

VII

PLV Proteins

Contents

VII-1	Introduction	411
VII-2	Sequences	412
VII-3	Alignments	420
	VII-3.1 Gag	420
	VII-3.2 Pol	424
	VII-3.3 Vif	431
	VII-3.4 Vpx	434
	VII-3.5 Vpr	435
	VII-3.6 Tat	436
	VII-3.7 Rev	438
	VII-3.8 Vpu	440
	VII-3.9 Env	441
	VII-3.10Nef	448

VII-1 Introduction

The selection of Primate Lentivirus protein sequences for the following alignments was based on the sequences in the complete genome alignment as a starting alignment, and complete or nearly complete genes from other isolates were added if they increased the diversity of samples represented.

The annotation is mainly based on knowledge from HIV-1, and should therefore be taken with caution for HIV-2 and SIV sequences.

VII-2 Sequences

Sequences included in the PLV protein alignments.

Name	Accession	Proteins	Author	Reference
H1B.FR.83.HXB2	K03455	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Wong-Staal, F.	<i>Nature</i> 313 (6000):277-284 (1985)
H1A1.UG.85.U455_U455A	M62320	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Oram, J.D.	<i>ARHR</i> 6 (9):1073-1078 (1990)
H1C.ET.86.ETH2220	U46016	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Salminen, M.O.	<i>ARHR</i> 12 (14):1329-1339 (1996)
H1F1.BE.93.VI850	AF077336	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T.	<i>Virology</i> 269 (1):95-104 (2000)
H1O.BE.87.ANT70	L20587	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, M.	<i>J Virol</i> 68 (3):1586-1596 (1994)
H1O.CM.91.MVP5180	L20571	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gurtler, L.G.	<i>J Virol</i> 68 (3):1581-1585 (1994)
H1O.CM.98.98CMA105	AY169816	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Yamaguchi, J.	<i>ARHR</i> 19 (11):979-988 (2003)
H1O.CM.98.98CMABB197	AY169811	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Yamaguchi, J.	<i>ARHR</i> 19 (11):979-988 (2003)
H1O.SN.99.99SE_MP1299	AJ302646	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Toure-Kane, C.	<i>ARHR</i> 17 (12):1211-1216 (2001)
H1O.US.99.99USTWLA	AY169814	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Yamaguchi, J.	<i>ARHR</i> 19 (11):979-988 (2003)
H1N.CM.02.DJO0131	AY532635	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bodelle, P.	<i>ARHR</i> 20 (8):902-908 (2004)
H1N.CM.04.04CM_1015_04	DQ017382	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Yamaguchi, J.	<i>ARHR</i> 22 (1):83-92 (2006)

Name	Accession	Proteins	Author	Reference
H1N.CM.04.04CM_1131_03	DQ017383	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Yamaguchi, J.	<i>ARHR</i> 22 (1):83-92 (2006)
H1N.CM.95.YBF30	AJ006022	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Simon, F.	<i>Nat Med</i> 4 (9):1032-1037 (1998)
H1P.CM.06.U14788	HQ179987	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vallari, A.	<i>J Virol</i> 85 (3): 1403-7 (2011)
CPZ.CD.06.BF1167	JQ866001	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Li, Y.	<i>J Virol</i> 86 (19):10776-10791 (2012)
CPZ.CD.90.ANT	U42720	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, M.M.	<i>Virology</i> 221 (2):346-350 (1996)
CPZ.CM.01.SIVcpzCAM13	AY169968	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Nerrienet, E.	<i>J Virol</i> 79 (2):1312-9 (2005)
CPZ.CM.05.LB715	KP861923	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Barbian, H.J.	<i>MBio</i> 2015 Apr 21;6(2) pii: e00296-15 doi: 10.1128/mBio00296-15
CPZ.CM.05.SIVcpzDP943	EF535993	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Van Heuverswyn, F.	<i>Virology</i> 368 (1):155-171 (2007)
CPZ.CM.05.SIVcpzEK505	DQ373065	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	<i>Science</i> 313 (5786):523-526 (2006)
CPZ.CM.05.SIVcpzLB7	DQ373064	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	<i>Science</i> 313 (5786):523-526 (2006)
CPZ.CM.05.SIVcpzMB66	DQ373063	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	<i>Science</i> 313 (5786):523-526 (2006)
CPZ.CM.05.SIVcpzMB897	EF535994	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Van Heuverswyn, F.	<i>Virology</i> 368 (1):155-171 (2007)
CPZ.GA.88.SIVcpzGAB2	AF382828	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bibollet-Ruche, F.	<i>ARHR</i> 20 (12):1377-1381 (2004)

Name	Accession	Proteins	Author	Reference
CPZ.TZ.00.TAN1	AF447763	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Santiago, M.L.	<i>J Virol</i> 77 (3):2233-2242 (2003)
CPZ.TZ.06.SIVcpzTAN13	JQ768416	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	Unpublished
CPZ.TZ.09.UG38	JN091690	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Rudicell, R.S.	<i>J Virol</i> 85 (19): 9918-28 (2011)
CPZ.US.85.US_Marilyn	AF103818	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	<i>Nature</i> 397 (6718):436-441 (1999)
GOR.CM.04.SIVgorCP684con	FJ424871	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	<i>J Virol</i> 83 (4):1635-1648 (2009)
GOR.CM.07.SIVgor2139_287	FJ424866	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	<i>J Virol</i> 83 (4):1635-1648 (2009)
GOR.CM.07.SIVgorCP2135con	FJ424863	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	<i>J Virol</i> 83 (4):1635-1648 (2009)
GOR.CM.12.SIVgor_BPID1	KP004989	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	D'arc, M.	<i>PNAS USA</i> 112 (11); E1343-52 (2015)
GOR.CM.12.SIVgor_BQID2	KP004991	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	D'arc, M.	<i>PNAS USA</i> 112 (11); E1343-52 (2015)
GOR.CM.13.SIVgor_BPID15	KP004990	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	D'arc, M.	<i>PNAS USA</i> 112 (11); E1343-52 (2015)
MAC.US.x.239	M33262	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kestler, H.	<i>Science</i> 248 (4959):1109-1112 (1990)
H2A.DE.x.BEN	M30502	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kirchhoff, F.	<i>Virology</i> 177 (1):305-311 (1990)

Name	Accession	Proteins	Author	Reference
H2A.PT.x.ALI	AF082339	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Reeves, J.D.	<i>J Virol</i> 73 (9); 7795-804 (1999)
H2A.SN.86.ST_JSP4_27	M31113	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kumar, P.	<i>J Virol</i> 64 (2):890-901 (1990)
H2B.CI.x.EHO	U27200	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Rey-Cuille, M.A.	<i>Virology</i> 202 (1):471-476 (1994)
H2B.GH.86.D205_ALT	X61240	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	<i>Nature</i> 342 (6252):948-950 (1989)
H2G.CI.92.Abt96	AF208027	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Brennan, C.A.	<i>ARHR</i> 13 (5):401-404 (1997)
H2U.CI.07.07IC_TNP3	KC693505	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Ayouba, A.	<i>AIDS</i> 27 (15):2488-2491 (2013)
H2U.FR.96.12034	AY530889	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Damond, F.	<i>ARHR</i> 20 (6):666-672 (2004)
ASC.UG.10.RT03	KJ461716	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Lauck, M.	<i>Retrovirology</i> 11 (1):55 (2014)
COL.CM.x.CGU1	AF301156	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	<i>J Virol</i> 75 (2):857-866 (2001)
COL.UG.10.BWC01	KF214240	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Lauck, M.	<i>Retrovirology</i> 10 (1):107 (2013)
COL.UG.10.BWC07	KF214241	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Lauck, M.	<i>Retrovirology</i> 10 (1):107 (2013)

Name	Accession	Proteins	Author	Reference
DEB.CM.04.SIVdeb04CMPF3061	FJ919724	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Aghokeng, A.F.	<i>Infect Genet Evol</i> 10 (3); 386-96 (2010)
DEB.CM.99.CM40	AY523865	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	<i>J Virol</i> 78 (14):7748-7762 (2004)
DEB.CM.99.CM5	AY523866	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	<i>J Virol</i> 78 (14):7748-7762 (2004)
DRL.DE.11.D3	KM378563	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	<i>ARHR</i> 31 (3); 353-7 (2015)
DRL.DE.11.D4	KM378564	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	<i>ARHR</i> 31 (3); 353-7 (2015)
GRV.ET.x.GRI_677_gri_1	M66437	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Fomsgaard, A.	<i>Virology</i> 182 (1):397-402 (1991)
GSN.CM.99.CN166	AF468659	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	<i>J Virol</i> 76 (16):8298-8309 (2002)
GSN.CM.99.CN71	AF468658	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	<i>J Virol</i> 76 (16):8298-8309 (2002)
LST.CD.88.SIVlhoest485	AF188115	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	<i>J Virol</i> 74 (8):3892-3898 (2000)
LST.CD.88.SIVlhoest524	AF188116	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	<i>J Virol</i> 74 (8):3892-3898 (2000)
LST.KE.x.lho7	AF075269	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	<i>J Virol</i> 73 (2):1036-1045 (1999)
MAL.ZM.x.SIVagmMAL_ZMB	LC114462	Gag, Pol, Vif, Vpr, Env	Carr, M.	<i>Arch Virol</i> 162 (2); 543-548 (2017)
MND-1.GA.x.MNDGB1	M27470	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Tsujimoto, H.	<i>Nature</i> 341 (6242):539-541 (1989)

Name	Accession	Proteins	Author	Reference
MND-2.CM.98.CM16	AF367411	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Takehisa, J.	<i>ARHR</i> 17 (12):1143-1154 (2001)
MND-2.GA.x.M14	AF328295	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Souquiere, S.	<i>J Virol</i> 75 (15):7086-7096 (2001)
MND-2.x.x.5440	AY159322	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	<i>J Virol</i> 77 (8):4867-4880 (2003)
MNE.US.x.MNE027	U79412	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kimata, J.T.	<i>J Virol</i> 72 (1):245-256 (1998)
MON.CM.99.L1_99CML1	AY340701	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	<i>J Virol</i> 77 (23):12523-12534 (2003)
MUS-1.CM.01.CM1239	EF070330	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	<i>Virology</i> 360 (2):407-418 (2007)
MUS-2.CM.01.CM1246	EF070329	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	<i>Virology</i> 360 (2):407-418 (2007)
MUS-3.GA.09.09GabOI81	KF304707	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Liegeois, F.	<i>Viruses</i> 6 (7):2880-2898 (2014)
OLC.CI.97.97CI12	FM165200	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	<i>J Virol</i> 83 (1):428-439 (2009)
RCM.CM.00.SIVagi_00CM312	HM803690	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Ahuka-Mundeke, S.	<i>J Gen Virol</i> 2010 Dec;91(Pt 12):2959-64 doi: 101099/vir0025767-0 Epub 2010 Aug 25
RCM.CM.02.SIVrcm_02CM8081	HM803689	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Ahuka-Mundeke, S.	<i>J Gen Virol</i> 2010 Dec;91(Pt 12):2959-64 doi: 101099/vir0025767-0 Epub 2010 Aug 25

Name	Accession	Proteins	Author	Reference
RCM.GA.x.SIVRCMGAB1	AF382829	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	<i>J Virol</i> 78 (14):7748-7762 (2004)
SAB.SN.x.SAB1	U04005	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Jin, M.J.	<i>EMBO J</i> 13 (12):2935-2947 (1994)
SMM.CI.79.SIVsmCI2	JX860430	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	<i>J Virol</i> 86 (24):13217-13231 (2012)
SMM.SL.92.SIVsmSL92A	JX860432	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	<i>J Virol</i> 86 (24):13217-13231 (2012)
SMM.SL.92.SL92B	AF334679	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Chen, Z.	<i>J Virol</i> 70 (6):3617-3627 (1996)
SMM.US.06.FTq	JX860414	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	<i>J Virol</i> 86 (24):13217-13231 (2012)
SMM.US.11.SIVsmE660_FL8	JQ864086	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Wu, F.	<i>J Virol</i> 86 (16): 8835-47 (2012)
SMM.US.86.CFU212	JX860407	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	<i>J Virol</i> 86 (24):13217-13231 (2012)
SMM.US.x.H9	M80194	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	<i>J Virol</i> 66 (1):414-419 (1992)
SUN.GA.98.L14	AF131870	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	<i>J Virol</i> 73 (9):7734-7744 (1999)
SYK.KE.x.KE51	AY523867	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	<i>J Virol</i> 78 (14):7748-7762 (2004)

Name	Accession	Proteins	Author	Reference
SYK.KE.x.SYK173_COMGNM	L06042	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	<i>J Virol</i> 67 (3):1517-1528 (1993)
TAL.CM.00.266	AY655744	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	<i>Virology</i> 349 (1):55-65 (2006)
TAL.CM.01.8023	AM182197	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	<i>Virology</i> 349 (1):55-65 (2006)
TAN.UG.x.SIVagmTAN1	U58991	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Soares, M.A.	<i>Virology</i> 228 (2):394-399 (1997)
VER.KE.x.AGM155	M29975	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Johnson, P.R.	<i>J Virol</i> 64 (3):1086-1092 (1990)
VER.ZA.10.VSAC4004	KR862336	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bailey, A.L.	<i>J Virol</i> 2016 May 11 pii: JVI00573-16
VER.ZA.10.VSAG1003	KR862363	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bailey, A.L.	<i>J Virol</i> 2016 May 11 pii: JVI00573-16
VER.ZA.10.VSAM0022	KR862356	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bailey, A.L.	<i>J Virol</i> 2016 May 11 pii: JVI00573-16
WRC.CI.98.98CI04	AM713177	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	<i>J Virol</i> 83 (1):428-439 (2009)
WRC.GM.05.Pbt_05GM_X02	AM937062	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Locatelli, S.	<i>Virology</i> 376 (1):90-100 (2008)

H1B.FR.83.HXB2
H1A1.UG.85.U455.U455A
H1E1.ET.86.ETH2220
H1F1.BE.93.V1850
H10.BE.87.ANT70
H10.CM.91.MVP5180
H10.CM.98.98CMA105
H10.CM.98.98CMA187
H10.CM.98.98CMA8195
H10.SN.99.99S5.MP1299
H10.US.99.99TSM1A
H1N.CM.02.J08131
H1N.CM.04.04CM.1015.04
H1N.CM.04.04CM.1131.03
H1N.CM.95.YBF30
H1P.CM.06.U14788
CPZ.CM.06.P1167
CPZ.CM.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.09.UG38
CPZ.TZ.09.UG38
CPZ.TZ.09.UG38
CPZ.US.85.US.Marilyn
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139.287
GOR.CM.07.SIVgorCP135.con
GOR.CM.12.SIVgor.BPID1
GOR.CM.12.SIVgor.BPID2
GOR.CM.13.SIVgor.BPID15

MAC.US.x.239
H2A.DE.x.BEN
H2A.PT.x.ALI
H2A.SN.86.ST.JSP4.27
H2B.CI.x.EH0
H2B.GH.86.0295.ALT
H2G.CI.92.Abt96
H2U.CI.07.07IC.TMP3
H2U.FR.96.12034
ASC.UG.10.RT03
COL.CM.x.C011
COL.UG.10.BWC01
COL.UG.10.BWC07
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DRL.DE.11.D3
DRL.DE.11.D3
DRL.DE.11.D4
GRV.ET.x.GRI.677.gri.1
GSN.CM.99.CN166
GSN.CM.99.CN171
LST.CD.88.SIVLhoest485
LST.CD.88.SIVLhoest524
LST.KE.x.lho7
MAL.ZM.x.SIVagmMAL.ZMB
MND.1.GA.x.MNDCB1
MND.2.GA.x.CM16
MND.2.GA.x.M14
MND.2.x.x.5440
MNE.US.x.MNE027
MON.CM.99.L1.99CML1
MUS.1.CM.01.CM1239
MUS.2.CM.01.CM1246
MUS.3.GA.09.09gab081
OLC.CI.97.97C112
RCM.CM.00.SIVa1.00CM312
RCM.CM.02.SIVrcm.02CM8081
RCM.GA.x.SIVrcmGAB1
SAB.SN.x.SAB1
SMM.CI.79.SIVsmC12
SMM.SL.92.SIVsmS192A
SMM.SL.92.SIVS192B
SMM.US.06.CFU212
SMM.US.11.SIVsmE660.FL8
SMM.US.86.CFU212
SMM.US.x.H9
SUN.GA.98.L14
SVK.KE.x.KC51
SVK.KI.x.SVK173.COMGNM
TAL.CM.00.266
TAL.CM.01.8023
TAN.UG.x.SIVagmTAN1
VER.KE.x.AGM155
VER.ZA.10.VSA4094
VER.ZA.10.VSA1093
VER.ZA.10.VSAM0022
WRC.CI.98.RC104
WRC.GM.05.Pbt.05GM.X02

membrane binding
Gag start, p17 start
MGARASV.LSSGGLDRWEKIRLRPGGKKYKLVKHIVASRELERFAVNPGLLETSEGRQILGLOPSLQTSSEELRSYNTVATLYCVHQRIEIKDTKEALDKIEEEO.....NKSKKKAQAAADTG.HSNVQSNYPVIONIG 140

phosphorylation site
p17 end_p24 start

Gag p15 start
p15 end_p27 start

membrane binding

Table with columns for protein names, amino acid sequences, and accession numbers. The table is organized into two main sections: 'CyPA binding' and 'major homology region'. It lists various HIV-1 sequences from different sources and their corresponding amino acid alignments.

H1B, FR. 83, HXB2
H1A1.UG.85.U455.U455A
H1C.ET.86.ETH2220
H1F1.BE.93.VI850
H10.BE.87.ANT70
H10.CM.91.MVP5180
H10.CM.98.98CMA105
H10.CM.98.98CMA197
H10.CM.98.98CMA8195
H10.SN.99.99SE.MP1299
H10.US.99.99STWLA
H1N.CM.02.DJ08131
H1N.CM.04.04CM.1015.04
H1N.CM.04.04CM.1131.03
H1N.CM.95.YBF30
H1P.CM.06.U14788
CPZ.CD.06.BT167
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.09.TAN1
CPZ.TZ.06.SIVcpzTAN13
CPZ.TZ.09.UG38
CPZ.US.85.US.Marilyn
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139.287
GOR.CM.07.SIVgorCP2135con
GOR.CM.12.SIVgor.BPID1
GOR.CM.12.SIVgor.BOID2
GOR.CM.13.SIVgor.BPID15

MAC.US.x.239
H2A.DE.x.BEN
H2A.PT.x.ALI
H2A.SN.86.ST.JSP4.27
H2B.CI.x.EH0
H2B.GH.86.0295.ALT
H2G.CI.92.Abt96
H2U.CI.07.07IC.TMP3
H2U.FR.96.12034
ASC.UG.10.RT03
COL.CM.x.G011
COL.UG.10.BWC01
COL.UG.10.BWC07
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM49
DRL.DE.11.D3
DRL.DE.11.D4
GRV.ET.x.GRI.677.gri.1
GSN.CM.99.CN166
GSN.CM.99.CN171
LST.CD.88.SIVlhoest485
LST.CD.88.SIVlhoest524
LST.KE.x.lho7
MAL.ZM.x.SIVagmMAL.ZMB
MND.1.GA.x.MNDGB1
MND.2.CM.98.CM16
MND.2.GA.x.M14
MND.2.x.x.5440
MNE.US.x.MNE027
MON.CM.99.L1.99CML1
MUS.1.CM.01.CM1239
MUS.2.CM.01.CM1246
MUS.3.GA.09.09gab0181
OLC.CI.97.97C112
RCM.CM.00.SIVagi.00CM312
RCM.CM.02.SIVrcm.02CM8081
RCM.GA.x.SIVRCMGA81
SAB.SN.x.SAB1
SMM.CI.79.SIVsmC12
SMM.SL.92.SIVsmSL92A
SMM.SL.92.SIVsmSL92B
SMM.US.06.FIC0
SMM.US.11.SIVsmE660.FL8
SMM.US.86.CFU212
SMM.US.x.H9
SUN.GA.98.L14
SVK.KE.x.KC51
SVK.KE.x.SVK173.COMGNM
TAL.CM.00.266
TAL.CM.01.8023
TAN.UG.x.SIVagmTAN1
VER.KE.x.AGM155
VER.ZA.10.VSAG4004
VER.ZA.10.VSAG1093
VER.ZA.10.VSAG0022
WRC.CI.98.98CI04
WRC.GM.05.Pbt.05GM.X02

p24 end_p2 start
RAEQASQEVKNWMTETLLVQNANPDCKITLKALGPAATLEEMMTACQGVGPGHKARVLAEAMSOVTVS.AT...IMMQRGNFRNOR...KIVKCFNCGKEGHTARNCRAPRKK...GWCKCGKEGHOMKDOCT. 427

p27 end_p2 start
---TDA--Q-I-LV-G-VNP-L-Q-LM-LKEALAP.VTIP.FAAA-ORGP...R-PT-W-S-Q--RQ...MD-V-AK-P. 428

p27 end_p2 start
---TDA--Q-I-LV-G-VNP-L-Q-LM-LKEALAP.VTIP.FAAA-ORGP...R-PT-W-S-Q--RQ...MD-V-AK-P. 428

p2 end_p8 start
---TDA--Q-I-LV-G-VNP-L-Q-LM-LKEALAP.VTIP.FAAA-ORGP...R-PT-W-S-Q--RQ...MD-V-AK-P. 428

Table with columns for protein names, p7 end p1 start, p1 end p6 start, Vpr binding, PSAP in HIV-2 B, Vpr binding p6 end, Gag end, and protein IDs (500-515). Rows include sequences for various HIV strains and proteins like ERQANFLGKI, WPSYKGRPGNLF, QSRPE, etc.

