Crystal structure of HERV-K envelope glycoprotein surface subunit

Nikos Nikolopoulos[a,b] & Yorgo Modis[a,b,#]

[a]*Molecular Immunity Unit, Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom*

[b]*Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Department of Medicine, University of Cambridge, Cambridge CB2 0AW, United Kingdom*

[#]*Correspondence and requests for materials should be addressed to Y.M. at MRC-LMB, Francis Crick Avenue, Cambridge CB2 0QH, United Kingdom. Phone: +44 1223 267282. e-mail: ymodis@mrc-lmb.cam.ac.uk*

Running title: Crystal structure of HERK-K Env SU

Keywords: endogenous retrovirus, betaretrovirus, envelope glycoprotein, N-linked glycosylation, ligand binding site, X-ray crystallography, receptor binding protein

ABSTRACT

The most recently acquired and transcriptionally active family of human endogenous retroviruses (HERVs) is HERV-K. Of the approximately 100 copies of HERV-K in our genome, many retain the potential to proliferate by retrotransposition, express viral proteins, and form functional virus particles. Aberrant expression of the HERV-K envelope glycoprotein (Env) has been associated with cancer and neurodegeneration. Autoantibodies against HERV-K Env have been found in patients with various autoimmune diseases. Here, we report the crystal structures of the Env surface subunit (SU) from HERV-K HML-2, determined at 2.25 Å resolution. The overall fold is somewhat similar to Syncycin-2 SU and distantly related to HIV-1 gp120. The structure contains five disulfide bonds and four N-linked glycans, including one glycan that appears important for structural stability. Two extended loops form a surface for potential interactions with cell-surface receptors or other cellular factors. The structure also contains three detergent molecules, similar in structure to cholesterol, bound to hydrophobic surface patches. This crystal structure provides a platform for future studies to map autoantigenic epitopes, identify small-molecules that interfere with HERV-K activity, and extend our mechanistic understanding of retroviruses.

IMPORTANCE 15% of the human genome consists of endogenous retroviruses and other virus-derived elements inherited from ancestral viral infections. Many endogenous retroviruses from the HERV-K family retain the ability to proliferate across the genome and produce virus-like particles. Aberrant expression of the HERV-K envelope glycoprotein is associated with cancer, neurodegeneration, and autoimmune disease. Here, we report the crystal structure of the HERV-K envelope glycoprotein surface subunit. The structure provides a view in atomic-level detail of the molecular components in HERV-K most likely to trigger autoimmune responses. The structure also identifies binding sites for drug-like molecules, paving the way for future studies to develop new therapeutics.

44 **INTRODUCTION**

45 Endogenous retroviruses (ERVs) and other endogenous viral elements, relics of
46 ancient germline infections, constitute approximately 15% of the human genome (1, 2).
47 Most human ERV (HERV) loci have been rendered nonfunctional by mutations,
48 deletions, and recombination events but HERV-K, belonging to a betaretrovirus-like
49 group within the *Retroviridae* family and representing the most recent wave of
50 endogenization, stands out as remarkably intact (3, 4). Many HERV-K, particularly, the
51 human MMTV-like family 2 (HML-2) proviruses maintain full-length open reading
52 frames (ORFs) for *gag*, *pol* and *env* genes (5). Notably, reconstructed consensus
53 HERV-K genomes can produce infectious virions *in vitro* (6, 7).

54 The HERV-K envelope glycoprotein (Env) has been implicated in both health
55 and disease contexts. During early embryonic development, HERV-K Env is expressed
56 on the surface of pluripotent stem cells regulating their function via its association with
57 CD98 and impacting neuronal differentiation (8). In adult somatic cells, HERV-K is
58 generally repressed (9). Reactivation of HERV-K and expression of HERV-K Env is
59 associated with a range of cancers, neurodegeneration, autoimmune diseases (10-15).
60 Expression of HERV-K Env has been reported in melanoma cells (10), breast cancer
61 (11), and ovarian cancer (12). Autoantibodies against HERV-K Env have been reported
62 in patients with systemic lupus erythematosus (SLE) (14), rheumatoid arthritis (RA)
63 (15) and amyotrophic lateral sclerosis (14).

64 As a functional class I fusion protein, HERV-K Env follows a canonical structure
65 organization and is synthesized as a precursor which is cleaved into surface (SU) and
66 transmembrane (TM) subunits (16). SU engages host factors for cell entry, heparan
67 sulfate glycosaminoglycans serve as primary attachment factors while the CD98 heavy
68 chain, a component of amino acid transporter LAT1, acts as a receptor (17, 18). Here
69 we report the crystal structure of the receptor-binding unit (SU) of a human

70 endogenous retrovirus at 2.25 Å resolution using a consensus sequence derived from
71 ten HERV-K (HML-2) proviruses (7). Our crystal structure complements two cryo-EM
72 reconstructions of the trimeric Env SU-TM (19, 20), published while this manuscript
73 was in preparation, to provide a high-resolution view of the SU subunit including
74 glycosylation and disulfide bonds. In addition, we identify detergent molecules bound
75 within distinct SU pockets, which shed light on potential mechanisms of receptor
76 engagement at the cell surface, SU-TM interactions and provide starting points for
77 structure-guided therapeutic development. These insights establish an atomic
78 framework for dissecting HERV-K Env function and exploring strategies to counteract
79 HERV-K Env associated disease activity.

80

81 **RESULTS**

82 **Crystal structure of HERV-K Env SU.** Env SU (residues 111–430) of a consensus-
83 sequence HERV-K construct, HERV-K_{CON} (7), was expressed in HEK293S GnTI^{-} cells
84 to produce a uniform glycosylation pattern to facilitate glycoprotein crystallization.
85 HEK293S GnTI^{-} lack N-acetylglucosaminyltransferase I activity and accordingly
86 secrete proteins with Man5GlcNAc2 (mannopentaose-di-(N-acetyl-D-glucosamine)) as
87 the predominant N-linked glycosylation. The long N-terminal native signal peptide of
88 HERV-K Env (residues 1–110) was replaced by the signal peptide-pro-peptide domain
89 from human Trypsin-1 to enhance protein secretion (21). Substitution of cysteine at
90 position 141 with alanine (C141A) abolished non-specific protein cross-linking and
91 aggregation through aberrant disulfide bond formation. The abovementioned protein
92 engineering procedures produced a high yield of soluble HERV-K Env SU for
93 subsequent purification, crystallization, and crystallographic structure determination.

