Check for updates

HIV Transmembrane Glycoprotein Conserved Domains and Genetic Markers Across HIV-1 and HIV-2 Variants

Ana Valadés-Alcaraz, Roberto Reinosa and África Holguín*

HIV-1 Molecular Epidemiology Laboratory, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Microbiology Department, Hospital Universitario Ramón y Cajal, CIBER en Epidemiología y Salud Pública (CIBERESP), Red en Investigación Translacional en Infecciones Pediátricas (RITIP), Madrid, Spain

HIV envelope transmembrane glycoproteins gp41 (HIV-1) and gp36 (HIV-2) present high variability and play a key role in the HIV-host cell membrane's fusion, as a target for human broadly neutralizing antibodies (bnAbs) and drugs. Thus, a better knowledge of amino acid (aa) conservation across structural domains and HIV variants can help to identify conserved targets to direct new therapeutic and diagnostic strategies. All available gp41/gp36 nucleotide sequences were downloaded from Los Alamos National Laboratory (LANL) HIV Sequence Database, selecting 17,078 sequences ascribed to HIV-1 and HIV-2 variants with ≥3 sequences. After aligning and translating into aa with MEGAv6.0, an in-house bioinformatics program (EpiMolBio) was used to identify the most conserved aa and the aa changes that were specific for each variant (V-markers) *vs.* HXB2/BEN (HIV-1/HIV-2) reference sequence. We analyzed the presence of specific aa changes among V-markers affecting infectivity, gp41 structure, function, or resistance to the enfuvirtide viral fusion inhibitor (T-20). We also inferred the consensus sequences per HIV variant, describing in each HIV-1 group (M, N, O, P) the conservation level along the complete gp41 per structural domain and locating in each binding site the anti-gp41 human Abs (bnAbs and non bnAbs) described in LANL. We found 38.3/59.7% highly conserved aa present in ≥90% of the 16,803/275 gp41/gp36 sequences ascribed to 105/3 HIV-1/HIV-2 variants, with 9/12.6% of them showing complete conservation across LANL sequences. The fusion peptide, its proximal region, the N-heptad repeat, and the membrane-proximal external region were the gp41 domains with ≥84% of conserved aa in the HIV-1 consensus sequence, the target of most Abs. No natural major resistance mutations to T-20 were observed. Our results show, for the first time, a complete conservation study of gp41/gp36 per variant in the largest panel of HIV variants analyzed to date, providing useful information for a more rational design of drugs, vaccines, and molecular detection tests targeting the HIV transmembrane glycoprotein.
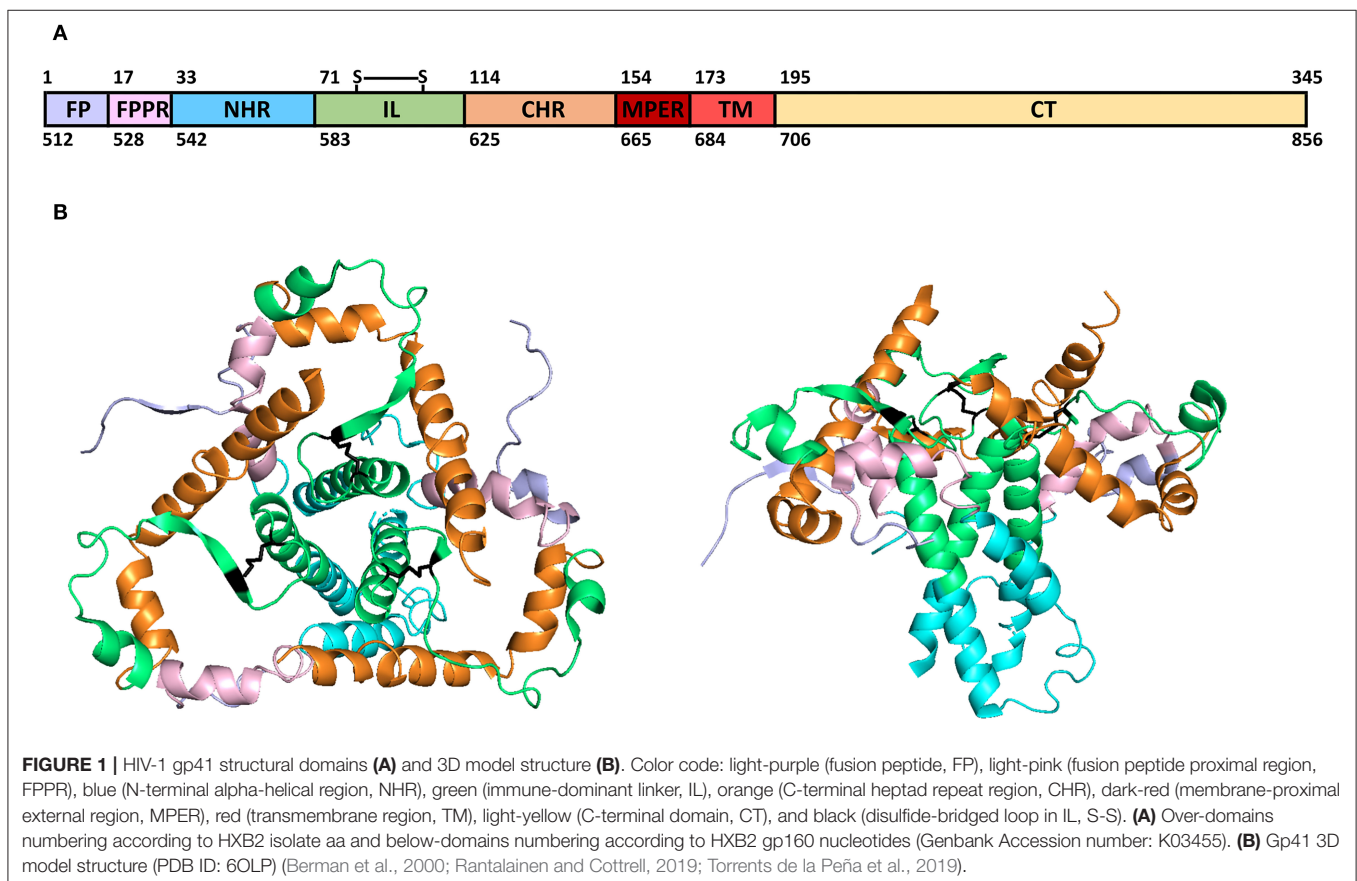
Keywords: HIV, transmembrane glycoprotein gp41, transmembrane glycoprotein gp36, conservation, variants, antibody binding domains

INTRODUCTION

The human immunodeficiency virus (HIV) envelope transmembrane glycoproteins gp41 and gp36 are located on the HIV-1 and HIV-2 virion's membrane forming trimers with gp120 and gp105 glycoproteins, respectively. These proteins mediate the viral fusion with the host cell's membrane, allowing the entry of genetic material and viral proteins into the cells (Blumenthal et al., 2012). Therefore, they are important targets for the development of fusion inhibitors, such as antiretrovirals (ARV) (Qadir and Malik, 2010), antibodies (Abs) (Caillat et al., 2020), and aptamers (Li et al., 2016) used as HIV treatment. HIV-1 gp41, with 345 amino acids (aa), can be segmented into three domains (**Figure 1A**): one exposed ectodomain (aa 1-172), a transmembrane region (TM, aa 173-194), and an intraviral not exposed C-terminal domain (CT, aa 195-345). The ectodomain is exposed and can be divided further into distinct functional regions important for fusion and virus infectivity: an N-terminal hydrophobic region termed as fusion peptide (FP, aa 1-16), necessary to bind the virus to the cell membrane, followed by an N-terminal alpha-helical region or N-heptad repeat region (NHR, aa 33-70). These domains were linked by a fusion peptide proximal region (FPPR, aa 17-32) rich in polar aa and critical for HIV-1 fusion and infectivity because it stabilizes the envelope trimers (Lu et al., 2019). A loop immune-dominant linker with

a disulfide bridge (IL, aa 71-113) links the NHR to a C-heptad repeat region (CHR, aa 114-153) (**Figure 1B**). A membrane-proximal external region (MPER, aa 154-172), conformationally flexible, connects the CHR to the TM region (Louis et al., 2016). The gp41 NHR domain is the target of ARV, such as enfuvirtide (T-20), the only clinically approved viral fusion inhibitor for the treatment of HIV infection (Lazzarin, 2005; Oldfield et al., 2005), and MPER for immunogens, as it contains epitopes to broadly neutralizing antibodies (bnAbs), such as 2F5, 4E10, Z13, and 10E8 (Los Alamos HIV Molecular Immunology Database, 2021a).

HIV is one of the most genetically diverse pathogens due to its high mutation and recombination rates, large population size, and rapid replication rate (Hemelaar, 2012). The HIV epidemic is the result of two types of viruses: HIV-1 and HIV-2, which are closely related to SIVcpz (Gao et al., 1999) and SIVsm (Gao et al., 1992), respectively. HIV-1 causes most of the HIV infections worldwide and has been divided according to genetic homology into four groups: M (major or main), N (non-M, non-O) (Simon et al., 1998), O (outlier) (De Leys et al., 1990), and P (Plantier et al., 2009). However, the global HIV epidemic is related to group M (Hemelaar et al., 2019), which has been subdivided into 10 subtypes (A–D, F–H, J–L) and eight sub-subtypes (A1, A2, A3, A4, A5, A6, F1, F2) (Robertson et al., 2000; Salminen, 2000; Leitner et al., 2005; Yamaguchi et al., 2020), at least 118 circulating

FIGURE 1 | HIV-1 gp41 structural domains **(A)** and 3D model structure **(B)**. Color code: light-purple (fusion peptide, FP), light-pink (fusion peptide proximal region, FPPR), blue (N-terminal alpha-helical region, NHR), green (immune-dominant linker, IL), orange (C-terminal heptad repeat region, CHR), dark-red (membrane-proximal external region, MPER), red (transmembrane region, TM), light-yellow (C-terminal domain, CT), and black (disulfide-bridged loop in IL, S-S). **(A)** Over-domains numbering according to HXB2 isolate aa and below-domains numbering according to HXB2 gp160 nucleotides (Genbank Accession number: K03455). **(B)** Gp41 3D model structure (PDB ID: 6OLP) (Berman et al., 2000; Rantalainen and Cottrell, 2019; Torrents de la Peña et al., 2019).

recombinant forms (CRF) (Los Alamos HIV Sequence Database, 2021a) and uncountable unique recombinant forms (URF). HIV-2 has been classified into nine groups (A-I) and two recombinants (CRF01_AB and URF) (Visseaux et al., 2016).

