VII

PLV Proteins

Contents

VII-1 Introduction **405**
VII-2 Sequences **406**
VII-3 Alignments **413**
VII-3.1 Gag 413
VII-3.2 Pol 417
VII-3.3 Vif 424
VII-3.4 Vpx 426
VII-3.5 Vpr 427
VII-3.6 Tat 428
VII-3.7 Rev 430
VII-3.8 Vpu 432
VII-3.9 Env 433
VII-3.10 Nef 440

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)
H1O.CM.91.MVP5180	L20571	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gurtler, L.G.	*J Virol* **68**(3):1581-1585 (1994)
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)
CPZ.CD.06.BF1167	JQ866001	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Li, Y.	*J Virol* **86**(19):10776-10791 (2012)
CPZ.TZ.00.TAN1	AF447763	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Santiago, M.L.	*J Virol* **77**(3):2233-2242 (2003)
GOR.CM.07.SIVgorCP2135con	FJ424863	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	*J Virol* **83**(4):1635-1648 (2009)
MAC.US.x.239	M33262	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kestler, H.	*Science* **248**(4959):1109-1112 (1990)
H2A.SN.86.ST_JSP4_27	M31113	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kumar, P.	*J Virol* **64**(2):890-901 (1990)
H2B.CI.x.EHO	U27200	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Rey-Cuille, M.A.	*Virology* **202**(1):471-476 (1994)
H2G.CI.92.Abt96	AF208027	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Brennan, C.A.	*ARHR* **13**(5):401-404 (1997)

Name	Accession	Proteins	Author	Reference
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)
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)

Name	Accession	Proteins	Author	Reference
DRL.DE.11.D3	KM378563	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_1A11	M76764	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Marthas, M.L.	*J Med Primatol* **18**(3-4):311-9 (1989)
MND-1.GA.x.MNDGB1	M27470	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Tsujimoto, H.	*Nature* **341**(6242):539-541 (1989)
MND-2.CM.98.CM16	AF367411	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Takehisa, J.	*ARHR* **17**(12):1143-1154 (2001)

Name	Accession	Proteins	Author	Reference
MND-2.GA.x.M14	AF328295	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Souquiere, S.	*J Virol* **75**(15):7086-7096 (2001)
MND-2.x.x.5440	AY159322	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	*J Virol* **77**(8):4867-4880 (2003)
MNE.US.x.MNE027	U79412	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kimata, J.T.	*J Virol* **72**(1):245-256 (1998)
MON.CM.99.L1_99CML1	AY340701	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)
MUS-1.CM.01.CM1239	EF070330	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
MUS-1.CM.01.SIVmus_01CM1085	AY340700	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)
MUS-2.CM.01.CM1246	EF070329	Gag, Pol, Vif, Vpr, Tat, Rev, 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, Env, Nef	Liegeois, F.	*Viruses* **6**(7):2880-2898 (2014)
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

Name	Accession	Proteins	Author	Reference
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.SIVsmSL92B	JX860433	Gag, Pol, Vif, Vpx, Vpr	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)
SMM.US.04.G078	JX860415	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	*J Virol* **86**(24):13217-13231 (2012)
SMM.US.05.D215	JX860413	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	*J Virol* **86**(24):13217-13231 (2012)

Name	Accession	Proteins	Author	Reference
SMM.US.06.FTq	JX860414	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	*J Virol* **86**(24):13217-13231 (2012)
SMM.US.86.CFU212	JX860407	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Fischer, W.	*J Virol* **86**(24):13217-13231 (2012)
SMM.US.x.H9	M80194	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	*J Virol* **66**(1):414-419 (1992)
SMM.US.x.PGM53	AF077017	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Novembre, F.J.	*J Virol* **72**(11):8841-8851 (1998)
SMM.x.x.pE660.CG7G	JX648292	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Lopker	Unpublished
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)
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)

