Supplementary Data

Supplementary Methods

Structural data

The four-stranded β-sheet conformations of the V1V2 domain were derived from previously published co-crystal structures of V1V2 with BNAbs PG9 and PG16.[1,2] The conformation of the gp140 trimer was derived from its respective co-crystal structure in complex with the BNAb PGT122.[3]

V1V2 bioinformatics

V1V2 loop sequence information was obtained from the Los Alamos National Laboratory HIV Database and filtered according to a protocol described elsewhere.[4] Briefly, the initial set of sequences was filtered by removing problematic sequences, sequences with non-amino acid symbols, sequences with V1V2 entirely deleted, and sequences with unspecified side chains. Only one sequence per patient was randomly selected from the resulting set, while all other sequences, including all the sequences with no reported patient ID, were also filtered out. No subtype-based filtering was performed at this point. A set of 5303 diverse V1V2 sequences was retained for the further analysis. Subtype composition of this set was the following: 39% of subtype B, 21% of subtype C, 9% subtype A (including A, A1, and A2), 5% of CRF01_AE, 4% CRF02_AG, 22% other subtypes and CRFs. Amino acid composition at each position of V1V2 (according to the HxB2 numbering) was calculated in the alignment of all filtered sequences. Each position of V1V2 was then scored according to its amino acid variability as described elsewhere.[4] Specifically, the score was calculated as $1 - p/n$, where p is the number of occurrences of the most common amino acid on a given position of V1V2 alignment and n is the total number of V1V2 sequences in the alignment. The scoring was then normalized by assigning the most variable position of V1V2 to 100%. The same analysis was also performed separately for three subtypes (A, B, and C) and two CRFs (01_AE and 02_AG). The analysis for other subtypes and CRFs was omitted due to a small number of sequences of those clades available in the study set. Insertion and deletion patterns were obtained by calculating a proportion of sequences in the filtered set, which have a deletion or insertion at a given position of V1V2. Finally, the resulting positional variability scores were mapped onto the crystallographic conformations of V1V2.

V1V2 accessibility estimation

Accessibility of amino acids to antibodies within the V1V2 domain was quantified based on the relative solvent accessibility of each amino acid within the 3D crystallographic structure of the fully glycosylated and non-glycosylated gp140 trimer. Solvent accessible areas were calculated using the ICM Pro software (Molsoft, LLC, La Jolla, CA)[5] with a probe radius of 1.4Å. For each amino acid X, relative solvent accessibility was calculated by normalizing its solvent accessible area calculated in the context of the gp140 trimer to the maximum solvent accessible area calculated in a GXG tripeptide. If relative solvent accessibility of a given residue was greater than 20%, the residue was considered to be solvent exposed.[6] As the resolution of the gp140 trimer structure is not high we also provide solvent accessibility values for each amino acid in the context of the scaffolded glycosylated V1V2 domain from the co-crystal structures of V1V2 with BNAbs PG9 and PG16 (see Supplementary Fig. S3). Note also that solvent accessibility for residues directly adjacent to the non-resolved parts of V1V2 structures cannot be calculated accurately in a same way as for other residues, and, therefore, should be accepted with caution.