Pol start
Gag-Pol TF start

Gag-Pol TF end protease start

H1B.FR.83.HXB2	FFREDLALFLOGKA	REFSSEQT	RANSPTRRELOVWGRDN	NSPSEAG	ADROGT	VSFNFQVTLWQRPLVTLIKIGGOLKEALDITGADDTVLEEMSL	PGRWKPKMIGGIGGFIKRVQYD01LIEICGH
H1A1.UG.85.U455.U455A	N	Q	E	S	W	D	G
H1C.ET.86.ETH2220	N	Q	E	S	W	D	G
H1F1.BE.93.V1850	N	Q	E	S	W	D	G
H10.95.99.99T870	Q1	SGGHE	OLCA	TS	TPISPTDGGGSE	TGESGT	ERPE
H10.CM.91.MVP5180	V	SGGHE	OLCA	TS	VPISTPTGGGSE	TRES	EGEESG
H10.CM.98.98CMA105	Q1	SGGHE	OLCA	TS	APLSPDTGGGNG	TGESGT	EGESE
H10.CM.98.98CMA8197	Q1	SGGHE	OLCA	TSAPS	PISPTDGGGNG	TGESGT	E.PE
H10.SN.99.99SE.MP1299	T	SGGHE	OLCA	TS	VPISTPTDGGGNE	TRESGT	EGGPE
H10.US.99.99STWLA	Q	Y	ANAPS	SPISTPTGGGSE	AX	YEGSGG	
H1N.CM.02.DJ08131	KG	VS	RET	LPDPNN	KE	H	AT
H1N.CM.04.04CM.1015.04	KG	VS	RET	LPDPNN	KE	H	AT
H1N.CM.04.04CM.1131.03	G	VS	RET	L	PDNN	KE	H
H1N.CM.05.YBF30	E	VS	RET	K	LPDPNN	KE	H
H1N.CM.05.U14788	CS	SGGQ	OLCA	TS	TIDSPTGGGDD	SEAGG	KQ
CPZ.CD.06.BF1167	THPL	VGVT	K	CA	SESDES	GT	ESGES
CPZ.CD.09.ANT	TDPHVV	VOT	L	CA	GGG	GANS	THR
CPZ.CM.01.SIVcpzCAM13	G	P	RE	L	CT	D	S
CPZ.CM.05.LB715	N	S	RE	L	CT	D	S
CPZ.CM.05.SIVcpzDP943	R	VS	VERGI	K	TREI	PAG	EGTH
CPZ.CM.05.SIVcpzEK505	G	S	RE	K	PPDMN	KE	SN
CPZ.CM.05.SIVcpzLB7	T	P	RE	K	L	AN	W
CPZ.CM.05.SIVcpzMB66	T	O	N	P	E	H	S
CPZ.CM.05.SIVcpzMB897	N	SP	RE	PP	E	RTT	TKC
CPZ.GA.88.SIVcpzGAB2	GG	PERE	L	K	PEGG	GT	S
CPZ.TZ.09.TG70	THPL	VGVT	K	CA	HP	EGE	AP
CPZ.TZ.09.UG38	THPL	VGVT	CA	HP	EGE	SGT	D
CPZ.US.05.U14788	T	VP	IVERGI	K	TREI	PAG	EGTH
GOR.CM.04.SIVgorCP684con	I	SRGRE	OLCA	AS	ATNSPTNGGGRG	NEE	TRKGGT
GOR.CM.07.SIVgorCP2139.287	I	SRGRE	OLCA	AS	ATNSPTNGGGRG	DEE	TRKGGT
GOR.CM.07.SIVgorCP2135con	I	SRGRE	OLCA	AS	ATNSPTNGGGRG	DEE	TRKGGT
GOR.CM.12.SIVgorBPID1	V	SGGHE	OLCA	AS	TTNSPTNGGSD	PE	QTE
GOR.CM.12.SIVgorBQID2	V	SGGHE	OLCA	AS	TTNSPTNGGSD	PE	QTE
GOR.CM.13.SIVgorBPID15	N	SGGQ	OLCA	AS	AANSPTNGGSD	PE	QTE

Pol p15 start

p15 end protease start

MAC.US.x.239	PWSMG	KE	P	PHGSSASGAD	ANCS	RGPCSGSAKELHAVQQAERKA	RKOREALOGDRG
H2A.DE.x.BEN	VGPTG	KE	S	QLPRDPPSPGAD	TNSTSPR	SSSGTIVGEIYAAREKA	EGEATIRQDGGGLA
H2A.PT.x.ALI	AWPMG	KE	S	QLPRNPPSSAGIN	TNSTSPR	ASSGPAAGVYAAEKAKRA	ERAIOGDGGGLA
H2A.SN.86.ST.JSP4.27	VGPMG	KE	P	PCGPNPAGAD	TNSTPDR	PSRPTPREVHAAREKA	ERAIOJRSDRGLPAAR
H2B.CI.x.EH0	VRLG	KE	S	Q	PRG	TPG	AICAPTEP
H2C.GH.86.D295.1LT	VRLG	KE	S	Q	PRG	TPG	AICAPTEP
H2C.CI.92.Abt96	VWTLG	KE	S	Q	PHDPPSPGSD	SISTT	DGSPSRPEELHAXGEAKG
H2U.CI.07.071C.TMP3	VWPMG	KE	S	Q	PHDPPSPGSD	TNSA	SSRTPSSGADPSPSSGSAKELHAGTQETK
H2U.FR.96.12034	AWPLG	K	S	Q	PHGSPSPGAD	PNSP	SSRTPSSGADPSPSSGSAKELHAGTQETK
ASC.UG.10.RT03	GFR	Q	ASG	AQEGQEL	PSSTAISSTFDAGRHFSGGEL	EDORATD	RR
COL.CM.x.C0U1	GMOG	ATD	S	HLPP		F	EGGE
COL.UG.10.BWC01	GRGD	RP	D	NL	PS	NEODAAGT	EERREH
COL.UG.10.BWC07	DTRGD	PT	D	SCLPPK		GDGOGNLSLLETP	R
DEB.CM.04.SIVdeb04CMPF3061	TSLG	EESP	QF	HRRAAT	DMGHREVTGGGRD	LLLFQGTGTG	PAAGGDEERTPT
DEB.CM.99.CM40	YSLG	LOET	QELSRNRN	EV	HPGHS	SGILD	YOGTSSGEGOPPE
DRL.CM.99.CM70	YSLG	LPET	QELSRNRN	EV	HPRHS	SGLFN	YOGTSSGEGOPPE
DRL.DE.11.D3	YSLG	ORET	QEF	GL		ADPISSSIARIPPGSGREDAPKVHGAGSAEETT	TATAEE
DRL.DE.11.D4	YSLG	ORET	QEF	GL		ADPISSSIARIPPGSGREDAPKVHGAGSAEETT	TATAEE
GRV.ET.x.GRI.677.gri.1	VWPLG	RSET	KKF	CAIQRHRSWGTNSP	NGNSLRSSKEAPPVCREGTAEPRE	RTDKETE	ERSGG
GSN.CM.99.CM166	PWPGA	PPSP	Q	L	NFPAA	NSPG	GGRTPA
GSN.CM.99.CM171	PWPGA	TSP	Q	L	NFPAA	ISPG	GGGTPA
LST.CD.88.SIVhoest485	LWTL	ENSS	KL	FDGGE	CRS	SISPP	NGKELNKSRESLNRIQEPRTTTEE
LST.CD.88.SIVhoest524	VWLSL	EWTF	LQ	NHGRG	CRS	DSFSA	HGESASMSKRESLNRIQEPRTTTEE
LST.KE.x.lho7	IWLSL	QTSR	QLS	TGRS	CRE	SISPP	NGESASMSKRESLNRIQEPRTTTEE
MAL.ZM.x.SIVagmMAL.ZMB	VWMPD	YEA	EKFSRRHSWGG	TPAQCPISIDG		LRPCKEAPTTI	CREKTIETEAR
MND.1.GA.x.MNDGB1	VWPLG	SLQT	G	L		GTRDGSNSSTIREGTSAEINSEITERA	QAE
MND.2.CM.98.CM16	YPLG	QWOT	Q	LP		SNADGRSRENAVRLHERATAESGKO	IGKE
MND.2.GA.x.M14	CSLG	QWQA	Q	L		DARDGSSREDVIRLYKEHSTEGSKGD	QAGEG
MND.2.x.x.5440	YPLG	QWQA	Q	LP		CHAIPISTP	DARVGRPKRDVAVRLEYAERTTEGSR
MNE.US.x.MNE027	LWPMG	KE	P	PHGNASGAD	TNCSP	RGPCSGSAKELHAVQQAERKA	RKOREALOGDRG
MON.CM.99.L1.99CML1	SFRS	QEGA	K	LP	HLCAA	ORTDASA	EDG
MUS.1.CM.01.CM1239	GSSWP	PE	EA	K	LPD	LAFI	NGPARAGRPQGSILA
MUS.2.CM.01.CM1246	GSRP	PVSEA	E	LP	YIAHA	YGTPTRY	AGELLPTSRAG
MUS.3.GA.09.09GAB0181	SFW	PE	GT	K	LP	HPVGA	DG
OLC.CI.97.97C112	MEGR	ESRASPHATEDRDG	HL	LCTYGP		TDGEGMG	AEEGE
RCM.CM.00.SIVa1.00CM312	CSLE	AGET	E	NFLR	D	TLCRGG	SVPRCGSHVKEITL
RCM.CM.02.SIVrcm.02CM8081	KCALE	SGET	E	NFDFL	D	SLCRGG	GSQSLPVRSRSHAKEPTGQNAETEEG
RCM.GA.x.SIVRCMGAB1	TPLG	SGEA	E	NFDFL	D	SLCRGG	GSQSLPVRSRSHAKEPTGQNAETEEG
SAB.SN.x.SAB1	VWPLG	ORET	Q	EPFL	D	LHOTN	SSPNG
SHM.CI.79.SIVsmC12	IWPMG	KE	S	Q	PHDPP	PAGAD	NGAPNCSGCA
SHM.SL.92.SIVsmS192A	IWSLG	KE	S	Q	PHDPP	NID	TDSP
SHM.SL.92.SIVsmS192B	WPMG	KE	S	Q	PHDPP	NID	TDSP
SHM.US.06.CM1246	WPMG	KE	S	Q	PHDPP	NID	TDSP
SHM.US.11.SIVsmE660.FL8	AWPMG	KE	S	Q	PHDPP	NID	TDSP
SHM.US.86.CFU212	XWPMG	KE	S	Q	PHDPP	NID	TDSP
SHM.US.x.H9	AWPMG	KE	S	Q	PHDPP	NID	TDSP
SUN.GA.98.L14	VWPL	UG	A	L	HAFCS	A	SRGSDIGEGMDT
VK.KE.x.KC51	RVGE	QETP	SK	CD	EGDT	NSPTSGGLA	MAATDMGK
VYK.KC.x.SYK173.COMGM	RMG	LQ	PS	OLP	SEEARICS	P	GHH
TAL.CM.00.100263	VYFR	ENPT	G	ORTE	FS	SGATIGTAN	RASHISIMAGDGHYDPDTSO
TAL.CM.01.8023	GFR	EDPP	G	GTTE	LP	SPG	ATHRAD
TAN.UG.x.SIVagmTAN1	NANL	NKA	AQF	RARRSRSDGSDACA	IPH	RFASCGSIRPSQETAGAVCKQK	RSTKEAE
VER.KE.x.AGM155	VWVD	DKT	KKFSR	HWSGG	TKCAPSTEG	LYTLRPSKEAPAVCRERETNEKSEQ	KPSPSE
VER.ZA.10.VS4C094	VWVD	DT	EKFSRRHSWGG		RSNL	SHCSPTDGS	IRPQEAAPATCRSRETAGT
VER.ZA.10.VS4A093	VWVD	EQA	EELPR	HSGGG	S	N	D
VER.ZA.10.VS4M022	VWTVG	EQT	EKFSRRHSWGG		W	E	F
WRC.CI.98.CI84	WERA	DALK	PLSSK	ELEG	FWAG	NNTVCSNDEGSKH	TRIREIPPSRSTEGAGTK
WRC.GM.05.Pbt.05GM.X02	WPR	AE	LQOL	SS		KN	CRGEYLWASG

Table of HIV protease end p66, p51 RT start sequences. Columns include sequence identifiers (e.g., H1B, H1A1, H1C), amino acid sequences, and protein names (protease end p66, p51 RT start, M41L, K70R D67N, D110 catalytic site). Residue numbers 281-376 are listed on the right.

Table of HIV protease end p51 RT start sequences. Columns include sequence identifiers (e.g., MAC, H2A, H2A), amino acid sequences, and protein names (protease end p51 RT start, D catalytic site). Residue numbers 293-396 are listed on the right.

Table with columns for protein names (e.g., H1B, H1A1, H1C), sequence alignments, and accession numbers. Includes a section for p51 RT end_p15 RNase H start.

Table with columns for protein names (e.g., MAC, H2A, H2B), sequence alignments, and accession numbers. Includes a section for p51 RT end_p15 RNase H start.

H1B.FR.83.HXB2
H1A1.UG.85.U455.U455A
H1C.ET.86.ETH2220
H1D.BE.93.VI850
H1E.87.AMT70
H10.CM.91.MVP5180
H10.CM.98.98CMA105
H10.CM.98.98CMA8197
H10.CM.98.98CMA8197
H10.SN.99.99SE.MP1299
H10.US.99.99STWLA
H1N.CM.02.DJ08131
H1N.CM.04.04CM.1015.04
H1N.CM.04.04CM.1131.03
H1N.CM.95.YBF38
H1P.CM.06.U14788
CPZ.CM.06.BF1167
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.GA.88.SIVcpzGAB2
YRSJ.TA.09.WD
CPZ.TZ.09.UG38
CPZ.US.85.US.Marilyn
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139.287
GOR.CM.07.SIVgorCP2135con
GOR.CM.12.SIVgor.BPID1
GOR.CM.12.SIVgor.BPID2
GOR.CM.13.SIVgor.BPID15

RQKVVTLTDTTNDKTELQAIYALALQDSGLEVINIVTDSQYALGIQAQPDQSESELVNIIIEQLIKKVKYLWVPAHKGIGGNEQDKLVSA GIER. KVLFLDGDIAQDQHEKHYKSHNWRAMASDFNLPPVVAKEIVASCDKCLQK. GEAMHGQDVCSPGIWLDCTHLE
H1B.FR.83.HXB2
H1A1.UG.85.U455.U455A
H1C.ET.86.ETH2220
H1D.BE.93.VI850
H1E.87.AMT70
H10.CM.91.MVP5180
H10.CM.98.98CMA105
H10.CM.98.98CMA8197
H10.CM.98.98CMA8197
H10.SN.99.99SE.MP1299
H10.US.99.99STWLA
H1N.CM.02.DJ08131
H1N.CM.04.04CM.1015.04
H1N.CM.04.04CM.1131.03
H1N.CM.95.YBF38
H1P.CM.06.U14788
CPZ.CM.06.BF1167
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.GA.88.SIVcpzGAB2
YRSJ.TA.09.WD
CPZ.TZ.09.UG38
CPZ.US.85.US.Marilyn
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139.287
GOR.CM.07.SIVgorCP2135con
GOR.CM.12.SIVgor.BPID1
GOR.CM.12.SIVgor.BPID2
GOR.CM.13.SIVgor.BPID15

MAC.US.x.239
H2A.DE.x.BEN
H2A.PT.x.ALI
H2A.SN.86.ST.JSP4.27
H2B.CI.x.EH0
H2B.GH.86.D295.ALT
H2G.CI.92.Abt96
H2U.CI.07.07IC.TMP3
H2U.FR.96.12034
ASC.UG.10.RT03
COL.CM.x.G001
COL.UG.10.BWC01
COL.UG.10.BWC07
DEB.CM.04.SIVdeb04CMP3061
DEB.CM.99.CM40
DEB.CM.99.CM40
DRL.DE.11.D3
DRL.DE.11.D4
GRV.ET.x.GRI.677.gri.1
GSN.CM.99.CN166
GSN.CM.99.CN171
LST.CD.88.SIVhoest485
LST.CD.88.SIVhoest524
LST.KE.x.lho7
MAL.ZM.x.SIVagmMAL.ZMB
MND.1.GA.x.MNDGB1
MND.2.CM.98.CM16
MND.2.GA.x.M14
MND.2.x.x.5440
MNE.US.x.MNE027
MON.CM.99.L1.99CML1
MUS.1.CM.01.CM1239
MUS.2.CM.01.CM1246
MUS.3.GA.09.09gab0181
OLC.CI.97.97C112
RCM.CM.00.SIVag1.00CM312
RCM.CM.02.SIVrcm.02CM8081
RCM.GA.x.SIVRCMGAB1
SAB.SN.x.SAB1
SMM.CI.79.SIVsmC12
SMM.SL.92.SIVsmS192A
SMM.SL.92.S192B
SMM.SL.92.S192B
SMM.US.11.SIVsmE660.FL8
SMM.US.86.CFU212
SMM.US.x.H9
SUN.GA.98.L14
SVK.KE.x.KC51
VER.ZA.10.VSA40904
VER.ZA.10.VSA61093
VER.ZA.10.VSA0022
WRC.CI.98.98CI04
WRC.GM.05.Pbt.05GM.X02

KD-KV.EQ-QA-E.FLM-T.PKA-IV-VM-TGC-TE-R-EM-SEI-V-0-0-EK-EP-E-D-VKELVFK-G-RI-RQ-DT-HO-I-ANSDL-T-M-795
KD-KV.EQ-QA-EVFERM-A-PK-IV-VM-VAG-TE-NRI-EM-A-V-0-0-EK-EP-E-D-IIKELTHK-GI-LL-RQ-NT-AQ-Q-I-NAEL-V-M-Y-808
-D-KV.EQ-QA-E.FA-VT-PKA-IV-VM-VAG-TE-NRI-EM-AI-V-0-0-EK-EP-E-D-VKELSHK-G-RL-RQ-NT-AQ-Q-I-NAEL-T-M-808
-D-RL.EQ-QA-E.FAM-VT-PKA-IV-VM-VAG-TE-KI-EM-AI-V-0-0-EK-EP-E-D-VKELSHK-G-KL-RQ-NT-AQ-Q-I-NAEL-T-M-808
KD-KV.EQ-QA-E.FA-PO-TV-VM-VAG-TE-TI-EM-AI-VG-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-808
KD-KV.EQ-QA-E.FA-T.EPO-IV-VM-VAG-TE-TI-EM-AI-VG-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-807
KD-KV.EQ-QA-E.FAM-PK-IV-VM-VAG-TE-P-K-EM-XL-VG-0-0-EK-EP-E-D-VKELTHK-GI-QL-RQ-N-Q-I-NAEL-T-M-791
KD-KV.EQ-QA-E.FLM-SQA-V-VM-VAG-TE-P-K-EM-XL-VG-LE-H-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-791
KDR.KP.EQ-QA-E.FAM-E-PK-IM-AG-TE-T-P-SK-E-R-A-VG-0-0-EK-EP-E-D-VKELQHK-GI-QL-RQ-N-Q-I-NAEL-T-M-804
KE-A.EH-A-L-LL-G.TKO-T-VM-VAG-TE-TI-EM-AI-VG-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-793
BYE-N.HEGS-QA-TGL-KHGPKRM-L-L-V-TG-YOQLPSSRGE-T-OAGMA-AIHWX-C-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-792
EEE-I.HEGS-QA-R.VLM-EHGPKRM-L-L-F-VTG-EE-N-P-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-764
EEIEEV.HEGS-QA-KG-LM-KHGPKRM-L-L-VSG-EEEN-P-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-764
K-S.KA.EN-A-T.VLM-E.EK-L-VN-LTHE-TTT-H-EK-O-Q-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-792
K-S.KC.EN-A-E-L.EGPKSM-L-VM-LT-CE-V-T-A-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-789
K-S.KS.EN-A-E-L.EGPKSM-L-VM-LT-CE-V-T-A-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-789
KE-IA.EE-A-L-K-PK-M-SSS-ET-DNPI-M-S-A-N-N-E-R-O-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-798
KE-IA.EE-A-L-K-PK-M-SSS-ET-DNPI-M-S-A-N-N-E-R-O-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-798
KE-IA.EE-A-L-K-PK-M-SSS-ET-DNPI-M-S-A-N-N-E-R-O-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-798
KSRIRI.EE-QA-T.VKM-E-EN-VMN-LT-CE-V-T-A-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-806
KYRAIE.EE-QA-H.VL-KEGPPKM-L-VM-LT-CE-V-T-A-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-795
KYRAIE.EE-QA-H.VL-KEGPPKM-L-VM-LT-CE-V-T-A-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-788
K-KS.EE-QA-E.VL-K-KSR-VMKVSOR-TET-HPV-K-KE-O-DA-G-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-803
KE-KE-QA-E.VL-K-KSR-VMKVSOR-TET-HPV-K-KE-O-DA-G-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-803
K-R.K-EN-QA-E.LM-E-SK-LT-H-T-D-A-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-800
K-R.K-EN-QA-E.LM-E-SK-LT-H-T-D-A-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-800
K-F.EEVLN-Q-H.VK-F-SY-VV-LASR-TETDPTI-KE-LMKG-I-S-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-793
KSR-IA.EN-A-E-KM-F-PR-M-LS-A-DNPI-RE-LM-H-A-A-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-796
KSR-IA.EN-A-E-L-T-R-DNPI-RE-N-M-A-A-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-797
KS-A.EE-A-T-KM-PR-LS-A-DNPI-RE-LM-G-TEI-S-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-797
KD-RV.EQ-QA-E.FLM-T.PKA-IV-VM-TGC-TE-R-EM-SEI-V-0-0-EK-EP-E-D-VKELVFK-G-RI-RQ-DT-HO-I-ANSDL-T-M-791
AR-A.EQ-A-E.VL-R-GPPQM-L-V-LASC-ET-PI-EA-0-L-DO-F-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-791
KE-S.EN-A-E.VL-KEGPPSM-I-V-VAS-0-ET-P-EE-Q-LT-A-S-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-803
QERAIN.EN-QA-L-M-GPP-M-L-V-T-A-VE-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-790
-AID.EN-QA-L-K-GPPKM-L-V-T-CE-A-P-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-786
KH-I.HQESS-A-IR-IE-R-KS-L-M-VLEK-EV-EP-IQEL-KE-E-RI-IG-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-785
KE-KE.EE-A-E.VL-K-KR-VF-LAGS-T-P-O-G-E-S-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-797
K-R.K-EN-A-E.VL-K-KR-VF-LAGS-T-P-O-G-E-S-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-803
K-R.K-EN-A-E.VL-K-KR-VF-LAGS-T-P-O-G-E-S-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-797
D-A-EN-A-E-L-R-SK-I-M-AGE-TE-DNPI-O-E-A-I-V-EI-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-816
E-KP.EQ-QA-E-AP-R-A-PKA-IV-VM-VAGH-TETD-P-E-E-AI-VG-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-800
KD-KV.EQ-QA-E.FAM-A-PKA-IV-VM-VLG-TE-PI-E-NHMLL-EL-V-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-790
KF-L.EQA-QA-E-FM-VT-A-IV-V-N-TG-SE-D-T-A-VQ-A-IG-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-788
KF-L.EQA-QA-E-FM-VT-A-IV-V-N-TG-SE-D-T-A-VQ-A-IG-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-788
D-ANL.EQ-QA-E-FM-VT-A-IV-V-N-TG-SE-D-T-A-VQ-A-IG-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-791
D-ANL.EQ-QA-E-FM-VT-A-IV-V-N-TG-SE-D-T-A-VQ-A-IG-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-791
KDR-KV.EQ-QA-E.FLM-T.PKA-IV-VM-VAG-TE-NKI-S-EM-TAI-VT-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-791
G-TKL.EQ-QA-E-FM-VT-A-IV-V-N-TG-SE-D-T-A-VQ-A-IG-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-791
SR-KK.EN-QA-E-KM-E-RSS-RLLSKR-TETD-KE-V-LIRO-DO-G-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-803
K-TLKR.EN-A-E.VLM-LMWC-TS-HPV-E-MOKAME-I-TT-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-823
T-R.OY-QA-EGLLH-KDK-V-SY-LMTC-TNT-PI-E-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-797
KEH-IK.EK-A-E-LM-K-PK-VY-LAGS-TETDNKI-ED-LS-A-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-792
KEH-G-SN-A-E-VL-KH-PK-VY-LVGN-TETDNKI-ED-LS-A-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-793
ESO-IE.EE-QA-E-VL-KH-KSK-VM-LT-T-D-V-EM-N-DAI-T-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-803
K-R.OQ.EN-QA-T-KM-E-PK-M-LT-T-D-P-E-A-MVO-AT-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-803
VER-ZA.10.VSA40904-K-R-KE-EN-QA-T-LM-K-PK-VY-LAGS-TETDNKI-ED-LS-A-L-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-801
K-I-EN-QA-H-VL-KR-V-SSH-T-D-P-E-A-Q-L-N-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-803
K-RIR-EN-QA-VLM-E-KO-I-V-VSSH-T-D-P-E-AA-O-A-N-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-803
WRC-IN.YN-QA-A-LK-E-G-SKI-I-VM-VKDH-T-A-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-807
-N.IQKI-N-Q-I-LRM-E-SDK-I-V-KDH-T-P-MQE-DL-KG-EI-TT-0-0-EK-EP-E-D-VKELVFK-GI-QL-RQ-N-Q-I-NAEL-T-M-799

