VII

PLV Proteins

Contents

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 a grain of salt 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.	*Nature* **313**(6000):277-284 (1985)
H1A1.UG.85.U455_U455A	M62320	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Oram, J.D.	*ARHR* **6**(9):1073-1078 (1990)
H1B.US.90.WEAU160_GHOSH	U21135	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Wei, X.	*Nature* **422**(6929):307-312 (2003)
H1C.ET.86.ETH2220	U46016	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Salminen, M.O.	*ARHR* **12**(14):1329-1339 (1996)
H1D.CD.84.84ZR085	U88822	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*J Virol* **72**(7):5680-5698 (1998)
H1F1.BE.93.VI850	AF077336	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T.	*Virology* **269**(1):95-104 (2000)
H1G.SE.93.SE6165_G6165	AF061642	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Carr, J.K.	*Virology* **247**(1):22-31 (1998)
H1H.CF.90.056	AF005496	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*J Virol* **72**(7):5680-5698 (1998)
H1J.SE.93.SE9280_7887	AF082394	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T.	*ARHR* **15**(3):293-297 (1999)
H1K.CM.96.96CM_MP535	AJ249239	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Triques, K.	*ARHR* **16**(2):139-151 (2000)
H101_AE.TH.90.CM240	U54771	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Carr, J.K.	*J Virol* **70**(9):5935-5943 (1996)
H102_AG.NG.x.IBNG	L39106	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Howard, T.M.	*ARHR* **10**(12):1755-1757 (1994)

Name	Accession	Proteins	Author	Reference
H103_AB.RU.97.KAL153_2	AF193276	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Liitsola, K.	*ARHR* **16**(11):1047-1053 (2000)
H104_cpx.CY.94.94CY032_3	AF049337	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*J Virol* **72**(12):10234-10241 (1998)
H1O.BE.87.ANT70	L20587	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, M.	*J Virol* **68**(3):1586-1596 (1994)
H1O.CM.91.MVP5180	L20571	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gurtler, L.G.	*J Virol* **68**(3):1581-1585 (1994)
H1O.FR.92.VAU	AF407418	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vartanian, J.P.	*J Gen Virol* **83**(Pt 4):801-805 (2002)
H1O.SN.99.99SE_MP1299	AJ302646	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Toure-Kane, C.	*ARHR* **17**(12):1211-1216 (2001)
H1O.SN.99.99SE_MP1300	AJ302647	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Toure-Kane, C.	*ARHR* **17**(12):1211-1216 (2001)
H1O.US.99.99USTWLA	AY169814	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.US.x.I_2478B	AB485668	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takekawa	Unpublished
H1N.CM.02.DJO0131	AY532635	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bodelle, P.	*ARHR* **20**(8):902-908 (2004)
H1N.CM.02.SJGddd	GQ324959	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vallari, A.	*ARHR* **26**(1):109-115 (2010)
H1N.CM.04.04CM_1015_04	DQ017382	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Yamaguchi, J.	*ARHR* **22**(1):83-92 (2006)
H1N.CM.04.04CM_1131_03	DQ017383	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Yamaguchi, J.	*ARHR* **22**(1):83-92 (2006)

Name	Accession	Proteins	Author	Reference
H1N.CM.06.U14296	GQ324962	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vallari, A.	*ARHR* **26**(1):109-115 (2010)
H1N.CM.06.U14842	GQ324958	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vallari, A.	*ARHR* **26**(1):109-115 (2010)
H1N.CM.95.YBF30	AJ006022	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Simon, F.	*Nat Med* **4**(9):1032-1037 (1998)
H1N.CM.97.YBF106	AJ271370	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Roques, P.	*AIDS* **18**(10):1371-1381 (2004)
H1P.FR.06.RBF168	GQ328744	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Plantier, J.-C.	*Nat Med* **15**(8); 871-2 (2009)
CPZ.CD.90.ANT	U42720	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, M.M.	*Virology* **221**(2):346-350 (1996)
CPZ.CM.01.SIVcpzCAM13	AY169968	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Nerrienet, E.	*J Virol* **79**(2):1312-9 (2005)
CPZ.CM.05.SIVcpzDP943	EF535993	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Van Heuverswyn, F.	*Virology* **368**(1):155-171 (2007)
CPZ.CM.05.SIVcpzEK505	DQ373065	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.CM.05.SIVcpzLB7	DQ373064	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.CM.05.SIVcpzMB66	DQ373063	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.CM.05.SIVcpzMB897	EF535994	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Van Heuverswyn, F.	*Virology* **368**(1):155-171 (2007)
CPZ.CM.05.SIVcpzMT145	DQ373066	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	*Science* **313**(5786):523-526 (2006)

Name	Accession	Proteins	Author	Reference
CPZ.GA.88.GAB1	X52154	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Huet, T.	*Nature* **345**(6273):356-359 (1990)
CPZ.GA.88.SIVcpzGAB2	AF382828	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bibollet-Ruche, F.	*ARHR* **20**(12):1377-1381 (2004)
CPZ.TZ.00.TAN1	AF447763	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Santiago, M.L.	*J Virol* **77**(3):2233-2242 (2003)
CPZ.TZ.09.UG38	JN091690	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Rudicell, R.S.	*J Virol* **85**(19); 9918-28 (2011)
CPZ.US.85.US_Marilyn	AF103818	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*Nature* **397**(6718):436-441 (1999)
MAC.US.x.239	M33262	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kestler, H.	*Science* **248**(4959):1109-1112 (1990)
H2A.DE.x.BEN	M30502	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kirchhoff, F.	*Virology* **177**(1):305-311 (1990)
H2A.GW.x.ALI	AF082339	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Reeves, J.D.	*J Virol* **73**(9); 7795-804 (1999)
H2A.SN.x.ST_HIV_2_ST	M31113	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kumar, P.	*J Virol* **64**(2):890-901 (1990)
H2B.CI.x.EHO	U27200	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Rey-Cuille, M.A.	*Virology* **202**(1):471-476 (1994)
H2B.GH.86.D205_ALT	X61240	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	*Nature* **342**(6252):948-950 (1989)

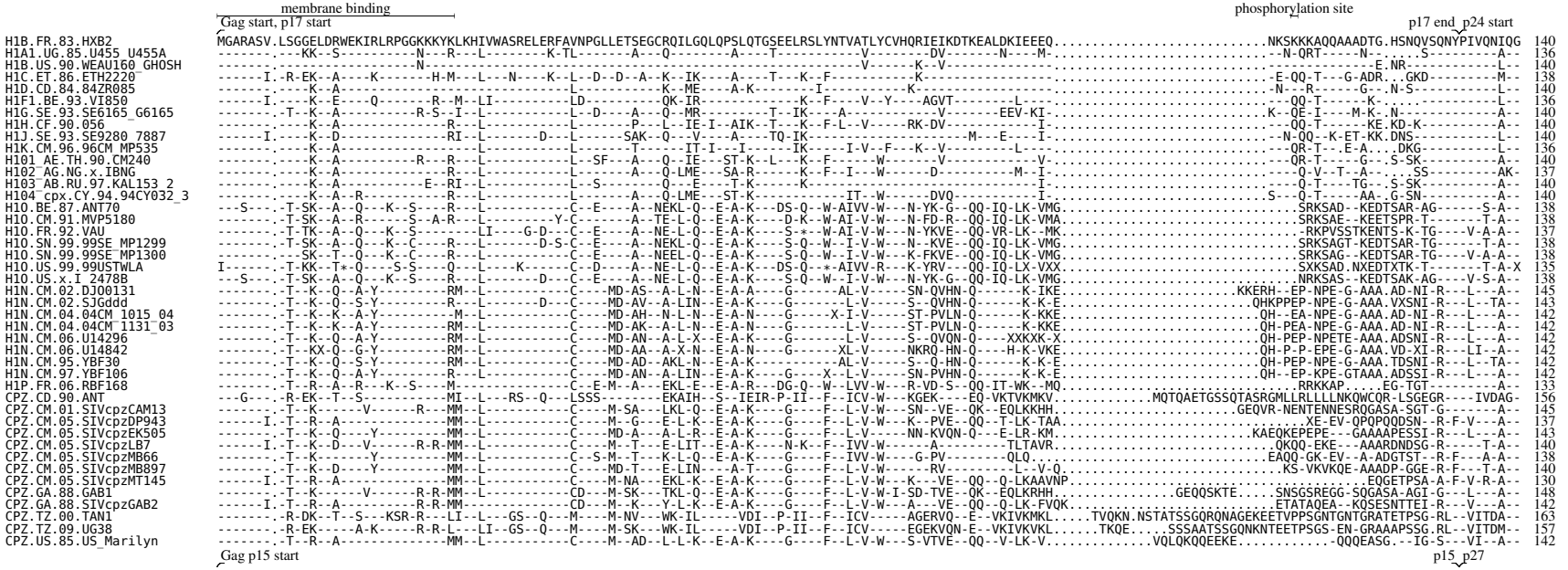
Name	Accession	Proteins	Author	Reference
H2G.CI.92.Abt96	AF208027	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Brennan, C.A.	*ARHR* **13**(5):401-404 (1997)
H2U.FR.96.12034	AY530889	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Damond, F.	*ARHR* **20**(6):666-672 (2004)
COL.CM.x.CGU1	AF301156	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	*J Virol* **75**(2):857-866 (2001)
DEB.CM.04.SIVdeb04CMPF3061	FJ919724	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Aghokeng, A.F.	*Infect Genet Evol* **10**(3); 386-96 (2010)
DEB.CM.99.CM40	AY523865	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
DEB.CM.99.CM5	AY523866	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
DEN.CD.x.CD1_CMO580407	AJ580407	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Dazza, M.C.	*J Virol* **79**(13):8560-8571 (2005)
DRL.x.x.FAO	AY159321	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	*J Virol* **77**(8):4867-4880 (2003)
GOR.CM.04.SIVgorCP684con	FJ424871	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	*J Virol* **83**(4):1635-1648 (2009)
GOR.CM.07.SIVgor2139_287	FJ424866	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	*J Virol* **83**(4):1635-1648 (2009)
GOR.CM.07.SIVgorCP2135con	FJ424863	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	*J Virol* **83**(4):1635-1648 (2009)
GRV.ET.x.GRI_677_gri_1	M66437	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Fomsgaard, A.	*Virology* **182**(1):397-402 (1991)

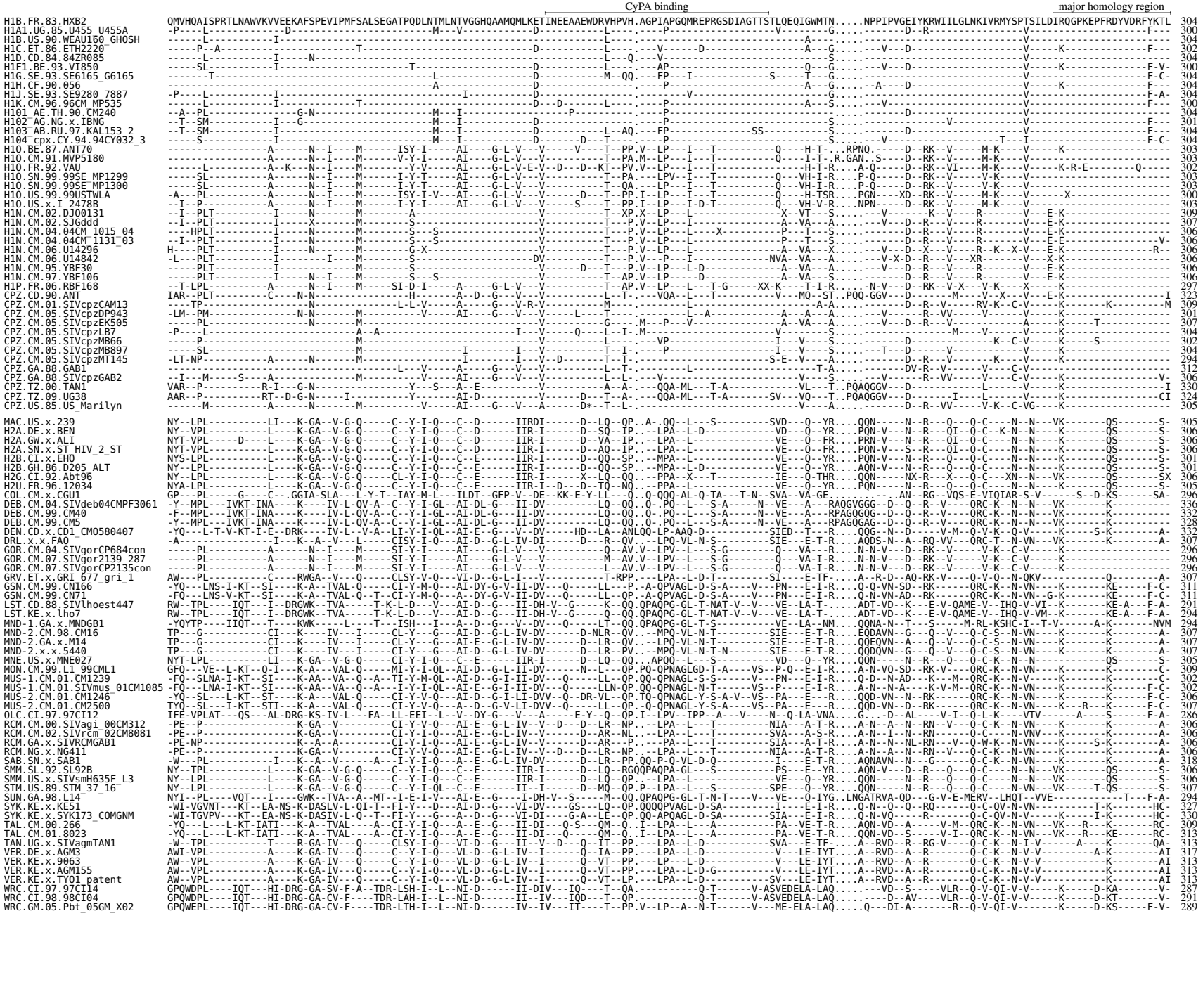
Name	Accession	Proteins	Author	Reference
GSN.CM.99.CN166	AF468659	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **76**(16):8298-8309 (2002)
GSN.CM.99.CN71	AF468658	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **76**(16):8298-8309 (2002)
LST.CD.88.SIVlhoest447	AF188114	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	*J Virol* **74**(8):3892-3898 (2000)
LST.KE.x.lho7	AF075269	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	*J Virol* **73**(2):1036-1045 (1999)
MND-1.GA.x.MNDGB1	M27470	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Tsujimoto, H.	*Nature* **341**(6242):539-541 (1989)
MND-2.CM.98.CM16	AF367411	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Takehisa, J.	*ARHR* **17**(12):1143-1154 (2001)
MND-2.GA.x.M14	AF328295	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Souquiere, S.	*J Virol* **75**(15):7086-7096 (2001)
MND-2.x.x.5440	AY159322	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	*J Virol* **77**(8):4867-4880 (2003)
MNE.US.x.MNE027	U79412	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kimata, J.T.	*J Virol* **72**(1):245-256 (1998)
MON.CM.99.L1_99CML1	AY340701	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)
MUS-1.CM.01.CM1239	EF070330	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
MUS-1.CM.01.SIVmus_01CM1085	AY340700	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)

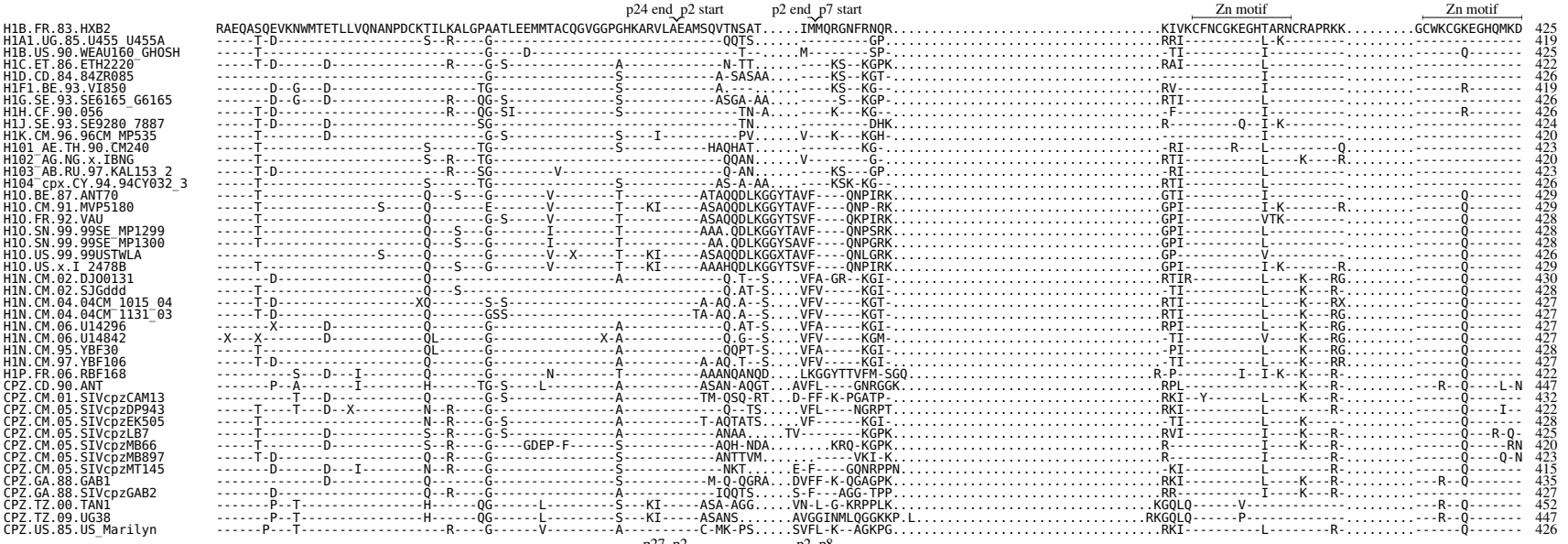
Name	Accession	Proteins	Author	Reference
MUS-2.CM.01.CM1246	EF070329	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
MUS-2.CM.01.CM2500	EF070331	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
OLC.CI.97.97CI12	FM165200	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
RCM.CM.00.SIVagi_00CM312	HM803690	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Ahuka-Mundeke, S.	*J Gen Virol* **2010** Aug 25
RCM.CM.02.SIVrcm_02CM8081	HM803689	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Ahuka-Mundeke, S.	*J Gen Virol* **2010** Aug 25
RCM.GA.x.SIVRCMGAB1	AF382829	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
RCM.NG.x.NG411	AF349680	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	*J Virol* **75**(24):12014-12027 (2001)
SAB.SN.x.SAB1	U04005	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Jin, M.J.	*EMBO J* **13**(12):2935-2947 (1994)
SMM.SL.92.SL92B	AF334679	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Chen, Z.	*J Virol* **70**(6):3617-3627 (1996)
SMM.US.x.SIVsmH635F_L3	DQ201172	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kuwata, T.	*J Virol* **80**(3):1463-75 (2006)
STM.US.89.STM_37_16	M83293	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Novembre, F.J.	*Virology* **186**(2):783-787 (1992)

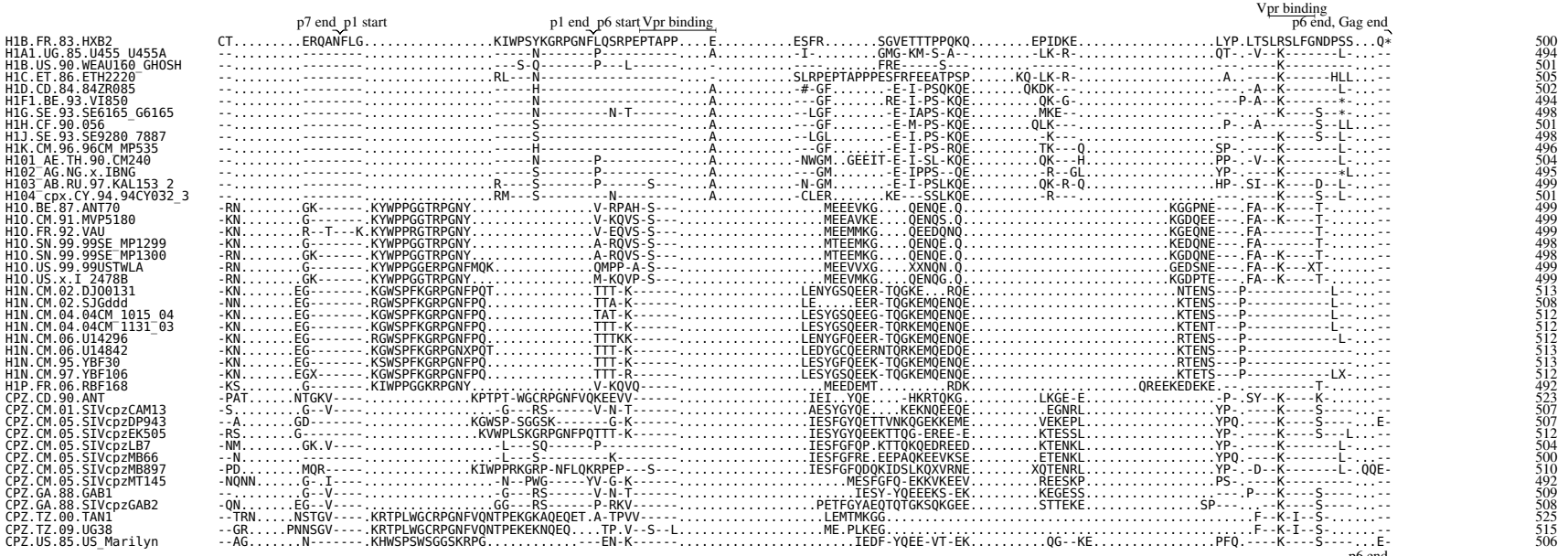
Name	Accession	Proteins	Author	Reference
SUN.GA.98.L14	AF131870	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	*J Virol* **73**(9):7734-7744 (1999)
SYK.KE.x.KE51	AY523867	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
SYK.KE.x.SYK173_COMGNM	L06042	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	*J Virol* **67**(3):1517-1528 (1993)
TAL.CM.00.266	AY655744	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*Virology* **349**(1):55-65 (2006)
TAL.CM.01.8023	AM182197	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*Virology* **349**(1):55-65 (2006)
TAN.UG.x.SIVagmTAN1	U58991	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Soares, M.A.	*Virology* **228**(2):394-399 (1997)
VER.DE.x.AGM3	M30931	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Baier, M.	*Virology* **176**(1):216-221 (1990)
VER.KE.x.9063	L40990	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	*J Virol* **69**(2):955-967 (1995)
VER.KE.x.AGM155	M29975	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Johnson, P.R.	*J Virol* **64**(3):1086-1092 (1990)
VER.KE.x.TYO1_patent	DJ048201	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Omori, T.	Patent: WO 2007049749-A 13 03-MAY-2007; DNAVEC CORPORATION
WRC.CI.97.97CI14	AM745105	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
WRC.CI.98.98CI04	AM713177	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
WRC.GM.05.Pbt_05GM_X02	AM937062	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Locatelli, S.	*Virology* **376**(1):90-100 (2008)

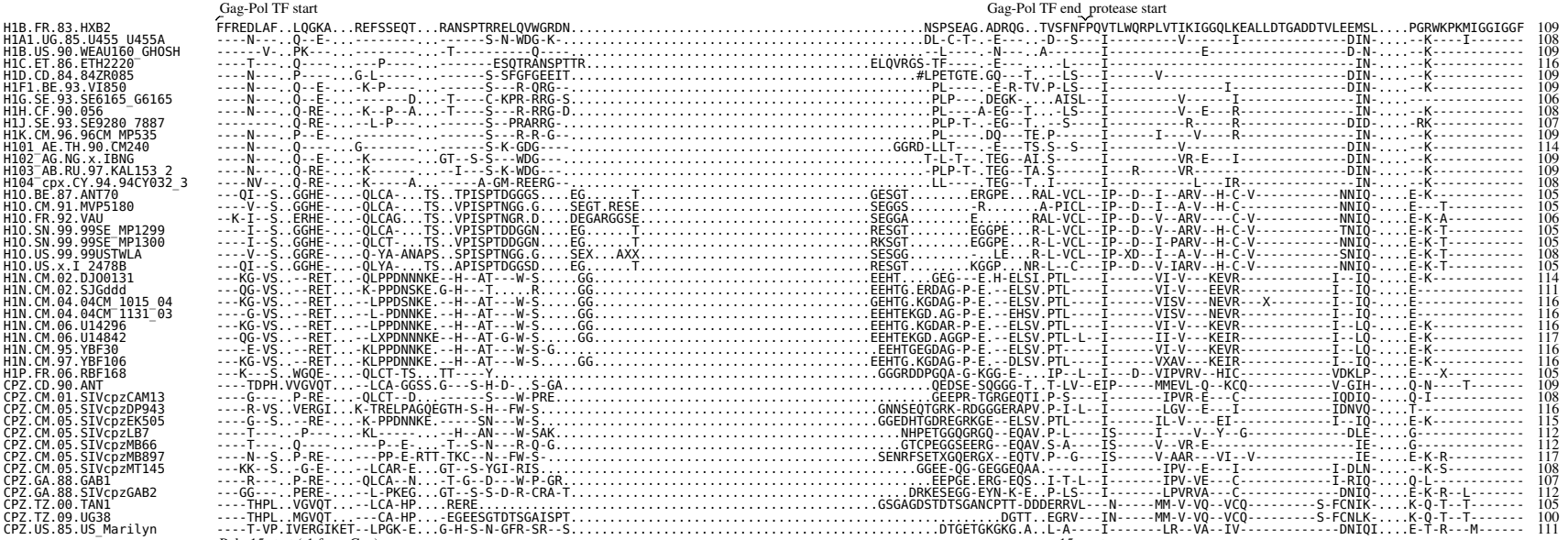
Name	Accession	Proteins	Author	Reference
MAC.US.x.251_1A11	M76764	Vpx	Marthas, M.L.	*J Med Primatol* **18**(3-4):311-9 (1989)
MAC.US.x.251_BK28	M19499	Vpx	Hirsch, V.	*Cell* **49**(3):307-319 (1987)
MAC.US.x.EMBL_3	Y00295	Vpx	Franchini, G.	*Nature* **328**(6130):539-543 (1987)
SMM.US.x.H9	M80194	Vpx	Courgnaud, V.	*J Virol* **66**(1):414-419 (1992)
SMM.US.x.PGM53	AF077017	Vpx	Novembre, F.J.	*J Virol* **72**(11):8841-8851 (1998)
H1O.CM.94.BCF06	AB485666	Vpu	Takekawa	Unpublished
H1O.CM.96.96CMA102	AY169803	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.96.96CMABB009	AY169806	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.96.96CMABB637	AY169810	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.97.97CMABB447	AY169813	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.97.97CMABB497	AY169809	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.98.98CMA104	AY169802	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.98.98CMA105	AY169816	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.98.98CMABB141	AY169807	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.98.98CMABB197	AY169811	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.98.98CMABB212	AY169804	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.98.98CMU2901	AY169812	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.98.98CMU5337	AY169808	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.99.99CMU4122	AY169815	Vpu	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.x.pCMO2_3	AY618998	Vpu	Tebit, D.M.	*Virology* **326**(2):329-339 (2004)
H1O.CM.x.pCMO2_5	AY623602	Vpu	Tebit, D.M.	*Virology* **326**(2):329-339 (2004)
CPZ.CM.98.CAM3	AF115393	Vpu	Corbet, S.	*J Virol* **74**(1):529-534 (2000)
CPZ.CM.98.CAM5	AJ271369	Vpu	Muller-Trutwin, M.C.	*J Med Primatol* **29**(3-4); 166-72 (2000)
CPZ.TZ.01.TAN2	EF394357	Vpu	Takehisa, J.	*J Virol* **81**(14):7463-7475 (2007)
CPZ.TZ.02.TAN3_1	DQ374658	Vpu	Takehisa, J.	*J Virol* **81**(14):7463-7475 (2007)
CPZ.TZ.06.TAN5	JN091691	Vpu	Rudicell, R.S.	*J Virol* **85**(19); 9918-28 (2011)

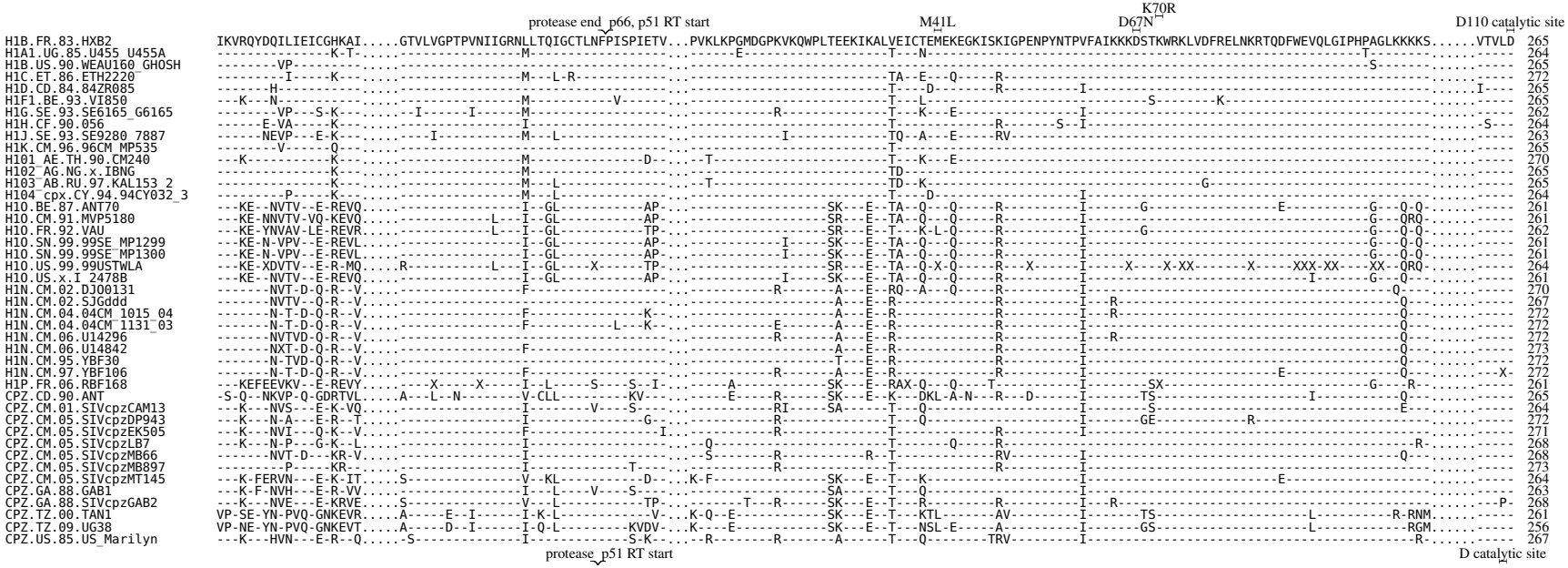

membrane binding
Gag start, p17 start
phosphorylation site
p17 end_p24 start
Gag p15 start
p15_p27