94 The crystal structure of HERV-K_{CON} Env SU was determined at 2.25 Å
95 resolution by molecular replacement using an atomic search model generated with
96 AlphaFold 3 (22). See **Table 1** for crystallographic data collection, refinement, and
97 validation statistics. The fold is composed primarily of β-sheets. An extended β-sheet
98 network at the core of the protein provides structural stability and organizes the fold
99 (**Fig. 1A**). The SU subunit can be subdivided into an inner and an outer domain,
100 adopting terminology from the structure of the gp120 envelope glycoprotein from HIV-1
101 (23, 24). The inner domain (residues 111-171 and 379-430) forms a membrane-
102 proximal structural unit expected to interface with the transmembrane (TM) subunit,
103 analogous to the interaction between HIV-1 gp120 and gp41 (the SU and TM subunits,
104 respectively). The HERV-K SU inner domain consists of β-strands and loops and
105 contains both the N- and C-termini of SU. The main structural feature of the inner
106 domain is a distorted β-barrel of six β-strands, stabilized by three disulfide bonds,
107 linking cysteines 164 and 171, 382 and 413, and 405 and 408. Notably, residues 405-
108 408 are part a CXXC motif, considered a hallmark of Envs from gamma-type
109 retroviruses. The outer domain (residues 172-378) encompasses the expected
110 receptor-binding region at the membrane-distal apex. It contains one long β-sheet of
111 five antiparallel β-strands and one short β-sheet of three antiparallel β-strands,
112 separated by two long loops, each stabilized by a disulfide bond (Cys227-Cys246 and
113 Cys275-Cys282). An α-helix is also present near the interface with the inner domain.

114 Four of the five predicted N-linked sequons (Asn128, Asn153, Asn274, and
115 Asn355) are glycosylated in HERV-K Env SU expressed in HEK293S GnTI^{-} cells; the
116 inner-domain sequon at Asn372 shows no evidence of glycosylation in the electron
117 density map. At Asn128 the map contained clear density for the chitobiose core (two N-
118 acetylglucosamines) plus four mannose moieties. Together these carbohydrates
119 residues form a hexasaccharide that is tightly lodged in a surface cavity between the
120 inner and outer domains, forming hydrophobic contacts with proline residues in each

121 domain (**Fig. 1B**). The unusually well-ordered glycan at Asn128 therefore behaves
122 more like an integral structural element (filling a void and making stabilizing contacts)
123 than a flexible surface decoration. By contrast, the glycans at positions 274 and 355 sit
124 on the exposed outer domain and display more limited, peripheral density consistent
125 with solvent-accessible, mobile oligomannose adducts that are unlikely to form a
126 continuous steric cloak. The glycan at position 153, although located in the inner
127 domain, lacks the structure seen at position 128 but may nevertheless contribute more
128 subtly to transient interactions with the TM subunit or modulate local folding dynamics.
129 These observations define a spatially sparse glycan landscape on HERV-K SU that
130 contrasts sharply with the dense glycan shield of HIV-1 gp120 (23-25) and more
131 closely resembles the lighter glycosylation reported for Syncytin-2 SU subunit (26). The
132 architecture implies relatively large protein surfaces will be available in HERV-K SU for
133 receptor engagement and antibody recognition, with the glycan at Asn128 representing
134 a localized structural exception rather than part of an immune-evasive glycan cloak.

135 Within the outer domain, two salt bridges play key roles in structural
136 stabilization. The first, between Asp264 and Arg268, reinforces the short β-sheet of 3
137 antiparallel β-strands and its adjoining loop (**Fig. 2**). The second, between Asp294,
138 located within a short α-helical segment, and Lys305, stabilizes the longer five-
139 stranded β-sheet together with the adjoining extended loop region (**Fig. 2**). Together,
140 these interactions delimit and stabilize the surface-exposed loop. A third salt bridge is
141 present between Glu182 of the longer β-sheet and Lys318 of a smaller loop between
142 the two β-sheets (**Fig. 2**). Other notable structural features are three relatively rare cis
143 peptide bonds involving residues Gly135-Pro136, Ser329-Pro330, and Pro385-Pro386.

144 **Detergent molecules bind hydrophobic surfaces at the TM interaction interface.**
145 A molecule of CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate),
146 a cholic acid derivative included in the crystallization solution (27), was modeled within

147 a hydrophobic pocket at the interface between the outer and inner SU domains (**Fig.**
148 **3**). The four-ring steroid nucleus of CHAPS – specifically rings C and D and one
149 protruding methyl group on its convex face – nestles into this pocket, which is formed
150 primarily by Phe259 (outer domain) and Pro398 (inner domain).

151 Two additional CHAPS molecules were modeled on concave hydrophobic
152 surface patches on the inner SU domain adjacent to the trypsin-derived signal-peptide
153 tail. One of these molecules makes hydrophobic contacts with a hydrophobic surface
154 clusters formed by Met150, Met388, Leu112, and Leu410, along with a phenylalanine
155 from the trypsin-derived signal-peptide tail (**Fig. 3**). The second CHAPS molecule is
156 stabilized by residues from an adjacent SU molecule in the crystal lattice
157 (crystallographic symmetry packing interactions), as well as by interactions with the
158 neighboring CHAPS molecule. These CHAPS molecules stabilize structural features in
159 the SU fragment that, in the intact envelope complex, are maintained via interactions
160 with the TM subunit (19, 20).

