VII

PLV Proteins

Contents

VII-1 Introduction **399**
VII-2 Sequences **400**
VII-3 Alignments **408**
VII-3.1 Gag 408
VII-3.2 Pol 412
VII-3.3 Vif 419
VII-3.4 Vpx 422
VII-3.5 Vpr 423
VII-3.6 Tat 424
VII-3.7 Rev 426
VII-3.8 Vpu 428
VII-3.9 Env 429
VII-3.10 Nef 436

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. For example, several diverse African green monkey virus isolates have been sequenced only in a region of the *env* gene, and recent Gorilla sequences are only available in the pol and env genes. When necessary, some of the more common sequences (such as HIV-1 M group) were removed to make room on the Compendium pages for these diverse virus sequences. More complete alignments are available from our web site `http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html` where space limitations are not an issue.

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

VII-2 Sequences

Sequences included in the PLV protein alignments.

Name	Accession	Proteins	Author	Reference
H1B.FR.83.HXB2	K03455	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Wong-Staal, F.	*Nature* **313**(6000):277-284 (1985)
H1A1.UG.85.U455	M62320	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Oram, J.D.	*ARHR* **6**(9):1073-1078 (1990)
H1C.ET.86.ETH2220	U46016	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Salminen, M.O.	*ARHR* **12**(14):1329-1339 (1996)
H1D.CD.84.84ZR085	U88822	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*J Virol* **72**(7):5680-5698 (1998)
H1F1.BE.93.VI850	AF077336	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T.	*Virology* **269**(1):95-104 (2000)
H1G.SE.93.SE6165	AF061642	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Carr, J.K.	*Virology* **247**(1):22-31 (1998)
H1H.CF.90.056	AF005496	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*J Virol* **72**(7):5680-5698 (1998)
H1J.SE.93.SE9280_7887	AF082394	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T.	*ARHR* **15**(3):293-297 (1999)
H1K.CM.96.96CM_MP535	AJ249239	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Triques, K.	*ARHR* **16**(2):139-151 (2000)
H1O.BE.87.ANT70	L20587	Gag, Tat, Rev	Vanden Haesevelde, M.	*J Virol* **68**(3):1586-1596 (1994)
H1O.CM.94.BCF06	AB485666	Gag, Tat, Rev	Takekawa	Unpublished
H1O.CM.96.96CMA102	AY169803	Gag, Tat, Rev	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.96.96CMABB009	AY169806	Gag, Tat, Rev	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.97.97CMABB447	AY169813	Gag, Tat, Rev	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.98.98CMA104	AY169802	Gag, Tat, Rev	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)

Name	Accession	Proteins	Author	Reference
H1O.CM.98.98CMABB197	AY169811	Gag, Tat, Rev	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.98.98CMU2901	AY169812	Gag, Tat, Rev	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.CM.99.99CMU4122	AY169815	Gag, Tat, Rev	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.FR.92.VAU	AF407418	Gag, Tat, Rev	Vartanian, J.P.	*J Gen Virol* **83**(Pt 4):801-805 (2002)
H1O.US.99.99USTWLA	AY169814	Gag, Tat, Rev	Yamaguchi, J.	*ARHR* **19**(11):979-988 (2003)
H1O.US.x.I_2478B	AB485668	Gag, Tat, Rev	Takekawa	Unpublished
H1N.CM.02.DJO0131	AY532635	Gag, Tat, Rev	Bodelle, P.	*ARHR* **20**(8):902-908 (2004)
H1N.CM.02.SJGddd	GQ324959	Gag, Tat, Rev	Vallari, A.	*ARHR* **26**(1):109-115 (2010)
H1N.CM.04.04CM_1131_03	DQ017383	Gag, Tat, Rev	Yamaguchi, J.	*ARHR* **22**(1):83-92 (2006)
H1N.CM.06.U14296	GQ324962	Gag, Tat, Rev	Vallari, A.	*ARHR* **26**(1):109-115 (2010)
H1N.CM.06.U14842	GQ324958	Gag, Tat, Rev	Vallari, A.	*ARHR* **26**(1):109-115 (2010)
H1N.CM.95.YBF30	AJ006022	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Simon, F.	*Nat Med* **4**(9):1032-1037 (1998)
H1N.CM.97.YBF106	AJ271370	Gag, Tat, Rev	Roques, P.	*AIDS* **18**(10):1371-1381 (2004)
H1P.FR.06.RBF168	GQ328744	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env	Plantier, J.-C.	*Nat Med* **15**(8); 871-2 (2009)
CPZ.CD.90.ANT	U42720	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, M.M.	*Virology* **221**(2):346-350 (1996)
CPZ.CM.01.SIVcpzCAM13	AY169968	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Nerrienet, E.	*J Virol* **79**(2):1312-9 (2005)
CPZ.CM.05.SIVcpzEK505	DQ373065	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.CM.05.SIVcpzLB7	DQ373064	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.CM.05.SIVcpzMB66	DQ373063	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	*Science* **313**(5786):523-526 (2006)

Name	Accession	Proteins	Author	Reference
CPZ.CM.05.SIVcpzMT145	DQ373066	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.CM.98.CAM3	AF115393	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Corbet, S.	*J Virol* **74**(1):529-534 (2000)
CPZ.CM.98.CAM5	AJ271369	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Muller-Trutwin, M.C.	*J Med Primatol* **29**(3-4); 166-72 (2000)
CPZ.CM.x.SIVcpzDP943	EF535993	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Heuverswyn, F.V.	*Virology* **368**(1):155-171 (2007)
CPZ.CM.x.SIVcpzMB897	EF535994	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Heuverswyn, F.V.	*Virology* **368**(1):155-171 (2007)
CPZ.GA.88.GAB1	X52154	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Huet, T.	*Nature* **345**(6273):356-359 (1990)
CPZ.GA.88.GAB2	AF382828	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bibollet-Ruche, F.	*ARHR* **20**(12):1377-1381 (2004)
CPZ.TZ.00.TAN1	AF447763	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Santiago, M.L.	*J Virol* **77**(3):2233-2242 (2003)
CPZ.TZ.01.TAN2	EF394357	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	*J Virol* **81**(14):7463-7475 (2007)
CPZ.TZ.02.TAN3_1	DQ374658	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	*J Virol* **81**(14):7463-7475 (2007)
CPZ.US.85.US_Marilyn	AF103818	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*Nature* **397**(6718):436-441 (1999)
MAC.US.x.239	M33262	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kestler, H.	*Science* **248**(4959):1109-1112 (1990)

Name	Accession	Proteins	Author	Reference
H2A.DE.x.BEN	M30502	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kirchhoff, F.	*Virology* **177**(1):305-311 (1990)
H2B.CI.x.EHO	U27200	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Rey-Cuille, M.A.	*Virology* **202**(1):471-476 (1994)
H2G.CI.92.ABT96	AF208027	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Brennan, C.A.	*ARHR* **13**(5):401-404 (1997)
H2U.FR.96.12034	AY530889	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Damond, F.	*ARHR* **20**(6):666-672 (2004)
COL.CM.x.CGU1	AF301156	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	*J Virol* **75**(2):857-866 (2001)
DEB.CM.04.SIVdeb04CMPF3061	FJ919724	Gag, Pol, Vif, Vpr, 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	AJ580407	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Dazza, M.C.	*J Virol* **79**(13):8560-8571 (2005)
DRL.x.x.FAO	AY159321	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	*J Virol* **77**(8):4867-4880 (2003)
GOR.CM.04.SIVgorCP684con	FJ424871	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	*J Virol* **83**(4):1635-1648 (2009)
GOR.CM.07.SIVgor2139_287	FJ424866	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	*J Virol* **83**(4):1635-1648 (2009)

Name	Accession	Proteins	Author	Reference
GRV.ET.x.GRI_677	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.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)
MND-1.GA.x.MNDGB1	M27470	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Tsujimoto, H.	*Nature* **341**(6242):539-541 (1989)
MND-2.CM.98.CM16	AF367411	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Takehisa, J.	*ARHR* **17**(12):1143-1154 (2001)
MND-2.GA.x.M14	AF328295	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Souquiere, S.	*J Virol* **75**(15):7086-7096 (2001)
MND-2.x.x.5440	AY159322	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	*J Virol* **77**(8):4867-4880 (2003)
MNE.US.x.MNE027	U79412	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kimata, J.T.	*J Virol* **72**(1):245-256 (1998)
MON.CM.99.L1	AY340701	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)

Name	Accession	Proteins	Author	Reference
MUS-1.CM.01.CM1239	EF070330	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
MUS-1.CM.01.SIVmus_01CM1085	AY340700	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)
MUS-2.CM.01.CM1246	EF070329	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
MUS-2.CM.01.CM2500	EF070331	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
OLC.CI.97.97CI12	FM165200	Gag, Pol, Vif, Vpr, Tat, Rev, Env	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
RCM.CM.00.SIVagi_00CM312	HM803690	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env	Ahuka-Mundeke, S.	*J Gen Virol* **2010** Aug 25
RCM.CM.02.SIVrcm_02CM8081	HM803689	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env	Ahuka-Mundeke, S.	*J Gen Virol* **2010** Aug 25
RCM.GA.x.SIVRCMGAB1	AF382829	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
RCM.NG.x.NG411	AF349680	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env	Beer, B.E.	*J Virol* **75**(24):12014-12027 (2001)
SAB.SN.x.SAB1	U04005	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Jin, M.J.	*EMBO J* **13**(12):2935-2947 (1994)
SMM.SL.92.SL92B	AF334679	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Chen, Z.	*J Virol* **70**(6):3617-3627 (1996)
SMM.US.x.H9	M80194	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	*J Virol* **66**(1):414-419 (1992)

Name	Accession	Proteins	Author	Reference
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.US.x.SIVsmH635F_L3	DQ201172	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kuwata, T.	*J Virol* **80**(3):1463-75 (2006)
STM.US.89.STM_37_16	M83293	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Novembre, F.J.	*Virology* **186**(2):783-787 (1992)
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	L06042	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	*J Virol* **67**(3):1517-1528 (1993)
TAL.CM.00.266	AY655744	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*Virology* **349**(1):55-65 (2006)
TAL.CM.01.8023	AM182197	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*Virology* **349**(1):55-65 (2006)
TAN.UG.x.SIVagmTAN1	U58991	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Soares, M.A.	*Virology* **228**(2):394-399 (1997)
VER.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)

Name	Accession	Proteins	Author	Reference
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)
H1B.US.90.WEAU160	U21135	Pol, Vif, Vpr, Vpu, Env, Nef	Wei, X.	*Nature* **422**(6929):307-312 (2003)
H101_AE.TH.90.CM240	U54771	Pol, Vif, Vpr, Vpu, Env, Nef	Carr, J.K.	*J Virol* **70**(9):5935-5943 (1996)
H102_AG.NG.x.IBNG	L39106	Pol, Vif, Vpr, Vpu, Env, Nef	Howard, T.M.	*ARHR* **10**(12):1755-1757 (1994)
H103_AB.RU.97.KAL153_2	AF193276	Pol, Vif, Vpr, Vpu, Env, Nef	Liitsola, K.	*ARHR* **16**(11):1047-1053 (2000)
H104_cpx.CY.94.CY032	AF049337	Pol, Vif, Vpr, Vpu, Env, Nef	Gao, F.	*J Virol* **72**(12):10234-10241 (1998)
H2A.GW.x.ALI	AF082339	Pol, Vif, Vpx, Vpr, Env, Nef	Azevedo-Pereira	Unpublished
H2A.SN.x.ST	M31113	Pol, Vif, Vpx, Vpr, Env, Nef	Kumar, P.	*J Virol* **64**(2):890-901 (1990)
H2B.GH.86.D205	X61240	Pol, Vif, Vpx, Vpr, Env, Nef	Dietrich, U.	*Nature* **342**(6252):948-950 (1989)
GOR.CM.2007.SIVgorCP2135con	FJ424863	Pol, Vif, Vpr, Vpu, Env, Nef	Takehisa, J.	*J Virol* **83**(4):1635-1648 (2009)
LST.CD.88.485	AF188115	Pol, Vif, Vpr, Env, Nef	Beer, B.E.	*J Virol* **74**(8):3892-3898 (2000)
MAC.US.x.251_1A11	M76764	Pol, Vif, Vpx, Vpr, Env, Nef	Marthas, M.L.	*J Med Primatol* **18**(3-4):311-9 (1989)
MAC.US.x.251_BK28	M19499	Pol, Vif, Vpx, Vpr, Env, Nef	Hirsch, V.	*Cell* **49**(3):307-319 (1987)
MAC.US.x.EMBL_3	Y00295	Pol, Vif, Vpx, Env, Nef	Franchini, G.	*Nature* **328**(6130):539-543 (1987)
MON.NG.x.NG1	AJ549283	Pol, Vif, Vpr, Vpu, Env, Nef	Barlow, K.L.	*J Virol* **77**(12):6879-6888 (2003)
VER.DE.x.AGM3	M30931	Pol, Vif, Vpr, Env, Nef	Baier, M.	*Virology* **176**(1):216-221 (1990)
VER.KE.x.9063	L40990	Pol, Vif, Vpr, Tat, Env, Nef	Hirsch, V.M.	*J Virol* **69**(2):955-967 (1995)

```
                              membrane binding                                                                                          phosphorylation site
                              Gag start, p17 start  nuclear localization                                          nuclear localization                      p17 end  p24 start
H1B.FR.83.HXB2               MGARASV.LSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQ..........................NKSKKKAQQAAADTG.HSNQVSQNYPIVQNIQG  140
H1A1.UG.85.U455              -------.---KK--S----------N---R---L--------K-TL-------A---Q---------A----T-----------V--------DV-------N----M-..........................--N-QRT-----N-......S---------A--  136
H1C.ET.86.ETH2220            ------I.-R-EK--A----K-------H-M---L---N----K--L--D--D--A--K--IK----A----T---K--F-----------K-----------------..........................-E-QQ-T---G-ADR...GKD---------M--  138
H1D.CD.84.84ZR085            -------.----K--A--------------------------------L----------K--ME----A-K--------I-----------K-----------------..........................-N---R------G--.N-S-----------L--  140
H1F1.BE.93.VI850             ------I.----K--E----Q-------R--M--LI----------LD----------QK-IR------------K--F---V--Y---AGVT--------L----..........................---QQ-T------K-.........---------L--  136
H1G.SE.93.SE6165             -------.-T--K--A----------R-S--I--L-----------L--D----A---Q--MR---------T--IK----A-----------V-------EEV-KI-..........................K--QE-I----M-K-.N------------A--  140
H1H.CF.90.056                -------.----K--A--------------R---L-----------L-------P---L--IE-I--AIK--T---K--F-L--V-----RK-DV-----------I-..........................---QQ-T------KE.KD-K----------A--  140
H1J.SE.93.SE9280_7887        ------I.----K--D--------------RI--L-------D---L------SAK--Q---V---A----TQ-IK--------------------M---E----I-..........................--N-QQ--K-ET-KK.DNS----------L--  140
H1K.CM.96.96CM_MP535         -------.----K--A-------------------L-----------L-------T------IT-I---I------IK-----I-V--F---K--V---------L----..........................---QR-T--.E-A...DKG----------L--  136
H1O.BE.87.ANT70              ---S---.-T-SK--A--Q---K--S----R---L-----------C--E----A--NEKL-Q--E-A-K---DS-Q--W-AIVV-W---N-YK-G--QQ-IQ-LK-VMG..........................SRKSAD--KEDTSAR-AG------S-A--  138
H1O.CM.94.BCF06              -------.-T-SK--A--R------S--T-----L---------N--CD----SA--NE-L-Q--E-A-K----D-K--W-AI-V-W---N-F-VQN-QQ-IQ-LK-VME..........................NRKSAD--KGETS-K-T-----V-A-A--  138
H1O.CM.96.96CMA102           -------.-T-SK--A--Q---K--S----R---L-------D---C--E----A--NEKL-Q--E-A-K----S-Q--W--I-V-W---N-YK-E--QQ-IQRLR--KG..........................NRKSAG--KEDTXTR-TX-----I--A--  138
H1O.CM.96.96CMABB009        -------.-T-SK--A--H---K--S--H-R---L-------D---C--E----A--NE-L-Q--E-A-K---DS-Q--W--I-V-W---K-YKVE--QQ-IQ-LK-VMR..........................NRKS...-NEDTXXK-T-----VXASA--  135
H1O.CM.97.97CMABB447        -------.-T-SK--D--Q-------S----R---L-----------C--E----A--NEKL-Q--E-A-K----D-K--W-AI-V-W---N-YKVG--QQ-VQ-LK-VMG..........................NRKSADT-KETTS-K-T-----V-A-M--  138
H1O.CM.98.98CMA104          -------.-T-SK--A--Q-K-K--C----R---L-----------C--E----A--NEKL-Q--E-A-K----S-Q--W--I-V-W---N-FKVE--QQ-IQ-LR--KG..........................GRKSAG--KEDTSAR-TG----V---A--  138
H1O.CM.98.98CMABB197        ---S---.-T-SK--A--Q---K--S---------------T--C--E----A--NEKL-Q--E-A-K--TRD-Q--W--IVV-W---N-YKVE--QQ-IQ-LK-VME..........................S-KSAS--KEDTSAK-TX-----I--A--  138
H1O.CM.98.98CMU2901         ---S---.-T-SK--A--R---K--S----R---L-------D---C--E----A--NEKL-Q--E-A-K----S-Q--W--I-V-W---N-YKVE--QQ-IQ-LK-VMG..........................SRKSAS--KEETSAR-TG----V-P-A--  138
H1O.CM.99.99CMU4122         ---S---.-K-SK--A--Q---K--S----R---L-----------C--E----A--NEKL-Q--E-A-K---DS-Q--W--IVV-W--XN-YR-E--QQ-IK-LR-VMX..........................SXKSASD-KEDTXAR-TG-----IP-A--  138
H1O.FR.92.VAU                -------.-T-TK--A--Q---K--S--------LI----G-D---C--E----A--NE-L-Q--E-A-K----S-*--W-AI-V-W---N-YKVE--QQ-VR-LK--MK..........................-RKPVSSTKENTS-K-TG----V-A-A--  137
H1O.US.99.99USTWLA           I------.-T-KK--T*-Q----S-S----Q---L----K------C--D----A--NE-L-Q--E-A-K---DS-Q--*-AIVV-R---K-YRV---QQ-IQ-LX-VXX..........................SXKSAD.NXEDTXTK-T-------T-A-X  135
H1O.US.x.I_2478B             ---S---.-T-SK--A--Q---K--S----R---L-------D---C--E----A--NE-L-Q--E-A-K----S-Q--W--I-V-W---N-YK-G--QQ-IQ-LK-VMG..........................NRKSAS--KEDTSAK-AG----V-S-A--  138
H1N.CM.02.DJ00131            -------.-T--K--Q--A-Y---------RM--L-----------C----MD-AS--A-L-N--E-A-A----G------AL-V-----SN-QVHN-Q------K-IKE...................KKERH#SPSQ-T-#-G-AAA.AD-NI-R---L---A--  144
H1N.CM.02.SJGddd             -------.-T--K--Q--S-Y---------R---L-------D---C----MD-AV--A-LIN--E-A-K----G-------L-V-----S--QVHN-Q------K-K-E....................QHKPP.EP.-NPE-G-AAA.VXSNI-R---L--TA--  143
H1N.CM.04.04CM_1131_03       -------.-T--K--K--A-Y---------RM--L-----------C----MD-AK--A-L-N--E-A-N----G-------L-V-----ST-PVLN-Q------K-KKE....................QHKP..EA.-NPE-G.AAAAAD-NI-R---L---A--  142
H1N.CM.06.U14296             -------.-T--K--Q--A-Y---------RM--L-----------C----MD-AN--A-L-X--E-A-K----G-------L-V-----S--QVQN-Q----XXKXK-X....................QHKP..EP.-NPETE-AAAA.DSNI-R---L---A--  142
H1N.CM.06.U14842             -------.-T--KX-Q--G-Y---------RM--L-----------C----MD-AA--A-X-N--E-A-N----G------XL-V-----NKRQ-HN-Q----H-K-VKE....................QHKP..KP.-EPE-G.AAAAVD-XI-R---LI--A--  142
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....................QHKP..EP.-NPE-G-AAA.TDSNI-R---L--TA--  142
H1N.CM.97.YBF106             -------.-T--K--Q--A-Y---------R---L-----------C----MD-AN--A-LIN--E-A-K----G----X--L-V-----SN-PVHN-Q------K-K-E....................QHKS..EP.-KPE-GTAAAA.DSSI-R---L---A--  142
H1P.FR.06.RBF168             -------.-T--R--A--R---K--S----M---------------C--E-M--A---EKL-E--E-A-R---DG-Q--W--LVV-W---R-VD-S--QQ-IT-WK--MQ..........................RRKKAP.....EG-TGT---------A--  133
CPZ.CD.90.ANT                ---G---.-R-EK--T--S-----------MI--L---RS--Q---LSSS--------EKAIH--S--IEIR-P-II--F--ICV-W---KGEK----EQ-VKTVKMKV............MQTQAETGSSQTASRGMLLRLLLLNKQWCQR-LSGEGR----IVDAG-  156
CPZ.CM.01.SIVcpzCAM13        -------.-T--K------V------R---MM--L-----------C----M-SA---LKL-Q--E-A-K----G----F--L-V-W---SN--VE--QK--EQLKKHH....................GEQVR-NENTENNESRQGASA-SGT-G-------A--  145
CPZ.CM.05.SIVcpzEK505        -------.-T--K--Q----Y---------MM--L-----------C----MD-A---A-L-R--E-A-K----G----F--L-V-----NN-KVQN-Q---E-LR-KM.......................KAEQKEPEPE---GAAAAPESSI-R---L---A--  143
CPZ.CM.05.SIVcpzLB7          ------I.-T--K--D---V------R-R-MM--L-----------C----M--T---E-LIT--E-A-K----N-K--F--IVV-W------A---------TLTAVR...............QKQQKEKE...............---AAARDNDSG-R-----T-A--  140
CPZ.CM.05.SIVcpzMB66         -------.-T--K-------Y---------MM--L-----------C--S-M--T---K-L-Q--E-A-K----G----F--IVV-W----G-PV---------QLQ..............EAQQKGKQEVAAATADG................TST--R-F---A-A--  138
CPZ.CM.05.SIVcpzMT145        ------I.-T--R--A--------------MM--L-----------C----M-NA---EKL-K--E-A-K----G----F--L-V-W---K---VE--QQ--Q-LKAAVNP...................................EQGETPSA-A-F-V-R-A--  130
CPZ.CM.98.CAM3               -------.-T--R--A--------------MM--L-------D---C----M--A---E-L-R--E-A-K----G----F--L-V-W---K--SVE--QQ--T-LK-AV.........................TASREQEV-QPQQQQQDSA--R---V---A--  140
CPZ.CM.98.CAM5               -------.-T--R--A------X-------MM--L-----------C----M--A---ENL-K--E-A-R----G----F--L-V-W---K--QVE--QQ--S-LK-TV..........................AAQEQEV-QPQQQQQESSI-R-------A--  139
CPZ.CM.x.SIVcpzDP943         ------I.-T--R--A--------------MM--L-----------C----M--G---E-L-K--E-A-K----G----F--L-V-W---K--PVE--QQ--T-LK-TAA.............................XE-EV-QPQPQQDSN--R-F-V---A--  137
CPZ.CM.x.SIVcpzMB897         -------.-T--K--D----Y---------MM--L-----------C----MD-T---E-LIN----A-T----G----F--L-V-W------RV----------L--V-Q..........................KS-VKVKQE-AAADP-GGE-R-F---T-A--  140
CPZ.GA.88.GAB1               -------.-T--K------V------R-R-MM--L-----------CD---M-SK---TKL-Q--E-A-K----G----F--L-V-W-I-SD-TVE--QK--EQLKRHH...............GEQQSKTE.....SNSGSREGG-SQGASA-AGI-G---L---A--  148
CPZ.GA.88.GAB2               ------I.-T--R--A----------R-R-MM---------------CD---M--K---Y-L-K--E-A-K----G----F--L-V-W---A---VE--QQ--Q-LK-FVQK.........................ETATAQEA--KQSESNTTEI-R---V---A--  142
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
CPZ.TZ.01.TAN2               -------.-R-DK--T--S---K-R-R---LI--L---GS--Q---M----M-NV---WK-IL-----VDI--P-II--F--ICV-----AGEKVH--E--VKIVKVKLS.......INKNNSAAASSGQGQNAGGRETVPSSDNTENTGRTTATPSG-RL---ITDA--  162
CPZ.TZ.02.TAN3_1             -------.-K-DK--T--S---KSR-R---LI--L---GS--Q---M----M-NV---WK-IL-----VDI--P-II--F--ICV-----AGEKVQ--E--VKIVKVKLR.......MQTDNSAATSSGQDQNAGGRETVPSSGNTGN-GRATTTPGG-RV---ITDA--  162
CPZ.US.85.US_Marilyn         -------.-T--R--A--------------MM--L-----------C----M--AD--L-L-K--E-A-K----G----F--L-V-W---S-VTVE--QQ--V-LK-V..............VQLQKQQEEKE............-QQQEASG.--IG-S---VI--A--  142
                              Gag p15 start                                                                                                                           p15  p27
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.DE.x.BEN                 ----N--.-R-KKA-EL--V----------R-------AN--DK-GLAES---SK---QK--RV-D-LVP----N-K--F---CVIW-L-AEEKV---E--K.-LAQRHLV............AETGTAEK............MPNTSRPTAPPSGKRG---VQ-AG.-  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.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
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........................KEAEKLDM---TGP.Q  139
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                 --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.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....VDKNENAASKNENGE.....................TAT-SGR-R----QVVN.Q  143
GOR.CM.04.SIVgorCP684con     --------.-T--R--A--R---K--S----M---VX----------C--E-M--A---EKL----E-A-K---DG----W--LVV-W---R--DVS--QQ--E-FK-AMS..............................KRKSDSRPET-----V-T-A--  131
GOR.CM.07.SIVgor2139_287     --------.-T--R--A--R---K--S----M---V-----------C--E-M--A---EKL----ESA-K---DG----W--LVV-W---R--DVS--QQ--E-FK-VMS............................-RKSDSRP.......ET-----V-A-A--  131
GRV.ET.x.GRI_677             --GGH-A.---RS--TF-------N-----QI--LI--GK-M---GLHEK----K---QK-IEV-T-LEP----G-KA-F-LCCVIW-I-AEQKV---E--VVTVKQHYHL...........VDKNEKAA..............KKKNET-A.PPGGE-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.........EETKGKEKQ..............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.........VGTKEKETQ..............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-PT.............QEER...........................QPK-G---L-REN.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--N..............K.............................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-P..........A.PEAAGKK.......................Q-TGG---LIREN.Q  130
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-LKVIT.............RKEEKQEDE...........................-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.ENAASEKEQRA....................IVTPSGR-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.ESAASEKEEKA....................TATPSGR-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.ENAASEEEKGA....................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                 ----H-AM---TK--KY--V-----R-----LI------AK--D--GLSDS----QD--KK--EVIL-LQP----SIK--FGIASV---I-AG---E--EQ-KQQVKIRCHL.........AGEQGEQK.................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
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..K.EKGAT.................MT-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..K.EKGAT.................AT-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..K.EKGAT.................APPSGQRG----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..Q.EKGAT.................VT-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........ELQ-KKKNEP-VTP-GG.-R-----SVN.N  153
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.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.US.x.SIVsmH635F_L3       ----N--.---KKT-EL-------N-----M---V---AN--D--GLAES---DNK---QK--SV-A-LVP----N-K-----ICVIW-I-AEEKV-H-E--K.Q-VQRHLV............VETGTADK............MP-TSRPTAPPSGRGG---VQ-VG.-  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-P..........AKNEEAQA.............................---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...EDEETVTSSGQKENSNDTA....TS-GREGKMQLPAAMPS-GG.-G---LIR-P-N  162
SYK.KE.x.SYK173              ---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...KDDPPATSGGQSENSSQNM.ASETS-GQ-VV-QEKQKAA.TPPPRG---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.........GEKKETTEEEK..............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...........VEKEK...TAA....APSGGQ--NYNTAATP-GR.HG---V--QN.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...........VDKEK...TAV....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...........VEKEK...SAT....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--..............KKP..............................---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-VKD...........TKQKEKP...........................---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---VRP..............REEKA...........................-V-M-VLRGP.Q  125
```

```
                                                                                              CyPA binding                                                 major homology region
H1B.FR.83.HXB2               QMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVH.AGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTN.....NPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTL  304
H1A1.UG.85.U455              -P----L-------------D-------------------------M---V------------D-----------L----.----P------------------V--------G.....-------D--R----------------V--------------------F---  300
H1C.ET.86.ETH2220            -----P--A-----------------------T-----------------------------D-----------L----.---V------D-----------------A---G.....---V---D-------------------V-----K--------------F---  302
H1D.CD.84.84ZR085            ------L-----------I-----N-------------------------------------------------------L---Q.---V----------------------------S.....---------------------------V------------------------  304
H1F1.BE.93.VI850             -----SL-----------I-----------------------T---------------------D-----------L----.---AP-----------------------Q---G.....---V---D-------------------V-----K--------------F-V-  300
H1G.SE.93.SE6165             -------T------------------------------------L-----------------D-----------M--QQ.---FP---I-------------S-----T---G.....---------------------------V--------------------F-C-  304
H1H.CF.90.056                --------------------------------------------A-----------------D----------------.----P-----------------------A---G.....--A----D-------------------V-----K--------------F---  304
H1J.SE.93.SE9280_7887        -P----L-----------I-------------------------------------------I------------D----------------.--------V-----------------------G.....---------------------------V--------------------F-A-  304
H1K.CM.96.96CM_MP535         ------L-----------I---------------T-----------------------------D---D-------L----.----P-----------------------A---S.....---V-----------------------V--------------------F---  300
H1O.BE.87.ANT70              -----------------A------N--I----M------ISY-I-----AI----G-L-V---V-----V----T--PP.V--LP---I---T----------Q----H-T-...RPNQ.-----D--RK--V-----M-K----V-----K------------------  303
H1O.CM.94.BCF06              ------L----------A------N--I----M------V-Y-I-----AI----G-L-V---V----------T--PA.M--LP---L-D-T----------Q----N-I-...RPN-.-V---D--RK--V-----M-K----V-----K------------------  303
H1O.CM.96.96CMA102           ------L----------A------N--I----M------V-Y-I-----AI----G-L-V---V----------T--IP.V--LP---I---T----------Q----H-T-...RPN-##----D--RK--VX----M-K----V-----K------------------  302
H1O.CM.96.96CMABB009         ---Y--L----------A------N--I----M------ISH-V-----AI----G-L-V---V--D-------I--QP.V--LP---I---T----------Q----H-T-...RAN-.-----D--RK--V-----M-K----V---V------------------I-  300
H1O.CM.97.97CMABB447         --I--PL----------A------N--I----M------I-Y-I-----AI----G-L-V---V------D---T--PV.V--LP---I---T----------Q---VH-I-...RANA.-----D--RK--V-----M-K----V-----K------------------  303
H1O.CM.98.98CMA104           ------L----------A------N--I----M---X---SY-X-----AI----G-L-V---V----------T--IA.V--LP---I---T-------------VH-T-...RAN-.-----D--RK--V-----M-K----V-----K------------------  303
H1O.CM.98.98CMABB197         HQ---XL----------A------N--I----M------I-Y-IX----AI----G-L-I---V--D---D---T--PP.I--LP---I---T---------NQA--H-T-...RPNQ.S----D--RK--V-----M-K----V-----K------------------  303
H1O.CM.98.98CMU2901          --I---L----------A------N--I----M------I-Y-I-----AI----G-L-V---V--D---D---S--PV.V--LP---I---T---------SQ--VH-I-...RANH.-----D--RK--V-----M-K----V-----K------------------  303
H1O.CM.99.99CMU4122          -----P-----------A------N--I----M------I-Y-I-----AI----G-L-V---V-----S----T--PA.M--LP-------T----------Q---VH-T-...RAN-.-----D--RK--V-----M-K----V-----K------------------  303
H1O.FR.92.VAU                ------L----------A--K---N--I----M--------Y-V-----AI----G-L-V-E-V--D---D--KT--PV.V--LP---I---T-------------H-T-...RANQ.-----D--RK--VI----M-K----V-----K-R-E--------Q-----  302
H1O.US.99.99USTWLA           -A---PL----------A------N--I----M------ISY-I-V---AI----G-L-V---V------D---T--PP.I--LP---I---T----------Q----H-TS.R.PGN..----XD--RK--V-----M-K----V------X-----------------  300
H1O.US.x.I_2478B             --I--P-----------A------N--I----M------I-Y-I-----AI----G-L-V---V-----S----T--PP.I--LP---I-D-T----------Q---VH-V-....RNPN-----D--RK--V-----M-K----V------------------------  303
H1N.CM.02.DJO0131            --I--PLT----------I-----N-------M--------A-------------------V---------T--XP.X--LP---L---------------X--VT---S.....---V------K---V----R-------V---E-K------------------  308
H1N.CM.02.SJGddd             --I--PLT----------I-----X-------M--------S-------------------V---------T---P.V--LP---I---------------A--VA---A.....---V---D--R---V----R-------V---E-K------------------  307
H1N.CM.04.04CM_1131_03       --I--PLT----------I-----N-------M--------S--S----------------V---------T---P.V--LP---L---------------P---T---S.....-------D--R---V----R-------V---E-E----------------V-  306
H1N.CM.06.U14296             H----PLT----------I-----N-------M--------G-X-----------------V---------T---P.V--LP---L---------------A--VA---X.....---V---D--X---V----R--K--X-V---E-K---------------R--  306
H1N.CM.06.U14842             -L---PLT----------I--------I----M--------S------------------DV---------T---P.V---P---I-------------NVA--VA---A.....---V-X-D--R---V---XR-------V---X-K------------------  306
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
H1N.CM.97.YBF106             -----PLT----------I-----N--I----M--------S--S----------------V---------T--AP.V--LP-----D-------------A--VA---S.....-------D--R---V----R-------V---E-K------------------  306
H1P.FR.06.RBF168             --T-LPL----------A------N--I----M-----SI-D-I-----A----G-L-V---V---------T--AP.V--LP---L---T-G-----XX-K----T-I-R.....-N-V---D--RK--V-X---V-K----X-----K------------------  297
CPZ.CD.90.ANT                IAR--PLT---------C----N-N----------------H------A--D--G---V---V---------L--T-.---VQA--L---T---------V---MQ--ST..PQQ-GGV---D-------M----V--X---V---E-K-----------------I  323
CPZ.CM.01.SIVcpzCAM13        ----TP------------------N-----------------L-L-V-----A-----G---V-R-V---------M---.--------L------------------A-A.....-------D--R--V-----RV-K--C-V-----K-------K--------M  309
CPZ.CM.05.SIVcpzEK505        -----PL-----------------N-------M-------------------------V---------G----.M---P---V---------------A--VA---A.....---V---D--R---V-------------A-----K-----T-----------  307
CPZ.CM.05.SIVcpzLB7          -P----L-------------------------A--A-----------------------I---V----Q---L--I-.M------------------V-------S.....-----------------M----V------V-----K------------------  304
CPZ.CM.05.SIVcpzMB66         -----P--------------------------M------------------------V---------L----.---VP-----------------I---V---S.....-------D--------------K--C-V-----K-----S-----------  302
CPZ.CM.05.SIVcpzMT145        -LT-NP-----------A------N-------M-----------------I--------I---V--D-------T--T-.-----------------------S-E--V---A.....-------D--R--V-------K---V-----K------------------  294
CPZ.CM.98.CAM3               -L---PM-----------I---N-N-------M----------V-----AI----G---V---V---------T----.---V----L--A---------------V---A---A.....---V------R--VV-----V----C-V-----K------------------  304
CPZ.CM.98.CAM5               --I---M-----------I---N-N-------M----------V-----A-----G---V---V---------L----.---------L--A----------------VA---A.....------------R--VV-----V-K--C-V-----K------------------  303
CPZ.CM.x.SIVcpzDP943         -LM--PM---------------N-N-------M----------V-----AI----G---V---V----L----T----.---------L--A--------------A---A---A.....---V------R--VV-----V-------V-----K------------------  301
CPZ.CM.x.SIVcpzMB897         -----SL-------------------------M-------------------I--------I---V---------T--I-.----P-----------------I---V----S.....---T---D------V---------------V-----K------------------  304
CPZ.GA.88.GAB1               ---------------------------------------L---V-----A-----G---V---V---------L--T-.---------L-------------------T-A.....-------DV-R--V------V---C-V-----------------------  312
CPZ.GA.88.GAB2               --I---M-----S----A--------------M----------V-----AI----G---V---V---------L--L-.---V----------------------V---S.....---V------R--VV-----V---C-V-----K----------------V-  306
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
CPZ.TZ.01.TAN2               VAR--PL----------T--D-G-K------------------Y---S---AI-E------I---V---------A-XA-.---MQA-ML---S-A--------V---VM---T..PQAQGGV---D-------------L------V-----K-----S-----------I  329
CPZ.TZ.02.TAN3_1             IAR--PL----------T----G-K------------------Y---S---AI-D-----------V--D-------Q--A-.---MPA-ML---T-A-------SV----L---T..PQGQGGV---D-------------L------V-----K-----------------I  329
CPZ.US.85.US_Marilyn         ------M----------A------N-------M----------V-----AI----G---V---A------D*--T--L-.-------------------------V----A.....-------D--R--VV-----V-K--C-VG----K------------------  305

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.DE.x.BEN                 NY--VPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--D------IIR-I------D--SQ--IP..--LPA--L-D-----------VD---Q--YR....PQN-V---N--R---QI--Q-C--K-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-----NX-R---X---Q-C---XN--N---VK-------QS-------SX  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
COL.CM.x.CGU1                GP---PL-----G----C--.GGIA-SLA---L-Y-T--IAY-M-L---ILDT--GFP-V--DE--KK-E-Y-LL---..QQ-QQQ-AL-Q-TA---T-N--SVA--VA-GE............AN--RG--VQS-E-VIQIAR-S-V------S--D-KS------SA-  296
DEB.CM.04.SIVdeb04CMPF3061   -Y--MPL---IVKT-INA----K----IV-L-QV-A--C--Y-I-GL--AI-DL-G---II-DV----------LQ--QQ.QP..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.QP..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.QP..PQ--L---S-A-----N--VE---A---RPAGQGAG--D--Q--R--V----QRC-K--N--N---VK-------K----------  328
DEN.CD.x.CD1                 -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.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
GOR.CM.04.SIVgorCP684con     -----PL----------A------N--I---M-----SI-Y-I-----AI----G-L-V---V----------Q--AV.V--LPV--L---S-G--------Q---VA---R....N-N-V---D--RK--V-----V-K--C-V-----K------------------  296
GOR.CM.07.SIVgor2139_287     -----PL----------A------N--I---M-----SI-Y-I-----AI----G-L-V---V----------M--AV.V--LPV--L---S-G--------Q---VA-I-R....N-N-V---D--RK--V-----V-K--C-V-----K------------------  296
GRV.ET.x.GRI_677             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-A....D.T-VD--K---E-V-QAME-V--IHQ-V-VI--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-A....D.T--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-A....D.T-VD--K---E-V-QAME-V--IHQ-V-VM--K-------KE-A---F-A-  294
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..A-.QQ--L---S-----------VD---Q--YR....QQN-----N--R---Q---Q-C-K--N--N-----------QS-------S-  305
MON.CM.99.L1                 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
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.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.x.H9                  NYT-LPL----------L----K-GA--V-G-Q-----C--X-I-Q---C--E------IIR-I------D--LQ-XQP..---PA--L--X-----------XD---Q--XR....QQN-----N--X---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.US.x.SIVsmH635F_L3       NY--LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--QP..--LPA--L-------------VE---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              -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.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....N.--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....N.---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
```


