VII

PLV Proteins

Contents

VII-1 Introduction **395**
VII-2 Sequences **396**
VII-3 Alignments **403**
VII-3.1 Gag 403
VII-3.2 Pol 407
VII-3.3 Vif 414
VII-3.4 Vpx 417
VII-3.5 Vpr 418
VII-3.6 Tat 419
VII-3.7 Rev 421
VII-3.8 Vpu 423
VII-3.9 Env 424
VII-3.10 Nef 431

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	van Beveren, C.P.	*(in) Weiss,RL, Teich,N, Varmus,H and Coffin,J(Eds); RNA TUMOR VIRUSES, SECOND EDITION,* **2**, Vol 2: 1102-1123; Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
H1A1.UG.85.U455	M62320	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Oram, J.D.	*ARHR* **6**(9):1073-1078 (1990)
H1B.US.90.WEAU160	U21135	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Wei, X.	*Nature* **422**(6929):307-312 (2003)
H1C.ET.86.ETH2220	U46016	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Salminen, M.O.	*ARHR* **12**(14):1329-1339 (1996)
H1D.CD.84.84ZR085	U88822	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*J Virol* **72**(7):5680-5698 (1998)
H1F1.BE.93.VI850	AF077336	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T.	*Virology* **269**(1):95-104 (2000)
H1G.SE.93.SE6165	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.SE7887	AF082394	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T.	*ARHR* **15**(3):293-297 (1999)
H1K.CM.96.MP535	AJ249239	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Triques, K.	*ARHR* **16**(2):139-151 (2000)
H101_AE.TH.90.CM240	U54771	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Carr, J.K.	*J Virol* **70**(9):5935-5943 (1996)

Name	Accession	Proteins	Author	Reference
H102_AG.NG.x.IBNG	L39106	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Howard, T.M.	*ARHR* **10**(12):1755-1757 (1994)
H103_AB.RU.97.KAL153_2	AF193276	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Liitsola, K.	*ARHR* **16**(11):1047-1053 (2000)
H104_cpx.CY.94.CY032	AF049337	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*J Virol* **72**(12):10234-10241 (1998)
H1N.CM.95.YBF30	AJ006022	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Simon, F.	*Nat Med* **4**(9):1032-1037 (1998)
H1O.BE.87.ANT70	L20587	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, M.	*J Virol* **68**(3):1586-1596 (1994)
H1O.CM.91.MVP5180	L20571	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gurtler, L.G.	*J Virol* **68**(3):1581-1585 (1994)
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.SIVcpzMB66	DQ373063	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.GA.88.GAB2	AF382828	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bibollet-Ruche, F.	*ARHR* **20**(12):1377-1381 (2004)
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)

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.TZ.00.TAN1	AF447763	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Santiago, M.L.	*J Virol* **77**(3):2233-2242 (2003)
CPZ.CM.05.SIVcpzLB7	DQ373064	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.US.85.CPZUS	AF103818	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	*Nature* **397**(6718):436-441 (1999)
CPZ.GA.88.GAB1	X52154	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Huet, T.	*Nature* **345**(6273):356-359 (1990)
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)
MUS_1.CM.01.1085	AY340700	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)
MUS_1.CM.01.CM1239	EF070330	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
MUS_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)
MAC.US.x.239	M33262	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kestler, H.	*Science* **248**(4959):1109-1112 (1990)
H2A.GW.x.ALI	AF082339	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Azevedo-Pereira	Unpublished

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)
H2A.SN.x.ST	M31113	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kumar, P.	*J Virol* **64**(2):890-901 (1990)
H2B.GH.86.D205	X61240	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	*Nature* **342**(6252):948-950 (1989)
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)
MAC.US.x.251_BK28	M19499	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hirsch, V.	*Cell* **49**(3):307-319 (1987)
MAC.US.x.251_1A11	M76764	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Marthas, M.L.	*J Med Primatol* **18**(3-4):311-9 (1989)
MAC.US.x.EMBL_3	Y00295	Gag, Pol, Vif, Tat, Rev, Env, Nef	Franchini, G.	*Nature* **328**(6130):539-543 (1987)
SMM.US.x.H9	M80194	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	*J Virol* **66**(1):414-419 (1992)
SMM.US.x.PGM53	AF077017	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Novembre, F.J.	*J Virol* **72**(11):8841-8851 (1998)
SMM.US.x.SIVsmH635F_L3	DQ201172	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kuwata, T.	*J Virol* **80**(3):1463-75 (2006)
SMM.SL.92.SL92B	AF334679	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Chen, Z.	*J Virol* **70**(6):3617-3627 (1996)

Name	Accession	Proteins	Author	Reference
STM.US.x.STM	M83293	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Novembre, F.J.	*Virology* **186**(2):783-787 (1992)
MNE.US.x.MNE027	U79412	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kimata, J.T.	*J Virol* **72**(1):245-256 (1998)
GRV.ET.x.GRI_677	M66437	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Fomsgaard, A.	*Virology* **182**(1):397-402 (1991)
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
VER.KE.x.9063	L40990	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	*J Virol* **69**(2):955-967 (1995)
VER.KE.x.AGM155	M29975	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Johnson, P.R.	*J Virol* **64**(3):1086-1092 (1990)
VER.DE.x.AGM3	M30931	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Baier, M.	*Virology* **176**(1):216-221 (1990)
TAN.UG.x.TAN1	U58991	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Soares, M.A.	*Virology* **228**(2):394-399 (1997)
SAB.SN.x.SAB1C	U04005	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Jin, M.J.	*EMBO J* **13**(12):2935-2947 (1994)
OLC.CI.97.97CI12	FM165200	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
DRL.x.x.FAO	AY159321	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	*J Virol* **77**(8):4867-4880 (2003)
RCM.NG.x.NG411	AF349680	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	*J Virol* **75**(24):12014-12027 (2001)
RCM.GA.x.GAB1	AF382829	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
MND_1.GA.x.MNDGB1	M27470	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Tsujimoto, H.	*Nature* **341**(6242):539-541 (1989)
MND_2.x.x.5440	AY159322	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	*J Virol* **77**(8):4867-4880 (2003)

Name	Accession	Proteins	Author	Reference
MND_2.CM.98.CM16	AF367411	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Takehisa, J.	*ARHR* **17**(12):1143-1154 (2001)
MND_2.GA.x.M14	AF328295	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Souquiere, S.	*J Virol* **75**(15):7086-7096 (2001)
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)
TAL.CM.01.8023	AM182197	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*Virology* **349**(1):55-65 (2006)
TAL.CM.00.266	AY655744	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*Virology* **349**(1):55-65 (2006)
MON.CM.99.L1	AY340701	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)
SUN.GA.98.L14	AF131870	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	*J Virol* **73**(9):7734-7744 (1999)
WRC.CI.97.97CI14	AM745105	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
WRC.CI.98.98CI04	AM713177	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
WRC.GM.05.Pbt_05GM_X02	AM937062	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Locatelli, S.	*Virology* **376**(1):90-100 (2008)
LST.CD.88.524	AF188116	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	*J Virol* **74**(8):3892-3898 (2000)
LST.CD.88.485	AF188115	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	*J Virol* **74**(8):3892-3898 (2000)
LST.CD.88.447	AF188114	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	*J Virol* **74**(8):3892-3898 (2000)
LST.KE.x.lho7	AF075269	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	*J Virol* **73**(2):1036-1045 (1999)

Name	Accession	Proteins	Author	Reference
SYK.KE.x.SYK173	L06042	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	*J Virol* **67**(3):1517-1528 (1993)
SYK.KE.x.KE51	AY523867	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
COL.CM.x.CGU1	AF301156	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	*J Virol* **75**(2):857-866 (2001)
MON.NG.x.NG1	AJ549283	Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Barlow, K.L.	*J Virol* **77**(12):6879-6888 (2003)

```
                          membrane binding                                                                                                     phosphorylation site
                          Gag start, p17 start   nuclear localization                                                                     nuclear localization                          p17 end  p24 start
H1B.FR.83.HXB2          MGARASV.LSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQ.......................NKSKKKAQQAAADTG.HSNQVSQNYPIVQNIQGQMVH  144
H1A1.UG.85.U455         -------.---KK--S----------N---R---L--------K-TL-------A---Q--------A----T-----------V--------DV-------N----M-.......................--N-QRT-----N--.....S---------A---P--  140
H1B.US.90.WEAU160       -------.------------------N-----------------------------------------------------V------K--V---------------.......................--------------E.NR-----------L------  144
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------  142
H1D.CD.84.84ZR085       -------.----K--A----------------------------L----------K--ME---A-K-------I----------K---------------.......................-N---R-----G--.N-S-----------L------  144
H1F1.BE.93.VI850        ------I.----K--E----Q-------R--M--LI----------LD----------QK-IR-------------K--F---V--Y---AGVT-------L----.......................---QQ-T------K-.....-----------L------  140
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------  144
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------  144
H1J.SE.93.SE7887        ------I.----K--D-------------RI--L-------D---L------SAK--Q---V----A----TQ-IK---------------------M---E----I-.......................--N-QQ--K-ET-KK.DNS----------L---P--  144
H1K.CM.96.MP535         -------.----K--A------------------L-----------L-------T------IT-I---I------IK----I-V--F---K--V---------L----.......................---QR-T--.E-A....DKG---------L------  140
H101_AE.TH.90.CM240     -------.----K--A----------R---R---L-----------L--SF---A---Q--IE---ST-K--L---K--F------W-------V-----------V-.......................---QR-T-----G--.S-SK----------A----A-  144
H102_AG.NG.x.IBNG       -------.----K--A--------------R---L-----------L-------A---Q-LME---SA-R------K--F--I---W------D---------M--I-.......................----Q-V--T--A--....SS---------AK---T-  141
H103_AB.RU.97.KAL153_2  -------.----K--A-----------E--RI--L-----------L--S--------Q---E----T-K------K-----------------------------I-.......................----Q-T----TG--.S-SK----------A----T-  144
H104_cpx.CY.94.CY032    -------.----K--A--R-----------R---L-----------L-------A---Q-LME---ST-K------------IT--W------DVQ-----------I-.......................S---Q-T-----AA-.G-SN----------A------  144
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.......................QHKPEPKNPE-GAAAATDSNI-R---L--TA------  146
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------  142
H1O.CM.91.MVP5180       -------.-T-SK--A--R------S--A-R---L---------Y-C-------A--TE-L-Q--E-A-K----D-K--W-AI-V-W---N-FD-R--QQ-IQ-LK-VMA..........................SRKSAE--KEETSPR-T-------T-A------  142
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-IAR-  160
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------  149
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------  147
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------  142
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---L--  144
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----I-  146
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----I-  143
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---LT-  134
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..TVQKNNSTATSSGQRQNAGEKEETVPPSGNTGNTGRATETPSG-RL--VITDA--VAR-  167
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......................QKQQ-EKE---AAARDNDSG-R-----T-A---P--  144
CPZ.US.85.CPZUS         -------.-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.................VQLQ-QQEEKE-QQQEASG.--IG-S---VI--A------  146
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--EQLKRHHGE...............QQSKTESNSGSREGG-SQGASA-AGI-G---L---A------  152
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-KQQVRIRCHLEE...................T-G-E-QNKDPPGAAGGQAV------VIR-A---YQ-  150
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-KKQVQIRCHLVG...................T-E-ETQNKDPPGAAGGQAV--------R-A---FQ-  150
MUS_1.CM.01.1085        ----H-AM-T-NK---Y--V----K-----LI---I------D--GLSDS---SRD--QK-IEVIL-LEPA---S-K--FGI-SV-W-I-AEVHVE--E--KKQ-RARCHLEE..........................KKGEKESQAQAA.-A---VIR-A---FQ-  141
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-RIRCHLAD...........................KQGEEKTAQ--T-A----IR-P---FQ-  141
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-RIRCHLAE.........................-KEDSKG-S-AKATTSG----VIRTA---YQ-  145
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-RMRCHLGE........................--E--EQR-QPSGAGAS-----VIRTA--TYQ-  146
                          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-VQRHLVVE.....................TGTTETMPKTSRPTAP-SGRGG---VQ-IG.-NY--  146
H2A.GW.x.ALI            ----N--.-R-RKA-EL-R-----------Q-------AN--D--GLAES---SK---QR--KV-E-LVP----N-K--F---CVVW---AEEKV---EG-K.Q-IQRHLAAE.....................IETAEKMPSTSRPTAPPSEQGG-F-VQ-VA.-NYT-  146
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.-LAQRHLVAE.....................TGTAEKMPNTSRPTAPPSGKRG---VQ-AG.-NY--  146
H2A.SN.x.ST             ----N--.-R-KKA-EL-------------R-------AN--D--GLAES---SK---QK--TV-D-LVP----N-K--F---CVIW-I-AEEKA---E--K.QKVQRHLVAE.....................TKTTEKMPSTSRPTAPPSGNGG-F-VQ-VA.-NYT-  146
H2B.GH.86.D205          ----G--.---KKT-EL--V----------M---V---VN--D--GLAES---SK---QK--KV-A-LVP----N-K--F-I-CVIF-L-AEEKV---E--K.--AQRHLAAD..........................TEKMP-TNKPTAPPSGG.----VQ-LA.-NY--  141
H2B.CI.x.EHO            ----G--.---KKT-EL--V--------R-M-------VN-----GLAESR-GSK----K-RKV-G-LVP----N-K------CVIF-L-AEEKV---E--K.--AQRHLAAD..........................TEKMP-MSKPSKPTSRL..A--VQ-IA.-NYS-  141
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.QVVQRHLVVE.....................TGTTEKVP-TSRPIAPPSGRGG---VQ-VG.-NY--  146
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-VQRHLAVE.....................TETAEKMP-TSRPTXPPSGGRG---VQ-IG.-NYA-  145
MAC.US.x.251_BK28       ----N--.---KKA-EL-------------M---V---AN--D--GLAES---NK---QK--SV-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.Q-VQRHLVVE.....................TGTAETMPKTSRPTAP-SGRGG---VQ-IG.-NY--  146
MAC.US.x.251_1A11       ----N--.---KKA-EL-------N-----M---V---AN--D--GLAES---NK---QK--SV-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.Q-VQRHLVVE.....................TGTAEIMPKTSRPTAP-SGRGG---VQ-IG.-NY--  146
MAC.US.x.EMBL_3         ----NA-.---KKA-EL-------------M---V---AN--D--GLAES---NK---QK--SV-A-LVP----N-K------CVIW-I-AEEKV-H-E--K.Q-VQRHLVVE.....................TGTAETMPKTSRPTAP-SGRGG---VQ-IG.-NY--  146
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-VQRHLVVE.....................TGTADKMP-TSRPTAPPSGRGG---VQ-VG.-NYT-  146
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-VQRHLVVE.....................TGTADKMPITSRPTAPPSGRGG---VQ-IG.-NYT-  146
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-VQRHLVVE.....................TGTADKMP-TSRPTAPPSGRGG---VQ-VG.-NY--  146
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.-TVQSHLVVE.....................SGTAEKLP-QSRPTAPPSGG..---VQ-VG.NNY--  144
STM.US.x.STM            ----S--.---KKA-EL--V----------M---V---AN--D--GLAES---SK---QK-ITV-E-LVP----N-K--F---CVIW-I-AEEKV-H-E--K.QVVKRHLVVE.....................TGTANKMP-TSRPTAPPSGRGG---VQ-VG.-NY--  146
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-VQRHLVVE.....................TGTAETMPKTSRPTAP-SGKRG---VQ-IG.-NYT-  146
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--VVTVKQHYHLVD...............KNEKAA.......KKKNE-TAPPGGE-R---V-NQN.NAW--  146
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--VATVRQHCHLVE...............KEKSATETSSGQKKNDKGIAAPPGG.---F-AQ-QG.NAW--  152
VER.KE.x.9063           ---AT-A.-K-RQ--QF-------N-----QI--LI--GK-M---GLHDR----ED--KK-IEV-F-LEP----G-K--F-L-CV---I-REQKV---E--VATVRQCCHLVE...............KEKRATEPSSGEKKSNRE-TAPPGG.---F-AQ-QG.NAW--  152
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---AIVRQCCHLVD...............KEKTAVTPPGGQ-KNNTGGTATPGG.---F-AQ-QG.NAW--  152
VER.DE.x.AGM3           ---AT-A.-NRRQ--KF-H-----T-----QI--LI--GK-M---GLHER---SE---KK-IEV-Y-LEP----G-K--F-L-CV-F---KDK-V---E--VAIVRQCCHLVEKERNAERNT............TETSSG-KKNDKGVTVPPGG---F-AQ-QG.NAWI-  156
TAN.UG.x.TAN1           ---GH-A.---RN--TF-------N-----Q---LI--GKKM---GLHEK----K---QK-IEV-S-LEP----G-K--F-LCCVIW-I-AEQKV---E--VVLVKQRCHLVE...............KEKTAAAPSGGQ--NYNTAATP-GR.HG---V--QN.N-W--  152
SAB.SN.x.SAB1C          ---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--VPAEMTESATATSSGQTK...........ELQAKKKNEPTVTPSGG-R-----SVN.N-W--  157
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...................................KKEEQLAM--IR-A--IFE-  128
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-VKQHCHLVD.......................-NENA-SKNENGETAT-SGR-R----QVVN.Q-A--  147
RCM.NG.x.NG411          ------I.---KK--A--QV----------M---L---C------GLSDT---SA--IQK-I-VIL-LVP----G-K--F-LCCV-W---KEVKV---E--VAHVKKQCHLVDKDENAGEQEKGAT......................VT-SGQRG-----TIN.Q-PE-  146
RCM.GA.x.GAB1           ------L.---KK--A--SV----------M---L---CKK-NK-GLSDH----AT--EK---V-L-LVP----G-K--F-LCCV-W---KEVKV---E--VA-VK-CCHLVEKAENTTEKEKGAT......................APPSGQRG----ITIN.Q-PE-  146
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-LKVITRKEEKQEDE...............................-K-F-VQRDAA--YQY  133
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-VKQRYHLVVERENAASEEEKGA.....................TATPAVR-K----QVIN.QTP--  147
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-VKQKLHLVAEKENAASEKEQR....................AIVTPSGR-K----QIIN.QTP--  147
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-QCFHLAAKGESAASEKEEK....................ATATPSGR-K----QIIN.QTP--  147
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--KKAVYRRIKQVEQEEELEMALIQRQKEKEVEQKQQ-QQE-RQQEPQAAPQTTP-V----LRQG.Q-F--  169
DEB.CM.99.CM5           ---GR-AP-T-AA---F-------K-----LI--L-------D--GLAGA----K---QE--TR---LEAS---G-K---GI-CVVWAC-ANLSVE--E--KKAVARRIKQAE....QEEELEMALIQKKAEKELE--QKT----ADTQQKTNQPA-F--LRQG.Q-Y--  165
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---REVRKRIGAKKKNIVIGNADQGDDQPPQGAAGGGSSAE-G-S-GGSSFRDPN-------RTA---YQ-  170
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-KAEVKKRCHPV.................EKKEESKGTTE-G-VRSAIVT-SGH-G----IR-A-Q-YQ-  153
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-KAEVRKRCHLGEKKETTEEEK.....................VRKATVT-SGQTG----IR-A-Q-YQ-  149
MON.CM.99.L1            ----H-AM---TK--KY--V----R-----LI------AK--D--GLSDS----QD--KK--EVIL-LQP----SIK--FGIASV---I-AG---E--EQ-KQQVKIRCHLA.....................GEQGEQKA---AAPPTGGVP-G---V-RTQG-GFQ-  148
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-PAKNEEAQA...............................---F-VQREG.QNYI-  131
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.QGPQW  127
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-VKDTKQKEKP............................---M-VLRTA.QGPQW  131
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---VRPREEKA...............................-V-M-VLRGP.QGPQW  129
LST.CD.88.524           --SGN--.--RQIEKDFCNV-----S--T-QKR-VD--TK--D--GLGAN----A---KK---VRW-LYP---KN-KA-VG--CVIC-C-LG-RVN--Q--IN-VKI--NK...................................PVPE-K-F-LIREN.QRW--  129
LST.CD.88.485           --SGN--.--RQIEGEFCS------S--T-QKR-----TK--D--GLGAH----AD--KR---VCW-LYS---KN-KA-VG--CV-C-C-LG-K-AN-A--MQ--KI-PTQE..................................DRQ-K-G---LIREN.QRW--  131
LST.CD.88.447           --SGN--.--RQIEGEFCS-----DS--T-QKR-L---T---D--GLGAH----AD--KK---VCW-LYS---KN-KA-VG--CV-C-C-LG-KVA--Q--MQ--KI-PTQE..................................ERQPK-G---L-REN.QRW--  131
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-PAPEAAGKK..............................Q-TGG---LIREN.QRW--  134
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-ACNWKDDPPATSGGQSENSSQNM.ASET-SGQKVVQQEKQKAATPPPRG---LLR-P-N-WI-  169
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-ACNWEDEETVTSSGQKENSNDTA....T-SGREGKMQLPAAMP-SGG-G---LIR-P-N-WI-  166
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--AYKKQAMIEMASKEEEKAK.......................KEAEKLDM---TGP.QGP--  143
```

