

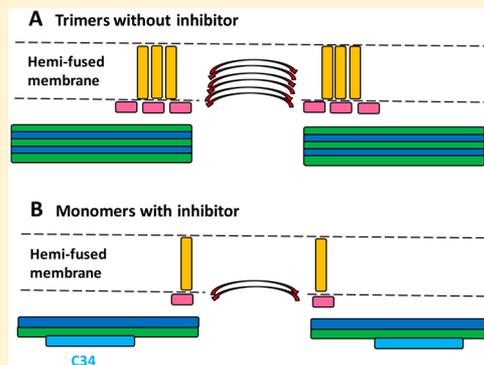
# Efficient Fusion at Neutral pH by Human Immunodeficiency Virus gp41 Trimers Containing the Fusion Peptide and Transmembrane Domains

S. Liang, P. U. Ratnayake, C. Keinath, L. Jia, R. Wolfe, A. Ranaweera, and D. P. Weliky\*<sup>1</sup>

Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

## Supporting Information

**ABSTRACT:** Human immunodeficiency virus (HIV) is membrane-enveloped, and an initial infection step is joining/fusion of viral and cell membranes. This step is catalyzed by gp41, which is a single-pass integral viral membrane protein. The protein contains an ~170-residue ectodomain located outside the virus that is important for fusion and includes the fusion peptide (FP), N-helix, loop, C-helix, and viral membrane-proximal external region (MPER). The virion initially has noncovalent complexes between three gp41 ectodomains and three gp120 proteins. A gp120 contains ~500 residues and functions to identify target T-cells and macrophages via binding to specific protein receptors of the target cell membrane. gp120 moves away from the gp41 ectodomain, and the ectodomain is thought to bind to the target cell membrane and mediate membrane fusion. The secondary and tertiary structures of the ectodomain are different in the initial complex with gp120 and the final state without gp120. There is not yet imaging of gp41 during fusion, so the temporal relationship between the gp41 and membrane structures is not known. This study describes biophysical and functional characterization of large gp41 constructs that include the ectodomain and transmembrane domain (TM). Significant fusion is observed of both neutral and anionic vesicles at neutral pH, which reflects the expected conditions of HIV/cell fusion. Fusion is enhanced by the FP, which in HIV/cell fusion likely contacts the host membrane, and the MPER and TM, which respectively interfacially contact and traverse the HIV membrane. Initial contact with vesicles is made by protein trimers that are in a native oligomeric state that reflects the initial complex with gp120 and also is commonly observed for the ectodomain without gp120. Circular dichroism data support helical structure for the N-helix, C-helix, and MPER and nonhelical structure for the FP and loop. Distributions of monomer, trimer, and hexamer states are observed by size-exclusion chromatography (SEC), with dependences on solubilizing detergent and construct. These SEC and other data are integrated into a refined working model of HIV/cell fusion that includes dissociation of the ectodomain into gp41 monomers followed by folding into hairpins that appose the two membranes, and subsequent fusion catalysis by trimers and hexamers of hairpins. The monomer and oligomer gp41 states may therefore satisfy dual requirements for HIV entry of membrane apposition and fusion.



Human immunodeficiency virus (HIV) is enveloped by a membrane obtained during budding from an infected host cell. Infection of a new cell begins with joining (fusion) of membranes of the virus and host cell, and this process is catalyzed by the ~41 kDa glycoprotein “gp41”, which is single-pass integral viral membrane protein.<sup>1,2</sup> gp41 also contains an ~170-residue ectodomain and an ~150-residue endodomain that are located outside and inside the virus, respectively (Figure 1A). gp41 is synthesized as the second subunit of a larger gp160 precursor protein, and following proteolytic cleavage, the first subunit gp120 forms a noncovalent complex with the gp41 ectodomain and contains three gp41 and three gp120 molecules. We use the residue numbering scheme for gp41 based on the gp160 precursor, so that the N-terminus of gp41 is residue 512. Host cells are identified by HIV via gp120 binding to primary CD4 and secondary CXCR4 and CCR5 receptors, followed by separation of gp120 from gp41 and a structural rearrangement of the gp41 ectodomain. Muta-

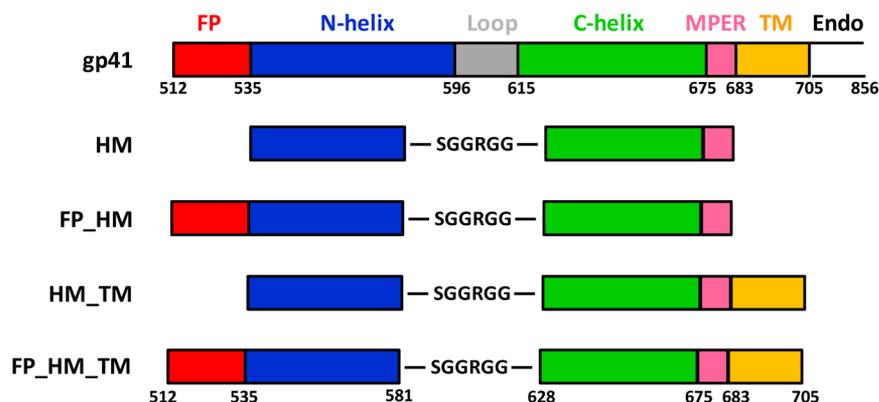
genesis/fusion relationships for gp160-mediated cell/cell fusion support a primary role for the gp41 ectodomain in fusion.<sup>3,4</sup> There are structures of the initial complex of the gp41 ectodomain with gp120, with typical resolution of 3–5 Å,<sup>5–10</sup> and high-resolution structures of soluble regions of the gp41 ectodomain without gp120.<sup>11–14</sup> For the initial complex with gp120, the gp41 ectodomain exhibits a set of distinct  $\alpha$  helices connected by loops. The central  $\alpha 7$  helix (gp41<sub>572–595</sub>) forms an interior parallel trimeric bundle with the N-terminal end pointing toward gp120 and the C-terminal end likely pointing toward the viral membrane. A loop from the  $\alpha 7$  N-terminus connects to the  $\alpha 6$  helix (gp41<sub>530–543</sub>) and then the N-terminal fusion peptide (FP) domain (gp41<sub>512–529</sub>), with extended structure for the FP. The FP is exposed and an epitope of

Received: August 4, 2017

Revised: January 16, 2018

Published: January 18, 2018

## A Protein constructs



## B Protein sequences

<b>HM</b>			CTLTVQA	RQLLSGIVQQ	QNNLLRAIEA	QQHLLQLTVW
	GIKQLQARIL	SGGRGGWMEW	DREINNYTSL	IHSLIEESQN	QQEKNEQELL	ELDKWASLWN
	WFNITNWLWY	IKGGGGGGL	HHHHHH			
<b>HM_TM</b>			CTLTVQA	RQLLSGIVQQ	QNNLLRAIEA	QQHLLQLTVW
	GIKQLQARIL	SGGRGGWMEW	DREINNYTSL	IHSLIEESQN	QQEKNEQELL	ELDKWASLWN
	WFNITNWLWY	IKLFIMIVGG	LVGLRIVFAV	LSIVGGGGGGL	HHHHHH	
<b>FP_HM</b>	AVGIGALFLG	FLGAAGSTMG	ARSMTLTVQA	RQLLSGIVQQ	QNNLLRAIEA	QQHLLQLTVW
	GIKQLQARIL	SGGRGGWMEW	DREINNYTSL	IHSLIEESQN	QQEKNEQELL	ELDKWASLWN
	WFNITNWLWY	IKGGGGGGL	HHHHHH			
<b>FP_HM_TM</b>	AVGIGALFLG	FLGAAGSTMG	ARSMTLTVQA	RQLLSGIVQQ	QNNLLRAIEA	QQHLLQLTVW
	GIKQLQARIL	SGGRGGWMEW	DREINNYTSL	IHSLIEESQN	QQEKNEQELL	ELDKWASLWN
	WFNITNWLWY	IKLFIMIVGG	LVGLRIVFAV	LSIVGGGGGG	GGLEHHHHHH	

**Figure 1.** (A) Schematic diagrams of full-length HIV gp41 and the four truncated constructs of this study with domains and corresponding colors: FP  $\equiv$  fusion peptide, red; N-helix, blue; loop, gray; C-helix, green; MPER  $\equiv$  membrane-proximal external region, pink; TM  $\equiv$  transmembrane domain, orange; and endo = endodomain, white. The four constructs have non-native SGGRGG replacing native residues 582–627. (B) Amino acid sequences with colors matching segments in panel A and the non-native C-terminal  $G_6LEH_6$  or  $G_8LEH_6$  segments colored black. The  $H_6$  is for  $Co^{2+}$  affinity chromatography, and the  $G_6LE$  and  $G_8LE$  are necessary spacers for exposure of the  $H_6$  tag. The sequence is from the HXB2 laboratory strain of HIV.

neutralizing antibodies.<sup>15,16</sup> A loop from the  $\alpha 7$  C-terminus connects to the  $\alpha 8$  helix (gp41<sub>619–623</sub>) and then the  $\alpha 9$  helix (gp41<sub>628–664</sub>). The  $\alpha 6$ ,  $\alpha 8$ , and  $\alpha 9$  helices are outside and approximately perpendicular to the  $\alpha 7$  bundle and are also on three sides of the bundle. Comparison of this initial gp41 ectodomain structure in complex with gp120 to the later gp41 structure without gp120 shows large-scale extensions and topological rearrangements of helical segments. The latter structure is a “trimer of hairpins” with three N-helices (gp41<sub>536–596</sub>) that form an interior bundle, followed by loops, and then C-helices (gp41<sub>616–675</sub>) that run antiparallel in the exterior grooves of the bundle.<sup>11–14</sup> The hairpin is probably the final structure during fusion based on a  $T_m$  of  $\approx 110$  °C.<sup>17,18</sup> The color coding in Figure 1A reflects the N- and C-helices of the ectodomain structure without gp120.

Membrane structures during fusion have also been characterized by electron and fluorescence microscopies and include (1) “hemifusion” in which a single bilayer diaphragm maintains separation of the HIV and host cell contents, (2) “pore formation” of channels in the diaphragm that permit diffusion of  $<1$  kDa species, and (3) “pore expansion” in which the pores irreversibly coalesce to break the diaphragm and permit complete mixing of the viral and cell contents.<sup>2,19</sup> There has not yet been imaging of the gp41 ectodomain during fusion, so there is only limited knowledge of the temporal relationship between the membrane states and gp41 structural states during

fusion, and therefore only limited knowledge of which fusion step(s) is catalyzed by which gp41 structure(s). Current fusion mechanistic models are often based on the observation that individual gp41 N- and C-helix peptides inhibit pore formation but not pore expansion.<sup>19</sup> Such peptides would likely not tightly bind to a final-stage trimer-of-hairpins structure but could bind to a variety of other gp41 structures, which supports the hypothesis that the peptides bind to earlier-stage gp41 structures and prevent formation of the final structure.<sup>20</sup> Electron microscopy from the 1980s onward has shown that HIV can enter cells via direct fusion with the plasma membrane or via endocytosis followed by fusion with the endocytic membrane.<sup>1,21</sup> The HIV-containing endosomes likely remain at neutral pH, which contrasts with endocytic entry by other enveloped viruses, e.g., influenza, for which reduction of endosomal pH triggers conformational changes in the viral fusion protein and subsequent fusion of the viral and endocytic membranes.<sup>22</sup> Some of the most recent data support direct fusion of HIV and plasma membranes as the primary route of HIV infection.<sup>23</sup>

This study compares different gp41 constructs in the absence of gp120 and provides data about their structures, oligomeric states, and catalysis of vesicle fusion. The data presented here and earlier are combined to create an integrated model of gp41-mediated membrane fusion. Figure 1A displays gp160-based residue numbering of gp41 with color coding of different

domains with defined structures and/or functions. The “fusion peptide” (FP) includes 16 apolar residues at the gp41 N-terminus, and the name reflects impairment of gp160-mediated fusion when there are deletions and/or mutations in the FP.<sup>3,24</sup> The “N-helix” and “C-helix” regions are each ~60-residue continuous helices in the final-state six-helix bundle structure adopted by the gp41 ectodomain in the absence of gp120.<sup>12–14</sup> As discussed above, shorter helical regions with a different tertiary structure are formed in the initial ectodomain complex with gp120. The membrane-proximal external region (MPER) is proposed to interfacially bind the viral membrane and is adjacent to the transmembrane (TM) region.<sup>25,26</sup> Although the MPER is commonly defined as gp41<sub>662–683</sub>, gp41<sub>662–675</sub> is structurally the C-terminal end of the C-helix in large ectodomain constructs.<sup>13</sup> The structural distinction between MPER and TM regions is also unclear, as fairly continuous helical structure is adopted by peptides corresponding to gp41<sub>671–693</sub>, gp41<sub>683–704</sub>, and gp41<sub>677–716</sub>.<sup>27,28</sup>

Models of gp160-mediated fusion usually show movement of gp120 away from gp41 after binding of the receptor to gp120, followed by a conformational change in the freed gp41 and then binding of the gp41 FP in the target membrane.<sup>29</sup> These steps all occur prior to membrane fusion. The fusion relevance of gp41 without gp120 is also supported by the common observation of fusion of vesicles after addition of gp41 constructs to the vesicle solution.<sup>30</sup> gp41 constructs catalyze fusion for a variety of lipid compositions that often include phosphatidylcholine, as well as other lipids such as cholesterol, phosphatidylethanolamine, phosphatidylserine, phosphatidylglycerol, and sphingomyelin.<sup>30–32</sup> There are increased rates and extents of vesicle fusion when the FP and MPER are included in the N-helix/loop/C-helix construct.<sup>29</sup>

gp41-only constructs with longer N- and C-helices are typically soluble at only pH <4 and visibly aggregate at physiologic pH, so most biophysical work has been done at low pH. Atomic-resolution structures of N-helix/loop/C-helix constructs with >2 mM protein show trimers of hairpins with an interior bundle of three parallel N-helices, and three C-helices packed in the exterior grooves of the bundle and antiparallel to the N-helix direction.<sup>11,12</sup> Analytical ultracentrifugation (AUC) sedimentation equilibrium data have been interpreted to support  $K_a \approx 10^{12} \text{ M}^{-2}$  for the three monomers  $\leftrightarrow$  one trimer equilibrium, which corresponds to equal trimer and monomer masses when the total protein concentration is  $\approx 1 \mu\text{M}$ .<sup>33,34</sup> The equilibrium has also been observed in elution peaks via size-exclusion chromatography (SEC), but the data are consistent with  $K_a \approx 10^8 \text{ M}^{-2}$ , with equal monomer and trimer masses when the total protein concentration is  $\approx 60 \mu\text{M}$ .<sup>29,35</sup> We do not understand the discrepancy of AUC versus SEC. Dodecylphosphocholine detergent reduces  $K_a$ , i.e., favors monomer versus trimer.<sup>36–39</sup> A hairpin structure for the monomer similar to that in the trimer is supported by the negligible dependence of circular dichroism (CD) and nuclear magnetic resonance (NMR) signals on protein concentration, as well as retention of a hyperthermostable  $T_m$  of >100 °C.<sup>18,40</sup> Monomers are appealing in fusion mechanistic models because after gp120 moves away from gp41, asynchronous conformational changes of individual gp41 monomers into hairpins are topologically easier than concerted changes of a trimer.<sup>29,36,41</sup> Monomers may also be a binding target of gp41 N- and C-helix peptides that have been shown to be effective at fusion inhibition up to the final pore expansion step.<sup>19</sup>

Detergent-associated peptides with sequences corresponding to FP, MPER, and TM sequences are typically monomers with  $\alpha$  helical structure.<sup>27,42–44</sup> The membrane-associated FP often forms small oligomers with intermolecular antiparallel  $\beta$  sheet structure.<sup>31,45</sup> Peptides corresponding to FP or MPER sequences often catalyze vesicle fusion, and the level of fusion is increased when FP or MPER is appended to a hairpin construct with an N-helix/loop/C-helix sequence.<sup>25,29,31</sup> One caveat is that significant fusion with these constructs required a pH of  $\leq 4$  and anionic vesicles.<sup>46</sup> Fusion by hairpin constructs therefore correlated with electrostatic attraction between highly cationic protein (charge of  $\approx 10$ ) and anionic vesicles. There is likely much less attraction for HIV/cell fusion at physiologic pH because the ectodomain is approximately neutral. One important result of this study is significant fusion at physiologic pH by hairpin constructs that contain the FP, MPER, and TM regions. There are similar results with neutral and anionic vesicles, which supports a major contribution to fusion from hydrophobic rather than electrostatic protein/membrane interaction, which is similar to HIV/cell fusion.

## ■ MATERIALS AND METHODS

**Materials.** Materials were purchased from the following companies: DNA from GenScript (Piscataway, NJ), *Escherichia coli* BL21(DE3) strain from Novagen (Gibbstown, NJ), Luria-Bertani (LB) medium from Dot Scientific (Burton, MI), isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) from Goldbio (St. Louis, MO), cobalt affinity resin from Thermo Scientific (Waltham, MA), and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (sodium salt) (POPG) from Avanti Polar Lipids (Alabaster, AL). Most other materials were obtained from Sigma-Aldrich (St. Louis, MO).