PLV Proteins

```
                                                                                p24 end p2 start     p2 end p7 start                                      Zn motif                           Zn motif
H1B.FR.83.HXB2              RAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSAT.....IMMQRGNFRNQR...................................KIVKCFNCGKEGHTARNCRAPRKK.........GCWKCGKEGHQMKD 425
H1A1.UG.85.U455             -----T-D-------------------S--R----G------------------------------QQTS...............GP....................................RRI----------L-K--------.........-------------- 419
H1C.ET.86.ETH2220           -----T-D------D---------------R----G-S--------------A-------------N-TT......----KS--KGPK...................................RAI----------L----------.........-------------- 422
H1D.CD.84.84ZR085           -----------------------------------G----------------S------------A-SASAA....----KS--KGT-...................................-------------I----------.........-------------- 426
H1F1.BE.93.VI850            -------D--G---D-------------------TG----------------S------------ANSA.......----KS--KG--...................................RV-----------I----------.........------R------- 419
H1G.SE.93.SE6165            -------D--G---D---------------R---QG-S--------------S------------ASGA-AA....-----S--KGP-...................................RTI----------L----------.........-------------- 426
H1H.CF.90.056               -----T-D----------------------R---QG-SI-------------S---------------TN-A........K---KG--...................................-F-----------I----------.........------R------- 426
H1J.SE.93.SE9280_7887       -----T-D------D-------------------SG--------------------------------TN...............DHK...................................R---------Q--I-K--------.........-------------- 424
H1K.CM.96.96CM_MP535        -----T--------D--------------------G-S--------------S----I----------PV......V---K---KGH-...................................-------------I----------.........-------------- 420
H1O.BE.87.ANT70             -----T---------------------Q---S---G-------V--------T----------ATAQQDLKGGYTAVF----QNPIRK...................................GTI----------I----------.........------Q------- 429
H1O.CM.94.BCF06             -----T---------------S-----Q-------G-------V--------T---KI-----VTAQQELKGGYTAVF----QNP-RK...................................GPI-------D--I-K------RR.........------Q------- 429
H1O.CM.96.96CMA102          -------------------------X-Q---S---G-------V--------T----------XSAQQDIKGGYTAVF----QNP-R-...................................GPI----------I-K------R-.........------Q------X 428
H1O.CM.96.96CMABB009        -----T---------------------Q-------G-------V--------T---KI-----ASAHQDIKGGYTAVF----QNPIRK...................................RPI----------V----------.........------Q------- 426
H1O.CM.97.97CMABB447        -----T---------------------Q---S---G-------V--------T---KI-----AS-QQDLKGGYTAVF----QNPTRG...................................RPI-------D--L-K--------.........------Q------- 429
H1O.CM.98.98CMA104          ---------------------------Q-------G-------V--------T----------ASAQQDLKGGYTAVF----QNPIR-...................................GP-----------I----------.........---N--QD------ 429
H1O.CM.98.98CMABB197        ---------------------------Q---S---G-S-----V--------S----------ATAQQDLKGGYTSVF----QNP-R-...................................GAI----------I-K------R-.........------Q------- 429
H1O.CM.98.98CMU2901         -------D-------------------Q-------G-------V--------T---KL-----VTAHQDLKGGYTSVF----QNPMRK...................................GPI----------I-K--------.........------Q-----R- 429
H1O.CM.99.99CMU4122         ---------------------------Q---S---G-------V--------T----------ATAQQDLKGGYTAVF----QNPGRK...................................GPI----------L----------.........------Q------- 429
H1O.FR.92.VAU               -----T---------------------Q-------G-S-----V--------T----------ASAQQDLKGGYTSVF----QKPIRK...................................GPI----------VTK--------.........------Q------- 428
H1O.US.99.99USTWLA          ---------------------S-----Q-------G-------V--X-----T---KI-----ASAQQDLKGGXTAVF----QNLGRK...................................GP-----------V----------.........------Q------- 426
H1O.US.x.I_2478B            -----T---------------------Q---S---G-------V--------T---KI-----AAAHQDLKGGYTSVF----QNPIRK...................................GPI----------I-K-------R.........------Q------- 429
H1N.CM.02.DJO0131           -------D-------------------Q------------------------A-------------Q.T--S....VFA-GR--KGI-...................................RTIR---------L----K---RG.........------Q------- 429
H1N.CM.02.SJGddd            -----T---------------------Q---S----------------------------------Q.AT-S....VFV-----KGI-...................................-TI----------L----K---R-.........------Q------- 428
H1N.CM.04.04CM_1131_03      -----T-D-------------------Q-------GSS------------------------TA-AQ.A--S....VFV-----KGT-...................................RTI----------L----K---RG.........------Q------- 427
H1N.CM.06.U14296            -------X------D------------Q-------G----------------A-------------Q.AT-S....VFA-----KGI-...................................RPI----------L----K---RG.........------Q------- 427
H1N.CM.06.U14842            -X---X--------D------------QL------G--------------X-A-------------QG.--S....VFV-----KGM-...................................-TI----------V----K---RG.........------Q------- 427
H1N.CM.95.YBF30             -----T---------------------QL------G----------------A-------------QQPT-S....VFA-----KGI-...................................-PI----------L----K---RG.........------Q------- 428
H1N.CM.97.YBF106            -----T-D-------------------Q-------G----------------A----------A-AQ.T--S....VFV-----KGI-...................................-TI----------L----K---RR.........------Q------- 427
H1P.FR.06.RBF168            ----------S---D---I--------Q-------G-------N--------T----------AAANQANQD.......LK-GYTTVFMQSGQ.............................R-P--------I--I-K--K---R-.........------Q------- 422
CPZ.CD.90.ANT               -------P-A--------I--------H------TG-S----L---------A----------ASAN-AQGT...AVFL----GNRGGK..................................RPL---------------K---R-.........---R--Q----L-N 447
CPZ.CM.01.SIVcpzCAM13       ----------T---D------------Q-------G-S--------------A----------TM-QSQ-RT...D-FF-K-PGATP-...................................RKI--Y-------L----K---R-.........------Q------- 432
CPZ.CM.05.SIVcpzEK505       -----T---------------------N--R----G-S--------------A----------T-AQTATS.....VF------KGI-...................................-TI----------L----K-----.........------Q------- 428
CPZ.CM.05.SIVcpzLB7         -----T--------D------------S--R----G-S--------------A------------ANAA.....TV--------KGPK...................................RVI----------I----K---R-.........------Q---R-Q- 425
CPZ.CM.05.SIVcpzMB66        -----T--------D------------S--R----G----GDEP-F------S------------AQH-NDA....KRQFK-PK.......................................R------------I----K---R-.........------Q-----RN 420
CPZ.CM.05.SIVcpzMT145       -------D------D---I--------N--R----G----------------S-------------NKT......E-F---GQNRPPN...................................-KI----------L--------R-.........------Q------- 415
CPZ.CM.98.CAM3              -----T--------D------------N--R----G----------------A------------LQ-PT.....GVFL----NGKPT...................................RKI----------L--------R-.........------Q------N 425
CPZ.CM.98.CAM5              -----T--------D------------S--R----G----------------A-------------Q--S.....NVF-----NGRVN...................................-KI--Y-------V--------R-.........------Q---N--E 424
CPZ.CM.x.SIVcpzDP943        -----T----T---D--X---------N--R----G----------------A-------------Q--TS.....VFL----NGRPT...................................RKI----------L--------R-.........------Q----I-- 422
CPZ.CM.x.SIVcpzMB897        -----T-D-------------------Q--R----G----------------S------------ANTTVM.......-----VKI-K...................................R------------I--------R-.........------Q----Q-N 423
CPZ.GA.88.GAB1              --------------D------------Q-------G----------------S-----------M-Q-QGRA...DVFF-K-QGAGPK...................................RKI----------L----K---R-.........---R--Q------- 435
CPZ.GA.88.GAB2              -------D-------------------Q--R----G----------------A------------IQQTS.....S-F---AGG-TPP...................................RR-----------I----K---R-.........-------------- 427
CPZ.TZ.00.TAN1              -------P--T----------------H------QG------L---------S---KI-----ASA-AGG.....VN-L-G-KRPPLK..................................KGQLQ------V-------------.........---R--Q------- 452
CPZ.TZ.01.TAN2              -------P-------------------H------TG-S----L---------S---KI-----AAANTG..GGIN...-LQ-GR-QPVKK.................................GQLQ------I-------------.........---R--Q------- 452
CPZ.TZ.02.TAN3_1            -------P--T----------------H------TG-S----L---------AY--KI-----AAANTT..GGIN...-LQ-GR-PPGKK.................................GQLQ------I-------------.........---R--Q------- 452
CPZ.US.85.US_Marilyn        ------P---T-------------------R----G------V---------A----------C-MK-PS.....SVFL-K--AGKPG...................................RKI----------L--------R-.........------Q------- 426
                                                                                   p27 p2               p2 p8
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 426
H2A.DE.x.BEN                ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEAMGPSPIP..FAAA-Q.........................................R-AIRYW-------S--Q-----RQ.........-------P--I-AN 424
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 422
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 429
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 427
COL.CM.x.CGU1               ---P-AG-I-A--ANN--I-H------R---G-QK.PS--D-LA--------D---K------Q-FQQER.....TN-IEVKTAKCFNCQGIGHLARMCPKRPIGGAGRGRGRGRGGFRGAPRRP-R--T-NQ---MQ-D-P............................ 432
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 461
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 456
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 453
DEN.CD.x.CD1                -----D-S--L---TS-----------Q--E---NP------LQ---------Q---PT---LASAFQKQGLCHGSRKGAQPPVER....................................RGLR------QI--L-KD--K-KRVT.....QPGA-F----L---A-N 460
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 434
GOR.CM.04.SIVgorCP684con    -------D----------I--------Q-------G-------N--------T----------AAAHSDLKGGYTAVF--K-GNNQGIKK.................................GP--------V--I-K------R-.........------QD------ 424
GOR.CM.07.SIVgor2139_287    -------D----------I--------Q-------G-------N--------T----------AAAHSDLKGGYTAVF--K-GNNQGIKK.................................GP--------V--I-K------R-.........------QD------ 424
GRV.ET.x.GRI_677            -----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-- 425
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 433
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 432
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 426
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 424
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 429
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 421
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 432
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 432
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 432
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 426
MON.CM.99.L1                ----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-- 430
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 428
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 432
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 435
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 436
OLC.CI.97.97CI12            ----G-----T----K--I--------GV----.VKPS----L----------Y--------LRP-QMVQQRPAP.........................................GGGPQQGRLR-Y---RF---KN--SQKGP...........T-Y----P--IA-N 403
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-- 433
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--- 433
RCM.GA.x.SIVRCMGAB1         -----DPA------QS--I--------MV--G--MNPS----L---------Q----------QMMQSN.......--A-QSAN.............................RGPPRRSGGNPNLR-Y----P--IS-Y-K---R-.........------SPD-LL-- 431
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--- 432
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----- 451
SMM.SL.92.SL92B             ----TDPA------Q---I--------LV--G--MNP-----L----------Q---LM----KDALTGSLV.....AA-FRGAAKGQ................................GNKP-IR------T--S--Q-----R-.........------E--RIQAN 430
SMM.US.x.H9                 ----TDPAX-----Q---I--------LV--G-XINP-----L------X---Q-X-LM---LKDALTQGPLP..FXAV-QKGQ......................................X--IX-X--X----S--QF----RX.........-------A--V-AK 427
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 427
SMM.US.x.SIVsmH635F_L3      ----TDPA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LKEALRPDQLP..FAAV-QKGQ......................................RRTI--W-------S--Q-----RQ.........---G---T--V-AK 427
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 427
SUN.GA.98.L14               ---SG-E---E--K-KM---------LVI-----ETPS-----R--------S--GKI-----ASAFRQQVG..RQA-V-QNLPPRNSQ......................GRFVRIGGGGPR-PMT------P--L-NQ--E-K-G......PPGS------M--KQAQ 434
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 448
SYK.KE.x.SYK173             -----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 447
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 448
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 451
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-- 435
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-- 433
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-- 432
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-PVPMRRGQGPT--G--EV--VQRN 414
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-PVPLRRGQGPT--G--EI--VQRN 418
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 414
```

```
                                    p7 end p1 start        p1 end p6 start       Vpr binding                                                                                          Vpr binding
                                                                                                                                                                                       p6 end, Gag end
H1B.FR.83.HXB2               CT.........ERQANFLGKIWP.......SYKGRPGNFL.........QSRPEPTAPPEE..........................S....................FRSGVETTTPPQKQE..PIDKE....LYPLTSLRSLFGNDPSSQ..*   500
H1A1.UG.85.U455              --.........---------------........-N-------P.........----------A-..........................I....................-GM-EKM-S-A.----..LK-R-....QT--V--K-------L--..-   494
H1C.ET.86.ETH2220            --.........--------RL--........-N--------.........-----------------S.......LRPEPTAPPPES..........................FRF-EA--SP---..LK-R-....A.-----K------HLL-..-   505
H1D.CD.84.84ZR085            --.........---------------........-H--------.........----------A-..........................A....................#GFG-EI--S----..QK--DK..E----A--K-------L--..-   502
H1F1.BE.93.VI850             --.........---------------........-N--------.........----------A-..........................-....................FGFR-EI--SP---..QK-G-L...YP--A--K-------*--..-   494
H1G.SE.93.SE6165             --.........---------------........-N--------........-N-T------A-..........................-....................LGFG-EIA-SP---..MKE--......LYP--K----S--*--..-   498
H1H.CF.90.056                --.........---------------........-S--------.........----------A-..........................-....................FGFG-EM--SP---.QLK---.....P--A-------S--LL-..-   501
H1J.SE.93.SE9280_7887        --.........---------------........-S--------.........----------A-..........................-....................LGLG-EI.-SP---..-K---.......-------K----S--L--..-   498
H1K.CM.96.96CM_MP535         --.........---------------........-H--------.........----------A-..........................-....................FGFG-EI--SPR--..TK---Q...SP-----K-------L--..-   496
H1O.BE.87.ANT70              -RN........GK-------Y--........PGGT----YV.......QRPA..H-S---M-..........................E..................EVKG......QENQE..Q.KG.GPNELYPF.A--K----T-Q.....-   499
H1O.CM.94.BCF06              -KN........G--------Y--........PGGT----YV.......QK-V..S-S---M-..........................E..................EVKE......LEKQS..Q.GG.-KEELYPF.A--K----T-Q.....-   499
H1O.CM.96.96CMA102           -RN........GK-------Y--........PGGT----YV.......QK-V..P-S---M-..........................E..................VMQGQK....NQNQE..X.RE.DPNELYPF.A--K----T-Q.....-   500
H1O.CM.96.96CMABB009         -KH...........------RY--........PGGT----Y-.......QK-V..S-S---M-..........................E..................VVKE......QEDQN..Q.GK.DQNELYPF.A--K----T-Q.....-   494
H1O.CM.97.97CMABB447         -RN........GK------RY--........PGGT----YV.......QR-V..P-S---M-..........................E..................VVKGQE....NQNQN..Q.GG.DQNELYPF.A--K----T-Q.....-   501
H1O.CM.98.98CMA104           -RN........GK-------Y--........PGGT-----M.......QK-V..A-S---M-..........................E..................VVKG......QENQE..Q.KE.DSNELYPF.A--K----T-Q.....-   499
H1O.CM.98.98CMABB197         -KN........GK-------Y--........PGGA----YV.......QK-V..P-AT-SAP..........................P...............MEEEMKG......QENRE..Q....NPNELYPF.A--K----T-Q.....-   500
H1O.CM.98.98CMU2901          -RN........GK-------Y--........PGGT----YA.......QR-X..A-S---M-..........................D..................EMKGQE....KENQE..Q.KG.DPNELYPF.A--K----T-Q.....-   501
H1O.CM.99.99CMU4122          -RN........GK-------Y--........PGGT----YA.......QR-V..S-S---M-..........................E..................ERKG......-ESQE..Q.KG.DPNELYPF.A-------T-Q.....-   499
H1O.FR.92.VAU                -KN........R--T---K-Y--........PRGT----YV.......QE-V..S-S---M-..........................E..................MMKG......QEED-N.Q.KG.-QNELYPF.A-------T-Q.....-   499
H1O.US.99.99USTWLA           -RN........G--------Y--........PGGE-----M.......QK-M..P-P--SAP..........................P...............MEEVVXG......XXNQN..Q.GE.DSNELYPF.A--K---XT-Q.....-   499
H1O.US.x.I_2478B             -RN........GK-------Y--........PGGT----YM.......QK-V..P-S---M-..........................E..................VMKG......QENQG..Q.KGDPTELYPFASLKSLFG..T-Q.....-   499
H1N.CM.02.DJ00131            -KN.......EG--------G-S.......PF-------PQ......TTTT-K------L-..........................N....................YGSQ-ERSTQG-E...RQENTENSLYP-----------L---..-   512
H1N.CM.02.SJGddd             -NN.......EG-------RG-S.......PF-------P.......QTTA-K------L-..........................-.........................-ERSTQG-EMQENQE-TENSLYP-----------L---..-   508
H1N.CM.04.04CM_1131_03       -KN.......EG--------G-S.......PF-------P.......QTTT-K------L-..........................-....................YGSQ-ERSTQR-EMQENQE-TENTLYP-----------L---..-   512
H1N.CM.06.U14296             -KN.......EG-------RG-S.......PF-------P.......QTTTKK------L-..........................N....................YGFQ-ERSTQG-EMQENQERTENSLYP------------L--..-   512
H1N.CM.06.U14842             -KN.......EG--------G-S.......PF------XPQ......TTTT-K------L-..........................D....................YGCQ-ERNTQR-EMQEDQE-TENSLYP----------------..-   513
H1N.CM.95.YBF30              -KN.......EG--------S-S.......PF-------P.......QTTT-K------L-..........................-....................YGFQ-EKSTQG-EMQENQERTENSLYP----------------..-   513
H1N.CM.97.YBF106             -KN.......EGX-------G-S.......PF-------P.......QTTT-R------L-..........................-....................YGSQ-EKSTQG-EMQENQE-TETSLYP-----------LX--..-   512
H1P.FR.06.RBF168             -KS........G--------------........PGGK----YV.......QK-V..Q-----M-..........................E......DEMTRDKQR...EEKED...............E--...YPLTSLRSLFG..T-Q.....-   492
CPZ.CD.90.ANT                -PAT......NTGKV-----PT-.............TWWGCRPGNFVQKEEVV------I-..........................I.....YQEEHKRTQKGLK..................GEE-..LPPSYSLKSLFG..K-Q.....-   523
CPZ.CM.01.SIVcpzCAM13        -S.........G--V-----G--.........-RS------V.......QNRT.-------A-..........................-.....YGYQEKEKNQ....EEQEE.................GNRLYPPLTSLKSLFG..S-------.-   507
CPZ.CM.05.SIVcpzEK505        -RS.......G---------V--.........LS-------PQ......TTTRK.------I-..........................-.....YGYQEEKTTQGTEREEKEKT...............ESSLYPPLTSLKSLFG..S---L-..-   512
CPZ.CM.05.SIVcpzLB7          -NM........GK.V-----L--.........-SQ------P.......QSRP.-------I-..........................-.....FGFQPKTTQKQEDREEDKT................ENKLYPPLTSLKSLFG..---L--..-   504
CPZ.CM.05.SIVcpzMB66         --N...............-----L--.........-S--------.......QKRP.-------I-..........................-.....FGFREEEPAQKEEVKSEET................ENKLYPQLTSLKSLFG..---L--..-   500
CPZ.CM.05.SIVcpzMT145        -NQNN......G-.I-----N--.........PWG-----YV.......QGRK.-------M-..........................-.....FGFQEEKKVKEEVRE....................ESKPPSPLTSLKSLFG..-------..-   492
CPZ.CM.98.CAM3               -PA.......G---------G-S..............PWSGGSKRPGNFL-G-K------L-..........................-.....YGYQEATGKRQEEEREK...............E--PLYPPLTSLKSLFG..S-----..E-   508
CPZ.CM.98.CAM5               --TG.......G--------D-S..............PWSGRSKRPGNFP-GGK-----M-..........................-.....YGFQETTMEKKKE...................E--PLYPQLTSLKSLFG..S---L..E-   503
CPZ.CM.x.SIVcpzDP943         --A.......GD--------G-S....PWSGGSK-------.......QGRK.-------I-..........................-.....FGYQETTVNKQGEKKEMEVE.................--PLYPQLTSLKSLFG..S-----..E-   507
CPZ.CM.x.SIVcpzMB897         -PD........MQR-------------.........PR--------QKR....PEPP..--S---I-..........................-.....FGFQDQKIDSLKQXVRNEXQT.............ENRLYPPLDSLKSLFG..---L--QE.   509
CPZ.GA.88.GAB1               --.........G--V-----G--..................SRSGRPGNFV-N-T------I-..........................-.....YGYQEEEKSQEKK.................EG-SSLYPPTSLKSLFG..S-----..Q-   509
CPZ.GA.88.GAB2               -QN.......EG--V----GG--..................SRSGRPGNFP-RKV------P-..........................T.....FGYAEQTQTGKSQKGEEST...............TE--..SP---SLKSLFG..S-----..Q-   508
CPZ.TZ.00.TAN1               --TRN.....NSTGV-----RT-LWGCRPGNFVQNTPEKGKAQEQETAQTP.VV-----L-..........................M.....................................TMKGGFSLKSIFG..S-Q.....-   525
CPZ.TZ.01.TAN2               --T....KNNNSPGV-----RT-LWGCRPGNFVQNTPEKGKNQEQEQTQTPA.V-S---MT..........................DL....................................PLKDGFSLKSIFG..S-Q.....-   527
CPZ.TZ.02.TAN3_1             --T....KNNNSPGV-----RT-LWGCRPGNFVQNTPEKGKNQEQEQTQTPAVV-S---MD..........................L.....................................PLKDGFSLKSIFG..S-Q.....-   527
CPZ.US.85.US_Marilyn         --AG.......N--------H-S............-SWSGGSKRPGNFLEN-K------I-..........................D.....FGYQEETVTQEKQG..................KE--..PFQLTSLKSLFG..S-----..E-   506

                                       p8 p1                     p1 p6       PTAP motif                                    PSAP in HIV-2 B, U                                   p6 end
                                                                                                                                                                                   Gag end
MAC.US.x.239                 -P.........D---G---LGPW..............GKKPRNFPMAQVHQGLM-----.-..........................DPAVDLLKNYMQLGKQQREKQRE-REKP...............YKEVTED-LH-N----G-Q.....-   511
H2A.DE.x.BEN                 -P.............----G---LGPR..............GKKPRNFPVTQAPQGLI-----AD...........................PAAELLERYMQQGRKQREQRE-PYK-V-EDLLHL-QRETPHREETEDLLHLNSLFG..K-Q.....-   522
H2B.CI.x.EHO                 -P.............----G---FGPW..............GKKPRNFPVQA.PQGIV-S---MNPAFGMTPQGAIPSAPPAD..........PAEEMLKNYMQLGKKQKENRE-P................YKEVTEDLLHLNSLFG..E-Q.....-   520
H2G.CI.92.ABT96              -P.............----G---FGPW..............GKKPRNFPMTQVPQGLT-S---MD...........................PAVDLLKNYMXLGRKQKEQRNKP.................YKEVTXXLLHL-SLFG..D-Q.....-   510
H2U.FR.96.12034              -P.........-TKVG---LGPW..............EKKPRNFPMAQVPQGLT-----A-....PAVDLLTPTAPPAD..........PAVDLLKSYMQQGKKQKENRE-P................YKEVTEDLLHLNSLFG..--Q.....-   522
COL.CM.x.CGU1                ...........NK------ATRG.........................................................................VELQTAIFPPKMSKDLPPR-EKG....................ES---SLKSLFG..D-Q.....-   483
DEB.CM.04.SIVdeb04CMPF3061   -P.....KAAPQSKN-----LPW..............-RRAPRNFIEQQH-TWGIGKYI-..........................EGGDEIFSSSRVPEQKPQQQEETRRG..................KYPDLSDLKSLFG..E-Q.....-   545
DEB.CM.99.CM40               -G........QTPR-----NTPW............ASRR.PRNFLEGIGKYI--S--DQ.V.........................YSTTRVPVQGQENNPQREETTRRG....................KYPDLMGLKSLFG..E-Q.....-   536
DEB.CM.99.CM5                -G........QVPR-----NAPW............ASRR.PRNFLEGIGKYI--S--DQ.V.........................YSTTRVPVQGERNAPQEDTTTRRG....................KYPDLMGLKSLFG..E-Q.....-   533
DEN.CD.x.CD1                 -R....SNTNDQG-V----NGLG.........RKN-PRNFVQYYPQMNQDLPWS-VRVQQQ.........................PQYPQNFQEGGNRQSCQEDKRQNKQSHQEDEETT...........RES---SLKSLFG..D-Q.....-   559
DRL.x.x.FAO                  -..........PK-V----NTPW...........GS-KPRNFPQA......SLT-S---LP....GYLQEDPAERMLQ..........KYMEQGAQQKRQQRQQQQKR.....................GPYEEA-NSL-SLFG..S-QLQ...-   522
GOR.CM.04.SIVgorCP684con     -RS........G--------Y--..................PGGERPGNYV-KQVQ-----M-..........................E...........EEKETRKP...................GKE-..LYPLTSLKSLFG..T-Q.....-   490
GOR.CM.07.SIVgor2139_287     -RS........G--------Y--..................PGGERPGNYV-KQVQ-----M-..........................E...........GEKEMRKP...................EKE-..LYPLTSLKSLFG..T-Q.....-   490
GRV.ET.x.GRI_677             -K.........NG------YGHW...........GGAKPRNFVQYRGDTV.GL------M-TAYDPAKK.................LLQQYAEKGQRLREEREQTRKQKEKE-.......................EDVSL-SLFG..G-Q.....-   514
GSN.CM.99.CN166              -R.....TGTPIKELY--DLGRG............-RRAPRNFPVT......FLL----A--EEPQ.......................QLNLSTPVGEPQSKGQKEES.....................RNT.---SL-SLFG..E-Q.....-   513
GSN.CM.99.CN71               -KT.....GTPAKEL---DLGRG............RRAPRNFPVT......SLL-S--A--VEPQ.......................QLNLSPPVGEPQNKEQKAEK.....................RNT.---SL-SLFG..E-Q.....-   511
LST.CD.88.SIVlhoest447       -PKP.......KGKV----YGPW.............RSGPPGNFPLMGGNAGVV-S---M-..........................RS..PTQAERALETYRNLGQQLR-QQQIPQK...............CVDEPC-SFFFF..PD-Q.....-   513
LST.CD.88.SIVlhoest524       -PKP.......KGKV----YGPW.............R-GPSGNFPIMGGAAGVI-S---M-..........................KV.PPTRAERAIETYRNLGQQLRKQQQVPQK...............CVDEPC-NFFFF..PD-Q.....-   512
LST.KE.x.lho7                -P.......KPKGKV----YGPW.............SSKPPGNYPLLGGAAGRI-S---M-..........................-A..PTKAERALETYRTLGQQLK-QQQQVPQ...............KCVDE-CLNSLF..PD-Q.....-   516
MND-1.GA.x.MNDGB1            -P.....KPAQQQRV----YGPW.................GPSKPGNYPAQEVT-----L-..........................E..KPLQKTLSTYQKLGRGLRQKMKEEK.................REEDFHSL-TLF..QE-Q.....-   503
MND-2.CM.98.CM16             -P..........KMV----NTPW...........GS-KPRNFPAM......PLT-----MP....GM..EDPAERMLLD........YMKKGQQQK.A.ESKQEKKER...................GPYEAA-NSL-SLFGTDQLQ.....-   518
MND-2.GA.x.M14               -P..........KMV----NAPW...........GS-KPRNFPAM......PLT-S---MP....GM..EDPAEKMLLD........YMRKGQQQRAAKETKQEK.DK...................GPYEAA-NSL-SLFGTDQLQ.....-   519
MND-2.x.x.5440               -...........PKMV----NTPW...........GS-KPRNFPAM......PLT-S---MP....GL..EDPAEKMLLD........YMKKGQQQRAAAGAQKEEKK...................GPYEAA-NSL-SLFG..T-QLQ...-   519
MNE.US.x.MNE027              -P.........D---G---FGPW..............GKKPRNFPMAQMHQGLT-----.-..........................DPAVDLLKNYMQLGKQQRESKRKPY.................KEVAEDLLHLNSLFG..E-Q.....-   507
MON.CM.99.L1                 -PQPP....KQNKGV----NPFG...........PKKGPRNFPLT......SVQ-S--TLPAGTA.......................EPVNLNQEIGSPKTGTPKET.....................RRD.---SLASLFG..E-Q.....-   511
MUS-1.CM.01.CM1239           -R......SEGAKS-----DPLG..........LQRRRPRNFPQT......SLL-----.-..........................QEGDLVNLSPPVGPKDGASDR....................KS.---SL-SLFQ..D--LEK..-   506
MUS-1.CM.01.SIVmus_01CM1085  -R......SDNTKAV----DPFG..........LQKRRPRNFPQT......SLL-----.-..........................QEGELVNLSPAVGPGEGKQDR....................KS.---SL-SLFQ..D-Q-D-..-   510
MUS-2.CM.01.CM1246           -R.....TDTGKSAV-----DLG..........F-SRKPRNFPVT......SLM-----.D..........................TAEPVNFSPLVGPKEEPPKDRKDR.................KT.---SL-SLFG..D-Q-K-..-   517
MUS-2.CM.01.CM2500           -R.....TNSDKGAV-----GF#..........#FHRKPRNFPVT......SLM-S---.D..........................TAEIVNFSPEVGPRKENQTPQAPPDR...............KN.---SL-SLFG..D---E-..-   518
OLC.CI.97.97CI12             -R...VGGQQQQSKV-----WRG.............-KAEPPLNMQQRTEMDIF-S--TL-...............................QILEKGTREKREMKE.....................EKKP---SLASLF..-E-Q.....-   481
RCM.CM.00.SIVagi_00CM312     -P..........K-V----NAPW.............KQGKPRNFPLTS....VT-S--GMD..........................Q......YDPAEAMLKRYLE.QGKKQKQQKRQEE-SQQ....GRAYEEALSSLNSLFG..--Q.....-   518
RCM.CM.02.SIVrcm_02CM8081    --..........K-V----SAPW.............SRGKPRNFPLTS....LT-S--GV-..........................QDNHASQYDPAEAMLRKYLE.QG-MQKQQQRKEE-AQK....GKAYEEALASLNSLFG..S-Q.....-   524
RCM.GA.x.SIVRCMGAB1          --..........K-I----RLPW.............-QGKPRNFPLTS....LT-S--GM-..........................-.....NYDPAEEMLKNYLRRAGEQ.KRQQRQEESKKR....EGAYQEALTSLNSLFG..S-Q.....L   517
RCM.NG.x.NG411               -P..........KEV----NAPW.............KQGKPRNLPLTS....LR----GR-..........................E.GWNPQYDPAEEMLRKYLA.LG-QHKQEQRKEN-EKV....GRAYEDALSSLNSLFG..S-L.....-   522
SAB.SN.x.SAB1                -Q..........K-V----FGPW...........GR-KPRNFPL..T....SIR-----M-RDYSRPEENWYADRPPTRGPGPDDP.ATALLKQYAVQGKRQKQQWQNHSPQQS.................PYEEA-SSLRSLFG..E-Q.....-   555
SMM.SL.92.SL92B              -P.........NQK-G---LGPW..............GKKPRNFPMQTT..SLT-S---.D...........................PAARIVKEYLEKAQREKTRRS-P................YKEVTEDLLHLNSLFG..E-Q.....-   508
SMM.US.x.H9                  -P.............----G---LGPW..............GKKPRNFPMAQMPQGLT---X-.-..........................DPAVDLLKNYMKVGRRQRENRE-P................YKEVTEDLLHLNSLFG..E-Q.....-   508
SMM.US.x.PGM53               -P.............----G---LGPW..............GKKPRNFPMAQMPQGLT-----.-..........................DPAVDLLRNYMKMGRKQRENRE-P................YKEVTEDLLHLNSLFG..G-Q.....-   508
SMM.US.x.SIVsmH635F_L3       -P.............----G---FGPW..............GKKPRNFPMAQMPQGLT-----.-..........................DPAVDLLKNYMKMGRKQRENRE-P................YKEVTEDLLHLNSLFG..E-Q.....-   508
STM.US.89.STM_37_16          -P...........---VG---FGPW..............GKKPHNFPMAQIPQGLT-----.-..........................DPAADLLRSYMQLGKKQRESRKTPY.................KEVTEDLLHLNSLFG..E-Q.....-   508
SUN.GA.98.L14                -P........QGK-V----YGPW.......N.............RGPPGNFPVM-S---L-.....DLTLGNRMTPPQS.......K...AERALE.TYRLLGQGLRAQQKRK-R..................GECQE-CLNMLF..PE--.....-   522
SYK.KE.x.KE51                -RQPRKGQGNPPP-------G-G...........NRKPPANFPVM.......-ET-----A-DWPWQQQTWGNYASPQQKHIAMATPLQPQSSPKTVKDLLVPGKPSEKKEEMKE..................KGP---SLQSLFG..D-Q.....-   562
SYK.KE.x.SYK173              -RQPKRNQGPPVA-------G-G..........VSRRPPANFPV........-S--S---L-DIEDGPWLTWSAQMSQQAQAKAQNSP-KKPPTNREVLSPKESSGKEETKSLYP..........................SL-SLFG..E-Q.....-   555
TAL.CM.00.266                -S.........SA-------SFL.......GKIPPVRGPRNFPV........VQ-S---I-..........................PLTASWQEGTTTWTPPAKKPQETTTTR...................ES.---SLASLFG..D-Q.....-   528
TAL.CM.01.8023               -R.........QP-------AFL.......GKIPPVKGPRNFPV........VQ-----I-..........................PLTASWQEGTTTWTPPTKDAAAPKQQTR..................ES.---SLASLFG..D-Q.....-   532
TAN.UG.x.SIVagmTAN1          -R.........SG------RMPT............WGTKPRNFLEQG.....GAV-----MPAHGFPTGSPA................AGAYDPARKLLEQYAKKGDQLRKQKEK...................ELEDYSL-SLFG..G-Q.....-   524
VER.KE.x.AGM155              -R..........G-V----YGRW............MGTKPRNFPAATL....GA--S---PPNNSTPYDPA.................KKLLQQYAEKGKQMRNQNRNPPANNPD...................WNEGYSLNSLFG..E-Q.....-   521
VER.KE.x.TYO1_patent         -R..........G-V----YGRW............MGAKPRNFPAATL....GA--S---PPSGTTPYDPA.................KKLLQQYAEKGKQLREQKRNPPAMNPD...................WTEGYSLNSLFG..E-Q.....-   520
WRC.CI.97.97CI14             -P.......KIR--V----NGRG.......QMPSNNYPVQKLRSLNNFGPTTTM-S---MR.........EDPNIQELEKY.......LLKGAQLREQQEHKKTEGQK.................R-VGVPEW-SLKSLF..EP-Q.....-   512
WRC.CI.98.98CI04             -P.......KSR--I----NGRG.......QMPSNNYPVQKVRSLNNFGPAT.I-S---MR.........EDPNIQELEKY.......LLKGAQLKEQAQAKKEEKKR...................-TGLPEW-SLKSLF..ET-Q.....-   514
WRC.GM.05.Pbt_05GM_X02       -P.......ETR--V----NGRG..QRPSNN-PVQKM-AGVNTFGPV...VPTA-VD-TMQ..........................ELEQYLLRGAQMKREKEKEM...................QKEGKVE--SLKSLF..EP-Q.....-   503
```

```
                               Gag-Pol TF start                                                                       Gag-Pol TF end protease start
H1B.FR.83.HXB2                 FFREDLAF..LQGKA...REFSSEQT...RANSPTRRELQVWGRDN...............................................NSPSEAG.ADRQG..TVSFNFPQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSL....PGRWKPKMIGGIGGF  109
H1A1.UG.85.U455                ----N---..Q--E-...--------...------S-N-WDG-K-................................................DL-C-T-.-E---..-D--S---I---------V------I--------------DIN-....--K----I-------  108
H1B.US.90.WEAU160              ------V-..PK---...--------...-T----------Q----...............................................--L----.-N---..A-------I------------E-----------------D-N-....--K------------  109
H1C.ET.86.ETH2220              ----T---..Q----...----P---...-------ESQTRANSPTTR......................................ELQVRGS-TF----.-E---..---L----I--------------------------------IN-....--K------------  116
H1D.CD.84.84ZR085              ----N---..P----...G-L-----...------S-SFGFGEEIT..............................................#LPETGTE.GQ---T.--LS----I------V------------------------DIN-....--K------------  109
H1F1.BE.93.VI850               ----N---..Q--E-...-K-P----...------S---R-QRG--...............................................PL-----.-E-R-TV.P-LS---I---------------I---------------DIN-....--K------------  109
H1G.SE.93.SE6165               ----N---..Q--E-...-------D...-T----C-KPR-RRG-S...............................................PLP---.DEGK-....AISL---I---------V------I---------------IN-....---------------  106
H1H.CF.90.056                  ----N---..Q-RE-...-K--P--A...-T----S---R-RRG-D.................................................PL----A-EG--T..-LS---I---------V--E---R---------------IN-....--K------------  108
H1J.SE.93.SE7887               --------..Q-RE-...--L-P---...-------S--PRARRG-.................................................PLP-T-.-EG--T..--S---I----------R-----R--------------DID-....-RK------------  107
H1K.CM.96.MP535                ----N---..P--E-...--------...------S---R-R-G--.................................................PL----.DQ---TE.P-----I------I----V----R---------------IN-....--K------------  109
H101_AE.TH.90.CM240            ----N---..Q----...G-------...------S-K-GDG----...........................................GGRD-LLT---.-E---TS.S--S---I---------V---------------------DIN-....--K------------  114
H102_AG.NG.x.IBNG              ----N---..Q--E-...-K-------..GT--S-S---WDG---................................................T-L-T--.TEG--AI.S------I---------VR-E---I--------------DIN-....--K------------  109
H103_AB.RU.97.KAL153_2         ----N---..Q-RE-...-K-------..--I---S-K-WDG---................................................-PLP-T-.TEG--TA.S------I---R-----VR--------------------DIN-....--K------------  109
H104_cpx.CY.94.CY032           ----NV--..Q-RE-...-K-----A...-----A-GM-REERG--...............................................LL-----.TEG--T..I------I-----------L---IR---------------IN-....--K------------  108
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------------  116
H1P.FR.06.RBF168               ---K---S..WGQE-...-QLCT-TS...TT----Y...................................................GGGRDDPGQA-G-KGG-E-...IP--L--I---D--VIPVRV--HIC------------VDKLP-....E---X----------  105
CPZ.CD.90.ANT                  ----TDPH....VVGVQT--LCA-GGSS.G---S-H-D-..S-GA.................................................QEDSE-SQGGG-T..T-LV--EIP-----MMEVL-Q--KCQ-----------V-GIH-....Q-N----T-------  109
CPZ.CM.-.SIVcpzDP943           ----R-VS..VERGI...K-TRELPAGQEGTH-S-H--FW-S.............................................GNNSEQTGRK-RDGGGERAPV.P-I-L--I-------LGV--E---I------------IDNVQ-....T--------------  116
CPZ.CM.-.SIVcpzMB897           ----N--S..P-RE-...---PP-E-RTT-TKC--N--FW-S-............................................SENRFSETXGQERGX--EQTV.P--G---IS-----V-AAR---VI--V-------------IE-....E-K-R----------  117
CPZ.CM.01.SIVcpzCAM13          ----G---..P-RE-...-QLCT--D...------S---W-PRE.....................................................GEEPR-TGRGEQTI.P-S----I-------IPVR-E---C------------IQDIQ-....Q-I------------  108
CPZ.CM.05.SIVcpzEK505          ----G--S..--RE-...-K-PPDNNKE.-----SN---W-S.............................................GGEDHTGDREGRKGE--ELSV.PTL----I------IL-V----EI-------------I--IQ-....E-K------------  115
CPZ.CM.05.SIVcpzLB7            ----T---..-P---...-KL------...--H--AN---W-SAK.............................................NHPETGGQGRGQ--EQAV.P-L-----IS-----I----V--Y--G-------------DLE-....G--------------  112
CPZ.CM.05.SIVcpzMB66           ----T---..Q----...---P--E-...-T--S-N---R-Q-G..............................................GTCPEGGSEERG--EQAV.S-A-----IS-----V--VR-E-------------------IE-....G--------------  112
CPZ.CM.05.SIVcpzMT145          ---KK--S..-G-E-...--LCAR-E...GT--S-YGI-RIS................................................GGEE-QG-GEGGEQAA.---------I-------IPV--E---I------------I-DLN-....--K-S----------  108
CPZ.CM.98.CAM3                 ----R-VSV..ERGIKET--LPAG-E...GTH-SAP---W-S-SNR...............................................KETGG-EGEGK-TPVS-V-L---I-------LGV--E---I------------IDNVQ-....T-K-R----------  114
CPZ.CM.98.CAM5                 ----G-VSV..EREIKKT--L-AGRE...GTH-S-HG--W-S-NN-...................................................GEKEGGE-TPVS-I-L---I-------LGV-VE---I------------IDNVQ-....T-K-R-R--------  110
CPZ.GA.88.GAB1                 ----R---P..-RE-...-QLCA--N...-T-G--D---W-P-GR................................................EEPGE.ERG-EQS..I-T-L---I-------IPV-VE---C------------I-RIQ-....Q-L------------  107
CPZ.GA.88.GAB2                 ---GG---P..ERE-...--L-PKEG...GT--S-S-D-R-CRA-T.............................................DRKESEGG-EYN-K-E..P-LS---I-------LPVRVA---C-------------DNIQ-....E-K-R--L-------  112
CPZ.TZ.00.TAN1                 ----THPL.....VGVQT--LCA-HP....RERE.....................................................GSGAGDSTDTSGANCPTT-DDDERRVL---N-----MM-V-VQ--VCQ---------S-FCNIK-....K-Q-T--T-------  105
CPZ.TZ.2001.TAN2               ----AHPL..VGVQT...G-LCT-HP...-EREESGAGTDTDASST.IGPTDDG..................................................SAT.ERRVL---IN-----MM-V-VQ--VCQ---------S-FCNIK-....K-Q-T--T-------  106
CPZ.TZ.2002.TAN3_1             ----THPL..VGVQT...G-LCA-HP...-EREESGTGTNTDASSSAIGSTD.G..................................................PTT.ERRVL---IN-----MM-V-VQ--VCQ---------S-FCNIK-....K-Q-T--T-------  106
CPZ.US.85.CPZUS                ----T-VPIV.ERGIKET--LPGK-E...G-H-S-N-GFR-SR--S..................................................DTGETGKGKG.A..L-A---I-------LR--VA--IV-------------DNIQI....E-T-R---M------  111
                               Pol p15 start (-1 from Gag)                                                                   p15 protease
MAC.US.x.239                   ---PWSMG....KE-...PQ-PHGS.......-ASGADANCSP...............................RGPSCGSAKELHAVGQAAERKA-RKQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV------  121
H2A.DE.x.BEN                   ---VGPTG....KE-...SQLPRDP.......--SGADTNSTS..............GRSSSGTVGEIYAAREKAEGAEGETIQRGDGGLAAPR....AER-TSQRGDRGLAA--FS--K--V--AY-ED-PV-V--------SIVAGIE-....GDNYT--IV------  134
H2A.GW.x.ALI                   ---AWPMG....KE-...SQLPRNP.......-SAGINTNSTP..............SRASSGPAGAVYAAGEKAKRAEREAIQRGDGGLTAPR....AGR-TTQRGDRGLAA--FS--K--V--AY-E--PV-V--------SIVAGIE-....GSNYT--IV------  134
H2A.SN.x.ST                    ---VGPMG....KE-...PQ-PCGP.......N-AGADTNSTP..............DRPSRGPTREVHAAREKAERAEREAIQRSDRGLPAAR....ETR-TMQRDDRGLAA--FS--K--V--AHVE--PV-V--------SIVAGVE-....GSNYS--IV------  134
H2B.CI.x.EHO                   ---VRPLG....KE-...SQ-PRPG.......T-.GDSAICAP.............DEPSIRHDTSGCDSICTPCRSSRGDAKELHATREEAEG....EQRETLQGGDRG-AA--FS--R--V-KAT-E--SV-V--------SIVAGIE-....GSNYT--IV------  134
H2B.GH.86.D205                 ---VRTLG....KE-...SQLPHDP.......-ASGSDTICTP.............DEPSRGHDTSGGDTICAPCRSSSGDAEKLHADGETTER....EPRETLQGGDRG-AA--FS--R--V-KAC-E--SV-V-----V--SIVAGIE-....GSNYT--IV------  135
H2G.CI.92.ABT96                ---VWTLG....KE-...SQLPHDP.......--SGSDSISTT...............................DGPSSRPTEELHAXGEEAKG....AEKQTLQGGDRXXAA--LS--R--V--AY-ES-PV-V--------SIVAGIE-....GSNYS--IV------  117
H2U.FR.96.12034                ---AWPLG....K--...SQLPHGP.......--AGTDPNSP.................SSRTSSGSADPDSPSSGPSSGSAKELHATGQETKG....EQRETIQGGDGG-AA--FS--K--I--AY-E--PVKV--N-----SIVAGIK-....EPYYT--TV-E-E--  131
COL.CM.x.CGU1                  ---GNQGGG.....-TDSHLPP....................................................................EDEQR...PATQEGEGGE.-LSLLE-S-RR--I-IVEVE--KCQ-----------FK-EDV..QLT-A-TAVQ-Q----A   91
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---  117
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---  117
DEB.CM.SIVdeb04CMPF3061        ----TSLG....EESP..Q-..........FHRA................................ATPDMGHREVHTGGRGRDLLLFQGTGTET..PAAGGDEEREIPR-IGLE-P--R--VT--E-E-RRVQV---------IIH-KDI..QLTSPFV--TV--L---  118
DEN.CD.x.CD1                   ----WSWE....EESS..Q-LCTVLPSDESGFAMEPS-G...........................PAATTVPTKFSRGGKQTILSGRQEAEQAV...PSGG-GDNEGE-LS-VE-S--R--IGQ-WVQ--RV-V-I-------IIA-QDI.NLGTEG-I--TV-----Y  133
DRL.x.x.FAO                    ----YSLG....QRET.QEFS-GL...................................ADPISSSIARIPPGGSGREDAPKVHGAGSSAEETTE...TTTAEE-PLRGGLQL--FS-RL--TTIVEVE--KV------------IQGLE-....T-K---QI------T  123
GOR.CM.2004.SIVgorCP684con     ----I--S..RGRE-...-QLCA-AS............................................................ATNSPT-GGGRE-NEETRKGG-L-S----I---D--VIPVR---H-C------------V-HLP-....E-----R--------  101
GOR.CM.2007.SIVgor2139_287     ----I--S..RGRE-...-QLCA-AS..AT----N.......................................................GGGRKGDEETRKGGT...L-S----I---D--VIPV----H-C------------V-HL--....E-----R--------  101
GOR.CM.2007.SIVgorCP2135con    ----I--S..RGRE-...-QLCA-AS..AT----N.......................................................GGGRERNEETREGGT...L-S----I---D--IIPV----H-C------------V-HL--....E-----R--------  101
GRV.ET.x.GRI_677               ---VWPLG....RSE.TKKFCAIQRRHS.WSGTNSPP...............................NGNSLRSSKEAPPAVCREGTAPERGE....RTDKETEGERSGGC-LELP--R--MKRVI-E-TPVQ----------IIQ-KD-HFPPHKP-RS-VV-----G  129
GSN.CM.99.CN166                ---PWPGA....PPSP..Q-L--NFPAA.NSPGG.GGRTPA..............................................AQSLDT-RGTPEQGTEG-EQEH-LSL-VIS--G--MKEVS----VVSM---------IVQDSAI..ELDTP-T---V-----L  114
GSN.CM.99.CN71                 ---PWPGA.....TSP..Q-L--NFPAA.ISPGG.GGGTPA..............................................AQPVSAGGGAPEQGAEGRKEEH-LSL-VIS--G--MREVSM---VVSM---------IVQ-SAI..ELDTP-T--TV-----L  113
LST.CD.88.447                  ---LWTLE....EWSS..-KL-FDGGE..CRS.S-ISSP.............................NGEEPNTSRESLRNIQEPRTTAEE.....ATADPTEMCGRAMPEFLL-SR--IEN-LVN--PVS-----------IFS-N-V..RIE-PYT-RVV-----Q  123
LST.CD.88.485                  ---LWTLE....EWSS..-KL-FDGGE..CRS.S-ISSP.............................NGKELNKSRESLRNIQEPRTTTEE.....ATAGSTEMCGRAMPEFLL-SR--IEN-LVNR-PVS-----R-----IFS-N-V..KIE-PYT-RVV-----Q  123
LST.CD.88.524                  ---VWSLE....EWTF..-QL-NHGRG..CRS.DSFSAT............................HGESASNKSREGNRDIQEPRTTTEE.....ATTSSTEVCGRALPEFLL-SR--IEKVLVN--EVT-----------IFS-N-V..RIE-PYT-RVV-----Q  124
LST.KE.x.lho7                  ---IWSLE...-QTSR..QLS-TGRS...CRE.NSISSP..............................NGECSNKSRESTGDLQDFRTTAEK...AA-ASPTEMCGRALPEFTL.SR--IES-LVN--PVS-----------IFS-N-V..RIE-PYT-RVV-----Q  123
MAC.US.x.251_1A11              ---PWSMG....KE-...PQ-PHGS.......-ASGADANCSP...............................RGPSCGSAKELHAVGQAAERKT-RKQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV------  121
MAC.US.x.251_BK28              ---PWSMG....KE-...PQ-PHGS.......-ASGADANCSP...............................RGPSCGSAKELHAVGQAAER....KQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV------  117
MAC.US.x.EMBL_3                ---PWSMG....KE-...PQ-PHGS.......-ASGADANCSP...............................RGPSCGSAKELHAVGQAAER....KQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV------  117
MND-1.GA.x.MNDGB1              ---VWPLG....SLQT..G-L-.........................................GTRGDSNSSTIRGETSAENSEHLSEIRERA.....QAE-EG-EERGG-S--EYS-SR--IEEVSVD-VTIR----------IFN-RNI..KLK-N-Q--I------N  116
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  121
MND-2.GA.x.M14                 ----CSLG....QWQA..Q-L-.......CHATDPISTP...........................DARDGGSSREDVIRLYEKGATTEGSKGD....QAGEG--SLRG-LQL--FS--N--TTVVE-E--KV------------IKDVD-....T-N---QI------A  122
MND-2.x.x.5440                 ----YPLG....QWQA..Q-LP.......CHAIDPISTP...........................DARVGRPGRKDAVRLYEERATTEGSSR-....PERGKE-SLRG-LQL--FS--N--TTVVE-E--KV------------IKDLE-....T-N---RI------A  122
MNE.US.x.MNE027                ---LWPMG....KE-...PQ-PHGP.......NASGADTNCSP...............................RGPSCGSAKKLHAVGQTAER....KQKEALQGGGRG-AA--FS--R--V--AH-E--PA-V--------SIVAGIE-....GPHYT--IV------  117
MON.CM.99.L1                   ----SFRS....QEGA..K-LP-HLCAA.QRTDSAS............................................RDCGASQSQPRDWF-...EDG-PK-DEAG-LSLLSIS--G--IRSVL----KVSM---------IIQDQHV..NLGEN-T--VV-----M  114
MON.NG.x.NG1                   ----SFW.....PRQE..GAKELSSNLL.A----SGTN...................................PXGXQGELLTSSGTTRDP.......SADSPE-DEEG-LSLLSIP--G--VR-VE-A--SVSM----X----IVQXQDI..NLGEN-T---V-----L  118
MUS-1.CM.01.1085               ---GS-RP..PEKE-...E-LPPDLSLA.NCSAR-GGGIGQSLSSS....................................................-SRGGETGQEE-LS--VIP-SG--VRPVE-E--VVSM---------IIQDAHI..HLE-P-N--SV-----N  110
MUS-1.CM.01.CM1239             ---GSSWP..PEKE-...K-LPPDLAFT.NGPARAGGRPGQSLPAS....................................................-SQGWSERQEE-LS--VIS-SG--IREVM-E--VVHM---------IVQDSNI..QIDKP-N--LV-----N  110
MUS-2.CM.01.CM1246             ----GSRL..PVSE-...E-LP-YIAHA.YGTPRYS-AGELLP.TSRAQG.................................................RATQG-ERQKD-LS--VIS--G--IK-VD-N--PVSM---------IIQDSDI..KIS-ELT--VV--L--L  113
MUS-2.CM.01.CM2500             ----RFR...-PQE-...E-LPCYIAHA.ISSPRYS-DS-LLP-GGSTEGE...............................................PNPPSPTGQEE-LS--VIP--G--IR-VMVH--PVSM---------IIQ-SEI..EIDKE-T--VV--L--L  114
OLC.CI.97.97CI12               ----MEGR...ESR-SPQHATEDRDGHL.SLCTYPG................................................TDFGEGNK......-KERDE-GEET-LSL-SIS--R--MIELY-E--QVT-----------IKGKDI..YME-NGI-RI------A  110
RCM.CM.SIVagi-00CM312          ----CSLE....AGET.EEFSFNLRDTL.CSRDGSV-P..................................CGSHVKEILGAGEKAEAAKEAG....-EKPAGESIRRGIVL--FS--K--VTWVT-E--KV------------ISDIE-....GDN----I------Y  122
RCM.CM.SIVrcm-02CM8081         ---KCALE....SGET.EELSVDFLDSL.CSRGG-GQS................................CLPVRSSRSHAKEIPGTGQNAETATE-GREGSEGESLRG-TCL--FS--K--VTMVN-E--KV------------INGIE-....GEN----I------Y  128
RCM.GA.x.GAB1                  ----TPLG....SGEA.EELSFDFLDSL.CSRDGEQLRP.................................CRRDAKELSEEGRGTKETTEAG....REQEER-SISG-L-L--FA--K--TTIAQ-E--KV-V----------I-GIE-....GND-T--I------Y  123
RCM.NG.x.NG411                 ----CSLE....AGET.EEPSLDFLETY......CS-........................ERGRMESPVRSSRRDAQEIPSIGETAQTGAEER....EQGESGESLRGCVVL--FS--K--VTWVE-E--KV---------G--IQNLD-....GEN----I------Y  127
SAB.SN.x.SAB1C                 ---VWPLG....QRET.QEFP-DLHQTN.SSPNG-GLQQ.............AGGKLVCRQTSDQRTRARRSSNSPVKAVCCSGETAETAVA......KPLATTEPLRGGLQL---S--R--MK-VY-E--KVT---------S-IQGIE-....GDN---RI------C  141
SMM.SL.92.SL92B                ---PWPMG....KE-...SQLPHAD.........NIIDTVSTP...............................RSSSPDSQ.GVSRESTEGKD....KEEQALQGGDGG-AA--FS-RR--V-RAL-E--PL-V--------SIVAGIE-....GPGYT-RIV------  114
SMM.US.x.H9                    ---AWPMG....KE-...PQ-PHGP.......DASGADTNCSX...............................RGSSCGSTEELHEGGQKAEG....EQRETLQGGNGG-AA--FS--R--V--AY-EE-PI-V--------SIVAGIE-....GPNYT--IV------  117
SMM.US.x.PGM53                 ---AWPMG....KE-...PQ-PHGP.......DASGADANCSP...............................RGSSCGSTEELHEDGQKAEG....EQRETLQGGDGG-AA--FS--R--V-AAY-EE-PV-V--------SIVAGIE-....GPNYT--IV------  117
SMM.US.x.SIVsmH635F_L3         ---VWPMG....KE-...PQ-PHGP.......DASGADTNCSP...............................RGSSCGSTEELHEDGQKAEG....EQRETLQGGDRG-AA--FS--R--V--AY-EE-PV-V--------SIVTGIE-....GPNYT--IV------  117
STM.US.x.STM                   ---LWPMG....KE-...PQLPHGP.......NTSGADANCPS...............................RRPSCGSAEKLHAAGQEAER....EQE-TLQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVAGIE-....GLQYT--VV------  117
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  125
SYK.KE.x.KE51                  ----RVGE......QETPSK--CDEGDT.NSPTSGG...............LAMAATDMGKLCFTPTEAHSNGHPATATVLSKDSQGPLGPWETF...REER-DEGEGT-LSL-SIS--R--MME-NVE-TKVTM---------IIR-QDI..QLHQP-T--IV--L--N  143
SYK.KE.x.SYK173                ----RMGG...LQE-P..SQLP--ERAI.C.......SPSG...............GHRRWAMAHMVSTDEPTSTSEGTELSLEETPHQQGSALA...EGEQWE-GDQE-LSLS-LS--R--MIEVDVE-D-VQM-V-------IIR-EDI..QLHQP-S--IV--L--N  137
TAL.CM.00.266                  ----VFFR....ENPT..GQRTE-FSSG.ATIGTAN......................................RASHSIMAGGDYHVDPTSQE.....APRNNHNKGE-LSLFGQS--R--I-KVL-E--EV-V---------IIQVGEI..ELE-VPHQ-TV--LA-Y  118
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  119
TAN.UG.x.TAN1                  ----NANL....-NKA.AQF-R.ARRSR...SDGSPDACARIP...............................HRFASCGSIRPSQETAGAVCQKG..RSIEKAEGEGTGGL-LEFP--R--MSK-Y-Q--PVQV---------IIK-EDI..HLE---R--VV-----G  127
VER.DE.x.AGM3                  ---VWPVD....-SET.QKFSRRYSWGG.ANCA-STES..................................IRPCKEAPAAICRQGEAVEGTKE...KTTSSESRLDRGIF-.ELP--R--IK-VY-E-VPIR----------IIK-AD-..QLS-T----I------G  124
VER.KE.x.9063                  ---VWTVD....-DKT.EKLSRCHSWSG.TERA-STDT..................................IRPSKEAPAAVCREGKTVEGTRE...ENTTNKSRLGGRIF-.ELP--R--IK-VI-E-VPVKV---------IIK-TDI..QLT-P-R--V------G  124
VER.KE.x.AGM155                ---VWPVD....-DKT.KKFSR-HSWGG.TKCA-STEQ...............................LYTLRPSKEAPAAVCRERETNEKSEQ...KPPSE-SRLERGIF-.ELP--R--IR-CI---TAVK----------IIKDTD-..QLR-S-R--IV-----G  127
VER.KE.x.TYO1_patent           ---VWTVD....-GKT.EKFSRRYSWSG.TECASSTERHHP...............................IRPSKEAPAAICRERETTEGAKE...ESTGNESGLDRGIF-.ELP--R--IK-VY-E-VPIK----------IIK-ND-..QLS-P-R--I------G  127
WRC.CI.97.97CI14               ----WQRA....DALEQLSSSEA-ELEQ.FWANNNNAFCSSNE-GSKH.......................SGAGKIFIEGSTIEGTTGTQE......DRGTEE-SGGAGMAL--IS-RT--MIQCI-E-HSVS-----------IDVNL.....NKN-Q-RI------T  131
WRC.CI.98.98CI04               ----WERA....DALKQLPSSKG-ELEQ.FWAG.NNTVCSSNEGGSKH.......................TRIREIPPERSSIEGAGTGK.......KGREEERNGAARVA---IS-RN--MIQCI-E-NSVS-----------IDIHLN....NKN-Q-RI------T  130
WRC.GM.05.Pbt_05GM_X02         ----WPRA....EALQQLSS.............SKN-CRGEYLWASG....................ANSTSGSYHARTRAISSERSSDEEG......ERERNAEGGESRISL--IS-RT--MIQCI--RESVS-----------VDVQL.....E-N-Q-RI------T  122
```

```
                                                                                protease end p66, p51 RT start                                 M41L                  K70R
                                                                                                                                                                    D67N
H1B.FR.83.HXB2             IKVRQYDQILIEICGHKAI.....GTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETV...PVKLKPGMDGPKVKQW..PLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKS......VTV  263
H1A1.UG.85.U455            ---------------K-T-.....----------------M--------------------...-------E----------..----------T---N-------------------------------------------------------------T--------......---  262
H1B.US.90.WEAU160          --------VP---------.....-------------------------------------...------------------..-----------------------------------------------------------------------------S--------......---  263
H1C.ET.86.ETH2220          ---------I-----K---.....----------------M---L-R--------------...------------------..----------TA--E---Q----R------------------------------------------------------------------......---  270
H1D.CD.84.84ZR085          -------H-----------.....-------------------------------------...------------------..----------T----D--------R---------I---------------------------------------------------------......I--  263
H1F1.BE.93.VI850           ---K---N-----------.....----------------M----------V---------...------------------..----------T---L-------------------------------S--------K-------------------------------------......---  263
H1G.SE.93.SE6165           --------VP---S-K---.....---I------I-----M--------------------...------------R-----..----------T---K---E-----------------I---------------------------------------------------------......---  260
H1H.CF.90.056              ------E-VA-----K---.....----------------I--------------------...------------------..----------T-----------------R------S--I---------------------------------------------------------......-S-  262
H1J.SE.93.SE7887           ------NEVP---E-K---.....----I-----------M---L----------------...-------------I----..----------TQ--A---E-----RV------------------------------------------------------------------......---  261
H1K.CM.96.MP535            --------V------Q---.....-------------------------------------...------------------..----------T-----------------------------------------------------------------------------------......---  263
H101_AE.TH.90.CM240        ---K-----------K---.....----------------M-------------D------...--T---------------..----------T---K---E---------------------------------------------------------------------------......---  268
H102_AG.NG.x.IBNG          ---------------K---.....----------------M--------------------...------------------..----------TD----------------------------------------------------------------------------------......---  263
H103_AB.RU.97.KAL153_2     ---------------K---.....----------------M---L----------------...--T---------------..----------TD--K----------------------------------------------G---------------------------------......---  263
H104_cpx.CY.94.CY032       ---------P-----K---.....----------------M---L----------------...------------------..----------T----D-----------------I---------------------------------------------------------......---  262
H1N.CM.95.YBF30            -------N-TVD-Q-R--V.....-------------------------------------...------------------..---T---E--R-------------R---------I----------------------------------------------Q-------------......---  270
H1P.FR.06.RBF168           ---KEFEEVKV--E-REVY.....----X-----X-----I--L-----S----S--I-..------A------------..--SK---E--RAX-Q---Q----T-----------I--------SX------------------------------G----R--......---  259
CPZ.CD.90.ANT              -S-Q--NKVP-Q-GDRTVL.....A---L--N--------V-CLL---------KV-----...------E-----R-----..--SK---E--K---DKL-A-N---R---D------I-------TS-------------------I-----------Q-------......---  263
CPZ.CM.-.SIVcpzDP943       ---K---N-A---E-R--T.....----------------I---------------G----...-------------R----..----------T---Q-------------------I-------GE------------R---------------------------------------......---  270
CPZ.CM.-.SIVcpzMB897       ---------P-----KR--.....----------------I-------------T------...------------R-----..----------T-------------R---------I---------------------------------------------------------......---  271
CPZ.CM.01.SIVcpzCAM13      ---K---NVS---E-K-VQ.....----------------I--------V----S------...------------RI----..--SA------T---Q-----------------I-------S---------------------------------E---------......---  262
CPZ.CM.05.SIVcpzEK505      ---K---NVI---Q-K--V.....----------------F----------------I---...------------R-----..----------T-------------R---------I-----------------------------------------------------------......---  269
CPZ.CM.05.SIVcpzLB7        ---K---N-P---G-K--L.....----------------I-------------------..-Q------------------..----------T------Q-----R---------------------------------------------------R-------......---  266
CPZ.CM.05.SIVcpzMB66       -------NVT-D---KR-V.....----------------I-------------------..-S--------R---------..--------R-T-------------RV--------I--------------------------------------------Q----------......---  266
CPZ.CM.05.SIVcpzMT145      ---K-FERVN---E-K-IT.....S---------------V--KL-----------D---...K-F----------------..--SK---E--T---K-----------------I-------------------------E-----------------------......---  262
CPZ.CM.98.CAM3             ---K---N-A---E-R-TT.....----------------I--------------------...------------R-----..---A------T---Q--------R---------I---------------------------------------------------------......---  268
CPZ.CM.98.CAM5             ---K---N-V---E-R-TT.....----------------I--------------V-----...------------R-----..---A------T---Q-------------------I--------------------------E-----------------------......---  264
CPZ.GA.88.GAB1             ---K-F-NVH---E-R-VV.....----------------I---L---V----S-------...------------------..--SA------T---Q-------------------I---------------------------------------------------------......---  261
CPZ.GA.88.GAB2             ---K---NVE---E-KRVE.....S---------------V---L----------TP----...--------T---R-----..--SK---E--T---R--------R---------I---R-----------------------------------------------......---  266
CPZ.TZ.00.TAN1             VP-SE-YN-PVQ-GNKEVR.....A-----E--I------I-K-L-----------V----...K-Q--E------------..--SK---E--T---KTL-------AV-----------I-------TS--------------------L----------R-RNM......---  259
CPZ.TZ.2001.TAN2           VS-SE-YN-PVQ-GNKEVR.....A-----E--I------I-K-L----------V-V---...K-Q--E------------..--SK---E--T---NTL-------A------------I--------S--------------------L-------S--R-RNM......---  260
CPZ.TZ.2002.TAN3_1         VP-SE-YN-PVQ-GNKEVR.....A-----D--I------V-R-L----------V-V---...K-Q--E------------..--SR---E--T---NTL-------A------------I---R---TS--------------------L-------S--R-RNM......---  260
CPZ.US.85.CPZUS            ---K---HVN---E-R--Q.....-S--------------I--------------S-K---...--R--------R------..---A------T---Q--------TRV---------I----------------------------------------------R-......---  265
                                                                                 protease p51 RT start
MAC.US.x.239               -NTKE-KNVE--VL-KRIK.....--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-R......I--  275
H2A.DE.x.BEN               -NTKE-KNVE-KVLNKRVR.....A-IMT-D--I--F---I--AL-MS--L-VAK--PI...K-T----K---RL------..----K---E--K---EK-----QLEEAP-T------T-------KN---M-I-------V----T-I-----------A---R......ISI  288
H2A.GW.x.ALI               -NTKE-EDVE-KVLNKRVK.....A-IMT-D--I--F---I--AL-MS--L-VAK--PI...E-R----K----LR-----..----K---E--K---EKT-R--QLEEAP-T------T-------KN---M-I-------V----T-I-----------A--RR......I--  288
H2A.SN.x.ST                -NTKE-KNVE-RVLNKRVR.....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---R......I--  288
H2B.CI.x.EHO               -NTNE-KNVE--VV-KRVR.....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--R......I--  288
H2B.GH.86.D205             -NTKE-KDVE--VV-KRVR.....A-IMT-D--I--F---I-NTL-M-----VAKV-P-...K-E----K----IR-----..--SR---L--K---EK-----QLEEAP-T------T-------KN---M-I-------V----T--N#-------#AE-RR......I--  287
H2G.CI.92.ABT96            -NTKE-KNVE-KVV-KRIQ.....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-RR......I--  271
H2U.FR.96.12034            -NTKKFHNVN-KVL-KTIR.....--IMT-D--I--F------EL-MS-----GKEAL-...IGG----K----LR-----..--SK--LE--K---EK-----QLEEAP-T------T-------KN---M-I------RV----S------------AE--R......I--  285
COL.CM.x.CGU1              -R-K--RDKILKMG-KEYR.....-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-TH......I--  245
DEB.CM.99.CM40             -N-KC-PG-E-AMANKV-V.....SDL---D--I--L---Y-AKM-VSI---VLDYTP-.........---E-A----I----..--SR---E--KD-LKP-IEA-Q-VPAA-T----S---V-----KS---M-I-L-A---A--EV----T-----SA-PQMEQ......I--  269
DEB.CM.99.CM5              -N-KC-PG-E-KMADKV-V.....SD----D--I--L---Y-AKM-VSI---VLDFTP-.........---E-A---RI----..--SR---E--K--IRP-VEA-Q-VPTE-T---SS---V-----KN---M-I-L-A---A---V----T-----S--PQMEQ......I--  269
DEB.CM.SIVdeb04CMPF3061    -N--C-PG-E-QCQ-KV-V.....-E----E--I--L---F-APL-FSI-..MLESQIQ.YT-----E-KS-------ASSQGKDR..GSQGNSGKDGGGRPMAPAS-T----S---------KN---M-I-L-----A--PV----T---------PQ-RQ......I--  272
DEN.CD.x.CD1               -N-QA-PG-E-KLMDKV-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-EH......I--  289
DRL.x.x.FAO                -N----FNCE-K-A-KITH.....AS--------------V-CKL-A-----V-QV-P-...K-T---------I-------..--SK---E--K---EQ-----Q--R-------------C-----G----------Q---A----F---------G---ACEQ......I--  277
GOR.CM.2004.SIVgorCP684con ---KE-EDVE---EHR-VV.....---------A------V--K--------------V-...--S----------------..--SK---E--TA--Q---Q-------------------I---------------------------------------G--------......---  255
GOR.CM.2007.SIVgor2139_287 ---KE-EDVE---EHR-VV.....----------------I--K----------S-DV-...--S----------------..--SK---E--TA--Q---Q-------------------I---------------------------------------G--------......---  255
GOR.CM.2007.SIVgorCP2135con ---KE-EDVE---EHR-VV.....----------------I--K----------S-DV-...--S----------------..--SK---E--TA--Q---Q-------------------I--------X------------------------------G-----R-......---  255
GRV.ET.x.GRI_677           -H-KE-QGVQVQLEDKIIT.....-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---Q......I--  285
GSN.CM.99.CN166            -Q--EHKHVEVIFNEKRIK.....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-----------IPHPAGLEQKEH--I  274
GSN.CM.99.CN71             -Q-KEHRHVEVIFNEKRIK.....A-----------L---C-SKL-V---MVQQKL-P-...E-H--S-KE---I------..--SK---E--TQ-TQ--L-L-QLE---------S---------KSQ--M-I---K--EA---LA-----------------......EQKEH--I  267
LST.CD.88.447              -R-KE-KDVF---A-KTTV.....----L-----D-V---I--NV-AR-VLSQLSDKIP.ITK------C---R-------..--SK---EG-QA--DRL---------ADAG------I-------KNE----I---Q--AL----H-----------I--C-Q..........---  279
LST.CD.88.485              -R-KE-KDVF---A-KTTV.....----L-----D-V---I--SV-AK-VLAQLSDKIP.ITK-R----C---R-------..--SK---EG-QA--DRL---------ADAG------I-------KNE----I---Q--AL---LH-----------I--C-Q......I--  279
LST.CD.88.524              ---KEFRDVF-Q-A-KTTV.....----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-Q......I--  280
LST.KE.x.lho7              -R-KE-RDVF---A-KTTV.....----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-Q......I--  279
MAC.US.x.251_1A11          -NTKE-KNVE--VL-KRIR.....--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-R......I--  275
MAC.US.x.251_BK28          -NTKE-KNVK--VL-KRIK.....--IMT-D--I--F------AL-MS--L--AKV-P-...K-T----KV---L------..--SK---V--R---EK---D-QLEEAP-T------T-------KN---M-I------RV----T------------A-R-R......I--  271
MAC.US.x.EMBL_3            -NTKE-KNVE--VL-KRIK.....R-IMT-D--I--F------AL-MS--L--AKV-P-...K-A----KV---L------..--SK---V--R---EK---D-QLEEAP-T------T-------KN---M-IH-----RV--ELYRSPIR--------A-R-R......I--  271
MND-1.GA.x.MNDGB1          LR-K---NVYV--R-KGTF.....-----I-----ID-----IMEKL-GK-ILAQLSDKIP.ITK------V---RI----..--SK---VG-QK--DRL-E-----RVD-G-N----I-------KNE----I-------L----H-L---------I--C-R......I--  272
MND-2.CM.98.CM16           -N-K-FFNCKVT-A-KTTH.....AS---------V----V-KKL-------V-KV---...K-T----T----I------..--SK---L--Q---SQ-----Q--------------C-----G-S--------Q---V----F---------G---QCEQ......I--  275
MND-2.GA.x.M14             -R-K--FNCK-TVA-KSTH.....AS--------------V-KKL---------K----...K-T----------RI----..--SK---L--QT--D---------R------------C-----G-S--------Q--RV----F---------G---QCEQ......I-I  276
MND-2.x.x.5440             -R----FNCKV--A-KITH.....AA--------------V-KKL---------K----...K-T----T---RI------..--SK---L--Q---NQ--------RV-----------C-----G-S--------Q---V----F------T--G---QCEQ......I--  276
MNE.US.x.MNE027            -NTKE-KNVE--VL-KRIK.....--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-R......I--  271
MON.CM.99.L1               -E-K--KA-QVIFEDREVW.....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........-I  271
MON.NG.x.NG1               -Q-K-FKGVTVSFQ-KTVR.....X---L-X--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------------QNMAN......I-I  274
MUS-1.CM.01.1085           -S-KE-RGVQVCFN-KTIR.....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--I  264
MUS-1.CM.01.CM1239         -S--E-RGVQVSFNEKTIK.....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--I  264
MUS-2.CM.01.CM1246         -N-K--KGVIVKFNEKRIV.....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-...A...H--I  269
MUS-2.CM.01.CM2500         -N-K--KAVKV-FNDKRI-.....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--...A...H--I  268
OLC.CI.97.97CI12           -R-M--RCRIRIF.EKEIQ.....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--------------------EREQ......I--  265
RCM.CM.SIVagi-00CM312      -N-K--YNCK---A-KVTH.....AH--------------V-HKL-AS------KA-II...K-E--E-Q---R-------..--SK---E--T---NA---------R-----------I-C-R-----------------------F---------G----C-Q......I--  276
RCM.CM.SIVrcm-02CM8081     -N-K--YNCK---A-KVTH.....AH--------------V-HKL-AS------KA--I...E------Q-----------..--SK---E--T---NA-----------------------I-C-R-----------------------F---------G----C-Q......I--  282
RCM.GA.x.GAB1              -N-K--KNCE---A-KRTH.....AH--------------V-KKL-A-------Q---I...K-E--S-Q---R-------..--SK---E--T---NA-----------------------I-C-------------------------F---------G---QCER......I--  277
RCM.NG.x.NG411             -N-K-FNHCK-T-A-KTTH.....AH--------------V-SKL-AS------KA--I...K-E----Q---R-------..--SK---E--T---NA--------R-------------I-C-------------------------F---------G---QCRQ......I--  281
SAB.SN.x.SAB1C             -N-KA-HNQEVK-EDKTCK.....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------------QQREQ......I--  295
SMM.SL.92.SL92B            -NTKE-RTVK--VL-KVIR.....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-------------EMER......I--  268
SMM.US.x.H9                -NTKE-KDVK-KVL-KVIK.....--IMT-D--I--F------AM-MS--L--AKVXPI...K-T----K----LR-----..--SK---I--R---EK---D-QLEXAP-T------T----XX-GN---M-I-------V----T---------X--A-RRR......I--  271
SMM.US.x.PGM53             -NTKE-KDVK-KVL-KVIK.....--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-RRR......I--  271
SMM.US.x.SIVsmH635F_L3     -NTKE-KDVK-KVL-KVIK.....--IMT-D--I--F------AM-MS-----AKV-PI...K-T----KE----R-----..--SK---I--R---EK---D-QLEEAP-T------T-------KN---M-I-------V----T------------A-RRR......I--  271
STM.US.x.STM               -NTKEFKNVN--VL-K-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-RRR......I--  271
SUN.GA.98.L14              -R-KE-ADVFV--RDKT-T.....-----------D-----I-SAL-GR-VLATLSEKIP.ITK----E-AA---------..--SR---EG-QK--DRL-A-----RAD-G------I-------KNE----I---V--GM----H-L------------C-Q......I--  281
SYK.KE.x.KE51              -R----N--RFQ-ERPDGRIKEVE-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--  301
SYK.KE.x.SYK173            -T----RN-RFTVVKPSGKRKQVE--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-I  295
TAL.CM.00.266              -Q-QS-SA-E-QWENKR-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-Q......I--  272
TAL.CM.01.8023             -Q-QS-TGVE--WEHKR-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--Q......I--  273
TAN.UG.x.TAN1              -S-K---N--VN-D-KETR.....-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--Q......I--  283
VER.DE.x.AGM3              LN-KE-SDREVRLEDKILR.....--I-I-S--I------I-APA-AK-VMGQLSEQIP.IT-----E-AR--FL-----..--SK------Q---DQL----------G--A------C-----KSQ--M----------A----F---------S-FE-MTE......I--  280
VER.KE.x.9063              LN-KE-SNVEVK-EDKILR.....-----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-Q......I--  280
VER.KE.x.AGM155            LN-KE--NVEVQLEDKILR.....-----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-Q......I-I  283
VER.KE.x.TYO1_patent       LN-KE-NDREVK-EDKILR.....---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-MRQ......I--  283
WRC.CI.97.97CI14           MN-KLHR-VK-R-L-KTIE.....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-N......IS-  287
WRC.CI.98.98CI04           MN-KL-RK-Q-R-L-KTVE.....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-N......IS-  286
WRC.GM.05.Pbt_05GM_X02     MQ-KL-KDVR-K-L-KEVR.....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--RN......I--  278
```

K219Q
D110 catalytic site polymerase motif T215Y

H1B.FR.83.HXB2	LDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDS..WTVNDIQKLVGKLNWASQIYPGIKV	431
H1A1.UG.85.U455	--------------S------------------V---------------S--------------S-H------------------------A------A---S--FI-----------------------------Q------..--------------------A----	430
H1B.US.90.WEAU160	---F---------D--------------------K----E-..--------------------A----	431
H1C.ET.86.ETH2220	--------------G-----------T----------------------P-----PQ-------AP--E----------------------AP-----E---K--F------------------------------Q------..-------------------------	438
H1D.CD.84.84ZR085	---------I--C--I------------EV---------------------A---K--E-----------------------------------S-T----E-..-------------------------	431
H1F1.BE.93.VI850	-------------K--K---------V--------------------------C----------MK--------------------------------E------F-----------------H------------Q--N---..-------------------------	431
H1G.SE.93.SE6165	---------------------------------V-----------------------R------AN--EM---------------------A------E---K--F------------------------------Q--D-E-..-------------------------	428
H1H.CF.90.056	-------------KE---A---E---EM---------------------A------A---K--F----------------------------TVK------..---------------------N---	430
H1J.SE.93.SE7887	------------Y--C------K---ER--E------------------E---R--K---E---K--F------------------------------Q----ED..-------------------------	429
H1K.CM.96.MP535	-------------K-------------------V-------------------H----------IK--EM-------------------P--------E---K--F------------------------------Q--D---..-------------------------	431
H101_AE.TH.90.CM240	--------------S---IK--EM----K-----------------------A---S--F-----------------------R------E------..--------------------A----	436
H102_AG.NG.x.IBNG	-------------K------------V--------------------------A----------TK--E----------------------A------G---K--F-----------------------------VE------..--------------------A---I	431
H103_AB.RU.97.KAL153_2	-------------Q------------T---E-----------------------E-----E------F-------------------------------------..--------------------A----	431
H104_cpx.CY.94.CY032	-------------PE-----------T--------------------------C----------FK--E----------------------A------E------F------------------------------QPA----..-------------------------	430
H1N.CM.95.YBF30	---------C---K--T---------EKH-E-I---------------LA---EAV-D--D---K--F------------------------------K-----V..-----------------------R-	438
H1P.FR.06.RBF168	---------C---P-----X------V------------------X-X---X-Y----------ES--EVE--------------PLSE--QRV-K--E--YV--F--------X---------------------K----EA..-----------------------R-	427
CPZ.CD.90.ANT	---------I---Q------------V---------C----------------A------A---DKY-AVE-------------M--TA--EM--K-----QV---E------------Q--------------K-K--P-D-..------------------------T	431
CPZ.CM.-.SIVcpzDP943	---------C----S-----------V-------------------------------------Q---ELT------------X----E--K-V-L------T--F------------------------------Q--Q-ET..-------------------------	438
CPZ.CM.-.SIVcpzMB897	---------C----E-----------V------------------------------------Y----E----------------H-E---E-V----A---K--F------------------------------Q--X---..-------------------------	439
CPZ.CM.01.SIVcpzCAM13	---------C---K------------V-------------------------------------AK---TV--------------D-T---Q-V--------H--F------------------------------K----EV..-----L---I--------------I	430
CPZ.CM.05.SIVcpzEK505	---------C----N-----------V---------------------------T----------N--EL-----------------T---EAV-R--S---T--F----------------------------T-Q-----T..-------Q-----------------	437
CPZ.CM.05.SIVcpzLB7	---------C----N---R--------------------K-E---N-V----A---K--F------------------------------Q----E-..------------------------I	434
CPZ.CM.05.SIVcpzMB66	---------C----N-----------V---E-I--------------K-EL--E-V----A---K--F------------------------------Q----E-..---------I------C------R-	434
CPZ.CM.05.SIVcpzMT145	---------C---P----F------------------------------S--------------TKH-E-I--------------P-EE--KRV----A---K--F----------------------------Q-K---QEV..------------------------I	430
CPZ.CM.98.CAM3	---------C----N-----------V-----------------------------I---T---Q-H-LI----------------LRE--E-V-L------T--F------------------------------Q--Q-ET..-------------------------	436
CPZ.CM.98.CAM5	---------C----N-----------V------V-------X------------------D----H--LI------------------K--E-V-L------T--F------------------------------Q--Q-EI..--------------------S----	432
CPZ.GA.88.GAB1	---------C---K-------------------V---------------S--------------EK---T-----------------D---K-V--------K--F------------------------------Q----EV..---------I--------------I	429
CPZ.GA.88.GAB2	P--------C----N-----------V-------------------------------V-D---QK--EL---------------SLEE--KRV-Q--E------F------------------------------Q--Q-EN..------------------------I	434
CPZ.TZ.00.TAN1	---------I---P------------L--N---K-F------------------------D---EH--VD----------I----NEEE--KL-KK------T---E-----Y-EK---M--------N-----N-T---PEQ..----H---------------H---T	427
CPZ.TZ.2001.TAN2	---------I---P------------L--N---K-F------------------------D---EH--VD----------I----SEEE--KL-KK------V---E-----Y-EK---M--------H-----N-T---PEQ..----HX--------------HX--T	428
CPZ.TZ.2002.TAN3_1	---------I---P------------L--N---K-F------------------------D---EH--VD----------I----SEEE--KL-KK------L---E-----Y-EK---M--------N-----N-T---PEQ..----H---------------H---T	428
CPZ.US.85.CPZUS	---------C---KE-----------V---------------------------------D----H--VI---------------NLEK--E-V-L---Y--T--F-------E----------------------Q--Q-EI..------------------------I	433

D catalytic site DD catalytic site

MAC.US.x.239	--I------I----E--Q-----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-------T	443
H2A.DE.x.BEN	---------I--H----Q-----L-AV--ME--K--I-K--------------YT-RQV------A---VILI-----ILIA--RTGLE-DKVVLQ-KEL-NGL-FS---E-F--D---Q---C--W-T--KL-KLQ--Q--I..-----------V----A---S---T	456
H2A.GW.x.ALI	---------I--H-S--Q-----L--V--AE--K--I-K--------------HT-RQ-------A-Q-VILI-----ILIA--RTDLE-DKVVLQ-KEL-NGL-FS---E-F--D--YK----G-W-T--KL-K-Q--Q-EV..--------H--V----A-------T	456
H2A.SN.x.ST	---------I--H----Q-----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	456
H2B.CI.x.EHO	-------------P---Q-----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	456
H2B.GH.86.D205	I--------I---PN--Q-----L--V--AE--K--I-K---------QS-C-Y--R-V-D----A-S-VI-I-----ILIA--RSDLE-DRVVSQ-KEL-NDM-FS--EE-F--D---K------W-K--KL-K-Q----EV..----A------V----A-LF----T	455
H2G.CI.92.ABT96	--I------I---V---Q-----L--V--AE--K--I-K--------------HT-R-V-D----A-E-XIXI-----IL-A--RTDLE-DQVVLQ-KEL-NKX-FS---E-F--D--YQ-----XW-K--KL-K-N-X--ET..-----------V-X--A-L-----T	439
H2U.FR.96.12034	---------I--C----Q-----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-----T	453
COL.CM.x.CGU1	--I------I---P-YQ---------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----RT	412
DEB.CM.99.CM40	--LA---Y-IA--PT-A------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---T	437
DEB.CM.99.CM5	--LA---Y-I---PQ-AR--------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---T	437
DEB.CM.SIVdeb04CMPF3061	--IA---Y-I---PS-AQ--------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---T	440
DEN.CD.x.CD1	--LK---Y-----P--AQ-----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---T	459
DRL.x.x.FAO	---------C---P------------V--QG----------------------ATAD---Q---EK---VI---------F---RPKHL-EQM-K---T--KF---E-----F-DN--YE----Q---K-----E-R--D-EI..-----------V--------S---T	445
GOR.CM.2004.SIVgorCP684con	---------C---P------------V------V-------------------H------D---K--E-E---------------P-TE--Q-V-K--E--YA--F------------------------------K----E-..-----------------------R-	423
GOR.CM.2007.SIVgor2139_287	---------C---P------------V------V-------------------H------D---K--E-E---------------P-TE--Q-V-K--E--YA--F------------------------------K----E-..-------R---------------R-	423
GOR.CM.2007.SIVgorCP2135con	---------C---P------------V------V-------------------H------D---K--E-E---------------P-TE--Q-V-K--E--YA--F------------------------------K----E-..-----------------------R-	423
GRV.ET.x.GRI_677	I-I----Y-I--CKE-----------V--TG------F-C---------T---NTAAN---EIKRHT-GLE-V------WLA--HDETR-NQQVDIV-KM--EK--E-----V-R---WE----K---N---INK-E--PLEG.E----K---V--V------------T	454
GSN.CM.99.CN166	I-MK---Y-I--Y---------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---RT	442
GSN.CM.99.CN71	I-MK---Y----Y-E-------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---RT	435
LST.CD.88.447	V-I------I---P-Y-Q-----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--RT	449
LST.CD.88.485	V-I------I---P-Y-Q-----V-----QE--R---------------C---ATVAGL-SE---L----IL-------LI---RDKKG-DQAVK---EL-MT-N-E--E--F-A---YH----V----R-EIEKVK--DI-LQKT---N------V----A-L-E--RT	449
LST.CD.88.524	V-I------I---P-Y-Q--------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--RT	450
LST.KE.x.lho7	V-I------I---PNY----------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--RT	449
MAC.US.x.251_1A11	--I------I----E--Q-----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-------T	443
MAC.US.x.251_BK28	--I------I----E--Q-----L--V--AE--K--I-K--------------YT-RHV------A---VTLV-----ILIA--RTDLE-DRVVLQ-KEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--QRET..-----------V----A-------T	439
MAC.US.x.EMBL_3	--I------I----E--Q-----L--V--AE--K--I-K--------------YT-RHV------A---VTLV-----ILIA--RTDLE-DRVVLQ-KEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--QRET..-----------V----A-------T	439
MND-1.GA.x.MNDGB1	--I------I---P-Y-P-----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-----T	441
MND-2.CM.98.CM16	--I------C----------------V--QG----------------------ATAD---Q---ERH--V---------F----RVAPE-SRM-Q---D---F---E-----F------E----I---K-----KVQ----EE..--------------------S---T	443
MND-2.GA.x.M14	--I------C----S-----------V--QG----------------------ATAD---KT-KEEY-EVL--------F-----SASE-NKMVNR--E--RF---E-----F------E----V---Q-----K-Q----ET..--------------------S---T	444
MND-2.x.x.5440	--I------C---V------------V--QG----------------------ATAD---KA-KEKH-EVL--------F-----SASE-SKM-V---E--QF---E-----F------E----V---K-----K-Q----E-..--------------------S---T	444
MNE.US.x.MNE027	---------I----E--Q-----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-------T	439
MON.CM.99.L1	V-MK--FY-I--W-P-------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---T	439
MON.NG.x.NG1	V-MK--FY-I----N-------S---V--AE-AV---FR-----X-----T--Y-ASELXKXI-RKY-EVTLI------LI---HSLPE--KIVTAI-TX-SEK-IQ--IE-Y-DKY-LQ-L------TH----K-E---Q-Q..----G-----------A-------T	442
MUS-1.CM.01.1085	V-IK--FY-----P--------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-----T	432
MUS-1.CM.01.CM1239	V-IK--FY-----PA-------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-T	432
MUS-2.CM.01.CM1246	V-MQ--FY-----KE--P----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---T	437
MUS-2.CM.01.CM2500	V-MQ--FY-I--Y-E--P----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---T	436
OLC.CI.97.97CI12	--I------I---PEY-E--------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---T	433
RCM.CM.SIVagi-00CM312	--I------C--Y-P----------A---QG----------------------TTAN-----V------L----------F---RTELE-SQM-KK--E---A--FE--E--F-DK---E----V---------E-K----EE..--------------------A---T	444
RCM.CM.SIVrcm-02CM8081	--I------C--Y-P----------A---QG--V-------------------TTANR-------E---V----------F---RTELE-NQM-KR--E---I--FE-----F-DK---E----V---------E-K----EE..--------------------T---T	450
RCM.GA.x.GAB1	--I------CL-Y-P----------AV--QG--V-------------------A-AN---Q---EE---VI---------F---RTKLE-DKM-KQ--D---F--FE-----F-DK--Y---------KS----E-K----EE..--------------------S-LRT	445
RCM.NG.x.NG411	--I------C--Y-P---------LA---QG----------------------TTAN--------E---V----------F---RTAHE-SQM-KR--E---T--FE-----F-DK---E----V---------E-K-L--EE..--------------------S---T	449
SAB.SN.x.SAB1C	--I------C---P--Q---------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---T	463
SMM.SL.92.SL92B	--I----------PE--Q--------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-------T	436
SMM.US.x.H9	---------I----E--Q-----L--V--AE--K--I-K-------------XHTXRNV------A---VTLI-----ILIA--RTDLE-DRVVLQ-KEL-NSI-FS--EE-F--D---Q------W-T--KL-K-E--QRET..----N------V----A-------T	439
SMM.US.x.PGM53	---------I----E--Q-----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-------T	439
SMM.US.x.SIVsmH635F_L3	---------I----E--Q-----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-------T	439
STM.US.x.STM	---------I----G--Q-----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-------T	439
SUN.GA.98.L14	V--------I---P-Y----------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-LRT	450
SYK.KE.x.KE51	--LK---YT-----E--Q-----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------------C	470
SYK.KE.x.SYK173	I-LK---Y-----KE--P-----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---T	464
TAL.CM.00.266	--MK---Y-I---P--SQ--------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---T	440
TAL.CM.01.8023	--MK---Y-I---P--AR--------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---T	441
TAN.UG.x.TAN1	--I----Y-I---PE--P-------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-----T	451
VER.DE.x.AGM3	--I----Y-I---PE-----------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---LRT	448
VER.KE.x.9063	--I----Y-I---PE-----------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---LRT	448
VER.KE.x.AGM155	I------Y-I---PE----------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---LRT	451
VER.KE.x.TYO1_patent	-------Y-I---PN----------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---LRT	451
WRC.CI.97.97CI14	V-I------I---PEY-Q-----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--QT	454
WRC.CI.98.98CI04	V-I------I---P-Y-Q-----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	453
WRC.GM.05.Pbt_05GM_X02	--I-----FI---PEY-Q-----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---T	445

```
                                                                                                                                   p51 RT end p15 RNase H start
H1B.FR.83.HXB2              RQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEP.FKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTP.KFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDG  599
H1A1.UG.85.U455             K--------A----DIVT-----------------D-------------V--------D----------.-----------K-S-----------V---VS---------I-.--R---------A--M-----------------------------D--A---------  598
H1B.US.90.WEAU160           K-------------I--I--------------------------------L-----------------.-----------V------------------------------------.---------------------------------------------------------  599
H1C.ET.86.ETH2220           ---------A----DIVT-----------------------F---------------ND---F-F----.---------F-KRGT----------AV----AL----------.--R---------A--D-----------------------------A-V-------  606
H1D.CD.84.84ZR085           K------------------------------------M-------------L------------------.-----------------------------AI----------.--R-----------ID-----------------------------I---------  599
H1F1.BE.93.VI850            -P-------A----DIV---A-------K-----R---------------------D---------N-.----------KV-S-----------------AL--------RSP------L----D---D---------------------------T---A--D------  600
H1G.SE.93.SE6165            TH-------A----DIVS--A---M---------R-------------E----V----LD----------.Y---------KRGS-----------V----A-----------.------R-----I---D---------R--T---P----Y----  596
H1H.CF.90.056               K--------A----DI----K-------------R--I---------------R------------------.----------K--T-----I----------S---------I-.--R-----------------------------------H---------T---A----Y-I--  598
H1J.SE.93.SE7887            K------K-A----DIV---R---------K-------------SA-E----V----LD-----------.-----------KR-S---------A-V----AL-A--------.--R----R--------D-----------------------------M---------  597
H1K.CM.96.MP535             K--------V----DIV---A-------------------------------------ND-----------.H-----------S-----------------A--G--------.--R-------------------------------------T------------  599
H101_AE.TH.90.CM240         K--------A----DIV------------------T-------------V--V-----D-----------.-----------RGS------R----V---VA-----------.--R----R-------M-----------------------------D-----------  604
H102_AG.NG.x.IBNG           K---R----A----DIVA----------------------------T---V--L-----D----------.-----------KK-S------------V---VAM----------.--R----R-------M-----------------------------D-----------  599
H103_AB.RU.97.KAL153_2      ---------A---------A-----------------------------V--------------------.-----------L-----------------A-------------.---------------------------------------------------------  599
H104_cpx.CY.94.CY032        K--------A----DIV---T-------------------A------------------------------.H----------T-S------R---------AM-C---------.--R--------D---------------------------TD--A---------  598
H1N.CM.95.YBF30             K-----I---------VNF-------------------L------G-E-V-------------------L.H----------K--S-----I---V-V-R-VA-----------.--R--V---V--A---DH---------------------------T---S---------  606
H1P.FR.06.RBF168            KE----I----S----VA-XR------E--K-----------Q-E-E--VD-----E----------E.H-----X---XQKAT----IR--A-VI--TSQ---------L-.--R--VNRNV--N--SD-------------S----I---------D--P-I-------  595
CPZ.CD.90.ANT               K---R-I--V-S--DRVQM-R------E--KQ--QQKIE-Y--Q-GLP-K-T-----S---------NE.G-L--A-----PT-T---E-R--AGV----GL---------EV-.--Q---TR---DA--SD-------------------IR---N-LAD--PE--------  599
CPZ.CM.-.SIVcpzDP943        K-----IK-A------VAM-Q---M--E-------D---------E-E----V----NS------F--Q.H-----------Q-S-----IR--A-V----VA-------------.--R--V---V----S---------D-------------X--T---P---------  606
CPZ.CM.-.SIVcpzMB897        K-----I---------VX--Q--------------RD----A------E------------------------Q.H----------K--A--S---------V---VAL-----------.--R---------A--D-----------Y-------------T-------------  607
CPZ.CM.01.SIVcpzCAM13       K-----I----R--D-V---P------------VST--------N-D-E----V-----E--------F--Q.HR----------Q-ST----IR--A-VI---AV------------.R-R--V---S-----A-------------I----------S--TD--PT-D-Y----  598
CPZ.CM.05.SIVcpzEK505       K-----I--A------VT--R------------------------A--N-D-E---------------------DL.H----------K--ST----IR----V---VAL-----------.--R--V---V---------------D-----------------T---S----Y----  605
CPZ.CM.05.SIVcpzLB7         K-I-R----V-NS---VTF-R-------------------N----------A------V----S---------Q.-----------QGS----------A-VM---AI--------I-.--R------A--A---D------------Y----SH--F-----Q---P---------  602
CPZ.CM.05.SIVcpzMB66        K-----I---------VTF-T--------------------------A------E--------------------F--Q.Y-----------S----------V---VAL--------V-.R-R---------A--D-----------Y--------------QD--P---------  602
CPZ.CM.05.SIVcpzMT145       K-----I--A-G----VA--P--------Q---------A----E-E------------------F---.H---------GKQ-A-----IR--A-T----A------------.--R--V---M--A--AD-----------I-------------T---E----Y----  598
CPZ.CM.98.CAM3              K-----IK------DIVTM-Q---M--E--X---XD---------E-E----X----NS------F--Q.H-----------Q-S-----IR--A-V----A-------------.--R--V---V----S--------D-X-----------------T---P---------  604
CPZ.CM.98.CAM5              K-----IK--------VAM-Q---M--E-------D---------E-E----V----NS------F--Q.H-----------Q-S-----IR--A-V----A-------------.--R--V---V----S---------D-------------------P---P---------  600
CPZ.GA.88.GAB1              K-----I----K--D-V---P------------VST---------D-E---------NC------F---.H-----------Q-S-----IR--A-----A-------------.--R--V---S--A--A-------------I----------S--T---PTTD-Y----  597
CPZ.GA.88.GAB2              K-----I--A-----IVEM-R------E-------------T----E-E----V---NN-------F--T.GR---------Q-A-P---IR--A-V-P-VSP---IL--TV-.------V-R-V--A---D--------------R------------------Q----Y----  602
CPZ.TZ.00.TAN1              KE----I--V-G---PVEM-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--------  595
CPZ.TZ.2001.TAN2            KE----X--VXG--DPVE--R------E--KQ----N-Q-A----KLP-Q-AV-----------------E.G---------KLP-T---EIR--AGLI---GN--------IV-.--L--VS----SQ---D---V--V-----I-----IR---N-LTD--PE--------  596
CPZ.TZ.2002.TAN3_1          KE----I--V-G--DPVEM-R------E--KQ----E-Q-A----KLP-Q-AV-----------------E.G---------KLP-T---EIR--AGLI---GN---I---IV-.--L--VS----SQ---D---V--V-----I-----IR---N-LTD--PE--------  596
CPZ.US.85.CPZUS             K-----IK-A------VNF-H---M--E------------------E-E-V--V----RS------F--R.H-----------Q-S-----IR--V-V----A---------V-.----LV---V-----S---------D--------------K--T-A-E---------  601
                                                                                                                                   p51 RT end p15 RNase H start
MAC.US.x.239                KH--R-I--KMT---EVQW--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--  610
H2A.DE.x.BEN                KH--R-I--KMT---EVQW--L--A--E--KI--SQEQE-Y--QEE-E-E-T---SQGH----K-H--E..-I--V----KIKNT---G-RL-AQV----GK-AL----RI-.--H--VER----Q--DN---V------D--S-----R-TFN-VGD--P------T--  623
H2A.GW.x.ALI                KH--R-I--KMT---GVQW--L--A--E---I--SQEQE-H--QEE-E-E-TV--DQDN----K-H-GE..-I--VE---K-KNT---G-RL-AQV----GK-AL----RI-.R-H--VER----Q--DD---V----D-D--S-----R-AFN-V-D--L------T--  623
H2A.SN.x.ST                 -N--R-I--KMT---EVQW--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--  623
H2B.CI.x.EHO                -HI---I--KMT---EVQW--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--  623
H2B.GH.86.D205              -HI---I--KMT---EVQW--L--A--Q--KI--EQEQE-S--KERVP-E-TV--NLAN----K-H-GN..-V--V----KVKNT---G-RL-AHV----GK-AL----EI-.V-H--VER---DQ---D---V------D--S----IR-A-N-V-D-LE-R--Y-T--  622
H2G.CI.92.ABT96             KN-X--I--XMT---EVQW--L--A--Q--KI--XQXQE-A--KEXEPXE-TV--NLDN---KVH-GN..RI--V----KIKNT---G-RL-AHV--X-GK-AL-----L-.F-H--VERD---Q---D---V----D-D--S--H--R-V-N-V---LEQT--Y-T--  606
H2U.FR.96.12034             KH----IK-KMT---EVQW--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--  619
COL.CM.x.CGU1               KEIS-CTK-K-H-L-EVQ-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--  579
DEB.CM.99.CM40              KH--RAI--VPG--DPVT-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----  604
DEB.CM.99.CM5               KH---AI--VPG--DKVS-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----  604
DEB.CM.SIVdeb04CMPF3061     KH---SI--VPG--DPVK-----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----  607
DEN.CD.x.CD1                KE---SI--V-D-G-T-S---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--  626
DRL.x.x.FAO                 KE----I--A-P-D-EVEW-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----  613
GOR.CM.2004.SIVgorCP684con  KE----I----S----VT-SK------E----------------Q-E-E--VNV---EE-------F-DE.H----------Q-AT----IR--A-VI--VSQ--------L-.--R---CRS--DAF--D----S-------IS----I-----------T---------  591
GOR.CM.2007.SIVgor2139_287  KE----I----S----VT-SK------E----------------Q-E-E--VNV----E-------F-DE.Y----------QKAT----IR--A-VI--VSQ--------L-.--R---CRS--DAF--D----S-------IS----I-----------P---------  591
GOR.CM.2007.SIVgorCP2135con KE----I----S----VT-SK------E----------------Q-E-E--VNV----E-------F-DE.Y----------QKAT----IR--A-VI--VSQ--------L-.--R---CRS--DAF--D----S--------S----I-----------P---------  591
GRV.ET.x.GRI_677            KHT-AM---K-N-L-E-VW-----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----  621
GSN.CM.99.CN166             K----CI--A-G--DQVE-----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----  610
GSN.CM.99.CN71              K---RCI--A-G---QVE-----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----  603
LST.CD.88.447               KE----I--V---D--VTW-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----  616
LST.CD.88.485               KE----I--V---E--VTW-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----  616
LST.CD.88.524               KE----I--V-P-EQLVTW----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----  617
LST.KE.x.lho7               KE----I--V-P-E-I-NW----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----  616
MAC.US.x.251_1A11           KH--R-I--KMT---EVQW--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--  610
MAC.US.x.251_BK28           KH--R-I--KMT---EVQW--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-SQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QV-.--H--VERDV--Q---D---V------D-IS-----R-VFN-V-D--E-E--Y-T--  606
MAC.US.x.EMBL_3             KH--R-I--KMT---AVQW--M--A-YE--NI--SQEQE-C--QEG-P-E-TVI-SQDN----K-H--D..-I--VR-F-KIKNT---G-RL-AHVI---GK-A---V-QV-.--H--VERDV--Q---D---V------D-IS-----R-VFN-V-D--EVE--Y-T--  606
MND-1.GA.x.MNDGB1           K-----I--GLNI--KVTM----R--YEQ-K---A-EQE-S----N-E-YVRF--TTG-DISF-WK-G..N-V-RA---GKQKT--S--LMK-AG-T--VGR-------FV-.-MQI-TTR-I--D--H----C-----V--IS--M-ERE--S-SP--LE-V--Y----  608
MND-2.CM.98.CM16            KE--R-I--A-P-D-KVEW-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----  611
MND-2.GA.x.M14              KE----I--A-P-D--VEW-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----  612
MND-2.x.x.5440              KE----I--ARP-DKIVEW-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----  612
MNE.US.x.MNE027             KH--R-I--KMT---EVQW--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----  606
MON.CM.99.L1                KN---SI--ARS--DTVV-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----  606
MON.NG.x.NG1                KH--XCIX-X----DIVE-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----  609
MUS-1.CM.01.1085            K----CI--V-S---EVQ-----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----  600
MUS-1.CM.01.CM1239          K----CI--V-S---EVQ-----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----  600
MUS-2.CM.01.CM1246          KH--RCI--I-G---LVE-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----  605
MUS-2.CM.01.CM2500          KH---CI--V-----TVE---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----  604
OLC.CI.97.97CI12            KA----IK-P-G-LDKVEW----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-------  598
RCM.CM.SIVagi-00CM312       KE----I--I-R-DDKVEF-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----  612
RCM.CM.SIVrcm-02CM8081      KE----I--V-R-D-KVEF-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----  618
RCM.GA.x.GAB1               KE----I--A---D-KVEM-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----  612
RCM.NG.x.NG411              KE----I--V---DDKVEF-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----  617
SAB.SN.x.SAB1C              KH--R-I--ARP---IVQW-------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----  631
SMM.SL.92.SL92B             KN---MI--KMT---EVQW--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--  603
SMM.US.x.H9                 KH--R-I--KMT---EVQW--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-------R  606
SMM.US.x.PGM53              KH----I--KMT---EVQW--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---------  606
SMM.US.x.SIVsmH635F_L3      KH--R-I--KMT---EVQW--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---------  606
STM.US.x.STM                KH----I--KM----EVQW--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-------  606
SUN.GA.98.L14               KE----I--N-P-A-KVKM----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----  616
SYK.KE.x.KE51               KA-S-CI--A----DIVE------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--------  638
SYK.KE.x.SYK173             KE---CI----Q----LT---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------  633
TAL.CM.00.266               KA--RY---V-G-L-EVQ-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----  607
TAL.CM.01.8023              KA---C---V-G-L-EVT------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----  608
TAN.UG.x.TAN1               KA-------AQD--A-VQW--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--  618
VER.DE.x.AGM3               KN----I--K-N-L-TVTW-----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----  615
VER.KE.x.9063               K-I-R-I--K-N-L-EVQW-----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----  615
VER.KE.x.AGM155             KNI------K-N-LD-VEW-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----  618
VER.KE.x.TYO1_patent        KNI---I--K-N-L-LVTW-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----  618
WRC.CI.97.97CI14            K---R-I--N-P---EVEW----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--  622
WRC.CI.98.98CI04            ------I--N-P---EVTW----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--  621
WRC.GM.05.Pbt_05GM_X02      KE----I--N-P--AEVEW-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--  613
```

