

Supplementary Information

Conformational plasticity of the HIV-1 gp41 immunodominant region is recognized by multiple non-neutralizing antibodies

Jonathan D. Cook^{1,2}, Adree Khondker¹, and Jeffrey E. Lee^{1,*}

¹Department of Laboratory Medicine and Pathobiology, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, M5S 1A8, Canada

²Present address: Department of Medicine, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, M5S 1A8, Canada

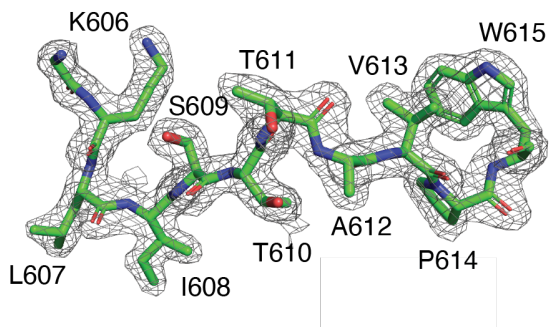
*Correspondence to: J.E.L.

Email: jeff.lee@utoronto.ca

Phone: (416)-978-1060

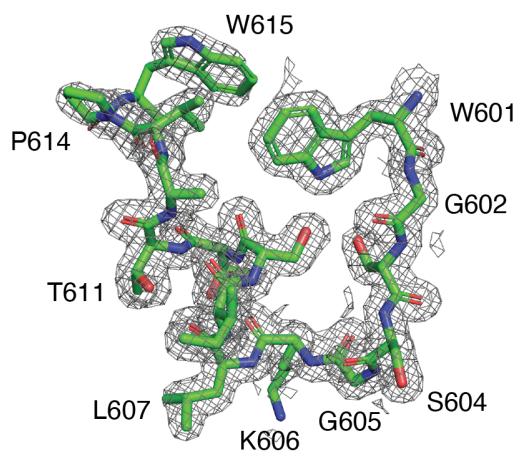
Fax: (416)-978-5959

a



3D6-bound PID

b

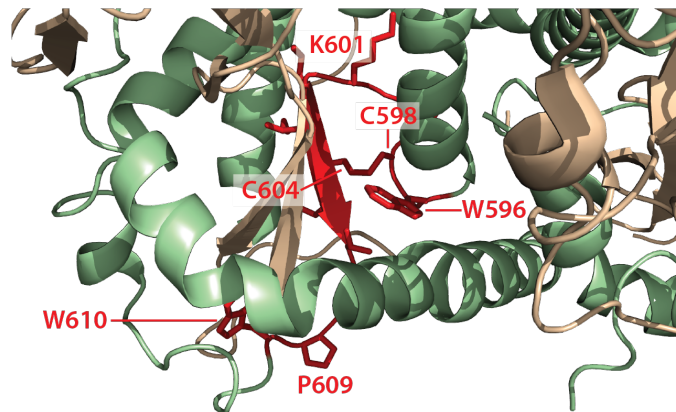
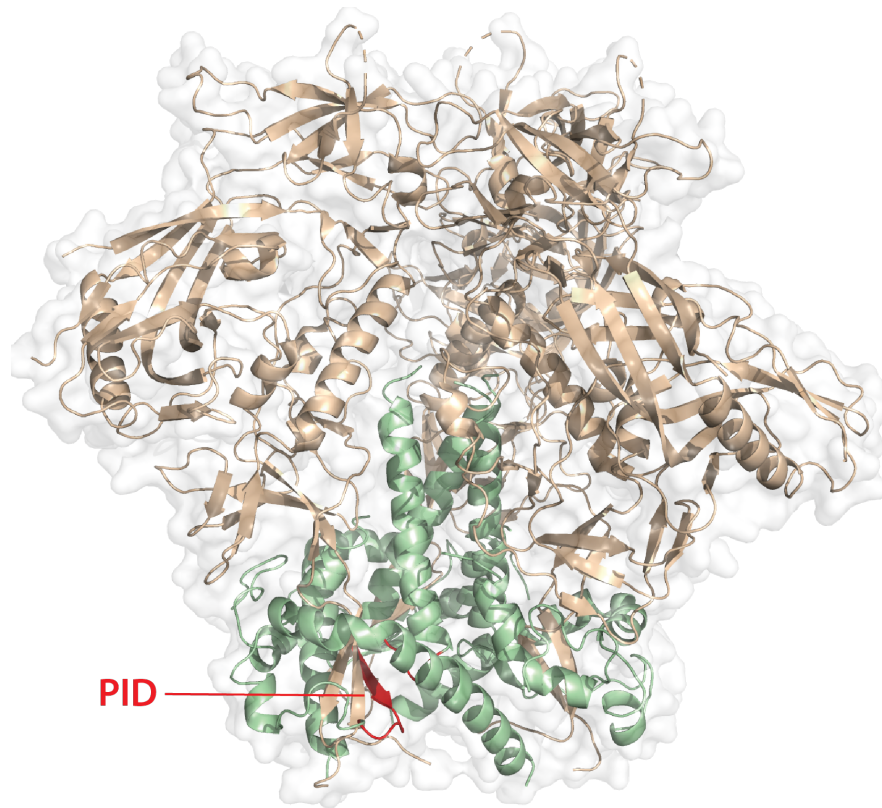


