VII

PLV Proteins

Contents

VII-1 Introduction **409**
VII-2 Sequences **410**
VII-3 Alignments **418**
VII-3.1 Gag 418
VII-3.2 Pol 423
VII-3.3 Vif 430
VII-3.4 Vpx 432
VII-3.5 Vpr 433
VII-3.6 Tat 435
VII-3.7 Rev 437
VII-3.8 Vpu 439
VII-3.9 Env 440
VII-3.10 Nef 447

VII-1 Introduction

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

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

VII-2 Sequences

Sequences included in the PLV protein alignments.

Name	Accession	Proteins	Author	Reference
H1B.FR.83.HXB2	K03455	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Wong-Staal, F.	*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)
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)
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H1O.CM.94.BCF06	AB485666	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takekawa, N.	Unpublished
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)

Name	Accession	Proteins	Author	Reference
H1N.CM.02.DJO0131	AY532635	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bodelle, P.	*ARHR* **20**(8):902-908 (2004)
H1N.CM.04.04CM_1131_03	DQ017383	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Yamaguchi, J.	*ARHR* **22**(1):83-92 (2006)
H1N.CM.95.YBF30	AJ006022	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Simon, F.	*Nat Med* **4**(9):1032-1037 (1998)
H1P.CM.06.U14788	HQ179987	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vallari, A.	*J Virol* **85**(3); 1403-7 (2011)
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.06.BF1167	JQ866001	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Li, Y.	*J Virol* **86**(19):10776-10791 (2012)
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.LB715	KP861923	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Barbian, H.J.	*MBio* **2015** Apr 21;6(2) pii: e00296-15 doi: 101128/mBio00296-15
CPZ.GA.88.SIVcpzGAB2	AF382828	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bibollet-Ruche, F.	*ARHR* **20**(12):1377-1381 (2004)
CPZ.US.85.US_Marilyn	AF103818	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*Nature* **397**(6718):436-441 (1999)
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)

Name	Accession	Proteins	Author	Reference
GOR.CM.07.SIVgorCP2135con	FJ424863	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	*J Virol* **83**(4):1635-1648 (2009)
GOR.CM.12.SIVgor_BPID1	KP004989	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	D'arc, M.	*PNAS USA* **112**(11); E1343-52 (2015)
GOR.CM.12.SIVgor_BQID2	KP004991	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	D'arc, M.	*PNAS USA* **112**(11); E1343-52 (2015)
GOR.CM.13.SIVgor_BPID15	KP004990	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	D'arc, M.	*PNAS USA* **112**(11); E1343-52 (2015)
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)
H2B.CI.x.EHO	U27200	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Rey-Cuille, M.A.	*Virology* **202**(1):471-476 (1994)
H2G.CI.92.Abt96	AF208027	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Brennan, C.A.	*ARHR* **13**(5):401-404 (1997)
H2U.CI.07.07IC_TNP3	KC693505	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Ayouba, A.	*AIDS* **27**(15):2488-2491 (2013)
H2U.FR.96.12034	AY530889	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Damond, F.	*ARHR* **20**(6):666-672 (2004)
ASC.UG.10.RT03	KJ461716	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Lauck, M.	*Retrovirology* **11**(1):55 (2014)

Name	Accession	Proteins	Author	Reference
ASC.UG.10.RT08	KJ461715	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Lauck, M.	*Retrovirology* **11**(1):55 (2014)
ASC.UG.10.RT11	KJ461714	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Lauck, M.	*Retrovirology* **11**(1):55 (2014)
COL.CM.x.CGU1	AF301156	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	*J Virol* **75**(2):857-866 (2001)
COL.UG.10.BWC01	KF214240	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Lauck, M.	*Retrovirology* **10**(1):107 (2013)
COL.UG.10.BWC07	KF214241	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Lauck, M.	*Retrovirology* **10**(1):107 (2013)
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.DE.11.D3	KM378563	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	*ARHR* **31**(3); 353-7 (2015)
DRL.DE.11.D4	KM378564	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	*ARHR* **31**(3); 353-7 (2015)
DRL.DE.11.D5	KM378565	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	*ARHR* **31**(3); 353-7 (2015)

Name	Accession	Proteins	Author	Reference
DRL.DE.11.D6	KM378566	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	*ARHR* **31**(3); 353-7 (2015)
DRL.x.x.FAO	AY159321	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	*J Virol* **77**(8):4867-4880 (2003)
GRV.ET.x.GRI_677_gri_1	M66437	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Fomsgaard, A.	*Virology* **182**(1):397-402 (1991)
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.CD.88.SIVlhoest485	AF188115	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	*J Virol* **74**(8):3892-3898 (2000)
LST.CD.88.SIVlhoest524	AF188116	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)
MAC.US.x.251_BK28	M19499	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hirsch, V.	*Cell* **49**(3):307-319 (1987)
MAC.US.x.EMBL_3	Y00295	Gag, Pol, Vif, Vpx, Vpr, Rev, Env, Nef	Franchini, G.	*Nature* **328**(6130):539-543 (1987)
MND-1.GA.x.MNDGB1	M27470	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Tsujimoto, H.	*Nature* **341**(6242):539-541 (1989)

Name	Accession	Proteins	Author	Reference
MND-2.CM.98.CM16	AF367411	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Takehisa, J.	*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)
MON.NG.x.NG1	AJ549283	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Barlow, K.L.	*J Virol* **77**(12):6879-6888 (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)
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)
MUS-3.GA.09.09GabOI81	KF304707	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Liegeois, F.	*Viruses* **6**(7):2880-2898 (2014)
MUS-3.GA.11.11GabPts02	KF304708	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Liegeois, F.	*Viruses* **6**(7):2880-2898 (2014)

