the crystal structure of domain II of *T. cruzi* HAP2 does display an  $\alpha_0$  helix in the same location and orientation (Figure 4A, right panel). In contrast to the fusion helices,  $\alpha_0$  runs perpendicular to the target membrane. The only non-polar bulky residues present at the membrane insertion region are in the second turn of the *T. cruzi* HAP2  $\alpha_0$  helix, and are not fully exposed (Figure 4A, third panel), suggesting that this segment may have adopted an alternative conformation in the aqueous environment used for crystallization (Fedry et al., 2018).

**The non-polar residues of the fusion helices are required for liposome binding.**

Our previous results demonstrated that gamete fusion *in vivo* correlates with the ability of HAP2 to insert into liposomes *in vitro* (Fedry et al., 2018; Fedry et al., 2017). To confirm that the exposed residues of the fusion helices are indeed important for membrane insertion, we examined the behavior of the HAP2 ectodomain and of selected mutants in liposome co-floitation in density gradients, as described in the STAR Methods section. All the mutants tested were similar to wild type in terms of expression yields and transit through the quality control system of the producer cells for secretion. The SEC profiles of the mutants were also similar to wild type, indicating that the introduced mutations did not induce any obvious misfolding, as expected for exposed residues such as those being tested.

As P196 induces an important kink causing the shift from  $\eta$ F to  $\alpha$ F2 in loop *cd2*, we tested the effect of the P196A mutation and found that there were no statistically significant differences with wild type (Figure 4C) in liposome co-floitation. The kink induced by P196 in the conformation of the fusion helices is thus not essential for insertion into membranes. In contrast, we found a more drastic effect when we substituted the non-polar residues by histidine, a polar aa. Indeed, the double mutants W173H/F177H ( $\alpha$ F1), W192H/F193H ( $\eta$ F) and the triple mutant L197H/L200H/I201H ( $\alpha$ F2) displayed essentially no detectable liposome binding. The triple mutant L197Q/L200Q/I201Q had the same effect, suggesting that any polar side chain replacing the non-polar bulky side chains of these helices will interfere with membrane binding, as expected. This last result is also an important control, as the introduction of multiple neighboring histidines could induce local repulsion upon their protonation, although the membrane binding assays were done at pH 7.4 and the pKa of solvent-exposed histidine side chains is 6.8. Of note, in contrast to virus fusion driven by class II fusion proteins, which takes place in the acidic environment of the endosomes, *Chlamydomonas* gamete fusion occurs in the extracellular environment at neutral pH.

When the same bulky non-polar residues were replaced by alanine, which has a small non-polar side chain, the effect on membrane insertion was less striking: the difference in binding of double mutant W192A/F193A in  $\eta$ F was less significant with respect to wild type, and W173A/F177A ( $\alpha$ F1) or L197A/L200A/I201A ( $\alpha$ F2) displayed reduced binding although not as significant as the mutants with polar residues. It required a mutant with alanine substitution in both,  $\alpha$ F1 and  $\eta$ F, W176A/F177A/W192A/F193A, for a complete loss of binding comparable to the experiments involving substitution by polar residues (Figure 4C). These experiments therefore demonstrate that the bulky non-polar residues of the fusion loops indeed control the insertion of the HAP2 ectodomain into lipid bilayers. The differences observed when substituting them to alanine instead of polar residues suggest that certain

positions are key anchoring sites requiring a bulky non-polar side chain, whereas others are necessary only to maintain the non-polar character of the surface. This is evidenced in particular by the significant difference in insertion to liposomes of the mutant W173A/F177A compared to W192A/F193A (Figure 4C, bottom panels).

## Discussion

The X-ray structure of the *C. reinhardtii* HAP2 ectodomain described here has higher resolution and provides views of important functional regions that were not resolved in the previously reported structure. One crucial segment that was missing is the tip of domain II and its fusion loops, which must insert into the target membrane and are critical for the membrane fusion function of HAP2. The new structure shows that the domain II tip exposes a hydrophobic platform formed by three short helices projecting ten bulky non-polar residues that are used for membrane insertion (Figure 1A, steps ii to v). This region is involved in hydrophobic head-to-head interactions between two HAP2 trimers to form the crystals (Figure S1B), making a dimer of trimers that was also observed in solution (Figure S1A, left panel).

The arrangement of the membrane insertion surface in *C. reinhardtii* HAP2 is reminiscent of the  $\alpha$ -helical hairpin of the influenza virus HA fusion peptide, which projects nine bulky non-polar residues (Figure 5, second panel) into the target membrane (reviewed in (Apellaniz et al., 2014)). This structure forms only upon interaction of the HA fusion peptide with the outer leaflet of the membrane, whereas in HAP2 the helical arrangement appears to be pre-formed in the protein. An important difference is that in HA the two antiparallel amphipathic helices pack very tightly against each other, whereas in HAP2 they have a rather loose, horseshoe-shaped arrangement. The HAP2 fusion loops are also very different to those of the viral class II proteins that have been characterized, such as the fusion proteins of flaviviruses - i.e, dengue virus E (Klein et al., 2013; Modis et al., 2004), of Gc from bunyaviruses such as RVFV (Guardado-Calvo et al., 2017) (Figure 5, right panels), and also to the RV E1 fusion loops (Figure 3B). Molecular dynamics analyses using RVFV Gc showed that insertion of the fusion loop involves a rearrangement of the membrane such that the bulky lipid head groups move apart to avoid clashes with the main chain of the inserted loop. The presence of lipids with very small polar heads were found to be required in the membrane to fill the void created in the aliphatic moiety at the site of insertion. This effect would be more drastic in the case of *C. reinhardtii* HAP2 as the inserted motif is much larger (compare panels in Figures 4A and

5). We note that the tips of the three domains II in the HAP2 trimer leave an empty space at the center (Figure 4B), which we propose is important to accommodate the displaced bulky lipid head groups upon insertion of the fusion helices. In influenza virus HA2, the tight  $\alpha$ -helical platform of the fusion peptide is loosely tethered via a flexible linker to the central, rigid trimeric  $\alpha$ -helical coiled coil of the HA2 subunit of HA (Bullough et al., 1994; Chen et al., 1999), providing flexibility and leaving space to accommodate the displaced lipid head groups. In HAP2, not only the lateral packing of the fusion helices appears to be loose (the minor effect of the P196A mutation in liposome binding shown in Figure 4C indeed suggest that this region displays conformational malleability), but their connection to the rigid *bdc*  $\beta$ -sheet of the domain II tip also displays flexibility (Figure 4A) that is potentially important for membrane insertion. The relatively flexible connection to the *bdc*  $\beta$ -sheet could combine with a potential rocking about the strictly conserved R185-E125 salt bridge (Figure 4A). The malleability of this region is further exemplified by formation or not of a vertical  $\alpha$  helix around the R185 position. We also note that sequence analysis of the HAP2 fusion loop region further suggests that the three fusion helices are not conserved in many algae (Figure S5), and some appear to have a single fusion helix, similar to HAP2 from the flowering plants (Fedry et al., 2018). Future studies using molecular dynamics simulations by taking advantage of the currently available X-ray structures of three orthologs should be useful to better understand the HAP2 membrane insertion process.