V1-loop sampling

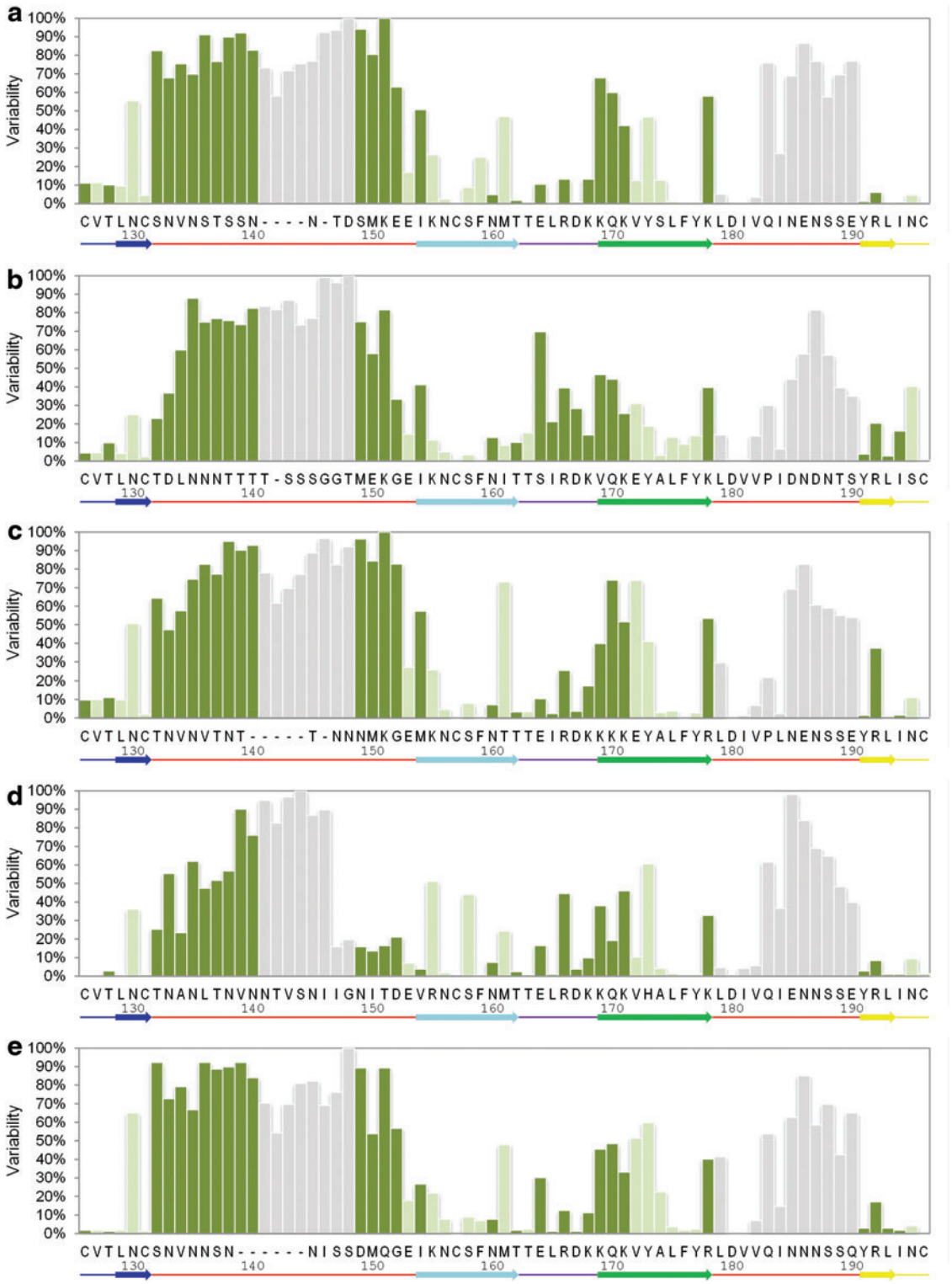
Conformational sampling of the $V1^{132–153}$ of BG505 (including its N137-linked glycan) was performed in the context of the fully glycosylated gp140 trimer using the loop modeling tool of the ICM Pro and a previously described protocol.[7] The resulting conformational stack was compressed to obtain the 10 most structurally diverse $V1^{132–153}$ conformations (including the one observed in the crystallographic structure of the gp140 trimer[3]). The ensemble of these 10 conformations indicates the perimeter of the 3D volume that any of the dynamic conformations of the V1 loop can potentially occupy.

Supplementary References

1. McLellan JS, Pancera M, Carrico C, *et al.:* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature 2011;480(7377):336–343.
2. Pancera M, Shahzad-Ul-Hussan S, Doria-Rose NA, *et al.:* Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. Nat Struct Mol Biol. Jul 2013;20(7):804–813.
3. Julien JP, Cupo A, Sok D, *et al.:* Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. Science. Oct 31 2013.
4. Almond D, Kimura T, Kong X, Swetnam J, Zolla-Pazner S, Cardozo T: Structural conservation predominates over sequence variability in the crown of HIV type 1's V3 loop. AIDS Res Hum Retroviruses. Jun 2010;26(6):717–723.
5. Abagyan R, Totrov M, Kuznetsov D: ICM–a new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. Journal of Computational Chemistry. 1994;15(5):488–506.
6. Chen H, Zhou HX: Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. Nucleic Acids Res. 2005;33(10):3193–3199.
7. Arnautova YA, Abagyan RA, Totrov M: Development of a new physics-based internal coordinate mechanics force field and its application to protein loop modeling. Proteins. Feb 2011;79(2):477–498.