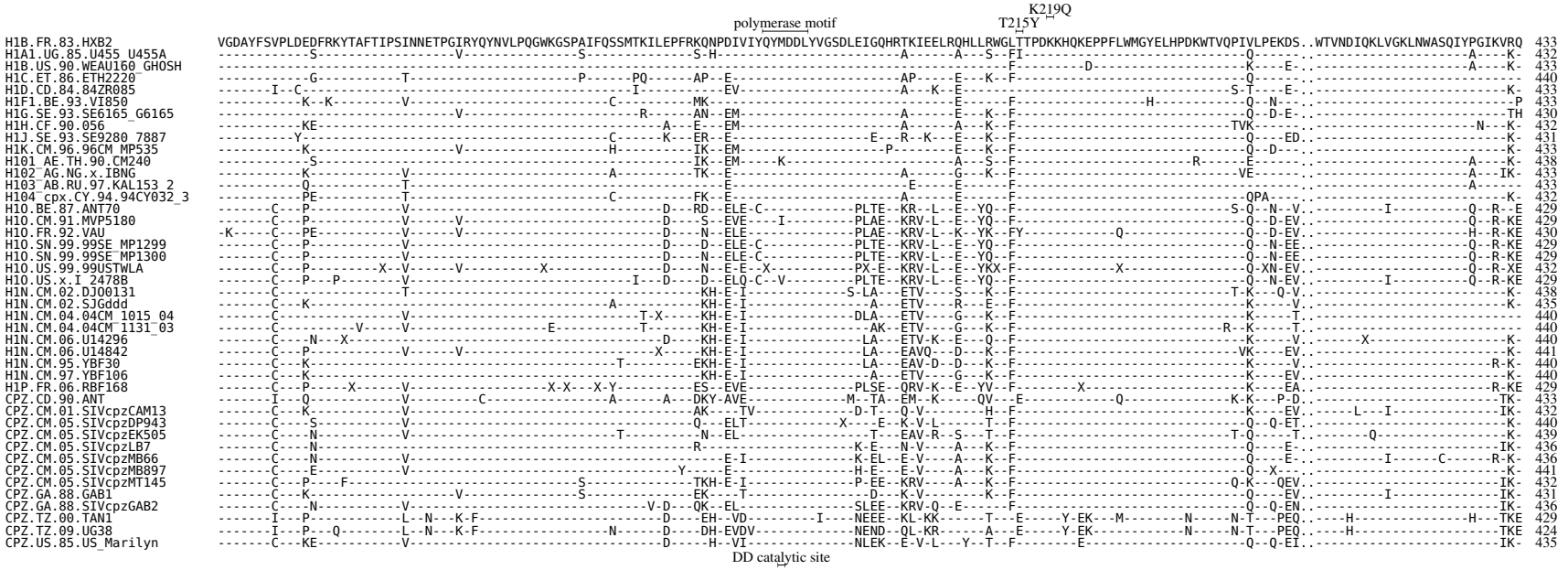

CyPA binding
major homology region

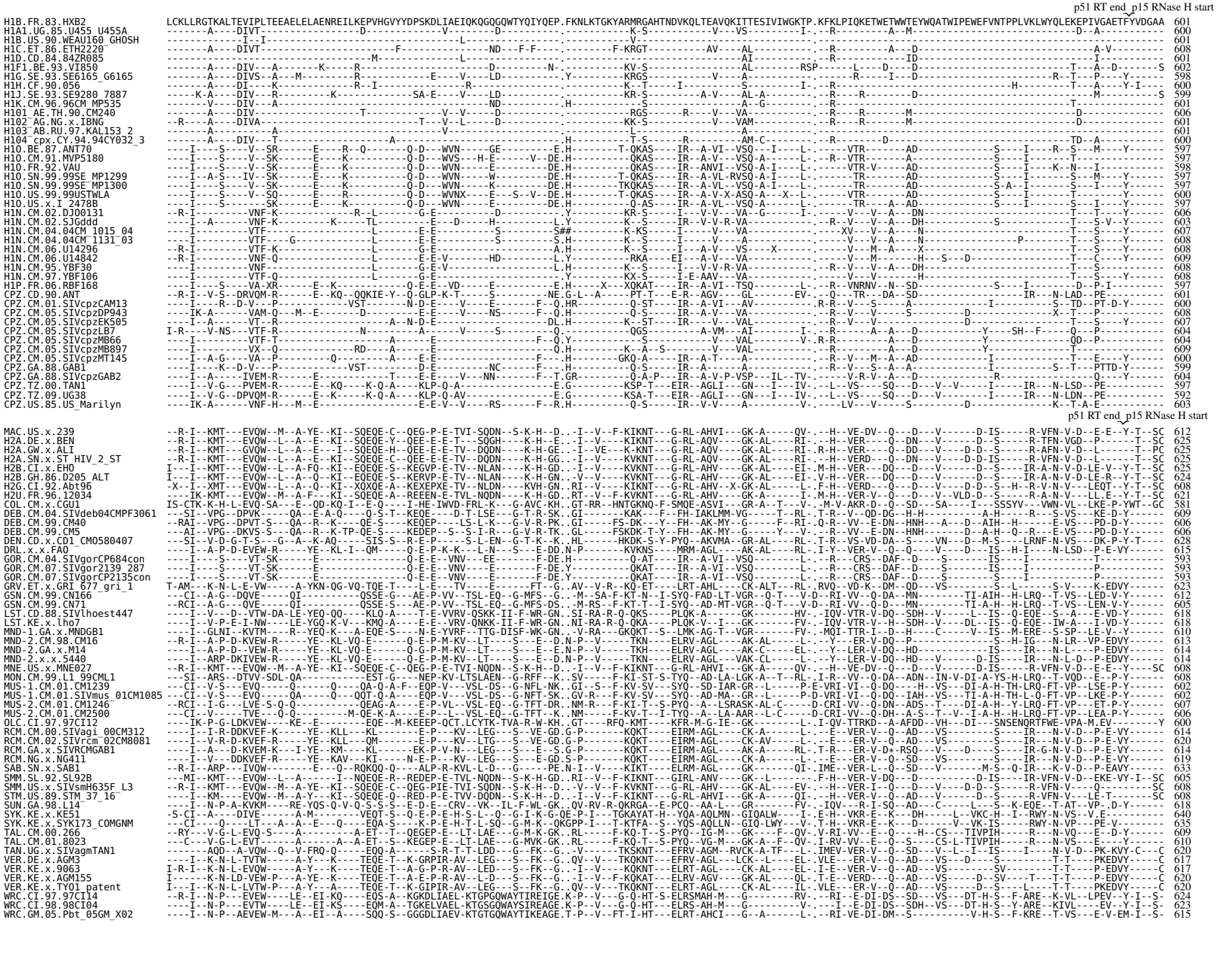

p24 end p2 start
p2 end p7 start
Zn motif
Zn motif
p27 p2
p2 p8

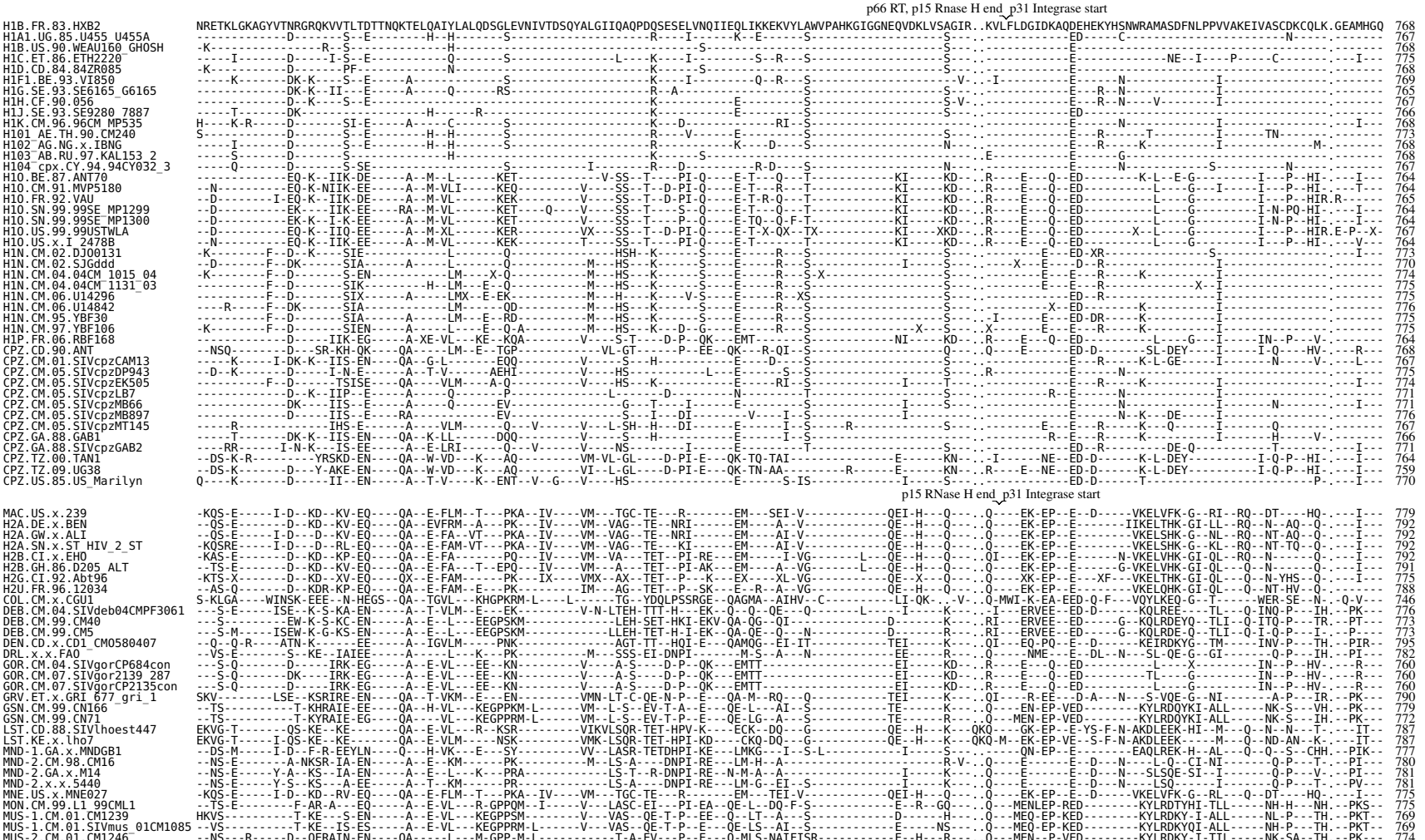

p7 end p1 start
p1 end p6 start Vpr binding
Vpr binding
p6 end, Gag end
p8 p1
p1 p6
PTAP motif
PSAP in HIV-2 B, U
p6 end
Gag end


Gag-Pol TF start
Gag-Pol TF end_protease start
H1B.FR.83.HXB2
H1A1.UG.85.U455_U455A
H1B.US.90.WEAU160_GHOSH
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165_G6165
H1H.CF.90.056
H1J.SE.93.SE9280_7887
H1K.CM.96.96CM_MP535
H101_AE.TH.90.CM240
H102_AG.NG.x.IBNG
H103_AB.RU.97.KAL153_2
H104_cpx.CY.94.94CY032_3
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
H1O.FR.92.VAU
H1O.SN.99.99SE_MP1299
H1O.SN.99.99SE_MP1300
H1O.US.99.99USTWLA
H1O.US.x.I_2478B
H1N.CM.02.DJO0131
H1N.CM.02.SJGddd
H1N.CM.04.04CM_1015_04
H1N.CM.04.04CM_1131_03
H1N.CM.06.U14296
H1N.CM.06.U14842
H1N.CM.95.YBF30
H1N.CM.97.YBF106
H1P.FR.06.RBF168
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.CM.05.SIVcpzMT145
CPZ.GA.88.GAB1
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.00.TAN1
CPZ.TZ.09.UG38
CPZ.US.85.US_Marilyn
Pol p15 start (-1 from Gag)
p15_protease
MAC.US.x.239
H2A.DE.x.BEN
H2A.GW.x.ALI
H2A.SN.x.ST_HIV_2_ST
H2B.CI.x.EHO
H2B.GH.86.D205_ALT
H2G.CI.92.Abt96
H2U.FR.96.12034
COL.CM.x.CGU1
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM5
DEN.CD.x.CD1_CMO580407
DRL.x.x.FAO
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139_287
GOR.CM.07.SIVgorCP2135con
GRV.ET.x.GRI_677_gri_1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest447
LST.KE.x.lho7
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-1.CM.01.SIVmus_01CM1085
MUS-2.CM.01.CM1246
MUS-2.CM.01.CM2500
OLC.CI.97.97CI12
RCM.CM.00.SIVagi_00CM312
RCM.CM.02.SIVrcm_02CM8081
RCM.GA.x.SIVRCMGAB1
RCM.NG.x.NG411
SAB.SN.x.SAB1
SMM.SL.92.SL92B
SMM.US.x.SIVsmH635F_L3
STM.US.89.STM_37_16
SUN.GA.98.L14
SYK.KE.x.KE51
SYK.KE.x.SYK173_COMGNM
TAL.CM.00.266
TAL.CM.01.8023
TAN.UG.x.SIVagmTAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TYO1_patent
WRC.CI.97.97CI14
WRC.CI.98.98CI04
WRC.GM.05.Pbt_05GM_X02


protease end_p66, p51 RT start
M41L
K70R
D67N
D110 catalytic site
protease_p51 RT start
D catalytic site