A driver for the high variability observed in certain regions of HAP2 may be the requirement of barriers that avoid cross-species fertilization. In the viral class II proteins, the fusion loops are exposed only at the time of fusion, and until this moment the protein is maintained in a metastable conformation via interaction with itself (as diagrammed in Figure 1A, panel i) or most commonly with other proteins, such that that the fusion loops are masked (Guardado-Calvo and Rey, 2017; Kielian, 2006). The high sequence variability of the HAP2 fusion loops could thus be driven by its necessary interaction with species-specific proteins that recognize and bind to this region. The controlled release of their grip on HAP2 could then be a way to trigger its fusogenic conformational change and exposure of the fusion helices (corresponding to panel ii in Figure 1A). In our *in vitro* experiments, in the absence of the putative interacting molecules normally present in the *in vivo* situation, the recombinant HAP2 ectodomain is triggered spontaneously to change conformation, with monomers trimerizing irreversibly in solution to form post-fusion trimers that interact head-to-head (Figure S1). The remaining monomers in our preparations remain competent for membrane insertion, by forming

membrane-inserted post-fusion trimers when incubated with liposomes (Fedry et al., 2018; Fedry et al., 2017). Cellular fusion events are normally controlled so that they can take place fast but only at the required time and space. This is evidenced, for instance, in neurotransmitter release, where the fusogenic SNARE proteins are controlled by an array of other synaptic proteins that allow for very fast fusion of vesicles that are pre-bound at the cytosolic side of the plasma membrane. The rate-limiting steps involved in priming the fusion machinery have already taken place when a signal provided by calcium influx activates fusion (Brunger et al., 2018). In the case of HAP2-driven gamete fusion, our results now suggest that in addition to fusion loop masking, there may be additional controls at other steps - such as trimerization, which involves rearrangements in the bottom end of domain I. Our data further suggest additional potential means of identifying interaction partners of HAP2 *in vivo*, for instance by comparative co-immunoprecipitation using HAP2 wild-type and variants mutated at key sites at the domain I bottom.

**Acknowledgements** Funding for this project was provided by the ERC Advanced grant CelCelFus (number 340371) to FAR, as well as recurrent funding from Institut Pasteur and the CNRS. JF had a PhD fellowship from Ecole Polytechnique AMX.

We thank I. Fernandez for technical help and discussion; W.J. Snell for helpful discussions and for testing the effect antibodies against HAP2 on *C. reinhardtii* gamete fusion *in vivo*; Ahmed Haouz and staff of the Pasteur crystallography facility as well as the staff of Proxima-1 and -2 at the SOLEIL synchrotron for help during crystallization screening and diffraction data collection.

**Data Deposition:** The atomic coordinates and structure factors have been deposited in the PDB under accession code 6E18

**Authors contributions:** TK made the HAP2 constructs, JF optimized expression. TK and FJ immunized mice with the recombinant HAP2 ectodomain to obtain HAP2-specific monoclonal antibodies, and determined the aa sequence from the hybridomas. JF made constructs expressing Fab and EB the constructs for scFv from two of the antibodies. EB and FAR designed the HAP2 mutants in the membrane insertion region. EB obtained crystals using the HAP2/A20 scFv complex, determined the new X-ray structure and conducted the

liposome binding experiments. PL completed the crystallographic refinement. TK, JF, EB and FAR conceived the research.

**Declaration of interests:** The authors declare no financial interest. FAR is a consultant with *Flagship Pioneering*.

## REFERENCES

- Altgarde, N., Eriksson, C., Peerboom, N., Phan-Xuan, T., Moeller, S., Schnabelrauch, M., Svedhem, S., Trybala, E., Bergstrom, T., and Bally, M. (2015). Mucin-like Region of Herpes Simplex Virus Type 1 Attachment Protein Glycoprotein C (gC) Modulates the Virus-Glycosaminoglycan Interaction. *J Biol Chem* *290*, 21473-21485.
- Angrisano, F., Sala, K.A., Da, D.F., Liu, Y., Pei, J., Grishin, N.V., Snell, W.J., and Blagborough, A.M. (2017). Targeting the Conserved Fusion Loop of HAP2 Inhibits the Transmission of *Plasmodium berghei* and *falciparum*. *Cell Rep* *21*, 2868-2878.
- Apellaniz, B., Huarte, N., Largo, E., and Nieva, J.L. (2014). The three lives of viral fusion peptides. *Chem Phys Lipids* *181*, 40-55.
- Bork, P., Downing, A.K., Kieffer, B., and Campbell, I.D. (1996). Structure and distribution of modules in extracellular proteins. *Q Rev Biophys* *29*, 119-167.
- Bressanelli, S., Stiasny, K., Allison, S.L., Stura, E.A., Duquerroy, S., Lescar, J., Heinz, F.X., and Rey, F.A. (2004). Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *The EMBO journal* *23*, 728-738.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, P., Roversi, P., Sharff, A., Smart, O., Vornrhein, C., *et al.* (2010). BUSTER version 2.9., Cambridge, United Kingdom, Global Phasing Ltd.
- Brunger, A.T., Choi, U.B., Lai, Y., Leitz, J., and Zhou, Q. (2018). Molecular Mechanisms of Fast Neurotransmitter Release. *Annu Rev Biophys* *47*, 469-497.
- Bullough, P.A., Hughson, F.M., Skehel, J.J., and Wiley, D.C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* *371*, 37-43.
- Cai, L., Gochin, M., and Liu, K. (2011). Biochemistry and biophysics of HIV-1 gp41 - membrane interactions and implications for HIV-1 envelope protein mediated viral-cell fusion and fusion inhibitor design. *Curr Top Med Chem* *11*, 2959-2984.
- Carpenter, G., and Cohen, S. (1990). Epidermal growth factor. *J Biol Chem* *265*, 7709-7712.
- Chao, L.H., Klein, D.E., Schmidt, A.G., Pena, J.M., and Harrison, S.C. (2014). Sequential conformational rearrangements in flavivirus membrane fusion. *Elife* *3*, e04389.
- Chen, J., Skehel, J.J., and Wiley, D.C. (1999). N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc Natl Acad Sci U S A* *96*, 8967-8972.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* *66*, 12-21.
- Cole, E.S., Cassidy-Hanley, D., Fricke Pinello, J., Zeng, H., Hsueh, M., Kolbin, D., Ozzello, C., Giddings, T., Winey, M., and Clark, T.G. (2014). Function of the Male-Gamete-Specific Fusion Protein HAP2 in a Seven-Sexed Ciliate. *Current biology : CB* *24*, 2168-2173.
- Conibear, A.C., Rosengren, K.J., Harvey, P.J., and Craik, D.J. (2012). Structural characterization of the cyclic cystine ladder motif of theta-defensins. *Biochemistry* *51*, 9718-9726.