161 **Comparison to other retroviral Env SU structures.** A structural similarity search with
162 DALI (28) identified Syncytin-2 SU as the protein with the most similar structure to
163 HERV-K SU (Z-score 7.5, rmsd 5.9 Å, PDB 7OIX (26)). The next most similar
164 structures were those of HIV-1 gp120 (Z-score 3.2, rmsd 5.4 Å, PDB 6PWU (25)) and
165 SIV gp120 (Z-score 2.8, rmsd 3.4 Å, PDB 8DUA (29)), although the similarity scores for
166 these were only marginally higher than for clearly unrelated proteins (Z-score 2.7 and
167 below). Syncytin-2, which drives cell-cell fusion of trophoblasts fusion in placental
168 development, is derived from the Env of HERV-FRD, a gammaretrovirus-like
169 endogenous retrovirus (26, 30, 31). Most of the structural similarity between HERV-K
170 SU, Syncytin-2 SU and HIV-1 gp120 maps to the inner domain, which has a similar
171 distorted β-barrel as its core architectural element in all three proteins (**Fig. 4**). Upon
172 superposition on HERV-K SU inner domain, the root mean square deviations of the

173 inner domain Cα positions of Syncytin-2 SU and HIV-1 gp120 were 2.89 Å and 4.01 Å,
174 respectively. Given that the inner domain is proximal to the TM membrane fusion
175 subunit, the structural conservation of this domain suggests a common mode of
176 intramolecular communication within the heterodimeric Env complexes.

177 The outer domain of HERV-K SU incorporates a mixed architecture comprising
178 β-sheets, loop insertions, and helical elements. The main structural feature is a long β-
179 sheet composed of five antiparallel β-strands, a topological element also found in the
180 outer domain of Syncytin-2 SU. However, the structures of the outer domains, and the
181 insertions within them, are otherwise highly divergent in HERV-K SU, Syncytin-2 SU
182 and HIV-1 gp120 (**Fig. 4A**).

183

184 **DISCUSSION**

185 Here we report the crystal structure and associated atomic model of the envelope
186 glycoprotein Surface (SU) subunit from the human endogenous betaretrovirus HERV-K
187 (HML-2). The HERV-K SU fold retains the two-domain topology, with inner and outer
188 domains, seen in other retroviral SU subunits including gammaretroviruses (Syncytin-2)
189 and lentiviruses (HIV-1). This topological conservation demonstrates that a common
190 architectural blueprint underlies Env proteins across the *Orthoretrovirinae* subfamily of
191 *Retroviridae* (which includes the genera alpharetrovirus, betaretrovirus,
192 gammaretrovirus, lentivirus, and others), with genus-specific refinements (glycan
193 decorations, loop insertions, receptor specificities) overlaying the common framework.
194 However, despite the similarity in secondary structure topology to other retroviral SUs,
195 perhaps the most striking feature of the structure of HERV-K SU is how different its
196 three-dimensional structure is from those of its closest structural homologs, Syncytin-2
197 SU and HIV-1 gp120. This indicates that HERV-K SU has diverged further from the

198 SUs of other gamma- and gamma-like retroviruses at the structural level than might
199 have been expected from phylogenetic analysis.

200 The crystal structure of HERV-K SU, determined at 2.25 Å resolution, reveals
201 various HERV-K-specific structural elements. The fold is stabilized by five disulfide
202 bonds and three key salt bridges. Two loop regions provide an extensive surface for
203 potential interactions with cell-surface receptors or other cellular factors. The structure
204 offers a detailed view of the glycan landscape. Four N-linked glycans make for a
205 relatively sparse coverage of the molecular surface of SU, standing in contrast with the
206 dense glycan shield of HIV-1 gp120. An unusually well-ordered, cavity-embedded
207 branched carbohydrate chain at Asn128 of HERV-K SU appears to function as an
208 integral structural element within the inner domain. The crystal structure also contains
209 three amphiphilic CHAPS molecules, similar in structure to cholesterol, bound to
210 hydrophobic surface patches of SU. These could prove to be attractive starting points
211 for small-molecule or fragment screening campaigns to identify and compounds that
212 bind tightly to SU and interfere with the membrane fusion activity of HERV-K Env or its
213 ability to interact with other cellular proteins.

214 The HERV-K SU crystal structure reveals the conserved structural features
215 shared with other orthoretroviral Envs while also identifying HERV-K-specific features,
216 namely its sparse glycocalyx, structure-stabilizing glycan, hydrophobic surface patches,
217 and extended surface-exposed loops. Autoantibodies against HERV-K Env have been
218 reported in patients with various autoimmune diseases (14, 15). Given that SU is the
219 receptor binding subunit in retroviruses and contains most of the known antibody
220 binding epitopes, the epitopes of disease-causing autoantibodies against HERV-K Env
221 are likely to map to the solvent accessible surface of SU. The structure of SU will be an
222 invaluable tool in future studies to identify these epitopes. More broadly, the availability
223 of a high-resolution crystal structure for HERV-K SU, along with our optimized

224 expression construct and purification protocol, open paths toward extending our
225 mechanistic understanding of betaretroviral envelope proteins and developing
226 translational applications targeting HERV-K elements, the most active and recently
227 acquired endogenous retroviruses in humans.

228

229 **MATERIALS AND METHODS**

230 **Cloning of expression vectors.** The gene encoding the SU domain of HERV-K_{CON}
231 Env (residues 111-430) (7) was cloned into the pcDNA3.1(+) plasmid. To enhance
232 protein secretion, an N-terminal signal peptide–pro-peptide domain derived from
233 human Trypsin-1 was fused to the gene (21), followed by a human rhinovirus (HRV) 3C
234 protease cleavage site. A C-terminal hexahistidine (His6) tag was added for purification
235 purposes, preceded by a Pro-Gly sequence to facilitate efficient removal of the His6 tag
236 by carboxypeptidase A. Additionally, the sequence was engineered to substitute
237 cysteine at position 141 with alanine to prevent oligomerization caused by the free
238 cysteine residue.