Name	Accession	Proteins	Author	Reference
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** Dec;91(Pt 12):2959-64 doi: 101099/vir0025767-0 Epub 2010 Aug 25
RCM.CM.02.SIVrcm_02CM8081	HM803689	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Ahuka-Mundeke, S.	*J Gen Virol* **2010** Dec;91(Pt 12):2959-64 doi: 101099/vir0025767-0 Epub 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.CI.79.SIVsmCI2	JX860430	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	*J Virol* **86**(24):13217-13231 (2012)
SMM.LR.89.SIVsmLIB1	JX860431	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	*J Virol* **86**(24):13217-13231 (2012)
SMM.SL.92.SIVsmSL92A	JX860432	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	*J Virol* **86**(24):13217-13231 (2012)
SMM.SL.92.SL92B	AF334679	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Chen, Z.	*J Virol* **70**(6):3617-3627 (1996)
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)

```
                                   membrane binding                                                                                              phosphorylation site
                                  Gag start, p17 start
H1B.FR.83.HXB2              MGARASV.LSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEAL.DKIEEEQ...............NKS....................KKKAQQAA.............  119
H1A1.UG.85.U455_U455A       -------.---KK--S----------N---R---L--------K-TL-------A---Q--------A----T-----------V--------DV-------.N---M-.................--N....................-QRT----.............  119
H1C.ET.86.ETH2220           ------I.-R-EK--A----K-------H-M---L---N-----K--L--D--D--A--K--IK----A----T---K--F-----------K---------.-------.................-E-....................QQ-T---G.............  119
H1D.CD.84.84ZR085           -------.----K--A----------------------------L-----------K--ME----A-K--------I-----------K---------.-------.................-N-....................--R-----.............  119
H1F1.BE.93.VI850            ------I.----K--E----Q-------R--M--LI----------LD----------QK-IR-------------K--F----V--Y----AGVT------.--L----.................---....................QQ-T----.............  119
H1G.SE.93.SE6165_G6165      -------.-T--K--A----------R-S--I--L-----------L--D----A---Q--MR---------T--IK----A------------V--------.EEV-KI-.................K--....................QE-I----.............  119
H1H.CF.90.056               -------.----K--A-----------------R---L-----------L-------P---L--IE-I--AIK--T---K--F-L--V-----RK-DV-------.-----I-.................---....................QQ-T----.............  119
H1J.SE.93.SE9280_7887       ------I.----K--D---------------RI--L--------D---L------SAK--Q---V----A----TQ-IK---------------------M---.E----I-.................--N....................-QQ--K-E.............  119
H1K.CM.96.96CM_MP535        -------.----K--A-----------------------L-----------L--------T-------IT-I---I------IK------I-V--F---K--V--------.--L----.................---....................QR-T--.E.............  118
H1O.BE.87.ANT70             ---S---.-T-SK--A--Q---K--S----R---L-----------C--E----A--NEKL-Q--E-A-K---DS-Q--W-AIVV-W---N-YK-G--QQ-I.Q-LK-VMG...........................................SRKSAD-.............  116
H1O.CM.94.BCF06             -------.-T-SK--A--R------S--T-----L----------N--CD------SA--NE-L-Q--E-A-K----D-K--W-AI-V-W---N-F-VQN-QQ-I.Q-LK-VME...........................................NRKSAD-.............  116
H1O.FR.92.VAU               -------.-T-TK--A--Q---K--S---------LI----G-D---C--E----A--NE-L-Q--E-A-K----S-*--W-AI-V-W---N-YKVE--QQ-V.R-LK--MK...........................................-RKPVSS.............  115
H1N.CM.02.DJO0131           -------.-T--K--Q--A-Y---------RM--L-----------C----MD-AS--A-L-N--E-A-A----G------AL-V-----SN-QVHN-Q---.---K-IKE..........KKERH-.....................EP-NPE-G.............  124
H1N.CM.04.04CM_1131_03      -------.-T--K--K--A-Y---------RM--L-----------C----MD-AK--A-L-N--E-A-N----G-------L-V-----ST-PVLN-Q---.---K-KKE...............QH-.....................PEA-NPE-G.............  121
H1N.CM.95.YBF30             -------.-T--K--Q--S-Y---------RM--L-----------C----MD-AD--AKL-N--E-A-K-----------AL-V-----S--Q-HN-Q---.---K-K-E...............QH-.....................PEP-NPE-G.............  121
H1P.CM.06.U14788            -------.-T--R--A--R---------------V-----------C----M---A---EKL-E--E-A-R---DG-Q--W--LVV-W---K----S--QQ-I.T-WK--MQ...........................................-RKKTS.............  115
H1P.FR.06.RBF168            -------.-T--R--A--R---K--S----M---------------C--E-M--A---EKL-E--E-A-R---DG-Q--W--LVV-W---R-VD-S--QQ-I.T-WK--MQ...........................................RRKKAP.............  115
CPZ.CD.06.BF1167            -------.-R-EK-----SV-----------LI--L---GN--Q---L----M--AD---K-IM--S--VEI--P-II--F--I-V-H-I-EGEK-Q--EA-V.KIVK-KLK.ITSSE....TTGNEETAKGKKETAVTS..........SGQTEKI.............  139
CPZ.CD.90.ANT               ---G---.-R-EK--T--S------------MI--L---RS--Q---LSSS---------EKAIH--S--IEIR-P-II--F--ICV-W---KGEK----EQ-V.KTVKMKVMQTQAET......GSSQTASRGMLLRLLL..............LNK.............  134
CPZ.CM.01.SIVcpzCAM13       -------.-T--K---------V------R---MM--L-----------C----M-SA---LKL-Q--E-A-K----G----F--L-V-W---SN--VE--QK--.EQLKKHH.............GEQVRKNE........NTENNE..SRQGAS-S.............  130
CPZ.CM.05.LB715             -------.-T-SK--D---------------MM--L--------D---C----M--G----E-LIK--E-A-G----GF---F--LVV-W-------QV--------.-QLA-W-..................Q-Q..........................---.E--.............  117
CPZ.GA.88.SIVcpzGAB2        ------I.-T--R--A------------R-R-MM-------------CD----M--K---Y-L-K--E-A-K----G----F--L-V-W---A---VE--QQ--.Q-LK-FVQK.........ETATAQEAAAK.................QSESNT.............  127
CPZ.US.85.US_Marilyn        -------.-T--R--A---------------MM--L-----------C----M--AD--L-L-K--E-A-K----G----F--L-V-W---S-VTVE--QQ--.V-LK-V...........VQLQKQQEEKE.................QQ--E.............  123
GOR.CM.04.SIVgorCP684con    -------.-T--R--A--R---K--S----M---V-----------C--E-M--A---EKL----E-A-K---DG----W--LVV-W---R--DVS--QQ--.E-FK-AMS...........................................-R.................  111
GOR.CM.07.SIVgor2139_287    -------.-T--R--A--R---K--S----M---V-----------C--E-M--A---EKL----ESA-K---DG----W--LVV-W---R--DVS--QQ--.E-FK-VMS...........................................-R.................  111
GOR.CM.07.SIVgorCP2135con   -------.-T--R--A--R---K--S----M---V-C---------C--E-M--A---EKL----E-A-K---DG----W--LVV-W---R--DVS--QQ--.E-FK-AMS...........................................-R.................  111
GOR.CM.12.SIVgor_BPID1      -------.-T--R--A--R---K--S--T-M---V----K------C--E-M--AD--XKL----E-A-K---DG-Q--W--LVV-W---R-Q-VS--QQ-I.A-WK--M...........................................Q-RKK.................  113
GOR.CM.12.SIVgor_BQID2      -------.---SK--E----------S----L---L-----------C--E----A--TEKL-Q--E-A-K----S-Q--W--IVV-W---N-FKVA--QQ-I.Q-LK-VK...........................................GH-NTS-Q.............  116
GOR.CM.13.SIVgor_BPID15     -------.-T--R--A--R---K-----T-M---V-----------C--E-M--A---EKL----E-A-K----G-Q--W--LVV-W---K--DVA--QQ-I.T-WK--M...........................................Q-RRK.................  113
                                  Gag p15 start

MAC.US.x.239                --V-N--.---KKA-EL-------N-----M---V---AN--D--GLAES---NK---QK--SV-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.QIVQRHLV.............VETGTTET................MP-TSR................  123
H2A.DE.x.BEN                ----N--.-R-KKA-EL--V---------R-------AN--DK-GLAES---SK---QK--RV-D-LVP----N-K--F---CVIW-L-AEEKV---E--K.KLAQRHLV.............AETGTAEK................MPNTSR................  123
H2B.CI.x.EHO                ----G--.---KKT-EL--V--------R-M-------VN-----GLAESR-GSK----K-RKV-G-LVP----N-K------CVIF-L-AEEKV---E--K.KIAQRHLA.............AD...TEK................MPAMSK................  120
H2G.CI.92.Abt96             ----S--.---KKA-EL--V----------M---V---AN--D--GLAES---SK---QK--SV-A-LVP----N-K------CVIW-L-AEQKV-H-E--K.QVVQRHLV.............VETGTTEK................VPATSR................  123
H2U.CI.07.07IC_TNP3         ----H--.---KKT-EL--V---N---R-M---V---AN--D--GLAES---SK---QK--SV-A-LVP----N-K-----TCVIW-L-AEEKV-H-E--K.QLVQRHLV.............VETGTADK................MPATSR................  123
H2U.FR.96.12034             ----N#-.---KKV-EL--V---N------M---V---AN--D--GLAES---AK---QK--SV-A-LVP----N-K------CVIW-L-AEEKV-H-E--K.QIVQRHLA.............VETETAEK................MPATSR................  122
ASC.UG.10.RT03              ----H-AM-T-KK--KY-------K-----QI-------K-----GLSDS---NK---QK--EVIV-LEE----S-K--FGL-SV-W---ADVTVE--E--KKQIRI..................RCHM....................SG-KEETSA............  120
ASC.UG.10.RT08              ----H-AM-T-KK--KY-------K-----QI-------K-----GLSDS---NK---QK--EVII-LEE----S-K--FGL-SV-W---ADVTVE--E--KKQIRI..................RCHI....................SD-KEETSA............  120
ASC.UG.10.RT11              ----H-AM-T-KK--KY-------K---R-QI-------K-----GLSDS---NK---QK--EVIV-LED----S-K--FGL-SV-W---ADVTVE--E--KKQIRI..................RCHM....................SG-KEETSA............  120
COL.CM.x.CGU1               --NEQGL.-GKKT-EDLQ-V--KK-K-GC--I--VR-MCT-VS-CVLIFE--KSAT-VA---EKVT-LVD----V-----G.CSVC--L-RKWN-E--Q--EKKVE-AYKK..............QAMIEMAS...........KEEEKA--EAEK............  129
COL.UG.10.BWC01             --NEQGL.-GKKT-EKLQEVK-KR-K-GC--V--VR-MCT-VN-LCL-IE--KS-Q-IAV--EKVR-LKEA---V-K--WGMLCVC--L-RSWD----Q--ESKVE-AFKKQ..............AMIEQA............EEQERA-AKEEK............  129
COL.UG.10.BWC07             --NEQGL.-GKKTVEKLQ-VK-KK-K-GC--I---R-MCT-V--LQL-IE--KSAQ-IGK--E-V--LVV----V-K--WGMLCVC--L-RGWDVEH-Q--EGRVA-AMKKQ..............AMIEQA............ERTEQQAAQEELEKRQ............  133
DEB.CM.04.SIVdeb04CMPF3061  ---GR-AP-T-KA--KF--V----K-----MI--L---A---D--GLSEA---NK---QE--TR-T-LEA----N-K---G-CCVIWAC-AGLQ-E--E--KKAVKLRIAQ............VEAEDQLEMQVLQ............K-AE-EA-KKQGA........VS  139
DEB.CM.99.CM40              ---GR-AP-T-AA---F--V----K-----MI--L---GK--D--GLAAA---NK---QE--TR---LEA----N-K---GI-CVIWAC-ANLSVD--E--KKAVYRRIKQ............VEQEEELEMALIQ............RQ-EKEVEQK............Q  135
DEB.CM.99.CM5               ---GR-AP-T-AA---F-------K-----LI--L-------D--GLAGA----K---QE--TR---LEAS---G-K---GI-CVVWAC-ANLSVE--E--KKAVARRIKQ............AEQEEELEMALIQ............K-AEKELEKK.............  134
DEN.CD.x.CD1_CMO580407      --SGS--PMT-AA---F--V-----S----MI--LI-SKK-M---GLGEQ----Y---QT-IET-A-LEPQ---S-K--FR-CCVVW-C-RGL-VA--E---REVRKRIGA.............KK-NIVIGNAD............QGDDQPPQGAAG..........G  135
DRL.DE.11.D3                ---S--G.-R-EK--AF--V----S---------L--VAK--D--GLHER---SQ---QK---V-F-LVP----N-I--F--CCVVW---SKVKVT--E--VQKVKQHCHL.......VDK...NENAA...................S-NENG...............  124
DRL.DE.11.D4                ---S--G.-R-EK--AF--V----S-----R---L--VAK--D--GLHER---SQ---QK---V-F-LVP----N-I--F--CCVVW---SKVKVS--E--VQKVKQHCHL.......VDK...NENAA...................S-NENG...............  124
DRL.DE.11.D5                ---S--G.-R-EK--AF--V----S---------L--VAK--D--GLHER---SQ---QK---V-F-LVP----N-I--F--CCVVW-I-SKVKVT--E--IQEVKRHCHL.......VDK...KENAA...................S-NENG...............  124
DRL.DE.11.D6                ---S--G.-R-EK--AF--V----S--R------L--VAK--D--GLHER---SQ---QK---V-F-LVP----N-I--F--CCVVW---SKVKVT--E--VQKVKQHCHL.......VDK...NENAA...................S-NENG...............  124
DRL.x.x.FAO                 ---S--G.-R-EK--AF--V----S---------L--VAK--D--GLHER---SQ---QK---V-F-LVP----N-I--F--CCVVW---SKVKVT--E--VQKVKQHCHL.......VDK...NENAA...................S-NENG...............  124
GRV.ET.x.GRI_677_gri_1      --GGH-A.---RS--TF-------N-----QI--LI--GK-M---GLHEK----K---QK-IEV-T-LEP----G-KA-F-LCCVIW-I-AEQKV---E--VVTVKQHYHL.............VD..KNEKAA.................-KKNE...............  123
GSN.CM.99.CN166             ----H-AM-T-TK---Y--V----K-----MI--L----K----Y-LSDA---NQ----R--EIVF-LEP----A-K--FGI-SV---I-AD-QVE--EQ-KQQVRI..................RCHL....................EETKG.KEKQ............  120
GSN.CM.99.CN71              ----H-AM-T-TK---Y--VS---R-----AI--L----K----Y-LSGS---SK---Q-V-EVTS-LEPN---S-K---GI-SV---I-AD-WVE--EQ-KKQVQI..................RCHL....................VGTKE.KETQ............  120
LST.CD.88.SIVlhoest447      --SGN--.--RQIEGEFCS-----DS--T-QKR-L---T---D--GLGAH----AD--KK---VCW-LYS---KN-KA-VG--CV-C-C-LG-KVA--Q--MQKIKI-PTQ.............................................EER-P..K............  116
LST.CD.88.SIVlhoest485      --SGN--.--RQIEGEFCS------S--T-QKR-----TK--D--GLGAH----AD--KR---VCW-LYS---KN-KA-VG--CV-C-C-LG-K-AN-A--MQKIKI-PTQ.............................................EDR---.K............  116
LST.CD.88.SIVlhoest524      --SGN--.--RQIEKDFCNV-----S--T-QKR-VD--TK--D--GLGAN----A---KK---VRW-LYP---KN-KA-VG--CVIC-C-LG-RVN--Q--INKVKI-QNK.............................................PVPESK..............  116
LST.KE.x.lho7               --SGN--.--RQIEKDFCSV-----S--T-QKR-VE--TK--D--GLGSQ----A---KK--SVCW-LYA---KN-KA-VG--CVIC-C-LGVR-S--Q--ISKVKI-PAP.............................................EA-GK..K............  116
MAC.US.x.251_BK28           ----N--.---KKA-EL-------------M---V---AN--D--GLAES---NK---QK--SV-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.QIVQRHLV.............VETGTAET................MP-TSR................  123
MAC.US.x.EMBL_3             ----NA-.---KKA-EL-------------M---V---AN--D--GLAES---NK---QK--SV-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.QIVQRHLV.............VETGTAET................MP-TSR................  123
MND-1.GA.x.MNDGB1           --NGN-A.-L-TD--KF-----KR----C-R---LC-CKG--D--GLSDK----QQ--EK--SVCW-LYDQ--DN-KA-VG--CVVA-I-AG----S-QD--KKLKVITRK.............................................EEK-...EDE............  117
MND-2.CM.98.CM16            ---S--G.-R-EK--EL-------S-----Q----I-V-K--D--GLHEK---SK---EK--SV-F-LVP----N-I-----CCCVW---AKEKVT--E--KEKVKQKLHL......VAEK....ENAASE...................-EQR................  124
MND-2.GA.x.M14              ---S--G.-R-EK--EL-T-----S-----Q----I-V-K--D--GLHEK----Q---EK--SV-F-LVP----N-LA----CCCIW---AKVKVT--E--KKKVEQCFHL......AAKG....ESAASE...................-EEK................  124
MND-2.x.x.5440              ---S--G.-R-EK--EL-------S-----Q---VI-V-K--D--GLHEK---SQ---EK--SV-F-LVP----N-I-----CCCIW---AKVKVT--E--KEKVKQRYHL......VVER....ENAASE...................EEKGA...............  125
MNE.US.x.MNE027             ----N--.---KKA-EL-------------M---V---AN--D--GLADS---NK---QK--SA-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.QIVQRHLV.............VETGTAET................MP-TSR................  123
MON.CM.99.L1_99CML1         ----H-AM---TK--KY--V----R-----LI------AK--D--GLSDS----QD--KK--EVIL-LQP----SIK--FGIASV---I-AG---E--EQ-KQQVKIRCHL.............................................AGEQG.EQK............  119
MON.NG.x.NG1                ..........................................................................................................................................................................    0
MUS-1.CM.01.CM1239          ----H-AM-T-TK---Y--V----K---R-LI-------K-----GLSDS----K---QK--EV-I-LEP----S-K--FGITSV-W-I-AEV-VE--EQ-RKQIRI..................RCHL....................AD-QGE.E............  118
MUS-1.CM.01.SIVmus_01CM1085 ----H-AM-T-NK---Y--V----K-----LI---I------D--GLSDS---SRD--QK-IEVIL-LEPA---S-K--FGI-SV-W-I-AEVHVE--E--KKQIRA..................RCHL....................EE-KGEKE............  119
MUS-2.CM.01.CM1246          ----H-AM-K-TK-NKY--V----K-----LI-------------GLSDT---SKD--QR--EVIC-LEE----A-K--FGI-SV-W-I-ADVTV---E--KKQIRI..................RCHL....................AE-KEDSKG............  120
MUS-2.CM.01.CM2500          ----H-AM-T-TK--KY--V----K---R-LI---------K--D--GLSDS----K---QR--EVIL-LEP----S-K--FGI-SV---I-AEVSVE--E--KKQIRM..................RCHL....................GE-KE-QEQ............  120
MUS-3.GA.09.09GabOI81       ---KH-AM-T-TK--KY--V----K-----LI---------K--D--GLSDA---SK---QR--EVII-LES----N-K--FGI-SV-W---ADVTVE--E--KRQVRI..................RCHL....................VD-KAG..............QEX  120
MUS-3.GA.11.11GabPts02      ---KH-AM-T-AK---Y--V----K-----LI---------K-----GLSDS----K---QK--EVII-LEP----A-K--FGI-SV-W-I-AEVAVE--E--KKQIRI..................RCHL....................GE-KEE..............EA  119
OLC.CI.97.97CI12            --GTE-K.IDPEL-KKL-RV---ET---C-RRA-VK--VQ--KHLCID-DD-G-K---KR-VEK-L-LEEG-TQK-K--L-ICIVIL-LQEGLPVRN-I--KKRVKVK-..............................................--EE...............  112
RCM.CM.00.SIVagi_00CM312    ------I.-T-KK--A--Q------S----M---LA--C--------GLSDT--DSAG--EK---V-L-LVP----G-K--F-LCCA-W---KEVKV---E--VQYVKKQCHL.......VEKE...DTAGE...................-EK................  122
RCM.CM.02.SIVrcm_02CM8081   -------.---KK--A--Q------S----M---L-----------GLSDT--DSAV--QK---V-L-LVP----G-K--F-LCCA-W---SEVKVR--E--IQYVKKQCHL.......EDRE...DSAAE...................-EK................  122
RCM.GA.x.SIVRCMGAB1         ------L.---KK--A--SV----------M---L---CKK-NK-GLSDH----AT--EK---V-L-LVP----G-K--F-LCCV-W---KEVKV---E--VAKVK-CCHL.......VEKA...ENTTE...................-EK................  122
RCM.NG.x.NG411              ------I.---KK--A--QV----------M---L---C------GLSDT---SA--IQK-I-VIL-LVP----G-K--F-LCCV-W---KEVKV---E--VAHVKKQCHL.......VDKD...ENAGE...................QEK................  122
SAB.SN.x.SAB1               ---SN--.---RK--AF-SV---N--------R-L----K--D--SLSAN----K--VVK--SV-L-LVP----N-IA-F-LCCV-A-I-AE-KV---E--KAKVK--VPA......EMTE....SATATSSGQTK.........ELQA--KNE................  134
SMM.CI.79.SIVsmCI2          ----N--.---KKA-EL--V----------M---L---AN--D--GLADS---SK---HK--SV-A-LVP----N-K--F---CVIW-I-AEEKV-H-E--K.QIVQRHLV.............VETGTADK................MPATSR................  123
SMM.LR.89.SIVsmLIB1         ----N--.---KKA-QL--V----N-R---M---V---AN--D--GLADS---DK---QK--SV-V-LVP----N-K------CVIW-I-AEEKV-H-E--K.QIVQRHLV.............VETGTADT................VP-TSR................  123
SMM.SL.92.SIVsmSL92A        ----N--.---KKA-EL--V----------M---VI--AN--D--GLAES---SK---QR--SV-V-LVP----N-K------CVIW-I-AEMKV-N-E--K.QIVQKHLA.............VENETADQ................LPATSR................  123
SMM.SL.92.SL92B             ----G--.---KKA-EL--V------R---M----I--A---D--GSAES---SK---QR--AV-A-LMP----N-K--FS--CVVW-L-AEMKV---E--K.KTVQSHLV.............VESGTAEK................LPAQSR................  123
STM.US.89.STM_37_16         ----S--.---KKA-EL--V----------M---V---AN--D--GLAES---SK---QK-ITV-E-LVP----N-K--F---CVIW-I-AEEKV-H-E--K.QVVKRHLV.............VETGTANK................MPATSR................  123
SUN.GA.98.L14               ---GN-T.VDREVVRSF-RVA-K-----T-QAR-V---GK--D--GL-KE--R-V---QK--SVCW-LYAS--KN-KA-VG--CV-A-C--G-PV---Q---KKVRL-PAK.............................................NEEA-..A.............  116
SYK.KE.x.KE51               ---GG-AI-T-RK--KY-------K---R-LVR-L---GK--D--GLSDQ---SK---E--IKTIL-LEKH---N-K--FGIT-VVWA--AEKDVE--EQ-KQKVK-ACNW..................EDEET................VTSSGQKENSNDTA.........T  131
SYK.KE.x.SYK173_COMGNM      ---AG-AI-T-R----Y-------K---R-LVR-L---KK--D--GLSDQ-M-SK---EK--TV-L-LEAN---N-K--FGIISVVWA--AKK-VE--EQ-KQKVK-ACNW..................KDDPP................ATSGGQSENSSQNM.......ASET  134
TAL.CM.00.266               ----H-AM-T-RA--KY--V----K-----MV-------K-M---GLSDA----K---VK-IEAIYMLEPN---GIK--FGI-CV---I-AG-D-E--EQ-KAEVRK..................RCHL....................GE-KE.......TTEEEK....  122
TAL.CM.01.8023              ----H-AM-T-RA--KY--V----K-----EVR-V----K-M---GLSDA----K---VK-IEAIS-LEPN---GIK--FGI-CV---I-AG---E--EQ-KAEVKK..................RCHP....................VE-KEESKG...TTEQGA....  126
TAN.UG.x.SIVagmTAN1         ---GH-A.---RN--TF-------N-----Q---LI--GKKM---GLHEK----K---QK-IEV-S-LEP----G-K--F-LCCVIW-I-AEQKV---E--VVLVKQRCHL.......VEKE...KTAA.................APSGGQQQNYN................  129
VER.DE.x.AGM3               ---AT-A.-NRRQ--KF-H-----T-----QI--LI--GK-M---GLHER---SE---KK-IEV-Y-LEP----G-K--F-L-CV-F---KDK-V---E--VAIVRQCCHL.......VEKE...RNAE..........RNTTETSSGQ--NDKG................  134
VER.KE.x.9063               ---AT-A.-K-RQ--QF-------N-----QI--LI--GK-M---GLHDR----ED--KK-IEV-F-LEP----G-K--F-L-CV---I-REQKV---E--VATVRQCCHL.......VEKE...KRAT................EPSSGE--SNRE................  130
VER.KE.x.AGM155             ---AT-A.-NRRQ--EF-H-----N-----QI--LI--GKKMD--GLHEK----E---KK-IEV-S-LEP----GMK----L-CV-L----EKKV---E---AIVRQCCHL.......VDKE...KTAV................TPPGGQQ-NNTG................  130
VER.KE.x.TYO1_patent        ---AT-A.-NRRQ--QF-------N-----QI--LI--GK-M---GLHER----E---KR-IEV-Y-LEP----G-K--F-L-CV---L-KEQKV---E--VATVRQHCHL.......VEKE...KSAT................ETSSGQ--NDKG................  130
WRC.CI.97.97CI14            --SGN-IS-APHICEGLQRVH--KDS----QK-----LVK---KLCLDSA-IS-A--VKKLMSH-A-L-PDA-K-VK-IWGMLCV-A-L-E--PCRN-Q--KEKVNV-Q...............................................--P...............  112
WRC.CI.98.98CI04            --SGN--S-APHICEGLQ-VH--K-S----QK--VI-LVK---KLCLDST-IA-ET-TKKLMEH-A-LVPD--K-VK--WGMMCV-A-L-A--PCTN-Q--KEKVEVKD...............................................T-QKEKP............  116
WRC.GM.05.Pbt_05GM_X02      --SGN-IS-APEICESLQRV---KDS--T--R--L--LVK---KLCLDGK-ISSQ--VKKLCEHIA-LVPE--KTI--IWGMISV-A-LYNN-QVS--QA-IKKIEVRP.............................................RE..EK-............  114
```

```
                                    p17 end p24 start                                                                                   CyPA binding
H1B.FR.83.HXB2              .........ADTG.HSNQVSQNYPIVQNIQ..GQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVH.AGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTN.....NPPIPVGEIYKRWIILGLN  271
H1A1.UG.85.U455_U455A       .........-N--.....S---------A-..--P----L-------------D-----------------------M---V-----------D----------L----.----P-------------------V--------G....-------D--R--------  267
H1C.ET.86.ETH2220           .........-ADR...GKD---------M-..------P--A---------------T---------------------------------D----------L----.----V------D--------------A---G....---V---D----------  269
H1D.CD.84.84ZR085           .........-G--.N-S-----------L-..------L------------I-----N---------------------------------D----------L---Q.---V------------------------S....-------------------  271
H1F1.BE.93.VI850            .........--K-.....----------L-..------SL-----------I----------------------T----------------D----------L----.---AP---------------------Q---G....---V---D----------  267
H1G.SE.93.SE6165_G6165      .........M-K-.N-------------A-..--------T--------------------------------L-----------------D----------M--QQ.---FP---I-------------S----T--G....-------------------  271
H1H.CF.90.056               .........--KE.KD-K----------A-..---------------------------------------A-------------------D---------------.----P--------------------A---G....--A----D----------  271
H1J.SE.93.SE9280_7887       .........T-KK.DNS-----------L-..--P----L-----------I----------------------------I----------D------------------------V-------------------G....-------------------  271
H1K.CM.96.96CM_MP535        .........-A....DKG----------L-..-------L-----------I---------------T-----------------------D---D--------L----.----P--------------------A---S....---V--------------  267
H1O.BE.87.ANT70             .........-KEDTSAR-AG------S-A-..-----------------A------N--I---M------ISY-I-----AI----G-L-V---V-----V----T--PP.V--LP---I---T----------Q----H-T-R....P-Q-----D--RK--V----  270
H1O.CM.94.BCF06             .........-KGETS-K-T-----V-A-A-..-------L---------A------N--I---M------V-Y-I-----AI----G-L-V---V----------T--PA.M--LP---L-D-T----------Q----N-I-R....P-N-V---D--RK--V----  270
H1O.FR.92.VAU               .........TKENTS-K-TG----V-A-A-..-------L---------A--K---N--I---M--------Y-V-----AI----G-L-V-E-V--D---D--KT--PV.V--LP---I---T----------------H-T-R....A-Q-----D--RK--VI---  269
H1N.CM.02.DJO0131           .........-AAA.AD-NI-R---L---A-..---I--PLT--------I------N-------M-------A----------------------V----------T--XP.X--LP---L----------------X--VT---S....----V--------K--V----  276
H1N.CM.04.04CM_1131_03      .........-AAA.AD-NI-R---L---A-..---I--PLT--------I------N-------M-------S---S------------------V----------T---P.V--LP---L----------------P---T---S....---------D--R---V----  273
H1N.CM.95.YBF30             .........-AAA.TDSNI-R---L--TA-..-------PLT-------I--------------M-------S----------------------V-------D---T---P.V--LP---L-D-------------A--VA---A....----V---D--R---V----  273
H1P.CM.06.U14788            .............EG-TGT---------A-..---T-MPL---------A------N--I---M------I-D-I------A----G-L-V---V----------T---P.V--LP---L---T-G--------K----T---R.....-N-V---D--RK--V----  264
H1P.FR.06.RBF168            .............EG-TGT---------A-..---T-LPL---------A------N--I---M-----SI-D-I------A----G-L-V---V----------T--AP.V--LP---L---T-G--------K----T-I-R.....-N-V---D--RK--V----  264
CPZ.CD.06.BF1167            .........YP-LPS-G-TGR-----FDAA..Q-PR--P----------T----N-K------------A---I-H-------AI-D--G---V---V--------D---L--T-.---LV--TL---T-----------IV---Q-V-T..PQNQGGV---D-----------  295
CPZ.CD.90.ANT               .........QWCQR-LSGEGR----IVDAG..-IAR--PLT--------C-----N-N------------------H---------A--D--G---V---V-------------L--T-.---VQA--L---T-----------V---MQ--ST..PQQ-GGV---D---------M---  290
CPZ.CM.01.SIVcpzCAM13       .........SG.......T-G-------A-..------TP-------------------N-----------------L-L-V------A-----G---V-R-V-------------M-----.------------L---------------------A-A.........---D--R---V----  276
CPZ.CM.05.LB715             .........-TAA..DSS--R---V-A-A-..------PL----------------------------M--A-------------------------I---V------S----I--I-.M--------------------------I--------S......-------------V----  268
CPZ.GA.88.SIVcpzGAB2        .........TE.......I-R---V---A-..---I----M-----S----A-------------------M----------V-----AI-----G---V---V-----------------L--L-.---V-----------------------V----S......----V--------R--VV----  273
CPZ.US.85.US_Marilyn        .........-SGS....IG-S---VI--A-..-------M--------------A-------N--------M-----------V-----AI-----G---V---A-------D*---T--L-.------------------------------V-----A......---------D--R--VV----  272
GOR.CM.04.SIVgorCP684con    ..........KSDSRPET-----V-T-A-..------PL----------A------N--I---M-----SI-Y-I-----AI----G-L-V---V-----------Q--AV.V--LPV--L---S-G---------Q---VA---R....N-N-V---D--RK--V----  263
GOR.CM.07.SIVgor2139_287    ..........KSDSRPET-----V-A-A-..------PL----------A------N--I---M-----SI-Y-I-----AI----G-L-V---V-----------M--AV.V--LPV--L---S-G---------Q---VA-I-R....N-N-V---D--RK--V----  263
GOR.CM.07.SIVgorCP2135con   ..........KSDSRPET-----V-A-A-..------PL----------A------N--I---M-----SI-Y-I-----AI----G-L-V---V-----------L--AV.V--LPV--L---S-G---------Q---VA-I-R....N-N-V---D--RK--V----  263
GOR.CM.12.SIVgor_BPID1      ..........AQEG-TGT----------A-..------PL----------A------N--I---M-----SI-Y-I-----A-----G-L-V--DV----------T--AP.V--LP---I---T-G----S---Q----A-I-R.....AN------D--RK--V----  264
GOR.CM.12.SIVgor_BQID2      .........QTDQRS-Q-TG------P-A-..---T--PL----------A------N--I---M-------I-Y-I-----AI----G-L-V---V---------D---T--PV.V--LP---I---A-------------Q---VH-T-R....A-N-----D--RK--VM---  270
GOR.CM.13.SIVgor_BPID15     ..........EKES-PGT----------A-..---I--PL----------A------N--I---M-----SI-Y-I-----AI----G-L-V--DV----------L--AP.V--LP---I---T-G---------Q----A-I-K.............D--RK--V----  264
                                    p15 end p27 start
MAC.US.x.239                .........PTAPSSGRGG..---VQ-IG...-NY--LPL---------LI---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIRDI------D--LQ--QP..A-.QQ--L---S---------SVD---Q--YR....QQN-----N--R---Q---Q  272
H2A.DE.x.BEN                .........PTAPPSGKRG..---VQ-AG...-NY--VPL----------L---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIR-I------D--SQ--IP..--LPA--L-D-------------VD---Q--YR....PQN-V---N--R---QI--Q  273
H2B.CI.x.EHO                .........PSKPTSRL....A--VQ-IA...-NYS-LPL----------L---K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--QQ--SP..--MPA--L-----------------VE---Q--YR....PQN-V---N--R---Q---Q  268
H2G.CI.92.Abt96             .........PIAPPSGRGG..---VQ-VG...-NY--LPL----------L---K-GA--V-G-Q-----CL-Y-I-Q---C--E------IIR-I------X--LQ--QQ..--PPA--X---T-------------IE---Q--THR....QQN-----N--R---Q---Q  273
H2U.CI.07.07IC_TNP3         .........PTAPPSGRRG..---VQ-VG...-NY--LPL----------L---K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--IQN-QP..--LPA--L-----------------VE---Q--YR....QQN-----N--R---Q---Q  273
H2U.FR.96.12034             .........PTXPPSGGRG..---VQ-IG...-NYA-LPL----------L---K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I--D---D--TQ--NQ..--PPA--L-----------------VE---Q--YR....PQN-----N--R---Q---Q  272
ASC.UG.10.RT03              ..............GGGASG------R-A-..--HQ---L---I-KT-ITTI---K--A--VAL-Q-----CI-Y-I-QLM-AI-E--G-V-LI-DVV-----A--QL--QP.-Q-QPNAGL-Y-S------VS-SI---LE---R....AVD-VN-----RK--VM--Q  269
ASC.UG.10.RT08              ..............GGGASG----V-R-A-..--HQ---L---I-KT-ITTI---K--A--VAL-Q-----CI-Y-I-QLM-AI-E--G-V-LI-DVV----HT--QL--QP.-Q-QPNAGL-Y-S------VS-SI---LE---R....AVDQVN-----RK--VM--Q  269
ASC.UG.10.RT11              ..............GGGASG------R-A-..--HQ---L---I-KT-ITTI---K--A--VAL-Q-----CI-Y-I-QLM-AI-E--G-V-LI-DVV-----N--QL--QP.VQ-QPNAGL-Y-S------VS-SI---LE---R....TVDQVN-----RK--VM--Q  269
COL.CM.x.CGU1               ...................LDM.---T.GP.Q-PV.--PL-----G----C--G.GIA-SLA---L-Y-T--IAY-M-L---ILDT--GFP-V--DE--KK-E-Y-LL---Q..Q-QQQ-AL-Q-TA---T-N--SVA--VA-GE..........-IAN--RG--VQS-E  263
COL.UG.10.BWC01             ...................LDM.---N.TP.Q-PA.--PL-----A----C--G.GIA-SLA---L-Y-T--IAY-M-L---ILDT--GFL-V--DE--KK-E-Y-LQN--P..Q-QQQ--L-Q-TA---T-N--SVA--VA-GE..........--AN--RG--VQS-E  263
COL.UG.10.BWC07             ...................EM.---T.G.PQ-P-.-SPL-----A----C--Q.GITASLA---L-Y-T--IAY-M-L---ILDT--GFL-V---EL-KK-E-Y-LT---P..Q-QQQ-AL-Q-TA---M----SVA---A-GE..........--AN--RG--VQS-E  266
DEB.CM.04.SIVdeb04CMPF3061  AQAQPAQVQSAAPQQGTV-RG----LRQG...--Y--MPL---IVKT-INA----K----IV-L-QV-A--C--Y-I-GL--AI-DL-G---II-DV----------LQ--QQ..Q-.PQ--L---S-A-----N--VE---A---RAQGVGGG--D--Q--R--V----Q  303
DEB.CM.99.CM40              QKQQEQRQQEPQAAPQTTP-V----LRQG...Q-F--MPL---IVKT-INA----K----IV-L-QV-A--C--Y-I-GL--AI-DL-G---II-DV----------LQ--QQ..Q-.PQ--L---S-A-----N--VE---A---RPAGQGQG--D--Q--R--V----Q  299
DEB.CM.99.CM5               QKTQ...AA-ADTQQKTNQPA-F--LRQG...Q-Y--MPL---IVKT-INA----K----IV-L-QV-A--C--Y-I-GL--AI-DL-G---II-DV----------LQ--QQ..Q-.PQ--L---S-A-----N--VE---A---RPAGQGAG--D--Q--R--V----Q  295
DEN.CD.x.CD1_CMO580407      GSSAEQGAS-GGSSFRDPN-------RTA-..--YQ---L-T-V-KT-I-E--DRK-----IV-L--V-A--LI-Y-I-QL--AI-E--G---V--DV-----HD--LA--ANLQQ-LP-AAQ-D-------------SIED--T---R....QQG--N--D-----V-M--Q  299
DRL.DE.11.D3                .........ETATSSGRSR..----QVVN...Q-A----------------I---K--A--V---L-----CISY-I-Q---AI-D--G-L-IV-DI------D---R--QV..--LPQ-VL-N-S-----------SIE---E-T-R....AQDS-N--A--RQ-VV---Q  274
DRL.DE.11.D4                .........ETATSSGRSR..----QVVN...Q-A----------------I---K--A--V---L-----CISY-I-Q---AI-D--G-L-IV-DI------D---R--QV..--LPQ-VL-N-S-----------SIE---E-T-R....AQDA-N--A--RQ-VV---Q  274
DRL.DE.11.D5                .........ETATPSGRSR..----QVVN...Q-A-------K--------I---K--A--V---L-----CISY-I-Q---AI-D--G-L-IV-DI------D---R--QV..--LPQ-VL-N-S-----------SIE---E-T-R....AQDA-N--A--RQ-VV---Q  274
DRL.DE.11.D6                .........ETVTSSGRSR..----QVVN...Q-A----------------I---K--A--V---L-----CISY-I-Q---AI-D--G-L-IV-DI------D---R--QV..--LPQ-VL-N-S-----------SIE---E-T-R....AQDS-N--A--RQ-VV---Q  274
DRL.x.x.FAO                 .........ETATSSGRSR..----QVVN...Q-A----------------I---K--A--V---L-----CISY-I-Q---AI-D--G-L-IV-DI------D---R--QV..--LPQ-VL-N-S-----------SIE---E-T-R....AQDS-N--A--RQ-VV---Q  274
GRV.ET.x.GRI_677_gri_1      .........TTAPPGGE..-R---V-NQN...NAW---PL----------C----RWGA--V---Q-----CLSY-V-Q---VI-D--G-L-I---V-----------T-RPP.---LPA--L-D-T-----------SI----E-TF-....A--R-D--AQ-RK-V----Q  274
GSN.CM.99.CN166             .....NKDPPGAAGGQAV------VIR-A-..--YQ---LNS-I-KT--SI----K-A--TVAL-Q-----CI-Y-M-Q---AI-DY-G-V-II-DV---Q-----LL---P..A-QPVAGL-D-S-A----V----PN---E-I-R....Q-Q-VN-SD--RK------Q  278
GSN.CM.99.CN71              .....NKDPPGAAGGQAV--------R-A-..--FQ---LNS-V-KT--SI----K-A--TVAL-Q--T--CI-Y-M-Q---A--DY-G-V-II-DV---Q-----LL---QP..A-QPVAGL-D-S-A----V----PN---E-I-R....Q-N-VN-AD--RK------Q  278
LST.CD.88.SIVlhoest447      ...................-G---L-REN...QRW--TPL----IQT---I--DRGWK--TVA-----T-K-L-D---V---AI-D--G---II-DH-V--G-----K--QQ.QPAQPG-GL-T-NAT-V--V---VE--LA-T-.....ADT-VD--K---E-V-QAME  258
LST.CD.88.SIVlhoest485      ...................-G---LIREN...QRW--TPL----IQT---I--DRGWK--TVA-----T-K-L-D---V---AI-D--V---II-DH-V--G-----T--QQ.QPAQPG-GL-T-NAT-V--V---VE--LA-T-.....ADT-VD--K---E-V-QAME  258
LST.CD.88.SIVlhoest524      ...................-F-LIREN...QRW--TPL----IQT---I--DRGWK--TVA-----T-K-L-E---V---AI-D--G---II-DH-A--G-----Q--QQ.QPAQPG-GL-T-NAT-V--I---VE--QA-T-.....ADT--D--K---E-V-QAME  256
LST.KE.x.lho7               .................Q-TGG---LIREN...QRW--TPL----IQT---I--DRGWK--TVA-----T-K-L-D---V---AI-D--G---II-DH-V--G-----Q--QQ.QPAQPG-GL-T-NAT-V--V---VE--LA-T-.....ADT-VD--K---E-V-QAME  261
MAC.US.x.251_BK28           .........PTAPSSGRGG..---VQ-IG...-NY--LPL---------LI---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIRDI------D--LQ--QP..A-.QQ--L---S---------SVD---Q--YR....QQN-----N--R---Q---Q  272
MAC.US.x.EMBL_3             .........PTAPSSGRGG..---VQ-IG...-NY--LPL---------LI---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIRDI------D--LQ--QP..A-.QQ--L---S---------SVD---Q--YR....QQN-----N--R---Q-R-Q  272
MND-1.GA.x.MNDGB1           ...................-K-F-VQRDAA..--YQYTP----IIQT---T----KWK-----L----T---ISH---I---A--D--G---V--DV---Q-----LT--QQ.QPAQPG-GL-T-S-----------VE--LA--NM....QQNA-N--T---S-----M-  261
MND-2.CM.98.CM16            .........-IVTPSGRSK..----QIIN...QTP---G-----------CI---K----IV---I-----CL-Y---G---AI-D--G-L-IV-DV------D-NLR--QV..--MPQ-VL-N-T-----------SIE---E-T-R....EQDAVN--G---Q---V---Q  274
MND-2.GA.x.M14              .........-TATPSGRSK..----QIIN...QTP---G-----------CI---K----IV---I-----CL-Y---G---AI-E--G-L-IV-DV------D--LR--QV..--LPQ-VL-N-T-----------SIE---E-T-R....QQEQVN--A---Q---V---Q  274
MND-2.x.x.5440              .........-TATPAVRSK..----QVIN...QTP---G-----------CI---K----IV---I-----CI-Y---G---AI-D--G-L-IV-DV------D--LR--PV..--MPQ-VL-N-T-N-------SIE---E-T-R....QQDQVN--G---Q---V---Q  274
MNE.US.x.MNE027             .........PTAPSSGKRG..---VQ-IG...-NYT-LPL---------LI---K-GA--V-G-Q-----CI-Y-I-Q---C--E------IIR-I------D--LQ--QQ..A-.QQ--L---S---------VD---Q--YR....QQN-----N--R---Q---Q  272
MON.CM.99.L1_99CML1         ......AAA-AAPPTGGVP-G---V-RTQG..-GFQ---VE--L-KT--Q-I---K-A---VAL-Q-----MI-Y-I-QL--AI-D--G-L-II-DV------N--L---QP.PQ-QPNAGLGD-T-A----VS--P-Q--E-I-R....A-N-VQ-SD--RK-V----Q  276
MON.NG.x.NG1                ..........................................................................................................................................................  0
MUS-1.CM.01.CM1239          ............KTAQ--T-A----IR-P-..--FQ--SLNA-I-KT--SI----K-AA---VA--Q--A--TI-Y-M-QL--AI-D--G-I-II-DV---Q-----LL--QP.QQ-QPNAGL-S-S-----V----PN---E-I-R....Q-D--N-AD---K---M--Q  269
MUS-1.CM.01.SIVmus_01CM1085 ...........SQAQAA.-A---VIR-A-..--FQ---LNA-I-KT--SI----K-AA---VA--Q--A---I-Y-I-QL--AI-E--G-I-II-DV---Q-----LLN-QP.QQ-QPNAGL-N-T------VS--P-----E-I-R....A-N--N-A----K-V-M--Q  269
MUS-2.CM.01.CM1246          .........-S-AKATTSG----VIRTA-..--YQ--SL---L-KT--ST----K-A---VAL-Q-----CI-Y-V-Q---AI-D--G-I-LI-DVV--Q--DR-VL--QP.TQ-QPNAGL-Y-S-A-V--VS--PA---E---R....QQD-VN--N--RK------Q  273
MUS-2.CM.01.CM2500          .........R-QPSGAGAS-----VIRTA-..-TYQ--SL---I-KT--STI---K-A---VAL-Q-----CI-Y-V-Q---A--D--G-V-LI-DVV--Q--------LL--QP.-Q-QPNAGL-Y-S-A----VS--PA---E---R....QQD-VN--D--RK------Q  274
MUS-3.GA.09.09GabOI81       GGATGGA.TGGVN....TT-------IRTTN..--FQ---LE--L-KS--ST----K-A---VAL-Q--T--LI-Y-M-Q---AI-E--G-L-II-DV---Q--D--LL--QP.-Q-QPNAGL-S-S------VS--PA---E-I-R....Q-E--N-SD--RK------Q  278
MUS-3.GA.11.11GabPts02      KGAEAPSPKGAEAPRPGAS-Y---VIRAAN..--YQ---L-T-V-KS--NTI---K-A--TVAL-Q-----LIAY-M-Q---AI-E--G-L-II-DV---Q--D--ML--QP.-Q-QPNAGL-S-S------VS--PA---E---R....Q-D--N-S---RK---V--Q  282
OLC.CI.97.97CI12            ...................QLAM--IR-A-.GIFE.-VPLAT---QS---AL-DRG-KS-IV-L---FA--LL-EEI--L--V--DY-G---V----A------E-Y--Q--QP.I--LPV--IPP--A---V-----N--Q-LA-VNA....G...-D--AL----V-I--Q  253
RCM.CM.00.SIVagi_00CM312    ........G-TMTSSGQRG..-----TIN...Q-PE--P-----------------K-GA--V---------CI-Y-V-Q---AI-E--G-L-IV--V--D---D--LR--NP..--LPA--L---T-----------NIA---A-T-R....A-N--A--N--RN--V---Q  273
RCM.CM.02.SIVrcm_02CM8081   ........G-TATSSGQKG..----ITVN...Q-PE--P-----------------K-GA--V---------CI-Y-I-Q---AI-E--G-L-IV--V-------D--AR--NL..--LPA--L---T-----------SVA---A-S-R....A-N--I--N--RN------Q  273
RCM.GA.x.SIVRCMGAB1         ........G-TAPPSGQRG..----ITIN...Q-PE-NP-----------------K--A--A---------CI-Y-I-Q---AI-E--G-L-IV--V-------D--AR---P..---PA--L---T-----------SIA---A-T-R....A-N--N--NL-RN---V--Q  273
RCM.NG.x.NG411              ........G-TVTSSGQRG..-----TIN...Q-PE--P-----------------K-GA--V---------CI-Y-V-Q---AI-E--G-L-IV--V--D---D--LR--NP..--LPA--L---T-----------NIA---A-T-R....A-N--A--N--RN--V---Q  273
SAB.SN.x.SAB1               .........PTVTPSGGSR..-----SVN...N-W---PL----------I---K--A--V-------A---I-Y-I-Q---A--E--G-L-IV-DV------D--LR--PP.QQ-P-Q-VL-D-Q-------------I----E-T-R....AQNAVN--N---G-------Q  285
SMM.CI.79.SIVsmCI2          .........PTAPPGGKGG..---VQ-VG...-NYT-LPL---------LI---K-GA--V-G-Q-----C--Y-I-Q---C--E------II--I------D--LQ--NV..--LPA--L---S---------VD---Q--YR....QQN-V---N--R---Q---Q  273
SMM.LR.89.SIVsmLIB1         .........PTAPPSDRRG..---VQ-MG...-NY--LPL--------*--L---K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D*-LQ--QP..S--PA--L------------------VD---Q--YR....QQN-----N--R---Q-R-Q  271
SMM.SL.92.SIVsmSL92A        .........PTAPPSGKRG..---VQ-VG...-NY--LPL----------L---K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--QP..--LPA--L-----------------VE---Q--YR....AAN-----N--R---Q---Q  273
SMM.SL.92.SL92B             .........PTAPPS..GG..---VQ-VG...NNY--TPL----------L---K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--RGQQPAQPA-GL---S-----------PS---E--YR....AQN-V---D--R---Q---Q  273
STM.US.89.STM_37_16         .........PTAPPSGRGG..---VQ-VG...-NY--LPL----------L---K-GA--V-G-Q-----C--Y-I-QL--C--E------II--I------D--MQ--QP.P--LPA--L---S---------SPE---Q--YR....QQN-----N--R---Q---Q  274
SUN.GA.98.L14               ...................---F-VQREG...QNYI--PL----VQT---I----GWK--TVA--A--MT--I-E-I-V---AI-E--G----I-DH-V--S-----M--QQ.QPAQPG-GL-T-N-T----V----VE---Q-IYG..LNGATRVA-QD---G-V-E-ME  261
SYK.KE.x.KE51               SSGREGKMQLPAAMP-SGG-G---LIR-P-..N-WI-VGVNT---KT--EA-NS-K-DASLV-L-QI-T--FI-Y---D---AI-D--G---VI-DV----GS---LQ--QP.QQQQPVAGL-D-SA----------I-----E-I-R....Q-N--Q--Q--RQ-------Q  294
SYK.KE.x.SYK173_COMGNM      SSGQKVVQQEKQKAATPPPRG---LLR-P-..N-WI-TGVPV---KT--EA-NS-K-DASIV-L-Q--T--FI-Y---G---A--D--G---VI-DI-----G-A--LE--QP.QQ-APQAGL-D-SA----------SIA---E-I-R....Q-N-VQ-----R-------Q  297
TAL.CM.00.266               ........VRKATVT-SGQTG----IR-A-..Q-YQ---L---L-KT-IATI---K-A--TVAL----A--CI-Y-I-Q---A--E--G---II-DI---Q-S---QM--Q..I--LPA--L---A-----------PA--VE-T-R....AQN-VD--A-----V-M--Q  276
TAL.CM.01.8023              ........VRSAIVT-SGH-G----IR-A-..Q-YQ---L---L-KT-IATI---K-A--TVAL----A--CI-Y-I-Q---A--E--G---II-DI---Q-----QM--Q..I--LPA--L---A-----------PA--VE-T-R....QQN-VD--S-----V-I--Q  280
TAN.UG.x.SIVagmTAN1         .........TAATPSGR..HG---V--QN...N-W--TPL----------T----R-GA-IV---Q-----CLSY-I-Q---VI-D--G---II--V--D---Q--IT--PP.---LPA--L-D-------------SVA---E-TF-....A--RVD--R--RG-V----Q  280
VER.DE.x.AGM3               .........VTVPPGG...---F-AQ-QG...NAWI-VPL-----------A----K-GA-IV---Q-----C--Y-I-Q---VL-D--G-L-IV--I------Q--IA--PP.---LPA--L-D--------------V---LE-IYT....A--RVD--A--R-------Q  284
VER.KE.x.9063               .........TTAPPGG...---F-AQ-QG...NAW--VPL-----------A----K-GA-IV---Q-----C--Y-I-Q---VL-D--G-L-IV--I------Q--VT--PP.---LP---L-D--------------V---LE-IYT....A--RVD--A--R-------Q  280
VER.KE.x.AGM155             .........GTATPGG...---F-AQ-QG...NAW--VPL-----------A----K-GA-IV---Q-----C--Y-I-Q---VL-D--G-L-IV--I------Q--VT--PP.---LPA--L-D-G------------V---LE-IYT....A--RVD--A--R-------Q  280
VER.KE.x.TYO1_patent        .........IAAPPGG...---F-AQ-QG...NAW--VPL-----------A----K-GA-IV---Q-----C--Y-I-Q---VL-D--G-L-IV--I------Q--VT--LP.---LPA--L-D-------------SV---LE-IYT....A--RVD--A--R-------Q  280
WRC.CI.97.97CI14            ...................---M-VLRTA..Q-PQ.WDPL----IQT---HI-DRG-GA-SV-F-A--TDR-LSH-I--L--NI-D-------II-DIV---IQ----T--QA.-----------Q-T------V-ASVEDELA-LAQ.........-VD--S------VLR--Q  254
WRC.CI.98.98CI04            ...................---M-VLRTA..Q-PQ.WDPL----IQT---HI-DRG-GA-CV-F----TDR-LAH-I--L--NI-D-------II--IV---IQD----T--QP.-----------Q-T------V-ASVEDELA-LAQ.........---D--AV-----VLR--Q  258
WRC.GM.05.Pbt_05GM_X02      ...................-V-M-VLRGP..Q-PQ.WEPL----IQT---HI-DRG-GA-CV-F----TDR-LTH-I--L--NI-D-------IV--IV---IT----T--PP.V--LP--A--N-T------V---ME-ELA-LAQ.....Q---DI-A-------R--Q  256
```

```
                               major homology region                                                          p24 end p2 start       p2 end p7 start                              Zn motif
H1B.FR.83.HXB2             KIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSAT.........IMMQRGN................FRNQR.............KIVKCFNCGKEGHTAR 403
H1A1.UG.85.U455_U455A      --------V--------------------F--------T-D-------------------S--R----G-------------------------------QQTS.........-------................--GP..............RRI----------L-K 397
H1C.ET.86.ETH2220          --------V-----K--------------F--------T-D------D----------------R----G-S--------------A-------------N-TT..........----KS-...............-KGPK............RAI----------L-- 400
H1D.CD.84.84ZR085          --------V-----------------------------------------------------------G----------------S-----------A-SASAA.........----KS-...............-KGT-..............-------------I-- 404
H1F1.BE.93.VI850           --------V-----K--------------F-V--------D--G--D-------------------TG----------------S-----------A.---..........----KS-...............-KG--............RV-----------I-- 397
H1G.SE.93.SE6165_G6165     --------V--------------------F-C--------D--G--D----------------R--QG-S---------------S-----------ASGA-AA............-S-...............-KGP-............RTI----------L-- 404
H1H.CF.90.056              --------V-----K--------------F--------T-D-----------------------R--QG-SI--------------S--------------TN-A..........---K--...............-KG--............-F----------I-- 404
H1J.SE.93.SE9280_7887      --------V--------------------F-A------T-D------D------------------SG--------------------------------TN........................--DHK............R--------Q--I-K 402
H1K.CM.96.96CM_MP535       --------V--------------------F--------T--------D-------------------G-S---------------S----I----------PV.........V---K--...............-KGH-..............-------------I-- 398
H1O.BE.87.ANT70            -M-K----V-----K-----------------------T-------------------Q---S---G-------V--------T----------ATAQQDL....KGGYTAVF----................QNP...IRK.........GTI-----------I-- 407
H1O.CM.94.BCF06            -M-K----V-----K-----------------------T----------------S----Q-------G-------V--------T---KI-----VTAQQEL....KGGYTAVF----................QNP...NRK.........GPI-------D--I-K 407
H1O.FR.92.VAU              -M-K----V-----K-R-E---------Q----------T-------------------Q-------G-S-----V--------T----------ASAQQDL....KGGYTSVF----................QKP...IRK.........GPI---------VTK 406
H1N.CM.02.DJO0131          R-------V---E-K-------------------------D-------------------Q------------------------A-------------Q.T--S.......VFA-GR-...............-KGI-............RTIR---------L-- 408
H1N.CM.04.04CM_1131_03     R-------V---E-E----------------V------T-D-------------------Q-------GSS-------------------------TA-AQ.A--S.......VFV---...............-KGT-............RTI----------L-- 405
H1N.CM.95.YBF30            R-------V---E-K-----------------------T---------------------QL------G----------------A--------------QQPT-S.......VFA---...............-KGI-............-PI----------L-- 406
H1P.CM.06.U14788           -V-K--C-V-----K-----------------------------S--D------------Q-------G-------N--------T----------AAANQASQELK.GGYTTVF--S-Q...............R..................-P--------V--I-K 400
H1P.FR.06.RBF168           -V-K----I-----K-----------------------------S--D---I--------Q-------G-------N--------T----------AAANQANQDLK.GGYTTVF--S-Q...............R..................-P--------I--I-K 400
CPZ.CD.06.BF1167           -L-K----V----------------------I-------P--------------------------G------L----------A----------Q-TQQNVN..........-V-GRA...............APQGKRGA...........GNI-------I----- 429
CPZ.CD.90.ANT              -V--X---V---E-K-----------------I-------P--A------I--------H-------TG-S----L----------A----------ASAN-AQGT......AVFL---...............GNRGGK............RPL------------- 425
CPZ.CM.01.SIVcpzCAM13      RV-K--C-V-----K-------K---------M---------T--D-------------Q-------G-S----------------A----------TM-QSQ-RT......D-FF-K-P...............GATP-............RKI--Y-------L-- 410
CPZ.CM.05.LB715            --------V-----K--------------------------T----T--D-------------N--R---G----------------A-------------ANATV.........................-KGP.............KRTI--Y-------L-- 398
CPZ.GA.88.SIVcpzGAB2       -V----C-V-----K----------------V------D-------------------Q--R---G----------------A-------------IQQTS.........S-F---AG...............G-TPP............RR-----------I-- 405
CPZ.US.85.US_Marilyn       -V-K--C-VG----K---------------------P---T----------------------R----G------V-------A----------C-MK-PS.........SVFL-K--...............AGKPG...........RKI----------L-- 404
GOR.CM.04.SIVgorCP684con   -V-K--C-V-----K-----------------------D---------I--------Q-------G-------N--------T----------AAAHSDLK....GGYTAVF--K-G...............NNQ..GI..........KKGP--------V--I-K 402
GOR.CM.07.SIVgor2139_287   -V-K--C-V-----K-----------------------D---------I--------Q-------G-------N--------T----------AAAHSDLK....GGYTAVF--K-G...............NNQ..GI..........KKGP--------V--I-K 402
GOR.CM.07.SIVgorCP2135con  -V-K--C-V-----K-----------------------D---------I--------Q-------G-------N--------T----------AAAHSDLK....GGYTAVF--K-G...............XNQ..GI..........KKGP--------V--I-K 402
GOR.CM.12.SIVgor_BPID1     -V-K----I-----K-----------------------D---------X--------Q-------G-------N--------T----------AAANQASQDLK.GGYTAVF----...............Q-................GP--------V--I-K 400
GOR.CM.12.SIVgor_BQID2     -M-K----V-----K----------------------T---------------------Q---S---G-------I--------T----------AS-HQDL....KGGYTAVF----...............QNP...NKR.........GPI-----------L-K 407
GOR.CM.13.SIVgor_BPID15    -V-K----I-----K-----------------------D------D--------------Q-------G-------N--------T----------AAANQASQDLK.GGYTAVF----...............Q-................GP--------L--I-K 400
                                                                                                          p27 end p2 start       p2 end p8 start

MAC.US.x.239               -C----N--N---VK-------QS-------S-----TDAA------Q---I--------LV--G--VNP-----L----------Q---LM---LKEALAPVPIP......FAAA-Q..................-GP.............R-PI--W-------S-- 404
H2A.DE.x.BEN               -C--K-N--N----K-------QS-------S-----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEAMGPSPIP......FAAA-Q.......................................R-AIRYW-------S-- 402
H2B.CI.x.EHO               -C----N--N----K-------QS-------S-----TDPA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LKEALTPS-NP......FAAA-PR.................AG..............KRT-T-W----A----- 400
H2G.CI.92.Abt96            -C----N--N---VK-------QS-------S-----TDPA------Q---I--------LV--G--MNP-----L-----X----Q---LM---LKEAL-PTALP......FAAA-QK.................TGG.............KRSTI--W--------V- 407
H2U.CI.07.07IC_TNP3        -C----N--N---VK-----S-QS-------S-----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEALAPGPLP......FAAV-QR.................TGG.............KKSMI--W-------S-- 407
H2U.FR.96.12034            -C----N--N----K-------QS-------S-----TDPA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LKEALQPMPIP......FAAV-QR.................GGP.............RRP-R-W---------- 405
ASC.UG.10.RT03             RC-K--N-VN----------S-K--------C------D-S------T---------E--A---M.-G-S----L---------HY-TKL-----ASAF.QK-GN......YN-V-GP..................-GPKG....ARRGPMKNPNPR-----QF--MSK 409
ASC.UG.10.RT08             RC-K--N-VN----------S-K--------C------D-S------S--I-----E--A---M.-G-S----L---------HY-TKL-----ASAF.QR-GN......YN-A-GP..................QGPKG....ARRGPMKNPNPR-----RF--MS- 409
ASC.UG.10.RT11             RC-K--N-VN----------S-K--------C------D-S------T---------E--A---M.-G-S----L---------HY-TKL-----ASAF.QK-GN......YN-V-GP..................-GPKG....ARRGPMKNPNPR-----QF--MS- 409
COL.CM.x.CGU1              -VIQIAR-S-V------S--D-KS------SA----P-AG-I-A--ANN--I-H------R---G-.QKPS--D-LA--------D---K------Q-FQQER........TN-IE.......................................V-TA-----QGI--L-- 387
COL.UG.10.BWC01            -VIQIAR---V------S--D-KS------TA----P-GS-I-Q--ANN--I-H---E--R---G-.QKPS--D-LA--------DY--K------Q-FRQE.........TMN-IE.......................................V-TA---Q--EI--LQK 387
COL.UG.10.BWC07            -VIQIAR-S-V------A--D-KS------FSA----P-GG-I-Q--ANN--I-H---E--R---G-.QKPS--D-LS--------DY--K------QGMQKIG............LVE.......................................VSTAR-----QV--L-- 388
DEB.CM.04.SIVdeb04CMPF3061 RC-K--N--N---VK-------K----------------D-A------T--MI-------R---------NP-----LH--------QQ---LM----VSAMREGGG.......LT-V-QR..................GP........GGPRPGNRIR-Y---QI--VQK 439
DEB.CM.99.CM40             RC-K--N--N---VK-------K----------------D-A------T--MI-------RI---G--QNP-----LH--------QQ---LM----ASALKE-GS.......LG-V-QR..................-GP........PGS..RRQIR-----QI--LQK 434
DEB.CM.99.CM5              RC-K--N--N---VK-------K----------------D-A------T--MI-------RI---G--QNP-----LH--------QQ---LM----ASALKE-GS.......LG-V-QR..................GG-........GGG.PRRQLR-----QI--VQ- 431
DEN.CD.x.CD1_CMO580407     -V-K--Q-V-----K-----S-K--------A------D-S--L---TS-----------Q--E----NP-----LQ---------Q---PT---LASAFQKQGLC......HGSRKGAQPPVER..........................RGLR-----QI--L-K 434
DRL.DE.11.D3               RC-T--N-VN---VK-------K--------A----HTDAA------Q---I---------V---G--TNPS----LL---------------M----R...EQQ-A...........----QN..PP...................-GPP.....RGPRGPPPRNPR-P---QY---L- 412
DRL.DE.11.D4               RC-T--N-VN---VK-------K--------A----HTDAA------Q---I---------V---G--MNPS----LL---------------M----K...EQQAA...........----QN..SL...................-GPP.....RGPRGPPPRNP--P---QF--NL- 412
DRL.DE.11.D5               RC-T--N-VN---VK-------K--------A----HTDAA------Q---I---------V---G--MNPS----LL---------------M----K...EQQAA...........----QN..PP...................-GPP.....RGPRGPPPRNPR-P---QF--.L- 411
DRL.DE.11.D6               RC-T--N-VN---VK-------K--------A----HTDAA------Q---I---------V---G--MNPS----LL---------------M----R...EQQ-A...........----QN..PP...................-GPP.....RGPRGPPPRNPR-P---QY---L- 412
DRL.x.x.FAO                RC-T--N-VN---VK-------K--------A----HTDAA------Q---I---------V---G--MNPS----LL---------------M----R...EQQAA...........----QN..PP...................-GPP.....RGPRGPPP-NPR-P---QF---L- 412
GRV.ET.x.GRI_677_gri_1     -V-Q--N-QKV------------Q--------A------P-D------Q---I--------L---G--MNP-----LI---------Q---KLMV-M--NGQ..............N-V-....V...................GPQ.......KKGP.RGPL-------F--MQ- 403
GSN.CM.99.CN166            RC-K--N-VN----K-------KE-----F-C------D-A------QS----------E--L----M.-G-N----L---------T--S-L-----ATALKGTSS......YN-V-....K...................GPP......GRQG...-TP--Y---QF--L-- 411
GSN.CM.99.CN71             RC-K--N-VN--G-K-------KE-----F-C------D-A------QL----------E--L----M.-G-S----L---------T--S-L-----ATALKG.SS......YN-V-....R...................GPP......GRQG...-TP--Y----F--I-- 410
LST.CD.88.SIVlhoest447     -V--IHQ-V-VI--K-------KE-A---F-A----GG-H---E--K-KM---------RMVI----EG-------R---------A--G-I-----ATAMQ-QMR.......QN-V-V.TPPRNAQ...........GRFVRT...GGGG.PR-PLT------P----- 404
LST.CD.88.SIVlhoest485     -V--IHQ-V-VM--K-------KE-A---F-A----GG-H---E--K-KM---------RMVI----EG-S--D--K---------A--GKI-----ATAMQ-QMR.......QN-V-V.TPPRNEQ...........GRFVRT...GGGG.PR-PLT------P----- 404
LST.CD.88.SIVlhoest524     -V--IHQ-V-VM--K-------K--A---F-A----GG-H---E--K-KM----------MVI----EG-------K---------A--RKI-----ATAMQGQMK.......HL-V-Q.TPPRNAQ...........GRFVRT...GGGG.PRRPLT------P--S-K 402
LST.KE.x.lho7              -V--IHQ-V-VM--K-------KE-A---F-A----GG-H---E--K-KM---------RLVI----EG-S-----K---------A--GKI-----ASAIQQQVR.......QN-V-V.TPLRNAQ...........GRFVRT...GGGG.PR-PLT------P----- 407
MAC.US.x.251_BK28          -C----N--N---VK-------QS-------S-----TDAA------Q---I--------LV--G--VNP-----L----------Q---LM---LKEALAPVPIP......FAAA-K..................-GP.............R-PI--W-------S-- 404
MAC.US.x.EMBL_3            -C----N-IN---VK-R-----QS-------S-----TDAA------Q---I--------LV--G--VNP-----L----------Q---LM---LKEALRPVPTP......FAAA-Q..................-GP.............R-PI--W-------S-- 404
MND-1.GA.x.MNDGB1          RL-KSHC-I--T-V------A-K-------NVM------G---M--QQH--IE----E--Q--RS--KG------LE--------Q----LM--M-RT-VGQ..S......QNFV-QRGPQR....GPVRQP....TG..............R-PI.-----N----V-- 399
MND-2.CM.98.CM16           -C-S--N-VN----K-------K--------A----RTDPQ--T---Q---I--------S---G--MNPS----PL--------KY--QMM----K...EAQS........AV---NSGGPP.................-G..........PPRQPPRNIR-P----F--GL- 410
MND-2.GA.x.M14             -C-S--N-VN----K-------K--------A----RTDPQ--T---Q---I--------S---G--MNPS----LL--------KY--QMM----Q...KMQS........EV---NSGGPP.................-G..........PPRQPPRNPR-P---RY--VL- 410
MND-2.x.x.5440             -C-S--N-VN----K-------K--------A----RTDPQ--T---Q---I--------AT--G--MNP-----LL--------KY--QMM----Q...QAQAA........V---NSGGPP.................-G..........PPRQPPRNPR-P----F--VL- 410
MNE.US.x.MNE027            -C-K--N--N------------QS-------S-----TDPA------Q---I--------LV--G--MNP-----L----------Q--KLM---LKEALAPGPLP......FAAA-Q..................KGP.............R-PI--W-------S-- 404
MON.CM.99.L1_99CML1        RC-K--N-VN----K-------K--------C-----TD-A------Q-------------L----M.-G-S--D-LQ--------M--S-I-----AGAIAN........MPMN-V-ARGPPQ.......................................RKG.QP-------F--M-K 408
MON.NG.x.NG1               .................................................---------L----M.-G-S----LQ--------A--S-L-----ATAI--N........MPMN-V-GRGGXQPR.....................................RQGXQIR-Y----F--V-K  81
MUS-1.CM.01.CM1239         RC-K--N-V-----K-------K--------C-----SD-A------S---------E--L---SM.-G-------LQ--------A--S-LM----ASAL.RQNS.....QL.N-V-G...ARGK..........GSQGGP..............RGNPR-Y---QF--M-- 406
MUS-1.CM.01.SIVmus_01CM1085 RC-K--N-VN----K-------K------F-C-----SD-A------S---------E--L---SM.-G-------LQ--------T--S-LM---LASAF.KQAG.....TL.N-V-GAKGARGR........GNGSQGN..............RGNPR-Y---QY--V-- 410
MUS-2.CM.01.CM1246         RC-K--N-VN----K-------K------F-C------D-A------QS--------E--LV---M.-G---K--LQ----IE--M--S-LM----TNAL.RQNTI....NTIN-V--...QSPR.........GVMGKK-.............ENSTR-Y---QF--L-- 413
MUS-2.CM.01.CM2500         RC-K--N-VN----K---R---K------F-C------D-A------QS--------E--LV---M.-G-------LQ--------I--SKLM----ASAL.RQNTI....ETIN-V-G...QGPR.........GGAGGGP.............RRTPR-----QF--I-K 414
MUS-3.GA.09.09GabOI81      RC-K--N-VN------------KE------C-----TD-A------T---------E--A---M.-G-------LQ--------AS---L-----VDAF-KTGAT.......A-V--AA.PIPR...............GPN.............NRQPR-----QF--L-K 414
MUS-3.GA.11.11GabPts02     RC-K--N-VN------------K--------C-----TD-A------S---------E------M.-G-------LQ--------AQ---L-----VSAF-KSGAS......TA-----P.PPGRNP..............Q.............KRAPR-----QF--L-- 419
OLC.CI.97.97CI12           -L-K----VTV------A----S------F-A-----G-----T----K--I--------GV-----.VKPS----L----------Y--------LRP-Q.................V-QRPAP..................GGGPQ............QGRLR-Y---RF---KN 382
RCM.CM.00.SIVagi_00CM312   -C-K--N-VN----K-------K--------A------DPA------QS--I---------V--G--MNP-----L---------QY---------Q...QANSA.........--A-MRGGQR...............SGP.......PRGQGGKTFL--Y----L--IS- 411
RCM.CM.02.SIVrcm_02CM8081  -C----N-VNV---K-------K--------A-----TDPA------Q---I---------V--G--MNPS----LL--------Q----------Q...HANSM........M-A-MRGQGR...............SGP.......PRRGGERPVP--Y----T--IS- 411
RCM.GA.x.SIVRCMGAB1        -W-K--N-VN----K-----S-K--------A------DPA------QS--I--------MV--G--MNPS----L---------Q----------Q...MMQSN.........--A-QSAN...................-GP.......PRRSGGNPNLR-Y----P--IS- 409
RCM.NG.x.NG411             -C-K--N-VN----K-------K--------A------DPA------QS--I---------V--G--MNP-----L-----A---Q----------Q...MVQ.........SN--A-Q-.GPR...............KGP.......PKLGGGPRFL--Y----T----- 410
SAB.SN.x.SAB1              -C-K--N-VN----K-------K--------A-----TDPA------QS--I---------V--G--MNP-----L-----I--AQ----LM----TAAFQQQ-V......GN-FV-Q-.ARP..................-GP.....LGGRGRPLNPNI--Y----P--L-- 429
SMM.CI.79.SIVsmCI2         -C----N--N---VK-------QS-------S-----TDAA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LREAFQPGPLP......FAAA-Q..................-GGGVR............RPI--W-------S-- 407
SMM.LR.89.SIVsmLIB1        -C----N--N---VK-------QS-------S-----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEAFQP-SLP......FAAV-QR..................GGTK.............RT-R-W-R-----S-- 404
SMM.SL.92.SIVsmSL92A       -C----N-VN----K-------QS-------S-----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEALGPTQLP......FAAA-QK..................G..P.............SK-R-W-------S-- 404
SMM.SL.92.SL92B            -C----N--N---VK-------QS-------S-----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM----KDALTGSLV.........AA-FR..................GAAKGQ........GNKP-IR------T--S-- 408
STM.US.89.STM_37_16        -C----N-VN---VK-----T-QS-------S------DPS------R---I--------LV--G--MNP-----L----------Q---LM---LKE-FQPDPLP......FAAA-QQ..................G...............RRT---W----A----K 405
SUN.GA.98.L14              RV--LHQT--VVE------------T---F-A----SG-E---E--K-KM----------LVI----ETPS-----R---------S--GKI-----ASAFRQQVG......RQA-V-QNLPPRNSQ...........G-FV-I......GGGGPR-PMT-------P--L-N 409
SYK.KE.x.KE51              -C-QV-N-VN----------T-K-------HC------DPA----L-QS--I-----E-RS----MVK.P------LQ--------QY--KL-----VMTQH-L...........G-I-GPRQGSNP......................................RRGPTR-----QL--LQK 426
SYK.KE.x.SYK173_COMGNM     -C-QV-N-V-----K-----I-K-------HC------DPS--G-L-Q---I-----E-RQ----MVK.P------LQ--------L---KLM----VMAQQ-V...........N-V-GPSKG........................................RSMI--Y---QI--MQK 425
TAL.CM.00.266              RC-K--N-VN---VK---R---K-------RC-----TD-A------Q---------E--A----M--G------LQ--------Q--S-LM----AEALK.QNTQ....QVIA-V-Q..GQGGP.........RGRGGP-R.........NPPGQIR-Y----F--I-K 421
TAL.CM.01.8023             RC-K--N-VN---VK---R---KE------RC-----TD-A------Q---------E--A----M--Q------LQ--------Q--S-LM----AEALK.QNTQ....QVIA-V-...GQGGP.........KGRGGP-R.........TPPGQIR-Y----F--I-K 424
TAN.UG.x.SIVagmTAN1        -C-K--N-I-V------A----K-------QA-----TP-D----------I--------LV--G--IHP-----L-------------KLMV---Q-MQG..........VN-V-G....APR...............GGRG..........RGRGPPR--K--QI--IQK 413
VER.DE.x.AGM3              -C-K--N-V-V---------A-K--------AI------G---Q----S--I---------V---G--MHP-----L---------SY--K-M--M-QNMQSQ...........N---Q....GGQ...............-GRP.............RPP---Y----F--MQ- 415
VER.KE.x.9063              -C-K--N-V-V-----------K--------AI------G---Q----S--I---------V---G--MHP-----L---------SY---M--M-QNMQTQ...........N-V-Q....GAA...............GGV-.............RQRPPL--Y----F--MQ- 413
VER.KE.x.AGM155            -C-K--N-V-V-----------K--------AI------G---Q----S--I---------V---G--MHP-----L---------SY--K-M--M-QNLQSQ...........N-V-Q....GGG...............-GRP.............RPPP--Y----F--MQ- 411
VER.KE.x.TYO1_patent       -C-K--N-V-V-----------K--------AI------G---Q----S--I---------V---G--MHP-----L---------SY--K-M--M-QTMQ-Q...........N-V-Q....GGP...............KR-..............RPPLR-Y----F--MQ- 410
WRC.CI.97.97CI14           -V-QI-V-V-----K-----D-KA-------V-----GAP---Q---DK--I-------------G-.QQP-----L----------Y-SKIM-Q-LQEALH-Q..........A-V-QAGAKP.......................................KGPIR-----QI--M-K 383
WRC.CI.98.98CI04           -V-QI-V-V-----K-----D-KT-------V-----GAP---Q---DK--I--------A---G-.QQP-----L----------Y-SKIM-Q-LQEAL.-.SQ..........A-V-QQPVKP.......................................KGPIR-----QI--M-K 387
WRC.GM.05.Pbt_05GM_X02     -V-QI-V-------K-----D-KS-----F-V-----GVP---Q---DK--I--------Q---G-.QSP-----L----------Y-SKIM-Q-LQEAL.-QSQ.........A-V-QQAVK........................................RGPLR-----QI--M-K 385
```

```
                            Zn motif                             Zn motif                 p7 end p1 start              p1 end p6 start                         Vpr binding
H1B.FR.83.HXB2              NCRAPRKK.............................GCWK.CGKEGHQMKDCT............ERQANFLG.KIWP..SYKGR................PGNFLQ......................SRP...EPTAPP..........................EES  462
H1A1.UG.85.U455_U455A       --------.............................----.------------............--------.----..-N---................----P-......................---...------..........................A-I  456
H1C.ET.86.ETH2220           --------.............................----.------------............--------.RL--..-N---................------......................---...------ESLRPEPTAPP...............P--  470
H1D.CD.84.84ZR085           --------.............................----.------------............--------.----..-H---................------......................---...------..........................A#-  462
H1F1.BE.93.VI850            --------.............................----.--R---------............--------.----..-N---................------......................---...------..........................A--  456
H1G.SE.93.SE6165_G6165      --------.............................----.------------............--------.----..-N---................------......................N-T...------..........................A--  463
H1H.CF.90.056               --------.............................----.--R---------............--------.----..-S---................------......................---...------..........................A--  463
H1J.SE.93.SE9280_7887       --------.............................----.------------............--------.----..-S---................------......................---...------..........................A--  461
H1K.CM.96.96CM_MP535        --------.............................----.------------............--------.----..-H---................------......................---...------..........................A--  457
H1O.BE.87.ANT70             --------.............................----.--Q-------RN............GK------.-Y--..PGGT-................---YV-......................RPA...H-S---..........................M-E  467
H1O.CM.94.BCF06             ------RR.............................----.--Q-------KN............G-------.-Y--..PGGT-................---YV-......................KQV...S-S---..........................M-E  467
H1O.FR.92.VAU               --------.............................----.--Q-------KN............R--T--K.-Y--..PRGT-................---YV-......................EQV...S-S---..........................M-E  466
H1N.CM.02.DJO0131           --K---RG.............................----.--Q-------KN...........EG-------.-G-S..PF---................----P-....................TTTT-K...------..........................L-N  472
H1N.CM.04.04CM_1131_03      --K---RG.............................----.--Q-------KN...........EG-------.-G-S..PF---................----P-.....................TTT-K...------..........................L--  468
H1N.CM.95.YBF30             --K---RG.............................----.--Q-------KN...........EG-------.-S-S..PF---................----P-.....................TTT-K...------..........................L--  469
H1P.CM.06.U14788            --K---RR.............................----.--Q-------KS............G-------.----..PGGK-................---YV-......................KQV...Q-----..........................M-E  460
H1P.FR.06.RBF168            --K---R-.............................----.--Q-------KS............G-------.----..PGGK-................---YV-......................KQV...Q-----..........................M-E  460
CPZ.CD.06.BF1167            ------R-.............................----.--Q----L-----.........QSKSGV----.-RT-..WWGC-................-----V-.NQSQTQNQEQSQTQEKLV...EKA...--S---.......................MEEVNH  511
CPZ.CD.90.ANT               --K---R-..........................---R.--Q----L-N-PAT.........NTGKV----.-PT-T.WWGC.............R----V-....................KEEVV...------..........................I-I  490
CPZ.CM.01.SIVcpzCAM13       --K---R-.............................----.--Q-------S............G--V----.-G--..-RS--................----V-......................N-T...------..........................A--  469
CPZ.CM.05.LB715             --K---R-.............................----.--Q---R-Q--KLS............KV-----.-----..LNS--................----P-......................---...------..........................M--  457
CPZ.GA.88.SIVcpzGAB2        --K---R-.............................----.------------QN...........EG--V----.GG--..-RS--................----P-......................RKV...------..........................P-T  466
CPZ.US.85.US_Marilyn        ------R-.............................----.--Q----------AG............N-------.-H-S.P-WS-G.............SKR-----E.......................N-K...------..........................I-D  469
GOR.CM.04.SIVgorCP684con    ------R-.............................----.--QD------RS............G-------.-Y--..PGGE-................---YV-......................KQV...Q-----..........................M-E  462
GOR.CM.07.SIVgor2139_287    ------R-.............................----.--QD------RS............G-------.-Y--..PGGE-................---YV-......................KQV...Q-----..........................M-E  462
GOR.CM.07.SIVgorCP2135con   ------R-.............................----.--Q-------RS............G-------.-Y--..PGGE-................---YV-......................KQV...Q-----..........................M-E  462
GOR.CM.12.SIVgor_BPID1      --K---R-.............................----.--Q-------K.............SG-------.-----..PGGK-................---YX-......................KQV...Q-----..........................M-E  460
GOR.CM.12.SIVgor_BQID2      ------R-.............................----.--Q-------K.............NG-------.-Y--..LGGT-................---YV-......................KQA...Q-S---..........................M-E  467
GOR.CM.13.SIVgor_BPID15     ------R-.............................----.--Q-------K.............IGK------.-----..PGGK-................---YV-......................KQV...Q-----..........................M-E  460
                                                                                          p8 end p1 start              p1 end p6 start                   PTAP motif

MAC.US.x.239                Q-----RQ.............................----.---MD-V-AK-P............D---G---LGP-.....GKK................-R--PM.....................AQVHQGLM-----ED..........................  464
H2A.DE.x.BEN                Q-----RQ.............................----.---P--I-AN-P............----G---LGPR.....GKK................-R--PV.....................TQA-QGLI-----..........................AD  462
H2B.CI.x.EHO                Q-K---RQ.............................----.--QQ--I-SK-P............----G---FGP-.....GKK................-R--PV......................QA-QGIV-S---MNPAFGMTPQG........AIPSAPPAD  477
H2G.CI.92.Abt96             Q-----RQ.............................----.---P--I-AK-P............----G---FGP-.....GKK................-R--PM.....................TQV-QGLT-S---..........................MD  467
H2U.CI.07.07IC_TNP3         Q-----RQ.............................----.---P--I-AK-P............----G---FGP-.....GKK................-R--PM.....................TQI-QGLT-----..........................AD  467
H2U.FR.96.12034             Q-K---RQ.............................----.-E-P--N-AS-P............-TKVG---LGP-.....EKK................-R--PM.....................AQV-QGLT-----AEPAVDLLTPTAPP............AD  479
ASC.UG.10.RT03              Q-PH-K-I.........................K-F-.--R---MARQ-RAS.......EGAQGG-----KAFL.....DKHL..............G-RKGPR......................NFPVVQ-S---...................LTLQAD  478
ASC.UG.10.RT08              Q-PQ-K-M.........................K-F-.--R---MARQ-RAS.......DSAQGG-----KAFL.....DKHL..............G-RKGPR......................NFPVVQ-S---...................LTLQAD  478
ASC.UG.10.RT11              Q-PQ---I.........................K-F-.--R---MARQ-RAS.......EGAQGG-----KAFL.....DKHL..............G-RKGPR......................NFPVVQ-S---...................LTLQAD  478
COL.CM.x.CGU1               M-PK.-PIG...GAGR..GRGRGRGGFRGAPRRP..VR-FT.-NQ--MQR--P............NK------ATRG....VEL...............QT...............................AIF--..........................  454
COL.UG.10.BWC01             Q-PK.-GAGGFRGGGMR.GRGRGGRGAWGRGRNPGPLK-FG.--QL--IQR--Q............NK------AGRG....IDL...............QT...............................AIF--..........................  460
COL.UG.10.BWC07             --PK--QFP...TRG..RGSSRGRGIMRGGMRRGGQAK-FG.--QL--IQAT-P............NK-V----TPGG....MDL...............QT...............................AVF--..........................  458
DEB.CM.04.SIVdeb04CMPF3061  D--K--TV.........................R-F-.-QQT--MA-N-P.........KAAPQSKN----KLP-.....GR-...............A-R--IEQ.....................QHQTWGIGKYI-EG.....................GD  506
DEB.CM.99.CM40              D-KR---T.........................K-F-.--Q---IA-N-G............QTPR-----NTP-.....AS-...............R-R---EG..........................IGKYI-EST.....................AD  493
DEB.CM.99.CM5               D-KK---V.........................K-F-.------IA-N-G............QVPR-----NAP-.....AS-...............R-R---EG..........................IGKYI-EST.....................AD  490
DEN.CD.x.CD1_CMO580407      D--K-KRVT.....................QPGA-F-.---L---A-N-R.......SNTNDQG-V----NGLG.....RKN...............P-R--V-Y.............YPQMNQDLPWSPVRVQQQ-QYP......................QN  514
DRL.DE.11.D3                Q-TK----.........................--FR.--ALD-MLRN-.............PK-V----NTP-.....GSG...............K-R--P-A.........................SLT-S---LPGY.................LQ-D  474
DRL.DE.11.D4                Q-TK----.........................--FR.--ALD-MLRN-.............PK-V----NTP-.....GSG...............K-R--P-A.........................SLT-S---MPGF.................LQ-D  474
DRL.DE.11.D5                Q-TK----.........................--FR.--ALD-MLRN-.............PK-V----NTP-.....GSG...............K-R--P-A.........................SLT-S---LPGY.................LQ-D  473
DRL.DE.11.D6                Q-TK----.........................--FR.--ALD-MLRN-.............PK-V----NTP-.....GSG...............K-R--P-A.........................SLT-S---LPGY.................LQ-D  474
DRL.x.x.FAO                 Q-TK----.........................--FR.--ALD-MLRN-.............PK-V----NTP-.....GSG...............K-R--P-A.........................SLT-S---LPGY.................LQ-D  474
GRV.ET.x.GRI_677_gri_1      E-K---QI.........................K-F-.---I--MA---K.............NG------YGH-.....GGA...............K-R--V-Y.....................RGDTVGL------.....................METAYD  469
GSN.CM.99.CN166             D-PK-KER.........................K-F-.--RA--FS-P-R.....T...GTPIKELY--DLGRG.....RR-...............A-R--PV..........................TFLL----A..........................--E  473
GSN.CM.99.CN71              D-PK-KER.........................K-F-.---A--LARQ-KT.........GTPAKEL---DLGRG......R-...............A-R--PV..........................TSLL-S--A..........................--V  471
LST.CD.88.SIVlhoest447      M--Q--QE.........................---N.--SKE-RFAQ-PKP.........KGKV----YGP-.....RSG...............P----PL.....................MGGNAGVV-S---.....................MERSP..T  471
LST.CD.88.SIVlhoest485      M--Q--QE.........................---N.--SKE-RFAQ-PKP.........KGKV----YGP-.....RNG...............P----PL.....................MGGNAGVV-S---.....................MERNS..T  471
LST.CD.88.SIVlhoest524      M--Q--QE.........................---N.--AKD-R-AQ-PKP.........KGKV----YGP-.....RNG...............PS---PI.....................MGGAAGVI-S---.....................MEKVPP.T  470
LST.KE.x.lho7               M--Q--QE.........................---N.--SKE-RFAQ-P..........KPKGKV----YGP-.....SSKP..................---YPLL...................GGAAGRI-S---.....................MESAP..T  474
MAC.US.x.251_BK28           Q-----RQ.............................----.---MD-V-AK-P............D---G---LGP-.....GKK................-R--PM.....................AQVHQGLT-----ED..........................  464
MAC.US.x.EMBL_3             Q-----RQ.........................R---.---MD-V-AK-P............D---G---LGP-.....GKK................-R--PM.....................AQVHQGLT-----ED..........................  464
MND-1.GA.x.MNDGB1           FFK---R-.........................---N.--AMD--KAQ-P.........KPAQQQRV----YGP-.....GPS...............K---YPA..........................QEVT-----L.....................EEKP...  463
MND-2.CM.98.CM16            D-IS----.........................--F-.--DL--I-RN-P.............KMV----NTP-.....GSG...............K-R--PA..........................MPLT-----MPG...................M-D  470
MND-2.GA.x.M14              D--L----.........................--F-.--DT--M-RN-P.............KMV----NAP-.....GSG...............K-R--PA..........................MPLT-S---MPG...................M-D  470
MND-2.x.x.5440              D------R.........................--F-.--DP--L-RN-.............PKMV----NTP-.....GSG...............K-R--PA..........................MPLT-S---MPG...................L-D  470
MNE.US.x.MNE027             Q-----RQ.............................----.--QM--V-AK-P............D---G---FGP-.....GKK................-R--PM.....................AQMHQGLT-----ED..........................  464
MON.CM.99.L1_99CML1         --K--QRR.........................K-YN.--QP--LA---PQPP.......KQNKGV----NPFG....P-K...............G-R--PL..........................TSVQ-S--TLP......................AG  472
MON.NG.x.NG1                --T----T.........................--F.R------XS-N-PNGGQNQ.....PRNRV----NPFG....PGK-...............G-R--PV..........................TSLQ----AEP......................TQ  148
MUS-1.CM.01.CM1239          S-PK-KTR.........................K-F-.--R---LA-Q-R.........SEGAKS-----DPLG....LQR-...............R-R--P-...........................TSLL-----.........................-Q  467
MUS-1.CM.01.SIVmus_01CM1085 D-PK--N-.........................K-F-.--R---LARQ-R.........SDNTKAV----DPFG....LQK-...............R-R--P-...........................TSLL-----.........................-Q  471
MUS-2.CM.01.CM1246          D-PK-KST.........................R-F-.------LARQ-R.....T...DTGKSAV----KDLG....FRS-...............K-R--PV..........................TSLM-----.........................DT  475
MUS-2.CM.01.CM2500          D-PK--VR.........................K-F-.--R---LA-Q-R.....T...NSDKGAV----KGFG....FH##................K-R--PV..........................TSLM-S---.........................DT  474
MUS-3.GA.09.09GabOI81       --PK--VV.........................K-F-.--R---MA-Q-N..........TSTGRV----NPFG....PQK-...............G-R--PV..........................TQLV-----.........................GT  474
MUS-3.GA.11.11GabPts02      S-PK-KVV.........................K-F-.--R---MA-Q--...........SQGKV----NPFG....PQK-...............G-R--PV..........................TQLV-----.........................DF  478
OLC.CI.97.97CI12            --SQ.KGP.........................T-Y-.---P--IA-N-R.....VGGQQQQSKV----KWRG..........GKAEP.......PLNMQQR..............................TEMD.......................IF  441
RCM.CM.00.SIVagi_00CM312    Y-K---R-.............................----.--E-D-IL---P.............K-V----NAP-.....-QG...............K-R--PL..........................TSVT-S--GMD......................QYD  470
RCM.CM.02.SIVrcm_02CM8081   F-K---R-.............................----.--D-D-L-----.............K-V----SAP-.....SRG...............K-R--PL..........................TSLT-S--GVEQDNH.................ASQYD  476
RCM.GA.x.SIVRCMGAB1         Y-K---R-.............................----.--SPD-LL----.............K-I----RLP-.....GQG...............K-R--PL..........................TSLT-S--GMES.....................NYD  469
RCM.NG.x.NG411              Y-KTS---.........................---R.--E---L----P.............KEV----NAP-.....-QG...............K-R-LPL..........................TSLR----GREE.GW.................NPQYD  474
SAB.SN.x.SAB1               F-K---RQ.............................----.--SPD------Q.............K-V----FGP-.....GRG...............K-R--PLT...........................SIR-----MERDYSRPEENWYADRP..PTRGPGPDD  509
SMM.CI.79.SIVsmCI2          Q-K---RQ.............................----.---S--V-AK-P............----G---FGP-.....GKK................-R--PM.....................TQLQQGLT-----MDPAVAAPP.................MD  476
SMM.LR.89.SIVsmLIB1         Q-K---RQ.............................----.---T--I-AK-P............----G---FGP-.....GKK................-R--PM.....................AQMHQGLT-----..........................AD  464
SMM.SL.92.SIVsmSL92A        Q-K---RQ.............................----.---M--I-AK-P............----G---FGP-.....GKK................-R--PM.....................AQM-QGLT-----..........................AD  464
SMM.SL.92.SL92B             Q-----R-.............................----.--E--RIQAN-P............NQK-G---LGP-.....GKK................-R--PM......................QTTSLT-S---..........................D  465
STM.US.89.STM_37_16         Q-KG--RQ.............................----.--QQ----AK-P...............VG---FGP-.....GKK................-H--PM.....................AQI-QGLT-----..........................D  465
SUN.GA.98.L14               Q--E-K-G......................PPGS---.---M--KQAQ-P...........QGK-V----YGP-.....NRG...............P----PV.....................MPSAPPLEDLTLG...................NRMTPPQ-  480
SYK.KE.x.KE51               D-PR-K-L.........................K-FN.--GT--IARQ-RQPRKGQ...GNPPP------KGWG....NRK...............P-A--PV..........................MRET-----AEDWPW............QQQTWGNYASP  507
SYK.KE.x.SYK173_COMGNM      D-KK-L-A.........................K-FN.---T--LARA-RQPKRNQ...GPPVA------KGWG....VSR-...............P-A--PV............................RS--S---LEDI.............EDGPWLTWSAQ  504
TAL.CM.00.266               --P--PRQKV....................PPGT-Y-.---P--IA-Q-S.............SA------KSFL.....GKI...............P-VRGPR......................NFPVVQ-S---.........................I-PL  487
TAL.CM.01.8023              --P--Q-PRG....................PPGS-F-.--QM--RAAQ-R.............QP------KAFL.....GKI...............P-VKGPR......................NFPVVQ-----.........................I-PL  490
TAN.UG.x.SIVagmTAN1         D-PRAGPN.........................K-L-.---P--LA---R.............SG------RMPT....WGT...............K-R---E.....................QGGAVPTA-PM-AHGFPT..............GSPAAGAYD  486
VER.DE.x.AGM3               Q-PE---M.........................R-L-.---P--LA---R...............G-V----YGR-....MGA...............K-R--PAA.........................TLGV------.....................PPS...PYD  477
VER.KE.x.9063               Q-TE---I.........................K-L-.---L--LA---R...............G-V----YGR-....TGT...............K-R---AA.........................THGV--S---.....................PPT...PYD  475
VER.KE.x.AGM155             Q-PE---I.........................K-L-.---P--LA---R...............G-V----YGR-....MGT...............K-R--PAA.........................TLGA--S---.....................PPNNSTPYD  476
VER.KE.x.TYO1_patent        Q-PE---T.........................K-L-.---L--LA---R...............G-V----YGR-....MGA...............K-R--PAA.........................TLGA--S---.....................PPSGTTPYD  475
WRC.CI.97.97CI14            E-PK-K-PVPMR.................RGQGPT--G.--EV--VQRN-P.........KIR--V----NGRG.....QMPS...NNYPVQK..LRSLN--GPT..........................TTM-S---MR.........................-D  463
WRC.CI.98.98CI04            D-PK-K-PVPLR.................RGQGPT--G.--EI--VQRN-P.........KSR--I----NGRG.....QMPS...NNYPVQK..VRSLN--GP...........................ATI-S---MR.........................-D  466
WRC.GM.05.Pbt_05GM_X02      D-KK-KRSPQQR....................GPPT--G.---T--IQ-N-P.........ETR--V----NGRG.....QR..PS.NNYPVQKMN.AGVNT-GP............................VV---..........................VD  459
```

```
                                                                                    Vpr binding
                                                                                               p6 end, Gag end
H1B.FR.83.HXB2              FRSGVET.....TTPPQKQE....................PIDKE..LYPL.................TSLRSLFG.NDPSSQ............*   500
H1A1.UG.85.U455_U455A       -GM-EK......M-S-A---....................LK-R-..QT--.................V--K----.---L--............-   494
H1C.ET.86.ETH2220           --F.E-......A--SP---....................LK-R-....A-.................---K----.--HLL-............-   505
H1D.CD.84.84ZR085           -GF-E-......I--S----..................QKDK---.-----.................A--K----.---L--............-   502
H1F1.BE.93.VI850            -GFRE-......I--SP---....................QK-G-L.YP--.................A--K----.------............-   492
H1G.SE.93.SE6165_G6165      LGF-E-......IA-SP---....................MKE--.-----.................---K----.S-----............-   496
H1H.CF.90.056               -GF-E-......M--SP---...................QLK---...P--.................A-------.S--LL-............-   501
H1J.SE.93.SE9280_7887       LGL-E-.......I-SP---....................-K---.-----.................---K----.S--L--............-   498
H1K.CM.96.96CM_MP535        -GF-E-......I--SPR--....................TK---.QSP--.................---K----.---L--............-   496
H1O.BE.87.ANT70             EVK-Q.........EN-E-K.....................GGPNELYPF.................A--K----.T-Q...............-   499
H1O.CM.94.BCF06             EVKEL.........EK-S-G.....................GE--ELYPF.................A--K----.T-Q...............-   499
H1O.FR.92.VAU               MMK-Q-........ED-N-K.....................GEQNELYPF.................A-------.T-Q...............-   499
H1N.CM.02.DJO0131           YG-QE-R......STQ.GK-..................RQENTENS.LYP--.................--------.--L---............-   513
H1N.CM.04.04CM_1131_03      YG-QE-R......STQ.RK-MQE...............NQEKTENT.LYP--.................--------.--L---............-   512
H1N.CM.95.YBF30             YGFQE-K......STQ.GK-MQE...............NQERTENS.LYP--.................--------.------............-   513
H1P.CM.06.U14788            EEMTQ.........NK-XE-K....................E-EKELYPL.................A--K----.T-Q...............-   493
H1P.FR.06.RBF168            DEMTR.........DK-RE-K....................E-EKE.YPL.................--------.T-Q...............-   492
CPZ.CD.06.BF1167            WN-KDG...............................................................F--K-I--.S-Q*WK............-   531
CPZ.CD.90.ANT               YQEEH.........KRTQKGLKG....................-EE-P-S.................Y--K----.K-Q*WKFSYRDKNVRL.Y-   536
CPZ.CM.01.SIVcpzCAM13       YGYQEKEK....NQEE-EEG.......................NRLYP--.................---K----.S-------.........-   507
CPZ.CM.05.LB715             -GFQP........KTA--M-VKE...............EEQTEE--SLYP--.................---K----.S-Q---............-   501
CPZ.GA.88.SIVcpzGAB2        -GYAEQ-Q....-GKS--G-....................ESTT-KESPLY.................P--K----.S-------.........-   508
CPZ.US.85.US_Marilyn        -GYQE--V....-QEK-GK-.......................--PF..Q-.................---K----.S---E...........-   506
GOR.CM.04.SIVgorCP684con    EEKET.........RKPGK-.........................ELYPL.................---K----.T-Q...............-   490
GOR.CM.07.SIVgor2139_287    GEKEM.........RKPEK-.........................ELYPL.................---K----.T-Q...............-   490
GOR.CM.07.SIVgorCP2135con   GEKEM.........RKPEK-.........................ELYPL.................---K----.T-Q...............-   490
GOR.CM.12.SIVgor_BPID1      EEMTQ.........SK--E-Q....................E--KELYPL.................A--K----.T-Q...............-   493
GOR.CM.12.SIVgor_BQID2      VMK-K.........EN-D-G.....................G-QGELYPF.................A--K----.T-Q...............-   499
GOR.CM.13.SIVgor_BPID15     EEVTQ.........SR--EG......................P--ELYPL.................---K----.T-Q...............-   491
                                                                                                              p6 end
                                     PSAP in HIV-2 B, U                                                        Gag end
MAC.US.x.239                PAVDLLKN....YMQLG--QR............EKQRESREKPYKEVTED.................LLH-N----.G-Q...............-   511
H2A.DE.x.BEN                PAAELLER....YMQQGRKQ...............REQRERPYKEVTED-LHLEQRETPHREETEDLLH-N----.K-Q...............-   522
H2B.CI.x.EHO                PAEEMLKN....YMQLG-KQ...............KENRERPYKEVTED.................LLH-N----.E-Q...............-   520
H2G.CI.92.Abt96             PAVDLLKN....YMQLGRKQ...............KEQRNKPYKEVTEX.................LLH-S----.D-Q...............-   510
H2U.CI.07.07IC_TNP3         PAVDLLKN....YMQMGRKQ...............KESRERPYKEVTED.................LLH-N----.D-Q...............-   510
H2U.FR.96.12034             PAVDLLKS....YMQQG-KQ...............KENRERPYKEVTED.................LLH-N----.--Q...............-   522
ASC.UG.10.RT03              TFQEENLG.....PAVGI-KIP.............E...GEKTR-SLYPS....................-S----.D-Q...............-   517
ASC.UG.10.RT08              TFQEENLG.....PAVGI-KTP.............EDNNGEKTR-SLYPS....................-S----.D-Q...............-   520
ASC.UG.10.RT11              TFQEENLG.....PAVGT-KTP.............EGNNGEKTR-SLYPS....................-S----.D-Q...............-   520
COL.CM.x.CGU1               ....KMSK...........................DLPPRRE-GES---....................--K----.D-Q...............-   483
COL.UG.10.BWC01             .AMSKMQK...........................QEGRKR-E.NT---....................-------.E-Q-*.............-   492
COL.UG.10.BWC07             KM-QRKEP...........................SKGEMDREG..I--....................--K----.E-LL-GQN..........-   494
DEB.CM.04.SIVdeb04CMPF3061  EIFSSSRV.....PEQKP-Q.................QEE..TRRGKYPD.................LSD-K----.E-Q...............-   545
DEB.CM.99.CM40              QVYSTTRV.....PVQGQEN...............NPQREE.TTRRGKYPD.................LMG-K----.E-Q*P-YT..........-   540
DEB.CM.99.CM5               QVYSTTRV.....PVQG.ER...............NAPQEDTTTRRGKYPD.................LMG-K----.E-Q*P-YMWKEKGSRC..-   545
DEN.CD.x.CD1_CMO580407      -QE-GNRQ.....SCQEDKRQN..........KQSHQEDEETTR-SLYP.....................--K----.D-Q...............-   559
DRL.DE.11.D3                PAERMLQK....YMEQGA-QK.RQ........QRQQQQKKG.PYKEAYNS....................-S----.S-QLQ.............-   522
DRL.DE.11.D4                PAERMLQK....YMEQGA-QK.RQ........QRQQQQKRG.PY-EAYNS....................-S----.S-QLQ.............-   522
DRL.DE.11.D5                PAERMLRK....YMEQGA-QK.RQ........QRQQQQKKG.PY-EAYNS....................-S----.S-QLQ.............-   521
DRL.DE.11.D6                PAERMLQK....YMEQGA-QK.RQ........QRQQQQKKG.PY-EAYNS....................-S----.S-QLQ.............-   522
DRL.x.x.FAO                 PAERMLQK....YMEQGA-QK.RQ........QRQQQQKRG.PY-EAYNS....................-S----.S-QLQ.............-   522
GRV.ET.x.GRI_677_gri_1      PAKKLLQQ....YAEKGQRLRE.............EREQTRKQ--KEVED.................V--S----.G-Q*NESS..........-   518
GSN.CM.99.CN166             EPQQLNLS.....--VGEPQS..............KGQKEESRNTLYPS....................-S----.E-Q*RKSP..........-   517
GSN.CM.99.CN71              EPQQLNLS.....P-VGEPQN..............KEQKAEKRNTLYPS....................-S----.E-Q*EKSPWGDKWCPCY.-   524
LST.CD.88.SIVlhoest447      QAERALETYRNLGQQLRR-QQ....................-PQKCVDE-...................C-SFF-FPD-Q...............-   513
LST.CD.88.SIVlhoest485      KAERALETYRNLGQQLR--QQ....................VPQKCVDE-...................C-SFF-FPD-Q...............-   513
LST.CD.88.SIVlhoest524      RAERAIETYRNLGQQLR--QQ....................VPQKCVDE-...................C-NFF-FPD-Q...............-   512
LST.KE.x.lho7               KAERALETYRTLGQQLKR-QQQ..................VPQKCVDE-...................C-N---.PD-Q...............-   516
MAC.US.x.251_BK28           PAVDLLKN....YMQLG--Q...............RESREKPYKEVTED.................LLH-N----.G-Q...............-   507
MAC.US.x.EMBL_3             PAVDLLKN....YMQLG--Q...............RESREKPYKEVTED.................LLH-N----.G-Q...............-   507
MND-1.GA.x.MNDGB1           .LQKTLSTYQKLGRGL.R-KMK...................EE-REEDFH..................--ST--.QE-Q...............-   503
MND-2.CM.98.CM16            PAERMLLD....YMKKGQ-QK.AE........SKQEKKERG.PY-AAYNS....................-S----.T-QLQ.............-   518
MND-2.GA.x.M14              PAEKMLLD....YMRKGQ-QRAAK........ETKQEKDKG.PY-AAYNS....................-S----.T-QLQ.............-   519
MND-2.x.x.5440              PAEKMLLD....YMKKGQ-QRAAA........GAQKEEKKG.PY-AAYNS....................-S----.T-QLQ.............-   519
MNE.US.x.MNE027             PAVDLLKN....YMQLG--Q...............RESKRKPYKEVAED.................LLH-N----.E-Q...............-   507
MON.CM.99.L1_99CML1         TAEP-NLN.....QEIGSPKT...............GTPKETRRDLYPS....................-A----.E-Q...............-   511
MON.NG.x.NG1                GAYK-NFS.....P-VGPP-TP.............APTAPKETRKDLYPS....................-A----.D-Q-E-............-   193
MUS-1.CM.01.CM1239          EGDL-NLS.....P-VGPKDG...................AS-RKSLYPS....................-S---Q.D--LEK............-   506
MUS-1.CM.01.SIVmus_01CM1085 EGEL-NLS.....PAVGPG-G...................KQ-RKSLYPS....................-S---Q.D-Q-D-............-   510
MUS-2.CM.01.CM1246          .AEP-NFS.....PLVGPK-E...............PPKDRK-RKTLYPS....................-S----.D-Q-K-............-   517
MUS-2.CM.01.CM2500          .AEI-NFS.....PEVGPRKEN.............QTPQA-P-RKNLYPS....................-S----.D---E-............-   518
MUS-3.GA.09.09GabOI81       .GEP-NFS.....PKVGEPQK..............EKEE...TRKTLYPS....................-S----.D-Q-E-............-   514
MUS-3.GA.11.11GabPts02      .GKP-NFA.....PKVGPPTT..............SGEEGETTRNSLYPS....................-S----.D---E-*A..........-   522
OLC.CI.97.97CI12            PSAPTLEQILEKG-REKR....................EMKEEKKP---....................--A----.E-Q*..............-   481
RCM.CM.00.SIVagi_00CM312    PAEAMLKR....YLEQG-KQKQQK........RQEEKSQQGRAY-EA-SS....................-N----.--Q...............-   518
RCM.CM.02.SIVrcm_02CM8081   PAEAMLRK....YLEQGRMQKQQQ........RKEEKAQKGKAY-EA-AS....................-N----.S-Q...............-   524
RCM.GA.x.SIVRCMGAB1         PAEEMLKN....YLRRAGEQKRQQ........RQEESKKREGAYQEA-TS....................-N----.S-QLQ.............-   519
RCM.NG.x.NG411              PAEEMLRK....YLALGR-HKQEQ........RKENKEKVGRAY-DA-SS....................-N----.S-L...............-   522
SAB.SN.x.SAB1               PATALLKQ....YAVQG-RQK..Q........QWQNHSPQQSPY-EAYSS....................------.E-Q...*..KLCT......-   559
SMM.CI.79.SIVsmCI2          PAVDMLRN....YLKLGQ-Q...............RERKERPYKEVTED.................LLH-N----.D-Q...............-   519
SMM.LR.89.SIVsmLIB1         PAVDLLKN....YMQLG-KQ...............RENRERPYKEVTED.................LLH-N----.D-Q...............-   507
SMM.SL.92.SIVsmSL92A        PAVDLLKD....YMKLGRKQ...............KESRERPYKEVTED.................LLH-N----.E-Q...............-   507
SMM.SL.92.SL92B             PAARIVKE....YLEKAQR-...............KTRRSRPYKEVTED.................LLH-N----.E-Q...............-   508
STM.US.89.STM_37_16         PAADLLRS....YMQLG-KQ...............RESRKTPYKEVTED.................LLH-N----.E-Q...............-   508
SUN.GA.98.L14               KAERALETYRLLGQGLRA-QKR...................KERGECQE-...................C-NM--P.E--..............-   522
SYK.KE.x.KE51               QQKHIAMA.....--L-P-SSPKTVKDLLVPGKPSEKKEEMKE-GPLYPS....................-Q----.D-Q*WK............-   564
SYK.KE.x.SYK173_COMGNM      MSQQAQAK.....AQNSPSKKPPTNREVLSP.KESSGKEETKS...LYPS....................-S----.E-Q*..............-   555
TAL.CM.00.266               TA-WQ-GT.....-TWTPPAKK.............PQE.TTTTR-SLYPS....................-A----.D-Q...............-   528
TAL.CM.01.8023              TA-WQ-GT.....-TWTPPTKD.............AAAPKQQTR-SLYPS....................-A----.D-Q...............-   532
TAN.UG.x.SIVagmTAN1         PARKLLEQ....YAKKGD-LR....................KQ--KE-ED.................Y--S----.G-Q*VRSTYRASQYKFC.-   537
VER.DE.x.AGM3               PAKKLLQQ....YADKG--LRE.............QRKKP-AVNPDWTEG.................Y--N----.E-Q...............-   522
VER.KE.x.9063               PAKKLLQQ....YAEKG--LRE.............QGKRT-PTNPDWAEG.................Y--N----.E-Q...............-   520
VER.KE.x.AGM155             PAKKLLQQ....YAEKG--MRN.............QNRNP-ANNPDWNEG.................Y--N----.E-Q...............-   521
VER.KE.x.TYO1_patent        PAKKLLQQ....YAEKG--LRE.............QKRNP-AMNPDWTEG.................Y--N----.E-Q...............-   520
WRC.CI.97.97CI14            PNIQELEKYLLKGAQLRE-QEHK............KTEGQKREVGVPEW-...................--K---E.P-Q*YNA...........-   515
WRC.CI.98.98CI04            PNIQELEKYLLKGAQLKE-AQAK............KEE.KKRETGLPEW-...................--K---E.T-Q*YNV...........-   517
WRC.GM.05.Pbt_05GM_X02      PTMQELEQYLLRGAQMKREKEK.................EMQKEGKVE--...................--K---E.P-Q*YNVSLVERVSLRY.-   516
```

```
                                Pol start
                                Gag-Pol TF start                                                                Gag-Pol TF end_protease start
H1B.FR.83.HXB2                  FFREDLAFLQGKA...REFSSEQT...RANSPTRRELQVWGRDN.............................NSPSEAG.ADRQG..TVSFNFPQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSL....PGRWKPKMIGGIGGFIKVRQYDQILIEICGH  125
H1A1.UG.85.U455_U455A           ----N---Q--E-...-------...------S-N-WDG-K-...............................DL-C-T-.-E-----.-D--S---I---------V------I--------------DIN-.....-K----I----------------------K  124
H1C.ET.86.ETH2220               ----T---Q---------P-------------ESQTRANSPTT.......................RELQVRGS-TF-----.-E-------L----I-------------------------------IN-.....-K-----------------------I-----K  132
H1D.CD.84.84ZR085               ----N---P-------G-L-------------S-SFGFGEEIT.........................#LPETGTE.GQ---T.--LS---I------V-------------------------DIN-.....-K-------------------H--------  125
H1F1.BE.93.VI850                ----N---Q--E-....-K-P-------------S---R-QRG-...............................PL-----.-E-R-TV.P-LS---I---------------I---------------DIN-.....-K---------------K---N--------  125
H1G.SE.93.SE6165_G6165          ----N---Q--E-...-------D...-T----C-KPR-RRG-S...............................PLP---.DEGK-....AISL--I---------V------I---------------IN-.....-------------------------VP---S-K  122
H1H.CF.90.056                   ----N---Q-RE-....-K--P--A...-T----S---R-RRG-D...............................PL----A-EG--T...-LS---I---------V--E---R---------------IN-.....-K-------------------E-VA-----K  124
H1J.SE.93.SE9280_7887           --------Q-RE-....-L-P-------------S--PRARRG-...............................PLP-T-.-EG--T...-S----I----------R-----R---------------DID-.....-RK-------------------NEVP---E-K  123
H1K.CM.96.96CM_MP535            ----N---P--E-...-------...------S---R-R-G-...............................PL-----.DQ---TE.P------I------I----V----R---------------IN-.....-K-------------------V------Q  125
H1O.BE.87.ANT70                 ---QI--SGGHE-...-QLCA-...TS..TPISPTDGGGSEG.........................TGESGT........ERGPE...RAL-VCL--IP--D--I--ARV--H-C-V------------NNIQ-....E-K---------------KE--NVTV--E-R  121
H1O.CM.94.BCF06                 ----V--SRGHE-...-QLCA-...TS..VPISSTNGGGS.........................EGAREAESGGGEG......R--PVCL--IP--D--I--A-V--H-C-------------NNIQ-....E---T-------------KE-NNVTV-VQ-K  121
H1O.FR.92.VAU                   --K-I--SERHE-...-QLCAG...TS..VPISPTNGRDDEG.....................ARGGSESEGGA........E.......RAL-VCL--IP--D--V--ARV----C-V------------NNIQ-....E-K-A-------------KE-YNVAV-LE-R  122
H1N.CM.02.DJO0131               ---KG-VS--RET...-QLPPDNNNKE--H--AT---W-S-G.........................EEHT....GEG---E.H-ELSI.PTL----I------VI-V---KEVR------------I--IQ-....E-K-------------------NVT-D-Q-R  130
H1N.CM.04.04CM_1131_03          ----G-VS--RET...-L-PDNNKE.--H--AT---W-S-G.........................EEHTEKGD.AG-P-E.--EHSV.PTL----I------VISV---NEVR------------I--IQ-....E-------------------N-T-D-Q-R  132
H1N.CM.95.YBF30                 ----E-VS--RET...-KLPPDNNKE.--H--AT---W-S-G.........................EEHTGEGDAG-P-E.--ELSV.PT-----I------VI-V---KEVR------------I--LQ-....E-K-------------------N-TVD-Q-R  132
H1P.CM.06.U14788                -------SGGQE-...-QLCT-TS...TTD---Y.................................GGGGDDSEQA-G-KGG-ERV...P-SL--I---D--VIPV----HIC------------VDNLP-....E-----------------KEFEEVKX--E-R  122
H1P.FR.06.RBF168                ---K---SWGQE-...-QLCT-TS...TT----Y.................................GGGRDDPGQA-G-KGG-E....IP--L--I---D--VIPVRV--HIC------------VDKLP-....E-----------------KEFEEVKV--E-R  121
CPZ.CD.06.BF1167                ----THPLVGVQT...-K-CA-SESDSESGTESDTGETSGES...............RTLSSS.................N.GGGESLEQQGWL--EIH-----MMEVTVQ--VCT-----------FADLK-....K-SCN-RT-------VP-QE-YNVP-T-A-K  128
CPZ.CD.90.ANT                   ----TDPHVV-VQT...--LCA-GGSS.G---S-H-D-..S-GA.........................QEDSEGSQGGGGT..........T-LV--EIP-----MMEVL-Q--KCQ-----------V-GIH-....Q-N----T--------S-Q--NKVP-Q-GDR  125
CPZ.CM.01.SIVcpzCAM13           ----G---P-RE-...-QLCT--D.........S---W-PRE.................................GEEPR-TGRGEQTI.P-S----I------IPVR-E---C------------IQDIQ-....Q-I-----------------K---NVS---E-K  124
CPZ.CM.05.LB715                 ----N--SQ-RE-....-L-------------NG--W-STK..................................DSSE-GSEGG-TDRGE-ESVP------IS-----VL-V----H-----------------LE-....G-K-----------------K--EN-P-----K  130
CPZ.GA.88.SIVcpzGAB2            ---GG---PERE-....-L-PKEG...GT--S-S-D-R-CRA-T..............................DRKESEGG-EYN-K-E..P-LS---I-------LPVRVA---C-------------DNIQ-....E-K-R--L----------K---NVE---E-K  128
CPZ.US.85.US_Marilyn            ----T-VP..IVERGIK-TRELPGKQEG-H-S-N-GFR-SR--S.................................DTGETGKGKG.A..L-A----I-------LR--VA--IV-------------DNIQI....E-T-R---M----------K---HVN---E-R  127
GOR.CM.04.SIVgorCP684con        ----I--SRGRE-...-QLCA-AS.....................................ATNSPT-GGGRE-NEETRKGG-L-S----I---D--VIPVR---H-C------------V-HLP-....E-----R-----------KE-EDVE---EHR  117
GOR.CM.07.SIVgor2139_287        ----I--SRGRE-...-QLCA-AS...AT----N.................................GGGRKGDEETRKGGT...L-S----I---D--VIPV----H-C------------V-HL-....E-----R-----------KE-EDVE---EHR  117
GOR.CM.07.SIVgorCP2135con       ----I--SRGRE-...-QLCA-AS...AT----N.................................GGGRERNEETREGGT...L-S----I---D--IIPV----H-C------------V-HL-....E-----R-----------KE-EDVE---EHR  117
GOR.CM.12.SIVgor_BPID1          -------SWGQE-...-QLXT-...ASTT----Y.................................GGGGDDPE..QT-G-TGGQE-V.....SL--I---D--VIPV----HIC------------VDKIP-....E-----------------KEFENVK---E-R  122
GOR.CM.12.SIVgor_BQID2          ----V--SGGHE-...-QLCA-...ASTTV---N.................................GGSDERKG.....ESGSRGGSGRAL-ICL--IP--D--V--A----H-C-V------------HDIQ-....E-K-T-------------KEFNNVTVQ-E-R  121
GOR.CM.13.SIVgor_BPID15         ----N--SGGQE-...-QLCA-...ASA-----H.................................GGGGSDPE....QTEGRAEG-T......L--I---D--VIPV----HIC------------VDKIP-....E---V-------------KEFENVK---E-R  120
                                Pol p15 start                                                                p15 end_protease start
MAC.US.x.239                    ---PWSMG..KE-...PQ-PHGS.......-ASGADANCSP.................RGPSCGSAKELHAVGQAAERKA-RKQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV-------NTKE-KNVE--VL-K  137
H2A.DE.x.BEN                    ---VGPTG..KE-...SQLPRDP.......--SGADTNSTSGR..SSSGTVGEIYAAREKAEGAEGETIQRGDGGLAA-R..AER-TSQRGDRGLAA--FS--K--V--AY-ED-PV-V--------SIVAGIE-....GDNYT--IV-------NTKE-KNVE-KVLNK  150
H2B.CI.x.EHO                    ---VRPLG..KE-...SQ-PRPG.......T-.GDSAICAPDEP.SIRHDTSGCDSICTPCRSSRGDAKELHATREEAEG..EQRETLQGGDRG-AA--FS--R--V-KAT-E--SV-V--------SIVAGIE-....GSNYT--IV-------NTNE-KNVE--VV-K  150
H2G.CI.92.Abt96                 ---VWTLG..KE-...SQLPHDP.......--SGSDSISTT.................DGPSSRPTEELHAXGEEAKG....AEKQTLQGGDRXXAA--LS--R--V--AY-ES-PV-V--------SIVAGIE-....GSNYS--IV-------NTKE-KNVE-KVV-K  133
H2U.CI.07.07IC_TNP3             ---VWPMG..KE-...SQLPHDP.......D-SGVDTNSA..................PSRPSSGSAEELHADGQETKG....EQGKTIQGGDGG-AA--FS--R--V--AH-E--PV-V--------SIVAGIE-....GVNYH--IV-------NTKE-TNVE--VL-K  133
H2U.FR.96.12034                 ---AWPLG..K-....SQLPHGP.......--AGTDPNSP.....SSRTSSGSADPDSPSSGPSSGSAKELHATGQETKG..EQRETIQGGDGG-AA--FS--K--I--AY-E--PVKV--N-----SIVAGIK-....EPYYT--TV-E-E---NTKKFHNVN-KVL-K  147
ASC.UG.10.RT03                  ----GFFR..QASGAQEGPQEL.........................PSSTAISSTFDSAGRHFSGGELRPSCRDTKDP..RR...GEDKG--LSL-VIP--G--I-SCLVER-EVDL---------IIHDQDI..ELGDN-H--I---V--N-R-KA-HSVE--WQ-K  135
ASC.UG.10.RT08                  ----GFFR..QASGAQEGPQEL.........................PSSTAISSTFDSAGRHFSGGEPRPSCRDTEDP..RRQQWGKNKG-PLSL-VIP--G--I-SCLVE--EVDL---------IIHDQDI..ELGDN-H--I---V--N-R-KA-HNVD--WQ-K  138
ASC.UG.10.RT11                  ----GFFR..QAPGAQEGPQEL.........................PSSTAISSTSDSAGRHFSGGELRPSCRDTEDP..RREQWGEDKG-PLSL-VIP--G--I-SCLVE--EVDL---------IIHDQDI..DLGDN-H--I---V--N-R-KA-HSVS--WQ-K  138
COL.CM.x.CGU1                   ---GNQGG..-ATD......-HLPP.......EDEQRPATQ.....................................E-EGGE-LSLLE-S-RR--I-IVEVE--KCQ-----------FK-EDV..QLT-A-TAVQ-Q----A-R-K--RDKILKMG-K  107
COL.UG.10.BWC01                 ---GRQGD..RP-...DRNL-PS.......NEQDA-AGGT.....................................EERREH-VSLLE-S-RR--I-IAE-E--KCQV----------IK-SDI..DIK-T-TAVN-Q----N-R-K--KDKKVL-G-K  109
COL.UG.10.BWC07                 ---DTRGD..-PT...DSCLPPK.......DESEEGTLQG.....................................GDGQGRNLSLLEIP--R--F-WAELE--RFQV-V---------K-EEI..QLT-T-TVVT------S-RAK--KNRVMK-G-K  109
DEB.CM.04.SIVdeb04CMPF3061      ----TSLG..EESP.QEFHRAATP.DMGHREVHTGGRGRD.........................LLLFQGTGTET..PAAGGDEEREIPR-IGLE-P--R--VT--E-E-RRVQV---------IIH-KDI..QLTSPFV--TV--L----N--C-PG-E-QCQ-K  134
DEB.CM.99.CM40                  ----YSLG..LQET.QELSRRNW..EV.HPGEHS-SGILD.........................YQGTSTGAGEQPPE-GNNEEREV.PG--GAEIS--R--MT-VH-E-RKVQV-I-------IIS-KDI..DLEAP-T--TV--L----N-KC-PG-E-AMANK  133
DEB.CM.99.CM5                   ----CSLG..LPET.QEF-GRDW..EI.HPREHS-SGLFN.........................NQGTCARGKEC-PG-HN-QEREV.PR--GAEIS--R--MTAVYVE-KRVQV-I-------IIS-KD...DLDAP-T--TV--L----N-KC-PG-E-KMADK  133
DEN.CD.x.CD1_CMO580407          ----WSWE..EESS.QELCTVLPS.DESGFAMEPS-G.............PAATTVPTKFSRGGKQTILSGRQEAEQAV...PSGG-GDNEGE-LS-VE-S--R--IGQ-WVQ--RV-V-I-------IIA-QDI.NLGTEG-I--TV-----Y-N-QA-PG-E-KLMDK  149
DRL.DE.11.D3                    ----YSLG..QRET.QEFS-GL.......................ADPISSSIARIPPGGSGREDAPKVHGAGSSAEETT-.TTTAEE-PLQGGLQL--FS-RL--TTIVE-E--KV------------IQGLE-....T-K---QI------T-N----FNCE-K-A-K  139