HxB2 #	HxB2 sequence	Glycosylated trimer					Non-glycosylated trimer					Relative exposure gain
		4NCO.a	4NCO.e	4NCO.i	Mean	Relative SAA	4NCO.a	4NCO.e	4NCO.i	Mean	Relative SAA	
126	C	37.8	41.7	43.1	40.9	26%	37.8	41.7	43.1	40.9	26%	0%
127	V	21.1	25.5	23.1	23.2	14%	21.5	25.5	23.4	23.5	14%	0%
128	S	44.9	44.7	31.0	40.2	26%	49.5	51.2	41.4	47.4	30%	5%
129	L	9.1	10.0	10.7	9.9	5%	9.1	10.0	10.7	9.9	5%	0%
130	K	45.8	34.5	30.2	36.8	19%	86.2	80.5	81.4	82.7	42%	23%
131	C	7.1	7.1	7.6	7.3	5%	7.1	7.1	7.6	7.3	5%	0%
132	T	68.0	71.8	66.1	68.6	44%	93.9	93.7	94.6	94.1	60%	16%
133	D	56.7	62.7	62.5	60.6	35%	56.7	62.7	62.5	60.6	35%	0%
134	L	116.9	123.3	118.1	119.4	70%	124.2	130.1	133.5	129.2	76%	6%
135	K	49.8	38.6	50.2	46.2	29%	49.8	38.6	50.2	46.2	29%	0%
136	N	152.7	146.1	153.2	150.7	87%	152.7	146.1	153.2	150.7	87%	0%
137	D	73.5	75.3	79.0	75.9	44%	130.1	134.5	137.3	134.0	77%	33%
138	T	80.5	98.5	97.0	92.0	46%	99.4	115.3	110.8	108.5	55%	8%
139	N	63.4	56.2	51.5	57.0	36%	108.3	111.9	108.8	109.7	70%	34%
140	T	132.4	131.4	128.0	130.6	80%	132.4	131.4	128.0	130.6	80%	0%
141	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
142	S	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
143	S	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
144	S	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
145	G	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
146	R	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
147	M	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
148	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
149	M	94.9	90.1	103.0	96.0	59%	94.9	90.1	103.0	96.0	59%	0%
150	E	98.9	136.0	147.1	127.3	60%	98.9	136.0	147.1	127.3	60%	0%
151	K	136.6	155.4	164.8	152.3	58%	136.6	155.4	164.8	152.3	58%	0%
152	G	20.2	20.1	21.2	20.5	23%	20.2	20.1	21.2	20.5	23%	0%
153	E	30.1	48.5	32.4	37.0	19%	30.1	48.5	32.4	37.0	19%	0%
154	I	65.5	69.2	63.5	66.1	32%	65.5	69.2	63.5	66.1	32%	0%
155	K	21.4	20.9	22.0	21.4	9%	21.4	20.9	22.0	21.4	9%	0%
156	N	8.3	12.4	8.8	9.8	6%	49.7	53.8	51.3	51.6	30%	24%
157	C	0.0	0.0	0.0	0.0	0%	1.8	0.5	0.0	0.8	0%	0%
158	S	15.8	17.0	17.9	16.9	13%	38.7	36.4	36.0	37.0	28%	15%
159	F	20.4	20.6	18.8	20.0	8%	20.4	20.6	18.8	20.0	8%	0%