F240-bound PID

Supplementary Figure 1. Electron density map of Fab 3D6- and F240-bound PID regions. A representative composite omit $2|F_o| - |F_c|$ electron density map of the **a.** Fab 3D6-bound and **b.** Fab F240-bound PID region contoured at 1σ and superimposed with its final refined model.

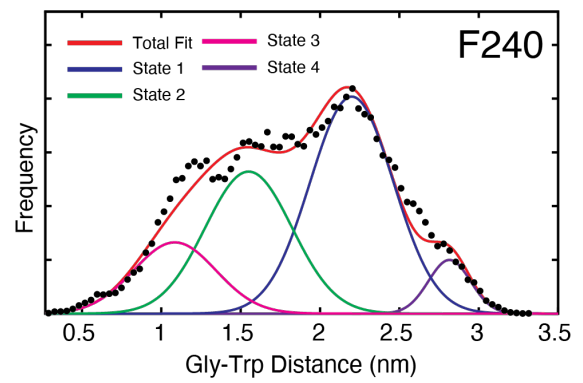
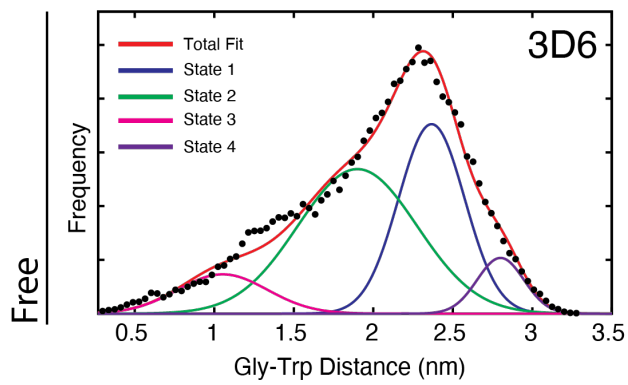
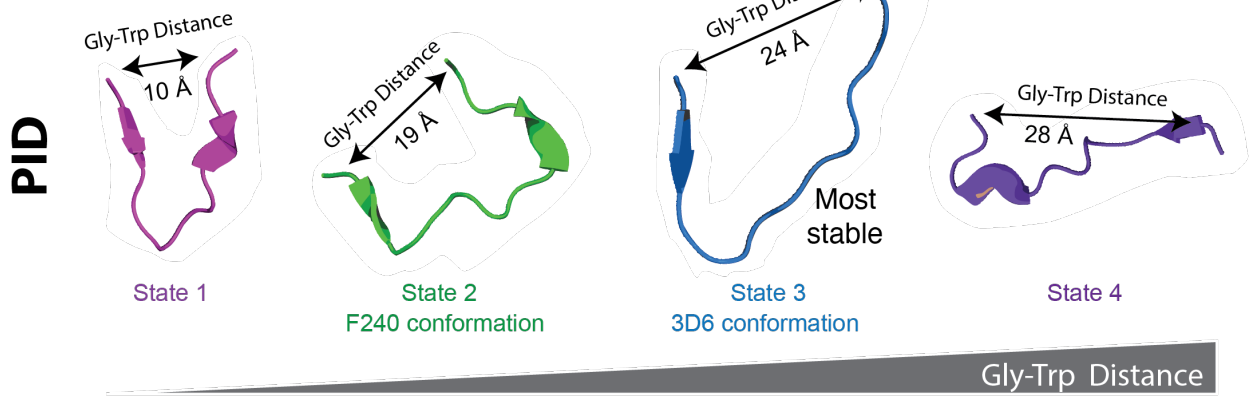
	FR1-IMGT (1-26)	CDR1-IMGT (27-38)	FR2-IMGT (39-55)	CD (
F240 Heavy M99652 H sapiens IGHV3-11*01 F	1 10 20	30	40 50)
	Q V A A S	T	T Q L A T	
R2-IMGT 56-65)		FR3-IMGT (66-104)		
F240 Heavy M99652 H sapiens IGHV3-11*01 F	60 70 80 90 100			
	KD . GSEK KYADSLQ . GRFAVSRDNANNLVFLQLNTVEDDDTGVYYC AR			
	SS . GSTI YYADSVK . GRFTISRDNAKNSLYLQMNSLRAEDTAVYYC AR			
	KD EK K LQ AV N L V F L T V E D D G			
F240 kappa Z00023 H sapiens IGKV4-1*01 F	1 10 20	30	40 50)
	EFLLTQSPDLSAVTLGETATITCRSS RNIHLHSLNKNY LAWYQQRPQAPKLLVI WA			
	DIVMTQSPDLSAVSLGERATINCKSS QSVLYSSNNKNY LAWYQQRPQAPKLLIY WA			
	EFL L T T T R R N I H L R A V I			