Since the HIV transmembrane glycoprotein is a key target for human bnAbs and anti-HIV drugs, a better knowledge of aa conservation across structural domains and HIV variants can help to identify conserved targets to direct new therapeutic and diagnostic strategies. Furthermore, each HIV variant presents natural polymorphisms and unique aa changes (V-markers) along the viral genome fixed during viral evolution (Arenas et al., 2016), which have not been described to date in each HIV variant, mainly in HIV-1 group M non-B subtypes and recombinants, which are the majority in the pandemic (Hemelaar et al., 2019), and in HIV-2.

We present, for the first time, the most conserved gp41 domains in each HIV-1 variant per structural domain and anti-gp41 antibody binding domains in the largest panel of HIV-1 variants analyzed to date, identifying the V-markers and the consensus transmembrane glycoprotein sequence for each HIV variant (type, group, subtype, sub-subtype, and CRF).

MATERIALS AND METHODS

HIV Transmembrane Glycoprotein Sequences

In October and November of 2020, we downloaded all available gp41 (HIV-1, 345 aa) and gp36 (HIV-2, 350 aa) nucleotides sequences from Los Alamos National Laboratory (LANL) HIV Sequence Database (Los Alamos HIV Sequence Database, 2021b) selecting one sequence per patient and grouping them per HIV variant (types, groups, subtypes, sub-subtypes, and CRF). URF sequences were not included in this study. They were aligned, edited, and translated into aa with the MEGAv6.0 program (Molecular Evolutionary Genetics Analysis: https://www.megasoftware.net/) (Tamura et al., 2013). MUSCLE function (Multiple Sequence Comparison by Log Expectation) (Edgar, 2004) was used for alignments with HXB2 (HIV-1 subtype B, GenBank accession number: K03455) or BEN (HIV-2 subtype A, GenBank accession number: M30502) reference sequences, removing nucleotides insertions. Sequences with stop codons in unusual positions and groups, subtypes, sub-subtypes, and CRF with <3 sequences were excluded from the study, except group P, as it was necessary to establish the HIV-1 aa sequence consensus.

Gp41/Gp36 aa Conservation and Inferred Consensus Sequences

Using a bioinformatics tool developed in our laboratory (EpiMolBio program), we analyzed the gp41/gp36 aa conservation of HIV variants with at least three available sequences (except group P, with two sequences). We also inferred the aa consensus sequences for gp41 HIV-1/gp36 HIV-2 and each HIV-1/HIV-2 variant, providing the most conserved aa in each residue.

The new EpiMolBio bioinformatics tool reported the percentage of sequences with a conserved aa in each position of any protein, establishing a color code for this study depending on the frequency of each conserved aa in gp41/gp36: white (<90%), light-blue/light-pink (≥90–<100%), and dark blue-green/fuchsia (100% or complete conservation across considered sequences), respectively. We also studied the level of conserved aa per gp41 structural domain in each HIV-1 non-M group, group M, and HIV-1 consensus sequences. For the analysis, we summed the aa conservation percentages (percentages of most conserved aa) of gp41 residues involved in each secondary structural domain and then divided the sum by the total number of residues per domain.

Finally, we used WebLogo (https://weblogo.berkeley.edu/logo.cgi) (Crooks et al., 2004) to generate a figure showing the HIV-1 and HIV-1 group M gp41, as well as the HIV-2 gp36 consensus sequences, including in each protein position the most frequent aa present in the corresponding alignment. The HIV-1 group M gp41 consensus sequence was generated after the alignment of group M variants consensus sequences, the HIV-1 gp41 consensus sequence was generated after the alignment of the HIV-1 groups (M, N, O, P) consensus sequences, and the HIV-2 gp36 consensus sequence was generated after the alignment of the HIV-2 groups A, B, and the CRF01_AB consensus sequences. The aa letters were represented as large as their conservation percentage with a color code according to their side-chain: in black, non-polar aliphatic (glycine, G; alanine, A; valine, V; leucine, L; methionine, M; isoleucine, I); in dark blue-green, aromatic (phenylalanine, F; tyrosine, Y; tryptophan, W); in fuchsia, polar uncharged (serine, S; threonine, T; cysteine, C; proline, P; asparagine, N; glutamine; Q); in light-blue, positively charged (lysine, K; arginine, R; histidine, H); and in light-pink, negatively charged (aspartic acid, D; glutamic acid, E). Deletions were represented by a yellow "X".

HIV-1 Monoclonal Human Antibodies Location

We analyzed the overall conservation in each gp41 secondary-structure domain across the four HIV-1 groups and in the HIV-1 consensus sequence. We also identified the aa conservation level in each HIV-1 group and each anti-gp41 human Abs (bnAbs and non-bnAbs) binding domain described in the LANL HIV Immunology Database (Los Alamos HIV Molecular Immunology Database, 2021a), showing lineal epitopes in blue and non-linear epitopes recognized by bnAbs in orange.

Gp41/Gp36 Natural Polymorphisms and V-Markers Across HIV Variants

We described the gp41/gp36 natural polymorphisms and V-markers in HIV-1 and HIV-2 variants using the EpiMolBio program. To identify the aa changes present in ≥90% sequences in each HIV-1 or HIV-2 variant (natural polymorphisms), we compared all gp41 sequences with the HXB2 isolate, HIV-1 consensus, and HIV-1 group M consensus sequences, and all gp36 sequences with BEN isolate, and the HIV-2 consensus sequence. Among the natural polymorphisms found, we identified the exclusive V-markers of each HIV-1 or HIV-2

variant, not present in any other HIV variant. The color code for that analysis was light-blue (≥90–<100%) and dark blue-green (100%) for HIV-1 and light-pink (≥90–<100%) and fuchsia (100%) for HIV-2.

We also studied the presence of major T-20 resistance mutations according to the 2019 edition of the International Antiviral Society–USA (2019 IAS-USA) drug resistance mutations list (Wensing et al., 2019) among the V-markers found. Moreover, we looked for L44M change due to association with 1.8-fold resistance to T-20 *in vitro* (Mink et al., 2005).

Furthermore, we examined the presence of natural polymorphisms and specific V-markers on four gp41 positions (N160, W161, F162, and W169), described as key for HIV-1 neutralization by 10E8 (Huang et al., 2012; Kwon et al., 2016), which is a highly potent bnAb-recognizing gp41 MPER (epitope NWFDISNWLWYIK, gp41 positions 160-172) (Los Alamos HIV Molecular Immunology Database, 2021b). Other key reasons to study 10E8 bnAbs were that it was not detected by some serological diagnostic tests targeting gp41 (Smith et al., 2021) and was recently used to design new strategies for the development of a more efficient HIV-1 vaccine (Kuchar et al., 2021).

Finally, some gp41 changes (S23P, S23A, T25A, T27A, and I48P) affecting infectivity, gp41 structure or function (Alsahafi et al., 2015; Lu et al., 2019), were also studied across HIV-1 variants.

RESULTS

Analyzed Gp41/Gp36 Sequences and Inferred Consensus Sequences

We downloaded all 18,348 HIV transmembrane glycoprotein sequences from the LANL database. Once sequences with stop codons in unusual positions were discarded and after excluding variants with <3 sequences (except HIV-1 group P), a total of 17,078 gp41/gp36 sequences from 108 variants, including types, groups, subtypes, sub-subtypes, and CRF, were finally used in this study: 16,803 gp41 (HIV-1, 105 variants) and 275 gp36 (HIV-2, three variants) (**Table 1**). Among the HIV-1 gp41 sequences, 99 belonged to non-M groups (N, O, and P) and 16,704 were ascribed to group M (nine subtypes, six sub-subtypes, and 87 CRF). The gp41 sequences from group M sub-subtype A5, subtype F, CRF30_0206, CRF84_A1D, CRF91_01C, CRF94_cpx, CRF97_01B, CRF101_01B, and CRF102_0107 were not available in LANL. Gp36 sequences from groups E, H, and I were also absent. The variants with the highest representation in HIV-1 group M were subtype B (48.5%), subtype C (23.9%), and recombinant CRF01_AE (13.1%). In HIV-2, the most represented group was A (85.5%).

Consensus sequences at aa level were inferred by EpiMolBio for HIV-1 group M (**Figure 2A**), HIV-1 (**Figure 2B**), and HIV-2 (**Figure 2C**) to study the homology at aa level across variants. The gp41 HIV-1 consensus sequence was generated after aligning the four HIV-1 groups (M, N, O, P) consensus sequences and the HIV-1 group M consensus after aligning 102 group M variants with at least three sequences. The gp36 HIV-2

TABLE 1 | Gp41/gp36 LANL sequences were analyzed in this study.