160	N	33.0	38.5	36.5	36.0	21%	83.2	94.0	93.8	90.3	52%	31%
161	I	4.4	6.1	6.1	5.5	3%	4.4	6.1	6.1	5.5	3%	0%
162	S	48.8	66.6	63.2	59.5	38%	48.8	66.6	63.2	59.5	38%	0%
163	T	7.5	8.0	8.5	8.0	5%	7.5	8.0	8.5	8.0	5%	0%
164	S	53.9	58.8	49.7	54.1	28%	65.3	67.2	63.2	65.2	34%	6%
165	I	55.9	61.5	51.9	56.4	27%	55.9	61.5	51.9	56.4	27%	0%
166	R	129.5	136.6	79.6	115.2	44%	148.9	157.9	103.4	136.7	52%	8%
167	G	84.5	89.4	93.3	89.1	55%	84.5	89.4	93.3	89.1	55%	0%
168	K	132.6	143.3	137.1	137.7	59%	132.6	143.3	137.1	137.7	59%	0%
169	V	96.4	104.4	108.9	103.2	44%	118.9	136.7	135.0	130.2	56%	12%
170	Q	105.6	104.8	98.7	103.0	53%	105.6	104.8	98.7	103.0	53%	0%
171	K	138.8	137.2	144.3	140.1	60%	177.8	171.1	176.6	175.1	75%	15%
172	E	8.3	13.5	13.4	11.8	7%	8.3	13.5	13.4	11.8	7%	0%
173	Y	49.9	45.8	51.3	49.0	20%	126.1	118.8	124.5	123.1	49%	30%
174	A	8.9	8.9	8.5	8.8	7%	8.9	8.9	8.5	8.8	7%	0%
175	F	7.9	7.4	7.4	7.6	4%	7.9	7.4	7.4	7.6	4%	0%
176	F	33.8	25.4	33.9	31.0	13%	33.8	25.4	33.9	31.0	13%	0%
177	Y	9.9	8.3	9.6	9.3	4%	9.9	8.3	9.6	9.3	4%	0%
178	K	165.9	174.8	161.7	167.5	64%	165.9	174.8	161.7	167.5	64%	0%
179	L	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
180	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
181	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
182	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
183	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
184	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
185	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
186	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
187	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
188	T	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
189	T	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
190	S	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
191	Y	137.0	146.4	134.8	139.4	56%	137.0	146.4	134.8	139.4	56%	0%
192	K	144.2	175.5	153.6	157.8	60%	154.3	176.0	158.2	162.8	62%	2%
193	L	107.6	103.3	108.7	106.5	52%	107.6	103.3	108.7	106.5	52%	0%
194	T	76.9	90.6	80.9	82.8	42%	76.9	90.6	80.9	82.8	42%	0%
195	S	29.2	35.4	23.0	29.2	17%	29.2	35.4	23.0	29.2	17%	0%
196	C	16.5	14.0	22.7	17.7	11%	16.5	14.0	22.7	17.7	11%	0%