**Protein Sequences.** Figure 1 displays schematic diagrams and amino acid sequences of the four gp41 constructs used in this study. The sequences are from the HXB2 laboratory strain of HIV and have the gp160 precursor residue numbering, residues 1–511 and 512–856 for the gp120 and gp41 subunits, respectively. Figure S1 displays the corresponding DNA sequences. In the absence of gp120, the gp41 soluble ectodomain (SE) adopts thermostable hairpin structure, i.e., N-helix, 180° turn, and C-helix, consisting of approximately residues 536–596, 597–615, and 616–675, respectively.<sup>12,13</sup> Our constructs have residues 582–627 replaced with the six-residue non-native SGGRGG, and SE and FP+SE constructs with this replacement still adopt highly helical and thermostable SE structure.<sup>47</sup> The four constructs used in this study are as follows: (1) HM (“hairpin” + MPER), a SE construct, gp41<sub>535–581/SGGRGG/628–683</sub>; (2) FP\_HM, a FP+SE construct, gp41<sub>512–581/SGGRGG/628–683</sub>; (3) HM\_TM, a SE+TM construct, gp41<sub>535–581/SGGRGG/628–705</sub>; and (4) FP\_HM\_TM, a FP+SE+TM construct, gp41<sub>512–581/SGGRGG/628–705</sub>. All constructs have a non-native H<sub>6</sub> affinity tag at their C-termini that is preceded by a G<sub>6</sub>LE or G<sub>8</sub>LE spacer. The spacer is needed for exposure of the affinity tag during purification. HM and HM\_TM have a M535C mutation, which is needed for native chemical ligation with the FP, which is not part of this study.

**Protein Expression.** Each DNA insert was subcloned into a pET-24a(+) vector that contained the Lac operon and kanamycin antibiotic resistance. The plasmid was transformed into the *E. coli* BL21(DE3) strain. Typical bacterial culture conditions included use of (1) LB medium that contained 50 mg kanamycin/L to select for bacterial cells that contained the

plasmid and (2) 180 rpm shaking at 37 °C. Bacterial stocks were prepared by mixing equal (0.5 mL) volumes of overnight culture and 50% (v/v) glycerol, followed by freezing and storage at -80 °C. New cultures were prepared with overnight growth of 50  $\mu$ L of a bacterial glycerol stock in 50 mL of medium, followed by addition of 1 L of fresh medium and growth for 2 h until the OD<sub>600</sub> reached  $\approx$ 0.8. Protein expression was induced at 37 °C for 5 h with addition of 2 mmol of isopropyl  $\beta$ -D-1-thiogalactopyranoside.

**Separation of the Cellular Fraction Enriched with Inclusion Bodies.** Subsequent purification of recombinant protein (RP) was mostly performed at 4 °C. First, the cell pellet was harvested by centrifugation (9000g for 10 min) with subsequent storage at -20 °C. Wet cells (5 g) were lysed by tip sonication in 40 mL of phosphate-buffered saline (PBS) at pH 7.4. Initial separation of RP from other cellular components was based on most RP being in inclusion bodies, which are insoluble in PBS and can be pelleted by modest centrifugation (48000g for 20 min). The supernatant contains soluble proteins and other molecules, as well as suspended membrane fragments that are only effectively pelleted by >200000g centrifugation. A solid that is highly enriched with RP inclusion bodies was therefore obtained by initial modest centrifugation, and then twofold repetition of resuspension in 40 mL of fresh PBS and modest centrifugation.

**<sup>13</sup>C NMR of the Inclusion Body Fraction.** The relative expression quantity of each RP construct was assessed using a 50 mL expression in medium with [1-<sup>13</sup>C]Gly, followed by separation of the cell fraction enriched with inclusion bodies, and <sup>13</sup>C NMR of this fraction.<sup>48</sup> The RP is synthesized mostly during the expression period, so the RP quantity correlates with labeled Gly incorporation and <sup>13</sup>C NMR signal intensity. Cells were grown overnight in 50 mL of LB medium and then transferred to 50 mL of minimal medium containing 250  $\mu$ L of 50% glycerol, and 10 mg each of [1-<sup>13</sup>C]glycine and the 19 other unlabeled amino acids. These unlabeled amino acids inhibit amino acid biosynthesis and therefore metabolic scrambling of the [1-<sup>13</sup>C]Gly. After growth for an additional 2 h, expression was induced for 5 h, followed by sonication in PBS and centrifugation. This pellet was lyophilized, and an  $\sim$ 50  $\mu$ L aliquot was then transferred to a magic-angle-spinning (MAS) NMR rotor with a 4 mm diameter. A <sup>13</sup>C NMR spectrum was then acquired using an instrument with a 9.4 T magnet and an Agilent Infinity Plus console. The pulse sequence was <sup>1</sup>H  $\rightarrow$  <sup>13</sup>C cross-polarization followed by <sup>13</sup>C acquisition with <sup>1</sup>H decoupling, and a spectrum was the sum of  $\sim$ 11000 scans.

**Protein Purification.** The pellet enriched with inclusion bodies was completely solubilized by tip sonication in 40 mL of PBS at pH 7.4 that also contained 8 M urea, 0.5% sodium dodecyl sulfate (SDS), and 0.8% sodium lauroyl sarcosinate (Sarkosyl). Co<sup>2+</sup> resin (1 mL) was added, and the resin/solution was agitated for 2 h at ambient temperature to bind RP to resin. The solution containing unbound protein was removed by gravity filtration of the suspension.

At this point, separate protocols were used to obtain RP soluble in either SDS or dodecylphosphocholine (DPC). The first step in the SDS protocol was further removal of unbound protein with three resin washes with 1 mL aliquots of the PBS/urea/SDS/Sarkosyl solution, and subsequent gravity filtration. Bound protein was eluted from the resin by 0.5 mL aliquots (four times) of the PBS/urea/SDS/Sarkosyl solution with 250 mM imidazole, and gravity filtration. The eluent fractions were

pooled and then mixed with an equal volume of buffer that contained 10 mM Tris-HCl (pH 8.0), 0.17% *n*-decyl- $\beta$ -D-maltoside, 2 mM EDTA, and 1 M L-arginine, with subsequent agitation overnight at 4 °C.<sup>49</sup> Arginine, urea, and other detergents were removed by dialysis against 10 mM Tris (pH 7.4) with 0.2% SDS. If the intermediate mixing step with the arginine solution was skipped, RP precipitated during dialysis, which suggests that RP aggregates are broken up by the arginine solution.

The procedure described above was attempted using DPC rather than SDS in the dialysis buffer, but RP consistently precipitated, presumably because aggregation associated with loss of urea and arginine occurred faster than solubilization by DPC. We therefore tried buffer exchange with RP still bound to the column.<sup>41</sup> Exchange was accomplished by a threefold suspension of the resin in 1 mL of 20 mM sodium phosphate buffer (pH 7.4) that also contained 0.25% DPC, with subsequent gravity filtration. RP was then eluted in the pH 7.4 phosphate buffer with 0.25% DPC and 250 mM imidazole. The eluent fractions were pooled, and imidazole was removed by dialysis against pH 7.4 phosphate buffer with 0.25% DPC, or by dialysis against 20 mM sodium acetate buffer (pH 3.2) with 0.25% DPC. Purified RP concentrations were determined using A<sub>280</sub>.

**Circular Dichroism (CD).** CD spectra at ambient temperature were recorded with a Chirascan instrument (Applied Photophysics) equipped with a quartz cuvette with a 1 mm path length and a 190–260 nm wavelength range scanned in 0.5 nm steps. The RP concentration was  $\approx$ 10  $\mu$ M, and each spectrum was the (protein + buffer) - buffer difference. For some samples, a temperature series of CD spectra were acquired in 5 °C increments over a 25–90 °C range. These latter spectra were acquired using a J-810 instrument (Jasco) and a water circulation bath. There was no visible precipitation at any temperature, or after cooling, which reveals retention of soluble RP.

**Size-Exclusion Chromatography (SEC).** The RP solution was first dialyzed against the SEC running buffer and then concentrated to  $\sim$ 1 mg/mL RP, with subsequent separation of particulates using centrifugation at 9000g for 10 min. There was no visible precipitate for samples in SDS and a very small precipitate for samples in DPC. An  $\sim$ 100  $\mu$ L aliquot was injected into a DuoFlow Pathfinder 20 instrument (Bio-Rad) equipped with a Tricorn Superdex 200 Increase 10/300 GL column (GE Technologies). Instrument parameters included a flow rate of 0.3 mL/min, A<sub>280</sub> detection, and  $\sim$ 10-fold dilution in the column, i.e.,  $\sim$ 0.1 mg/mL RP.

**Protein-Induced Vesicle Fusion.** Fusion activities of the different gp41 constructs were assayed by protein-induced lipid mixing between unilamellar vesicles. Assays were performed for vesicles with two different lipid compositions: (1) 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho(1'-*rac*-glycerol) (POPG), and cholesterol (Chol) at an 8:2:5 molar ratio and (2) POPC and Chol at a 2:1 molar ratio. The Chol mole fraction in both compositions is close to that of the plasma membrane of cells infected with HIV.<sup>50</sup> Compositions 1 and 2 have net negative and neutral charges, respectively, so comparison between them allows assessment of the impact of protein/vesicle electrostatics on fusion. HIV gp41 likely makes initial contact with the outer leaflet of the cell membrane that normally has little anionic lipid, but recent data support scramblase-mediated transport of anionic phosphatidylserine lipid from the inner to outer leaflet

prior to gp160-mediated fusion.<sup>51</sup> Vesicle preparation began with dissolution of POPC, POPG, and Chol (1.6, 0.4, and 1.0  $\mu\text{mol}$ , respectively) or POPC and Chol (2.0 and 1.0  $\mu\text{mol}$ , respectively) in 1 mL of chloroform, followed by solvent removal using dry nitrogen gas and then overnight vacuum. The dry lipid films were then suspended in 2 mL of aqueous buffer and subjected to 10 freeze/thaw cycles to create unilamellar vesicles. Subsequent extrusion (10-fold) of the suspension through a polycarbonate membrane with 100 nm diameter pores resulted in vesicles with typical diameters of 200–300 nm, as observed by electron microscopy.<sup>52</sup> For each composition, a set of companion “labeled” vesicles were also prepared that contained additional 2 mol % fluorescent lipid *N*-(7-nitro-2,1,3-benzoxadiazol-4-yl)phosphatidylethanolamine and 2 mol % quenching lipid *N*-(lissamine rhodamine B sulfonyl) phosphatidylethanolamine. Systematic differences among vesicle compositions were minimized by weighing all lipids at the same time using the same scale, and lipid loss was estimated to be  $\sim 30\%$  during extrusion. The final vesicle solution contained POPC, POPG, and Chol at a total concentration of  $\approx 225 \mu\text{M}$  and a 1:9 labeled vesicle:unlabeled vesicle ratio. The solution was transferred to a quartz cuvette in a fluorimeter and subjected to constant stirring at 37 °C. Fluorescence was monitored using 467 nm excitation, 530 nm detection, and 1 s time increment. The initial baseline fluorescence,  $F_0$ , was determined, and an aliquot of a protein stock solution was then added ( $t = 0$ ). The stock solution contained 40  $\mu\text{M}$  protein in 10 mM Tris buffer (pH 7.4) with 150 mM NaCl and 0.2% SDS. The time-dependent fluorescence increase [ $\Delta F(t) = F(t) - F_0$ ] was diagnostic of protein-mediated fusion between a labeled and unlabeled vesicle. Relative to the initial labeled vesicle, the fused vesicle has a higher fluorescence because of the longer average fluorophore–quencher distance. The dead time after protein addition was  $\sim 5$  s, and the final asymptotic fluorescence change was usually achieved by 600 s. A 12  $\mu\text{L}$  aliquot of 10% Triton X-100 was then added to solubilize the vesicles, with a corresponding maximum fluorescence change ( $\Delta F_{\text{max}}$ ). The percent fusion parameter was calculated as  $[\Delta F(t)/\Delta F_{\text{max}}] \times 100$ . Assignment of protein as the cause of fusion is supported by negligible fusion after addition of an aliquot of a stock solution without protein.

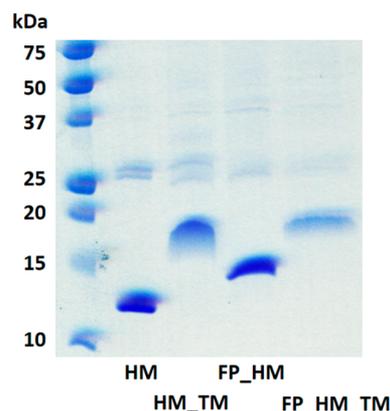
## RESULTS

### Protein Expression, Solubilization, and Purification.

After expression, the first step in RP purification was cell lysis in PBS. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) of the soluble lysate of each of the constructs did not show any obvious band corresponding to RP, which evidenced low RP solubility in PBS, as would be expected for membrane proteins. The pellet size was visibly reduced by resuspension in PBS and centrifugation (twice), which is consistent with removal of non-RP material. Similar RP expression levels were evidenced by similar pellet sizes, and by similar  $^{13}\text{C}$  NMR intensities of pellets obtained from expression in minimal medium with added  $[1-^{13}\text{C}]\text{Gly}$  (Figure S2). We performed the inexpensive  $^{13}\text{C}$  labeling followed by separation of inclusion-body-enriched pellet and  $^{13}\text{C}$  NMR because this approach provides an accurate estimate of RP expression when most RP is in inclusion bodies.<sup>48</sup> Less accurate results are obtained by the alternate method of solubilization of the pellet in SDS followed by SDS–PAGE. The RP band intensity in the gel generally decreases with protein hydro-

phobicity, which could be explained by incomplete solubilization of aggregates by SDS.

Our next task was to find other conditions that completely solubilized the RP-rich pellet. Solubilization was achieved by tip sonication of mixtures that contained PBS (pH 7.4) and either (1) 8 M urea, (2) 6 M GuHCl, (3) 8 M urea and 0.8% Sarkosyl, or (4) 8 M urea, 0.5% SDS, and 0.8% Sarkosyl. Solubilization after tip sonication was assessed both visually and also by pellet size after centrifugation at 45000g for 20 min. Each solution was then subjected to  $\text{Co}^{2+}$  affinity chromatography, followed by SDS–PAGE (Figure S3). The darkest RP bands at the highest purity were obtained for mixture 4 that was then used for all subsequent protein purification (Figure 2).



**Figure 2.** SDS–PAGE of the purified HM (MW = 13.7 kDa), HM\_TM (MW = 16.7 kDa), FP\_HM (MW = 16.5 kDa), and FP\_HM\_TM (MW = 18.9 kDa).

All constructs have a non-native C-terminal  $\text{H}_6$  affinity tag preceded by either a  $\text{G}_6\text{LE}$  or  $\text{G}_8\text{LE}$  spacer, where the spacer is required for binding of RP to the  $\text{Co}^{2+}$  resin. The SE by itself likely still adopts helical hairpin structure in high concentrations of denaturant, so we expect that the spacer affords greater solvent exposure of the  $\text{H}_6$  tag.<sup>29</sup> The longer  $\text{G}_8\text{LE}$  spacer was required for binding of FP\_HM\_TM to the resin.

Our next task was to find conditions without urea and Sarkosyl for which solubility was retained for all purified RPs. Dialysis was performed against a variety of buffers, and soluble RPs were obtained for 10 mM Tris (pH 7.4) and 0.2% SDS. We were also interested in studying the RPs in DPC detergent because it has the phosphatidylcholine headgroup common to a significant fraction of the lipid in membranes of cells infected with HIV.<sup>50</sup> Although the RPs precipitated if dialyzed directly against low- or neutral-pH buffers containing 0.25% DPC, all RPs were soluble if the initial exchange into buffer with 0.25% DPC was performed with RP bound to  $\text{Co}^{2+}$  resin, followed by elution using buffer with DPC and 250 mM imidazole, and then dialysis to remove the imidazole.<sup>41</sup> The final solutions contained 0.25% DPC and either 20 mM sodium acetate buffer (pH 3.2) or 20 mM sodium phosphate buffer (pH 7.4). RP solubility when initial exchange was done with resin-bound RP versus RP precipitation when initial exchange was done by dialysis of a RP solution suggests that RP aggregates can form quickly but are not the lowest-free energy state in 0.25% DPC.