DRL.DE.11.D4                    ----YSLG..QRET.QEFS-GL.......................ADPISSSNARIPPGGSGREDAPKVHGAGSPAEETT-.TATAEE-PLRGGLQL--FS-RL--TTIVE-E--KV------------IQGLE-....T-K---QI------A-N-K--FNCE-KVA-K  139
DRL.DE.11.D5                    ----YSLG..QRET.QEFS-GL.......................ADPISSSTARIPPGGSGREDAPKVHGARSSAEETT-.TATAEE-PLRGGLQL--FS-RL--TTIVE-E--KV------------IQGLE-....K-K---QI------T-N----FNCE-KVA-K  139
DRL.DE.11.D6                    ----YSLG..QRET.QEFS-GL.......................ADPISSPIARIPPGGSGREDAPKVHGAGSSAEETT-.TTTAEE-PLRGGLQL--FS-RL--TTIVE-E--KV------------IQGLE-....T-K---QI------T-N----FNCE-K-A-K  139
DRL.x.x.FAO                     ----YSLG..QRET.QEFS-GL.......................ADPISSSIARIPPGGSGREDAPKVHGAGSSAEETT-.TTTAEE-PLRGGLQL--FS-RL--TTIVEVE--KV------------IQGLE-....T-K---QI------T-N----FNCE-K-A-K  139
GRV.ET.x.GRI_677_gri_1          ---VWPLG..RSET.KKFCAIQRRHS.WSGTNSPP.................NGNSLRSSKEAPPAVCREGTAPERGE....RTDKETEGERSGGC-LELP--R--MKRVI-E-TPVQ----------IIQ-KD-HFPPHKP-RS-VV-----G-H-KE-QGVQVQLEDK  145
GSN.CM.99.CN166                 ---PWPGA..PPSP..Q-L--NFPAA.NSPGG.GGRTPA.............................AQSLDT-RGTPEQGTEG-EQEH-LSL-VIS--G--MKEVS----VVSM---------IVQDSAI..ELDTP-T---V-----L-Q--EHKHVEVIFNEK  130
GSN.CM.99.CN71                  ---PWPGA...TSP..Q-L--NFPAA.ISPGG.GGGTPA.............................AQPVSAGGGAPEQGAEGRKEEH-LSL-VIS--G--MREVSM---VVSM---------IVQ-SAI..ELDTP-T--TV-----L-Q-KEHRHVEVIFNEK  129
LST.CD.88.SIVlhoest447          ---LWTLE..EWSS..-KL-FDGGE..CRS.S-ISSP.................NGEEPNTSRESLRNIQEPRTTAEE.....ATADPTEMCGRAMPEFLL-SR--IEN-LVN--PVS----------IFS-N-V..RIE-PYT-RVV-----Q-R-KE-KDVF---A-K  139
LST.CD.88.SIVlhoest485          ---LWTLE..EWSS..-KL-FDGGE..CRS.S-ISSP.................NGKELNKSRESLRNIQEPRTTTEE.....ATAGSTEMCGRAMPEFLL-SR--IEN-LVNR-PVS-----R----IFS-N-V..KIE-PYT-RVV-----Q-R-KE-KDVF---A-K  139
LST.CD.88.SIVlhoest524          ---VWSLE..EWTF..-QL-NHGRG..CRS.DSFSAT.................HGESASNKSREGNRDIQEPRTTTEE.....ATTSSTEVCGRALPEFLL-SR--IEKVLVN--EVT----------IFS-N-V..RIE-PYT-RVV-----Q---KEFRDVF-Q-A-K  140
LST.KE.x.lho7                   ---IWSLE..QTSR..QLS-TGRS...CRE.NSISSP..................NGECSNKSRESTGDLQDFRTTAEK...AA-ASPTEMCGRALPEFTL.SR--IES-LVN--PVS----------IFS-N-V..RIE-PYT-RVV-----Q-R-KE-RDVF---A-K  139
MAC.US.x.251_BK28               ---PWSMG..KE-...PQ-PHGS.......-ASGADANCSP.................RGPSCGSAKELHAVGQAAER....KQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV-------NTKE-KNVK--VL-K  133
MAC.US.x.EMBL_3                 ---PWSMG..KE-...PQ-PHGS.......-ASGADANCSP.................RGPSCGSAKELHAVGQAAER....KQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV-------NTKE-KNVE--VL-K  133
MND-1.GA.x.MNDGB1               ---VWPLG..SLQT..G-L-.........................GTRGDSNSSTIRGETSAENSEHLSEIRERA.....QAE-EG-EERGG-S--EYS-SR--IEEVSVD-VTIR----------IFN-RNI..KLK-N-Q--I------NLR-K---NVYV--R-K  132
MND-2.CM.98.CM16                ----YPLG..QWQT..Q-LP.......GDAIDPN.GP............SNARDGRPSRENAVRLHEERATAESGKQ.....TGKE-E-SIRG-LQL--FS--N--TTVVE-E--KV------------IKDLD-....K-N---QI------S-N-K-FFNCKVT-A-K  137
MND-2.GA.x.M14                  ----CSLG..QWQA..Q-L-.......CHATDPISTP............DARDGGSSREDVIRLYEKGATTEGSKGD....QAGEG--SLRG-LQL--FS--N--TTVVE-E--KV------------IKDVD-....T-N---QI------A-R-K--FNCK-TVA-K  138
MND-2.x.x.5440                  ----YPLG..QWQA..Q-LP.......CHAIDPISTP............DARVGRPGRKDAVRLYEERATTEGSSR-....PERGKE-SLRG-LQL--FS--N--TTVVE-E--KV------------IKDLE-....T-N---RI------A-R----FNCKV--A-K  138
MNE.US.x.MNE027                 ---LWPMG..KE-...PQ-PHGP.......NASGADTNCSP..................RGPSCGSAKKLHAVGQTAER....KQKEALQGGGRG-AA--FS--R--V--AH-E--PA-V--------SIVAGIE-....GPHYT--IV-------NTKE-KNVE--VL-K  133
MON.CM.99.L1_99CML1             ----SFRS..QEGA..K-LP-HLCAA.QRTDSAS...........................RDCGASQSQPRDWF----EDG-PK-DEAG-LSLLSIS--G--IRSVL----KVSM---------IIQDQHV..NLGEN-T--VV-----M-E-K--KA-QVIFEDR  130
MON.NG.x.NG1                    ----SFWP.RQEGA..K-L--NLLAA.NSP-G-NPXGLQGELLT.........................SSGTTRD---SADSPE-DEEG-LSLLSIP--G--VR-VE-A--SVSM---------IVQ-QDI..NLGEN-T---V-----L-Q-K-FKGVTVSFQ-K  134
MUS-1.CM.01.CM1239              ---GSSWP.PE-EA..K-LPPDLAFT.NGPARAGGRPGQSLPAS.......................................-SQGWSERQEE-LS--VIS-SG--IREVM-E--VVHM---------IVQDSNI..QIDKP-N--LV-----N-S--E-RGVQVSFNEK  126
MUS-1.CM.01.SIVmus_01CM1085     ---GS-RP.PE-EA..E-LPPDLSLA.NCSAR-GGGIGQSLSSS.......................................-SRGGETGQEE-LS--VIP-SG--VRPVE-E--VVSM---------IIQDAHI..HLE-P-N--SV-----N-S-KE-RGVQVCFN-K  126
MUS-2.CM.01.CM1246              ----GSRL.PVSEA..E-LP-YIAHA.YGTPRYS-AGELLPTSRAQG.......................................RATQG-ERQKD-LS--VIS--G--IK-VD-N--PVSM---------IIQDSDI..KIS-ELT--VV--L--L-N-K--KGVIVKFNEK  129
MUS-2.CM.01.CM2500              ----RFRL.PQ.EA..E-LPCYIAHA.ISSPRYS-DS-LLP.............................GGSTEGE...PNPPSPTGQEE-LS--VIP--G--IR-VMVH--PVSM---------IIQ-SEI..EIDKE-T--VV--L--L-N-K--KAVKV-FNDK  130
MUS-3.GA.09.09GabOI81           ----SFWS.PE-GT..K-LP-HPVGA...DG---DWGASQLFA...............................EGRG....AAEGEGRNQEN-LS--VIP--G--IR-VE-E--RVSV---------II--SHI..VLK-E-T---V-----L-N--E-RNVA-RFQDK  127
MUS-3.GA.11.11GabPts02          ----SFRS.PEEGT..K--PCHPAGTHSPTGFWQASQLCPESR...........................TS-H....LGGRGRDDKEL-LSL-V-P--G--FR-VS-E--EVSV---------II--G-V..QIK---T---V-----M-S--E-ENVQ-KFYDK  130
OLC.CI.97.97CI12                ----MEGR.ESR-SPQHATEDRDGHL.SLCTYPG.....................................TDFGEGNK.......KERDE-GEET-LSL-SIS--R--MIELY-E--QVT----------IKGKDI..YME-NGI-RI------A-R-M--RCRIRIF.EK  125
RCM.CM.00.SIVagi_00CM312        ----CSLE..AGET..E---FNLRDTLCSRDGSV-P.....................CGSHVKEILGAGEKAEAAKEAG.....EKPAGESIRRGIVL--FS--K--VTWVT-E--KV------------ISDIE-....GDN----I------Y-N-K--YNCK---A-K  138
RCM.CM.02.SIVrcm_02CM8081       ---KCALE..SGET..E-L-VDFLDSLCSRGG-GQS..............CLPVRSSRSHAKEIPGTGQNAETATEEG....REGSEGESLRG-TCL--FS--K--VTMVN-E--KV------------INGIE-....GEN----I------Y-N-K--YNCK---A-K  144
RCM.GA.x.SIVRCMGAB1             ----TPLG..SGEA..E-L-FDFLDSLCSRDGEQLRP.................CRRDAKELSEEGRGTKETTEAG....REQEER-SISG-L-L--FA--K--TTIAQ-E--KV-V----------I-GIE-....GND-T--I------Y-N-K--KNCE---A-K  139
RCM.NG.x.NG411                  ----CSLE..AGET..E-P-LDFLETY.....CS-..........ERGRMESPVRSSRRDAQEIPSIGETAQTGAEER....EQGESGESLRGCVVL--FS--K--VTWVE-E--KV---------G--IQNLD-....GEN----I------Y-N-K-FNHCK-T-A-K  143
SAB.SN.x.SAB1                   ---VWPLG..QRET.QEFP-DLHQTN.SSPNG-GLQQAGGKLVCRQTSDQRTRARRSSNSPVKAVCCSGETAETAVA.....KPLATTEPLRGGLQL---S--R--MK-VY-E--KVT---------S-IQGIE-....GDN---RI------C-N-KA-HNQEVK-EDK  157
SMM.CI.79.SIVsmCI2              ---IWPMG..KE-...SQLPHDP.......T-AGADTNGAPNGSS........CGGAPNGSSCGHVEELSETRSATER....EERKALQRGNGRLAA--FS--R--V--AY-E--PV-V--------SIVAGIE-....GLGYT--IV-------NT-E-TNVN--VL-K  142
SMM.LR.89.SIVsmLIB1             ---LWSLG..KE-...SK-PHGP.......NASGADTNCSP.................SRPSCGSAEELHAIGQEAEG....EQRKTLQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVAGIE-....GPNYT--IV-------NTKE-ENVE-TVL-K  133
SMM.SL.92.SIVsmSL92A            ---IWSLG..KE-...SQLPHGP.......NASGPDTDSPP.................SGPSSGSAEGLHEVGKETER....EQRETIQGGDGGLAA--FS--R--V--AH-E--PV-V--------SIVAGIE-....GLNY---IV-------NTKE-QKVK-QVL-K  133
SMM.SL.92.SL92B                 ---PWPMG..KE-...SQLPHAD.........NIIDTVSTP.................RSSSPDSQ.GVSRESTEGKD....KEEQALQGGDGG-AA--FS-RR--V-RAL-E--PL-V--------SIVAGIE-....GPGYT-RIV-------NTKE-RTVK--VL-K  130
STM.US.89.STM_37_16             ---LWPMG..KE-...PQLPHGP.......NTSGADANCPS.................RRPSCGSAEKLHAAGQEAER....EQE-TLQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVAGIE-....GLQYT--VV-------NTKEFKNVN--VL-K  133
SUN.GA.98.L14                   ---VWPLE..QG-A...--L--HAFCS.A..................SRGSDIGEQNDTTPVKGRESSGDLQTVRTGPQ....ST-EEE-KGGM-GAMLEYA-SR---EKVF-N--PVT----------IVS-D-V..SIE-H---RVV-----Q-R-KE-ADVFV--RDK  141
SYK.KE.x.KE51                   ----RVGE..QETP..SK--CDEGDT.NSPTSGGLAMAATDMGK.LCFTPTEAHSNGHPATATVLSKDSQGPLGPWETF...REER-DEGEGT-LSL-SIS--R--MME-NVE-TKVTM---------IIR-QDI..QLHQP-T--IV--L--N-R----N--RFQ-ERP  159
SYK.KE.x.SYK173_COMGNM          ----RMGG.LQE-P..SQLP--ERAI.CSP-GGH.......-RWAMAHMVSTDEPTSTSEGTELSLEETPHQQGSALA....EGEQWE-GDQE-LSLS-LS--R--MIEVDVE-D-VQM-V-------IIR-EDI..QLHQP-S--IV--L--N-T----RN-RFTVVKP  153
TAL.CM.00.266                   ----VFFR..ENPT..GQRTE-FS........SGATIGTAN.................RASHSIMAGGDYHVDPTSQE.....APRNNHNKGE-LSLFGQS--R--I-KVL-E--EV-V---------IIQVGEI..ELE-VPHQ-TV--LA-Y-Q-QS-SA-E-QWENK  134
TAL.CM.01.8023                  ----GFFR..EDPP..GQGTE-LPSG.ATHRTAD.......................RASHSIMAGGDYHMDPSDQG....CRSTQATDEGE-LSLVSQS--R--I-KVL-E-REV-V---------IIQ-GDI..ELE-VPHQ-TV--LA-Y-Q-QS-TGVE--WEHK  135
TAN.UG.x.SIVagmTAN1             ----NANL..-NKA.AQF-R.ARRSR...SDGSPDACARIP.................HRFASCGSIRPSQETAGAVCQKG..RSIEKAEGEGTGGL-LEFP--R--MSK-Y-Q--PVQV---------IIK-EDI..HLE---R--VV-----G-S-K---N--VN-D-K  143
VER.DE.x.AGM3                   ---VWPVD..-SET.QKFSRRYSWGG.ANCA-STES.....................IRPCKEAPAAICRQGEAVEGTKE...KTTSSESRLDRGIF-.ELP--R--IK-VY-E-VPIR----------IIK-AD...QLS-T----I------GLN-KE-SDREVRLEDK  140
VER.KE.x.9063                   ---VWTVD..-DKT.EKLSRCHSWSG.TERA-STDT.....................IRPSKEAPAAVCREGKTVEGTRE...ENTTNKSRLGGRIF-.ELP--R--IK-VI-E-VPVKV---------IIK-TDI..QLT-P-R--V------GLN-KE-SNVEVK-EDK  140
VER.KE.x.AGM155                 ---VWPVD..-DKT.KKFSR-HSWGG.TKCA-STEQ.................LYTLRPSKEAPAAVCRERETNEKSEQ...KPPSE-SRLERGIF-.ELP--R--IR-CI---TAVK----------IIKDTD...QLR-S-R--IV-----GLN-KE--NVEVQLEDK  143
VER.KE.x.TYO1_patent            ---VWTVD..-GKT.EKFSRRYSWSG.TECASSTERHHP.................IRPSKEAPAAICRERETTEGAKE...ESTGNESGLDRGIF-.ELP--R--IK-VY-E-VPIK----------IIK-ND...QLS-P-R--I------GLN-KE-NDREVK-EDK  143
WRC.CI.97.97CI14                ----WQRA..DALEQLSSSEA-ELEQ.FWANNNNAFCSSNE-GSKH.........SGAGKIFIEGSTIEGTTGTQE......DRGTEE-SGGAGMAL--IS-RT--MIQCI-E-HSVS----------IDVNL.....NKN-Q-RI------TMN-KLHR-VK-R-L-K  147
WRC.CI.98.98CI04                ----WERA..DALKQLPSSKG-ELEQ.FWAG.NNTVCSSNEGGSKH.........TRIREIPPERSSIEGAGTGK.......KGREEERNGAARVA---IS-RN--MIQCI-E-NSVS----------IDIHLN....NKN-Q-RI------TMN-KL-RK-Q-R-L-K  146
WRC.GM.05.Pbt_05GM_X02          ----WPR...AE-...LQQL-SS..........KN-CRGEYLWASG.....ANSTSGSYHARTRAISSERSSDEEG......ERERNAEGGESRISL--IS-RT--MIQCI--RESVS----------VDVQL.....E-N-Q-RI------TMQ-KL-KDVR-K-L-K  138
```

```
                                                                                                               K70R
                                               protease end p66, p51 RT start              M41L            D67N                               D110 catalytic site
H1B.FR.83.HXB2                KAI.....GTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETV...PVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL......KKKKSVTVLDVGDAYFSVPLDEDFRK 281
H1A1.UG.85.U455_U455A         -T-.....----------------M------------------...------E-------------------T---N---------------------------------------------------------T---......----------------------S--- 280
H1C.ET.86.ETH2220             ---.....----------------M---L-R------------...--------------------------TA--E---Q-----R---------------------------------------------------......----------------------G--- 288
H1D.CD.84.84ZR085             ---.....-----------------------------------...--------------------------T----D--------R----------I----------------------------------------......-----I-----------I--C----- 281
H1F1.BE.93.VI850              ---.....----------------M-----------V------...--------------------------T---L-----------------------------S--------K----------------------......---------------------K--K- 281
H1G.SE.93.SE6165_G6165        ---.....--I------I------M------------------...-----------R--------------T---K---E----------------I----------------------------------------......-------------------------- 278
H1H.CF.90.056                 ---.....----------------I------------------...--------------------------T-------------R-------S--I----------------------------------------......------S--------------KE--- 280
H1J.SE.93.SE9280_7887         ---.....----I-----------M---L--------------...------------I-------------TQ--A---E-----RV--------------------------------------------------......--------------------Y----- 279
H1K.CM.96.96CM_MP535          ---.....-----------------------------------...--------------------------T-----------------------------------------------------------------......---------------------K---- 281
H1O.BE.87.ANT70               EVQ.....----------------I--GL-----------AP-...------------------SK---E--TA--Q---Q-----R----------I-------G-----------------E-----------G--......-Q-Q-------------C---P---- 277
H1O.CM.94.BCF06               EVQ.....------------L---I--GL-----------AP-...------------------SK---E--TA--Q---Q----------------I-------G-----------------------------G--......-QRQ-------------C---P---- 277
H1O.FR.92.VAU                 EVR.....------------L---I--GL-----------TP-...------------------SR---E--T---K-L-Q-----R----------I-------G-----------------------------G--......-Q-Q-------K-----C---PE--- 278
H1N.CM.02.DJO0131             --V.....----------------F------------------...-----------R-------A---E--RQ--A---Q-----R----------I----------------------------------------......Q----------------C-------- 286
H1N.CM.04.04CM_1131_03        --V.....----------------F-----------L---K--...-----------E-------A---E--R-------------R----------I----------------------------------------......-Q---------------C-------- 288
H1N.CM.95.YBF30               --V.....-----------------------------------...-------------------T---E--R-------------R----------I----------------------------------------......-Q---------------C---K---- 288
H1P.CM.06.U14788              QVY.....----------------I--L-----S-------I-...-----A-----R------SK---E--RA--Q---Q----T-----------I-------GS----------------------------G--......Q-R----I---------C---P---- 278
H1P.FR.06.RBF168              EVY.....----------------I--L-----S---S---I-...-----A------------SK---E--RA--Q---Q----T-----------I--------S----------------------------G--......--R--------------C---P---- 277
CPZ.CD.06.BF1167              TVK.....TK--I-----------V-R-L---------TV-I-...K----E------------SR---E--T---ENLRA----IEV--D------I----------------------------------------......--R-VI-----------I---PE-Q- 284
CPZ.CD.90.ANT                 TVL.....A---L--N--------V-CLL---------KV---...-----E-----R------SK---E--K---DKL-A-N---R---D------I-------TS--------------------I----------......-Q---------------I---Q---- 281
CPZ.CM.01.SIVcpzCAM13         -VQ.....----------------I--------V----S----...-----------RI-----SA------T---Q--------------------I--------S-------------------------------......-E---------------C---K---- 280
CPZ.CM.05.LB715               R-V.....----------------I------------------...--A--------R--------------T-----------------------------------------------------------------......-----------------C----N--- 286
CPZ.GA.88.SIVcpzGAB2          RVE.....S---------------V---L-----------TP-...-------T---R------SK---E--T---R---------R----------I---R------------------------------------......--------P--------C----N--- 284
CPZ.US.85.US_Marilyn          --Q.....-S--------------I-------------S-K--...--R--------R-------A------T---Q--------TRV---------I----------------------------------------......---R-------------C---KE--- 283
GOR.CM.04.SIVgorCP684con      -VV.....---------A------V--K-------------V-...--S---------------SK---E--TA--Q---Q----------------I-------------------------------------G--......-----------------C---P---- 273
GOR.CM.07.SIVgor2139_287      -VV.....----------------I--K----------S-DV-...--S---------------SK---E--TA--Q---Q----------------I-------------------------------------G--......-----------------C---P---- 273
GOR.CM.07.SIVgorCP2135con     -VV.....----------------I--K----------S-DV-...--S---------------SK---E--TA--Q---Q----------------I-------------------------------------G--......---R-------------C---P---- 273
GOR.CM.12.SIVgor_BPID1        EVS.....------------V---I--LV------------V-...-----------R------SK---E--TA--Q---Q----T-----------I--------S----------------------------G--......-QR--------------C---P---- 278
GOR.CM.12.SIVgor_BQID2        EVQ.....------------V---I--GL-----------AP-...------------------SR---E--TA--Q---Q-----R----------I-------------------------------------G--......---Q-------------C---P---- 277
GOR.CM.13.SIVgor_BPID15       EVH.....----------------I--S-------------V-...------------------SK---E--TA--QD--Q----TR----------I--------S----------------------------G--......-Q---------------C---P---- 276
                                               protease end p51 RT start                                                                                      D catalytic site

MAC.US.x.239                  RIK.....--IMT-D--I--F------AL-MS-----AKV-P-...K-A----K----L-----SK---V--R---EK---D-QLEEAP-T------T-------KN---M-I------RV----T---------------......A-R-RI----I------I----E--Q 293
H2A.DE.x.BEN                  RVR.....A-IMT-D--I--F---I--AL-MS--L-VAK--PI...K-T----K--RL-------K---E--K---EK-----QLEEAP-T------T-------KN---M-I-------V----T-I-------------......A---RISI---------I--H----Q 306
H2B.CI.x.EHO                  RVR.....A--MT-D--I--F---I-NSL-M-----VAR--P-...K-Q---EK----IR---SK---L--K---EK-----QLEEAP-T----S-T-------KN---M-I-------V--E-T---------------......AS--RI---------------P---Q 306
H2G.CI.92.Abt96               RIQ.....A--MT-D--I--F---I-VKL-MS-----AKV-P-...K-R---------IR---SK---Q--R---DK--Q--QLEEAP-T------T-T-----KN---M-I------RV----T---------------......AE-RRI----I-----I---V---Q 289
H2U.CI.07.07IC_TNP3           -VR.....A-IMT-D--I--F---I-VKL-MS---VAKV----...K-T---------IR---SK---E--K---EK--Q--QLEEAP-T------T-----R-KN---M-I------RV----T---------------......AE-RRI-----------I----Q--P 289
H2U.FR.96.12034               TIR.....--IMT-D--I--F------EL-MS-----GKEAL-...IGG----K----LR---SK-LE--K---EK-----QLEEAP-T------T-------KN---M-I------RV----S---------------......AE--RI-----------I--C----Q 303
ASC.UG.10.RT03                RVT.....EE----D--I--M---F--KF-V---MVVGKFKPT...E-R-RE-K----------SA---Q--K--VQD-VDK-QLE---------S---C-R---K----M-I---K--EH---LA-----------------......P--AQ-SIV-IK---YAI--H---Q- 291
ASC.UG.10.RT08                RVT.....-E----D--I--M---F--KF-V---MVVGKFKPT...E-R--E-K---R------SA---Q--K--VQD-VNK-QLE---------S---C-----K----M-I---K--EH---LA-----------------......P--AQ-SIV-IK---YA---H---Q- 294
ASC.UG.10.RT11                RVT.....EE----D--I--M---F--KF-V---MVVGKFKPT...E-R--E-K----------SA---Q--K--VQD-VDK-QLE-V-------S---C-R---K----M-I---K--EH---LA-----------------......P--AQ-SIV-IK---YA---H---QR 294
COL.CM.x.CGU1                 EYR.....-DI-I-H--I--V---M-KEGKLVVAQLSDR-PVT...K-T--E------------SK---EG-QK--ERL-A----E-AELG------I-C-R---KNE----I------------M---------G--......IE-THI----I------I---P-YQ- 263
COL.UG.10.BWC01               EY-.....-DI------I------I-KDGKIV-AQLSDR-PIT...K----E-----R------SK---EG-QK--ERL------E-AAHG------I-C-R---KNE----I-------------F---Q-----G--......VE--NI----I------I---P-YQQ 265
COL.UG.10.BWC07               EYK.....-DI-LAD--I--L----IGEGKIV-AQLSAN-PVT...K-S----L---R------SK---EG-QK--DRL-A----E-AALG------I-C-----KNE----I---------------I-----G--......HE--NI----I----------PEYQ- 265
DEB.CM.04.SIVdeb04CMPF3061    V-V.....-E----E--I--L---F-APL-FSI-MLE-Q-QYT...-----E-KS-------ASSQG-DRGSQGNSGKDGGGRPMAPAS-T----S----------KN---M-I-L-----A--PV----T------------......PQ-RQI----IA---Y-I---PS-AQ 290
DEB.CM.99.CM40                V-V.....SDL---D--I--L---Y-AKM-VSI---VLDYT.....-----E-A----I-----SR---E--KD-LKP-IEA-Q-VPAA-T----S---V-----KS---M-I-L-A---A--EV----T-----SA-......PQMEQI----LA---Y-IA--PT-A- 287
DEB.CM.99.CM5                 V-V.....SD----D--I--L---Y-AKM-VSI---VLDFT.....-----E-A---RI-----SR---E--K--IRP-VEA-Q-VPTE-T---SS---V-----KN---M-I-L-A---A---V----T-----S--......PQMEQI----LA---Y-I---PQ-AR 287
DEN.CD.x.CD1_CM0580407        V-V.....D--------I--L---Y-SKL-V---MAVQQMRPQ.VT-------F--RI-----T--L---QQ-V-Q-----RLE-AP-T---------V-P---K-QY-M-I---K--EA--------------------------......QQ-EHI----LK---Y-----P--AQ 307
DRL.DE.11.D3                  ITH.....AS--------------V-CKL-A-----V-QV-P-...K-T---------I-----SK---E--K---EQ-----Q--R-------------C-----G----------Q---A----F---------G--......-ACEQI-----------C---P---- 295
DRL.DE.11.D4                  ITH.....AS--------------V-CKL-A-----V-QV-P-...K-T---------I-----SK---E--K---EQ-----Q--R-----A------C-----G----------Q---A----F---------G--......-ECEQI-----------C---P---- 295
DRL.DE.11.D5                  ITH.....AS--------------V-CKL-A-----V-QV-P-...K-T---------I-----SK---E--K---EQ-----Q--R-------------C-----G----------Q---A----F---------G--......-ACEQI-----------C---P---- 295
DRL.DE.11.D6                  ITH.....AS--------------V-CKL-A-----V-QV-P-...K-T---------I-----SK---E--K---EQ-----Q--R-------------C-----G----------Q---A----F---------G--......-ACEQI-----------C---P---- 295
DRL.x.x.FAO                   ITH.....AS--------------V-CKL-A-----V-QV-P-...K-T---------I-----SK---E--K---EQ-----Q--R-------------C-----G----------Q---A----F---------G--......-ACEQI-----------C---P---- 295
GRV.ET.x.GRI_677_gri_1        IIT.....-SI-I-S--I------I-A-A-MK-VMGVLSSQIE.ETK-Q--E-K----L-----SR---E--T---KQ--E---L-R--G-----------------K-Q--M----------A----F---------------......Q---QI--I-I----Y-I--CKE--- 303
GSN.CM.99.CN166               RIK.....A-----------L---C-SKL-V---MVQQK--PI...E-H--A-KE---I-----SR---E--TQ-TQ--LRL-QLE---------S---------KSQ--M-I---K--EA---LA------------IPHPAGLEQ-EH--II-MK---Y-I--Y----- 292
GSN.CM.99.CN71                RIK.....A-----------L---C-SKL-V---MVQQKL-P-...E-H--S-KE---I-----SK---E--TQ-TQ--L-L-QLE---------S---------KSQ--M-I---K--EA---LA-----------------......EQ-EH--II-MK---Y----Y-E--- 285
LST.CD.88.SIVlhoest447        TTV.....----L-----D-V---I--NV-AR-VLSQLSDKIP.ITK------C---R------SK---EG-QA--DRL--------ADAG------I-------KNE----I---Q--AL----H------------I.......--C-Q---V-I-------I---P-Y-Q 297
LST.CD.88.SIVlhoest485        TTV.....----L-----D-V---I--SV-AK-VLAQLSDKIP.ITK-R----C---R------SK---EG-QA--DRL--------ADAG------I-------KNE----I---Q--AL---LH------------I.......--C-QI--V-I-------I---P-Y-Q 297
LST.CD.88.SIVlhoest524        TTV.....----L-----D-V---I-NPL-AR-VLAQLSEKIP.ITK-H----C---R------SK---EG-QA--DRL-------PVD-G------I-------KNE----I---K--EL----H-L-----------I.......--C-QI--V-I-------I---P-Y-Q 298
LST.KE.x.lho7                 TTV.....----L-----D-V---I--AM-AK-ILAQLSDKIP.ITK-S----C---R------SK---EG-QA--DRL-------PVD-G------I-------KNE----I---K--EL----H-L-----------I.......--C-QI--V-I-------I---PNY-- 297
MAC.US.x.251_BK28             RIK.....--IMT-D--I--F------AL-MS--L--AKV-P-...K-T----KV---L-----SK---V--R---EK---D-QLEEAP-T------T-------KN---M-I------RV----T---------------......A-R-RI----I------I----E--Q 289
MAC.US.x.EMBL_3               RIK.....R-IMT-D--I--F------AL-MS--L--AKV-P-...K-A----KV---L-----SK---V--R---EK---D-QLEEAP-T------T-------KN---M-IH-----RV--ELYRSPIR----------......A-R-RI----I------I----E--Q 289
MND-1.GA.x.MNDGB1             GTF.....----I----ID-----IMEKL-GK-ILAQLSDKIP.ITK------V--RI-----SK---VG-QK--DRL-E-----RVD-G-N----I-------KNE----I--------L----H-L----------I.......--C-RI----I-----I---P-Y-P 290
MND-2.CM.98.CM16              TTH.....AS----------V---V-KKL-------V-KV---...K-T----T---I-----SK---L--Q---SQ-----Q-----------------C-----G-S---------Q---V----F---------G--......-QCEQI----I-----C-------- 293
MND-2.GA.x.M14                STH.....AS--------------V-KKL----------K---...K-T--------RI-----SK---L--QT--D----------R-------------C-----G-S---------Q--RV----F---------G--......-QCEQI-I--I-----C----S--- 294
MND-2.x.x.5440                ITH.....AA--------------V-KKL----------K---...K-T----T---RI-----SK---L--Q---NQ--------RV-----------C-----G-S---------Q---V----F-----T--G--......-QCEQI----I-----C---V---- 294
MNE.US.x.MNE027               RIK.....--IMT-D--I--F------AL-MS-----AKV-P-...K-T----K----LR---SK---V--R---EK---D-QLEEAP-T------T-------KN---M-I-------V----T---------------......A-R-RI-----------I----E--Q 289
MON.CM.99.L1_99CML1           EVW.....A--------I--L---V-AKM-V---MAAGGDLQPPSLE-T--A-KE---L-----SR---E--TQ-TQ--L-L-QLEPTE-N----S-C-------KS---M-I-------A----F---------------......E-MDH--IV-MK--FY-I--W-P--- 289
MON.NG.x.NG1                  TVR.....A---L----I--L---V-SRM-I---MXQGIAQSP.SLT-T----K----L-----SK---E--TT-XR--IEL-QLEEAE-T----S-C---R---K----M-I-------A----F---------------......QNMANI-IV-MK--FY-I----N--- 292
MUS-1.CM.01.CM1239            TIK.....A--------I--M---C-SKF-I---MIQEK--PI...K-A--E-AK--M-------Q------EG-VQQ-L-LDQ-EE---D----S-C---R---KS---M-I---Q--EA--E-T---------------......AEHEH--IV-IK--FY-----PA--- 282
MUS-1.CM.01.SIVmus_01CM1085   TIR.....A--------I--M---C-SKF-I---MIQGKL-PI...K-T--E-AK--M-------Q---R--ED-TQQ-L-V-Q-EE--------S-C---R---KS---M-I---Q--EA--ELA---------------......IEHEH--IV-IK--FY-----P---- 282
MUS-2.CM.01.CM1246            RIV.....A----S---I--L---C-SKLEI---MVMAGQQLK.PTK-S--ERKK--M------SK---E--K--TQK-I-L-QLEEA--N----S-----R---K----M-I---K--EV--E-A---T-----S--......AQ-AH--IV-MQ--FY-----KE--P 287
MUS-2.CM.01.CM2500            RI-.....A--------I--L---C-SKL-I---MVGGQL-PT...K-G--A-KE--M------SK---E--E--TQR-L-L-QLEET--Q----S-A---R---KS---M-I---K--EA----S---T-----S--......S--AH--IV-MQ--FY-I--Y-E--P 286
MUS-3.GA.09.09GabOI81         YIV.....A-----------M---C--KL-F---CVQTSL-PL...K-R--E-KE---------A---D--IN-TQ--L-L-QLEEA--D----S-C-------KS---M-I---K--ES----S---------------......SR-RH--IV-MK---Y-----P---- 283
MUS-3.GA.11.11GabPts02        EVY.....AN--I-------L---V-S-M-V----VTG-K-I.....-----E-K---------A------EQ-TQ--V-L-QLEE--------S-C---R---K----M-I---K--EN---LF---------S--......AEHEH-----MK---Y-I-------- 286
OLC.CI.97.97CI12              EIQ.....EA----D----VV---V-AAV-IR-VCAQLSDKIP.ITK----E-HKP-QI-----SK---E--EK--DRL-E-----Q-AEAS-QW---I-C-----KNEY---I---A--AA--------------------......-EREQI----I------I---PEY-E 283
RCM.CM.00.SIVagi_00CM312      VTH.....AH--------------V-HKL-AS------KA-II...K-E--E-Q---R------SK---E--T---NA--------R-----------I-C-R-------------------------F---------G--......--C-QI----I-----C--Y-P--- 294
RCM.CM.02.SIVrcm_02CM8081     VTH.....AH--------------V-HKL-AS------KA--I...E-----Q---------SK---E--T---NA--------------------I-C-R-------------------------F---------G--......--C-QI----I-----C--Y-P--- 300
RCM.GA.x.SIVRCMGAB1           RTH.....AH--------------V-KKL-A-------Q---I...K-E--S-Q---R------SK---E--T---NA--------------------I-C---------------------------F---------G--......-QCERI----I------CL-Y-P--- 295
RCM.NG.x.NG411                TTH.....AH--------------V-SKL-AS------KA--I...K-E---Q---R------SK---E--T---NA--------R-----------I-C---------------------------F---------G--......-QCRQI----I------C--Y-P--- 299
SAB.SN.x.SAB1                 TCK.....A-I---E---------V-A-L-V---LTQRE--PI...K-H----Q---RIR----SK---E--KA--EDL--Q-HLER--------------R---K-Q--I-M---Q---S----Q---------------......QQREQI----I------C---P--Q- 313
SMM.CI.79.SIVsmCI2            RVK.....--IMT-D--I--F---I--AL-MS--C-LAKV-P-...K-T----KE---L-----SK---E--R---EK-----QLEEAP-T------T-------KN---M-I-S-----V----N---------------......AE-RRI----I------I--W-E--Q 298
SMM.LR.89.SIVsmLIB1           RIR.....--IMT-D--I--F------EL-MS----VAKV-P-...K------K----LR---SK---E--K---EK-----QLEEAP-T------T-------KN---M-I-------V----T-I-------------......A-RRRI-----------I----E--Q 289
SMM.SL.92.SIVsmSL92A          EVR.....A-LMT-D--I--F---I--AL-IS--Y-VAKVQP-...K-T----K----QR---SK---E--K---EK-----QLEEAP-T------T-------KN---M-I-------V----T-I-------------......QQRRRI-----------I---P---Q 289
SMM.SL.92.SL92B               VIR.....E-LMT-D--I--F---I-AAL-VS--Y-VAKV-YT...K----E-----RL----SK---Q--T---EK-----QLERAP-T------T------NKD---M-I-------M--E-T---------------......-EMERI----I----------PE--Q 286
STM.US.89.STM_37_16           -IK.....--IMT-D--I--F------AL-MS----VAKV-P-...K-T----K----I-----SK---E--K---EK---D-QLEEAP-T------T-------KN---M-I-------V----T-I-------------......A-RRRI-----------I----G--Q 289
SUN.GA.98.L14                 T-T.....----------D-----I-SAL-GR-VLATLSEKIP.ITK----E-AA---------SR---EG-QK--DRL-A-----RAD-G------I-------KNE----I---V--GM----H-L-----------------......--C-QI--V---------I---P-Y-- 299
SYK.KE.x.KE51                 DGRIKEVE-SL---------L---I-SKL-AK-VMIAQKL-PM...K-Q----KEL--L------R--LE--RA-VSD-LEK-QLE-AS-T--------V-Q---KL---MPM---K--EA----------------G--......ST-R.M----LK---YT-----E--Q 319
SYK.KE.x.SYK173_COMGNM        SGKRKQVE--L---------L---I--KL-VK-VMVQTAL-P-...K-S---DKEL-RL----SV--LE--KA-VED-L-A-QLE-AS-T--------V-R---KK---M-I---K--EA----F---------G--......-RQ-.L-II-LK---Y-----KE--P 313
TAL.CM.00.266                 R-R.....D---I----I--L---F-AKFSV---MVS-D-PIT...K----E-KEP--I------R---EG-QI-IDD-V-A-QLEEV-------S-----R---K-Q--M-I---A--QA----A---Q-----S--......PQM-QI----MK---Y-I---P--SQ 290
TAL.CM.01.8023                R-R.....D---I----I--L---F-AKF-V---MVSTE-PVI...K----E-KEP--I------K---EG-QA-IDD-V-A-QLEE--------S-----R---K-Q--M-I---A--QA----A---Q-----S--......PQ--QI----MK---Y-I---P--AR 291
TAN.UG.x.SIVagmTAN1           ETR.....-N--I----I--V---I-APL-AR-IMGQLNEKIP.ITK-Q-----S---L------S-------T--TQAL--------V-G---------C-R---K-Q--M-I---A---A----F-------------------......-R--QI----I----Y-I---PE--P 301
VER.DE.x.AGM3                 ILR.....--I-I-S--I------I-APA-AK-VMGQLSEQIP.IT-----E-AR--FL-----SK------Q---DQL-----------G--A------C-----KSQ--M----------A----F---------S-F......E-MTEI----I----Y-I---PE--- 298
VER.KE.x.9063                 ILR.....----L-A--I--------APA-AR-VMGQLSEQIP.IT-----D-AR--HI-----SK------T---EQL-----------G--A----I-C-R---K-Q--M---------V----F-------------------......Q-M-QI----I----Y-I---PE--- 298
VER.KE.x.AGM155               ILR.....----I-A--I------F-A-A-AK-VMGQLSQTIP.IT--R--E-AR--RL-----SK---I--Q---KTL-E---L-RV-GD-A------C-R---KSQ--M----------A----F-------------------......--M-QI-II------Y-I---PE--- 301
VER.KE.x.TYO1_patent          ILR.....--I-L-A--I--------APA-AR-VMGQLSEKIP.VT-----E-AR--C-R----SK---E--Q---SQL-Q-----RV-G--A----I-C-----KSQ--M----------A----F-------------------......R-MRQI----------Y-I---PN--- 301
WRC.CI.97.97CI14              TIE.....ADIMT-K--I------V-NKM-MH-A-AQLSDKIP.ITK-S----CS--R------SKK--EG-QK--DRL-A-----RAD-G------I-------KNE----I---K--EV----F---------G--......R-N-NIS-V-I------I---PEY-Q 305
WRC.CI.98.98CI04              TVE.....ADIMT-K--I------V-EKM-MH-A-AQLSDKIP.ITK-S----CT--R------SK---EG-QK--DRL-A------AS-G------I-------KNE----I---K--EV----F---------G--......R-N-NIS-V-I------I---P-Y-Q 304
WRC.GM.05.Pbt_05GM_X02        EVR.....ANLMT-K--I------V-EKL-MH-A-AQLSDKIP.ITK-S----CQ--RI-----SK---EG-QK--DRL-A-----RAEAN------I--R---KNE----I---G--AV----F---------G--......H--RNI----I-----FI---PEY-Q 296
```

```
                                                                                     K219Q
                                                   polymerase motif              T215Y
H1B.FR.83.HXB2              YTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDS..WTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIP 449
H1A1.UG.85.U455_U455A       ---------------V---------------S-------------S-H----------------------A-----A---S--FI------------------------------Q------..-------------------A---K--------A---DIVT 448
H1C.ET.86.ETH2220           --------T---------------------P------PQ------AP--E--------------------AP-----E---K--F------------------------------Q------..-------------------------------A---DIVT 456
H1D.CD.84.84ZR085           ------------------------------------I----------EV---------------------A---K--E------------------------------------S-T----E-..-----------------------K---------------- 449
H1F1.BE.93.VI850            --------V-------------------------C---------MK------------------------------E-----F---------------H-------------Q--N---..-------------------------P------A---DIV- 449
H1G.SE.93.SE6165_G6165      ---------------V--------------------R------AN--EM---------------------A-----E---K--F------------------------------Q--D-E-..----------------------TH-------A---DIVS 446
H1H.CF.90.056               ----------------------------------------A---E---EM--------------------A-----A---K--F----------------------------TVK------..-----------------N---K--------A---DI-- 448
H1J.SE.93.SE9280_7887       ----------------------------------C------K---ER--E-------------------E---R--K---E---K--F------------------------------Q----ED..----------------------K------K-A---DIV- 447
H1K.CM.96.96CM_MP535        ---------------V------------------H---------IK--EM--------------------P-------E---K--F------------------------------Q--D---..----------------------K---------V---DIV- 449
H1O.BE.87.ANT70             --------V---------------------------------D---RD--ELE-C-----------PLTE--KR--L--E--YQ--F----------------------------S-Q--N--V.---------I---------Q--R--E----I----S----V- 445
H1O.CM.94.BCF06             --------V------V-------------------I-------S--E-E-----------------PLTE--K-V-L--E--YQ--F----------------------------T-Q--N-EV.-----------------Q--R-KE----I----S----V- 445
H1O.FR.92.VAU               --------V------V--------------------------D---N--ELE--------------PLAE--KRV-L--K--YK--FY-----------Q-----------------Q--D-EV.-----------------H--R-KE----I----S----V- 446
H1N.CM.02.DJO0131           --------T---------------------------------------KH-E-I------------S-LA---ETV---S---K--F----------------------------T-K---Q-V.----------------------K---R-I---------VN 454
H1N.CM.04.04CM_1131_03      --V-----V-----------------E----------T------KH-E-I--------------------AK--ETV---G---K--F-------------------------R--K-----T.------------------------------I---------VT 456
H1N.CM.95.YBF30             ---------------------------------------T---------EKH-E-I-------------------LA---EAV-D--D---K--F------------------------------K-----V.----------------------R-K-----I---------VN 456
H1P.CM.06.U14788            --------V--D---L--V-----------------H-----------S--EVE------------PLSE--QRV-K--E--YV--F------------------------------K----E-..----------------------R-KE----I----S----VA 446
H1P.FR.06.RBF168            -X------V---------------------------Y----------ES--EVE------------PLSE--QRV-K--E--YV--F------------------------------K----EA..----------------------R-KE----I----S----VA 445
CPZ.CD.06.BF1167            --------V---R--K--C-T---------------Y------D---KH-E-DV------------TAEK-KEVVQ---N--KA---E-----Y-EK-----------E-----Q-K---PET.----------------------RT-E----I--V-G--D-VQ 452
CPZ.CD.90.ANT               --------V---------C-----------------A------A---DKY-AVE------------M--TA--EM--K-----QV---E-----------Q--------------K-K---P-D.----------------------TK---R-I--V-S--DRVQ 449
CPZ.CM.01.SIVcpzCAM13       --------V---------------------------------------AK----TV----------D-T---Q-V--------H--F------------------------------K----EV.-----L---I------------IK-----I----R--D-V- 448
CPZ.CM.05.LB715             --------V---------------------------A--------Y-----E-I------------K-EL--Q-V----D-----F------------------A----Q--D-E-..----------------------K-----I----S----VT 454
CPZ.GA.88.SIVcpzGAB2        --------V----------------------------------V-D---QK--EL-----------SLEE--KRV-Q--E-----F------------------------------Q--Q-EN.----------------------IK-----I--A-----IVE 452
CPZ.US.85.US_Marilyn        --------V-----------------------------------D-----H--VI-----------NLEK--E-V-L---Y--T--F-------E----------------------Q--Q-EI.----------------------IK-----IK-A-----VN 451
GOR.CM.04.SIVgorCP684con    --------V------V--------------------H-----D---K--E-E--------------P-TE--Q-V-K--E--YA--F------------------------------K----E-..----------------------R-KE----I----S----VT 441
GOR.CM.07.SIVgor2139_287    --------V------V--------------------H-----D---K--E-E--------------P-TE--Q-V-K--E--YA--F------------------------------K----E-..-------R--------------R-KE----I----S----VT 441
GOR.CM.07.SIVgorCP2135con   --------V------V--------------------H-----D---K--E-E--------------P-TE--Q-V-K--E--YA--F------------------------------K----E-..----------------------R-KE----I----S----VT 441
GOR.CM.12.SIVgor_BPID1      --------V------V--------------------H-----------E--E-E------------PLPE--QRV-K--E--YA--F------------------------------K---RE-..----------------------R-KE----I---------- 446
GOR.CM.12.SIVgor_BQID2      --------V------V--------------------------D---D--E-E--------------PLAE--R-V-V--E--YQ--F------------------------------K--D-EV.-------X-----------Q--R-KE----I----S----V- 445
GOR.CM.13.SIVgor_BPID15     --------V------V--------------------H-----D---N--E-E--------------PLSE--QRV-K--E--YA--F------------------------------K----E-..-------G--------------RIKE----I---------VT 444
                                                   DD catalytic site

MAC.US.x.239                -----L--V--AE--K--I-K-------------YT-RHV------A---VTLV-----ILIA--RTDLE-DRVVLQSKEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--QRET.-----------V----A-------TKH--R-I--KMT---EVQ 461
H2A.DE.x.BEN                -----L-AV--ME--K--I-K-------------YT-RQV------A---VILI-----ILIA--RTGLE-DKVVLQ-KEL-NGL-FS---E-F--D---Q---C--W-T--KL-KLQ--Q--I.-----------V----A---S---TKH--R-I--KMT---EVQ 474
H2B.CI.x.EHO                -----L-AV--AE--K--L-K-------------YT-A-V-D----A-N-VT-I-----IL-A--RSDLE-DRVVSQ-KEL-NNM-FS--EE-F--D---K------W-K--KL-K-Q----EV.-----------V----A-LF----T-HI---I--KMT---EVQ 474
H2G.CI.92.Abt96             -----L--V--AE--K--I-K-------------HT-R-V-D----A-E-XI-I-----IL-A--RTDLE-DQVVLQ-KEL-NKT-FS---E-F--D--YQ------W-K--KL-K-N----ET.-----------V----A-L-----TKN----I--KMT---EVQ 457
H2U.CI.07.07IC_TNP3         -----L--V--AE--K--I-K-------------YT-R-V-D----A-E-VI-I-----IL-A--RSDLE-DKVVLQ-KDL-NQL-FS---E-F--D--YQ------W-K--KL-K-N----EV.-----------V----A-------TKN----I--KMT---EVQ 457
H2U.FR.96.12034             -----L-----AE--K--I-K---------T---YT-RNL------A-E-A--I-----ILIA--RNDLE-DQLVSQ-KEL-NKT-FS---E-F--D--HK------W-K--KL-K-E---R-V.-----------V----A-V-----TKH----IK-KMT---EVQ 471
ASC.UG.10.RT03              --------V--MG--E---FK-------A--T--NTIGDL-QQI--KY-Q-LMI-------LI---ETVTE--KIV-QI-NI--KQ--Q--EE-Y-PQR-VN-L----R-RE--IPR-T--V-QV..Y---E-------Q-------------TKA----I----G---EVQ 459
ASC.UG.10.RT08              --------V--MG--E---FK-------A--T--NTVGDL-QQI--KY-Q-LMI-------LI---ETVTE--KIV-QI-NI--KR--Q--ED-Y-PQR-VN-L----R-RE--IPR-T--V-QV..Y---E-------Q-------------TKA----I----G---EVQ 462
ASC.UG.10.RT11              --------V--MG--E---FK-------A--T--NTIGDL-QQI--KY-QVLMI-------LI---ET-TE--KIV-QI-NI--KQ--Q--EE-Y-PQR-VN-L----R-RE--IPK-T--V-QA..Y---E-------Q-------------TKA----I----G---EVQ 462
COL.CM.x.CGU1               --------V--LA--T--V-K---------T---GTVAR-------RGQL.Q-A---------L----PLEE--RAVQ---DA--YY--E--E--Y-AD--YK-------KQ-KL-KVKI--Q-K.------------A---L--L----RTKEIS-CTK-K-H-L-EVQ 430
COL.UG.10.BWC01             --------V--LS--T--I-K---------T---ATVA-----I--KGKL.Q-A--------------PKEE--KE-Q---DA--YY--E--E--Y-A---YK-------QE-KL-KVKI--QET.------------G---L--V----RTKEIS-CTK-K-H-L-EVQ 432
COL.UG.10.BWC07             -----V--V--MS--Q--V-K---------T---GTVA-----V-REGKL.H-A-------I----D-KE--QEVQKV-TL--HY--E--E--Y-ET---K-------G--KL-KVKI---EN.------------A---L-------RTKE-S-CTK-K-G-LDEVQ 432
DEB.CM.04.SIVdeb04CMPF3061  --------V--TA--E--EFR------SA-----HTVHRV--G---KY--VL-V-------L----RTEE---LVVKT--DY--ER--K--PE-F-A---YQ----V---K---T----QI--EEE.----K------I-------S---TKH---SI--VPG--DPVK 458
DEB.CM.99.CM40              -----V--V--IQ--E--EFR------SA-----A-VGRQ-QI--EK--ELI-V-------L-----RKD--LK-VAQ---F--ER--R--PE-Y-ED---H-------K--RLM-VT--DEEE.---HK------Q-------S---TKH--RAI--VPG--DPVT 455
DEB.CM.99.CM5               --------V--TQ--E--EFR------SA-----ATVGRQ-KL-KEKY--LI-V-------L---N-DKVS-LEQVKQ--KF--EK--R--PE-Y-ED--YH-------K--HLL--E---EEE.---HK------Q-------S---TKH---AI--VPG--DKVS 455
DEN.CD.x.CD1_CMO580407      -----V------G--Q--IFK------V---T--ATAD-M-SQ---AH--VILI-------L-A--RSLE---KMVK---DF-AV--FE---A-F-ET--IQ-------KH-QI-K-D---LPEDGQ----A-------V-------L-S---TKE---SI--V-D-G-T-S 477
DRL.DE.11.D3                --------V--QG---------------------ATAD---Q---EK---VI--------F----RPKHL-EQM-K---T--KF--PE-----F-DN--YE----Q---K-----E-R--D-EI.------------V-------S---TKE----I--A-P-D-EVE 463
DRL.DE.11.D4                --------V--QG---------------------ATAD---Q---EK---VI--------F----RPKHL-DQM-K---T--KF---E-----F-DN--YE----Q---K-----E-R--D-EI.------------I-------S---TKE----I--A-P-D-EVE 463
DRL.DE.11.D5                --------V--QG---------------------ATAD---Q---EK---VI--------F----RPKHL-DQM-K---T--RF---E-----F-DN--YE----Q---K-----E-R--D-EI.------------V-------S---TKE----I--A-P-D-EVE 463
DRL.DE.11.D6                --------V--QG---------------------ATAD---Q---EK---VI--------F----RPKHL-EQM-K---T--RF---E-----F-DN--YE----Q---R-----E-R--D-EI.------------V-------S---TKE----I--A-P-D-EVE 463
DRL.x.x.FAO                 --------V--QG---------------------ATAD---Q---EK---VI--------F----RPKHL-EQM-K---T--KF---E-----F-DN--YE----Q---K-----E-R--D-EI.------------V-------S---TKE----I--A-P-D-EVE 463
GRV.ET.x.GRI_677_gri_1      --------V--TG------F-C--------T---NTAAN---EIKRHT-GLE-V------WLA--HDETR-NQQVDIV-KM--EK--E-----V-R---WE----K---N---INK-E--PLEGE.----K---V--V-----------TKHT-AM---K-N-L-E-V 472
GSN.CM.99.CN166             ----SV--V--QA-AE---FK-------A--T--NIVASL-RQI-D-E--V-LI-------LI---RKL-E--QVVHKI-NL-TSYNIQ--EA---PDY-VK-L------KG-RL--VE--DQ-T.--------------V---H---RTK----CI--A-G--DQVE 460
GSN.CM.99.CN71              ----SV--V--Q--AE---FK-------A--T--ATVASL-YQI-D-E--V--I-------LI---RKLAE--QVVHKI-NL-TSYNIQ--EA---QDY-VR-L------KG-RL--VE--DQ-I.--------------V---H---RTK---RCI--A-G---QVE 453
LST.CD.88.SIVlhoest447      -----V-----QE--R--------------C---ATVAGL-SE---L---IL--------FI---RDKKG-DQAVK---DL-VT-N-K--E--F-A---YH----V----R-EIEKTK--DI-LQKT---N-------V----A-L-E--RTKE----I--V---D--VT 467
LST.CD.88.SIVlhoest485      -----V-----QE--R--------------C---ATVAGL-SE---L---IL--------LI---RDKKG-DQAVK---EL-MT-N-E--E--F-A---YH----V----R-EIEKVK--DI-LQKT---N-------V----A-L-E--RTKE----I--V---E--VT 467
LST.CD.88.SIVlhoest524      --------L--QE--K--------------C---ETVSGL-AE---L---LL--------FI---R-KEG-SQAVK---EL-IT-N-E--E--F-DK--YH----V----R-EIERVK---I-TRKT---E-------V----A-L-D--RTKE----I--V-P-EQLVT 468
LST.KE.x.lho7               --------L--QE--K--------------C---GTVAGL-SE---L---MI--------FI---R-RKG-DQAVK---EL-MT-N-E--E--F-A---YH----V----R-EIEK-K---M-LTKT---Q-------V----A-L-D--RTKE----I--V-P-E-I-N 467
MAC.US.x.251_BK28           -----L--V--AE--K--I-K-------------YT-RHV------A---VTLV-----ILIA--RTDLE-DRVVLQ-KEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--QRET.-----------V----A-------TKH--R-I--KMT---EVQ 457
MAC.US.x.EMBL_3             -----L--V--AE--K--I-K-------------YT-RHV------A---VTLV-----ILIA--RTDLE-DRVVLQ-KEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--QRET.-----------V----A-------TKH--R-I--KMT---AVQ 457
MND-1.GA.x.MNDGB1           -----V--V--QA--K--M-----------C---GTVASL--V---NH-TVQL-------F----YTAEE-EKA-V---AL-MT-N-E--E--Y------H-----------KIEKVQ---LAEQ.P---E-----------A-L-----TK-----I--GLNI--KVT 459
MND-2.CM.98.CM16            --------V--QG---------------------ATAD---Q---ERH--V---------F----RVAPE-SRM-Q---D---F---E-----F-----E----I---K-----KVQ----EE.---------------------S---TKE--R-I--A-P-D-KVE 461
MND-2.GA.x.M14              --------V--QG---------------------ATAD---KT-KEEY-EVL--------F-----SASE-NKMVNR--E--RF---E-----F-----E----V---Q-----K-Q----ET.---------------------S---TKE----I--A-P-D--VE 462
MND-2.x.x.5440              --------V--QG---------------------ATAD---KA-KEKH-EVL--------F-----SASE-SKM-V---E--QF---E-----F-----E----V---K-----K-Q----EE.---------------------S---TKE----I--ARP-DKIVE 462
MNE.US.x.MNE027             -----L--V--AE--K--I-K-------------HT-RNV------A---VTLV-----IL-A--RTDLE-DRVVLQ-KEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--Q-ET.-----------V----A-------TKH--R-I--KMT---EVQ 457
MON.CM.99.L1_99CML1         ----S---L--AE-AK---FK---------TV--HTAAELPQEI--KY-EVTLI-------LI---NHSLPE--KIVGVI-AT---K-IQ--PE-F-DDY-VQ-L------SH-KIAK-E--DQEQ.-----E---I-------A---S---TKN---SI--ARS--DTVV 457
MON.NG.x.NG1                ----S---V--AE-AV---FR---------T---Y-ASEL-KXI-RKY-EVTLI-------LI---HSLPE--KIVTAI-TL-SEK-IQ--IE-Y-DKY-LQ-L------TH-----K-E---Q-Q.----G------------A-------TKH--XCI--X----DIVE 460
MUS-1.CM.01.CM1239          ----SL-QV--QG-AR---FK---------T---YTAA-L-QEI-EA----TLI-------LI--NR--SG--RVVAQI-NM--KL-IQ--ED-Y--DY-VQ-L------RG-KI--VT--QQEN.---------------A-A---V-TK----CI--V-S---EVQ 450
MUS-1.CM.01.SIVmus_01CM1085 ----SL-QV--LG-AR---FK---------T---FTAA-L--EIK-A---VTLI-------LI--NR-LSS--KVVAQI-NM--KL-IQ--QE-F-EDY-VQ-L------RG-KI--VA--QQEV.------V--------A-A-----TK----CI--V-S---EVQ 450
MUS-2.CM.01.CM1246          ----SV-AV--MG-AK-F-FK---------T---ATTA-L--QV--D-Q-AL-V-------LI---R---E--RLVKKI-DL-SSK-IQ--ED---PDY-VE-L------KG-RIK-VE--DQ-T.----E-------------A-V-S---TKH--RCI--I-G---LVE 455
MUS-2.CM.01.CM2500          ----SV--V--LG-AK---FK---------T---ATTA-L--QI--E-Q-AL-V-------LI---R---E--RLVKRV-DL-TSK-IQ--EE---QDY-VE-L------GG-RIR-VE--DQET.----E-------------A---S---TKH---CI--V-----TVE 454
MUS-3.GA.09.09GabOI81       ----S--AV--QM-AR---FK---------T---HTVSNL-KDI-EKYK--ILI-------LIA--RDL-D--KVV-AI-NM--KYNIQ--EH---QDY-VN-L------QG-KI--VS---QAE.----E-------------A-V----RTKH--HCI--A----DRVV 451
MUS-3.GA.11.11GabPts02      ----S---------Q-AK---FK-------A--T--NTAASL-KQI-DKYK-VT-I------ML----R-K-E--KIV--I--L-VKY-IQ--EE-Y-R-Y-VN-L------KG-KI---Q---Q-V.--------------------RTK---ACI--V----DIVQ 454
OLC.CI.97.97CI12            --------V--DR--R--I-----------C----TVAHLM--W--E--E-DL-------------PRAE-LE-VKQ--EL-RSYN-E--P--Y-ED--YE----T-Y-KR-KIAR-K----EE.-----------Q--------S---TKA----IK-P-G-LDKVE 451
RCM.CM.00.SIVagi_00CM312    -------A---QG---------------------TTAN-----V-------L--------F----RTELE-SQM-KK--E---A--FE--E-F-DK--E----V---------E-K----EE.------------------A---TKE----I--I-R-DDKVE 462
RCM.CM.02.SIVrcm_02CM8081   -------A---QG--V------------------TTANR-------E----V--------F----RTELE-NQM-KR--E---I--FE-----F-DK--E----V---------E-K----EE.------------------T---TKE----I--V-R-D-KVE 468
RCM.GA.x.SIVRCMGAB1         -------AV--QG--V------------------A-AN---Q---EE---VI--------F----RTKLE-DKM-KQ--D---F--FE-----F-DK--Y---------KS----E-K----EE.-------------------S-LRTKE----I--A---D-KVE 463
RCM.NG.x.NG411              ------LA---QG---------------------TTAN--------E----V--------F----RTAHE-SQM-KR--E---T--FE-----F-DK--E----V---------E-K-L--EE.---------------------S---TKE----I--V---DDKVE 467
SAB.SN.x.SAB1               --------V--RE-------K---------T---TTAN---QE--QK---VD-------MLIA--RPKAE-LVMVQQ--DY-ET--FK--E--F--D--Y-------Y-K--QL-E-T---REE.-------------------------T---TKH--R-I--ARP---IVQ 481
SMM.CI.79.SIVsmCI2          ----AL--V--AE--K--I-K-------------DT-R-V------A-E-AT-I-----ILIA--RNDLE-DQLVAQ-KEL-NGL-FS---D-F--D--HK------W-K--KL-K-K---R-R.-----------V---VA-L-----TKH----IK-KMT---EVQ 466
SMM.LR.89.SIVsmLIB1         -----L--V--AE--K--V-K-------------FT-K-V------T---VTLI--I--ILIA--RTDLE-DKVVLQ-KEL-NGL-FS--EE-F--D--YQ--R-K-W-T-*KL-K-E--Q-ER.-----------V----AR-----TKH----I--K-T---EVQ 456
SMM.SL.92.SIVsmSL92A        -----L--V-SAE--K--I-K-------------HT-RQV------A-S-VTLV-----ILIA--RSDLE-DKVVTQ-KEL-NDL-FS--EE-F--D---Q-------Y-K--KL-K-T----EV.-----------V----A-------TKH----I--KVT---EVQ 457
SMM.SL.92.SL92B             --------V--QE--K--I-K-------------AT-RQV------A---VLLV-------LI--NRGLTE-DKMVTQ--DM-NNL-FS--ED-F--N--LQ-----L-Y-K--KL-K-E-----ER.-----------V----A-------TKN---MI--KMT---EVQ 454
STM.US.89.STM_37_16         -----L--V--AE--K--I-K-------------YT-RN------RA---VTLI-----ILIA--RTDLE-DRVVLQ-KEL-NNL-FS--EE-F--D---Q------W-T--KL-K-E--QR-V.-----------V----A-------TKH----I--KM----EVQ 457
SUN.GA.98.L14               --------V--QA--K--V-----------C---GTVASL--K--R-Y-EVQL-------LI---Y-KKK-EEIVKQ---L-ME-N-E--E--Y-G---YK----I----R-EIEK-K---PLEEE.P-------I--VI-----L-E-LRTKE----I--N-P-A-KVK 468
SYK.KE.x.KE51               -----V-----SS--E---FK---------T---ATVNQ--Q-I-EK-S--I-V-------L---NRSEKE-GQIVS-IVKS--AV-FSI-PE-W-DKF-LQ-L--T------SL-KVQ---ITES.P---EL--II-V-------------CKA-S-CI---A----DIVE 488
SYK.KE.x.SYK173_COMGNM      -----V-----AS--E---FT---------T----TINQ--Q----KYS-LTLI-------LI-T-RSEKA-QEIVQQIVTA--KV-FKV-KE-W-DQY-MQ-L--T------QL-K-E--NI-DE.I---QL---I-V-------S---TKE---CI----Q----LT 482
TAL.CM.00.266               --------L--DR--R---FK---------T---HTSDQL-RK--DKY-QVALF-------L---NESLEE--RIVG---AL-NKV-IQ--EA-Y-DK--IK-L----Y----K-----E--D-LT.----EL---I-------C---S---TKA--RY---V-G-L-EVQ 458
TAL.CM.01.8023              --------L--DR--R---FK---------T---YTSDQL-RK--EKY-QVALF-------LI---ESLE---RIVG---AL-NKV-IQ--EA-F-DK--IK-L----Y-E--K--N-E----QT.----L-------------S---TKA---C----V-G-L-EVT 459
TAN.UG.x.SIVagmTAN1         -------QV--VG------F-C--------T---HTAQRV-KEIG-KVQ-LE-I------MWI----PEKE-DE-V--V-KL--LR-FE-----V--R--YD----K-Y-SH-SLNK-EI--QEH.-----------V---------L-----TKA-------AQD--A-VQ 469
VER.DE.x.AGM3               --------V--QG--T---F-C--------T---NTAAS----EIK-ELKPLT-V------W---QEDEYT-DRLV-Q--MK-SA---E-----V--K--YE----K-W-H--QISS-E-ED-EE.-------R--------A-L---LRTKN----I--K-N-L-TVT 466
VER.KE.x.9063               --------V--QG--V---F-C--------T---NTAA----KIKEELKPLT-V------W---QEDEYT-DKLV-Q--VK-KT---E-----V--K--YE----K-W-Q--QLNS-N-E--EQ.-----------------A-L---LRTK-I-R-I--K-N-L-EVQ 466
VER.KE.x.AGM155             -------TV---G------F-C--------T---NTAS----EIK-ELKQLT-V------W---QE-GPK-DQLVQT--NR-QE---E--E--V-R---E----K-W-H--KL-S-E-EK-EQ.----L------------A-L---LRTKNI------K-N-LD-VE 469
VER.KE.x.TYO1_patent        -------TV--QG------F-C--------T---NTAAS---EIKRNL-ALT-V------W---QENEHT-DKLV-Q--TK-QA---E--E--V-----YE----K-W-H--ELSR-Q-E---E.------------------A-L---LRTKNI---I--K-N-L-LVT 469
WRC.CI.97.97CI14            -----V--V---S--K--V-----------C----TVASL--I--GK--N-QVL-------I----DKTE-LN-V----KL--T-N-E--E--L-L---CN----I------QI-E-K---EVK..----V-----------A-L-E--QTK---R-I--N-P---EVE 472
WRC.CI.98.98CI04            -----V--V---S--K--V-----------C---ATVASL-DV-K----TMQ-L-------I---F-KPR-LE-V-----L--T-N-E--E--L-L---CK----I----R-QI-E-K---EVR..----V-----------A---E--QT------I--N-P---EVT 471
WRC.GM.05.Pbt_05GM_X02      -----V------K--T--V-----------C---ATVASL--I---KH-KVQ-V----------Y-KKE-LQM-Q---DL--T-N-E--E--L-L---CK----I------EI-E-K---EIR..----V---I-------A---E---TKE----I--N-P--AEVE 463
```

```
                                                                                                                                                                           p51 RT end p15 RNase H start
H1B.FR.83.HXB2             LTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEP.FKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTP.KFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTNRG  617
H1A1.UG.85.U455_U455A      -----------------D-------------V--------D----------.-----------K-S-----------V--VS--------I-.--R----------A--M------------------------------D--A------------------------D--  616
H1C.ET.86.ETH2220          ---------------------F---------------ND---F-F----.---------F-KRGT---------AV----AL---------.--R----------A---D--------------------------------A-V-------------I-------D--  624
H1D.CD.84.84ZR085          ------------------M-------------L-----------------.-------------------------AI----------.--R----------ID---------------------------------I-------------K----------D--  617
H1F1.BE.93.VI850           --A-------K-----R--------------------D---------N-.----------KV-S----------------AL--------RSP------L----D---D-------------------------T---A--D-------S-----K-------DK-  618
H1G.SE.93.SE6165_G6165     --A---M---------R------------E----V----LD----------.Y-----------KRGS-----------V----A----------.--------R----I---D------------------R--T---P----Y-----------------DK-  614
H1H.CF.90.056              --K-------------R--I---------------R-----------------.----------K--T-----I----------S---------I-.--R--------------------------------H-----------T---A----Y-I-------------D--  616
H1J.SE.93.SE9280_7887      --R---------K-------------SA-E----V----LD-----------.-----------KR-S-----------A-V----AL-A---------.--R----R------------D--------------------------M------------S-----T-------DK-  615
H1K.CM.96.96CM_MP535       --A--------------------------------ND-----------.H------------S-----------------A--G--------.--R------------------------------------------T-----------------H-----K-R-----D--  617
H1O.BE.87.ANT70            -SR------E----R--Q-------Q-D---WVN-----GE---------E.H---------T-QKAS----IR--A-VI--VSQ---I----L-.-----VTR-------AD---------------S----I----R--S----M-----Y---------------------EQ-  613
H1O.CM.94.BCF06            -SK------E----R----------Q-D---WVN---H--------V--DE.H----------QKAS---EIR-MA-V---VSQ-A------L-.-----VTR-------A----------------S----I-------T---------Y----------N----------EQ-  613
H1O.FR.92.VAU              -SK------E----K----------Q-D---WVN-----E---------DE.H----------QKAS----IR--ANVI--VSQ-A-I----L-.-----VTR-V------AD---------------S----I----K--N---I--------------D----------I-EQ-  614
H1N.CM.02.DJ00131          F-K--------------R--L-------G-E----------D----------.Y-----------KR-S------I---V-V---VA--G------I-.------V---V--A---DN----------------------T---T------Y-------K----------F--D--  622
H1N.CM.04.04CM_1131_03     F----G-----------------L-------E-E-----------S-----------S.H------------K--S------I-----V-----A-----------.------V---V--A---N------------------P--------T---S-----Y-------------F--D--  624
H1N.CM.95.YBF30            F-----------------------L-------G-E-V------------------L.H------------K--S-----I---V-V-R-VA-----------.--R---V---V--A---DH------------------------T---S-------------------------F--D--  624
H1P.CM.06.U14788           F-R------E--K-------------Q-E-E--VD-----A-------VF--E.H----------QKAT----IR--A-VI--VSQ-----------L-.--R--VNRNV----SD------------S----I-----------D--P-T---------------------D--  614
H1P.FR.06.RBF168           --R------E--K-------------Q-E-E--VD-----E----------E.H----------QKAT----IR--A-VI--TSQ-----------L-.--R--VNRNV--N--SD------------S----I-----------D--P-I---------------------D--  613
CPZ.CD.06.BF1167           --A------E--KQ----T-Q-T----KEP-K-AV-----K-------V--KE.NQI--------NH-T----EL----AGV-----G-----I---RV-.V-Q--VT--V-QQ---SD---VS-------------I-----N-LT--MPE----Y---------DS---------D--  620
CPZ.CD.90.ANT              M-R------E--KQ--QQKIE-Y--Q-GLP-K-T-----S----------NE.G-L--A------PT-T---E-R--AGV-----GL--------EV-.--Q---TR---DA---SD-------------------IR---N-LAD--PE-----------NSQ---------D--  617
CPZ.CM.01.SIVcpzCAM13      --P-------------VST--------N-D-E-----V-----E--------F--Q.HR-----------Q-ST----IR--A-VI---AV------------.R-R---V----S-----A-----------------I--------------S--TD---PT-D-Y----------K------I-DK-  616
CPZ.CM.05.LB715            --K-------------AN--------------E-----V------S--------Q.--------------KT-S--------------V------AL--------I-.--R-------------A----N---------------Y---------------T----T-----------------D----------D--  622
CPZ.GA.88.SIVcpzGAB2       M-R------E------------------T------E-E----V----NN---------F---T.GR----------Q-A-P----IR--A-V-P-VSP----IL--TV-.------V-R-V---A----D-------------R----------------------Q-----Y------------RR-------I-N-  620
CPZ.US.85.US_Marilyn       F-H---M--E----------------------E-E-V---V----RS---------F---R.H-----------Q-S-----IR--V-V-----A-----------V-.-----LV---V------S-------------D--------------K--T-A-E----------------Q-----K--------D--  619
GOR.CM.04.SIVgorCP684con   -SK------E----------------Q-E-E--VNV-----EE---------F-DE.H----------Q-AT----IR--A-VI--VSQ-----------L-.--R----CRS---DAF---D-----S--------IS-----I-----------------T----------------S-Q--------D--  609
GOR.CM.07.SIVgor2139_287   -SK------E----------------Q-E-E--VNV-----E----------F-DE.Y----------QKAT----IR--A-VI--VSQ-----------L-.--R----CRS---DAF---D-----S--------IS-----I-----------------P----------------S-Q--------DK-  609
GOR.CM.07.SIVgorCP2135con  -SK------E----------------Q-E-E--VNV-----E----------F-DE.Y----------QKAT----IR--A-VI--VSQ-----------L-.--R----CRS---DAF---D-----S--------------S-----I-----------------P----------------S-Q--------D--  609
GOR.CM.12.SIVgor_BPID1     --R------E--K-------------Q-E----VD------E----------F---E.H----------QKAT----IR--A-VI--VSQ-----------L-.--R--VNRTL---A--SD---------------S-----I-----------D--P-T----------------------D--  614
GOR.CM.12.SIVgor_BQID2     -SK------E----R----------Q-D---WVN------E---------V--DE.H----------QKVS----IR--A-VI--VSQ-----------L-.------VTR-----A--AD---------------S-----I----------S---S-----Y-----------D----------EQ-  613
GOR.CM.13.SIVgor_BPID15    F-R------E--K-------R------Q-E-E--VD------E----------F---E.H----------QKAT----IR--A-VI--VSQ-----------L-.--R--VNRTL---A--SD-----------------Y-S-----I-----------D--PET------------------------D--  612
                                                                                                                                                                          p51 RT end p15 RNase H start
MAC.US.x.239               W--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-SQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QV-.--H--VE-DV--Q---D---V------D-IS-----R-VFN-V-D--E-E--Y-T--SC-KQS-E-----I-D--  628
H2A.DE.x.BEN               W--L--A--E--KI--SQEQE-Y--QEE-E-E-T---SQGH----K-H--E..-I--V----KIKNT---G-RL-AQV----GK-AL----RI-.--H--VER----Q--DN---V------D--S-----R-TFN-VGD--P-------T--SC--QS-E-------D--  641
H2B.CI.x.EHO               W--L--A-FQ--KI--EQEQE-S--KEGVP-E-TV--NLAN----K-H-GD..-I--V----KVKNT---G-RL-AHV----GK-AL----EI-.M-H--VER----DQ---D---V------D--S----IR-A-N-V-D-LE-V--Y-T--SC-KAS-E-------D--  641
H2G.CI.92.Abt96            W--L--A--Q--KI--XQEQE-A--KEXEP-E-TV--NLDN----KVH-GN..RI--V----KIKNT---G-RL-AHV----GK-AL-----L-.F-H--VERD---Q---D---V----D-D--S--H--R-V-N-V---LEQT--Y-T--SC-KTS-E-------D--  624
H2U.CI.07.07IC_TNP3        W--L--A--Q--KI--EQEQE-T--REGEP-E-TV--NLDN----K-H-GN..-I--V--F-KIKNT---G-RL-AHVI---GK-A------V-.I-H--VER----Q---D---V------D--S-----R-A-N-V---LERV--Y-T--SC--NS-E-------D--  624
H2U.FR.96.12034            W--M--A-F---KI--SQEQE-A--REEN-E-TVL-NQDN----K-H-GD..RT--V--F-KVKNT---G-RL-AHV----GK-A------I-.M-H--VER-V--Q---D---V--VLD-D--S-----R-A-N-V---LL.E--Y-T--SC--AS-Q-------D--  637
ASC.UG.10.RT03             V-----T------QA--DQEAK-G---AE-P-EVD-IQL-GT--G-TVR-DK..EV-----F-KQ-S--S-PF---VD-IM--GK---L-Y--RI-.Q-NV-VN--Q-DA---N-H--VS---DIKP-H----IQM-----VQD--SE---W-I--------S---------D--  626
ASC.UG.10.RT08             V-----T--V--QA--DQEAK-G---AE-P-EVD-TQL-ET--G-TVR-DK..EV-----F-KQ-S--S-PF---VD-IM--GK---L-Y--RI-.Q-NV-VN--Q-DA---S-H--VS---DIKP-H----IQM-----VQD--P----W-I--------S---------D--  629
ASC.UG.10.RT11             V-----T------QA--DQEAK-G---AE-P-EVD-IQL-GT--G-TVR-DK..EV-----F-KQ-S--S-PF---AD-IM--GK---L-Y--RI-.Q-NV-VN--Q-DA---S-H--VS---DIKP-H----IQM-----VQ---PE---W-I--------S---------D--  629
COL.CM.x.CGU1              -SA---E--QD-KQ-I--E-Q-----I-HE-IWVD-FRL-K---G-AVC-KH..GT-RR--HNTGKNQ-F-SMQE-ASVI---GR-A--T---V-.-M-V-AKR-D--Q--SD---SA-----I---SSSYV---VWN-VL--LKE-P-YWT--GCS-KLGA----WINSK-  597
COL.UG.10.BWC01            -SA---E--E--KK-V--EQK----R-EE-IWV--MRL-K---G-AVQ--H..GT--R--INTGKS--F-SMQE-ATVI---GR-A------V-.RM-V-AKR-D--Q--SD---S-----I---S--YV---VWN-ML--VRS-P-YWT--GYS-KDNW----WINSS-  599
COL.UG.10.BWC07            -SGQ--E--EQ-------EQE----K-EEEIWVT-FRL-K---G-AVE-VH..GT--R--VNTGKS--F-SMQE-AMVI---GR-AL-----V-.LM-I-AKR-D--Q--SD---S-----I---S-S-V---VWN-VLS-LEKEV-YWT--GYS-KEGI----W-NSK-  599
DEB.CM.04.SIVdeb04CMPF3061 -----QA--E-A-Q----Q-S-T--KEQE----D-T-LSE---G-T-R-SK..GI--------KAK---F--FH-IAKLMM-VG-----T--RL-.T-R--V--QD-DG--H-H----------A-H-----R---S-VS---KE-D-Y-----------S-E------ISE--  625
DEB.CM.99.CM40             -S---QA--R--K----QE-S----KEQEP----LS-L-K---G-V-R-PK..GI------FS-DK---Y--FH--AK-MY--G-----F--RI-.Q-R--VV--E-DN--HNH---A---D--AIH--H------E-VS---PD-D-Y-----------S---------EW-  622
DEB.CM.99.CM5              -S---QA--R--K-TP-QE-S----KEDEP--S--S-I-R---G-V-R-TK..GL-----FSKDK-T-Y--FH--AK-MY--G-----Y---V-.--R--VV--E-DN--HNH-------D--A-H--Q--R---E-VS---PD-D-Y-----------S-M------ISEW-  622
DEN.CD.x.CD1_CMO580407     ---G--A--K-AQ-----SIS-S--R-E-P------S-L-EN--G-T-K--K..HL------HKDK-S-Y-PYQ--AKVMA--GR-AL-----RL-.T-R--VS-VD-DA--S-----VN---D--M-S-----LRNF-N-VS---DK-P-Y-T------Q--Q-R----ATN-  644
DRL.DE.11.D3               W-R-----YE--KL----QM-----Q-E-P-K-K---L-N---S---E-DD.N-P-------KVKNS----MRM-AGL----AK-AL-----RL-.T-Y--VER-V--Q--Q-----V-----D----S--H-I-----N-LSD--P-E-VY-----------VS-E-------S--  631
DRL.DE.11.D4               WSR-----YE--KL----QM-----Q-E-P-K-K---L-N---S---E-DD.N-P-------KVKNS----MRM-AGL---VAK-AL-----RL-.I-Y--VER-V--Q--Q-----V-----D----S-----I-----N-LSD--P-E-VY-----------VS-E-------S--  631
DRL.DE.11.D5               W-R-----YE--KLT---QM-----Q-E-P-K-K---L-N---S---E-DD.N-P-------KVKNS----MRM-AGL---VAK-AL-----RL-.I-Y--VERDV--Q--Q-----V-----D----S-----I-----N-LSD--P-E-VY-----------VS-E-------S--  631
DRL.DE.11.D6               W-R-----YE--KL---XQM-----Q-E-P-K-K---L-N---S---E-DD.N-P-------KVKNS----MRM-AGL----AK-AL-----RL-.T-Y--VER-V--Q--Q-----V-----D----S--H-I-----N-LSD--P-E-VY-----------VS-E-------S--  631
DRL.x.x.FAO                W-R-----YE--KL-I--QM-----Q-E-P-K-K---L-N---S---E-DD.N-P-------KVKNS----MRM-AGL----AK-AL-----RL-.I-Y--VER-V--Q--Q-----V-----D---IS--H-I-----N-LSD--P-E-VY-----------VS-E-------S--  631
GRV.ET.x.GRI_677_gri_1     W-----A-YKN-QG-VQ-TQE-T-----L-E---TV-----E-------FT--G..AV--V-R--KQ-ET-----LRT-AHL-----CK-ALT----RL-.RVQ--VD-K--DM--QD---VS---------S--L-------S-V-----K-EDVY-------SKV----------LSE--  639
GSN.CM.99.CN166            -----QI-----------QSSE-G---AE-P-VV--TSL-EQ--G-MFS--G..-M--SA-F-KT-N---I-SYQ-FAD-LT-VGR---Q-T----V-D--RI-VV--Q-DA---MN--------TI-AIH--H-LRQ--T-VS---LED-V-Y-----------TS---------T-  628
GSN.CM.99.CN71             -----QI-----------QSSE-S---AE-P-VV--TSL-EQ--G-MFS-DS..-M-RS--F-KT-T---I-SYQ--AD-MT-VGR---Q-T----V-D--RI-VV--Q-D----MN--------TI-A-H--H-LRQ--T-VS---LEN-V-Y-----------TS---------T-  621
LST.CD.88.SIVlhoest447     W-DA-LE-YEQ-QQ----KLQ-A----T-E-VVRV-QSKK-II-F-WR-GN..SI-RA-R-Q-QKS-----PLQK-A-------GK--------HV-.-IQV-VTR-V-DQ--SDH--V-------L--IS--Q-EQE--S---A----E-VD-Y--------EKVG-T--------QS-  634
LST.CD.88.SIVlhoest485     W-DA-LE-YEQ-QQ----KLQ-A----T-E-VVRV-QNKK-VI-F-WR-GN..SI-RA-R-Q-QKA-----PLQK-A-------GK--------HV-.-IQV-VTR-V-DQ--SDH--V-------L--IS--Q-EQE--S---A----E-VD-Y--------EKVG-T--------QS-  634
LST.CD.88.SIVlhoest524     W-----LE-YET-KT-----K-Q-A------E-VVRV-QNKK-VI-F-WR-GN..AI-RA-R-Q-QKS--M-PLQ--A-------GK--------FV-.-IQV-VTR-----N--INF---V-----DL--IS--Q-EQE--S---A----E-TD-Y-------E-VG-T------I-QS-  635
LST.KE.x.lho7              W-----LE-YGQ-K-V---KMQ-A-----E-E--VRV-QNKK-II-F-WR-GN..NI-RA-R-Q-QKA-----PLQK-V--I---GK--------FV-.-IQV-VTR-V--H--SDH--V-----DL---IS--Q-EQE--IW-A----I-VD-Y--------EKVG-T------I-QS-  634
MAC.US.x.251_BK28          W--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-SQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QV-.--H--VERDV--Q---D---V------D-IS-----R-VFN-V-D--E-E--Y-T--SC-KQS-E-----I-D--  624
MAC.US.x.EMBL_3            W--M--A-YE--NI--SQEQE-C--QEG-P-E-TVI-SQDN----K-H--D..-I--VR-F-KIKNT---G-RL-AHVI---GK-A---V-QV-.--H--VERDV--Q---D---V------D-IS-----R-VFN-V-D--EVE--Y-T--SC-KQS-E-----I-D--  624
MND-1.GA.x.MNDGB1          M-----R--YEQ-K---A-EQE-S-----N-E-YVRF--TTG-DISF-WK-GN..-V-RA----GKQKT---S---LMK-AG-T--VGR----------FV-.-MQI-TTR-I--D---H-----C------V--IS---M-ERE---S-SP--LE-V---Y--------DS-M------I-D--  626
MND-2.CM.98.CM16           W-R-----YE--KL-VQ-E-------Q-E-P-M-KV--LT----S---E--D.N-P--V-----TKN-----ELRV-AGL----AK-AL-----L-.--Y----ER-V-DQ--P-----------------------S--H-IG---N-LR--VP-EDVY-----------NS-E---------A-N  629
MND-2.GA.x.M14             W-R-----YE--KL-VQ-E-------Q-G-P-M-KV--LT----S---E--E.N-P--V-----TKH-----ELRV-AGL----AK-C------EL-.--Y----LER-V-DQ--HD--------------IS----IR---N-L-----P-EDVY-----------NS-E-------Y-A--  630
MND-2.x.x.5440             W-R-----YE--KL-VQ-E-------Q-E-P-M-KV--LT----S---E--D.N-P--V-----TKN-----ELRV-AGL----VAK-CL------L-.--Y----LER-V-DQ--HD---V--------D-IS-----IR---N-L-D--P-EDVY-----------NS-E-------Y-S--  630
MNE.US.x.MNE027            W--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-NQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QV-.--H--VE-DV--Q---D---V------D-IS-----R-VFN-V-D--E-E--Y-----SC-KQS-E-----I-D--  624
MON.CM.99.L1_99CML1        -SDL-QA-----------EST-G----NEP-KV-LTSLAEN--G-RFF--K..SV-----F-KI-ST-S-TYQ--AD-LA-LGK-A--T--RL-.I-R--VV--Q-DA---ADN---IN-V-DI-A-YS-H-LRQ--T-VQD--E--P-Y-----------TS-E---------F-  624
MON.NG.x.NG1               -SDL-QA----------SQQST-G----QL--KV--TSLXE---G-XYF--X..RV-----F-KIKN--S-TYQ--AD--T-LGK-AL-----L-.I-R--VN--Q-D----ADN---V-------I-P-HS-H-LRQ--T-VQX--EH-P-Y-----------AS-E---------ST-  627
MUS-1.CM.01.CM1239         -----Q------Q-----QA-Q-A-F---EQP-V---VSL-DS--G-NFL-NK..GI--S--F-KV-SV----SYQ--SD-IAR-GR--L--------P-E-VRI-VI--Q-DQ----H--VS---DI-A-H-TH-LRQ-FT-VP--LSE-P-Y-----------HKVS-----------T-  618
MUS-1.CM.01.SIVmus_01CM1085 -----QA-----Q---QQT-Q-A----EQP-V---VSL-DS--G-NFT-SK..GV-R---F-KV-SV----SYQ--AD-MA--GR--L--------P-D-VRI-VI--Q-DQ---IAH--VS---TI-A-H-TH-L-Q-FT-VP--LKE-P-Y-----------VS------------T-  618
MUS-2.CM.01.CM1246         -S-Q-Q------------QEAG-A-----E-P-VL---VSL-EQ--G-TFT-DR..NM-R---F-KI-T---S-PYQ--A--LSRASK-AL-C-----D-CRI-VV--Q-DN---ADS--T-----DI-A-H--Y-LRQ-FT-VP---ET-P-Y-----------NS---R------D--  623
MUS-2.CM.01.CM2500         ---Q-Q---------M-QE-K-A------E-P--L--VSL-EQ--G-TFT--K..NM-----F-KV-T--I-TYQ--A--LA-AAR--L-C-----D-CRI-VV--Q-DH--A-S--T--V---I-A-H--H-LRQ-FT-VP--LEA-P-Y-----------NS---R------D--  622
MUS-3.GA.09.09GabOI81      -S-L-----M-------M-QE-Q-D----E---VV--TSL-G---G-MFS-NS..-P-----F-KV-SV---PYQ---D-MT-LGR-A--Y---V-D--RI-VI--Q--K--ADH--V-----DV-A-HS-H-LRQY-T-VQ---EK-P-Y-----------TS-A-R-------T-  619
MUS-3.GA.11.11GabPts02     ---A--A---------I-QE-Q-R-----G-E-TV--TSL-EQ--G-TFS--N..-H-----F-KV-TT-S-SYQ--AD--TRLGK-AL-Y---I-DV-RI-VI--Q-DN---A-----V--V-NV-AIH--H-LRQY-T-VQ---E--P-Y-T---------NS-------------T-  622
OLC.CI.97.97CI12           W----KE--E---------EQE--M-KEEEP-QCT.LCYTK-TVA-R-W-KH..GT----RFQ-KMT------KFR-M-G-IE---GK---------L-.I-QV-TTRKD---A-AFDD---VH---DI---SNSENQRTFWE-VPA-M.EV-----------Y---KS-E----------SSG-  616
RCM.CM.00.SIVagi_00CM312   F-K-----YE--KLL---KL--------E-P---KV--LEG---S--VE-GD.G-P--------KQKT-----EIRM-AGL----CK-A-----L-.--E---VER-V--Q--AD---VS----------S-----IR---N-V-D--P-E-VY-----------NS----------I-D--  630
RCM.CM.02.SIVrcm_02CM8081  F-R-----YE--KLL---QM--------E-P---KV--LTG---S--VE-GD.G-P--------KQKT-----EIRM-AGL----CK-AV----L-.--E----ER-V--Q--AD---VS----------S-----IR---N-V-D--P-E-VY-----------NS-Q-------I-D--  636
RCM.GA.x.SIVRCMGAB1        M-K---I-YE--KM-----KL-------EK-P-V-N---LEG---S---E--S.G-P--------KQKT-----EIRM-AGL----AK-A-----RL-.T-R----ER-V-D*-RSQ---V-----D----S-----IR-G-N-V-D--P-E-VY-----------NS-I--------------  630
RCM.NG.x.NG411             F-R-----YE--KAV---KI-----N-E-P---KV--LEG---S---E-GD.S-P--------KQKT-----EIRM-AGL----CK-A-----L-.--E----ER-V--Q--SD---VS----------S-----IR---N-V-D--P-E-VY-----------NS-----------D--  635
SAB.SN.x.SAB1              W--------E---Q--RQKQQ-Q----ALP-R-KVL-L-D---G------PE.N-I---V-----KIKT-----ELRM-AGL----GK---------QI-.IME---VER-L--Q--SD---V---------M-S--Q-IR---K-V-D--P-EAVY-----------NS-E--------L-D--  649
SMM.CI.79.SIVsmCI2         W--L--A------I---EQEQE-AF-REEEN-E-TV--NQDN----K-H-GS..RV--V--F-KIKNT---G-RL-AHV----GK-AL-----I-.F-H--VER-V--Q---D---V-----D-D--S-----R-VFN-V-D--P-E-----T--SC--VS-E-------D--  633
SMM.LR.89.SIVsmLIB1        W--M--A-YE-SKI--SQEQE-C--QED-P-E-TVV-NQDS--S-K-H--D..-I-RV--F-KVKNT---G-RL-AHV----GK-AL--*-QI-.--H--VER-I*-Q---D---V------*D-IS-----R-VFNPV---LE-V-------RSC----S-I-------I----  620
SMM.SL.92.SIVsmSL92A       W--L--A-------KI--SQEQE-S--QEGEP-E-TAI-NQEN--S-K-H-GE..-I--V-----KIKNT---G-RL-AQVI---GK-A------V-.--H--VER-L--Q---D---V--V-D-D--S-----IRMAFN-V-D--P-E--Y-----SC-KQS-A---------D--  624
SMM.SL.92.SL92B            W--L--A-------I--NQEQE-R--REDEP-E-TVL-NQDN--S-K-H-GD..RI--V--F-KIKNT---GIRL-ANV----GK--L--------.F-H--VER-V-DQ---D------------D-IS-----IR-VFN-V-D--EKE-VY-I--SC--NS-E-------D--  621
STM.US.89.STM_37_16        W--M--A-Y---KI--SQEQE-Q--RED-P-E-TVV-DQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QI-.--H--VER-V--Q--AD---V------D--S-----R-VFN-V---LE-T--------SC--QS-E-----I-D--  624
SUN.GA.98.L14              M----RE-YQS-Q-V-Q-S-S-S--E-D-E--CRV--VK--IL-F-WL-GK..QV-RV-R-QKRGA--E-PCQ---AA-L---GR---------FV-.-IQV---R-I-SQ--AD---C-----L---S--K-EQE---T-AT---VP-.D-Y---------EKLE-R------I-QG-  634
SYK.KE.x.KE51              ------A-M--------VEQT-S--Q-E-P-E-H-S-L--Q--G-I-K-G-QE-P-I---TGKAYAT-H--YQA-AQLMN--GIQALW-----I-.E-H--VKR-E--K---DH------L--VKC-H--I--RWY-N-VS--V.E----------------NS-E-Q----I-D--  656
SYK.KE.x.SYK173_COMGNM     ---A--A--E---Q----EQA-S----K-P-E-H-T-L-SQ--G-M-K--QKGPP-I---T-KTFA--S--YQS-AQLLN--GIQ-LWY---V-.T-H--VKR-E--K---D-------V--VK-IS------RWY-N-VP---PE-V---------------DS-T-N-----ASD-  651
TAL.CM.00.266              -S----A---------A-ET--T--QEGEP-E---LT-LAE---G-M-K-GK..RL------F-KQ-T--S-PYQ---IG-M---GK-----F--QV-.V-RI-VV--E--Q-----H---CS----TIVPIH------R----N-VQ----E--D-Y---------H--S-E-----------AT-  625
TAL.CM.01.8023             ------A---------A--A-ET--S--KEGEP-E---LT-LAE---G-MVK-GK..RL------F-KQ-T--S-PYQ---VG-M---GK-A--F--QV-.I-RV-VV--E--Q--S----CS-L-TIVPIH------R----N-VS----E----Y---------H--S-E-----------TS-  626
TAN.UG.x.SIVagmTAN1        W--Q--V-FRQ-Q-----EQQ-A------S-R-T-T-LDD---G--FK--G..-V-------TKSKNT----EFRV-AGM--RVCK-A-TF---L-.IMEV-VER-V--Q--SD---V--L--I--IS-----I-----N-V-D--PK-KVY-C----C--NS-------FL--T-  636
VER.DE.x.AGM3              W-----A-Y---K----TEQE-T--K-GRPIR-AV---LEG---S--FK--G..QV--V-----TKQKNT----EFRV-AGL----LCK--L-----EL-.VLE----ER-V--Q--AD---VS---D-----S-----------T-T-----PKEDVY-------C---NSRE------I-QY-  633
VER.KE.x.9063              W-----A-Y---K----TEQE-T--A-G-P-R-AV---LED---S--FK--G..-I--V-----KQKNT----ELRT-AGL----CK-AL-----EL-.I-E---VER-V--Q--AD---VS----------SV-----------T-T-----P-EDVY-------C---NS-E-------I-QY-  633
VER.KE.x.AGM155            W-P---A-YE--K----TEQE-T--A-E-P-R-AV---L-D---S--FK--G..-I--V--F-KQKAT----ELRV-AGV----GK-AL-----QL-.T-E---VERD----Q--AD---VS-----D--SV-----T----T-T-----P-EDVY-------C---QS-E-------I-QQ-  636
VER.KE.x.TYO1_patent       W-P---A-Y---A----TEQE-T--K-GIPIR-AV---LEG---S--FK--G..QV--V-----TKQKNT----ELRT-AGL----CK-AL-----IL-.VLE----ER-V--Q--AD---VS-----D--S-----L-----T-T-----PKEDVY-------C---NS-E------ISQY-  636
WRC.CI.97.97CI14           W----LE--EI-KQ----EQS-A--KGKDLIAEL-KTGPGQWAYTIREIGE.K-P---V---G-Q-HT-S-ELRSMAH-M---G---------RV-.--RI--E-DI-DS--SD---VS---DT-H-S--F-ARE---K-VL--LPEV---Y-I--S---DS-E------.WTKS  639
WRC.CI.98.98CI04           W----LE--EI-KS----EQM-A--TGKELVAEL-KTGSGQWAYSIREAGE.K-P---V---G-Q-HT---ELRS-AH-M---G----------V-.----I--E-DI-DS--SDH--VS---DT-H-S--Y-ARE--KIVL-----EV---Y-I--S---DS-E-----F.WTNS  638
WRC.GM.05.Pbt_05GM_X02     W-M---A--EI--A----SQQ-S--GGGDLIAEV-KTGTGQWAYTIKEAGE.T-P---V---FT-I-HT---ELRT-AHCI---G--A------L-.--RI-VE-DI-DM--S----------V-H-S--F-KRE--T-VS---E-V-EM-I--S--K-S-E------.WTKT  630
```

```
                                                                                                                    p66 RT, p15 Rnase H end p31 Integrase start
H1B.FR.83.HXB2           RQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIR..KVLFLDGIDKAQDEHEK.YHSNWRAMASDFNLPPVVAKEIVASCDKCQLK.GEAMHGQVDCSPGIWQLDCTHL  783
H1A1.UG.85.U455_U455A    -----S--E--------H--H-------S-------------------R----I------K--E------S-------------------S---..-----------ED---.--C-----------------------------N-----.--------------------  782
H1C.ET.86.ETH2220        ---I-S--E-----------Q-------S-----------L----K----I---------S--R---S-------------------S---..-----------E----.---------NE--I----P-----C-------.---I----N-------------  790
H1D.CD.84.84ZR085        -----PF-------------N---------------------------K------S-------------------------------S---..-----------E----.------------------------------------.--------------------  783
H1F1.BE.93.VI850         K----S--E-----A-------------S-------------------K----I---------Q--R---S--------------------V-..-I---------E----.--N-------------I-----------------.--------------------  784
H1G.SE.93.SE6165_G6165   K--II---E-----A-----Q------RS-------------------R--A------------------S-------------------S---..-----------E---R.--N-------------I-----------------.--------------------  780
H1H.CF.90.056            K----S--E---------------------------------------K-----------E---------S-------------------S-V-..-----------E---R.--N----V--------I-----------------.--------------------  782
H1J.SE.93.SE9280_7887    -----------------H------R-----------------------K-----------E---------S-------------------S---..-----------ED---.------------------------------------.--------------------  781
H1K.CM.96.96CM_MP535     -----SI-E-----A-----C-------S-------------------K---D-------------RI--S-------------------S---..-----------E----.--N-------------I-----------------.---I----------------  783
H1O.BE.87.ANT70          K--IIK-DE-----A--M--L------KET------------V-SS--T----PI-Q-----E-T---Q---T-----------KI-----KD--..R----E---Q--ED---.------K-L--E-G---------I---P--HI-.---I---------EV--I----M  779
H1O.CM.94.BCF06          K---IK-EE-----A--M-VLI-----KEA---------V----S---T--D-PI-Q-----E-T---Q---T-----------KI-----QD--..R----E---Q--ED---.--------L----G---------I---P--HI-.----------Y--E---M-----  779
H1O.FR.92.VAU            K--IIK-DE-----A--M-VL------KEK---------V----SS--T--D-PI-Q-----E-T-R-Q---T-----------KI-----KD--..R----E---Q--ED---.--------L----G---------I---P--HIR.R-------------E---M-----  780
H1N.CM.02.DJO0131        K----SIE------------L-------Q--------------HSH--K------S-----E-----R---S-------------------S---..-------E---ED--R.----------------S-----------------.---I----N----V--------  788
H1N.CM.04.04CM_1131_03   -----SIK---------H--LM---E--Q-----------M---HS---K------S-----E-----R---S-------------------S---..-------E---E---R.----------------I-----------------.----------N----V--------  790
H1N.CM.95.YBF30          -----SIA------A-----LM---E--RD----------M---HS---K------S-----E-----R---S-------------------S---..-I-------E---ED-DR.-----K------------I-----------------.----------N----V--------  790
H1P.CM.06.U14788         ---IIK-EE-----A--E-VL---KE--EQA---------V----S-T----D-P--QK---EMT-------S-----------NI-----KD--..R----E---Q--ED---.--------L--E-G---I-----INN-P--HV-.--------------------M  780
H1P.FR.06.RBF168         ---IIK-EG-----A--E-VL---KE--KQA---------V----S-T----D-P--QK---EMT-------S-----------NI-----KD--..R----E---Q--ED---.--------L----G---I-----IN--P---V-.--------------------T  779
CPZ.CD.06.BF1167         KSR-IE--E----QA-----L---K-AKEK---------V----L-GT----D-P--EE--QE-MN--A---N-----------E-------KN--..-----E---Q--ED--R.F----KML-DEY----I-----I-Q-N--HR-.---I--------E--------I  786
CPZ.CD.90.ANT            -SR-KH-QK----QA-----LM---E--TGP-------------VL-GT-------P--EE--QK---R-QI--S-------------------Q---..Q----E-------ED-D-.------SL-DEY----I-----I-Q----HV-.---R-----------V-----  783
CPZ.CM.01.SIVcpzCAM13    K--IIS-EN----QA--G-L-------EQQ---------V-----S---H-------------E-----D----S-------------------S---..-----------E---R.-----K-L-GE----I-----------N-----V-.---L----------T--M-----  782
CPZ.CM.05.LB715          K--IIS--E-----A-----Q-------S-------------------G---K-----I------S--N--R---S-----------E-------S---..R-----------E----.--N--K--------I--I--------------.--------------------  788
CPZ.GA.88.SIVcpzGAB2     K---IS-EE-----A--E-LRI------Q----V--------V----NS-------I--------E----------T-------------S---..-----------ED--R.---------DE-Q--------------T--------.---I----------------  786
CPZ.US.85.US_Marilyn     ---II--EN-----A--T-V----K--ENT--V--G---V----HS-------------E-------S-IS-----------I------S---..-----------ED-D-.------T----------------------P-.---I----------------  785
GOR.CM.04.SIVgorCP684con ---IRK-EG-----A--E-VL---EE--KN----------V-----A-S----D-P--QK---EMTT-----------------EI-----KD--..R----E---Q--ED---.--------L----G---------IN--P--HV-.---R----------------  775
GOR.CM.07.SIVgor2139_287 ---IRK-EG-----A--E-VL---EE--KN----------V-----A-S----D-P--QK---EMTT-----------------EI-----KD--..R----E---Q--ED---.-------TL----G---------IN--P--HV-.---R----------------  775
GOR.CM.07.SIVgorCP2135con ---IRK-EG-----A--E-VL---EE--KN----------V-----A-S----D-P--QK---EMTT-----------------EI-----KD--..R----E---Q--ED---.--------L----G---------IN--P--HV-.---R----------------  775
GOR.CM.12.SIVgor_BPID1   ---IIK-EE-----A--E-VL---KE--KQA---------V-----S-T----D-P--QR---EMT--D----S-----------EI-----KD--..R----E---Q--ED---YHSNWKTLASDFGLP-I-A.---IN--P--.HVK------------------T  780
GOR.CM.12.SIVgor_BQID2   K--IIK-EE-----A--M-VLV-----KK-----------V--V-S---T--D-P--Q----E-T---R----------------KI-----KD--..R----E---Q--ED---YHSNWRALASDFGLP-V-A.---I---P--.HVK-------------------  779
GOR.CM.13.SIVgor_BPID15  ---I-K-EG-----A--E-VL---KE--KQA---------V-----A-T----D-P--QK---EMT-------S-----------EI-----KD--..R----E---Q--ED---YHSNWRTLASDFGLP-I-A.---IN--P--.HVK------------------T  778
                                                                                                                    p15 RNase H end p31 Integrase start
MAC.US.x.239             KD--KV-EQ----QA--E-FLM--T---PKA--IV----VM---TGC-TE---R-------EM---SEI-V-----------QEI-H---Q---..Q----EK-EP--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.---I---ANSDL-T--M-----  794
H2A.DE.x.BEN             KD--KV-EQ----QA--EVFRM--A---PK---IV----VM---VAG--TE--NRI-------EM----A--V-----------QE--H---Q---..Q----EK-EP--E----.---IIKELTHK-GI-LL--RQ--N--AQ--Q-.---I----NAEI-V--M-Y---  807
H2B.CI.x.EHO             KD--KP-EQ----QA--E-FA-------PQ---IV----VM---VA---TET--PI-RE---EM-----I-VG-------L---QE--H---Q---..QI---EK-EP--E----.--N-VKELVHK-GI-QL--RQ--N-------Q-.---I----NSEL-T--M-----  807
H2G.CI.92.Abt96          KD--KV-EQ----QA--E-FAM------PK---IV----VM---AG---TET--P---K---EX-----XL-VG-----------QE--X---Q---..Q----XK-EP--E----.F---VKELTHK-GI-QL---Q--N-YHS--Q-.---I----NAEL-T--M-----  790
H2U.CI.07.07IC_TNP3      KD--KV-EQ----QA--E-FLM------SQA---V----VM---AG---TET--P---K---EM-M--AL-XG-----------LE--H---Q---..Q----EK-EP--E--D-.F---VKELVFK-GI-RL---Q--DT----HQ-.---I----NTEL-T--M-----  790
H2U.FR.96.12034          KDR-KP-EQ----QA--E-FAM--E---PK---------IM---AG---TET--P--SK---E--R--A--VG-----------QE--H---Q---..Q----EK-EP--E----.----VKELQHK-GI-QL---Q--NT-HV--Q-.--------NTEV-T--M-----  803
ASC.UG.10.RT03           KE---A-EH----A-----LL-----G-TKQ--------V----TGA-TETDHPILE---T--QS--AI--S----------------A------K---..R-----EQ-PQ--E---R.--N--QDLRDR-QI-ALI----LKA-P---G-.--P-----NMEV-L--M-----  792
ASC.UG.10.RT08           KE---A-EH----A-----LL-----G-TKQ--------V----TGA-TETNHPILE---T--QS--AI--S----------------A------K---..------EQ-PQ--E---R.--N--QDLRDR-RI-ALI----LKA-P---G-.--P-----NMEV-L--M-----  795
ASC.UG.10.RT11           KE---A-EH----A-----LL-----G-TKQ--------V--VVTGA-TETDHPILE---T--QS--AI--S----------------A------K---..R-----EQ-PQ--E---R.--N--QDLRDR-QI-ALI----LKA-P---G-.--P-----NMEV-L--M-----  795
COL.CM.x.CGU1            EEE--N-HEGS--QA--TGVL---KHGPKRM-L----L------TG--YDQLPSSRGE--QAGMA--AIHV--C-----------LI-QK-..-V-..Q-MWI-K-EA-EED-Q-.F---VQYLKEQ-G--T------WER-SE--N-.-Q-V---L-Y-Y-L------E  761
COL.UG.10.BWC01          EEE-I--KEGS--QA--R-VLM--EHGPKRM-L----L--F--VTG--EE-N-P--E---QKGLE--AIHVS-C-----------II--K-..-V-..Q-MWMEK-EE-ELD-Q-.----VQY-QEQ-GI-QL-----WER-PQ--NR.-QP---SL-Y-Y-V--M----E  763
COL.UG.10.BWC07          EEEIEV-HEGS--QA--KG-LM--KHGPKRM-L----L------VSG--EETN-P--E---QEG----ALHV--C-----------LI-QK-..-V-..-AMWIEK-EE-EED-S-.F---VQHLRETYG--AI-----WDR-SQ--NR.--PV--SL-Y-Y----M----E  763
DEB.CM.04.SIVdeb04CMPF3061 K-S-KA-EN----A--T-VLM--E---EK---------V-N-LTEH-TTT-H---EK--Q--Q--QE---Q---------L-----I-----K---..-I---ERVEE--ED-D-.-----KQLREE----TL---Q-INQ-P---IH.--PK----NADL-----M-----  791
DEB.CM.99.CM40           K-S-KC-EN----A--E--L---EEGPSKM------------LEH-SET-HKI-EKV-QA-QG--QI------------------D-------K---..RI---ERVEE--ED---.--G--KQLRDEYQ--TLI--Q-ITQ-P---TR.--PT-----A-M-----M-----  788
DEB.CM.99.CM5            K-G-KS-EN----A--E--L---EEGPSKM------------LLEH-TET-H-I-EK--QA-QE--Q---N------------D-------R---..RI---ERVEE--ED---.--G--KQLRDE-Q--TLI--Q-I-Q-P---I-.--P------A-M-V--I-----  788
DEN.CD.x.CD1_CMO580407   K------EE----A--IGVLM-----PNK-------------AGT-TT--HQI-E---QAMQG--EI-IT-----------TEI------K---..QI---EQ-PQ--E--D-.-----KEIRDKYG--TM-----INV-P---TH.--PIR----A-L-T--M-----  810
DRL.DE.11.D3             KE--IA-EE-----A-----L---K---PK----------M---SSS-EI-DNPI-------M-S--A---N-------------E-------R---..Q-----NME---E--DL.--N---SL-QE-G--GI-------Q-P---IH.--PI-----A---T--M-----  797
DRL.DE.11.D4             KE--IA-EE-----A-----L---K---PK----------M---SSS-EI-DNPI-------M-S--A---N-------------E-------R---..Q-----NME---E--DL.--N---SL-QE-G--GI-------Q-P---IH.--PI-----A---T--M-----  797
DRL.DE.11.D5             KE--IA-EG-----A--*--L---K---PK----------M---SSS-EI-DNPI-------M-S--A---N-------------E-------R---..Q-----NME---E--DL.--N---SL-QE-G--GI-------Q-P---IH.--PI-----A---T--M-----  796
DRL.DE.11.D6             KE--IAXEE-----A-----L---K---PK----------M---SSS-EI-DNPI-------M-S--A---N-------------E-------R---..Q-----NME---E--DL.--N---SL-QE-G--GI-------Q-P---IH.--PI-----A---T--M-----  797
DRL.x.x.FAO              KE--IAIEE-----A-----L---K---PK----------M---SSS-EI-DNPI-------M-S--A---N-------------EE------R---..Q-----NME---E--DL.--N---SL-QE-G--GI-------Q-P---IH.--PI-----A---T--M-----  797
GRV.ET.x.GRI_677_gri_1   KSRIRE-EN----QA--T-VKM--E---EN---------VMN-LT-C-QE-N-P--E---QA-M--RQ---Q-----------TEI------K---..QI----R-EE---D-A-.--N---S-VQE-G--NI-------A-P---IR.--PK-----A-IET--M-----  805
GSN.CM.99.CN166          KHRAIE-EE----QA--H-VL---KEGPPKM-L------VM--L-S--EV-T-A--E---QE-L---AI--S-----------TE-------K---..Q----EN-EP-VED---.-----KYLRDQYKI-ALL-----NK-S---VH.--PK----NAEL-M--M-----  794
GSN.CM.99.CN71           KYRAIE-EG----QA----VL---KEGPPRM-L------VM--L-S--EV-T-P--E---QE-LG--A---S-----------TE-------K---..Q---MEN-EP-VED---.-----KYLRDQYKI-ALL-----NK-S---IH.--PK----NAEL---M-----  787
LST.CD.88.SIVlhoest447   KE--KE-------QA--E-VL---R--KSR---------VIKVLSQR-TET-HPV-K----ECK--DQ---G------------QE--H---K---QKQ----GK-EP--E-YS-.F-N-AKDLEEK-HI--M---Q--N--N---T-.---IT----T------I----M  802
LST.CD.88.SIVlhoest485   KE--KE-------QA--E-VL---K--KSR---------VMKVLSQR-TET-HPV-K----ECK--DQ---G------------QE--H---K---QIQ----EK-EP--E--S-.F-N-AKDLEEK-H---M---Q--N--N---T-.---IT----T------I----M  802
LST.CD.88.SIVlhoest524   KE-IKE-Q-----QA--E-VL------KST---------VMKVLSQR-TET-HPI-KN--KE-Q--DA---G------------QEI-Q--IK---TKQ-M--EK-EP-VE--G-.F-N-AKDLEEKYGI--M---Q--NE-A---Q-.---IT----V-V---------  803
LST.KE.x.lho7            KE--KE-------QA--E-VLM-----NSK---------VMK-LSQR-TET-HPI-KD----CKQ-DQ---G------------QE--H---K---QKQ-M--EK-EP-VE--S-.F-N-AKDLEEK-----M---Q--ND-AN--K-.---IT----V-V---------  802
MAC.US.x.251_BK28        KD--KV-EQ----QA--E-FLM--T---PKT--IV----VM---TGC-TE---R-------EM---SEI-V-----------QEI-H---Q---..Q----EK-EP--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.---I----NSDL-T--M-----  790
MAC.US.x.EMBL_3          KDI-KV-.T----QA--E---HGIE---PKR--IVEL-VCY-NNNRF-TE---R-------EM--.VR--V-----LE------QEIGP---Q-F-..Q----EK-EP--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.---I----NSDL-T--M-----  788
MND-1.GA.x.MNDGB1        F-R-EEYLN----Q---H-VK---E---SY---------VV--LASR-TETDHPI-KE---LMKG---I--S-L-----------I------S---..------QN-EP--E----.----EAQLREK-H--AL---Q--Q--S--CHH.--PIK--T-A-L-V--I-----  792
MND-2.CM.98.CM16         KSR-IA-EN-----A--E--KM------PK-------------M--LS-A----DNPI-RE---LM-H--A--------------------R-V-..Q-----E--------E--D-.--N-----L-Q--CI-NI-------Q-P---T-.--PI-----T---T--M----M  795
MND-2.GA.x.M14           KS--IA-EN-----A--E--L---K---PRA-------------LS-T--R-DNPI-RE--N-M-A--A-------------------I------K---..Q-----E--------E--D-.--N---SLSQE-SI--I-------Q-P---V-.--PI-----AA--T--M-----  796
MND-2.x.x.5440           KS---A-EE-----A--T--KM------PR--------------LS-A----DNPI-RE---LM-G--EI--S--------------I------K---..Q-----E--------E--D-.--N---LSQ-----I-------Q-P---T-.--PV----NAD--T--M-----  796
MNE.US.x.MNE027          KD--RV-EQ----QA--E-FLM--T---PKA--IV----VM---TGC-TE---R-------EM---TEI-V-----------QEI-H---Q---..Q----EK-EP--E--D-.----VKELVFK-G--RL---Q--DT----HQ-.---I----NSDL-T--M-----  790
MON.CM.99.L1_99CML1      AR-A---EQ-----A--E-VL---R-GPPQM--I-----V---LASC-EI---PI-EA--QE-L--DQ-F-S-------------E--R--GQ---..Q---MENLEP-RED---.-----KYLRDTYHI-TLL-----NH-H---NH.--PKS----NAEV-V--M-----  790
MON.NG.x.NG1             QR--IP-EQ-----A--E-VL----NGPASM--------VM--LASK-EV---PI-EK--Q--LG-KQ---S-------------E------Q---..Q---MEN-EP-RED---.-----KYLRDTYGI-TLL-----NH-P--HTQ.--PK-----SDI----M----Y  793
MUS-1.CM.01.CM1239       KE---S-EN-----A--E-VL---KEGPPSM----------V---VAS--QE-T-P--EE--Q--LT--A---S-------------D-------H---..Q---MEQ-EP-KED---.-----KYLRDKY-I-ALL-----NL-P---TH.--PKT----NAEL-V--M-----  784
MUS-1.CM.01.SIVmus_01CM1085 KE--IS-ES-----A--E-VL---KEGPPRM-L------V---VAS--QE-T-P--E---QE-LS--AI--S-------------E------NS---..Q---MEQ-EP-KED---.-----KYLRDKYQI-ALL-----NH-P---TH.--PKT----NAEI-V--M-----  784
MUS-2.CM.01.CM1246       QERAIN-EN----QA-----L---M-GPP-M-L------------T-A-EV---P--E---Q-MLS-NAIFISR------------E--H---R---..Q---MEN--P-VED---.-----KYLRDKY-I-TIL-----NK-SA--TH.--PK----NADL-V--M-----  789
MUS-2.CM.01.CM2500       QE-AID-EN-----A-----LM--T-GPPKM------------T-A-ET---P--EK--Q--LT--AI--S-----------QE--H---K---..Q---MEN-EP-IED-D-.-----KYLRDKY-I-TLL-----NR-SA--TH.--PK----NADI-V--M-----  788
MUS-3.GA.09.09GabOI81    ---AID-EN----QA-----L---K-GPPKM-L----------VT-C-EA---P--E---Q--LT--EI-IT-----------TE--R---Q---..Q---ENVEP-KED-D-.-----KYLRDVYQI-ALL-----NH-S---TQ.--PKT----NAEI-----M-----  785
MUS-3.GA.11.11GabPts02   QR---E-QG-----A--E-VL---K--PSKL-L------V----LSE-ET-D-P--E---Q-MLQ-QAI-VS-----R-----ME--R---Q-V-..QI--MEN-EP-VED---.-----KYLRDTYKI-TLL-----NK-G---TQ.--PK----NAEL----M-----  788
OLC.CI.97.97CI12         KH-I-HIQESS---A-IR--IE--R--KS--------L--M--VLEK-EV--EP-IQEL-KE-E---RI-IG--------L---QE--Q---K---TKQ-M--AD-EP--ES-N-.--Q-AQSLQQE-QI-LI---R--DE-EE-KG-.-VPR--I-AVGVYR--I-----  784
RCM.CM.00.SIVagi_00CM312 KE--KE-E------A--E-VL---K---KR---------VF--LAGS--T---P--Q--------G--E---S-------------E------R---..Q-------E---E--D-.-----------E-RI-Q-------Q-P---V-.---V---T-A---T--M-----  796
RCM.CM.02.SIVrcm_02CM8081 -E--KE-E------A--E-VL---K---KK---------VF--LAGS--T-D-P--Q--------G--E---S-------------E------K---..Q-------E---E----.--N------E--QI-Q-------Q-P---V-.-----------A--KT--M-----  802
RCM.GA.x.SIVRCMGAB1      KE--KE-EE-----A--E-VL---K---PK---------VY--LE----T-D-G--TE--N-M-G--A---S-------------E------K---..Q-------E---E----.--N----L-E--QI-QI-------Q-P---V-.---I-----A---T--M-----  796
RCM.NG.x.NG411           KE--KE-EN-----A--G-VL---K---KR-------RHVF--LAGS--T---P--Q-----M-G--E---S-------------------Q---..Q--------G---E----.---------EE-QI-QI-------Q-P---V-.---V-----A---T--M-----  801
SAB.SN.x.SAB1            D----A-EN-----A--E--L---R---SK---I------M---AGE-TE-DNNI-Q-----E-----A--I--------V----EI-------Q---..Q-----R-EE--E--D-.--A---S-QQE-G--AI-------A-P---I-.--SV-----A---V--M-----  815
SMM.CI.79.SIVsmCI2       -E--KP-EQ----RA--E--AP--R-A-PKA-V------VM--VAGH-TETD-P---E---E-----AI-VG-------L---QE--H---Q---..Q----EQ-EP--E----.----VKQLVHK-GI-QL---Q--NT-TQ--Q-.---I----NAEL-T--M-----  799
SMM.LR.89.SIVsmLIB1      -D--KA-EQ-----A--E-FLM-M----P-A--IV----VM--VAG--TE--NN-------EM---TAI-V-*----------RE--H---QR--..Q--#-EK-EP--E----.----IKELVFKYGI-RL---Q--DT----HQ-.---I----NAEL-T--I-----  784
SMM.SL.92.SIVsmSL92A     KD--KV-EQ----QA--E-FAM--A---PKA--------V---VLG--TE---PI--E--NHMML--EL-V-----------L---QE--H---Q---..Q----EN-EP--E--D-.----IKELVYK-GI-#L---Q--NT------Q-.---I----NADI-T--M-----  789
SMM.SL.92.SL92B          KE--LP-EQA---QA-----LL---K--PSK--V------V-N--TG--SE-D-DI-A--------VQ--A--IG-----------NE--R---Q---..Q----ES-EP--ED-D-.----VKEL-QKY-I-QL---Q--NA-N---Q-.---I---TNAEV-T--M-----  787
STM.US.89.STM_37_16      KN--KA-EQ----QA--E-FAM--A---PKA---V----VM---TG--TE---K-------EM----AI-V-----------QE--H---Q---..Q----EK-EP--E----.----VKELVFK-GI-RL---Q--DT----HQ-.---I----NAEL-T--M-----  790
SUN.GA.98.L14            -SR-KK-EN----QA--E--KM--E--RSS------------RLLSKR-TETD----KE-V-LIRQ-DQ---G-----------QEI-Q---Q---KRQ-M-IEK-EP-VE--G-.F-N-AASLQEM-DI-L----Q--NE-AQ--Q-.---IT----A-V-----I----M  802
SYK.KE.x.KE51            K-TLKR-EN-----A--E-VLM-------P----------------LMNC-TS--HPV-E--MQKAME---I-IT-----------QE--Q---K---..-I----N-P---E----.--T-VEYIRQE-H--RQ---A--EM-P---IR.--PK-----TDL-T--M-----  822
SYK.KE.x.SYK173_COMGNM   T-R-QY-EQ----QA--EGLLM-----KDK----V----SY--LMTC-TNT-HPI-E---QEA----AI-VT--------------A------K---..-I---ER-PQ--ED--R.----MEYLRQE-H--RQ---A-IQQ-P---NR.--PK-----VDIYN--M----E  817
TAL.CM.00.266            KEH-IK-EK-----A--E-VL---K---PK---------VY--LAGS-TETDNKIIED--QL-LS--A-----L--------------------Q---..-I--VEQ-PE--E---R.--N---DLKAR-K--TI---A-IEA-P---VQ.--PKT--NNAAV-T--M-----  791
TAL.CM.01.8023           KEH--G-SN-----A--E-VL---KH--PK---------VY--LVGN-TETDNKI-ED--Q--LG--A---T-L-------------------Q---..RI--VEQ-PE--E---R.--N--KDLKAR-K--TI---A-IEA-P---VQ.--PKT--TNAAV-T--M-----  792
TAN.UG.x.SIVagmTAN1      ESQ-IE-E-----QA-----VL---KH-KSK--------VM--LT---T--D-V--------EM-N-DAI--T-----------QEI------Q---..Q-----RMEE--ES-D-.--T--QFIRDA-GI-AL-------A-P---IR.--PI-----A-V-V--M-----  802
VER.DE.x.AGM3            K-R-EK-EN----QA--M--KM--E---PN----------M--LT---T--D-P-IE---ALMVQ-HQI--Q----D--------EI------Q-M-..-I---EK-EE--E---R.--N---NL-DTYG--QI-------M-P---I-.--PV-----A---V--M-----  799
VER.KE.x.9063            K-R-EK-EQ----QA--V--KM--E---EK----------M--LT---T--D-P--E---ALMV--RAI--Q-----------EI------Q---..R---IGR-EE--E--DR.------KNL-DT-G--QI-------M-PNAKI-.--PI-----A---V--M----V  799
VER.KE.x.AGM155          K-R-QQ-EN----QA--T--KM--E---PK----------M--LT---T--D-P--E---A-MVQ--AI--Q-----------EI-----K-V-..RI--IGR-EE--E--DR.------NL-DT-G--QI-------M-P---V-.--PI-----A---V--M----I  802
VER.KE.x.TYO1_patent     K-R-E--EN----QA--T--KM--E---PN----------M--LT---T--D-P--E---ALM-Q-QQI--Q-----------EI------K---..R----EK-EE--E---R.--N--KNL-DTYG--QI-------M-P---I-.--PV-----A---T--M-----  802
WRC.CI.97.97CI14         KE---SVQN----Q---I-LK---E-G-QKM--I-----VM--VKDH-TM---P--QE---L-KG--EI-ID-----------TEI-R-------S.-IM--EN-TP-IES-T-.--QG-K-LQEE-H--QLI--R-IEE-ST--G-G-AP---M--YEV-T--M-----  807
WRC.CI.98.98CI04         KE--IN-YN----QA--A-LK---E-G-SKI--I-----VM--VKDH-TI---A--QE--DL-KQ--EI-IE-----------TEI---------S.--M--EN-TP-IES-T-.--QG-K-LQEE-HI-QLI--R--EE-ST--G-G-APL--M--YEV-T--M-----  806
WRC.GM.05.Pbt_05GM_X02   -N-IQKI-N----Q---I-LRM--E--SDK---I-----V----KDH-T----P-MQE--DL-KG--EI-IT-----------TEI-A-------S.R-M--EN-TP-IES-T-.--QG-G-LEEE-H--TIL--R--EE-AT--G-G-HPI--M--YEI-T--M-----  798
```

```
H1B.FR.83.HXB2              EGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTI.HTDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQI.TKIQNFRVYYRDSRNPLWK 951
H1A1.UG.85.U455_U455A       --------------------------------I----------V-.---------S-A-K-V----N-Q------------------------------E---------------------------------I----------------.S-------------D-I-- 950
H1C.ET.86.ETH2220           ---I--------------------------------------RV-.---------SNA-K--------Q------------------------------E--------------------R------------I----S--------N--.L-------------D-I-- 958
H1D.CD.84.84ZR085           ----------------------------A--------------VV.---------S---K-------------------------------------------------------------------------I----S----R------.--------------D-I-- 951
H1F1.BE.93.VI850            -------------E------------------I----------I-.---------S-A-K-S------Q-----------------I----------------------------------------------I---S-----R------.----------------V-- 952
H1G.SE.93.SE6165_G6165      ---I-I--------------------------I---------TV-.---------S-A-K------N-T----------------------------------------------------------------I----S-----------.--------------D-V-- 948
H1H.CF.90.056               --Q-----------------------K----------S-----V-.---------S-A-K------D-Q----------------------------------------------------------------I----------------.SN--K---------D-I-- 950
H1J.SE.93.SE9280_7887       ----------------------------A-F-I----------V-.---------SGA-K------D--------------------------------E---------------------------------I---------R------.--------------D-I-- 949
H1K.CM.96.96CM_MP535        ---I----------------------------I----------V-.-----T---STV-K-------V-----------------------------------------------------------------I----------------.LN--K---------E-I-- 951
H1O.BE.87.ANT70             ---I-I--------F----------------------A-----V-.-----P---ST-MK------N-QH---------------A------S--Q---------R--------V------------T-----I--L-SQ---T------.L-XHK---------D-I-- 947
H1O.CM.94.BCF06             ---I-I--------F----------------------A-----I-.-----P---S-AMK-----T--QH---------------A------S--Q------------------V------------T----LI--L-SQ---T------.L-----Q-------D-I-- 947
H1O.FR.92.VAU               ---I-I--------F-----------------------A--I-IL.-----P---S-AMK------N-QH---------------------ES-MQ------------------V------------T----LI--L-SQ---T------.L-------------D-I-- 948
H1N.CM.02.DJO0131           ---I-----------L----------------I----------V-.-----P---S---K------N-T----------------------------I---------------------------E-T-----I---------TK--T--.L-------------D-I-- 956
H1N.CM.04.04CM_1131_03      ---I-----------L----------------I----------V-.-----P---S---K------S---------------A--------------I-----------------------------T-----I---------TK--T--.L-V-----------D-I-- 958
H1N.CM.95.YBF30             ---I-----------L----------------I----------V-.-----P---S---K------N---------------A--------------I-----------------------------T-----I---------TN--T--.L-V-----------D-I-- 958
H1P.CM.06.U14788            ---I----------FM------V-----A---V----S-----V-.-----P---S-A-K-----LN-TH-------------------------H------------------V------------T--D--I--L--Q---T------.L-----Q-------D-I-- 948
H1P.FR.06.RBF168            ---I----------FL----------------V----A-----V-.-----P---S-A-K-----LN-TH-------------------------L------------------V------------T--D--I--M--Q---T------.L-------------D-I-- 947
CPZ.CD.06.BF1167            ----V---------F-----LSS---K---F-I----------K-.-----A---S-A-K------N-Q--------------------Q--I-VD-I-E-----R---A---Y-------------TP-D--L--LT------------.F-L-----------D-V-- 954
CPZ.CD.90.ANT               -----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---Q---H-----D-V-- 951
CPZ.CM.01.SIVcpzCAM13       ---I-I----I---------------------I----S-----VL.-----T---SNA-K------N---K------------------------E------------------S------------T-----I----------------.L---K---------D-I-- 950
CPZ.CM.05.LB715             ---I----------------------------I----------V-.---------SNA-K------Q-Q----------------------------------------------------------------I---------T------.L-V-K---------D-I-- 956
CPZ.GA.88.SIVcpzGAB2        ---I-----------L----------------I----------K-.-----T---S-A-K-------------------------------------------------------------------T-------ML---L--Q------.L---K---------D-I-- 954
CPZ.US.85.US_Marilyn        ---I-I---------L----------------I----------V-.---------SS-----------Q----------------------------I-----------------------------T----I----SEL--DL------.L-V-----------D-I-- 953
GOR.CM.04.SIVgorCP684con    ---I----------F----------------------A-----V-.-----P---S---K-----LNVTH-------------------------Q------------------V------------T--D--I-ML--Q---T------.L-------------D-I-- 943
GOR.CM.07.SIVgor2139_287    ---I----------F----------------------A-----I-.-----P---S---K-----LNVTH-------------------------Q------------------V------------T--D--I-ML--Q---T------.L-------------D-I-- 943
GOR.CM.07.SIVgorCP2135con   ---I----------F----------------------A-----I-.-----P---S---K-----LNVTH-------------------------Q------------------V------------T--D--I-ML--Q---T------.L-------------D-I-- 943
GOR.CM.12.SIVgor_BPID1      ---I----------FL---------------------A-----V-.-----P---S-A-K-----LN-TH-------------------------L------------------V------------T--D--I-MM--Q---T------.L-------------D-I-- 948
GOR.CM.12.SIVgor_BQID2      ---I----------F----------------------A-----V-.-----P---S-A-K-----LN-TH-------------------------Q-------Q----------V------------T--D--I----SQ---T------.L-------------D-I-- 947
GOR.CM.13.SIVgor_BPID15     ---I----------FL---------------------A-----A-.-----P--AS-A-K-----LN-TH-------------------------L------------------V------------T--D--I--M--Q---T------.L-------------D-I-- 946

MAC.US.x.239                ---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-------EG-DQ--- 962
H2A.DE.x.BEN                ---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--Q----EG-DQ--- 975
H2B.CI.x.EHO                -----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--------EG-DQ--- 975
H2G.CI.92.Abt96             -----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-------EG-DQ--- 958
H2U.CI.07.07IC_TNP3         -----I--------F-------Q---RQ--L-------T--ISHL.-----A---SQK-KMVA--V--E-S--V-----------A--HH--NQ-DRI---NSME-I-L--THCM----R----DMTPA--LINM-T-EQEIQFQ-TKN.S-FK-------EG-DQ--- 958
H2U.FR.96.12034             ----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-------EG-DQ--- 971
ASC.UG.10.RT03              ---I-I---------T--KL--Q---K---I---Q-CA---TQ-.-----P---SQELA--A---NVQHST-V-----------NK--Q--ET-SKI--EVQY-E---A--LL-L---KR-----MTPA--LINM-H--LELQQ..TKF.P-FS-------TGKS-E-L 958
ASC.UG.10.RT08              ---I-I---------T--KL--Q---K---I---Q-CA---ITQ.-----P---SQELA--A---N-QHST-V-----------NK-RQ--ET-SKI--EVQY-E---A--LL-L---KR-----MTPA--LINM-H--LELQQ..TKF.P-FS-------TGKS-E-L 961
ASC.UG.10.RT11              ---I-I---------T--KL--Q---K---I---Q-CA---TQ-.-----P---SQELA--A---NVQHST-V-----------NK-RQ--ET---KI--EVQY-E---A--LL-L---KR-----MTPA---INM-H-ELELQQ..TK-.P-FS-------TGKS-E-L 961
COL.CM.x.CGU1               -----------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--KA-WKEH.TGE-Q 929
COL.UG.10.BWC01             -----------CTLFCW-TILKR-S-E--GRA-I---SQ-E--QV.-----P---SQAF---V--L---HTT-----------I--QR--MI-EG-KKMEE---T-ESK-A---YAL-------L--K-PW--Q-ERAIIELE-QNITQIQNK-F---K--WKEH.TGE-Q 931
COL.UG.10.BWC07             --AI-------CTLFCQ-FMLKR--AE---RG-I---SM-E--QV.-----T-----QAFK--V--L--THTT-T---------I--QR--M--ET-RKMK----T-ESR-A---YAL-------L--K-PW--Q-ERAIIELD-QNITQIQNK-FK--K--WKEH.TGE-Q 931
DEB.CM.04.SIVdeb04CMPF3061  -----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-FSG-K----EGAD-S-- 959
DEB.CM.99.CM40              -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-LLG-K----EGAD-T-- 956
DEB.CM.99.CM5               -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-LLGYK----EGAD-T-- 956
DEN.CD.x.CD1_CM0580407      -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--YINMLN-HLEIQH..TKQ.P-FS--K----EGA--Q-Q 976
DRL.DE.11.D3                ---I-I-------------------EK-------------SHL.-----P---SEK-ATV----K-EHTT-V---------------NQ--------I-----K-E-------L--------------E---A---------S-LL-NK--QN-.L----------EG-DQ--- 965
DRL.DE.11.D4                ---I-I-------------------EK-------------SHL.-----P---SEK-ATV----K-EHTT-V---------------NQ--------I-----K-E-------L--------------E---A-----------LL-SK--QN-.L----------EG-DQ--- 965
DRL.DE.11.D5                ---I-I-------------------EK-------------SHL.-----P---SEK-ATV----K-EHTT-V---------------NQ--------I-----K-E-------L--------------D---A-----------LL-SK--QN-.L----------EG-DQ--- 964
DRL.DE.11.D6                ---I-I-------------------EK-------------SHL.-----P---SEK-ATV----K-EHTT-V---------------NQ--------I-----K-E-------L--------------E---A---------S-LL-NK--QN-.L----------EG-DQ--- 965
DRL.x.x.FAO                 ---I-I--------H----------EK-------------SHL.-----P---SEK-ATV----K-EHTT-V---------------NQ--------I-----K-E-------L--------------E---A---------S-LL-SK--QN-.L----------EG-DQ--- 965
GRV.ET.x.GRI_677_gri_1      -----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--K----EG-D-V-- 973
GSN.CM.99.CN166             ---------------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-K------QGA----Q 962
GSN.CM.99.CN71              ---------------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-K------QGA--H-Q 955
LST.CD.88.SIVlhoest447      --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--LK-K----QG-D-Q-- 970
LST.CD.88.SIVlhoest485      --Q--IN-I---C-FMV-----D---KT--N-----CS----QQV.-----P--VSKE-Q-VT--L--EHST-------------AK--V--Q--ER--ED-QQ------L---H------QR--L--L-PA--FINM-NA-LE-QY---LN.S--LK-K----QG-D-Q-- 970
LST.CD.88.SIVlhoest524      --QI-MN---M---FMV-----D---KT--N-----CS-----Q-.-----P--VSKD-Q-VT--LS-EHTT-------------AK--V--Q--ER--ED-QQ------L---H------QR--L--L-PA--YINM-NA-LE-QYI--LN.S--LK-K----QG-D-Q-R 971
LST.KE.x.lho7               --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--LK-K----QG-D-Q-- 970
MAC.US.x.251_BK28           ---IVI--------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-------EG-DQ--- 958
MAC.US.x.EMBL_3             ---IVI--------F-------Q---RQHY.----------YLH-YTHS--A--ASQE-KMVT-----EAHLWV-----------A--HH--NQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--LINM-T-EQEIQFQ-SKN.S-FK-------EG-DQ--- 956
MND-1.GA.x.MNDGB1           -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--K--F-EG-DQ--- 960
MND-2.CM.98.CM16            -----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---------EG-DQQ-- 963
MND-2.GA.x.M14              -----I--------L------------K---H----------HL.-----P--VSEK-ATV----Q-EHTT------------AK-HH--V--E---------R-E-------L---------------E--P---------T---L-TK--QN-.S----------EG-EQ--- 964
MND-2.x.x.5440              ---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-----Q----EG-DQQ-- 964
MNE.US.x.MNE027             ---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-------EG-DQ--- 958
MON.CM.99.L1_99CML1         --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--.--FSK-----F-TGAD-Q-- 958
MON.NG.x.NG1                --SI-----------VW-KIL-R---K--GIA--E--AT---IHHV.-----P---SQEFA--A---N-EHST-V-----------N---QI-TTL-KI--EVQF-S--LAQTLY-L----X----DI-PI--L-NM-H--LELQHSNN--.--FSK-----F-TGAD-S-- 961
MUS-1.CM.01.CM1239          ---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---------TGSD-S-- 952
MUS-1.CM.01.SIVmus_01CM1085 ---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---------TGSD-S-- 952
MUS-2.CM.01.CM1246          -------------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.--FSK-------RGA--Y-L 957
MUS-2.CM.01.CM2500          ---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..FSK-------RGA--S-Q 954
MUS-3.GA.09.09GabOI81       -E-I-----------WTKIL-R---RN--LA--Q--SQ---SH-.-----T---SEEFA--A-----EHTT-------------NA--Q--TQ-KKI--EVQY-E---AQ-TY-L----R--M-DL-P----INM-H-ELEIQY-NS-N.--F---------QGSD---- 953
MUS-3.GA.11.11GabPts02      ---------------VW-RLL-R---RD--LA--Q--AM---TH-.-----P---S-QFA--T--GN-EHTT-V-----------NI--Q--NT-SKI-EEV-Y-E---AQ-A--L---------DL-PI---INMLH-ELELQHT-N--.S-F---------QGGD---- 956
OLC.CI.97.97CI12            -----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-EL-NQ-Y--T.QI-YTQNK.F-KFSGYRV-YKNKEGQ-C 951