K219Q
T215Y
polymerase motif
DD catalytic site


p51 RT end_p15 RNase H start
p51 RT end_p15 RNase H start


p66 RT, p15 Rnase H end_p31 Integrase start
p15 RNase H end_p31 Integrase start

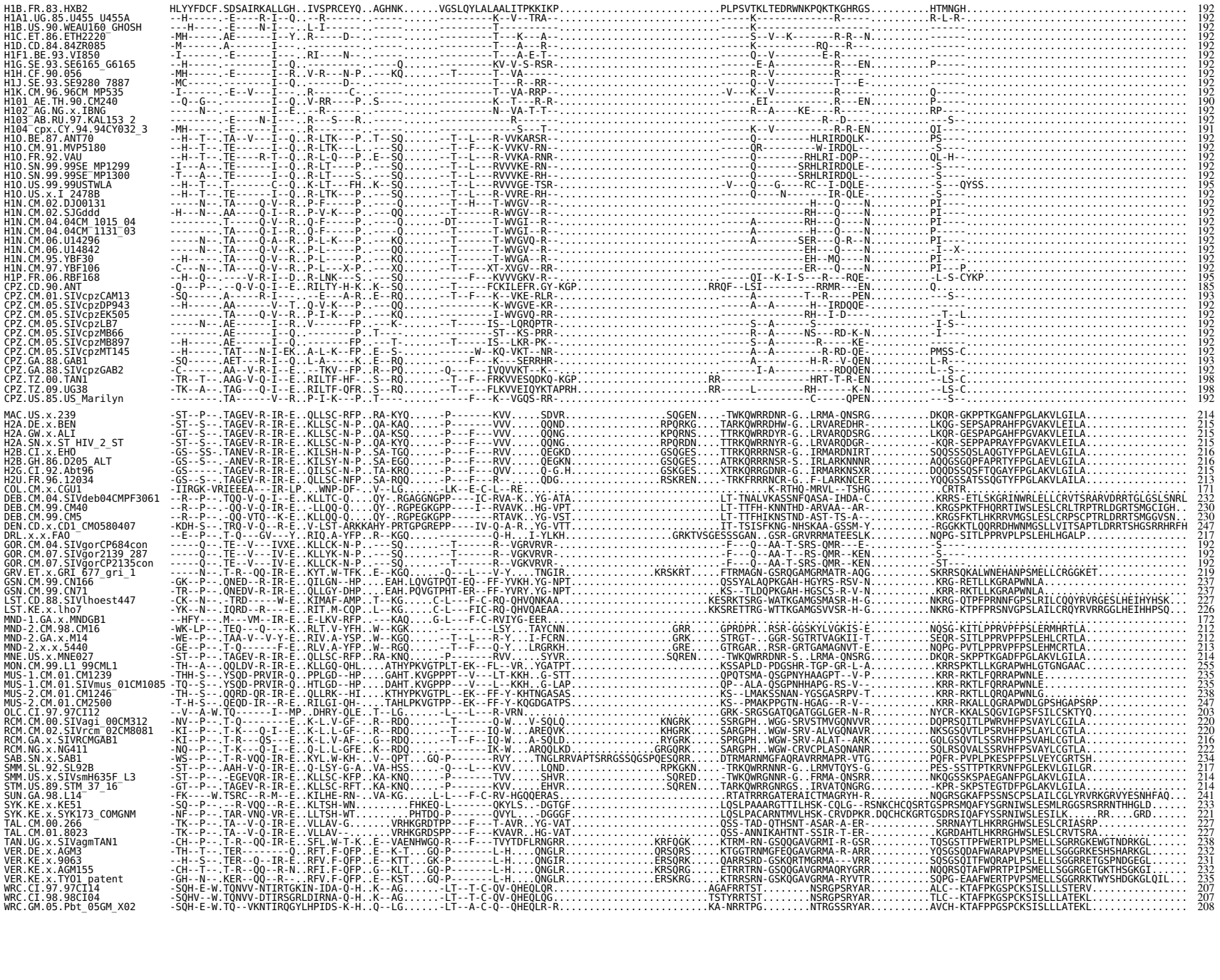
```
H1B.FR.83.HXB2              VDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTIHTDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGER.IVDIIATDIQTKELQKQI.TKIQ 936
H1A1.UG.85.U455_U455A       -----------------------------------------------I----------V----------S-A-K-V----N-Q------------------------------E--------------------------------.-I----------------.S--- 935
H1B.US.90.WEAU160_GHOSH     ------------------I----------------------------I---------------------ST--K-----------------------I------------------------------------------------.-I---------QQ-----.---- 936
H1C.ET.86.ETH2220           -N----------------I--------------------------------------RV----------SNA-K--------Q------------------------------E--------------------R-----------.-I----S--------N--.L--- 943
H1D.CD.84.84ZR085           -------------------------------------------A--------------VV---------S---K---------------------------------------E--------------------------------.-I----S----R------.---- 936
H1F1.BE.93.VI850            ----------------------------E------------------I----------I----------S-A-K-S------Q-----------------I---------------------------------------------.-I---S-----R------.---- 937
H1G.SE.93.SE6165_G6165      ------------------I-I--------------------------I----------TV---------S-A-K------N-T---------------------------------------------------------------.-I----S-----------.---- 933
H1H.CF.90.056               -----------------Q-----------------------K----------S-----V----------S-A-K------D-Q---------------------------------------------------------------.-I----------------.SN-- 935
H1J.SE.93.SE9280_7887       -------------------------------------------A-F-I----------V----------SGA-K------D--------------------------------E--------------------------------.-I---------R------.---- 934
H1K.CM.96.96CM_MP535        ------------------I----------------------------I----------V------T---STV-K-------V----------------------------------------------------------------.------------------.LN-- 936
H101_AE.TH.90.CM240         ----------------------------------------------------------V----------S-A-K------NVQ------------------------------E--------------------------------.-I----------------.---- 941
H102_AG.NG.x.IBNG           ---G--------------I----------------------------I----------V----------S-A-K------NVT---------------------------------------------------------------.-I----S-----------.I--- 936
H103_AB.RU.97.KAL153_2      ------------------I----------------------------V----------I----------STA-K-----------------------------Q--QT--------------------------------------.-I----------------.I--- 936
H104_cpx.CY.94.94CY032_3    --------------------M--------------------------I----------M--A---P---S-A-K------D-N---------------------------------------------------------------.-I----S-----------.---- 935
H1O.BE.87.ANT70             -----EV--I----M---I-I--------F----------------------A-----V------P---ST-MK------N-QH---------------A------S--Q---------R--------V------------T----.-I--L-SQ---T------.L-XH 932
H1O.CM.91.MVP5180           --Y--E---M--------I-I-------DF----------------------A-----V------P---S-AMK-----T--QH---------------A------S--Q------------------V------------T----.LI--L-SQ---T------.L--N 932
H1O.FR.92.VAU               -----E---M--------I-I--------F----------------------A--I-IL------P---S-AMK------N-QH---------------------ES-MQ------------------V------------T----.LI--L-SQ---T------.L--- 933
H1O.SN.99.99SE_MP1299       -----EV--M--------I-I--------F----------------------A-----V------P---S-AMK------N--H---------------A------S--Q------------------V--Y---------T----.-I--L-SQ---T------.F--- 932
H1O.SN.99.99SE_MP1300       --Y--EV--I--------I-I--------F----------------------A-----I------P---S--MK-----T--H----------------A------S--Q------------------V--Y---------T----.-I--L-SQ---T------.F--- 932
H1O.US.99.99USTWLA          -----E---M--------I-I--------F----------------------A-----I-----RP---S--MK-----T--QH--X---------X--A------S--Q------------------V------------T----.LI--L-SQ---T------.L--- 935
H1O.US.x.I_2478B            -----EV--I----V---I----------F----------------------A-----I------P---S--MK-----TN-QH---------------A------S--Q---------R--------V------------T----.-I--L-SH---T------.L--- 932
H1N.CM.02.DJO0131           -N----V-----------I-----------L----------------I----------V------P---S---K------N-T----------------------------I---------------------------E-T----.-I---------TK--T--.L--- 941
H1N.CM.02.SJGddd            -N----V-----------I-----------L----------------I----------V------P---S---K------N-Q-------------A----------------------R---------------------T----.-I---------TK--T--.LQV- 938
H1N.CM.04.04CM_1015_04      -N----V-----------I-----------L----------X-----I----------V------P---S---K------S---------------A--------------I-----------------------------T----.-I---------TK--T--.L-V- 942
H1N.CM.04.04CM_1131_03      -N----V-----------I-----------L----------------I----------V------P---S---K------S---------------A--------------I-----------------------------T----.-I---------TK--T--.L-V- 943
H1N.CM.06.U14296            -N----V-----------I-----------L------------A---I----------V------X---S---K------N---------------X--------------I-----------------------------T----.-I-----------N-T--.L-V- 943
H1N.CM.06.U14842            -N----V-----------I---X-------L----------------I----A-----V----------S---K----------------------A--------------I-----------------------------T----.-I-------------T--.L-V- 944
H1N.CM.95.YBF30             -N----V-----------I-----------L----------------I----------V----------S---K------N---------------A--------------I-----------------------------T----.-I---------TN--T--.L-V- 943
H1N.CM.97.YBF106            IN----V-----------I-----------L----------------I----------V------P--IS---K----------------------A--------------I----------------------------XT----.-I---------TK--T--.L-V- 943
H1P.FR.06.RBF168            --------------T---I----------FL----------------V----A-----V------P---S-A-K-----LN-TH-------------------------L------------------V------------T--D-.-I--M--Q---T------.L--- 932
CPZ.CD.90.ANT               ---------V----------I-----S--F-----MAD---KS---------S------------A---S-A-K------N-Q--------------------Q--Q--I---Q-----V---H-----------------TP-Q-.-L--L------TQ--N--.L--- 936
CPZ.CM.01.SIVcpzCAM13       ------T--M--------I-I----I---------------------I----S-----VL-----T---SNA-K------N---K------------------------E------------------S------------T----.-I----------------.L--- 935
CPZ.CM.05.SIVcpzDP943       ------M-----------------------L----------------I----------V----------SN--K--------Q----------------------------I-E---------------------------T----.-I---------TN-----.L-V- 943
CPZ.CM.05.SIVcpzEK505       IN----V-----------I-----------L----------------I----------V----------S---K------N-Q-------------A--------------I-----------------------------T----.-I---------TK--T--.L-V- 942
CPZ.CM.05.SIVcpzLB7         ------------------------------M----------------I----S-----I----------STA-K------Q-Q---------------------------------------------------------------.-I---------T------.S-V- 939
CPZ.CM.05.SIVcpzMB66        ------------------I-----------M----------------I----------V----------SS--K--------Q----------------------------I----------------------------------.-I--L------TK--Q--.SN-- 939
CPZ.CM.05.SIVcpzMB897       ------------------I----------------------------V----------V----------S-A-K------N-Q---------------------------------------------------------------.-I---------R------.S-V- 944
CPZ.CM.05.SIVcpzMT145       ---------I--------I-I--------------------------V-----------------P---SNA-K--------R--------------------------E------------------------------------.-I----------------.L--- 935
CPZ.GA.88.GAB1              ---------V--------I----------------------------------------------P---S-A-K------D-------------------L----------------------------------------T----.-I---------S------.L-V- 934
CPZ.GA.88.SIVcpzGAB2        ------------------I-----------L----------------I----------K------T---S-A-K-------------------------------------------------------------------T----.---ML---L--Q------.L--- 939
CPZ.TZ.00.TAN1              --Y--E---I--------I----------F-------E---R-----I----------K------P---STA-K------Q-QH-------------------Q--Q--E-------Q-R---I---Y-----------E-T----.LL--LT-N---Q------.L-V- 932
CPZ.TZ.09.UG38              -----EV-----------I-I--------F-------E---K-----I----------K------P---STA-K------Q-QH-------------------Q--Q--Q---E---Q-----V---Y-----------E-T----.LL-LLT-N---Q------.L--- 927
CPZ.US.85.US_Marilyn        ------------------I-I---------L----------------I----------V----------SS-----------Q----------------------------I-----------------------------T----.-I----SEL--DL-----.L-V- 938

MAC.US.x.239                ANSDL-T--M--------I-I--------F-------Q---RQ--L----------ITHL-----A--ASQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQFQ-SKN.S-FK 947
H2A.DE.x.BEN                -NAEI-V--M-Y------I-I--------F-------Q-S-RQ--L------S---ITHL-----P---SQE-KMVA--V--E-S--V-----------A--HH--NQ-SRI-E--NTIE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQF--RKN.SNFK 960
H2A.GW.x.ALI                -NAEL-T--M--------II---------F-------Q-S-RQ--L------S---ITHL---S-V---SQE-KMVA--V--E-S--V-----------A--HH--NQ-SRI-E--NTVE-I-L---HCM----R----DMTPA--.LINM-S-EQEIQF--TKN.L-FK 960
H2A.SN.x.ST_HIV_2_ST        -NAEL-T--M--------I-I--------F-------Q-S-RQ--L------S---ITHL-----A---SQE-KMVA--I--E-S--V-----------A--HH--NQ-SRI-E--NTVE-I-L---HCM----R----DMTPA--.LINMVTAEQEIQF--AKN.S-L- 960
H2B.CI.x.EHO                -NSEL-T--M----------I--------F-------Q---RQ--L------S---ITHL-----A---SQD-KM-A--I--E-T--V----E------A--HH--NQ-DRI----VSIE-V-L--THCM----R----DMTPA--.--NM-T-EQEIQF--TKN.L-F- 960
H2B.GH.86.D205_ALT          -NADL-T--M--------I-I--------F-------Q---RQ--L------S---ITHL-----A---SPS-KMVA--V--E-T--V-----------A--HH--NQ-DRL----VSIE-V-L--THCM----R----DMTPA--.L-NM-T-EQEIQFF-AKN.L-F- 959
H2G.CI.92.Abt96             -NAEL-T--M----------I--------FV------Q---RQ--L-----------ITHL-----A---SQE-KMVA--V--E-A--V--------------HH---Q-DKI----NSIE-I-L---HCM----R----DMTPA--.-INM-T-EQEIQFQ-TKN.S-FK 943
H2U.FR.96.12034             -NTEV-T--M---------VI--------F-------Q---RQ--L-----------ITHL-----A---SQE-KMVA--V--EHT--V-----------A--HH--NQ-ERI-E--NTME-I-L--AHCM----R----DMTPA--.-INM-S-EQEIQF--SKN.S-FK 956
COL.CM.x.CGU1               L-Y-Y-L-------E-----------CTLFCW-TILKR---E--GRA-I---SQ-E-RQV-----P--VSQHFK--V--L--AHTT-H-----------QR--DV-RK-KKMK----T-ESK-A---YAL-------L--K-PW--.Q-ERAIIELD-QN-T-LQNQ-FK 915
DEB.CM.04.SIVdeb04CMPF3061  -NADL----M----------I---------FT-TQ-LKE---KQ--L---Q-GA---IQQV-----P--ISQAFA-----L--EHTT-V-----------NK-RQ--DT-T---ED-QR-E---A--TH-L----R----DMTPS--.--NM-N-ELEI-Y--QK-.S-FS 944
DEB.CM.99.CM40              --A-M----M------N-I-I------------TKIL-----K---L---Q-GA---TQ------P--ISQP-K-----L--EHTT-V-------------K-RI--DT-K-I--D--R-E---A--TY-I----R----DM-PI--.L-NM-T-E-E--T--QK-.S-LL 941
DEB.CM.99.CM5               --A-M-V--I------N---I---------FM-TK-LS-----K---L---Q-GA---I-Q------P--ISQP-K-----L--DHTT-V-------------K-RV--ET-K-I-ED--R-E---A--T--I----R----DMA-I--.L-NM-T--LE-QT--QK-.S-LL 941
DEN.CD.x.CD1_CMO580407      --A-L-T--M------N---I---NA----L-TKIL-----K---L---QI-A---ITKL-----P--ISQAMQ------N--HTT-V---------I--NK--Q--EV-AL--ED--R-E--LA--TL-L----R--L-DMT-A--Y-NMLNTHLEIQH..T--..P-FS 961
DRL.x.x.FAO                 --A---T--M--------I-I--------H------------EK-------------SHL-----P---SEK-ATV----K-EHTT-V------------NQ--------I-----K-E-------L-------------E---A--.------S-LL-SK--QN-.L--- 950
GOR.CM.04.SIVgorCP684con    ------------------I----------F-----------------------A-----V-----P---S---K-----LNVTH-------------------------Q-------------------V-----------T--D-.-I-ML---Q---T-------.L--- 928
GOR.CM.07.SIVgor2139_287    ------------------I----------F-----------------------A-----I-----P---S---K-----LNVTH-------------------------Q-------------------V-----------T--D-.-I-ML---Q---T-------.L--- 928
GOR.CM.07.SIVgorCP2135con   ------------------I----------F-----------------------A-----I-----P---S---K-----LNVTH-------------------------Q-------------------V-----------T--D-.-I-ML---Q---T-------.L--- 928
GRV.ET.x.GRI_677_gri_1      --A-IET--M----------I---------F--------R---K---H-----LA-----HL-----P---SQN-A-V---GN-EHTT-----------S-----RQ--E--S-I--DC-R-E-------TH--------------I-SA--.L-NMLT-QLELNT--N--.Q--L 958
GSN.CM.99.CN166             -NAEL-M--M---------------------VW-RI--Q---RQ--LK--E--AT---THL-----P--VSKELE------D-QHTT-V-----------N---Q--ET-TKI-EEVTY-E---AQ-CY------------DMCPT--.--NM-H-ELE-QH-NT-N.S-F- 947
GSN.CM.99.CN71              -NAEL----M---------------------VW-RI--Q---RQ--LK--E--AT---THL-----P--ISKELE------N-QHST-V-----------N---Q--ET-QKI--EVTY-E---AQ-CY------------DMCPT--.L-NM-H-ELE-QH-NT-S.S-F- 940
LST.CD.88.SIVlhoest447      --T------I----M--Q--IN-------FMV------D---KT--H------CS----QQV-----P--VSKE-Q-VI--L--EHTT---------------AK--A--Q--ER--ED-QQ------L--TH-----QR--L--L-PA--.FINM-NA-LE-QY---LN.S--L 955
LST.KE.x.lho7               --V-V-------------Q--IN-------FMV------D---KT-SH------CS-----Q------P--VSKE-Q-VT--I--EHTT---------------AK--V--S--ER--ED-QQ------L-----H-----QR--L--LTPA--.FINM-NAELE-QY---LN.S--L 955
MND-1.GA.x.MNDGB1           T-A-L-V--I------NQI-I--------FMK-----T----KK--E-------AQ--ISKL-----P---SQE-ETM---L--EHT---------------NK--Y--EL-EKI-EDCKE-----A--T-------QR--L--MT-----.--NM-N-ELEYQYQ-N--.S-NL 945
MND-2.CM.98.CM16            --T---T--M----M------IA------R-L-------T---K---H---------------HL-----P---SEK-ATV-----Q-EHTT---------------I-AK-HH--Q---------------K-E-------L-------------E--P---.---------LL-TK--HN-.Q--- 948
MND-2.GA.x.M14              --AA--T--M----------I--------L------------K---H---------------HL-----P--VSEK-ATV-----Q-EHTT---------------AK-HH--V--E---------R-E-------L-------------E--P---.------T---L-TK--QN-.S--- 949
MND-2.x.x.5440              -NAD--T--M--------I-I--------L------------K---H---------------HL-----P--VSEK-ATV-----Q-EHTT-V-------------AK-HH-----E---------R-E-------L-------------E--P---.------T---L-TK-RQN-.S--- 949
MNE.US.x.MNE027             -NSDL-T--M--------I-I--------F-------Q---RQ--L------S---ITHL-----A---SQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSME-I-L---HCM----R----DMTPA--.LLNM-T-EQEIQFQ-SKN.S-FK 943
MON.CM.99.L1_99CML1         -NAEV-V--M--------SI-----------VW-KIL-R---KC-GIA--E--AM---TQ------P---SQEFE--A---N-HHTT-V-----------N--RQ--ET-KKI--EV-Y-P--LAQ-L--L----------DMAPVD-.FINM-H-ELELQTSNN--.--FS 943
MUS-1.CM.01.CM1239          -NAEL-V--M----------I------------TW-KIL-R---RQ--LG--E--AL---TQ------A--ISEEFG------S-DHTT-V-----------NK--Q--ET-QKI-EEVTY-E---AQ----M---K------DLA-A--.INMLH-ELELQH--S-K.S-F- 937
MUS-1.CM.01.SIVmus_01CM1085 -NAEI-V--M----------I------------MW-KIL-R---RQ--LG--E--SL---ATQ-----A--IS-EFE------S-EHTT-V-----------NK--Q--EA-HKI-EEVTY-E---AQ----L---K---V-DL-NA--.INMLH-ELELQH--Q-K.Y-F- 937
MUS-2.CM.01.CM1246          -NADL-V--M---------------------SF-W-RI-SQ---RL--LE-SN--AT--ISQ--P---T---SKEFQ-VA---N-QHTT-V-----------NA--Q--ET-HK--EEVTY-E---AQ----L-N--------DMTPT--.L-NMLY-ELEIQQ--NHN.--FS 942
MUS-2.CM.01.CM2500          -NADI-V--M----------I-----------F-W-RI-SQ-N-RL--LE--N--AT---ITQ------T---SKEFQ-VA-----EHST-V-----------NA--Q--ET-NK--EEVQY-E---AQ----L-N--------DMTPT--.L-NMLY-E-E-QQI-NTK..FS 939
OLC.CI.97.97CI12            -AVGVYR--I-----------CL-A-TG--FLYG-ILR--D-EN-VK-V-R--SQ-----I------T--VNQKMQ-V---L--EH-T---W------II-AK-RV--E-L-RI-E--T--E--LA--I------N---Q--GTVA-E.L-NQ-Y--T.QI-YTQNK.F-KF 936
RCM.CM.00.SIVagi_00CM312    T-A---T--M----------I--------------------K----------A-----HL-----A---S-A-Q-V----Q-EHT--V---------------HQ--T----I-----K-E-------L------------------.-I----SEL--NK--N--.S--- 949
RCM.CM.02.SIVrcm_02CM8081   --A--KT--M----------I----------------L-----K---H------A-----HL-----A---SSA-Q-V----Q-EHT--V---------------HQ--T--T-I-----KIE-------L------------------.-I----S-L--TK--N--.S--- 955
RCM.GA.x.SIVRCMGAB1         --A---T--M--------I-I--------------------K---H------A----RKL-----A---S-A-Q-V----Q-EHA--V----------------Q--I--E---E---K-E-------LV-----------------.-I------LA-NK--N--.S--- 949
RCM.NG.x.NG411              --A---T--M--------I-I--------------------K---H------A-----HL-----T---S-A-Q-D----Q-EHT--V---------------HQ--I----I-----K-E-------L------------------.-I----S-L--TK--N--.S--- 954
SAB.SN.x.SAB1               --A---V--M--------I-I--------F-------Q---KA--H------S---ITQL-----T---SQQ-A-I---GK-EHT--V-------------------Q--E----I--D--R-E---I---H-----------------A--.LIN--H-ELE--T--QK-.S--- 968
SMM.SL.92.SL92B             TNAEV-T--M----------I--------F-------R---RQ--L----I-S---I-HL-----A---SQE-KMVA--L-VE-S--V------------A-DLH---N-DKI-E---SVE-L-L--AHCM----R----DMTPA--.--NM-T-ELE-QY-NS-N.S-F- 940
SMM.US.x.SIVsmH635F_L3      -NAEL-T--M--------I-I--------F-------Q---RQ--L-----------ITHL-----A---SQE-KMVA-----E-T--V-----------A--HH--TQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.L-NM-T-EQEIQFQ-SKN.S-FK 943
STM.US.89.STM_37_16         -NAEL-T--M--------I-I--------F-------Q---RQ--L------S---THL-----A---SQE-KMVA-----E-T--V-----------A--HH--TQ-DRI----NTVE-V-L---HCM----KR---L-DMTPA--.L-NM-T-EQEIQFQ-SKN.S-FK 943
SUN.GA.98.L14               --A-V----I----M-E---I---------MV---L-N-Q-KT--TW----CAM----Q------P--ISKD-E-V---L--QHTT-------------AK--V--Q--SRI-ED-QE------L--LH-----QR--L--MT-A--.FINM-NA-LE-QY---IN.S--L 955
SYK.KE.x.KE51               --TDL-T--M-----------C---NT----T-TKILKR-------GL---QI-A---I-H------P--ISDAFA--------VEHTT------------I--NK-RQ--ET-N-I-EEV-R-E---A--T--L----------TTPAD-.YINMLY-EL-LQNTHT...Q-FS 973
SYK.KE.x.SYK173_COMGNM      --VDIYN--M----E-----C---NT------TKILKR---D---L--MQI-S---I-Q------P--VSDKFK-------C--EHTT-------------I---K-RY--EA-S-I--DVT--Q---A--T--L------------I-P---.YINMLY-EL-LQQNTTS..P-FS 969
TAL.CM.00.266               NNAAV-T--M-------QI-C------------TKIL-R---R---L---QV-S---ISHL-----P--VS-EMQ-MV--LK-EHSA-V----------A--NK--Q----T-T-I--EVQY-S---AQ-T--L-Y--R--L-DMCPA-A.-INM-Y-EL--TQ--N--.QNFS 944
TAL.CM.01.8023              TNAAV-T--M-------Q--C------------TKIL-R---R---L---QV-S---IEHL-----P--VS-EMQ-TA--LK-EHTT-V----------S--NK--Q----T-Q-I--EVQY-S---AQ-T--L----R--L-DMCPA-A.LINM-Y-EL--TT--N--.HNFS 945
TAN.UG.x.SIVagmTAN1         --A-V-V--M--------I-I--------FL------AR-----K---K----IIS----ITKL-----P--VSQE-QTI---GQVEHTT-------------S-------RQ--ET-EKI-EDCAF-E---L--CH---------------MTPA--.LINM-T-QLEIQHI-T-Q.Q--S 955
VER.DE.x.AGM3               --A---V--M----------I--------F-------R---K---K----ILS----ITQL-----P---SQE-A-M---GK-EHTT-V----------SI------Q--E---KI--DCQYTE---L--CH--------------LTPA--.LINM-T-QLELQH--TK-.Q--L 952
VER.KE.x.9063               --A---V--M----V---IVI--------F-------R---R---K----ILS----IVQL-----P---SQE-A-I---GK-EHTT-V----------S-------Q--E---KI--DCQYTE---L--CH--------------LTPA--.LINM-T-QLELQT--TK-.Q--L 952
VER.KE.x.AGM155             --A---V--M----I---IVI--------F-------R---K---K----II-----ITHL-----P---SQE-A-M---GKVEHTT-V----------SI------Q--E---KI--DCQYTE---L--CH--------------LT-A--.LINM-T-QLEINT--TK-.Q--L 955
VER.KE.x.TYO1_patent        --A---T--M---------VI--------F-------R---K---K----ILS----ITQL-----P---SQE-A-I---GK-EHTT------------SI------Q--E---KI--DCQYTE---L--CH--------------QTSA--.LIN--T-QLEIQH--TK-.Q--L 955
WRC.CI.97.97CI14            --YEV-T--M--------I--N---------L--KI-S--DAYQ--L-T-Q--S---I-K------T--KSQKMQ-I---L--E---------------I--K-RR--D--HSI-ED--K-E--LA--L-------E---L--.-PA--.L-NM-TS-LE-QQT-Q-K.L-FK 959
WRC.CI.98.98CI04            --YEV-T--M--------I--N---------L--KI----DAYQ--L---QI-----M-K------T--KSQKMQTI---L--E---------------I--K-RR--E--HRI-ED--K-E--LA--L-------E---L--.-PA--.L-NM-TAELE-NQT-Q-K.L-FK 958
WRC.GM.05.Pbt_05GM_X02      --YEI-T--M--------I--N---------L--KI--T-DAYQ--L-TIQ---------K------T--KSQKME-I---L--E---------------I--K-RV--Q--QRI-ED--K-E--LA--IY------E---L--.-PA--.--NM-QS--E-NQT-Q-K.F-FK 950
```

```
                                                                                                    p31 Integrase end
                                                                                                              Pol end
H1B.FR.83.HXB2                NFRVYYRDSRNPLWKGPAKLLWKGEGAVVIQDNS.DIKVVPRRKAKIIRDYGKQMAGDDCVASR.....QDED..........*  1003
H1A1.UG.85.U455_U455A         ----------D-I---------------------.-------------------------M-G-.....----..........-  1003
H1B.US.90.WEAU160_GHOSH       ----------D-----------------------.-----------------------------.....----..........-  1004
H1C.ET.86.ETH2220             ----------D-I---------------------.----------------------A----G-.....----..........-  1011
H1D.CD.84.84ZR085             ----------D-I---------------------.-----------------------------.....----..........-  1004
H1F1.BE.93.VI850              ------------V---------------------.E--I-----------------V-----G-.....----..........-  1005
H1G.SE.93.SE6165_G6165        ----------D-V--------------------N.E--------------------------G-.....----..........-  1001
H1H.CF.90.056                 K---------D-I---------------------.E-------E--------------------.....----..........-  1003
H1J.SE.93.SE9280_7887         ----------D-I-------P-------------.E--------------------------G-.....----..........-  1002
H1K.CM.96.96CM_MP535          K---------E-I---------------------.E--------------------------G-.....----..........-  1004
H101_AE.TH.90.CM240           ----------D-I---------------------.---------------------------G-.....----..........-  1009
H102_AG.NG.x.IBNG             ----------D-I---------------------.---------------------------G-.....----..........-  1004
H103_AB.RU.97.KAL153_2        ----------D-I--------------------N.-----------------------------.....----..........-  1004
H104_cpx.CY.94.94CY032_3      ----------E-I---------------------.----------------------N---G-.....----..........-  1003
H1O.BE.87.ANT70               K---------D-I-----Q-------------KG.--------------E-------T-SM--G.....-T-SESVEQPSEIP-  1010
H1O.CM.91.MVP5180             ----------D-I-----Q-------------KG.----------------------T-SM-N-.....-T-SESMEQPGEIP-  1010
H1O.FR.92.VAU                 ----------D-I-----Q-------------KG.----------------------A-S---G.....-T-GESVEQPSEIP-  1011
H1O.SN.99.99SE_MP1299         K-Q-------D-I-----Q-------------KG.--------------H-------T-SM--G.....-T-SENVEQPGEIP-  1010
H1O.SN.99.99SE_MP1300         K-Q-------D-I-----Q-------------KG.--------------H-------T-SM--G.....-T-SENVEQPGEIP-  1010
H1O.US.99.99USTWLA            ----------D-I-----Q-------------KG.E---------------------T--M--G.....-T-SENMEQXSEIS-  1013
H1O.US.x.I_2478B              ----------D-I-----Q-------------KG.----------------------T-SM---.....-T-SESVEQPGEIP-  1010
H1N.CM.02.DJO0131             ----------D-I--------------------G.-----------------------G----G.....-A-NQDMA......-  1013
H1N.CM.02.SJGddd              ----------D-I-------------------KG.-----------------------G----G.....---NQEME......-  1010
H1N.CM.04.04CM_1015_04        ----------D-I--------------------G.-----------------------G----G.....---NQDME......-  1014
H1N.CM.04.04CM_1131_03        ----------D-I--------------------G.-----------------------G----G.....---NQDME......-  1015
H1N.CM.06.U14296              ----------D-I--------------------G.-----------------------G----G.....---NQDMD......-  1015
H1N.CM.06.U14842              -----F----D-I--------------------G.-----------------------G----G.....---NQDME......-  1016
H1N.CM.95.YBF30               ----------D-I--------------------G.-----------------------G----G.....---NQEME......-  1015
H1N.CM.97.YBF106              --X-------D-I--------------------G.-----------------------G----G.....X--NQDME......-  1015
H1P.FR.06.RBF168              ----------D-I-----T-------------KG.---------------------------DI.....-R-SESLE......-  1004
CPZ.CD.90.ANT                 Q---H-----D-V-----Q-----------K-QE.E------------KE-R-KIEDR-DL-G-.....-N--..........-  1004
CPZ.CM.01.SIVcpzCAM13         K---------D-I-----T-------------QG.E----------------------------.....----..........-  1003
CPZ.CM.05.SIVcpzDP943         ----------D-I-----R-----------K-KE.EV---------------------GSM--G.....---SQNLE......-  1015
CPZ.CM.05.SIVcpzEK505         ----------E-T--------------------G.-----------------------G----G.....---SQDME......-  1014
CPZ.CM.05.SIVcpzLB7           K-------R-D-I----------------L--QE.E-------------------E-AN-L-D-.....---N..........-  1007
CPZ.CM.05.SIVcpzMB66          K---------D-I----------------L--QE.E----------------------------.....---NQNME......-  1011
CPZ.CM.05.SIVcpzMB897         K---------E-I----------------L--KE.-------------------VE-N--L-D-.....---S..........-  1012
CPZ.CM.05.SIVcpzMT145         Q--------KD-I-------------------QD.E---------------------A-S----.....---S..........-  1003
CPZ.GA.88.GAB1                K---------D-I-----T-------------QG.EL---------------------------.....-N--..........-  1002
CPZ.GA.88.SIVcpzGAB2          K---------D-I-----S--------------G.-------------KH----V-----L-GG.....---NQNME......-  1011
CPZ.TZ.00.TAN1                --------A-D-I-----R-----------KEGE.-------------KE-------AGGMDD-.....-N-T..........-  1000
CPZ.TZ.09.UG38                --------G-D-V-----R-----------KEGE.---------V---K-------SAGSMDD-.....-N-AQNME......-   999
CPZ.US.85.US_Marilyn          ----------D-I-----------------KE-E.EV-----------K--------A-SM-G-.....---S..........-  1006
                                                                                                    p31 Integrase end
                                                                                                              Pol end

MAC.US.x.239                  -------EG-DQ-----GE---------ILKVGT.-------------K---GGKEV-SSSHME.....DTGEAREVA.....-  1020
H2A.DE.x.BEN                  --Q----EG-DQ-----GE---------IVKVGT.-----------------GRQEL-SSPHLE.....GAREDGEMA.....C  1033
H2A.GW.x.ALI                  --P----EG-DQ-----GE-----D---IVKVGT.-----------------GRQEL-SGPHLE.....GAREDGEVA.....-  1033
H2A.SN.x.ST_HIV_2_ST          -----F-EG-DQ-----GE-----D---IVKVGA.---II--------K---GRQEM-SGSNLE.....GAREDGEVA.....-  1033
H2B.CI.x.EHO                  -------EG-DQ-----GD---------I-KVGT.E---I---------N--GGKEL-CSADVE.....DTMQAREVAQSN..-  1036
H2B.GH.86.D205_ALT            --Q----EG-DQ-----GE---------I-KVGT.E-------------H--GGKGL-CSADME.....DTRQAREMAQSD..-  1035
H2G.CI.92.Abt96               -------EG-DQ-----GE---------I-KVGT.E------------K---GGKEV-SSTNME.....DTRQTGEVA.....-  1016
H2U.FR.96.12034               -------EG-DQ------D---------L-KVGT.----I------------GGKEL-SSSNLE.....GA-KVREMALPD..Q  1032
COL.CM.x.CGU1                 --KA-WKEH.TGE-Q--GE-V---------RNSQGTLF-K----V--T-LQYGEDV-SENLL-N.....GQKEAETVKGMD..-   991
DEB.CM.04.SIVdeb04CMPF3061    G-K----EGAD-S-----T-----------KLDTG-L---------V-K----DVGSKIDPQDT.....HE............-  1011
DEB.CM.99.CM40                G-K----EGAD-T-----T----------CKTEVG-----------V-KE---D-DSKVNTQNS.....NE............-  1008
DEB.CM.99.CM5                 GYK----EGAD-T-----T----------CKTEVG-----------V-K----GVDSKTDTQNP.....SE............-  1008
DEN.CD.x.CD1_CMO580407        --K----EGA--Q-Q---E-----------KTQAG--FP---------KP-NAE.SRKGEQNKG.....M-SETDI.......-  1032
DRL.x.x.FAO                   -------EG-DQ------E-V---------KEGT.-L--I--------K----NVDSHTNME-....................-  1013
GOR.CM.04.SIVgorCP684con      ----------D-I-----T-------------KG.---------------------------DT.....-G-SESLEQSS...-  1003
GOR.CM.07.SIVgor2139_287      ----------D-I-----T-------------KG.---------------------------DT.....-G-SESLEQSS...-  1003
GOR.CM.07.SIVgorCP2135con     ----------D-I-----T-------------KG.---------------------------DT.....-G-SESLEQSS...-  1003
GRV.ET.x.GRI_677_gri_1        --K----EG-D-V-----R-I---------KEGE.-------------K---ERKTM-SEGSMEGVREANKQMEGDSDLQDQE-  1041
GSN.CM.99.CN166               K------QGA----Q---V---------V-TQAGE-IT----------KP--EAKTE-VGG-TH.....TNN-..........-  1016
GSN.CM.99.CN71                K------QGA--H-Q---V---------V-TQAGE-IT----------KP--EAK-E-VGS-AH.....TSN-REEGRMAD..-  1017
LST.CD.88.SIVlhoest447        K-K----QG-D-Q-----Q----------VKEGD.T-FS-------LVK---EGPKDSEGGLHN.....N.............-  1020
LST.KE.x.lho7                 K-K----QG-D-Q-----Q----------VKEGE.N-FS-------LVK---E.GPK-SESSLD.....NN............-  1020
MND-1.GA.x.MNDGB1             --K--F-EG-DQ-----GI----------LKYQE.E--I-----C---K---E.SGKNSQ-NLE.....SV............-  1010
MND-2.CM.98.CM16              -------EG-DQQ-----E-I---------KEGT.-L----------------TVDS-PN-EA....................-  1011
MND-2.GA.x.M14                -------EG-EQ------E-I---------KEGT.-L----------------TVDSNPHMED-.....-ETA..........-  1017
MND-2.x.x.5440                --Q----EG-DQQ-----E-I---------KEGV.-L-----------K----TVDSNPHMGR....................-  1012
MNE.US.x.MNE027               -------EG-DQ-----GE---------ILKVGT.-------------K---GGKEV-SSSHME.....DTGEAREVA.....-  1016
MON.CM.99.L1_99CML1           K----F-TGAD-Q-----R-----------KTEEG--LT---------KP--TTENVGGDTNQY.....NLRKQDGLAN....-  1018
MUS-1.CM.01.CM1239            -------TGSD-S-----A----------VKTEQGQVIT---------KP--AKENVGSKSNTG.....DHRKEDGLDN....-  1012
MUS-1.CM.01.SIVmus_01CM1085   -------TGSD-S-----V---R-------KTD-GEVIT---------KP--AKEDVGSKSDTG.....DLRKEDRLDN....-  1012
MUS-2.CM.01.CM1246            K------RGA--Y-LR--R--------L--KTKEGE-VT---------K---TRQNV-SEPD-V.....HVRKEDGLAD....-  1017
MUS-2.CM.01.CM2500            K------RGA--S-Q---Q--------I--KTDQGE-IT---------K---TKENV-SDPNPV.....Y-RKEDGLAD....-  1014
OLC.CI.97.97CI12              SGYRV-YKNKEGQ-C---E---------LLKEGD.KYFS--L-----VKP--VDSSANV........................-   995
RCM.CM.00.SIVagi_00CM312      -----F-EG-DQQ-----T-I-----------GQ.-L-------C--VK---RKNVDSETNMEG.....GQ-EN.........-  1018
RCM.CM.02.SIVrcm_02CM8081     -----F-EG-DQQ-----T-I-----------GQ.-L-------C---K---RKDVDSETSMEG.....RQ-KD.........-  1024
RCM.GA.x.SIVRCMGAB1           -------EG-DQ--R-----I----------EETG-L-----------KE--RKDVDSEANLAG.....RQ-EN.........-  1019
RCM.NG.x.NG411                -----F-EG-DQQ-------I-E---------QE.-L-------C---K---RKEVDRETNMEG.....RQ-ES.........-  1023
SAB.SN.x.SAB1                 -------EG-D-V-------I----------EQG.EL-TI--------K----..-L-SQAPLE.....GNGRTAGEVD....-  1040
SMM.SL.92.SL92B               -------EG-DQ------E-----------KVGT.E------------K---GRQEMGSSASME.....DQ............-  1006
SMM.US.x.SIVsmH635F_L3        -------EG-DQ--R--GE---------ILKVGT.E------------K---GGKEL-SGSHLE.....DTGEAREVA.....-  1016
STM.US.89.STM_37_16           -------EG-DQ-----GE---------IVKVGT.-------------K---GGKEV-SGSHLE.....DTREAGEVA.....-  1016
SUN.GA.98.L14                 K-K----EG-D-Q-----------------KQGE.N-L--------LVK---GESSSVEM-G.....................-  1017
SYK.KE.x.KE51                 --K----HGT.SD-Q---S------------TPDQQVIA---------SSD-ERVDSGTHLEIT.....SKSN..........-  1041
SYK.KE.x.SYK173_COMGNM        -------QGK.NE-----R----------V-TEEG--FA---------T-H-ER-DSGSH-END.....PKT-..........-  1037
TAL.CM.00.266                 D-K----KGA----Q---H-I--------LRTDEGEVIT---------KP--QALGNKIDLE-S.....KEQ-AEMGRDN...-  1020
TAL.CM.01.8023                D-K----KGA----Q---H-V--------LRTDEGEVIT---------KP--QA-GNKTDLEGS.....KEQ-AEMGRDN...-  1021
TAN.UG.x.SIVagmTAN1           --K----EG-E-V-----T-----------KEGE.EL-----------KE--ERKTVG-KTNME.....G.............-  1020
VER.DE.x.AGM3                 -------EG-D-V----GQ-I---------KGGV.EL-EY--------K--EPRKRMG-ESNLE.....GAGGADN.......-  1023
VER.KE.x.9063                 -------EG-D-V-----Q-I---------KEGE.EL-I---------K--EPRKRMG-EGNME.GLRGP-NQMARNSQILDD-  1034
VER.KE.x.AGM155               -------EG-D-V-----R-I--------LKEGE.EL-----------K--EPRKTLG-ETHLE.....GAGGSDHQMAG.DS-  1032
VER.KE.x.TYO1_patent          -------EG-D-V-----Q-I--------LK.DGS-L-----------K--EPKQRVGNEGDVE.....GTRGSDN.......-  1026
WRC.CI.97.97CI14              --Q----TGA-QQ-Q--G--P------L--ETPE.G-IT-------L-KVWNGEGMDRSSSDKN...................-  1023
WRC.CI.98.98CI04              --Q----TGA-QQ-Q--G---------L-VETPE.G-IT--K------KVWNGEGMDRSSSNKD...................-  1022
WRC.GM.05.Pbt_05GM_X02        --Q----TGASQQ-Q--G--V------L-V-TPEG-LLT---------KTWDGKGMDSSTTN.....................-  1013
```

```
                          Vif start
H1B.FR.83.HXB2            MEN..................RWQVMIVWQVDRM..RIRTWKSLVK...HHMYVSGKARGWFYRHHYESPHPRISSEVHIPLGD..........ARLVITTYWGLHTGE.....RDWHLGQGVSIEWRKK.......................RYSTQVDPELADQLI  107
H1A1.UG.85.U455_U455A     ---..................-------------..K----N----..------K--Q---------R-S-V---------E..........----VR----------.....K-----H-------L-.......................--------D---H--  107
H1B.US.90.WEAU160_GHOSH   ---..................-------------..---------..-----I-K--K--S-------T-----------E..........GK--------------.....-----------------Q.......................--------D------  107
H1C.ET.86.ETH2220         ---..................----L--------..K----N----..---HI-RR-N--V-----D-R--KV---------E..........---I-K-----Q---.....------H-------LR.......................S-N-----G---H--  107
H1D.CD.84.84ZR085         ---..................-------------..--N--------..Y--HI-K--K---------D---K-----------..........----V-----------.....-E-----------R.......................--------G------  107
H1F1.BE.93.VI850          ---..................---L---------..--N--------..Y-----K--K--S----FQ-R---V--------EE..........VK---------P----.....-E-----------QG.......................K-R--I--G------  107
H1G.SE.93.SE6165_G6165    ---..................-------------..-----H----..------K--------P--A-R---V-----------..........-T--V-----------.....K--Q--H-------QR.......................--R---E-D---H--  107
H1H.CF.90.056             ---..................-------------..--N--------..Y--HI-R----------F--T------------E..........---------N------.....-E-----------L-.......................-------E-G------  107
H1J.SE.93.SE9280_7887     ---..................-------------..--N--------..Y--N--K---K-L-----D-N--K---------E..........-I---------Q----.....--------------QR.......................--R-----G------  107
H1K.CM.96.96CM_MP535      ---..................-------------..--N--------..Y--HI-K--NR-Y-------R--K-----------..........-E--V-----L-----.....--------------L-.......................--R---E-D------  107
H101_AE.TH.90.CM240       ---..................-------------..-----N----..----I-K--KK---------Q--KV---------E..........-----R-----Q----.....K-----H-------QR.......................K----I--D---K--  107
H102_AG.NG.x.IBNG         ---..................-------------..-----N----..------K--K-------F--R--KVC----------..........----VR----------.....------H-------KQ-.......................-------I--D------  107
H103_AB.RU.97.KAL153_2    ---..................-------------..-----N----..--I-I-K-----V-K-----RN--------------..........-K---K----------.....--------A------E.......................--------N------  107
H104_cpx.CY.94.94CY032_3  -A-..................-------------..K----N----..------K--K-*---------R--KV---------E..........----VR-----QP---.....Q-----H-------LR.......................--------D------  106
H10.BE.87.ANT70           ---..................----L-------Q..KVKA-N-----..Y-K-R-R-TEN-W-------RN--V--S-Y--V-V..........-HV-V------MP---.....--E---H-------Y-.......................K-K--I---T--RM-  107
H10.CM.91.MVP5180         ---..................----L----I--Q..KVKA-N-----..Y-K-M-K--AN-R-------RN-KV--A-Y--VAE..........-DI-V------MP---.....-EE---H-------QY-.......................E-K--I---T--RM-  107
H10.FR.92.VAU             ---..................----L-------Q..KVKA-N-----..Y-K-R-R--TD-Y-K------IN--V--R-Y--V-P..........-CIMV------MP---.....-EEQ--H------QCR.......................K-I--I---T--K--  107
H10.SN.99.99SE_MP1299     ---..................----L-------Q..-VK--N-----..Y-K-R---T-D-Y----F--RN--V--G-Y--V-G..........PWI-V------MP---.....--E---H-------Y-.......................K-K--I---T--RM-  107
H10.SN.99.99SE_MP1300     ---..................----L-------Q..-VK--N-----..Y-K-R---T-D-Y----F--KN--V--G-Y--V-G..........PWI-V------MP---.....--E---H----D--YR.......................K-T--I---T--RM-  107
H10.US.99.99USTWLA        ---..................----L-------Q..KVK--NX----..Y-K-R-K--KD-Y------TRN-----G-Y--V-G..........-MI-V------MP---.....--E---H-------EY-.......................N-K--I---T--RM-  107
H10.US.x.I_2478B          ---..................K---L-------Q..KVKV-N-----..Y-K-R-R-TKD-Y------CRN-----S-Y--I-P..........-FI-V------MP---.....--E-M-H-------QY-.......................K-A--I---T--RM-  107
H1N.CM.02.DJ00131         ---..................-----V------R..-K--H-----..------R--E--Y------TN--K---------V-L..........-----------T----.....-S----H-------KLG.......................K-K------M--K--  107
H1N.CM.02.SJGddd          -A-..................-----V--------..K-KK-N-----..Y-----K--Q--Y------TR--K---------V-Q..........-----V-----T----.....QS----H-------LR.......................K-K------M-----  107
H1N.CM.04.04CM_1015_04    ---..................-----V--------..---I-N-----..---H--K--K--Y------R--KT--------V-Q..........-T---V-----T----.....QP----H-------LR.......................K-K------M--K--  107
H1N.CM.04.04CM_1131_03    ---..................-----V--------..K--I-N-----..---H--K--K--Y------TR--K---------V-Q..........-----V-----T----.....QP----H-------LR.......................K-K------M--K--  107
H1N.CM.06.U14296          ---..................-----V--------..K----T-----..-------KR-K--Y------IN---K--------X-Q..........-K--VI-----T----.....QP----H-------LG.......................K-K------M--R--  107
H1N.CM.06.U14842          ---..................-----V--------..K----N-----..-------K--T--H------TXX-KK--------V-Q..........-K---I-----T----.....QP----H-------LG.......................K-K------M--R--  107
H1N.CM.95.YBF30           ---..................-----V--------..K--K-N-----..-------K--K--Y------TH--K---------V-Q..........----TV-----T----.....QS----H-------LR.......................K-K------M--K--  107
H1N.CM.97.YBF106          ---..................-----V-----X-..K----N-----..-------K--K--Y------TK--KT---X----V-X..........-X---V-----T----.....QP----HX------QG.......................K-K------M--K--  107
H1P.FR.06.RBF168          ---..................----------IS-E..KVKA-N-----..Y-K-R-K-TKN--------XV------G-Y--V-A..........-II-V------MP---.....KEEQ--H-X-V---QG.......................K-T--I---T--R--  107
CPZ.CD.90.ANT             -..........................I------..--NI---------..-IWETKVLKP-K------ND--KKGE------PTLD..........KK--V-VF---QC---.....-P----H-------CG.......................K-I------T---M-  101
CPZ.CM.01.SIVcpzCAM13     ---..................K-------------..--KA-N-----..Y---KTK--KD--------H-N--V--------FK-N..........SK-I-----A-T---.....-A----H----Q--MG.......................S-V-----YV--K--  108
CPZ.CM.05.SIVcpzDP943     ---..................-----V--------..-VK--N-----..Y---RTK---K----------SN--V---I----EK..........-K--V------MP---.....HE----H-------YG.......................S-R--I---M--R--  107
CPZ.CM.05.SIVcpzEK505     ---..................-----V--------..-V---N-----..-------K------------TDN-K----I-----E..........-K---V-----MP---.....-P----H-------QG.......................I-R--I-------K--  107
CPZ.CM.05.SIVcpzLB7       ---..................--K-L----I----..--KA-N-----..Y-I----R------------EN-KV---I-----E..........-K---V-----TP---.....------H------KQ-.......................K----I------HM-  107
CPZ.CM.05.SIVcpzMB66      ---..................-------------..--K--N-----..-------R---N---------NN-K----------E..........-K-----------T----.....-A----H-------QR.......................S-R--I--D------  107
CPZ.CM.05.SIVcpzMB897     ---..................--K------I----..-VKA-N-----..-------K------------EN-KV---I---V--..........-K---V-----MP---.....------H-------Q-.......................------I------H--  107
CPZ.CM.05.SIVcpzMT145     ---..................-----Q--------..-VK--N-----..Y---R-K--KN-------DGTN--MA-QIY--INQ..........-K-IV------MP---.....------H--A----ER.......................G--------T--R--  107
CPZ.GA.88.GAB1            ---..................-------------..--K--N-----..Y-I-R-K-----------DH-N-KVA--I---FR-Y..........SK-IV----A-SP---.....-A----H----Q--LG.......................S-V-----FT--R--  108
CPZ.GA.88.SIVcpzGAB2      --S..................-----------L----..-----N-----..F---K-K-----------D--N-KV--------IR-..........-KI-VC---N-MP---.....QP----H-------QR.......................T-R--I--QT--R--  107
CPZ.TZ.00.TAN1            ---..................-----QV--MI----..-L---T-----..--IFTTKCCKD-K------TDT-KRAG-I----TER..........SK--VLH----AC---.....-P----H-IGL---QG.......................K----I---T-----  108
CPZ.TZ.09.UG38            ---..................-----QV--MI----..-L---N-----..--IFT-KCCSS-K------TDI-KRAG-I---INEK..........-KI-VLH----AC---.....-P----H--GL---QG.......................K-V--L--AT-----  108
CPZ.US.85.US_Marilyn      ---..................-----L--------..-VKA-N-----..Y---R-K--K---------TN--VN------IDK..........-KI--I-----K----.....QA----H-------QG.......................T-K--I---M--K--  107
                          Vif start
MAC.US.x.239              --E...............EK--IAVPT-RIPER...LER-H--I-..YLKYK.TKDLQKVC-VP-FKVGWAWWTCSRV-FPLQEG.........SH-EVQG--H-TPE......KG-LSTYA-R-T-YS-.......................NFW-D-T-NY--I-L  109
H2A.DE.x.BEN              --E...............DRN-I-VPT-R-PGR...MEK-HA----..YLKYR.TKDLEEVR-VP-HKVGWAWWTCSRV-FPLQGK.........SH-E-QA--N-TPE......KG-LSSHA-RLT-YTE.......................KFW-D-T-DC--I--  109
H2A.GW.x.ALI              --E...............GKS-I-VPT-R-PGR...MEK-H-----..YLKYR.TKDLEKVC-VP-HKVGWAWWTCSRV-FPLQGR.........SH-E-QA--N-TPE......KG-LSSYA-R-T-YTE.......................KFW-D-T-DC--S--  109
H2A.SN.x.ST_HIV_2_ST      --E...............GK--IAVPT-R-PGR...MER-H--I-..YLKYR.T-DLEKVC-VP-HKVGWAWWTCSRV-FPLKGE.........SH-E-QA--N-TPE......KG-LSSYS-RLT-YTE.......................KFW-D-T-DC--S--  109
H2B.CI.x.EHO              --E...............EKN-IAVPT-RIPCR...LER-H--I-..YLKYR.TKDLQQVS-VP-HKVGWAWWTCSRV-FPLKEG.........-H-EVQG--N-TPE......-GFLSSYA-RLT-YER.......................SFY-D-T-DV--R-L  109
H2B.GH.86.D205_ALT        --E...............EKD-I-VPT-RIPGR...LER-H--I-..YLKYR.T-ELQQVS-VP-HKVGWAWWTCSRI-FPLNKG.........-W-EVQG--N-TPE......-GFLSSYA-RLT-YER.......................NFY-D-T-DV----L  109
H2G.CI.92.Abt96           --E...............EK--IAVPT-RIPGR...LEK-H--I-..FLKFK.TKDLQKAV-VP-HKVGWAWWTCSRV-FPLTKE.........-H-E-QG--N-TPE......KG-LSQYA-RLT-YTR.......................KFY-D-T--T----L  109
H2U.FR.96.12034           --E...............EKN-I-VPT-R-P-R...LER-H--I-..HL#YN.TKELQKVC-VP-HKVGWAWWTCSRV-FPLKKG.........SH-EVQA--N-TPE......KG-LSSYA-RLT-YTR.......................GFW-D-T-DC----L  108
COL.CM.x.CGU1             -.................EKM-A-R-CYPMGKR..KPKPLREWI-.......KQFPECQYY-LGKT---EWWTEGKIEFY--S..........AIK-S-IYLGT-TPDR.......Y-R-PP--Y-LQIG.......................KWE-DLSWQQ-VGYM  102
DEB.CM.04.SIVdeb04CMPF3061 -.................EKM-V-RLTHKIPMN..KLDR-C-AI-...KGK-KTKELAPITWQ--F-LRNSYFTQTRI-FPITK.........EGKV-VELF-C-APE.....KGWLPGMA.-GLS-E.......................QGNWQ-ELS----EH--  108
DEB.CM.99.CM40            -.................EKT-I-RLTHKIPMN..KLDR-C-AI-...REK-KTKELDTVQWI--F-LRNSYYTQTKIVFPITK.........EGGV-VELL-C-TPE.....KGWLPTMA.-A-A-E.......................TKTWK-ELT-D---H--  108
DEB.CM.99.CM5             -.................EKE-I-RLTHKIPVN..KLDK-C-AI-...RGK-KTKELQAIKWI----LRRAFYTQTKIVFPVTE.........EGKV-VEIF-S-APE.....KGWMPSMA.-G-S-E.......................QTGWQ-ELT-D---H--  108
DEN.CD.x.CD1_CMO580407    -..........QNLGRENRTKE-I-RLTYKIGET..KLEH-M-TL-.....SIIYKTNELKGTVIIPHFRITWEWYTMTKIHIPLR.........A-YMEVKL--H-TPE.....KGYLSQVAISIAYVH.......................LSTGWA-E---VT--EI-  117
DRL.x.x.FAO               -.................EKM-I-IPT-R-E-Y..MVER-N--C-..Y-KHKGEKYLER-H-AP-FQCS.GWWTHSQWTIPFKDK.........SKI-V-VL-N-TPD......KG-LSKYAIT--HIQE.......................KFC-LI--TT--RI-  108
GOR.CM.04.SIVgorCP684con  ---..................---------IP-E..-VKA-N-----..Y-K-RTKQ-KQ--------AIN--V--G-Y--V-P..........-EIIV------MP---.....KEEQ--N--G----QG.......................K-I--L---T--RI-  107
GOR.CM.07.SIVgor2139_287  ---..................---------IP-E..-VKA-N-----..Y-K-RTKQ-KQ--------AVN--V--G-Y--V-P..........-EIIV------MP---.....KEEQ--H--G----QG.......................K-I--L---T--RI-  107
GOR.CM.07.SIVgorCP2135con ---..................---------IP-E..-VKA-N-----..Y-K-RTKQ-KQ--------AVN--V--G-Y--V-P..........-EIIV------MP---.....KEEQ--H--G----QG.......................K-I--L---T--RI-  107
GRV.ET.x.GRI_677_gri_1    --R...............EKQ-I-RV--R-SER..Q-SR-RGI-T..YKIR...N-QLP-E----WQVQWQFWTYSQF-IPLSKD.........DYIEVNI-HN-TPE......-G-LSSH--GLSYYHQKG.......................-K-E---GT--RM-  109
GSN.CM.99.CN166           --R.............QRQKM-V-QPIQIMTER..KVDWLLRAT-...--IWSGK..TPFV-V---QLQ-Q-FTQNKIRLAM-TRKVGE.EVEATYIE--IL-DTTSHGPASLS-STYYQ-A-I---VYNRSKLG...........AREGDI.IW---NLT-GV-M-I-  136
GSN.CM.99.CN71            --R.............QRQKM-V-QPIQVMTER..KVEWLIRAT-...--IWSGK..TPFV-I---QLQ-Q-FTQNKIRLAL-TRQVGE.EVEATYIE--IL-DTTSHGPASLS-STYYQ-A-T---VYNRSKLG...........ARETDI.TW---NLT-DV-M-I-  136
LST.CD.88.SIVlhoest447    --................RDQRI-RVAYITT-S..LVEKFLDIRR......AHKETQND-VSL-GTGVGWEFYTYNNIVIPLETG.........T.--VRC-AH-AA-.......-G-ISQWA-----IYG.......................N-Q-EI--I--N-M-  105
LST.KE.x.lho7             --................RDQRI-RVA-ITT-S..L-EKFLDI-R......LNTEAKEN-VGL-GTGVGWEFYTYHKLVIPLESG.........T.--VRI-AH-AA-.......-G-ISQWA-----VYR.......................D-Q-EI--I----M-  105
MND-1.GA.x.MNDGB1         -.................RVERI-RLT-K-SSQ.---EK-HW--RRQMAWATANNEEGC-WLYP-FMAYNEWYTCSKVVIIINRD.........I--IVRS--H-QIE......VGCLSTYA----AVVRPPPFEK.......................EWC-EIT--V--H--  118
MND-2.CM.98.CM16          -.................EKQ-I-IPT-KL--H..AVER-N-----..Y-KYKGEKHLDK-DLKA-FQCS.GWWTHSQKYIPLEED.........EQII--IL-N-TPE......KG-LSTYAMT--YKFT.......................N-Y-HI--QT--RM-  108
MND-2.GA.x.M14            -.................EKQ-I-IPT-RI--R..QLEK-N-----..Y-KYKEEKHLDK-ELF--FQCS.GWWTHSQK-IPFKDG.........SKII--AL-N-TPE......KG-LSQYAITA-YN-G.......................D-Y-HI--VT--RM-  108
MND-2.x.x.5440            -.................EKQ-IAIPT-E.--QDVKL-D-N-I--..Y-KYKGEKHLEN-QLYP-FQCS.GWWTHSQKKIPFKDG.........SKII--VL-N-TPE......KG-LSQYAFTV-YQ-E.......................N-F-YI--VT--RI-  109
MNE.US.x.MNE027           --E...............EK--IAVPT-RIPER...LER-H--I-..YLKYK.TKDLQKVC-VP-HKVGWAWWTCSRV-FPLQEE.........SQ-EVQG--N-TPE......-G-LSTYA-R-T-YSR.......................NFW-D-T-DY--I-L  109
MON.CM.99.L1_99CML1       -A................PRKM-V-TPTNIISES..KMDWLIRCT-...W-ILTGS..APFV-V---QLHNQ-F-QNKIKLPL-LGITQEGESWATY-E--I--D-TNVGPSALSPTVYNK-AYT---VYWMQKVERVPLDPVWNPNPRDREVRTW-A-HLT-D--N-IV  147
MUS-1.CM.01.CM1239        -A................PRKM-V-SPIQETTER..KMDWIIRAT-...--IFSGK..TPFV-V---QVQ-Q-F-QNKIKLALTQNTLESGETEVTFIE--VL-DVTNVGPASLSKSTYWK-SYIL--VYMRKGPR...........DRERDY.VW-H-YLT--I-MKI-  135
MUS-1.CM.01.SIVmus_01CM1085 -A................PRKM-V-SPIQETSER..K-DWIIRAT-...YYILSGR..AAFT-V---QLY-Q-F-QSKIRIPLTQNTRESGAVESTYIEL-VL-DVTNVGPASCSKSTWWK-SFIL--VYLRRHPQ...........DREQDY.IW---YLS-T--MRI-  135
MUS-2.CM.01.CM1246        -A................PGKM-I-SPIQYMSER..KMDWLIRCT-...-YIFSGK..APFT-V---QLQ-Q-FTQNKIKIPLSINRIEGGPTETTYIE--ILLDVTNVGPASLS-STYWQ-SYILK--YIRAAPQ...........NREIDM.VH-E-FL---VTM-I-  135
MUS-2.CM.01.CM2500        -A................QRKM-I-TPIQFMTER..-MDWLIRCT-...--ILSGK..APFS-I---QVQ-Q-FTQNQIKIPLTIKKREEDKVESTYIQ--IL-DVTNVGPASLSKSTYWQ-SYIL---VYVRAAPQ...........DREVDM.VW---YL-----M-I-  135
OLC.CI.97.97CI12          --....................IAVPMYRLNNG..QVEHV-FMA-.....LAYKFWGYH-IWQNC-KY-REWWTHGKIYIPVPG.........GEIE--L--HIRH-......KGLQNSG--NFKYIVG.......................ISWMYLT--YV-FMV  103
RCM.CM.00.SIVagi_00CM312  -A................EKM-I-RPI-R---R..K-EQ-H-----..Y-QYKGKKE-KE-E-VP-FKV-WGWW-HSEIHVPLGKN.........TKVK-----N-TAE......KG-LSTY--G-AYIDNTCDP.......................C-F-DI--I---KM-  114
RCM.CM.02.SIVrcm_02CM8081 -A................EKM-I-RP--R---R..K-EQ-H-----..Y-MYKGKKE--E-E-VP-FKV-WGWW-HSEVHIPLGNN.........TKIKV----N-T-E......KG-LGTY-AALAYIDQKCDP.......................P-F-DI--IV--S--  114
RCM.GA.x.SIVRCMGAB1       -A................EKM-I-RPI-R---R..K-EQ-H-----..Y-QYKGKKA-KE-E-VP-FKV-WGWW-HSEVHIPLEEG.........SK-K-----N-TVE......KG-LGTY--G-L-I-G.......................D-V-D-F-WT--S--  110
RCM.NG.x.NG411            -A................EK--IERPI-K---R..KVEQ-H-----..Y-QYKGEKE-KD-E-VP-FKV-WGWW-RSEVHIPLGNN.........TKIKV----S-TIE......KG-LGTY--G-AYIDNRCDP.......................C-F-DI--TV--K-V  114
SAB.SN.x.SAB1             -.................EKH-I-RPL-K-TGG..QQER-T-----..Y-MHVS.KQCVH-R-TP-TKIRWNWY-YQEWVIPLKDG.........-LIKV-N--H-TPE......KG-LETYATG-GYS...KG.......................EWF-EL--WT--HI-  108
SMM.SL.92.SL92B           --E...............GK--VAVP--RIS-RI..VER-H-CI-..F-KYK.TRELEKAC-VP-HKVGWAWYTASRV-FPLEEG.........SH-EVQV--N-TPE......KG-LSSYA-R-T-YSE.......................KFW-D-T-DV----A  110
SMM.US.x.SIVsmH635F_L3    --E...............EKN-I-VPT-RIPER...LER-H--I-..HLKYN.TKDLQMAC-VP-HKVGWAWWTCSRV-FPLRDK.........TH-EVQG--N-TPE......KG-LSTHA-R-T-YSR.......................NFW-D-T-DC--T-L  109
STM.US.89.STM_37_16       --E...............EK--I-VPT-RIPGR...LER-H--I-..HLKYN.TKELSKAC-VP-HKVGWAWWTCSRV-FPLQGE.........-H-EVQG--N-TPE......KG-LSEYA-R-T-YTR.......................NFWSD-T-DC----L  109
SUN.GA.98.L14             --ENPPQWRWWDEREWEDRRQYKI-R---LI--I..AVEKFLD-RR......MHRETKDD-VSMYGTGTGWEWYTYNKI-IPVTYG.........T.V-VRI-GH-TPA......KGLINQW-C-M--IYN.......................S-Q-EI--LV---M-  122
SYK.KE.x.KE51             -.................EKE-I-VLT-KLHPN..Q-DKMQHIL-...V-K-K-KQLEKAT-K---QITWEWWTRAQWEIPVGQ.........GV-I-KF-HN-TPE......KGLLRTEG.-GLI-AHPS.......................GW--E-T-QT--A-V  108
SYK.KE.x.SYK173_COMGNM    -.................EKE-I-VPT-RMTPR..Q-DRLQHII-...T-K-K-KELEKAT-K---QIEWQWYTYCQWTIPVGD.........GTIW--F-HN-APE......RGWL-MQG.IR-QYQWN.......................QWN-DLT-AV--R--  107
TAL.CM.00.266             -.................DKL-VTRLT-K-P-N..-MQR-VGTIR...F-IHKTKYLQKLIWV--FQLR-ASYTQAKW-LPLEEK.....DTRVEY-T--Q--C-APD.....HGYTSDWDTI----SVSDIVRD.................EKV-A-E-T-NV--H--  120
TAL.CM.01.8023            -.................DKL-VTRLT-K-P-N..-MQR-VGTIR...Y---KTKYLQKLIWV--FQLRMASFTQAKW-LPLEEK.....ENRVEYIT--Q--C-APD.....KGYTSDWDTI----SVSDVTLD.................AKV---ELT--V--H--  120
TAN.UG.x.SIVagmTAN1       --R...............EKL-VTRLT-R-SGE..H-DK--GI---..Y-MR...NRLQD-T-LM--QCGWAWYTCSRFLIPLGGE.........GKI-VDC--H-TPE......QG-LSTYA-A-SFENWQN.......................T-K-E-T-DV--HM-  110
VER.DE.x.AGM3             -NQ...............EKE-VMRVT-K-PEE..L-TK-QGI-R..YWMR...TRKLD-K--M--QITWAWYTMSRYEIPLGQH.........GSIHVDL--H-TPE......KG-LSTYAEG-QYLSNRDP.......................W-R-EL--AT--S--  111
VER.KE.x.9063             -NQ...............EKE-VMRVT-K-SGD..L-TK-QGI---..YWMT...KRNLK-Q-MM--QINWAWYTMCRY-IPIGLE.........GEIQ-DC--H-TPE......-G-LSTYAIG-QYLGAVG.......................NFR-EL--AT--GM-  110
VER.KE.x.AGM155           -SQ...............EKH-VMRLT-K-QEE..V-TK-QGI-R..YWMN...KRNLK-E-KM--QITWAWYTMSRYVIPLPGS.........GEIHVDI--H-APK......QG-LSTYA-G-QYVSLVND.......................K-R-EL--NT--SM-  111
VER.KE.x.TYO1_patent      -NP...............NKE-VMRVT-K-PGD..L-TK-QGI-R..YWMR...QRNLK-N-YM--QITWAWYTMSRYVIPIGKH.........GEICVDL--H-TPE......QG-LSTYA-G-QYVSNLES.......................K-R-EL--AT--SI-  111
WRC.CI.97.97CI14          --..................KE-IGVPLIRISKG..KF-R-TYYIA......NLIKAGQAN-LWGSWHAAAECFTRHIVKIPIQGH.........WEMQ-RC--N-SPE......KG-MDTM-FN-RMVYEN.......................-WE-FIM-SE-NKI-  106
WRC.CI.98.98CI04          --..................KE-IGVPLIKISKG..-FKR-AYWIA......HLMQNKKAS-LWGTWHTQAECFTRYLLRIPLQGH.........WE-Q-RC--N-APE......KG-LDTM-FN-RMVYEN.......................-WE-FIM-SE--VM-  106
WRC.GM.05.Pbt_05GM_X02    --..................KE-IAVPLIKISKG..-F-R-SYWIA......ELIQEKQAR-LFGSWHVHAECFTRHIVRIPLQGH.........WEMQ-RC--N-APE......IG-IDTL-FN-RMVWEN.......................KWE-FITVK---V--  106
```