PLV Proteins

```
                                                                                         CyPA binding                                                      major homology region
H1B.FR.83.HXB2              QAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVH.AGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTN.....NPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQ 308
H1A1.UG.85.U455             --L-------------D------------------------M---V------------D-----------L----.----P------------------V--------G.....-------D--R----------------V--------------------F------- 304
H1B.US.90.WEAU160           --L-----------I-------------------------------------------------------L----.----------------------------A----.....---------------------------V---------------------------- 308
H1C.ET.86.ETH2220           -P--A-----------------------T-----------------------------D-----------L----.---V------D-----------------A---G.....---V---D-------------------V-----K--------------F------- 306
H1D.CD.84.84ZR085           --L-----------I-----N-------------------------------------------------L---Q.---V----------------------------S.....---------------------------V---------------------------- 308
H1F1.BE.93.VI850            -SL-----------I----------------------T--------------------D-----------L----.---AP-----------------------Q---G.....---V---D-------------------V-----K--------------F-V----- 304
H1G.SE.93.SE6165            ---T-------------------------------------L----------------D-----------M--QQ.---FP---I-------------S-----T---G.....---------------------------V--------------------F-C----- 308
H1H.CF.90.056               -----------------------------------------A----------------D----------------.----P-----------------------A---G.....--A----D-------------------V-----K--------------F------- 308
H1J.SE.93.SE7887            --L-----------I-------------------------------I-----------D----------------.--------V-----------------------G.....---------------------------V--------------------F-A----- 308
H1K.CM.96.MP535             --L-----------I-------------T-----------------------------D---D-------L----.----P-----------------------A---S.....---V-----------------------V--------------------F------- 304
H101_AE.TH.90.CM240         -PL---------------G-N--------------------M---I------------------P----------.----P----------------------------.....-------D-------------------V---------------------------- 308
H102_AG.NG.x.IBNG           -SM-----------I---G----------------------M---I------------D----------------.----P---------------------------S.....--------------V------------V--------------------F------- 305
H103_AB.RU.97.KAL153_2      -SM-----------I--------------------------M---I------------D-----------L--AQ.---FP--------------SS-----------S.....-------D-------------------V--------------------F------- 308
H104_cpx.CY.94.CY032        -S------------I--------------------------M---I------------D-------D---T----.----P---------------------------S.....---V-------------------T---I--------------------F-C----- 308
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---------------------- 310
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-RP.....-Q-----D--RK--V-----M-K----V-----K---------------------- 307
H1O.CM.91.MVP5180           -------------A------N--I----M------V-Y-I-----AI----G-L-V---V----------T--PA.M--LP---I---T----------Q---I-T-RG.....ANS----D--RK--V-----M-K----V---------------------------- 307
CPZ.CD.90.ANT               -PLT---------C----N-N----------------H-------A--D--G---V---V----------L--T-.---VQA--L---T----------V---MQ--STP..QQ-GGV---D-------M----V--X---V---E-K-----------------I---- 327
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---- 313
CPZ.CM.05.SIVcpzEK505       -PL-----------------N-------M------------------------------V----------G----.M---P---V---------------A--VA---A.....---V---D--R---V------------A-----K-----T---------------- 311
CPZ.CM.05.SIVcpzMB66        -P--------------------------M------------------------------V----------L----.---VP------------------I---V----S.....-------D--------------K--C-V-----K-----S---------------- 306
CPZ.CM.98.CAM3              -PM-----------I---N-N-------M----------V-----AI----G---V---V----------T----.---V----L--A------------V---A---A.....---V------R--VV-----V----C-V-----K---------------------- 308
CPZ.GA.88.GAB2              --M-----S----A--------------M----------V-----AI----G---V---V----------L--L-.---V-----------------------V----S.....---V------R--VV-----V----C-V-----K----------------V----- 310
CPZ.CM.98.CAM5              --M-----------I---N-N-------M----------V-----A-----G---V---V----------L----.--------L--A---------------VA---A.....----------R--VV-----V-K--C-V-----K---------------------- 307
CPZ.CM.05.SIVcpzMT145       NP-----------A------N-------M-----------------I--------I---V--D-------T--T-.----------------------S-E--V----A.....-------D--R---V-------K----V-----K---------------------- 298
CPZ.TZ.00.TAN1              -P----------R-I---G-N----------------Y---S---A--E----------V----------A--A-.---QQA-ML---T-A------------VL---TPQA..QGGV---D-----------L-------V-----K-----------------I---- 334
CPZ.CM.05.SIVcpzLB7         --L-------------------------A--A-----------------------I---V-----Q----L--I-.M----------------------V--------S.....---------------M----V------V-----K---------------------- 308
CPZ.US.85.CPZUS             --M----------A------N-------M----------V-----AI----G---V---A------D*--T--L-.---------------------------V----A.....-------D--R--VV-----V-K--C-VG----K---------------------- 309
CPZ.GA.88.GAB1              -----------------------------------L---V-----A-----G---V---V----------L--T-.--------L---------------------T-A.....-------DV-R--V------V----C-V---------------------------- 316
GSN.CM.99.CN166             --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-RQ....-Q-VN-SD--RK------QRC-K--N-VN----K-------KE-----F-C----- 315
GSN.CM.99.CN71              --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-RQ....-N-VN-AD--RK------QRC-K--N-VN--G-K-------KE-----F-C----- 315
MUS_1.CM.01.1085            --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-RA....-N--N-A----K-V-M--QRC-K--N-VN----K-------K------F-C----- 306
MUS_1.CM.01.CM1239          -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-RQ....-D--N-AD---K---M--QRC-K--N-V-----K-------K--------C----- 306
MUS_2.CM.01.CM1246          -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---RQ....QD-VN--N--RK------QRC-K--N-VN----K-------K------F-C----- 310
MUS_2.CM.01.CM2500          -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---RQ....QD-VN--D--RK------QRC-K--N-VN----K---R---K------F-C----- 311

MAC.US.x.239                LPL----------LI---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIRDI------D--LQ--QP..A-.QQ--L---S---------SVD---Q--YRQ....QN-----N--R---Q---Q-C----N--N---VK-------QS-------S----- 309
H2A.GW.x.ALI                VPL-----D----L----K-GA--V-G-Q-----C--Y-I-Q---C--D------IIR-I------D--VA--IP..--LPA--L--------------VE---Q--FRP....RN-V---N--R---QI--Q-C----N--N----K-------QS-------S----- 310
H2A.DE.x.BEN                VPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--D------IIR-I------D--SQ--IP..--LPA--L-D------------VD---Q--YRP....QN-V---N--R---QI--Q-C--K-N--N----K-------QS-------S----- 310
H2A.SN.x.ST                 VPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--D------IIR-I------D--AQ--IP..--LPA--L--------------VE---Q--FRP....QN-V---S--R---QI--Q-C----N--N----K-------QS-------S----- 310
H2B.GH.86.D205              LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--QQ--SP..--MPA--L-D------------VE---Q--YRA....QN-V---N--R---Q---Q-C----N--N----K-------QS-------S----- 305
H2B.CI.x.EHO                LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--QQ--SP..--MPA--L--------------VE---Q--YRP....QN-V---N--R---Q---Q-C----N--N----K-------QS-------S----- 305
H2G.CI.92.ABT96             LPL----------L----K-GA--V-G-Q-----CL-Y-I-Q---C--E------IIR-I------X--LQ--QQ..--PPA--X---T----------IE---Q-THRQ....QN-----NX-R---X---Q-C---XN--N---VK-------QS-------SX---- 310
H2U.FR.96.12034             LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I--D---D--TQ--NQ..--PPA--L--------------VE---Q--YRP....QN-----N--R---Q---Q-C----N--N----K-------QS-------S----- 309
MAC.US.x.251_BK28           LPL----------LI---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIRDI------D--LQ--QP..A-.QQ--L---S---------SVD---Q--YRQ....QN-----N--R---Q---Q-C----N--N---VK-------QS-------S----- 309
MAC.US.x.251_1A11           LPL----------LI---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIRDI------D--LQ--QP..A-.QQ--L---S---------SVD---Q--YRQ....QN-----N--R---Q---Q-C----N--N---VK-------QS-------S----- 309
MAC.US.x.EMBL_3             LPL----------LI---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIRDI------D--LQ--QP..A-.QQ--L---S---------SVD---Q--YRQ....QN-----N--R---Q-R-Q-C----N-IN---VK-R-----QS-------S----- 309
SMM.US.x.H9                 LPL----------L----K-GA--V-G-Q-----C--X-I-Q---C--E------IIR-I------D--LQ-XQP..---PA--L---X----------XD---Q--XRQ....QN-----N--X---Q---Q-C----N--N---VK-----S-QS-------S----- 310
SMM.US.x.PGM53              LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--QP..---PA--L-D------------VE---Q--YRQ....QN-----N--R---Q---Q-C----N--N---VK-------QS-------S----- 310
SMM.US.x.SIVsmH635F_L3      LPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--QP..--LPA--L--------------VE---Q--YRQ....QN-----N--R---Q---Q-C----N--N---VK-------QS-------S----- 310
SMM.SL.92.SL92B             TPL----------L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--LQ--RGQQPAQPA-GL---S----------PS---E--YRA....QN-V---D--R---Q---Q-C----N--N---VK-------QS-------S----- 310
STM.US.x.STM                LPL----------L----K-GA--V-G-Q-----C--Y-I-QL--C--E------II--I------D--MQ--QP.P--LPA--L---S---------SPE---Q--YRQ....QN-----N--R---Q---Q-C----N-VN---VK-----T-QS-------S----- 311
MNE.US.x.MNE027             LPL----------LI---K-GA--V-G-Q-----CI-Y-I-Q---C--E------IIR-I------D--LQ--QQ..A-.QQ--L---S----------VD---Q--YRQ....QN-----N--R---Q---Q-C-K--N--N------------QS-------S----- 309
GRV.ET.x.GRI_677            -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----- 311
VER.KE.x.TYO1_patent        VPL----------A----K-GA-IV---Q-----C--Y-I-Q---VL-D--G-L-IV--I------Q--VT--LP.---LPA--L-D-----------SV---LE-IYTA....--RVD--A--R-------Q-C-K--N-V-V-----------K--------AI---- 317
VER.KE.x.9063               VPL----------A----K-GA-IV---Q-----C--Y-I-Q---VL-D--G-L-IV--I------Q--VT--PP.---LP---L-D------------V---LE-IYTA....--RVD--A--R-------Q-C-K--N-V-V-----------K--------AI---- 317
VER.KE.x.AGM155             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-IYTA....--RVD--A--R-------Q-C-K--N-V-V-----------K--------AI---- 317
VER.DE.x.AGM3               VPL----------A----K-GA-IV---Q-----C--Y-I-Q---VL-D--G-L-IV--I------Q--IA--PP.---LPA--L-D------------V---LE-IYTA....--RVD--A--R-------Q-C-K--N-V-V---------A-K--------AI---- 321
TAN.UG.x.TAN1               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----- 317
SAB.SN.x.SAB1C              -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-RA....QNAVN--N---G------Q-C-K--N-VN----K-------K--------A----- 322
OLC.CI.97.97CI12            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-VNAG.......-D--AL----V-I--Q-L-K----VTV------A----S------F-A----- 290
DRL.x.x.FAO                 --------------I---K--A--V---L-----CISY-I-Q---AI-D--G-L-IV-DI------D---R--QV..--LPQ-VL-N-S---------SIE---E-T-RA....QDS-N--A--RQ-VV---QRC-T--N-VN---VK-------K--------A----H 311
RCM.NG.x.NG411              -P----------------K-GA--V---------CI-Y-V-Q---AI-E--G-L-IV--V--D---D--LR--NP..--LPA--L---T---------NIA---A-T-RA....-N--A--N--RN--V---Q-C-K--N-VN----K-------K--------A----- 310
RCM.GA.x.GAB1               NP----------------K--A--A---------CI-Y-I-Q---AI-E--G-L-IV--V------D--AR---P..---PA--L---T---------SIA---A-T-RA....-N--N--NL-RN---V--Q-W-K--N-VN----K-----S-K--------A----- 310
MND_1.GA.x.MNDGB1           TP----IIQT---T----KWK-----L----T---ISH---I---A--D--G---V--DV---Q-----LT--QQ.QPAQPG-GL-T-S----------VE--LA--NMQ....QNA-N--T---S-----M-RL-KSHC-I--T-V------A-K-------NVM---- 298
MND_2.x.x.5440              -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-RQ....QDQVN--G---Q--V---Q-C-S--N-VN----K-------K--------A----R 311
MND_2.CM.98.CM16            -G-----------CI---K----IV---I-----CL-Y---G---AI-D--G-L-IV-DV------D-NLR--QV..--MPQ-VL-N-T---------SIE---E-T-RE....QDAVN--G---Q--V---Q-C-S--N-VN----K-------K--------A----R 311
MND_2.GA.x.M14              -G-----------CI---K----IV---I-----CL-Y---G---AI-E--G-L-IV-DV------D--LR--QV..--LPQ-VL-N-T---------SIE---E-T-RQ....QEQVN--A---Q--V---Q-C-S--N-VN----K-------K--------A----R 311
DEB.CM.99.CM40              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-------------- 336
DEB.CM.99.CM5               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-------------- 332
DEN.CD.x.CD1                --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---RQ....QG--N--D-----V-M--Q-V-K--Q-V-----K-----S-K--------A----- 336
TAL.CM.01.8023              --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-RQ....QN-VD--S-----V-I--QRC-K--N-VN---VK---R---KE------RC----- 317
TAL.CM.00.266               --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-RA....QN-VD--A-----V-M--QRC-K--N-VN---VK---R---K-------RC----- 313
MON.CM.99.L1                --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-RA....-N-VQ-SD--RK-V----QRC-K--N-VN----K-------K--------C----- 313
SUN.GA.98.L14               -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-IYGL..NGATRVA-QD---G-V-E-MERV--LHQT--VVE------------T---F-A----S 298
WRC.CI.97.97CI14            DPL----IQT---HI-DRG-GA-SV-F-A--TDR-LSH-I--L--NI-D------II-DIV---IQ----T--QA.----------Q-T------V-ASVEDELA-LAQN.......--VD--S-----VLR--Q-V-QI-V-V-----K-----D-KA-------V----- 291
WRC.CI.98.98CI04            DPL----IQT---HI-DRG-GA-CV-F----TDR-LAH-I--L--NI-D------II--IV---IQD---T--QP.----------Q-T------V-ASVEDELA-LAQN.....---D--AV----VLR--Q-V-QI-V-V-----K-----D-KT-------V----- 295
WRC.GM.05.Pbt_05GM_X02      EPL----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-LAQQ.....---DI-A-------R--Q-V-QI-V-------K-----D-KS-----F-V----- 293
LST.CD.88.524               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-AD.....T--D--K---E-V-QAME-V--IHQ-V-VM--K-------K--A---F-A----G 293
LST.CD.88.485               TPL----IQT---I--DRGWK--TVA-----T-K-L-D---V---AI-D--V---II-DH-V--G-----T--QQ.QPAQPG-GL-T-NAT-V--V---VE--LA-T-AD.....T-VD--K---E-V-QAME-V--IHQ-V-VM--K-------KE-A---F-A----G 295
LST.CD.88.447               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-AD.....T-VD--K---E-V-QAME-V--IHQ-V-VI--K-------KE-A---F-A----G 295
LST.KE.x.lho7               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-AD.....T-VD--K---E-V-QAME-V--IHQ-V-VM--K-------KE-A---F-A----G 298
SYK.KE.x.SYK173             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-RQ....-N-VQ-----R-------Q-C-QV-N-V-----K-----I-K-------HC----- 334
SYK.KE.x.KE51               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-RQ....-N--Q--Q--RQ------Q-C-QV-N-VN----------T-K-------HC----- 331
COL.CM.x.CGU1               -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----P 300
```

```
                                                                    p24 end p2 start      p2 end p7 start                         Zn motif                      Zn motif
H1B.FR.83.HXB2          ASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSAT.....IMMQRGNFRNQR...............KIVKCFNCGKEGHTARNCRAPRKK.........GCWKCGKEGHQMKDCT..  427
H1A1.UG.85.U455         -T-D------------------S--R----G---------------------------QQTS.....---------GP...............RRI----------L-K---------.........----------------..  421
H1B.US.90.WEAU160       -T---------------------------G---D-----------------------T-.....M--------SP-...............-TI---------I---------.........-----Q----------..  427
H1C.ET.86.ETH2220       -T-D------D--------------R----G-S--------------A------------N-TT.........----KS--KGPK...............RAI---------L---------.........----------------..  424
H1D.CD.84.84ZR085       -----------------------------G---------------S-----------A-SASAA.....----KS--KGT-...............-------------I---------.........----------------..  428
H1F1.BE.93.VI850        ---D--G---D-----------------TG---------------S-----------ANSA.........----KS--KG--...............RV----------I---------.........-----R----------..  421
H1G.SE.93.SE6165        ---D--G---D--------------R---QG-S------------S-----------ASGA-AA.....-----S--KGP-...............RTI---------L---------.........----------------..  428
H1H.CF.90.056           -T-D---------------------R---QG-SI-----------S---------------TN-A.......----K---KG--...............-F----------I---------.........-----R----------..  428
H1J.SE.93.SE7887        -T-D------D-----------------SG---------------------------------TN.........----------DHK...............R---------Q--I-K--------.........----------------..  426
H1K.CM.96.MP535         -T--------D------------------G-S-------------S----I----------PV.....V---K---KGH-...............-------------I---------.........----------------..  422
H101_AE.TH.90.CM240     -T---------------------S-----TG--------------S-----------HAQHAT.........---------KG-...............-RI------R---L---------Q.........----------------..  425
H102_AG.NG.x.IBNG       -T---------------------S--R--TG------------------------------QQAN.....V--------G-...............RTI---------L----K----R.........----------------..  422
H103_AB.RU.97.KAL153_2  -T-D---------------------R---SG--------V--------------------Q-AN.........----KS---GP...............-RI---------L---------.........----------------..  425
H104_cpx.CY.94.CY032    -T---------------------S-----TG--------------S-----------AS-A-AA.......-----KSK-KG--...............RTI---------L---------.........----------------..  428
H1N.CM.95.YBF30         -T---------------------QL------G---------------A-------------QQPT-S...VFA-----KGI-...............-PI---------L----K---RG.........------Q--------KNE  432
H1O.BE.87.ANT70         -T---------------------Q---S---G-------V--------T----------ATAQQDLKGGYTAVF----QNPIRK...............GTI---------I---------.........------Q--------RN.  432
H1O.CM.91.MVP5180       -T----------------S-----Q-------E-------V--------T---KI-----ASAQQDLKGGYTAVF----QNP-RK...............GPI---------I-K-------R.........------Q--------KN.  432
CPZ.CD.90.ANT           ---P--A-------I--------H------TG-S----L---------A----------ASAN-AQGT...AVFL----GNRGGK...............RPL--------------K---R-.........---R--Q----L-N-PAT  451
CPZ.CM.01.SIVcpzCAM13   ------T---D-------------Q--------G-S---------------A----------TM-QSQ-RT...D-FF-K-PGATP-...............RKI--Y-------L----K---R-.........------Q--------S..  434
CPZ.CM.05.SIVcpzEK505   -T---------------------N--R-----G-S---------------A----------T-AQTATS.....VF------KGI-...............-TI---------L----K-----.........------Q--------RS.  431
CPZ.CM.05.SIVcpzMB66    -T--------D-------------S--R-----G----GDEP-F------S------------AQH-NDA.....KR-FKGPK...............R------------I----K---R-.........------Q-----RN--N.  423
CPZ.CM.98.CAM3          -T--------D-------------N--R-----G----------------A-------------LQ-PT.....GVFL----NGKPT...............RKI---------L--------R-.........------Q------N-PA.  428
CPZ.GA.88.GAB2          ---D--------------------Q--R-----G----------------A-------------IQQTS.....S-F---AGG-TPP...............RR----------I----K---R-.........---------------QN.  430
CPZ.CM.98.CAM5          -T--------D-------------S--R-----G----------------A--------------Q--S.....NVF-----NGRVN...............-KI--Y-------V--------R-.........------Q---N--E--TG  428
CPZ.CM.05.SIVcpzMT145   ---D------D---I---------N--R-----G----------------S--------------NKT......E-F----GQNRPPN...............-KI---------L--------R-.........------Q--------NQN  419
CPZ.TZ.00.TAN1          ---P--T----------------H------QG------L---------S---KI-----ASA-AGG.....VN-L-G-KRPPLK...............KGQLQ------V---------------.........---R--Q---------TR  456
CPZ.CM.05.SIVcpzLB7     -T--------D-------------S--R-----G-S---------------A-------------ANAA.....TV--------KGPK...............RVI---------I----K---R-.........------Q---R-Q--NM.  428
CPZ.US.85.CPZUS         --P---T------------------R-----G------V--------A----------C-MK-PS.....SVFL-K--AGKPG...............RKI---------L--------R-.........------Q----------AG  430
CPZ.GA.88.GAB1          ----------D-------------Q-------G---------------S-----------M-Q-QGRA...DVFF-K-QGAGPK...............RKI---------L----K---R-.........---R--Q----------..  437
GSN.CM.99.CN166         -D-A------QS--------E--L----M.-G-N----L---------T--S-L-----ATALKGTSS...YN-V-K-PPGR-...........G-TP--Y---QF--L--D-PK-KER.........K-F---RA--FS-P-R..  435
GSN.CM.99.CN71          -D-A------QL--------E--L----M.-G-S----L---------T--S-L-----ATALKG.SS...YN-V---PPGR-...........G-TP--Y----F--I--D-PK-KER.........K-F----A--LARQ-KT.  435
MUS_1.CM.01.1085        SD-A------S---------E--L---SM.-G------LQ--------T--S-LM---LASAF.KQAG.TL.N-V-GAKG..........ARGRGNGSQGNRGNPR-Y---QY--V--D-PK--N-.........K-F---R---LARQ-R..  434
MUS_1.CM.01.CM1239      SD-A------S---------E--L---SM.-G------LQ--------A--S-LM---ASAL.RQNS.QL.N-V-G..........ARG.KGSQGGPRGNPR-Y---QF--M--S-PK-KTR.........K-F---R---LA-Q-R..  430
MUS_2.CM.01.CM1246      -D-A------QS--------E--LV---M.-G---K--LQ----IE--M--S-LM----TNAL.RQNTINTIN-V---.QSPR..........GVMGKKRENSTR-Y---QF--L--D-PK-KST.........R-F-------LARQ-R..  437
MUS_2.CM.01.CM2500      -D-A------QS--------E--LV---M.-G------LQ--------I--SKLM----ASAL.RQNTIETIN-V-G..QGPR..........GGAGGGPRRTPR-----QF--I-KD-PK--VR.........K-F---R---LA-Q-R..  438
                                                                       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-P..  428
H2A.GW.x.ALI            TDPA------Q------------LV--G--MNP-----L----------Q---LM---LKEAMTP-PIP..FAAA-Q...........RRTI--W-------S--Q------Q.........-------P--L-AN-P..  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-P..  426
H2A.SN.x.ST             TDPA------Q------------LV--G--INP-----L----------Q---LM---LKEAMAP-PIP..FAAA-Q...........RRTI--W-------S--Q-----RQ.........-------A--I-AK-P..  426
H2B.GH.86.D205          TDPA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LKEALTP-PIP..FAAV-QKAGK...........RGT-T-W----Q-----Q-----RQ.........-------T--I-SK-P..  425
H2B.CI.x.EHO            TDPA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LKEALTPS-NP..FAAA-PRAG...........KRT-T-W----A-----Q-K---RQ.........------QQ--I-SK-P..  424
H2G.CI.92.ABT96         TDPA------Q---I--------LV--G--MNP-----L-----X----Q---LM---LKEAL-PTALP..FAAA-QKTGG...........KRSTI--W--------V-Q-----RQ.........-------P--I-AK-P..  431
H2U.FR.96.12034         TDPA------Q---I--------LV--G--MNP-----L-----I----Q---LM---LKEALQPMPIP..FAAV-QRGGP...........RRP-R-W----------Q-K---RQ.........-----E-P--N-AS-P..  429
MAC.US.x.251_BK28       TDAA------Q---I--------LV--G--VNP-----L----------Q---LM---LKEALAPVPIP..FAAA-KRGP...........R-PI--W-------S--Q-----RQ.........-------MD-V-AK-P..  428
MAC.US.x.251_1A11       TDAA------Q---I--------LV--G--VNP-----L----------Q---LM---LKEALAPVPIP..FAAA-QRGP...........R-PI--W-------S--Q-----RQ.........-------MD-V-AK-P..  428
MAC.US.x.EMBL_3         TDAA------Q---I--------LV--G--VNP-----L----------Q---LM---LKEALRPVPTP..FAAA-QRGP...........R-PI--W-------S--Q-----RQ.........R------MD-V-AK-P..  428
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-P..  429
SMM.US.x.PGM53          TDPA------Q---I--------LV--G--MNP-----L----------Q---LM---LKEALTPGQLP..FAAV-QKGP...........R-TI--W-------S-KQ-----RQ.........-------T--V-AK-P..  429
SMM.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-P..  429
SMM.SL.92.SL92B         TDPA------Q---I--------LV--G--MNP-----L----------Q---LM----KDALTGSLV....AA-FRGAAKGQ...........GNKP-IR------T--S--Q-----R-.........------E--RIQAN-P..  432
STM.US.x.STM            -DPS------R---I--------LV--G--MNP-----L----------Q---LM---LKE-FQPDPLP..FAAA-QQG...........RRT---W----A----KQ-KG--RQ.........------QQ----AK-P..  429
MNE.US.x.MNE027         TDPA------Q---I--------LV--G--MNP-----L----------Q--KLM---LKEALAPGPLP..FAAA-QKGP...........R-PI--W-------S--Q-----RQ.........------QM--V-AK-P..  428
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---K..  427
VER.KE.x.TYO1_patent    --G---Q----S--I--------V---G--MHP-----L---------SY--K-M--M-QTMQ-Q.......N-V-Q-GPKR-...........RPPLR-Y----F--MQ-Q-PE---T.........K-L----L--LA---R..  434
VER.KE.x.9063           --G---Q----S--I--------V---G--MHP-----L---------SY---M--M-QNMQTQ.......N-V-Q-AAGGV-...........RQRPPL--Y----F--MQ-Q-TE---I.........K-L----L--LA---R..  437
VER.KE.x.AGM155         --G---Q----S--I--------V---G--MHP-----L---------SY--K-M--M-QNLQSQ.......N-V-Q-GG-GRP...........RPPP--Y----F--MQ-Q-PE---I.........K-L----P--LA---R..  435
VER.DE.x.AGM3           --G---Q----S--I--------V---G--MHP-----L---------SY--K-M--M-QNMQSQ.......N---Q-GQ-GRP...........RPP---Y----F--MQ-Q-PE---M.........R-L----P--LA---R..  439
TAN.UG.x.TAN1           TP-D----------I--------LV--G--IHP-----L-------------KLMV---Q-MQG.......VN-V-GAPRGGRG...........RGRGPPR--K--QI--IQKD-PRAGPN.........K-L----P--LA---R..  437
SAB.SN.x.SAB1C          TDPA------QS--I---------V--G--MNP-----L-----I--AQ----LM----TAAFQQQ-V..GN-FV-Q-ARPRGP...........LGGRGRPLNPNI--Y----P--L--F-K---RQ.........------SPD------Q..  453
OLC.CI.97.97CI12        G-----T----K--I--------GV----.VKPS----L----------Y--------LRP-QMVQQRPAP...........GGGPQQGRLR-Y---RF---KN--SQKGP.........T-Y----P--IA-N-R..  405
DRL.x.x.FAO             TDAA------Q---I---------V---G--MNPS----LL--------------M----REQQAA.........----QNPP-GPP...........RGPRGPPP-NPR-P---QF---L-Q-TK---.........---FR--ALD-MLRN-...  435
RCM.NG.x.NG411          -DPA------QS--I---------V--G--MNP-----L-----A---Q----------QM-Q.......SN--A-Q-GP-KGP...........PKLGGGPRFL--Y----T-----Y-KTS---.........----R--E---L----P..  434
RCM.GA.x.GAB1           -DPA------QS--I--------MV--G--MNPS----L---------Q----------QMMQSN.........--A-QSAN...........RGPPRRSGGNPNLR-Y----P--IS-Y-K---R-.........------SPD-LL-----..  433
MND_1.GA.x.MNDGB1       --G---M--QQH--IE----E--Q--RS--KG------LE--------Q----LM--M-RT-VGQ..S....QNFVQQRGPQR...........GPVRQPTGRKPI----N----V--FFK---R-.........---N--AMD--KAQ-P..  423
MND_2.x.x.5440          TDPQ--T---Q---I--------AT--G--MNP-----LL--------KY--QMM----Q-AQAA.......V---NSGGPPRG...........PPRQPPRNPR-P----F--VL-D------R.........---F---DP--L-RN-...  433
MND_2.CM.98.CM16        TDPQ--T---Q---I--------S---G--MNPS----PL--------KY--QMM----KEAQS.......AV---NSGGPPRG...........PPRQPPRNIR-P----F--GL-D-IS----.........---F---DL--I-RN-P..  434
MND_2.GA.x.M14          TDPQ--T---Q---I--------S---G--MNPS----LL--------KY--QMM----QKMQS.......EV---NSGGPPRG...........PPRQPPRNPR-P---RY--VL-D--L----.........---F---DT--M-RN-P..  434
DEB.CM.99.CM40          -D-A------T--MI-------RI---G--QNP-----LH--------QQ---LM----ASALKE-GS...LG-V-QRRGPPGS...........RRQIR-----QI--LQKD-KR---T.........K-F---Q---IA-N-G..  458
DEB.CM.99.CM5           -D-A------T--MI-------RI---G--QNP-----LH--------QQ---LM----ASALKE-GS...LG-V-QRGGQGGG...........PRRQLR-----QI--VQ-D-KK---V.........K-F-------IA-N-G..  455
DEN.CD.x.CD1            -D-S--L---TS-----------Q--E----NP-----LQ---------Q---PT---LASAFQKQGLCHGSRKGAQPPVER...........RGLR-----QI--L-KD--K-KRVT.....QPGA-F----L---A-N-R..  462
TAL.CM.01.8023          TD-A------Q---------E--A----M--Q------LQ--------Q--S-LM----AEALK.QNTQQVIA-V-.-QGGP...........KGRGGPRRTPPGQIR-Y----F--I-K--P--Q-PRG....PPGS-F---QM--RAAQ-R..  453
TAL.CM.00.266           TD-A------Q---------E--A----M--G------LQ--------Q--S-LM----AEALK.QNTQQVIA-V-Q-QGGP...........RGRGGPRRNPPGQIR-Y----F--I-K--P--PRQKV....PPGT-Y----P--IA-Q-S..  450
MON.CM.99.L1            TD-A------Q-------------L----M.-G-S--D-LQ--------M--S-I-----AGAIAN...MPMN-V-...........ARGPPQRKGQP--------F--M-K--K--QRR.........K-YN--QP--LA---PQP  434
SUN.GA.98.L14           G-E---E--K-KM-----------LVI----ETPS-----R--------S--GKI-----ASAFRQQVG..RQA-V-QNLPPRNSQ...........GRFVRIGGGGPR-PMT------P--L-NQ--E-K-G......PPGS------M--KQAQ-P..  436
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-P..  416
WRC.CI.98.98CI04        GAP---Q---DK--I--------A---G-.QQP-----L----------Y-SKIM-Q-LQEAL.-.SQ....A-V-QQPVKP...........KGPIR-----QI--M-KD-PK-K-PVPLRRGQGPT--G--EI--VQRN-P..  420
WRC.GM.05.Pbt_05GM_X02  GVP---Q---DK--I--------Q---G-.QSP-----L----------Y-SKIM-Q-LQEAL.-QSQ....A-V-QQAVK...........RGPLR-----QI--M-KD-KK-KRSPQQR..GPPT--G---T--IQ-N-P..  416
LST.CD.88.524           G-H---E--K-KM----------MVI----EG-------K--------A--RKI-----ATAMQGQMK...HL-V-QTPP--AQ...........GRFVRTGGGGPRRPLT------P--S-KM--Q--QE...........---N--AKD-R-AQ-PKP  428
LST.CD.88.485           G-H---E--K-KM---------RMVI----EG-S--D--K--------A--GKI-----ATAMQ-QMR...QN-V-VTPP--EQ...........GRFVRTGGGGPR-PLT------P-----M--Q--QE...........---N--SKE-RFAQ-PKP  430
LST.CD.88.447           G-H---E--K-KM---------RMVI----EG-------R--------A--G-I-----ATAMQ-QMR...QN-V-VTPP--AQ...........GRFVRTGGGGPR-PLT------P-----M--Q--QE...........---N--SKE-RFAQ-PKP  430
LST.KE.x.lho7           G-H---E--K-KM---------RLVI----EG-S-----K--------A--GKI-----ASAIQQQVR...QN-V-VTPL--AQ...........GRFVRTGGGGPR-PLT------P-----M--Q--QE...........---N--SKE-RFAQ-P..  431
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-RQP  451
SYK.KE.x.KE51           -DPA----L-QS--I-----E-RS----MVK.P-----LQ--------QY--KL-----VMTQH-L.......G-I-GPRQGSNP...........RRGPTR-----QL--LQKD-PR-K-L.........K-FN--GT--IARQ-RQP  452
COL.CM.x.CGU1           -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
```

```
                              p7 end p1 start        p1 end p6 start      Vpr binding                                                                  Vpr binding
                                                                                                                                                     p6 end, Gag end
H1B.FR.83.HXB2             .......ERQANFLGKIWPS....YKGRP...GNFLQSRP..........EPTAPPEESFRSGVETT.....................................TPPQK.QEPIDKELY..PLTSLRSLFGNDPSSQ*   500
H1A1.UG.85.U455            .......--------------....N----...---P----..........------A-I-GM-EKM.....................................-S-A-.--LK-R-QT..--V--K-------L---   494
H1B.US.90.WEAU160          .......-----------S-....Q----...---P---L..........------------FRE---...................................--S--.---------..-----K----------   501
H1C.ET.86.ETH2220          .......--------RL---....N----...--------..........-------SLRP......EPTAPPPE.....................SFRFEEA--SP-.--LK-R-....A----K------HLL--   505
H1D.CD.84.84ZR085          .......--------------....H----...--------..........------A-ASGL-RR*P..................................#SQK-E.-KDK-----..--A--K-------L---   501
H1F1.BE.93.VI850           .......--------------....N----...--------..........------A---GFRE-I-....................................-SP-.--QK-G---.P--A--K-------*---   494
H1G.SE.93.SE6165           .......--------------....N----...-----N-T..........------A--LGF-E-IA....................................-SP-.--MKE-------...P--K----S--*---   498
H1H.CF.90.056              .......--------------....S----...--------..........------A---GF-E-M-....................................-SP-QEQLK---PP...-A-------S--LL--   501
H1J.SE.93.SE7887           .......--------------....S----...--------..........------A--LGL-E-I.....................................-SP-.---K------.-----K----S--L---   498
H1K.CM.96.MP535            .......--------------....H----...--------..........------A---GF-E-I-....................................-SPR.--TK---QS.P-----K-------L---   496
H101_AE.TH.90.CM240        .......--------------....N----...---P----..........------A-NWGM-E-I-..............................GEEI-SLP-.--QK---HP.P--V--K-------L---   504
H102_AG.NG.x.IBNG          .......--------------....S----...---P----..........------A---GM-E-I.....................................P-SPQ.---R--G--.P-----K-------*L--   495
H103_AB.RU.97.KAL153_2     .......--------R-----....S----...---P----.......--S---A-N-GM-E-I.....................................--SL-.--QK-R-QH.P-SI--K----D--L---   499
H104_cpx.CY.94.CY032       .......--------RM----....S----...-----N--..........------A-CLERKE---...................................SSL-.---R------.-----K----S--L---   501
H1N.CM.95.YBF30            .......G--------S-SP....F----...---P-TTTRK..........------L--YGFQE-..............................KSTQGKEM-EN--RTENS--.P---------------   513
H1O.BE.87.ANT70            .......GK-------Y--P....GGT--...--YV-RPA..........H-S---.................................MEEEVKGQENQE-KGGPN----..-FA--K----T-Q*L-Q   501
H1O.CM.91.MVP5180          .......G--------Y--P....GGT--...--YV-KQV..........S-S---.................................MEEAVKEQENQS-KGDQE----..-FA--K----T-Q*--Q   501
CPZ.CD.90.ANT              ......NTGKV-----PT-T...WWGC--...---V-KEEVV..........------I-IYQEEH................................KR-QKGLKG-EELPPS-......--K----K-Q*WKF   525
CPZ.CM.01.SIVcpzCAM13      .......G--V-----G----....RS----...---V-N-T..........------A--YGYQEK.....................................EKNQEEQEEGNR--.P-----K----S------   507
CPZ.CM.05.SIVcpzEK505      ......G---------V--L....S-----...---P-TTTRK..........------I--YGYQ....................................EEKTTQGTEREEK-KTESS--.P-----K----S---L--   512
CPZ.CM.05.SIVcpzMB66       .......---------L----....S-----...-----K--..........------I---GFRE-..................................EPAQKEEV-SE-TE.NK--.PQ----K-------L---   500
CPZ.CM.98.CAM3             ......G---------G-SP.WSGGSK--...-----G-K..........------L--YGYQEA-G..................................KRQEEEREK-KE..P--.P-----K----S----E-   508
CPZ.GA.88.GAB2             ......EG--V----GG---..R..S----...---P-RKV..........------P-T-GYAEQ-Q...............................TGKSQKGEESTT-KE.SP--....P--K----S------   508
CPZ.CM.98.CAM5             .......G--------D-SP.WSGRSK--...---P-GGK..........------M--YGFQET-M.....................................EK-KE-KE..P--.PQ----K----S---LE-   503
CPZ.CM.05.SIVcpzMT145      N......G-.I-----N--P....WG-----...--YV-G-K..........------M---GFQ.....................................EEKKV-EEVREESKPPS.-----K----------   492
CPZ.TZ.00.TAN1             N.....NSTGV-----RT-L....WGC--...---V-NT-...EKGKAQEQE--QTP..........VVPTAPPLE......................................MTMKGGF--K-I--S-Q**--   526
CPZ.CM.05.SIVcpzLB7        .......GK.V-----L-----....SQ----...---P-----..........------I---GFQPK................................TTQKQEDREEDKTENK--.P-----K-------L---   504
CPZ.US.85.CPZUS            .......N--------H-SPSWSGGSK--...----EN-K..........------I-D-GYQE--V....................................-QEKQGK-KE...PF..Q----K----S----E-   506
CPZ.GA.88.GAB1             .......G--V-----G----..R..S----...---V-N-T..........------I--YGYQE-E....................................KSQEKKEGESS---..-P---K----S------   509
GSN.CM.99.CN166            ...TGTPIKELY--DLGRGR.....RRA-...R--PVTFL..........L----A-.............................EEEPQQLNLSTPVGEPQSKGQK-ESRNT--.....P--S----E-Q...-   513
GSN.CM.99.CN71             ....GTPAKEL---DLGRGR......RA-...R--PVTSL..........L-S--A-.............................EVEPQQLNLSPPVGEPQNKEQKAEKRNT--.....P--S----E-Q...-   511
MUS_1.CM.01.1085           ....SDNTKAV----DPFGL....Q-R--...R--P-TSL..........L------QEGELVNLSPAV.................................GPGEGKQDR-S--.....P--S---QD-Q-D--   510
MUS_1.CM.01.CM1239         ....SEGAKS-----DPLGL....QRR--...R--P-TSL..........L------QEGDLVNLSPPV.................................GP-DGASDR-S--.....P--S---QD--LEK-   506
MUS_2.CM.01.CM1246         ...TDTGKSAV-----DLGF....RSRK-...R--PVTSL..........M-----DTAEPVNFSPLVGPKE...............................EPPKDRKDR-T--.....P--S----D-Q-K--   517
MUS_2.CM.01.CM2500         ...TNSDKGAV-----GFGF.....HRK-...R--PVTSL..........M-S---DTAEIVNFSPEVGPRKEN............................QTPQAP-DR-N--.....P--S----D---E--   519
                                                                                                                                                       p6 end
                                 p8 p1                  p1 p6         PTAP motif                              PSAP in HIV-2 B, U                      Gag end

MAC.US.x.239               .......D---G---LGPWG......KK-...R--PMAQ......VHQGLM-----............................EDPAVDLLKNYMQLGKQQREKQRES..REKPYKEVTED-LH-N----G-Q...-   511
H2A.GW.x.ALI               ..........----G---LGPWG......KK-...R--PVT-.......VPQGLT-----..............AEPAADLLEQYMQQGRKQREQRERPYKEVTEDLLHLE....Q..GETPHKEVTED-LH-N----K-Q...-   522
H2A.DE.x.BEN               ..........----G---LGPRG......KK-...R--PVTQ......APQGLI-----..............ADPAAELLERYMQQGRKQREQRERPYKEVTEDLLHLE....Q..RETPHREETED-LH-N----K-Q...-   522
H2A.SN.x.ST                ..........----G---LGPWG......KK-...R--PVAQ......IPQGLT-----..............IDPVEDLLEKYMQQGKRQREQRERPYKEVTEDFLQLE....-..QETPCRETTED-LH-N----K-Q...-   522
H2B.GH.86.D205             ..........----G---LGPWG......KK-...R--PMTQ......VPQGVT-S---MNPAEGMTPRGATPSAPPAD...........PAVEMLKSYMQMGRQQRE....S..RERPYKEVTED-LH-N----E-Q...-   522
H2B.CI.x.EHO               ..........----G---FGPWG......KK-...R--PVQA.......PQGIV-S---MNPAFGMTPQGAIPSAPPAD...........PAEEMLKNYMQLGKKQKE....N..RERPYKEVTED-LH-N----E-Q...-   520
H2G.CI.92.ABT96            ..........----G---FGPWG......KK-...R--PMTQ......VPQGLT-S---..................MD...........PAVDLLKNYMXLGRKQKE....Q..RNKPYKEVTXX-LH-S----D-Q...-   510
H2U.FR.96.12034            ........-TKVG---LGPWE......KK-...R--PMAQ......VPQGLT-----....AEPAVDLLTPTAPPAD...........PAVDLLKSYMQQGKKQKE....N..RERPYKEVTED-LH-N------Q...-   522
MAC.US.x.251_BK28          .......D---G---LGPWG......KK-...R--PMAQ......VHQGLT-----............................EDPAVDLLKNYMQLGKQQRE....S..REKPYKEVTED-LH-N----G-Q...-   507
MAC.US.x.251_1A11          .......D---G---LGPWG......KK-...R--PMAQ......VHQGLT-----............................EDPAVDLLKNYMQLGKQQREKQRES..REKPYKEMTED-LH-N----G-Q...-   511
MAC.US.x.EMBL_3            .......D---G---LGPWG......KK-...R--PMAQ......VHQGLT-----............................EDPAVDLLKNYMQLGKQQRE....S..REKPYKEVTED-LH-N----G-Q...-   507
SMM.US.x.H9                ..........----G---LGPWG......KK-...R--PMAQ......MPQGLT---X-............................EDPAVDLLKNYMKVGRRQRE....N..RERPYKEVTED-LH-N----E-Q...-   508
SMM.US.x.PGM53             ..........----G---LGPWG......KK-...R--PMAQ......MPQGLT-----............................EDPAVDLLRNYMKMGRKQRE....N..RERPYKEVTED-LH-N----G-Q...-   508
SMM.US.x.SIVsmH635F_L3     ..........----G---FGPWG......KK-...R--PMAQ......MPQGLT-----............................EDPAVDLLKNYMKMGRKQRE....N..RERPYKEVTED-LH-N----E-Q...-   508
SMM.SL.92.SL92B            .......NQK-G---LGPWG......KK-...R--PMQT........TSLT-S---.............................DPAARIVKEYLEKAQREKT....R..RSRPYKEVTED-LH-N----E-Q...-   508
STM.US.x.STM               ..........---VG---FGPWG......KK-...H--PMAQ......IPQGLT-----............................EDPAADLLRSYMQLGKKQRE....S..RKTPYKEVTED-LH-N----E-Q...-   508
MNE.US.x.MNE027            .......D---G---FGPWG......KK-...R--PMAQ......MHQGLT-----............................EDPAVDLLKNYMQLGKQQRE....S..KRKPYKEVAED-LH-N----E-Q...-   507
GRV.ET.x.GRI_677           .......NG------YGHWG.....GAK-...R--V-Y-.....GDTVGL------........................METAYDPAKKLLQQYAEKGQRLREER....EQTRKQKEKEVEDV--S----G-Q...-   514
VER.KE.x.TYO1_patent       ........G-V----YGRWM.....GAK-...R--PAATL........GA--S---PP-GTTP.........................YDPAKKLLQQYAEKGKQLREQKR...NP-AMNPDWTEGYSLNS.---E-Q...-   520
VER.KE.x.9063              ........G-V----YGRWT.....GTK-...R---AATH........GV--S---PPTP.........................YDPAKKLLQQYAEKGKQLREQGK...RT-PTNPDWAEGYSLNS.---E-Q...-   520
VER.KE.x.AGM155            ........G-V----YGRWM.....GTK-...R--PAATL........GA--S---PPNN....................STPYDPAKKLLQQYAEKGKQMRNQNR...NP-ANNPDWNEGYSLNS.---E-Q...-   521
VER.DE.x.AGM3              ........G-V----YGRWM.....GAK-...R--PAATL........GV------PP-P.........................YDPAKKLLQQYADKGKQLREQRK...KP-AVNPDWTEGYSLNS.---E-Q...-   522
TAN.UG.x.TAN1              .......SG------RMPTW..G...TK-...R---EQGGA.........V-----MPAHGFP.........................TGSPAAGAYDPARKLLEQYAKKG..DQLRKQKEKELEDY--S----G-Q...-   524
SAB.SN.x.SAB1C             ........K-V----FGPWG.....R-K-..RNFP-T-..........IR-----M-RDY-R..PEENWYADRPPTRGPGPDDPATALLKQYAVQGKRQKQQWQ......NHSPQQSP-EEAYS-------E-Q...-   555
OLC.CI.97.97CI12           .VGGQQQQSKV-----WRGG.....-AE-PLNMQQRTEMD.........IF-S--TL-..................................QILEKGTR......EKREM--EKKPLYP--A----E-Q...-   481
DRL.x.x.FAO                .......PK-V----NTPWG.....S-K-...R--P-AS..........LT-S---LPG....................YLQEDPAERMLQKYMEQGAQQKRQQR...-QQQKRGP-EEAYN--S----S-QLQ.-   522
RCM.NG.x.NG411             ........KEV----NAPWK.....Q-K-...R-LPLTS..........LR----GR-EG.................WNPQYDPAEEMLRKYLALGRQHKQEQRKE....NKEKVGRA-EDA-S--N----S-L...-   522
RCM.GA.x.GAB1              ........K-I----RLPWG.....Q-K-...R--PLTS..........LT-S--GM--NYDP.......................AEEMLKNYLRRAGEQKRQQRQE....ESKKREGA-QEA----N----S-QLQ.-   519
MND_1.GA.x.MNDGB1          ...KPAQQQRV----YGPWG.....PSK-....--YPAQE.........VT-----L.........................EEKPLQKTLSTYQKLGRGL......RQKMKE-KREEDFH--ST--QE-Q...-   503
MND_2.x.x.5440             .......PKMV----NTPWG.....S-K-...R--PAMP..........LT-S---MPGL....................EDPAEKMLLDYMKKGQQQRAAAGA....-KEEK-GP-EAAYN--S----T-QLQ.-   519
MND_2.CM.98.CM16           ........KMV----NTPWG.....S-K-...R--PAMP..........LT-----MPGM....................EDPAERMLLDYMKKGQQQKAESK......--KKERGP-EAAYN--S----T-QLQ.-   518
MND_2.GA.x.M14             ........KMV----NAPWG.....S-K-...R--PAMP..........LT-S---MPGM....................EDPAEKMLLDYMRKGQQQRAAKET....KQEK--GP-EAAYN--S----T-QLQ.-   519
DEB.CM.99.CM40             ......QTPR-----NTPWA.....SR--...R---EGIG......KYIP-S--DQVY-T..................................TRVPVQGQENNPQRE-TTRRGK-.PD-MG-K----E-Q...-   536
DEB.CM.99.CM5              ......QVPR-----NAPWA.....SR--...R---EGIG......KYIP-S--DQVY-T..................................TRVPVQGERNAPQEDTTTRRGK-.PD-MG-K----E-Q...-   533
DEN.CD.x.CD1               ..SNTNDQG-V----NGLGR.....-NP-...R--V-YY-Q......MNQDLPWS-VRV...................QQQPQYPQNFQEGGNRQSCQEDKRQNKQS...HQEDEETTRESLYP--K----D-Q...-   559
TAL.CM.01.8023             .......QP-------AFLG.....-IP-...VKGPRNF-.........VVQ-----I...........................EPLTASWQEGTTTWTPPTKD....AAAPKQQTRESLYP--A----D-Q...-   532
TAL.CM.00.266              .......SA-------SFLG.....-IP-...VRGPRNF-.........VVQ-S---I...........................EPLTASWQEGTTTWTPPAKK.....PQETTTTRESLYP--A----D-Q...-   528
MON.CM.99.L1               P....KQNKGV----NPFGP.....-KG-...R--PLT..........SVQ-S--TLPA.................................GTAEPVNLNQEIGS-...KTGTP--TRRDLYP--A----E-Q...-   511
SUN.GA.98.L14              ......QGK-V----YGPWN.....R-P-....---P.............VM-S---...................LEDLTLGNRMTPPQSKAERALETYRLLGQGLR....AQQKR--RGECQEPC-NM--PE--....-   522
WRC.CI.97.97CI14           .....KIR--V----NGRGQ.....MPSNNYPVQKLRSLNNFGP..TTTM-S---MREDPNI...................QELEKYLLKGAQLREQQEHKK......T-GQKR-VGVPEWP--K---EP-Q...-   512
WRC.CI.98.98CI04           .....KSR--I----NGRGQ.....MPSNNYPVQKVRSLNNFGP...ATI-S---MREDPNI...................QELEKYLLKGAQLKEQAQAK.......K-EKKR-TGLPEWP--K---ET-Q...-   514
WRC.GM.05.Pbt_05GM_X02     .....ETR--V----NGRGQ.....RPSNNYPVQKMNAGVNTFG...PVV----V.........................DPTMQELEQYLLRGAQMKRE......K-KEMQKEGKVEYP--K---EP-Q...-   503
LST.CD.88.524              .......KGKV----YGPWR.....NG-S.---PIMGGAAG......VI-S---....................MEKVPPTRAERAIETYRNLGQQLRK.....QQQVPQKCVDEPCLNFFF-PD-Q...-   512
LST.CD.88.485              .......KGKV----YGPWR.....NG-P.---PLMGGNAG......VV-S---.....................MERNSTKAERALETYRNLGQQLRK.....QQQVPQKCVDEPCLSFFF-PD-Q...-   513
LST.CD.88.447              .......KGKV----YGPWR.....SG-P.---PLMGGNAG......VV-S---.....................MERSPTQAERALETYRNLGQQLRR.....QQQIPQKCVDEPCLSFFF-PD-Q...-   513
LST.KE.x.lho7              .....KPKGKV----YGPW-.....SK-P.--YPLLGGAAG......RI-S---.....................MESAPTKAERALETYRTLGQQLKR....-QQQVPQKCVDEPCLNSLF.PD-Q...-   516
SYK.KE.x.SYK173            KRNQGPPVA-------G-GV.....SR--P..A--PV..........RS--S---L-..........DIEDGPWLTWSAQMSQQAQAKAQNSPSKKPPTNREVLS-...K-SSG--ETKSLYP--S----E-Q...-   555
SYK.KE.x.KE51              RKGQGNPPP-------G-GN......RK-P..A--PV.........MRET-----A-D.....WPWQQQTWGNYASPQQKHIAMATPLQPQSSPKTVKDLLVPGK-S...EKKEEMKEKGPLYP--Q----D-Q...-   562
COL.CM.x.CGU1              .......NK------ATRGV..E..........................LQTAIF--..........................................KMSKD...LP-RREKGES.LYP--K----D-Q...-   483
```

```
                           Gag-Pol TF start                                                                                   Gag-Pol TF end  protease start
H1B.FR.83.HXB2             FFREDLAF..LQGKA...REFSSEQT..RANSPTRRELQVWGRDN..............................NSPSEAGADRQGTV.....SFNFPQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSL....PGRWKPKMIGGIGGFIKVRQYDQILIEIC  123
H1A1.UG.85.U455            ----N---..Q--E-...--------.-------S-N-WDG-K-..............................DL-C-T--E----D.......--S---I---------V------I---------------DIN-.....--K----I--------------------  122
H1B.US.90.WEAU160          ------V-..PK--.............-T---------Q-...................................--L-----N---A-..........---I-----------E-----------------D-N-.....--K-----------------VP----  123
H1C.ET.86.ETH2220          ----T---..Q---------P---.--------ESQTRANSPTTR..........................ELQVRGSNTFSEA-AERQG..-L----I---------------------------------IN-.....--K--------------------I----  130
H1D.CD.84.84ZR085          ----N---..P-----..G-L----.-------S-SFGFGEEIT..........................#LPETGTEGQRQGTV..-LS---I------V-----------------------------DIN-.....--K-------------------H------  123
H1F1.BE.93.VI850           ----N---..Q--E-...-K-P----.-------S---R-QRG--.........................PLSEAGAERRGTV.P-LS---I---------------I-----------------DIN-.....--K--------------K---N------  123
H1G.SE.93.SE6165           ----N---..Q--E-..--------D..-T----C-KPR-RRG-S.........................PLPEA-DEGKG..AISL--I---------V------I---------------IN-.....--------------------------VP---S  120
H1H.CF.90.056              ----N---..Q-RE-...-K--P--A..-T----S---R-RRG-D.........................PLSE-GAAEGQGT...-LS---I---------V--E---R----------------IN-.....--K------------------E-VA----  122
H1J.SE.93.SE7887           ---------..Q-RE-...--L-P----.-------S--PRARRG-........................PLPETGAEGQGTV..-S----I----------R-----R---------------DID-.....-RK-----------------NEVP---E  121
H1K.CM.96.MP535            ----N---..P--E-..................S---R-R-G--..........................PLSEAGDQRQGTE.P------I------I----V----R----------------IN-.....--K----------------------V-----  123
H101_AE.TH.90.CM240        ----N---..Q------..G----------.------S-K-GDG---......................GGRDNLLTEAGAERQGTS.S--S---I---------V------------------DIN-.....--K--------------K----------  128
H102_AG.NG.x.IBNG          ----N---..Q--E-...-K-------.GT--S-S---WDG---.........................TSLSTAGTEGQGAI.S------I---------VR-E---I---------------DIN-.....--K------------------------  123
H103_AB.RU.97.KAL153_2     ----N---..Q-RE-...-K--------.--I---S-K-WDG---.........................NPLPETGTEGQGTA.S------I---R-----VR---------------------DIN-.....--K------------------------  123
H104_cpx.CY.94.CY032       ----NV--..Q-RE-...-K-----A.------A-GM-REERG--.........................LLSEAGTEGQGTI..------I-----------L---IR--------------IN-.....--K--------------------P----  122
H1N.CM.95.YBF30            ----E-VS..--RET...-KLPPDNNKE--H--AT---W-S-G.......................EEHTGEGD--EPGEDRELSV.PT-----I------VI-V---KEVR------------I--LQ-....E-K------------------N-TVD-Q  130
H1O.BE.87.ANT70            ---QI--S..GGHE-...-QLCA-TS..TPI---.................................DGGGSEGTGESGTER-PERAL..-VCL--IP--D--I--ARV--H-C-V-----------NNIQ-....E-K--------------KE--NVTV--E  119
H1O.CM.91.MVP5180          ----V--S..GGHE-...-QLCA-TS..VPI---.................................NGGGSEGTRESESEG-SGRAV..PICL--IP--D--I--A-V--H-C-------------NNIQ-....E---T-------------KE-NNVTV-VQ  119
CPZ.CD.90.ANT              ----TDPH....VVGVQT--LCA-GGSSG---S-H-D-..S-GA.........................Q-DSEGS--GGGTT..-LV--EIP-----MMEVL-Q--KCQ-----------V-GIH-....Q-N----T--------S-Q--NKVP-Q-G  123
CPZ.CM.01.SIVcpzCAM13      ----G---..P-RE-...-QLCT--D..------S---W-PRE..........................G-EPRGTGRGEQTI.P-S----I-------IPVR-E---C------------IQDIQ-....Q-I--------------K---NVS---E  122
CPZ.CM.05.SIVcpzEK505      ----G--S..--RE-...-K-PPDNNKE-----SN---W-S..........................GGEDHTGDRE-RKGEDRELSV.PTL----I------IL-V----EI------------I--IQ-....E-K--------------K---NVI---Q  129
CPZ.CM.05.SIVcpzMB66       ----T---..Q--------.---P--E-..-T--S-N---R-Q-G.......................GTC-EGGSEE-GDREQAV.S-A----IS-----V--VR-E------------------IE-....G------------------------NVT-D--  126
CPZ.CM.98.CAM3             ----R-VS..VERGIKET--LPAG-E..GTH-SAP---W-S-SNR.......................K-T-GGEGEGKGTPVS-V-L--I-------LGV--E---I------------IDNVQ-....T-K-R--------------K---N-A---E  128
CPZ.GA.88.GAB2             ---GG---..PERE-...--L-PKEG..GT--S-S-D-R-CRA-T.......................DRKESE-GGEYNRKGEP..-LS---I-------LPVRVA---C-------------DNIQ-....E-K-R--L-----------K---NVE---E  126
CPZ.CM.98.CAM5             ----G-VS..VEREIKKT--L-AGRE..GTH-S-HG--W-S-NN-........................GEKE-GEGTPVS-I-L--I-------LGV-VE---I------------IDNVQ-....T-K-R-R-------------K---N-V---E  124
CPZ.CM.05.SIVcpzMT145      ---KK--S..-G-E-...--LCAR-E..GT--S-YGI-RIS.........................GGE-SQGGGE-GEQAAV.------I-------IPV--E---I------------I-DLN-.....--K-S--------------K-FERVN---E  122
CPZ.TZ.00.TAN1             ----THPL.....VGVQT--LCA-HP...RERE.......................GSGAGDSTDTSGANCP-TGDDDERRVL---N-----MM-V-VQ--VCQ----------S-FCNIK-....K-Q-T--T--------VP-SE-YN-PVQ-G  119
CPZ.CM.05.SIVcpzLB7        ----T---..-P-----..-KL--------H--AN---W-SAK.......................NHPETGGQGRG-DREQAV.P-L----IS-----I----V--Y--G---------------DLE-....G------------------K---N-P---G  126
CPZ.US.85.CPZUS            ----T-VPI.VERGIKET--LPGK-E..G-H-S-N-GFR-SR--S.......................DTGETGKGKGAL..-A----I-------LR--VA--IV-------------DNIQI....E-T-R---M---------K---HVN---E  125
CPZ.GA.88.GAB1             ----R---P..-RE-...-QLCA--N..-T-G--D---W-P-GR.......................-EPGEER-REQSI..-T-L--I-------IPV-VE---C------------I-RIQ-....Q-L-----------------K-F-NVH---E  121
GSN.CM.99.CN166            ---PWPGA....PPSP..Q-L--NFPAANSPGG.......................GGRTPAAQSLDT-RGTPEQGTEGGEQEH-LSL-VIS--G--MKEVS----VVSM----------IVQDSAI..ELDTP-T---V-----L-Q--EHKHVEVIFN  128
GSN.CM.99.CN71             ---PWPGA.....TSP..Q-L--NFPAAISPGG.......................GGGTPAAQPV-AGGGAPEQGAEGRKEEH-LSL-VIS--G--MREVSM---VVSM----------IVQ-SAI..ELDTP-T--TV-----L-Q-KEHRHVEVIFN  127
MUS_1.CM.01.1085           ---GS-RP.....PEKEAE-LPPDLSLANCSAR-GGGIGQSLSSS.......................GSR-GETGQEE-LS--VIP-SG--VRPVE-E--VVSM----------IIQDAHI..HLE-P-N--SV-----N-S-KE-RGVQVCFN  124
MUS_1.CM.01.CM1239         ---GSSWP.....PEKEAK-LPPDLAFTNGPARAGGRPGQSLPAS.......................GS--WSERQEE-LS--VIS-SG--IREVM-E--VVHM----------IVQDSNI..QIDKP-N--LV-----N-S--E-RGVQVSFN  124
MUS_2.CM.01.CM1246         ----GSRL.....PVSEAE-LP-YIAHAYGTPRYS-AGELLP.TSRAQG.......................RATQGQERQKD-LS--VIS--G--IK-VD-N--PVSM----------IIQDSDI..KIS-ELT--VV--L--L-N-K--KGVIVKFN  127
MUS_2.CM.01.CM2500         ----RFR......LPQEAE-LPCYIAHAISSPRYS-DS-LLP-GGSTEGE.......................PNPPSPTGQEE-LS--VIP--G--IR-VMVH--PVSM----------IIQ-SEI..EIDKE-T--VV--L--L-N-K--KAVKV-FN  128
                           Pol p15 start (-1 from Gag)                                                                          p15  protease

MAC.US.x.239               ---PWSMG....KE-...PQ-PHGS......-ASGADANCSP.................RGPSCGSAKELHAVGQAAERK-ERKQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV-------NTKE-KNVE--VL  135
H2A.GW.x.ALI               ---AWPMG....KE-...SQLPRNP......-SAGINTNSTP.SRASSGPAGAVYAAGEKAKRAEREAIQRGDGGLTA-R....AGRD-TQRGDRGLAA--FS--K--V--AY-E--PV-V--------SIVAGIE-....GSNYT--IV-------NTKE-EDVE-KVL  148
H2A.DE.x.BEN               ---VGPTG....KE-...SQLPRDP.......--SGADTNSTS.GRSSSGTVGEIYAAREKAEGAEGETIQRGDGGLAA-R....AERD-SQRGDRGLAA--FS--K--V--AY-ED-PV-V--------SIVAGIE-....GDNYT--IV-------NTKE-KNVE-KVL  148
H2A.SN.x.ST                ---VGPMG....KE-...PQ-PCGP.....N-AGADTNSTP.-RPSRGPTREVHAAREKAERAEREAIQRSDRGLPAAR....ETRD-MQRDDRGLAA--FS--K--V--AHVE--PV-V--------SIVAGVE-....GSNYS--IV-------NTKE-KNVE-RVL  148
H2B.GH.86.D205             ---VRTLG....KE-...SQLPHDP......-ASGSDTICTPDEPSRGHDTSGGDTICAPCRSSSGDAEKLHADGETTER....EPRE-LQGGDRG-AA--FS--R--V-KAC-E--SV-V-----V--SIVAGIE-....GSNYT--IV-------NTKE-KDVE--VV  149
H2B.CI.x.EHO               ---VRPLG....KE-...SQ-PRPG.....T-GDSAICAP.DEPSIRHDTSGCDSICTPCRSSRGDAKELHATREEAEG....EQRE-LQGGDRG-AA--FS--R--V-KAT-E--SV-V--------SIVAGIE-....GSNYT--IV-------NTNE-KNVE--VV  148
H2G.CI.92.ABT96            ---VWTLG....KE-...SQLPHDP.......--SGSDSISTT.................DGPSSRPTEELHAXGEEAKG....AEKQ-LQGGDRXXAA--LS--R--V--AY-ES-PV-V--------SIVAGIE-....GSNYS--IV-------NTKE-KNVE-KVV  131
H2U.FR.96.12034            ---AWPLG....K--...SQLPHGP.......--AGTDPNSP....SSRTSSGSADPDSPSSGPSSGSAKELHATGQETKG....EQRE-IQGGDGG-AA--FS--K--I--AY-E--PVKV--N-----SIVAGIK-....EPYYT--TV-E-E---NTKKFHNVN-KVL  145
MAC.US.x.251_1A11          ---PWSMG....KE-...PQ-PHGS......-ASGADANCSP.................RGPSCGSAKELHAVGQAAERKTERKQREALQGDDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV-------NTKE-KNVE--VL  135
MAC.US.x.251_BK28          ---PWSMG....KE-...PQ-PHGS......-ASGADANCSP.................RGPSCGSAKELHAVGQAAER....KQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV-------NTKE-KNVK--VL  131
MAC.US.x.EMBL_3            ---PWSMG....KE-...PQ-PHGS......-ASGADANCSP.................RGPSCGSAKELHAVGQAAER....KQREALQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVTGIE-....GPHYT--IV-------NTKE-KNVE--VL  131
SMM.US.x.H9                ---AWPMG....KE-...PQ-PHGP.....DASGADTNCSX.................RGSSCGSTEELHEGGQKAEG....EQRE-LQGGNGG-AA--FS--R--V--AY-EE-PI-V--------SIVAGIE-....GPNYT--IV-------NTKE-KDVK-KVL  131
SMM.US.x.PGM53             ---AWPMG....KE-...PQ-PHGP.....DASGADANCSP.................RGSSCGSTEELHEDGQKAEG....EQRE-LQGGDGG-AA--FS--R--V-AAY-EE-PV-V--------SIVAGIE-....GPNYT--IV-------NTKE-KDVK-KVL  131
SMM.US.x.SIVsmH635F_L3     ---VWPMG....KE-...PQ-PHGP.....DASGADTNCSP.................RGSSCGSTEELHEDGQKAEG....EQRE-LQGGDRG-AA--FS--R--V--AY-EE-PV-V--------SIVTGIE-....GPNYT--IV-------NTKE-KDVK-KVL  131
SMM.SL.92.SL92B            ---PWPMG....KE-...SQLPHAD.......NIIDTVSTP.................RSSSPDSQGVSRESTEGKD....KEEQALQGGDGG-AA--FS-RR--V-RAL-E--PL-V--------SIVAGIE-....GPGYT-RIV-------NTKE-RTVK--VL  128
STM.US.x.STM               ---LWPMG....KE-...PQLPHGP.....NTSGADANCPS.................RRPSCGSAEKLHAAGQEAER....EQED-LQGGDRG-AA--FS--R--V--AH-E--PV-V--------SIVAGIE-....GLQYT--VV-------NTKEFKNVN--VL  131
MNE.US.x.MNE027            ---LWPMG....KE-...PQ-PHGP.....NASGADTNCSP.................RGPSCGSAKKLHAVGQTAER....KQKEALQGGGRG-AA--FS--R--V--AH-E--PA-V--------SIVAGIE-....GPHYT--IV-------NTKE-KNVE--VL  131
GRV.ET.x.GRI_677           ---VWPLG....RSE.TKKFCAIQRRHSWSGTNSPP.................NGNSLRSSKEAPPAVCREGTAPERGE....RTDKETEGERSGGC-LELP--R--MKRVI-E-TPVQ----------IIQ-KD-HFPPHKP-RS-VV-----G-H-KE-QGVQVQLE  143
VER.KE.x.TYO1_patent       ---VWTVD....-GK.TEKFSRRYS.......WSG..TECASSTERHHP.........IRPSKEAPAAICRERETTEGAK-...EST-NESGLDRGIF-EL.P--R--IK-VY-E-VPIK----------IIK-ND-..QLS-P-R--I------GLN-KE-NDREVK-E  141
VER.KE.x.9063              ---VWTVD....-DK.TEKLSRCHS.......WSG..TERAPSTDT............IRPSKEAPAAVCREGKTVEGTR-...ENTTNKSRLGGRIF-EL.P--R--IK-VI-E-VPVKV---------IIK-TDI..QLT-P-R--V------GLN-KE-SNVEVK-E  138
VER.KE.x.AGM155            ---VWPVD....-DK.TKKFSR-HS.......WGG..TKCAPSTEQ.........LYTLRPSKEAPAAVCRERETNEKSEQ...KPPSEQSRLERGIF-EL.P--R--IR-CI---TAVK----------IIKDTD-..QLR-S-R--IV-----GLN-KE--NVEVQLE  141
VER.DE.x.AGM3              ---VWPVD....-SE.TQKFSRRYS.......WGG..ANCAPSTES............IRPCKEAPAAICRQGEAVEGTK-...KTTSSESRLDRGIF-EL.P--R--IK-VY-E-VPIR----------IIK-AD-..QLS-T----I------GLN-KE-SDREVRLE  138
TAN.UG.x.TAN1              ----NANL....-NK.AAQF-RARRSR.....SDGSPDACARIP...............HRFASCGSIRPSQETAGAVCQKG..RSIEKAEGEGTGGL-LEFP--R--MSK-Y-Q--PVQV---------IIK-EDI..HLE---R--VV-----G-S-K---N--VN-D  141
SAB.SN.x.SAB1C             ---VWPLG....QRE.TQEFP-DLHQTNSSPNG-GLQQAGGKLVCRQTSDQRTRARRSSNSPVKAVCCSGETAETAVA......KPLA-TEPLRGGLQL---S--R--MK-VY-E--KVT---------S-IQGIE-....GDN---RI------C-N-KA-HNQEVK-E  155
OLC.CI.97.97CI12           ----MEGR...ESR-SPQHATEDRDGHLSLCTYPG..............................TDFGEG-K......GKERDEGGEET-LSL-SIS--R--MIELY-E--QVT-----------IKGKDI..YME-NGI-RI------A-R-M--RCRIRIF.  123
DRL.x.x.FAO                ----YSLG......QRETQ----GL.................ADPISSSIARIPPGGSGREDAPKVHGAGSSAEETT-...TTTAEEGPLRGGLQL--FS-RL--TTIVEVE--KV------------IQGLE-....T-K---QI------T-N----FNCE-K-A  137
RCM.NG.x.NG411             ----CSLE.....AGETEEPSLDFLETY.....CS-...........ERGRMESPVRSSRRDAQEIPSIGETAQTGAEER....EQGESGESLRGCVVL--FS--K--VTWVE-E--KV---------G--IQNLD-....GEN----I------Y-N-K-FNHCK-T-A  141
RCM.GA.x.GAB1              ----TPLG.....SGEAEELSFDFLDSLCSRDGEQLRP..................CRRDAKELSEEGRGTKETTEAG....RE-EERGSISG-L-L--FA--K--TTIAQ-E--KV-V-----------I-GIE-....GND-T--I------Y-N-K--KNCE---A  137
MND_1.GA.x.MNDGB1          ---VWPLG......SLQTG-L-.........................GTRGDSNSSTIRGETSAENSEHLSEIRERA.....QAEDEGGEERGG-S--EYS-SR--IEEVSVD-VTIR----------IFN-RNI..KLK-N-Q--I------NLR-K---NVYV--R  130
MND_2.x.x.5440             ----YPLG......QWQAQ-LPCHA..................IDPISTPDARVGRPGRKDAVRLYEERATTEGS-R-....PER-KEGSLRG-LQL--FS--N--TTVVE-E--KV------------IKDLE-....T-N--RI------A-R----FNCKV--A  136
MND_2.CM.98.CM16           ----YPLG......QWQTQ-LP...GDAIDPNGP.................SNARDGRPSRENAVRLHEERATAESGKQ.....TGKEREGSIRG-LQL--FS--N--TTVVE-E--KV------------IKDLD-....K-N---QI------S-N-K-FFNCKVT-A  135
MND_2.GA.x.M14             ----CSLG......QWQAQ-L-....CHATDPISTP...............DARDGGSSREDVIRLYEKGATTEGSKGD....QAGEGQGSLRG-LQL--FS--N--TTVVE-E--KV------------IKDVD-....T-N---QI------A-R-K--FNCK-TVA  136
DEB.CM.99.CM40             ----YSLG.....LQET.Q-L-RRNWEVHPGEHS-SGILD..............................YQGTSTGAGEQPPEGGNNEEREV.PG--GAEIS--R--MT-VH-E-RKVQV-I-------IIS-KDI..DLEAP-T--TV--L-----N-KC-PG-E-AMA  131
DEB.CM.99.CM5              ----CSLG.....LPET.Q--FGRDWEIHPREHS-SG-FN..............................NQGTCARGKECSPGGHNDQEREV.PR--GAEIS--R--MTAVYVE-KRVQV-I-------IIS-KD-..DLDAP-T--TV--L-----N-KC-PG-E-KMA  131
DEN.CD.x.CD1               ----WSWE....EESSQELCTVLPSDESGFAMEPSEG...............PAATTVPTKFSRGGKQTILSGRQEAEQAV...PSG-RGDNEGE-LS-VE-S--R--IGQ-WVQ--RV-V-I-------IIA-QDI.NLGTEG-I--TV-----Y-N-QA-PG-E-KLM  147
TAL.CM.01.8023             ----GFFR......EDPPGQGTE-LPSGATHRTAD.........................RASHSIMAGGDYHMDPSDQG....C-STQATDEGE-LSLVSQS--R--I-KVL-E-REV-V---------IIQ-GDI..ELE-VPHQ-TV--LA-Y-Q-QS-TGVE--WE  133
TAL.CM.00.266              ----VFFR......ENPTGQRTE-FSSGATIGTAN.........................RASHSIMAGGDYHVDPT-QE.....APRNNHNKGE-LSLFGQS--R--I-KVL-E--EV-V---------IIQVGEI..ELE-VPHQ-TV--LA-Y-Q-QS-SA-E-QWE  132
MON.CM.99.L1               ----SFRS......QEGAK-LP-HLCAAQRTDSAS..............................RDCGASQSQPRDWFP...EDGDPKGDEAG-LSLLSIS--G--IRSVL----KVSM---------IIQDQHV..NLGEN-T--VV-----M-E-K--KA-QVIFE  128
MON.NG.x.NG1               ----SFW.......PRQEGAKELSSNLLA----SGTN.......................PXGXQGELLTSSGTTRDP.......SADSPEGDEEG-LSLLSIP--G--VR-VE-A--SVSM----X----IVQXQDI..NLGEN-T---V-----L-Q-K-FKGVTVSFQ  132
SUN.GA.98.L14              ---VWPLE......QGAA--L--HAFCSA..................SRGSDIGEQNDTTPVKGRESSGDLQTVRTG-Q....STAEEEGKGGM-GAMLEYA-SR---EKVF-N--PVT----------IVS-D-V..SIE-H---RVV-----Q-R-KE-ADVFV--R  139
WRC.CI.97.97CI14           ----WQRA....DALEQLSSSEA-ELEQFWANNNNAFCSSNE-GSKH.........SGAGKIFIEGSTIEGTTGTQE......--GTEEGSGGAGMAL--IS-RT--MIQCI-E-HSVS-----------IDVNL.....NKN-Q-RI------TMN-KLHR-VK-R-L  145
WRC.CI.98.98CI04           ----WERA....DALKQLPSSKG-ELEQFWAG.NNTVCSSNEGGSKH.........TRIREIPPERSSIEGAGTGK.......KGREEERNGAARVA---IS-RN--MIQCI-E-NSVS-----------IDIHLN....NKN-Q-RI------TMN-KL-RK-Q-R-L  144
WRC.GM.05.Pbt_05GM_X02     ----WPRA....EALQQLSS............SKN-CRGEYLWASG.....ANSTSGSYHARTRAISSERSSDEEG......E-ERNAEGGESRISL--IS-RT--MIQCI--RESVS-----------VDVQL.....E-N-Q-RI------TMQ-KL-KDVR-K-L  136
LST.CD.88.524              ---VWSLE......EWTF-QL-NHGRGC-SD-FSAT...................HGESASNKSREGNRDIQEPRTTTEE.....ATTSSTEVCGRALPEFLL-SR--IEKVLVN--EVT----------IFS-N-V..RIE-PYT-RVV-----Q---KEFRDVF-Q-A  138
LST.CD.88.485              ---LWTLE......EWSS-KL-FDGGEC-SSTISSP...................NGKELNKSRESLRNIQEPRTTTEE.....ATAGSTEMCGRAMPEFLL-SR--IEN-LVNR-PVS-----R----IFS-N-V..KIE-PYT-RVV-----Q-R-KE-KDVF---A  137
LST.CD.88.447              ---LWTLE......EWSS-KL-FDGGEC-SSTISSP...................NGEEPNTSRESLRNIQEPRTTAEE.....ATADPTEMCGRAMPEFLL-SR--IEN-LVN--PVS----------IFS-N-V..RIE-PYT-RVV-----Q-R-KE-KDVF---A  137
LST.KE.x.lho7              ---IWSLE.....QQTSRQLS-TGRS....CRE.NSISSP................NGECSNKSRESTGDLQDFRTTAEK...AAAASPTEMCGRALPEFTL.SR--IES-LVN--PVS----------IFS-N-V..RIE-PYT-RVV-----Q-R-KE-RDVF---A  137
SYK.KE.x.SYK173            ----RMGG.....LQEAPSQLP--ERAICSP-G.........GHRRWAMAHMVSTDEPTSTSEGTELSLEETPHQQG-ALA...EGEQWEGGDQE-LSLS-LS--R--MIEVDVE-D-VQM-V-------IIR-EDI..QLHQP-S--IV--L--N-T----RN-RFTVV  151
SYK.KE.x.KE51              ----RVGE......QETPSK--CDEGDTNSPTSGG..-AMAAT-MGKLCFTPTEAHSNGHPATATVLSKDSQGPLGPWETF...REERRDEGEGT-LSL-SIS--R--MME-NVE-TKVTM---------IIR-QDI..QLHQP-T--IV--L--N-R----N--RFQ-E  157
COL.CM.x.CGU1              ---GNQGGG.....-TDSHLPP..................................................EDEQR...PATQEGEGGE.-LSLLE-S-RR--I-IVEVE--KCQ-----------FK-EDV..QLT-A-TAVQ-Q----A-R-K--RDKILKMG  105
```

```
                                                                                 K70R
                          protease end  p66, p51 RT start            M41L   D67N                                          D110 catalytic site
H1B.FR.83.HXB2 GHKAI.....GTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETV...PVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKS......VTVLDVGDAYFSVPLDEDF 279
H1A1.UG.85.U455 -K-T-.....----------------M------------------...------E-------------------T---N---------------------------------------------------------T--------......-----------------S- 278
H1B.US.90.WEAU160 -----.....-----------------------------------...-----------------------------------------------------------------------------------------S-------......------------------- 279
H1C.ET.86.ETH2220 -K---.....----------------M---L-R------------...--------------------------TA--E---Q-----R--------------------------------------------------------......-----------------G- 286
H1D.CD.84.84ZR085 -----.....-----------------------------------...--------------------------T----D--------R----------I---------------------------------------------......I-----------I--C--- 279
H1F1.BE.93.VI850 -----.....----------------M-----------V------...--------------------------T---L-----------------------------S--------K---------------------------......----------------K-- 279
H1G.SE.93.SE6165 -K---.....--I------I------M------------------...-----------R--------------T---K---E----------------I---------------------------------------------......------------------- 276
H1H.CF.90.056 -K---.....----------------I------------------...--------------------------T-------------R-------S--I---------------------------------------------......-S--------------KE- 278
H1J.SE.93.SE7887 -K---.....----I-----------M---L--------------...------------I-------------TQ--A---E-----RV-------------------------------------------------------......---------------Y--- 277
H1K.CM.96.MP535 -Q---.....-----------------------------------...--------------------------T----------------------------------------------------------------------......----------------K-- 279
H101_AE.TH.90.CM240 -K---.....----------------M---------------D--...--T-----------------------T---K---E--------------------------------------------------------------......-----------------S- 284
H102_AG.NG.x.IBNG -K---.....----------------M------------------...--------------------------TD---------------------------------------------------------------------......----------------K-- 279
H103_AB.RU.97.KAL153_2 -K---.....----------------M---L--------------...--T-----------------------TD--K------------------------------------G-----------------------------......----------------Q-- 279
H104_cpx.CY.94.CY032 -K---.....----------------M---L--------------...--------------------------T----D-------------------I---------------------------------------------......----------------PE- 278
H1N.CM.95.YBF30 -R--V.....-----------------------------------...-------------------T---E--R-------------R----------I-----------------------------------------Q---......------------C---K-- 286
H1O.BE.87.ANT70 -REVQ.....----------------I--GL-----------AP-...------------------SK---E--TA--Q---Q-----R----------I-------G-----------------E-----------G---Q-Q-......------------C---P-- 275
H1O.CM.91.MVP5180 -KEVQ.....------------L---I--GL-----------AP-...------------------SR---E--TA--Q---Q-----R----------I-------------------------------------G---QRQ-......------------C---P-- 275
CPZ.CD.90.ANT DRTVL.....A---L--N--------VCLL----------KV---...-----E-----R------SK---E--K---DKL-A-N---R---D------I-------TS--------------------I-----------Q---......------------I---Q-- 279
CPZ.CM.01.SIVcpzCAM13 -K-VQ.....----------------I--------V----S----...-----------RI-----SA------T---Q--------------------I--------S--------------------------------E---......------------C---K-- 278
CPZ.CM.05.SIVcpzEK505 -K--V.....----------------F-----------------I...-----------R--------------T-------------R----------I---------------------------------------------......------------C----N- 285
CPZ.CM.05.SIVcpzMB66 -KR-V.....----------------I------------------...--S--------R-----------R--T-------------RV---------I-----------------------------------------Q---......------------C----N- 282
CPZ.CM.98.CAM3 -R-TT.....----------------I------------------...-----------R-------A------T---Q---------R----------I---------------------------------------------......------------C----N- 284
CPZ.GA.88.GAB2 -KRVE.....S---------------V---L-----------TP-...-------T---R------SK---E--T---R---------R----------I---R-----------------------------------------......---P--------C----N- 282
CPZ.CM.98.CAM5 -R-TT.....----------------I--------------V---...-----------R-------A------T---Q--------------------I-------------------------E-------------------......------------C----N- 280
CPZ.CM.05.SIVcpzMT145 -K-IT.....S---------------V--KL-----------D--...K-F---------------SK---E--T---K--------------------I-------------------------E-------------------......------------C---P-- 278
CPZ.TZ.00.TAN1 NKEVR.....A-----E--I------I-K-L------------V-...K-Q--E------------SK---E--T---KTL-------AV---------I-------TS--------------------L----------R-RNM......------------I---P-- 275
CPZ.CM.05.SIVcpzLB7 -K--L.....----------------I------------------...--Q-----------------------T-------Q-----R------------------------------------------------------R-......------------C----N- 282
CPZ.US.85.CPZUS -R--Q.....-S--------------I-------------S-K--...--R--------R-------A------T---Q--------TRV---------I-------------------------------------------R-......------------C---KE- 281
CPZ.GA.88.GAB1 -R-VV.....----------------I---L----V----S----...------------------SA------T---Q--------------------I---------------------------------------------......------------C---K-- 277
GSN.CM.99.CN166 EKRIK.....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--II-MK---Y-I--Y--- 290
GSN.CM.99.CN71 EKRIK.....A----------L---C-SKL-V---MVQQKL-P-...E-H--S-KE---I-----SK---E--TQ-TQ--L-L-QLE---------S---------KSQ--M-I---K--EA---LA------------EQ-EH......--II-MK---Y----Y-E- 283
MUS_1.CM.01.1085 -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......--IV-IK--FY-----P-- 280
MUS_1.CM.01.CM1239 EKTIK.....A--------I--M---C-SKF-I---MIQEK--PI...K-A--E-AK--M-------Q------EG-VQQ-L-LDQ-EE---D----S-C---R---KS---M-I---Q--EA--E-T------------AEHEH......--IV-IK--FY-----PA- 280
MUS_2.CM.01.CM1246 EKRIV.....A----S---I--L---C-SKLEI---MVMAGQQLK.PTK-S--ERKK--M------SK---E--K--TQK-I-L-QLEEA--N----S-----R---K----M-I---K--EV--E-A---T-----S--AQ-AH......--IV-MQ--FY-----KE- 285
MUS_2.CM.01.CM2500 DKRI-.....A--------I--L---C-SKL-I---MVGGQL-PT...K-G--A-KE--M------SK---E--E--TQR-L-L-QLEET--Q----S-A---R---KS---M-I---K--EA----S---T-----S--S--AH......--IV-MQ--FY-I--Y-E- 284
                            protease  p51 RT start                                                                                 D catalytic site

MAC.US.x.239 -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----I------I----E- 291
H2A.GW.x.ALI NKRVK.....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-----------I--H-S- 304
H2A.DE.x.BEN NKRVR.....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---------I--H--- 304
H2A.SN.x.ST NKRVR.....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-----------I--H--- 304
H2B.GH.86.D205 -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#-------fAE-RR......I--I--------I---PN- 303
H2B.CI.x.EHO -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---------------P-- 304
H2G.CI.92.ABT96 -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----I------I---V-- 287
H2U.FR.96.12034 -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-----------I--C--- 301
MAC.US.x.251_1A11 -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----I------I----E- 291
MAC.US.x.251_BK28 -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----I------I----E- 287
MAC.US.x.EMBL_3 -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----I------I----E- 287
SMM.US.x.H9 -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-----------I----E- 287
SMM.US.x.PGM53 -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-----------I----E- 287
SMM.US.x.SIVsmH635F_L3 -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-----------I----E- 287
SMM.SL.92.SL92B -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----I----------PE- 284
STM.US.x.STM -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-----------I----G- 287
MNE.US.x.MNE027 -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-----------I----E- 287
GRV.ET.x.GRI_677 DKIIT.....-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--I-I----Y-I--CKE- 301
VER.KE.x.TYO1_patent DKILR.....--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---------Y-I---PN- 299
VER.KE.x.9063 DKILR.....----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----I----Y-I---PE- 296
VER.KE.x.AGM155 DKILR.....----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-II------Y-I---PE- 299
VER.DE.x.AGM3 DKILR.....--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----I----Y-I---PE- 296
TAN.UG.x.TAN1 -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----I----Y-I---PE- 299
SAB.SN.x.SAB1C DKTCK.....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----I------C---P-- 311
OLC.CI.97.97CI12 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----I------I---PEY 281
DRL.x.x.FAO -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-----------C---P-- 293
RCM.NG.x.NG411 -KTTH.....AH--------------V-SKL-AS-----KA--I-...K-E----Q---R------SK---E--T---NA--------R----------I-C-------------------------F---------G---QCRQ......I----I------C--Y-P- 297
RCM.GA.x.GAB1 -KRTH.....AH--------------V-KKL-A------Q---I-...K-E--S-Q---R------SK---E--T---NA-------------------I-C-------------------------F---------G---QCER......I----I------CL-Y-P- 293
MND_1.GA.x.MNDGB1 -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----I------I---P-Y 288
MND_2.x.x.5440 -KITH.....AA--------------V-KKL---------K----...K-T----T---RI-----SK---L--Q---NQ--------RV-----------C-----G-S--------Q---V----F------T--G---QCEQ......I----I------C---V-- 292
MND_2.CM.98.CM16 -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----I------C------ 291
MND_2.GA.x.M14 -KSTH.....AS--------------V-KKL---------K----...K-T--------RI-----SK---L--QT--D---------R------------C-----G-S--------Q--RV----F---------G---QCEQ......I-I--I------C----S- 292
DEB.CM.99.CM40 NKV-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----LA---Y-IA--PT- 285
DEB.CM.99.CM5 DKV-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----LA---Y-I---PQ- 285
DEN.CD.x.CD1 DKV-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----LK---Y-----P-- 305
TAL.CM.01.8023 HKR-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----MK---Y-I---P-- 289
TAL.CM.00.266 NKR-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----MK---Y-I---P-- 288
MON.CM.99.L1 DREVW.....A--------I--L---V-AKM-V---MAAGGDLQPPSLE-T--A-KE---L-----SR---E--TQ-TQ--L-L-QLEPTE-N----S-C-------KS---M-I-------A----F------------E-MDH......--IV-MK--FY-I--W-P- 287
MON.NG.x.NG1 -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-IV-MK--FY-I----N- 290
SUN.GA.98.L14 DKT-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--V--------I---P-Y 297
WRC.CI.97.97CI14 -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-V-I------I---PEY 303
WRC.CI.98.98CI04 -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-V-I------I---P-Y 302
WRC.GM.05.Pbt_05GM_X02 -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----I-----FI---PEY 294
LST.CD.88.524 -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--V-I------I---P-Y 296
LST.CD.88.485 -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--V-I------I---P-Y 295
LST.CD.88.447 -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......---V-I------I---P-Y 295
LST.KE.x.lho7 -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--V-I------I---PNY 295
SYK.KE.x.SYK173 KPSGKRKQVE--L---------L---I--KL-VK-VMVQTAL-P-...K-S---DKEL-RL-----SV--LE--KA-VED-L-A-QLE-AS-T--------V-R---KK---M-I---K--EA----F---------G---RQ-.......L-II-LK---Y-----KE- 311
SYK.KE.x.KE51 RPDGRIKEVE-SL---------L---I-SKL-AK-VMIAQKL-PM...K-Q----KEL--L------R--LE--RA-VSD-LEK-QLE-AS-T--------V-Q---KL---MPM---K--EA--------------G--ST-R.......M----LK---YT-----E- 317
COL.CM.x.CGU1 -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----I------I---P-Y 261
```

```
                                                                                                          K219Q
                                                                    polymerase motif                T215Y
H1B.FR.83.HXB2            RKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDS..WTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEV 447
H1A1.UG.85.U455           -----------------V---------------S-------------S-H-------------------------A------A---S--FI------------------------------Q------.---------------------A----K--------A----DI 446
H1B.US.90.WEAU160         ---------------------------------------------------------------------------------------F---------D-----------------------K----E-.---------------------A----K--------------I 447
H1C.ET.86.ETH2220         ----------T-------------------P------PQ------AP--E----------------------AP-----E---K--F------------------------------Q------.-----------------------------------A----DI 454
H1D.CD.84.84ZR085         ----------------------------------I----------EV----------------------A---K--E---------------------------------------S-T----E-.---------------------------K-------------- 447
H1F1.BE.93.VI850          K---------V---------------------C---------MK------------------------------E------F----------------H------------Q--N---.----------------------------P-------A----DI 447
H1G.SE.93.SE6165          -----------------V---------------------R------AN--EM----------------------A------E---K--F------------------------------Q--D-E-.---------------------------TH-------A----DI 444
H1H.CF.90.056             ----------------------------------------A---E---EM----------------------A------A---K--F----------------------------TVK------.--------------------N---K---------A----DI 446
H1J.SE.93.SE7887          --------------------------------C------K---ER--E-------------------E---R--K---E---K--F------------------------------Q----ED.----------------------------K------K-A----DI 445
H1K.CM.96.MP535           -----------------V---------------H----------IK--EM-------------------P--------E---K--F------------------------------Q--D---.---------------------------K--------V----DI 447
H101_AE.TH.90.CM240       ------------------------------------------IK--EM-----K----------------------A---S--F-----------------------R------E------.--------------------A----K--------A----DI 452
H102_AG.NG.x.IBNG         ----------V--------------------A-----------TK--E---------------------A------G---K--F------------------------------VE------.---------------------A---IK---R----A----DI 447
H103_AB.RU.97.KAL153_2    ----------T---------------------------------------E-----------------------E-----E-----F--------------------------------------.---------------------A--------------A------ 447
H104_cpx.CY.94.CY032      ----------T---------------------C----------FK--E---------------------A------E-----F------------------------------QPA----.----------------------------K--------A----DI 446
H1N.CM.95.YBF30           ---------------------------------T---------EKH-E-I--------------LA---EAV-D--D---K--F------------------------------K-----V..------------------------R-K-----I--------- 454
H1O.BE.87.ANT70           ----------V-----------------------D---RD--ELE-C------------PLTE--KR--L--E--YQ--F---------------------------S-Q--N--V..----------I----------Q--R--E----I----S---- 443
H1O.CM.91.MVP5180         ----------V------V-----------------D----S--EVE----I----------PLAE--KRV-L--E--YQ--F------------------------------Q--D-EV..-----------------------Q--R-KE----I----S---- 443
CPZ.CD.90.ANT             ----------V---------C---------------A------A---DKY-AVE-------------M--TA--EM--K-----QV---E------------Q--------K-K---P-D..--------------------------TK---R-I--V-S--DR 447
CPZ.CM.01.SIVcpzCAM13     ----------V-----------------------------------AK----TV--------------D-T---Q-V--------H--F------------------------------K----EV..-----L---I--------------IK-----I----R--D- 446
CPZ.CM.05.SIVcpzEK505     ----------V-------------------T---------N--EL---------------------T---EAV-R--S---T--F------------------------------T-Q-----T..--------Q------------------K-----I--A------ 453
CPZ.CM.05.SIVcpzMB66      ----------V--------------------------------------E-I----------------K-EL--E-V----A---K--F------------------------------Q----E-..----------I------C------R-K-----I--------- 450
CPZ.CM.98.CAM3            ----------V---------------------I---T---Q-H--LI--------------LRE--E-V-L------T--F------------------------------Q--Q-ET..---------------------------K-----IK------DI 452
CPZ.GA.88.GAB2            ----------V--------------------------V-D---QK--EL---------------SLEE--KRV-Q--E------F------------------------------Q--Q-EN..-------------------------IK-----I--A-----I 450
CPZ.CM.98.CAM5            ----------V------V-------X-----------------D-----H--LI--------------------K--E-V-L------T--F------------------------------Q--Q-EI..---------------------S----K-----IK-------- 448
CPZ.CM.05.SIVcpzMT145     --F-----------------------------S-------------TKH-E-I--------------P-EE--KRV----A---K--F----------------------------Q-K---QEV..--------------------------IK-----I--A-G---- 446
CPZ.TZ.00.TAN1            ----------L--N---K-F--------------------D----EH--VD---------I----NEEE--KL-KK------T---E-----Y-EK---M--------N-----N-T---PEQ..----H---------------H---TKE----I--V-G---P 443
CPZ.CM.05.SIVcpzLB7       -------------------------------------------R---------------------K-E---N-V----A---K--F------------------------------Q----E-..---------------------------IK-I-R----V-NS--- 450
CPZ.US.85.CPZUS           ----------V------------------------------D-----H--VI--------------NLEK--E-V-L---Y--T--F--------E-------------------Q--Q-EI..---------------------------IK-----IK-A------ 449
CPZ.GA.88.GAB1            -----------------V---------------S-------------EK----T-----------------D---K-V--------K--F------------------------------Q----EV..----------I--------------IK-----I----K--D- 445
GSN.CM.99.CN166           ------SV--V--QA-AE---FK-------A--T---NIVASL-RQI-D-E--V-LI------LI---RKL-E--QVVHKI-NL-TSYNIQ--EA---PDY-VK-L------KG-RL--VE--DQ-T..---------------V---H--RTK---CI--A-G--DQ 458
GSN.CM.99.CN71            ------SV--V--Q--AE---FK-------A--T---ATVASL-YQI-D-E--V--I------LI---RKLAE--QVVHKI-NL-TSYNIQ--EA---QDY-VR-L------KG-RL--VE--DQ-I..---------------V---H--RTK---RCI--A-G---Q 451
MUS_1.CM.01.1085          ------SL-QV--LG-AR---FK----------T---FTAA-L--EIK-A---VTLI------LI--NR-LSS--KVVAQI-NM--KL-IQ--QE-F-EDY-VQ-L------RG-KI--VA--QQEV..-----V----------A-A-----TK----CI--V-S---E 448
MUS_1.CM.01.CM1239        ------SL-QV--QG-AR---FK----------T---YTAA-L-QEI-EA----TLI------LI--NR--SG--RVVAQI-NM--KL-IQ--ED-Y--DY-VQ-L------RG-KI--VT--QQEN..-----------------A-A---V-TK----CI--V-S---E 448
MUS_2.CM.01.CM1246        -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---TKH--RCI--I-G---L 453
MUS_2.CM.01.CM2500        -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---TKH---CI--V-----T 452
                                                                    DD catalytic site

MAC.US.x.239              -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-------TKH--R-I--KMT---E 459
H2A.GW.x.ALI              -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-------TKH--R-I--KMT---G 472
H2A.DE.x.BEN              -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---TKH--R-I--KMT---E 472
H2A.SN.x.ST               -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-N--R-I--KMT---E 472
H2B.GH.86.D205            -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-HI---I--KMT---E 471
H2B.CI.x.EHO              -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-HI---I--KMT---E 472
H2G.CI.92.ABT96           -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-----TKN-X--I--XMT---E 455
H2U.FR.96.12034           -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-----TKH----IK-KMT---E 469
MAC.US.x.251_1A11         -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-------TKH--R-I--KMT---E 459
MAC.US.x.251_BK28         -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-------TKH--R-I--KMT---E 455
MAC.US.x.EMBL_3           -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-------TKH--R-I--KMT---A 455
SMM.US.x.H9               -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-------TKH--R-I--KMT---E 455
SMM.US.x.PGM53            -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-------TKH---I--KMT---E 455
SMM.US.x.SIVsmH635F_L3    -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-------TKH--R-I--KMT---E 455
SMM.SL.92.SL92B           -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-------TKN---MI--KMT---E 452
STM.US.x.STM              -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-------TKH----I--KM----E 455
MNE.US.x.MNE027           -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-------TKH--R-I--KMT---E 455
GRV.ET.x.GRI_677          ----------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------------TKHT-AM---K-N-L-E 470
VER.KE.x.TYO1_patent      ---------TV--QG------F-C---------T---NTAAS---EIKRNL-ALT-V------W---QENEHT-DKLV-Q--TK-QA---E--E--V-----YE----K-W-H--ELSR-Q-E---E..----------------A-L---LRTKNI---I--K-N-L-L 467
VER.KE.x.9063             ----------V--QG--V---F-C---------T---NTAA----KIKEELKPLT-V------W---QEDEYT-DKLV-Q--VK-KT---E----V--K--YE----K-W-Q--QLNS-N-E--EQ..----------------A-L---LRTK-I-R-I--K-N-L-E 464
VER.KE.x.AGM155           ---------TV---G------F-C---------T---NTAS----EIK-ELKQLT-V------W---QE-GPK-DQLVQT--NR-QE---E--E--V-R----E----K-W-H--KL-S-E-EK-EQ..-----L----------A-L---LRTKNI------K-N-LD- 467
VER.DE.x.AGM3             ----------V--QG--T---F-C---------T---NTAAS---EIK-ELKPLT-V------W---QEDEYT-DRLV-Q--MK-SA---E----V--K--YE----K-W-H--QISS-E-ED-EE..-------R-------A-L---LRTKN----I--K-N-L-T 464
TAN.UG.x.TAN1             -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-----TKA-------AQD--A- 467
SAB.SN.x.SAB1C            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---TKH--R-I--ARP---I 479
OLC.CI.97.97CI12          -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---TKA----IK-P-G-LDK 449
DRL.x.x.FAO               ----------V--QG---------------------ATAD---Q---EK---VI--------F----RPKHL-EQM-K---T--KF---E-----F-DN--YE----Q---K-----E-R--D-EI..-----------V--------S---TKE----I--A-P-D-E 461
RCM.NG.x.NG411            --------LA---QG---------------------TTAN--------E---V---------F----RTAHE-SQM-KR--E---T--FE-----F-DK---E----V----------E-K-L--EE..--------------------S---TKE----I--V---DDK 465
RCM.GA.x.GAB1             ---------AV--QG--V------------------A-AN---Q---EE---VI--------F----RTKLE-DKM-KQ--D---F--FE-----F-DK--Y----------KS----E-K----EE..--------------------S-LRTKE----I--A---D-K 461
MND_1.GA.x.MNDGB1         -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------TK-----I--GLNI--K 457
MND_2.x.x.5440            ----------V--QG---------------------ATAD---KA-KEKH-EVL---------F-----SASE-SKM-V---E--QF---E-----F------E----V---K-----K-Q----E-..--------------------S---TKE----I--ARP-DKI 460
MND_2.CM.98.CM16          ----------V--QG---------------------ATAD---Q---ERH--V----------F----RVAPE-SRM-Q---D---F---E-----F------E----I---K-----KVQ----EE..--------------------S---TKE--R-I--A-P-D-K 459
MND_2.GA.x.M14            ----------V--QG---------------------ATAD---KT-KEEY-EVL---------F-----SASE-NKMVNR--E--RF---E-----F------E----V---Q-----K-Q----ET..--------------------S---TKE----I--A-P-D-- 460
DEB.CM.99.CM40            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---TKH--RAI--VPG--DP 453
DEB.CM.99.CM5             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---TKH---AI--VPG--DK 453
DEN.CD.x.CD1              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---TKE---SI--V-D-G-T 475
TAL.CM.01.8023            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---TKA---C---V-G-L-E 457
TAL.CM.00.266             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---TKA--RY---V-G-L-E 456
MON.CM.99.L1              ------S---L--AE-AK---FK----------TV--HTAAELPQEI--KY-EVTLI-------LI--NHSLPE--KIVGVI-AT---K-IQ--PE-F-DDY-VQ-L-------SH-KIAK-E--DQEQ..----E----I------A---S---TKN---SI--ARS--DT 455
MON.NG.x.NG1              ------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-------TKH--XCIX-X----DI 458
SUN.GA.98.L14             ----------V--QA--K--V------------C---GTVASL--K--R-Y-EVQL--------LI---Y-KKK-EEIVKQ---L-ME-N-E--E--Y-G---YK----I----R-EIEK-K--PLEEE.P-------I--VI-----L-E-LRTKE----I--N-P-A-K 466
WRC.CI.97.97CI14          -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--QTK---R-I--N-P---E 470
WRC.CI.98.98CI04          -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-------I--N-P---E 469
WRC.GM.05.Pbt_05GM_X02    -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---TKE----I--N-P--AE 461
LST.CD.88.524             -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--RTKE----I--V-P-EQL 466
LST.CD.88.485             -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--RTKE----I--V---E-- 465
LST.CD.88.447             -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--RTKE----I--V---D-- 465
LST.KE.x.lho7             ----------L--QE--K---------------C---GTVAGL-SE---L---MI--------FI---R-RKG-DQAVK---EL-MT-N-E--E--F-A---YH----V-----R-EIEK-K---M-LTKT---Q------V----A-L-D--RTKE----I--V-P-E-I 465
SYK.KE.x.SYK173           -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---TKE---CI----Q---- 480
SYK.KE.x.KE51             -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------------CKA-S-CI--A----DI 486
COL.CM.x.CGU1             Q----------V--LA--T--V-K----------T---GTVAR-------RGQL.Q-A--------L-----PLEE--RAVQ---DA--YY--E--E--Y-AD--YK--------KQ-KL-KVKI--Q-K..------------A---L--L----RTKEIS-CTK-K-H-L-E 428
```

```
                                                                                                                                                                                        p51 RT end  p15 RNase H start
H1B.FR.83.HXB2             IPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEP.FKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTP.KFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTN  615
H1A1.UG.85.U455            VT-----------------D-------------V--------D----------.-----------K-S-----------V---VS---------I-.--R---------A--M----------------------------D--A-------------------------D  614
H1B.US.90.WEAU160          --I------------------------------------L----------------.-----------V-----------------------------------------.---------------------------------------------------------K-----------  615
H1C.ET.86.ETH2220          VT---------------------F------------------ND---F-F----.--------F-KRGT----------AV----AL----------.--R---------A---D----------------------------A-V-------------I-------D  622
H1D.CD.84.84ZR085          --------------------M-------------L------------------.-----------------------------AI----------.--R------------ID----------------------------I-----------K-----------D  615
H1F1.BE.93.VI850           V---A-------K-----R-----------------------D----------N-.-----------KV-S----------------AL--------RSP------L----D----D------------------------T---A--D-------S-----K-------D  616
H1G.SE.93.SE6165           VS--A---M---------R-----------E----V----LD----------.Y---------KRGS-----------V----A---------------------R-----I---D--------------------R--T---P----Y------------------D  612
H1H.CF.90.056              ----K-------------R--I-----------------R-----------------.-----------K--T-----I---------S---------I-.--R-----------------------------------H---------T---A----Y-I--------------------D  614
H1J.SE.93.SE7887           V---R---------K-------------SA-E----V----LD----------.-----------KR-S---------A-V----AL-A--------.--R----R-------D--------------------------------M---------S-----T-------D  613
H1K.CM.96.MP535            V---A------------------------------------ND----------.H-----------S----------------A--G--------.--R--------------------------------------T----------------H----K-R-----D  615
H101_AE.TH.90.CM240        V-----------------T-------------V--V-----D----------.-----------RGS------R----V---VA-----------.--R----R-------M-------------------------D--------------S-------------D  620
H102_AG.NG.x.IBNG          VA--------------------------T---V--L-----D----------.-----------KK-S-----------V---VAM----------.--R----R-------M-------------------------D--------------------I-------D  615
H103_AB.RU.97.KAL153_2     ----A---------------------------V------------------------.-------------L-----------------A-------------.---------------------------------------------------------------S-------D  615
H104_cpx.CY.94.CY032       V---T-------------------A------------------------------.H----------T-S------R---------AM-C--------.--R---------D-----------------------------TD--A--------------Q-------D  614
H1N.CM.95.YBF30            VNF-----------------L-------G-E-V-------------------L.H---------K--S-----I---V-V-R-VA-----------.--R--V---V--A---DH------------------------T---S----------------------F--D  622
H1O.BE.87.ANT70            V--SR------E----R--Q-------Q-D---WVN-----GE---------E.H--------T-QKAS----IR--A-VI--VSQ---I----L-.-----VTR-------AD-------------S----I----R--S---M----Y-------------------E  611
H1O.CM.91.MVP5180          V--SK------E----K-----------Q-D---WVS---H-E------V--DE.H-----------QKAS----IR--A-V---VSQ-A-----L-.--R--VTR-------A--------------S----I-------T-------------------N---------E  611
CPZ.CD.90.ANT              VQM-R------E--KQ--QQKIE-Y--Q-GLP-K-T-----S---------NE.G-L--A-----PT-T---E-R--AGV----GL-------EV-.--Q---TR---DA--SD------------------IR---N-LAD--PE-----------NSQ--------D  615
CPZ.CM.01.SIVcpzCAM13      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-----------K-----I-D  614
CPZ.CM.05.SIVcpzEK505      VT--R-------------------A--N-D-E------------------DL.H---------K--ST----IR----V---VAL----------.--R--V---V----------------D----------------T---S----Y----------------F--D  621
CPZ.CM.05.SIVcpzMB66       VTF-T-------------------A------E-----------------F--Q.Y-------------S-----------V---VAL--------V-.R-R---------A---D-----------Y-------------QD--P--------------------------D  618
CPZ.CM.98.CAM3             VTM-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------------------K-------D  620
CPZ.GA.88.GAB2             VEM-R------E-----------T----E-E----V---NN-------F--T.GR---------Q-A-P---IR--A-V-P-VSP---IL--TV-.-----V-R-V--A---D--------------R----------------Q----Y----------RR------I-  618
CPZ.CM.98.CAM5             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------------------K-----I-D  616
CPZ.CM.05.SIVcpzMT145      VA--P----------Q--------A----E-E-------------------F---.H---------GKQ-A-----IR--A-T----A-------------.--R--V---M--A--AD-----------I--------------T---E----Y------------R--------  614
CPZ.TZ.00.TAN1             VEM-R------E--KQ----K-Q-A----KLP-Q-A-----------------E.G----------KSP-T---EIR--AGLI---GN---I---IV-.--L--VS----SQ---D---V--V-----I------IR---N-LSD--PE-------------DS-K-R------  611
CPZ.CM.05.SIVcpzLB7        VTF-R---------------N--------A-------V-----S----------Q.------------QGS----------A-VM---AI--------I-.--R-------A---A---D--------------Y----SH--F-----Q---P--------------------------D  618
CPZ.US.85.CPZUS            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-------------Q----K-------D  617
CPZ.GA.88.GAB1             V---P------------VST--------D-E---------NC-------F---.H----------Q-S-----IR--A--------A-------------.--R--V---S--A--A-------------I-----------S--T---PTTD-Y-----------T-------D  613
GSN.CM.99.CN166            VE-----QI----------QSSE-G---AE-P-VV--TSL-EQ--G-MFS--G..-M--SA-F-KT-N--I-SYQ-FAD-LT-VGR--Q-T---V-D--RI-VV--Q-DA--MN--------TI-AIH--H-LRQ--T-VS--LED-V-Y---------TS----------  626
GSN.CM.99.CN71             VE-----QI----------QSSE-S---AE-P-VV--TSL-EQ--G-MFS-DS..-M-RS--F-KT-T--I-SYQ--AD-MT-VGR--Q-T---V-D--RI-VV--Q-D---MN--------TI-A-H--H-LRQ--T-VS--LEN-V-Y---------TS----------  619
MUS_1.CM.01.1085           VQ-----QA-----Q---QQT-Q-A----EQP-V---VSL-DS--G-NFT-SK..GV-R---F-KV-SV---SYQ--AD-MA--GR--L-----P-D-VRI-VI--Q-DQ--IAH--VS---TI-A-H-TH-L-Q-FT-VP--LKE-P-Y---------VS----------  616
MUS_1.CM.01.CM1239         VQ-----Q------Q----QA-Q-A-F--EQP-V---VSL-DS--G-NFL-NK..GI--S--F-KV-SV---SYQ--SD-IAR-GR--L-----P-E-VRI-VI--Q-DQ----H--VS---DI-A-H-TH-LRQ-FT-VP--LSE-P-Y-------HKVS----------  616
MUS_2.CM.01.CM1246         VE-S-Q-Q-----------QEAG-A----E-P-VL--VSL-EQ--G-TFT-DR..NM-R---F-KI-T--S-PYQ--A--LSRASK-AL-C-----D-CRI-VV--Q-DN--ADS--T----DI-A-H--Y-LRQ-FT-VP---ET-P-Y---------NS---R-----D  621
MUS_2.CM.01.CM2500         VE---Q-Q---------M-QE-K-A----E-P--L--VSL-EQ--G-TFT--K..NM-----F-KV-T--I-TYQ--A--LA-AAR--L-C-----D-CRI-VV--Q-DH--A-S--T--V--I-A-H--H-LRQ-FT-VP--LEA-P-Y---------NS---R-----D  620
                                                                                                                                                                                       p51 RT end  p15 RNase H start
MAC.US.x.239               VQW--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-SQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QV-.--H--VE-DV--Q---D---V------D-IS-----R-VFN-V-D--E-E--Y-T--SC-KQS-E-----I-D  626
H2A.GW.x.ALI               VQW--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--PC--QS-E-----I-D  639
H2A.DE.x.BEN               VQW--L--A--E--KI--SQEQE-Y--QEE-E-E-T---SQGH----K-H--E..-I--V----KIKNT---G-RL-AQV----GK-AL----RI-.--H--VER----Q--DN---V------D--S-----R-TFN-VGD--P------T--SC--QS-E-------D  639
H2A.SN.x.ST                VQW--L--A--E--KI--SQEQE-C--QEE-E-E-TV--DQDN----K-H-GG..-I--V----KVKNT---G-RL-AQV----GK-AL----RI-.--H--VERD---Q--DN---V----D-D-IS-----R-VFN-V-D--L------T--SC-KQSRE-----I-D  639
H2B.GH.86.D205             VQW--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--SC--TS-E-------D  638
H2B.CI.x.EHO               VQW--L--A-FQ--KI--EQEQE-S--KEGVP-E-TV--NLAN----K-H-GD..-I--V----KVKNT---G-RL-AHV----GK-AL----EI-.M-H--VER---DQ---D---V------D--S----IR-A-N-V-D-LE-V--Y-T--SC-KAS-E-------D  639
H2G.CI.92.ABT96            VQW--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--SC-KTS-X-------D  622
H2U.FR.96.12034            VQW--M--A-F---KI--SQEQE-A--REEEN-E-TVL-NQDN----K-H-GD..RT--V--F-KVKNT---G-RL-AHV----GK-A------I-.M-H--VER-V--Q---D---V--VLD-D--S-----R-A-N-V---LL.E--Y-T--SC--AS-Q-------D  635
MAC.US.x.251_1A11          VQW--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-NQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QV-.--H--VE-DV--Q---D---V------N-IS-----R-VFN-V-D--E-E--Y-T--SC-KQS-E-----I-D  626
MAC.US.x.251_BK28          VQW--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-SQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QV-.--H--VERDV--Q---D---V------D-IS-----R-VFN-V-D--E-E--Y-T--SC-KQS-E-----I-D  622
MAC.US.x.EMBL_3            VQW--M--A-YE--NI--SQEQE-C--QEG-P-E-TVI-SQDN----K-H--D..-I--VR-F-KIKNT---G-RL-AHVI---GK-A---V-QV-.--H--VERDV--Q---D---V------D-IS-----R-VFN-V-D--EVE--Y-T--SC-KQS-E-----I-D  622
SMM.US.x.H9                VQW--M--A-YE--KI-PSQEPE-C--QEG-P-E-PVI-SQDN--S-K-H--D..-I--V--F-KIKNT---G-SL-AHV----GK-A-I---QV-.R-H--VEX-I--Q---D---V------D--S-----R-VFN-V----Q--------RSC--QSRE-------D  622
SMM.US.x.PGM53             VQW--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-SQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHV-E--GK-A-----QV-.--H--VER-I--Q---D---V------D--S-----R-VFN-V----Q---------SC--QSRE-------D  622
SMM.US.x.SIVsmH635F_L3     VQW--M--A-YE--KI--SQEQE-C--QEG-PIE-TVI-SQDN--S-K-H--D..-V--V--F-KVKNT---G-RL-AHV----GK-AL----EV-.--H--VER-I--Q---D---V----D-D--S-----R-VFN-V----Q---------SC--QSRE-------D  622
SMM.SL.92.SL92B            VQW--L--A------I--NQEQE-R--REDEP-E-TVL-NQDN--S-K-H-GD..RI--V--F-KIKNT---GIRL-ANV----GK--L--------.F-H--VER-V-DQ---D----------D-IS----IR-VFN-V-D--EKE-VY-I--SC--NS-E-------D  619
STM.US.x.STM               VQW--M--A-Y---KI--SQEQE-Q--RED-P-E-TVV-DQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QI-.--H--VER-V--Q--AD---V------D--S-----R-VFN-V---LE-T-------SC--QS-E-----I-D  622
MNE.US.x.MNE027            VQW--M--A-YE--KI--SQEQE-C--QEG-P-E-TVI-NQDN--S-K-H--D..-I--V--F-KIKNT---G-RL-AHVI---GK-A-----QV-.--H--VE-DV--Q---D---V------D-IS-----R-VFN-V-D--E-E--Y----SC-KQS-E-----I-D  622
GRV.ET.x.GRI_677           -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------SKV--------LSE  637
VER.KE.x.TYO1_patent       VTW-P---A-Y---A----TEQE-T--K-GIPIR-AV--LEG---S--FK--G..QV--V----TKQKNT---ELRT-AGL----CK-AL----IL-.VLE---ER-V--Q--AD---VS------D--S----L-----T-T----PKEDVY-----C--NS-E-----ISQ  634
VER.KE.x.9063              VQW-----A-Y---K----TEQE-T--A-G-P-R-AV--LED---S--FK--G..-I--V----KQKNT---ELRT-AGL----CK-AL----EL-.I-E--VER-V--Q--AD---VS--------SV--------T-T----P-EDVY-----C--NS-E-----I-Q  631
VER.KE.x.AGM155            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-----C--QS-E-----I-Q  634
VER.DE.x.AGM3              VTW-----A-Y---K----TEQE-T--K-GRPIR-AV--LEG---S--FK--G..QV--V----TKQKNT---EFRV-AGL---LCK--L----EL-.VLE---ER-V--Q--AD---VS---D---S---------T-T----PKEDVY-----C--NSRE-----I-Q  631
TAN.UG.x.TAN1              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---C--NS------FL--  634
SAB.SN.x.SAB1C             VQW--------E---Q--RQKQQ-Q----ALP-R-KVL-L-D---G-----PE.N-I--V----KIKT----ELRM-AGL----GK-------QI-.IME--VER-L--Q--SD---V-------M-S--Q-IR---K-V-D--P-EAVY--------NS-E-----L-D  647
OLC.CI.97.97CI12           VEW----KE--E--------EQE--M-KEEEP-QCTLCYTKGTVAYRIWQKHG...T----RFQ-KMT----KFR-M-G-IE--GK--------L-.I-QV-TTRKD--A-AFDD--VH---DI---SNSENQRTFWE-VPA-M.EV--------Y--KS-E------SS  614
DRL.x.x.FAO                VEW-R-----YE--KL-I--QM-----Q-E-P-K-K---L-N---S---E-DD.N-P-------KVKNS----MRM-AGL----AK-AL----RL-.I-Y--VER-V--Q--Q-----V----D---IS--H-I----N-LSD--P-E-VY--------VS-E-------S  629
RCM.NG.x.NG411             VEF-R-----YE--KAV---KI-----N-E-P---KV--LEG---S---E-GD.S-P-------KQKT----EIRM-AGL----CK-A------L-.--E---ER-V--Q--SD---VS--------S----IR---N-V-D--P-E-VY--------NS---------D  633
RCM.GA.x.GAB1              VEM-K---I-YE--KM----KL-------EK-P-V-N---LEG---S---E--S.G-P-------KQKT----EIRM-AGL----AK-A-----RL-.T-R---ER-V-D*-RSQ---V----D-----S----IR-G-N-V-D--P-E-VY--------NS-I--------  628
MND_1.GA.x.MNDGB1          VTM----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--------DS-M-----I-D  624
MND_2.x.x.5440             VEW-R-----YE--KL-VQ-E------Q-E-P-M-KV--LT----S---E--D.N-P--V------TKN----ELRV-AGL---VAK-CL-----L-.--Y--LER-V-DQ--HD---V------D-IS----IR---N-L-D--P-EDVY--------NS-E-----Y-S  628
MND_2.CM.98.CM16           VEW-R-----YE--KL-VQ-E------Q-E-P-M-KV--LT----S---E--D.N-P--V------TKN----ELRV-AGL----AK-AL-----L-.--Y---ER-V-DQ--P-------------S--H-IG---N-LR--VP-EDVY--------NS-E-------A  627
MND_2.GA.x.M14             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--------NS-E-----Y-A  628
DEB.CM.99.CM40             VT-S---QA--R--K----QE-S----KEQEP----LS-L-K---G-V-R-PK..GI-----FS-DK---Y--FH--AK-MY--G-----F--RI-.Q-R--VV--E-DN--HNH---A---D--AIH--H------E-VS---PD-D-Y----------S---------E  620
DEB.CM.99.CM5              VS-S---QA--R--K-TP-QE-S----KEDEP--S--S-I-R---G-V-R-TK..GL-----FSKDK-T-Y--FH--AK-MY--G-----Y---V-.--R--VV--E-DN--HNH-------D--A-H--Q--R---E-VS---PD-D-Y----------S-M-----ISE  620
DEN.CD.x.CD1               -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------Q--Q-R----AT  642
TAL.CM.01.8023             VT------A------A--A-ET--S--KEGEP-E--LT-LAE---G-MVK-GK..RL-----F-KQ-T--S-PYQ--VG-M---GK-A--F--QV-.I-RV-VV--E--Q--S----CS-L-TIVPIH-----R---N-VS---E----Y------H--S-E-------T  624
TAL.CM.00.266              VQ-S----A---------A-ET--T--QEGEP-E--LT-LAE---G-M-K-GK..RL-----F-KQ-T--S-PYQ--IG-M---GK-----F--QV-.V-RI-VV--E--Q----H--CS---TIVPIH-----R---N-VQ---E--D-Y------H--S-E-------A  623
MON.CM.99.L1               VV-SDL-QA-----------EST-G----NEP-KV-LTSLAEN--G-RFF--K..SV-----F-KI-ST-S-TYQ--AD-LA-LGK-A--T--RL-.I-R--VV--Q-DA--ADN--IN-V-DI-A-YS-H-LRQ--T-VQD--E--P-Y---------TS-E---------  622
MON.NG.x.NG1               VE-SDL-QA---------SQQST-G----QL--KV--TSLXE---G-XYF--..XRV-----F-KIKN--S-TYQ--AD--T-LGK-AL-----L-.I-R--VN--Q-D---ADN--V-----I-P-HS-H-LRQ--T-VQX--EH-P-Y---------AS-E-------S  625
SUN.GA.98.L14              VKM----RE-YQS-Q-V-Q-S-S-S--E-D-E--CRV--VK--IL-F-WL-G..KQV-RV-R-QKRGA--E-PCQ--AA-L---GR-------FV-.-IQV---R-I-SQ--AD---C-----L---S--K-EQE--T-AT--VP-.D-Y------EKLE-R-----I-Q  632
WRC.CI.97.97CI14           VEW----LE--EI-KQ----EQS-A--KGKDLIAEL-KTGPGQWAYTIREIGE.K-P--V---G-Q-HT-S-ELRSMAH-M---G--------RV-.--RI--E-DI-DS--SD---VS---DT-H-S--F-ARE--K-VL--LPEV--Y-I--S----DS-E------.WT  637
WRC.CI.98.98CI04           VTW----LE--EI-KS----EQM-A--TGKELVAEL-KTGSGQWAYSIREAGE.K-P--V---G-Q-HT---ELRS-AH-M---G---------V-.---I--E-DI-DS--SDH--VS---DT-H-S--Y-ARE--KIVL----EV--Y-I--S---DS-E----F.WT  636
WRC.GM.05.Pbt_05GM_X02     VEW-M---A--EI--A----SQQ-S--GGGDLIAEV-KTGTGQWAYTIKEAGE.T-P--V--FT-I-HT---ELRT-AHCI---G--A------L-.--RI-VE-DI-DM--S----------V-H-S--F-KRE--T-VS---E-V-EM-I--S--K-S-E------.WT  628
LST.CD.88.524              VTW----LE-YET-KT----K-Q-A------E-VVRV-QNKK-VI-F-WR-G..NAI-RA-R-Q-QKS--M-PLQ--A------GK-------FV-.-IQV-VTR----N--INF--V----DL--IS--Q-EQE--S--A---E-TD-Y------E-VG-T-----I-Q  633
LST.CD.88.485              VTW-DA-LE-YEQ-QQ----KLQ-A----T-E-VVRV-QNKK-VI-F-WR-G..NSI-RA-R-Q-QKA----PLQK-A------GK-------HV-.-IQV-VTR-V-DQ--SDH--V-----L--IS--Q-EQE--S--A---E-VD-Y------EKVG-T-------Q  632
LST.CD.88.447              VTW-DA-LE-YEQ-QQ----KLQ-A----T-E-VVRV-QSKK-II-F-WR-G..NSI-RA-R-Q-QKS----PLQK-A------GK-------HV-.-IQV-VTR-V-DQ--SDH--V-----L--IS--Q-EQE--S--A---E-VD-Y------EKVG-T-------Q  632
LST.KE.x.lho7              -NW----LE-YGQ-K-V---KMQ-A----E-E--VRV-QNKK-II-F-WR-G..NNI-RA-R-Q-QKA----PLQK-V--I---GK-------FV-.-IQV-VTR-V--H--SDH--V----DL--IS--Q-EQE--IW-A---I-VD-Y------EKVG-T-----I-Q  632
SYK.KE.x.SYK173            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-----------DS-T-N----AS  649
SYK.KE.x.KE51              VE------A-M--------VEQT-S--Q-E-P-E-H-S-L--Q--G-I-K-G-QE-P-I---TGKAYAT-H--YQA-AQLMN--GIQALW----I-.E-H--VKR-E--K---DH-----L--VKC-H--I--RWY-N-VS--V.E-------------NS-E-Q---I-D  654
COL.CM.x.CGU1              VQ-SA---E--QD-KQ-I--E-Q----I-HE-IWVD-FRL-K---G-AVC-K..HGT-RR--HNTGKNQ-F-SMQE-ASVI---GR-A--T---V-.-M-V-AKR-D--Q--SD---SA----I---SSSYV---VWN-VL--LKE-P-YWT--GCS-KLGA----WINS  595
```