Name	Accession	Proteins	Author	Reference
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)
MON.NG.x.NG1	AJ549283	Pol, Vif, Vpr, Vpu, Env, Nef	Barlow, K.L.	*J Virol* **77**(12):6879-6888 (2003)

```
                                membrane binding                                                                                        phosphorylation site
                                Gag start, p17 start                                                                                                   p17 end p24 start
H1B.FR.83.HXB2                  MGARASV.LSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQ............................NKSKKKAQQAAADTG.HSNQVSQNYPIVQNIQG 140
H1O.CM.91.MVP5180               -------.-T-SK--A--R------S--A-R---L---------Y-C-------A--TE-L-Q--E-A-K----D-K--W-AI-V-W---N-FD-R--QQ-IQ-LK-VMA...............................SRKSAE--KEETSPR-T-------T-A-- 138
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-AAA.TDSNI-R---L--TA-- 142
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-IT-WK--MQ...............................-RKKTS.....EG-TGT---------A-- 133
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-VKIVK-KLKITSSE....TTGNEETAKGKKETAVTS....SGQTEKIYP-LPS-G-TGR-----FDAAQ 161
CPZ.TZ.00.TAN1                  -------.-R-DK--T--S---KSR-R---LI--L---GS--Q---M----M-NV---WK-IL-----VDI--P-II--F--ICV-----AGERVQ--E--VKIVKMKL.....TVQKN.NSTATSSGQRQNAGEKEETVPPSGNTGNTGRATETPSG-RL--VITDA-- 163
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......................................KRKSDSRPET-----V-A-A-- 131
                                Gag p15 start                                                                                                          p15 end p27 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.Q-VQRHLV.............VETGTTET............MPKTSRPTAP-SGRGG---VQ-IG.- 142
H2A.SN.86.ST_JSP4_27            ----N--.-R-KKA-EL-------------R-------AN--D--GLAES---SK---QK--TV-D-LVP----N-K--F---CVIW-I-AEEKA---E--K.QKVQRHLV.............AETKTTEK............MPSTSRPTAPPSGNGG-F-VQ-VA.- 142
H2B.CI.x.EHO                    ----G--.---KKT-EL--V--------R-M-------VN-----GLAESR-GSK----K-RKV-G-LVP----N-K------CVIF-L-AEEKV---E--K.--AQRHLA.............AD...TEK............MP-MSKPSKPTSRL..A--VQ-IA.- 137
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............VP-TSRPIAPPSGRGG---VQ-VG.- 142
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............MP-TSRPTAPPSGRRG---VQ-VG.- 142
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.Q-VQRHLA.............VETETAEK............MP-TSRPTXPPSGGRG---VQ-IG.- 141
ASC.UG.10.RT03                  ----H-AM-T-KK--KY-------K-----QI-------K-----GLSDS---NK---QK--EVIV-LEE----S-K--FGL-SV-W---ADVTVE--E--KKQ-RIRCHM.............SGKKEETSA....................GGGASG------R-A-- 137
ASC.UG.10.RT08                  ----H-AM-T-KK--KY-------K-----QI-------K-----GLSDS---SK---QK--EVII-LEE----S-K--FGL-SV-W---ADVTVE--E--KKQ-RIRCHI.............SDKKEETSA....................GGGASG----V-R-A-- 137
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--KKQ-RIRCHM.............SGKKEETSA....................GGGASG------R-A-- 137
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--EK-V--AYKK...QAMIE.MASKEEEKAK......................KEAEK.LDM.---TGP.Q 139
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--ES-V--AFKKQ...........................AMIEQAEEQERAKAKEEK.LDM.---NTP.Q 139
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.........................AMIEQAERTEQQAAQEELEKRQEM.---T.G.P 141
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--KKAVKLRIAQVEAEDQLEMQVLQKKAEAEAAKKQGAVSAQAQPA-VQSAAPQQGTV-RG----LRQG.- 169
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...VEQEEELEMALIQRQKEKEVEQK.QQ-QQE-RQQEPQAAPQTTP-V----LRQG.Q 165
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...AEQEEELEMALIQKKAEKELEKK..Q-TQ...---ADTQQKTNQPA-F--LRQG.Q 161
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...KKKNIVIGNADQGDDQPPQGAAG.GGSSAE-G-S-GGSSFRDPN-------RTA-- 166
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--VQ-VKQHCHL.............VD..K.NENAA.........SKNENGETAT-SGR-R----QVVN.Q 143
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--VQ-VKQHCHL.............VD..K.NENAA.........SKNENGETAT-SGR-R----QVVN.Q 143
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...........KKKNE-TAPPGGE-R---V-NQN.N 142
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-KQQVRIRCHL.............EETKG.KEKQ..........NKDPPGAAGGQAV------VIR-A-- 146
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-KKQVQIRCHL.............VGTKE.KETQ..........NKDPPGAAGGQAV--------R-A-- 146
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--MQ--KI-PTQ.........................................EERQPK-G---L-REN.Q 127
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--MQ--KI-PTQ.........................................EDRQ-K-G---LIREN.Q 127
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--IN-VKI--NK...........................................PVPE-K-F-LIREN.Q 125
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--IS-VKI-PAP...EAAGK.K.................................Q-TGG---LIREN.Q 130
MAC.US.x.251_1A11               ----N--.---KKA-EL-------N-----M---V---AN--D--GLAES---NK---QK--SV-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.Q-VQRHLV.............VETGTAEI............MPKTSRPTAP-SGRGG---VQ-IG.- 142
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--K-LKVITRK...EEKQ...EDE...................................K-F-VQRDAA- 129
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--KE-VKQKLHL...VAEK...ENAASE.KEQR..................AIVTPSGR-K----QIIN.Q 143
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--KK-V-QCFHL...AAKG...ESAASE.KEEK..................ATATPSGR-K----QIIN.Q 143
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--KE-VKQRYHL...VVER...ENAASE.EEKGA..................TATPAVR-K----QVIN.Q 143
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.Q-VQRHLV.............VETGTAET............MPKTSRPTAP-SGKRG---VQ-IG.- 142
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......................A---AAPPTGGVP-G---V-RTQG- 144
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-RKQ-RIRCHL.............ADKQGE.E...................KTAQ--T-A----IR-P-- 137
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--KKQ-RARCHL.............EEKKGEKE...................SQAQAA.-A---VIR-A-- 137
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--KKQ-RIRCHL.............AEKKEDSKG.................-S-AKATTSG---VIRTA-- 141
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--KKQ-RMRCHL.............GEKKEQQEQ...............R-QPSGAGAS-----VIRTA-- 142
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--KRQVRIRCHL................VDKKAG.......QEXGGATGGA.TGGV-TT------IRTTN- 146
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--KKQ-RIRCHL................GEKKEE....EAKGAE-PSPKGAEAPRPGAS-Y---VIRAAN- 150
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--KKRVKVKE....................KKEE.........................QLAM--IR-A-- 124
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....KEK................GATMT-SGQRG-----TIN.Q 142
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....KEK................GATAT-SGQKG----ITVN.Q 142
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--VA-VK-CCHL...VEKA...ENTTE....KEK.................GATAPPSGQRG---ITIN.Q 142
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................GATVT-SGQRG-----TIN.Q 142
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--KA-VK--VPA...EMTE...SATATSSGQTK.........ELQAKKKNEPTVTPSGG-R-----SVN.N 153
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.Q-VQRHLV.............VETGTADK............MP-TSRPTAPPGGKGG---VQ-VG.- 142
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.Q-VQRHLV.............VETGTADT............VPKTSRPTAPPSDRRG---VQ-MG.- 142
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.Q-VQKHLA.............VENETADQ............LP-TSRPTAPPSGKRG---VQ-VG.- 142
SMM.SL.92.SIVsmSL92B            ----G--.---KKA-EL--V-------R--M----I--A---D--GLAES---SK---QR--TV-A-LMP----N-K--F---CVVW-L-AEMKV---E--K.-TVQSHLV.............VESGTAEK............LP-QSRPTAPPS..GG---VQ-VG.N 140
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.-TVQSHLV.............VESGTAEK............LP-QSRPTAPPSGG..---VQ-VG.N 140
SMM.US.04.G078                  ----S--.---KKA-EL---C---N---R-M---V---AN--D--GLAES--DNK---QK--SV-A-LVP----N-K------CVIW-I-AEEKV-H-E--R.Q-VQRHLV.............VETGTADK............MPTTSRPTAPPSGKGG---VQ-VG.- 142
SMM.US.05.D215                  ----C--.---KKA-EL--V----------M---V---AN--D--GLAES---NK---QR-ISV-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.Q-VQRHLV.............VETETAGT............MPKTSRPTAPPSGMRG---VQ-VG.- 142
SMM.US.06.FTq                   ----S--.---KKT-EL--V----N-----M---V---AN--D--GLAES---NK---QK--SV-A-LVP----N-K------CVIW-I-AEQKV-H-E--K.Q-VQRHLV.............VETGTADT............MP-TSRPTAPPSGRGG---VQ-IG.- 142
SMM.US.86.CFU212                ----S--.---KKA-EL--V----------M---V---AN--D--GLAES---XK---QK--SV-A-LTP----N-K------CVIW-I-AEEKV-H-E--K.Q-VQRHLV.............VETGTADP............MPKSSRPTAPPSGMKG---VQ-VG.- 142
SMM.US.x.H9                     --V-N--.---KKA-EL-------------M-------AN--D--GLAES---NK---QK--SV-A-LVP----N-K------RV-W-I-AEEKV-H-E--K.Q-VQRHLV.............VETGTADK............MP-TSRPTAPPSGRGG---VQ-VG.- 142
SMM.US.x.PGM53                  ----N--.---KKA-EL-------N-----M---V---AN--D--GLAES----K---QK-ISV-A-LVPA---N-K------CVIW-I-AEEKV-H-E--K.Q-VQRHLV.............VETGTADK............MPITSRPTAPPSGRGG---VQ-IG.- 142
SMM.x.x.pE660.CG7G              ----N--.---KKA-EL-------N-----M-------AN--D--GLAES--DNK---QK--SV-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.Q-VQRHLV.............VETGTADK............MP-TSRPTAPPSGRGG---VQ-IG.- 142
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............MP-TSRPTAPPSGRGG---VQ-VG.- 142
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---K-VRL-PAK...NEEAQ.A.....................................---F-VQREG.Q 127
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-KQ-VK-ACNW...EDEET.VTSSGQKENSNDTA....T-SGREGKMQLPAAMP-SGG-G---LIR-P-N 162
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-KQ-VK-ACNW...KDDPP.ATSGGQSENSSQNM.ASET-SGQKVVQQEKQKAATPPPRG---LLR-P-N 165
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-KAEVRKRCHL...GEKKE.TTEEEK.....................VRKATVT-SGQTG----IR-A-Q 145
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-KAEVKKRCHP.............VEKKEESKG........TTE-G-VRSAIVT-SGH-G----IR-A-Q 149
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.............VEKEKTAA......APSGGQ--NYNTAATP-GR.HG---V--QN.N 148
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.............VEKERNAE...RNTTETSSG-KKNDKGVTVPPGG---F-AQ-QG.N 152
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.............VEKEKRAT......EPSSGEKKSNRE-TAPPGG.---F-AQ-QG.N 148
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.............VDKEKTAV......TPPGGQ-KNNTGGTATPGG.---F-AQ-QG.N 148
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.............VEKEKSAT......ETSSGQKKNDKGIAAPPGG.---F-AQ-QG.N 148
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--KE-VNV--KK...P...........................................---M-VLRTA.Q 123
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--KE-V-VKDTK...QKEKP.......................................---M-VLRTA.Q 127
WRC.GM.05.Pbt_05GM_X02          --SGN-IS-APEICESLQRV---KDS--T--R--L--LVK---KLCLDGK-ISSQ--VKKLCEHIA-LVPE--KTI--IWGMISV-A-LYNN-QVS--QA-IK---VRPRE...EKA.........................................-V-M-VLRGP.Q 125
```

```
                                                                                                                         CyPA binding                                                 major homology region
H1B.FR.83.HXB2              QMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVH.AGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTN.....NPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTL  304
H1O.CM.91.MVP5180           -----------------A------N--I----M------V-Y-I-----AI----G-L-V---V----------T--PA.M--LP---I---T----------Q----I-T-R....GANS----D--RK--V-----M-K----V------------------------  303
H1N.CM.95.YBF30             -----PLT----------I------------M--------S------------------V------D---T---P.V--LP---L-D-------------A--VA---A.....---V---D--R---V----R-------V---E-K------------------  306
H1P.CM.06.U14788            --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-----V-K--C-V-----K-----------------  297
CPZ.CD.06.BF1167            -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------------L-K----V----------------------I  328
CPZ.TZ.00.TAN1              VAR--P----------R-I---G-N----------------Y---S---A--E----------V----------A--A-.---QQA-ML---T-A-------------VL---T..PQAQGGV---D------------L-----V-----K----------------I  330
GOR.CM.07.SIVgorCP2135con   -----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-----V-K--C-V-----K------------------  296

MAC.US.x.239                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-C----N--N---VK-------QS-------S-  305
H2A.SN.86.ST_JSP4_27        NYT-VPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--D------IIR-I------D--AQ--IP..--LPA--L--------------VE---Q--FR....PQN-V---S--R---QI--Q-C----N--N----K-------QS-------S-  306
H2B.CI.x.EHO                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-C----N--N----K-------QS-------S-  301
H2G.CI.92.Abt96             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-C----N--N---VK-------QS-------S-  306
H2U.CI.07.07IC_TNP3         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-C----N--N---VK-----S-QS-------S-  306
H2U.FR.96.12034             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-C----N--N----K-------QS-------S-  305
ASC.UG.10.RT03              -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--QRC-K--N-VN----------S-K--------C-  302
ASC.UG.10.RT08              -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--QRC-K--N-VN----------S-K--------C-  302
ASC.UG.10.RT11              -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--QRC-K--N-VN----------S-K--------C-  302
COL.CM.x.CGU1               GP-.HQPLSPRTLGAWVKCV-GGIA-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-VIQIAR-S-V------S--D-KS------SA-  296
COL.UG.10.BWC01             GPA.HQPLSPRTL-AWVKCV-GGIA-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-VIQIAR---V------S--D-KS------TA-  296
COL.UG.10.BWC07             -GPMHSPLSPRTL-AWVKCV-QGITASLA---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-VIQIAR-S-V------A--D-KS-----FSA-  299
DEB.CM.04.SIVdeb04CMPF3061  -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----QRC-K--N--N---VK-------K---------  336
DEB.CM.99.CM40              -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----QRC-K--N--N---VK-------K---------  332
DEB.CM.99.CM5               -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----QRC-K--N--N---VK-------K---------  328
DEN.CD.x.CD1_CMO580407      -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-V-K--Q-V-----K-----S-K--------A-  332
DRL.DE.11.D3                -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---QRC-T--N-VN---VK-------K--------A-  307
DRL.x.x.FAO                 -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---QRC-T--N-VN---VK-------K--------A-  307
GRV.ET.x.GRI_677_gri_1      AW---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-V-Q--N-QKV------------Q--------A-  307
GSN.CM.99.CN166             -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------QRC-K--N-VN----K-------KE-----F-C-  311
GSN.CM.99.CN71              -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------QRC-K--N-VN--G-K-------KE-----F-C-  311
LST.CD.88.SIVlhoest447      RW--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-V--IHQ-V-VI--K-------KE-A---F-A-  291
LST.CD.88.SIVlhoest485      RW--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-V--IHQ-V-VM--K-------KE-A---F-A-  291
LST.CD.88.SIVlhoest524      RW--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-V--IHQ-V-VM--K-------K--A---F-A-  289
LST.KE.x.lho7               RW--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-V--IHQ-V-VM--K-------KE-A---F-A-  294
MAC.US.x.251_1A11           NY--LPL----------LI---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIRDI------D--LQ--QP...APQQ--L---S---------SVD---Q--YR....QQN-----N--R---Q---Q-C----N--N---VK-------QS-------S-  305
MND-1.GA.x.MNDGB1           -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-RL-KSHC-I--T-V-------A-K-------NVM  294
MND-2.CM.98.CM16            TP---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-C-S--N-VN----K-------K--------A-  307
MND-2.GA.x.M14              TP---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-C-S--N-VN----K-------K--------A-  307
MND-2.x.x.5440              TP---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-C-S--N-VN----K-------K--------A-  307
MNE.US.x.MNE027             NYT-LPL----------LI---K-GA--V-G-Q-----CI-Y-I-Q---C--E------IIR-I------D--LQ--QQ...APQQ--L---S----------VD---Q--YR....QQN-----N--R---Q---Q-C-K--N--N-----------QS------S-  305
MON.CM.99.L1_99CML1         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----QRC-K--N-VN----K-------K--------C-  309
MUS-1.CM.01.CM1239          -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--QRC-K--N-V-----K-------K--------C-  302
MUS-1.CM.01.SIVmus_01CM1085 -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--QRC-K--N-VN----K-------K------F-C-  302
MUS-2.CM.01.CM1246          -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------QRC-K--N-VN----K-------K------F-C-  306
MUS-2.CM.01.CM2500          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------QRC-K--N-VN----K---R---K------F-C-  307
MUS-3.GA.09.09GabOI81       -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------QRC-K--N-VN-------------KE------C-  311
MUS-3.GA.11.11GabPts02      -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--QRC-K--N-VN-------------K--------C-  315
OLC.CI.97.97CI12            IFE-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-L-K----VTV------A----S------F-A-  286
RCM.CM.00.SIVagi_00CM312    -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-C-K--N-VN----K-------K--------A-  306
RCM.CM.02.SIVrcm_02CM8081   -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-C----N-VNV---K-------K--------A-  306
RCM.GA.x.SIVRCMGAB1         -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-W-K--N-VN----K-----S-K--------A-  306
RCM.NG.x.NG411              -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-C-K--N-VN----K-------K--------A-  306
SAB.SN.x.SAB1               -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-C-K--N-VN----K-------K--------A-  318
SMM.CI.79.SIVsmCI2          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-C----N--N---VK-------QS-------S-  306
SMM.LR.89.SIVsmLIB1         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-C----N--N---VK-------QS-------S-  304
SMM.SL.92.SIVsmSL92A        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-C----N-VN----K-------QS-------S-  306
SMM.SL.92.SIVsmSL92B        NY--TPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-M------D--LQ--RGQQPAQPA-GL---S---------NPN---E--YR....AQN-V---D--R---Q---Q-C----N--N---VK-------QS-------S-  306
SMM.SL.92.SL92B             NY--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-C----N--N---VK-------QS-------S-  306
SMM.US.04.G078              NYT-LPL----------L----R-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--QP..--LPA--L--------------VD---Q--YR....QQN-----N--R---Q---Q-C----N-MN---VK-------QS-------C-  306
SMM.US.05.D215              NYI--PL----------L----K-GA--V-G-Q-----C--Y-V-Q---C--E------IIR-I------D--LQ--QP..--LPA--L---T----------VD---Q--YR....QQN-V---N--R---Q---Q-C----N--N----K----K--QS-------S-  306
SMM.US.06.FTq               NY--LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--D------IIR-I------D--LQ--QP.M---PA--L-D------------VE---E--YR....QQN-----N--R---Q---Q-C----N--N---VK-------QA-------S-  307
SMM.US.86.CFU212            NY--LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--QP..---PA--L-D------------VE---Q--YR....QQN-----N--R---Q---Q-C----N--N---VK-------QS-------S-  306
SMM.US.x.H9                 NYT-LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--QP..---PA--L--------------XD---Q--YR....QQN-----N--R---Q---Q-C----N--N---VK-----S-QS-------S-  306
SMM.US.x.PGM53              NYT-LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--QP..---PA--L-D------------VE---Q--YR....QQN-----N--R---Q---Q-C----N--N---VK-------QS-------S-  306
SMM.x.x.pE660.CG7G          NY--LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--QP..--LPV--L--------------VD---Q--YR....QQN-----N--R---Q---Q-C----N--N---VK-------QS-------S-  306
STM.US.89.STM_37_16         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-C----N-VN---VK-----T-QS-------S-  307
SUN.GA.98.L14               NYI--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-MERV--LHQT--VVE------------T---F-A-  294
SYK.KE.x.KE51               -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-C-QV-N-VN----------T-K-------HC-  327
SYK.KE.x.SYK173_COMGNM      -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-C-QV-N-V-----K-----I-K-------HC-  330
TAL.CM.00.266               -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--QRC-K--N-VN---VK---R---K-------RC-  309
TAL.CM.01.8023              -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--QRC-K--N-VN---VK---R---KE------RC-  313
TAN.UG.x.SIVagmTAN1         -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-C-K--N-I-V------A----K-------QA-  313
VER.DE.x.AGM3               AWI-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-C-K--N-V-V----------A-K--------AI  317
VER.KE.x.9063               AW--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-C-K--N-V-V------------K--------AI  313
VER.KE.x.AGM155             AW--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-C-K--N-V-V------------K--------AI  313
VER.KE.x.TYO1_patent        AW--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-C-K--N-V-V------------K--------AI  313
WRC.CI.97.97CI14            GPQWDPL----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-V-QI-V-V------K-----D-KA-------V-  287
WRC.CI.98.98CI04            GPQWDPL----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-V-QI-V-V------K-----D-KT-------V-  291
WRC.GM.05.Pbt_05GM_X02      GPQWEPL----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-V-QI-V--------K-----D-KS-----F-V-  289
```

```
                                                                                  p24 end p2 start  p2 end p7 start           Zn motif                                         Zn motif
H1B.FR.83.HXB2               RAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSAT.....IMMQRGNFRNQR.............KIVKCFNCGKEGHTARNCRAPRKK...........................GCWKCGKEGHQMKDCT..  427
H1O.CM.91.MVP5180            -----T---------------S-----Q-------E-------V--------T---KI-----ASAQQDLKGGYTAVF----QNP-RK.............GPI----------I-K-------R...........................------Q--------KN.  432
H1N.CM.95.YBF30              -----T--------------------QL------G----------------A-------------QQPT-S....VFA-----KGI-.............-PI----------L---K---RG...........................------Q--------KN.  431
H1P.CM.06.U14788             ----------S---D-----------Q-------G-------N--------T----------AAANQASQELK.GGYTTVFMQSG--.............-P--------V--I-K--K---RR...........................------Q--------KS.  425
CPZ.CD.06.BF1167             -------P--------------------------G------L---------A----------Q-TQQNVNMVQG..RAAPQGK-GA.............GNI-------I-----------R-...........................------Q----L----..  453
CPZ.TZ.00.TAN1               -------P--T-----------------H------QG------L---------S---KI-----ASA-AGG.....VN-L-.-GK-PPLK...........KGQLQ------V-------------...........................---R--Q---------TR  456
GOR.CM.07.SIVgorCP2135con    -------D----------I--------Q-------G-------N--------T----------AAAHSDLK.GGYTAVFMQKGGX--GI..........KKGP--------V--I-K------R-...........................------Q--------RS.  427
                                                                                  p27 end p2 start  p2 end p8 start
MAC.US.x.239                 ----TDAA------Q---I--------LV--G--VNP-----L----------Q---LM---LKEALAPVPIP..FAAA-QRGP................R-PI--W-------S--Q-----RQ...........................-------MD-V-AK-P..  428
H2A.SN.86.ST_JSP4_27         ----TDPA------Q------------LV--G--INP-----L----------Q---LM---LKEAMAP-PIP..FAAA-Q...................RRTI--W-------S--Q-----RQ...........................-------A--I-AK-P..  426
H2B.CI.x.EHO                 ----TDPA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LKEALTPS-NP..FAAA-PRAG................KRT-T-W----A-----Q-K---RQ...........................------QQ--I-SK-P..  424
H2G.CI.92.Abt96              ----TDPA------Q---I--------LV--G--MNP-----L-----X----Q---LM---LKEAL-PTALP..FAAA-QKTGG..............KRSTI--W--------V-Q-----RQ...........................-------P--I-AK-P..  431
H2U.CI.07.07IC_TNP3          ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEALAPGPLP..FAAV-QRTGG..............KKSMI--W-------S--Q-----RQ...........................-------P--I-AK-P..  431
H2U.FR.96.12034              ----TDPA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LKEALQPMPIP..FAAV-QRGGP...............RRP-R-W----------Q-K---RQ...........................-----E-P--N-AS-P..  429
ASC.UG.10.RT03               -----D-S------T---------E--A----M.-G-S----L---------HY-TKL-----ASAF.QK-GNYNMVQGP--PKGAR-GPM........KNPNPR-----QF--MSKQ-PH-K-I.........................K-F---R---MARQ-RAS  435
ASC.UG.10.RT08               -----D-S------S---I-----E--A----M.-G-S----L---------HY-TKL-----ASAF.QR-GNYNMAQGPQ-PKGAR-GPM........KNPNPR-----RF--MS-Q-PQ-K-M.........................K-F---R---MARQ-RAS  435
ASC.UG.10.RT11               -----D-S------T---------E--A----M.-G-S----L---------HY-TKL-----ASAF.QK-GNYNMVQGP--PKGAR-GPM........KNPNPR-----QF--MS-Q-PQ---I.........................K-F---R---MARQ-RAS  435
COL.CM.x.CGU1                ---P-AG-I-A--ANN--I-H------R---G-.QKPS--D-LA--------D---K------Q-FQQER.....TN-IE..................V-TA-----QGI--L--M-PKRPIGGAGR.....GRGRGRGGFRGAPRRPVR-FT-NQ---MQR--P..  432
COL.UG.10.BWC01              ---P-GS-I-Q--ANN--I-H---E--R---G-.QKPS--D-LA--------DY--K------Q-FRQE.......T-NMIE.................V-TA---Q--EI--LQKQ-PKRGAGGFRGGGMRGRGRGGRGAWGRGRNPGPLK-FG--QL--IQR--Q..  438
COL.UG.10.BWC07              ---P-GG-I-Q--ANN--I-H---E--R---G-.QKPS--D-LS--------DY--K------QGMQKIGLVE.........................VSTAR-----QV--L----PK--QFPTRG...RGSSRGRGIMRGGMRRGGQAK-FG--QL--IQAT-P..  436
DEB.CM.04.SIVdeb04CMPF3061   -----D-A------T--MI-------R---------NP-----LH--------QQ---LM----VSAMREGGG...LT-V-QR.GPGGP.........RPGNRIR-Y---QI--VQKD--K--TV.........................R-F--QQT--MA-N-P..  463
DEB.CM.99.CM40               -----D-A------T--MI-------RI---G--QNP-----LH--------QQ---LM----ASALKE-GS...LG-V-QRRGPPGS...........RRQIR-----QI--LQKD-KR---T.........................K-F---Q---IA-N-G..  458
DEB.CM.99.CM5                -----D-A------T--MI-------RI---G--QNP-----LH--------QQ---LM----ASALKE-GS...LG-V-QRGGQGGG..........PRRQLR-----QI--VQ-D-KK---V.........................K-F-------IA-N-G..  455
DEN.CD.x.CD1_CMO580407       -----D-S--L---TS-----------Q--E----NP-----LQ---------Q---PT---LASAFQKQGLCHGSRKGAQPPVER..............RGLR-----QI--L-KD--K-KRVT.....................QPGA-F----L---A-N-R..  462
DRL.DE.11.D3                 ---HTDAA------Q---I--------V---G--TNPS----LL---------------M----REQQTA..........----QNPP-GPP....RGPRGPPPRNPR-P---QY---L-Q-TK-----.......................--FR--ALD-MLRN-...  435
DRL.x.x.FAO                  ---HTDAA------Q---I--------V---G--MNPS----LL---------------M----REQQAA..........----QNPP-GPP....RGPRGPPP-NPR-P---QF---L-Q-TK-----.......................--FR--ALD-MLRN-...  435
GRV.ET.x.GRI_677_gri_1       -----P-D------Q---I--------L---G--MNP-----LI--------Q---KLMV-M--NGQ.........N-V-V-PQKKGP...........RGPL-------F--MQ-E-K---QI.........................K-F----I--MA---K..  427
GSN.CM.99.CN166              -----D-A------QS--------E--L----M.-G-N----L---------T--S-L-----ATALKGTSS...YN-V-K-PPGR-.............G-TP--Y---QF--L--D-PK-KER.........................K-F---RA--FS-P-R..  435
GSN.CM.99.CN71               -----D-A------QL--------E--L----M.-G-S----L---------T--S-L-----ATALKG.SS...YN-V---PPGR-.............G-TP--Y----F--I--D-PK-KER.........................K-F----A--LARQ-KT.  435
LST.CD.88.SIVlhoest447       ---GG-H---E--K-KM---------RMVI----EG-------R--------A--G-I-----ATAMQ-QMR...QN-V-VTPP--AQ.GRFVRTGGGGPR-PLT------P-----M--Q--QE...........................---N--SKE-RFAQ-PKP  430
LST.CD.88.SIVlhoest485       ---GG-H---E--K-KM---------RMVI----EG-S--D--K--------A--GKI-----ATAMQ-QMR...QN-V-VTPP--EQ.GRFVRTGGGGPR-PLT------P----M--Q--QE...........................---N--SKE-RFAQ-PKP  430
LST.CD.88.SIVlhoest524       ---GG-H---E--K-KM----------MVI----EG-------K--------A--RKI-----ATAMQGQMK...HL-V-QTPP--AQ.GRFVRTGGGGPRRPLT------P--S-KM--Q--QE...........................---N--AKD-R-AQ-PKP  428
LST.KE.x.lho7                ---GG-H---E--K-KM---------RLVI----EG-S-----K--------A--GKI-----ASAIQQQVR...QN-V-VTPL--AQ.GRFVRTGGGGPR-PLT------P-----M--Q--QE...........................---N--SKE-RFAQ-P..  431
MAC.US.x.251_1A11            ----TDAA------Q---I--------LV--G--VNP-----L----------Q---LM---LKEALAPVPIP..FAAA-QRGP................R-PI--W-------S--Q-----RQ...........................-------MD-V-AK-P..  428
MND-1.GA.x.MNDGB1            ------G---M--QQH--IE----E--Q--RS--KG------LE--------Q----LM--M-RT-VGQ..S...QNFV-QRGPQR.......GPVRQPTGRKPI----N----V--FFK---R-...........................---N--AMD--KAQ-P..  423
MND-2.CM.98.CM16             ---RTDPQ--T---Q---I--------S---G--MNPS----PL--------KY--QMM----KEAQS.......AV---NSGGPPRG.......PPRQPPRNIR-P----F--GL-D-IS----...........................--F---DL--I-RN-P..  434
MND-2.GA.x.M14               ---RTDPQ--T---Q---I--------S---G--MNPS----LL--------KY--QMM----QKMQS.......EV---NSGGPPRG.......PPRQPPRNPR-P---RY--VL-D--L----...........................--F---DT--M-RN-P..  434
MND-2.x.x.5440               ---RTDPQ--T---Q---I--------AT--G--MNP-----LL--------KY--QMM----Q-AQAA.......V---NSGGPPRG.......PPRQPPRNPR-P----F--VL-D------R...........................--F---DP--L-RN-...  433
MNE.US.x.MNE027              ----TDPA------Q---I--------LV--G--MNP-----L----------Q--KLM---LKEALAPGPLP..FAAA-QKGP................R-PI--W-------S--Q-----RQ...........................-------QM--V-AK-P..  428
MON.CM.99.L1_99CML1          ----TD-A------Q------------L----M.-G-S--D-LQ--------M--S-I-----AGAIAN....MPMN-V-ARGPPQ.............RKGQP-------F--M-K--K--QRR.........................K-YN--QP--LA---PQP  434
MUS-1.CM.01.CM1239           ----SD-A------S---------E--L---SM.-G------LQ--------A--S-LM----ASAL.RQNS.QL.N-V-G...ARGK......GSQGGPRGNPR-Y---QF--M--S-PK-KTR.........................K-F---R---LA-Q-R..  430
MUS-1.CM.01.SIVmus_01CM1085  ----SD-A------S---------E--L---SM.-G------LQ--------T--S-LM---LASAF.KQAG.TL.N-V-GAKGARG-.....GNGSQGNRGNPR-Y---QY--V--D-PK--N-.........................K-F---R---LARQ-R..  434
MUS-2.CM.01.CM1246           -----D-A------QS--------E--LV---M.-G---K--LQ----IE--M--S-LM----TNAL.RQNTINTIN-V--..QSPR......GVMGKKRENSTR-Y---QF--L--D-PK-KST.........................R-F-------LARQ-R..  437
MUS-2.CM.01.CM2500           -----D-A------QS--------E--LV---M.-G------LQ--------I--SKLM----ASAL.RQNTIETIN-V-G..QGPR......GGAGGGPRRTPR-----QF--I-KD-PK--VR.........................K-F---R---LA-Q-R..  438
MUS-3.GA.09.09GabOI81        ----TD-A------T---------E--A----M.-G------LQ--------AS---L-----VDAF-KTGATA...-V--AAPIPR..........GPNNRQPR-----QF--L-K--PK--VV.........................K-F---R---MA-Q-N..  438
MUS-3.GA.11.11GabPts02       ----TD-A------S---------E-------M.-G------LQ--------AQ---L-----VSAF-KSGASTAMMQRGPPPG--P............QKRAPR-----QF--L--S-PK-KVV.........................K-F---R---MA-Q--..  443
OLC.CI.97.97CI12             ----G-----T----K--I--------GV----.VKPS----L----------Y--------LRP-Q...........-V-QRPAP..........GGGPQQGRLR-Y---RF---KN--SQKGP........................T-Y----P--IA-N-R..  405
RCM.CM.00.SIVagi_00CM312     -----DPA------QS--I---------V--G--MNP-----L---------QY---------Q-ANSA..........--A-MRGGQR.....SGPPRGQGGKTFL--Y----L--IS-Y-K---R-...........................------E-D-IL---P..  435
RCM.CM.02.SIVrcm_02CM8081    ----TDPA------Q---I---------V--G--MNPS----LL--------Q----------QHANSM.......M-A-MRGQGR.....SGPPRRGGERPVP--Y----T--IS-F-K---R-...........................------D-D-L-----..  435
RCM.GA.x.SIVRCMGAB1          -----DPA------QS--I--------MV--G--MNPS----L---------Q----------QMMQSN.........--A-QSAN.......RGPPRRSGGNPNLR-Y----P--IS-Y-K---R-...........................------SPD-LL----..  433
RCM.NG.x.NG411               -----DPA------QS--I---------V--G--MNP-----L-----A---Q----------QM-Q.......SN--A-Q-GP-KGP......PKLGGGPRFL--Y----T-----Y-KTS---...........................---R--E---L----P..  434
SAB.SN.x.SAB1                ----TDPA------QS--I---------V--G--MNP-----L-----I--AQ---LM----TAAFQQQ-V..GN-FV-Q-ARPRGP....LGGRGRPLNPNI--Y----P--L--F-K---RQ...........................------SPD------Q..  453
SMM.CI.79.SIVsmCI2           ----TDAA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LREAFQPGPLP..FAAA-QRGGGVR..............RPI--W-------S--Q-K---RQ...........................-------S--V-AK-P..  431
SMM.LR.89.SIVsmLIB1          ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEAFQP-SLP..FAAV-QRGGTK...............RT-R-W-R-----S--Q-K---RQ...........................-------T--I-AK-P..  428
SMM.SL.92.SIVsmSL92A         ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEALGPTQLP..FAAA-QKG..P...............SK-R-W-------S--Q-K---RQ...........................-------M--I-AK-P..  428
SMM.SL.92.SIVsmSL92B         ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM----KDALTGPLV.....AA-FRGAAKGQ..........GNKP-IR------T--S--Q-----R-...........................------E---IQAN-P..  432
SMM.SL.92.SL92B              ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM----KDALTGSLV.....AA-FRGAAKGQ..........GNKP-IR------T--S--Q-----R-...........................------E--RIQAN-P..  432
SMM.US.04.G078               ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEALAPGPLP..FAAA-QRGGPR................-T---W----------Q-K---RQ...........................-------T--V-AK-P..  430
SMM.US.05.D215               ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKDALQPGHLP..FAAA-QKGGPR................-MI--W-------S--Q-----RQ...........................-------M--T-AN-P..  430
SMM.US.06.FTq                ----TDPA------Q---I--------V---G--MNP------S---------Q---LM---LKSALAPGVLP..FAAA-QKGGLR...............HT---W-------S-KQ-----RQ...........................-------M--I-AK-P..  431
SMM.US.86.CFU212             ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEALTPGPLP..FAAA-LRGGPK...............RT---W---D-----Q-K---RQ...........................-------M--V-AK--..  430
SMM.US.x.H9                  ----TDPAX-----Q---I--------LV--G--INP-----L----------Q-X-LM---LKDALTQGPLP..FAAV-QKGQ................X--IX-X--X----S--QF----RQ...........................-------A--V-AK-P..  429
SMM.US.x.PGM53               ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEALTPGQLP..FAAV-QKGP................R-TI--W-------S-KQ-----RQ...........................-------T--V-AK-P..  429
SMM.x.x.pE660.CG7G           ----TDPA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LKEALRPDQLP..FAAV-Q.KGQR..............RTI--W-------S--Q-----RQ...........................---G---T--V-AK-P..  429
STM.US.89.STM_37_16          -----DPS------R---I--------LV--G--MNP-----L----------Q---LM---LKE-FQPDPLP..FAAA-QQG.................RRT---W----A----KQ-KG--RQ...........................------QQ----AK-P..  429
SUN.GA.98.L14                ---SG-E---E--K-KM----------LVI----ETPS-----R--------S--GKI-----ASAFRQQVG..RQA-V-QNLPPRNSQGRFVRIGGGGPR-PMT------P--L-NQ--E-K-G......................PPGS------M--KQAQ-P..  436
SYK.KE.x.KE51                -----DPA----L-QS--I-----E-RS----MVK.P-----LQ--------QY--KL-----VMTQH-L......G-I-GPRQGSNP...........RRGPTR-----QL--LQKD-PR-K-L.........................K-FN--GT--IARQ-RQP  452
SYK.KE.x.SYK173_COMGNM       -----DPS--G-L-Q---I-----E-RQ----MVK.P-----LQ--------L---KLM----VMAQQ-V......N-V-GPSKG..............RSMI--Y---QI--MQKD-KK-L-A.........................K-FN---T--LARA-RQP  451
TAL.CM.00.266                ----TD-A------Q---------E--A----M--G------LQ--------Q--S-LM----AEALK.QNTQQVIA-V-Q-QGGP....RGRGGPRRNPPGQIR-Y----F--I-K--P--PRQKV......................PPGT-Y----P--IA-Q-S..  450
TAL.CM.01.8023               ----TD-A------Q---------E--A----M--Q------LQ--------Q--S-LM----AEALK.QNTQQVIA-V-.-QGGP....KGRGGPRRTPPGQIR-Y----F--I-K--P--Q-PRG......................PPGS-F---QM--RAAQ-R..  453
TAN.UG.x.SIVagmTAN1          ----TP-D----------I--------LV--G--IHP-----L--------------KLMV---Q-MQG.......VN-V-GAPRGGRG..........RGRGPPR--K--QI--IQKD-PRAGPN.........................K-L----P--LA---R..  437
VER.DE.x.AGM3                ------G---Q----S--I--------V---G--MHP-----L---------SY--K-M--M-QNMQSQ.......N---Q-GQ-GRP............RPP---Y----F--MQ-Q-PE---M.........................R-L----P--LA---R..  439
VER.KE.x.9063                ------G---Q----S--I--------V---G--MHP-----L---------SY---M--M-QNMQTQ.......N-V-Q-AAGGV-...........RQRPPL--Y----F--MQ-Q-TE---I.........................K-L----L--LA---R..  437
VER.KE.x.AGM155              ------G---Q----S--I--------V---G--MHP-----L---------SY--K-M--M-QNLQSQ.......N-V-Q-GG-GRP............RPPP--Y----F--MQ-Q-PE---I.........................K-L----P--LA---R..  435
VER.KE.x.TYO1_patent         ------G---Q----S--I--------V---G--MHP-----L---------SY--K-M--M-QTMQ-Q.......N-V-Q-GPKR-.............RPPLR-Y----F--MQ-Q-PE---T.........................K-L----L--LA---R..  434
WRC.CI.97.97CI14             ----GAP---Q---DK--I-----------G-.QQP-----L----------Y-SKIM-Q-LQEAL.H.SQ....A-V-QAGAKP..............KGPIR-----QI--M-KE-PK-K-PVPMR....................RGQGPT--G--EV--VQRN-P..  416
WRC.CI.98.98CI04             ----GAP---Q---DK--I--------A---G-.QQP-----L----------Y-SKIM-Q-LQEAL.-.SQ....A-V-QQPVKP..............KGPIR-----QI--M-KD-PK-K-PVPLR....................RGQGPT--G--EI--VQRN-P..  420
WRC.GM.05.Pbt_05GM_X02       ----GVP---Q---DK--I--------Q---G-.QSP-----L----------Y-SKIM-Q-LQEAL.-QSQ....A-V-QQAVK...............RGPLR-----QI--M-KD-KK-KRSPQQR......................GPPT--G---T--IQ-N-P..  416
```

```
                                    p7 end p1 start     p1 end p6 start  Vpr binding                                                                    Vpr binding
                                                                                                                                                        p6 end, Gag end
H1B.FR.83.HXB2              .......ERQANFLGKIWPSYKGRP...GNFLQSRP...EPTAPPEESFRSGVETT................................TPPQKQE.PIDKE..LYPLTSLRSLFGNDPSSQ..*    500
H1O.CM.91.MVP5180           .......G--------Y--PGGT--....--YV-K...QV..SPSAPPMEEAVK--....................QENQ.SQKGDQ.........-E.......---FA--K----T-Q.....-    499
H1N.CM.95.YBF30             ......EG--------S-SPF---......--P-TTTRK.------L--YGFQE-..........................KSTQGKEMQEN--R.TENSLY.P---------------.-    513
H1P.CM.06.U14788            .......G-----------PGGK--....--YV-KQV...Q-----M..................................EEEEMTQNK-XE-KEDE--L.----A--K----T-Q.....-    493
CPZ.CD.06.BF1167            .....QSKSGV-----RT-WWGC--......--V-NQSQTQNQEQSQTQEKLVEKA...............EPSAPP.................M.EEVNHWNSKDGF--K-I--S-Q.....-    529
CPZ.TZ.00.TAN1              N.....NSTGV-----RT-LWGC--......--V-NT-....EKGK....................................AQEQETAQTPVVPTAPPL-MTMKGGF--K-I--S-Q.....-    525
GOR.CM.07.SIVgorCP2135con   .......G--------Y--PGGE--....--YV-KQV...Q-----M.......................................EEGEKEMRKPEKEEL...------K----T-Q.....-    490

                                                                                                                                                        p6 end
                                    p8 end p1 start     p1 end p6 start     PTAP motif                       PSAP in HIV-2 B, U                             Gag end
MAC.US.x.239                .......D---G---LGPWG..KK-...R--PMAQ.......VHQGLMPTAPP.................EDPAVDLLKNYMQLGKQQREK-RESREKPYKEVTED-LH-N----G-Q.....-    511
H2A.SN.86.ST_JSP4_27        ..........----G---LGPWG..KK-...R--PVAQ.......I-QGLTPTAPPID..PVEDLLEKYMQQGKRQREQRERPYKEVTEDFLQLE..KQETPCRETTED-LH-N----K-Q.....-    522
H2B.CI.x.EHO                ..........----G---FGPWG..KK-...R--PVQA.......PQ.GIVPSAPPMNP.AFGMTPQGAIPSAPPADPAEEMLKNYMQLGKKQKE..NRERPYKEVTED-LH-N----E-Q.....-    520
H2G.CI.92.Abt96             ..........----G---FGPWG..KK-...R--PMTQ.......V-QGLTPSAPP................MDPAVDLLKNYMQLGRKQKE....QRNKPYKEVTEX-LH-S----D-Q.....-    510
H2U.CI.07.07IC_TNP3         ..........----G---FGPWG..KK-...R--PMTQ.......I-QGLTPTAP.................PADPAVDLLKNYMQMGRKQKE....SRERPYKEVTED-LH-N----D-Q.....-    510
H2U.FR.96.12034             ........-TKVG---LGPWE..KK-...R--PMAQ.......V-QGLTPTAP.....PAEPAVDLLTPTAPPADPAVDLLKSYMQQGKKQKE..NRERPYKEVTED-LH-N------Q.....-    522