p66 RT, p15 Rnase H end p31 Integrase start

```
H1B.FR.83.HXB2             AANRETKLGKAGYVTNRGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAG.IR..KVLFLDGID.KAQDEHEK.YHSNWRAMASDFNLPPVVAKEIVASCDKCQLK.GE  763
H1A1.UG.85.U455            ---------------D-------S--E--------H--H------S-------------------R----I-----K--E------S--------------------S-.--..----------.---ED---.--C----------------------N------.--  762
H1B.US.90.WEAU160          ---K---------------R--S-------------H---------------------------------S-------------------------------------S-.--..----------.---E----.----------------------------------.--  763
H1C.ET.86.ETH2220          -------I-------D-----I-S--E-----------Q-------S-------------L----K----I--------S--R---S--------------------S-.--..----------.---E----.---------NE--I----P-----C-------.--  770
H1D.CD.84.84ZR085          ---K-----------D-------PF-------------N-----------------------K------S-------------------------------------S-.--..----------.---E----.----------------------------------.--  763
H1F1.BE.93.VI850           -S-----K-------DK-K----S--E-----A-------------S-----------------K----I---------Q--R---S---------------------.V-..-I--------.---E----.--N------------I-----------------.--  764
H1G.SE.93.SE6165           ---------------DK-K--II---E-----A-----Q-----RS------------------R--A-----------------S--------------------S-.--..----------.---E---R.--N------------I-----------------.--  760
H1H.CF.90.056              ---------------D--K----S--E------------------------------------K---------E--------S--------------------S-.V-..----------.---E---R.--N----V--------I-----------------.--  762
H1J.SE.93.SE7887           -S-----T-------DK-----------------H------R----------------------K---------E--------S--------------------S-.--..----------.---ED---.----------------------------------.--  761
H1K.CM.96.MP535            --H----K-R-----D-------SI-E-----A-----C------S------------------K---D------------RI--S-----------------------.--..----------.---E----.--N------------I-----------------.--  763
H101_AE.TH.90.CM240        --S------------D-------S--E--------H--H------S-------------------R----V------E--------S--------------------S-.--..----------.---E---R.------T---------I------TN---------.--  768
H102_AG.NG.x.IBNG          -------I-------D-------S--E--------H--H------S-------------------R-----------K----D----S--------------------N-.--..----------.---E---R.-----K----------I-------------M-.--  763
H103_AB.RU.97.KAL153_2     -------S-------D-------S-------------H--------------------------K------S----------------------------------------.--..E---------.---E----.--G-------------------------------.--  763
H104_cpx.CY.94.CY032       -------Q-------D-------S-SE------------------S-----------I--------R---D---------R-D----S--------------------N-.--..----------.---E----.--N-----------S-----------N------.--  762
H1N.CM.95.YBF30            ------------F--D-------SIA------A-----LM---E--RD---------M---HS---K------S----E-----R---S--------------------S-.--..-I------E.---ED-DR.-----K----------I-----------------.--  770
H1P.FR.06.RBF168           ---------------D-----IIK-EG-----A-XE-VL---KE--KQA--------V----S-T----D-P--QK---EMT--------S--------------NI-----KD.--..R----E---.Q--ED---.-------L----G---I-----IN--P---V-.--  759
CPZ.CD.90.ANT              ----NSQ--------D---SR-KH-QK----QA-----LM--E--TGP-----------VL-GT------P--EE--QK---R-QI--S--------------------Q-.--..Q----E---.---ED-D-.-------SL-DEY----I-----I-Q----HV-.--  763
CPZ.CM.-.SIVcpzDP943       ----D--K-------D-----I-N-E------A--T-V------AEHI---------V----HS-----------L---E------S--S--------------------S-.--..R---------.---E----.--N----------------------------------.--  770
CPZ.CM.-.SIVcpzMB897       ---------------D-----IIS--E----RA-------------EV-------------------S---I---DI--------V----I--S--------------I-----------.--..----------.---E----.--N--K---DE-----I-----------------.--  771
CPZ.CM.01.SIVcpzCAM13      -------K-----I-DK-K--IIS-EN----QA--G-L-------EQQ---------V-----S---H-----------E----D----S--------------------S-.--..----------.---E---R.-----K-L-GE-----I-------N-----V-.--  762
CPZ.CM.05.SIVcpzEK505      ------------F--D------TSISE----QA----VLM----A-Q----------V----HS---K-----------E-----RI--S--------------I-----T-.--..----------.---E---R.-----K----------I-----------------.--  769
CPZ.CM.05.SIVcpzLB7        ---------------D--K--IIP--E-----A-----Q-------P--------------L-------D---------N--------T--------------------S-.--..----------.R--E----.--N------------I-----------------.--  766
CPZ.CM.05.SIVcpzMB66       ---------------DK----IIS--E-----A-----Q------EV----------------G---T----I------E--------S--------------I-----S-.--..----------.---E----.--N------------I-------N---------.--  766
CPZ.CM.05.SIVcpzMT145      -------R-------------IHS-E------A----VLM------Q---V------V---L-SH--H---DI------E------I--S----R---------------S-.--..---------E.---E---R.-----K---Q------I-------Q---------.--  762
CPZ.CM.98.CAM3             -------K-------D-----I-N-EN-----A--T-V------AEHI---------V----HS-----------L---E------S--S------V-----------S-#-..---------#-HKE---E#--N----------I-HI----L--R---------.--  768
CPZ.CM.98.CAM5             -------K-----I-D-----IIN-EN-----A--T-V------AEHI---------V----HS---------GS---E------S--S--------------------S-.--..----------.---E----.--N--K-------------------------------.--  764
CPZ.GA.88.GAB1             -------T-------DK-K--IIS-EN----QA--K-LL------DQQ---------V-----S---H-----------E------I--S----------------------.--..----------.R--E---R.-----K----------I-------H-----V-.--  761
CPZ.GA.88.GAB2             ------RR------I-N-K---IS-EE-----A--E-LRI------Q---V------V----NS-----------I---E--------T--------------------S-.--..----------.---ED--R.---------DE-Q-----------T---------.--  766
CPZ.TZ.00.TAN1             ----DS-K-R---------YRSKD-EN----QA--W-VD---K---AQ---------VM-VL-GL----D-PI-E---QK-TQ-TAI---------------E------KN.--..-I------N.E--ED-D-.-----K-L-DEY----------I-Q-P--HI-.--  759
CPZ.TZ.2001.TAN2           -----S-K-R-----D---YRTKD-EN----QA--W--D---R---AQ---------VI-VL-GL----D-PI-E---Q--TQ-TAI---------------E------KN.-X..-I------N.E--ED-D-.-----K-L-DEYH---I-----I-Q-P--HI-.--  760
CPZ.TZ.2002.TAN3_1         -----S-K-R-----D---Y-TRD-EN----QA--W--D---R---AQ---------VM-VL-GL----D-PI-E---Q--TQ-AAI---------------E------KN.--..R-------N.E--ED-D-.-------L-DEY----I-----I-Q-P--HI-.--  760
CPZ.US.85.CPZUS            --Q----K-------D-----II--EN-----A--T-V----K--ENT--V--G---V----HS---------------E------S-IS--------------I-----S-.--..----------.---ED-D-.-----T-----------------------P-.--  765
```