R2-IMGT 56-65)		FR3-IMGT (66-104)		
F240 kappa Z00023 H sapiens IGKV4-1*01 F	60 70 80 90 100			
S MRVSGVA . DRFSGSG . . SGTDFAITLISLQPEDAAVYYC QHYHTTH			
S TRESGVP . DRFSGSG . . SGTDFTLTISLQAEDEVAVYYC QQYYSTP			
	M V A A P A H T H			
3D6 heavy KC713947 H sapiens IGHV3-9*03 F	1 10 20	30	40 50)
	EVQLVESGG . GLVQPGRSLRLS CAAS GFTF . . . NDYA MHWVRQAPGKGLEWVSG IS			
	EVQLVESGG . GLVQPGRSLRLS CAAS GFTF . . . DDYA MHWVRQAPGKGLEWVSG IS			
	N			
R2-IMGT 56-65)		FR3-IMGT (66-104)		
3D6 heavy KC713947 H sapiens IGHV3-9*03 F	60 70 80 90 100			
	WD . SSSI GYADSVK . GRFTISRDNAKNSLYLQMNSLRAEDMALYYC VK			
	WN . SSGI GYADSVK . GRFTISRDNAKNSLYLQMNSLRAEDMALYYC AKD			
	D S V			
3D6 light X72813 H sapiens IGKV1-5*03 F	1 10 20	30	40 50)
	DIQMTQSPSTLSASVGDRTVITCRAS QSI SRW LAWYQQKPKVAPKLLIY KA			
	DIQMTQSPSTLSASVGDRTVITCRAS QSI SSW LAWYQQKPKVAPKLLIY KA			
	R V			
R2-IMGT 56-65)		FR3-IMGT (66-104)		
3D6 light X72813 H sapiens IGKV1-5*03 F	60 70 80 90 100			
S SLESGVP . SRFSGSG . . SGTEFTLTISLQPDFATYYC QQYNYSYS			
S SLESGVP . SRFSGSG . . SGTEFTLTISLQPDFATYYC QQYNYSYS			