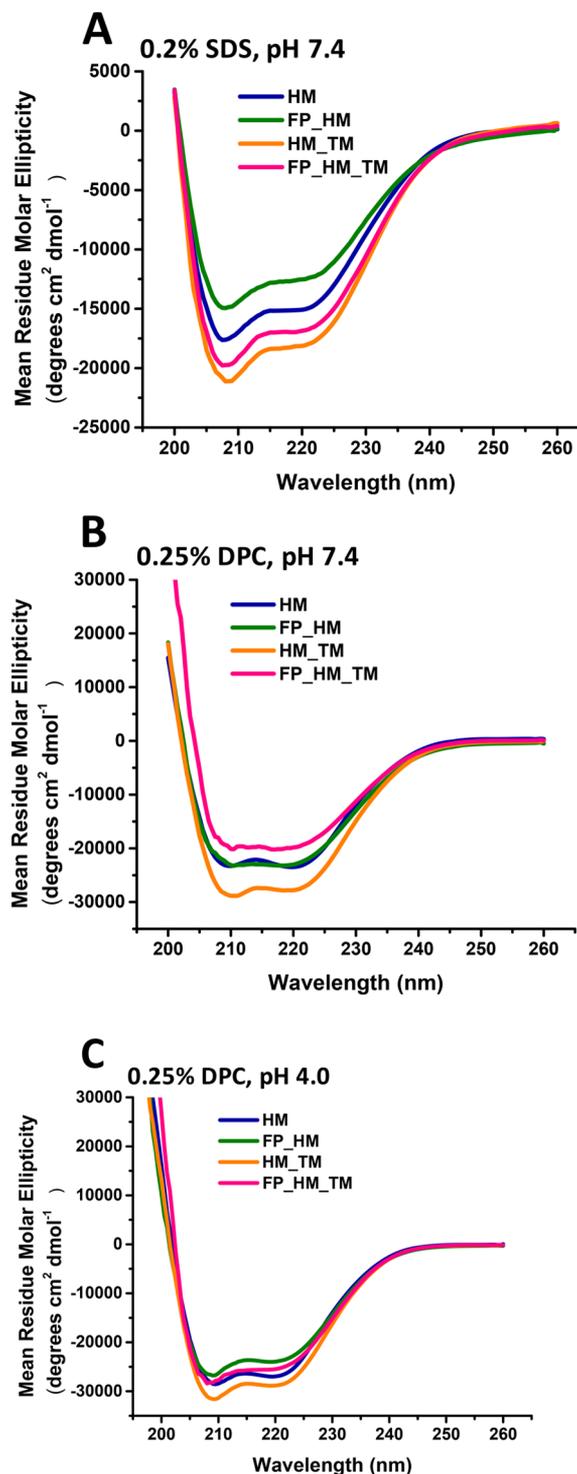
Figure 2 displays SDS–PAGE of the four RPs after dialysis into 10 mM Tris (pH 7.4) with 0.2% SDS. Bands corresponding to FP\_HM, HM\_TM, and FP\_HM\_TM were cut from a gel and subjected to trypsin digestion and mass

spectrometry. Peptides were identified from each of the three bands that provided 80, 55, and 75% sequence coverage, respectively (Figure S4). Anti-H<sub>6</sub> Western blots also supported protein identities (data not shown). The rate of migration of HM-TM in the gel is slower than that of FP-HM even though the two proteins have similar molecular weights (MWs). The rate of migration of a protein in SDS-PAGE is inversely correlated with the mass-bound SDS, and this mass is likely greater for TM than for FP because of the greater hydrophobicity of the TM segment.<sup>53</sup> This interpretation is supported by SEC of the proteins in SDS in Figure 6A and accompanying analysis in Table 2.

Typical purified yields of HM, FP\_HM, HM\_TM, and FP\_HM\_TM were ~10, 0.5, 0.5, and 0.3 mg/L bacterial culture, respectively. The higher yield for HM may be due in part to an ~2-fold higher level of expression (Figure S2) and may also have a contribution from better binding to the Co<sup>2+</sup> resin. Although there were not obvious RP bands in washes of resin prior to elution, binding of resin of FP\_HM\_TM required a glycine spacer longer than those of the other constructs. This suggests occlusion of the H<sub>6</sub> tag by FP and TM segments, which is sterically plausible because of SE helical hairpin folding even with a denaturant at high concentrations.<sup>29</sup> The lower FP\_HM\_TM yield was not due to incomplete elution from the Co<sup>2+</sup> resin. The resin was magenta color before protein binding and changed to pale pink after protein binding. After protein elution, the resin changed back to the original magenta color. The resin was then boiled, and SDS-PAGE of the boiled solution did not show a band at the MW<sub>FP\_HM\_TM</sub>.

**Influence of FP and TM on Hyperthermal  $\alpha$  Helical Hairpin Structure.** Figure 3 displays CD spectra at ambient temperature of the four gp41 constructs in (A) 0.2% SDS (pH 7.4), (B) 0.25% DPC (pH 7.4), and (C) 0.25% DPC (pH 4.0). We tried to obtain the most quantitative comparison between constructs by acquiring all spectra for a single buffer during the same day. All spectra have the characteristic shape of  $\alpha$  helical secondary structure with minima near 208 and 222 nm. This helicity is consistent with the helical hairpin structure of gp41 SE. The  $|\theta|$  values for a single construct in DPC are similar at low and neutral pH, which supports a pH-independent hairpin structure. The average  $\theta_{222\text{-DPC}}/\theta_{222\text{-SDS}} \approx 3/2$ , with variation of this ratio among constructs. This ratio suggests higher helicity in DPC than in SDS. The shapes of the CD curves are typically different in DPC versus SDS, as reflected by a typical  $\theta_{208\text{-DPC}} \approx \theta_{222\text{-DPC}}$  relation, whereas  $\theta_{208\text{-SDS}} \approx 1.2\theta_{222\text{-SDS}}$ .

The HM\_TM construct has the largest  $|\theta|$  values under all three buffer conditions, with a  $|\theta_{222}|$  value of  $\approx 28000$  deg cm<sup>2</sup> dmol<sup>-1</sup> in DPC that correlates with 85% average helicity. This is equal to the average helicity calculated using a model of 100% helicity of 123 of 125 of the native residues and 0% helicity for two native residues as well as 20 non-native residues, i.e., SGGRGG loop and C-terminal G<sub>6</sub>LEH<sub>6</sub> (Table 1). Nearly complete native helicity is consistent with a fully folded protein containing SE helical hairpin and TM helix structural elements. Similar analysis of  $|\theta_{222}|$  of HM yields ~76% experimental helicity and ~10 nonhelical native residues. These residues are most likely at the N- and C-termini of the hairpin, based on reasoning that the TM stabilizes hairpin helical structure. Addition of the FP segment decreases average helicity, and this correlates with a similar loss of helicity deduced from comparative CD spectra of the shorter "HP" (gp41<sub>535-581</sub>/SGGRGG/628-666) and "FP\_HP" (gp41<sub>512-581</sub>/SGGRGG/628-666) constructs.<sup>47</sup> HP and HM share



**Figure 3.** Circular dichroism spectra at ambient temperature of samples containing  $\sim 10$   $\mu$ M protein in different buffer/detergent solutions: (A) 10 mM Tris (pH 7.4) and 0.2% SDS, (B) 20 mM phosphate (pH 7.4) and 0.25% DPC, and (C) 20 mM acetate (pH 4.0) and 0.25% DPC. All spectra for a single buffer/detergent condition were recorded on the same day. The 0.20% SDS is  $\sim 5 \times$  the critical micelle concentration (CMC), and the 0.25% DPC is  $\sim 8 \times$  CMC.

the same N-helix and SGGRGG non-native loop, but HP lacks the 17 native C-terminal residues of HM, as well as the non-native C-terminal G<sub>6</sub>LEH<sub>6</sub>. The average helicity of FP\_HM and FP\_HM\_TM in DPC is  $\sim 73\%$ , which corresponds to

Table 1. Analysis of CD Spectra

construct	average fractional helicity <sup>a</sup>	no. of nonhelical residues <sup>b</sup>
HM	0.76	10
HM_TM	0.85	2
FP_HM	0.73	20
FP_HM_TM	0.73	24

<sup>a</sup>Calculated using 100% helicity as  $|\theta_{222}| = 33000 \text{ deg cm}^2 \text{ dmol}^{-1}$ .<sup>54,55</sup>

<sup>b</sup>Calculated using  $N_{\text{nonhelical}} = N_{\text{tot}} - N_{\text{non-native}} - f_{\text{helix}}/N_{\text{tot}}$ .

nonhelical structure for ~20 (FP\_HM) and ~24 (FP\_HM\_TM) native residues. Most of these residues are likely within the FP, which matches earlier CD data in DPC consistent with ~20-more nonhelical residues for gp41<sub>512-705</sub> versus gp41<sub>538-705</sub>.<sup>34,37</sup> CD spectra in SDS at pH 7.4 were recorded in 5 °C increments between 25 and 90 °C. The temperature series for a single construct were all acquired during the same day. Figure 4 displays representative spectra at 25, 60, and 90 °C, and Figure 5 displays plots of  $\theta_{222}$  versus temperature. The spectra in Figures 4 and 3A were acquired on different instruments, but reproducibility is evidenced by similar shapes of ambient-temperature spectra, and by the same ordering of  $\theta_{222}$  values with the construct. Figure 5 shows only moderate decreases in  $|\theta_{222}|$  with an increase in temperature, which supports  $T_m$  values of  $\geq 90$  °C for the four constructs. Similar moderate CD changes have also been observed for the

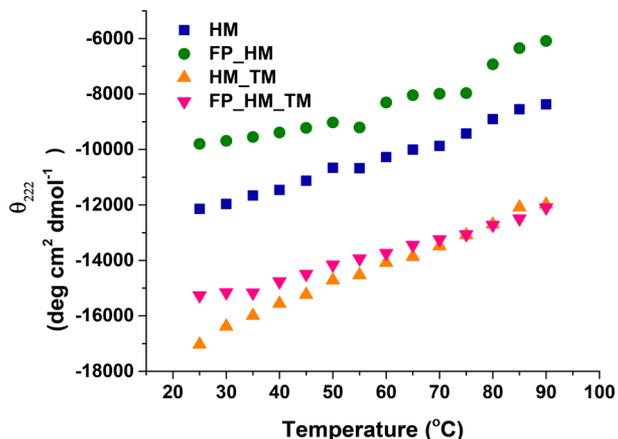


Figure 5.  $\theta_{222}$  vs temperature in 0.2% SDS (pH 7.4). All spectra for a single construct were recorded on the same day.

shorter HP and FP\_HP constructs, and the combined data support hyperthermostable hairpin SE structure for all constructs.<sup>47</sup>

HM and FP\_HM exhibit similar  $\Delta(|\theta_{222}|)$  values of  $\approx 4000 \text{ deg cm}^2 \text{ dmol}^{-1}$  over the 25–90 °C range, which is consistent with assignment of the helical structure to the HM region. In some contrast, HM\_TM exhibits a  $\Delta(|\theta_{222}|)$  of  $\approx 5000$  and

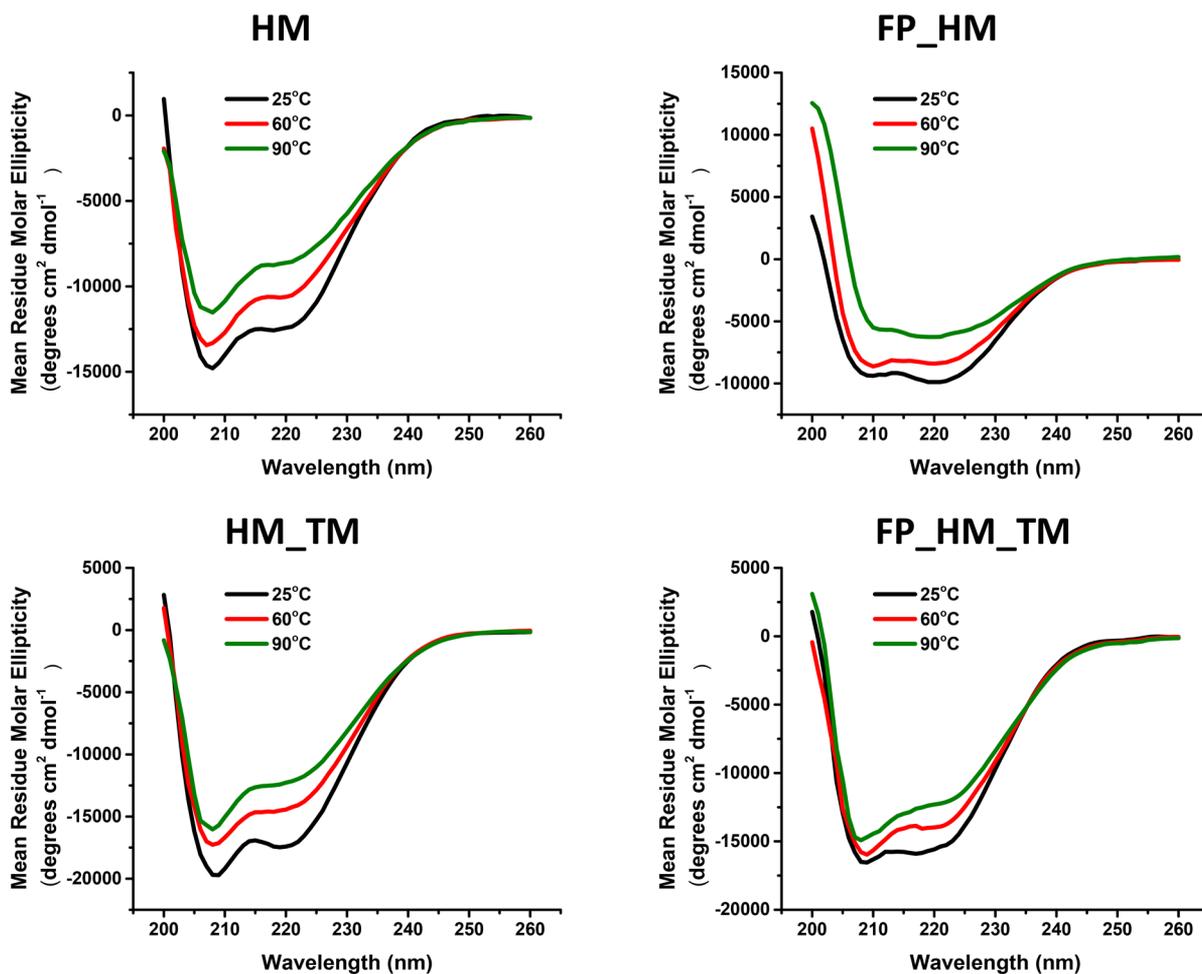


Figure 4. Circular dichroism in 0.2% SDS (pH 7.4) at 25, 60, and 90 °C. Differences between the ambient-temperature spectra of the same construct for these data vs Figure 3A may be partly due to the use of different CD instruments.

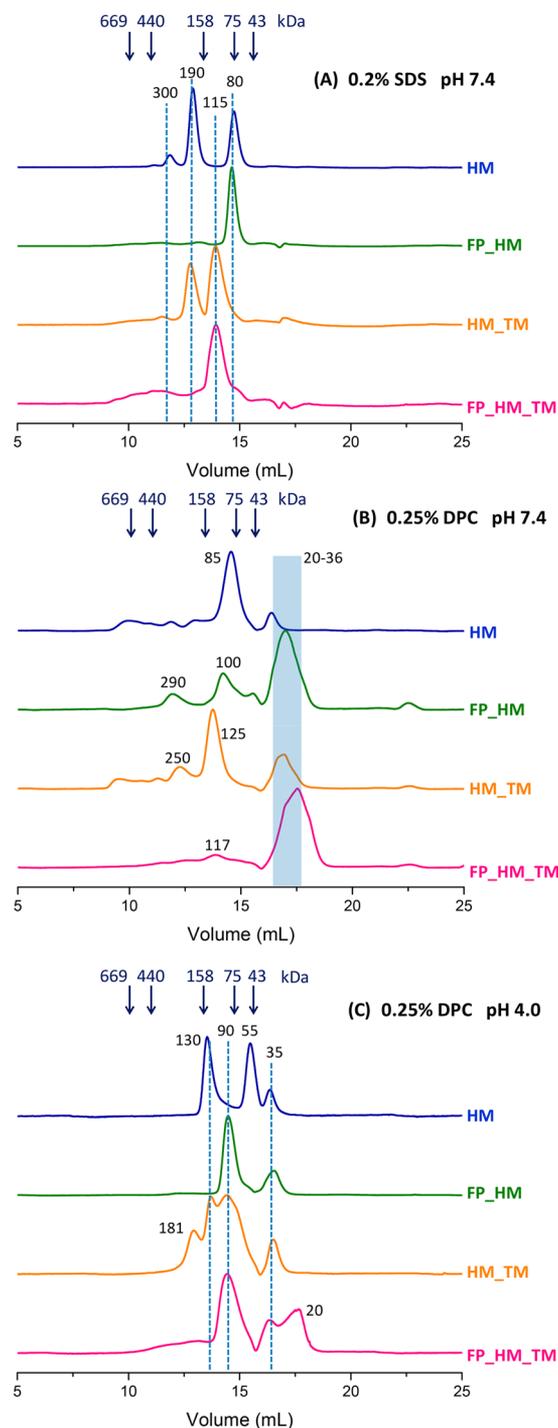
FP\_HM\_TM exhibits a  $\Delta(\theta_{222})$  of  $\approx 3000 \text{ deg cm}^2 \text{ dmol}^{-1}$ , so appending the FP and TM appears to stabilize the helical hairpin structure of HM. This may be due to the location of the FP and TM segments in the same micelle, so opening of the hairpin requires micelle deformation to maintain detergent contact with both FP and TM. There is not this penalty if the construct contains only the FP or the TM.