239 **Protein expression and purification.** The HERV-K Env SU domain was expressed in
240 HEK293S GnTI$^-$ cells adapted for suspension culture and maintained at 37 °C in a
241 humidified atmosphere containing 8% CO_2 with shaking. Cells were grown in serum-
242 free Expi293 Expression Medium (Gibco) to a final density of 3×10^6 viable cells/mL
243 with 95–99% viability before transfection. Transfection was performed using 1 mg/mL
244 polyethylenimine (PEI) at a DNA:PEI ratio of 1:3 (w/w). Secreted protein was harvested
245 5 days post-transfection by collecting the supernatant after centrifugation at 300 × g for
246 5 min at 4 °C to pellet cells. The clarified medium was applied to Ni Sepharose Excel
247 resin (Cytiva), washed, and eluted with 50 mM Tris pH 7.4, 300 mM NaCl containing 10
248 mM and 250 mM imidazole, respectively. The protein was further purified by size-

249 exclusion chromatography using a Superdex 200 Increase 10/300 GL column (Cytiva)
250 equilibrated in 50 mM Tris pH 7.4, 100 mM NaCl. (**Fig. S1**).

251 **Crystallographic structure determination.** Crystals were grown at 18 °C by sitting-
252 drop vapor diffusion. Purified protein was concentrated to 11.7 mg/mL, and 1 volume of
253 seed stock was added to 5 volumes of protein solution. The final solution was mixed
254 with an equal volume of reservoir solution optimized from the Morpheus Fusion Screen
255 (Molecular Dimensions) (27): 32.7% PEG 8K/ethylene glycol (1:2), 0.1 M
256 MES/imidazole pH 6.5, 90 mM Morpheus NPS, and 1.572% (w/v) Morpheus III cholic
257 acids. Seed stock was prepared by crushing small crystals from a previous optimized
258 crystallization experiment conducted with varying PEG 8K/ethylene glycol (1:2)
259 concentrations (18–36%) and Morpheus III cholic acids (0.24–1.8% w/v), while
260 maintaining constant 0.1 M MES/imidazole pH 6.5 and 90 mM Morpheus NPS. X-ray
261 diffraction data were collected at 100 K using Pilatus3 6M and Eiger2 XE 16M
262 detectors at Diamond Light Source beamlines I24 and I04, respectively. Data were
263 processed with autoPROC v1.0.5 (32), and Xia2 (Dials, Aimless) (33). Molecular
264 replacement was performed with Phaser (34) as implemented in PHENIX v1.21.1 (35)
265 using an atomic search model of the SU domain of HERV-K_{CON} Env from AlphaFold 3
266 (22). An initial model was built using AutoBuild in PHENIX, manually completed with
267 COOT v.0.9.6 (36), and iteratively refined with PHENIX. See **Table 1** for
268 crystallographic data collection, refinement, and validation statistics.

269 **Structural analysis.** Coordinate files were submitted to PDBsum (37) for structural
270 feature analysis. Molecular visualization and figure generation were performed using
271 UCSF ChimeraX (38). Structural homologs were identified using the DALI server (39),
272 and 3D structural comparisons were carried out using the TM-align algorithm (40).

Data availability. The atomic coordinates of the HERV-K SU glycoprotein fragment were deposited in the Protein Data Bank with accession codes 9SYA [https://doi.org/10.2210/pdb9sya/pdb].

Acknowledgements. We thank Fabrice Gorrec (MRC-LMB Crystallization Facility) for advice on protein crystallization, and Dom Bellini at the MRC-LMB X-ray Crystallography Facility for advice in crystallographic data collection. Crystallographic data were collected on beamlines I03 and I04 at Diamond Light Source (DLS). Access to DLS was supported by the Wellcome Trust, MRC and BBSRC. This work was supported by Wellcome Trust Senior Research Fellowship 217191/Z/19/Z to Y.M. Access to Diamond Light Source was supported by the Wellcome Trust; the Medical Research Council; and UKRI (proposal number MX36838). Open access publication was funded by the University of Cambridge. To ensure open access, the University of Cambridge has applied a CC BY public copyright license to any Author Accepted Manuscript version arising.

References

1. Frank JA, Feschotte C. 2017. Co-option of endogenous viral sequences for host cell function. Curr Opin Virol 25:81–89.
2. Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. PLoS Genet 6:e1001191.
3. Tonjes RR, Lower R, Boller K, Denner J, Hasenmaier B, Kirsch H, Konig H, Korbmacher C, Limbach C, Lugert R, Phelps RC, Scherer J, Thelen K, Lower J, Kurth R. 1996. HERV-K: The biologically most active human endogenous retrovirus family. Journal of Acquired Immune Deficiency Syndromes 13:S261–S267.

298 4. Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J. 2001.
299 Insertional polymorphisms of full-length endogenous retroviruses in humans. Curr
300 Biol 11:1531–5.
301 5. Hohn O, Hanke K, Bannert N. 2013. HERV-K(HML-2), the Best Preserved Family
302 of HERVs: Endogenization, Expression, and Implications in Health and Disease.
303 Front Oncol 3:246.
304 6. Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, Pierron G, Heidmann T.
305 2006. Identification of an infectious progenitor for the multiple-copy HERV-K human
306 endogenous retroelements. Genome Res 16:1548–56.
307 7. Lee YN, Bieniasz PD. 2007. Reconstitution of an infectious human endogenous
308 retrovirus. PLoS Pathog 3:e10.
309 8. Wang T, Medynets M, Johnson KR, Doucet-O'Hare TT, DiSanza B, Li W, Xu Y,
310 Bagnell A, Tyagi R, Sampson K, Malik N, Steiner J, Hadegan A, Kowalak J,
311 O'Malley J, Maric D, Nath A. 2020. Regulation of stem cell function and neuronal
312 differentiation by HERV-K via mTOR pathway. Proc Natl Acad Sci U S A
313 117:17842–17853.
314 9. Lavie L, Kitova M, Maldener E, Meese E, Mayer J. 2005. CpG methylation directly
315 regulates transcriptional activity of the human endogenous retrovirus family HERV-
316 K(HML-2). J Virol 79:876–83.
317 10. Buscher K, Trefzer U, Hofmann M, Sterry W, Kurth R, Denner J. 2005. Expression
318 of human endogenous retrovirus K in melanomas and melanoma cell lines. Cancer
319 Res 65:4172–80.
320 11. Wang-Johanning F, Radvanyi L, Rycaj K, Plummer JB, Yan P, Sastry KJ,
321 Piyathilake CJ, Hunt KK, Johanning GL. 2008. Human endogenous retrovirus K
322 triggers an antigen-specific immune response in breast cancer patients. Cancer
323 Res 68:5869–77.