p66 RT, p15 RNase H end, p31 Integrase start

p15 RNase H end, p31 Integrase start

PLV Proteins

Pol

Table of HIV sequences for PLV Proteins, including identifiers like H1B, H1A1, H1C, H1F1, H1G, H10, H11, H12, H13, H14, H15, H16, H17, H18, H19, H20, H21, H22, H23, H24, H25, H26, H27, H28, H29, H30, H31, H32, H33, H34, H35, H36, H37, H38, H39, H40, H41, H42, H43, H44, H45, H46, H47, H48, H49, H50, H51, H52, H53, H54, H55, H56, H57, H58, H59, H60, H61, H62, H63, H64, H65, H66, H67, H68, H69, H70, H71, H72, H73, H74, H75, H76, H77, H78, H79, H80, H81, H82, H83, H84, H85, H86, H87, H88, H89, H90, H91, H92, H93, H94, H95, H96, H97, H98, H99, H100, H101, H102, H103, H104, H105, H106, H107, H108, H109, H110, H111, H112, H113, H114, H115, H116, H117, H118, H119, H120, H121, H122, H123, H124, H125, H126, H127, H128, H129, H130, H131, H132, H133, H134, H135, H136, H137, H138, H139, H140, H141, H142, H143, H144, H145, H146, H147, H148, H149, H150, H151, H152, H153, H154, H155, H156, H157, H158, H159, H160, H161, H162, H163, H164, H165, H166, H167, H168, H169, H170, H171, H172, H173, H174, H175, H176, H177, H178, H179, H180, H181, H182, H183, H184, H185, H186, H187, H188, H189, H190, H191, H192, H193, H194, H195, H196, H197, H198, H199, H200, H201, H202, H203, H204, H205, H206, H207, H208, H209, H210, H211, H212, H213, H214, H215, H216, H217, H218, H219, H220, H221, H222, H223, H224, H225, H226, H227, H228, H229, H230, H231, H232, H233, H234, H235, H236, H237, H238, H239, H240, H241, H242, H243, H244, H245, H246, H247, H248, H249, H250, H251, H252, H253, H254, H255, H256, H257, H258, H259, H260, H261, H262, H263, H264, H265, H266, H267, H268, H269, H270, H271, H272, H273, H274, H275, H276, H277, H278, H279, H280, H281, H282, H283, H284, H285, H286, H287, H288, H289, H290, H291, H292, H293, H294, H295, H296, H297, H298, H299, H300, H301, H302, H303, H304, H305, H306, H307, H308, H309, H310, H311, H312, H313, H314, H315, H316, H317, H318, H319, H320, H321, H322, H323, H324, H325, H326, H327, H328, H329, H330, H331, H332, H333, H334, H335, H336, H337, H338, H339, H340, H341, H342, H343, H344, H345, H346, H347, H348, H349, H350, H351, H352, H353, H354, H355, H356, H357, H358, H359, H360, H361, H362, H363, H364, H365, H366, H367, H368, H369, H370, H371, H372, H373, H374, H375, H376, H377, H378, H379, H380, H381, H382, H383, H384, H385, H386, H387, H388, H389, H390, H391, H392, H393, H394, H395, H396, H397, H398, H399, H400, H401, H402, H403, H404, H405, H406, H407, H408, H409, H410, H411, H412, H413, H414, H415, H416, H417, H418, H419, H420, H421, H422, H423, H424, H425, H426, H427, H428, H429, H430, H431, H432, H433, H434, H435, H436, H437, H438, H439, H440, H441, H442, H443, H444, H445, H446, H447, H448, H449, H450, H451, H452, H453, H454, H455, H456, H457, H458, H459, H460, H461, H462, H463, H464, H465, H466, H467, H468, H469, H470, H471, H472, H473, H474, H475, H476, H477, H478, H479, H480, H481, H482, H483, H484, H485, H486, H487, H488, H489, H490, H491, H492, H493, H494, H495, H496, H497, H498, H499, H500, H501, H502, H503, H504, H505, H506, H507, H508, H509, H510, H511, H512, H513, H514, H515, H516, H517, H518, H519, H520, H521, H522, H523, H524, H525, H526, H527, H528, H529, H530, H531, H532, H533, H534, H535, H536, H537, H538, H539, H540, H541, H542, H543, H544, H545, H546, H547, H548, H549, H550, H551, H552, H553, H554, H555, H556, H557, H558, H559, H560, H561, H562, H563, H564, H565, H566, H567, H568, H569, H570, H571, H572, H573, H574, H575, H576, H577, H578, H579, H580, H581, H582, H583, H584, H585, H586, H587, H588, H589, H590, H591, H592, H593, H594, H595, H596, H597, H598, H599, H600, H601, H602, H603, H604, H605, H606, H607, H608, H609, H610, H611, H612, H613, H614, H615, H616, H617, H618, H619, H620, H621, H622, H623, H624, H625, H626, H627, H628, H629, H630, H631, H632, H633, H634, H635, H636, H637, H638, H639, H640, H641, H642, H643, H644, H645, H646, H647, H648, H649, H650, H651, H652, H653, H654, H655, H656, H657, H658, H659, H660, H661, H662, H663, H664, H665, H666, H667, H668, H669, H670, H671, H672, H673, H674, H675, H676, H677, H678, H679, H680, H681, H682, H683, H684, H685, H686, H687, H688, H689, H690, H691, H692, H693, H694, H695, H696, H697, H698, H699, H700, H701, H702, H703, H704, H705, H706, H707, H708, H709, H710, H711, H712, H713, H714, H715, H716, H717, H718, H719, H720, H721, H722, H723, H724, H725, H726, H727, H728, H729, H730, H731, H732, H733, H734, H735, H736, H737, H738, H739, H740, H741, H742, H743, H744, H745, H746, H747, H748, H749, H750, H751, H752, H753, H754, H755, H756, H757, H758, H759, H760, H761, H762, H763, H764, H765, H766, H767, H768, H769, H770, H771, H772, H773, H774, H775, H776, H777, H778, H779, H780, H781, H782, H783, H784, H785, H786, H787, H788, H789, H790, H791, H792, H793, H794, H795, H796, H797, H798, H799, H800, H801, H802, H803, H804, H805, H806, H807, H808, H809, H810, H811, H812, H813, H814, H815, H816, H817, H818, H819, H820, H821, H822, H823, H824, H825, H826, H827, H828, H829, H830, H831, H832, H833, H834, H835, H836, H837, H838, H839, H840, H841, H842, H843, H844, H845, H846, H847, H848, H849, H850, H851, H852, H853, H854, H855, H856, H857, H858, H859, H860, H861, H862, H863, H864, H865, H866, H867, H868, H869, H870, H871, H872, H873, H874, H875, H876, H877, H878, H879, H880, H881, H882, H883, H884, H885, H886, H887, H888, H889, H890, H891, H892, H893, H894, H895, H896, H897, H898, H899, H900, H901, H902, H903, H904, H905, H906, H907, H908, H909, H910, H911, H912, H913, H914, H915, H916, H917, H918, H919, H920, H921, H922, H923, H924, H925, H926, H927, H928, H929, H930, H931, H932, H933, H934, H935, H936, H937, H938, H939, H940, H941, H942, H943, H944, H945, H946, H947, H948, H949, H950, H951, H952, H953, H954, H955, H956, H957, H958, H959, H960, H961, H962, H963, H964, H965, H966, H967, H968, H969, H970, H971, H972, H973, H974, H975, H976, H977, H978, H979, H980, H981, H982, H983, H984, H985, H986, H987, H988, H989, H990, H991, H992, H993, H994, H995, H996, H997, H998, H999, H1000.

H1B.FR.83.HXB2
 H1A1.UG.85.U455.U455A
 H1C.ET.86.ETH2220
 H1F1.BE.93.VI850
 H1O.BE.87.ANT70
 H1O.CM.91.MVP5180
 H1O.CM.98.98CMA105
 H1O.CM.98.98CMA107
 H1O.CM.98.98CMA1197
 H1O.SN.99.99SE.MP1299
 H1O.US.99.99USTWLA
 H1N.CM.02.DJ08131
 H1N.CM.04.04CM.1015_04
 H1N.CM.04.04CM.1131_03
 H1N.CM.95.YBF30
 H1P.CM.06.U14788
 CPZ.CD.06.BF167
 CPZ.CD.90.ANT
 CPZ.CM.01.SIVcpzCAM13
 CPZ.CM.05.LB715
 CPZ.CM.05.SIVcpzDP943
 CPZ.CM.05.SIVcpzEK505
 CPZ.CM.05.SIVcpzLB7
 CPZ.CM.05.SIVcpzMB66
 CPZ.CM.05.SIVcpzMB897
 CPZ.GA.88.SIVcpzGAB2
 CPZ.TZ.09.TAN1
 CPZ.TZ.06.SIVcpzTAN13
 CPZ.TZ.09.UG38
 CPZ.US.85.US.Marilyn
 GOR.CM.04.SIVgorCP684con
 GOR.CM.07.SIVgor2139.287
 GOR.CM.07.SIVgorCP2135con
 GOR.CM.12.SIVgor.BPID1
 GOR.CM.12.SIVgor.BQID2
 GOR.CM.13.SIVgor.BPID15

MAC.US.x.239
 H2A.DE.x.BEN
 H2A.PT.x.ALI
 H2A.SN.86.ST.J5P4_27
 H2B.CI.x.EH0
 H2B.GH.86.D295.ALT
 H2G.CI.92.Abt96
 H2U.CI.07.07IC.TNP3
 H2U.FR.96.12034
 ASC.UG.10.RT03
 COL.CM.x.CG01
 COL.UG.10.BWC01
 COL.UG.10.BWC07
 DEB.CM.04.SIVdeb04CMPF3061
 DEB.CM.99.CM40
 DEB.CM.99.CM5
 DRL.DE.11.D3
 DRL.DE.11.D4
 GRV.ET.x.GRI.677.gri_1
 GSN.CM.99.CN166
 GSN.CM.99.CN71
 LST.CD.88.SIVlhoest485
 LST.CD.88.SIVlhoest524
 LST.KE.x.lho7
 MAL.ZM.x.SIVagmMAL.ZMB
 MND-1.GA.x.MNDCB1
 MND-2.CM.98.CM16
 MND-2.GA.x.M14
 MND-2.x.x.5440
 MNE.US.x.MNE027
 MON.CM.99.L1.99CM1
 MUS-1.CM.01.CM1239
 MUS-2.CM.01.CM1246
 MUS-3.GA.09.09gab0I81
 OLC.CI.97.97CI12
 RCM.CM.00.SIVag1.00CM312
 RCM.CM.02.SIVrcm.02CM8081
 RCM.GA.x.SIVRCMGAB1
 SAB.SN.x.SAB1
 SMM.CI.79.SIVsmCI2
 SMM.SL.92.SIVsmSL92A
 SMM.SL.92.SL92B
 SMM.US.06.CFU212
 SMM.US.11.SIVsmE660_FL8
 SMM.US.86.CFU212
 SMM.US.x.H9
 SUN.GA.98.L14
 SVK.KE.x.KE51
 SVK.KE.x.SVK173.COMGMN
 TAL.CM.00.266
 TAL.CM.01.8023
 TAN.UG.x.SIVagmTAN1
 VER.KE.x.AGM155
 VER.ZA.10.VSAC4004
 VER.ZA.10.VSAG1003
 VER.ZA.10.VSAM0022
 WRC.CI.98.98CI04
 WRC.GM.05.Pbt.05GM.X02

AKLLWKGEAVVIQDNS.DIKVVP RRKAKIIRDY GKQMGADCCVASRODED.....
 M-G
 A-G
 E-I
 V-G
 -Q.....-KG.....E.....T-SM-G-T-SESVEOPSEIP
 -Q.....-KG.....E.....T-SM-N-T-SESMEOPSEIP
 -Q.....-KG.....E.....T-SM-G-T-SETVEOPSEIP
 -Q.....-KG.....E.....X-SM-G-T-GESMEOPSEIP
 -Q.....-KG.....H.....T-SM-G-T-SENVEOPSEIP
 -Q.....-KG.E.....T-M-G-T-SENMEOXSEIS
G.....G-A-NQDMA
G.....G-G-NQDME
G.....G-G-NQDME
G.....G-G-NOEME
 -T.....-KG.....DT-R-SESLEQSG
 -R.....-KEGE.E.....N.....V-AGG-DT-R-SESLEQSG
 -Q.....-K-QE.E.....KE-R-KIEDR-DL-G-N
-QGE.E
L-QE.E.....V-N-M-GG-SQGM
 -R.....-K-KE.EV.....GSM-G-SONLE
L-QE.E.....E-AN-L-D-N
L-QE.E.....NONME
L-KE.....VE-N-L-D-S
 -S.....-KG.....KH.....V-L-GG-NONME
 -R.....-KEGE.....KE.....N-AGGMD-N
 -R.....-KEGE.....KE.....AGS-DD-N-A
 -R.....-KEGE.....V-K.....SAGSMD-N-AQNME
-KE.E.EV.....K.....A-SM-G-S
 -T.....-KG.....DT-G-SESLEQSS
 -T.....-KG.....DT-G-SESLEQSS
 -T.....-KG.....DT-G-SESLEQSS
 -T.....-KG.....DT-G-SESMEQSG.....0
 -Q.....-KG.....X-SM-XG-T-GESMEQSG.....E
 -T.....-KG.....N.....DT-G-SESLEQSG.....0

p31 Integrase end
 Pol end
 GE.....-ILKVG T.....K.....GGKEV-SSSHMEDTGEAREVA.....
 GE.....-IVKVG T.....GROEL-SSPHLEGAREDEMAC.....
 GE.....-D-IVKVG T.....GROEL-SGPHLEGAREDEVA.....
 GE.....-D-IVKVG A-II.....K.....GROEM-SGSHLEGAREDEVA.....
 GE.....-I-KVGT E-I.....N.....GGKEL-CSADVEDTMOAREVAQSN
 GE.....-I-KVGT E.....H.....GGKEL-CSADMEDTROAREMAQSD
 GE.....-I-KVGT E.....K.....GGKEV-SSSNMEDTROTGEVA
 GE.....-I-KVGT.....I.....K.....GGKEV-SGSHLEGARDGEMAQLEIP
 -D.....L-KVGT.....I.....K.....GGKEL-SSNLEGA-KVREMALPDQTP
 GE.....-V.....KTTTGLT-K.....K.....ASQNVGKKGMDSPKREDTSLD
 GE.....-V.....RNSQGLF-K-V-T-LQYGEDV-SENL-NGQKEATYK.GMD
 GO-V.....-KNOEGSLF-K-V-TKVN YGERMVSKALLPNEQ-EAGEVK..EVD
 GO-V.....-RSTEGTLF-K-V-S-ITHGTQD-VVSEDLPLNGKEKPGTETLD
 -T.....-KLDTG-L.....V-K.....DVGSKIDPDQTHE
 -T.....-CKTEVG.....V-KE.....D-BSKVNTQNSNE
 -T.....-CKTEVG.....V-K.....GVDSKTDTMPSE
 -E-V.....-KEGT-L-I.....K.....NVDSHTNME
 -E-V.....-KEGT-L-I.....K.....N-DSHTNME
 -R-I.....-KEGE.....K.....ERKTM-SEGSMEGVREANKQMEGSDLODQE
 -V.....-V-TQAGE-IT.....KP.....EAKTE-VGG-THTN
 -V.....-V-TQAGE-IT.....KP.....EAK-F-VGS-AHTSN-REEGRMAD
 -QI.....-VKEGD-TFS.....LVK.....EGPKDSEGLDNN
 -Q.....-VKEGE-N-FSI.....LVK.....EGPKDSSMDN
 -Q.....-VKEGE-N-FS.....LVK.....EGPKDSESSLN
 -Q-I.....-KEADG-L-I.....K.....GGKEL-GEDHVESQ+TYK
 GI.....-LKYOE E-I.....C.....K.....E-SGKNSQ-NLESV
 -E-I.....-KEGT-L.....K.....TVDS-PI-EA
 -E-I.....-KEGT-L.....K.....TVDSNPMMED-ETA
 -E-I.....-KEGV-L.....K.....TVDSNPMMGR
 GE.....-ILKVG T.....K.....GGKEV-SSSHMEDTGEAREVA
 -R.....-KTEEG-LT.....KP.....TENVGGDTQYNYLRKODGLAN
 -A.....-VKTQGVIT.....K.....AKENVGSKSNTGDHRKEDLDN
 -R.....-L-KTKEG-VT.....K.....TRQNV-SEPD-VHVYKEDGLAD
 -T.....-KTDLGEV-T.....KP-NA-KNVGSKSNTGDIREEDGLVDQGN
 -F.....-LLKEGD-KYFS-L.....VKP.....VDSSANV
 -T-I.....-GO-L.....C-VK.....RKNV-SETNMEG-GOEN
 -T-I.....-GO-L.....C-K.....RKDV-SETSMEG-ROEKD
 -T.....-EET-L.....K.....RKDV-SEANLAG-ROEEN
I.....-EDG-EL-TI.....K.....-L-SOAPLEGNRTRAG...EVD
 -N.....-I-KVGT.....I.....K.....GGKELGNSPYLENP-EDGKMAQPD
 GE.....-I-KVG.....I.....K.....GGKEM-SSADMENTTQVRETAQLSEISKV
 GE.....-KVGT E.....K.....GROEMGSSASMEDO
 GE.....-I-KVGT E.....K.....AGKEL-SGSHLEDTREAEMA
 GE.....-ILKVG T E.....K.....GGKEL-SGSHLEDTGEAREVA
 GE.....-ILKVG T.....K.....GGKEL-SGSHLEDTREAEMA
 GE.....-ILKVG T E.....K.....GGKEL-SGSHLEDTGEAREVA
 -S.....-KQGE-N-L.....LVK.....GESSSVEH-G
 -R.....-TPDQVIA.....SSD.....ERVSDSHLETSSKN
 -R.....-V-TEEG-FA.....T-H-ER-DSGSH-ENDPKT
 -H-I.....-LRTDEGEVIT.....KP-QALGNKIDLE-SKEQ-AEMGRDN
 -H-V.....-LRTDEGEVIT.....KP-QA-GNKTDLEGSKEQ-AEMGRDN
 -T.....-KEGE-EL.....KE-ERKTVG-KTNMEG
 -R-I.....-LKEGE-EL.....K-EPRKTLG-ETHLEGAGSSDQHQAG.DS
 -T-I.....-K-GE-EL.....K-ERK-MG-ENHLEGGTNNL
 -T-I.....-K-QE-EL.....K-##KRMG-ENHVEGIRRPDN.....E
 -V-I.....-KEQG-EL-I.....K-GKGVNEDHMEGIRGRADN
 -G.....-L-VETPE-G-IT-K.....KVNNGEGMDRSSSNKD
 -G-V.....-L-V-TPEG-LLT.....K.....KTWDGKGM#

1003
1003
1011
1005
1010
1010
1010
1010
1011
1010
1010
1013
1014
1015
1015
1008
1007
1004
1003
1013
1015
1014
1007
1011
1012
1011
1000
999
999
1006
1003
1003
1003
1008
1008
1007
1006

1020
1033
1033
1033
1036
1035
1016
1022
1034
1018
991
993
995
1011
1008
1008
1013
1013
1041
1016
1017
1020
1021
1020
1023
1010
1011
1017
1012
1016
1018
1012
1017
1017
1012
1016
1018
1024
1019
1040
1028
1023
1006
1020
1016
1016
1016
1017
1024
1037
1020
1021
1020
1020
1032
1024
1024
1026
1022
1007

Table with columns for protein names (e.g., H1B, H1A1, H1C), accession numbers, and amino acid sequences. Includes a 'Vif start' marker and a 'Vif' label on the right side.