Variants				N° SEQS
HIV-1	Non-M Groups		N	11
			O	86
			P*	2
HIV-1	Group M	Subtypes	A	20
			A1	752
			A2	8
			A3	3
			A4	2
			A5	0
			A6	167
			B	8106
			C	3985
			D	183
			F	0
			F1	122
			F2	15
			G	153
			H	11
			J	8
			K	3
			L	3
HIV-1	Group M	CRF	CRF01_AE	2186
			CRF02_AG	291
			CRF03_AB	5
			CRF04_cpx	8
			CRF05_DF	4
			CRF06_cpx	14
			CRF07_BC	126
			CRF08_BC	62
			CRF09_cpx	5
			CRF10_CD	3
			CRF11_cpx	26
			CRF12_BF	18
			CRF13_cpx	10
			CRF14_BG	12
			CRF15_01B	9
			CRF16_A2D	4
			CRF17_BF	6
			CRF18_cpx	5
			CRF19_cpx	5
			CRF20_BG	4
			CRF21_A2D	4
			CRF22_01A1	13
			CRF23_BG	2
			CRF24_BG	4
			CRF25_cpx	5
			CRF26_A5U	5
			CRF27_cpx	4
			CRF28_BF	5
			CRF29_BF	7
			CRF30_0206	0

(Continued)

TABLE 1 | Continued

Variants				N° SEQS
HIV-1	Group M	CRF	CRF31_BC	3
			CRF32_06A6	3
			CRF33_01B	7
			CRF34_01B	3
			CRF35_AD	21
			CRF36_cpx	3
			CRF37_cpx	4
			CRF38_BF	1
			CRF39_BF	3
			CRF40_BF	4
			CRF41_CD	3
			CRF42_BF	13
			CRF43_02G	5
			CRF44_BF	3
			CRF45_cpx	5
			CRF46_BF	8
			CRF47_BF	3
			CRF48_01B	3
			CRF49_cpx	5
			CRF50_A1D	4
			CRF51_01B	7
			CRF52_01B	3
			CRF53_01B	4
			CRF54_01B	3
			CRF55_01B	9
			CRF56_cpx	4
			CRF57_BC	7
			CRF58_01B	6
			CRF59_01B	8
			CRF60_BC	5
			CRF61_BC	3
			CRF62_BC	3
			CRF63_02A	15
			CRF64_BC	8
			CRF65_cpx	6
			CRF66_BF	3
			CRF67_01B	2
			CRF68_01B	3
			CRF69_01B	7
			CRF70_BF	3
			CRF71_BF	15
			CRF72_BF	5
			CRF73_BG	2
			CRF74_01B	3
			CRF75_BF	3
			CRF76_01B	2
			CRF77_cpx	4
			CRF78_cpx	3
			CRF79_0107	3
			CRF80_0107	2
			CRF81_cpx	2

(Continued)

TABLE 1 | Continued

Variants				N° SEQS
HIV-1	Group M	CRF	CRF82_cpx	6
			CRF83_cpx	11
			CRF84_A1D	0
			CRF85_BC	11
			CRF86_BC	3
			CRF87_cpx	3
			CRF88_BC	3
			CRF89_BF	3
			CRF90_BF1	6
			CRF91_01C	0
			CRF92_C2U	5
			CRF93_cpx	3
			CRF94_cpx	0
			CRF95_02B	5
			CRF96_cpx	3
			CRF97_01B	0
			CRF98_06B	1
			CRF99_BF	2
			CRF100_01C	3
			CRF101_01B	0
			CRF102_0107	0
			CRF103_01B	4
HIV-2	Groups		A	235
			B	34
			C	1
			D	1
			E	0
			F	2
			G	1
			H	0
			I	0
HIV-2	CRF		CRF01_AB	6

In red, variants with <3 sequences. Only HIV-1 group P, with 2 sequences, was included in the study (with an asterisk). N°, number; SEQS, sequences; CRF, circulating recombinant forms.

consensus sequence was inferred after aligning the 275 gp36 LANL sequences ascribed to groups A, B, and CRF01_AB. **Supplementary Tables 1**, **2** show the inferred HIV-1 and HIV-2 transmembrane consensus sequences in this study, respectively.

Amino Acid Conservation of Gp41 and Gp36 Across HIV Variants

We identified the gp41 and gp36 residues with ≥90% and 100% conservation across sequences from each analyzed variant (**Table 2**). The gp41 HIV-1 consensus aa sequence had 38.3% of the 345 gp41 residues conserved in ≥90% of gp41 sequences, and 9% of aa were 100% conserved. Higher conservation of gp36 HIV-2 aa consensus sequence was observed compared to HIV-1 gp41 aa consensus sequence, with 59.7% of 350 gp36 residues conserved in ≥90% of gp36

FIGURE 2 | Inferred consensus transmembrane sequences for HIV-1 group M **(A)**, HIV-1 **(B)**, and HIV-2 **(C)**. Each residue includes the most frequent aa present in the corresponding alignment. The aa code letters were represented in proportion to their percentage of conservation. Color code according to their side-chain: non-polar aliphatic (glycine, G; alanine, A; valine, V; leucine, L; methionine, M; isoleucine, I) in black; aromatic (phenylalanine, F; tyrosine, Y; tryptophan, W) in dark blue-green; polar uncharged (serine, S; threonine, T; cysteine, C; proline, P; asparagine, N; glutamine, Q) in fuchsia; positively charged (lysine, K; arginine, R; histidine, H) in light-blue, and negatively charged (aspartic acid, D; glutamic acid, E) in light-pink. Deletions were represented by "X" in yellow.

sequences, and 12.6% of aa totally conserved. Thus, 4 out of 10 aa positions were ≥90% conserved in the HIV-1 gp41 consensus sequence, rising to 6 out of 10 in the HIV-2 gp36 consensus sequence.

For HIV-1 groups, the highest percentage of conserved aa in ≥90%/100% of their HIV-1 gp41 LANL sequences were found in groups P (88.7%/88.7%) and N (85.8%/73%), followed by O (62.3%/33.6%) and M (60.3%/7.5%). Therefore, group P was the HIV-1 non-M group with the highest number of highly conserved gp41 residues, while group O had the fewest.

Among the 16,704 group M gp41 sequences, CRF46_BF had the lowest number of aa present in ≥90% of sequences (64.1%, eight sequences) and CRF41_CD the highest (98.8%, three sequences). When considering the completely conserved (100%) residues, subtype C showed the lowest number (3.2%, 3,985 sequences) and CRF41_CD the highest (98.8%, three sequences). Thus, among group M variants, CRF41_CD was the most conserved variant, while subtype C was the least conserved.

Regarding HIV-2, the variant with the highest percentage of ≥90%/100% conserved aa in gp36 protein was CRF01_AB

TABLE 2 | Number and percentage of highly (≥90%, 100%) conserved aa per HIV-1/HIV-2 variant.

Variants			SEQS	N° AA	%	N° AA	%
				≥90%		100%	
HIV-1	HIV-1 Consensus		4	132	38.3	31	9
	Non-M Groups	N	11	296	85.8	252	73
		O	86	215	62.3	116	33.6
		P	2	306	88.7	306	88.7
HIV-1	Group M Consensus		102	208	60.3	26	7.5
	Group M Subtypes	A	20	272	78.8	188	54.5
		A1	752	241	69.9	60	17.4
		A2	8	236	68.4	236	68.4
		A3	3	296	85.8	296	85.8
		A6	167	287	83.2	112	32.5
		B	8106	247	71.6	22	6.4
		C	3985	240	69.6	11	3.2
		D	183	257	74.5	58	16.8
		F1	122	263	76.2	95	27.5
		F2	15	241	69.9	199	57.7
		G	153	247	71.6	90	26.1
		H	11	274	79.4	204	59.1
		J	8	241	69.9	241	69.9
		K	3	289	83.8	289	83.8
		L	3	276	80	276	80
HIV-1	Group M CRF	CRF01_AE	2186	260	75.4	43	12.5
		CRF02_AG	291	260	75.4	66	19.1
		CRF03_AB	5	311	90.1	311	90.1
		CRF04_cpx	8	249	72.2	249	72.2
		CRF05_DF	4	270	78.3	270	78.3
		CRF06_cpx	14	255	73.9	204	59.1
		CRF07_BC	126	286	82.9	142	41.2
		CRF08_BC	62	281	81.4	173	50.1
		CRF09_cpx	5	290	84.1	290	84.1
		CRF10_CD	3	291	84.3	291	84.3
		CRF11_cpx	26	246	71.3	170	49.3
		CRF12_BF	18	245	71	185	53.6
		CRF13_cpx	10	285	82.6	236	68.4
		CRF14_BG	12	306	88.7	268	77.7
		CRF15_01B	9	243	70.4	243	70.4
		CRF16_A2D	4	273	79.1	273	79.1
		CRF17_BF	6	246	71.3	246	71.3
		CRF18_cpx	5	270	78.3	270	78.3
		CRF19_cpx	5	271	78.6	271	78.6
		CRF20_BG	4	291	84.3	291	84.3
		CRF21_A2D	4	272	78.8	272	78.8
		CRF22_01A1	13	266	77.1	202	58.6
		CRF24_BG	4	299	86.7	299	86.7
		CRF25_cpx	5	270	78.3	270	78.3
		CRF26_A5U	5	257	74.5	257	74.5
		CRF27_cpx	4	254	73.6	254	73.6
		CRF28_BF	5	268	77.7	268	77.7

(Continued)

TABLE 2 | Continued

Variants	SEQS	N° AA	%	N° AA	%
		≥90%		100%	
CRF29_BF	7	236	68.4	236	68.4
CRF31_BC	3	291	84.3	291	84.3
CRF32_06A6	3	321	93	321	93
CRF33_01B	7	266	77.1	266	77.1
CRF34_01B	3	327	94.8	327	94.8
CRF35_AD	21	289	83.8	219	63.5
CRF36_cpx	3	304	88.1	304	88.1
CRF37_cpx	4	277	80.3	277	80.3
CRF39_BF	3	283	82	283	82
CRF40_BF	4	276	80	276	80
CRF41_CD	3	341	98.8	341	98.8
CRF42_BF	13	322	93.3	301	87.2
CRF43_02G	5	277	80.3	277	80.3
CRF44_BF	3	286	82.9	286	82.9
CRF45_cpx	5	265	76.8	265	76.8
CRF46_BF	8	221	64.1	221	64.1
CRF47_BF	3	295	85.5	295	85.5
CRF48_01B	3	299	86.7	299	86.7
CRF49_cpx	5	268	77.7	268	77.7
CRF50_A1D	4	298	86.4	298	86.4
CRF51_01B	7	286	82.9	286	82.9
CRF52_01B	3	302	87.5	302	87.5
CRF53_01B	4	288	83.5	288	83.5
CRF54_01B	3	310	89.9	310	89.9
CRF55_01B	9	282	81.7	282	81.7
CRF56_cpx	4	338	98	338	98
CRF57_BC	7	247	71.6	247	71.6
CRF58_01B	6	294	85.2	294	85.2
CRF59_01B	8	295	85.5	295	85.5
CRF60_BC	5	288	83.5	288	83.5
CRF61_BC	3	330	95.7	330	95.7
CRF62_BC	3	322	93.3	322	93.3
CRF63_02A	15	302	87.5	259	75.1
CRF64_BC	8	252	73	252	73
CRF65_cpx	6	294	85.2	294	85.2
CRF66_BF	3	282	81.7	282	81.7
CRF68_01B	3	320	92.8	320	92.8
CRF69_01B	7	280	81.2	280	81.2
CRF70_BF	3	286	82.9	286	82.9
CRF71_BF	15	253	73.3	212	61.4
CRF72_BF	5	257	74.5	257	74.5
CRF74_01B	3	300	87	300	87
CRF75_BF	3	316	91.6	316	91.6
CRF77_cpx	4	304	88.1	304	88.1
CRF78_cpx	3	279	80.9	279	80.9
CRF79_0107	3	318	92.2	318	92.2
CRF82_cpx	6	299	86.7	299	86.7
CRF83_cpx	11	319	92.5	298	86.4