Supplementary Figure 2. Comparison of Fab 3D6 and Fab F240 heavy and light chains with the inferred germline. The light and heavy chain variable regions of Fab F240 and Fab 3D6 compared with the inferred germline. Antibody amino acid changes from germline are listed below the sequence alignment and highlighted in bold. The amino acids are labeled according to the Kabat numbering scheme.

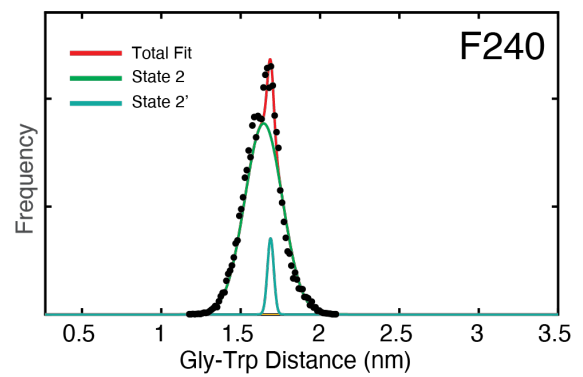
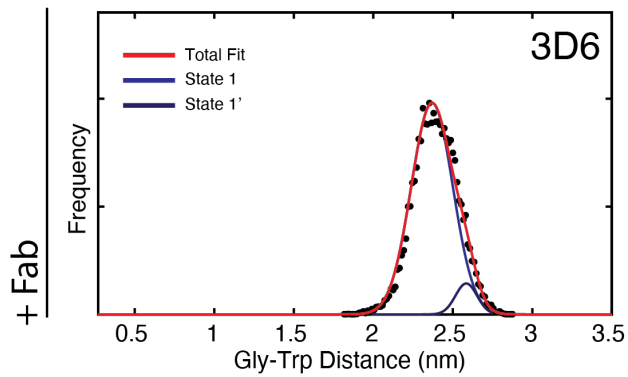


Supplementary Figure 3. Pre-fusion HIV-1 gp140 PID conformation. The crystal structure of pre-fusion HIV-1 SOSIP gp140 bound to human Fab PGT122 and 35O22 (PDB: 4TVP) revealed that PID residues 596-610 (highlighted in red) is buried, with only residue 608-611 partially accessible. Residues 602-606 forms an intermolecular β -sheet with the gp120 subunit. In the Fab F240 and 3D6-bound structures, residues 602-606 forms a short helix and random coil, respectively.