324 12. Wang-Johanning F, Liu J, Rycaj K, Huang M, Tsai K, Rosen DG, Chen DT, Lu DW,
325 Barnhart KF, Johanning GL. 2007. Expression of multiple human endogenous
326 retrovirus surface envelope proteins in ovarian cancer. Int J Cancer 120:81–90.
327 13. Li W, Lee MH, Henderson L, Tyagi R, Bachani M, Steiner J, Campanac E, Hoffman
328 DA, von Geldern G, Johnson K, Maric D, Morris HD, Lentz M, Pak K, Mammen A,
329 Ostrow L, Rothstein J, Nath A. 2015. Human endogenous retrovirus-K contributes
330 to motor neuron disease. Sci Transl Med 7:307ra153.
331 14. Khadjinova AI, Wang X, Laine A, Ukadike K, Eckert M, Stevens A, Bengtsson AA,
332 Lood C, Mustelin T. 2022. Autoantibodies against the envelope proteins of
333 endogenous retroviruses K102 and K108 in patients with systemic lupus
334 erythematosus correlate with active disease. Clin Exp Rheumatol 40:1306–1312.
335 15. Mameli G, Erre GL, Caggiu E, Mura S, Cossu D, Bo M, Cadoni ML, Piras A,
336 Mundula N, Colombo E, Buscetta G, Passiu G, Sechi LA. 2017. Identification of a
337 HERV-K env surface peptide highly recognized in Rheumatoid Arthritis (RA)
338 patients: a cross-sectional case-control study. Clin Exp Immunol 189:127–131.
339 16. Dewannieux M, Blaise S, Heidmann T. 2005. Identification of a functional envelope
340 protein from the HERV-K family of human endogenous retroviruses. J Virol
341 79:15573–7.
342 17. Robinson-McCarthy LR, McCarthy KR, Raaben M, Piccinotti S, Nieuwenhuis J,
343 Stubbs SH, Bakkers MJG, Whelan SPJ. 2018. Reconstruction of the cell entry
344 pathway of an extinct virus. PLoS Pathog 14:e1007123.
345 18. Steiner JP, Bachani M, Malik N, DeMarino C, Li WX, Sampson K, Lee MH, Kowalak
346 J, Bhaskar M, Doucet-O'Hare T, Garcia-Montojo M, Cowen M, Smith B, Reoma LB,
347 Medina J, Brunel J, Pierquin J, Charvet B, Perron H, Nath A. 2022. Human
348 Endogenous Retrovirus K Envelope in Spinal Fluid of Amyotrophic Lateral Sclerosis
349 Is Toxic. Annals of Neurology 92:545–561.

350 19. Shaked R, Katz M, Cohen-Dvashi H, Diskin R. 2025. The prefusion structure of the
351 HERV-K (HML-2) Env spike complex. Proc Natl Acad Sci U S A 122:e2505505122.
352 20. Shek J, Sun C, Wilson EM, Moadab F, Hastie KM, Rajamanickam RR, Penalosa
353 PJ, Harkins SS, Parekh D, Hariharan C, Zyla DS, Yu C, Shaffer KCL, Lewis VI,
354 Avalos RD, Mustelin T, Saphire EO. 2025. Human endogenous retrovirus K
355 (HERV-K) envelope structures in pre- and postfusion by cryo-EM. Sci Adv
356 11:eady8168.
357 21. Mullin MJ, Wilkinson C, Hiles I, Smith KJ. 2021. Improved secretion of recombinant
358 human IL-25 in HEK293 cells using a signal peptide-pro-peptide domain derived
359 from Trypsin-1. Biotechnol Lett 43:757–765.
360 22. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O,
361 Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill
362 M, Reiman D, Tunyasuvunakool K, Wu Z, Zemgulyte A, Arvaniti E, Beattie C,
363 Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A,
364 Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K,
365 Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S,
366 Zhong ED, Zielinski M, Zidek A, Bapst V, Kohli P, Jaderberg M, Hassabis D,
367 Jumper JM. 2024. Accurate structure prediction of biomolecular interactions with
368 AlphaFold 3. Nature 630:493–500.
369 23. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. 1998.
370 Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor
371 and a neutralizing human antibody. Nature 393:648–59.
372 24. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA,
373 Sodroski JG. 1998. The antigenic structure of the HIV gp120 envelope glycoprotein.
374 Nature 393:705–11.
375 25. Pan J, Peng H, Chen B, Harrison SC. 2020. Cryo-EM Structure of Full-length HIV-1
376 Env Bound With the Fab of Antibody PG16. J Mol Biol 432:1158–1168.