SUPPLEMENTARY FIG. S1. Difference in V1V2 position-specific accessibility between glycosylated and non-glycosylated trimers. The surface-accessible area (in $Å^2$) and mean relative accessibility (relative SAA, in %) are shown for each HxB2 position between 126 and 196 in the context of glycosylated and nonglycosylated gp140 trimers. Exposure gain upon loss of glycosylation is shown in the rightmost column.



SUPPLEMENTARY FIG. S2. Position-specific variability and accessibility in the V1V2 domains of different HIV-1 subtypes and circulating recombinant forms (CRFs). Plots are made in the same way as the plot in Fig. 2a. Data for three subtypes and two CRFs are shown: subtype A (including A, A1, and A2) **(a)**, subtype B **(b)**, subtype C **(c)**, CRF01_AE **(d)**, and CRF02_AG **(e)**.

HxB2 #	HxB2 sequence	4NCO.a	4NCO.e	4NCO.i	3U2S.g	3U2S.c	3U4E.g	3U4E.j	4DQO	Mean
126	C	24%	27%	27%	4%	5%	9%	5%	11%	14%
127	V	12%	15%	13%	20%	15%	17%	17%	31%	18%
128	S	29%	28%	20%	17%	26%	1%	1%	35%	20%
129	L	4%	5%	5%	2%	3%	3%	4%	12%	5%
130	K	23%	18%	15%	12%	17%	15%	17%	25%	18%
131	C	4%	5%	5%	1%	1%	4%	4%	4%	3%
132	T	43%	46%	42%	62%	57%	61%	54%	45%	51%
133	D	33%	36%	36%	n/a	n/a	32%	35%	n/a	34%
134	L	68%	72%	69%	n/a	n/a	2%	1%	n/a	42%
135	K	32%	25%	32%	n/a	n/a	46%	41%	n/a	35%
136	N	88%	84%	88%	n/a	n/a	44%	36%	n/a	68%
137	D	42%	43%	45%	n/a	n/a	79%	85%	n/a	59%
138	T	40%	49%	49%	n/a	n/a	n/a	n/a	n/a	46%
139	N	40%	36%	33%	n/a	n/a	n/a	n/a	n/a	36%
140	T	81%	81%	79%	n/a	n/a	n/a	n/a	n/a	80%
141	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
142	S	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
143	S	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
144	S	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
145	G	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
146	R	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
147	M	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
148	I	n/a	n/a	n/a	n/a	n/a	78%	93%	n/a	86%
149	M	58%	55%	63%	n/a	n/a	59%	44%	n/a	56%
150	E	46%	64%	69%	n/a	n/a	61%	71%	n/a	62%
151	K	52%	59%	63%	n/a	n/a	21%	8%	n/a	41%
152	G	23%	23%	24%	n/a	n/a	66%	61%	n/a	39%
153	E	16%	25%	17%	90%	93%	59%	61%	101%	58%
154	I	32%	34%	31%	47%	39%	8%	11%	42%	31%
155	K	9%	9%	9%	37%	39%	11%	15%	47%	22%
156	N	5%	7%	5%	41%	40%	10%	10%	46%	20%
157	C	0%	0%	0%	6%	6%	0%	0%	7%	2%
158	S	12%	13%	13%	16%	18%	19%	14%	21%	16%
159	F	8%	9%	8%	17%	16%	12%	12%	20%	13%
160	N	19%	22%	21%	15%	13%	13%	12%	13%	16%
161	I	2%	3%	3%	41%	35%	35%	32%	29%	22%
162	S	31%	42%	40%	30%	10%	14%	15%	36%	27%
163	T	5%	5%	5%	28%	28%	42%	38%	29%	22%
164	S	28%	30%	26%	95%	80%	44%	45%	75%	53%
165	I	27%	30%	25%	44%	42%	31%	30%	35%	33%
166	R	49%	52%	30%	78%	61%	66%	68%	67%	59%
167	G	52%	55%	57%	77%	81%	66%	63%	70%	65%
168	K	57%	61%	59%	59%	57%	53%	48%	57%	56%
169	V	41%	45%	47%	51%	39%	43%	48%	49%	45%
170	Q	54%	53%	50%	58%	58%	54%	55%	49%	54%
171	K	60%	59%	62%	51%	51%	59%	57%	52%	56%
172	E	5%	8%	8%	25%	29%	18%	23%	28%	18%
173	Y	20%	18%	20%	42%	46%	31%	32%	30%	30%
174	A	7%	7%	6%	16%	17%	34%	32%	29%	18%
175	F	4%	4%	4%	18%	21%	44%	46%	34%	22%
176	F	14%	11%	14%	46%	44%	32%	38%	52%	31%
177	Y	4%	3%	4%	56%	58%	52%	55%	87%	40%
178	K	63%	66%	61%	95%	95%	60%	56%	n/a	71%
179	L	n/a	n/a	n/a	n/a	n/a	89%	56%	n/a	72%
180	D	n/a	n/a	n/a	n/a	n/a	n/a	104%	n/a	104%
181	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
182	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
183	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
184	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
185	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
186	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
187	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
188	T	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
189	T	n/a	n/a	n/a	n/a	n/a	130%	109%	n/a	120%
190	S	n/a	n/a	n/a	75%	78%	62%	55%	59%	66%
191	Y	55%	58%	54%	65%	42%	29%	54%	41%	50%
192	K	55%	67%	58%	58%	73%	21%	20%	80%	54%
193	L	52%	50%	53%	48%	39%	21%	25%	38%	41%
194	T	39%	46%	41%	51%	49%	40%	38%	45%	44%
195	S	17%	20%	13%	38%	37%	53%	51%	43%	34%
196	C	11%	9%	14%	41%	42%	39%	37%	46%	30%

SUPPLEMENTARY FIG. S3. Position-specific accessibility in the V1V2 domain. Relative solvent accessibility is shown for each HxB2 position between 126 and 196 in the context of the glycosylated gp140 trimer and in the context of scaffolded glycosylated V1V2 from the cocrystal structures of V1V2 with broadly neutralizing antibodies (BNAbs) PG9 and PG16.