(Continued)

TABLE 2 | Continued

Variants				SEQS	Conservation			
					≥90%		100%	
					N° AA	%	N° AA	%
HIV-1	Group M	CRF	CRF85_BC	11	308	89.3	251	72.8
			CRF86_BC	3	305	88.4	305	88.4
			CRF87_cpx	3	312	90.4	312	90.4
			CRF88_BC	3	303	87.8	303	87.8
			CRF89_BF	3	290	84.1	290	84.1
			CRF90_BF1	6	258	74.8	258	74.8
			CRF92_C2U	5	234	67.8	234	67.8
			CRF93_cpx	3	277	80.3	277	80.3
			CRF95_02B	5	294	85.2	294	85.2
			CRF96_cpx	3	305	88.4	305	88.4
			CRF100_01C	3	293	84.9	293	84.9
			CRF103_01B	4	302	87.5	302	87.5
HIV-2		HIV-2 Consensus		3	209	59.7	44	12.6
	Groups	A		235	231	66	55	15.7
		B		34	251	71.7	159	45.4
	CRF	CRF01_AB		6	290	82.9	290	82.9

Data showed the number of aa conserved in ≥90 or 100% of sequences per analyzed variant. Gp41 (HIV-1), 345 aa; gp36 (HIV-2), 350 aa. Gp41 HIV-1 consensus sequence was generated after the alignment of the four HIV-1 groups (M, N, O, P) consensus sequences, the HIV-1 group M consensus after the alignment of 102 group M variants with ≥3 sequences, and the HIV-2 consensus after the alignment of three HIV-2 variants. N°, number; SEQS, sequences; AA, amino acids; %, aa percentage; CRF, circulating recombinant forms.

(82.9%/82.9%), followed by group B (71.7%/45.4%) and group A (66%/15.7%).

Gp41 Conservation in Each HIV-1 Group per Structural Domain and Anti-Gp41 Antibody Binding Sites

When we studied the level of conserved aa in each gp41 structural domain, we observed that gp41 conservation differed between HIV-1 groups and structural domains (**Table 3**), ranging from 66.4% to 98.9% conservation. The gp41 domain in the HIV-1 consensus sequence with the highest conservation was NHR (86.2%), followed by FPPR (84.9%), MPER (84.2%), and FP (84%). We observed a high conservation percentage in FPPR (96%) and NHR (95.4%) domains in the group M consensus sequence, despite the lower overall gp41 conservation *vs.* the other three HIV-1 groups. Group N gp41 presented the most conserved FP (98.9%) and FPPR (98.8%) domains, group O the best conserved MPER domain (95.7%), and group P the most conserved NHR (98.7%), TM (95.5%), IL (95.3%), and CT (93.7%) gp41 domains (**Table 3**). CHR domain presented the highest conservation (95.0%) in groups N and P. Considering all analyzed HIV-1 variants, the FP, FPPR, NHR, and MPER were the gp41 domains with the highest (≥84%) number of conserved aa, being the target for most anti-gp41 Abs. The least conserved

domains in the HIV-1 consensus sequence were TM (69%) and CT (66.4%) located inside the virus.

The conservation level in each domain for human anti-gp41 bnAbs and non-bnAbs across HIV-1 groups was also described (**Figure 3**), located between the FP and MPER domains since TM and CT (the less conserved gp41 domains according to our data) are not exposed. The bnAbs are directed to CHR (2F5) and mainly to the MPER domain (2F5, Z13, 4E10, 10E8, and derivatives). No anti-gp41 Ab-binding sites were shown in LANL in the first 13 aa of the FP domain, in the 17 first positions of NHR, in the last 9 residues of IL, and the first 14 aa of CHR. The same was true in the target peptide to the fusion inhibitor T-20 (GIVQQQNNLL, NHR, residues 36-45), even though it presented high conservation across HIV-1 groups (**Figure 3**).

Natural Polymorphisms and V-Markers

All-natural polymorphisms (aa present in ≥90% sequences) found in each HIV variant are shown in **Supplementary Table 3** (HIV-1) and **Supplementary Table 4** (HIV-2). The number of polymorphisms in gp41 across HIV-1 non-B variants (different from subtype B) increased when using the HXB2 subtype B sequence as reference. HIV-2 group A presented fewer polymorphisms in gp36 when HIV-2 BEN subtype A group was used as reference. **Figure 4** shows the natural polymorphisms that could be considered as exclusive V-markers for HIV-1 gp41 non-M groups (86 V-markers, **Figure 4A**), for HIV-1 group M gp41 variants (120 V-markers, **Figure 4B**), and HIV-2 gp36 variants (24 V-markers, **Figure 4C**). The specific V-markers per variant can be found in **Supplementary Tables 3, 4**.

No V-markers associated with T-20 major resistance were found in HIV-1 groups N, O, P, group M, and HIV-2 variants. Only M44 was present in the group P consensus gp41 sequence (**Supplementary Tables 1, 3**). Group O gp41 consensus sequence carried D42 V-marker, but no T42, a residue associated with high T-20 resistance.

After the analysis of the four gp41 key positions (N160, W161, F162, and W169) for HIV-1 neutralization by the anti-MPER bnAb 10E8, we observed high (90–100%) conservation of W161 and W169 in the transmembrane consensus sequences across HIV-1 variants. L162 appeared in group O and L/M162 in group P gp41 consensus sequence. Residue 160 showed the highest variability, carrying the group P and many HIV-1 group M variants another aa, mainly serine.

Finally, regarding changes in gp41 residues (A23, P23, A25, A27, and P48) affecting infectivity, gp41 structure, or function, we observed that A23 appeared in groups P and O and in CRF63_02A gp41 consensus sequences, while A25 appeared in group O, subtype L, CRF41_CD, and CRF47_BF group M variants. V48 appeared in the CRF90_BF1 gp41 consensus sequence but P48 did not.

DISCUSSION

This is the most up-to-date descriptive study related to HIV-1/HIV-2 transmembrane envelope proteins, providing the conservation level and the V-markers in each HIV variant, identifying the conserved gp41 domains in each HIV-1 group

TABLE 3 | Percentage of aa conservation in each gp41 structural domain across HIV-1 groups and HIV-1 consensus sequences.

Gp41 structural domains	% AA Conservation HIV-1 GP41								
	FP	FPPR	NHR	IL	CHR	MPER	TM	CT	GP41
Gp41 residues	1–16	17–32	33–70	71–113	114–153	154–172	173–194	195–345	1–345
Group N	**98.9**	98.8	96.5	95.0	95.0	90.9	94.6	93.5	94.6
Group O	**97.7**	87.6	95.3	79.2	83.8	95.7	89.0	86.0	87.3
Group P	93.8	90.6	98.7	95.3	95.0	89.5	95.5	93.7	94.3
Cons. Group M	86.7	**96.0**	95.4	86.6	79.6	86.1	91.9	81.0	85.1
Cons. HIV-1	84.0	84.9	**86.2**	73.3	70.7	84.2	69.0	66.4	72.8

For the analysis, we used the aa conservation percentages per residue (percentages of most conserved aa) of each HIV-1 group and HIV-1 consensus sequence provided in **Supplementary Table 1** *with color code. After identifying the residues of each secondary structural domain, the sum of the aa conservation percentages of each of them was divided by the total number of residues per domain. In color, the highest conservation level in each structural domain. In bold, the most conserved structural domain in each HIV-1 group and HIV-1 consensus sequence. Gp41 residues according to HXB2 HIV-1 isolate. N°, number; AA, amino acids; Cons, aa consensus sequence; FP, fusion peptide; FPPR, fusion peptide proximal region; NHR, N-terminal alpha-helical region; IL, immune-dominant linker; CHR, C-terminal heptad repeat region; MPER, membrane-proximal external region; TM, transmembrane region; CT, C-terminal domain.*

per structural domain and per anti-gp41 antibody binding site in the most extensive panel of HIV-1 gp41/HIV-2 gp36 sequences (n = 17,078) and variants (105 HIV-1 variants, three HIV-2 variants) analyzed to date. Our in-house bioinformatics tool was also used to establish consensus sequences for gp41/gp36 proteins to study the aa level conservation across HIV variants.

The results have shown that the degree of conservation of the protein can differ across HIV variants and transmembrane structural domains, as our group previously described with fewer sequences and variants (Holguín et al., 2007). The higher level of conserved residues across variants in gp36 (275 sequences) vs. gp41 (16,803 sequences) could be explained by the lower number of gp36 sequences in LANL due to the lower HIV-2 prevalence and worldwide infections (1–2 of 38 million HIV infections) explained by its lower transmissibility and virulence vs. HIV-1 (Azevedo-Pereira and Santos-Costa, 2016; Clinical Info and gov, 2019; Kapoor and Padival, 2021). A higher level of fully conserved aa observed in HIV-1 non-M groups (99 sequences) vs. group M (16,704 sequences) could also be explained by the lower prevalence of groups N, O, and P in the pandemic (Mourez et al., 2013). The same happens in most CRF, with few available gp41 sequences in LANL, due to their low prevalence and the absence of gp41 sequencing in countries where they circulate, with sequencing or research not always available.