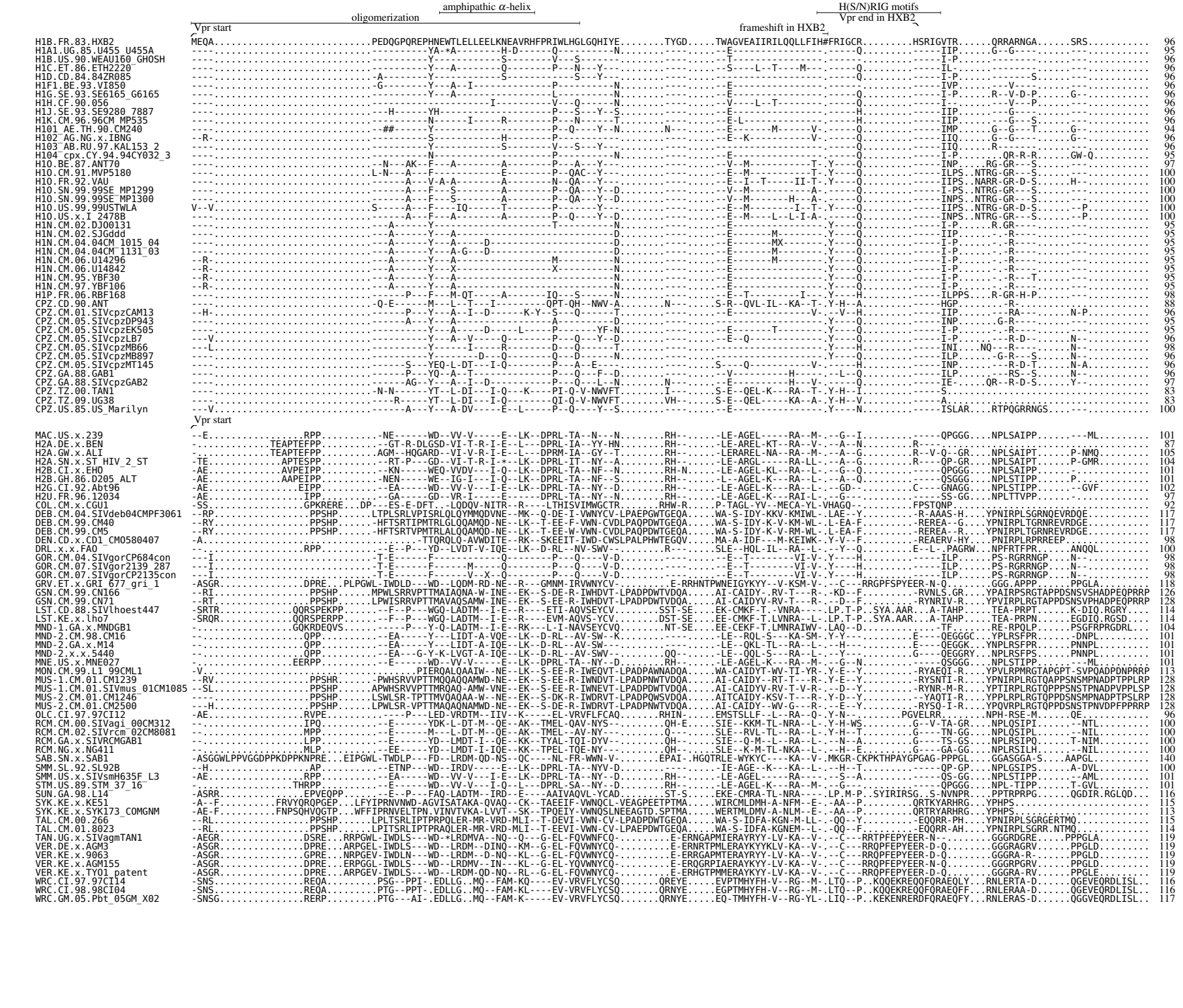
Vif end

```
H1B.FR.83.HXB2                ..................*  192
H1A1.UG.85.U455_U455A         ..................-  193
H1B.US.90.WEAU160_GHOSH       ..................-  193
H1C.ET.86.ETH2220             ..................-  193
H1D.CD.84.84ZR085             ..................-  193
H1F1.BE.93.VI850              ..................-  193
H1G.SE.93.SE6165_G6165        ..................-  193
H1H.CF.90.056                 ..................-  193
H1J.SE.93.SE9280_7887         ..................-  193
H1K.CM.96.96CM_MP535          ..................-  193
H101_AE.TH.90.CM240           ..................-  191
H102_AG.NG.x.IBNG             ..................-  193
H103_AB.RU.97.KAL153_2        ..................-  193
H104_cpx.CY.94.94CY032_3      ..................-  192
H1O.BE.87.ANT70               ..................-  193
H1O.CM.91.MVP5180             ..................-  193
H1O.FR.92.VAU                 ..................-  193
H1O.SN.99.99SE_MP1299         ..................-  193
H1O.SN.99.99SE_MP1300         ..................-  193
H1O.US.99.99USTWLA            ..................-  196
H1O.US.x.I_2478B              ..................-  193
H1N.CM.02.DJO0131             ..................-  193
H1N.CM.02.SJGddd              ..................-  193
H1N.CM.04.04CM_1015_04        ..................-  193
H1N.CM.04.04CM_1131_03        ..................-  193
H1N.CM.06.U14296              ..................-  193
H1N.CM.06.U14842              ..................-  193
H1N.CM.95.YBF30               ..................-  193
H1N.CM.97.YBF106              ..................-  193
H1P.FR.06.RBF168              ..................-  196
CPZ.CD.90.ANT                 ..................-  186
CPZ.CM.01.SIVcpzCAM13         ..................-  194
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.CM.05.SIVcpzMT145         ..................-  193
CPZ.GA.88.GAB1                ..................-  194
CPZ.GA.88.SIVcpzGAB2          ..................-  193
CPZ.TZ.00.TAN1                ..................-  199
CPZ.TZ.09.UG38                ..................-  199
CPZ.US.85.US_Marilyn          ..................-  193
```

Vif end

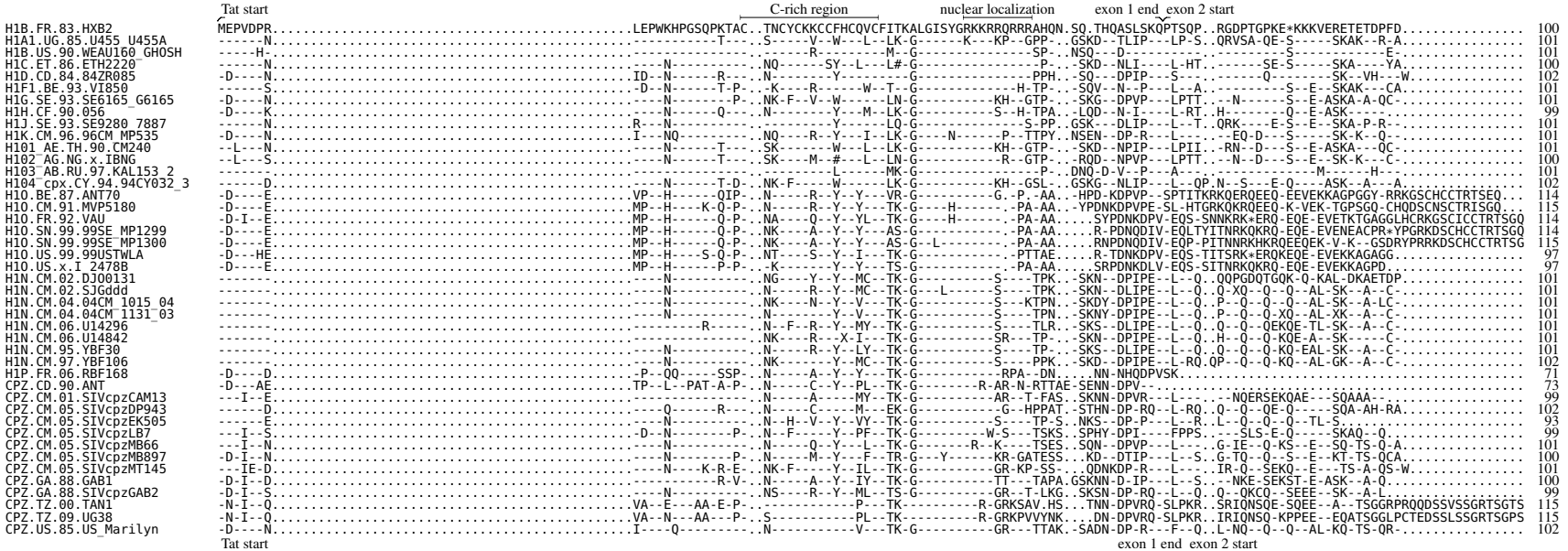
```
MAC.US.x.239                  ..................-  215
H2A.DE.x.BEN                  ..................-  216
H2A.GW.x.ALI                  ..................-  216
H2A.SN.x.ST_HIV_2_ST          ..................-  216
H2B.CI.x.EHO                  ..................-  217
H2B.GH.86.D205_ALT            ..................-  217
H2G.CI.92.Abt96               ..................-  216
H2U.FR.96.12034               ..................-  214
COL.CM.x.CGU1                 ..................-  172
DEB.CM.04.SIVdeb04CMPF3061  HKEGPENDLATSGRRVLS-  251
DEB.CM.99.CM40                ..................-  231
DEB.CM.99.CM5                 ..................-  231
DEN.CD.x.CD1_CMO580407      KNHAERNMEAL.......-  259
DRL.x.x.FAO                   ..................-  218
GOR.CM.04.SIVgorCP684con      ..................-  193
GOR.CM.07.SIVgor2139_287      ..................-  193
GOR.CM.07.SIVgorCP2135con     ..................-  193
GRV.ET.x.GRI_677_gri_1        ..................-  220
GSN.CM.99.CN166               ..................-  238
GSN.CM.99.CN71                ..................-  238
LST.CD.88.SIVlhoest447        ..................-  228
LST.KE.x.lho7                 ..................-  227
MND-1.GA.x.MNDGB1             ..................-  173
MND-2.CM.98.CM16              ..................-  213
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-1.CM.01.SIVmus_01CM1085   ..................-  236
MUS-2.CM.01.CM1246            ..................-  239
MUS-2.CM.01.CM2500            ..................-  248
OLC.CI.97.97CI12              ..................-  204
RCM.CM.00.SIVagi_00CM312      ..................-  221
RCM.CM.02.SIVrcm_02CM8081     ..................-  221
RCM.GA.x.SIVRCMGAB1           ..................-  217
RCM.NG.x.NG411                ..................-  223
SAB.SN.x.SAB1                 ..................-  235
SMM.SL.92.SL92B               ..................-  218
SMM.US.x.SIVsmH635F_L3        ..................-  215
STM.US.89.STM_37_16           ..................-  215
SUN.GA.98.L14                 ..................-  242
SYK.KE.x.KE51                 ..................-  234
SYK.KE.x.SYK173_COMGNM        ..................-  222
TAL.CM.00.266                 ..................-  228
TAL.CM.01.8023                ..................-  228
TAN.UG.x.SIVagmTAN1           ..................-  239
VER.DE.x.AGM3                 ..................-  233
VER.KE.x.9063                 ..................-  232
VER.KE.x.AGM155               ..................-  233
VER.KE.x.TYO1_patent          ..................-  236
WRC.CI.97.97CI14              ..................-  208
WRC.CI.98.98CI04              ..................-  208
WRC.GM.05.Pbt_05GM_X02        ..................-  209
```

```
                           Vpx start                                                                                                    Vpx end
MAC.US.x.239               MSD..PRERIP..PGNSGEETIGEAFEWLNRTVEEINREAVNHLPRELIFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGW.RPGPPPPPPPGLA*   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.GW.x.ALI               -AN..---TV-..----------------D----AL---------------------R---------Q--T-------M---MYT-FM---T---G---P---.-S-----------V-   113
H2A.SN.x.ST_HIV_2_ST       -AG..---T--..----------------D----A----------------------R---------I--T-------M---M-I-S-R--T---G---P---.-S-----------V-   113
H2B.CI.x.EHO               -.-..----V-..--------V-------ET-L-HL--V---------------K--A--RE-----I--T-------M---M-I-FA---G--R----P---.-S-------------   112
H2B.GH.86.D205_ALT         -.-..----V-..----D---V----A--E--IT-L--V------------------A--RE-----I--T-----L-M---M-V-YT------Q----P---.-S-------------   112
H2G.CI.92.Abt96            -.-..----X-..-E--D-----------E--XT----I--------------R-------------A--T-----L-M---M-V------T--QK---P---.-Q-------------   112
H2U.FR.96.12034            -G-..------..----E---V----S--E---RD---A--Q---------------Q--R------E--T--------------------Q-------P---.-S-----------I-   113
DRL.x.x.FAO                -AERQSV--A-AE-MGA--...V-LE---Q-SLLR--Q--RL-FHP-FL-RL-NTCM-HY--ALQL-FT-S-----L-L---M---FQQ--S--QGR-PPPLRPAGDRL------PP.-   114
MAC.US.x.251_1A11          ---..------..------------------------------------------------------------------------------------------.---------------   113
MAC.US.x.251_BK28          ---..------..------------------------------------------------------Q----------M------------------------.---------------   113
MAC.US.x.EMBL_3            ---..------..-----------------------------------------------------X....................................................   63
MND-2.CM.98.CM16           -AE.....GA-EI-EGA--...VDLNT--E-SL-K--Q--RL-FHP-FL-RL-NACI-H---RHQR-L--A-----L-MN--M-T-MQQE-P-RSG.-P......-.-...-----MV-   100
MND-2.GA.x.M14             -AE.....-A-EA-QGA--...VGLEQ--E-SL-QV----QL-FHP-FL-RL-NTCV-H---RLRRTLN-A-----L-M---M-V-MQQ--P-RSG.-S......-.-...--------   100
MND-2.x.x.5440             -AE.....-A-EA-EGA--...VGLEQ--ETSL-R-----RL-FHP-FL-RL-NTCV-H---RHQR-LD-A-----L-MH--MYT-MQQ--P-RNG.RPR.....-.-...-----M--   100
MNE.US.x.MNE027            ---..------..------------------------------------------------------------------------------------------.---------------   113
RCM.CM.00.SIVagi_00CM312   -AE.....-A-..E-PT-AGEVEFQPWLARMLY-INQEA.RL-F-G-F--HL-RTCV--------RTLE-AG----L-M-----T-MRS--KLR...--DPPR.QR-ERV-IL--MQ-   106
RCM.CM.02.SIVrcm_02CM8081  -AE.....-A-..EVPT-AGEVEFQPWLARMLY-INQEA.RL-FHP-F--RL-RTCV-H----L-R-LE-AG----L-M-----T-MRS---LR...--DPPR.QR-ERV-IL--MQ-   106
RCM.GA.x.SIVRCMGAB1        -AE.....-A-..EVPT-AGEAEFQPWLRDMLEKVNLEA.RL-FHP-F--RL-RTCV-H---VHQR-LE-AA----L-M-----I--QT--SQR...--PNPR.AV-ERITIL--M.-   105
RCM.NG.x.NG411             -AE..G---V-..EAPT-AGDVEF-PWLHRMLT-VNLEA.RL-FHP-F--RL-RTCV-H---RL-R-LE-AG----L-M-----I--QS--SQR...--Q-QAREA-ERIQIL--M.-   109
SMM.SL.92.SL92B            -T-..------..----------------HN---AL-QT--Q-----------R-C----V----Y----A----VQ-M---M-Q-FR---T-R----SQ---.-T-------------   113
SMM.US.x.H9                -XX..---X--..-------XX--X-D--H--X-----A-----X--------R-X-------M---V--T-X----------------------GE------.---------------   113
SMM.US.x.PGM53             ---..------..-------------------------A--------------R------R--M---E--T-------------V---R-------E------.-S-------------   113
SMM.US.x.SIVsmH635F_L3     ---..------..---------E-----------G---A------------------------M---E--T--------------------------------.-T-------------   113
STM.US.89.STM_37_16        ---..------..-------A-E------H----D------------------R------------PG------------------S-R--------------.---------------   113
```


oligomerization
amphipathic α-helix
H(S/N)RIG motifs
Vpr start
Vpr end in HXB2
frameshift in HXB2
H1B.FR.83.HXB2
H1A1.UG.85.U455_U455A
H1B.US.90.WEAU160_GHOSH
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165_G6165
H1H.CF.90.056
H1J.SE.93.SE9280_7887
H1K.CM.96.96CM_MP535
H101_AE.TH.90.CM240
H102_AG.NG.x.IBNG
H103_AB.RU.97.KAL153_2
H104_cpx.CY.94.94CY032_3
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
H1O.FR.92.VAU
H1O.SN.99.99SE_MP1299
H1O.SN.99.99SE_MP1300
H1O.US.99.99USTWLA
H1O.US.x.I_2478B
H1N.CM.02.DJO0131
H1N.CM.02.SJGddd
H1N.CM.04.04CM_1015_04
H1N.CM.04.04CM_1131_03
H1N.CM.06.U14296
H1N.CM.06.U14842
H1N.CM.95.YBF30
H1N.CM.97.YBF106
H1P.FR.06.RBF168
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.CM.05.SIVcpzMT145
CPZ.GA.88.GAB1
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.00.TAN1
CPZ.TZ.09.UG38
CPZ.US.85.US_Marilyn
MAC.US.x.239
H2A.DE.x.BEN
H2A.GW.x.ALI
H2A.SN.x.ST_HIV_2_ST
H2B.CI.x.EHO
H2B.GH.86.D205_ALT
H2G.CI.92.Abt96
H2U.FR.96.12034
COL.CM.x.CGU1
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM5
DEN.CD.x.CD1_CMO580407
DRL.x.x.FAO
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139_287
GOR.CM.07.SIVgorCP2135con
GRV.ET.x.GRI_677_gri_1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest447
LST.KE.x.lho7
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-1.CM.01.SIVmus_01CM1085
MUS-2.CM.01.CM1246
MUS-2.CM.01.CM2500
OLC.CI.97.97CI12
RCM.CM.00.SIVagi_00CM312
RCM.CM.02.SIVrcm_02CM8081
RCM.GA.x.SIVRCMGAB1
RCM.NG.x.NG411
SAB.SN.x.SAB1
SMM.SL.92.SL92B
SMM.US.x.SIVsmH635F_L3
STM.US.89.STM_37_16
SUN.GA.98.L14
SYK.KE.x.KE51
SYK.KE.x.SYK173_COMGNM
TAL.CM.00.266
TAL.CM.01.8023
TAN.UG.x.SIVagmTAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TYO1_patent
WRC.CI.97.97CI14
WRC.CI.98.98CI04
WRC.GM.05.Pbt_05GM_X02