Table with columns for protein names (e.g., MAC, H2A, H2B), accession numbers, and amino acid sequences. Includes a 'Vif start' marker and a 'Vif' label on the right side.

	Vif end	
H1B.FR.83.HXB2	192
H1A1.UG.85.U455_U455A	193
H1C.ET.86.ETH2220	193
H1F1.BE.93.VI850	193
H10.BE.87.ANT70	193
H10.CM.91.MVP5180	193
H10.CM.98.98CMA105	193
H10.CM.98.98CMABB197	193
H10.SN.99.99SE_MP1299	193
H10.US.99.99STWLA	196
H1N.CM.02.DJ08131	193
H1N.CM.04.04CM_1015_04	193
H1N.CM.04.04CM_1131_03	193
H1N.CM.95.YBF30	193
H1P.CM.06.U14788	196
CPZ.CD.06.BF1167	197
CPZ.CD.90.ANT	197
CPZ.CM.01.SIVcpzCAM13	194
CPZ.CM.05.LB715	193
CPZ.CM.05.SIVcpzDP943	193
CPZ.CM.05.SIVcpzEK505	193
CPZ.CM.05.SIVcpzLB7	193
CPZ.CM.05.SIVcpzMB66	193
CPZ.CM.05.SIVcpzMB897	193
CPZ.GA.88.SIVcpzGAB2	193
CPZ.TZ.00.TAN1	199
CPZ.TZ.06.SIVcpzTAN13	199
CPZ.TZ.09.UG38	199
CPZ.US.85.US_Marilyn	193
GOR.CM.04.SIVgorCP684con	193
GOR.CM.07.SIVgor2139_287	193
GOR.CM.07.SIVgorCP2135con	193
GOR.CM.12.SIVgor_BPID1	193
GOR.CM.12.SIVgor_BQID2	198
GOR.CM.13.SIVgor_BPID15	193
	Vif end	
MAC.US.x.239	215
H2A.DE.x.BEN	216
H2A.PT.x.ALI	216
H2A.SN.86.ST_JSP4_27	216
H2B.CI.x.EH0	217
H2B.GH.86.0205_ALT	217
H2G.CI.92.Abt96	216
H2U.CI.07.07IC_TNP3	215
H2U.FR.96.12034	214
ASC.UG.10.RT03	242
COL.CM.x.C0U1	172
COL.UG.10.BWC01	173
COL.UG.10.BWC07	172
DEB.CM.04.SIVdeb04CMPF3061	SNRLHKEGPENDLATSGRRVLS-	251
DEB.CM.99.CM40	GH	231
DEB.CM.99.CM5	SN	231
DRL.DE.11.D3	218
DRL.DE.11.D4	218
GRV.ET.x.GRI_677_gri_1	220
GSN.CM.99.CN166	238
GSN.CM.99.CN71	238
LST.CD.88.SIVlhoest485	K	228
LST.CD.88.SIVlhoest524	227
LST.KE.x.lho7	Q	227
MAL.ZM.x.SIVagmMAL_ZMB	227
MND-1.GA.x.MNDGB1	223
MND-2.CM.98.CM16	223
MND-2.GA.x.M14	213
MND-2.x.x.5440	214
MNE.US.x.MNE027	215
MON.CM.99.L1_99CML1	256
MUS-1.CM.01.CM1239	236
MUS-2.CM.01.CM1246	239
MUS-3.GA.09.09Gab0I81	241
OLC.CI.97.97CI12	204
RCM.CM.00.SIVagi_00CM312	221
RCM.CM.02.SIVrcm_02CM8081	221
RCM.GA.x.SIVRCMGAB1	217
SAB.SN.x.SAB1	235
SMM.CI.79.SIVsmCI2	217
SMM.SL.92.SIVsmSL92A	214
SMM.SL.92.SL92B	218
SMM.US.06.Ftq	218
SMM.US.11.SIVsmE660_FL8	215
SMM.US.86.CFU212	215
SMM.US.x.H9	215
SUN.GA.98.L14	Q	242
SVK.KE.x.KE51	224
SVK.KE.x.SVK173_COMGNM	222
TAL.CM.00.266	228
TAL.CM.01.8023	228
TAN.UG.x.SIVagmTAN1	239
VER.KE.x.AGM155	223
VER.ZA.10.VSAC4004	228
VER.ZA.10.VSAG1003	224
VER.ZA.10.VSAM0022	223
WRC.CI.98.98CI04	208
WRC.GM.05.Pbt_05GM_X02	209

	Vpx start	Vpx end
MAC.US.x.239	MSD..PRERIP..PGNSGEETIGEAFEWLNRTVEEINREAVNHLPRELIFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRLGEGHGAGGW.RPG..PPPPPPGLA*	112
H2A.DE.x.BEN	-T.....V.....E-I-AL.....R.....A-T.....M---I-T-F-R-T-W-DM-RE-LEDQ.....V-	114
H2A.PT.x.ALI	-AN...TV.....D---AL.....R.....Q---T.....M---MYT-FM-T---G---P---S.....V-	113
H2A.SN.86.ST.JSP4_27	-AG...T.....D---A.....R.....I---T.....M---M-I-S-R-T---G---P---S.....V-	113
H2B.CI.x.EH0V.....D---V---A-E-IT-L-V.....K-A-RE---I-T.....M---M-I-FA-G-R---P---S.....	113
H2B.GH.86.D205.ALTV.....E-D---V---A-E-IT-L-V.....R-A-RE---I-T.....L-M-M-V-YT---Q---P---S.....	112
H2G.CI.92.Abt96V.....E-D---V---E-T---I.....R-A-RE---I-T.....L-M-M-V---T-QK---P---Q.....	112
H2U.CI.07.071C.TNP3D-VE---E-S-TA-Q.....G---T---T.....Y---R-Q-R---P.....	113
H2U.FR.96.12034	-G.....E-V---S-E-RD-A-Q.....Q-R---E-T.....F---O---P---S.....	113
DRL.DE.11.D3	-AEGQ-A-A-AE-VGA...V-LE-Q-SLLR-Q-RL-FHP-FL-RL-NTCM-HY-ALQL-FT-S---L-LH-M-T-FOO-S-VQGR-PPPLRPAGDRL	116
DRL.DE.11.D4	-AEGQSV-A-AE-VGA...V-LE-Q-SLLR-Q-QL-FHP-FL-RL-NTCM-HY-ALQL-FT-S---L-LH-M-T-FOO-S-VRGR-PPPLRPAGDRL	117
MND-2.CM.98.CM16	-AE..GAPE-E..GA...V-DLNT-E-SL-K-Q-RL-FHP-FL-RL-NACI-H-RHOR-L-A---L-MN-M-T-MQOE-P-RS---P.....MV-	100
MND-2.GA.x.M14	-AE.....A-EA-QGA...V-GLEQ-E-SL-QV---QL-FHP-FL-RL-NTCV-H-RLRRTLN-A---L-M-M-V-MQO-P-RS---S.....	100
MND-2.x.x.5440	-AE.....A-EA-EGA...V-GLEQ-ETSL-R---RL-FHP-FL-RL-NTCV-H-RHOR-LD-A---L-MH-MYT-MQO-P-RN.....RP-	100
MNE.US.x.MNE027A-EG-TGA...VEFOPWLARMLY...Q-RL-F-G-F-HL-RTCV---RTL-AG---L-M---T-MRS-KLR...DPPRO...ERV...IL-MQ-	113
RCM.CM.00.SIVagi.00CM312	-AE.....A-EV-TGA...VEFOPWLARMLY...Q-RL-FHP-F-RL-RTCV-H---L-R-LE-AG---L-M---T-MRS-LR...DPPRO...ERV...IL-MQ-	106
RCM.CM.02.SIVrcm.02CM8081	-AE.....A-EV-TGA...AEFOPWL.RDML-KV-L-RL-FHP-F-RL-RTCV-H---VHOR-LE-AA---L-M---I-OT-SQR...PNPR.AV-ERI...TIL-M-	106
RCM.GA.x.SIVRCMGAB1	-GS..A.....T.....YS---AV-G-K.....R---F---QA-TQ---L---C-V---T-R---S---A.....	105
SMM.CI.79.SIVsmCI2	-T.....Q---D---D---E---V---T.....R-C---V---Y---A---VQ-M---M-Q-FR---T-R---SQ---T.....	113
SMM.SL.92.SIVsmSL92AE.....E---D---A---Q.....R.....R-A-T.....V---R---T.....	111
SMM.SL.92.SL92BE.....G---A.....R.....M---E---T.....T.....	113
SMM.US.06.FI0A.....L---V.....R-C-F---A-T---M---M---R---N---R---E---Q.....	113
SMM.US.11.SIVsmE660_FL8	-X.....X---D---H---X---A.....R-X---M---V---T-X---G-E.....	113
SMM.US.86.CFU212		
SMM.US.x.H9		

Vpr start	oligomerization		amphipathic α -helix		H(S/N)RIG motifs		Vpr end in HXB2		Vpr end
	frameinsert in HXB2				frameinsert in HXB2				
H1B.FR.83.HXB2	MEQA	PEDQGPQREPHNEWTLLELLEELKNEAVRHFPRILWLHGLQHIYETGVD	TWAGVEAIRILQLQLFIH#FRIGCRHSRIGVTR	QRRRANGASRS	96
H1A1.UG.85.U455.U455A			96
H1C.ET.86.ETH2220			97
H1F1.BE.93.VI850			97
H10.BE.87.ANT70			98
H10.CM.91.MVP5180			101
H10.CM.98.98CMA105			101
H10.CM.98.98CMA197			101
H10.CM.98.98CMA1105			101
H10.SN.99.99SE.MP1299			101
H10.US.99.99STWLA			101
H1N.CM.02.DJ08131	V-V	S	96
H1N.CM.04.04CM.1015.04			96
H1N.CM.04.04CM.1131.03			96
H1N.CM.05.YBF30			96
H1P.CM.06.U14788			99
CPZ.CD.06.BF1167			96
CPZ.CD.90.ANT			89
CPZ.CM.01.SIVcpzCAM13	-H		97
CPZ.CM.05.LB715			96
CPZ.CM.05.SIVcpzDP943			96
CPZ.CM.05.SIVcpzEK505			96
CPZ.CM.05.SIVcpzLB7			96
CPZ.CM.05.SIVcpzMB66	-L		97
CPZ.CM.05.SIVcpzMB897			99
CPZ.GA.88.SIVcpzGAB2			98
CPZ.TZ.09.TAN1			84
CPZ.TZ.06.SIVcpzTAN13			84
CPZ.TZ.09.UG38			84
CPZ.US.85.US.MariLyn			99
GOR.CM.04.SIVgorCP684con			101
GOR.CM.07.SIVgor2139.287			99
GOR.CM.07.SIVgorCP2135con			99
GOR.CM.12.SIVgor.BPID1			99
GOR.CM.12.SIVgor.BQID2			101
GOR.CM.13.SIVgor.BPID15			99
MAC.US.x.239			102
H2A.DE.x.BEN			105
H2A.PT.x.ALI			106
H2A.SN.86.S2.JSP4.27			106
H2B.CI.x.EH0			105
H2D.GH.86.0295.167			105
H2G.CI.92.Abt96			102
H2U.CI.07.071C.TMP3			103
H2U.FR.96.12034			103
ASC.UG.10.RT03			104
COL.CM.x.C0U1			95
COL.UG.10.BWC01			92
COL.UG.10.BWC07			95
DEB.CM.04.SIVdeb04CMPF3061	-RPPPSHP		118
DEB.CM.09.CM40	-RPPPSHP		118
DEB.CM.09.CM5	-RPPPSHP		118
DRL.DE.11.D4			101
DRL.DE.11.D4			101
GRV.ET.x.GRI.677.gri_1	-ASGRDPE		119
GSN.CM.99.CN166	-RIPPSHP		134
GSN.CM.99.CN171	-RTPPSHP		136
LST.CD.88.SIVlhoest485	SR-R		115
LST.CD.88.SIVlhoest524	SR-R		115
LST.KE.x.lho7	SR-R		115
MAL.ZM.x.SIVagmMAL.ZMB	-ASGG5		122
MND.1.GA.x.MNDGB1			121
MND.2.CM.98.CM16			102
MND.2.GA.x.M14			102
MND.2.x.x.5440			102
MNE.US.x.MNE027			102
MON.CM.99.L1.99CM1			121
MUS.1.CM.01.CM1239	-RVPPSHR		136
MUS.2.CM.01.CM1246	-PPSHR		136
MUS.3.GA.09.09Ga081	-SNPRERIPPSHPRPWH		140
OLC.CI.97.97C112	-AE		97
RCM.CM.00.SIVa01.00CM312			101
RCM.CM.02.SIVrcm.02CM8081			101
RCM.GA.x.SIVrcmGAB1			101
SAB.SN.x.SAB1			141
SMM.CI.79.SIVsmC12	-AE		102
SMM.SL.92.SIVsmS192A	-AE		103
SMM.SL.92.SIVsmS192B	-H		103
SMM.SL.92.SIVsmS192C	-AE		103
SMM.US.11.SIVsmE660.FL8	-AE		102
SMM.US.86.CFU212	-AE		102
SMM.US.x.H9	-TE		102
SUN.GA.98.L14	-AS		117
SVK.KE.x.KC51	-A-F		116
SVK.KE.x.SVK173.COMGNM	-AE-F		116
TAL.CM.00.266	-RL		114
TAL.CM.01.8023	-RL		115
TAN.UG.x.SIVagmTAN1	-AEGR		120
VER.KE.x.AGM155	-ASGR		116
VER.ZA.10.VSAG094	-ASGR		119
VER.ZA.10.VSAG1093	-SSG#		120
VER.ZA.10.VSAG0022	-ASGT		119
WRC.CI.98.98CI04			117
WRC.GM.05.Pbt.05GM.X02			118

Tat start	C-rich region	nuclear localization	exon 1 end_exon 2 start	
H1B.FR.83.HXB2	MEPVPD	RLEPWKHPGSOPTKAC	TNCYCKKCFHCQVCFITKALGISYGRKRRORRAHONRGDPTGPKK+KVKVERETET.DPFD
H1A1.UG.85.U455 U455A	N	N	V	W
H1C.ET.86.ETH2220	N	N	N	N
H1F.BE.93.VI850	S	D	N	N
H1O.BE.87.ANT70	D	D	N	N
H1O.CM.91.MVP5180	D	N	N	N
H1O.CM.98.98CMA105	D	N	N	N
H1O.CM.98.98CMA197	D	N	N	N
H1O.CM.98.98CMA18105	D	N	N	N
H1O.SN.99.99S5E.MP1299	D	N	N	N
H1O.US.09.99STWLA	D	N	N	N
H1N.CM.02.DJ08131	D	N	N	N
H1N.CM.04.04CM.1015_04	D	N	N	N
H1N.CM.04.04CM.1131_03	D	N	N	N
H1N.CM.05.YBF30	D	N	N	N
H1P.CM.06.U14788	D	N	N	N
CPZ.CD.06.BP1167	D	N	N	N
CPZ.CD.90.ANT	D	N	N	N
CPZ.CM.01.SIVcpzCAM13	D	N	N	N
CPZ.CM.05.LB715	D	N	N	N
CPZ.CM.05.SIVcpzDP943	D	N	N	N
CPZ.CM.05.SIVcpzEK505	D	N	N	N
CPZ.CM.05.SIVcpzLB7	D	N	N	N
CPZ.CM.05.SIVcpzMB66	D	N	N	N
CPZ.CM.05.SIVcpzMB897	D	N	N	N
CPZ.GA.88.SIVcpzGAB2	D	N	N	N
CPZ.TZ.09.TAN1	D	N	N	N
CPZ.TZ.06.SIVcpzTAN13	D	N	N	N
CPZ.TZ.09.UG38	D	N	N	N
CPZ.US.85.US.Marilyn	D	N	N	N
GOR.CM.04.SIVgorCP684con	D	N	N	N
GOR.CM.07.SIVgor2139_287	D	N	N	N
GOR.CM.07.SIVgorCP2135con	D	N	N	N
GOR.CM.12.SIVgor.BPID1	D	N	N	N
GOR.CM.12.SIVgor.BQID2	D	N	N	N
GOR.CM.13.SIVgor.BPID15	D	N	N	N

Tat start	C-rich region	nuclear localization	exon 1 end_exon 2 start	
MAC.US.x.239ETPLRQENSLSSNERSSCISEADASTPESANLGETLSQYLRLE
H2A.DE.x.ALIETPLKAPESLKPYNPSSCTSERVDTAQLAKQGELLAQLHRLEP
H2A.PT.x.BENETPLKPGSSLMYPNPPSSCTSEDDAVAOELAKQGETLSQYLRLEP
H2A.SN.86.ST.JSP4_27LDMETPLKAPESLGSYNPSSCTSEDDAAAGLQVSPGDEILYLYLLEDNK
H2B.CI.x.EH0ETPLKEQESSLSSSSSSSSSTSEGAVNTQGLDNGETLSQYLR
H2B.GH.86.D295_17ETPLKEQESSLSSSSSSSSSTSEPVNTQGLDNGETLSQYLR
H2C.CI.92.071CIGMETPSREOENSLKSCRELSSTFEETVDAPGLGTOAELLYLYLLESNK
H2U.CI.07.07IC TMP3ETPLKEQESSLKSCKEPLSCTSEEDVRIPESSGKTEETLSQYLRLET
H2U.FR.96.12034ETPLKEQESSLKSCKEPLSSTSEGDVGTGSAAGETLSQYLRLE
ASC.UG.10.RT83SLKHRAPAL
COL.UG.x.G01+1SFONERLTKN
COL.UG.10.BWC01+1SFONERLTKN
COL.UG.10.BWC07+1SFONERLTKN
DEB.CM.04.SIVdeb04CMPF3061+1SFONERLTKN
DEB.CM.09.CM40+1SFONERLTKN
DEB.CM.09.CM16+1SFONERLTKN
DRL.DE.11.D3+1SFONERLTKN
DRL.DE.11.D4+1SFONERLTKN
GRV.ET.x.GRI.677_gri_1+1SFONERLTKN
GSN.CM.09.CN166+1SFONERLTKN
GSN.CM.09.CN171+1SFONERLTKN
LST.CD.88.SIVLhoest485+1SFONERLTKN
LST.CD.88.SIVLhoest524+1SFONERLTKN
LST.KE.x.lho7+1SFONERLTKN
MND.1.GA.x.MNDG81+1SFONERLTKN
MND.2.CM.98.CM16+1SFONERLTKN
MND.2.GA.x.M14+1SFONERLTKN
MND.2.x.x.5440+1SFONERLTKN
MNE.US.x.MNE627+1SFONERLTKN
MON.CM.99.L1.99CML1+1SFONERLTKN
MUS.1.CM.01.CM1239+1SFONERLTKN
MUS.2.CM.01.CM1246+1SFONERLTKN
MUS.3.GA.09.09GabaD181+1SFONERLTKN
OLC.CI.97.97C112+1SFONERLTKN
RCM.CM.00.SIVagi_00C312+1SFONERLTKN
RCM.CM.02.SIVrcm_02C80881+1SFONERLTKN
RCM.GA.x.SIVRCMGAB1+1SFONERLTKN
SAB.SN.x.SAB1+1SFONERLTKN
SMM.CI.79.SIVsmC12+1SFONERLTKN
SMM.SL.92.SIVsmSL92A+1SFONERLTKN
SMM.SL.92.SL92B+1SFONERLTKN
SMM.US.06.FT0+1SFONERLTKN
SMM.US.11.SIVsmE660_FLB+1SFONERLTKN
SMM.US.86.CFU212+1SFONERLTKN
SMM.US.x.H9+1SFONERLTKN
SUN.GA.98.L14+1SFONERLTKN
SVK.KE.x.KE51+1SFONERLTKN
SVK.KE.x.SYK173_COMGNM+1SFONERLTKN
TAL.CM.00.266+1SFONERLTKN
TAL.CM.01.8023+1SFONERLTKN
TAN.UG.x.SIVagmTAN1+1SFONERLTKN
VER.KE.x.AGM155+1SFONERLTKN
VER.ZA.10.VSAC4004+1SFONERLTKN
VER.ZA.10.VSAG1003+1SFONERLTKN
VER.ZA.10.VSAM0022+1SFONERLTKN
WRC.CI.98.RC104+1SFONERLTKN
WRC.CM.05.Pbt_05GM_X02+1SFONERLTKN