p66 RT, p15 Rnase H end p31 Integrase start

```
H1B.FR.83.HXB2          RGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAG.IR..KVLFLDGID.KAQDEHEK.YHSNWRAMASDFNLPPVVAKEIVASCDKCQLK.GEAMHGQVDCSPGIWQLD  779
H1A1.UG.85.U455         -------S--E--------H--H-------S--------------------R----I------K--E------S------------------S-.--..---------.---ED---.--C---------------------N-----.----------------- 778
H1B.US.90.WEAU160       ----R--S-------------H----------------------------------S-----------------------------------S-.--..---------.---E----.------------------------------.----------------- 779
H1C.ET.86.ETH2220       -----I-S--E-----------Q-------S---------------L----K----I---------S--R---S------------------S-.--..---------.---E----.---------NE--I----P-----C-------.---I----N--------- 786
H1D.CD.84.84ZR085       --------PF-------------N---------------------------K------S-----------------------------------S-.--..---------.---E----.------------------------------.----------------- 779
H1F1.BE.93.VI850        K-K----S--E-----A-------------S--------------------K----I---------Q--R---S-------------------.V-..-I-------.---E----.--N------------I---------------.----------------- 780
H1G.SE.93.SE6165        K-K--II---E-----A-----Q------RS--------------------R--A-------------------S------------------S-.--..---------.---E---R.--N------------I---------------.----------------- 776
H1H.CF.90.056           --K----S--E----------------------------------------K----------E--------S------------------S-.V-..---------.---E---R.--N----V--------I---------------.----------------- 778
H1J.SE.93.SE7887        K-----------------H------R-------------------------K----------E--------S------------------S-.--..---------.---ED---.------------------------------.----------------- 777
H1K.CM.96.MP535         -------SI-E-----A-----C-------S--------------------K---D--------------RI--S-------------------.--..---------.---E----.--N------------I---------------.---I------------- 779
H101_AE.TH.90.CM240     -------S--E--------H--H-------S--------------------R----V------E--------S------------------S-.--..---------.---E---R.------T---------I------TN-------.----------------- 784
H102_AG.NG.x.IBNG       -------S--E--------H--H-------S--------------------R----------K----D----S------------------N-.--..---------.---E---R.------K---------I-------------M-.----------G------- 779
H103_AB.RU.97.KAL153_2  -------S------------H------------------------------K-----S-------------------------------------.--..E--------.---E----.--G---------------------------.----------------- 779
H104_cpx.CY.94.CY032    -------S-SE-----------------S-----------I-------R---D---------R-D---S------------------N-.--..---------.---E----.--N-----------S---------N-----.----------------- 778
H1N.CM.95.YBF30         -------SIA------A-----LM---E--RD---------M---HS---K------S----E-----R---S------------------S-.--..-I------E.---ED-DR.------K-----------I----------------.---------N----V---- 786
H1O.BE.87.ANT70         Q-K--IIK-DE-----A--M--L------KET-----------V-SS--T----PI-Q-----E-T---Q---T-----------KI-----KD.--..R----E---.Q--ED---.------K-L--E-G---------I---P--HI-.---I---------EV--I- 775
H1O.CM.91.MVP5180       Q-K-NIIK-EE-----A--M-VLI-----KEQ---------V----SS--T--D-PI-Q-----E-T---R---T-----------KI-----KD.--..R----E---.Q--ED---.--------L----G---I-----I---P--HI-.---T-----Y--E---M- 775
CPZ.CD.90.ANT           ---SR-KH-QK----QA-----LM--E--TGP------------VL-GT------P--EE--QK---R-QI--S------------------Q-.--..Q----E---.---ED-D-.------SL-DEY----I-----I-Q----HV-.---R-----------V- 779
CPZ.CM.01.SIVcpzCAM13   K-K--IIS-EN----QA--G-L-------EQQ---------V-----S---H-------------E----D---S------------------S-.--..---------.---E---R.------K-L-GE-----I-------N-----V-.---L---------T--M- 778
CPZ.CM.05.SIVcpzEK505   ------TSISE----QA----VLM----A-Q----------V----HS---K-------------E-----RI--S-------------I-----T-.--..---------.---E---R.------K----------I--------------.---I---IN----V---- 785
CPZ.CM.05.SIVcpzMB66    K----IIS--E-----A-----Q------EV----------------G---T----I-------E---------S-------------I-----S-.--..---------.---E----.--N-------------I-------N-------.---I------------- 782
CPZ.CM.98.CAM3          -----I-N-EN-----A--T-V------AEHI---------V----HS-----------L---E------S--S-------V-----------S-#--..---------f-HKE---Ef--N----------I-HI----L--R-------.----------------- 784
CPZ.GA.88.GAB2          N-K---IS-EE-----A--E-LRI-----Q---V------V----NS---------I-------E-----------T------------------S-.--..---------.---ED--R.---------DE-Q-----------T-------.---I------------- 782
CPZ.CM.98.CAM5          -----IIN-EN-----A--T-V------AEHI---------V----HS----------GS---E------S--S------------------S-.--..---------.---E----.--N--K----------------------------.----------------- 780
CPZ.CM.05.SIVcpzMT145   -----IHS-E------A----VLM------Q---V------V---L-SH--H---DI-------E------I--S----R-------------S-.--..--------E.---E---R.------K---Q------I-------Q-------.----------------I- 778
CPZ.TZ.00.TAN1          ---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-.---I-----Y--E---I- 775
CPZ.CM.05.SIVcpzLB7     --K--IIP--E-----A-----Q------P-------------L-------D---------N---------T------------------S-.--..---------.R--E----.--N-------------I---------------.----------------- 782
CPZ.US.85.CPZUS         -----II--EN-----A--T-V----K--ENT--V--G---V----HS---------------E------S-IS-------------I-----S-.--..---------.---ED-D-.------T----------------------P-.---I------------- 781
CPZ.GA.88.GAB1          K-K--IIS-EN----QA--K-LL------DQQ---------V-----S---H-------------E------I--S-------------------.--..---------.R--E---R.------K----------I-------H-----V-.----------------V- 777
GSN.CM.99.CN166         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.--PK----NAEL-M--M- 790
GSN.CM.99.CN71          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.--PK----NAEL----M- 783
MUS_1.CM.01.1085        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.--PKT---NAEI-V--M- 780
MUS_1.CM.01.CM1239      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.--PKT---NAEL-V--M- 780
MUS_2.CM.01.CM1246      --QERAIN-EN----QA-----L---M-GPP-M-L------------T-A-EV---P--E---Q-MLS-NAIFISR-----------E--H---R-.--..Q---MEN--.P-VED---.------KYLRDKY-I-TIL-----NK-SA--TH.--PK---NADL-V--M- 785
MUS_2.CM.01.CM2500      --QE-AID-EN-----A-----LM--T-GPPKM--------------T-A-ET---P--EK--Q--LT--AI--S-----------QE--H---K-.--..Q---MEN-E.P-IED-D-.------KYLRDKY-I-TLL-----NR-SA--TH.--PK----NADI-V--M- 784
```