	Vpr end	
H1B.FR.83.HXB2	.......*	96
H1A1.UG.85.U455_U455A	.......-	96
H1B.US.90.WEAU160_GHOSH	.......-	97
H1C.ET.86.ETH2220	.......-	97
H1D.CD.84.84ZR085	.......-	97
H1F1.BE.93.VI850	.......-	97
H1G.SE.93.SE6165_G6165	.......-	97
H1H.CF.90.056	.......-	97
H1J.SE.93.SE9280_7887	.......-	97
H1K.CM.96.96CM_MP535	.......-	97
H101_AE.TH.90.CM240	.......-	95
H102_AG.NG.x.IBNG	.......-	97
H103_AB.RU.97.KAL153_2	.......-	97
H104_cpx.CY.94.94CY032_3	.......-	96
H1O.BE.87.ANT70	.......-	98
H1O.CM.91.MVP5180	.......-	101
H1O.FR.92.VAU	.......-	101
H1O.SN.99.99SE_MP1299	.......-	101
H1O.SN.99.99SE_MP1300	.......-	101
H1O.US.99.99USTWLA	.......-	101
H1O.US.x.I_2478B	.......-	101
H1N.CM.02.DJO0131	.......-	96
H1N.CM.02.SJGddd	.......-	96
H1N.CM.04.04CM_1015_04	.......-	96
H1N.CM.04.04CM_1131_03	.......-	96
H1N.CM.06.U14296	.......-	96
H1N.CM.06.U14842	.......-	96
H1N.CM.95.YBF30	.......-	96
H1N.CM.97.YBF106	.......-	96
H1P.FR.06.RBF168	.......-	99
CPZ.CD.90.ANT	.......-	89
CPZ.CM.01.SIVcpzCAM13	.......-	97
CPZ.CM.05.SIVcpzDP943	.......-	96
CPZ.CM.05.SIVcpzEK505	.......-	96
CPZ.CM.05.SIVcpzLB7	.......-	97
CPZ.CM.05.SIVcpzMB66	.......-	99
CPZ.CM.05.SIVcpzMB897	.......-	97
CPZ.CM.05.SIVcpzMT145	.......-	97
CPZ.GA.88.GAB1	.......-	97
CPZ.GA.88.SIVcpzGAB2	.......-	98
CPZ.TZ.00.TAN1	.......-	84
CPZ.TZ.09.UG38	.......-	84
CPZ.US.85.US_Marilyn	.......-	101

	Vpr end	
MAC.US.x.239	.......-	102
H2A.DE.x.BEN	.......-	88
H2A.GW.x.ALI	.......-	106
H2A.SN.x.ST_HIV_2_ST	.......-	105
H2B.CI.x.EHO	.......-	102
H2B.GH.86.D205_ALT	.......-	102
H2G.CI.92.Abt96	.......-	103
H2U.FR.96.12034	.......-	98
COL.CM.x.CGU1	.......-	93
DEB.CM.04.SIVdeb04CMPF3061	.......-	118
DEB.CM.99.CM40	.......-	118
DEB.CM.99.CM5	.......-	118
DEN.CD.x.CD1_CMO580407	.......-	99
DRL.x.x.FAO	.......-	101
GOR.CM.04.SIVgorCP684con	.......-	99
GOR.CM.07.SIVgor2139_287	.......-	99
GOR.CM.07.SIVgorCP2135con	.......-	99
GRV.ET.x.GRI_677_gri_1	.......-	119
GSN.CM.99.CN166	SRYRMDE-	134
GSN.CM.99.CN71	SRYRMDE-	136
LST.CD.88.SIVlhoest447	.......-	115
LST.KE.x.lho7	.......-	115
MND-1.GA.x.MNDGB1	.......-	105
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_99CML1	SRYRMDE-	121
MUS-1.CM.01.CM1239	SRYRMDE-	136
MUS-1.CM.01.SIVmus_01CM1085	SRYRMDE-	136
MUS-2.CM.01.CM1246	SRYRMDE-	136
MUS-2.CM.01.CM2500	SRYRMDE-	136
OLC.CI.97.97CI12	.......-	97
RCM.CM.00.SIVagi_00CM312	.......-	101
RCM.CM.02.SIVrcm_02CM8081	.......-	101
RCM.GA.x.SIVRCMGAB1	.......-	101
RCM.NG.x.NG411	.......-	101
SAB.SN.x.SAB1	.......-	141
SMM.SL.92.SL92B	.......-	101
SMM.US.x.SIVsmH635F_L3	.......-	102
STM.US.89.STM_37_16	.......-	102
SUN.GA.98.L14	.......-	117
SYK.KE.x.KE51	.......-	116
SYK.KE.x.SYK173_COMGNM	.......-	114
TAL.CM.00.266	.......-	116
TAL.CM.01.8023	.......-	115
TAN.UG.x.SIVagmTAN1	.......-	120
VER.DE.x.AGM3	.......-	120
VER.KE.x.9063	.......-	120
VER.KE.x.AGM155	.......-	120
VER.KE.x.TYO1_patent	.......-	120
WRC.CI.97.97CI14	.......-	117
WRC.CI.98.98CI04	.......-	117
WRC.GM.05.Pbt_05GM_X02	.......-	118

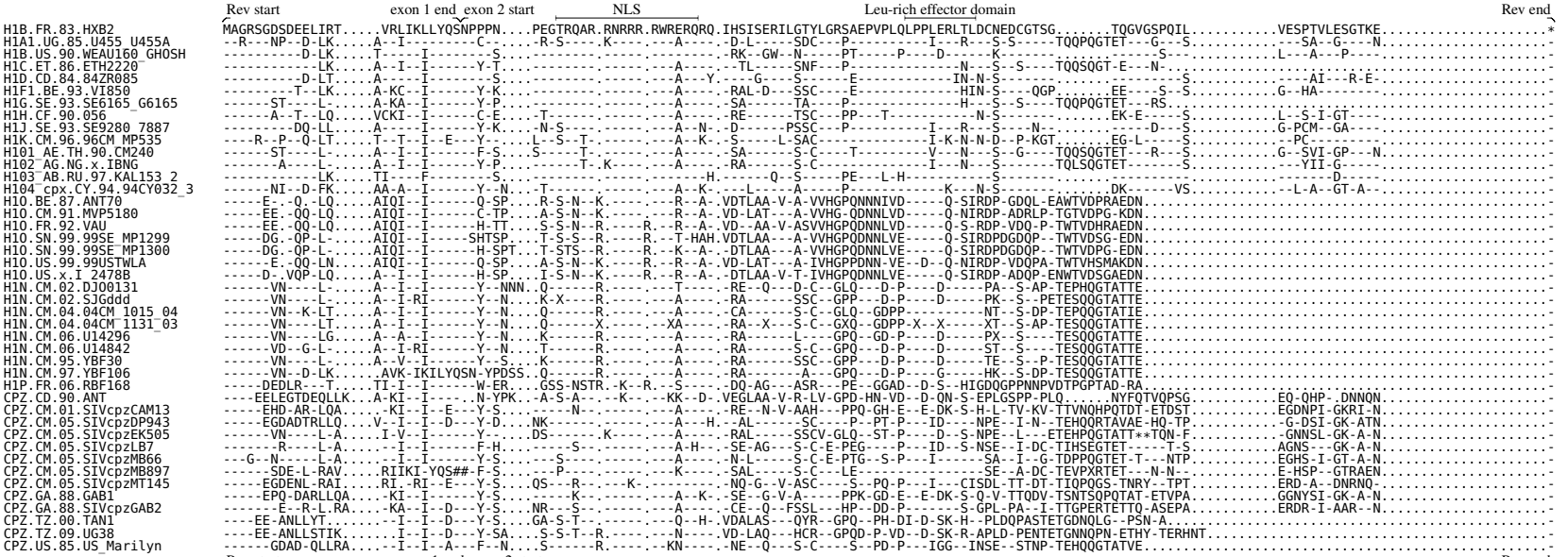

Tat start
C-rich region
nuclear localization
exon 1 end
exon 2 start
H1B.FR.83.HXB2
H1A1.UG.85.U455_U455A
H1B.US.90.WEAU160_GHOSH
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165_G6165
H1H.CF.90.056
H1J.SE.93.SE9280_7887
H1K.CM.96.96CM_MP535
H101_AE.TH.90.CM240
H102_AG.NG.x.IBNG
H103_AB.RU.97.KAL153_2
H104_cpx.CY.94.94CY032_3
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
H1O.FR.92.VAU
H1O.SN.99.99SE_MP1299
H1O.SN.99.99SE_MP1300
H1O.US.99.99USTWLA
H1O.US.x.I_2478B
H1N.CM.02.DJO0131
H1N.CM.02.SJGddd
H1N.CM.04.04CM_1015_04
H1N.CM.04.04CM_1131_03
H1N.CM.06.U14296
H1N.CM.06.U14842
H1N.CM.95.YBF30
H1N.CM.97.YBF106
H1P.FR.06.RBF168
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.CM.05.SIVcpzMT145
CPZ.GA.88.GAB1
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.00.TAN1
CPZ.TZ.09.UG38
CPZ.US.85.US_Marilyn
Tat start
exon 1 end
exon 2 start
MAC.US.x.239
H2A.DE.x.BEN
H2A.GW.x.ALI
H2A.SN.x.ST_HIV_2_ST
H2B.CI.x.EHO
H2B.GH.86.D205_ALT
H2G.CI.92.Abt96
H2U.FR.96.12034
COL.CM.x.CGU1
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM5
DEN.CD.x.CD1_CMO580407
DRL.x.x.FAO
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139_287
GOR.CM.07.SIVgorCP2135con
GRV.ET.x.GRI_677_gri_1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest447
LST.KE.x.lho7
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-1.CM.01.SIVmus_01CM1085
MUS-2.CM.01.CM1246
MUS-2.CM.01.CM2500
OLC.CI.97.97CI12
RCM.CM.00.SIVagi_00CM312
RCM.CM.02.SIVrcm_02CM8081
RCM.GA.x.SIVRCMGAB1
RCM.NG.x.NG411
SAB.SN.x.SAB1
SMM.SL.92.SL92B
SMM.US.x.SIVsmH635F_L3
STM.US.89.STM_37_16
SUN.GA.98.L14
SYK.KE.x.KE51
SYK.KE.x.SYK173_COMGNM
TAL.CM.00.266
TAL.CM.01.8023
TAN.UG.x.SIVagmTAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TYO1_patent
WRC.CI.97.97CI14
WRC.CI.98.98CI04
WRC.GM.05.Pbt_05GM_X02

Tat end

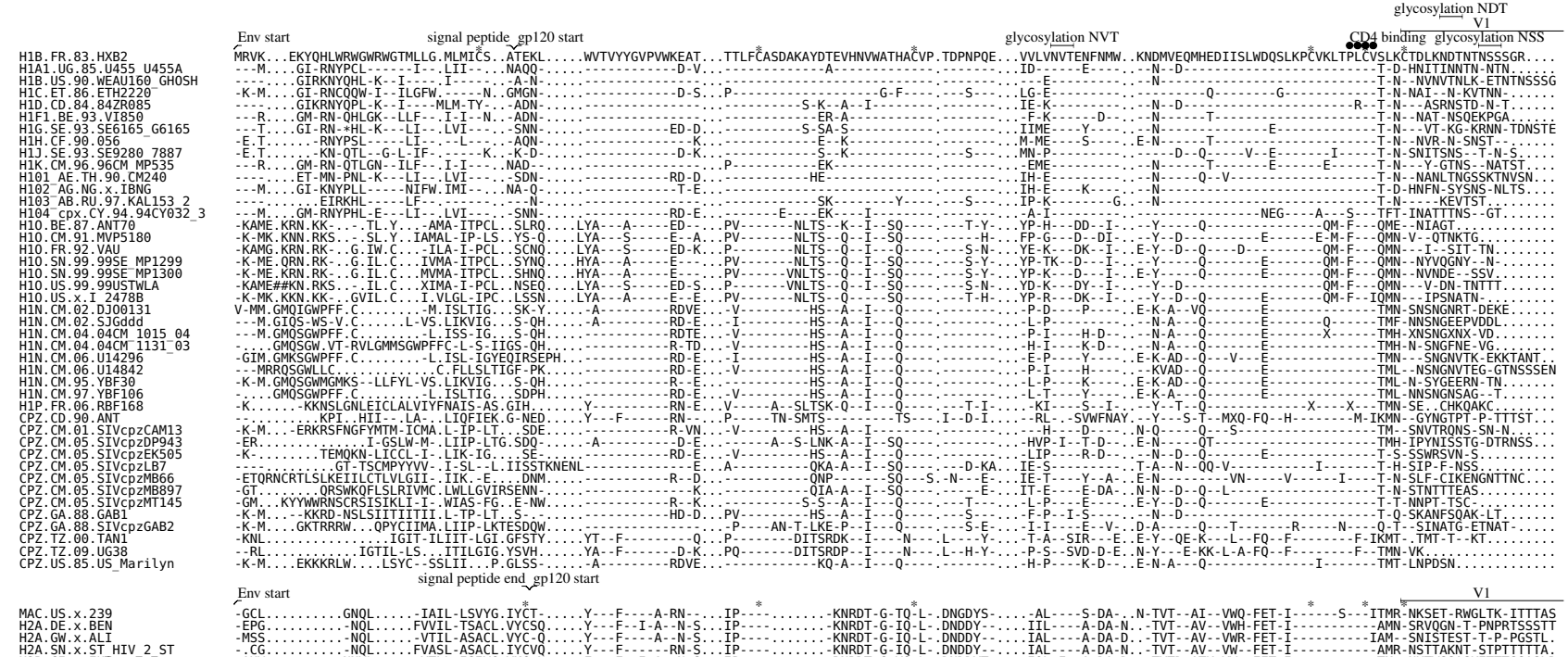
H1B.FR.83.HXB2	.....*	100
H1A1.UG.85.U455_U455A	.....	102
H1B.US.90.WEAU160_GHOSH	.....	102
H1C.ET.86.ETH2220	.....	101
H1D.CD.84.84ZR085	.....	103
H1F1.BE.93.VI850	.....	102
H1G.SE.93.SE6165_G6165	.....	102
H1H.CF.90.056	.....	100
H1J.SE.93.SE9280_7887	.....	102
H1K.CM.96.96CM_MP535	.....	102
H101_AE.TH.90.CM240	.....	102
H102_AG.NG.x.IBNG	.....	101
H103_AB.RU.97.KAL153_2	.....	102
H104_cpx.CY.94.94CY032_3	.....	103
H1O.BE.87.ANT70	.....	115
H1O.CM.91.MVP5180	.....	116
H1O.FR.92.VAU	.....	115
H1O.SN.99.99SE_MP1299	.....	115
H1O.SN.99.99SE_MP1300	Q.....	117
H1O.US.99.99USTWLA	.....	98
H1O.US.x.I_2478B	.....	98
H1N.CM.02.DJO0131	.....	102
H1N.CM.02.SJGddd	.....	102
H1N.CM.04.04CM_1015_04	.....	102
H1N.CM.04.04CM_1131_03	.....	102
H1N.CM.06.U14296	.....	102
H1N.CM.06.U14842	.....	102
H1N.CM.95.YBF30	.....	102
H1N.CM.97.YBF106	.....	103
H1P.FR.06.RBF168	.....	72
CPZ.CD.90.ANT	.....	74
CPZ.CM.01.SIVcpzCAM13	.....	100
CPZ.CM.05.SIVcpzDP943	.....	103
CPZ.CM.05.SIVcpzEK505	.....	94
CPZ.CM.05.SIVcpzLB7	.....	100
CPZ.CM.05.SIVcpzMB66	.....	102
CPZ.CM.05.SIVcpzMB897	.....	101
CPZ.CM.05.SIVcpzMT145	.....	102
CPZ.GA.88.GAB1	.....	101
CPZ.GA.88.SIVcpzGAB2	.....	100
CPZ.TZ.00.TAN1	SSGYTRPFKTSSGSSGSACKH-	137
CPZ.TZ.09.UG38	SCGSTRPVKTSAGSSGRS...-	134
CPZ.US.85.US_Marilyn	.....	103

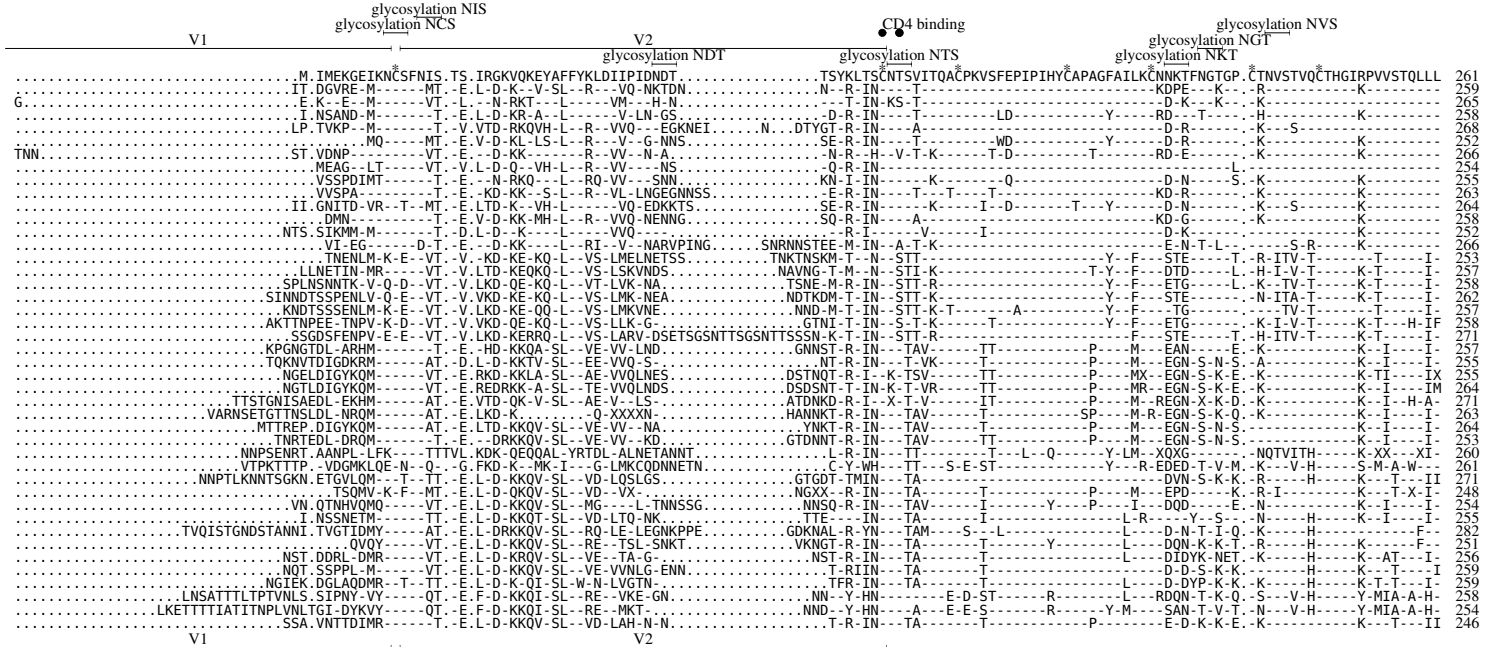
Tat end

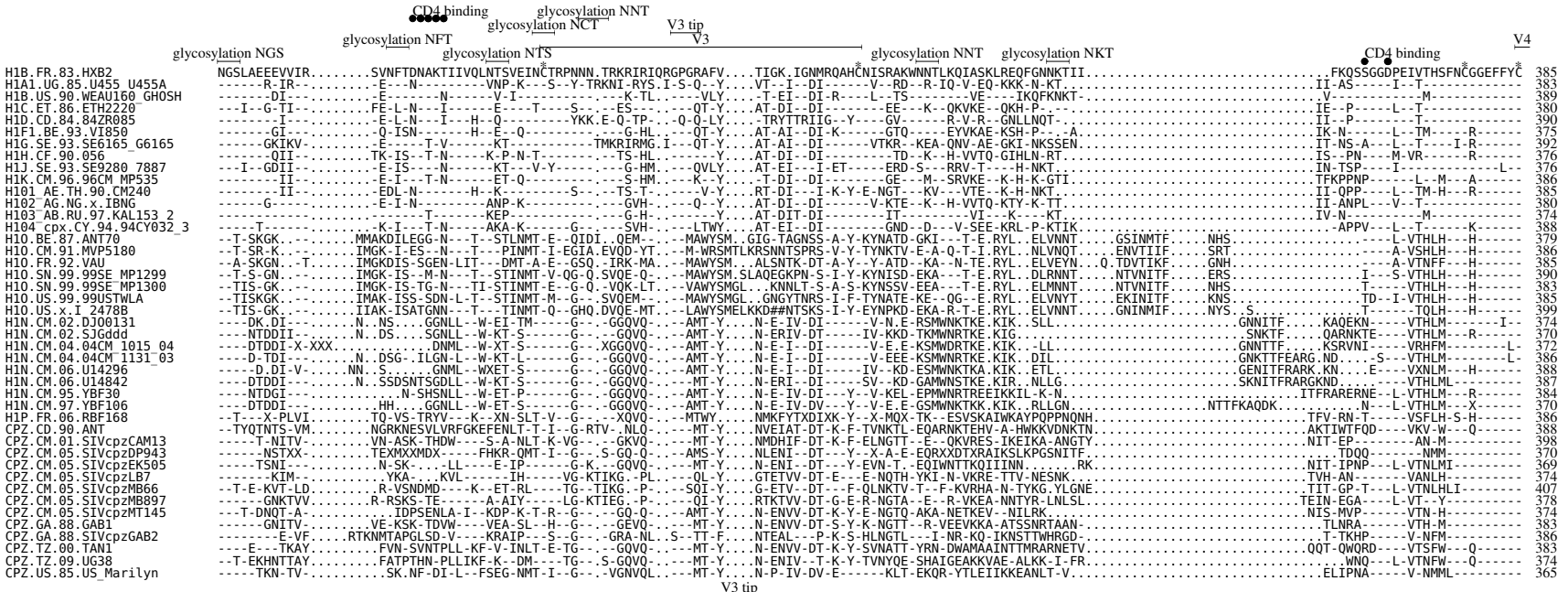
MAC.US.x.239	.....	131
H2A.DE.x.BEN	.....	131
H2A.GW.x.ALI	.....	137
H2A.SN.x.ST_HIV_2_ST	.....	131
H2B.CI.x.EHO	.....	139
H2B.GH.86.D205_ALT	.....	129
H2G.CI.92.Abt96	.....	129
H2U.FR.96.12034	.....	128
COL.CM.x.CGU1	.....	80
DEB.CM.04.SIVdeb04CMPF3061	.....	99
DEB.CM.99.CM40	C.....	116
DEB.CM.99.CM5	.....	102
DEN.CD.x.CD1_CMO580407	.....	100
DRL.x.x.FAO	.....	119
GOR.CM.04.SIVgorCP684con	.....	83
GOR.CM.07.SIVgor2139_287	.....	83
GOR.CM.07.SIVgorCP2135con	.....	83
GRV.ET.x.GRI_677_gri_1	.....	95
GSN.CM.99.CN166	.....	151
GSN.CM.99.CN71	.....	135
LST.CD.88.SIVlhoest447	.....	111
LST.KE.x.lho7	.....	112
MND-1.GA.x.MNDGB1	.....	108
MND-2.CM.98.CM16	.....	113
MND-2.GA.x.M14	.....	101
MND-2.x.x.5440	.....	101
MNE.US.x.MNE027	.....	131
MON.CM.99.L1_99CML1	.....	104
MUS-1.CM.01.CM1239	.....	102
MUS-1.CM.01.SIVmus_01CM1085	.....	101
MUS-2.CM.01.CM1246	.....	109
MUS-2.CM.01.CM2500	.....	102
OLC.CI.97.97CI12	.....	108
RCM.CM.00.SIVagi_00CM312	.....	116
RCM.CM.02.SIVrcm_02CM8081	.....	113
RCM.GA.x.SIVRCMGAB1	.....	106
RCM.NG.x.NG411	.....	116
SAB.SN.x.SAB1	.....	107
SMM.SL.92.SL92B	.....	105
SMM.US.x.SIVsmH635F_L3	.....	129
STM.US.89.STM_37_16	.....	129
SUN.GA.98.L14	.....	103
SYK.KE.x.KE51	.....	99
SYK.KE.x.SYK173_COMGNM	.....	111
TAL.CM.00.266	.....	111
TAL.CM.01.8023	SSATVTRP.....	128
TAN.UG.x.SIVagmTAN1	.....	102
VER.DE.x.AGM3	SD.....	120
VER.KE.x.9063	NA.....	120
VER.KE.x.AGM155	GN.....	120
VER.KE.x.TYO1_patent	.....	101
WRC.CI.97.97CI14	.....	132
WRC.CI.98.98CI04	.....	132
WRC.GM.05.Pbt_05GM_X02	.....	129