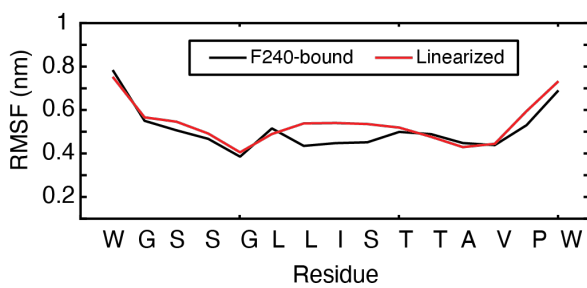
a



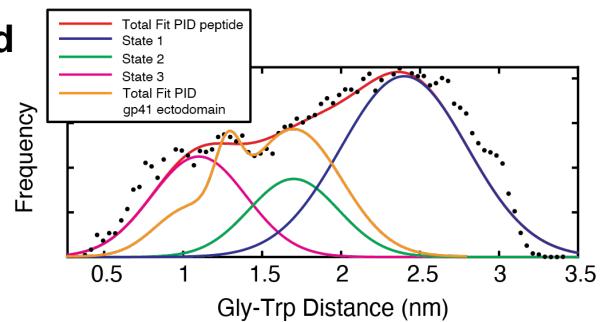
b



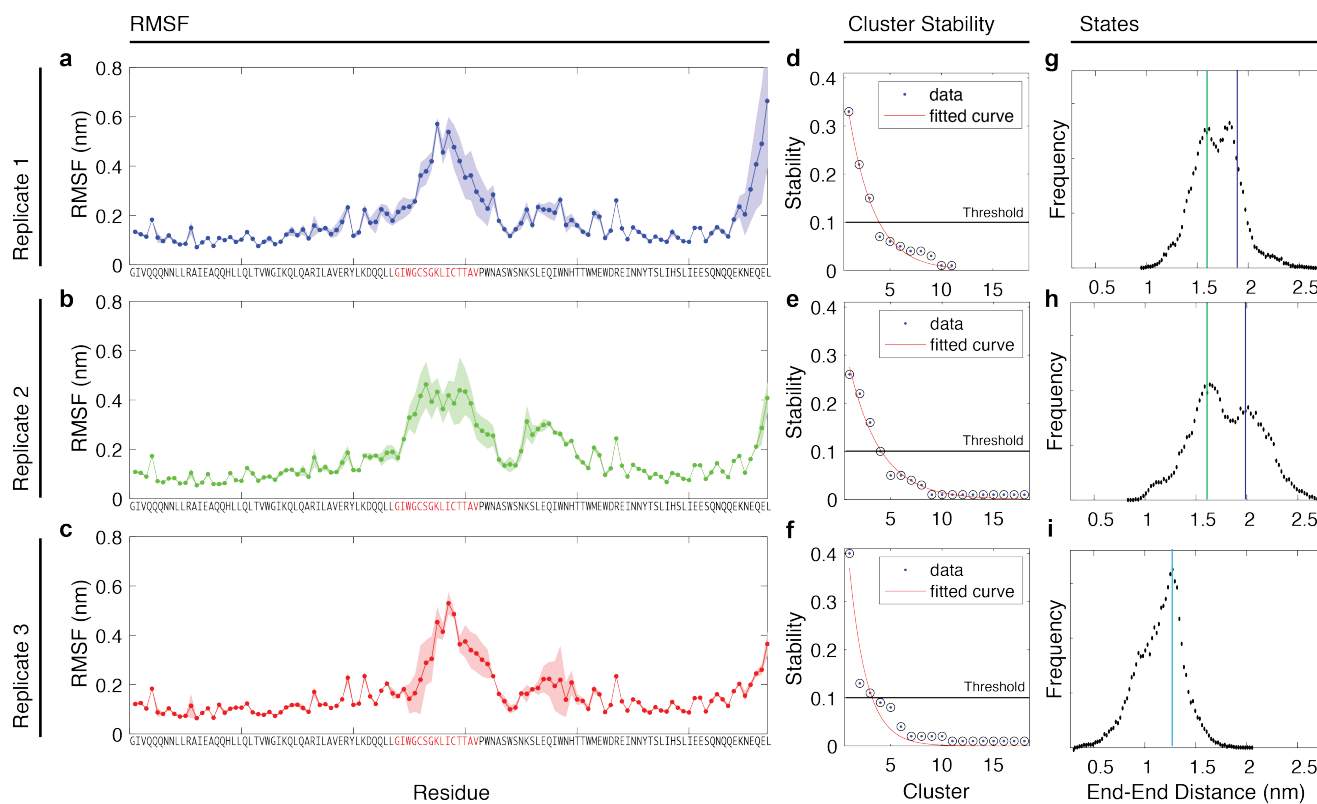
c



d



Supplementary Figure 4. Molecular dynamics simulations of the HIV-1 gp41 PID. a. Analysis of the MD simulations of the PID peptide extracted from the 3D6-bound and F240-bound structures. Raw data showing the frequency of end-to-end distance measurements between Gly600 and Trp610 (black dots) are used to calculate an overall total fit observed in the simulation. Potential metastable states are deconvoluted from the total fit, as estimated by distinct Gaussian distributions. Upon fitting, the analyses revealed four metastable conformations, with end-to-end distances of 1.2, 1.9, 2.4, and 2.8 nm (blue, green, pink and purple curves). The metastable states with 2.4 and 1.9 nm Gly-Trp end-to-end distances agree with the PID conformations observed in the 3D6 and F240-bound structures, respectively. **b.** MD simulations of the PID region bound to Fab 3D6 and Fab F240 showed state selectivity. A single major PID conformation is observed with high frequency when bound to Fab 3D6 or F240. This shows that the PID region is locked into a specific conformation upon binding the Fab. Raw data is shown as black dots with the total fit displayed as the red curve. **c.** Root-mean squared fluctuation (RMSF) analysis of PID region MD simulations starting from either a linearized or F240-bound peptide conformation reveals no difference in positional fluctuation of each residue based on starting conditions. This suggests that the results of MD simulations do not differ between using a linearized or Fab-bound conformation of the PID region. **d.