RCM.CM.00.SIVagi_00CM312    -----I--------------------K---------A-----HL.-----A---S-A-Q-V----Q-EHT--V---------------HQ--T----I-----K-E-------L-------------------I----SEL--NK--N--.S---------F-EG-DQQ-- 964
RCM.CM.02.SIVrcm_02CM8081   -----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---------F-EG-DQQ-- 970
RCM.GA.x.SIVRCMGAB1         ---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----------EG-DQ--R 964
RCM.NG.x.NG411              ---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---------F-EG-DQQ-- 969
SAB.SN.x.SAB1               ---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----------EG-D-V-- 983
SMM.CI.79.SIVsmCI2          -----I--------F-------Q---RQ--L------S---ITHL.-----A---SQE-KMVA--V--E-T--V---------------HH--NQ-DRI-E--NSIE-I-L--CHCM----R----DMTPA---NM-S-EQEIQF--TKN.S-FK-------EG-DQ--- 967
SMM.LR.89.SIVsmLIB1         ---I-I--------F-------Q---RQ--L-----VS---ITHL.-----A---SQE-KMVA--N-E-T-RV-----------A--HH--TQ-DKIKE--NSIE-I-L--THCM----R----DMTPA--L-NM-T-EQEIQFQ-SKN.S-FK-------EG-DQ--- 952
SMM.SL.92.SIVsmSL92A        ---I-I--------F-------Q---RQ--L------A-----HL.-----A---SQE-KMVA-----E-S--V-----------A--HH--NQ-DRI----VS-E-V-L--THCM----R----DMTPA--LINM-TAEQEIQYH-TKN.S-FKK------EG-DQ--- 957
SMM.SL.92.SL92B             -----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--------EG-DQ--- 955
STM.US.89.STM_37_16         ---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-------EG-DQ--- 958
SUN.GA.98.L14               -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--LK-K----EG-D-Q-- 970
SYK.KE.x.KE51               -----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--K----HGT.SD-Q 987
SYK.KE.x.SYK173_COMGNM      -----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-------QGK.NE-- 983
TAL.CM.00.266               --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--.QNFSD-K----KGA----Q 959
TAL.CM.01.8023              --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-ALINM-Y-EL--TT--N--.HNFSD-K----KGA----Q 960
TAN.UG.x.SIVagmTAN1         ---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--K----EG-E-V-- 970
VER.DE.x.AGM3               -----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-------EG-D-V-- 967
VER.KE.x.9063               ---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-------EG-D-V-- 967
VER.KE.x.AGM155             ---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-------EG-D-V-- 970
VER.KE.x.TYO1_patent        ----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-------EG-D-V-- 970
WRC.CI.97.97CI14            ---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--Q----TGA-QQ-Q 974
WRC.CI.98.98CI04            ---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--Q----TGA-QQ-Q 973
WRC.GM.05.Pbt_05GM_X02      ---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--Q----TGASQQ-Q 965
```

```
                                                                                           p31 Integrase end
                                                                                                  Pol end
H1B.FR.83.HXB2               GPAKLLWKGEGAVVIQDNS.DIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED...............*   1003
H1A1.UG.85.U455_U455A        -------------------.-------------------------M-G-----...............-   1003
H1C.ET.86.ETH2220            -------------------.----------------------A----G-----...............-   1011
H1D.CD.84.84ZR085            -------------------.---------------------------------...............-   1004
H1F1.BE.93.VI850             -------------------.E--I-----------------V-----G-----...............-   1005
H1G.SE.93.SE6165_G6165       ------------------N.E--------------------------G-----...............-   1001
H1H.CF.90.056                -------------------.E-------E------------------------...............-   1003
H1J.SE.93.SE9280_7887        -----P-------------.E--------------------------G-----...............-   1002
H1K.CM.96.96CM_MP535         -------------------.E--------------------------G-----...............-   1004
H1O.BE.87.ANT70              ---Q-------------KG.--------------E-------T-SM--G-T-SESVEQPSEIP.....-   1010
H1O.CM.94.BCF06              ---Q-------------KG.--------------N----V--T-S-----T-SESMEQPGETP.....-   1010
H1O.FR.92.VAU                ---Q-------------KG.----------------------A-S---G-T-GESVEQPSEIP.....-   1011
H1N.CM.02.DJO0131            ------------------G.-----------------------G----G-A-NQDMA...........-   1013
H1N.CM.04.04CM_1131_03       ------------------G.-----------------------G----G--NQDME...........-   1015
H1N.CM.95.YBF30              ------------------G.-----------------------G----G--NQEME...........-   1015
H1P.CM.06.U14788             ---T-------------KG.--------------N------------DT-R-SESLEQSG........-   1008
H1P.FR.06.RBF168             ---T-------------KG.---------------------------DI-R-SESLE...........-   1004
CPZ.CD.06.BF1167             ---R-----------KEGE.E------------K-----V--AGG--D--N--...............-   1007
CPZ.CD.90.ANT                ---Q-----------K-QE.E------------KE-R-KIEDR-DL-G--N--...............-   1004
CPZ.CM.01.SIVcpzCAM13        ---T-------------QG.E--------------------------------...............-   1003
CPZ.CM.05.LB715              --------------L--QE.E------------------V--N--M-GG---SQGME...........-   1013
CPZ.GA.88.SIVcpzGAB2         ---S--------------G.-------------KH----V-----L-GG---NQNME...........-   1011
CPZ.US.85.US_Marilyn         ---------------KE-E.EV-----------K--------A-SM-G----S...............-   1006
GOR.CM.04.SIVgorCP684con     ---T-------------KG.---------------------------DT-G-SESLEQSS........-   1003
GOR.CM.07.SIVgor2139_287     ---T-------------KG.---------------------------DT-G-SESLEQSS........-   1003
GOR.CM.07.SIVgorCP2135con    ---T-------------KG.---------------------------DT-G-SESLEQSS........-   1003
GOR.CM.12.SIVgor_BPID1       ---T-------------KG.---------------------------DT-G-SESMEQSG.....QIS-   1011
GOR.CM.12.SIVgor_BQID2       ---Q-------------KG.----------------------X-SM-XG-T-GESMEQPG.....EIP-   1010
GOR.CM.13.SIVgor_BPID15      ---T-------------KG.----------------------N----DT-G-SESLEQSG.....QVS-   1009
                                                                                           p31 Integrase end
                                                                                                  Pol end
MAC.US.x.239                 --GE---------ILKVGT.-------------K---GGKEV-SSSHMEDTGEAREVA..........-   1020
H2A.DE.x.BEN                 --GE---------IVKVGT.-----------------GRQEL-SSPHLEGAREDGEMAC..........   1033
H2B.CI.x.EHO                 --GD---------I-KVGT.E---I---------N--GGKEL-CSADVEDTMQAREVAQSN.......-   1036
H2G.CI.92.Abt96              --GE---------I-KVGT.E------------K---GGKEV-SSTNMEDTRQTGEVA..........-   1016
H2U.CI.07.07IC_TNP3          --GE---------I-KVGT.----I--------K---GGKEV-SGSNLEGARQDGEMAQLGEIP....-   1022
H2U.FR.96.12034              ---D---------L-KVGT.----I------------GGKEL-SSSNLEGA-KVREMALPDQTP.....   1034
ASC.UG.10.RT03               ---Q-----------KTTTGELLT--K------KP--ASQNVGNKEGMDSPRKEDTSLD.........-   1018
ASC.UG.10.RT08               ---Q-----------KTTTGELLT--K------KP--ANQNVG-TENMDSPRKEDTSLD.........-   1021
ASC.UG.10.RT11               ---Q-----------KTTTGELLT---------KP--ASQSVG-KENMDSPRKEDTSLD.........-   1021
COL.CM.x.CGU1                --GE-V---------RNSQGTLF-K----V--T-LQYGEDV-SENLL-NGQKEAETVKGMD.......-    991
COL.UG.10.BWC01              --GQ-V---------KNQEGSLF-K----V--TKVNYGERMVSKALLPNEQ-EAGEVKEVD.......-    993
COL.UG.10.BWC07              --GQ-V---------RSTEGTLF-K----V--S-ITHGTDQ-VVSEDLLPNGKEKPGETETLD.....-    995
DEB.CM.04.SIVdeb04CMPF3061   ---T-----------KLDTG-L---------V-K----DVGSKIDPQDTHE.................-   1011
DEB.CM.99.CM40               ---T----------CKTEVG-----------V-KE---D-DSKVNTQNSNE.................-   1008
DEB.CM.99.CM5                ---T----------CKTEVG-----------V-K----GVDSKTDTQNPSE.................-   1008
DEN.CD.x.CD1_CMO580407       ---E-----------KTQAG--FP---------KP-NAE.SRKGEQNKGM-SETDI............-   1032
DRL.DE.11.D3                 ---E-V---------KEGT.-L--I--------K----NVDSHTNME-....................-   1013
DRL.DE.11.D4                 ---E-V---------KEGT.-L--I--------K----N-DSHTNME-....................-   1013
DRL.DE.11.D5                 ---E-V---------KEGT.-L--I--------K----NVDSHTNME-....................-   1012
DRL.DE.11.D6                 ---E-V---------KEGT.-L--I--------K----NVDSHTNME-....................-   1013
DRL.x.x.FAO                  ---E-V---------KEGT.-L--I--------K----NVDSHTNME-....................-   1013
GRV.ET.x.GRI_677_gri_1       ---R-I---------KEGE.-------------K---ERKTM-SEGSMEGVREANKQMEGDSDLQDQE.   1040
GSN.CM.99.CN166              ---V----------V-TQAGE-IT---------KP--EAKTE-VGG-THTNN-...............-   1016
GSN.CM.99.CN71               ---V----------V-TQAGE-IT---------KP--EAK-E-VGS-AHTSN-REEGRMAD.......-   1017
LST.CD.88.SIVlhoest447       ---Q----------VKEGD.T-FS-------LVK---EGPKDSEGGLHNN..................-   1020
LST.CD.88.SIVlhoest485       ---QI---------VKEGD.T-FS-------LVK---EGPKDSEGSLDNN..................-   1020
LST.CD.88.SIVlhoest524       ---Q----------VKEGE.N-FSI------LVK---EGPKDSQSSMDN-..................-   1021
LST.KE.x.lho7                ---Q----------VKEGE.N-FS-------LVK---EGPKDSESSLDNN..................-   1020
MAC.US.x.251_BK28            --GE---------ILKVGT.-------------K---GGKEV-SSSHMEDTGEAREVA..........-   1016
MAC.US.x.EMBL_3              --GE---------ILKVGT.-------------K---GGKEV-SSSHMEDTGEAREVA..........-   1014
MND-1.GA.x.MNDGB1            --GI----------LKYQE.E--I-----C---K---E.SGKNSQ-NLESV.................-   1010
MND-2.CM.98.CM16             ---E-I---------KEGT.-L----------------TVDS-PN-EA....................-   1011
MND-2.GA.x.M14               ---E-I---------KEGT.-L----------------TVDSNPHMED--ETA...............-   1017
MND-2.x.x.5440               ---E-I---------KEGV.-L-----------K----TVDSNPHMGR....................-   1012
MNE.US.x.MNE027              --GE---------ILKVGT.-------------K---GGKEV-SSSHMEDTGEAREVA..........-   1016
MON.CM.99.L1_99CML1          ---R-----------KTEEG--LT---------KP--TTENVGGDTNQYNLRKQDGLAN.........-   1018
MON.NG.x.NG1                 ---H--------L--KTDQGEVIT----------P--SKEDVGSKPSAH-IREVDGMAD.........-   1021
MUS-1.CM.01.CM1239           ---A----------VKTEQGQVIT---------KP--AKENVGSKSNTGDHRKEDGLDN.........-   1012
MUS-1.CM.01.SIVmus_01CM1085  ---V---R-------KTD-GEVIT---------KP--AKEDVGSKSDTGDLRKEDRLDN.........-   1012
MUS-2.CM.01.CM1246           R--R--------L--KTKEGE-VT---------K---TRQNV-SEPD-VHVRKEDGLAD.........-   1017
MUS-2.CM.01.CM2500           ---Q--------I--KTDQGE-IT---------K---TKENV-SDPNPVY-RKEDGLAD.........-   1014
MUS-3.GA.09.09GabOI81        ---T-----------KTDLGEV-T---------KP-NA-KNVGSKSNTGDIREEDGLVDQGN......-   1016
MUS-3.GA.11.11GabPts02       ---A-----------KTDLGE-LT-------V-KP--PKKDVGGIPD-GD-GKTPGLVD.........-   1016
OLC.CI.97.97CI12             ---E---------LLKEGD.KYFS--L-----VKP--VDSSANV........................-    995
RCM.CM.00.SIVagi_00CM312     ---T-I-----------GQ.-L------C--VK---RKNVDSETNMEGGQ-EN..............-   1018
RCM.CM.02.SIVrcm_02CM8081    ---T-I-----------GQ.-L------C---K---RKDVDSETSMEGRQ-KD..............-   1024
RCM.GA.x.SIVRCMGAB1          -----I----------EETG-L-----------KE--RKDVDSEANLAGRQ-EN..............-   1019
RCM.NG.x.NG411               -----I-E---------QE.-L------C---K---RKEVDRETNMEGRQ-ES..............-   1023
SAB.SN.x.SAB1                -----I----------EQG.EL-TI--------K----..-L-SQAPLEGNGRTAGEVD.........-   1040
SMM.CI.79.SIVsmCI2           ---N---------I-KVGT.----I--------K---GGKELGNSPYLENP-EDGKMAQPD.......-   1028
SMM.LR.89.SIVsmLIB1          --GE---------ILKVGT.-------------K---GGKEV-SGSHLEDTREIREMA..........-   1010
SMM.SL.92.SIVsmSL92A         --GE---------I-KVG-.----I--------K---GGKEM-SSADMENTTQVRETAQLSEISKV..-   1023
SMM.SL.92.SL92B              ---E-----------KVGT.E------------K---GRQEMGSSASMEDQ.................-   1006
STM.US.89.STM_37_16          --GE---------IVKVGT.-------------K---GGKEV-SGSHLEDTREAGEVA..........-   1016
SUN.GA.98.L14                ---------------KQGE.N-L--------LVK---GESSSVEM-G.....................-   1017
SYK.KE.x.KE51                ---S------------TPDQQVIA---------SSD-ERVDSGTHLEITSKSN...............-   1041
SYK.KE.x.SYK173_COMGNM       ---R----------V-TEEG--FA---------T-H-ER-DSGSH-ENDPKT-...............-   1037
TAL.CM.00.266                ---H-I--------LRTDEGEVIT---------KP--QALGNKIDLE-SKEQ-AEMGRDN........-   1020
TAL.CM.01.8023               ---H-V--------LRTDEGEVIT---------KP--QA-GNKTDLEGSKEQ-AEMGRDN........-   1021
TAN.UG.x.SIVagmTAN1          ---T-----------KEGE.EL-----------KE--ERKTVG-KTNMEG..................-   1020
VER.DE.x.AGM3                --GQ-I---------KGGV.EL-EY--------K--EPRKRMG-ESNLEGAGGADN............-   1023
VER.KE.x.9063                ---Q-I---------KEGE.EL-I---------K--EPRKRMG-EGNMEGLRGPDNQMARNSQILDD.-   1034
VER.KE.x.AGM155              ---R-I--------LKEGE.EL-----------K--EPRKTLG-ETHLEGAGGSDHQMAG.DS.....-   1032
VER.KE.x.TYO1_patent         ---Q-I--------LK.DGS-L-----------K--EPKQRVGNEGDVEGTRGSDN............-   1026
WRC.CI.97.97CI14             --G--P------L--ETPE.G-IT-------L-KVWNGEGMDRSSSDKN...................-   1023
WRC.CI.98.98CI04             --G---------L-VETPE.G-IT--K------KVWNGEGMDRSSSNKD...................-   1022
WRC.GM.05.Pbt_05GM_X02       --G--V------L-V-TPEG-LLT---------KTWDGKGM#..........................-   1007
```

```
                             Vif start
H1B.FR.83.HXB2               ME................NR.WQVMIVWQVDRM..RIRTWKSLVKHHMYV....SGKARGWFYRHHYESP.HPRISSEVHIPLG............DARLVITTYWGLHTGERDWHLGQG.....VSIEWRKKR...............................YS  95
H1A1.UG.85.U455_U455A        --.................--.------------..K----N---------....-K--Q---------R.-S-V---------------------.E----VR---------K-----H-.....------L--...............................--  95
H1C.ET.86.ETH2220            --.................--.---L--------..K----N-------HI....-RR-N--V-----D-R.--KV--------------------.E---I-K-----Q---------H-.....------LRS...............................-N  95
H1D.CD.84.84ZR085            --.................--.------------..--N-------Y--HI....-K--K-------D--.--K-----------------------.------V-----------E-------.....------R-...............................--  95
H1F1.BE.93.VI850             --.................--.--L---------..--N-------Y-------....-K--K--S----FQ-R.---V--------E..........EVK----------P---E-------.....------QGK..............................-R  95
H1G.SE.93.SE6165_G6165       --.................--.------------..------H-----------....-K--------P--A-R.---V------------------.--T--V----------K--Q--H-.....------QR-...............................-R  95
H1H.CF.90.056                --.................--.------------..--N-------Y--HI....-R-----------F--T.-----------------------.E-----------N----E-------.....------L--...............................--  95
H1J.SE.93.SE9280_7887        --.................--.------------..--N-------Y--N-....-K---K-L-----D-N.--K---------------------.E-I---------Q-----------------.....------QR-..............................-R  95
H1K.CM.96.96CM_MP535         --.................--.------------..--N-------Y--HI....-K--NR-Y-------R.--K---------------------.--E--V-----L----------------.....------L--...............................-R  95
H1O.BE.87.ANT70              --.................--.---L-------Q..KVKA-N----Y-K-R....-R-TEN-W-------R.N--V--S-Y--V-............V-HV-V------MP----E---H-.....------Y-K..............................-K  95
H1O.CM.94.BCF06              --.................--.---L-------Q..-VKA-N-----K-M....-K--NK-Q------CR.N-KV--V-Y--VA...........E-DIMV------MP---EE---H-.....------QY-K..............................-R  95
H1O.FR.92.VAU                --.................--.---L-------Q..KVKA-N----Y-K-R....-R--TD-Y-K-----I.N--V--R-Y--V-............P-CIMV------MP---EEQ--H-.....-----QCRK..............................-I  95
H1N.CM.02.DJ00131            --.................--.----V------R..--K--H-----------....-R--E--Y------TN.--K---------V-............L-----------T----S----H-.....-----KLGK..............................-K  95
H1N.CM.04.04CM_1131_03       --.................--.----V--------..K--I-N-------H-....-K--K--Y------TR.--K---------V-............Q-----V-----T---QP----H-.....------LRK..............................-K  95
H1N.CM.95.YBF30              --.................--.----V--------..K--K-N----------....-K--K--Y------TH.--K---------V-............Q----TV-----T---QS----H-.....------LRK..............................-K  95
H1P.CM.06.U14788             --.................--.--------IP-E..KVKA-N----Y-K-R....-K-T-K--------TT.------A-Y--V-............T-TIIV------MP---EEQ--H-.....A-V---QGK..............................-T  95
H1P.FR.06.RBF168             --.................--.--------IS-E..KVKA-N----Y-K-R....-K-TKN--------TV.------G-Y--V-............A-II-V------MP--KEEQ--H-.....X-V---QGK..............................-T  95
CPZ.CD.06.BF1167             --.................--.---RV---I----..---A-N------I-K....-KHCKE-K-----DID.---KGG-----I.TE..........K-K----I-----C--KA----H-.....A-----QG-..............................-W  96
CPZ.CD.90.ANT                I-.................K-.*KIEMI------..--NI--------IWE....TKVLKP-K------ND.--KKGE------PTL...........-KK--V-VF---QC---P----H-.....------CGK..............................-I  96
CPZ.CM.01.SIVcpzCAM13        --.................-K.------------..--KA-N----Y---K....TK--KD--------H-.N--V-------FKD...........NSK-I-----A-T----A----H-.....----Q--MGS..............................-V  96
CPZ.CM.05.LB715              --.................--.------------..-VKA-N----------....-K--KN---------T.S-KV---I---I-............--K---V-----QP--------H-.....------QR-...............................--  95
CPZ.GA.88.SIVcpzGAB2         --.................S-.-------L----..-----N----F---K....-K-----------D--.N-KV-------IR............--KI-VC---N-MP--QP----H-.....------QRT..............................-R  95
CPZ.US.85.US_Marilyn         --.................--.---L--------..-VKA-N----Y---R....-K--K---------T.N--VN------ID............K-KI--I-----K---QA----H-.....------QGT..............................-K  95
GOR.CM.04.SIVgorCP684con     --.................--.--------IP-E..-VKA-N----Y-K-R....TKQ-KQ--------AI.N--V--G-Y--V-............P-EIIV------MP--KEEQ--N-.....-G---QGK..............................-I  95
GOR.CM.07.SIVgor2139_287     --.................--.--------IP-E..-VKA-N----Y-K-R....TKQ-KQ--------AV.N--V--G-Y--V-............P-EIIV------MP--KEEQ--H-.....-G---QGK..............................-I  95
GOR.CM.07.SIVgorCP2135con    --.................--.--------IP-E..-VKA-N----Y-K-R....TKQ-KQ--------AV.N--V--G-Y--V-............P-EIIV------MP--KEEQ--H-.....-G---QGK..............................-I  95
GOR.CM.12.SIVgor_BPID1       --.................--.--------IP-E..-VKA-N----Y-K-R....-KE-KN--------AT.N--V--G-Y--V-............T-TIIV------MP---KEQ--H-.....A-V---QG-..............................-I  95
GOR.CM.12.SIVgor_BQID2       --.................-K.---L----X--Q..KVKA-N----Y-K-R....-K--KD--------AT.N--V--G-Y--V-............G-CIIV------MPX-XAEQ--N-.....-----QLRN..............................-R  95
GOR.CM.13.SIVgor_BPID15      --.................--.--------IP-E..KVKA-N----Y-K-R....-KR-KN--------AV.N--V--G-Y--V-............T-MIIV------MP--KEDQ--H-.....A-VV-SQGK..............................--  95
                             Vif start
MAC.US.x.239                 --................EEKR-IAVPT-RIPER...LER-H--I-YLK-K....TKDLQKVC-VP-FKVGWAWWTC-R-IF--QE...........GSH-EVQG--H-TP.-KG-LSTYA.....-R-T-YS-N..............................FW  97
H2A.DE.x.BEN                 --................ED-N-I-VPT-R-PGR...MEK-HA---YLK-R....TKDLEEVR-VP-HKVGWAWWTC-R-IF--QG...........KSH-E-QA--N-TP.-KG-LSSHA.....-RLT-YTEK..............................FW  97
H2B.CI.x.EHO                 --................EEKN-IAVPT-RIPCR...LER-H--I-YLK-R....TKDLQQVS-VP-HKVGWAWWTC-R-IF--KE...........G-H-EVQG--N-TP.--GFLSSYA.....-RLT-YERS..............................FY  97
H2G.CI.92.Abt96              --................EEKR-IAVPT-RIPGR...LEK-H--I-FLKFK....TKDLQKAV-VP-HKVGWAWWTC-R-IF--TK...........E-H-E-QG--N-TP.-KG-LSQYA.....-RLT-YTRK..............................FY  97
H2U.CI.07.07IC_TNP3          --................EEKR-I-VPT-R-PGR...MER-H-----YLKHR....TKELSKVC-VP-HKVGWAWWTC-R-IF--EG...........ESH-E-QA--N-TP.-KG-LSTYA.....-R-T-YTRG..............................FW  97
H2U.FR.96.12034              --................EEKN-I-VPT-R-P-R...LER-H--I--LSIT....#KELQKVC-VP-HKVGWAWWTC-R-IF--KK...........GSH-EVQA--N-TP.-KG-LSSYA.....-RLT-YTRG..............................FW  96
ASC.UG.10.RT03               -A................PAKM-VIRR--IALER..--QA-I--T-YLK-....KTKELEEVQWI---QLYA-RYTQNKAI---TLKVSQDYV....EATAIEVQIL-D-SS.C-SRDPPGTRDTQALT-S-TYT-L..................TKGDREIDRIS-E 123
ASC.UG.10.RT08               -A................PTKM-VIRRI-IAPER..--QA-I--T--LK-....KTKELEEVQWI---QLYA-RYTQNKAI---TLKVSQDYV....EATAIEVQIL-E-SS.C-TRNPPGVRDTQALT-S-TYT-L..................NKGDREIDRIS-E 123
ASC.UG.10.RT11               -A................PAKV-VIRRI-IALER..--QA-I--T-YLK-....KTKELEEVQWI---QLYA-RYTQNK-I---TLKVSQDYV....EATAIEVQIL-E-SS.C-SRDPPGVRDTQGLT-S-TYT-L..................HKGDREIDRIS-E 123
COL.CM.x.CGU1                --..................KM-A-R-CYPMGKR..KPKPLREWI-.......KQFP..ECQY-YLGKTESPEWWTEGKIEFY--...........SAIK-S-IYLGT-TP.D-.Y-R-PP.......--Y-LQIG..............................KWE  90
COL.UG.10.BWC01              --..................KE-S-RLCYPMSKK..KQEKLRKWI-.......QQFP..NCRYLYKGKTEAPEWWTEGR-VV--...........EAIK-H-IML-H-G-.D-.Y-R-TA.......--Y-VYIG..............................GWC  90
COL.UG.10.BWC07              --................QTKE-S-R-CYPMG-K..NQEKLRLWI-.......KNFP..ECTYHYKGKTESPEWWTEGRIV-R--...........DSIK-TVIML-H-S-.D-.F-R-SA.......--Y-LQIG..............................KWE  92
DEB.CM.04.SIVdeb04CMPF3061   --..................KM-V-RLTHKIPMN..KLDR-C-AI-KGK-....KTKELAPITWQ--F-LRNSYFTQTRIIF-ITK..........EGKV-VELF-C-AP.-KG-LP-MA.....-GLS-E.............................QGNWQ  96
DEB.CM.99.CM40               --..................KT-I-RLTHKIPMN..KLDR-C-AI-REK-....KTKELDTVQWI--F-LRNSYYTQTKIVF-ITK..........EGGV-VELL-C-TP.-KG-LPTMA.....-A-A-E.............................TKTWK  96
DEB.CM.99.CM5                --..................KE-I-RLTHKIPVN..KLDK-C-AI-RGK-....KTKELQAIKWI----LRRAFYTQTKIVF-VTE..........EGKV-VEIF-S-AP.-KG-MPSMA.....-G-S-E.............................QTGWQ  96
DEN.CD.x.CD1_CMO580407       -Q.........NLGRENRTKE-I-RLTYKIGET..KLEH-M-TL-SII-....KTNELK-TVIIP-FRITWEWYTMTKI----R............A-YMEVKL--H-TP.-KGYLSQVA.....I--AYVH............................LSTGWA 105
DRL.DE.11.D3                 --..................KM-I-IPT-R-E-Y..MVER-N--C-Y-KHK...GEKYLER-Y-AP-FQCS.GWWTH-QWT--FKD...........KSKI-V-VL-N-TP.DKG-LSKYA.....IT--HIQEK...............................FC  96
DRL.DE.11.D4                 --..................KI-I-IPT-R-E-Y..MVER-N--C-Y-K-K...GEKYLER-H-AP-FQCS.GWWTH-QWT--FQD...........KSKI-V-VL-N-TP.DKG-LSKYA.....IT--HIQEK...............................FC  96
DRL.DE.11.D5                 --..................KM-I-IPT-R-E-Y..MVER-N--C-Y-K-K...GEKYLEK-H-AP-FQCS.GWWTH-QWT--FQD...........KSKI-V-VL-N-TP.-KG-LSKYA.....IT--HIQEK...............................FC  96
DRL.DE.11.D6                 --..................KM-I-IPT-R-E-Y..MVER-N--C-Y-KHK...GEKYLEK-H-AP-FQCS.GWWTH-QWT--FKD...........KSKI-V-VL-N-TP.-KG-LSKYA.....IT--HIQEK...............................FC  96
DRL.x.x.FAO                  --..................KM-I-IPT-R-E-Y..MVER-N--C-Y-KHK...GEKYLER-H-AP-FQCS.GWWTH-QWT--FKD...........KSKI-V-VL-N-TP.DKG-LSKYA.....IT--HIQEK...............................FC  96
GRV.ET.x.GRI_677_gri_1       --................REKQ-I-RV--R-SER..Q-SR-RGI-TYKIR......N-QLP-E----WQVQWQFWTY-QFI---SK............-DYIEVNI-HN-TP.--G-LSSH-VG...L-YYH.Q-G...............................-K  97
GSN.CM.99.CN166              --..............RQRQKM-V-QPIQIMTER..KVDWLLRAT---IW....SGK..TPFV-V---QLQHQRFTQNKIRLAMDTRKVG.EE...VEATYIE--IL-DTTSHGPASLSRSTYYQQA-I---VYN-S.................KLGAREGDIIW--  124
GSN.CM.99.CN71               --..............RQRQKM-V-QPIQVMTER..KVEWLIRAT---IW....SGK..TPFV-I---QLQHQRFTQNKIRLA-DTRQVG.EE...VEATYIE--IL-DTTSHGPASLSRSTYYQQA-T---VYN-S.................KLGARETDITW--  124
LST.CD.88.SIVlhoest447       --................RDQRI-RVAYITT-S..LVEKFLDIRR.......AHKETQND-VSL-GTGVGWEFYTYNNIV---E...........TGT.--VRC-AH-AA.G-G-ISQWA.....-----IYG...............................N-Q  93
LST.CD.88.SIVlhoest485       --................RDQRI-RVA-ITT-N..L-EKFLDIRR.......AHKETQVD-VSL-GTGVGWEFYTYNKIV---E...........IGT.--VRC-AH-AA.G-G-ISQWA.....-----IYG...............................N-Q  93
LST.CD.88.SIVlhoest524       --................RDQRI-RVA-ITN-S..LVEKFLDIKR.......NH-ESKND-VSL-GTGVGWEFYTYNKLL---D...........TGT.--VRI-AH-AA.G-G-ISQWA.....-----IYG...............................N-Q  93
LST.KE.x.lho7                --................RDQRI-RVA-ITT-S..L-EKFLDI-R.......LNTEAKEN-VGL-GTGVGWEFYTYHKLV---E...........SGT.--VRI-AH-AA.G-G-ISQWA.....-----VYR...............................D-Q  93
MAC.US.x.251_BK28            --................EEKR-IAVPT-RIPER...LER-H--I-YLK-K....TKDLQKVC-VP-FKVGWAWWTC-R-IF--QE...........GSH-EVQG--H-TP.--G-LSTYA.....-R-T-YSRN..............................FW  97
MAC.US.x.EMBL_3              --................EEKR-IAVPT-RIPER...LER-H--I-YLK-K....TKDLQKVC-VP-FKVGWAWWTC-R-IF--QE...........GSH-EVQG--H-TP.--G-PSTYA.....-R-T-YSRD.............................#FW  97
MND-1.GA.x.MNDGB1            --................RVERI-RLT-K-SSQ.---EK-HW--RRQ-AWATANNEEGCW-L-P-FMAYN.EWYTC-K-V-IIN............R-I--IVRS--H-QI.-VGCLSTYA.....----AVVRPPP.........................FEKEWC 106
MND-2.CM.98.CM16             --..................KQ-I-IPT-KL--H..AVER-N----Y-K-K...GEKHLDK-DLKA-FQCS.GWWTH-QKY---EE...........-EQII--IL-N-TP.-KG-LSTYA.....MT--YKFTN...............................-Y  96
MND-2.GA.x.M14               --..................KQ-I-IPT-RI--R..QLEK-N----Y-K-K...EEKHLDK-ELF--FQCS.GWWTH-QKI--FKD...........GSKII--AL-N-TP.-KG-LSQYA.....ITA-YN-GD...............................-Y  96
MND-2.x.x.5440               --..................KQ-IAIPT-EDRQD.VKL-D-N-I--Y-K-K...GEKHLEN-QLYP-FQCS.GWWTH-QKK--FKD...........GSKII--VL-N-TP.-KG-LSQYA.....FTV-YQ-EN...............................-F  97
MNE.US.x.MNE027              --................EEKR-IAVPT-RIPER...LER-H--I-YLK-K....TKDLQKVC-VP-HKVGWAWWTC-R-IF--QE...........ESQ-EVQG--N-TP.--G-LSTYA.....-R-T-YSRN..............................FW  97
MON.CM.99.L1_99CML1          -A..............PR..KM-V-TPTNIISES..KMDWLIRCT-W-IL....TGS..APFV-V---QLHNQRFSQNKIKL--DLGITQEGE...SWATY-E--I--D-TNVGPSALSPTVYNKQAYT---VYWMQKVERVP.....LDPVWNPNPR..DREV.RTW-A 135
MON.NG.x.NG1                 -A..............PR..KM-V-SP-HIRSER..WMEWLIRAT-YNI-....TGK..VPFV-V---QLTHQRFSQCKIKL--EMGITQXXE...QWATYXE--IL-DXTNVGPASLSPSVKWKQAYI---VYW-Q..PRVPQQEGVVDRVWFPSPRRPDREIDRVW-A 141
MUS-1.CM.01.CM1239           -A..............PR..KM-V-SPIQETTER..KMDWIIRAT---IF....SGK..TPFV-V---QVQHQRFSQNKIKLA-TQNTLESGE...TEVTFIE--VL-DVTNVGPASLSKSTYWKQSYIL--VYM-K.................GPRDRERDYVW-H 123
MUS-1.CM.01.SIVmus_01CM1085  -A..............PR..KM-V-SPIQETSER..K-DWIIRAT-YYIL....SGR..AAFT-V---QLYHQRFSQ-KIR---TQNTRESGA...VESTYIEL-VL-DVTNVGPASCSKSTWWKQSFIL--VYL-R.................HPQDREQDYIW--  123
MUS-2.CM.01.CM1246           -A..............PG..KM-I-SPIQYMSER..KMDWLIRCT--YIF....SGK..APFT-V---QLQHQRFTQNKIK---SINRIEGGP...TETTYIE--ILLDVTNVGPASLSRSTYWQQSYILK--YI-A.................APQNREIDMVH-E 123
MUS-2.CM.01.CM2500           -A..............QR..KM-I-TPIQFMTER..-MDWLIRCT---IL....SGK..APFS-I---QVQHQRFTQNQIK---TIKKREEDK...VESTYIQ--IL-DVTNVGPASLSKSTYWQQSYIL--VYV-A.................APQDREVDMVW--  123
MUS-3.GA.09.09GabOI81        --..............PR..KM-V-SPIQETSER..KMDWLIRAT-YLIF....SGX..AP-T-V-A-QLQSQRFSQNKIK---AIQLRDIEGQQWIEATHIE--VL-DVTNVGYASLSRSTQWRQTYIL-FVYQ-R.................QMESRDIDVIW-H 125
MUS-3.GA.11.11GabPts02       -A..............PR..KM-V-SPIQVTTER..HLDWLIRCT-Y-IL....SGQ..TP-S-V-A-QLQHQRFSQNKIR---TQETTEREK...VSTHIE--IL-DVTNVGPASLSKSTYYEQSYIL--VY--N.................RKGGRETDTVW-H 122
OLC.CI.97.97CI12             --......................-IAVPMYRLNNGQVEHVKFMAKLAYKFWG........YH-IWQNC-KY-REWWTHGKIY--VPG............GEIE--L--HIRH.GKGLQNSG-.....-NFKYIVGIS..............................W  91
RCM.CM.00.SIVagi_00CM312     -A................EKM-I-RPI-R---R..K-EQ-H----Y-Q-K...GKKE-KE-E-VP-FKV-WGWWSH--I-V---K...........NTKVK-----N-TA.-KG-LSTY-.....-G-AYIDNTCD..........................P..C-F 102
RCM.CM.02.SIVrcm_02CM8081    -A................EKM-I-RP--R---R..K-EQ-H----Y---K...GKKE--E-E-VP-FKV-WGWWSH--------N...........NTKIKV----N-T-.-KG-LGTY-.....AALAYIDQKCD..........................P..P-F 102
RCM.GA.x.SIVRCMGAB1          -A................EKM-I-RPI-R---R..K-EQ-H----Y-Q-K...GKKA-KE-E-VP-FKV-WGWWSH-------EE...........GSK-K-----N-TV.-KG-LGTY-.....-G-L-I-G...............................D-V  98
RCM.NG.x.NG411               -A................EKR-IERPI-K---R..KVEQ-H----Y-Q-K...GKKE-KD-E-VP-FKV-WGWWSR--------N...........NTKIKV----S-TI.-KG-LGTY-.....-G-AYIDN-CD..........................P..C-F 102
SAB.SN.x.SAB1                --..................KH-I-RPL-K-TGG..QQER-T----Y--H-....-KQCVH-R-TP-TKIRWNWYSYQ-WV---KD...........G-LIKV-N--H-TP.-KG-LETYATG...IGYS...-G..............................EWF  96
SMM.CI.79.SIVsmCI2           --................EEKS-VTVPT-RIP-R...MER-H--I--LKFN....TKELQKVC-VP-NKVGWAWWTC-R-IF--EK...........ESH-EVQG--N-TP.-KG-LSSYA.....-R-T-YSRN..............................FW  97
SMM.LR.89.SIVsmLIB1          --................EEKR-I-VPT-KIPGR...LER-H--I--LK-N....TKELNKAC-VP-HKVGWAWWTC-R-IF--QG...........E-Y-EVQG--N-TP.-KG-LSTYA.....-RLT-YSRN..............................FW  97
SMM.SL.92.SIVsmSL92A         --................EEKR-IAVPT-KIPHR...LERRH----YLK-K....TEDLKKVV-VP-HKVGWAWWTC-R-IF--EK...........E-H-EVQG--N-TP.-KG-LSCYA.....-RLT-YTRN..............................FW  97
SMM.SL.92.SL92B              --................EGKR-VAVP--RIS-RI..VER-H-CI-F-K-K....TRELEKAC-VP-HKVGWAWYTA-R-IF--EE...........GSH-EVQV--N-TP.-KG-LSSYA.....-R-T-YSEK..............................FW  98
STM.US.89.STM_37_16          --................EEKR-I-VPT-RIPGR...LER-H--I--LK-N....TKELSKAC-VP-HKVGWAWWTC-R-IF--QG...........E-H-EVQG--N-TP.-KG-LSEYA.....-R-T-YTRN..............................FW  97
SUN.GA.98.L14                --ENPPQWRWWDEREWEDRRQYKI-R---LI--I..AVEKFLD-RR.......MHRETKDD-VSMYGTGTGWEWYTYNKII--VT...........YGT.V-VRI-GH-TP.AKGLINQW-.....C-M--IYN..............................S-Q 110
SYK.KE.x.KE51                --..................KE-I-VLT-KLHPN..Q-DKMQHIL-V-K-K....-KQLEKAT-K---QITWEWWTRAQWE--V-...........QGV-I-KF-HN-TP.-KGLLR..TEG...-GLI-AHPS.............................GW-  96
SYK.KE.x.SYK173_COMGNM       --..................KE-I-VPT-RMTPR..Q-DRLQHII-T-K-K....-KELEKAT-K---QIEWQWYTYCQWT--V-...........-GTIW--F-HN-AP.--G-LH..MQG...IR-QYQWN.............................QWN  95
TAL.CM.00.266                -D..................KL-VTRLT-K-P-N..-MQR-VGTIRF-IH....KTKYLQKLIWV--FQLRHASYTQAKWIL--EEKDTR........VEY-T--Q--C-APDHGYTSDWDT.....I----SVSDI.....................VRDEKV-A 108
TAL.CM.01.8023               -D..................KL-VTRLT-K-P-N..-MQR-VGTIRY---....KTKYLQKLIWV--FQLRMASFTQAKWIL--EEKENR........VEYIT--Q--C-APDKGYTSDWDT.....I----SVSDV.....................TLDAKV-- 108
TAN.UG.x.SIVagmTAN1          --................REKL-VTRLT-R-SGE..H-DK--GI--Y--R......NRLQD-T-LM--QCGWAWYTC-RFL----G...........EGKI-VDC--H-TP.-QG-LSTYAVA...I-F-N.WQN.............................T-K  98
VER.DE.x.AGM3                -N................QEKE-VMRVT-K-PEE..L-TK-QGI-RYW-R......TRKLD-K--M--QITWAWYTM-RYE----Q...........HGSIHVDL--H-TP.-KG-LSTYAEG...IQYLS.NRDP............................W-R  99
VER.KE.x.9063                -N................QEKE-VMRVT-K-SGD..L-TK-QGI--YW-T......KRNLK-Q-MM--QINWAWYTMCRYI--I-L...........EGEIQ-DC--H-TP.--G-LSTYAIG...IQYLG.AVG.............................NFR  98
VER.KE.x.AGM155              -S................QEKH-VMRLT-K-QEE..V-TK-QGI-RYW-N......KRNLK-E-KM--QITWEWYTM-RYV---PG...........SGEIHVDI--H-AP.KQG-LSTYAVG...IQYVS.LVND............................K-R  99
VER.KE.x.TYO1_patent         -N................P-KE-VMRVT-K-PGD..L-TK-QGI-RYW-R......QRNLK-N-YM--QITWAWYTM-RYV--I-K...........HGEICVDL--H-TP.-QG-LSTYAVG...IQYVS.NLES............................K-R  99
WRC.CI.97.97CI14             --..................KE-IGVPLIRISKG..KF-R-TYYIA.......NLIKAGQAN-LWGSWHAAAECFTRHI-K--IQ...........GHWEMQ-RC--N-SP.-KG-MDTM-.....FN-RMVYEN.............................RWE  94
WRC.CI.98.98CI04             --..................KE-IGVPLIKISKG..-FKR-AYWIA.......HLMQNKKAS-LWGTWHTQAECFTRYLLR---Q...........GHWE-Q-RC--N-AP.-KG-LDTM-.....FN-RMVYEN.............................RWE  94
WRC.GM.05.Pbt_05GM_X02       --..................KE-IAVPLIKISKG..-F-R-SYWIA.......ELIQEKQAR-LFGSWHVHAECFTRHI-R---Q...........GHWEMQ-RC--N-AP.-IG-IDTL-.....FN-RMVWEN.............................KWE  94
```

```
                                                                                                                                                                      Vif end
H1B.FR.83.HXB2              TQVDPELADQLIHLYYFDCFSDSAIRKALLGH..IVSPRCEYQA.GHNK........VGSLQYLALAALITPK..........KIKPPLPS............VTKLTEDR.......WNKPQKTKGHRGS.HTMNGH...............*  192
H1A1.UG.85.U455_U455A       -----D---H----H------E----R-I--Q..--R-------.----........---------K--V--T..........RA------............-K------.......-------R-----.R-L-R-...............-  193
H1C.ET.86.ETH2220           -----G---H---MH-----AE-----I--Y..R-----D---.----........---------T---K--..........-A------............-S--V--K.......-------R-R--N.------...............-  193
H1D.CD.84.84ZR085           -----G-------M------A-------I---..-----------.----........---------T---A--..........-R------............-K------.......----RQ---R---.------...............-  193
H1F1.BE.93.VI850            --I--G-------I------E------I---..RI---N----.----........---------T---A-E..........-T------............-Q--V---.......-----E-R-----.------...............-  193
H1G.SE.93.SE6165_G6165      ---E-D---H----H-------------I--Q..-----------.---Q........---------KV-V-S-..........RSR-----............--E-A---.......-------R---EN.P-----...............-  193
H1H.CF.90.056               ---E-G-------MH-----E------I--R..V-R--N-P-.--KQ........--T------T--VA--..........--------............-R--V---.......-------R-----.------...............-  193
H1J.SE.93.SE9280_7887       -----G-------MC-------------I--Q..------D---.----........---------T---R--..........RR------............-Q--V---.......-------T---E-.------...............-  193
H1K.CM.96.96CM_MP535        ---E-D-------I------E--V---I---..R------C--.----........---------T--VA-R..........RP---V--............-K--V---.......-------R-----.Q-----...............-  193
H1O.BE.87.ANT70             --I---T--RM---H--T--TA--V---I--Q..R-LTK---PT.--SQ........--T--L---R-VVKAR..........SR------............-Q------.......---HLRIRDQLK-.PS----...............-  193
H1O.CM.94.BCF06             --I---T--RM---H--N--TE----R-I--Q..R-LT----S-.--SQ........--T--F---K--VKV-..........RN------............-Q---N--.......-------IR-QL--.-S--E-...............-  193
H1O.FR.92.VAU               --I---T--K----H--T--TE----R-T--Q..R-L-Q---PE.--SQ........--T--L---R-VVKA-..........RNR-----............-Q------.......--RHLRI-DQP--.QL-H--...............-  193
H1N.CM.02.DJO0131           ------M--K-------N--TA----Q-V--R..P-F-----P-.---Q........--T--H---T-WVGV-..........-R------............--------.......---H--Q---N.PI----...............-  193
H1N.CM.04.04CM_1131_03      ------M--K---------TA----Q-I--R..Q-F-----P-.---Q........--T------T-WVGI-..........-R------............-A------.......---H--Q---N.P-----...............-  193
H1N.CM.95.YBF30             ------M--K----H----TA----Q-V--R..P-L-----P-.--KQ........--T------T-WVGA-..........-R------............--------.......--EH--MQ---N.PI----...............-  193
H1P.CM.06.U14788            --I---T--R----H--Q------V-R-I--D..R-LNX---S-.--SQ........---------KVVVGKV..........-R------............-QI--Q-I.......-SN--R---XQE-.-S-S-C............YKP-  196
H1P.FR.06.RBF168            --I---T--R----H--Q------V-R-I--D..R-LNK---S-.--SQ........-----F---KVVVGKV..........-R------............-QI--K-I.......-S--R---RQE-.-L-S-C............YKP-  196
CPZ.CD.06.BF1167            -E----T---I--SR--P--TER---Q-I--E..KILQH-HFPT.--S-........-----F--FKKVLEQN..KQ.....KPHR----............--V-A---.......-----RI-A-QEN.RGI--Y...............-  197
CPZ.CD.90.ANT               ------T---M--Q---P----Q-V-Q-I--E..RILTY-H-KK.--SQ........--T-----FCKILEFRG.YP.....KGPRRQF--............LSI-----.......----RRMR---EN.Q*-S-C...............-  197
CPZ.CM.01.SIVcpzCAM13       -----YV--K---SQ-----A-----R-I---..--E---A-RE.--RQ........--T--F---K--VKE-..........RLR-----............-A------.......--T--R---PEN.---S--...............-  194
CPZ.CM.05.LB715             --I------H----H--N--AE-----------..----K---P-.--L-........--------IS--FKSN..........PKR-----............-S------.......-----R-----E-.------...............-  193
CPZ.GA.88.SIVcpzGAB2        --I--QT--R---C------AA--V-R-I--E..--TKV--FPR.--PQ........-Q------IVQVVKT-..........-K------............--I-A---.......-------RDQQEN.L--S--...............-  193
CPZ.US.85.US_Marilyn        --I---M--K----------TA------V--R..P-I-K---PT.----........-----F---K--VGQS..........-RR-----............--------.......---C-----QPEN.---S--...............-  193
GOR.CM.04.SIVgorCP684con    --L---T--RI------Q--TE--V---IV-E..KLLCK-N-P-.--SQ........--T------R--VGRV..........RVR--F--............-Q--AA-T.......-SRS-QMR---E-.-S----...............-  193
GOR.CM.07.SIVgor2139_287    --L---T--RI------Q--TE--V---IV-E..KLLYK-N-P-.--SQ........--T------R--VGKV..........RVR--F--............-Q--AA-T.......--RS-QMR--KEN.-S----...............-  193
GOR.CM.07.SIVgorCP2135con   --L---T--RI------Q--TE--V---IV-E..KLLCK-N-P-.--SQ........--T------R--VGKV..........RVR--F--............-Q--AA-T.......-SRS-QMR--KEN.-ST---...............-  193
GOR.CM.12.SIVgor_BPID1      --I---T--R----H--Q------V---I--D..RLLH----S-.--SQ........---------RV-VGKV..........-RR-----............-QV-----.......-S---RI---QE-.-L-S-C...............-  193
GOR.CM.12.SIVgor_BQID2      --I---T--RM-----XS--TE------I--Q..K-L-T---P-.--SQ........--T--F---RV-VKTR..........-FR-----............-Q--A---.......--RYPRIRDQLE-.-S----.........WKFWKN-  199
GOR.CM.13.SIVgor_BPID15     --I---T--R----H--Q------V---I--D..RLLT----S-.--SQ........---------RV-VGKV..........RR------............-QL----K.......---R--I---Q-N.-L-S-C...............-  193
                                                                                                                                                                      Vif end
MAC.US.x.239                -D-T-NY--I-L-ST--P--TAGEV-R-IR-E..QLLSC-RFPR.A-KY.......Q-P-------KVVSDVRS.........QGEN-TWKQWRRDNRR..GLRMA-QNSRGDKQ...RGG--PTKGANFPGLAKVL-ILA.............-  215
H2A.DE.x.BEN                -D-T-DC--I---ST--S--TAGEV-R-IR-E..KLLSC-N-PQ.A-KA.......Q-P-------VVVQQ.NDR.......PQR-GTARKQWRRDHWR..GLR-AREDHRSLKQ...GGSE-SAPRA-FPGVAKVLEILA.............-  216
H2B.CI.x.EHO                -D-T-DV--R-L-GS--SS-TANEV-R-IR-E..KILSH-N-PS.A-TG.......Q-P---F---RVVQEG-DG.......SQGESTTRKQRRRNSRR..GIRMARDNIRTSQQ...SSSQSLAQGTYFPGLAEVL-ILA.............-  217
H2G.CI.92.Abt96             -D-T--T----L-GS-----TAGEV-R-IR-E..QILSC-N-PT.A-KR.......Q-P---F---QVVQKGHG.......S-GESXTRKQRRGDNRR..GIRMARKNSX-DQQ...DSSQSFTQGAYFPGLAKVL-ILA.............-  216
H2U.CI.07.07IC_TNP3         SD-T-DY----L-GT--S-YTAGEV-R-IR-E..QLLSC-NFPS.A-KG.......Q-P---F---KIVQDVRP.........QR-NTTWKLRRRNSGR..SLRMARVNSHSHKP...RGS-SSAQGTYFPGVAKVL-ILA.............-  215
H2U.FR.96.12034             -D-T-DC----L-GS--S--TAGEV-R-IR-E..QLLSC-NFPS.A-RQ.......Q-P---F---R--QDGRS.........-REN-TRKFRRRNCRR..GFQLARKNCE-YQQ...GSSATSSQGTYFPGLAKVLAILA.............-  214
ASC.UG.10.RT03              -EIT-H---AM--VEH-S--TAQDFTR-IR-Q..QLLGE-HHP..HYLGKE.....IP.T------R-IQD..............GRSTAIT..........PNTLA-QNSGDVPKT....-PGNTMGAK-RGRK-LQ.................  223
ASC.UG.10.RT08              -EIT-Q---AMV-VEH-S--TAQDFTR-IR-Q..QLLGE-HHP..HYLGKE.....IP.T------R-IQD..............GRN-AIT..........PTTLV-QNR.NVPKT....-PGNTMGAK-RGRK-FQ.................  222
ASC.UG.10.RT11              -EIT-Q---AMV-VEH-S--TAQDFNR-IR-Q..QLLGE-HHP..HYLGKE.....VP.T------R-IQD..............GRN-AIT..........PTTLARQNSRDVPKT....-SGNTMGVK-RGRK-LQ.................  223
COL.CM.x.CGU1               -DLSWQQ-VGYM-IIRGK-VRIEEEARKAIRG..LPWNP-DF-V.--LG.........-LK--E-C-L--R.E-ERTH..............................QKMRV.........L-GTSHGCRTR......................-  172
COL.UG.10.BWC01             SDLTWQM-TGYL-IIRGK-IEIEKEVRKAIRG..LPWDP-DF-V.--LG.........-R---E-C-SK-QKDQEREY..............................QKMR-.........W-ETS-GCKSR......................-  173
COL.UG.10.BWC07             -DLTWQ--VGYM-IIRGK-VEIQEEARRA-KG..LPWCP-DF-V.--LG.........-LT--E-C-N-ID.ENK..P...............................-RTRP.........E-GITPGYRSR......................-  172
DEB.CM.04.SIVdeb04CMPF3061  -ELS----EH----R--P--TQQ-V-Q-I--E..KLLTC-Q....QYH.RGAGGN.GPP----IC-RVA-KY.............G-ATALT..........PTNALVKASSNFQAS...ATIHDAGCK-R-HE-LS.................  199
DEB.CM.99.CM40              -ELT-D---H----R--P---QQ-V-Q-IR-E..-LLQQ-Q....QYH.RGPEGK.GPP----I--RVAVKH.............G-V-TLT..........PTTFHLKNNTHDKAR...VAAGHARSK---PK-FH.................  199
DEB.CM.99.CM5               -ELT-D---H----R--P---QQ-VTQ--K-E..KLLQQ-Q....QYH.RGPEGK.GPP-------RTAVKY.............G-VSTLT..........PTTFHIKNSTNDKAS...TTTSHAGSK---FK-LH.................  199
DEN.CD.x.CD1_CMO580407      -E---VT--EI--KDH-S--TRQ-V-Q--R-E..V-LST-ARKKAHYHPRTGPGR.EPP----IV-Q-A-RY.............G-VTTIT..........PTSISFKNGRNHSKA...ATGSSMGY---GKK-LQ.................  213
DRL.DE.11.D3                -FI--TT--RI---E--P--T-Q---GV---Y...RIQA-YFPRG.-KGQ.........----------Q-H-KYLRHGRRTASGESSSGASG-RRG.........R-RRVAAEESLKNQPGGSITLPPRVPLP-LEHLH-ALP.............-  218
DRL.DE.11.D4                -FI--TT--KI---E--P--T-Q---GV---F...RIQA-YFPRG.-KGQ.........----------Q-H-KYLKHGRRTVSGESSSGASG-RRG.........R-RRVATKESLENQPRGPVTLPPRVPFP-LEHLH-ALP.............-  218
DRL.DE.11.D5                -FI--TT--KIL--E--P--T-Q---GV---F...RIQA-YFPRG.-KGQ.........----------Q-H-KYLKHGRRTANGESSSGANG-RRG.........R-RRMAAKEPLKDQPRGPITLPPRVPFP-LEH-H-ALP.............-  218
DRL.DE.11.D6                -FX--TT--RI---E--P--T-Q---GV---Y...RIQA-YFPRG.-KGQ.........----------Q-H-KYLKHGRRTASGESSSGASG-RRG.........R-RRVAAEESLKNQPGGSITLPPRVPLP-LEHLH-ALP.............-  218
DRL.x.x.FAO                 -LI--TT--RI---E--P--T-Q---GV---Y...RIQA-YFPRG.-KGQ.........----------Q-H-KYLKHGRKTVSGESSSGANG-RRG.........R-RRMATEESLKNQPGGSITLPPRVPLP-LEHLH-ALP.............-  218
GRV.ET.x.GRI_677_gri_1      -E---GT--RM------N--T-R--QQ-IR-E...KYTW-TFKE.--KGQ.......-Q---L---V-Y..................TNGIR.................KRSKRTF.........TRMAGNL-S-QGAMGRMATRHAQGSKRRSQKALW.  201
GSN.CM.99.CN166             -NLT-GV-M-I--GK--P--QNED--R-IR-E..QILGN--HPE.A-L.QVG...TPQT-EQ--FF-YVKH............YG-N-TQ-..........SYALAQPKGAHNHGY...RSKRSVGNK--GRE-LL.................  228
GSN.CM.99.CN71              -NLT-DV-M-I--TR--P--QNEDV-R-IR-E..QLLGY-DHPE.A-P.QVG...TPHT-ER--FF-YVRY............YG-N-TK-..........PSTLDQPKGAHNHGS...CSKRHVGNK-RGRK-LL.................  228
LST.CD.88.SIVlhoest447      -EI--I--N-M--CK--N---TRD-----W-E..KIMAF-AMPT.--KG......C-L---F-C-RQ-QHVQ...........N-AAK...............ESRK-SRGLWATK..GAMGSMASR-H-G.NKRG-QTPFP............  197
LST.CD.88.SIVlhoest485      -EI--I----M--CK--N---TRD-----W-E..KIMAI-AMPT.--KG......C-L---F-C-RQ-QHVQ...........S-AAK...............ESRK-SRGLWTTK..RAMGSMVSR-H-G.NKGG-QTPFP............  197
LST.CD.88.SIVlhoest524      -E---I----ML-VK----WVHRD-VR-IR-E..KILHS-KMPT.--KG......C-L-----C-RQ-QHVQ...........T-TAA...............QPRKAPRGLWSSK..RAMGSMASR-N-G.NQGR-KTPFP............  197
LST.KE.x.lho7               -EI--I----M--YK--N--IQRD--R----E..RIT.M-CQPL.--KG......C-L---FIC-RQ-QHVQ...........AEAAK...............KSRE-TRGLWTTK..GAMGSVVSR-H-G.NKRG-KTPFP............  196
MAC.US.x.251_BK28           -D-T-DY--I-L-ST--P--TAGEV-R-IR-E..QLLSC-KFPR.A-RY.......Q-P-------KVVSDVRS.........QGEN-TWKQWRRDNRR..GLRMA-QNSRGDKQ...RGS--PTKGADFPGLAKVL-ILA.............-  215
MAC.US.x.EMBL_3             -D-T-DY-GIFPA*HL-P#-TAGEV-R-IR-E..QLLSC-KFPR.A-RY.......Q-P-------KVVSDVRS.........QGEN-TWKQWRRDNRR..GLRMA-QNSRGDKQ...RGG--PTKGADFPGLAKVL-ILA.............-  213
MND-1.GA.x.MNDGB1           -EIT--V--H----HFY---M---VM--IR-E..E-LKV-RFP-.--KAQ.....G-L---F-C-RVIYG-E...........ER*A......................SIR-SRTSQ..RAIQSVASRYY-G.NKGRSKKALP............  205
MND-2.CM.98.CM16            -HI--QT--RM--WK-LP--TEQ---Q----K...RLTV-YFHWG.-KGK.................LSYTAYCNNGR.....RGPRD-RR-RRG.........GSKY-VGKISGENQSGGKITLPPRVPFP-LER-HRTLA.............-  213
MND-2.GA.x.M14              -HI--VT--RM--WE--P--TAA-V--V-Y-E...RIVA-YSPWG.-KGQ.......--T--L---R-Y-KFCRNGR.....KSTRGTTGGRRS.........GTRTVAGKIIGTSEQRGSITLPPRVPFP-LEHLCRTLA.............-  213
MND-2.x.x.5440              -YI--VT--RI--GE--P--T-Q------F-E...RLVA-YFPWG.-RGQ.......--T--F---Q-YLRGRKHGR.....EGTRGARR-RRG.........RTGAMAGNVTGENQPGGPVTLPPRVPFP-LEH-CRTLA.............-  214
MNE.US.x.MNE027             -D-T-DY--I-L-ST--P--TAGEV-R-IR-E..QLLSC-RFPR.A-KN.......Q-P-------RVVSYVRS.........QREN-TWKQWRRDNRR..SLRMA-QNSRGDKQ...RGS--PTKGADFPGLAKVL-ILA.............-  215
MON.CM.99.L1_99CML1         -HLT-D--N-IV-TH--A--QQLDV-R-IR-E..KLLGQ-QHL-.T-YPKVG...TPLT-EK--FL--VR............YGAT-TK-..........SAPLDLPDGSHRKTG...PTGRHLGAK-R-PK-LL.................  239
MON.NG.x.NG1                -YLT-DV---I--TR--S--QA-DV-R-IR-E..AXLGN-HHIQ.T-YNKVG...TPPT-EK--FL-CFK............YG-CTTEP..........SSTLDLPRGSHRARR...SAGRHMGPK--GPFALL.................  245
MUS-1.CM.01.CM1239          -YLT--I-MKI--THH-S--YSQD-PRVIR-Q..PPLGD--HPG.A-T.KVG...PPPT--V---LT-KKH............G-STTQP..........QTSMALQSGPNYHAA...GPTGHVGPK-RGRK-LF.................  226
MUS-1.CM.01.SIVmus_01CM1085 -YLS-T--MRI--TQ--S--YSQD-PRVIR-Q..HTLGD--HPD.A-T.KVG...PPP---V---L--KKH............G-LAPQP..........PSALALQSGPNHHAP...GTRSHVGSK-RGRK-LF.................  226
MUS-2.CM.01.CM1246          -FL---VTM-I--TH--S--QQRD-QR-IR-E..QLLRK--HIK.T-YPKVG...TPL--EK--FF-Y-KHT..........NGASASK-..........PSLMA-SSNANNYGS...GASRPVGTK-RGRK-LL.................  229
MUS-2.CM.01.CM2500          -YL-----M-I--T-H-S--QEQD-IR--R-E..RILGI-QH-T.A-LPKVG...TPP--EK--FF-Y-KQG..........DGAT-SK-..........PSPMA-PPGTNNHGA...GTKRHVGSK-RGRKALL.................  229
MUS-3.GA.09.09GabOI81       -YLTNX--M-I--C---P--QEED--R-IR-E..QLLGH--HPK.A-FPKVG...TPL---Q--FFQYVKS............QR-N-TV-..........PSTLALQASANHHAT...GTRSHVG-K-RGRK-LL.................  230
MUS-3.GA.11.11GabPts02      -HLTVG--MRM--YK--S--QEED--R-IR-E..QLLGH-QH-Q.A-FSKVG...TPLT-EK--FL-Y-KH............YG-STPLT..........YSPMALQACSHNHAT...GSGGHVG-K-RGRK-LL.................  227
OLC.CI.97.97CI12            MYLT--YV-FMV--V--A-WTQ------I--MP..DHRY-QLET.--LG.........-L---L---R-VRNGR.............-SSRG-..............GA-QGATGG.......LGERHNGRNYCRGKKALS..............  186
RCM.CM.00.SIVagi_00CM312    -DI--I---KM--NV--P--T-Q--------E...KVLV-GF-RG.-RDQ.......--T------Q-WVKSQLQ...KNGRKSSRG-HWGGRS.........R-STMVGQNVVRDQPRSQITLPWRV-FP-VAYLC-ILA.............-  221
RCM.CM.02.SIVrcm_02CM8081   -DI--IV--S---KI--P--T-K---Q-I--E...KVLL-GF-RG.-RDQ.......--T-----IQ-WAREQVK...KHGRKSARG-HWGWRS.........R-PA-VGQNAVRNKSGSQVTLPSRV-FP-LAYLC-TLA.............-  221
RCM.GA.x.SIVRCMGAB1         -D-F-WT--S---KI--P--T-R---QS---E...KVLV-AF-GG.-RDQ.......--T--F--IQ-WAKSQLD...RYGRKSPRG-HWGWRS.........R-PA-ATGHARKGQLGSQVTLSSRV-FP-VAHLC-TLA.............-  217
RCM.NG.x.NG411              -DI--TV--K-V-NQ--P--T-K---Q-I--E...QVLL-GFEKG.-RDQ.......-----------IK-WARQQLKDG.RGQRKSARG-HWGWRC.........R-CP-ASQNANRSQLRSQVALSSRV-FP-VAYLC-TLA.............-  223
SAB.SN.x.SAB1               -EL--WT--HI--WS--P--T-R-VQQ-IR-E...KYLW-KH-V.--QPT....GQ-P-------RVY..................TNG-RRVAPTSRRGSSQGSPQESQRRD........TRMARNM-FAQRAVRRMAPRHVTGPQFRGPVPLP.  217
SMM.CI.79.SIVsmCI2          -D-T-DV------ST--S-YTQGEV-R-IR-E..QILSC-K-PS.A-KR.......Q-P---F---RI-HKQNGK.......C-REN-TRNQWGRNNRR..GFR-AIQHSGSSKQ...GSS-TSAQGADFPGLAKVL-ILA.............-  217
SMM.LR.89.SIVsmLIB1         -D-T-DY--V-L-GT--P---EGEV-R-IR-E..QLLSC-RFPK.A-KY.......Q-P---F---KVVEHVRS.........QREN-TWKQWRRNYRR..GLR-ARKNSG-NKQ...TGSE-STKGIDFPGLAKVL-ILA.............-  215
SMM.SL.92.SIVsmSL92A        -D-T-D---I---GA--S-YTQGEV-R-IR-E..QILSC-NFPS.A-QY.......K-S---F---R--Q...DG.......S-RTD-SREQWGRDNRR..GLRLARKNSR-SQQ...NSSE-FAKGA-FPGLAKVL-ILA.............-  214
SMM.SL.92.SL92B             -D-T-DV----A-ST--P--AAH-V-Q-IR-E..Q-LSY-G-AV.A-HS.......S-Q---L---KVVLQNDR........P-G-N-TRKQWRRNNRR..GLRMVTQYSGGPES...NSSTTPTKRVNFPGLEKVL-ILGR............-  218
STM.US.89.STM_37_16         SD-T-DC----L-GT--P--TAGEV-R-IR-E..KLLSC-RFTK.A-KN.......Q-P-------KVVEHVRS.........QRENTARKQWRRGNRG..SIR-ATQNGRGHKP...RGS--STEGTDFPGLAKVL-ILA.............-  215
SUN.GA.98.L14               -EI--LV---M--FK----WTSRC--R-M--E..KILHE-RN-V.A-KG.......L-L---F-C-RV-HGQQ............ERAS....................RTATR-RGATE..RAICTMAGRYH-R.NQGRSGKAFP............  211
SYK.KE.x.KE51               -E-T-QT--A-V-SQ--P----R-VQQ--R-E...KLTSHCWNF..-KEQ.......-L-------QKYLSKD............GTGFLQS............LPAAARGT........TILHSK-CQL-RGRSNKC-CQSRTGSPRSMQAFY.  206
SYK.KE.x.SYK173_COMGNM      -DLT-AV--R---NF--P--TAR-VNQ-VR-E...LLTSHCWTP..-TDQ.......-P-------QVYL.KD............GGGFLQS............LPACARNT........MVLHSK-CRVDP.KRDQC-CKGRTGSDRSIQAFY.  203
TAL.CM.00.266               -E-T-NV--H---TK--P--TA--V-Q-IR-E..VLLAV-G.....VRHKGRD..TPP---F---T-AVRY............G-VATQ-..........SSTADLQ-HSNTKAS...ARKAHERSS-RN..AYTL-...............  209
TAL.CM.01.8023              -ELT--V--H---TK--P--TA--V-Q-IR-E..VLLAV--.....VRHKGRD..SPP---F---KVAVRH............G-VATQ-..........SSANNIKAHTNTKSS...IRKTHERSKGRD..AHTL-...............  209
TAN.UG.x.SIVagmTAN1         -E-T-DV--HM--CH--P--T-R--QQ-IR-E...SFLW-T-KE.--VAENHW..GQ-R---F---TVYTDFL............RNGRR...............KRFQGKK.........TRMVRNL-SQQGAVGRMIKRHGSRTQSGSTTPFW.  211
VER.DE.x.AGM3               -EL--AT--S---TH--T--TER--------Q...RFTF-QFPE.--K-T....GQ-P-------L-H................QNG-R...............QRSQRSK.........TGGTRNM-FEQGAVGRMAKRHARRYQSGSQDAFW.  205
VER.KE.x.9063               -EL--AT--GM---H--S--TER--Q--IR-E...RFVF-QFPE.--KTT....GK-P-------L-H................QNGIR...............ERSQRKQ.........ARRSRDL-SKQRTMGRMAKGHVRRSQSGSQITFW.  204
VER.KE.x.AGM155             -EL--NT--SM--CH--T--T-R--QQ--R-N...RFIF-QFPG.--KLT....GQ-P-------L-H................QNG-R...............KRSQRGE.........TRRTRNL-SQQGAVGRMAQRYGRRNQQRSQTAFW.  205
VER.KE.x.TYO1_patent        -EL--AT--SI--GH--N--KER--QQ--R--...RFVF-QFPE.--KST....GQ-P-------L-H................QNG-R...............ERSKRGK.........TRRSRNL-SKQGAVGRMAKRYVTRSQPGGEAAFW.  205
WRC.CI.97.97CI14            -FIM-SE-NKI--SQH-E-WTQNVV-NTIRTGKIN-IDA-Q-HK.--AG.........-LT--T-C-QV-QHEQLQR..............................AGAFR-.........TSTNSRGPS-Y.ARALCNGKTAFP...........  190
WRC.CI.98.98CI04            -FIM-SE--VM--SQHV--WTQNVV-DTIRSGRLDIRNA-Q-HK.--AG.........-LT--T-C-QV-QHEQLQG..............................TS-YR-.........TSTNSRGPS-Y.AR-LCNGKTAFP...........  190
WRC.GM.05.Pbt_05GM_X02      -FITVK---V---SQH-E-WTQ--VKNTIRQGYLHPIDS-K-HQ.--LG.........-LT--A-C-Q--QHEQLRKR.............................-APNR-.........TPGNTRGSS-Y.ARAVCHGKTAFP...........  191
```

```
                          Vpx start                                                                                                                              Vpx end
MAC.US.x.239              MSD..PRERIP..PGNSGEETIGEAFEWLNRTVEEINREAVNHLPRELIFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGW.RPG..PPPPPPPGLA*   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
H2B.CI.x.EHO              -.-..----V-..---------V-------ET-L-HL--V---------------K--A--RE-----I--T-------M---M-I-FA---G--R----P---.-S-..-----------   112
H2G.CI.92.Abt96           -.-..----V-..-E--D-----------E---T----I-------------R-------------A--T-----L-M---M-V------T--QK---P---.-Q-..-----------   112
H2U.CI.07.07IC_TNP3       ---..-------..-----D--VE------ES--TA--Q-------------------G--------T---T-----------Y----R--Q--R----P---.---..-----------   113
H2U.FR.96.12034           -G-..-------..----E---V----S--E---RD---A--Q---------------Q--R------E--T-------------------Q-------P---.-S-..---------I-   113
DRL.DE.11.D3              -AEGQ-A--A-AE-VGA--...V-LE---Q-SLLR--Q--RL-FHP-FL-RL-NTCM-HY--ALQL-FI-S-----L-L---M---FQQ--S-VQGR-PPPLRPAGDRL------.--V-   116
DRL.DE.11.D4              -AEGQSV--A-AE-VGA--...V-LE---Q-SLLR--Q--QL-FHP-FL-RL-NTCM-HY--ALQL-FT-S-----L-LH--M-T-FQQ--S-VRGR-PPPLRPAGDRL---------V-   117
DRL.DE.11.D5              -AEGQ-M--A-AE-MGA--...V-LE---Q-SLLR--Q--QL-FHP-FL-RL-NTCM-HY--ALQL-FT-S-----L-L---M---FQL--S-VQGR-PPPLRPAGDRL---------V-   117
DRL.DE.11.D6              -AEGQ-A--A-AE-VGA--...V-LE---Q-SLLR--Q--RL-FHP-FL-RL-NTCM-HY--ALRL-FT-S-----L-L---M---FQQ--S-VQGR-PPPLRPAGDRL-----------   117
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-------*-V-   116
MAC.US.x.251_BK28         ---..-------..--------------------------------------------------Q---------M------------------------.---..-----------   113
MAC.US.x.EMBL_3           ---..-------..-------------------------------------------------X.............................................   63
MND-2.CM.98.CM16          -AE..GAPE--E...GA--...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-EV-TGA--...A-FQP--RDML-KV-L--RL-FHP-F--RL-RTCV-H---VHQR-LE-AA----L-M-----I--QT--SQR...--PNPR.AV-ERI..TIL--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-ERI..QIL--M.-   109
SMM.CI.79.SIVsmCI2        -GS..A-----..--T-------------YS---AV--G--K-----------R----F--------QA-TQ----L-------C-V----T--R----S---.-A-..-----------   113
SMM.LR.89.SIVsmLIB1       ---..-------..-----------------E--------Q-------------R----F--------A--T-----------M-L-------------.S---.-S-..-----------   112
SMM.SL.92.SIVsmSL92A      -.-..---Q--..-----D-------D--E-----V--T-------------R-------------AA-T-------------------------.---P---.---..--------F--   111
SMM.SL.92.SL92B           -T-..-------..-----------------HN---AL-QT--Q-----------R-C----V----Y----A----VQ-M---M-Q-FR---T-R----SQ---.-T-..-----------   113
STM.US.89.STM_37_16       ---..-------..-------A-E------H----D-----------------R-------------PG------------------S-R-------------.---..-----------   113
```

```
                                                                 amphipathic α-helix                                         H(S/N)RIG motifs
                                                  oligomerization                                             Vpr end in HXB2
                              Vpr start                                                              frameshift in HXB2
H1B.FR.83.HXB2               MEQA........................PEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYETYGD........TWAGVEAIIRILQQLLFIH#FRIGCRHSRIGVTR.....QRRARNGASRS..................................  96
H1A1.UG.85.U455_U455A        ----........................----------YA-*A-------H-D------Q---------N-----........--E----------------.-----Q-----IIP.....G--G-------..................................  95
H1C.ET.86.ETH2220            ----........................----------Y---A-------Q--------P--N---Y-------........--S---L--T---M---.-----Q-----IL-......------------..................................  96
H1D.CD.84.84ZR085            ----........................-A--------Y-----------S-----------S---Y-------........-------------------.-----Q-----I-P........-------S---..............................  96
H1F1.BE.93.VI850             ----........................-G--------Y---A--I------------P---------N-----........--E-------------------------IVP.......---V-------..............................  96
H1G.SE.93.SE6165_G6165       ----........................----------Y---A----------------L---------N-----........--E----------------.-----Q-----I-P....R--V-D-PG--..............................  96
H1H.CF.90.056                ----........................--------------------I---------V---Q------N-----........--V---L--T---------.-----Q-----I--.......--V---P---..............................  96
H1J.SE.93.SE9280_7887        ----........................---H------YH-----------------P---S---Y--S-----........--E----------------.-----H-----IIP.......---G-------..............................  96
H1K.CM.96.96CM_MP535         ----........................----------N------I-----R--------P---N-----T-----........--E-L--------------.-----H-----IIP.......---G---S---..............................  96
H1O.BE.87.ANT70              ----........................--N---AK--F---A-------A--------P---A---Y-------........--V--M-----------T-.Y----Q-----INP....RG-GR---S---*GAPLAPPWKSAPDPL.................  112
H1O.CM.94.BCF06              ----........................------A---F--*A-------E--------P--QAC--Y-------........--E--M---------M-T-.Y----Q-----IIPS.NTRG-GR---S---*NTSLASPWEQTPNPL.................  114
H1O.FR.92.VAU                ----........................------A---V-A-A-------A--------N--QA---Y-------........--E--I--T-----II-T-.Y----Q-----IIPS.NARR-GR-D-SH--*NAPLASPWKPTPDPL.................  115
H1N.CM.02.DJO0131            ----........................---A------Y--------------------T---------N-----........--E----------------.Y----Q-----I-P.....R.GR-------..............................  95
H1N.CM.04.04CM_1131_03       ----........................---A------Y---A-G---D------------------------D-----........--E-------M-------.Y----Q-----I-P.......-R-------..............................  95
H1N.CM.95.YBF30              --R-........................---A------Y---A------------------------------N-----........--E----------------.Y----Q-----I-P.......-R---T---..............................  95
H1P.CM.06.U14788             ---P........................------P---F---M-QT-----E--------P---S---Y--N-----........--E--T---------I---.Y----Q-----ILTP...SR-GR-H-P---*SASMATARESALKPM.................  113
H1P.FR.06.RBF168             ----........................------P---F---M-QT-----A-------IQ---S------N-----........--E--T---------I---.Y----H-----ILPP...SR-GR-H-P---*PASMAAARESALKPM.................  113
CPZ.CD.06.BF1167             ----........................--N---P---W-Q-L-DT---I----M-----NT-QNI-NWVFQN---........S-E--QVL-TL--KA--T-.Y-Y--A-----.........IGPR-------RS*ALVTPKLYTKHALQFLLL.................  113
CPZ.CD.90.ANT                ----........................-Q-E------M---L--T---I---------QPT-QH--NWV-AN---........S-R--QVL-IL--KA--T-.Y-H--A-----HGP.............R-R*NTSMVTPSCYTSNAMQ.................  104
CPZ.CM.01.SIVcpzCAM13        --H-........................------P---Y---A--I--D------K-Y--S---Q------T-----........--E--------------V-.--V--H-----IIP.........---RA---N-P*VGAMEASRQPAKNSL.................  111
CPZ.CM.05.LB715              ----........................------P---Y---A--V-----Q--------Q-------Y--N-----........-------------------.-----Q-----IVV.......--RT--SN-P..............................  95
CPZ.GA.88.SIVcpzGAB2         ----........................------AG--Y---A--I--D-----------P---Q---L--NN-----........--E---------H---V-.-----Q-----IE-.....QR--R-D-SY--..............................  97
CPZ.US.85.US_Marilyn         ---V........................------A---Y---A-DV-----E--L-----P--Q----Y--S-----........--E----------------.Y----N-----ISLARRTP-.GR---S---..............................  100
GOR.CM.04.SIVgorCP684con     ---I........................-T-E------F-------------Q--------P---Q----V-D-----........--E--T--------VI-V-.Y----H-----ILP..PSR-GR---PN--*AATMATAREPAPNPMQHMLL.................  118
GOR.CM.07.SIVgor2139_287     ---I........................-T-E------F-----M-----Q--------P---Q----V-D-----........--E--T--------VI-V-.Y----H-----ILP..PSR-GR---PN--*AATMAATREPAPNPMQHMLL.................  118
GOR.CM.07.SIVgorCP2135con    ---I........................-T-E------F-----V-----Q--------P---Q----V-D-----........--E--T--------VI-V-.Y----H-----ILP..PSR-GR---PN--*AATMAAAREPAPNPMQHMLL.................  118
GOR.CM.12.SIVgor_BPID1       ----........................------P---F---M--T-----A--------P--QL------D-----........--E--T---------V-V-.Y----Q-----ILP..PSR-GR-H-PN--*AASMAAARESALKPM.................  113
GOR.CM.12.SIVgor_BQID2       ---V........................------A---F---A--V-----A--LK----P---A------------........--E--T---------I-T-.Y----H-----IIP.SNARG--R---P---*GAPLASSWESTPNPL.................  115
GOR.CM.13.SIVgor_BPID15      ----........................----R-P---F---M--T-----A--------P---S------D-----........--E--T---------I---.Y----P-----IAP..HP--GR-H-PN--*FASMATARESALKPV.................  113
                              Vpr start

MAC.US.x.239                 --E.......................RP--NE------WD--VV-V-----E--LK--DPRL-TA--N---NRH--.........-LE-AGEL-----RA--M-.--G--I-----QPGGG...NPLSAIPP---ML...............................  101
H2A.DE.x.BEN                 -TE-...................PTEFP---GT-R-DLGSD-VI-T-R-I-E--L---DPRL-IA--YY-HNRH--.........-LE-AREL-KT--RA--V-.--A--NR----..........................................  87
H2B.CI.x.EHO                 -AE-...................VPEIP---KN-----WEQ-VVDV---I-Q--LK--DPRL-TA--NF--NRH-N.........-LE-AGEL-KL--RA--L-.--G--Q-----QPGGG...NPLSAIPP-...................................  101
H2G.CI.92.Abt96              -AE.......................EIP---EA-----WD--VV-V---I-E--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-K---RA--L-.--GD--C----GNAGG...NPLSTIPP--GVF...............................  102
H2U.CI.07.07IC_TNP3          -AE.......................RP---E------WD--VV-V-----Q--LK--DPSL-TA--NY--DR---.........-LE-AGEL-K---RA--V-.--G--A-----GQ-GG...NPL-AVPP--NVL...............................  101
H2U.FR.96.12034              -AE.......................IP---GA-----GD--VR-I-----E-----DPRL-TA--NY--NRH--.........-LE-AGEL-K---RAI-L-.--G--------SS-GG...NPLTTVPP-*GVL...............................  100
ASC.UG.10.RT03               --G-PPSHP.................THWLSRTV-MSQRQIQ-T-W--NE--EK-CS-EE-R-VW--CLSLPA-SNWTEQQAMA-TAIDY---V-K--WD-.YKK--Y-RYHARV-RFPQLRPLRGTAGQDTP..HDSDPPQVLRPSRYHPDN.............  133
ASC.UG.10.RT08               --GIPPSHP.................-HWLSRT.-MSQRQIQ-T-W--NE--EK--S-EE-R-VW--CLSLPA-INWTEQQAWA-TAIDY-K-I-K--WD-.YKK--F-RYHVRV-RFPQLRPLRGTAGQETP..HDSDPPQVLRPSRYHPDN*WIHPKRG.....  139
ASC.UG.10.RT11               --GIPPSHP.................-HWL-RTV-MSQRQIQ-T-W--NE--EK-CS-EE---VW--CLSLPA-SNWTEQQAWA-TAIDY-K-I-K--WD-.YKK--Y-RYHARVKRYPHLRPLRGTAGQDTP..HDSDPPQVLRPSRYHPDN*WIPPKRR.....  140
COL.CM.x.CGU1                -SSGPKR...................EREDPPEDESP-RDFT-LQDQV-NITR--R----LTHISVIMWGCTRRHWGR.......P-TAGL-YV--MECA-YL-.VHA-QCRFPSTQNP...........................................  92
COL.UG.10.BWC01              -SEQAPK...................RPREGVPEDETV-RDFN-LQ-QV-LITQ--R----LTHISVIMMGCVRRHWPD.......Q--AAI-YV---ESA-YL-.IHA-QCKFPSRTNAP*RLATIN-GVRVVAFTVSCAFSPKAWASAIRRIRRLTAASKYV*.....  136
COL.UG.10.BWC07              -S-RHRR...................EQAPEDET--RDYT-LQ-QV-FITQ--R----LTHIAVIWWGCTRRHWPD.......Q--IGM-YV--MEIA-YL-.LHA-QCKFPSRTNNP*QLA-IS-GAKRVCITVSCAFFQKAWASATKRIRSLHKASKYET.....  134
DEB.CM.04.SIVdeb04CMPF3061   --RPPPSHP.................LTPLSRLVPISRLQLQYMMQDVNE--MK--Q-DE-I-VWNYCV-LPAEPGWTGEQAWA-S-IDY-KKV-KMIWL-.LAE--Y-R-AAAS-H.YPNI-PLSGRNQEVRDQE............................  117
DEB.CM.99.CM40               --RYPPSHP.................-HFTSRTIPMTRLGLQQAMQD-NE--LK--T-EE-F-VWN-CVDLPAQPDWTGEQAWA-S-IDY-K-V-KM-WL-.L-EA-F-REREA--G.YPNI-PLTGRNREVRDGE............................  117
DEB.CM.99.CM5                --RYPPSHP.................-HFTSRTVPMTRLALQQAMQD-NE--LK--T-EE-W-VWN-CVDLPAQPDWTGEQAWA-S-IDY-K-V-RM-WL-.L-EA-F-REREA--R.YPNI-PLTGRNREVRDGE............................  117
DEN.CD.x.CD1_CMO580407       -AE.......................................TTQRQLQ-AVWDITE--RK--SKEEIT-IWD-CWSLPALPHWTEGQVMA-A-IDF---M-KEIWK-.Y-V--F-REAERV-H...YPNIRPLRPR-EEP*TRTA...........................  102
DRL.DE.11.D3                 --R.......................P---E--P---YD--LVDT-V-IQE--LK--D-RL--NV-SWV--R---.........SLE--QQL-----RA--L-.--Y--QEG-F-.PAGRW..NPFRTFPRANQQL...............................  100
DRL.DE.11.D4                 --R.......................P---E--P---YD--LVDT-V-IQE--LK--D-RL--NV-SWV--R---.........SLE--QQL-T---RA--L-.--Y--QEG-L-.PAGRW..NPFRTFPRANQQL...............................  100
DRL.DE.11.D5                 V-R.......................P---E--P---YD--LVDT-V-IQE--LK--D-RL--NV-SWV--R---.........SLE--QQL-T---KA--L-.--Y--QE--L-.PAGRW..NPFRTFPRANQQL...............................  100
DRL.DE.11.D6                 --R.......................P---E--P---YD--LIDT-V-IQE--LK--D-RL--NV-SWV--R---.........SLE--QKL-T---RA--L-.--Y--QEG-L-.PAGRW..NPFRSFPRANQQL...............................  100
DRL.x.x.FAO                  --R.......................P---E--P---YD--LVDT-V-IQE--LK--D-RL--NV-SWV--R---.........SLE--HQL-IL--RA--L-.--Y--QE--L-.PAGRW..NPFRTFPRANQQL...............................  100
GRV.ET.x.GRI_677_gri_1       -ASGRDPRE.................PLPGWL-IWDLD---WD--LQDM-RD-NE--R---GMNM-IRVWNYCV-EGRR....HNTPWNEIGYKYY--V-KSM-V-.--C---RRGPFSPYEERRNGQGGG.APPP......PPGLA.......................  118
GSN.CM.99.CN166              --RIPPSHP.................MPWLSRRVPTTMAIAQNA-W-INE--EK--S-DE-R-IWHDVT-LPA-PDWTVDQAAI-CAIDY-.RV-T---R-.-KD--F-RVNLS.GRYPAIRPSRGTAPPD-NSVSHADPEQPRRPSRYRMDE.............  133
GSN.CM.99.CN71               --RTPPSHP.................LPWISRRVPTTMAVAQSAMW-INE--EK--S-EE-R-IWHDVT-LPA-PDWTVDQAAI-CAIDYV-RV-T---R-.--D--F-RYNRIV-RYPVIRPLRGTAPPD-NSVPHADPEQPRRPSRYRMDE.............  135
LST.CD.88.SIVlhoest447       -SRTR.....................QQRSPEKP---F--P---WGQ-LADTM--I-E--R----ETI-AQVSEYCVSST-SE........EKACMKF-T-VNRA-----.LPT-PSYAAAR...AGTAHPTEAAPRPTKRDIQRGRY...........................  114
LST.CD.88.SIVlhoest485       -SR-R.....................QQRSPERP---F--P---WGQ-LADTM--I-E--R----EAI-AQVSEYCVSST-SE........EEACMRF-T-VNRA-----.LPT-PSYAATR...AGTAHPTEAAPKPTERDIQRGRY...........................  114
LST.CD.88.SIVlhoest524       -SRQR.....................QQRSPERP---F--P---WAQ-LADTM--I-E--R----EAI-AQVSEYCVSST-SE........EEACMRF-TLVNRA-YL--.LPQ-PSYAAAR...AGTAHPTEAAPRPTGRDIQRGHY*STPILPE.....................  121
LST.KE.x.lho7                -SRQR.....................QQRSPERP---F--P---WGQ-LADTM--I-E--R----EVM-AQVS-YCVDST-SE........EEACMKF-TLVNRA--L-.LP.T-PSYA.AAR..AGTAHPTEAAPRPNEGDIQRGSD...........................  114
MAC.US.x.251_BK28            --E.......................RP--NE------WD--VV-V-----E--LK--DPRL-TA--N---NRH--T.........LE-AGEL-----RA--M-.--G--N-----QPGGG...NPLSTIPP-*GVL...............................  100
MAC.US.x.EMBL_3              -AKQNSRGD.............................................KQRGGKPP-K-ADFPGLAKVLGILA**TRDV--RA--M-.--G--N-----QPGGG...NPLSTIPP-*GVL...............................  74
MND-1.GA.x.MNDGB1            -.........................GQKRDEQVS-----P---Y-Q-LADTM--I-E--RK---L-I-NAVSEYCVQNT-SE.......EE-CEKF-T.LMNRAIWV-.LAQ--DGTF....-ERRPQLPPSGFRPRGDRL...............................  104
MND-2.CM.98.CM16             ---.......................P---EA-----Y---LIDT-A-VQE--LK--D-RL--AV-SW--K-----.........-LE--RQL-S---KA-SM-.Y-Y---E----QEGGG..CYPLRSFPR-DNPL*HMLL............................  105
MND-2.GA.x.M14               ---.......................P---EA-----Y---LIDT-A-IQE--LK--D-RL--AV-SW--------.........-LE--QKL-TL--RA--L-.--H---E----QEGGK..YNPLRSFPRPNNPL...............................  101
MND-2.x.x.5440               ---.......................P---EA---G-Y-K-LVGT-A-IQE--LK--D-RL--AV-SWV--QQ---.........-LE--QQL-S---RA--L-.--Y---G----QEGGR..YNPLRSFPSPNNPL...............................  101
MNE.US.x.MNE027              -.........................EERP---E------WD--VV-V-----E--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-K---RA--M-.--G--N-----QSGGG...NPLSTIPP---ML...............................  101
MON.CM.99.L1_99CML1          -.........................................VPIERQALQAAIW--NE--LK--S-EE-R-IWEQVT-LPA-PAWNADQAWA-CAIDYT-WV-TI-YR-.Y-E--Y-RYAEQI-RYPVLRPMRGTAPGPTSSVPQADPDNPRRPSRYRMDE*WNLLTQTYPKNNT  133
MON.NG.x.NG1                 --N.......................APPSHP-HWTSRVVPIERAALQAAIWD-NE--LX--SKEE-R-IWDQVT-LPA-PGWNADQAWA-CAIDYT-RV-TXI-R-.M-D--Y-RHAR-X-RYPXLRPLRGTAPGPTSSVPHADPERPLHPSRYHHDE.............  135
MUS-1.CM.01.CM1239           --R.......................VPPSHR-PWHSRVVPTTMQQAQQAMWD-NE--EK--S-EE-R-IWNDVT-LPA-PNWTVDQAAI-CAIDY--RT-T---R-.Y-E--Y-RYSNTI-RYPNIRPLRGTQ-PP-NSMPNADPTPPLRPSRYRMDE.............  135
MUS-1.CM.01.SIVmus_01CM1085 --S.......................LPPSHPAPWHSRVVPTTMRQAQ-AMW-VNE--EK--S-EE-R-IWNEVT-LPA-PDWTVDQAAI-CAIDYV-RV-T-V-R-.--D--Y-RYNR-M-RYPTIRPLRGTQPPP-NSTPNADPVPPLSPSRYRMDE.............  135
MUS-2.CM.01.CM1246           ---.......................APPSHPLSWLSR-TPTTMVQAQAA-W--NE--EK--S-DK-R-IWDRVT-LPA-PQWSVDQAAITCAIDY-KSV-T---R-.Y-D--Y--YAQTI-RYPPLRPLRGTQPPD-NSMPNADPTPSLRPSRYRMDE.............  135
MUS-2.CM.01.CM2500           ---.......................HPPSHPLPWLSR-VPTTMAQAQNAMWD-NE--EK--S-DE-R-IWDRVT-LPA-PNWTVDQAAI-CAIDY--WV-G---R-.--E--Y-RYSQ-I-RYPQVRPLRGTQPPD-NSTPNVDPFPPRRPSRYRMDE.............  135
MUS-3.GA.09.09GabOI81        -SNPRERIPPSHPR................PWHSRLVPTTMQQAQ-AMW-VNA--EK--S-EE-R-IWDRVT-LPA-PQWTVDQAAI-CAIDY--WV-TII-R-.Y-E--Y-RYANTIQRYPILRPLRGTQPPE-NSMPHADPTPPLRPSRYRMDE.............  139
MUS-3.GA.11.11GabPts02       --R-PPSHTL....................PWHSRRVPTTMQQAQ-AMW-VNE--E---S-EE-R-IWNDVT-LPA-PDWSVDQAAI-CAIDYV-RVNT-I-R-.--E--Y-RYNQII-KYPIIRPLRGTMPPDTNSTPHADPEIPRRPSRYRMDE*WKWYRQIHLTGNT  148
OLC.CI.97.97CI12             -AE.......................RV------P---LED-VRDTM--IIV--K-----EL-VRVFLFCAQRHIN.......DEMSTSLLF--L--RA--Q-.Y-N--PGVELRRNPHRRSEGMQE..........................................  96
RCM.CM.00.SIVagi_00CM312     --I.......................PQ--E------YDK-L-DT-M--QE--AK--TMEL-QAV-NYS--QH-E.........SIE--KKM-TL-NRA--L-.Y-H-WSG--V-TA-GR...NPLQSIPI--NTL...............................  100
RCM.CM.02.SIVrcm_02CM8081    --M.......................P---E------M---L-DT-M--QE--AK--TMEL--AV-NY---Q---.........SLE--RVL-TL--RA--L-.Y-H--TG----TN-GG...NPLQSIPL--NIL...............................  100
RCM.GA.x.SIVRCMGAB1          --L.......................P---E------YD--LMDT-I--QE--KK--TYAL-TQI-DYV--QH--.........SIE--Q-M--L--RA--L-.--N--AG----TS-GS...NPLRSIPQT-NIM...............................  100
RCM.NG.x.NG411               --M.......................L---EE-----YD--LMDT-I-IQE--KK--TPEL-TQE-NY---QH--.........SLE--K-M-TL-NKA--L-.--H--EG----GA-GG...NPLRSILH--NIL...............................  100
SAB.SN.x.SAB1                -ASGGWLPPVGGDPPKDPPKNPREEIPGWL-TWDLP---FD--LRDM-QD-NS--QC----NL-FR-WWN-V-EPAI...DHGQTRLE-WYKYC----KA--V-.MKGR-CKPKTHPAYGPGAGGPPPGL-GASGGAASAAPGL*.........................  140
SMM.CI.79.SIVsmCI2           -AE.......................HP---E--P---WDA-VVDV---I-E--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-K---RA-----.--A--L-----HSGGG...NPLSTVPP--GVL...............................  101
SMM.LR.89.SIVsmLIB1          -AE.......................RP---E--P---WD--VV-V---V-E--LK--DPRL-NA--NY--DR---.........-LE-AGEL-K---RA--M-.--G--------QSGGR...NPL-TIPP--GVL...............................  101
SMM.SL.92.SIVsmSL92A         -SLS......................PP---EA-----WD--VT-V---I-T--LK--DPRL-TA--NY--DR---.........-LE-AGEL-K---RA--L-.--G--TN----CR-GG...NPL-TIPP--NVL...............................  102
SMM.SL.92.SL92B              --H.......................A---ETNP---WD--IRDV-----E--LK--DPRL-TA--NYV-D----.........-IE-AGE--K---KA--L-.--H--T-----QP-GP...NPLGSIPSA-DVL...............................  100
STM.US.89.STM_37_16          -.........................THRP---E------WD--VV-V---I-Q--L---DPRL-SA--NY--DRH--.........-LE-AGEL-K---RA--M-.--G--------QPGGG...NPL-TIPPT-GVL...............................  101
SUN.GA.98.L14                -AS...................RREPVEQP---E--P---FAQ-LADTM--IRD--E----AAIVAQVL-YCADST-S........EKE-CMRA-TL-NRA-----LP.M-P............SYI-I-S-SGNVNPRPPTRPRPGQGDIRRGLQD.............  116
SYK.KE.x.KE51                -A--F..........FRVYQRQPGEPLFYIPRNVNWD-AGVISATAKA-QVAQ--CK--TAEEIF-VWNQCL-VEAGPEETPTMAWIRCMLDMH-A-NFM--E-.-AA--PQRTKYARHRGYPHPS..........................................  115
SYK.KE.x.SYK173_COMGNM       -AE-F..........FNPSQHVQGTPWFFIPRNVELT-NVINVTVKA-LVVT--SK--TPQEIY-VWNQSLNEEAGTD.SPTMAWERTMLDMV-A-NLM--E-.-AA--PQRTRYARHRGYPHPS..........................................  113
TAL.CM.00.266                --RL......................PPSHPLPLTSRLIPTPRPQLER-MR-VRD-MLI--T-DEVI-VWN-CV-LPA-PDWTGEQAWA-S-IDFA-KGN-M-LL-.-QQ--Y-EQQRR-PHYPNIRPLSGR-ERTMQ...............................  115
TAL.CM.01.8023               --RL......................PPSHPLPITSRLIPTPRAQLER-MR-VRD-MLI--T-EEVI-VWN-CV-LPAEPDWTGEQAWA-S-IDFA-KGNEM--L-.-QQ--F-EQQRR-AHYPNIRPLSGRR.NTMQ...............................  114
TAN.UG.x.SIVagmTAN1          -AEGR.................DSRERRPGWL-IWDLS---WD-*LRDMVA--NQ--Q---G-EL-FQVWNFCQ-EGER....NGAPMIERAYRYY-LV-KA--V-.--C---RRTPFEPYEERRNGVGGGRDGRE.....PPPGLA.......................  119
VER.DE.x.AGM3                -ASGR.................DPREARPGEL-IWDLS---WD--LRDM--DINQ--KM--G-EL-FQVWNYCQ-EGER....NRTPMLERAYKYYKLV-KA--V-.--C---RRQPFEPYEERRDGQGGGRAGRV......PPGLD*TTQSTPEAVYKQVLL.........  134
VER.KE.x.9063                -ASGR.................GPRENRPGEV-IWDLN---WD--LRDM--D-NQ--KL--G-EL-FQVWNYCQ-EGER....RGAPMTERAYRYY-LV-KA--V-.--C---RRQPFEPYEERRDGQGGGRANRA......PPGLD............................  119
VER.KE.x.AGM155              -ASGR.................DPREERPGGL-IWDLS---WD--LRDMV--IN---KL--G-EL-YQVWNYCQ-EGER....QGRPIAERAYKYY-LV-KA--V-.--C---RRQPFEPYEERRNGQGGGRPGRV......PPGLD............................  119
VER.KE.x.TYO1_patent         -ASGR.................DPREARPGEV-IWDLS---WD--LRDM-QD-NQ--RL--G-EL-FQVWNYCQ-EGER....HGTPMMERAYKYY-LV-KA--V-.--C---RRQPFEPYEERRDGQGGGRANRV......PPGLE*RLSEASADL..............  128
WRC.CI.97.97CI14             -SN.......................SR-QAPSGGP-PIPED-LGMQ--FAM-KQ----EV-VRVFLYCSQ.......QREYEEVPTMHYFH-V--RG--M-.LTQ--PKQQEKREQQFQRAEQLY--LERTAGDQGEVEQRDLISL*SSMFLCCLCV...........  126
WRC.CI.98.98CI04             -SN.......................SR-QAPTGGP-PTPED-LGMQ--FAM-KL----EV-VRVFLYCSQ.......QRNYEEGPTMHYFH-V--RG--M-.LTQ--PKQQEKREQQFQRAEQFF--LERAAGDQGEVEQRDLISL.......................  116
WRC.GM.05.Pbt_05GM_X02       -SN.......................SGR-RPPTGGPQAIPED-LGMQ--FAM-K-----EV-VRVFLYCSQ.......QRNYEEQ-TMHYFH-V--RG-YL-.LIQ--PKEKENRE-DFQRAEQFY--LERA-GDQGGVEQRDLISL.......................  117
```

```
                                                                Vpr end
H1B.FR.83.HXB2                                                         *  96
H1A1.UG.85.U455_U455A                                                  -  96
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
H1O.BE.87.ANT70                                                        - 113
H1O.CM.94.BCF06                                                        - 115
H1O.FR.92.VAU                                                          - 116
H1N.CM.02.DJO0131                                                      -  96
H1N.CM.04.04CM_1131_03                                                 -  96
H1N.CM.95.YBF30                                                        -  96
H1P.CM.06.U14788                                                       - 114
H1P.FR.06.RBF168                                                       - 114
CPZ.CD.06.BF1167                                                       - 114
CPZ.CD.90.ANT                                                          - 105
CPZ.CM.01.SIVcpzCAM13                                                  - 112
CPZ.CM.05.LB715                                                        -  96
CPZ.GA.88.SIVcpzGAB2                                                   -  98
CPZ.US.85.US_Marilyn                                                   - 101
GOR.CM.04.SIVgorCP684con                                               - 119
GOR.CM.07.SIVgor2139_287                                               - 119
GOR.CM.07.SIVgorCP2135con                                              - 119
GOR.CM.12.SIVgor_BPID1                                                 - 114
GOR.CM.12.SIVgor_BQID2                                                 - 116
GOR.CM.13.SIVgor_BPID15                                                - 114
                                                                Vpr end
MAC.US.x.239                                                           - 102
H2A.DE.x.BEN                                                           -  88
H2B.CI.x.EHO                                                           - 102
H2G.CI.92.Abt96                                                        - 103
H2U.CI.07.07IC_TNP3                                                    - 102
H2U.FR.96.12034                                                        - 101
ASC.UG.10.RT03                                                         - 134
ASC.UG.10.RT08                                                         - 140
ASC.UG.10.RT11                                                         - 141
COL.CM.x.CGU1                                                          -  93
COL.UG.10.BWC01                                                        - 137
COL.UG.10.BWC07                                                        - 135
DEB.CM.04.SIVdeb04CMPF3061                                             - 118
DEB.CM.99.CM40                                                         - 118
DEB.CM.99.CM5                                                          - 118
DEN.CD.x.CD1_CMO580407                                                 - 103
DRL.DE.11.D3                                                           - 101
DRL.DE.11.D4                                                           - 101
DRL.DE.11.D5                                                           - 101
DRL.DE.11.D6                                                           - 101
DRL.x.x.FAO                                                            - 101
GRV.ET.x.GRI_677_gri_1                                                 - 119
GSN.CM.99.CN166                                                        - 134
GSN.CM.99.CN71                                                         - 136
LST.CD.88.SIVlhoest447                                                 - 115
LST.CD.88.SIVlhoest485                                                 - 115
LST.CD.88.SIVlhoest524                                                 - 122
LST.KE.x.lho7                                                          - 115
MAC.US.x.251_BK28                                                      - 101
MAC.US.x.EMBL_3                                                        -  75
MND-1.GA.x.MNDGB1                                                      - 105
MND-2.CM.98.CM16                                                       - 106
MND-2.GA.x.M14                                                         - 102
MND-2.x.x.5440                                                         - 102
MNE.US.x.MNE027                                                        - 102
MON.CM.99.L1_99CML1         PQQPHAHLVLTAFVEFVAFTVRFASSQKASGFHMAGNGRDEERRLLSLALAAVRILQESSE- 195
MON.NG.x.NG1                                                           - 136
MUS-1.CM.01.CM1239                                                     - 136
MUS-1.CM.01.SIVmus_01CM1085                                            - 136
MUS-2.CM.01.CM1246                                                     - 136
MUS-2.CM.01.CM2500                                                     - 136
MUS-3.GA.09.09GabOI81                                                  - 140
MUS-3.GA.11.11GabPts02      PRQHRRRLATLVFVKSVLITVY...............CASRKKH- 178
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.CI.79.SIVsmCI2                                                     - 102
SMM.LR.89.SIVsmLIB1                                                    - 102
SMM.SL.92.SIVsmSL92A                                                   - 103
SMM.SL.92.SL92B                                                        - 101
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                                                          - 135
VER.KE.x.9063                                                          - 120
VER.KE.x.AGM155                                                        - 120
VER.KE.x.TYO1_patent                                                   - 129
WRC.CI.97.97CI14                                                       - 127
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             MEPVDP............................................RLEPWKHPGSQPKTAC.TNCYCKKCCFHCQVCFITKALGISYGRKKRQRRRAHQN....SQTHQASLSK.QPTSQP....RGDPTGPKE*KKKVERETET.DPFD............. 100
H1A1.UG.85.U455_U455A      ------............................................N-----------T---.S-----V--W---L--LK-G------K---KP--GPP-....G-KD--TLIP-.--LP-S....QRVSA-QE-S-----SKAK-.-R-A............ 101
H1C.ET.86.ETH2220          ------............................................N----N----------.NQ------SY--#SLLSD-G---------------P-....S-KD--NLI--.--L-HT....------SE-S-----SKA--.--YA............ 100
H1D.CD.84.84ZR085          -D----............................................NID--N------R---.N--------Y---------G--------------PPH....S-----DPIP-.--S---....-------Q--------SK--V.H---W........... 102
H1F1.BE.93.VI850           ------............................................S-D--N------T-P-.-K----R------W--T--G-------------H-TP-....S--V--N--P-.--L--A....---------S--E--SKAK-.--CA............ 101
H1G.SE.93.SE6165_G6165     -D----............................................N----N--------P-.NK-F--V--W------LN-G----------KH--GTP-....S-KG--DPVP-.--LPTT....--N------S--E-ASKA-A.-QC-............ 101
H1H.CF.90.056              -D----............................................K----N------Q---.N--------Y---M--LK-G----------S--H-TPA....SL-D--N-I--.--L-RT....H--------Q--E-ASK---.-............. 99
H1J.SE.93.SE9280_7887      ------............................................NR---N----------.---------Y-----LQ-G---------------S-PP....G-K---DLIP-.--L--T....QRK----E-S--E--SKA-P.-R-............ 101
H1K.CM.96.96CM_MP535       -D----............................................NI---NQ---------.NQ----R--Y---I--LK-G----N------P--TTPY....N-EN--DP-R-.--L---....--EQ-D---S-----SK-K-.-Q-............ 101
H1O.BE.87.ANT70            -D----............................................EVP--H------QIP-.N-----R--Y--Y---VR-G----------G.-P.-AA....SHPD-KDPVP-.-SPTIT....KRKQERQE-QEEE--KKAGP.GGYPRRKGSCHCCTRTS 112
H1O.CM.94.BCF06            -D----............................................EIP--H----K-Q-P-.N-----R--Y--F---TK-G----H------.-PA-AA....SYPDNKDPVPE.-SP-HT....-RKQKRQE-QE----A*-GA.SGHPCNQDSYNCCTRIS 112
H1O.FR.92.VAU              -D-I--............................................EMP--H------Q-P-.NA----Q--Y--YL--TK-G----H------RPAA-AS.....YPDNKDPVPE.-SP-NN....KRK*ERQE-QE-E--TK-GA.GGLHCRKGSCICCTRTS 112
H1N.CM.02.DJO0131          ------............................................-----N----------.NG----Y--Y--MC--TK-G----------S----TPK....S-KN--DPIPE.--L--QQQ..P--Q--Q-KQ--AL-DKA--.--............ 101
H1N.CM.04.04CM_1131_03     ------............................................-----N----------.N---------Y--V---TK-G----------S----TPN....S-KNY-DPIPE.--L--Q....P--Q--Q-XQ--AL-XK--A.--C-............ 101
H1N.CM.95.YBF30            ------............................................-----N----------.N-----R--Y--LY--TK-G----------S----TP-....S-KS--DLIPE.--L--Q....Q--Q--Q-KQ-EAL-SK--A.--C-............ 101
H1P.CM.06.U14788           -D----............................................D-P--QQ-----SSP-.N-----A--Y--Y---TK-G-----------RPA--DN....NN-N--DPV--.*-FACT....GAQQE*HQ-QT----KQATA.-RS-CWKDYCLSSGTT. 111
H1P.FR.06.RBF168           -D----............................................D-P--QQ-----SSP-.N-----A--Y--Y---TK-G-----------RPA--DN....NN-N--DPV--.*-LA*T....GIQQE*HQ-QT----KQATA.-RS-CWTDSCLSSGTT. 110
CPZ.CD.06.BF1167           ------............................................D----L--S-T-S-P-.NS----V--Y---L--TK----------R-GK..PSAS.....NKNN-NPVRE.-SL-KQR.....-QVSQ--Q--E-ANQA--.GGG-IQ........... 100
CPZ.CD.90.ANT              -D---A............................................ETP--L--PAT-A-P-.N-----C--Y--PL--TK-G--------R-AR-N-RTTAE..S-ENN-DPV--.*SLPKT....SRIQSSQ-K-E----EK-GS.GGGPCSTGSSHSCGWT. 114
CPZ.CM.01.SIVcpzCAM13      ---I--............................................E---------------.N-----A-----MY--TK-G-----------AR--T-FAS.....-KNN-DPVR-.--L----....--NQERSEKQAE---SQAAA.---............ 99
CPZ.CM.05.LB715            -D-I--............................................NI---N----------.N-----S-S---PL--MK-G---------------G-SKS.....N-N--D-IPE.--F--S.....---QSS-EKQE----SK-TSDPFGC*TGSCH....... 107
CPZ.GA.88.SIVcpzGAB2       -D-I--SLEPWNHPGSQPKTAC....................................................NS---R--Y--ML--TS-G----------GR--T-LKG.....-KSN-DP-RQ.--L--Q....Q--QKCQ--SEEE--SK--A.-L.............. 99
CPZ.US.85.US_Marilyn       -D----............................................NI----Q----------.N-----------V---TK-G----------GR---TTAKS.....-ADN-DP-R-.--F--Q....L-NQ--Q--Q--AL-KQ-TS.-QR-.......... 102
GOR.CM.04.SIVgorCP684con   -D-T--............................................E-P--QQ-----P-P-.NT-------Y--Y---TK-G-----------RPA-TADK.....D-DN-DPV--.-SLAGT....-SQQE*YQK-E----GQ-A-.-RS-CWQDSCLSSGTIS 113
GOR.CM.07.SIVgor2139_287   -D-T--............................................E-P--QQ-----P-P-.NT-------Y--Y---TK-G-----------RPA-TADK.....D-NN--PV--.-SLAGT....-SQQE............................. 82
GOR.CM.07.SIVgorCP2135con  -D-T--............................................E-P--QQ-----P-P-.NT-------Y--Y---TK-G-----------RPA-TADK.....D-NN--PV--.-SLPGT....-SQQE............................. 82
GOR.CM.12.SIVgor_BPID1     -D-I--............................................E-P--QQ-----SSP-.NT-F--V--Y--Y---VK-G-----------RPT-TNN-.....H-N--DPV--.--LART....GAQQE............................. 82
GOR.CM.12.SIVgor_BQID2     -D----............................................EVP--H------Q-P-.N--X--X-AY--YL--TK-G-----------RPAV-AS.....H-GDKDPVPE.--LPHT....PRNQERQ-KQEE---K--S-SGLHCHQDSCISYTWTSG 114
GOR.CM.13.SIVgor_BPID15    -D-T--............................................D-P--QQ-----SSP-.N--F--A--Y--Y--LVK-G----------GRAA-PNN-.....N-N--DPV--.................................................. 71
                           Tat start                                                                                                   exon 1 end exon 2 start
MAC.US.x.239               -...ETPLREQENSLESSNERSSCI............SEADASTPESANLG-EILSQLYR-LE--YNT-------Y---F--LK-G---C-EQSRK-R-TPKK.......AKANTS-A-N..KPISN...RTRHCQPE-AK-ET--KAV...ATAPGLGR......... 130
H2A.DE.x.BEN               -...ETPLKAPESSLKPYNEPSSCT............SERDVTAQELAKQG-ELLAQLHR-LEP-TNK----R-S----L--SK-G-----E--G--R-TPRK.......TK-PSP-APD..KSIST...RT--SQPT--Q--TS-ATV...VTTCGLGQ.......... 130
H2B.CI.x.EHO               -...EIPLKEQESSLNSSSGHSSST............SEGVANTQGLDN-G-EILSQLYR-LK--SNT------SY---L--LK-G---C-E..RS-K-SSKR.......AK-TTS-APN.E.SLSA...RT--SQPT-KQ--E--TTR...ATDLGPGRSNTSTSRFA 137
H2G.CI.92.Abt96            -IGMETPSREQENSLKSCRELSSCT............FEETVDAPGLEGTQAEILYQLYR-LE--SNK-------Y---L--LN-G--VC-E.QP--R-TPKK.......TKANSF-A-N.DRSISA..RTRNXQPK-K--E---TKV...ATDLGLGR......... 133
H2U.CI.07.07IC_TNP3        -...ETPLKEQESLSKSCKEPSLCT............SEEDVRIPESGDKE-EILSQLYR-LET-YNK-------Y---L--LK-G---C-E..QS-R-SPKK.......AKVNTFPA-N..KSIPT...RTRNSQPK-KQ-ET--TTL...ASTPGLGK......... 128
H2U.FR.96.12034            -...EIPLKEQESSLKSCKEQSSST............SEGDVGTQGSAAAG-EILSQLYR-LE--SNK-------Y---L--LK-G---C-E..QP-R-SPKK.......VKAYSSAP-N..KSLSN...GTRNSQPKEK---AL-TAV...ATDLGLGR......... 128
ASC.UG.10.RT03             -D-SQEG............................................--K-HRAPA--P---.-P-F--N-IY--IL--QR-G---R---FRK-R-PG-APTP..ESL-NN-D-VRQ.-Y...........QVDEEKQT----T--...TSDPCDC*TGS...... 101
ASC.UG.10.RT08             -D-SQEG............................................--K-HRAPA--P-P-.NS-F--S-IY--IL--QQ-G---R---FRK-R-PG-APTP..ESL-NN-D-VRQ.-S...........QVDE-KQAA---T--...TSDPCDC......... 98
ASC.UG.10.RT11             -D-SQEE............................................V-K-HRAPA--P-P-.NS-F--S-IY--IL--QR-G---R---FRK-R-PG-APTP..ESL-NN-DFVRQ.-F...........QV-EEKQAE---T--...TSNPCDC......... 98
COL.CM.x.CGU1              -.SL....................................................SLYAK-LE--NDK-W--A------L--LS-G---R-KKN-SAT...................QT.-TNVSS....QIRGQDREWEQPDNAE-IHREQENSSSRFWR#...... 85
COL.UG.10.BWC01            -*IS....................................................FQNER-LK--NNK-W--G------L--LS-G---C-KKN-KIN...................RC.-EPLSS....QIRGQNQEWEQPDNGE-LDSEQEDNNSRLVQL#..... 86
COL.UG.10.BWC07            -QIS....................................................FQNQ--L---TNK-W--A-VY---L--LS-G---C-KKN-KSP...................QS.KEEELTTIL.QIPTQTQW-QPDNA-ELDS#.IQEAAAADFGYW..... 89
DEB.CM.04.SIVdeb04CMPF3061 --EEMDLFQGRGRGEANH.................................................PS----I-T---YL--TQ-G---T--TRR-KR-APESA..TSDS-NQDIV-KQ...R.........RISSFQEKE-EAL-AAPGD.GGLPRRKDS....... 98
DEB.CM.99.CM40             --SI--FEERANT...PH.................................................PA-M--A-IL---L---Q-G------PRR-KRKC-REYTPAAESN-PD-DFVPE.-..........RKCFDSQKKEAE--ETAGP.GGQHCREDSSVSSGRIS 107
DEB.CM.99.CM5              --EI--FKETANT...PH..................................................P--M--A-TM---L--MQ-G---Y--SRR-KR--GSSFA...ANDKD--EFIPE.-..........WKCFNSE-KET---KASRA.GGRHRPEDS........ 96
DEN.CD.x.CD1_CMO580407     -QKEIWKHYRVGCFHREAERVRHYP.............NIRP#TPSQRRAMNADSID-FAGN--P-N.T----H-AY---L--LQ-G---V--.RAP-K-P--R.....QSGTPD-EAV--.-SL-D-R.....RQQISQKKE-EKVETATGSSGGNIREDS......... 135
DRL.DE.11.D3               -DARKVDLDQQDA..GIH.....................................FE-FQGRTNN-.NK-W--H-AY---L--LQ-G---R-NVSRP-K-TASVV....SRVSETGDTPVA.RTHTAQRR.PTTTDSQT-KKE-SKTE-SGI.SDQSPPSQNLQDLGAKK 123
DRL.DE.11.D4               -DARKVDLDQQDA..GIH.....................................FE-FQGRTSN-.NK-H--H-AY---L--LQ-G---R-HVSRP-K-TA-VV....SRVFETGDTPVA.GTHTAQRR.PPTTDSQT-KKE-SQAETSG-.PDQSPPGQNPPGLGAKK 123
DRL.DE.11.D5               -DARKVDLDQQDA..GIH.....................................LE-FQGRTNN-.NT----H-AY---L--LQ-G---R-NVSRP-K-TA-VV....SRVSETGDTPVA.RTHTAQRR.PTTTDSQT-KKE-SQTETSG-.SNQNPPSQNPPSLGAKK 123
DRL.DE.11.D6               -DARKVGLDQQDA..GIH.....................................FD-FQGRTNN-.NK-W--Q-AY---L--LQ-G---R-NVSRP-K-TASVV....SRVSETGDTPVA.RTHTAQRR.PTTTDSQTEKKE-SKTEASGI.SDQSPPSQNLQDLGAKK 123
DRL.x.x.FAO                -DARKVDLDQQDA..GTH.....................................FE-FQGRTNN-.NT-W--Y-AY---L--LQ-G---R-NVSRP-K-TATVV....SRVSETGDTPVA.RTHTAQRR.PTTTDSQTEKKE-SKTEASG-.SDQNPPSQDPPKLGAKK 123
GRV.ET.x.GRI_677_gri_1     -DKEEEP...HPLLQDLHR.........................................-LQP-TNK-------Y---EL--LQ-G--VR-HVSRK-RKTSTQ.........DN-DPIRQ.-SI-TV......QRN-QTTEEG-T-V-KAAAAN.............. 94
GSN.CM.99.CN166            -NRDPEECLLPNWQQ..........................................--AA-A-P-.SA------AL---L--LR-G--L-LHGRS-KRK-PKATEP.DPSVSSN-D-A*V.-SL-HKQW....NAVRQTQP--ALA-KARS.G-.............. 104
GSN.CM.99.CN71             -LILSSPEDLQDTAW............................MNRDPAECL-PK-QQ--AA-N-P-.SA----R-A----L--LR-G--L-LHGRS-KRK-SEASEP.APSVP-N-DPIQI.-SLPHSKW....NAIRQT-QE-T-A-KATS.G-S-ISQDSRVSCGTT. 134
LST.CD.88.SIVlhoest447     -QQQEQEQLTRQKQHQDPLKETYKEAVT.............................K-LS--NNK-C--------IL--QQ-G---R-YVHR--K-VAQVF........EDKQDPVD.GTA-GRKE...RTKNSQTKKER*RKTKKTSTST.............. 110
LST.CD.88.SIVlhoest485     -QQQEQEQLTRQKQHQNPLKEIYKEAVT.............................K-LSV-DNK-C--R------IL--QQ-G---R-YVHR--R-VTQVF........EDKQDPVD.GGRKE.....RTQNSQTKKER*RKTKKTSTST.............. 107
LST.CD.88.SIVlhoest524     -QQPEQEQHTQQKQHLDQLEEIYKEAIT.............................D-LQS-QNK-H--V------IL--QQ-----R-YVHRT-R-VTYIL........ETEQDPVD.SGRTE.....RAKNSQTKKERQRKT-KTSTST.............. 108
LST.KE.x.lho7              -QQPEQEQPIQQKQHQDLMKETYNEAVT.............................KALQ--QNT-C--------IL--QQ-G--LH-YVHRK-R-VTSIL........ET-QDPVD.SGRKK.....RTKNSQT-K-R-EQIQRLSTST.............. 108
MAC.US.x.251_BK28          -...ETPLREQENSLESSNERSSCI............LEADATTPESANLG-EILSQLYR-LE--YNT-------Y---F--LK-G---C-EQSRK-R-TPKK.......AKANTS-A-N..KLIPN...RTRHCQPE-AK-ET--KAV...ATAPGLGR......... 130
MND-1.GA.x.MNDGB1          ---SGKEDHNCPPQDSGQEEIDYK.............................QLLEEYY--LQ--ENK-W--------ML--QK-G---R-HVYRK-VPGTNKK......IPGSGEEAIR.RTA-RT......YTANGQTTE--K.......ATA............. 107
MND-2.CM.98.CM16           -DAGKAVSDKK......................................EGDVTPYD-FRDRT-P-.DT----R--Y---L--LQ----VH-HAYRN-RS-QRLL....EKISEDS-APVG.RE-ISVPT....TSNSQAKKE--SPTQKNR.V-NQTAPK......... 112
MND-2.GA.x.M14             -DAGKAELDKK......................................EANITPLD-FQGRT-P-.NK-------Y---C--LQ-----H-HVYRT-RT-KRLL....GKISQDSEAAVG.RE-ISDLA....TSDSQTKKE-ESST*KNR.TSSQRAPRKNLADLGGQ 120
MND-2.x.x.5440             -DVGEVASDKK......................................EEDITHFD-FRART-P-.NK-F--V--Y---L--LQ----VH-HVYRQ-RP--GLL....EKVPSNSQAALG.RE-ISGPT....TSN-QT-KEE-SQT*KNR.TSSQRAPDKNLGSLGSQ 120
MNE.US.x.MNE027            -IDMETPLREQENSLKSSNGRSSCT.............SEAAATTLESANLE-EILSQLYR-LE--YNT-------Y---F--LK-G---C-EQSR--R-TPKK.......VKANTS-A-N.NRPISDR....TKHCQPKKEQ-ETV--AV.ATVPGLGR......... 134
MON.CM.99.L1_99CML1        -MEPVD...........................................PD-PKEQ--PAT-R-P-.---F-RV------I--L--G--------.--KR--ATSPVP...GLSSSKNPAR-.-SL-H-....TR-SQR-T-QAQA-ATAATP.-RQHI........... 104
MON.NG.x.NG1               -NSAXI.........................................VNPA-P-EQ--SAT-L-P-.N--F--I-AL-X----LQ-G--------GRK-R-GPR-PT....IGDRSYQDNPG.VKSLP....KSLGNQERQKKQE-AVESKAGTGXQSMREDF......... 111
MUS-1.CM.01.CM1239         -D-SVE............................................N-P-EQR--AA-T-P-.SS----R-A---ML--QK---------RF-KRP-SDQN......V-N-PNPVPI.KSIPNL....QRQQSRQ--Q-EALAAAAGP.--C-I........... 101
MUS-1.CM.01.SIVmus_01CM1085 -D-SVE...........................................E-PKEQR--AA-A-P-.-S-F----A---VL--QK---------RF-KRS-ADSS......LPS-SDPVQI.KSL-DR....-R-QKRQ--Q-ET--A-A--.--............ 98
MUS-2.CM.01.CM1246         -D-SVE............................................G-P-DQR--AA-P-P-.-------C-A---IL--QK---------RI-KRP-TDSL......LPSD-NPVRF.-SIPD............SQQN-SNKKKQKKA.LEEAAGTDPCRCFQKTS 106
MUS-2.CM.01.CM2500         -D-EVE............................................N-P--QR-S-A-T-P-.-P----A-AY--MQ--L-----------NRK-H--TPEG......LPNN-DPIQL.KSIPE-.....R-QER--KQ--E-AP-AVS.--CAI............ 101
MUS-3.GA.09.09GabOI81      -QIRL-L*DHPGIAWMNNLE............................EEN-P--QRGTAA-A-P-.NT-F--S-AY--IL--QK----------R--P--SDKG......L-SD-DPVPF.KTL-I-.....R-KDSQ----ETL-T-ATS.NQCAC............ 116
MUS-3.GA.11.11GabPts02     --V-P-D............................................TPN-Q--PAT-Q-P-.NS-F----AY--IL--QK-----L---SRK-R-GTDSG......LPNN-N-VQL.KSLPKH....E-L-VSQ-KQ--E-AATPA-.--............ 98
OLC.CI.97.97CI12           -DAQEWS..............................................LE-TLT....EDLKE-KNK-------Y---L--LQ-G----STESP-K--KKVV........S--QDSRD.TSHF-IPS...RTRNSQTTE-E-TASSSRA..-STSGTVPQSVQGH.. 107
RCM.CM.00.SIVagi_00CM312   -DGQEAGL.........................................ERQEE-TLYN-FQSVE-P-.NA------AY---L--LQ--I--Q-ASRR--RQ-DKKEA........KIDTLATPDKSL-ATS....RRRNSKPKE--ESSV-KAVATNHTTGRKDLSIS..... 115
RCM.CM.02.SIVrcm_02CM8081  -DVRAVGS.........................................ERIEE-TLYN-FHYQE-S-.NK------AY---L--LQ-----N-ASRR-GS-DKKKTQ........VDQL-ATEHRPLSATGR...NSQSK--E-ETL-KKVV..TNRPAGREDISTP..... 113
RCM.GA.x.SIVRCMGAB1        -DVQGVGLEH.......................................PEEVILYD-FRKRE-S-.NT-------Y---L--LQ-G---N-ASRA--R-SKEENK.......ADKFPVPN.HRSISTTR...-NRKLQEKKE-T--KKVA..TSTTIG........... 106
RCM.NG.x.NG411             -DVKGAGLEE.......................................REE-ILLD-FFTRE-S-.NS----R-VY---L--LQ-----N-ASRR--RG-PGNKK....KN-INPFDT-E.-SL-AI....GRNSQSK--K-EA--KAVD-.NITTGREDIFIS......... 115
SAB.SN.x.SAB1              -DQEQEARPQ...........................................VWE-LQEEL.HR-LQ--DNT-F--V-----IL--HK-----R-YVPRP-RASKK.........IS-NQVSLH.KSI-TW.....TRDSQTKK-S-AAVG*T..ATAGHTLGRENS..... 106
SMM.CI.79.SIVsmCI2         -IGMETPLKEQGNLSRSCNEHSLST.............SEQDVFTPELGTVG-EILSQLYR-LEE-SNR-------Y---L--LK-G---W-E.RS--R-APKK.......AKAYTF-A-N.HRSIST...RDRNGQPT--Q--T--TTV...ATDLGLGK......... 133
SMM.LR.89.SIVsmLIB1        -IGMETPSKEQGSSLRSCNEHSSCT.............SEEAVGTPELASQE-EILWQLYR-LEE-FNK---RS--Y---L--VK-G---T-A....-P-TPKK.......AKA-TPFT-N.HRSVPN...RTRHSQSK-----E-ATTV...ATDLGLGR......... 131
SMM.SL.92.SIVsmSL92A       -...ETPLKEQENLLKSCREPCSST.............SEEGVQTQELDAEE-EILSQLYR-LET-CNK-F-----Y---L--LK-G---C-EPERT-R--AKKK.......IK-DSFPS-N.YRSIPT...RSRNCQPK---T----TAV...V-DPGLGK......... 132
SMM.SL.92.SL92B            -........................................DAPIQGLDSQGDQIPWDLYR-LET-YNS-F---------L--LR-G---T-A..-P-R-VKKK.......AK.ANPFDTSNHRSISS...GTRNCQPK--Q-E---T-V...-TDCGLGR......... 105
STM.US.89.STM_37_16        -IDMETPLKEQESSLRSSSEPSSCT.............SEAVAATPGLANQE-EILWQLYR-LEE-CNK-F-----Y---L--V--G---T-E-SR--VKKK-K..........TYPISASNNRSLST...RARNSQPK--Q--E--T-V...ESTPGLGK......... 132
SUN.GA.98.L14              -STQGHQQDQDQGKGTLE...................EAYKTN...................LE--DNK-W-RR------L--LQ-G---H-Y..VY-G-QHQS.......ISKDKQEAVR.RIG-RK....P.A-NPQTKK---RQ-KK-S.TST............. 102
SYK.KE.x.KE51              -ST...................................EDKVRETQGIPTSFLEGTFLSNG-Q-P-.SR-------Y--YR--LQ-G---T-V-ERK-RHA-TTA......EDFAAH-STT.RRSPL-.....STTRPQQGQ--E*KQIQTI.TSPASLRSAENF..... 115
SYK.KE.x.SYK173_COMGNM     -S...................................STDQICQTQRVPPSFLEGTFLEKG-P-P-.NK-F--N--Y---L--LQ-G---T-A-PRK-AA-SI..........SEDD-APT.GTLP.......RAGRTQANPQT-KKAVETE.KDSTSLPSAENL..... 110
TAL.CM.00.266              -SSKEEL..........................................RTTPISDPFQEEGRGP-.NK-------Y---L--LQ-G---R.YVPD--F-K-RRSTP..SSN-EDKDPV--.--L-RA....T-IQKRQE-TLQ---SKAA-.SGT-RKQDS......... 110
TAL.CM.01.8023             -SSKEELPTTPTYDPF...........................................PAGG-P-NK.-------Y---L--LQ-G---R.YVPD--I-K-RRSTS..SSN-KDKDPV--.--L-RA....T-NQECQ-KKVQ---SP-GA.GGEHRLKDSLPLSGRTS 117
TAN.UG.x.SIVagmTAN1        --SEGDGMAESLLQDLHR...........................................-L-P-TNK-F--R--Y------LQ-G---T-HVSRI-RPKKN-.........SN--NLV-Q.-SI-AW.....GGNSQTTQEE-TKIPAAA.ETSRRPQ.......... 101
VER.DE.x.AGM3              -DKGEDEQGAYHQDLIE........................................QLKA-LKR-TNK----C--Y---L--LQ-G--VT-HAPRI-RKKI-P.........LDRFPEQ-.-SI-TR.....GRDSQTTQ-GQEKV--SARTAPSLGRKNLAQQSGRA 114
VER.KE.x.9063              -DKGEEERTVLHQDLIR........................................QYKK-L-T-RNK-F-----Y---L--LQ-G--VT-HAPRT-RKKSVQ.........PNRL-QQD.-SI-TR.....GRD-QATQESQKKV-R--TTAQILGRKDLERDKREA 114
VER.KE.x.AGM155            -DKGEEDQDVSHQDLIK........................................QYRK-LET-TNK-F-----Y---F--LR-G---T-HAFRT-RKKI-S.........ADRIPVPQ.-SI-IR.....GRDSQTTQESQKKV-EQAKANLRISRKNLGDETRGP 114
VER.KE.x.TYO1_patent       -DKGEAEQIVSHQDLSE........................................DYQK-LQT-KNK-F-----Y---L--LQ-G--VT-HAPRT-RKKI-S.........LDL-P-QH.-SI-TK.....WGRD-QTTPTSQEKV--TAGSN............... 100
WRC.CI.97.97CI14           -RKSLLCIISTGYCKEGSSCISPK.GVLNNKRKGNNNFKEQNNCIETWREQQ-IREKLSKETLSP-DHR-S-VA-VYS-TL--LQ-G-----RSYSKKTKPDTTT.............AAS.SR-TSTT....WKRDSKTAK-E*KKIT-V................. 131
WRC.CI.98.98CI04           -KK-PLCIIFTGYYKEGSSCTLPK.GVQNNKKRENNNFKEQNNSLETWKEQQ-IRERLSKETLSP-NHK-G-Y--V-S--L--LQ-G-----RSYSKKIGASTAK.............ATT.SRATSTT....WKRDCQT-K--QKKTTVL.................. 132
WRC.GM.05.Pbt_05GM_X02     -KNRPLCTIFTGSFKEDYIYTSYK.DVPKKKRTEKEIFKEQNNFIEIWKGLQ-IREELSRETLSP-NHQ-R----I-S--L--LQ-G-----HSYSKKTNKDKSD.............A-T.SRVIS*T....WKRDCQTTK-Q-KKIS..................... 129
```

```
                                    Tat end
H1B.FR.83.HXB2                 ..........*
H1A1.UG.85.U455_U455A          ..........-
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           ..........-
H1O.BE.87.ANT70                EQ........-
H1O.CM.94.BCF06                GQ........-
H1O.FR.92.VAU                  GQ........-
H1N.CM.02.DJO0131              ..........-
H1N.CM.04.04CM_1131_03         ..........-
H1N.CM.95.YBF30                ..........-
H1P.CM.06.U14788               ..........-
H1P.FR.06.RBF168               ..........-
CPZ.CD.06.BF1167               ..........-
CPZ.CD.90.ANT                  ..........-
CPZ.CM.01.SIVcpzCAM13          ..........-
CPZ.CM.05.LB715                ..........L
CPZ.GA.88.SIVcpzGAB2           ..........-
CPZ.US.85.US_Marilyn           ..........-
GOR.CM.04.SIVgorCP684con       EPWLCGSTGH-
GOR.CM.07.SIVgor2139_287       ..........-
GOR.CM.07.SIVgorCP2135con      ..........-
GOR.CM.12.SIVgor_BPID1         ..........-
GOR.CM.12.SIVgor_BQID2         Q.........-
GOR.CM.13.SIVgor_BPID15        ..........-
                                    Tat end
MAC.US.x.239                   ..........-
H2A.DE.x.BEN                   ..........-
H2B.CI.x.EHO                   N.........-
H2G.CI.92.Abt96                ..........-
H2U.CI.07.07IC_TNP3            ..........-
H2U.FR.96.12034                ..........-
ASC.UG.10.RT03                 ..........-
ASC.UG.10.RT08                 ..........-
ASC.UG.10.RT11                 ..........-
COL.CM.x.CGU1                  ..........-
COL.UG.10.BWC01                ..........-
COL.UG.10.BWC07                ..........-
DEB.CM.04.SIVdeb04CMPF3061     ..........-
DEB.CM.99.CM40                 NNC.......-
DEB.CM.99.CM5                  ..........-
DEN.CD.x.CD1_CMO580407         ..........-
DRL.DE.11.D3                   .Q........-
DRL.DE.11.D4                   .Q........-
DRL.DE.11.D5                   .Q........-
DRL.DE.11.D6                   .Q........-
DRL.x.x.FAO                    .Q........-
GRV.ET.x.GRI_677_gri_1         ..........-
GSN.CM.99.CN166                ..........-
GSN.CM.99.CN71                 ..........-
LST.CD.88.SIVlhoest447         ..........-
LST.CD.88.SIVlhoest485         ..........-
LST.CD.88.SIVlhoest524         ..........-
LST.KE.x.lho7                  ..........-
MAC.US.x.251_BK28              ..........-
MND-1.GA.x.MNDGB1              ..........-
MND-2.CM.98.CM16               ..........-
MND-2.GA.x.M14                 SK........-
MND-2.x.x.5440                 SKSGAG....-
MNE.US.x.MNE027                ..........-
MON.CM.99.L1_99CML1            ..........-
MON.NG.x.NG1                   ..........-
MUS-1.CM.01.CM1239             ..........-
MUS-1.CM.01.SIVmus_01CM1085    ..........-
MUS-2.CM.01.CM1246             TT........-
MUS-2.CM.01.CM2500             ..........-
MUS-3.GA.09.09GabOI81          ..........-
MUS-3.GA.11.11GabPts02         ..........-
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.CI.79.SIVsmCI2             ..........-
SMM.LR.89.SIVsmLIB1            ..........-
SMM.SL.92.SIVsmSL92A           ..........-
SMM.SL.92.SL92B                ..........-
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                 TASSATVTRP-
TAN.UG.x.SIVagmTAN1            ..........-
VER.DE.x.AGM3                  TGASD.....-
VER.KE.x.9063                  VGANA.....-
VER.KE.x.AGM155                VGAGN.....-
VER.KE.x.TYO1_patent           ..........-
WRC.CI.97.97CI14               ..........-
WRC.CI.98.98CI04               ..........-
WRC.GM.05.Pbt_05GM_X02         ..........-
```