p15 RNase H end p31 Integrase start

```
MAC.US.x.239            --KD--KV-EQ----QA--E-FLM--T---PKA--IV----VM---TGC-TE---R-------EM---SEI-V------------QEI-H---Q-.--..Q----EK-E.P--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.---I---ANSDL-T--M- 790
H2A.GW.x.ALI            ---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-.---I----NAEL-T--M- 803
H2A.DE.x.BEN            --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-.---I----NAEI-V--M- 803
H2A.SN.x.ST             ---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-.---I----NAEL-T--M- 803
H2B.GH.86.D205          --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-.---I----NADL-T--M- 802
H2B.CI.x.EHO            --KD--KP-EQ----QA--E-FA-------PQ---IV----VM--VA---TET--PI-RE---EM-----I-VG-------L---QE--H---Q-.--..QI---EK-E.P--E----.--N-VKELVHK-GI-QL--RQ--N------Q-.---I----NSEL-T--M- 803
H2G.CI.92.ABT96         --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-.---I----NAEL-T--M- 786
H2U.FR.96.12034         --KDR-KP-EQ----QA--E-FAM--E---PK---------IM---AG--TET--P--SK---E--R--A--VG-----------QE--H---Q-.--..Q----EK-E.P--E----.----VKELQHK-GI-QL---Q--NT-HV--Q-.---------NTEV-T--M- 799
MAC.US.x.251_1A11       --KD--KV-EQ----QA--E-FLM--T---PKA--IV----VM---TGC-TE---R------KEM---SEI-V------------QE--H---Q-.--..Q----EK-E.P--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.---I----NSDL-T--M- 790
MAC.US.x.251_BK28       --KD--KV-EQ----QA--E-FLM--T---PKT--IV----VM---TGC-TE---R-------EM---SEI-V------------QEI-H---Q-.--..Q----EK-E.P--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.---I----NSDL-T--M- 786
MAC.US.x.EMBL_3         --KDI-KV-.T----QA--E---HGIE---PKR--IVEL-VCY-NNNRF-TE---R-------EM--.VR--V-----LE-----QEIGP---Q-.F-..Q----EK-E.P--E--D-.----VKELVFK-G--RI--RQ--DT----HQ-.---I----NSDL-T--M- 784
SMM.US.x.H9             ---G-TKL-EQ----QA--E-F----A---PKA--IV----VM--VAG--TE---R-------KM----AI-V-------X--E-QE--H---QE.--..Q----EK-E.P--E----.----VKKLVFK-X--RL---Q--DT----H--.--XI----NXXL-T--M- 786
SMM.US.x.PGM53          ---D-AKP-EQ----QA--E-F----A---PKA--IV----VM---AG--TE---R-------EM----AI-V------------QE--H---Q-.--..Q----EK-E.P--E----.----VKELVFK-GI-RL---Q--DT-H--HQ-.---I----NAEL-T--M- 786
SMM.US.x.SIVsmH635F_L3  ---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-.---I----NAEL-T--M- 786
SMM.SL.92.SL92B         --KE--LP-EQA---QA----LL---K--PSK--V------V-N--TG--SE-D-DI-A------VQ--A--IG-----------NE--R---Q-.--..Q----ES-E.P--ED-D-.----VKEL-QKY-I-QL---Q--NA-N---Q-.---I---TNAEV-T--M- 783
STM.US.x.STM            --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-.---I----NAEL-T--M- 786
MNE.US.x.MNE027         --KD--RV-EQ----QA--E-FLM--T---PKA--IV----VM---TGC-TE---R-------EM---TEI-V------------QEI-H---Q-.--..Q----EK-E.P--E--D-.----VKELVFK-G--RL---Q--DT----HQ-.---I----NSDL-T--M- 786
GRV.ET.x.GRI_677        --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.--PK-----A-IET--M- 801
VER.KE.x.TYO1_patent    Y-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-.--PV-----A---T--M- 798
VER.KE.x.9063           Y-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-.--PI-----A---V--M- 795
VER.KE.x.AGM155         Q-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-.--PI-----A---V--M- 798
VER.DE.x.AGM3           Y-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-.--PV-----A---V--M- 795
TAN.UG.x.TAN1           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.--PI-----A-V-V--M- 798
SAB.SN.x.SAB1C          --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-.--SV-----A---V--M- 811
OLC.CI.97.97CI12        G-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-.-VPR--I-AVGVYR--I- 780
DRL.x.x.FAO             --KE--IAIEE-----A-----L---K---PK----------M---SSS-EI-DNPI--------M-S--A---N-------------EE-----R-.--..Q-----NME.---E--DL.--N---SL-QE-G--GI-------Q-P---IH.--PI-----A---T--M- 793
RCM.NG.x.NG411          --KE--KE-EN-----A--G-VL---K---KR-------RHVF--LAGS--T---P--Q-----M-G--E---S------------------Q-.--..Q------G.---E----.-----------EE-QI-QI-------Q-P---V-.---V-----A---T--M- 797
RCM.GA.x.GAB1           --KE--KE-EE-----A--E-VL---K---PK---------VY--LE----T-D-G--TE--N-M-G--A---S-----------E------K-.--..Q------E.---E----.--N----L-E--QI-QI-------Q-P---V-.---I-----A---T--M- 792
MND_1.GA.x.MNDGB1       --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.--PIK--T-A-L-V--I- 788
MND_2.x.x.5440          --KS---A-EE-----A--T--KM------PR-------------LS-A----DNPI-RE---LM-G--EI--S-------------I-----K-.--..Q----E---.---E--D-.--N----LSQ------I-------Q-P---T-.--PV----NAD--T--M- 792
MND_2.CM.98.CM16        -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-.--PI-----T---T--M- 791
MND_2.GA.x.M14          --KS--IA-EN-----A--E--L---K---PRA------------LS-T--R-DNPI-RE--N-M-A--A-----------------I-----K-.--..Q----E---.---E--D-.--N---SLSQE-SI--I-------Q-P---V-.--PI-----AA--T--M- 792
DEB.CM.99.CM40          W-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.--PT-----A-M----M- 784
DEB.CM.99.CM5           W-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-.--P------A-M-V--I- 784
DEN.CD.x.CD1            N-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.--PIR----A-L-T--M- 806
TAL.CM.01.8023          S-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.--PKT--TNAAV-T--M- 788
TAL.CM.00.266           T-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.--PKT--NNAAV-T--M- 787
MON.CM.99.L1            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.--PKS---NAEV-V--M- 786
MON.NG.x.NG1            T-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.--PK----XSDI----M- 789
SUN.GA.98.L14           G--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-.---IT----A-V----I- 798
WRC.CI.97.97CI14        KSKE---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-AP---M--YEV-T--M- 803
WRC.CI.98.98CI04        NSKE--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-APL--M--YEV-T--M- 802
WRC.GM.05.Pbt_05GM_X02  KT-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-HPI--M--YEI-T--M- 794
LST.CD.88.524           S-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-.---IT----V-V------ 799
LST.CD.88.485           S-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-.---IT----T------I- 798
LST.CD.88.447           S-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-.---IT----T------I- 798
LST.KE.x.lho7           S-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-.---IT----V-V------ 798
SYK.KE.x.SYK173         D-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.--PK-----VDIYN--M- 813
SYK.KE.x.KE51           --K-TLKR-EN-----A--E-VLM------P--------------LMNC-TS--HPV-E--MQKAME---I-IT-----------QE--Q---K-.--..-I----N-P.---E----.--T-VEYIRQE-H--RQ---A--EM-P---IR.--PK-----TDL-T--M- 818
COL.CM.x.CGU1           K-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-V---L-Y-Y-L---- 757
```

```
H1B.FR.83.HXB2               CTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTI.HTDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGER.IVDIIATDIQTKELQKQI.TKIQNFRVYYRDSR 946
H1A1.UG.85.U455              ------------------------------------I----------V-.---------S-A-K-V----N-Q------------------------------E------------------------------------.-I----------------.S------------- 945
H1B.US.90.WEAU160            -------I----------------------------I------------.---------ST--K-----------------------I------------------------------------------------.-I---------QQ-----.-------------- 946
H1C.ET.86.ETH2220            -------I--------------------------------------RV-.---------SNA-K--------Q------------------------------E-------------------R-----------.-I----S--------N--.L------------- 953
H1D.CD.84.84ZR085            --------------------------------A--------------VV.---------S---K------------------------------------------------------------------------.-I----S----R------.-------------- 946
H1F1.BE.93.VI850             -----------------E------------------I----------I-.---------S-A-K-S------Q-----------------I---------------------------------------------.-I---S-----R------.-------------- 947
H1G.SE.93.SE6165             -------I-I--------------------------I---------TV-.---------S-A-K------N-T-----------------------------------------------------------------.-I----S-----------.-------------- 943
H1H.CF.90.056                ------Q-----------------------K----------S-----V-.---------S-A-K------D-Q-----------------------------------------------------------------.-I----------------.SN--K--------- 945
H1J.SE.93.SE7887             --------------------------------A-F-I----------V-.---------SGA-K------D--------------------------------E------------------------------------.-I---------R------.-------------- 944
H1K.CM.96.MP535              -------I----------------------------I----------V-.-----T---STV-K-------V------------------------------------------------------------------.------------------.LN--K--------- 946
H101_AE.TH.90.CM240          -----------------------------------------------V-.---------S-A-K------NVQ------------------------------E------------------------------------.-I----------------.-------------- 951
H102_AG.NG.x.IBNG            -------I----------------------------I----------V-.---------S-A-K------NVT-----------------------------------------------------------------.-I----S-----------.I------------- 946
H103_AB.RU.97.KAL153_2       -------I----------------------------V----------I-.---------STA-K-----------------------------Q--QT----------------------------------------.-I----------------.I------------- 946
H104_cpx.CY.94.CY032         ---------M--------------------------I----------M-.-A---P---S-A-K------D-N-----------------------------------------------------------------.-I----S-----------.-------------- 945
H1N.CM.95.YBF30              -------I-----------L----------------I----------V-.---------S---K------N---------------A--------------I-----------------------------T----.-I---------TN--T--.L-V----------- 953
H1O.BE.87.ANT70              ---M---I-I--------F----------------------A-----V-.-----P---ST-MK------N-QH---------------A------S--Q---------R-------V-------------T----.-I--L-SQ---T------.L-XHK--------- 942
H1O.CM.91.MVP5180            -------I-I-------DF----------------------A-----V-.-----P---S-AMK-----T--QH---------------A------S--Q-----------------V-------------T----.LI--L-SQ---T------.L--N---------- 942
CPZ.CD.90.ANT                ---------I-----S--F-----MAD---KS---------S-------.-----A---S-A-K------N-Q--------------------Q--Q----I-----Q-----V---H-------------TP-Q-.-L--L------TQ--N--.L---Q---H----- 946
CPZ.CM.01.SIVcpzCAM13        -------I-I----I---------------------I----S-----VL.-----T---SNA-K------N---K------------------------E-----------------S-------------T----.-I----------------.L---K--------- 945
CPZ.CM.05.SIVcpzEK505        -------I-----------L----------------I----------V-.---------S---K------N-Q-------------A--------------I-----------------------------T----.-I---------TK--T--.L-V----------- 952
CPZ.CM.05.SIVcpzMB66         -------I-----------M----------------I----------V-.---------SS--K--------Q----------------------------I----------------------------------.-I--L------TK--Q--.SN--K--------- 949
CPZ.CM.98.CAM3               -------------------L----------------I----------V-.-----P---SN--K--------Q----------------------------I-E-----R-----------------------T----.-I---------TN-----.L-V----------- 951
CPZ.GA.88.GAB2               -------I-----------L----------------I----------K-.-----T---S-A-K--------------------------------------------------------------------T----.---ML---L--Q------.L---K--------- 949
CPZ.CM.98.CAM5               -------I-----------L----------------I----------V-.-----P---SN--K--------Q----------------------------I-----Q-----------------------T----.-I---------TN-----.L-V----------- 947
CPZ.CM.05.SIVcpzMT145        -------I-I--------------------------V------------.-----P---SNA-K--------R--------------------------E------------------------------------.-I----------------.L---Q--------K 945
CPZ.TZ.00.TAN1               ---------I--------F-------E---R-----I----------K-.-----P---STA-K------Q-QH-------------------Q--Q--E-------Q-R---I---Y-----------E-T----.LL--LT-N----Q-----.L-V---------A- 942
CPZ.CM.05.SIVcpzLB7          -------------------M----------------I----S-----I-.---------STA-K------Q-Q-----------------------------------------------------------------.-I---------T------.S-V-K-------R- 949
CPZ.US.85.CPZUS              -------I-I---------L----------------I----------V-.---------SS-----------Q----------------------------I-----------------------------T----.-I----SEL--DL-----.L-V----------- 948
CPZ.GA.88.GAB1               ---------I---------------------------------------.-----P---S-A-K------D-------------------L-------------------------------------------T----.-I---------S------.L-V-K--------- 944
GSN.CM.99.CN166              -------------------VW-RI--Q---RQ--LK--E--AT---THL.-----P--VSKELE------D-QHTT-V-----------N---Q--ET-TKI-EEVTY-E---AQ-CY-----------DMCPT--.--NM-H-ELE-QH-NT-N.S-F-K------QGA 957
GSN.CM.99.CN71               -------------------VW-RI--Q---RQ--LK--E--AT---THL.-----P--ISKELE------N-QHST-V-----------N---Q--ET-QKI--EVTY-E---AQ-CY-----------DMCPT--.L-NM-H-ELE-QH-NT-S.S-F-K------QGA 950
MUS_1.CM.01.1085             -------I-----------MW-KIL-R---RQ--LG--E--SL--ATQ-.-----A--IS-EFE------S-EHTT-V-----------NK--Q--EA-HKI-EEVTY-E---AQ----L---K---V-DL-NA--.-INMLH-ELELQH--Q-K.Y-F--------TGS 947
MUS_1.CM.01.CM1239           -------I-----------TW-KIL-R---RQ--LG--E--AL---TQ-.-----A--ISEEFG------S-DHTT-V-----------NK--Q--ET-QKI-EEVTY-E---AQ----M---K-----DLA-A--.-INMLH-ELELQH--S-K.S-F--------TGS 947
MUS_2.CM.01.CM1246           -----------------SF-W-RI-SQ---RL--LE-SN--AT--ISQ-.-P---T---SKEFQ-VA---N-QHTT-V-----------NA--Q--ET-HK--EEVTY-E---AQ----L-N-------DMTPT--.L-NMLY-ELEIQQ--NHN.--FSK------RGA 952
MUS_2.CM.01.CM2500           -------I----------F-W-RI-SQ-N-RL--LE--N--AT--ITQ-.-----T---SKEFQ-VA-----EHST-V-----------NA--Q--ET-NK--EEVQY-E---AQ----L-N-------DMTPT--.L-NMLY-E-E-QQI-NTK...FSK------RGA 949

MAC.US.x.239                 -------I-I--------F-------Q---RQ--L----------ITHL.-----A--ASQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQFQ-SKN.S-FK-------EG- 957
H2A.GW.x.ALI                 ---------II-------F-------Q-S-RQ--L------S---ITHL.---S-V---SQE-KMVA--V--E-S--V-----------A--HH--NQ-SRI-E--NTVE-I-L---HCM----R----DMTPA--.LINM-S-EQEIQF--TKN.L-FK--P----EG- 970
H2A.DE.x.BEN                 Y------I-I--------F-------Q-S-RQ--L------S---ITHL.-----P---SQE-KMVA--V--E-S--V-----------A--HH--NQ-SRI-E--NTIE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQF--RKN.SNFK--Q----EG- 970
H2A.SN.x.ST                  -------I-I--------F-------Q-S-RQ--L------S---ITHL.-----A---SQE-KMVA--I--E-S--V-----------A--HH--NQ-SRI-E--NTVE-I-L---HCM----R----DMTPA--.LINMVTAEQEIQF--AKN.S-L------F-EG- 970
H2B.GH.86.D205               -------I-I--------F-------Q---RQ--L------S---ITHL.-----A---SPS-KMVA--V--E-T--V-----------A--HH--NQ-DRL----VSIE-V-L--THCM----R----DMTPA--.L-NM-T-EQEIQFF-AKN.L-F---Q----EG- 969
H2B.CI.x.EHO                 ---------I--------F-------Q---RQ--L------S---ITHL.-----A---SQD-KM-A--I--E-T--V----E------A--HH--NQ-DRI----VSIE-V-L--THCM----R----DMTPA--.--NM-T-EQEIQF--TKN.L-F--------EG- 970
H2G.CI.92.ABT96              ---------I--------FV------Q---RQ--L----------ITHL.-----A---SQE-KMVA--V--E-A--V--------------HH---Q-DKI----NSIE-I-L---HCM----R----DMTPA--.-INM-T-EQEIQFQ-TKN.S-FK-------EG- 953
H2U.FR.96.12034              --------VI--------F-------Q---RQ--L----------ITHL.-----A---SQE-KMVA--V--EHT--V-----------A--HH--NQ-ERI-E--NTME-I-L--AHCM----R----DMTPA--.-INM-S-EQEIQF--SKN.S-FK-------EG- 966
MAC.US.x.251_1A11            -------I-I--------F-------Q---RQ--L----------ITHL.-----A--ASQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSVE-I-L---HSM----R--T-DMTPA--.LINM-T-EQEIQFQ-SKN.S-FK-------EG- 957
MAC.US.x.251_BK28            -------IVI--------F-------Q---RQ--L----------ITHL.-----A--ASQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQFQ-SKN.S-FK-------EG- 953
MAC.US.x.EMBL_3              -------IVI--------F-------Q---RQHY.----------YLH-YTHS--A--ASQE-KMVT-----EAHLWV-----------A--HH--NQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.LINM-T-EQEIQFQ-SKN.S-FK-------EG- 951
SMM.US.x.H9                  -------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-SKN.S-FK-------EG- 953
SMM.US.x.PGM53               -------I-I--------F-------Q---RQ--L------S---ITHL.-----A---SQE-KMVA-----E-T--V-----------A--HH--TQ-DRI-E--NSIE-I-L---HCM----R----DMTPA--.L-NM-T-EQEIQFQ-SKN.S-FK-------EG- 953
SMM.US.x.SIVsmH635F_L3       -------I-I--------F-------Q---RQ--L----------ITHL.-----A---SQE-KMVA-----E-T--V-----------A--HH--TQ-DRI-E--NSVE-I-L---HCM----R----DMTPA--.L-NM-T-EQEIQFQ-SKN.S-FK-------EG- 953
SMM.SL.92.SL92B              ---------I--------F-------R---RQ--L----I-S---I-HL.-----A---SQE-KMVA--L-VE-S--V-----------A-DLH---N-DKI-E---SVE-L-L--AHCM----R----DMTPA--.--NM-T-ELE-QY-NS-N.S-F--------EG- 950
STM.US.x.STM                 -------I-I--------F-------Q---RQ--L------S----THL.-----A---SQE-KMVA-----E-T--V-----------A--HH--TQ-DRI----NTVE-V-L---HCM---KR--L-DMTPA--.L-NM-T-EQEIQFQ-SKN.S-FK-------EG- 953
MNE.US.x.MNE027              -------I-I--------F-------Q---RQ--L------S---ITHL.-----A---SQE-KMVA-----EHT--V-----------A--HH--NQ-DRI-E--NSME-I-L---HCM----R----DMTPA--.LLNM-T-EQEIQFQ-SKN.S-FK-------EG- 953
GRV.ET.x.GRI_677             ---------I--------F-------R---K---H-----LA-----HL.-----P---SQN-A-V---GN-EHTT----------S-----RQ--E--S-I--DC-R-E------TH------------I-SA--.L-NMLT-QLELNT--N--.Q--L--K----EG- 968
VER.KE.x.TYO1_patent         --------VI--------F-------R---K---K----ILS---ITQL.-----P---SQE-A-I---GK-EHTT----------SI-----Q--E---KI--DCQYTE---L--CH------------QTSA--.LIN--T-QLEIQH--TK-.Q--L-------EG- 965
VER.KE.x.9063                ---V---IVI--------F-------R---R---K----ILS---IVQL.-----P---SQE-A-I---GK-EHTT-V--------S------Q--E---KI--DCQYTE---L--CH------------LTPA--.LINM-T-QLELQT--TK-.Q--L-------EG- 962
VER.KE.x.AGM155              ---I---IVI--------F-------R---K---K----II----ITHL.-----P---SQE-A-M---GKVEHTT-V--------SI-----Q--E---KI--DCQYTE---L--CH------------LT-A--.LINM-T-QLEINT--TK-.Q--L-------EG- 965
VER.DE.x.AGM3                ---------I--------F-------R---K---K----ILS---ITQL.-----P---SQE-A-M---GK-EHTT-V--------SI-----Q--E---KI--DCQYTE---L--CH------------LTPA--.LINM-T-QLELQH--TK-.Q--L-------EG- 962
TAN.UG.x.TAN1                -------I-I--------FL-----AR---K---K----IIS---ITKL.-----P--VSQE-QTI---GQVEHTT----------S-----RQ--ET-EKI-EDCAF-E---L--CH------------MTPA--.LINM-T-QLEIQHI-T-Q.Q--S--K----EG- 965
SAB.SN.x.SAB1C               -------I-I--------F-------Q---KA--H------S---ITQL.-----T---SQQ-A-I---GK-EHT--V---------------Q--E----I--D--R-E---I---H---------------A--.LIN--H-ELE--T--QK-.S----------EG- 978
OLC.CI.97.97CI12             ---------CL-A-TG--FLYG-ILR--D-EN-VK-V-R--SQ----I-.-----T--VNQKMQ-V---L--EH-T---W------II-AK-RV--E-L-RI-E--T--E--LA--I------N---Q--GTVA-E.L-NQ-Y--T.QI-YTQNK.F-KFSGYRV-YKNK 946
DRL.x.x.FAO                  -------I-I--------H-----------EK--------------SHL.-----P---SEK-ATV---K--EHTT-V--------------NQ-------I-----K-E-------L-----------E---A--.------S-LL-SK--QN-.L----------EG- 960
RCM.NG.x.NG411               -------I-I--------------------K---H------A-----HL.-----T---S-A-Q-D----Q-EHT--V--------------HQ--I----I-----K-E-------L------------------.-I----S-L--TK--N--.S--------F-EG- 964
RCM.GA.x.GAB1                -------I-I--------------------K---H------A----RKL.-----A---S-A-Q-V----Q-EHA--V---------------Q--I--E---E---K-E-------LV-----------------.-I------LA-NK--N--.S----------EG- 959
MND_1.GA.x.MNDGB1            -----NQI-I--------FMK----T----KK--E------AQ--ISKL.-----P---SQE-ETM---L--EHT--------------NK--Y--EL-EKI-EDCKE-----A--T------QR--L--MT----.--NM-N-ELEYQYQ-N--.S-NL--K--F-EG- 955
MND_2.x.x.5440               -------I-I---------L----------K---H------------HL.-----P--VSEK-ATV----Q-EHTT-V-----------AK-HH-----E-------R-E-------L-----------E--P---.-----T---L-TK-RQN-.S-----Q----EG- 959
MND_2.CM.98.CM16             ---M---IA--------R-L------T---K---H------------HL.-----P---SEK-ATV----Q-EHTT-----------I-AK-HH--Q----------K-E-------L-----------E--P---IVDI-ATDLLT--LQHN.-.Q----------EG- 958
MND_2.GA.x.M14               ---------I---------L----------K---H------------HL.-----P--VSEK-ATV----Q-EHTT-------------AK-HH--V--E-------R-E-------L-----------E--P---.-----T---L-TK--QN-.S----------EG- 959
DEB.CM.99.CM40               -----N-I-I-----------TKIL-----K---L---Q-GA----TQ-.-----P--ISQP-K-----L--EHTT-V------------K-RI--DT-K-I--D--R-E---A--TY-I----R----DM-PI--.L-NM-T-E-E--T--QK-.S-LLG-K----EGA 951
DEB.CM.99.CM5                -----N---I--------FM-TK-LS----K---L---Q-GA---I-Q-.-----P--ISQP-K-----L--DHTT-V------------K-RV--ET-K-I-ED--R-E---A--T--I----R----DMA-I--.L-NM-T--LE-QT--QK-.S-LLGYK----EGA 951
DEN.CD.x.CD1                 -----N---I---NA----L-TKIL-----K---L---QI-A---ITKL.-----P--ISQAMQ------N--HTT-V--------I--NK--Q--EV-AL--ED--R-E--LA--TL-L----R--L-DMT-A--Y-NMLNTHLEIQH..T--..P-FS--K----EGA 971
TAL.CM.01.8023               ------Q--C-----------TKIL-R---R---L---QV-S---IEHL.-----P--VS-EMQ-TA--LK-EHTT-V--------S--NK--Q---T-Q-I--EVQY-S---AQ-T--L----R--L-DMCPA-A.LINM-Y-EL--TT--N--.HNFSD-K----KGA 955
TAL.CM.00.266                ------QI-C-----------TKIL-R---R---L---QV-S---ISHL.-----P--VS-EMQ-MV--LK-EHSA-V--------A--NK--Q---T-T-I--EVQY-S---AQ-T--L-Y--R--L-DMCPA-A.-INM-Y-EL--TQ--N--.QNFSD-K----KGA 954
MON.CM.99.L1                 ------SI-----------VW-KIL-R---KC-GIA--E--AM---TQ-.-----P---SQEFE--A---N-HHTT-V-----------N--RQ--ET-KKI--EV-Y-P--LAQ-L--L---------DMAPVD-.FINM-H-ELELQTSNN--.--FSK----F-TGA 953
MON.NG.x.NG1                 ---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--.--FSK----F-TGA 956
SUN.GA.98.L14                ---M-E---I---------MV---L-N-Q-KT--TW----CAM----Q-.-----P--ISKD-E-V---L--QHTT-------------AK--V--Q--SRI-ED-QE-----L--LH-----QR--L--MT-A--.FINM-NA-LE-QY---IN.S--LK-K----EG- 965
WRC.CI.97.97CI14             -------I--N--------L--KI-S--DAYQ--L-T-Q--S---I-K-.-----T--KSQKMQ-I---L--E--------------I--K-RR--D--HSI-ED--K-E--LA--L------E---L--.-PA--.L-NM-TS-LE-QQT-Q-K.L-FK--Q----TGA 969
WRC.CI.98.98CI04             -------I--N--------L--KI----DAYQ--L---QI-----M-K-.-----T--KSQKMQTI---L--E--------------I--K-RR--E--HRI-ED--K-E--LA--L------E---L--.-PA--.L-NM-TAELE-NQT-Q-K.L-FK--Q----TGA 968
WRC.GM.05.Pbt_05GM_X02       -------I--N--------L--KI-T--DAYQ--L-TIQ--------K-.-----T--KSQKME-I---L--E--------------I--K-RV--Q--QRI-ED--K-E--LA--IY-----E---L--.-PA--.--NM-QS--E-NQT-Q-K.F-FK--Q----TGA 960
LST.CD.88.524                ------QI-MN---M---FMV-----D---KT--N-----CS-----Q-.-----P--VSKD-Q-VT--LS-EHTT-------------AK--V--Q--ER--ED-QQ-----L---H-----QR--L--L-PA--.YINM-NA-LE-QYI--LN.S--LK-K----QG- 966
LST.CD.88.485                ---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---LN.S--LK-K----QG- 965
LST.CD.88.447                ---M--Q--IN-------FMV-----D---KT--H-----CS----QQV.-----P--VSKE-Q-VI--L--EHTT-------------AK--A--Q--ER--ED-QQ-----L--TH-----QR--L--L-PA--.FINM-NA-LE-QY---LN.S--LK-K----QG- 965
LST.KE.x.lho7                ------Q--IN-------FMV-----D---KT-SH-----CS-----Q-.-----P--VSKE-Q-VT--I--EHTT-------------AK--V--S--ER--ED-QQ-----L---H-----QR--L--LTPA--.FINM-NAELE-QY---LN.S--LK-K----QG- 965
SYK.KE.x.SYK173              ---E-----C---NT------TKILKR---D---L--MQI-S---I-Q-.-----P--VSDKFK-----C--EHTT----------I---K-RY--EA-S-I--DVT--Q---A--T--L----------I-P---.YINMLY-EL-LQQNTTS..P-FS-------QGK 979
SYK.KE.x.KE51                ---------C---NT----T-TKILKR------GL---QI-A---I-H-.-----P--ISDAFA-------VEHTT----------I--NK-RQ--ET-N-I-EEV-R-E---A--T--L----------TTPAD-.YINMLY-EL-LQNTHT...Q-FS--K----HGT 983
COL.CM.x.CGU1                ---E-----------CTLFCW-TILKR---E--GRA-I---SQ-E-RQV.-----P--VSQHFK--V--L--AHTT-H-----------QR--DV-RK-KKMK----T-ESK-A---YAL-------L--K-PW--.Q-ERAIIELD-QN-T-LQNQ-FK--KA-WKEH. 924
```

```
                                                                p31 Integrase end
                                                                      Pol end
                                                                                 ↘
H1B.FR.83.HXB2          NPLWKGPAKLLWKGEGAVVIQDNS.DIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED*  1003
H1A1.UG.85.U455         D-I----------------------.------------------------M-G------  1003
H1B.US.90.WEAU160       D-----------------------.-------------------------------  1004
H1C.ET.86.ETH2220       D-I----------------------.----------------------A----G------  1011
H1D.CD.84.84ZR085       D-I----------------------.-------------------------------  1004
H1F1.BE.93.VI850        --V----------------------.E--I----------------V-----G------  1005
H1G.SE.93.SE6165        D-V--------------------N.E-------------------------G------  1001
H1H.CF.90.056           D-I----------------------.E-------E------------------------  1003
H1J.SE.93.SE7887        D-I-------P-------------.E-------------------------G------  1002
H1K.CM.96.MP535         E-I----------------------.E-------------------------G------  1004
H101_AE.TH.90.CM240     D-I----------------------.-------------------------G------  1009
H102_AG.NG.x.IBNG       D-I----------------------.-------------------------G------  1004
H103_AB.RU.97.KAL153_2  D-I--------------------N.--------------------------------  1004
H104_cpx.CY.94.CY032    E-I----------------------.----------------------N---G------  1003
H1N.CM.95.YBF30         D-I-------------------G.----------------------G----G--NQ  1011
H1O.BE.87.ANT70         D-I-----Q-------------KG.--------------E-------T-SM--G-T-SE  1000
H1O.CM.91.MVP5180       D-I-----Q-------------KG.----------------------T-SM-N--T-SE  1000
CPZ.CD.90.ANT           D-V-----Q-----------K-QE.E------------KE-R-KIEDR-DL-G--N---  1004
CPZ.CM.01.SIVcpzCAM13   D-I-----T-------------QG.E-------------------------------  1003
CPZ.CM.05.SIVcpzEK505   E-T-----------------G.----------------------G----G---SQ  1010
CPZ.CM.05.SIVcpzMB66    D-I---------------L--QE.E-----------------------------NQ  1007
CPZ.CM.98.CAM3          D-I-----R----------KERE.EV--I-------------------SM-GG---SQ  1009
CPZ.GA.88.GAB2          D-I-----S--------------G.-------------KH----V-----L-GG---NQ  1007
CPZ.CM.98.CAM5          D-I-----R----------KE-E.EV--------------------GSM-G--N--Q  1005
CPZ.CM.05.SIVcpzMT145   D-I-------------------QD.E--------------------A-S-------S-  1003
CPZ.TZ.00.TAN1          D-I-----R----------KEGE.------------KE-------AGGMDD--N-T-  1000
CPZ.CM.05.SIVcpzLB7     D-I---------------L--QE.E------------------E-AN-L-D----N-  1007
CPZ.US.85.CPZUS         D-I----------------KE-E.EV----------K--------A-SM-G----S-  1006
CPZ.GA.88.GAB1          D-I-----T-------------QG.EL----------------------------N---  1002
GSN.CM.99.CN166         ----Q---V----------V-TQAGE-IT---------KP--EAKTE-VGG-THTNN--  1016
GSN.CM.99.CN71          --H-Q---V----------V-TQAGE-IT---------KP--EAK-E-VGS-AHTSN-R  1009
MUS_1.CM.01.1085        D-S-----V---R-------KTD-GEVIT---------KP--AKEDVGSKSDTGDLRKE  1006
MUS_1.CM.01.CM1239      D-S-----A----------VKTEQGQVIT---------KP--AKENVGSKSNTGDHRKE  1006
MUS_2.CM.01.CM1246      --Y-LR--R--------L--KTKEGE-VT---------K---TRQNV-SEPD-VHVRKE  1011
MUS_2.CM.01.CM2500      --S-Q---Q--------I--KTDQGE-IT---------K---TKENV-SDPNPVY-RKE  1008

MAC.US.x.239            DQ-----GE---------ILKVGT.-------------K---GGKEV-SSSHMEDTGEA  1015
H2A.GW.x.ALI            DQ-----GE-----D---IVKVGT.-----------------GRQEL-SGPHLEGARED  1028
H2A.DE.x.BEN            DQ-----GE---------IVKVGT.-----------------GRQEL-SSPHLEGARED  1028
H2A.SN.x.ST             DQ-----GE-----D---IVKVGA.---II--------K---GRQEM-SGSNLEGARED  1028
H2B.GH.86.D205          DQ-----GE---------I-KVGT.E------------H--GGKGL-CSADMEDTRQA  1027
H2B.CI.x.EHO            DQ-----GD---------I-KVGT.E---I---------N--GGKEL-CSADVEDTMQA  1028
H2G.CI.92.ABT96         DQ-----GE---------I-KVGT.E------------K---GGKEV-SSTNMEDTRQT  1011
H2U.FR.96.12034         DQ-----D----------L-KVGT.----I------------GGKEL-SSSNLEGA-KV  1024
MAC.US.x.251_1A11       DQ-----GE---------ILKVGT.-------------K---GGKEM-SSSHMEDTGEA  1015
MAC.US.x.251_BK28       DQ-----GE---------ILKVGT.-------------K---GGKEV-SSSHMEDTGEA  1011
MAC.US.x.EMBL_3         DQ-----GE---------ILKVGT.-------------K---GGKEV-SSSHMEDTGEA  1009
SMM.US.x.H9             DQ-----GE---------ILKVGT.E------------K---GGKEL-SGSHLEDTGEA  1011
SMM.US.x.PGM53          DQ-----GE---------ILKVGT.E------------K---GGKEL-SGPHLEDTGEA  1011
SMM.US.x.SIVsmH635F_L3  DQ--R--GE---------ILKVGT.E------------K---GGKEL-SGSHLEDTGEA  1011
SMM.SL.92.SL92B         DQ------E----------KVGT.E------------K---GRQEMGSSASMEDQ*KN  1007
STM.US.x.STM            DQ-----GE---------IVKVGT.-------------K---GGKEV-SGSHLEDTREA  1011
MNE.US.x.MNE027         DQ-----GE---------ILKVGT.-------------K---GGKEV-SSSHMEDTGEA  1011
GRV.ET.x.GRI_677        D-V-----R-I--------KEGE.-------------K---ERKTM-SEGSMEGVREA  1026
VER.KE.x.TYO1_patent    D-V-----Q-I--------LK.DGS-L-----------K--EPKQRVGNEGDVEGTRGS  1023
VER.KE.x.9063           D-V-----Q-I--------K.EGEEL-I---------K--EPRKRMG-EGNMEGLRGP  1020
VER.KE.x.AGM155         D-V-----R-I--------LK.EGEEL-----------K--EPRKTLG-ETHLEGAGGS  1023
VER.DE.x.AGM3           D-V----GQ-I--------K.GGVEL-EY--------K--EPRKRMG-ESNLEGAGGA  1020
TAN.UG.x.TAN1           E-V-----T----------KEGE.EL-----------KE--ERKTVG-KTNMEG*...  1019
SAB.SN.x.SAB1C          D-V-------I----------EQG.EL-TI--------K------L-SQAPLEGNGRT  1034
OLC.CI.97.97CI12        EGQ-C---E---------LLKEGD.KYFS--L-----VKP--VDSSANV.....*....  994
DRL.x.x.FAO             DQ------E-V--------KEGT.-L--I--------K----NVDSHTNME-*.....  1012
RCM.NG.x.NG411          DQQ------I-E--------QE.-L-------C---K---RKEVDRETNMEGRQ-ES  1022
RCM.GA.x.GAB1           DQ--R-----I----------EETG-L-----------KE--RKDVDSEANLAGRQ-EN  1018
MND_1.GA.x.MNDGB1       DQ-----GI---------LKYQE.E--I---C---K---E.SGKNSQ-NLESV*..  1009
MND_2.x.x.5440          DQQ-----E-I--------KEGV.-L----------K----TVDSNPHMGR*.....  1011
MND_2.CM.98.CM16        DQQ-----E-I--------KEGT.-L----------------TVDS-PN-EA**....  1010
MND_2.GA.x.M14          EQ------E-I--------KEGT.-L----------------TVDSNPHMED--ETA-  1017
DEB.CM.99.CM40          D-T-----T----------CKTEVG----------V-KE---D-DSKVNTQNSNE*..  1007
DEB.CM.99.CM5           D-T-----T----------CKTEVG----------V-K----GVDSKTDTQNPSE*..  1007
DEN.CD.x.CD1            --Q-Q---E----------KTQAG--FP---------KP-NAE.SRKGEQNKGM-SET  1029
TAL.CM.01.8023          ----Q---H-V--------LRTDEGEVIT---------KP--QA-GNKTDLEGSKEQ-A  1014
TAL.CM.00.266           ----Q---H-I--------LRTDEGEVIT---------KP--QALGNKIDLE-SKEQ-A  1013
MON.CM.99.L1            D-Q-----R----------KTEEG--LT---------KP--TTENVGGDTNQYNLRKQ  1012
MON.NG.x.NG1            D-S-----H--------L--KTDQGEVIT----------P--SKEDVGSKPSAH-IREV  1015
SUN.GA.98.L14           D-Q----------------KQ.GEN-L--------LVK---GESSSVEM-G*......  1016
WRC.CI.97.97CI14        -QQ-Q--G--P------L--ETPEG.-IT-------L-KVWNGEGMDRSSSDKN*....  1022
WRC.CI.98.98CI04        -QQ-Q--G---------L-VETPEG.-IT--K------KVWNGEGMDRSSSNKD*....  1021
WRC.GM.05.Pbt_05GM_X02  SQQ-Q--G--V------L-V-TPEG-LLT---------KTWDGKGMDSSTTN..*....  1012
LST.CD.88.524           D-Q-R---Q----------VKEGE.N-FSI------LVK---EGPKDSQSSMDN-*...  1020
LST.CD.88.485           D-Q-----QI---------VKEGD.T-FS-------LVK---EGPKDSEGSLDNN*...  1019
LST.CD.88.447           D-Q-----Q----------VKEGD.T-FS-------LVK---EGPKDSEGGLHNN*...  1019
LST.KE.x.lho7           D-Q-----Q----------VKEGE.N-FS-------LVK---E.GPK-SESSLDNN*..  1019
SYK.KE.x.SYK173         .NE-----R----------V-TEEG--FA--------T-H-ER-DSGSH-ENDPKT--  1037
SYK.KE.x.KE51           .SD-Q---S------------TPDQQVIA---------SSD-ERVDSGTHLEITSKSN-  1041
COL.CM.x.CGU1           TGE-Q--GE-V---------RNSQGTLF-K----V--T-LQYGEDV-SENLL-NGQKEA  983
```

```
                                Vif start
H1B.FR.83.HXB2                  MEN..................RWQ..VMIVWQVDRM..RIRTWKSLVK...HHMYVSGKARGWFYRHHYESPHPRIS.SEVHIPL.GDAR.........LVITTYWGLHTGE.....RDWHLGQGVSIEWRKK...........................RYSTQV   98
H1A1.UG.85.U455                 ---..................---..----------..K----N----..------K--Q---------R-S-V-.--------.-E--.........--VR-----------.....K-----H-------L-...........................------   98
H1B.US.90.WEAU160               ---..................---..----------..----------..----I-K--K--S------T----.--------.-EGK.........---------------.....----------------Q...........................------   98
H1C.ET.86.ETH2220               ---..................---..-L--------..K----N----..---HI-RR-N--V-----D-R--KV-.--------.-E--.........-I-K-----Q---.....-------H-------LR...........................S-N---   98
H1D.CD.84.84ZR085               ---..................---..----------..--N-------..Y--HI-K--K-------D----K--.--------.----.........--V-----------.....-E------------R...........................------   98
H1F1.BE.93.VI850                ---..................---..L---------..--N-------..Y-----K--K--S----FQ-R---V-.--------.EEVK.........-----------P--.....-E-----------QG...........................K-R--I   98
H1G.SE.93.SE6165                ---..................---..----------..-----H----..------K--------P--A-R---V-.--------.---T.........--V-----------.....K--Q--H-------QR...........................--R---   98
H1H.CF.90.056                   ---..................---..----------..--N-------..Y--HI-R---------F--T----.--------.-E--.........---------N---.....-E-----------L-...........................------   98
H1J.SE.93.SE7887                ---..................---..----------..--N-------..Y--N--K---K-L----D-N--K--.--------.-E-I.........---------Q---.....---------------QR...........................--R---   98
H1K.CM.96.MP535                 ---..................---..----------..--N-------..Y--HI-K--NR-Y------R--K--.--------.---E.........--V-----L---.....---------------L-...........................--R---   98
H101_AE.TH.90.CM240             ---..................---..----------..-----N----..----I-K--KK---------Q--KV-.--------.-E--.........----R-----Q---.....K-----H-------QR...........................K----I   98
H102_AG.NG.x.IBNG               ---..................---..----------..-----N----..------K--K-------F--R--KVC.--------.----.........--VR-----------.....------H------KQ-...........................-----I   98
H103_AB.RU.97.KAL153_2          ---..................---..----------..-----N----..--I-I-K-----V-K-----RN----.--------.---K.........---K-----------.....---------A------E...........................------   98
H104_cpx.CY.94.CY032            -A-..................---..----------..K----N----..------K--K-*--------R--KV-.--------.-E--.........--VR-----QP--.....Q----H-------LR...........................------   97
H1N.CM.95.YBF30                 ---..................---..--V-------..K--K-N----..------K--K--Y------TH--K--.------V.-Q--.........--TV-----T---.....QS----H-------LR...........................K-K---   98
H1O.BE.87.ANT70                 ---..................---..-L-------Q..KVKA-N-----..Y-K-R-R-TEN-W-------RN--V-.-S-Y--V.-V-H.........V-V------MP--.....--E---H-------Y-...........................K-K--I   98
H1O.CM.91.MVP5180               ---..................---..-L----I--Q..KVKA-N-----..Y-K-M-K--AN-R-------RN-KV-.-A-Y--V.AE-D.........I-V------MP--.....-EE---H------QY-...........................E-K--I   98
CPZ.CD.90.ANT                   I-K..................-*K..IEMI------..--NI--------..--IWETKVLKP-K-------ND--KKG.E------PTLDKK.........--V-VF---QC--.....-P----H-------CG...........................K-I---   99
CPZ.CM.01.SIVcpzCAM13           ---..................K--..----------..--KA-N-----..Y---KTK--KD--------H-N--V-.-------FKDNSK.........-I-----A-T---.....-A----H----Q--MG...........................S-V---   99
CPZ.CM.05.SIVcpzEK505           ---..................---..--V-------..-V---N----..------K-------------TDN-K--.--I-----.-E-K.........----V-----MP--.....-P----H-------QG...........................I-R--I   98
CPZ.CM.05.SIVcpzMB66            ---..................---..----------..--K--N----..------R---N----------NN-K--.--------.-E-K.........-----------T---.....-A----H-------QR...........................S-R--I   98
CPZ.CM.98.CAM3                  ---..................--L..----------..-VKA-N-----..C---KTK--KD---------S--KV-.--I---I.EK-K.........--V------MP-Q.....HA----H-------CG...........................S-----   98
CPZ.GA.88.GAB2                  --S..................---..------L---..------N-----..F---K-K-----------D--N-KV-.------I.R--K.........I-VC---N-MP--.....QP----H-------QR...........................T-R--I   98
CPZ.CM.98.CAM5                  ---..................---..--V-------..--K--N-----..T---RTK--KK---------DN-KV-.--------.DT-K.........--V------T--Q.....HA----H-------CG...........................S-K---   98
CPZ.CM.05.SIVcpzMT145           ---..................---..-Q--------..-VK--N-----..Y---R-K--KN-------DGTN--MA.-QIY--I.NQ-K.........-IV------MP--.....------H--A----ER...........................G-----   98
CPZ.TZ.00.TAN1                  ---..................---..-QV--MI---..-L---T-----..--IFTTKCCKD-K------TDT-KRA.G-I-----.TERSK.........--VLH---AC--.....-P----H-IGL---QG...........................K----I   99
CPZ.CM.05.SIVcpzLB7             ---..................--K..-L----I---..--KA-N-----..Y-I---R-----------EN-KV-.--I-----.-E-K.........----V-----TP--.....------H------KQ-...........................K----I   98
CPZ.US.85.CPZUS                 ---..................---..-L--------..-VKA-N-----..Y---R-K--K----------TN--VN.------I.DK-K.........I--I-----K---.....QA----H-------QG...........................T-K--I   98
CPZ.GA.88.GAB1                  ---..................---..----------..--K--N-----..Y-I-R-K-----------DH-N-KVA.--I---FRDYSK.........-IV----A-SP--.....-A----H----Q--LG...........................S-V---   99
MUS_1.CM.01.1085                -APR...............KM-V.VSP-QETSE-...K-DWIIRAT-Y..YILS.GR..AAFT-V---QLY-Q-F-Q-KIR---TQNT-ESGAVESTYIEL-VL-DVTNVGPASCSKSTWWK-SFIL--VYLRR..................HPQDREQDYIW---YL  126
MUS_1.CM.01.CM1239              -APR...............KM-V.VSP-QETTE-...KMDWIIRAT-H..-IFS.GK..TPFV-V---QVQ-Q-F-QNKIKLA-TQNTLESGETEVTFIE--VL-DVTNVGPASLSKSTYWK-SYIL--VYMRK..................GPRDRERDYVW-H-YL  126
MUS_2.CM.01.CM1246              -APG...............KM-I.VSP-QYMSE-...KMDWLIRCT-H..YIFS.GK..APFT-V---QLQ-Q-FTQNKIK---SINRIEGGPTETTYIE--ILLDVTNVGPASLS-STYWQ-SYILK--YIRA..................APQNREIDMVH-E-FL  126
MUS_2.CM.01.CM2500              -AQR...............KM-I.VTP-QFMTE-...-MDWLIRCT-H..-ILS.GK..APFS-I---QVQ-Q-FTQNQIK---TIKK-EEDKVESTYIQ--IL-DVTNVGPASLSKSTYWQ-SYIL--VYVRA..................APQDREVDMVW---YL  126
                                Vif start
MAC.US.x.239                    --E...............EK--I.A-PTWRIPE-...LE-WHSLIKY...LKYK.TKDLQKVC-VP-FKVGWAWWTC-R-IF--QE.........GSH-EVQG--H-TPE......KG-LSTYA-R-T-YS-...........................NFW-D-  100
H2A.GW.x.ALI                    --E...............GKS-I.V-PTWRVPG-...MEKWHSLVKY...LKYR.TKDLEKVC-VP-HKVGWAWWTC-R-IF--Q-.........RSH-E-QA--N-TPE......KG-LSSYA-R-T-YTE...........................KFW-D-  100
H2A.DE.x.BEN                    --E...............DRN-I.V-PTWRVPG-...MEKWHALVKY...LKYR.TKDLEEVR-VP-HKVGWAWWTC-R-IF--Q-.........KSH-E-QA--N-TPE......KG-LSSHA-RLT-YTE...........................KFW-D-  100
H2A.SN.x.ST                     --E...............GK--I.A-PTWRVPG-...ME-WHSLIKY...LKYR.T-DLEKVC-VP-HKVGWAWWTC-R-IF--K-.........ESH-E-QA--N-TPE......KG-LSSYS-RLT-YTE...........................KFW-D-  100
H2B.GH.86.D205                  --E...............EKD-I.V-PTWRIPG-...LE-WHSLIKY...LKYR.T-ELQQVS-VP-HKVGWAWWTC-RIIF--NK.........GAW-EVQG--N-TPE......-GFLSSYA-RLT-YER...........................NFY-D-  100
H2B.CI.x.EHO                    --E...............EKN-I.A-PTWRIPC-...LE-WHSLIKY...LKYR.TKDLQQVS-VP-HKVGWAWWTC-R-IF--KE.........GAH-EVQG--N-TPE......-GFLSSYA-RLT-YER...........................SFY-D-  100
H2G.CI.92.ABT96                 --E...............EK--I.A-PTWRIPG-...LEKWHSLIKF...LKFK.TKDLQKAV-VP-HKVGWAWWTC-R-IF--TK.........EAH-E-QG--N-TPE......KG-LSQYA-RLT-YTR...........................KFY-D-  100
H2U.FR.96.12034                 --E...............EKN-I.V-PTWRVPR-...LE-WHSLIKH...L#YN.TKELQKVC-VP-HKVGWAWWTC-R-IF--KK.........GSH-EVQA--N-TPE......KG-LSSYA-RLT-YTR...........................GFW-D-   99
MAC.US.x.251_1A11               --E...............EK--I.A-PTWRIPE-...LE-WHSLIKY...LKYK.TKDLQKVC-VP-FKVGWAWWTC-R-IF--QE.........GSH-EVQG--H-TPE......KG-LSTYA-RVT-YS-...........................NFW-D-  100
MAC.US.x.251_BK28               --E...............EK--I.A-PTWRIPE-...LE-WHSLIKY...LKYK.TKDLQKVC-VP-FKVGWAWWTC-R-IF--QE.........GSH-EVQG--H-TPE......-G-LSTYA-R-T-YSR...........................NFW-D-  100
MAC.US.x.EMBL_3                 --E...............EK--I.A-PTWRIPE-...LE-WHSLIKY...LKYK.TKDLQKVC-VP-FKVGWAWWTC-R-IF--QE.........GSH-EVQG--H-TPE......-G-PSTYA-R-T-YSR...........................DLLDRC  100
SMM.US.x.H9                     --E...............EKN-I.V-PTWRIPG-...LEKWHSLIKH...LKYN.TKDLQKAC-VP-HKVGWAWWTC-R-IF--RD.........EAH-EVQG--N-TPE......XG-LSTYA-R-T-YSR...........................NFW-DA  100
SMM.US.x.PGM53                  --E...............EKN-I.V-PTWKIPG-...LEKWHSLIKY...LKFR.TKDLQKAC-VP-HKVGWAWWTC-R-IF--RD.........ESH-EVQG--N-APE......KG-LSTYA-R-T-YSR...........................KFW-D-  100
SMM.US.x.SIVsmH635F_L3          --E...............EKN-I.V-PTWRIPE-...LE-WHSLIKH...LKYN.TKDLQMAC-VP-HKVGWAWWTC-R-IF--RD.........KTH-EVQG--N-TPE......KG-LSTHA-R-T-YSR...........................NFW-D-  100
SMM.SL.92.SL92B                 --E...............GK--V.A-PVWRISR-I..VE-WHSCIKF...-KYK.TRELEKAC-VP-HKVGWAWYTA-R-IF--EE.........GSH-EVQV--N-TPE......KG-LSSYA-R-T-YSE...........................KFW-D-  101
STM.US.x.STM                    --E...............EK--I.V-PTWRIPG-...LE-WHSLIKH...LKYN.TKELSKAC-VP-HKVGWAWWTC-R-IF--Q-.........EAH-EVQG--N-TPE......KG-LSEYA-R-T-YTR...........................NFWSD-  100
MNE.US.x.MNE027                 --E...............EK--I.A-PTWRIPE-...LE-WHSLIKY...LKYK.TKDLQKVC-VP-HKVGWAWWTC-R-IF--QE.........ESQ-EVQG--N-TPE......-G-LSTYA-R-T-YSR...........................NFW-D-  100
GRV.ET.x.GRI_677                --R...............EKQ-I.VRVVWRVSE-..QIS-WRGIVTY...KIR..NKQLP.-E----WQVQWQFWTY-QFI---SK.........DDYIEVNI-HN-TPE......-G-LSSH--GLSYYHQKG.........................-K-E-  100
VER.KE.x.TYO1_patent            -NP...............NKE-V.MRVTWKVPGD..LITKWQGIVRY...WMR..QRNLK.-N-YM--QITWAWYTM-RYV--IGK.........HGEICVDL--H-TPE......QG-LSTYA-G-QYVSNLES..........................K-R-EL  102
VER.KE.x.9063                   -NQ...............EKE-V.MRVTWKVSGD..LITKWQGIVKY...WMT..KRNLK.-Q-MM--QINWAWYTMCRYI--IGL.........EGEIQ-DC--H-TPE......-G-LSTYAIG-QYLGAVG...........................NFR-EL  101
VER.KE.x.AGM155                 -SQ...............EKH-V.MRLTWKVQEE..VITKWQGIVRY...WMN..KRNLK.-E-KM--QITWAWYTM-RYV---P-.........SGEIHVDI--H-APK......QG-LSTYA-G-QYVSLVND..........................K-R-EL  102
VER.DE.x.AGM3                   -NQ...............EKE-V.MRVTWKVPEE..LITKWQGIVRY...WMR..TR-LD.-K--M--QITWAWYTM-RYE---GQ.........HGSIHVDL--H-TPE......KG-LSTYAEG-QYLSNRDP..........................W-R-EL  102
TAN.UG.x.TAN1                   --R...............EKL-V.TRLTWRVSGE..HIDKWKGIVKY...-MR..NRLQD.-T-LM--QCGWAWYTC-RFL---G-.........EGKI-VDC--H-TPE......QG-LSTYA-A-SFENWQN...........................T-K-E-  101
SAB.SN.x.SAB1C                  --................KH-I.VRPLWKVTGG..QQE-WTSLVKY...-MHVSKQCVH.-R-TP-TKIRWNWY-YQ-WV---KD.........GALIKV-N--H-TPE......KG-LETYATG-GYS...KG..........................EWF-EL   99
OLC.CI.97.97CI12                --..................-I.A-PMYRLNNG..QVEHVKFMAK......LAYKFWGYH-IWQNC-KY-REWWTHGKIY--VP-GE.........IE--L--HIRH-......KGLQNSG--NFKYIVG...........................ISWMYL   94
DRL.x.x.FAO                     --................KM-I.VIPTWRVERY..MVE-WNSLCKY...-KHKGEKYLER-H-AP-FQCS.GWWTH-QWT--FKD.........KSKI-V-VL-N-TPD......KG-LSKYAIT--HIQE...........................KFC-LI   99
RCM.NG.x.NG411                  -A................EK--I.ERP-WKVDR-..KVEQWHSLVKY...-QYKGKKE-KD-E-VP-FKV-WGWW-R-------GN.........NTKIKV----S-TIE......KG-LGTY--G-AYIDNRCDP.........................C-F-DI  105
RCM.GA.x.GAB1                   -A................EKM-I.VRP-WRVDR-..KIEQWHSLVKY...-QYKGKKA-KE-E-VP-FKV-WGWW-H-------EE.........GSK-K-----N-TVE......KG-LGTY--G-L-I-G...........................D-V-D-  101
MND_1.GA.x.MNDGB1               --................RVERI.VRLTWKVSSQ..RIEKWHWLVRRQMAWATANNEEGCW-L-P-FMAYNEWYTC.-K-V-IINR-I-.........-IVRS--H-QIE......VGCLSTYA----AVVRPPPFEK........................EWC-EI  109
MND_2.x.x.5440                  --................KQ-I.AIPTWE.DRQDVKLRDWNSIVKY...-KYKGEKHLEN-QLYP-FQCS.GWWTH-QKK--FKD.........GSKII--VL-N-TPE......KG-LSQYAFTV-YQ-E...........................N-F-YI  100
MND_2.CM.98.CM16                --................KQ-I.VIPTWKLDRH..AVE-WNSLVKY...-KYKGEKHLDK-DLKA-FQCS.GWWTH-QKY---EE.........DEQII--IL-N-TPE......KG-LSTYAMT--YKFT...........................N-Y-HI   99
MND_2.GA.x.M14                  --................KQ-I.VIPTWRIDR-..QLEKWNSLVKY...-KYKEEKHLDK-ELF--FQCS.GWWTH-QKI--FKD.........GSKII--AL-N-TPE......KG-LSQYAITA-YN-G...........................D-Y-HI   99
GSN.CM.99.CN166                 --RQR.............QKM-V.VQP-QIMTE-..KVDWLLRATKH...-IWS.GK..TPFV-V---QLQ-Q-FTQNKIRLAMDTRKVG.EEVEATYIE--IL-DTTSHGPASLS-STYYQ-A-I---VYNRS..................KLGAREGDIIW---NL  127
DEB.CM.99.CM40                  --................KT-I.VRLTHKIPMN..KLD-WCSAIKR...EKYK.TKELDTVQWI--F-LRNSYYTQTKIVF-ITK.........EGGV-VELL-C-TPE.....KGWLPTMA.-A-A-ET...........................KTWK-EL   99
DEB.CM.99.CM5                   --................KE-I.VRLTHKIPVN..KLDKWCSAIKR...GKYK.TKELQAIKWI----LRRAFYTQTKIVF-VTE.........EGKV-VEIF-S-APE.....KGWMPSMA.-G-S-EQ...........................TGWQ-EL   99
GSN.CM.99.CN71                  --RQR.............QKM-V.VQP-QVMTE-..KVEWLIRATKH...-IWS.GK..TPFV-I---QLQ-Q-FTQNKIRLA-DTRQVG.EEVEATYIE--IL-DTTSHGPASLS-STYYQ-A-T---VYNRS..................KLGARETDITW---NL  127
DEN.CD.x.CD1                    -Q-LGRENR.........TKE-I.VRLTYKIGET..KLEHWMST-KS...IIYK.TNELK-TVIIP-FRITWEWYTMTKI----R..........AAYMEVKL--H-TPE.....KGYLSQVA.I--AYVHL...........................STGWA-E-  108
TAL.CM.01.8023                  -D................KL-V.TRLTWKVPRN..RMQ-WVGTIRY...-MYK.TKYLQKLIWV--FQLRMASFTQAKWIL--EEKENR.....VEYIT--Q--C-APD.....KGYTSDWDTI----SVSDV.........................TLDAKV---EL  111
TAL.CM.00.266                   -D................KL-V.TRLTWKVPRN..RMQ-WVGTIRF...-IHK.TKYLQKLIWV--FQLR-ASYTQAKWIL--EEKDTR.....VEY-T--Q--C-APD.....HGYTSDWDTI----SVSDI.........................VRDEKV-A-E-  111
MON.CM.99.L1                    -APR..............KM-V.VTPTNIISES..KMDWLIRCTKW...-ILT.GS..APFV-V---QLHNQ-F-QNKIKL--DLGITQEGESWATY-E--I--D-TNVGPSALSPTVYNK-AYT---VYWMQKVERVP.....LDPVWNPNPR..DREV.RTW-A-HL  138
MON.NG.x.NG1                    -APR..............KM-V.VSPVHIRSE-..WMEWLIRATKY...NIYT.GK..VPFV-V---QLT-Q-F-QCKIKL--EMGITQXXEQWATYXE--IL-DXTNVGPASLSPSVKWK-AYI---VYWRQ..PRVPQQEGVVDRVWFPSPRRPDREIDRVW-A-YL  144
WRC.CI.97.97CI14                --................KE-I.G-PLIRISKG..KFR-WTYYIA......NLIKAGQANFLWGSWHAAAECFTR.HI-K--IQ-HWE.........MQ-RC--N-SPE......KG-MDTM-FN-RMVYEN...........................-WE-FI   97
WRC.CI.98.98CI04                --................KE-I.G-PLIKISKG..RFK-WAYWIA......HLMQN-KASFLWGTWHTQAECFTR.YLLR---Q-HWE.........-Q-RC--N-APE......KG-LDTM-FN-RMVYEN...........................-WE-FI   97
WRC.GM.05.Pbt_05GM_X02          --................KE-I.A-PLIKISKG..RFR-WSYWIA......ELIQE-QARFLFGSWHVHAECFTR.HI-R---Q-HWE.........MQ-RC--N-APE......IG-IDTL-FN-RMVWEN...........................KWE-FI   97
SUN.GA.98.L14                   --ENPPQWRWWDEREWEDRRQYKIVRIVWLIDRI..AVEKFLDLRR..M-RET...-DDWVSMYGTGTGWEWYTY.NKII--VTYGT.........V-VRI-GH-TPA......KGLINQW-C-M--IYN...........................S-Q-EI  113
LST.CD.88.524                   --................RDQRI.VRVAWITNRS..LVEKFLDIKR...N-SES...-NDWVSLHGTGVGWEFYTY.NKLL---DTGT...........--VRI-AH-AA-.......-G-ISQWA-----IYG...........................N-Q-E-   96
LST.CD.88.485                   --................RDQRI.VRVAWITTRN..LIEKFLDIRR...A-KET...QVDWVSLHGTGVGWEFYTY.NKIV---EIGT...........--VRC-AH-AA-.......-G-ISQWA-----IYG...........................N-Q-EI   96
LST.CD.88.447                   --................RDQRI.VRVAYITTRS..LVEKFLDIRR...A-KET...QNDWVSLHGTGVGWEFYTY.NNIV---ETGT...........--VRC-AH-AA-.......-G-ISQWA-----IYG...........................N-Q-EI   96
LST.KE.x.lho7                   --................RDQRI.VRVAWITTRS..LIEKFLDIVR...LNTEA...-ENWVGLHGTGVGWEFYTY.HKLV---ESGT...........--VRI-AH-AA-.......-G-ISQWA-----VYR...........................D-Q-EI   96
SYK.KE.x.SYK173                 --................KE-I.V-PTWRMTP-..QID-LQHIIKT...-KYK.-KELEKAT-K---QIEWQWYTYCQWT--VG..........DGTIW--F-HN-APE.....RGWL-MQG.IR-QYQWN...........................QWN-DL   98
SYK.KE.x.KE51                   --................KE-I.V-LTWKLHPN..QIDKMQHI-KV...-KYK.-KQLEKAT-K---QITWEWWTRAQWE--VG..........QGV-I-KF-HN-TPE.....KGLLRTEG.-GLI-AHPS...........................GW--E-   99
COL.CM.x.CGU1                   --K.................M-A.VRICYPMGK-..KPKPLREWIK-.........QFPECQY-YLGKTESPEWWTEGKIEFY-GSAIK..........-S-IYLGT-TPDR.......Y-R-PP--Y-LQIG...........................KWE-DL   93
```

```
                                                                                                                                                                    Vif end
H1B.FR.83.HXB2          DPELADQLIHLYYFDCF.SDSAIRKALLGH..IVSPRCEYQ..AGHN...KVG...SLQYLALAALI.................................TPKKIKPPLPSVTKLTEDRWNKPQKTKGHRGSHTMNGH*.........................  192
H1A1.UG.85.U455         --D---H----H-----.-E----R-I--Q..--R------.----...---...-------K--V.................................--TRA-------K-------------R-----R-L-R--.........................  193
H1B.US.90.WEAU160       --D---------H----.-E---N-I---.L-I------.----...---...-------T---.................................-------------K---------------------------.........................  193
H1C.ET.86.ETH2220       --G---H---MH-----.AE-----I--Y..R-----D--.----...---...-------T---.................................K---A-------S--V--K-------R-R--N-------.........................  193
H1D.CD.84.84ZR085       --G-------M------.A------I---..---------.----...---...-------T---.................................A---R-------K----------RQ---R----------.........................  193
H1F1.BE.93.VI850        --G-------I------.-E-----I---..RI----N--.----...---...-------T---.................................A-E-T-------Q--V--------E-R------------.........................  193
H1G.SE.93.SE6165        E-D---H----H-----.-------I--Q..---------.----...Q--...-------KV-V.................................-S-RSR-------E-A----------R---ENP------.........................  193
H1H.CF.90.056           E-G-------MH-----.-E-----I--R..V-R---N-P.---K...Q--...T------T--V.................................A-----------R--V----------R------------.........................  193
H1J.SE.93.SE7887        --G-------MC-----.-------I--Q..------D--.----...---...-------T---.................................R--RR-------Q--V----------T---E--------.........................  193
H1K.CM.96.MP535         E-D-------I------.-E--V---I---..R------C-.----...---...-------T--V.................................A-RRP---V---K--V----------R-----Q------.........................  193
H101_AE.TH.90.CM240     --D---K----Q--G--.-------I--Q..V-RR----P..S---...---...-------K--T.................................---R-R------.EI-----------.R---ENP------.........................  191
H102_AG.NG.x.IBNG       --D------------N--.-------I--E..--R------.----...---...-------N--V.................................A-T-T-------R--A----KE----R-----RP-----.........................  193
H103_AB.RU.97.KAL153_2  --N--------------.-E---N-I---..R---S---R..----...---...----------R.................................--------------------------R--D------S---.........................  193
H104_cpx.CY.94.CY032    --D-------MH-----.-E-----I---..R--------.----...---...-----------.................................S---T-------K--V----------R-R-ENQI-----.........................  192
H1N.CM.95.YBF30         ---M--K----H-----.TA----Q-V--R..P-L-----P..---K...Q--...T------T-WV.................................GA--R---------------EH--MQ----NPI-----.........................  193
H1O.BE.87.ANT70         ---T--RM---H--T--.TA--V---I--Q..R-LTK---P..T--S...Q--...T--L---R-VV.................................KARSR-------Q---------HLRIRDQLK-PS-----.........................  193
H1O.CM.91.MVP5180       ---T--RM---H--T--.TE-----I--Q..R-LTK---L..---S...Q--...T--F---K-VV.................................KV-RN-------QR---------W-IRDQL---S-----.........................  193
CPZ.CD.90.ANT           ---T---M--Q---P--.--Q-V-Q-I--E..RILTY-H-K..K--S...Q--...T-----FCKILEFRG.................................Y--GPRRQF--LSI---------RRMR---ENQ*.........................  192
CPZ.CM.01.SIVcpzCAM13   --YV--K---SQ-----.A-----R-I---..--E---A-R..E--R...Q--...T--F---K--V.................................KE-RLR------A--------T--R----PEN---S---.........................  194
CPZ.CM.05.SIVcpzEK505   ------K----------.TA----Q-V--R..P-I-K---P..---K...Q--...-------I-WV.................................GVQ-RR-------------RH--I-D------T--L-.........................  193
CPZ.CM.05.SIVcpzMB66    --D--------------.AE-----I--Q..--------P..T---...---...-------ST--.................................KS-PRR------R--A-----NS---RD-K-N-I-----.........................  193
CPZ.CM.98.CAM3          ---M--K----Q-----.AA-----V--K..Q-Y-K---P..---Q...Q--...-------R-WV.................................RVG-K--------------RH--IRV-QE--------.........................  193
CPZ.GA.88.GAB2          --QT--R---C------.AA--V-R-I--E..--TKV--FP..R--P...Q-Q...------IVQVV.................................KT--K--------I-A----------RDQQENL--S---.........................  193
CPZ.CM.98.CAM5          ---T--R----H-----.AA-----V--K..Q-Y-K---P..T--Q...QX-...-------R-WV.................................GVQ-KT-----IX--A------Y---R--QEN----X--.........................  193
CPZ.CM.05.SIVcpzMT145   ---T--R----H-----.TAT---N-I-EK..A-L-K--FP..E--S...---...------W--KQ-V.................................KT--NR------A--A--------R-RD-QE-PMSS-C-.........................  193
CPZ.TZ.00.TAN1          ---T------TR--T--.AAG-V-Q-I--E..RILTF-HF-..S--R...Q--...T--F--FRKVVESQDK.................................Q--GPRR--------------HRT-T-R-EN--LS-C-.........................  199
CPZ.CM.05.SIVcpzLB7     ------HM------N--.AE-----I--R..V------FP..---K...---...T-----IS--L.................................QRQPTR------S--A------S----------I-S---.........................  193
CPZ.US.85.CPZUS         ---M--K----------.TA------V--R..P-I-K---P..T---...---...---F---K--V.................................GQS-RR-------------C-----QPEN---S---.........................  193
CPZ.GA.88.GAB1          --FT--R---SQ-----.AET---R-I--Q..L-A-----K..E--R...Q--...---F---K---.................................SERRHR------A---------H-R--V-QENL-R----.........................  194
MUS_1.CM.01.1085        S-T--MRI--TQ--S--.YSQD-PRVIR-QH..TLGD--H....PDAHT.---PPP---V---L--KK.....................H..GKLAP......QP-SALALQSGPNHHAPGT-SHVGS-RR-RKTLFQRRAPWNLE*.....................  235
MUS_1.CM.01.CM1239      T--I-MKI--THH-S--.YSQD-PRVIR-QP..PLGD--H....PGAHT.---PPPT--V---LT-KK.....................H..GKSTT......QPQTSMALQSGPNYHAAGPTGHVGP-RR-RKTLFQRRAPWNLE*.....................  235
MUS_2.CM.01.CM1246      ---VTM-I--TH--S--.QQRD-QR-IR-EQ..LLRK--H....IKTHYP---TPL--EK--FF-Y-K.....................HTNGASAS......KS-SLMAKSSNANNYGSGASRPVGT-RR-RKTLLQRQAPWNLG*.....................  238
MUS_2.CM.01.CM2500      -----M-I--T-H-S--.QEQD-IR--R-ER..ILGI-QH....QTAHLP---TPP--EK--FF-Y-K.....................QGDGATPS......KS-SPMAK-PGTNNHGAGTKRHVGS-RR-RKALLQGRAPWDLGPSHGAPSRP*.............  247
                                                                                                                                                                                  Vif end
MAC.US.x.239            T-NY--I-L-ST--P--.TAGEV-R-IR-EQ..LLSC-RFP..RA-KYQ.....VP-------KVVS.......................DVRSQGE......NPTW-QWRRDNRRGLRMAKQNSRGD-QR-GKPPTKGANFPGLAKVLGILA*...............  214
H2A.GW.x.ALI            T-DC--S---GT--S--.TAGEV-R-IR-EK..LLSC-N-P..QA-KSQ.....VP---F---VVVQQ.......................NGKPQRN......S-TR-QWRRDYRRGLRVARQDSRGL-QR-GESPAPGAHFPGVAKVLEILA*...............  215
H2A.DE.x.BEN            T-DC--I---ST--S--.TAGEV-R-IR-EK..LLSC-N-P..QA-KAQ.....VP-------VVVQQ.......................NDRPQRK......G-AR-QWRRDHWRGLRVAREDHRSL-QG-SEP-APRAHFPGVAKVLEILA*...............  215
H2A.SN.x.ST             T-DC--S---ST--S--.TAGEV-R-IR-EK..LLSC-N-P..QA-KYQ.....VP---F---VVVQQ.......................NGRPQRD......N-TR-QWRRNYRRGLRVARQDGRSH-QR-SEPPAPRAYFPGVAKVLEILA*...............  215
H2B.GH.86.D205          T-DV----L-GS--S--.-ANEV-R-IR-EK..ILSY-N-P..SA-EGQ.....VP---F---RVVQE......................GKNGSQGE.....SATR-QRRRNSRRSIRLARKN-NRAQQGSGQPFAPRTYFPGLAEVLGILA*...............  216
H2B.CI.x.EHO            T-DV--R-L-GS--SS-.TANEV-R-IR-EK..ILSH-N-P..SA-TGQ.....VP---F---RVVQE......................GKDGSQGE.....S-TR-QRRRNSRRGIRMARDNIRTSQQSSSQSLAQGTYFPGLAEVLGILA*...............  216
H2G.CI.92.ABT96         T--T----L-GS-----.TAGEV-R-IR-EQ..ILSC-N-P..TA-KRQ.....VP---F---QVVQK......................G.HGSKGE.....SXTR-QRRGDNRRGIRMARKNSXRDQQDSSQSFTQGAYFPGLAKVLGILA*...............  215
H2U.FR.96.12034         T-DC----L-GS--S--.TAGEV-R-IR-EQ..LLSC-NFP..SA-RQQ.....VP---F---R--QD......................G..RSKRE.....NPTR-FRRRNCRRGFQLARKNCERYQQGSSAT-SQGTYFPGLAKVLAILA*...............  213
MAC.US.x.251_1A11       T-NY--I-L-ST--P--.TAGEV-R-IR-EQ..LLSC-RFP..RA-KYQ.....VP-------KVVS.......................DVRSQGE......NPTW-QWRRDNRRGLRMAKQNSRGD-QR-GKPPTKGANFPGLAKVLGILA*...............  214
MAC.US.x.251_BK28       T-DY--I-L-ST--P--.TAGEV-R-IR-EQ..LLSC-KFP..RA-RYQ.....VP-------KVVS.......................DVRSQGE......NPTW-QWRRDNRRGLRMAKQNSRGD-QR-SKPPTKGADFPGLAKVLGILA*...............  214
MAC.US.x.EMBL_3         NTR-CRHFSCIALISL-.TAGEV-R-IR-EQ..LLSC-KFP..RA-RYQ.....VP-------KVVS.......................DVRSQGE......NPTW-QWRRDNRRGLRMAKQNSRGD-QR-GKPPTKGADFPGLAKVLGILA*...............  214
SMM.US.x.H9             T-DY--T-L-GT--P--.-EGEV-R-IR-EK..LLSC-KFP..KA-KNQ.....VP-------TVVS.......................HVXSQGE......XPTW-QWRRNXRRGXRLATQNXRRN-QGSSESFAXGANFPGLAKXLGILA*...............  214
SMM.US.x.PGM53          T-DY--T-L-GT--P--.-EGEV-R-IR-EK..LLSC-KFP..KA-KNQ.....VP-------TVVS.......................HVRSQRE......DPTW-QWRRNNRRGIRMAKQNSRRN-QGSSESLAEGANFPGLAKVLGILA*...............  214
SMM.US.x.SIVsmH635F_L3  T-DC--T-L-ST--P--.-EGEVQR-IR-EK..LLSC-KFP..KA-KNQ.....VP-------TVVS.......................HVRSQRE......DPTW-QWRGNNRRGFRMAKQNSRRN-QGSSKSPAEGANFPGLAKVLGILA*...............  214
SMM.SL.92.SL92B         T-DV----A-ST--P--.AAH-V-Q-IR-EQ..-LSY-G-A..VA-HSS.....VQ---L---KVVLQ.......................NDRPKGK......NPTR-QWRRNNRRGLRMVTQYSGGPESNSSTTPTKRVNFPGLEKVLGILGR*...............  217
STM.US.x.STM            T-DC----L-GT--P--.TAGEV-R-IR-EK..LLSC-RFT..KA-KNQ.....VP-------KVVE.......................HVRSQRE......N-AR-QWRRGNRGSIRVATQNGRGH-PR-SKP-TEGTDFPGLAKVLGILA*...............  214
MNE.US.x.MNE027         T-DY--I-L-ST--P--.TAGEV-R-IR-EQ..LLSC-RFP..RA-KNQ.....VP-------RVVS.......................YVRSQRE......NPTW-QWRRDNRRSLRMAKQNSRGD-QR-SKPPTKGADFPGLAKVLGILA*...............  214
GRV.ET.x.GRI_677        --GT--RM------N--.T-R--QQ-IR-E...KYTW-TFK..E--KGQ.....VQ---L---V-Y.....TNGIR..............KRSKRT.....FTRMAGNLGSRQGAMGRMAT-HAQGS-RRSQKALWNEHANPSMELLCRGGKET*.............  219
VER.KE.x.TYO1_patent    --AT--SI--GH--N--.KER--QQ--R--...RFVF-QFP..E--KST...-QVP-------L-H.....QNGLR..............ERSKRG.....KTRRSRNLGSKQGAVGRMAK-YVTRSQPG-EAAFWERTPVPSMELLSGGRRKTWYSHDGKGLQIL*.  235
VER.KE.x.9063           --AT--GM---H--S--.TER--Q--IR-E...RFVF-QFP..E--KTT...-KVP-------L-H.....QNGIR..............ERSQRK.....QARRSRDLGSKQRTMGRMAKGHVRRSQSGSQITFWQRAPLPSLELLSGGRRETGSPNDGEGL*....  231
VER.KE.x.AGM155         --NT--SM--CH--T--.T-R--QQ--R-N...RFIF-QFP..G--KLT...-QVP-------L-H.....QNGLR..............KRSQRG.....ETRRTRNLGSQQGAVGRMAQ-YGRRNQQRSQTAFWPRTPIPSMELLSGGRGETGKTHSGKGI*....  232
VER.DE.x.AGM3           --AT--S---TH--T--.TER--------Q...RFTF-QFP..E--KKT...-QVP-------L-H.....QNGLR..............QRSQRS.....KTGGTRNMGFEQGAVGRMAK-HARRYQSGSQDAFWARAPVPSMELLSGGGRKESHSHARKGL*....  232
TAN.UG.x.TAN1           T-DV--HM--CH--P--.T-R--QQ-IR-E...SFLW-T-K..E--VAENHW-QVR---F---TVYTDF..LRNGRR..............KRFQGK.....KTRMVRNLGSQQGAVGRMIK-HGSRTQSGSTTPFWERTPLPSMELLSGRRGKEWGTNDRKGL*....  238
SAB.SN.x.SAB1C          --WT--HI--WS--P--.T-R-VQQ-IR-E...KYLW-KH-..V--QPT...-QVP-------RVY......TNGLRRVAPTSRRGSSQGSPQESQRR.....DTRMARNMGFAQRAVRRMAP-HVTGPQFR-PVPLPKESPFPSLVEYCGRTSH*.............  234
OLC.CI.97.97CI12        T--YV-FMV--V--A-W.TQ------I--MP..DHRY-QLET..--LG......VL---L---R-VRNG......................................RKSSRGSGATQGA-GGLGE-H-GRNYCR-KKALSQGVIGPSFSILCSKTYQ*.............  203
DRL.x.x.FAO             --TT--RI---E--P--.T-Q---GV---Y...RIQA-YFP..R--KGQ.....VG-------Q-H.....IKYLKH..................GRKTVSGESSSGAN..GSRRGRVRRMATEESLKNQPG-SITLPPRVPLPSLEHLHGALP*.............  217
RCM.NG.x.NG411          --TV--K-V-NQ--P--.T-K---Q-I--E...QVLL-GFE..K--RDQ.....VG------IK-W.....ARQQLKD.............GRGQRK.....SARGPH..WGWRCRVCPLASQNANRSQLRSQVALSSRVHFPSVAYLCGTLA*.............  222
RCM.GA.x.GAB1           F-WT--S---KI--P--.T-R---QS---E...KVLV-AF-..G--RDQ.....VGT--F--IQ-W.....AKSQLD...............RYGRK.....SPRGPH..WGWR-RVPALATGHARKGQLGSQVTLSSRVHFPSVAHLCGTLA*.............  216
MND_1.GA.x.MNDGB1       T--V--H----HFY---.M---VM--IR-E..E-LKV-RFP..---KAQ....GVL---F-C-RVIYGP.....................................EER*ASIRRSRT.....SQRAIQSVASRYY-GNKGRSKKALPSHYPKCSIRILCAKHRE*.....  222
MND_2.x.x.5440          --VT--RI--GE--P--.T-Q------F-E...RLVA-YFP..W--RGQ.....VGT--F---Q-Y.....LRGRKH..................GRE.....GTRGAR..RSRRGR-GAMAGNVTGENQPG-PVTLPPRVPFPSLEHMCRTLA*.............  213
MND_2.CM.98.CM16        --QT--RM--WK-LP--.TEQ---Q----K...RLTV-YFH..W--KGK.....VG-------LSY.....TAYCNN..................GRR.....GPRDPR..RSRRGGS-YLVGKISGENQSG-KITLPPRVPFPSLERMHRTLA*.............  212
MND_2.GA.x.M14          --VT--RM--WE--P--.TAA-V--V-Y-E...RIVA-YSP..W--KGQ.....VGT--L---R-Y.....IKFCRN..................GRK.....STRGTT..GGRR-G-RTVAGKIIGTSEQR-SITLPPRVPFPSLEHLCRTLA*.............  212
GSN.CM.99.CN166         T-GV-M-I--GK--P--.QNED--R-IR-EQ..ILGN--H....PEAHL.Q--TPQT-EQ--FF-YVK.....................H.YGKNPT......QSSYALAQ-KGAHNHGYRSKRSVGN-RG-RETLLKGRAPWNLA*......................  237
DEB.CM.99.CM40          T-D---H----R--P--.-QQ-V-Q-IR-EI..LLQQ-Q....QY-.RGPEGKGPP----I--RVAVK.....................HGKVPTLT.........-TTFHLKNNTHD-ARVAAGHARS-RGSPKTF-QRRTIWSLESLCRLTRPTRLDGRTSMGCIGH*  230
DEB.CM.99.CM5           T-D---H----R--P--.-QQ-VTQ--K-EK..LLQQ-Q....QY-.RGPEGKGPP------RTAVK.....................YGKVSTLT.........-TTFHIKNSTND-ASTTTSHAGS-RGSFKTL-KRRVMGSLESLCRPSCPTRLDRRTSMGGVSN*  230
GSN.CM.99.CN71          T-DV-M-I--TR--P--.QNEDV-R-IR-EQ..LLGY-DH....PEAHP.Q--TPHT-ER--FF-YVR.....................Y.YGKNPT......KS-STLDQ-KGAHNHGSCSKRHVGN-RR-RKTLLKGRAPWNLA*......................  237
DEN.CD.x.CD1            --VT--EI--KDH-S--.TRQ-V-Q--R-EV..-LST-ARKKAHY-PRTGPGREPP----IV-Q-A-R.....................YGKVTTIT.........-TS-SFKNGRNHSKAATGSSMGYHRG-KKTLQQRRD-WNMGSLLVITSAPTLDRRTSHGSRRHR  245
TAL.CM.01.8023          T--V--H---TK--P--.TA--V-Q-IR-EV..LLAV--.......VRHKGRDSPP---F---KVAVR.....................HGKVATQS........SSANNIKAHTN--SSIRKTHERS-GRDAHTL-KRR--WSLESLCRVTSRA*.............  227
TAL.CM.00.266           T-NV--H---TK--P--.TA--V-Q-IR-EV..LLAV-G.......VRHKGRDTPP---F---T-AVR.....................YGKVATQS........SSTADLQTH-N--ASARKAHERSSRRNAYTL-KRR--WSLESLCRIASRP*.............  227
MON.CM.99.L1            T-D--N-IV-TH--A--.QQLDV-R-IR-EK..LLGQ-QH....LATHYP---TPLT-EK--FL--VR.......................YGATPT......KSSAPLDL-DG-HR-TGPTGRHLGA-RRSPKTLLKGRAPWHLGTGNGAAC*...............  255
MON.NG.x.NG1            T-DV---I--TR--S--.QA-DV-R-IR-EA..XLGN-HH....IQTHYN---TPPT-EK--FL-CFK.......................YGKCTT......EPSSTLDL-RG-HRARRSAGRHMGP-RG-PXALLKGRAPWHLGSGH*...................  257
WRC.CI.97.97CI14        M-SE-NKI--SQH-E-W.TQNVV-NTIRTGKIN-IDA-Q-H..K--AG......VLT--T-C-QV-QHE.......................................QLQR.AGAFRRT........STNSRGPSRYARALC--KTAFPKGSPCKSISLLLSTERV*.....  207
WRC.CI.98.98CI04        M-SE--VM--SQHV--W.TQNVV-DTIRSGRLDIRNA-Q-H..K--AG......VLT--T-C-QV-QHE.......................................QLQG.TSTYRRT........STNSRGPSRYARTLC--KTAFPKGSPCKSISLLLATEKL*.....  207
WRC.GM.05.Pbt_05GM_X02  TVK---V---SQH-E-W.TQ--VKNTIRQGYLHPIDS-K-H..Q--LG......VLT--A-C-Q--QHE.......................................QLR-R-A-NRRT........PGNTRGSSRYARAVCH-KTAFPPGSPCKSISLLLATEKL*.....  208
SUN.GA.98.L14           --LV---M--FK----W.TSRC--R-M--E..KILHE-RN-..VA-KG.....LVL---F-C-RV-HGQ.......................................QERAS...RTATRRRGATERAICTMAGRYH-RNQGRSGKAFPSSNSCPSLAILCGLYRVRKGRVY  234
LST.CD.88.524           --I----ML-VK----W.VHRD-VR-IR-E..KILHS-KMP..T--KG.....CVL-----C-RQ-QHV.......................................QT-TAAQ-RKAPRG-WSSKRAMGSMASR-N-GNQGR-KTPFPRSNFGPSLGILCQQYRVRRGGLH  220
LST.CD.88.485           --I----M--CK--N--.-TRD-----W-E..KIMAI-AMP..T--KG.....CVL---F-C-RQ-QHV.......................................QS-AA-ESRKTSRG-WTTKRAMGSMVSR-H-GNKGG-QTPFPRSHFGPSLGILCQQYRVRGGSLH  220
LST.CD.88.447           --I--N-M--CK--N--.-TRD-----W-E..KIMAF-AMP..T--KG.....CVL---F-C-RQ-QHV.......................................QN-AA-ESRKTSRG-WATKGAMGSMASR-H-GNKRG-QTPFPRNNFGPSLRILCQQYRVRGESLH  220
LST.KE.x.lho7           --I----M--YK--N--.IQRD--R----E..RIT.M-CQP..L--KG.....CVL---FIC-RQ-QHV.......................................QAEAA-KSRETTRG-WTTKGAMGSVVSR-H-GNKRG-KTPFPRSNVGPSLAILCRQYRVRRGGLH  219
SYK.KE.x.SYK173         T-AV--R---NF--P--.TAR-VNQ-VR-EL..LTSH-WT....P-TD.....QVP-------QVYL.....................KDGGGFLQSLPACARNTMVLHSKKCR-DPKR.-QCHCKGR-GSD-SIQAFYSSRNIWSLESILK.....RR....GRD*.  221
SYK.KE.x.KE51           T-QT--A-V-SQ--P--.--R-VQQ--R-EK..LTSH-WN....F-KE.....QVL-------QKYLS.....................KDGTGFLQSLPAAARGTTILHSKKCQLGRGRSNKCHCQSR-GSP-SMQAFYSGRNIWSLESMLRGGSRSRRNTHHGLD*.  233
COL.CM.x.CGU1           SWQQ-VGYM-IIRGK-VRIEEEA---IR-LP...WNP-DF-..V--LG.....VLK--E-C-L--R..........................................................EKERTH--MRVLK-TSHGCRTR*.........................  171
```