**SEC Supports Coexisting Protein Monomers, Trimers, Hexamers, and Higher-Order Oligomers.** Figure 6 displays SEC of the gp41 constructs in (A) SDS (pH 7.4), (B) DPC (pH 7.4), or (C) DPC (pH 4.0). These detergents were chosen in part because all proteins are visually soluble in them and are also thermostable, as assessed by their CD spectra. A running protein concentration of  $\approx 5 \mu\text{M}$  is comparable to that in the CD experiments. The 0.20% SDS is  $\sim 5 \times$  critical micelle concentration (CMC), and the 0.25% DPC is  $\sim 8 \times$  CMC. All buffers also contained 150 mM NaCl to inhibit adsorption of protein to the column material. There is  $<10\%$  statistical probability of protein occupation of a micelle, so oligomerization is likely due to protein/protein interaction rather than crowding. Reproducibility is evidenced by similar traces for replicate samples (Figure S5A,B) and for samples with and without a DTT reducing agent in the injection sample and running buffer (Figure S5C).

SEC measures particle migration (elution) volumes, which reflect particle sizes and shapes, and an elution volume is converted to a particle MW using interpolation between the elution volumes of the soluble protein standards. One advantage of SEC is the reproducibility of the elution volume values for a single column type across different chromatography instruments, and therefore different research studies. Table 2 lists our proposed assignments of many peaks in the SEC traces. Most assignments are protein monomer, trimer, or hexamer because these ectodomain species have been commonly observed in earlier studies using SEC, SEC-MALS, and AUC.<sup>29,33–40</sup> The peak SEC elution volumes in Figure 6 reflect protein + detergent masses, so assignment of a peak to a particular protein oligomer correlates to an estimated detergent mass = total mass – protein mass. Assignment of peak to a specific oligomer is also based on having a reasonable value for the detergent mass contribution, which takes into account (1) the typical 20–25 kDa mass of a detergent micelle without protein, (2) an  $\sim 5:1$  SDS:hydrophobic protein segment mass ratio, and (3) a SEC-MALS-based report of  $\sim 17$  kDa DPC per monomer gp41 ectodomain segment.<sup>36,53</sup>

The SEC traces in 0.2% SDS (pH 7.4) exhibit the smallest number and narrowest peaks. A dominant peak is observed for FP\_HM at 80 kDa and for FP\_HM\_TM at 115 kDa. The 80 kDa peak is also observed for HM and the 115 kDa peak for HM\_TM, and they are accompanied by a peak of approximately equal intensity at 190 kDa for both constructs. The reproducibility of elution volumes among different constructs supports separation of discrete species.

The 78 kDa peak is proposed to include HM trimer (41 kDa) and SDS (37 kDa), and the 82 kDa peak includes FP\_HM trimer (49 kDa) and SDS (33 kDa). The 115 kDa peak includes either HM\_TM (50 kDa) or FP\_HM\_TM (57 kDa) trimer, and SDS (60 kDa). Relative to the 80 kDa peaks, the additional  $\sim 25$  kDa SDS mass in the 115 kDa peaks is consistent with solvation of the TM segment by SDS. There is also correlation with the 15% SDS-PAGE gel depicted in Figure 2 that shows an  $\sim 5$  kDa increase in monomer MW with inclusion of TM versus much smaller MW change for inclusion



**Figure 6.** SEC of gp41 constructs under the following conditions: (A) 10 mM Tris (pH 7.4), 150 mM NaCl, and 0.2% SDS at temperature; (B) 20 mM phosphate (pH 7.4), 150 mM NaCl, and 0.25% DPC at 4 °C; and (C) 20 mM acetate (pH 4.0), 150 mM NaCl, and 0.25% DPC at 4 °C. SEC was performed with a Superdex 200-increase column, 1 mg/mL protein loading with  $\sim 10$ -fold dilution in the column, and  $A_{280}$  detection. The arrows in the plots are at the elution volumes of the MW standards, and some of the peaks are identified with dashed lines and with MWs calculated from interpolation between MW standards.

of the FP. The  $\sim 60$  kDa detergent mass for the FP\_HM\_TM trimer matches the detergent mass determined by AUC for the influenza virus HA2 fusion protein trimer, where both proteins contain a FP domain and a TM.<sup>56</sup> Hexamer assignment of the 190 kDa peak correlates with HM (80 kDa) or HM\_TM (100

Table 2. SEC Peak Interpretation<sup>a</sup>

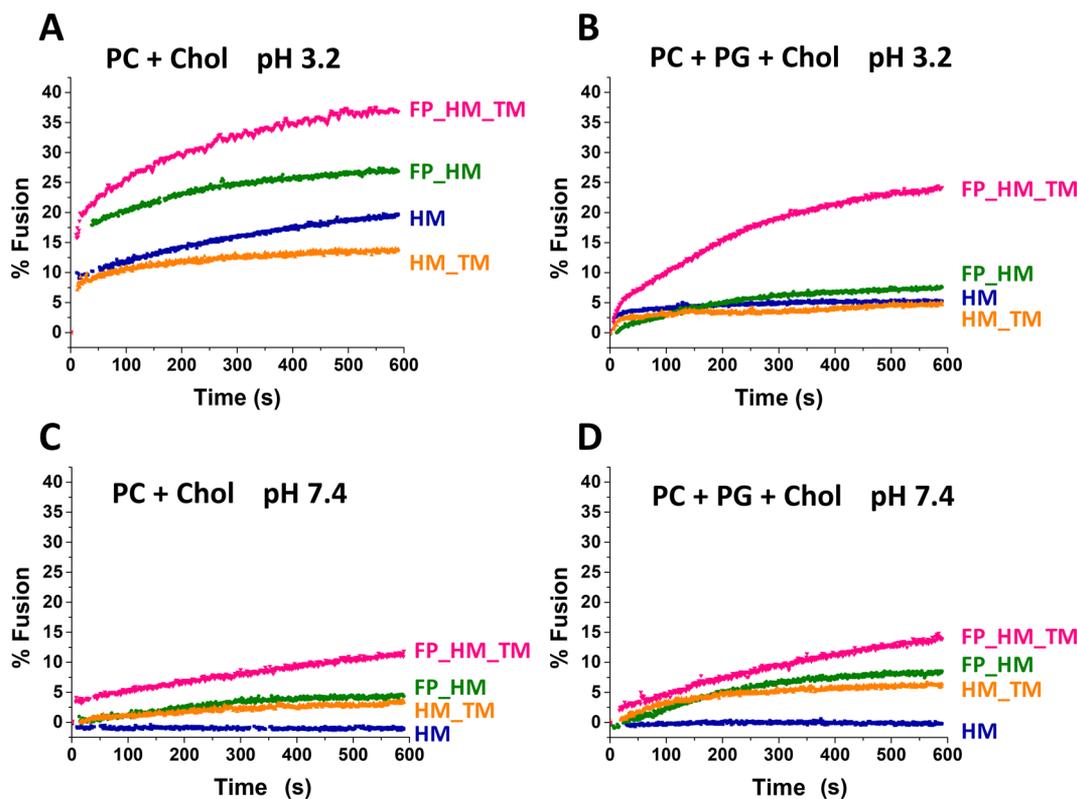
condition	construct(s)	MW <sub>peak</sub>	oligomer	MW <sub>protein</sub>	MW <sub>detergent</sub>
SDS, pH 7.4	HM, FP_HM	80	trimer	45	35
	HM_TM, FP_HM_TM	115	trimer	55	60
DPC, pH 7.4	HM, HM_TM	190	hexamer	100	90
	HM, FP_HM, HM_TM, FP_HM_TM	35	monomer	15	20
DPC, pH 4.0	HM, FP_HM, HM_TM, FP_HM_TM	100	trimer	50	50
	HM, FP_HM, HM_TM, FP_HM_TM	35	monomer	15	20
DPC, pH 4.0	HM	55	trimer	40	15
	FP_HM, HM_TM, FP_HM_TM	90	trimer	50	40
	HM, HM_TM	130	hexamer	100	30
	HM_TM	180	hexamer	100	80

<sup>a</sup>Peaks are grouped by oligomer assignments, and MW<sub>peak</sub> and MW<sub>protein</sub> are typical values for the assignment.

kDa), and SDS (100 kDa), and dodecamer assignment of the 300 kDa peak correlates with HM (160 kDa) and SDS (140 kDa).

SEC traces in DPC (panels B and C of Figure 6 at pH 7.4 and 4.0, respectively) exhibit some peaks with MWs comparable to those in SDS, and assignment is done by similar reasoning. For example, HM at pH 7.4 has a dominant peak at 85 kDa that we assign to HM trimer (41 kDa) + DPC (44 kDa). Our trimer assignment correlates with previous sedimentation velocity AUC analysis of the related gp41<sub>528–579</sub>/SGGRGG/628–683 construct in DPC at pH 7.0.<sup>39</sup> This construct includes the MPER but lacks most of the FP and all the TM. The major species was a trimer (45 kDa) + DPC (23 kDa). Major trimer peaks in other SEC traces include the 125 kDa peak at pH 7.4 of HM\_TM (50 kDa) + DPC (75 kDa) and the 90 kDa peaks of FP\_HM, HM\_TM, or FP\_HM\_TM (50 kDa), and DPC (40 kDa) at pH 4.0. Minor peaks assigned to trimers include the 100 kDa peak of FP\_HM (50 kDa) + DPC (50 kDa) and the 115 kDa peak of FP\_HM\_TM (55 kDa) + DPC (60 kDa).

All DPC traces have peaks at lower MWs that are assigned to monomer species. There is often a defined peak at 35 kDa with calculated monomer protein (15 kDa) + DPC (20 kDa) contributions, and sometimes a broader peak at lower MWs. The monomer peaks are dominant for FP\_HM and FP\_HM\_TM at pH 7.4. Our SEC monomer assignment correlates with earlier sedimentation velocity AUC data in DPC at pH 7 for several shorter constructs like gp41<sub>528–579</sub>/SGGRGG/628–655 and gp41<sub>546–579</sub>/SGGRGG/628–683.<sup>39</sup>

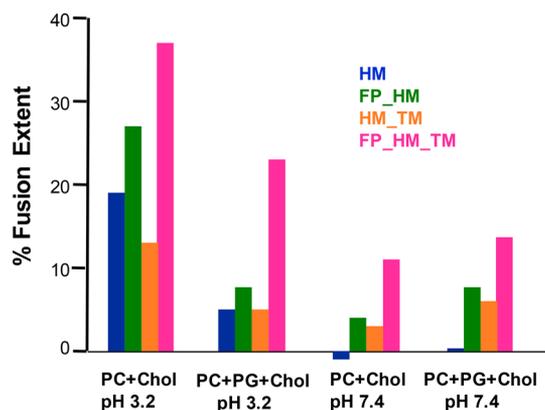


**Figure 7.** Vesicle fusion assays of gp41 proteins. Fusion was initiated by addition of an aliquot of a protein stock solution at 0 s, and subsequent fusion was monitored by increased fluorescence associated with intervesicle lipid mixing. The stock contained 40  $\mu$ M protein in buffer (pH 7.4) with 0.2% SDS, and the protein/vesicle mixture contained 0.5  $\mu$ M protein, POPC, POPG, and Chol at a total concentration of 225  $\mu$ M, and vesicle molar compositions and pH values: (A) 2:1 POPC:Chol at pH 3.2, (B) 8:2:5 POPC:POPG:Chol at pH 3.2, (C) 2:1 POPC:Chol at pH 7.4, and (D) 8:2:5 POPC:POPG:Chol at pH 7.4. Fusion was not observed for any vesicle composition after addition of an aliquot of buffer with 0.2% SDS without protein. All data were obtained on the same day with the same protein stocks. The assay dead time was  $\sim$ 5 s. Small negative values of fusion in a few cases were due to decreased fluorescence associated with the volume increase from addition of the protein aliquot.

Sedimentation coefficients evidenced monomer protein (10 kDa) + DPC (15 kDa).

For constructs in DPC, there are high-MW peaks at pH 7.4 that are weak and broad, while there are defined peaks at pH 4.0, including 55 kDa for HM, 90 kDa for FP\_HM, HM\_TM, and FP\_HM\_TM, 130 kDa for HM and HM\_TM, and 180 kDa for HM\_TM. The 90 kDa peak is assigned to trimer protein (50 kDa) + DPC (40 kDa) and the 180 kDa peak to hexamer protein (100 kDa) and DPC (80 kDa), like similar MW peaks in SDS at pH 7.4. The 55 kDa peak, unique to HM, is tentatively assigned to trimer protein (40 kDa) + nonmicellar DPC (15 kDa). HM is less hydrophobic than the other proteins, and the large  $\sim 30$  positive charge may enable solubility with less detergent. The 130 kDa peak is similarly assigned to hexamer protein (100 kDa) + DPC (30 kDa).

**Hairpin Protein-Induced Vesicle Fusion at Physiologic and Low pH.** Figure 7 shows the time courses of vesicle fusion induced by the four different gp41 constructs for vesicles composed of either POPC, POPG, and Chol (8:2:5) or POPC and Chol (2:1), at pH 3.2 or 7.4. The displayed traces in Figure 7 were recorded on the same day with the same protein stocks. Figure 8 displays long-time ( $\sim 600$  s) fusion extents based on



**Figure 8.** Long-time fusion extents (after  $\sim 600$  s) based on the data depicted in Figure 7 with a 1:450 protein:total lipid mole ratio. Replicate data were acquired on different days with the same protein stocks and vesicles and exhibited  $\pm 2\%$  typical variation in extents among assay replicates. Figure S6 shows extents from experiments using different protein stocks and vesicle preparations, and also comparative fusion extents at protein:lipid ratios of 1:450 and 1:225.

the Figure 7 data. There was  $\pm 2\%$  typical variation in the long-time fusion extent for replicate assays that were performed on different days with the same protein stocks and vesicle preparations.<sup>47</sup> Assays performed with different protein stocks and vesicle preparations exhibited larger systematic variations in absolute fusion extents but retained similar trends in comparative extents with respect to construct, pH, and lipid composition (Figure S6). POPC and Chol are included to represent some of the physicochemical characteristics of the plasma membrane of cells infected by HIV, including PC as a common lipid headgroup, and Chol as  $\sim 0.3$  mole fraction of total lipid.<sup>50</sup> POPG has a calculated charge of  $-1$  at both pH values and is included to represent the  $\sim 0.15$  mole fraction of anionic lipids.<sup>51</sup>

Vesicle fusion was probed by intervesicle lipid mixing detected after addition of  $40 \mu\text{M}$  protein solubilized in  $10 \text{ mM}$  Tris buffer (pH 7.4), with  $0.2\%$  SDS and  $150 \text{ mM}$  NaCl. No fusion was observed for buffer without protein, so fusion

was due to protein/vesicle rather than SDS/vesicle interaction. Protein stocks in SDS at neutral pH were used because the trimer is a dominant oligomeric state (Figure 6A), which matches the likely initial and final gp41 ectodomain states during HIV/cell fusion.<sup>10,12</sup> Vesicle fusion was not done using protein stocks in DPC because of greater variation in oligomer states among constructs, with a large monomer fraction in several cases (Figure 6B,C). Earlier vesicle fusion studies were performed with stock protein with a large monomer fraction.<sup>29,46</sup> It is unlikely that fusion can be examined with a single oligomeric species, based on earlier work in which a second SEC was run on a fraction representing a single oligomeric state of an initial SEC.<sup>29</sup> This second SEC experiment was similar to the original SEC experiment and showed a mixture of oligomeric states that was consistent with thermodynamic equilibrium.

HIV/host cell fusion occurs at pH 7.4, and gp41-induced vesicle fusion was also detected at this pH (Figure 7 C,D, Figures 8, and Figure S6). Most earlier studies were performed with shorter constructs and showed that fusion required low pH and anionic vesicles.<sup>30,46</sup> These previous results were interpreted to support attractive electrostatics as a requirement for protein/vesicle binding. For the study presented here, HM did not induce fusion at neutral pH, which is consistent with these previous studies, but larger constructs induced appreciable fusion.<sup>29</sup> FP\_HM\_TM exhibited the greatest fusion extent, while FP\_HM and HM\_TM had smaller extents that were similar to one another, with that of FP\_HM being consistently higher than that of HM\_TM. These data support a positive correlation between protein hydrophobicity and vesicle fusion. The level of fusion was moderately higher for anionic versus neutral vesicles, which evidence that electrostatic repulsion between anionic vesicles does not affect fusion. These data are also consistent with little bulk electrostatic interaction between a vesicle and the nearly neutral protein, whose calculated charge is approximately  $-1$  at pH 7.4.

The pH 3.2 data provide a comparison with earlier studies of shorter gp41 constructs at low pH. One common feature of all four conditions of Figures 7 and 8 is the highest fusion extent for FP\_HM\_TM, which supports the hydrophobicity/fusion correlation. The fusion extent at low pH was generally comparable or greater than at neutral pH. This correlates with different magnitudes of protein charge at low versus neutral pH, approximately  $+10$  versus approximately  $-1$ . Interestingly, the greater extent at low pH is likely not due to direct protein/vesicle attraction, because the extent was also greater for neutral versus anionic vesicles. This result contrasts with the reverse trend at neutral pH.