- Daly, N.L., and Craik, D.J. (2011). Bioactive cystine knot proteins. *Curr Opin Chem Biol* 15, 362-368.
- DuBois, R.M., Vaney, M.C., Tortorici, M.A., Kurdi, R.A., Barba-Spaeth, G., Krey, T., and Rey, F.A. (2013). Functional and evolutionary insight from the crystal structure of rubella virus protein E1. *Nature* 493, 552-556.
- Ebchuqin, E., Yokota, N., Yamada, L., Yasuoka, Y., Akasaka, M., Arakawa, M., Deguchi, R., Mori, T., and Sawada, H. (2014). Evidence for participation of GCS1 in fertilization of the starlet sea anemone *Nematostella vectensis*: implication of a common mechanism of sperm-egg fusion in plants and animals. *Biochemical and biophysical research communications* 451, 522-528.
- Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501.
- Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* 69, 1204-1214.
- Farmer, T., Naslavsky, N., and Caplan, S. (2018). Tying trafficking to fusion and fission at the mighty mitochondria. *Traffic*.
- Fedry, J., Forcina, J., Legrand, P., Pehau-Arnaudet, G., Haouz, A., Johnson, M., Rey, F.A., and Krey, T. (2018). Evolutionary diversification of the HAP2 membrane insertion motifs to drive gamete fusion across eukaryotes. *PLoS Biol* 16, e2006357.
- Fedry, J., Liu, Y., Pehau-Arnaudet, G., Pei, J., Li, W., Tortorici, M.A., Traincard, F., Meola, A., Bricogne, G., Grishin, N.V., *et al.* (2017). The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein. *Cell* 168, 904-915 e910.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., *et al.* (2014). Pfam: the protein families database. *Nucleic Acids Res* 42, D222-230.
- Georgadaki, K., Khoury, N., Spandidos, D.A., and Zoumpourlis, V. (2016). The molecular basis of fertilization (Review). *Int J Mol Med* 38, 979-986.
- Gibbons, D.L., Vaney, M.C., Roussel, A., Vigouroux, A., Reilly, B., Lepault, J., Kielian, M., and Rey, F.A. (2004). Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* 427, 320-325.
- Gilmartin, A.A., Lamp, B., Rumenapf, T., Persson, M.A., Rey, F.A., and Krey, T. (2012). High-level secretion of recombinant monomeric murine and human single-chain Fv antibodies from *Drosophila* S2 cells. *Protein Eng Des Sel* 25, 59-66.
- Guardado-Calvo, P., Atkovska, K., Jeffers, S.A., Grau, N., Backovic, M., Perez-Vargas, J., de Boer, S.M., Tortorici, M.A., Pehau-Arnaudet, G., Lepault, J., *et al.* (2017). A glycerophospholipid-specific pocket in the RVFV class II fusion protein drives target membrane insertion. *Science* 358, 663-667.
- Guardado-Calvo, P., and Rey, F.A. (2017). The Envelope Proteins of the Bunyavirales. *Adv Virus Res* 98, 83-118.
- Halldorsson, S., Behrens, A.J., Harlos, K., Huiskonen, J.T., Elliott, R.M., Crispin, M., Brennan, B., and Bowden, T.A. (2016). Structure of a phleboviral envelope glycoprotein

reveals a consolidated model of membrane fusion. *Proc Natl Acad Sci U S A* *113*, 7154-7159.

Harrison, S.C. (2008). Viral membrane fusion. *Nat Struct Mol Biol* *15*, 690-698.

Harrison, S.C. (2015). Viral membrane fusion. *Virology* *479-480*, 498-507.

Harter, C., James, P., Bachi, T., Semenza, G., and Brunner, J. (1989). Hydrophobic binding of the ectodomain of influenza hemagglutinin to membranes occurs through the "fusion peptide". *J Biol Chem* *264*, 6459-6464.

Hernandez, J.M., and Podbilewicz, B. (2017). The hallmarks of cell-cell fusion. *Development* *144*, 4481-4495.

Hu, J., and Rapoport, T.A. (2016). Fusion of the endoplasmic reticulum by membrane-bound GTPases. *Semin Cell Dev Biol* *60*, 105-111.

Igonet, S., and Rey, F.A. (2012). SnapShot: Viral and eukaryotic protein fusogens. *Cell* *151*, 1634-1634 e1631.

Jahn, R., and Scheller, R.H. (2006). SNAREs--engines for membrane fusion. *Nat Rev Mol Cell Biol* *7*, 631-643.

Johnson, M.A., von Besser, K., Zhou, Q., Smith, E., Aux, G., Patton, D., Levin, J.Z., and Preuss, D. (2004). Arabidopsis hapless mutations define essential gametophytic functions. *Genetics* *168*, 971-982.

Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *J Appl Crystallogr*, 67-72.

Kawai-Toyooka, H., Mori, T., Hamaji, T., Suzuki, M., Olson, B.J., Uemura, T., Ueda, T., Nakano, A., Toyoda, A., Fujiyama, A., *et al.* (2014). Sex-specific posttranslational regulation of the gamete fusogen GCS1 in the isogamous volvocine alga *Gonium pectorale*. *Eukaryot Cell* *13*, 648-656.

Kielian, M. (2006). Class II virus membrane fusion proteins. *Virology* *344*, 38-47.

Kirchdoerfer, R.N., Wasserman, H., Amarasinghe, G.K., and Saphire, E.O. (2017). Filovirus Structural Biology: The Molecules in the Machine. *Curr Top Microbiol Immunol* *411*, 381-417.

Klein, D.E., Choi, J.L., and Harrison, S.C. (2013). Structure of a dengue virus envelope protein late-stage fusion intermediate. *J Virol* *87*, 2287-2293.

Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J., *et al.* (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes Dev* *22*, 1051-1068.