ASC.UG.10.RT03              ....EGAQGG------AFLD..KHLGPRKGPRNF.........................PVVQPSAPPLTLQADTFQEENLGPAVGIQKIP..-G...EKTRESLYP--S----D-Q.....-    517
ASC.UG.10.RT08              ....DSAQGG------AFLD..KHLGPRKGPRNF.........................PVVQPSAPPLTLQADTFQEENLGPAVGIQKTP..-DNNGEKTRESLYP--S----D-Q.....-    520
ASC.UG.10.RT11              ....EGAQGG------AFLD..KHLGPRKGPRNF.........................PVVQPSAPPLTLQADTFQEENLGPAVGTQKTP..-GNNGEKTRESLYP--S----D-Q.....-    520
COL.CM.x.CGU1               .......NK------ATRGV..ELQT.....AIFP-.......KMSKDLPPR.............................................R-KGESLYP--K----D-Q.....-    483
COL.UG.10.BWC01             .......NK------AGRGI..DLQ...TAIFPPA.......MSKMQKQEGR..............................................-RDENTLYP-------E-Q-.....-    492
COL.UG.10.BWC07             .......NK-V----TPGGM..DLQ...TAVFPPK.......MSQRKEPSK-............................................EMDREGIYP--K----E-LL-GQN-    494
DEB.CM.04.SIVdeb04CMPF3061  ...KAAPQSKN-----LPWG..R-AP..R-...........FIEQ...............QHQTWGIGKYIPEGGDEIFSSSRVPEQK-..-QQEETRRGKYPD-SD-K----E-Q.....-    545
DEB.CM.99.CM40              ......QTPR-----NTPWA..S-RP..R---EGIGKY.I-ESTADQVYST............................TRVPVQGQENNPQR-ETTRRGKY.PD-MG-K----E-Q.....-    536
DEB.CM.99.CM5               ......QVPR-----NAPWA.SR--..R---EGIGKY.I-ESTADQVYST............................TRVPVQGERNAPQEDTTTRRGKY.PD-MG-K----E-Q.....-    533
DEN.CD.x.CD1_CMO580407      ..SNTNDQG-V----NGLGR..KN-P..R--V-YY-QMNQDLPWSPVRV.............QQQPQYPQNFQEGGNRQSCQEDKRQNKQS...HQEDEETTRESLYP--K----D-Q.....-    559
DRL.DE.11.D3                .......PK-V----NTPWG..SGKP.RNFPQA-.......................LTPSAPPLPGYLQEDPAERMLQKYMEQGAQQKRQ-R.QQQQK-GPYKEAYN--S----S-QLQ...-    522
DRL.x.x.FAO                 .......PK-V----NTPWG..SGKP.RNFPQA-.......................LTPSAPPLPGYLQEDPAERMLQKYMEQGAQQKRQ-R.QQQQKRGPYEEAYN--S----S-QLQ...-    522
GRV.ET.x.GRI_677_gri_1      .......NG------YGHWG..-AKP.RNFVQYRGDTV.GLEPTAPP.................METAYDPAKKLLQQYAEKGQRLREER....-QTRKQKEKEVEDV--S----G-Q.....-    514
GSN.CM.99.CN166             ...TGTPIKELY--DLGRGR..R-AP..R--PVTFLLP.TAP-E.--EPQQ..............................LNLSTPVGEP-SKGQKE-SRNTLYP--S----E-Q.....-    513
GSN.CM.99.CN71              ....GTPAKEL---DLGRGR...-AP..R--PVTSLLP.SAP-E.-VEPQQ..............................LNLSPPVGEP-NKEQKA-KRNTLYP--S----E-Q.....-    511
LST.CD.88.SIVlhoest447      .......KGKV----YGPWR..SG-P..---PLMGGN..AGV.V-SAPP................MERSPTQAERALETYRNLGQQLRR.....QQQIPQKCVDEPCLSFFF-PD-Q.....-    513
LST.CD.88.SIVlhoest485      .......KGKV----YGPWR..NG-P..---PLMGGN..AGV.V-SAPP................MERNSTKAERALETYRNLGQQLRK.....QQQVPQKCVDEPCLSFFF-PD-Q.....-    513
LST.CD.88.SIVlhoest524      .......KGKV----YGPWR..NG-S..---PIMGGA..AGV.I-SAPP...............MEKVPPTRAERAIETYRNLGQQLRK.....QQQVPQKCVDEPCLNFFF-PD-Q.....-    512
LST.KE.x.lho7               .....KPKGKV----YGPW-.SKP-G..NYP-LGGA...AGR.I-SAPP.................MESAPTKAERALETYRTLGQQLKR...QQQQVPQKCVDEPCLNSLF.PD-Q.....-    516
MAC.US.x.251_1A11           .......D---G---LGPWG..KK-...R--PMAQ.......VHQGLTPTAPP.................EDPAVDLLKNYMQLGKQQREK-RESREKPYKEMTED-LH-N----G-Q.....-    511
MND-1.GA.x.MNDGB1           ...KPAQQQRV----YGPWG..PSKP..--YP............................AQEVTPTAPPLEEKPLQKTLSTYQKLGRGL.....RQKMKE-KREEDFH--ST--QE-Q.....-    503
MND-2.CM.98.CM16            ........KMV----NTPWG..SGKP..R--P.......AMPLT-T.AP............PMPGMEDPAERMLLDYMKKGQQQKAESK.....QEKKERGPYEAAYN--S----T-QLQ...-    518
MND-2.GA.x.M14              ........KMV----NAPWG..SGKP..R--P.......AMPLT-SAPP.............MPGMEDPAEKMLLDYMRKGQQQRAAKET....KQEK--GPYEAAYN--S----T-QLQ...-    519
MND-2.x.x.5440              .......PKMV----NTPWG..SGKP..R--P.......AMPLT-SAPP.............MPGLEDPAEKMLLDYMKKGQQQRAAAGA....QKEEK-GPYEAAYN--S----T-QLQ...-    519
MNE.US.x.MNE027             .......D---G---FGPWG..KK-...R--PMAQ.......MHQGLTPTAPP.................EDPAVDLLKNYMQLGKQQRE....SKRKPYKEVAED-LH-N----E-Q.....-    507
MON.CM.99.L1_99CML1         P....KQNKGV----NPFGP..KKGP..R--PLTSVQP.SAPTL-A...............................GTAEPVNLNQEIGS-...KTGTP--TRRDLYP--A----E-Q.....-    511
MUS-1.CM.01.CM1239          ....SEGAKS-----DPLGL.QR-RP..R--P-TSLLP.TAPPEQ-GDLVNLSPPV.......................................GPKDGASDRKSLYP--S---QD--LEK..-    506
MUS-1.CM.01.SIVmus_01CM1085 ....SDNTKAV----DPFGL.QK-RP..R--P-TSLLP.TAPPEQ-GELVNLSPAV.......................................GPGEG-QDRKSLYP--S---QD-Q-D-..-    510
MUS-2.CM.01.CM1246          ...TDTGKSAV-----DLGF.RS-KP..R--PVTSLMP.TAPPDTA-PVNFSPLVGPKE.....................................-P-K-RKDRKTLYP--S----D-Q-K-..-    517
MUS-2.CM.01.CM2500          ...TNSDKGAV-----GFGF.HR.KP..R--PVTSLMP.SAPPDTA-IVNFSP-.............................VGPRKEN...QT-QAPPDRKNLYP--S----D---E-..-    519
MUS-3.GA.09.09GabOI81       .....TSTGRV----NPFGP.QK-GP..R--PVTQLVP...-----GTGEPVNFSP..................................KVGEPQKEKE-TRKTLYP--S----D-Q-E-..-    514
MUS-3.GA.11.11GabPts02      ......SQGKV----NPFGP.QK-GP..R--PVTQLVPTA-PDFGKPVNFAPKVG...................................-PTTSGEEGETTRNSLYP--S----D---E-..-    521
OLC.CI.97.97CI12            .VGGQQQQSKV-----WRGG.-AE-PLNMQQRTEMDIF.PSAPTL-..........................QILEKGTR.......-KREM--EKKPLYP--A----E-Q.....-    481
RCM.CM.00.SIVagi_00CM312    ........K-V----NAPWK..QGKP..R--PLTSVTPSA-GMDQYDP..................AEAMLKRYLEQGKKQKQQKRQE....-KSQQGRAYEEA-S--N------Q.....-    518
RCM.CM.02.SIVrcm_02CM8081   ........K-V----SAPW-..RGKP..R--PLTSLTPSA-GVEQDNH.................ASQYDPAEAMLRKYLEQGRMQKQQQR-E-KAQKGKAYEEA-A--N----S-Q.....-    524
RCM.GA.x.SIVRCMGAB1         ........K-I----RLPWG..QGKP..R--PLTSLTPSA-GMESNYDP..................AEEMLKNYLRRAGEQKRQQRQE....-SKKREGAYQEA----N----S-QLQ...-    519
RCM.NG.x.NG411              ........KEV----NAPWK..QGKP..R-LPLTSLRPT.....APG...........REEGWNPQYDPAEEMLRKYLALGRQHKQEQRKE...NKEKVGRAYEDA-S--N----S-L.....-    522
SAB.SN.x.SAB1               ........K-V----FGPWG..RGKP.RNFP-T-IRPT.A-PMERDY-RPEENWYADRPPTRGPGPDDPATALLKQYAVQGKRQKQQWQ.....NHSPQQSPYEEAYS-------E-Q.....-    555
SMM.CI.79.SIVsmCI2          ..........----G---FGPWG..KK-...R--PMTQ.......LQQGLTPTAPPMDPA........VAAPPMDPAVDMLRNYLKLGQQQRE....RKERPYKEVTED-LH-N----D-Q.....-    519
SMM.LR.89.SIVsmLIB1         ..........----G---FGPWG..KK-...R--PMAQ.......MHQGLTPTAPP.................ADPAVDLLKNYMQLGKKQRE....NRERPYKEVTED-LH-N----D-Q.....-    507
SMM.SL.92.SIVsmSL92A        ..........----G---FGPWG..KK-...R--PMAQ.......M-QGLTPTAPP.................ADPAVDLLKDYMKLGRKQKE....SRERPYKEVTED-LH-N----E-Q.....-    507
SMM.SL.92.SIVsmSL92B        .......NQK-G---LGPWG..KK-...R--PMQT.........TSLTPSAPP................DPAAQMVK.EYLQKAQREKT....RRSRPYKEVTED-LH-N----E-Q.....-    508
SMM.SL.92.SL92B             .......NQK-G---LGPWG..KK-...R--PMQT.........TSLTPSAPP................DPAARIVK.EYLEKAQREKT....RRSRPYKEVTED-LH-N----E-Q.....-    508
SMM.US.04.G078              .......D---G---FGPWG..KK-...R--PMAQ.......V-QRLVPTAPP.................EDPAVDLLKNYMQLGRKQRE....NRERPYKEVTED-LH-S----E-Q.....-    509
SMM.US.05.D215              .......D---G---FGPWG..KK-...R--PMAQ.......VHQGLTPTAPP.................ADPAVDLLKNYMQLGRKQRE....NRERPYKEVTED-LH-N----E-Q.....-    509
SMM.US.06.FTq               .........---VG---FGPWG..KK-...R--PMTQ.......M-QGLTPTAPP-DLA.............RDPAEDLLKRYMQQGRKQRE....SRERPYKEVRED-LH-S----E-Q.....-    514
SMM.US.86.CFU212            ..........----G---FGPWG..KK-...R--PMAQ.......VHQGLTPTAPP.................ADPAVDLLKEYMKQGRKQRE....NRERPYKEVTED-LH-N----D-Q.....-    509
SMM.US.x.H9                 ..........----G---LGPWG..KK-...R--PMAQ.......M-QGLTPTAPP.................EDPAVDLLKNYMKVGRRQRE....NRERPYKEVTED-LH-N----E-Q.....-    508
SMM.US.x.PGM53              ..........----G---LGPWG..KK-...R--PMAQ.......M-QGLTPTAPP.................EDPAVDLLRNYMKMGRKQRE....NRERPYKEVTED-LH-N----G-Q.....-    508
SMM.x.x.pE660.CG7G          ..........----G---LGPWG..KK-...R--PMAQ.......M-QGLIPTAPP.................EDPAVDLLKNYMKMGRKQRE....NRERPYKEVTED-LH-N----E-Q.....-    508
STM.US.89.STM_37_16         .........---VG---FGPWG..KK-...H--PMAQ.......I-QGLTPTAPP.................EDPAADLLRSYMQLGKKQRE....SRKTPYKEVTED-LH-N----E-Q.....-    508
SUN.GA.98.L14               ......QGK-V----YGPWN..RG-P..---PVMPSAP.P.................LEDLTLGNRMTPPQSKAERALETYRLLGQGLR....AQQKR--RGECQEPC-NM--PE--.....-    522
SYK.KE.x.KE51               RKGQGNPPP-------G-GN..RK-P..A--PVM-ETP.TAPPAEDWPWQQQTWGN.YASPQQKHIAMATPLQPQSSPKTVKDLLVPGK-S...-KKEEMKEKGPLYP--Q----D-Q.....-    562
SYK.KE.x.SYK173_COMGNM      KRNQGPPVA-------G-GV.SR--P..A--PVRSEPS.A-PLEDI........-DGPWLTWSAQMSQQAQAKAQNSPSKKPPTNREVLS-....KESSG--ETKSLYP--S----E-Q.....-    555
TAL.CM.00.266               .......SA-------SFLG..KI-P..VRGPRNF-.........VVQPSAPPI................EPLTASWQEGTTTWTPPAKK.....PQETTTTRESLYP--A----D-Q.....-    528
TAL.CM.01.8023              .......QP-------AFLG..KI-P..VKGPRNF-VV.Q-----I..........................EPLTASWQEGTTTWTPPTKD....AAAPKQQTRESLYP--A----D-Q.....-    532
TAN.UG.x.SIVagmTAN1         .......SG------RMPTW..-TKP.RNFLE.QGGAV...PTA-PMPAHGFP................TGSPAAGAYDPARKLLEQYAKKG..DQLRKQKEKELEDY--S----G-Q.....-    524
VER.DE.x.AGM3               ........G-V----YGRWM..-AKP.RNFPAATLGVE.PTAP--PSP....................YDPAKKLLQQYADKGKQLREQRK...KP-AVNPDWTEGYSL.N----E-Q.....-    522
VER.KE.x.9063               ........G-V----YGRWT..-TKP.RNFLAATHGVE.PSAP--PTP....................YDPAKKLLQQYAEKGKQLREQGK...RT-PTNPDWAEGYSL.N----E-Q.....-    520
VER.KE.x.AGM155             ........G-V----YGRWM..-TKP.RNFPAATLGAE.PSAP--PNN.................STPYDPAKKLLQQYAEKGKQMRNQNR...NP-ANNPDWNEGYSL.N----E-Q.....-    521
VER.KE.x.TYO1_patent        ........G-V----YGRWM..-AKP.RNFPAATLGAE.PSAP--PSGTTP.................YDPAKKLLQQYAEKGKQLREQKR...NP-AMNPDWTEGYSL.N----E-Q.....-    520
WRC.CI.97.97CI14            .....KIR--V----NGRGQ..MPSNNYPVQKLRSLNN.FGPTTTMP-APPMR-DPNI.........QELEKYLLKGAQLREQQEHKK......TEGQKR-VGVPEWP--K---EP-Q.....-    512
WRC.CI.98.98CI04            .....KSR--I----NGRGQ..MPSNNYPVQKVRSLNN.FGP-.TIP-APPMR-DPNI.........QELEKYLLKGAQLKEQAQAK.......KEEKKR-TG-PEWP--K---ET-Q.....-    514
WRC.GM.05.Pbt_05GM_X02      .....ETR--V----NGRG...Q--...S-NYPVQ..........KMNAGVNTFGPVV.....PTAPVDPTMQELEQYLLRGAQMKRE......KEKEMQKEKGKVEYP--K---EP-Q.....-    503
```

```
                              Pol start
                              Gag-Pol TF start                                                                   Gag-Pol TF end protease start
H1B.FR.83.HXB2                FFREDLAFLQGKA...REFSSEQT...RANSPTRRELQVWGRDN................................NSPSEAG.ADRQG..TVSFNFPQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSL....PGRWKPKMIGGIGGFIKVRQYDQILIEICGH  125
H1O.CM.91.MVP5180             ----V--SGGHE-...-QLCA-...TS..VPISPTNGGGSEG....................TRES.ESEGGS........-R.......A-PICL--IP--D--I--A-V--H-C------------NNIQ-....E---T------------KE-NNVTV-VQ-K  121
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
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.TZ.00.TAN1                ----THPLVGVQT..--LCA-HP....RERE...................................GSGAGDSTDTSGANCPTT-DDDERRVL---N-----MM-V-VQ--VCQ---------S-FCNIK-....K-Q-T--T-------VP-SE-YN-PVQ-GNK  121
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
                              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.SN.86.ST_JSP4_27          ---VGPMG..KE-...PQ-PCGP.......N-AGADTNSTPDR..PSRGPTREVHAAREKAERAEREAIQRSDRGLPAAR..ETR-TMQRDDRGLAA--FS--K--V--AHVE--PV-V--------SIVAGVE-....GSNYS--IV-------NTKE-KNVE-RVLNK  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----------IIHDQDIEL..GDN-H--I---V--N-R-KA-HSVE--WQ-K  135
ASC.UG.10.RT08                ----GFFR..QASGAQEGPQEL........................PSSTAISSTFDSAGRHFSGGEPRPSCRDTEDP..RRQQWGKNKG-PLSL-VIP--G--I-SCLVE--EVDL----------IIHDQDIEL..GDN-H--I---V--N-R-KA-HNVD--WQ-K  138
ASC.UG.10.RT11                ----GFFR..QAPGAQEGPQEL........................PSSTAISSTSDSAGRHFSGGELRPSCRDTEDP..RREQWGEDKG-PLSL-VIP--G--I-SCLVE--EVDL----------IIHDQDIDL..GDN-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-EDVQ..LT-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-SDID..IK-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-EEIQ..LT-T-TVVT------S-RAK--KNRVMK-G-K  109
DEB.CM.04.SIVdeb04CMPF3061    ----TSLG..EESP..Q-..........FHRA................ATPDMGHREVHTGGRGRDLLLFQGTGTET..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..Q-L-RRNWEV.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...Q--FGRDWEI.HPREHS-SG-FN..........................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..Q-LCTVLPSDESGFAMEPS-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.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_1A11             ---PWSMG..KE-...PQ-PHGS.......-ASGADANCSP................RGPSCGSAKELHAVGQAAERKT-RKQREALQGDDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV-------NTKE-KNVE--VL-K  137
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                  -RESFWPR..E.GA..K-L--NLLAA.NSP-G-NPXGLQGELLT..........................SSGTTRDP....SADSPE-DEEG-LSLLSIP--G--VR-VE-A--SVSM---------IVQ-QDI..NLGEN-T---V-----L-Q-K-FKGVTVSFQ-K  133
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.....................................EGRGAAEGEGRNQEN-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.....................................T-NHLGGRGRDDKEL-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..................CLPVRSSRSHAKEIPGTGQNAETATE-GREGSEGESLRG-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.SIVsmSL92B          ---PWPMG..KE-...SQLPHAD.........NIIDTVSTP................RSSSPDGQ.GVSTESTEGKD....KEEQALQGGDGG-AA--FS-RR--V-RAL-E--PL-V--------SIVAGIE-....GPGYT--IV-------NTKE-RTVK--VL-K  130
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
SMM.US.04.G078                ---VWPMG..KE-...SQ-PHGP.......-TSEAGANGSP................RGSGSGSAKELHAAREEAEG....EQRETLQGGDGG-AA--FS--R--V--AY-E--PV-V--------SIVAGIE-....GPQYT--IV-------NTKE-QDVE-KVL-K  133
SMM.US.05.D215                ---IWPLG..KE-...PQ-PHGP.......-TSGADADCPS................GGSSGRPAEELHAARQEAEG....EQRETLQGGDRG-AA--FS--R--I--AY-E--PV-V--------SIVAGIE-....GPNYT--IV-------NTKE-TEVE-KVL-K  133
SMM.US.06.FTq                 ---LWSMG..KE-...SQLPHDP.......DASGAYANGPS-RPS............KGSGGGSLEEVHAAGQEAEG....EQRETIQGGEGG-AA--FS--R--V--AH-E--PV-V--------SIVAGIE-....GLNYT--VV-------NTKE-KNVA--VL-K  137
SMM.US.86.CFU212              ---XWPMG..KE-...PQ-PHGP.......-TSGADANSTP................SRSSCRXVKGIHETGQETER....EQRETLQGGDGG-AA--FS--R--V--AQVE--PV-V--------SIVAGVE-....GTNYT--IV-------NTKE-KNVE--VL-K  133
SMM.US.x.H9                   ---AWPMG..KE-...PQ-PHGP.......DASGADTNCSX................RGSSCGSTEELHEGGQKAEG....EQRETLQGGNGG-AA--FS--R--V--AY-EE-PI-V--------SIVAGIE-....GPNYT--IV-------NTKE-KDVK-KVL-K  133
SMM.US.x.PGM53                ---AWPMG..KE-...PQ-PHGP.......DASGADANCSP................RGSSCGSTEELHEDGQKAEG....EQRETLQGGDGG-AA--FS--R--V-AAY-EE-PV-V--------SIVAGIE-....GPNYT--IV-------NTKE-KDVK-KVL-K  133
SMM.x.x.pE660.CG7G            ---AWPMG..KE-...PQLPHGP.......DASGADTNCSP................RGSSCGSTEELHEDGQKAEG....EQRETLQGGDRG-AA--FS--R--V--AY-EE-PV-V--------SIVAGIE-....GPNYT--VV-------NTKE-KDVK-KVL-K  133
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
H1O.CM.91.MVP5180          EVQ....------------L---I--GL-----------AP-...------------------SR---E--TA--Q---Q-----R----------I----------------------------------G-.......-QRQ-------------C---P----  277
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
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.TZ.00.TAN1             EVR.....A-----E--I------I-K-L------------V-...K-Q--E-----------SK---E--T---KTL-------AV----------I-------TS-----------------L-------------.......R-RNM-----------I---P----  277
GOR.CM.07.SIVgorCP2135con  -VV....-----------------I--K----------S-DV-...--S--------------SK---E--TA--Q---Q---------------I----------------------------------G-.......---R------------C---P----  273
                                     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.SN.86.ST_JSP4_27       RVR.....A-IMT-D--I--F---I--AL-MS--L-VAK--PI...KIM-----K----LR-----K---E--K---EK--R--QLEEAP-T------T-------KN---M-I-------V----T-I-----------.......A---RI------------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.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-----SR---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_1A11          RIR.....--IMT-D--I--F------AL-MS--L--AKV-P-...K-A----KV---L-----SK---V--R---EK---D-QLEEAP-T------T-------KN---M-I------RV----T-----------------.......A-R-RI----I------I----E--Q  293
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---  291
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.SIVsmSL92B       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----T-----------------.......-EMERI----I-----------PE--Q  286
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
SMM.US.04.G078             -IK.....--IMT-D--I--F------AM-MS--L-VAKV-P-...K-------KE---L-----SK---E--K---EK-----QLEEAP-T------T-------KS---M-I------RV----T-I-------------.......A-RRRI------------I----E--Q  289
SMM.US.05.D215             -IK.....--IMT-D--I--F------TM-MS-----VAKV-P-...K-T-----K----LR----SK---E--K---EK---D-QLEEAP-T------T-------KN---M-I-------V----T-----------------.......A-RRRI------------I----N--Q  289
SMM.US.06.FTq              TIR.....---MT-D--I--F------RL-MS-----VARV-P-...K-T-----Q----LR----SK---E--K---EK-----QLEEAP-T------T-------KN---M-I------RV----T-----------------.......A-RRRI------------I---PE--Q  293
SMM.US.86.CFU212           RIR.....--LMT-D--I--F------AL-MS-----MAQV-P-...K-T-----K-----R----SK---Q--K---EQ---D-QLEEAP-T------T-------KN---M-I-------V----T-----------------.......A-R-RI------------I--------Q  289
SMM.US.x.H9                VIK.....--IMT-D--I--F------AM-MS--L--AKV-PI...K-T-----K----LR----SK---I--R---EK---D-QLEXAP-T------T----XX-GN---M-I-------V----T---------X------.......A-RRRI------------I----E--Q  289
SMM.US.x.PGM53             VIK.....--IMT-D--I--F------AM-MS--L-VAKV-PI...K-T-----K----LR----SK---V--R---EK---D-QLEEAP-T------T-------KN---M-I------RV----T-----------------.......A-RRRI------------I----E--Q  289
SMM.x.x.pE660.CG7G         VIK.....--IMT-D--I--F------AM-MS--L--AKV-PI...K-T-----KE---LR----SK---I--R---EK---D-QLEEAP-T------T-------KN---M-I-------V----T-----------------.......A-RRRI------------I----E--Q  289
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
H1O.CM.91.MVP5180           --------V------V--------------------------D----S--EVE----I---------PLAE--KRV-L--E--YQ--F------------------------------Q--D-EV..---------------------Q--R-KE----I----S----V-  445
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
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.TZ.00.TAN1              --------L--N---K-F----------------------D----EH--VD---------I----NEEE--KL-KK------T---E-----Y-EK---M---------N-----N-T---PEQ..----H-----------------H---TKE----I--V-G---PVE  445
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
                                                         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.SN.86.ST_JSP4_27        -----L-----AE--K--I-K-S-----------YT-RQV------A----ILI-----ILIA--RTDLE-DRVVLQ-KEL-NGL-FS---E-F--D--YQ------W-T--KL-R-Q--Q-EV..-----------V----A-------T-N--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.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-------FI---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_1A11           -----L--V--AE--K--I-K-------------YT-RHV------A---VTLV-----ILIA--RTNLE-DRVVLQ-KEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--QRET..-----------V----A-------TKH--R-I--KMT---EVQ  461
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----E-..--------------------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  459
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.SIVsmSL92B        --------V--QE--K--I-K-----------V--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
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
SMM.US.04.G078              -----L--V--AE--K--V-K-------------HT-RQV-D----AH--VTLI-----ILIA--RTDLE-DKVVLQ-KEL-NGL-FS---E-F--D--YQ------W-T--KL-K-E--Q-EN..-----------V----A-------TKH--R-I--KMT---EVQ  457
SMM.US.05.D215              -----L--V--AE--K--I-K-------------YT-RNV------A---VTLI-----ILIA--RTDLE-DKVVLQ-KEL-NGI-FS---E-F--D--YH------W-T--KL-K-E--QRET..-----------V----A-------TKH----I--KVS---EVQ  457
SMM.US.06.FTq               -----L--V--AE--K--I-K-------------HT-RNV------A----TLV-----ILIA--RTDLE-DRVVLQ-KEL-NGL-FS---E-F--D--YH------W-T--KL-K-E--QRE-..-----------V----A-------TKH----I--KM----EVQ  461
SMM.US.86.CFU212            -----L--V--AE--K--I-K-------------HT-RGV------A---VTLI-----ILIA--RTDLE-DRVVLQ-KEL-NGM-FS---E-F--D--YQ------W-T--KL-K-E--Q--N..-----------V----A-V-----TKH----I--KMT---EVQ  457
SMM.US.x.H9                 -----L--V--AE--K--I-K-------------HTXRNV------A---VTLI-----ILIA--RTDLE-DRVVLQ-KEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--QRET..----N------V----A-------TKH--R-I--KMT---EVQ  457
SMM.US.x.PGM53              -----L--V--AE--K--I-K-------------HT-RNV-G----A---VTLI-----ILIA--RTDLE-DRVVLQ-KEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--QRET..-----------V----A-------TKH----I--KMT---EVQ  457
SMM.x.x.pE660.CG7G          -----L--V--AE--K--I-K-------------YT-RNV------A---VTLI-----ILIA--RTDLE-DRVVLQ-KEL-NGI-FS--EE-F--D---Q------W-T--KL-K-E--QRET..-----------V----A-------TKH--R-I--KMT---EVQ  457
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
H1O.CM.91.MVP5180         -SK------E----K----------Q-D---WVS---H-E------V--DE.H----------QKAS----IR--A-V---VSQ-A------L-.--R--VTR-------A--------------S----I-------T-----------------N---------EQ-  613
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
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.TZ.00.TAN1            M-R------E--KQ----K-Q-A----KLP-Q-A----------------E.G---------KSP-T---EIR--AGLI---GN---I---IV-.--L--VS----SQ---D---V--V-----I-----IR---N-LSD--PE-----------DS-K-R--------  613
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
                                                                                                                                                                                        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.SN.86.ST_JSP4_27      W--L--A--E--KI--SQEQE-C--QEE-E-E-TV--DQDN----K-H-GG..-I--V----KVKNT---G-RL-AQV----GK-AL----RI-.--H--VERD---Q--DN---V----D-D-IS-----R-VFN-V-D--L------T--SC-KQSRE-----I-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--REEEN-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.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_1A11         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-------N-IS-----R-VFN-V-D--E-E--Y-T--SC-KQS-E-----I-D--  628
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-  626
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.SIVsmSL92B      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
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
SMM.US.04.G078            W--M--A-YE--KI--SQEQE-C--QED-P-E-T-V-SQDN--S-K-H-GE..-V--V----KIKNT---G-RL-AHV----GK-A-----QV-.--HM-VER----Q---D---V------D--S-----R-VFN-V-D--E---------SC-KQS-E-----I-D--  624
SMM.US.05.D215            W--M--A-YE---I--SQEQE-C--QED-Q-E-TVV-NQDN--S-K-H--G..-T--V--F-KIKNT---G-RL-AHVI---GK-A-----QV-.--H--VER-V--Q---D----------D--S-----R-VFN-V-D--E-E-------SC--QS-E-----I-D--  624
SMM.US.06.FTq             W--M--A-FE---I--NQEQE-C--QED-P-E-T-V-SQDN--S-K-H--N..-I--V----KIKNT---G-RL-AHV----GK-A------V-.--HM-VER-I--Q---D---V------D-IS-----R-VFN-V-D--E---------SC--Q--V-----I-D--  628
SMM.US.86.CFU212          W--L--A-YE--KI--SQEQE-C--QEG-P-E-TVV-NQDN--S-K-H--D..-V--V--F-KIKNT---G-RL-AHV----GK-A-----RI-.--H--VER-I--Q---D---V----D-D--S-----R-VFN-V-D-LE-E----T--SC--QS-E--------DG-  624
SMM.US.x.H9               W--M--A-YE--KI-PSQEPE-C--QEG-P-E-PVI-SQDN--S-K-H--D..-I--V--F-KIKNT---G-SL-AHV----GK-A-I---QV-.R-H--VEX-I--Q---D---V------D--S-----R-VFN-V----Q--------RSC--QSRE-------D--  624
SMM.US.x.PGM53            W--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-SQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHV-E--GK-A-----QV-.--H--VER-I--Q---D---V------D--S-----R-VFN-V----Q---------SC--QSRE-------D--  624
SMM.x.x.pE660.CG7G        W--M--A-YE--KI--SQEQE-C--QEG-PIE-TVI-SQDN--S-K-H--D..-V--V--F-KVKNT---G-RL-AHV----GK-AL----EV-.--H--VER-I--Q---D---V----D-D--S-----R-VFN-V----Q---------SC--QSRE-------D--  624
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..KVLFLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLK.GEAMHGQVDCSPGIWQLDCTHLE 784
H1O.CM.91.MVP5180         K-NIIK-EE-----A--M-VLI-----KEQ---------V----SS--T--D-PI-Q----E-T---R---T------------KI-----KD--..R----E---Q--ED----------L----G---I-----I---P--HI-.---T-----Y--E---M------ 780
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---------- 791
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- 781
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--RF----KML-DEY----I-----I-Q-N--HR-.---I--------E--------I- 787
CPZ.TZ.00.TAN1            -YRSKD-EN----QA--W-VD---K---AQ---------VM-VL-GL----D-PI-E---QK-TQ-TAI---------------E-----KN--..-I------NE--ED-D------K-L-DEY----------I-Q-P--HI-.---I-----Y--E---I------ 780
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------------------ 776
                                                                                                                 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------ 795
H2A.SN.86.ST_JSP4_27      -D--RL-EQ----QA--E-FAM-VT---PKA--IV----VM--VAG--TE---KI------EM----AI-V------------QE--H---Q---..Q----EK-EP--E--------VKELSHK-G--KL--RQ--NT-TQ--Q-.---I----NAEL-T--M------ 808
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------ 808
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------ 791
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------ 791
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------ 804
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------ 793
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------ 796
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------ 796
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- 762
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- 764
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- 764
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------ 792
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------ 789
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------ 789
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------ 811
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------ 798
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------ 798
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------ 806
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------ 795
GSN.CM.99.CN71            KYRAIE-EG----QA----VL---KEGPPRM-L------VM--L-S--EV-T-P--E---QE-LG--A---S------------TE------R---..Q---MEN-EP-VED--------KYLRDQYKI-ALL-----NK-S---IH.--PK----NAEL----M------ 788
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- 803
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- 803
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---------- 804
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---------- 803
MAC.US.x.251_1A11         KD--KV-EQ----QA--E-FLM--T---PKA--IV----VM---TGC-TE---R------KEM---SEI-V------------QE--H---Q---..Q----EK-EP--E--D-----VKELVFK-G--RI--RQ--DT----HQ-.---I----NSDL-T--M------ 795
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------ 793
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- 796
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------ 797
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------ 797
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------ 791
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------ 791
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------ 785
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------ 785
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------ 790
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------ 789
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------ 786
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------ 789
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------- 785
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------ 797
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------ 803
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------ 797
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------ 802
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------ 816
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------ 800
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------ 785
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------ 790
SMM.SL.92.SIVsmSL92B      KE--LP-EQ----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------ 788
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------ 788
SMM.US.04.G078            -N-IKS-EQ----QA--E-FLM-VT---T-A--IV----VM--VTG--TE---KI------EM---TAI-V---------L---QEI-H---Q---..Q----EK-EP--E--------VKELVFK-GI-RL---Q--DT-----Q-.---I----NSEL-V--M------ 791
SMM.US.05.D215            -DR-K--EQ----QA--E-FLM-VK---PKA---V----VM--VAG--TE--NSI------EM---TAI-V------------QE--H---Q---..Q----EK-EP--ED-------VKELVFKYGI-RL---Q--DT----HQ-.---I----NAEL-T--M------ 791
SMM.US.06.FTq             KD--KL-EQ----QA--E-FLM-VT---P-A--IV----VM--VAG--TE---RV------EM---TAI-V------------QEI-H---Q---..Q----EK-EP--E--------VKELVFK-RI-RL---Q-IDT----HQ-.---I----NAEL-T--M------ 795
SMM.US.86.CFU212          KDR-KV-EQ----QA--E-FLM--T---PKA--IV----VM--VAG--TE--NKI-S----EM---TAI-VT-------L---QE--H---Q---..Q----EK-EP--E--------VKELVFKYGI-RL---Q--DT----HQ-.---I----NADL-T--M------ 791
SMM.US.x.H9               -G-TKL-EQ----QA--E-F----A---PKA--IV----VM--VAG--TE---R-------KM----AI-V-------X--E-QE--H---QE--..Q----EK-EP--E--------VKKLVFK-X--RL---Q--DT----H--.---I----NAXL-T--M------ 791
SMM.US.x.PGM53            -D-AKP-EQ----QA--E-F----A---PKA--IV----VM---AG--TE---R-------EM----AI-V------------QE--H---Q---..Q----EK-EP--E--------VKELVFK-GI-RL---Q--DT-H--HQ-.---I----NAEL-T--M------ 791
SMM.x.x.pE660.CG7G        -D-AKL-EQ----QA--E-F----A---PKA--IV----VM---AG--TE---R-------EM----AI-V------------QE--H---Q---..Q----EK-EP--E--------VKELVFK-S--RL---Q--DT----HQ-.---I----NAEL-T--M------ 791
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------ 791
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- 803
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------ 823
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- 818
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------ 792
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------ 793
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------ 803
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------ 800
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- 800
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- 803
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------ 803
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------ 808
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------ 807
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------ 799
```

```
H1B.FR.83.HXB2              GKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTIHTDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQI.TKIQNFRVYYRDSRNPLWKGP 953
H1O.CM.91.MVP5180           --I-I-------DF---------------------A-----V------P---S-AMK-----T--QH--------------A------S--Q-----------------V-----------T----LI--L-SQ---T------.L--N----------D-I---- 949
H1N.CM.95.YBF30             --I----------L--------------I----------V---------S---K------N-----------------A-----------I--------------------------T----I---------TN--T--.L-V----------D-I---- 960
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---- 950
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---- 956
CPZ.TZ.00.TAN1              ----I--------F--------E---R-----I---------K------P---STA-K------Q-QH--------------------Q--Q--E-------Q-R---I---Y-----------E-T----LL--LT-N----Q------.L-V---------A-D-I---- 949
GOR.CM.07.SIVgorCP2135con   --I----------F---------------------A-----I------P---S---K-----LNVTH-------------------------Q-----------------V-----------T--D--I-ML--Q---T------.L-------------D-I---- 945

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----- 964
H2A.SN.86.ST_JSP4_27        --I-I--------F-------Q-S-RQ--L------S---ITHL-----A---SQE-KMVA--I--E-S--V-----------A--HH--NQ-SRI-E--NTVE-I-L---HCM----R----DMTPA--LINMVTAEQEIQF--AKN.S-L------F-EG-DQ----- 977
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----- 977
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----- 960
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----- 960
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----- 973
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-- 960
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-- 963
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-- 963
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-- 931
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-- 933
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-- 933
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---- 961
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---- 958
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---- 958
DEN.CD.x.CD1_CMO580407      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-- 978
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----- 967
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----- 967
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---- 975
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-- 964
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-- 957
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---- 972
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---- 972
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-- 973
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---- 972
MAC.US.x.251_1A11           --I-I--------F-------Q---RQ--L----------ITHL-----A--ASQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSVE-I-L---HSM----R--T-DMTPA--LINM-T-EQEIQFQ-SKN.S-FK-------EG-DQ----- 964
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----- 962
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---- 965
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---- 966
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---- 966
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----- 960
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---- 960
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---- 962
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---- 954
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---- 954
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-LR- 959
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-- 956
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------ 955
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------ 958
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-- 953
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---- 966
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---- 972
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-- 966
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---- 971
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---- 985
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----- 969
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----- 954
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----- 959
SMM.SL.92.SIVsmSL92B        ----I--------SF------R---RQ--L----I-S---I-HL-----A---SQE-KMVA--L-VE-S--V-----------A--LH---N-DKI-E----SVE-L-L--AHCM----R----DMTPA----NM-T-ELE-QY-TS-N.S-F-----------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----- 957
SMM.US.04.G078              --IVI--------F-------Q---RQ--L------S---ITHL-----A--ASQE-KMVA-----EHT--V-----------A--HH--NQ-ERI-E--NSVE-I-L---HCM----R----DMTPA--LINM-T-EQEIQFQ-SKN.S-FK-----F-EG-DQ----- 960
SMM.US.05.D215              ----I--------F-------Q---RQ--L------S---ITHL-----A--ASQE-KMVA------E-T--V-----------A--HH--TH-DRI----NSIE-V-L--THCM----R----DMTPA--L-NM-T-EQEIQFQ-SKN.S-FK-------EG-DQ----- 960
SMM.US.06.FTq               ----I--------F-------Q---RQ--L------S---ITHL-----A---SQE-KMVA------E-T--V-----------A--HH--TQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--LINM-T-EQEIQFQ-SKN.S-FK-------EG-DQ----- 964
SMM.US.86.CFU212            --IVI--------F-------Q---RQ--L------S---ITHL-----A---SQE-KMVA--TN-EHT--V-----------A--HH--TQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--LINM-T-EQEIQFQ-SKN.S-FK-------EG-DQ----- 960
SMM.US.x.H9                 --I-I-----G--F-------X---RQ--L------S---ITHL-----A---SQE-KMVA------E-T-XV-------X---A--HH--TQ-DRI-E--NSIE-I-L--IHCI----R----DMTPA--L-NM-T-EQEIQFQ-SKN.S-FK-------EG-DQ----- 960
SMM.US.x.PGM53              --I-I--------F-------Q---RQ--L------S---ITHL-----A---SQE-KMVA------E-T--V-----------A--HH--TQ-DRI-E--NSIE-I-L---HCM----R----DMTPA--L-NM-T-EQEIQFQ-SKN.S-FK-------EG-DQ----- 960
SMM.x.x.pE660.CG7G          --I-I--------F-------Q---RQ--L----------ITHL-----A---SQE-KMVA------E-T--V-----------A--HH--TQ-DRI-E--NSIE-I-L---HCM----R----DMTPA--L-NM-T-EQEIQFQ-SKN.S-FK-------EG-DQ----- 960
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----- 960
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---- 972
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-- 989
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---- 985
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-- 961
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-- 962
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---- 972
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---- 969
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---- 969
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---- 972
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---- 972
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-- 976
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-- 975
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-- 967
```

```
                                                                                p31 Integrase end
                                                                                        Pol end
H1B.FR.83.HXB2              AKLLWKGEGAVVIQDNS.DIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED................*  1003
H1O.CM.91.MVP5180           -Q-------------KG.----------------------T-SM-N--T-SESMEQPGEIP......-  1010
H1N.CM.95.YBF30             ---------------G.-----------------------G----G---NQEME............-  1015
H1P.CM.06.U14788            -T-------------KG.-------------N-------------DT-R-SESLEQSG.........-  1008
CPZ.CD.06.BF1167            -R----------KEGE.E------------K-----V--AGG--D--N-..................-  1007
CPZ.TZ.00.TAN1              -R----------KEGE.-------------KE-------AGGMDD--N-T..................-  1000
GOR.CM.07.SIVgorCP2135con   -T-------------KG.--------------------------DT-G-SESLEQSS.........-  1003
                                                                                p31 Integrase end
                                                                                        Pol end
MAC.US.x.239                GE---------ILKVGT.-------------K---GGKEV-SSSHMEDTGEAREVA...........-  1020
H2A.SN.86.ST_JSP4_27        GE-----D---IVKVGA.---II--------K---GRQEM-SGSNLEGAREDGEVA...........-  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.x.x.FAO                 -E-V---------KEGT.-L--I--------K----NVDSHTNME......................-  1013
GRV.ET.x.GRI_677_gri_1      -R-I---------KEGE.------------K---ERKTM-SEGSMEGVREANKQMEGDSDLQDQE-  1041
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_1A11           GE---------ILKVGT.-------------K---GGKEM-SSSHMEDTGEAREVA...........-  1020
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..........-  1020
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--------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.SIVsmSL92B        -E----------KVGT.E------------K---GRQEMGSSASMEDQ..................-  1006
SMM.SL.92.SL92B             -E----------KVGT.E------------K---GRQEMGSSASMEDQ..................-  1006
SMM.US.04.G078              GE---------ILKVGT.E------------K--AGGKEVGSGSHLEDPGEAREMA...........-  1016
SMM.US.05.D215              GE---------ILKVGT.---------V--VK--AGGKEV-SSSHLEDTRETGEVA...........-  1016
SMM.US.06.Ffq               GE---------I-KVGT.-------------K--AGGKEL-SGTHVEDTREAGEMA...........-  1020
SMM.US.86.CFU212            GE---------ILKVGT.------------K---GGKEV-SGSHLEDTREAGEMA...........-  1016
SMM.US.x.H9                 GE---------ILKVGT.E------------K---GGKEL-SGSHLEDTGEAREVA...........-  1016
SMM.US.x.PGM53              GE---------ILKVGT.E------------K---GGKEL-SGPHLEDTGEAREVA...........-  1016
SMM.x.x.pE660.CG7G          GE---------ILKVGT.E------------K---GGKEL-SGSHLEDTGEAREVA...........-  1016
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
```