## DISCUSSION

This paper presents a systematic structural and functional comparison of protein constructs that contain folded domains of the gp41 protein. We integrate our data with existing data and propose some new interpretations of these data. Key results include (1) helical SE and TM, and nonhelical FP in SDS and DPC detergents; (2) hyperthermostability with a  $T_m$  of  $>90^\circ\text{C}$ ; (3) trimer and sometimes hexamer protein in SDS at pH 7.4, and mixtures of monomer, trimer, and higher-order oligomer protein in DPC at pH 4.0 and 7.4; and (4) substantial protein-induced vesicle fusion, including fusion of neutral and anionic vesicles at neutral pH, which are the conditions similar to those of HIV/cell fusion. Vesicle fusion by a large gp41 construct has rarely been observed under these conditions and is aided by

inclusion of both the FP and TM, and by protein which is predominantly trimer rather than monomer.

**Protein Constructs.** The constructs had a non-native six-residue sequence that replaced the SE loop and adjoining terminal regions of the SE N- and C-helices (Figure 1). This modified SE was chosen because of its better solubility properties while retaining hyperthermostable helical structure with a  $T_m$  of  $>100$  °C (Figures 3–5). All constructs could be dialyzed without precipitation into a SDS solution at pH 7.4, which was supported by SEC traces with small peaks assigned to aggregates, and sharp peaks assigned to either trimer or hexamer protein (Figure 6A and Table 2). Direct dialysis of FP\_HM\_TM into a DPC solution resulted in rapid precipitation. There was not precipitation if the DPC solution was introduced with protein bound to  $\text{Co}^{2+}$  resin, followed by elution. Our observation appears to be consistent with one previous study, but not others, and we do not understand the reason for this apparent discrepancy.<sup>34,41</sup> Relative to those in SDS, the SEC traces in DPC exhibit more peaks, with most intense ones assigned to monomer, trimer, and hexamer protein (Figure 6B,C). Weaker peaks are due to larger species, but not assigned to specific oligomers. The general similarity between the SEC traces of different constructs in SDS (pH 7.4) is consistent with thermodynamic equilibrium, and there is similar consistency between traces of different constructs in DPC at both pH 7.4 and 4.0, which also shows equilibrium. Precipitation upon direct dialysis into DPC thus appears to be due to a rapid rate of aggregation rather than thermodynamic stability of large aggregates.

**Synthesis of Oligomeric and Structural Data.** We and other groups have worked on characterizing the structural and oligomeric states of gp41 constructs, and we integrate our current data with earlier data. Common features of most studies are the very high helicity of the SE and a  $T_m$  of  $>100$  °C, particularly if the construct contains longer N- and C-helix segments with complementarity between their hydrophobic surfaces. These features are exhibited with a native or non-native loop, for solutions without detergent at low pH, and for solutions with detergent at low and neutral pH.<sup>17,18,29,34,41,47,57</sup> The CD data of this work correlate with helical TM and with nonhelical FP, where the latter result is consistent with an earlier CD study, but not consistent with earlier NMR chemical shifts.<sup>34,47</sup> The protein and detergent concentrations are  $\sim 50$ -fold lower in CD than in NMR. In a membrane, the FP often adopts intermolecular antiparallel  $\beta$  sheet structure.<sup>31,45,58</sup> Our CD data at higher temperatures support greater thermostability of FP\_HM\_TM versus that of more truncated constructs. The same thermostability correlation was also observed for the influenza virus HA2 protein, whose sequence is nonhomologous with gp41 but performs a similar fusion function.<sup>49</sup> Close FP/TM contact has not been detected by NMR, so increased thermostability may be due to greater micelle deformation needed to surround both FP and TM for an unfolded SE.<sup>34</sup>

SEC peaks from this study are interpreted to support gp41 with SE, which is either monomer, trimer, hexamer, or larger oligomer. We compare our data to SEC data from earlier studies as well as equilibrium and sedimentation velocity analytical ultracentrifugation (AUC) data. There is variability among SE lengths in these studies, and we initially consider longer constructs for which  $T_m > 100$  °C, which matches the stability of full-length SE.<sup>17,57</sup> The typical total protein concentration was  $\approx 5$   $\mu\text{M}$  for all these SEC and AUC studies,

which makes it more reasonable to compare oligomeric interpretations.

Our SEC data support predominant trimer fraction for all four of our constructs in SDS at neutral pH, with some hexamer fraction for HM and HM\_TM, and no monomer fraction for any construct. There is a significant monomer fraction for all constructs in DPC at both low and neutral pH, as well as trimer and larger oligomer fractions. These data suggest that monomer charges are stabilized by the zwitterionic charges of DPC but not by anionic SDS. We had previously observed a predominant monomer fraction for the SE HP construct (gp41<sub>535–581</sub>/SGGRGG/628–666) at low pH without detergent, whereas large aggregates are predominant at neutral pH.<sup>29,35</sup> These different oligomeric states correlate with different magnitudes of calculated protein charge, approximately +10 at low pH versus approximately –1 at neutral pH, with corresponding large differences in intermonomer electrostatic repulsion. We do not observe a greater monomer fraction in DPC at low versus neutral pH, which supports the significance of electrostatic interactions between the protein and the DPC headgroups with pH-independent zwitterionic charges.

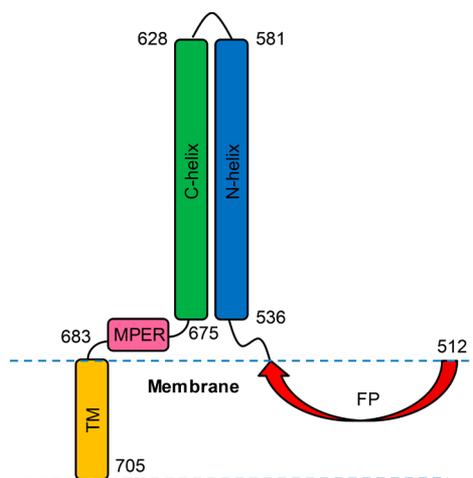
Our SEC in DPC at low pH is generally consistent with an earlier study of the gp41<sub>512–705</sub> construct in DPC at low pH.<sup>34</sup> The previous SEC exhibited a monomer:trimer ratio of  $\approx 3:1$ , which contrasts with sedimentation equilibrium AUC data that were interpreted to support monomer  $\leftrightarrow$  trimer ( $K_a \approx 10^{12} \text{ M}^{-2}$ ) that would correspond to a monomer:trimer ratio of  $\approx 1:4$ . Our SEC of the closely related FP\_HM\_TM (gp41<sub>512–581</sub>/SGGRGG/628–705) in DPC at low pH exhibits a monomer:trimer ratio of  $\approx 1:1$  that is intermediate between the two previous results. For the gp41<sub>538–705</sub> SE+MPER+TM construct, and the gp41<sub>538–665</sub> SE-only construct, sedimentation equilibrium data in DPC at low pH are interpreted to support a  $K_a$  of  $\approx 10^{10} \text{ M}^{-2}$ .<sup>37</sup> Combined consideration of all  $K_a$ 's suggests that FP and TM segments stabilize the trimer versus the monomer, even though they do not directly interact. There is correlation with our observation that FP\_HM\_TM is the most thermostable of the four fully trimeric constructs in SDS at neutral pH. For gp41<sub>538–665</sub> at low pH without detergent, sedimentation equilibrium  $K_a$ 's of  $5 \times 10^{11}$  and  $7 \times 10^{12} \text{ M}^{-2}$  have been independently reported and likely reflect the magnitude of uncertainty for this technique.<sup>33,37</sup> The general consensus from the literature is that monomer is stabilized by low pH and also stabilized by DPC detergent.

Sedimentation velocity AUC has been applied to a construct very similar to FP\_HM\_TM in DPC.<sup>41</sup> Although  $K_a$ 's were not reported from sedimentation velocity, the data exhibited a sedimentation coefficient that increased abruptly at pH 5 and were interpreted to support monomer below pH 5 and trimer above pH 5. We did not observe this behavior in SEC and do not understand the discrepancy between methods. The AUC result correlates with the reduction in calculated protein charge from +9 to +3 between pH 4 and 5.

The constructs of this study were designed based on SE structures to have N- and C-helix segment lengths that yielded a  $T_m$  of  $>100$  °C, like the full-length protein. There are earlier studies of constructs with similar or shorter helix lengths, with some surprising and sometimes conflicting results about monomer versus trimer fractions in DPC. For example, the longer gp41<sub>528–579</sub>/SGGRGG/628–683 construct at 400  $\mu\text{M}$  and low pH exhibits a tumbling time derived from NMR that is consistent with predominant monomer.<sup>38</sup> This corresponds to a  $K_a$  of  $\leq 10^6 \text{ M}^{-2}$  that is 100-fold smaller than the  $K_a$

determined by sedimentation equilibrium AUC for the shorter gp41<sup>538–579/SGGRGG/628–665</sup> construct.<sup>37</sup> At neutral pH, a much larger  $K_a$  of  $\geq 4 \times 10^{10} \text{ M}^{-2}$  has been reported for the longer construct based on sedimentation velocity data, whereas much smaller  $K_a$  values of  $\leq 2 \times 10^8 \text{ M}^{-2}$  were reported for either gp41<sup>546–579/SGGRGG/628–683</sup> or gp41<sup>528–579/SGGRGG/628–655</sup>, each of which is a construct with a shorter N- or C-helix, respectively.<sup>39</sup>

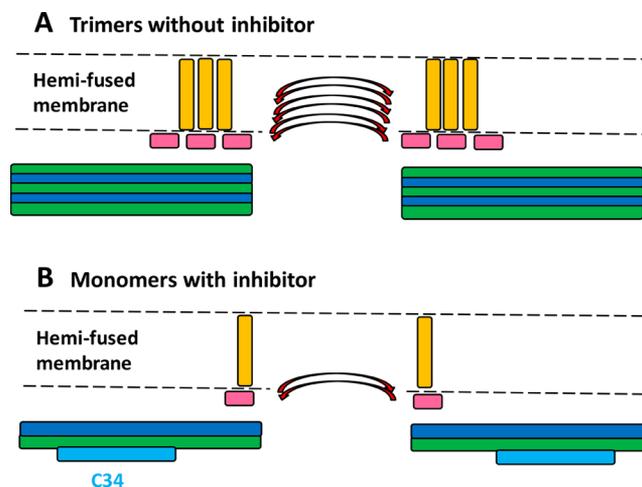
**FP\_HM\_TM Structural Model.** Figure 9 displays a medium-resolution structural model for FP\_HM\_TM that



**Figure 9.** Structural model of FP\_HM\_TM based on circular dichroism spectra of the four constructs, and other data. A monomer is shown for the sake of clarity, but the model should be valid for trimers and hexamers. For a trimer, each interior N-helix of a parallel bundle contacts two exterior C-helices, and each exterior C-helix contacts two interior N-helices. Approximate residue numbers are displayed.

incorporates the Table 1 analysis of the CD spectra as well as previous results. This model likely reflects the final hairpin gp41 state during fusion, based on a  $T_m > 90^\circ \text{C}$  (Figure 5). There are not yet clear experimental data that follow the temporal relationship between gp41 structure and membrane structure during membrane fusion, but efficacies of peptide inhibitors added at different stages of membrane fusion are reasonably interpreted to support the existence of the hairpin state of gp41 at intermediate stages of membrane fusion.<sup>19,29</sup> A monomer is displayed for the sake of clarity, but the model should also be valid for a trimer bundle or a hexamer (dimer of trimer bundles). The HM SE region is primarily hairpin structure that contains helices of residues 536–581 and 628–675. This structure is supported by very high helicity of HM and HM\_TM, and by a previous crystal structure.<sup>13</sup> The C-terminal MPER and TM regions are also highly helical, based on the HM\_TM CD spectrum, and on structures in detergent of peptides corresponding to MPER and/or TM sequences.<sup>27,44</sup> There are likely breaks in helical structure, based on the topological constraints of the probable membrane interface and traversal locations for the MPER and TM, respectively.

The FP region is represented as extended and  $\beta$  strand structure, based both on the reduced average helicity for constructs that include the FP and on earlier NMR and infrared data that provide evidence of FP antiparallel  $\beta$  sheet structure in the membrane.<sup>31,45,58</sup> Such structure is reasonable for a hexamer for which the strands from the two trimers are interleaved (Figure 10A). Quantitative analysis of the <sup>13</sup>C NMR spectrum of gp41<sup>512–581/SGGRGG/628–666</sup> in membrane bilayers



**Figure 10.** Schematic illustrating (A) trimer and (B) monomer favored in the absence and presence of the peptide inhibitor, respectively. Panel B displays the “C34” inhibitor that contains C-helix residues 628–661. Binding of C34 to the N-helix may require dissociation of the C-helix from the N-helix. The sequence color coding matches that of Figures 1 and 9, and loops between structured regions are not displayed for the sake of clarity. The FPs from different trimers or monomers adopt antiparallel  $\beta$  sheet structure, and the trimeric TM bundle is based on a TM peptide structure.<sup>28,58</sup> Fusion is enhanced in panel A vs B because of greater clustering of membrane-perturbing protein regions in the trimer than in the monomer. This enhancement exists for the displayed hemifusion state as well as membrane states that precede hemifusion.

also supports predominant  $\alpha$  helical structure for non-FP regions and  $\beta$  structure for the FP segment.<sup>18</sup> NMR data also support a distribution of N-terminal antiparallel  $\beta$  sheet registries for the FP, with significant populations of registries like  $512 \rightarrow 527/527 \rightarrow 512$  for adjacent strands.<sup>45</sup> FP insertion in a single leaflet is evidenced by NMR contacts between multiple FP residues and lipid tails.<sup>59</sup> There is not close contact between the FP and the MPER or between the FP and the TM, as evidenced by the absence of NMR crosspeaks between these domains.<sup>34</sup>

Similar structures of the monomer and trimer forms of FP\_HM\_TM in Figures 9 and 10 are supported by previous observations that both oligomeric species exhibit similar structural properties that include (1)  $\theta_{222}$  values that correlate with high fractional helicity and (2)  $T_m$  values of  $>90^\circ \text{C}$ .<sup>11,17,29,34,37,40,47</sup> These properties underlie the models in Figures 9 and 10 of the soluble ectodomain regions showing well-defined secondary and tertiary structures that match the high-resolution structure of the trimer.<sup>12–14</sup> One conundrum is understanding how gp41 can be structurally hyperthermostable with a  $T_m$  of  $\approx 110^\circ \text{C}$  and a  $\Delta H_m$  of  $\approx 60 \text{ kcal/mol}$  and also exhibit NMR relaxation data (for the trimeric species) that are consistent with large-amplitude internal motions on the nanosecond to millisecond time scale.<sup>17,18,34,37</sup> These NMR data have been interpreted to support a model of dynamic equilibrium between a compact trimer of hairpins and a trimer of extended N- and C-helices.

**Relationship between Vesicle Fusion and HIV/Cell Fusion.** Mature HIV virions contain a complex of the gp41 ectodomain trimer in noncovalent association with three gp120 subunits.<sup>7,10</sup> It is not known whether gp41 in the initial complex is its lowest-free energy state. However, for the influenza virus hemagglutinin fusion protein, calorimetric data

are consistent with a lowest-free energy state for the initial trimeric HA1/HA2 complex at neutral pH.<sup>60</sup> Binding of the HIV gp120 subunits to cellular receptors leads to separation of gp120 from gp41, and then reformation of the secondary and tertiary structure of the gp41 ectodomain into the final hyperthermostable hairpin state.<sup>11</sup> Virus/cell membrane fusion is catalyzed by the gp41 ectodomain, and the FP likely binds to the target cell membrane early in the fusion process.<sup>3,29,59</sup> There is not yet imaging that probes gp41 structure during fusion, and it therefore is not known which gp41 structure(s) mediates the different steps of membrane fusion. Gp41 is unlikely to catalyze fusion in its initial complex with gp120 because the FP is not positioned to reach the cell membrane and because there is not a clear path for lipids to flow around the large complex. This study describes vesicle fusion induced by gp41 in its final hairpin state in the absence of gp120, which is a candidate structure for catalyzing some steps of HIV/cell fusion. The hairpin role in fusion is supported by (1) monomer hairpins as a potential target of fusion inhibitors (Figure 10B) and (2) the reasonable hypothesis that hairpin formation apposes the viral and cell membranes prior to fusion via attachment of the TM to the viral membrane and attachment of the FP to the cell membrane.<sup>19,29</sup> In addition, many steps of cell/cell fusion are induced by the influenza virus hemagglutinin subunit II (HA2) fusion protein in its final hairpin conformation.<sup>61,62</sup> HA2 is significantly structurally similar to gp41, including its trimeric prefusion state with receptor protein (gp41/gp120 and HA2/HA1 complexes) and final hairpin structure in the absence of receptor protein.<sup>63,64</sup>

The gp41 proteins of this study and many previous studies are recombinantly produced in bacteria and lack the post-translational glycosylations of gp41 produced in human cells. The lack of glycosylation likely does not impact virus/cell fusion, as detailed in two recent studies that examined properties of the five glycosylation sites of the gp41 ectodomain.<sup>65,66</sup> N616 is the only site for which there can be a large reduction in viral infectivity with a loss of glycan, although there are HIV strains for which there is no reduction associated with this loss. Loss of infectivity with removal of N616 glycan is highly correlated with the inability to form a gp41/gp120 complex, rather than with loss of gp41 fusion function. The lack of glycans at other sites resulted in an ~6-fold reduction in the  $IC_{50}$  for viral replication by the T-20 fusion inhibitor, which is consistent with efficient fusion in the absence of these glycans.