** Analysis of PID conformations from MD simulations of the HIV-1 gp41 ectodomain reveal that the Gly-Trp end-to-end distances are observed in the free PID peptide MD simulations as well (Total Fit, State 1, 2 and 3). Raw data showing the frequency of end-to-end distance measurements for the free PID peptide conformation are shown as black dots, with the total fit shown as the red curve. Three metastable conformations (State 1, 2 and 3) are deconvoluted from the total fit. The total fit of the PID region from the MD simulation of the complete gp41 ectodomain (orange curve; also in Figure 6b) reveals a similar conformation to State 2 and 3 in the free PID peptide simulation.



Supplementary Figure 5. Replicates of HIV-1 gp41 ectodomain MD simulation. a-c. RMSF of each residue in three replicate of the complete HIV-1 gp41 ectodomain MD simulation over 100 ns after 100 ns equilibration, **d-f.** Stability (% time spent in conformation) for each cluster from most stable to least stable, fit with decaying exponential, which shows fully the different states found in the PID domain in solution. **g-i.** PID end-to-end distance over the last 100 ns of simulation showing multiple exhibited PID conformations in each replicate. Together, this shows that each replicate has end-to-end distances that correspond with the metastable state conformation end-to-end distances (highlighted lines) found in the full-length HIV-1 gp41 ectodomain suggesting this state reliably occurs in simulation.