p15 RNase H end p31 Integrase start

```
MAC.US.x.239               SC-KQS-E-----I-D--KD--KV-EQ----QA--E-FLM--T---PKA--IV----VM---TGC-TE---R-------EM---SEI-V-----------QEI-H---Q-.--..Q----EK-E.P--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.--  774
H2A.DE.x.BEN               SC--QS-E-------D--KD--KV-EQ----QA--EVFRM--A---PK---IV----VM--VAG--TE--NRI-------EM----A--V-----------QE--H---Q-.--..Q----EK-E.P--E----.---IIKELTHK-GI-LL--RQ--N--AQ--Q-.--  787
H2A.GW.x.ALI               PC--QS-E-----I-D---D--KV-EQ----QA--E-FA--VT---PKA--IV----VM--VAG--TE--NRI-------EM----AI-V-----------QE--H---Q-.--..Q----EK-E.P--E----.----VKELSHK-G--NL--RQ--NT-AQ--Q-.--  787
H2A.SN.x.ST                SC-KQSRE-----I-D---D--RL-EQ----QA--E-FAM-VT---PKA--IV----VM--VAG--TE---KI-------EM----AI-V-----------QE--H---Q-.--..Q----EK-E.P--E----.----VKELSHK-G--KL--RQ--NT-TQ--Q-.--  787
H2B.CI.x.EHO               SC-KAS-E-------D--KD--KP-EQ----QA--E-FA-------PQ---IV----VM--VA---TET--PI-RE---EM-----I-VG-------L---QE--H---Q-.--..QI---EK-E.P--E----.--N-VKELVHK-GI-QL--RQ--N------Q-.--  787
H2B.GH.86.D205             SC--TS-E-------D--KD--KV-EQ----QA--E-FA---T--EPQ---IV----VM---A---TET--PI-AK---EM----A--VG-------L---QE--H---Q-.--..Q----EK-E.P--E----.--G-VKELVHK-GI-QL---Q--N------Q-.--  786
H2G.CI.92.ABT96            SC-KTS-X-------D--KD--XV-EQ----QX--E-FAM------PK---IX----VMX--AX--TET--P---K---EX----XL-VG-----------QE--X---Q-.--..Q----XK-E.P--E---X.F---VKELTHK-GI-QL---Q--N-YHS--Q-.--  770
H2U.FR.96.12034            SC--AS-Q-------D--KDR-KP-EQ----QA--E-FAM--E---PK---------IM---AG--TET--P--SK---E--R--A--VG-----------QE--H---Q-.--..Q----EK-E.P--E----.----VKELQHK-GI-QL---Q--NT-HV--Q-.--  783
COL.CM.x.CGU1              GCS-KLGA----WINSK-EEE--N-HEGS--QA--TGVL---KHGPKRM-L----L------TG--YDQLPSSRGE--QAGMA--AIHV--C----------LI-QK-..-.V-..Q-MWI-K-E.A-EED-Q-.F---VQYLKEQ-G--T------WER-SE--N-.-Q  741
DEB.CM.99.CM40             -----S---------EW-K-S-KC-EN-----A--E--L---EEGPSKM------------LEH-SET-HKI-EKV-QA-QG--QI--------------D-------K-.--..RI---ERVE.E--ED---.--G--KQLRDEYQ--TLI--Q-ITQ-P---TR.--  768
DEB.CM.99.CM5              -----S-M-----ISEW-K-G-KS-EN-----A--E--L---EEGPSKM-----------LLEH-TET-H-I-EK--QA-QE--Q---N-------------D-------R-.--..RI---ERVE.E--ED---.--G--KQLRDE-Q--TLI--Q-I-Q-P---I-.--  768
DEB.CM.SIVdeb04CMPF3061    -----S-E-----ISE--K-S-KA-EN-----A--T-VLM--E---EK---------V-N-LTEH-TTT-H---EK--Q--Q--QE---Q-------L-----I-----K-.--..-I---ERVE.E--ED-D-.-----KQLREE----TL---Q-INQ-P---IH.--  771
DEN.CD.x.CD1               ----Q--Q-R----ATN-K------EE-----A--IGVLM-----PNK--------------AGT-TT--HQI-E---QAMQG--EI-IT-----------TEI-----K-.--..QI---EQ-P.Q--E--D-.-----KEIRDKYG--TM-----INV-P---TH.--  790
DRL.x.x.FAO                ----VS-E-------S--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.--  777
GOR.CM.2004.SIVgorCP684con -----S-Q-------D-----IRK-EG-----A--E-VL---EE--KN---------V----A-S----D-P--QK---EMTT-------------------EI-----KD.--..R----E---.Q--ED---.-------L-----X---------IN--P--HV-.--  755
GOR.CM.2007.SIVgor2139_287 -----S-Q-------DK----IRK-EG-----A--E-VL---EE--KN---------V----A-S----D-P--QK---EMTT-------------------EI-----KD.--..R----E---.Q--ED---.------TL----G---------IN--P--HV-.--  755
GOR.CM.2007.SIVgorCP2135con -----S-Q-------D-----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-.--  755
GRV.ET.x.GRI_677           --SKV--------LSE--KSRIRE-EN----QA--T-VKM--E---EN---------VMN-LT-C-QE-N-P--E---QA-M--RQ---Q-----------TEI-----K-.--..QI----R-E.E---D-A-.--N---S-VQE-G--NI-------A-P---IR.--  785
GSN.CM.99.CN166            ----TS----------T-KHRAIE-EE----QA--H-VL---KEGPPKM-L------VM--L-S--EV-T-A--E---QE-L---AI--S-----------TE------K-.--..Q----EN-E.P-VED---.-----KYLRDQYKI-ALL-----NK-S---VH.--  774
GSN.CM.99.CN71             ----TS----------T-KYRAIE-EG----QA----VL---KEGPPRM-L------VM--L-S--EV-T-P--E---QE-LG--A---S-----------TE------R-.--..Q---MEN-E.P-VED---.-----KYLRDQYKI-ALL-----NK-S---IH.--  767
LST.CD.88.447              --EKVG-T-------QS-KE--KE-------QA--E-VL---R--KSR---------VIKVLSQR-TET-HPV-K----ECK--DQ---G-----------QE--H---K-.--QKQ----GK-E.P--E-YS-.F-N-AKDLEEK-HI--M---Q--N--N---T-.--  782
LST.CD.88.485              --EKVG-T-------QS-KE--KE-------QA--E-VL---K--KSR---------VMKVLSQR-TET-HPV-K----ECK--DQ---G-----------QE--H---K-.--QIQ----EK-E.P--E--S-.F-N-AKDLEEK-H---M---Q--N--N---T-.--  782
LST.CD.88.524              --E-VG-T-----I-QS-KE-IKE-Q-----QA--E-VL------KST---------VMKVLSQR-TET-HPI-KN--KE-Q--DA---G-----------QEI-Q--IK-.--TKQ-M--EK-E.P-VE--G-.F-N-AKDLEEKYGI--M---Q--NE-A---Q-.--  783
LST.KE.x.lho7              --EKVG-T-----I-QS-KE--KE-------QA--E-VLM-----NSK---------VMK-LSQR-TET-HPI-KD----CKQ-DQ---G-----------QE--H---K-.--QKQ-M--EK-E.P-VE--S-.F-N-AKDLEEK-----M---Q--ND-AN--K-.--  782
MAC.US.x.251_1A11          SC-KQS-E-----I-D--KD--KV-EQ----QA--E-FLM--T---PKA--IV----VM---TGC-TE---R------KEM---SEI-V-----------QE--H---Q-.--..Q----EK-E.P--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.--  774
MAC.US.x.251_BK28          SC-KQS-E-----I-D--KD--KV-EQ----QA--E-FLM--T---PKT--IV----VM---TGC-TE---R-------EM---SEI-V-----------QEI-H---Q-.--..Q----EK-E.P--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.--  770
MAC.US.x.EMBL_3            SC-KQS-E-----I-D--KDI-KV-.T----QA--E---HGIE---PKR--IVEL-VCY-NNNRF-TE---R-------EM--.VR--V-----LE-----QEIGP---Q-.F-..Q----EK-E.P--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.--  768
MND-1.GA.x.MNDGB1          ----DS-M-----I-D--F-R-EEYLN----Q---H-VK---E---SY---------VV--LASR-TETDHPI-KE---LMKG---I--S-L----------I-----S-.--..-----QN-E.P--E----.----EAQLREK-H--AL---Q--Q--S--CHH.--  772
MND-2.CM.98.CM16           ----NS-E-------A-NKSR-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-.--  775
MND-2.GA.x.M14             ----NS-E-----Y-A--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-.--  776
MND-2.x.x.5440             ----NS-E-----Y-S--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-.--  776
MNE.US.x.MNE027            SC-KQS-E-----I-D--KD--RV-EQ----QA--E-FLM--T---PKA--IV----VM---TGC-TE---R-------EM---TEI-V-----------QEI-H---Q-.--..Q----EK-E.P--E--D-.----VKELVFK-G--RL---Q--DT----HQ-.--  770
MON.CM.99.L1               ----TS-E---------F-AR-A---EQ-----A--E-VL---R-GPPQM--I-----V---LASC-EI---PI-EA--QE-L--DQ-F-S-----------E--R--GQ-.--..Q---MENLE.P-RED---.-----KYLRDTYHI-TLL-----NH-H---NH.--  770
MON.NG.x.NG1               ----AS-E-------ST-QR--IP-EQ-----A--E-VL----NGPASM--------VM--LASK-EV---PI-EK--Q--LG-KQ---S-----------EX----XQ-.--..Q---MEN-E.X-RED---.-----KYLRDTYGI-TLL-----NH-P--HTQ.--  773
MUS-1.CM.01.1085           ----VS----------T-KE--IS-ES-----A--E-VL---KEGPPRM-L------V---VAS--QE-T-P--E---QE-LS--AI--S-----------E-----NS-.--..Q---MEQ-E.P-KED---.-----KYLRDKYQI-ALL-----NH-P---TH.--  764
MUS-1.CM.01.CM1239         --HKVS----------T-KE---S-EN-----A--E-VL---KEGPPSM--------V---VAS--QE-T-P--EE--Q--LT--A---S-----------D------H-.--..Q---MEQ-E.P-KED---.-----KYLRDKY-I-ALL-----NL-P---TH.--  764
MUS-2.CM.01.CM1246         ----NS---R-----D--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.--  769
MUS-2.CM.01.CM2500         ----NS---R-----D--QE-AID-EN-----A-----LM--T-GPPKM--------------T-A-ET---P--EK--Q--LT--AI--S-----------QE--H---K-.--..Q---MEN-E.P-IED-D-.-----KYLRDKY-I-TLL-----NR-SA--TH.--  768
OLC.CI.97.97CI12           -Y--KS-E------SSG-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-E.P--ES-N-.--Q-AQSLQQE-QI-LI---R--DE-EE-KG-.-V  764
RCM.CM.SIVagi-00CM312      ----NS-------I-D--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-.--  776
RCM.CM.SIVrcm-02CM8081     ----NS-Q-----I-D---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-.--  782
RCM.GA.x.GAB1              ----NS-I----------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-.--  776
RCM.NG.x.NG411             ----NS---------D--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-.--  781
SAB.SN.x.SAB1C             ----NS-E-----L-D--D----A-EN-----A--E--L---R---SK---I------M---AGE-TE-DNNI-Q----E-----A--I--------V----EI-----Q-.--..Q-----R-E.E--E--D-.--A---S-QQE-G--AI-------A-P---I-.--  795
SMM.SL.92.SL92B            SC--NS-E-------D--KE--LP-EQA---QA----LL---K--PSK--V------V-N--TG--SE-D-DI-A------VQ--A--IG------------NE--R---Q-.--..Q----ES-E.P--ED-D-.----VKEL-QKY-I-QL---Q--NA-N---Q-.--  767
SMM.US.x.H9                SC--QSRE-------D---G-TKL-EQ----QA--E-F----A---PKA--IV----VM--VAG--TE---R-------KM----AI-V-------X--E-QE--H---QE.--..Q----EK-E.P--E----.----VKKLVFK-X--RL---Q--DT----H--.--  770
SMM.US.x.PGM53             SC--QSRE-------D---D-AKP-EQ----QA--E-F----A---PKA--IV----VM---AG--TE---R-------EM----AI-V-----------QE--H---Q-.--..Q----EK-E.P--E----.----VKELVFK-GI-RL---Q--DT-H--HQ-.--  770
SMM.US.x.SIVsmH635F_L3     SC--QSRE-------D---D-AKL-EQ----QA--E-F----A---PKA--IV----VM--VAG--TE---R-------EM----AI-V-----------QE--H---Q-.--..Q----EK-E.P--E----.----VKELVFK-GI-RL---Q--DT--R-HQ-.--  770
STM.US.x.STM               SC--QS-E-----I-D--KN--KA-EQ----QA--E-FAM--A---PKA---V----VM---TG--TE---K-------EM----AI-V-----------QE--H---Q-.--..Q----EK-E.P--E----.----VKELVFK-GI-RL---Q--DT----HQ-.--  770
SUN.GA.98.L14              --EKLE-R-----I-QG--SR-KK-EN----QA--E--KM--E--RSS-----------RLLSKR-TETD----KE-V-LIRQ-DQ---G-----------QEI-Q---Q-.--KRQ-M-IEK-E.P-VE--G-.F-N-AASLQEM-DI-L----Q--NE-AQ--Q-.--  782
SYK.KE.x.KE51              ----NS-E-Q---I-D--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.--  802
SYK.KE.x.SYK173            ----DS-T-N----ASD-T-R-QY-EQ----QA--EGLLM-----KDK----V----SY--LMTC-TNT-HPI-E---QEA----AI-VT-----------A------K-.--..-I---ER-P.Q--ED--R.----MEYLRQE-H--RQ---A-IQQ-P---NR.--  797
TAL.CM.00.266              --H--S-E-------AT-KEH-IK-EK-----A--E-VL---K---PK---------VY--LAGS-TETDNKIIED--QL-LS--A-----L------------------Q-.--..-I--VEQ-P.E--E---R.--N---DLKAR-K--TI---A-IEA-P---VQ.--  771
TAL.CM.01.8023             --H--S-E-------TS-KEH--G-SN-----A--E-VL---KH--PK---------VY--LVGN-TETDNKI-ED--Q--LG--A---T-L-----------------Q-.--..RI--VEQ-P.E--E---R.--N--KDLKAR-K--TI---A-IEA-P---VQ.--  772
TAN.UG.x.TAN1              -C--NS------FL--T-ESQ-IE-E-----QA----VL---KH-KSK---------VM--LT---T--D-V-------EM-N-DAI--T-----------QEI-----Q-.--..Q-----RME.E--ES-D-.--T--QFIRDA-GI-AL-------A-P---IR.--  782
VER.DE.x.AGM3              -C--NSRE-----I-QY-K-R-EK-EN----QA--M--KM--E---PN----------M--LT---T--D-P-IE---ALMVQ-HQI--Q----D-------EI-----Q-.M-..-I---EK-E.E--E---R.--N---NL-DTYG--QI-------M-P---I-.--  779
VER.KE.x.9063              -C--NS-E-----I-QY-K-R-EK-EQ----QA--V--KM--E---EK----------M--LT---T--D-P--E---ALMV--RAI--Q------------EI-----Q-.--..R---IGR-E.E--E--DR.-----KNL-DT-G--QI-------M-PNAKI-.--  779
VER.KE.x.AGM155            -C--QS-E-----I-QQ-K-R-QQ-EN----QA--T--KM--E---PK----------M--LT---T--D-P--E---A-MVQ--AI--Q------------EI-----K-.V-..RI--IGR-E.E--E--DR.------NL-DT-G--QI-------M-P---V-.--  782
VER.KE.x.TYO1_patent       -C--NS-E-----ISQY-K-R-E--EN----QA--T--KM--E---PN----------M--LT---T--D-P--E---ALM-Q-QQI--Q------------EI-----K-.--..R----EK-E.E--E---R.--N--KNL-DTYG--QI-------M-P---I-.--  782
WRC.CI.97.97CI14           S---DS-E-----.WTKSKE---SVQN----Q---I-LK---E-G-QKM--I-----VM--VKDH-TM---P--QE---L-KG--EI-ID-----------TEI-R------.--S.-IM--EN-T.P-IES-T-.--QG-K-LQEE-H--QLI--R-IEE-ST--G-G-A  787
WRC.CI.98.98CI04           S---DS-E----F.WTNSKE--IN-YN----QA--A-LK---E-G-SKI--I-----VM--VKDH-TI---A--QE--DL-KQ--EI-IE-----------TEI--------.--S.--M--EN-T.P-IES-T-.--QG-K-LQEE-HI-QLI--R--EE-ST--G-G-A  786
WRC.GM.05.Pbt_05GM_X02     S--K-S-E-----.WTKT-N-IQKI-N----Q---I-LRM--E--SDK---I-----V----KDH-T----P-MQE--DL-KG--EI-IT-----------TEI-A------.--S.R-M--EN-T.P-IES-T-.--QG-G-LEEE-H--TIL--R--EE-AT--G-G-H  778
```

```
H1B.FR.83.HXB2                AMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTI.HTDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGER.IVDIIATDIQTKELQKQ  931
H1A1.UG.85.U455               ------------------------------------------------------I---------V-.---------S-A-K-V----N-Q----------------------------------E--------------------------------------.-I---------------  930
H1B.US.90.WEAU160             -----------------------I------------------------------I----------.---------ST--K--------------------------I--------------------------------------------------------.-I---------QQ----  931
H1C.ET.86.ETH2220             -I----N----------------I---------------------------------------RV-.---------SNA-K-------Q----------------------------------E-------------------R-----------.-I----S--------N-  938
H1D.CD.84.84ZR085             --------------------------------------------------A-------------VV.---------S---K---------------------------------------------------------------------------------.-I----S----R-----  931
H1F1.BE.93.VI850              ---------------------------------E-----------------I---------I-.---------S-A-K-S------Q----------------I------------------------------------------------------.-I---S-----R-----  932
H1G.SE.93.SE6165              -----------------------I-I----------------------------I---------TV-.---------S-A-K-----N-T---------------------------------------------------------------------.-I----S----------  928
H1H.CF.90.056                 ----------------------Q----------------------K---------S----V-.---------S-A-K-----D-Q-------------------------------------------------------------------------.-I---------------  930
H1J.SE.93.SE7887              ------------------------------------------------A-F-I---------V-.---------SGA-K-----D-----------------------------------E--------------------------------------.-I---------R-----  929
H1K.CM.96.MP535               -I---------------------I------------------------------I---------V-.-----T---STV-K-------V-----------------------------------------------------------------------.-----------------  931
H101_AE.TH.90.CM240           ------------------------------------------------------------V-.---------S-A-K-----NVQ-----------------------------------E--------------------------------------.-I---------------  936
H102_AG.NG.x.IBNG             --------G--------------I------------------------------I---------V-.---------S-A-K-----NVT---------------------------------------------------------------------.-I----S----------  931
H103_AB.RU.97.KAL153_2        -----------------------I------------------------------V---------I-.---------STA-K-------------------------Q--QT---------------------------------------------------.-I---------------  931
H104_cpx.CY.94.CY032          -------------------------M----------------------------I---------M-.-A---P---S-A-K-----D-N-----------------------------------------------------------------------.-I----S----------  930
H1N.CM.95.YBF30               ------N----V-----------I----------L-------------------I---------V-.---------S---K-----N-------------A----------I---------------------------T----.-I---------TN--T-  938
H1P.FR.06.RBF168              -------------------T---I----------FL------------------V----A----V-.-----P---S-A-K-----LN-TH-------------------L-------------------V-----------T--D-.-I--M--Q---T-----  927
CPZ.CD.90.ANT                 -R------------V---------I-----S--F-----MAD---KS---------S-------.-----A---S-A-K-----N-Q-----------------Q--Q---I-----Q-----V---H-------------TP-Q-.-L--L------TQ--N-  931
CPZ.CM.-.SIVcpzDP943          -----------M----------------------L-------------------I---------V-.---------SN--K-------Q-----------------------I-E--------------------------T----.-I---------TN----  938
CPZ.CM.-.SIVcpzMB897          -----------------------I------------------------------V---------V-.---------S-A-K-----N-Q---------------------------------------------------------------.-I---------R-----  939
CPZ.CM.01.SIVcpzCAM13         -L---------T--M--------I-I----I-----------------------I----S----VL.-----T---SNA-K-----N---K-------------------E-------------------S------------T----.-I---------------  930
CPZ.CM.05.SIVcpzEK505         -I---IN----V-----------I----------L-------------------I---------V-.---------S---K-----N-Q-------------A----------I---------------------------T----.-I---------TK--T-  937
CPZ.CM.05.SIVcpzLB7           ----------------------------------M-------------------I----S----I-.---------STA-K-----Q-Q---------------------------------------------------------------.-I---------T-----  934
CPZ.CM.05.SIVcpzMB66          -I---------------------I----------M-------------------I---------V-.---------SS--K-------Q------------------------I-----------------------------------.-I--L------TK--Q-  934
CPZ.CM.05.SIVcpzMT145         --------------I--------I-I----------------------------V------------.-----P---SNA-K-------R------------------------E----------------------------------.-I---------------  930
CPZ.CM.98.CAM3                ----------------------------------L-------------------I---------V-.-----P---SN--K-------Q-------------------------I-E-----R-----------------T----.-I---------TN----  936
CPZ.CM.98.CAM5                -----------------------I----------L-------------------I---------V-.-----P---SN--K-------Q-------------------------I-----Q---------------------T----.-I---------TN----  932
CPZ.GA.88.GAB1                --------------V---------I-----------------------------------------.-----P---S-A-K-----D-----------------L--------------------------------------T----.-I---------S-----  929
CPZ.GA.88.GAB2                -I---------------------I----------L-------------------I---------K-.-----T---S-A-K-------------------------------------------------------------T----.---ML---L--Q-----  934
CPZ.TZ.00.TAN1                -I-----Y--E---I----------I--------F-------E---R-----I---------K-.-----P---STA-K-----Q-QH-----------------Q--Q--E-------Q-R---I---Y-----------E-T----.LL--LT-N----Q----  927
CPZ.TZ.2001.TAN2              -I-----Y--E---I----------I--------FL------E---R-----I---------K-.-----P---S-A-K-----Q-QH-----------------Q--Q--E-I-E-----X---X---Y-----------E-T----.LL--LT-N---QQ----  928
CPZ.TZ.2002.TAN3_1            -I-----Y--E---I----------I--A-----FL------E---K-----I---------K-.-----P---S-A-K-----Q-QH-----------------Q--Q----I-E-----R---T---Y-----------E-T----.LL--LT-N---QQ----  928
CPZ.US.85.CPZUS               -I---------------------I-I--------L-------------------I---------V-.---------SS----------Q-------------------------I----------------------------T----.-I----SEL--DL----  933

MAC.US.x.239                  -I---ANSDL-T--M--------I-I--------F-------Q---RQ--L----------ITHL.-----A--ASQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQFQ-SK  942
H2A.DE.x.BEN                  -I----NAEI-V--M-Y------I-I--------F-------Q-S-RQ--L------S---ITHL.-----P---SQE-KMVA--V--E-S--V-----------A--HH--NQ-SRI-E--NTIE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQF--RK  955
H2A.GW.x.ALI                  -I----NAEL-T--M----------II-------F-------Q-S-RQ--L------S---ITHL.---S-V---SQE-KMVA--V--E-S--V-----------A--HH--NQ-SRI-E--NTVE-I-L---HCM----R----DMTPA--.LINM-S-EQEIQF--TK  955
H2A.SN.x.ST                   -I----NAEL-T--M--------I-I--------F-------Q-S-RQ--L------S---ITHL.-----A---SQE-KMVA--I--E-S--V-----------A--HH--NQ-SRI-E--NTVE-I-L---HCM----R----DMTPA--.LINMVTAEQEIQF--AK  955
H2B.CI.x.EHO                  -I----NSEL-T--M----------I--------F-------Q---RQ--L------S---ITHL.-----A---SQD-KM-A--I--E-T--V----E------A--HH--NQ-DRI----VSIE-V-L--THCM----R----DMTPA--.--NM-T-EQEIQF--TK  955
H2B.GH.86.D205                -I----NADL-T--M--------I-I--------F-------Q---RQ--L------S---ITHL.-----A---SPS-KMVA--V--E-T--V-----------A--HH--NQ-DRL----VSIE-V-L--THCM----R----DMTPA--.L-NM-T-EQEIQFF-AK  954
H2G.CI.92.ABT96               -I----NAEL-T--M----------I--------FV------Q---RQ--L----------ITHL.-----A---SQE-KMVA--V--E-A--V--------------HH---Q-DKI----NSIE-I-L---HCM----R----DMTPA--.-INM-T-EQEIQFQ-TK  938
H2U.FR.96.12034               ------NTEV-T--M---------VI--------F-------Q---RQ--L----------ITHL.-----A---SQE-KMVA--V--EHT--V-----------A--HH--NQ-ERI-E--NTME-I-L--AHCM----R----DMTPA--.-INM-S-EQEIQF--SK  951
COL.CM.x.CGU1                 -V---L-Y-Y-L-------E-----------CTLFCW-TILKR---E--GRA-I---SQ-E-RQV.-----P--VSQHFK--V--L--AHTT-H-----------QR--DV-RK-KKMK----T-ESK-A---YAL-------L--K-PW--.Q-ERAIIELD-QN-T-L  909
DEB.CM.99.CM40                PT-----A-M----M------N-I-I-----------TKIL-----K---L---Q-GA---TQ-.-----P--ISQP-K-----L--EHTT-V-----------K-RI--DT-K-I--D--R-E---A--TY-I----R----DM-PI--.L-NM-T-E-E--T--QK  936
DEB.CM.99.CM5                 P------A-M-V--I------N---I--------FM-TK-LS----K---L---Q-GA---I-Q-.-----P--ISQP-K-----L--DHTT-V-----------K-RV--ET-K-I-ED--R-E---A--T--I----R----DMA-I--.L-NM-T--LE-QT--QK  936
DEB.CM.SIVdeb04CMPF3061       PK----NADL----M----------I--------FT-TQ-LKE---KQ--L---Q-GA---IQQV.-----P--ISQAFA-----L--EHTT-V----------NK-RQ--DT-T---ED-QR-E---A--TH-L----R----DMTPS--.--NM-N-ELEI-Y--QK  939
DEN.CD.x.CD1                  PIR----A-L-T--M------N---I---NA----L-TKIL-----K---L---QI-A---ITKL.-----P--ISQAMQ------N--HTT-V--------I--NK--Q--EV-AL--ED--R-E--LA--TL-L----R--L-DMT-A--.YINMLN-HLEIQH..TK  956
DRL.x.x.FAO                   PI-----A---T--M--------I-I--------H-------------EK--------------SHL.-----P---SEK-ATV----K-EHTT-V----------------NQ--------I-----K-E-------L-----------E---A--.------S-LL-SK--QN  945
GOR.CM.2004.SIVgorCP684con    -R---------------------I----------F---------------------A-----V-.-----P---S---K-----LNVTH---------------------------Q------------------V------------T--D-.-I-ML--Q---T-----  923
GOR.CM.2007.SIVgor2139_287    -R---------------------I----------F---------------------A-----I-.-----P---S---K-----LNVTH---------------------------Q------------------V------------T--D-.-I-ML--Q---T-----  923
GOR.CM.2007.SIVgorCP2135con   -R---------------------I----------F---------------------A-----I-.-----P---S---K-----LNVTH---------------------------Q------------------V------------T--D-.-I-ML--Q---T-----  923
GRV.ET.x.GRI_677              PK-----A-IET--M----------I--------F------R---K---H-----LA-----HL.-----P---SQN-A-V---GN-EHTT----------S-----RQ--E--S-I--DC-R-E------TH------------I-SA--.L-NMLT-QLELNT--N-  953
GSN.CM.99.CN166               PK----NAEL-M--M-------------------VW-RI--Q---RQ--LK--E--AT---THL.-----P--VSKELE------D-QHTT-V-----------N---Q--ET-TKI-EEVTY-E---AQ-CY-----------DMCPT--.--NM-H-ELE-QH-NT-  942
GSN.CM.99.CN71                PK----NAEL----M-------------------VW-RI--Q---RQ--LK--E--AT---THL.-----P--ISKELE------N-QHST-V-----------N---Q--ET-QKI--EVTY-E---AQ-CY-----------DMCPT--.L-NM-H-ELE-QH-NT-  935
LST.CD.88.447                 -IT----T------I----M--Q--IN-------FMV-----D---KT--H-----CS----QQV.-----P--VSKE-Q-VI--L--EHTT--------------AK--A--Q--ER--ED-QQ-----L--TH-----QR--L--L-PA--.FINM-NA-LE-QY-.QK  949
LST.CD.88.485                 -IT----T------I----M--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-.QK  949
LST.CD.88.524                 -IT----V-V-----------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.QK  950
LST.KE.x.lho7                 -IT----V-V-----------Q--IN-------FMV-----D---KT-SH-----CS-----Q-.-----P--VSKE-Q-VT--I--EHTT--------------AK--V--S--ER--ED-QQ-----L---H-----QR--L--LTPA--.FINM-NAELE-QY-.QK  949
MAC.US.x.251_1A11             -I----NSDL-T--M--------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-SK  942
MAC.US.x.251_BK28             -I----NSDL-T--M--------IVI--------F-------Q---RQ--L----------ITHL.-----A--ASQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQFQ-SK  938
MAC.US.x.EMBL_3               -I----NSDL-T--M--------IVI--------F-------Q---RQHY.----------YLH-YTHS--A--ASQE-KMVT-----EAHLWV-----------A--HH--NQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQFQ-SK  936
MND-1.GA.x.MNDGB1             PIK--T-A-L-V--I------NQI-I--------FMK----T----KK--E------AQ--ISKL.-----P---SQE-ETM---L--EHT--------------NK--Y--EL-EKI-EDCKE-----A--T------QR--L--MT----.--NM-N-ELEYQYQ-N-  940
MND-2.CM.98.CM16              PI-----T---T--M---M-----IA------R-L------T---K---H-------------HL.-----P---SEK-ATV----Q-EHTT-----------I-AK-HH--Q-----------K-E-------L-----------E--P---IVDI-ATDLLT--LQH##  942
MND-2.GA.x.M14                PI-----AA--T--M----------I---------L----------K---H-------------HL.-----P--VSEK-ATV----Q-EHTT------------AK-HH--V--E-------R-E-------L-----------E--P---.-----T---L-TK--QN  944
MND-2.x.x.5440                PV----NAD--T--M--------I-I---------L----------K---H-------------HL.-----P--VSEK-ATV----Q-EHTT-V----------AK-HH-----E-------R-E-------L-----------E--P---.-----T---L-TK-RQN  944
MNE.US.x.MNE027               -I----NSDL-T--M--------I-I--------F-------Q---RQ--L------S---ITHL.-----A---SQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSME-I-L---HCM----R----DMTPA--.LLNM-T-EQEIQFQ-SK  938
MON.CM.99.L1                  PKS---NAEV-V--M-------SI-----------VW-KIL-R---KC-GIA--E--AM---TQ-.-----P---SQEFE--A---N-HHTT-V-----------N--RQ--ET-KKI--EV-Y-P--LAQ-L--L----------DMAPVD-.FINM-H-ELELQTSNN-  938
MON.NG.x.NG1                  PK----XSDI----M----Y--SI-----------VW-KIL-R---K--GIA--E--AT--IHHV.-----P---SQEFAX-A---N-EHST-V-----------N---QI-TTL-KI--EVQF-S--LAQTLY-L----X----DI-PI--.L-NM-H--LELQHSNN-  941
MUS-1.CM.01.1085              PKT---NAEI-V--M--------I-----------MW-KIL-R---RQ--LG--E--SL--ATQ-.-----A--IS-EFE------S-EHTT-V-----------NK--Q--EA-HKI-EEVTY-E---AQ----L---K---V-DL-NA--.-INMLH-ELELQH--Q-  932
MUS-1.CM.01.CM1239            PKT---NAEL-V--M--------I-----------TW-KIL-R---RQ--LG--E--AL---TQ-.-----A--ISEEFG------S-DHTT-V-----------NK--Q--ET-QKI-EEVTY-E---AQ----M---K-----DLA-A--.-INMLH-ELELQH--S-  932
MUS-2.CM.01.CM1246            PK----NADL-V--M-------------------SF-W-RI-SQ---RL--LE-SN--AT--ISQ-.-P---T---SKEFQ-VA---N-QHTT-V-----------NA--Q--ET-HK--EEVTY-E---AQ----L-N-------DMTPT--.L-NMLY-ELEIQQ--NH  937
MUS-2.CM.01.CM2500            PK----NADI-V--M--------I-----------F-W-RI-SQ-N-RL--LE--N--AT--ITQ-.-----T---SKEFQ-VA-----EHST-V-----------NA--Q--ET-NK--EEVQY-E---AQ----L-N-------DMTPT--.L-NMLY-E-E-QQI-NT  936
OLC.CI.97.97CI12              PR--I-AVGVYR--I----------CL-A-TG--FLYG-ILR--D-EN-VK-V-R--SQ----I-.-----T--VNQKMQ-V---L--EH-T---W------II-AK-RV--E-L-RI-E--T--E--LA--I------N---Q--GTVA-E.L-NQ-Y--T.QI-YTQN  931
RCM.CM.SIVagi-00CM312         -V---T-A---T--M----------I-----------------------K----------A-----HL.-----A---S-A-Q-V----Q-EHT--V----------------HQ--T----I-----K-E-------L------------------.-I----SEL--NK--N-  944
RCM.CM.SIVrcm-02CM8081        -------A--KT--M----------I--------------L-----K---H------A-----HL.-----A---SSA-Q-V----Q-EHT--V----------------HQ--T--T-I-----KIE-------L------------------.-I----S-L--TK--N-  950
RCM.GA.x.GAB1                 -I-----A---T--M--------I-I-----------------------K---H------A----RKL.-----A---S-A-Q-V----Q-EHA--V-----------------Q--I--E---E---K-E-------LV-----------------.-I------LA-NK--N-  944
RCM.NG.x.NG411                -V-----A---T--M--------I-I-----------------------K---H------A-----HL.-----T---S-A-Q-D----Q-EHT--V----------------HQ--I----I-----K-E-------L------------------.-I----S-L--TK--N-  949
SAB.SN.x.SAB1C                SV-----A---V--M--------I-I--------F-------Q---KA--H------S---ITQL.-----T---SQQ-A-I---GK-EHT--V-----------------Q--E----I--D--R-E---I---H---------------A--.LIN--H-ELE--T--QK  963
SMM.SL.92.SL92B               -I---TNAEV-T--M----------I--------F-------R---RQ--L----I-S---I-HL.-----A---SQE-KMVA--L-VE-S--V-----------A-DLH---N-DKI-E---SVE-L-L--AHCM----R----DMTPA--.--NM-T-ELE-QY-NS-  935
SMM.US.x.H9                   XI----NXXL-T--M--------I-I-----G--F-------X---RQ--L------S---ITHL.-----A---SQE-KMVA---X-E-T-XV-------X---A--HH--TQ-DRI-E--NSIE-I-L--IHCI----R----DMTPA--.L-NM-T-EQEIQFQ-SK  938
SMM.US.x.PGM53                -I----NAEL-T--M--------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-SK  938
SMM.US.x.SIVsmH635F_L3        -I----NAEL-T--M--------I-I--------F-------Q---RQ--L----------ITHL.-----A---SQE-KMVA-----E-T--V-----------A--HH--TQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.L-NM-T-EQEIQFQ-SK  938
STM.US.x.STM                  -I----NAEL-T--M--------I-I--------F-------Q---RQ--L------S----THL.-----A---SQE-KMVA-----E-T--V-----------A--HH--TQ-DRI----NTVE-V-L---HCM---KR--L-DMTPA--.L-NM-T-EQEIQFQ-SK  938
SUN.GA.98.L14                 -IT----A-V----I----M-E---I----------MV---L-N-Q-KT--TW----CAM----Q-.-----P--ISKD-E-V---L--QHTT--------------AK--V--Q--SRI-ED-QE-----L--LH-----QR--L--MT-A--.FINM-NA-LE-QY-.QK  949
SYK.KE.x.KE51                 PK-----TDL-T--M----------C---NT----T-TKILKR------GL---QI-A---I-H-.-----P--ISDAFA-------VEHTT----------I--NK-RQ--ET-N-I-EEV-R-E---A--T--L-----------TTPAD-.YINMLY-EL-LQNT..H  968
SYK.KE.x.SYK173               PK-----VDIYN--M----E-----C---NT------TKILKR---D---L--MQI-S---I-Q-.-----P--VSDKFK------C--EHTT----------I---K-RY--EA-S-I--DVT--Q---A--T--L----------I-P---.YINMLY-EL-LQQNTTS  965
TAL.CM.00.266                 PKT--NNAAV-T--M-------QI-C-----------TKIL-R---R---L---QV-S---ISHL.-----P--VS-EMQ-MV--LK-EHSA-V---------A--NK--Q---T-T-I--EVQY-S---AQ-T--L-Y--R--L-DMCPA-A.-INM-Y-EL--TQ--N-  939
TAL.CM.01.8023                PKT--TNAAV-T--M-------Q--C-----------TKIL-R---R---L---QV-S---IEHL.-----P--VS-EMQ-TA--LK-EHTT-V---------S--NK--Q---T-Q-I--EVQY-S---AQ-T--L----R--L-DMCPA-A.LINM-Y-EL--TT--N-  940
TAN.UG.x.TAN1                 PI-----A-V-V--M--------I-I--------FL-----AR---K---K----IIS---ITKL.-----P--VSQE-QTI---GQVEHTT----------S-----RQ--ET-EKI-EDCAF-E---L--CH-----------MTPA--.LINM-T-QLEIQHI-T-  950
VER.DE.x.AGM3                 PV-----A---V--M----------I--------F-------R---K---K----ILS---ITQL.-----P---SQE-A-M---GK-EHTT-V---------SI-----Q--E---KI--DCQYTE---L--CH-----------LTPA--.LINM-T-QLELQH--TK  947
VER.KE.x.9063                 PI-----A---V--M----V---IVI--------F-------R---R---K----ILS---IVQL.-----P---SQE-A-I---GK-EHTT-V---------S------Q--E---KI--DCQYTE---L--CH-----------LTPA--.LINM-T-QLELQT--TK  947
VER.KE.x.AGM155               PI-----A---V--M----I---IVI--------F-------R---K---K----II----ITHL.-----P---SQE-A-M---GKVEHTT-V---------SI-----Q--E---KI--DCQYTE---L--CH-----------LT-A--.LINM-T-QLEINT--TK  950
VER.KE.x.TYO1_patent          PV-----A---T--M---------VI--------F-------R---K---K----ILS---ITQL.-----P---SQE-A-I---GK-EHTT----------SI-----Q--E---KI--DCQYTE---L--CH-----------QTSA--.LIN--T-QLEIQH--TK  950
WRC.CI.97.97CI14              P---M--YEV-T--M--------I--N---------L--KI-S--DAYQ--L-T-Q--S---I-K-.-----T--KSQKMQ-I---L--E--------------I--K-RR--D--HSI-ED--K-E--LA--L------E---L--.-PA--.L-NM-TS-LE-QQT-Q-  954
WRC.CI.98.98CI04              PL--M--YEV-T--M--------I--N---------L--KI----DAYQ--L---QI-----M-K-.-----T--KSQKMQTI---L--E--------------I--K-RR--E--HRI-ED--K-E--LA--L------E---L--.-PA--.L-NM-TAELE-NQT-Q-  953
WRC.GM.05.Pbt_05GM_X02        PI--M--YEI-T--M--------I--N---------L--KI--T-DAYQ--L-TIQ--------K-.-----T--KSQKME-I---L--E--------------I--K-RV--Q--QRI-ED--K-E--LA--IY-----E---L--.-PA--.--NM-QS--E-NQT-Q-  945
```

```
                                                                           p31 Integrase end
                                                                                   Pol end
H1B.FR.83.HXB2                I.TKIQNFRVYYRDSRNPLWKGPAKLLWKGEGAVVIQDNS.DIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED...............*.   1003
H1A1.UG.85.U455               -.S-------------D-I---------------------.-------------------------M-G-----...............-.   1003
H1B.US.90.WEAU160             -.--------------D-----------------------.---------------------------------...............-.   1004
H1C.ET.86.ETH2220             -.L-------------D-I---------------------.----------------------A----G-----...............-.   1011
H1D.CD.84.84ZR085             -.--------------D-I---------------------.---------------------------------...............-.   1004
H1F1.BE.93.VI850              -.----------------V---------------------.E--I-----------------V-----G-----...............-.   1005
H1G.SE.93.SE6165              -.--------------D-V--------------------N.E--------------------------G-----...............-.   1001
H1H.CF.90.056                 -.SN--K---------D-I---------------------.E-------E------------------------...............-.   1003
H1J.SE.93.SE7887              -.--------------D-I-------P-------------.E--------------------------G-----...............-.   1002
H1K.CM.96.MP535               -.LN--K---------E-I---------------------.E--------------------------G-----...............-.   1004
H101_AE.TH.90.CM240           -.--------------D-I---------------------.---------------------------G-----...............-.   1009
H102_AG.NG.x.IBNG             -.I-------------D-I---------------------.---------------------------G-----...............-.   1004
H103_AB.RU.97.KAL153_2        -.I-------------D-I--------------------N.---------------------------------...............-.   1004
H104_cpx.CY.94.CY032          -.--------------E-I---------------------.----------------------N----G-----...............-.   1003
H1N.CM.95.YBF30               -.L-V-----------D-I--------------------G.-----------------------G----G---NQEME...........-.   1015
H1P.FR.06.RBF168              -.L-------------D-I-----T-------------KG.---------------------------DI-R-SESLE...........-.   1004
CPZ.CD.90.ANT                 -.L---Q---H-----D-V-----Q-----------K-QE.E------------KE-R-KIEDR-DL-G--N--...............-.   1004
CPZ.CM.-.SIVcpzDP943          -.L-V-----------D-I-----R-----------K-KE.EV---------------------GSM--G---SQNLE...........-.   1015
CPZ.CM.-.SIVcpzMB897          -.S-V-K---------E-I----------------L--KE.-------------------VE-N--L-D----S...............-.   1012
CPZ.CM.01.SIVcpzCAM13         -.L---K---------D-I-----T-------------QG.E--------------------------------...............-.   1003
CPZ.CM.05.SIVcpzEK505         -.L-V-----------E-T--------------------G.-----------------------G----G---SQDME...........-.   1014
CPZ.CM.05.SIVcpzLB7           -.S-V-K-------R-D-I----------------L--QE.E-------------------E-AN-L-D----N...............-.   1007
CPZ.CM.05.SIVcpzMB66          -.SN--K---------D-I----------------L--QE.E-------------------------------NQNME...........-.   1011
CPZ.CM.05.SIVcpzMT145         -.L---Q--------KD-I-------------------QD.E---------------------A-S-------S...............-.   1003
CPZ.CM.98.CAM3                -.L-V-----------D-I-----R-----------KERE.EV--I-------------------SM-GG---SQGLE...........-.   1013
CPZ.CM.98.CAM5                -.L-V-----------D-I-----R-----------KE-E.EV---------------------GSM-G--N--QNLE...........-.   1009
CPZ.GA.88.GAB1                -.L-V-K---------D-I-----T-------------QG.EL----------------------------N--...............-.   1002
CPZ.GA.88.GAB2                -.L---K---------D-I-----S--------------G.-------------KH----V-----L-GG---NQNME...........-.   1011
CPZ.TZ.00.TAN1                -.L-V---------A-D-I-----R-----------KEGE.-------------KE-------AGGMDD--N-T...............-.   1000
CPZ.TZ.2001.TAN2              -.L-V---------A-D-I-----R-----------KEGE.-------------KE-------AGSMDD--N-T...............-.   1001
CPZ.TZ.2002.TAN3_1            -.L-V---------A-D-I-----R-----------KEGE.-------------KE----V--ANSMDD--N-TQNME...........-.   1005
CPZ.US.85.CPZUS               -.L-V-----------D-I-----------------KE-E.EV-----------K--------A-SM-G----S...............-.   1006
                                                                                p31 Integrase end
                                                                                        Pol end
MAC.US.x.239                  N.S-FK-------EG-DQ-----GE---------ILKVGT.-------------K--.GGGKEV-SSSHME-TGEAREVA.........-.   1020
H2A.DE.x.BEN                  N.SNFK--Q----EG-DQ-----GE---------IVKVGT.-----------------GRQEL-SSPHLEGAREDGEMA..........C.   1033
H2A.GW.x.ALI                  N.L-FK--P----EG-DQ-----GE-----D---IVKVGT.-----------------GRQEL-SGPHLEGAREDGEVA..........-.   1033
H2A.SN.x.ST                   N.S-L------F-EG-DQ-----GE-----D---IVKVGA.---II--------K---GRQEM-SGSNLEGAREDGEVA..........-.   1033
H2B.CI.x.EHO                  N.L-F--------EG-DQ-----GD---------I-KVGT.E---I---------N--GGKEL-CSADVEDTMQAREVAQSN.......-.   1036
H2B.GH.86.D205                N.L-F---Q----EG-DQ-----GE---------I-KVGT.E-------------H-.GGGK-L--S-DME-TRQAREMAQSD......-.   1035
H2G.CI.92.ABT96               N.S-FK-------EG-DQ-----GE---------I-KVGT.E------------K--.GGGKEV-SSTNME-TRQTGEVA.........-.   1016
H2U.FR.96.12034               N.S-FK-------EG-DQ------D---------L-KVGT.----I-----------.GGGKEL-SSSNLEGAEKVREMALPDQTP#V.-.   1036
COL.CM.x.CGU1                 QNQ-FK--KA-WKEH.TGE-Q--GE-V---------RNSQGTLF-K----V--T-LQYGED..VGSENLLSNGQKEAETVKGMD.....-.    991
DEB.CM.99.CM40                -.S-LLG-K----EGAD-T-----T---------CKTEVG------------V-KE-.GKDMDSKVNTQNSN-................-.   1008
DEB.CM.99.CM5                 -.S-LLGYK----EGAD-T-----T---------CKTEVG------------V-K--.GKGVDSKTDTQNPS-................-.   1008
DEB.CM.SIVdeb04CMPF3061       -.S-FSG-K----EGAD-S-----T-----------KLDTG-L---------V-K--.GKDV-SKIDPQDTH-................-.   1011
DEN.CD.x.CD1                  Q.P-FS--K----EGA--Q-Q---E-----------KTQAG--FP---------KP-.NAESRKGEQNKGM-SETDI............-.   1032
DRL.x.x.FAO                   -.L----------EG-DQ------E-V---------KEGT.-L--I--------K--.GKNVDSHTNMES...................-.   1013
GOR.CM.2004.SIVgorCP684con    -.L-------------D-I-----T-------------KG.---------------------------DT-G-SESLEQSS........-.   1003
GOR.CM.2007.SIVgor2139_287    -.L-------------D-I-----T-------------KG.---------------------------DT-G-SESLEQSS........-.   1003
GOR.CM.2007.SIVgorCP2135con   -.L-------------D-I-----T-------------KG.---------------------------DT-G-SESLEQSS........-.   1003
GRV.ET.x.GRI_677              -.Q--L--K----EG-D-V-----R-I---------KEGE.-------------K--.GERKTM-SEG-MEGVREANKQMEGDSDLQDQE*   1040
GSN.CM.99.CN166               N.S-F-K------QGA----Q---V----------V-TQAGE-IT---------KP-.GEAKTE-VGGATHTNND..............-.   1016
GSN.CM.99.CN71                S.S-F-K------QGA--H-Q---V----------V-TQAGE-IT---------KP-.GEAKAE-VGSAAHTSNDREEGRMAD......-.   1017
LST.CD.88.447                 LNS--LK-K----QG-D-Q-----Q----------VKEGD.T-FS-------LVK--.GEGPK-SEGGLHNN.................-.   1020
LST.CD.88.485                 LNS--LK-K----QG-D-Q-----QI---------VKEGD.T-FS-------LVK--.GEGPK-SEGSLDNN.................-.   1020
LST.CD.88.524                 LNS--LK-K----QG-D-Q-R---Q----------VKEGE.N-FSI------LVK--.GEGPK-SQSSMDNQ.................-.   1021
LST.KE.x.lho7                 LNS--LK-K----QG-D-Q-----Q----------VKEGE.N-FS-------LVK--.GEGPK-SESSLDNN.................-.   1020
MAC.US.x.251_1A11             N.S-FK-------EG-DQ-----GE---------ILKVGT.-------------K--.GGGKEM-SSSHME-TGEAREVA.........-.   1020
MAC.US.x.251_BK28             N.S-FK-------EG-DQ-----GE---------ILKVGT.-------------K--.GGGKEV-SSSHME-TGEAREVA.........-.   1016
MAC.US.x.EMBL_3               N.S-FK-------EG-DQ-----GE---------ILKVGT.-------------K--.GGGKEV-SSSHME-TGEAREVA.........-.   1014
MND-1.GA.x.MNDGB1             -.S-NL--K--F-EG-DQ-----GI----------LKYQE.E--I-----C---K--.GESGKNSQ-NLESV.................-.   1010
MND-2.CM.98.CM16              -.Q----------EG-DQQ-----E-I---------KEGT.-L--------------.GKTVDS-PNVEA...................-.   1010
MND-2.GA.x.M14                -.S----------EG-EQ------E-I---------KEGT.-L--------------.GKTVDSNPHMEDRQ-TA..............-.   1017
MND-2.x.x.5440                -.S-----Q----EG-DQQ-----E-I---------KEGV.-L-----------K--.GKTVDSNPHMG-...................-.   1012
MNE.US.x.MNE027               N.S-FK-------EG-DQ-----GE---------ILKVGT.-------------K--.GGGKEV-SSSHME-TGEAREVA.........-.   1016
MON.CM.99.L1                  -.--FSK----F-TGAD-Q-----R-----------KTEEG--LT---------KP-.GTTENVGGDTNQYNLRKQDGLAN........-.   1018
MON.NG.x.NG1                  -.--FSK----F-TGAD-S-----H--------L--KTDQGEVIT----------P-.GSKEDVGSKP-AHQIREVDGMAD........-.   1021
MUS-1.CM.01.1085              K.Y-F--------TGSD-S-----V---R-------KTD-GEVIT---------KP-.GAKEDVGSKSDTG-LRKEDRLDN........-.   1012
MUS-1.CM.01.CM1239            K.S-F--------TGSD-S-----A----------VKTEQGQVIT---------KP-.GAKENVGSKSNTG-HRKEDGLDN........-.   1012
MUS-2.CM.01.CM1246            N.--FSK------RGA--Y-LR--R--------L--KTKEGE-VT---------K--.GTRQNV-SEPDSVHVRKEDGLAD........-.   1017
MUS-2.CM.01.CM2500            K...FSK------RGA--S-Q---Q--------I--KTDQGE-IT---------K--.GTKENV-SDPNPVYDRKEDGLAD........-.   1014
OLC.CI.97.97CI12              K.F-KFSGYRV-YKNKEGQ-C---E---------LLKEGD.KYFS--L-----VKP-.GVDSSANV.......................-.    995
RCM.CM.SIVagi-00CM312         -.S--------F-EG-DQQ-----T-I-----------GQ.-L-------C--VK---RKNVDSETNMEGGQ-EN..............-.   1018
RCM.CM.SIVrcm-02CM8081        -.S--------F-EG-DQQ-----T-I-----------GQ.-L-------C---K---RKDVDSETSMEGRQ-KD..............-.   1024
RCM.GA.x.GAB1                 -.S----------EG-DQ--R-----I----------EETG-L-----------KE--RKDVDSEANLAGRQ-EN..............-.   1019
RCM.NG.x.NG411                -.S--------F-EG-DQQ-------I-E---------QE.-L-------C---K---RKEVDRETNMEGRQ-ES..............-.   1023
SAB.SN.x.SAB1C                -.S----------EG-D-V-------I----------EQG.EL-TI--------K--.GK.-L-SQAPLEGNGRTAGEVD.........-.   1040
SMM.SL.92.SL92B               N.S-F--------EG-DQ------E-----------KVGT.E------------K--.GGRQEMGSS--ME-Q................-.   1006
SMM.US.x.H9                   N.S-FK-------EG-DQ-----GE---------ILKVGT.E------------K--.GGGKEL-SGSHLE-TGEAREVA.........-.   1016
SMM.US.x.PGM53                N.S-FK-------EG-DQ-----GE---------ILKVGT.E------------K--.GGGKEL-SGPHLE-TGEAREVA.........-.   1016
SMM.US.x.SIVsmH635F_L3        N.S-FK-------EG-DQ--R--GE---------ILKVGT.E------------K--.GGGKEL-SGSHLE-TGEAREVA.........-.   1016
STM.US.x.STM                  N.S-FK-------EG-DQ-----GE---------IVKVGT.-------------K--.GGGKEV-SGSHLE-TREAGEVA.........-.   1016
SUN.GA.98.L14                 -NS--LK-K----EG-D-Q-------------------KQ.GEN-L--------LVK--.GGESSSVEMVG..................-.   1017
SYK.KE.x.KE51                 T.Q-FS--K----HGT.SD-Q---S-----------TPDQQVIA----------SSD.GERVDSGTHLEITSKSN..............-.   1041
SYK.KE.x.SYK173               ..P-FS-------QGK.NE-----R----------V-TEEG--FA---------T-H.GERMDSGSHVENDPKTD..............-.   1037
TAL.CM.00.266                 -.QNFSD-K----KGA----Q---H-I--------LRTDEGEVIT---------KP-.G-AL-NKIDLESSK-QDAEMGRDN.......-.   1020
TAL.CM.01.8023                -.HNFSD-K----KGA----Q---H-V--------LRTDEGEVIT---------KP-.G-AM-NKTDLEGSK-QDAEMGRDN.......-.   1021
TAN.UG.x.TAN1                 Q.Q--S--K----EG-E-V-----T-----------KEGE.EL-----------KE-.GERKTVGDKTNMEG.................-.   1020
VER.DE.x.AGM3                 -.Q--L-------EG-D-V----GQ-I---------K.GGVEL-EY--------K--EPRKR.MGDESNLEGAGGADN...........-.   1023
VER.KE.x.9063                 -.Q--L-------EG-D-V-----Q-I---------K.EGEEL-I---------K--EPRKR.MGDEGNMEGLRGPDNQMARNSQILDD-.   1034
VER.KE.x.AGM155               -.Q--L-------EG-D-V-----R-I--------LK.EGEEL-----------K--EPRKT.LGDETHLEGAGGSDHQMAGDS.....-.   1032
VER.KE.x.TYO1_patent          -.Q--L-------EG-D-V-----Q-I--------LK.DGS-L-----------K--EPKQR.VGNEGDVEGTRGSDN...........-.   1026
WRC.CI.97.97CI14              K.L-FK--Q----TGA-QQ-Q--G--P------L--ETPEG.-IT-------L-KVWNGEGMDRSSSDKN...................-.   1023
WRC.CI.98.98CI04              K.L-FK--Q----TGA-QQ-Q--G---------L-VETPEG.-IT--K------KVWNGEGMDRSSSNKD...................-.   1022
WRC.GM.05.Pbt_05GM_X02        K.F-FK--Q----TGASQQ-Q--G--V------L-V-TPEG-LLT---------KTWDGKGMDSSTTN.....................-.   1013
```

```
                                    Vif start
H1B.FR.83.HXB2                .........M.ENR.........WQVMIVWQ..VDRMRIRTWK.SLVKH.HMYVSGK.ARGWFY.RHHYESPHPRISSEV.HIP.......LG...DA.RLVITTYWGLHTGERDWHLGQG.........VSIEWR...........................KKRY  94
H1A1.UG.85.U455               .........-.---.........--------..----K---N.-----.-----K-.-Q----.------R-S-V----.---.......---...E-.---VR---------K-----H-.........------...........................L---  94
H1B.US.90.WEAU160             .........-.---.........--------..----------.-----.---I-K-.-K--S-.------T---------.---.......---...EG.K-------------------------.........------...........................-Q--  94
H1C.ET.86.ETH2220             .........-.---.........---L----..----K---N.-----.--HI-RR.-N--V-.-----D-R--KV-----.---.......---...E-.--I-K-----Q---------H-.........------...........................LRS-  94
H1D.CD.84.84ZR085             .........-.---.........--------..------N---.----Y.--HI-K-.-K-----.-----D----K-----.---.......---...---.----V-----------E------.........------...........................-R--  94
H1F1.BE.93.VI850              .........-.---.........--L-----..------N---.----Y.-----K-.-K--S-.---FQ-R---V------.---.......-E...EV.K----------P---E------.........------...........................QGK-  94
H1G.SE.93.SE6165              .........-.---.........--------..---------H.-----.-----K-.------.-P--A-R---V------.---.......---...---.T--V---------K--Q--H-.........------...........................QR--  94
H1H.CF.90.056                 .........-.---.........--------..------N---.----Y.--HI-R-.------.---F--T---------.---.......---...E-.----------N---E------.........------...........................L---  94
H1J.SE.93.SE7887              .........-.---.........--------..------N---.----Y.--N--K-.--K-L-.----D-N--K-------.---.......---...E-.I---------Q------------.........------...........................QR--  94
H1K.CM.96.MP535               .........-.---.........--------..------N---.----Y.--HI-K-.-NR-Y-.------R--K-------.---.......---...---.E--V-----L------------.........------...........................L---  94
H101_AE.TH.90.CM240           .........-.---.........--------..---------N.-----.---I-K-.-KK---.-------Q--KV-----.---.......---...E-.----R-----Q--K-----H-.........------...........................QRK-  94
H102_AG.NG.x.IBNG             .........-.---.........--------..---------N.-----.-----K-.-K----.---F--R--KVC----.---.......---...---.----VR--------------H-.........-----K...........................Q---  94
H103_AB.RU.97.KAL153_2        .........-.---.........--------..---------N.------.-I-I-K-.----V-.K-----RN--------.---.......---...---.K---K-----------------.........A-----...........................-E--  94
H104_cpx.CY.94.CY032          .........-.A--.........--------..----K---N.-----.-----K-.-K-*--.------R--KV-----.---.......---...E-.---VR-----QP--Q-----H-.........------...........................LR--  93
H1N.CM.95.YBF30               .........-.---.........---V----..----K--K-N.-----.-----K-.-K--Y-.-----TH--K-------.---.......V-...Q-.---TV-----T---QS----H-.........------...........................LRK-  94
H1P.FR.06.RBF168              .........-.---.........--------..IS-EKVKA-N.----Y.-K-R-K-.TKN---.-----XV------G-.Y--.......V-...A-.II-V------MP--KEEQ--H-.........X-V---...........................QGK-  94
CPZ.CD.90.ANT                 .........-..I..........--------..------NI--.------.-IWETKV.LKP-K-.-----ND--KKGE--.---L.......PT...LDKK--V-VF---QC---P----H-.........------...........................CGK-  88
CPZ.CM.-.SIVcpzDP943          .........-.---.........---V----..-----VK--N.----Y.---RTK-.--K---.---------SN--V---I.---.......-E...K-.K--V------MP--HE----H-.........------...........................YGS-  94
CPZ.CM.-.SIVcpzMB897          .........-.---.........-K------..I----VKA-N.-----.-----K-.------.---------EN-KV---I.---.......V-...---.K---V-----MP--------H-.........------...........................Q---  94
CPZ.CM.01.SIVcpzCAM13         .........-.--K.........--------..-------KA-N.----Y.---KTK-.-KD---.-----H-N--V------.---.......FK...-NSK-I-----A-T----A----H-.........---Q--...........................MGS-  95
CPZ.CM.05.SIVcpzEK505         .........-.---.........---V----..-----V---N.-----.-----K-.------.--------TDN-K----I.---.......---...E-.K---V-----MP---P----H-.........------...........................QGI-  94
CPZ.CM.05.SIVcpzLB7           .........-.---.........-K-L----..I-----KA-N.----Y.-I---R-.------.--------EN-KV---I.---.......---...E-.K---V-----TP--------H-.........-----K...........................Q-K-  94
CPZ.CM.05.SIVcpzMB66          .........-.---.........--------..-------K--N.-----.-----R-.--N---.--------NN-K------.---.......---...E-.K---------T----A----H-.........------...........................QRS-  94
CPZ.CM.05.SIVcpzMT145         .........-.---.........---Q----..-----VK--N.----Y.---R-K-.-KN---.----DGTN--MA-QI.Y--.......IN...Q-.K-IV------MP--------H-.........-A----...........................ERG-  94
CPZ.CM.98.CAM3                .........-.---.........-L------..-----VKA-N.----C.---KTK-.-KD---.-------S--KV---I.---.......IE...K-.K--V------MP-QHA---H-.........------...........................CGS-  94
CPZ.CM.98.CAM5                .........-.---.........---V----..-------K--N.----T.---RTK-.-KK---.--------DN-KV------.---.......-D...T-.K--V------T--QHA---H-.........------...........................CGS-  94
CPZ.GA.88.GAB1                .........-.---.........--------..-------K--N.----Y.-I-R-K-.------.-----DH-N-KVA--I.---.......FR...-YSK-IV----A-SP---A----H-.........---Q--...........................LGS-  95
CPZ.GA.88.GAB2                .........-.-S-.........-------L..---------N.----F.---K-K-.------.-----D--N-KV------.---.......IR...---.KI-VC---N-MP--QP----H-.........------...........................QRT-  94
CPZ.TZ.00.TAN1                .........-.---.........---QV--M..I----L---T.------.-IFTTKC.CKD-K-.-----TDT-KRAG-I.---.......-T...ERSK--VLH----AC---P----H-.........IGL---...........................QGK-  95
CPZ.TZ.2001.TAN2              .........-.---.........---QV--M..I----L---N.------.-IFTTRC.C-E-K-.-----TDV-KRAG-I.---.......IT...GKAKI-VLH----AC---P----H-.........-GL---...........................QGK-  95
CPZ.TZ.2002.TAN3_1            .........-.---.........---Q---M..I----L---N.------.-I-TTKC.CKE-K-.-----TDV-KRAG-I.---.......-T...EKSK--VLH----AC---P----H-.........-GL---...........................QGK-  95
CPZ.US.85.CPZUS               .........-.---.........---L----..-----VKA-N.----Y.---R-K-.-K----.----------TN--VN---.---.......ID...K-.KI--I-----K---QA----H-.........------...........................QGT-  94
                                    Vif start
MAC.US.x.239                  .........-E-EKR........-IAVPT-R..I.PE-LER-H.--I-Y.LK-KT.-DLQKVC-VP-FKVGWAWWTC-R-.IF-........-Q...EGSH-EVQG--H-T.P-KG-LSTYA.........-R-T-Y...........................S-NF  96
H2A.DE.x.BEN                  .........-E-D-N........-I-VPT-R..-.PG-MEK-H.A---Y.LK-RT.-DLEEVR-VP--KVGWAWWTC-R-.IF-........-Q...GKSH-E-QA--N-T.P-KG-LSSHA.........-RLT-Y...........................TEKF  96
H2A.GW.x.ALI                  .........-E-GKS........-I-VPT-R..-.PG-MEK-H.----Y.LK-RT.-DLEKVC-VP--KVGWAWWTC-R-.IF-........-Q...GRSH-E-QA--N-T.P-KG-LSSYA.........-R-T-Y...........................TEKF  96
H2A.SN.x.ST                   .........-E-GKR........-IAVPT-R..-.PG-MER-H.--I-Y.LK-RT.GDLEKVC-VP--KVGWAWWTC-R-.IF-........-K...GESH-E-QA--N-T.P-KG-LSSYS.........-RLT-Y...........................TEKF  96
H2B.CI.x.EHO                  .........-E-EKN........-IAVPT-R..I.PC-LER-H.--I-Y.LK-RT.-DLQQVS-VP--KVGWAWWTC-R-.IF-........-K...EGAH-EVQG--N-T.P--GFLSSYA.........-RLT-Y...........................ERSF  96
H2B.GH.86.D205                .........-E-EKD........-I-VPT-R..I.PG-LER-H.--I-Y.LK-RT.GELQQVS-VP--KVGWAWWTC-RI.IF-........-N...KGAW-EVQG--N-T.P--GFLSSYA.........-RLT-Y...........................ERNF  96
H2G.CI.92.ABT96               .........-E-EKR........-IAVPT-R..I.PG-LEK-H.--I-F.LKFKT.-DLQKAV-VP--KVGWAWWTC-R-.IF-........-T...KEAH-E-QG--N-T.P-KG-LSQYA.........-RLT-Y...........................TRKF  96
H2U.FR.96.12034               .........-E-EKN........-I-VPT-R..-.PR-LER-H.--I--.L#-NT.-ELQKVC-VP--KVGWAWWTC-R-.IF-........-K...KGSH-EVQA--N-T.P-KG-LSSYA.........-RLT-Y...........................TRGF  95
COL.CM.x.CGU1                 .........-..EK.......M-A-R-CY.......PMGKR-...P-PLREWIKKQFPECQY-YLGKT---EWWTEGKIEFY.........--S...-IK-S-IYLGT-TPDRYH...........RGPP--Y-LQ...........................IGKW  89
DEB.CM.99.CM40                ..........M.EKT........-I-RLTHKIP..MNKLDR-C.-AI-R.EK-KT.-ELDTVQWIH-FELRNSYYTQTKIV.F-........IT..KEGGV-VELL-C-T.P-KG-LPTMA.........-A-A-E........................TKTW-..  96
DEB.CM.99.CM5                 ..........M.EKE........-I-RLTHKIP..VNKLDK-C.-AI-R.GK-KT.-ELQAIKWIH-YELRRAFYTQTKIV.F-........VT..EEGKV-VEIF-S-A.P-KG-MPSMA.........-G-S-E........................QTGWQ..  96
DEB.CM.SIVdeb04CMPF3061       ..........M.EKM........-V-RLTHKIP..MNKLDR-C.-AI-K.GK-KT.-ELAPITWQH-FELRNSYFTQTRII.F-........IT..KEGKV-VELF-C-A.P-KG-LP-MA.........-GLS-E........................QGNWQ..  96
DEN.CD.x.CD1                  ..MQNLGRENR.TKE........-I-RLTYKIG..ETKLEH-M.-TL-S.II-KT.NELK-TVIIP-FRITWEWYTMTKIH..........IPL.R-AYMEVKL--H-T.P-KGYLSQVA.........I--AYVH.....................LSTGWA..  105
DRL.x.x.FAO                   ..........M.EKM........-I-IPT-RVE..-YMVER-N.--C-Y.-KHKGE-YLER-H-AP-FQC-.GWWTH-QW.T--........FK...-KSKI-V-VL-N-T.PDKG-LSKYAI.........T--HI...........................QEKF  95
GOR.CM.2004.SIVgorCP684con    .........-.---.........--------..IP-E-VKA-N.----Y.-K-RTKQ.-KQ---.-----AIN--V--G-.Y--........V-...P-.EIIV------MP--KEEQ--N-.........-G----...........................QGK-  94
GOR.CM.2007.SIVgor2139_287    .........-.---.........--------..IP-E-VKA-N.----Y.-K-RTKQ.-KQ---.-----AVN--V--G-.Y--........V-...P-.EIIV------MP--KEEQ--H-.........-G----...........................QGK-  94
GOR.CM.2007.SIVgorCP2135con   .........-.---.........--------..IP-E-VKA-N.----Y.-K-RTKQ.-KQ---.-----AVN--V--G-.Y--........V-...P-.EIIV------MP--KEEQ--H-.........-G----...........................QGK-  94
GRV.ET.x.GRI_677              ........M.EREKQ........-I-RV--R..-SERQ-SR-R.GI-TY.KIR..N-QLP.-E-RH-WQVQWQFWTY-QFI.--L.......S...KDDYIEVNI-HN-T.P--G-LSSH-.........-GLSYY..HQ.....................-G.-  96
GSN.CM.99.CN166               ......M.ERQRQKM........-V-QPIQI...MTE-KVD-LLRAT--.-IWSG...KTPFV-VH-YQLQHQRFTQNKIRL.AMDTRKVGE..EVE-TYIE--IL-DTTSHGPASLSRSTYYQQA....-I---VYN................RSKLGAREGDIIW-  123
GSN.CM.99.CN71                ......M.ERQRQKM........-V-QPIQV...MTE-KVE-LIRAT--.-IWSG...KTPFV-IH-YQLQHQRFTQNKIRL.ALDTRQVGE..EVE-TYIE--IL-DTTSHGPASLSRSTYYQQA....-T---VYN................RSKLGARETDITW-  123
LST.CD.88.447                 MER........DQ-..........I-RVAYI..TT-SLVEKFLDIRRA-....KETQNDWVSLH.GTG.VGWEFYTY.NNIV--LE..........TGT--VRC-AH-A.AG-G-ISQWA..............-----IYG.........................N-  92
LST.CD.88.485                 MER........DQ-..........I-RVA-I..TT-NL-EKFLDIRRA-....KETQVDWVSLH.GTG.VGWEFYTY.NKIV--LE..........IGT--VRC-AH-A.AG-G-ISQWA..............-----IYG.........................N-  92
LST.CD.88.524                 MER........DQ-..........I-RVA-I..TN-SLVEKFLDIKRN-....SES-NDWVSLH.GTG.VGWEFYTY.NKLL--LD..........TGT--VRI-AH-A.AG-G-ISQWA..............-----IYG.........................N-  92
LST.KE.x.lho7                 MER........DQ-..........I-RVA-I..TT-SL-EKFLDIVRLN....TEA-ENWVGLH.GTG.VGWEFYTY.HKLV--LE..........SGT--VRI-AH-A.AG-G-ISQWA..............-----VYR.........................D-  92
MAC.US.x.251_1A11             .........-E-EKR........-IAVPT-R..I.PE-LER-H.--I-Y.LK-KT.-DLQKVC-VP-FKVGWAWWTC-R-.IF-........-Q...EGSH-EVQG--H-T.P-KG-LSTYA.........-RVT-Y...........................S-NF  96
MAC.US.x.251_BK28             .........-E-EKR........-IAVPT-R..I.PE-LER-H.--I-Y.LK-KT.-DLQKVC-VP-FKVGWAWWTC-R-.IF-........-Q...EGSH-EVQG--H-T.P--G-LSTYA.........-R-T-Y...........................SRNF  96
MAC.US.x.EMBL_3               .........-E-EKR........-IAVPT-R..I.PE-LER-H.--I-Y.LK-KT.-DLQKVC-VP-FKVGWAWWTC-R-.IF-........-Q...EGSH-EVQG--H-T.P--G-PSTYA.........-R-T-Y...........................SRDL  96
MND-1.GA.x.MNDGB1             MER........VE-..........I-RLT-KV..SSQ--EK-HWLVRRQMAWATANNEEGC-WLYP-FMAYNEWYTC.SKVV-IIN..........RDI--IVRS--H-Q.I-VGCLSTYAV.........---AVVR......................PPPFE-EW  105
MND-2.CM.98.CM16              ..........M.EKQ........-I-IPT-KLD..-HAVER-N.----Y.-K-KGE-HLDK-DLKA-FQC-.GWWTH-QKY.--L.......E...EDEQII--IL-N-T.P-KG-LSTYA.........MT--Y.....K.......................FTN-  95
MND-2.GA.x.M14                ..........M.EKQ........-I-IPT-RID..-RQLEK-N.----Y.-K-KEE-HLDK-ELFH-FQC-.GWWTH-QKI.--F.......K...-GSKII--AL-N-T.P-KG-LSQYA.........ITA-Y.....N.......................-GD-  95
MND-2.x.x.5440                ..........M.EKQ........-IAIPT-EDR.QDVKL-D-N.-I--Y.-K-KGE-HLEN-QLYP-FQC-.GWWTH-QK.K--........FK...-GSKII--VL-N-T.P-KG-LSQYA.........FTV-Y.....Q.......................-EN-  96
MNE.US.x.MNE027               .........-E-EKR........-IAVPT-R..I.PE-LER-H.--I-Y.LK-KT.-DLQKVC-VP--KVGWAWWTC-R-.IF-........-Q...EESQ-EVQG--N-T.P--G-LSTYA.........-R-T-Y...........................SRNF  96
MON.CM.99.L1                  ........M.APRKM........-V-TPTNI...ISESKMD-LIRCT-W.-ILTG...SAPFV-VH-YQLHNQRFSQNKIK.L-L.DLGITQEGESW-TY-E--I--D-TNVGPSALSPTVYNKQA....YT---VYWMQKVER...VPLDPVW..NPNPRDREV.RTW-  134
MON.NG.x.NG1                  ........M.APRKM........-V-SP-HI...RSE-WME-LIRAT-Y.NI-TG...KVPFV-VH-YQLTHQRFSQCKIK.L-L.EMGITQXXEQW-TYXE--IL-DXTNVGPASLSPSVKWKQA....YI---VYWRQPRVPQQEGVVDRVWFPSPRRPDREIDRVW-  140
MUS-1.CM.01.1085              ........M.APRKM........-V-SPIQE...TSE-KID-IIRAT-Y.YILSG...RAAFT-VH-YQLYHQRFSQ-KIR.--L.TQNTRESGAVESTYIEL-VL-DVTNVGPASCSKSTWWKQS....FIL--VYL................RRHPQDREQDYIW-  122
MUS-1.CM.01.CM1239            ........M.APRKM........-V-SPIQE...TTE-KMD-IIRAT--.-IFSG...KTPFV-VH-YQVQHQRFSQNKIK.LAL.TQNTLESGETEVTFIE--VL-DVTNVGPASLSKSTYWKQS....YIL--VYM................RKGPRDRERDYVW-  122
MUS-2.CM.01.CM1246            ........M.APGKM........-I-SPIQY...MSE-KMD-LIRCT--.YIFSG...KAPFT-VH-YQLQHQRFTQNKIK.--L.SINRIE-GPTETTYIE--ILLDVTNVGPASLSRSTYWQQS....YILK--YI................RAAPQNREIDMVH-  122
MUS-2.CM.01.CM2500            ........M.AQRKM........-I-TPIQF...MTE-RMD-LIRCT--.-ILSG...KAPFS-IH-YQVQHQRFTQNQIK.--L.TIKKREEDKVESTYIQ--IL-DVTNVGPASLSKSTYWQQS....YIL--VYV................RAAPQDREVDMVW-  122
OLC.CI.97.97CI12              ............M........E-IAVPMYRL.......NNGQVEH--FMAKLAYKFWGYH-IWQNCYKYPREWWTHGKIYI.-VP...........GGEIE--L--HIRH-KGLQNS-GV..........NFKYIV..........................GISW  91
RCM.CM.SIVagi-00CM312         .........-.AEKM........-I-RPI-R..---RK-EQ-H.----Y.-Q-KGK-E-KE-E-VP-FKVPWGWWSH--I.-V-..........-...KNTKVK-----N-T.A-KG-LSTY-.........-G-AYID.NTC.....................DPC-  101
RCM.CM.SIVrcm-02CM8081        .........-.AEKM........-I-RP--R..---RK-EQ-H.----Y.---KGK-E--E-E-VP-FKVPWGWWSH---.---.........-...NNTKIKV----N-T.T-KG-LGTY-.........AALAYID.QKC.....................DPP-  101
RCM.GA.x.GAB1                 .........-.AEKM........-I-RPI-R..---RK-EQ-H.----Y.-Q-KGK-A-KE-E-VP-FKVPWGWWSH---.---.........-E...EGSK-K-----N-T.V-KG-LGTY-.........-G-L-I..KG.....................D..-  97
RCM.NG.x.NG411                .........-.AEKR........-IERPI-K..---RKVEQ-H.----Y.-Q-KGK-E-KD-E-VP-FKVPWGWWSR---.---.........-...NNTKIKV----S-T.I-KG-LGTY-.........-G-AYID.NRC.....................DPC-  101
SAB.SN.x.SAB1C                ..........M.EKH........-I-RPL-KVT..GGQQER-T.----Y.--H--KQCVH.-R-TP-TKIRWNWYSYQ-WV.--L.......K...-GALIKV-N--H-T.P-KG-LETYA.........TG-GY.....S.......................-GEW  95
SMM.SL.92.SL92B               .........-E-GKR........-VAVP--R..IS-RIVER-H.-CI-F.-K-KT.RELEKAC-VP--KVGWAWYTA-R-.IF-........-E...EGSH-EVQV--N-T.P-KG-LSSYA.........-R-T-Y...........................SEKF  97
SMM.US.x.H9                   .........-E-EKN........-I-VPT-R..I.PG-LEK-H.--I--.LK-NT.-DLQKAC-VP--KVGWAWWTC-R-.IF-........-R...-EAH-EVQG--N-T.P-XG-LSTYA.........-R-T-Y...........................SRNF  96
SMM.US.x.PGM53                .........-E-EKN........-I-VPT-K..I.PG-LEK-H.--I-Y.LKFRT.-DLQKAC-VP--KVGWAWWTC-R-.IF-........-R...-ESH-EVQG--N-A.P-KG-LSTYA.........-R-T-Y...........................SRKF  96
SMM.US.x.SIVsmH635F_L3        .........-E-EKN........-I-VPT-R..I.PE-LER-H.--I--.LK-NT.-DLQMAC-VP--KVGWAWWTC-R-.IF-........-R...-KTH-EVQG--N-T.P-KG-LSTHA.........-R-T-Y...........................SRNF  96
STM.US.x.STM                  .........-E-EKR........-I-VPT-R..I.PG-LER-H.--I--.LK-NT.-ELSKAC-VP--KVGWAWWTC-R-.IF-........-Q...GEAH-EVQG--N-T.P-KG-LSEYA.........-R-T-Y...........................TRNF  96
SUN.GA.98.L14                 MEENPPQWRWWDE-.EWEDRRQYKI-R---L..I--IAVEKFLDLRRM-....RET-DDWVSM-.GTG.TGWEWYTY.NKII--VT..........YGTV-VRI-GH-T.PAKGLINQW-.........C-M--IYN.........................S-  109
SYK.KE.x.KE51                 ..........M.EKE........-I-VLT-KLH..PNQ-DKMQ.HIL-V.-K-K-.-QLEKAT-KH-YQITWEWWTRAQWE.--V..........-...QGV-I-KF-HN-T.P-KGLLRTE-.........-GLI-A........................HPSGW..  95
SYK.KE.x.SYK173               ..........M.EKE........-I-VPT-RMT..PRQ-DRLQ.HII-T.-K-K-.-ELEKAT-KH-YQIEWQWYTYCQWT.--V..........-...DGTIW--F-HN-A.P--G-LHM--.........IR-QYQ........................WNQWN..  95
TAL.CM.00.266                 ..........M.DKL........-VTRLT-KVP..-N-MQR-V.GTIRF.-IHKT.-YLQKLIWVH-FQLRHASYTQAKWI.L-L......EEKDTRVEY-T--Q--C-A.PDHGYTSDWD........TI----SVSDIV....................RDEKVYA..  108
TAL.CM.01.8023                ..........M.DKL........-VTRLT-KVP..-N-MQR-V.GTIRY.---KT.-YLQKLIWVH-FQLRMASFTQAKWI.L-L......EEKENRVEYIT--Q--C-A.PDKGYTSDWD........TI----SVSDVT....................LDAKVYS..  108
TAN.UG.x.TAN1                 ........M.EREKL........-VTRLT-R..-SGEH-DK--.GI--Y.--R..NRLQD.-T-LM-YQCGWAWYTC-RFL.--L.........-...GEGKI-VDC--H-T.P-QG-LSTYA.........-A-SF...ENW.....................QNT-  97
VER.DE.x.AGM3                 ........M.NQEKE........-VMRVT-K..-PEEL-TK-Q.GI-RY.W-R..TRKLD.-K-RM-YQITWAWYTM-RYE.--L.........-...QHGSIHVDL--H-T.P-KG-LSTYA.........EG-QYL..SNR.....................DPW-  98
VER.KE.x.9063                 ........M.NQEKE........-VMRVT-K..-SGDL-TK-Q.GI--Y.W-T..KRNLK.-Q-MM-YQINWAWYTMCRYI.--I.........-...LEGEIQ-DC--H-T.P--G-LSTYA.........IG-QYL..GAV.....................G.NF  97
VER.KE.x.AGM155               ........M.SQEKH........-VMRLT-K..-QEEV-TK-Q.GI-RY.W-N..KRNLK.-E-KM-YQITWAWYTM-RYV.--L.......P...GSGEIHVDI--H-A.PKQG-LSTYA.........-G-QYV..SLV.....................NDK-  98
VER.KE.x.TYO1_patent          ........M.NP-KE........-VMRVT-K..-PGDL-TK-Q.GI-RY.W-R..QRNLK.-N-YM-YQITWAWYTM-RYV.--I.........-...KHGEICVDL--H-T.P-QG-LSTYA.........-G-QYV..SNL.....................ESK-  98
WRC.CI.97.97CI14              .........-.-KE.........-IGVPLIR..ISKGKF-R-TYYIANLIKAGQANFLWGS-HAAAECFTRHIVK-.PIQ...............GHWEMQ-RC--N-S.P-KG-MDTM-F.......NIRMVY..........................EN-W  93
WRC.CI.98.98CI04              .........-.-KE.........-IGVPLIK..ISKG-FKR-AYWIAHLMQNKKASFLWGT-HTQAECFTRYLL--.PLQ...............GHWE-Q-RC--N-A.P-KG-LDTM-F.......NIRMVY..........................EN-W  93
WRC.GM.05.Pbt_05GM_X02        .........-.-KE.........-IAVPLIK..ISKG-F-R-SYWIAELIQEKQARFLFGS-HVHAECFTRHIV--.PLQ...............GHWEMQ-RC--N-A.P-IG-IDTL-F.......NIRMV-..........................ENKW  93
```

```
                                                                                                                                                                              Vif end
H1B.FR.83.HXB2              STQVDPELADQLIHLYYFDCF.SDSAIRKALLGHI..VSPRCEYQAGH...........NKVGSLQ.YLALA..........................ALITPKKIKP.....PLPSVTKLTEDRWNK..................PQKT...KGHRGSHTMNGH....   192
H1A1.UG.85.U455             ------D---H----H-----.-E----R-I--Q-..-R---------...........-------.----K..........................--V--TRA--.....-----K---------..................----...R-----R-L-R-....   192
H1B.US.90.WEAU160           ------D---------H----.-E----N-I---L..-I---------...........-------.----T..........................----------.....-----K---------..................----...------------....   192
H1C.ET.86.ETH2220           N-----G---H---MH-----.AE------I--YR..-----D-----...........-------.----T..........................---K---A--.....-----S--V--K---..................----...R-R--N------....   192
H1D.CD.84.84ZR085           ------G-------M------.A-------I----..-----------...........-------.----T..........................---A---R--.....-----K---------..................-RQ-...--R---------....   192
H1F1.BE.93.VI850            R--I--G-------I------.-E------I---R..I----N-----...........-------.----T..........................---A-E-T--.....-----Q--V------..................--E-...R-----------....   192
H1G.SE.93.SE6165            R---E-D---H----H-----.--------I--Q-..-----------...........-Q-----.----K..........................V-V-S-RSR-.....------E-A------..................----...R---ENP-----....   192
H1H.CF.90.056               ----E-G-------MH-----.-E------I--RV..-R---N-P---...........KQ--T--.----T..........................--VA------.....-----R--V------..................----...R-----------....   192
H1J.SE.93.SE7887            R-----G-------MC-----.--------I--Q-..-----D-----...........-------.----T..........................---R--RR--.....-----Q--V------..................----...T---E--------....   192
H1K.CM.96.MP535             R---E-D-------I------.-E--V---I---R..------C----...........-------.----T..........................--VA-RRP--.....-V---K--V------..................----...R-----Q-----....   192
H101_AE.TH.90.CM240         ---I--D---K----Q--G--.--------I--QV..-RR----PS--...........-------.----K..........................--T---R-R-.....------.EI------..................----...R---ENP-----....   190
H102_AG.NG.x.IBNG           ---I--D-----------N--.--------I--E-..-R---------...........-------.----N..........................--VA-T-T--.....-----R--A---KE..................----...R-----RP----....   192
H103_AB.RU.97.KAL153_2      ------N--------------.-E----N-I---R..---S---R---...........-------.-----..........................--R-------.....---------------..................--R-...-D-------S--....   192
H104_cpx.CY.94.CY032        ------D-------MH-----.-E------I---R..-----------...........-------.-----..........................---S---T--.....-----K--V------..................----...R-R-ENQI----....   191
H1N.CM.95.YBF30             K------M--K----H-----.TA----Q-V--RP..-L-----P---...........KQ--T--.----T..........................-WVGA--R--.....--------------E..................H--M...Q----NPI----....   192
H1P.FR.06.RBF168            T--I---T--R----H--Q--.----V-R-I--DR..-LNK---S---...........SQ-----.F---K..........................VVVGKV-R--.....-----QI--K-I-S-..................--R-...--RQE--L-S-CYKP.   195
CPZ.CD.90.ANT               I------T---M--Q---P--.--Q-V-Q-I--ER..ILTY-H-KK--...........SQ--T--.---FC..........................KIL.EFRGY-KGPRRQF--LSI--------..................-RRM...R---ENQ.........   185
CPZ.CM.-.SIVcpzDP943        R--I---M--R----H-----.AA------V--TQ..-V-K---P---...........QQ-----.----K..........................-WVGVE-KR-.....-----A--A------..................H--I...RDQQE-------....   192
CPZ.CM.-.SIVcpzMB897        ---I------H----H-----.AE------I--Q-..------FP---...........T---T--.---IS..........................--LKR-PK--.....-----S--A------..................-R--...----KE------....   192
CPZ.CM.01.SIVcpzCAM13       V-----YV--K---SQ-----.A-----R-I----..-E---A-RE--...........RQ--T--.F---K..........................--VKE-RLR-.....-----A--------T..................--R-...---PEN---S--....   193
CPZ.CM.05.SIVcpzEK505       R--I------K----------.TA----Q-V--RP..-I-K---P---...........KQ-----.----I..........................-WVGVQ-RR-.....--------------R..................H--I...-D-----T--L....   192
CPZ.CM.05.SIVcpzLB7         ---I------HM------N--.AE------I--RV..------FP---...........K---T--.---IS..........................--LQRQPTR-.....-----S--A------..................S---...--------I-S--....   192
CPZ.CM.05.SIVcpzMB66        R--I--D--------------.AE------I--Q-..-------PT--...........-------.----S..........................T--KS-PRR-.....-----R--A-----N..................S---...RD-K-N-I----....   192
CPZ.CM.05.SIVcpzMT145       -------T--R----H-----.TAT---N-I-EKA..-L-K--FPE--...........S------.-W--K..........................Q-VKT--NR-.....-----A--A------..................--R-...RD-QE-PMSS-C....   192
CPZ.CM.98.CAM3              -------M--K----Q-----.AA------V--KQ..-Y-K---P---...........QQ-----.----R..........................-WVRVG-K--.....--------------R..................H--I...RV-QE-------....   192
CPZ.CM.98.CAM5              K------T--R----H-----.AA------V--KQ..-Y-K---PT--...........QQX----.----R..........................-WVGVQ-KT-.....----IX--A------..................Y---...R--QEN----X-....   192
CPZ.GA.88.GAB1              V-----FT--R---SQ-----.AET---R-I--QL..-A-----KE--...........RQ-----.F---K..........................---SERRHR-.....-----A---------..................H-R-...-V-QENL-R---....   193
CPZ.GA.88.GAB2              R--I--QT--R---C------.AA--V-R-I--E-..-TKV--FPR--...........PQ-Q---.---IV..........................QVVKT--K--.....------I-A------..................----...RDQQENL--S--....   192
CPZ.TZ.00.TAN1              ---I---T------TR--T--.AAG-V-Q-I--ER..ILTF-HF-S--...........RQ--T--.F--FR..........................KVVESQDKQ-KGPRR---------------..................HRT-...T-R-EN--LS-C....   198
CPZ.TZ.2001.TAN2            C--L---T------TR--T--.TAG-V-Q-I--ER..ILTF-HF-T--...........RQ--T--.F--FK..........................KXXEIQYK--QGPRR---------------..................H-RI...T-PK-N-I---C....   198
CPZ.TZ.2002.TAN3_1          C------T------TR--T--.AAG-V-Q-I--ER..ILTF-HF-T--...........RQ--T--.F--FK..........................KVVETQYK--KGPRR---------------..................H-R-...T--K-N-----C....   198
CPZ.US.85.CPZUS             K--I---M--K----------.TA------V--RP..-I-K---PT--...........-------.F---K..........................--VGQS-RR-.....---------------..................C---...--QPEN---S--....   192
                                                                                                                                                                                 Vif end
MAC.US.x.239                W-D-T-NY--I-L-ST--P--.TAGEV-R-IR-EQ..LLSC-RFPRA-..........KYQ-P---.----KVVSDVR..SQGEN.............PTWKQWRRDNRR.GLRMAKQNSRGDKQRGG................KP-T-G...ANFP-LAKVL-ILA..   214
H2A.DE.x.BEN                W-D-T-DC--I---ST--S--.TAGEV-R-IR-EK..LLSC-N-PQA-..........KAQ-P---.----VVVQQNDR.PQRKG.............TARKQWRRDHWR.GLRVAREDHRSLKQGGS...............EPSAPR...AHFP-VAKVLEILA..   215
H2A.GW.x.ALI                W-D-T-DC--S---GT--S--.TAGEV-R-IR-EK..LLSC-N-PQA-..........KSQ-P---.F---VVVQQNGK.PQRNS.............TTRKQWRRDYRR.GLRVARQDSRGLKQRGG...............ES-APG...AHFP-VAKVLEILA..   215
H2A.SN.x.ST                 W-D-T-DC--S---ST--S--.TAGEV-R-IR-EK..LLSC-N-PQA-..........KYQ-P---.F---VVVQQNGR.PQRDN.............TTRKQWRRNYRR.GLRVARQDGRSHKQRGS...............EP-APR...AYFP-VAKVLEILA..   215
H2B.CI.x.EHO                Y-D-T-DV--R-L-GS--SS-.TANEV-R-IR-EK..ILSH-N-PSA-..........TGQ-P---.F---RVVQEGKDGSQGES.............TTRKQRRRNSRR.GIRMARDNIR-SQQSSS...............QSLAQG...TYFP-LAEVL-ILA..   216
H2B.GH.86.D205              Y-D-T-DV----L-GS--S--.-ANEV-R-IR-EK..ILSY-N-PSA-..........EGQ-P---.F---RVVQEGKNGSQGES.............-TRKQRRRNSRR.SIR-ARKNNNRAQQGSG...............QPFAPR...TYFP-LAEVL-ILA..   216
H2G.CI.92.ABT96             Y-D-T--T----L-GS-----.TAGEV-R-IR-EQ..ILSC-N-PTA-..........KRQ-P---.F---QVVQKG.HGSKGES.............XTRKQRRGDNRR.GIRMARKNSXRDQQDSS...............QSFTQG...AYFP-LAKVL-ILA..   215
H2U.FR.96.12034             W-D-T-DC----L-GS--S--.TAGEV-R-IR-EQ..LLSC-NFPSA-..........RQQ-P---.F---RALQDGR..SKREN.............PTRKFRRRNCRR.GFQ-ARKNCERYQQGSS...............ATSSQG...TYFP-LAKVLAILA..   213
COL.CM.x.CGU1               E-DLSWQQ-VGYM-IIRGK-VRIEEEA---IR-L...PWNP-DF-V--...........LG-LK--.E-C-L.........................--REKERTHQ.....KMRVLKGTSHGCRTR................................   171
DEB.CM.99.CM40              .-ELT-D---H----R--P--.-QQ-V-Q-IR-E-..LLQQ-Q..QY-RG.....PEGKGPP---.-I--R.........................VAVKHG-VPTLT...-TTFHL-NNTH.DKAR.....VAAGHA.....RSKRGSPKTF-QRRTIWSLESLCR   212
DEB.CM.99.CM5               .-ELT-D---H----R--P--.-QQ-VTQ--K-EK..LLQQ-Q..QY-RG.....PEGKGPP---.----R.........................TAVKYG-VSTLT...-TTFHI-NSTN.DKAS.....TTTSHA.....GSKRGSFKTL-KRRVMGSLESLCR   212
DEB.CM.SIVdeb04CMPF3061     .-ELS----EH----R--P--.TQQ-V-Q-I--EK..LLTC-Q..QY-RG.....AGG-GPP---.-IC-R.........................VA-KYG-ATALT...-TNALV-ASSN.FQA......SATIHD....AGCKRRSHETLSK-RINWRLELLCR   212
DEN.CD.x.CD1                .-E---VT--EI--KDH-S--.TRQ-V-Q--R-EV..-LST-ARKKA-YHPR...TGPGREPP---.-IV-Q.........................-A-RYG-VTTIT...-TSISF-NGRNHSKAA.....TGSSMG......YHRGGKKTLQQRRDHW-MGSLLV   226
DRL.x.x.FAO                 C-LI--TT--RI---E--P--.T-Q---GV---YR..IQA.-YFPR--..........KGQ-----.----QAHIKYLKHGRK...............TVSGESSSGAN..GSRRGR-RRMATEES........LKNQ.PGGSITL-PRV....PLPSLEHLH-ALP..   217
GOR.CM.2004.SIVgorCP684con  I--L---T--RI------Q--.TE--V---IVXEK..LLCK-N-P---..........SQ--T--.----R.........................--VGRVRVR-......-F---Q--AA-T-SR.................S-QM...R---E--S-------....   192
GOR.CM.2007.SIVgor2139_287  I--L---T--RI------Q--.TE--V---IV-EK..LLYK-N-P---..........SQ--T--.----R.........................--VGKVRVR-......-F---Q--AA-T--R.................S-QM...R--KEN-S-------....   192
GOR.CM.2007.SIVgorCP2135con I--L---T--RI------Q--.TE--V---IV-EK..LLCK-N-P---..........SQ--T--.----R.........................--VGKVRVR-......-F---Q--AA-T-SR.................S-QM...R--KEN-ST------....   192
GRV.ET.x.GRI_677            K-E---GT--RM------N--.T-R--QQ-IR-EK..YTW.-TFKE--.........K..GQ-Q---.L---VAY....TNGIR..............KRSKRTFTRMAGNLGSRQGAMGRMATRHA........QGSK.RRSQKALWNEH....ANPSMELLCRGGKET   219
GSN.CM.99.CN166             --NLT-GV-M-I--GK--P--.QNED--R-IR-EQ..ILGN--HPEA-........LQVGTPQT-E.Q--FF.........................-YVKH.YG-NPT...QSSYALAQPKGAH-HG.......YRSK..RSVGNKRGGRETLLK-RAPW-LA....   237
GSN.CM.99.CN71              --NLT-DV-M-I--TR--P--.QNEDV-R-IR-EQ..LLGY-DHPEA-........PQVGTPHT-E.R--FF.........................-YVRY.YG-NPT...KS--TLDQPKGAH-HG.......SCSK..RHVGNKRRGRKTLLK-RAPW-LA....   237
LST.CD.88.447               Q-EI--I--N-M--CK--N--.-TRD-----W-EK..IMAF-AMPT--..........KG.C-L---.F-C-RQLQHVQN.................KAAKESRKTSRGLWATKGAMGSMASRHHGGNKR.............GGQTPFPRNNFGPSLRILCQQYRVR   215
LST.CD.88.485               Q-EI--I----M--CK--N--.-TRD-----W-EK..IMAI-AMPT--..........KG.C-L---.F-C-RQLQHVQS.................KAAKESRKTSRGLWTTKRAMGSMVSRHHGGNKG.............GGQTPFPRSHFGPSLGILCQQYRVR   215
LST.CD.88.524               Q-E---I----ML-VK----W.VHRD-VR-IR-EK..ILHS-KMPT--..........KG.C-L---.--C-RQLQHVQT.................KTAAQPRKA-RGLWSSKRAMGSMASRHNGGNQG.............RGKTPFPRSNFGPSLGILCQQYRVR   215
LST.KE.x.lho7               Q-EI--I----M--YK--N--.IQRD--R----ER..ITMC-Q.PL--..........KG.C-L---.FIC-RQLQHVQA.................EAAKKSRETTRGLWTTKGAMGSVVSRHHGGNKR.............GGKTPFPRSNVGPSLAILCRQYRVR   214
MAC.US.x.251_1A11           W-D-T-NY--I-L-ST--P--.TAGEV-R-IR-EQ..LLSC-RFPRA-..........KYQ-P---.----KVVSDVR..SQGEN.............PTWKQWRRDNRR.GLRMAKQNSRGDKQRGG................KP-T-G...ANFP-LAKVL-ILA..   214
MAC.US.x.251_BK28           W-D-T-DY--I-L-ST--P--.TAGEV-R-IR-EQ..LLSC-KFPRA-..........RYQ-P---.----KVVSDVR..SQGEN.............PTWKQWRRDNRR.GLRMAKQNSRGDKQRGS................KP-T-G...ADFP-LAKVL-ILA..   214
MAC.US.x.EMBL_3             LDRCNTR-CRHFSCIALISL-.TAGEV-R-IR-EQ..LLSC-KFPRA-..........RYQ-P---.----KVVSDVR..SQGEN.............PTWKQWRRDNRR.GLRMAKQNSRGDKQRGG................KP-T-G...ADFP-LAKVL-ILA..   214
MND-1.GA.x.MNDGB1           C-EIT--V--H----HFY---.M---VM--IR-EE..-LKV-RFP---..........KAQG-L---.F-C-RVIYGPEE.................R*ASIRRSR.....TSQRAIQSVASRYYGGNKG.............RSKKALPSHYPKCSIRILCAKHRE.   222
MND-2.CM.98.CM16            Y-HI--QT--RM--WK-LP--.TEQ---Q----KR..LTV.-YFHW--..........KG-----.----LSYTAYCNNG...................RRGPRD-R..RSRRGGSKY-VGKIS........GENQ.SGGKITL-PRV....PFPSLER-HRTLA..   212
MND-2.GA.x.M14              Y-HI--VT--RM--WE--P--.TAA-V--V-Y-ER..IVA.-YSPW--..........KGQ--T--.L---RAYIKFCRNG...................RKSTRGTT..GGRRSGTRTVAGKII........GTSE.QRGSITL-PRV....PFPSLEHLCRTLA..   212
MND-2.x.x.5440              F-YI--VT--RI--GE--P--.T-Q------F-ER..LVA.-YFPW--..........RGQ--T--.F---QAYLRGRKHG...................REGTRGAR..RSRRGRTGAMAGNVT........GENQ.PGGPVTL-PRV....PFPSLEH-CRTLA..   213
MNE.US.x.MNE027             W-D-T-DY--I-L-ST--P--.TAGEV-R-IR-EQ..LLSC-RFPRA-..........K-Q-P---.----RVVSYVR..SQREN.............PTWKQWRRDNRR.SLRMAKQNSRGDKQRGS................KP-T-G...ADFP-LAKVL-ILA..   214
MON.CM.99.L1                A-HLT-D--N-IV-TH--A--.QQLDV-R-IR-EK..LLGQ-QHL-T-.......YPKVGTPLT-E.K--FL.........................--VRY..GATPT...KSSAPLD-PDGSHR-T.......GPTG..RHLGAKRRSPKTLLK-RAPWHLGTGNG   252
MON.NG.x.NG1                A-YLT-DV---I--TR--S--.QA-DV-R-IR-EA..XLGN-HHIQT-.......YNKVGTPPT-E.K--FL.........................-CFKY..G-CTT...EPS-TLD-PRGSHRAR.......RSAG..RHMG-KRGGPXALLK-RAPWHLGSGH.   257
MUS-1.CM.01.1085            --YLS-T--MRI--TQ--S--.YSQD-PRVIR-QH..TLGD--HPDA-........TKVGPPP---.V---L.........................--KKH..G-LAP...QP--ALA-QSGPNHHA.......PGTR..SHVGSKRRGRKTLFQRRAPW-LE....   235
MUS-1.CM.01.CM1239          H-YLT--I-MKI--THH-S--.YSQD-PRVIR-QP..PLGD--HPGA-........TKVGPPPT--.V---L.........................T-KKH..G-STT...QPQTSMA-QSGPNYHA.......AGPT..GHVG-KRRGRKTLFQRRAPW-LE....   235
MUS-2.CM.01.CM1246          E-FL---VTM-I--TH--S--.QQRD-QR-IR-EQ..LLRK--HIKT-.......YPKVGTPL--E.K--FF.........................-Y-KHTNGASAS...KS--LMAKSSNAN-YG.......SGAS..RPVGTKRRGRKTLLQRQAPW-LG....   238
MUS-2.CM.01.CM2500          --YL-----M-I--T-H-S--.QEQD-IR--R-ER..ILGI-QH-TA-.......LPKVGTPP--E.K--FF.........................-Y-KQGDGATPS...KS--PMAKPPGTN-HG.......AGTK..RHVGSKRRGRKALLQ-RAPWDLGPSHG   242
OLC.CI.97.97CI12            M.YLT--YV-FMV--V--A-W.TQ------I--MP..DHRY-QLET--.....L....GVLSLQL-AL..........................RAVRNGRKSSRGSGATQGATGG-G-RHNGRN.............YCRGKKALSQGVIGPSFSILCSKTYQ.   203
RCM.CM.SIVagi-00CM312       F-DI--I---KM--NV--P--.T-Q--------EK..-LVCGF.-R--..........RDQ--T--.----QAWVKSQL.QKNG...RKS.........SRGPHW..GGRSRVSTMVGQNVVRDQPRSQ................ITLPWR...VHFPSVAYLC-ILA..   220
RCM.CM.SIVrcm-02CM8081      F-DI--IV--S---KI--P--.T-K---Q-I--EK..-LLCGF.-R--..........RDQ--T--.---IQAWAREQV.KKHG...RKS.........-RGPHW..GWRSRVPA-VGQNAVRNKSGSQ................VTLPSR...VHFPSLAYLC-TLA..   220
RCM.GA.x.GAB1               V-D-F-WT--S---KI--P--.T-R---QS---EK..-LVCAF.-G--..........RDQ--T--.F--IQAWAKSQL.DRYG...RKS.........PRGPHW..GWRSRVPA-ATGHARKGQLGSQ................VTLSSR...VHFPSVAHLC-TLA..   216
RCM.NG.x.NG411              F-DI--TV--K-V-NQ--P--.T-K---Q-I--EQ..-LLCGF.EK--..........RDQ-----.---IKAWARQQLKDGRG..QRKS..........-RGPHW..GWRCRVC--A-QNANRSQLRSQ................VALSSR...VHFPSVAYLC-TLA..   222
SAB.SN.x.SAB1C              F-EL--WT--HI--WS--P--.T-R-VQQ-IR-EK..YLW.-KH-V--.........QPTGQ-P---.----RVY....TNGLRRVAPTSRRGSSQGSPQESQRRDTRMARNMGFAQRA-RRMAPRHV........TGPQ.FRGPVPL-KES....PFPSLVEYC-RTSH.   234
SMM.SL.92.SL92B             W-D-T-DV----A-ST--P--.AAH-V-Q-IR-EQ..-LSY-G-AVA-..........HSS-Q---.L---KVVLQNDR.PKGKN.............PTRKQWRRNNRR.GLRMVTQYSGGPESNSS................TT-T-R...VNFP-LEKVL-ILGR.   217
SMM.US.x.H9                 W-DAT-DY--T-L-GT--P--.-EGEV-R-IR-EK..LLSC-KFPKA-..........K-Q-P---.----TVVSHVX..SQGEX.............PTWKQWRRNXRR.GXR-ATQNXRRNKQGSS................ESFAXG...ANFP-LAKXL-ILA..   214
SMM.US.x.PGM53              W-D-T-DY--T-L-GT--P--.-EGEV-R-IR-EK..LLSC-KFPKA-..........K-Q-P---.----TVVSHVR..SQRED.............PTWKQWRRNNRR.GIRMAKQNSRRNKQGSS................ESLAEG...ANFP-LAKVL-ILA..   214
SMM.US.x.SIVsmH635F_L3      W-D-T-DC--T-L-ST--P--.-EGEVQR-IR-EK..LLSC-KFPKA-..........K-Q-P---.----TVVSHVR..SQRED.............PTWKQWRGNNRR.GFRMAKQNSRRNKQGSS................KS-AEG...ANFP-LAKVL-ILA..   214
STM.US.x.STM                WSD-T-DC----L-GT--P--.TAGEV-R-IR-EK..LLSC-RFTKA-..........K-Q-P---.----KVVEHVR..SQREN.............TARKQWRRGNRG.SIRVATQNGRGHKPRGS................KPSTEG...TDFP-LAKVL-ILA..   214
SUN.GA.98.L14               Q-EI--LV---M--FK----W.TSRC--R-M--EK..ILHE-RN-VA-.........KG.L-L---.F-C-RVLHGQQE.................RASRTATRR.RGATE..RAICTMAGRYHGRNQGR.............SGK.AFPSSNSCPSLAILC-LYRVR   229
SYK.KE.x.KE51               --E-T-QT--A-V-SQ--P--.--R-VQQ--R-EK..LTSH-WN..F-KE..........Q-L---.----Q.........................KYLSKDGTGFLQ...S--AAARG-TILHS-KCQLGRGRSNKCHCQSRTGSPRSMQAFYS-RNIWSLESMLR   219
SYK.KE.x.SYK173             .-DLT-AV--R---NF--P--.TAR-VNQ-VR-EL..LTSH-WT..P-TD..........Q-P---.----Q.........................VYL.KDGGGFLQ...S--ACARN-MVLHS-KCRVDPKR.DQCHCKGRTGSDRSIQAFYSSRNIWSLESILK   216
TAL.CM.00.266               .-E-T-NV--H---TK--P--.TA--V-Q-IR-EV..LLAV-G..VR-KG.........RDTPP---.F---T.........................-AVRYG-VATQS...SSTADLQTHSN.TKAS.....ARKAHE.....RSSRRNAYTL-KRRGHWSLESLCR   222
TAL.CM.01.8023              .-ELT--V--H---TK--P--.TA--V-Q-IR-EV..LLAV--..VR-KG.........RDSPP---.F---K.........................VAVRHG-VATQS...SSANNI-AHTN.TKSS.....IRKTHE.....RSKGRDAHTL-KRRGHWSLESLCR   222
TAN.UG.x.TAN1               K-E-T-DV--HM--CH--P--.T-R--QQ-IR-ES..FLW.-T-KE--.....VAENHWGQ-R---.F---TVYTDFLRNGRR..............KRFQGK-TRMVRNLGSQQGA-GRMIKRHG........SRTQ.SGSTTPFWER-....PLPSMELLS-RRGKE   229
VER.DE.x.AGM3               R-EL--AT--S---TH--T--.TER--------QR..FTF.-QFPE--.........KKTGQ-P---.----LAH....QNGLR..............QRSQRS-TGGTRNMGFEQGA-GRMAKRHA........RRYQ.SGSQDAFWARA....PVPSMELLS-GGRKE   223