H1B.FR.83.HXB2
H1A1.UG.85.U455_U455A
H1C.ET.86.ETH2220
H1F1.BE.93.VI850
H10.BE.87.ANT70
H10.CM.91.MVP5180
H10.CM.98.98CMA105
H10.CM.98.98CMABB197
H10.SN.99.99SE_MP1299
H10.US.99.99USTWLA
H1N.CM.02.DJ08131
H1N.CM.04.04CM_1015_04
H1N.CM.04.04CM_1131_03
H1N.CM.95.YBF30
H1P.CM.06.U14788
CPZ.CD.06.8F1167
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715L
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.00.TAN1	SSCGVTRPKTSSGSSGSAKHK
CPZ.TZ.06.SIVcpzTAN13	SSSGSTRPVSTSSGCGWPYSKP
CPZ.TZ.09.UG38	SSCGSTRPVKTSAGSSGR.S
CPZ.US.85.US_Marilyn
GOR.CM.04.SIVgorCP684con	WLCGSTGH
GOR.CM.07.SIVgor2139_287
GOR.CM.07.SIVgorCP2135con
GOR.CM.12.SIVgor_BP1D1
GOR.CM.12.SIVgor_BQ1D2
GOR.CM.13.SIVgor_BP1D15
Tat end	
MAC.US.x.239
H2A.DE.x.BEN
H2A.PT.x.ALI
H2A.SN.86.ST_JSP4_27
H2B.CI.x.EH0
H2B.GH.86.0205_ALT
H2G.CI.92.Abt96
H2U.CI.07.07IC_TNP3
H2U.FR.96.12034
ASC.UG.10.RT03
COL.CM.x.C01
COL.UG.10.BWC01
COL.UG.10.BWC07
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40	C.....
DEB.CM.99.CM5
DRL.DE.11.D3
DRL.DE.11.D4
GRV.ET.x.GRI_677_gri_1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest485
LST.CD.88.SIVlhoest524
LST.KE.x.lho7
MND-1.GA.x.MNDGB1
MND-2.CM.98.CM16
MND-2.GA.x.M14
MND-2.x.x.5440	SGAG
MNE.US.x.MNE027	SGAG
MON.CM.99.L1_99CML1
MUS-1.CM.01.CM1239
MUS-2.CM.01.CM1246
MUS-3.GA.09.09Gab0I81
OLC.CI.97.97CI12
RCM.CM.00.SIVagi_00CM312
RCM.CM.02.SIVrcm_02CM8081
RCM.GA.x.SIVRCMGAB1
SAB.SN.x.SAB1
SMM.CI.79.SIVsmCI2
SMM.SL.92.SIVsmSL92A
SMM.SL.92.SL92B
SMM.US.06.Ftq
SMM.US.11.SIVsmE660_FL8
SMM.US.86.CFU212
SMM.US.x.H9
SUN.GA.98.L14
SYK.KE.x.KE51
SYK.KE.x.SYK173_COMGNM
TAL.CM.00.266
TAL.CM.01.8023	SSATVTRP
TAN.UG.x.SIVagmTAN1
VER.KE.x.AGM155	AGN
VER.ZA.10.VSAC4004	SAD
VER.ZA.10.VSAG1903	TSD
VER.ZA.10.VSAM0022	PSD
WRC.CI.98.98CI04
WRC.GM.05.Pbt_05GM_X02

100
102
101
103
115
116
114
97
98
98
102
102
102
102
112
101
115
100
108
103
101
100
102
101
100
137
137
137
134
103
124
83
83
83
83
116
72

131
131
137
134
139
129
134
129
129
102
86
87
90
99
111
97
125
125
95
105
135
108
109
109
108
108
113
127
135
105
102
109
117
108
114
114
107
107
134
133
106
131
133
130
133
103
116
111
111
128
102
120
113
113
113
133
133

H1B.FR.83.HXB2
H1A1.UG.85.U455 U455A
H1C.ET.86.EHT2220
H1F.BE.93.VI850
H1D.BE.87.AMT70
H10.CM.01.MVP5180
H10.CM.98.98CMA105
H10.CM.98.CMAB197
H10.CM.98.98CMA8185
H10.SN.99.99SPE MP1299
H11.US.99.99SHTWLA
H1N.CM.02.DJ08131
H1N.CM.04.04CM 1015 04
H1N.CM.04.04CM 1131 03
H1N.CM.95.YBF30
H1P.CM.06.U14788
CPZ.CM.07.R1167
CPZ.CM.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.09.U14788
CPZ.TZ.09.UG38
CPZ.US.85.US Marilyn
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139 287
GOR.CM.07.SIVgorCP2135con
GOR.CM.12.SIVgor BP1D1
GOR.CM.12.SIVgor B0ID2
GOR.CM.13.SIVgor BPID15

MAC.US.x.239
H2A.DE.x.BEN
H2A.PT.x.ALI
H2A.SN.86.ST JSP4 27
H2B.CI.x.EH0
H2D.GH.86.0295 ALT
H2G.CI.92.Abt196
H2U.CI.07.07IC TMP3
H2U.FR.96.12034
ASC.UG.10.RT03
COL.CM.x.C0U1
COL.UG.10.BWC01
COL.UG.10.BWC07
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM40
DRL.DE.11.D3
DRL.DE.11.D4
GRV.ET.x.GRI 677 gri_1
GSN.CM.99.CN166
GSN.CM.99.CN171
LST.CD.88.SIVlhoest485
LST.CD.88.SIVlhoest524
LST.KE.x.lho7
MMD-1.GA.x.MNDG81
MMD-2.CM.98.CM16
MMD-2.GA.x.M14
MMD-2.x.x.5440
MNE.US.x.MNE627
MON.CM.99.L1 09CML1
MUS-1.CM.01.CM1239
MUS-2.CM.01.CM1246
MUS-3.GA.09.09gag0181
OLC.CI.97.97CI12
RCM.CM.00.SIVagi 00C312
RCM.CM.02.SIVrcm 02CM8081
RCM.GA.x.SIVrcmGAB1
SAB.SN.x.SAB1
SHM.CI.79.SIVsmC12
SHM.SL.92.SIVsmSL92A
SHM.SL.92.SL92B
SHM.US.06.FT0
SHM.US.11.SIVsmE660_FL8
SHM.US.86.CFU212
SHM.US.x.H9
SUN.GA.98.L14
SVK.KE.x.KE51
SVK.KE.x.SYK173_COMGNM
VER.ZA.10.VSAC1400A
VER.ZA.10.VSAG1903
VER.ZA.10.VSAM0022
WRC.CI.98.98CI04
WRC.GM.05.Pbt 05GM X02

Rev start	exon 1 end	exon 2 start	NLS	Leu-rich effector domain		
MAGRSQDSE	. . . ELIRTVRL	KLKLYQ	. . . SNPNNP	. . . EGTROARRNR	RRWREROR	QIHSIRERLGTYL . . .
. . . -R--NP	. . . D-LKA	-I	. . . C-R	. . . S-K	. . . A-D	. . . L-SDC . . .
. . . LKA	-I	-I	. . . Y-T	. . . A-TL	. . . D-SNF	. . . P-N
. . . LKA	-I	-I	. . . Y-K	. . . A-R	. . . L-D	. . . V-ASC . . .
. . . F-D	. . . Q-LOAIQI	-I	. . . C-A	. . . S-N-K	. . . R-A	. . . V-D-LAT
. . . EED	. . . QO-LOAIQI	-I	. . . T-A	. . . S-N-K	. . . R-A	. . . V-D-LAT
. . . D-D	. . . L-AI-I	-RI	. . . Q-P	. . . R-S-N-K	. . . R-A	. . . V-DNLA
. . . D-T	. . . O-LOAIQI	-I	. . . H-A	. . . G-S-S	. . . R-A	. . . VETLAA
. . . DGD	. . . QP-LOAIQI	-I	. . . SHTS	. . . T-S-S	. . . T-A	. . . HAVDTLAA
. . . ED	. . . Q-LOAIQI	-I	. . . Y-S	. . . A-S-N-K	. . . R-A	. . . V-D-LAT
. . . VN	. . . Q-LOAIQI	-I	. . . Y-S	. . . NNNO	. . . T	. . . RE-Q
. . . VN	. . . K-LTA	-I	. . . Y-N	. . . Q	. . . A	. . . CA-S
. . . VN	. . . LTA	-I	. . . Y-N	. . . Q	. . . A	. . . RA-S
. . . VN	. . . L-A	-V	. . . Y-S	. . . G	. . . K	. . . A-SSC . . .
. . . EED	. . . LR-T	-I	. . . Y-S	. . . AR	. . . GPS-NST	. . . S-DQ
. . . EEE	. . . T-LT	-I	. . . Y-S	. . . S	. . . I	. . . VEG-S
. . . EEELEGT	. . . DEQ-LKA	-KI	. . . Y-K	. . . A-S	. . . A-K	. . . K-K-D
. . . EHD-A	. . . R-LOA	-KI	. . . Y-S	. . . N	. . . A	. . . RE-N
. . . EADAT	. . . R-LOW	-I	. . . Y-D	. . . K	. . . A	. . . H-AL
. . . R	. . . L-AI	-V	. . . Y-S	. . . S	. . . A	. . . RAL
. . . G-N	. . . L-A	-I	. . . Y-S	. . . S	. . . A	. . . H-SE-AG
. . . SDE	. . . L-A	-I	. . . F	. . . S	. . . P	. . . K-SAL
. . . EE	. . . R-L-A	-KA	. . . Y-S	. . . N	. . . R	. . . A-CE
. . . EE	. . . R-L-A	-Y	. . . Y-S	. . . S	. . . A	. . . H-V
. . . EE	. . . R-L-A	-Y	. . . Y-S	. . . S	. . . A	. . . H-V
. . . EE	. . . R-L-A	-Y	. . . Y-S	. . . S	. . . A	. . . H-V
. . . EE	. . . R-L-A	-Y	. . . Y-S	. . . S	. . . A	. . . H-V
. . . EE	. . . R-L-A	-Y	. . . Y-S	. . . S	. . . A	. . . H-V
. . . EE	. . . R-L-A	-Y	. . . Y-S	. . . S	. . . A	. . . H-V
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST
. . . DE	. . . LQ-L	-I	. . . W-E	. . . G	. . . P	. . . S-NST

Rev start	exon 1 end	exon 2 start				
. . . SINHEREEE	. . . RKRL	-H	. . . T	. . . Y	. . . TG	. . . P-ANO
. . . SE-ADEEG	. . . QGKL	-LR	. . . H	. . . T	. . . Y	. . . QG-P-ASO
. . . TE-A-EE	. . . Q-KL	-R	. . . H	. . . T	. . . Y	. . . QG-P-ANO
. . . NE-AELEE	. . . R-KL	-R	. . . H	. . . T	. . . Y	. . . QG-P-ASO
. . . NARER	. . . QKGL	-LH	. . . H	. . . T	. . . Y	. . . QG-P-ASO
. . . H2G	. . . GH	-86	. . . 0295	. . . ALT		
. . . SNLEEEE	. . . RKRL	-HF	. . . H	. . . T	. . . Y	. . . QG-P-ANO
. . . NNPEEEE	. . . VR	-RL	. . . HF	. . . H	. . . T	. . . Y
. . . SNPEEG	. . . VO	-RL	. . . HF	. . . H	. . . T	. . . Y
. . . SGREE-P	. . . ELRO	-L	. . . AC	. . . I	. . . RT	. . . D-F
. . . IR	-IR	-TAATV	. . . F	. . . TD	. . . W-AR	. . . G-ARO
. . . TK	-IRS	-H	. . . ATNH	. . . YFAD	. . . Y	. . . DT-V-ARO
. . . HAG	-RGE	-L	. . . AIAA	. . . I	. . . F	. . . EG-S
. . . HAG	-RGSAAENR	. . . LKVIS	. . . I	. . . Y	. . . KG	. . . S-AST
. . . TAGRE	. . . R	. . . AD	. . . NI	. . . I	. . . S	. . . Y-KG
. . . SAGPREP	. . . PAWFOEYL	. . . VTR	. . . W	. . . DDLL	. . . PO	. . . AR-R
. . . SAGPEREQ	. . . PWFQOYL	. . . ATR	. . . W	. . . DDLL	. . . PO	. . . AR-R
. . . SGLKGE	. . . KOALKI	. . . T	. . . GR	. . . Y	. . . QF	. . . S-ARO
. . . DHARGN	. . . Q	. . . KLON	. . . LAC	. . . T	. . . HR	. . . Y-TS
. . . DPANGR	. . . Q	. . . KLON	. . . LAC	. . . T	. . . K	. . . Y-TA
. . . STGEE	. . . P	. . . YLKIS	. . . I	. . . WT	. . . G	. . . LOVGERRR
. . . STGPEG	. . . P	. . . TYLK	. . . SRI	. . . WT	. . . G	. . . LOVGERRR
. . . STGN	. . . P	. . . Y	. . . YL	. . . SRI	. . . WT	. . . G
. . . STGNVY	. . . QEL	. . . IR	. . . YLVVV	. . . K	. . . E	. . . G
. . . PTET	. . . L	. . . R	. . . DF	. . . PW	. . . K	. . . Y
. . . STPE	. . . ELG	. . . R	. . . DF	. . . W	. . . K	. . . Y
. . . STDR	. . . L	. . . R	. . . GF	. . . W	. . . K	. . . Y
. . . SSHA	. . . E	. . . R	. . . RL	. . . H	. . . H	. . . T
. . . NGR	. . . EE	. . . RL	. . . SL	. . . AAVR	. . . I	. . . Y
. . . A-ENV	. . . Q	. . . F	. . . C	. . . I	. . . RI	. . . Y
. . . A-ND	. . . Q	. . . F	. . . C	. . . I	. . . RI	. . . Y
. . . KEE	. . . L	. . . AA	. . . KACKV	. . . I	. . . H	. . . K
. . . ST	. . . SP	. . . KR	. . . KKVVSH	. . . QDSR	. . . DT	. . . Y
. . . LLGEE	. . . DR	. . . IR	. . . LR	. . . H	. . . L	. . . N
. . . LLGEE	. . . DO	. . . IR	. . . LR	. . . NF	. . . L	. . . N
. . . LPDQ	. . . ED	. . . R	. . . PK	. . . I	. . . NF	. . . I
. . . SLDGEE	. . . L	. . . RF	. . . I	. . . F	. . . T	. . . Y
. . . SDOGEE	. . . RKRL	. . . HF	. . . H	. . . T	. . . Y	. . . Q
. . . SO-GOEE	. . . RKRL	. . . HF	. . . H	. . . T	. . . Y	. . . Q
. . . QNPPEE	. . . R	. . . RL	. . . HF	. . . H	. . . T	. . . Y
. . . SNP	. . . EEE	. . . R	. . . RL	. . . HF	. . . H	. . . T
. . . S	. . . T	. . . R	. . . RL	. . . HF	. . . H	. . . T
. . . TDDEEE	. . . RKRL	. . . HF	. . . H	. . . T	. . . Y	. . . Q
. . . SSN	. . . EEE	. . . R	. . . RL	. . . HF	. . . H	. . . T
. . . STGDOS	. . . INQV	. . . IS	. . . R	. . . E	. . . GL	. . . A
. . . S-EREDT	. . . QEO	. . . L	. . . LL	. . . R	. . . A	. . . Q
. . . PDGGS	. . . ED	. . . V	. . . F	. . . W	. . . AH	. . . Q
. . . SLTG	. . . SGRG	. . . ED	. . . LA	. . . A	. . . R	. . . I
. . . SLTGSGKG	. . . ED	. . . LA	. . . A	. . . R	. . . I	. . . Y
. . . SLG	. . . E	. . . R	. . . I	. . . I	. . . R	. . . I
. . . PLGPEE	. . . R	. . . LL	. . . AF	. . . R	. . . Y	. . . S
. . . PLGPDE	. . . R	. . . LL	. . . L	. . . W	. . . N	. . . T
. . . LLGPEE	. . . R	. . . LL	. . . L	. . . R	. . . F	. . . US
. . . PLGPDE	. . . R	. . . LL	. . . L	. . . R	. . . F	. . . US
. . . WRC	. . . CI	. . . 98	. . . R	. . . I	. . . C	. . . I
. . . WRC	. . . GM	. . . 05	. . . P	. . . B	. . . T	. . . 0

	Rev end	
H1B.FR.83.HXB2*	116
H1A1.UG.85.U455_U455A*	124
H1C.ET.86.ETH2220*	108
H1F1.BE.93.VI850*	117
H10.BE.87.ANT70*	103
H10.CM.91.MVP5180*	104
H10.CM.98.98CMA105*	102
H10.CM.98.98CMA8B197*	103
H10.SN.99.99SE_MP1299*	104
H10.US.99.99USTWLA*	104
H1N.CM.02.DJ08131*	107
H1N.CM.04.04CM_1015_04*	105
H1N.CM.04.04CM_1131_03*	105
H1N.CM.95.YBF30*	105
H1P.CM.06.U14788*	103
CPZ.CD.06.8F1167	NSN.....*	126
CPZ.CD.90.ANT*	119
CPZ.CM.01.SIVcpzCAM13*	124
CPZ.CM.05.LB715*	111
CPZ.CM.05.SIVcpzDP943*	125
CPZ.CM.05.SIVcpzEK505*	104
CPZ.CM.05.SIVcpzLB7*	124
CPZ.CM.05.SIVcpzMB66*	124
CPZ.CM.05.SIVcpzMB897*	123
CPZ.GA.88.SIVcpzGAB2*	124
CPZ.TZ.00.TAN1*	106
CPZ.TZ.06.SIVcpzTAN13*	96
CPZ.TZ.09.UG38*	111
CPZ.US.85.US_Marilyn*	106
GOR.CM.04.SIVgorCP684con*	105
GOR.CM.07.SIVgor2139_287*	105
GOR.CM.07.SIVgorCP2135con*	105
GOR.CM.12.SIVgor_BPID1*	105
GOR.CM.12.SIVgor_BQID2*	109
GOR.CM.13.SIVgor_BPID15*	105
	Rev end	
MAC.US.x.239*	108
H2A.DE.x.BEN*	104
H2A.PT.x.ALI*	108
H2A.SN.86.ST_JSP4_27*	108
H2B.CI.x.EH0	QTRG.....*	151
H2B.GH.86.0205_ALT	QTRT.....*	154
H2G.CI.92.Abt96*	101
H2U.CI.07.07IC_TNP3*	103
H2U.FR.96.12034*	99
ASC.UG.10.RT03*	97
COL.CM.x.CG11*	94
COL.UG.10.BWC01*	98
COL.UG.10.BWC07	EVATICDECRA.....*	117
DEB.CM.04.SIVdeb04CMPF3061*	98
DEB.CM.99.CM40*	91
DEB.CM.99.CM5*	88
DRL.DE.11.D3*	87
DRL.DE.11.D4*	87
GRV.ET.x.GRI_677_gri_1	SKKSAARP.....*	148
GSN.CM.99.CN166*	125
GSN.CM.99.CN71*	121
LST.CD.88.SIVlhoest485	TKEKMEIQI0REKWSQ.....*	174
LST.CD.88.SIVlhoest524*	149
LST.KE.x.lho7*	149
MND-1.GA.x.MNDGB1*	88
MND-2.CM.98.CM16*	91
MND-2.GA.x.M14*	89
MND-2.x.x.5440*	88
MNE.US.x.MNE027*	108
MON.CM.99.L1_99CML1*	113
MUS-1.CM.01.CM1239*	107
MUS-2.CM.01.CM1246*	100
MUS-3.GA.09.09Gab0I81*	114
OLC.CI.97.97CI12*	124
RCM.CM.00.SIVagi_00CM312*	108
RCM.CM.02.SIVrcm_02CM8081*	105
RCM.GA.x.SIVRCMGAB1*	94
SAB.SN.x.SAB1*	100
SMM.CI.79.SIVsmCI2*	104
SMM.SL.92.SIVsmSL92A*	104
SMM.SL.92.SL92B*	100
SMM.US.96.FTq*	107
SMM.US.11.SIVsmE660_FL8*	101
SMM.US.86.CFU212*	106
SMM.US.x.H9*	101
SUN.GA.98.L14*	151
SYK.KE.x.KE51*	85
SYK.KE.x.SYK173_COMGNM*	76
TAL.CM.00.266*	126
TAL.CM.01.8023*	112
TAN.UG.x.SIVagmTAN1*	104
VER.KE.x.AGM155*	90
VER.ZA.10.VSAC4004*	90
VER.ZA.10.VSAG1903*	90
VER.ZA.10.VSAM0022*	90
WRC.CI.98.98CI04*	124
WRC.GM.05.Pbt_05GM_X02	KQQHQGNYS..QTLHKP.....*	146