F240 heavy chain codon-optimized DNA and protein sequence

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gaccaggtgcagctgggtgcagtcaggtggcggcggtggaagccggcgccagcctgaga
D Q V Q L V Q S G G G V V K P G A S L R
cttgcattgtagcgttccggctttacattcaccgattactatgtcatggatcagacag
L A C S A S G F T F T D Y Y M S W I R Q
accocccgcaaaggcctgcagtggtggcgtacataactaaagatgggtccgagaagaaa
T P G K G L Q W L A Y I T K D G S E K K
tacgggattcactgcaaggaagattcgcggtgtccagagacaacgccaataatctcgtc
Y A D S L Q G R F A V S R D N A N N L V
ttctgcagctcaatactgtagaggacgacgataccggcgtgtattactgccaagagac
F L Q L N T V E D D D T G V Y Y C A R D
gacgggtattacgacagatccggctattacggggtcttcgacctgtggggacaggaatt
D G Y Y D R S G Y Y G V F D L W G Q G I
agagtgcagtttagcagctcaaccaaggccctcagtttcccactggccccagc
R V T V S S A S T K G P S V F P L A P S
tctaagtctacaagcggagcactgccgcgctcggctgtctcgtgaaagattatttccca
S K S T S G G T A A L G C L V K D Y F P
gagcctgttactgtgagctggaactccgggtgcccttactgtctggcgttcatactttccg
E P V T V S W N S G A L T S G V H T F P
ggcgtcctccagtcctctgggctgtactccctctcatcagttgtgacctccaagctca
A V L Q S S G L Y S L S S V V T V P S S
agtctgggtactcagacatcctgtaacgttaatcacaagccatccaataccaagtg
S L G T Q T Y I C N V N H K P S N T K V
gacaaaaaagtgagccaaagtcagcggcggcgtggcaccggggtcccacatcaccat
D K K V E P K S C G R L V P R G S H H H
caccatcaccatcaccattga
H H H H H H H *
```

F240 light chain codon-optimized DNA and protein sequence

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gacgagtttctgctgaccagtgctcagactcactggcagtgactctcgggaaactgcc
D E F L L T Q S P D S L A V T L G E T A
acaatcactgtagaagcagccgaatattctccacagtctcaacaataagaattacctc
T I T C R S S R N I L H S L N N K N Y L
gcatggtaaccaacagagcctggacagggcacctaaactccttgttatttgggcttctatg
A W Y Q Q R P G Q A P K L L V I W A S M
cgagtctctggagtgccgacagattctctggttccggctcaggtacagatttcgccctc
R V S G V A D R F S G S G S G T D F A L
acaatctccagctgcagcagaggacgctgcaagtgtattatgccagcattactacaca
T I S S L Q P E D A A V Y Y C Q H Y Y T
acacacaggacatttggacaaggtacacgggttgagatccgaaggacctggcgtgccccc
T H R T F G Q G T R V E I R R T V A A P
tccgtcttcatttccctccctcagatgaacagttgaaatcaggaacggccttctgagtt
S V F I F P P S D E Q L K S G T A S V V
tgtctcctgaataattttaccocccgagaagcacaagtgcaagtggaagtggaacgcc
C L L N N F Y P R E A K V Q W K V D N A
ttgcagtcagggaaactcacaagaatcagtgactgagcaggacagcaaggacagcacgtat
L Q S G N S Q E S V T E Q D S K D S T Y
tcattgagctccacacttactgagcaaggcagattacgagaagcataaagtgatgct
S L S T L S T L S K A D Y E K H K V Y A
tgtgaggttaccatcagggactgtcctctccagtcacaagctttcaaccggggggag
C E V T H Q G L S S P V T K S F N R G E
tgctga
C -
```

Supplementary Figure 6. Nucleotide and translated amino acid sequences for Fab F240 heavy and light chains. The native F240 sequence is colored in black, while regions colored in red, green and blue correspond to the Ig-kappa signal peptide, thrombin cleavage tag, and C-terminal 10-histidine purification tag, respectively.

3D6 heavy chain codon-optimized DNA and protein sequence

atggagacagacacactcctgctatgggtactgctgctctgggtccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gacgaggtgcagttggtggaatccggggcggtggtcaaccgggtcggagtccttagg
D E V Q L V E S G G G L V Q P G R S L R
ctttcatgcgcgcaagcgggtttacattcaatgactatgctatgcattgggttcgacag
L S C A A S G F T F N D Y A M H W V R Q
gcaccgggaaaaggcttgaatgggtttccggcatctcatgggattctagttccattgga
A P G K G L E W V S G I S W D S S S I G
tatgccgattccgtgaaaggaagggtttacgatctcacgagacaatgccaagaacagcctt
Y A D S V K G R F T I S R D N A K N S L
taccttcaaatgaacagctgcgcgccaagatagggcctctattattgtgtgaaagga
Y L Q M N S L R A E D M A L Y Y C V K G
cgggactactatgactccggcgatacttaccggttgctttgatatatggggtcagggg
R D Y Y D S G G Y F T V A F D I W G Q G
acaatggctcactgtgtcatcagcatcaacgaaggccatcagcttcccgcctggccccg
T M V T V S S A S T K G P S V F P L A P
agtagtaagtctacttccggcggaactgccgcactgggctgccttgtaaaggactatttt
S S K S T S G G T A A L G C L V K D Y F
ccggagccagtcacccgtgagttggaacagtggtgcgcttacatctggagtcacatccttt
P E P V T V S W N S G A L T S G V H T F
cctgccgtgttgagtcagtcagtgctttagcttgcctgtggtcacagtcacatct
P A V L Q S S G L Y S L S S V V T V P S
tcctcactgggacacacatatataatgtaacgtgaaccataagccatctaataccaaa
S S L G T Q T Y I C N V N H K P S N T K
gtggataaaaaagtagaccacaagtcctgcggcgcctgggtgccacggggctcccatcac
V D K K V E P K S C G R L V P R G S H H
catcacccatcacccatcacccattga
H H H H H H H H *