100
102
101
103
102
102
100
102
102
115
115
115
102
102
102
112
111
101
115
100
108
100
103
124
83
83
83
116
72

131
131
139
134
129
129
102
99
99
86
87
90
99
111
97
136
125
125
125
125
125
95
105
135
111
108
109
109
131
108
113
123
127
135
105
112
102
99
109
102
117
99
108
116
114
107
116
107
134
132
133
106
133
103
116
111
111
128
102
120
120
120
101
132
133
130

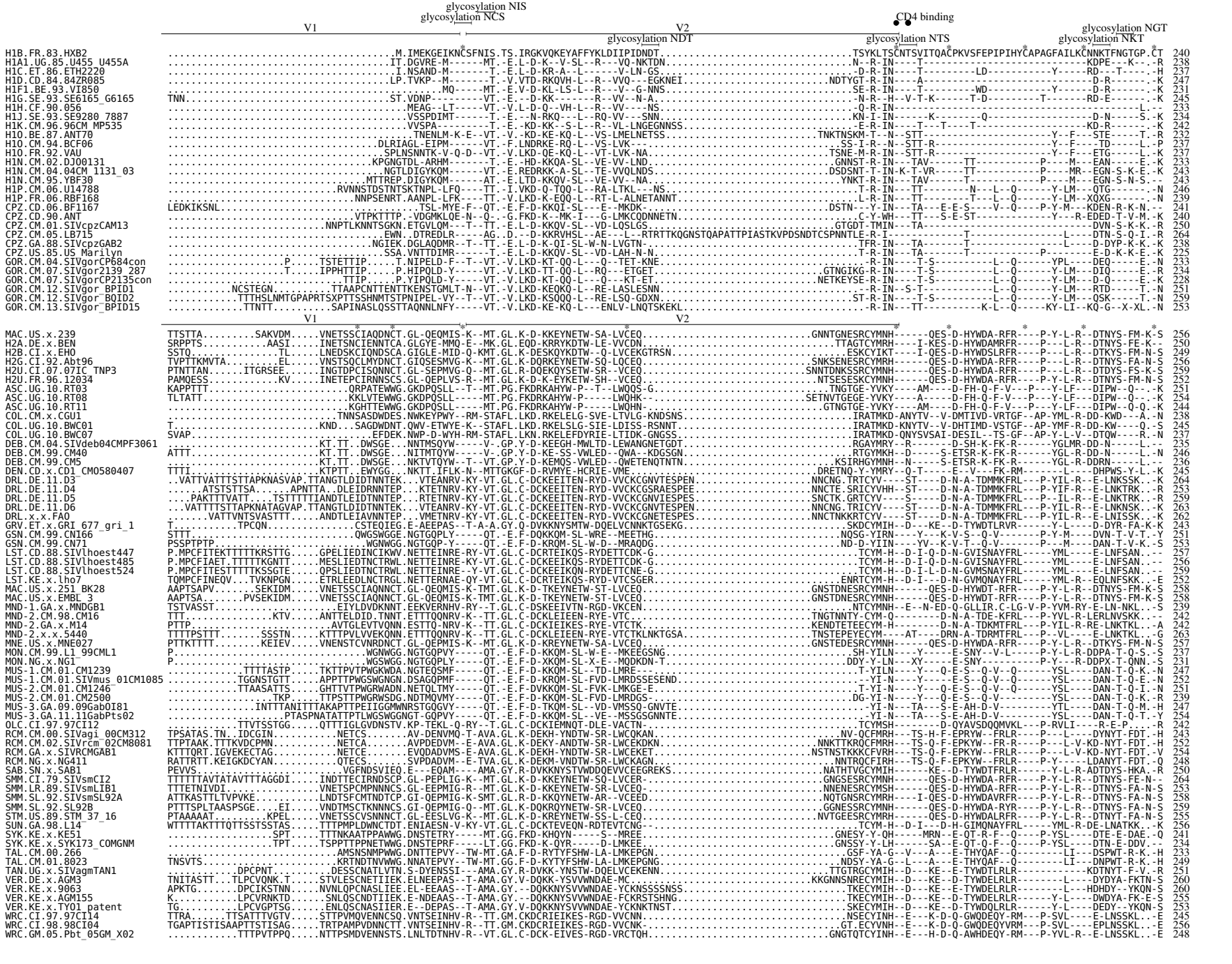
PLV Proteins

```
                            Rev start               exon 1 end exon 2 start   NLS                                       Leu-rich effector domain
H1B.FR.83.HXB2              MAGRSGDSDE.....ELIRTVRLIKLLYQ...SNPPPNP..EGTRQARRNRRRRWRERQRQIHSISERILGTYL.......GRSAEPVPLQLPPLERLTLDCNEDCGTS...........................GTQ.......GVGSPQILVESPTVLESGTKE..  116
H1A1.UG.85.U455_U455A       --R---NP--.....D-LKA--I------.....---C---..R-S----K-------A-----D-L-----SDC-........--P---------I---R---S-S---..........................---QPQGTET---G---S---SA--G----N..  123
H1C.ET.86.ETH2220           ----------.....--LKA--I--I---.....---Y-T-..---------------A------TL-----SNF-........--P-------------N---S--S---..........................---QSQGTTE---N-.................  107
H1D.CD.84.84ZR085           ----------.....D-LTA-----I---.....------S-..---------------A---Y----G---S---........---E-----------IN-N-S------..........................-------------------S----AI---R-E-..  116
H1F1.BE.93.VI850            ---------T.....--LKA-KC--I---.....---Y-K-..---------------A-----RAL-D---SSC-........---E-------------HIN-S----QG..........................PEE..........----S--SG--HA---------..  116
H1G.SE.93.SE6165_G6165      ------ST--.....--L-A-KA--I---.....---Y-P-..---------------A-----SA------TA--........--P-------------H---S--S---..........................---QPQGTET---RS.................  107
H1H.CF.90.056               ------A--T.....--LQVCKI--I---.....---C-E-..T--------------A-----RE------TSC-........--PP---T-----------N-S------..........................-EK........-E-----SL--S-I-GT----..  116
H1J.SE.93.SE9280_7887       ---------D.....Q-LLA-----I---.....---Y-K-..N-S------------A--N--D------PSSC-........--P---------I---R---S----N-..........................--------------D---SG-PCM--GA----..  116
H1K.CM.96.96CM_MP535        ----R--P--.....Q-LT---T--I--E.....---Y--L..--S--T---------A--K--S-----L-SAC-........-------------I-K-N-N-D--P-KG..........................TEG........L-----S--PC----------..  116
H1O.BE.87.ANT70             -----E-D--.....Q-LQAIQI--I---.....---Q-S-..R-S-N--K-------R--A-VDTLAA-V-A-VV.......HGPQNNNIVD-----Q-SIRDP-GDQL-..........................EAWTVDPRAEDN......................  102
H1O.CM.94.BCF06             -----DED--.....QQ-LQAIQI--I---.....---H-T-..G-S-N--K-------R--A-VDTLAT---T-VV.......HG-QDNNLVD-----Q-SIRDP-VDQLP..........................--WTVDPGTKDN......................  103
H1O.FR.92.VAU               -----EED--.....QQ-LQAIQI--I---.....---H-TT..S-S-N----------R--A-VD--AA-V-ASVV.......HGPQDNNLVD-----Q-S-RDP-VDQ-P..........................--WTVDHRAEDN......................  103
H1N.CM.02.DJO0131           ------VN--.....--L-A--I--I---.....---Y--NNNQ--------------T-----RE--Q---D-C-........-GLQ---D-P----D-----PA--S-AP..........................--EPHQGTATTE......................  106
H1N.CM.04.04CM_1131_03      ------VN--.....--LTA--I--I---.....---Y--N..Q--------------A-----RA------S-C-........-GLQ--GDPP----X-----XT--S-AP..........................--ESQQGTATTE......................  104
H1N.CM.95.YBF30             ------VN--.....--L-A--V--I---.....---Y--S..K--------------A-----RA------SSC-........-GPP---D-P----D-----TE--S--P..........................--ESQQGTATTE......................  104
H1P.CM.06.U14788            -----DED....LR---T-I-I--I---.....---S-AR..GPS-NST--------S-----DQ-AG--IASH-........--PE-LGGAD--DISQ-HIEDQGPPDN-..........................VDTPGPTTEQ.......................  102
H1P.FR.06.RBF168            -----DED....LR---T-I-I--I---.....---W-ER..GSS-NST--K-----S-----DQ-AG---ASR-........--PE--GGAD--D-S--HIGDQGPPNNP..........................VDTPGPTADQRA.....................  104
CPZ.CD.06.BF1167            ----EEEN.......-LQAIKI--I--E.....---Y--S..----S----------I----VEG--S---A----........--LPTAP-VD--D-SK---APLDIDQVD..........................PSVEPIAVGDCETIAAQAHTP-DTISTP-GNLG  122
CPZ.CD.90.ANT               ----EELEGT...DEQ-LKA-KI--I---.....---Y-K-..A-S-A---K---KK--D-VEGLAA-V-R-LV.......-GPD-HN-VD--D-QN-S-EPLGSPP-P..........................LQNYFQTVQPSGEQS-HPLDNNQN.........  118
CPZ.CM.01.SIVcpzCAM13       -----EHD-A.....R-LQA-KI--I--E.....---Y-S-..----N----------A-----RE--N-V-AAH-........--PPQ-GH-E--E-DK-S-H-L-TV-KV..........................--TVNQHPQTDT-ETDSTEGDNPI-GKRI-N..  123
CPZ.CM.05.LB715             ---------A.......--L-AI-I--I---.....---F--..----A----------A-------VA--V-A-C-........E-PEG--S-P-------S---Q--F--C..........................--TPAEGTET---D-T-S..............  110
CPZ.GA.88.SIVcpzGAB2        -------E--.....R-L-A-KA--I--D.....---Y-SN..R---S----------A-----CE--Q--FSSL-........--HP--DD-P-------S-GPL-PA--I..........................--TGPERTETTQ-ASEPAERDR-I-AAR--N..  123
CPZ.US.85.US_Marilyn        ------GDADE....Q-L-A--I--I--A.....---F--N..S-------------KN----NE--Q---S-C-........---S--PD-P---IGG--INSE--STNP..........................--EHQQGTATVE......................  105
GOR.CM.04.SIVgorCP684con    -----DED....LQ--LTKI-I-RI---.....---W-E-..GPS-NST-------KG-----DQ-AG---ASH-........---QN-GCVD--DISH--IGDQG-NTQP..........................DRXPDQTTEKSE......................  104
GOR.CM.07.SIVgor2139_287    -----DED....LQ--LTKI-I-R----.....---W-E-..GPS-NST-------KG-----DQ-AG---TSH-........--YQN-GCVD--DISH--IGDQG-NTQP..........................DRVPDQTTEKSE......................  104
GOR.CM.07.SIVgorCP2135con   -----DED....LQ--LTKI-I-R----.....---C-E-..GPS-NST-------KG-----DQ-AG---AAH-........---QN-DCVD--DISH--IGDQGNNTQP..........................DRAPDQTTEKSE......................  104
GOR.CM.12.SIVgor_BPID1      -----DED....LQ---T-I-I-RI---.....---W-EQ..GPS-NSN--------S-----DQ-AG---T-H-........--PE-LGGVD--D-GQ-HIGDQGPPSNP..........................VDTPETPAEQRT......................  104
GOR.CM.12.SIVgor_BQID2      -----EED.....QQ-LQAIKV--I---.....---F-T-..H---N--K-------R--A-VD--AT---ASLI.......PGPQDNN-VE--S--Q-SIQDP--TKPP..........................--GTADPRTKD-*-L-NT..............  108
GOR.CM.13.SIVgor_BPID15     -----EEE....LQD--T-I-I-RI---.....---W-EQ..GPS-NST--------D-----DQ-AG--ITNR-........---E-LGGVD--DISH-HIGDQRPPEHP..........................VDTPETPTEQ-N......................  104
                            Rev start                exon 1 end exon 2 start

MAC.US.x.239                -SNHEREEE.......-RKRL---H--H-...T--Y-TG..P--ANQ--Q-K----R-WQ-LLALAD--YSFPD.......PPTDT-LD-AIQQ-QN-AIESIP-PP-N..........................TPEALCDPTEDSR---D...............  107
H2A.DE.x.BEN                -SE-ADEEG.......-QGKL--LR--H-...T--Y-QG..P--ASQ-------R-RQWLRLVALANKLCAVPD.......PPTDS-LDRAIQH-Q---IQELP-PP-D..........................LPES...NSNQ-LAET...............  103
H2B.CI.x.EHO                -..NARER-.......-QKGL--LH--H-...T--Y-QG..P--ASQ--------KQ-GL--LALAD--HPLPD.......SPTEG-LD-AIQR-QN-IIKDLPNPP--...TPTAQASTCIPPIWDQLVPRSNPSSSQGCGRDSCERGEDLVGSPQESGRRDHCNTQED  146
H2G.CI.92.Abt96             -.SNLEEEE.......-RKRL---HF-H-...T--Y-QG..P--ANQ------K-KQ-WL--LALAD--YSFPX.......PPAR--LD-AIQQ-QS--IQDXP-SP-A..........................VPKNLQDPPSN.......................  100
H2U.CI.07.07IC_TNP3         -..NNPEEE......VR-RL---HF-H-...T--Y-RG..P--ANQ---------Q-WL-LLALAN--SSFPD.......PPADS-LD-AIQQFQS-AIAELPEPPSG..........................LPESTDSNQRTSEA....................  102
H2U.FR.96.12034             -..SNPEEG......VQ-RL---H--H-...T--Y-TG..P--ANQ---------Q-WL--LALADK-YTFPD.......PPTDS--D-AIQQ-QG-AIQDLPEPPAV..........................VPEHQRTSAP........................  98
ASC.UG.10.RT03              ---SGREE-P..ELRQ-L-AC-I-RT--D..---F-PA..S---ST---------R-----RA-A--V-SSC-........--LESDR-VD--DIGQ-R-ADPSPVER-D..............................................  96
ASC.UG.10.RT08              ---SGR-E-P..ELRQ-L-AC-I-RT--D..---F-PA..S-P-ST---------R-----RA-A--V-SSC-........--PESDRSVD--DISQ-R-ADPPPVEQ-I..............................................  96
ASC.UG.10.RT11              ---SGREE-P..ELRQ-L-AC-I--I--D..---F-PA..S-S-S----------R--H--RA-A--V-SSC-........--PESDRLVDP-DIGQ-R-ADSSSVEQ-N..............................................  96
COL.CM.x.CGU1               ....................I-R-RV-HK...H--W-R-.GV--ARQ--RNTQ-A-KQ-Q--LA-DG-VVEQLV.......ASLQQWD-YCPASTS.............................ANNLRQEEECLADSSVSIGTQGSKP-DS....  93
COL.UG.10.BWC01             ................IR-IR--TAATFV..FTD-W-E-.GV--ARQ--RA-Q-A-RQ-Q-ARA-A---VE-LC.......SSVQQWA-YSHVSE..............................ARDLQHQEDC-PDSLE-TGTEVSRPIDS....  97
COL.UG.10.BWC07             ................TK-IRS-H-ATNH.YFAD-Y-DT..V--ARQ--RA-Q-F-KQ-Q--LA-G--VVD-LV.......QNLQKWD-YSPGGATVT..........................RWRDPPAQEDASPGSLASIGTQECKPGDSSDTS  105
DEB.CM.04.SIVdeb04CMPF3061  --HAG-RGE....LQN-L-AIAA--I---.....---F-EG..GSRASR--R--.---QH-ETVD-LAA---ES--........--P-AA-E-A--D-QE-H-AAD-GSRDP..........................ARTNI.............................  97
DEB.CM.99.CM40              --HAG-RGSAEENTRQ-LKVIS---I---.....---Y-KG..G-SAST--R--QK--R--G-VD--A----QSH-........-G-PTTADVA--D-SQ-H-AD.......................................................  90
DEB.CM.99.CM5               --HAG-RG-AD...QA-L-MI-I--S---.....---Y-KG..G-SAST--R--QK--R--E-VDD-AQ---EAH-........--HLTASDVA--N--E-R-AD.......................................................  87
DEN.CD.x.CD1_CMO580407      ---LQERD......Q--VKA-----K---.....---Y-T-..--SKS---R---K--RQ-A-VVA------EYH-........---LTAADVE--D--K-SVSD.......................................................  84
DRL.DE.11.D3                -SAGPEREP....PAWFQEYL--VTR-W-...DDQL.....PQ-ARQ-KR--A-Q-RVEY--RALQA--FKILE........QRNSRDLVE..GI--IH-AEERESSS..................................................  86
DRL.DE.11.D4                -SAGPEREQ....PPWFQEYL--ATR-W-...DDLL.....PQ-ARQ-KR--A-Q-RVEH--RALQA---QVLE........QRNSRDLVE..GI--IH-AEEL-SSS..................................................  86
DRL.DE.11.D5                -SAGPEREQ....PPWFQEYL--ATR-W-...DDQL.....PQ-ARQ-KR--A-Q-RVEH--RTLQA---QALE........QRNSRDLVE..GI--IH-AEERESSS..................................................  86
DRL.DE.11.D6                -SAGPEREP....PAWFQEYQ--VTR-W-...DDQL.....PQ-ARQ-KR--A-Q-RVEY--RALQA--FKILE........QRNSRDLVE..GI--IH-AEERESSS..................................................  86
DRL.x.x.FAO                 -SAGPEREP....PPWFQEYL--VTR-W-...DDQL.....PQ-ARQ-KR--A-Q-RVEH--RTLQA---QSLE........-RNSRDLVE..GI--IH-AEERESSS..................................................  86
GRV.ET.x.GRI_677_gri_1      -SLGKEE..........KQALKI--T--G.....---Y-QF..S--ARQ--RA-Q---KQ-Q--DK-AG-V-N-FE......DQQLVAQLQELQLENKD-V-QHLP-PPHIH......QDSSGIPAVWAPATPRG....SNRACSSSGE-CEGSLGQTGCYCPIRLSGSHQQ  139
GSN.CM.99.CN166             --DHARGN-Q..KLQN--LAC----T-HR.....---Y-.TS.S---S----------A--G-VRE--N---ESLV........--PE--GD-D--D-GQ-S--SPW-REVP..........................VGTAAESNPESNPASETAQG....SS-QSDSLH  124
GSN.CM.99.CN71              --DPANGR-Q..KLQN-LLAC---T--K.....---Y-TA..S---S----KK----A----VRQ--H---ESLV........--PE-SGN-D--D-GH-S--SP--RQVP..........................VGAAAEGVPTSETAGG........SG-QSSSLY  120
LST.CD.88.SIVlhoest447      -STGGEK.........-SP-YLKIS-I-WT...GLQVGEKREPR-ARQ--RD-E-Q-KHLH-LRTVQ---FHSTVERGL..ERAFTRLTVCDS-EVDQGLWNSSPQPCVAVPF.VARPFSDPFLPTWATSSESLQQTG-GKRSEDCELELEQD-KEQKRHISSSCKKPQTG  156
LST.CD.88.SIVlhoest485      -STGGEE.........--P-YLKIS-I-WT...GLQVGERREPR-ARQ--RD-E-Q-RHLH-LRTVQ---FHSTVERGL..ERAFTGLTVCDS-EVDQGIWNSSPQPCVAVPF.VARPLSDPFLPTWATSLESQQRMD-GKRSEDCALELEQD-KEQKRHISSSCERHQTG  156
LST.CD.88.SIVlhoest524      -STGPEG.........--PTYLK-SRI-WT...GLQVGERREPR-ARQ--RD-E-Q-KHLH-LRAVQ---FQATVERGL..ERAFTGLTVCDS-EVAEGVGNTAPHPRVAIPH.LAGTYSDPFLPPWATSLANPQQIP-GECSENCELDQEQE-KEQRRAIKN........  148
LST.KE.x.lho7               -STGN--.........--P-YL--SRI-WT...GLQVGERREPR-ARQ--RE-S-Y-DYLH-LRAVQ---FQATVERGL..ERAFTRLAVSDS-EVAQGRGNTPPITSVAEPQ.LAVAFVDPFLPKWATPLADQQQMD-GKRSEDSELAQ-EM-KEQRTVIEH........  148
MAC.US.x.251_BK28           -SSHEREEE.......-RKRL---H--H-...T-SY-TG..P--ANQ--Q------R-WQ-LLALAD--YSFPD.......PPTDT-LD-AIQQ-QN-AIESIP-PP-N..........................TPEALCDPTK-SR---D...............  107
MAC.US.x.EMBL_3             -SSHEREEE.......-RKRL---H--H-...T-SY-TG..P--ANQ--Q------R-WQ-LLALAD--YSFPD.......PPTDT-LD-AIQQ-QN-AIESIP-PP-N..........................TPEALCDPTKDSR---D...............  107
MND-1.GA.x.MNDGB1           -STGNVY......QELIR-YLVVV-K--E...GL-AEPI..PQ-ARQ--R-KQQL-T-RA-LRELEG---KQI-DRG....PDQLCQGVTN-ALA-KSESSN.......................................................  87
MND-2.CM.98.CM16            -PTET--L-....RDFWKKYQKIVQR-WEG..RLS-SPQ..PQ-ARQ-KR--A-L-KTEY--RQLQS--WEALEQRA....NQRLAEDFKG-RVA-ES-SSN.......................................................  90
MND-2.GA.x.M14              -STEP-ELG....RDFWEKYH-IV-Q-WE....LS-TSQ..PQ-ARR-KR--AQL-RTEH-VRELQ---WQ-LEARA....NKELVEGVNN-HLA-ESAGSN.......................................................  88
MND-2.x.x.5440              -STDR--L-....EGFWRKYQAIV-Q-WE....LS-APQ..PQ-ARQ--R--A-L-KTEH-VRELQT--*EALEARA....SQELAEKVDA-HLA-EPPSSN.......................................................  87
MNE.US.x.MNE027             -SSHAE.......EE--R-RL---H--H-...T--Y-TG..PS-ANQ---------Q-WQ-FLALAD-VYSFPD.......PPTDT-LD-AIQQ-QN-AIESIP-PP-N..........................IPEILHDPTESPR---D...............  107
MON.CM.99.L1_99CML1         ---NGR-EE....RRL-SLALAAVRI-QE.....---Y-I-..R--AN----------RQ-L--A------FYH--........---Q--C--DI-D----SISDLP-PPE-..........................VPEAATPPAHTPAPTVGKP.............  112
MON.NG.x.NG1                ---AERGAA....VQDIL-LAIGAIRIIRE.....---Y-S-..S---S----------A--E-VX-LC---FNH--........D-LPR-NS-D--AI-G-R--DLPEPPE-..........................AQPPDPSSPRP-PTASGSQ.............  113
MUS-1.CM.01.CM1239          ---A-ENV......QD----C-I-RI---.....---Y-TS..R-S-A---------QQ--D--RA------VAV-........--PPQ-SD-V--DISQ-SINPSL-TPEG..........................SSPDTEVVSTT-DRD.................  106
MUS-1.CM.01.SIVmus_01CM1085 ---V-END......R----AC-A-QI--K.....---Y-TA..D---S---------KRK----RE------LSL-........--PPQ-SD-E--D-NK-S-HPLVATSE-..........................SPPDTEGVNKI-VHA-GTD.............  110
MUS-2.CM.01.CM1246          ---A--ND......Q----FC-Q-RI--D.....---Y-TV..SR--AT-K--K---KK--A-........--VAVF.......R-PPQ-ND-D--D-DQ-R-EPLVVDSKN..........................LPPPDKGASKS-D-A.................  99
MUS-2.CM.01.CM2500          ---TGR-I......E--LKACQT--I--N.....---Y-S-..G--KN----------Q--C--RAL-----VS--........--PPQ-DSVE--D--K-S-KPLVADPAD..........................SPSPPEVEHLS-AQGESNQHP-SGSP.......  117
MUS-3.GA.09.09GabOI81       ---KEE-L......AA--KACKV--I--H...-K-Y-S-..G--KT----K-----R-----SALA-----VC-........E-PPRLDSVD--DI-K---NPCVGDSGK..........................CSEDLDTAKSPLAN-EGTGST-..........  113
MUS-3.GA.11.11GabPts02      ---VER-A......EG---AC-I-RI--S.....---Y-ST..R-S-S---------QQ--Q--RE----V-SSL-........--HP--GD-E--D--K-S-EPLVVDEAG..........................-NSAGTASQSSKRDNRESP.............  110
OLC.CI.97.97CI12            ............#ST-SP-KR-KKVVSH-QDSRDTY-PG..Q--ARQ--R--Q-VLREHL-RVELYH-VYRDIE.......SKLERLTITPVEGSSESHSLDP-ELA-RL......ENLKTEETTWDQFAAE.....E-EKGKSSTE-S-AIGERRS.............  124
RCM.CM.00.SIVagi_00CM312    -LLGEEE......DR-IR-RL---H--HL...T--Y-Q-QEG--ASQ--R--VQ--R-WL--IQLA---FLYPD.....PPPNSD-LSNAVEQ-NQ--IAELPEPPIN..........................PFTNPASSVVDSS...................  107
RCM.CM.02.SIVrcm_02CM8081   -LLGEE.......DQ-IR-RL---NF-L-.....---Y-Q-..G--ANQ--R---L--R-WL--AQLA---FLLPD.....TPPDRDLIEGAVEQ-QQ-AISDLPEPPVN..........................PFSNPADSAVDNS...................  104
RCM.GA.x.SIVRCMGAB1         -LPGQD.......EED-RKKI---NF--L...I--Y-QH..G--ANS--KK-----R-WQ-VQQLA----LDST.....DPPVEQDLDAAIAD-QK-Q-NNLPEPPVD..........................FS..............................  93
RCM.NG.x.NG411              -LLGEEEEA....DQ-IR-RI---H-IHL.....---Y-QS..G--ANQ--RK--Q--R-WT--LQLA---FLYPD.....TPPDRD-LQEA-SN-QQ---SDLPEPPVN..........................PFSNPSSFAVDRS...................  107
SAB.SN.x.SAB1               -SLGQE..........--L-RF-I--F--T...T--Y-PG..Q--ARQ--RA-Q--AKQRQ-VIHLA----E-P.......VSQIDHLAQEFDQLVLDNLQQPPSLPPGH..........................PTENQTANS-S.....................  99
SMM.CI.79.SIVsmCI2          -.SDQEGEE.......-RKRL---HF-H-...T--Y-QG..T--ANQ---------Q-WL--LALAN--YSFPD.......PPANSSLD-AVQQ-QG-SIEDLPEPPRN..........................LPTSLSSSQEQDQA..................  103
SMM.LR.89.SIVsmLIB1        -...HE-RE........-Q-RL---H--H-...T--Y-TG..P--ANQ---------Q-WL--LALAD--YSFPD.......PPADT-LDVAVQQ-QS-AIQDLPEPPAT..........................VPETLRSNTTSQR-N.................  102
SMM.SL.92.SIVsmSL92A        -SQ-GQEEE-......-RKRL---HF-H-...T--Y-QG..P--ANQ--------KQ-WY--LALAN--YSFPD.......PPADTSFD-AIQQ-QS--IEDLPEPP-N..........................IPETLKPDKQDP....................  103
SMM.SL.92.SL92B             -..QNPEEE.......-R-RL---H-IH-...T--Y-AG..P--ASQ--------KR-WI--VALAD--YSFPD.......PPADTDID-AIQQFDS-SIQSLPEPP-T..........................VPETLRDQSAD.....................  99
STM.US.89.STM_37_16         -..SDQEEE.......-RKRL---QF-H-...T--Y-QG..P--ANQ---------Q-WN-LLALAN--YSFPD.......PPAST-LD-AVQQ-QG-SIQDLP-PPPN..........................LPKDLQDTAEN.....................  99
SUN.GA.98.L14               -STGDDS.........INQYL-IS-R--E...GL.A-GNL.PQ-HRQ--R--D-E-KNLH-LRAVQ---FA-T-DSRL..GRAFERLSVSDSSQVAESLGNSPSTKHLPPAK.FLVAPTYDFLPSWATPLADPQRLA-FAPYSGYEQDQERV-NQQGESIIVSE-K-...  150
SYK.KE.x.KE51               -S--EREDT....QEQ-L--LLR-AQQLE..AAA-Y-L-..Q-PSR---RN-N-Y-QL-A-RLYVQQ--FE-IA.......R----TLEQSFGE-QITD..............................................................  84
SYK.KE.x.SYK173_COMGNM      -PDQGS.........EQ-AVFL-M-AH-QE.....-Y-G-.---P--T--R--Q--Q-RT-RLYLQQ--FEAIF.......-SRTAALEDS-QQ-QISD..............................................................  77
TAL.CM.00.266               -SLTG-SGRG...EDP-LAAI-R-RI---.....---Y-EQ..Q-S-S---R-Y---KA--L-VEQ-AS---EH--........---DRSDQIH--D-SQ-S-ADLGP-ESC..........................ES-EPTGSASHIRL-TTPPD-SSSDSPVWVQ..  125
TAL.CM.01.8023              -SLTGESGKG...EDP-LAAI-R-RI---.....---Y-EQ..Q---S---R-Y---KA--E-VEN-AS---CH--........-GPQQ-PQ----D--K-S-AELGPPEAP..........................ESDHPTGSASHIRL-EN.................  111
TAN.UG.x.SIVagmTAN1         -SLG-E-.........-R-IIQI-RI--H.....--QY-PG.---ARQ--R--Q-FQQQ---VAAL----FIAQQRRDPSGGESL-AAFDQLVLDNQQ-VIETLP-PPQEP..........................HDSSSTA.........................  103
VER.DE.x.AGM3               -PLG-EE.........R-LL---AF-NK..N--Y-PV.---ARQ--RA-----QA-E-LRALA---WHSRV......EEQLVQAIDQLVLDQQH-AIQQLP-PPS-S..............................................  89
VER.KE.x.9063               -PLGPEE.........RKVFS--AY-NR...T--Y-PV.---ARQ--RA----------LRFLA---WNETK......ERQLEQTLDQLALDVQH-AIQHLP-PPN-S..............................................  89
VER.KE.x.AGM155             -PLGPEE.........R-LL---AF--R.....---Y-SV.---ARQ--RA----KN--K--YALA---W--RQ......EDQLVQAIDQLVLDTQH-VTQQLP-PPSQA..............................................  89
VER.KE.x.TYO1_patent        -PLGPEE.........R-LL---W---S...T--Y-PSG.---ARQ--RA-----QQ-D--RVLV--LQ...........EQVYA-DRLADEAQH-AIQQLP-PPH-A..............................................  84
WRC.CI.97.97CI14            ...............IVLTARK-SQIQQ-.LLQE--QQLG.R--ARQ--RN---LQQYRN-RLAVQ---CNSL-.......EQL-NLTIQH--TPPSHCVESAFVPPVV............DKRWVESSQRHQ...WKSIWSDTQEDCKEQEGEYAGRGISNWQQ-DSS  130
WRC.CI.98.98CI04            ...............IVPTAR--VRVQQK.LQQE--QQLG.R--ARQ--RS---QQYFR--RLAVQ---CNNLV.......EQLDKLLICS--NPPSNSVESPCVPLLV............DPAWVESSQKNQ...WSSACMHTQEDYKEQEGNNGERDIRK.......  123
WRC.GM.05.Pbt_05GM_X02      ...............IIHTAR---RISQT.LQQESY-E-G.R--ARQ-------FHSHR-IRLAVQ---AD-L-.......EQLGNLTLQH--EPPAAK-QPS-LSVDR............NTEWVESSQKTL...WSSVCCATGEDCCASAEKKGEGDIHAWQQRDKQ  130
```

```
                                                Rev end
H1B.FR.83.HXB2                  ................*
H1A1.UG.85.U455_U455A           ................-
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            ................-
H1O.BE.87.ANT70                 ................-
H1O.CM.94.BCF06                 ................-
H1O.FR.92.VAU                   ................-
H1N.CM.02.DJO0131               ................-
H1N.CM.04.04CM_1131_03          ................-
H1N.CM.95.YBF30                 ................-
H1P.CM.06.U14788                ................-
H1P.FR.06.RBF168                ................-
CPZ.CD.06.BF1167                NSN.............-
CPZ.CD.90.ANT                   ................-
CPZ.CM.01.SIVcpzCAM13           ................-
CPZ.CM.05.LB715                 ................-
CPZ.GA.88.SIVcpzGAB2            ................-
CPZ.US.85.US_Marilyn            ................-
GOR.CM.04.SIVgorCP684con        ................-
GOR.CM.07.SIVgor2139_287        ................-
GOR.CM.07.SIVgorCP2135con       ................-
GOR.CM.12.SIVgor_BPID1          ................-
GOR.CM.12.SIVgor_BQID2          ................-
GOR.CM.13.SIVgor_BPID15         ................-
                                                Rev end
MAC.US.x.239                    ................-
H2A.DE.x.BEN                    ................-
H2B.CI.x.EHO                    QTRG............-
H2G.CI.92.Abt96                 ................-
H2U.CI.07.07IC_TNP3             ................-
H2U.FR.96.12034                 ................-
ASC.UG.10.RT03                  ................-
ASC.UG.10.RT08                  ................-
ASC.UG.10.RT11                  ................-
COL.CM.x.CGU1                   ................-
COL.UG.10.BWC01                 ................-
COL.UG.10.BWC07                 EVATICDEC......RA-
DEB.CM.04.SIVdeb04CMPF3061      ................-
DEB.CM.99.CM40                  ................-
DEB.CM.99.CM5                   ................-
DEN.CD.x.CD1_CMO580407          ................-
DRL.DE.11.D3                    ................-
DRL.DE.11.D4                    ................-
DRL.DE.11.D5                    ................-
DRL.DE.11.D6                    ................-
DRL.x.x.FAO                     ................-
GRV.ET.x.GRI_677_gri_1          SKKSAARP.........-
GSN.CM.99.CN166                 ................-
GSN.CM.99.CN71                  ................-
LST.CD.88.SIVlhoest447          TKEEMEIQIQRKKWVSQ-
LST.CD.88.SIVlhoest485          TKEKMEIQIQREKWVSQ-
LST.CD.88.SIVlhoest524          ................-
LST.KE.x.lho7                   ................-
MAC.US.x.251_BK28               ................-
MAC.US.x.EMBL_3                 ................-
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             ................-
MON.NG.x.NG1                    ................-
MUS-1.CM.01.CM1239              ................-
MUS-1.CM.01.SIVmus_01CM1085     ................-
MUS-2.CM.01.CM1246              ................-
MUS-2.CM.01.CM2500              ................-
MUS-3.GA.09.09GabOI81           ................-
MUS-3.GA.11.11GabPts02          ................-
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.CI.79.SIVsmCI2              ................-
SMM.LR.89.SIVsmLIB1             ................-
SMM.SL.92.SIVsmSL92A            ................-
SMM.SL.92.SL92B                 ................-
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                KQ..............-
WRC.CI.98.98CI04                ................-
WRC.GM.05.Pbt_05GM_X02          KQQHQGNYCQTLHKP..-
```

```
                                                                                  phos phos
                                transmembrane domain                              cytoplasmic domain
                                Vpu start                                α-helix          α-helix                Vpu end
H1B.FR.83.HXB2                  TQPIP.....IVAIVALVVAIIIA.IVVWSIVIIEY....RKILRQRKIDRLIDRLIERAEDSGNESEG.EISALVEMG..............VEMGHHAPWDV.DDL.*  82
H1A1.UG.85.U455_U455A           MT-LE......-W--TG-I--L-L-.----T--G--.......................................................-  30
H1C.ET.86.ETH2220               MVDLLAKVDYRIV---FI--L---.----T-AY--------L----R-----K-TR----------D-.DTEE-ST................D--NLRLL--.N--.-  87
H1D.CD.84.84ZR085               M-SLQ.....-L------L-L---.----T--F--------R-K------W----IR--E----------.DKEE-STL.......................D----.-  83
H1F1.BE.93.VI850                MSYLLA.....IG-A--I--L---.----T--Y-------.K-LV-----NK-YK-IR------------.DAEE-AAL................G---PFI-G-I.NN-.-  82
H1G.SE.93.SE6165_G6165          M-SLV......-L---G-I--F-A-.----T--F-------.-E-RK-KR-GK-L--IR----------D-.DTEE--TL.......------DFD--VG.-N-.-  82
H1H.CF.90.056                   MYILG......LG-G----TF----.VI--T--Y-------.K-LV--K------E-IG----------D-.DTEE-SKL...............M----LNLGY-A.--.-  81
H1J.SE.93.SE9280_7887           MI-LQ......A----FI---FL--.-GM-T--Y-------.K-L-------K----IR----------D-.DTEE-ADL................-R-P-DL-N-N.--.-  82
H1K.CM.96.96CM_MP535            MVSLA......IS-------L-L--.-I--T--Y---------LVK-KR-NW----IR------------.DAEE-ADI...............G-L--LILGNI.-N-.-  81
H1O.BE.87.ANT70                 MHHRD......LLAIIIIS-LLFINVIL-GFILRK-LEQKEQDRKE-E-LERLR-IR-IRD--DY--N-E-EQEVM.................DLVLSHGF-NPMFEP-  86
H1O.CM.94.BCF06                 MHQRD......LLAIIALS-LCLIS-I--I-NLRN-LEQRKQDR-EQE-LERLR-VR-IKD--DYD-N-E-EQEVM.................DLM-SHGF-NPMFEW-  86
H1O.FR.92.VAU                   MHHQG......LLVLIVII-LLLINVLL-MFNLRA-LEQKKQDRTA-E-LER-R-IR-VED--DY--N-E-EQEVR.................DLVLDFGF-NPMFEL-  86
H1N.CM.02.DJO0131               ML..E......WGFA--G------.VII-VLLYK------.K-LKL-E--EQIRQ-IRD-T----S-ND-.DADW-..................AILLS-DKF-..NWV.-  75
H1N.CM.04.04CM_1131_03          ML..E......LGFI-FG-------.VII-VLLYK------.KE-KL-E--EKIRQ-IRD--KE--K--D-.DAEW-GDKESVG......DDEWLAILLSPDKL-..NWV.-  87
H1N.CM.95.YBF30                 ML..S......LGFI--GA-VS--.VI--ALLYR------.K--KL-E--KHIRQ-IR--E-------D-.DAEW-DG..........DEEWL-TLLSSSKL-QGNWV.-  84
H1P.CM.06.U14788                MH-RD......E-VLIIAGVLLLCC--X-GK-LLLVLK....ERERD-FVQRLA-WR-GQ--E-Y--NEE-EEQ-R..................-L-NLLGF-HVL....-  79
H1P.FR.06.RBF168                MH-RD......E-VLIIAGVLL-CI----GK-LLFVLK....ERERD-FVQRLV-WR--Q--G-Y--NEE-EEQ-R..................-L-NLLGFGNIL....-  79
CPZ.CD.06.BF1167                MLWQ.......FLQW-QYLGWGGA--I-I-ALLVIRKAYLY-KQVQEE--IEQAI-D-YQRR-STDS-I-EDENG..................LINWMDP-NYFL--.-  82
CPZ.CD.90.ANT                   MTN-F.....EY-FL-FSIVLW-IC-PILYKLYKI-KQQQIDNK-NQR-IEVLS-RLSIDSAIEEDE-A.DTYY-G................SGFANPVYREGDE....-  84
CPZ.CM.01.SIVcpzCAM13           MI.LI......-LGC-AI-L-LLN-FI-RNL....WRLCKQNKLEVE-EN-AL--T----------D-G-EEK-QRLL..............HDHDNMHGFANPLFDI-  84
CPZ.CM.05.LB715                 MTGLEIIG.....LIGI-IELS--.-GA-IVAYNN-....QEYK-GER--C--K-IR----------D-.DKEE-CTLL..............LPSD.PEFVLINN....-  81
CPZ.GA.88.SIVcpzGAB2            ML..S..MWVAIGLIGIGTLLV-N.----G--G-SVYKRWKRHKEEQR-ID--IKIT----------D-EDKET-ATLL.............HNNGFDNPMFEDRI....-  89
CPZ.US.85.US_Marilyn            ML.....NWFEIGLI--GIEG-LV.VII-GL-ARLWRQIKI-ENT-QE-QN-LE-IRI-E-------D-E-EET-AKLL.............SSLELDNPRIV.......-  85
GOR.CM.04.SIVgorCP684con        MH-RD......ILVIIIGIILLAVTVIS-LKALAL-LR....DR-E-RFFDRLE---SNK--E-Y--NEE-AAE-M..................-R-NELGFAFNLH..-  80
GOR.CM.07.SIVgor2139_287        MH-RD......IIVIIIGITLLAVTVII-LKIFAL-LR....DR-E-RFFDRLE--LSNK--E-Y--NEE-AAE-M...................---NELGF-FNLH..-  80
GOR.CM.07.SIVgorCP2135con       MH-RD......IIVIIIGITLLA-TVI--LKALAL-LR....DR-E-RFFNRLE--LSTK--E-Y--NEE-AAE-M..................-R-NELGF-FNLH..-  80
GOR.CM.12.SIVgor_BPID1          MH-RD......QGAL-IAGVL-VCC----GRL-LYLLE....ER-RDRFVQRLE-WRGKQ--E-YD-NEE-EEQ-R..................KL-NLLGF-NIL....-  79
GOR.CM.12.SIVgor_BQID2          MHYRG......QIVIIILC-LLLIXVLI-MLILKL-LDNK-QDR-E-E-ISRLR-IRXIRD--DY--NEE-EREXM................DLX-SYGFYNPMFEL-  86
GOR.CM.13.SIVgor_BPID15         MHYRD......EIVLLIAGVL-GGCLII-GYLLLYWLK....ERKRDIFVQRVA--RK-Q--E-Y--NEE-EEK-R..................-L-AQLGFAYYML..-  80
DEN.CD.x.CD1_CMO580407          MLK-K......LGPIEY-CLVFAV.VIT-IAA-GVGYLAY-AYKSY-EEL-Y-R.-RLWSI---Y--SQEDP.......................................-  65
GSN.CM.99.CN166                 MH-AA.....VWWWG-AIITF-YLCVALLA..LYLAWDKWV-GKPKPTQVAV-R.---D.-ED.SGIYDDAS-E-.....................TGFNGFANPGFEV-  79
GSN.CM.99.CN71                  MSAAA.....LWWWG-A-ITF-YFCLAIFA..LYLAWDKWI-GKPK.IPVAV-R.-V-D.DEE.SGIFEDAS-EP.....................NAY.GFANPGFEV-  77
MON.CM.99.L1_99CML1             MNYWWS.....LVAITYSLIL-ALPVAA-AW.WRY-KITK-FKRIDQE-Q---QIHER-RH---VDT-SE..-EQ.............HEETHGFVNPVFND-FGEWV.-  88
MON.NG.x.NG1                    MNTYW......I-AISVWCV-VLLAPCTLYLLYQS-KEHK-VSRFEQDFQ--VQXYQ-X..---Y-D-DEDHNSFDNPLFDDGDPDQWV*FRKQI---SIL-SAGL....-  98
MUS-1.CM.01.CM1239              MN........YWYLA-AL-TG-YFL-ALFA..FVLAYQRWCRP...KVEVSV-R.-L-EGDSD.SGIFEDAEDEPN...................GDG-HAFANPAFEQ-  76
MUS-1.CM.01.SIVmus_01CM1085     MN........YWYLA-VI-TG-YFV-AIFA..FVLAYQRWC-P..KKVEVSV-R.-L-EGDGD.SGIFEDAEDDMA...................ESE-HAFANPAFEQ-  77
MUS-2.CM.01.CM1246              MH........YWYLG-AI-TA-YLV-ALVA..F-LAYQRWCQP.KPPIEVNVLR.-L-EGDSD.SGIFEDACDGED..................EDS-RAFANPSFEP-  78
MUS-2.CM.01.CM2500              MN........WWWFA-A--TA-YFV-ALVA..FVLAYQRWCQPQKGQVEVNV-R.-L-EGDTD.SGIFEDAEDGSD..................DQR-.GFLNPAFEL-  78
MUS-3.GA.09.09GabOI81           MN........VWYLA-AI-TA-YXA--IVA..LWLLWDKYVRK..EPTVVSV-R.-L-EGDSGYE-VFEDAPQEP..GD.............PLIN...AF-NPAFEQ-  78
MUS-3.GA.11.11GabPts02          MN........VWYLA-A--TG-YFIVAIVA..LWLIWDKYV-K..-PTMVSV-R.-L-EGDSGYE-VFEDAMDQDNINQ.............PLVDNPHGF-NPAFEQ-  83
```

```
                                                                                                                                                   glycosylation NDT
                                                                                                                                                           V1
                          Env start             signal peptide end  gp120 start                                     glycosylation NVT                      CD4 binding  glycosylation NSS
H1B.FR.83.HXB2            MRVK...EKYQHLWRWGWRWGTMLLG.MLMICS..ATEKL.....WVTVYYGVPVWKEAT...TTLFCASDAKAYDTEVHNVWATHACVP.TDPNPQE...VVLVNVTENFNMWK..NDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGR....  146
H1A1.UG.85.U455_U455A     ---M...GI-RNYPCL------I--.LII--..NAQQ-.....------------D-V...-------------A------------------.--------...ID------E-----..-N--D---------------------------T-D-HNITINNTN-NTN......  142
H1C.ET.86.ETH2220         -K-M...GI-RNCQQW-I--ILGFW.-----N..GMGN-.....------------D-S...P------------------G-F----.---S---...LG-E-----------..---------Q--------G-----------T-N-NAI--N-KVTNN-......  143
H1D.CD.84.84ZR085         -----...GIKRNYQPL-K--I----MLM-TY-..-ADN-.....-----------------...----------S-K--A--I---------.--------...IE-K-----------..-N--D----------------------R--T-N---ASRNSTD-N-T......  144
H1F1.BE.93.VI850          ---R...GM-RN-QHLGK--LLF--.I-I--N..-ADN-.....-----------------...------------ER-A--------------.--------...-F-K------D---..-N-----T--------------------T-N--NAT-NSQEKPGA......  143
H1G.SE.93.SE6165_G6165    ---T...GI-RN-*HL-K---LI--.LVI---..-SNN-.....-------------ED-D...----------S-SA-S--------------.--------...IIME---Y------..-N--------------E-----------T-N---VT-KG-KRNN-TDNSTE  148
H1H.CF.90.056             -E.T.....-RNYPSL-----LI--.--L---..-AQN-.....--------------K...------------E--K--------------.--------...M-ME---S----E..-N-----T--------------------T-N--NVR-N-SNST--......  140
H1J.SE.93.SE9280_7887     -E.T.....-KN-QTL--G-L-IF-.-----K..-K-D-.....-------------D-K...------------S--K--------------.---S---...MN-P-------------.----D--Q----V--E-------I-----T-N-SNITSNS--T-N-S.....  141
H1K.CM.96.96CM_MP535      ---R...GM-RN-QTLGN--ILF--.I-I---..NAD--.....-----------------...P-----------EK---------------.--------...-EME-----------..-N-----T------E------E------T-N---Y-GTNS--NATST....  145
H1O.BE.87.ANT70           -KAME.KRN.KK-..-.TL.Y..-AMALITP-L..SLRQ-.....YA---A-----ED--...PV-------NLTS--K--I--SQ-----.---T-Y-...YP-H---DD--I---..-Y-----Q----------------QM-F---QME--NIAGT............  133
H1O.CM.94.BCF06           -T-ME.KKS.KKS..-.IL.C..IAMALIIP-L..SGRQ-.....YI---S-----ED--...PV-------NLTS--K-----SQ-----.---T-H-...YP-----DK-DI---..-Y--D-------------------QM-F---QMN--E-S-SSAS-TT........  139
H1O.FR.92.VAU             -KAMG.KRN.RK-..G.IW.C..-ILALIIP-L..SCNQ-.....YA---S-----ED-K...P--------NLTS--Q--I---Q-----.---S-N-...YE-K---DK--I-E..-Y--D--Q---D-----------QM-F---QMN---I--SIT-TN........  138
H1N.CM.02.DJO0131         -GMQ....IGWPFF.C........-M.ISLTIG...SK-Y.....-A---------RDVE...-V---------HS--A--I---Q-----.---------...-P-D---P----E..-K-A--VQ-------E---------------TMN-SNSNGNRT-DEKE......  133
H1N.CM.04.04CM_1131_03    -GMQ..SGWVTRRVLGMMSGWPFFCLLISS-IG...S-QH.....-----------R-TD...-V---------HS--A--I---Q-----.---------...-H-I----K-D----..-N-A---Q-------E---------------TMH-N-SNGFNE-VG.......  143
H1N.CM.95.YBF30           -K-M.GMQSGWMGMKS--LLFYL-VS.LIKVIG...S-QH.....-----------R--E...------------HS--A--I---Q-----.---------...-L-P---K----E..-K-AD--Q-------E---------------TML-N-SYGEERN-TN.......  144
H1P.CM.06.U14788          -K..K.KKN..S-..GNLEICLA-IMYFNA-SYT..SGTR.....Y----------RD-K...I-----A--SLTSX-Q--I---Q-----.---R-I-...-RID---S--I-D..-Y--T--Q-------------------V---TMX-S-..CS-V.DCTNNSS...  141
H1P.FR.06.RBF168          -K..K.KKN..S-..GNLEICLA-VIYFNA-SCA..SGIH.....Y----------RN-E...V-----A--SLTSK-Q--I---Q-----.---T-I-....-KI----S--I---..-Y--T--Q-------------------I---TMN-SE..CHKQAKC........  136
CPZ.CD.06.BF1167          -KMA..............LLIGWILTIIF-CKLGIGTY.....YT---------------...P------S--VAS-QP--I----N----.L--Q-Y-...IPI.-I-TE---ET..-Y---E-KK-L---FQ-----------F--KMN-SEFTRNST-GNNATKNNTS  139
CPZ.CD.90.ANT             --.........KPI..HII--.LA-.LIQF-EKGTNED......Y---F------RN---...P-----TN-SMTS-----------TS---.I--D-I.....--RL-TSVW--AY-..-Y---S-T--MXQ-FQ--H---------M-IKMN--GYNGTPT-P-TTTST...  137
CPZ.CM.01.SIVcpzCAM13     -K-MEERKRS.FNGFYMTMTIC-A...L-IP-LT.SD-.....-----------R-VN...-V---------HS--A--I------------.---------......---H----D----N..-Q-----Q---S-------------------TM--SNVTRQNS-SN-N......  143
CPZ.CM.05.LB715           -GT...........KKSCVLYYC-VILS-FSLTIDSK-E-.....-----------RD-E...------------QE--A--I--SQ------.-----E-...IH-P----Y-D----..-N-----Q--------------------------T-N--NCIWGNG-FR........  137
CPZ.GA.88.SIVcpzGAB2      -K-M....GKTRRRW...QPYCIIMA.LIIP-LKTESDQW.....................-P----AN-T-LKE-P--I---Q-----.---S-E-...-I-I----E--V-D..-A-----Q---T-------R-----N----Q-T--SINATG-ETNAT-.....  143
CPZ.US.85.US_Marilyn      -K-M..EK-.KR-..-.LSYC..--SSLIIPG....LSS-.....-A---------RDVE...------------KQ-A--I---Q-----.---------...-H-P---K-D--E..-N-A---Q----------------I-------TMT-LNPDSN............  131
GOR.CM.04.SIVgorCP684con  -SL.......................-.L-LT-IE.-KTEH.....YA---------R--K...V-----A--SFVSK-Q--I---Q-----.---S-T-...FL-E----T-DI---..-Y-----Q-------------------VM--TMT-SNDTKGS-KSTATTTSANT  132
GOR.CM.07.SIVgor2139_287  -SL.......................-.LILT-IE..ATH.....YA---------R--N...V-----A--NYVSK-Q--I---Q-----.---R-T-...FP-E----T--I---..-Y--D--QD--V----------------VM--T-N--Q-NTTNT-QTEMTNT...  127
GOR.CM.07.SIVgorCP2135con -SL.......................-.LILT-IE..ATH.....YA---------Q--N...V-----A--NYVSQ-Q--I---Q-----.---S-I-...FP-EX---T-DI-N..-N--D--QD-------------------VM--T-N--KTTET-TQTNXTN.....  125
GOR.CM.12.SIVgor_BPID1    -KD.....MTVMKKKKSSLGNWGICLALIIYFNAISYVNCTX...Y----------QD-K...-------A--DLASK-Q--I---Q-----.L--T-I-....LKL-I--S--I-E..-Y-----Q---V------------------F---TMN-SEY-GY............  139
GOR.CM.12.SIVgor_BQID2    -T-....TMKVMKKKKEKS-ILXIVMAFIIP--SSRPL.......YA---------RD-Q...-------A---MASS-M------Q-----.---Q-I-...LK-T----T--I---..SG-----Q-------------------VM--TMN-XRXTPN............  137
GOR.CM.13.SIVgor_BPID15   -KD.....MRVMRKKKRSLGNWG-SLAL-I--FNTGSYVMGTR..Y----------R--K...-------A--NLASK-Q--I---Q-----.---T-IK....-KL-IS-PY-I-E..-YI-K--QKN----LN-----------VI--TIN-SKF-PA............  140

                                            signal peptide end  gp120 start
                          Env start                                                                                                                                          V1
MAC.US.x.239              -GCL.........GNQL.....-IAIL-LSVYG.IYCT-.....Y---F----A-RN--...IP---.........-KNRDT-G-TQ-L-.DNGDYS-.....-AL----S-DA-N..-TVT--AI--VWQ-FET-I------S---ITMR-NKSET-RWGLTK-ITTTAS  132
H2A.DE.x.BEN              -EPG.........-NQL....FVVILLTSA-L..VYCSQ.....Y---F--I-A--N-S...IP-----.........-KNRDT-G-IQ-L-.DNDDY--.....IIL----A-DA-N..-TVT--AV--VWH-FET-I-----------AMN-SRVQGN-T-PNPRTSSSTT  132
H2B.CI.x.EHO              -AHV.........NNYL.....-VTLL-ISIYG.YMG-N.....F---F--I-A--N-S...IP-----.........-RNRDT-G-VQ-L-.DNDDYT-.....IQL-I--A-DA-D..-TVTD-ATK-VW--FET-I-----------TM--NKTWSSASKETTTSSASLR  132
H2G.CI.92.Abt96           -AYL.........GNQL.....-IATL-LSTYG.FYCTK.....Y---F--I-A-RN-S...VP-----.........-KNRDT-G-TQ-L-.DNDDYS-.....LAL----A-DA-D..-TVT--AV--VWN-FET-T----------IAMR-NKTETNKWGLTGTTTVTPA  132
H2U.CI.07.07IC_TNP3       -AYL.........GNQL.....-IALLCLNACV.IYCTQ.....Y---F--I-A-RN--...VP-----.........-KNRDT-G-TQ-L-.DNDDYS-.....LAL-I--A-DA-D..-TVT--AI--VWN-FET-I----------IAMR-NKTET-KWGLTGTVATTTS  132
H2U.FR.96.12034           -ACL.........GNQL.....-IALLCLNACV.IYCTQ.....Y---F--I-A-RN--...IP-----.........-NNRDT-G-TQ-L-.DNGDYS-.....LAL----A-DA-N..-TVT--AV--VW--FET-I----------IAMR-SKTETNRWGLTGTAATTTS  132
ASC.UG.10.RT03            -TKYKLIVLGIAIIIGLITIDLQQGDENDG.............-----F------HN--...PP---T.......ADARHT-V-TN-L-.----A-I-...TPMNISG-W--VHT..-Y--DH-DK-MAA-FL-AT---T----M--KM--QNYTKEAT-TTTPTSSTTS  139
ASC.UG.10.RT08            -TKHKLIVLGIAIIIGLIIVDLQQGEENNG.............-----F------HN--...PP---T.......ADARHT-V-TN-L-.----A-I-...TPMNISG-W--VHT..-Y--DH-DQ-MAA-FL-AT---TR---M--KM--SNFTKP-S-PQPTTSSKPT  139
ASC.UG.10.RT11            -TKYTLIALGIAIIIGLVIIDLQQGEENNG.............-----F------QN-S...PP---T.......ADARHT-V-MN-L-.----A-V-...TPMNITG-W--VHT..-Y--DH-DQ-MAA-FL-AT---T----M--KM--SNYTREAT-AAP-TTAPTT  139
COL.CM.x.CGU1             -LRY....................-FINL-VLGIV.L.SNK..........-----Q---A-E--DVNDQQF--F-SSPE.......IQQVLG-L-PPPGK-V-...QNMP---A-DLF-..-SFSGEVWIITQTTLE-R-R--A---AY-APMI--KVNRTE-GT-TVAPTT..  130
COL.UG.10.BWC01           -FD-..........KYIR....G-MIGIIC-EIV.V.GSK.....Y----Q---A-E--AAVDQQI--F-SSVD.......IQQVLG-L-PPPGK-V-...QNMP-I--S-DLF-..-SFSAEVWSITQTTLE-R-R--A--SAY-APML--EAQTNET-T-TPTTTTLT  137
COL.UG.10.BWC07           -KLS....................VILVI-G-SIV.I.S-K..........-----Q---A-ED-EITDQQF--F-SSIE.......MQQVLG-L-PPPGD-VV...QSME-IS-T--AF-..-PFSMEVWTVTQ-VLE-RIR--A--SAY-SWMN--KVQDN-TQATTTAVTSAT  132
DEB.CM.04.SIVdeb04CMPF3061 -LAF...................I-AIIVI-LGQ.K--QVSQQQQY---F-------N-S...VG---TANGSE......SG--VT--L-HSVIRE....Q-HMP--SQY-DAF-NP..IED-LW--MTG-YK-TYR-------Y--KMR-KTETTL-SNLTTP......  129
DEB.CM.99.CM40            --LI..........QLLL...LI-GLIVIGTQGK.QKKQQ.....YL---------VD-K...VD---TANSSE......SG--VT--L-HALVREE....-PMP---Q---AFDNP..IE--LWQ-MT--YK--F--------Y---MQ-IKTSTNPTPTNT-TTTTTI  136
DEB.CM.99.CM5             -LSL...................LIIIASLTLVGAA.P-KQQ..........-----------HN--...VG---TANSSE......SG--VT--L-HAMVRE....Q-PMP--SQY-DSFSNV..-EQ-LWD-MT--YR--F--------Y--RMN-SSTTTT-TAATTTI.....  127
DEN.CD.x.CD1_CMO580407    -SQV......RRTL....KVLLLI-W..TIGIIC.T---Q.....Y---F--I-A-RNSA...AP---T.......AMDRG---AAS---.E--YQR-....-WME-A-DY-DIYHNPIVDQVVQNMKD..-YK-TF--------V--RMN--RKEDNSTNAVTVSSTTTT  136
DRL.DE.11.D3              -IRH..........I......VIG-RILGFLGIS.IG.-H..........-----F--T-K-RP-S...-H-I--.........-DN-SF-V-TS-I-.SLLHYE-...TKIE-IE---TVPMTE-EVIK-AWGALS-MI-AV------IN-Y--KMI-N-T-EIPT-TT........  125
DRL.DE.11.D4              -IRH..........I......VIG-CILGFLGLS.IG.-H..........-----F--T-K-RP-S...-H-I--.........-DN-SF-V-TS-I-.SLLHYE-...TQIE-IE---TVPMTK-EVIK-AWGALS-MI-AV------IN-Y--KME---GPPQPT-TT........  125
DRL.DE.11.D5              -IRH..........I......VIG-CILGFLGIG.LG.QH..........-----F--T-K-RP-S...-H-I--.........-DN-SF-V-TS-I-.SLLHYE-...TQIK-IE---TVPMTE-EVIK-AWGALS-MI-AV------IN-Y--KME-D-IEVINGSST........  125
DRL.DE.11.D6              -IRH..........I......VIG-CILGFLGIS.IG.-H..........-----F--T-K-RP-S...-H-I--.........-DN-SF-V-TS-I-.SLLHYE-...TTIE-IE---TVPMTE-EVIK-AWGALS-MI-AV------IN-Y--KMI-N-T-ETPPPPT........  125
DRL.x.x.FAO               -IRH..........I......IIG-CILGFLGIS.IG.-H..........-----F--T-K-RP-S...-H-I--.........-DN-SF-V-TS-I-.SLLHYE-...TEIN-IE---TVPMTE-EVIK-AWGALS-MI-AV------IN-Y--KME-S---IVN.AA-TTPTPTT.  131
GRV.ET.x.GRI_677_gri_1    -GRL..........LIKI....LIIAIGI.SI.G..IGN-.....Y---F--I----NS-...VQA--M.......-PNT-M---TN-I-.D-HDNT-...-P-.-I--A-EA-D..-PL-K-AESN-HL-FE-TMR-----S-I-IKMS-VE-NGTAT-KATTTATTTM  130
GSN.CM.99.CN166           --KT.........VESMMMLAVS--AL-ALLIPG.LKCEN.....-T---------RD--...PP------PDIASN-PG-I-IST--L-.S--S-A-....-PL-I--K--IY-..-Y--DEVRD-MV--FN-A----------M--GMY-NLTNTS-SEPTTTP-PPNV  145
GSN.CM.99.CN71            -MR-.........VGFLKMLVVS-TLMA-L-PGS.RSKGN.....-T--F------RD-K...PP-------DITSN-PG-I-IST--L-.S--S-S-....-PL-I--E--IY-..-Y--DEVRD-MV--FN-A-R--------M--RM--KLPNTS-T-TP-TTTATTS  145
LST.CD.88.SIVlhoest447    -NCP..........GLIL......-LII-RVQGE.K-K-Q.....Y---F----K-E--V...VP-I--.........SANNSL-V-TS-I-.DLQSYAQ...IPIH-ISL--TKEIKD-QIIQ-AWSAMT-MV-TIM-----VN-Y--RM--EGEGIG-TKATKTKT.STT  133
LST.CD.88.SIVlhoest485    -DCL..........KLLL......FLLF-GAQGE.-.K-Q.....Y---F----K-ED-V...VP-I--.........SANNSL-V-TS-L-.DLQSYAQ...IPIH-ISL--TQEIKD-QIIQ-AWSAMT-MV-TIM-----IN-Y--RM---GGGIGPK-TP-PIGKTTT  133
LST.CD.88.SIVlhoest524    -ACP..........GNIL.....--LLC-GLTAG.EK--Q.....Y---F----K-ED-V...VP-I--.........SANNSL-V-TS-L-.DLQSYAQ...IPIH-I-L--TQEIKD-QIIQ-AWGAMT-MV-AIM---T-IN-Y--RM---GEGLGKKRTTTTTTPTTT  135
LST.KE.x.lho7             -ACP..........GL-.........ILL-LLGII..WG-Q.....Y---F----N-DDNVS..VP-I--.........SANTSL-V-TS-L-.DLQSYA-....-PIY-IS---TIPVKD-QVIQ-AWSAMNAMV-SIM-----IN-Y--RMQ-GEV......-KTPTTTPKTT  127
MAC.US.x.251_BK28         -GCL..........GNQL.....-IAIL-LSVYG.IYCTQ.....Y---F----A-RN--...IP---.........-KNRDT-G-TQ-L-.DNGDYS-....LAL----S-DA-E..-TVT--AI--VWQ-FET-I------S---ITMR-NKSET-RWGLTK-STTITT  132
MAC.US.x.EMBL_3           -GCL..........GNQL.....-IAIL-LSVYG.IYCTQ.....Y---F----A-RN--...IP---.........-KNRDT-G-TQ-L-.DNGDYS-....LAL----S-DA-E..-TVT--AI--VWQ-FET-I------S---ITMR-NKSET-RWGLTK-STTITT  132
MND-1.GA.x.MNDGB1         -KCP..........GLFL.....I-CLA#LV..G.IIG-Q.....Y---F---------K...-H-I--.........-DNSSL-V-TN-I-..SLPDYD..E-EIPDIK---TGLIRE-QI-Y-AWHAMG-ML-TI------IN-Y--KMQ-QETE-VSA-TAKPITTPTT  132
MND-2.CM.98.CM16          -LRY....................IILGI-VGL.GLGNQ..........--------T-K-HK-E...-H------.........--NDSL-V-TS---.SLLHYE-...QPIP-I-W--TGPMEE-EV-M-AWGA-S-MI-AV---------Y--KMQ--KG-DLATPIPTTSTTTTT  129
MND-2.GA.x.M14            -LRH....................-ITAIV-IGI.GIG-Q..........--------T-K-HP-R...-H------.........-DNNSL-V-TS---.SLLHYE-...QLIP-I----TVPIQE-EV-R-AWGA-S-MI-AV---------Y--KMQ--KGE-SSTSTTTPMPTTTT  130
MND-2.x.x.5440            -LRY....................LRYIVLGI-V-V.IVGEQ..........--------T-K-HK-R...-H------.........-DNNSF-V-TS---.SLLHYE-...QHIP-I----TGPIEE-EI-T-AWGA-S-MI-AV-----R---Y--KM---EGQ-E-EQATAKTTTPVP  132
MNE.US.x.MNE027           -GCL..........GNQL.....-IAILFLSAYG.IYCIQ.....Y---F----A-RN--...IP---V.......-RNRDT-G-TQ-L-.DNDDYS-....LAL-I--S-DA-E..-TVT--AI--VWH-FET-I----------ITM--NKSET-KWGLTK-STTTAA  132
MON.CM.99.L1_99CML1       --KL......MALSTLCLMM-LVN-.............SN.....-T---------P--...PP------PNYGSK-AG-N-LASS-L-.---T--S....LYL-I--E--AYQ..-Y---E-V--MK--FS-A----------M--RML-VEVNTVS-ASTTPAPSTPT  136
MON.NG.x.NG1              -GTL......TSGSXSGN....................-S..........-----F-----Y-P--...PP------PDYGSK-AG-N-LASS-L-.S--V-RA....LPL-I--S--AYN..-Y---E-V---K--FS-A----------M--RMR-QXINES-TVTPTTTTXATT  129
MUS-1.CM.01.CM1239        -PRM..........S.LMGM--T--LTL-LSS-I.IWGNQ..........-----F-------D--...PP-------NIANR-PG---I-T--L-.---A-------...-QL-ISGEYFNVYK.-Y--T--VD-VR-IFN-A----------M--KM--SRINVTRPS.........  134
MUS-1.CM.01.SIVmus_01CM1085 -PRM.........T-LRA.S-T--LTL-LSS-I.IWGRQ..........-----F-------D--...PP-------NIANR-PG---I-T--L-.---A-------...-QL-ISGEKFDIYK.-Y--D--TD-VR-IFN-A----------M--KM--AKVT-Y-RS.........  134
MUS-2.CM.01.CM1246        -PAM........ERTRILTALSLT-ALS-SLILR.KHA-Q..........-----F-R---S-N--...PP-----N-NIANR-PG--LI-T--L-.---A-------...-QL-ISNEHFNIYK.-Y--D--TD-VK-IFT-A----------M--KM--DL-NISMSS.........  137
MUS-2.CM.01.CM2500        -A-M...............T.SA-VFLTL-LS--L.IWG-Q..........-----F-------D--...PP-------NIANR-PG---I-T--L-.---A-------...-QL-I-GEYFDIYK.-Y--D--TD-VR-IFN-A----------M--RM---K-NSTRP-.........  130
MUS-3.GA.09.09GabOI81     -LLT..........ILP.........LSSEFPLPKSK-STE.....-T--F-------D-K...PP------VNIAS-RPG-I-IAT--L-.---S-------...-VL-IS-Y-.DIYK.-Y--DEVTD--R-IFN-AY--------M--RMA-VLQ-ENKTA.........  127
MUS-3.GA.11.11GabPts02    -PWT.......KITLTNLWLT-LMALIT-PLSNEASEQ-KE..........-----F-------D-K...PP------VNIAS-RPG---IST--L-.---S-------...-PL-I--E-.NIYK.-Y--DEVTD--K-IFQ-AY--------M--RM--ALI--NVTG.........  139
OLC.CI.97.97CI12          --K-....................LV-CLLWLSILGISIG.Q.....Y---F--T-K-RDSN...D--M-V.......-ANKSL-V-T--L-.SFDEDEI..EQPIPG-E-V-YNVS.D-EV-R-AWFG-A--L-AV----------F--KMQ-E-PSANFTEPITPISTAAP  132
RCM.CM.00.SIVagi_00CM312  -KSA................I-IVIIA..SL-SAI..VS-Q..........-----F-------NS-...APM--V.......-DNA-S-G-LN-I-.EGGLSP-....-P---S-Q-DA-E..-SLY--AKDNVWH-Y-ST-----R-S---ITMN--A-NGSWDGSVTTTTQPRT  129
RCM.CM.02.SIVrcm_02CM8081 -KCL..............SLIVLSS-I..-LALAK..QKPQ.....Y---F------VNS-...VPM--V.......-DNT-S-G-LN-I-.EGGISP-....-P---S-K-DA-E..-ALY--AKDNVWN-Y-ST-----R-S----TMN--VINGSWDGITTPAPSPAP  131
RCM.GA.x.SIVRCMGAB1       -DK-.............LVIVLIVVI-..IILVQG.SQKPQ.....YI--F------RNS-...VPM--V.......-DNTQS-G-LN-I-.EGGISP-....-SI--S-R-DA-N..-SLY--AKDNVWN-Y-ST-----R-S---ITMN-SAINGSWDGIPT-APPTTT  133
RCM.NG.x.NG411            -AKL...............ISIGIII-..-LGIII.VG--Q.....Y---F-------NS-...VPM---.......-DNAQS-G-LN-I-.EGGISP-....-P---S-K-DA-E..-SLY--AKSNVWH-Y-ST-----R--M-ITMY--A-NGSWDGTATQSPTTTT  131
SAB.SN.x.SAB1             -KLL..........TVLL...WLSGCWS-VWLVQ...........Y---F--I----NSS..VQA--K.......-PNT-L--STN-I-.D-EPEGTIAE-PIP-I--K-DA--NR-PL-G-AESN-HL-FEST-------S-M-IKMN-YR-EGGAA-TT-PSTSTAR  135
SMM.CI.79.SIVsmCI2        -ECL..........GNQL.....-IAVL-RSVCG.IWCTQ.....Y---F--I-A-RN-S...VP---.........-KNRDT-G-TQ-L-.DNGDY--....IRM----S-DA-D..-TVT--AI--VWN-FET-I----------IAMR-NKTET-KWGLTGNTTAAPT  132
SMM.LR.89.SIVsmLIB1       -ACL..........GNQL.....-IAIL-VSAYG.IYCTQ.....Y---F----A-RN--...IP---.........-KNRDT-G-TQ-L-.DNDDYS-....ISI----A-DA-D..-TVT--AI--VWN-FET-I----------IAM--NKNETNRWGLTRTSTTTPS  132
SMM.SL.92.SIVsmSL92A      -GCL..........GNQL.....-IAIL-LSA-E.IFCTQ.....Y---F----A-RN--...IP---.........-QNRDT-G-TQ-L-.DNDDYS-....LAI----S-DA-D..-TVT--AI--VWN-FET-I----------IAMQ-DRNETRRWGLTGGTATTTA  132
SMM.SL.92.SL92B           -ACP..........GLHL.....-IDILFLSVLG.TWCAQ.....Y--IF--I-A-RN--...IP---.........-QNRDT-G-VQ-L-.DNGDYS-....LAL----A-DA-D..-TVT--AI--VWN-FET-I----------IAM--NKNET-RWGLTRAATTTSS  132
STM.US.89.STM_37_16       -ACP..........GNQL.....-IAIL-LSACL.TYCTQ.....Y---F----A-RN--...IP---.........-KNRDT-G-TQ-L-.DNGDYS-....LAI----A-DA-D..-TVT--AI--VWN-FET-I----------IAMR-NKNET-KWGLTGKTVTTVT  132
SUN.GA.98.L14             --CP....................EILIGFSLLLG.SIAIQ.....Y---F--T-K-EP-V...VP-I--.........SANNSL-V-TS-L-.DLQTYA-....-PITGLE---TEGISN-QI-Q-AWQAMT-MV-AIM-----IN-Y--KM--PPKPTTPSS--TVKSSCDY  131
SYK.KE.x.KE51             -KLF..........NIIS....VV--IV-SLGIV.-RLEQ.....Y---F----N-ED-Q...VPM---.........-PHTGG---KN---.SSEQIEV....RVNISG-Y-TA-NSSHGIRQ-LLQ-MSD-FL-ADR-----S-M-IRTL-VED-DSSSG-LTTTI....  131
SYK.KE.x.SYK173_COMGNM    -AAF.........-TYI.....VCLFS-ISLGF.MEKQQ.....Y---F--I-H-ED-Y...AP---T.......-SHKGG---KN---.SADQIEV....RVNITG-Y-PA-NSSHMIRQ-IL--MSA-FL-ANR-----A-M-IRML--LDNSPATS.........  124
TAL.CM.00.266             -KYL..........AVI.........FCLAYLGRF.IIGNQ.....Y--I---------N--...PP---.........-MARDQ-S-SQ---.S--QNL-....-P--L--Y-DVY-..-Y--DHIA--MD--FLTATR--T----I--RM--E-ITKNLTISANTTTPAPT  129
TAL.CM.01.8023            -KCL..........QVLCLAICVIWVANIGKSLT.SA-QQ.....FL-----------N--...PP---.........-MARDQ-S-SQ---.S--QNL-....-P--L--Y-DVY-..-Y--DHIAD-MDA-FLTAT---T----M--RMA---ITKNLTGTVAPPTTNST  137
TAN.UG.x.SIVagmTAN1       -GPL..........-GKG...VL-VILG-SLIGL.LYGTQ.....YI--F--I----NSS..VQA--M.......-PNT-L---TN-I-.D-HDYT-....-QL--S-K-EA--DR-PL-A-AESN-HL-FEST----------M-IKMN--K-TSTAP-S-TPTSSSTT  136
VER.DE.x.AGM3             -KLT..........LLIG....I-LIGI.GVVL.N-RQQ..........-----F-------NSS..VQA--M.......-PTTRL---TNSI-.D-HDYT-....-PL-I--P-EA-ADR-PL-A-AGSN-HL-FE-T-------S---IKMS-VE-NSSEP-TTPKSTTAST  133
VER.KE.x.9063             -KPL..........LGIV....LIAIIGI.SSL..LS-RQ..........-----F--I----NSS..VQA--M.......-PTTRL---TN-I-.D-HDYT-....-PL-I--P-EA-ADR-PL-A-AGSN-HL-FE-T-------S---IKMS-VE-NSSRRNPTPTSASTTK  133
VER.KE.x.AGM155           -TKF..........LGIF....IV--IGI.GIGI.S-KQQ.....-I--F-------NSS..VQA--M.......-PTTRL---TN-I-.D-HDYT-....-PL-I--P-EA-ADR-PL-A-AGSN-HL-FE-T-------S---IKMN-VE--GSATSTPATSTTAGT  134
VER.KE.x.TYO1_patent      --YT..........IITL....GIIVIGI.GI...VLS-Q.....-I--F--I----NSS..VQA--M.......-PTTSL---TN-I-.D-HDYT-....-PL-I--P-EA-GDR-PLIA-AASN-HL-FE-TM------S---IKMN-VE-NSTRERATTPTTTPKS  132
WRC.CI.97.97CI14          -IF-.....................-VLG-IGIAI.SLAVQ.....Y--------Q-EDS-...DF-L--.........--NRSL-VSTR-I-.HLEHLE-...QYMPEIK---TRKVAE-QI-Q-AWTGVT-MI-TL------I--Y--KMQ--PGEGSETF....DPCESN  126
WRC.CI.98.98CI04          -LRF.....................ILISIV-IGL.GIG-Q.....YI-------Q-E-SK...DF-L--.........--NRSL-VSTH-I-.HIEHWE-...QY-P--K---TRKVAD-QIIQ-AWTGVT-MI-TI------I--Y--KMT--PGAP-S-NTKDFNPCEIS  130
WRC.GM.05.Pbt_05GM_X02    -NTF....................IFIAVVIGIGL.GLG-Q.....YI-------Q-E--K...DFML--.........--NRSL-V-TH-L-.HIEEIT-...QYMEGIK---TKKVEE-EV-Q-AWTGVT-MI-TL------IN-Y--KMS--LGDEKS-YTEDIELTTVK  131
```