VER.KE.x.9063               R-EL--AT--GM---H--S--.TER--Q--IR-ER..FVF.-QFPE--.........KTTG--P---.----LAH....QNGIR..............ERSQRKQARRSRDLGSKQRTMGRMAKGHV........RRSQ.SGSQITFW-RA....PLPSLELLS-GRRET   222
VER.KE.x.AGM155             R-EL--NT--SM--CH--T--.T-R--QQ--R-NR..FIF.-QFPG--.........KLTGQ-P---.----LAH....QNGLR..............KRSQRGETRRTRNLGSQQGA-GRMAQRYG........RRNQ.QRSQTAFWPR-....PIPSMELLS-GRGET   223
VER.KE.x.TYO1_patent        R-EL--AT--SI--GH--N--.KER--QQ--R--R..FVF.-QFPE--.........KSTGQ-P---.----LAH....QNGLR..............ERSKRG-TRRSRNLGSKQGA-GRMAKRYV........TRSQ.PGGEAAFWER-....PVPSMELLS-GRRKT   223
WRC.CI.97.97CI14            E-FIM-SE-NKI--SQH-E-W.TQNVV-NTIRTGKIN-IDA-Q-HK--...........AG-LT--T.-C-QV..........................LQHEQLQRAGA.FRRT-TNSRGPS-YARALCNG.............KTAFPKGSPCKSISLLLSTERV.   207
WRC.CI.98.98CI04            E-FIM-SE--VM--SQHV--W.TQNVV-DTIRSGRLDIRNA-Q-HK--...........AG-LT--T.-C-QV..........................LQHEQLQGTST.YRRT-TNSRGPS-YARTLCNG.............KTAFPKGSPCKSISLLLATEKL.   207
WRC.GM.05.Pbt_05GM_X02      E-FITVK---V---SQH-E-W.TQ--VKNTIRQGYLHPIDS-K-HQ--...........LG-LT--A.-C-QA..........................LQHEQLRKRKAPNRRTPGNTRGSS-YARAVCHG.............KTAFPPGSPCKSISLLLATEKL.   208
```

```
H1B.FR.83.HXB2              ......................................*  192
H1A1.UG.85.U455             ......................................-  193
H1B.US.90.WEAU160           ......................................-  193
H1C.ET.86.ETH2220           ......................................-  193
H1D.CD.84.84ZR085           ......................................-  193
H1F1.BE.93.VI850            ......................................-  193
H1G.SE.93.SE6165            ......................................-  193
H1H.CF.90.056               ......................................-  193
H1J.SE.93.SE7887            ......................................-  193
H1K.CM.96.MP535             ......................................-  193
H101_AE.TH.90.CM240         ......................................-  191
H102_AG.NG.x.IBNG           ......................................-  193
H103_AB.RU.97.KAL153_2      ......................................-  193
H104_cpx.CY.94.CY032        ......................................-  192
H1N.CM.95.YBF30             ......................................-  193
H1P.FR.06.RBF168            ......................................-  196
CPZ.CD.90.ANT               ......................................-  186
CPZ.CM.-.SIVcpzDP943        ......................................-  193
CPZ.CM.-.SIVcpzMB897        ......................................-  193
CPZ.CM.01.SIVcpzCAM13       ......................................-  194
CPZ.CM.05.SIVcpzEK505       ......................................-  193
CPZ.CM.05.SIVcpzLB7         ......................................-  193
CPZ.CM.05.SIVcpzMB66        ......................................-  193
CPZ.CM.05.SIVcpzMT145       ......................................-  193
CPZ.CM.98.CAM3              ......................................-  193
CPZ.CM.98.CAM5              ......................................-  193
CPZ.GA.88.GAB1              ......................................-  194
CPZ.GA.88.GAB2              ......................................-  193
CPZ.TZ.00.TAN1              ......................................-  199
CPZ.TZ.2001.TAN2            ......................................-  199
CPZ.TZ.2002.TAN3_1          ......................................-  199
CPZ.US.85.CPZUS             ......................................-  193

MAC.US.x.239                ......................................-  215
H2A.DE.x.BEN                ......................................-  216
H2A.GW.x.ALI                ......................................-  216
H2A.SN.x.ST                 ......................................-  216
H2B.CI.x.EHO                ......................................-  217
H2B.GH.86.D205              ......................................-  217
H2G.CI.92.ABT96             ......................................-  216
H2U.FR.96.12034             ......................................-  214
COL.CM.x.CGU1               ......................................-  172
DEB.CM.99.CM40              LTRPTRLDGR..TSMGCIGH..................-  231
DEB.CM.99.CM5               PSCPTRLDRR..TSMGGVSN..................-  231
DEB.CM.SIVdeb04CMPF3061     VTSRARVDRR..TGLGSLSNRLHKEGPENDLATSGRRVLS-  251
DEN.CD.x.CD1                ITSAPTLDRR..TSHGSRRHRFHKNHAERNMEAL......-  259
DRL.x.x.FAO                 ......................................-  218
GOR.CM.2004.SIVgorCP684con  ......................................-  193
GOR.CM.2007.SIVgor2139_287  ......................................-  193
GOR.CM.2007.SIVgorCP2135con ......................................-  193
GRV.ET.x.GRI_677            ......................................-  220
GSN.CM.99.CN166             ......................................-  238
GSN.CM.99.CN71              ......................................-  238
LST.CD.88.447               GESLHEIHYHSK..........................-  228
LST.CD.88.485               GGSLHEIHYHSK..........................-  228
LST.CD.88.524               RGGLHEIHHFS...........................-  227
LST.KE.x.lho7               RGGLHEIHHPSQ..........................-  227
MAC.US.x.251_1A11           ......................................-  215
MAC.US.x.251_BK28           ......................................-  215
MAC.US.x.EMBL_3             ......................................-  215
MND-1.GA.x.MNDGB1           ......................................-  223
MND-2.CM.98.CM16            ......................................-  213
MND-2.GA.x.M14              ......................................-  213
MND-2.x.x.5440              ......................................-  214
MNE.US.x.MNE027             ......................................-  215
MON.CM.99.L1                AAC...................................-  256
MON.NG.x.NG1                ......................................-  258
MUS-1.CM.01.1085            ......................................-  236
MUS-1.CM.01.CM1239          ......................................-  236
MUS-2.CM.01.CM1246          ......................................-  239
MUS-2.CM.01.CM2500          APSRP.................................-  248
OLC.CI.97.97CI12            ......................................-  204
RCM.CM.SIVagi-00CM312       ......................................-  221
RCM.CM.SIVrcm-02CM8081      ......................................-  221
RCM.GA.x.GAB1               ......................................-  217
RCM.NG.x.NG411              ......................................-  223
SAB.SN.x.SAB1C              ......................................-  235
SMM.SL.92.SL92B             ......................................-  218
SMM.US.x.H9                 ......................................-  215
SMM.US.x.PGM53              ......................................-  215
SMM.US.x.SIVsmH635F_L3      ......................................-  215
STM.US.x.STM                ......................................-  215
SUN.GA.98.L14               KGRVYESNHFAQ..........................-  242
SYK.KE.x.KE51               GGSRSRRNTH....HGLD....................-  234
SYK.KE.x.SYK173             .....RR.........GRD...................-  222
TAL.CM.00.266               IASRP.................................-  228
TAL.CM.01.8023              VTSRA.................................-  228
TAN.UG.x.TAN1               WGTNDRKGL.............................-  239
VER.DE.x.AGM3               SHSHARKGL.............................-  233
VER.KE.x.9063               GSPNDGEGL.............................-  232
VER.KE.x.AGM155             GKTHSGKGI.............................-  233
VER.KE.x.TYO1_patent        WYSHDGKGLQIL..........................-  236
WRC.CI.97.97CI14            ......................................-  208
WRC.CI.98.98CI04            ......................................-  208
WRC.GM.05.Pbt_05GM_X02      ......................................-  209
```

```
                        Vpx start                                                                                                   Vpx end
MAC.US.x.239            MSD..PRERIP..PGNSGEETIGEAFEWLNRTVEEINREAVNHLPRELIFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGW.RPGPPPPPPPPGLA*   112
H2A.DE.x.BEN            -T-..----V-..----------------E--I-AL--------------------R---------A--T-------M---I-T-F-R--T-W--DM-RE-LEDQ----------V-   114
H2A.GW.x.ALI            -AN..---TV-..----------------D----AL--------------------R---------Q--T-------M---MYT-FM---T---G---P---.-S----------V-   113
H2A.SN.x.ST             -AG..---T--..----------------D----A---------------------R---------I--T-------M---M-I-S-R--T---G---P---.-S----------V-   113
H2B.CI.x.EHO            -.-..----V-..---------V-------ET-L-HL--V---------------K--A--RE-----I--T-------M---M-I-FA---G--R----P---.-S------------   112
H2B.GH.86.D205          -.-..----V-..----D---V----A--E--IT-L--V------------------A--RE-----I--T-----L-M---M-V-YT------Q----P---.-S------------   112
H2G.CI.92.ABT96         -.-..----X-..-E--D-----------E--XT----I--------------R-------------A--T-----L-M---M-V------T--QK---P---.-Q------------   112
H2U.FR.96.12034         -G-..-------..----E---V----S--E---RD---A--Q---------------Q--R------E--T--------------------Q-------P---.-S----------I-   113
DRL.x.x.FAO             -AERQSV--A-AE-MGA--...V-LE---Q-SLLR--Q--RL-FHP-FL-RL-NTCM-HY--ALQL-FT-S-----L-L---M---FQQ--S--QGR-PPPLRPAGDRL-----PP.-   114
MAC.US.x.251_1A11       ---..-------..-------------------------------------------------------------------------------------.--------------   113
MAC.US.x.251_BK28       ---..-------..--------------------------------------------------Q----------M-----------------------.--------------   113
MAC.US.x.EMBL_3         ---..-------..-------------------------------------------------X...................................................   63
MND-2.CM.98.CM16        -AE.....GA-EI-EGA--...VDLNT--E-SL-K--Q--RL-FHP-FL-RL-NACI-H---RHQR-L--A-----L-MN--M-T-MQQE-P-RSG.-P.....-.-...-----MV-   100
MND-2.GA.x.M14          -AE.....-A-EA-QGA--...VGLEQ--E-SL-QV----QL-FHP-FL-RL-NTCV-H---RLRRTLN-A-----L-M---M-V-MQQ--P-RSG.-S.....-.-...--------   100
MND-2.x.x.5440          -AE.....-A-EA-EGA--...VGLEQ--ETSL-R-----RL-FHP-FL-RL-NTCV-H---RHQR-LD-A-----L-MH--MYT-MQQ--P-RNG.RPR.....-.-...-----M--   100
MNE.US.x.MNE027         ---..-------..-------------------------------------------------------------------------------------.--------------   113
RCM.CM.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.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.GAB1           -AE.....-A-..EVPT-AGEAEFQPWLRDMLEKVNLEA.RL-FHP-F--RL-RTCV-H---VHQR-LE-AA----L-M-----I--QT--SQR...--PNPR.AV-ERITIL--M.-   105
RCM.NG.x.NG411          -AE..G---V-..EAPT-AGDVEF-PWLHRMLT-VNLEA.RL-FHP-F--RL-RTCV-H---RL-R-LE-AG----L-M-----I--QS--SQR...--Q-QAREA-ERIQIL--M.-   109
SMM.SL.92.SL92B         -T-..-------..----------------HN---AL-QT--Q-----------R-C----V----Y----A----VQ-M---M-Q-FR---T-R----SQ---.-T------------   113
SMM.US.x.H9             -XX..---X--..-------XX--X-D--H--X-----A-----X--------R-X-------M---V--T-X---------------------GE------.--------------   113
SMM.US.x.PGM53          ---..-------..---------------------------A-------------R------R--M---E--T-------------V---R-------E------.-S------------   113
SMM.US.x.SIVsmH635F_L3  ---..-------..---------E-----------G---A------------------------M---E--T-----------------------------------.-T------------   113
STM.US.x.STM            ---..-------..-------A-E------H----D------------------R------------PG-----------------S-R--------------.--------------   113
```

```
                                                                      Vpr end in HXB2
                                          amphipathic α-helix     frameshift in HXB2
          Vpr start              oligomerization                          H(S/N)RIG motifs                Vpr end
H1B.FR.83.HXB2            MEQA.................PEDQGP........QREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYE.......TYGDTW.AGVEAIIRILQQLLFIH#FRIGCRHSR......IGV.........TRQRRARNGASR...........S*   96
H1A1.UG.85.U455           ----.................------........----YA-*A--------H-D------Q----------N.......------.E-----------------.-----Q---......--I.........IPG--G------...........--   96
H1B.US.90.WEAU160         ----.................------........----Y------------S--------V---S--------.......------.T-----------------.---------......--I.........-P----------...........--   97
H1C.ET.86.ETH2220         ----.................------........----Y---A--------Q--------P---N---Y---.......------.S----L--T----M---.-----Q---......--I.........L-----------...........--   97
H1D.CD.84.84ZR085         ----.................-A----........----Y------------S-----------S---Y---.......------.-----------------.-----Q---......--I.........-P-------S--...........--   97
H1F1.BE.93.VI850          ----.................-G----........----Y---A--I-------------P----------N.......------.E-----------------.---------......--I.........VP---V------...........--   97
H1G.SE.93.SE6165          ----.................------........----Y---A-----------------L----------N.......------.E-----------------.-----Q---......--I.........-PR--V-D-PG-...........--   97
H1H.CF.90.056             ----.................------........-----------------I---------V---Q------N.......------.V----L--T---------.-----Q---......--I.........-----V---P--...........--   97
H1J.SE.93.SE7887          ----.................---H--........----YH--------------------P---S---Y--S.......------.E-----------------.-----H---......--I.........IP---G------...........--   97
H1K.CM.96.MP535           ----.................------........----N------I-----R--------P---N------T.......------.E-L---------------.-----H---......--I.........IP---G---S--...........--   97
H101_AE.TH.90.CM240       ----.................--##--........----Y---------------------P--Q----Y--N......N-----.E-------M-----V-.-----Q---......--I.........MPG--G---TG-...........--   95
H102_AG.NG.x.IBNG         --R-.................------........----S------------H--------P-----------.......------.E--K----------V-.-----Q---......--I.........IQG--G----G-...........--   97
H103_AB.RU.97.KAL153_2    ----.................------........----Y------------S--------V---S---Y---.......------.-----------------.-----Q---......--I.........IQR---------...........--   97
H104_cpx.CY.94.CY032      ----.................------........----N---------------------P----------N.......------.E-----------------.-----Q---......--I.........-P---..-R-RGWS..........Q-   96
H1N.CM.95.YBF30           --R-.................---A--........----Y---A-----------------------------N.......------.E-----------------.Y---Q---......--I.........-P---.---T--...........--   96
H1P.FR.06.RBF168          ----.................------........P---F---M-QT-----A-------IQ---S------N.......------.E--T---------I---.Y----H---......--ILPPS.......R-GR-H-P--...........--   99
CPZ.CD.90.ANT             ----.................-Q-E--........----M---L--T---I---------QPT-QH--NWV-A.......N---S-.R--QVL-IL--KA--T-.Y-H--A---......--H................-PR-...........R-   89
CPZ.CM.-.SIVcpzDP943      ----.................------........A---Y---A--------------------A----------D.......------.-----------------.Y---Q---......--INP.........G--.------...........--   96
CPZ.CM.-.SIVcpzMB897      ----.................------........----Y--------D---Q--------Q---D------N.......------.-----------------.Y---Q---......--ILP.........-G-R---SN-...........--   97
CPZ.CM.01.SIVcpzCAM13     --H-.................------........P---Y---A--I--D------K-Y--S---Q------T.......------.E-------------V-.--V--H---......--IIP........QR--.A---N-...........P-   97
CPZ.CM.05.SIVcpzEK505     ----.................------........----Y---A-----D-----L-----P--------YF-N.......------.E-------T---------.Y---Q---......--ITP.........---.------...........--   96
CPZ.CM.05.SIVcpzLB7       ---V.................------........----Y---A--V------Q--------P---Q---Y--D.......------.E--Q-------------.Y---Q---......--ITP.........---R-D--N-...........--   97
CPZ.CM.05.SIVcpzMB66      ---L.................------........----Y------I-----R--------D---Q------T.......------.-----------------.Y---H---......--ININ.......Q---R----N-...........--   99
CPZ.CM.05.SIVcpzMT145     ----.................------........S---YEQ-L-DT---I-Q--------P---A--E---.......-----S-.---Q---------V-.-----H---......--INP.........---R-D-TN-...........A-   97
CPZ.CM.98.CAM3            ----.................------........P---Y---A--------------------A-------F--D.......------.-----------------.--L--Q---......--I.........IPG--.------...........--   96
CPZ.CM.98.CAM5            ---I.................------........P---Y---X--I-----T--L-----P---R---Y---.......------.I---------H---V-.--L--Q---......--IL........XER--.--T--...........X-   96
CPZ.GA.88.GAB1            ----.................------........P---YQ--A--T-------------P---Q---F--D.......------.V---------H------.--L--Q---......--IL........PQR--.S--SN-...........--   97
CPZ.GA.88.GAB2            ----.................------........AG--Y---A--I--D------------P---Q---L--N.......N-----.E-----------H---V-.-----Q---......--IE........RQR--R-D-SY-...........--   98
CPZ.TZ.00.TAN1            ----.................-N-N--........----YT--L-DI---I-Q---K-----PI-Q-V-NWVFT.......I---S-.E--QEL-K---RA--T-.Y-H--I---......--..................................--   84
CPZ.TZ.2001.TAN2          ----.................--NH--........----Y---L-DT---I-Q--M-----SV-Q-V-NWV-A.......IH--S-.E--QEL-----RAI-T-.Y-H--V---......--..................................--   84
CPZ.TZ.2002.TAN3_1        ----.................---H--........----Y---L-DT---I-Q---------PI-Q-V-NWVFT.......IH--S-.E--QEL-----RA--T-.--H--V---......--..................................--   84
CPZ.US.85.CPZUS           ---V.................------........A---Y---A-DV------E--L-----P--Q-----Y--S.......------.E-----------------.Y----N---......--ISLA....RR-P-G-.---S--...........--  101
          Vpr start                                                                                                                                              Vpr end

MAC.US.x.239              --ER................P--NE--.............----WD--VV-V-----E--LK--DPRL-TA--N---N.......RH---L.E-AGEL-----RA--M-.--G--I---.........--QPG.....GGNPLSAIPPSR-M...........L-  102
H2A.DE.x.BEN              -TE-PT............EFP---GT-.........R-DLGSD-VI-T-R-I-E--L---DPRL-IA--YY-HN.......RH---L.E-AREL-KT--RA--V-.--A--NR--.........--*TR.....RR-PCPA-PTPRAM...........H-  105
H2A.GW.x.ALI              -TE-PT............EFP-AGM--.........HQGARD--VI-V-R-I-E--L---DPRM-IA--GY--T.......RH---L.ERAREL-NA--RA--M-.--A--GR--.......V-QTR.....GRNPLSAIPTPRNM...........Q-  106
H2A.SN.x.ST               -TE-PT............ESP---RT-.........P---GD--VI-T-R-I-*--LK--DPRL-IT--NY--A.......RH---L.E-ARGL-----RA-LL-.--A--GR--.........--QPR.....GRNPLSAIPTPRGM...........R-  105
H2B.CI.x.EHO              -AE-VP............EIP---KN-.............----WEQ-VVDV---I-Q--LK--DPRL-TA--NF--N.......RH-N-L.E-AGEL-KL--RA--L-.--G--Q---.........--QPG.....GGNPLSAIPPS.............-  102
H2B.GH.86.D205            -AE-AP............EIP--NEN-.............----WE--IG-I---I-Q--LK--DPRL-TA--NF--S.......RH---L.--AGEL-K---RA--L-.--A--Q---.........--QSG.....GGNPLSTIPPP.............-  102
H2G.CI.92.ABT96           -AEEI..............P---EA-.............----WD--VV-V---I-E--LK--DPRL-TA--NY--D.......RH---L.E-AGEL-K---RA--L-.--GD--C--.........--GNA.....GGNPLSTIPPSRGV...........F-  103
H2U.FR.96.12034           -AEI...............P---GA-.............----GD--VR-I-----E------DPRL-TA--NY--N.......RH---L.E-AGEL-K---RAI-L-.--G----------.....--SSR.....GGNPLTTVPPS.............-   98
COL.CM.x.CGU1             -SSGPK.........REREDP---ES-.........E-D.FT.-LQDQV-NITR--R----LTHISVIMWGCTRR......HW-RP-.TAGL-YV--MECA-YL-VHAGQ--FPS......TQN...........................P-   93
DEB.CM.99.CM40            --RY...............P-SHPPH.FTSRTIPMTRL...GLQQAMQD-NE--LK--T-EE-F-VWN-CVDLPAQPDW-GEQA-A-S-IDY-K-V-KM-WL-.L-EA-F-RERE.....ATRG...YPNIRPLTGRNREVRD....G......E-  118
DEB.CM.99.CM5             --RY...............P-SHPPH.FTSRTVPMTRL...ALQQAMQD-NE--LK--T-EE-W-VWN-CVDLPAQPDW-GEQA-A-S-IDY-K-V-RM-WL-.L-EA-F-RERE.....ATRR...YPNIRPLTGRNREVRD....G......E-  118
DEB.CM.SIVdeb04CMPF3061   --RP...............P-SHPLT.PLSRLVPISRL...QLQYMMQDVNE--MK--Q-DE-I-VWNYCV-LPAEPGW-GEQA-A-S-IDY-KKV-KMIWL-.LAE--Y-R-AA.....ASRH...YPNIRPLSGRNQEVRD....Q......E-  118
DEN.CD.x.CD1              -AE............................TTQR...QLQ-AVWDITE--RK--SKEEIT-IWD-CWSLPALPHW-E-QVMA-A-IDF---M-KEIWK-.Y-V--F-REAE.....R-RH...YPNIRPL-PR-EEP...............-   99
DRL.x.x.FAO               --R................P---E--.........P---YD--LVDT-V-IQE--LK--D-RL--NV-SWV--.........R---SL.E--HQL-IL--RA--L-.--Y--QE--.......L-PAG.....RWNPF-TFPRANQQ...........L-  101
GOR.CM.2004.SIVgorCP684con ---I.................-T-E--.............----F------------Q--------P---Q----V-D.......-------.E--T--------VI-V-.Y----H----.......--ILPP....S...R-GR---PN-...........--   99
GOR.CM.2007.SIVgor2139_287 ---I.................-T-E--.............----F------M-----Q--------P---Q----V-D.......-------.E--T--------VI-V-.Y----H----.......--ILPP....S...R-GR---PN-...........--   99
GOR.CM.2007.SIVgorCP2135con ---I.................-T-E--.............----F------V--X--Q--------P---Q----V-D.......-------.E--T--------VI-V-.Y----H----.......--ILPP....S...R-GR---PN-...........--   99
GRV.ET.x.GRI_677          -ASG..............RD-REPL-GWLEIWDLD---WD--LQDM-RD-NE--R---GMNM-IRVWNYCV-E.GR...RHNTP-NEIGYKYY--V-KSM-V-.--C---RRGPFSPYEERRNGQ....GGGAPPP...PPGL...........A-  119
GSN.CM.99.CN166           --RI...............P-SHPM-.WLSRRVPTTMA...IAQNA-W-INE--EK--S-DE-R-IWHDVT-LPADPDW-VDQAAI-CAIDY--.V-T---R-.-KD--F-RVN.LSGRYPAIRPSRGTAPPDSNSVSHADPEQPRRPSRYRMDE-  134
GSN.CM.99.CN71            --RT...............P-SHPL-.WISRRVPTTMA...VAQSAMW-INE--EK--S-EE-R-IWHDVT-LPADPDW-VDQAAI-CAIDYV-RV-T---R-.--D--F-RYNRIVRRYP-IRPLRGTAPPDSNSVPHADPEQPRRPSRYRMDE-  136
LST.CD.88.447             -SRTRQQ........RSPEKP---F-.........P---WGQ-LADTM--I-E--R----ETI-AQVSEYCVS.......ST-SEE.KACMKF-T-VNRA----.LPT.-PSYA......AARAG..TAHP.-EAAPRPTKRDIQRGR.......Y-  115
LST.CD.88.485             -SR-RQQ........RSPERP---F-.........P---WGQ-LADTM--I-E--R----EAI-AQVSEYCVS.......ST-SEE.EACMRF-T-VNRA----.LPT.-PSYA......ATRAG..TAHP.-EAAPKPTERDIQRGR.......Y-  115
LST.CD.88.524             -SRQRQQ........RSPERP---F-.........P---WAQ-LADTM--I-E--R----EAI-AQVSEYCVS.......ST-SEE.EACMRF-TLVNRA-YL-.LPQ.-PSYA......AARAG..TAHP.-EAAPRPT-RDIQRGH.......Y-  115
LST.KE.x.lho7             -SRQRQQ........RSPERP---F-.........P---WGQ-LADTM--I-E--R----EVM-AQVS-YCVD.......ST-SEE.EACMKF-TLVNRA--L-.LPT.-PSYA......AARAG..TAHP.-EAAPRP-EGDIQRGS.......D-  115
MAC.US.x.251_1A11         --ER................P--NE--.........*---WG--VV-V-----EG-LK--DPRL-TA--N---N.......RH---L.E-AGEL-----RA--M-.--G--I---.........--QPG.....GRNPLSAIPPSR-M...........L-  101
MAC.US.x.251_BK28         --ER................P--NE--.............----WD--VV-V-----E--LK--DPRL-TA--N---N.......RH---L.E-AGEL-----RA--M-.--G--N---.........--QPG.....GGNPLSTIPPS.............-   98
MND-1.GA.x.MNDGB1         -G-KRD............EQVS-----.........P---Y-Q-LADTM--I-E--RK---L-I-NAVSEYCVQ.......NT-SEE.EAC-KF-TLMNRAIWV-.LAQ--DGTF......RERRP...QLP..PSGFRPR-DRL.............-  105
MND-2.CM.98.CM16          ---................P---EA-.............----Y---LIDT-A-VQE--LK--D-RL--AV-SW--K.............-----L.E--RQL-S---KA-SM-.Y-Y---E--.........--QEGG....GCYPL-SFPRSDNP...........L-  102
MND-2.GA.x.M14            ---................P---EA-.............----Y---LIDT-A-IQE--LK--D-RL--AV-SW---.............-----L.E--QKL-TL--RA--L-.--H---E--.........--QEGG....KYNPL-SFPRPNNP...........L-  102
MND-2.x.x.5440            ---................P---EA-...........--G-Y-K-LVGT-A-IQE--LK--D-RL--AV-SWV--.........QQ---L.E--QQL-S---RA--L-.--Y---G--.........--QEGG....RYNPL-SFPSPNNP...........L-  102
MNE.US.x.MNE027           --ER................P---E--.............----WD--VV-V-----E--LK--DPRL-TA--NY--D.......RH---L.E-AGEL-K---RA--M-.--G--N---.........--QSG.....GGNPLSTIPPSR-M...........L-  102
MON.CM.99.L1              ---P...............PQSHPLH.WTSRMVPIERQ...ALQAAIW--NE--LK--S-EE-R-IWEQVT-LPADPAWNADQA-A-CAIDYT-WV-TI-YR-.Y-E--Y-RYAEQIRRYP-LRPMRGTAPGPTSSVPQADPDNPRRPSRYRMDE-  136
MON.NG.x.NG1              --N-...............P-SHPPH.WTSRVVPIERA...ALQAAIWD-NE--LX--SKEE-R-IWDQVT-LPADPGWNADQA-A-XAIDYT-RV-TXI-R-.M-D--Y-RHARVXRRYPXLRPLRGTAPGPTSSVPHADPERPLHPSRYHHDE-  136
MUS-1.CM.01.1085          --SL...............P-SHPA-.WHSRVVPTTMR...QAQ-AMW-VNE--EK--S-EE-R-IWNEVT-LPADPDW-VDQAAI-CAIDYV-RV-T-V-R-.--D--Y-RYNRVMRRYPTIRPLRGTQPPPSNSTP-ADPVPPLSPSRYRMDE-  136
MUS-1.CM.01.CM1239        --RV...............P-SHRP-.WHSRVVPTTMQ...QAQQAMWD-NE--EK--S-EE-R-IWNDVT-LPADPNW-VDQAAI-CAIDY--RT-T---R-.Y-E--Y-RYSNTIRRYPNIRPLRGTQAPPSNSMP-ADPTPPLRPSRYRMDE-  136
MUS-2.CM.01.CM1246        ----...............P-SHPLS.WLSRQTPTTMV...QAQAA-W--NE--EK--S-DK-R-IWDRVT-LPADPQWSVDQAAITCAIDY-KSV-T---R-.Y-D--Y--YAQTIRRYPPLRPLRGTQPPDSNSMP-ADPTPSLRPSRYRMDE-  136
MUS-2.CM.01.CM2500        ---H...............P-SHPL-.WLSRQVPTTMA...QAQNAMWD-NE--EK--S-DE-R-IWDRVT-LPADPNW-VDQAAI-CAIDY--WV-G---R-.--E--Y-RYSQVIRRYPQVRPLRGTQPPDSNSTP-VDPFPPRRPSRYRMDE-  136
OLC.CI.97.97CI12          -A...............ERV-----------.........P---LED-VRDTM--IIV--K-----EL-VRVFLFCAQR......HIN-EM.STSLLF--L--RA--Q-.Y-N--PGVE......LRRNP............H-RSEGMQE...........-   97
RCM.CM.SIVagi-00CM312     --I................PQ--E--.............----YDK-L-DT-M--QE--AK--TMEL-QAV-NYS--.........QH-ESI.E--KKM-TL-NRA--L-.Y-H-WSG--.......V-TAR.....GRNPLQSIPISRNT...........L-  101
RCM.CM.SIVrcm-02CM8081    --M................P---E--.............----M---L-DT-M--QE--AK--TMEL--AV-NY---.........Q---SL.E--RVL-TL--RA--L-.Y-H--TG--.........--TNR.....GGNPLQSIPLSRNI...........L-  101
RCM.GA.x.GAB1             --L................P---E--.............----YD--LMDT-I--QE--KK--TYAL-TQI-DYV--.........QH--SI.E--Q-M--L--RA--L-.--N--AG--.........--TSR.....GSNPL-SIPQTRNI...........M-  101
RCM.NG.x.NG411            --M................L---EE-.............----YD--LMDT-I-IQE--KK--TPEL-TQE-NY---.........QH--SL.E--K-M-TL-NKA--L-.--H--EG--.........--GAR.....GGNPL-SILHSRNI...........L-  101
SAB.SN.x.SAB1C            -ASGGWLPPVGGDPPKDPPKN-REEI-GWLETWDLP---FD--LRDM-QD-NS--QC----NL-FR-WWN-V-EPAI...DH-Q-RLE-WYKYC----KA--V-.MKGR-CKPKTHPAYGP-AGGPPPGLGGASGGA-SAAPG............L-  141
SMM.SL.92.SL92B           --H-.................---ETN........P---WD--IRDV-----E--LK--DPRL-TA--NYV-D.............-----I.E-AGE--K---KA--L-.--H--T----.........--QPR.....GPNPLGSIPSARDV...........L-  101
SMM.US.x.H9               -TER...............P---EA-.............----WD--XV-V---X-E--LNX-DPRL-XA--NY--D.......RH---L.E-AGEL-----GA----.--G----------.....--QSG.....GXNPLSTIPPXRGV...........L-  102
SMM.US.x.PGM53            -TER...............P---EA-.............----WD--VVKV---V-E--LK--DPRL-TA--NY--D.......RH---L.E-AGEL-----RA----.--G--A---.........--QSR.....GGNPLSTIPPSRGV...........L-  102
SMM.US.x.SIVsmH635F_L3    -AER...............P---EA-.............----WD--VV-V---I-E--LK--DPRL-TA--NY--D.......RH---L.E-AGEL-----RA----.--S--A---.........--QSR.....GGNPLSTIPPSRAM...........L-  102
STM.US.x.STM              -THR...............P---E--.............----WD--VV-V---I-Q--L---DPRL-SA--NY--D.......RH---L.E-AGEL-K---RA--M-.--G----------.....--QPG.....GGNPLATIPPTRGV...........L-  102
SUN.GA.98.L14             -ASRRE..........PVEQP---E--.........P---FAQ-LADTM--IRD--E----AAIVAQVL-YCAD.......ST-SEK.E-CMRA-TL-NRA----.LPM.-PSYI......RIRSGSGNVNPRPPT-PRPGQGDIRRGLQ......D-  117
SYK.KE.x.KE51             -A--FFRVYQR.........Q-GEPLF.YIPRNVNWD-AGVISATAKA-QVAQ--CK--TAEEIF-VWNQCL-VEAGPEE-PTMA-IRCMLDMH-A-NFM--E-.-AA--PQRTKY.....ARHRG..YPHP......................--  116
SYK.KE.x.SYK173           -AE-FFNPSQH.........VQGTPWF.FIPRNVELTPN.VINVTVKA-LVVT--SK--TPQEIY-VWNQSLNEEAG.TDSPTMA-ERTMLDMV-A-NLM--E-.-AA--PQRTRY.....ARHRG..YPHP......................--  114
TAL.CM.00.266             --RL...............P-SHPL-.LTSRLIPTPR-...QLER-MR-VRD-MLI--T-DEVI-VWN-CV-LPADPDW-GEQA-A-S-IDFA-KGN-M-LL-.-QQ--Y-EQQRRTPHYPNIRPLSGRGERTM...................Q-  116
TAL.CM.01.8023            --RL...............P-SHPL-.ITSRLIPTPRA...QLER-MR-VRD-MLI--T-EEVI-VWN-CV-LPAEPDW-GEQA-A-S-IDFA-KGNEM--L-.-QQ--F-EQQRRTAHYPNIRPLSGR.RNTM...................Q-  115
TAN.UG.x.TAN1             -AEG..............RDSRERR-GWLEIWDLS---WD-*LRDMVA--NQ--Q---G-EL-FQVWNFCQ-E.GE...RN-APMIERAYRYY-LV-KA--V-.--C---RRTPFEPYEERRNGV....GGG-DG-E.PPPGL...........A-  120
VER.DE.x.AGM3             -ASG..............RD-REAR-GELEIWDLS---WD--LRDM--DINQ--KM--G-EL-FQVWNYCQ-E.GE...RNRTPMLERAYKYYKLV-KA--V-.--C---RRQPFEPYEERRDGQ....GGG-AG-..VPPGL...........D-  120
VER.KE.x.9063             -ASG..............RG-RENR-GEVEIWDLN---WD--LRDM--D-NQ--KL--G-EL-FQVWNYCQ-E.GE...RR-APMTERAYRYY-LV-KA--V-.--C---RRQPFEPYEERRDGQ....GGG-AN-..APPGL...........D-  120
VER.KE.x.AGM155           -ASG..............RD-REER-GGLEIWDLS---WD--LRDMV--IN---KL--G-EL-YQVWNYCQ-E.GE...RQ-RPIAERAYKYY-LV-KA--V-.--C---RRQPFEPYEERRNGQ....GGG-PG-..VPPGL...........D-  120
VER.KE.x.TYO1_patent      -ASG..............RD-REAR-GEVEIWDLS---WD--LRDM-QD-NQ--RL--G-EL-FQVWNYCQ-E.GE...RH-TPMMERAYKYY-LV-KA--V-.--C---RRQPFEPYEERRDGQ....GGG-AN-..VPPGL...........E-  120
WRC.CI.97.97CI14          -SNS.................RE-A-S.......GGP-PIPED-LGMQ--FAM-KQ----EV-VRVFLYCSQ.......QREYEEVPTMHYFH-V--RG--M-.LTQ--PKQQEKREQQFQRAEQLYRNLERTAGDQGEVEQ-DLIS.......L-  117
WRC.CI.98.98CI04          -SNS.................RE-A-T.......GGP-PTPED-LGMQ--FAM-KL----EV-VRVFLYCSQ.......QRNYEEGPTMHYFH-V--RG--M-.LTQ--PKQQEKREQQFQRAEQFFRNLERAAGDQGEVEQ-DLIS.......L-  117
WRC.GM.05.Pbt_05GM_X02    -SNSG................RERP-T.......GGPQAIPED-LGMQ--FAM-K-----EV-VRVFLYCSQ.......QRNYEEQ-TMHYFH-V--RG-YL-.LIQ--PKEKENRERDFQRAEQFYRNLERASGDQGGVEQ-DLIS.......L-  118
```

```
                          Tat start                                                                disulfide bonding           NLS          exon 1 end exon 2 start              Tat end
H1B.FR.83.HXB2            ME................PVDPRLEPWKHPGSQPK..........................TAC.TNCYCKKCCFHCQVCFITKALGIS.YGRKKRRQRR...RAHQNSQ.THQASLSKQPTS..QPRGDPTGPKE*KKKVERETETDPFD.................  100
H1A1.UG.85.U455           --................----N----------T..........................---.S----V--W---L--LK-G---.--K---KP--....GPP-G-K.D--TLIP---LP.QSQ-.VSA-QE-S-----SKAK--R-A.................  101
H1C.ET.86.ETH2220         --................----N----N------..........................---.NQ------SY--L---L#-G---...............--P-S-K.D--NLI----L-..HT------SE-S-----SKA----YA.................  100
H1D.CD.84.84ZR085         -D................----NID--N------R.........................---.N--------Y--------G---................-PPHS--.---DPIP---S-..--------Q--..........................   86
H1F1.BE.93.VI850          --................----S-D--N------T........................-P-.-K----R------W--T--G---.---------H....-TP-S--.V--N--P---L-..-A---------S--E--SKAK---CA.................  101
H1G.SE.93.SE6165          -D................----N----N------.........................-P-.NK-F--V--W------LN-G---.------KH--....GTP-S-K.G--DPVP---LP..TT--N------S--E-ASKA-A-QC-.................  101
H1H.CF.90.056             -D................----K----N-----Q.........................---.N--------Y---M--LK-G---.------S--H....-TPASL-.D--N-I----L-..RTH--------Q--E-ASK-----...................   99
H1J.SE.93.SE9280_7887     --................----NR---N------.........................---.---------Y------LQ-G---...............S-PPG-K.---DLIP---L-..-TQRK----E-S--E--SKA-P-R--.................  101
H1K.CM.96.96CM_MP535      -D................----NI---NQ------........................---.NQ---R--Y---I--LK-G---.N------P--....TTPY--E.N--DP-R---L-..-----EQ-D---S-----SK-K--Q--.................  101
H1N.CM.95.YBF30           --................---------N-------........................---.N----R--Y--LY--TK-G---.------S---....-TP-S-K.S--DLIPE--L-..-QQ--Q--Q-KQ-EAL-SK--A--C-.................  101
H1N.CM.97.YBF106          --................---------N-------........................---.NK-------Y--MC--TK-G---.------S---....-PPKS-K.D--DPIPE--L-.R-QP--Q--Q-KQ--AL-GK--A--C-.................  102
H1N.CM.02.DJO0131         --................---------N-------........................---.NG---Y--Y--MC--TK-G---.------S---....-TPKS-K.N--DPIPE--L-QQ-QP--Q--Q-KQ--AL-DKA----.................  101
H1N.CM.02.SJGddd          --................---------N-------........................---.N----R--Y--MC--TK-G--L.------S---....-TPKS-K.N--DLIPE--L-..-QQ-XQ--Q--Q--AL-SK--A--C-.................  101
H1N.CM.04.04CM_1131_03    --................---------N-------........................---.N--------Y--V---TK-G---.------S---....-TPNS-K.NY-DPIPE--L-..-QP--Q--Q-XQ--AL-XK--A--C-.................  101
H1N.CM.06.U14296          --................-------------R--.........................---.N--F--R--Y--MY--TK-G---.------S---....-TLRS-K.S--DLIPE--L-..-QQ--Q--QEKQE-TL-SK--A--C-.................  101
H1N.CM.06.U14842          --................-----------------........................---.NK---R---X-I---TK-G---.------SR--....-TP-S-K.N--DPIPE--L-..-QH--Q--Q-KQE-A--SK-----C-.................  101
H1O.BE.87.ANT70           -D................----EVP--H------Q........................IP-.N----R--Y--Y---VR-G---.------GRP......-AASHP.D-KDPVP--SPT..ITKRKQERQE-QEEE--KKAGPGGYPRRKGSCHCCTRTSEQ..  114
H1O.CM.96.96CMABB009      -D................----E-P--H----K-Q........................-P-.N----S--Y--YL--ER-G---.-------RPA....VAAGVP.NNTNPVP--SLP..NASWKQEQQ--QE----K---AGGASCSQGSCNCCTRTAGQSS  117
H1O.CM.94.BCF06           -D................----EIP--H----K-Q........................-P-.N----R--Y--F---TK-G----.H------RPA.....-AASYP.DNKDPVPE-SP-..HT-RKQKRQE-QE----A.........................   92
H1O.CM.96.96CMA102        -D................----EVP--H------Q........................-X-.NK----S--Y--Y---TR-G---.-------RSA.....-A.GHS.DNKDLVPE-SP-..ITXRK.......................................   78
H1O.CM.97.97CMABB447      -D................----EVP--H------P........................-P-.NA-------Y--YL--TK-G----.H------RPA.....-AASCS.NNKDLV----ST..STNRKXKRQE-QE----KK-G-GGDPCSKVSCRYCTQTSGR..  115
H1O.CM.98.98CMA104        -D................----EMP--H------Q........................-P-.N----S--Y--Y--LVR-G---.-------RPA....TAS.HP.DNKDPVPE-SP-..ITNRK.......................................   78
H1O.US.99.99USTWLA        -D................---HEMP--H----S-Q........................-P-.NT----S--Y--I---TK-G---.-------RPT....TAERHT.DNKDPVPE-SPT..ITSRK.......................................   79
H1O.CM.98.98CMABB197      -D................----EVP--Q------Q........................-P-.NA-------Y--P---TK-G---.-------HPA.....-AS.NP.D-KDPVP--SP-..STGRKQKRQE-QE-E--KKAGPS.....................   96
H1O.CM.98.98CMU2901       -D................----KMP--H------X........................-P-.NT----R--Y--Y---TR-G---.-------RPA....TVERH-.DNKDLVPE-SPP..VTXRKQKRQE-QE-E--KKAGP-.....................   97
H1O.CM.99.99CMU4122       -D................----EMP--H------Q........................-P-.NK-------Y--Y---TS-G---.-------RPA.....-AH.HP.DNKDLVPE-SPT..ITNRKQERQE-QE----K--GP-.....................   96
H1O.FR.92.VAU             -D................-I--EMP--H------Q........................-P-.NA---Q--Y--YL--TK-G----.H------RPA.....-AASYP.DNKDPVPE-SP-..NNKRK.......................................   79
H1O.US.x.I_2478B          -D................----EMP--H------P........................-P-.-K-------Y--Y---TS-G---.-------RPA.....-AASRP.DNKDLVPE-SP-..ITNRKQKRQE-QE-E--KKAGP-.....................   97
H1P.FR.06.RBF168          -D................----D-P--QQ-----S........................SP-.N----A--Y--Y---TK-G---.-------RPARADN....-N-.N--DPV--..........................................   71
CPZ.CM.01.SIVcpzCAM13     --................-I--E-------------.......................---.N----A-----MY--TK-G---.-------.A--.....-TAFA-SKNN-DPVR---L-..-----NQERSEKQAE---SQAAA--.................   99
CPZ.CM.05.SIVcpzEK505     --................----E----N-------........................---.N--H--V--Y--VY--TK-G---.------S---....-TP-SNK.S--DP-P---L-..-RL--Q--Q--Q--TL-S.........................   93
CPZ.CM.05.SIVcpzLB7       --................-I--S-D--N-------........................-P-.N--F-----Y--PF--TK-G---.-----W-S--....-TSKS-P.HY-DPI----FP..PS---SLS-E-Q-----SKAQ--Q....................   99
CPZ.CM.05.SIVcpzMB66      --................-I--N----N-------........................---.N----Q--Y---L--TK-G---.---R--K---....-TSES--.N--DPVP---L-..--G-IE--Q-KS--E--SQ-TS-Q..FA.................  101
CPZ.CM.05.SIVcpzMT145     --................-IE-D----N----K-R........................-E-.NK-F-----Y--IL--TK-G---.------GR-K...P-.SSN-.DNKDP-R---L-..--IR-Q--SEKQ--E---TS-A-QS.DW.................  101
CPZ.CM.x.SIVcpzDP943      --................----D----Q------R........................---.N----C-----M---EK-G---.-------G--HPPAT...S-T.HN-DP-RQ--L-.R.QQ--Q--QE-Q-----SQA-AH-RA.................  102
CPZ.CM.x.SIVcpzMB897      -D................-I--N----N--------.......................-P-.N----M--Y---F--TR-G--Y.------KR-GATESS....KD.HQDTIPKQ.-L-.Q.SG-TQ--Q--S--E--KT-TS-QCA.................  100
CPZ.TZ.01.TAN2            -N................-I--QVA------AA-E........................-P-.------Q-----PL--TK------.----R-GRKSAANNK.....NNQDPVRQQ..-LP.K.RSRIQNS-E-SQ-E--A-ATSGGRPRQQDSSVSSGRTSGTSS  116
CPZ.TZ.02.TAN3_1          -N................-I--QVA------AA-E........................-P-.-------------PL--TK------.----R-GRKSATDNK.....NNQDPVRQQ..-LP.K.RSRIQSSQE-PQEE--A-ATSGGLHRQQDSSVSSGRTSGTSS  116
CPZ.TZ.00.TAN1            -N................-I--QVA--E---AA-E........................-P-.-------------P---TK------.----R-GRKS...AV..H-T.NN-DPVRQ-SLP..KRSRIQNSQE-SQEE--A--TSGGRPRQQDSSVSSGRTSGTSS  116
CPZ.CM.98.CAM3            --................----D----L------R........................---.N----M-----I---EK-G---.-------G--....-LSAA-STHN-G--QQ--L-..RQQ-NQ--QE-Q--TL-SQAKA--..FA.................  102
CPZ.CM.98.CAM5            --................---XN----L--------.......................---.N----C-----M---EK-G---.-------G--...HSLASS-T.HN-D-IRQ--L-..RQQ-NQ--EE-Q--T--SQA-A--R-WSAGS............  107
CPZ.GA.88.GAB1            -D................-I--D-----------R........................-V-.N----A--Y--IY--TK-G---.------TT--....-TAPAGSKNN-D-IP---L-..-S--NKE-SEKST-E-ASK--A-Q..................  100
CPZ.GA.88.GAB2            -D................-I--S----N-------........................---.NS----R--Y--ML--TS-G---.------.G--....-TALKGSKSN-DP-RQ--L-..-QQ--QKCQ--SEEE--SK--A-L..................   99
CPZ.US.85.US_Marilyn      -D................----NI----Q------........................---.N----------V---TK-G---.------GR--....-TTAK-SADN-DP-R---F-..-QL-NQ--Q--Q--AL-KQ-TS-QR-.................  102
CPZ.CD.90.ANT             -D................---AETP--L--PAT-A........................-P-.N----C--Y--PL--TK-G---.----R-AR-...............NR-TTAESSENNQDP-SK...............................   73
GOR.CM.04.SIVgorCP684con  -D................-T--E-P--QQ-----P........................-P-.NT-------Y--Y---TK-G----YGRK-R--PA-TADK.....D-DNQDPVSKQSLAG..TRSQQE......................................   82
GOR.CM.07.SIVgor2139_287  -D................-T--E-P--QQ-----P........................-P-.NT-------Y--Y---TK-G----YGRK-R--PA-TADK.....D-NNQAPVSKQSLAG..TRSQQE......................................   82
RCM.CM.02.SIVrcm_02CM8081 -D................VRAVGS-RIEEETLYNPFHYQE.....................-S-.NK------AY---L--LQ-----N.-ASRR-GS-.....DKKKTQV.DQLLATE.--L-..ATGRNSQSK--E-ETL-KKVV-NRPAGREDISTP.........  112
RCM.CM.00.SIVagi_00CM312  -D................GQEAG--RQEEETLYNPFQSVE.....................-P-.NA------AY---L--LQ--I--Q.-ASRR--RQ-....DKKEAKI.DTL-TPD-S.L-..ATSRRRNSKPKE--ESSV-KAVATNHTTGRKDLSIS.......  115
RCM.GA.x.SIVRCMGAB1       -D..............................VQGVGLEHPEEVILYDPFRKRE-S-.NT-------Y---L--LQ-G--N.-ASRA--R-.......SKEENKADKFPVPN-SI-..TT--NRKLQEKKE-T--KKVA-STTIG................  105
RCM.NG.x.NG411            -D..............................VKGAGLEEREEEILLDPFFTRE-S-.NS----R-VY---L--LQ-----N.-ASRR--RG-P....GNKKKNQINPFDT-E-SL-..AIGRNSQSK--K-EA--KAVD-NITTGREDIFIS.........  115
                          Tat start                                                                                                    exon 1 end exon 2 start                  Tat end
MAC.US.x.239              --.............TPLREQENSLESSNER-SCISEADASTPESANLGEEILSQLYRP..LE--YNT-------Y---F--LK-G---C.-EQSRK-R......-TP.KKAKANTS-A-NK-I-..NRTRHCQPE-AK-ET--KAVA-A-GLGR...............  130
H2A.DE.x.BEN              --.............TPLKAPESSLKPYNEP-SCTSERDVTAQELAKQGEELLAQLHRP..LEP-TNK----R-S----L--SK-G----.-E--G--R......-TP.RKTK-PSP-APDKSI-..TRT--SQPT--Q--TS-ATVV-TCGLGQ...............  130
H2B.CI.x.EHO              --.............IPLKEQESSLNSSSGH-SSTSEGVANTQGLDNRGEEILSQLYRP..LK--SNT------SY---L--LK-G---C.-E-S.-KR......SS.KRAK-TTS-APNESL-..ART--SQPT-KQ--E--TTRA--LGPGRSNTSTSRFAN.....  138
H2G.CI.92.ABT96           --.............TPSREQENSLKSCREL-SCTFEETVDAPGLEGTQAEILYQLYRP..LE--SNK-------Y---L--LN-G--VC.-EQP.--R......-TP.KKTKANSF-A-NESI-..ARTRNXQPK-K..........................  112
H2U.FR.96.12034           --.............IPLKEQESSLKSCKEQ-SSTSEGDVGTQGSAAAGEEILSQLYRP..LE--SNK-------Y---L--LK-G---C.-EQP.--R......SP.KKVKAYSSAP-NKSL-..NGTRNSQPKEK..........................  111
SMM.SL.92.SL92B           TG.............IQ*KGQGK*LRYYKRHCSYTLGMDAPIQGLDSQGDQIPWDLYRP..LET-YNS-F---------L--LR-G---T.-AKPR--.........VKKKAKANPFDT-N-SI-..SGTRNCQPK--Q-E---T-VD--CGLGR...............  126
SMM.US.x.H9               --.............TPLKEQESSLESSREH-SSISEVDADTPESASLEXEILSQLYRX..VE--YNKX------Y---H--XK-G---C.-EQQR--........T.PKKTKANTF-A-DKSLX..XRARNCQPK--K-ET--X-VA--LGLXR...............  128
SMM.US.x.PGM53            --.............TPLKEQENSLGSSNER-SYTSEEDVLTPESAKVEEEILCQLYRP..LE--SNT-------Y---L--LK-G---C.-EQQR--R-......T.PKKAKANTF-APNESL-..TRARNSQPK--K-ET--T-VA--LGLGR...............  130
SMM.US.x.SIVsmH635F_L3    --.............TPLKEQENSLESYREH-SSISEVDVPTPESANLEEEILSQLYRP..LEP-YNK---R--Y---H--LK-G---C.-EQHR--.......T.PKKTK-NPLPA-NKSL-..TRTRNRQPK--K-E---T-VAA-LGLGR...............  128
STM.US.89.STM_37_16       --.............TPLKEQESSLRSSSEP-SCTSEAVAATPGLANQEEEILWQLYRP..LEE-CNK-F-----Y---L--V--G---T.-E-SR--.........VKKKAK-YPI-A-NKSL-..TRARNSQPK--Q--E---T-V-ST-GLGK...............  128
MNE.US.x.MNE027           --.............TPLREQENSLKSSNGR-SCTSEAAATTLESANLEEEILSQLYRP..LE--YNT-------Y---F--LK-G---C.-EQSR--R-......T.PKKVKANTS-A-NK-I-..DRTKHCQPK--Q-ET--TAVA-V-GLGR...............  130
DEB.CM.99.CM40            --.......................................SIDPFEERAN............-PH.PA-M--A-IL---L---Q-G-----.--PRR-KRKCRREYTPAAE-NQPD-DFVPE--LP..-R-RKCFDSQKKEAE--ETAGPGGQHCREDSSVSSGRISNNC.  115
DEB.CM.99.CM5             --.......................................EIDPFKETAN............-PH.P--M--A-TM---L--MQ-G---Y.--SRR-KR--..GSSFAA-.DKD--EFIPE--LP..-R-WKCFNSE-KET---KASRAGGRHRPEDS............  101
DEN.CD.x.CD1              -N....................................ADSIDPFAG.NK............-P-.NT----H-AY---L--LQ-G---V.---AP-KRP-......ARQ-GTPD-EAV---SL-..D--RQQISQ-KE-E---TA-GSSGGNIREDS............   99
MND-1.GA.x.MNDGB1         --...................-SGKEDHNCPP-DSGQEEIDYKQLLEEYYQPLQ.........--ENK-W--------ML--QK-G---R.-HVYRK-VP.....GTNKKIPGSGEEAIRR-GYR..PVFSQDCQQNLYR-RPDNGE-ESNSLELGEHNSENLKEESLNR  134
LST.CD.88.SIVlhoest447    -Q...................QQEQEQLT...RQ-QHQDPLKETYKEAVTKPLS.........--NNK-C--------IL--QQ-G---R.-YVHR--K.......RVAQVFEDK-DPVDGTASG.RKE-TKNSQT-KER............................  100
LST.CD.88.SIVlhoest524    -Q...................QPEQEQHT...-Q-QHLDQLEEIYKEAITDPLQ.........S-QNK-H--V------IL--QQ-----R.-YVHRT-R.......RVTYILE-E-DPVDRTSSG.RTE-AKNSQT-KERQRKT..EK-STST................  111
LST.KE.x.lho7             -Q...................QPEQEQPI...-Q-QHQDLMKETYNEAVTKALQ.........--QNT-C--------IL--QQ-G--LH.-YVHRK-R.......RVTSILE-Q-DPVDRTTSG.RKK-TKNSQTEKKR-EQI..QRLSTST................  111
DRL.x.x.FAO               -D.............ARKVDLDQQDAG.THFEP.FQ.......................GRTNN-.NT-W--Y-AY---L--LQ-G---R.-NVSRP-K-T....ATVVSRVSETGDTPVARTHT.A-R-PTT-DSQTE--EKSKTEASGTS-QNPPSQDPPKLGAKKQ.  124
GSN.CM.99.CN166           -P.............ILSSLGDPQDIAWMNMNRDPEECLLPNWQQPGAAPA...........-P-.SA------AL---L--LR-G---L-.LHGRS-KRK-PK...-TEPDPSVSSNQD-A...........................................   92
GSN.CM.99.CN71            -N...............................RDPAECLLPKWQQPGAAPN...........-P-.SA----R-A----L--LR-G---L-.LHGRS-KRK-SE...-SEPAPSVPTNQDPI-IQ-.LP.HSKWNAIRQTEQEKTVARKATSGPSDISQDSRVSCGTT...  119
MND-2.CM.98.CM16          -D.......................................AGKAVSDKKEGDVTPYDP.FR..DRT-P-.DT----R--Y---L--LQ----VH.-HAYRN-RS-Q....RLLEKISEDS-APVGRVIA.LFLMSSRETISVPTTSNSQAKKEKKSPTQKNRVPNQTAPK....  122
MND-2.GA.x.M14            -D.......................................AGKAELDKKEANITPLDP.FQ..GRT-P-.NK-------Y---C--LQ-----H.-HVYRT-RT-K....RLLGKISQDSEAAVGRVIA.LFLMS.......................................   87
MND-2.x.x.5440            -D.......................................VGEVASDKKEEDITHFDP.FR..ART-P-.NK-F--V--Y---L--LQ----VH.-HVYRQ-RP-R....GLLEKVPSNSQAALGRDIA.LFLMSSRETISGPTTSNGQTEKEEKSQT.............  110
MON.CM.99.L1              -ME......................................PVDPDLPKEQHPPATPR..........-P-.---F-RV------I--L--G---.-----RK-R-....ATSPVPGLSSSKNPAR--SL-..H-TR-SQR-T-QAQA-ATAATP-RQHI...............  104
SUN.GA.98.L14             -S...................TQGHQQD.........QDQGKGTLEEAYKTNLE..........--DNK-W-RR------L--LQ-G---H.-YVYRG-........Q--SISKDK-EAVRRIGSR.KPAPNPQ-KK-KKRQR....KKESTST.............  102
MUS-1.CM.01.SIVmus_01CM1085 -D................-SVEE-PKEQR--AA-A........................-P-.-S-F---A---VL--QK------.---RF-KRS-......ADS-LPS-SDPVQIKSL-..DR-R-QKRQ--Q-ET--A-A----.....................   98
MUS-1.CM.01.CM1239        -D................-SVEN-P-EQR--AA-T........................-P-.SS---R-A---ML--QK------.---RF-KRP-......SDQNVQN-PNPVPIKSIP..NLQRQQSRQ--Q-EALAAAAGP--C-I.................  101
MUS-2.CM.01.CM1246        -D................-SVEG-P-DQR--AA-P........................-P-.------C-A---IL--QK------.---RI-KRP-......TDSLLPSD-NPVRF-SIP..DSQQNKSNK-KQ--AL-EAAG---CRCFQKTSTT..........  108
MUS-2.CM.01.CM2500        -D................-EVEN-P--QR-S-A-T........................-P-.-P----A-AY--MQ--L------.---NRK-H--......T.PEGLPNN-DPIQLKSIP..E--R-QER--KQ--E-AP-AVS--CAI.................  101
TAL.CM.00.266             -S......................................SKEELRTTPISDPFQEEGR........GP-.NK-------Y---L--LQ-G---R.-VPDR-FRK-RR..STPS-NQEDKDPV----L-..RAT-IQKRQE-TLQ---SKAA-SGT-RKQDS............  110
SYK.KE.x.KE51             -S.......................................TEDKVRETQGIPTSFLEGTFLSNGPQ-P-.SR-------Y--YR--LQ-G---T.-V-ERK-RHAR...TTAEDFAA-SSTT....R-.RS-LPST-R-QQGQ--E.........................   98
SYK.KE.x.SYK173           -S.......................................STDQICQTQRVPPSFLEGTFLEKGPP-P-.NK-F--N--Y---L--LQ-G---T.-A-PRK-AA-....SISEDDSAPTGT-PRAGRT.QANPQTKKKAV-TE-DSTSLPSAENL.................  110
TAL.CM.01.8023            -S......................................SKEELPTTPTYDPFPAGG.........-P-.NK-------Y---L--LQ-G---R.-VPDR-IRK-RR..STSS-NQKDKDPV----L-..RAT-NQECQ-KKVQ---SP-GAGGEHRLKDSLPLSGRTSTASS  121
GRV.ET.x.GRI_677          -D...............K.EEE....PHPLLQDL.....................HRP..LQP-TNK-------Y--EL--LQ-G--VR.-HVSRK.-----...KT...-TQDN-DPIRQ-SI-.TV.QRNGQTTE-G-TE--KAAAAN.....................   94
SAB.SN.x.SABI             -D..............QEQEARPQVWEEL.QEEL.....................HRP..LQ--DNT-F--V-----IL--HK-----R.-YVPRP-R.......-SKK..IS-NQVSLHN............................................   73
TAN.UG.x.SIVagmTAN1       --...............S...EGDGMAESLLQDL.....................HRP..L-P-TNK-F--R--Y------LQ-G---T.-HVSRI-RP.......KKNHSN--NLV-Q-SI-.AW.G-NSQTTQ-E-T-IPAAA--SRRPQ................  101
VER.KE.x.9063             -D...............KGEEE-TVLHQDLIR-Y.....................KKP..L-T-RNK-F-----Y---L--LQ-G--VT.-HAPRT-R........KK-VQPNRL-QQD-SI-.TR.GR-GQATQ-SQ-------T-AQILGRKDLERDKREAVGANA  119
VER.KE.x.AGM155           -D...............KGEEDQDVSHQDLIK-Y.....................RKP..LET-TNK-F-----Y---F--LR-G---T.-HAFRT-R........KKIASADRIPVPQ-SI-.IR.GR-SQTTQ-SQ----EQAKANLRISRKNLGDETRGPVGAGN  119
OLC.CI.97.97CI12          -D...................AQEWSLEETLTEDL-..........................E-KNK-------Y---L--LQ-G----.-STESP-K--.....KKVV-HQQDSRDT-HFQIP..SRTRNSQTTEKE-TASSSRAPSTSGTVPQSVQGH.........  107
VER.KE.x.TYO1_patent      -D...............KGEAEQIVSHQDLSEDY.....................QKP..LQT-KNK-F-----Y---L--LQ-G--VT.-HAPRT-R........KKIASLDL-P-QH-SI-.TKWGR-GQTTPTSQE---TTAGSN.....................  100
WRC.CI.97.97CI14          -RKSLLCIISTGYCKEGSSCISPKGVLNNKRKGNNNFKEQNNCIETWREQQEIREKLSKETLSP-DHR-S-VA-VYS-TL--LQ-G----.-RSYSKKTKP............DTTTAA-SRT--.TTWKR-SKTA-KE...........................  125
WRC.CI.98.98CI04          -KKVPLCIIFTGYYKEGSSCTLPKGVQNNKKRENNNFKEQNNSLETWKEQQEIRERLSKETLSP-NHK-G-Y--V-S--L--LQ-G----.-RSYSKKIGA............STAKATTSRA--.TTWKR-CQTE-KKQ--TTVL......................  132
WRC.GM.05.Pbt_05GM_X02    -KNRPLCTIFTGSFKEDYIYTSYKDVP-KKRTEKEIFKEQNNFIEIWKGLQEIREELSRETLSP-NHQ-R----I-S--L--LQ-G----.-HSYSKKTNK............DKSDASTSRVI-............................................  112
COL.CM.x.CGU1             -S............LS..................................LYAKP..LE--NDK-W--A------L--LS-G---R.-.K-..NK..........SAT..QTQ-VA-TGSG.....NSQ-TQ-KYTES....KK-AAA-FGD..............   79
```

```
H1B.FR.83.HXB2              ....................*  100
H1A1.UG.85.U455             ....................-  102
H1C.ET.86.ETH2220           ....................-  101
H1D.CD.84.84ZR085           ....................-   87
H1F1.BE.93.VI850            ....................-  102
H1G.SE.93.SE6165            ....................-  102
H1H.CF.90.056               ....................-  100
H1J.SE.93.SE9280_7887       ....................-  102
H1K.CM.96.96CM_MP535        ....................-  102
H1N.CM.95.YBF30             ....................-  102
H1N.CM.97.YBF106            ....................-  103
H1N.CM.02.DJO0131           ....................-  102
H1N.CM.02.SJGddd            ....................-  102
H1N.CM.04.04CM_1131_03      ....................-  102
H1N.CM.06.U14296            ....................-  102
H1N.CM.06.U14842            ....................-  102
H1O.BE.87.ANT70             ....................-  115
H1O.CM.96.96CMABB009        CGFTTS..............-  124
H1O.CM.94.BCF06             ....................-   93
H1O.CM.96.96CMA102          ....................-   79
H1O.CM.97.97CMABB447        ....................-  116
H1O.CM.98.98CMA104          ....................-   79
H1O.US.99.99USTWLA          ....................-   80
H1O.CM.98.98CMABB197        ....................-   97
H1O.CM.98.98CMU2901         ....................-   98
H1O.CM.99.99CMU4122         ....................-   97
H1O.FR.92.VAU               ....................-   80
H1O.US.x.I_2478B            ....................-   98
H1P.FR.06.RBF168            ....................-   72
CPZ.CM.01.SIVcpzCAM13       ....................-  100
CPZ.CM.05.SIVcpzEK505       ....................-   94
CPZ.CM.05.SIVcpzLB7         ....................-  100
CPZ.CM.05.SIVcpzMB66        ....................-  102
CPZ.CM.05.SIVcpzMT145       ....................-  102
CPZ.CM.x.SIVcpzDP943        ....................-  103
CPZ.CM.x.SIVcpzMB897        ....................-  101
CPZ.TZ.01.TAN2              SGSTRPVETSSGSCGRSCKH-  137
CPZ.TZ.02.TAN3_1            SGSTRPVKTSSGSCGQSCKR-  137
CPZ.TZ.00.TAN1              SGYTRPFKTSSGSSGSACKH-  137
CPZ.CM.98.CAM3              ....................-  103
CPZ.CM.98.CAM5              ....................-  108
CPZ.GA.88.GAB1              ....................-  101
CPZ.GA.88.GAB2              ....................-  100
CPZ.US.85.US_Marilyn        ....................-  103
CPZ.CD.90.ANT               ....................-   74
GOR.CM.04.SIVgorCP684con    ....................-   83
GOR.CM.07.SIVgor2139_287    ....................-   83
RCM.CM.02.SIVrcm_02CM8081   ....................-  113
RCM.CM.00.SIVagi_00CM312    ....................-  116
RCM.GA.x.SIVRCMGAB1         ....................-  106
RCM.NG.x.NG411              ....................-  116

MAC.US.x.239                ....................-  131
H2A.DE.x.BEN                ....................-  131
H2B.CI.x.EHO                ....................-  139
H2G.CI.92.ABT96             ....................-  113
H2U.FR.96.12034             ....................-  112
SMM.SL.92.SL92B             ....................-  127
SMM.US.x.H9                 ....................-  129
SMM.US.x.PGM53              ....................-  131
SMM.US.x.SIVsmH635F_L3      ....................-  129
STM.US.89.STM_37_16         ....................-  129
MNE.US.x.MNE027             ....................-  131
DEB.CM.99.CM40              ....................-  116
DEB.CM.99.CM5               ....................-  102
DEN.CD.x.CD1                ....................-  100
MND-1.GA.x.MNDGB1           SLIEDLTSFARE........-  147
LST.CD.88.SIVlhoest447      ....................-  101
LST.CD.88.SIVlhoest524      ....................-  112
LST.KE.x.lho7               ....................-  112
DRL.x.x.FAO                 ....................-  125
GSN.CM.99.CN166             ....................-   93
GSN.CM.99.CN71              ....................-  120
MND-2.CM.98.CM16            ....................-  123
MND-2.GA.x.M14              ....................-   88
MND-2.x.x.5440              ....................-  111
MON.CM.99.L1                ....................-  105
SUN.GA.98.L14               ....................-  103
MUS-1.CM.01.SIVmus_01CM1085 ....................-   99
MUS-1.CM.01.CM1239          ....................-  102
MUS-2.CM.01.CM1246          ....................-  109
MUS-2.CM.01.CM2500          ....................-  102
TAL.CM.00.266               ....................-  111
SYK.KE.x.KE51               ....................-   99
SYK.KE.x.SYK173             ....................-  111
TAL.CM.01.8023              ATVTRP..............-  128
GRV.ET.x.GRI_677            ....................-   95
SAB.SN.x.SAB1               ....................-   74
TAN.UG.x.SIVagmTAN1         ....................-  102
VER.KE.x.9063               ....................-  120
VER.KE.x.AGM155             ....................-  120
OLC.CI.97.97CI12            ....................-  108
VER.KE.x.TYO1_patent        ....................-  101
WRC.CI.97.97CI14            ....................-  126
WRC.CI.98.98CI04            ....................-  133
WRC.GM.05.Pbt_05GM_X02      ....................-   13
COL.CM.x.CGU1               ....................-   80
```

```
                                Rev start                 exon 1 end exon 2 start    NLS                                    Leu-rich effector domain                      Rev end
H1B.FR.83.HXB2                  M........AGRS..GDSDEELIRTVRLIKLLYQSN..PPPN..PEGTRQARRNRRRRWRERQRQIHSISERILGTYLG........RSAEPVPLQLPPLERLTLDCN....................EDCGTSG.......TQGVGSPQILVESPTVLESGTKE.....  116
H1A1.UG.85.U455                 -.........-R-...-NP--D-LKA--I--------..-C--..-R-S---K-------A-----D-L-----SDC--.........-P---------I---R---S....................-S-----TQQPQGTET---G---S---SA--G----N.....  123
H1C.ET.86.ETH2220               -.........----.------LKA--I--I------..-Y-T..----------------A------TL-----SNF--.........-P-------------N---S....................--S----TQQSQGT-E---N-....................  107
H1D.CD.84.84ZR085               -.........----.-----D-LTA----I------..---S..----------------A---Y----G----S---.........--E------------IN-N-S....................-------.......-----------S----AI---R-E-.....  116
H1F1.BE.93.VI850                -.........----.----T--LKA-KC--I------..-Y-K..----------------A-----RAL-D---SSC--.........--E-------------HIN-S....................----QGP.......EE----S--SG--HA--------.....  116
H1G.SE.93.SE6165                -.........----..-ST----L-A-KA--I------..-Y-P..----------------A-----SA------TA---.........-P--------------H---S....................--S----TQQPQGTET---RS....................  107
H1H.CF.90.056                   -.........----..-A--T--LQVCKI--I------..-C-E..-T-------------A-----RE------TSC--.........-PP---T-----------N-S....................-------------.EK-E-----SL--S-I-GT----.....  116
H1J.SE.93.SE9280_7887           -.........----.-----DQ-LLA-----I------..-Y-K..-N-S-----------A--N--D------PSSC--.........-P---------I---R---S....................----N--..........---D---SG-PCM--GA----.....  116
H1K.CM.96.96CM_MP535            -.........---R..--P--Q-LT---T--I--E--..-Y--..L--S--T---------A--K--S-----L-SAC--.........------------I-K-N-N-D....................--P-KGT.......EG-L-----S--PC---------.....  116
H1O.BE.87.ANT70                 -.........----....E-DQ-LQAIQI--I------..-Q-S..-R-S-N--K-------R--A-VDTLAA-V-A-VVH.......GPQNNNIVD-----Q-SIRDP....................-GDQL-E........AWTVD-RAEDN................  102
H1O.CM.94.BCF06                 -.........----....-E-QQ-LQAIQI--I------..-H-T..-G-S-N--K-------R--A-VDTLAT---T-VVH.......G-QDNNLVD-----Q-SIRDP....................-VDQLP-........TWTVD-GTKDN................  103
H1O.CM.96.96CMA102              -.........----....D-DQ-LQAI-I--I------..-H-S..-X-S-N-----------R--A-VDT-AA---A-VVH.......GXQDNNLVD-----Q-SIRDS....................-CDQP-E........NWTVD-RTENN................  102
H1O.CM.96.96CMABB009            -.........----....-E-QQ-LQAC-I-QI------..-Y-T..-V-S-NS-K-------K----VEHLAA-V-A-VVH.......GQQDNHLVD-----Q-NIRDP....................-AEQ-T-........TLSMDSGAKDN................  103
H1O.CM.97.97CMABB447            -.........----....-D-RQ-LQA--I-RI------..-Q-A..-T-X-N----K-----R--A-VATLAAKF-ADIVH.......-PPDDN-VE--S-DQ-SIRDL....................-SEQLP-........TWT-DSGTEDN................  103
H1O.CM.98.98CMA104              -.........----....D-DQ-LQAI-I--I------..-H-S..-I-S-S-----------R--A-VDT-AA---A-VVH.......GPQNNNLVD-----Q-SIRDP....................-SDQP--........NWTVD-RAEDN................  102
H1O.CM.98.98CMABB197            -.........----....D-TQ-LQAIQI--I------..-H-A..-G-S-S-----------R--A-VETLAA-F-A-VVH.......GPQDNNIVD------SIRDP....................-SDQL--........TWT-D-RAEDN................  102
H1O.CM.98.98CMU2901             -.........----....-A-QP-LNAIKI--I------..-H-S..-X-S-N--K-------R--A--DTLAA-V-A-VVH.......GPQDNNLVE-----Q-NIRDX....................DGDQP-E........KWTVD-RTEDN................  103
H1O.CM.99.99CMU4122             -.........----....D-VP-LH-I-I--I------..-Q-S..-I-S-S-----------R--A--DTLAA-F-ASVVH.......GPQDNNLVE-----Q-SIRDP....................DGDQP--........T-TVD-RTEDN................  102
H1O.FR.92.VAU                   -.........----...EE-QQ-LQAIQI--I------..-H-T..TS-S-N-----------R--A-VD--AA-V-ASVVH.......GPQDNNLVD-----Q-S-RDP....................-VDQ-P-........TWTVDHRAEDN................  103
H1O.US.99.99USTWLA              -.........----...GE-QQ-LNAIQI--I------..-Q-S..-A-S-N--K-------R--A-VD-LAT---A-IVH.......GPPDNN-VE--D--Q-NIRDP....................-VDQPA-........TWTVHSMAKDN................  103
H1O.US.x.I_2478B                -.........----....-DVQP-LQA--I--I------..-H-S..-I-S-N--K-------R--A--DTLAA-V-T-IVH.......GPQDNNLVE-----Q-SIRDP....................-ADQP-E........NWTVDSGAEDN................  103
H1N.CM.02.DJO0131               -.........----..-VN----L-A--I--I------..-Y--NNNQ-------------T-----RE--Q---D-C--.........GLQ---D-P----D-----PA....................--S-AP-......TEPHQ-TATTE..................  106
H1N.CM.02.SJGddd                -.........----..-VN----L-A--I-RI------..-Y---.NK-X------------A-----RA------SSC--.........GPP---D-P----D-----PK....................--S--PE......TESQQ-TATTE..................  104
H1N.CM.04.04CM_1131_03          -.........----..-VN----LTA--I--I------..-Y---.NQ------X------XA-----RA--X---S-C--.........GXQ--GDPP-X--X-----XT....................--S-AP-......TESQQ-TATTE..................  104
H1N.CM.06.U14296                -.........----..-VN----LGA--A--I------..-Y---.NK-------------A-----RA------L----.........GPQ--GD-P----D-----PX....................--S-----.....TESQQ-TATTE..................  104
H1N.CM.06.U14842                -.........----..-VD--G-L-A--I-RI------..-Y---.NT-------------A-----RA------S-C--.........GPQ---D-P----D-----ST....................--S-----.....TESQQ-TATTE..................  104
H1N.CM.95.YBF30                 -.........----..-VN----L-A--V--I------..-Y---.SK-------------A-----RA------SSC--.........GPP---D-P----D-----TE....................--S--P-......TESQQ-TATTE..................  104
H1N.CM.97.YBF106                -.........----..-VN--D-LKA-KI--I------..-Y-D.SSQ-------------A-----RA-------A---.........GPQ---D-P----G-----HK....................--S-DP-......TESQQ-TATTE..................  105
H1P.FR.06.RBF168                -.........----...DEDLR---T-I-I--I------..-W-E..RGSS-NST--K-----S-----DQ-AG---ASR--.........-PE--GGAD--D-S--HIGD....................QGPPNNP........VDTPG-TADQRA................  104
CPZ.CM.01.SIVcpzCAM13           -.........----...EHD-AR-LQA-KI--I--E--..-Y-S..-----N-----------A-----RE--N-V-AAH--.........-PPQ-GH-E--E-DK-S-H-L....................-TV-KV-TTVNQHPQTDT-ETDSTEGDNPI-GKRI-N.....  123
CPZ.CM.05.SIVcpzEK505           -.........----..-VN----L-AI-V--I------..-Y---.DS------K-------A-----RAL-----SSCV-.........GLQ--ST-P----D--S-NPE....................--L---E......TEHPQ-TATT..................  103
CPZ.CM.05.SIVcpzLB7             -.........----..--R----L-A--I--I------..-F-H..-----S-----------A-H---SE-AG---S-C-E.........-PEG----P---ID--S-NSE....................--I-DC-TIHSEGTET-----T-SAGNS---GK-A-N.....  123
CPZ.CM.05.SIVcpzMB66            -.........--G-..-N-----L-A--I--I------..-Y-S..---S-----------A-----N-L-----S-C-E.........-PTG--S-P---I------SA....................--I--G-TDPPQGTET-T---NTPEGHS-I-GT-A-N.....  123
CPZ.CM.05.SIVcpzMT145           -.........----.EGDENL-RAIR##-RI--E--..-Y-S..QS---R-----K----------NQ-G--V-ASC--.........--S--PQ-P---I---CISDL....................-TT-DT-TIQPQGS-TNRY--TPTERD-A--DNRNQ-.....  121
CPZ.CM.98.CAM3                  -.........----.E--D-AC-LQA--I--V--S--..-Y-D..NK-------------A-------L-----SS---.........--P--PA-L---I----I-TG....................--I-KP-......TEPQQ-TATAEELQQIPVG-DNLLGKGAT  123
CPZ.CM.98.CAM5                  -.........----.E--E-TR-LQA--I--I--D--..-Y-D..NK-------------A------E-GQ-V-SSC--.........--S--PD-L---ID----NPV....................--SAN--......TEHQQ-TATVEGL..............  107
CPZ.CM.x.SIVcpzDP943            -.........----.E--A-TR-LQV--I--I--D--..-Y-D..NK-------------A---H--AL------SC--.........--P--PT-P---ID----NPE....................--I-N--......TEHQQRTAVAEGHQQ-PVG-DSILGKGAT  123
CPZ.CM.x.SIVcpzMB897            -.........----..-SDE-L-RAVR##--I------..-F-S..---P------------K-----SAL-----S-C--.........-LE-----------------SE....................--A-DC-TEVPXRTET---N-N--E-HSP--GTRAEN.....  121
CPZ.GA.88.GAB1                  -.........----.EPQD-AR-LQA-KI--I------..-Y-S..-----K----------A--K--SE--G-V-A----.........-PPK-GD-E--E-DK-S-Q-V....................-TTQDV-TSNTSQPQTAT-ETVPAGGNYSI-GK-A-N.....  124
CPZ.GA.88.GAB2                  -.........----..--E--R-L-A-KA--I--D--..-Y-S..NR---S----------A-----CE--Q--FSSL--.........-HP--DD-P-------S-GPL....................-PA--I-......T-GPERTETTQGA-EPAERDR-ILAARTK  122
CPZ.TZ.00.TAN1                  -.........----...EE-AN-LY---I--I--D--..-Y-S..GA-S-T----------Q--H-VDALAS---QYR--.........GPQ--PH-DI-D-SK-H--PL.....................-QPA-.......TET-DNQLGTQPSNSA............  105
CPZ.TZ.01.TAN2                  -.........----...EE-AN-LQ-I-I--I--D--..-Y-S..AA-S-T----------Q--H-VDALAN---QYR--.........GPQ--PH-A--D-SK-H-GPV.....................-DPA-.......TET-DNQ-SKQPSNSA............  105
CPZ.TZ.02.TAN3_1                -.........----...EE-AN-LQ-I-I--I--D--..-Y-S..GA-S-A----------R--L-VD--AN---QYR--.........GPQ--PH-D--D-SK-H-GPV.....................-NPA-.......AET-DNQSSTQPSNSA............  105
CPZ.US.85.US_Marilyn            -.........----.G--A--Q-L-A--I--I--A--..-F---.NS------------KN-----NE--Q---S-C--.........--S--PD-P---IGG--INSE....................--STNP-......TEHQQ-TATVE..................  105
CPZ.CD.90.ANT                   -.........---EELEGT--Q-LKA-KI--I------..-Y-K..-A-S-A---K---KK--D-VEGLAA-V-R-LV-.........GPD-HN-VD--D-QN-S-EPL.....................GSPPTPLQNYFQTV-PS-EQSQHPLDNNQN...........  118
GOR.CM.04.SIVgorCP684con        -.........----..DEDLQ--LTKI-I-RI------..-W-E..-GPS-NST-------KG-----DQ-AG---ASH--.........--QN-GCVD--DISH--IGD....................QGDN-QP........DRXPDQTTXKSE...............  104
GOR.CM.07.SIVgor2139_287        -.........----..DEDLQ--LTKI-I-R-------..-W-E..-GPS-NST-------KG-----DQ-AG---TSH--.........-YQN-GCVD--DISH--IGD....................QGDN-QP........DR-PDQTTEKSE...............  104
RCM.CM.00.SIVagi_00CM312        -...........LLGEEE-R-IR-RL---H--HLT-..-Y-QPQEG--ASQ--R--VQ--R-WL--IQLA---FLYPDP.......PPNSD-LSNAVEQ-NQ--IAELPEPPINP.............FTNPA-SVV...........DSS..................  107
RCM.CM.02.SIVrcm_02CM8081       -...........LL-EE-Q-IR-RL---NF-L----..-Y-Q..-G--ANQ--R---L--R-WL--AQLA---FLLPDT.......PPDRDLIEGAVEQ-QQ-AISDLPEPPVNP.............FSNPADSAV...........DNS..................  104
RCM.GA.x.SIVRCMGAB1             -...........LP-QDE-D-RKKI---NF--LI-..-Y-QH..G--ANS--KK-----R-WQ-VQQLA----LDSTD.......PPVEQDLDAAIAD-QK-Q-NNLPEPPVD.............FS.......................................   93
RCM.NG.x.NG411                  -.........LLGEEEEA-Q-IR-RI---H-IHL--..-Y-QS..G--ANQ--RK--Q--R-WT--LQLA---FLYPDT.......PPDRD-LQEA-SN-QQ---SDLPEPPVNP.............FSNPS-FAV...........DRS..................  107
                                Rev start                  exon 1 end  exon 2 start                                                                                          Rev end
MAC.US.x.239                    -............SNHERE---RKRL---H--H-T-..-Y-TG..P--ANQ--Q-K----R-WQ-LLALAD--YSFPD........PPTDT-LD-AIQQ-QN-AIESIPDPP................TNTPEALCDPTE........DSRSPQD...............  107
H2A.DE.x.BEN                    -............SERADE-G-QGKL--LR--H-T-..-Y-QG..P--ASQ-------R-RQWLRLVALANKLCAVPD........PPTDS-LDRAIQH-Q---IQELPDPP................T-LPE-NS............NQGLAET...............  103
H2B.CI.x.EHO                    -..............NARERD-QKGL--LH--H-T-..-Y-QG..P--ASQ--------KQ-GL--LALAD--HPLPD........SPTEG-LD-AIQR-QN-IIKDLPNPP................TSTP-AQASTCIPPIWDQLV-RSNPS-SQGCGRDSC-RGEDL  128
H2G.CI.92.ABT96                 -.............SNLEE---RKRL---HF-H-T-..-Y-QG..P-XANQ------K-KQ-WL--LALAD--YSFPX........PPAR--LD-AIQQ-QS--IQDXPDSP................TAVPKNLQD...........-PSN..................  100
H2U.FR.96.12034                 -..............SNPE-GVQ-RL---H--H-T-..-Y-TG..P--ANQ---------Q-WL--LALADK-YTFPD........PPTDS--D-AIQQ-QG-AIQDLPEPP................AVVPEHQR............TSAP..................   98
SMM.SL.92.SL92B                 -..............QNPE---R-RL---H-IH-T-..-Y-AG..P--ASQ--------KR-WI--VALAD--YSFPD........PPADTDID-AIQQFDS-SIQSLPEPP................TTVPETLRD...........QSAD..................   99
SMM.US.x.H9                     -..............SSNE---R-RL---HF-H-T-..-YXXG..P--ANQ--R-----XQ-WQ--LAXAD--YSFPD........PPADT-LD-AIQQ-QG-AIEXLPNPP................ASAPEPLKD..........AAESP..................  100
SMM.US.x.PGM53                  -.............SSSAGE---R-RL---HF-H-T-..-Y-QG..P--ANQ--R------Q-WQ--LALAD--YSFPD........PPADS-LD-AVQQ-QD-S-ENIPDPP................TSVPEALTE..........N-RSRQD...............  104
SMM.US.x.SIVsmH635F_L3          -..............SSTE---RKRL---HF-H-T-..-Y-QG..P--ANQ--R------Q-WQ--LALAD--YSFPN........PPTDT-LD-AIQQ-QG-AIEDLPDPP................TSAPETLKD..........AAKSS..................  100
STM.US.89.STM_37_16             -..............S-QE---RKRL---QF-H-T-..-Y-QG..P--ANQ---------Q-WN-LLALAN--YSFPD........PPAST-LD-AVQQ-QG-SIQDLPDPP................PNLPKDLQD..........TAEN...................   99
DEB.CM.04.SIVdeb04CMPF3061      -.....AHAG--G..ELQNLLRAIAAI..-I------..-F-EG..GSRAS.--R------QH-ETVD-LAA---ES---...........-P-AA-E-A--D-QE-H-AAD.....................EGSRD.............PARTNI.................   97
DEB.CM.99.CM40                  -.....AHAG--GSAEENTRQ-LKVIS---I------..-Y-KG..G-SAST--R--QK--R--G-VD--A----QSH--..........G-PTTADVA--D-SQ-H-AD...........................................................   90
DEB.CM.99.CM5                   -.....AHAG--G...-A-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                    -........AG....LQER-Q--VKA-----K------..-Y-T..---SKS---R---K--RQ-A-VVA------EYH--..........--LTAADVE--D--K-SVSD...........................................................   84
DRL.x.x.FAO                     -..........SAGPEREPPPWFQEYL--VTR-W-GRIP-RDDQ.LPQ-ARQ-KR--A-Q-RVEH--RTLQA---QSLER.........-NSRDLV..............................EGIERIHLAEERESS-................   92
GRV.ET.x.GRI_677                -.............SL-KEE...KQALKI--T--G--..-Y-Q..FS--ARQ--RA-Q---KQ-Q--DK-AG-V-N-FED.......QQLVAQLQELQLENKD-V-QHLPDPPHIHQ............DSSGIPAVWAPATPRGSNRACSSSGEGCEGS-GQTGCYCPIRL  133
GSN.CM.99.CN166                 -........ADHARGNDQKLQN--LAC----T-HR--..-Y-T..SS---S----------A--G-VRE--N---ESLV-........-PE--GD-D--D-GQ-S--SPWDREV..............PVGTAAESNPESNPASETAQGSSSQSD-LH..............  124
GSN.CM.99.CN71                  -........ADPANGRDQKLQN-LLAC----T--K--..-Y-T..AS---S----KK----A----VRQ--H---ESLV-........-PE-SGN-D--D-GH-S--SPEDRQV..............PVGAAAE-....VPTSETA-GSGSQSS-LY..............  120
MND-1.GA.x.MNDGB1               -.......STGN.....VYQ-LIR-YLVVV-K--EGL.PAE-I...PQ-ARQ--R-KQQL-T-RA-LRELEG---KQI-D.........-GPDQLC...............................QGV-NLALAEKSESSN...............   87
MND-2.CM.98.CM16                -..........PTETGDLDRDFWKKYQKIVQR-WEGRLS-S-Q...PQ-ARQ-KR--A-L-KTEY--RQLQS--WEALEQ.........-ANQRLA...............................EDFKGLRVAEEESTSSN..............   90
MND-2.GA.x.M14                  -..........STEPGELGRDFWEKYH-IV-Q-WEGRLS-TSQ...PQ-ARR-KR--AQL-RTEH-VRELQ---WQ-LEA.........-ANKELV...............................EGVNNLHLAEESAGSN...............   90
MND-2.x.x.5440                  -..........STDRGDLDEGFWRKYQAIV-Q-WEGRLS-A-Q...PQ-ARQ--R--A-L-KTEH-VRELQT--...............................................................   61
MNE.US.x.MNE027                 -.............SSHAEE---R-RL---H--H-T-..-Y-TG..PS-ANQ---------Q-WQ-FLALAD-VYSFPD........PPTDT-LD-AIQQ-QN-AIESIPDPP................TNIPEILHDPTE........--RSPQD...............  107
MON.CM.99.L1                    -..........--NGR.DEERRL-SLALAAVRI-QE--..-Y-I..-R--AN-----------RQ-L--A------FYH---..........--Q--C--DI-D----SISDLPDP...............P-SVPEAATP.....PAHTPA-TVGKP..............  112
MUS-1.CM.01.CM1239              -............AGASENVQD----C-I-RI------..-Y-T..SR-S-A---------QQ--D--RA------VAV--..........-PPQ-SD-V--DISQ-SINPS................L-TPEGSSPDTEVVSTT-DRD..................  106
MUS-1.CM.01.SIVmus_01CM1085     -............AGVSEN-R----AC-A-QI--K---..-Y-T..AD---S---------KRK----RE------LSL--..........-PPQ-SD-E--D-NK-S-HPL................VATSE-SPPDTEGVNKI-VHA-GTD...............  110
MUS-2.CM.01.CM1246              -............AGASGN-Q----FC-Q-RI--D---..-Y-T..VSR--AT-K--K---KK--.......AQ--VAVFR........-PPQ-ND-D--D-DQ-R-EPL................VVDSKNLPPPDKGASKS-D-A..................   99
MUS-2.CM.01.CM2500              -............AGT-RDI---LKACQT--I--N---..-Y-S..-G--KN-----------Q--C--RAL-----VS---.........-PPQ-DSVE--D--K-S-KPL................VADPADSPSPPEVEHLS-AQGESNQHP-SGSP...........  117
OLC.CI.97.97CI12                ..............QSHLGRGER-WY-IN-I-GTPPIFRY-P..GQ--ARQ--R--Q-VLREHL-RVELYH-VYR.DIESKLERLTITPV-GSSESHSLDPEELATRLEN.............LKT-ETTWDQFAAEETEKGKSSTEGSGAIGERRSRNNWSLRTSIK  138
SYK.KE.x.KE51                   -...........-GRERE-TQEQLLRT-LRIAQ-LA..-Y-L..-Q-PSR---RN-N-Y-QL-A-RLYVQQ--FE-IAR.............----TLE...QSFGE-QITD...............................................   81
SYK.KE.x.SYK173                 -..............PDQGSEQLA-F-RMIAHLQE..-Y-G.----P--T--R---Q--Q-RT-RLYLQQ--FEAIF-.........SRTAALE...DS-QQ-QISD...............................................   77
TAL.CM.00.266                   -.........SLTGGS-RGEDP-LAAI-R-RI------..-Y-E..QQ-S-S---R-Y---KA--L-VEQ-AS---EH---..........--DRSDQIH--D-SQ-S-ADL.................GPCE-.......CESQEPTGSASHIRLPTTPPDSSSSDSPV  122
TAL.CM.01.8023                  -.........SLTGES-KGEDP-LAAI-R-RI------..-Y-E..QQ---S---R-Y---KA--E-VEN-AS---CH---.........GPQQ-PQ----D--K-S-AEL.................GPPEA.......PESDHPTGSASHIRLPEN...........  111
SAB.SN.x.SAB1                   -.............SL-QEE..-L-RF-I--F--.T-..-Y-P..GQ--ARQ--RA-Q--AKQRQ-VIHLA----E-P............VSQIDHLAQEFDQ-V--NLQQPPSLP........................PGHPTENQTANSSS..................   98
TAN.UG.x.SIVagmTAN1             -.............SL-SE-..-R-IIQI-RI--H--..QY-P..G---ARQ--R--Q-FQQQ---VAAL----FIAQQRRDP.SGGESL-AAFDQLVLDNQQ-VIETLPDPPQEP...............................HDSSSTA.................  103
VER.KE.x.AGM155                 -.............PL-PEE...R-LL---AF--R--..-Y-S..V---ARQ--RA----KN--K--YALA---W--RQE.......DQLVQAIDQLVLDTQH-VTQQLPDPP...................................S-A.................   89
VER.KE.x.TYO1_patent            -.............PL-PEE...R-LL---W---ST-..-Y-PS.G---ARQ--RA-----QQ-D--RVLV--LQ.............EQVYA-DRLADEAQH-AIQQLPDPP...................................HSA.................   84
WRC.CI.97.97CI14                ..............GFLIVLTARKLSQIQQQ-L-AE..--QQL.GR--ARQ--RN---LQQYRN-RLAVQ---CNSL-E........QL-NLTIQH--TPPSHCVESAFVPPVVDK.........RWV-SSQRHQWKSIWSD--EDCKE-EGEYAGRGISNWQQKDS...  133
WRC.CI.98.98CI04                ..............GFLIVPTAR-L--VQQK-Q-AE..--QQL.GR--ARQ--RS---QQYFR--RLAVQ---CNNLVE........QLDKLLICS--NPPSNSVESPCVPLLVDP.........AWV-SSQKNQWSSACMH--EDYKE-EGNNGERDIRK..........  127
WRC.GM.05.Pbt_05GM_X02          ..............GSLIIHTAR-LI-ISQT-Q-AE..SY-EP.GR--ARQ-------FHSHR-IRLAVQ---AD-L-E........QLGNLTLQH--EPPAAK-QPSELSVDRNT.........EWV-SSQKTLWSSVCCA-GEDCCASAEKKGEGDIHAWQQRDK...  133
LST.CD.88.SIVlhoest447          -.......STG-.....EKE..SP-YLKIS-I-WTGL.QVGEKR.EPR-ARQ--RD-E-Q-KHLH-LRTVQ---FHSTVE.........-GL-RAFTR-TVCDSPEV-QGLWNSSPQPCVAVPFVARPFSDPFLPTWATSSESLQ-TG-GKRSEDCELELEQDQKEQKRHIS  147
LST.CD.88.SIVlhoest524          -.......STGP.....EGE..-PTYLK-SRI-WTGL.QVGERR.EPR-ARQ--RD-E-Q-KHLH-LRAVQ---FQATVE.........-GL-RAFTG-TVCDSPEVAEGVGNTAPHPRVAIPHLAGTYSDPFLPPWATSLANPQ-IP-GECSENCELDQEQEQKEQRRAIK  147
LST.KE.x.lho7                   -.......STGN.....GDE..-P-YL--SRI-WTGL.QVGERR.EPR-ARQ--RE-S-Y-DYLH-LRAVQ---FQATVE.........-GL-RAFTR-AVSDSPEVAQGRGNTPPITSVAEPQLAVAFVDPFLPKWATPLADQQ-MD-GKRSEDSELAQG-MQKEQRTVIE  147
SUN.GA.98.L14                   -.......ST.......GD-.SINQYL-IS-R--EGL.A-GNL...PQ-HRQ--R--D-E-KNLH-LRAVQ---FA-T-D.........SRLGRAFER-SVSDSSQVAESLGNSPSTKHLPPAKFLVAPTY-FLP-WATPLADPQRLA-FAPYSGYEQDQERVQNQQGESII  144
COL.CM.x.CGU1                   -TNAGVRPV-FTVSYAFFQKGWESAI-R-RV-HKH-..-W-RP.GV--ARQ--RNTQ-A-KQ-Q--LA-DG-VVEQLVAS......LQQWD-YCPASTSANN-RQEEE..................CLADSSVS........I-TQGSKPSDS................  118
```

```
H1B.FR.83.HXB2              ........................*   116
H1A1.UG.85.U455             ........................-   124
H1C.ET.86.ETH2220           ........................-   108
H1D.CD.84.84ZR085           ........................-   117
H1F1.BE.93.VI850            ........................-   117
H1G.SE.93.SE6165            ........................-   108
H1H.CF.90.056               ........................-   117
H1J.SE.93.SE9280_7887       ........................-   117
H1K.CM.96.96CM_MP535        ........................-   117
H1O.BE.87.ANT70             ........................-   103
H1O.CM.94.BCF06             ........................-   104
H1O.CM.96.96CMA102          ........................-   103
H1O.CM.96.96CMABB009        ........................-   104
H1O.CM.97.97CMABB447        ........................-   104
H1O.CM.98.98CMA104          ........................-   103
H1O.CM.98.98CMABB197        ........................-   103
H1O.CM.98.98CMU2901         ........................-   104
H1O.CM.99.99CMU4122         ........................-   103
H1O.FR.92.VAU               ........................-   104
H1O.US.99.99USTWLA          ........................-   104
H1O.US.x.I_2478B            ........................-   104
H1N.CM.02.DJO0131           ........................-   107
H1N.CM.02.SJGddd            ........................-   105
H1N.CM.04.04CM_1131_03      ........................-   105
H1N.CM.06.U14296            ........................-   105
H1N.CM.06.U14842            ........................-   105
H1N.CM.95.YBF30             ........................-   105
H1N.CM.97.YBF106            ........................-   106
H1P.FR.06.RBF168            ........................-   105
CPZ.CM.01.SIVcpzCAM13       ........................-   124
CPZ.CM.05.SIVcpzEK505       ........................-   104
CPZ.CM.05.SIVcpzLB7         ........................-   124
CPZ.CM.05.SIVcpzMB66        ........................-   124
CPZ.CM.05.SIVcpzMT145       ........................-   122
CPZ.CM.98.CAM3              N.......................-   125
CPZ.CM.98.CAM5              ........................-   108
CPZ.CM.x.SIVcpzDP943        N.......................-   125
CPZ.CM.x.SIVcpzMB897        ........................-   122
CPZ.GA.88.GAB1              ........................-   125
CPZ.GA.88.GAB2              N.......................-   124
CPZ.TZ.00.TAN1              ........................-   106
CPZ.TZ.01.TAN2              ........................-   106
CPZ.TZ.02.TAN3_1            ........................-   106
CPZ.US.85.US_Marilyn        ........................-   106
CPZ.CD.90.ANT               ........................-   119
GOR.CM.04.SIVgorCP684con    ........................-   105
GOR.CM.07.SIVgor2139_287    ........................-   105
RCM.CM.00.SIVagi_00CM312    ........................-   108
RCM.CM.02.SIVrcm_02CM8081   ........................-   105
RCM.GA.x.SIVRCMGAB1         ........................-    94
RCM.NG.x.NG411              ........................-   108

MAC.US.x.239                ........................-   108
H2A.DE.x.BEN                ........................-   104
H2B.CI.x.EHO                VGSPQESGRRDHCNTQEDQTRG....-   151
H2G.CI.92.ABT96             ........................-   101
H2U.FR.96.12034             ........................-    99
SMM.SL.92.SL92B             ........................-   100
SMM.US.x.H9                 ........................-   101
SMM.US.x.PGM53              ........................-   105
SMM.US.x.SIVsmH635F_L3      ........................-   101
STM.US.89.STM_37_16         ........................-   100
DEB.CM.04.SIVdeb04CMPF3061  ........................-    98
DEB.CM.99.CM40              ........................-    91
DEB.CM.99.CM5               ........................-    88
DEN.CD.x.CD1                ........................-    85
DRL.x.x.FAO                 ........................-    93
GRV.ET.x.GRI_677            SGSHQQSKKSAARP............-   148
GSN.CM.99.CN166             ........................-   125
GSN.CM.99.CN71              ........................-   121
MND-1.GA.x.MNDGB1           ........................-    88
MND-2.CM.98.CM16            ........................-    91
MND-2.GA.x.M14              ........................-    91
MND-2.x.x.5440              ........................-    62
MNE.US.x.MNE027             ........................-   108
MON.CM.99.L1                ........................-   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
OLC.CI.97.97CI12            QIWQ......................-   143
SYK.KE.x.KE51               ........................-    82
SYK.KE.x.SYK173             ........................-    78
TAL.CM.00.266               WVQ.......................-   126
TAL.CM.01.8023              ........................-   112
SAB.SN.x.SAB1               ........................-    99
TAN.UG.x.SIVagmTAN1         ........................-   104
VER.KE.x.AGM155             ........................-    90
VER.KE.x.TYO1_patent        ........................-    85
WRC.CI.97.97CI14            SKQ.......................-   137
WRC.CI.98.98CI04            ........................-   128
WRC.GM.05.Pbt_05GM_X02      QKQQHQGNYCQTLHKP..........-   150
LST.CD.88.SIVlhoest447      SSCKKPQTGTKEEMEIQIQRKKWVSQ-   174
LST.CD.88.SIVlhoest524      N.......................-   149
LST.KE.x.lho7               H.......................-   149
SUN.GA.98.LI4               VSEGKK....................-   151
COL.CM.x.CGU1               ........................-   119
```

```
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
```