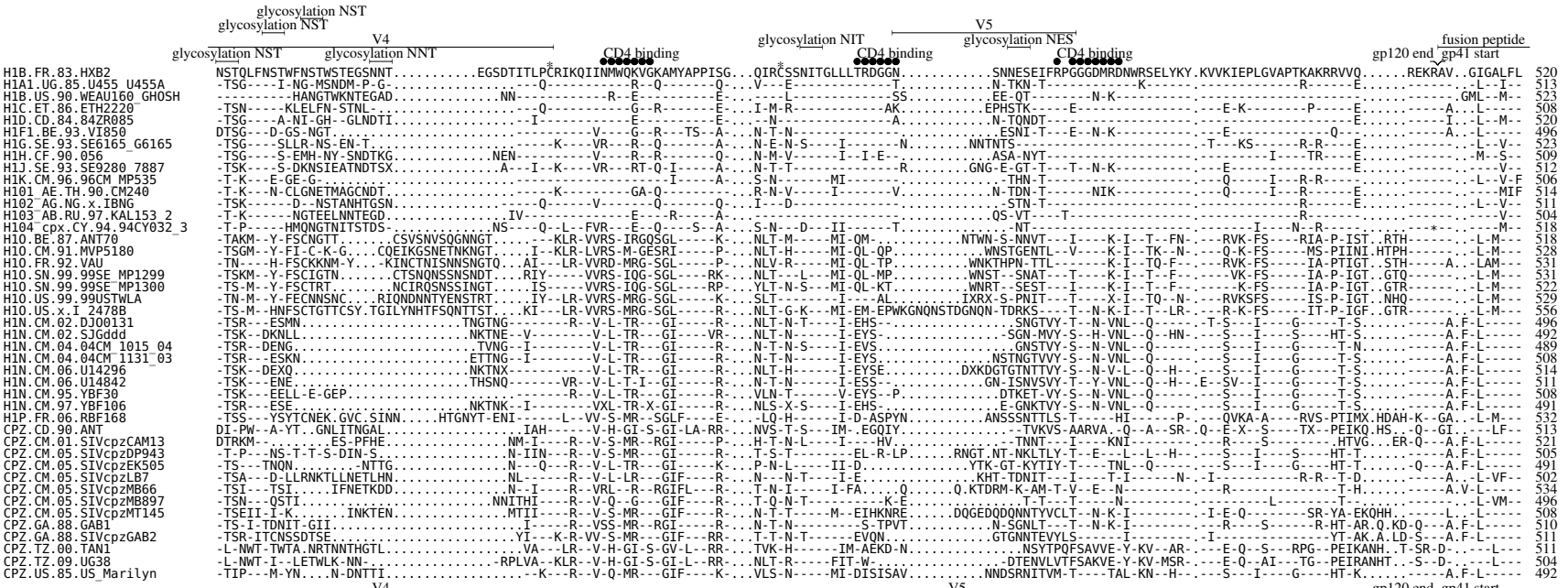

Rev start
exon 1 end
exon 2 start
NLS
Leu-rich effector domain
Rev end
H1B.FR.83.HXB2
H1A1.UG.85.U455_U455A
H1B.US.90.WEAU160_GHOSH
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165_G6165
H1H.CF.90.056
H1J.SE.93.SE9280_7887
H1K.CM.96.96CM_MP535
H101_AE.TH.90.CM240
H102_AG.NG.x.IBNG
H103_AB.RU.97.KAL153_2
H104_cpx.CY.94.94CY032_3
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
H1O.FR.92.VAU
H1O.SN.99.99SE_MP1299
H1O.SN.99.99SE_MP1300
H1O.US.99.99USTWLA
H1O.US.x.I_2478B
H1N.CM.02.DJO0131
H1N.CM.02.SJGddd
H1N.CM.04.04CM_1015_04
H1N.CM.04.04CM_1131_03
H1N.CM.06.U14296
H1N.CM.06.U14842
H1N.CM.95.YBF30
H1N.CM.97.YBF106
H1P.FR.06.RBF168
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.CM.05.SIVcpzMT145
CPZ.GA.88.GAB1
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.00.TAN1
CPZ.TZ.09.UG38
CPZ.US.85.US_Marilyn
Rev start
exon 1 end
exon 2 start
Rev end
MAC.US.x.239
H2A.DE.x.BEN
H2A.GW.x.ALI
H2A.SN.x.ST_HIV_2_ST
H2B.CI.x.EHO
H2B.GH.86.D205_ALT
H2G.CI.92.Abt96
H2U.FR.96.12034
COL.CM.x.CGU1
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM5
DEN.CD.x.CD1_CMO580407
DRL.x.x.FAO
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139_287
GOR.CM.07.SIVgorCP2135con
GRV.ET.x.GRI_677_gri_1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest447
LST.KE.x.lho7
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-1.CM.01.SIVmus_01CM1085
MUS-2.CM.01.CM1246
MUS-2.CM.01.CM2500
OLC.CI.97.97CI12
RCM.CM.00.SIVagi_00CM312
RCM.CM.02.SIVrcm_02CM8081
RCM.GA.x.SIVRCMGAB1
RCM.NG.x.NG411
SAB.SN.x.SAB1
SMM.SL.92.SL92B
SMM.US.x.SIVsmH635F_L3
STM.US.89.STM_37_16
SUN.GA.98.L14
SYK.KE.x.KE51
SYK.KE.x.SYK173_COMGNM
TAL.CM.00.266
TAL.CM.01.8023
TAN.UG.x.SIVagmTAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TYO1_patent
WRC.CI.97.97CI14
WRC.CI.98.98CI04
WRC.GM.05.Pbt_05GM_X02

```
                                                                              phos phos
                                  transmembrane domain                              cytoplasmic domain
                           Vpu start                                   α-helix                        α-helix                      Vpu end
H1B.FR.83.HXB2             T..QPIPIVA....IVALVVAI.IIAIVVWSIV....IIEYRK.ILRQRKIDRLIDRLIERAEDSGNESEG.......EIS..............ALVEMGV.EMGHH.APWDVDDL*   82
H1A1.UG.85.U455_U455A      M..T-LE-W-....-TG-I--L.-L-----T-.....G--#K-KL-K-K-----LN-IR----------D-.......DTEEL...........SL---------NY.DLGVDNN--   82
H1B.US.90.WEAU160_GHOSH    M..-SLQ-L-...-------G.-------------.L------.--------------IRD------------.......DQEEL...........S-----------------I----   82
H1C.ET.86.ETH2220          M.VDLLAK-DYRIV---FI--L.-------T-A....Y-----.L----R-----K-TR----------D-.......DTEEL...........STM-D--.....NL.RLL--N---   87
H1D.CD.84.84ZR085          M..-SLQ-L-...------L-L.-------T--....F----R.-K------W----IR--E----------.......DKEEL...........ST-----------APWDVD----   83
H1F1.BE.93.VI850           M..SYLLAIG....-A--I--L.-------T--....Y---K-.LV-----NK-YK-IR-------------.......DAEEL...........A--G--.......PFI-G-INN--   82
H1G.SE.93.SE6165_G6165     M..-SLV-L-....--G-I--F.-A-----T--....F----E.-RK-KR-GK-L--IR----------D-.......DTEEL...........VT-----D.....FD--VG-N--   82
H1H.CF.90.056              M...Y-LGLG....-G----TF.---VI--T--....Y---K-.LV--K------E-IG----------D-.......DTEEL...........SK-M---.......-LNLGY-A---   81
H1J.SE.93.SE9280_7887      M..I-LQ-A-....----FI---.FL--GM-T--...Y---K-.L-------K----IR----------D-.......DTEEL...........AD---R-.......PHDL-N-N---   82
H1K.CM.96.96CM_MP535       M...VSLAIS....--------L.-L--I--T--...Y-----.LVK-KR-NW----IR-------------.......DAEEL...........ADIG-L-.......-LILGNI-N--   81
H101_AE.TH.90.CM240        M..T-LE-S-....--G-I--L.-L----*T--....A--VK-.----------VK-IR-------------.......DTDEL...........AK---------.DFD--VG-N--   81
H102_AG.NG.x.IBNG          M..--LT-T-....--G----F.-A-----------.Y-----.-RK-K-----L--IR----------D-.......DTEEL...........ST-M---.......YEYIL-N-N--   82
H103_AB.RU.97.KAL153_2     M..-SLA-A-...---------VG.------G----..F-------------------IR-------------.......DQE...........--M---.......-LV---A----   79
H104_cpx.CY.94.94CY032_3   M..LFWE-W-....--G----L.--V----TL-....F---K-.LR---R--S-YN-IR----------D-.......DAEEL...........ST--G--.......NFD--VG-N--   82
H1O.BE.87.ANT70            M...HHRDLL...A-IIISALL.F-NVIL-GFILRK.YL-QKEQDRKE-E-LERLR-IR-IRD--DY--N-E......-EQEVM...........DLVLSH-........FDNPMFEP-   86
H1O.CM.91.MVP5180          M...HQENLL...ALI--SALC.L-NVLI-LFNLRI.YLVQ--QDR-EQE-LERLR-IK-IRD--DY--NEE......-QQEVM...........ELIHSH-........FANPMFE--   86
H1O.CM.94.BCF06            M...HQRDLL...A-I--SALC.L-S-I--I-NLRN.YL-Q--QDR-EQE-LERLR-VR-IKD--DYD-N-E......-EQEVM...........DLMHSH-........FDNPMFEW-   86
H1O.CM.96.96CMA102         M...HYRDLS...TLIIVSALL.L-NVXL-MFILRX.YL-HKRQER-E-E-LERLR-IR-IKD--DY--N-E......-EQEVM...........DLVHSH-........FDNPMFE--   86
H1O.CM.96.96CMABB009       M...YKDQII...L-IIFC-VF.L--ACI-LFILKT.YX-QK-QDR-EKELL-RLQ-I--IRD--DY--N-E......-EQEVM...........DLVH-H-........FVNPMFE--   86
H1O.CM.96.96CMABB637       M...NYKELL...SLIVVS-LL.LA---I-MFILKK.YL-QKEQDR-E-ELLKR-E--X-XRD--DY--N-D......-EQEVM...........HLVHTH-........FANPMFE--   86
H1O.CM.97.97CMABB447       M...HYRDLL...ILIVVSALL.L-N--I-TFILKK.YL-QK-QDR-E-ELLNR-K-IR-VRD--DY--N-E......-EQXVM...........HLVHTH-........FANPMFEF-   86
H1O.CM.97.97CMABB497       M...QYKGLL...L-II..ALL.L-NVX--MFNLRK.YL-QK-QER-E-EVINRLR-IR-VKD--DY--N-E......-EQEVM...........ELVHSH-........FDNPMFE--   84
H1O.CM.98.98CMA104         M...HYRDLL...ALIIVSALL.L-NVLL--FILRK.YL-QK-QDR-E-E-LERLRKIR-IKD--DY--NEE......-EQEVM...........DLIHSH-........FDNPMFE--   86
H1O.CM.98.98CMA105         M...HHRDLL...TLIIGSALL.L-NVLL-MFMLRK.YL-QKEQDR-EXE-LERLR-IX-IRD--DY--N-E......-EQEVM...........XLVLSH-........FDNPMFEP-   86
H1O.CM.98.98CMABB141       M...LHRDLL...LLIIISALL.LTN-IL-MF-LRK.YL-IK-QER-E-E-LERLRXIR-IRD--DY--NEE......-EQEVRD..........HLVHTF-........FANPMFEI-   87
H1O.CM.98.98CMABB197       M...HHRDLL....TLIA-S-LLF-N-IL-IY-LRK.YL-Q--QDR-E-E-LERLR-IX-IGD--DY--NEE......-EQEVM...........DLVHNH-........FDNPMFEP-   86
H1O.CM.98.98CMABB212       M...HQRDLL....-LIA-SILCL-C-L--TFNLRK.YL-H--QDK-E-E-LERLR-VR-IRD--DY--X-E......-EQEXM...........DLIHSH-........FANPLFE--   86
H1O.CM.98.98CMU2901        M...QHRDLL....-II..S-LL--NVIL-MY-LRQ.CL-HK-QAR-E-E-IDRLR-IR-IRD--DY--N-E......-EQEVR...........DLVHSY-........FDNPMFE--   84
H1O.CM.98.98CMU5337        M...HHRDLL....XXIIIS-LLX-NVIL-MFILRQ.XXKXX-QDR-E-E-LERLR-IR-IKD--DY--NEE......XEQEVR...........DLVHSH-........FDNPMFE--   86
H1O.CM.99.99CMU4122        M...QHKDLL....-LIITS-LLL-NVIL-LF-LKQ.CL-QK-QTK-E-E-I-RLR-IR-IED--DY--N-E......-EQTVR...........DLIHSH-........FDNPMFE--   86
H1O.CM.x.pCMO2_3           M...QHRDLI....TLIIIS-LLL-N-IL-MFILKQ.C--QK-QSR-E-E-IERLR-IRKIQD--DY--N-E......-EQEVR...........DLVHSH-........FDNPMFEP-   86
H1O.CM.x.pCMO2_5           M...QHRDLI....TLIIIS-LLL-N-IL-MFILKQ.C--QK-QSR-E-E-IERLR-IRKIQD--DY--N-E......-EQEVR...........DLVHSH-........FDNPMFEP-   86
H1O.FR.92.VAU              M...HHQGLL....VLIVII-LLL-NVLL-MFNLRA.YL-QK-QDRTA-E-LER-R-IR-VED--DY--N-E......-EQEVR...........DLVLDF-........FDNPMFE--   86
H1O.SN.99.99SE_MP1299      M...HHRDLL....ALITTS-LLLTNV-L-TFILRQ.YLKQK-QDK-E-E-LERLR-IRQIED--DY--D-T......-EQEVR...........DLVHSY-........FDNPMFE--   86
H1O.SN.99.99SE_MP1300      M...HHRDLL....TLIIIS-LLLTNVIL-AFILRQ.YLRQK-QDR-E-E-LERLR-IRQIED--DY--D-K......-EQEVR...........DLVHGY-........FDNPMFEP-   86
H1O.US.99.99USTWLA         M...QYKDLL....FLLAYS-LFLTTALI-MF-LKL.CL-QKRQGK-E-E-LDRLR-IXXIRD--DY--N-X......-EQEVM...........DLVHXH-........FDNPMFE--   86
H1O.US.x.I_2478B           M...HHRDLL....LLIIIS-LLL-N-IL-MFILRK.YL-QKRQDR-E-E-LERLR-IR-IRD--DY--NEE......-EQEVRG..........HLVHSF-........FDNPMFE--   87
H1N.CM.02.DJO0131          M....LE.WG....FA--G-------VII-VLL....YK--K-.LKL-E--EQIRQ-IRD-T----S-ND-.......DADWL...........AI-LSHD.......KFD..NWV.-   75
H1N.CM.02.SJGddd           M....LF.LG....FL--GL-L.---VII-VLL....YR--KE.-KL-ER-KHIRQ-IRD---------D-.......DTELL...........VTRLSLS.......KLDQGDWV.-   77
H1N.CM.04.04CM_1015_04     M....LS.LG....FS--G---.---VII-VLL....YK--KE.-KL-E--EQIKQ-VRD---------D-.......DAEWL...........AI-LSPD.......KLD..NWV.-   75
H1N.CM.04.04CM_1131_03     M....LE.LG....FI-FG---.---VII-VLL....YK--KE.-KL-E--EKIRQ-IRD--KE--K--D-.......DAEWLGDKESVGDDEWLAI-LSPD.......KLD..NWV.-   87
H1N.CM.06.U14296           M....LQSLG....FI--GL--.---V-I-VLI....YR--K-.-KL-GR-KTI.............GNN-.......NEEWL...........AI-LSPD.......KLD..NWV.-   66
H1N.CM.06.U14842           M....LF.LG....FI--G---.---V-I-VLL....YK--K-.-KL-EE-KQ-RQ-IRS-----D---DE.......-AEWV...........AT-LFPT.......KLD..NWV.-   75
H1N.CM.95.YBF30            M....LS.LG....FI--GA-V.S--VI--ALL....YR--K-.-KL-E---KHIRQ-IR--E-------D-DAEWLDGDEEWL...........VT-LSSS.......KLDQGNWV.-   84
H1N.CM.97.YBF106           M....LW.LG....FI--G---.---AII-VLL....YK--K-.-KL-E--EQIRQKIRD-T--R-K--D-.......DAEWL...........AI-LSPD.......KLD..NWV.-   75
H1P.FR.06.RBF168           M...HPRDE-....VLIIAGVLL-CI----GK-LLF.VLK....ERERD-FVQRLV-WR--Q--G-Y--NEE......-EEQLR...........ELGNLL-.......FGNIL...-   79
CPZ.CD.90.ANT              M..TN-FEY-....FL-FSIVL.W-ICIPILYKLYKIYKQQQIDNK-NQR-IEVLS-..RLSI--AI-EDE.......-ADTY.............YLGS-.....FANPVYREG-E-   84
CPZ.CM.01.SIVcpzCAM13      M...ILIALG.....CLAIAL-.LLN-FI-RNL....WRLCKQ.NKLEVE-EN-AL--T----------D-G......-EEKL...........QR-LHDHD...NM-GFANPLF-I-   84
CPZ.CM.05.SIVcpzDP943      M..LTWEQIG....LI--GIEG.---T---G-AFIT.WRRRKIEED-AQ--ID-LE-IRA-QD-------D-E......DQRQL..............MAY-........FDNPMF-W-   82
CPZ.CM.05.SIVcpzEK505      M..LLLIKLG....FIG-AIET.L-V----A--....YRI--E.VKVEE--SQ-RQ-IRD---------D-.......DAEEL...........AN-LPPD........RIDQDNWV-   80
CPZ.CM.05.SIVcpzLB7        M..D.LIELG....LIG--IEL.--V----LKA....YQL-KE.NI--KA-NK--E-IR--G-------------.......DMDEL...........H-ILRS-D.......PELVLI-N-   80
CPZ.CM.05.SIVcpzMB66       M..DIVQQ-GL...L-V-IIEL.V-V--I-VK-....YKLCKE.DR--K------A-IR----------D-.......DTEEL...........QD-ITE-DN..LM-IGIR-NRNN-   87
CPZ.CM.05.SIVcpzMB897      M..EIFI-LG....LIGI-IEL.V------LKA....YEC-KA.LK--ERR-Q----IR----------D-.......DTEEL...........E-ILIPED.....RHVLVAIRGY-   83
CPZ.CM.05.SIVcpzMT145      M...QLE--L....-ILFIALM.LV--FA-IAA....YK--K-.LQQV-R-E--Q--IRS---------D-.......DEILL..........VEE-MQVHQH.QVNPDWMDRILFW-   87
CPZ.CM.98.CAM3             M..LTWEQIG....LIGIGIE-.------A-G-AFKE.WRKGKEEEN-AQ--RT--E-I-S-Q-------D-E......DQRQLD..........REIHVY-........FDNPMF-W-   86
CPZ.CM.98.CAM5             M..LIWEQIG....LI--GIEL.--V----G-AYKK.WR-IKEERN-TQR-YN-AE-IRI-Q-------N-E......-EAQLD..........NL-HAN-........FDNPMF-WP-   87
CPZ.GA.88.GAB1             M..TLLVGLV....LILVGLIA.WNICIWGY-IKWG.YRR-KR.HRLETE-E--NLI-R----------N-E......-EERL...........EQ-IHNYN...HNNHFANPMF---   89
CPZ.GA.88.SIVcpzGAB2       M.LSMWVAIG....LIGIGTLL.V-N----G--GIS.VYKRW-RHKEEQR-ID--IKIT----------D-E......DKETLA..........TL-HNN-......FDN-MFE-RI-   89
CPZ.TZ.00.TAN1             MIKIVVGS-S....TNVIGILC.-LL-LIGGGLLIG.-GIR-ELERERQHQRV-ERLARRLSI---V-EDEE......FNWNN..............FDPH.......NYNPRDWI.-   84
CPZ.TZ.01.TAN2             MVKLVVGS-L....TNVIGIFC.-LL-LIGGGLLII.T-IK-EIERERQHQRI-ERLIRRLSI---V-EDEE......LNWNN..............FDPH.......NYNPRDWI.-   84
CPZ.TZ.02.TAN3_1           MVKIVVGS-L....TNVIGAFC.-LL-LIGGGLLII.AFVR-ELERERQHQRVIERLVRRLSI---IDEDEE......LNWNN..............FDPH.......NFNPRDWI.-   84
CPZ.TZ.06.TAN5             MVKLVVGSLL....TNVIGIFC.-LL-LIGGSLLIV.--IK-ELRRTRQQQRV-ERLIRRLSI---I-EDEE......LNWNN..............FDPH.......NYNPRDWI.-   84
CPZ.TZ.09.UG38             M.RLVGGSII....QNVVGILF.--VVI-GGGALIG.WGIR-EIHRERLHQRV-ERLTRRLSI---I-EDEE......INWNN..............FDPQ.......HYNPRDWI.-   83
CPZ.US.85.US_Marilyn       M.L.NWFEIG....LI--GIEG.-LVVII-GL-ARL.WRQIKIKENT-QE-QN-LE-IRI-E-------D-E......-EETLA...........KL-SSLE.......LDNPRIV.-   85
DEN.CD.x.CD1_CMO580407     M...LKIKLG.....PIEY-CLVFAVVIT-IAAIGVGYLA--AYKSYREELRYI..--RLWSI---Y--SQE......DP.........................................-   65
GOR.CM.04.SIVgorCP684con   M...HPRDIL....VIIIGIILLAVTVIS-LKALALYLRDR-E.....-RFFDRLE---SNK--E-Y--NEE......-AAELM...........ERGN-L-........FAFNLH..-   80
GOR.CM.07.SIVgor2139_287   M...HPRDII....VIIIGITLLAVTVII-LKIFALYLRDR-E.....-RFFDRLE--LSNK--E-Y--NEE......-AAELM...........EMGN-L-........FDFNLH..-   80
GOR.CM.07.SIVgorCP2135con  M...HPRDII....VIIIGITLLAATVI--LKALALYLRDR-E.....-RFFNRLE--LSTK--E-Y--NEE......-AAELM...........ERGN-L-........FDFNLH..-   80
GSN.CM.99.CN166            MHPAAVWWWG......-AIITFIYLCVALLA..LYLAWDKWV-GKPKPTQVAVIR.LIED.E----IYDDAS......SEL.................TGFN.......GFANPGFEV-   79
GSN.CM.99.CN71             MSAAALWWWG......-A-ITFIYFCLAIFA..LYLAWDKWI-GKPK.IPVAVIR.LVED.D-E--IFEDAS......SEP.................NAY........GFANPGFEV-   77
MON.CM.99.L1_99CML1        M..NYWWSLV....AITYSLILIALPVAA-AWWRY..YKITKRFKRIDQE-Q---QIHER-RH---VDT-SE......SEQHE...............ETH-..FVNPVFNDDFGEWV-   88
MUS-1.CM.01.CM1239         MN...YWYL-.......-AL-TGIYFL-ALFA..FVLAYQRWCRP...KVEVSVIR.L-E-GDS---IFEDAE......DEPNG...............DGHH.......AFANPAFEQ-   76
MUS-1.CM.01.SIVmus_01CM1085 MN...YWYL-.......-VI-TGIYFV-AIFA..FVLAYQRWC-P..KKVEVSVIR.L-E-GDG---IFEDAE......DDMAE...............S-HH.......AFANPAFEQ-   77
MUS-2.CM.01.CM1246         MH...YWYLG.......-AI-TAIYLV-ALVA..FILAYQRWCQP.KPPIEVNV-R.L-E-GDS---IFEDAC......DGEDE...............DSHR.......AFANPSFEP-   78
MUS-2.CM.01.CM2500         MN...WWWF-.......-A--TAIYFV-ALVA..FVLAYQRWCQPQKG-VEVNVIR.L-E-GDT---IFEDAE......DG-DD...............QRH........GFLNPAFE--   78
```

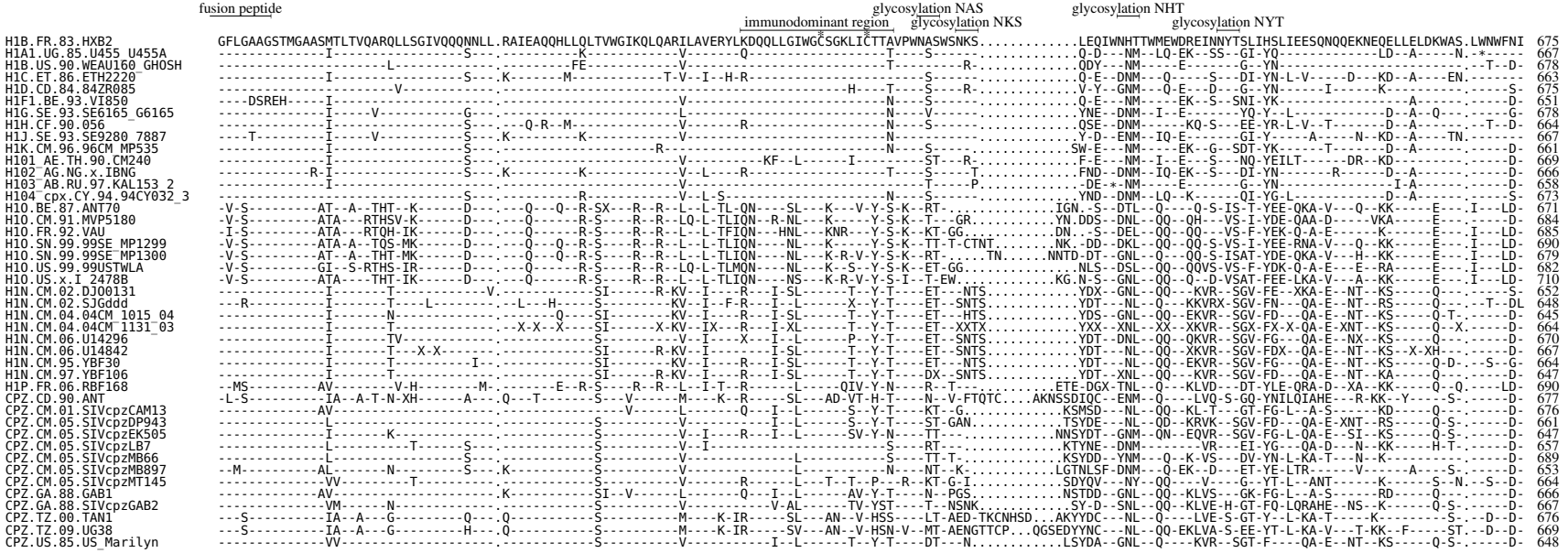

Env start
signal peptide
gp120 start
glycosylation NVT
glycosylation NDT
V1
CD4 binding
glycosylation NSS
H1B.FR.83.HXB2
H1A1.UG.85.U455_U455A
H1B.US.90.WEAU160_GHOSH
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165_G6165
H1H.CF.90.056
H1J.SE.93.SE9280_7887
H1K.CM.96.96CM_MP535
H101_AE.TH.90.CM240
H102_AG.NG.x.IBNG
H103_AB.RU.97.KAL153_2
H104_cpx.CY.94.94CY032_3
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
H1O.FR.92.VAU
H1O.SN.99.99SE_MP1299
H1O.SN.99.99SE_MP1300
H1O.US.99.99USTWLA
H1O.US.x.I_2478B
H1N.CM.02.DJO0131
H1N.CM.02.SJGddd
H1N.CM.04.04CM_1015_04
H1N.CM.04.04CM_1131_03
H1N.CM.06.U14296
H1N.CM.06.U14842
H1N.CM.95.YBF30
H1N.CM.97.YBF106
H1P.FR.06.RBF168
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.CM.05.SIVcpzMT145
CPZ.GA.88.GAB1
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.00.TAN1
CPZ.TZ.09.UG38
CPZ.US.85.US_Marilyn
Env start
signal peptide end
gp120 start
V1
MAC.US.x.239
H2A.DE.x.BEN
H2A.GW.x.ALI
H2A.SN.x.ST_HIV_2_ST
H2B.CI.x.EHO
H2B.GH.86.D205_ALT
H2G.CI.92.Abt96
H2U.FR.96.12034
COL.CM.x.CGU1
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM5
DEN.CD.x.CD1_CMO580407
DRL.x.x.FAO
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139_287
GOR.CM.07.SIVgorCP2135con
GRV.ET.x.GRI_677_gri_1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest447
LST.KE.x.lho7
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-1.CM.01.SIVmus_01CM1085
MUS-2.CM.01.CM1246
MUS-2.CM.01.CM2500
OLC.CI.97.97CI12
RCM.CM.00.SIVagi_00CM312
RCM.CM.02.SIVrcm_02CM8081
RCM.GA.x.SIVRCMGAB1
RCM.NG.x.NG411
SAB.SN.x.SAB1
SMM.SL.92.SL92B
SMM.US.x.SIVsmH635F_L3
STM.US.89.STM_37_16
SUN.GA.98.L14
SYK.KE.x.KE51
SYK.KE.x.SYK173_COMGNM
TAL.CM.00.3
TAL.CM.01.8023
TAN.UG.x.SIVagmTAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TYO1_patent
WRC.CI.97.97CI14
WRC.CI.98.98CI04
WRC.GM.05.Pbt_05GM_X02


glycosylation NIS
glycosylation NCS
CD4 binding
glycosylation NVS
glycosylation NGT
glycosylation NKT
V1
V2
glycosylation NDT
glycosylation NTS
H1B.FR.83.HXB2 261
H1A1.UG.85.U455_U455A 259
H1B.US.90.WEAU160_GHOSH 265
H1C.ET.86.ETH2220 258
H1D.CD.84.84ZR085 268
H1F1.BE.93.VI850 252
H1G.SE.93.SE6165_G6165 266
H1H.CF.90.056 254
H1J.SE.93.SE9280_7887 255
H1K.CM.96.96CM_MP535 263
H101_AE.TH.90.CM240 264
H102_AG.NG.x.IBNG 258
H103_AB.RU.97.KAL153_2 252
H104_cpx.CY.94.94CY032_3 266
H1O.BE.87.ANT70 253
H1O.CM.91.MVP5180 257
H1O.FR.92.VAU 258
H1O.SN.99.99SE_MP1299 262
H1O.SN.99.99SE_MP1300 257
H1O.US.99.99USTWLA 258
H1O.US.x.I_2478B 271
H1N.CM.02.DJO0131 257
H1N.CM.02.SJGddd 255
H1N.CM.04.04CM_1015_04 255
H1N.CM.04.04CM_1131_03 264
H1N.CM.06.U14296 271
H1N.CM.06.U14842 263
H1N.CM.95.YBF30 264
H1N.CM.97.YBF106 253
H1P.FR.06.RBF168 260
CPZ.CD.90.ANT 261
CPZ.CM.01.SIVcpzCAM13 271
CPZ.CM.05.SIVcpzDP943 248
CPZ.CM.05.SIVcpzEK505 254
CPZ.CM.05.SIVcpzLB7 255
CPZ.CM.05.SIVcpzMB66 282
CPZ.CM.05.SIVcpzMB897 251
CPZ.CM.05.SIVcpzMT145 256
CPZ.GA.88.GAB1 259
CPZ.GA.88.SIVcpzGAB2 259
CPZ.TZ.00.TAN1 258
CPZ.TZ.09.UG38 254
CPZ.US.85.US_Marilyn 246
V1
V2
MAC.US.x.239 277
H2A.DE.x.BEN 271
H2A.GW.x.ALI 268
H2A.SN.x.ST_HIV_2_ST 265
H2B.CI.x.EHO 269
H2B.GH.86.D205_ALT 271
H2G.CI.92.Abt96 277
H2U.FR.96.12034 273
COL.CM.x.CGU1 259
DEB.CM.04.SIVdeb04CMPF3061 256
DEB.CM.99.CM40 267
DEB.CM.99.CM5 257
DEN.CD.x.CD1_CMO580407 266
DRL.x.x.FAO 283
GOR.CM.04.SIVgorCP684con 254
GOR.CM.07.SIVgor2139_287 255
GOR.CM.07.SIVgorCP2135con 249
GRV.ET.x.GRI_677_gri_1 264
GSN.CM.99.CN166 272
GSN.CM.99.CN71 278
LST.CD.88.SIVlhoest447 278
LST.KE.x.lho7 273
MND-1.GA.x.MNDGB1 260
MND-2.CM.98.CM16 263
MND-2.GA.x.M14 263
MND-2.x.x.5440 284
MNE.US.x.MNE027 278
MON.CM.99.L1_99CML1 258
MUS-1.CM.01.CM1239 268
MUS-1.CM.01.SIVmus_01CM1085 273
MUS-2.CM.01.CM1246 272
MUS-2.CM.01.CM2500 260
OLC.CI.97.97CI12 263
RCM.CM.00.SIVagi_00CM312 264
RCM.CM.02.SIVrcm_02CM8081 273
RCM.GA.x.SIVRCMGAB1 275
RCM.NG.x.NG411 269
SAB.SN.x.SAB1 271
SMM.SL.92.SL92B 280
SMM.US.x.SIVsmH635F_L3 280
STM.US.89.STM_37_16 276
SUN.GA.98.L14 277
SYK.KE.x.KE51 262
SYK.KE.x.SYK173_COMGNM 255
TAL.CM.00.266 254
TAL.CM.01.8023 280
TAN.UG.x.SIVagmTAN1 272
VER.DE.x.AGM3 281
VER.KE.x.9063 281
VER.KE.x.AGM155 274
VER.KE.x.TYO1_patent 274
WRC.CI.97.97CI14 266
WRC.CI.98.98CI04 277
WRC.GM.05.Pbt_05GM_X02 269


CD4 binding
glycosylation NNT
glycosylation NCT
V3 tip
glycosylation NFT
V3
V4
glycosylation NGS
glycosylation NTS
glycosylation NNT
glycosylation NKT
CD4 binding
H1B.FR.83.HXB2
H1A1.UG.85.U455_U455A
H1B.US.90.WEAU160_GHOSH
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165_G6165
H1H.CF.90.056
H1J.SE.93.SE9280_7887
H1K.CM.96.96CM_MP535
H101_AE.TH.90.CM240
H102_AG.NG.x.IBNG
H103_AB.RU.97.KAL153_2
H104_cpx.CY.94.94CY032_3
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
H1O.FR.92.VAU
H1O.SN.99.99SE_MP1299
H1O.SN.99.99SE_MP1300
H1O.US.99.99USTWLA
H1O.US.x.I_2478B
H1N.CM.02.DJO0131
H1N.CM.02.SJGddd
H1N.CM.04.04CM_1015_04
H1N.CM.04.04CM_1131_03
H1N.CM.06.U14296
H1N.CM.06.U14842
H1N.CM.95.YBF30
H1N.CM.97.YBF106
H1P.FR.06.RBF168
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzDP943
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMB897
CPZ.CM.05.SIVcpzMT145
CPZ.GA.88.GAB1
CPZ.GA.88.SIVcpzGAB2
CPZ.TZ.00.TAN1
CPZ.TZ.09.UG38
CPZ.US.85.US_Marilyn
V3 tip
V3
V4
MAC.US.x.239
H2A.DE.x.BEN
H2A.GW.x.ALI
H2A.SN.x.ST_HIV_2_ST
H2B.CI.x.EHO
H2B.GH.86.D205_ALT
H2G.CI.92.Abt96
H2U.FR.96.12034
COL.CM.x.CGU1
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM5
DEN.CD.x.CD1_CMO580407
DRL.x.x.FAO
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139_287
GOR.CM.07.SIVgorCP2135con
GRV.ET.x.GRI_677_gri_1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest447
LST.KE.x.lho7
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-1.CM.01.SIVmus_01CM1085
MUS-2.CM.01.CM1246
MUS-2.CM.01.CM2500
OLC.CI.97.97CI12
RCM.CM.00.SIVagi_00CM312
RCM.CM.02.SIVrcm_02CM8081
RCM.GA.x.SIVRCMGAB1
RCM.NG.x.NG411
SAB.SN.x.SAB1
SMM.SL.92.SL92B
SMM.US.x.SIVsmH635F_L3
STM.US.89.STM_37_16
SUN.GA.98.L14
SYK.KE.x.KE51
SYK.KE.x.SYK173_COMGNM
TAL.CM.00.TAL1
TAL.CM.01.8023
TAN.UG.x.SIVagmTAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TYO1_patent
WRC.CI.97.97CI14
WRC.CI.98.98CI04
WRC.GM.05.Pbt_05GM_X02



PLV Proteins


fusion peptide
immunodominant region
glycosylation NAS
glycosylation NKS
glycosylation NHT
glycosylation NYT