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
H1O.BE.87.ANT70	.....	193
H1O.CM.91.MVP5180	.....	193
CPZ.CD.90.ANT	.....	192
CPZ.CM.01.SIVcpzCAM13	.....	194
CPZ.CM.05.SIVcpzEK505	.....	193
CPZ.CM.05.SIVcpzMB66	.....	193
CPZ.CM.98.CAM3	.....	193
CPZ.GA.88.GAB2	.....	193
CPZ.CM.98.CAM5	.....	193
CPZ.CM.05.SIVcpzMT145	.....	193
CPZ.TZ.00.TAN1	.....	199
CPZ.CM.05.SIVcpzLB7	.....	193
CPZ.US.85.CPZUS	.....	193
CPZ.GA.88.GAB1	.....	194
MUS_1.CM.01.1085	.....	235
MUS_1.CM.01.CM1239	.....	235
MUS_2.CM.01.CM1246	.....	238
MUS_2.CM.01.CM2500	.....	247
MAC.US.x.239	.....	214
H2A.GW.x.ALI	.....	215
H2A.DE.x.BEN	.....	215
H2A.SN.x.ST	.....	215
H2B.GH.86.D205	.....	216
H2B.CI.x.EHO	.....	216
H2G.CI.92.ABT96	.....	215
H2U.FR.96.12034	.....	213
MAC.US.x.251_1A11	.....	214
MAC.US.x.251_BK28	.....	214
MAC.US.x.EMBL_3	.....	214
SMM.US.x.H9	.....	214
SMM.US.x.PGM53	.....	214
SMM.US.x.SIVsmH635F_L3	.....	214
SMM.SL.92.SL92B	.....	217
STM.US.x.STM	.....	214
MNE.US.x.MNE027	.....	214
GRV.ET.x.GRI_677	.....	219
VER.KE.x.TYO1_patent	.....	235
VER.KE.x.9063	.....	231
VER.KE.x.AGM155	.....	232
VER.DE.x.AGM3	.....	232
TAN.UG.x.TAN1	.....	238
SAB.SN.x.SAB1C	.....	234
OLC.CI.97.97CI12	.....	203
DRL.x.x.FAO	.....	217
RCM.NG.x.NG411	.....	222
RCM.GA.x.GAB1	.....	216
MND_1.GA.x.MNDGB1	.....	222
MND_2.x.x.5440	.....	213
MND_2.CM.98.CM16	.....	212
MND_2.GA.x.M14	.....	212
GSN.CM.99.CN166	.....	237
DEB.CM.99.CM40	.....	230
DEB.CM.99.CM5	.....	230
GSN.CM.99.CN71	.....	237
DEN.CD.x.CD1	FHKNHAERNMEAL*	258
TAL.CM.01.8023	.....	227
TAL.CM.00.266	.....	227
MON.CM.99.L1	.....	255
MON.NG.x.NG1	.....	257
WRC.CI.97.97CI14	.....	207
WRC.CI.98.98CI04	.....	207
WRC.GM.05.Pbt_05GM_X02	.....	208
SUN.GA.98.L14	ESNHFAQ*.....	241
LST.CD.88.524	EIHHFS*.....	226
LST.CD.88.485	EIHYHSK*.....	227
LST.CD.88.447	EIHYHSK*.....	227
LST.KE.x.lho7	EIHHPSQ*.....	226
SYK.KE.x.SYK173	.....	221
SYK.KE.x.KE51	.....	233
COL.CM.x.CGU1	.....	171

```
                       Vpx start                                                                                      Vpx end
MAC.US.x.239           MSD.PRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPRELIFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGW.RPGPPPPPPPGLA*  112
H2A.GW.x.ALI           -AN.---TV-----------------D----AL---------------------R---------Q--T-------M---MYT-FM---T---G---P---.-S----------V-  113
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.SN.x.ST            -AG.---T------------------D----A----------------------R---------I--T-------M---M-I-S-R--T---G---P---.-S----------V-  113
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
H2B.CI.x.EHO           -.-.----V---------V-------ET-L-HL--V---------------K--A--RE-----I--T-------M---M-I-FA---G--R----P---.-S------------  112
H2G.CI.92.ABT96        -.-.----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
MAC.US.x.251_1A11      ---.------------------------------------------------------------------------------------------------.--------------  113
MAC.US.x.251_BK28      ---.------------------------------------------------------------Q----------M-------------------------.--------------  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
SMM.SL.92.SL92B        -T-.----------------------HN---AL-QT--Q-----------R-C----V----Y----A----VQ-M---M-Q-FR---T-R----SQ---.-T------------  113
STM.US.x.STM           ---.-------------A-E------H----D------------------R------------PG------------------S-R---------------.--------------  113
MNE.US.x.MNE027        ---.------------------------------------------------------------------------------------------------.--------------  113
DRL.x.x.FAO            -AERQSVERA-AEPMGAGEV-LE---Q-SLLR--Q--RL-FHP-FL-RL-NTCM-HY--ALQL-FT-S-----L-L---M---FQQ--S--QGR-PPPLRPAGDRL-----PP.-  114
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
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
MND_2.x.x.5440         -AE....-A-EAPE-AGE.VGLEQ--ETSL-R-----RL-FHP-FL-RL-NTCV-H---RHQR-LD-A-----L-MH--MYT-MQQ--P-RNG.RPR......-...-----M--  100
MND_2.CM.98.CM16       -AE....GA-EIPE-AGE.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-EAPQ-AGE.VGLEQ--E-SL-QV----QL-FHP-FL-RL-NTCV-H---RLRRTLN-A-----L-M---M-V-MQQ--P-RSG.-S.......-...--------  100
```

```
                                                                                                              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........................PEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYETYGD........TWAGVEAIIRILQQLLFIH#FRIGCRHSRIGVTR..............................QRRARNGASRS*   96
H1A1.UG.85.U455           ----........................----------YA-*A--------H-D------Q----------N----........--E-----------------.-----Q-----IIP..............................G--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-----IL-..............................------------   97
H1K.CM.96.MP535           ----........................----------N------I-----R--------P---N------T----........--E-L---------------.-----H-----IIP..............................---G---S----   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-----------------.-----------IVP..............................---V--------   97
H1G.SE.93.SE6165          ----........................----------Y---A-----------------L----------N----........--E-----------------.-----Q-----I-P..............................R--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-----IIP..............................---G--------   97
H101_AE.TH.90.CM240       ----........................--#f------Y---------------------P--Q----Y--NN----........--E-------M------V-.-----Q-----IMP..............................G--G---TG---   95
H102_AG.NG.x.IBNG         --R-........................----------S-----------H--------P------------------........--E--K-----------V-.-----Q-----IIQ..............................G--G----G---   97
H103_AB.RU.97.KAL153_2    ----........................----------Y------------S-------V---S---Y--------........-------------------.-----Q-----IIQ..............................R-----------   97
H104_cpx.CY.94.CY032      ----........................----------N---------------------P----------N----........--E-----------------.-----Q-----I-P...........................#KGEE-GD------   98
H1N.CM.95.YBF30           --R-........................---A------Y---A------------------------N----........--E-----------------.Y----Q-----I-P..............................---R---T----   96
H1O.CM.91.MVP5180         ----........................L-N---A---F------------E--------P--QAC--Y--------........--E--M-----------T-.Y----Q-----ILP.........................SNTRG-GR---S----  101
H1O.BE.87.ANT70           ----........................--N---AK--F---A---------A--------P---A---Y--------........--V--M-----------T-.Y----Q-----INP.............................RG-GR---S----   98
CPZ.CM.98.CAM3            ----........................------P---Y---A-----------------A-------F--D------........-------------------.--L--Q-----IIP..............................G.-R--------   96
CPZ.CM.05.SIVcpzLB7       ---V........................----------Y---A--V-----Q--------P---Q---Y--D------........--E--Q------------.Y----Q-----I-P..............................---R-D--N---   97
CPZ.CM.05.SIVcpzMB66      ---L........................----------Y------I-----R--------D---Q------T------........-------------------.Y----H-----INI...........................NQ---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.05.SIVcpzEK505     ----........................----------Y---A----D-----L-----P-------YF-N-------........--E------T--------.Y----Q-----I-P..............................-.-R--------   96
CPZ.GA.88.GAB1            ----........................------P---YQ--A--T--------------P---Q---F--D------........--V----------H------.--L--Q-----ILP..............................---RS--SN---   97
CPZ.GA.88.GAB2            ----........................------AG--Y---A--I--D-----------P---Q---L--NN-----........--E----------H---V-.-----Q-----IE-...........................QR--R-D-SY---   98
CPZ.CM.01.SIVcpzCAM13     --H-........................------P---Y---A--I--D------K-Y--S---Q-------T------........--E-------------V-.--V--H-----IIP..............................---RA---N-P-   97
CPZ.US.85.CPZUS           ---V........................------A---Y---A-DV-----E--L-----P--Q----Y--S-----........--E----------------.Y----N-----ISL.......................ARRTP-.GR---S----  101
CPZ.CM.98.CAM5            ---I........................------P---Y---X--I-----T--L-----P---R---Y--------........--I----------H---V-.--L--Q-----ILX..............................E.-R---T--X-   96
CPZ.CD.90.ANT             ----........................-Q-E------M---L--T---I---------QPT-QH--NWV-AN-----........S-R--QVL-IL--KA--T-.Y-H--A-----HGP..........................................R-R-   89
CPZ.TZ.00.TAN1            ----........................-N-N------YT--L-DI---I-Q---K----PI-Q-V-NWVFTI----........S-E--QEL-K---RA--T-.Y-H--I-----S.................................................-   84
RCM.NG.x.NG411            -.......................EML---EE-----YD--LMDT-I-IQE--KK--TPEL-TQE-NY---QH--.........SLE--K-M-TL-NKA--L-.--H--EG----GA-...........................GG.NPLR.SILHSRNIL-  101
                          Vpr start                                                                                                                                        Vpr end
MAC.US.x.239              -........................EERP--NE------WD--VV-V-----E--LK--DPRL-TA--N---NRH--.........-LE-AGEL-----RA--M-.--G--I-----QPG..............................GG..NPLSAIPPSR-ML-  102
H2A.GW.x.ALI              -.....................TEAPTEFP-AGM--HQGARD--VI-V-R-I-E--L---DPRM-IA--GY--TRH--.........-LERAREL-NA--RA--M-.--A--GR--V-Q--..............................GR..NPLSAIPTPRNMQ-  106
H2A.DE.x.BEN              -.....................TEAPTEFP---GT-R-DLGSD-VI-T-R-I-E--L---DPRL-IA--YY-HNRH--.........-LE-AREL-KT--RA--V-.--A--NR----*--..............................RR..TPCPA-PTPRAMH-  105
H2A.SN.x.ST               -.....................TEAPTESP---RT-P---GD--VI-T-R-I-*--LK--DPRL-IT--NY--ARH--.........-LE-ARGL-----RA-LL-.--A--GR---QP-..............................GR..NPLSAIPTPRGMR-  105
H2B.CI.x.EHO              -.....................AEAVPEIP---KN-----WEQ-VVDV---I-Q--LK--DPRL-TA--NF--NRH-N........-LE-AGEL-KL--RA--L-.--G--Q-----QPG..............................GG..NPLSAIPPS*GMQ-  105
H2B.GH.86.D205            -.....................AEAAPEIP--NEN-----WE--IG-I---I-Q--LK--DPRL-TA--NF--SRH--.........-L--AGEL-K---RA--L-.--A--Q-----QSG..............................GG..NPLSTIPPP*GMR-  105
H2G.CI.92.ABT96           -........................AEEIP---EA-----WD--VV-V---I-E--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-K---RA--L-.--GD--C-----GNA..............................GG..NPLSTIPPSRGVF-  103
H2U.FR.96.12034           -.........................AEIP---GA-----GD--VR-I-----E-----DPRL-TA--NY--NRH--.........-LE-AGEL-K---RAI-L-.--G---------SS-..............................GG..NPLTTVPPS*GVL-  101
MAC.US.x.251_BK28         -........................EERP--NE------WD--VV-V-----E--LK--DPRL-TA--N---NRH--.........-LE-AGEL-----RA--M-.--G--N-----QPG..............................GG..NPLSTIPPS*GVL-  101
MAC.US.x.251_1A11         -........................EERP--NE--*---WG--VV-V-----EG-LK--DPRL-TA--N---NRH--.........-LE-AGEL-----RA--M-.--G--I-----QPG..............................GR..NPLSAIPPSR-ML-  101
SMM.US.x.H9               -........................TERP---EA-----WD--XV-V---X-E--LNX-DPRL-XA--NY--DRH--.........-LE-AGEL-----GA-----.--G--------QSG..............................GX..NPLSTIPPXRGVL-  102
SMM.US.x.PGM53            -........................TERP---EA-----WD--VVKV---V-E--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-----RA-----.--G--A-----QS-..............................GG..NPLSTIPPSRGVL-  102
SMM.US.x.SIVsmH635F_L3    -........................AERP---EA-----WD--VV-V---I-E--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-----RA-----.--S--A-----QS-..............................GG..NPLSTIPPSRAML-  102
MNE.US.x.MNE027           -........................EERP---E------WD--VV-V-----E--LK--DPRL-TA--NY--DRH--.........-LE-AGEL-K---RA--M-.--G--N-----QSG..............................GG..NPLSTIPPSR-ML-  102
STM.US.x.STM              -........................THRP---E------WD--VV-V---I-Q--L---DPRL-SA--NY--DRH--.........-LE-AGEL-K---RA--M-.--G--------QPG..............................GG..NPLATIPPTRGVL-  102
MND_1.GA.x.MNDGB1         -.....................GQKRDEQVS-----P---Y-Q-LADTM--I-E--RK---L-I-NAVSEYCVQNT-SEE........EAC-KF-TLMNRAIWV-.LAQ--DGTF.RER..............................RPQLPPS.GF-PRGD-L-  105
MND_2.CM.98.CM16          -..........................EQP---EA-----Y---LIDT-A-VQE--LK--D-RL--AV-SW--K-----.........-LE--RQL-S---KA-SM-.Y-Y---E----QEG............................GGC.YPLR.SFPRSDNPL-  102
MND_2.GA.x.M14            -..........................EQP---EA-----Y---LIDT-A-IQE--LK--D-RL--AV-SW-----------.........-LE--QKL-TL--RA--L-.--H---E----QEG..............................GKYNPLR.SFPRPNNPL-  102
GRV.ET.x.GRI_677          -ASGRDPRE..............PLPGWL-IWDLD---WD--LQDM-RD-NE--R---GMNM-IRVWNYCV-EGRR....HNTPWNEIGYKYY--V-KSM-V-.--C---RRGPFSPY.......................EERRNGQGGGA.....PPPPPGLA-  119
VER.KE.x.TYO1_patent      -ASGR..............DPREARPGEV-IWDLS---WD--LRDM-QD-NQ--RL--G-EL-FQVWNYCQ-EGER....HGTPMMERAYKYY-LV-KA--V-.--C---RRQPFEPY.......................EERRDGQGGGR....AN-VPPGLE-  120
VER.KE.x.AGM155           -ASGR..............DPREERPGGL-IWDLS---WD--LRDMV--IN---KL--G-EL-YQVWNYCQ-EGER....QGRPIAERAYKYY-LV-KA--V-.--C---RRQPFEPY.......................EERRNGQGGGR....PG-VPPGLD-  120
VER.KE.x.9063             -ASGR..............GPRENRPGEV-IWDLN---WD--LRDM--D-NQ--KL--G-EL-FQVWNYCQ-EGER....RGAPMTERAYRYY-LV-KA--V-.--C---RRQPFEPY.......................EERRDGQGGGR....AN-APPGLD-  120
SAB.SN.x.SAB1C            -ASGGWLPPVGGDPPKDPPKNPREEIPGWL-TWDLP---FD--LRDM-QD-NS--QC----NL-FR-WWN-V-EPAI...DHGQTRLE-WYKYC----KA--V-.MKGR-CKPKTHPAY...................GPGAGGPPPGLGGASGG..A-SAAPGL.-  141
TAN.UG.x.TAN1             -AEGR..............DSRERRPGWL-IWDLS---WD-*LRDMVA--NQ--Q---G-EL-FQVWNFCQ-EGER....NGAPMIERAYRYY-LV-KA--V-.--C---RRTPFEPY.......................EERRNGVGGGR...DGREPPPGLA-  120
VER.DE.x.AGM3             -ASGR..............DPREARPGEL-IWDLS---WD--LRDM--DINQ--KM--G-EL-FQVWNYCQ-EGER....NRTPMLERAYKYYKLV-KA--V-.--C---RRQPFEPY.......................EERRDGQGGGR....AG-VPPGLD-  120
SYK.KE.x.KE51             -A--F..........FRVYQRQPGEPLFYIPRNVNWD-AGVISATAKA-QVAQ--CK--TAEEIF-VWNQCL-VEAGPEETPTMAWIRCMLDMH-A-NFM--E-.-AA--PQRTKYARH.............................RGYPHPS..........-  116
LST.KE.x.lho7             -SRQR................QQRSPERP---F--P---WGQ-LADTM--I-E--R----EVM-AQVS-YCVDST-SEE........EACMKF-TLVNRA--L-.LP.T-PSYA.AAR........................AGTAHPTEAAPRPNEGDIQRGSD-  115
LST.CD.88.524             -SRQR................QQRSPERP---F--P---WAQ-LADTM--I-E--R----EAI-AQVSEYCVSST-SEE........EACMRF-TLVNRA-YL-..LPQ-PSYAAAR........................AGTAHPTEAAPRPTGRDIQRGHY-  115
WRC.CI.97.97CI14          -.....................SNS.REQAPSG--PPI-E...D-LGMQ--FAM-KQ----EV-VRVFLYCSQQREYEE.......VPTMHYFH-V--RG--M-.LTQ--PKQQEKREQ.............Q.......FQRAEQLYRNLERTAGDQGEVEQRDLISL-  117
WRC.GM.05.Pbt_05GM_X02    -.....................SNSGRER-PTG---AI-E...D-LGMQ--FAM-K-----EV-VRVFLYCSQQRNYEE.......Q-TMHYFH-V--RG-YL-.LIQ--PKEKENRE-.............D.......FQRAEQFYRNLERASGDQGGVEQRDLISL-  118
WRC.CI.98.98CI04          -.....................SNS.REQAPTG--PPT-E...D-LGMQ--FAM-KL----EV-VRVFLYCSQQRNYEE.......GPTMHYFH-V--RG--M-.LTQ--PKQQEKREQ.............Q.......FQRAEQFFRNLERAAGDQGEVEQRDLISL-  117
OLC.CI.97.97CI12          -........................AERV------P---LED-VRDTM--IIV--K-----EL-VRVFLFCAQRHINDE.......MSTSLLF--L--RA--Q-.Y-N--PGVELRRNP...........................................HR-SEGMQE-   97
COL.CM.x.CGU1             -..............................RVLKGTS#-LQDQV-NITR--R----LTHISVIMWGCTRRHWGR.......P-TAGL-YV--MECA-YL-VHAGQ-..................................RFPST-NP.........-   78
DEN.CD.x.CD1              -A..................................-TTQRQLQ-AVWDITE--RK--SKEEIT-IWD-CWSLPALPHWTEGQVMA-A-IDF---M-KEIWK-.Y-V--F-REAERV-...........HYPNIRPLRP.................RR...EEP.....-   99
DEB.CM.99.CM5             --RYPPSHP...................-HFTSRTVPMTRLALQQAMQD-NE--LK--T-EE-W-VWN-CVDLPAQPDWTGEQAWA-S-IDY-K-V-RM-WL-.L-EA-F-REREA--...........RYPNIRPLTG.................RN...-EVRDGE-  118
DEB.CM.99.CM40            --RYPPSHP...................-HFTSRTIPMTRLGLQQAMQD-NE--LK--T-EE-F-VWN-CVDLPAQPDWTGEQAWA-S-IDY-K-V-KM-WL-.L-EA-F-REREA--...........GYPNIRPLTG.................RN...-EVRDGE-  118
TAL.CM.00.266             --RL...................PPSHPLPLTSRLIPTPRPQLER-MR-VRD-MLI--T-DEVI-VWN-CV-LPA-PDWTGEQAWA-S-IDFA-KGN-M-LL-.-QQ--Y-EQQRR-P...........HYPNIRPLSG.................RG...ER.TMQ.-  116
GSN.CM.99.CN166           --RI...................PPSHPMPWLSRRVPTTMAIAQNA-W-INE--EK--S-DE-R-IWHDVT-LPA-PDWTVDQAAI-CAIDY--.V-T---R-.-KD--F-RVNLS.G...........RYPAIRPSRGTAPPDSNSVSHADPEQPRR..PSRYRMDE-  134
GSN.CM.99.CN71            --RT...................PPSHPLPWISRRVPTTMAVAQSAMW-INE--EK--S-EE-R-IWHDVT-LPA-PDWTVDQAAI-CAIDYV-RV-T---R-.--D--F-RYNRIV-...........RYPVIRPLRGTAPPDSNSVPHADPEQPRR..PSRYRMDE-  136
MUS_1.CM.01.CM1239        --RV...................PPSHR-PWHSRVVPTTMQQAQQAMWD-NE--EK--S-EE-R-IWNDVT-LPA-PNWTVDQAAI-CAIDY--RT-T---R-.Y-E--Y-RYSNTI-...........RYPNIRPLRGTQAPPSNSMPNADPTPPLR..PSRYRMDE-  136
MUS_1.CM.01.1085          --SL...................PPSHPAPWHSRVVPTTMRQAQ-AMW-VNE--EK--S-EE-R-IWNEVT-LPA-PDWTVDQAAI-CAIDYV-RV-T-V-R-.--D--Y-RYNR-M-...........RYPTIRPLRGTQPPPSNSTPNADPVPPLS..PSRYRMDE-  136
MUS_2.CM.01.CM2500        ---H...................PPSHPLPWLSR-VPTTMAQAQNAMWD-NE--EK--S-DE-R-IWDRVT-LPA-PNWTVDQAAI-CAIDY--WV-G---R-.--E--Y-RYSQ-I-...........RYPQVRPLRGTQPPDSNSTPNVDPFPPRR..PSRYRMDE-  136
MUS_2.CM.01.CM1246        ----...................PPSHPLSWLSR-TPTTMVQAQAA-W--NE--EK--S-DK-R-IWDRVT-LPA-PQWSVDQAAITCAIDY-KSV-T---R-.Y-D--Y--YAQTI-...........RYPPLRPLRGTQPPDSNSMPNADPTPSLR..PSRYRMDE-  136
MON.CM.99.L1              ---PPQSHP..............LHWTSRMV.....PIERQALQAAIW--NE--LK--S-EE-R-IWEQVT-LPA-PAWNADQAWA-CAIDYT-WV-TI-YR-.Y-E--Y-RYAEQI-...........RYPVLRPMRGTAPGPTSSVPQADPDNPRR..PSRYRMDE-  136
MON.NG.x.NG1              --N-PPSHP..............PHWTSRVV.....PIERAALQAAIWD-NE--LX--SKEE-R-IWDQVT-LPA-PGWNADQAWA-XAIDYT-RV-TXI-R-.M-D--Y-RHAR-X-...........RYPXLRPLRGTAPGPTSSVPHADPERPLH..PSRYHHDE-  136
TAL.CM.01.8023            --RL...................PPSHPLPITSRLIPTPRAQLER-MR-VRD-MLI--T-EEVI-VWN-CV-LPAEPDWTGEQAWA-S-IDFA-KGNEM--L-.-QQ--F-EQQRR-A...........HYPNIRPLSG.......................RRNTMQ-  115
RCM.GA.x.GAB1             -M................#PLFTEMELP---E------YD--LMDT-I--QE--KK--TYAL-TQI-DYV--QH--........SIE--Q-M--L--RA--L-.--N--AG----TS-..................................GS..NPLR.SIPQTRNIM-  108
SMM.SL.92.SL92B           -..........................EHA---ETNP---WD--IRDV-----E--LK--DPRL-TA--NYV-D----.........-IE-AGE--K---KA--L-.--H--T-----QP-..................................GP..NPL.GSIPSARDVL-  101
MND_2.x.x.5440            -..........................EQP---EA---G-Y-K-LVGT-A-IQE--LK--D-RL--AV-SWV--QQ--.........-LE--QQL-S---RA--L-.--Y---G----QEG..................................GRY.NPL.-SFPSPNNPL-  102
DRL.x.x.FAO               -..........................ERP---E--P---YD--LVDT-V-IQE--LK--D-RL--NV-SWV--R----.........SLE--HQL-IL--RA--L-.--Y--QE--L-.PA..................................GRW.NPFR.TFPRANQQL-  101
SUN.GA.98.L14             -AS...................RREPVEQP---E--P---FAQ-LADTM--IRD--E----AAIVAQVL-YCADST-SE........KE-CMRA-TL-NRA----.LP.M-PSYIRIRSG.......................SGNVNPRPPTRPRPGQGDIRRGLQD-  117
LST.CD.88.447             -SRTR................QQRSPEKP---F--P---WGQ-LADTM--I-E--R----ETI-AQVSEYCVSST-SEE........KACMKF-T-VNRA----.LPT-PSYAAAR.......................AGTA.HPTEAAPRPTK-DIQRGRY.-  115
LST.CD.88.485             -SR-R................QQRSPERP---F--P---WGQ-LADTM--I-E--R----EAI-AQVSEYCVSST-SEE........EACMRF-T-VNRA----.LPT-PSYAATR........................AGTAHPTEAAPKPTERDIQRG-Y-  115
SYK.KE.x.SYK173           -AE-F..........FNPSQHVQGTPWFFIPRNVELT-NVINVTVKA-LVVT--SK--TPQEIY-VWNQSLNEEAGTD.SPTMAWERTMLDMV-A-NLM--E-.-AA--PQRTRYARH.....................................-GYPHP--  114
```

```
                              Tat start                                                        disulfide bonding              NLS                exon 1 end exon 2 start
H1B.FR.83.HXB2            M.............E...PVDPRLEPWKHPGSQPK..........................TAC.TNCYCKKCCFHCQVCFITKALGIS.YGRKKRRQRR....RA.HQNSQTHQASLSKQPTS.QPRGDPTGPKE.................................  86
H1A1.UG.85.U455           -.............-......----N-----------T..........................---.S-----V--W---L--LK-G----.--K---KP--....GP.P-G-KD--TLIP---LPQSQ-.VSA-QE-SKKKVESKAKTDRFA................... 101
H1B.US.90.WEAU160         -.............-......---H-------------..........................---.-----R---------M--G---.----------------.-S.P-------D--------.----------SKKKVERETETDPED................... 101
H1C.ET.86.ETH2220         -.............-......----N----N------..........................---.NQ------SY--L---L#-G---.----------------.---.P-S-KD--NLI----L-.HT------SE-SKKKVESKAETDPYA................... 100
H1D.CD.84.84ZR085         -.............D......----NID--N------R..........................---.N--------Y---------G---.---------------.-P.PHS-----DPIP---S-.--------Q--........................................  86
H1F1.BE.93.VI850          -.............-......----S-D--N------T..........................-P-.-K----R------W--T--G---.----------H....-T.P-S--V--N--P---L-.-A---------SKKEVESKAKTDPCA................... 101
H1G.SE.93.SE6165          -.............D......----N----N-------..........................-P-.NK-F--V--W------LN-G---.-------KH--....GT.P-S-KG--DPVP---LP.TT--N------SKKEVASKAEADQCD................... 101
H1H.CF.90.056             -.............D......----K----N------Q..........................---.N--------Y---M--LK-G---.------S--H....-T.PASL-D--N-I----L-.RTH--------QKKEVASKTETDP.....................  99
H1J.SE.93.SE7887          -.............-......----NR---N-------..........................---.---------Y-----LQ-G---.---------------.S-.PPG-K---DLIP---L-.-TQRK----E-SKKEVESKAEPDRFD................... 101
H1K.CM.96.MP535           -.............D......----NI---NQ------..........................---.NQ---R--Y---I--LK-G---.N------P---....TT.PY--EN--DP-R---L-.-----EQ-D---SKKKVESKTKTDQFD................... 101
H101_AE.TH.90.CM240       -.............-...L---N----N------T..........................---.SK-------W---L--LK-G---.-------KH--....GT.P-S-KD--NPIP---LP.II-RN--D---SKKEVASKAETDQCD................... 101
H102_AG.NG.x.IBNG         -.............-...L---S----N------T..........................---.SK----M--#---L--LN-G---.--------R--....GT.P-SR-D--NPVP---LP.TT--N--D---SKKEVESKTKTDPCD................... 100
H103_AB.RU.97.KAL153_2    -.............-......-----------------..........................---.---------L------MK-G---.----------------.---.P-DN--D-V--P---A-.-----------........................................  86
H104_cpx.CY.94.CY032      -.............-......----D----N------T..........................-D-.NK-F-----W------LK-G---.-------KH--....GS.L-G-KG--NLIP---L-Q--N--S---E-QKKKVASKTEADPFA................... 102
H1N.CM.95.YBF30           -.............-......---------N-------..........................---.N----R--Y--LY--TK-G---.-------S---....-T.P-S-KS--DLIPE--L-.-QQ--Q--Q-KQKEALESKTEADPCD.................. 101
H1O.BE.87.ANT70           -.............D......----EVP--H------Q..........................IP-.N----R--Y--Y---VR-G---.-------GRP....A-..ASHPD-KDPVP--SPT.ITKRKQERQE-QEEEVEKKAGPGGYPRRKGSCHCCTRTSEQ... 114
H1O.CM.91.MVP5180         -.............D......----EMP--H----K-Q..........................-P-.N----R--Y--Y---TK-G----.H------RPA....A-..ASYPDNKDPVPE-SL-.HTGRKQKRQE-QEKKVEKETGPSGQPCHQDSCNSCTRISGQ... 115
CPZ.CD.90.ANT             -.............D......---AETP--L--PAT-A..........................-P-.N----C--Y--PL--TK-G---.----R-AR-...NR-TTAES-ENN-DPV--..........................................  73
CPZ.CM.01.SIVcpzCAM13     -.............-....-I--E--------------..........................---.N----A-----MY--TK-G---.--------.A--....-TAFAS-KNN-DPVR---L-.----NQERSEKQAEKVESQAAADP..................  99
CPZ.CM.05.SIVcpzEK505     -.............-......----E----N-------..........................---.N--H--V--Y--VY--TK-G---.-------S---....-T.P-SNKS--DP-P---L-.-RL--Q--Q--QKKTLES.......................  93
CPZ.CM.05.SIVcpzMB66      -.............-....-I--N----N-------..........................---.N----Q--Y---L--TK-G---.----R--K----....-T.SES--N--DPVP---L-.--G-IE--Q-KSKKEVESQTTSDQFA................ 101
CPZ.CM.98.CAM3            -.............-......----D----L------R..........................---.N----M-----I---EK-G---.--------G---...-LSAAS-THN-G--QQ--L-.RQQ-NQ--QE-QKKTLESQAKADPFA................ 102
CPZ.GA.88.GAB2            -.............D....-I--S----N-------..........................---.NS---R--Y--ML--TS-G---.--------.G--...-TALKG-KSN-DP-RQ--L-.-QQ--QKCQ--SEEEVESKTEADL..................  99
CPZ.CM.98.CAM5            -.............-......---XN----L-------..........................---.N----C----M---EK-G---.--------G---...HSLASS-THN-D-IRQ--L-.RQQ-NQ--EE-QKKTVESQAEADPRDWSAGS............. 107
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--SEKQKKEVERTSEADQSDW................ 101
CPZ.TZ.00.TAN1            -.............N....-I--QVA--E---AA-E..........................-P-.------------P---TK--------.---R-GRKS...AV..HST.NN-DPVRQ-SLP.KRSRIQNSQE-SQEEVEAETTSGGRPRQQDSSVSSGRTSGTSSS 117
CPZ.CM.05.SIVcpzLB7       -.............-....-I--S-D--N--------..........................-P-.N--F-----Y--PF--TK-G---.------W-S--....-T.SKS-PHY-DPI----FP.PS---SLS-E-QKKKVESKAQTDQ...................  99
CPZ.US.85.CPZUS           -.............D......----NI----Q------..........................---.N-----------V---TK-G---.-------GR--....-TTAKS-ADN-DP-R---F-.-QL-NQ--Q--QKKALEKQTTSDQRD................ 102
CPZ.GA.88.GAB1            -.............D....-I--D-----------R..........................-V-.N----A--Y--IY--TK-G---.-------TT--....-TAPAG-KNN-D-IP---L-.-S--NKE-SEKSTKEVASKTEADQ.................. 100
MUS_1.CM.01.1085          -.............D...-SVEE-PKEQR--AA-A..........................-P-.-S-F---A---VL--QK-------.---RF-KRS-......ADSSLPS-SDPVQIKSL-.DR-R-QKRQ--QKETVEAEAETDP.....................  98
MUS_1.CM.01.CM1239        -.............D...-SVEN-P-EQR--AA-T..........................-P-.SS----R-A---ML--QK-------.---RF-KRP-......SD--V-N-PNPVPIKSIP.NLQRQQSRQ--QKEALAAAAGPDPCDI.................. 101
MUS_2.CM.01.CM1246        -.............D...-SVEG-P-DQR--AA-P..........................-P-.------C-A---IL--QK-------.---RI-KRP-......TDSLLPSD-NPVRF-SIP.DSQQNKSNK-KQKKALEEAAGTDPCRCFQKTSTT.......... 108
MUS_2.CM.01.CM2500        -.............D...-EVEN-P--QR-S-A-T..........................-P-.-P----A-AY--MQ--L--------.---NRK-H--......TPEGLPNN-DPIQLKSIP.E--R-QER--KQKKEVAPEAVSDPCAI................ 101
                              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-.......TPKKAKANTS-A-NK-I-.NRTRHCQPE-AKKETVEKAVATAPGLGR................ 130
H2A.GW.x.ALI              -.............-TPLKEPGSSLMPYNEP-SCTSEQDVAVQELAKQGEEILSQLYRP..LET-NNT----E--Y---L--LN-G---W.-D--G--R-.......SPKKIKA-SS-A-DKSI-.TRTRNSQPEEKQKKTLETTLGTDCGPGRSHIYIS.......... 136
H2A.DE.x.BEN              -.............-TPLKAPESSLKPYNEP-SCTSERDVTAQELAKQGEELLAQLHRP..LEP-TNK---R-S----L--SK-G---.-E--G--R-.......TPRKTK-PSP-APDKSI-.TRT--SQPT--QKKTSEATVVTTCGLGQ................ 130
H2A.SN.x.ST               -.............-TPLKAPEGSLGSYNEP-SCTSEQDAAAQGLVSPGDEILYQLYQP..LE--DNK-------Y---M--LN-G---W.-E--G--R-.......TPKKTKA-SS-A-DKSI-.TRT-NSQPE-KQKKTLETALETIGGPGR................ 130
H2B.GH.86.D205            -.............-IPLQEQESSLKSSNEP-SSTSEPVVNTQGLDNQGEEILSQLYRP..LK--DNT-------Y---L--LK-G---C.-D-SR--.........SAKRAK-TAP-APDKSL-.ART--SQPT-KQKKEVETTGTTDPGPGR................ 128
H2B.CI.x.EHO              -.............-IPLKEQESSLNSSSGH-SSTSEGVANTQGLDNRGEEILSQLYRP..LK--SNT------SY---L--LK-G---C.-E-SRK-.........SSKRAK-TTS-APNESL-.ART--SQPT-KQKKEVETTRATDLGPGRSNTSTSRFAN...... 138
H2G.CI.92.ABT96           -.............-TPSREQENSLKSCREL-SCTFEETVDAPGLEGTQAEILYQLYRP..LE--SNK-------Y---L--LN-G--VC.-EQPR--R.......TPKKTKANSF-A-NESI-.ARTRNXQPK-K........................................ 112
H2U.FR.96.12034           -.............-IPLKEQESSLKSCKEQ-SSTSEGDVGTQGSAAAGEEILSQLYRP..LE--SNK-------Y---L--LK-G---C.-EQPR--.........SPKKVKAYSSAP-NKSL-.NGTRNSQPKEK........................................ 111
MAC.US.x.251_1A11         -.............-TPLREQENSLESSSGH-SCTSEAAASTPELANLGEEILSQLYRP..LE--YNT-------Y---F--LK-G---C.-EQSRK-R-.......TPKKAKANTS-A-NK-I-.NRTRHCQPE-AKKETVEKAVATAPGLGR................ 130
MAC.US.x.251_BK28         -.............-TPLREQENSLESSNER-SCILEADATTPESANLGEEILSQLYRP..LE--YNT-------Y---F--LK-G---C.-EQSRK-R-.......TPKKAKANTS-A-NKLIP.NRTRHCQPE-AKKETVEKAVATAPGLGR................ 130
MAC.US.x.EMBL_3           .............................#SNER-SCILEADATTPESANLGEEILSQLYRP..LE--YNT-------Y---F--LK-G---C.-EQSRK-R-.......TPKKAKANTS-A-NKLIP.NRTRHCQPE-AKKETVEKAVATAPGLGR................ 116
SMM.US.x.H9               -.............-TPLKEQESSLESSREH-SSISEVDADTPESASLEXEILSQLYRX..VE--YNKX------Y---H--XK-G---C.-EQQR--.........TPKKTKANTF-A-DKSLX.XRARNCQPK--KKETVEXEVATDLGLXR................ 128
SMM.US.x.PGM53            -.............-TPLKEQENSLGSSNER-SYTSEEDVLTPESAKVEEEILCQLYRP..LE--SNT-------Y---L--LK-G---C.-EQQR--R-.......TPKKAKANTF-APNESL-.TRARNSQPK--KKETVETEVATDLGLGR................ 130
SMM.US.x.SIVsmH635F_L3    -.............-TPLKEQENSLESYREH-SSISEVDVPTPESANLEEEILSQLYRP..LEP-YNK----R--Y---H--LK-G---C.-EQHR--.........TPKKTK-NPLPA-NKSL-.TRTRNRQPK--KKEKVETEVAADLGLGR................ 128
SMM.SL.92.SL92B           T.............GIQ*KGQGK*LRYYKRHCSYTLGMDAPIQGLDSQGDQIPWDLYRP..LET-YNS-F---------L--LR-G---T.-AKPR--.........VKKKAKANPFDT-N-SI-.SGTRNCQPK--QKEKVETEVDTDCGLGR................ 126
STM.US.x.STM              -.............-TPLKEQESSLRSSSEP-SCTSEAVAATPGLANQEEEILWQLYRP..LEE-CNK-F-----Y---L--V--G---T.-E-SR--.........VKKKAK-YPI-A-NKSL-.TRARNSQPK--QKKEVETEVESTPGLGK................ 128
MNE.US.x.MNE027           -.............-TPLREQENSLKSSNGR-SCTSEAAATTLESANLEEEILSQLYRP..LE--YNT-------Y---F--LK-G---C.-EQSR--R-.......TPKKVKANTS-A-NK-I-.DRTKHCQPK--QKETVETAVATVPGLGR................ 130
GRV.ET.x.GRI_677          -.............D..K.EEE....PHPLLQDL.....................HRP..LQP-TNK-------Y--EL--LQ-G--VR.-HVSRK-R.........KTST-DN-DPIRQ-SI-TV.QRNGQTTE-GKTEVEKAAAAN........................  94
VER.KE.x.TYO1_patent      -.............D..KGEAEQIVSHQDLSEDY.....................QKP..LQT-KNK-F-----Y---L--LQ-G--VT.-HAPRT-R.........KKIASLDL-P-QH-SI-TKWGR-GQTTPTSQEKVETTAGSN..................... 100
VER.KE.x.9063             -.............D..KGEEE-TVLHQDLIR-Y.....................KKP..L-T-RNK-F-----Y---L--LQ-G--VT.-HAPRT-R.........KKSV-PNRL-QQD-SI-TR.GR-GQATQ-SQKKVERETTTAQILGRKDLERDKREAVGANA. 119
VER.KE.x.AGM155           -.............D..KGEEDQDVSHQDLIK-Y.....................RKP..LET-TNK-F-----Y---F--LR-G---T.-HAFRT-R.........KKIASADRIPVPQ-SI-IR.GR-SQTTQ-SQKKVEEQAKANLRISRKNLGDETRGPVGAGN. 119
VER.DE.x.AGM3             -.............D..KGE-EQGAYHQDLIE-L.....................KAP..LKR-TNK----C--Y---L--LQ-G--VT.-HAPRI-R.........KKIAPLDRFPEQ--SI-TR.GR-SQTTQKGQEKVETSARTAPSLGRKNLAQQSGRATGASD. 119
TAN.UG.x.TAN1             -.............-..S...EGDGMAESLLQDL.....................HRP..L-P-TNK-F--R--Y------LQ-G---T.-HVSRI-RP........KK-HSN--NLV-Q-SI-AW.G-NSQTTQ-EKTKIPAAAETSRRPQ................ 101
SAB.SN.x.SAB1C            -.............D.QEQEARPQVWEEL.QEEL.....................HRP..LQ--DNT-F--V-----IL--HK-----R.-YVPRP-R........-SKK..IS-NQVSLHN..........................................  73
OLC.CI.97.97CI12          -.............D.....AQEWSLEETLTEDL-...........................E-KNK-------Y---L--LQ-G----.-STESP-K--.....KKVVSH-QDSRDT-HFQIP.SRTRNSQTTEKEKTASSSRAPSTSGTVPQSVQGH.......... 107
DRL.x.x.FAO               -.............DARKVDLDQQDAG.THFEP.FQ....................GRT...NN-.NT-W--Y-AY---L--LQ-G---R.-NVSRP-K-T...ATVVSRVSETGDTPVARTHTA-R-PTT-DSQTEKKEKSKTEASGTSDQNPPSQDPPKLGAKKQ.. 124
RCM.NG.x.NG411            -.............D.........................VKGAGLEEREEEILLDPFFTRE-S-.NS---R-VY---L--LQ-----N.-ASRR--RG-P...GNKKKN-INPFDT-E-SL-.AIGRNSQSK--KKEAVEKAVDTNITTGREDIFIS.......... 115
RCM.GA.x.GAB1             -.............D.........................VQGVGLEHPEEVILYDPFRKRE-S-.NT-------Y---L--LQ-G---N.-ASRA--R-.......SKEENKADKFPVPN-SI-.TT--NRKLQEKKEKTVEKKVATSTTIG................ 105
MND_1.GA.x.MNDGB1         -.............-.......-SGKEDHNCPP-DSGQEEIDYKQLLEEYYQPLQ..........--ENK-W-------ML--QK-G---R.-HVYRK-VP.....GTNKKIPGSGEEAIRR-GYR.PVFSQDCQQNLYRKRPDNGEEESNSLELGEHNSENLKEESLNRS 135
MND_2.x.x.5440            -.............D.....................VGEVASDKKEEDITHFDPFRART...-P-.NK-F--V--Y---L--LQ----VH.-HVYRQ-RP-R...GLLEKVPSNSQAALGRDIALFLMSSRETISGPTTSNGQTEKEEKSQT................ 110
MND_2.CM.98.CM16          -.............D.....................AGKAVSDKKEGDVTPYDPFRDRT...-P-.DT----R--Y---L--LQ----VH.-HAYRN-RS-Q...RLLEKISEDS-APVGRVIALFLMSSRETISVPTTSNSQAKKEKKSPTQKNRVPNQTAPK..... 122
MND_2.GA.x.M14            -.............D.....................AGKAELDKKEANITPLDPFQGRT...-P-.NK-------Y---C--LQ-----H.-HVYRT-RT-K...RLLGKISQDSEAAVGRVIALFLMS..........................................  87
GSN.CM.99.CN166           -.............PILSSLGDPQDIAWMNMNRDPEECLLPNWQQPGAAPA............-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
DEB.CM.99.CM40            -.............-.........................SIDPFEERAN............-PH.PA-M--A-IL---L---Q-G-----.--PRR-KRKCRREYTPAAESN-PD-DFVPE--LP.-R-RKCFDSQKKEAEVEETAGPGGQHCREDSSVSSGRISNNC.. 115
DEB.CM.99.CM5             -.............-.........................EIDPFKETAN............-PH.P--M--A-TM---L--MQ-G---Y.--SRR-KR--..GSSFAAN.DKD--EFIPE--LP.-R-WKCFNSE-KETKVEKASRAGGRHRPEDS............. 101
DEN.CD.x.CD1              -.............N.......................ADSIDPFAG.NK............-P-.NT----H-AY---L--LQ-G---V.---AP-KRP-......AR-SGTPD-EAV---SL-.D--RQQISQ-KEKEKVETATGSSGGNIREDS.............  99
TAL.CM.01.8023            -.............S....................SKEELPTTPTYDPFPAGG..........-P-.NK-------Y---L--LQ-G---R.-VPDR-IRK-RR..STSSSN-KDKDPV----L-.RAT-NQECQ-KKVQKVESPTGAGGEHRLKDSLPLSGRTSTASSA 122
TAL.CM.00.266             -.............S....................SKEELRTTPISDPFQEEGR.........GP-.NK-------Y---L--LQ-G---R.-VPDR-FRK-RR..STPSSN-EDKDPV----L-.RAT-IQKRQE-TLQKVESKAATSGTDRKQDS............. 110
MON.CM.99.L1              -............M-.....................PVDPDLPKEQHPPATPR..........-P-.---F-RV------I--L--G---------RK-R-....ATSPVPGLSSSKNPAR--SL-.H-TR-SQR-T-QAQAVATAATPDRQHI................. 104
MON.NG.x.NG1              -.................QTQRGLCTRPDTTTMNSAXIVNPALPPEQHPSATPL..........-P-.N--F--I-AL-X----LQ-G--------GRK-R-G...P-HPTIGDRSY-DNPGVKSLP.KSL-NQERQ-KQEKAVESKAGTGXQSMREDF.............. 126
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-KKQKKTTVL...................... 132
WRC.GM.05.Pbt_05GM_X02    -KNRPLCTIFTGSFKEDYIYTSYKDVP-KKRTEKEIFKEQNNFIEIWKGLQEIREELSRETLSP-NHQ-R----I-S--L--LQ-G----.-HSYSKKTNK...........DKSDASTSRVI-.................................... 112
SUN.GA.98.L14             -.....................STQGHQQD.........QDQGKGTLEEAYKTNLE..........--DNK-W-RR------L--LQ-G---H.-YVYRG-........QHQSI-KDK-EAVRRIGSRKPAPNPQ-KK-KKRQR....KKESTST.................. 102
LST.CD.88.524             -.....................QQPEQEQHT...-Q-QHLDQLEEIYKEAITDPLQ.........S-QNK-H--V-----IL--QQ-----R.-YVHRT-R.......RVTYILE-E-DPVDRTSSGRTE-AKNSQT-KERQRKT..EKTSTST.................. 111
LST.CD.88.485             -.....................QQQEQEQLT...RQ-QHQNPLKEIYKEAVTKPLS.........V-DNK-C--R-----IL--QQ-G---R.-YVHR--R.......RVT-VFEDK-DPVDGTTSGRKE-TQNSQT-KER............................ 100
LST.CD.88.447             -.....................QQQEQEQLT...RQ-QHQDPLKETYKEAVTKPLS.........--NNK-C--------IL--QQ-G---R.-YVHR--K.......RVA-VFEDK-DPVDGTASGRKE-TKNSQT-KER............................ 100
LST.KE.x.lho7             -.....................QQPEQEQPI...-Q-QHQDLMKETYNEAVTKALQ.........--QNT-C--------IL--QQ-G--LH.-YVHRK-R.......RVTSILE-Q-DPVDRTTSGRKK-TKNSQTEKKRKEQI..QRLSTST.................. 111
SYK.KE.x.SYK173           -....................................S.STDQICQTQRVPPSFLEGTFLEKGPP-P-.NK-F--N--Y---L--LQ-G---T.-A-PRK-AA-.....SISEDDSAPTGT-PRAGRTQANPQTKKKAV-TEKDSTSLPSAENL.................. 110
SYK.KE.x.KE51             -....................................S.TEDKVRETQGIPTSFLEGTFLSNGPQ-P-.SR-------Y--YR--LQ-G---T.-V-ERK-RHAR....TTAEDFAA-SSTT....R-RS-LPST-R-QQGQKKE*............................  98
COL.CM.x.CGU1             -.....................SLS.......................LYAKPLE.........--NDK-W--A------L--LS-G---R.-....-KNK..........SAT..QTQ-VA-TGSG....NSQ-TQ-KYTES....KKTAAADFGD*..............  79
```