PLV Proteins

Vpu

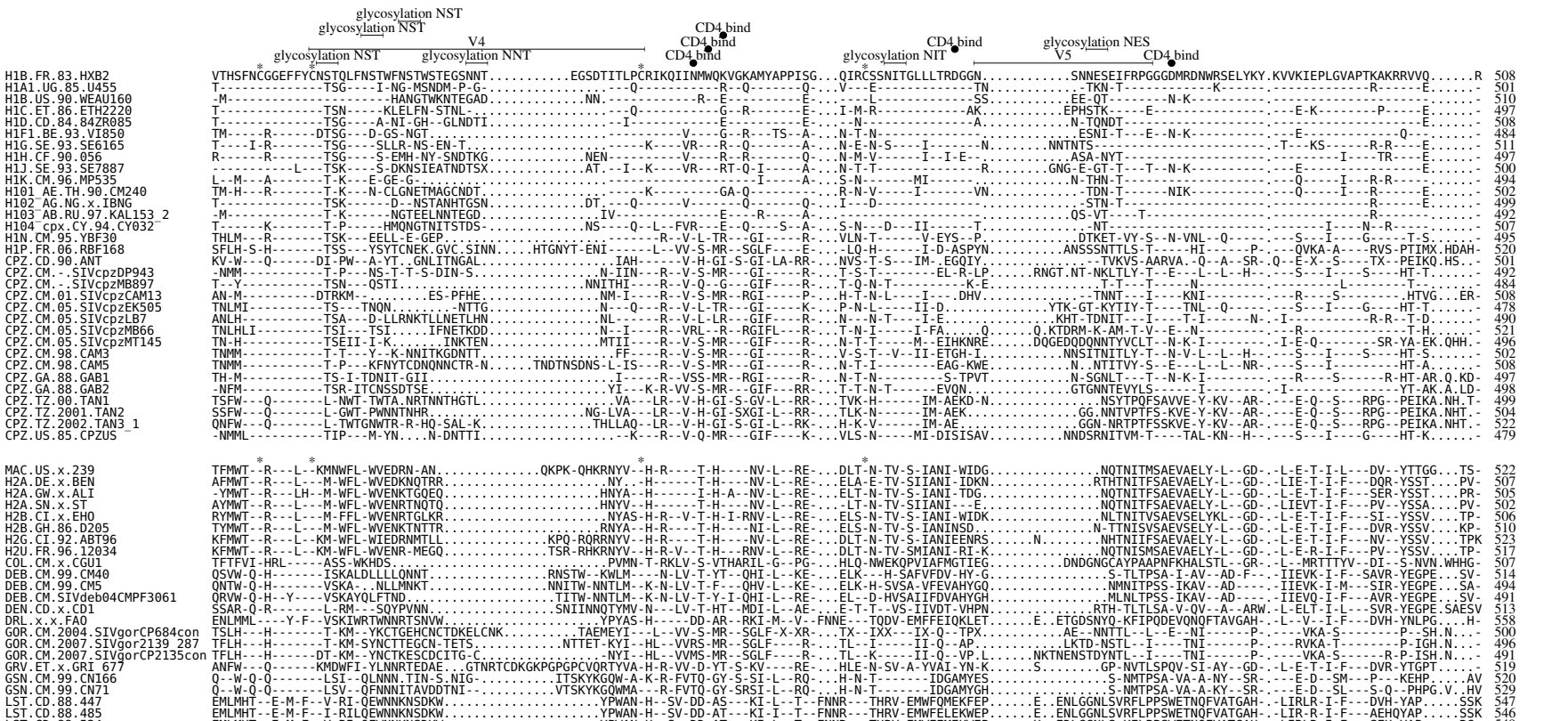
```
                                                                                                                                        glycosylation NDT
                                                                                                                                               V1
              Env start                 signal peptide gp120 start                                     glycosylation NVT                CD4 bind   glycosylation NSS
H1B.FR.83.HXB2           MRVK...EKYQHLWRWGWRWGTMLLG.MLMICS..ATEKL.....WVTVYYGVPVWKEAT...TTLFCASDAKAYDTEVHNVWATHACVP.TDPNPQE...VVLVNVTENFNMW..KNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGR....  146
H1A1.UG.85.U455          ---M....GI-RNYPCL------I--.LII--...NAQQ-.....------------D-V..-------------A------------------.--------...ID------E----..--N--D------------------------T-D-HNITINNTN-NTN......  142
H1B.US.90.WEAU160        -------...GIRKNYQHL-K--I----.I-------..-A-N-.....-------------...--------------------------.--------...---E---------.--N----------------------------T-N--NVNVTNLK-ETNTNSSSG  149
H1C.ET.86.ETH2220        -K-M....GI-RNCQQW-I--ILGFW.-----N..GMGN-.....------------D-S..P-----------------G-F---.---S---...LG-E---------.-----------Q--------G-----------T-N-NAI--N-KVTNN-......  143
H1D.CD.84.84ZR085        -------...GIKRNYQPL-K--I----MLM-TY-..-ADN-.....-----------------..----------S-K--A--I---------.--------...IE-K---------.--N--D----------------------R--T-N---ASRNSTD-N-T......  144
H1F1.BE.93.VI850         ---R....GM-RN-QHLGK--LLF--.I-I--N..-ADN-.....-----------------..------------ER-A-----------------.--------...-F-K-----D--.--N------T-------------------T-N--NAT-NSQEKPGA......  143
H1G.SE.93.SE6165         ---T....GI-RN-*HL-K---LI--.LVI----..-SNN-.....-------------ED-D..----------S-SA-S----------------.--------...IIME---Y----.--N-------------E-----------T-N---VT-KG-KRNN-TDNSTE  148
H1H.CF.90.056            -E.T......-RNYPSL-----LI--.--L----..-AQN-.....----------------K..------------E--K----------------.--------...M-ME----S----.E-N------T----------------------T-N--NVR-N-SNST--......  140
H1J.SE.93.SE7887         -E.T......-KN-QTL--G-L-IF-.-----K..-K-D-.....-------------D-K..------------S--K-----------.---S----...MN-P----------.-----D--Q-----V--E------I-----T-N-SNITSNS--T-N-S.....  141
H1K.CM.96.MP535          ---R....GM-RN-QTLGN--ILF--.I-I----.NAD--.....-----------------..P-----------EK-----------------.--------...-EME---------.--N------T-------E------E------T-N---Y-GTNS--NATST....  145
H101_AE.TH.90.CM240      -------...ET-MN-PNL-K---LI--.LVI----..-SDN-.....-------------RD-D..-----------HE---------------.--------...IH-E---------.--N-----Q--V-------------------T-N--NANLTNGSSKTNVSN...  146
H102_AG.NG.x.IBNG        ---M....GI-KNYPLL-----NIFW.IMI--...NA-Q-.....--------------T-E..-------------------------.--------...IH-E----K----.--N----------------------------T-D-HNFN-SYSNS-NLTS....  144
H103_AB.RU.97.KAL153_2   ----------...EIRKHL-----LF--.-----------N-.....-----------------..-----------SK---------Y----.---S----...IP-K-------G..--N----------------------------T-N-----KEVTST.........  137
H104_cpx.CY.94.CY032     ---M....GM-RNYPHL-E---LI--.LVI----..-SNN-.....-------------RD-E..--------E----EK----I-----------.--------...-A-I-------------.-----------------NEG----A---S---TFT-INATTTNS--GT.......  142
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.FR.06.RBF168         -K......-KKNSLGNLEICLALVIYFNAIS-AS.GIH.......Y-----------RN-E..V-----A--SLTSK-Q--I---Q----.---T-I-....-KI----S--I-..--Y--T--Q------------X----X---TMN-SE..CHKQAKC........  136
CPZ.CD.90.ANT            --.........KPI..HII--.LA-..LIQFIEK.G-NED.....Y---F-------RN--..P-----TN-SMTS----------TS---.I--D-I.....-RL-.-SVWFNAY.--Y---S-T--MXQ-FQ--H---------M-IKMN--GYNGTPT-P-TTTST...  137
CPZ.CM.-.SIVcpzDP943     -ER.............I-GSLW-M-.LIIP-LTG.SDQ-.....-A-----------D-E..-------A--S-LNK-A--I--SQ-----------.--------...-HVP-I--T-D--..E-N-----QT----------------------TMH-IPYNISSTG-DTRNSS...  137
CPZ.CM.-.SIVcpzMB897     -GT........QRSWKQFLSLRIVMC.LWLLGVIRSENN-.....-------------K..------------QIA-A--I--SQ----.------E-...IT-E----E-DA-..N-N--D--Q--L-------------------------T-N-STNTTTEAS..........  137
CPZ.CM.01.SIVcpzCAM13    -K-M...-ERKRSFNGFYMTM-ICMA.L-IP-LT...SDE.....-------------R-VN..-V----------HS--A--I-----------.--------...---H----D-----.N-Q------Q---S--------------------TM--SNVTRQNS-SN-N......  143
CPZ.CM.05.SIVcpzEK505    -K-........TEMQKN-LICCL-I-.LIK-IG....SE-.....-------------RD-E..-V----------HS--A--I---Q---------.--------...-LIP----R-D--..--N--D--Q--------E-------------T-S-SSWRSVN-S..........  133
CPZ.CM.05.SIVcpzLB7      -----.......GT-TSCMPYYVV-.I-SL--L.IISSTKNENL-------------E..A------------QKA-A--I--SQ----.---D-KA...IE-S-----------.T-A--N--QQ-V-------------I-------T-H-SIP-F-NSS..........  140
CPZ.CM.05.SIVcpzMB66     -ETQRNCRTLSLKEIILCTLVLGII-.IIK.-E....DNM.....-------------R--D..-------------QNP--------SQ--S.-N---E-...IE-T----Y--A-..E-N----------VN-----V------I----T-N-SLF-CIKENGNTTNC....  146
CPZ.CM.05.SIVcpzMT145    -GM...KYYWWRNSCRSISIKLI-I-.WIAS-FG..E-NW.....-------------R--K..-----------S-S--A--I---Q-----.---T------...-L-P----E-----.E-Y--D--Q--------E--------------T-T-NNPT-TSC-..........  140
CPZ.CM.98.CAM3           -K-M.......EKIKGSLIEKY-FM-.LIIP-LT.GSDQ-.....-A-----------D..--------E-S-LNK-A--I--SQ---------.---------...-QIP------------.N-T-A---NT---------------------I--TME-RKVTFNSTS-RNKTSTMTT  147
CPZ.CM.98.CAM5           -KAM.......ERKKLN-IICY--M-.LITP-LT.GH-.W.....-A-----------E..--------E-S-LNK-A--I--SQ-----.V--------...-YMP--S-S-------.--N-A---Q---S-------------------I--TMQ--T---IST--TN-......  140
CPZ.GA.88.GAB1           -K-M...--KKRD-NSLSIITIITII.L-TP-LT..S-.-.....-------------HD-D..PV-----------HS--A--I---Q---------.---S-----...-F-P--I-S--------.--N--D-------------------------T-Q-SKANFSQAK-LT.......  142
CPZ.GA.88.GAB2           -K-M....GKTRRRW...QPYCIIMA.LIIP-LKTESDQW...................-P----AN-T-LKE-P--I---Q-----.---S-E-...-I-I----E--V-..D-A-----Q---T-------R-----N----Q-T--SINATG-ETNAT-.....  143
CPZ.TZ.00.TAN1           -KNL...............IGIT-ILIIT-LGI.GFSTY.....YT--F---------Q..P--------DITSRDK--I----N---.L----Y-...-T-A--SIR---E..E-Y--QE-K---L--FQ--F---------F-IKMT-.TMT-T--KT.........  129
CPZ.TZ.2001.TAN2         --N............LIGTTL-LI...IIILGIG..SKEY.....YT--F----I------..PS-----N-DIASRDK--I----N---.L----Y-...-X-A--SIX-D-E..--Y--XE-K--LL--FQ--F---------F--TM---VVNTTVVNATAPT.....  134
CPZ.TZ.2002.TAN3_1       --N............LIGITLILI...I-TLGIG..SRDY.....YT--F----I--D--..PS-------DVTSR-K--I----N---.L----Y-...-PIA--SIK---E..E-Y--KE-K--LL--FQ--F---------F--VMN--KSQ-V-TPA-T-TAA...  136
CPZ.US.85.CPZUS          -K-M....EKKKRLW....LSYC--SSLII...P.GLSS-.....-A----------RDVE..-------------KQ-A--I---Q-----.---------...-H-P----K-D--..E-N-A---Q--------------I-------TMT-LNPDSN.............  131
                                            signal peptide end gp120 start
              Env start
MAC.US.x.239             -GCL..........GNQL.....-IAIL-LSVYG.IYCT-.....Y---F----A-RN--...IP---.........-KNRDT-G-TQ-L-.DNGDYS-.....-AL----S-DA-..N-TVT--AI--VWQ-FET-I------S---ITMR-NKSET-RWGLTK-ITTTAS  132
H2A.DE.x.BEN             -EPG..........-NQL.....FVVIL-TSACL.VYCSQ.....Y---F--I-A--N-S...IP----.........-KNRDT-G-IQ-L-.DNDDY--....IIL---A-DA-N..-TVT--AV--VWH-FET-I----------AMN-SRVQGN-T-PNPRTSSSTT  132
H2A.GW.x.ALI             -MSS..........-NQL.....-VTIL-ASACL.VYC-Q.....Y---F---A--N-S...IP----.........-KNRDT-G-IQ-L-.DNDDY--....IAL---A-DA-D..-TVT--AV--VWR-FET-I----------IAM--SNISTEST-T-P-PGSTL.  131
H2A.SN.x.ST              -.CG..........-NQL.....FVASL-ASACL.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
H2B.GH.86.D205           -AYF..........SSRL.....PIALL-IGI-G.FVC-Q.....Y---F--I-A-RN--...VP-I--.........-TNRDT-G-VQ-L-.DNGDYT-....IRL-I--A-DA-D..-TVTQ-AVD-VWR-FET-I-----------AMN-SKTETNPGNA--TTTTKPT  132
H2G.CI.92.ABT96          -AYL..........GNQL.....-IATX-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.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
COL.CM.x.CGU1            -LRY..........LFIN..........L-VLGIV.L.SNK.....-----Q---A-E--DVNDQQF--F-SSPE.......IQQVLG-L-PPPGK-V-...QNMP----A-DLF..--SFSGEVWIITQTTLE-R-R--A---AY-APMI--KVNRTE-GT-TVAPTT..  130
DEB.CM.99.CM40           --LI..........QLLL...LI-GLIVIGTQGK.QKKQQ.....YL---------VD-K...VD---T......ANSSESG--VT--L-..HALVR-....EVPMPNVTQ-FNAFD-PIE--LWQ-MT--YK--F---------Y---MQ-IKTSTNPTPTNT-TTTTTI  136
DEB.CM.99.CM5            -LSL..................LIIIASLTLVGAA.P-KQQ.....-----------HN--...VG---T......ANSSESG--VT--L-..HAMVR-....QVPMPNVSQYFDSFS-V-EQ-LWD-MT--YR--F---------Y--RMN-SSTTTT-TAATTTI.....  127
DEB.CM.SIVdeb04CMPF3061  -LAF...................I-AIIVI-LGQ.K--QVSQQQQY---F-------N-S...VG---T......ANGSESG--VT--L-..HSVIR-....QVHMPNVSQYFDAF--PIED-LW--MTG-YK-TYR-------Y--KMR-KTETTL-SNLTTP......  129
DEN.CD.x.CD1             -SQV..........-RTL...KV--LILWT-GIICT---Q.....Y---F--I-A-RNSA...AP---T.......AMDRG---AAS---.E--YQR-.....VWMENATDYFDIYH-PI-D-VVQNMKD-YK-TF---------V--RMN--RKEDNSTNAVTVSSTTTT  136
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---IVNAASTTPTPTT..  131
GOR.CM.2004.SIVgorCP684con -SL......................GL-LT-IE.-KTEH.....YA---------R--K...V-----A--SFVSK-Q--I---Q--------.---S-T-...FL-E----T-DI-K..-Y-----Q-----------------------VM--TMT-SNDTKGS-KSTATTTSANT  132
GOR.CM.2007.SIVgor2139_287 -SL.....................-.LILT-IE...ATH.....YA---------R--N...V-----A--NYVSK-Q--I---Q--------.---R-T-...FP-E----T--I-K..-Y--D--QD--V---------------VM--T-N--Q-NTTNT-QTEMTNT...  127
GOR.CM.2007.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-XQTNXTN.....  125
GRV.ET.x.GRI_677         -GRL..........LIKI....LIIAIGI.SI.G..IGN-.....Y---F--I----NS-...VQA--M.........-PNT-M---TN-I-.D-HDNT-...-P-.-I--A-EA-D..-PL-K-AESN-HL-FE-TMR-----S-I-IKMS-VE-NGTAT-KATTTATTTM  130
GSN.CM.99.CN166          --KT.........VESMMMLAVS--AL-ALLIPG.LKCEN.....-T----------RD--...PP------PDIASN-PG-I-IST--L-.S--S-A-.....-PL-I--K--IY..--Y--DEVRD-MV--FN-A----------M--GMY-NLTNTS-SEPTTTP-PPNV  145
GSN.CM.99.CN71           -MR-.........VGFLKMLVVS-TLMA-L-PGS.RSKGN.....-T--F-------RD-K..PP-------DITSN-PG-I-IST--L-.S--S-S-.....-PL-I--E--IY..--Y--DEVRD-MV--FN-A-R--------M--RM--KLPNTS-T-TP-TTTATTS  145
LST.CD.88.447            -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.88.485            -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.88.524            -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
MAC.US.x.251_BK28        -GCL..........GNQL.....-IAIL-LSVYG.IYCTQ.....Y---F----A-RN--...IP----.........-KNRDT-G-TQ-L-.DNGDYS-....LAL----S-DA-E..-TVT--AI--VWQ-FET-I------S---ITMR-NKSET-RWGLTK-STTITT  132
MAC.US.x.EMBL_3          -GCL..........GNQL.....-IAIL-LSVYG.IYCTQ.....Y---F----A-RN--...IP----.........-KNRDT-G-TQ-L-.DNGDYS-....LAL----S-DA-E..-TVT--AI--VWQ-FET-I------S---ITMR-NKSET-RWGLTK-STTITT  132
MND-1.GA.x.MNDGB1        -KCP..........GLFL.....I-CLA#LV..G.IIG-Q.....Y---F---------K...-H-I--.........-DNSSL-V-TN-I-..SLPDYDE..-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             --KL..................-A-STLCLMMTL.VNGSN.....-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             --MR.......MKTIIVLTILCLMM-.T-TSG-X.SGN-S.....-----F-----Y-P--...PP------PXYGSK-AG-N-LASS-L-.S--V-RA...LPL-I--S--AYN..-Y---E-V---K--FS-A----------M--RMR-QXINES-TVXPTTTXX.AT  145
MUS-1.CM.01.1085         -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-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-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
OLC.CI.97.97CI12         --K-..................LV-CLLWLSILGISIG.Q.....Y---F--T-K-RDSN..D--M-V.........-ANKSL-V-T--L-.SFDEDEI..EQPIPGV-EVFYNVSD-EV-R-AWFG-A--L-AV-----------F--KMQ-E-PSANFTEPITPISTAAP  132
RCM.CM.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.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.GAB1            -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.SAB1C           -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.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.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.US.x.SIVsmH635F_L3   -GCL..........GNQL.....-IALL-VSVLE.ICCVQ.....Y---F----A--N--...IP----.........-KNRDT-G-TQ-L-.DNDDYS-....LA--I--A-DA-N..-TVT--AI--VWN-FET-I----------IAMR-NKTET-RWGLTGRAETTTT  132
STM.US.x.STM             -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          -AAF..........-TYI.....VCLFS-ISLGF.MEKQQ.....Y---F--I-H-ED-Y..AP---T.......-SHKGG---KN---.SADQIEV...RVNITG-Y-PA-NSSHMIRQ-IL--MSA-FL-ANR-----A-M-IRML--LDNSPATS..........  124
TAL.CM.00.266            -KYL..........AVI.........FCLAYLGRF.IIGNQ.....Y--I---------N--...PP-----------.......-MARDQ-S-SQ---.S--QNL-....-P--L--Y-DVY..--Y--DHIA--MD--FLTATR--T----I--RM--E-ITKNLTISANTTTPAPT  129
TAL.CM.01.8023           -KCL..........QVLCLAICVIWVANIGKSLT.SA-QQ.....FL-----------N--...PP-----------.......-MARDQ-S-SQ---.S--QNL-....-P--L--Y-DVY..--Y--DHIAD-MDA-FLTAT---T----M--RMA---ITKNLTGTVAPPTTNST  137
TAN.UG.x.TAN1            -GPL..........-GKG...VL-VILG-SLIGL.LYGTQ.....YI--F--I----NSS...VQA--M.........-PNT-L---TN-I-.D-HDYT-...-Q-.--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-...-P-.-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-...-P-.-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-...-P-.-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
```

V1 — glycosylation NCS — glycosylation NIS — V2 — glycosylation NDT — glycosylation NTS — glycosylation NKT — glycosylation NGT — glycosylation NVS

```
H1B.FR.83.HXB2             ..................................M.IMEKGEIKNCSFNIS.TS.IRGKVQKEYAFFYKLDIIPIDNDT.................TSYKLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGP..CTNVSTVQCTHGIRPVVSTQL  259
H1A1.UG.85.U455            .................................IT.DGVRE-M------MT.-E.L-D-K--V-SL--R---VQ-NKTDN................N--R-IN----T----------------------------KDPE---K--..-R-----------K-------  257
H1B.US.90.WEAU160          G................................E.K--E--M------VT.-L.--N-RKT---L-----VM---H-N.................---T-IN-KS-T----------------------------D-K---K--..-K------------------  263
H1C.ET.86.ETH2220          .................................I.NSAND-M-------T.-E.L-D-KR-A--L------V-LN-GS..................-D-R-IN----T--------LD-----------Y-----RD---T----..-H-----------K-------  256
H1D.CD.84.84ZR085          ................................LP.TVKP--M-------T.-V.VTD-RKQVH-L--R--VVQ---EGKNEIN..........DTYGT-R-IN----A----------------------------D-R------..-K---S---------------  266
H1F1.BE.93.VI850           .......................................MQ-----MT.-E.V-D-KL-LS-L--R---V--G-NNS.................SE-R-IN----T--------WD-----------Y------D-R------..-K-----------K-------  250
H1G.SE.93.SE6165           TNN..............................ST.VDNP---------VT.-E.--D-KK-------R--VV--N-A..................-N-R--H--V-T-K------T-D---------T------RD-E------..-K-----------K-------  264
H1H.CF.90.056              ...................................MEAG--LT-----VT.-V.L-D-Q--VH-L--R--VV----NS...................-Q-R-IN--------------------------------------------L..-------------------  252
H1J.SE.93.SE7887           ...................................VSSPDIMT------T.-E.--N-RKQ---L--RQ-VV---SNN.................KN-I-IN------K--------Q-----------------D-N-----S..-K-----------K-------  253
H1K.CM.96.MP535            ...................................VVSPA---------T.-E.-KD-KK--S-L--R--VL-LNGEGNNSS...............-E-R-IN----T---T----T-------------------KD-R-------..-K-----------K-------  261
H101_AE.TH.90.CM240        ................................II.GNITD-VR--T--MT.-E.LTD-K--VH-L------VQ-EDKKTS.................SE-R-IN------K-----I--D-------T---Y------D-N-------..-K---S--------K-------  262
H102_AG.NG.x.IBNG          ....................................DMN---------T.-E.V-D-KK-MH-L--R--VVQ-NENNG..................SQ-R-IN----A----------------------------KD-G-------..-K-----------K-------  256
H103_AB.RU.97.KAL153_2     ...............................NTS.SIKMM-M-------T.-D.L-D--K----L-----VVQ-...................-R-I------V-----I------------------------D-K-------..------------K-------  250
H104_cpx.CY.94.CY032       ....................................VI-EG------D-T.-E.--D-KK----L--RI--V--NARVPING.......SNRNNSTEE-M-IN--A-T-K------------------------E-N-T-L--..------S-R-----K-------  264
H1N.CM.95.YBF30            ..........................MTTREP.DIGYKQM------AT.-E.LTD-KKQV-SL--VE-VV--NA..................YNKT-R-IN---TAV------T-------------P----M---EGN-S-N-S..-------------K--I----  262
H1P.FR.06.RBF168           ........................NNPSENRT.AANPL-LFK----TT.-V.LKD-K-EQQ-L--RT-L-ALNETANNT.................L-R-IN---TT--------T---L--Q------Y-LM--XQXG--------..-NQTVITH-----K-XX---X  258
CPZ.CD.90.ANT              ........................VTPKTTTP.-VDGMKLQE-N--Q-.-G.FKD-K--MK-I---G-LMKCQDNNETN...............C-Y-WH---TT---S-E-ST--------------Y---R-EDED-T-V-M..-K---V-H-----S-M-A-W-  259
CPZ.CM.-.SIVcpzDP943       ..................................TSQMV-K-F--MT.-E.L-D-QKQV-SL--VD--VX-..............NGXX--R-IN---TA-------T------------P----M---EPD-----K..-R-I----------K---T-X-  246
CPZ.CM.-.SIVcpzMB897       ....................................QVQY-----VT.-E.L-D-KKQV-SL--RE--TSL-SNKT...............VKNGT-R-IN---TA-------T-------Y--------L----DQN-K-K-T..-R-----H-----K------F  249
CPZ.CM.01.SIVcpzCAM13      .....................NNPTLKNNTSGKN.ETGVLQM---T--TT.-E.L-D-KKQV-SL--VD-LQSLGS...............GTGDT-TMIN---TA---------------------------DVN-S-K-K..-R-----H-----K---T---  269
CPZ.CM.05.SIVcpzEK505      .................................VN.QTNHVQMQ-----VT.-E.L-D-KKQV-SL--MG----L-TNNSSG..............NNSQ-R-IN---TAV------I-------Y----P----I---DQD-----E..-N-----------K--I----  252
CPZ.CM.05.SIVcpzLB7        ..................................I.NSSNETM------TT.-E.L-D-KKQT-SL--VD-LTQ-NK................TTE----IN---TA-------I-----------------L-R-----Y--S--..-N-----H-----K--I----  253
CPZ.CM.05.SIVcpzMB66       ....................TVQISTGNDSTANNI.TVGTIDMY-----AT.-E.L-DRKKQV-SL--RQ-LE-LEGNKPPE..........GDKNAL-R-YN---TAM----S---L----------------L----D-N-T-I-Q..-K-----H------------F  280
CPZ.CM.05.SIVcpzMT145      ...............................NST.DDRL-DMR-----VT.-E.L-D-KRQV-SL--VE--TA-G-..................NST-R-IN---TA-------T-----------------L----DIDYK-NET..-K-----H-----K--AT---  254
CPZ.CM.98.CAM3             ..............................NSPN.EKXDSTV---T--MT.-E.V-D-EK-V-SL--VD-LVL-----D...............T-R-IN---TA-------T-----------TP-Y--M---MPN-----TGR-N-I-------------T---  265
CPZ.CM.98.CAM5             ...............................TEL.KNI-EIM---T--MT.-E.V-D-KKEV-SL--VD--V--GGNKSND................T-R-IN---TTV------T------------P----M---DQK--------..---I-------------T--V  257
CPZ.GA.88.GAB1             ...............................NQT.SSPPL-M------VT.-E.L-D-KKQV-SL--VE-VVNLG-ENN................T-RIIN---TA-------T-----------------------D-D-S-K-K..-------H-----K---T---  257
CPZ.GA.88.GAB2             .............................NGIEK.DGLAQDMR--T--TT.-E.L-D-K-QI-SL-W-N-LVGTN-.................TFR-IN---TA-------T-----------------L----D-DYP-K-K..-K-----H-----K-T-T---  257
CPZ.TZ.00.TAN1             ....................LNSATTTLTPTVNLS.SIPNY-VY-----QT.-E.F-D-KKQI-SL--RE--VKE-GN..............NN--Y-HN--------E-D-ST------R---------L---RDQN-T-K-Q..-S---V-H-----Y-MIA-A-  256
CPZ.TZ.2001.TAN2           ..................PKVSATVATTP.TINLT.NIKEY--F-----QT.-E.F-D-KKQI-SLX-RE--MKVGSEST..............KN-D-Y-HN----A---E-D-ST------R--------M----DQGYT-N-T..-S--TV-H-----F-MIA-A-  265
CPZ.TZ.2002.TAN3_1         ....................PVNATTIP..VNLT.DIPNY-LF----QQT.-E.F-D-KKQI-SL--RE--MKDNRSMQDNRSN..SSKKSNNSNNSG-Y-HN----A---E-D-ST------R------Y-M---KDQNYT-N-I..-Y--TV-H-----F-MIA-A-  276
CPZ.US.85.CPZUS            ...............................SSA.VNTTDIMR------T.-E.L-D-KKQV-SL--VD-LAH-N-N.................T-R-IN---TA-------T-------------P---------E-D-K-K-E..-K-----------K---T---  244

MAC.US.x.239               TTSTTA.........SAKVDMVNETSSCIAQDNCT.GL-QEQMIS-K--MT.GL.K-D-KKEYNETW-SA-LVCEQ.............GNNTGNESRCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTNYS-FM-K.-SK-VVSS--RMMETQT--WF  275
H2A.DE.x.BEN               SRPPTS...........AASIINETSNCIENNTCA.GLGYE-MMQ-E--MK.GL.EQD-KRRYKDTW-LE-VVCDN................TTAGTCYMRH----I-KES-D-HYWDAMRFR----P---L-R--DTNYS-FE-K.--K-VAAS--RMMETQT--WF  269
H2A.GW.x.ALI               .................KPLINESDPCIKADNCPRGLGDE-MV--R--MT.GL.Q-D-PKQYNETW-SK-VVCEP............FNTTTNQ-RCYMNH-------ES-D-HYWDA-RFR----P-Y-L-R-DDINYS-FA-N.-SK-VAAT--RMMETQT--WF  266
H2A.SN.x.ST                .................NTTIGENSSCIRTDNCT.GLGEE-MVD-Q--MT.GL.E-D-KKLYNETW-SK-VVCES.............NDTKKEKTCYMNH-------ES-D-HYWDTMRFR----P---L-R--DTNYS-FE-N.-SK-VAAT--RMMETQT--WF  263
H2B.CI.x.EHO               SSTQ..............TLLNEDSKCIQNDSCA.GIGLE-MID-Q-KMT.GL.K-DESKQYKDTW--Q-LVCEKGTRSN...............ESKCYIKT----I-QES-D-HYWDSLRFR----P---L-R--DTKYS-FM-N.-SK-VVSL.YRMMETQT--WF  267
H2B.GH.86.D205             TTS.............RGLKTINETDPCIKNDSCT.GLGEE--MQ-N-SMT.GL.R-DELKQYKDTW-SE-LECNN...............TRKYTSRCYIRT---TI-QES-D-HYWDSLRFR----P--FL-R--DTNYS-FM-N.-SK-VASS--RMMETQS--WF  269
H2G.CI.92.ABT96            TVPTTKMVTA.........ELVNSTSQCLMYDNCT.GIQSESMVG-K--MT.GL.K-DQRKEYNETW-SQ-LQXEQ.............SNKSENESRCYMRH------QES-D-HYWDA-RFR----P---L-R--DTNYS-FA-N.-SK-VVST--RMMETQT--WF  275
H2U.FR.96.12034            PAMQESS............KVINETEPCIRNNSCS.GL-QEPLVS-R--MT.GL.K-D-K-EYKETW-SH--VCEQ.............NTSESESKCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTNYS-FM-N.-SK-VVST--RMRGTQT--WF  271
COL.CM.x.CGU1              ........................TNNSASDWDES.NWKEYPWY--RM-STAFL.LKDRKEL-LG-SVEDLTVLGNKNDSNS..............IRATMKD-ANYTV--V-DMTIVD-VRTGF--AP-YML-R-DD-KWD---A..-N--TA-S---EFNIT-MSHV  257
DEB.CM.99.CM40             ATTT.....................KTTTDWSGE.NITMTQYW-----V-.GP.Y-D-KE-SS-VWLED--QWA--KDGSGN.............RTGYMKH--D-----S-ETSR-K-FK-R------YGL-R-DD-N-----L..-N--TA-A--NL-HTMA--WV  265
DEB.CM.99.CM5              .........................KTTTDWSGE.NKTVTQYW--T--VT.GP.Y-D-KEMQS-VWLED--QWETENQTNTN...........KSIRHGYMNH--N-----S-ETSR-K-FK-R------YGL-R-DDRN-----L..----TA-A--NL-HTMA--WM  255
DEB.CM.SIVdeb04CMPF3061    .........................KTTTDWSGE.NNTMSQYW-----V-.GP.Y-D-KEEGH-MWLTD-LEWANGNETGDT.............RGAYMRY--R-------D-SH-K-FK-R------YGLMR-DD-N-----L..-----A-A--NL-HTMA--W-  254
DEN.CD.x.CD1               TTTIKTPTT.....................EWYGGNKTT.IFLK-N--MTTGKGF-D-RVMYE-HCRIE-VME...............DRETNQ-Y-YMRY--Q-T-----E--V---FK-RM-------L----DHPWS-Y-L..-K-M-ARV--DE-HTMA--W-  264
DRL.x.x.FAO                .......VATTVNTSVASTTTANDTLEIAVNNTEP...VMETNRV-KY-VT.GL.C-DCKEEITEN-RYD-VVCKCGNETESPES....NNCTNKKRTCYV----ST----D-N-A-TDMMKFRL---P-YIL-R--E-LNISSK...-K-ITA----RHMPATI-SHF  281
GOR.CM.2004.SIVgorCP684con ....................PTSTETTIP.....T.NIPELD-F--T--VT.-V.LKD-KT-QQ-L---Q--TET-KNE......................-R-IN----T-S---------L--Q------YPL----DEQ-----E..-N---I-H-------T-----  252
GOR.CM.2007.SIVgor2139_287 ....................TIPPHTTIP.....P.HIPQLD-Y-----VT.-V.LKD-TT-QQ-L--RQ---ETGET............GTNGIKG-R-IN----T-S---------L--Q------Y-LM---DIQ-----E..-R---I-H-------T-----  253
GOR.CM.2007.SIVgorCP2135con .........................TTIP.....P.YIPQLD-Y-----VT.-V.LKD-KT-QQ-L---Q---KT-ET............NETKEYSE-R-IN----T-S---------L--Q----X-Y-LM---D-Q-----E..-K---I-H-------T-----  247
GRV.ET.x.GRI_677           T...........TPCQN.........CSTEQIEG.E-AEEPAS--T-A-A.GY.Q-DVKKNYSMTW-DQELVCNNKTGSEKG............SKDCYMIH--D---KE--D-TYWDTLRVR------Y-L----D-DYR-FA-K.-K---V-H--RL-NTTIT-GI  262
GSN.CM.99.CN166            STTT.......................QWGSWGGE.NGTGQPLY-----QT.-E.F-DQKKQM-SL-WRE--MEETHG................NQSG-YIRN----Y---K-V-S--Q-V--------P-Y-M----DVN-T-V-T..-Y---A-T-----Q-L-A-W-  270
GSN.CM.99.CN71             PSSPTPTP...................WGNWGG.NGTGQP-Y-----QT.-E.F-D-KRQM-SL-W-D--MRAQDG................ND-D-YIIN----YV--K-V-T--Q-V-------P---M----DAN-T-V-K..-S---A-T-----Q-L-A-W-  272
LST.CD.88.447              P.MPCFITEKTTTTTKRSTTGGPELIEDINCIKWV.NETTEINRE-RY-VT.GL.C-DCRTEIKQS-RYDETTCDK-G..................TCYM-H--D-I-Q-D-N-GVISNAYFRL-----YML----E-LNFSAN...---ITATP--DYMISS--SFF  276
LST.CD.88.485              P.MPCFIAET.TTTTTKGNTTMESLIEDTNCTRWL.NETTEINRE-KY-VT.GL.C-DCKEEIKQS-RYDETTCDK-G..................TCYM-H--D-I-Q-D-N-GVISNAYFRL-----YML----E-LNFSAN...---ITATP--DYMIST--SFF  275
LST.CD.88.524              P.MPCFITESTTTTTKSSGTEQPSLIEDTNCTRWL.NETTEINRE--Y-VT.GL.C-DCKEEIKQN-RYDETTCNE-G..................TCYM-H--D-I-L-D-N-GVMSNAYFRL-----YML----E-LNFSAN...---ITATP--DYMISS--SFF  278
LST.KE.x.lho7              TQMPCFINEQV...TVKNPGNETRLEEDLNCTRGL.NETTERNAE-QY-VT.GL.C-DCRTEIKQS-RYD-VTCSGER................ENRTCYM-H--D-I---D-N-GVMQNAYFRL-----YML-R--EQLNFSKK...-E-ITATP--GYMLSS--SFF  271
MAC.US.x.251_1A11          AAPKAM.........SEKINMINETSSCIVHDNCT.GL-QEQMIG-K--MT.GL.K-D-KKEYNETW-SA-LVCEQ.............GNSTGNESRCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTNYS-FM-K.-SK-VVSS--RMMETQT--WF  274
MAC.US.x.251_BK28          AAPTSAPV.......SEKIDMVNETSSCIAQNNCT.GL-QEQMIS-K-TMT.GL.K-D-TKEYNETW-ST-LVCEQ.............GNSTDNESRCYMNH------QES-D-HYWDT-RFR----P-Y-L-R--DTNYS-FM-K.-SK-VVSS--RMMETQT--WF  277
MAC.US.x.EMBL_3            AAPTSA.......PVSEKIDMVNETSSCIAQNNCT.GL-QEQMIS-K-TMT.GL.K-D-TKEYNETW-ST-LVCEQ.............GNSTDNESRCYMNH------QES-D-HYWDT-RFR----P-Y-L-R--DTNYS-FM-K.-SK-VVSS--RMMETQT--WF  277
MND-1.GA.x.MNDGB1          TSTVASST..............EIYLDVDKNNT.EEKVERNHV-RY--T.GL.C-DSKEEIVTN-RGD-VKCEN...................NTCYMNH--E--N-ED-Q-GLLIR.C-LG-V-P-YVM-RY-E-LN-NKL...-S-I-A----QHLVAT--SFF  258
MND-2.CM.98.CM16           TTT..............KTVANTTELDID.TNNT.ETTTQQNRV-K--TT.GL.C-DCKLEIEEN-RYE-VTC...............TNGTNNTY-CYM-Q--------D-N-A-TDE-KFRL---P-YVL-R-LERLNVSKK.....-ITA----QPLPATT--MC  261
MND-2.GA.x.M14             PTTP..................AVTGLEVTVQNN.ESTTQ-NRV-K--TT.GL.C-DCKIEIKES-RYE-VTCTK.............KENDTETEECYM-H--------D-N-A-TDKMTFRL---P-YIL-R-RE-LNKTKL...-A---A----DPMPATI--MF  261
MND-2.x.x.5440             TTTTPSTTT.......SSSTNKTTTPVLVVEKQNN.ETTTQQNRV-K--TT.GL.C-DCKLEIEEN-RYE-VTCTKLNKTGSA.......TNSTEPEYECYM----AT----DRN-A-TDRMTFRL---P--VL----E-LNKTKL...-G---A----APLPATI--MF  282
MNE.US.x.MNE027            PTTKTTTT........KEIEVVNENSTCVNRDNCT.GL-QEPMIS-K--MT.GL.K-D-KREYNETW-SA-LVCEQ.............GNSTEDESRCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTKYS-FM-N.-SK-VVSS--RMMETQT--WF  276
MON.CM.99.L1               P............................WGNWGG.NGTGQPVY-----QT.-E.F-D-KKQM-SL-W-E--MKEEGSNG...............SH-YILN----Y-----E-SNY--V-L-----P-Y-L-R-DDPA-T-Q-S..-S---A-T---A-Q-I-A-WF  256
MON.NG.x.NG1               TP...........................WGSWGG.NGTGQPLY-----QT.-E.F-D-XKQM-SL-X-E--MQDK...............DNTTDDY-Y-LN---XYX----E-SNY----------P-Y---R-DDPX-T-QNN..-X---A-T-----M-I---WF  267
MUS-1.CM.01.1085           ............TGGNSTGTTAPPTTPWGSWGNGN.DSAGQPMF-----QT.-E.F-D-KRQM-SL-FVD-LMRDSSESEND.................--YI-N----Y-----E-S--Q-V--Q------YSL----DAN-T-Q-E..-N--TARH-----L-L-A-W-  271
MUS-1.CM.01.CM1239         .............TTTTASTPTKTTPVTPWGKWDA.NGTEQSMF-----QT.-E.F-D-KKQM-SL--TD-LMRE---.................T-YILN----Y---Q-E-S--Q-V--Q------YSL----DAN-T-Q-K..-N--TATH-----L-L-A-W-  266
MUS-2.CM.01.CM1246         ............TTAASATTSGHTTVTPWGRWADN.NETQLTMY-----QT.-E.F-DVKKQM-SL-FVK-LMKGE-E...................T-YI-N----Y---Q-E-S--Q-V--Q------YSL----DAN-T-Q-I..-N--TARH---D-L-LIA-W-  270
MUS-2.CM.01.CM2500         ..................TKPTTPSTTPWGRWSDG.NDTMQVMY-----QT.-E.F-D-KKQM-SL-FVD-LMRDGS-................DG-YI-N----Y---Q-E-S--Q-V---------YSL----DAN-T-Q-K..-R--TATH-----L-L-A-W-  258
OLC.CI.97.97CI12           ............TTVTSSTGGQTTTIGLGVDNSTV.KP-TEKL-Q-RY--T.GL.C-DCKIEMNQT-DLE-VACTN-...................TCYMSH--------D-QYAVSDQQMVKL---P-RVLI----R-E-P......-R-I-A-T--PPLSAT--SFF  261
RCM.CM.SIVagi-00CM312      TPSATAS.TN..IDCGIN......NETCS.......AV-DENVMQ-T-AVA.GL.K-DEKH-YNDTW-SR-LWCQKAN................NV-QCFMRH---TS-H-F-EPRYW--FRLR----P---L----DYNYT-FDT..-H--TATS--DM-NTT---SF  262
RCM.CM.SIVrcm-02CM8081     TTPTAAK.TTTKVDCPMN......NETCA.......AVPDEDVM--E-AVA.GL.K-DEKY-ANDTW-SR-LWCEKDKN............NNKTTKRQCFMRH---TS-Q-F-EPKYW--FR-R----P---L-V-KD-NYT-FDT..-H--TATS--DM-NTT---SF  271
RCM.GA.x.GAB1              KTTTQRT.IGVEKECTAG......NETCE.......EVQDADVMS-E-AVA.GL.K-DEKH-YNDTW-SR-LWCEKET............NSTNSTKKKCFVRH---TS-Q-F-EPKYW--FRLR----P---L-V-KD-NYT-FDT..-V--TATS---M-NTT-ASGF  273
RCM.NG.x.NG411             RATTRTT.KEIGKDCYAN......QTECS.......SVPDADVM--E-TVA.GL.K-DEKM-VNDTW-SR-LWCKAGN................NNTRQCFIRH---TS-Q-F-EPKYW--FRLR----P-Y-----LDANYT-FDT..-Q--TATS---M-NTT---SF  267
SAB.SN.x.SAB1C             PEVVS...................VGFNDSVIEQ.E---EQAM----AMA.GY.R-DVKKNYSTVWDDQEVVCEEGREKSNA...........THTVGCYMIH------KE--D-TYWDTFRLR------Y-L-R-ADTDYS-HKA..-R--TVSA--RL-NTT---GI  269
SMM.SL.92.SL92B            PTTTSPLTAASPSGE....EIVNDTMSCTKNNNCS.GI-QEPMIG-Q--MT.GL.K-DQKRQYNETW-SR-LVCEQ...............GGNESSRCYMNH------QES-D-HYWDA-RYR----P-Y-L-R--DTNYS-FA-N.-SK-VVSS--RMMETQT--WF  278
SMM.US.x.H9                QTSTTPPSPIIA.......KVVNDSDPCIRSNNCT.GL-QEPMVS-K--MT.GL.K-D-KREYNETW-SR-LVCEQ-..............SNGNETDSKCYMNH------QES-D-HYWDA-RXR----P-Y-L-R--DTNYS-FA-N.--K-VVSS--RMMETQT--WF  278
SMM.US.x.PGM53             TTTTTKSPKA....EAITAKVINESDPCISNNNCT.GL-QEPMVS-R--MT.GL.KKD-KKEYNETW-SR--VCEQ.............STTSRNESKCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTNYS-FA-N.-SK-VVSS--RMMETQT--WF  280
SMM.US.x.SIVsmH635F_L3     AKSTTSTTTTT.....VTPKVINEGDPCIKNNSCA.GL-QEPMIG-K--MT.GL.K-D-KIEYNETW-SR-L-CEQ.............PANGSESKCYMQH------QES-D-HYWDA-RFR----P-Y-L-R--DSNYS-FA-K.-SK-VVSS--RMMETQT--WF  278
STM.US.x.STM               PTAAAAAT.........KPELVNETSSCVSNNNCT.GL-EESLVG-K--MT.GL.K-D-KREYNETW-SS-L-CEQ.............NVTGEESRCYMRH------QES-D-HYWDALRFR----P-Y-L-R--DTNYT-FA-N.-SK-VVSS--RMMETQT--WF  274
SUN.GA.98.L14              WTTTTAKTTTQTTSSTSSTASTTTPMPLDWNCTDT.ENIAESN-V-KY-VT.GL.C-DCKTEVEQN-RDTEVTCNG--..................TCYM-H--D-I---D-H-GIMQNAYFRL-----YML-R-DE-LNATKK...-K-ITATP--NYMTST--SFF  275
SYK.KE.x.KE51              .................SPTTTTNKAATPPAWWG.DNSTETRY-----MT.GG.FKD-KHQYN-----S--MREE...............GNESY-Y-QH-----MRN--E-QT-R-F--Q----P-YSL----DTE-E-DAE..-Q--TA-S---PFNTLA--WF  260
SYK.KE.x.SYK173            .................TPTTSPPTTPPNETWWG.DNSTEPRF-----LT.GG.FKD-K-QYR-----D-LMKEE...............GNSSY-Y-LH------SA--E-QT-Q-F--Q----P-YSL----DTN-E-DDV.-----TA-S--QEFNTLA--WF  253
TAL.CM.00.266              ........................AMSNSNMPWWG.DNTTEPVY--TW-MT.GA.F-D-RYTYFSHW-LA-LMKEPGNGS...................F-YA-G--V---A---E-THYQAF--Q---------LI---DSPWT-R-K..-H-I-A-H--DP-TTIAA-W-  252
TAL.CM.01.8023             TNSVTS..................KRTNDTNVWWG.NNATEPVY--TW-MT.GG.F-D-KYTYFSHW-LA-LMKEPGNGN................DSY-YA-G--L---A---E-THYQAF--Q---------LI---DNPWT-R-K..-H-I-A-H--DP-TTIAA-W-  268
TAN.UG.x.TAN1              ............DPCPNT.....DESSCNATLVTN.S-DYENSSI---AMA.GY.R-DVKK-YNSTW-DQELVCEKENN.TTG.............TRGCYMIH--D---KE--E-TYWDTLRLR------------KDTNYT-F-V..-R---V-S--GLMNTT--SAF  270
VER.DE.x.AGM3              TNITASTT...TLPCVQNK.TSTVLESCNETIIEK.ELNEEPAS--T-AMA.GY.V-DQKK-YSVVWNDAE-MC...............KKGNNSNRECYMIH--D---KE--D-TYWDELRLR---------L----DYDYA-FKTN.-S---V-H--NL-NTT-T-G-  279
VER.KE.x.9063              APKTG.......DPCIKSTNNNVNLQPCNASLIEE.EL-EEAAS--T-AMA.GY.--DQKKNYSVVWNDAE-YCKNSSSSNSS...........TKECYMIH--D---KE--E-TYWDELRLR---------L---HDHDY--YKQN.-S---V-H--GLMNTT---G-  279
VER.KE.x.AGM155            K...........LPCVRNKTD.SNLQSCNDTIIEK.E-NDEAAS--T-AMA.GY.--DQKKNYSVVWNDAE-FCKRSTSHNG...........TKECYMIH--D---KE--D-TYWDELRLR------Y-L----DWDYA-FK-E.-S---V-H--TLMNTT-T-G-  274
VER.KE.x.TYO1_patent       TG..........LPCVGPTSG.ENLQSCNASIIER.E--DEPAS--T-AMA.GY.V-DQKKNYSVVWNDAE-YCKNKTNST...........SKECYMIH--D---KE--D-TYWDQLRLR------Y-L----DEDY--YKQN.-S---V-H--GLMNTT-T-G-  272
WRC.CI.97.97CI14           TTRA......TTSATTTVGTVSTTPVMQVENNCSQ.VNTSEINHV-R--TT.GM.CKDCRIEIKES-RGD-VVCNN-S...................ECYINH--E---K-D-Q-GWQDEQY-RM---P-SVL----E-LNSSKL...-E---A-M--DMMTATI-SFF  264
WRC.CI.98.98CI04           TGAPTISTISAAPTTSTISAGTRTPAMPVDNNCTT.VNTSEINHV-R--TT.GM.CKDCRIEIKES-RGD-VVCNK-G-...................ECYVNH--E---K-D-Q-GWQDEQYVRM---P-SVL----EPLNSSKL...-E---A-L--DMMTATI-SFF  275
WRC.GM.05.Pbt_05GM_X02     ...........TTTPVTPPQNTTPSMDVENNSTS.LNLTDTNHV-R--VT.GL.C-DCK-EIVES-RGD-VRCTQH...............GNGTQTCYINH--E---H-D-Q-AWHDEQY-RM---P-YVL-R--E-LNSSKL...-E---A-T--DLMTATI-SFF  267
```

```
                                                          glycosylation NNT
                                          CD4 bind     glycosylation NCT        V3 tip
                                     glycosylation NFT                              V3                                                                  CD4 bind
                    glycosylation NGS              glycosylation NTS                                   glycosylation NNT   glycosylation NKT          CD4 bind
                                                           *                                     *
H1B.FR.83.HXB2               LLNGSLAEEEVVIR........SVNFTDNAKTIIVQLNTSVEINCTRPNNN.TRKRIRIQRGPGRAFV....TIGK.IGNMRQAHCNISRAKWNNTLKQIASKLREQFGNNKTII..........................FKQSSGGDPEI  371
H1A1.UG.85.U455              --------R-IR--.........-E---N---------VNP-K---S--Y-TRKNI-RYS.I-S-Q--Y....VT--I--DI------V--RD--R-IQ-V-EQ-KKK-N-KT..........................II-AS-----I--  369
H1B.US.90.WEAU160            ---------DI---.........-E-------N------V-I-------------.---K-TL.----VLY....-T-EI--DI-R----L--TS--------VE----IKQFKNKT-......................V----------  375
H1C.ET.86.ETH2220            -----I--G-TI--.........FE-L-N---I------E----T----S--.--ES----.---QT-Y....AT-DI--DI--------EE---K--QKVKE--QKH-P-.--..........................IE--P-----L--  366
H1D.CD.84.84ZR085            ----------I---.........-E-L-N---I---H--Q---------YKK.E-Q-TP-..-Q-Q-LY....-TRYTTRIIG--Y----GV------R-V-R--GNLLNQT-..........................II--P--------  376
H1F1.BE.93.VI850             ---------GI---.........-Q-ISN-------H--E--Q---------.----G-HL.---QT-Y....AT-AI--DI-K------GTQ-----EYVKAE-KSH-P--.-A..........................IK-N------L--  361
H1G.SE.93.SE6165             --------GKIKV-.........-E-----T-V------KT-----------TMKRIRMG.I---QT-Y....AT-AI--DI------VTKR--KEA-QNV-AE-GKI-NKSSEN..........................IT-NS-A---L--  378
H1H.CF.90.056                ---------QII--.........TK-IS--T-N-----K-P-N-T-------.--TS-HL.------Y....AT-DI--DI---------TD--K--H-VVTQ-GIHLN-RT..........................IS--PN----M-V  362
H1J.SE.93.SE7887             -----I--GDII--.........-E-IS---N------KT---V-Y-----.---G-HM.---QVLY....AT-EI---I-ET-----ERD-S---RRV-T----H-NKT..........................IN-TSP----I--  362
H1K.CM.96.MP535              ----------II--.........-E-I---T-N------ET-Q---------.---S-HM.---K--Y....-T-DI--DI--------GE---M--SRVKE--K-H-K-GTI..........................TFKPPNP------  372
H101_AE.TH.90.CM240          ----------II--.........-EDL-N-------H--K--------S--.--TS-T-.-----V-Y....RT-DI---I-K-Y-E-NGT---KV---VTE--K-H-NKT..........................II-QPP----L--  371
H102_AG.NG.x.IBNG            --------G-----.........-E-I-N---------ANP-K---------.---GVH-.---Q--Y....AT-DI--DI------V-KTE--K--H-VVTQ-KTY-K-TT..........................II-ANPL---V--  366
H103_AB.RU.97.KAL153_2       --------------.........--------T-------KEP-----------.---G-H-.------Y....AT-DIT-DI--------IT---------VI---K----KT..........................IV-N---------  360
H104_cpx.CY.94.CY032         -------T------.........-K-I---T-N-----AKA-K------G--.---SVH-.---LTWY....AT-EI--DI--------GND--D---V-SEE-KRL-P-KTIK.........................-APPV---L--  374
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-V  370
H1P.FR.06.RBF168             I---T---X-PLVI.........TQ-VS-TRYV---K--XN-SLT-V--G--.--XQVQ-.--MTWY....NMKFYTXDIXK-Y--X-MQX-TK--ESVSKAIWKAYPQPPNQNH......................TFV-RN-T-----V  372
CPZ.CD.90.ANT                ----TYQTNTS-VM.........NGRKNESVLVRFGKEFENLT-T-I--G-RTV-.NLQ-.---MT-Y....NVEIAT-DT-K-F-TVNKTL-EQARNKTEHV-A-HWKKVDNKTN......................AKTIWTFQD----V  374
CPZ.CM.-.SIVcpzDP943         I-------NSTXX-.........TEXMXXMDX-----FHKR-QMT-I--G--.S-GQ-Q-.--AMS-Y....NLENI--DT---Y--X-A-E-EQRXXDTXRAIKSLKPGSNITF.........................TDQQ------  356
CPZ.CM.-.SIVcpzMB897         --------GNKTVV.........R-RSKS-TE------A-AIY-----LG-KTIEG.-P-.---QI-Y....RTKTVV-DT-G-E-R-NGTA--E--R-VKEA-NNTYR-LNLSL......................TEIN-EGA----L-V  364
CPZ.CM.01.SIVcpzCAM13        II-----T-NITV-.........VN-ASK-THDW----S-A-NLT-K-VG--.-GKVQ-.---MT-Y....NMDHIF-DT-K-F-ELNGTT--E--QKVRES-IKEIKA-ANGTY......................NIT-EP--------  384
CPZ.CM.05.SIVcpzEK505        I------TSNI---.........N-SK-...-LL----E-IP------G-K.--GQVQ-.---MT-Y....N-ENI--DT---Y-EVN-T.-EQIWNTTKQIIINN.......RK......................NIT-IPNP---L-V  355
CPZ.CM.05.SIVcpzLB7          I--------KIM--..........YKA-...KVL------IH-----VG-KTIKG.-PL.---QL-Y....GTETVV-DT-E---E-NQTH-YKI-N-VKRE-TTV-NESNK.........................TVH-AN-------V  360
CPZ.CM.05.SIVcpzMB66         ----T-E-KVT-LD.........R-VSNDMD----K--ET-RL----TG--TIKG.-P-.--SQI-Y....G-ETV--DT---F-QLNKTV-T--F-KVRHA-N-TYKG.YLGNE......................TIT-GP-T---L-V  393
CPZ.CM.05.SIVcpzMT145        I----T-DNQT-A-...........IDPSENLA-I--KDP-K-T-R--G--.--GQ-Q-.--AMT-Y....N-ENVV-DT-K-Y-E-NGTQ-AKA-NETKEV--NILRK.........................NIS-MVP------V  360
CPZ.CM.98.CAM3               I----V--NKTIA-.............R-GYXFLXXFQXX-S------G-X.S-GQ-Q-.---MT-Y....N-ENIV-DT-R-Y-Q-N-TV-DER-NETGQA---L-T-LTQV.........................N-TV-P-----V  371
CPZ.CM.98.CAM5               I--------GPAVAR.........SEDLRH-I--I-FKEGIN-I-M--G--.S-GQLQ-.---VS-Y....N-ENIV-DP------V-K-Q-MKK-NET-LAIAKYDKSIAKV.........................T-TPNP------  367
CPZ.GA.88.GAB1               -I------GNITV-.........VE-KSK-TDVW----VEA-SL--H--G--.--GEVQ-.---MT-Y....N-ENVV-DT-S-Y-K-NGTT--R-VEEVKKA-ATSSNRTAAN-........................TLNRA------V  369
CPZ.GA.88.GAB2               I---------E-VF....RTKNMTAPGLSD-V----KRAIP---S--G--.-GRA-NL..S--TT-F....NTEAL---P-K-S-HLNGTL---I-NR-KQ-IKNSTTWHRGD-........................T-TKHP-----V  372
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----V  369
CPZ.TZ.2001.TAN2             H---T-E-N-TKAY.........FADVNVNPPLL-KF-E--SLT-E-TG--.--GQVQ-.---MT-Y....N-ENVV-DT-K-Y-HVNATL-DQS-SRAMEAINKTLALYNKAV......................NEKFEW-K----V  377
CPZ.TZ.2002.TAN3_1           H---T-E-K-TKAY.........FVNAT-NPPLL-KF-E--PLT-G-TG--.V-GQVQ-.---MT-Y....NTENVV-DT-K-Y-WVNATL-YKS-DNAMEAITKTLKKSGLNV......................TRR-DWHK---L-V  389
CPZ.US.85.CPZUS              II-----TKN-TV-..........SK.NF-DI-L--FSEG-NMT-I--G--.-VGNVQL.---MT-Y....N-P-IV-DV-E------KLT-EKQR-YTLEIIKKEANLT-V.........................ELIPNA-----V  351
                                                                                     V3
                                                           *                                     *
MAC.US.x.239                 GF--TR--NRTY-Y.........WHGR--RTI-SLNKYYNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINDRPK--W-WF.GG--KDAI-EVKQTIVKHPRYTG-NNT.....................DKINLTAPGG.----V  388
H2A.DE.x.BEN                 GF--TR--NRTY-Y.........WHGR--RTI-SLNKYYNLTMR-K--G-K.-VLP-TL..MS-LV-.....HSQPINTRP---W-RF.GGR-REAMQEVKQT-VQHPRYKGINDT.....................GKINFT-PGA-S---V  383
H2A.GW.x.ALI                 GF--TR--NRTY-Y.........WHGR--RTI-SLNKHYNLTMH-K--G-K.-VVP-TL..MS-LI-.....HSQPINKRP---W-WF.KGE-RKAMQEVKET-VKHPRYKG-NDT.....................NQINFT-PGR-S-A-V  380
H2A.SN.x.ST                  GF--TR--NRTY-Y.........WHGR--RTI-SLNKFYNLTVH-K--G-K.-VVP-TL..MS-LV-.....HSQPINRRP---W-WF.KGE-KEAM-EVKLT-AKHPRYKG-NDT.....................EKIRFIAPGERS---V  377
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---V  381
H2B.GH.86.D205               GF--TR--NRTY-Y.........WHEK--RTI-SLNTYYNLS-H-K--G-K.-VVP--T..VS-LL-.....HSQPINKRP---W-WF.KGN-TEAI-EVKRTIIKHPRYKGGAKNIT..................SVKLVSEHGK-S---T  385
H2G.CI.92.ABT96              GF--TR--NRTY-Y.........WHGR--RTI-SLNKYYNLT-X-R--G-K.-VLP-T-..MS-LV-.....HSQPINERP---W-WF.GG--KEAMQEVKETVVKHPSYKG-NDT.....................KXITFTTPGE-S---V  389
H2U.FR.96.12034              GF--TRS-NRTY-Y.........WHGR--RTI-SLNKYYNLTMY-R--G-K.-VLPVT-..MS-LV-.....HSQAINN-PK--W-WF.GGN-KGAIQEVKETIANHPRYHG-KNI.....................SQINLAEPAG.--S-V  384
COL.CM.x.CGU1                -V-A-KELSDWAKD......REG-WKN-SGTIEYYWFPKDIALG-I-RG-S.SHRNLNT..AN-AK-Y....YELIPYSKGIYGR-QFVPMTGQ-KKNKTQFAIEIKKNLTAWLERISRKNI...............TITPRNGNRTS---A  381
DEB.CM.99.CM40               QF---DE-RAEELHI..IRKEVKGEVQNGSI--R-PAKYNLTLT-V--G-K.-YRA-HM..AT-LS-Y....-TFIQRLRIKR---RL.NGS-A-AT-EMRQ-IL-I--KANRTNN.......................LTIHYPK--R-V  383
DEB.CM.99.CM5                QF---DE-RAED-HI....IRK..GLENKTI-VRIPPKYNLSLD-V--G-K.-YRAVHM..AT-LS-Y....-TFIPRLRIKR---RM.-GN-TGAV-E-R-R-K-IY...NVTK.......................ITVHYP---R-V  365
DEB.CM.SIVdeb04CMPF3061      QF---DE-RAEE-HV....IRK..GLENKTI--RIPPEYNLRLD-V--G-K.-YRATHM..AT-LS-Y....-TFVERLRIKR---RM.-GN-SGAVSD-KQV-KGIY...NVTN.......................ITMHYP---R-V  364
DEN.CD.x.CD1                 IM-STVELLDNHTQI....IRN.TSKDNQTIA-LFRKSSVIKLK----G-V.-L-GLNM..AA-IVMR....ATIIPRRAI-R-A-K-VGGN-TR-RQE-EKAVK-E-NI-G-VKW.......................DVPRV-----T  379
DRL.x.x.FAO                  GF--TKHGQDEL-ET.....RGKPD-LNHKYVFR-NKKWNLK-I-R-KG-R.SIISTPS..AS-LL-Y....HGLEPGK-LKRGM-KF.EGQ-GKA-GSL-KE-KSVS-SIWRNTTNSTCISLKGKKI........QDRNVTGCM.......LKQLKVDNYTTK-.-LAS  422
GOR.CM.2004.SIVgorCP684con   I---T--KGNNATI.......I-S-SS-S-TD---K-ANP---T-E-TG--.-.RGQIQV.--LTIY....NSENI---T-K-F-KYNXTA-QKA-QATVRA-KTNLTHYFN............................K-------V  359
GOR.CM.2007.SIVgor2139_287   I---T---GNAT-I.........-K-SS-SG-D---K-ARP---T-E-TG--.-.RGQIQI.--MTIY....NSENIV--T-K-F-KYNETN-QGA--DTVQA-N-SLIHFR............................N-------V  358
GOR.CM.2007.SIVgorCP2135con  I---T--K-HAT-I.........-K-S--SG-D---K-AR----T-E-TS--.-.RGQIQV.--MTIY....NSENIV--T-K-F-KYNKTN-Q-A--DTVRA-KANQIRFQ............................N-------V  352
GRV.ET.x.GRI_677             G----RS-NRTE-W.......QKGGNDNDTVI-KLNKFYNLTVR-R--G-K.-VLPVT-..MA-LV-.....HSQ-YNTRLK--W-HF.QGD-KGAW-EVREEVKKVKNLTEVS-E.......................NIHLRRIW..----S  375
GSN.CM.99.CN166              H---TYQPGNNTRIM.......MNGMKNESIV-GFGEDYHLNLT-I--G-K.-IRNLQ-..-A-MT-Y....SQVIVG--T-K-Y-RLDPDR--RAIREAMKAMN-HWE-KTG.....................RNDTQIRWTSEPK--L-V  387
GSN.CM.99.CN71               H---TYQPGNNTRVM.......MNGKKNESIV-GFGEDYQLTLT-I--G-K.-L-NLQ-..-A-MT-Y....SQIIVG-DT-K-Y-KLNHT--DIAIR-AMKAMKNHWEKINNDTT....................PNKTQIRWTSEPK--L-V  392
LST.CD.88.447                GF--TNHTDDEL-PL...TPKKMG-LLEAKFVYK-AGKWGLI-R-I-KG-R.SEVSTIS..ST-LL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFHNLGNM-KRLNATAMNYTH..GNCIDNT...............KPCGRQ.....LKGLPIANMTRR--NLAT  412
LST.CD.88.485                GF--TNHT-DKL-PL...TPKKMG-LL-AKFVYK-SGKWGLI-R-I-KG-R.SEVSTIS..ST-LL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFNNLGKM-QKLNATAMNYTN..GQCPEGK...............QPCGRH.....LKGLPIANMTRR-EGLAT  411
LST.CD.88.524                GF--TNHT-DEL-PL...TPKKMGDLNGAKFVYK-AGKWGLI-R-I-KG-R.SEVSTIS..ST-LL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFNNLGRM-KKLNETAMNYKE..GNCSTPG...............KPCGRQ.....LKGLPIANMTRR-V-LAT  414
LST.KE.x.lho7                GF--TNHTRDEL-PL...TPNKMEDLNGAKFVYK-AGKWGLI-R-I-KG-R.SEVSTIS..ST-YL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFNNLGKM-K-LNAEAMNYTEGTGTCDSKK..............TTCGRK.....LKGLPIANMTRH-A-LAT  409
MAC.US.x.251_1A11            GF--TR--NRTY-Y.........WHGR--RTI-SLNKYYNLT-K-R--G-K.-VLPVT-..MSQLV-.....HSQPINDRPK--W-WF.-G--KDAI-EVKQTIVKHPRYTG-NNT.....................DKINLTAPGG.----V  387
MAC.US.x.251_BK28            GF--TR--NRTY-Y.........WHGR--RTI-SLNKYYNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINDRPK--W-WF.GG--KDAI-EVKQTIVKHPRYTG-NNT.....................DKINLTAPGG.----V  390
MAC.US.x.EMBL_3              GF--TR--NRTY-Y.........WHGR--RTI-SLNKYYNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPLTDRPK--W-WF.GG--KDAI-EVKQTIVKHPRYTG-NNT.....................DKINLTAPGG.----V  390
MND-1.GA.x.MNDGB1            GF--TMHK.-GELIPIDDKYRGPEE-HQRKFVYK-PGKYGLK-E-H-KG-R.SVVSTPS..AT-LL-Y....HGLEPGK-LKKGM-TF.KGR-GLA-WSL-KE-NKLNDSI-VNQT....CKNFTSTGEENKQNTDKQKEFAKCIKT........LKIDNYTT-.--RAA  406
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.KGYC..KGLKRK.........ENRTGCA.......LKTIKVSDYTTK-.E-GA  401
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.........GVVNKTGCA.......LKTIEVSNYTTI-.E-GA  401
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...........KNTTGCH.......LKTISISESTVK-.E-GA  422
MNE.US.x.MNE027              GF--TR--NRTY-Y.........WHSK--RTI-SLNKYNNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-RF.EGN-KEAI-EVKQTIVKHPRYTG-NNT.....................DKINLTAPRG.----V  389
MON.CM.99.L1                 Q--STGNAPNTTVM........MNKQKNESIVVRLAKHLH-N-T-I--G-K.-IRNLQ-..-A-MT-Y....SQLIVG--T-K-Y-KVNKTQ-ETA-QAVHEAVKTEWEKKNNGTN.....................VTTISWRFQPQ--K-V  373
MON.NG.x.NG1                 Q--ATS-TNXTTVMM........NGKRNESIV-RLXKTL--N-T-I--G-K.-LRNLQ-..-A-MT-Y....SQIIVG--X-K-Y-QVNKTQ-DAA-HAAHXAIXXEWESKNNNVT.....................VANITWRYQPK--N-V  384
MUS-1.CM.01.1085             Q---TYL-GNNTAVM.......MNGDKNESIG-KFGE-FR-NLT-I--G-K.-IRNLQ-..-A-MT-Y....SQLIVD-DT-R-Y-K-NKTQ-DPA-R-AMLAM-KHWHGRLQKQNHSSINET..............KIKIRWRSEPK--Q-V  396
MUS-1.CM.01.CM1239           Q---TYL-GNNTAVM.......MNGDKNESIG-KFGE-FR-NLT-L--G-K.-IRNLQ-..-A-MT-Y....SQLIVG-DT-R-Y-K-NKTQ-DVA-R-AMLAM-KHWHGRLEKQGE.NITEE..............KIKIRWTSQPK--N-V  390
MUS-2.CM.01.CM1246           Q---TYLPGNDTAVM.......MNGDKNESIA-KFGE-FR-NLT-I--G-K.-IRNLQ-..-A-MT-Y....SQLIVD-DT-R-Y-K-NKTQ-DPA-R-AMIAM-DHWH-RLQKANK.TINKT..............EIKIRWTSEPK--R-V  394
MUS-2.CM.01.CM2500           Q---TYLPGNDTAVM.......MNGKKNESIA-QFGE-FR-NLT-L--G-K.-IRNLQ-..-A-MS-Y....SQLIVD-DT-R-Y-K-NKTQ-DPA-R-AMKAM-AHWA.KMD.....NRSES..............DIKIRWTSEPK--R-V  377
OLC.CI.97.97CI12             GF--TRHKASRL-PI.TGDLNHKPREGTGVYVMAIADKYKL--E-I-KG-R.SIVSVNS..AS-LVYY....AGLEPY--IDKGL-RF.KGQ-GWM-YALGKG-EKENNTWANYTGVNITGIGR...........NSTTFAKAIKR........FKFDE-LKR---DAA  402
RCM.CM.SIVagi-00CM312        G----IS-NRTW-Y.........QRKQS-RTI-GLNSFYNLTVS-R--S-R.-I-G-SL..AT-VF......ISL-VEKRPKG-W-RF.YGN-KGAW-EVKE-VIQTK-YKG-NDT.....................TKINIRSVY---D-T  374
RCM.CM.SIVrcm-02CM8081       G----IS-NRTW-Y.........QRKQS-RTV-GLNREYNLTVE-R--S-R.-V-G-SL..AT-VF......ISLRVEKRPKG-W-RF.YGN-TGAWEEVKKEVIKTK-YKG-NDT.....................KKIYIRSHY---D-A  383
RCM.GA.x.GAB1                G----INVN-TW-Y.........QRRQS-RTV-GLNSFYNLSVT-R--S-R.-V-G-SL..AT-VF......ISLRVEKRPKG-W-RF.EGN-TDAW-EVKERVKTTK-YRG-SNT.....................DKIKIRTVY---D-A  385
RCM.NG.x.NG411               G----IS-NRTW-Y.........QRKQS-RTV-GLNRHFNLTVT-R--S-R.-V-G-SL..AT-VF......ISLRVQKRPKG-W-RF.DGD-KGAW-EVKEAINKTK-YKG-RQP.....................ENITIRSVY---D-A  379
SAB.SN.x.SAB1C               GI---YVANRTE-W.......QKNGNSNDSVI-RLNRYFNLT-R-R--G-K.-VLPVT-..MA-LV-.....HSQ-YNTRLK--W-WF.GGN-RGAW-EVKETIVRLPPKKYSGTN.....................DTNKI-L-RQW----S  385
SMM.SL.92.SL92B              GF--TR--NRTY-Y.........WHGGS-RTI-SLNKYYNLT-K-R--G-K.-VLPVT-..MS-LV-.....HSQPINERP---W-WF.GG--REAMQEVKKTIVKHPRYTG-NDT.....................RKINLTAPGG.----V  391
SMM.US.x.H9                  GF--TR--NRTY-Y.........WHXRS-RTI-SLNKYXNLTMR-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGE-KEAIREVKXX-VKHPRYKG-NXX.....................XQIKLTAPGG.----V  391
SMM.US.x.PGM53               GF--TR--NRTY-Y.........WHSRS-RTI-SLNKYYNLT-R-R--G-K.-VLPVT-..MS-LI-.....HSQPINERPK--W-WF.GGS-REAIQEVKETIVKHPRYTG-NET.....................KKINLTAPGG.----V  393
SMM.US.x.SIVsmH635F_L3       GF--TR--NRTY-Y.........WHGNS-RTI-SLNKYYNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGN-SEAIQEVKET-VKHPRYTG-NDT.....................RKINLTAPAG.----V  391
STM.US.x.STM                 GF--TR--NRTY-Y.........WHGR--RTI-SLNKYYNLTMS-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGE-RGAI-EVKET-VKHPRYTG-NDT.....................AKIRIVAPGG.----V  387
SUN.GA.98.L14                GF--TRHK-DEL-PI....NNKVGQGAEGEYVWK-AAKWGLV-Q-I-KG-R.SQVSTIS..ST-LL-Y....YGLEPGSKL-L-Q-KF.VGQ-GRAFATLGKL--QVEPLANMSNG....CTFDNITKT.............CRFTNGTDFKKMIKFKPMEKH-A-AAT  415
SYK.KE.x.KE51                Q---TYKAKDK-RFI.......RQKDKNESVI-L-PEKLGLQLV-E--G-E.SI-N-QL..AA-FFLPVI.QGRL-TGKAAKR-F-KV.KGE-GKFFE-VHNESIKVWK-.........................VTSTSWRSQPQ--L-V  374
SYK.KE.x.SYK173              Q---TYKAKDK-RFI.......KQKDKNESVI-L-PEALRLQ-I-E--G-E.SI-N-QL..AA-YFLPVI.QGKL-TGRDAKR-F-RV.TGN-TEFF--VHEQATKTWK-.........................VTNTTWRSQP---L-V  367
TAL.CM.00.266                I---TK..-NHTQVI.......HQAPKNESAI-LLGKEHNLT-T-I--G-R.-I-DLQ-..AA-LM-H....SQIIAGK-LNR-Y-KL.EGN-TKAMESANEEIAKRYKEHINNTD.....................KVNITWTAVRR--L-V  367
TAL.CM.01.8023               V---TK..-NQTQII.......HQAPKNESAI-LLGKEHNIT-Q-I--G-R.-I-DLQ-..AA-LM-H....SQIIAGKDLKR-Y-K-.EGN-TRAMEDVNE-IAKAYKDHINSSS....................GTAQIRWQGVRK--L-V  384
TAN.UG.x.TAN1                GI---Q--NRTE-W.......QKHGVSN-SVI-KLNKHYKLK-V-R--G-K.-VLPVT-..MA-LV-.....HSQQYNTKL---W-HF.QGD-KGAWREVRKTIV-LPKEKYRGTN.....................NTRQIWLSRQW----A  386
VER.DE.x.AGM3                -----YS-NRTQ-W.......QKHRVSNDSVLVLFNKHYNLTVT-K--G-K.-VLPVT-..MA-LV-.....HSQRYNTRL---W-HF.QGN-RGAW-EVKNEIVKLPKDRYQGTN.....................DTEEIYL-RLF----A  395
VER.KE.x.9063                -----YS-NRTQ-W.......QKHRVNNDSVIVLFNKYYNLTVT-K--G-K.-VLPVT-..MA-LV-.....HSQ-YNTKL---W-HF.QGR-REAW-EVKEEIIKLPKDRYRGTN.....................DTEKIYL-RQF----A  395
VER.KE.x.AGM155              -----YS-NRTQ-W.......QKHGVSNDSVL-LLNKHYNLTVT-K--G-K.-VLPVT-..MA-LV-.....HSQ-YNTRL---W-HF.QGN-KGAW-EVQEEIVKLPKERYQGTN.....................DTNKI-L-RQF----A  390
VER.KE.x.TYO1_patent         -----YH-NRTQ-W.......QKHRVNN-TVL-LFNKHYNLSVT-R--G-K.-VLPVT-..MA-LV-.....HSQ-YNMKL---W-HF.EGN-RGAWREVKQ-IV-LPKDRYKGTN.....................NTEHIYL-RQW----A  388
WRC.CI.97.97CI14             GF--TRHK-GEL-EV.IGNINHKPKEGSG-FVYK-PDKYN-K-E-H-KG-R.SLVSVNS..AS-LIYY....AGLEPHR-I-KGL-KF.TGL-GKM-YDLGKN-EKYNSSWVNYTE..GDCNSKN...........NSADFAKCIKR........FQIRNFTRQ-A-RAT  404
WRC.CI.98.98CI04             GF--TRYKKGELVEI.IGNINHRPHEGTG-FVYK-PSRYK---E-N-KG-R.SVVSVNS..AS-LIYY....AGLEPHR-I-KGL-KF.KGL-GHM-YDLGKN-Q-YNNTWVNYTE...DCNQKN...........DSREFAKCIKQ........FQIRNFTRK-A-RAT  414
WRC.GM.05.Pbt_05GM_X02       GF--TKYDKGEL-EV.IGNINHKPKDGTG-FVYK-PAKYN-A-E-H-KG-R.SLVSVSS..AS-LIYY....AGLEPYK-I-K-L-KF.TGQ-GHM-YGLGKE-QAYNSSWVNYTE...ECNKKG...........NSSAFATCIKQ........FQIRNFTKQ-A-RAT  406
```