377 26. Martinez-Molledo M, Nji E, Reyes N. 2022. Structural insights into the
378 lysophospholipid brain uptake mechanism and its inhibition by syncytin-2. Nature
379 Structural & Molecular Biology 29:604–612.
380 27. Gorrec F, Bellini D. 2022. The FUSION protein crystallization screen. J Appl
381 Crystallogr 55:310–319.
382 28. Holm L, Rosenstrom P. 2010. Dali server: conservation mapping in 3D. Nucleic
383 Acids Res 38:W545–9.
384 29. Gorman J, Wang C, Mason RD, Nazzari AF, Welles HC, Zhou T, Bess JW, Jr.,
385 Bylund T, Lee M, Tsybovsky Y, Verardi R, Wang S, Yang Y, Zhang B, Rawi R,
386 Keele BF, Lifson JD, Liu J, Roederer M, Kwong PD. 2022. Cryo-EM structures of
387 prefusion SIV envelope trimer. Nat Struct Mol Biol 29:1080–1091.
388 30. Friedli M, Trono D. 2015. The developmental control of transposable elements and
389 the evolution of higher species. Annu Rev Cell Dev Biol 31:429–51.
390 31. Dupressoir A, Lavialle C, Heidmann T. 2012. From ancestral infectious retroviruses
391 to bona fide cellular genes: role of the captured syncytins in placentation. Placenta
392 33:663–71.
393 32. Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T,
394 Bricogne G. 2011. Data processing and analysis with the autoPROC toolbox. Acta
395 Crystallogr D Biol Crystallogr 67:293–302.
396 33. Winter G. 2010. xia2: an expert system for macromolecular crystallography data
397 reduction. Journal of Applied Crystallography 43:186–190.
398 34. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ.
399 2007. Phaser crystallographic software. J Appl Cryst 40:658–674.
400 35. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ,
401 Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R,
402 Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010.

403 PHENIX: a comprehensive Python-based system for macromolecular structure
404 solution. Acta Crystallogr D Biol Crystallogr 66:213–21.
405 36. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. Acta
406 Crystallogr D Biol Crystallogr 60:2126–32.
407 37. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM.
408 1997. PDBsum: a Web-based database of summaries and analyses of all PDB
409 structures. Trends Biochem Sci 22:488–90.
410 38. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE.
411 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis.
412 Protein Sci 27:14–25.
413 39. Holm L, Laakso LM. 2016. Dali server update. Nucleic Acids Res 44:W351–5.
414 40. Zhang Y, Skolnick J. 2005. TM-align: a protein structure alignment algorithm based
415 on the TM-score. Nucleic Acids Res 33:2302–9.

416

417 **Figure Legends.**

418 **Fig. 1.** Overall fold and structural features of HERV-K Env SU. (**A**) Overall fold of
419 (residues 111-430), consisting of inner and outer domains. Four N-linked glycans are
420 shown in ball-and-stick representation with carbon atoms in grey. Five disulfides are
421 shown in ball-and-stick representation with yellow sulfur atoms (SS_{1-5}). The TM domain
422 and viral membrane would be located below the N- and C-termini (labeled N', C'). (**B**)
423 Closeup of the hexasaccharide glycan linked to Asn128, at the interface between the
424 inner and outer domains. NAG, N-acetyl glucosamine; MAN, α-*D*-mannose; BMA, β-*D*-
425 mannose. See **Table 1** for crystallographic data collection, refinement, and validation
426 statistics. (**C**) Closeups showing three salt bridges that stabilize the fold of the outer
427 domain.

428 **Fig. 2.** Binding mode of three sterol molecules to HERV-K SU. Three cholic acid
429 derivative molecules were clearly identifiable in the electron density map and modeled
430 as CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate). Cholic acid
431 derivatives were present as additives in the crystallization condition (see **Materials and**
432 **Methods**) (27). CHAPS molecules are shown in ball-and-stick representation with pale
433 yellow carbon atoms. "*A"* and "*B"* denote two HERV-K SU chains in the crystallographic
434 asymmetric unit. Proteins shown in grey are related to the asymmetric unit by
435 crystallographic symmetry. Their chain and residue identifiers are show in grey and
436 followed by a prime symbol.

437 **Fig. 3.** Comparison of HERV-K Env SU to Syncytin-2 SU and HIV-1 gp120. Structures
438 of HIV-1 gp120 (PDB 6PWU; (25)), left, and Syncytin-2 SU (PDB 7OIX; (26)), right, the
439 closest structural homologs of HERV-K SU. The relative orientations of the proteins
440 reflect structural alignment of their inner domain using the HERV-K SU inner domain as
441 the reference. The root mean square deviation of the inner domain Cα positions after
442 superposition onto the HERV-K SU inner domain was 2.89 Å for Syncytin-2 SU and
443 4.01 Å for HIV-1 gp120.