Lorieau, J.L., Louis, J.M., and Bax, A. (2010). The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface. *Proc Natl Acad Sci U S A* *107*, 11341-11346.

Martens, S., and McMahon, H.T. (2008). Mechanisms of membrane fusion: disparate players and common principles. *Nat Rev Mol Cell Biol* *9*, 543-556.

Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* *427*, 313-319.

- Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat Cell Biol* 8, 64-71.
- Pan, J., Misamore, M.J., Wang, Q., and Snell, W.J. (2003). Protein transport and signal transduction during fertilization in chlamydomonas. *Traffic* 4, 452-459.
- Pinello, J.F., Lai, A.L., Millet, J.K., Cassidy-Hanley, D., Freed, J.H., and Clark, T.G. (2017). Structure-Function Studies Link Class II Viral Fusogens with the Ancestral Gamete Fusion Protein HAP2. *Curr Biol* 27, 651-660.
- Rajcani, J., and Vojvodova, A. (1998). The role of herpes simplex virus glycoproteins in the virus replication cycle. *Acta Virol* 42, 103-118.
- Rao, Z., Handford, P., Mayhew, M., Knott, V., Brownlee, G.G., and Stuart, D. (1995). The structure of a Ca(2+)-binding epidermal growth factor-like domain: its role in protein-protein interactions. *Cell* 82, 131-141.
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42, W320-324.
- Sampath, S.C., Sampath, S.C., and Millay, D.P. (2018). Myoblast fusion confusion: the resolution begins. *Skelet Muscle* 8, 3.
- Satake, M., Coligan, J.E., Elango, N., Norrby, E., and Venkatesan, S. (1985). Respiratory syncytial virus envelope glycoprotein (G) has a novel structure. *Nucleic Acids Res* 13, 7795-7812.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9, 671-675.
- Schrodinger, LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.
- Shogren, R., Gerken, T.A., and Jentoft, N. (1989). Role of glycosylation on the conformation and chain dimensions of O-linked glycoproteins: light-scattering studies of ovine submaxillary mucin. *Biochemistry* 28, 5525-5536.
- Steele, R.E., and Dana, C.E. (2009). Evolutionary history of the HAP2/GCS1 gene and sexual reproduction in metazoans. *PLoS ONE* 4, e7680.
- Stenflo, J. (1991). Structure-function relationships of epidermal growth factor modules in vitamin K-dependent clotting factors. *Blood* 78, 1637-1651.
- Sudhof, T.C., and Rothman, J.E. (2009). Membrane fusion: grappling with SNARE and SM proteins. *Science* 323, 474-477.
- Tabak, L.A. (1995). In defense of the oral cavity: structure, biosynthesis, and function of salivary mucins. *Annu Rev Physiol* 57, 547-564.
- Tang, Y.Q., Yuan, J., Osapay, G., Osapay, K., Tran, D., Miller, C.J., Ouellette, A.J., and Selsted, M.E. (1999). A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. *Science* 286, 498-502.
- Tran, D.T., and Ten Hagen, K.G. (2013). Mucin-type O-glycosylation during development. *J Biol Chem* 288, 6921-6929.
- Valansi, C., Moi, D., Leikina, E., Matveev, E., Grana, M., Chernomordik, L.V., Romero, H., Aguilar, P.S., and Podbilewicz, B. (2017). Arabidopsis HAP2/GCS1 is a gamete fusion protein homologous to somatic and viral fusogens. *J Cell Biol* 216, 571-581.

- Wertz, G.W., Collins, P.L., Huang, Y., Gruber, C., Levine, S., and Ball, L.A. (1985). Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein. *Proc Natl Acad Sci U S A* 82, 4075-4079.
- Wickner, W., and Schekman, R. (2008). Membrane fusion. *Nat Struct Mol Biol* 15, 658-664.
- Wilson, I.A., Skehel, J.J., and Wiley, D.C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* 289, 366-373.
- Wong, J.L., and Johnson, M.A. (2010). Is HAP2-GCS1 an ancestral gamete fusogen? *Trends in cell biology* 20, 134-141.
- Zito, F., Lampiasi, N., Kireev, I., and Russo, R. (2016). United we stand: Adhesion and molecular mechanisms driving cell fusion across species. *Eur J Cell Biol* 95, 552-562.

## Figure Legends

### Figure 1. The new structure reveals functional species-specific regions of HAP2.

(A) Diagram illustrating the rearrangement of viral class II fusion proteins to catalyze membrane fusion. The three domains are colored red, yellow and blue (for domains I, II and III, respectively), the fusion loop orange, the stem magenta and the TM segment grey. Viral and cellular membranes are in light and dark grey, respectively. i) Prefusion arrangement, with the fusion loop buried at an oligomeric contact - a homo-dimer in the flaviviruses (as shown here, with the dimer partner striped for clarity, except that the flavivirus fusion protein spans the membrane twice) or a heterodimer with a companion protein (as in alphaviruses and bunyaviruses). ii) Triggering of the fusogenic conformational change. The acidic environment of an endosome leads to dimer dissociation and exposure of the fusion loops. iii) Extended intermediate bridging the two membranes. Multiple monomers at the viral surface insert into the target membrane with concomitant trimerization into an extended intermediate form, separating the two membranes at distances of about 15 nm. The curved arrows indicate the subsequent relocation of domain III to make a hairpin. iv) Hairpin formation and pulling of the two membranes against each other. Domain III relocation to the side of the trimer redirects the stem toward the fusion loops, thereby pulling the two membranes against each other (grey arrows) to a distance within 1nm, forcing outer leaflet dehydration to allow membrane contacts. Domain III relocation is accompanied by a domain I internal reorganization, indicated by a red square instead of a red circle. Several adjacent trimers are required to drive fusion (Chao et al., 2014). v) Fusion pore formation. Upon contact of the two membranes, fusion proceeds via an initial merger of the outer leaflets of the two membranes (“hemifusion”, not shown) followed by formation of a fusion pore, as illustrated here.