All gp41 secondary structure ectodomains presented >70% of conservation (HIV-1 consensus sequence), which supports their key role in the viral cycle and the importance of structure maintenance for virus-cell membranes fusion and viral entry (Pancera et al., 2014). Our data revealed that the best-conserved gp41 domains were N-HR (86.2%), FPPR (84.9%), MPER (84.2%), and FP (84%). It is essential to highlight that precisely NHR, FP, and FPPR gp41 domains have recently been implicated in the interaction with the fusion inhibitor T-20, which prevents the virus from entering the cell (Xu et al., 2019).

Regarding the gp41-CT region (aa 195-345), located inside the virus, it was the least conserved domain in the HIV-1 and group

M gp41 consensus sequences, despite its important role in gp41 structure and function (Fernandez and Freed, 2018). CT domain is involved in envelope conformation (Castillo-Menendez et al., 2018), being essential for efficient envelope incorporation into budding HIV-1 particles (Murakami and Freed, 2000), and requiring interaction with gag matrix protein (MA) (Wyma et al., 2000; Eastep et al., 2021). However, the domains in each protein involved in this interaction are still unknown (Fernandez and Freed, 2018). A link between the matrix trimers' formation and the binding between MA and gp41 CT has also been reported (Alfadhli et al., 2016).

The critical role of the FPPR or polar region at the N terminus of gp41 for HIV-1 fusion and infectivity by stabilizing envelope trimers (Lu et al., 2019) could explain the high gp41 aa conservation percentage found in this domain. S23 within FPPR is structurally essential for maintaining HIV-1 envelope trimer, viral fusogenicity, and infectivity (Lu et al., 2019). Single or combined mutations S23P, T25A, and T27A in the FPPR region abolished or significantly decreased HIV-1 infectivity without affecting viral production, and S23A change significantly reduced HIV-1 infectivity and fusogenicity but not envelope expression and cleavage (Lu et al., 2019). In our study, only S23A and T25A substitutions appeared in a low number of variants with low prevalence in the pandemic. The absence of I48P gp41 change in our sequence set could be explained by its high impact on the quaternary conformation and function of the envelope glycoprotein trimer (Alsahafi et al., 2015).

Analyzing the genetic variability of HIV glycoprotein transmembrane within its immunodominant epitopes is important for understanding its possible impact on HIV Abs detection (Dorn et al., 2000; Dong et al., 2005; Smith et al., 2021). HIV-1 bnAbs can neutralize most HIV-1 strains from diverse genetic and geographic backgrounds (Binley et al., 2004; Wang and Zhang, 2020). The anti-gp41 bnAbs can recognize the MPER (Huang et al., 2012) and FP domains (Yuan et al., 2019), as well as the gp120/gp41 interphase (Huang et al., 2014; Scharf et al., 2014; Wang and Zhang, 2020). The conformational plasticity of FP could facilitate the recognition
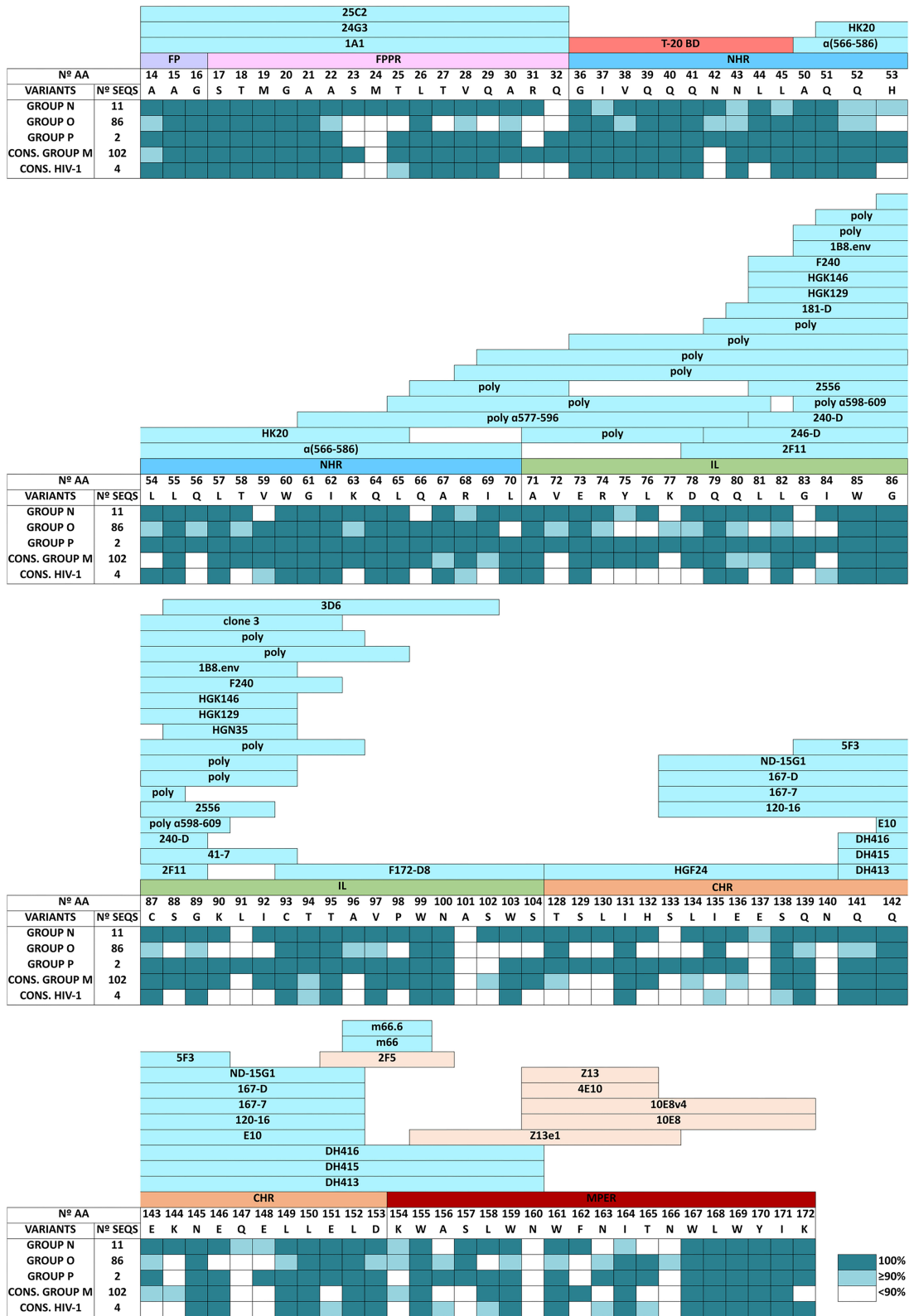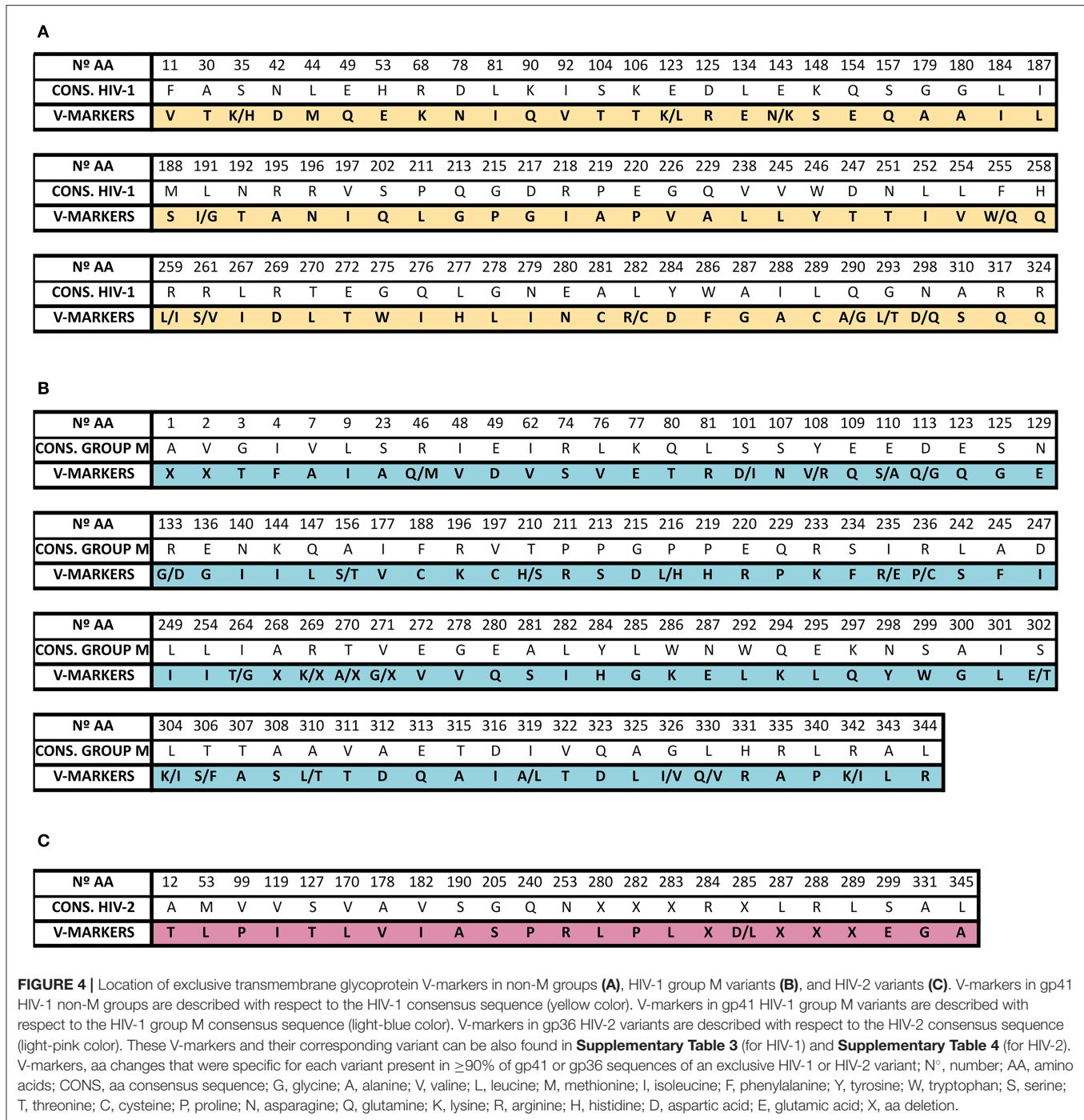
FIGURE 3 | Amino acid conservation level in gp41 residues involved in anti-gp41 human antibody binding domains described in LANL HIV Immunology Database across HIV-1 groups and HIV-1 consensus sequences. Data showed the percentage of the most conserved aa with the following conservation color code: white (aa