444

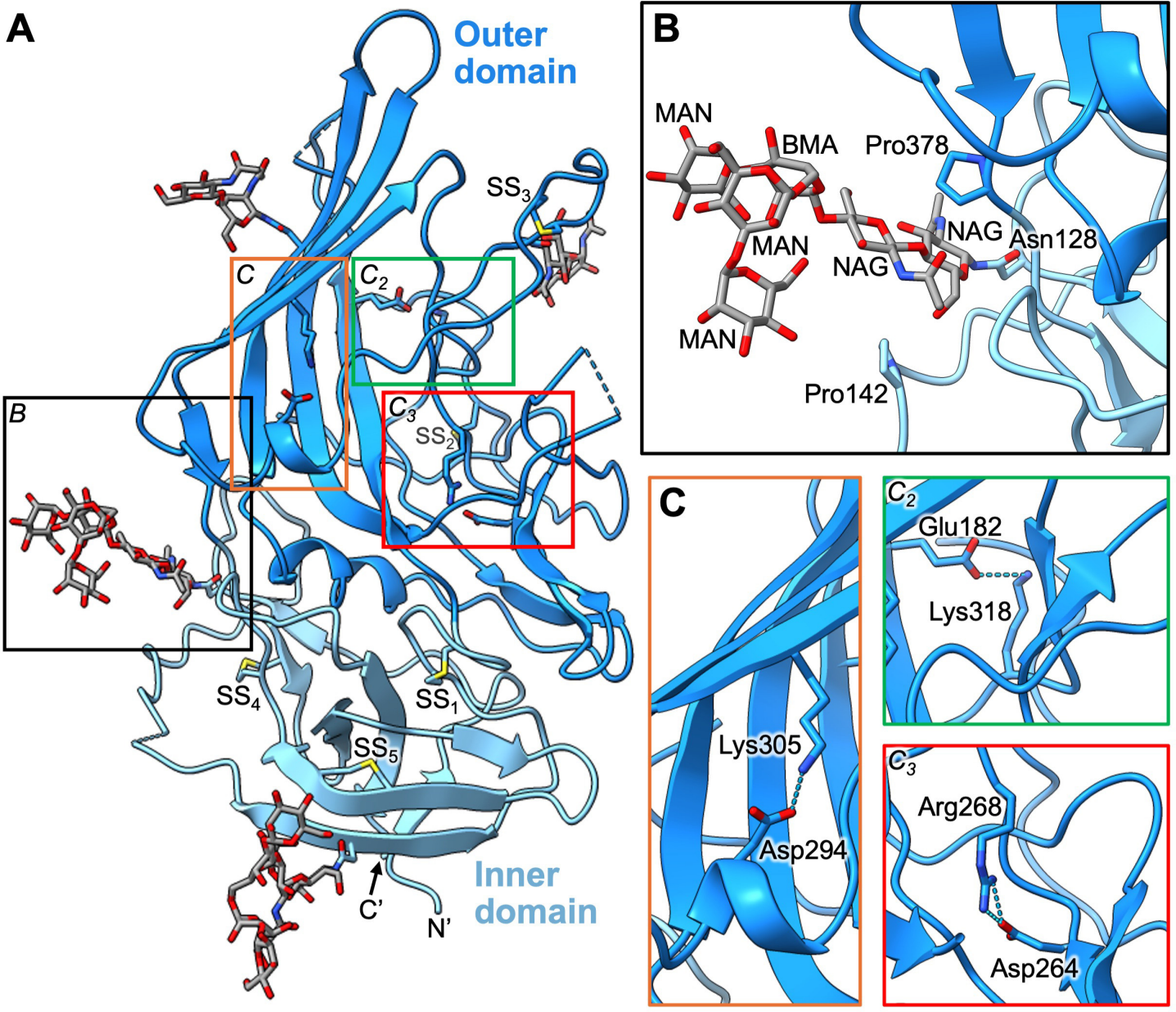
445 **Table 1. Crystallographic data collection, refinement, and validation statistics**

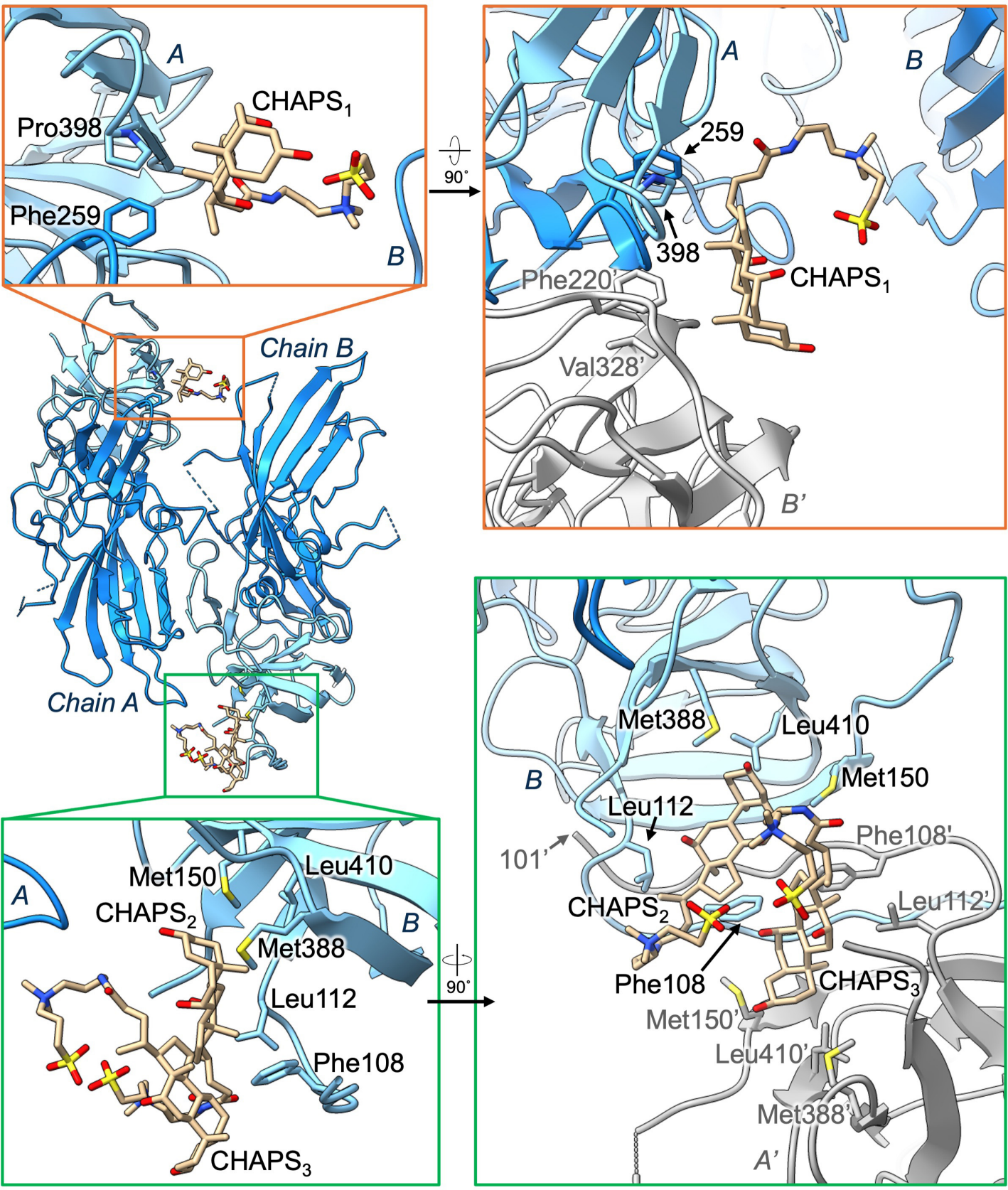
Data collection	HERV-K Env SU
Space group	*P* 1
Cell dimensions	
a, b, c (Å)	60.04 63.52 70.68
α, β, γ (°)	101.94, 101.83, 104.06
Resolution (Å)	59.32 – 2.254 (2.31 – 2.25)
Total reflections	162,480 (7,933)
Unique reflections	44,339 (3,144)
Multiplicity[a]	3.7 (3.5)
Completeness	98.2% (97.4%)
$\langle I \rangle / \sigma(I)$	11.4 (0.7)
Wilson *B*-factor	47.2
R_{merge}	0.114 (1.50)
R_{pim}	0.070 (0.931)
CC1/2	0.979 (0.317)
Refinement and Validation	
Resolution (Å)	59 – 2.25 (2.31 – 2.25)
No. reflections, working set	44,339 (3,144)
No. reflections, test set	2,018 (145)
R_{work}	0.2265 (0.3852)
R_{free}	0.2864 (0.4396)