B) The HAP2 ectodomain, crystallized in the post-fusion form (corresponding to panels iv and v in A). Left panel: the previously reported X-ray structure of a partially proteolysed *C. reinhardtii* HAP2 ectodomain to 3.3Å resolution. The foreground subunit is colored by domains as in A). The intervening  $\alpha 0$  helix is highlighted in brown (brown arrow). Disordered domain II loops are indicated by arrowheads. Black arrows at the bottom end mark the observed limits of the domain I C<sub>0</sub>D<sub>0</sub> and E<sub>0</sub>F<sub>0</sub> connections. Middle panel: Ribbon representation of the new 2.6Å resolution structure. The foreground subunit is shown in dark grey with regions not resolved previously highlighted in colors according to the domain they belong to (as in Figure 1A). A central light blue bar marks the molecular 3-fold axis of the

trimer. Right panel: the foreground subunit rotated as indicated and colored by domains to allow labeling the secondary structure elements buried in the trimer. In all panels, disulfide bonds are drawn as green sticks and N-linked glycan chains as sticks colored according to atom type (carbon white, nitrogen blue, oxygen red). A brown arrow in the middle and right panels marks the region of extended conformation that corresponds to the  $\alpha 0$  helix in the left panel.

(C) *C. reinhardtii* HAP2 aa sequence on a background colored by domains (as in A and B) and secondary structure elements marked. Regions not resolved in the previous structure are boxed. The remaining unstructured regions are in white background. N-linked glycosylation are marked by full or empty blue stars for resolved sugar residues or disordered glycans, respectively. The grey arrow marks the last residue (592) in the HAP2e expression construct. A dashed box frames the MPER.

### **Figure 2 MLR and CLL motifs in domain I.**

(A) The HAP2 trimer is shown at the center with boxes indicating the two insets, the  $E_0F_0$  (left) and  $C_0D_0$  (right) loops. Left panel: the dashed curved line represents the disordered central region of the  $E_0F_0$  loop (residues 245-270), with its mucin-like sequence displayed underneath with N-linked glycosylation sequons marked in blue. Right panel: The CLL motif in the  $C_0D_0$  loop with the disulfides drawn in green and labeled as in Figure 1C.

(B) Sequence alignment of the CLL motif showing its conservation across green algae, with cysteines in green the disulfide connectivity indicated. See STAR methods for abbreviations.

(C) Surface electrostatic potential of the CLL motif displayed on a scale of - 1 to 1 kT/e, from red (acidic) through neutral (white) to basic (blue).

(D) The  $C_0D_0$  loop CLL is shown at the center, next to an authentic cystine ladder motif as observed in  $\theta$ -defensins (Conibear et al., 2012) (left panel) and to an EGF-like module. The cysteines are numbered sequentially to show the different connectivity of the disulfide bonds (green).

### **Figure 3. The stem inserts into an adjacent subunit domain II groove.**

HAP2 displayed next to the viral class II glycoproteins for which the stem is resolved: RVFV Gc, (PDB code 6EGU) and RV E1 (PDB accession 4ADI) in their post-fusion conformation. In HAP2, the stem interacts with domain II of the “left” subunit in the trimer. In RVFV, it interacts with domain II of the same subunit (in yellow). In the RV E1 trimer, both domain III and the stem interact with the subunit to the right. For context, a fused membrane is schematized at the top, with hydrophilic and aliphatic moieties in light and dark grey, respectively. The inset (left panel) shows a more detailed view of the stem (magenta) inserted

in the groove formed by the *bc* loop and *ae fg*  $\beta$ -sheet of the adjacent protomer (grey). Dashed lines represent inter-chain hydrogen bonds between the stem and domain II.

#### **Figure 4. The *C. reinhardtii* HAP2 membrane insertion surface**

(A) The tip of *C. reinhardtii* HAP2 domain II (left) compared to the corresponding region in *A. thaliana* (center) and *T. cruzi* (right) (Fedry et al., 2018). The helices in fusion loops 1 and 2 are colored light orange and yellow, respectively (as in Figure 1B). Broken orange lines mark a 2-residue disordered segment in the *bc* loop in *T. cruzi* HAP2 that could potentially also insert into the target membrane. The panels were aligned on the conserved salt bridge (indicated by a continuous horizontal line), showing that the membrane would be at a variable distance from the *bdc* core  $\beta$ -sheet (the horizontal dashed lines mark roughly the limits of the polar moiety of the membrane's outer leaflet). This figure highlights the potential malleability of the membrane insertion surface.

(B) Surface representation of the HAP2 trimer tip colored by hydrophobicity. The grey stripe indicates roughly polar head groups layer of the membrane. Note the space left at the center by the projecting domain II tips of the three trimer subunits, seen on the right panel by depth cuing with white fog.

(C) Membrane binding analysis of *C. reinhardtii* WT and fusion helix mutants of HAP2 by liposome co-floitation assays on iodixanol gradients (Optiprep<sup>TM</sup>). Top (T) and bottom (B) fractions of the gradients were analyzed by WB for the presence of the HAP2 ectodomain. The right panels show the quantification of HAP2e detected in the gradient fractions. The values represent the average of three independent experiments  $\pm$  SD, analyzed by Student's *t* test. Level of significance are:  $P < 0.05$ , \*;  $P < 0.01$ , \*\*; not significant, NS.

#### **Figure 5**

**Comparison of the *C. reinhardtii* HAP2 fusion surface with viral counterparts.** The HAP2 fusion loops (left panel) form a platform projecting hydrophobic residues. As similar platform is formed by the fusion peptide of the class I influenza virus hemagglutinin (PDB code 2KXA). (second panel). The fusion loop of the RVFV (PDB code 6EGU) or the flavivirus fusion protein (PDB code 1OK8) (right panels) is significantly less extensive.

## **STAR Methods**

### **CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact, Félix A. Rey (felix.rey@pasteur.fr).

### **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

#### **Insect cell culture:**

*Drosophila* S2 cells were maintained in flasks at 28°C in HyClone medium (GE Healthcare Life Sciences) supplemented with 50 units/mL of penicillin, 50 units/mL of streptomycin and 7 µg/mL of puromycin for selection of stable transfectants. Large-scale cultures for protein production were grown in spinner flasks at 100 rpm. No information on the sex of these cells is available.

#### **Mammalian cell culture:**

Expi293F cells were grown at 37°C in Expi293 expression medium (Thermo Fisher) with rotary agitation at 135 rpm. No information on the sex of this cell line is available.

### **METHOD DETAILS**

#### **Expression construct for the single chain variable fragment (scFv) of monoclonal antibody A20**

Monoclonal antibodies targeting *C. reinhardtii* HAP2 were generated as described previously (Fedry et al., 2017). For sequencing of the antibody variable regions total RNA was prepared from A20 hybridoma cells and cDNA was synthesized using separate constant region primers for heavy and light chain. Variable regions of heavy and light chain genes were subsequently amplified using gene-specific primers and subcloned into a plasmid for insect cell expression of scFvs described previously (Gilmartin et al., 2012).