glycosylation NST
glycosylation NST
glycosylation NST
V4
glycosylation NNT
CD4 bind
CD4 bind
CD4 bind
glycosylation NIT
CD4 bind
glycosylation NES
V5
CD4 bind
H1B.FR.83.HXB2
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.x.IBNG
H103_AB.RU.97.KAL153_2
H104_cpx.CY.94.CY032
H1N.CM.95.YBF30
H1P.FR.06.RBF168
CPZ.CD.90.ANT
CPZ.CM.-.SIVcpzDP943
CPZ.CM.-.SIVcpzMB897
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMT145
CPZ.CM.98.CAM3
CPZ.CM.98.CAM5
CPZ.GA.88.GAB1
CPZ.GA.88.GAB2
CPZ.TZ.00.TAN1
CPZ.TZ.2001.TAN2
CPZ.TZ.2002.TAN3_1
CPZ.US.85.CPZUS
MAC.US.x.239
H2A.DE.x.BEN
H2A.GW.x.ALI
H2A.SN.x.ST
H2B.CI.x.EHO
H2B.GH.86.D205
H2G.CI.92.ABT96
H2U.FR.96.12034
COL.CM.x.CGU1
DEB.CM.99.CM40
DEB.CM.99.CM5
DEB.CM.SIVdeb04CMPF3061
DEN.CD.x.CD1
DRL.x.x.FAO
GOR.CM.2004.SIVgorCP684con
GOR.CM.2007.SIVgor2139_287
GOR.CM.2007.SIVgorCP2135con
GRV.ET.x.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.x.lho7
MAC.US.x.251_1A11
MAC.US.x.251_BK28
MAC.US.x.EMBL_3
MND-1.GA.x.MNDGB1
MND-2.CM.98.CM16
MND-2.GA.x.M14
MND-2.x.x.5440
MNE.US.x.MNE027
MON.CM.99.L1
MON.NG.x.NG1
MUS-1.CM.01.1085
MUS-1.CM.01.CM1239
MUS-2.CM.01.CM1246
MUS-2.CM.01.CM2500
OLC.CI.97.97CI12
RCM.CM.SIVagi-00CM312
RCM.CM.SIVrcm-02CM8081
RCM.GA.GAB1
RCM.NG.x.NG411
SAB.SN.x.SAB1C
SMM.SL.92.SL92B
SMM.US.x.H9
SMM.US.x.PGM53
SMM.US.x.SIVsmH635F_L3
STM.US.x.STM
SUN.GA.98.L14
SYK.KE.x.KE51
SYK.KE.x.SYK173
TAL.CM.00.266
TAL.CM.01.8023
TAN.UG.x.TAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TYO1_patent
WRC.CI.97.97CI14
WRC.CI.98.98CI04
WRC.GM.05.Pbt_05GM_X02

```
                                       fusion peptide                                                                                     glycosylation NAS           glycosylation NHT
                           gp120 end gp41 start                                                                     immunodominant region   glycosylation NKS              glycosylation NYT
H1B.FR.83.HXB2             EKRAV..GIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLL.RAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKS..........LEQIWNHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDK  665
H1A1.UG.85.U455            -----..-L--I----------------I----------------S---.----------K-----------V-------Q-----------------T----S------..........Q-D---NM--LQ-EK--SS--GI-YQ------------LD--A---  658
H1B.US.90.WEAU160          -----G.ML--M----------------------L-------------.---------FE-----------V----------------------------T--------R-..........QDY---NM----E-------G--YN--------------------  668
H1C.ET.86.ETH2220          ----A...L------------------I----------------S---.K-------M-----------T-V--I--H-R------------------------S------..........Q-E--DNM---Q-----S---DI-YN-L-V-----D---KD--A---  653
H1D.CD.84.84ZR085          ----I..-L--M---------------------V--------------.---------------------------------------H---T----S---R-..........V-Y--GNM---Q-E---D---G--YN------I------K-------  665
H1F1.BE.93.VI850           ----A..-L---------DSREH-----I-------------------.----------------------V----------------------N----S------..........Q-E---NM-----EK--S--SNI-YK----------------A---  641
H1G.SE.93.SE6165           -----..-L--V----------------I-----V-----------G---.------------------L----------------------N----V------..........YNE--DNM--I--E-------YQ-Y--L------------D--A--Q  668
H1H.CF.90.056              -----..-M--S----------------I----------------S---.----Q-R--M-----------V-------R-----------------N----S------..........QSE--DNM------KQ-S---EE-YR-L-V--T-------D--A---  654
H1J.SE.93.SE7887           -----.----V------T---------I-----V-----------S---.K---------K----------V----------------------N-----------..........Y-D--ENM--IQ-E-------GI-Y-----A-----N--KD--A---  657
H1K.CM.96.MP535            -----..-L--V-F--------------I----------------S---.----------------R-------------------------N----S------..........SW-E---NM-----EK--G--SDT-YK------T-------D--A---  651
H101_AE.TH.90.CM240        -----.----MIF--------------I----------------S---.----------------------V---------KF--L------I---------ST---R-..........F-E---NM--I--E---S---NQ-YEILT------DR--KD------  659
H102_AG.NG.x.IBNG          -----..-L--V-------------R-I----------------S---.K---------K-----------V--L---R-----------------T----S-----T.........FND--DNM--IQ-EK--S---DI-YN-------R------D--A---  656
H103_AB.RU.97.KAL153_2     -----.----V----------------I-------------------.----------------------V---------------------------T-----P..........-DE-*-NM-----E-------G--YN--------------I-A---  648
H104_cpx.CY.94.CY032       --*--.----M------------------------------------S---.---------R-----------V--L-S---------------------N----S------..........YND--DNM--LQ--K------QI-YG-L------------D--A---  663
H1N.CM.95.YBF30            ----AF.-L------------------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  654
H1P.FR.06.RBF168           K--GA..-L-M-----MS---------AV--------V-H--------M-.-------E--R-S----R--R--L--I-T--R------L-----QIV-Y-N----R--T---.......ETE-DGX-TNL--Q---KLVD---DT-YLE-QRA-D--XA--KK-----Q  680
CPZ.CD.90.ANT              Q--GI..---LF---L-S----------IA--A-T-N-XH------A---.Q---T-------S---V------M----K--R-----SL---AD-VT-H-T----N--V-FTQTC.AKNSSDIQC--ENM--Q----LVQ-S-GQ-YNILQIAHE---R-KK--Y----  667
CPZ.CM.-.SIVcpzDP943       ----AF.-L------------------L-------------------.-------------S----------V------------I--L------S--Y-T----ST-GAN.........TSYDE---NL--QD--KRVK--SGV-FD---QA-E-XNT--RS-----Q  651
CPZ.CM.-.SIVcpzMB897       -----..-L-VM----M----------AL-------N---------S---.K------------S----------V--------------L------------N----NT--K-.........LGTNLSF-DNM---Q-EK--D---ET-YE-LTR------V------A---  643
CPZ.CM.01.SIVcpzCAM13      Q---AF.-L------------------AV-------------------.-----------------V-----L------Q--I--L-----S--Y-T----KT--G.........KSMSD---NL--QQ--KL-T---GT-FG-L--A-S------KD-----Q  666
CPZ.CM.05.SIVcpzEK505      Q---AF.-L------------------I-------K-------------.-------------S---------V--I---R--I--L-----SV-Y-N----TT---.........NNSYDT--GNM--QN--EQVR--SGV-FG-L-QA-E--SI--KS-----Q  637
CPZ.CM.05.SIVcpzLB7        ----A..-L-VF----------------L---------T-----S---.-------------S----------V--I--------------------------S----RT---.........KTYNE--DNM--------VR---EI-YG---QA-D---N--KK-----H  647
CPZ.CM.05.SIVcpzMB66       ----AV.-L------------------L----------------S---.-------------S----------V--------------L-----------S----TT-T-.........KSYDD--YNM---Q--K-VS---DV-YN-L-KA-T---N--K-------  679
CPZ.CM.05.SIVcpzMT145      ----L..-L-----------------VV---------T--------------.-------------S----------V-------R------L---T--T--P---R--KT-G-I.........SDYQV---NY--QQ----V----G--YT-L--ANT------K------S  654
CPZ.CM.98.CAM3             ----AF.-L------------------VV-------------------.-------------S----------V-------R--I--L------A--Y-T----NT--AN.........TSFDE---NL--QD--KRVK--SGV-F----QA-E--NT--KS-----Q  661
CPZ.CM.98.CAM5             ----AF.-L------------------VV------H-------------.-------------S----------V-------------I-SL------A--Y-T-----T--AN.........TSYDE---NL--QD--KKVK--SGV-F----QA-E--NT--KD-----Q  667
CPZ.GA.88.GAB1             Q---AF.-L------------------AV-------------------.K------------SI--V-----L------Q--I--L------AV-Y-T----N--PG.-.......NSTDD--GNL--QQ--KLVS---GK-FG-L--A-S------RD-----Q  656
CPZ.GA.88.GAB2             S---AF.-L------------------VM------N-------------.-------------S----------V-----------V-AL-----TV-YST----T--NSNK.........SY-D--SNL--QQ--KLVE-H-GT-FQ-LQRAHE--NS--K------Q  657
CPZ.TZ.00.TAN1             SR-D..V---L------S----------IA--A---G---------Q---.Q-------------S----------M----K-IR-----SL---AN--V-HSS----LT-AED-TKCNHSDAKYYDC---NL--Q----LVE-S-GT-Y--L-KA-T-----K--------  666
CPZ.TZ.2001.TAN2           SR-D-..---L------S----------IA--A---G---D-----Q---.Q-------------S----------M----X-IR---Y-S----AN--V-HSS----L--AED-QKCNSTNTKYY-C---NL--Q----LVQ-ASET-Y--L-IA-T-----K--------  671
CPZ.TZ.2002.TAN3_1         SR-D-..---L------S----------LA--A---G---------Q---.Q-------------S----------M----K-IR---Y-S----AN--V-HSS----L--AGT-SKCNYTGAKYY-C---NL--Q---KLVQ-SSET-Y--L-TA-T---R-K--------  689
CPZ.US.85.CPZUS            ----AF.-L------------------VV-------------------.-------------S----------V------------I--L------T--Y-T----DT---N.........LSYDA--GNL--Q----KVR--SGT-F----QA-E--NT--KS-----Q  638

                           gp120 end gp41 start
MAC.US.x.239               N--G-.FVL-..-----AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS.............-TPK--NE--Q--E-KVDFLEEN-TA-L--A-I-----MY--QK-NS  674
H2A.DE.x.BEN               N--G-.FVL-..-----AT---A---R-L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N----VT-I-K---H-AQ-NS---AFRQV-H-T---VNDS.............-SPD-KNM--Q--EKQVRYLEAN-SQSL--A-I-----MY--QK-NS  659
H2A.GW.x.ALI               N--G-.FVL-..-----AT---A--T-AL--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N----VT-I-K-----AR-NS---AFRQV-H-T---VNNS.............-KPD-DNM--Q--EQQVRYLEAN-SEQL-RA-I-----TY--QK-NS  657
H2A.SN.x.ST                N--G-.FVL-..-----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  654
H2B.CI.x.EHO               N--.GVLVL-..-----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  658
H2B.GH.86.D205             N--G-.MVL-..-----AM---A---T-L--SA-S-T--A------QQPV.DVVKR--E--R-----T-N----VT-I-K-----AQ-NS---AFRQV-H-T---PNET.............-TPN--NM--QQ-EKQVHFLDAN-TA-L--A-I-----MY--QKINS  662
H2G.CI.92.ABT96            N--G-.FVL-..-----AM---A--T--L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N--T-VT-I-K-----AR-NS---AFRQV-H-T---D-LGA--T..........--PQ--NM--Q--EKQ--FLEDN-TR-L--A-I-----MY--QK-NS  679
H2U.FR.96.12034            N--G-.FVL-..-----AT---A-----L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N--T-VT-I-K-----AS-NA---AFRQV-H-T---INDT.............-TPN-DNM--Q--EEKV-YLEEN-TQ-L-AA-I-----MY--QK-NN  669
COL.CM.x.CGU1              Q--...GIFAFSI-AL-SG--AA--S--VA--I--QS-NGRASASS-RM-LKLV-T-SA--------V-N--V-VATI-G--EE-AK-ASI--ANMQ--R-I----KT-............GE-DP-QNM--KQ-HERVR---DI-EADLV-AYDL--E--KK-A--GD  661
DEB.CM.99.CM40             H---AGIAF-..LVA--ST--AA-----TA----S-S---------QE--.K-V--HGQ--T--A--VRN-NT-LT-I-K-----AK-NE---AF-Q--H-T----N-L.............EDPD-DNM--Q--EMKVA---DEWEGALQRA-E---R-VHA-QS-QD  667
DEB.CM.99.CM5              Q---AGIAF-..LVA--ST--AA-----TA----S-S---------QE--.K-V--HG---S--A--VRN-NT-LT-I-K-----SK-NE---AF-Q--H-T----HT-.............G-PD--NM--Q--E-KVA---DEWEGALQRA-E---R-VHA-QS-TD  647
DEB.CM.SIVdeb04CMPF3061    Q----GMAL-..LVA--ST--AA-----TA----S-S---------QE--.K-V--HG---T--A--V-N-NT-LT-L-K-----SK-NE---AF-Q--H-T----G-I.............GQPD--NM--Q--EQKVA---DIWEEALQKA-EK--Q-VHA-QS-QD  644
DEN.CD.x.CD1               SR-RREVSLVLGLI---S---TV---VGTG-A--S-D--A------QE--.--V-GHSA----S-----N-N--LT-I-K-----A--NE---AW-Q--H-S-E-SNQ........SIPGLALYQPDFQNM--QQ-ETD-AALFGNVSDALKKA-A---R-VH-VQS-YD  674
DRL.x.x.FAO                Q--GA.VVL-..I--L-SL---A-CSVATAM---SQA--T-M-E--KQ--.-LV-Q--E--K--I--V-N----LT-L-E-IG--AM-SL----FAQV-H-N-V-PNES.............VTPN-TSE-----QKRVDSISNN-TLDLQKAYE-EQ--IF--QK-GD  710
GOR.CM.2004.SIVgorCP684con ----A..-L-M-----MS---------AI--------V-H--------M-.---A---E--R-S----R--R--L--I-T--R------L---T-----Y-N----TT-TG-.......SDSE-DG--GNL--QQ--KLVD---DT-FLE-QKA-E---A--KA------  660
GOR.CM.2007.SIVgor2139_287 ----A..-L-M-----MS---------AV--------V-H--------M-.---A---E--R-S----R--R--L--I-T--R------L---T-R---Y-N----KT-TG-.......NDTE-DD--GNM--QQ--KLVD---DT-FLE-QRA-E---A--KA------  656
GOR.CM.2007.SIVgorCP2135con ----A..-L-M-----MS---------AV--------V-H-------X-M-.---A---E--R-S----R--R--L--I-T--R------L---T-R---Y-N----KT-TG-.......NETE-DD--GNM--QQ--KLVD---DT-FLE-QKA-E---V--KA------  651
GRV.ET.x.GRI_677           ---VPF.VL-..---------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  670
GSN.CM.99.CN166            -----..LSL-IS--T--S---T-----ATA--E-S-S--A--M---E---.--V----S---PS--------T-LSSL-K--R--TI-QA---ANRP--H-I----T-...........WANGSLPD-ENM--QK-SMLVE-D-YT-QQ-L-QANQ--AS-LN--MK-S-  676
GSN.CM.99.CN71             K---..ISL-IS--T--S---A-----ATA----S-S--A------E---.--V----S---S---------LSSL-K--R--TI-QA---ANQP--H-I----D-...........WAKNSTPD-E-M--Q--SKL-E-D-YT-QQ-L-NANH--S--MND--K-S-  685
LST.CD.88.447              Q---APLAL-..A--L-S---TA--LV-TI-----QVVIQ--L---KQ--.VLV-K--E--R--I--V-N----LT-I-E-----T--AS---QW-QV-H-N-E--YN.............ITPN-TRD--I----QVGVLEAN-ST-LQ-AYTTELE-RNAFKK-QE  699
LST.CD.88.485              Q---APLAL-..A--L-S---TA--LV-TI-----QVVIQ--L---KQ--.VLV-K--E--R--I--V-N----LT-I-E-----A--AS---QW-Q--H-N-E--YN.............ITPN-TRD--I----RVGVLEAN-ST-LQ-AYTTELE-RNAFKK-QE  698
LST.CD.88.524              QT--APLAL-..A--L-S---TA--LV-TI-----QVVIQ--L---KQ--.VLV-K--E--R--I--V-N----LT-I-E-----A--AS---QW-QV-H-N----YN.............VTPN-TRD--I--E-QVGSLEAN-TT-LQ-AYTTELE-RNNFKK-QD  701
LST.KE.x.lho7              QT--APLAL-..A--L-S---TA--LV-TI-----QAV-Q--L---KQ--.VLV-K--E--R--I--V-N----LT-L-E-V-H-A--AS---QW-QV-H-N-E-TYN.............ITPN-TKD--R--ESKVAI-DKN-T--LQ-AYTTELE-QNKFKK-QE  696
MAC.US.x.251_1A11          N--G-.FVL-..-----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  673
MAC.US.x.251_BK28          N--G-.FVL-..-----AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS.............-TPD--ND--Q--E-KVDFLEEN-TA-L--A-I-----MY--QK-NS  676
MAC.US.x.EMBL_3            N--G-.FVL-..-----AT---A-----FR--A-S-T--A------QQ--.GVVKR--E--R-----T-N--T-VT-I-K--E--AQ-NA---AFRQV-H-T---PNAS.............-TPD--ND--Q--E-KVDFLEEN-TA-L--A-I-----MY--QK-NS  676
MND-1.GA.x.MNDGB1          ---GA.VIL-..I--L-SL---A--SV-VA----SQS-VT---E--KQ--.KL--Q-SE--K--I--V-N--T-LTSL-N-I---A--SQ----WAQV-H-S-E-TNTS.............ITPN-TSE--K--ETRTDYLQQN-TEMLKQAYDREQR-TY--QK-GD  692
MND-2.CM.98.CM16           Q--GA.VLL-..MF-L--L------SVAVA----SQA--N---E--KV--.SL-DQHSE--K--I--V-N----LT-L-D-VA--AR-SM----FAQV-H-H---PNDS.............ITPN-TSE--L---KRVTAL-DNMTVNLQKAYELEQ--IY--EK-GD  683
MND-2.GA.x.M14             Q--GA.VLL-..M--L--L---A--SVAVA----SQA--N---E--KI--.SL-DQHSE--K--I--V-N----LT-L-D-VA--SR-AV----FSQV-H-N---PNES.............ITPN-TSE--L----RVTAI-NNMTIDLQRAYELEQ--MY--QK-GD  684
MND-2.x.x.5440             Q--GA.VLL-..M--L--L---A--SVAVA----SQT--N---E--KV--.SL-DQHSE--K--I--V-N--V-LT-L-E-VA--SR-SV----FSQV-H-S-K-PNNS.............IVPN-TSE--L----RV-SIVTNMTIDLQRAYELEQR-IF--QK-GD  706
MNE.US.x.MNE027            N--G-.FVL-..-----AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS.............-TPN--NE--Q--E-KVDFLEEN-TA-L--A-I-----MY--QK-NS  677
MON.CM.99.L1               -----..ITL-MA-----ST--G-----ATA----S-S--A------E---.--VT---S-------V------LT---KFI---T--NA---AN-AV-H-T----N-...........WAKGHFPE-DNM--QQ-SELVD-D-MT-QQ-L-AA-E--G--QH--MKPGQ  667
MON.NG.x.NG1               -R--..ITL-MA-----ST--G-----ATA----S-S--A------E---.--VT---S---S--------LT---KFI---T--NS---ANRAX-H-Q-L--NT-..........WAKGHFPE-DNM--QQ-SMLVD-D-A--QX-L--A-E--G--AH--MK-GQ  681
MUS-1.CM.01.1085           HA--..ISL-IT-----S---G-----ATA----S-S--A------A---.--V---------S---------LT-L-KFI---A--NL---ANRQ--H-R----D-...........WANHTQPG-ENM--QQ-S-LVD-D-TT-QE-L-LA-R---E-QHK-QK-LE  688
MUS-1.CM.01.CM1239         KA--..ISL-IT-----S---G-----ATA----S-S--A------A---.--V---------S---------LT-L-KFI---A--NL---ANRQ--H-R----T-...........WANHTLPG-ENM--QQ-SKLVD-D-NT-QG-L-AA-K---E-QHK-QK-LE  682
MUS-2.CM.01.CM1246         KA--..ISL-IT-----S---G-----ATA----S-S--A------A---.--V---------S---------LT-L-KFV---AI-NL---ANRQ--H-S----D-...........WANHTQPG-DNM--QQ-S-LVD-D-AV--G-L-MA-Q---E-QHK-QK-LE  684
MUS-2.CM.01.CM2500         QA--..ITL-IT-----S---G-----ATA----S-S--A------A---.--V-T-------S---------LT-L-KFI---A--NL---ANRQ--H-R----N-...........WANHTQPE-DNM--Q--S-LVD-D-AT-QG-L-LA-K---E-QHK-QK-LE  670
OLC.CI.97.97CI12           N--G.ALVL-..L--V--L---A--TV-TI----SSH--A---R--KHM-.EM--K--E--K--I--V-N----L-G--K-----A--KQ---EFMQV-H-T----YT-............QSPD-GND--LK-Q-KVD-L-QE-DDRLTIAYQE-QR-QY--QRFQE  680
RCM.CM.SIVagi-00CM312      Q--VP.LVL-..-----S---TA----ATA----S-H--A--L---KQ--.DIV-Q-----K-----T-N----VT-I-K--A--S--NTF--AWRQV-H-S---SF.............NKTPD-QKQ--L--E-NVSYLEAN-TASLQ-A-D-H---VH--EK-NN  653
RCM.CM.SIVrcm-02CM8081     Q--VP.LVL-..-----S---TA----ATA----S-H--A--L---KK--.DIV-Q-----R-----T-N----VT-I-K--A--S--NTF--AWRQV-H-T-Q-TF.............NATPE--KQ--L--E-N-SIIEGN-SLALQ-A-D-H-R-VH--EK-NN  662
RCM.GA.x.GAB1              Q--VP.LVL-..-----S---TA----ATA----S-H--A--L---K---.DIVKR--N--K-----T-N----VT-I-K--A--S--NTF--AWRQV-H-V---TF.............NKTPE-QKES-LQ-E-N-SYLEAN-TIALQ-A-D-H---VH--EK-SN  664
RCM.NG.x.NG411             Q--VP.LVL-..-----S---TA----ATA----S-H--A--L---KK--.DIV-Q--E--K-----T-N----VT-I-K--A--S--NTF--AWRQV-H-T-E-IY-Q...............TPE--KQ--L--E-N-SRLEGN-SVALQDA-E-H-R-VHD-EK-NS  658
SAB.SN.x.SAB1C             Q--V.PFVL-..---------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  675
SMM.SL.92.SL92B            N--G-.FVL-..-----AT---A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T-L-PNDS.............-VPD--NM--Q--EKKVEFLEAN-TQML--ARL-----MY--QK-NS  681
SMM.US.x.H9                N--G-.FVL-..-----AT---A-----L--SA-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AXXNS---AFRQV-H-X---PNDT.............-TPX--NMX-Q--EKQV-FLEAN-TXXL--A-I---X-MY--QK-NX  681
SMM.US.x.PGM53             N--G-.FVL-..-----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  683
SMM.US.x.SIVsmH635F_L3     N--G-.FVL-..-----AT---A-----L--SA-S-T--A------QQ--.DVVKR-HE--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T---PNDS.............-VPN-DNM--Q--E-KVDFLEAN-TQ-L--A-I-----MY--QK-NS  682
STM.US.x.STM               T--G-.FVL-..-----AT---A-----L---A-S-T--T------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T---PNDS.............-VPD--NM--Q--E-KVDFLEAN-TQ-L--A-V-----MY--QK-NS  677
SUN.GA.98.L14              ----APVAL-..A-AL-S---TA--LV-TI-----QAV-Q--L---KQ--.VLV-K--E--R--I--V-N----LT-L-E-VQ--S--AS---QW-QV-H-N----YN.............ITPN-TKD-------QVKM-DDN-TA-LQ-AYVTELE-QNKFKQ-QE  702
SYK.KE.x.KE51              Q---APLAL-..---L-S---TA--S-ATA--L-SQT--A------QK--.E-V-------G-----V-N-N--LT-L-T--R--AIMSN---AF-Q--H---T-QQACG-N-........RCPTPQ-ENM--HT-E-QVD-L-DH-DN-LR-A-E-----VHD-TK-QE  662
SYK.KE.x.SYK173            ----APLAL-..---L-S---TA--G-ATA--L-SQT--A------QK--.E-V-------G-----V-N-N--LT-L-T--R--AI-SN---AF-Q--H---T-EKACG-N-N.......FCPKPQ-KNM--HR-EQ-VD-L-DH-DG-LR-A-E---R-VHD-TK-QE  657
TAL.CM.00.266              Q---APLAL-..-I---ST--TA---VATA----S-S---------EH--.----H-------------N-N--LT-L-K--E--AK-NS---AW-Q--Y-S----KT-..........WSNYTDPQ-QNM--Q--EMKVD-H-G--SQ-LQ-A-E---R-VH--QK-ND  670
TAL.CM.01.8023             ----APLAL-..-I---ST--TA---VATA----S-S---------EH--.----H-------------N-N--LT-L-K--E--AR-NS---AW-Q--Y-S----KT-..........WTNSTNPD-QNM--Q--EKLVD-ASDT-TV-LQ-A-E---R-VH--QK-ND  688
TAN.UG.x.TAN1              ---VPF.VL-..---------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  681
VER.DE.x.AGM3              T--V.PFVL-..---------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  692
VER.KE.x.9063              S--V.PFVL-..---------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  692
VER.KE.x.AGM155            Q--V.PFVL-..---------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  686
VER.KE.x.TYO1_patent       Q--V.PFVL-..---------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  684
WRC.CI.97.97CI14           ---G.AIIL-..I-----L---A--TV-TV--I-SQH--A------K---.ELV-K--E--R--I--V-N----LS-I-GF-E--LK-KQ---ELTQV-H-----Q-............WNATPN-ENE--QH-YKRTDELQAN-SAILVIAHD-EQA-MF--QK-SD  684
WRC.CI.98.98CI04           ---G.AIVL-..I-----L---A--TV-TV--I-SQH--A--L---K---.ELV-K--E--R--I--V-N----LS-I-GF-E--LK-KQ---ELTQV-H-N----E-............WNATPN--NE--QH-Y-RTDELQAN-SV-LV-AH-TEQ--MF--QK-QD  694
WRC.GM.05.Pbt_05GM_X02     ---G.AIVL-..L-----L---A--TV-TV--I-SQH--A------K---.ELV-K--E--R--I--V-N----LS---G--E--VK-KQ---ELTQV-H-----EEA............WNMTPA--NE--QQ-HLRTDTLM-N-SDMLVQAHA-EQS-MF--QK-TD  686
```