PLV Proteins

```
                          Vif start
H1B.FR.83.HXB2            ME..................NRWQVMIVWQVDRM.RIRTWKSLVKHHMYV...SGKARGWFYRHHYESPHPRISSEVHIPL..............GDARLVITTYWGLHTGERDWHLGQGV.....SIEW........................RKKR......YSTQVD   99
H1O.CM.91.MVP5180         --..................-----L----I--Q.KVKA-N----Y-K-M...-K--AN-R-------RN-KV--A-Y--V..............AE-DI-V------MP---EE---H--.....----........................QY-E.......-K--I-   99
H1N.CM.95.YBF30           --..................------V-------.K--K-N----------...-K--K--Y------TH--K--------V..............-Q----TV-----T---QS----H--.....----........................-LRK.......-K----   99
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--I-   99
CPZ.CD.06.BF1167          --..................-----RV---I---.---A-N------I-K...-KHCKE-K-----DID---KGG-----I.............TEK-K----I-----C--KA----H-A.....----........................-QG-.......-W-E--  100
CPZ.TZ.00.TAN1            --..................-----QV--MI---.-L---T------IFT...TKCCKD-K------TDT-KRAG-I----.............TERSK--VLH----AC---P----H-I.....GL--........................-QGK.......----I-  100
GOR.CM.07.SIVgorCP2135con --..................----------IP-E.-VKA-N----Y-K-R...TKQ-KQ--------AVN--V--G-Y--V..............-P-EIIV------MP--KEEQ--H--.....G---........................-QGK.......-I--L-   99
                          Vif start
MAC.US.x.239              --EE................K--IAVPT-RIPER..LER-H--I-YLK-K...TKDLQKVC-VP-FKVGWAWWTCSRV-FPL............QEGSH-EVQG--H-TPEKG.-LSTYA-.....R-T-........................YS-N......FW-D-T  101
H2A.SN.86.ST_JSP4_27      --EG................K--IAVPT-R-PGR..MER-H--I-YLK-R...T-DLEKVC-VP-HKVGWAWWTCSRV-FPL............K-ESH-E-QA--N-TPEKG.-LSSYS-.....RLT-........................YTEK......FW-D-T  101
H2B.CI.x.EHO              --EE................KN-IAVPT-RIPCR..LER-H--I-YLK-R...TKDLQQVS-VP-HKVGWAWWTCSRV-FPL............KEG-H-EVQG--N-TPERG.FLSSYA-.....RLT-........................YERS......FY-D-T  101
H2G.CI.92.Abt96           --EE................K--IAVPT-RIPGR..LEK-H--I-FLKFK...TKDLQKAV-VP-HKVGWAWWTCSRV-FPL............TKE-H-E-QG--N-TPEKG.-LSQYA-.....RLT-........................YTRK......FY-D-T  101
H2U.CI.07.07IC_TNP3       --EE................K--I-VPT-R-PGR..MER-H----YLKHR...TKELSKVC-VP-HKVGWAWWTCSRV-FPL............E-ESH-E-QA--N-TPEKG.-LSTYA-.....R-T-........................YTRG......FWSD-T  101
H2U.FR.96.12034           --EE................KN-I-VPT-R-P-R..LER-H--I--L#-N...TKELQKVC-VP-HKVGWAWWTCSRV-FPL............KKGSH-EVQA--N-TPEKG.-LSSYA-.....RLT-........................YTRG......FW-D-T  100
ASC.UG.10.RT03            -APA................KM-VIRR--IALER.--QA-I--T-YLK-K...TKELEEVQWI---QLYAH-YTQNKA-IPLTLKVSQ....DYVEATAIEVQIL-D-SSCRSRDPP-TRD.TQALT-S-TYTRLTK.............GDREIDRIS......-E-EIT  127
ASC.UG.10.RT08            -APT................KM-VIRRI-IAPER.--QA-I--T--LK-K...TKELEEVQWI---QLYAH-YTQNKA-IPLTLKVSQ....DYVEATAIEVQIL-E-SSCRTRNPP-VRD.TQALT-S-TYTRLNK.............GDREIDRIS......-E-EIT  127
ASC.UG.10.RT11            -APA................KV-VIRRI-IALER.--QA-I--T-YLK-K...TKELEEVQWI---QLYAH-YTQNKV-IPLTLKVSQ....DYVEATAIEVQIL-E-SSCRSRDPP-VRD.TQGLT-S-TYTRLHK.............GDREIDRIS......-E-EIT  127
COL.CM.x.CGU1             --..................KM-A-R-CYPMGKR.KPKPLREWI-KQFPE...CQYYYLGKTESP.....EWWTEGKIEFYL...........GSAIK-S-IYLGT-TPDR..Y-R-PP-......-Y-L........................QIGK......WE-DLS   94
COL.UG.10.BWC01           --..................KE-S-RLCYPMSKK.KQEKLRKWI-QQFPN...CRYLYKGKTEAP.....EWWTEGRVVVPL...........GEAIK-H-IML-H-G-DR..Y-R-TA-......-Y-V........................YIGG......WCSDLT   94
COL.UG.10.BWC07           --QT................KE-S-R-CYPMG-K.NQEKLRLWI-KNFPE...CTYHYKGKTESP.....EWWTEGRIVIRL...........GDSIK-TVIML-H-S-DR..F-R-SA-......-Y-L........................QIGK......WE-DLT   96
DEB.CM.04.SIVdeb04CMPF3061 --.................KM-V-RLTHKIPMN.KLDR-C-AI-KGK-K...TKELAPITWQ--F-LRNSYFTQTRI-FPI............TKEGKV-VELF-C-APEKG.-LP-MA-.....GLS-........................EQGN......WQ-ELS  100
DEB.CM.99.CM40            --..................KT-I-RLTHKIPMN.KLDR-C-AI-REK-K...TKELDTVQWI--F-LRNSYYTQTKIVFPI............TKEGGV-VELL-C-TPEKG.-LPTMA-.....A-A-........................ET-T.......WK-ELT  100
DEB.CM.99.CM5             --..................KE-I-RLTHKIPVN.KLDK-C-AI-RGK-K...TKELQAIKWI----LRRAFYTQTKIVFPV............TEEGKV-VEIF-S-APEKG.-MPSMA-.....G-S-........................EQTG......WQ-ELT  100
DEN.CD.x.CD1_CMO580407    -QN..........LGRENRTKE-I-RLTYKIGET.KLEH-M-TL-SII-K...TNELK-TVIIP-FRITWEWYTMTKIHIPL............RA-YMEVKL--H-TPEKG.YLSQVAI.....--AYVH.......................LSTG......WA-E--  109
DRL.DE.11.D3              --..................KM-I-IPT-R-E-Y.MVER-N--C-Y-KHK..GEKYLER-Y-AP-FQ.CSGWWTHSQWTIPF............KDKSKI-V-VL-N-TPDKG.-LSKYAI.....T--H.........................IQEK......FC-FI-  100
DRL.x.x.FAO               --..................KM-I-IPT-R-E-Y.MVER-N--C-Y-KHK..GEKYLER-H-AP-FQ.CSGWWTHSQWTIPF............KDKSKI-V-VL-N-TPDKG.-LSKYAI.....T--H.........................IQEK......FC-LI-  100
GRV.ET.x.GRI_677_gri_1    --RE................KQ-I-RV--R-SER.Q-SR-RGI-TYKI.....RN-QLP-E----WQVQWQFWTYSQF-IPL............SK-DYIEVNI-HN-TPERG.-LSSH--.....GLSYY.......................HQ-G......-K-E--  101
GSN.CM.99.CN166           --RQRQ..............KM-V-QPIQIMTER.KVDWLLRAT---I.....WSGKTPFV-V---QLQ-Q-FTQNKIRLAMDTRKVG....EEVEATYIE--IL-DTTSHGPASLSRSTYYQQAVI---VYNRS.KL............GAREGDIIW.......---NLT  128
GSN.CM.99.CN71            --RQRQ..............KM-V-QPIQVMTER.KVEWLIRAT---I.....WSGKTPFV-I---QLQ-Q-FTQNKIRLALDTRQVG....EEVEATYIE--IL-DTTSHGPASLSRSTYYQQAVT---VYNRS.KL............GARETDITW.......---NLT  128
LST.CD.88.SIVlhoest447    --R.................DQRI-RVAYITT-S.LVEKFLDIRRA-K......ETQND-VSL-GTGVGWEFYTYNNIVIPL............ETG.T--VRC-AH-AA-RG.-ISQWA-.....----........................IYGN.......-Q-EI-   97
LST.CD.88.SIVlhoest485    --R.................DQRI-RVA-ITT-N.L-EKFLDIRRA-K......ETQVD-VSL-GTGVGWEFYTYNKIVIPL............EIG.T--VRC-AH-AA-RG.-ISQWA-.....----........................IYGN.......-Q-EI-   97
LST.CD.88.SIVlhoest524    --R.................DQRI-RVA-ITN-S.LVEKFLDIKRN-S......ESKND-VSL-GTGVGWEFYTYNKLLIPL............DTG.T--VRI-AH-AA-RG.-ISQWA-.....----........................IYGN.......-Q-E--   97
LST.KE.x.lho7             --R.................DQRI-RVA-ITT-S.L-EKFLDI-RLNT......EAKEN-VGL-GTGVGWEFYTYHKLVIPL............ESG.T--VRI-AH-AA-RG.-ISQWA-.....----........................VYRD.......-Q-EI-   97
MAC.US.x.251_1A11         --EE................K--IAVPT-RIPER..LER-H--I-YLK-K...TKDLQKVC-VP-FKVGWAWWTCSRV-FPL............QEGSH-EVQG--H-TPEKG.-LSTYA-.....RVT-........................YS-N......FW-D-T  101
MND-1.GA.x.MNDGB1         --R.................VERI-RLT-K-SSQ.--EK-HW--RRQ-AWATANNEEGC-WLYP-FMAYNEWYTCSKVVIII............NR-I--IVRS--H-QIEVG.CLSTYA-.....---AVVRPP.P....................FE-E......WC-EIT  110
MND-2.CM.98.CM16          --..................KQ-I-IPT-KL--H.AVER-N----Y-K-K..GEKHLDK-DLKA-FQ.CSGWWTHSQKYIPL............EE-EQII--IL-N-TPEKG.-LSTYAM.....T--Y.........................KFTN.......-Y-HI-  100
MND-2.GA.x.M14            --..................KQ-I-IPT-RI--R.QLEK-N----Y-K-K..EEKHLDK-ELF--FQ.CSGWWTHSQK-IPF............KDGSKII--AL-N-TPEKG.-LSQYAI.....TA-Y.........................N-GD.......-Y-HI-  100
MND-2.x.x.5440            --..................KQ-IAIPT-EDRQDVKL-D-N-I--Y-K-K..GEKHLEN-QLYP-FQ.CSGWWTHSQKKIPF............KDGSKII--VL-N-TPEKG.-LSQYAF.....TV-Y.........................Q-EN.......-F-YI-  101
MNE.US.x.MNE027           --EE................K--IAVPT-RIPER..LER-H--I-YLK-K...TKDLQKVC-VP-HKVGWAWWTCSRV-FPL............QEESQ-EVQG--N-TPERG.-LSTYA-.....R-T-........................YSRN......FW-D-T  101
MON.CM.99.L1_99CML1       -APR................KM-V-TPTNIISES.KMDWLIRCT-W-ILT.....GSAPFV-V---QLHNQ-F-QNKIKLPLDLGITQ...EGESWATY-E--I--D-TNVGPSALSPTVYNKQAYT---VYWMQ.KVERVPLDPVWNPNPRDREVRTW......-A-HLT  139
MON.NG.x.NG1              -APR................KM-V-SP-HIRSER.WMEWLIRAT-YNI-T.....GKVPFV-V---QLT-Q-F-QCKIKLPLEMGITQ...XXEQWATYXE--IL-DXTNVGPASLSPSVKWKQAYI---VYWRQPRVPQQEGVVDRVWFPSPR-PD-EIDRVW-A-YLT  145
MUS-1.CM.01.CM1239        -APR................KM-V-SPIQETTER.KMDWIIRAT---IFS.....GKTPFV-V---QVQ-Q-F-QNKIKLALTQNTLE...SGETEVTFIE--VL-DVTNVGPASLSKSTYWKQSYIL--VYMRKGP.............RDRE-DYVW......-H-YLT  127
MUS-1.CM.01.SIVmus_01CM1085 -APR..............KM-V-SPIQETSER.K-DWIIRAT-YYILS.....GRAAFT-V---QLY-Q-F-QSKIRIPLTQNTRE...SGAVESTYIEL-VL-DVTNVGPASCSKSTWWKQSFIL--VYLRRHP.............QDREQDYIW.......---YLS  127
MUS-2.CM.01.CM1246        -APG................KM-I-SPIQYMSER.KMDWLIRCT--YIFS.....GKAPFT-V---QLQ-Q-FTQNKIKIPLSINRIE...GGPTETTYIE--ILLDVTNVGPASLSRSTYWQQSYILK-RYIRAAP.............QNREIDMVH......-E-FL-  127
MUS-2.CM.01.CM2500        -AQR................KM-I-TPIQFMTER.-MDWLIRCT---ILS.....GKAPFS-I---QVQ-Q-FTQNQIKIPLTIKKRE...EDKVESTYIQ--IL-DVTNVGPASLSKSTYWQQSYIL--VYVRAAP.............QDREVDMVW.......---YL-  127
MUS-3.GA.09.09GabOI81     .MPR................KM-V-SPIQETSER.KMDWLIRAT-YLIFS.....GXAP-T-V-A-QLQSQ-F-QNKIKIPLAIQLRDIEGQQWIEATHIE--VL-DVTNVGYASLSRSTQWRQTYIL-FVYQRRQM.............ESRDIDVIW......-H-YLT  129
MUS-3.GA.11.11GabPts02    -APR................KM-V-SPIQVTTER.HLDWLIRCT-Y-ILS.....GQTP-S-V-A-QLQ-Q-F-QNKIRIPLTQETTE....REKVSTHIE--IL-DVTNVGPASLSKSTYYEQSYIL--VYKRNRK.............GGRETDTVW......-H-HLT  126
OLC.CI.97.97CI12          --....................-IAVPMYRLNNG.QVEHV-FMA-LAYK.....FWGYH-IWQNC-KY-REWWTHGKIYIPV............P-G.EIE--L--HIRH-KG.LQNSG--.....NFKY........................IVGI......SWMYLT   95
RCM.CM.00.SIVagi_00CM312  -AE.................KM-I-RPI-R---R.K-EQ-H----Y-Q-K..GKKE-KE-E-VP-FKV-WGWW-HSEIHVPL............GKNTKVK-----N-TAEKG.-LSTY--.....G-AY........................IDNT.CDPC.-F-DI-  106
RCM.CM.02.SIVrcm_02CM8081 -AE.................KM-I-RP--R---R.K-EQ-H----Y---K..GKKE--E-E-VP-FKV-WGWW-HSEVHIPL............GNNTKIKV----N-T-EKG.-LGTY-A.....ALAY........................IDQK.CDPP.-F-DI-  106
RCM.GA.x.SIVRCMGAB1       -AE.................KM-I-RPI-R---R.K-EQ-H----Y-Q-K..GKKA-KE-E-VP-FKV-WGWW-HSEVHIPL............EEGSK-K-----N-TVEKG.-LGTY--.....G-L-........................I-GD.......-V-D-F  102
RCM.NG.x.NG411            -AE.................K--IERPI-K---R.KVEQ-H----Y-Q-K..GKKE-KD-E-VP-FKV-WGWW-RSEVHIPL............GNNTKIKV----S-TIEKG.-LGTY--.....G-AY........................IDN-.CDPC.-F-DI-  106
SAB.SN.x.SAB1             --..................KH-I-RPL-K-TGG.QQER-T----Y--H-...-KQCVH-R-TP-TKIRWNWY-YQEWVIPL............KDG-LIKV-N--H-TPEKG.-LETYAT.....G-GY........................S-GE......WF-EL-  100
SMM.CI.79.SIVsmCI2        --EE................KS-VTVPT-RIP-R..MER-H--I--LKFN...TKELQKVC-VP-NKVGWAWWTCSRV-FPL............EKESH-EVQG--N-TPEKG.-LSSYA-.....R-T-........................YSRN......FW-D-T  101
SMM.LR.89.SIVsmLIB1       --EE................K--I-VPT-KIPGR..LER-H--I--LK-N...TKELNKAC-VP-HKVGWAWWTCSRV-FPL............Q-E-Y-EVQG--N-TPEKG.-LSTYA-.....RLT-........................YSRN......FW-D-T  101
SMM.SL.92.SIVsmSL92A      --EE................K--IAVPT-KIPHR..LERRH----YLK-K...TEDLKKVV-VP-HKVGWAWWTCSRV-FPL............EKE-H-EVQG--N-TPEKG.-LSCYA-.....RLT-........................YTRN......FW-D-T  101
SMM.SL.92.SIVsmSL92B      --ED................K--VAVP--RIS-R.IVER-H-CI-F-K-K...TRELEKAC-VP-HKVGWAWYTASRV-FPL............EEGSH-EVQV--N-TPEKG.-LSSYA-.....R-T-........................YSEK......FW-D-T  102
SMM.SL.92.SL92B           --EG................K--VAVP--RIS-R.IVER-H-CI-F-K-K...TRELEKAC-VP-HKVGWAWYTASRV-FPL............EEGSH-EVQV--N-TPEKG.-LSSYA-.....R-T-........................YSEK......FW-D-T  102
SMM.US.04.G078            -QEE................K--V-VPT-RIPGR..LER-H--I--LK-N...TKELQKVC-VP-HKVGWAWWTCSRV-FPL............Q-ETH-E-QG--N-TPEKG.-LSTYA-.....R-T-........................YTRG......FW-D-T  101
SMM.US.05.D215            -QEE................K--I-VPT-RIPGR..LEK-H--I--LKFN...TKDLQKAC-VP-HKVGWAWWTCSRV-FPL............EKESH-EVQG--N-TPEKG.-LSTYA-.....R-T-........................YTRN......FW-D-T  101
SMM.US.06.FTq             -QEE................KN-I-VPT-RIPGR..LER-H--I--LK-N...TKELQKAC-VP-HKVGWAWWTCSRV-FPL............E-ESH-EVQG--N-TPEKG.-LSTYA-.....R-T-........................YTR-......FWSD-T  101
SMM.US.86.CFU212          --EE................K--I-VPT-RIPGR..LEK-H--I--LK-N...TKDLQQAC-VP-HKVGWAWWTCSRV-FPL............Q-ESH-EVQG--N-TPEKG.-LSTYA-.....R-T-........................YSRN......FW-D-T  101
SMM.US.x.H9               --EE................KN-I-VPT-RIPGR..LEK-H--I--LK-N...TKDLQKAC-VP-HKVGWAWWTCSRV-FPL............RDE-H-EVQG--N-TPEXG.-LSTYA-.....R-T-........................YSRN......FW-DAT  101
SMM.US.x.PGM53            --EE................KN-I-VPT-KIPGR..LEK-H--I-YLKFR...TKDLQKAC-VP-HKVGWAWWTCSRV-FPL............RDESH-EVQG--N-APEKG.-LSTYA-.....R-T-........................YSRK......FW-D-T  101
SMM.US.x.pE660.CG7G       --EE................KN-I-VPT-RIPER..LER-H--I--LK-N...TKDLQMAC-VP-HKVGWAWWTCSRV-FPL............RDETH-EVQG--N-APEKG.-LSTHA-.....R-T-........................YSRN......FW-D-T  101
STM.US.89.STM_37_16       --EE................K--I-VPT-RIPGR..LER-H--I--LK-N...TKELSKAC-VP-HKVGWAWWTCSRV-FPL............Q-E-H-EVQG--N-TPEKG.-LSEYA-.....R-T-........................YTRN......FWSD-T  101
SUN.GA.98.L14             --ENPPQWRWWDEREWEDRRQYKI-R---LI--I.AVEKFLD-RRM-R......ETKDD-VSMYGTGTGWEWYTYNKI-IPV............TYG.TV-VRI-GH-TPAKG.LINQW-C.....-M--........................IYNS.......-Q-EI-  114
SYK.KE.x.KE51             --..................KE-I-VLT-KLHPN.Q-DKMQHIL-V-K-K...-KQLEKAT-K---QITWEWWTRAQWEIP............V-QGV-I-KF-HN-TPEKG.LLRTE--.....GLI-........................AHPSG.....W--E-T  100
SYK.KE.x.SYK173_COMGNM    --..................KE-I-VPT-RMTPR.Q-DRLQHII-T-K-K...-KELEKAT-K---QIEWQWYTYCQWTIP............V--GTIW--F-HN-APERG.-LHM--I.....R-QY........................QWNQ......WN-DLT   99
TAL.CM.00.266             -D..................KL-VTRLT-K-P-N.-MQR-VGTIRF-IHK...TKYLQKLIWV--FQLR-ASYTQAKW-LPL.........EEKDTRVEY-T--Q--C-APDHGYTSDWDTI.....----SV...................SDIV-DEKV......-A-E-T  112
TAL.CM.01.8023            -D..................KL-VTRLT-K-P-N.-MQR-VGTIRY---K...TKYLQKLIWV--FQLRMASFTQAKW-LPL.........EEKENRVEYIT--Q--C-APDKGYTSDWDTI.....----SV...................SDVTLDAKV......---ELT  112
TAN.UG.x.SIVagmTAN1       --RE................KL-VTRLT-R-SGE.H-DK--GI--Y--.....RNRLQD-T-LM--QCGWAWYTCSRFLIPL............G-EGKI-VDC--H-TPEQG.-LSTYA-.....A-SF.EN......................WQNT.......-K-E-T  102
VER.DE.x.AGM3             -NQE................KE-VMRVT-K-PEE.L-TK-QGI-RYW-.....RTRKLD-K--M--QITWAWYTMSRYEIPL............GQHGSIHVDL--H-TPEKG.-LSTYAE.....G-QYLSN......................-DPW.......-R-EL-  103
VER.KE.x.9063             -NQE................KE-VMRVT-K-SGD.L-TK-QGI--YW-.....TKRNLK-Q-MM--QINWAWYTMCRY-IPI............GLEGEIQ-DC--H-TPERG.-LSTYAI.....G-QYLG.......................AVGN......FR-EL-  102
VER.KE.x.AGM155           -SQE................KH-VMRLT-K-QEE.V-TK-QGI-RYW-.....NKRNLK-E-KM--QITWAWYTMSRYVIPL............P-SGEIHVDI--H-APKQG.-LSTYA-.....G-QYVSL......................VNDK.......-R-EL-  103
VER.KE.x.TYO1_patent      -NPN................KE-VMRVT-K-PGD.L-TK-QGI-RYW-.....RQRNLK-N-YM--QITWAWYTMSRYVIPI............GKHGEICVDL--H-TPEQG.-LSTYA-.....G-QYVSN......................LESK.......-R-EL-  103
WRC.CI.97.97CI14          --..................KE-IGVPLIRISKG.KF-R-TYYIANL......IKAGQAN-LWGSWHAAAECFTRHIVKIPI............Q-HWEMQ-RC--N-SPEKG.-MDTM-F.....N-RM........................VYEN-......WE-FIM   98
WRC.CI.98.98CI04          --..................KE-IGVPLIKISKG.-FKR-AYWIA-L......MQNKKAS-LWGTWHTQAECFTRYLLRIPL............Q-HWE-Q-RC--N-APEKG.-LDTM-F.....N-RM........................VYEN-......WE-FIM   98
WRC.GM.05.Pbt_05GM_X02    --..................KE-IAVPLIKISKG.-F-R-SYWIAEL......IQEKQAR-LFGSWHVHAECFTRHIVRIPL............Q-HWEMQ-RC--N-APEIG.-IDTL-F.....N-RM........................VWENK......WE-FIT   98
```

```
                                                                                                                                             Vif end
H1B.FR.83.HXB2              PELADQLIHLYYFDCFSDSAIRKALLGHIVSPRC...EYQAGH......NKVGSLQYLALAALITPKKIKP.......PLPSVTKLTEDRWNKPQKTKGHRGSH....TMNGH.......................................*  192
H1O.CM.91.MVP5180           --T--RM---H--T--TE------I--QR-LTK-...-L--.......SQ--T--F---K-VVKV-RN--.............-----QR---------W-IRDQL---....S----......................................-  193
H1N.CM.95.YBF30             --M--K----H-----TA----Q-V--RP-L---...-P--.......KQ--T------T-WVGA--R--.....................-EH--MQ----NP....I----......................................-  193
H1P.CM.06.U14788            --T--R----H--Q------V-R-I--DR-LNX-...--S--......SQ---------KVVVGKV-R--.............-----QI--Q-I-SN--R---XQE--....S-S-CYKP..................................-  196
CPZ.CD.06.BF1167            --T---I--SR--P--TER---Q-I--EKILQH-...HFPT--.....S------F--FKKVLEQN-Q--HRP..........-V-A--------RI-A-QENR....GI--Y.......................................-  197
CPZ.TZ.00.TAN1              --T------TR--T--AAG-V-Q-I--ERILTF-...HF-S--.....RQ--T--F--FRKVVESQDKQ-KGPRR..---------------HRT-T-R-EN-....-LS-C.......................................-  199
GOR.CM.07.SIVgorCP2135con   --T--RI------Q--TE--V---IV-EKLLCK-...N-P---.....SQ--T------R--VGKVRVR-.......-F---Q--AA-T-SRS-QMR--KEN-....ST---.......................................-  193
                                                                                                                                             Vif end
MAC.US.x.239                -NY--I-L-ST--P--TAGEV-R-IR-EQLLSC-...RFPRA-K.....YQ-P-------KVV...............SDVRSQGENPT-KQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA.......................-  215
H2A.SN.86.ST_JSP4_27        -DC--S---ST--S--TAGEV-R-IR-EKLLSC-...N-PQA-K.....YQ-P---F---VVVQ..............QNGRPQRDNTTRKQWRRNYRRGLRVARQDGRSHKQRGSEPPAPRAYFPGVAKVLEILA......................-  216
H2B.CI.x.EHO                -DV--R-L-GS--SS-TANEV-R-IR-EKILSH-...N-PSA-T.....GQ-P---F---RVVQ..............EGKDGSQGESTTRKQRRRNSRRGIRMARDNIRTSQQSSSQSLAQGTYFPGLAEVLGILA......................-  217
H2G.CI.92.Abt96             --T----L-GS-----TAGEV-R-IR-EQILSC-...N-PTA-K.....RQ-P---F---QVV...............QKGHGS-GESXTRKQRRGDNRRGIRMARKNSXRDQQDSSQSFTQGAYFPGLAKVLGILA......................-  216
H2U.CI.07.07IC_TNP3         -DY----L-GT--S-YTAGEV-R-IR-EQLLSC-...NFPSA-K.....GQ-P---F---KIV...............QDVRPQRKNTT-KLRRRNS-RSLRMARVNSHSHKPRGSKSSAQGTYFPGVAKVLGILA......................-  215
H2U.FR.96.12034             -DC----L-GS--S--TAGEV-R-IR-EQLLSC-...NFPSA-R.....QQ-P---F---R-................QDGRS-RENPTRKFRRRNCRRGFQLARKNCERYQQGSSATSSQGTYFPGLAKVLAILA......................-  214
ASC.UG.10.RT03              -H---AM--VEH-S--TAQDFTR-IR-QQLLGE-...HHPHYLGKEI.....PT------R-I............QDGRSTAITPNTLAKQNS..GDVPKTNPGNTMGAKRRGRKTLQPRRTPWSLAALPLPPCRQ...............-  242
ASC.UG.10.RT08              -Q---AMV-VEH-S--TAQDFTR-IR-QQLLGE-...HHPHYLGKEI.....PT------R-I............QDGRNPAITP-TLVKQN..RNVPKTNPGNTMGAKRRGRKTFQPRRTPWSLAALPLPPCGH...............-  241
ASC.UG.10.RT11              -Q---AMV-VEH-S--TAQDFNR-IR-QQLLGE-...HHPHYLGKEV.....PT------R-I............QDGRNPAITP-TLARQNS..RDVPKTNSGNTMGVKRRGRKTLQPRRTPWSLAALPLPPCRQ...............-  242
COL.CM.x.CGU1               WQQ-VGYM-IIRGK-VRIEEEARKAIRGLPWNP-...DF-V--L......G-LK--E-C-L--...............REKERTHQKMRVLKG......TSH-CRTR.........................................-  172
COL.UG.10.BWC01             WQM-TGYL-IIRGK-IEIEKEVRKAIRGLPWDP-...DF-V--L......G-R---E-C-SK-...............QKDQEREYQKMRRWKE......TST-CKSR.........................................-  173
COL.UG.10.BWC07             WQ--VGYM-IIRGK-VEIQEEARRA-KGLPWCP-...DF-V--L......G-LT--E-C-N-I...............DENKPLR-RPEKG......ITP-YRSR.........................................-  172
DEB.CM.04.SIVdeb04CMPF3061  ----EH----R--P--TQQ-V-Q-I--EKLLTC-....Q-YHRGAGG..-GPP----IC-RVA-K..........YGKATAL-PTNALVKASSNFQASATIHDAGCKR..RSHETLSKGRINWRLELLCRVTSRARVDRRTGLGSLSNRLHKEGPENDLATSGRRVLS-  251
DEB.CM.99.CM40              -D---H----R--P---QQ-V-Q-IR-E-LLQQ-....Q-YHRGPEG..KGPP----I--RVAVK..........HGKV-TL-PT-FHLK-NTHDKARVAAG-ARSKR..-SPKTFHQRRTIWSLESLCRLTRPTRLDGRTSMGCIGH.....................-  231
DEB.CM.99.CM5               -D---H----R--P---QQ-VTQ--K-EKLLQQ-....Q-YHRGPEG..KGPP-------RTAVK..........YGKVSTL-PT-FHIK-STNDKASTTT--AGSKR..-SFKTLHKRRVMGSLESLCRPSCPTRLDRRTSMGGVSN.....................-  231
DEN.CD.x.CD1_CMO580407      -VT--EI--KDH-S--TRQ-V-Q--R-EV-LST-ARKKAHYHPRTGPG.REPP----IV-Q-A-R..........YGKVTTI-PTSISFK-GRNHS-AAT--SMGYHR..-GKKTLQQRRDHWNMGSLLVITSAPTLDRRTSHGSRRHRFHKNHAERNMEAL......-  259
DRL.DE.11.D3                -TT--RI---E--P--T-Q---GV---YRIQA.-...YFPR--K.....GQ---------Q-H-K........YLRHGRRTASGESSSGASGSRRGRVRRVAAEESLKNQP-GSITLPPRVPLPSLEHLHGALP............................-  218
DRL.x.x.FAO                 -TT--RI---E--P--T-Q---GV---YRIQA.-...YFPR--K.....GQ---------Q-H-K........YLKHGRKTVSGESSSGANGSRRGRVRRMATEESLKNQP-GSITLPPRVPLPSLEHLHGALP............................-  218
GRV.ET.x.GRI_677_gri_1      -GT--RM------N--T-R--QQ-IR-EKYTW.-...TFKE--K.....GQ-Q---L---V-YTNGI......RKRSKRTFTRMAGNLGSRQGA...M-RMATRHAQGSKRRSQKALWNEHANPSMELLCRGGKET......................-  220
GSN.CM.99.CN166             -GV-M-I--GK--P--QNED--R-IR-EQILGN-...-HPEA-.LQV..GTPQT-EQ--FF-Y..........VKHYGKNPTQSSYALAQP-G..AHN-GYRSKRSVGNKRGGRETLLKGRAPWNLA...........................-  238
GSN.CM.99.CN71              -DV-M-I--TR--P--QNEDV-R-IR-EQLLGY-...DHPEA-.PQV..GTPHT-ER--FF-Y..........VRYYGKNPTKSPSTLDQP-G..AHN-GSCSKRHVGNKRRGRKTLLKGRAPWNLA...........................-  238
LST.CD.88.SIVlhoest447      -I--N-M--CK--N---TRD-----W-EKIMAF-...AMPT--K.....GC-L---F-C-RQ-QHVQNKAA.......KESRK-SRGLWATKGA....MGSMASRHHGGNKRGGQTPFPRNNFGPSLRILCQQYRVRGESLHEIHYHSK............-  228
LST.CD.88.SIVlhoest485      -I----M--CK--N---TRD-----W-EKIMAI-...AMPT--K.....GC-L---F-C-RQ-QHVQSKAA.......KESRK-SRGLWTTKRA....MGSMVSRHHGGNK-GGQTPFPRSHFGPSLGILCQQYRVRGGSLHEIHYHSK............-  228
LST.CD.88.SIVlhoest524      -I----ML-VK----WVHRD-VR-IR-EKILHS-...KMPT--K.....GC-L-----C-RQ-QHVQTKTA.......AQ-RKAPRGLWSSKRA....MGSMASRHHGGNQ-RGKTPFPRSNFGPSLGILCQQYRVRRGGLHEIHHFS............-  227
LST.KE.x.lho7               -I----M--YK--N--IQRD--R----ERITMC-....QPL--K.....GC-L---FIC-RQ-QHVQAEAA.......KKSRE-TRGLWTTKGA....MGSVVSRHHGGNKRGGKTPFPRSNVGPSLAILCRQYRVRRGGLHEIHHPSQ...........-  227
MAC.US.x.251_1A11           -NY--I-L-ST--P--TAGEV-R-IR-EQLLSC-...RFPRA-K.....YQ-P-------KVV...............SDVRSQGENPT-KQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA.......................-  215
MND-1.GA.x.MNDGB1           --V--H----HFY---M---VM--IR-EE-LKV-...RFP---KAQ...G-L---F-C-RVI...............YG-EER*ASIR-SRTS-RAIQSVA-RYYGGNKGRSKKALPSHYPKCSIRILCAKHRE.......................-  223
MND-2.CM.98.CM16            -QT--RM--WK-LP--TEQ---Q----KRLTV.-...YFHW--K.....G-----------LSY...............TAYCNNGRRGPRDPRRSRR-GSKYLVGKISGENQSGGKITLPPRVPFPSLERMHRTLA....................-  213
MND-2.GA.x.M14              -VT--RM--WE--P--TAA-V--V-Y-ERIVA.-...YSPW--K.....GQ--T--L---R-Y...............IKFCRNGRKSTRGTTGGRRSGTRTVAGKIIGTSEQRGSITLPPRVPFPSLEHLCRTLA....................-  213
MND-2.x.x.5440              -VT--RI--GE--P--T-Q------F-ERLVA.-...YFPW--R.....GQ--T--F---Q-Y...............-RGRKHGR-GTRGARRSRR-RT-AMAGNV-GENQPGGPVTLPPRVPFPSLEHMCRTLA....................-  214
MNE.US.x.MNE027             -DY--I-L-ST--P--TAGEV-R-IR-EQLLSC-...RFPRA-K.....-Q-P-------RVV...............SYVRSQRENPTWKQWRRDNRRSLRMAKQNSRGDKQRGSKPPTKGADFPGLAKVLGILA.......................-  215
MON.CM.99.L1_99CML1         -D--N-IV-TH--A--QQLDV-R-IR-EKLLGQ-...QHL-T-YPKV..GTPLT-EK--FL-............LVRYGATPTKSSAPLDLPDG..SHRKT-PTGRHLGAKRRSPKTLLKGRAPWHLGTGNGAAC.................-  256
MON.NG.x.NG1                -DV---I--TR--S--QA-DV-R-IR-EAXLGN-...HHIQT-YNKV..GTPPT-EK--FL-............CFKYGKCTTEPSSTLDLPRG..SHRA-R-AGRHMGPKRGGPFALLKGRAPWHLGSGH.....................-  258
MUS-1.CM.01.CM1239          --I-MKI--THH-S--YSQD-PRVIR-QPPLGD-...-HPGA-.TKV..GPPPT--V---LT............LKKHGKSTTQPQ-SMALQSG..PNY-AAGPTGHVGPKRRGRKTLFQRRAPWNLE.........................-  236
MUS-1.CM.01.SIVmus_01CM1085 -T--MRI--TQ--S--YSQD-PRVIR-QHTLGD-...-HPDA-.TKV..GPPP---V---L-............LKKHGKLAPQPPSALALQSG..PNH-APGTRSHVGSKRRGRKTLFQRRAPWNLE.........................-  236
MUS-2.CM.01.CM1246          --VTM-I--TH--S--QQRD-QR-IR-EQLLRK-...-HIKT-YPKV..GTPL--EK--FF-Y..........IKHTNGASASKSPSLMAKSSN..ANNYGSGASRPVGTKRRGRKTLLQRQAPWNLG.........................-  239
MUS-2.CM.01.CM2500          ----M-I--T-H-S--QEQD-IR--R-ERILGI-...QH-TA-LPKV..GTPP--EK--FF-Y..........IKQGDGATPSKSPSPMAKPPG..-NN-GAGTKRHVGSKRRGRKALLQGRAPWDLGPSHGAPSRP................-  248
MUS-3.GA.09.09GabOI81       NX--M-I--C---P--QEED--R-IR-EQLLGH-...-HPKA-FPKV..GTPL---Q--FFQY...........VKSQRKNPTVSPSTLALQAS..ANH-ATGTRSHVGGKRRGRKTLLKGRAPWNLG.........................-  240
MUS-3.GA.11.11GabPts02      VG--MRM--YK--S--QEED--R-IR-EQLLGH-...QH-QA-FSKV..GTPLT-EK--FL-Y..........IKHYGKSTPLTYSPMALQAC..SHN-ATGSGGHVGGKRRGRKTLLKRRAPWNLERRNGATR...................-  244
OLC.CI.97.97CI12            --YV-FMV--V--A-WTQ------I--MPDHRY-...QLET--L.....G-L---L---R-V...............RNGRKSSRGSGATQGA....TGGLGERHNGRNYCRGKKALSQGVIGPSFSILCSKTYQ......................-  204
RCM.CM.00.SIVagi_00CM312    -I---KM--NV--P--T-Q--------EK-LV.-...GF-R--R.....DQ--T------Q-WVK........SQLQKNG...R-SSRGPHWGGRSRVSTMVGQNVVRDQPRSQITLPWRVHFPSVAYLCGILA.........................-  221
RCM.CM.02.SIVrcm_02CM8081   -IV--S---KI--P--T-K---Q-I--EK-LL.-...GF-R--R.....DQ--T-----IQ-WAR........EQVKKHG...R-SARGPHWGWRSRVPALVGQNAVRNKS-SQVTLPSRVHFPSLAYLCGTLA.........................-  221
RCM.GA.x.SIVRCMGAB1         -WT--S---KI--P--T-R---QS---EK-LV.-...AF-G--R.....DQ--T--F--IQ-WAK........SQLDRYG...R-SPRGPHWGWRSRVPALATGHARKGQL-SQVTLSSRVHFPSVAHLCGTLA.........................-  217
RCM.NG.x.NG411              -TV--K-V-NQ--P--T-K---Q-I--EQ-LL.-...GFEK--R.....DQ--------IK-WAR........QQLKDGRG.QR-SARGPHWGWRCRVCPLA-QNANRSQLRSQVALSSRVHFPSVAYLCGTLA.........................-  223
SAB.SN.x.SAB1               -WT--HI--WS--P--T-R-VQQ-IR-EKYLW.-...KH-V--QPT...GQ-P-------RVYTNGL...............RRVAPTSRRGSSQGSPQESQRRDTRMARNMGFAQRAVRRMAPRHVT.GPQFRGPVPLPKESPFPSLVEYCGRTSH...........-  235
SMM.CI.79.SIVsmCI2          -DV------ST--S-YTQGEV-R-IR-EQILSC-...K-PSA-K.....RQ-P---F---RI-H..............KQNGKC-RENPTR-QWGRNNRRGFRVAIQHSGSSKQGSSKTSAQGADFPGLAKVLGILA......................-  217
SMM.LR.89.SIVsmLIB1         -DY--V-L-GT--P---EGEV-R-IR-EQLLSC-...RFPKA-K.....YQ-P---F---KVV...............EHVRSQRENPT-KQWRRNYRRGLRVARKNSGRNKQTGSEPSTKGIDFPGLAKVLGILA......................-  215
SMM.SL.92.SIVsmSL92A        -D---I---GA--S-YTQGEV-R-IR-EQILSC-...NFPSA-Q.....Y--S---F---R--Q................DGS-R-DPSREQWGRDNRRGLRLARKNSRRSQQNSSEPFAKGAHFPGLAKVLGILA......................-  214
SMM.SL.92.SIVsmSL92B        -DV----V-ST--P--AAH-V-Q-IR-EQ-LSY-...G-AVA-H.....SS-Q---L---KVVL..............QNDRP-GKNPTRKQWRRNNRRGLRMVTQYSGGPESNSSTTPTKRVNFPGLEKVLGILGR....................-  218
SMM.SL.92.SL92B             -DV----A-ST--P--AAH-V-Q-IR-EQ-LSY-...G-AVA-H.....SS-Q---L---KVVL..............QNDRP-GKNPTRKQWRRNNRRGLRMVTQYSGGPESNSSTTPTKRVNFPGLEKVLGILGR....................-  218
SMM.US.04.G078              -DY--V-L-GT--T---EGEV-R-IR-EQLLSC-...RFPKA-K.....-Q-P-------KVV...............SYVRSQRENPT-KQWRRDNRRGLRVAAQNGSRDKQGGCKPSTKGTDFPGLAKVLGILV......................-  215
SMM.US.05.D215              -DC--V-L-NT--P---AGEV-R-IR-EQLLSC-...RFPKA-T.....-Q-P-------KVV...............EHVKSQRENPTRKQWRRDNWGSFRMARKNSGRHKQSSSKSSTKRADFPGLAKVLGILA......................-  215
SMM.US.06.FTq               -DY--T-L-GT--P--TAGEV-R-IR-EQLLSS-...RFPKA-K.....-Q-P-------RIV...............SHVRSQRENPA-KQWRRNNRRGLRMAREDSRRYKQGSSKPPTKGTDFPGLAKVLGILA......................-  215
SMM.US.86.CFU212            -DY--T-L-GS--S--TAGEV-R-IR-EQLLSC-...KFPRA-R.....-Q-P-------SVV...............SHGRSQGEDPTRKQWRRNNRRSLRVAEQNGGRTKQGSSKSPAKRANFPGLAKVLGILA......................-  215
SMM.US.x.H9                 -DY--T-L-GT--P---EGEV-R-IR-EKLLSC-...KFPKA-K.....-Q-P-------TVV...............SHVXSQGEXPT-KQWRRNXRRGXRLATQNXRRNKQGSSESFAXGANFPGLAKXLGILA......................-  215
SMM.US.x.PGM53              -DY--T-L-GT--P---EGEV-R-IR-EKLLSC-...KFPKA-K.....-Q-P-------TVV...............SHVRSQREDPT-KQWRRNNRRGIRMAKQNSRRNKQGSSESLAEGANFPGLAKVLGILA......................-  215
SMM.US.x.pE660.CG7G         -DY--T-L-ST--P---EGEV-R-IR-EKLLSC-...KFPKA-K.....-Q-P-------TVV...............SHVRSQGEDPT-KQWGRNNRRGLRMAKQNSRRNKQGSSKSPAEGANFPGLAKVLGILA......................-  215
STM.US.89.STM_37_16         -DC----L-GT--P--TAGEV-R-IR-EKLLSC-...RFTKA-K.....-Q-P-------KVV...............EHVRSQRENTARKQWRRGNRGSIRVATQNGRGHKPRGSKPSTEGTDFPGLAKVLGILA......................-  215
SUN.GA.98.L14               -LV---M--FK----WTSRC--R-M--EKILHE-...RN-VA-K.....GL-L---F-C-RV-.HGQQERA.......SRTATRRRG..ATERA....ICTMAGRYHGRNQ-RSGKAFPSSNSCPSLAILCGLYRVRKGRVYESNHFAQ...........-  242
SYK.KE.x.KE51               -QT--A-V-SQ--P----R-VQQ--R-EKLTSH-..WNFHKE........Q-L-------QKYLSKDGTGFLQSLPAAARGTTILHSKKCQLGR...GRSNKC-CQSR-GSPRSMQAFYSGRNIWSLESMLRGGSRSRRNTHHGLD...........-  234
SYK.KE.x.SYK173_COMGNM      -AV--R---NF--P--TAR-VNQ-VR-ELLTSH-..WTPHTD........Q-P-------QVY.LKDGGGFLQSLPACARNTMVLHSKKCRVD-....KRDQC-CKGR-GSDRSIQAFYSSRNIWSLESILKRRGRD....................-  222
TAL.CM.00.266               -NV--H---TK--P--TA--V-Q-IR-EVLLAV-.....GVR-KGR...DTPP---F---T-...........AVRYGKVATQSSSTADLQTH..SNTKASARKAHERSSRRNAYTLHKRRGHWSLESLCRIASRP.....................-  228
TAL.CM.01.8023              --V--H---TK--P--TA--V-Q-IR-EVLLAV-.....EVR-KGR...DSPP---F---KV...........AVRHGKVATQSSSANNIKAH..-NTKSSIRKTHERSK-RDAHTLHKRRGHWSLESLCRVTSRA.....................-  228
TAN.UG.x.SIVagmTAN1         -DV--HM--CH--P--T-R--QQ-IR-ESFLW.-...T-KE--VAENHWGQ-R---F---TVYTDFLRNGR..RKRFQGKKTRMVRNLGSQQGA...V-RMIKRHGSR-QS-STTPFWERTPLPSMELLSGRRGKEWGTNDRKGL.............-  239
VER.DE.x.AGM3               -AT--S---TH--T--TER--------QRFTF.-...QFPE--KKT...GQ-P-------L-HQNGL......RQRSQRSKTGGTRNMGFEQGA...V-RMAKRHARRYQS-SQDAFWARAPVPSMELLSGGGRKESHSHARKGL.............-  233
VER.KE.x.9063               -AT--GM---H--S--TER--Q--IR-ERFVF.-...QFPE--KTT...G--P-------L-HQNGI......RERSQRKQARRSRDLGSKQRT...M-RMAKGHVRRSQS-SQITFWQRAPLPSLELLSGGRRETGSPNDGEGL.............-  232
VER.KE.x.AGM155             -NT--SM--CH--T--T-R--QQ--R-NRFIF.-...QFPG--KLT...GQ-P-------L-HQNGL......RKRSQRGETRRTRNLGSQQGA...V-RMAQRYGRRNQQRSQTAFWPRTPIPSMELLSGGRGETGKTHSGKGI..............-  233
VER.KE.x.TYO1_patent        -AT--SI--GH--N--KER--QQ--R--RFVF.-...QFPE--KST...GQ-P-------L-HQNGL......RERSKRGKTRRSRNLGSKQGA...V-RMAKRYVTRSQP-GEAAFWERTPVPSMELLSGGRRKTWYSHDGKGLQIL...........-  236
WRC.CI.97.97CI14            -SE-NKI--SQH-E-WTQNVV-NTIRTGKINVID....ACQY-KGHA..G.-LT--T-C-QV-QH.........EQL.QRAGAFRR-STNSRG-..........SRYARALCNGKTAFPKGSPCKSISLLLSTERV............................-  208
WRC.CI.98.98CI04            -SE--VM--SQHV--WTQNVV-DTIRSGRLDI-N....ACQY-KGHA..G.-LT--T-C-QV-QH.........EQL.QGT-TYRR-STNSRG-..........SRYARTLCNGKTAFPKGSPCKSISLLLATEKL............................-  208
WRC.GM.05.Pbt_05GM_X02      VK---V---SQH-E-WTQ--VKNTIRQGYLH-ID....SCKY-QGHL..G.-LT--A-C-Q--QH.........EQLRKRKAPNRR-PGNTRGS..........SRYARAVC-GKTAFPPGSPCKSISLLLATEKL............................-  209
```

```
                             Vpx start                                                                                                    Vpx end
MAC.US.x.239                 MSD..PRERIP..PGNSGEETIGEAFEWLNRTVEEINREAVNHLPRELIFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGW.RPGPPPPPPPG.LA*   112
H2A.SN.86.ST_JSP4_27         -AG..---T--..----------------D----A----------------------------R---------I--T-------M---M-I-S-R--T---G---P---.-S---------.-V-   113
H2B.CI.x.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-----PG-V-   116
DRL.x.x.FAO                  -AERQSV--A-AE-MGA--...V-LE---Q-SLLR--Q--RL-FHP-FL-RL-NTCM-HY--ALQL-FT-S-----L-L---M---FQQ--S--QGR-PPPLRPAGDRL-----PP..-   114
MAC.US.x.251_1A11            ---..------..----------------------------------------------------------------------------------------.-----------.---   113
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-ERITIL--.M.-   105
RCM.NG.x.NG411               -AE..G---V-..EAPT-AGDVEF-PWLHRMLT-VNLEA.RL-FHP-F--RL-RTCV-H---RL-R-LE-AG----L-M-----I--QS--SQR...--Q-QAREA-ERIQIL--.M.-   109
SMM.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.SIVsmSL92B         -T-..------..----------------HN---AL-QT--Q-------S---R-C----V----Y----A----VQ-M---M-Q-FR---T-R----SQ---.-T---------.---   113
SMM.SL.92.SL92B              -T-..------..----------------HN---AL-QT--Q-----------R-C----V----Y----A----VQ-M---M-Q-FR---T-R----SQ---.-T---------.---   113
SMM.US.04.G078               ---..------..----------------H---A-------------------R------Y----R-E--T------------W---M------R--------.-----------.---   113
SMM.US.05.D215               --N..------..----------------E----D---V--------------R----------------T------------YI----------------D-.-----------.---   113
SMM.US.06.FTq                ---..------..----------------E----D---A--------------R----------R--A--T-------------V---------R--------.T----------.---   113
SMM.US.86.CFU212             -A-..------..----------------------L--V--------------R-C--F--------A--T-------M---M----R---N--R---E----.-Q---------.---   113
SMM.US.x.H9                  --X..------..--------X----D--H--X-----A--------------R-X-------M---V--T-X----------------------GE------.-----------.---   113
SMM.US.x.PGM53               ---..------..-------------------------A--------------R------R--M---E--T-------------V---R-------E------.-S---------.---   113
SMM.x.x.pE660.CG7G           ---..------..---------E-----------G---A--------------R---------M---E--T-------------V-----------E------.-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                            Vpr end
H1B.FR.83.HXB2             MEQA.........................PEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYETYGD........TWAGVEAIIRILQQLLFIH#FRIGCRHSRIGVTR.....QRRARNGASRS.....................*     96
H1O.CM.91.MVP5180          ----.........................L-N---A---F------------E--------P--QAC--Y-------........--E--M-----------T-.Y----Q-----ILPS.NTRG-GR---S---......................-    101
H1N.CM.95.YBF30            --R-.........................---A------Y---A-----------------------------N-------........--E----------------.Y----Q-----I-P.....-.-R---T---......................-     96
H1P.CM.06.U14788           ---P.........................------P---F---M-QT-----E--------P---S---Y--N--------........--E--T---------I---.Y----Q-----ILTP...SR-GR-H-P---.....................-     99
CPZ.CD.06.BF1167           ----.........................--N---P---W-Q-L-DT---I----M-----NT-QNI-NWVFQN---........S-E--QVL-TL--KA--T-.Y-Y--A-----IGPRRNGAS-SRS.........................-     96
CPZ.TZ.00.TAN1             ----.........................-N-N------YT--L-DI---I-Q---K----PI-Q-V-NWVFTI---........S-E--QEL-K---RA--T-.Y-H--I-----S.....................................-     84
GOR.CM.07.SIVgorCP2135con  ---I.........................-T-E------F------V-----Q--------P---Q----V-D--------........--E--T--------VI-V-.Y----H-----ILP...PSR-GR---PN--.....................-     99
                            Vpr start                                                                                                                        Vpr end

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...................-    102
H2A.SN.86.ST_JSP4_27       -TE.....................PTESP---RT-P---GD--VI-T-R-I-*--LK--DPRL-IT--NY--ARH--.........-LE-ARGL-----RA-LL-.--A--GR----QP-GR...NPLSAIPTP-GMR..................-    105
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-*GMQ..................-    105
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..................-    103
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..................-    102
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..................-    101
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-    134
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-    133
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-    134
COL.CM.x.CGU1              -SS.................GPKREREDPPEDESP-RDFT-LQDQV-NITR--R----LTHISVIMWGCTRRHWGR.......P-TAGL-YV--MECA-YL-VHAGQ--FPSTQNP........................................-     93
COL.UG.10.BWC01            -SE................QAPKRPREGVPEDETV-RDFN-LQ-QV-LITQ--R----LTHISVIMMGCVRRHWPD.......Q--AAI-YV---ESA-YL-IHAGQ-KFPSRTNAP......................................-     95
COL.UG.10.BWC07            -S-...................RHRREQAPEDET--RDYT-LQ-QV-FITQ--R----LTHIAVIWWGCTRRHWPD.......Q--IGM-YV--MEIA-YL-LHAGQ-KFPSRTNNP......................................-     92
DEB.CM.04.SIVdeb04CMPF3061 --RPPPSHP...................LTPLSRLVPISRLQLQYMMQDVNE--MK--Q-DE-I-VWNYCV-LPAEPGWTGEQAWA-S-IDY-KKV-KMIWL-.LAE--Y-R-AAAS-H.YPNI-PLSGRNQEVRDQE.................-    118
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.................-    118
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.................-    118
DEN.CD.x.CD1_CMO580407     -AE..............................TTQRQLQ-AVWDITE--RK--SKEEIT-IWD-CWSLPALPHWTEGQVMA-A-IDF---M-KEIWK-.Y-V--F-REAERV-H...YPNIRPLRPR-EEP......................-     99
DRL.DE.11.D3               --RP...........................---E--P---YD--LVDT-V-IQE--LK--D-RL--NV-SWV--R---------.........SLE--QQL-----RA--L-.--Y--QEG-F-.PAGRW..NPFRTFPRANQQL...................-    101
DRL.x.x.FAO                --RP...........................---E--P---YD--LVDT-V-IQE--LK--D-RL--NV-SWV--R---------.........SLE--HQL-IL--RA--L-.--Y--QE--L-.PAGRW..NPFRTFPRANQQL...................-    101
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...........-    119
GSN.CM.99.CN166            --RIPPSHP...................MPWLSRRVPTTMAIAQNA-W-INE--EK--S-DE-R-IWHDVT-LPA-PDWTVDQAAI-CAIDY-.RV-T---R-.-KD--F-RVNLS.GRYPAIRPSRGTAPPD-NSVSHADPEQPRRPSRYRMDE-    134
GSN.CM.99.CN71             --RTPPSHP...................LPWISRRVPTTMAVAQSAMW-INE--EK--S-EE-R-IWHDVT-LPA-PDWTVDQAAI-CAIDYV-RV-T---R-.--D--F-RYNRIV-RYPVIRPLRGTAPPD-NSVPHADPEQPRRPSRYRMDE-    136
LST.CD.88.SIVlhoest447     -SRTR...............QQRSPEKP---F--P---WGQ-LADTM--I-E--R----ETI-AQVSEYCV........SSTGSEEKACMKF-T-VNRA-----.LPT-PSYAAAR...AGTAHPTEAAPRPTKRDIQRGRY.............-    115
LST.CD.88.SIVlhoest485     -SR-R...............QQRSPERP---F--P---WGQ-LADTM--I-E--R----EAI-AQVSEYCV........SSTGSEEEACMRF-T-VNRA-----.LPT-PSYAATR...AGTAHPTEAAPKPTERDIQRGRY.............-    115
LST.CD.88.SIVlhoest524     -SRQR...............QQRSPERP---F--P---WAQ-LADTM--I-E--R----EAI-AQVSEYCV........SSTGSEEEACMRF-TLVNRA-YL-.LPQ-PSYAAAR...AGTAHPTEAAPRPTGRDIQRGHY.............-    115
LST.KE.x.lho7              -SRQR...............QQRSPERP---F--P---WGQ-LADTM--I-E--R----EVM-AQVS-YCVDST-.........SEEEACMKF-TLVNRA--L-.LP.T-PSYA.AAR..AGTAHPTEAAPRPNEGDIQRGSD.............-    115
MAC.US.x.251_1A11          --E.......................RP--NE--*---WG--VV-V-----EG-LK--DPRL-TA--N---NRH--.........-LE-AGEL-----RA--M-.--G--I-----QPGGR...NPLSAIPP---ML...................-    101
MND-1.GA.x.MNDGB1          -...................GQKRDEQVS-----P---Y-Q-LADTM--I-E--RK---L-I-NAVSEYCVQNT-SE.......EE-CEKF-T.LMNRAIWV-.LAQ--DGTF....-ERRPQLPPSGFRPRGDRL.....................-    105
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...................-    102
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...................-    102
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...................-    102
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...................-    102
MON.CM.99.L1_99CML1        -V...............................PIERQALQAAIW--NE--LK--S-EE-R-IWEQVT-LPA-PAWNADQAWA-CAIDYT-WV-TI-YR-.Y-E--Y-RYAEQI-RYPVLRPMRGTAPGPTSSVPQADPDNPRRPSRYRMDE-    121
MON.NG.x.NG1               VV...............................PIERAALQAAIWD-NE--LX--SKEE-R-IWDQVT-LPA-PGWNADQAWA-CAIDYT-RV-TXI-R-.M-D--Y-RHAR-X-RYPXLRPLRGTAPGPTSSVPHADPERPLHPSRYHHDE-    121
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-    136
MUS-1.CM.01.SIVmus_01CM1085 --SLPPSHP.....................APWHSRVVPTTMRQAQ-AMW-VNE--EK--S-EE-R-IWNEVT-LPA-PDWTVDQAAI-CAIDYV-RV-T-V-R-.--D--Y-RYNR-M-RYPTIRPLRGTQPPP-NSTPNADPVPPLSPSRYRMDE-    136
MUS-2.CM.01.CM1246         ----..................PPSHPLSWLSR-TPTTMVQAQAA-W--NE--EK--S-DK-R-IWDRVT-LPA-PQWSVDQAAITCAIDY-KSV-T---R-.Y-D--Y--YAQTI-RYPPLRPLRGTQPPD-NSMPNADPTPSLRPSRYRMDE-    136
MUS-2.CM.01.CM2500         ---H..................PPSHPLPWLSR-VPTTMAQAQNAMWD-NE--EK--S-DE-R-IWDRVT-LPA-PNWTVDQAAI-CAIDY--WV-G---R-.--E--Y-RYSQ-I-RYPQVRPLRGTQPPD-NSTPNVDPFPPRRPSRYRMDE-    136
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-    140
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-    136
OLC.CI.97.97CI12           -AE.......................RV------P---LED-VRDTM--IIV--K-----EL-VRVFLFCAQ.......RHINDEMSTSLLF--L--RA--Q-.Y-N--......PGVELRRNPH-RSE-MQE.......................-     97
RCM.CM.00.SIVagi_00CM312   --........................IPQ--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..................-    101
RCM.CM.02.SIVrcm_02CM8081  --........................MP---E------M---L-DT-M--QE--AK--TMEL--AV-NY---Q---------........SLE--RVL-TL--RA--L-.Y-H--TG----TN-GG...NPLQSIPL--NIL..................-    101
RCM.GA.x.SIVRCMGAB1        --........................LP---E------YD--LMDT-I--QE--KK--TYAL-TQI-DYV--QH--------........SIE--Q-M--L--RA--L-.--N--AG----TS-GS...NPLRSIPQT-NIM..................-    101
RCM.NG.x.NG411             --........................ML---EE-----YD--LMDT-I-IQE--KK--TPEL-TQE-NY---QH--------........SLE--K-M-TL-NKA--L-.--H--EG----GA-GG...NPLRSILH--NIL..................-    101
SAB.SN.x.SAB1              -ASGGWLPPVGGDPPKDPPKNPREEIPGWL-TWDLP---FD--LRDM-QD-NS--QC----NL-FR-WWN-V-EPAI...DHGQTRLE-WYKYC----KA--V-.MKGR-CKPKTHPAYGPGAGGPPPGL-GASGGAASAAPGL............-    141
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..................-    102
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..................-    102
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..................-    103
SMM.SL.92.SIVsmSL92B       --H........................A---ETNP---WD--IRDV-----E--LK--DPRL-TA--NY--D----------.........-IE-AGE--K---KA--L-.--H--T-----QP-GP...NPLGSIPSA-NVL..................-    101
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..................-    101
SMM.US.04.G078             -AE.......................RP---E------WD--VT-V---I-Q--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-K---RA--M-.--N--V-----QSGGG...NPL-TIPP--EVL..................-    102
SMM.US.05.D215             -AG.......................VP---E------WD--VG-V---I-E--LK--DPRL-NA--NY--DRH--.........-LE-AGEL-K---RA--M-.--G--Q-----QPGGG...NPL-TIPP--GVL..................-    102
SMM.US.06.FTq              -AE.......................RP---E------WD--VV-V-----E--LK--DTRL-TA--NY--DRH--.........-LE-AGEL-K---RA--L-.--G--T-----TN-GG...NPLSTVPT--NVL..................-    102
SMM.US.86.CFU212           -AE.......................AP---EA-----WD--VM-V-----E--K---DLRL-TA--NY--DRH--.........-LE-AGEL-K---RA----.--G--A-----HP-GG...NPL-TIPPT-DVL..................-    102
SMM.US.x.H9                -TE.......................RP---EA-----WD--VV-V---X-E--LNX-DPRL-IA--NY--DRH--.........-LE-AGEL-----GA----.--G--------QSGGX...NPLSTIPPX-GVL..................-    102
SMM.US.x.PGM53             -TE.......................RP---EA-----WD--VVKV---V-E--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-----RA----.--G--A-----QS-GG...NPLSTIPP--GVL..................-    102
SMM.x.x.pE660.CG7G         -AE.......................RP---EA-----WD--VV-V-----E--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-----RA----.--S--A-----QS-GG...NPLSTIPP--DML..................-    102
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..................-    102
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-    117
SYK.KE.x.KE51              -A--F..........FRVYQRQPGEPLFYIPRNVNWD-AGVISATAKA-QVAQ--CK--TAEEIF-VWNQCL-VEAGPEETPTMAWIRCMLDMH-A-NFM--E-.-AA--PQRTKYARHRGYPHPS.............................-    116
SYK.KE.x.SYK173_COMGNM     -AE-F..........FNPSQHVQGTPWFFIPRNVELT-NVINVTVKA-LVVT--SK--TPQEIY-VWNQSLNEEAGTD.SPTMAWERTMLDMV-A-NLM--E-.-AA--PQRTRYARHRGYPHPS.............................-    114
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..................-    116
TAL.CM.01.8023             --RL....................PPSHPLPITSRLIPTPRAQLER-MR-VRD-MLI--T-EEVI-VWN-CV-LPAEPDWTGEQAWA-S-IDFA-KGNEM--L-.-QQ--F-EQQRR-AHYPNIRPLSGRR.NTMQ..................-    115
TAN.UG.x.SIVagmTAN1        -AEGR..............DSRERRPGWL-IWDLS---WD-*LRDMVA--NQ--Q---G-EL-FQVWNFCQ-EGER....NGAPMIERAYRYY-LV-KA--V-.--C---RRTPFEPYEERRNGVGGGRDGRE.....PPPGLA...........-    120
VER.DE.x.AGM3              -ASGR..............DPREARPGEL-IWDLS---WD--LRDM--DINQ--KM--G-EL-FQVWNYCQ-EGER....NRTPMLERAYKYYKLV-KA--V-.--C---RRQPFEPYEERRDGQGGGRAGRV......PPGLD...........-    120
VER.KE.x.9063              -ASGR..............GPRENRPGEV-IWDLN---WD--LRDM--D-NQ--KL--G-EL-FQVWNYCQ-EGER....RGAPMTERAYRYY-LV-KA--V-.--C---RRQPFEPYEERRDGQGGGRANRA......PPGLD...........-    120
VER.KE.x.AGM155            -ASGR..............DPREERPGGL-IWDLS---WD--LRDMV--IN---KL--G-EL-YQVWNYCQ-EGER....QGRPIAERAYKYY-LV-KA--V-.--C---RRQPFEPYEERRNGQGGGRPGRV......PPGLD...........-    120
VER.KE.x.TYO1_patent       -ASGR..............DPREARPGEV-IWDLS---WD--LRDM-QD-NQ--RL--G-EL-FQVWNYCQ-EGER....HGTPMMERAYKYY-LV-KA--V-.--C---RRQPFEPYEERRDGQGGGRANRV......PPGLE...........-    120
WRC.CI.97.97CI14           -SN.......................SR-QAPSGGP-PIPED-LGMQ--FAM-KQ----EV-VRVFLYCSQ.......QREYEEVPTMHYFH-V--RG--M-.LTQ--PKQQEKREQQFQRAEQLY--LERTAGDQGEVEQRDLISL.......-    117
WRC.CI.98.98CI04           -SN.......................SR-QAPTGGP-PTPED-LGMQ--FAM-KL----EV-VRVFLYCSQ.......QRNYEEGPTMHYFH-V--RG--M-.LTQ--PKQQEKREQQFQRAEQFF--LERAAGDQGEVEQRDLISL.......-    117
WRC.GM.05.Pbt_05GM_X02     -SNS......................GR-RPPTGGPQAIPED-LGMQ--FAM-K-----EV-VRVFLYCSQ.......QRNYEEQ-TMHYFH-V--RG-YL-.LIQ--PKEKENRE-DFQRAEQFY--LERA-GDQGGVEQRDLISL.......-    118
```

```
                                 Tat start                                                                   C-rich region          nuclear localization           exon 1 end  exon 2 start
H1B.FR.83.HXB2                   MEPVDP...............................................RLEPWKHPGSQPKTAC.TNCYCKKCCFHCQVCFITKALGISYGRKKRRQRRRAHQN.....SQTHQASLSK.QPTSQP...RGDPTGPKE*KKKVERETET.DPFD.........  100
H1O.CM.91.MVP5180                -D----...............................................EMP--H----K-Q-P-.N-----R--Y--Y---TK-G----H-------.-PA-AA.....SYPDNKDPVPE.-SL-HT...GRKQKRQE-QE----K--GP.SGQPCHQDSCNSC  109
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-CWKDYCLSS  108
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.TZ.00.TAN1                   -N-I--...............................................QVA--E---AA-E-P-.-------------P---TK-----------.-RG-KSAVH.....-TNN-DPVRQ.-SLPKR...SRIQNSQE-SQEE--A--TS.GGRPRQQDSSVSS  108
GOR.CM.07.SIVgorCP2135con        -D-T--...............................................E-P--QQ-----P-P-.NT-------Y--Y---TK-G-----------RPA-TADK.....D-NN--PV--.-SLPGT...-SQQE*YQK-E----GQ-A-.-RS-CWQDSCRSS  109
                                 Tat start                                                                                                                     exon 1 end  exon 2 start

MAC.US.x.239                     --...TPLREQENSLESSNERSSCI................SEADASTPESANLGE-ILSQLYRPLEACY.NT-------Y---F--LK-G---C-EQSRK-R-TPKK.......AKANTS-A-N.K-I-NR.....TRHCQP-KA--ETV-KAV.ATAPGLGR.....  130
H2A.SN.86.ST_JSP4_27             -LDMETPLKAPEGSLGSYNEPSSCT................SEQDAAAQGLVSPGD-ILYQLYQPLEACD.NK-------Y---M--LN-G---W-E--G--R-TPKK.......TKA-SS-A-D.KSI-TR.....TGNSQP-K-Q-KTL--AL.ETIGGPGR.....  133
H2B.CI.x.EHO                     --IPLKEQESSLNSSSGHSSST...................SEGVANTQGLDNRGE-ILSQLYRPL-ACS.NT------SY---L--LK-G---C-E..RS-K-SSKR.......AK-TTS-APN.ESL-AR.....TGDSQPTK-Q-K-V--TR.ATDLGPGRSNTST  133
H2G.CI.92.Abt96                  -IGMETPSREQENSLKSCRELSSCT................FEETVDAPGLEGTQA-ILYQLYRPLEACS.NK-------Y---L--LN-G--VC-EQPR--RTPKKT........KANSF-A-N.DRSISAR....TRNXQPKK-*-EKV--KV.ATDLGLGR.....  132
H2U.CI.07.07IC_TNP3              --TPLKEQESLSKSCKEPSLCTSEEDVRIPESGDKEEEILSQLYRPLETCYN.....................K-------Y---L--LK-G---C-E.QS--RSPKK........AKVNTFPA-N.KSIPTR.....TRNSQPKK-Q-ETV--TL.ASTPGLGK.....  128
H2U.FR.96.12034                  --IPLKEQESSLKSCKEQSSSTSEGDVGTQGSAAAGEEILSQLYRPLEACSN.....................K-------Y---L--LK-G---C-E.QP--RSPKKVK........AYSSAP-N.KSL-NG.....TRNSQPKE-*-KAL--AV.ATDLGLGR.....  127
ASC.UG.10.RT03                   -D-SQEGLEKWHRAPAQPPTAC..................................................P-F--N-IY--IL--QR-G---R---FRK-R-PG-APTP..ESL-NN-D-VRQ.-Y.........QVDEEKQT----T--TS.--C-C*TGS....  101
ASC.UG.10.RT08                   -D-SQEGLEKWHRAPAQPPTPC.................................................NS-F--S-IY--IL--QQ-G---R---FRK-R-PG-APTP..ESL-NN-D-VRQ.-S.........QVDE-KQAA---T--TS.--C-C........  98
ASC.UG.10.RT11                   -D-SQEEVEKWHRAPAQPPTPC.................................................NS-F--S-IY--IL--QR-G---R---FRK-R-PG-APTP..ESL-NN-DFVRQ.-F.........QV-EEKQAE---T--TS.N-C-C........  98
COL.CM.x.CGU1                    -SLSLYAKPLEAC..........................................................NDK-W--A------L--LS-G---R-..--NKSAT..................QT.-TNVSS...QIRGQDREWEQPDNAE-IHR#A#KTAAADFGD..  84
COL.UG.10.BWC01                  -*ISFQNERPLKAC.........................................................NNK-W--G------L--LS-G---C-..--NKKIN..................RC.-EPLSS...QIRGQNQEWEQPDNGE-LDSEQEDNNSRLVQL#.  86
COL.UG.10.BWC07                  -QISFQNQQPLTAC.........................................................TNK-W--A-VY---L--LS-G---C-..--NKKSP..................QS.KEEELTTILQIPTQTQW-QPDNA-ELDS#.IQEAAAADFGYW.  89
DEB.CM.04.SIVdeb04CMPF3061       --EEMDLFQGRGRGEAN......................................................HPS----I-T---YL--TQ-G---T--TRR-KR-APESA..TSDS-NQDIV-KQ...R.........RISSFQEKE-EAL-AAPGD.GGLPRRKDS....  98
DEB.CM.99.CM40                   --SI--FEERANTPH........................................................PA-M--A-IL---L---Q-G------PRR-KRKC-REYTPAAESN-PD-DFVPE.-.........-RKCFDSQKKEAE--ETAGP.GGQHCREDSSVSS  103
DEB.CM.99.CM5                    --EI--FKETANTPH........................................................P--M--A-TM---L--MQ-G---Y--SRR-KR--GSSFA...ANDKD--EFIPE.-.........-WKCFNSE-KET---KASRA.GGRHRPEDS....  96
DEN.CD.x.CD1_CMO580407           -QKEIWKHYRVGCFHREAERVRHYPNIRP#TPSQRRAMNADSIDPFAGNKTPCN..................T----H-AY---L--LQ-G---V---AP-KRP-AR......QSGTPD-EAV--.-SL-D....PRRQQISQKKE-EKVETATGSSGGNIREDS....  135
DRL.DE.11.D3                     -DARKVDLDQQDAGIHFEPFQGRTNNCN...........................................K-W--H-AY---L--LQ-G---R-NVSRP-K-TASVV....SRVSETGDTPVA.RTHTAQRRPTTTDSQT-KKE-SKTE-SGI.SDQSPPSQNLQDL  119
DRL.x.x.FAO                      -DARKVDLDQQDAGTHFEPFQGRTNNCN...........................................T-W--Y-AY---L--LQ-G---R-NVSRP-K-TATVV....SRVSETGDTPVA.RTHTAQRRPTTTDSQTEKKE-SKTEASG.TSDQNPPSQDPPKL  119
GRV.ET.x.GRI_677_gri_1           -DKEEEPHPL.LQDLHR................................................-LQP-TNK-------Y--EL--LQ-G--VR-HVSRK-RKTSTQ.........DN-DPIRQ.-SI-TV.....QRN-QTTEEG-T-V-KAAAAN...........  94
GSN.CM.99.CN166                  -NRDPEECLLPNWQQPGAAPATPC...............................................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............................MNRDPAECLLPKWQQPGAAPNTPCS..A----R-A----L--LR-G--L-LHGRS-KRK-SEASEP.APSVP-N-DPIQI.-SLPHSKW...NAIRQT-QE-T-A-KATS.G-S-ISQDSRVSC  131
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_1A11                --TPLREQENSLESSSGHSSCTSEAAASTPELANLGEEILSQLYRPLEACYN.....................T-------Y---F--LK-G---C-EQSRK-R-TPKK.......AKANTS-A-N.NRPISNR....TRHCQP-KA--ETV-KAV.ATAPGLGR.....  131
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                 -DAGKAVSDKKEGDVTPYDPFRDRTTPCD..........................................T----R--Y---L--LQ----VH-HAYRN-RS-QRLL....EKISEDS-APVG.RE-ISVPT...TSNSQAKKE--SPTQKNR.V-NQTAPK.....  112
MND-2.GA.x.M14                   -DAGKAELDKKEANITPLDPFQGRTTPCN..........................................K-------Y---C--LQ----H-HVYRT-RT-KRLL....GKISQDSEAAVG.RE-ISDLA...TSDSQTKKE-ESST*KNR.TSSQRAPRKNLAD  116
MND-2.x.x.5440                   -DVGEVASDKKEEDITHFDPFRARTTPCN..........................................K-F--V--Y---L--LQ----VH-HVYRQ-RP--GLL....EKVPSNSQAALG.RE-ISGPT...TSN-QT-KEE-SQT*KNR.TSSQRAPDKNLGS  116
MNE.US.x.MNE027                  -IDMETPLREQENSLKSSNGRSSCT................SEAAATTLESANLEE-ILSQLYRPLEACY.NT-------Y---F--LK-G---C-EQSR--R-TPKK.......VKANTS-A-N.NRPISDR....TKHCQPKKEQ-ETV--AV.ATVPGLGR.....  134
MON.CM.99.L1_99CML1              -MEPVDPDLPKEQHPPATPRTPC..................................................---F-RV------I--L--G----.YGR--KR--ATSPVP...GLSSSKNPAR-.-SL-H-...TR-SQR-T-QAQA-ATAATP.-RQHI........  104
MUS-1.CM.01.CM1239               -D-SVENLPPEQRPGAAPTTPC.................................................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-SVEELPKEQRPGAAPATPC.................................................-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-SVEGLPPDQRPGAAPPTPC.................................................------C-A---IL--QK---------RI-KRP-TDSL......LPSD-NPVRF.-SIPD.........SQQN-SNKKKQKKA.LEEAAGTDPCRCF  102
MUS-2.CM.01.CM2500               -D-EVENLPPWQRPSSAPTTPC.................................................-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*DHPGIAWMNNLEEENLPPWQRGTAAPATPC.................................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-DTPNWQHPPATPQTPC.................................................NS-F----AY--IL--QK-----L---SRK-R-GTDSG......LPNN-N-VQL.KSLPKH...E-L-VSQ-KQ--E-AATPA-.-.........  98
OLC.CI.97.97CI12                 -DAQEWSLEETLTEDLKEC....................................................KNK-------Y---L--LQ-G-----STESP-K--KKVV........S--QDSRD.TSHF-IPS..RTRNSQTTE-E-TASSSRA..-STSGTVPQSVQ  105
RCM.CM.00.SIVagi_00CM312         -DGQEAGLERQEEETLYNPFQSVETPCN...........................................A------AY---L--LQ--I--Q-ASRR--RQ-DKKEA.......KIDTLATPDKSL-ATS...RRRNSKPKE--ESSV-KAVATNHTTGRKDLSIS  115
RCM.CM.02.SIVrcm_02CM8081        -DVRAVGSERIEEETLYNPFHYQETSCN...........................................K------AY---L--LQ----N-ASRR-GS-DKKKTQ........VDQL-ATEHRPLSATG..RNSQSK--E-ETL-KKVV..TNRPAGREDISTP  113
RCM.GA.x.SIVRCMGAB1              -DVQGVGLEHPEEVILYDPFRKRETSCN...........................................T-------Y---L--LQ-G----N-ASRA--R-SKEENK.......ADKFPVPN.HRSISTTR..-NRKLQEKKE-T--KKVA..TSTTIG........  106
RCM.NG.x.NG411                   -DVKGAGLEEREEEILLDPFFTRETSCN...........................................S----R-VY---L--LQ----N-ASRR--RG-PGNKK....KN-INPFDT-E.-SL-AI...GRNSQSK--K-EA--KAVD-.NITTGREDIFIS.  115
SAB.SN.x.SAB1                    -DQEQEARPQVWEELQEELHRPLQACD............................................NT-F--V-----IL--HK-----R-YVPRP-RASKK.........IS-NQVSLH.KSI-TW.....TRDSQTKK-S-AAVG*T..ATAGHTLGRENS.  106
SMM.CI.79.SIVsmCI2               -IGMETPLKEQGNLSRSCNEHSLST................SEQDVFTPELGTVGE-ILSQLYRPLEECS.NR-------Y---L--LK-G---W-E.RS--R-APKK.......AKAYTF-A-N.HRSISTR....-RN-QPTKEQ-KTV--TV.ATDLGLGK.....  133
SMM.LR.89.SIVsmLIB1              -IGMETPSKEQGSSLRSCNEHSSCT................SEEAVGTPELASQEE-ILWQLYRPLEECF.NK---RS--Y---L--VK-G---T-A...-P-TPKK.......AKA-TPFT-N.HRSVPNR....TRHSQSKKE*-K-VA-TV.ATDLGLGR.....  130
SMM.SL.92.SIVsmSL92A             --TPLKEQENLLKSCREPCSSTSEEGVQTQELDAEEEEILSQLYRPLETCCN.....................K-F-----Y---L--LK-G---C-EPERT-R--AKKK.......IK-DSFPS-N.YRSIPTR....SRNCQPKKE*TKKV--AV.V-DPGLGK.....  131
SMM.SL.92.SL92B                  -DAPIQGLDSQGDQIPWDLYRPLETCYN...........................................S-F---------L--LR-G---T-AKPR--VKKK-K..........ANPFDTSNHRSISSG....TRNCQPKKEQ-EKV---V.-TDCGLGR.....  105
SMM.US.04.G078                   -IDMETPLREQESSLKSCNEPSSCT................SEMDVSTPELAKAEE-ILWQLYRPLEKC-.NT----------F--LR-G---F-ESSR--R-TPKK.......AKANTSVA-D.YRSLSER....ARNC-PKEE--EKV--AV.ATDLGLGR.....  134
SMM.US.05.D215                   -IDMETPSKEQENSLRSFNELSSCT................SGADANTPELANLEE-ILWQLYRPLEACY.NR-------Y---Y--LK-G---C-ETAG--R-AKKK.......ITPYSL-SRN.NRSISTS....SRDSQPKEE*-EKVA-AV.ATDLGLGR.....  133
SMM.US.06.FTq                    --TPLREQESSLKSYKEHSSSTSEEAAPIQESAQIEEEILFQLYQPLETCCN.....................T-------Y---L--LK-G---C-E.QS--R-TPKK.......VK-NSF-APN.NRSISTR....-RNSQPKKE*-KTL--TL.ATDLGLGR.....  129
SMM.US.86.CFU212                 --TRSREPESSLRYSSERSSSTSEVDVPTQESVTQEEEILWQLYRPLEMCCN.....................S----R--Y---L--LR-G--VC-D..RP-R-TPKK.......VK-NTLFT-D.NRSISTR....TRNSQP-KE*-K-VVSTV.ASDLGLGR.....  128
SMM.US.x.H9                      -IDMETPLKEQESSLESSREHSSSI................SEVDADTPESASLEE-ILSQLYRPVEACY.NK-------Y---H--LK-G---C-EQQR--TPKKTK..........ANTFSASDNRSLSXR....ARNCQPKKE--ETV-X-V.ATDLGLXR.....  132
SMM.US.x.PGM53                   --TPLKEQENSLGSSNERSSYTSEEDVLTPESAKVEEEILCQLYRPLEACSN.....................T-------Y---L--LK-G---C-EQQR--R-TPKK.......AKANTF-APN.DRSLSTR....ARNSQPKKE--ETV---V.ATDLGLGR.....  131
SMM.x.x.pE660.CG7G               -IDMETPLKEQENSLESCREHSSSI................SEVDVPTPESANLEE-ILYQLYRPLE-CY.NK----R--Y---H--LK-G---C-E..QH-R-TPKK.......TKANPLPA-N.KSL-TR.....ARNRQPKKE--EKV---V.AADLGLGR.....  131
STM.US.89.STM_37_16              -IDMETPLKEQESSLRSSSEPSSCT................SEAVAATPGLANQEE-ILWQLYRPLEEC-.NK-F-----Y---L--V--G---T-E-SR--VKKK-K..........TYPISASNNRSLSTR....ARNSQPKKEQ-K-V---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                   -SSKEELPTTPTYDPFPAGG.TPCNK................................................-------Y---L--LQ-G---R.YVPD--I-K-RRSTS..SSN-KDKDPV--.--L-RA...T-NQECQ-KKVQ---SP-GA.GGEHRLKDSLPLS  113
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--SARTAPSLGRKNLAQQ  110
VER.KE.x.9063                    -DKGEEERTVLHQDLIR.....................................QYKK-L-T-RNK-F-----Y---L--LQ-G--VT-HAPRT-RKKSVQ.........PNRL-QQD.-SI-TR.....GRD-QATQESQKKV-R--TTAQILGRKDLERD  110
VER.KE.x.AGM155                  -DKGEEDQDVSHQDLIKQYRKPLETC.............................................TNK-F-----Y---F--LR-G---T-HAFRT-RKKI-S.........ADRIPVPQ.-SI-IR.....GRDSQTTQESQKKV-EQAKANLRISRKNLGDE  110
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                 -RKSLLCIISTGYCKEGSSCISPKGVLNNKRKGNNNFKEQNNCIETWREQQEIREK-SKETLSPC......DHR-S-VA-VYS-TL--LQ-G-----RSYSKKTKPDTTT.............AAS.SR-TSTT...WKRDSKTAK-E*KKIT-V..............  131
WRC.CI.98.98CI04                 -KK-PLCIIFTGYYKEGSSCTLPKGVQNNKKRENNNFKEQNNSLETWKEQQEIRE--SKETLSPC......NHK-G-Y--V-S--L--LQ-G-----RSYSKKIGASTAK.............ATT.SRATSTT...WKRDCQT-K--QKKTTVL..............  132
WRC.GM.05.Pbt_05GM_X02           -KNRPLCTIFTGSFKEDYIYTSYKDVPKKKRTEKEIFKEQNNFIEIWKGLQEIREE-SRETLSPC......NHQ-R----I-S--L--LQ-G-----HSYSKKTNKDKSD.............A-T.SRVIS*T...WKRDCQTTK-Q-KKIS................  129
```

```
                                                  Tat end
H1B.FR.83.HXB2                  ..........................*  100
H1O.CM.91.MVP5180               TRISGQ....................-  116
H1N.CM.95.YBF30                 ..........................-  102
H1P.CM.06.U14788                GTT.......................-  112
CPZ.CD.06.BF1167                ..........................-  101
CPZ.TZ.00.TAN1                  GRTSGTSSSGYTRPFKTSSGSSGSACKH-  137
GOR.CM.07.SIVgorCP2135con       GTISEP....................-  116
                                                  Tat end
MAC.US.x.239                    ..........................-  131
H2A.SN.86.ST_JSP4_27            ..........................-  134
H2B.CI.x.EHO                    SRFAN.....................-  139
H2G.CI.92.Abt96                 ..........................-  133
H2U.CI.07.07IC_TNP3             ..........................-  129
H2U.FR.96.12034                 ..........................-  128
ASC.UG.10.RT03                  ..........................-  102
ASC.UG.10.RT08                  ..........................-   99
ASC.UG.10.RT11                  ..........................-   99
COL.CM.x.CGU1                   ..........................-   85
COL.UG.10.BWC01                 ..........................-   87
COL.UG.10.BWC07                 ..........................-   90
DEB.CM.04.SIVdeb04CMPF3061      ..........................-   99
DEB.CM.99.CM40                  GRISNNC...................-  111
DEB.CM.99.CM5                   ..........................-   97
DEN.CD.x.CD1_CMO580407          ..........................-  136
DRL.DE.11.D3                    GAKK.Q....................-  125
DRL.x.x.FAO                     GAKK.Q....................-  125
GRV.ET.x.GRI_677_gri_1          ..........................-   95
GSN.CM.99.CN166                 ..........................-  105
GSN.CM.99.CN71                  GTT.......................-  135
LST.CD.88.SIVlhoest447          ..........................-  111
LST.CD.88.SIVlhoest485          ..........................-  108
LST.CD.88.SIVlhoest524          ..........................-  109
LST.KE.x.lho7                   ..........................-  109
MAC.US.x.251_1A11               ..........................-  132
MND-1.GA.x.MNDGB1               ..........................-  108
MND-2.CM.98.CM16                ..........................-  113
MND-2.GA.x.M14                  LGGQSK....................-  123
MND-2.x.x.5440                  LGSQSKSGAG................-  127
MNE.US.x.MNE027                 ..........................-  135
MON.CM.99.L1_99CML1             ..........................-  105
MUS-1.CM.01.CM1239              ..........................-  102
MUS-1.CM.01.SIVmus_01CM1085     ..........................-   99
MUS-2.CM.01.CM1246              QKTSTT....................-  109
MUS-2.CM.01.CM2500              ..........................-  102
MUS-3.GA.09.09GabOI81           ..........................-  117
MUS-3.GA.11.11GabPts02          ..........................-   99
OLC.CI.97.97CI12                GH........................-  108
RCM.CM.00.SIVagi_00CM312        ..........................-  116
RCM.CM.02.SIVrcm_02CM8081       ..........................-  114
RCM.GA.x.SIVRCMGAB1             ..........................-  107
RCM.NG.x.NG411                  ..........................-  116
SAB.SN.x.SAB1                   ..........................-  107
SMM.CI.79.SIVsmCI2              ..........................-  134
SMM.LR.89.SIVsmLIB1             ..........................-  131
SMM.SL.92.SIVsmSL92A            ..........................-  132
SMM.SL.92.SL92B                 ..........................-  106
SMM.US.04.G078                  ..........................-  135
SMM.US.05.D215                  ..........................-  134
SMM.US.06.FTq                   ..........................-  130
SMM.US.86.CFU212                ..........................-  129
SMM.US.x.H9                     ..........................-  133
SMM.US.x.PGM53                  ..........................-  132
SMM.x.x.pE660.CG7G              ..........................-  132
STM.US.89.STM_37_16             ..........................-  133
SUN.GA.98.L14                   ..........................-  103
SYK.KE.x.KE51                   ..........................-  116
SYK.KE.x.SYK173_COMGNM          ..........................-  111
TAL.CM.00.266                   ..........................-  111
TAL.CM.01.8023                  GRTSTASSATVTRP............-  128
TAN.UG.x.SIVagmTAN1             ..........................-  102
VER.DE.x.AGM3                   SGRATGASD.................-  120
VER.KE.x.9063                   KREAVGANA.................-  120
VER.KE.x.AGM155                 TRGPVGAGN.................-  120
VER.KE.x.TYO1_patent            ..........................-  101
WRC.CI.97.97CI14                ..........................-  132
WRC.CI.98.98CI04                ..........................-  133
WRC.GM.05.Pbt_05GM_X02          ..........................-  130
```