H1B.FR.83.HXB2	.....*	86
H1A1.UG.85.U455	.....-	102
H1B.US.90.WEAU160	.....-	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.SE7887	.....-	102
H1K.CM.96.MP535	.....-	102
H101_AE.TH.90.CM240	.....-	102
H102_AG.NG.x.IBNG	.....-	101
H103_AB.RU.97.KAL153_2	.....-	87
H104_cpx.CY.94.CY032	.....-	103
H1N.CM.95.YBF30	.....-	102
H1O.BE.87.ANT70	.....-	115
H1O.CM.91.MVP5180	.....-	116
CPZ.CD.90.ANT	.....-	74
CPZ.CM.01.SIVcpzCAM13	.....-	100
CPZ.CM.05.SIVcpzEK505	.....-	94
CPZ.CM.05.SIVcpzMB66	.....-	102
CPZ.CM.98.CAM3	.....-	103
CPZ.GA.88.GAB2	.....-	100
CPZ.CM.98.CAM5	.....-	108
CPZ.CM.05.SIVcpzMT145	.....-	102
CPZ.TZ.00.TAN1	GYTRPFKTSSGSSGSACKH-	137
CPZ.CM.05.SIVcpzLB7	.....-	100
CPZ.US.85.CPZUS	.....-	103
CPZ.GA.88.GAB1	.....-	101
MUS_1.CM.01.1085	.....-	99
MUS_1.CM.01.CM1239	.....-	102
MUS_2.CM.01.CM1246	.....-	109
MUS_2.CM.01.CM2500	.....-	102
MAC.US.x.239	.....-	131
H2A.GW.x.ALI	.....-	137
H2A.DE.x.BEN	.....-	131
H2A.SN.x.ST	.....-	131
H2B.GH.86.D205	.....-	129
H2B.CI.x.EHO	.....-	139
H2G.CI.92.ABT96	.....-	113
H2U.FR.96.12034	.....-	112
MAC.US.x.251_1A11	.....-	131
MAC.US.x.251_BK28	.....-	131
MAC.US.x.EMBL_3	.....-	117
SMM.US.x.H9	.....-	129
SMM.US.x.PGM53	.....-	131
SMM.US.x.SIVsmH635F_L3	.....-	129
SMM.SL.92.SL92B	.....-	127
STM.US.x.STM	.....-	129
MNE.US.x.MNE027	.....-	131
GRV.ET.x.GRI_677	.....-	95
VER.KE.x.TYO1_patent	.....-	101
VER.KE.x.9063	.....-	120
VER.KE.x.AGM155	.....-	120
VER.DE.x.AGM3	.....-	120
TAN.UG.x.TAN1	.....-	102
SAB.SN.x.SAB1C	.....-	74
OLC.CI.97.97CI12	.....-	108
DRL.x.x.FAO	.....-	125
RCM.NG.x.NG411	.....-	116
RCM.GA.x.GAB1	.....-	106
MND_1.GA.x.MNDGB1	LIEDLTSFARE.....-	147
MND_2.x.x.5440	.....-	111
MND_2.CM.98.CM16	.....-	123
MND_2.GA.x.M14	.....-	88
GSN.CM.99.CN166	.....-	93
GSN.CM.99.CN71	.....-	120
DEB.CM.99.CM40	.....-	116
DEB.CM.99.CM5	.....-	102
DEN.CD.x.CD1	.....-	100
TAL.CM.01.8023	TVTRP.....-	128
TAL.CM.00.266	.....-	111
MON.CM.99.L1	.....-	105
MON.NG.x.NG1	.....-	127
WRC.CI.97.97CI14	.....-	126
WRC.CI.98.98CI04	.....-	133
WRC.GM.05.Pbt_05GM_X02	.....-	113
SUN.GA.98.L14	.....-	103
LST.CD.88.524	.....-	112
LST.CD.88.485	.....-	101
LST.CD.88.447	.....-	101
LST.KE.x.lho7	.....-	112
SYK.KE.x.SYK173	.....-	111
SYK.KE.x.KE51	.....	98
COL.CM.x.CGU1	.....	79

```
                              Rev start                              exon 1 end exon 2 start           NLS                                 Leu-rich effector domain                                      Rev end
H1B.FR.83.HXB2                M.AGRS.......GDSDE...ELIRTVRLIKLLYQSNPPP...........NPEGTRQARRNRRRRWRERQRQIHSISERIL.GTYLGRSAE.......PVPLQLPPLERLTLDCNEDCGTSGTQ..................GVGSPQILVESPTVLESGTKE*.....   116
H1A1.UG.85.U455               -.-R-........-NP--...D-LKA--I---------C-...........--R-S---K-------A----D-L-----.SDC---P--.......---------I---R---S-S-------QPQGTE...........T---G---S---SA--G----N-......   124
H1B.US.90.WEAU160             -.----.........-------...D-LK------I--------...........S--------------------RK--GW--.N-----PT-.......----P---D-----K---------..................----S---L---A---P---------.....   117
H1C.ET.86.ETH2220             -.----.........-------...--LKA--I--I------Y-...........T---------------A------TL-----.SNF---P--.......---------------N---S--S------QSQGTT...........E---N-*.......................   107
H1D.CD.84.84ZR085             -.----.........-------...D-LTA-----I---------...........S---------------A---Y----G----.S------E-.......--------------IN-N-S---------..................-------S----AI---R-E--.....   117
H1F1.BE.93.VI850              -.----.........----T...--LKA-KC--I------Y-...........K---------------A-----RAL-D---.SSC----E-.......--------------HIN-S----QGPEE..................----S--SG--HA----------.....   117
H1G.SE.93.SE6165              -.----.........-ST--...--L-A-KA--I------Y-...........P---------------A-----SA------.TA----P--.......--------------H---S--S------QPQGTE...........T---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---------.....   117
H1J.SE.93.SE7887              -.----.........------D...Q-LLA-----I------Y-...........K-N-S-----------A--N--D-----P.SSC---P--.......---------I---R---S----N----..................----D---SG-PCM--GA---------.....   117
H1K.CM.96.MP535               -.----R.........--P--...Q-LT---T--I--E---Y-...........-L--S--T--------A--K--S-----L-.SAC--------.......---------I-K-N-N-D--P-K.--EG..................-L-----S--PC-----------.....   117
H101_AE.TH.90.CM240           -.----.........-ST--...--L-A--I--I------F-...........SS----T---------A-----SA------.S-C----T-.......---------V---N---S--G------QSQGTE...........T---R---SG--SVI-GP---N-......   124
H102_AG.NG.x.IBNG             -.----.........--A--...--L-A--I--I------Y-...........P------T-K-------A-----RA------.S-C--------.......---------I---N---S---------LSQGTE...........T-------S---YII-G----------.....   124
H103_AB.RU.97.KAL153_2        -.----.........-------...--LK-I---F--------...........S--------------------H------Q--.S-----PE-.......--L-H----------S---------..................------------------D---------.....   117
H104_cpx.CY.94.CY032          -.----.........-NI--...D-FKAA-A--I------Y-N...........--T-------------A--K----L-----.A-----P--.......-----------K---N-S-------DK..................------VS--L-A--GT-A-------.....   118
H1N.CM.95.YBF30               -.----.........-VN--...--L-A--V--I------Y-...........-SK-------------A-----RA------.SSC--GPP-.......--D-P----D-----TE--S--P--E..................-Q-.........GTA-T--......   105
H1O.BE.87.ANT70               -.----E........D-Q....-LQAIQI--I------Q-...........S-R-S-N--K-------R--A-VDTLAA-V-.A-VVHGPQN.......NNIVD-----Q-SIR...-PEGD........Q...........LSEAWTVDPRAEDN*.............   102
H1O.CM.91.MVP5180             -.----E........E-QQ...-LQAIQI--I------C-...........T-A-S-N--K-------R--A-VD-LAT---.A-VVHG-QD.......NNLVD-----Q-NIR...-PEAD........R...........LPGTGTVDPGTKDN*.............   103
CPZ.CD.90.ANT                 -.----EEL.....EGT--Q...-LKA-KI--I------Y-...........K-A-S-A----K----KK--D-VEGLAA-V-.R-LV-GPD-.......HN-VD--D-QN-S-EPLGSPP-PLQNYFQTVQ...........PSGEQS-HPLDNNQN*............   118
CPZ.CM.01.SIVcpzCAM13         -.----E.......HD-A...R-LQA-KI--I--E---Y-...........S-----N-----------A-----RE--N-V-.AAH---PPQ.......-GH-E--E-DK-S-H-L-TV-KV--TVNQHPQ...........TDT-ETDSTEGDNPI-GKRI-N-......   124
CPZ.CM.05.SIVcpzEK505         -.----.........-VN--...--L-AI-V--I------Y-...........-DS------K-------A-----RAL-----.SSCV-GLQ-.......-ST-P----D--S-NPE--L---E-E...................H--..........GTA-T*......   103
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--DPPQGTE...........T-T---NTPEGHS-I-GT-A-N-......   124
CPZ.CM.98.CAM3                -.----E.......--D-A...C-LQA--I--V--S---Y-...........DNK-------------A--------L-----.SS-----P-.......-PA-L---I----I-TG--I-KP--EPQQGTA...........TAEELQ--P-G-DNL-GK-ATN-......   125
CPZ.GA.88.GAB2                -.----.........--E--...R-L-A-KA--I--D---Y-...........SNR---S---------A-----CE--Q--F.SSL---HP-.......-DD-P-------S-GPL-PA--I--TGPERTE...........TTQ-ASEPAERDR-I-AAR--N-......   124
CPZ.CM.98.CAM5                -.----E.......--E-T...R-LQA--I--I--D---Y-...........DNK-------------A------E-GQ-V-.SSC----S-.......-PD-L---ID----NPV--SAN---EHQQGTA...........TVE-L*.......................   107
CPZ.CM.05.SIVcpzMT145         -.----E.......--E....N-L-AI-I-RI--E---Y-...........SQS---R-----K---------NQ-G--V-.ASC----S-.......-PQ-P---I---CISDL-TT-DT--IQPQGST...........TNRY--TPTERD-A--DNRNQ--......   123
CPZ.TZ.00.TAN1                -.----EE..........-AN...-LY---I--I--D---Y-...........SGA-S-T----------Q--H-VDALAS---.QYR--GPQ-.......-PH-DI-D-SK-H--PLDQPAS.........T...........ETGDNQLGTQP-NSA*............   105
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--IHSEGTE...........T-----T-SAGNS---GK-A-N-......   124
CPZ.US.85.CPZUS               -.----G.......--A--...Q-L-A--I--I--A---F-...........-NS------------KN-----NE--Q---.S-C----S-.......-PD-P---IGG--INSE--STNP--EHQQGTA...........TVE*........................   105
CPZ.GA.88.GAB1                -.----E.....PQD-A...R-LQA-KI--I------Y-...........S-----K---------A--K--SE--G-V-.A-----PPK.......-GD-E--E-DK-S-Q-V-TTQDV--SNTSQPQ...........TAT-ETVPAGGNYSI-GK-A-N-......   125
MUS_1.CM.01.1085              -.-...........-V-ENDR.----AC-A-QI--K---Y-...........TAD---S---------KRK----RE------.LSL---PPQ.......-SD-E--D-NK-S-H.......PLVATSES.............SPPDTEGVNKIVVHAQGTD*.......   110
MUS_1.CM.01.CM1239            -.-...........-A-ENVQ.D----C-I-RI------Y-...........TSR-S-A---------QQ--D--RA------.VAV---PPQ.......-SD-V--DISQ-SIN.......PSLDTPEG.............SSPDTEVVSTTVDRD*..........   106
MUS_2.CM.01.CM1246            -.-...........-A-GNDQ.----FC-Q-RI--D---Y-...........TVSR--AT-K--K---KK--.......AQ--.VAVFR-PPQ.......-ND-D--D-DQ-R-E.......PLVVDSKN.............LPPPDKGASK-VDSA*...........    99
MUS_2.CM.01.CM2500            -.-...........-TGRDIE.--LKACQT--I--N---Y-...........S-G--KN-----------Q--C--RAL-----.VS----PPQ.......-DSVE--D--K-S-K.......PLVADPAD.............SPSP-EVEHL-VAQG--NQHPPSGSP*   117
                              Rev start                              exon 1 end exon 2 start                                                                                              Rev end

MAC.US.x.239                  -.SNHE.........RE-...--RKRL---H--H-T--Y-...........TGP--ANQ--Q-K----R-WQ-LLALAD--Y.SFPD.PPTDT......-LD-AIQQ-QN-AIESIP-PP...-NTPEALC...............D-TEDSR--QD*............   107
H2A.GW.x.ALI                  -.TE-A.........GE-...D-Q-KL---R--H-T--Y-...........QGP--ANQ---------Q-WG--VALAD--F.TFPD.PP-SS......-LDRAVQH-QG--IQDLP-PP...-DLPESSE...............-ADNNQGLAET*............   107
H2A.DE.x.BEN                  -.SE-A.........DE-...G-QGKL--LR--H-T--Y-...........QGP--ASQ-------R-RQWLRLVALANKLC.AVPD.PPTDS......-LDRAIQH-Q---IQELP-PP...-DLPES..................NSNQGLAET*............   103
H2A.SN.x.ST                   -.NE-A.........EE-....--R-KL---R--H-T--Y-...........QGP--ASQ--------KQ-W--LVALADK-Y.TFPD.PP-DS......-LEQTIQH-QG--IQELP-PP...-NLPESSE...............-IDSSQRLAEI*............   107
H2B.GH.86.D205                -.TA-...........EG...D-QKEL--LH--H-T--Y-...........QGP--ASQ--------KR-GL--LALAD--H.SFST.AP--E......-LN-AIQR-QN--VEDLPNPPPNLN-SPTTQAPGCVPPVWDQLVPRSA-SGSKGYGRNSCECRRDLMGGSQ   136
H2B.CI.x.EHO                  -.NA-...........ER...D-QKGL--LH--H-T--Y-...........QGP--ASQ--------KQ-GL--LALAD--H.PLPD.SPT-G......-LD-AIQR-QN-IIKDLPNPP...-STPTAQASTCIPPIWDQLVPRSN-SSSQGCGRDSCERGEDLVGSPQ   133
H2G.CI.92.ABT96               -.SNL..........EE-....--RKRL---HF-H-T--Y-...........QGP-XANQ------K-KQ-WL--LALAD--Y.SFPX.PP-RE......-LD-AIQQ-QS--IQDXP-SP...-AVPKNLQ...............D-PSN*...............   100
H2U.FR.96.12034               -.SNP...........E-...GVQ-RL---H--H-T--Y-...........TGP--ANQ---------Q-WL--LALADK-Y.TFPD.PPTDS......--D-AIQQ-QG-AIQDLPEPP...AVVPEHQR..................TSAP*...............    98
MAC.US.x.251_1A11             -.SSHE.........RE-....--RKRL---H--H-T--Y-...........TGP--ANQ--Q------R-WQ-LLALAD--Y.SFPD.PPTDT......-LD-AIQQ-QN-AIESIP-PP...-NTPEALC...............D-TKDSR--QD*............   107
MAC.US.x.251_BK28             -.SSHE.........RE-....--RKRL---H--H-T-SY-...........TGP--ANQ--Q------R-WQ-LLALAD--Y.SFPD.PPTDT......-LD-AIQQ-QN-AIESIP-PP...-NTPEALC...............D-TKGSR--QD*............   107
MAC.US.x.EMBL_3               -.SSHE.........RE-....--RKRL---H--H-T-SY-...........TGP--ANQ--Q------R-WQ-LLALAD--Y.SFPD.PPTDT......-LD-AIQQ-QN-AIESIP-PP...-NTPEALC...............D-TKDSR--QD*............   107
SMM.US.x.H9                   -.SSN...........E-....--R-RL---HF-H-T--YX...........XGP--ANQ--R-----XQ-WQ--LAXAD--Y.SFPD.PP-DT......-LD-AIQQ-QG-AIEXLPNPP...ASAPEPLK...............DAAESP*..............   100
SMM.US.x.PGM53                -.SSSA.........GE-....--R-RL---HF-H-T--Y-...........QGP--ANQ--R------Q-WQ--LALAD--Y.SFPD.PP-DS......-LD-AVQQ-QD-S-ENIP-PP...-SVPEALT...............E...NPR-RQD*............   104
SMM.US.x.SIVsmH635F_L3        -.SST...........E-....--RKRL---HF-H-T--Y-...........QGP--ANQ--R------Q-WQ--LALAD--Y.SFPN.PPTDT......-LD-AIQQ-QG-AIEDLP-PP...-SAPETLK...............DAAKSS*..............   100
SMM.SL.92.SL92B               -.QNP...........E-....--R-RL---H-IH-T--Y-...........AGP--ASQ--------KR-WI--VALAD--Y.SFPD.PP-DT.....DID-AIQQFDS-SIQSLPEPP...-TVPETLR...............DQSAD*...............    99
STM.US.x.STM                  -.SDQ..........EE-.....-RKRL---QF-H-T--Y-...........QGP--ANQ---------Q-WN-LLALAN--Y.SFPD.PP-ST......-LD-AVQQ-QG-SIQDLP-PP...PNLPKDLQ...............DTAEN*...............    99
MNE.US.x.MNE027               -.SSHA.........EE-....--R-RL---H--H-T--Y-...........TGPS-ANQ---------Q-WQ-FLALAD-VY.SFPD.PPTDT......-LD-AIQQ-QN-AIESIP-PP...-NIPEILH...............D-TESPR--QD*............   107
GRV.ET.x.GRI_677              -..SLG.........KE-K.....QALKI--T--G---Y-Q...........FS--ARQ--RA-Q---KQ-Q--DK-AG-V-NTFED.QQLVA.QLQELQLENKDLV-QH-PDPPHIHQDS--IPAVWAPATPRGSNRACSSS-E-CEGS-GQTGCYCPIRLSGSHQQSK   141
VER.KE.x.TYO1_patent          -..PLG.........PE-R.....-LL---W---ST--Y-P...........SG---ARQ--RA-----QQ-D--RVLV--LQ.............EQVYA-DRLADEAQH-AIQQLP-PPH-A*..................................    84
VER.KE.x.9063                 -..PLG.........PE-R....KVFS--AY-NRT--Y-P...........V---ARQ--RA----------LRFLA---W.NE..TKE....RQLEQTLDQLALDVQH-AIQHLP-PPN-S*..................................    89
VER.KE.x.AGM155               -..PLG.........PE-R.....-LL---AF--R---Y-S...........V---ARQ--RA----KN--K--YALA---W.---.RQE....DQLVQAIDQLVLDTQH-VTQQLP-PPSQA*..................................    89
VER.DE.x.AGM3                 -..PLG.........-E-R.....-LL---AF-NKN--Y-P...........V---ARQ--RA-----QA-E-LRALA---W.HS..RVE....EQLVQAIDQLVLDQQH-AIQQLP-PPS-S*..................................    89
TAN.UG.x.TAN1                 -..SLG.........-EDLR....-IIQI-RI--H--QY-P...........G---ARQ--R--Q-FQQQ---VAAL----F.IAQQR-DPSGGESLAAAFDQLVLDNQQ-VIETLP-PPQEPHDSSSTA*..........................   103
SAB.SN.x.SAB1C                -..SLG.........QE-LL....-RF-I--F--TT--Y-P...........GQ--ARQ--RA-Q--AKQRQ-VIHLA----.E-PVSQI....DHLAQEFDQLVLDNLQQPPSLPPGHP-ENQTANSSS*..........................    99
OLC.CI.97.97CI12              LQSHLGR..............GER-WY-IN-I-GTPPIFRYP.........PGQ--ARQ--R--Q-VLREHL-RVELYH-VY.RDIESKLERLTITPVEGSSESHSLDPEELATRL-NLK-EE-TWDQFAAEETEKGKSSTEGSGAIGERRSRNNWS-RTSI-QIWQ*..   143
DRL.x.x.FAO                   -..SAGPERE........PPPWFQEYL--VTR-W-GRI--.........RDDQLPQ-ARQ-KR--A-Q-RVEH--RTLQA---.QSLER-NSRD.....LVEGIERIH-AEERESSS*.................................    92
RCM.NG.x.NG411                -..LLGE........EE-ADQ-IR-RI---H-IHL---Y-...........QSG--ANQ--RK--Q--R-WT--LQLA---F.LYPDTPPDRD......-LQEA-SN-QQ---SDLPEPPVNPFSNPSSFA................VDRS*..................   107
RCM.GA.x.GAB1                 -..LPGQ........DE-...D-RKKI---NF--LI--Y-...........QHG--ANS--KK-----R-WQ-VQQLA----.LDSTDPPV-Q......DLDAAIAD-QK-Q-NNLPEPPVD.FS*................................    93
MND_1.GA.x.MNDGB1             -.STGNV...........YQELIR-YLVVV-K--EGLSTCLFT...GLPAEPIPQ-ARQ--R-KQQL-T-RA-LRELEG---.KQI-D-GPDQ.....LCQGVTNLA-AEKSESS-*..................................    95
MND_2.x.x.5440                -STD-GDL..........DEGFWRKYQAIV-Q-WEGIS-SSLCLQGRLSPAPQPQ-ARQ--R--A-L-KTEH-VRELQT--*............................................................    71
MND_2.CM.98.CM16              -PTETGDL..........DRDFWKKYQKIVQR-WEGLS-SSLCLQGRLSPSPQPQ-ARQ-KR--A-L-KTEY--RQLQS--WEALEQRANQRLAEDFKGLRVAEESTSSN*...............................   100
MND_2.GA.x.M14                -STEPGEL..........GRDFWEKYH-IV-Q-WEGLS-SSLCLKGRLSPTSQPQ-ARR-KR--AQL-RTEH-VRELQ---WQTLEARANK-LVEGVNNLHLAEESAGSN*...............................   100
GSN.CM.99.CN166               -ADHAR........-NDQKLQ.N--LAC----T-HR-RLLTSLPA.DPYPTSSGTRSARRN-R--W-A-QG-VREI-NRILESLVGRPEEPGDLDLPDLGQLS-DS-WDREVPVGTAAESNPESNPASETAQGSSSQSDSLH*..........................   132
GSN.CM.99.CN71                -.-DPAN.......-RDQKLQ.N-LLAC----T--K---Y-...........TAS---S----KK----A----VRQ--H---.ESLV--PE-.......SGN-D--D-GH-S--SP--RQVPVGAAAEG..............-PTSETAGG-GSQSS-LY*.......   120
DEB.CM.99.CM40                -.-HAGGR......-SAE-NTRQ-LKVIS---I------Y-...........KGG-SAST--R--QK--R--G-VD--A----.QSH--G-PT.......TADVA--D-SQ-H-AD*....................................    90
DEB.CM.99.CM5                 -.-HAGGR......--A-Q...A-L-MI-I--S------Y-...........KGG-SAST--R--QK--R--E-VDD-AQ---.EAH---HLT.......ASDVA--N--E-R-AD*....................................    87
DEN.CD.x.CD1                  -.--LQER.........-Q.....-VKA-----K------Y-...........T---SKS---R---K--RQ-A-VVA------.EYH----LT.......AADVE--D--K-SVSD*....................................    84
TAL.CM.01.8023                -.SLTGES......-KGED...P-LAAI-R-RI------Y-...........EQQ---S---R-Y---KA--E-VEN-AS---.CH---GPQQ.......-PQ----D--K-S-AELGPPEAP.....ESDH............PTGSASH-RLPEN*..............   111
TAL.CM.00.266                 -.SLTGGS......-RGED...P-LAAI-R-RI------Y-...........EQQ-S-S---R-Y---KA--L-VEQ-AS---.EH-----DR.......SDQIH--D-SQ-S-ADLGP-ESC.....ESQE............PTGSASH-RLPTTPPDS-SSDSPVWVQ*   125
MON.CM.99.L1                  -.--NG.R......DEERR...L-SLALAAVRI-QE---Y-...........I-R--AN-----------RQ-L--A------F.YH-----Q-.......-C--DI-D----SISDLP-PPE-VP...................EAAT-PAHTPA---GKP*.........   112
MON.NG.x.NG1                  -.--AE.RG....AAVQD...I-RLAIGA-RIIRE---Y-...........S-S---S-----------A--E-VX-LC---F.NH--D-LPR.......-NS-D--AI-G-R--DLPEPPE-A...................QPPD-SSPRPV--ASG-Q*.........   113
WRC.CI.97.97CI14              ..IVLTAR................KLSQIQQQ-L-AE--QQ...........LGR--ARQ--RN---LQQYRN-RLAVQ---C.NSL-EQL-N.......LTIQH--TPPSHCVESAFVPPVVDKRWVESSQRHQWKSIWSDTQEDCKE-EGEYAGRGISNWQQKDSSKQ*   133
WRC.CI.98.98CI04              ..IVPTAR................-L--VQQK-Q-AE--QQ...........LGR--ARQ--RS---QQYFR--RLAVQ---C.NNLVEQLDK.......LLICS--NPPSNSVESPCVPLLVDPAWVESSQKNQWSSACMHTQEDYKE-EGNNGERDIRK*.........   124
WRC.GM.05.Pbt_05GM_X02        ..IIHTAR................-LI-ISQT-Q-AESY-E...........PGR--ARQ-------FHSHR-IRLAVQ---A.D-L-EQLGN.......LTLQH--EPPAAK-QPS-LSVDRN-EWVESSQKTLWSSVCCATGEDCCASAEKKGEGDIHAWQQRDKQKQQ   134
SUN.GA.98.L14                 -.STG..............DDSINQYL-IS-R--EGLA-G.........N..LPQ-HRQ--R--D-E-KNLH-LRAVQ---F.A-T-DSRLGR.....AFER-SVSDSSQVAESLGNSPS-KHLPPAKFLVAPTYDFLPSWATPLAD--R-AGFAPYSGYEQDQERVQNQ   138
LST.CD.88.524                 -.STGP.............EG--PTYLK-SRI-WTGLQVG.........ERREPR-ARQ--RD-E-Q-KHLH-LRAVQ---F.QATVE-GL-R.....AFTG-TVCDSPEVAEGVGNTAPHPRVAIPHLAGTYSDPFLPPWATSLAN--QIPGGECSENCELDQEQEQKE   141
LST.CD.88.485                 -.STGG.............EE--P-YLKIS-I-WTGLQVG.........ERREPR-ARQ--RD-E-Q-RHLH-LRTVQ---F.HSTVE-GL-R.....AFTG-TVCDSPEVDQGIWNSSPQPCVAVPFVARPLSDPFLPTWATSLE-Q-RMDGGKRSEDCALELEQDQKE   141
LST.CD.88.447                 -.STGG.............EK-SP-YLKIS-I-WTGLQVG.........EKREPR-ARQ--RD-E-Q-KHLH-LRTVQ---F.HSTVE-GL-R.....AFTR-TVCDSPEVDQGLWNSSPQPCVAVPFVARPFSDPFLPTWATSSE-L-QTGGGKRSEDCELELEQDQKE   141
LST.KE.x.lho7                 -.STGN.............GD--P-YL--SRI-WTGLQVG.........ERREPR-ARQ--RE-S-Y-DYLH-LRAVQ---F.QATVE-GL-R.....AFTR-AVSDSPEVAQGRGNTPPITSVAEPQLAVAFVDPFLPKWATPLADQ-QMDGGKRSED-ELAQGEMQKE   141
SYK.KE.x.SYK173               -.P..........DQGS-QLAVFL-MIAHL.....QE-Y-...........G---P--T--R---Q--Q-RT-RLYLQQ--F.EAIF-SRTA.......ALEDS-QQ-QISD*....................................    77
SYK.KE.x.KE51                 -.S--ERE.....DTQEQLLRT-L-IAQQL#Q-EAAA-Y-...........L-Q-PSR---RN-N-Y-QL-A-RLYVQQ--F.E-IAR----.......TLEQSFGE-QITD*....................................    86
COL.CM.x.CGU1                 -.TNAGVRPVAFTVSYAFFQKGWESAI-R-RV-HKH--W-R..........PGV--ARQ--RNTQ-A-KQ-Q--LA-DG-VV.EQLVASLQQWD.....-YCPASTSANN-RQEEECLADS-VSIGTQGSK................-SDS*..................   118
```