#### **Protein expression and Purification**

*Drosophila* S2 stable cell lines expressing soluble C-terminally truncated *Chlamydomonas reinhardtii* HAP2 ectodomain (residues 23-592), were generated to obtain high yields of recombinant proteins for crystallization experiments (Fedry et al., 2017). The expression construct included an enterokinase cleavage site (sequence DDDDK) immediately downstream HAP2 G592. at the C-terminal end of the ectodomain, followed by a double strep-tag for purification. The expression of the protein is under the control of metallothionein promoter, which is induced by adding 4 µM CdCl<sub>2</sub> to the cell culture medium when the cell

density reaches approximately  $7 \times 10^6$  cells/ml. At 5 days post-induction, the cells were pelleted and the soluble ectodomain was purified by affinity chromatography from the supernatant on a *StrepTactin Superflow* column. The strep tag was then removed by treatment with enterokinase and the protein was further purified by SEC on a *Superdex200* column equilibrated with 10 mM bicine pH 9.3. As we reported previously (Fedry et al., 2017), low ionic strength conditions at pH 9.3 reduce formation of HAP2 hexamers. As shown in Figure S1, hexamers appear to form by head-to-head interaction between two trimers via the hydrophobic fusion loops, thus interfering with HAP2 binding to membranes.

The yields of HAP2 ectodomain obtained from S2 cells was around 15 mg/l of cell culture. The purified protein was concentrated to approximately 5 mg/ml.

We also generated an S2 stable cell line expressing antibody A20 scFv, which was expressed following the same procedures used for the HAP2 ectodomain. The complex between the HAP2 ectodomain and scFv A20 was obtained mixing both proteins in a molar ratio 1:2 and incubating overnight at 4°C. The HAP2-scFv A20 complex was then separated from unbound scFv on a *Superdex200* column. The fractions in Figure S1, right panel, corresponding to the complex were pooled and then concentrated to approximately 7 mg/ml.

### **Crystallization and Structure Determination**

Crystals were grown using the sitting drop method at 293K by mixing HAP2-scFv A20 protein solution in a ratio 1:1 with reservoir solution containing 0.2M  $\text{LiSO}_4$ , 0.1M Tris-HCl pH 8.5 and 25% w/v PEG 5000 MME. Small hexagonal plates obtained after 5 days were cryoprotected with reservoir solution containing 15% glycerol and plunged in liquid nitrogen. Data collection was carried out at the SOLEIL synchrotron (St Aubin, France). Data were processed, scaled and reduced with XDS (Kabsch, 1988) and AIMLESS (Evans and Murshudov, 2013). Diffraction anisotropy was corrected with DEBYE and STARANISO programs using the STARANISO server (<http://staraniso.globalphasing.org/>). These programs perform an anisotropic cut-off of merged intensity data, a Bayesian estimation of structure amplitudes and apply an anisotropic correction to the data. The HAP2 structure was determined by molecular replacement using PHASER and the coordinates of the previous crHAP2e ectodomain (PDB accession code 5MF1) as search model. The model was corrected and completed using COOT (Emsley et al., 2010) and the corrected anisotropic amplitudes were used for model refinement in BUSTER (Bricogne et al., 2010). The stereochemical properties of the model were validated with MolProbity (Chen et al., 2010).

### **Design and expression of HAP2 fusion loops mutants**

Recombinant HAP2 ectodomain and fusion loops mutants for liposome co-flotation experiments were obtained by transient transfection of mammalian cells, by using the pcDNA3.1(+) plasmid with the gene coding for the HAP2 ectodomain (residues 23-592) followed by a double strep-tag at the C-terminal end of the ectodomain. Mutants at positions 173, 177, 192, 193, 197, 200 and 201 were generated by standard PCR methods. Expi293 cells were transfected with WT or mutant constructs using polyethylenimine (PEI, 1 mg/ml) in a DNA:PEI ratio 1:4. Cell culture supernatants were harvested at 5 days post-transfection and protein was purified by affinity chromatography as mentioned above. As for the protein produced in S2 cells, the purified proteins were buffer exchanged to 10 mM bicine pH 9.3 and concentrated to 1 mg/ml, to minimize hexamer formation.

The yields of HAP2 WT expressed in mammalian cells was 200 µg of pure protein per 30 ml of cell culture. The yields of all HAP2 fusion loop mutants tested was in the range of 1 to 2 times the yield of the WT protein.

### **Liposome Co-flotation Experiments**

Liposomes composed of DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine), DOPC(1,2-dioleoyl- sn-glycero-3-phosphocholine), cholesterol and sphingomyelin in a molar ratio 1/1/3/1, were prepared by the freeze-thaw and extrusion method. 1µM protein was mixed with 1 mM freshly prepared liposomes and incubated for 1h at 25°C in 100 mL PBS. Samples were adjusted to a final concentration of 40% Optiprep™ (in PBS), loaded in centrifuge tubes and overlaid with 4.5 mL 20% Optiprep™ and 0.3 mL PBS. Centrifugations were performed overnight at 4°C and 192000 x g on SW55ti rotor. Top and bottom fractions of the gradient were analyzed by immunoblotting using anti-Strep tag antibody.

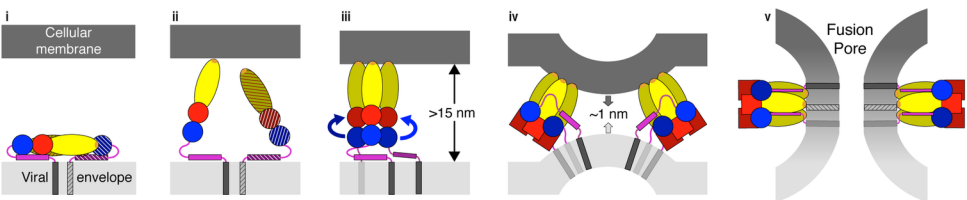
**Illustrations:** Cartoon representations were made with PyMol (Schrodinger, 2015. ). The HAP2 amino acid sequence alignments displayed were made by MUSCLE (Edgar, 2004) and rendered with ESPript 3.0 (Robert and Gouet, 2014). The organisms represented in Figure 2 have the following abbreviations: *Chlamydomonas reinhardtii* (Cr), *Gonium pectoral* (Gp), *Volvox carterii* (forms A and B; VcA and VcB respectively), *Monoraphidium neglectum* (Mn), *Coccomyxa subellipsoidea*, (Cs), *Auxenochlorella protothecoides* (Ap), *Chlorella variabilis* (Cv), *Chlorella sorokiniana* (Csor) and *Micractinium conductrix* (Mc).