PLV Proteins

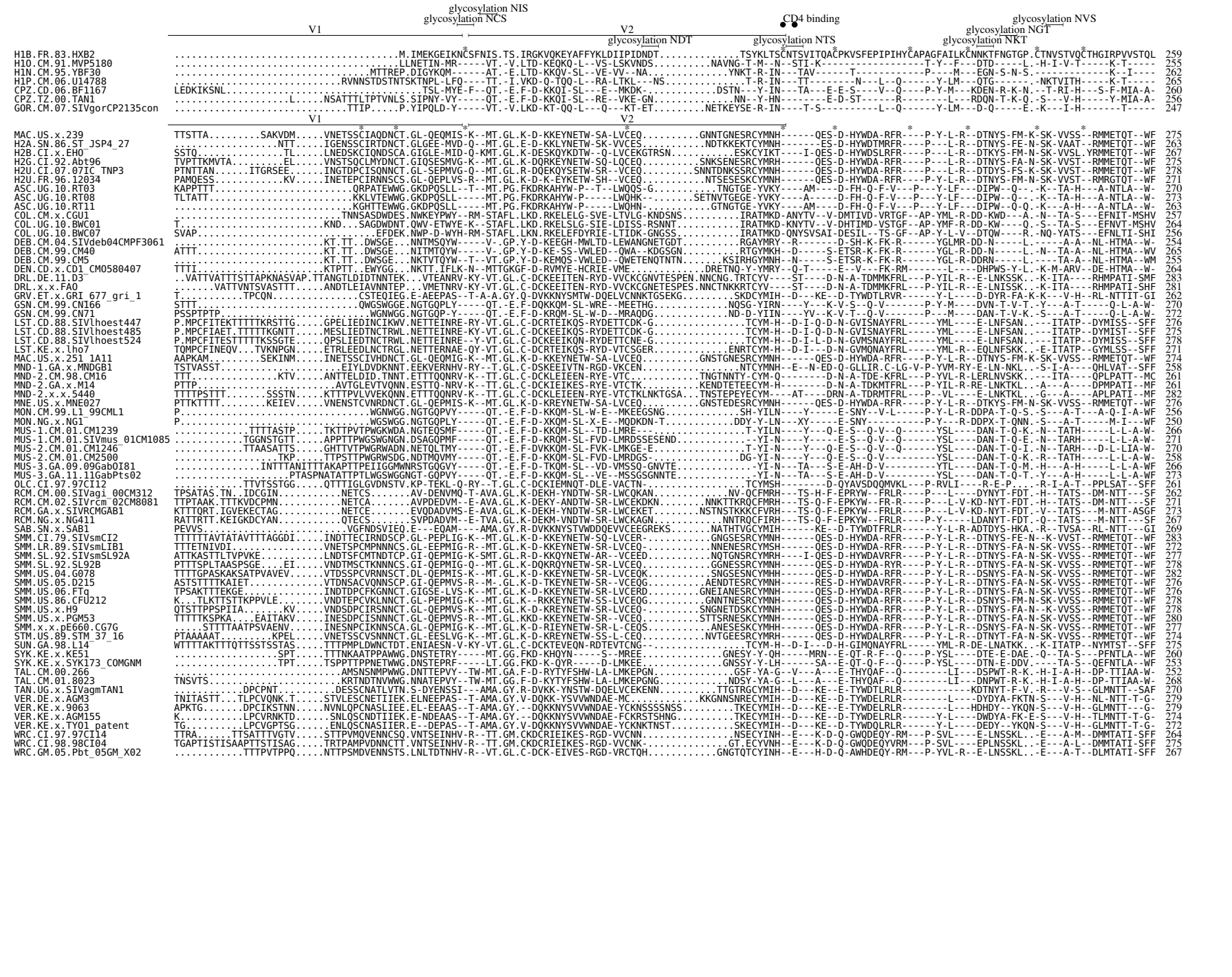
```
                                Rev start                     exon 1 end exon 2 start   NLS                                       Leu-rich effector domain
H1B.FR.83.HXB2                 MAGRSGDSDE.....ELIRTVRLIKLLYQ...SNPPPNP..EGTRQARRNRRRRWRERQRQIHSISERILGTYL.......GRSAEPVPLQLPPLERLTLDCNEDCGTS..........................GTQ.......GVGSPQILVESPTVLESGTKE...  116
H1O.CM.91.MVP5180              -----EED......QQ-LQAIQI--I---.....---C-T-..A-S-N--K-------R--A-VD-LAT---A-VV.......HG-QDNNLVD-----Q-NIRDP-ADRLP..........................--GTVDPGTKDN.....................  103
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
CPZ.CD.06.BF1167               ----EEEN.........-LQAIKI--I--E.....---Y--S.-----S----------I----VEG--S---A---........--LPTAP-VD--D-SK---APLDIDQVD..........................PSVEPIAVGDCETIAAQAHTP-DTISTP-GNLGN  123
CPZ.TZ.00.TAN1                 ----EE-A.......N-LY---I--I--D.....---Y-SG..A-S-T----------Q--H-VDALAS---QYR-........-GPQ--PH-DI-D-SK-H--PLDQPAST..........................E-GDNQLGTQPSN-A...................  105
GOR.CM.07.SIVgorCP2135con      -----DED.....LQ--LTKI-I-R----.....---C-E-..GPS-NST-------KG-----DQ-AG---AAH-........---QN-DCVD--DISH--IGDQGNNTQP..........................DRAPDQTTEKSE.....................  104
                                Rev start                       exon 1 end exon 2 start

MAC.US.x.239                   -SNHER.......EE--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.SN.86.ST_JSP4_27           -NE-AE.......EE--R-KL---R--H-...T--Y-QG..P--ASQ--------KQ-W--LVALADK-YTFPD.......PPADS-LEQTIQH-QG--IQELP-PP-N..........................LPESSESIDSSQRLAE-.................  107
H2B.CI.x.EHO                   -..NAR.......ERD-QKGL--LH--H-...T--Y-QG..P--ASQ--------KQ-GL--LALAD--HPLPD.......SPTEG-LD-AIQR-QN-IIKDLPNPP--..TPTAQASTCIPPIWDQLVPRSNPSSSQGCGRDSCERGEDLVGSPQESGRRDHCNTQEDQ  147
H2G.CI.92.Abt96                -.SNLE.......EE--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            -..NNP.......EE-VR-RL---HF-H-...T--Y-RG..P--ANQ---------Q-WL-LLALAN--SSFPD.......PPADS-LD-AIQQFQS-AIAELPEPPSG..........................LPESTDSNQRTSEA....................  102
H2U.FR.96.12034                -..SNP.......EEGVQ-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............................RWRDPPAQEDASPGSLASIGTQECKPGDSSDTSE  106
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.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-CEGSLGQTGCYCPIRLSGSHQQS  140
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-EVDQGLWNSSPQPCVAVPFVARPFSDPFLPTWATSSESLQQTG-GKRSEDCELELEQD-KEQKRHISSSCKKPQTGT  157
LST.CD.88.SIVlhoest485         -STGGEE.........--P-YLKIS-I-WT...GLQVGERREPR-ARQ--RD-E-Q-RHLH-LRTVQ---FHSTVERGL..ERAFTGLTVCDS-EVDQGIWNSSPQPCVAVPFVARPLSDPFLPTWATSLESQQRMD-GKRSEDCALELEQD-KEQKRHISSSCERHQTGT  157
LST.CD.88.SIVlhoest524         -STGPEG.........--PTYLK-SRI-WT...GLQVGERREPR-ARQ--RD-E-Q-KHLH-LRAVQ---FQATVERGL..ERAFTGLTVCDS-EVAEGVGNTAPHPRVAIPHLAGTYSDPFLPPWATSLANPQQIP-GECSENCELDQEQE-KEQRRAIKN.........  148
LST.KE.x.lho7                  -STGN--.........--P-YL--SRI-WT...GLQVGERREPR-ARQ--RE-S-Y-DYLH-LRAVQ---FQATVERGL..ERAFTRLAVSDS-EVAQGRGNTPPITSVAEPQLAVAFVDPFLPKWATPLADQQQMD-GKRSEDSELAQ-EM-KEQRTVIEH.........  148
MAC.US.x.251_1A11              -SSHER.......EE--RKRL---H--H-...T--Y-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
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             -.SDQE.......GE--RKRL---HF-H-...T--Y-QG..T--ANQ---------Q-WL--LALAN--YSFPD.......PPANSSLD-AVQQ-QG-SIEDLPEPPRN..........................LPTSLSSSQEQDQA....................  103
SMM.LR.89.SIVsmLIB1            -...HE.......DR--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-GQE......EE--RKRL---HF-H-...T--Y-QG..P--ANQ--------KQ-WY--LALAN--YSFPD.......PPADTSFD-AIQQ-QS--IEDLPEPP-N..........................IPETLKPDKQDP......................  103
SMM.SL.92.SL92B                -..QNP.......EE--R-RL---H-IH-...T--Y-AG..P--ASQ--------KR-WI--VALAD--YSFPD.......PPADTDID-AIQQFDS-SIQSLPEPP-T..........................VPETLRDQSAD.......................   99
SMM.US.04.G078                 -SLQGE.......EE--R-RL---H--H-...T--Y-KE..P--AHQ--R------Q-WL--LALAD--YSFPD.......PPVSS-LD-AVQQ-QS-AIEDLPEPP-N..........................VPETLRDSTENPRVH-A.................  107
SMM.US.05.D215                 -RQQEE.......EE--RKRL---HF-HA..T--Y-QA..P--ANR---------Q-WL--LALAD-VSSLST.......PPV-T-LD-AIQQ-QS-AITDLP-PPA-..........................VPEDLRSSPDSPRDH-D.................  107
SMM.US.06.FTq                  -.SNP-.......EE--R-RL---HF-H-...T-LY-QG..T--ANQ---------Q-WL--LALAD--YSFPD.......PPADS-ID-AVRQ-QN-AIEQLPEPPA-..........................APGDLQSVAESPR-H-T.................  106
SMM.US.86.CFU212               -..TDQ.......EE--RKRL---HF-H-...T--Y-QG..P--ANQ--------YR-WH--LALAD--YSFPD.......PPVHS-LD-AVQQ-QN-AIEDLPEPP--..........................APETLQCAASNK-DH-N.................  105
SMM.US.x.H9                    -SSN.........EE--R-RL---HF-H-...T--YXXG..P--ANQ--R------Q-WQ--LALAD--YSFPD.......PPADT-LD-AIQQ-QG-AIEXLPNPPA-..........................APEPLKDAAESP......................  100
SMM.US.x.PGM53                 -SSSA-.......EE--R-RL---HF-H-...T--Y-QG..P--ANQ--R------Q-WQ--LALAD--YSFPD.......PPADS-LD-AVQQ-QD-S-ENIP-PP--..........................VPEAL...TENPR-R-D.................  104
SMM.x.x.pE660.CG7G             -SST.........EE--R-RL---HF-H-...T--Y-QG..P--ANQ--R------Q-WQ--LALAD--YSFPN.......PPADT-LD-AIQQ-QG-AIEDLP-PP--..........................APEPLNDVAKSP......................  100
STM.US.89.STM_37_16            -..SDQ.......EE--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..GRAFERLSVSDSSQVAESLGNSPSTKHLPPAKFLVAPTYDFLPSWATPLADPQRLA-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-DSSK  131
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...WSSVCCATGEDCCASAEKKGEGDIHAWQQRDKQK  131
```

```
                                    Rev end
                                    ⌒
H1B.FR.83.HXB2                .................*  116
H1O.CM.91.MVP5180             .................-  104
H1N.CM.95.YBF30               .................-  105
H1P.CM.06.U14788              .................-  103
CPZ.CD.06.BF1167              SN...............-  126
CPZ.TZ.00.TAN1                .................-  106
GOR.CM.07.SIVgorCP2135con     .................-  105
                                    Rev end
                                    ⌒
MAC.US.x.239                  .................-  108
H2A.SN.86.ST_JSP4_27          .................-  108
H2B.CI.x.EHO                  TRG..............-  151
H2G.CI.92.Abt96               .................-  101
H2U.CI.07.07IC_TNP3           .................-  103
H2U.FR.96.12034               .................-  99
ASC.UG.10.RT03                .................-  97
ASC.UG.10.RT08                .................-  97
ASC.UG.10.RT11                .................-  97
COL.CM.x.CGU1                 .................-  94
COL.UG.10.BWC01               .................-  98
COL.UG.10.BWC07               VATICDECRA......-  117
DEB.CM.04.SIVdeb04CMPF3061    .................-  98
DEB.CM.99.CM40                .................-  91
DEB.CM.99.CM5                 .................-  88
DEN.CD.x.CD1_CMO580407        .................-  85
DRL.DE.11.D3                  .................-  87
DRL.x.x.FAO                   .................-  87
GRV.ET.x.GRI_677_gri_1        KKSAARP.........-  148
GSN.CM.99.CN166               .................-  125
GSN.CM.99.CN71                .................-  121
LST.CD.88.SIVlhoest447        KEEMEIQIQRKKWVSQ-  174
LST.CD.88.SIVlhoest485        KEKMEIQIQREKWVSQ-  174
LST.CD.88.SIVlhoest524        .................-  149
LST.KE.x.lho7                 .................-  149
MAC.US.x.251_1A11             .................-  108
MND-1.GA.x.MNDGB1             .................-  88
MND-2.CM.98.CM16              .................-  91
MND-2.GA.x.M14                .................-  89
MND-2.x.x.5440                .................-  88
MNE.US.x.MNE027               .................-  108
MON.CM.99.L1_99CML1           .................-  113
MUS-1.CM.01.CM1239            .................-  107
MUS-1.CM.01.SIVmus_01CM1085   .................-  111
MUS-2.CM.01.CM1246            .................-  100
MUS-2.CM.01.CM2500            .................-  118
MUS-3.GA.09.09GabOI81         .................-  114
MUS-3.GA.11.11GabPts02        .................-  111
OLC.CI.97.97CI12              .................-  124
RCM.CM.00.SIVagi_00CM312      .................-  108
RCM.CM.02.SIVrcm_02CM8081     .................-  105
RCM.GA.x.SIVRCMGAB1           .................-  94
RCM.NG.x.NG411                .................-  108
SAB.SN.x.SAB1                 .................-  100
SMM.CI.79.SIVsmCI2            .................-  104
SMM.LR.89.SIVsmLIB1           .................-  103
SMM.SL.92.SIVsmSL92A          .................-  104
SMM.SL.92.SL92B               .................-  100
SMM.US.04.G078                .................-  108
SMM.US.05.DZ15                .................-  108
SMM.US.06.FTq                 .................-  107
SMM.US.86.CFU212              .................-  106
SMM.US.x.H9                   .................-  101
SMM.US.x.PGM53                .................-  105
SMM.x.x.pE660.CG7G            .................-  101
STM.US.89.STM_37_16           .................-  100
SUN.GA.98.L14                 .................-  151
SYK.KE.x.KE51                 .................-  85
SYK.KE.x.SYK173_COMGNM        .................-  78
TAL.CM.00.266                 .................-  126
TAL.CM.01.8023                .................-  112
TAN.UG.x.SIVagmTAN1           .................-  104
VER.DE.x.AGM3                 .................-  90
VER.KE.x.9063                 .................-  90
VER.KE.x.AGM155               .................-  90
VER.KE.x.TYO1_patent          .................-  85
WRC.CI.97.97CI14              Q................-  133
WRC.CI.98.98CI04              .................-  124
WRC.GM.05.Pbt_05GM_X02        QQHQGNYCQTLHKP...-  146
```

```
                                                                  phos phos
                                transmembrane domain              cytoplasmic domain
                                Vpu start                                α-helix          α-helix                                Vpu end
H1B.FR.83.HXB2                  TQP..IPIVAIVALVVAIIIAIVVWSIVIIE....YRKILRQRK.IDRLIDRLIERAEDSGNESEGEIS.........ALVEMGVEMGH...HAPWDVDDL....*   82
H1O.CM.91.MVP5180               MHQ..ENLL-LI--SALCL-NVLI-LFNLRIYLVQRKQDR-EQE.-LERLR-IK-IRD--DY--NE-EQ...QE......-MELIHSHGF..AN-MFEL......-   86
H1N.CM.95.YBF30                 MLS.....LGFI--GA-VS--VI--ALLYR-.....K--KL-E--KHIRQ-IR--E-------D-DAEWLDGDEEW...-TLLSSSKLD..QGN-V........-   84
H1P.CM.06.U14788                MH-..RDEAVLIIAG-LLLCC--X-GK-LLL.....VL-ERERD-.FVQRLA-WR-GQ--E-Y--NE-EE...EQLRELG.............NLLGFDHV......-   79
CPZ.CD.06.BF1167                MLWQFLQWLQYLGWGG--V-W-IALLVIRKAYLY.IKQVQEEDR.-EQA-IDRYQ-RGSTDSGI-EDEN...GLI..................NWMDPDNYFLL..-   82
CPZ.TZ.00.TAN1                  MIKIVVGS-STNVIGILC-LL-LIGGGLL-GIG..I-RE-ERER.QHQRVLERLA-RLSIDSGV-EDEE..............FNWNNFDP...-NYNPR-WI....-   84
GOR.CM.07.SIVgorCP2135con       MH-..RD-IV-IIGITLLAATVI--LKALAL....-LRDR-E-R.FFNRLE--LSTK--E-Y--NE-EA...AELMERG.............NELGFDFN-H...-   80
DEN.CD.x.CD1 CMO580407          MLK...IKLGPIEY-CLVFAVVIT-IAA-GVGYL.AYRAYKSYR.EELRYI--RLWSI---Y--SQ-DP.....................................-   65
GSN.CM.99.CN166                 MH-.....A-VWWWGA---TF-YLCVALLALYLAWDKWVKGKP-PTQVAVI----DE----IYDDASSE..............LTGFNGF..AN-GFEV......-   79
GSN.CM.99.CN71                  MSA.....A-LWWWGA-V-TF-YFCLAIFALYLAWDKW-KGKP-.-PVAVI--V-DD-E--IFEDASSE...............PNAYGF..AN-GFEV......-   77
MON.CM.99.L1 99CML1             MNY..WWSLVAITYSLIL-ALP-AAWAWWRY....-KITK-FKR.--QE-Q---QIH-RRRHD-GVDTE.........SES-QHE-THGF..VN-VFN--FGEWV-   88
MON.NG.x.NG1                    MNT..YW-A--SVWC-IVLL-PCTLYLLYQS....-KEHK-VSR.FEQDFQ--VQXYQEXDSGY-D-DEDHNSFDNPLFDDGDPDQWV*FRKQIMGHSILWSAGL-   98
MUS-1.CM.01.CM1239              MNY..WYLA-ALVTGIYFL--LFAFVLAYQRWCRP..........KVEVSVIRLLEEG--DSGIFEDAE.....DEP.....N-DGHHAF..AN-AFEQ......-   76
MUS-1.CM.01.SIVmus_01CM1085     MNY..WYLA-VIVTGIYFV---FAFVLAYQRWCKP.........-.KVEVSVIRLLEEG-GDSGIFEDAE.....DDM.....AES-HHAF..AN-AFEQ......-   77
MUS-2.CM.01.CM2500              MNW..WWFA-A-VTAIYFV--L-AFVLAYQR....WCQP...KG.QVEVNVIRLLEEG-TDSGIFEDAE.....DG......SDDQRHGF..LN-AFEL......-   78
MUS-3.GA.09.09GabOI81           MNV..WYLA-AIVTAIYXA-V--ALWLL.......WD-Y.VRKEPTVVSVI--L-EGDSGYE-VFEDAP.....QEP.....GDPLINAF..DN-AFEQ......-   78
```

```
                                                                                                                     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...VVLVNVTENFNMW..KNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGR....  146
H1O.CM.91.MVP5180       -K-MK.KNN.RKS..-.SL.Y..IAMAL-IP-LS.YSKQ-.....YA---S-----E--A...PV-------NLTS--Q--I--SQ---.-----H-...FP-G---D--DI-..--Y--D---------E------E-M-F---QMN-V--QTNKTG..........  137
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-SYTS.GTR.....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
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.TZ.00.TAN1          -KNL...............IGIT-ILIIT-LGI.GFSTY.....YT--F---------Q...P--------DITSRDK--I----N---.L---Y-...-T-A--SIR---EE..-Y--QE-K---L--FQ--F--------F-IKMT-.TMT-T--KT.........  129
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
                                          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.SN.86.ST_JSP4_27    -.CG..........-NQL....FVASLLASA-L..IYCVQ.....Y---F------RN-S...IP---........-KNRDT-G-IQ-L-.DNDDY--....IAL----A-DA-N..-TVT--AV--VW--FET-I-----------AMR-NSTTAKNT-STPTTTTTA.  130
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-DLFK..-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-DLFK..-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--AFK..-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-DAFKNP..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.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--IYK..-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--IYK..-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_1A11       -GCL..........GNQL.....-IAIL-LSVYG.IYCT-.....Y---F----A-RN--...IP---........-KNRDT-G-TQ-L-.DNGDYS-....LAL----S-DA-E..-TVT--AI--VWQ-FET-I----------ITMR-NKSET-RWGLTK-STT.TT  131
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-KNR-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
SMM.US.04.G078          -GCL..........GNQL.....-IAIL-VSAYG.IYCTQ.....Y---F----A-RN--...IP---........-KNRDT-G-VQ-L-.DNDDYS-....LAI----A-DA-D..-TVT--AI--VWN-FET-I----------IAM--NKTDT-KWGLTGTTTTTQK  132
SMM.US.05.D215          -GCL..........GNQL.....-IAIL-VSACG.IWCTQ.....Y---F----A-RN--...IP---........-SNRDT-G-TQ-L-.DNGDYS-....MPI----A-DA-D..-TVT--AI--VWN-FES-I----------IAMR-NKSDT-RWGLTGVQATTPT  132
SMM.US.06.FTq           -GCL..........GNQL.....-IASL-LSACS.IYC-K.....Y---F--I-A-RN--...VP---........-KNRDT-G-VQ-L-.DNDDYS-....LAL----A-DA-E..-TVT--AV--VWN-FKT-I----------IAMR-NRTETNRWGLTGNSDTTTV  132
SMM.US.86.CFU212        -GCL..........GNQL.....-IAIL-VSAYG.IYCAQ.....Y---F----A-RN--...IP---.........--NRDT-G-TQ-L-.DNDDYS-....LAI----A-DA-N..-TVT--AI--VWN-FET-I----------IAMR-NKAET-KWGLTGTAAPPTT  132
SMM.US.x.H9             -GCL..........GNQL.....-IALS-XSA-G.IYCVQ.....Y---F----A-RN--...VP---........-RNRDT-G-TQ-L-.DNGDYS-....LAI----A-DA-D..-TVT--AI--VWN-FET-I----------ITMR-NKSEXXRWGLTGXPAPTTT  132
SMM.US.x.PGM53          -GCL..........GNQL.....-IAIL-LSAYG.IYCTQ.....Y---F----A--N--...IP---........-KNRDT-G-TQ-L-.DNDDYS-....LAI----R-DA-N..-TVT--AI--VWN-FET-I----------ITMR-NKTET-RWGLTKNEIPTTT  132
SMM.x.x.pE660.CG7G      -GCL..........GNQL.....-IALL-VSALE.IYCIQ.....Y---F----A--N--...IP---........-KNRDT-G-TQ-L-.DNDDYS-....LAI-I--A-DA-N..-TVT--AI--VWN-FET-I----------IAMR-NKTET-RWGLTRNAGTTTT  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-DVYK..-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-DVYK..-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-KDR-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
glycosylation NVS
V1
V2
glycosylation NDT
glycosylation NTS
glycosylation NGT
glycosylation NKT
H1B.FR.83.HXB2 259
H1O.CM.91.MVP5180 255
H1N.CM.95.YBF30 262
H1P.CM.06.U14788 265
CPZ.CD.06.BF1167 260
CPZ.TZ.00.TAN1 256
GOR.CM.07.SIVgorCP2135con 247
V1
V2
MAC.US.x.239 275
H2A.SN.86.ST_JSP4_27 263
H2B.CI.x.EHO 267
H2G.CI.92.Abt96 275
H2U.CI.07.07IC_TNP3 278
H2U.FR.96.12034 271
ASC.UG.10.RT03 270
ASC.UG.10.RT08 273
ASC.UG.10.RT11 263
COL.CM.x.CGU1 257
COL.UG.10.BWC01 264
COL.UG.10.BWC07 256
DEB.CM.04.SIVdeb04CMPF3061 254
DEB.CM.99.CM40 265
DEB.CM.99.CM5 255
DEN.CD.x.CD1_CM0580407 264
DRL.DE.11.D3 283
DRL.x.x.FAO 281
GRV.ET.x.GRI_677_gri_1 262
GSN.CM.99.CN166 270
GSN.CM.99.CN71 272
LST.CD.88.SIVlhoest447 276
LST.CD.88.SIVlhoest485 275
LST.CD.88.SIVlhoest524 278
LST.KE.x.lho7 271
MAC.US.x.251_1A11 274
MND-1.GA.x.MNDGB1 258
MND-2.CM.98.CM16 261
MND-2.GA.x.M14 261
MND-2.x.x.5440 282
MNE.US.x.MNE027 276
MON.CM.99.L1_99CML1 256
MON.NG.x.NG1 250
MUS-1.CM.01.CM1239 266
MUS-1.CM.01.SIVmus_01CM1085 271
MUS-2.CM.01.CM1246 266
MUS-2.CM.01.CM2500 258
MUS-3.GA.09.09GabOI81 266
MUS-3.GA.11.11GabPts02 273
OLC.CI.97.97CI12 261
RCM.CM.00.SIVagi_00CM312 262
RCM.CM.02.SIVrcm_02CM8081 271
RCM.GA.x.SIVRCMGAB1 273
RCM.NG.x.NG411 267
SAB.SN.x.SAB1 269
SMM.CI.79.SIVsmCI2 283
SMM.LR.89.SIVsmLIB1 272
SMM.SL.92.SIVsmSL92A 277
SMM.SL.92.SL92B 278
SMM.US.04.G078 282
SMM.US.05.D215 276
SMM.US.06.FTq 276
SMM.US.86.CFU212 278
SMM.US.x.H9 278
SMM.US.x.PGM53 280
SMM.x.x.pE660.CG7G 277
STM.US.89.STM_37_16 274
SUN.GA.98.L14 275
SYK.KE.x.KE51 260
SYK.KE.x.SYK173_COMGNM 253
TAL.CM.00.266 252
TAL.CM.01.8023 268
TAN.UG.x.SIVagmTAN1 270
VER.DE.x.AGM3 279
VER.KE.x.9063 279
VER.KE.x.AGM155 274
VER.KE.x.TYO1_patent 272
WRC.CI.97.97CI14 264
WRC.CI.98.98CI04 275
WRC.GM.05.Pbt_05GM_X02 267

```
                                                                            CD4 binding       glycosylation NNT
                                                                                      glycosylation NCT              V3 tip
                                                            glycosylation NFT                                          V3
                                        glycosylation NGS                    glycosylation NTS                                          glycosylation NNT   glycosylation NKT                                                          CD4 binding
H1B.FR.83.HXB2                LLNGSLAEEEVVIR........SVNFT.DNAKTIIVQLNTSVEINCTRPNNN.TRKRIRIQRGPGRAFV....TIGK.IGNMRQAHCNISRAKWNNTLKQIASKLREQFGNNKTII................................FKQSSGGDPEIVTHSFNCGGEF  382
H1O.CM.91.MVP5180             I---T-SR-K..--......IMGK-I-.ES--N---T---PINMT-I-EGIA.EVQD-YT..--M-WRSMTLKRSNNTSPRS-V-Y-TYNKTV-E-A-Q-T-I.RYL..NLVNQT.....ENVTIIF.....SRT......................-A-VSHLH---H---  383
H1N.CM.95.YBF30               I-----NTDGI---............N.-SHSNLL--W-ET-P------G--.-GGQVQ-..--AMT-Y....N-E-IV-DI---Y--V-KEL-EPMWNRTREEIKKIL-K-N............................ITFRARERNE--L-VTHLM---R---  381
H1P.CM.06.U14788              I---T--KG-PLVI........TQ-VS.-TG-V---K--E--SLT-I--G--.-GQVQL..-V.MTWY....NMKHYV-DI-A----V--EN-KR--EWVSEAIWKAYPQHSNHTSK.......................HNHTFV--N-T-----VSSLH-S-H---  394
CPZ.CD.06.BF1167              H----VE-G-TKAY........Y-EAKHNAPLL-KFNKA-VFN-T-V--G--.--GQVQ-..---MT-Y....N-RNV--DT-K-F-YVNKTL-E-ATEAAKVAIN-TLK-TTFKN.....................TTYTFRTDGDL-WI----Q-TQ--W---Q---  393
CPZ.TZ.00.TAN1                H-----E---TKAY........FVN-.SVNTPLL-KF-V-INLT-E-TG--.--GQVQ-..---MT-Y....N-ENVV-DT-K-Y-SVNATT-YRN-DWAMAAINTTMRARNETV...............................QQT-QWQRD----VTSFW---Q---  380
GOR.CM.07.SIVgorCP2135con     I---T--K-HAT-I........-K-S-.-SG-D---K-AR----T-E-TS--.--GQ-QV..--MTIY....NSENIV--T-K-F-KYNKTN-Q-A--DTVRA-KANQIRFQ......................................N-------VTFLH---H---  363