PLV Proteins

H1B.FR.83.HXB2	.....	116
H1A1.UG.85.U455	.....	124
H1B.US.90.WEAU160	.....	117
H1C.ET.86.ETH2220	.....	107
H1D.CD.84.84ZR085	.....	117
H1F1.BE.93.VI850	.....	117
H1G.SE.93.SE6165	.....	107
H1H.CF.90.056	.....	117
H1J.SE.93.SE7887	.....	117
H1K.CM.96.MP535	.....	117
H101_AE.TH.90.CM240	.....	124
H102_AG.NG.x.IBNG	.....	124
H103_AB.RU.97.KAL153_2	.....	117
H104_cpx.CY.94.CY032	.....	118
H1N.CM.95.YBF30	.....	105
H1O.BE.87.ANT70	.....	102
H1O.CM.91.MVP5180	.....	103
CPZ.CD.90.ANT	.....	118
CPZ.CM.01.SIVcpzCAM13	.....	124
CPZ.CM.05.SIVcpzEK505	.....	103
CPZ.CM.05.SIVcpzMB66	.....	124
CPZ.CM.98.CAM3	.....	125
CPZ.GA.88.GAB2	.....	124
CPZ.CM.98.CAM5	.....	107
CPZ.CM.05.SIVcpzMT145	.....	123
CPZ.TZ.00.TAN1	.....	105
CPZ.CM.05.SIVcpzLB7	.....	124
CPZ.US.85.CPZUS	.....	105
CPZ.GA.88.GAB1	.....	125
MUS_1.CM.01.1085	.....	110
MUS_1.CM.01.CM1239	.....	106
MUS_2.CM.01.CM1246	.....	99
MUS_2.CM.01.CM2500	.....	117
MAC.US.x.239	.....	107
H2A.GW.x.ALI	.....	107
H2A.DE.x.BEN	.....	103
H2A.SN.x.ST	.....	107
H2B.GH.86.D205	ESGESNHRDPQENQTRT*.....	153
H2B.CI.x.EHO	ESGRRDHCNTQEDQTRG*.....	150
H2G.CI.92.ABT96	.....	100
H2U.FR.96.12034	.....	98
MAC.US.x.251_1A11	.....	107
MAC.US.x.251_BK28	.....	107
MAC.US.x.EMBL_3	.....	107
SMM.US.x.H9	.....	100
SMM.US.x.PGM53	.....	104
SMM.US.x.SIVsmH635F_L3	.....	100
SMM.SL.92.SL92B	.....	99
STM.US.x.STM	.....	99
MNE.US.x.MNE027	.....	107
GRV.ET.x.GRI_677	KSAARP*.....	147
VER.KE.x.TYO1_patent	.....	84
VER.KE.x.9063	.....	89
VER.KE.x.AGM155	.....	89
VER.DE.x.AGM3	.....	89
TAN.UG.x.TAN1	.....	103
SAB.SN.x.SAB1C	.....	99
OLC.CI.97.97CI12	.....	143
DRL.x.x.FAO	.....	92
RCM.NG.x.NG411	.....	107
RCM.GA.x.GAB1	.....	93
MND_1.GA.x.MNDGB1	.....	95
MND_2.x.x.5440	.....	71
MND_2.CM.98.CM16	.....	100
MND_2.GA.x.M14	.....	100
GSN.CM.99.CN166	.....	132
GSN.CM.99.CN71	.....	120
DEB.CM.99.CM40	.....	90
DEB.CM.99.CM5	.....	87
DEN.CD.x.CD1	.....	84
TAL.CM.01.8023	.....	111
TAL.CM.00.266	.....	125
MON.CM.99.L1	.....	112
MON.NG.x.NG1	.....	113
WRC.CI.97.97CI14	.....	133
WRC.CI.98.98CI04	.....	124
WRC.GM.05.Pbt_05GM_X02	HQGNYCQTLHKP*.....	146
SUN.GA.98.L14	QGESIIVSEGKK*.....	150
LST.CD.88.524	QRRAIKN*.....	148
LST.CD.88.485	QKRHISSSCERHQTGTKEKMEIQIQREKWVSQ*	173
LST.CD.88.447	QKRHISSSCKKPQTGTKEEMEIQIQRKKWVSQ*	173
LST.KE.x.lho7	QRTVIEH*.....	148
SYK.KE.x.SYK173	.....	77
SYK.KE.x.KE51	.....	86
COL.CM.x.CGU1	.....	118

```
                                                                                   phos          phos
                                transmembrane domain                               cytoplasmic domain
                                Vpu start                                α-helix                                α-helix                Vpu end
H1B.FR.83.HXB2                  TQP...IPIVAIVALVVAIIIAIVVWSIVIIEY......RKILRQRKIDRLIDRLIERAEDSG.........NESE...GEISALVEMGVEMGHHAPW.DVDDL*   82
H1A1.UG.85.U455                 MT-...LE-W--TG-I--L-L-----T--G--*#.....K-L-K-K-----LN-IR-----------------D....GDTEELSLL----NYDLG.VDNN-X   81
H1B.US.90.WEAU160               M-S...LQ-L---------G----------L---.....-----------IRD-----------------....GDQEELSAL----------.-I----   82
H1C.ET.86.ETH2220               DLLAKVDYRIV---FI--L-------T-AY---.....--L----R----K-TR-----------------D....GDTEELSTM-D--NLRLL.--N---   85
H1D.CD.84.84ZR085               M-S...LQ-L------L-L-------T--F---.....-R-K-----W----IR--E-----------------....GDKEELSTL---------DVD----   83
H1F1.BE.93.VI850                MSY...LLAIG-A--I--L-------T--Y---.....K-LV-----NK-YK-IR-------------------....GDAEELAALG---PFI-G.-INN--   82
H1G.SE.93.SE6165                M-S...LV-L---G-I--F-A-----T--F---.....-E-RK-KR-GK-L--IR-----------------D....GDTEELVTL----DFD--.VG-N--   82
H1H.CF.90.056                   MYI...LGLGIGALV-TF--A.VI--T--Y---.....K-LV--K------E-IG-------NES......DG.D......TEELSKLM----LNLG.Y-A---   81
H1J.SE.93.SE7887                MI-...LQ-A----FI---FL--GM-T--Y---.....K-L-------K----IR-------NES......DG.D......TEELADL--R-P-DL-.N-N---   82
H1K.CM.96.MP535                 MVS...LA-SIVALV-AL-LA.-I--T--Y---.....--LVK-KR-NW----IR-------------------....GDAEELADIG-L--LILG.NI-N--   81
H101_AE.TH.90.CM240             MT-...LE-S---G-I--L-L----*T--A--V......K-----------VK-IR-------------------....GDTDELAKL----DFD--.VG-N--   81
H102_AG.NG.x.IBNG               M--...LT-T---G----F-A--------Y---.....---RK-K-----L--IR-----------------D....GDTEELSTLM---YEYIL.-N-N-X   82
H103_AB.RU.97.KAL153_2          M-S...LA-A-------VG------G---F---.....----------IR-------------------.....GDQ-ALM----LV--.-A----   79
H104_cpx.CY.94.CY032            MLF...WE-W---G----L--V----TL-F---.....K-LR---R--S-YN-IR-----------------D...GDAEELSTL-G--NFD--.VG-N--   82
H1N.CM.95.YBF30                 ML......SLGFI--GA-VS--VI--ALLYR-.......K--KL-E--KHIRQ-IR--E----------D.GDA-WLDGD-EWLVTLLSSSK.LDQGNW   82
H1O.BE.87.ANT70                 MHH...RDLL--IIISALLF-NVIL-GFILRK-..LEQKEQDRKE-E-LERLR-IR-IRD--DYES......-G.........EEEQ.E-MDLVLSHG.FDNPMF   83
H1O.CM.91.MVP5180               MHQ...ENLL-LI--SALCL-NVLI-LFNLRI-..LVQRKQDR-EQE-LERLR-IK-IRD--DYES.......-.........EEQQ.E-MELI-SHG.FANPMF   83
CPZ.CD.90.ANT                   MTNIFEYAFL-FSIVLWI-C-P..ILYKLYKI-.KQQQIDNKRN--I-EV-SR--SIDSAIEE.........D-.........EAD..TYYL-.SGF..ANPVYR   79
CPZ.CM.01.SIVcpzCAM13           MIL......-ALGCLAIAL-LLN-FI-RNLW......RLCKQNKLEVE-EN-AL--T----------------D...-GEEEKLQRLLHDHDNMHG.FANP-F   81
CPZ.CM.05.SIVcpzEK505           MLL...LIKLGFIG-AIETL-V----A--YRI-.......-EVKVEE--SQ-RQ-IRD-----------------.....DGDAEELANLLPPDR.IDQ-NW   78
CPZ.CM.05.SIVcpzMB66            MDI..VQQVGLL-V-IIELV-V--I-VK-Y......KLCKEDR--K------A-IR-------NES......DGDTEELQDLITEGDNLMHI-IRDNR.NN*.V-   88
CPZ.CM.98.CAM3                  MLT..W.EQIGLIGIGIE------A-G-AFK-W..RKGKEEEN-AQ--RT--E-I-S-Q----------------D.......GEDQRQLDREI-VYG.FDNPMF   83
CPZ.GA.88.GAB2                  MLS..MWVAIGLIGIGTLLV-N----G--G-SVYKRWKRH-EEQRI.--LI.IKIT------------------D....GEDKETL...ATLL-NNG.FDNPMF   84
CPZ.CM.98.CAM5                  MLI..W.EQIGLI--GIEL--V----G-AYKKW..REIKEERN-TQR-YN-AE-IRI-Q----------------N.......GEE-AQLDNLL--NG.FDNPMF   83
CPZ.CM.05.SIVcpzMT145           M-L....E--L-ILFIALMLV--FA-IAAY......KEYK-LQQV-R-E--Q--IRS------NES......DGD...EILLVEELMQVHQHQVNPD-.MDRI-F   85
CPZ.TZ.00.TAN1                  MIK..IVVGSVSTNVIGILC-LLILIGG.GLLIGIGIRRELERERQHQRV-ERLARRLSI---VEE......D-EF...............NWNNFDPHNYNPR   80
CPZ.CM.05.SIVcpzLB7             MDL....IELGLIG--IEL--V----LKAYQL-......KENI--KA-NK--E-IR--G-----------------.........GD-DELHAILRSG.-PELVL   76
CPZ.US.85.CPZUS                 MLN..WFE-GL-ALGI.EG-LVVII-GL-ARLW..RQIKI-ENT-QE-QN-LE-IRI-E----------------D....GEEEETL....AKLLSSLELDNX..   81
CPZ.GA.88.GAB1                  MTL..LV.GLVLILVGL.-AWN-CI-GYI-KWGYRRYKRHRLETE..-E--NLI-R------------------N...--EEERL-QLIHNYN-NNH.FANPMF   86
GSN.CM.99.CN166                 MH-.....AAVWWWGAAI-TFIYLCVALLALY.....LAWDKWVKG-PKPTQVAV-RLI-............D-ED.........SGIYDDASSELTGFNGFANP   74
GSN.CM.99.CN71                  MSA.....AALWWWGAAV-TFIYFCLA-FALY.....LAWDKWIKG-PK.IPVAV-RLV-............DDE-.........SGIFEDASSEPNAYGFANPG   73
MUS_1.CM.01.1085                MNY......WYLA-VI-TG-YFVIAIFAF..VLAYQRW...CKPKKVEVSV-RL-E-GDG---IFE......DAED..........D-AES.E---FANPAFEQ-   77
MUS_1.CM.01.CM1239              MNY.WYLAAALVTGIYFL-ALFAF-LAYQ...RWCRPK........VEVSV-RL-E-GDS---IFE......DAED...........PNGD----FANPAFEQ-   76
MUS_2.CM.01.CM1246              MHY.WYLGAAIVT-IYLV-ALVAFILAYQ...RWCQPK......PPIEVNVLRL-E-GDS---IFE......DACD........GEDEDSHRAFAN-SFEP*SNI   80
MUS_2.CM.01.CM2500              MNW.WWFAAAVVT-IYFV-ALVAF-LAYQ...RWCQPQ.....KGQVEVNV-RL-E-GDT---IFEDAE...DG-D..............D.QR-GFLNPAFE--   78
DEN.CD.x.CD1                    MLK.......IKLGPIEYYCLVFA-VITW-AAIGVGYLAYRAYKSYREELRYI--RLWSI---..........Y--S......QEDP....*GPPPY-V..DNWNN   76
MON.CM.99.L1                    MNY.WWSLVAITYS-ILIALPVAAWAW.W...RYYKITK-FKRIDQE-Q---QIHER-RH---VDTESESEQH-ET...................GFVNP-FNDD   82
MON.NG.x.NG1                    MNT.....YWIAAIS-WCV-VLLAPCTLY..LLYQSYKEH-RVSRFEQ-FQRLVQXYQEX---..........Y-D-.......DEDHNSFDNPLFDDG.-P-QWV   81
```

```
                                                                                                                                                   glycosylation NDT
                                                                                                                                                          V1
                             Env start                signal peptide gp120 start                                                    glycosylation NVT                                                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
H1O.BE.87.ANT70              -E.........KRNK.KLWTLYLAMA.LITP-LS..LRQ-.....YA---A-----ED--...PV-------NLTS--K--I--SQ----.----T-Y-...YP-H---DD--I-.--Y-----Q----------------QM-F---QME--NIAGT--E..........  133
H1O.CM.91.MVP5180            -K.........KNN-.KSWSLYIAMA.L-IP-LS.YSKQ-.....YA---S-----E--A...PV-------NLTS--Q--I--SQ----.-------H-...FP-G---D--DI-.--Y--D----------E------E-M-F---QMN-V--QTNKTG.........  134
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.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--K..-N--D--Q-------E--------------T-S-SSWRSVN-S..........  133
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.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-RNKTST...  144
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.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.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.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.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.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
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
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-SEPTTT......  139
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-TTTA...  142
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
                                                      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.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.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.SN.x.ST                  KMCG..........-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.  131
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
H2B.CI.x.EHO                 -AHV..........NNYL.....-VTLL-ISIYG.YMG-N....F---F--I-A--N-S...IP---.........-RNRDT-G-VQ-L-.DNDDYT-...IQL-I--A-DA-D..-TVTD-ATK-VW--FET-I-----------TM--NKTWSSASKETTTSSASLR  132
H2G.CI.92.ABT96              -AYL..........GNQL.....-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
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
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
SMM.SL.92.SL92B              -ACP..........GLHL.....-IDILFLSVLG.TWCAQ....Y--IF--I-A-RN--...IP---.........-QNRDT-G-VQ-L-.DNGDYS-....LAL----A-DA-D..-TVT--AI--VWN-FET-I----------IAM--NKNET-RWGLTRAATTTSS  132
STM.US.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
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
GRV.ET.x.GRI_677             -GRL..........LIKI.......LIIAIGI-I.GIGN-.....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
VER.KE.x.TYO1_patent         --YT..........IITL.......-IIVIGIGI.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
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.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
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
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
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
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---IVNAASTTPTPTTVA  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
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
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.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
MND_2.CM.98.CM16             -LRY....................IILGI-VGL.GLGNQ....--------T-K-HK-E...-H--------........--NDSL-V-TS---SLLHYEEQ..PIPNITWNF.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
DEB.CM.99.CM40               --LI..........QLLL...LI-GLIVIGTQGK.QKKQQ....YL----------VD-K...VD---TANSSE......SG--VT--L-HALVREEVPMPN-TQ-.FNA-DNP.....IE--LWQ-MT--YK--F---------Y---MQ-IKTSTNPTPTNT-TTTTTI  136
DEB.CM.99.CM5                -LSL.................LIIIASLTLVGAA.P-KQQ....------------HN--...VG---TANSSE......SG--VT--L-HAMVRE-VPMPN-SQ.YFDS-SNV.....-EQ-LWD-MT--YR--F---------Y--RMN-SSTTTT-TAATTTI.....  127
DEN.CD.x.CD1                 -SQV......RRTL....KVLLLI-W..TIGIIC.T---Q.....Y---F--I-A-RNSA...AP---T.......AMDRG---AAS---.E--YQR-VWMENATDYFDIYHNP...IVDQVVQNMKD..-YK-TF---------V--RMN--RKEDNSTNAVTVSSTTTT  136
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
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
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
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
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.....DPCES  125
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-NTKDF.NPCEI  129
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
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---GEGLGKKRTTTTTPTTT  135
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.88.447                -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.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....TKTPTTTPKTTT  128
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
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
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-........  124
```

```
                                                             glycosylation NIS
                                                         glycosylation NCS
                                        V1                                       V2
                                                                                glycosylation NDT        glycosylation NTS              glycosylation NKT   glycosylation NGT   glycosylation NVS            glycosylation NGS
                                                                     *                                               *         *              *              *          *       *
H1B.FR.83.HXB2            ..................................M.IMEKGEIKNCSFNIS.TSIRGKVQKEYAFFYKLDIIPIDNDT.............TSYKLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGP..CTNVSTVQCTHGIRPVVSTQLLLNGS  264
H1A1.UG.85.U455           .................................IT.DGVRE-M------MT.-EL-D-K--V-SL--R---VQ-NKTDN............N--R-IN----T---------------------------KDPE---K--..-R-----------K-----------  262
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-----------------------  268
H1C.ET.86.ETH2220         .................................I.NSAND-M-------T.-EL-D-KR-A--L------V-LN-GS............-D-R-IN----T---------LD-----------Y-----RD---T---..-H-----------K-----------  261
H1D.CD.84.84ZR085         ................................LP.TVKP--M-------T.-VVTD-RKQVH-L--R--VVQ---EGKNEIN.....DTYGT-R-IN----A---------------------------D-R------..-K---S-------------------  271
H1F1.BE.93.VI850          ..........................................MQ-----MT.-EV-D-KL-LS-L--R---V--G-NNS............SE-R-IN----T---------WD-----------Y------D-R------..-K-----------K-----------  255
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-----------  269
H1H.CF.90.056             ..................................MEAG--LT-----VT.-VL-D-Q--VH-L--R--VV----NS.............-Q-R-IN---------------------------------------L..------------------------  257
H1J.SE.93.SE7887          ..................................VSSPDIMT------T.-E--N-RKQ---L--RQ-VV---SNN............KN-I-IN------K--------Q-----------------D-N-----S..-K-----------K-----------  258
H1K.CM.96.MP535           ..................................VVSPA---------T.-E-KD-KK--S-L--R--VL-LNGEGNNSS.........-E-R-IN----T---T----T----------------KD-R------..-K-----------K-----------  266
H101_AE.TH.90.CM240       .................................II.GNITD-VR--T--MT.-ELTD-K--VH-L------VQ-EDKKTS...........SE-R-IN------K----I--D------T---Y------D-N------..-K---S-------K-----------  267
H102_AG.NG.x.IBNG         ...................................DMN---------T.-EV-D-KK-MH-L--R--VVQ-NENNG............SQ-R-IN----A--------------------------KD-G------..-K-----------K-----------  261
H103_AB.RU.97.KAL153_2    ...............................NTS.SIKMM-M-------T.-DL-D--K----L-----VVQ-----------------D--R-I------V------I------------------D-K------..-------------K-----------  255
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-----------  269
H1N.CM.95.YBF30           ..............................MTTREP.DIGYKQM------AT.-ELTD-KKQV-SL--VE-VV--NA.............YNKT-R-IN---TAV------T------------P----M---EGN-S-N-S..-------------K--I----I----  267
H1O.BE.87.ANT70           .........................................NLM-K-E--VT.-V-KD-KE-KQ-L--VS-LMELNETSSTN.....KTNSKM-T--N--STT-----------------------Y--F---STE-----T..-R-ITV-T-------T-----I---T  253
H1O.CM.91.MVP5180         .................................LL.NETIN-MR-----VT.-VLTD-KEQKQ-L--VS-LSKVNDSNAVN.........G-T-M--N--STI-K-----------------T-Y--F---DTD-----L..-H-I-V-T-----K-T-----I---T  257
CPZ.CD.90.ANT             ...........................VTPKTTTP.-VDGMKLQE-N--Q-.-GFKD-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-----T  264
CPZ.CM.01.SIVcpzCAM13     .....................NNPTLKNNTSGKN.ETGVLQM---T--TT.-EL-D-KKQV-SL--VD-LQSLGS............GTGDT-TMIN---TA--------------------------DVN-S-K-K..-R-----H-----K---T---II---  274
CPZ.CM.05.SIVcpzEK505     .................................VN.QTNHVQMQ-----VT.-EL-D-KKQV-SL--MG----L-TNNSS........GNNSQ-R-IN---TAV------I------Y----P----I---DQD-----E..-N-----------K--I----I----  257
CPZ.CM.05.SIVcpzMB66      ...................TVQISTGNDSTANNI.TVGTIDMY-----AT.-EL-DRKKQV-SL--RQ-LE-LEGNKPPE.....GDKNAL-R-YN---TAM----S---L-------------L----D-N-T-I-Q..-K-----H-----------F----T  285
CPZ.CM.98.CAM3            ...........................MTTNSPN.EKXDSTV---T--MT.-EV-D-EK-V-SL--VD-LVL-----D.............T-R-IN---TA-------T----------TP-Y--M---MPN-----TGR-N-I-------------T---I----  270
CPZ.GA.88.GAB2            ............................NGIEK.DGLAQDMR--T--TT.-EL-D-K-QI-SL-W-N-LVGTN-..............TFR-IN---TA-------T---------------L----D-DYP-K-K..-K-----H-----K-T-T---I----  262
CPZ.CM.98.CAM5            ................................TEL.KNI-EIM---T--MT.-EV-D-KKEV-SL--VD--V--GGNKSND..........T-R-IN---TTV------T-------------P----M---DQK-------..---I------------T--VI----  262
CPZ.CM.05.SIVcpzMT145     ................................NST.DDRL-DMR-----VT.-EL-D-KRQV-SL--VE--TA-G-.............NST-R-IN---TA-------T--------------L----DIDYK-NET..-K-----H-----K--AT---I----  259
CPZ.TZ.00.TAN1            ...................LNSATTTLTPTVNLS.SIPNY-VY-----QT.-EF-D-KKQI-SL--RE--VKE-GN.............NN--Y-HN--------E-D-ST------R---------L---RDQN-T-K-Q..-S---V-H-----Y-MIA-A-H----  261
CPZ.CM.05.SIVcpzLB7       ..................................I.NSSNETM------TT.-EL-D-KKQT-SL--VD-LTQ-NK.............TTE----IN---TA------I---------------------L-R-----Y--S--..-N-----H-----K--I----I----  258
CPZ.US.85.CPZUS           ................................SSA.VNTTDIMR------T.-EL-D-KKQV-SL--VD-LAH-N-N...............T-R-IN---TA-------T-------------P---------E-D-K-K-E..-K-----------K---T---II---  249
CPZ.GA.88.GAB1            ................................NQT.SSPPL-M------VT.-EL-D-KKQV-SL--VE-VVNLG-ENN..............T-RIIN---TA-------T-------------------D-D-S-K-K..--------H-----K---T----I----  262
GSN.CM.99.CN166           ..................PRPPNVSTTTQWGSWG.GE.NGTGQPLY-----QT.-EF-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-H---T  275
GSN.CM.99.CN71            ..................TTSPSSPTPTPWGNWG..G.NGTGQP-Y-----QT.-EF-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-H---T  277
MUS_1.CM.01.1085          ..............TGGNSTGTTAPPTTPWGSWGNGN.DSAGQPMF-----QT.-EF-D-KRQM-SL-FVD-LMRDSSESEND..............--YI-N----Y-----E-S--Q-V--Q------YSL----DAN-T-Q-E..-N--TARH-----L-L-A-W-Q---T  276
MUS_1.CM.01.CM1239        ..............TTTTASTPTKTTPVTPWGKW.DA..NGTEQSMF-----QT.-EF-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-Q---T  271
MUS_2.CM.01.CM1246        ..............TTAASATTSGHTTVTPWGRWADN..NETQLTMY-----QT.-EF-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-Q---T  275
MUS_2.CM.01.CM2500        ..................TKPTTPSTTPWGRWSDG..NDTMQVMY-----QT.-EF-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-Q---T  263

                                          *         *            *                              *                *          *             *              *          *       *
MAC.US.x.239              TTSTTA..........SAKVDMVNETSSCIAQDNCT.GL-QEQMIS-K--MT.GLK-D-KKEYNETW-SA-LVCEQ.........GNNTGNESRCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTNYS-FM-K.-SK-VVSS--RMMETQT--WFGF--T  280
H2A.GW.x.ALI              ...................KPLINESDPCIKADNCPRGLGDE-MV--R--MT.GLQ-D-PKQYNETW-SK-VVCEP.........FNTTTNQ-RCYMNH-------ES-D-HYWDA-RFR----P-Y-L-R-DDINYS-FA-N.-SK-VAAT--RMMETQT--WFGF--T  271
H2A.DE.x.BEN              SRPPTS............AASIINETSNCIENNTCA.GLGYE-MMQ-E--MK.GLEQD-KRRYKDTW-LE-VVCDN.............TTAGTCYMRH----I-KES-D-HYWDAMRFR----P---L-R--DTNYS-FE-K.--K-VAAS--RMMETQT--WFGF--T  274
H2A.SN.x.ST               ...................NTTIGENSSCIRTDNCT.GLGEE-MVD-Q--MT.GLE-D-KKLYNETW-SK-VVCES..........NDTKKEKTCYMNH-------ES-D-HYWDTMRFR----P---L-R--DTNYS-FE-N.-SK-VAAT--RMMETQT--WFGF--T  269
H2B.GH.86.D205            TTS..............RGLKTINETDPCIKNDSCT.GLGEE--MQ-N-SMT.GLR-DELKQYKDTW-SE-LECNN...........TRKYTSRCYIRT---TI-QES-D-HYWDSLRFR----P--FL-R--DTNYS-FM-N.-SK-VASS--RMMETQS--WFGF--T  274
H2B.CI.x.EHO              SSTQ.................TLLNEDSKCIQNDSCA.GIGLE-MID-Q-KMT.GLK-DESKQYKDTW--Q-LVCEK..........GTRSNESKCYIKT----I-QES-D-HYWDSLRFR----P---L-R--DTKYS-FM-N.-SK-VVSL.YRMMETQT--WFGF--T  272
H2G.CI.92.ABT96           TVPTTK..........MVTAELVNSTSQCLMYDNCT.GIQSESMVG-K--MT.GLK-DQRKEYNETW-SQ-LQXEQ.........SNKSENESRCYMRH------QES-D-HYWDA-RFR----P---L-R--DTNYS-FA-N.-SK-VVST--RMMETQT--WFGF--T  280
H2U.FR.96.12034           PAMQESS.............KVINETEPCIRNNSCS.GL-QEPLVS-R--MT.GLK-D-K-EYKETW-SH--VCEQ..........NTSESESKCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTNYS-FM-N.-SK-VVST--RMRGTQT--WFGF--T  276
MAC.US.x.251_1A11         AAPKAM..........SEKINMINETSSCIVHDNCT.GL-QEQMIG-K--MT.GLK-D-KKEYNETW-SA-LVCEQ.........GNSTGNESRCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTNYS-FM-K.-SK-VVSS--RMMETQT--WFGF--T  279
MAC.US.x.251_BK28         AAPTSA........PVSEKIDMVNETSSCIAQNNCT.GL-QEQMIS-K-TMT.GLK-D-TKEYNETW-ST-LVCEQ.........GNSTDNESRCYMNH------QES-D-HYWDT-RFR----P-Y-L-R--DTNYS-FM-K.-SK-VVSS--RMMETQT--WFGF--T  282
MAC.US.x.EMBL_3           AAPTSA........PVSEKIDMVNETSSCIAQNNCT.GL-QEQMIS-K-TMT.GLK-D-TKEYNETW-ST-LVCEQ.........GNSTDNESRCYMNH------QES-D-HYWDT-RFR----P-Y-L-R--DTNYS-FM-K.-SK-VVSS--RMMETQT--WFGF--T  282
SMM.US.x.H9               QTSTTPPS......PI..IAKVVNDSDPCIRSNNCT.GL-QEPMVS-K--MT.GLK-D-KREYNETW-SR-LVCEQ........NSNGNETDSKCYMNH------QES-D-HYWDA-RXR----P-Y-L-R--DTNYS-FA-N.--K-VVSS--RMMETQT--WFGF--T  283
SMM.US.x.PGM53            TTTTTKSPKA.....EAITAKVINESDPCISNNNCT.GL-QEPMVS-R--MT.GLKKD-KKEYNETW-SR--VCEQ.........STTSRNESKCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTNYS-FA-N.-SK-VVSS--RMMETQT--WFGF--T  285
SMM.US.x.SIVsmH635F_L3    AKSTTSTTTTT......VTPKVINEGDPCIKNNSCA.GL-QEPMIG-K--MT.GLK-D-KIEYNETW-SR-L-CEQ..........PANGSESKCYMQH------QES-D-HYWDA-RFR----P-Y-L-R--DSNYS-FA-K.-SK-VVSS--RMMETQT--WFGF--T  283
SMM.SL.92.SL92B           PTTTSPLTAAS.....PSGEEIVNDTMSCTKNNNCS.GI-QEPMIG-Q--MT.GLK-DQKRQYNETW-SR-LVCEQ...........GGNESSRCYMNH------QES-D-HYWDA-RYR----P-Y-L-R--DTNYS-FA-N.-SK-VVSS--RMMETQT--WFGF--T  283
STM.US.x.STM              PTAAAAAT..........KPELVNETSSCVSNNNCT.GL-EESLVG-K--MT.GLK-D-KREYNETW-SS-L-CEQ..........NVTGEESRCYMRH------QES-D-HYWDALRFR----P-Y-L-R--DTNYT-FA-N.-SK-VVSS--RMMETQT--WFGF--T  279
MNE.US.x.MNE027           PTTKTTTT.........KEIEVVNENSTCVNRDNCT.GL-QEPMIS-K--MT.GLK-D-KREYNETW-SA-LVCEQ.........GNSTEDESRCYMNH------QES-D-HYWDA-RFR----P-Y-L-R--DTKYS-FM-N.-SK-VVSS--RMMETQT--WFGF--T  281
GRV.ET.x.GRI_677          T............TPCQN..........CSTEQIEG.E-AEEPAS--T-A-A.GYQ-DVKKNYSMTW-DQELVCNNKTGSEKGSK........DCYMIH--D---KE--D-TYWDTLRVR------Y-L----D-DYR-FA-K.-K---V-H--RL-NTTIT-GIG----  267
VER.KE.x.TYO1_patent      TG...........LPCVGPTSG.ENLQSCNASIIER.E--DEPAS--T-AMA.GYV-DQKKNYSVVWNDAE-YCKNKTNSTSK..........ECYMIH--D---KE--D-TYWDQLRLR------Y-L----DEDY--YKQN.-S---V-H--GLMNTT-T-G------  277
VER.KE.x.9063             APKTG........DPCIKSTNNNVNLQPCNASLIEE.EL-EEAAS--T-AMA.GY--DQKKNYSVVWNDAE-YCKNSSSSNSSTK........ECYMIH--D---KE--E-TYWDELRLR-------L---HDHDY--YKQN.-S---V-H--GLMNTT---G------  284
VER.KE.x.AGM155           K............LPCVRNKTD.SNLQSCNDTIIEK.E-NDEAAS--T-AMA.GY--DQKKNYSVVWNDAE-FCKRSTSHNGTK.........ECYMIH--D---KE--D-TYWDELRLR------Y-L----DWDYA-FK-E.-S---V-H--TLMNTT-T-G------  279
VER.DE.x.AGM3             TNITASTT....TLPCVQNK.TSTVLESCNETIIEK.ELNEEPAS--T-AMA.GYV-DQKK-YSVVWNDAE-MC...........KKGNNSNRECYMIH--D---KE--D-TYWDELRLR-------L----DYDYA-FKTN.-S---V-H--NL-NTT-T-G------  284
TAN.UG.x.TAN1             ..............DPCPNT.....DESSCNATLVTN.S-DYENSSI---AMA.GYR-DVKK-YNSTW-DQELVCEKENNTTGTR.........GCYMIH--D---KE--E-TYWDTLRLR-----------KDTNYT-F-V..-R---V-S--GLMNTT--SAFGI---  275
SAB.SN.x.SAB1C            .....................PEVVSVGFNDSVIEQ.E---EQAM----AMA.GYR-DVKKNYSTVWDDQEVVCEEGREKSNAT......HTVGCYMIH------KE--D-TYWDTFRLR------Y-L-R-ADTDYS-HKA..-R--TVSA--RL-NTT---GIGI---  274
OLC.CI.97.97CI12          .............TTVTSSTGGQTTTIGLGVDNSTV.KP-TEKL-Q-RY--T.GLC-DCKIEMNQT-DLE-VACTN-................TCYMSH--------D-QYAVSDQQMVKL---P-RVLI----R-E-P......-R-I-A-T--PPLSAT--SFFGF--T  266
DRL.x.x.FAO               TTVNTS............VASTTTANDTLEIAVNNT.EPVMETNRV-KY-VT.GLC-DCKEEITEN-RYD-VVCKCGNETESPESNNCTNKKRTCYV----ST----D-N-A-TDMMKFRL---P-YIL-R--E-LNISSK...-K-ITA----RHMPATI-SHFGF--T  286
RCM.NG.x.NG411            RATTRTT..KEIGKDCYAN......QTECS.......SVPDADVM--E-TVA.GLK-DEKM-VNDTW-SR-LWCKA...........GNNNTRQCFIRH---TS-Q-F-EPKYW--FRLR----P-Y-----LDANYT-FDT..-Q--TATS---M-NTT---SFG----  272
RCM.GA.x.GAB1             KTTTQRT..IGVEKECTAG......NETCE.......EVQDADVMS-E-AVA.GLK-DEKH-YNDTW-SR-LWCEK.......ETNSTNSTKKKCFVRH---TS-Q-F-EPKYW--FRLR----P---L-V-KD-NYT-FDT..-V--TATS---M-NTT-ASGFG----  278
MND_1.GA.x.MNDGB1         TSTVASST.................EIYLDVDKNNT.EEKVERNHV-RY--T.GLC-DSKEEIVTN-RGD-VKCEN...............NTCYMNH--E--N-ED-Q-GLLIR.C-LG-V-P-YVM-RY-E-LN-NKL...-S-I-A----QHLVAT--SFFGF--T  263
MND_2.x.x.5440            TTTTPSTT........TSSSTNKTTTPVLVVEKQNN.ETTTQQNRV-K--TT.GLC-DCKLEIEEN-RYE-VTCTKLNKTGS..ATNSTEPEYECYM----AT----DRN-A-TDRMTFRL---P--VL----E-LNKTKL...-G---A----APLPATI--MFGF--T  287
MND_2.CM.98.CM16          TTT................KTVANTTELDIDTNNT.ETTTQQNRV-K--TT.GLC-DCKLEIEEN-RYE-VTC...........TNGTNNTY-CYM-Q--------D-N-A-TDE-KFRL---P-YVL-R-LERLNVSKK..----ITA----QPLPATT--MCGSS-T  266
MND_2.GA.x.M14            PTTP...................AVTGLEVTVQNN.ESTTQ-NRV-K--TT.GLC-DCKIEIKES-RYE-VTCT.........KKENDTETEECYM-H--------D-N-A-TDKMTFRL---P-YIL-R-RE-LNKTKL...-A---A----DPMPATI--MFGF--T  266
DEB.CM.99.CM40            ATTT...................KTTTD...WSGE.NITMTQYW-----V-.GPY-D-KE-SS-VWLED--QWA--KDGSGNR.........TGYMKH--D-----S-ETSR-K-FK-R------YGL-R-DD-N-----L..-N--TA-A--NL-HTMA--WVQF---  270
DEB.CM.99.CM5             .......................KTTTD...WSGE.NKTVTQYW--T--VT.GPY-D-KEMQS-VWLED--QWETENQTNTNK......SIRHGYMNH--N-----S-ETSR-K-FK-R------YGL-R-DDRN-----L..----TA-A--NL-HTMA--WMQF---  260
DEN.CD.x.CD1              TTTI..................KTPTTE...WYG..GNKTTIFLK-N--MT.-GKGFRDKRVMYEA.....HCRIE-VME.....DRETNQ-Y-YMRY--Q-T-----E--V---FK-RM-------L----DHPWS-Y-L..-K-M-ARV--DE-HTMA--W-IM-ST  269
TAL.CM.01.8023            TNSVTS.................KRTNDTNVWWG..NNATEPVY--TW-MT.GGF-D-KYTYFSHW-LA-LMKEPGNGN...........DSY-YA-G--L---A---E-THYQAF--Q--------LI---DNPWT-R-K..-H-I-A-H--DP-TTIAA-W-V---T  273
TAL.CM.00.266             .......................AMSNSNMPWWG..DNTTEPVY--TW-MT.GAF-D-RYTYFSHW-LA-LMKEPGNGS.............F-YA-G--V---A---E-THYQAF--Q--------LI---DSPWT-R-K..-H-I-A-H--DP-TTIAA-W-I---T  257
MON.CM.99.L1              P...........................WGNWGG.NGTGQPVY-----QT.-EF-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-WFQ--ST  261
MON.NG.x.NG1              TP..........................WGSWGG.NGTGQPLY-----QT.-EF-D-XKQM-SL-X-E--MQDK...........DNTTDDY-Y-LN---XYX----E-SNY----------P-Y---R-DDPX-T-QNN..-X---A-T-----M-I---WFQ--AT  272
SUN.GA.98.L14             WTTTTAKT.TTQTTSSTSSTASTTTP.MPLDWNCTDTENIAESN-V-KY-VT.GLC-DCKTEVEQN-RDTEVTCNG--...............TCYM-H--D-I---D-H-GIMQNAYFRL-----YML-R-DE-LNATKK...-K-ITATP--NYMTST--SFFGF--T  280
WRC.CI.97.97CI14          NTTRA......TTSATTTVGTVSTTPVMQVENNCSQ.VNTSEINHV-R--TT.GMCKDCRIEIKES-RGD-VVCNN-S................ECYINH--E---K-D-Q-GWQDEQY-RM---P-SVL----E-LNSSKL...-E---A-M--DMMTATI-SFFGF--T  269
WRC.CI.98.98CI04          STGAPTISTISAAPTTSTISAGTRTPAMPVDNNCTT.VNTSEINHV-R--TT.GMCKDCRIEIKES-RGD-VVCNK-G-..............ECYVNH--E---K-D-Q-GWQDEQYVRM---P-SVL----EPLNSSKL...-E---A-L--DMMTATI-SFFGF--T  280
WRC.GM.05.Pbt_05GM_X02    .............TTTPVTPPQNTTPSMDVENNSTS.LNLTDTNHV-R--VT.GLC-DCK-EIVES-RGD-VRCTQHG...........NGTQTCYINH--E---H-D-Q-AWHDEQY-RM---P-YVL-R--E-LNSSKL...-E---A-T--DLMTATI-SFFGF--T  272
LST.CD.88.524             PMPCFIT..ESTTTTTKSSGTEQPSLIEDTNCTRWL.NETTEINRE--Y-VT.GLC-DCKEEIKQN-RYDETTCNE-G...............TCYM-H--D-I-L-D-N-GVMSNAYFRL-----YML----E-LNFSAN..----ITATP--DYMISS--SFFGF--T  283
LST.CD.88.485             PMPCFIA..ET.TTTTTKGNTTMESLIEDTNCTRWL.NETTEINRE-KY-VT.GLC-DCKEEIKQS-RYDETTCDK-G...............TCYM-H--D-I-Q-D-N-GVISNAYFRL-----YML----E-LNFSAN..----ITATP--DYMIST--SFFGF--T  280
LST.CD.88.447             PMPCFIT..EKTTTTTKRSTTGGPELIEDINCIKWV.NETTEINRE-RY-VT.GLC-DCRTEIKQS-RYDETTCDK-G...............TCYM-H--D-I-Q-D-N-GVISNAYFRL-----YML----E-LNFSAN..----ITATP--DYMISS--SFFGF--T  281
LST.KE.x.lho7             QMPCFIN..EQV...TVKNPGNETRLEEDLNCTRGL.NETTERNAE-QY-VT.GLC-DCRTEIKQS-RYD-VTCSGER............ENRTCYM-H--D-I---D-N-GVMQNAYFRL-----YML-R--EQLNFSKK...-E-ITATP--GYMLSS--SFFGF--T  276
SYK.KE.x.SYK173           ...................TPTTSPPTTPPNETWWG.DNSTEPRF-----LT.GGFKD-K-QYR-----D-LMKEEGN.............SSY-Y-LH------SA--E-QT-Q-F--Q----P-YSL----DTN-E-DDV.-----TA-S--QEFNTLA--WFQ---T  258
SYK.KE.x.KE51             ...................SPTTTTNKAATPPAWWG.DNSTETRY-----MT.GGFKD-KHQYN-----S--MREEGN.............ESY-Y-QH-----MRN--E-QT-R-F--Q----P-YSL----DTE-E-DAE..-Q--TA-S---PFNTLA--WFQ---T  265
COL.CM.x.CGU1             ...................TVAPTTTNNSASDWDES.NWKEYPWY--RM-STAFLLKDRKEL-LG-SVEDLTVLGNKNDSNS..........IRATMKD-ANYTV--V-DMTIVD-VRTGF--AP-YML-R-DD-KWD---A..-N--TA-S---EFNIT-MSHV-V-A-  262
```

```
                                        glycosylation NNT
                                  glycosylation NCT        V3 tip
              glycosylation NFT                              V3
                        glycosylation NTS                                          glycosylation NNT   glycosylation NKT                                   CD4                 V4
                                                                                                                                                          CD4        glycosylation NST
H1B.FR.83.HXB2              LAEEEVVIR........SVNFTDNAKTIIVQLNTSVEINCTRPNNN.TRKRIRIQRGPGRAFV....TIGK.IGNMRQAHCNISRAKWNNTLKQIASKLREQFGNNKTII...............................FKQSSGGDPEIVTHSFNCGGEFFYCNST  388
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--T-------------TS  386
H1B.US.90.WEAU160           ----DI--..........-E-------N------V-I------------.---K-TL.----VLY....-T-EI--DI-R----L--TS--------VE----IKQFKNKT-.............................V------------M--------------  392
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--T-------------TS  383
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--------T-------------TS  393
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--TM-----R------DTS  378
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--T----I-R-------TS  395
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-VR------R-------TS  379
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-------------L---TS  379
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------L--M---A-------T-  389
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--TM-H---R-------T-  388
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--T-------------TS  383
H103_AB.RU.97.KAL153_2      -----------------------T-------KEP-------------.---G-H-.-------Y....AT-DIT-DI--------IT--------VI---K----KT.............................IV-N----------M-------------T-  377
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--T------K-------T-  391
H1N.CM.95.YBF30             -NTDGI---..............N-SHSNLL--W-ET-P------G--.-GGQVQ-..--AMT-Y....N-E-IV-DI---Y--V-KEL-EPMWNRTREEIKKIL-K-N.........................ITFRARERNE--L-VTHLM---R-------TS  387
H1O.BE.87.ANT70             -SKGKIRMM........AKDILEGG-N---T--STLNMT-E--QI..DIQEM--.--MAWYSMG..IG-TAGNSS-A-Y-KYNATD-GKI---T-ERYL-LVN-TG.......................SINMT-NH-----L-VTHLH---H-------TA  379
H1O.CM.91.MVP5180           -SR-KIR-M........GK-I-ES--N---T---PINMT-I-EGIA.EVQD-YT.--M-WRSMTLKRSNNTSPRS-V-Y-TYNKTV-E-A-Q-T-IRYLNLVNQTEN.......................VTII-SRT----A-VSHLH---H-------TS  386
CPZ.CD.90.ANT               YQTNTS-VM........NGRKNESVLVRFGKEFENLT-T-I--G-RTV-.NLQ-.---MT-Y....NVEIAT-DT-K-F-TVNKTL-EQARNKTEHV-A-HWKKVDNKTN.......................AKTIWTFQD----VKV-W---Q------DI-  391
CPZ.CM.01.SIVcpzCAM13       --T-NITV-........VN-ASK-THDW----S-A-NLT-K-VG--.-GKVQ-.---MT-Y....NMDHIF-DT-K-F-ELNGTT--E--QKVRES-IKEIKA-ANGTY.......................NIT-EP--------AN-M----------DTR  401
CPZ.CM.05.SIVcpzEK505       --TSNI---.........N-SK-...-LL----E-IP------G-K.--GQVQ-.---MT-Y....N-ENI--DT---Y-EVN-T.-EQIWNTTKQIIINN.......RK.......................NIT-IPNP---L-VTNLMI----------TS  372
CPZ.CM.05.SIVcpzMB66        -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-VTNLHLI---------TS  410
CPZ.CM.98.CAM3              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-----VTNMM-----------T-  388
CPZ.GA.88.GAB2              -----E-VF.....RTKNMTAPGLSD-V----KRAIP---S--G--.-GRA-NL..S--TT-F....NTEAL---P-K-S-HLNGTL---I-NR-KQ-IKNSTTWHRGD-.....................T-TKHP-----V-NFM-----------TS  389
CPZ.CM.98.CAM5              ----GPAVAR.........SEDLRH-I--I-FKEGIN-I-M--G--.S-GQLQ-.---VS-Y....N-ENIV-DP------V-K-Q-MKK-NET-LAIAKYDKSIAKV.......................T-TPNP------TNMM-----------T-  384
CPZ.CM.05.SIVcpzMT145       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------VTN-H-----------TS  377
CPZ.TZ.00.TAN1              -E---TKAY........FVN-SVNTPLL-KF-V-INLT-E-TG--.--GQVQ-.---MT-Y....N-ENVV-DT-K-Y-SVNATT-YRN-DWAMAAINTTMRARNETV.......................QQT-QWQRD----VTSFW---Q-------L-  386
CPZ.CM.05.SIVcpzLB7         ----KIM--..........YKA-...KVL------IH-----VG-KTIKG.-PL.---QL-Y....GTETVV-DT-E---E-NQTH-YKI-N-VKRE-TTV-NESNK.........................TVH-AN-------VANLH-----------TS  377
CPZ.US.85.CPZUS             --TKN-TV-..........SK.NF-DI-L--FSEG-NMT-I--G--.-VGNVQL.---MT-Y....N-P-IV-DV-E------KLT-EKQR-YTLEIIKKEANLT-V.........................ELIPNA-----V-NMML-----------TI  368
CPZ.GA.88.GAB1              ---GNITV-........VE-KSK-TDVW----VEA-SL--H--G--.--GEVQ-.---MT-Y....N-ENVV-DT-S-Y-K-NGTT--R-VEEVKKA-ATSSNRTAAN-.......................TLNRA------VTH-M-----------TS  386
GSN.CM.99.CN166             YQPGNNTRIM.......MNGMKNESIV-GFGEDYHLNLT-I--G-K.-IRNLQ-..-A-MT-Y....SQVIVG--T-K-Y-RLDPDR--RAIREAMKAMN-HWE-KTG.........................RNDTQIRWTSEPK--L-VQ--W-Q-Q-------LS  404
GSN.CM.99.CN71              YQPGNNTRVM.......MNGKKNESIV-GFGEDYQLTLT-I--G-K.-L-NLQ-..-A-MT-Y....SQIIVG-DT-K-Y-KLNHT--DIAIR-AMKAMKNHWEKINNDTT.......................PNKTQIRWTSEPK--L-VQ--W-Q-Q-------LS  409
MUS_1.CM.01.1085            YL-GNNTAVM.......MNGDKNESIG-KFGE-FR-NLT-I--G-K.-IRNLQ-..-A-MT-Y....SQLIVD-DT-R-Y-K-NKTQ-DPA-R-AMLAM-KHWHGRLQKQNHSSINET...................KIKIRWRSEPK--Q-VQ--W-----------LS  413
MUS_1.CM.01.CM1239          YL-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-VQ--W---A-------LS  407
MUS_2.CM.01.CM1246          YLPGNDTAVM.......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-VQ--W---Q-------LS  411
MUS_2.CM.01.CM2500          YLPGNDTAVM.......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-VQ--W-----------LS  394
                                                              V3

MAC.US.x.239                R--NRTY-Y.........WHGR--RTI-SLNKYYNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINDRPK--W-WF.GG--KDAI-EVKQTIVKHPRYTG-NNT.........................DKINLTAPGG.----VTFMWT--R---L--KMN  405
H2A.GW.x.ALI                R--NRTY-Y.........WHGR--RTI-SLNKHYNLTMH-K--G-K.-VVP-TL..MS-LI-.....HSQPINKRP---W-WF.KGE-RKAMQEVKET-VKHPRYKG-NDT.........................NQINFT-PGR-S-A-V-YMWT--R---LH--M-  397
H2A.DE.x.BEN                R--NRTY-Y.........WHGR--RTI-SLNKYYNLTMR-K--G-K.-VLP-TL..MS-LV-.....HSQPINTRP---W-RF.GGR-REAMQEVKQT-VQHPRYKGINDT.........................GKINFT-PGA-S---VAFMWT--R---L---M-  400
H2A.SN.x.ST                 R--NRTY-Y.........WHGR--RTI-SLNKFYNLTVH-K--G-K.-VVP-TL..MS-LV-.....HSQPINRRP---W-WF.KGE-KEAM-EVKLT-AKHPRYKG-NDT.........................EKIRFIAPGERS---VAYMWT--R---L---M-  395
H2B.GH.86.D205              R--NRTY-Y.........WHEK--RTI-SLNTYYNLS-H-K--G-K.-VVP--T..VS-LL-.....HSQPINKRP---W-WF.KGN-TEAI-EVKRTIIKHPRYKGGAKNIT.......................SVKLVSEHGK-S---TTYMWT--R---L---M-  402
H2B.CI.x.EHO                R--NRTY-Y.........WHGK--RTI-SLNSYYNLTMH-K--G-K.MVVP--T..VS-IL-.....HSQPINKRPK--W-WF.KGN-TEAIQEVKETIKNHPRYSG-TNI.........................SQIRLAEHARSS---VRYMWT--R---L---M-  398
H2G.CI.92.ABT96             R--NRTY-Y.........WHGR--RTI-SLNKYYNLT-X-R--G-K.-VLP-T-..MS-LV-.....HSQPINERP---W-WF.GG--KEAMQEVKETVVKHPSYKG-NDT.........................KXITFTTPGE-S---VKFMWT--R---L--KM-  406
H2U.FR.96.12034             RS-NRTY-Y.........WHGR--RTI-SLNKYYNLTMY-R--G-K.-VLPVT-..MS-LV-.....HSQAINN-PK--W-WF.GGN-KGAIQEVKETIANHPRYHG-KNI.........................SQINLAEPAG.--S-VKFMWT--R---L--KM-  401
MAC.US.x.251_1A11           R--NRTY-Y.........WHGR--RTI-SLNKYYNLT-K-R--G-K.-VLPVT-..MSQLV-.....HSQPINDRPK--W-WF.-G--KDAI-EVKQTIVKHPRYTG-NNT.........................DKINLTAPGG.----VTFMWT--R---L--KMN  404
MAC.US.x.251_BK28           R--NRTY-Y.........WHGR--RTI-SLNKYYNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINDRPK--W-WF.GG--KDAI-EVKQTIVKHPRYTG-NNT.........................DKINLTAPGG.----VTFMWT--R---L--KMN  407
MAC.US.x.EMBL_3             R--NRTY-Y.........WHGR--RTI-SLNKYYNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPLTDRPK--W-WF.GG--KDAI-EVKQTIVKHPRYTG-NNT.........................DKINLTAPGG.----VTFMWT--R---L--KMN  407
SMM.US.x.H9                 R--NRTY-Y.........WHXRS-RTI-SLNKYXNLTMR-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGE-KEAIREVKXX-VKHPRYKG-NXX.........................XQIKLTAPGG.----VTFMWT--R---L--KMN  408
SMM.US.x.PGM53              R--NRTY-Y.........WHSRS-RTI-SLNKYYNLT-R-R--G-K.-VLPVT-..MS-LI-.....HSQPINERPK--W-WF.GGS-REAIQEVKETIVKHPRYTG-NET.........................KKINLTAPGG.----VTFMWT--R---L--KMN  410
SMM.US.x.SIVsmH635F_L3      R--NRTY-Y.........WHGNS-RTI-SLNKYYNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGN-SEAIQEVKET-VKHPRYTG-NDT.........................RKINLTAPAG.----VTFMWT--R---L--KMN  408
SMM.SL.92.SL92B             R--NRTY-Y.........WHGGS-RTI-SLNKYYNLT-K-R--G-K.-VLPVT-..MS-LV-.....HSQPINERP---W-WF.GG--REAMQEVKKTIVKHPRYTG-NDT.........................RKINLTAPGG.----VTFMWT--R---L--KMN  408
STM.US.x.STM                R--NRTY-Y.........WHGR--RTI-SLNKYYNLTMS-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-WF.GGE-RGAI-EVKET-VKHPRYTG-NDT.........................AKIRIVAPGG.----VTFMWT--R---L--KMN  404
MNE.US.x.MNE027             R--NRTY-Y.........WHSK--RTI-SLNKYNNLTMK-R--G-K.-VLPVT-..MS-LV-.....HSQPINERPK--W-RF.EGN-KEAI-EVKQTIVKHPRYTG-NNT.........................DKINLTAPRG.----VTFMWT--R---L--KMN  406
GRV.ET.x.GRI_677            RS-NRTE-W.......QKGGNDNDTVI-KLNKFYNLTVR-R--G-K.-VLPVT-..MA-LV-.....HSQ-YNTRLK--W-HF.QGD-KGAW-EVREEVKKVKNLTEVS-E.........................NIHLRRIW..----SANFW---Q------KMD  392
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----ASNLW---Q------KMD  405
VER.KE.x.9063               YS-NRTQ-W.......QKHRVNNDSVIVLFNKYYNLTVT-K--G-K.-VLPVT-..MA-LV-.....HSQ-YNTKL---W-HF.QGR-REAW-EVKEEIIKLPKDRYRGTN.........................DTEKIYL-RQF----AANLW---Q------KMD  412
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----AANLW---Q------KMD  407
VER.DE.x.AGM3               YS-NRTQ-W.......QKHRVSNDSVLVLFNKHYNLTVT-K--G-K.-VLPVT-..MA-LV-.....HSQRYNTRL---W-HF.QGN-RGAW-EVKNEIVKLPKDRYQGTN.........................DTEEIYL-RLF----AANLW---Q------KMD  412
TAN.UG.x.TAN1               Q--NRTE-W.......QKHGVSN-SVI-KLNKHYKLK-V-R--G-K.-VLPVT-..MA-LV-.....HSQQYNTKL---W-HF.QGD-KGAWREVRKTIV-LPKEKYRGTN.........................NTRQIWLSRQW----AANIWL--Q------TPD  403
SAB.SN.x.SAB1C              YVANRTE-W.......QKNGNSNDSVI-RLNRYFNLT-R-R--G-K.-VLPVT-..MA-LV-.....HSQ-YNTRLK--W-WF.GGN-RGAW-EVKETIVRLPPKKYSGTN.........................DTNKI-L-RQW----SEFFF---Q------KMD  402
OLC.CI.97.97CI12            RHKASRL-PI.TGDLNHKPREGTGVYVMAIADKYKL--E-I-KG-R.SIVSVNS..AS-LVYY....AGLEPY--IDKGL-RF.KGQ-GWM-YALGKG-EKENNTWANYTGVNITGIGR.....NSTTFAKAIKR...FKFDE-LKR---DAAENMMMM---YMWF--I-  419
DRL.x.x.FAO                 KHGQDEL-ET.....RGKPD-LNHKYVFR-NKKWNLK-I-R-KG-R.SIISTPS..AS-LL-Y....HGLEPGK-LKRGM-KF.EGQ-GKA-GSL-KE-KSVS-SIWRNTTNSTCISLKGKKI....QDRNVTGCMLKQLKVDNYTTK--LASENLMML----Y-F--VS  439
RCM.NG.x.NG411              IS-NRTW-Y.........QRKQS-RTV-GLNRHFNLTVT-R--S-R.-V-G-SL..AT-VF......ISLRVQKRPKG-W-RF.DGD-KGAW-EVKEAINKTK-YKG-RQP.........................ENITIRSVYG.--D-AKCFWL--D---L--KMN  396
RCM.GA.x.GAB1               INVN-TW-Y.........QRRQS-RTV-GLNSFYNLSVT-R--S-R.-V-G-SL..AT-VF......ISLRVEKRPKG-W-RF.EGN-TDAW-EVKERVKTTK-YRG-SNT.........................DKIKIRTVYG.--D-ARYFWL--N---L--KLN  402
MND_1.GA.x.MNDGB1           MHK.-GELIPIDDKYRGPEE-HQRKFVYK-PGKYGLK-E-H-KG-R.SVVSTPS..AT-LL-Y....HGLEPGK-LKKGM-TF.KGR-GLA-WSL-KE-NKLNDSI-VNQTCKNFTSTGEENKQNTDKQKEFAKCIKTLKIDNYTT-.--RAAEMMMMT-Q--M-F--V-  423
MND_2.x.x.5440              KHDYDEL-QT..NPRKGKDE-H-HKYVYR-DKKWGLQVR-R-KG-R.SIISTPS..ST-LL-Y....HGLEPGK-L-KGK-QL.EG--GQA-HSLSLE--KINDSIYKDNH.NMTCKSSNNK.......KNTTGCHLKTISISESTVK-E-GAE-IMLL----Y-F--W-  439
MND_2.CM.98.CM16            KHDYNEL-QT..NTKKGKEE-H-HKYVYR-DEKYNLQVV-R-KG-R.SVISTPS..AT-LI-Y....SGLEPGK-LKKGM-QL.KGQ-GIAMHDL-IE--KIDSSIWRNVT.KGYCKGLKRK.......ENRTGCALKTIKVSDYTTK-E-GAE-IMLL----Y-F--W-  418
MND_2.GA.x.M14              KHNYDEL-LV..NPQ...KE-H-HKYVYR-NKKWGLKVV-D-KG-R.SIVSVPS..AS-LI-Y....HGLEPGR-LKKGM-RL.IGQ-GRA-NALSRE-KKVNASIYRGFNASGPCRGINNR.....GVVNKTGCALKTIEVSNYTTI-E-GAE-IMIL----Y-F--W-  418
DEB.CM.99.CM40              DE-RAEELHI..IRKEVKGEVQNGSI--R-PAKYNLTLT-V--G-K.-YRA-HM..AT-LS-Y....-TFIQRLRIKR---RL.NGS-A-AT-EMRQ-IL-I--KANRTNN.............................LTIHYPK--R-VQSVW-Q-H-------IS  400
DEB.CM.99.CM5               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-VQNTW-Q-H-------VS  382
DEN.CD.x.CD1                VELLDNHTQI.....IRNTSKDNQTIA-LFRKSSVIKLK----G-V.-L-GLNM..AA-IVMR....ATIIPRRAI-R-A-K-VGGN-TR-RQE-EKAVK-E-NI-G-VKW..............................DVPRV-----TSSAR-Q-R-------L-  396
TAL.CM.01.8023              K..-NQTQII.......HQAPKNESAI-LLGKEHNIT-Q-I--G-R.-I-DLQ-..AA-LM-H....SQIIAGKDLKR-Y-K-.EGN-TRAMEDVNE-IAKAYKDHINSSS.........................GTAQIRWQGVRK--L-VSLFWHA-Q-------IS  401
TAL.CM.00.266               K..-NHTQVI.......HQAPKNESAI-LLGKEHNLT-T-I--G-R.-I-DLQ-..AA-LM-H....SQIIAGK-LNR-Y-KL.EGN-TKAMESANEEIAKRYKEHINNTD.........................KVNITWTAVRR--L-VQLFWHA-Q-------IS  384
MON.CM.99.L1                GNAPNTTVM........MNKQKNESIVVRLAKHLH-N-T-I--G-K.-IRNLQ-..-A-MT-Y....SQLIVG--T-K-Y-KVNKTQ-ETA-QAVHEAVKTEWEKKNNGTN.........................VTTISWRFQPQ--K-VQ--W---Q-------VS  390
MON.NG.x.NG1                S-TNXTTVMM........NGKRNESIV-RLXKTL--N-T-I--G-K.-LRNLQ-..-A-MT-Y....SQIIVG--X-K-Y-QVNKTQ-DAA-HAAHXAIXXEWESKNNNVT.........................VANITWRYQPK--N-VQ--W---Q-------LS  401
SUN.GA.98.L14               RHK-DEL-PI....NNKVGQGAEGEYVWK-AAKWGLV-Q-I-KG-R.SQVSTIS..ST-LL-Y....YGLEPGSKL-L-Q-KF.VGQ-GRAFATLGKL--QVEPLANMSNG....CTFDNITKT..CRFTNGTDFKKMIKFKPMEKH-A-AATEMLMMT--E-M-F--L-  432
WRC.CI.97.97CI14            RHK-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-RATENIMMI----M-F--I-  421
WRC.CI.98.98CI04            RYKKGELVEI.IGNINHRPHEGTG-FVYK-PSRYK---E-N-KG-R.SVVSVNS..AS-LIYY....AGLEPHR-I-KGL-KF.KGL-GHM-YDLGKN-Q-YNNTWVNYTE...DCNQKN....DSREFAKCIKQ...FQIRNFTRK-A-RATENIMMI----M-F--I-  431
WRC.GM.05.Pbt_05GM_X02      KYDKGEL-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-RATENLMMI----M-F--I-  423
LST.CD.88.524               NHT-DEL-PL...TPKKMGDLNGAKFVYK-AGKWGLI-R-I-KG-R.SEVSTIS..ST-LL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFNNLGRM-KKLNETAMNYKE..GNCSTPG.........KPCGRQLKGLPIANMTRR-V-LATEMLMHT--E-M-F--V-  431
LST.CD.88.485               NHT-DKL-PL...TPKKMG-LL-AKFVYK-SGKWGLI-R-I-KG-R.SEVSTIS..ST-LL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFNNLGKM-QKLNATAMNYTN..GQCPEGK.........QPCGRHLKGLPIANMTRR-EGLATEMLMHT--E-M-F--I-  428
LST.CD.88.447               NHTDDEL-PL...TPKKMG-LLEAKFVYK-AGKWGLI-R-I-KG-R.SEVSTIS..ST-LL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFHNLGNM-KRLNATAMNYTH..GNCIDNT.........KPCGRQLKGLPIANMTRR--NLATEMLMHT--E-M-F--V-  429