glycosylation NIS
glycosylation NCS
CD4 binding
V1
V2
glycosylation NDT
glycosylation NTS
glycosylation NGT
glycosylation NKT
H1B.FR.83.HXB2
H1A1.UG.85.U455_U455A
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
H1O.BE.87.ANT70
H1O.CM.94.BCF06
H1O.FR.92.VAU
H1N.CM.02.DJO0131
H1N.CM.04.04CM_1131_03
H1N.CM.95.YBF30
H1P.CM.06.U14788
H1P.FR.06.RBF168
CPZ.CD.06.BF1167
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715
CPZ.GA.88.SIVcpzGAB2
CPZ.US.85.US_Marilyn
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139_287
GOR.CM.07.SIVgorCP2135con
GOR.CM.12.SIVgor_BPID1
GOR.CM.12.SIVgor_BQID2
GOR.CM.13.SIVgor_BPID15
V1
V2
MAC.US.x.239
H2A.DE.x.BEN
H2B.CI.x.EHO
H2G.CI.92.Abt96
H2U.CI.07.07IC_TNP3
H2U.FR.96.12034
ASC.UG.10.RT03
ASC.UG.10.RT08
ASC.UG.10.RT11
COL.CM.x.CGU1
COL.UG.10.BWC01
COL.UG.10.BWC07
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM5
DEN.CD.x.CD1_CMO580407
DRL.DE.11.D3
DRL.DE.11.D4
DRL.DE.11.D5
DRL.DE.11.D6
DRL.x.x.FAO
GRV.ET.x.GRI_677_gri_1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest447
LST.CD.88.SIVlhoest485
LST.CD.88.SIVlhoest524
LST.KE.x.lho7
MAC.US.x.251_BK28
MAC.US.x.EMBL_3
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
MON.NG.x.NG1
MUS-1.CM.01.CM1239
MUS-1.CM.01.SIVmus_01CM1085
MUS-2.CM.01.CM1246
MUS-2.CM.01.CM2500
MUS-3.GA.09.09GabOI81
MUS-3.GA.11.11GabPts02
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.CI.79.SIVsmCI2
SMM.LR.89.SIVsmLIB1
SMM.SL.92.SIVsmSL92A
SMM.SL.92.SL92B
STM.US.89.STM_37_16
SUN.GA.98.L14
SYK.KE.x.KE51
SYK.KE.x.SYK173_COMGNM
TAL.CM.00.262
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 NVS
CD4 binding
glycosylation NNT
glycosylation NCT
V3 tip
V3
glycosylation NFT
glycosylation NGS
glycosylation NTS
glycosylation NNT
glycosylation NKT
H1B.FR.83.HXB2
H1A1.UG.85.U455_U455A
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
H1O.BE.87.ANT70
H1O.CM.94.BCF06
H1O.FR.92.VAU
H1N.CM.02.DJO0131
H1N.CM.04.04CM_1131_03
H1N.CM.95.YBF30
H1P.CM.06.U14788
H1P.FR.06.RBF168
CPZ.CD.06.BF1167
CPZ.CD.90.ANT
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.LB715
CPZ.GA.88.SIVcpzGAB2
CPZ.US.85.US_Marilyn
GOR.CM.04.SIVgorCP684con
GOR.CM.07.SIVgor2139_287
GOR.CM.07.SIVgorCP2135con
GOR.CM.12.SIVgor_BPID1
GOR.CM.12.SIVgor_BQID2
GOR.CM.13.SIVgor_BPID15
MAC.US.x.239
H2A.DE.x.BEN
H2B.CI.x.EHO
H2G.CI.92.Abt96
H2U.CI.07.IC_TNP3
H2U.FR.96.12034
ASC.UG.10.RT03
ASC.UG.10.RT08
ASC.UG.10.RT11
COL.CM.x.CGU1
COL.UG.10.BWC01
COL.UG.10.BWC07
DEB.CM.04.SIVdeb04CMPF3061
DEB.CM.99.CM40
DEB.CM.99.CM5
DEN.CD.x.CD1_CMO580407
DRL.DE.11.D3
DRL.DE.11.D4
DRL.DE.11.D5
DRL.DE.11.D6
DRL.x.x.FAO
GRV.ET.x.GRI_677_gri_1
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.SIVlhoest447
LST.CD.88.SIVlhoest485
LST.CD.88.SIVlhoest524
LST.KE.x.lho7
MAC.US.x.251_BK28
MAC.US.x.EMBL_3
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
MON.NG.x.NG1
MUS-1.CM.01.CM1239
MUS-1.CM.01.SIVmus_01CM1085
MUS-2.CM.01.CM1246
MUS-2.CM.01.CM2500
MUS-3.GA.09.09GabOI81
MUS-3.GA.11.11GabPts02
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.CI.79.SIVsmCI2
SMM.LR.89.SIVsmLIB1
SMM.SL.92.SIVsmSL92A
SMM.SL.92.SL92B
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

```
                                                 glycosylation NST
                                           glycosylation NST
                                                                V4                                                            V5
                                                                                             glycosylation NIT            glycosylation NES
                 CD4 binding        glycosylation NST     glycosylation NNT          CD4 binding             CD4 binding              CD4 binding
H1B.FR.83.HXB2               SGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTEGSNNT..........EGSDTITLPCRIKQIINMWQKVGKAMYAPPISG...QIRCSSNITGLLLTRDGGN.................SNNESEIFRPGGGDMRDNWRSELYKY.KVVKIEPLGVAPTKA  501
H1A1.UG.85.U455_U455A        ----I--T-------------TSG----I-NG-MSNDM-P-G-..................--Q-----------R--Q-------Q-...V---E------------T.................N-TKN-T-------------K------.------------R-  494
H1C.ET.86.ETH2220            ----L--T-------------TSN-----KLELFN-STNL-..................--Q-----------G--R-------E-...I-M-R-----------AK.................EPHSTK-----E----------------.---E-K--------P  490
H1D.CD.84.84ZR085            -------T-------------TSG----A-NI-GH--GLNDTI..................-I-----------E-------E-...--N-------------A.................N-TQNDT-------------------.---------------  501
H1F1.BE.93.VI850             ----L--TM-----R------DTSG---D-GS-NGT..........................----------V----G--R---TS--A-...N-T-N-----------------.................ESNI-T---E--N-K-----------.---E-----------  477
H1G.SE.93.SE6165_G6165       A---L--T----I-R-------TSG----SLLR-NS-EN-T.......................-----K----VR---R--Q-------A-...N-E-N-S----I-------N.................NNTNTS--------------------.-T---KS------R-  504
H1H.CF.90.056                ----M-VR------R-------TSG----S-EMH-NY-SNDTKG.............NEN---------V----R--R-------Q-...N-M-V------I--I-E--.................ASA-NYT-----------------.-----------I---T  490
H1J.SE.93.SE9280_7887        ----I-------------L---TSK----S-DKNSIEATNDTSX..............A---I--K----VR---RT-Q-I-----A-...N-T-T------------R.................GNG-E-GT-T---T--N-K------.---E-----------  493
H1K.CM.96.96CM_MP535         P------L--M---A-------T-K---E-GE-G-..........................------------I-----A-...S-N-------MI----------.................THN-T--------------------.--Q-----I---R-  487
H1O.BE.87.ANT70              ----L-VTHLH---H-------TAKM--Y.T-SCNGT-....CSVS..NVSQGNNGT........KLR-VVRS-IRGQSGL-----K-...NLT-M-----MI-QM-.................NTWN-S-NNVT---I----K-I--T--FN-.---RVK-FS----RI  495
H1O.CM.94.BCF06              ----I-ASSLH---H------DTSR---Y.T-KCNG-Q....C.NETNKTQTNKTQDT...II--K-R-VVRS-I-GELGL-----P-...DLT-K-----MI-QL-TP.................YN-SCDNVT---T------I--T---N-.--IQVK-FS-----I  512
H1O.FR.92.VAU                ----A-VTNFF---H-------TN----H.T-SCKKNMTYNKINCTNI..SNNSNGTQ...AI---LR-VVRD-MRG-SGL-----P-...NLV-R-----MI-QL-TP.................WNKTHPN-TTL-------K-I--TQ-F--.---RVK-FS-----I  508
H1N.CM.02.DJO0131            N-----VTHLM-------I---TSR---ESMN.......................TNGTNG---------R--V-L-TR---GI-----R-...NLT-N-T----I-EHS--.................SNGTVY-T--N-VNL--Q------.-T-S---I----G--  473
H1N.CM.04.04CM_1131_03       D-S---VTHLM--------L--TSR---ESKN.......................ETTNG--I---------V-L-TR---GI-----R-...N-T-N------I-EYS.................NSTNGTVVY-S--N-VNL--Q--------.---S---I----G--  488
H1N.CM.95.YBF30              NE--L-VTHLM---R-------TSK---EELL-E-GEP..........................--------R--V-L-TR---GI-----R-...VLN-T------V-EYS--P.................DTKET-VY-S--N-VNL--Q--------.---S---I----G--  488
H1P.CM.06.U14788             T-----VSSLH-S-H-------TSS---FSYTWNDTTWRAAG-.......HNT...SENI.--S--L--VV-S-MR--SGLF-T--E-...ELK-Q------M-E-ELPYN.................DSSKNTTLS-I-----TNI------P-.----QVKA-A-----I  514
H1P.FR.06.RBF168             T-----VSFLH-S-H-------TSS---YSYTCNEKGVCSI--.......HTGNYT-ENI------L--VV-S-MR--SGLF----E-...-LQ-H------I-D-ASPYN.................ANSSSNTTLS-T-----HI------P-.----QVKA-A----RV  508
CPZ.CD.06.BF1167             ----Q-TQ--W---Q------DLKNWT-NGIEID.......................GMLIA---LR--V-H-GI-SRGI-L--RE-...-VK-V-S---FIM-AETNDTDG.................T-TPTFSAKVE-Y-KV-MSR-.---E-Q--S-----G  497
CPZ.CD.90.ANT                QD----VKV-W---Q------DI-PW--A-YTGNLITNGAL................IAH-------V-H-GI-S-GI-LA-RR-...NVS-T-S---IM-.EGQIY.................-TVKVS-AARVA.-Q--A--SR-.Q--E-X--S----TG  491
CPZ.CM.01.SIVcpzCAM13        -------AN-M----------DTRKM--.........ES-PFHE..............NM-I-----R--V-S-MR--RGI-----P-...H-T-N-L----I----HV.................-TNNT---I----KNI--------.---R----S------  499
CPZ.CM.05.LB715              C--E......F-Y-ETSSLFNA.-IDITPNTTLN..........ETTWDYPAENNINQEMIIH-------VS---R---GIFS---R-...T-Q-N-T----------N-KADSENKTRIETGQNKAEEI-RTD-----A--E-----------R-.--------L-----  547
CPZ.GA.88.SIVcpzGAB2         P-----V-NFM-----------TSR-ITCNSSDTSE......................YI---K-R-VV-S-MR---GIF---RR-...T-T-N-T------EVQN.................GTGNNTEVYLS------I---------.-I------------  487
CPZ.US.85.US_Marilyn         A-----V-NMML----------TIP---M-YN....N-DNTTI..................--K---R--V-Q-MR---GIF----K-...VLS-N-----MI-DISISAV.................NNDSRNITVM-T----TAL-KN--H--.---S---I----G--  472
GOR.CM.04.SIVgorCP684con     ------VTSLH---H-------T-KM--Y.KCTGEHCN....C-DKELCNKTAEM-Y.....I---L--VV-S-MR--SGLF---XR-...TX--IXX---IX-Q--TPX.................AE--NNTTL--L--E--NI------P-.----VKA-S------  491
GOR.CM.07.SIVgor2139_287     ------VTFLH---H-------T-KM-SY.N..C-TEG....CNTTETSNTTETTKY.....I--HL--VVRS-MR--SGLF----R-...TL--I-----II-Q--AP.................LKTD-NSTL--I----TNI------P-.---RVKA-T------  487
GOR.CM.07.SIVgorCP2135con    ------VTFLH---H------DT-KM--YNCTKESCDCITG-C...............NYI--HL--VVMS-MR--SGLF----R-...TL--K-----II-Q--VPLNK.................TNENSTDYNTL--I----TNI------P-.----VKA-S------  482
GOR.CM.12.SIVgor_BPID1       ------VASLH---H-------TSA---H...SC-K.NMT.K-EWICT.....PNNATG-LR----L--VV-S-MR--SGLF----P-...SLT-K-----I--E--LP..LGNMTNT.................TL--I----KNI------P.Y---QVKA-------I  520
GOR.CM.12.SIVgor_BQID2       ----L-VAMLH---H-------T-TM--Y-Y.-CSH.H.......SCKAQSKNIS---GDSYI--KL--VV-S-MR--SGLF----R-...TLK-M--------E--VP..XNITAT.................XRNITL--T--E-K-I------P.Y---QVKA-S-----I  521
GOR.CM.13.SIVgor_BPID15      ------VTHLH---H-------TSK---Y-Y.VCNQTDQ-N.ANYSCH...STSTLENG-LVI--KLR-VV-S-MR--SGLF---VP-...SLT-H------I-Q--WP..LNNNTTT.................-L--E----K-I------P.Y---QVKA-A-----I  522

                                                                V4                                                        V5
MAC.US.x.239                 ..----VTFMWT--R---L--KMNWFL-WVEDRN-AN.............QKPK-QHKRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANI-WIDG.....................NQTNITMSAEVAELY-L--GD-.-L-E-T-I-L---DV  512
H2A.DE.x.BEN                 ..S---VAFMWT--R---L---M-WFL-WVEDKNQTRR.....................NY..-H------T-H----NV-L--RE-...ELA-E-TV-S-IIANI-IDKN.....................RTHTNITFSAEVAELY-L--GD-.-LIE-T-I-F---DQ  498
H2B.CI.x.EHO                 RSS---VRYMWT--R---L---M-FFL-WVENRTGL...................KRNYAS-H-R--V-T-H-I-RNV-L--RE-...ELS-N-TV-S-IANI-WIDK.....................NLTNITVSAEVSELYKL--GD-.-L-E-T-I-F---SI  497
H2G.CI.92.Abt96              ..S---VKFMWT--R---L--KM-WFL-WIEDRNMTLL.............KPQ-RQRRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANIEENRS.....................NNHTNIIFSAEVAELY-L--GD-.-L-E-T-I-F---NV  514
H2U.CI.07.07IC_TNP3          G-----ATFMWT--R---L--KMNWFL-WVEDRGEND...........TKDQNIR-RHKRNYV--H-R-V--T-NR---NV-L--RE-...DLT-N-TV-S-IANI-WTDT.....................NRTNITMSAEVAELY-L--GD-.-L-E-T-I-F---SV  518
H2U.FR.96.12034              ..--S-VKFMWT--R---L--KM-WFL-WVENR-MEGQ.................TSR-RHKRNYV--H-R-V--T-H----RNV-L--RE-...DLT-N-TV-S-SMIANI-RI-K.....................NQTNISMSAEVAELY-L--GD-.-L-E-R-I-F---PV  508
ASC.UG.10.RT03               ..--R-VES-W-Q-Q-------LSRI-QK-N-TNDTFYPRNI.........TKEGAELSN-WFT-T-R-VV-R-SH-E-L--L--RR-...HVQ-T---V--I-VESEYYPG.....................-PFNMTPSANIK-L-KID-RR-.---E-D-I-L---PI  514
ASC.UG.10.RT08               ..--R-VES-W-Q-Q-------LSRI-QK-N-NLTNNTFYPQNI........TQEG-ALSN-WFT-T-R-VV-R-SH-E-L--L--RR-...HVQ-T-----I-VESEYYPG.....................-PFNMTPSANIK-L-KID-RR-.---E-D-I-L---PL  517
ASC.UG.10.RT11               ..--R-VES-W-Q-Q-------LSRI-QQ-N-TNDTFYPRNI.........TEEG-RLSRTWFT-T-R-VV-R-SH-E-L--L--RR-...HVQ-T-----V-VESEYYPG.....................-PFNMTPSANIK-L-KID-RR-.---E-D-I-L---PI  507
COL.CM.x.CGU1                RTS---ATFTFVI-HRL-----ASS-WKHDS...........................PVMN-T-RKLV-S-VTHARIL-G--PG-...HLQ-NWEKQPVIAFMGTIEG...............DNDGNGCAYPAAPNFKHALSTL--GR-.-L--MRTTTYV--DI  495
COL.UG.10.BWC01              KTS---ATFTFLI-HRL----DASS-WAKNSSVMN...........................-T-RKLV-S-VSHARLL-G--PD-...HLQ-NWEKQPVIAFMGTVEK...............GG-ETGCAYPAAPNFKHALSTL--GR-.-L--MRTTVY---DV  502
COL.UG.10.BWC07              TYTFITFHGMFLY-NASSL.......WSG-T.........................TVMN-T-RKLV-S-VTHA-IL-G--LK-...TLQTYWEKQPV-AFMGSVEK...............GE-KTGCAYPAAVNWKHAASTL--GN-.-L---T-TAFV--DH  491
DEB.CM.04.SIVdeb04CMPF3061   P---R-VQRVW-Q-H--Y----VSKAYQLFTND.................TITW-NNTLM--K-N-LV-T-Y-I-QHI-L--RE-...EL--D-HVSAIIFDVAHYGH.....................MLNLTPSS-IKAV--AD-----.IIEVQ-I-F---AV  481
DEB.CM.99.CM40               PK--R-VQSVW-Q-H-------ISKALDLLLLQNNT...............RNSTW--KWLM----N-LV-T-YT--QHI-L--KE-...ELK---H-SAFVFDV-HY-G.....................S-TLTPSA-I-AV--AD-F--.-IIEVK-I-F---SAV  504
DEB.CM.99.CM5                P---R-VQNTW-Q-H-------VSKA-..NLLMNKT...............NNITW-NNTLM--K-N-LV-T-F---QHV-L--KE-...ELK-H-SVSA-VFEVAHYGQ.....................NMNITPSS-IKAV--AD-----.IIEVK-I-M---SI  484
DEN.CD.x.CD1_CMO580407       V-----TSSAR-Q-R-------L-RM---SQYPVNN................SNIINNQTYMV-N---LV-T-HT--MDI-L--AE-...E-T-T--VS-IIVDT-VHPN.....................RTH-TLTLSA-V-QV--A--ARW.-L-ELT-I-L---SV  501
DRL.DE.11.D3                 TK--LASENLMML----Y-F--ISKIWRTWNNR-SNVW....................YPYAS-H----VDD-AR--RKI-M--V--FNNE---TQDV-EMFFEIQKLETN.............EINDSNYQ-KFVPQDEVQNQFTAIGAH-.-L--V--I-F---DV  553
DRL.DE.11.D4                 AE--LASENLMMM----Y-F--V-KIWKTWNNR-SSVW....................YPYAS-H-----DD-A---RKI-M--V--FNNE---TQEV-EMFFEVHRLDTN.............VTDDSHYQ-KFIPQDEVQNQSTAVGAH-.-L--V--I-F---DV  537
DRL.DE.11.D5                 TK--LAAENLMML----Y-F--ISKIWRTWNNR-SNLW....................YPYAS-H----VDD-A---RKI-M--V--FNNE---TQDV-EMFFEIQKLRTN.............ETNDSNWQ-KFIPQDEVQNQFTAIGAH-.-L--V--I-F---DV  545
DRL.DE.11.D6                 TK--LASENLMML----Y-F--ISKIWKTWNDR-SNVW....................YPYAS-H-----DD-AR--RKI-M--V--FNNE---TQDV-EMFFEIQKLETE.............EDNDSHYQ-KFVPQDEVQNQFTAIGAH-.-L--V--I-F---DV  550
DRL.x.x.FAO                  TK--LASENLMML----Y-F--VSKIWRTWNNRTSNVW....................YPYAS-H-----DD-AR--RKI-M--V--FNNE---TQDV-EMFFEIQKLETE.............ETGDSNYQ-KFIPQDEVQNQFTAVGAH-.-L--V--I-F---DV  549
GRV.ET.x.GRI_677_gri_1       ..----SANFW---Q------KMDWFI-YLNNRTEDAE...GTNRTCDKGKPGPGPCVQRTYVA-H-R-VV-D-YT-S-KV----RE-...HLE-N-SV-A-YVAI-YN-K.....................SGP-NVTLSPQV-SI-AY--GD-.-L-E-T-I-F---DV  511
GSN.CM.99.CN166              PK--L-VQ--W-Q-Q-------LSI--QLNNN.TIN-S.NIG-..........ITSKYKGQW-A-K-R-FVTQ-GY-S-SI-L--RQ-...H-N-T-------IDGAMYES.....................S-NMTPSA-VA-A-NY--SR-.---E-D--SM---P-  512
GSN.CM.99.CN71               PK--L-VQ--W-Q-Q-------LSV--QFNNNITAVDDTNI--..........VTSKYKGQWMA---R-FVTQ-GY-SRSI-L--RQ-...H-N-T-------IDGAMYGH.....................S-NMTPSA-VA-A-KY--SR-.---E-D--SL---S-  519
LST.CD.88.SIVlhoest447       R--NLATEMLMHT--E-M-F--V-RI-QEWNNKNSDKW....................YPWAN-H--SV-DD-AS---KI-L--T--FNNR---THRV-EMWFQMEKFEP.............EENLGGNLSVRFLPPSWETNQFVATGAH-.-LIRLR-I-F---DV  539
LST.CD.88.SIVlhoest485       R-EGLATEMLMHT--E-M-F--I-RILQEWNNKNSDKW....................YPWAN-H--SV-DD-AT---KI-I--T--FNNR---THRV-EMWFELEKWEP.............EENLGGNLSVRFLPPSWETNQFVATGAH-.-LIR-R-I-F---AE  538
LST.CD.88.SIVlhoest524       R-V-LATEMLMHT--E-M-F--V-RI-QEWNNKNSDKW....................YPWAN-H--S--DD-AT---KI-L--T--FNNR---THRV-EMWFEMEKWEP.............HEDLGGNLS-KFLPPSWETNQFVATGAH-.-LIRLR-I-F---SE  541
LST.KE.x.lho7                H-A-LATEMLMHT--E-M-F--V-RI-QEWNNKNSDKW....................YPWAN-H--S--DD-ATI--KI-L--T--FNNR---THRV-EMFFEMEKWEP.............HEDLGGNLS-KFLPPSWETNQFVA-GS--.-LI-LN-I-F---DE  536
MAC.US.x.251_BK28            ..----VTFMWT--R---L--KMNWFL-WVEDRDVTT.............QRPK-RHRRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANI-WTDG.....................NQTSITMSAEVAELY-L--GD-.-L-E-T-I-L---DV  514
MAC.US.x.EMBL_3              ..----VTFMWT--R---L--KMNWFL-WVEDRDVTTQ...............RPK-RHRRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANI-WTDG.....................NQTSITMSAEVAELY-L--GD-.-L-E-T-I-L---DV  514
MND-1.GA.x.MNDGB1            -.--RAAEMMMMT-Q--M-F--V-RIMRAWNDPNEKKW....................YPYAS-Q-R--VDD-MQ--RKI-L--T--FNNH---THRV-EMYFEMQKIDS.............NETKMQ-KFLPPSETSNQFVAYGAH-.-L---M-I-I---DV  530
MND-2.CM.98.CM16             TK-E-GAE-IMLL----Y-F--W-KIWRTWNDQNSSVW....................YPWMS-N-R---DD-H---KI-M--V--FNNE----NDV-EMFFEVQKTEE.............GYI-KFVPQDEVQNRFTAVGAH-.QL--VD-I-F---EV  522
MND-2.GA.x.M14               TI-E-GAE-IMIL----Y-F--W-RIWKTWNNQNSNVW....................YPYMS-K-R--VDD-H---RKI-M--V--FNNH---TNDV-EMFFEVQLVEG.............KRYL-KFLPQDEVQNQFTAVGAH-.-L--VD-I-F---DV  523
MND-2.x.x.5440               VK-E-GAE-IMLL----Y-F--W-KIWKAWNSKQSSVW....................YPYMSRN-R--VGD-H---KI-M--V--FNNE---TNDV-EMFFEVQKTDDD.............GYI-KFIPQDWIQNQYTAVGAH-.-L--VD-I-F---DI  545
MNE.US.x.MNE027              ..----VTFMWT--R---L--KMNWFL-WVEDKNLTG.............TTQKPQ-QHKRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANI-WIDG.....................NQTNITMSAEVAELY-L--GD-.-L-E-T-I-L---NV  515
MON.CM.99.L1_99CML1          PQ--K-VQ--W---Q-------VSA--INRRT-K-DG......ISPFDVNNKPNTTYHGGW-A-T-R-MVTQ-GY-S-SI-L--RK-...HVQ-T----A--I-GELYQN.....................NVTLVPSAQVS-S-----SR-.---E-D--SM---T-  503
MON.NG.x.NG1                 PK--N-VQ--W---Q-------LSR--I-SDPXKNKNETFE....STFNVSTNASTAYNGEW-T--LR-LVTQ-GYIS-SX-L--RQ-...HVQ-T----A--VDGELYGS.....................TVTLIPSARVS-A-----S--.---E-D--S----A-  500
MUS-1.CM.01.CM1239           PK--N-VQ--W---A-------LSV--QFNNNITQINE.........SNIHNVTTKYHQDQWMV---R-FVTQ-GY-S-SI-L--RR-...HVQ-T--V----IDGAMYKD.....................TVNMTPSANVM-A--M--KN-.---E-D--SM---EV  518
MUS-1.CM.01.SIVmus_01CM1085  PK--Q-VQ--W-----------LSV--QFVNHTEINE..........SNIHTVTTKYHKDQWMV---R-FV-Q-GY-S-SI-L--RK-...LV--E-------IDGAMYRD.....................S-NMTPSANVL-A--M-PK-VQ--.-E-D--SM---EV  524
MUS-2.CM.01.CM1246           PK--R-VQ--W---Q-------LSV--QFENNTREIND.........SNIHTVKTKYRHDQWMV---R-FVTQ-GY-S-SI-L--RK-...HVQ-T-HV----IDGAMYGN.....................T-NMTPSANV..WT-ASRLRD.T--E-D--S----EV  520
MUS-2.CM.01.CM2500           PK--R-VQ--W-----------LSV--QFENDTRQIHN.........SNIDTVTTKYHKDQWMV---R-FVTQ-GYTS-SI-L--RQ-...H-Q-T--V----IDGAMYGD.....................TVNRLTPSANVL-A--M--KR-.---E-D--S----EV  506
MUS-3.GA.09.09GabOI81        PK----VQ--W-A-D-------LSA--QFENNVTEINSG........NIDNVTSKYKENQWMT-K---FVTQ-GFTS-SI-L--KR-...H-Q-T-------IEGAMYRD.....................S-NMTPSANVM-A--A--SR-.---E-D--S----RV  519
MUS-3.GA.11.11GabPts02       PR--R-VE--W-A-D-------LSV--QFDSVREITDE.........NIHNLTSKYKKNQWML---R-FVTQ-GYTS-SI-L--KQ-...H-N-T-----M--ENAMYGN.....................S-NMTPSANVI-A--A--SR-.---E-D--SM---TV  520
OLC.CI.97.97CI12             R---DAAENMMMM---YMWF--I-KILKTWNNA-DPKW....................YPWSQ--LR-NVDD-M---RRI-L--QP-FNNR---.....QRVTEMFLEIQ.............KVE-GTQVSLLPPS-NVNQFV-VASD-.-L-EL-LI-F---ET  523
RCM.CM.00.SIVagi_00CM312     Y---D-TKYFWL--D---L--KLNWFL-YLNNKTEG...............NEK-RRQAMFV--VT-M-V-D-YT-SRKV-T--RQD...ALT-NATVSY--ADIEYV-Q.....................N-TNVTLSAEVG-L-AA--GR-.---E-K-I-Y---AI  496
RCM.CM.02.SIVrcm_02CM8081    Y---D-AKYFWL--D---L--KLNWFL-LLNNRTDG...............NIN-RRQAMFV--IT-MVV-D-YT-S-KV-T--RPD...ALK-RATV-Y--ADI-YIHT.....................N-TNVTLTAEVG-L-AA--GR-.---E-K-I-Y---DV  505
RCM.GA.x.SIVRCMGAB1          Y---D-ARYFWL--N---L--KLNWFL-LLNNETVG...............TTN-KRKAPFV--IT-M-V-D-YT-SRKV-T--RPD...ALK--AQVSY--ADI-YI-D.....................--TNITLSAEVG-Y-AA--GR-.-AIE-R-I-F---EI  507
RCM.NG.x.NG411               Y---D-AKCFWL--D---L--KMNWFL-YLNNKTEGNK.............N-ARQAMFV--IT-MVV-D-YT-SRKV-T--RPD...AL--NATV-Y--ADI-YTDN.....................MTTNVTLSAEVG-I-AA--GR-.---E-M-I-Y---NV  501
SAB.SN.x.SAB1                QW----SEFFF---Q------KMDWFL-YLNNK-VDPD..H--CAKNN.....TKPCWQRTYV--H-R-VV-D-YTLS-KT----RE-...HLE-N-TA-A-YVELNYNSK.....................NRTNVTLSPQIESI-AN--GD-.---E-K-I-F----V  516
SMM.CI.79.SIVsmCI2           ..S---VAFMWT--R---L--KM-WFL-WVEERN..............MTEQNPE-KKRHNYA--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANINWND.....................GNGTNITFSAEVAELY-L--GD-.-L-E-T-I-F---NV  522
SMM.LR.89.SIVsmLIB1          ..----VTFMWT--R---L--KMNWFL-WVEDRD-NHA.........RWTSQTSK-RKKKNYV--H-R-V--T-HR---NV-L--RE-...DLT-N-TV-S-IANI-WI-.....................NN-TNITMSAEIAELY-L--GD-.-L-E-T-I-L---NV  514
SMM.SL.92.SIVsmSL92A         ..----VTFMWT--R---L--KMNWFL-WVEDRNLTLP.........RWKNQKPE-KAKRNYV--H-R-V--T-HR---NV-L--RE-...DLQ-N-TV-S-IADI-WI-.....................-N-TNITMSAEVAELY-L--GD-.-L-E-T-I-L---NV  519
SMM.SL.92.SL92B              ..----VTFMWT--R---L--KMNWFL-WVEDRN-S-P.........RWTTQTKK-QHKRNYV--H-R----T-HR---NV-L--RE-...DLT-N-TV-S-IANI-WIDN.....................N-TNITMSAEVAELY-L--GD-.-L-E-T-I-M---HV  520
STM.US.89.STM_37_16          ..----VTFMWT--R---L--KMNWFL-WIENR--SE.........MRDWNKNKK-QQKRNYV--H-R-V--T-H----NV-L--RQ-...DLT-N-TV-S-SIIANI-WT-N.....................N-TNITASAEVAELY-L--GD-.-L-E-T-I-L---NV  516
SUN.GA.98.L14                H-A-AATEMLMMT--E-M-F--L-RI-KVWNDTTSNKW....................YPWAN-H--SV-DD-AS---KI-L--T--FNNR---ANRV-EAWF-LERVEDW.............KVNGS-I-VVAFQPPTNTLNQFV-TGAH-.-L-R-R-I-F---DE  543
SYK.KE.x.KE51                PQ--L-VR--W-Q---------VSKI-ANVSNGYANP..................SNYAKNLY-S-A-R----Y-GY-T-L--L--TE-...H-K-T----AV-TDIEYYPN.....................-QLNFAPTANVE-V--AD-FN-.-LIR-K-I-F---SQ  495
SYK.KE.x.SYK173_COMGNM       P---L-VR--W-Q---------VSK--ANITNGNASKN..................NYASNLR-S-A-R---D-RY-R-LI-L--TA-...H-K-T--V-AV-TDIEYYPG.....................-TLNFTPTANVE-V--AD-FN-.-LIQ-K-I-F---DQ  488
TAL.CM.00.266                RR--L-VQLFWHA-Q-------ISSM-LN-SYEYNYTSNQGKLHQGVIGK.KPRSSDRTKFYMS-QLR-M--R-SR-E-L--L--RE-...HLQ-K-------VDINHHRG.....................S-D-TVEPSSQI--S--AAVSR-.--IE-V-F-Y---NI  505
TAL.CM.01.8023               RK--L-VSLFWHA-Q-------ISSM-MK-NYTYNYTA-KGVLHQEKIGRQEHRYSDRTKFFMS-QLR-M--R-S--E-L--L--RE-...HLS-K--V--I-VDINHYKG.....................NDI-SVEPSSQI--A--AAVAR-.--IE-T-I-F---NI  523
TAN.UG.x.SIVagmTAN1          QW----AANIWL--Q------TPDWFV-WLNNE-NSGR..NVDVEGNNCTTGKDKRCYKRTYV--H-RS-V-D-YTLS-KT----RE-...HLE-T-TV-SMMVSL-YNSK.....................ERTNVTLTANLENI-AY--GR-.-LIE---I-F---EI  522
VER.DE.x.AGM3                LF----AANLW---Q------KMDWFL-YLNNRTVDPD..H-PCNGTKGKGKAPGPCAQRTYVA-H-RSV--D-YTLSRKT----RE-...HLQ-T-TV--MSVELNYNSK.....................NRTNVTLSPQIETI-AA--GR-.-L-E-T-I-F---EV  531
VER.KE.x.9063                QF----AANLW---Q------KMDWFL-YLNNQ-VDAN..H-QCSNETKKGNAPGPCVQRTYVA-H-RSV--D-YTIS-KV----RE-...HLE-T-TV--MTVELNYN-N.....................YRTNVTLSPQIEGI-AA--GR-.-L-E-T-I-F---NV  531
VER.KE.x.AGM155              QF----AANLW---Q------KMDWFL-YLNNLTVDAD..H-HCKNNAGKGRSPGPCVQRTYVA-H-RSV--D-YTIS-KT----RE-...HLQ-T-TV--MTVELNYN-Q.....................NRTNVTLSPQIETI-AA--GR-.-L-E-T-I-F---EV  526
VER.KE.x.TYO1_patent         QW----ASNLW---Q------KMDWFL-YLNNKTWDAD..H-FCSSKK.KGHAPGPCVQRTYVA-H-RSV--D-YTLS-KT----RE-...HLQ-R-TV--MTVELNYNSK.....................NRTNVTLSPQIESI-AA--GR-.-L-E-T-I-F---EV  523
WRC.CI.97.97CI14             Q-A-RATENIMMI----M-F--I-RIADAWNNR-ETKW....................YPWS--T-RN-VED-AH--IKV-L--V--FNNH---TQR....VTEILFNWE.....................RNQVMFQAPSNIQNSFVAVGQ--.-L--LKMI-M--GDG  521
WRC.CI.98.98CI04             K-A-RATENIMMI----M-F--I-KIADAWNNKNESKW....................YPWS--T-RN-VDD-AH--VKV-L--V--FRNH---TQR....VTEILFNWE.....................NKQVEFQAPSNIQNSFVAVGQ--.-L-RLKMI-L--GDG  531
WRC.GM.05.Pbt_05GM_X02       Q-A-RATENLMMI----M-F--I-KIADWWNNKTEKRW....................YPWA--T-RS-VDD-ASI--KI-L--V--FNNH---TQR....VSEVLFNWE.....................NGQVEFQAPSH-QNSFVAVGQ--.-L--LKLI-----DG  523
```