                                                                                            V3 tip
                                                                                       V3
MAC.US.x.239                  GF--TR--NRTY-Y........WHGR.--RTI-SLNKYYNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINDRPK--W-WF.GG--KDAI-EVKQTIVKHPRY...TGTNNT.DKINLTAPGG........................----VTFMWT--R---  399
H2A.SN.86.ST_JSP4_27          GF--TR--NRTY-Y........WHGR.--RTI-SLNKFYNLTVH-K--G-K.-VVP-TL..MS-LV-.....HSQPINRRP---W-WF.KGE-KEAM-EVKLT-AKHPRY...KGTNDT.EKIRFIAPGER....................S---VAYMWT--R---  388
H2B.CI.x.EHO                  GF--TR--NRTY-Y........WHGK.--RTI-SLNSYYNLTMH-K--G-K.MVVP--T..VS-IL-.....HSQPINKRPK--W-WF.KGN-TEAIQEVKETIKNHPRYSG-TNI.......................SQIRLAEHARSS---VRYMWT--R---  392
H2G.CI.92.Abt96               GF--TR--NRTY-Y........WHGR.--RTI-SLNKYYNLT-X-R--G-K.-VLP-T-..MS-LV-.....HSQPINERP---W-WF.GG--KEAMQEVKETVVKHPSY...KGTNDTKXITFTTPGEG....................S---VKFMWT--R---  400
H2U.CI.07.07IC_TNP3           GF--TR--NRTY-Y........WHGR.--RTI-SLNKYYNLTVR-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGS-KEAIQEVKET-VKHPRYTGINDTKKINLTQ.........................PG-----ATFMWT--R---  402
H2U.FR.96.12034               GF--TRS-NRTY-Y........WHGR.--RTI-SLNKYYNLTMY-R--G-K.-VLPVT-..MS-LV-.....HSQAINN-PK--W-WF.GGN-KGAIQEVKETIANHPRY...HGTKNI.SQINLAEPAG.......................--S-VKFMWT--R---  395
ASC.UG.10.RT03                Q---TYEKNTDEVQV.......MRKYNSNYSVAF-FSESQI-NLT-V--G-K.SIRNLQ-..-A-MT-Y....SQLIVG--T-R-Y-K-NGTQ-Y-A-NATYRAIKKEYNLTRDQPITWRSQPR..............................--R-VES-W-Q-Q---  396
ASC.UG.10.RT08                Q---TYEKNTDKVQV.......MRKYNSNYSVAF-FSESQI-NLT-V--G-K.SIRNLQ-..-A-MT-Y....SQLIVG--T-R-Y-K-NGTQ-YSA-NATYRAIQKEYNLTENQSITWRSQPR..............................--R-VES-W-Q-Q---  399
ASC.UG.10.RT11                Q---TYEKNTDEVQV.......MRKYNSNYSVAF-FSESQI-NLT-V--G-K.-IRNLQ-..-S-MT-Y....SQLIVG-DT-R-F-R-NGTQ-YHA-NATYRAIKKEYNLTQNQPITWRSQPR..............................--R-VES-W-Q-Q---  389
COL.CM.x.CGU1                 -V-A-KELSDWAKD......REG-WKN.-SGTIEYYWFPKDIALG-I-RG-S.SHRNLNT..AN-AK-Y....YELIPYSKGIYGR-QF.VPMTGQNK-NKTQFAI-IKK-LTAWLERISRKNI..................TITPRNGNRTS---ATFTFVI-HRL-  392
COL.UG.10.BWC01               -V-A-EELK-WSKD......REG-WKN.ESGTFEYYWFPKDIALG-I-RG-S.SHRNLNT..AN-AK-Y....YEVRPEADRWKGR-QF.VPMNQKNMTKK.EFAGKIVE-LTNWLQNITEYGN.................ITIKPRNGNKTS---ATFTFLI-HRL-  399
COL.UG.10.BWC07               VV---KTLVDYAKT......REGIWAN.ESGNIEYYWFPKDLYLG-K-RG-V.-NRNLNT..AN-AV-Y....YELRGQDADWK-R-QF.VKIGM-KTTLAQEIKKNLTYWL-N-TSYDNITIK...............PRNGNKTSDPEATYTFITFHGMFLY-NASS  395
DEB.CM.04.SIVdeb04CMPF3061    QF---DE-RAEE-HV....IRK..GLE.NKTI--RIPPEYNLRLD-V--G-K.-YRATHM..AT-LS-Y....-TFVERLRIKR---RM.-GN-SGAVSD-KQV-KGIY...NVTN............................ITMHYP---R-VQRVW-Q-H--Y  375
DEB.CM.99.CM40                QF---DE-RAEELHI..IRKEVKGEVQ.NGSI--R-PAKYNLTLT-V--G-K.-YRA-HM..AT-LS-Y....-TFIQRLRIKR---RL.NGS-A-AT-EMRQ-IL-I--KANRTNN.........................LTIHYPK--R-VQSVW-Q-H---  394
DEB.CM.99.CM5                 QF---DE-RAED-HI....IRK..GLE.NKTI-VRIPPKYNLSLD-V--G-K.-YRAVHM..AT-LS-Y....-TFIPRLRIKR---RM.-GN-TGAV-E-R-R-K-IY...NVTK............................ITVHYP---R-VQNTW-Q-H---  376
DEN.CD.x.CD1_CMO580407        IM-STVELLDNHTQI.....IRNTSKD.NQTIA-LFRKSSVIKLK----G-V.-L-GLNM..AA-IVMR....ATIIPRRAI-R-A-K-VGGN-TR-RQE-EKAVK-E-NI-G-VKW...........................DVPRV-----TSSAR-Q-R---  390
DRL.DE.11.D3                  GF--TKHDQDEL-ET.....RGRPE-V.-HKYVFR-NKKWNLK-M-R-KG-R.SVISTPS..AT-LL-Y....HGLEPGK-LNRGM-KF.EGQ-GKA-ESL-KE-KAVNASIWRNTTIKGKECKRLKETKD..TKTVGTGCMLKQLKVDNYTTK--LASENLMML----Y  437
DRL.x.x.FAO                   GF--TKHGQDEL-ET.....RGKPD-L.NHKYVFR-NKKWNLK-I-R-KG-R.SIISTPS..AS-LL-Y....HGLEPGK-LKRGM-KF.EGQ-GKA-GSL-KE-KSVS-SIWRNTT..NSTCISLKGKKI..QDRNVTGCMLKQLKVDNYTTK--LASENLMML----Y  433
GRV.ET.x.GRI_677_gri_1        G----RS-NRTE-W.......QKGGND.NDTVI-KLNKFYNLTVR-R--G-K.-VLPVT-..MA-LV-.....HSQ-YNTRLK--W-HF.QGD-KGAW-EVREEVKKVKNLTEVS-E..........................NIHLRRIW.-----SANFW---Q---  386
GSN.CM.99.CN166               H---TYQPGNNTRIM.......MNGMK.NESIV-GFGEDYHLNLT-I--G-K.-IRNLQ-..-A-MT-Y....SQVIVG--T-K-Y-RLDPDR--RAIREAMKAMN-HWE-KTG.........................RNDTQIRWTSEPK--L-VQ--W-Q-Q---  398
GSN.CM.99.CN71                H---TYQPGNNTRVM.......MNGKK.NESIV-GFGEDYQLTLT-I--G-K.-L-NLQ-..-A-MT-Y....SQIIVG-DT-K-Y-KLNHT--DIAIR-AMKAMKNHWEKINNDTT.......................PNKTQIRWTSEPK--L-VQ--W-Q-Q---  403
LST.CD.88.SIVlhoest447        GF--TNHTDDEL-PL...TPKKMG-LL.EAKFVYK-AGKWGLI-R-I-KG-R.SEVSTIS..ST-LL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFHNLGNM-KRLNATAMNYTHGNCIDNT...........KPCGRQLKGLPIANMTRR--NLATEMLMHT--E-M  423
LST.CD.88.SIVlhoest485        GF--TNHT-DKL-PL...TPKKMG-LL.-AKFVYK-SGKWGLI-R-I-KG-R.SEVSTIS..ST-LL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFNNLGKM-QKLNATAMNYTNGQCPEGK...........QPCGRHLKGLPIANMTRR-EGLATEMLMHT--E-M  422
LST.CD.88.SIVlhoest524        GF--TNHT-DEL-PL...TPKKMGDLN.GAKFVYK-AGKWGLI-R-I-KG-R.SEVSTIS..ST-LL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFNNLGRM-KKLNETAMNYKEGNCSTPG...........KPCGRQLKGLPIANMTRR-V-LATEMLMHT--E-M  425
LST.KE.x.lho7                 GF--TNHTRDEL-PL...TPNKMEDLN.GAKFVYK-AGKWGLI-R-I-KG-R.SEVSTIS..ST-YL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFNNLGKM-K-LNAEAMNYTEGTGTCDSKK.........TTCGRKLKGLPIANMTRH-A-LATEMLMHT--E-M  420
MAC.US.x.251_1A11             GF--TR--NRTY-Y........WHGR.--RTI-SLNKYYNLT-K-R--G-K.-VLPVT-..MSQLV-.....HSQPINDRPK--W-WF.-G--KDAI-EVKQTIVKHPRY...TGTNNT.DKINLTAPGG........................----VTFMWT--R---  398
MND-1.GA.x.MNDGB1             GF--TMHK.-GELIPIDDKYRGPEE-H.QRKFVYK-PGKYGLK-E-H-KG-R.SVVSTPS..AT-LL-Y....HGLEPGK-LKKGM-TF.KGR-GLA-WSL-KE-NKLNDSI-VNQTCKNFTSTGEENKQNTDKQKEFAKCIKTLKIDNYTT-.--RAAEMMMMT-Q--M  417
MND-2.CM.98.CM16              GSS-TKHDYNEL-QT..NTKKGKEE-H.-HKYVYR-DEKYNLQVV-R-KG-R.SVISTPS..AT-LI-Y....SGLEPGK-LKKGM-QL.KGQ-GIAMHDL-IE--KIDSSIWRNVT.KGYCKGLKRK.......ENRTGCALKTIKVSDYTTK-E-GAE-IMLL----Y  412
MND-2.GA.x.M14                GF--TKHNYDEL-LV..NPQ...KE-H.-HKYVYR-NKKWGLKVV-D-KG-R.SIVSVPS..AS-LI-Y....HGLEPGR-LKKGM-RL.IGQ-GRA-NALSRE-KKVNASIYRGFNASGPCRGINNR.....GVVNKTGCALKTIEVSNYTTI-E-GAE-IMIL----Y  412
MND-2.x.x.5440                GF--TKHDYDEL-QT..NPRKGKDE-H.-HKYVYR-DKKWGLQVR-R-KG-R.SIISTPS..ST-LL-Y....HGLEPGK-L-KGK-QL.EG--GQA-HSLSLE--KINDSIYKDNH.NMTCKSSNNK.......KNTTGCHLKTISISESTVK-E-GAE-IMLL----Y  433
MNE.US.x.MNE027               GF--TR--NRTY-Y........WHSK.--RTI-SLNKYNNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-RF.EGN-KEAI-EVKQTIVKHPRY...TGTNNT.DKINLTAPRG........................----VTFMWT--R---  400
MON.CM.99.L1_99CML1           Q--STGNAPNTTVMM........NKQK.NESIVVRLAKHLH-N-T-I--G-K.-IRNLQ-..-A-MT-Y....SQLIVG--T-K-Y-KVNKTQ-ETA-QAVHEAVKTEWEKKNNGTN..........................VTTISWRFQPQ--K-VQ--W---Q---  384
MON.NG.x.NG1                  Q--ATS-TNXTTVMM........NGKR.NESIV-RLXKTL--N-T-I--G-K.-LRNLQ-..-A-MT-Y....SQIIVG--T-K-Y-QVNKTQ-DAA-HAAHXAIXXEWESKNNNVT..........................VANITWRYQPK--N-VQ--W---Q---  378
MUS-1.CM.01.CM1239            Q---TYL-GNNTAVM.......MNGDK.NESIG-KFGE-FR-NLT-L--G-K.-IRNLQ-..-A-MT-Y....SQLIVG-DT-R-Y-K-NKTQ-DVA-R-AMLAM-KHWHGRLEKQGE.NITEE..................KIKIRWTSQPK--N-VQ--W---A---  401
MUS-1.CM.01.SIVmus_01CM1085   Q---TYL-GNNTAVM.......MNGDK.NESIG-KFGE-FR-NLT-I--G-K.-IRNLQ-..-A-MT-Y....SQLIVD-DT-R-Y-K-NKTQ-DPA-R-AMLAM-KHWHGRLQKQNHSSINET..................KIKIRWRSEPK--Q-VQ--W-------  407
MUS-2.CM.01.CM1246            Q---TYLPGNDTAVM.......MNGDK.NESIA-KFGE-FR-NLT-I--G-K.-IRNLQ-..-A-MT-Y....SQLIVD-DT-R-Y-K-NKTQ-DPA-R-AMIAM-DHWH-RLQKANK.TINKT..................EIKIRWTSEPK--R-VQ--W---Q---  405
MUS-2.CM.01.CM2500            Q---TYLPGNDTAVM.......MNGKK.NESIA-QFGE-FR-NLT-L--G-K.-IRNLQ-..-A-MS-Y....SQLIVD-DT-R-Y-K-NKTQ-DPA-R-AMKAM-AHWA.KMD.....NRSES..................DIKIRWTSEPK--R-VQ--W-------  388
MUS-3.GA.09.09GabOI81         Q---T-QPGNDTAVV.......MNGKK.NESVV-RLGE-YK-NLT-I--G-K.-IRNLQ-..-A-MT-Y....SQLIVG-DT-R-Y-K-DKTQ-DAAIR-SMIAM-DHWI.KINKNDSTHNYTK.................ENITIRWISQPK----VQ--W-A-D---  402
MUS-3.GA.11.11GabPts02        Q---TYQ-GNNTAVM.......MNGKK.NESIVVRLGE-HR-QLT-I--G-K.-IRNLQ-..-A-MS-Y....SQLIVD-DT-R-Y-K-DKD--D-AIR-VTLAV-DHWV.KINNGTNA.....................EKVTMRWTSQPR--R-VE--W-A-D---  404
OLC.CI.97.97CI12              GF--TRHKASRL-PI.TGDLNHKPREG.TGVYVMAIADKYKL--E-I-KG-R.SIVSVNS..AS-LVYY....AGLEPY--IDKGL-RF.KGQ-GWM-YALGKG-EKENNTWANYTGVNITGIGR.....NSTTFAKAIKR...FKFDE-LKR---DAAENMMMM---YM  413
RCM.CM.00.SIVagi_00CM312      G----IS-NRTW-Y........QRKQ.S-RTI-GLNSFYNLTVS-R--S-R.-I-G-SL..AT-.V-.....ISL-VEKRPKG-W-RF.YGN-KGAW-EVKE-VIQTK-YKG-NDT.........................TKINIRSVY---D-TKYFWL--D---  385
RCM.CM.02.SIVrcm_02CM8081     G----IS-NRTW-Y........QRKQ.S-RTV-GLNREYNLTVE-R--S-RTVKGIS.L..AT-.V-.....ISLRVEKRPKG-W-RF.YGN-TGAWEEVKKEVIKTK-YKG-NDT.........................KKIYIRSHY---D-AKYFWL--D---  394
RCM.GA.x.SIVRCMGAB1           G----INVN-TW-Y........QRRQ.S-RTV-GLNSFYNLSVT-R--S-R.-V-G-SL..AT-.V-.....ISLRVEKRPKG-W-RF.EGN-TDAW-EVKERVKTTK-YRG-SNT.........................DKIKIRTVY---D-ARYFWL--N---  396
RCM.NG.x.NG411                G----IS-NRTW-Y........QRKQ.S-RTV-GLNRHFNLTVT-R--S-RTVKGIS.L..AT-.V-.....ISLRVQKRPKG-W-RF.DGD-KGAW-EVKEAINKTK-YKG-RQP.........................ENITIRSVY---D-AKCFWL--D---  390
SAB.SN.x.SAB1                 GI---YVANRTE-W.......QKNGNS.NDSVI-RLNRYFNLT-R-R--G-K.-VLPVT-..MA-LV-.....HSQ-YNTRLK--W-WF.GGN-RGAW-EVKETIVRLPPKKYSGTN........................DTNKI-L-RQW----SEFFF---Q---  396
SMM.CI.79.SIVsmCI2            GF--TR--NRTY-Y........WHGR.--RTI-SLNKYYNLTMR-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GG--KEA--EVKET-VKHPRY..KGGTNDT.RKINFTEPGRK.....................S---VAFMWT--R---  409
SMM.LR.89.SIVsmLIB1           GF--TR--NRTY-Y........WHGR.--RTI-SLNKYYNLT-R-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGS-RKAIQEVKET-VKHPRY...TGTNDT.NKINLTAPGG........................----VTFMWT--R---  396
SMM.SL.92.SIVsmSL92A          GF--TR--NRTY-Y........WHGR.S-RTI-SLNKTYGL-LH-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.KGE-KKAIREVKETIVQHPRY...TGTNRT.EKIYLTAPGG........................----VTFMWT--R---  401
SMM.SL.92.SL92B               GF--TR--NRTY-Y........WHGG.S-RTI-SLNKYYNLT-K-R--G-K.-VLPVT-..MS-LV-.....HSQPINERP---W-WF.GG--REAMQEVKKTIVKHPRY...TGTNDT.RKINLTAPGG........................----VTFMWT--R---  402
SMM.US.04.G078                GF--TR--NRTY-Y........WHGR.--RTI-SLNKFYNLT-R-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGS-KKAIQEVKETIVKHPRY...TGTNDT.EKINLTAPRG........................----VTFMWT--R---  406
SMM.US.05.D215                GF--TR--NRTY-Y........WHGK.S-RTI-SLNKYYNLT-R-R--G-K.-VLP-T-..MS-LV-.....HAQPINERP---W-WF.GG--KQAI-EVKET-VKHPRY...TGTNDT.SKIKLTKPGG........................----VTFMWT--R---  400
SMM.US.06.FTq                 GF--TR--NRTY-Y........WHGR.--RTI-SLNKYYNLT-K-R--G-K.-VMPVT-..MS-LV-.....HSQPINDRP---W-WF.GG--KSAMQEVKQT-VKHPRY...TGTNDTKNINFTEPGKT.....................S---VAFMWT--R---  401
SMM.US.86.CFU212              GF--TR--NRTY-Y........WHGH.R-STI-SLNSYYNLT-R-R--G-K.-VLPVT-..MS-LV-.....HSQPINERP---W-WF.GG--KSAI-EVKET-VKHPRY...TGTNDT.SKINLTAPGG........................----VTFMWT--R---  402
SMM.US.x.H9                   GF--TR--NRTY-Y........WHXR.S-RTI-SLNKYXNLTMR-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGE-KEAIREVKXX-VKHPRY...KGTNXX.XQIKLTAPGG........................----VTFMWT--R---  402
SMM.US.x.PGM53                GF--TR--NRTY-Y........WHSR.S-RTI-SLNKYYNLT-R-R--G-K.-VLPVT-..MS-LI-.....HSQPINERPK--W-WF.GGS-REAIQEVKETIVKHPRY...TGTNET.KKINLTAPGG........................----VTFMWT--R---  404
SMM.x.x.pE660.CG7G            GF--TR--NRTY-Y........WHGK.S-RTI-SLNKYYNLTMR-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGS-KEAIQEVKET-VKHPRY...TGTNDT.KKINLTAPAG........................----VTFMWT--R---  401
STM.US.89.STM_37_16           GF--TR--NRTY-Y........WHGR.--RTI-SLNKYYNLTMS-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGE-RGAI-EVKET-VKHPRY...TGTNDT.AKIRIVAPGG........................----VTFMWT--R---  398
SUN.GA.98.L14                 GF--TRHK-DEL-PI....NNKVGQGA.EGEYVWK-AAKWGLV-Q-I-KG-R.SQVSTIS..ST-LL-Y....YGLEPGSKL-L-Q-KF.VGQ-GRAFATLGKL--QVEPLANMSNG..CTFDNITKT....CRFTNGTDFKKMIKFKPMEKH-A-AATEMLMMT--E-M  426
SYK.KE.x.KE51                 Q---TYKAKDK-RFI.......RQKDK.NESVI-L-PEKLGLQLV-E--G-E.SI-N-QL..AA-FFLPVI.QGRL-TGKAAKR-F-KV.KGE-GKFFE-VHNESIKVWK-..........................VTSTSWRSQPQ--L-VR--W-Q----  385
SYK.KE.x.SYK173_COMGNM        Q---TYKAKDK-RFI.......KQKDK.NESVI-L-PEALRLQ-I-E--G-E.SI-N-QL..AA-YFLPVI.QGKL-TGRDAKR-F-RV.TGN-TEFF--VHEQATKTWK-..........................VTNTTWRSQP---L-VR--W-Q----  378
TAL.CM.00.266                 I---TK.-NHTQVI........HQAPK.NESAI-LLGKEHNLT-T-I--G-R.-I-DLQ-..AA-LM-H....SQIIAGK-LNR-Y-KL.EGN-TKAMESANEEIAKRYKEHINNTD.........................KVNITWTAVRR--L-VQLFWHA-Q---  378
TAL.CM.01.8023                V---TK.-NQTQ-I........HQAPK.NESAI-LLGKEHNIT-Q-I--G-R.-I-DLQ-..AA-LM-H....SQIIAGKDLKR-Y-K-.EGN-TRAMEDVNE-IAKAYKDHINSSS........................GTAQIRWQGVRK--L-VSLFWHA-Q---  395
TAN.UG.x.SIVagmTAN1           GI---Q--NRTE-W.......QKHGVS.N-SVI-KLNKHYKLK-V-R--G-K.-VLPVT-..MA-LV-.....HSQQYNTKL---W-HF.QGD-KGAWREVRKTIV-LPKEKYRGTN........................NTRQIWLSRQW----AANIWL--Q---  397
VER.DE.x.AGM3                 -----YS-NRTQ-W.......QKHRVS.NDSVLVLFNKHYNLTVT-K--G-K.-VLPVT-..MA-LV-.....HSQRYNTRL---W-HF.QGN-RGAW-EVKNEIVKLPKDRYQGTN........................DTEEIYL-RLF----AANLW---Q---  406
VER.KE.x.9063                 -----YS-NRTQ-W.......QKHRVN.NDSVIVLFNKYYNLTVT-K--G-K.-VLPVT-..MA-LV-.....HSQ-YNTKL---W-HF.QGR-REAW-EVKEEIIKLPKDRYRGTN........................DTEKIYL-RQF----AANLW---Q---  406
VER.KE.x.AGM155               -----YS-NRTQ-W.......QKHGVS.NDSVL-LLNKHYNLTVT-K--G-K.-VLPVT-..MA-LV-.....HSQ-YNTRL---W-HF.QGN-KGAW-EVQEEIVKLPKERYQGTN........................DTNKI-L-RQF----AANLW---Q---  401
VER.KE.x.TYO1_patent          -----YH-NRTQ-W.......QKHRVN.N-TVL-LFNKHYNLSVT-R--G-K.-VLPVT-..MA-LV-.....HSQ-YNMKL---W-HF.EGN-RGAWREVKQ-IV-LPKDRYKGTN........................NTEHIYL-RQW----ASNLW---Q---  399
WRC.CI.97.97CI14              GF--TRHK-GEL-EV.IGNINHKPKEG.SG-FVYK-PDKYN-K-E-H-KG-R.SLVSVNS..AS-LIYY....AGLEPHR-I-KGL-KF.TGL-GKM-YDLGKN-EKYNSSWVNYTEGDCNSKN......NSADFAKCIKR...FQIRNFTRQ-A-RATENIMMI----M  415
WRC.CI.98.98CI04              GF--TRYKKGELVEI.IGNINHRPHEG.TG-FVYK-PSRYK---E-N-KG-R.SVVSVNS..AS-LIYY....AGLEPHR-I-KGL-KF.KGL-GHM-YDLGKN-Q-YNNTWVNYTE.DCNQKN......DSREFAKCIKQ...FQIRNFTRK-A-RATENIMMI----M  425
WRC.GM.05.Pbt_05GM_X02        GF--TKYDKGEL-EV.IGNINHKPKDG.TG-FVYK-PAKYN-A-E-H-KG-R.SLVSVSS..AS-LIYY....AGLEPYK-I-K-L-KF.TGQ-GHM-YGLGKE-QAYNSSWVNYTE.ECNKKG......NSSAFATCIKQ...FQIRNFTKQ-A-RATENLMMI----M  417
```

```
                                   glycosylation NST
                               glycosylation NST
                                               V4                                                                                V5                                                                    fusion peptide
                          glycosylation NST      glycosylation NNT                CD4 binding            glycosylation NIT   CD4 binding  glycosylation NES  CD4 binding                      gp120 end gp41 start
H1B.FR.83.HXB2                 FYCNSTQLFNSTWFNSTWSTEGSNNT...........EGSDTITLPCRIKQIINMWQKVGKAMYAPPISG...QIRCSSNITGLLLTRDGGN......SNNESEIFRPGGGDMRDNWRSELYKY.KVVKIEPLGVAPTKAKRRVVQ.......REKRAV..GIGALFLGF  522
H1O.CM.91.MVP5180              ----TSGM--Y.T-INCTKSG...CQEIKGSNETNKNGT.....I--KLR-LVRS-M-GESRI-----P-...NLT-H-----MI-QL-QP......WNSTGENTL--V----K-I--TK--N-.---Q-K-FS-----MS-PIINI..HTPH-------..-L-M----V  530
H1N.CM.95.YBF30                ----TSK---EELL-E-GEP.....................-------R--V-L-TR---GI-----R-...VLN-T------V-EYS--P.......DTKET-VY-S--N-VNL--Q-------.---S---I----G-----T-S.............-A.F-L-------  510
H1P.CM.06.U14788               ----TSS---FSYTWNDTTWRAAG-.......HNT...SENI.--S--L--VV-S-MR--SGLF-T--E-...ELK-Q------M-E-ELPYN.....DSSKNTTLS-I----TNI------P-.---QVKA-A-----IS-PTIMA..HDGH-K---A..-L-M------  540
CPZ.CD.06.BF1167               ---DLKNWT-NGIEID.........................GMLIA---LR--V-H-GI-SRGI-L--RE-...-VK-V-S---FIM-AETNDTDG........T-TPTFSAKVE-Y-KV-MSR-.---E-Q--S-----G--P--GDP.EAKS-S--E..I-M-MI-M--  524
CPZ.TZ.00.TAN1                 ----L-NWT-TWTA.NRTNNTHGTL..................VA---LR--V-H-GI-S-GV-L--RR-...TVK-H------IM-AEKD-N.........NSYTPQFSAVVE-Y-KV--AR-.---E-Q--S---RPG--PEIKAN...HT-SR-D-..---L-----  513
GOR.CM.07.SIVgorCP2135con      ---DT-KM--YNCTKESCDCITG-C................-NYI--HL--VVMS-MR--SGLF----R-...TL--K-----II-Q--VPLNK.TNENSTDYNTL--I----TNI------P-.----VKA-S------R-P-IS.....HN-----A..-L-M-----  505
                                               V4                                                    V5                                                         gp120 end gp41 start
MAC.US.x.239                   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--YTTGG....TS-N--G-.FVL-..-----  535
H2A.SN.86.ST_JSP4_27           L---M-WFL-WVENRTNQTQ.....................HNYV--H------T-H----NV-L--RE-...-LT-N-TV-SIIANI---E..........NQTNITFSAEVAELY-L--GD-.-LIEVT-I-F---PV--YSSA.....PV-N--G-.FVL-..-----  515
H2B.CI.x.EHO                   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--YSSV.....TP-N--G-.LVL-..-----  519
H2G.CI.92.Abt96                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--YSSV.....TPKN--G-.FVL-..-----  536
H2U.CI.07.07IC_TNP3            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--YSSA.....TP-N--G-.FVL-..-----  540
H2U.FR.96.12034                L--KM-WFL-WVENR-MEGQ.............TSR-RHKRNYV--H-R-V--T-H---RNV-L--RE-...DLT-N-TV-SMIANI-RI-K.........NQTNISMSAEVAELY-L--GD-.-L-E-R-I-F---PV--YSSV.....TP-N--G-.FVL-..-----  530
ASC.UG.10.RT03                 ----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--YEPPPS..KSVVKRA-A.LTL-..-----  539
ASC.UG.10.RT08                 ----LSRI-QK-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--YEPPPS..KSIVKRA-A.LTL-..-----  542
ASC.UG.10.RT11                 ----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--YEPPSNT.KSVVKRA-A.LTL-..-----  533
COL.CM.x.CGU1                  ----ASS-WKHDS............................PVMN-T-RKLV-S-VTHARIL-G--PG-...HLQ-NWEKQPVIAFMGTIEG....DNDGNGCAYPAAPNFKHALSTL--GR-.-L--MRTTTYV--DI--S-NVN..WHHG-Q--GI.FAFS..I-AL  520
COL.UG.10.BWC01                ---DASS-WAKNSSVMN..............................-T-RKLV-S-VSHARLL-G--PD-...HLQ-NWEKQPVIAFMGTVEK....GG-ETGCAYPAAPNFKHALSTL--GR-.-L--MRTTVY---DV--Q-AIN..WHHG-Q--GI.FAFS..I-AL  527
COL.UG.10.BWC07                L.......WSG-T............................TVMN-T-RKLV-S-VTHA-IL-G--LK-...TLQTYWEKQPV-AFMGSVEK....GE-KTGCAYPAAVNWKHAASTL--GN-.-L---T-TAFV--DHR-GGTA.....VV----GI.FAFS..LIAL  513
DEB.CM.04.SIVdeb04CMPF3061     ----VSKAYQLFTND...................TITW-NNTLM--K-N-LV-T-Y-I-QHI-L--RE-...EL--D-HVSAIIFDVAHYGH..........MLNLTPSS-IKAV--AD-----.IIEVQ-I-F---AVR-YEGPE....SV-Q----GMAL-..LVA-  505
DEB.CM.99.CM40                 ----ISKALDLLLLQNNT................RNSTW--KWLM----N-LV-T-YT--QHI-L--KE-...ELK---H-SAFVFDV-HY-G..........S-TLTPSA-I-AV--AD-F--.IIEVK-I-F--SAVR-YEGPE....SV-H---AGIAF-..LVA-  528
DEB.CM.99.CM5                  ----VSKA-..NLLMNKT................NNITW-NNTLM--K-N-LV-T-F---QHV-L--KE-...ELK-H-SVSA-VFEVAHYGQ..........NMNITPSS-IKAV--AD-----.IIEVK-I-M---SIR-YEGPE....SA-Q---AGIAF-..LVA-  508
DEN.CD.x.CD1_CMO580407         ----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---SVR-YEGPESAESVS-RR-E-SLVL-..LI--  529
DRL.DE.11.D3                   -F--ISKIWRTWNNR-SNVW.....................YPYAS-H----VDD-AR--RKI-M--V--FNNE---TQDV-EMFFEIQKLETN..EINDSNYQ-KFVPQDEVQNQFTAIGAH-.-L--V--I-F---DVH-YNLPE.....H-Q--GA.VVL-..I--L  575
DRL.x.x.FAO                    -F--VSKIWRTWNNRTSNVW.....................YPYAS-H-----DD-AR--RKI-M--V--FNNE---TQDV-EMFFEIQKLETE..ETGDSNYQ-KFIPQDEVQNQFTAVGAH-.-L--V--I-F---DVH-YNLPG.....H-Q--GA.VVL-..I--L  571
GRV.ET.x.GRI_677_gri_1         ---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---DVR-YTGPT.......----VP.FVL-..-----  532
GSN.CM.99.CN166                ----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---KEHP.......AV-----.LSL-IS--T-  534
GSN.CM.99.CN71                 ----LSV--QFNNNITAVDDTNI--..........VTSKYKGQWMA---R-FVTQ-GY-SRSI-L--RQ-...H-N-T-------IDGAMYGH..........S-NMTPSA-VA-A-KY--SR-.---E-D--SL---S-Q--PHPG.....VHVK----.ISL-IS--T-  543
LST.CD.88.SIVlhoest447         -F--V-RI-QEWNNKNSDKW.....................YPWAN-H--SV-DD-AS---KI-L--T--FNNR---THRV-EMWFQMEKFEP..EENLGGNLSVRFLPPSWETNQFVATGAH-.-LIRLR-I-F---DVH-YAP......SSKQ---APLAL-..A--L  561
LST.CD.88.SIVlhoest485         -F--I-RILQEWNNKNSDKW.....................YPWAN-H--SV-DD-AT---KI-I--T--FNNR---THRV-EMWFELEKWEP..EENLGGNLSVRFLPPSWETNQFVATGAH-.-LIR-R-I-F---AEHQYAP.....SSKQ---APLAL-..A--L  560
LST.CD.88.SIVlhoest524         -F--V-RI-QEWNNKNSDKW.....................YPWAN-H--S--DD-AT---KI-L--T--FNNR---THRV-EMWFEMEKWEP..HEDLGGNLS-KFLPPSWETNQFVATGAH-.-LIRLR-I-F---SEH-YAP.....RG-QT--APLAL-..A--L  563
LST.KE.x.lho7                  -F--V-RI-QEWNNKNSDKW.....................YPWAN-H--S--DD-ATI--KI-L--T--FNNR---THRV-EMFFEMEKWEP..HEDLGGNLS-KFLPPSWETNQFVA-GS--.-LI-LN-I-F---DEH-YAP.....RG-QT--APLAL-..A--L  558
MAC.US.x.251_1A11              L--KMNWFL-WVEDRN-AN..............QKPK-QHKRNYV--H-R----T-H----NV-L--RE-...DLT-I-TV-S-IANI-WTDG.........NQTNITMSAEVAELY-L--GD-.-L-E-T-I-L---NVR-YTTGG....TS-N--G-.FVL-..-----  534
MND-1.GA.x.MNDGB1              -F--V-RIMRAWNDPNEKKW.....................YPYAS-Q-R--VDD-MQ--RKI-L--T--FNNH---THRV-EMYFEMQKIDS.....NETKMQ-KFLPPSETSNQFVAYGAH-.-L---M-I-I---DV--HTLPE....HHK---GA.VIL-..I--L  553
MND-2.CM.98.CM16               -F--W-KIWRTWNDQNSSVW.....................YPWMS-N-R---DD-H----KI-M--V--FNNE----NDV-EMFFEVQKTEE........GYI-KFVPQDEVQNRFTAVGAH-.QL--VD-I-F---EVA-YHLPE.....A-Q--GA.VLL-..MF-L  544
MND-2.GA.x.M14                 -F--W-RIWKTWNNQNSNVW.....................YPYMS-K-R--VDD-H---RKI-M--V--FNNH---TNDV-EMFFEVQLVEG........KRYL-KFLPQDEVQNQFTAVGAH-.-L--VD-I-F---DVH-YHLPD.....AKQ--GA.VLL-..M--L  545
MND-2.x.x.5440                 -F--W-KIWKAWNSKQSSVW.....................YPYMSRN-R--VGD-H----KI-M--V--FNNE---TNDV-EMFFEVQKTDDD.......GYI-KFIPQDWIQNQYTAVGAH-.-L--VD-I-F---DIH-HHLPN.....T-Q--GA.VLL-..M--L  567
MNE.US.x.MNE027                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--YTTGG....TP-N--G-.FVL-..-----  538
MON.CM.99.L1_99CML1            ----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-Q--TGV.......H-----I..TL-MA----  525
MON.NG.x.NG1                   ----LSR--I-SDPXKNKNETFESTFNVSTNASTAYNGEWL-CR-RQLVT-WGYISKSXYLPPRQGHVQC.......T----A--VDGELYGS..........TVTLIPSARVS-A-----S-.---E-D--S---A-Q--HGL.......M--R--I..TL-MA----  522
MUS-1.CM.01.CM1239             ----LSV--QFNNNITQINE.........SNIHNVTTKYHQDQWMV---R-FVTQ-GY-S-SI-L--RR-...HVQ-T--V----IDGAMYKD..........TVNMTPSANVM-A--M--KN-.---E-D--SM---EVQ--P-P.......HVKA--I..SL-IT----  540
MUS-1.CM.01.SIVmus_01CM1085    ----LSV--QFVNHTEINE..........SNIHTVTTKYHKDQWMV---R-FV-Q-GY-S-SI-L--RK-...LV--E-------IDGAMYRD..........S-NMTPSANVL-A--M-PK-VQ---E-D--SM---EVQ--P-P.......HVHA--I..SL-IT----  546
MUS-2.CM.01.CM1246             ----LSV--QFENNTREIND.........SNIHTVKTKYRHDQWMV---R-FVTQ-GY-S-SI-L--RK-...HVQ-T-HV----IDGAMYGN..........T-NMTPSANV..WT-ASRLRD.T--E-D--S---EVR--P-P.......HVKA--I..SL-IT----  542
MUS-2.CM.01.CM2500             ----LSV--QFENDTRQIHN.........SNIDTVTTKYHKDQWMV---R-FVTQ-GYTS-SI-L--RQ-...H-Q-T--V----IDGAMYGD.........TVNRLTPSANVL-A--M--KR-.---E-D--S---EVQ--P-P.......HVQA--I..TL-IT----  528
MUS-3.GA.09.09GabOI81          ----LSA--QFENNVTEINSG........NIDNVTSKYKENQWMT-K---FVTQ-GFTS-SI-L--KR-...H-Q-T--------IEGAMYRD..........S-NMTPSANVM-A--A--SR-.---E-D--S---RVP--EAP.......RLHT--I..SL-IT--T-  541
MUS-3.GA.11.11GabPts02         ----LSV--QFDSVREITDE.........NIHNLTSKYKKNQWML---R-FVTQ-GYTS-SI-L--KQ-...H-N-T-----M--ENAMYGN..........S-NMTPSANVI-A--A--SR-.---E-D--SM---TVQ--ETP.......RLHT--I..SL-IT--T-  542
OLC.CI.97.97CI12               WF--I-KILKTWNNA-DPKW.....................YPWSQ--LR-NVDD-M---RRI-L--QP-FNNR---.....QRVTEMFLEIQ....KVE-GTQVSLLPPS-NVNQFV-VASD-.-L-EL-LI-F---ETQ-YQP..........-N--GA.LVL-..L--V  542
RCM.CM.00.SIVagi_00CM312       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--YET........KQ--VP.LVL-..-----  515
RCM.CM.02.SIVrcm_02CM8081      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--YET........-Q--VP.LVL-..-----  524
RCM.GA.x.SIVRCMGAB1            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--YQT........KQ--VP.LVL-..-----  526
RCM.NG.x.NG411                 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---NVR-YET........KQ--VP.LVL-..-----  520
SAB.SN.x.SAB1                  ---KMDWFL-YLNNK-VDPD..H--CAKNN.....TKPCWQRTYV--H-R-VV-D-YTLS-KT----RE-...HLE-N-TA-A-YVELNYNSK.........NRTNVTLSPQIESI-AN--GD-.-L-E-K-I-F----VR-YTG......PE-Q--VP.FVL-..-----  537
SMM.CI.79.SIVsmCI2             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--YSSAT....PA-N--G-.FVL-..-----  545
SMM.LR.89.SIVsmLIB1            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--YST.......M-P--G-.FVL-..-----  534
SMM.SL.92.SIVsmSL92A           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--YTTST.....S-N--G-.FVL-..-----  541
SMM.SL.92.SL92B                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--YTTS.....TSKN--G-.FVL-..-----  542
SMM.US.04.G078                 L--KMNWFL-WVEDRNEKGS.........IWEQQNKK-QKKRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANI-WID.........NNQTNITMSAEVAELY-L--GD-.-L-E-T-I-M---SV--YTTGT....S-N--G-.FVL-..-----  546
SMM.US.05.D215                 L--KMNWFL-WVEERNATEM.........NARNKNNR-KKKHNYV--H-R----T-HR---NV-L--RE-...DLT-N-TV-SIIAEI-WL-.........NN-TNITMSA-VAELY-L--GD-.-L-E-T-I-L---SV--YTT-.....TS-QR-G-.FVL-..-----  540
SMM.US.06.FTq                  L--KM-WFL-WVEDRNMS..............SQNKA-QKRRNYV--H-R----T-H----NV-L--RE-...ELT-N-TV-S-IANIYWID.........GNRTNITFSAEVAELY-L--GD-.-L-E-T-I-F---DV--YSSVT....TP-N--G-.FVL-..-----  537
SMM.US.86.CFU212               L--KMNWFL-WVEDRNRTQM.........EARNNNDK-RRKRNYV--H-R-V--T-HR---NV-L--RE-...DLT-N-TV-S-IAEI-WYD.........KNQTNITMSAEVAELY-L--GD-.-L-E-T-I-I---NV--YTTTT....TS-A--G-.FVL-..-----  543
SMM.US.x.H9                    L--KMNWFL-WVEXIKNG-..........RWTSQNQK-RYQKNYV--H-X----T-H----NVXL--RE-...DLT-N-TV-S-IAEI-WI-G.........NKTNITMSAEVAELY-L--GD-.-L-E-T-I-F---SV--YTTTG....AS-N--G-.FVL-..-----  542
SMM.US.x.PGM53                 L--KMNWFL-WVEDRN-SGS.........LWTQQTKADEKKRNYV--H-R----T-HR---NV-L--RE-...DLT-N-TV-S-IADI-WIDG.........NQTNITMSAEVAELY-L--GD-.-L-E-T-I-L---SVR-YTTGA....S-N--G-.FVL-..-----  544
SMM.x.x.pE660.CG7G             L--KMNWFL-WVEDRDQK-S.........RWRQQNTR-RQKKNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IAEI-WT-.........NN-TNITMSAEVAELY-L--GD-.-L-E-T-I-L---SVR-YTTT....GAS-N--G-.FVL-..-----  542
STM.US.89.STM_37_16            L--KMNWFL-WIENR--SE.........MRDWNKNKK-QQKRNYV--H-R-V--T-H----NV-L--RQ-...DLT-N-TV-SIIANI-WT-N.........N-TNITASAEVAELY-L--GD-.-L-E-T-I-L---NV--YTTS.....TS-T--G-.FVL-..-----  538
SUN.GA.98.L14                  -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---DEH-YAP.......RK----APVAL-..A-AL  564
SYK.KE.x.KE51                  ----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---SQR-YELP.....TKQ---APLAL-.----L  517
SYK.KE.x.SYK173_COMGNM         ----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---DQR-YELP.....NT-----APLAL-.----L  511
TAL.CM.00.266                  ----ISSM-LN-SYEYNYTSNQGKLHQGVIGK.KPRSSDRTKFYMS-QLR-M--R-SR-E-L--L--RE-...HLQ-K-------VDINHHRG........S-D-TVEPSSQI--S--AAVSR-.--IE-V-F-Y---NIR-YDGP.....ANKQ---APLAL-..-I--  528
TAL.CM.01.8023                 ----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---NIR-YDGP.....ATK----APLAL-..-I--  546
TAN.UG.x.SIVagmTAN1            ---TPDWFV-WLNNE-NSGR..NVDVEGNNCTTGKDKRCYKRTYV--H-RS-V-D-YTLS-KT----RE-...HLE-T-TV-SMMVSL-YNSK.........ERTNVTLTANLENI-AY--GR-.-LIE---I-F---EIR-Y-GPT......-----VP.FVL-..-----  543
VER.DE.x.AGM3                  ---KMDWFL-YLNNRTVDPD..H-PCNGTKGKGKAPGPCAQRTYVA-H-RSV--D-YTLSRKT----RE-...HLQ-T-TV--MSVELNYNSK.........NRTNVTLSPQIETI-AA--GR-.-L-E-T-I-F---EVR-YTGG.....HD-T--VP.FVL-..-----  553
VER.KE.x.9063                  ---KMDWFL-YLNNQ-VDAN..H-QCSNETKKGNAPGPCVQRTYVA-H-RSV--D-YTIS-KV----RE-...HLE-T-TV--MTVELNYN-N.........YRTNVTLSPQIEGI-AA--GR-.-L-E-T-I-F---NVR-YTGG.....QD-S--VP.FVL-..-----  553
VER.KE.x.AGM155                ---KMDWFL-YLNNLTVDAD..H-HCKNNAGKGRSPGPCVQRTYVA-H-RSV--D-YTIS-KT----RE-...HLQ-T-TV--MTVELNYN-Q.........NRTNVTLSPQIETI-AA--GR-.-L-E-T-I-F---EVR-YTGG.....QE-Q--VP.FVL-..-----  548
VER.KE.x.TYO1_patent           ---KMDWFL-YLNNKTWDAD..H-FCSSKK.KGHAPGPCVQRTYVA-H-RSV--D-YTLS-KT----RE-...HLQ-R-TV--MTVELNYNSK.........NRTNVTLSPQIESI-AA--GR-.-L-E-T-I-F---EVR-YTGG.....HE-Q--VP.FVL-..-----  545
WRC.CI.97.97CI14               -F--I-RIADAWNNR-ETKW.....................YPWS--T-RN-VED-AH--IKV-L--V--FNNH---TQR....VTEILFNWE........RNQVMFQAPSNIQNSFVAVGQ--.-L--LKMI-M--GDG--V-PPE....HR----GA.IIL-..I---  544
WRC.CI.98.98CI04               -F--I-KIADAWNNKNESKW.....................YPWS--T-RN-VDD-AH--VKV-L--V--FRNH---TQR....VTEILFNWE........NKQVEFQAPSNIQNSFVAVGQ--.-L-RLKMI-L--GDGR-VTPKE....HH----GA.IVL-..I---  554
WRC.GM.05.Pbt_05GM_X02         -F--I-KIADWWNNKTEKRW.....................YPWA--T-RS-VDD-ASI--KI-L--V--FNNH---TQR....VSEVLFNWE........NGQVEFQAPSH-QNSFVAVGQ--.-L--LKLI-----DG--VTPPD....HH----GA.IVL-..L---  546
```

```
                                  fusion peptide                                 immunodominant region   glycosylation NAS / glycosylation NKS        glycosylation NHT / glycosylation NYT
H1B.FR.83.HXB2                LGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLL.RAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKS.............LEQIWNHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWAS.LWNWFNITN 677
H1O.CM.91.MVP5180             -S---------ATA---RTHSV-K------D---.---Q------R-S----R--R--LQ-L-TLIQN--R-NL---K-----Y-S-K--T---GRY............NDDS--DNL--QQ--QH---VS-I-YDE-QAA-D-----VKA-----E---.-----D--K 686
H1N.CM.95.YBF30               ------------I-------T----------I--.-----------SI--------KV--I---R---I-SL------T--Y-T----ET---NTS............YDT---NL--QQ--EKVR--SGV-FG---QA-E--NT--KS-----Q-D-.--S--G--K 666
H1P.CM.06.U14788              -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-QKA-E--KE--RK--------Q.--S-MD--K 698
CPZ.CD.06.BF1167              -S----------IA------G---------A---.Q-----------S---------L----K--R------L---AN----HSSLE--S--L--TLCGNKTNNYENNYEC.--SQY--QQ--E--R-ISDV-YLSL-TA-V---R-HKQ-------S-.-----D--- 691
CPZ.TZ.00.TAN1                -S----------IA--A---G---------Q---.Q-----------S---------M----K-IR-----SL---AN--V-HSS----LT-AED-TKCNHSD...AKYYDC---NL--Q----LVE-S-GT-Y--L-KA-T-----K---------S-.--D--D--Q 678
GOR.CM.07.SIVgorCP2135con     MS---------AV---------V-H------X-M-.---A---E--R-S----R--R--L--I-T--R------L---T-R---Y-N----KT-TG-N..........ETE-DD--GNM--QQ--KLVD---DT-FLE-QKA-E---V--KA--------D.--S-LD--Q 663

MAC.US.x.239                  -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--A-I-----MY--QK-NS-DV.FG---DLAS 686
H2A.SN.86.ST_JSP4_27          -TT--AA-----L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N----VT-I-K-----AQ-NS---AFRQV-H-T---VNDT.................-TPD--NM--Q--EQR-R-LEAN-SESL-QA-I-----MY--QK-NS-DV.FG---DL-S 666
H2B.CI.x.EHO                  -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--A-I-----MY--QK-NQ-DI.FS---DF-S 670
H2G.CI.92.Abt96               -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--A-I-----MY--QK-NS-DV.FG---DL-S 691
H2U.CI.07.07IC_TNP3           -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--A-I-----MYQ-QK-NS-DV.FG--LDL-S 691
H2U.FR.96.12034               -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-AA-I-----MY--QK-NN-DI.FG---DL-S 681
ASC.UG.10.RT03                -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-A-RR--E-THK-QK-GE-DN.-----D-SK 693
ASC.UG.10.RT08                -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-A-RR--E-THK-QK-GE-DN.--S--D-SR 697
ASC.UG.10.RT11                -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-A-RR--E-THK-QK-GE-DN.--S--D-SR 688
COL.CM.x.CGU1                 -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-AYDL--E--KK-A--GD-TN.WFSG-GLF- 673
COL.UG.10.BWC01               -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-AYDL--A--KK-A--GD-TN.WFSG-GLF- 679
COL.UG.10.BWC07               -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-AYEL--E--KK-A--GD-T-.WFSG-G-FS 665
DEB.CM.04.SIVdeb04CMPF3061    -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---DIWEEALQKA-EK--Q-VHA-QS-QD-D-.-----DLSR 656
DEB.CM.99.CM40                -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---DEWEGALQRA-E---R-VHA-QS-QD-D-.-----DLSR 679
DEB.CM.99.CM5                 -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---DEWEGALQRA-E---R-VHA-QS-TD-D-.-----DLSR 659
DEN.CD.x.CD1_CMO580407        -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-AALFGNVSDALKKA-A---R-VH-VQS-YD-DN.-----DLSK 686
DRL.DE.11.D3                  -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-TLDLQKAYE-EQ--IF--QK-GDLT-.WA---DF-W 726
DRL.x.x.FAO                   -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-TLDLQKAYE-EQ--IF--QK-GDLT-.WA---DF-W 722
GRV.ET.x.GRI_677_gri_1        -----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--A---ES--LDLYQK--D-SG.F-S--SLST 682
GSN.CM.99.CN166               -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-QANQ--AS-LN--MK-S--D-.--S--D-SD 688
GSN.CM.99.CN71                -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-NANH--S--MND--K-S--D-.--S--D-S- 697
LST.CD.88.SIVlhoest447        -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-AYTTELE-RNAFKK-QEFN..F---LD-LS 710
LST.CD.88.SIVlhoest485        -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-AYTTELE-RNAFKK-QEFN..F-S-LD-LS 709
LST.CD.88.SIVlhoest524        -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-AYTTELE-RNNFKK-QDFN..F-S-MDL-T 712
LST.KE.x.lho7                 -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-AYTTELE-QNKFKK-QEFN..F-S-LD-SH 707
MAC.US.x.251_1A11             -AT---A-----L---A-S-T--A------QL--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS.................-TPK--NE--Q--E-KVDFLEEN-TA-L--A-I-----MY--QK-NS-DV.FG---DLAS 685
MND-1.GA.x.MNDGB1             -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-TEMLKQAYDREQR-TY--QK-GDLT-.WAS--DF-W 704
MND-2.CM.98.CM16              --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-DNMTVNLQKAYELEQ--IY--EK-GD-T-.WAS--DF-W 695
MND-2.GA.x.M14                --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-NNMTIDLQRAYELEQ--MY--QK-GDLT-.WAS--DL-W 696
MND-2.x.x.5440                --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-SIVTNMTIDLQRAYELEQR-IF--QK-GDLNFHGLTG-DL-W 719
MNE.US.x.MNE027               -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--A-I-----MY--QK-NS-DV.FG---DLAS 689
MON.CM.99.L1_99CML1           -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-AA-E--G--QH--MKPGQ-DF.-----D-SK 679
MON.NG.x.NG1                  -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--A-E--G--AH--MK-GQ-DW.-----D-SK 676
MUS-1.CM.01.CM1239            -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-AA-K---E-QHK-QK-LE-D-.--E--D-SK 694
MUS-1.CM.01.SIVmus_01CM1085   -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-LA-R---E-QHK-QK-LE-D-.--E--D-SK 700
MUS-2.CM.01.CM1246            -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-MA-Q---E-QHK-QK-LE-D-.--E--D-SK 696
MUS-2.CM.01.CM2500            -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-LA-K---E-QHK-QK-LE-DA.--E--D-SK 682
MUS-3.GA.09.09GabOI81         -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--AEK---R-NHD-MK-AE-D-.--S--D-SH 695
MUS-3.GA.11.11GabPts02        -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-LANK---E-NHK-MK-AE-D-.--S--D-SH 696
OLC.CI.97.97CI12              --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-DDRLTIAYQE-QR-QY--QRFQELSD.WTK---LA- 692
RCM.CM.00.SIVagi_00CM312      -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-A-D-H---VH--EK-NN-GD.AFS-LRLDW 665
RCM.CM.02.SIVrcm_02CM8081     -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-A-D-H-R-VH--EK-NN-GD.AFS-LRLDW 674
RCM.GA.x.SIVRCMGAB1           -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-A-D-H---VH--EK-SN-GD.AFS-L-LDW 676
RCM.NG.x.NG411                -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-SVALQDA-E-H-R-VHD-EK-NS-GD.MLS-L-MDW 670
SAB.SN.x.SAB1                 -----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-QAHE-EQ--LDSYQK-VS-SD.F-S--DL-K 687
SMM.CI.79.SIVsmCI2            -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--A-I-----MY--QK-NS-DV.FG---DF-S 696
SMM.LR.89.SIVsmLIB1           -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--A-I-----MY--QK-NS-DV.FG---DL-S 685
SMM.SL.92.SIVsmSL92A          -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--A-I-----MY--QK-NS-DV.FG---DLAS 692
SMM.SL.92.SL92B               -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--ARL-----MY--QK-NS-DV.FG---DL-S 693
SMM.US.04.G078                -AT---A-----L--SA-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T---PNDS.................-TPN--NM--Q--E-KVDYLEAN-TEML--A-I-----MY--QK-NS-SV.FG---DLAS 697
SMM.US.05.D215                -AT---A-----L--SA-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T---PNDT.................-TPN--NM--Q--E-QVDFLEAN-TQML--A-I-----MY--QK-NS-DV.FG---DL-S 691
SMM.US.06.FTq                 -AT---A-----L--SA-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K--Q--AK-NS---AFRQV-H-T---VNDS.................-QPD--NM--Q--EKQVAYLEAN-TQQL--A-IL----MYD-QK-NS-DV.FG---DL-S 688
SMM.US.86.CFU212              -AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T-E-P-AN.................VTPD--NM--Q--E-KVDYLEAN-TL-L--A-I-----MY--QK-NS-DV.FG---DL-S 694
SMM.US.x.H9                   -AT---A-----L--SA-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----APXNS---AFRQV-H-T---PNDT.................-TPX--NMX-Q--EKQV-FLEAN-TXXL--A-I---X-MY--QK-NXXDX.FG--XDL-X 693
SMM.US.x.PGM53                -AT---A-----V-RSA-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-R---AQ-NS---AFRQV-H-T---PNAS.................-VPN--NM--Q--E-QVDDLEAN-TQAL--A-I-----MY--QK-NS-DI.FG---DL-S 695
SMM.x.x.pE660.CG7G            FAT---A-----L--SA-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AH-NS---AFRQV-H-T---PNET.................-VPN-SNM--Q--E-QVDFLEAN-TQ-L--A-I-----MY--QK-NS-DI.FG---DL-S 693
STM.US.89.STM_37_16           -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--A-V-----MY--QK-NS-DV.FG---DL-S 689
SUN.GA.98.L14                 -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-AYVTELE-QNKFKQ-QEFN..F-S-LDLSQ 713
SYK.KE.x.KE51                 -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-A-E-----VHD-TK-QE-D-.--S--DLSK 674
SYK.KE.x.SYK173_COMGNM        -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-A-E---R-VHD-TK-QE-D-.--S--DLSK 669
TAL.CM.00.266                 -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-A-E---R-VH--QK-ND-D-.--S---LSA 682
TAL.CM.01.8023                -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-A-E---R-VH--QK-ND-D-.--S---LSA 700
TAN.UG.x.SIVagmTAN1           -----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-STTLVKAYE-EQ--MDTYQK-GD-T-.W--I-DVSS 693
VER.DE.x.AGM3                 -----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--ARA-E---LDAYQK-SS-SD.F-S--DFSK 704
VER.KE.x.9063                 -----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-TEQLVKARE-E---LDAYQR-TS-SN.F-S--DFSK 704
VER.KE.x.AGM155               -----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-QARE-E---LDAYQK-SD-S-.F-S--DFSK 698
VER.KE.x.TYO1_patent          -----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-TRQLVKARE-E---LDAYQK-TS-SD.F-S--DFSK 696
WRC.CI.97.97CI14              --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-SAILVIAHD-EQA-MF--QK-SD-TA.WGS-LQWF- 696
WRC.CI.98.98CI04              --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-AH-TEQ--MF--QK-QD-TT.WGS-LQWF- 706
WRC.GM.05.Pbt_05GM_X02        --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-SDMLVQAHA-EQS-MF--QK-TD-TQ.WGS-LSWF- 698
```