The vesicle fusion data of this study provide information about the relative membrane perturbations of different hairpin constructs that were added as trimer, or trimer + hexamer, without a detectable monomer fraction (Figure 6A). Membrane binding will likely be in these initial oligomeric states, which reflect trimer as the initial gp41 ectodomain state in complex with gp120 and probably also the final hairpin state without gp120. Vesicle binding by gp41 trimers contrasts with the results of many previous vesicle fusion studies for which the gp41 ectodomain was likely predominantly monomeric.<sup>29,46</sup>

For this study, stock protein was principally hyperthermostable trimers and the protein:lipid ratio of ~1:450 implies ~60 protein trimers per vesicle when there is quantitative protein binding, and a smaller copy number with weakened binding. This is close to the typical ~15 trimers per virion, with significant microscopy and functional evidence that trimers are spatially clustered during HIV/cell fusion.<sup>67–69</sup>

Earlier studies of vesicle fusion induced by large gp41 ectodomain constructs like HM and FP\_HM often showed strong dependences on pH and membrane charge that were interpreted as supporting a large contribution to fusion from protein/vesicle electrostatic attraction.<sup>17,29,30,46</sup> There was similarly much greater leakage of anionic vesicles at low pH than at neutral pH.<sup>41</sup> For the study presented here, electrostatic effects were much less pronounced, as evidenced by weaker dependences of fusion extents on low versus neutral pH and anionic versus neutral vesicles (Figures 7, 8, and S6). The fusion extent is a little higher at low pH than at neutral pH, and the extent at low pH is a little higher with PC + Chol than with PC + PG + Chol; on the other hand, the extent at neutral pH exhibits the opposite trend with vesicle composition. A higher fusion extent at low pH versus neutral pH correlates with a large difference in magnitude of protein charge, approximately +10 versus approximately –1. A higher extent at low pH with neutral versus anionic vesicles correlates with weakened bulk protein/vesicle electrostatic attraction that may allow for greater protein and lipid conformational flexibilities. There is also correlation to vesicle fusion observed at neutral pH with short FP (gp41<sub>512–534</sub>) or MPER/TM (gp41<sub>671–693</sub>) peptides that have large positive charges from non-native lysines appended to increase their aqueous solubility.<sup>70,71</sup>

For the study presented here, large ectodomain constructs with FP and/or TM segments induced significant fusion at neutral pH of both neutral and anionic vesicles, which reflects expected physiologic conditions of HIV/host cell fusion.<sup>51</sup> Fusion under physiologic conditions for the study presented here but not earlier studies may be due to inclusion of more hydrophobic segments in the study presented here, and also stock solutions with predominant trimer in the study presented here versus monomer in previous studies.<sup>29,46</sup> The contribution of the hydrophobic effect is evidenced by highest fusion extent for FP\_HM\_TM, which contains the highly conserved FP and TM segments.<sup>72,73</sup> The fusion efficiency of trimer versus monomer may be due to the increased local concentration of FP and TM and correlates with a much higher vesicle fusion extent induced by FPs that are cross-linked at their C-termini, with topology similar to that in the hairpin trimer.<sup>70</sup> Membrane perturbation is magnified in the fusion rate by the Arrhenius law, assuming that the perturbations reduce the activation energy by making membranes more like the fusion transition state. Previous studies have evidenced contributions of the FP and TM to fusion that are sequence-specific or reflect overall hydrophobicity. Example sequence-specific effects are large reductions in gp160-mediated cell/cell fusion associated with mutations like V513E (FP) and R696D (TM).<sup>3,73</sup> Hydrophobicity effects are evidenced by comparable extents of vesicle fusion induced by WT- and residue-scrambled FPs, and by efficient cell/cell fusion and HIV infectivity after replacement of the gp41 TM with the TM from another membrane protein.<sup>74,75</sup> There are also non-fusion-related effects of FP and TM mutations that include reductions in the levels of HIV-associated immune suppression and intracellular trafficking of gp160.<sup>76,77</sup>

Efficient vesicle fusion at physiologic pH by the trimer of hairpins may provide insight into how peptides corresponding to N- or C-helix regions inhibit fusion up to the final pore expansion step, which is after intermembrane lipid mixing and pore formation (Figure 10).<sup>19</sup> These peptides would likely not tightly bind to the trimer of hairpins, and we propose that they instead bind to monomer hairpins formed from dissociation of

the trimer. Such dissociation has been known for 20 years, with additional recent data from our group and other groups, and could plausibly occur during the ~30 min HIV/cell fusion time.<sup>2,29,33,36,40,41</sup> The bound peptides would prevent reassociation into the fusion-efficient trimer. For constructs like FP\_HM\_TM in a DPC detergent at low pH, the three monomer  $\leftrightarrow$  trimer  $K_a \approx 10^{12} \text{ M}^{-2}$ , so that  $\text{mass}_{\text{monomer}} \approx \text{mass}_{\text{trimer}}$  when the total protein concentration approximates 1  $\mu\text{M}$ . The membrane  $K_a$  has not been measured, but there are comparable statistical-average interprotein molecular separations of ~100 nm for  $\approx 1 \mu\text{M}$  bulk protein and ~50 nm for ~15 protein trimers in a 100 nm diameter virion. Previous work from our group and other groups supports the monomer as the hyperthermostable helical folding unit, and it is therefore reasonable that the hairpin is the lowest-free energy structure.<sup>18,29,34,40,41,47</sup> Asynchronous folding of monomer protein into hairpin brings the two membranes into apposition and is likely topologically easier than concerted folding of a trimer into a six-helix bundle. It may therefore be evolutionarily advantageous to retain some stability of the monomer hairpin relative to the trimer of hairpins.

$K_d$  values have been measured for several systems containing a C-helix peptide and a trimeric gp41 construct. There have been corollary measurements of  $\text{IC}_{50}$  values of peptide inhibitions of (1) gp160-mediated cell/cell fusion, (2) entry of HIV into cells, and (3) HIV infection of these cells. For example, the  $K_d \approx 1 \mu\text{M}$  for C34 (gp41<sub>628–661</sub>) binding to a shorter “six-helix” gp41 (gp41<sub>546–579/loop/628–655</sub>) ectodomain trimer that has a  $T_m$  of  $\approx 80^\circ \text{C}$  versus  $>100^\circ \text{C}$  for the full ectodomain.<sup>20</sup>  $K_d \approx 1 \mu\text{M}$  for C34 binding to a “five-helix” gp41 ectodomain (gp41<sub>543–582/loop/625–662</sub>) for which one of the C-helices is absent, so that the interior N-helix bundle is exposed for C34 binding.<sup>78,79</sup> The five-helix bundle was developed as a surrogate of a hypothesized “pre-fusion intermediate” in which the gp41 ectodomain is extended with C-helices fully separated from the trimeric bundle of N-helices.<sup>80</sup> These  $K_d$  values are much larger than  $\text{IC}_{50} \approx 10 \text{ nM}$  for C34 inhibition of gp160-mediated cell/cell fusion and  $\text{IC}_{50} \approx 2 \text{ nM}$  for C34 inhibition of HIV entry and infection.<sup>81,82</sup> By contrast, the  $K_d$  for binding to the five-helix form was much smaller than  $\text{IC}_{50}$ 's when two or three amino acids were added to the N-terminus of C34, where  $K_d \approx 1 \text{ pM}$  for binding to the five-helix form (gp41<sub>543–582/loop/625–662</sub>) and  $\text{IC}_{50}$  values for cell/cell fusion, HIV entry, and HIV infection are ~1, 0.5, and 0.5 nM, respectively.<sup>82,83</sup> For comparison, the clinically prescribed T20 (gp41<sub>638–673</sub>) has a  $K_d$  of  $\approx 30 \text{ nM}$  for binding with an extended five-helix form (gp41<sub>530–582/loop/625–669</sub>), and  $\text{IC}_{50}$  values for cell/cell fusion, HIV entry, and HIV infection of ~25, 25, and 5 nM, respectively.<sup>83,84</sup> Given the relative lack of correlation between fusion inhibition and peptide binding to a five- or six-helix gp41 ectodomain, it is plausible that fusion inhibition also has a contribution from peptide binding to monomer hairpins, which we and other groups commonly observe for gp41 ectodomain constructs.

There are similarities between our observations for HIV gp41 and those for nonhomologous fusion proteins from other viruses. For example, trimers of hairpins of the full-length HA2 protein of the influenza virus catalyze vesicle fusion, and HA2 also exhibits a monomer fraction in SEC.<sup>49</sup> There is also evidence for a monomer intermediate during alphavirus fusion.<sup>22</sup> There may be a common mechanism of the monomer folding into a hairpin with subsequent association into trimers that catalyze fusion. This mechanism is supported by hairpin

trimers of the HA2 ectodomain inducing lipid mixing and pore formation between docked cells.<sup>61,62</sup> There are similar effects of point mutations for both HA2 ectodomain-induced cell/cell fusion and virus/cell fusion.

## SUMMARY

This study reports vesicle fusion at physiologic pH by a hyperthermostable HIV gp41 hairpin trimer that includes the FP and TM segments. This final gp41 state may catalyze HIV/cell fusion steps that follow apposition of the membranes, where the latter step is likely concurrent with hairpin formation. In addition, this study and earlier studies report partial dissociation of the hairpin trimer into monomers. The monomers may be evolutionarily advantageous because they aid initial hairpin formation and may also be the target of gp41 N- and C-helix peptide fusion inhibitors.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.7b00753.

DNA sequences corresponding to protein inserts, <sup>13</sup>C NMR of bacterial cell pellets enriched with inclusion bodies, SDS–PAGE of affinity purification eluents for different solubilization conditions of cell pellets, sequence coverages of proteins from peptides created by trypsin digestion and identified by mass spectrometry, original and replicate traces of size-exclusion chromatography, and additional fusion extent data, including dose dependence (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: weliky@chemistry.msu.edu.

### ORCID

D. P. Weliky: 0000-0002-2765-5950

### Funding

Supported by National Institutes of Health Grant AI47153.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors acknowledge assistance from the Michigan State University Genomics and Mass Spectrometry facilities.

## ABBREVIATIONS

AUC, analytical ultracentrifugation; CD, circular dichroism; Chol, cholesterol; DPC, dodecylphosphocholine; DTT, dithiothreitol; FP, fusion peptide; HP, hairpin construct; HM, hairpin + MPER construct; MPER, membrane-proximal external region; POPC, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine; POPG, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (sodium salt); RP, recombinant protein; Sarkosyl, sodium lauroyl sarcosinate; SDS, sodium dodecyl sulfate; SE, soluble ectodomain; SEC, size-exclusion chromatography; TM, transmembrane domain.

## REFERENCES

- (1) Grewe, C., Beck, A., and Gelderblom, H. R. (1990) HIV: early virus-cell interactions. *JAIDS, J. Acquired Immune Defic. Syndr.* 3, 965–974.
- (2) Blumenthal, R., Durell, S., and Viard, M. (2012) HIV entry and envelope glycoprotein-mediated fusion. *J. Biol. Chem.* 287, 40841–40849.
- (3) Freed, E. O., Delwart, E. L., Buchsacher, G. L., Jr., and Panganiban, A. T. (1992) A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc. Natl. Acad. Sci. U. S. A.* 89, 70–74.
- (4) Postler, T. S., and Desrosiers, R. C. (2013) The tale of the long tail: the cytoplasmic domain of HIV-1 gp41. *J. Virol.* 87, 2–15.
- (5) Julien, J.-P., Cupo, A., Sok, D., Stanfield, R. L., Lyumkis, D., Deller, M. C., Klasse, P.-J., Burton, D. R., Sanders, R. W., Moore, J. P., Ward, A. B., and Wilson, I. A. (2013) Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* 342, 1477–1483.
- (6) Lyumkis, D., Julien, J.-P., de Val, N., Cupo, A., Potter, C. S., Klasse, P.-J., Burton, D. R., Sanders, R. W., Moore, J. P., Carragher, B., Wilson, I. A., and Ward, A. B. (2013) Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* 342, 1484–1490.
- (7) Pancera, M., Zhou, T. Q., Druz, A., Georgiev, I. S., Soto, C., Gorman, J., Huang, J. H., Acharya, P., Chuang, G. Y., Ofek, G., Stewart-Jones, G. B. E., Stuckey, J., Bailer, R. T., Joyce, M. G., Louder, M. K., Tumba, N., Yang, Y. P., Zhang, B. S., Cohen, M. S., Haynes, B. F., Mascola, J. R., Morris, L., Munro, J. B., Blanchard, S. C., Mothes, W., Connors, M., and Kwong, P. D. (2014) Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* 514, 455–461.
- (8) Garces, F., Lee, J. H., de Val, N., Torrents de la Pena, A., Kong, L., Puchades, C., Hua, Y., Stanfield, R. L., Burton, D. R., Moore, J. P., Sanders, R. W., Ward, A. B., and Wilson, I. A. (2015) Affinity maturation of a potent family of HIV antibodies is primarily focused on accommodating or avoiding glycans. *Immunity* 43, 1053–1063.
- (9) Lee, J. H., Ozorowski, G., and Ward, A. B. (2016) Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. *Science* 351, 1043–1048.
- (10) Ozorowski, G., Pallesen, J., de Val, N., Lyumkis, D., Cottrell, C. A., Torres, J. L., Copps, J., Stanfield, R. L., Cupo, A., Pugach, P., Moore, J. P., Wilson, I. A., and Ward, A. B. (2017) Open and closed structures reveal allosteric and pliability in the HIV-1 envelope spike. *Nature* 547, 360–363.
- (11) Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M., and Clore, G. M. (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *EMBO J.* 17, 4572–4584.
- (12) Yang, Z. N., Mueser, T. C., Kaufman, J., Stahl, S. J., Wingfield, P. T., and Hyde, C. C. (1999) The crystal structure of the SIV gp41 ectodomain at 1.47 Å resolution. *J. Struct. Biol.* 126, 131–144.
- (13) Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog.* 6, e1000880.
- (14) Shi, W. X., Bohon, J., Han, D. P., Habte, H., Qin, Y. L., Cho, M. W., and Chance, M. R. (2010) Structural characterization of HIV gp41 with the membrane-proximal external region. *J. Biol. Chem.* 285, 24290–24298.
- (15) Kong, R., Xu, K., Zhou, T. Q., Acharya, P., Lemmin, T., Liu, K., Ozorowski, G., Soto, C., Taft, J. D., Bailer, R. T., Cale, E. M., Chen, L., Choi, C. W., Chuang, G. Y., Doria-Rose, N. A., Druz, A., Georgiev, I. S., Gorman, J., Huang, J. H., Joyce, M. G., Louder, M. K., Ma, X. C., McKee, K., O'Dell, S., Pancera, M., Yang, Y. P., Blanchard, S. C., Mothes, W., Burton, D. R., Koff, W. C., Connors, M., Ward, A. B., Kwong, P. D., and Mascola, J. R. (2016) Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody. *Science* 352, 828–833.
- (16) van Gils, M. J., van den Kerkhof, T. L. G. M., Ozorowski, G., Cottrell, C. A., Sok, D., Pauthner, M., Pallesen, J., de Val, N., Yasmee, A., de Taeye, S. W., Schorcht, A., Gumbs, S., Johanna, I., Saye-Francisco, K., Liang, C.-H., Landais, E., Nie, X., Pritchard, L. K., Crispin, M., Kelsoe, G., Wilson, I. A., Schuitemaker, H., Klasse, P. J., Moore, J. P., Burton, D. R., Ward, A. B., and Sanders, R. W. (2016) An HIV-1 antibody from an elite neutralizer implicates the fusion peptide as a site of vulnerability. *Nat. Microbiol.* 2, No. 16199.
- (17) Lev, N., Fridmann-Sirkis, Y., Blank, L., Bitler, A., Epanand, R. F., Epanand, R. M., and Shai, Y. (2009) Conformational stability and membrane interaction of the full-length ectodomain of HIV-1 gp41: Implication for mode of action. *Biochemistry* 48, 3166–3175.
- (18) Sackett, K., Nethercott, M. J., Epanand, R. F., Epanand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative analysis of membrane-associated fusion peptide secondary structure and lipid mixing function of HIV gp41 constructs that model the early pre-hairpin intermediate and final hairpin conformations. *J. Mol. Biol.* 397, 301–315.
- (19) Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation. *Mol. Biol. Cell* 14, 926–938.
- (20) Louis, J. M., Baber, J. L., and Clore, G. M. (2015) The C34 peptide fusion inhibitor binds to the six-helix bundle core domain of HIV-1 gp41 by displacement of the C-terminal helical repeat region. *Biochemistry* 54, 6796–6805.
- (21) Miyachi, K., Kim, Y., Latinovic, O., Morozov, V., and Melikyan, G. B. (2009) HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell* 137, 433–444.
- (22) Kielian, M. (2014) Mechanisms of virus membrane fusion proteins. *Annu. Rev. Virol.* 1, 171–189.
- (23) Herold, N., Anders-Osswein, M., Glass, B., Eckhardt, M., Muller, B., and Krausslich, H. G. (2014) HIV-1 entry in SupT1-R5, CEM-ss, and primary CD4(+) T Cells occurs at the plasma membrane and does not require endocytosis. *J. Virol.* 88, 13956–13970.
- (24) Schaal, H., Klein, M., Gehrmann, P., Adams, O., and Scheid, A. (1995) Requirement of N-terminal amino acid residues of gp41 for human immunodeficiency virus type 1-mediated cell fusion. *J. Virol.* 69, 3308–3314.
- (25) Suarez, T., Gallaher, W. R., Agirre, A., Goni, F. M., and Nieva, J. L. (2000) Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency virus type 1 envelope glycoprotein: putative role during viral fusion. *J. Virol.* 74, 8038–8047.
- (26) Montero, M., van Houten, N. E., Wang, X., and Scott, J. K. (2008) The membrane-proximal external region of the human immunodeficiency virus type 1 envelope: Dominant site of antibody neutralization and target for vaccine design. *Microbiol. Mol. Biol. Rev.* 72, 54–84.
- (27) Apellaniz, B., Rujas, E., Serrano, S., Morante, K., Tsumoto, K., Caaveiro, J. M. M., Angeles Jimenez, M., and Nieva, J. L. (2015) The atomic structure of the HIV-1 gp41 transmembrane domain and its connection to the immunogenic membrane-proximal external region. *J. Biol. Chem.* 290, 12999–13015.
- (28) Dev, J., Park, D., Fu, Q., Chen, J., Ha, H. J., Ghantous, F., Herrmann, T., Chang, W., Liu, Z., Frey, G., Seaman, M. S., Chen, B., and Chou, J. J. (2016) Structural basis for membrane anchoring of HIV-1 envelope spike. *Science* 353, 172–175.
- (29) Banerjee, K., and Weliky, D. P. (2014) Folded monomers and hexamers of the ectodomain of the HIV gp41 membrane fusion protein: Potential roles in fusion and synergy between the fusion peptide, hairpin, and membrane-proximal external region. *Biochemistry* 53, 7184–7198.
- (30) Sackett, K., TerBush, A., and Weliky, D. P. (2011) HIV gp41 six-helix bundle constructs induce rapid vesicle fusion at pH 3.5 and little fusion at pH 7.0: understanding pH dependence of protein aggregation, membrane binding, and electrostatics, and implications for HIV-host cell fusion. *Eur. Biophys. J.* 40, 489–502.
- (31) Pereira, F. B., Goni, F. M., Muga, A., and Nieva, J. L. (1997) Permeabilization and fusion of uncharged lipid vesicles induced by the HIV-1 fusion peptide adopting an extended conformation: dose and sequence effects. *Biophys. J.* 73, 1977–1986.
- (32) Yang, J., Prorok, M., Castellino, F. J., and Weliky, D. P. (2004) Oligomeric  $\beta$ -structure of the membrane-bound HIV-1 fusion peptide formed from soluble monomers. *Biophys. J.* 87, 1951–1963.