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

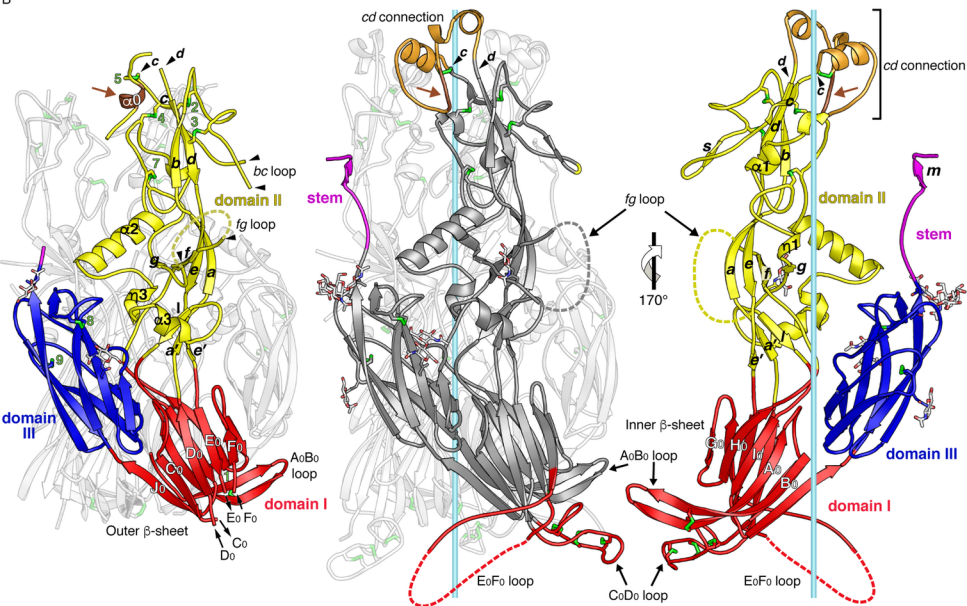
Densitometry analysis of western blots from liposome co-flotation assays was performed with ImageJ software (Schneider et al., 2012) and expressed as percentage of HAP2e in top and bottom fractions of iodixanol gradients. Results are shown as mean  $\pm$  SD of three independent experiments (n=3). Differences in liposome binding between HAP2e WT and mutants were analyzed by a Student's t-test (P<0.05, \*; P<0.01, \*\*).

## **DATA AND SOFTWARE AVAILABILITY**

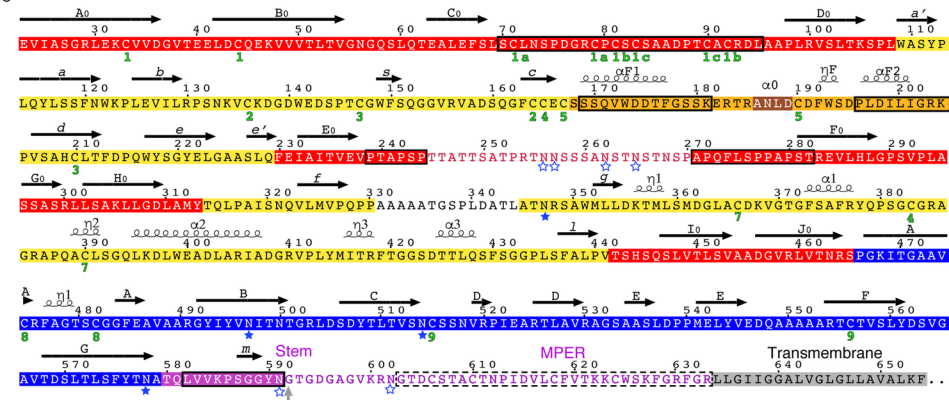
The atomic coordinates and structure factors of the have been deposited in the protein data bank under the accession code 6E18.



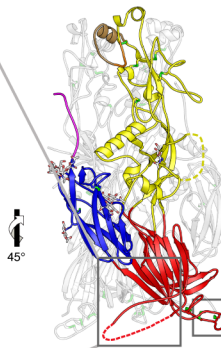
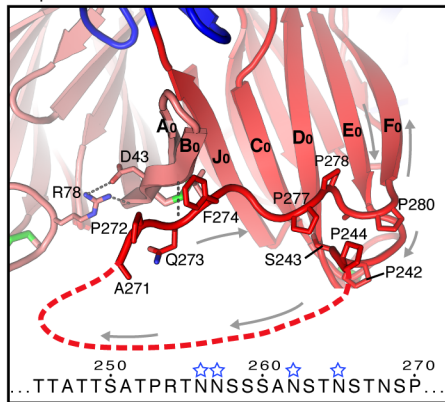
**B**



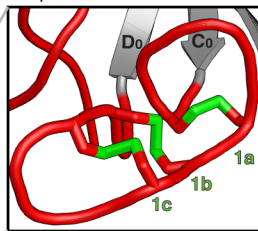
**C**



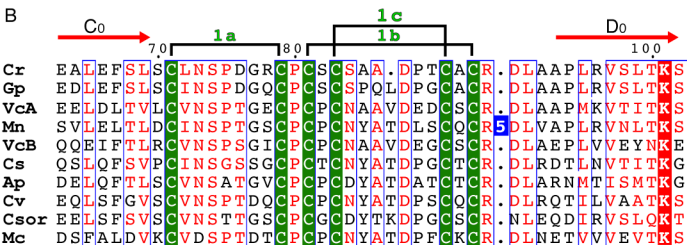
### Loop EoF0



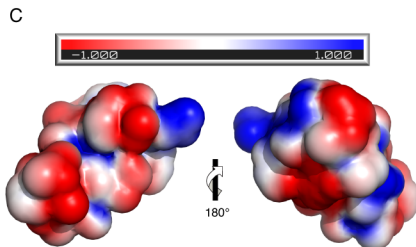
### Loop CoD0



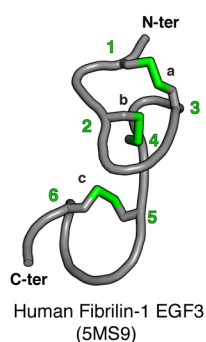
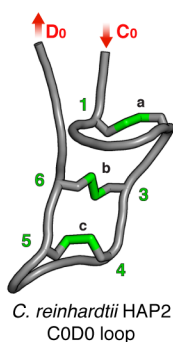
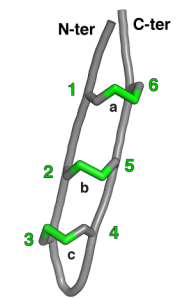
B