```
                                                       gp41 cytoplasmic tail start
                                 transmembrane domain                                                                   glycosylation NGS                                                                                     glycosylation NAT
                                                                                                                                   *
H1B.FR.83.HXB2              WLWYIKLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPT.......PRGPDRPEGIEEEGGERDRD.RSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLG.......RRGWEALKYW...........WNLLQYWSQELKNSAVSLLNATAIA  821
H1O.CM.91.MVP5180           ------IA-I---A-I-I-VIMII-NL-KNI----Q---L-IPV-H.......RQEAET-GRTG-----G--P.KWTA-PP-F-QQLYT---TII-WT--L-SN-ISGIR-........LIDYLGL-LWI-.GQ.KTIEACRLC.GAVM---L-------TN--DTI-VS  837
H1N.CM.95.YBF30             ------IA----A-I--I--ISI-IT-IA----------L--LI--.......A-------ET-GGV--Q--G.--V---S-FS--V-E---N-LI-L----T-S---LR-TL----QSLSRGLQLLN-LRTHL...........-GI-A--GK--RD--I----T---V  817
H1P.CM.06.U14788            ------I-------II--K-LM-IGNV-RK-------V-L--LRLH.......GGPAGIAP-TD-----AGNG.-----LE-L-P-------N-VVQI-QI-VSCI-RIKDLLTI....LWIHLG.QLLNS........NCLRDCFAACG--T----Q--T----TV--S  847
CPZ.CD.06.BF1167            ------IA-I--AS-------MFIINL-GKL-----L--P-IPIQ-A.........---GQ---T--G---SG--.-WRGYPV-F-Q---E-C-Q-L-WI-QTSVN--WHLWT.....IKLIHQWSLLL--TVRQLLRRLIHLLTQFLH--ET-GILIRQHCIVCIDTL-EF  849
CPZ.TZ.00.TAN1              ------IA-I--A-------LMFIVNV-KQ-----T--FS-IPTQA.......EQD-EQ-G--AGG--G--NI.-WTPSPA-FFSIV-E---N-LIWI-QTFQNFIW-LWISLQA-KQGIISLAHSLVIVHRTIIVGVRQI....IEWSSNTYAS-RVLLIQAIDRL-NF  836
GOR.CM.07.SIVgorCP2135con   ------I----IA-------LM-IINMCH------------ILARN.......QVPAGIVPE-G----RA-S-.-----LA-F-P-------T-IVWI-RT-VI--SGIKE-TLS-IEHLTRLL--VNNLTRDC...........FAFIA--G----Q--I---DCV-VW  814

MAC.US.x.239                -IK--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIQQ..D-AL-TREGK-RDGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRVYQI-QPILQRLSATLQRIREVLRTE.........LTY---GWSYFHEAVQAVWRSATET  844
H2A.SN.86.ST_JSP4_27        -IK--QYGVY--V-IIV----IY-VQMLS-L-K--R-VFSSPPAYFQQIHIHK..DREQ-ARE-T-EDVGNSV.GDNWWPWPIRYIHFLIRQLIR-LNRLYNICRD-LSRSFQT-QLISQSLRRALTAVRDWLRFN.........TAY---GGEWIQEAFRAFAR--GET  824
H2B.CI.x.EHO                -MA--R-GLY--I-I-V---AIYIIQMLA-L-K--R-VFSSPPSY-QQIPIRK..DRGQ-ANE-T-E-GGNNE.GYRSWPWQIEYIHFPIRQLRD-LIWLYSGCRT-LSKTFQT-QPVLQPLRLPPA....................Y-R-GISWFQEAIQAAAR-AGET  817
H2G.CI.92.Abt96             -VK-VY-GLYV-A-VIV----IY-VQ-LG-L----R-VXSSPPSYVQQIHIRK..DQEX-TKE-I--ESGNKG.GYRSWPWQIEYIHFLXRQLGN-LTWLYSNCRA-LSRIXQT-QPLFQRISRTLQAIREHLRLE.........AAYFS-GFRW-QEACTAATR-AQET  849
H2U.CI.07.07IC_TNP3         -VK-VY-GFYV-A-II---VIIY-VQLLS-L-K--R-VFSSPP-YVQQIHTHE..DQEQ-TKE-T--DGG-NV.GFNSWPWQIEY-HFLIRQLIR-LTWLYSNFRAWLSQSFQS-QAACQRVLTAIREHLRLEGAYLQYGCQWF.QEA--AGWKAXRETLA-AWRDLWET  857
H2U.FR.96.12034             -VK-VY-GLYV-A-III----IY-VQLLGSL-K--R-VFSSHPSYVEQIPIQR..DQEQ-TKG-I-E-AG-SG.GYRSWPWQIRYIHFLIHQLIR-LTWLYSSCRD-LSRICQS-QPLF.......QSIRERLHLE.........IAY---GW-YF-EAFQAFGK-ARET  832
ASC.UG.10.RT03              -FQW--IAVL--I--IA----MWLIN-LKM--A--L--AQTPFHQPQV........-G-RGETD----DG-NV.--V--L--F--VV-----AIV-WI-RT-VN-GWLIRHLWNG-TDWWFGRAYPRLILAAEATAEATRRA...VAYIH-GL--F-AAVI-AWD-L--F  851
ASC.UG.10.RT08              -FQW--IAVL--I--IA----MWLIN-LKM--A--L--AQTPFHQPQV........-G-R-ETGG---DG-NV.--V--L--F--VV-----AIV-WI-RT-VN-GWLIRHLWSS-SDWWFGSAYPRLILAAEATAEATSRA...VAYIH-GF--F-AAVI-AWD-L--F  855
ASC.UG.10.RT11              -FQW--IAVL--I--IA----MWLIN-LKM--A--L--VQTPFHQPQV........-G-RGETGG---DG-NI.K-V--L--F--VV-----AIV-WIHRT-VN-GWLIRHLWSS-TDWWFGSAYPRLILAAEATAEATRRA...IAYI--GL--F-AAVI-GWDTL--F  846
COL.CM.x.CGU1               IFK-VLYAAYV----I----IMV-IACIRGAFRVKGFQQIGRTNVSS......QIRGQDR-WEQPDNA-EIHREQ.ENSSSRF...............WRLMEEWWSSLL-HCSNGILTVLPQHPPTTSDKRRS................A-.-T-RYL..LVPRGPSLQ  802
COL.UG.10.BWC01             IFK-VLYAAYVI--IIAF--LML-I-CIRGAYRLRGFQRLGREPLSS......QIRGQNQ-WEQPDN--ELDSEQEDNNSRLVQ...............LLKGL.WKRFAQACSNGLHTLMSQRPETFNTRRTVFQT..................-W-L..LVPRCPDLS  807
COL.UG.10.BWC07             -FK--MYAGYAIV--I----GWLLI-CFSSGFRRRKFQRLGREEELTTIL...QIPTQTQ.WEQPDNA-ELD..SDSGSSSSRF..............WLLVKEWWTHLS--CRNGIHTLQGEQPLPDGETHQPRR..............MPL-A-WRL..LVPRSANL-  799
DEB.CM.04.SIVdeb04CMPF3061  -F-W-RIVVYVIA---L---IMF-VN-ISKLCR-----QTPFQK..........EDLEL-GEG-GGV-GST-R..WTPSPQ-F-S---G--QQ-LSWL-QTF-NFIWLL--GA-I-QEQIYRLSFAI-TL-QRLGEQLRVL....-GY---GWK-F-DTI-WAGGELGQT  811
DEB.CM.99.CM40              -F-W-R-VVY-IAA-IL---AMFGVN-GSKLCR-----QTL-PK.........EEEVL-LAEEGGRS-GDG-A.-WTA-PR-FFS---E--QQ-LTWL-QTF-NCIWLIR-GF---QEGIYRLSFAL-QR-RLLGERFRVL....GGY---GWK-F-DTITWAGGELWQW  835
DEB.CM.99.CM5               -F-W-R-VVY-IAA-IL---AMFGVN-GTKLCR-----QTL-PK.........EVEVLQLG-EGDKS--GVKS.-WTTSPR-F-R-----ISQ-LTWL-QT--NCAWLIR-GF---QETIYRLSFAF-QR-RLLGEHLRVV....-GYC--GWK-F-DTIIWAGQEIWEW  815
DEN.CD.x.CD1_CMO580407      -F-W-R-VVYVIA---I---IMFIINVLS-LCR-----LQIPIR..........-QKAANQPEEG-GESGDGNRLKWWQYPR-F-SI--E--SQ-LTW--LT--N-VFLIRTACSRIWDFV..........HRHSLILLG........YF--GL---QTGLRD-GTSAIQQ  829
DRL.DE.11.D3                -FK---IGFF--IAII----LA-LW-T-G-F----R--PYVFKDAYRPET..TNYHRQPDKEKG--QDRGEWN.IRSEPSKPESSRS-SKE.TVGTWL.KES-GYIWLKNVKAV-EYGWSE..........................-QEAGRGIC.TLLREAAQRLWRG  864
DRL.x.x.FAO                 -SK---IGFF--MAII----LA-LW-T-G-F----R--PYIFKDAYRPET..TNYHRQPDREKG--QDRGEW.NIRSEPSKPGSSKA-SEE.TVGTWL.KES-GYIWLKNVKAV-EYGWSE..........................-QEAGRGIC.TLLQEVAQRIWRE  860
GRV.ET.x.GRI_677_gri_1      --G-V-IGFLVIVIIL---FAWVLWGCIRNI----N--P.........QIHIHS..SAE--DNGGGQDRGGESSSSKLIRLQEESSTPSRINNWWLN-KSCS--IRTWCYNICLT--IFIRTAVG........................Y---GL-Q-QEA-TG-AQ-L-R-  819
GSN.CM.99.CN166             -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-WEQLQKAIQKAIQQARQLREVAARV....IAYIS-GI---QAA-TGI-DSL--F  846
GSN.CM.99.CN71              ------I---V-AA--A---IMF--NMLR---------P-ILT-QQV.......ER-P-D-TRKD--AQGNV.--V-YLT-FSS-L-------VIWI-QI-AT-VWTLQ--GKF-WEQ....LQKAFQQARQLGEVAARA....AAYIS-GI---QAAVSGI-DSL-VF  851
LST.CD.88.SIVlhoest447      -FQ---YAVL-II-IIV--V-SFIVQNIVKMCR--RV-APSAYVEQDCKWEKR.ENQEQ-DKEG-IEKDKENIYINLEQYKKESSTPPWNVDWSEPLQDSLFVT--KWIKAFGI--LSLVWQFLSWLGHLVILFFQ.HGQHLRKACSRRVVENAQKIA-WSWNKIRRNRR  878
LST.CD.88.SIVlhoest485      -FQ---YAVL-II-IIV---SFIVQNIVKMCR--RV-APSAYVEQDYKWEKG.EN-EQ-DKEG-IEKDKEDIYINLEQCKKESSIPPWNVDWSEPLQDSLFVT--KWIKAFGI--LSLVWQFLSWLGHLVILFFQ.HGQHLWKASSGWMVENAQKIARWSWNKIRRNRR  877
LST.CD.88.SIVlhoest524      -FQ---YAVL-II-III---LSFIIQS-VKMCR--RV-APSAYVEQDFKWENG.ESQEQ-DKEG-TEKD-ENIYINLEQCKKESFRPPWNVDWNEPLQDSLFVT--KWLKAWGI--LTLVWQFLTWLGHIVILSYH.HGQHLWQTLSRFLVE-AQKTA-WIRNKSRRNRG  880
LST.KE.x.lho7               -FT-V-YAVL-ILVII---VLSFIIQN-VKMCR--RV--PSVYIEQDYKWEKE.ENQEQ-DRE--K-ADTETIYINLEQCKKESSRPLWNVDWNEPLQDSL-VT--KWLKEGGI--LSLVWQSLSWL-HL-ILFFQ.NGQRLWQTSSRWMVENAQKIQ-WLREKCRRNRG  875
MAC.US.x.251_1A11           -IK--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQ*THIQQ..DQAL-TREGK--DGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRVYQI-QPILQRLSATLRRIREVLRTE.........LTY---GWNYFHEAVQVDWRSATET  842
MND-1.GA.x.MNDGB1           -VQ-L-WGVFL-L-II----LL-LWNTIS-F----R-VFS........QDCQQN.LYRK--DNG---SNSLELG.EHNSENL...KEESLNRSLIEDLTSFARE*PIWLWLKNLRAAIEYGF..............LAVQEA...IRSLGS-LVSFAAISWR-GYTGWLQ-NR  846
MND-2.CM.98.CM16            --K-V-IGLL--IVIIV---LACLW-VLGKF----R--...PYVFKGDYLRPH..NLKQ-GKER--EPDSEKQ.STKSDSSKVEFGKPWSKE...QIRDWLKTSRGYVWLKNLQAVIEYGWQE..........................-KTAGGKIF-VLQGYAQGLWSRG  832
MND-2.GA.x.M14              --K-V-IG-L-IMVVI----LACLW-TIG-F----R--.PYVLKGDYLRPR....NLRQ-DEER-GELNLEEQ.NIKSESSKKEFGRPWRPE...QIRNWLKESTIY-WLKNLQAVIEYGWQE..........................-KAAGARICQVLQSFAQRLWSGG  833
MND-2.x.x.5440              --K-V-IGLLV-VVII---MLACLW-VLGKF----R--.PYVFKGDYLRPH....NLK--DREGG-EPDLEKQ.NIKSESSRQESRKPWKPE...QVRSWLK-ST-Y-WLKNLQAVIEYGWQE..........................-KAAGAALYQVLQGFAQRLWSRG  856
MNE.US.x.MNE027             -IR--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIRQ..DQAL-TKEGT--DGG-SG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPIFQRFSTTLQRVREVLRTE.........LTY---GWSYFQEAVQVAWRSATET  847
MON.CM.99.L1_99CML1         ------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.......RRLLAYC--GI---QAAVT---D-L-RF  834
MON.NG.x.NG1                ------I--IV-AA-----VLMFI-G-IR--G----L--P-IPTQV.........PREPGAPE-TGE-GGEQG.-NRST-YARGF-TIIWTDCRD-IVWTFQLLRDSGL-IYRS-QRVRSHLIPLLRDLCRQLREAS......SRLLAY---GL--XQXACTGAID-L-RF  831
MUS-1.CM.01.CM1239          ------I-C-V-A---LF--FMF--G-LR-I--------P-I-TQP........EAAEP---T-GGV-SSG-T.--V-YL--F-S-F-----N-VIWF-QT-VSSVSTLRW-LQRAAHLIQKWSV-L-IVIREQARRCSELLARLLAYI--GI---QAAGTAA-DQL--F  856
MUS-1.CM.01.SIVmus_01CM1085 ------I-C-V-A---LF-L-MF--G-LR----------P-ILIRP.......QT--EA---T-GD--SGS--.--V-YLS-FFS-F-----N-VIW--QT-ISSASTLWWLLQRAVHLIQKGSI-L-FT-RALIEPLREALQRALAC--HGI---QTASTAV-DQL-TF  862
MUS-2.CM.01.CM1246          ----V-I-C-V-A--IA----MFI-GMLR-I--------P-I-TR...............QSAEQ-QQEKTEKG.AGRSGRHR--S-FSE--HN-MIWT-QI-TSSASNLWWLTQRIYHHLTKGLVNP-TL--SFREALGRFIAH....A--GI-K-QAAVSGA-D-L--F  847
MUS-2.CM.01.CM2500          --F-V-I-CLV-A--IA---IMFL-G-LRK--------PP-I-TRA.........-E--RT--ET--G--A-GSV.--V-YL--F-S-------N-IVW--QI--SS-SNLWWLIQ-IVHHLQKWSISL-LRARATSTHLLEVLRRLLAY---GL--FQEGVTAAIDGL--W  844
MUS-3.GA.09.09GabOI81       ------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-REQAARLREVIRRIIAY---GL---QEA-TGF-D-L--F  857
MUS-3.GA.11.11GabPts02      ----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-TTGNLLSSIWEACQRAYAY---GF---QAAVSGI--QL--F  858
OLC.CI.97.97CI12            -MT-V-IGLLV-LAY---N--LGLVRKIRIM----VS-QLP.......FSDTL..QDKEQ-DNG-G-DS-FFES..TFNEW-CTTECTGTLRVNWRDLLS-LWKGVVRVTR..................................WI.QKSLQHVWRTLKQKRQ...  810
RCM.CM.00.SIVagi_00CM312    -ME---IG-FVIL-II------LLW-S-SKL-A--T--LSPPPYYHSQIPIRNLKKEEQQAKG--GEFSGEGGGYKSYNWQKGSFY-LILRPIVT--QTLWNNLIS-LLLSYQS-QSILSRIQHPLL-IVA-VQ................DGW--F-AF--W-GEMA-EN  819
RCM.CM.02.SIVrcm_02CM8081   -ME-L-IGVF-IL-VI----F-LLWGCISKI-A--T--LSPPPYYHQQTPIRNREEQPIK--G-GDS--EGG..YKSPSWQRGYFYSLILRPIEI-LRALWNSCNS-LSVIFQS-RSILSQIQQTLR-IIA-IQ................DGW----EFSGW-AEMAQQN  826
RCM.GA.x.SIVRCMGAB1         -MQ---IGFF--I-II---VAWLLWNCLSNL----R--.SPPSYVQQIHIHNTGE-QT-GEKR-D---EGG..NKYNNWLREYCW-QLIHPLSRIWTQLSQICRSCSSIIFQS-RWILA...........................KI--GW--F-EFSSWFAEMALQN  815
RCM.NG.x.NG411              --K--RIG-F-IL-II----I-LLW-C-SKL-A--T--LSPPPCYHQQIPIRNREEQPIK--EKGGS--GGGH.KYYNWQR-........YFYILILRPIEILCRR-YQICSNS-SVIYQSLQSIPSRIL-HLRWIGA..........K--DGW--F-GFSSW-AEMARTN  822
SAB.SN.x.SAB1               -FGWM-IA--VIA-IIVA-VLLVIIG-LRKF-K--A---SLPSSH*.QIHIHL..DKGQ-DKE--Q-SGGLNSDSRSYTWQREFL............RHLCH-LITWLRN.LTSWFSTIFSNLH-CLQDIQQRTRQLTA........H-E-GW----AA-A--TV-VVQ-  832
SMM.CI.79.SIVsmCI2          -VK--YIGFYV-V-II----IVYLIQLLGKL-K--R-VFSSPPSYVQQIHIHK..-QE--TNEGT-EDGG-NG.GYRSWPWQIEYIHFLIRQLIR-LTWLYNNCRD-VLRIFQS-Q-IFQLVS...QAARNRIRPR..........IAY---GW-YFQEACQEAAR-ARET  851
SMM.LR.89.SIVsmLIB1         -VK--QYGV-V-L-VI----MIY-IQMLL-L----R-VFSSPPSYHQQIRTQQ..DQAQ-IKEGI-E-GG-NG.GYRSWPWQIEYIHFLIHQLIR-LTWLYNNCRAWLFRICQN-RQLFQRLSEAIQPVREVIRRE.........FTY---GWSWFEDALQAAWR-ARET  843
SMM.SL.92.SIVsmSL92A        -VK--Q-GVY--I-I-L---IIY-IQMLS-LSK--R-VFSSPPSYHQQIHTHK..VQEL-TKEGIDEEGGNSG.GTRSWPWQIEYIHFLIRQLIH-LTWLYNSCRA-LLRIFQS-QPTYQRLLSLISRIRDILRRA.........VAYIR-GW-YFIE-LQEAAI-ARQT  850
SMM.SL.92.SL92B             -VR--QYGVFL-I-I-L----IY-VQMLS-L----R-VFSSPPSYHQQIHIQR..DQEL-AKEGT--EGGNGG.GYRLWPWQIEYIHFLIRQLIRILTWLYNNLTR-ASRAYQN-QQLCQRLSEISQPIRELVRREAG.........YIR-GWNYFIEACQEAWRSAQE-  851
SMM.US.04.G078              -IK--QYGVYL-V-IIF----IY-VQMLAKL-K--R-VFSSPPTYHQQIPIRK..SQEL-TKGG---EGG-SG.GYRSWPWQIEYIHFLIRQLVR-LTWLFNNCRAWLSRIYQS-QPTFQRLSETAQRIREFIRLR.........VTY---GWSWF-EAVQAAWKSARET  855
SMM.US.05.D215              -IR--QYGVI--L-VI----IIY-VQLLGQF-K--R-VFSSPPSYHQQIHIHK..LQGQ-TEGGI--EGG-SG.GYRSWPWQIEY-HFLLRQL-R-LTWLYNSCRAWLSQICQT-QPVFQRISEVLQTVREIIRTE.........TTY---GWSWFQEAIQAFFKFARET  849
SMM.US.06.FTq               -VK-VYFGFYV-A-VIV----IY-IQMLGKL-K--R-VFSSPPTYRQQIYIHK..-QEQ-TKEGI-EDVG-NV.GYRSWPWQIEYIHFLIRQLIR-LTWLYDNCRTWLLNSYQS-QPVLRGISRALQRVREVIRLE.........LTY---GWSWFQEA-QAAFK-ARET  846
SMM.US.86.CFU212            -VK--QYG-Y--V-IIL----IY-VQMLA-L-K--R-VFSSPPNYRQQIHIHK..DQEQ-TREGI-E-GGIDG.GIRSWPWQIEYIHFLIRQFIR-LTWLYSSCRT-LLRIYQN-QPVLQRLSSVLQAIRETIRTE.........LTYIG-GCLWF-EALQAAGR-AGQT  852
SMM.US.x.H9                 -IK--QYGVL--L-VI-----IY-VQMLA-L----R-VFSSPPAXVXQIPIXT..-QEL-TKEG---DGGX-G.GNRSWPXQIEYIHFLIRQLIR-LTWLFSSCRDWLLRXCQI-QPVLQSLSRTLQRAREVIRVEIT.........Y---GWRY-QEA-QXWWKFXRET  851
SMM.US.x.PGM53              -IK--QYGVL--L-V------IY-VQMIA-L----R-VFSSPPVYVQQIPIHK..-QEQ-TKEG---DGG--G.GNRSWPWQIEYIHFLIRQLIR-LTWLFNNCRTYLLRIYQI-QPVFQRL...LQRIREVVRTE.........ISY---GCRWFQEAVQAAWR-RGET  850
SMM.x.x.pE660.CG7G          -IR--QYGVL--L-V------IYIVQMLA-L----R-VFSSPPAYVQQIPIHK..-QEP-TKEG---EGG--G.G-RSWPWQIEYIHFLIRQLIR-LTWLFSSCRDWLLRTYQI-QPVLQSLSTTLQRVREVIRIG.........IAY---GWRYFQEAVQAWWKFARET  851
STM.US.89.STM_37_16         -VR--QYGVYL-I---M--VAIYIMQLLA-L-K--R-VFSSPPSCRQQIPIHK..-QEQ-TKEGT-E-GG--G.GINSWPWQIEYTHFLIRQLVR-LTWLYNNFRACLSRIYQT-HPTFQRISRILQRIREVVRLG.........AAY---GCIWIQEA-QAAWR-AGET  847
SUN.GA.98.L14               -FL---YAVL-I-IIIAA--LSFIIQQIY-MC---R..VLSPSAYVEQDWLQET.C-KPTDKE---ET-KE-I.YINLEQSKKES-PPPWTVDWDEPLRDS-LVT--KWLKAWGIVLAQNIYHLLSFLWH-LTTSFHHGQRLWQTLRGWLGSHLIQAT-RIRNACR-SRE  879
SYK.KE.x.KE51               -FQ-L-IGFFAIAAI-I--VLSFAWG--RNMLG-----LQKPQP--........LYHKA-A-P--GIETDTDN.YKPSVSTF-REFLRQS-AEAQKLWNRVSESCRSLIRGL-IAW-..............................FIS-GVT---EA-I--GREV-HQ  805
SYK.KE.x.SYK173_COMGNM      -FF-L-IGFYVI-A--L--L-SFSVG-IKNLLG--V-ILQNPT...........Q-RKD-GKPAD-EEGSGDR.EGLNVSTF-RESLRQS-EAGQQLWRTVCSSFRSLIRQLTITW-..............................FIS-GFN---IA-A--GREIRDW  797
TAL.CM.00.266               -FRWLRIAVIV-AS-IL----MYIVNVLKLL----L--P-K-PI-S.......N-D-EA-G-DAT---KQGSY.KWN-SQA-F-SI--E--TG-IRYI-QT-AN-AWLIWDLAKAAKAKSQQVAQATYAF-RLHLTAAAATLQY.......GFS-IQAAWRDGIT-L-RF  837
TAL.CM.01.8023              -FRWLRIAVIV-AS-IL----MYIINVLKLL----L--P-K-PI-S.......N-E-GV--EEGT---KP--S.-WRTSPQ-FF-I--E--N--LSY--QT--N-VWLSWDLLKR-KATTQQAAQATYAY-RIKFTAAV.......ATI--GFS-IQAAWGDCVT-L-RF  855
TAN.UG.x.SIVagmTAN1         -F-W--WGFY--I--ILF-MAWLIWGCIA------F---P.......QINIRL..-REQ-DNAGG-DKDSSSS.-DKSPPSVKES-LPNRGGIQAEERAW-QHLTNWCL-ISSW--RLYQILRRSLTTLLQLLRQE.........CQYI--GW-QF-EG-ARSFE-L-S-  844
VER.DE.x.AGM3               --NIL-IGFLD-L-II---LLYT-Y-CIA----------P.......QIHIHP..WKGQ-DNA-GP-EGG-KRKN-SEPWQKESGTAEWKSNWCKRLTNWCSISSIWLYNSCLT--VHLRSAFQ........................YI--GLG---AA-QEAVV-L-RL  842
VER.KE.x.9063               --NIL-IGFLV-V-II---LLYTIY-CIG----------P.......QIHIHP..WKG--GNAG-PEEGGE--ND-SDSWQKGSGTRQKRGSWSKRLTNWLSTFSTWLYNTFLT--IHLRGAWQ........................YI--GLG---TA-QEARL-V-RF  842
VER.KE.x.AGM155             --NIL-IGFLA-I-VI---LLYTLYTCIA----------P.......QIHIHP..WKGQ-DNAG-PEEGGRTGKSKSTH*QKEFGGRDKRTSWCRQLTNWF-TLSIWLHNSCLT--LKLRSAWQ........................Y---GLG---AA-QEAV-HL-SF  835
VER.KE.x.TYO1_patent        --NIL-MGFLVIV-II---LLYT-YGCIV------V---P.......QIHIHQ.V-KG--DNAD-P-EGG-NSRIKLESW*KDSKSRCMQ-TAWLTRLN......TWLYNSCLT--IQLRKAFQ........................Y---GLA---TG-QEI-QTL-GV  828
WRC.CI.97.97CI14            IWNWL-WGLLVI-IV-----LACLMNSIRIA-----K-IS-......QNHLNN.LEEGQQDSE-GIEEDYNSIEINDWQYRRES-IVFLNSLLTS---ICQHLLHIVWSLLSCL-WW......IND--SL--DI...........SGRAYGAIRKKIAKSRKENTQAE.V  841
WRC.CI.98.98CI04            IWNWL-WGLL-I-IV--I--LGWLMTSIRIA-----K-IS-......QSHLNN.LEEGL-DRE--AEEDNSTLDV-DWQFKKES-ITLLNSLINF-SAVCQTLLQIVWSLLACLFWWIQHGW........NLLKRI..............SGAAHACIRKKITKSRKETT  847
WRC.GM.05.Pbt_05GM_X02      IWNWL-WVLLVI-LV--I-VLGCLFTSLRIA-----K-VS-......QSHILN.LEEGL-DNE-T-EEDFIAIEE-DWQFKKESLILFWNSLATL-Y-ICLSLLQ-SYSLLSYLWI-IQNGW........SLLKRL..............CGAAYAALQEKIAARRQKRK  839
```

```
                                                   Env end
                                                 gp41 end
                                       cytoplasmic tail end
                                                        *
H1B.FR.83.HXB2              VAEGTDRVIEVV...QGACRAIRHIPRRIRQGLER..ILL*   856
H1O.CM.91.MVP5180           --NW--GI-LGL...-RIGQGFL---------A--...-V-   873
H1N.CM.95.YBF30             ------I--LA...-RIG-G-L------------.A-I-   853
H1P.CM.06.U14788            --NW--Q--A-G...-QIG-GFLN----L-------..S---   883
CPZ.CD.06.BF1167            TGWW--G---AL...RV-VDI-----T------I..A-N-   885
CPZ.TZ.00.TAN1              TGWW--LI--G-...VYIA-G--N----------L..A-N-   872
GOR.CM.07.SIVgorCP2135con   T-DW--Q--AIA...-RIG-G-LN-----------..S---   850
                                                   Env end
                                                 gp41 end

MAC.US.x.239                L-GAWGDLW-TL...RRGG-W-LA----------L..T---   880
H2A.SN.86.ST_JSP4_27        LTNAWRGFWGTL...GQIG-G-LAV-------A-I..A---   860
H2B.CI.x.EHO                L-SAARTSWG-L...RR-AGE-IA--------A-L..A---   853
H2G.CI.92.Abt96             LTSTWRALWKTL...GRVG-G-LA----------L..T---   885
H2U.CI.07.07IC_TNP3         LGRVGRGILAIP...................F-L..A---   882
H2U.FR.96.12034             LSRTGRELW-TL...GRVG-WL-A--------F-L..A---   868
ASC.UG.10.RT03              TWNW-EA-LAYL...RRLAQG-IA--------A-I..L-N-   887
ASC.UG.10.RT08              TWNW-EA-LAAL...RQLAQG-IA--------A-I..L-N-   891
ASC.UG.10.RT11              TWNW-EA-LVAL...RRLAQEFIA--------A-I..L-N-   882
COL.CM.x.CGU1               IL-TLQ-WLRSA...ARGW-RAPEYL-GWIYDRPQ..GPA-   838
COL.UG.10.BWC01             IP-TLS-WLRSA...TRRW-RASEYL-GWLFDRPE..SPA-   843
COL.UG.10.BWC07             IPQTLQ-WLRSA...TS-G-RALESCG-WIR-DPQ..GPA-   835
DEB.CM.04.SIVdeb04CMPF3061  FQRWAEVALQ-L...GRGI-E-LA--A-----A-L..F-N-   847
DEB.CM.99.CM40              LQGIAEMALQGL...VR-GG--LRV-A-----A-L..--N-   871
DEB.CM.99.CM5               LQGIAQVA-QGL...VWGG-NLLA--A-----A-L..L-N-   851
DEN.CD.x.CD1_CMO580407      GRATAEV-LAAL...TR-A-EVVA----------I..V-N-   865
DRL.DE.11.D3                GRQFGLSAARAL...R-LAQGVAN--------A-I..L-N-   900
DRL.x.x.FAO                 GRQLGLSSARAL...RALAQEVAA--------A-V..LFN-   896
GRV.ET.x.GRI_677_gri_1      AR-AWG-LGAI-...RS-Y--VINS---V-----K..V-G-   855
GSN.CM.99.CN166             TWNW-EA-LHAC...RRVW-EFLA--------A-I..LFN-   882
GSN.CM.99.CN71              TWNW-EPLVQT-...GRVW-EFLA--------A-I..L-N-   887
LST.CD.88.SIVlhoest447      DISAAHARNLRL...GQKK-WRFRFRG-SGFPS-TTETA--   916
LST.CD.88.SIVlhoest485      DISAAHARDIRL...GQKK-WRFRFRG-SGFPS-TTETA--   915
LST.CD.88.SIVlhoest524      ELSK-NGTNIRL...GQNR-RW.RFRF-SGLPS-TTETT--   917
LST.KE.x.lho7               QLSS---KNIQL...GKKK-WRLRFGG-SGISS-ATETA--   913
MAC.US.x.251_1A11           L-GAWGDL*-AL...RRGG-W-LA----------L..T---   877
MND-1.GA.x.MNDGB1           IFT-CREA-IAA.....GTC-W--L-----SA---.P-N-   879
MND-2.CM.98.CM16            HQW-LSTAACFR....AIA-G-IN--------A-V..L-N-   867
MND-2.GA.x.M14              HQL-LSVIRGAA....AIG---GN--------A-V..L-N-   868
MND-2.x.x.5440              YQL-LSG-RGAA....AFG-G-WN--------A-A..L-N-   891
MNE.US.x.MNE027             L-GAWGDLW-TL...GRVG-W-LA-------E--L..T---   883
MON.CM.99.L1_99CML1         TIVW--ALLHAG...GRLW---VA--------A-I..F-N-   870
MON.NG.x.NG1                TVIW--A-LRLG...GRLW-GLVAV-------A-I..L-N-   867
MUS-1.CM.01.CM1239          TWNW-EPILQIG...RRVW-EFLA--------A-I..L-N-   892
MUS-1.CM.01.SIVmus_01CM1085 TWNW-ET-LQAG...RRVW-EFLA----V---A-I..L-N-   898
MUS-2.CM.01.CM1246          TWNW-EPILQFS...-KIW-KFLA--------A-I..L-N-   883
MUS-2.CM.01.CM2500          T-NW--A-LQ-S...KQL*K-FLA--K-----A-I..L-N-   879
MUS-3.GA.09.09GabOI81       TWNW-EA-LQ-L...RRVA-EFLA--------A-I..LFN-   893
MUS-3.GA.11.11GabPts02      TWNW-EPLFQAX...RRVW-EFLA--------A-I..L-N-   894
OLC.CI.97.97CI12            ...HGINLQQKK...-RRGSPLQKAQEQLGREGAE.TTGV-   844
RCM.CM.00.SIVagi_00CM312    AYYTWRGLQSLA...RQLAGWPATCG----------..L-N-   855
RCM.CM.02.SIVrcm_02CM8081   AYFVWRGLCVTA...-DIA-WPGTVC------F--..L-I-   862
RCM.GA.x.SIVRCMGAB1         AYYTWRGLCA-A...RDFAGWPAMVC----------..LCN-   851
RCM.NG.x.NG411              AYYTWRGLCA-C...RDPAGWPATLC------F--..F-N-   858
SAB.SN.x.SAB1               ATSAS-S-RHAC...RSIV--VIAH---M--E---..WFN-   868
SMM.CI.79.SIVsmCI2          L-SAGRDLW-TL...GRIG-GLLAA---------L..AF--   887
SMM.LR.89.SIVsmLIB1         L-SAWRDLWQTL...GRIG-GLWA----------L..T---   879
SMM.SL.92.SIVsmSL92A        LVSTGRTIWQTL...GHIG-G-AA----------L..T-N-   886
SMM.SL.92.SL92B             IVGAWGLIW-TL...GRVG-G-AA----------L..M-N-   887
SMM.US.04.G078              L-SAWRDIW-TL...GRVG-GLLA--------F-L..T---   891
SMM.US.05.D215              L-SAWRDLW-TL...GRVG-G-AA--------F-L..TM--   885
SMM.US.06.FTq               L-SAWRDLW-IL...RRVGGWMLA----------L..T---   882
SMM.US.86.CFU212            L-SAWRDLW-TL...GRVG-G-LN----------L..T---   888
SMM.US.x.H9                 L-SAWRDLW-TL...GRVG-G-LA--X-------L..T---   887
SMM.US.x.PGM53              L-SAWRDSW-TL...GRVG-WVLA----------L..A---   886
SMM.x.x.pE660.CG7G          L-SAWRDIW-TL...GRVG-G-LA--------F-L..A---   887
STM.US.89.STM_37_16         L-SAGRDLW-TL...GRVG-R-GA----------L..T---   883
SUN.GA.98.L14               RVSSSQKARSRTFSLGRKW-PKWNRT-GS-IPS-TTETT--   920
SYK.KE.x.KE51               MVAIWQALLAYA...RRVAENVAAL---L-----I..Y-N-   841
SYK.KE.x.SYK173_COMGNM      --AIWQAIYAAT...RRVVE-VAAL---L-----I..Y-N-   833
TAL.CM.00.266               TWWW-NG-LAAA...RW-GE--AA----T-----Y..CFT-   873
TAL.CM.01.8023              TWRW-SG-LAAA...RW-VQ--AA----T-----Y..C-A-   891
TAN.UG.x.SIVagmTAN1         AQSASRTLWNAC...RS-Y---LEH---M--E---..WFN-   880
VER.DE.x.AGM3               AQNAGYQIWLAC...RS-Y---INS---V-----G..--N-   878
VER.KE.x.9063               AQNAGHQIWLAC...RSTY-H-ISS---V-----E..--N-   878
VER.KE.x.AGM155             ARNAAHQIWLAC...RS-Y---INS---V-----E..V-N-   871
VER.KE.x.TY01_patent        AQNACHQIWLAC...RS-Y-N-VNS---V-----E..--N-   864
WRC.CI.97.97CI14            S-T-SRKTAASS...RIRML-T-LCASFDKCRRSFSERRQ-   879
WRC.CI.98.98CI04            EK-ISENSNRET...T-SS-VWKL-T-LCSSLNKSRRR-F-   885
WRC.GM.05.Pbt_05GM_X02      EK-IYTPGSRET...SRSSSIRGITA-LCTNLNPCRRGQQ-   877
```

```
                                 myristoylation                                                                                                           phosphorylation
                                  Nef start                                                                                              acidic cluster  poly-P helix
H1B.FR.83.HXB2                   MGGKWSKSSV...IGWPTVRERMRR...A..................EPAADRVG.AASRDLEKHGAITSSNT......AATNAACAWLEA...................QEEE...EVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDLWIY  115
H1O.CM.91.MVP5180                --NA----KF...A--SE--D-----........SSS.........DPQQ-C-PG--.-V--E-ATR-G-S--H-......PQN---L-F-DS...................HKD.-...D-----R---------F---F---F--------D---Y-HK-AE-------  120
H1N.CM.95.YBF30                  --KI-----L...V---EI------...RQTQEP.............AVEP-VGA-.---Q--ANR----IR--......RDN-ESI-----...................QE---..------R-------I---Q-F---F---D-------VW-RK--------M-  121
H1P.CM.06.U14788                 --NA-K---L...V---A---KIKQ...TTPTTPDP.........TTPVTP-PG--.EI-KE-AQGKG-P-KFS......SKN---L-F-D-...................H---..------R------C------F---F--------D--VY-PE-AE-------  125
CPZ.CD.06.BF1167                 --AV........L-TRFPGIDNML...................NRIRNTPAADA.PG-Q--HNR-GL-TNTI......GTEHDVLQHSGD...................HT--..A-----R----M----E-L-----W----------YW-PK-AA---T-MF  111
CPZ.TZ.00.TAN1                   --NIFGRWPG...ARKAIE..DLH......................NTSSEP--.Q--Q--QNK-GL-TNTL......GTSADVLEYSAD...................HT--..------R-A--M----E-L-I---W--------D--FF-PK-AA---T-M-  111
GOR.CM.07.SIVgorCP2135con        --NT-K---F...----A------Q...T................SPEPKP-PG--.EV--V-AQR-G-PAKY-......PRN-Q-L-F---...................H-D-..------X---------F-G-F---F----------VF--A-SE-----V-  118
                                  Nef start     R17Y mutation                                                          premature stop in Mac239