fusion peptide

gp120 end gp41 start

glycosylation NAS

immunodominant region

glycosylation NKS

glycosylation NHT

glycosylation NYT

```
H1B.FR.83.HXB2            KRRVVQ.......REKRAV..GIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLL.RAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKS.............LEQIWNHTTWMEWDREINNYTSLIHSLIEE  648
H1A1.UG.85.U455_U455A     -----E.......------..-L--I-----------------I-----------------S---.----------K------------V------Q-----------------T---S-------------.......Q-D---NM--LQ-EK--SS--GI-YQ----  641
H1C.ET.86.ETH2220         -----E.......-----A..-L-----------------I-----------------S---.K-------M-----------T-V--I--H-R------------------------S------------.......Q-E--DNM---Q-----S---DI-YN-L-V  636
H1D.CD.84.84ZR085         -----E.......-----I..-L--M----------------------V---------------.----------------------------------------------H----T---S----R-........V-Y--GNM---Q-E---D---G--YN----  648
H1F1.BE.93.VI850          --Q---.......-----A..-L---------DSREH----I-----------------------.-----------------------V----------------------N---S-------------.......Q-E---NM-----EK--S--SNI-YK----  624
H1G.SE.93.SE6165_G6165    R----E.......------..-L--V-----------------I-----V---------G---.-----------------------L-----------------------N---V-------------.......YNE--DNM--I--E-------YQ-Y--L--  651
H1H.CF.90.056             R----E.......------..-M--S-----------------I-----------------S---.---Q-R--M-------------V--------R-------------N---S-------------.......QSE--DNM------KQ-S---EE-YR-L-V  637
H1J.SE.93.SE9280_7887     -----E.......------..----V------T--------I-----V-----------S---.K-------K--------------V-----------------------N------------------.......Y-D--ENM--IQ-E--------GI-Y-----  640
H1K.CM.96.96CM_MP535      R-----.......------..-L--V-F-------------I-----------------S---.------------------R------------------------------N---S-------------.......SW-E---NM-----EK--G--SDT-YK----  634
H1O.BE.87.ANT70           A-P-IST...RTH------..-L-M----V-S---------AT--A--THT--K-----D---.---Q---Q--R-SX---R--R--L--L-TL-QN----SL---K---V-Y-S-K--RT-..............IGN.-S--DTL--Q----Q-S-IS-T-YEE-QK  644
H1O.CM.94.BCF06           S-PIIGL..NTTH-G-----..-L-M----V-S---------AT--A-RTQGV-K-----D---.---Q-----R-S---R--R--LQ-L-TLIQN--R-SL---K-RI--Y-SAK--NT-G.............NW.TDSA--NL--QQ--QQ-DE-STT-YTK-Q-  663
H1O.FR.92.VAU             A-PTIGT...STH-----A..-LAM----I-S---------ATA---RTQH-IK-----D---.---Q-----R-S---R--R--L--L-TFIQN---HNL---KNR---Y-S-K--KT-GG............DN..-S--DEL--QQ--QQ---VS-F-YEK-Q-  658
H1N.CM.02.DJ00131         ---T-S.......-----A.F-L----------------I-------T-----------V.-----------SI------R-KV--I----R---I-SL------T--Y-T----ET---NTS...........YDT--GNL--QQ---KVR--SGV-FE---K  622
H1N.CM.04.04CM_1131_03    ---T-S.......-----A.F-L----------------I-------T------------.--X----X----SI------X-KV--I----R---I-XL------T--Y-T----ET---XXTS..........YXX--GNL--QX--XKVR--SGX-FX-X-Q  637
H1N.CM.95.YBF30           ---T-S.......-----A.F-L----------------I-------T----------I--.-----------SI--------KV--I----R---I-SL------T--Y-T----ET---NTS...........YDT---NL--QQ--EKVR--SGV-FG---Q  637
H1P.CM.06.U14788          S-PTIMA..HDGH-K---A..-L-M------S---------GI------K---H--------M-.-------E--R-S----R--R--L--I-T--R------L-----QIX-Y-N----ST-T--N..........ETE-DG--GNL--Q---KLVD---DT-YLE-QK  669
H1P.FR.06.RBF168          S-PTIMX..HDAH-K--GA..-L-M-----MS---------AV--------V-H-------M-.-------E--R-S----R--R--L--I-T--R------L-----QIV-Y-N----R--T-............ETE-DGX-TNL--Q---KLVD---DT-YLE-QR  663
CPZ.CD.06.BF1167          --P--GDP.EAKS-S--E..I-M-MI-M---S---------IA------G----------A---.Q-----------------S--------------L-----K--R------L---AN----HSSLE--S--L--TLCGNKTNNYENNYEC.--SQY--QQ--E--R-ISDV-YLSL-T  662
CPZ.CD.90.ANT             --PEIKQ....HS-Q--GI..---LF---L-S---------IA--A-T-N-XH------A---.Q---T--------S---V-------M----K--R-----SL---AD-VT-H-T-----N--V-FTQTC....AKNSSDIQC--ENM--Q----LVQ-S-GQ-YNILQI  650
CPZ.CM.01.SIVcpzCAM13     --HT-G.....ER-Q---A.F-L------------------AV---------------------.-----------------------V------L--------Q---I--L------S--Y-T----KT--G.............KSMSD---NL--QQ--KL-T---GT-FG-L--  649
CPZ.CM.05.LB715           ---T-N.......-----P..-L-LM--------------L-----------T-----S---.--------------S--------------V--I--------------------D-K--ET-...............KSYNE--DNM---Q--K-VR---ET-YD---Q  694
CPZ.GA.88.SIVcpzGAB2      --YT-AKA...LD-S---A.F-L------------------VM------N-------------.-----------------S--------------V----------V-AL------TV-YST-----T--NSNK.............SY-D--SNL--QQ--KLVE-H-GT-FQ-LQR  640
CPZ.US.85.US_Marilyn      --HT-K.......-----A.F-L------------------VV---------------------.-----------------S--------------V-----------I--L------T--Y-T----DT---N............LSYDA--GNL--Q----KVR--SGT-F----Q  621
GOR.CM.04.SIVgorCP684con  --P--S.....HN-----A..-L-M-----MS---------AI--------V-H--------M-.---A---E--R-S----R--R--L--I-T--R------L---T------Y-N----TT-TG-..........SDSE-DG--GNL--QQ--KLVD---DT-FLE-QK  643
GOR.CM.07.SIVgor2139_287  --P-IG.....HN-----A..-L-M-----MS---------AV--------V-H--------M-.---A---E--R-S----R--R--L--I-T--R------L---T-R---Y-N----KT-TG-..........NDTE-DD--GNM--QQ--KLVD---DT-FLE-QR  639
GOR.CM.07.SIVgorCP2135con R-P-IS.....HN-----A..-L-M-----MS---------AV--------V-H------X-M-.---A---E--R-S----R--R--L--I-T--R------L---T-R---Y-N----KT-TG-..........NETE-DD--GNM--QQ--KLVD---DT-FLE-QK  634
GOR.CM.12.SIVgor_BPID1    S-PTIMG....PH----GA..-L-M-------S---------AV-------K---H---------.---Q---E--R-S----R--R--L--I-T-----------L-------V-Y-N----QKN-TTY.........QSDSQ-DA--DNL--Q---KQV----D--FLE-QI  674
GOR.CM.12.SIVgor_BQID2    --PIIG.....LN----GA..-L-M-----MS---------L---------K---H-------M-.---A---E--R-S----R--R--L--I-T--R------L---A---X-Y-N----NT-T-............SXTE--H--ENL--Q---KLVD---XD-FLK-QX  673
GOR.CM.13.SIVgor_BPID15   S-PTIMAQ..THQ----GA..-L-M-----MS---------AV--------K---H---------.-------E--R-S----R--R--L--I-T--R------L---T-QIV-Y-N----QT-TGR..........NDTE-DS---KL--Q---KLVD---DT-FVE-QK  677
```