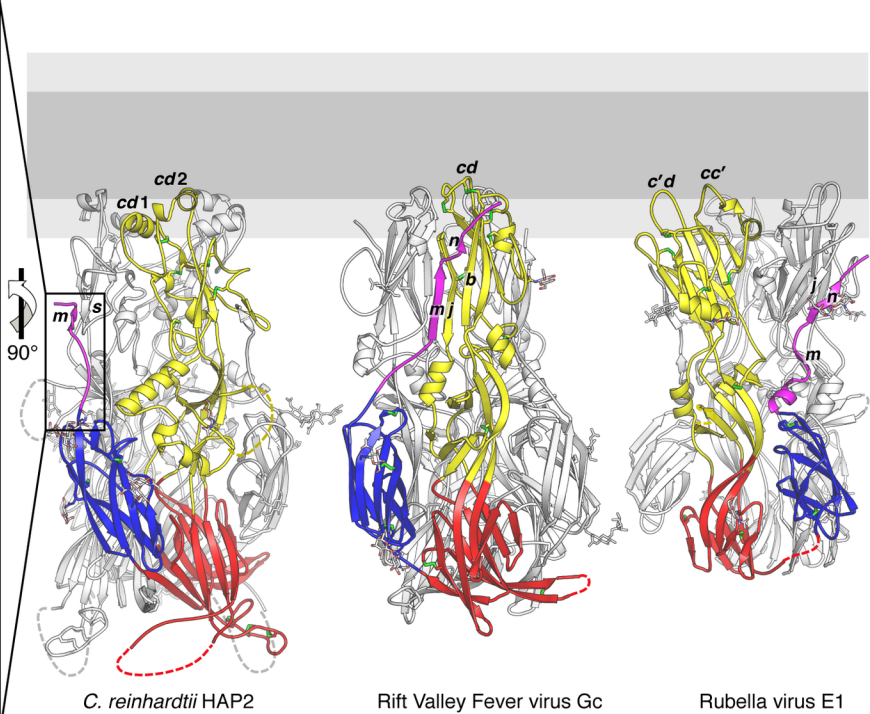
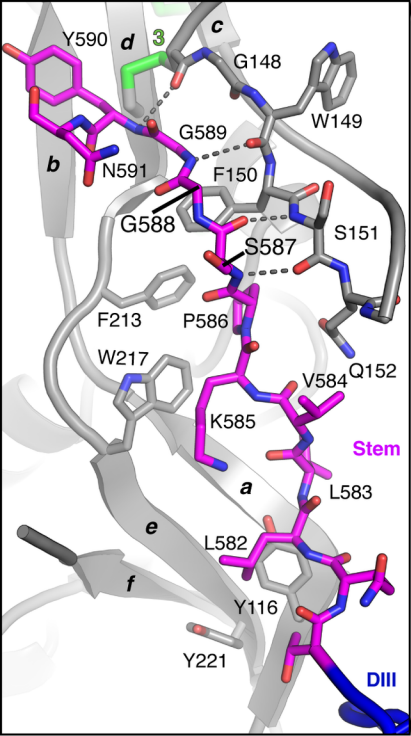


C

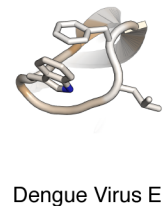
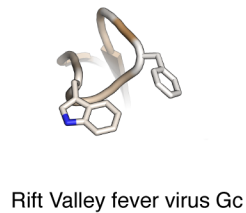
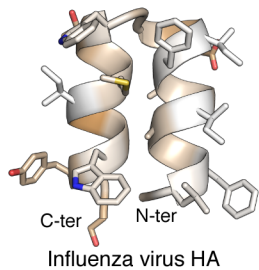
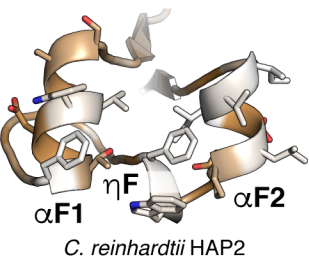
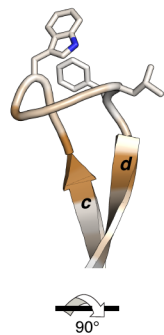
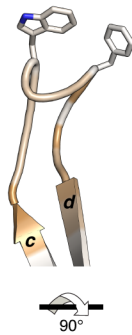
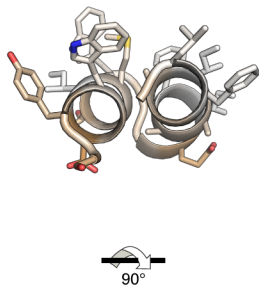
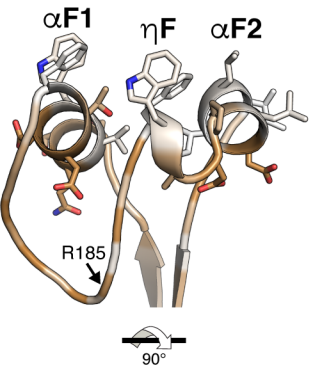


D









**Table 1. Data collection and refinement statistics.**

<b>Wavelength</b>	0.9785
<b>Resolution range</b>	38.99 - 2.601 (2.694 - 2.601)
<b>Space group</b>	P 63 2 2
<b>Unit cell</b>	125.64 125.64 367.99 90 90 120
<b>Total reflections</b>	75352 (1152)
<b>Unique reflections</b>	37676 (604)
<b>Multiplicity</b>	2.0 (2.0)
<b>Completeness (%)</b>	70.66 (11.48)
<b>Mean I/sigma(I)</b>	12.32 (1.58)
<b>Wilson B-factor</b>	54.09
<b>R-merge</b>	0.05924 (0.4293)
<b>R-meas</b>	0.08377 (0.6071)
<b>R-pim</b>	0.05924 (0.4293)
<b>CC1/2</b>	0.995 (0.66)
<b>CC*</b>	0.999 (0.892)
<b>Reflections used in refinement</b>	38050 (604)
<b>Reflections used for R-free</b>	1909 (28)
<b>R-work</b>	0.2244 (0.3323)
<b>R-free</b>	0.2431 (0.3191)
<b>CC(work)</b>	0.921 (0.743)
<b>CC(free)</b>	0.866 (0.598)
<b>Number of non-hydrogen atoms</b>	4136
macromolecules	3943
ligands	111
solvent	82
<b>Protein residues</b>	529
<b>RMS(bonds)</b>	0.010
<b>RMS(angles)</b>	1.60
<b>Ramachandran favored (%)</b>	97.32
<b>Ramachandran allowed (%)</b>	2.49
<b>Ramachandran outliers (%)</b>	0.19
<b>Rotamer outliers (%)</b>	1.37
<b>Clashscore</b>	1.75