MAC.US.x.239                     ---AI-MRRS...RPSGDL-Q-LL-...-RGETYG...........RLLGEVEDGY.SQ-PGGLDK-LSSL-CE..GQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDE*DDD...L--VS-R-K----T-S--L-I-M---I--------IYY-A--HR---IYLE  146
H2A.SN.86.ST_JSP4_27             --ASG--KRS...EPSRGL---LLQ...TPGEASG...........GHWDKLGGEY.LQ-QEGSGR-QKSP-CE..GRRYQQGDFMNTPWR-PA.EGEKGSYKQQNMDDVDSDDDD...L--V----R----E---RL-R-M--LI---------YY-D--RRV--IYLE  146
H2B.CI.x.EHO                     --SAG--KQS...KQQ-GL---LL-...-RRGPRG...........ESSGERQERS.LQYPGGSDK-LNSP-CD.............GQKTLGAEGGGKQ.......DSDEDD-DN..---VR-R-G-------F-L---M--------E---IFY-E--HK---TYLE  129
H2G.CI.92.Abt96                  --SAG--KRA...PQQLGL-K-LLQ...-RGEPYG...........KLWEGLEEGY.SQY-EESGK-LSSL-CE..GQQYTQGQFMNTPWRNPATERAKLAYRXQNNDDVDSDDND...L--V--Y-K----V-S--L-I-M---I--------IYY-E--HR---IYME  147
H2U.CI.07.07IC_TNP3              --ASG--KRS...KQAGRL---LLQ...-RGETYG...........RLWDGLEEGY.SQ--GGSGRDLNLP-YE..GQAYSEGQFMNTPWRNPAGEGEKLKYRQQNMDDVDSDDDD...L-KVS-R-R---------WP--M------------IYYN---HR---MYLE  147
H2U.FR.96.12034                  ---SI--KHS...KRSGKL---LLA...-RGENYG...........RLWGGLEDGF.ELYQGGSGKDLSSRFCEAQGQGYSEGEFMNTPWRTPATGKDKLQYKQQNMDDVDEEDND...L--VA-R-R-A--T-----L---M---I------D-IFY----HR---IYLE  149
ASC.UG.10.RT03                   --S-N--QQS...SQPGMPSPYSPG...TGQKLYW...........RIFGDLP-GL.LP--GESGREPRYY-IR..............EET-ETAKRGEVVPVRYNEGDSTRDDD-E..G-----C-------Q--F-T---F-W----------FY-PE-HKK---YA-  136
ASC.UG.10.RT08                   --S-N--QQS...SQPG-PSPYSPG...TGQKLCW...........RLCGSLP-EL.LP-QGGSGREPRY--IK..................EEAAQPGEVVPLRYNEGDYTRTDD-E..------R----I-T--F-T---F-W----------FY-PE-HQK---YA-  133
ASC.UG.10.RT11                   --S-N--QQS...SQAG-PLPYSPG...IGQRLYW...........WLCGDLP-SL.LP--GESGR-PRY--IE...............E--AAAKRGEVVPMRYNEGDNTRDDD-E..------R----I-NL-F-I---F-W----------YY-PE-HKK---YA-  135
COL.CM.x.CGU1                    --SLL-CLNI...RQQ-QT-GGVLG...RLFGIYWYPGVQ.....AFRFLRHF-G-YDLR-VAGVERLNICEDG.......YMTVPKVLPNCPCPLCYRHPDVYYHNPEPEDV-DV...-A-CY-MSRL-V-QA-I-LL--I-C----------I-RTPE-D-LIEQYA-  149
COL.UG.10.BWC01                  --SIL-CLR....GQR-STPGGLCS...RLFGTYWYRGVQ.....TYRFLRHF-G-YDLR-VAGVEHLNILEDG.......YLTVPR-LPNCPCPLCYRHPSVYYHSPEPEDAD--...-G-CF-QAR--V-QG-I-LL--I-C----------I-RTPE-D-LIE-YA-  148
COL.UG.10.BWC07                  --SIL-RG-NRYQMER--SPGGCLS...RLFGVYWYPGVQ.....TWRFLRHF-G-YDLR-VQGVEHLNLVADG.......YEEIPRDLPNCPCPLCYRHPRVYFHSPEPEDLG-....NNSCF-CAR--V-RP-I-LLF-I-F----------I-RTPE-DRL-ERYC-  151
DEB.CM.04.SIVdeb04CMPF3061       ---RN--TQS...S-QEENWGKHFK...GGQRWRY...........RYLGEESEKF.LPYQPESDKELNC-LTE.......E-FIGNPRR-IK..........CTQRQQQDFDQQ-...------K-R--I-DP---LMI-Y----------DIFY-A--HA--E-HAQ  132
DEB.CM.99.CM40                   ----N--IPS...R-QVGSFGSGS-...GLLRWRY...........RDLSEQEEQF.SECLLESDREQSS--TE..........FWGSPRR-IK..........CTNKQQQDLDQT-...A-----R---SI-DP---LMI-Y----------DVFY-A--HA--E-HAQ  130
DEB.CM.99.CM5                    ----N--TPS...F-QGKKSGSGSK...GLLRWRY...........KDLSGEAETF.SPFLQESDKEQSCY-TE.......EPLY-DPHR-IK..........CTQRAQECYD--H...GI--A-R-R--I-EP---LMI-Y--Y-------DIFY-A--HA--EIHAH  132
DEN.CD.x.CD1_CM0580407           --S-N--QAS...GI-GLAPSNKEG...RQRKWYW...........RLL-EQQEKW.WP-P-GSDRALKL--TRLLEHDEDEETFVD-EEIIQSSPALDPQHPRIHQYVVDHDD--...------R-RR--HAP---DMI-M----------QDIFY-P--HA----YAQ  149
DRL.DE.11.D3                     ---QN--KRE...EAYARFYG-L----...GYGAEGGNLD....YQQLEPSE-LL-EW.RTFPEGLDKAQKY--TSSLDTTPIQ.A--A-EKDVTLPTEEEQPS.........EE---..------Y--C---EP---EL--F----------K-IWW-L--ES----YAQ  146
DRL.x.x.FAO                      ---QN--KRE...EAYARFYK-L----...GYGAKGGNLD....YHQLEPSEPLL-KW.QPFPGGLDKAQRS--TSSLDTTHMQ.--A-EKDVTLPTEEEQPS.........EE---..------C--R---EP---DL--F----------K-IWW-L--ET----YAQ  145
GRV.ET.x.GRI_677_gri_1           --SSN--RQQ...Q-LLKLWRGL-G...KPGADWV...........LLSDPLIGQS.STVQEECGKALKK-WGK.................GKMTPDGRRLQEGDTFDEWDDD--..------Q-R----Q----L---F-----S---D-IYY-E--EK--N-YAL  131
GSN.CM.99.CN166                  --S-N--QQQ...QESS-ALLSS-G...TGQRPYF...........TLVDEYGENF.WL-P-ASDK-RRYYLTE......EPKPKRGSL........................--Y...-PSC--R-R----DP---VM--------------MFYCED-HQK-EQYA-  119
GSN.CM.99.CN71                   --S-N--QQS...QESS-ALLSSHG...TGQSPSF...........RLLDEYGENS.WL-P-ASDR-RRYYLTE......ESPSRQNCI........................DY...-PSC--R-R------DP---LM------------AMFYCED-HQK-ESYC-  118
LST.CD.88.SIVlhoest447           --NIFG-PAA...D--WKTLR-L-.......................AGAGTRSEGT.EETYQQLMQETSDWDKR...................................RDGDSDS--..------R--R--CSP---TLI-----I---------YW--------C-YCE  108
LST.CD.88.SIVlhoest485           --NIFG-PAA...D--WKTLR-L-.......................AGAGTRSEGT.EETYQQLMRETSDWDKR...................................KDGDSDS-G..------R--R--CSP---TLI-----I---------YW--------C-YCE  108
LST.CD.88.SIVlhoest524           --NIFG-P-A...DS-WR-LRKL-.......................AGSGTRAEGT.EE-YQKLMEQTLGWDK-..................................EEDGDSDS..------R--R--CSS---TLI-----I---------YW----N---T-YCE  107
LST.KE.x.lho7                    --NAFGRPAA...D--WKTLR-F-.......................AGSGRNAEGT.ED-YRALIEKTSNWAKR...................................RDGDSDS--..------K--R--CSP---TLI-----I---------YW--------I-YCE  108
MAC.US.x.251_1A11                ---TI-MRRS...RSTGDL-Q-LL-...-RGETYE...........RLLGEVEDGS.SQ-LGELDK-LSSL-CE..GQKYNQEQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDD...L--VS-R-K----T-S--L-I-M---I--------IYY-A--HR---IYLE  147
MND-1.GA.x.MNDGB1                --SSQ--KRS...EA-VRYSSAL-QLVGGPVTPDGYKQIESSQGAEKQSLLRG-AY.GTYSEGLDKVQNDPLTK...........................DEKLDLTQQDPE---..------CR--S--VPS--DLI-F---I-------G-IYY-R--EE----YAE  136
MND-2.CM.98.CM16                 ----S--QQE...EKYLKYYKA----...GYGAEGTNGD....YQQLHASEPLLGAL.ST-QEEFDREQKS--TD...................................--...-T----Y--C-V-EP---DL--MP-----------IW--K--EE----YAQ  117
MND-2.GA.x.M14                   ----S--QRE...Q-YAKYYKAL----...GYGAEGTNLD....YQLLEEQRLLEGPL.GT-LGASDKELKSC-TK..........................VDDEQK-...-R----Y-TQ-V-EA---NLI-M-----------IWF-R--EE--N-YAQ  121
MND-2.x.x.5440                   ----S--QRE...QHYIKYYKAL----...GYGAEGTNLD....YQVLEEQQPLAGAS.GT-LAASDKVLRPY-TE.........RDLRHQQDVTLPSEKEQP.........SD---..------Y-TQ-VPEA---DLI-M-----------IWW-K--EE----YAQ  137
MNE.US.x.MNE027                  ---AT--KRS...KSPGDL-Q-LL-...-RGETYG...........RLWEGLEDGY.SQ-LGGSDKSLSLL-CE..GQKYNQGQFMNTPWKNPAGEREKLAYRKQNMDDIDEEDDG...L--V--R-R----A----L----M---I--------IYY-E--HK---IYLE  147
MON.CM.99.L1_99CML1              --S-N--QQS...PLSS-PLLGSQS...SGRMRYF...........ML-DDYGEQS.WL-P-ASDRERKY-LTE......GRNGKQRRQ.................PLDDDDDDD...G--C--R-R----DP-W-LMM----Y-------GEMFYCED-HRKIEQYA-  125
MON.NG.x.NG1                     --S-S--QPA...QEPLMPSHGSQS...SGQTRSF...........AW-DDYGEDS.WL-P-ASDR-RRF-LTE......GRNHRSNRRPTTV.............VDDADEDQD...LP-C--R-M----DP-W-IMM-----------DK-FYCVD-HQK-EQYA-  129
MUS-1.CM.01.CM1239               --S-N--QQE...QQPL-SSPSSPG...TGQSPYF...........RLVDEYGENS.WL-P-ASDR-RKY-LTE......GSNERPLIE..........................H...-RSC--R-R----DP---MMI----Y---------MFYCEE-HQK-ETYT-  118
MUS-1.CM.01.SIVmus_01CM1085      --S-S--QPA...QQSLISSPPSPG...TGRKQYF...........KLVEEYGENS.WL-P-ASGR-RRY-LTE......G-SKRPVIE..........................H...-RTC--R-R---I-DP---LM--M------------MFV-TE-LQK-ET-A-  118
MUS-2.CM.01.CM1246               --S-N--PQS...QERS-PLRSSHG...TGPNRYF...........SLVKKYGENS.WL-P-ASAR-RKY-LTE......NKGRRQIIE..........................H...-CSC--R-K--I-DVN--LMI----Y--K-R----MFVCID-LQK-ETYA-  118
MUS-2.CM.01.CM2500               --S-N--KES...QQPLMGSPSGPQ...TGQTPYF...........KLVNNYKKHS.WL-P-PNASAR-RKY-LTK......GSNKK-IIK........................KH...KRSC--R--H-IQNP---LIIN-------------MFY-EE-HQK-ETYA-  118
MUS-3.GA.09.09GabOI81            --S-N--KQQ...QDSSMPLLSSPG...TGQRPCF...........KFCDEWPENS.WL-P-ASGRERKF-LTE......GKDRRPVID.........................---...DRSCH---KCG--DP-F-LMI------------DRMFYCTD-HQK-ETYA-  118
MUS-3.GA.11.11GabPts02           --S-S--QQS...QESS-SLLSSHG...TGQNRYF...........RLCDEYGENS.WL-P-ASARERRY-LTE......GKDRRPVIE.........................C...-RSC--R-RCGIQAV-F-RLI-------------MFYCEE-HQK-ETYAF  118
OLC.CI.97.97CI12                 --SIC-RRNR...E-EVLY-.........................RLRSNWGEKE.QKQLES-DQHQTNLAIE.............................SLMTGDTDTDD...LE----K----V-QA---TLIE------------MYW-P--ENLIY-YAL  106
RCM.CM.00.SIVagi_00CM312         ----S--H-Q...S-LVRW-QK-LT...TPGEGYS...........RWHDNLQDGQ.PRVAEGSGKASRDF-IR......................GGFTIETQQSVDAMDWY--TDDTL-----K------A-SF-L-I-M-----------IYY-I--HR---IYLE  130
RCM.CM.02.SIVrcm_02CM8081        ----S------...A-LQKW-N--LT...SSGEAYA...........LLHKTLQDGP.-RCAEGSGRASRDFLSR......................GGITIETQQSVDAIDWY--TDDTL-----K---------F-L-I------------IYY-T--HR---IYLE  130
RCM.GA.x.SIVRCMGAB1              ----S--N-A...A-LLRW-FK-LT...TPGEGYV...........RWHETLLDGQ.PWCAEGSGRASRDFVIR......................GGITAETQASIDDIDWY-DTDDTL-----K--------S--L-I-M-----------IYW-I---R--MYLE  130
RCM.NG.x.NG411                   ----S--A-A...A-LRRW-GL-LT...TPGEDYA...........RFAETLQDGQ.PRCAEGSGRASRDFLTR......................GGFTIETQQSVDAIDWY-DTDDTL-----R----I---S--L-I------------IYF-E--RR----YME  130
SAB.SN.x.SAB1                    ----S--QQQ...RHSLWLWSKL-Q...-PVIQYD...........MLADPLLGQS.SHIQEECAKSLRDGLIR...............QGDSSRTEEGVKMKHQGRQPSWYD-D-E..------R-CL---A----L-I-FG-----------IYY-E--KK----YAL  135
SMM.CI.79.SIVsmCI2               ---NI--KRA...RKPHGP---LLQ...-RGETYG...........RLWEGLEEGY.LQP-GELGRDWSSL-CE..GQTYSEGQFMNTPWKNPVKERGKLKYGRQCMNNVGEDDKD...L--VA-R-KI---E----L-I-M---I---#----IYY-E--HK---MYLE  146
SMM.LR.89.SIVsmLIB1              ---AG--MRS...KRHGGL---LL-...-RGETYG...........RLWEGLEEGY.GQ-LGASGK-LNLPCCE..PQRYNQGDFVNTPWRNPAREKEK..YRQQNMDDVDEEDD-...LI-VA-MNRI---A-S--L-I-M---I--------IYY-H--HR---IYLE  145
SMM.SL.92.SIVsmSL92A             ---NT-SKAC...KKPQSL-DKLLS...-RGEQYG...........RLWDTLEEGS.QR-QGGSGRDWNLH-IE...QKYSPGEYMNTPWRNPAREREK.......#ENMDDID--..D--IA-R-KI---A----L-I-M---I--------IYY-E--HK---SYL-  138
SMM.SL.92.SL92B                  ---IT-SKPA...RKHGDL-KKLLS...-RGDSYG...........RLWDGLEEGS.QQ-QGGSDRALNSCLTR.DGYKYNVGEFMNTPWRNPATERQKQQYRQQCMDDIDEDDD-...LA-VV-W-K----A-S--L-I-M---I------D-IYY-E--HR---IYLE  148
SMM.US.04.G078                   ---AG--KQS...KRRGNL---LLQ...-RGETYG...........RLWEGLEEGY.SQ--GASGR-LSSL-CE..AQKYSEGQYMNTPWRNPATERAKLGYRQQNMDDVDDEDDD...L--VS-Y-HA---A-S--LVI-M---I--------IYY----HR---IYLE  147
SMM.US.05.D215                   ---VG--KRS...KRSSNLQ--LLQ...-RGETYG...........RLWEGLEEGS.QQ--GGSGK-LNSLCCE...QVYSEGDYMNTPWRTPAREREKRMYKQQQMDDVDEEDND...LI-VA-R-K----A----L-I-M---I--------IYYCP--HR---IYLE  146
SMM.US.06.FTq                    ---AG--KQH...KRHLKL---LLQ...-RGETYG...........RFYDGLEDGC.LQ--GASGRALSSLCCE..GESYSQGQFMNTPWRNPAREREKLAYRQQNIDDIDDDNDD...LI-V--HSK----C----L---M---I--------IYYN---HR---IYLE  147
SMM.US.86.CFU212                 --AFG--KRC...RQQEGL-D-LL-...-RGETYG...........RLWEGLEEGY.LT--GESGK-LSSL-CE..PQMYNEGQYMNTPWRNPARERAKLAYRKQNMDDVDEEDD-...L--VA-H-K----A-S--L-I-M---I------D-IYY-E--HR---MYLE  147
SMM.US.x.H9                      ---VT--KQR...XAGGNLX--LLQ...-RGETYG...........RLWEGLEGEY.SQ-X-ASGK-LSSL-CE..PQKYCEGQFMNTPWRNPATEGAKLGYRQQNMDDVDNED.....L--C--S-R--V-I----L-I-M---I--------VYY-I--HK---YLE  145
SMM.US.x.PGM53                   --AAG--KQS...RQRGGLG-KLLQ...-RGETHG...........KLWEGLEDGY.SQ--GELGRDWNLH-FE..GQGYSEGQFMNTPWRNPAREREKLKYRQQNMDDVDDDDD-...LI-VS-H-K----A-S--L-I-M---I--------IYY-E--HR---IYLE  147
SMM.x.x.pE660.CG7G               ---VI--KQC...RRGGNL---LLQ...-RGETYG...........RLWEGLEEGY.SQ--GASGK-LSSL-CE..PQKYSEGQFMNTPWRNPATEREKLGYRQQNMDDVDDEDDD...L--VS-H-R----A----L-I-M---I--------IYYNEK-HR--MYME  147
STM.US.89.STM_37_16              --ASG--KQR...KQHGEL---LL-...-RGETYG...........KLLEGLGEGS.GP-QGASDK-LNSH-CE..PQRYNEGQFMNTPWKNPAAESAKLEYRQQNMDDVDEEDDN...L--VA-H-R----E----L-I-----I-S------IYY-E--HR---MYLE  147
SUN.GA.98.L14                    --NAFGRP-E...V--VRTLF-L-.......................AGSGTRAEPA.GREYHRLRRQEVEPLVS.............................AENGGPNGI-Q-...--E---R--R--CKP---QLI-----I---------WY-RT-EE----YAE  111
SYK.KE.x.KE51                    --SQS--KQP...SR-DEKWRTRWW...PFGKPYS...........PMPDELL-MS.QPYHEDFDR-WRSTLTE...............PILDPKRDFDSGGKKWNAGDICHD-GDED.L-----L-------L---L-I-----I-R----Q-MFFC-K-HE--Q-YLK  136
SYK.KE.x.SYK173_COMGNM           --STS---QQ...LRSEGKYAIGW-...LFGKQYT...........PLPDELS-PL.QPC-GGFDKAWRSTLTE............PIDPHGPDRDWGHSGGQKFSPGDIVQD-GDTG.L-----C--T---TL---L-I-----I-N----Q-MNYCEK-DE--H-YLQ  139
TAL.CM.00.266                    --SVK--LLG...ETASQRLQGSHG...GGRMVYW...........RLRDGQVKRS.QL-PGARGKDLNIALLD......EEWQ-TG-DYMGKALTQGEVW......YEKEQE-NDME.L-----R-K----TC---LGI-F-F--------K-IFYNK--HA--N-YAH  139
TAL.CM.01.8023                   --SVK--LLG...ETASQRLQGSPG...-GRLVYW...........RLRDGQC-QL.QR--GERGK-LNIALLD......-GWQ-SD-NWMGKALSEGEVW......YEKEQE-NDME.L-----R--------CN--LGI-F-F--------K-IFYNN--HA--N-YAH  139
TAN.UG.x.SIVagmTAN1              ---SN--REQ...Q-LLRLWRAL-K...-PVVRYG...........MLADPLIGQS.SNIQEECDKNWNGG-TR..................RGKSTPEGRKLAADDTWDDWEP---..------R-R----Q----L---F------------IYW-PK-EQ--N-YAL  132
VER.DE.x.AGM3                    --LGN--PQH...KKQLSLWHALHK...TRATRYG...........LLADPLIGQS.STLQEECDKALKE-LIR...............KRNGKMTPEGRKLQEGDKWDEWSD--D...------R-R----Q----L---F----------D-IYY-D--NQ--N-YAL  134
VER.KE.x.9063                    --LGN--PQH...KKHVSLWHALHK...TQVTKYG...........LLADPLIGTS.S-VQEEYDK-LRK-LIR...............KRNGNMTPDGRRLQEGDEWDEWSDD-D...------K-R----Q----L---F----------D-IYY-E--NR--N-YAL  134
VER.KE.x.AGM155                  --LGS--PQH...KKQL-IWRALHA...TRHTRYG...........LLADPLIGQS.STLQEECDK-LRK-LIR...............KRNGNMTPEGRRLQDGDQWDEWSD--D...------R-R----QI---L---F----------D-IYY-D--NK--N-YAL  134
VER.KE.x.TYO1_patent             --SQN--PAH...KKYSKLWQALHK...THVTRYG...........LLADPLIGTS.STVQEECDKALRK-LIR...............KQNGNMTEEGRRLQEGDTWEEWSDD--..------R-R----Q----L---F----------D-IYY-D--NK--N-YAL  134
WRC.CI.97.97CI14                 ---VF--T--...EEHMERYA-RLQR...-GRRIRRQRY........QQLA-ERQQQ.-VESGCLLPDSAQALIN.............................AEEASVKEDSSD-D-...G----------I-QP-F-LL-----LF-------DIFW-A--EE-VN-YAH  126
WRC.CI.98.98CI04                 ---IF--E--...EQRMHAYA-RLQR...-GRKQRRKRY........QKIATERQQV.-VESGNLLPDSAQALIR.............................AEEDSSKEDSSDD-..------H----I-QA-F-LL-----LF-------YIFW-P--EE-VN-YAF  125
WRC.GM.05.Pbt_05GM_X02           ---VF--D--...EQRMLRYR-RLLR..VGRKERRRRY........TRLA-ERQAE.--ASGELLPDSAQTLIR.............................AEEDSSKEDSSDDD...AI----R------QA-F-IL-----LF-------DIFW-P--EE-VN-YA-  125
```

```
                              HXB2 premature Nef end                                                                                                      normal Nef end
H1B.FR.83.HXB2              HTQGYFPD*QNYTPGPG.......VRYPLTFGWCYKLVPVEPDKIEE..ANKGENTSLLHPVSL....HGMDD.P...EREVLEWRFDSRLAFHHVARELHPEYFKNC.......................................*   205
H1O.CM.91.MVP5180           ----F---W-C-----------.....P-F------LF-----SAEEA-R.LG-TN-DA-----ACN....--AE-.A...HG-I-K-Q--RS-GLT-I-LQK---L-PK..........................................-   212
H1N.CM.95.YBF30             ----IL--WH------------.....I---V-----F----LSAEEV--.--E-D-NA----ICQ....--A--.D...HK---V-----S--RR---------FY---.........................................-   213
H1P.CM.06.U14788            ----F---W-------------.....E--------LF-----SEVEA--.MGD-Q-KAK----ACT....Y-FS-.-...HK-I-V-K---S-GRE---LQK---L-IKD.........................................-   218
CPZ.CD.06.BF1167            N---I---W-------------.....-----CR--LF----AD-PEED-.......RNL-M--ACS....Y-R--.-...NG-L-V-K---E--RR-I---R------............................................-   196
CPZ.TZ.00.TAN1              N---V---W-------------.....I----CR--LF-----D-PEDD-.......KNI----ACS....--TT-.-...DG-T-I-----S--RR-I---RY-----...........................................-   196
GOR.CM.07.SIVgorCP2135con   --------W-------------.....I--------LF-----SEAEA--.MGTAS-RAI----ACS....--R--.-...HG-I-V----RA-GTT---LQK---L--RE.........................................-   211
                                        max HIV-1 similarity                                                                                               normal Nef end
MAC.US.x.239                KEE-II--W-D--S----------.....I---K----LW-----NVSDEAQ..EDE..EHY-M--AQT....SQW--.-...WG---A-K--PT--YTYE-YVRY--E-GSKSGLSEEEVRRRLT...ARGLLNMADKKETR.........-   263
H2A.SN.86.ST_JSP4_27        KEE-IIG-W----H----------.....----KF---LW-----DVPQEGD..DSE..THC-V--AQT....SRF--.-...HG-T-V----PT---SYE-FIRY--E-GYKSGLPEDEWKARLK...ARGIPFS................-   256
H2B.CI.x.EHO                NEE-IVSGW----H----------.....----KF---LW----INMIAEP-..DEE..THC-V--AQT....SAW--.-...HE-T-V-Q---L--YDY--FSRF--E-GYQSGMPEKEWKAKLR...ARGIPTE................-   239
H2G.CI.92.Abt96             KE--II--W----A----------.....I---X----LW-----TVSDEAQ..EDE..THC-V--AQT....SPW--.-...-T---A-K--PT--YDYR-FI----E-RWKSGLPEAVWKEKLK...QRGLPIE................-   257
H2U.CI.07.07IC_TNP3         NEE-IVA-W----E----------.....--F-K----LW-----HVSDEAS..NDE..THC-V--AQT....SQEE-.R...WG---A-K--AT--YDYK-FI-Y--E-GSKSGLPEKEVKRRLA...ARGLLNMADKKETS.........-   264
H2U.FR.96.12034             NEE-II--W----S----------.....T-W-MC---LW--E--DANDKAQ..EDE..RLY-VGSAQT....SCEE-.H...WG-A-V-K---S--YSYQ-FIKC--E-GSKSGLSEEEVKRRLT...ARGLPVKNC..............-   261
ASC.UG.10.RT03              NVW-LV-GW-G--Q----------.....T-F-TC--ILWE-----IGPQGLGEGDE..RA----AGQQVYQD..........HC-T-V-H-NPT---EPGILKADKLGQ-PLPGQKTFLTGRD...........................-   238
ASC.UG.10.RT08              -VW-LT-GW-G--Q----------.....P-F-TC--ILWE-----VGPQGLGEGDE..RAK---AGQQSFQD..........YG---A-H-NPT---EPGILKADKLGQQPLPGQETFLTGRD...........................-   235
ASC.UG.10.RT11              -VW-VV-GW-G-------------.....P-F-TC--ILWE-----DVGPQGLGEGDE..RAL---AGQQAYQD.......L...HG-T-V-H-NPT---EPGILKADKLGQQPLPGMKSCITGRD...........................-   237
COL.CM.x.CGU1               IEW-CLKGWLE-EDEL-EDGALKED-K--VA--LW---YI-QLGEYAY......SYD-SLLSVT....SRKKKK-QQ.......VAIEMVD............................................................-   224
COL.UG.10.BWC01             IEW-VLKGWLE-EDETDENMQIREE-K--VA--LW---FI-QLGQWAY......SYDMSLLSIN....NRKKKPQQQQ.VA-DTVAIETPD........................................................-   229
COL.UG.10.BWC07             IEW-VLS-WLE-EDETNEDGTPKEE-L--AC--LW---LLDSIGKWAY......CFDPALLNRP....VTKKKKQDAPN-STVAMEIETVD........................................................-   233
DEB.CM.04.SIVdeb04CMPF3061  NEW-II-GWLQ--E----------.....I---KY--FLF------IADPDY.ENDE..RNI---DAHQ....GQ-E-.-...YK-R-V-K---Q--YCYK-GHEAHTKETHTRRCMFPKRK.............................-   232
DEB.CM.99.CM40              NEW-II-GWLQ--K----------.....----IY--FLF------IADPDY.ENDE..RNI---DAHQ....GM-E-.-...H--R-V-K---T--YCYK-GH..A-QREHTRRCMFPKRK.............................-   228
DEB.CM.99.CM5               -EW-II-GWLK--E----------.....P---TY--FLF------IADPDY.ENDE..RNI---DAHQ....GQAE-.-...HK-R-V-K---S--YCYK-GHE.AHKPEHNRRCMFPKRK.............................-   231
DEN.CD.x.CD1_CMO580407      NEH-FITGW-T--D----------.....I----E--FLW-----TIEEEYD..NKE..ENC---DRYE....GQQA-.-...W--T-V----PE--WCYK-GHKKLQTSQHNNYCQRAKPRK...........................-   249
DRL.DE.11.D3                NEW-FIKGW-S--K----------.....I---KV--FLF-----DV-DDLI..DQP..CNR--NSSQM....GPI--.S...-G-R-M-Q--AS--YTFQ-CIH---E-GHVTSIRHD...KREEPDCCKRKWWQX................-   256
DRL.x.x.FAO                 NEW-FIKGW-S--K----------.....I---KM--FLF------V-EDLV..NQQ..CNR--NSSQM....GPI--.-...-G-R-M-Q--PG--YTFQ-CIH---E-GHVKSIK.....SVEEPGCCKRKWWQFKPTTEGCHGDNLQKC-   269
GRV.ET.x.GRI_677_gri_1      NEW-IID-W-A-S-----------.....I---RV--F-F-----DLHEEAR..NCE..RHC-M--AQMG..EDPDGI.D...HG---V-K--PK--VEYRPDM..FKDMHEHAKR....................................-   224
GSN.CM.99.CN166             LEW-LV-GWLSF------------.....T---TIP-F-IC-R--ATTEDS-.PGDD..EYL-T--AYQ....GRSE-.Q...HK-F-VFS-C-K--IKSGIQLDQLQQEERKMRLTANRFL.............................-   219
GSN.CM.99.CN71              YEW-IV-GWLQW------------.....I---TMP-F-WC-R--AMTEDS-.PGDD..QYL-N--AYQ....GQQE-.H...H--I-VFS-C----LKSGWQMNQLQQEERKKRLTANRFL.............................-   218
LST.CD.88.SIVlhoest447      NEW-LIG-FMT--D----------.....T-F--A---LWQ-E--AC-EYTD.PSDY..RQC---SSQ-....GVQE-.-...WG-R-I-H-NPM--VDY--LRKQ-KSIQASAIGL..QRK............................-   206
LST.CD.88.SIVlhoest485      NEW-LIG-FMT--D----------.....T--------LWQ-E--AC-EYTD.PSDA..SQC---SSQ-....GVQE-.-...WG-R-I-H-SPM--VDF--LRKQ-KSIQASSLSLNCKRK.............................-   208
LST.CD.88.SIVlhoest524      NEW-LIG-FM---D----------.....T----A---LWQ-E--AC-EFKD.PSDE..RQC---SSQ-....GVLE-.-...WG-R-I-H-NPM--VDY--LRKQ-ASVQATAMRFNCERK............................-   207
LST.KE.x.lho7               NEW-LIG-FM---D----------.....T--------LWQ-E--AC-EYKD.PSDE..TQC---SSQ-....GVLE-.-...WG-R-I-H-NPM--VDFI-LKKQ-AKIQ-TAFAFDCKRK............................-   208
MAC.US.x.251_1A11           KEE-II--W-D--S----------.....I---K----LW-----NVSDEAQ..EDE..EHY-M--AQT....SQW--.-...WG--PA-K--PT--YTYE-YVRY--E-GSKSGLSEEEVRRRLT...ARGLLNMADKKETR.........-   264
MND-1.GA.x.MNDGB1           NEW-FE-GW-Q--T----------.....T---K---FLF--E--SRAIGD-..YAA..-NH---SSQ-....CPQE-.-...-G-T-M-#W-PH--YDFA-LX...............................................-   215
MND-2.CM.98.CM16            NEW-FI-TW-S--D----------.....I---K--RFLF--C--AVPPDQ-..N-E..CNK--QSSQ-....GIQEE#-...WG-R-V-K---G--YTFY-PIKR-GE--HVQSISHIAYAKEHKPECCKRKWWQF...............-   231
MND-2.GA.x.M14              NEW-FI--W-E--S----------.....I---KR--FLF--I--QVPPDQ-..NQE..CNR--NSSQ-....GIQEE#-...WG-R-M-K--PE--YTFY-PIKC-KEYRHVTSLSYTAY..QEKSDCCKRKWWQF...............-   233
MND-2.x.x.5440              NEW-LI-NW----K----------.....I---K---FLF-----AVSPPL-..EDE..CNR--NSSQM....GIQE-.-...-G-R-M-K---G--YTFY-PIIR--EY-CVTSLSYEAYKKEEKPDCCKRKWWQF...............-   251
MNE.US.x.MNE027             KEE-IV--W-D--S----------.....P---K----LW-----NVSDEAQ..EGE..ENY----AQT....SQW--.-...WG---V-K--PT--YTYE-YIRY--E-GSKSGLSEEEVRRRLT...ARGLLKMADKKETS.........-   264
MON.CM.99.L1_99CML1         LEW-LI-GWLQ--E----------.....----TMP-FLWC-R--ATTEDS-.EGDE..DFL-T--AYQ....GR-E-.-...H-QF-VFS-C-K--VKSGRQLAQLQQEERKKRLAANRIL.............................-   225
MON.NG.x.NG1                VMW-LV-GWLQF-X----------.....----TIP-FLWC-R--AIQEDS-.DGDD..EFL-T--AYQ....GRDE-.-...HKQF-VFS-C-K-GVKSGPQLDELQQEERKRRLTANRIL.............................-   229
MUS-1.CM.01.CM1239          VDW-IV-GWLQF-E----------.....T---TLP-FLWC-RE-AITEDS-.EGDE..EYL-T-AAYQ....GKEE-.-...HKQF-VFS-C----MKSGRQLDQIQQEERKRRLTANRIL.............................-   218
MUS-1.CM.01.SIVmus_01CM1085 VEW-II-GWLM--D----------.....T---KIP-FLWC-RE-AMTEGS-.EGDD..QYL-D--AFQ....GREE-.H...HKQF-VFS-C----LKSGLQLDQMQQEERKKRLATNHIL.............................-   218
MUS-2.CM.01.CM1246          VKSKIV-RSLI--D----------.....T---RQP-FLWC-R--AMTEES-.PGDD..QYL-T--AYV....GRDE-.H...HK-F-VFS-C----LKSGPQLNQIQQKERKRRLTANRIL.............................-   218
MUS-2.CM.01.CM2500          VEW-LI-GWLM--E----------.....------MP-VLFC-R--HITENS-.LGDD..EHL-C--AFI....GREE-.-...HK-F-VFS-C----LKSGRQLNQMQQEERKRRLAANRILSQSP.........................-   222
MUS-3.GA.09.09GabOI81       MEW-IV-GWLRF------------.....T---TMP-FLWC-R--AMQEDS-.AGDD..EHL-T--AFQ....GRDE-.I...HK-F-VFS-C-K--IKSGKQLDQLQQEERKKRLTANRFL.............................-   218
MUS-3.GA.11.11GabPts02      VEW-IV-GWLS-------------.....----RIP-FLWC-R---IAEDPG.EEDD..TFL-A--AYQ....GREE-.H...HK-F-VFS-C----LKSGQQLDQLQQAERRKRLTANRIL.............................-   218
OLC.CI.97.97CI12            NCW-IIGNF-ATEEVEEG.....TL-R-K-P--LWD-RR---LPETLEERG.......................................................................................-   152
RCM.CM.00.SIVagi_00CM312    NEH-II--W---------------.....P---TF---LWM----DVSDEAK..EDE..EH-----AES....S-IE-.-...WG-T-A-K-NPM--VDYIGYR---DF-GERKNKTQ.................................-   225
RCM.CM.02.SIVrcm_02CM8081   NEH-II--W----K----------.....I---MM---LW-----DVIDEAK..EDE..EH-----AET....C--E-.-...WG---A-K--PM--IDY-GYK----D-GEKKNKTQ.................................-   225
RCM.GA.x.SIVRCMGAB1         NEH-II--W---------------.....I---TL---LWQ----DVSDEAR..EDE..EH-----AET....S--E-.-...WG---A-K-NPM--VDYIGYR----F-GERKNKTQ.................................-   225
RCM.NG.x.NG411              NEH-II--W---------------.....I---TM---LWQ----DVSDEAT..E-E..RHC----AET....S----.-...WG---A-K-KPL--VDYAGYR----F-GERKNKTK.................................-   225
SAB.SN.x.SAB1               NEW-IVDGW----D----------.....T---KC----F-----DLSEEAK..NSE..-HC----AQV....AYE--.A...WK-T-V-K--PL--VDY--WR----QVPSAQG...................................-   227
SMM.CI.79.SIVsmCI2          NEE-II--W----S----------.....I---MM---LW-----NVSDEAQ..EDE..RHC-V--AQI....SHL--.-...WG---A-K--PQ--YNYE-FIRY--E-GSKSGLSEDEVKKRLT...ARGLLKMADKKENS.........-   263
SMM.LR.89.SIVsmLIB1         KEE-IV--W----S----------.....I---MF---LW----MDVSDEAQ..EDE..THC-V--AQT....-Q---.-...WG---A-S--PK--YTYE-YVRY--E-GSKSGLSEDEVKRRLT...ARGLLKMADKKETR.........-   262
SMM.SL.92.SIVsmSL92A        NEE-II--W----R----------.....I---MF---LWQ----DVSDEAN..EDE..-HC--#-AQT....YK-S-.-...WG-T-A-K-NPM--YTYE-FVRY--E-GWKSGLTEEEVKRRLAS..RKKPNHKKQHGAEEGS........-   257
SMM.SL.92.SL92B             KEE-II--W----S----------.....I---KF---LWQ----DVSDEAN..NDE..THC-V--AQT....YQHS-.-...WGK--A-K-NPH--YTYE-FVR---E-GWKSGLTEEEVERRLANK.PKPQKKMADKKETS.........-   267
SMM.US.04.G078              KEE-IV--W----S----------.....----MF---LW-----DVSDEAQ..EDE..THC-V--AQT....-QW--.-...WG---A-K--PQ--YTYE-FIRY--E-GMQSGLSEDEVQRRLT...ARGLLKMADKGGNX.........-   263
SMM.US.05.D215              NEE-IV--W----S----------.....I---M----LW-----DVSDEAQ..EDE..LHC-V--AQT....AQW--.-...WG---A-K--AT--YNYE-FIKY--E-GSKSGLSEEEVRRRLT...ARGLLKMADKGKQ..........-   261
SMM.US.06.FTq               KEE-II--W----S----------.....I---MM---LW------VSDEAQ..NDE..THC-V--AQT....GQW--.-...WGK--A-K--PT--YTYE-YIKY--D-GDKSGLSEEEVKRRLT...ARGLLKMADKKETX.........-   263
SMM.US.86.CFU212            KEE-II--W----S----------.....I---MF---LW-----DVSDEAQ..EDE..THC-V--AQT....YQW--.-...WG---A-K--PT--YTYQ-FVKY--E-GYKSGLSEEEVKRRLT...ARGL#KMADKKETX.........-   262
SMM.US.x.H9                 KEE-II--W----A----------.....I---MF---LW--X--NVSDEAQ..EDE..THC-M--AQT....SQW--.-...WG---A-K--PX--YXYK-FVE---E-GSQSGLSKEEVQRRLT...ARGLLKMADKKKTS.........-   262
SMM.US.x.PGM53              KEE-IT--W----S----------.....I---MF---LW-----SASDEAQ..EDE..THY-V--AQI....SQW--.-...WG---A-K---Q--YRYE-FIR---E-GSKSGLSEEEVKRRLT...ARGLLKMADKKETS.........-   264
SMM.x.x.pE660.CG7G          KEE-II--W----S----------.....T---MYY--LW-----DVSDEAQ..EDE..THC-V--AQT....-QW--.-...WG---A-K--PE--YSYK-FIKY--E-GSKSDLSEEEVKRRLT...ARGLIKMADKKETS.........-   264
STM.US.89.STM_37_16         KEE-IV--W----A----------.....I---KQ---LW-----DMSNEAQ..EDD.GTHY-V--AQT....-QW--.-...WG---V-K--PL--HTYE-FVR---E-GSKSGLPKEEVERRLT...ARGLLKMADKKETS.........-   265
SUN.GA.98.L14               NEW-FITGW-D--K----------.....----KA---LW--A--TI-EDRD.PNHP..CQA---SSQQ....GVNE-.-...WG-R-I-T--PT--YDFR-IQK---E--HVTSLQWEV..............................-   209
SYK.KE.x.KE51               NEH-VI.-DIT--S----------.....T----I---LWE-A-N-IEGYLV..DEE..D-LMM--AAGV....GASE-.-...H--N-M-N-NPH--YTPGWEMARQQLERQTGKR................................-   229
SYK.KE.x.SYK173_COMGNM      NEH-II.-RI---S----------.....T----I---LWE---N-IEGCL-..YEE..H-L----A-GQ...GSSS#......W-N-M-N-NPH--I-Q-G-WPACS*RDKQENHKSCSQHLVRTSPX......................-   240
TAL.CM.00.266               NEW-IL--W----E----------.....T----C--ILW--C---IHEDDT.....EDGHL-P--AYD....GQAE-.-...WG-S-V-V--EK--YTPG-KMAEWDRLEREKRMLLAPPQTASS........................-   241
TAL.CM.01.8023              NEW-IL--W-S--E----------.....T----C--ILW--C---IHEDDT.....EDGHL-M--AYD....GQQE-.-...WG-A-V-V--EK--YTPG-KMAEYDRLERKKKELLAPPQTASS........................-   241
TAN.UG.x.SIVagmTAN1         NEW-IID-W-A-S-----------.....T-K-RC--F-FE----DVSQEAQ...DE..RHC----AQI....EWES-.-...WK-T-V-K-NPL--VQYNPDS..FKDMHGLVKRK..................................-   223
VER.DE.x.AGM3               NEW-IID-WNAWSE----------.....I---RC--F-F-----DLHEEA-..TCE..RHC-V--AQVR..EDPDGI.N...HG---V-K--PM--VQYDPNRKYLTDMHDLGKRK................................-   230
VER.KE.x.9063               NEW-IID-WNAWST----------.....----RC--F-F------MHEEA-..TCE..RHY-A--AQIK..EDPDGI.S...HG-T-V-K--PM--VQYDPNRQYF-DMHAIVKRK.................................-   230
VER.KE.x.AGM155             NEW-IID-WNAWSK----------.....I---RC--F-F-----ALHEEA-..TCE..RHC-V--AQ-H..EDPDGI.N...HG-I-A-K--PM--VQYDPSREYFTDLYSTVGTGN...............................-   231
VER.KE.x.TYO1_patent        NEW-IID-WNAWSK----------.....I-F-KC--F-F-----DLHEEAQ..TCE..RHC-V--AQMG..EDPDGI.S...HG-I-V-K--PM--IQYDPNREYFTDMHGLVKRK................................-   230
WRC.CI.97.97CI14            NEW-LIDGW-A--E----------.....I-W-K-----W-----HFSEESDNTDPYY-RNK----AC.....AEPE-.-...WG-H-V-K--PK--CDF--GR-PGQSIATGKEAIESLYKNKKQCK.........................-   234
WRC.CI.98.98CI04            NEW-LIEGW-A--E----------.....I-W-K-----W------FSEESNDSSQYY-KNK----GC.....QESE-.-...WG-H-V-K--P---CDF--SR-PGHIIAQGEDAIETLHKEKLKRHK.........................-   234
WRC.GM.05.Pbt_05GM_X02      -EW-IVEGW-A--D----------.....----K-----W-----HFSEESPNSDQGF-KNL----AV.....AESE-.-...WK-Y-V-K--P---CDF--GR-ADGQIATGIETLYEMKKRNNK...........................-   231
```