PLV Proteins

```
                                                                                                            glycosylation NGS
                                                                                                                          *
H1B.FR.83.HXB2                WAS.LWNWFNITNWLWYIKLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPT.......PRGPDRPEGIEEEGGERDRD.RSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLG.......RRGWEALKYW...........WNLLQYWSQELK  808
H1A1.UG.85.U455               --N.-*-----S------R--VI-----I------T----I--------------LA-I.......-E-LG--GR-------QGK-.------S-F--IA----N----------FA---A-A-----RSSLKGL-L---G---L..............----L--GR---  807
H1B.US.90.WEAU160             ---.--T--D-S-------I-------I------T----------------------A.......--------------------.--G---D-F-T---V--------L----I-----AK--------...............C..............---------------  811
H1C.ET.86.ETH2220             -EN.---------------I-------VI----I--------------------LI-H.......------LG--------QG--.--------F--IF-----------------I--AA-T-----RSSLKGLQ----T---L..........GS-V---GL---  803
H1D.CD.84.84ZR085             ---.-----S--Q------I--------I------------------------L--A.......-----------------Q--G.T-------FS--------N----------E-I--AA------........---------L..........---------R---  808
H1F1.BE.93.VI850              ---.-----D-S-------I--------I----------------K------L--LI-S.......------------G---QGK-.--V---T-F---A-----N------RH---FI--AA---DRGL.......---------L..........G--TR-------  784
H1G.SE.93.SE6165              ---.-----G--R------I--------I------------------------LTHH.......Q-E---------G---QG-G.--V---S-F-P-------------------SI---A-T-----RSSLKGL-L---G---L..............----L--GR---  818
H1H.CF.90.056                 ---.--T--D-SH------I--------I----I--------------------LV-N.......----------T--G---Q---.--V----F-PVV-------S----RL--------V-T-------........-----R----L..............--------G----  797
H1J.SE.93.SE7887              -TN.-------S-------I----I---I----I----A----------------LI-N.......-TEA---G----G---QG-T.--------F---A-----N-----------FV--AA-T-GT--........L----I---L..........V--VW--G----  800
H1K.CM.96.MP535               ---.-----D--K------I----I---I----A-----V--------------LI--.......S--A-------------Q-KN.--V---S-F---A-----N------RQ--N-I------L-RGL.......-G-------L..............---V--------  794
H101_AE.TH.90.CM240           ---.-----D---------I--------I----I--------------------PSHH.......QKE---------G---QG--.--V---S-F---A--------------------T--AA-T-----HSSLKGL-----G---L..........G---L--G----  809
H102_AG.NG.x.IBNG             ---.-----D--------RI--------I---------T-I--------------LTHH.......Q-E-----R---G---Q-K-.--V---S-F---A--------------------I--AA-T-----HNCLKGL-L--G----L..............---IS--V----  806
H103_AB.RU.97.KAL153_2        ---.-----D-SK------I------------I---------G-----------R--A.......Q-------------D------.T------RF-------------F-I--H-------AA-------...........-----------..............---------I----  791
H104_cpx.CY.94.CY032          ---.-----S--K------I--------I----I------------------L--LI--T.......Q--L---G-T------Q--S.--------F-P-------N------RH--N-----A-T----.......I--------L..............--F-L--G----  807
H1N.CM.95.YBF30               -D-.--S--G--K------IA----A-I--I--ISI-IT-IA----------L--LI--.......A-------ET-GGV--Q--G.--V---S-FS--V-E---N-LI-L----T-S---LR-TL----QSLSRGLQLLN-LRTHL..........-GI-A--GK--R  804
H1P.FR.06.RBF168              --Q.----LD-XQ------I-------II----LL-XXNV-R-I------V-L--LGLN.......GDPAGIAP-TN-----AGNG.-----LD-F-P-V----KN-VVQI-QI-VGCI-GIKDLLTI-WIHLGQLLT--LNC-RDC..........FAACG--T----  830
CPZ.CD.90.ANT                 -S-.-----D--Q------I------AI-----LLVLV-CLRK-----H------..I--QN.....QQD-EQ--E-R----RK--I.-WRA-QH-FF--L-V--T-IIQWI-QIC-TC--NLWAVLQH-CR......ITFRLCNHLE..........N--STL-.TIIR  810
CPZ.CM.-.SIVcpzDP943          -S-.-----D-S-------I-L-A-A--I-I---GVIM-LIAK----------L--LI--.......T------G-T------PG-G.T-T---S-F---A-----N-L--L----I---S-LR-TL-I--QNINKGLQWLRDISRHL..........-GVIA--GR--Q  801
CPZ.CM.-.SIVcpzMB897          -S-.-----D--Q--------V------I-I-----M--------------------PF-V.......R-D---------G---N-NV.--V---S-F---A-----N-------------W--R-TL-IV-QKFXAGL-LV--T-T-L..........KSI----G----  793
CPZ.CM.01.SIVcpzCAM13         ---.-----D---------I-L-A---II----IMS-V-VIR----------L--LI-A.......Q-E-GTLGETGG----PGSG.--V---T-CWP-------N-VIW--RS-TS-AC--W-QL-K--QLLINIL-LIQ-K-TLL..........RGII---GR--R  816
CPZ.CM.05.SIVcpzEK505         -S-.-----D--K------I---V-A-I--I--ISIIM-M-A----------L--LI--.......T--------RT--DA--L-NG.--V---S-F---A-E-F-N-L--L----T-C-S-LR-TL---RQNIHKGLQLLN-LRI-L..........-GIIA--GR---  787
CPZ.CM.05.SIVcpzLB7           -T-.-----D-SH------I----I---IVC--I----A----------------LS-I.......---LAQ-G-T------QGT-.--V--LD-F---A-N---D----L----T--V--AR-TL-IV-QYTLKGL-LV----L-L..........QGI----GK---  797
CPZ.CM.05.SIVcpzMB66          ---.-----D--S------I--I-----I-------L------------------LI-A.......R-DR----E---G---P-NV.------S-F---A-N---D----L-----------L-TL--V-QTLLKGL---R---IHL..........RGI----G----  829
CPZ.CM.05.SIVcpzMT145         --N.--S--D---------M-LIV---II----C--IG-LI----K------L--LI-A.......NQ---GLGET-KG---NV-G.------S-F-P-V-E---N-LS-L--Q---CASLIW-LL-I--QYSLRGVQQIGTL-HQL..........RGT----TT-I-  804
CPZ.CM.98.CAM3                -S-.-----D--R-----------A----I---G-II-F-AK----------L--LI--.......T-E----G-T--DV--PGKG.------S-F----E---N-L--C---------LG-TL-N--QSLNKGLQQLRNFSR-L..........-GVIT--GR--Q  811
CPZ.CM.98.CAM5                -S-.--S-----Q------I-LIV-A--I-F-LIGI-M-VIAK----------L--LI--.......T-E---RG-T--D---PG-G.--T---S-F---V-E---N-LI-F----I---S-LW-T-QI--QSINKGLQQLKDFST-L..........KGV-T--GR-IQ  817
CPZ.GA.88.GAB1                ---.-----D--K------I-L-A---II----IMT-F-V-R-----------L--LI-V.......Q-EQG-LGE-D-G---Q--S.--V---E-C-P--------N-GIW--QS-TS-ACN-W-QLKT--HLILHSL-LLR-R-CLL..........GGII---GK---  806
CPZ.GA.88.GAB2                -S-.-----D---------M--IL---VI-----IG-VN-IR-S------------LI-A.......TG--EV----GGG---Q--G.--V---S-F---C---I-N-TI-L-RL---C-SVLWSLL-Q--QQVLRGL-LLR-L-SQL..........KGIG---L---R  807
CPZ.TZ.00.TAN1                -S-.--D--D--Q------IA-I--A-------LMFIVNV-KQ-----T--FS-IPTQA.......EQD-EQ-G--AGG--G--NI.-WTPSPA-FFSIV-E---N-LIWI-QTFQNFIW-LWISLQA-KQGIISLAHSLVIVHRTIIVGVRQI....IEWSSNTYAS-R  823
CPZ.TZ.2001.TAN2              -S-.--D--D--Q-----R-A-I--A-------LMFIVNV-KK-----M--FS--PTQA.......QQD-EQ-G--A-G--G-GNI.-WTPSPT-FFSIV-E---N-LIWL-QTC-NFIWVLWT-LQA-KQGTISLANNLVIVHR-IVVKIRQI....IEWCHNTYAS-R  828
CPZ.TZ.2002.TAN3_1            -S-.--D--D--Q------IA-I--A-------LMFIVNV-KK-----M--FS--PTQA.......EQD-EQ-G-TAGG--GGGNF.-WTPSPT-FFSIV-E---N-LIWI-QTCQNFIWVLWT-LQA-KQGTISLAHNLVIVHR-IIVRVRQT....IEWCGNTYAS-R  846
CPZ.US.85.CPZUS               -S-.-----D---------I-LIV-AS---I---GVIF-L-AK---------L--LF--.......T-E------T--GA-KT-NV.--T---S-F---V-E---N-LI-L----E-----LR-T-QI--QNINKGLQLLN-LRARC..........-GVIA--AR---  788

MAC.US.x.239                  -DV.FG---DLAS-IK--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIQQ..D-AL-TREGK-RDGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRVYQI-QPILQRLSATLQRIREVLRTE.........LTY---GWSYFH  831
H2A.DE.x.BEN                  -DI.-G---DL-S-VK--QYGVH--V-IIA---AIY-VQLLS-F-K--R-VFSSPPGYLQQIHIHK..DRGQ-ANEGT-EDVGGDS.GYDLWPWPINYVQFLIHLLTR-LIGLYNICRD-LSKNSP....TRRLISQSLTAIRDWLRLK.........AAQ---GCEWIQ  812
H2A.GW.x.ALI                  -DV.FT--LDL-A-VK--QYGVY-IV-I-A----IY-VQMLS-L-K--R-VFSSPPGYIQQIHIHK..DQEQ-TRG-T-EDVG-NV.GDRLWPWPIAY-HFLIHLLAR-LIGLYSICRD-LSRISPI-QPIFRSLQRALTTIRDWLRLK.........AAY---GCEWIQ  814
H2A.SN.x.ST                   -DV.FG---DL-S-IK--QYGVY--V-IIV----IY-VQMLS-L-K--R-VFSSPPAYFQQIHIHK..DREQ-ARE-T-EDVGNSV.GDNWWPWPIRYIHFLIRQLIR-LNRLYNICRD-LSRSFQT-QLISQSLRRALTAVRDWLRFN.........TAY---GGEWIQ  811
H2B.CI.x.EHO                  -DI.FS---DF-S-MA--R-GLY--I-I-V---AIYIIQMLA-L-K--R-VFSSPPSY-QQIPIRK..DRGQ-ANE-T-E-GGNNE.GYRSWPWQIEYIHFPIRQLRD-LIWLYSGCRT-LSKTFQT-QPVLQPLRLPPA...................Y-R-GISWFQ  804
H2B.GH.86.D205                -DV.FG---DL-S-IK--H-GLY--A---V----VYIVQMLA-L-K--R-VFSSPPSY-QQIPIRK..DRGQ-ANE-T-E-GGNDG.DYRSWPWQIEYIHFLLRQLRN-LIWLYNGCRT-LLKTFQ........ILHQISTNLQPLRLP.........VAY---GISWFQ  811
H2G.CI.92.ABT96               -DV.FG---DL-S-VK-VY-GLYV-A-VIV----IY-VQ-LG-L----R-VXSSPPSYVQQIHIRK..DQEX-TKE-I--ESGNKG.GYRSWPWQIEYIHFLXRQLGN-LTWLYSNCRA-LSRIXQT-QPLFQRISRTLQAIREHLRLE.........AAYFS-GFRW-Q  836
H2U.FR.96.12034               -DI.FG---DL-S-VK-VY-GLYV-A-III----IY-VQLLGSL-K--R-VFSSHPSYVEQIPIQR..DQEQ-TKG-I-E-AG-SG.GYRSWPWQIRYIHFLIHQLIR-LTWLYSSCRD-LSRICQS-QPLF.......QSIRERLHLE.........IAY---GW-YF-  819
COL.CM.x.CGU1                 -TN.WFSG-GLF-IFK-VLYAAYV----I----IMV-IACIRGAFRVKGFQQIGRTNVSS.....QIRGQDR-WEQPDNA-EIHR.EQENSSSRFW..............RLMEEWWSSLL-HCSNGILT............VLPQHPPTTS....DKRRSA-.-T-R  791
DEB.CM.99.CM40                -D-.-----DLSR-F-W-R-VVY-IAA-IL---AMFGVN-GSKLCR-----QTL-PK.........EEEVL-LAEEGGRS-GDG-A.-WTA-PR-FFS---E--QQ-LTWL-QTF-NCIWLIR-GF---QEGIYRLSFAL-QR-RLLGERFRVL....GGY---GWK-F-  822
DEB.CM.99.CM5                 -D-.-----DLSR-F-W-R-VVY-IAA-IL---AMFGVN-GTKLCR-----QTL-PK.........EVEVLQLG-EGDKS--GVKS.-WTTSPR-F-R-----ISQ-LTWL-QT--NCAWLIR-GF---QETIYRLSFAF-QR-RLLGEHLRVV....-GYC--GWK-F-  802
DEB.CM.SIVdeb04CMPF3061       -D-.-----DLSR-F-W-RIVVYVIA---L---IMF-VN-ISKLCR-----QTPFQK..........EDLEL-GEG-GGV-GST-R.-WTPSPQ-F-S---G--QQ-LSWL-QTF-NFIWLL--GA-I-QEQIYRLSFAI-TL-QRLGEQLRVL....-GY---GWK-F-  798
DEN.CD.x.CD1                  -DN.-----DLSK-F-W-R-VVYVIA---I---IMFIINVLS-LCR-----LQIPIR.........-QKAANQPEEG-GESGDGNRLKWWQYPR-F-SI--E--SQ-LTW--LT--N-VFLIRTACSRIWDFV..........HRHSLILLG........YF--GL---Q  816
DRL.x.x.FAO                   LT-.WA---DF-W-SK---IGFF--MAII----LA-LW-T-G-F----R--.PYIFKDAYRPET..TNYHRQ-DREKG-EQD-GEW.NIRSEPSKPGSSKAWSEE..TVGTWLKESRGYIWLKNVKAV-EYGWSE.........................-QEAGRGIC.  847
GOR.CM.2004.SIVgorCP684con    --D.--S-LD--Q-I---RI--IVIA-----K-LM-IIN-CH------------ILGRN.......QVPAGIVPE-G----RA-S-.------LA-F-P-------T-VVWI-RT-VI--SGIKE-TLS-IEXLTRLLX-VNNLTRDC...........FAFIA--G----  810
GOR.CM.2007.SIVgor2139_287    --D.--S-LD--Q------I--I-IA-------LM-IINMCH------------ILGRN.......QVPAGIVPE-G----RA-N-.------LA-F-P-----I-T-VVWI-RT-VI--SGIKE-TLS-IEYLTRLL--VNNLTRDC...........FAFIA--G----  806
GOR.CM.2007.SIVgorCP2135con   --D.--S-LD--Q------I----IA-------LM-IINMCH------------ILARN.......QVPAGIVPE-G----RA-S-.------LA-F-P-------T-IVWI-RT-VI--SGIKE-TLS-IEHLTRLL--VNNLTRDC...........FAFIA--G----  801
GRV.ET.x.GRI_677              -SG.F-S--SLST--G-V-IGFLVIVIIL---FAWVLWGCIRNI----N--P........QIHIHS..SAE--DNGGGQDRGGESSSSKLIRLQEESSTPSRINNWWLN-KSCS--IRTWCYNICLT--IFIRTAVG.......................Y---GL-Q-Q  806
GSN.CM.99.CN166               -D-.--S--D-SD-QR---I-VIV-AA-IA----MFI-NMLR-I---------P-ILI-Q.......AVERGP-DAT--GV-AQGKV.--V-YLT-FSS-L-----N-VIWI-QI-AS-AWTLRG-GKF-WEQLQKAIQKAIQQARQLREVAARV....IAYIS-GI---Q  833
GSN.CM.99.CN71                -D-.--S--D-S-------I---V-AA--A---IMF--NMLR----------P-ILT-Q.......QVER-P-D-TRKD--AQGNV.--V-YLT-FSS-L-------VIWI-QI-AT-VWTLQ--GKF-WEQ....LQKAFQQARQLGEVAARA....AAYIS-GI---Q  838
LST.CD.88.447                 FN..F---LD-LS-FQ---YAVL-II-IIV--V-SFIVQNIVKMCR--RV-APSAYVEQDCKWEKR.ENQEQ-DKEG-IEKDKENIYINLEQYKKESSTPPWNVDWSEPLQDSLFVT--KWIKAFGI--LSLVWQFLSWLGHLVILFFQ.HGQHLRKACSRRVVENAQKI  865
LST.CD.88.485                 FN..F-S-LD-LS-FQ---YAVL-II-IIV---SFIVQNIVKMCR--RV-APSAYVEQDYKWEKG.EN-EQ-DKEG-IEKDKEDIYINLEQCKKESSIPPWNVDWSEPLQDSLFVT--KWIKAFGI--LSLVWQFLSWLGHLVILFFQ.HGQHLWKASSGWMVENAQKI  864
LST.CD.88.524                 FN..F-S-MDL-T-FQ---YAVL-II-III---LSFIIQS-VKMCR--RV-APSAYVEQDFKWENG.ESQEQ-DKEG-TEKD-ENIYINLEQCKKESFRPPWNVDWNEPLQDSLFVT--KWLKAWGI--LTLVWQFLTWLGHIVILSYH.HGQHLWQTLSRFLVE-AQKT  867
LST.KE.x.lho7                 FN..F-S-LD-SH-FT-V-YAVL-ILVII---VLSFIIQN-VKMCR--RV--PSVYIEQDYKWEKE.ENQEQ-DRE--K-ADTETIYINLEQCKKESSRPLWNVDWNEPLQDSL-VT--KWLKEGGI--LSLVWQSLSWL-HL-ILFFQ.NGQRLWQTSSRWMVENAQKI  862
MAC.US.x.251_1A11             -DV.FG---DLAS-IK--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQ*THIQQ..DQAL-TREGK--DGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRVYQI-QPILQRLSATLRRIREVLRTE.........LTY---GWNYFH  829
MAC.US.x.251_BK28             -DV.FG---DLAS-IK--QYG-YV-V-VIL----IYIVQMLAKL----R-VFSSPPSYFQ*THTQQ..D-AL-TREGK--DGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPILQRLSATLRRVREVLRTE.........LTY---GWSYFH  832
MAC.US.x.EMBL_3               -DV.FG---DLAS-IK--QYG-YV-V-VIL----IYIVQMLAKL----R-VFSSPPSYFQXTHTQQ..D-AL-TREGK--DGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPILQRLSATLRRIREVLRTELT.........Y---GWSYFH  833
MND-1.GA.x.MNDGB1             LT-.WAS--DF-W-VQ-L-WGVFL-L-II----LL-LWNTIS-F----R-VFS.......QDCQQN.LYRK--DNG---SNSLELG.EHNSENL...KEESLNRSLIEDLTSFARE*PIWLWLKNLRAAIEYGF............LAVQEA...IRSLGS-LVSFAAIS  833
MND-2.CM.98.CM16              -T-.WAS--DF-W--K-V-IGLL--IVIIV---LACLW-VLGKF----R--...PYVFKGDYLRPH..NLKQ-GKER--EPDSEKQ.STKSDSSKVEFGKPWSKE...QIRDWLKTSRGYVWLKNLQAVIEYGWQE.........................-KTAGGKIF-  819
MND-2.GA.x.M14                LT-.WAS--DL-W--K-V-IG-L-IMVVI----LACLW-TIG-F----R--.PYVLKGDYLRPR...NLRQ-DEER-GELNLEEQ.NIKSESSKKEFGRPWRPE...QIRNWLKESTIY-WLKNLQAVIEYGWQE.........................-KAAGARICQ  820
MND-2.x.x.5440                LNFHGLTG-DL-W--K-V-IGLLV-VVII---MLACLW-VLGKF----R--.PYVFKGDYLRPH...NLK--DREGG-EPDLEKQ.NIKSESSRQESRKPWKPE...QVRSWLK-ST-Y-WLKNLQAVIEYGWQE.........................-KAAGAALYQ  843
MNE.US.x.MNE027               -DV.FG---DLAS-IR--QYGVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIRQ..DQAL-TKEGT--DGG-SG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPIFQRFSTTLQRVREVLRTE.........LTY---GWSYFQ  834
MON.CM.99.L1                  -DF.-----D-SK------I--IV-AA-I----LMFI-GVIS-LG----L--S-IPI-S.......HA-QPT-D-TGAG--DGSNS.--PAYLK-FFTI--E---N-V-WT-QI-K-SV-VIY--LQRVSQRLPPLLHIRLLQ-WESLRRLLAYCQY.......GI---Q  821
MON.NG.x.NG1                  -DW.-----D-SK------I--IV-AA-----VLMFI-G-IR--G----L--P-...IPTQVPR....E-GA--ETG-G---QG-NRXTVYARGFLTIIWT-CRDLIVWTFQLLRDSG-T-YRSLQRVRSHLIPLLRDLCRQLREASSRLL.........AY---GL--XQ  835
MUS-1.CM.01.1085              -D-.--E--D-SK------I-C-V-A---LF-L-MF--G-LR----------P-ILIRP.......QT--EA---T-GD--SGS--.--V-YLS-FFS-F-----N-VIW--QT-ISSASTLWWLLQRAVHLIQKGSI-L-FT-RALIEPLREALQRALAC--HGI---Q  849
MUS-1.CM.01.CM1239            -D-.--E--D-SK------I-C-V-A---LF--FMF--G-LR-I---------P-I-TQP.......-EAAEP---T-GGV-SSG-T.--V-YL--F-S-F-----N-VIWF-QT-VSSVSTLRW-LQRAAHLIQKWSV-L-IVIREQARRCSELLARLLAYI--GI---Q  843
MUS-2.CM.01.CM1246            -D-.--E--D-SK----V-I-C-V-A--IA----MFI-GMLR-I---------P-I-TR..............QSAEQ-QQEKTEKG.AGRSGRHR--S-FSE--HN-MIWT-QI-TSSASNLWWLTQRIYHHLTKGLVNP-TL--SFREALGRFIAH....A--GI-K-Q  834
MUS-2.CM.01.CM2500            -DA.--E--D-SK--F-V-I-CLV-A--IA---IMFL-G-LRK---------PP-I-TRA.......-E--RT--ET--G--A-GSV.--V-YL--F-S-------N-IVW--QI--SS-SNLWWLIQ-IVHHLQKWSISL-LRARATSTHLLEVLRRLLAY---GL--FQ  831
OLC.CI.97.97CI12              LSD.WTK---LA--MT-V-IGLLV-LAY---N--LGLVRKIRIM----VS-QLP......FSDTL..QDKEQ-DNG-G-DS-FFES..TFNEW-CTTECTGTLRVNWRDLLS-LWKGVVRVTR.....................................WI.QKSLQH  800
RCM.CM.SIVagi-00CM312         -GD.AFS-LRLDW-ME---IG-FVIL-II------LLW-X-SKL-A--T--LSPPPYYHSQIPIRNLKKEEQQAKG--GEFSGEGGGYKSYNWQKGSFY-LILRPIVT--QTLWNNLIS-LLLSYQS-QSILSRIQHPLL-IVA-VQ...............DGW--F-  806
RCM.CM.SIVrcm-02CM8081        -GD.AFS-LRLDW-ME-L-IGVF-IL-VI----F-LLWGCISKI-A--T--LSPPPYYHQQTPIRNREEQPIK--G-GDS--EGG..YKSPSWQRGYFYSLILRPIEI-LRALWNSCNS-LSVIFQS-RSILSQIQQTLR-IIA-IQ...............DGW----  813
RCM.GA.x.GAB1                 -GD.AFS-L-LDW-MQ---IGFF--I-II---VAWLLWNCLSNL----R--..SPPSYVQQIHIHNTGE-QT-GEKR-D---EGG..NKYNNWLREYCW-QLIHPLSRIWTQLSQICRSCSSIIFQS-RWILA.........................KI--GW--F-  802
RCM.NG.x.NG411                -GD.MLS-L-MDW--K--RIG-F-IL-II----I-LLW-C-SKL-A--T--LSPPPCYHQQIPIRNREEQPIK--EKGGS--GGGH.KYYNWQR-.........YFYILILRPIEILCRR-YQICSNS-SVIYQSLQSIPSRIL-HLRWIGA.........K--DGW--F-  809
SAB.SN.x.SAB1C                -SD.F-S--DL-K-FGWM-IA--VIA-IIVA-VLLVIIG-LRKF-K--A---SLPSSH*.QIHIHL..DKGQ-DKE--Q-SGGLNSDSRSYTWQREFL...........RHLCH-LITWLRN.LTSWFSTIFSNLH-CLQDIQQRTRQLTA........H-E-GW----  819
SMM.SL.92.SL92B               -DV.FG---DL-S-VR--QYGVFL-I-I-L----IY-VQMLS-L----R-VFSSPPSYHQQIHIQR..DQEL-AKEGT--EGGNGG.GYRLWPWQIEYIHFLIRQLIRILTWLYNNLTR-ASRAYQN-QQLCQRLSEISQPIRELVRREAG.........YIR-GWNYFI  838
SMM.US.x.H9                   XDX.XG--XDL-X-IK--QYGVL--L-VI-----IY-VQMLA-L----R-VFSSPPAXVXQIPIXT..-QEL-TKEG---DGGX-G.GNRSWPXQIEYIHFLIRQLIR-LTWLFSSCRDWLLRXCQI-QPVLQSLSRTLQRAREVIRVEIT.........Y---GWRY-Q  838
SMM.US.x.PGM53                -DI.FG---DL-S-IK--QYGVL--L-V------IY-VQMIA-L----R-VFSSPPVYVQQIPIHK..-QEQ-TKEG---DGG--G.GNRSWPWQIEYIHFLIRQLIR-LTWLFNNCRTYLLRIYQI-QPVFQRL...LQRIREVVRTE.........ISY---GCRWFQ  837
SMM.US.x.SIVsmH635F_L3        -DI.FG---DL-S-IR--QYGVL--L-V------IY-VQMLA-L----R-VFSPPPAYVQQIPIHK..DQEP-TKEG---EGG--G.G-RSWPWQIEYIHFLIRQLIR-LTWLFSSCRDWLLRIYQI-QPVLQRLSRTLQRVREVIRIE.........ITY---GWSYFQ  839
STM.US.x.STM                  -DV.FG---DL-S-VR--QYGVYL-I---M--VAIYIMQLLA-L-K--R-VFSSPPSCRQQIPIHK..-QEQ-TKEGT-E-GG--G.GINSWPWQIEYTHFLIRQLVR-LTWLYNNFRACLSRIYQT-HPTFQRISRILQRIREVVRLG.........AAY---GCIWIQ  834
SUN.GA.98.L14                 FN..F-S-LDLSQ-FL---YAVL-I-IIIAA--LSFIIQQIY-MC---RV--PSAYVEQDWLQETC.PK-TDK-EE--TEK--I..YINLEQSKKESLPPPWTVDWDEPLRDSL-VT--KWLKAWGIV-AQNIYHLLSFL-HL-TTSFH.HGQRLWQTLRGWLGSHLIQA  866
SYK.KE.x.KE51                 -D-.--S--DLSK-FQ-L-IGFFAIAAI-I--VLSFAWG--RNMLG-----LQKPQP--........LYHKA-A-P--GIETDTDN.YKPSVSTF-REFLRQS-AEAQKLWNRVSESCRSLIRGL-IAW-...........................FIS-GVT---  792
SYK.KE.x.SYK173               -D-.--S--DLSK-FF-L-IGFYVI-A--L--L-SFSVG-IKNLLG--V-ILQNPT...........Q-RKD-GKPAD-EEGSGDR.EGLNVSTF-RESLRQS-EAGQQLWRTVCSSFRSLIRQLTITW-...........................FIS-GFN---  784
TAL.CM.00.266                 -D-.--S---LSA-FRWLRIAVIV-AS-IL----MYIVNVLKLL----L--P-K-PI-S.......N-D-EA-G-DAT---KQGSY.KWN-SQA-F-SI--E--TG-IRYI-QT-AN-AWLIWDLAKAAKAKSQQVAQATYAF-RLHLTAAAATLQY.......GFS-IQ  824
TAL.CM.01.8023                -D-.--S---LSA-FRWLRIAVIV-AS-IL----MYIINVLKLL----L--P-K-PI-S.......N-E-GV--EEGT---KP--S.-WRTSPQ-FF-I--E--N--LSY--QT--N-VWLSWDLLKR-KATTQQAAQATYAY-RIKFTAAV.......ATI--GFS-IQ  842
TAN.UG.x.TAN1                 -T-.W--I-DVSS-F-W--WGFY--I--ILF-MAWLIWGCIA------F---P.......QINIRL..-REQ-DNAGG-DKDSSSS.-DKSPPSVKES-LPNRGGIQAEERAW-QHLTNWCL-ISSW--RLYQILRRSLTTLLQLLRQE.........CQYI--GW-QF-  831
VER.DE.x.AGM3                 -SD.F-S--DFSK--NIL-IGFLD-L-II---LLYT-Y-CIA----------P.......QIHIHP..WKGQ-DNA-GP-EGG-KRKN-SEPWQKESGTAEWKSNWCKRLTNWCSISSIWLYNSCLT--VHLRSAFQ.......................YI--GLG---  829
VER.KE.x.9063                 -SN.F-S--DFSK--NIL-IGFLV-V-II---LLYTIY-CIG----------P.......QIHIHP..WKG--GNAG-PEEGGE--ND-SDSWQKGSGTRQKRGSWSKRLTNWLSTFSTWLYNTFLT--IHLRGAWQ.......................YI--GLG---  829
VER.KE.x.AGM155               -S-.F-S--DFSK--NIL-IGFLA-I-VI---LLYTLYTCIA----------P.......QIHIHP..WKGQ-DNAG-PEEGGRTGKSKSTH*QKEFGGRDKRTSWCRQLTNWF-TLSIWLHNSCLT--LKLRSAWQ.......................Y---GLG---  822
VER.KE.x.TYO1_patent          -SD.F-S--DFSK--NIL-MGFLVIV-II---LLYT-YGCIV------V---P.......QIHIHQ.V-KG--DNAD-P-EGG-NSRIKLESW*KDSKSRCMQ-TAWLTRLN......TWLYNSCLT--IQLRKAFQ.......................Y---GLA---  815
WRC.CI.97.97CI14              -TA.WGS-LQWF-IWNWL-WGLLVI-IV-----LACLMNSIRIA-----K-IS-......QNHLNN.LEEGQQDSE-GIEEDYNSIEINDWQYRRES-IVFLNSLLTS---ICQHLLHIVWSLLSCL-WWINDGW........SLLKDI..............SGRAYGA  824
WRC.CI.98.98CI04              -TT.WGS-LQWF-IWNWL-WGLL-I-IV--I--LGWLMTSIRIA-----K-IS-......QSHLNN.LEEGL-DRE--AEEDNSTLDV-DWQFKKES-ITLLNSLINF-SAVCQTLLQIVWSLLACLFWWIQHGW........NLLKRI..............SGAAHAC  834
WRC.GM.05.Pbt_05GM_X02        -TQ.WGS-LSWF-IWNWL-WVLLVI-LV--I-VLGCLFTSLRIA-----K-VS-......QSHILN.LEEGL-DNE-T-EEDFIAIEE-DWQFKKESLILFWNSLATL-Y-ICLSLLQ-SYSLLSYLWI-IQNGW........SLLKRL..............CGAAYAA  826
```

```
                                                                                    gp41 end
                              glycosylation NAT                                      Env end
                                                         *
H1B.FR.83.HXB2                NSAVSLLNATAIAVAEGTDRVIEVV...QGACRAIRHIPRRIRQGLER..ILL....*   856
H1A1.UG.85.U455               I--IT--D-V-V---GWI-----IG...-TIG---LN-----------..A--.....-  856
H1B.US.90.WEAU160             ----------------------I-...-RT----L------------..A--.....-   860
H1C.ET.86.ETH2220             K--IN---T---V-G-----F--LI...-RIW--FCN----------A..A-Q....-   852
H1D.CD.84.84ZR085             ---I--VD-----------I-DI-...RR--K-VL---T---------..A--.....-  857
H1F1.BE.93.VI850              ---I--F-T---V-------I---L...-R-G--VLN--------A--..A--.....-  833
H1G.SE.93.SE6165              ---I---DTV-----NW-------A...-R-----LN--T---------..A--.....- 867
H1H.CF.90.056                 ---ID---T----------GI-VI-...-R-W---L---------F--..S--.....-  846
H1J.SE.93.SE7887              ---I----T-----------I--IA...-R-F---L------------..A--.....-  849
H1K.CM.96.MP535               ---I----T-----G---I--IG...-R-F--LL------------..A--.....-    843
H101_AE.TH.90.CM240           I--I---D-----A-GW-------A...---W---L------------..T--.....-  858
H102_AG.NG.x.IBNG             ---IN---TI--V--NW---A--IG...-RVG----N--------F--..A--.....-  855
H103_AB.RU.97.KAL153_2        S--IN-IGTI-----GW------IG...-RF---M-N--------A-K..A-Q....-   840
H104_cpx.CY.94.CY032          ---IN-F-T-----------I--A-...-R-----CN-----------..A--.....-  856
H1N.CM.95.YBF30               D--I----T---V-------I--LA...-RIG-G-L------------..A-I....-   853
H1P.FR.06.RBF168              Q--T---DTV--S--GW--Q--I-G...-QIG-GFLN--------I--..S--.....-  879
CPZ.CD.90.ANT                 TEIIKNIDRL--W-G-K--SILLAL...-TIV-I--EV---------I..A-N....-   859
CPZ.CM.-.SIVcpzDP943          I--T---DT-----------I------...-RIG-G-L------------..S--.....- 850
CPZ.CM.-.SIVcpzMB897          T--I---D------------X--IA...-RVG-G-L------------..A--.....-  842
CPZ.CM.01.SIVcpzCAM13         T--T---D------G-----I-RA-...-IVF-I-GN-----------..T--.....-  865
CPZ.CM.05.SIVcpzEK505         I--IN--DT--V--------I--L-...-RIG-G-L------------..A--.....-  836
CPZ.CM.05.SIVcpzLB7           T--T---DT-----------I--I-...-RIG-G-L---------F--..A--.....-  846
CPZ.CM.05.SIVcpzMB66          T--I---DT-----------I--IA...-RFG-G-LN-----------..A--.....-  878
CPZ.CM.05.SIVcpzMT145         ---I---DT-----G-----IL-AI...TRLG-G-L------------..A--.....-  853
CPZ.CM.98.CAM3                T--I---D------------IL--A...-IIG-G-L------------..S--.....-  860
CPZ.CM.98.CAM5                I--T---DT------N----I--F-...-IVG-G-L------------..S--.....-  866
CPZ.GA.88.GAB1                I--I---D------------I--AF...-VTL-I--N-----------..A--.....-  855
CPZ.GA.88.GAB2                T--I---DT---I---R--TI---A...TRIG-G-L------------..A--E....-  857
CPZ.TZ.00.TAN1                VLLIQAIDRL-NFTGWW--LI--G-...VYIA-G--N----------L..A-N....-   872
CPZ.TZ.2001.TAN2              A-LIHAIDRL-DFTGWW--LI--GI...TYIG-G--N----------I..A-N....-   877
CPZ.TZ.2002.TAN3_1            A-LIHAIDRL-DFTGWW--LL--G-...VYIA-G--N----------I..A-N....-   895
CPZ.US.85.CPZUS               V--T---DT-----------I--LT...RRLFLG-I------------..S--.....-   837
                                                                             Env end

MAC.US.x.239                  EAVQAVWRSATETL-GAWGDLW-TL...RRGG-W-LA----------L..T--.....-  880
H2A.DE.x.BEN                  EAFQAFART-RETL-GAWGWLW-AA...RRIG-G-LAV-------A-L..A--.....-  861
H2A.GW.x.ALI                  EAFRAFARIARETLTNTWRDLWGA-...-WVG-R-LAV-------A-I..A--.....-  863
H2A.SN.x.ST                   EAFRAFAR--GETLTNAWRGFWGTL...GQIG-G-LAV-------A-I..A--.....-  860
H2B.CI.x.EHO                  EAIQAAAR-AGETL-SAARTSWG-L...RR-AGE-IA--------A-L..A--.....-  853
H2B.GH.86.D205                EALRAAAR--GETL-SAGETLW-AL...RR-A---IA----------L..T--.....-  860
H2G.CI.92.ABT96               EACTAATR-AQETLTSTWRALWKTL...GRVG-G-LA----------L..T--.....-  885
H2U.FR.96.12034               EAFQAFGK-ARETLSRTGRELW-TL...GRVG-WL-A--------F-L..A--.....-  868
COL.CM.x.CGU1                 YL..LVPRGPSLQIL-TLQ-WLRSA...ARGW-RAPEYL-GWIYDRPQ..GPA....-   838
DEB.CM.99.CM40                DTITWAGGELWQWLQGIAEMALQGL...VR-GG--LRV-A-----A-L..--N....-   871
DEB.CM.99.CM5                 DTIIWAGQEIWEWLQGIAQVA-QGL...VWGG-NLLA--A-----A-L..L-N....-   851
DEB.CM.SIVdeb04CMPF3061       DTI-WAGGELGQTFQRWAEVALQ-L...GRGI-E-LA--A-----A-L..F-N....-   847
DEN.CD.x.CD1                  TGLRD-GTSAIQQGRATAEV-LAAL...TR-A-EVVA----------I..V-N....-   865
DRL.x.x.FAO                   TLLQEVAQRIWREGRQLGLSSARAL...RALAQEVAA--------A-V..LFN....-   896
GOR.CM.2004.SIVgorCP684con    Q--I---DCV-VWT-DW--Q--AIA...-RIG-G-LN-----------..S--.....-  859
GOR.CM.2007.SIVgor2139_287    Q--I---DCV-VWT-NW--Q--AIA...-RIG-G-LN-----------..S--.....-  855
GOR.CM.2007.SIVgorCP2135con   Q--I---DCV-VWT-DW--Q--AIA...-RIG-G-LN-----------..S--.....-  850
GRV.ET.x.GRI_677              EA-TG-AQ-L-R-AR-AWG-LGAI-...RS-Y--VINS---V-----K..V-G....-   855
GSN.CM.99.CN166               AA-TGI-DSL--FTWNW-EA-LHAC...RRVW-EFLA--------A-I..LFN....-   882
GSN.CM.99.CN71                AAVSGI-DSL-VFTWNW-EPLVQT-...GRVW-EFLA--------A-I..L-N....-   887
LST.CD.88.447                 A-WSWNKIRRNRRDISAAHARNLRL...GQKK-WRFRFRG-SGFPS-TTETA-.....-  916
LST.CD.88.485                 ARWSWNKIRRNRRDISAAHARDIRL...GQKK-WRFRFRG-SGFPS-TTETA-.....-  915
LST.CD.88.524                 A-WIRNKSRRNRGELSK-NGTNIRL...GQNR-RW.RFRF-SGLPS-TTETT-.....-  917
LST.KE.x.lho7                 Q-WLREKCRRNRGQLSS---KNIQL...GKKK-WRLRFGG-SGISS-ATETA-.....-  913
MAC.US.x.251_1A11             EAVQVDWRSATETL-GAWGDL*-AL...RRGG-W-LA----------L..T--.....-  877
MAC.US.x.251_BK28             EAVQAGWRSATETL-GAWRDLW-TL...RRGG-W-LA----------L..T--.....-  881
MAC.US.x.EMBL_3               EAVQAGWRSATETL-GAWGDLW-TL...RRGG-W-LA----------L..T--.....-  882
MND-1.GA.x.MNDGB1             WR-GYTGWLQ-NRIFT-CREA-IAA......GTC-W--L-----SA--..P-N....-   879
MND-2.CM.98.CM16              VLQGYAQGLWSRGHQW-LSTAACFR....AIA-G-IN--------A-V..L-N....-   867
MND-2.GA.x.M14                VLQSFAQRLWSGGHQL-LSVIRGAA....AIG---GN--------A-V..L-N....-   868
MND-2.x.x.5440                VLQGFAQRLWSRGYQL-LSG-RGAA....AFG-G-WN--------A-A..L-N....-   891
MNE.US.x.MNE027               EAVQVAWRSATETL-GAWGDLW-TL...GRVG-W-LA-------E--L..T--.....-  883
MON.CM.99.L1                  AAVT---D-L-RFTIVW--ALLHAG...GRLW---VA--------A-I..F-N....-   870
MON.NG.x.NG1                  XACTGAID-L-RFTVIW--A-LRLG...GRLW-GLVAV-------A-I..L-N....-   884
MUS-1.CM.01.1085              TASTAV-DQL-TFTWNW-ET-LQAG...RRVW-EFLA----V---A-I..L-N....-   898
MUS-1.CM.01.CM1239            AAGTAA-DQL--FTWNW-EPILQIG...RRVW-EFLA--------A-I..L-N....-   892
MUS-2.CM.01.CM1246            AAVSGA-D-L--FTWNW-EPILQFS...-KIW-KFLA--------A-I..L-N....-   883
MUS-2.CM.01.CM2500            EGVTAAIDGL--WT-NW--A-LQ-S...KQL*K-FLA--K-----A-I..L-N....-   879
OLC.CI.97.97CI12              VWRTLKQKRQ......HGINLQQKK...-RRGSPLQKAQEQLGREGAE.TTGV....-   844
RCM.CM.SIVagi-00CM312         AF--W-GEMA-ENAYYTWRGLQSLA...RQLAGWPATCG---------..L-N....-   855
RCM.CM.SIVrcm-02CM8081        EFSGW-AEMAQQNAYFVWRGLCVTA...-DIA-WPGTVC------F--..L-I....-   862
RCM.GA.x.GAB1                 EFSSWFAEMALQNAYYTWRGLCA-A...RDFAGWPAMVC---------..LCN....-   851
RCM.NG.x.NG411                GFSSW-AEMARTNAYYTWRGLCA-C...RDPAGWPATLC------F--..F-N....-   858
SAB.SN.x.SAB1C                AA-A--TV-VVQ-ATSAS-S-RHAC...RSIV--VIAH---M--E----..WFN....-  868
SMM.SL.92.SL92B               EACQEAWRSAQE-IVGAWGLIW-TL...GRVG-G-AA----------L..M-N....-   887
SMM.US.x.H9                   EA-XXWWKFXRETL-SAWRDLW-TL...GRVG-G-LA-XX-------L..T--.....-  887
SMM.US.x.PGM53                EAVQAAWR-RGETL-SAWRDSW-TL...GRVG-WVLA----------L..A--.....-  886
SMM.US.x.SIVsmH635F_L3        EA-QAWWKVARETL-SAWRDIW-TL...GRVG-G-LA----V-----L..T--.....-  888
STM.US.x.STM                  EA-QAAWR-AGETL-SAGRDLW-TL...GRVG-R-GA----------L..T--.....-  883
SUN.GA.98.L14                 T-RIRNACR-SRERVSSSQKARSRTFSLGRKW-PKWNRT-GS-IPS-TTETT-.....-  920
SYK.KE.x.KE51                 EA-I--GREV-HQMVAIWQALLAYA...RRVAENVAAL---L-----I..Y-N....-   841
SYK.KE.x.SYK173               IA-A--GREIRDW--AIWQAIYAAT...RRVVE-VAAL---L-----I..Y-N....-   833
TAL.CM.00.266                 AAWRDGIT-L-RFTWWW-NG-LAAA...RW-GE--AA----T-----Y..CFT....-   873
TAL.CM.01.8023                AAWGDCVT-L-RFTWRW-SG-LAAA...RW-VQ--AA----T-----Y..C-A....-   891
TAN.UG.x.TAN1                 EG-ARSFE-L-S-AQSASRTLWNAC...RS-Y---LEH---M--E----..WFN....-  880
VER.DE.x.AGM3                 AA-QEAVV-L-RLAQNAGYQIWLAC...RS-Y---INS---V-----G..--N....-   878
VER.KE.x.9063                 TA-QEARL-V-RFAQNAGHQIWLAC...RSTY-H-ISS---V-----E..--N....-   878
VER.KE.x.AGM155               AA-QEAV-HL-SFARNAAHQIWLAC...RS-Y---INS---V-----E..V-N....-   871
VER.KE.x.TYO1_patent          TG-QEI-QTL-GVAQNACHQIWLAC...RS-Y-N-VNS---V-----E..--N....-   864
WRC.CI.97.97CI14              IRKKIAKSRKENTQ--VSATGSRKT...AASS-IRMLAT-LCASFDKCRRSFSERRQ-  879
WRC.CI.98.98CI04              IRKKITKSRKETTEK-ISENSNRET...T-SS-VWKL-T-LCSSLNKSRRR-F....-  885
WRC.GM.05.Pbt_05GM_X02        LQEKIAARRQKRKEK-IYTPGSRET...SRSSSIRGITA-LCTNLNPCRRGQQ....-   877
```

```
                            myristoylation                                                                                                                      phos
                             Nef start                                                                                                  acidic cluster    poly-P helix  phos
H1B.FR.83.HXB2              MGGKWSKSSVIGWPTVRERMRR..................AEPAADRVGAASRDLEKHGAITSSNTAATNAAC.........................AWLEAQ..............EEE........EVGFPVT.PQVPLRPMTYKAAVDLSHFLKEKGGLEGLIH  102
H1A1.UG.85.U455             -------K-RVE--E--K---E..................TPA--KG---V-Q--D-Y--V-----SS---S-.........................------..............--G........D-----R.-------------F---F--------D----  102
H1B.US.90.WEAU160           ---I---R-GS---AI----K-..................-----EG---V----A---------E..N--D-.........................V--K--..............-D-........------R.-------------H-----#----------Y   99
H1C.ET.86.ETH2220           ---TM--C-PV---AI---I--..................-A---EG--------D-Y--L-----P-N-PD-.........................------.............E---........------R.-------------F---L------------Y  103
H1D.CD.84.84ZR085           ---------IV---AI---I-K...........TDPRERRRP-----G---V-----R---------T-----.........................--V---.............E-D-........------R.--------------------RK-------VY  111
H1F1.BE.93.VI850            ---------IV---A-G----Q..................TPT--EG---V----DRR--------RT--PDL.........................------..............---........------R.----V------------------------Y  102
H1G.SE.93.SE6165            ---------IV---E----I-N..................TPT--EG---V-Q--DR-----------N-PD-.........................------.............E-DS........------R.---------F-G-F---F--------D---Y  103
H1H.CF.90.056               --------RMG--S-I------..................---V-EG---V----DRR--V-IN---S--RDA.........................------.............EDG-........------R.-----------G-F------------D---Y  103
H1J.SE.93.SE7887            --N----.....--Q-------................APA----G---V-Q--A---------------D-.........................------..............T--........------K.--I--------G-----F--------D---Y   99
H1K.CM.96.MP535             ---------IV---AI-------...........ARPAADRVGTQ----G---V-Q--AR---V-----SHN-PD-.........................------..............---........------R.-------------F--GF--------D---Y  111
H101_AE.TH.90.CM240         --S------IV---Q---KIKQ..................TP--TEG---V-Q--D----------.ID--D-.........................V--R--..............-D-........------M.-----------G-F---F--------D---Y  100
H102_AG.NG.x.IBNG           ---------IV---K-MK---Q..................TPT--TG-----Q--DR----------Q--PD-.........................------..............-D-........N-----R.-----------G--------------D---Y  102
H103_AB.RU.97.KAL153_2      ---------IV---Q----I--..................APA---RG--PV-Q--D-Y--V-------N--D-.........................------..............---........------R.-----------G-F------------D---Y  103
H104_cpx.CY.94.CY032        ---------IV---EI------AR......A.EPERMRRAQ-----AG---V-Q--D------IN------PDK.........................T-----.............E---........------R.---------F-G-L------------D---Y  115
H1N.CM.95.YBF30             --KI-----LV---EI-----...............QTQEP.-VEP-VGA----Q--ANR----IR--RDN-ESI.........................------.............E---........------R.-------I---Q-F---F---D-------VW  108
CPZ.CD.90.ANT               --SA---IKWV-A...-QAI-K..................IHETNPD-.I-PCGNE-ASR--L---TIGTEKDVI.........................TYS-DH..............T--........GT----R.----M----E-L-----X---------DX--X  100
CPZ.CM.-.SIVcpzDP943        --N------LV---E--A-I-Q..................-Q---PG--SV-----A-----TR--PQN-QTL.........................----EM............QGH--........------R.---------F-G-F--GY------------Y  104
CPZ.CM.-.SIVcpzMB897        --N------LV---A--D-L--..................TP---EG--PV-----R------R--PQ--ETL.........................----E-..............RD-.........------R.-------------F----------------Y  102
CPZ.CM.01.SIVcpzCAM13       --AQ-----LV---EI-K-IKE..................-PI--EG--EV-K----Y-----RH-PS--QTL.........................---DEM..............KHH......EE------R.-----------G-F------RG-------VY  104
CPZ.CM.05.SIVcpzEK505       --KV-----LV---Q-ID----..................QQDP-EG---V-Q--ANR----IR--KEN-Q-L.........................----E-..............K--.........A-----C.-----------E-F---F--R--------VW  102
CPZ.CM.05.SIVcpzLB7         --S------LV---E--R-I-Q..................ETV--EG--P--Q--AR---------PQ--ETI.........................----E-..............K--..........-IW---K.---------F-E-F--GY-----------VY  102
CPZ.CM.05.SIVcpzMB66        ---------LV---E--D-L--..................TQT--EG--PV----AR------R--SQ--ETL.........................----EV..............QN-.........------R.---------F-G-F--------------VY  102
CPZ.CM.05.SIVcpzMT145       --A------LV---EI---IK-..................-PT---G--EV-Q-IASR--V-IR--PQ--ETL.........................----EM..............QD-.........------K.---------F-G-F----------------W  102
CPZ.CM.98.CAM3              --NR-----LV---D----LSE..................LNLL-PG--PV-Q---A---V-TRH-PQN-QTL.........................----EM............QEHN-.........------R.---------F---F-------K--------Y  104
CPZ.CM.98.CAM5              --N------LV--AA--Q-I-Q..................TP---SG--PV-Q---AR----TR-H.TKYQTL.........................----EM............ENH--.........------R.-----------Q-F--------------VY  103
CPZ.GA.88.GAB1              --T------LV---E--R-I-E..................-PT--EG--EV-K---R------R--PE--QTL.........................----EM..............DN-.........------R.----T-------F--------------VY  102
CPZ.GA.88.GAB2              --N------IV---Q----L--..................TQE--EG--EV-Q--ARR-----RH-PQ--QTL.........................----EM..............-Q-........D-----R.-------------F--------------VY  102
CPZ.TZ.00.TAN1              --NIFG..RWP...GA-KAIEDL................HNTSSEP.--Q--Q--QNK-GL-TNTLGTSADVL.........................EYSAD..............HT-.......E------R.-A--M----E-L-I---W--------D--FF   98
CPZ.TZ.2001.TAN2            --NIFG..RWP...GA-RAIEDL................HKSSHEP.I-Q--T--QNR-GL-NNTIGTSADVV.........................EYSAD..............HT-.......E------R.-A--M----E-L-----W--------D--YF   98
CPZ.TZ.2002.TAN3_1          --NIFG..KWP...GAQRAIEEL................HKSSHEP.--Q--K--QNR-GL-TNTIGTPEDVL.........................EHSAD..............HT-.......E------R.-A--M----E-L-----W--------D--FY   98
CPZ.US.85.CPZUS             --N------IV---E--N-L-Q..................TQTTA--EG--PV-Q--AE-----TR--PQN-QTL.........................---DEM............TNH-S.........------R.-----------Q-F--GF----------VY  106
                             Nef start     R17Y mutation                                                                                premature stop in Mac239

MAC.US.x.239                ---AI-MRRSRPSGDL-Q-LL-A.........RGETYGRLLGEVE-GYSQSPGG-D-GLSS.L-CEGQKYNQG....QYMNTPWRNPAEEREKLAYRKQNMDDI..............D-*.....DDDL--VS-R.-K----T-S--L-I-M---I--------IYY  133
H2A.DE.x.BEN                --ASG--KLSKHSRGL---LL-A.........RGDGYGKQRDASGGEYSQFQEESGREQNS.P-CEGQQYQQG....EYMNSPWRNPATERQKDLYRQQNMDDV..............DSD.....DDDLI-V---.-R--R-E----L-I-M---I------Q-MFY  134
H2A.GW.x.ALI                --ASG--KRSGPLQGL---LLQT.........PGETCGGQCSGSGGGYSQSQGGSGRGQKL.P-CEGQRYQQG....DFMNTPWRTPATEREKELYKQQNMDDV..............DLDD....DDSL--VS--.-R-Q--T----L---M--LI--R-----MFY  135
H2A.SN.x.ST                 --ASG--KRSEPSRGL---LLQT.........PGEASGGHWDKLGGEYLQSQEGSGRGQKS.P-CEGRRYQQG....DFMNTPWRAPA.EGEKGSYKQQNMDDV..............DSD.....DDDL--V---.-R----E---RL-R-M--LI---------YY  133
H2B.CI.x.EHO                --SAG--KQSKQQ-GL---LL-A.........RRGPRGESSGERQE-SLQYPGGSD-GLNS.P-CDGQKTLGA...................EGGGKQDSD..............-DD.....EDN---VR-R.-G-------F-L---M--------E---IFY  116
H2B.GH.86.D205              --SAG--KRSERQQGL--KLL-V.........PERPYGRLSGERREQSSRSPGESD-DLNS.P-CEGQNARGA...................EGGGQQDADE..............SD-......DD---AICKT-I---------L---M---I--Q-----MYY  117
H2G.CI.92.ABT96             --SAG--KRAPQQLGL-K-LLQA.........RGEPYGKLW-GLEEGYSQYREESG-GLSS.L-CEGQQYTQG....QFMNTPWRNPATERAKLAYRXQNNDDV..............DSD.....DNDL--V--Y.-K----V-S--L-I-M---I--------IYY  134
H2U.FR.96.12034             ---SI--KHSKRSGKL---LLAA.........RGENYGRLWGGLE-GFELYQGGSG-DLSS.RFCE-QGQGYS..EGEFMNTPWRTPATGKDKLQYKQQNMDDV..............D--.....DNDL--VA-R.-R-A--T----L---M---I------D-IFY  136
COL.CM.x.CGU1               --SLL-CLNIRQQ-QT-GGVLGRLFG..IYWYPGVQAFRFLRHFRGGYDLRRVAGV-RLNICEDGYMTVPKVLPNCPCP...............LCYRHPDVYYHNPEP..........-DV.....EDV-A-CY-M.SRL-V-QA-I-LL--I-C-----------I-R  136
DEB.CM.99.CM40              ----N--IPSR-QVGSFGSGS-G...................LLR.WRYRDLSEQ-EQFSECLLESDREQSSS...............STE..FWGSPRR-IKCTNKQQQ........DLD.....QTEA-----R.---SI-DP---LMI-Y-----------DVFY  117
DEB.CM.99.CM5               ----N--TPSF-QGKKSGSGSKG...................LLR.WRYKDLSGEAETFSPFLQESDKEQSCY...............STEEPLYADPHR-IKCTQRAQE........CYD.....EEHGI--A-R.-R--I-EP---LMI-Y--Y--------DIFY  119
DEB.CM.SIVdeb04CMPF3061     ---RN--TQSS-QEENWGKHFKG...................GQR.WRYRYLGEESEKFLPYQPESDKELNCS...............LTEETFIGNPRR-IKCTQRQQQ........DFD.....QQE------K.-R--I-DP---LMI-Y-----------DIFY  119
DEN.CD.x.CD1                --S-N--QAS.-IWGLAPSNKEG...................RQRKWYWRLLPEQQEKWWPSPDGSDRALKLSSTR.LLEHDEDEETFVDAEEIIQSSPA-DP-HPRIHQYV.....VDHD.....DEE------R.-RR--HAP---DMI-M----------QDIFY  136
DRL.x.x.FAO                 ---QN--KREEAYARFYK-L--.............GYGAKGGNLDYHQLEPSEPL-R-WQPFPGGLDK-QRSS...................STSSLDTTHMQ-AAA.EKDVTLPT....---QPSEEEEE------C.--R---EP---DL--F----------K-IWW  132
GOR.CM.2004.SIVgorCP684con  --NA-K---FV---A------QT.............SPEP..DP-PG--EV-IA-AQR-G-PAKY-PKN-Q-L..........................-F---...............H-N.......E-----V.---------F-G-F---F-----------VF  105
GOR.CM.2007.SIVgor2139_287  --NT-K---F----A------QT.............NPEP..KP-PG--EV--A-AQR-G-PAKY-PQN-Q-L..........................-F---...............H-D.......E------R.---------F-G-F---F--N---------VF  105
GOR.CM.2007.SIVgorCP2135con --NT-K---F----A------QT.............SPEP..KP-PG--EV--V-AQR-G-PAKY-PRN-Q-L..........................-F---...............H-D.......E------X.---------F-G-F---F-----------VF  105
GRV.ET.x.GRI_677            --SSN--RQQQ-LLKLWRGL-GK.........PGADWVLLSD-LIGQSSTVQEECG-ALKKSWGK...GKMTP...................DGRRLQEGDTFDE...........WDD.....DEE------Q.-R----Q----L---F-----S----D-IYY  118
GSN.CM.99.CN166             --S-N--QQQ.QESSTALLSS-G...................TGQRPYFTLVDEYGENFWLSPDASDKGR.........................RYYLT-EPK..PKR.........GSL.....EEY-PSC--R.-R----DP---VM--------------MFY  106
GSN.CM.99.CN71              --S-N--QQS.QESSTALLSSHG...................TGQSPSFRLLDEYGENSWLSPDASDRGR.........................RYYLT-ES...PSR.........QNC.....IDY-PSC--R.------DP---LM-------------AMFY  105
LST.CD.88.447               --NIFG-PAAD--WKTLR-L-A................G-GTRSEGTEETYQQ-M.QETSD..WDKRRDG..........................................DSD.....SEE------R.--R--CSP---TLI-----I---------YW   95
LST.CD.88.485               --NIFG-PAAD--WKTLR-L-A................G-GTRSEGTEETYQQ-M.RETSD..WDKRKDG..........................................DSD.....SEG------R.--R--CSP---TLI-----I---------YW   95
LST.CD.88.524               --NIFG-P-ADS-WR-LRKL-A................GSGTR-EGTEESYQK-M.EQTLG..WDKTEE...........................................DGD.....SDS------R.--R--CSS---TLI-----I---------YW   94
LST.KE.x.lho7               --NAFGRPAAD--WKTLR-F-A................GSGRN-EGTEDSY-A-I.EKTSN..WAKRRDG..........................................DSD.....SEE------K.--R--CSP---TLI-----I---------YW   95
MAC.US.x.251_1A11           ---TI-MRRSRSTGDL-Q-LL-A.........RGETYERLLGEVE-GSSQSLGE-D-GLSS.L-CEGQKYNQE....QYMNTPWRNPAEEREKLAYRKQNMDDI..............D--.....DDDL--VS-R.-K----T-S--L-I-M---I--------IYY  134
MAC.US.x.251_BK28           ---AI-MRRSKPAGDL-QKLL-A.........RGETYGRLLGEVE-GSSQSLGG-G-GLSS.R-CEGQKYNQG....QYMNTPWRNPAEEKEKLAYRKQNMDDI..............D--.....DDDL--VS-R.-K----A----L-I-M---I--------IYY  134
MAC.US.x.EMBL_3             ---AI-MRRSKPAGDL-QKLL-A.........RGETYGRLLGEVE-GSSQSLGG-G-DLSS.R-CKGQKYNQG....QYMNTPWRNPAEEKEKLAYRKQNMDDI..............D--.....DDDL--VS-R.-K----A----L-I-M---I--------IYY  134
MND-1.GA.x.MNDGB1           --SSQ--KRSEA-VRYSSAL-Q......LVGGPVTPDGYKQI-SSQGAEKQSLLRGRAY-TYSEGLDKVQ-DPL.........................TKDEKLDLTQQ.........DP-.....EEE------C.R--S--VPS--DLI-F---I------G-IYY  123
MND-2.CM.98.CM16            ----S--QQEEKYLKYYKA---.............GYGAEGTNGDYQQLH-SEPL-GALSTSQEEFDREQKSS.........................STD..................---..........-T----Y.--C-V-EP---DL--MP-----------IW-  104
MND-2.GA.x.M14              ----S--QREQ-YAKYYKAL--.............GYGAEGTNLDYQLLEEQRLLEGPL-TSLGASDKELKSC.........................STKV...............DD-......QKE-R----Y.-TQ-V-EA---NLI-M------------IWF  108
MND-2.x.x.5440              ----S--QREQHYIKYYKAL--.............GYGAEGTNLDYQVLEEQQPLAGAS-TSLAASDKVLRPY...................STERDLRHQQDVTLPSEKEQ..PS....D--.........E------Y.-TQ-VPEA---DLI-M------------IWW  124
MNE.US.x.MNE027             ---AT--KRSKSPGDL-Q-LL-A.........RGETYGRLW-GLE-GYSQSLGGSD-SLSL.L-CEGQKYNQG....QFMNTPWKNPAGEREKLAYRKQNMDDI..............D--.....DDGL--V--R.-R----A----L---M---I--------IYY  134
MON.CM.99.L1                --S-N--QQS.PLSSTPLLGSQS...................SGRM-YFMLED-YGEQSWLSPDASDRER.........................KYSLT-GRNGKQRRQ..PL....DDD....DDDDG--C--R.-R----DP-W-LMM----Y-------GEMFY  112
MON.NG.x.NG1                --S-X-XQPA.QE-LMPSHGSQS...................SGQT-SF-WED-YGEDSWLSPDASDRGR.........................RFSLT-GRNHRSNRR..PTTVV.DDA...DEDQDLP-C--R.-M----DP-W-IMM------------DK-FY  116
MUS-1.CM.01.1085            --S-S--QPA.QQSLISSPPSPG...................TGRKQYFKLVEEYGENSWLSPDASGRGR.........................RYSLT-GA...SKR.........PVI.....EEH-RTC--R.-R--I-DP---LM--M------------MFV  105
MUS-1.CM.01.CM1239          --S-N--QQE.QQ-LTSSPSSPG...................TGQSPYFRLVDEYGENSWLSPDASDRGR.........................KYSLT-GS...NER.........PLI.....EEH-RSC--R.-R----DP---MMI----Y----------MFY  105
MUS-2.CM.01.CM1246          --S-N--PQS.QERSTPL-SSHG...................TGPN-YFSLVKKYGENSWLSPDAS-RGR.........................KYSLT-NK...GRR.........QII.....EEH-CSC--R.-K--I-DVN--LMI----Y--K-R----MFV  105
MUS-2.CM.01.CM2500          --S-N--KES.QQ-LMGSPSGPQ...................TGQTPYFKLVNNYK--SWLSPNAS-RGR.........................KYSLTKGS...NKK.........AII.....KKHKRSC--R.--H-IQNP---LIIN-------------MFY  105
SAB.SN.x.SAB1C              ----S--QQQRHSLWLWSKL-QA.........PVIQYDML-D-LLGQSSHIQEECA-SLRDGLIRQGDSSRTE...................EGVKMKHQGRQPSW...........YD-.....DEE------R.-CL---A----L-I-FG------------IYY  122
SMM.SL.92.SL92B             ---IT-SKPARKHGDL-KKLLSA.........RGDSYGRLWDGLEEGSQQSQGGSDRALNSCLTRDGYKYNV....GEFMNTPWRNPATERQKQQYRQQCMDDI..............D-D.....DDELA-VV-W.-K-----A-S--L-I-M---I------D-IYY  135
SMM.US.x.H9                 ---VT--KQRXAGGNLX--LLQA.........RGETYGRLW-GLEGEYSQSXDASG-GLSS.L-CEPQKYCEG....QFMNTPWRNPATEGAKLGYRQQNMDDV..............DN-.....D..L--C--S.-R--V-I----L-I-M---I--------VYY  132
SMM.US.x.PGM53              --AAG--KQSRQRGGLG-KLLQA.........RGETHGKLW-GLE-GYSQSRGE-GRDWNL.H-FEGQGYSEG....QFMNTPWRNPAREREKLKYRQQNMDDV..............DDD.....DDELI-VS-H.-K----A-S--L-I-M---I--------IYY  134
SMM.US.x.SIVsmH635F_L3      ---AI--KQHRRGGKL---LL-A.........RGETYGRLW-GLEEGYSQSLGASG-GLSS.L-CEPQKYSEG....QYMNTPWRNPTAEKAKLGYRQQNMDDV..............DD-.....DDDL--VS-H.-K----A----L-I-M---I--------IYY  134
STM.US.x.STM                --ASG--KQRKQHGEL---LL-A.........RGETYGKLL-GLGEGS-PSQGASD-GLNS.H-CEPQRYNEG....QFMNTPWKNPAAESAKLEYRQQNMDDV..............D--.....DDNL--VA-H.-R----E----L-I-----I-S------IYY  134
SUN.GA.98.L14               --NAFGRP-EV--VRTLF-L-A................GSGTR-EPA-REYHR-R.RQEVEPLVS-ENGGP.........................NGI.................-Q-.........--E---R.--R--CKP---QLI-----I---------WY   98
SYK.KE.x.KE51               --SQS--KQPSR-DEKWRTRWWPFG.......KPYSPMPDELLRMSQPYHEDF-RGWRSTL-EPILD..PKRDFD...................SGGKK-NAGDICH...........D-G.....DEDL-----L.-------L---L-I-----I-R----Q-MFF  123
SYK.KE.x.SYK173             --STS---QQLRSEGKYAIGW-LFG.......KQYTPLPDELSRPLQPCRGGF-KAWRSTL-EPIDPHGPDRDWGH...................SGGQKFSPGDIVQ...........D-G.....DTGL-----C.--T---TL---L-I-----I-N----Q-MNY  126
TAL.CM.00.266               --SVK--LLG.ETASQ-LQGSHG...................GGRMVYWRLRDGQV-RSQLSPGARGKDLNIALL........DEEWQATGADYMGK-LTQGEVWYEKEQ........--N.....DMEL-----R.-K----TC---LGI-F-F--------K-IFY  126
TAL.CM.01.8023              --SVK--LLG.ETASQ-LQGSPG...................AGRLVYWRLRDGQCRQLQRSRGERGKGLNIALL........DAGWQASDANWMGK-LS-GEVWYEKEQ........--N.....DMEL-----R.-------CN--LGI-F-F--------K-IFY  126
TAN.UG.x.TAN1               ---SN--REQQ-LLRLWRAL-KA.........PVVRYGML-D-LIGQSSNIQEECD-NWNGG-TRR..GKSTP...................EGRKLAADDTW...........DDW...EPEEE------R.-R----Q----L---F-------------IYW  119
VER.DE.x.AGM3               --LGN--PQHKKQLSLWHALHKT.........RATRYGLL-D-LIGQSSTLQEECD-ALKESLIRKRNGKMTP...................EGRKLQEGDKW...........D-W...SDEED------R.-R----Q----L---F----------D-IYY  121
VER.KE.x.9063               --LGN--PQHKKHVSLWHALHKT.........QVTKYGLL-D-LIGTSS-VQEEYD-GLRKSLIRKRNG-MTP...................DGRRLQEGDEW...........D-W...SDDED------K.-R----Q----L---F----------D-IYY  121
VER.KE.x.AGM155             --LGS--PQHKKQL-IWRALHAT.........RHTRYGLL-D-LIGQSSTLQEECD-GLRKSLIRKRNG-MTP...................EGRRLQDGD-W...........D-W...SDEED------R.-R----QI---L---F----------D-IYY  121
VER.KE.x.TYO1_patent        --SQN--PAHKKYSKLWQALHKT.........HVTRYGLL-D-LIGTSSTVQEECD-ALRKSLIRKQNG-MTE...................EGRRLQEGDTW...........--W...SDDEE------R.-R----Q----L---F----------D-IYY  121
WRC.CI.97.97CI14            ---VF--T--EEHMERYA-RLQR..........AGRRIRRQRYQQLAAERQQQAV-SGCLLPD-AQ-LI--EE.........................-SVKED..............SSD.....EDEG------.----I-QP-F-LL-----LF-------DIFW  113
WRC.CI.98.98CI04            ---IF--E--EQRMHAYA-RLQR..........AGRKQRRKRYQKIATERQQVAV-SGNLLPD-AQ-LIR-EE.........................DSSKE..............DSS.....DDE------H.----I-QA-F-LL-----LF-------YIFW  112
WRC.GM.05.Pbt_05GM_X02      ---VF--D--EQRMLRYR-RLLR..........VGRKERRRRYTRLAAERQAEAAASGELLPD-AQTLIR-EE.........................DSSKE..............DSS.....DDDAI----R.------QA-F-IL-----LF-------DIFW  112
```

```
                              HXB2 premature Nef end                                                                                                              normal Nef end
H1B.FR.83.HXB2                SQRRQDILDLWIYHTQGYFPD*QNYTPGPG.......VRYPLTFGWCYKLVPVEPDKIEEA.NK....GENTSLLHPVSLHGMDD...PEREVLEWRFDSRLAFHH...VARELHPEYFKNC.................................*..............  205
H1A1.UG.85.U455               --K--E-----V-----F---W--------.......I---------------D-AEV---.TG....---N-----ICQ--V--...E-K---M-K---T--LK-...R-Y-----FY-D..................................-..............  206
H1B.US.90.WEAU160             --K--------V-----F---W--------.......T----C----F-------E-V---.-E....---N-----M---------.H-K---M-K---K-----...-----------D-.................................-..............  204
H1C.ET.86.ETH2220             -KK--E-----V-N---F---W--------.......----------F-----D-SEV--I.-E....---NC----A-----E-...ED----K-K---H--RR-...M--------Y-D-.................................-..............  208
H1D.CD.84.84ZR085             --K--------V-----F---W--------.......I---------FE----D-EVV-K-.TE...--DNC----ICQ---E-...E-K---V---N-----E-...K-K-KY-------.................................-..............  216
H1F1.BE.93.VI850              -KK-G-T----V---------W--------.......I-----L---F-----D-EEV-K-.-E....---NC----M-Q---E-...ED----R-K---S--LR-...I---R---FYQD..................................-..............  206
H1G.SE.93.SE6165              -KK--E-----V-N-------W--------.......T-F-------F----MD-AEV---.--....---N-----ICQ---E-...ED----V-----S--RR-...I--------Y-D-.................................-..............  208
H1H.CF.90.056                 -KQ--------V-N-------W--------.......E-F-------F-----N-QEV-Q-.-E....---N-----M-----E-...DG----M-K-----LT-...L--VK----.-D-.................................-..............  207
H1J.SE.93.SE7887              -KK--E-----VHN-------W--------.......I---------------D-SEV---.-E....---NC----ICQ--IE-...E-----Q-K---S--RR-...I-------FY-D-.................................-..............  204
H1K.CM.96.MP535               -KK--E-----V-----F---W--------.......I---------------D-AEV--T.TE...--DNC----INQ---E-...EH--I-M-K---S--RR-...-------D-Y-D-.................................-..............  216
H101_AE.TH.90.CM240           -KK--E-----V-N---F---W--------.......I-F--C----F-----DQREV--D.--....---NC----M-Q--IE-...E-----M-K---A--RK-...----Q----Y-D-.................................-..............  205
H102_AG.NG.x.IBNG             -KK--E-----V-N---F---W--------.......T-F-------F----MD-AEV---.-E....---N-----ICQ---E-...DD----I---------R-...T--------Y-D-.................................-..............  207
H103_AB.RU.97.KAL153_2        -KK--E-----V---------W-...--IR.........F-------------D-AEV---.TE....---N-----ICQ------...E-K---M-K------LT-...R-------FY-D-.................................-..............  203
H104_cpx.CY.94.CY032          -KK--E-----V-----F---WD-------.......E-F--C----F-----D-QEV---.TE....----C----I-Q---E-...E-----K-K------YK-...--------FY-D-.................................-..............  220
H1N.CM.95.YBF30               -RK--------M-----IL--WH-------.......I---V-----F---LSAEEV---.-E...-D-NA---ICQ--A--...DHK---V-----S--RR-...--------FY---.................................-..............  213
CPZ.CD.90.ANT                 XLK-AA---M-MFN---I---W----EE--.......-----CR---F-----SPPD.........D-RNI---ACT--DG-...-HK-I-R-E--AS-MRR-...I--R-----RD..................................-..............  199
CPZ.CM.-.SIVcpzDP943          -R---E-----V-----F---W--------.......----------F---LT-EEV-K-.-E...-DTNI---ICQ-EIE-...EHG---V--Y-----LR-...I--Q----YQD..................................-..............  208
CPZ.CM.-.SIVcpzMB897          -----E-----V-----F---W--------.......----V-----F-----D-TDV---.-E...-D-NV---MCQ--QE-...EH---M-Q---H--LT-...R--------Y-D-.................................-..............  207
CPZ.CM.01.SIVcpzCAM13         --K--E-----V-----F---W--------.......E---PC----F-----LTTEQV-A-.-E...-D-NC----ICQ---E-...ESK---I---------LR-...I---K----Y-D..................................-..............  208
CPZ.CM.05.SIVcpzEK505         -TK--E-----V-----I---W--------.......----------F-----LS-EEV---.-Q...-D-NV---MCQ---E-...-DK---V-----S--RV-...R-------FYQ--.................................-..............  207
CPZ.CM.05.SIVcpzLB7           -KK--E----GV-----I---W--------.......I---I---G-F----LT-EEV---.--....---NL----ICQ---E-...EDK---I-KY--Q--LR-...I-----S--Y---.................................-..............  207
CPZ.CM.05.SIVcpzMB66          -KK--E-----V-----F---W--------.......I---------F----L--EEV-R-.-E...-D-NI----ICQ--QE-...EA----V-T------LK-...R--------Y-D-.................................-..............  207
CPZ.CM.05.SIVcpzMT145         -RK--E-----V---------W--------.......----------F----LP-EEV-K-.-E....---NC----MCQ---E-...E-----V-T---K--LR-...L---K----YRD..................................-..............  206
CPZ.CM.98.CAM3                -IXXXXXX-HGY-----F---W--------.......------Y-G-F----LT-EEV-Q-.-E...-D-NI----ICQ---E-...AHG---V--------LR-...I--Q----YQ-..................................-..............  208
CPZ.CM.98.CAM5                -TK--E-----V-----I---W--------.......----------F----LT-EEV-K-.-E...-DTNI----ICQ---E-...EHG---V--------LK-...L--Q----YGDHNPAPAQ...........................-..............  214
CPZ.GA.88.GAB1                -R---E-----V-----F---W----T---.......T-F--C----F----LTEEQV-Q-.-E...-D-NC----ICQ---E-...EDK---V--------LR-...I--Q----Y-D..................................-..............  206
CPZ.GA.88.GAB2                -K---E-----V---------WN-------.......------Y---F----S-EAV-----------DNL----LCT--FE-...E-K---I-K---Q--LR-...L---K----YRD..................................-..............  206
CPZ.TZ.00.TAN1                -PK-AA---T-M-N---V---W--------.......I----CR--LF-----D-PED..........D-KNI----ACS--TT-...-DG-T-I-----S--RR-...I--RY-----...................................-..............  196
CPZ.TZ.2001.TAN2              -TK-AA---T-MFN---I---W--------.......-------R--LF-----D-PED..........D-QNI--R-ACS--SN-...--G-T-I-----S--RR-...I--RY-----...................................-..............  196
CPZ.TZ.2002.TAN3_1            --K-AA---T-MFN---I---W--------.......-----VR--LF-----D-PED..........D-KNI----ACS--SN-...--G-T-I-----S--RR-...I--RF---Y-....................................-..............  196
CPZ.US.85.CPZUS               -R---E-----V-----I---W--------.......-------Y---F----LTEEEV-Q-.--.......--TNI----MCQ---E-...EHG---I-Q--TE--RR-...R-K-------R-..................................-..............  210
                                                                   max HIV-1 similarity                                                                                    normal Nef end
MAC.US.x.239                  -A--HR---IYLEKEE-II--W-D--S---.......I---K----LW-----N.VSD-AQ.ED.....-EHY-M--AQTSQW--...-WG---A-K--PT--YTY...E-YVRY--E-GSKSGLSEEEVRRR..LTARGLLNM....ADKKETR-..............  263
H2A.DE.x.BEN                  -R--HR---IYLEKEE-II--W----H---.......----MY---LW---S--.LSQ-AE.ED.....-ANC-V--AQTSRH--...EHG-T-V-Q---M--YNY...K-FT-Y--E-GHKSGLPEKEWKAK..LKARGIPYSE..........-..............  258
H2A.GW.x.ALI                  -E--HR---IYLEKEE-II--W----H---.......I---MF---LW-----D.VPQ-GE.DT.....-THC-----QTSRH--...THG-T-V----PK--HDY...K-FI----E-GYKSGLPEDEWKAR..LKARGIPFSKNR......NS-..............  263
H2A.SN.x.ST                   -D--RRV--IYLEKEE-IIG-W----H---.......----KF---LW-----D.VPQ-GD.DS.....-THC-V--AQTSRF--...-HG-T-V----PT---SY...E-FIRY--E-GYKSGLPEDEWKAR..LKARGIPFS...........-..............  256
H2B.CI.x.EHO                  -E--HK---TYLENEE-IVSGW----H---.......----KF---LW----IN.MIA-PE.DE.....-THC-V--AQTSAW--...-HE-T-V-Q---L--YDY...--FSRF--E-GYQSGMPEKEWKAK..LRARGIPTE...........-..............  239
H2B.GH.86.D205                -E--HR---TYFENEE-IVSGW----H---.......I---KY---LW------.VPAATR.EE....E-THC-M--AQISSW--...IHG-T-I-Q---L--YDY...--FNRF--E-GYQSGLPEEEWKAR..LKARGIPTD...........-..............  241
H2G.CI.92.ABT96               -E--HR---IYMEKE--II--W----A---.......I---X----LW-----T.VSD-AQ.ED.....-THC-V--AQTSPW--...--T---A-K--PT--YDY...R-FI----E-RWKSGLPEAVWKEK..LKQRGLPIE...........-..............  257
H2U.FR.96.12034               ----HR---IYLENEE-II--W----S---.......T-W-MC---LW--E--D.ANDKAQ.ED.....-RLY-VGSAQTSCEE-...HWG-A-V-K---S--YSY...Q-FIKC--E-GSKSGLSEEEVKRR..LTARGLPVK.........NC-..............  261
COL.CM.x.CGU1                 TPE-D-LIEQYA-IEW-CLKGWLE-EDEL-EDGALKED-K--VA--LW---YI-QLGEYAYSYD.......L---SVT-RKKKKP.....QQ-AIEMV-.....................*.........................................  223
DEB.CM.99.CM40                -A--HA--E-HAQNEW-II-GWLQ--K---.......----IY--FLF------.IADPDY.END....-RNI---DAHQGM-E-...-H--R-V-K---T--YCY.K.AGH.AEQR.EHTRRCMFPKRK..........................-..............  228
DEB.CM.99.CM5                 -A--HA--EIHAH-EW-II-GWLK--E---.......P---TY--FLF------.IADPDY.END....-RNI---DAHQGQAE-...-HK-R-V-K---S--YCY.K.AGH-A-KP.EH-RRCMFPKRK..........................-..............  231
DEB.CM.SIVdeb04CMPF3061       -A--HA--E-HAQNEW-II-GWLQ--E---.......I---KY--FLF------.IADPDY.END....-RNI---DAHQGQ-E-...-YK-R-V-K---Q--YCY.K.AGH-A-TKETHTRRCMFPKRK..........................-..............  232
DEN.CD.x.CD1                  -P--HA----YAQNEH-FITGW-T--D---.......I----E--FLW-----T.IEE-YD.--.....-ENC---DRYEGQQA-...-W--T-V----PE--WCY.K.AGHKKLQTSQH-NYCQRAKPRK........................-..............  249
DRL.x.x.FAO                   -L--ET----YAQNEW-FIKGW-S--K---.......I---KM--FLF------.VDEDLV.-Q.....QCNR--NSSQMGPI--...--G-R-M-Q--PG--YTF...Q-CIH---E-GHV......KSIKS..VEEPG............CCKRKWWQFKPTTEGCHG  262
GOR.CM.2004.SIVgorCP684con    --A-S------V---------W--------.......------Y--LF-----SEAEA--MGTA....S-RAI----ACS--R--...-HG-T-V----RA-GTT-...--LQK---L-QRD.................................-..............  211
GOR.CM.2007.SIVgor2139_287    --A-S------V---------W--------.......I--------LF-----SEAEA--MGTS....N-KAM----ACS--R--...-YG-I-V----RA-GTT-...--LQK---L--RD.................................-..............  211
GOR.CM.2007.SIVgorCP2135con   --A-SE-----V---------W--------.......I--------LF-----SEAEA--MGTA....S-RAI----ACS--R--...-HG-I-V----RA-GTT-...--LQK---L--RE.................................-..............  211
GRV.ET.x.GRI_677              -E--EK--N-YALNEW-IID-W-A-S----.......I---RV--F-F-----D.LHE-AR.-C.....-RHC-M--AQMGEDP-GI.DHG---V-K--PK--VEY........R-DM--DMHEHAKR...........................-..............  224
GSN.CM.99.CN166               CED-HQK-EQYA-LEW-LV-GWLSF-----.......T---TIP-F-IC-R--A.TTEDSE.PGD....DEYL-T--AYQGRSE-...QHK-F-VFS-C-K--IKS..GIQLDQLQQEERKMR..LTANRFL.........................-..............  219
GSN.CM.99.CN71                CED-HQK-ESYC-YEW-IV-GWLQW-----.......I---TMP-F-WC-R--A.MTEDSE.PGD....DQYL-N--AYQGQQE-...HH--I-VFS-C----LKS..GWQMNQLQQEERKKR..LTANRFL.........................-..............  218
LST.CD.88.447                 --------C-YCENEW-LIG-FMT--D---.......T-F--A---LWQ-E--A.CDEYTD.PS....DYRQC---SSQ-GVQE-...-WG-R-I-H-NPM--VDY...--LRKQ-KSIQ.......ASAIGL..QR................K-..............  206
LST.CD.88.485                 --------C-YCENEW-LIG-FMT--D---.......T--------LWQ-E--A.CDEYTD.PS....DASQC---SSQ-GVQE-...-WG-R-I-H-SPM--VDF...--LRKQ-KSIQ.......ASSLSLNCKR................K-..............  208
LST.CD.88.524                 ----N---T-YCENEW-LIG-FM---D---.......T----A---LWQ-E--A.CDEFKD.PS....D-RQC---SSQ-GVLE-...-WG-R-I-H-NPM--VDY...--LRKQ-ASVQ.......ATAMRFNCERK................-..............  207
LST.KE.x.lho7                 --------I-YCENEW-LIG-FM---D---.......T--------LWQ-E--A.CDEYKD.PS....D-TQC---SSQ-GVLE-...-WG-R-I-H-NPM--VDF...I-LKKQ-AKIQ.......NTAFAFDCKR................K-..............  208
MAC.US.x.251_1A11             -A--HR---IYLEKEE-II--W-D--S---.......I---K----LW-----N.VSD-AQ.ED.....-EHY-M--AQTSQW--...-WG--PA-K--PT--YTY...E-YVRY--E-GSKSGLSEEEVRRR..LTARGLLNM....ADKKETR-..............  264
MAC.US.x.251_BK28             -A--HR---MYLEKEE-II--W-D--S---.......I---K----LW-----N.VSD-AQ.ED.....-RHY-MQ-AQTSKW--...-WG---A-K--PT--YTY...E-YARY--E#GSKSGLSEEEVRRR..LTARGLLNM....ADKRETR-..............  263
MAC.US.x.EMBL_3               -A--HR---MYLE...-II--W-D--S---.......I---K----LW-----N.VSD-AQ.ED.....-RHY-MQ-AQTSKW--...-W#---T-K--PT--YTY...E-YVRY--E-GSKSGLSEEEVRRR..LTARGLLNM....ADKRETR-..............  260
MND-1.GA.x.MNDGB1             -R--EE----YAENEW-FE-GW-Q--T---.......T---K---FLF--E--SRAIGD-Y.AA......-NH---SSQ-CPQE-...--G-T-M-SGTLI-PMTL...QH.........................................*..........  214
MND-2.CM.98.CM16              -K--EE----YAQNEW-FI-TW-S--D---.......I---K--RFLF--C--A.VPPDQE.-N.....-CNK--QSSQ-GIQEE..#-WG-R-V-K---G--YTF...Y-PIKR-GE--HV......QSISH..IAYAK.......EHKPECCKRKWWQF..........  230
MND-2.GA.x.M14                -R--EE--N-YAQNEW-FI--W-E--S---.......I---KR--FLF--I--Q.VPPDQE.-Q.....-CNR--NSSQ-GIQEE..#-WG-R-M-K--PE--YTF...Y-PIKC-KEYRHV......TSLSY..TAYQE.........KSDCCKRKWWQF..........  232
MND-2.x.x.5440                -K--EE----YAQNEW-LI-NW----K---.......I---K---FLF----A.VSPPLE.ED.....-CNR--NSSQMGIQE-...--G-R-M-K---G--YTF...Y-PIIR--EY-CV......TSLSY..EAYKKE.......EKPDCCKRKWWQF..........  250
MNE.US.x.MNE027               -E--HK---IYLEKEE-IV--W-D--S---.......P---K----LW-----N.VSD-AQ.EG.....-ENY---AQTSQW--...-WG---V-K--PT--YTY...E-YIRY--E-GSKSGLSEEEVRRR..LTARGLLKM....ADKKETS-..............  264
MON.CM.99.L1                  CED-HRKIEQYA-LEW-LI-GWLQ--E---.......----TMP-FLWC-R--A.TTEDSE.EGD....-DFL-T--AYQGR-E-...-H-QF-VFS-C-K--VKS..GRQLAQLQQEERKKR..LAANRIL.........................-..............  225
MON.NG.x.NG1                  CXDXHQK-EQYA-VMW-LV-GWLQF-X---.......----TIP-FLWC-R--A.IQEDSE.DGD....XEFL-T--AYQGRDE-...-HKQF-VXS-C-K-GVKS..GPQLDELQQEERKRR..LTANRIL.........................-..............  229
MUS-1.CM.01.1085              -TE-LQK-ET-A-VEW-II-GWLM--D---.......T---KIP-FLWC-RE-A.MTEGSE.EGD....DQYL-D--AFQGREE-...HHKQF-VFS-C----LKS..GLQLDQMQQEERKKR..LATNHIL.........................-..............  218
MUS-1.CM.01.CM1239            CEE-HQK-ETYT-VDW-IV-GWLQF-E---.......T---TLP-FLWC-RE-A.ITEDSE.EGD....-EYL-T-AAYQGKEE-...-HKQF-VFS-C----MKS..GRQLDQIQQEERKRR..LTANRIL.........................-..............  218
MUS-2.CM.01.CM1246            CID-LQK-ETYA-VKSKIV-RSLI--D---.......T---RQP-FLWC-R--A.MTE-SE.PGD....DQYL-T--AYVGRDE-...HHK-F-VFS-C----LKS..GPQLNQIQQKERKRR..LTANRIL.........................-..............  218
MUS-2.CM.01.CM2500            -EE-HQK-ETYA-VEW-LI-GWLM--E---.......-----MP-VLFC-R--H.ITENSE.LGD....DEHL-C--AFIGREE-...-HK-F-VFS-C----LKS..GRQLNQMQQEERKRR..LAANRILSQS..P...................-..............  222
SAB.SN.x.SAB1C                -E--KK----YALNEW-IVDGW----D---.......T---KC----F-----D.LSE-AK.-S.....--HC----AQVAYE--...AWK-T-V-K--PL--VDY.......VAWRLHPEQVPSA.QG...........................-..............  227
SMM.SL.92.SL92B               -E--HR---IYLEKEE-II--W----S---.......I---KF---LWQ----D.VSD-AN.-D.....-THC-V--AQTYQHS-...-WGK--A-K-NPH--YTY...E-FVR---E-GWKSGLTEEEVERRLANKPKPQKKM....ADKKETS-..............  267
SMM.US.x.H9                   -I--HK---XYLEXEE-II--W----A---.......I---MF---LW--X--N.VSD-AQ.ED.....-THC-M--AQTSQW--...-WG---A-K--PX--YXY...K-FVE---E-GSQSGLSKEEVQRR..LTARGLLKM....ADKKKTS-..............  262
SMM.US.x.PGM53                -E--HR---IYLEKEE-IT--W----S---.......I---MF---LW-----S.ASD-AQ.ED.....-THY-V--AQISQW--...-WG---A-K---Q--YRY...E-FIR---E-GSKSGLSEEEVKRR..LTARGLLKM....ADKKETS-..............  264
SMM.US.x.SIVsmH635F_L3        NEK-HR---MYMEKEE-II--W----L---.......T---MY---LW-----D.VSD-AQ.ED.....-THC-V--AQT-QW--...-WG---A-K--PE--YSY...K-FIKY--E-GSKSGLSEEEVKRR..LTARGIYKM....ADKRETS-..............  264
STM.US.x.STM                  -E--HR---MYLEKEE-IV--W----A---.......I---KQ---LW-----D.MSN-AQ.ED....DGTHY-V--AQT-QW--...-WG---V-K--PL--HTY...E-FVR---E-GSKSGLPKEEVERR..LTARGLLKM....ADKKETS-..............  265
SUN.GA.98.L14                 -RT-EE----YAENEW-FITGW-D--K---.......----KA---LW--A--T.IDEDRD.PN....HPCQA---SSQQGVNE-...-WG-R-I-T--PT--YDF...R-IQK---E--H......VTSLQW...................EV*.............  208
SYK.KE.x.KE51                 C-K-HE--Q-YLKNEH-VID-.IT--S---.......T----I---LWE-A-N-.IEGYLV.DE.....-D-LMM--AAGV-ASED..-H--N-M-N-NPH--YTPGWEM-.RQQL-RQTGKR.........................-..............  229
SYK.KE.x.SYK173               CEK-DE--H-YLQNEH-II.-RI---S---.......T----I---LWE---N-.IEGCLE.YE.....-H-L----A-GQ-SSSMGE-HV-.....LQPPPGYTPGWEM-.R-QL-RQTGKPQE...LQSALSKN..IS..............-..............  241
TAL.CM.00.266                 NK--HA--N-YAHNEW-IL--W----E---.......T----C--ILW--C---.IHEDDT.ED.....GH.L-P--AYDGQAE-...-WG-S-V-V--EK--YTP..GAKMAEWDRLEREKRMLLAPPQTASS.....................-..............  241
TAL.CM.01.8023                NN--HA--N-YAHNEW-IL--W-S--E---.......T----C--ILW--C---.IHEDDT.ED.....GH.L-M--AYDGQQE-...-WG-A-V-V--EK--YTP..GAKMAEYDRLERKKKELLAPPQTAS....S.................-..............  241
TAN.UG.x.TAN1                 -PK-EQ--N-YALNEW-IID-W-A-S----.......T-K-RC--F-FE----D.VSQ-AQ.D......-RHC----AQIEWES-...-WK-T-V-K-NPL--VQY......NPD..S--DMHGLVKRK.........................-..............  223
VER.DE.x.AGM3                 -D--NQ--N-YALNEW-IID-WNAWSE---.......I---RC--F-F-----D.LHE-AE.TC.....-RHC-V--AQVREDP-GI.NHG---V-K--PM--VQY......DPNRK-LTDMHDL.GKRK.........................-..............  230
VER.KE.x.9063                 -E--NR--N-YALNEW-IID-WNAWST---.......----RC--F-F------.MHE-AE.TC.....-RHY-A--AQIKEDP-GI.SHG-T-V-K--PM--VQY......DPNRQ--EDMHAIVKRK.........................-..............  230
VER.KE.x.AGM155               -D--NK--N-YALNEW-IID-WNAWSK---.......I---RC--F-F-----A.LHE-AE.TC.....-RHC-V--AQ--EDP-GI.NHG-I-A-K--PM--VQY......DPSR---TDLYSTVGTGN........................-..............  231
VER.KE.x.TYO1_patent          -D--NK--N-YALNEW-IID-WNAWSK---.......I-F-KC--F-F-----D.LHE-AQ.TC.....-RHC-V--AQMGEDP-GI.SHG-I-V-K--PM--IQY......DPNR---TDMHGLVKRK.........................-..............  230
WRC.CI.97.97CI14              -A--EE-VN-YAHNEW-LIDGW-A--E---.......I-W-K-----W-----H.FSE-SD.-T.DPYY-RNK----ACAEPE-P....WG-H-V-K--PK--CDF...--GR-PGQSIATGKEAIESLYK.NK...KQCK..............-..............  234
WRC.CI.98.98CI04              -P--EE-VN-YAFNEW-LIEGW-A--E---.......I-W-K-----W-------.FSE-SN.DS.SQYY-KNK----GCQESE-P....WG-H-V-K--P---CDF...--SR-PGHIIAQGEDAIETLHK.EKL..KRHK..............-..............  234
WRC.GM.05.Pbt_05GM_X02        -P--EE-VN-YA--EW-IVEGW-A--D---.......----K-----W-----H.FSE-SP.-S.DQGF-KNL----AVAESE-P....WK-Y-V-K--P---CDF...--GR-ADGQIATGIETL..YEM.KK...RNNK..............-..............  231
```

```
H1B.FR.83.HXB2              .......  205
H1A1.UG.85.U455             .......  206
H1B.US.90.WEAU160           .......  204
H1C.ET.86.ETH2220           .......  208
H1D.CD.84.84ZR085           .......  216
H1F1.BE.93.VI850            .......  206
H1G.SE.93.SE6165            .......  208
H1H.CF.90.056               .......  207
H1J.SE.93.SE7887            .......  204
H1K.CM.96.MP535             .......  216
H101_AE.TH.90.CM240         .......  205
H102_AG.NG.x.IBNG           .......  207
H103_AB.RU.97.KAL153_2      .......  203
H104_cpx.CY.94.CY032        .......  220
H1N.CM.95.YBF30             .......  213
CPZ.CD.90.ANT               .......  199
CPZ.CM.-.SIVcpzDP943        .......  208
CPZ.CM.-.SIVcpzMB897        .......  207
CPZ.CM.01.SIVcpzCAM13       .......  208
CPZ.CM.05.SIVcpzEK505       .......  207
CPZ.CM.05.SIVcpzLB7         .......  207
CPZ.CM.05.SIVcpzMB66        .......  207
CPZ.CM.05.SIVcpzMT145       .......  206
CPZ.CM.98.CAM3              .......  208
CPZ.CM.98.CAM5              .......  214
CPZ.GA.88.GAB1              .......  206
CPZ.GA.88.GAB2              .......  206
CPZ.TZ.00.TAN1              .......  196
CPZ.TZ.2001.TAN2            .......  196
CPZ.TZ.2002.TAN3_1          .......  196
CPZ.US.85.CPZUS             .......  210

MAC.US.x.239                .......  263
H2A.DE.x.BEN                .......  258
H2A.GW.x.ALI                .......  263
H2A.SN.x.ST                 .......  256
H2B.CI.x.EHO                .......  239
H2B.GH.86.D205              .......  241
H2G.CI.92.ABT96             .......  257
H2U.FR.96.12034             .......  261
COL.CM.x.CGU1               .......  223
DEB.CM.99.CM40              .......  228
DEB.CM.99.CM5               .......  231
DEB.CM.SIVdeb04CMPF3061     .......  232
DEN.CD.x.CD1                .......  249
DRL.x.x.FAO                 DNLQKC*  268
GOR.CM.2004.SIVgorCP684con  .......  211
GOR.CM.2007.SIVgor2139_287  .......  211
GOR.CM.2007.SIVgorCP2135con .......  211
GRV.ET.x.GRI_677            .......  224
GSN.CM.99.CN166             .......  219
GSN.CM.99.CN71              .......  218
LST.CD.88.447               .......  206
LST.CD.88.485               .......  208
LST.CD.88.524               .......  207
LST.KE.x.lho7               .......  208
MAC.US.x.251_1A11           .......  264
MAC.US.x.251_BK28           .......  263
MAC.US.x.EMBL_3             .......  260
MND-1.GA.x.MNDGB1           .......  214
MND-2.CM.98.CM16            ......*  230
MND-2.GA.x.M14              ......*  232
MND-2.x.x.5440              ......*  250
MNE.US.x.MNE027             .......  264
MON.CM.99.L1                .......  225
MON.NG.x.NG1                .......  229
MUS-1.CM.01.1085            .......  218
MUS-1.CM.01.CM1239          .......  218
MUS-2.CM.01.CM1246          .......  218
MUS-2.CM.01.CM2500          .......  222
SAB.SN.x.SAB1C              .......  227
SMM.SL.92.SL92B             .......  267
SMM.US.x.H9                 .......  262
SMM.US.x.PGM53              .......  264
SMM.US.x.SIVsmH635F_L3      .......  264
STM.US.x.STM                .......  265
SUN.GA.98.L14               .......  208
SYK.KE.x.KE51               .......  229
SYK.KE.x.SYK173             .......  241
TAL.CM.00.266               .......  241
TAL.CM.01.8023              .......  241
TAN.UG.x.TAN1               .......  223
VER.DE.x.AGM3               .......  230
VER.KE.x.9063               .......  230
VER.KE.x.AGM155             .......  231
VER.KE.x.TYO1_patent        .......  230
WRC.CI.97.97CI14            .......  234
WRC.CI.98.98CI04            .......  234
WRC.GM.05.Pbt_05GM_X02      .......  231
```