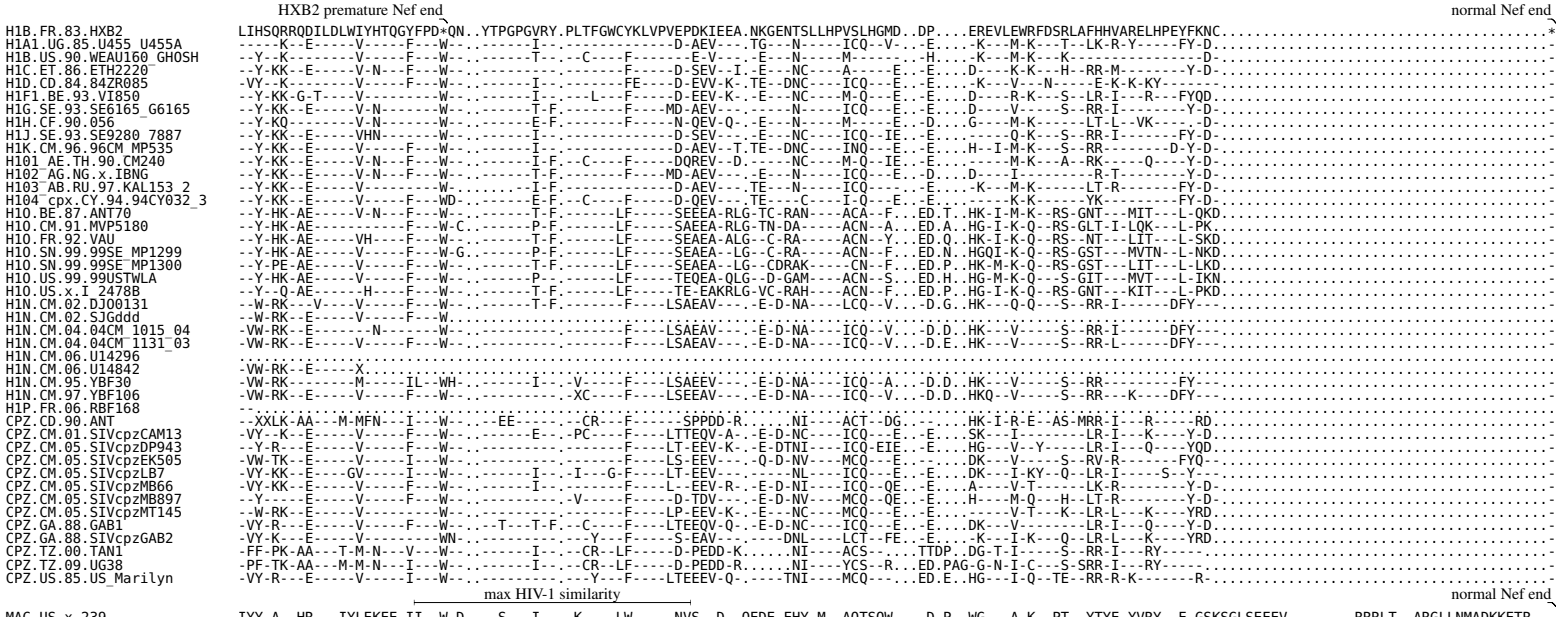
```
                                       transmembrane domain  cytoplasmic tail start                glycosylation NGS    *                                          glycosylation NAT
H1B.FR.83.HXB2               TNWLWYIKLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPT.......PRGPDRPEGIEEEGGERDRD.RSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLG.......RRGWEALKYW...........WNLLQYWSQELKNSAVSLLNATA 819
H1A1.UG.85.U455_U455A        S------R--VI-----I------T----I-------------LA-I.......-E-LG--GR-------QGK-.------S-F--IA-----N-----------FA---A-A-----RSSLKGL-L---G---L...........----L--GR---I--IT--D-V- 818
H1B.US.90.WEAU160_GHOSH      S-------I--------I------T----------------------A.......------------------------.--G---D-F-T---V--------L----I-----AK-------.......---------C...........----------------------- 822
H1C.ET.86.ETH2220            --------I-------VI----I---------------------LI-H.......------LG--------QG--.--------F--IF------------------I--AA-T-----RSSLKGLQ----T---L...........GS-V---GL---K--IN---T-- 814
H1D.CD.84.84ZR085            -Q------I--------I----------------------------L--A.......-------------------Q--G.T------FS--------N----------E-I--AA-------.......---------L...........--------R------I--VD--- 819
H1F1.BE.93.VI850             S-------I--------I---------------K------L--LI-S.......------------G---QGK-.--V---T-F---A-----N------RH---FI--AA---DRGL.......---------L...........G--TR----------I--F-T-- 795
H1G.SE.93.SE6165_G6165       -R------I--------I---------------------------LTHH.......Q-E---------G---QG-G.--V---S-F-P-------------------SI---A-T-----RSSLKGL-L---G---L...........----L--GR------I---DTV- 829
H1H.CF.90.056                SH------I--------I----I----------------------LV-N.......---------T--G---Q---.--V-----F-PVV-------S----RL--------V-T-----.......----R-----L...........-------G-------ID---T-- 808
H1J.SE.93.SE9280_7887        S-------I----I---I----I----A----------------LI-N.......-TEA---G----G---QG-T.--------F---A-----N-----------FV--AA-T-GT--.......L----I---L...........V--VW--G-------I----T-- 811
H1K.CM.96.96CM_MP535         -K------I----I---I----A-----V---------------LI--.......S--A------------Q-KN.--V---S-F---A-----N------RQ--N-I------L-RGL.......-G------L...........---V----------I----T-- 805
H101_AE.TH.90.CM240          --------I--------I----I---------------------PSHH.......QKE---------G---QG--.--V---S-F---A------------------T--AA-T-----HSSLKGL-----G---L...........G---L--G----I--I---D--- 820
H102_AG.NG.x.IBNG            -------RI--------I---------T-I--------------LTHH.......Q-E-----R---G---Q-K-.--V---S-F---A------------------I--AA-T-----HNCLKGL-L--G----L...........---IS--V-------IN---TI- 817
H103_AB.RU.97.KAL153_2       SK------I-------------I---------G-----------R--A.......Q------------D------.T------RF------------F-I--H-------AA-------.......----------...........--------I----S--IN-IGTI- 802
H104_cpx.CY.94.94CY032_3     -K------I--------I----I------------------L--LI--T......Q--L---G-T------Q--S.--------F-P-------N------RH--N-----A-T-----.......I-------L...........-F-L--G-------IN-F-T-- 818
H1O.BE.87.ANT70              -K------IA-I---A---V-VIMI--N--KNI----Q---L-IPNHH.......QEEAGT-GRTGGG---EG-P.-W-PSPQ-F-P-LYT---TII-WT--L-SN-ASGIQK........VISYL-L-LWI-.GQ.KIINVCRIC.AAVT---L---Q---T---DTL- 822
H1O.CM.91.MVP5180            -K------IA-I---A-I-I-VIMII-NL-KNI----Q---L-IPV-H.......RQEAET-GRTG-----G--P.KWTA-PP-F-QQLYT---TII-WT--L-SN-ISGIR-........LIDYLGL-LWI-.GQ.KTIEACRLC.GAVM---L-------TN--DTI- 835
H1O.FR.92.VAU                -K------IA-I---A-I-V-V-ML--NL-KNI----Q---L-IPIQQ.......QAEVGT-G-TG-G--DE--R.-WTP-PQ-F-H-LYT---TII-WI--L-SN-ASEIQK........LIRHLGL-LWII.GQ.RTIEACRLF.KAII---L---QT--TN--DTV- 836
H1O.SN.99.99SE_MP1299        -K------IA-I---A-I-V---MI--NL-KNI----Q----K--IHH.......QPEAEA-G-TG-G----GMP.TL-PWPQ-F-P-LYT---TII-W---L-SN-ASGIQT........VISHLGL-L.WTLGQ.KIISACRLC.IAVI---L---Q---T---DTI- 841
H1O.SN.99.99SE_MP1300        -K------IA-I---A-I-V---MI--NL-RNI-H--Q------PTHHQ......QPEAQA-G-TG-G------L.---PSPQ-F-P-LYT---TII-W---L-SN-ASGIQT........VISHLGL-LWI-.GQ.KIISACRIC.IAVI---L---Q---T---DTL- 831
H1O.US.99.99USTWLA           -K------IA-I---A-I-V---MI--NL-KNI----Q---L-IPNHH.......QQEVGT--RTG-G---EG-R.-W-ASPP-F-Q-LYT---TII-W--QI-SS-TSGIQK........LISLLGL-LCT-.WQ.RITEACRLC.AAVT---L----I--T---DTV- 833
H1O.US.x.I_2478B             -K------TA-I---A-I-I---MVI-NL-RNI----Q-----IPIHH.......QSEAET-GRTG-G---EG-P.-L-PSPQ-F-P-LYT---TII-W---L-SN-ASGIQK........LISHL-I-LWT-.GQ.KIINACRIC.AAVI---L---Q--TN-VDTI- 861
H1N.CM.02.DJO0131            --------IA--V-A-II-I--ISVIIT-IA----------L--LI--TT.....T-------ET--GV-GQG--.--V---S-F-T-V-E-F-N-LI-L----T-S---LQ-TL---RSLIRGLQLLN-LRTRL...........-GIIA--GK---D--I----TI- 805
H1N.CM.02.SJGddd             --------IA---A-XI-I--ISVIIX-IA----------L--LI--.......T--X-----T--GV--Q--G.--V---S-F---V-E--N-SI-L----T-S---LR-TL---RQSLSRGLQLLN-LITRL...........-GI-A--GG-XRD--I----T-- 799
H1N.CM.04.04CM_1015_04       S-------IA--V-A-II-I--ISVIIT-IA----------L--LI--.......T--------T--GV--Q--G.--V----F---V-E-C-N-VTLL---------TQ-TL-I--QSLNRGLQLLN-LR-RL...........-GI-A--GK---D--I----T-- 796
H1N.CM.04.04CM_1131_03       -S------IA--V-A-II-I--ISVIIT-IA----------L--LI--.......T-------XT--GV-XQ--G.--V---X-F---V-E-X-N-VILLCX--X-----XQ-TL----QSLSRGLQLLN-LXXRX...........-GX-A--GK---D--I---XTX- 815
H1N.CM.06.U14296             S-------IX---A-II-I-XISVIIT-IA----------L--LI--.......T------GETG-DV--Q--G.--V---S-FFS---E---N-VI-L-R--T-S---LX-TL----QSLSRGLQLLN-LRIHL...........-GIIA--GK---D--I---XT-- 821
H1N.CM.06.U14842             -K------IA---A-II-I--IS-IIT-IA----------L--LI--.......T-------ETG-G---Q---.--V----F---V-E---N-LT-L----T-S---LR-TL----QSLSKGLQLLN-LRIHL...........-GIIA--XK--XD--I----T-- 818
H1N.CM.95.YBF30              -K------IA---A-I--I--ISI-IT-IA----------L--LI--.......A-------ET-GGV--Q--G.--V---S-FS--V-E---N-LI-L----T-S---LR-TL----QSLSRGLQLLN-LRTHL...........-GI-A--GK--RD--I----T-- 815
H1N.CM.97.YBF106             -K------IA--V-A-II-I--IS-IIT-IA----------L--LI--A.....A-------ET--GV-GQ--G.--V---S-F----E---N-LI-L----A-S---IR-TL-I--QSLSRGLQLLN-LRIRL...........-GIIA--GK---D--I----T-- 799
H1P.FR.06.RBF168             XQ------I-------II----LL-XXNV-R-I------V-L--LGLN.......GDPAGIAP-TN----AGNG.-----LD-F-P-V----KN-VVQI-QI-VGCI-GIKDLLTI-WIHLGQLLT--LNC-RDC...........FAACG--T----Q--T---DTV- 841
CPZ.CD.90.ANT                -Q------I------AI-----LLVLV-CLRK-----H------..I--QN.....QQD-EQ--E-R----RK--I.-WRA-QH-FF--L-V--T-IIQWI-QIC-TC--NLWAVLQH-CR......ITFRLCNHLE...........N--STL-.TIIRTEIIKNIDRL- 821
CPZ.CM.01.SIVcpzCAM13        --------I-L-A---II----IMS-V-VIR----------L--LI-A.......Q-E-GTLGETGG----PGSG.--V---T-CWP-------N-VIW--RS-TS-AC--W-QL-K--QLLINIL-LIQ-K-TLL...........RGII---GR--RT--T---D--- 827
CPZ.CM.05.SIVcpzDP943        S-------I-L-A-A--I-I---GVIM-LIAK----------L--LI--.......T------G-T------PG-G.T-T---S-F---A-----N-L--L----I---S-LR-TL-I--QNINKGLQWLRDISRHL...........-GVIA--GR--QI--T---DT-- 812
CPZ.CM.05.SIVcpzEK505        -K------I---V-A-I--I--ISIIM-M-A----------L--LI--.......T-------RT--DA--L-NG.--V---S-F---A-E-F-N-L--L----T-C-S-LR-TL---RQNIHKGLQLLN-LRI-L...........-GIIA--GR---I--IN--DT-- 798
CPZ.CM.05.SIVcpzLB7          SH------I----I---IVC--I----A---------------LS-I.........---LAQ-G-T-----QGT-.---V--LD-F---A-N---D----L----T--V--AR-TL-IV-QYTLKGL-LV----L-L...........QGI---GK---T-T---DT-- 808
CPZ.CM.05.SIVcpzMB66         -S------I--I----I-------L-----------------LI-A.......R-DR----E---G---P-NV.------S-F---A-N---D----L-----------L-TL--V-QTLLKGL---R---IHL...........RGI----G---T--I---DT-- 840
CPZ.CM.05.SIVcpzMB897        -Q--------V------I-I-----M------------------PF-V.......R-D---------G---N-NV.--V---S-F---A-----N------------W--R-TL-IV-QKFXAGL-LV--T-T-L...........KSI----G---T--I---D--- 804
CPZ.CM.05.SIVcpzMT145        --------M-LIV---II----C--IG-LI----K------L--LI-A.......NQ---GLGET-KG---NV-G.------S-F-P-V-E---N-LS-L--Q---CASLIW-LL-I--QYSLRGVQQIGTL-HQL...........RGT----TT-I----I---DT-- 815
CPZ.GA.88.GAB1               -K------I-L-A---II----IMT-F-V-R----------L--LI-V.......Q-EQG-LGE-D-G---Q--S.--V---E-C-P-------N-GIW--QS-TS-ACN-W-QLKT--HLILHSL-LLR-R-CLL...........GGII---GK---I--I---D--- 817
CPZ.GA.88.SIVcpzGAB2         --------M--IL---VI-----IG-VN-IR-S-----------LI-A.......TG--EV---GGG---Q--G.--V---S-F---C---I-N-TI-L-RL---C-SVLWSLL-Q--QQVLRGL-LLR-L-SQL...........KGIG---L---RT--I---DT-- 818
CPZ.TZ.00.TAN1               -Q------IA-I--A-------LMFIVNV-KQ-----T--FS-IPTQA.......EQD-EQ-G--AGG--G--NI.-WTPSPA-FFSIV-E---N-LIWI-QTFQNFIW-LWISLQA-KQGIISLAHSLVIVHRTIIVGVRQI...IEWSSNTYAS-RVLLIQAIDRL- 834
CPZ.TZ.09.UG38               -Q------IA-L--A--------MFIVNVIRQ-----M--FS-IPTQA.......HQD-EQ--ETAGG--GTGNV.-WTP-HR-FFIVV-E---T-L-WI-QTCQNFGWLLWT-LKT-RQGIISLTQ-LIIVQRDIILKSRQL...FDW-SNTYSI-RT-LIQAIDRL- 827
CPZ.US.85.US_Marilyn         --------I-LIV-AS---I---GVIF-L-AK---------L--LF--.......T-E------T--GA-KT-NV.--T---S-F---V-E---N-LI-L----E-----LR-T-QI--QNINKGLQLLN-LRARC...........-GVIA--AR---V--T---DT-- 799

MAC.US.x.239                 AS-IK--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIQQ..D-AL-TREGK-RDGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRVYQI-QPILQRLSATLQRIREVLRTE.........LTY---GWSYFHEAVQAVWRSAT 842
H2A.DE.x.BEN                 -S-VK--QYGVH--V-IIA---AIY-VQLLS-F-K--R-VFSSPPGYLQQIHIHK..DRGQ-ANEGT-EDVGGDS.GYDLWPWPINYVQFLIHLLTR-LIGLYNICRD-LSKNSP....TRRLISQSLTAIRDWLRLK.........AAQ---GCEWIQEAFQAFART-R 823
H2A.GW.x.ALI                 -A-VK--QYGVY-IV-I-A----IY-VQMLS-L-K--R-VFSSPPGYIQQIHIHK..DQEQ-TRG-T-EDVG-NV.GDRLWPWPIAY-HFLIHLLAR-LIGLYSICRD-LSRISPI-QPIFRSLQRALTTIRDWLRLK.........AAY---GCEWIQEAFRAFARIAR 825
H2A.SN.x.ST_HIV_2_ST         -S-IK--QYGVY--V-IIV----IY-VQMLS-L-K--R-VFSSPPAYFQQIHIHK..DREQ-ARE-T-EDVGNSV.GDNWWPWPIRYIHFLIRQLIR-LNRLYNICRD-LSRSFQT-QLISQSLRRALTAVRDWLRFN.........TAY---GGEWIQEAFRAFAR--G 822
H2B.CI.x.EHO                 -S-MA--R-GLY--I-I-V---AIYIIQMLA-L-K--R-VFSSPPSY-QQIPIRK..DRGQ-ANE-T-E-GGNNE.GYRSWPWQIEYIHFPIRQLRD-LIWLYSGCRT-LSKTFQT-QPVLQPLRLPPA..................Y-R-GISWFQEAIQAAAR-AG 815
H2B.GH.86.D205_ALT           -S-IK--H-GLY--A---V----VYIVQMLA-L-K--R-VFSSPPSY-QQIPIRK..DRGQ-ANE-T-E-GGNDG.DYRSWPWQIEYIHFLLRQLRN-LIWLYNGCRT-LLKTFQ........ILHQISTNLQPLRLP.........VAY---GISWFQEALRAAAR--G 822
H2G.CI.92.Abt96              -S-VK-VY-GLYV-A-VIV----IY-VQ-LG-L----R-VXSSPPSYVQQIHIRK..DQEX-TKE-I--ESGNKG.GYRSWPWQIEYIHFLXRQLGN-LTWLYSNCRA-LSRIXQT-QPLFQRISRTLQAIREHLRLE.........AAYFS-GFRW-QEACTAATR-AQ 847
H2U.FR.96.12034              -S-VK-VY-GLYV-A-III----IY-VQLLGSL-K--R-VFSSHPSYVEQIPIQR..DQEQ-TKG-I-E-AG-SG.GYRSWPWQIRYIHFLIHQLIR-LTWLYSSCRD-LSRICQS-QPLF......QSIRERLHLE.........IAY---GW-YF-EAFQAFGK-AR 830
COL.CM.x.CGU1                F-IFK-VLYAAYV----I----IMV-IACIRGAFRVKGFQQIGRTNVSS.....QIRGQDR-WEQPDNA-EIHR.EQENSSSRFW..............RLMEEWWSSLL-HCSNGILT...........VLPQHPPTTS....DKRRSA-.-T-RYL..LVPRGPS 800
DEB.CM.04.SIVdeb04CMPF3061   SR-F-W-RIVVYVIA---L---IMF-VN-ISKLCR-----QTPFQK.........EDLEL-GEG-GGV-GST-R.-WTPSPQ-F-S---G--QQ-LSWL-QTF-NFIWLL--GA-I-QEQIYRLSFAI-TL-QRLGEQLRVL....-GY---GWK-F-DTI-WAGGELG 809
DEB.CM.99.CM40               SR-F-W-R-VVY-IAA-IL---AMFGVN-GSKLCR-----QTL-PK.........EEEVL-LAEEGGRS-GDG-A.-WTA-PR-FFS---E--QQ-LTWL-QTF-NCIWLIR-GF---QEGIYRLSFAL-QR-RLLGERFRVL...GGY---GWK-F-DTITWAGGELW 833
DEB.CM.99.CM5                SR-F-W-R-VVY-IAA-IL---AMFGVN-GTKLCR-----QTL-PK.........EVEVLQLG-EGDKS--GVKS.-WTTSPR-F-R-----ISQ-LTWL-QT--NCAWLIR-GF---QETIYRLSFAF-QR-RLLGEHLRVV....-GYC--GWK-F-DTIIWAGQEIW 813
DEN.CD.x.CD1_CMO580407       SK-F-W-R-VVYVIA---I---IMFIINVLS-LCR-----LQIPIR.........-QKAANQPEEG-GESGDGNRLKWWQYPR-F-SI--E--SQ-LTW--LT--N-VFLIRTACSRIWDFV..........HRHSLILLG........YF--GL---QTGLRD-GTSAI 827
DRL.x.x.FAO                  -W-SK---IGFF--MAII----LA-LW-T-G-F----R--.PYIFKDAYRPET..TNYHRQ-DREKG-EQD-GEW.NIRSEPSKPGSSKAWSEE..TVGTWLKESRGYIWLKNVKAV-EYGWSE........................-QEAGRGIC.TLLQEVAQRIW 858
GOR.CM.04.SIVgorCP684con     -Q-I---RI--IVIA-----K-LM-IIN-CH-----------ILGRN.......QVPAGIVPE-G----RA-S-.-----LA-F-P------T-VVWI-RT-VI--SGIKE-TLS-IEXLTRLLX-VNNLTRDC...........FAFIA--G----Q--I---DCV- 821
GOR.CM.07.SIVgor2139_287     -Q------I--I-IA------LM-IINMCH-----------ILGRN.......QVPAGIVPE-G----RA-N-.-----LA-F-P-----I-T-VVWI-RT-VI--SGIKE-TLS-IEYLTRLL--VNNLTRDC...........FAFIA--G----Q--I---DCV- 817
GOR.CM.07.SIVgorCP2135con    -Q------I----IA------LM-IINMCH-----------ILARN.......QVPAGIVPE-G----RA-S-.-----LA-F-P------T-IVWI-RT-VI--SGIKE-TLS-IEHLTRLL--VNNLTRDC...........FAFIA--G----Q--I---DCV- 812
GRV.ET.x.GRI_677_gri_1       ST--G-V-IGFLVIVIIL---FAWVLWGCIRNI----N--P........QIHIHS..SAE--DNGGGQDRGGESSSSKLIRLQEESSTPSRINNWWLN-KSCS--IRTWCYNICLT--IFIRTAVG.........................Y---GL-Q-QEA-TG-AQ-L- 817
GSN.CM.99.CN166              SD-QR---I-VIV-AA-IA----MFI-NMLR-I--------P-ILI-Q.......AVERGP-DAT--GV-AQGKV.--V-YLT-FSS-L-----N-VIWI-QI-AS-AWTLRG-GKF-WEQLQKAIQKAIQQARQLREVAARV...IAYIS-GI---QAA-TGI-DSL- 844
GSN.CM.99.CN71               S-------I---V-AA--A---IMF--NMLR----------P-ILT-Q.......QVER-P-D-TRKD--AQGNV.--V-YLT-FSS-L------VIWI-QI-AT-VWTLQ--GKF-WEQ....LQKAFQQARQLGEVAARA...AAYIS-GI---QAAVSGI-DSL- 849
LST.CD.88.SIVlhoest447       LS-FQ---YAVL-II-IIV--V-SFIVQNIVKMCR--RV-APSAYVEQDCKWEKR.ENQEQ-DKEG-IEKDKENIYINLEQYKKESSTPPWNVDWSEPLQDSLFVT--KWIKAFGI--LSLVWQFLSWLGHLVILFFQ.HGQHLRKACSRRVVENAQKIA-WSWNKIRRN 876
LST.KE.x.lho7                SH-FT-V-YAVL-ILVII---VLSFIIQN-VKMCR--RV--PSVYIEQDYKWEKE.ENQEQ-DRE--K-ADTETIYINLEQCKKESSRPLWNVDWNEPLQDSL-VT--KWLKEGGI--LSLVWQSLSWL-HL-ILFFQ.NGQRLWQTSSRWMVENAQKIQ-WLREKCRRN 873
MND-1.GA.x.MNDGB1            -W-VQ-L-WGVFL-L-II----LL-LWNTIS-F----R-VFS.......QDCQQN.LYRK--DNG---SNSLELG.EHNSENL...KEESLNRSLIEDLTSFARE*PIWLWLKNLRAAIEYGF...........LAVQEA...IRSLGS-LVSFAAISWR-GYTGWLQ- 844
MND-2.CM.98.CM16             -W--K-V-IGLL--IVIIV---LACLW-VLGKF----R--...PYVFKGDYLRPH..NLKQ-GKER--EPDSEKQ.STKSDSSKVEFGKPWSKE...QIRDWLKTSRGYVWLKNLQAVIEYGWQE.......................-KTAGGKIF-VLQGYAQGLWS 830
MND-2.GA.x.M14               -W--K-V-IG-L-IMVVI----LACLW-TIG-F----R--.PYVLKGDYLRPR...NLRQ-DEER-GELNLEEQ.NIKSESSKKEFGRPWRPE...QIRNWLKESTIY-WLKNLQAVIEYGWQE.......................-KAAGARICQVLQSFAQRLWS 831
MND-2.x.x.5440               -W--K-V-IGLLV-VVII---MLACLW-VLGKF----R--.PYVFKGDYLRPH...NLK--DREGG-EPDLEKQ.NIKSESSRQESRKPWKPE...QVRSWLK-ST-Y-WLKNLQAVIEYGWQE.......................-KAAGAALYQVLQGFAQRLWS 854
MNE.US.x.MNE027              AS-IR--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIRQ..DQAL-TKEGT--DGG-SG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPIFQRFSTTLQRVREVLRTE.........LTY---GWSYFQEAVQVAWRSAT 845
MON.CM.99.L1_99CML1          SK------I--IV-AA-I----LMFI-GVIS-LG----L--S-IPI-S.......HA-QPT-D-TGAG--DGSNS.--PAYLK-FFTI--E---N-V-WT-QI-K-SV-VIY--LQRVSQRLPPLLHIRLLQ-WESLRRLLAYCQY.......GI---QAAVT---D-L- 832
MUS-1.CM.01.CM1239           SK------I-C-V-A---LF--FMF--G-LR-I--------P-I-TQP.......-EAAEP---T-GGV-SSG-T.--V-YL--F-S-F-----N-VIWF-QT-VSSVSTLRW-LQRAAHLIQKWSV-L-IVIREQARRCSELLARLLAYI--GI---QAAGTAA-DQL- 854
MUS-1.CM.01.SIVmus_01CM1085  SK------I-C-V-A---LF-L-MF--G-LR---------P-ILIRP.......QT--EA---T-GD--SGS--.--V-YLS-FFS-F-----N-VIW--QT-ISSASTLWWLLQRAVHLIQKGSI-L-FT-RALIEPLREALQRALAC--HGI---QTASTAV-DQL- 860
MUS-2.CM.01.CM1246           SK---V-I-C-V-A--IA---MFI-GMLR-I--------P-I-TR.............QSAEQ-QQEKTEKG.AGRSGRHR--S-FSE--HN-MIWT-QI-TSSASNLWWLTQRIYHHLTKGLVNP-TL--SFREALGRFIAH....A--GI-K-QAAVSGA-D-L- 845
MUS-2.CM.01.CM2500           SK--F-V-I-CLV-A--IA---IMFL-G-LRK--------PP-I-TRA.......-E--RT--ET--G--A-GSV.--V-YL--F-S-------N-IVW--QI--SS-SNLWWLIQ-IVHHLQKWSISL-LRARATSTHLLEVLRRLLAY---GL--FQEGVTAAIDGL- 842
OLC.CI.97.97CI12             A--MT-V-IGLLV-LAY---N--LGLVRKIRIM----VS-QLP......FSDTL..QDKEQ-DNG-G-DS-FFES..TFNEW-CTTECTGTLRVNWRDLLS-LWKGVVRVTR...........................WI.QKSLQHVWRTLKQKRQ. 810
RCM.CM.00.SIVagi_00CM312     DW-ME---IG-FVIL-II------LLW-X-SKL-A--T--LSPPPYYHSQIPIRNLKKEEQQAKG--GEFSGEGGGYKSYNWQKGSFY-LILRPIVT--QTLWNNLIS-LLLSYQS-QSILSRIQHPLL-IVA-VQ...............DGW--F-AF--W-GEMA- 817
RCM.CM.02.SIVrcm_02CM8081    DW-ME-L-IGVF-IL-VI----F-LLWGCISKI-A--T--LSPPPYYHQQTPIRNREEQPIK--G-GDS--EGG..YKSPSWQRGYFYSLILRPIEI-LRALWNSCNS-LSVIFQS-RSILSQIQQTLR-IIA-IQ...............DGW----EFSGW-AEMAQ 824
RCM.GA.x.SIVRCMGAB1          DW-MQ---IGFF--I-II---VAWLLWNCLSNL----R--..SPPSYVQQIHIHNTGE-QT-GEKR-D---EGG..NKYNNWLREYCW-QLIHPLSRIWTQLSQICRSCSSIIFQS-RWILA.........................KI--GW--F-EFSSWFAEMAL 813
RCM.NG.x.NG411               DW--K--RIG-F-IL-II----I-LLW-C-SKL-A--T--LSPPPCYHQQIPIRNREEQPIK--EKGGS--GGGH.KYYNWQR-........YFYILILRPIEILCRR-YQICSNS-SVIYQSLQSIPSRIL-HLRWIGA.........K--DGW--F-GFSSW-AEMAR 820
SAB.SN.x.SAB1                -K-FGWM-IA--VIA-IIVA-VLLVIIG-LRKF-K--A---SLPSSH*.QIHIHL..DKGQ-DKE--Q-SGGLNSDSRSYTWQREFL...........RHLCH-LITWLRN.LTSWFSTIFSNLH-CLQDIQQRTRQLTA........H-E-GW----AA-A--TV-VV 830
SMM.SL.92.SL92B              -S-VR--QYGVFL-I-I-L----IY-VQMLS-L----R-VFSSPPSYHQQIHIQR..DQEL-AKEGT--EGGNGG.GYRLWPWQIEYIHFLIRQLIRILTWLYNNLTR-ASRAYQN-QQLCQRLSEISQPIRELVRREAG........YIR-GWNYFIEACQEAWRSAQ 849
SMM.US.x.SIVsmH635F_L3       -S-IR--QYGVL--L-V------IY-VQMLA-L----R-VFSPPPAYVQQIPIHK..DQEP-TKEG---EGG--G.G-RSWPWQIEYIHFLIRQLIR-LTWLFSSCRDWLLRIYQI-QPVLQRLSRTLQRVREVIRIE.........ITY---GWSYFQEA-QAWWKVAR 850
STM.US.89.STM_37_16          -S-VR--QYGVYL-I---M--VAIYIMQLLA-L-K--R-VFSSPPSCRQQIPIHK..-QEQ-TKEGT-E-GG--G.GINSWPWQIEYTHFLIRQLVR-LTWLYNNFRACLSRIYQT-HPTFQRISRILQRIREVVRLG.........AAY---GCIWIQEA-QAAWR-AG 845
SUN.GA.98.L14                SQ-FL---YAVL-I-IIIAA--LSFIIQQIY-MC---RV--PSAYVEQDWLQETC.PK-TDK-EE--TEK--I..YINLEQSKKESLPPPWTVDWDEPLRDSL-VT--KWLKAWGIV-AQNIYHLLSFL-HL-TTSFH.HGQRLWQTLRGWLGSHLIQAT-RIRNACR-S 877
SYK.KE.x.KE51                SK-FQ-L-IGFFAIAAI-I--VLSFAWG--RNMLG-----LQKPQP--........LYHKA-A-P--GIETDTDN.YKPSVSTF-REFLRQS-AEAQKLWNRVSESCRSLIRGL-IAW-..........................FIS-GVT---EA-I--GREV- 803
SYK.KE.x.SYK173_COMGNM       SK-FF-L-IGFYVI-A--L--L-SFSVG-IKNLLG--V-ILQNPT...........Q-RKD-GKPAD-EEGSGDR.EGLNVSTF-RESLRQS-EAGQQLWRTVCSSFRSLIRQLTITW-..........................FIS-GFN---IA-A--GREIR 795
TAL.CM.00.266                SA-FRWLRIAVIV-AS-IL----MYIVNVLKLL----L--P-K-PI-S.......N-D-EA-G-DAT---KQGSY.KWN-SQA-F-SI--E--TG-IRYI-QT-AN-AWLIWDLAKAAKAKSQQVAQATYAF-RLHLTAAAATLQY.......GFS-IQAAWRDGIT-L- 835
TAL.CM.01.8023               SA-FRWLRIAVIV-AS-IL---MYIINVLKLL----L--P-K-PI-S.......N-E-GV--EEGT---KP--S.-WRTSPQ-FF-I--E--N--LSY--QT--N-VWLSWDLLKR-KATTQQAAQATYAY-RIKFTAAV.......ATI--GFS-IQAAWGDCVT-L- 853
TAN.UG.x.SIVagmTAN1          SS-F-W--WGFY--I--ILF-MAWLIWGCIA------F---P.......QINIRL..-REQ-DNAGG-DKDSSSS.-DKSPPSVKES-LPNRGGIQAEERAW-QHLTNWCL-ISSW--RLYQILRRSLTTLLQLLRQE.........CQYI--GW-QF-EG-ARSFE-L- 842
VER.DE.x.AGM3                SK--NIL-IGFLD-L-II---LLYT-Y-CIA----------P.......QIHIHP..WKGQ-DNA-GP-EGG-KRKN-SEPWQKESGTAEWKSNWCKRLTNWCSISSIWLYNSCLT--VHLRSAFQ.....................YI--GLG---AA-QEAVV-L- 840
VER.KE.x.9063                SK--NIL-IGFLV-V-II---LLYTIY-CIG----------P.......QIHIHP..WKG--GNAG-PEEGGE--ND-SDSWQKGSGTRQKRGSWSKRLTNWLSTFSTWLYNTFLT--IHLRGAWQ.....................YI--GLG---TA-QEARL-V- 840
VER.KE.x.AGM155              SK--NIL-IGFLA-I-VI---LLYTLYTCIA----------P.......QIHIHP..WKGQ-DNAG-PEEGGRTGKSKSTH*QKEFGGRDKRTSWCRQLTNWF-TLSIWLHNSCLT--LKLRSAWQ.....................Y---GLG---AA-QEAV-HL- 833
VER.KE.x.TYO1_patent         SK--NIL-MGFLVIV-II---LLYT-YGCIV------V---P.......QIHIHQ.V-KG--DNAD-P-EGG-NSRIKLESW*KDSKSRCMQ-TAWLTRLN......TWLYNSCLT--IQLRKAFQ.....................Y---GLA---TG-QEI-QTL- 826
WRC.CI.97.97CI14             F-IWNWL-WGLLVI-IV-----LACLMNSIRIA-----K-IS-......QNHLNN.LEEGQQDSE-GIEEDYNSIEINDWQYRRES-IVFLNSLLTS---ICQHLLHIVWSLLSCL-WWINDGW........SLLKDI.............SGRAYGAIRKKIAKSRKE 835
WRC.CI.98.98CI04             F-IWNWL-WGLL-I-IV--I--LGWLMTSIRIA-----K-IS-......QSHLNN.LEEGL-DRE--AEEDNSTLDV-DWQFKKES-ITLLNSLINF-SAVCQTLLQIVWSLLACLFWWIQHGW........NLLKRI.............SGAAHACIRKKITKSRKE 845
WRC.GM.05.Pbt_05GM_X02       F-IWNWL-WVLLVI-LV--I-VLGCLFTSLRIA-----K-VS-......QSHILN.LEEGL-DNE-T-EEDFIAIEE-DWQFKKESLILFWNSLATL-Y-ICLSLLQ-SYSLLSYLWI-IQNGW........SLLKRL.............CGAAYAALQEKIAARRQK 837
```

```
                                                                 cytoplasmic tail end
                                                                            gp41 end
                                                                             Env end
H1B.FR.83.HXB2                IAVAEGTDRVIEVV...QGA*CRAIRHIPRRIRQGLER..ILL....*   856
H1A1.UG.85.U455_U455A         V---GWI-----IG...-TIG---LN-----------..A--....-    856
H1B.US.90.WEAU160_GHOSH       ------------I-...-RT----L------------..A--....-    860
H1C.ET.86.ETH2220             -V-G-----F--LI...-RIW--FCN----------A..A-Q....-    852
H1D.CD.84.84ZR085             ---------I-DI-...RR--K-VL---T--------..A--....-    857
H1F1.BE.93.VI850              -V-------I---L...-R-G--VLN--------A--..A--....-    833
H1G.SE.93.SE6165_G6165        ----NW-------A...-R-----LN--T--------..A--....-    867
H1H.CF.90.056                 --------GI-VI-...-R-W---L---------F--..S--....-    846
H1J.SE.93.SE9280_7887         ---------I--IA...-R-F---L------------..A--....-    849
H1K.CM.96.96CM_MP535          ----G----I--IG...-R-F--LL------------..A--....-    843
H101_AE.TH.90.CM240           --A-GW-------A......--W---L----------..T--....-    858
H102_AG.NG.x.IBNG             -V--NW---A--IG...-RVG----N--------F--..A--....-    855
H103_AB.RU.97.KAL153_2        ----GW------IG...-RF---M-N--------A-K..A-Q....-    840
H104_cpx.CY.94.94CY032_3      ---------I--A-...-R-----CN-----------..A--....-    856
H1O.BE.87.ANT70               V---NW--GI-AGI...-RIGTG--N-----------..S--....-    860
H1O.CM.91.MVP5180             VS--NW--GI-LGL...-RIGQGFL---------A--..--V....-    873
H1O.FR.92.VAU                 V---NW--SI-LGI...-SIG-G-LN-----------..L--....-    874
H1O.SN.99.99SE_MP1299         V---NW-VTI-LGI...-RIG-G-LN-----------..S--....-    879
H1O.SN.99.99SE_MP1300         V---NW--GI-LGL...-RIG-G-LN-----------..A--....-    869
H1O.US.99.99USTWLA            V--GSW--NI-SG-...-RIG-G-LN--------A--..--X....-    871
H1O.US.x.I_2478B              V---NW--SI-LG-...-RIG-G-WNV-------F--..S--....-    899
H1N.CM.02.DJO0131             -V-------L--LA...-RIG-G-L------------..A--....-    843
H1N.CM.02.SJGddd              -V-------L--LA...-RIG-G-L---------A--..A-V....-    837
H1N.CM.04.04CM_1015_04        ----GX---L--LA...-RIG-G-L------------..T--....-    834
H1N.CM.04.04CM_1131_03        ---------X--LA...-RIG-G-L-V----------..T-V....-    853
H1N.CM.06.U14296              -V-------L--LA...--IG-G-L------------..A--....-    859
H1N.CM.06.U14842              -V-------L--LA...-RIG-G-L------------..--V....-    856
H1N.CM.95.YBF30               -V-------I--LA...-RIG-G-L------------..A-I....-    853
H1N.CM.97.YBF106              -V-------F--LA...-RIG-G-L------------..A--....-    837
H1P.FR.06.RBF168              -S--GW--Q--I-G...-QIG-GFLN--------I--..S--....-    879
CPZ.CD.90.ANT                 -W-G-K--SILLAL...-TIV-I--EV---------I..A-N....-    859
CPZ.CM.01.SIVcpzCAM13         ---G-----I-RA-...-IVF-I-GN-----------..T--....-    865
CPZ.CM.05.SIVcpzDP943         ---------I----...-RIG-G-L------------..S--....-    850
CPZ.CM.05.SIVcpzEK505         V--------I--L-...-RIG-G-L------------..A--....-    836
CPZ.CM.05.SIVcpzLB7           ---------I--I-...-RIG-G-L---------F--..A--....-    846
CPZ.CM.05.SIVcpzMB66          ---------I--IA...-RFG-G-LN-----------..A--....-    878
CPZ.CM.05.SIVcpzMB897         ---------X--IA...-RVG-G-L------------..A--....-    842
CPZ.CM.05.SIVcpzMT145         ---G-----IL-AI...TRLG-G-L------------..A--....-    853
CPZ.GA.88.GAB1                ---------I--AF...-VTL-I--N-----------..A--....-    855
CPZ.GA.88.SIVcpzGAB2          -I---R--TI---A...TRIG-G-L------------..A--E...-    857
CPZ.TZ.00.TAN1                NFTGWW--LI--G-...VYIA-G--N----------L..A-N....-    872
CPZ.TZ.09.UG38                NFTGWW--L--AG-...AFVAQG--N----------I..A-N....-    865
CPZ.US.85.US_Marilyn          ---------I--LT...RRLFLG-I------------..S--....-    837
                                                                             Env end
MAC.US.x.239                  ETL-GAWGDLW-TL...RRGG-W-LA----------L..T--....-    880
H2A.DE.x.BEN                  ETL-GAWGWLW-AA...RRIG-G-LAV-------A-L..A--....-    861
H2A.GW.x.ALI                  ETLTNTWRDLWGA-....-WVG-R-LAV-------A-I..A--....-    863
H2A.SN.x.ST_HIV_2_ST          ETLTNAWRGFWGTL...GQIG-G-LAV-------A-I..A--....-    860
H2B.CI.x.EHO                  ETL-SAARTSWG-L...RR-AGE-IA--------A-L..A--....-    853