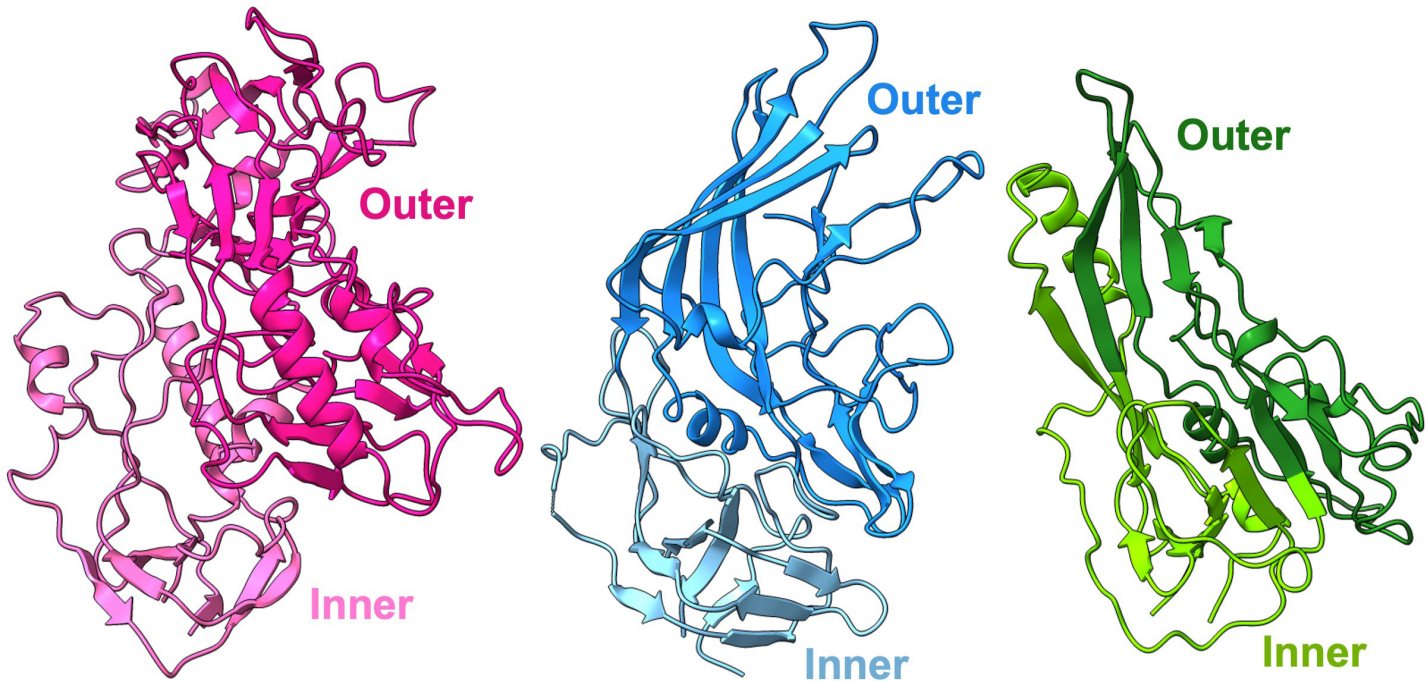
No. non-hydrogen atoms	5,903
Protein	5,199
Glycans	353
Ligands	126
Water	225
Average *B*-factors[a]	
Protein (Å^2)	57.53
Glycans (Å^2)	91
Ligand (Å^2)	87
Water (Å^2)	52
Root mean square deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.13
Ramachandran plot	
Favored (%)	92.23
Allowed (%)	7.29
Outliers (%)	0.48
Clashscore (Phenix 1.21.1)	14.28
PDB accession code	9SYA

446 Numbers in parentheses refer to highest resolution shell.

447 [a]Residual B-factors after TLS refinement. See PDB entry for TLS parameters.


A
Outer domain
SS3
C
C2
C3
SS2
B
SS4
SS1
SS5
C'
N'
Inner domain
B
MAN
BMA
Pro378
MAN
NAG
NAG
Asn128
MAN
Pro142
C
Lys305
Asp294
C2
Glu182
Lys318
C3
Arg268
Asp264


A
CHAPS1
Pro398
Phe259
B
90°
A
B
259
398
Phe220'
CHAPS1
Val328'
B'
Chain B
Chain A
A
Met150
Leu410
CHAPS2
B
Met388
Leu112
Phe108
CHAPS3
90°
Met388
Leu410
B
Met150
Leu112
Phe108'
101'
CHAPS2
Leu112'
Phe108
CHAPS3
Met150'
Leu410'
Met388'
A'



HIV-1 gp120 HERV-K Env SU Syncytin-2 SU