(Continued)

FIGURE 3 | conserved <90%), light-blue (aa conserved ≥90%–<100%), and dark blue-green (aa conserved 100%). Anti-gp41 human antibodies include bnAbs and non-bnAbs described in Los Alamos HIV Immune Database (Los Alamos HIV Molecular Immunology Database, 2021a). Linear epitopes are shown in blue and non-linear epitopes, recognized by broadly neutralizing antibodies (bnAbs), are shown in orange. We also indicated the T-20 fusion inhibitor binding domain in red (T-20BD). Nº, number; AA, amino acids; SEQS, sequences; Cons, aa consensus sequence; FP, fusion peptide; FPPR, fusion peptide proximal region; NHR, N-terminal alpha-helical region; IL, immune-dominant linker; CHR, C-terminal heptad repeat region; MPER, membrane-proximal external region; TM, transmembrane region; CT, C-terminal domain; G, glycine; A, alanine; V, valine; L, leucine; M, methionine; I, isoleucine; F, phenylalanine; Y, tyrosine; W, tryptophan; S, serine; T, threonine; C, cysteine; P, proline; N, asparagine; Q, glutamine; K, lysine; R, arginine; H, histidine; D, aspartic acid; E, glutamic acid; poly, polyclonal.

A

Nº AA	11	30	35	42	44	49	53	68	78	81	90	92	104	106	123	125	134	143	148	154	157	179	180	184	187
CONS. HIV-1	F	A	S	N	L	E	H	R	D	L	K	I	S	K	E	D	L	E	K	Q	S	G	G	L	I
V-MARKERS	V	T	K/H	D	M	Q	E	K	N	I	Q	V	T	T	K/L	R	E	N/K	S	E	Q	A	A	I	L

Nº AA	188	191	192	195	196	197	202	211	213	215	217	218	219	220	226	229	238	245	246	247	251	252	254	255	258
CONS. HIV-1	M	L	N	R	R	V	S	P	Q	G	D	R	P	E	G	Q	V	V	W	D	N	L	L	F	H
V-MARKERS	S	I/G	T	A	N	I	Q	L	G	P	G	I	A	P	V	A	L	L	Y	T	T	I	V	W/Q	Q

Nº AA	259	261	267	269	270	272	275	276	277	278	279	280	281	282	284	286	287	288	289	290	293	298	310	317	324
CONS. HIV-1	R	R	L	R	T	E	G	Q	L	G	N	E	A	L	Y	W	A	I	L	Q	G	N	A	R	R
V-MARKERS	L/I	S/V	I	D	L	T	W	I	H	L	I	N	C	R/C	D	F	G	A	C	A/G	L/T	D/Q	S	Q	Q

B

Nº AA	1	2	3	4	7	9	23	46	48	49	62	74	76	77	80	81	101	107	108	109	110	113	123	125	129
CONS. GROUP M	A	V	G	I	V	L	S	R	I	E	I	R	L	K	Q	L	S	S	Y	E	E	D	E	S	N
V-MARKERS	X	X	T	F	A	I	A	Q/M	V	D	V	S	V	E	T	R	D/I	N	V/R	Q	S/A	Q/G	Q	G	E

Nº AA	133	136	140	144	147	156	177	188	196	197	210	211	213	215	216	219	220	229	233	234	235	236	242	245	247
CONS. GROUP M	R	E	N	K	Q	A	I	F	R	V	T	P	P	G	P	P	E	Q	R	S	I	R	L	A	D
V-MARKERS	G/D	G	I	I	L	S/T	V	C	K	C	H/S	R	S	D	L/H	H	R	P	K	F	R/E	P/C	S	F	I

Nº AA	249	254	264	268	269	270	271	272	278	280	281	282	284	285	286	287	292	294	295	297	298	299	300	301	302
CONS. GROUP M	L	L	I	A	R	T	V	E	G	E	A	L	Y	L	W	N	W	Q	E	K	N	S	A	I	S
V-MARKERS	I	I	T/G	X	K/X	A/X	G/X	V	V	Q	S	I	H	G	K	E	L	K	L	Q	Y	W	G	L	E/T

Nº AA	304	306	307	308	310	311	312	313	315	316	319	322	323	325	326	330	331	335	340	342	343	344
CONS. GROUP M	L	T	T	A	A	V	A	E	T	D	I	V	Q	A	G	L	H	R	L	R	A	L
V-MARKERS	K/I	S/F	A	S	L/T	T	D	Q	A	I	A/L	T	D	L	I/V	Q/V	R	A	P	K/I	L	R

C

Nº AA	12	53	99	119	127	170	178	182	190	205	240	253	280	282	283	284	285	287	288	289	299	331	345
CONS. HIV-2	A	M	V	V	S	V	A	V	S	G	Q	N	X	X	X	R	X	L	R	L	S	A	L
V-MARKERS	T	L	P	I	T	L	V	I	A	S	P	R	L	P	L	X	D/L	X	X	X	E	G	A

FIGURE 4 | Location of exclusive transmembrane glycoprotein V-markers in non-M groups **(A)**, HIV-1 group M variants **(B)**, and HIV-2 variants **(C)**. V-markers in gp41 HIV-1 non-M groups are described with respect to the HIV-1 consensus sequence (yellow color). V-markers in gp41 HIV-1 group M variants are described with respect to the HIV-1 group M consensus sequence (light-blue color). V-markers in gp36 HIV-2 variants are described with respect to the HIV-2 consensus sequence (light-pink color). These V-markers and their corresponding variant can be also found in **Supplementary Table 3** (for HIV-1) and **Supplementary Table 4** (for HIV-2). V-markers, aa changes that were specific for each variant present in ≥90% of gp41 or gp36 sequences of an exclusive HIV-1 or HIV-2 variant; Nº, number; AA, amino acids; CONS, aa consensus sequence; G, glycine; A, alanine; V, valine; L, leucine; M, methionine; I, isoleucine; F, phenylalanine; Y, tyrosine; W, tryptophan; S, serine; T, threonine; C, cysteine; P, proline; N, asparagine; Q, glutamine; K, lysine; R, arginine; H, histidine; D, aspartic acid; E, glutamic acid; X, aa deletion.

of the virus by bnAbs (Yuan et al., 2019). It is also known that HIV-1 variability can impact on bnAbs reactivity in HIV diagnostic tests targeting gp41, leading to non-reactive results with different serological diagnostic assays (Smith et al., 2021). Furthermore, the identification of HIV antibody binding domains is critical for vaccine development studies (Kuchar et al., 2021). Many bnAbs directed to gp41 have been described (Los Alamos HIV Molecular Immunology Database, 2021a), and the identification of the aa conservation level across HIV variants on the recognized epitopes by each bnAb and in those key gp41 residues for viral neutralization is of particular interest. The high variability found in key gp41 residue 160 across some HIV-1 variants could explain the previously reported failure of 10E8 recognition by some different serological diagnostic tests (Smith et al., 2021). We could not analyze the gp41 aa conservation in the gp41 target sequence per each failing diagnostic test because manufacturers do not provide detailed information regarding which part of the gp41 sequence was targeted in their HIV diagnostic assays detecting the transmembrane protein, which can also differ across assays. The provided information in the **Supplementary Tables** of the manuscript can help manufacturers and other researchers design new gp41-based molecular and serological diagnostic tests to identify those HIV-1 variants whose diagnosis could be compromised by viral genetic variability.

Natural infection by HIV-2 also leads to the elicitation of high titers of bnAbs against primary HIV-2 strains (De Silva et al., 2012; Kong et al., 2012; Özkaya Sahin et al., 2012), although not all bnAbs to HIV-2 neutralize HIV-1 variants (Björling et al., 1993). In fact, MPER-specific Abs induced by vaccination with recombinant gp36 proteins in rats did not neutralize HIV-2 (Behrendt et al., 2012). It is known that HIV types present different mechanisms for the processing of envelope glycoproteins from a smaller env precursor in HIV-2 (gp140) than in HIV-1 (gp160) (Rey et al., 1989). To date, unfortunately, the exact aa residues of each secondary structure in HIV-2 gp36 have not been specified. However, some conserved gp36 epitopes have been reported (Jadhav et al., 2011), as well as immunogenic sites, antibody binding sites in the TM and IL region of HIV-2 transmembrane gp36 (Chiodi et al., 1993), and other epitopes recognized by bnAbs in HIV-2 envelope gp140 (Kong et al., 2012).

Previous studies showed the importance of consensus sequence establishment to guide vaccine development (Ellenberger et al., 2002; Sliepen et al., 2019). For the first time, our study also provides the aa consensus sequence of the transmembrane glycoprotein in each HIV variant (type, group, subtype, sub-subtype, and CRF). Moreover, we showed the conservation level across their sequences, which could be helpful to look for highly conserved peptides to direct new ARV, Abs, aptamers, probes, or primers to control or diagnose HIV infection regardless of the HIV variant. Furthermore, we showed the first identification of specific natural polymorphisms of gp41 and gp36 that can be considered as V-markers for all HIV-1 and HIV-2 variants, which should be considered in the new strategies for developing HIV-1 vaccines based

on epitopes recognized by bnAbs (Kuchar et al., 2021). The exclusive HIV V-markers identified in gp41/gp36 sequences could help in faster and preliminary HIV variant identification if required, before doing the phylogenetic study, the gold standard method for correct HIV variant characterization. New studies are required to evaluate the structural and biological impact of the different levels of aa conservation in gp41 across HIV-1 variants and the specific V-markers found in the viral transmembrane protein.