	transmembrane domain		phos cytoplasmic domain		Vpu end														
	Vpu start		α-helix	α-helix															
H1B.FR.83.HXB2	TOPIP	IVAIVALVVAIIA	IVVMSVIVIEY	...RKILRORKIDRLIDRLIERAEDSGNESEG	EISALVEMG	...VEMGHAPWDV	DDL	*											
H1A1.UG.85.U455.U455A	MT.LE	W-TG-I-L-L-L	...T-G-#	...K-L-K-K	...LN-IR	...D-DTEE-SLL	...NYDLGVD	NN											
H1C.ET.86.ETH2220	MVOLLAKVDYRIV	FT-L	...T-AY	...L-R	...K-TR	...D-DTEE-ST	...D-NLRL	N											
H1F1.BE.93.VI850	MSYLLA	IG-A-I-L	...T-Y	...K-LV	...NK-YK	...D-DAEE-AAL	...G-PFT	G-T-NW											
H10.BE.87.ANT70	MHHRD	LLAIIIS	LLFINVIL	GFILRK	LEQEKDORKE	E-LERLR	IR-IRD	DY-N-E-EQEVN	...DLVLSHGF	NPMFEP									
H10.CM.91.MVP5180	MHQEN	LLALIALS	LCLINVL	LFNLR	LVQRKODR	E-E-LERLR	IK-IRD	DY-N-E-EQEVN	...LI-SHG	FANPMFEL									
H10.CM.98.98CMA105	MHHRD	LLTLIIGS	LLLINVL	MFMLRK	LEQEKDOR	E-E-LERLR	IK-IRD	DY-N-E-EQEVN	...DLVLSHGF	NPMFEP									
H10.CM.98.98CMA105	MHHRD	LLTLIA	S-LLFIN	IL-TY	LRK-LEQRKODR	E-E-LERLR	IX-IGD	DY-N-E-EQEVN	...DLV-NHGF	NPMFEP									
H10.SN.99.99SE.MP1299	MHHRD	LLALITTS	LLLTNV	L-TFILRO	LKQKQDK	E-E-LERLR	IK-IRD	DY-N-E-EQEVN	...DLV-SYGF	NPMFEP									
H10.US.99.99USTWLA	M-YKD	LLFLAYS	LFLTALI	MF-LKLCLEQK	QSK	E-E-LRRLR	IK-IRD	DY-N-X-EQEVN	...DLV-XHGF	NPMFEL									
H1N.CM.02.DJ08131	ML.E	WGFA	-G	...VII-VLLYK	...K-LKL	E-E-ETIRO	IK-IRD	T-S-ND-DADW	...AILLS	DKF-NWV									
H1N.CM.04.04CM.1015_04	ML.S	LGFS	-G	...VII-VLLYK	...KE-KL	E-E-EQIKQ	VRD	...D-DAEW	...AILLS	SPDKL-NWV									
H1N.CM.04.04CM.1131_03	ML.E	LGFI	-FG	...VII-VLLYK	...KE-KL	E-E-EKIRO	IK-IRD	KE-K-D-DAEW	...GDKESVGD	EWLAILLS	SPDKL-NWV								
H1N.CM.95.YBF30	ML.S	LGFI	-GA-VS	...VI-ALLYR	...K-KL	E-E-KHIRO	IK-IRD	E-D-DAEW	...DGDEWL	TLLSSSKL	OGNWV								
H1P.CM.06.U14788	MH.RD	E-VLITAGVLLCC	X-GK	LLLVLK	...ERERD	FVORLA	WR-GQ	E-Y-NEE	EEQ	R	...L-NLLGF	FVL							
CPZ.CD.06.BF1167	MLWQ	FLWQ	OYLGWGA	-I-I	ALLVIRKAYLY	KVQEE	...TEQAT	D-YORR	STDS	I-EDENG	...LINMMP	NYFL							
CPZ.CD.90.ANT	MTN.F	EY-FL	FSIVLW	-IC	PILYKYKI	KQOQD	DNK-NOR	IEVLS	RLSIDS	AIEEDE	A-DTUY	G	...SGFAN	PVYREGDE					
CPZ.CM.01.SIVcpzCAM13	MI.LI	...LGC	AI-L-LLN	FI-RNL	...WRLCK	QNKLEVE	EN-AL	-T	...D-G	E-EK	QRL	...HDHDMHG	FANPLFDI						
CPZ.CM.05.LB715	MTGLEI	...IGLGI	IELS	...GA-IVAYN	...QEQY	GER	-C-K	IR	...D-DKEE	CTLL	...LPSD	PEFVLIN							
CPZ.CM.05.SIVcpzDP943	MLT.WE	QIGLI	-GIEG	...T-G	AF-TWRRK	TEED	AQ-ID	LE-IRA	OD	...D-EDQRQ	...MAYG	FDNPMF	W						
CPZ.CM.05.SIVcpzEK505	MLLLIK	...LGFIG	AIETL	W-A	YRI	...EVKVEE	SQ-RQ	IRD	...D-DAEE	...ANLL	PPDR	ODNWV							
CPZ.CM.05.SIVcpzLB7	MD.LIEL	...GLIG	-IEL	-V	...LKAYOL	...KENI	KA-KN	-E-IR	-G	...D-DMDE	HAIL	...RSGD	PELVLDIN						
CPZ.CM.05.SIVcpzMB66	MDIVQOV	...GLL	-V-IIELV	-V	...I-VK	YKLC	...KEDR	-K	...A-IR	...D-DTEE	QDLI	...T-GDN	MHIGIR	NRNN					
CPZ.CM.05.SIVcpzMB897	MEIFIL	...GLIGI	IELV	...LKAYE	...KALK	ERR	Q	...IR	...D-DTEE	EAIL	...IPEDR	VLVAIRGY							
CPZ.GA.88.SIVcpzGAB2	MLSMWV	AIGLIGITLLV	N	...G-G	SVYKRWKR	HKEEQR	ID	IKIT	...D-EDKET	ATLL	...HNNG	FDNPMFEDRI							
CPZ.TZ.00.TAN1	MIRK.V	VGSYSTN	IG-LCTLL	LIIIGGGLI	GIGIR	ELE	E-OHQ	VLE-AR	LSID	SGV	ED-EFNV	...NNFDP	PHYNPR	WI					
CPZ.TZ.06.SIVcpzTAN13	MIRK.V	VGNLEON	GVLVII	LVGGG	ALTAII	IKERN	E-OHQ	E-VR	LSID	SGI	ED-EFQW	...LNFD	PDHNPR	WI					
CPZ.TZ.09.UG38	M.RLV	...GG	ION-G	LFIIV	IVGGG	ALIGWIR	E-H	E-LHO	VLE-TR	LSID	SGI	ED-EINW	...NNFDP	PHYNPR	WI				
CPZ.US.85.US.Marilyn	ML.NWF	...ETGLI	-GIEG	LV-VII	GL-ARLW	RQIK	ENT	QE-ON	VLE-IRI	E	...D-E	EET	AKLL	...SSLED	PNPRIV				
GOR.CM.04.SIVgorCP684con	MH.RD	...ILVII	IGIILL	LAVTVIS	LKALAL	LR	...DR	E-RFFDRLE	...SNK	-E-Y-NEE	AAE	M	...R-NELG	FAPNLH					
GOR.CM.07.SIVgor2139.287	MH.RD	...IIVII	IGITL	LAVTVII	LKIFAL	LR	...DR	E-RFFDRLE	...LSNK	-E-Y-NEE	AAE	M	...R-NELG	FAPNLH					
GOR.CM.07.SIVgorCP2135con	MH.RD	...IIVII	IGITL	LAVTVII	LKALAL	LR	...DR	E-RFFDRLE	...LSTK	-E-Y-NEE	AAE	M	...R-NELG	FAPNLH					
GOR.CM.12.SIVgor.BPID1	MH.RD	...OGAL	IAGVL	VCC	...GRL	LYLLE	...ER	RORFVORLE	WRGKO	-E-YD	NEE	EEQ	R	...KL-NLLG	F-NIL				
GOR.CM.12.SIVgor.BPID2	MHYRG	...QIVII	IILC	LLLIX	VLI	MLILK	LDMK	QDR	E-E	ISRLR	IRXIRD	DY-NEE	EREXM	...DLX	SYGF	PNMFEL			
GOR.CM.13.SIVgor.BPID15	MHYRD	...EIVLI	IAGVL	GGCLTI	GYL	LWYK	...ERKRD	IFVORVA	RK-Q	-E-Y-NEE	EEK	R	...L-AOLG	FAYYML					
GSN.CM.09.CN166	MH-AA	...VWVG	-AIIIF	YLCVALLA	LYLAW	KWV	GKPK	TPQAV	R	...D-ED	SGIYD	DAS	E	...TGF	NGFANP	GFEV			
GSN.CM.09.CN71	MSAAA	...LWVG	-AIIIF	YFCLAIFA	LYLAW	KWV	GKPK	IPQAV	R	...V-D	DEE	SGIFED	AS	...MAY	G	FANP	GFEV		
MON.CM.09.L1.99CML1	MNYWNS	...LVAIT	YSLIL	ALPVAA	AW.WRY	KITK	FKRID	Q	...QIHER	RH	...SGVDT	SESE	HEETHG	...FVNP	VFD	FGEW			
MUS-1.CM.01.CM1239	MN	...YWYLA	-AL	YG	-ALFA	...FVLA	YQRWCP	...KVEVSV	R	...L-EGDSD	SGIFED	AEDEPN	...GDG	HAFAN	PAFEQ				
MUS-2.CM.01.CM1246	MH	...YWYLG	-AI	TA	-YLV	ALVA	...F	LAYQRWCP	...KPIEVN	VL	...L-EGDSD	SGIFED	ACDGED	...EDS	RAFAN	PSFEP			
MUS-3.GA.09.09Gabo181	MN	...VWYLA	-AI	TA	-YXA	...IVA	...LWLW	DKRYR	...EPTV	VS	...R-L-EGDS	GYE	VFEDAP	...QEP	...G	...PLIN	...AF	NP	AFEQ

Table with columns for Env start, signal peptide end_gp120 start, glycosylation NVT, glycosylation NVY, CD4 binding, glycosylation NNTS, and glycosylation NDT. Rows list various HIV-1 sequences and their corresponding amino acid annotations.

Table with columns for Env start, signal peptide end_gp120 start, and V1. Rows list various HIV-1 sequences and their corresponding amino acid annotations.