gp120 end gp41 start

```
MAC.US.x.239              --YTTGG....TS-N--G-.FVL-..-----AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS.................-TPK--NE--Q--E-KVDFLEEN-TA-L--  657
H2A.DE.x.BEN              R-YSST.....PV-N--G-.FVL-..-----AT---A---R-L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N----VT-I-K---H-AQ-NS---AFRQV-H-T---VNDS.................-SPD-KNM--Q--EKQVRYLEAN-SQSL--  642
H2B.CI.x.EHO              --YSSV.....TP-N--G-.LVL-..-----AT---A-----L--SA-S-T--A------QQ-V.DVVKR--E--R-----T-N----VT-I-K-----AQ-NS---AFRQV-H-T---VNES.................-KPD--NM--QQ-E-QVRFLDAN-TK-L--  641
H2G.CI.92.Abt96           --YSSV.....TPKN--G-.FVL-..-----AM---A--T--L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N--T-VT-I-K-----AR-NS---AFRQV-H-T---D-LGA--T...............--PQ--NM--Q--EKQ--FLEDN-TR-L--  662
H2U.CI.07.07IC_TNP3       --YSSA.....TP-N--G-.FVL-..-----AT--TA--T--L--SA-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AS-NA---AFRQV-H-T---VNNS.................-QPD--NM--Q--EVKVKDLEEN-TV-L--  662
H2U.FR.96.12034           --YSSV.....TP-N--G-.FVL-..-----AT---A-----L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N--T-VT-I-K-----AS-NA---AFRQV-H-T---INDT.................-TPN-DNM--Q--EEKV-YLEEN-TQ-L-A  652
ASC.UG.10.RT03            --YEPPPS..KSVVKRA-A.LTL-..-----ST--A---SVATA---S-S--E---A--KA--.E-V-H--Q----S----RN-N--LT-I-K-V---GI-ASY--QF-Q--Y-S----K-FA..............DGA-PAWH-.M--I--E-KVT-H-GI-NT-LV-  664
ASC.UG.10.RT08            --YEPPPS..KSIVKRA-A.LTL-..-----ST--A---SVATA---S-S--E---A--KA--.E-V-H--Q----S----RN-N--LT-I-K-V---GI-ASY--QF-Q--Y-S----K-FA..............DGEALPE--NM--I--E-KVT-H-GI-NG-LV-  668
ASC.UG.10.RT11            --YEPPSNT.KSVVKRA-A.LTL-..-----ST--A---SVATA---S-S--E---A--KA--.E-V-H--Q----S----RN-N--LT-I-K-V---GI-ASY--HF-Q--Y-S----I-FA..............DGGVLPD--NM--I--E-KVT-H-GI-NN-LV-  659
COL.CM.x.CGU1             --S-NVN..WHHG-Q--GI.FAFS..I-AL-SG--AA--S--VA--I--QS-NGRASASS-RM-LKLV-T-SA--------V-N--V-VATI-G--EE-AK-ASI--ANMQ--R-I-----KT-...............GE-DP-QNM--KQ-HERVR---DI-EADLV-  644
COL.UG.10.BWC01           --Q-AIN..WHHG-Q--GI.FAFS..I-AL-SG--AA--S--VA--I--QS--E-L----RM--.KLV-T-SA--------V-N--V-VATI-S--EE-AK-ASI---NMQ--R-I-----KT-...............GM-EP-QNY--KQ-YEQVA---NI-EADLV-  650
COL.UG.10.BWC07           R-GGTA.....VV----GI.FAFS..LIAL-SG--AA--S--VA--I--QS--D-L-R--RQ--.KLV-T-SA--------V-N--V-VATI----EE-AK-AVI--ANLQV-R------ST-...............GD--P-MNM--KQ-HERVK---DV-QEELQ-  636
DEB.CM.04.SIVdeb04CMPF3061 R-YEGPE....SV-Q----GMAL-..LVA--ST--AA-----TA----S-S---------QE--.K-V--HG---T--A--V-N-NT-LT-L-K-----SK-NE---AF-Q--H-T-----G-I...............GQPD--NM--Q--EQKVA---DIWEEALQK  627
DEB.CM.99.CM40            R-YEGPE....SV-H---AGIAF-..LVA--ST--AA-----TA----S-S---------QE--.K-V--HGQ--T--A--VRN-NT-LT-I-K-----AK-NE---AF-Q--H-T-----N-L...............EDPD-DNM--Q--EMKVA---DEWEGALQR  650
DEB.CM.99.CM5             R-YEGPE....SA-Q---AGIAF-..LVA--ST--AA-----TA----S-S---------QE--.K-V--HG---S--A--VRN-NT-LT-I-K-----SK-NE---AF-Q--H-T-----HT-...............G-PD--NM--Q--E-KVA---DEWEGALQR  630
DEN.CD.x.CD1_CMO580407    R-YEGPESAESVS-RR-E-SLVL-..LI---S---TV---VGTG-A--S-D--A------QE--.--V-GHSA----S-----N-N--LT-I-K-----A--NE---AW-Q--H-S-E-SNQS..........IPGLALYQPDFQNM--QQ-ETD-AALFGNVSDALKK  657
DRL.DE.11.D3              H-YNLPE.....H-Q--GA.VVL-..I--L-SL------SVATAM---SQA-------E--KQ--.-LV-Q--E--K--I--V-N----LT-L-E-VG--AR-SL----FAQV-H-N-V-PNES...............VTPN-TSE-----QKRVDSISNN-TLDLQK  697
DRL.DE.11.D4              H-YNLPE.....H-Q--GA.VVL-..I--L-SL------SVATAM---SQA--N---E--KQ--.-LV-Q--E--K--I--V-N----LT-L-E-VG--AK-SL----FAQV-H-N-E-PNET...............VTPN-TSE-----QKRVDSISNN-TLDLQK  681
DRL.DE.11.D5              H-YNLPE.....H-Q--GA.VVL-..I--L-SL---A--SVATAM---SQA--N---E--KQ--.-LV-Q--E--K--I--V-N----LT-L-E-IG--AK-SL----FAQV-H-N-L-PNES...............VTPN-TSE-----QKRVDSISDN-TLDLQK  689
DRL.DE.11.D6              H-YNLPE.....R-Q--GA.VVL-..I--L-SL------SVATAM---SQA-------E--KQ--.-LV-Q--E--K--I--V-N----LT-L-E-VG--AR-SL----FAQV-H-N-V-PNES...............VTPN-TSE-----Q-RVDSISNN-TLDLQK  694
DRL.x.x.FAO               H-YNLPG.....H-Q--GA.VVL-..I--L-SL---A-CSVATAM---SQA--T-M-E--KQ--.-LV-Q--E--K--I--V-N----LT-L-E-IG--AM-SL----FAQV-H-N-V-PNES...............VTPN-TSE-----QKRVDSISNN-TLDLQK  693
GRV.ET.x.GRI_677_gri_1    R-YTGPT.......----VP.FVL-..---------TA----AT-----S-H--A--L---K---.A-V-Q--Q--K--I--V-N-N--VT-L-K--E--AR-NS---AW-QV-H-T---KYNN.................TPK-DNM--L--E-Q--ALEGN-TQ-L--  653
GSN.CM.99.CN166           --KEHP......AV-----.LSL-IS--T--S---T-----ATA--E-S-S--A--M---E---.--V----S----PS---------T-LSSL-K--R--TI-QA---ANRP--H-I----T-..............WANGSLPD-ENM--QK-SMLVE-D-YT-QQ-L-Q  659
GSN.CM.99.CN71            Q--PHPG.....VHVK---.ISL-IS--T--S---A-----ATA---S-S--A-------E---.--V----S----S------------LSSL-K--R--TI-QA---ANQP--H-I----D-..............WAKNSTPD-E-M--Q--SKL-E-D-YT-QQ-L-N  668
LST.CD.88.SIVlhoest447    H-YAP......SSKQ---APLAL-..A--L-S---TA--LV-TI-----QVVIQ--L---KQ--.VLV-K--E--R--I--V-N----LT-I-E-----T--AS---QW-QV-H-N-E--YN..............ITPN-TRD--I----QVGVLEAN-ST-LQ-  682
LST.CD.88.SIVlhoest485    HQYAP......SSKQ---APLAL-..A--L-S---TA--LV-TI-----QVVIQ--L---KQ--.VLV-K--E--R--I--V-N----LT-I-E-----A--AS---QW-Q--H-N-E--YN..............ITPN-TRD--I----RVGVLEAN-ST-LQ-  681
LST.CD.88.SIVlhoest524    H-YAP......RG-QT--APLAL-..A--L-S---TA--LV-TI-----QVVIQ--L---KQ--.VLV-K--E--R--I--V-N----LT-I-E-----A--AS---QW-QV-H-N----YN..............VTPN-TRD--I--E-QVGSLEAN-TT-LQ-  684
LST.KE.x.lho7             H-YAP......RG-QT--APLAL-..A--L-S---TA--LV-TI-----QAV-Q--L---KQ--.VLV-K--E--R--I--V-N----LT-L-E-V-H-A--AS---QW-QV-H-N-E-TYN.............ITPN-TKD--R--ESKVAI-DKN-T--LQ-  679
MAC.US.x.251_BK28         --YTTGG....TS-N--G-.FVL-..-----AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS.................-TPD--ND--Q--E-KVDFLEEN-TA-L--  659
MAC.US.x.EMBL_3           --YTTGG....TS-N--G-.FVL-..-----AT---A-----FR--A-S-T--A------QQ--.GVVKR--E--R-----T-N--T-VT-I-K--E--AQ-NA---AFRQV-H-T---PNAS.................-TPD--ND--Q--E-KVDFLEEN-TA-L--  659
MND-1.GA.x.MNDGB1         --HTLPE....HHK---GA.VIL-..I--L-SL---A--SV-VA----SQS-VT---E--KQ--.KL--Q-SE--K--I--V-N--T-LTSL-N-I---A--SQ----WAQV-H-S-E-TNTS..............ITPN-TSE--K--ETRTDYLQQN-TEMLKQ  675
MND-2.CM.98.CM16          A-YHLPE.....A-Q--GA.VLL-..MF-L--L------SVAVA----SQA--N---E--KV--.SL-DQHSE--K--I--V-N----LT-L-D-VA--AR-SM----FAQV-H-H---PNDS..............ITPN-TSE--L---KRVTAL-DNMTVNLQK  666
MND-2.GA.x.M14            H-YHLPD.....AKQ--GA.VLL-..M--L--L---A--SVAVA----SQA--N---E--KI--.SL-DQHSE--K--I--V-N----LT-L-D-VA--SR-AV----FSQV-H-N---PNES..............ITPN-TSE--L----RVTAI-NNMTIDLQR  667
MND-2.x.x.5440            H-HHLPN.....T-Q--GA.VLL-..M--L--L---A--SVAVA----SQT--N---E--KV--.SL-DQHSE--K--I--V-N--V-LT-L-E-VA--SR-SV----FSQV-H-S-K-PNNS..............IVPN-TSE--L----RV-SIVTNMTIDLQR  689
MNE.US.x.MNE027           --YTTGG....TP-N--G-.FVL-..-----AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS.................-TPN--NE--Q--E-KVDFLEEN-TA-L--  660
MON.CM.99.L1_99CML1       Q--TGV......H-----I..TL-MA-----ST---G-----ATA----S-S--A-------E---.--VT---S---------V------LT---KFI---T--NA---AN-AV-H-T-----N-.............WAKGHFPE-DNM--QQ-SELVD-D-MT-QQ-L-A  650
MON.NG.x.NG1              Q--HGL......M--R--I..TL-MA-----ST---G-----ATA----S-S--A-------E---.--VT---S----S-----------LT---KFI---T--NS---ANRAV-H-Q-L--NT.............WAKGHFPE-DNM--QQ-SMLVD-D-A--QX-L--  647
MUS-1.CM.01.CM1239        Q--P-P......HVKA--I..SL-IT-----S---G-----ATA----S-S--A-------A---.--V---------S-----------LT-L-KFI---A--NL---ANRQ--H-R----T-..............WANHTLPG-ENM--QQ-SKLVD-D-NT-QG-L-A  665
MUS-1.CM.01.SIVmus_01CM1085 Q--P-P......HVHA--I..SL-IT-----S---G-----ATA----S-S--A-------A---.--V---------S-----------LT-L-KFI---A--NL---ANRQ--H-R----D-..............WANHTQPG-ENM--QQ-S-LVD-D-TT-QE-L-L  671
MUS-2.CM.01.CM1246        R--P-P......HVKA--I..SL-IT-----S---G-----ATA----S-S--A-------A---.--V---------S-----------LT-L-KFV---AI-NL---ANRQ--H-S----D-..............WANHTQPG-DNM--QQ-S-LVD-D-AV--G-L-M  667
MUS-2.CM.01.CM2500        Q--P-P......HVQA--I..TL-IT-----S---G-----ATA----S-S--A-------A---.--V-T-------S-----------LT-L-KFI---A--NL---ANRQ--H-R----N-..............WANHTQPE-DNM--Q--S-LVD-D-AT-QG-L-L  653
MUS-3.GA.09.09GabOI81     P--EAP......RLHT--I..SL-IT--T--S---G---S-AIA----S-S--A-------Q---.--VQT-------S-----------MT-I-K--R--T--NS---ANRA--H-NL---K-..............WANETMPG-DNM--Q--S-L-D-D-V--SN-L--  666
MUS-3.GA.11.11GabPts02    Q--ETP......RLHT--I..SL-IT--T--S---G---S-AIA----S-S--A-------Q---.--VQT------MS-----------LT-I-K-----T--N----ANRA--H-N-------..............WANHTQPG-DNM--Q--S-LVD-D-LE-ER-L-L  667
OLC.CI.97.97CI12          Q-YQP.........-N--GA.LVL-..L--V--L---A--TV-TI----SSH--A---R--KHM-.EM--K--E--K--I--V-N----L-G--K-----A--KQ---EFMQV-H-T-----YT.................QSPD-GND--LK-Q-KVD-L-QE-DDRLTI  663
RCM.CM.00.SIVagi_00CM312  --YET.........KQ--VP.LVL-..-----S---TA----ATA----S-H--A--L---KQ--.DIV-Q-----K-----T-N----VT-I-K--A--S--NTF--AWRQV-H-S---SF..................NKTPD-QKQ--L--E-NVSYLEAN-TASLQ-  636
RCM.CM.02.SIVrcm_02CM8081 --YET.........-Q--VP.LVL-..-----S---TA----ATA----S-H--A--L---KK--.DIV-Q-----R-----T-N----VT-I-K--A--S--NTF--AWRQV-H-T-Q-TF..................NATPE--KQ--L--E-N-SIIEGN-SLALQ-  645
RCM.GA.x.SIVRCMGAB1       --YQT.........KQ--VP.LVL-..-----S---TA----ATA----S-H--A--L---K---.DIVKR--N--K-----T-N----VT-I-K--A--S--NTF--AWRQV-H-V---TF..................NKTPE-QKES-LQ-E-N-SYLEAN-TIALQ-  647
RCM.NG.x.NG411            R-YET.........KQ--VP.LVL-..-----S---TA----ATA----S-H--A--L---KK--.DIV-Q--E--K-----T-N----VT-I-K--A--S--NTF--AWRQV-H-T-E-IY-Q..................TPE--KQ--L--E-N-SRLEGN-SVALQD  641
SAB.SN.x.SAB1             R-YTG......PE-Q--VP.FVL-..----------AA----ATA----SQ---A--L---K---.A-V-Q--QM-K--I--V-N-N--VT-L-K--E--AR-N----AFRQV-H-T-L-KYNN.................TPD-ENM--Q--E-Q-EK-EAN-SRIL-Q  658
SMM.CI.79.SIVsmCI2        --YSSAT....PA-N--G-.FVL-..-----AT---A-----L--SA-S-T--A------QQ--.DVVTR--E--R-A---T-N--T-VT-I-K-----AR-NA---AFRQV-H-T---ENNT.................-QPD--NM--Q--EIK-RDLEAN-SESL--  667
SMM.LR.89.SIVsmLIB1       --YST.......M-P--G-.FVL-..-----AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N----VT-I-K-----AQ-NS---AFRQV-H-T-E-PNKT.................-QPN-ENM--Q--E-KVDFLEAN-TQ-L--  656
SMM.SL.92.SIVsmSL92A      --YTTST....S-N--G-.FVL-..-----AT---A-----L---A-S-T--A------RQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T---PNDS.................-TPD--NM--Q--E-KVDFLEAN-TLML--  663
SMM.SL.92.SL92B           --YTTS.....TSKN--G-.FVL-..-----AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T-L-PNDS.................-VPD--NM--Q--EKKVEFLEAN-TQML--  664
STM.US.89.STM_37_16       --YTTS.....TS-T--G-.FVL-..-----AT---A-----L---A-S-T--T------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T---PNDS.................-VPD--NM--Q--E-KVDFLEAN-TQ-L--  660
SUN.GA.98.L14             H-YAP.......RK----APVAL-..A-AL-S---TA--LV-TI-----QAV-Q--L---KQ--.VLV-K--E--R--I--V-N----LT-L-E-VQ--S--AS---QW-QV-H-N----YN..............ITPN-TKD-------QVKM-DDN-TA-LQ-  685
SYK.KE.x.KE51             R-YELP.....TKQ----APLAL-..---L-S---TA--S-ATA--L-SQT--A------QK--.E-V-------G-----V-N-N--LT-L-T--R--AIMSN---AF-Q--H---T-QQACG-N-............RCPTPQ-ENM--HT-E-QVD-L-DH-DN-LR-  645
SYK.KE.x.SYK173_COMGNM    R-YELP.....NT-----APLAL-..---L-S---TA--G-ATA--L-SQT--A------QK--.E-V-------G-----V-N-N--LT-L-T--R--AI-SN---AF-Q--H---T-EKACG-N-N...........FCPKPQ-KNM--HR-EQ-VD-L-DH-DG-LR-  640
TAL.CM.00.266             R-YDGP.....ANKQ---APLAL-..-I---ST--TA---VATA----S-S---------EH--.-----H----------------N-N--LT-L-K--E--AK-NS---AW-Q--Y-S----KT-.............WSNYTDPQ-QNM--Q--EMKVD-H-G--SQ-LQ-  653
TAL.CM.01.8023            R-YDGP.....ATK----APLAL-..-I---ST--TA---VATA----S-S---------EH--.-----H----------------N-N--LT-L-K--E--AR-NS---AW-Q--Y-S----KT-.............WTNSTNPD-QNM--Q--EKLVD-ASDT-TV-LQ-  671
TAN.UG.x.SIVagmTAN1       R-Y-GPT.......----VP.FVL-..---------AA---TATA----SQ---A--L---K---.A-V-Q--QM-K--I--V-N-N--VT-L-K--E--TR-NL---AF-QV-H-T---TFNN.................TPD-DNM--Q--ESQ-TALEGN-STTLVK  664
VER.DE.x.AGM3             R-YTGG.....HD-T--VP.FVL-..---------TA----ATA----SQH--A--L---K---.A-V----QM-K--I--V-N-N--VT-L-K--E--AR-NA---AW-QV-H-T---QWNN...............RTPD--NM--L--E-Q-SYLEGN-TTQL--  675
VER.KE.x.9063             R-YTGG.....QD-S--VP.FVL-..---------TA----ATA----SQH--A--M---K---.A-V----QM-K--I--V-N-N--VT-L-K--E--AR-NV---AW-QV-H-T---QWQN...............MTPN-QNM--L--E-Q-GELEGN-TEQLVK  675
VER.KE.x.AGM155           R-YTGG.....QE-Q--VP.FVL-..---------TA----ATA----SQH--A--L---K---.A-VG---QM-K--I--V-N-N--VT-L-K--A--AR-NA---AW-QV-H-T---TWNN.................TPE--NM--L--EKQ-EGLEGN-TKQL-Q  669
VER.KE.x.TYO1_patent      R-YTGG.....HE-Q--VP.FVL-..---------TA----ATA----SQH--A--L---K---.A-V----QM-K--I--V-N-N--VT-L-K--E--AR-NS---AW-QV-H-T-E-PWTN...............RTPD-QNM--L--E-Q-ADLE-N-TRQLVK  667
WRC.CI.97.97CI14          --V-PPE....HR----GA.IIL-..I-----L---A--TV-TV--I-SQH--A------K---.ELV-K--E--R--I--V-N----LS-I-GF-E--LK-KQ---ELTQV-H------Q-...............WNATPN-ENE--QH-YKRTDELQAN-SAILVI  667
WRC.CI.98.98CI04          R-VTPKE....HH----GA.IVL-..I-----L---A--TV-TV--I-SQH--A--L---K---.ELV-K--E--R--I--V-N----LS-I-GF-E--LK-KQ---ELTQV-H-N----E-...............WNATPN--NE--QH-Y-RTDELQAN-SV-LV-  677
WRC.GM.05.Pbt_05GM_X02    --VTPPD....HH----GA.IVL-..L-----L---A--TV-TV--I-SQH--A------K---.ELV-K--E--R--I--V-N----LS---G--E--VK-KQ---ELTQV-H-----EEA...............WNMTPA--NE--QQ-HLRTDTLM-N-SDMLVQ  669
```

```
                                                                               gp41 cytoplasmic tail start
                                                          transmembrane domain                                                         glycosylation NGS
H1B.FR.83.HXB2                SQNQQEKNEQELLELDKWAS.LWNWFNITNWLWYIKLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPT.......PRGPDRPEGIEEEGGERDRD.RSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLG.......RRGWEALKYW...... 796
H1A1.UG.85.U455_U455A         ---------LD--A-----N.-*-----S------R--VI-----I------T----I---------------LA-I.......-E-LG--GR-------QGK-.------S-F--IA-----N-----------FA---A-A-----RSSLKGL-L---G---L...... 795
H1C.ET.86.ETH2220             -----D---KD--A----EN.---------------I-------VI----I-----------------------LI-H.......------LG--------QG--.---------F--IF------------------I--AA-T-----RSSLKGLQ----T---L...... 791
H1D.CD.84.84ZR085             --I------K----------.-----S--Q-----I--------I--------------------------L--A.......----------------Q--G.T-------FS-------N----------E-I--AA------...........-------L...... 796
H1F1.BE.93.VI850              -------------A------.-----D-S-------I--------I-----------------K------L--LI-S.......-------------G---QGK-.--V---T-F---A-----N------RH---FI--AA---DRGL........-------L...... 772
H1G.SE.93.SE6165_G6165        ----------D--A--Q---.-----G--R-----I--------I--------------------------LTHH.......Q-E---------G---QG-G.--V---S-F-P-------------------SI---A-T-----RSSLKGL-L---G---L...... 806
H1H.CF.90.056                 --T-------D--A------.--T--D-SH-----I--------I----I-------------------LV-N.......-----------T--G---Q---.--V-----F-PVV-------S----RL--------V-T--------........R----L...... 785
H1J.SE.93.SE9280_7887         A-----N--KD--A----TN.-------S-------I----I---I----I----A--------------LI-N.......-TEA---G----G---QG-T.--------F---A-----N-----------FV--AA-T-GT--.......L----I---L...... 788
H1K.CM.96.96CM_MP535          --T-------D--A------.-----D--K-----I----I---I----A-----V--------------LI--.......S--A------------Q-KN.--V---S-F---A-----N------RQ--N-I-------L-RGL.......-G-------L...... 782
H1O.BE.87.ANT70               A-V---Q--KK-----E---.I---LD--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.KIINV 795
H1O.CM.94.BCF06               A-D---Q--KT-----E---.-----D--K------IA-I---A-I-I-V-MI--NL-KNI----Q---L-IPI-H.......QEEAET-GRTG----GV--R.KWTP-QP-F-Q-LYT---TII-WT--L-SN-ASGIQ-........LISYLGL-LWI-.GQ.KTIEA 814
H1O.FR.92.VAU                 A-E------K------E---.I---LD--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.RTIEA 809
H1N.CM.02.DJO0131             A-E--NT--KS-----Q---.-----S---------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...... 779
H1N.CM.04.04CM_1131_03        A-E--NT--KS-----Q--X.-----D--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-LRXRL...... 792
H1N.CM.95.YBF30               A-E--NT--KS-----Q-D-.--S--G--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...... 792
H1P.CM.06.U14788              A-E--KE--RK---------Q.--S-MD--K------I-------II--K-LM-IGNV-RK--------V-L--LRLH.......GGPAGIAP-TD-----AGNG.-----LE-L-P-------N-VVQI-QI-VSCI-RIKDLLTI....LWIHLG.QLLNS........N 817
H1P.FR.06.RBF168              A-D--KA--KK-----Q--Q.----LD--Q------I-------II----LL-XXNV-R-I-------V-L--LGLN.......GDPAGIAP-TN-----AGNG.-----LD-F-P-V----KN-VVQI-QI-VGCI-GIKDLLTI....LWIHLG.QLLTR.GL.....N 813
CPZ.CD.06.BF1167              A-V---R-HKQ-------S-.-----D----------IA-I--AS-------MFIINL-GKL-----L--P-IPIQ-A.........----GQ---T--G---SG--.-WRGYPV-F-Q---E-C-Q-L-WI-QTSVN--WHLWTLIK-IHQWSLLLWETVRQL-RRLIHLLT. 822
CPZ.CD.90.ANT                 AHE---R-KK--Y-----S-.-----D--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...... 799
CPZ.CM.01.SIVcpzCAM13         A-S------KD-----Q---.-----D----------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...... 804
CPZ.CM.05.LB715               A-----Y--KS----NQ---.-----D-S-------I---------I-C-----L-------------------LF-I...........----EQ-GETG-----Q-NV.-----L--F-P-V-N---D--R-L------F---AR-TL--V-QHLLKGL-LV--T-L-L...... 849
CPZ.GA.88.SIVcpzGAB2          AHE--NS--K-------Q-S-.-----D---------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...... 795
CPZ.US.85.US_Marilyn          A-E--NT--KS------Q-S-.-----D---------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...... 776
GOR.CM.04.SIVgorCP684con      A-E---A--KA--------D.--S-LD--Q-I---RI--IVIA-----K-LM-IIN-CH-------------ILGRN.......QVPAGIVPE-G----RA-S-.------LA-F-P-------T-VVWI-RT-VI--SGIKE-TLS-IEXLTRLLX-VNNLTRDC...... 798
GOR.CM.07.SIVgor2139_287      A-E---A--KA--------D.--S-LD--Q------I--I-IA-------LM-IINMCH-------------ILGRN.......QVPAGIVPE-G----RA-N-.------LA-F-P-----I-T-VVWI-RT-VI--SGIKE-TLS-IEYLTRLL--VNNLTRDC...... 794
GOR.CM.07.SIVgorCP2135con     A-E---V--KA--------D.--S-LD--Q------I----IA-------LM-IINMCH-------------ILARN.......QVPAGIVPE-G----RA-S-.------LA-F-P-------T-IVWI-RT-VI--SGIKE-TLS-IEHLTRLL--VNNLTRDC...... 789
GOR.CM.12.SIVgor_BPID1        A-E---Q-QKK-----Q--Q.--S-LD--Q------I---------I----FI--VNV-K-----------L--LGQN.......RGPAGIAT-TD----GAGSV.------LD-F-P------KN-VVWI-QI-VSCI-GIKDLLAI-WIHLRHLLNK-LNCVRDC...... 829
GOR.CM.12.SIVgor_BQID2        ANT---V--KK--------D.--S--D--K------IA----XA-I---V-MV---L-KNI-K--Q---L--PS-H.......-TE-GT--KTGG----G-KH.KWTP-PP-F-H-LYL---TII-W---L-SS-VSRIQ--LS---LGLRILGQKTF--C-TL...... 828
GOR.CM.13.SIVgor_BPID15       ANE--KE--KK-----Q-TQ.--S--D--K------I---------I--KFLL-IIN--K-----------L-ILGLN.......RGLAGIAP-TD-----I-NG.-----LD-L-PIV----KN-VVWI-RI-ATCI-GTKDLLNI-WIHLRHLLNK-INC-RDC...... 832

MAC.US.x.239                  A-I-----MY--QK-NS-DV.FG---DLAS-IK--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIQQ..D-AL-TREGK-RDGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRVYQI-QPILQRLSATLQRIREVLRTE.... 819
H2A.DE.x.BEN                  A-I-----MY--QK-NS-DI.-G---DL-S-VK--QYGVH--V-IIA---AIY-VQLLS-F-K--R-VFSSPPGYLQQIHIHK..DRGQ-ANEGT-EDVGGDS.GYDLWPWPINYVQFLIHLLTR-LIGLYNICRD-LSKNSP....TRRLISQSLTAIRDWLRLK.... 800
H2B.CI.x.EHO                  A-I-----MY--QK-NQ-DI.FS---DF-S-MA--R-GLY--I-I-V---AIYIIQMLA-L-K--R-VFSSPPSY-QQIPIRK..DRGQ-ANE-T-E-GGNNE.GYRSWPWQIEYIHFPIRQLRD-LIWLYSGCRT-LSKTFQT-QPVLQPLRLPPA............ 794
H2G.CI.92.Abt96               A-I-----MY--QK-NS-DV.FG---DL-S-VK-VY-GLYV-A-VIV----IY-VQ-LG-L----R-VXSSPPSYVQQIHIRK..DQEX-TKE-I--ESGNKG.GYRSWPWQIEYIHFLXRQLGN-LTWLYSNCRA-LSRIXQT-QPLFQRISRTLQAIREHLRLE.... 824
H2U.CI.07.07IC_TNP3           A-I-----MYQ-QK-NS-DV.FG--LDL-S-VK-VY-GFYV-A-II---VIIY-VQLLS-L-K--R-VFSSPP-YVQQIHTHE..DQEQ-TKE-T--DGG-NV.GFNSWPWQIEY-HFLIRQLIR-LTWLYSNFRAWLSQSFQS-QAACQRVLTAIREHLRLEGAYLQYG 828
H2U.FR.96.12034               A-I-----MY--QK-NN-DI.FG---DL-S-VK-VY-GLYV-A-III----IY-VQLLGSL-K--R-VFSSHPSYVEQIPIQR..DQEQ-TKG-I-E-AG-SG.GYRSWPWQIRYIHFLIHQLIR-LTWLYSSCRD-LSRICQS-QPLF.......QSIRERLHLE.... 807
ASC.UG.10.RT03                A-RR--E-THK-QK-GE-DN.-----D-SK-FQW--IAVL--I--IA----MWLIN-LKM--A--L--AQTPFHQPQV.........-G-RGETD----DG-NV.--V--L--F--VV-----AIV-WI-RT-VN-GWLIRHLWNG-TDWWFGRAYPRLILAAEATAEATR 824
ASC.UG.10.RT08                A-RR--E-THK-QK-GE-DN.--S--D-SR-FQW--IAVL--I--IA----MWLIN-LKM--A--L--AQTPFHQPQV.........-G-R-ETGG---DG-NV.--V--L--F--VV-----AIV-WI-RT-VN-GWLIRHLWSS-SDWWFGSAYPRLILAAEATAEATS 828
ASC.UG.10.RT11                A-RR--E-THK-QK-GE-DN.--S--D-SR-FQW--IAVL--I--IA----MWLIN-LKM--A--L--VQTPFHQPQV.........-G-RGETGG---DG-NI.K-V--L--F--VV-----AIV-WIHRT-VN-GWLIRHLWSS-TDWWFGSAYPRLILAAEATAEATR 819
COL.CM.x.CGU1                 AYDL--E--KK-A--GD-TN.WFSG-GLF-IFK-VLYAAYV----I----IMV-IACIRGAFRVKGFQQIGRTNVSS......QIRGQDR-WEQPDNA-EIHREQ.ENSSSRF..............WRLMEEWWSSLL-HCSNGILTVLPQHPPTTSDKRRS...... 785
COL.UG.10.BWC01               AYDL--A--KK-A--GD-TN.WFSG-GLF-IFK-VLYAAYVI--IIAF--LML-I-CIRGAYRLRGFQRLGREPLSS......QIRGQNQ-WEQPDN--ELDSEQEDNNSRLVQ..............LLKGL.WKRFAQACSNGLHTLMSQRPETFNTRRTVFQT... 794
COL.UG.10.BWC07               AYEL--E--KK-A--GD-T-.WFSG-G-FS-FK--MYAGYAIV--I----GWLLI-CFSSGFRRRKFQRLGREEELTTIL...QIPTQTQ.WEQPDNA-ELD..SDSGSSSSRF.............WLLVKEWWTHLS--CRNGIHTLQGEQPLPDGETHQPRR.... 781
DEB.CM.04.SIVdeb04CMPF3061    A-EK--Q-VHA-QS-QD-D-.-----DLSR-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-QRLGEQLRV 785
DEB.CM.99.CM40                A-E---R-VHA-QS-QD-D-.-----DLSR-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-RLLGERFRV 809
DEB.CM.99.CM5                 A-E---R-VHA-QS-TD-D-.-----DLSR-F-W-R-VVY-IAA-IL---AMFGVN-GTKLCR-----QTL-PK.........EVEVLQLG-EGDKS--GVKS.-WTTSPR-F-R-----ISQ-LTWL-QT--NCAWLIR-GF---QETIYRLSFAF-QR-RLLGEHLRV 789
DEN.CD.x.CD1_CM0580407        A-A---R-VH-VQS-YD-DN.-----DLSK-F-W-R-VVYVIA---I---IMFIINVLS-LCR-----LQIPIR.........-QKAANQPEEG-GESGDGNRLKWWQYPR-F-SI--E--SQ-LTW--LT--N-VFLIRTACSRIWDFV..........HRHSLILLG. 806
DRL.DE.11.D3                  AYE-EQ--IF--QK-GDLT-.WA---DF-W-FK---IGFF--IAII----LA-LW-T-G-F----R--PYVFKDAYRPET..TNYHRQPDKEKG--QDRGEWN.IRSEPSKPESSRS-SKE.TVGTWL.KES-GYIWLKNVKAV-EYGWSE.................. 842
DRL.DE.11.D4                  AYE-EQ--IF--QK-GDLT-.WA---DF-W-FK---IGFFV-IVII----LA-LW-T-G-F----R--PYIFKDAYRPET..TSYHRQPDKEKG--P-RDEWN.TRSEPSRPESSRS-SKE.TVETWL.KES-GYIWLKNL-AV-EYGWSE.................. 826
DRL.DE.11.D5                  AYD-EQ--IF--QK-GDLT-.WA---DF-W-FK---IGFF-IVVIR----LA-LW-T-G-F----R--PYIFKDAYRPET..TNYHRQPDKEKG--PDRDEWN.IKSEPSKPESSKP-SKE.TVGTWL.KES-GYIWLKNVKAV-EYGGSE.................. 834
DRL.DE.11.D6                  AYE-EQ--IF--QK-GDLT-.WAD--DF-W-FK---IGFF--IAII----LA-LW-T-G-F----R--PYVFKDAYRPET..TNYHRQPDREKG--QDRGEWN.IRSEPSKPESSRS-SKE.TVGTWL.KES-GYIWLKNVKAV-EYGWSE.................. 839
DRL.x.x.FAO                   AYE-EQ--IF--QK-GDLT-.WA---DF-W-SK---IGFF--MAII----LA-LW-T-G-F----R--PYIFKDAYRPET..TNYHRQPDREKG--QDRGEW.NIRSEPSKPGSSKA-SEE.TVGTWL.KES-GYIWLKNVKAV-EYGWSE.................. 838
GRV.ET.x.GRI_677_gri_1        A---ES--LDLYQK--D-SG.F-S--SLST--G-V-IGFLVIVIIL---FAWVLWGCIRNI----N--P........QIHIHS..SAE--DNGGGQDRGGESSSSKLIRLQEESSTPSRINNWWLN-KSCS--IRTWCYNICLT--IFIRTAVG................ 796
GSN.CM.99.CN166               ANQ--AS-LN--MK-S--D-.--S--D-SD-QR---I-VIV-AA-IA----MFI-NMLR-I--------P-ILI-QAV.......ERGP-DAT--GV-AQGKV.--V-YLT-FSS-L-----N-VIWI-QI-AS-AWTLRG-GKF-WEQLQKAIQKAIQQARQLREVAAR 820
GSN.CM.99.CN71                ANH--S--MND--K-S--D-.--S--D-S-------I---V-AA--A---IMF--NMLR----------P-ILT-QQV.......ER-P-D-TRKD--AQGNV.--V-YLT-FSS-L-------VIWI-QI-AT-VWTLQ--GKF-WEQ....LQKAFQQARQLGEVAAR 825
LST.CD.88.SIVlhoest447        AYTTELE-RNAFKK-QEFN..F---LD-LS-FQ---YAVL-II-IIV--V-SFIVQNIVKMCR--RV-APSAYVEQDCKWEKR.ENQEQ-DKEG-IEKDKENIYINLEQYKKESSTPPWNVDWSEPLQDSLFVT--KWIKAFGI--LSLVWQFLSWLGHLVILFFQ.HGQ 848
LST.CD.88.SIVlhoest485        AYTTELE-RNAFKK-QEFN..F-S-LD-LS-FQ---YAVL-II-IIV----SFIVQNIVKMCR--RV-APSAYVEQDYKWEKG.EN-EQ-DKEG-IEKDKEDIYINLEQCKKESSIPPWNVDWSEPLQDSLFVT--KWIKAFGI--LSLVWQFLSWLGHLVILFFQ.HGQ 847
LST.CD.88.SIVlhoest524        AYTTELE-RNNFKK-QDFN..F-S-MDL-T-FQ---YAVL-II-III---LSFIIQS-VKMCR--RV-APSAYVEQDFKWENG.ESQEQ-DKEG-TEKD-ENIYINLEQCKKESFRPPWNVDWNEPLQDSLFVT--KWLKAWGI--LTLVWQFLTWLGHIVILSYH.HGQ 850
LST.KE.x.lho7                 AYTTELE-QNKFKK-QEFN..F-S-LD-SH-FT-V-YAVL-ILVII---VLSFIIQN-VKMCR--RV--PSVYIEQDYKWEKE.ENQEQ-DRE--K-ADTETIYINLEQCKKESSRPLWNVDWNEPLQDSL-VT--KWLKEGGI--LSLVWQSLSWL-HL-ILFFQ.NGQ 845
MAC.US.x.251_BK28             A-I-----MY--QK-NS-DV.FG---DLAS-IK--QYG-YV-V-VIL----IYIVQMLAKL----R-VFSSPPSYFQ*THTQQ..D-AL-TREGK--DGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPILQRLSATLRRVREVLRTE.... 820
MAC.US.x.EMBL_3               A-I-----MY--QK-NS-DV.FG---DLAS-IK--QYG-YV-V-VIL----IYIVQMLAKL----R-VFSSPPSYFQXTHTQQ..D-AL-TREGK--DGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPILQRLSATLRRIREVLRTELT.. 823
MND-1.GA.x.MNDGB1             AYDREQR-TY--QK-GDLT-.WAS--DF-W-VQ-L-WGVFL-L-II----LL-LWNTIS-F----R-VFS.....QDCQQNLY...RK--DNG---SNSLELG.EHNSENL...KEESLNRSLIEDLTSFARE*PIWLWLKNLRAAIEYGF...........LAVQEA. 818
MND-2.CM.98.CM16              AYELEQ--IY--EK-GD-T-.WAS--DF-W--K-V-IGLL--IVIIV---LACLW-VLGKF----R--.PYVFKGDYLRPH....NLKQ-GKER--EPDSEKQ.STKSDSSKVEFGKPWSKE...QIRDWLKTSRGYVWLKNLQAVIEYGWQE................. 809
MND-2.GA.x.M14                AYELEQ--MY--QK-GDLT-.WAS--DL-W--K-V-IG-L-IMVVI----LACLW-TIG-F----R--.PYVLKGDYLRPR....NLRQ-DEER-GELNLEEQ.NIKSESSKKEFGRPWRPE...QIRNWLKESTIY-WLKNLQAVIEYGWQE................. 810
MND-2.x.x.5440                AYELEQR-IF--QK-GDLNFHGLTG-DL-W--K-V-IGLLV-VVII---MLACLW-VLGKF----R--.PYVFKGDYLRPH....NLK--DREGG-EPDLEKQ.NIKSESSRQESRKPWKPE...QVRSWLK-ST-Y-WLKNLQAVIEYGWQE................. 833
MNE.US.x.MNE027               A-I-----MY--QK-NS-DV.FG---DLAS-IR--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIRQ..DQAL-TKEGT--DGG-SG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPIFQRFSTTLQRVREVLRTE.... 822
MON.CM.99.L1_99CML1           A-E--G--QH--MKPGQ-DF.-----D-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-WESL..... 806
MON.NG.x.NG1                  A-E--G--AH--MK-GQ-DW.-----D-SK------I--IV-AA-----VLMFI-G-IR--G----L--P-IPTQV.........PREPGAPE-TGE-GGEQG.-NRST-YARGF-TIIWTDCRD-IVWTFQLLRDSGL-IYRS-QRVRSHLIPLLRDLCRQLREAS.... 803
MUS-1.CM.01.CM1239            A-K---E-QHK-QK-LE-D-.--E--D-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-IVIREQARRCSE 826
MUS-1.CM.01.SIVmus_01CM1085   A-R---E-QHK-QK-LE-D-.--E--D-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-RALIEPLRE 832
MUS-2.CM.01.CM1246            A-Q---E-QHK-QK-LE-D-.--E--D-SK----V-I-C-V-A--IA----MFI-GMLR-I--------P-I-TR..............QSAEQ-QQEKTEKG.AGRSGRHR--S-FSE--HN-MIWT-QI-TSSASNLWWLTQRIYHHLTKGLVNP-TL--SFREALGR 821
MUS-2.CM.01.CM2500            A-K---E-QHK-QK-LE-DA.--E--D-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-LRARATSTHLLE 814
MUS-3.GA.09.09GabOI81         AEK---R-NHD-MK-AE-D-.--S--D-SH------I--I--AA------LMWLIG-LR---E---LV-P-NPIHP.......QE-QRQ----KGDV-DGGNV.K-V--L--F--CV-N---D-IVWI-QT--SS-STLVWAIQESARRIWTLLN-L-QT-REQAARLRE 827
MUS-3.GA.11.11GabPts02        ANK---E-NHK-MK-AE-D-.--S--D-SH----V-I--IV-AA---F-LLMV--G-LR-------A--T-IPTQA.......RGA-GQ--ET--G--SNASN.--V-YL--F---F---I---VIW--QT--SS-SNLWWLTK-VATQLVRLA-AVR-TTGNLLSSIWE 828
OLC.CI.97.97CI12              AYQE-QR-QY--QRFQELSD.WTK---LA--MT-V-IGLLV-LAY---N--LGLVRKIRIM----VS-QLP......FSDTL..QDKEQ-DNG-G-DS-FFES..TFNEW-CTTECTGTLRVNWRDLLS-LWKGVVRVTR............................. 792
RCM.CM.00.SIVagi_00CM312      A-D-H---VH--EK-NN-GD.AFS-LRLDW-ME---IG-FVIL-II------LLW-S-SKL-A--T--LSPPPYYHSQIPIRNLKKEEQQAKG--GEFSGEGGGYKSYNWQKGSFY-LILRPIVT--QTLWNNLIS-LLLSYQS-QSILSRIQHPLL-IVA-VQ...... 799
RCM.CM.02.SIVrcm_02CM8081     A-D-H-R-VH--EK-NN-GD.AFS-LRLDW-ME-L-IGVF-IL-VI----F-LLWGCISKI-A--T--LSPPPYYHQQTPIRNREEQPIK--G-GDS--EGG..YKSPSWQRGYFYSLILRPIEI-LRALWNSCNS-LSVIFQS-RSILSQIQQTLR-IIA-IQ...... 806
RCM.GA.x.SIVRCMGAB1           A-D-H---VH--EK-SN-GD.AFS-L-LDW-MQ---IGFF--I-II---VAWLLWNCLSNL----R--..SPPSYVQQIHIHNTGE-QT-GEKR-D---EGG..NKYNNWLREYCW-QLIHPLSRIWTQLSQICRSCSSIIFQS-RWILA................... 792
RCM.NG.x.NG411                A-E-H-R-VHD-EK-NS-GD.MLS-L-MDW--K--RIG-F-IL-II----I-LLW-C-SKL-A--T--LSPPPCYHQQIPIRNREEQPIK--EKGGS--GGG..HKYYNWQRG.......YFYILILRPIEILCRR-YQICSNS-SVIYQSLQSIPSRIL-HLRWIGA.. 799
SAB.SN.x.SAB1                 AHE-EQ--LDSYQK-VS-SD.F-S--DL-K-FGWM-IA--VIA-IIVA-VLLVIIG-LRKF-K--A---SLPSSH*.QIHIHL..DKGQ-DKE--Q-SGGLNSDSRSYTWQREFL.............RHLCH-LITWLRN.LTSWFSTIFSNLH-CLQDIQQRTRQLTA. 809
SMM.CI.79.SIVsmCI2            A-I-----MY--QK-NS-DV.FG---DF-S-VK--YIGFYV-V-II----IVYLIQLLGKL-K--R-VFSSPPSYVQQIHIHK..-QE--TNEGT-EDGG-NG.GYRSWPWQIEYIHFLIRQLIR-LTWLYNNCRD-VLRIFQS-Q-IFQLVS...QAARNRIRPR.... 826
SMM.LR.89.SIVsmLIB1           A-I-----MY--QK-NS-DV.FG---DL-S-VK--QYGV-V-L-VI----MIY-IQMLL-L----R-VFSSPPSYHQQIRTQQ..DQAQ-IKEGI-E-GG-NG.GYRSWPWQIEYIHFLIHQLIR-LTWLYNNCRAWLFRICQN-RQLFQRLSEAIQPVREVIRRE.... 818
SMM.SL.92.SIVsmSL92A          A-I-----MY--QK-NS-DV.FG---DLAS-VK--Q-GVY--I-I-L---IIY-IQMLS-LSK--R-VFSSPPSYHQQIHTHK..VQEL-TKEGIDEEGGNSG.GTRSWPWQIEYIHFLIRQLIH-LTWLYNSCRA-LLRIFQS-QPTYQRLLSLISRIRDILRRA.... 825
SMM.SL.92.SL92B               ARL-----MY--QK-NS-DV.FG---DL-S-VR--QYGVFL-I-I-L----IY-VQMLS-L----R-VFSSPPSYHQQIHIQR..DQEL-AKEGT--EGGNGG.GYRLWPWQIEYIHFLIRQLIRILTWLYNNLTR-ASRAYQN-QQLCQRLSEISQPIRELVRREAG.. 828
STM.US.89.STM_37_16           A-V-----MY--QK-NS-DV.FG---DL-S-VR--QYGVYL-I---M--VAIYIMQLLA-L-K--R-VFSSPPSCRQQIPIHK..-QEQ-TKEGT-E-GG--G.GINSWPWQIEYTHFLIRQLVR-LTWLYNNFRACLSRIYQT-HPTFQRISRILQRIREVVRLG.... 822
SUN.GA.98.L14                 AYVTELE-QNKFKQ-QEFN..F-S-LDLSQ-FL---YAVL-I-IIIAA--LSFIIQQIY-MC---R..VL.SPSAYVEQDWLQETC-KPTDKE---ET-KE-I.YINLEQSKKES-PPPWTVDWDEPLRDS-LVT--KWLKAWGIVLAQNIYHLLSFLWH-LTTSFHHGQ 849
SYK.KE.x.KE51                 A-E-----VHD-TK-QE-D-.--S--DLSK-FQ-L-IGFFAIAAI-I--VLSFAWG--RNMLG-----LQKPQP--.........LYHKA-A-P--GIETDTDN.YKPSVSTF-REFLRQS-AEAQKLWNRVSESCRSLIRGL-IAW-...................... 782
SYK.KE.x.SYK173_COMGNM        A-E---R-VHD-TK-QE-D-.--S--DLSK-FF-L-IGFYVI-A--L--L-SFSVG-IKNLLG--V-ILQNPT...........Q-RKD-GKPAD-EEGSGDR.EGLNVSTF-RESLRQS-EAGQQLWRTVCSSFRSLIRQLTITW-...................... 774
TAL.CM.00.266                 A-E---R-VH--QK-ND-D-.--S---LSA-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-RLHLTAAAA 814
TAL.CM.01.8023                A-E---R-VH--QK-ND-D-.--S---LSA-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. 831
TAN.UG.x.SIVagmTAN1           AYE-EQ--MDTYQK-GD-T-.W--I-DVSS-F-W--WGFY--I--ILF-MAWLIWGCIA------F---P.......QINIRL..-REQ-DNAGG-DKDSSSS.-DKSPPSVKES-LPNRGGIQAEERAW-QHLTNWCL-ISSW--RLYQILRRSLTTLLQLLRQE.... 819
VER.DE.x.AGM3                 ARA-E---LDAYQK-SS-SD.F-S--DFSK--NIL-IGFLD-L-II---LLYT-Y-CIA----------P.......QIHIHP..WKGQ-DNA-GP-EGG-KRKN-SEPWQKESGTAEWKSNWCKRLTNWCSISSIWLYNSCLT--VHLRSAFQ............... 819
VER.KE.x.9063                 ARE-E---LDAYQR-TS-SN.F-S--DFSK--NIL-IGFLV-V-II---LLYTIY-CIG----------P.......QIHIHP..WKG--GNAG-PEEGGE--ND-SDSWQKGSGTRQKRGSWSKRLTNWLSTFSTWLYNTFLT--IHLRGAWQ............... 819
VER.KE.x.AGM155               ARE-E---LDAYQK-SD-S-.F-S--DFSK--NIL-IGFLA-I-VI---LLYTLYTCIA----------P.......QIHIHP..WKGQ-DNAG-PEEGGRTGKSKSTH*QKEFGGRDKRTSWCRQLTNWF-TLSIWLHNSCLT--LKLRSAWQ............... 812
VER.KE.x.TYO1_patent          ARE-E---LDAYQK-TS-SD.F-S--DFSK--NIL-MGFLVIV-II---LLYT-YGCIV------V---P.......QIHIHQ.V-KG--DNAD-P-EGG-NSRIKLESW*KDSKSRCMQ-TAWLTRLN......TWLYNSCLT--IQLRKAFQ............... 805
WRC.CI.97.97CI14              AHD-EQA-MF--QK-SD-TA.WGS-LQWF-IWNWL-WGLLVI-IV-----LACLMNSIRIA-----K-IS-.......QNHLNN.LEEGQQDSE-GIEEDYNSIEINDWQYRRES-IVFLNSLLTS---ICQHLLHIVWSLLSCL-WWINDGW........SLLKDI.... 817
WRC.CI.98.98CI04              AH-TEQ--MF--QK-QD-TT.WGS-LQWF-IWNWL-WGLL-I-IV--I--LGWLMTSIRIA-----K-IS-.......QSHLNN.LEEGL-DRE--AEEDNSTLDV-DWQFKKES-ITLLNSLINF-SAVCQTLLQIVWSLLACLFWWIQHGW........NLLKRI.... 827
WRC.GM.05.Pbt_05GM_X02        AHA-EQS-MF--QK-TD-TQ.WGS-LSWF-IWNWL-WVLLVI-LV--I-VLGCLFTSLRIA-----K-VS-.......QSHILN.LEEGL-DNE-T-EEDFIAIEE-DWQFKKESLILFWNSLATL-Y-ICLSLLQ-SYSLLSYLWI-IQNGW........SLLKRL.... 819
```

```
                                                                                                         Env end
                                                                                                         gp41 end
                                      glycosylation NAT                                       cytoplasmic tail end
                                                                                         *
H1B.FR.83.HXB2              .....WNLLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVV...QGACRAIRHIPRRIRQGLER..ILL....*   856
H1A1.UG.85.U455_U455A       .....----L--GR---I--IT--D-V-V---GWI-----IG...-TIG---LN-----------..A--....-   856
H1C.ET.86.ETH2220           .....GS-V---GL---K--IN---T---V-G-----F--LI...-RIW--FCN---------A..A-Q....-   852
H1D.CD.84.84ZR085           .....--------R------I--VD-----------I-DI-...RR--K-VL---T--------..A--....-   857
H1F1.BE.93.VI850            .....G--TR----------I--F-T---V-------I---L...-R-G--VLN--------A--..A--....-   833
H1G.SE.93.SE6165_G6165      .....----L--GR------I---DTV-----NW-------A...-R-----LN--T--------..A--....-   867
H1H.CF.90.056               .....-------G-------ID---T----------GI-VI-...-R-W---L---------F--..S--....-   846
H1J.SE.93.SE9280_7887       .....V--VW--G-------I----T-----------I--IA...-R-F---L------------..A--....-   849
H1K.CM.96.96CM_MP535        .....---V-----------I----T------G----I--IG...-R-F--LL------------..A--....-   843
H1O.BE.87.ANT70             CRIC.AAVT---L---Q---T---DTL-V---NW--GI-AGI...-RIGTG--N-----------..S--....-   860
H1O.CM.94.BCF06             CRLC.GAVT---L---RA--TN--DTI-V--GNW--SI-LGI...-RIG-GFLN--------A--..A-N....-   879
H1O.FR.92.VAU               CRLF.KAII---L---QT--TN--DTV-V---NW--SI-LGI...-SIG-G-LN-----------..L--....-   874
H1N.CM.02.DJO0131           .....-GIIA--GK---D--I----TI--V-------L--LA...-RIG-G-L------------..A--....-   840
H1N.CM.04.04CM_1131_03      .....-GX-A--GK---D--I---XTX----------X--LA...-RIG-G-L-V----------..T-V....-   853
H1N.CM.95.YBF30             .....-GI-A--GK--RD--I----T---V-------I--LA...-RIG-G-L------------..A-I....-   853
H1P.CM.06.U14788            CLRDCFAACG--T----Q--T----TV--S--NW--Q--A-G...-QIG-GFLN----L-------..S--....-   883
H1P.FR.06.RBF168            CLRDCFAACG--T----Q--T---DTV--S--GW--Q--I-G...-QIG-GFLN--------I--..S--....-   879
CPZ.CD.06.BF1167            ...QFLH--ET-GILIRQHCIVCIDTL-EFTGWW--G---AL...RV-VDI-----T-------I..A-N....-   885
CPZ.CD.90.ANT               .....N--STL-.TIIRTEIIKNIDRL--W-G-K--SILLAL...-TIV-I--EV---------I..A-N....-   859
CPZ.CM.01.SIVcpzCAM13       .....RGII---GR--RT--T---D------G-----I-RA-...-IVF-I-GN-----------..T--....-   865
CPZ.CM.05.LB715             .....KGV----GR---V--I---DT-----------I---A...-RFG-G-LN--T--------..A--....-   910
CPZ.GA.88.SIVcpzGAB2        .....KGIG---L---RT--I---DT---I---R--TI---A...TRIG-G-L------------..A--E...-   857
CPZ.US.85.US_Marilyn        .....-GVIA--AR---V--T---DT-----------I--LT...RRLFLG-I------------..S--....-   837
GOR.CM.04.SIVgorCP684con    .....FAFIA--G----Q--I---DCV-VWT-DW--Q--AIA...-RIG-G-LN-----------..S--....-   859
GOR.CM.07.SIVgor2139_287    .....FAFIA--G----Q--I---DCV-VWT-NW--Q--AIA...-RIG-G-LN-----------..S--....-   855
GOR.CM.07.SIVgorCP2135con   .....FAFIA--G----Q--I---DCV-VWT-DW--Q--AIA...-RIG-G-LN-----------..S--....-   850
GOR.CM.12.SIVgor_BPID1      .....LAVCG--V---QR--TN--DT--V----W--Q--LIG...-RIG-G-LN-----------..S--....-   890
GOR.CM.12.SIVgor_BQID2      .....KATA---L---QR-XTN--DTV-V---NW--SI-LG-...-RFG-G-LN----------L..S--....-   889
GOR.CM.13.SIVgor_BPID15     .....LAVCG--A---QQ--T---DTV-VR--DW--Q--L-G...-RIG-G-LN----L-------..S--....-   893
```

```
                                                                                                         Env end
                                                                                                         gp41 end

MAC.US.x.239                .....LTY---GWSYFHEAVQAVWRSATETL-GAWGDLW-TL...RRGG-W-LA----------L..T--....-   880
H2A.DE.x.BEN                .....AAQ---GCEWIQEAFQAFART-RETL-GAWGWLW-AA...RRIG-G-LAV-------A-L..A--....-   861
H2B.CI.x.EHO                .......Y-R-GISWFQEAIQAAAR-AGETL-SAARTSWG-L...RR-AGE-IA--------A-L..A--....-   853
H2G.CI.92.Abt96             .....AAYFS-GFRW-QEACTAATR-AQETLTSTWRALWKTL...GRVG-G-LA----------L..T--....-   885
H2U.CI.07.07IC_TNP3         CQWF.QEA--AGWKAXRETLA-AWRDLWETLGRVGRGILAIP................-----F-L..A--....-   882
H2U.FR.96.12034             .....IAY---GW-YF-EAFQAFGK-ARETLSRTGRELW-TL...GRVG-WL-A--------F-L..A--....-   868
ASC.UG.10.RT03              RA...VAYIH-GL--F-AAVI-AWD-L--FTWNW-EA-LAYL...RRLAQG-IA--------A-I..L-N....-   887
ASC.UG.10.RT08              RA...VAYIH-GF--F-AAVI-AWD-L--FTWNW-EA-LAAL...RQLAQG-IA--------A-I..L-N....-   891
ASC.UG.10.RT11              RA...IAYI--GL--F-AAVI-GWDTL--FTWNW-EA-LVAL...RRLAQEFIA--------A-I..L-N....-   882
COL.CM.x.CGU1               ..........A-.-T-RYL..LVPRGPSLQIL-TLQ-WLRSA...ARGW-RAPEYL-GWIYDRPQ..GPA....-   838
COL.UG.10.BWC01             ...............-W-L..LVPRCPDLSIP-TLS-WLRSA...TRRW-RASEYL-GWLFDRPE..SPA....-   843
COL.UG.10.BWC07             ..........MPL-A-WRL..LVPRSANL-IPQTLQ-WLRSA...TS-G-RALESCG-WIR-DPQ..GPA....-   835
DEB.CM.04.SIVdeb04CMPF3061  L....-GY---GWK-F-DTI-WAGGELGQTFQRWAEVALQ-L...GRGI-E-LA--A-----A-L..F-N....-   847
DEB.CM.99.CM40              L....GGY---GWK-F-DTITWAGGELWQWLQGIAEMALQGL...VR-GG--LRV-A-----A-L..--N....-   871
DEB.CM.99.CM5               V....-GYC--GWK-F-DTIIWAGQEIWEWLQGIAQVA-QGL...VWGG-NLLA--A-----A-L..L-N....-   851
DEN.CD.x.CD1_CMO580407      .......YF--GL---QTGLRD-GTSAIQQGRATAEV-LAAL...TR-A-EVVA----------I..V-N....-   865
DRL.DE.11.D3                .......-QEAGRGIC.TLLREAAQRLWRGGRQFGLSAARAL...R-LAQGVAN--------A-I..L-N....-   900
DRL.DE.11.D4                .......-QEAGRSIC.ALLQEAAERLWRGGRQLGLSTARAL...RAFAQ-LAN--------A-V..L-N....-   884
DRL.DE.11.D5                .......-QEAGRSIC.TLLQEAAQRLWRGGRQLGLSTTRAL...RAFAQ-LAN--------A-V..L-N....-   892
DRL.DE.11.D6                .......-QEAGRGIC.TLLQEAAQRLWRGGRQFGLSAARAL...R-LAQGVAT--------A-I..L-N....-   897
DRL.x.x.FAO                 .......-QEAGRGIC.TLLQEVAQRIWREGRQLGLSSARAL...RALAQEVAA--------A-V..LFN....-   896
GRV.ET.x.GRI_677_gri_1      .......Y---GL-Q-QEA-TG-AQ-L-R-AR-AWG-LGAI-...RS-Y--VINS---V-----K..V-G....-   855
GSN.CM.99.CN166             V....IAYIS-GI---QAA-TGI-DSL--FTWNW-EA-LHAC...RRVW-EFLA--------A-I..LFN....-   882
GSN.CM.99.CN71              A....AAYIS-GI---QAAVSGI-DSL-VFTWNW-EPLVQT-...GRVW-EFLA--------A-I..L-N....-   887
LST.CD.88.SIVlhoest447      HLRKACSRRVVENAQKIA-WSWNKIRRNRRDISAAHARNLRL...GQKK-WRFRFRG-SGFPS-TTETA-.....-   916
LST.CD.88.SIVlhoest485      HLWKASSGWMVENAQKIARWSWNKIRRNRRDISAAHARDIRL...GQKK-WRFRFRG-SGFPS-TTETA-.....-   915
LST.CD.88.SIVlhoest524      HLWQTLSRFLVE-AQKTA-WIRNKSRRNRGELSK-NGTNIRL...GQNR-RW.RFRF-SGLPS-TTETT-.....-   917
LST.KE.x.lho7               RLWQTSSRWMVENAQKIQ-WLREKCRRNRGQLSS---KNIQL...GKKK-WRLRFGG-SGISS-ATETA-.....-   913
MAC.US.x.251_BK28           .....LTY---GWSYFHEAVQAGWRSATETL-GAWRDLW-TL...RRGG-W-LA----------L..T--....-   881
MAC.US.x.EMBL_3             .......Y---GWSYFHEAVQAGWRSATETL-GAWGDLW-TL...RRGG-W-LA----------L..T--....-   882
MND-1.GA.x.MNDGB1           ..IRSLGS-LVSFAAISWR-GYTGWLQ-NRIFT-CREA-IAA......GTC-W--L-----SA--..P-N....-   879
MND-2.CM.98.CM16            .......-KTAGGKIF-VLQGYAQGLWSRGHQW-LSTAACFR....AIA-G-IN--------A-V..L-N....-   867
MND-2.GA.x.M14              .......-KAAGARICQVLQSFAQRLWSGGHQL-LSVIRGAA....AIG---GN--------A-V..L-N....-   868
MND-2.x.x.5440              .......-KAAGAALYQVLQGFAQRLWSRGYQL-LSG-RGAA....AFG-G-WN--------A-A..L-N....-   891
MNE.US.x.MNE027             .....LTY---GWSYFQEAVQVAWRSATETL-GAWGDLW-TL...GRVG-W-LA-------E--L..T--....-   883
MON.CM.99.L1_99CML1         ..RRLLAYC--GI---QAAVT---D-L-RFTIVW--ALLHAG...GRLW---VA--------A-I..F-N....-   870
MON.NG.x.NG1                ..SRLLAY---GL--XQXACTGAID-L-RFTVIW--A-LRLG...GRLW-GLVAV-------A-I..L-N....-   867
MUS-1.CM.01.CM1239          LLARLLAYI--GI---QAAGTAA-DQL--FTWNW-EPILQIG...RRVW-EFLA--------A-I..L-N....-   892
MUS-1.CM.01.SIVmus_01CM1085 ALQRALAC--HGI---QTASTAV-DQL-TFTWNW-ET-LQAG...RRVW-EFLA----V---A-I..L-N....-   898
MUS-2.CM.01.CM1246          FIAH....A--GI-K-QAAVSGA-D-L--FTWNW-EPILQFS...-KIW-KFLA--------A-I..L-N....-   883
MUS-2.CM.01.CM2500          VLRRLLAY---GL--FQEGVTAAIDGL--WT-NW--A-LQ-S...KQL*K-FLA--K-----A-I..L-N....-   879
MUS-3.GA.09.09GabOI81       VIRRIIAY---GL---QEA-TGF-D-L--FTWNW-EA-LQ-L...RRVA-EFLA--------A-I..LFN....-   893
MUS-3.GA.11.11GabPts02      ACQRAYAY---GF---QAAVSGI--QL--FTWNW-EPLFQAX...RRVW-EFLA--------A-I..L-N....-   894
OLC.CI.97.97CI12            ........WI.QKSLQHVWRTLKQKRQ......HGINLQQKK...-RRGSPLQKAQEQLGREGAE.TTGV....-   844
RCM.CM.00.SIVagi_00CM312    ..........DGW--F-AF--W-GEMA-ENAYYTWRGLQSLA...RQLAGWPATCG----------..L-N....-   855
RCM.CM.02.SIVrcm_02CM8081   ..........DGW----EFSGW-AEMAQQNAYFVWRGLCVTA...-DIA-WPGTVC------F--..L-I....-   862
RCM.GA.x.SIVRCMGAB1         .......KI--GW--F-EFSSWFAEMALQNAYYTWRGLCA-A...RDFAGWPAMVC----------..LCN....-   851
RCM.NG.x.NG411              .......K--DGW--F-GFSSW-AEMARTNAYYTWRGLCA-C...RDPAGWPATLC------F--..F-N....-   858
SAB.SN.x.SAB1               .......H-E-GW----AA-A--TV-VVQ-ATSAS-S-RHAC...RSIV--VIAH---M--E---..WFN....-   868
SMM.CI.79.SIVsmCI2          .....IAY---GW-YFQEACQEAAR-ARETL-SAGRDLW-TL...GRIG-GLLAA---------L..AF-....-   887
SMM.LR.89.SIVsmLIB1         .....FTY---GWSWFEDALQAAWR-ARETL-SAWRDLWQTL...GRIG-GLWA----------L..T--....-   879
SMM.SL.92.SIVsmSL92A        .....VAYIR-GW-YFIE-LQEAAI-ARQTLVSTGRTIWQTL...GHIG-G-AA----------L..T-N....-   886
SMM.SL.92.SL92B             .......YIR-GWNYFIEACQEAWRSAQE-IVGAWGLIW-TL...GRVG-G-AA----------L..M-N....-   887
STM.US.89.STM_37_16         .....AAY---GCIWIQEA-QAAWR-AGETL-SAGRDLW-TL...GRVG-R-GA----------L..T--....-   883
SUN.GA.98.L14               RLWQTLRGWLGSHLIQAT-RIRNACR-SRERVSSSQKARSRTFSLGRKW-PKWNRT-GS-IPS-TTETT-.....-   920
SYK.KE.x.KE51               .......FIS-GVT---EA-I--GREV-HQMVAIWQALLAYA...RRVAENVAAL---L-----I..Y-N....-   841
SYK.KE.x.SYK173_COMGNM      .......FIS-GFN---IA-A--GREIRDW--AIWQAIYAAT...RRVVE-VAAL---L-----I..Y-N....-   833
TAL.CM.00.266               TLQY.......GFS-IQAAWRDGIT-L-RFTWWW-NG-LAAA...RW-GE--AA----T-----Y..CFT....-   873
TAL.CM.01.8023              ......ATI--GFS-IQAAWGDCVT-L-RFTWRW-SG-LAAA...RW-VQ--AA----T-----Y..C-A....-   891
TAN.UG.x.SIVagmTAN1         .....CQYI--GW-QF-EG-ARSFE-L-S-AQSASRTLWNAC...RS-Y---LEH---M--E---..WFN....-   880
VER.DE.x.AGM3               .......YI--GLG---AA-QEAVV-L-RLAQNAGYQIWLAC...RS-Y---INS---V-----G..--N....-   878
VER.KE.x.9063               .......YI--GLG---TA-QEARL-V-RFAQNAGHQIWLAC...RSTY-H-ISS---V-----E..--N....-   878
VER.KE.x.AGM155             .......Y---GLG---AA-QEAV-HL-SFARNAAHQIWLAC...RS-Y---INS---V-----E..V-N....-   871
VER.KE.x.TYO1_patent        .......Y---GLA---TG-QEI-QTL-GVAQNACHQIWLAC...RS-Y-N-VNS---V-----E..--N....-   864
WRC.CI.97.97CI14            ..........SGRAYGAIRKKIAKSRKENTQ--VSATGSRKT...AASS-IRMLAT-LCASFDKCRRSFSERRQ-   879
WRC.CI.98.98CI04            ..........SGAAHACIRKKITKSRKETTEK-ISENSNRET...T-SS-VWKL-T-LCSSLNKSRRR-F....-   885
WRC.GM.05.Pbt_05GM_X02      ..........CGAAYAALQEKIAARRQKRKEK-IYTPGSRET...SRSSSIRGITA-LCTNLNPCRRGQQ....-   877
```

```
                                 myristoylation                                                                                                                                      phosphorylation
                                 Nef start                                                                                                           acidic cluster                poly-P helix
H1B.FR.83.HXB2                  MGGKWSK.......SSVIGWPTVRERMR...............RAEPAADRVGAASRDL....................EKHGAITSSNTAATNAACAWLEAQEEEE.....................VGFPVT.PQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQ  104
H1A1.UG.85.U455_U455A           -------.......K-RVE--E--K---ETPA...............--KG---V-Q--....................D-Y--V-----SS---S---------GD.....................-----R.-------------F---F--------D------  104
H1C.ET.86.ETH2220               ---TM--.......C-PV---AI---I-RA...............A---EG--------....................D-Y--L-----P-N-PD----------E.....................-----R.-------------F---L------------Y-K  105
H1D.CD.84.84ZR085               -------.......--IV---AI---I-KTDPRE...RR....-P-----G---V----....................-R----------T-------V-----D-E.....................-----R.--------------------RK-------VY--  113
H1F1.BE.93.VI850                -------.......--IV---A-G----QTPT...............--EG---V----....................DRR--------RT--PDL----------.....................-----R.----V------------------------Y-K  104
H1G.SE.93.SE6165_G6165          -------.......--IV---E----I-NTPT...............--EG---V-Q--....................DR-----------N-PD---------DSE.....................-----R.---------F-G-F---F--------D---Y-K  105
H1H.CF.90.056                   -------.......-RMG--S-I-----RAEP...............V-EG---V----....................DRR--V-IN---S--RDA-------DG-E.....................-----R.-----------G-F------------D---Y-K  105
H1J.SE.93.SE9280_7887           --N----.......--Q-----RAPA...............------G---V-Q--....................A-------------D-------T---.....................-----K.--I--------G----F--------D---Y-K  101
H1K.CM.96.96CM_MP535            -------.......--IV---AI-----RARPAA...DRVG...TQ----G---V-Q--....................AR---V-----SHN-PD----------.....................-----R.-------------F--GF--------D---Y-K  113
H1O.BE.87.ANT70                 --NALR-.......GKFE--AA------RTR..T.FPESE.......C-PG--QI--E-....................AAR-G-P--H-PQN---L-F--SHQ---.....................-----A.-----------G-F---F------------Y-H  110
H1O.CM.94.BCF06                 --NA---.......RKFA--SE--Q---QSS.....PDPQ.......C-PG---V--E-....................AAR-G-PD-Y-PQN---L-F--SHT---.....................------.--M----I-----F---F--------D---Y-H  108
H1O.FR.92.VAU                   --NVLG-.......-KWA--SE--R---QTS.....PDPQSSDT..QP-PG-R-V--A-....................ADRRG-P--Y-PQN-Q-L-F-GSHQGKG.....................I---R.-----------G-F---F--------D---Y-H  112
H1N.CM.02.DJO0131               --KI---.......--IV---EI------.RQRP...HEPA....VEP-VG-----Q--....................ANR--L-T---RTN-PTV--V------GE.....................-----R.------------------F------------W-R  111
H1N.CM.04.04CM_1131_03          --KIL--.......NR-A--SEI------.RQTQ...EPAQ.......----VG-----Q--....................ANR----T---RDN-ETVT---------E.....................-----R.----I--------F---F-----------VW-R  111
H1N.CM.95.YBF30                 --KI---.......--LV---EI------.RQTQ...EPAV......-.P-VGA----Q--....................ANR----IR--RDN-ESI---------E.....................-----R.---------I---Q-F---F---D-------VW-R  110
H1P.CM.06.U14788                --NA-K-.......--LV---A---KIKQTT.....PTTPDPTTPVTP-PG--EI-KE-....................AQGKG-P-KFSSKN---L-F-D-H-----.....................-----R.------C------F---F--------D--VY-P  114
H1P.FR.06.RBF168                #-NA-K-.......--LV---A-----NQTA.....PETP.......-PG--EI-KE-....................AQGKG-R-KYNSKN---L-F---H-D--.....................-----R.----X--------F---F--------D----..  102
CPZ.CD.06.BF1167                --AVLGT.......RFP-IDNMLN-I-...............NTPAADAPG-Q--....................HNR-GL-TNTIGTEHDVLQHSGDHT--A.....................-----R.----M----E-L-----W-----------YW-P  100
CPZ.CD.90.ANT                   --SA---.......IKWV-ARQAIRKIHE...............TNP-DI-PCGNE-....................ASR--L---TIGTEKDVITYS-DHT--G.....................T---R.----M----E-L-----X----------DX--XXL  102
CPZ.CM.01.SIVcpzCAM13           --AQ---.......--LV---EI-K-IK...............E-PI--EG--EV-K--....................--Y-----RH-PS--QTL---DEMKHH-EE.....................-----R.-----------G-F-------RG-------VY--  106
CPZ.CM.05.LB715                 --A----.......--LV---E----I-...............QTQ---EG--PV----....................DN-----VR--PQ--QTL----E-K---.....................-----R.---------F-G-F--------R-------Y-R  104
CPZ.GA.88.SIVcpzGAB2            --N----.......--IV---Q----L-...............-TQE--EG--EV-Q--....................ARR-----RH-PQ--QTL----EM-Q-D.....................-----R.---------------F--------------VY-K  104
CPZ.US.85.US_Marilyn            --N----.......--IV---E--N-L-QT...............QTTA--EG--PV-Q--....................AE-----TR--PQN-QTL---DEMTNH-SE.....................-----R.-----------Q-F--GF-------------VY-R  108
GOR.CM.04.SIVgorCP684con        #-NA-K-.......--FV---A------QTS.....PEPDP.......-PG--EV-IA-....................AQR-G-PAKY-PKN-Q-L-F---H-N---.....................-----V.---------F-G-F---F------------VF--  106
GOR.CM.07.SIVgor2139_287        --NT-K-.......--F----A------QTN.....PEPKP.......-PG--EV--A-....................AQR-G-PAKY-PQN-Q-L-F---H-D---.....................-----R.---------F-G-F---F--N---------VF--  107
GOR.CM.07.SIVgorCP2135con       --NT-K-.......--F----A------QTS.....PEPKP.......-PG--EV--V-....................AQR-G-PAKY-PRN-Q-L-F---H-D---.....................-----X.---------F-G-F---F------------VF--  107
GOR.CM.12.SIVgor_BPID1          --NA-K-.......--LV---A--D--HQTT.....P##EA.......-PG--EV--E-....................A-GKG-P-KYN-RN---L-F-D-HR---.....................-----R.------C--F-G-F---F--N-----D---Y-P  105
GOR.CM.12.SIVgor_BQID2          --NA-K-.......--FV---A--D---QTS.....PDPE......C-PG--IV--E-....................AQK-G-P--Y-PQN---L-F-DSHT---.....................------.---------F-G-F---F--------D---Y-P  108
GOR.CM.13.SIVgor_BPID15         --NA-K-.......--LV---A--D-IHQTT.....PEIDP.......-PG--KI--E-....................A-EK--P-KYNSRN-E-L-F---H-D--.....................-----K.---------F-G-F---F--N-----D--VY-P  107
                                 Nef start            R17Y mutation                                                                     premature stop in Mac239

MAC.US.x.239                    ---AI-M.......RRSRPSGDL-Q-LLRARGET..YGRLLGEVEDGYSQSP-GLDKG-.SSLSCEG........QKYNQGQYMNTPWR-P-EEREKL-YRKQNMDDIDE*DDD...............L--VS-R.-K----T-S--L-I-M---I--------IYY-A  135
H2A.DE.x.BEN                    --ASG--.......KLSKHSRGL---LLRARGDG..YGKQRDASGGEYSQFQEESG-EQ.NSPSCE........GQQYQQGEYMNSPWR-P-TERQKDLYRQQNMDDVDSDDDD...............LI-V---.-R--R-E----L-I-M---I------Q-MFY-R  136
H2B.CI.x.EHO                    --SAG--.......KQSKQQ-GL---LLRARRGP..RGESSGE-Q-RSLQYP-GSDKG-.NSPSCDG........................QKTLGAEGGGKQDSDEDDEDNE...............--VR-R.-G-------F-L---M-------E---IFY-E  118
H2G.CI.92.Abt96                 --SAG--.......KRAPQQLGL-K-LLQARGEP..YGKLWEGLE-GYSQYREESGKG-.SSLSCEG........QQYTQGQFMNTPWR-P-TER-KL-YRXQNMDDVDSDDND...............L--V--Y.-K----V-S--L-I-M---I--------IYY-E  136
H2U.CI.07.07IC_TNP3              --ASG--.......KRSKQAGRL---LLQARGET..YGRLWDGLE-GYSQSR-GSG---.NLPSYEG........QAYSEGQFMNTPWR-P-GEGEKLKYRQQNMDDVDSDDDD...............L-KVS-R.-R--------WP--M-----------IYYN-  136
H2U.FR.96.12034                 ---SI--.......KHSKRSGKL---LLAARGEN..YGRLWGGLEDGFELYQ-GSGK--.SSRFCEAQ......GQGYSEGEFMNTPWRTP-TGKDKLQYKQQNMDDVDEEDND...............L--VA-R.-R-A--T----L---M---I------D-IFY--  138
ASC.UG.10.RT03                  --S-N--.......QQSSQPGMPS.PYSPGTGQK.LYWRIFGDLPRGLLPSR-ESG-EPRYY................SIR-ETAETAKRGEVVPVRYNEGDSTRDDD-EG...............-----C.------Q--F-T---F-W---------FY-P  125
ASC.UG.10.RT08                  --S-N--.......QQSSQPG-PS.PYSPGTGQK.LCWRLCGSLPRELLPSQ-GSG-EPRYS................SIK-EAAQ...PGEVVPLRYNEGDYTRTDD-EE...............-----R.----I-T--F-T---F-W---------FY-P  122
ASC.UG.10.RT11                  --S-N--.......QQSSQAG-PL.PYSPGIGQR.LYWWLCGDLPRSLLPSR-ESG-GPRYS................SIE-EAA-AK.RGEVVPMRYNEGDNTRDDD-EE...............-----R.----I-NL-F-I---F-W---------YY-P  124
COL.CM.x.CGU1                   --SLL-...CLNIRQQPQTRGG-LG-LFGIYWYPGVQAFRFLRHFRGGY-LRRV-GVERLNICEDG.........YMTVPKVLPNCPCPLCYRHPDVYYHNPEP-DV-D................VEAGCYVMSRL-V-QA-I-LL--I-C---------I-RTP  138
COL.UG.10.BWC01                 --SIL-...CLRG.QRPSTPGGLCS-LFGTYWYRGVQTYRFLRHFRGGY-LRRV-GVEHLNILEDG.........YLTVPRALPNCPCPLCYRHPSVYYHSPEP-DADEEE..................-GCFVQAR--V-QG-I-LL--I-C---------I-RTP  137
COL.UG.10.BWC07                 --SIL-RGSNRYQMERPTSPGGCLS-LFGVYWYPGVQTWRFLRHFRGGY-LRRVQGVEHLNLVADG.........YEEIPRDLPNCPCPLCYRHPRVYFHSPEP-DLGE...................NNS.CFVCAR--V-RP-I-LLF-I-F---------I-RTP  140
DEB.CM.04.SIVdeb04CMPF3061      ---RN--.......TQSS-QEENWGKHFKGGQR..WRYRYLGEES-KFLPYQPESDKE-NCS..............LTE-TFIGNPR..REIKCTQRQQ..QDFDQQEE...............-----K.-R--I-DP---LMI-Y----------DIFY-A  121
DEB.CM.99.CM40                  ----N--.......IPSR-QVGSFGSGSRGLLR..WRYRDLSEQE-QFSECLLESD-EQSSS..............STE..FWGSPR..REIKCTNKQQ..QDLDQTEA...............-----R.---SI-DP---LMI-Y----------DVFY-A  119
DEB.CM.99.CM5                   ----N--.......TPSF-QGKKSGSGSKGLLR..WRYKDLSGE--TFSPFLQESDKEQSCY..............STE-PLY-DPH..REIKCTQR-Q..ECYD--HG...............I--A-R.-R--I-EP---LMI-Y--Y-------DIFY-A  121
DEN.CD.x.CD1_CMO580407          --S-N--.......QAS.-IWGLAPSNKEGRQRK.WYWRLLPEQQ-KWWPSPDGSD-A-KLSSTRLLEHDEDEETFVDAE-IIQSSPA..LDPQHPRIHQYVVDHDD-EE...............-----R.-RR--HAP---DMI-M---------QDIFY-P  138
DRL.DE.11.D3                    ---QN--.......KREEAYARFYG-L-RGYGAEGGNLDYQQLEPSE-LL-..EWRTFPEGLDKAQKYSSTSSLDTTPIQ.AAA-AEKDV-LP-EEEQPSE-EE--...............-----Y.--C---EP---EL--F---------K-IWW-L  135
DRL.DE.11.D4                    ---QN--.......RREEAYARFYR-L-RGYGAEGGNLDYQQLEPSEPLLK..HWQTFPEGLDKAQKSSLTSSLDTTLMQ.AAA-AEKDV-LP-EEEQPSE-EE--...............-----Y.-RC---EP---DL--F---------K-IWW-H  134
DRL.DE.11.D5                    ---QN--.......KQEEAYARFYR-L-RGYGAEGGNLDYQQLEPSEPLLK..HWQTYPEGLDKAQKSSLTSSLDTTPMQ.AAA-AEKDV-LP-EEEQPSE-EE---...............-----W.--C---DP---DLI-F---------K-IWW-L  136
DRL.DE.11.D6                    ---QN--.......KREEAYARFYR-L-RGYGAEGGNLDYQQLEPSE-LL-..EWRPFPEGLDKAQKYSSTSSLDTTPIQAAAA-AEKDV-LP-EEEQPSE-EE----...............-----C.--R---EP---EL--F---------K-IWW-L  138
DRL.x.x.FAO                     ---QN--.......KREEAYARFYK-L-RGYGAKGGNLDYHQLEPSEPLL-..KWQPFPGGLDKAQRSSSTSSLDTTHMQ..AA-AEKDV-LP-EEEQPSE-EE--...............-----C.--R---EP---DL--F---------K-IWW-L  134
GRV.ET.x.GRI_677_gri_1          --SSN--.......RQQQ-LLKLWRGL-GKPGAD..WVLLSDPLIGQSSTVQEECGKA-.....KKSWGK..........G-MTPDGRRLQEGDTFDEWDDDEE-...............-----Q.-R---Q----L---F-----S----D-IYY-E  120
GSN.CM.99.CN166                 --S-N--.......QQQQESS-AL.LSSRGTGQR.PYFTLVDEYG-NFWLSPD-SDKGRRYY................LTE-PK............PKRGS...L--Y-PS...............C--R.-R----DP---VM------------MFYCE  108
GSN.CM.99.CN71                  --S-N--.......QQSQESS-AL.LSSHGTGQS.PSFRLLDEYG-NSWLSPD-SD-GRRYY................LTE-S.............PSRQN...CIDY-PS...............C--R.-R----DP---LM-----------AMFYCE  107
LST.CD.88.SIVlhoest447          --NIFG-.......PAAD--WKTLR-L-AGAGTR..............SEGTEETYQQ-MQET................SD...WDKRRDG....................DSDSEEE...............-----R.--R--CSP---TLI-----I---------YW--  97
LST.CD.88.SIVlhoest485          --NIFG-.......PAAD--WKTLR-L-AGAGTR..............SEGTEETYQQ-MRET................SD...WDKRKDG....................DSDSEGE...............-----R.--R--CSP---TLI-----I---------YW--  97
LST.CD.88.SIVlhoest524          --NIFG-.......P-ADS-WR-LRKL-AGSGTR..............-EGTEESYQK-MEQT................LG...WDKTEE.....................DGDSDSE...............-----R.--R--CSS---TLI-----I---------YW--  96
LST.KE.x.lho7                   --NAFGR.......PAAD--WKTLR-F-AGSGRN..............-EGTEDSY-A-IEKT................SN...WAKRRDG....................DSDSEEE...............-----K.--R--CSP---TLI-----I---------YW--  97
MAC.US.x.251_BK28               ---AI-M.......RRSKPAGDL-QKLLRARGET..YGRLLGEVEDGSSQSL-GLGKG-.SSRSCEG........QKYNQGQYMNTPWR-P-EEKEKL-YRKQNMDDIDEEDDD...............L--VS-R.-K----A----L-I-M---I--------IYY-A  136
MAC.US.x.EMBL_3                 ---AI-M.......RRSKPAGDL-QKLLRARGET..YGRLLGEVEDGSSQSL-GLGK--.SSRSCKG........QKYNQGQYMNTPWR-P-EEKEKL-YRKQNMDDIDEEDDD...............L--VS-R.-K----A----L-I-M---I--------IYY-A  136
MND-1.GA.x.MNDGB1               --SSQ--.......KRSEA-VRYSSAL-QLVG...............GPVTPDGYKQIESSQGAEK............QSLLRGR-YGTYSEGLDKVQNDP-TKD-KLDLTQQDPEEEEE...............-----C.R--S--VPS--DLI-F---I------G-IYY-R  125
MND-2.CM.98.CM16                ----S--.......QQEEKYLKYYKA--RGYG...............-EGTNGDYQQ-HASE...............PLLG..-LST-QEEFDREQK.SSS..........TDEEEE...............T---Y.--C-V-EP---DL--MP-----------IW--K  106
MND-2.GA.x.M14                  ----S--.......QREQ-YAKYYKAL-RGYG...............-EGTNLDYQL-EEQR...............LL-G..PLGT-LG-SDKELK.SCSTKVD.......DEQKEE...............R----Y.-TQ-V-EA---NLI-M-----------IWF-R  110
MND-2.x.x.5440                  ----S--.......QREQHYIKYYKAL-RGYG...............-EGTNLDYQV-EEQQ...............PLAG..-SGT-LA-SDKVLR.PYSTERDLRHQQDVTLPSEKEQPSDEEEE.......-----Y.-TQ-VPEA---DLI-M-----------IWW-K  126
MNE.US.x.MNE027                 ---AT--.......KRSKSPGDL-Q-LLRARGET..YGRLWEGLEDGYSQSL-GSDKS-.SLLSCEG........QKYNQGQFMNTPWK-P-GEREKL-YRKQNMDDIDEEDDG...............L--V--R.-R----A----L---M---I--------IYY-E  136
MON.CM.99.L1_99CML1             --S-N--.......QQSPLSS-PL.LGSQSSGRM.RYFMLEDDYG-QSWLSPD-SD-ERKYS................LTEGRN-KQ.........RRQP....LDDDDDD.D...............G--C--R.-R----DP-W-LMM----Y-------GEMFYCE  114
MON.NG.x.NG1                    --S-S--.......QPAQEPLMPS.HGSQSSGQT.RSFAWEDDYG-DSWLSPD-SD-GRRFS................LTEGRNHRS.........NRRPTTV.VDDADEDQD...............LP-C--R.-M----DP-W-IMM------------DK-FYCV  118
MUS-1.CM.01.CM1239              --S-N--.......QQEQQPL-SS.PSSPGTGQS.PYFRLVDEYG-NSWLSPD-SD-GRKYS................LTEGS.............NERPL...I--H-RS...............C--R.-R----DP---MMI----Y---------MFYCE  107
MUS-1.CM.01.SIVmus_01CM1085     --S-S--.......QPAQQSLISS.PPSPGTGRK.QYFKLVEEYG-NSWLSPD-SG-GRRYS................LTEGA.............SKRPV...I--H-RT...............C--R.-R--I-DP---LM--M-------------MFV-T  107
MUS-2.CM.01.CM1246              --S-N--.......PQSQERS-PL.-SSHGTGPN.RYFSLVKKYG-NSWLSPD-SA-GRKYS................LTEN-.............GRRQI...I--H-CS...............C--R.-K--I-DVN--LMI----Y--K-R----MFVCI  107
MUS-2.CM.01.CM2500              --S-N--.......KESQQPLMGS.PSGPQTGQT.PYFKLVNNYKKHSWLSPN-SA-GRKYS................LTKGS.............NKK-I...IKKHKRS...............C--R.--H-IQNP---LIIN-------------MFY-E  107
MUS-3.GA.09.09GabOI81            --S-N--.......KQQQDSSMPL.LSSPGTGQR.PCFKFCDEWP-NSWLSPD-SG-ERKFS................LTEG-.............DRRPV...ID--DRS...............CH--.-KCG--DP-F-LMI------------DRMFYCT  107
MUS-3.GA.11.11GabPts02          --S-S--.......QQSQESS-SL.LSSHGTGQN.RYFRLCDEYG-NSWLSPD-SA-ERRYS................LTEG-.............DRRPV...I--C-RS...............C--R.-RCGIQAV-F-RLI-------------MFYCE  107
OLC.CI.97.97CI12                --SIC--......................R-NREGEV........LYR.RLRSNW-EKEQKQ................LES....EDQHQ-NLAIESLMTGDTDTDDLE...............-----K.----V-QA---TLIE------------MYW-P  95
RCM.CM.00.SIVagi_00CM312        ----S--.......H-QS-LVRW-QK-LTTPGEG..YSRWHDNLQDGQPRVAEGSGKAS............................RDFSIRGGFTIETQQSVDAMDWYEETDD...............TL-----K.------A-SF-L-I-M------------IYY-I  119
RCM.CM.02.SIVrcm_02CM8081       ----S--.......---A-LQKW-N--LTSSGEA..YALLHKTLQDGP-RCAEGSG-AS............................RDFLSRGGITIETQQSVDAIDWYEETDD...............TL-----K.---------F-L-I-------------IYY-T  119
RCM.GA.x.SIVRCMGAB1             ----S--.......N-AA-LLRW-FK-LTTPGEG..YVRWHETLLDGQPWCAEGSG-AS............................RDFVIRGGIT-ETQ-SIDDIDWYEDTDD...............TL-----K.---------S--L-I-M-----------IYW-I  119
RCM.NG.x.NG411                  ----S--.......A-AA-LRRW-GL-LTTPGED..YARFAETLQDGQPRCAEGSG-AS............................RDFLTRGGFTIETQQSVDAIDWYEDTDD...............TL-----R.-----I---S-L-I-------------IYF-E  119
SAB.SN.x.SAB1                   ----S--.......QQQRHSLWLWSKL-QAPVIQ..YDMLADPLLGQSSHIQEECAKS-....................RDGLIRQGD-SRTEEGVKMKHQGR-PSWYD..EDEEE...............-----R.-CL---A----L-I-FG-----------IYY-E  124
SMM.CI.79.SIVsmCI2              ---NI--.......KRARKPHGP---LLQARGET..YGRLWEGLE-GYLQPR-ELG--W.SSLSCEG........QTYSEGQFMNTPWK-PVKERGKLKYGRQCMNNVGEDDKD...............L--VA-R.-KI---E----L-I-M---I-#-------IYY-E  135
SMM.LR.89.SIVsmLIB1             ---AG--.......MRSKRHGGL---LLRARGET..YGRLWEGLE-GYGQSL--SGKG-.NLPCCEP........QRYNQGDFVNTPWR-P-REKEK..YRQQNMDDVDEEDDE...............LI-VA-M.NRI---A-S--L-I-M---I---------IYY-H  134
SMM.SL.92.SIVsmSL92A            -ADSGTH.......WKRDRSDPKED...QAGIGT..YTQLSRNT-QG...................................NT..*IPHGETQLEKEKRENMDDIDEED...............--IA-R.-KI---A----L-I-M---I---------IYY-E  99
SMM.SL.92.SL92B                 ---IT-S.......KPARKHGDL-KKLLSARGDS..YGRLWDGLE-GSQQSQ-GSD-A-NSCLTRDG........YKYNVGEFMNTPWR-P-TERQKQQYRQQCMDDIDEDDDE...............LA-VV-W.-K----A-S--L-I-M---I-------D-IYY-E  137
STM.US.89.STM_37_16             --ASG--.......KQRKQHGEL---LLRARGET..YGKLLEGLG-GSGPSQ--SDKG-.NSHSCEP........QRYNEGQFMNTPWK-P--ES-KLEYRQQNMDDVDEEDDN...............L--VA-H.-R----E----L-I-----I-S-------IYY-E  136
SUN.GA.98.L14                   --NAFGR.......P-EV--VRTLF-L-AGSGTR..............-EPA-REYHR-RRQE................VEPLVSAENGGP.................NGIEQEE...............-E---R.--R--CKP---QLI-----I---------WY-R  100
SYK.KE.x.KE51                   --SQS--.......KQPSR-DEKW.-T-WWPFGK.PYSPMPDELLRMSQPYHEDFD-GWRST...............LTEPILD..PKRDFD.SGGKKWNAGDICHD-GDED...............L-----L.--------L---L-I-----I-R----Q-MFFC-  125
SYK.KE.x.SYK173_COMGNM          --STS--.......-QQLRSEGKY.AIGWRLFGK.QYTPLPDELSR-LQPCR-GFDKAWRST...............LTEPIDPHGPDRDWGHSGGQKFSPGDIVQD-GDTG...............L-----C.--T---TL---L-I-----I-N----Q-MNYCE  128
TAL.CM.00.266                   --SVK--.......LLGETASQRL.QGSHGGGRM.VYWRLRDGQVKRSQLSP--RGK--NIA...............LLD-EWQ-TGADYMGKALTQGEVWYEK-Q--NDME...............L-----R.-K----TC---LGI-F-F---------K-IFYNK  128
TAL.CM.01.8023                  --SVK--.......LLGETASQRL.QGSPGAGRL.VYWRLRDGQCRQLQRSR-ERGKG-NIA...............LLDAGWQ-SDANWMGKALSEGEVWYEK-Q--NDME...............L-----R.--------CN--LGI-F-F---------K-IFYNN  128
TAN.UG.x.SIVagmTAN1             ---SN--.......REQQ-LLRLWRAL-KAPVVR..YGMLADPLIGQSSNIQEECDKNW....................NGGSTRR..GKSTPEGRKL-ADDTWDDW-PEEEE...............-----R.-R----Q----L---F-------------IYW-P  121
VER.DE.x.AGM3                   --LGN--.......PQHKKQLSLWHALHKTRATR..YGLLADPLIGQSSTLQEECDKA-....................K-SLIRKRNGKMTPEGRKLQEGDKWD-WSDEEDE...............-----R.-R----Q----L---F-----------D-IYY-D  123
VER.KE.x.9063                   --LGN--.......PQHKKHVSLWHALHKTQVTK..YGLLADPLIGTSSAVQEEYDKG-....................RKSLIRKRNG-MTPDGRRLQEGDEWD-WSDDEDE...............-----K.-R----Q----L---F-----------D-IYY-E  123
VER.KE.x.AGM155                 --LGS--.......PQHKKQL-IWRALHATRHTR..YGLLADPLIGQSSTLQEECDKG-....................RKSLIRKRNG-MTPEGRRLQDGDQWD-WSDEEDE...............-----R.-R----QI---L---F-----------D-IYY-D  123
VER.KE.x.TYO1_patent            --SQN--.......PAHKKYSKLWQALHKTHVTR..YGLLADPLIGTSSTVQEECDKA-....................RKSLIRKQNG-MTEEGRRLQEGDTW--WSDDEEE...............-----R.-R----Q----L---F-----------D-IYY-D  123
WRC.CI.97.97CI14                ---VF--.......T--EEHMERYA-RLQRAGRR........I-RQRYQQLAAERQQQA...................VESGCLLPD-AQALINAEEASVK-DSSD-DEG...............------.----I-QP-F-LL-----LF--------DIFW-A  115
WRC.CI.98.98CI04                ---IF--.......E--EQRMHAYA-RLQRAGRK........Q-RKRYQKIATERQQVA...................VESGNLLPD-AQALIRAEEDSSK-DSSDD-E...............-----H.----I-QA-F-LL-----LF--------YIFW-P  114
WRC.GM.05.Pbt_05GM_X02          ---VF--.......D--EQRMLRYR-RLLRVGRK........E-RRRYTRLAAERQAEA...................AASGELLPD-AQ-LIRAEEDSSK-DSSDDDA...............I---R.------QA-F-IL-----LF--------DIFW-P  114
```

PLV Proteins

```
                                  HXB2 premature Nef end                                                                                        normal Nef end
H1B.FR.83.HXB2                    RRQDILDLWIYHTQGYFPD*QNYTPGPG.......VRYPLTFGWCYKLVPVEPDKIEEA.NKGENTSLLHPVSLHGMDDP..EREVLEWRFDSRLAFHHVARELHPE.YFK..NC...................................* 205
H1A1.UG.85.U455_U455A             K--E-----V-----F---W-------......I--------------D-AEV---.TG---N-----ICQ--V--E..-K---M-K---T--LK-R-Y------.FY-..D.................................- 206
H1C.ET.86.ETH2220                 K--E-----V-N---F---W-------......---------F-----D-SEV--I.-E---NC----A-----E-E..D----K-K---H--RR-M--------.-Y-..D-................................- 208
H1D.CD.84.84ZR085                 K--------V-----F---W-------......I---------FE----D-EVV-K-.TE--DNC----ICQ---E-E..-K---V---N-----E-K-K-KY-..---..-................................- 216
H1F1.BE.93.VI850                  K-G-T----V---------W-------......I-----L---F-----D-EEV-K-.-E---NC----M-Q---E-E..D----R-K---S--LR-I---R---.FYQD*D-..............................- 208
H1G.SE.93.SE6165_G6165            K--E-----V-N-------W-------......T-F-------F----MD-AEV---.------N-----ICQ---E-E..D----V-----S--RR-I--------.-Y-..D-................................- 208
H1H.CF.90.056                     Q--------V-N-------W-------......E-F-------F-----N-QEV-Q-.-E---N-----M-----E-D..G----M-K------LT-L--VK---.-Y-..D-................................- 207
H1J.SE.93.SE9280_7887             K--E-----VHN-------W-------......I--------------D-SEV---.-E---NC----ICQ--IE-E..-----Q-K---S--RR-I-------.FY-..D-................................- 204
H1K.CM.96.96CM_MP535              K--E-----V-----F---W-------......I--------------D-AEV--T.TE--DNC----INQ---E-E..H--I-M-K---S--RR--------D.-Y-..D-................................- 216
H1O.BE.87.ANT70                   K-AE-----V-N---F---W-------......T-F------LF-----SEEEA-RLG-TC-RAN----ACA--FE-T..HK-I-M-K--RS-GNT---MIT---LFQ-..D.................................- 214
H1O.CM.94.BCF06                   K-AE-----VH----F---W-------......P--------LF-----SAEEA-RLG-SS-DA-----ACNY-DE-E..HK-I-K-Q--RS-GLT---LQK---LFP-...................................- 211
H1O.FR.92.VAU                     K-AE-----VH----F---W-------......T-F------LF-----SEAEA-ALG--C-RA-----ACN--YE-Q..HK-I-K-Q--RS--NT---LIT---LFS-..D.................................- 216
H1N.CM.02.DJO0131                 K---V----V-----F---W-------......T-F-------F----LSAEAV---.-E-D-NA----LCQ--V--G..HK---Q-Q---S--RR-I------D.FY-...-................................- 214
H1N.CM.04.04CM_1131_03            K--E-----V-----F---W-------......----------F----LSAEAV---.-E-D-NA----ICQ--V--E..HK---V-----S--RR-L-------D.FY-...-................................- 214
H1N.CM.95.YBF30                   K--------M-----IL--WH------......I---V-----F----LSAEEV---.-E-D-NA----ICQ--A--D..HK---V-----S--RR---------.FY-...-................................- 213
H1P.CM.06.U14788                  E-AE-----------F---W-------......E--------LF-----SEVEA--MGD-Q-KAK----ACTY-FS--..HK-I-V-K---S-GRE---LQK---.L-I..KD................................- 218
H1P.FR.06.RBF168                  ..........................................................................................................................................  102
CPZ.CD.06.BF1167                  K-AA---T-MFN---I---W-------......------CR--LF----AD-PE......ED-RNL-M--ACSY-R----..NG-L-V-K---E--RR-I---R---.---...............................- 196
CPZ.CD.90.ANT                     K-AA---M-MFN---I---W----EE-......------CR---F------SPP......DD-RNI----ACT--DG--..HK-I-R-E--AS-MRR-I---R----.--R..D.................................- 199
CPZ.CM.01.SIVcpzCAM13             K--E-----V-----F---W-------......E---PC----F----LTTEQV-A-.-E-D-NC----ICQ---E-E..SK---I---------LR-I---K---.-Y-..D.................................- 208
CPZ.CM.05.LB715                   ---E-----V-----F---W-------......----------F-----P-EEV-R-.-E---NV----MCQ---E-E..D----Q-----H--LR-I---Q---.-Y-..D-................................- 207
CPZ.GA.88.SIVcpzGAB2              ---E-----V---------WN------......--------Y---F-----S-EAV---.-----DNL----LCT--FE-E..-K---I-K---Q--LR-L---K---.-YR..D.................................- 206
CPZ.US.85.US_Marilyn              ---E-----V-----I---W-------......--------Y---F----LTEEEV-Q-.----TNI----MCQ---E-E..HG---I-Q--TE--RR-R-K-----.-R..-................................- 210
GOR.CM.04.SIVgorCP684con          A-S------V---------W-------......--------Y--LF-----SEAEA--MGTAS-RAI----ACS--R----..HG-T-V----RA-GTT---LQK---LFQR..D.................................- 210
GOR.CM.07.SIVgor2139_287          A-S------V---------W-------......I--------LF-----SEAEA--MGTSN-KAM----ACS--R----..YG-I-V----RA-GTT---LQK---LFKR..D.................................- 211
GOR.CM.07.SIVgorCP2135con         A-SE-----V---------W-------......I--------LF-----SEAEA--MGTAS-RAI----ACS--R----..HG-I-V----RA-GTT---LQK---LFKR..E.................................- 211
GOR.CM.12.SIVgor_BPID1            E-AE-----V-----F---W-------......---------LF-----SEAEA--MG--N-RAK----ACAY-FS--..HK-I-M-K--RS-GIT---LQK---LFS-..D.................................- 209
GOR.CM.12.SIVgor_BQID2            Q-AE-----VX----F---W-------......X-----Y--LF-----SEAEA--MG--N-RAK----ACN--YG-E..HGQI-K-Q--RS-GNN---LA----YFKN...................................- 211
GOR.CM.13.SIVgor_BPID15           E-AE-----V-----F---W-------......-------W------MF-----SEAEA-YMG--D-RAK----ACTY-FS--..HK-I-V-KY-RS-GTQ---LMK---LFI-..D.................................- 211
                                            |---------- max HIV-1 similarity ----------|                                                            normal Nef end

MAC.US.x.239                      --HR---IYLEKEE-II--W-D--S---------......I---K----LW------..NVSD--QEDE.EHY-M--AQTSQW--..PWG---A-K--PT--YTYE-YVRY--EFGSKSGLSEEEVRRRLT..ARGLLNMADKKETR..........- 263
H2A.DE.x.BEN                      --HR---IYLEKEE-II--W----H---------......----MY---LW---S-..ELSQ--EEDE.ANC-V--AQTSRH--..EHG-T-V-Q---M--YNYK-FT-Y--EFGHKSGLPEKEWKAKLK..ARGIPYSE................- 258
H2B.CI.x.EHO                      --HK---TYLENEE-IVSGW----H---------......----KF---LW----I..NM-A-P.EDE-THC-V--AQTSAW--..PHE-T-V-Q---L--YDY--FSRF--EFGYQSGMPEKEWKAKLR..ARGIPTE.................- 239
H2G.CI.92.Abt96                   --HR---IYMEKE--II--W----A---------......I---X----LW------..TVSD--QEDE.THC-V--AQTSPW--..P-T---A-K--PT--YDYR-FI----EFRWKSGLPEAVWKEKLK..QRGLPIE.................- 257
H2U.CI.07.07IC_TNP3                --HR---MYLENEE-IVA-W----E---------......--F-K----LW------..HVSD--S-DE.THC-V--AQTSQEE-..RWG---A-K--AT--YDYK-FI-Y--EFGSKSGLPEKEVKRRLA..ARGLLNMADKKETS..........- 264
H2U.FR.96.12034                   --HR---IYLENEE-II--W----S---------......T-W-MC---LW--E--..-ANDK-QEDE.RLY-VGSAQTSCEE-..HWG-A-V-K---S--YSYQ-FIKC--EFGSKSGLSEEEVKRRLT..ARGLPVKNC...............- 261
ASC.UG.10.RT03                    E-HKK---YA-NVW-LV-GW-G--Q---------......T-F-TC--ILWE-----..EIGPQGLGE-DERAS-LHAGQQVYQ-..PHC-T-V-H-NPT---EPGILKADKLGQKPLPGQKTFLTGRD........................- 238
ASC.UG.10.RT08                    E-HQK---YA--VW-LT-GW-G--Q---------......P-F-TC--ILWE-----..EVGPQGLGE-DERAK-LHAGQQSFQ-..PYG---A-H-NPT---EPGILKADKLGQQPLPGQETFLTGRD........................- 235
ASC.UG.10.RT11                    E-HKK---YA--VW-VV-GW-G------------......P-F-TC--ILWE-----..-VGPQGLGE-DERA--LHAGQQAYQ-..LHG-T-V-H-NPT---EPGILKADKLGQQPLPGMKSCITGRD........................- 237
COL.CM.x.CGU1                     E-D-LIEQYA-IEW-CLKGWLE-EDEL-EDGALKED-K--VA--LW---YI..EQLG-YAYSYDLSL-SVTS.......................................RKKKKPQQVAIEMVD..................- 224
COL.UG.10.BWC01                   E-D-LIE-YA-IEW-VLKGWLE-EDETDENMQIREE-K--VA--LW---FI..EQLGQWAYSYDMSL-SINN.......................................RKKK.PQQQQVAVDTVAIETPD...........- 229
COL.UG.10.BWC07                   E-DRL-ERYC-IEW-VLS-WLE-EDETNEDGTPKEE-L--AC--LW---LL..-S-GKWAYCFDPAL-NR--.......................................TKKKKQDAPNRSTVAMEIETVD..........- 233
DEB.CM.04.SIVdeb04CMPF3061        --HA--E-HAQNEW-II-GWLQ--E---------......I---KY--FLF------..EIADPDYEND-RNI---DAHQGQ-E-..PYK-R-V-K---Q--YCYK-.GHEAHTKETHTRRCMFPK.RK..........................- 232
DEB.CM.99.CM40                    --HA--E-HAQNEW-II-GWLQ--K---------......----IY--FLF------..EIADPDYEND-RNI---DAHQGM-E-..PH--R-V-K---T--YCYK-.GHAEQR..EHTRRCMFPK.RK.........................- 228
DEB.CM.99.CM5                     --HA--EIHAH-EW-II-GWLK--E---------......P---TY--FLF------..EIADPDYEND-RNI---DAHQGQAE-..PHK-R-V-K---S--YCYK-.GHEAHKP.EHNRRCMFPK.RK.........................- 231
DEN.CD.x.CD1_CMO580407            --HA----YAQNEH-FITGW-T--D---------......I----E--FLW------..TIE-.EYDNK-ENC---DRYEGQQA-..PW--T-V----PE--WCYK-.GHKKLQTSQHN-YCQRAKPRK......................- 249
DRL.DE.11.D3                      --ES----YAQNEW-FIKGW-S--K---------......I---KV--FLF-----DV-DDLI...DQPCNR--NSSQMGPI--..S-G-R-M-Q--AS--YTFQ-CIH---EFGHVTS...IRHDKREEPDCCKRKWWQX................- 256
DRL.DE.11.D4                      K-ET----YAQNEW-FIKGW-S--R---------......I---KV--FLF------V-EDVV...NQPCN---NSSQMGPI--..PNG-R-M-Q--PT--YSFK-CIH---EFGHARS...IEHIA.EEPACCKRKWWQX................- 254
DRL.DE.11.D5                      K-EI----IYAQNEW-FIKGW-S--K---------......----MR--FLF-----QV-EDLV...NQPCN---NSSQMGPI--..P-G-R-M-Q--PS--YSFQ-CIH---EFGHVKS...IKPI..EEPACCKRKWWQ................- 255
DRL.DE.11.D6                      --ES----YAQNEW-FIKGW-S--K---------......I---KV--FLF-----DV-DDLV...DQPCNR--NSSQMGPI--..P---R-V-Q--AS--YTFQ-CIH---EFGHVTS...IRHDKREEPDCCKRKWWQX................- 259
DRL.x.x.FAO                       --ET----YAQNEW-FIKGW-S--K---------......I---KM--FLF------V-EDLV...NQQCNR--NSSQMGPI--..P-G-R-M-Q--PG--YTFQ-CIH---EFGHVKS...IKS..VEEPGCCKRKWWQFKPTTEGCHGDNLQKC- 269
GRV.ET.x.GRI_677_gri_1            --EK--N-YALNEW-IID-W-A-S---------......I---RV--F-F-------..-LH---R-CE.RHC-M--AQMGEDP-GIDHG---V-K--PK--VEYRP..DMFKDMH.......EHAKR.........................- 224
GSN.CM.99.CN166                   D-HQK-EQYA-LEW-LV-GWLSF-----------......T---TIP-F-IC-R--..ATT-DSEPGDDEYL-T--AYQGRSE-..QHK-F-VFS-C-K--IKSGI.Q-DQLQQEERKMR..LTANRFL.......................- 219
GSN.CM.99.CN71                    D-HQK-ESYC-YEW-IV-GWLQW-----------......I---TMP-F-WC-R--..AMT-DSEPGDDQYL-N--AYQGQQE-..HH--I-VFS-C----LKSGW.QMNQLQQEERKKR..LTANRFL.......................- 218
LST.CD.88.SIVlhoest447            ------C-YCENEW-LIG-FMT--D---------......T-F--A---LWQ-E--AC-EYTDP..SDYRQC---SSQ-GVQE-..PWG-R-I-H-NPM--VDY--LRKQ-K...........SIQASAIGLQRK..................- 206
LST.CD.88.SIVlhoest485            ------C-YCENEW-LIG-FMT--D---------......T--------LWQ-E--AC-EYTDP..SDASQC---SSQ-GVQE-..PWG-R-I-H-SPM--VDF--LRKQ-K...........SIQASSLSLNCKRK................- 208
LST.CD.88.SIVlhoest524            --N---T-YCENEW-LIG-FM---D---------......T----A---LWQ-E--AC-EFKDP..SD-RQC---SSQ-GVLE-..PWG-R-I-H-NPM--VDY--LRKQ-A...........SVQATAMRFNCERK................- 207
LST.KE.x.lho7                     ------I-YCENEW-LIG-FM---D---------......T--------LWQ-E--AC-EYKDP..SD-TQC---SSQ-GVLE-..PWG-R-I-H-NPM--VDFI-LKKQ-A...........KIQNTAFAFDCKRK................- 208
MAC.US.x.251_BK28                 --HR---MYLEKEE-II--W-D--S---------......I---K----LW------..NVSD--QEDE.RHY-MQ-AQTSKW--..PWG---A-K--PT--YTYE-YARY--E#GSKSGLSEEEVRRRLT..ARGLLNMADKRETR..........- 263
MAC.US.x.EMBL_3                   --HR---MYL...E-II--W-D--S---------......I---K----LW------..NVSD--QEDE.RHY-MQ-AQTSKW--..PW#---T-K--PT--YTYE-YVRY--EFGSKSGLSEEEVRRRLT..ARGLLNMADKRETR..........- 260
MND-1.GA.x.MNDGB1                 --EE----YAENEW-FE-GW-Q--T---------......T---K---FLF--E--SRAIGD-...YAA-NH---SSQ-CPQE-..P-G-T-M-SGTLI-PMTLQH.............................................- 215
MND-2.CM.98.CM16                  --EE----YAQNEW-FI-TW-S--D---------......I---K--RFLF--C--AVPPDQ-...NN-CNK--QSSQ-GIQEE#.PWG-R-V-K---G--YTFY-PIKR-GEFKHVQSISHIAYAKEHKPECCKRKWWQF...............- 231
MND-2.GA.x.M14                    --EE--N-YAQNEW-FI--W-E--S---------......I---KR--FLF--I--QVPPDQ-...NQ-CNR--NSSQ-GIQEE#.PWG-R-M-K--PE--YTFY-PIKC-KE-RHVTSLSYTAYQ..EKSDCCKRKWWQF...............- 233
MND-2.x.x.5440                    --EE----YAQNEW-LI-NW----K---------......I---K---FLF-----AVSPPL-...ED-CNR--NSSQMGIQE-..P-G-R-M-K---G--YTFY-PIIR--E-KCVTSLSYEAYKKEEKPDCCKRKWWQF...............- 251
MNE.US.x.MNE027                   --HK---IYLEKEE-IV--W-D--S---------......P---K----LW------..NVSD--QEGE.ENY----AQTSQW--..PWG---V-K--PT--YTYE-YIRY--EFGSKSGLSEEEVRRRLT..ARGLLKMADKKETS..........- 264
MON.CM.99.L1_99CML1               D-HRKIEQYA-LEW-LI-GWLQ--E---------......----TMP-FLWC-R--..ATT-DSEEGD-DFL-T--AYQGR-E-..PH-QF-VFS-C-K--VKSGR.Q-AQLQQEERKKR..LAANRIL.......................- 225
MON.NG.x.NG1                      D-HQK-EQYA-VMW-LV-GWLQF-X---------......----TIP-FLWC-R--..AIQ-DSEDGDDEFL-T--AYQGRDE-..PHKQF-VFS-C-K-GVKSGP.Q-DELQQEERKRR..LTANRIL.......................- 229
MUS-1.CM.01.CM1239                E-HQK-ETYT-VDW-IV-GWLQF-E---------......T---TLP-FLWC-RE-..AIT-DSEEGD-EYL-T-AAYQGKEE-..PHKQF-VFS-C----MKSGR.Q-DQIQQEERKRR..LTANRIL.......................- 218
MUS-1.CM.01.SIVmus_01CM1085       E-LQK-ET-A-VEW-II-GWLM--D---------......T---KIP-FLWC-RE-..AMT-GSEEGDDQYL-D--AFQGREE-..HHKQF-VFS-C----LKSGL.Q-DQMQQEERKKR..LATNHIL.......................- 218
MUS-2.CM.01.CM1246                D-LQK-ETYA-VKSKIV-RSLI--D---------......T---RQP-FLWC-R--..AMT--SEPGDDQYL-T--AYVGRDE-..HHK-F-VFS-C----LKSGP.Q-NQIQQKERKRR..LTANRIL.......................- 218
MUS-2.CM.01.CM2500                E-HQK-ETYA-VEW-LI-GWLM--E---------......------MP-VLFC-R--..HIT-NSELGDDEHL-C--AFIGREE-..PHK-F-VFS-C----LKSGR.Q-NQMQQEERKRR..LAANRILSQSP...................- 222
MUS-3.GA.09.09GabOI81             D-HQK-ETYA-MEW-IV-GWLRF-----------......T---TMP-FLWC-R--..AMQ-DSEAGDDEHL-T--AFQGRDE-..IHK-F-VFS-C-K--IKSGK.Q-DQLQQEERKKR..LTANRFL.......................- 218
MUS-3.GA.11.11GabPts02            E-HQK-ETYAFVEW-IV-GWLS------------......----RIP-FLWC-R--..EIA-DPGEEDDTFL-A--AYQGREE-..HHK-F-VFS-C----LKSGQ.Q-DQLQQAERRKR..LTANRIL.......................- 218
OLC.CI.97.97CI12                  --ENLIY-YALNCW-IIGNF-ATEEVEEGT.....L-R-K-P--LWD-RR---LPETLE.....ERG.............................................................................- 152
RCM.CM.00.SIVagi_00CM312          --HR---IYLENEH-II--W--------------......P---TF---LWM-----..-VSD--KEDE.EH-----AESS-IE-..PWG-T-A-K-NPM--VDYIGYR---DFFG......ERKNKTQ.........................- 225
RCM.CM.02.SIVrcm_02CM8081         --HR---IYLENEH-II--W----K---------......I---MM---LW------..-V-D--KEDE.EH-----AETC--E-..PWG---A-K--PM--IDY-GYK----DFG......EKKNKTQ.........................- 225
RCM.GA.x.SIVRCMGAB1               ---R---MYLENEH-II--W--------------......I---TL---LWQ-----..-VSD--REDE.EH-----AETS--E-..PWG---A-K-NPM--VDYIGYR----FFG......ERKNKTQ.........................- 225
RCM.NG.x.NG411                    --RR----YMENEH-II--W--------------......I---TM---LWQ-----..-VSD--TENE.RHC----AETS-----..PWG---A-K-KPL--VDYAGYR----FFG......ERKNKTK.........................- 225
SAB.SN.x.SAB1                     --KK----YALNEW-IVDGW----D---------......T---KC----F------..-LS---K-SE.-HC----AQVAYE--..AWK-T-V-K--PL--VDY--WR----QVP........SAQG..........................- 227
SMM.CI.79.SIVsmCI2                --HK---MYLENEE-II--W----S---------......I---MM---LW------..NVSD--QEDE.RHC-V--AQISHL--..PWG---A-K--PQ--YNYE-FIRY--EFGSKSGLSEDEVKKRLT..ARGLLKMADKKENS..........- 263
SMM.LR.89.SIVsmLIB1               --HR---IYLEKEE-IV--W----S---------......I---MF---LW----M..-VSD--QEDE.THC-V--AQT-Q---..PWG---A-S--PK--YTYE-YVRY--EFGSKSGLSEDEVKRRLT..ARGLLKMADKKETR..........- 262
SMM.SL.92.SIVsmSL92A              --HK---SYL-NEE-II--W----R---------......I---MF---LWQ-----..-VSD--NEDE.-HC--#-AQTYK-S-..PWG-T-A-K-NPM--YTYE-FVRY--EFGWKSGLTEEEVKRRLA..SRKKPNHKKQHGAEEGS.......- 229
SMM.SL.92.SL92B                   --HR---IYLEKEE-II--W----S---------......I---KF---LWQ-----..-VSD--N-DE.THC-V--AQTYQHS-..PWGK--A-K-NPH--YTYE-FVR---EFGWKSGLTEEEVERRLANKPKPQKKMADKKETS..........- 267
STM.US.89.STM_37_16               --HR---MYLEKEE-IV--W----A---------......I---KQ---LW------..-MSN--QEDDGTHY-V--AQT-QW--..PWG---V-K--PL--HTYE-FVR---EFGSKSGLPKEEVERRLT..ARGLLKMADKKETS..........- 265
SUN.GA.98.L14                     T-EE----YAENEW-FITGW-D--K---------......----KA---LW--A--TI-EDRDP..NHPCQA---SSQQGVNE-..PWG-R-I-T--PT--YDFR-IQK-------------EFKH.VTSLQWEV.....................- 209
SYK.KE.x.KE51                     K-HE--Q-YLKNEH-VID.DIT--S---------......T----I---LWE-A-N..EIEGYLVDEEDTLMMHPAAGVGASE-..PH--N-M-N-NPH--YTPGWEMARQQLERQTGKR.......................................- 229
SYK.KE.x.SYK173_COMGNM            K-DE--H-YLQNEH-IID.RI---S---------......T----I---LWE---N..EIEGCLEYEEHTLL-HPASGQGSSSMGEPHV-......LQPPPGYTPGWEMARLQLERQTGKPQELQSALSKNIS.....................- 241
TAL.CM.00.266                     --HA--N-YAHNEW-IL--W----E---------......T----C--ILW--C--..EIH-DDT..EDGHL-P--AYDGQAE-..PWG-S-V-V--EK--YTPG-.KMAEWDRLEREKRMLLAPPQTASS......................- 241
TAL.CM.01.8023                    --HA--N-YAHNEW-IL--W-S--E---------......T----C--ILW--C--..EIH-DDT..EDGHL-M--AYDGQQE-..PWG-A-V-V--EK--YTPG-.KMAEYDRLERKKKELLAPPQTASS......................- 241
TAN.UG.x.SIVagmTAN1               K-EQ--N-YALNEW-IID-W-A-S---------......T-K-RC--F-FE-----..-VSQ--QD.E.RHC----AQIEWES-..PWK-T-V-K-NPL--VQYNP..DSFKDMH.......GLVKRK........................- 223
VER.DE.x.AGM3                     --NQ--N-YALNEW-IID-WNAWSE---------......I---RC--F-F------..-LH---ETCE.RHC-V--AQVREDP-GINHG---V-K--PM--VQYDPNRKYLTDMH.......DLGKRK........................- 230
VER.KE.x.9063                     --NR--N-YALNEW-IID-WNAWST---------......----RC--F-F------..EMH---ETCE.RHY-A--AQIKEDP-GISHG-T-V-K--PM--VQYDPNRQYF-DMH.......AIVKRK........................- 230
VER.KE.x.AGM155                   --NK--N-YALNEW-IID-WNAWSK---------......I---RC--F-F------..ALH---ETCE.RHC-V--AQ--EDP-GINHG-I-A-K--PM--VQYDPSREYFTDLYS......TVGTGN........................- 231
VER.KE.x.TYO1_patent              --NK--N-YALNEW-IID-WNAWSK---------......I-F-KC--F-F------..-LH---QTCE.RHC-V--AQMGEDP-GISHG-I-V-K--PM--IQYDPNREYFTDMH.......GLVKRK........................- 230
WRC.CI.97.97CI14                  --EE-VN-YAHNEW-LIDGW-A--E---------......I-W-K-----W-----HFSEESDNTDPYYERNK-LHPACAEPE---..WG-H-V-K--PK--CDF--GR-PGQSIATGKEAIESLYKNK.KQCK.....................- 234
WRC.CI.98.98CI04                  --EE-VN-YAFNEW-LIEGW-A--E---------......I-W-K-----W------FSEESNDSSQYYEKNK-LHPGCQESE---..WG-H-V-K--P---CDF--SR-PGHIIAQGEDAIETLHKEKLKRHK......................- 234
WRC.GM.05.Pbt_05GM_X02            --EE-VN-YA--EW-IVEGW-A--D---------......----K-----W-----HFSEESPNSDQ-FEKN--LHPAVAESE---..WK-Y-V-K--P---CDF--GR-ADGQIATG...IETLYEMKKRNNK.......................- 231
```