H2B.GH.86.D205_ALT            ETL-SAGETLW-AL...RR-A---IA----------L..T--....-    860
H2G.CI.92.Abt96               ETLTSTWRALWKTL...GRVG-G-LA----------L..T--....-    885
H2U.FR.96.12034               ETLSRTGRELW-TL...GRVG-WL-A--------F-L..A--....-    868
COL.CM.x.CGU1                 LQIL-TLQ-WLRSA...ARGW-RAPEYL-GWIYDRPQ..GPA....-    838
DEB.CM.04.SIVdeb04CMPF3061    QTFQRWAEVALQ-L...GRGI-E-LA--A-----A-L..F-N....-    847
DEB.CM.99.CM40                QWLQGIAEMALQGL...VR-GG--LRV-A-----A-L..--N....-    871
DEB.CM.99.CM5                 EWLQGIAQVA-QGL...VWGG-NLLA--A-----A-L..L-N....-    851
DEN.CD.x.CD1_CMO580407        QQGRATAEV-LAAL...TR-A-EVVA----------I..V-N....-    865
DRL.x.x.FAO                   REGRQLGLSSARAL...RALAQEVAA--------A-V..LFN....-    896
GOR.CM.04.SIVgorCP684con      VWT-DW--Q--AIA...-RIG-G-LN-----------..S--....-    859
GOR.CM.07.SIVgor2139_287      VWT-NW--Q--AIA...-RIG-G-LN-----------..S--....-    855
GOR.CM.07.SIVgorCP2135con     VWT-DW--Q--AIA...-RIG-G-LN-----------..S--....-    850
GRV.ET.x.GRI_677_gri_1        R-AR-AWG-LGAI-...RS-Y--VINS---V-----K..V-G....-    855
GSN.CM.99.CN166               -FTWNW-EA-LHAC...RRVW-EFLA--------A-I..LFN....-    882
GSN.CM.99.CN71                VFTWNW-EPLVQT-...GRVW-EFLA--------A-I..L-N....-    887
LST.CD.88.SIVlhoest447        RRDISAAHARNLRL...GQKK-WRFRFRG-SGFPS-TTETA-....-    916
LST.KE.x.lho7                 RGQLSS---KNIQL...GKKK-WRLRFGG-SGISS-ATETA-....-    913
MND-1.GA.x.MNDGB1             NRIFT-CREA-IAA......GTC-W--L-----SA--..P-N....-    879
MND-2.CM.98.CM16              RGHQW-LSTAACFR....AIA-G-IN--------A-V..L-N....-    867
MND-2.GA.x.M14                GGHQL-LSVIRGAA....AIG---GN--------A-V..L-N....-    868
MND-2.x.x.5440                RGYQL-LSG-RGAA....AFG-G-WN--------A-A..L-N....-    891
MNE.US.x.MNE027               ETL-GAWGDLW-TL...GRVG-W-LA-------E--L..T--....-    883
MON.CM.99.L1_99CML1           RFTIVW--ALLHAG...GRLW---VA--------A-I..F-N....-    870
MUS-1.CM.01.CM1239            -FTWNW-EPILQIG...RRVW-EFLA--------A-I..L-N....-    892
MUS-1.CM.01.SIVmus_01CM1085   TFTWNW-ET-LQAG...RRVW-EFLA----V---A-I..L-N....-    898
MUS-2.CM.01.CM1246            -FTWNW-EPILQFS...-KIW-KFLA--------A-I..L-N....-    883
MUS-2.CM.01.CM2500            -WT-NW--A-LQ-S...KQL*K-FLA--K-----A-I..L-N....-    879
OLC.CI.97.97CI12              .....HGINLQQKK...-RRGSPLQKAQEQLGREGAE.TTGV....-    844
RCM.CM.00.SIVagi_00CM312      ENAYYTWRGLQSLA...RQLAGWPATCG---------..L-N....-    855
RCM.CM.02.SIVrcm_02CM8081     QNAYFVWRGLCVTA...-DIA-WPGTVC------F--..L-I....-    862
RCM.GA.x.SIVRCMGAB1           QNAYYTWRGLCA-A...RDFAGWPAMVC---------..LCN....-    851
RCM.NG.x.NG411                TNAYYTWRGLCA-C...RDPAGWPATLC------F--..F-N....-    858
SAB.SN.x.SAB1                 Q-ATSAS-S-RHAC...RSIV--VIAH---M--E---..WFN....-    868
SMM.SL.92.SL92B               E-IVGAWGLIW-TL...GRVG-G-AA----------L..M-N....-    887
SMM.US.x.SIVsmH635F_L3        ETL-SAWRDIW-TL...GRVG-G-LA----V-----L..T--....-    888
STM.US.89.STM_37_16           ETL-SAGRDLW-TL...GRVG-R-GA----------L..T--....-    883
SUN.GA.98.L14                 RERVSSSQKARSRTFSLGRKW-PKWNRT-GS-IPS-TTETT-....-    920
SYK.KE.x.KE51                 HQMVAIWQALLAYA...RRVAENVAAL---L-----I..Y-N....-    841
SYK.KE.x.SYK173_COMGNM        DW--AIWQAIYAAT...RRVVE-VAAL---L-----I..Y-N....-    833
TAL.CM.00.266                 RFTWWW-NG-LAAA...RW-GE--AA----T-----Y..CFT....-    873
TAL.CM.01.8023                RFTWRW-SG-LAAA...RW-VQ--AA----T-----Y..C-A....-    891
TAN.UG.x.SIVagmTAN1           S-AQSASRTLWNAC...RS-Y---LEH---M--E---..WFN....-    880
VER.DE.x.AGM3                 RLAQNAGYQIWLAC...RS-Y---INS---V-----G..--N....-    878
VER.KE.x.9063                 RFAQNAGHQIWLAC...RSTY-H-ISS---V-----E..--N....-    878
VER.KE.x.AGM155               SFARNAAHQIWLAC...RS-Y---INS---V-----E..V-N....-    871
VER.KE.x.TYO1_patent          GVAQNACHQIWLAC...RS-Y-N-VNS---V-----E..--N....-    864
WRC.CI.97.97CI14              NTQ--VSATGSRKT...AASS-IRMLAT-LCASFDKCRRSFSERRQ-    879
WRC.CI.98.98CI04              TTEK-ISENSNRET...T-SS-VWKL-T-LCSSLNKSRRR-F....-    885
WRC.GM.05.Pbt_05GM_X02        RKEK-IYTPGSRET...SRSSSIRGITA-LCTNLNPCRRGQQ....-    877
```

```
                               myristoylation
                                Nef start                                                                                                         acidic cluster                    poly-P helix  phosphorylation
H1B.FR.83.HXB2                 MGGKWSKSSVIGWPTVRERMR................RAEPAADRVGAASRDLE.....................KHGAITSSNTAATNAACAWLEAQE..............................EEE.VGFPVT.PQVPLRPMTYKAAVDLSHFLKEKGGLEG  99
H1A1.UG.85.U455_U455A          -------K-RVE--E--K---ETPA..................--KG---V-Q--D.....................-Y--V-----SS---S--------E..............................GD..-----R.------------F---F--------D-  99
H1B.US.90.WEAU160_GHOSH        ---I---R-GS---AI----K.....................------EG---V----A.....................----------.EN--D-V--K---..............................D--.-----R.-------------H-----#--------  96
H1C.ET.86.ETH2220              ---TM--C-PV---AI---I-RA...............A---EG--------D.....................-Y--L-----P-N-PD--------E..............................---.-----R.-------------F--L---------  100
H1D.CD.84.84ZR085              ---------IV---AI---I-KTDPRERR.........-P-----G---V-----.....................R----------T-------V----E..............................D--.-----R.--------------------RK------  108
H1F1.BE.93.VI850               ---------IV---A-G---QTPT..................---EG---V---D.....................RR--------RT--PDL-------E..............................---.-----R.---V---------------------  99
H1G.SE.93.SE6165_G6165         ---------IV---E----I-NTPT.................---EG---V-Q--D.....................R-----------N-PD--------E..............................DS-.-----R.---------F-G-F---F--------D-  100
H1H.CF.90.056                  --------RMG--S-I-----RA..................--V-EG---V---D.....................RR--V-IN---S--RDA-------D..............................G--.-----R.-----------G-F-----------D-  100
H1J.SE.93.SE9280_7887          --N-----.....--Q------RAPAP..................----G---V-Q--A.....................----------------D-------TE.............................---.-----K.--I-------G----F--------D-  96
H1K.CM.96.96CM_MP535           ---------IV---AI-----RARPAADRVG.......TQ----G---V-Q--A.....................R---V-----SHN-PD--------E..............................---.-----R.-------------F--GF--------D-  108
H101_AE.TH.90.CM240            --S------IV---Q---KIKQTPP................-TEG---V-Q--D.....................----------.ID--D-V--R---D..............................---.-----M.-----------G-F---F--------D-  97
H102_AG.NG.x.IBNG              ---------IV---K-MK---QTPT.................---TG-----Q--D.....................R----------Q--PD--------D.............................-N.-----R.-----------G------------D-  99
H103_AB.RU.97.KAL153_2         ---------IV---Q----I-RAPAP...................--RG--PV-Q--D.....................-Y--V-------N--D--------E..............................---.-----R.-----------G-F-----------D-  100
H104_cpx.CY.94.94CY032_3       ---------IV---EI-----RARAEPERMRR....AQ-----AG---V-Q--D.....................------IN------PDKT------E..............................---.-----R.---------F-G-L-----------D-  112
H1O.BE.87.ANT70                --NALR-GKFE--AA---------.....RTRT.FP....ES--C-PG--QI--E-A.....................AR-G-P--H-PQN---L-F--SHQEE.............................---.-----A.-----------G-F---F----------  105
H1O.CM.91.MVP5180              --NA----KFA--SE--D----.....RSSS.DP....QQ-C-PG---V--E-A.....................TR-G-S--H-PQN---L-F-DSHKDE............................D..-----R.---------F---F---F--------D-  104
H1O.FR.92.VAU                  --NVLG--KWA--SE--R----.....QTSPDPQ...SSDTQP-PG-R-V--A-A.....................DRRG-P--Y-PQN-Q-L-F-GSHQGK............................G..I----R.-----------G-F---F--------D-  107
H1O.SN.99.99SE_MP1299          --NVLG-DIFK--SA-------.....GTSP.......DP--C-PG--QI--E-A.....................AR-G-P--Y-PQN---L-F--SHQDE.............................---.-----R.-----------G-F---F--------D-  103
H1O.SN.99.99SE_MP1300          --NVLG-DKFK--SA-------.....KTSP.......EP--C-PG--QV--E-A.....................AR-G-SN-H-PQN---L-F--SHQDE.............................D..-----R.-----------G-F---F--------D-  103
H1O.US.99.99USTWLA             --NILT--KFE---A--R----.....QTSP.......DPQ-C-PG--EV--A-A.....................GR-G-PN-Y-PQN---L-F---YQDE............................D..------R.-----Q---F---H---------N---D-  103
H1O.US.x.I_2478B               --NALR--KFA---A-------.....RAS.......PKP-DC-PG--EV--Q-A.....................AR-G-P--Y-PQN---L-F--SHQDE............................D..-------.------------G-F---F--------D-  103
H1N.CM.02.DJO0131              --KI-----IV---EI-----.....RQRPHEPA.....VEP-VG-----Q--A.....................NR--L-T---RTN-PTV--V----EE............................G-.-----R.-----------------F----------  106
H1N.CM.02.SJGddd               --KA-----IV---A------RA...RQTQ.EXAAG...AE--VG---V-Q--A.....................NR-GV-T---KDN-ETV-------EE.............................---.-----C.---X---------F---F----------  108
H1N.CM.04.04CM_1015_04         --KS------V---EI-----.....RQRQ.TQAA....A.--VG-----Q--A.....................NR----T---RDN-ETV--V----EE.............................---.-----R.----------F---F---F----------  105
H1N.CM.04.04CM_1131_03         --KIL--NR-A--SEI-----.....RQTQ.EPAQ.......-----VG-----Q--A.....................NR----T---RDN-ETVT------EE.............................---.-----R.----I---------F---F----------  106
H1N.CM.06.U14296               --KI-----IV---EI-----.....RQTR.ETP.....VEP-AG---V-QE-A.....................TR----T---RDN-QTL-------EE.............................---E-----R.--------------X.............  93
H1N.CM.06.U14842               --KS-----IV---EI-----.....RQTQ..............----XG-----Q--A.....................NR--L-T---KDN-E-V-------EE.............................---.-----R.-------------F---F--D---R--  101
H1N.CM.95.YBF30                --KI-----LV---EI-----.....RQTQ.EPA.....VEP-VGA----Q--A.....................NR----IR--RDN-ESI-------EE.............................---.-----R.--------I---Q-F---F--D------  105
H1N.CM.97.YBF106               --KI-----LV---EI---I-.....RQTP..........EP-VG---V-Q--A.....................NR----T---KDN-QTV-------EX.............................---.-----R.--------------F---F----------  100
H1P.FR.06.RBF168               --NA-K---LV---A-----N.....QTAP........ETX.-PG--EI-KE-A.....................QGKG-R-KYNSKN---L-F---H-DE.............................---.----XR.----X--------F---F-------XD-  101
CPZ.CD.90.ANT                  --SA---IKWV-ARQAIRKIHE..................TNP-DI-PCGNE-A.....................SR--L---TIGTEKDVITYS-DHTEE............................G..T----R.----M----E-L-----X--------DX  97
CPZ.CM.01.SIVcpzCAM13          --AQ-----LV---EI-K-IKEAP..................I--EG--EV-K---.....................-Y-----RH-PS--QTL---DEMKHH.............................---.-----R.-----------G-F------RG-------  101
CPZ.CM.05.SIVcpzDP943          --N------LV---E--A-I-QAQ..................---PG--SV-----.....................A-----TR--PQN-QTL----EMQGH.............................---.-----R.---------F-G-F--GY---------  101
CPZ.CM.05.SIVcpzEK505          --KV-----LV---Q-ID---..RQQ...............DP-EG---V-Q--A.....................NR----IR--KEN-Q-L----E-KEE............................A..-----C.-----------E-F---F--R--------  99
CPZ.CM.05.SIVcpzLB7            --S------LV---E--R-I-QET..................V--EG--P--Q--A.....................R---------PQ--ETI----E-KE.............................---.IW---K.---------F-E-F--GY----------  99
CPZ.CM.05.SIVcpzMB66           ---------LV---E--D-L-RTQ..................T--EG--PV----A.....................R------R--SQ--ETL----EVQN.............................---.-----R.---------F-G-F---------------  99
CPZ.CM.05.SIVcpzMB897          --N------LV---A--D-L-RTP..................---EG--PV-----.....................R------R--PQ--ETL----E-RD.............................---.-----R.-------------F---------------  99
CPZ.CM.05.SIVcpzMT145          --A------LV---EI---IKRAP..................T---G--EV-Q-IA.....................SR--V-IR--PQ--ETL----EMQD.............................---.-----K.---------F-G-F---------------  99
CPZ.GA.88.GAB1                 --T------LV---E--R-I-EAP..................T--EG--EV-K---.....................R------R--PE--QTL----EMDN.............................---.-----R.----T--------F---------------  99
CPZ.GA.88.SIVcpzGAB2           --N------IV---Q----L-RTQ..................E--EG--EV-Q--A.....................RR-----RH-PQ--QTL----EM-Q.............................-D.-----R.-------------F---------------  99
CPZ.TZ.00.TAN1                 --NIFG..RWP-ARKAI-DLHN..................TSSEP--Q--Q--Q.....................NK-GL-TNTLGTSADVLEYSADHTEE.............................---.-----R.-A--M----E-L-I---W--------D-  95
CPZ.TZ.09.UG38                 --NIFG..KWP-AQRAIQAIHE..................SNPTIK-Q--Q--Q.....................AR-GL-T-TLGTSADVINYSHDHSEE.............................---.-----R.-A--M----E-L-----W--------D-  95
CPZ.US.85.US_Marilyn           --N------IV---E--N-L-QTQTT..................A--EG--PV-Q--A.....................E-----TR--PQN-QTL---DEMTNH............................-S-.-----R.-----------Q-F--GF----------  103
                                Nef start    R17Y mutation                                                          premature stop in Mac239

MAC.US.x.239                   ---AI-MRRSRPSGDL-Q-LLRARGETYGRLLG...EVEDGYSQSP-GLDKG-SS.LSCEGQK..YNQGQYMNTPWRNP-EEREKL-YRKQNMDDIDE*DDD..............................L--VS-R.-K----T-S--L-I-M---I--------  130
H2A.DE.x.BEN                   --ASG--KLSKHSRGL---LLRARGDGYGKQRD...ASGGEYSQFQEESG-EQNS.PSCEGQQ..YQQGEYMNSPWRNP-TERQKDLYRQQNMDDVDSDDDD..............................LI-V---.-R--R-E----L-I-M---I------Q-  131
H2A.GW.x.ALI                   --ASG--KRSGPLQGL---LLQTPGETCGGQCS...GSGGGYSQSQ-GSG-GQKL.PSCEGQR..YQQGDFMNTPWRTP-TEREKELYKQQNMDDVDLDDDDS.............................L--VS--.-R-Q--T----L--M--LI--R-----  132
H2A.SN.x.ST_HIV_2_ST           --ASG--KRSEPSRGL---LLQTPGEASGGHWD...KLGGEYLQSQEGSG-GQKS.PSCEGRR..YQQGDFMNTPWRAP-.EGEKGSYKQQNMDDVDSDDDD..............................L--V---.-R----E---RL-R-M--LI---------  130
H2B.CI.x.EHO                   --SAG--KQSKQQ-GL---LLRARRGPRGESSG...E-Q-RSLQYP-GSDKG-NS.PSCDGQK.............TLGAEGGG....KQDSDEDDEDNE..............................--VR-R.-G-------F-L---M-------E---  113
H2B.GH.86.D205_ALT             --SAG--KRSERQQGL--KLLRVPERPYGRLSG...E-R-QSSRSP-ESDK--NS.PSCEGQN.............ARGAEGGG....QQDADESDEDDE..............................--AICKT-I---------L---M---I--Q-----  114
H2G.CI.92.Abt96                --SAG--KRAPQQLGL-K-LLQARGEPYGKLWE...GLE-GYSQYREESGKG-SS.LSCEGQQ..YTQGQFMNTPWRNP-TERAKL-YRXQNNDDVDSDDND..............................L--V--Y.-K----V-S--L-I-M---I--------  131
H2U.FR.96.12034                ---SI--KHSKRSGKL---LLAARGENYGRLWG...GLEDGFELYQ-GSGK--SS.RFCEAQGQGYSEGEFMNTPWRTP-TGKDKLQYKQQNMDDVDEEDND..............................L--VA-R.-R-A--T----L--M---I-----D-  133
COL.CM.x.CGU1                  --SLL-CLNIRQQ-QT-GGVLGRLFGIYWYPGVQAFRFLRHFRGGYDLRRVAGVERLNICEDG........YMTVP-VLPNCPCPLCYRHPDVYYHNPEPED..........................V-DVEA-CY-M.SRL-V-QA-I-LL--I-C----------  133
DEB.CM.04.SIVdeb04CMPF3061     ---RN--TQSS-QEENWGKHFKGGQR.WRYRYLGEESEKFLPYQPESDKELNCS..........................LTEETFIGNPRREIKCTQR-Q.........................QDFDQQ-E-----K.-R--I-DP---LMI-Y----------D  116
DEB.CM.99.CM40                 ----N--IPSR-QVGSFGSGSRGLLR.WRYRDLSEQEEQFSECLLESDREQSSS..........................STE..FWGSPRREIKCTNK-Q.........................QDLDQT-A-----R.---SI-DP---LMI-Y----------D  114
DEB.CM.99.CM5                  ----N--TPSF-QGKKSGSGSKGLLR.WRYKDLSGEAETFSPFLQESDKEQSCY..........................STEEPLY-DPHREIKCTQRAQ.........................ECYD--HGI--A-R.-R--I-EP---LMI-Y--Y-------D  116
DEN.CD.x.CD1_CMO580407         --S-N--QAS.-IWGLAPSNKEGRQRKWYWRLLPEQQEKWWPSPDGSDRALK-SSTRLLEHDED........EETF..VDAEEIIQSSPALDPQHPRIHQY.........................VVDHDD--E-----R.-RR--HAP---DMI-M----------QD  133
DRL.x.x.FAO                    ---QN--KREEAYARFYK-L-R..............GYG..-KGGNLDYHQ--......................PSEPLLR..KWQPFPGGLDKAQRSSSTSSLDTTHMQAAAAEKDVTLPTEEEQPSEE---E-----C.--R---EP---DL--F---------K-  129
GOR.CM.04.SIVgorCP684con       --NA-K---FV---A---------.....QTSP.EP.......DP-PG--EV-IA-A......................QR-G-PAKY-PKN-Q-L-F---H-NE.............................----------V.---------F-G-F---F----------  102
GOR.CM.07.SIVgor2139_287       --NT-K---F----A---------.....QTNP.EP.......KP-PG--EV--A-A......................QR-G-PAKY-PQN-Q-L-F---H-DE.............................-----------R.---------F-G-F---F--N--------  102
GOR.CM.07.SIVgorCP2135con      --NT-K---F----A---------.....QTSP.EP.......KP-PG--EV--V-A......................QR-G-PAKY-PRN-Q-L-F---H-DE.............................-----------X.---------F-G-F---F----------  102
GRV.ET.x.GRI_677_gri_1         --SSN--RQQQ-LLKLWRGL-GKPGADWVLLSD...PLIGQSSTVQEECGKA-KK..SWGK.................GKMTPDGRRLQEGDTFDE-DD.D-EE.............................----------Q.-R---Q---L---F----S---D-  115
GSN.CM.99.CN166                --S-N--QQQQESS-ALLSS-GTGQR.PYFTLVDEYGENFWLSPDASDKG-RYY.........................LTEEPK..PKRGSLEEY-PS..............................C--R.-R----DP---VM--------------  103
GSN.CM.99.CN71                 --S-N--QQSQESS-ALLSSHGTGQS.PSFRLLDEYGENSWLSPDASDRG-RYY.........................LTEE-...PSRQN-IDY-PS..............................C--R.------DP---LM-------------A  102
LST.CD.88.SIVlhoest447         --NIFG-PAAD--WKTLR-L-A..............G-GTRSEGTEETYQQ-M.....................QETSD...........WDKRRDGDSD..........................S--E-----R.--R--CSP---TLI-----I--------  92
LST.KE.x.lho7                  --NAFGRPAAD--WKTLR-F-A..............GSGRN-EGTEDSY-A-I.....................EKTSN...........W-KRRDGDSD..........................S--E-----K.--R--CSP---TLI-----I--------  92
MND-1.GA.x.MNDGB1              --SSQ--KRSEA-VRYSSAL-Q..............LVG..GPVTPDGYKQI-S..................SQGAE-QSLLRGRAYGTYSEGLDKVQNDPLTKD.....EKLDLTQQD...........PE---E-----C.R--S--VPS--DLI-F---I------G-  120
MND-2.CM.98.CM16               ----S--QQEEKYLKYYKA--R..............GYG..-EGTNGDYQQ-H.....................ASEPLLG..ALSTSQEEFDREQKSSS...........................TD---ET----Y.--C-V-EP---DL--MP----------  101
MND-2.GA.x.M14                 ----S--QREQ-YAKYYKAL-R..............GYG..-EGTNLDYQL--.....................EQRLLEG..PLGTSLG-SDKELKSCST........KVD...............DEQK-ER----Y.-TQ-V-EA---NLI-M-----------  105
MND-2.x.x.5440                 ----S--QREQHYIKYYKAL-R..............GYG..-EGTNLDYQV--.....................EQQPLAG..ASGTSL--SDKVLRPYST........ERDLRHQQDVTLPSEKEQPSD---E-----Y.-TQ-VPEA---DLI-M-----------  121
MNE.US.x.MNE027                ---AT--KRSKSPGDL-Q-LLRARGETYGRLWE...GLEDGYSQSL-GSDKS-SL.LSCEGQK..YNQGQFMNTPW-NP-GEREKL-YRKQNMDDIDEEDDG..............................L--V--R.-R----A----L---M---I--------  131
MON.CM.99.L1_99CML1            --S-N--QQSPLSS-PLLGSQSSGRM.RYFMLEDDYGEQSWLSPDASDRE-KYS.........................LTEGR-GKQRRQPLDDDDDDD...........................G--C--R.-R----DP-W-LMM----Y-------GE  109
MUS-1.CM.01.CM1239             --S-N--QQEQQPL-SSPSSPGTGQS.PYFRLVDEYGENSWLSPDASDRG-KYS.........................LTEG-...NERPLIEEH-RS..............................C--R.-R----DP---MMI----Y---------  102
MUS-1.CM.01.SIVmus_01CM1085    --S-S--QPAQQSLISSPPSPGTGRK.QYFKLVEEYGENSWLSPDASGRG-RYS.........................LTEGA...SKRPVIEEH-RT..............................C--R.-R--I-DP---LM--M-----------  102
MUS-2.CM.01.CM1246             --S-N--PQSQERS-PLRSSHGTGPN.RYFSLVKKYGENSWLSPDAS-RG-KYS.........................LTENK...GRRQIIEEH-CS..............................C--R.-K--I-DVN--LMI----Y--K-R----  102
MUS-2.CM.01.CM2500             --S-N--KESQQPLMGSPSGPQTGQT.PYFKLVNNYKKHSWLSPNAS-RG-KYS.........................LTKG-...NKK-IIKKHKRS..............................C--R.--H-IQNP---LIIN------------  102
OLC.CI.97.97CI12               --SIC-RRNRE-EVLY-RLR..................SNWGEKEQKQLE...................SEDQHQ-NLAIESLMTGDTDTDD...........................LE----K.----V-QA---TLIE------------  90
RCM.CM.00.SIVagi_00CM312       ----S--H-QS-LVRW-QK-LTTPGEGYSRWHDNLQDGQPRV-EGS-K----FS.....................IR-GF.TIE-QQSVD-MD-Y-ETDDT...........................L-----K.------A-SF-L-I-M-----------  114
RCM.CM.02.SIVrcm_02CM8081      ----S-----A-LQKW-N--LTSSGEAYALLHKTLQDGPARC-EGS-R----FL.....................SR-G-.TIE-QQSVD-ID-Y-ETDDT...........................L-----K.---------F-L-I------------  114
RCM.GA.x.SIVRCMGAB1            ----S--N-AA-LLRW-FK-LTTPGEGYVRWHETLLDGQPWC-EGS-R----FV.....................IR-G-.TAE-Q-SIDDID-Y-DTDDT...........................L-----K.--------S--L-I-M-----------  114
RCM.NG.x.NG411                 ----S--A-AA-LRRW-GL-LTTPGEDYARFAETLQDGQPRC-EGS-R----FL.....................TR-GF.TIE-QQSVD-ID-Y-DTDDT...........................L-----R.----I---S--L-I------------  114
SAB.SN.x.SAB1                  ----S--QQQRHSLWLWSKL-QAPVIQYDMLAD...PLLGQSSHIQEECAKS-RD..GLIRQG...........DSSRTEEGVKMKHQGRQPS-YDED-EE...........................------R.-CL---A----L-I-FG----------  119
SMM.SL.92.SL92B                ---IT-SKPARKHGDL-KKLLSARGDSYGRLWD...GLE-GSQQSQ-GSD-A-NSCLTRDGYK..YNVGEFMNTPWRNP-TERQKQQYRQQCMDDIDEDDDE...........................LA-VV-W.-K----A-S--L-I-M---I-----D-  132
SMM.US.x.SIVsmH635F_L3         ---AI--KQHRRGGKL---LLRARGETYGRLWE...GLE-GYSQSL--SGKG-SS.LSCEPQK..YSEGQYMNTPWRNPTAEKAKLGYRQQNMDDVDDEDDD...........................L--VS-H.-K----A----L-I-M---I--------  131
STM.US.89.STM_37_16            --ASG--KQRKQHGEL---LLRARGETYGKLLE...GLG-GSGPSQ--SDKG-NS.HSCEPQR..YNEGQFMNTPW-NP-AE-AKLEYRQQNMDDVDEEDDN...........................L--VA-H.-R----E----L-I-----I-S------  131
SUN.GA.98.L14                  --NAFGRP-EV--VRTLF-L-A..............GSGTR-EPA-REYHR-R.....................RQEVEP........LVS-ENGGPNGI...........................-Q-E-E---R.--R--CKP---QLI-----I--------  95
SYK.KE.x.KE51                  --SQS--KQPSR-DEKWRTRWWPFGK.PYSPMPDELL-MSQPYHEDFDRGWRST.........................LTEPILD..PKRDFD.SGGKKWNAGDI.................CHDEGD-DL-----L.-------L---L-I-----I-R---Q-  120
SYK.KE.x.SYK173_COMGNM         --STS---QQLRSEGKYAIGWRLFGK.QYTPLPDELS-PLQPCRGGFDKAWRST.........................LTEPIDPHGPDRDWGHSGG-KFSPGDI.................VQDEGDTGL-----C.--T---TL---L-I-----I-N---Q-  123
TAL.CM.00.266                  --SVK--LLGETASQRLQGSHGGGRM.VYWRLRDGQVKRSQLSPGARGKDLNIALLDEEWQATG........ADYMGKALTQGEVWYEKEQEENDM-LVG...........................---R.-K----TC---LGI-F-F--------K-  123
TAL.CM.01.8023                 --SVK--LLGETASQRLQGSPGAGRL.VYWRLRDGQC-QLQRSRGERGKGLNIALLDAGWQASD........ANWMGKALSEGEVWYEKEQEENDM-LVG...........................---R.--------CN--LGI-F-F--------K-  123
TAN.UG.x.SIVagmTAN1            ---SN--REQQ-LLRLWRAL-KAPVVRYGMLAD...PLIGQSSNIQEECDKNWNG..GSTRR.............GKSTPEGRKL--DDTWDD-EP.E-EE...........................-----R.-R---Q----L---F-----------  116
VER.DE.x.AGM3                  --LGN--PQHKKQLSLWHALHKTRATRYGLLAD...PLIGQSSTLQEECDKA-KE..SLIRKR...........NGKMTPEGRKLQEGDKWDE-SD.E-DE...........................-----R.-R---Q----L---F---------D-  118
VER.KE.x.9063                  --LGN--PQHKKHVSLWHALHKTQVTKYGLLAD...PLIGTSSAVQEEYDKG-RK..SLIRKR...........NGNMTPDGRRLQEGDEWDE-SD.D-DE...........................-----K.-R---Q----L---F---------D-  118
VER.KE.x.AGM155                --LGS--PQHKKQL-IWRALHATRHTRYGLLAD...PLIGQSSTLQEECDKG-RK..SLIRKR...........NGNMTPEGRRLQDGDQWDE-SD.E-DE...........................-----R.-R---QI---L---F---------D-  118
VER.KE.x.TYO1_patent           --SQN--PAHKKYSKLWQALHKTHVTRYGLLAD...PLIGTSSTVQEECDKA-RK..SLIRKQ...........NGNMTEEGRRLQEGDTWEE-SD.D-EE...........................-----R.-R---Q----L---F---------D-  118
WRC.CI.97.97CI14               ---VF--T--EEHMERYA-RLQRAGRRIRRQ......-YQQL-AERQQQAVESGC.....................LLPDSAQALIN-EE-SVKEDSSD-DEG...........................-----.----I-QP-F-LL-----LF-------D  110
WRC.CI.98.98CI04               ---IF--E--EQRMHAYA-RLQRAGRKQRRK......-YQKI-TERQQVAVESGN.....................LLPDSAQALIR-EEDSSKEDSSDDEE...........................-----H.----I-QA-F-LL-----LF-------Y  109
WRC.GM.05.Pbt_05GM_X02         ---VF--D--EQRMLRYR-RLLRVGRKERRR......-YTRL-AERQ-EAAASGE.....................LLPDSAQTLIR-EEDSSKEDSSDDDA...........................I---R.------QA-F-IL-----LF-------D  109
```


HXB2 premature Nef end
normal Nef end
max HIV-1 similarity
normal Nef end