- (33) Wingfield, P. T., Stahl, S. J., Kaufman, J., Zlotnick, A., Hyde, C. C., Gronenborn, A. M., and Clore, G. M. (1997) The extracellular domain of immunodeficiency virus gp41 protein: Expression in *Escherichia coli*, purification, and crystallization. *Protein Sci.* 6, 1653–1660.
- (34) Lakomek, N. A., Kaufman, J. D., Stahl, S. J., Louis, J. M., Grishaev, A., Wingfield, P. T., and Bax, A. (2013) Internal dynamics of the homotrimeric HIV-1 viral coat protein gp41 on multiple time scales. *Angew. Chem., Int. Ed.* 52, 3911–3915.
- (35) Caffrey, M., Braddock, D. T., Louis, J. M., Abu-Asab, M. A., Kingma, D., Liotta, L., Tsokos, M., Tresser, N., Pannell, L. K., Watts, N., Steven, A. C., Simon, M. N., Stahl, S. J., Wingfield, P. T., and Clore, G. M. (2000) Biophysical characterization of gp41 aggregates suggests a model for the molecular mechanism of HIV-associated neurological damage and dementia. *J. Biol. Chem.* 275, 19877–19882.
- (36) Roche, J., Louis, J. M., Grishaev, A., Ying, J. F., and Bax, A. (2014) Dissociation of the trimeric gp41 ectodomain at the lipid-water interface suggests an active role in HIV-1 Env-mediated membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* 111, 3425–3430.
- (37) Lakomek, N.-A., Kaufman, J. D., Stahl, S. J., and Wingfield, P. T. (2014) HIV-1 envelope protein gp41: An NMR study of dodecyl phosphocholine embedded gp41 reveals a dynamic prefusion intermediate conformation. *Structure* 22, 1311–1321.
- (38) Roche, J., Louis, J. M., Aniana, A., Ghirlando, R., and Bax, A. (2015) Complete dissociation of the HIV-1 gp41 ectodomain and membrane proximal regions upon phospholipid binding. *J. Biomol. NMR* 61, 235–248.
- (39) Louis, J. M., Baber, J. L., Ghirlando, R., Aniana, A., Bax, A., and Roche, J. (2016) Insights into the conformation of the membrane proximal regions critical to the trimerization of the HIV-1 gp41 ectodomain bound to dodecyl phosphocholine micelles. *PLoS One* 11, e0160597.
- (40) Caffrey, M., Kaufman, J., Stahl, S., Wingfield, P., Gronenborn, A. M., and Clore, G. M. (1999) Monomer-trimer equilibrium of the ectodomain of SIV gp41: Insight into the mechanism of peptide inhibition of HIV infection. *Protein Sci.* 8, 1904–1907.
- (41) Dai, Z., Tao, Y. S., Liu, N. N., Brenowitz, M. D., Girvin, M. E., and Lai, J. R. (2015) Conditional trimerization and lytic activity of HIV-1 gp41 variants containing the membrane-associated segments. *Biochemistry* 54, 1589–1599.
- (42) Jaroniec, C. P., Kaufman, J. D., Stahl, S. J., Viard, M., Blumenthal, R., Wingfield, P. T., and Bax, A. (2005) Structure and dynamics of micelle-associated human immunodeficiency virus gp41 fusion domain. *Biochemistry* 44, 16167–16180.
- (43) Gabrys, C. M., and Weliky, D. P. (2007) Chemical shift assignment and structural plasticity of a HIV fusion peptide derivative in dodecylphosphocholine micelles. *Biochim. Biophys. Acta, Biomembr.* 1768, 3225–3234.
- (44) Sun, Z. Y. J., Oh, K. J., Kim, M. Y., Yu, J., Brusica, V., Song, L. K., Qiao, Z. S., Wang, J. H., Wagner, G., and Reinherz, E. L. (2008) HIV-1 broadly neutralizing antibody extracts its epitope from a kinked gp41 ectodomain region on the viral membrane. *Immunity* 28, 52–63.
- (45) Schmick, S. D., and Weliky, D. P. (2010) Major antiparallel and minor parallel beta sheet populations detected in the membrane-associated Human Immunodeficiency Virus fusion peptide. *Biochemistry* 49, 10623–10635.
- (46) Ratnayake, P. U., Sackett, K., Nethercott, M. J., and Weliky, D. P. (2015) pH-dependent vesicle fusion induced by the ectodomain of the human immunodeficiency virus membrane fusion protein gp41: Two kinetically distinct processes and fully-membrane-associated gp41 with predominant beta sheet fusion peptide conformation. *Biochim. Biophys. Acta, Biomembr.* 1848, 289–298.
- (47) Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin folding of HIV gp41 abrogates lipid mixing function at physiologic pH and inhibits lipid mixing by exposed gp41 constructs. *Biochemistry* 48, 2714–2722.
- (48) Vogel, E. P., and Weliky, D. P. (2013) Quantitation of recombinant protein in whole cells and cell extracts via solid-state NMR spectroscopy. *Biochemistry* 52, 4285–4287.
- (49) Ratnayake, P. U., Prabodha Ekanayaka, E. A., Komanduru, S. S., and Weliky, D. P. (2016) Full-length trimeric influenza virus hemagglutinin II membrane fusion protein and shorter constructs lacking the fusion peptide or transmembrane domain: Hyperthermostability of the full-length protein and the soluble ectodomain and fusion peptide make significant contributions to fusion of membrane vesicles. *Protein Expression Purif.* 117, 6–16.
- (50) Lorizate, M., Sachsenheimer, T., Glass, B., Habermann, A., Gerl, M. J., Krausslich, H. G., and Brugger, B. (2013) Comparative lipidomics analysis of HIV-1 particles and their producer cell membrane in different cell lines. *Cell. Microbiol.* 15, 292–304.
- (51) Zaitseva, E., Zaitsev, E., Melikov, K., Arakelyan, A., Marin, M., Villasmil, R., Margolis, L. B., Melikyan, G. B., and Chernomordik, L. V. (2017) Fusion stage of HIV-1 entry depends on virus-induced cell surface exposure of phosphatidyserine. *Cell Host Microbiol.* 22, 99–110.
- (52) Yang, R. (2005) Synthesis, enhanced fusogenicity, and solid state NMR measurements of oligomeric HIV-1 fusion peptide constructs. Ph.D. Thesis, Michigan State University, East Lansing, MI.
- (53) Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., and Deber, C. M. (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* 106, 1760–1765.
- (54) Chen, Y. H., Yang, J. T., and Chau, K. H. (1974) Determination of helix and beta-form of proteins in aqueous-solution by circular-dichroism. *Biochemistry* 13, 3350–3359.
- (55) Lu, M., Blacklow, S. C., and Kim, P. S. (1995) A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat. Struct. Mol. Biol.* 2, 1075–1082.
- (56) Swalley, S. E., Baker, B. M., Calder, L. J., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (2004) Full-length influenza hemagglutinin HA<sub>2</sub> refolds into the trimeric low-pH-induced conformation. *Biochemistry* 43, 5902–5911.
- (57) Vogel, E. P., Curtis-Fisk, J., Young, K. M., and Weliky, D. P. (2011) Solid-state nuclear magnetic resonance (NMR) spectroscopy of human immunodeficiency virus gp41 protein that includes the fusion peptide: NMR detection of recombinant Fgp41 in inclusion bodies in whole bacterial cells and structural characterization of purified and membrane-associated Fgp41. *Biochemistry* 50, 10013–10026.
- (58) Sackett, K., Nethercott, M. J., Zheng, Z. X., and Weliky, D. P. (2014) Solid-state NMR spectroscopy of the HIV gp41 membrane fusion protein supports intermolecular antiparallel beta sheet fusion peptide structure in the final six-helix bundle state. *J. Mol. Biol.* 426, 1077–1094.
- (59) Jia, L. H., Liang, S., Sackett, K., Xie, L., Ghosh, U., and Weliky, D. P. (2015) REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins. *J. Magn. Reson.* 253, 154–165.
- (60) Eband, R. F., and Eband, R. M. (2003) Irreversible unfolding of the neutral pH form of influenza hemagglutinin demonstrates that it is not in a metastable state. *Biochemistry* 42, 5052–5057.
- (61) Leikina, E., LeDuc, D. L., Macosko, J. C., Eband, R., Eband, R., Shin, Y. K., and Chernomordik, L. V. (2001) The 1–127 HA<sub>2</sub> construct of influenza virus hemagglutinin induces cell-cell hemifusion. *Biochemistry* 40, 8378–8386.
- (62) Kim, C. S., Eband, R. F., Leikina, E., Eband, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA<sub>2</sub> subunit of Influenza Hemagglutinin can by itself drive low pH-dependent fusion. *J. Biol. Chem.* 286, 13226–13234.
- (63) 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.
- (64) Chen, J., Skehel, J. J., and Wiley, D. C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA<sub>2</sub> subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc. Natl. Acad. Sci. U. S. A.* 96, 8967–8972.
- (65) Wang, W., Nie, J., Prochnow, C., Truong, C., Jia, Z., Wang, S., Chen, X. S., and Wang, Y. (2013) A systematic study of the N-

glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology* 10, 14.

(66) Mathys, L., and Balzarini, J. (2014) The role of N-glycans of HIV-1 gp41 in virus infectivity and susceptibility to the suppressive effects of carbohydrate-binding agents. *Retrovirology* 11, 107.

(67) Zhu, P., Liu, J., Bess, J., Chertova, E., Lifson, J. D., Grise, H., Ofek, G. A., Taylor, K. A., and Roux, K. H. (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* 441, 847–852.

(68) Sougrat, R., Bartesaghi, A., Lifson, J. D., Bennett, A. E., Bess, J. W., Zabransky, D. J., and Subramaniam, S. (2007) Electron tomography of the contact between T cells and SIV/HIV-1: Implications for viral entry. *PLoS Pathog.* 3, e63.

(69) Chojnacki, J., Staudt, T., Glass, B., Bingen, P., Engelhardt, J., Anders, M., Schneider, J., Muller, B., Hell, S. W., and Krausslich, H. G. (2012) Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science* 338, 524–528.

(70) Qiang, W., and Weliky, D. P. (2009) HIV fusion peptide and its cross-linked oligomers: efficient syntheses, significance of the trimer in fusion activity, correlation of  $\beta$  strand conformation with membrane cholesterol, and proximity to lipid headgroups. *Biochemistry* 48, 289–301.

(71) Apellaniz, B., and Nieva, J. L. (2015) Fusion-competent state induced by a C-terminal HIV-1 fusion peptide in cholesterol-rich membranes. *Biochim. Biophys. Acta, Biomembr.* 1848, 1014–1022.

(72) Durell, S. R., Martin, I., Ruyschaert, J. M., Shai, Y., and Blumenthal, R. (1997) What studies of fusion peptides tell us about viral envelope glycoprotein-mediated membrane fusion. *Mol. Membr. Biol.* 14, 97–112.

(73) Long, Y., Meng, F., Kondo, N., Iwamoto, A., and Matsuda, Z. (2011) Conserved arginine residue in the membrane-spanning domain of HIV-1 gp41 is required for efficient membrane fusion. *Protein Cell* 2, 369–376.

(74) Wilk, T., Pfeiffer, T., Bukovsky, A., Moldenhauer, G., and Bosch, V. (1996) Glycoprotein incorporation and HIV-1 infectivity despite exchange of the gp160 membrane-spanning domain. *Virology* 218, 269–274.

(75) Torres, O., and Bong, D. (2011) Determinants of membrane activity from mutational analysis of the HIV fusion peptide. *Biochemistry* 50, 5195–5207.

(76) Miyauchi, K., Curran, A. R., Long, Y. F., Kondo, N., Iwamoto, A., Engelman, D. M., and Matsuda, Z. (2010) The membrane-spanning domain of gp41 plays a critical role in intracellular trafficking of the HIV envelope protein. *Retrovirology* 7, 95.

(77) Faingold, O., Cohen, T., and Shai, Y. (2012) A GxxxG-like motif within HIV-1 fusion peptide is critical to its immunosuppressant activity, structure, and interaction with the transmembrane domain of the T-cell receptor. *J. Biol. Chem.* 287, 33503–33511.

(78) Root, M. J., Kay, M. S., and Kim, P. S. (2001) Protein design of an HIV-1 entry inhibitor. *Science* 291, 884–888.

(79) Deng, Y., Zheng, Q., Ketas, T. J., Moore, J. P., and Lu, M. (2007) Protein design of a bacterially expressed HIV-1 gp41 fusion inhibitor. *Biochemistry* 46, 4360–4369.

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

(81) Gustchina, E., Hummer, G., Bewley, C. A., and Clore, G. M. (2005) Differential inhibition of HIV-1 and SIV envelope-mediated cell fusion by C34 peptides derived from the C-terminal heptad repeat of gp41 from diverse strains of HIV-1, HIV-2, and SIV. *J. Med. Chem.* 48, 3036–3044.

(82) Chong, H., Yao, X., Sun, J., Qiu, Z., Zhang, M., Waltersperger, S., Wang, M., Cui, S., and He, Y. (2012) The M-T hook structure is critical for design of HIV-1 fusion inhibitors. *J. Biol. Chem.* 287, 34558–34568.

(83) Champagne, K., Shishido, A., and Root, M. J. (2009) Interactions of HIV-1 inhibitory peptide T20 with the gp41 N-HR coiled coil. *J. Biol. Chem.* 284, 3619–3627.

(84) Ding, X., Zhang, X., Chong, H., Zhu, Y., Wei, H., Wu, X., He, J., Wang, X., and He, Y. (2017) Enfuvirtide (T20)-based lipopeptide is a potent HIV-1 cell fusion inhibitor: Implications for viral entry and inhibition. *J. Virol.* 91, e00831-17.