glycosylation NIS
glycosylation NCS

H1B.FR.83.HXB2	..M IMEKGKIKNSFNI	TSIRGKVOKEYAFYKLDIIPINDT	..TSYKLTSCNTSVIQAQCPKVSFEPIPIHYCAPAGAILKCNKTKFNGTGP	..CTN	241						
H1A1.UG.85.U455.U455A	..IT DGVRE M	..EL-D-K-V-SL-R..-VQ-NKTDN	..N-R-IN-T	..KDPK-E-K	239						
H1C.ET.86.ETH2220	..I NSAND M	..V-L-N-S	..D-R-IN-T	..RD-K	238						
H1F1.BE.93.V1850	..T NEIN MR	..EL-D-KL-L-S-L-R..-V-G-NMS	..GE-R-IN-T	..LD	237						
H10.HO.87.ANT70	..T NEIN MR	..V-KD-L-S-L-R..-V-LMELNETSS	..TNKNSKM-T-N-SIT	..Y-F-D	232						
H10.CM.91.MVP5180	..LLNETIN MR	..V-LTD-KEQKQ-L-VS-LSKWNDS	..NAVNG-T-M-N-SII-K	..T-Y-F-DTD	L-H 232						
H10.CM.98.98CMA105	..GTSSESSEDMOR	..E-VT-VVKD-KE-KO-L-RS-LMELKDKD	..NSTMT-M-T-IN-SITT	..Q	F-F-STE-T-E 236						
H10.CM.98.98CMA8197	..AGTTTTSPDL QDL-Q	..E-MT-VVKD-KE-KO-L-A-LMELKDN	..TTMA-I-T-IN-SIT-K	..Y-F-STE-T-K	244						
H10.SN.99.99SSE MP1299	..SINNDTS SPENLV	..E-VT-VVKD-KE-KO-L-VS-LMK-NEA	..NDTKDM-T-IN-SIT-K	..Y-F-STE-T-N	242						
H10.US.99.99STWLA	..KPGMGTDL-TMYP-K-D	..VT-VVKD-KE-KO-L-VS-LMK-NEA	..GKTN-T-IN-S-K	..Y-F-STE-T-N	242						
H1N.CM.02.DJ08131	..KPGMGTDL-ARHM	..T-E-HD-KKQA-SL-VE-VV-LND	..GNMST-R-IN-TAV	..T	P-M-EAN-E-K 234						
H1N.CM.04.04CM 1015 04	..NGELDIGYKQM	..VT-ERKO-KKLA-SL-AE-VVLNSES	..DSTNQT-R-I-K-TSV	..TT	P-MX-EGN-S-K-E-K 235						
H1N.CM.04.04CM 1131 03	..NGTLDIGYKQM	..VT-ERDRKK-A-SL-TE-VVLDNS	..DSDNST-T-IN-K-T-VR	..TT	P-MR-EGN-S-K-E-K 244						
H1N.CM.95.YBF30	..MTTREPDIYGKOM	..AT-ELTD-KKQV-SL-VE-VV-NA	..YNKT-R-IN-TAV	..T	P-M-EGN-S-N-S 232						
H1P.CM.06.U14788	..RVNNDTSTMSKTNL-LFO	..VT-EFKD-Q-T00-L-RA-LD-L-NS	..TSTN-T-R-IN-TT	..S-N-L-Q	P-Y-LM-KEN-R-K-N 247						
CPZ.CO.06.ANT70	..LED	..VT-EFKD-Q-T00-L-E-MKDK	..DSTN-T-Y-IN-TA	..E-S-N-V-Q	P-Y-LM-KEN-R-K-N 244						
CPZ.CO.90.ANT	..VTFKTTTP-VDMKLOE-N-Q	..GFKD-K-MK-I-G-LMKCDNNETN	..C-Y-WH-TT-S-E-ST	..V-Q	P-Y-EDED-T-V-M-K 241						
CPZ.CO.01.SIVcpzCAM13	..NNPLTKNMSGKNETGVLOM	..T-TT-EL-D-KKQV-SL-VD-LQSLGS	..GTGDT-TMIN-TA	..T	P-DVN-S-K-K-R 251						
CPZ.CO.05.LB715	..ENW-DTREDLR	..AG-D-D-KKRHSL-AE-L-RTRTTRKOGNSTQAPATTPPIASTKSPMSNDTCSNNITL	..R-I-T	..T	P-L-DTN-S-Q-I-R 268						
CPZ.CO.05.SIVcpzD943	..E-TREDFM-K-F	..EL-D-KQKV-SL-VD-VX	..NGX-R-IN-TA	..T	P-M-EPD-K-R 238						
CPZ.CO.05.SIVcpzEK505	..VNOTM-QM-Q	..EL-D-KQKV-SL-VD-L-TNNSG	..TAV	..T	P-I-DOQ-E-R 234						
CPZ.CO.05.SIVcpzL87	..I NSSNETM	..TT-EL-D-KKQV-SL-VD-LTD-NK	..TTE--IN-TA	..T	P-L-R--Y-S-I-N 235						
CPZ.CO.05.SIVcpzMB66	..TVQISTGNDSTANNITVGTIDM	..AT-EL-DRKKQV-SL-RQ-LE-LEGNKPPE	..GKNAL-R-YN-TAM-S-L	..L	P-L-N-D-N-T-I-Q-K 262						
CPZ.CO.05.SIVcpzMB897	..QVOY	..VT-EL-D-KKQV-SL-RE-TSL-SNKT	..VKNGT-R-IN-TA	..T	P-L-DON-K-K-T-R 231						
CPZ.GA.88.SIVcpzGAB2	..NGIEKDLGADQMR	..T-TT-EL-D-K-QI-SL-W-N-LVGTN	..TFR-IN-TA	..T	P-L-D-DYP-K-K-K 239						
CPZ.TZ.09.TAN1	..NSATTTTPTPNL	..VT-EL-D-KQKV-SL-VD-L-TNNSG	..TAV	..T	P-L-DON-K-K-T-R 231						
CPZ.TZ.06.SIVcpzTAN13	..EPPTSKTTV	..VT-EL-D-KKQV-SL-VD-L-TNNSG	..TAV	..T	P-L-DON-K-K-T-R 231						
CPZ.TZ.09.UG38	..TTTIIATITNPLNLTGI-DYKVL	..QT-EF-D-KKQI-SL-RE-VMKESQ	..NKTNNSYE-Y-HN-TA	..E-D-TT	P-R--Y-H-L-R-DTYS-FM-N-SK 240						
CPZ.US.85.US Marilyn	..SSAVNTTDMR	..VT-EL-D-KKQV-SL-VD-LAH-N-N	..T-R-IN-TA	..T	P-L-DE-D-K-K-E-K 226						
GOR.CM.04.SIVgorCP684con	..PTSTETTTIPIPELD-F	..VT-VLKD-KT-Q0-L-Q-TET-KNE	..T-R-IN-TS	..L-Q	P-YPL-DEO-E-N 234						
GOR.CM.07.SIVgor2139 287	..TIPPHITPHIT	..VT-VLKD-KT-Q0-L-RQ-LTGT	..GTNGIK-R-IN-TS	..L-Q	P-YLM-DT-O-F-R 239						
GOR.CM.07.SIVgorCP2135con	..TTIPPTTIPD-LV	..VT-VLKD-KT-Q0-L-Q-TET-KNE	..NETKEYS-R-IN-TS	..L-Q	P-YLM-DT-O-F-R 239						
GOR.CM.12.SIVgor BPID1	..TTAAAPCNTENTTKENSTGMLT-N	..VT-VLKO-KEQKO-L-RE-LASLESNN	..R-IN-TS	..L-Q	P-YLM-RD-T-NE 252						
GOR.CM.12.SIVgor BQID2	..RTSXP	..TSSHNMSTPNPEL-VY	..ST-R-IN-TS	..L-Q	P-YLM-QSK-T-XL 260						
GOR.CM.13.SIVgor BPID15	..SAPINASLQSSSTAQNNLNFI	..VT-VLKO-KE-QO-L-ENLV-LNQTSEKLL	..R-IN-TT	..K-L-Q	P-KY-LI-KQ-G-X-XL-NX 254						
MAC.US.x.239	..TTSTTA	..SAKVM	..VNSTSSCIAQDNCT-GL-QEOMIS-K	..MT-GLK-D-KKEYNETW-SA-LVCEQ	..GNNTGNSRCYMH	..QES-D-HYWDA-RFR	..P-Y-L-R-DTNYS-FM-K-SK	257			
H2A.DE.x.BEN	..SRPPTS	..AAAI	..INETSNCIENNTCA-GLGYE-MMO-E	..MK-GLQD-KRRYKDTW-LE-VVCDN	..TTRAGTCYMRH	..I-KES-D-HYWDAMFR	..P-L-R-DTNYS-FE-K-K	251			
H2A.PT.x.ALI	..KPL	..INESDPCIKADNCPRLGDPE-MV-R	..MT-GLQ-D-PKQYNETW-SK-VCEPE	..GLED-DRKQVNETW-SK-VCEPE	..FNITTTNO-RCYMH	..ES-D-HYWDA-RFR	..P-L-R-DDTYS-FM-N-SK	248			
H2A.SN.86.ST JSP4 27	..NTT	..IGENSCIRDNCT-GLGEE-MVD-Q	..MT-GLD-E-KKLYNETW-SK-VVCESE	..VTGTEKQVYMRH	..ES-D-HYWDAMFR	..P-L-R-DTNYS-FE-N-SK	245				
H2B.CI.x.EH0	..SSSQ	..TL	..LNEDSKCIGNDSCA-GLGEE-MVD-Q	..KMT-GLK-DESKYLE-MQ-L	..ESKQYK-T	..TL-R-DTNYS-FM-N-SK	250				
H2B.GH.86.Q295 AL7	..RGLKT	..INEIDPCIKDNDSCT-GLGEE-MO-N	..SMT-GLK-DEKLYKDTW-SE-LECNM	..TRKYTSRCYLR	..TI	..QES-D-HYWDLRFR	..P-FL-DTNYS-FM-N-SK	251			
H2G.CI.92.Abt96	..TPVTKMVT	..EL	..VNSTSOCLMYDNCT-GI0SESMVG-K	..MT-GLK-DRKKEYNETW-SQ-LVCEQ	..SNKSNESRCYMRH	..QES-D-HYWDA-RFR	..P-L-R-DTNYS-FM-N-SK	257			
H2U.CI.07.071C TMP3	..PNTTAN	..ITGRSEE	..INGTDPICISQNNCT-GL-SEPMW-Q	..MT-GLR-DORKEYSETW-SR-VLCEO	..SNNDTNKSRCYMH	..QES-D-HYWDA-RFR	..P-L-R-DTDSYS-F5-K-SK	260			
H2U.FR.96.12034	..PAMQESS	..KVQ	..INETEPCIRNNCS-GL-QEPLVS-R	..MT-GLK-D-K-EYEKTE-SH-VCEO	..NTSESESKYMH	..QES-D-HYWDA-RFR	..P-Y-L-R-DTNYS-FM-N-SK	253			
ASC.UG.10.RT83	..KAPPTT	..LWS	..ORPATEWGW-KDQPSLL	..T-MT-PGFKDKAHYW-P-T	..TNGTGE-YKY	..AM-D-FH-Q-F-V	..P-L-DTDM-Q-I-K	239			
COL.UG.x.G01	..T	..KND	..SAGDWDNT-QWV-ETWYE-K	..STAFLLKN-RKLESLG-SIE-LDISS-R	..NSNIRATMKD-KNYTV	..V-DHTIMD-VSTGF	..AP-YMF-R-DD-KW--Q-S	236			
COL.UG.10.BWC07	..SVAP	..EFDEK-NWP-D-WYH-RM	..STAFLLKN-RKLESLG-SIE-LDISS-R	..SNNIRATMKD-QNYVSVAI	..DESIL-TS-GF	..AP-Y-L-V-DTQW	..R-NQ	238			
DEB.CM.04.SIVdeb04CMPF3061	..ATTT	..KT.TT	..DWSGE-NNTMSQYV	..V-GPY-D-KEEG-MMLTD-LEWANGNETGDT	..RGAYMR-R	..D-SH-K-FK-R	..YGLMR-DD-N--L-	246			
DEB.CM.99.CM40	..DR	..KT.TT	..DWSGE-NITMTQYV	..V-GPY-D-KE-SH-VLLED-OWA-KDQSGN	..RTGYMKH-D	..E-STR-K-FK-R	..YGLR-DRN--L-N	247			
DRL.DE.11.D3	..VAITVAITSTTAPKNASVAP	..TTIANGTLDITDITNTEK	..VTEANR-KY-VT	..GLC-DCKEIEITN-RYD-VCKGGSNTEPSE	..KSRIGYCYMH	..D-STR-K-FK-R	..YGLR-DRN--L-N	247			
DRL.DE.11.D4	..ATSTSTTSA	..APNTTA	..DLEIDRNTEP	..KTEINR-KY-VT	..GLC-DCKEIEITN-RYD-VCKGGSNTEPSE	..NCTE-SRICVYH	..ST-D-N-A-TDMMKFLR	..P-YIF-R-E-LNKTRK	265		
GRV.ET.x.GRI 677 gri 1	..GNSN.CM.99.CN166	..GNSN.CM.99.CN171	..STTT	..QWSSGGENGTGOPYL	..QT-EF-D-KKQV-SL-WRE-MEETHG	..NSQV-YLRN	..Y-K-V-S-Q-V-T	P-Y-M--DYN-T-V-T-Y 252			
LST.CD.88.SIVLhoest485	..P.MPCFIAET	..TTTTTKGNTT	..MESLIEDTNCIRLW-NETTEINRE	..Y-VT-GLC-DCKEIEITN-RYD-VCKGGSNTEPSE	..ND-D-YIIN	..D-YV-K-V-T-Q-V	..P-M--DAN-T-V-K	534			
LST.CD.88.SIVLhoest524	..P.MPCFIAET	..TTTTTKSSTE	..QPSLIEDTNCIRLW-NETTEINRE	..Y-VT-GLC-DCKEIEITN-RYD-VCKGGSNTEPSE	..TCYM-H	..D-I-L-D-N-GVSNAYFRL	..YML-E-LNFSAN	257			
LST.KE.x.lho7	..TMPQCFINEQV	..TVKPNQ	..ETRLEEDLNCTRL-NETTERNAE	..QY-VT-GLC-DCKEIEITN-RYD-VCKGGSNTEPSE	..ENRTCYM-H	..D-I-D-N-GVSNAYFRL	..YML-R-EDNFASK	253			
MAL.ZM.x.SIVagmMAL ZMB	..TKST	..QRPCKQYERK	..NTLQACNDTIIEE-E-DEEPAS-T	..AMA-GY-DOKKNSVVDMDRE-VCKNGSDN	..SKDCYMIH	..D-KE-D-TYWDLRLR	..Y-L--DHDY-YKON-S	261			
MMD.1.GA.x.MNDG81	..TSTVASST	..KTV	..ANTIELDITDITNTEK	..VTEANR-KY-VT	..GLC-DCKEIEITN-RYD-VCKGGSNTEPSE	..NCTE-SRICVYH	..ST-D-N-A-TDMMKFLR	..P-YIF-R-E-LNKTRK	265		
MMD.2.CM.98.CM16	..PTTP	..KTV	..AVTGLEVTVONN-ESTIQ-NRV-K	..TT-GLC-DCKEIEITN-RYD-VCKGGSNTEPSE	..KENDTEIECYM-H	..D-N-A-TDMMKFLR	..P-YIL-R-DTYS-FM-N-SK	243			
MMD.2.GA.x.M14	..TTTTPTTIT	..SSSTN	..KTTTPVLVVEKQNN-ETTQQNR-K	..TT-GLC-DCKEIEITN-RYD-VCKGGSNTEPSE	..TNSTPEYECYM-H	..AT-DRN-A-TDRMTFLR	..P-VL--E-LNKTKL	264			
MNE.US.x.MNE027	..PTTKITTTT	..KEIEV	..VNESTCVNRDNT-GL-QEPMIS-K	..MT-GLK-D-KREYNETW-SA-LVCEQ	..GNSTDESRCYMH	..QES-D-HYWDA-RFR	..P-Y-L-R-DTYS-FM-N-SK	258			
MON.CM.99.L1 99CM1	..P	..KTV	..AVTGLEVTVONN-ESTIQ-NRV-K	..TT-GLC-DCKEIEITN-RYD-VCKGGSNTEPSE	..SH-YLLN	..Y-Q-S-N-V-L	..P-Y-L-R-DDPA-T-Q-S	238			
MUS.1.CM.01.CM1239	..GHTTVPWGRWDA	..NCTEQSMF	..GHTTVPWGRWDA-NETQLTNY	..QT-EF-D-KKQV-SL-TD-LNRE	..T-YLLN	..Y-Q-S-N-V-L	..P-Y-L-R-DDPA-T-Q-S	248			
MUS.2.CM.01.CM1246	..TAKAP	..TTPEIIGMWMRSTGOGVY	..QT-EF-D-TKOM-SL-VD-VMSQ-GNVTE	..T-YI-N	..Y-Q-S-N-V-L	..YSL-DAN-T-Q-M-N	252				
MUS.3.GA.09.09gab0181	..TAKAP	..TTPEIIGMWMRSTGOGVY	..QT-EF-D-TKOM-SL-VD-VMSQ-GNVTE	..T-YI-N	..Y-Q-S-N-V-L	..YSL-DAN-T-Q-M-N	252				
OLC.CI.97.97C112	..TPSAS	..TN-IDCGIN	..NETCS	..AV-DENVMO-T-AVA	..GLK-DEKH-YNDTW-SR-LWCQAN	..NS-OCFMRH	..TS-H-F-EPRVY-FRLR	..P-L--DYNVT-FDT	244		
RCM.CM.00.SIVag1 00CM312	..TFTPAK	..ITKVDPCPM	..NETCA	..AVPQDVM-E-AVA	..GLK-DEKH-YNDTW-SR-LWCKDQV	..NKTTKROCFMRH	..D-F-EPRVY-FRLR	..P-L-V-KO-NVT-FDT	253		
RCM.GA.x.SIVRCMGA81	..KTTTQRT	..IGVEKCTAG	..NETCE	..EVQDAQVMS-E-AVA	..GLK-DEKH-YNDTW-SR-LWCKEET	..NSTNSTRKCCYMRH	..NS-TNSTRKCCYMRH	..Y-L-KO-NVT-FDT	253		
SAB.SN.x.SAB1	..PEVVS	..VGFND5VIEQ	..E-EQAM-K	..AMA-GYR-DVKNKNSVVDMDQEVVCEEGREKS	..NATHTVGCMYH	..KE-D-TYWDTRFR	..Y-L-R-ADTYS-HKA	251			
SMM.CI.79.SIVsmC12	..TTTTTAVTATAVTTTAGGDI	..INDTTEICIRNDS-GL-PEPLIG-K	..SMT-GLR-D-KKEYNETW-SQ-LVCEQ	..VNDTTEICIRNDS-GL-PEPLIG-K	..SMT-GLR-D-KKEYNETW-SQ-LVCEQ	..GNSERCYMH	..QES-D-HYWDA-RFR	..P-Y-L-R-DTNYS-FE-N-K	265		
SMM.SL.92.SIVsmS192A	..ATTKASPTLLTPVKE	..LNDTSCFMTNDTCP	..GL-QEPMIG-K	..SMT-GLR-D-KKEYNETW-SQ-LVCEQ	..VNDTTEICIRNDS-GL-PEPLIG-K	..SMT-GLR-D-KKEYNETW-SQ-LVCEQ	..GNSERCYMH	..I-QES-D-HYWDVFR	259		
SMM.SL.92.SIVsmS192B	..PTTTSPLTAASPSGE	..EI	..VNDTSCFMTNDTCP	..GL-QEPMIG-K	..SMT-GLR-D-KKEYNETW-SQ-LVCEQ	..VNDTTEICIRNDS-GL-PEPLIG-K	..SMT-GLR-D-KKEYNETW-SQ-LVCEQ	..GNSERCYMH	260		
SMM.SL.92.SIVsmS192C	..TPSKAKTTEKGE	..EI	..VNDTSCFMTNDTCP	..GL-QEPMIG-K	..SMT-GLR-D-KKEYNETW-SQ-LVCEQ	..VNDTTEICIRNDS-GL-PEPLIG-K	..SMT-GLR-D-KKEYNETW-SQ-LVCEQ	..GNSERCYMH	260		
SMM.US.11.SIVsmE660_FL8	..K	..TLKSTSTTTPVLE	..VNDTEPCVKNLNNCT-GL-PEPMIG-K	..MT-GLK-D-KREYNETW-SA-LVCEQ	..GNSERCYMH	..QES-D-HYWDA-RFR	..P-Y-L-R-DTNYS-FM-N-SK	260			
SMM.US.86.CFU212	..QTSVTPPSPITIA	..KV	..VNDSDPCIRSNCT-GL-QEPMIS-K	..MT-GLK-D-KREYNETW-SA-LVCEQ	..GNSERCYMH	..QES-D-HYWDA-RFR	..P-Y-L-R-DTNYS-FM-N-SK	260			
SUN.GA.98.L14	..WTITAKTTTQTSSTSTAS	..TTTPMPLDWNCTDT	..ENIAESN-V-KY	..VT-GLC-DCKEIEITN-RDTEVTCNG	..TCYM-H	..D-I-D-GIMQVAYFRL	..YML-R-DE-LNATKK	K 267			
VER.KE.x.KE51	..K	..LPCVRNKTD	..SNLQSCNDTIEE-E-NDEAAS-T	..AMA-GYR-DOKKNSVVDMDAE-FCKRST5	..GNSCY-Y-QH	..MRN-E-OT-R-F-V	..P-VSL--DTE-E-DIE	0 242			
VER.ZA.10.V5A40904	..SVK-KR	..SVK173 COMGNM	..NETCE	..EVQDAQVMS-E-AVA	..GLK-DEKH-YNDTW-SR-LWCKEET	..NSTNSTRKCCYMRH	..NS-TNSTRKCCYMRH	..Y-L-KO-NVT-FDT	253		
VER.ZA.10.V5A41093	..TAL	..CM.00.266	..TPT	..AMSNMPPWGW-DNTTEPZY-TW	..MT-GGF-D-KYTYFSHW-LA-LMKPEGN	..NSDY-YA-G	..L-A-E-THYOAF-Q	..LI--DNMPT-R-K-H	250		
VER.ZA.10.V5A0022	..TNSVTS	..KRTNDTNVWVG-NNATEPVY	..TW-MT-GGF-D-KYTYFSHW-LA-LMKPEGN	..NSDY-YA-G	..L-A-E-THYOAF-Q	..LI--DNMPT-R-K-H	250				
VER.ZA.10.V5A0022	..TNSVTS	..KRTNDTNVWVG-NNATEPVY	..TW-MT-GGF-D-KYTYFSHW-LA-LMKPEGN	..NSDY-YA-G	..L-A-E-THYOAF-Q	..LI--DNMPT-R-K-H	250				
WRC.CI.98.Q8C104	..TAN.UG.x.SIVagmTAN1	..TDPKPNTE	..SSCNATLVN-S-DYENSSI	..AMA-GYR-DVKK-YNSWT-DQELVCEKENNTT	..GTRGCMYH	..D-KE-E-TYWDTRLR	..KDNVT-F-V-R	252			
WRC.CI.98.Q8C104	..VER.KE.x.AGM155	..LPCVRNKTD	..SNLQSCNDTIEE-E-NDEAAS-T	..AMA-GYR-DOKKNSVVDMDAE-FCKRST5	..GNSERCYMH	..D-KE-E-TYWDTRLR	..Y-L--DNDYA-FK-S	256			
WRC.CI.98.Q8C104	..VER.ZA.10.V5A40904	..SVK-KR	..SVK173 COMGNM	..NETCE	..EVQDAQVMS-E-AVA	..GLK-DEKH-YNDTW-SR-LWCKEET	..NSTNSTRKCCYMRH	..NS-TNSTRKCCYMRH	253		
WRC.CI.98.Q8C104	..VER.ZA.10.V5A41093	..TAL	..CM.00.266	..TPT	..AMSNMPPWGW-DNTTEPZY-TW	..MT-GGF-D-KYTYFSHW-LA-LMKPEGN	..NSDY-YA-G	..L-A-E-THYOAF-Q	..LI--DNMPT-R-K-H	250	
WRC.CI.98.Q8C104	..VER.ZA.10.V5A0022	..TNSVTS	..KRTNDTNVWVG-NNATEPVY	..TW-MT-GGF-D-KYTYFSHW-LA-LMKPEGN	..NSDY-YA-G	..L-A-E-THYOAF-Q	..LI--DNMPT-R-K-H	250			
WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..TAN.UG.x.SIVagmTAN1	..TDPKPNTE	..SSCNATLVN-S-DYENSSI	..AMA-GYR-DVKK-YNSWT-DQELVCEKENNTT	..GTRGCMYH	..D-KE-E-TYWDTRLR	..KDNVT-F-V-R	252		
WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..VER.KE.x.AGM155	..LPCVRNKTD	..SNLQSCNDTIEE-E-NDEAAS-T	..AMA-GYR-DOKKNSVVDMDAE-FCKRST5	..GNSERCYMH	..D-KE-E-TYWDTRLR	..Y-L--DNDYA-FK-S	256		
WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..VER.ZA.10.V5A40904	..SVK-KR	..SVK173 COMGNM	..NETCE	..EVQDAQVMS-E-AVA	..GLK-DEKH-YNDTW-SR-LWCKEET	..NSTNSTRKCCYMRH	..NS-TNSTRKCCYMRH	253	
WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..VER.ZA.10.V5A41093	..TAL	..CM.00.266	..TPT	..AMSNMPPWGW-DNTTEPZY-TW	..MT-GGF-D-KYTYFSHW-LA-LMKPEGN	..NSDY-YA-G	..L-A-E-THYOAF-Q	..LI--DNMPT-R-K-H	250
WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..VER.ZA.10.V5A0022	..TNSVTS	..KRTNDTNVWVG-NNATEPVY	..TW-MT-GGF-D-KYTYFSHW-LA-LMKPEGN	..NSDY-YA-G	..L-A-E-THYOAF-Q	..LI--DNMPT-R-K-H	250		
WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..TAN.UG.x.SIVagmTAN1	..TDPKPNTE	..SSCNATLVN-S-DYENSSI	..AMA-GYR-DVKK-YNSWT-DQELVCEKENNTT	..GTRGCMYH	..D-KE-E-TYWDTRLR	..KDNVT-F-V-R	252	
WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..VER.KE.x.AGM155	..LPCVRNKTD	..SNLQSCNDTIEE-E-NDEAAS-T	..AMA-GYR-DOKKNSVVDMDAE-FCKRST5	..GNSERCYMH	..D-KE-E-TYWDTRLR	..Y-L--DNDYA-FK-S	256	
WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..WRC.CI.98.Q8C104	..VER.ZA								

Table of HIV protein sequences including glycosylation NYS, glycosylation NGS, glycosylation NNT, glycosylation NCT, V3 tip, glycosylation NNT, glycosylation NKT, and CD4 binding. Sequences are aligned and numbered from 350 to 380.

Table of HIV protein sequences including glycosylation NYS, glycosylation NGS, glycosylation NNT, glycosylation NCT, V3 tip, glycosylation NNT, glycosylation NKT, and CD4 binding. Sequences are aligned and numbered from 385 to 400.

Table of glycosylation NST, glycosylation NNT, CD4 binding, glycosylation NIT, glycosylation NES, and CD4 binding domains. Includes protein names like H1B, H1A1, H1F1, etc., and their corresponding amino acid sequences.

Table of glycosylation NST, glycosylation NNT, CD4 binding, glycosylation NIT, glycosylation NES, and CD4 binding domains. Includes protein names like MAC, H2A, H2B, etc., and their corresponding amino acid sequences.

Table with columns for protein names, amino acid sequences, and accession numbers. It is organized into sections: gp120 end_gp41 start, fusion peptide, glycosylation NAS, immunodominant region, glycosylation NKS, glycosylation NHT, and glycosylation NYT. Each row represents a different HIV sequence entry.

	transmembrane domain	gp41 cytoplasmic tail start	glycosylation	NGS
H1B.FR.83.HXB2	EEESNOQEKEQELLELDKQAS.LWNWFNITNMLWYLKLTIMVIGGLVGLRIVFAVLSIVNVRVROGYSPLSFQTHLPT	PRGDPDRPEGIEEEEGERRDR	RSIRLVNGLSLAIWDD.LRSLCLFSYHRLRDLLLIVTRIVELVGG.....RRGWALKYW..
H1A1.UG.85.U455.U455ALD-A-N*.....S.....R-VI-I-T-I.....LA-I.....E-LG-GR.....QGK.....S-F-IA.....N.....FA-A-A.....RSSLKG-L-L-G-L
H1C.ET.86.ETH2220V-D-KD-A-EN.....
H1F.ET.87.VI1850
H10.CM.91.MVP5180
H10.CM.98.98CM1A05
H10.CM.98.98CMAB197
H10.SN.99.99SMP1299
H10.US.99.99U171A
H1N.CM.02.J018131
H1N.CM.04.04CM.1015.04
H1N.CM.04.04CM.1131.03
H1N.CM.95.YBF30
H1P.CM.06.U14788
CPZ.CM.06.R1167
CPZ.CM.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.GA.88.SIVcpzGAB2
CPZ.GA.09.CM171
CPZ.TZ.06.SIVcpzTAN13
CPZ.TZ.09.UG38
CPZ.US.85.US.Marilyn
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139.287
GOR.CM.07.SIVgorCP2135con
GOR.CM.12.SIVgor.BPID1
GOR.CM.12.SIVgor.BPID2
GOR.CM.13.SIVgor.BPID15
MAC.US.x.239
H2A.DE.x.BEN
H2A.PT.x.ALI
H2A.SB.86.S2.JSP4.27
H2Z.CI.x.EH01
H2B.GH.86.D285.ALT
H2G.CI.92.Abt96
H2U.CI.07.07IC.TMP3
H2U.FR.96.12034
ASC.UG.10.RT03
COL.CM.x.C011
COL.UG.10.BW01
COL.UG.10.BW07
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.09.CM40
DEB.CM.09.CM5
DRL.DE.11.D3
DRL.DE.11.D4
GRV.ET.x.GRI.677.gri.1
GSN.CM.99.CN166
GSN.CM.99.CN171
LST.CD.88.SIVlhoest485
LST.CD.88.SIVlhoest524
LST.KE.x.lho7
MAL.ZM.x.SIVagmMAL.ZMB
MND.1.GA.x.MNDGB1
MND.2.CM.98.CM16
MND.2.GA.x.M14
MND.2.GA.x.5440
MNE.US.x.MNE027
MON.CM.99.L1.99CM1
MUS.1.CM.01.CM1239
SVK.KE.x.KC51
MUS.3.GA.09.09gab0181
OLC.CI.97.97C112
RCM.CM.00.SIVagi.00CM312
RCM.CM.02.SIVrcm.02CM8081
RCM.GA.x.SIVRCMGAB1
SAB.SN.x.SAB1
SHM.CI.79.SIVsmC12
SHM.SL.92.SIVsmSL92A
SHM.SL.92.SIVsmSL92B
SHM.SL.92.SIVsmSL92C
SHM.US.10.SIVsmE660_FLB
SHM.US.86.CFU212
SHM.US.x.H9
SUN.GA.98.L14
SVK.KE.x.KC51
VER.ZA.10.VS4C4094
VER.ZA.10.VS4G1093
VER.ZA.10.VS4M0022
WRC.CI.98.98C104
WRC.GM.05.PB05GM.X02

Env end
gp41 end
glycosylation NAT
cytoplasmic tail end

H1A.FR.83.HXB2
H1A1.UG.85.U455.U455A
H1C.ET.86.ETH2220
H1F1.BE.93.V1850
H10.BE.87.ANT70
H10.CM.91.MVP5180
H10.CM.98.98CMA105
H10.CM.98.98CMA187
H10.SN.99.99S5.MP1299
H10.US.99.99USTWLA
H1N.CM.02.DJ08131
H1N.CM.04.04CM.1015.04
H1N.CM.04.04CM.1131.03
H1N.CM.95.YBF30
H1P.CM.06.U14788
CPZ.CM.06.8P5167
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715
CPZ.CM.05.SIVcpzDP943
CPZ.CM.09.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.00.TAN1
CPZ.TZ.06.SIVcpzTAN13
CPZ.TZ.09.UG38
CPZ.US.85.US.Marilyn
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139.287
GOR.CM.07.SIVgorCP2135con
GOR.CM.12.SIVgor.BPID1
GOR.CM.12.SIVgor.BPID2
GOR.CM.13.SIVgor.BPID15