3D6 light chain codon-optimized DNA and protein sequence

atggagacagacacactcctgctatgggtactgctgctctgggtccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gacgacatacaaaatgacccaatccccctccactctctcagcctcagtcggagatagggtg
D D I Q M T Q S P S T L S A S V G D R V
actattacgtgcagagcgtcacagagcatttccagatggccttgcgtggtaccagcagaag
T I T C R A S Q S I S R W L A W Y Q Q K
ccgggcaaaagtcgcaaaactgctcatctataaggcgagtagcctggagtcggcgtaacct
P G K V P K L L I Y K A S S L E S G V P
tctaggttctccggctcaggtagtgacggaggttcacacttacaatttcaagtttgcag
S R F S G S G S G T E F T L T I S S L Q
ccggagcagcttccacatactactgtcaacaatataacagttactctttcgggcccggg
P D D F A T Y Y C Q Q Y N S Y S F G P G
actaaggtggacattaagcggaccgtagcggcgccttcagctttatcttcccgcctca
T K V D I K R T V A A P S V F I F P P S
gacgaacagttgaagtcaggtactgagtggtgtgtctccttaataacttttatccg
D E Q L K S G T A S V V C L L N N F Y P
cgagaggctaaagtacagtggaagtcgataacgctctccagagtggaattcacaggaa
R E A K V Q W K V D N A L Q S G N S Q E
tctgtaaccgagcaagattcaaggatagtagctacagtcctgtcatctaccctcactctt
S V T E Q D S K D S T Y S L S S T L T L
tctaaagctgactatgaaaaacacaaggtatagcttgtgaggttaccatcagggactg
S K A D Y E K H K V Y A C E V T H Q G L
tcttcacctgtcaogaagcttttcaataggggtgagtggtga
S S P V T K S F N R G E C -

Supplementary Figure 7. Nucleotide and translated amino acid sequences for Fab 3D6 heavy and light chains. The native 3D6 sequence is colored in black, while regions colored in red, green and blue correspond to the Ig-kappa signal peptide, thrombin cleavage tag, and C-terminal 10-histidine purification tag, respectively.

Supplementary Table 1. Parameters for molecular dynamic simulations

n _{replicates}	Peptide/protein	n _{atoms}	Simulation Time (Analyzed Time) /ns	Footnote
2	GylistTAVPW	9,651	1000 (700)	-
1	Fab 3D6 + WGSSGYLISTTAVPW	106,989	300 (150)	-
1	WGSSGYLISTTAVPW	8,584	1000 (700)	1
2	WGSSGYLISTTAVPW	49,842	1000 (700)	2
1	Fab F240 + WGSSGYLISTTAVPW	98,052	300 (150)	-
3	HIV-1 gp41	163,534	200 (100)	3

¹Initial conformation of the PID peptide was extracted from the complexed Fab structure

²Initial conformation of the PID peptide was linearized

³Initial model was based on a homology model created using Phyre2

Supplementary Movie. Molecular dynamic simulation (100 ns) of the complete HIV-1 gp41 ectodomain in one representative replicate. The video was generated with the PyMol in-build movie generator, with one frame recorded per ns and then overlaid. The PID region is highlighted in red, and an individual chain is highlighted in the chainbow coloring scheme from blue (N-terminus) to red (C-terminus).