## Supplementary Information for Publication

Efficient Fusion at Neutral pH by Human Immunodeficiency Virus gp41 Trimers containing the Fusion Peptide and Transmembrane Domains

S. Liang, P. U. Ratnayake, C. Keinath, L. Jia, R. Wolfe, A. Ranaweera, and D. P. Weliky\*

Department of Chemistry, Michigan State University, East Lansing, MI 48824 USA

**Figure S1.** DNA sequences of gp41 inserts. Each line is 75 nucleotides.

**HM**

TGTACGCTGACGGTCCAAGCACGTCAGCTGCTGAGCGGCATTGTGCAGCAACAGAACAATCTGCTGCGCGCGATC  
GAAGCCCAACAGCATCTGCTGCAGCTGACCGTTTGGGGTATTAAACAACCTGCAGGCTCGTATCCTGAGCGGCGGT  
CGCGGCGGTTGGATGGAATGGGATCGTGAAATTAACAATTATACGAGCCTGATTCACTCTCTGATCGAAGAAAGT  
CAAAACCAACAGGAGAAAAACGAACAGGAACCTGCTGGAACCTGGACAAATGGGCCTCCCTGTGGAACCTGGTTTAAC  
ATTACGAACTGGCTGTGGTACATCAAAGGCGGCGGTGGCGGTGGT

**HM\_TM**

TGTACCCTGACGGTCCAAGCTCGCCAACCTGCTGAGTGGTATCGTTCAACAACAAAACAATCTGCTGCGTGCTATC  
GAAGCCCAACAACATCTGCTGCAGCTGACCGTGTGGGGCATTAAACAGCTGCAGGCCCGTATCCTGAGCGGCGGT  
CGCGGCGGTTGGATGGAATGGGATCGTGAAATTAACAATTATACGAGCCTGATTCACTCTCTGATCGAAGAAAGT  
CAGAATCAGCAAGAGAAAAACGAACAAGAACTGCTGGAACCTGGACAAATGGGCCTCCCTGTGGAACCTGGTTTAAT  
ATTACCAACTGGCTGTGGTACATCAAACCTGTTTATTATGATCGTTGGCGGTCTGGTCGGTCTGCGTATTGTGTTT  
GCTGTCCTGAGTATTGTTCGGTGGCGGTGGCGGCGGT

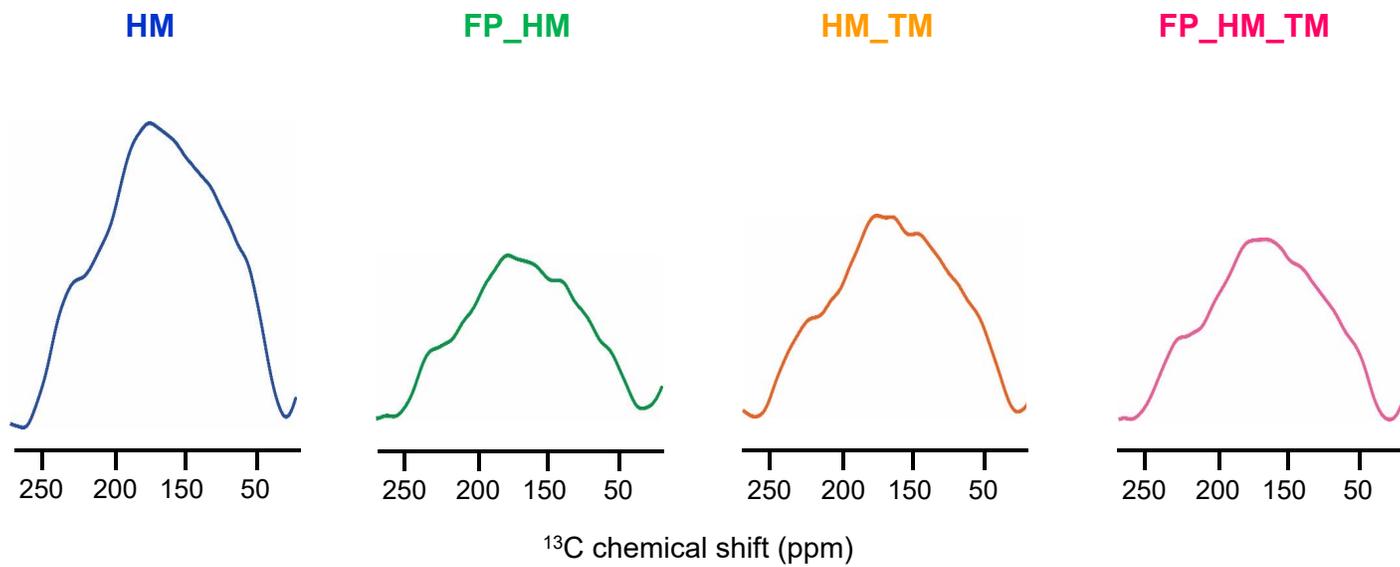
**FP\_HM**

GCCGTGGGTATCGGTGCTCTGTTTCTGGGTTTCTGGGTGCTGCTGGTTCGACGATGGGTGCCCGCTCAATGACG  
CTGACGGTCCAAGCACGTCAGCTGCTGAGCGGCATTGTGCAGCAACAGAACAATCTGCTGCGCGCGATCGAAGCC  
CAACAGCATCTGCTGCAGCTGACCGTTTGGGGTATTAAACAACCTGCAGGCTCGTATCCTGAGCGGCGGTTCGCGGC  
GGTTGGATGGAATGGGATCGTGAAATTAACAATTATACGAGCCTGATTCACTCTCTGATCGAAGAAAGTCAAAAC  
CAACAGGAGAAAAACGAACAGGAACCTGCTGGAACCTGGACAAATGGGCCTCCCTGTGGAACCTGGTTTAACATTACG  
AACTGGCTGTGGTACATCAAAGGCGGCGGTGGCGGTGGT

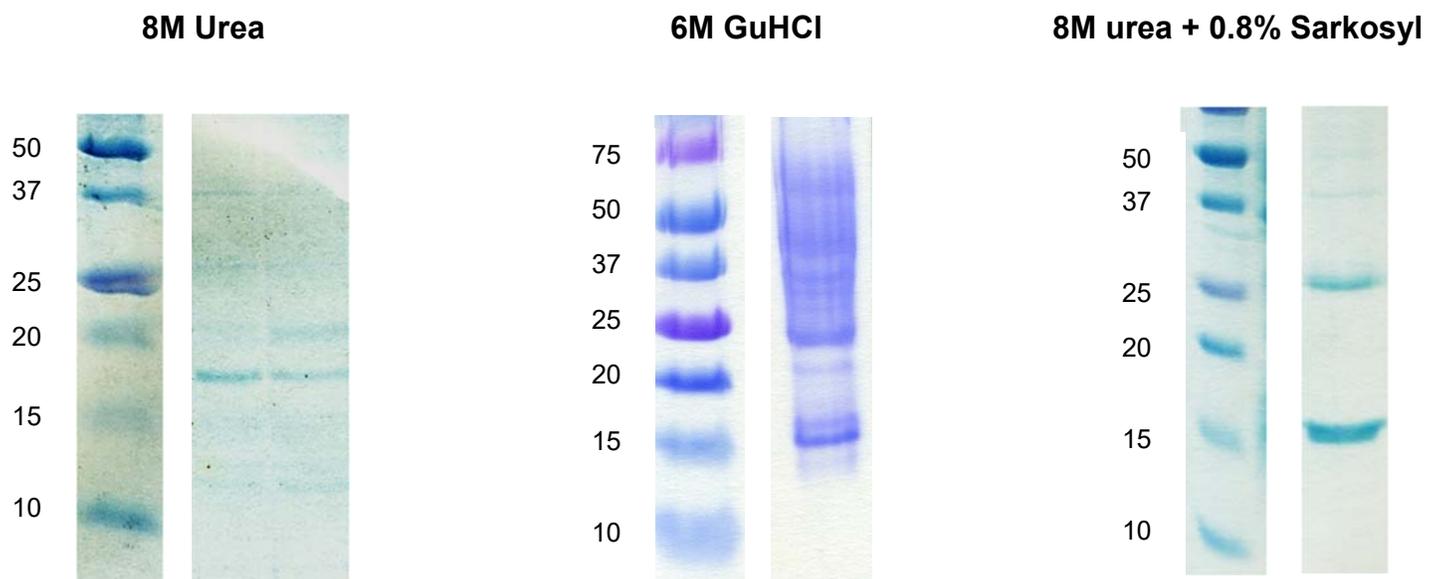
**FP\_HM\_TM**

GCGGTTGGCATTGGTTCGCTGTTTCTGGGCTTTCTGGGTGCGGCGGGTAGCACTATGGGTGCGCGTAGCATGACC  
CTGACCGTTCAAGCGCGTCAACTGCTGAGCGGTATCGTGCAGCAACAGAACAACCTGCTGCGTGCGATTGAGGCG  
CAACAGCACCTGCTGCAGCTGACCGTTTGGGGCATCAAGCAACTGCAGGCGCGTATTCTGAGCGGTGGCCGTGGT  
GGCTGGATGGAGTGGACCGTGAAATCAACAACCTACACCAGCCTGATCCACAGCCTGATTGAGGAAAGCCAAAAC  
CAACAGGAGAAGAACGAGCAGGAACCTGCTGGAACCTGGATAAATGGGCGAGCCTGTGGAACCTGGTTCAACATCACC  
AACTGGCTGTGGTACATCAAACCTGTTTATTATGATCGTTGGGTGGCCTGGTTGGTCTGCGTATCGTGTTCGCGGT  
CTGAGCATTGTTGGCGGTGGTGGTGGCGGTGGTGGT

**Figure S2.** Static  $^{13}\text{C}$  NMR spectra of cellular pellets enriched in inclusion bodies labeled with  $1\text{-}^{13}\text{C}$  Gly. Each spectrum is the sum of 11109 scans.



**Figure S3.** SDS-PAGE after  $\text{Co}^{2+}$ -affinity chromatography of the solubilized pellet enriched in inclusion body protein. The protein is FP\_HM and different solubilization conditions for the pellet are noted. Only the MW marker lane and relevant elution lane(s) are displayed.



**Figure S4.** Sequence coverage of peptides after trypsin digestion. The highlighted yellow residues were in a peptide. The highlighted green Met residues had masses consistent with oxidation. The highlighted green Asn and Gln residues had masses consistent with deamination.

### FP\_HM

FPHM (100%), 16,503.9 Da  
sequence | Weliky FPHM

3 exclusive unique peptides, 8 exclusive unique spectra, 52 total spectra, 117/146 amino acids (80% coverage)

AVGIGALFLG	FLGAAGSTMG	ARSMTLTVQA	RQLLSGIVQQ	QNNLLRAIEA	QQHLLQLTVW
GIKQLQARIL	SGGRGGWMEW	DREINNYTSL	IHSLIEESQN	QKEKNEQELL	ELDKWASLWN
WFNITNWLWY	IKGGGGGLE	HHHHH			

### HM\_TM

HMTM (100%), 16,681.3 Da  
sequence | Weliky HMTM

1 exclusive unique peptides, 1 exclusive unique spectra, 10 total spectra, 80/145 amino acids (55% coverage)

CTLTVQARQL	LSGIVQQQNN	LLRAIEAQQH	LLQLTVWGIK	QLQARILSGG	RGGWMEWDRE
INNYTSLIHS	LIEESQNQQE	KNEQELLELD	KWASLWNWFN	ITNWLWYIKL	FIMIVGGLVG
LRIVFAVLSI	VGGGGGLEH	HHHHH			

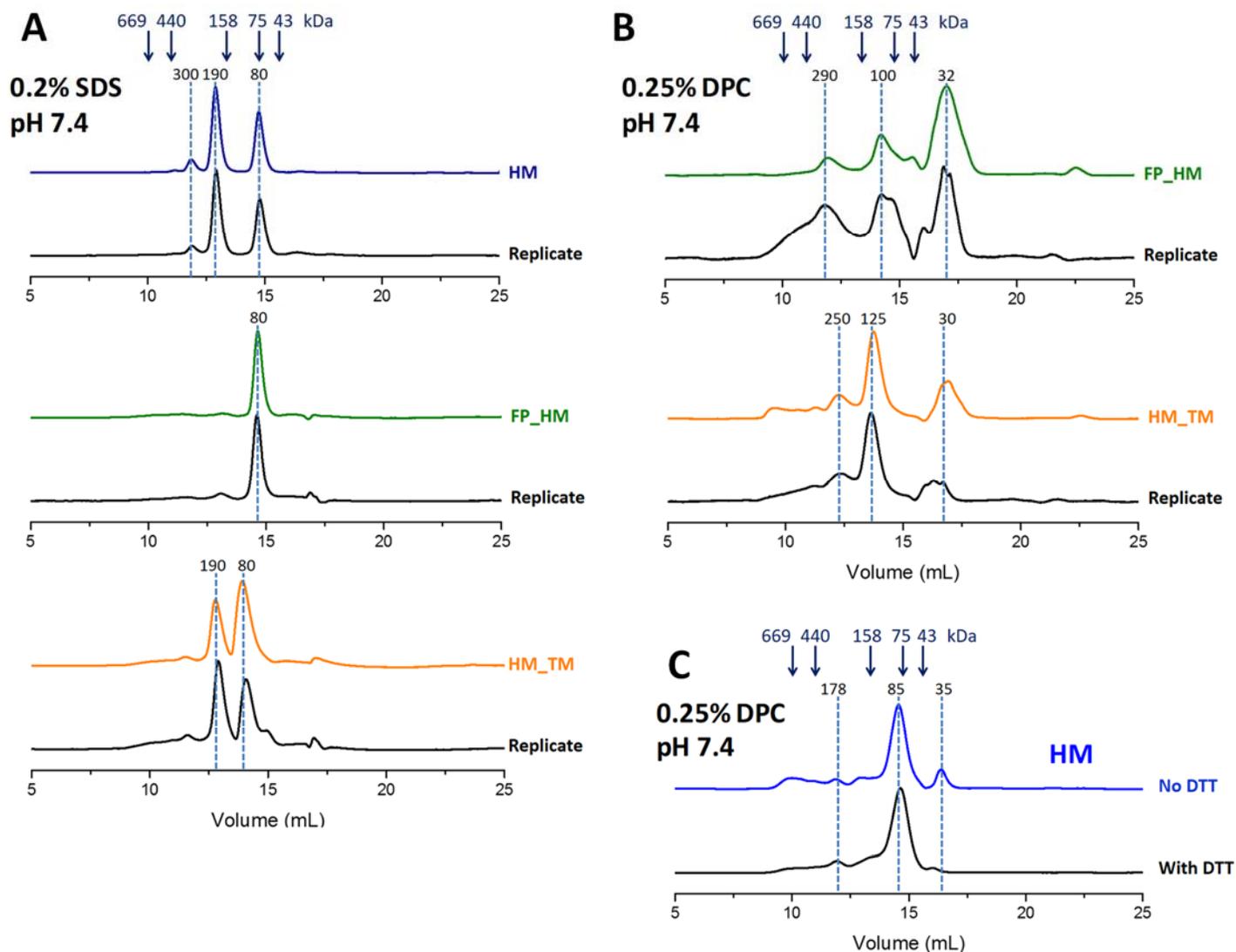
### FP\_HM\_TM

FPHMTM (100%), 18,930.1 Da  
sequence | Weliky FPHMTM

8 exclusive unique peptides, 14 exclusive unique spectra, 38 total spectra, 128/170 amino acids (75% coverage)

AVGIGALFLG	FLGAAGSTMG	ARSMTLTVQA	RQLLSGIVQQ	QNNLLRAIEA	QQHLLQLTVW
GIKQLQARIL	SGGRGGWMEW	DREINNYTSL	IHSLIEESQN	QKEKNEQELL	ELDKWASLWN
WFNITNWLWY	IKLIFIMIVGG	LVGLRIVFAV	LSIVGGGGGG	GGLEHHHHHH	

**Figure S5.** SEC traces of replicate samples in: (A) 10 mM Tris buffer at pH 7.4 with 0.2% SDS and 150 mM NaCl; and (B, C) 20 mM phosphate buffer at pH 7.4 with 0.25% DPC and 150 mM NaCl. The (C) replicates differ in the presence vs. absence of 2 mM DTT reducing agent. The replicates evidence the reproducibility of SEC data, and little dependence on reducing agent.



**Figure S6.** Comparison of vesicle fusion extents at ~600 s time. Panel A is the same as Figure 8 in the main manuscript. Panel B data were acquired about a year before the panel A data, which was prior to our successful purification of FP\_HM\_TM. The panel A data were acquired on the same day using the same protein stocks and the same vesicle preparations. The panel B data were acquired over two days using the same protein stocks and the same vesicle preparations. The panel A protein stocks and vesicles were different from the panel B protein stocks and vesicles. Replicates for panel A and for panel B assays showed typical variation of  $\pm 2\%$  in fusion extent. Panels A and B experiments were done using [protein] = 0.5  $\mu\text{M}$  and [total lipid] = 225  $\mu\text{M}$ . Panel C displays fusion extents for [protein] = 0.5 and 1  $\mu\text{M}$ , and [total lipid] = 225  $\mu\text{M}$  using the panel B protein stocks and vesicles.