MAC.US.x.239
H2A.DE.x.BEN
H2A.PT.x.ALI
H2A.SN.86.ST.JSP4.27
H2B.CI.x.EHO7
H2B.GH.86.D295.ALT
H2G.CI.92.Abt96
H2U.CI.07.07IC.TMP3
H2U.FR.96.12034
ASC.UG.10.RT03
COL.CM.x.CG1
COL.UG.10.BWC01
COL.UG.10.BWC07
DEB.CM.04.SIVdeb04MPCF3061
DEB.CM.99.CM40
DEB.CM.99.CM5
DRL.DE.11.D3
DRL.DE.11.D4
GRV.ET.x.GRI.677.gri.1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest485
LST.CD.88.SIVlhoest524
LST.KE.x.lho7
MAL.ZM.x.SIVagmMAL.ZMB
MND.1.GA.x.MNDGB1
MND.2.CM.98.CM16
MND.2.GA.x.M14
MND.2.x.x.5440
MNE.US.x.MNE027
MON.CM.99.L1.99CM1
MUS.1.CM.01.CM1239
MUS.2.CM.01.CM1246
MUS.3.GA.09.09GaB0181
OLC.CI.97.97C112
RCM.CM.00.SIVag1.00CM312
RCM.CM.02.SIVrcm.02CM8081
RCM.GA.x.SIVrcmGAB1
SAB.SN.x.SAB1
SMM.CI.79.SIVsmC12
SMM.SL.92.SIVsmSL92A
SMM.SL.92.SL92B
SMM.US.06.FL8
SMM.US.11.SIVsmE660.FL8
SMM.US.86.CFU212
SMM.US.x.H9
SUN.GA.98.L14
SVK.KE.x.KES1
SVK.KE.x.SVK173.COMGNM
TAL.CM.00.266
TAL.CM.01.8023
TAN.UG.x.SIVagmTAN1
VER.KE.x.AGM155
VER.ZA.10.VSAC4004
VER.ZA.10.VSAG1003
VER.ZA.10.VSAM0022
WRC.CI.98.98CI04
WRC.GM.05.Pbt.05GM.X02

.....WNLLQYWSOELKNSAVSLLNATAIAVAEGTRDRIEIVV...QGACRAIRHPRIRROGLER..ILL.*
.....NCLRDCAAFG...T...TV...S...NW...0...A...G...OTIG...GFLN...L...T...I...A...N...
.....N...STL...TIIRTEIKNIDRL...W...G...K...S...ILLAL...T...IV...I...EV...I...A...N...
.....RGII...GR...RT...T...D...G...I...RA...IVF...I...GN...T...
.....KGV...GR...V...I...DT...I...A...RFG...G...LN...T...A...
.....GVIA...GR...QI...T...DT...I...L...RIG...G...LN...T...S...
.....GIIA...GR...I...IN...DT...V...I...L...RIG...G...LN...T...S...
.....QGI...GK...T...I...DT...I...I...RIG...G...LN...F...A...
.....RGI...G...T...I...DT...I...IA...RFG...G...LN...A...
.....KSI...G...T...I...D...I...IA...RFG...G...LN...A...
.....KIGT...T...RT...T...DT...I...R...T...A...TVG...G...LN...A...E...
TIIVGVROI.IEWSNTYLS.RVLLIYAITL.NFTGWW...G...N...A...N...
KRWLF.RDW.ANAVIV.T.FIRN.DRL.NFT.WW...IL...GA...VNIF...G...N...T...I...A...N...
KSRQL...FDW.SNTYSI.RT.LIOAIDRL.NFTGWW...L...AG...AFVAQG...N...I...A...N...
.....GVIA...AR...V...T...DT...I...LT...RRLFLG...I...S...X...
.VNNLTRDC.FAFIA...G...0...I...DCV...VWT...DW...0...AIA...RIG...G...LN...S...
.VNNLTRDC.FAFIA...G...0...I...DCV...VWT...DW...0...AIA...RIG...G...LN...S...
.VNNLTRDC.FAFIA...G...0...I...DCV...VWT...DW...0...AIA...RIG...G...LN...S...
.GLNCVRDC.LAVCG...V...OR...TN...DT...V...W...0...LIG...RIG...G...LN...S...
.TFEACKTL.KATA...L...QR...XTN...DTV...V...NW...SI...LG...RFG...G...LN...L...S...
.GINCLRDC.LAVCG...A...QQ...T...DTV...VR...DW...0...LG...RIG...G...LN...L...S...

Env end
gp41 end

.....LTY...GWSYFHEAVQVWRSATETL...GAWGDLW...TL...RRGG...W...LA...L...T...
.....AAQ...GCEWIEAFAQAFART...RETL...GAWGWLW...AA...RRI...G...LAV...A...L...A...
.....AAY...GCEWIEAFAFAFARIARETL...TNRDRLWGA...WVG...R...LAV...A...I...A...
.....IAY...GGEWIEAFAFAFAR...GETL...TNAWGRFWGTL...GOIG...G...LAV...A...I...A...
.....AYR...GTSWFQEAIAAAR...AETL...SARTSG...L...RR...AGE...IA...A...L...A...
.....VAY...GTSWFQEAIAAAR...GETL...SAGETLW...AL...RR...A...IA...L...T...
.....AAYFS...GFRW...QEACTAATR...AQETL...TSTWRALWKT...L...T...
YGQWF...QEA...AGWKAXRETLA...AWRDWELT...GRVGLIA...F...L...A...
.....IAY...GW...YF...EAFQAFGK...ARETL...SRTGRELW...TL...GRVG...W...LA...F...L...A...
EATTRA...VAYTH...GL...F...AAVT...AWD...L...FTWNV...EA...LAYL...RRLAG...A...L...N...
QHPPFTSDK.RRS.A...T...RYL...LVRP...PDSLIP...T...LS...WLSA...TRRW...RASEYL...GWLFRDPE...SPA...
ORPETFNTR.RTVE...T...W...L...LVPR...CDLIP...T...LS...WLSA...TRRW...RASEYL...GWLFRDPE...SPA...
EQPLPDGET.HOPRRMPL.A.WRL...LVPR...SANL...IPOTLQ...WLSA...TS...G...RALESCG...WIR...DPO...GPA...
QLRVL...GY...GWK...F...DTI...WAGGELGOTFORWAEVALQ...L...GRGT...E...LA...A...A...L...F...N...
RFRVL...GGY...GWK...F...DTI...WAGGELWQWLOGIAEMALQGL...VR...GG...LRV...A...A...L...N...
HLRVV...GYC...GWK...F...DTI...WAGGELWQWLOGIAEMALQGL...VR...GG...LRV...A...A...L...N...
.....QEAGRIC...TLLREAQRLWRRGROGLSARAL...R...LAQVAN...A...I...L...N...
.....QEAGRIC...ALLQEAERLWRRGROGLSTARAL...RAFAQ...LAN...A...V...L...N...
.....Y...GL...Q...QEA...TG...AQ...L...R...AR...AWG...LGAI...RS...Y...VTNS...V...K...V...G...
VAARV...IAYIS...GI...QAA...TGI...DSL...FTWNV...EA...LHAC...RRWV...EFLA...A...I...LFN...
VAARA...IAYIS...GI...QAA...VSGI...DSL...FTWNV...EPLVOT...GRWV...EFLA...A...I...LFN...
FFHQGHLLKASSGMV.ENAOKI...RWSW...KTRNRNRDISAHARDIRLG...KRWRF...FRG...SGF...PSETTETL...
SYHHGHLLWOTLSRFLV.ESAOKT...SWIR...KSRNRG...LSKTNGTNRGL...NRR...W...FRF...SGL...PSETTETL...
FFQNGORLWOTSSRMV.ENAOKIQSW...REKCRNRNRLSSTDRKNIGLQKKRWRL...FGG...SGI...SSEATETAL...
.....Y...GFG...AA...QEAAR...V...SFAQNAAHQIWLAC...RS...Y...TNS...V...EC...P...N...
QEA...IRS...LGS...LVFPAISWR...GYTGWLO...NRIFT...CREA...IAA...GTC...W...L...SA...P...N...
.....KTAGKIF...VLOGYAOGLWSRGHGW...LSTAACFR...ATA...G...LN...A...V...L...N...
.....KAAGARICQVLSFAORLWSGGHQL...LSVIRGAA...AIG...GN...A...V...L...N...
.....KAAGAALYQVLOGFAORLWSRGYQL...LSG...RGA...AFG...G...WN...A...A...L...N...
LTY...GWSYFHEAVQVWRSATETL...GAWGDLW...TL...GRVG...W...LA...E...L...T...
RRL.LAYC...GI...QAAVT...D...L...RFTVW...ALLHAG...GRLW...VA...A...I...F...N...
RCELLARL.LAYT...GI...QAAVT...D...L...RFTVW...ALLHAG...GRLW...VA...A...I...F...N...
ALGRFAH...GI...K...QAAVSGA...D...FTWNV...EPLQFS...KIW...KFLA...A...I...LFN...
RLREVIRRI.IAY...GL...QEA...TGF...D...L...FTWNV...EA...LO...L...RRVA...EFLA...A...I...LFN...
.....WI...OKSLQHVWRTLKOKRO...HGINLQOKK...RRCSPLOKAOQLGREAG...TTG...
.....AKV...DGV...F...AF...W...GEMA...ENAYTYWRGLQSLA...ROLAGWPATCG...L...N...
.....AKI...DGV...EFS...GW...AEMAQONAYFYWRGLCVTA...DIA...HPCTVC...F...L...I...
AKI...GW...F...EFS...FAELANAYTYWRGLCA...A...R...G...W...P...M...
DIQQRTRQ...TAHLEYGWQELKAAASL...VAV...QAA...SASDS...RHACRSIV...VIAH...M...E...W...N...
.....IAY...GW...YFQEAQEAAR...ARETL...SAGRDW...TL...GRIG...GLLAA...L...AF...
VAYIR...GW...YFIE...LOEAAI...AROTL...VSTGRITWOTL...GHIG...G...AA...L...T...N...
.....YTR...GMWYFEACQEAWSAQE...IVGAWGLW...TL...GRVG...G...AA...L...M...N...
LTY...GWSYFHEAVQVWRSATETL...GAWGDLW...TL...GRVG...W...LA...E...L...T...
.....IAY...GWRV...QEA...QAWMK...FARETL...SAWRDIW...TL...GRVG...G...LA...V...L...T...
.....LTYIG...GCLWF...EALQAAGR...AQOTL...SAWRDIW...TL...GRVG...G...LN...L...T...
.....Y...GWRV...QEA...QXWMKFXRETL...SAWRDIW...TL...GRVG...G...LA...X...L...T...
SFHHQRLWTLRGWGL...HLIQATSRI...RACRTS...RERVSSQKARSRTFSL...RKWRPKWNR...TSRI...PSETTETL...
FIS...GVT...EA...I...GREV...HNVATWQALL...AYA...RRVAENVAAL...L...T...Y...N...
FIS...GFN...IA...A...GREIRDW...ATWQAIYAA...RRVVE...VAAL...L...I...Y...N...
AAAATLQY...GFS...IQAAWRDGT...L...RFTWVW...NG...LAAA...RW...GE...AA...T...Y...CFT...
AAV...ATI...GFS...IQAAWDCVT...L...RFTWVW...SG...LAAA...RW...VQ...AA...T...Y...C...A...
TLQLLQROE.COYI...GW...OF...EG...ARISFE...L...S...AQASRTLWNA...RS...Y...LEH...M...E...W...N...
.....Y...GLG...AA...QEA...V...HL...SFARNAHOIWLAC...RS...Y...TNS...V...E...V...N...
.....Y...H...GLR...QTA...QEA...V...S...ADONAAHOIWLAC...RS...Y...TVINS...V...A...Q...N...
.....Y...GLG...QTA...QEAARTV...RLAONLGHQIWLTC...RS...Y...VVD5...V...EV...N...
.....GFG...AA...QEAAR...V...GVAQNAHOIWL...L...I...T...SS...VWKL...T...L...CSSLNKSRRR...F...
S...GAAAHACIRKIKTSRKETTER...ISENSNRET...T...SS...VWKL...T...L...CSSLNKSRRR...F...
S...GAAAYALQEKIARROKRKEK...TYTPGSRRET...SRSSSRIGITA...L...CTLNLP...CRRGQO...

856
856
852
853
860
873
868
874
879
871
840
833
853
853
885
859
865
913
849
853
846
878
842
857
832
871
865
837
859
855
850
890
888
893

880
861
863
860
853
860
885
868
868
887
867
843
835
847
871
867
900
884
855
882
887
915
917
913
877
879
867
868
891
883
870
892
883
893
844
855
862
867
868
886
887
887
888
888
920
841
833
873
891
871
878
869
873
885
887

H1B_FR.83_HXB2
H1A1.UG.85_U455_U455A
H1C.ET.86_ETH2220
H1F1.BE.93_V1850
H10.FR.87_ANT70
H10.CM.91_MVP5180
H10.CM.98_98CMA105
H10.CM.98_98CMA197
H10.CM.98_98CMA197
H10.SN.99_99SME MP1299
H10.US.99_99USTWLA
H1N.CM.02_GJ08131
H1N.CM.04_04CM_1015_04
H1N.CM.04_04CM_1131_03
H1N.CM.95_YBF30
H1P.CM.06_U14788
CPZ.CD.06_06P1167
CPZ.CD.90_ANT
CPZ.CM.01_SIVcpzCAM13
CPZ.CM.05_LB715
CPZ.CM.05_SIVcpzDP943
CPZ.CM.05_SIVcpzEK505
CPZ.CM.05_SIVcpzLB7
CPZ.CM.05_SIVcpzMB66
CPZ.CM.05_SIVcpzMB897
CPZ.GA.88_SIVcpzGAB2
CPZ.TZ.09_TAN1
CPZ.TZ.09_UG38
CPZ.US.85_US_Marilyn
GOR.CM.04_SIVgorCP684con
GOR.CM.07_SIVgor2139_287
GOR.CM.07_SIVgorCP2135con
GOR.CM.12_SIVgor_BP1D1
GOR.CM.12_SIVgor_BP1D2
GOR.CM.13_SIVgor_BP1D5

Table of HIV-1 protein sequences for the Nef region, starting with 'HXB2 premature Nef end'. The table lists amino acid sequences for various HIV-1 strains, with columns for residue positions and specific amino acid substitutions. The sequences are aligned to show conserved regions and variations.

MAC.US.x.239
H2A.DE.x.BEN
H2A.PT.x.ALI
H2A.SN.86.ST_JSP4_27
H2B.CI.x.EH0
H2B.GH.86_0285_ALT
H2G.CI.92_Abt96
H2U.CI.07_07IC_TNP3
H2U.FR.96_12034
ASC.UG.10_RT03
COL.CM.x.G0U1
COL.UG.10_BWC01
COL.UG.10_BWC07
DEB.CM.04_SIVdeb04CMP3061
DEB.CM.99_CM40
LMI.Y.99_CM5
DRL.DE.11_D3
DRL.DE.11_D4
GRV.ET.x.GRI_677_gri_1
G5N.CM.99_CN166
G5N.CM.99_CN171
LST.CD.88_SIVlhoest485
LST.CD.88_SIVlhoest524
LST.KE.x.lho7
MND-1.GA.x.MNDG81
MND-2.CM.98_CM16
MND-2.GA.x.M14
MND-2.x.x.5440
MNE.US.x.MNE027
MON.CM.99_L1_99CML1
MUS-1.CM.01_CM1239
MUS-2.CM.01_CM1246
MUS-3.GA.09_09gag0181
OLC.CI.97_97C112
RCM.CM.00_SIVagi_00CM312
RCM.CM.02_SIVrcm_02CM8081
RCM.GA.x_SIVRCMgAB1
SAB.ZA.x.SAB1
SMM.CI.79_SIVsmC12
SMM.SL.92_SIVsmSL92A
SMM.SL.92_SIVsmSL92B
SMM.US.96_Ft0
SMM.US.11_SIVsmE660_FL8
SMM.US.86_CFU212
SMM.US.x.H9
SUN.GA.98_L14
SVK.KE.x.KE51
SVK.KE.x.SVK173_CONGMN
TAL.CM.00_266
TAL.CM.01_8023
TAN.UG.x_SIVagmTAN1
VER.KE.x.AEM155
VER.ZA.10_VSAc4004
VER.ZA.10_VSA61003
VER.ZA.10_VSAM0022
WRC.CI.98_98C104
WRC.CM.05_Pbt_05GM_X02

Table of HIV-1 protein sequences for the Nef region, starting with 'max HIV-1 similarity'. The table lists amino acid sequences for various HIV-1 strains, with columns for residue positions and specific amino acid substitutions. The sequences are aligned to show conserved regions and variations.

	normal Nef end	
H1B.FR.83.HXB2NC.....*	205
H1A1.UG.85.U455_U455AD.....	206
H1C.ET.86.ETH2220D.....	208
H1F1.BE.93.VI850D.....	206
H10.BE.87.ANT70D.....	214
H10.CM.91.MVP5180D.....	212
H10.CM.98.98CMA105D.....	214
H10.CM.98.98CMABB197D.....	216
H10.SN.99.99SE_MPI299D.....	212
H10.US.99.99USTWLAD.....	212
H1N.CM.02.DJ08131D.....	214
H1N.CM.04.04CM_1015_04D.....	213
H1N.CM.04.04CM_1131_03D.....	214
H1N.CM.95.YBF30D.....	213
H1P.CM.06.U14788KD.....	218
CPZ.CD.06.8F1167D.....	196
CPZ.CD.90.ANTD.....	199
CPZ.CM.01.SIVcpzCAM13D.....	208
CPZ.CM.05.LB715D.....	207
CPZ.CM.05.SIVcpzDP943D.....	208
CPZ.CM.05.SIVcpzEK505D.....	207
CPZ.CM.05.SIVcpzLB7D.....	207
CPZ.CM.05.SIVcpzMB66D.....	207
CPZ.CM.05.SIVcpzMB897D.....	207
CPZ.GA.88.SIVcpzGAB2D.....	206
CPZ.TZ.00.TAN1D.....	196
CPZ.TZ.06.SIVcpzTAN13D.....	197
CPZ.TZ.09.UG38D.....	198
CPZ.US.85.US_MarilynD.....	210
GOR.CM.04.SIVgorCP684conD.....	211
GOR.CM.07.SIVgor2139_287D.....	211
GOR.CM.07.SIVgorCP2135conE.....	211
GOR.CM.12.SIVgor_BP1D1D.....	209
GOR.CM.12.SIVgor_BQ1D2D.....	211
GOR.CM.13.SIVgor_BP1D15D.....	211
	normal Nef end	
MAC.US.x.239	GLLMADKKETR...-	263
H2A.DE.x.BEN	GIPYSE.....-	258
H2A.PT.x.ALI	GIPFSKNRNS.....-	263
H2A.SN.86.ST_JSP4_27	GIPFS.....-	256
H2B.CI.x.EH0	GIPTE.....-	239
H2B.GH.86.0205_ALT	GIPTE.....-	241
H2G.CI.92.Abt96	GLPIE.....-	257
H2U.CI.07.07IC_TNP3	GLLMADKKETS...-	264
H2U.FR.96.12034	GLPVKNC.....-	261
ASC.UG.10.RT03-	238
COL.CM.x.CG1-	228
COL.UG.10.BWC01-	232
COL.UG.10.BWC07-	232
DEB.CM.04.SIVdeb04CMPF3061-	228
DEB.CM.99.CM40-	231
DEB.CM.99.CM5-	256
DRL.DE.11.D3-	254
DRL.DE.11.D4-	224
GRV.ET.x.GRI_677_gri_1-	219
GSN.CM.99.CN166-	218
GSN.CM.99.CN71-	208
LST.CD.88.SIVlhoest485-	207
LST.CD.88.SIVlhoest524-	208
LST.KE.x.lho7-	215
MND-1.GA.x.MNDGB1-	231
MND-2.CM.98.CM16-	233
MND-2.GA.x.M14-	251
MND-2.x.x.5440-	264
MNE.US.x.MNE027	GLLMADKKETS...-	225
MON.CM.99.L1_99CML1-	218
MUS-1.CM.01.CM1239-	218
MUS-2.CM.01.CM1246-	218
MUS-3.GA.09.09Gab0I81-	152
OLC.CI.97.97CI12-	225
RCM.CM.00.SIVagi_00CM312-	225
RCM.CM.02.SIVrcm_02CM8081-	225
RCM.GA.x.SIVRCMGAB1-	227
SAB.SN.x.SAB1-	263
SMM.CI.79.SIVsmCI2	GLLMADKKENS...-	229
SMM.SL.92.SIVsmSL92A	KKPNHKKOHGAEEGS-	267
SMM.SL.92.SL92B	POKKMADKKETS...-	263
SMM.US.96.FTq	GLLMADKKETS...-	263
SMM.US.11.SIVsmE660_FL8	GLLMADKKETS...-	262
SMM.US.86.CFU212	GL#KADKKETX...-	262
SMM.US.x.H9	GLLMADKKETS...-	209
SUN.GA.98.L14-	229
SYK.KE.x.KE51-	241
SYK.KE.x.SYK173_COMGNM-	241
TAL.CM.00.266-	241
TAL.CM.01.8023-	223
TAN.UG.x.SIVagmTAN1-	231
VER.KE.x.AGM155-	234
VER.ZA.10.VSAC4004-	231
VER.ZA.10.VSAG1003-	230
VER.ZA.10.VSAM0022-	233
WRC.CI.98.98CI04*	230
WRC.GM.05.Pbt_05GM_X02*	233