Long-term exposure to the first entry inhibitor T-20 induces drug-resistant mutations (Pérez-Alvarez et al., 2006). Interestingly, none of the variants had V-markers associated with major resistance to T-20, as we previously reported testing 79 different HIV variants from naïve patients (Holguín et al., 2007). Thus, no natural major resistance mutations to T-20 were observed. The L44M change found in group P consensus sequence, previously associated with 1.8-fold resistance to T-20 *in vitro* (Mink et al., 2005), was previously found in T-20 naïve subjects from China (Chang et al., 2021), maybe reflecting a resistance transmission during primoinfection. New HIV-1 fusion inhibitors are under development (Luque and Camarasa, 2021) and previous studies showed that optimized T-20 derivates could have been effective inhibitors of infection for multiple HIV-1 variants (Chen et al., 2019).

The main limitation of the study was the absence of LANL sequences from some HIV-1 (sub-subtype A5, subtype F, CRF30_0206, CRF84_A1D, CRF91_01C, CRF94_cpx, CRF97_01B, CRF101_01B, CRF102_0107) and HIV-2 variants (groups E, H, I), as well as the scarce number of HIV gp41/gp36 sequences in other 14 HIV variants with <3 sequences in LANL, which meant that they could not be included in the analysis (except group P).

The information provided in this manuscript aims to help other researchers studying the biological, therapeutic, diagnostic, or structural role of gp41 to identify the natural polymorphisms and specific V-markers per variant in each gp41/gp36 residue or epitope according to their interest. This study will also be useful for a more rational design of anti-gp41 drugs and vaccines and future HIV molecular diagnostic tests directed to transmembrane HIV protein.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Files**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

AV-A downloaded and analyzed the HIV LANL sequences under study, validated some EpiMolBio functions necessary for sequences analyses, performed the computations, discussed results, and wrote the first draft of the manuscript. RR developed the in-house EpiMolBio bioinformatics program, validated some EpiMolBio functions necessary for sequences analyses, discussed results, and reviewed the final version of the manuscript. ÁH

designed and supervised the study, reviewed and edited the manuscript, funding application, and project administration. All authors approved the final version submitted.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.855232/full#supplementary-material

REFERENCES

Alfadhli, A., Mack, A., Ritchie, C., Cylinder, I., Harper, L., Tedbury, P. R., et al. (2016). Trimer enhancement mutation effects on HIV-1 matrix protein binding activities. *J. Virol.* 90, 5657–5664. doi: 10.1128/JVI.00509-16

Alsahafi, N., Debbeche, O., Sodroski, J., and Finzi, A. (2015). Effects of the I559P gp41 change on the conformation and function of the human immunodeficiency virus (HIV-1) membrane envelope glycoprotein trimer. *PLoS ONE* 10:e0122111. doi: 10.1371/journal.pone.0122111

Arenas, M., Lorenzo-Redondo, R., and Lopez-Galindez, C. (2016). Influence of mutation and recombination on HIV-1 *in vitro* fitness recovery. *Mol. Phylogenet. Evol.* 94, 264–270. doi: 10.1016/j.ympev.2015.09.001

Azevedo-Pereira, J. M., and Santos-Costa, Q. (2016). HIV interaction with human host: HIV-2 as a model of a less virulent infection. *AIDS Rev.* 18, 44–53. Available online at: https://www.aidsreviews.com/resumen.php?id=1327&indice=2016181&u=unp

Behrendt, R., Fiebig, U., Kurth, R., and Denner, J. (2012). Induction of antibodies binding to the membrane proximal external region of gp36 of HIV-2. *Intervirology* 55, 252–256. doi: 10.1159/000324483

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235

Binley, J. M., Wrin, T., Korber, B., Zwick, M. B., Wang, M., Chappey, C., et al. (2004). Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *J. Virol.* 78, 13232–13252. doi: 10.1128/JVI.78.23.13232-13252.2004

Björling, E., Scarlatti, G., von Gegerfelt, A., Albert, J., Biberfeld, G., Chiodi, F., et al. (1993). Autologous neutralizing antibodies prevail in HIV-2 but not in HIV-1 infection. *Virology* 193, 528–530. doi: 10.1006/viro.1993.1160

Blumenthal, R., Durell, S., and Viard, M. (2012). HIV entry and envelope glycoprotein-mediated fusion. *J. Biol. Chem.* 287, 40841–40849. doi: 10.1074/jbc.R112.406272

Caillat, C., Guilligay, D., Sulbaran, G., and Weissenhorn, W. (2020). Neutralizing antibodies targeting HIV-1 gp41. *Viruses* 12:E1210. doi: 10.3390/v12111210

Castillo-Menendez, L. R., Witt, K., Espy, N., Princiotto, A., Madani, N., Pacheco, B., et al. (2018). Comparison of uncleaved and mature human immunodeficiency virus membrane envelope glycoprotein trimers. *J. Virol.* 92, e00277–e00218. doi: 10.1128/JVI.00277-18

Chang, L., Zhao, J., Guo, F., Ji, H., Zhang, L., Jiang, X., et al. (2021). HIV-1 gp41 genetic diversity and enfuvirtide resistance-associated mutations among enfuvirtide-naïve patients in southern China. *Virus Res.* 292:198215. doi: 10.1016/j.virusres.2020.198215

Chen, G., Cook, J. D., Ye, W., Lee, J. E., and Sidhu, S. S. (2019). Optimization of peptidic HIV-1 fusion inhibitor T20 by phage display. *Protein. Sci.* 28, 1501–1512. doi: 10.1002/pro.3669

Chiodi, F., Björling, E., Samuelsson, A., and Norrby, E. (1993). Antigenic and immunogenic sites of HIV-2 glycoproteins. *Chem. Immunol.* 56, 61–77. doi: 10.1159/000319156

Clinical Info, H. I. V., and gov. (2019). *HIV-2 Infection*. Available online at: https://clinicalinfo.hiv.gov/es/node/9288 (accessed September 15, 2021).

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004

De Leys, R., Vanderborght, B., Vanden Haesevelde, M., Heyndrickx, L., van Geel, A., Wauters, C., et al. (1990). Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of West-Central African origin. *J. Virol.* 64, 1207–1216. doi: 10.1128/jvi.64.3.1207-1216.1990

De Silva, T. I., Aasa-Chapman, M., Cotten, M., Hué, S., Robinson, J., Bibollet-Ruche, F., et al. (2012). Potent autologous and heterologous neutralizing antibody responses occur in HIV-2 infection across a broad range of infection outcomes. *J. Virol.* 86, 930–946. doi: 10.1128/JVI.06126-11

Dong, X.-N., Ying, J., Wu, Y., and Chen, Y.-H. (2005). Genetic variability of principal neutralizing determinants on HIV-1 gp41 and its correlation with subtypes. *Immunol. Lett.* 101, 104–107. doi: 10.1016/j.imlet.2005.04.013

Dorn, J., Masciotra, S., Yang, C., Downing, R., Biryahwaho, B., Mastro, T. D., et al. (2000). Analysis of genetic variability within the immunodominant epitopes of envelope gp41 from human immunodeficiency virus type 1 (HIV-1) group M and its impact on HIV-1 antibody detection. *J. Clin. Microbiol.* 38, 773–780. doi: 10.1128/JCM.38.2.773-780.2000

Eastep, G. N., Ghanam, R. H., Green, T. J., and Saad, J. S. (2021). Structural characterization of HIV-1 matrix mutants implicated in envelope incorporation. *J. Biol. Chem.* 296:100321. doi: 10.1016/j.jbc.2021.100321

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Ellenberger, D. L., Li, B., Lupo, L. D., Owen, S. M., Nkengasong, J., Kadio-Morokro, M. S., et al. (2002). Generation of a consensus sequence from prevalent and incident HIV-1 infections in West Africa to guide AIDS vaccine development. *Virology* 302, 155–163. doi: 10.1006/viro.2002.1577

Fernandez, M. V., and Freed, E. O. (2018). Meeting Review: 2018 International workshop on structure and function of the lentiviral gp41 cytoplasmic tail. *Viruses* 10:613. doi: 10.3390/v10110613

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., et al. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* 397, 436–441. doi: 10.1038/17130

Gao, F., Yue, L., White, A. T., Pappas, P. G., Barchue, J., Hanson, A. P., et al. (1992). Human infection by genetically diverse SIV SM-related HIV-2 in West Africa. *Nature* 358, 495–499. doi: 10.1038/358495a0

Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18, 182–192. doi: 10.1016/j.molmed.2011.12.001

Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Fleminger, I., Kirtley, S., et al. (2019). Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infect. Dis.* 19, 143–155. doi: 10.1016/S1473-3099(18)30647-9

Holguín, A., De Arellano, E. R., and Soriano, V. (2007). Amino acid conservation in the gp41 transmembrane protein and natural polymorphisms associated

with enfuvirtide resistance across HIV-1 variants. *AIDS Res. Hum. Retroviruses* 23, 1067–1074. doi: 10.1089/aid.2006.0256

Huang, J., Kang, B. H., Pancera, M., Lee, J. H., Tong, T., Feng, Y., et al. (2014). Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-120 interface. *Nature* 515, 138–142. doi: 10.1038/nature13601

Huang, J., Ofek, G., Laub, L., Louder, M. K., Doria-Rose, N. A., Longo, N. S., et al. (2012). Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* 491, 406–412. doi: 10.1038/nature11544

Jadhav, S., Tripathy, S., Kulkarni, S., Chaturbhuj, D., Ghare, R., Bhattacharya, J., et al. (2011). Genetic conservation in gp36 transmembrane sequences of Indian HIV type 2 isolates. *AIDS Res. Hum. Retroviruses* 27, 1337–1343. doi: 10.1089/aid.2011.0063

Kapoor, A. K., and Padival, S. (2021). *HIV-2. In StatPearls*. StatPearls Publishing. Available online at: http://www.ncbi.nlm.nih.gov/books/NBK572083/~ [Internet].

Kong, R., Li, H., Bibollet-Ruche, F., Decker, J. M., Zheng, N. N., Gottlieb, G. S., et al. (2012). Broad and potent neutralizing antibody responses elicited in natural HIV-2 infection. *J. Virol.* 86, 947–960. doi: 10.1128/JVI.06155-11

Kuchar, M., Kosztyu, P., Daniel Lišková, V., Cerný, J., Petroková, H., and Vróblová, E., et al. (2021). Myomedin scaffold variants targeted to 10E8 HIV-1 broadly neutralizing antibody mimic gp41 epitope and elicit HIV-1 virus-neutralizing sera in mice. *Virulence* 12, 1271–1287. doi: 10.1080/21505594.2021.1920251

Kwon, Y. D., Georgiev, I. S., Ofek, G., Zhang, B., Asokan, M., Bailer, R. T., et al. (2016). Optimization of the solubility of HIV-1-neutralizing antibody 10E8 through somatic variation and structure-based design. *J. Virol.* 90, 5899–5914. doi: 10.1128/JVI.03246-15

Lazzarin, A. (2005). Enfuvirtide: the first HIV fusion inhibitor. *Exp. Opin. Pharmacother.* 6, 453–464. doi: 10.1517/14656566.6.3.453

Leitner, T., Korber, B., Daniels, M., Calef, C., and Foley, B. (2005). HIV-1 subtype and circulating recombinant form (CRF) reference sequences, 2005. *HIV Seq. Compend.* 2005, 41–48. Available online at: https://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2005/partI/leitner.pdf

Li, K., Xiu, C. L., Gao, L. M., Shi, M., and Zhai, Y. (2016). Subtractive SELEX using agar beads for screening DNA aptamers with specific affinity to HIV gp41 antigen. *Nan Fang Yi Ke Da Xue Xue Bao* 36, 1592–1598. doi: 10.3969/j.issn.1673-4254.2016.12.01

Los Alamos HIV Molecular Immunology Database (2021a). *Gp160 Ab Epitope Map*. Available online at: https://www.hiv.lanl.gov/content/immunology/maps/ab/gp160.html (accessed September 15, 2021).

Los Alamos HIV Molecular Immunology Database (2021b). *Search Antibody Database 10E8 MAb*. Available online at: https://www.hiv.lanl.gov/content/immunology/ab_search.html?results=Searchandprotein=gp160andinrange=671-683andmab_name=10E8 (accessed February 27, 2022).

Los Alamos HIV Sequence Database (2021a). *HIV Circulating Recombinant Forms (CRFs)*. Available online at: https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html#CRF01 (accessed September 15, 2021).

Los Alamos HIV Sequence Database (2021b). *Sequence Search Interface*. Available online at: https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html (accessed November 19, 2020).

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. doi: 10.1371/journal.pone.0160597

Lu, W., Chen, S., Yu, J., Behrens, R., Wiggins, J., Sherer, N., et al. (2019). The polar region of the HIV-1 envelope protein determines viral fusion and infectivity by stabilizing the gp120-gp41 association. *J. Virol.* 93, e02128–e02118. doi: 10.1128/JVI.02128-18

Luque, F. J., and Camarasa, M. (2021). HIV-1 envelope spike MPER: from a vaccine target to a new druggable pocket for novel and effective fusion inhibitors. *Chem. Med. Chem.* 16, 105–107. doi: 10.1002/cmdc.202000411

Mink, M., Mosier, S. M., Janumpalli, S., Davison, D., Jin, L., Melby, T., et al. (2005). Impact of human immunodeficiency virus type 1 gp41 amino acid substitutions selected during enfuvirtide treatment on gp41 binding and antiviral potency of enfuvirtide in vitro. *J. Virol.* 79, 12447–12454. doi: 10.1128/JVI.79.19.12447-12454.2005

Mourez, T., Simon, F., and Plantier, J. C. (2013). Non-M variants of human immunodeficiency virus type 1. *Clin. Microbiol. Rev.* 26, 448–461. doi: 10.1128/CMR.00012-13

Murakami, T., and Freed, E. O. (2000). The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. *Proc. Natl. Acad. Sci. U.S.A.* 97, 343–348. doi: 10.1073/pnas.97.1.343

Oldfield, V., Keating, G. M., and Plosker, G. (2005). Enfuvirtide: a review of its use in the management of HIV infection. *Drugs* 65, 1139–1160. doi: 10.2165/00003495-200565080-00007

Özkaya Sahin, G., Holmgren, B., da Silva, Z., Nielsen, J., Nowroozalizadeh, S., and Esbjörnsson, J., et al. (2012). Potent intratype neutralizing activity distinguishes human immunodeficiency virus type 2 (HIV-2) from HIV-1. *J. Virol.* 86, 961–971. doi: 10.1128/JVI.06315-11

Pancera, M., Zhou, T., Druz, A., Georgiev, I. S., Soto, C., Gorman, J., et al. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* 514, 455–461. doi: 10.1038/nature13808

Pérez-Alvarez, L., Carmona, R., Ocampo, A., Asorey, A., Miralles, C., Pérez de Castro, S., et al. (2006). Long-term monitoring of genotypic and phenotypic resistance to T20 in treated patients infected with HIV-1. *J. Med. Virol.* 78, 141–147. doi: 10.1002/jmv.20520

Plantier, J.-C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemée, V., et al. (2009). A new human immunodeficiency virus derived from gorillas. *Nat. Med.* 15, 871–872. doi: 10.1038/nm.2016

Qadir, M. I., and Malik, S. A. (2010). HIV fusion inhibitors. *Rev. Med. Virol.* 20, 23–33. doi: 10.1002/rmv.631

Rantalainen, K., and Cottrell, C. A. (2019). Full length HIV-1 Env AMC011 in complex with PGT151 Fab. *PLoS Pathog.* (2019) 15:e1007920. doi: 10.2210/pdb6olp/pdb

Rey, M. A., Krust, B., Laurent, A. G., Guétard, D., Montagnier, L., and Hovanessian, A. G. (1989). Characterization of an HIV-2-related virus with a smaller sized extracellular envelope glycoprotein. *Virology* 173, 258–267. doi: 10.1016/0042-6822(89)90242-0

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., et al. (2000). HIV-1 nomenclature proposal. *Science* 288, 55–55. doi: 10.1126/science.288.5463.55d

Salminen, M. (2000). HIV inter-subtype recombination – Consequences for the epidemic. *AIDS Rev.* 3, 178–189. Available online at: https://www.researchgate.net/profile/Mika-Salminen/publication/254903484_HIV_Inter-subtype_Recombination_-_Consequences_for_the_Epidemic/links/53df80660cf2aede4b4905b9/HIV-Inter-subtype-Recombination-Consequences-for-the-Epidemic.pdf

Scharf, L., Scheid, J. F., Lee, J. H., West, A. P., Chen, C., Gao, H., et al. (2014). Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Rep.* 7, 785–795. doi: 10.1016/j.celrep.2014.04.001

Simon, F., Mauclère, P., Roques, P., Loussert-Ajaka, I., Müller-Trutwin, M. C., Saragosti, S., et al. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* 4, 1032–1037. doi: 10.1038/2017

Sliepen, K., Han, B. W., Bontjer, I., Mooij, P., Garces, F., Behrens, A.-J., et al. (2019). Structure and immunogenicity of a stabilized HIV-1 envelope trimer based on a group-M consensus sequence. *Nat Commun.* 10:2355. doi: 10.1038/s41467-019-10262-5

Smith, T., Masciotra, S., Luo, W., Sullivan, V., Switzer, W. M., Johnson, J. A., et al. (2021). Broadly neutralizing HIV-1 antibody reactivity in HIV tests: implications for diagnostics. *AIDS* 35, 1561–1565. doi: 10.1097/QAD.0000000000002898

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Torrents de la Peña, A., Rantalainen, K., Cottrell, C. A., Allen, J. D., van Gils, M. J., Torres, J. L., et al. (2019). Similarities and differences between native HIV-1 envelope glycoprotein trimers and stabilized soluble trimer mimetics. *PLoS Pathog.* 15:e1007920. doi: 10.1371/journal.ppat.1007920

Visseaux, B., Damond, F., Matheron, S., Descamps, D., and Charpentier, C. (2016). HIV-2 molecular epidemiology. *Infect. Genet. Evol.* 46, 233–240. doi: 10.1016/j.meegid.2016.08.010

Wang, Q., and Zhang, L. (2020). Broadly neutralizing antibodies and vaccine design against HIV-1 infection. *Front. Med.* 14, 30–42. doi: 10.1007/s11684-019-0721-9

Wensing, A. M., Calvez, V., Ceccherini-Silberstein, F., Charpentier, C., Günthard, H. F., Paredes, R., et al. (2019). 2019 Update of the drug resistance mutations in

HIV-1. *Top. Antivir. Med.* 27, 111–121. Available online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6892618/pdf/tam-27-111.pdf

Wyma, D. J., Kotov, A., and Aiken, C. (2000). Evidence for a stable interaction of gp41 with Pr55Gag in immature human immunodeficiency virus type 1 particles. *J. Virol.* 74, 9381–9387. doi: 10.1128/JVI.74.20.9381-9387.2000

Xu, W., Pu, J., Su, S., Hua, C., Su, X., Wang, Q., et al. (2019). Revisiting the mechanism of enfuvirtide and designing an analog with improved fusion inhibitory activity by targeting triple sites in gp41. *AIDS* 33, 1545–1555. doi: 10.1097/QAD.0000000000002208

Yamaguchi, J., Vallari, A., McArthur, C., Sthreshley, L., Cloherty, G. A., Berg, M. G., et al. (2020). Brief report: complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. *J. Acquir. Immune Defic. Syndr.* 83, 319–322. doi: 10.1097/QAI.0000000000002246

Yuan, M., Cottrell, C. A., Ozorowski, G., van Gils, M. J., Kumar, S., Wu, N. C., et al. (2019). Conformational plasticity in the HIV-1 fusion peptide facilitates recognition by broadly neutralizing antibodies. *Cell Host Microbe* 25, 873–883. doi: 10.1016/j.chom.2019.04.011