LST.KE.x.lho7               NHTRDEL-PL...TPNKMEDLNGAKFVYK-AGKWGLI-R-I-KG-R.SEVSTIS..ST-YL-Y....YGLEHGSRL-L-Q-KF.EGQ-GRMFNNLGKM-K-LNAEAMNYTEGTGTCDSKK.........TTCGRKLKGLPIANMTRH-A-LATEMLMHT--E-M-F--V-  426
SYK.KE.x.SYK173             YKAKDK-RFI.......KQKDKNESVI-L-PEALRLQ-I-E--G-E.SI-N-QL..AA-YFLPVI.QGKL-TGRDAKR-F-RV.TGN-TEFF--VHEQATKTWK-.........................VTNTTWRSQP---L-VR--W-Q---------VS  384
SYK.KE.x.KE51               YKAKDK-RFI.......RQKDKNESVI-L-PEKLGLQLV-E--G-E.SI-N-QL..AA-FFLPVI.QGRL-TGKAAKR-F-KV.KGE-GKFFE-VHNESIKVWK-.........................VTSTSWRSQPQ--L-VR--W-Q---------VS  391
COL.CM.x.CGU1               KELSDWAKD......REG-WKN-SGTIEYYWFPKDIALG-I-RG-S.SHRNLNT..AN-AK-Y....YELIPYSKGIYGR-QFVPMTGQ-KKNKTQFAIEIKKNLTAWLERISRKNI....................TITPRNGNRTS---ATFTFVI-HRL-----AS  398
```

```
                         glycosylation NST
                      glycosylation NST
                                 V4                                    CD4                                                        glycosylation NES                                   fusion peptide
                              glycosylation NNT                      CD4         CD4      glycosylation NIT              V5                                    gp120 end  gp41 start
                                                                        *                           *
H1B.FR.83.HXB2            QLFNSTWFNSTWSTEGSNNT..........EGSDTITLPCRIKQIINMWQKVGKAMYAPPISG...QIRCSSNITGLLLTRDGGN......SNNESEIFRPGGGDMRDNWRSELYKY.....KVVKIEPLGVAPTKAKRRVVQ.....REKRAV..GIGALFLGFLGAA  526
H1A1.UG.85.U455           G----I-NG-MSNDM-P-G-.................---Q----------R--Q-------Q-...V---E------------TN......-TKN-T-----------K------.....-------------R------E.........-------..-L--I--------  519
H1B.US.90.WEAU160         -------HANGTWKNTEGAD............NN.------------R--E---------E-...----L------------SS....EE-QT--------N-K----------.....-----------------------------.....-----G.ML--M--------  529
H1C.ET.86.ETH2220         N-----KLELFN-STNL-................---Q----------G--R-------E-...I-M-R-----------AK.....EPHSTK-----E--------------.....---E-K--------P-----E.....-----A...L----------  514
H1D.CD.84.84ZR085         G----A-NI-GH--GLNDTI................--I-----------E----------E-...--N---------------A......N-TQNDT--------------------.....-----------------------E.....-----I..-L--M--------  526
H1F1.BE.93.VI850          G---D-GS-NGT.......................-----------V----G--R---TS--A-...N-T-N-----------------......ESNI-T---E--N-K----------.....---E-------------Q-------.....-----A..-L---------DS  502
H1G.SE.93.SE6165          G----SLLR-NS-EN-T...................-----K----VR---R--Q-------A-...N-E-N-S----I-------N..NNTNTS-------------------------.....-T---KS------R-R----E.....--------..-L--V--------  529
H1H.CF.90.056             G----S-EMH-NY-SNDTKG............NEN----------V----R--R-------Q-...N-M-V------I--I-E--......ASA-NYT---------------------.....---------------I----TR----E.....--------..-M--S--------  515
H1J.SE.93.SE7887          K----S-DKNSIEATNDTSX............AT.--I--K----VR---RT-Q-I-----A-...N-T-T--------------R..GNG-E-GT-T---T--N-K----------.....---E--------------------E.....--------.----V------T-  518
H1K.CM.96.MP535           K---E-GE-G-......................-----------------------I-----A-...S-N-------MI------......N-THN-T--------------------.....---Q-----I---R-R----------.....--------..-L--V-F------  512
H101_AE.TH.90.CM240       K---N-CLGNETMAGCNDT...............-----K--------GA-Q-----------...R-N-V-----I-------VN......-TDN-T------NIK------------.....---Q-----I---R------E.....--------.----MIF------  520
H102_AG.NG.x.IBNG         K------D--NSTANHTGSN............DT.---Q------V-------Q-------Q-...I---D---------------......-STN-T--------------------.....-------------------R------E.....--------..-L--V--------  517
H103_AB.RU.97.KAL153_2    K------NGTEELNNTEGD...............IV--------------E----R-----A-...----------------------......QS-VT----T-----------------.....-----------------------R-----------.....--------.----V--------  510
H104_cpx.CY.94.CY032      P-----HMQNGTNITSTDS-............NS----Q--L--FVR---E--Q----S--A-...S-N---D---II------T.......-NT--------------------------.....------------I----N--R---------.....*-.....----M--------  524
H1N.CM.95.YBF30           K---EELL-E-GEP.....................-------R--V-L-TR---GI-----R-...VLN-T-------V-EYS--P.......DTKET-VY-S--N-VNL--Q-------.....---S---I----G-----T-S.........-AF.-L----------  514
H1O.BE.87.ANT70           KM--Y-F.SCNGT-CSVS-V..........SQGNNG----KLR-VVRS-IRGQSGL-----K-...NLT-M-----MI-QM-NTWNS....S-NNVT---I----K-I--T--FN-.....---RVK-FS----RIA-P-ISTRTH.-------..-L-M----V-S--  521
H1O.CM.91.MVP5180         GM--Y-FI-C-K-GCQEIKG........SN-TNKNG-I--KLR-LVRS-M-GESRI-----P-...NLT-H-----MI-QL-QPW......NSTGENTL--V----K-I--TK--N-.....---Q-K-FS-----MS-PIINIHTPH-------..-L-M----V-S--  531
CPZ.CD.90.ANT             PW--A-YT..GNLITNGAL..................IAH------V-H-GI-S-GI-LA-RR-...NVS-T-S---IM-.EGQIY.........--TVKVS-AARVA.-Q--A--SR-.....Q--E-X--S----TX--PEIKQHS..-Q--GI.----LF---L-S--  519
CPZ.CM.01.SIVcpzCAM13     KM--.........ES-PFHE................NM-I----R--V-S-MR--RGI-----P-...H-T-N-L----I---.DHV.........-TNNT---I----KNI----------.....---R----S---------.HTVG..ER-Q---AF.-L----------  527
CPZ.CM.05.SIVcpzEK505     ---TNQN.........-NTTG.................N---Q---R--V-L-TR---GI-----K-...P-N-L-----II-D.........YTK-GT-KYTIY-T----TNL--Q---------.....---S---I----G----HT-T.......-Q---AF.-L----------  497
CPZ.CM.05.SIVcpzMB66      I---TSI.....IFNETKDD...............N--I----R--VRL--R--RGIFL---R-...T-N-I-----I-FA.....QQ.KTDRM-K-AM-T-V--E--N-----------.....---R-----------------T-H.........-----AV.-L----------  540
CPZ.CM.98.CAM3            T---Y--K-NNITKGDNTT..................FF----R--V-S-MR---GI-----R-...V-S-T--V--II-ETGH-I.....NNSITNITLY-T--N-V-L--L--H--.....---S---I----S----HT-S..........-----AF.-L----------  521
CPZ.GA.88.GAB2            R-ITCNSSDTSE......................YI---K-R-VV-S-MR---GIF---RR-...T-T-N-T------EVQN.........GTGNNTEVYLS------I----------.....-I--------------YT-AKA.LD-S---AF.-L----------  517
CPZ.CM.98.CAM5            P---KFNYTCDNQNNCTR-N......TNDTNSDNS-L-IS---R--V-S-MR---GI-----R-...N-T-I--------EAG-KWE......N..NTITVY-S--E---L--L--NR-.....---S---I------------HT-A.........-----AF.-L----------  527
CPZ.CM.05.SIVcpzMT145     EII-I-K.......INKTEN...............MTII----R--V-S-MR---GIF----R-...N-T-T-----M--EIHKNREDQGEDQDQNNTYVCLT--N-K-I-----------.....-I-E-Q--------SR-YA-EKQHH.-----L..-L----------  514
CPZ.TZ.00.TAN1            NWT-TWTA.NRTNNTHGTL..................VA---LR--V-H-GI-S-GV-L--RR-...TVK-H-------IM-AEKD-N.........NSYTPQFSAVVE-Y-KV--AR-.....---E-Q--S---RPG--PEIKANH.T-SR-D..V---L------S--  517
CPZ.CM.05.SIVcpzLB7       A---D-LLRNKTLLNETLHN...............NL------R--V-L-LR---GIF----R-...N---N-T----I-E.........KHT-TDNIT---I----T-I------N-.....-I-----------R-R--T-D.........-----A..-L-VF--------  508
CPZ.US.85.CPZUS           P---M-YN....N-DNTTI..................--K---R--V-Q-MR---GIF----K-...VLS-N-----MI-DISISAV...NNDSRNITVM-T----TAL-KN--H--.....---S---I----G----HT-K.........-----AF.-L----------  498
CPZ.GA.88.GAB1            -I-TDNIT-GII.......................I----R--VSS-MR--RGI-----R-...N-T-N---------S-TPVT.....N-SGNLT---T--N-K-I----------.....---R----S-------R-HT-ARQ.KD-Q---AF.-L----------  516
GSN.CM.99.CN166           I--QLNNN.TIN-S.NIG-..........ITSKYKGQW-A-K-R-FVTQ-GY-S-SI-L--RQ-...H-N-T--------IDGAMYES..........S-NMTPSA-VA-A-NY--SR-.....---E-D--SM---P---KEHP....AV-----..LSL-IS--T--S--  538
GSN.CM.99.CN71            V--QFNNNITAVDDTNI--..........VTSKYKGQWMA---R-FVTQ-GY-SRSI-L--RQ-...H-N-T--------IDGAMYGH..........S-NMTPSA-VA-A-KY--SR-.....---E-D--SL---S-Q--PHPGV..HVK----..ISL-IS--T--S--  547
MUS_1.CM.01.1085          V--QFVNHTEINE.........SNIHTVTTKYHKDQWMV---R-FV-Q-GY-S-SI-L--RK-...LV--E--------IDGAMYRD..........S-NMTPSANVL-A--M-PK-VQ.....---E-D--SM---EVQ--P-P....HVHA---..ISL-IT-----S--  550
MUS_1.CM.01.CM1239        V--QFNNNITQINE........SNIHNVTTKYHQDQWMV---R-FVTQ-GY-S-SI-L--RR-...HVQ-T--V----IDGAMYKD..........TVNMTPSANVM-A--M--KN-.....---E-D--SM---EVQ--P-P....HVKA---..ISL-IT-----S--  544
MUS_2.CM.01.CM1246        V--QFENNTREIND........SNIHTVKTKYRHDQWMV---R-FVTQ-GY-S-SI-L--RK-...HVQ-T-HV----IDGAMYGN..........T-NMTPSANV...-TRASRLRD....T--E-D--S----EVR--P-P....HVKA---..ISL-IT-----S--  546
MUS_2.CM.01.CM2500        V--QFENDTRQIHN........SNIDTVTTKYHKDQWMV---R-FVTQ-GYTS-SI-L--RQ-...H-Q-T--V----IDGAMYGD..........TVNRLTPSANVL-A--M--KR-.....---E-D--S----EVQ--P-P....HVQA---..ITL-IT-----S--  532
                                                                                                                                                               gp120 end  gp41 start
                                                                        *                           *
MAC.US.x.239              WFL-WVEDRN-AN..............QKPK-QHKRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANI-WIDG.........NQTNITMSAEVAELY-L--GD-......-L-E-T-I-L---DV--YTTGG..TS-N--G-.FVL-..------AT-  539
H2A.GW.x.ALI              WFL-WVENKTGQEQ.....................HNYA--H------I-H-A--NV-L--RE-...ELT-N-TV-S-IANI-TDG.........NQTNITFSAEVAELY-L--GD-......-L-E-T-I-F---SER-YSST...PR-N--G-.FVL-..------AT-  522
H2A.DE.x.BEN              WFL-WVEDKNQTRR......................NY..-H------T-H----NV-L--RE-...ELA-E-TV-SIIANI-IDKN.........RTHTNITFSAEVAELY-L--GD-......-LIE-T-I-F---DQR-YSST...PV-N--G-.FVL-..------AT-  524
H2A.SN.x.ST               WFL-WVENRTNQTQ.....................HNYV--H------T-H----NV-L--RE-...-LT-N-TV-SIIANI---E.........NQTNITFSAEVAELY-L--GD-......-LIEVT-I-F---PV--YSSA...PV-N--G-.FVL-..------TT-  520
H2B.GH.86.D205            WFL-WVENKTNTTR.....................RNYA--H-R----T-H----NI-L--RE-...ELS-N-TV-S-IANINSD..........N-TTNISVSAEVSELY-L--GD-......-L-E-T-I-F---DVR-YSSV...KP-N--G-.MVL-..------AM-  527
H2B.CI.x.EHO              FFL-WVENRTGLKR......................NYAS-H-R--V-T-H-I-RNV-L--RE-...ELS-N-TV-S-IANI-WIDK.........NLTNITVSAEVSELYKL--GD-......-L-E-T-I-F---SI--YSSV...TP-N--.GVLVL-..------AT-  523
H2G.CI.92.ABT96           WFL-WIEDRNMTLL..............KPQ-RQRRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANIEENRSN........NHTNIIFSAEVAELY-L--GD-......-L-E-T-I-F---NV--YSSV...TPKN--G-.FVL-..------AM-  540
H2U.FR.96.12034           WFL-WVENR-MEGQ..............TSR-RHKRNYV--H-R-V--T-H---RNV-L--RE-...DLT-N-TV-SMIANI-RI-K.........NQTNISMSAEVAELY-L--GD-......-L-E-R-I-F---PV--YSSV...TP-N--G-.FVL-..------AT-  534
MAC.US.x.251_1A11         WFL-WVEDRN-AN..............QKPK-QHKRNYV--H-R----T-H----NV-L--RE-...DLT-I-TV-S-IANI-WTDG.........NQTNITMSAEVAELY-L--GD-......-L-E-T-I-L---NVR-YTTGG..TS-N--G-.FVL-..------AT-  538
MAC.US.x.251_BK28         WFL-WVEDRDVTT..............QRPK-RHRRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANI-WTDG.........NQTSITMSAEVAELY-L--GD-......-L-E-T-I-L---DV--YTTGG..TS-N--G-.FVL-..------AT-  541
MAC.US.x.EMBL_3           WFL-WVEDRDVTTQ..............RPK-RHRRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANI-WTDG.........NQTSITMSAEVAELY-L--GD-......-L-E-T-I-L---DV--YTTGG..TS-N--G-.FVL-..------AT-  541
SMM.US.x.H9               WFL-WVEXIKNG-..........RWTSQNQK-RYQKNYV--H-X----T-H----NVXL--RE-...DLT-N-TV-S-IAEI-WI-G.........NKTNITMSAEVAELY-L--GD-......-L-E-T-I-F---SV--YTTTG..AS-N--G-.FVL-..------AT-  546
SMM.US.x.PGM53            WFL-WVEDRN-SGS.........LWTQQTKADEKKRNYV--H-R----T-HR---NV-L--RE-...DLT-N-TV-S-IADI-WIDG.........NQTNITMSAEVAELY-L--GD-......-L-E-T-I-L---SVR-YTTGA...S-N--G-.FVL-..------AT-  548
SMM.US.x.SIVsmH635F_L3    WFL-WVEDRDQN-N.........RWKQQKKR-QQKRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IAEI-WI-N.........N-TNITMSAEVAELY-L--GD-......-L-E-T-I-L---DVR-YTTTG..AS-N--G-.FVL-..------AT-  547
SMM.SL.92.SL92B           WFL-WVEDRN-S-P.........RWTTQTKK-QHKRNYV--H-R----T-HR---NV-L--RE-...DLT-N-TV-S-IANI-WIDN.........N-TNITMSAEVAELY-L--GD-......-L-E-T-I-M---HV--YTTS...TSKN--G-.FVL-..------AT-  546
STM.US.x.STM              WFL-WIENR--SE.........MRDWNKNKK-QQKRNYV--H-R-V--T-H----NV-L--RQ-...DLT-N-TV-SIIANI-WT-N.........N-TNITASAEVAELY-L--GD-......-L-E-T-I-L---NV--YTTS...TS-T--G-.FVL-..------AT-  542
MNE.US.x.MNE027           WFL-WVEDKNLTG............TTQKPQ-QHKRNYV--H-R----T-H----NV-L--RE-...DLT-N-TV-S-IANI-WIDG.........NQTNITMSAEVAELY-L--GD-......-L-E-T-I-L---NV--YTTGG..TP-N--G-.FVL-..------AT-  542
GRV.ET.x.GRI_677          WFI-YLNNRTEDAE..GT-RTCDKGK.PGPGPCVQRTYVA-H-R-VV-D-YT-S-KV----RE-...HLE-N-SV-A-YVAI-YN-KS........GP-NVTLSPQV-SI-AY--GD-......-L-E-T-I-F---DVR-YTGPT....----VPF.VL-..---------  536
VER.KE.x.TYO1_patent      WFL-YLNNKTWDAD..H-FCSSKK.KGHAPGPCVQRTYVA-H-RSV--D-YTLS-KT----RE-...HLQ-R-TV--MTVELNYNSK........NRTNVTLSPQIESI-AA--GR-......-L-E-T-I-F---EVR-YTGG...HE-Q--V.PFVL-..---------  549
VER.KE.x.9063             WFL-YLNNQ-VDAN..H-QCSNETKKGNAPGPCVQRTYVA-H-RSV--D-YTIS-KV----RE-...HLE-T-TV--MTVELNYN-N........YRTNVTLSPQIEGI-AA--GR-......-L-E-T-I-F---NVR-YTGG...QD-S--V.PFVL-..---------  557
VER.KE.x.AGM155           WFL-YLNNLTVDAD..H-HCKNNAGKGRSPGPCVQRTYVA-H-RSV--D-YTIS-KT----RE-...HLQ-T-TV--MTVELNYN-Q........NRTNVTLSPQIETI-AA--GR-......-L-E-T-I-F---EVR-YTGG...QE-Q--V.PFVL-..---------  552
VER.DE.x.AGM3             WFL-YLNNRTVDPD..H-PCNGTKGKGKAPGPCAQRTYVA-H-RSV--D-YTLSRKT----RE-...HLQ-T-TV--MSVELNYNSK........NRTNVTLSPQIETI-AA--GR-......-L-E-T-I-F---EVR-YTGG...HD-T--V.PFVL-..---------  557
TAN.UG.x.TAN1             WFV-WLNNE-NSGR..NVDVEGNNCTTGKDKRCYKRTYV--H-RS-V-D-YTLS-KT----RE-...HLE-T-TV-SMMVSL-YNSK........ERTNVTLTANLENI-AY--GR-......-LIE---I-F---EIR-Y-GPT....----VPF.VL-..---------  547
SAB.SN.x.SAB1C            WFL-YLNNK-VDPD..H--CAKNN.....TKPCWQRTYV--H-R-VV-D-YTLS-KT----RE-...HLE-N-TA-A-YVELNYNSK........NRTNVTLSPQIESI-AN--GD-......-L-E-K-I-F----VR-YTG....PE-Q--V.PFVL-..---------  541
OLC.CI.97.97CI12          KILKTWNNA-DPKW.....................YPWSQ--LR-NVDD-M---RRI-L--QP-FNNR---TQRV-EMF-EIQKVEN........GTQVSLLPPS-NVNQFV-VASD-......-L-EL-LI-F---ETQ-Y......QP-N--G.ALVL-..L--V--L-  546
DRL.x.x.FAO               KIWRTWNNRTSNVW.....................YPYAS-H-----DD-AR--RKI-M--V--FNNE---TQDV-EMFFEIQKLETE..ETGDSNYQ-KFIPQDEVQNQFTAVGAH-......-L--V--I-F---DVH-YNLPG...H-Q--GA.VVL-..I--L-SL-  575
RCM.NG.x.NG411            WFL-YLNNKTEGNK................N-ARQAMFV--IT-MVV-D-YT-SRKV-T--RPD...AL--NATV-Y--ADI-YTDN........MTTNVTLSAEVG-I-AA--GR-.......---E-M-I-Y---NVR-YET......KQ--VP.LVL-..------S--  524
RCM.GA.x.GAB1             WFL-LLNNETVG................TTN-KRKAPFV--IT-M-V-D-YT-SRKV-T--RPD...ALK--AQVSY--ADI-YI-D..........--TNITLSAEVG-Y-AA--GR-......-AIE-R-I-F---EI--YQT......KQ--VP.LVL-..------S--  530
MND_1.GA.x.MNDGB1         RIMRAWNDPNEKKW.....................YPYAS-Q-R--VDD-MQ--RKI-L--T--FNNH---THRV-EMYFEMQKIDSNE.....TKMQ-KFLPPSETSNQFVAYGAH-......-L---M-I-I---DV--HTLPE..HHK---GA.VIL-..I--L-SL-  557
MND_2.x.x.5440            KIWKAWNSKQSSVW.....................YPYMSRN-R--VGD-H----KI-M--V--FNNE---TNDV-EMFFEVQKTDDD......-GYI-KFIPQDWIQNQYTAVGAH-......-L--VD-I-F---DIH-HHLPN...T-Q--GA.VLL-..M--L--L-  571
MND_2.CM.98.CM16          KIWRTWNDQNSSVW.....................YPWMS-N-R---DD-H----KI-M--V--FNNE----NDV-EMFFEVQKTEEG........YI-KFVPQDEVQNRFTAVGAH-.....QL--VD-I-F---EVA-YHLPE...A-Q--GA.VLL-..MF-L--L-  548
MND_2.GA.x.M14            RIWKTWNNQNSNVW.....................YPYMS-K-R--VDD-H---RKI-M--V--FNNH---TNDV-EMFFEVQLVEG.......KRYL-KFLPQDEVQNQFTAVGAH-......-L--VD-I-F---DVH-YHLPD...AKQ--GA.VLL-..M--L--L-  549
DEB.CM.99.CM40            KALDLLLLQNNT................RNSTW--KWLM----N-LV-T-YT--QHI-L--KE-...ELK---H-SAFVFDV-HY-G..........S-TLTPSA-I-AV--AD-F--......-IIEVK-I-F--SAVR-YEGPE..SV-H---AGIAF-..LVA--ST-  532
DEB.CM.99.CM5             KA-..NLLMNKT................NNITW-NNTLM--K-N-LV-T-F---QHV-L--KE-...ELK-H-SVSA-VFEVAHYGQ..........NMNITPSS-IKAV--AD----......-IIEVK-I-M---SIR-YEGPE..SA-Q---AGIAF-..LVA--ST-  512
DEN.CD.x.CD1              RM---SQYPVNN.................SNIINNQTYMV-N---LV-T-HT--MDI-L--AE-...E-T-T--VS-IIVDT-VHPN........RTH-TLTLSA-V-QV--A--ARW......-L-ELT-I-L---SVR-YEGPESAESVSR-RREVSLVLGLI---S--  533
TAL.CM.01.8023            SM-MK-NYTYNYTA-KGVLHQEKIGRQEHRYSDRTKFFMS-QLR-M--R-S--E-L--L--RE-...HLS-K--V--I-VDINHYKGN........DI-SVEPSSQI--A--AAVAR-.......--IE-T-I-F---NIR-YDGP...ATK----APLAL-..-I---ST-  550
TAL.CM.00.266             SM-LN-SYEYNYTSNQGKLHQGVIGK.KPRSSDRTKFYMS-QLR-M--R-SR-E-L--L--RE-...HLQ-K-------VDINHHRGS.........-D-TVEPSSQI--S--AAVSR-.......--IE-V-F-Y---NIR-YDGP...ANKQ---APLAL-..-I---ST-  532
MON.CM.99.L1              A--INRRT-K-DG.......ISPFDVNNKPNTTYHGGW-A-T-R-MVTQ-GY-S-SI-L--RK-...HVQ-T----A--I-GELYQN..........NVTLVPSAQVS-S-----SR-.........---E-D--SM---T-Q--TGV....H------..ITL-MA-----ST-  529
MON.NG.x.NG1              R--X-SDPXKNKN....ETFESTFNVSTNASTAYNGEW-T--LR-LVTQ-GYIS-SX-L--RQ-...HVQ-T----A--VDGELYGS..........TVTLIPSARVS-A-----S--.........---E-D--S----A-Q--HGL.....MRE-RA.ITL-MA-----ST-  543
SUN.GA.98.L14             RI-KVWNDTTSNKW.....................YPWAN-H--SV-DD-AS---KI-L--T--FNNR---ANRV-EAWF-LERVEDW.KVNGS-I-VVAFQPPTNTLNQFV-TGAH-......-L-R-R-I-F---DEH-YAP.....RK----APVAL-..A-AL-S--  568
WRC.CI.97.97CI14          RIADAWNNR-ETKW.....................YPWS--T-RN-VED-AH--IKV-L--V--FNNH---.....-QRVTEILFNWE........RNQVMFQAPSNIQNSFVAVGQ--......-L--LKMI-M--GDG--V-PPE..HR----G.AIIL-..I-----L-  548
WRC.CI.98.98CI04          KIADAWNNKNESKW.....................YPWS--T-RN-VDD-AH--VKV-L--V--FRNH---.....-QRVTEILFNWE........NKQVEFQAPSNIQNSFVAVGQ--......-L-RLKMI-L--GDGR-VTPKE..HH----G.AIVL-..I-----L-  558
WRC.GM.05.Pbt_05GM_X02    KIADWWNNKTEKRW.....................YPWA--T-RS-VDD-ASI--KI-L--V--FNNH---.....-QRVSEVLFNWE........NGQVEFQAPSH-QNSFVAVGQ--......-L--LKLI-----DG--VTPPD..HH----G.AIVL-..L-----L-  550
LST.CD.88.524             RI-QEWNNKNSDKW.....................YPWAN-H--S--DD-AT---KI-L--T--FNNR---THRV-EMWFEMEKWEPH..EDLGGNLS-KFLPPSWETNQFVATGAH-......-LIRLR-I-F---SEH-YAP....RG-QT--APLAL-..A--L-S--  567
LST.CD.88.485             RILQEWNNKNSDKW.....................YPWAN-H--SV-DD-AT---KI-I--T--FNNR---THRV-EMWFELEKWEPE..ENLGGNLSVRFLPPSWETNQFVATGAH-......-LIR-R-I-F---AEHQYAP....SSKQ---APLAL-..A--L-S--  564
LST.CD.88.447             RI-QEWNNKNSDKW.....................YPWAN-H--SV-DD-AS---KI-L--T--FNNR---THRV-EMWFQMEKFEPE..ENLGGNLSVRFLPPSWETNQFVATGAH-......-LIRLR-I-F---DVH-YAP....SSKQ---APLAL-..A--L-S--  565
LST.KE.x.lho7             RI-QEWNNKNSDKW.....................YPWAN-H--S--DD-ATI--KI-L--T--FNNR---THRV-EMFFEMEKWEPH..EDLGGNLS-KFLPPSWETNQFVA-GS--......-LI-LN-I-F---DEH-YAP....RG-QT--APLAL-..A--L-S--  562
SYK.KE.x.SYK173           K--ANITNGNASKN................NYASNLR-S-A-R----D-RY-R-LI-L--TA-...H-K-T--V-AV-TDIEYYPG.........-TLNFTPTANVE-V--AD-FN-......-LIQ-K-I-F---DQR-YELP...NT-----APLAL-..----L-S--  515
SYK.KE.x.KE51             KI-ANVSNGYANP................SNYAKNLY-S-A-R----Y-GY-T-L--L--TE-...H-K-T----AV-TDIEYYPN.........-QLNFAPTANVE-V--AD-FN-......-LIR-K-I-F---SQR-YELP....TKQ---APLAL-..----L-S--  521
COL.CM.x.CGU1             S-WKHDS..........................PVMN-T-RKLV-S-VTHARIL-G--PG-...HLQ-NWEKQPVIAFMGTIEG....DNDGNGCAYPAAPNFKHALSTL--GR-......-L--MRTTTYV--DI--S-NVNWHHG-Q--...GIFAFSI-AL-SG-  524
```

```
                                        fusion peptide                                                                          glycosylation NAS            glycosylation NHT
                                                                                             immunodominant region            glycosylation NKS                        glycosylation NYT
H1B.FR.83.HXB2            GSTMGAASMTLTVQARQLLSGIVQQQNNLL.RAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKS..........LEQIWNHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWAS.LWNWFNITNWLWYIKL  684
H1A1.UG.85.U455           --------I-----------------S---.----------K------------V-------Q-------------------T----S-----..........Q-D---NM--LQ-EK--SS--GI-YQ------------LD--A-----N.-*-----S------R-  676
H1B.US.90.WEAU160         ---------------L--------------.---------FE------------V-------------------------T---------R-..........QDY---NM-----E-------G--YN-----------------------.--T--D-S-------I  687
H1C.ET.86.ETH2220         --------I-----------------S---.K-------M-----------T-V--I--H-R-------------------------S-----..........Q-E--DNM---Q-----S---DI-YN-L-V-----D---KD--A----EN.---------------I  672
H1D.CD.84.84ZR085         ----------------V-------------.-----------------------------------------H---T---S---R-..........V-Y--GNM---Q-E---D---G--YN-----I------K----------.-----S--Q------I  684
H1F1.BE.93.VI850          REH-----I---------------------.---------------------V-------------------------N---S-----..........Q-E---NM-----EK--S--SNI-YK------------------A------.-----D-S-------I  660
H1G.SE.93.SE6165          --------I-----V-----------G---.---------------------L-------------------------N---V-----..........YNE--DNM--I--E-------YQ-Y--L------------D--A--Q---.-----G--R------I  687
H1H.CF.90.056             --------I-----------------S---.---Q-R--M------------V-------R-----------------N---S-----..........QSE--DNM------KQ-S---EE-YR-L-V--T-------D--A------.--T--D-SH------I  673
H1J.SE.93.SE7887          --------I-----V-----------S---.K---------K-----------V-------------------------N---------..........Y-D--ENM--IQ-E-------GI-Y-----A-----N--KD--A----TN.-------S-------I  676
H1K.CM.96.MP535           --------I-----------------S---.--------------------R--------------------------N---S-----..........W-E---NM-----EK--G--SDT-YK------T------D--A------.-----D--K------I  670
H101_AE.TH.90.CM240       --------I-----------------S---.---------------------V---------KF--L------I---------ST---R-..........F-E---NM--I--E---S---NQ-YEILT------DR--KD---------.-----D---------I  678
H102_AG.NG.x.IBNG         ------R-I-----------------S---.K---------K-----------V--L---R------------------T---S----T..........FND--DNM--IQ-EK--S---DI-YN-------R------D--A------.-----D--------RI  675
H103_AB.RU.97.KAL153_2    --------I---------------------.---------------------V--------------------------T-----P..........-DE-*-NM-----E-------G--YN--------------I-A------.-----D-SK------I  667
H104_cpx.CY.94.CY032      --------------------------S---.----------R-----------V--L-S--------------------N---S-----..........YND--DNM--LQ--K------QI-YG-L------------D--A------.-----S--K------I  682
H1N.CM.95.YBF30           --------I------T----------I--.------------SI--------KV--I----R---I-SL------T--Y-T----ET---NTS........YDT---NL--QQ--EKVR--SGV-FG---QA-E--NT--KS-----Q-D-.--S--G--K------I  673
H1O.BE.87.ANT70           -------AT--A--THT--K------D---.---Q---Q--R-SX---R--R--L--L-TL-QN----SL---K---V-Y-S-K--RT-IG..........N-S--DTL--Q-----Q-S-IS-T-YEE-QKA-V---Q--KK-----E---.I---LD--K------I  677
H1O.CM.91.MVP5180         -------ATA---RTHSV-K------D---.---Q------R-S----R--R--LQ-L-TLIQN--R-NL---K-----Y-S-K--T---G.........RYNDDS--DNL--QQ--QH---VS-I-YDE-QAA-D-----VKA------E---.-----D--K------I  690
CPZ.CD.90.ANT             --------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-----S-.-----D--Q------I  686
CPZ.CM.01.SIVcpzCAM13     -------AV---------------------.-----------------V------L------Q---I--L------S--Y-T----KT--G.........KSMSD---NL--QQ--KL-T---GT-FG-L--A-S------KD-----Q---.-----D---------I  685
CPZ.CM.05.SIVcpzEK505     --------I-------K-------------.------------S----------V--I----R---I--L------SV-Y-N----TT--.........NNSYDT--GNM--QN--EQVR--SGV-FG-L-QA-E--SI--KS-----Q-S-.-----D--K------I  656
CPZ.CM.05.SIVcpzMB66      --------L-----------------S---.------------S----------V-------------L-----------S----TT-T-.........KSYDD--YNM---Q--K-VS---DV-YN-L-KA-T---N--K----------.-----D--S------I  698
CPZ.CM.98.CAM3            --------VV--------------------.------------S----------V-------R---I--L------A--Y-T----NT--AN.........TSFDE---NL--QD--KRVK--SGV-F----QA-E--NT--KS-----Q-S-.-----D--R-------  680
CPZ.GA.88.GAB2            --------VM------N-------------.------------S----------V-----------V-AL------TV-YST----T--NSNK........SY-D--SNL--QQ--KLVE-H-GT-FQ-LQRAHE--NS--K------Q-S-.-----D---------M  676
CPZ.CM.98.CAM5            --------VV------H-------------.------------S----------V-----------I-SL------A--Y-T-----T--AN.........TSYDE---NL--QD--KKVK--SGV-F----QA-E--NT--KD-----Q-S-.--S-----Q------I  686
CPZ.CM.05.SIVcpzMT145     --------VV--------T-----------.------------S----------V-------R------L---T--T--P---R--KT-G-I.........SDYQV---NY--QQ----V----G--YT-L--ANT------K------S--N.--S--D---------M  673
CPZ.TZ.00.TAN1            --------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----------S-.--D--D--Q------I  685
CPZ.CM.05.SIVcpzLB7       --------L----------T------S---.------------S----------V--I-------------------------S----RT--..........KTYNE--DNM--------VR---EI-YG---QA-D---N--KK-----H-T-.-----D-SH------I  666
CPZ.US.85.CPZUS           --------VV--------------------.------------S----------V-----------I--L------T--Y-T----DT---N.........LSYDA--GNL--Q-----KVR--SGT-F----QA-E--NT--KS-----Q-S-.-----D---------I  657
CPZ.GA.88.GAB1            -------AV---------------------.K-----------SI--V------L-------Q---I--L------AV-Y-T----N--PGS........NSTDD--GNL--QQ--KLVS---GK-FG-L--A-S------RD-----Q---.-----D--K------I  675
GSN.CM.99.CN166           -T-----ATA--E-S-S--A--M---E---.--V----S---PS--------T-LSSL-K--R--TI-QA---ANRP--H-I-----T--A-..........GSLPD-ENM--QK-SMLVE-D-YT-QQ-L-QANQ--AS-LN--MK-S--D-.--S--D-SD-QR---I  695
GSN.CM.99.CN71            -A-----ATA----S-S--A------E---.--V----S---S----------LSSL-K--R--TI-QA---ANQP--H-I-----D--AK..........NSTPD-E-M--Q--SKL-E-D-YT-QQ-L-NANH--S--MND--K-S--D-.--S--D-S-------I  704
MUS_1.CM.01.1085          -G-----ATA----S-S--A------A---.--V----------S----------LT-L-KFI---A--NL---ANRQ--H-R-----D--A-..........HTQPG-ENM--QQ-S-LVD-D-TT-QE-L-LA-R---E-QHK-QK-LE-D-.--E--D-SK------I  707
MUS_1.CM.01.CM1239        -G-----ATA----S-S--A------A---.--V----------S----------LT-L-KFI---A--NL---ANRQ--H-R-----T--A-..........HTLPG-ENM--QQ-SKLVD-D-NT-QG-L-AA-K---E-QHK-QK-LE-D-.--E--D-SK------I  701
MUS_2.CM.01.CM1246        -G-----ATA----S-S--A------A---.--V----------S----------LT-L-KFV---AI-NL---ANRQ--H-S-----D--A-..........HTQPG-DNM--QQ-S-LVD-D-AV--G-L-MA-Q---E-QHK-QK-LE-D-.--E--D-SK----V-I  703
MUS_2.CM.01.CM2500        -G-----ATA----S-S--A------A---.--V-T--------S----------LT-L-KFI---A--NL---ANRQ--H-R-----N--A-..........HTQPE-DNM--Q--S-LVD-D-AT-QG-L-LA-K---E-QHK-QK-LE-DA.--E--D-SK--F-V-I  689

MAC.US.x.239              --A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS..............-TPK--NE--Q--E-KVDFLEEN-TA-L--A-I-----MY--QK-NS-DV.FG---DLAS-IK--QY  693
H2A.GW.x.ALI              --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-DV.FT--LDL-A-VK--QY  676
H2A.DE.x.BEN              --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-DI.-G---DL-S-VK--QY  678
H2A.SN.x.ST               -AA-----L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N----VT-I-K-----AQ-NS---AFRQV-H-T---VNDT..............-TPD--NM--Q--EQR-R-LEAN-SESL-QA-I-----MY--QK-NS-DV.FG---DL-S-IK--QY  674
H2B.GH.86.D205            --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-DV.FG---DL-S-IK--H-  681
H2B.CI.x.EHO              --A-----L--SA-S-T--A------QQ-V.DVVKR--E--R-----T-N----VT-I-K-----AQ-NS---AFRQV-H-T---VNES..............-KPD--NM--QQ-E-QVRFLDAN-TK-L--A-I-----MY--QK-NQ-DI.FS---DF-S-MA--R-  677
H2G.CI.92.ABT96           --A--T--L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N--T-VT-I-K-----AR-NS---AFRQV-H-T---D-LGA--T...........--PQ--NM--Q--EKQ--FLEDN-TR-L--A-I-----MY--QK-NS-DV.FG---DL-S-VK-VY-  698
H2U.FR.96.12034           --A-----L--SA-S-T--A------QQ--.DVVKR--EM-R-----T-N--T-VT-I-K-----AS-NA---AFRQV-H-T---INDT..............-TPN-DNM--Q--EEKV-YLEEN-TQ-L-AA-I-----MY--QK-NN-DI.FG---DL-S-VK-VY-  688
MAC.US.x.251_1A11         --A-----L---A-S-T--A------QL--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS..............-TPK--NE--Q--E-KVDFLEEN-TA-L--A-I-----MY--QK-NS-DV.FG---DLAS-IK--QY  692
MAC.US.x.251_BK28         --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-DV.FG---DLAS-IK--QY  695
MAC.US.x.EMBL_3           --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-DV.FG---DLAS-IK--QY  695
SMM.US.x.H9               --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-NXXDX.XG--XDL-X-IK--QY  700
SMM.US.x.PGM53            --A-----V-RSA-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-R---AQ-NS---AFRQV-H-T---PNAS..............-VPN--NM--Q--E-QVDDLEAN-TQAL--A-I-----MY--QK-NS-DI.FG---DL-S-IK--QY  702
SMM.US.x.SIVsmH635F_L3    --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-DI.FG---DL-S-IR--QY  701
SMM.SL.92.SL92B           --A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T-L-PNDS..............-VPD--NM--Q--EKKVEFLEAN-TQML--ARL-----MY--QK-NS-DV.FG---DL-S-VR--QY  700
STM.US.x.STM              --A-----L---A-S-T--T------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NS---AFRQV-H-T---PNDS..............-VPD--NM--Q--E-KVDFLEAN-TQ-L--A-V-----MY--QK-NS-DV.FG---DL-S-VR--QY  696
MNE.US.x.MNE027           --A-----L---A-S-T--A------QQ--.DVVKR--E--R-----T-N--T-VT-I-K-----AQ-NA---AFRQV-H-T---PNAS..............-TPN--NE--Q--E-KVDFLEEN-TA-L--A-I-----MY--QK-NS-DV.FG---DLAS-IR--QY  696
GRV.ET.x.GRI_677          -TA----AT-----S-H--A--L---K---.A-V-Q--Q--K--I--V-N-N--VT-L-K--E--AR-NS---AW-QV-H-T---KYNN..............TPK-DNM--L--E-Q--ALEGN-TQ-L--A---ES--LDLYQK--D-SG.F-S--SLST--G-V-I  689
VER.KE.x.TYO1_patent      -TA----ATA----SQH--A--L---K---.A-V----QM-K--I--V-N-N--VT-L-K--E--AR-NS---AW-QV-H-T-E-PWTN.............RTPD-QNM--L--E-Q-ADLE-N-TRQLVKARE-E---LDAYQK-TS-SD.F-S--DFSK--NIL-M  703
VER.KE.x.9063             -TA----ATA----SQH--A--M---K---.A-V----QM-K--I--V-N-N--VT-L-K--E--AR-NV---AW-QV-H-T---QWQN.............MTPN-QNM--L--E-Q-GELEGN-TEQLVKARE-E---LDAYQR-TS-SN.F-S--DFSK--NIL-I  711
VER.KE.x.AGM155           -TA----ATA----SQH--A--L---K---.A-VG---QM-K--I--V-N-N--VT-L-K--A--AR-NA---AW-QV-H-T---TWNN..............TPE--NM--L--EKQ-EGLEGN-TKQL-QARE-E---LDAYQK-SD-S-.F-S--DFSK--NIL-I  705
VER.DE.x.AGM3             -TA----ATA----SQH--A--L---K---.A-V----QM-K--I--V-N-N--VT-L-K--E--AR-NA---AW-QV-H-T---QWNN.............RTPD--NM--L--E-Q-SYLEGN-TTQL--ARA-E---LDAYQK-SS-SD.F-S--DFSK--NIL-I  711
TAN.UG.x.TAN1             -AA---TATA----SQ---A--L---K---.A-V-Q--QM-K--I--V-N-N--VT-L-K--E--TR-NL---AF-QV-H-T---TFNN..............TPD-DNM--Q--ESQ-TALEGN-STTLVKAYE-EQ--MDTYQK-GD-T-.W--I-DVSS-F-W--W  700
SAB.SN.x.SAB1C            -AA----ATA----SQ---A--L---K---.A-V-Q--QM-K--I--V-N-N--VT-L-K--E--AR-N----AFRQV-H-T-L-KYNN..............TPD-ENM--Q--E-Q-EK-EAN-SRIL-QAHE-EQ--LDSYQK-VS-SD.F-S--DL-K-FGWM-I  694
OLC.CI.97.97CI12          --A--TV-TI----SSH--A---R--KHM-.EM--K--E--K--I--V-N----L-G--K-----A--KQ---EFMQV-H-T----YTQ..............SPD-GND--LK-Q-KVD-L-QE-DDRLTIAYQE-QR-QY--QRFQELSD.WTK---LA--MT-V-I  699
DRL.x.x.FAO               --A-CSVATAM---SQA--T-M-E--KQ--.-LV-Q--E--K--I--V-N----LT-L-E-IG--AM-SL----FAQV-H-N-V-PNES.............VTPN-TSE-----QKRVDSISNN-TLDLQKAYE-EQ--IF--QK-GDLT-.WA---DF-W-SK---I  729
RCM.NG.x.NG411            -TA----ATA----S-H--A--L---KK--.DIV-Q--E--K-----T-N----VT-I-K--A--S--NTF--AWRQV-H-T-E-IY-Q..............TPE--KQ--L--E-N-SRLEGN-SVALQDA-E-H-R-VHD-EK-NS-GD.MLS-L-MDW--K--RI  677
RCM.GA.x.GAB1             -TA----ATA----S-H--A--L---K---.DIVKR--N--K-----T-N----VT-I-K--A--S--NTF--AWRQV-H-V---TFNK..............TPE-QKES-LQ-E-N-SYLEAN-TIALQ-A-D-H---VH--EK-SN-GD.AFS-L-LDW-MQ---I  683
MND_1.GA.x.MNDGB1         --A--SV-VA----SQS-VT---E--KQ--.KL--Q-SE--K--I--V-N--T-LTSL-N-I---A--SQ----WAQV-H-S-E-TNTS.............ITPN-TSE--K--ETRTDYLQQN-TEMLKQAYDREQR-TY--QK-GDLT-.WAS--DF-W-VQ-L-W  711
MND_2.x.x.5440            --A--SVAVA----SQT--N---E--KV--.SL-DQHSE--K--I--V-N--V-LT-L-E-VA--SR-SV----FSQV-H-S-K-PNNS.............IVPN-TSE--L----RV-SIVTNMTIDLQRAYELEQR-IF--QK-GDLNFHGLTG-DL-W--K-V-I  726
MND_2.CM.98.CM16          -----SVAVA----SQA--N---E--KV--.SL-DQHSE--K--I--V-N----LT-L-D-VA--AR-SM----FAQV-H-H---PNDS.............ITPN-TSE--L---KRVTAL-DNMTVNLQKAYELEQ--IY--EK-GD-T-.WAS--DF-W--K-V-I  702
MND_2.GA.x.M14            --A--SVAVA----SQA--N---E--KI--.SL-DQHSE--K--I--V-N----LT-L-D-VA--SR-AV----FSQV-H-N---PNES.............ITPN-TSE--L----RVTAI-NNMTIDLQRAYELEQ--MY--QK-GDLT-.WAS--DL-W--K-V-I  703
DEB.CM.99.CM40            -AA-----TA----S-S---------QE--.K-V--HGQ--T--A--VRN-NT-LT-I-K-----AK-NE---AF-Q--H-T----N-L..............EDPD-DNM--Q--EMKVA---DEWEGALQRA-E---R-VHA-QS-QD-D-.-----DLSR-F-W-R-  686
DEB.CM.99.CM5             -AA-----TA----S-S---------QE--.K-V--HG---S--A--VRN-NT-LT-I-K-----SK-NE---AF-Q--H-T----HT-..............G-PD--NM--Q--E-KVA---DEWEGALQRA-E---R-VHA-QS-TD-D-.-----DLSR-F-W-R-  666
DEN.CD.x.CD1              -TV---VGTG-A--S-D--A------QE--.--V-GHSA----S-----N-N--LT-I-K-----A--NE---AW-Q--H-S-E-SNQS........IPGLALYQPDFQNM--QQ-ETD-AALFGNVSDALKKA-A---R-VH-VQS-YD-DN.-----DLSK-F-W-R-  693
TAL.CM.01.8023            -TA---VATA----S-S---------EH--.----H-------------N-N--LT-L-K--E--AR-NS---AW-Q--Y-S----KT-T-...........STNPD-QNM--Q--EKLVD-ASDT-TV-LQ-A-E---R-VH--QK-ND-D-.--S---LSA-FRWLRI  707
TAL.CM.00.266             -TA---VATA----S-S---------EH--.----H-------------N-N--LT-L-K--E--AK-NS---AW-Q--Y-S----KT--............YTDPQ-QNM--Q--EMKVD-H-G--SQ-LQ-A-E---R-VH--QK-ND-D-.--S---LSA-FRWLRI  689
MON.CM.99.L1              -G-----ATA----S-S--A------E---.--VT---S--------V------LT---KFI---T--NA---AN-AV-H-T----N--AK..........GHFPE-DNM--QQ-SELVD-D-MT-QQ-L-AA-E--G--QH--MKPGQ-DF.-----D-SK------I  686
MON.NG.x.NG1              -G-----ATA----S-S--A------E---.--VT---S----S----------LT---KFI---T--NS---ANRAX-H-Q-L--NT-AK..........GHFPE-DNM--QQ-SMLVD-D-A--QX-L--A-E--G--AH--MK-GQ-DW.-----D-SK------I  700
SUN.GA.98.L14             -TA--LV-TI-----QAV-Q--L---KQ--.VLV-K--E--R--I--V-N----LT-L-E-VQ--S--AS---QW-QV-H-N----YN..............ITPN-TKD-------QVKM-DDN-TA-LQ-AYVTELE-QNKFKQ-QEFN..F-S-LDLSQ-FL---Y  720
WRC.CI.97.97CI14          --A--TV-TV--I-SQH--A------K---.ELV-K--E--R--I--V-N----LS-I-GF-E--LK-KQ---ELTQV-H------Q--N..............ATPN-ENE--QH-YKRTDELQAN-SAILVIAHD-EQA-MF--QK-SD-TA.WGS-LQWF-IWNWL-W  703
WRC.CI.98.98CI04          --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--N..............ATPN--NE--QH-Y-RTDELQAN-SV-LV-AH-TEQ--MF--QK-QD-TT.WGS-LQWF-IWNWL-W  713
WRC.GM.05.Pbt_05GM_X02    --A--TV-TV--I-SQH--A------K---.ELV-K--E--R--I--V-N----LS---G--E--VK-KQ---ELTQV-H-----EEA-N..............MTPA--NE--QQ-HLRTDTLM-N-SDMLVQAHA-EQS-MF--QK-TD-TQ.WGS-LSWF-IWNWL-W  705
LST.CD.88.524             -TA--LV-TI-----QVVIQ--L---KQ--.VLV-K--E--R--I--V-N----LT-I-E-----A--AS---QW-QV-H-N----YN..............VTPN-TRD--I--E-QVGSLEAN-TT-LQ-AYTTELE-RNNFKK-QDFN..F-S-MDL-T-FQ---Y  719
LST.CD.88.485             -TA--LV-TI-----QVVIQ--L---KQ--.VLV-K--E--R--I--V-N----LT-I-E-----A--AS---QW-Q--H-N-E--YN..............ITPN-TRD--I----RVGVLEAN-ST-LQ-AYTTELE-RNAFKK-QEFN..F-S-LD-LS-FQ---Y  716
LST.CD.88.447             -TA--LV-TI-----QVVIQ--L---KQ--.VLV-K--E--R--I--V-N----LT-I-E-----T--AS---QW-QV-H-N-E--YN..............ITPN-TRD--I----QVGVLEAN-ST-LQ-AYTTELE-RNAFKK-QEFN..F---LD-LS-FQ---Y  717
LST.KE.x.lho7             -TA--LV-TI-----QAV-Q--L---KQ--.VLV-K--E--R--I--V-N----LT-L-E-V-H-A--AS---QW-QV-H-N-E-TYN..............ITPN-TKD--R--ESKVAI-DKN-T--LQ-AYTTELE-QNKFKK-QEFN..F-S-LD-SH-FT-V-Y  714
SYK.KE.x.SYK173           -TA--G-ATA--L-SQT--A------QK--.E-V-------G-----V-N-N--LT-L-T--R--AI-SN---AF-Q--H---T-EKACG-N-N.......FCPKPQ-KNM--HR-EQ-VD-L-DH-DG-LR-A-E---R-VHD-TK-QE-D-.--S--DLSK-FF-L-I  676
SYK.KE.x.KE51             -TA--S-ATA--L-SQT--A------QK--.E-V-------G-----V-N-N--LT-L-T--R--AIMSN---AF-Q--H---T-QQACG-N-........RCPTPQ-ENM--HT-E-QVD-L-DH-DN-LR-A-E-----VHD-TK-QE-D-.--S--DLSK-FQ-L-I  681
COL.CM.x.CGU1             -AA--S--VA--I--QS-NGRASASS-RM-LKLV-T-SA--------V-N--V-VATI-G--EE-AK-ASI--ANMQ--R-I----KT-.............GE-DP-QNM--KQ-HERVR---DI-EADLV-AYDL--E--KK-A--GD-TN.WFSG-GLF-IFK-VLY  680
```

```
                                                                                          glycosylation NGS                                                     glycosylation NAT
                                                                                                      *
H1B.FR.83.HXB2          FIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPT.......PRGPDRPEGIEEEGGERDRD.RSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLG.......RRGWEALKYW...........WNLLQYWSQELKNSAVSLLNATAIAVAEGTDR 828
H1A1.UG.85.U455         -VI-----I------T----I--------------LA-I.......-E-LG--GR-------QGK-.------S-F--IA-----N-----------FA---A-A-----RSSLKGL-L---G---L...........----L--GR---I--IT--D-V-V---GWI-- 827
H1B.US.90.WEAU160       --------I------T----------------------A.......--------------------.--G---D-F-T---V--------L---I------AK-------.......---------C...........-------------------------------- 831
H1C.ET.86.ETH2220       -------VI----I---------------------LI-H.......------LG--------QG--.--------F--IF------------------I--AA-T-----RSSLKGLQ----T---L...........GS-V---GL---K--IN---T---V-G----- 823
H1D.CD.84.84ZR085       --------I--------------------------L--A.......----------------Q--G.T-------FS--------N----------E-I--AA-------.......---------L...........---------R------I--VD----------- 828
H1F1.BE.93.VI850        --------I----------------K------L--LI-S.......------------G---QGK-.--V---T-F---A-----N------RH---FI--AA---DRGL.......---------L...........G--TR----------I--F-T---V------- 804
H1G.SE.93.SE6165        --------I--------------------------LTHH.......Q-E---------G---QG-G.--V---S-F-P-------------------SI---A-T-----RSSLKGL-L---G---L...........----L--GR------I---DTV-----NW--- 838
H1H.CF.90.056           --------I----I---------------------LV-N.......---------T--G---Q---.--V-----F-PVV-------S----RL--------V-T-----.......---R-----L...........-------G-------ID---T----------G 817
H1J.SE.93.SE7887        ----I---I----I----A----------------LI-N.......-TEA---G----G---QG-T.--------F---A-----N-----------FV--AA-T-GT--.......L----I---L...........V--VW--G-------I----T----------- 820
H1K.CM.96.MP535         ----I---I----A-----V---------------LI--.......S--A------------Q-KN.--V---S-F---A-----N------RQ--N-I------L-RGL.......-G-------L...........---V-----------I----T------G---- 814
H101_AE.TH.90.CM240     --------I----I---------------------PSHH.......QKE---------G---QG--.--V---S-F---A------------------T--AA-T-----HSSLKGL-----G---L...........G---L--G----I--I---D-----A-GW--- 829
H102_AG.NG.x.IBNG       --------I---------T-I--------------LTHH.......Q-E-----R---G---Q-K-.--V---S-F---A------------------I--AA-T-----HNCLKGL-L--G----L...........---IS--V-------IN---TI--V--NW--- 826
H103_AB.RU.97.KAL153_2  -------------I---------G-----------R--A.......Q------------D------.T------RF------------F-I--H-------AA-------.......----------...........-------I----S--IN-IGTI-----GW--- 811
H104_cpx.CY.94.CY032    --------I----I------------------L--LI--T......Q--L---G-T------Q--S.--------F-P-------N------RH--N-----A-T-----.......I--------L...........--F-L--G-------IN-F-T----------- 827
H1N.CM.95.YBF30         A----A-I--I--ISI-IT-IA----------L--LI--.......A-------ET-GGV--Q--G.--V---S-FS--V-E---N-LI-L----T-S---LR-TL----QSLSRGLQLLN-LRTHL...........-GI-A--GK--RD--I----T---V------- 824
H1O.BE.87.ANT70         A-I---A---V-VIMI--N--KNI----Q---L-IPNHH.......QEEAGT-GRTGGG---EG-P.-W-PSPQ-F-P-LYT---TII-WT--L-SN-ASGIQKVISY-RLGLWILGQKIINVCRIC...........AAVT---L---Q---T---DTL-V---NW--G 828
H1O.CM.91.MVP5180       A-I---A-I-I-VIMII-NL-KNI----Q---L-IPV-H.......RQEAET-GRTG-----G--P.KWTA-PP-F-QQLYT---TII-WT--L-SN-ISGIR-LIDY--LGLWILGQKTI--CRLC...........GAVM---L-------TN--DTI-VS--NW--G 841
CPZ.CD.90.ANT           ------AI-----LLVLV-CLRK-----H-----IPTQN.......QQD-EQ--E-R----RK--I.-WRA-QH-FF--L-V--T-IIQWI-QIC-TC--NLWAVLQH-CR......ITFRLCNHLE...........N--STL-.TIIRTEIIKNIDRL--W-G-K--S 830
CPZ.CM.01.SIVcpzCAM13   -L-A---II----IMS-V-VIR----------L--LI-A.......Q-E-GTLGETGG----PGSG.--V---T-CWP-------N-VIW--RS-TS-AC--W-QL-K--QLLINIL-LIQ-K-TLL...........RGII---GR--RT--T---D------G----- 836
CPZ.CM.05.SIVcpzEK505   ---V-A-I--I--ISIIM-M-A----------L--LI--.......T-------RT--DA--L-NG.--V---S-F---A-E-F-N-L--L----T-C-S-LR-TL---RQNIHKGLQLLN-LRI-L...........-GIIA--GR---I--IN--DT--V-------- 807
CPZ.CM.05.SIVcpzMB66    --I-----I-------L------------------LI-A.......R-DR----E---G---P-NV.------S-F---A-N---D----L-----------L-TL--V-QTLLKGL---R---IHL...........RGI----G----T--I---DT----------- 849
CPZ.CM.98.CAM3          -----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--QT--I---D------------ 831
CPZ.GA.88.GAB2          --IL---VI-----IG-VN-IR-S-----------LI-A.......TG--EV----GGG---Q--G.--V---S-F---C---I-N-TI-L-RL---C-SVLWSLL-Q--QQVLRGL-LLR-L-SQL...........KGIG---L---RT--I---DT---I---R--T 827
CPZ.CM.98.CAM5          -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-IQI--T---DT------N---- 837
CPZ.CM.05.SIVcpzMT145   -LIV---II----C--IG-LI----K------L--LI-A.......NQ---GLGET-KG---NV-G.------S-F-P-V-E---N-LS-L--Q---CASLIW-LL-I--QYSLRGVQQIGTL-HQL...........RGT----TT-I----I---DT-----G----- 824
CPZ.TZ.00.TAN1          A-I--A-------LMFIVNV-KQ-----T--FS-IPTQA.......EQD-EQ-G--AGG--G--NI.-WTPSPA-FFSIV-E---N-LIWI-QTFQNFIW-LWISLQA-KQGIISLAHSLVIVHRTIIVGVRQI....IEWSSNTYAS-RVLLIQAIDRL-NFTGWW--L 843
CPZ.CM.05.SIVcpzLB7     ----I---IVC--I----A----------------LS-I.......---LAQ-G-T------QGT-.--V--LD-F---A-N---D----L----T--V--AR-TL-IV-QYTLKGL-LV----L-L...........QGI----GK---T--T---DT----------- 817
CPZ.US.85.CPZUS         -LIV-AS---I---GVIF-L-AK---------L--LF--.......T-E------T--GA-KT-NV.--T---S-F---V-E---N-LI-L----E-----LR-T-QI--QNINKGLQLLN-LRARC...........-GVIA--AR---V--T---DT----------- 808
CPZ.GA.88.GAB1          -L-A---II----IMT-F-V-R----------L--LI-V.......Q-EQG-LGE-D-G---Q--S.--V---E-C-P-------N-GIW--QS-TS-ACN-W-QLKT--HLILHSL-LLR-R-CLL...........GGII---GK---I--I---D------------ 826
GSN.CM.99.CN166         -VIV-AA-IA----MFI-NMLR-I--------P-ILI-Q.......AVERGP-DAT--GV-AQGKV.--V-YLT-FSS-L-----N-VIWI-QI-AS-AWTLRG-GKF-WEQLQKAIQKAIQQARQLREVAARV....IAYIS-GI---QAA-TGI-DSL--FTWNW-EA 853
GSN.CM.99.CN71          ---V-AA--A---IMF--NMLR----------P-ILT-Q.......QVER-P-D-TRKD--AQGNV.--V-YLT-FSS-L-------VIWI-QI-AT-VWTLQ--GKF-WEQ....LQKAFQQARQLGEVAARA....AAYIS-GI---QAAVSGI-DSL-VFTWNW-EP 858
MUS_1.CM.01.1085        -C-V-A---LF-L-MF--G-LR----------P-ILIRP.......QT--EA---T-GD--SGS--.--V-YLS-FFS-F-----N-VIW--QT-ISSASTLWWLLQRAVHLIQKGSI-L-FT-RALIEPLREALQRALAC--HGI---QTASTAV-DQL-TFTWNW-ET 869
MUS_1.CM.01.CM1239      -C-V-A---LF--FMF--G-LR-I--------P-I-TQP.......-EAAEP---T-GGV-SSG-T.--V-YL--F-S-F-----N-VIWF-QT-VSSVSTLRW-LQRAAHLIQKWSV-L-IVIREQARRCSELLARLLAYI--GI---QAAGTAA-DQL--FTWNW-EP 863
MUS_2.CM.01.CM1246      -C-V-A--IA---MFI-GMLR-I---------P-I-TRQ.......SAEQEQQ-KT-KGA-RSG-H.-........-S-FSE--HN-MIWT-QI-TSSASNLWWLTQRIYHHLTKGLVNP-TL--SFREALGRFIAH....A--GI-K-QAAVSGA-D-L--FTWNW-EP 854
MUS_2.CM.01.CM2500      -CLV-A--IA---IMFL-G-LRK--------PP-I-TRA.......-E--RT--ET--G--A-GSV.--V-YL--F-S-------N-IVW--QI--SS-SNLWWLIQ-IVHHLQKWSISL-LRARATSTHLLEVLRRLLAY---GL--FQEGVTAAIDGL--WT-NW--A 851

MAC.US.x.239            GVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIQQ..D-AL-TREGK-RDGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRVYQI-QPILQRLSATLQRIREVLRTE.........LTY---GWSYFHEAVQAVWRSATETL-GAWGD 851
H2A.GW.x.ALI            GVY-IV-I-A----IY-VQMLS-L-K--R-VFSSPPGYIQQIHIHK..DQEQ-TRG-T-EDVG-NV.GDRLWPWPIAY-HFLIHLLAR-LIGLYSICRD-LSRISPI-QPIFRSLQRALTTIRDWLRLK.........AAY---GCEWIQEAFRAFARIARETLTNTWRD 834
H2A.DE.x.BEN            GVH--V-IIA---AIY-VQLLS-F-K--R-VFSSPPGYLQQIHIHK..DRGQ-ANEGT-EDVGGDS.GYDLWPWPINYVQFLIHLLTR-LIGLYNICRD-LSKNSP....TRRLISQSLTAIRDWLRLK.........AAQ---GCEWIQEAFQAFART-RETL-GAWGW 832
H2A.SN.x.ST             GVY--V-IIV----IY-VQMLS-L-K--R-VFSSPPAYFQQIHIHK..DREQ-ARE-T-EDVGNSV.GDNWWPWPIRYIHFLIRQLIR-LNRLYNICRD-LSRSFQT-QLISQSLRRALTAVRDWLRFN.........TAY---GGEWIQEAFRAFAR--GETLTNAWRG 832
H2B.GH.86.D205          GLY--A---V----VYIVQMLA-L-K--R-VFSSPPSY-QQIPIRK..DRGQ-ANE-T-E-GGNDG.DYRSWPWQIEYIHFLLRQLRN-LIWLYNGCRT-LLKTFQ........ILHQISTNLQPLRLP.........VAY---GISWFQEALRAAAR--GETL-SAGET 831
H2B.CI.x.EHO            GLY--I-I-V---AIYIIQMLA-L-K--R-VFSSPPSY-QQIPIRK..DRGQ-ANE-T-E-GGNNE.GYRSWPWQIEYIHFPIRQLRD-LIWLYSGCRT-LSKTFQT-QPVLQPLRLPPA....................Y-R-GISWFQEAIQAAAR-AGETL-SAART 824
H2G.CI.92.ABT96         GLYV-A-VIV----IY-VQ-LG-L----R-VXSSPPSYVQQIHIRK..DQEX-TKE-I--ESGNKG.GYRSWPWQIEYIHFLXRQLGN-LTWLYSNCRA-LSRIXQT-QPLFQRISRTLQAIREHLRLE.........AAYFS-GFRW-QEACTAATR-AQETLTSTWRA 856
H2U.FR.96.12034         GLYV-A-III----IY-VQLLGSL-K--R-VFSSHPSYVEQIPIQR..DQEQ-TKG-I-E-AG-SG.GYRSWPWQIRYIHFLIHQLIR-LTWLYSSCRD-LSRICQS-QPLF.......QSIRERLHLE.........IAY---GW-YF-EAFQAFGK-ARETLSRTGRE 839
MAC.US.x.251_1A11       GVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQ*THIQQ..DQAL-TREGK--DGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRVYQI-QPILQRLSATLRRIREVLRTE.........LTY---GWNYFHEAVQVDWRSATETL-GAWGD 849
MAC.US.x.251_BK28       G-YV-V-VIL----IYIVQMLAKL----R-VFSSPPSYFQ*THTQQ..D-AL-TREGK--DGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPILQRLSATLRRVREVLRTE.........LTY---GWSYFHEAVQAGWRSATETL-GAWRD 852
MAC.US.x.EMBL_3         G-YV-V-VIL----IYIVQMLAKL----R-VFSSPPSYFQXTHTQQ..D-AL-TREGK--DGGEGG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPILQRLSATLRRIREVLRTELT.........Y---GWSYFHEAVQAGWRSATETL-GAWGD 853
SMM.US.x.H9             GVL--L-VI-----IY-VQMLA-L----R-VFSSPPAXVXQIPIXT..-QEL-TKEG---DGGX-G.GNRSWPXQIEYIHFLIRQLIR-LTWLFSSCRDWLLRXCQI-QPVLQSLSRTLQRAREVIRVEIT.........Y---GWRY-QEA-XXWWKFXRETL-SAWRD 858
SMM.US.x.PGM53          GVL--L-V------IY-VQMIA-L----R-VFSSPPVYVQQIPIHK..-QEQ-TKEG---DGG--G.GNRSWPWQIEYIHFLIRQLIR-LTWLFNNCRTYLLRIYQI-QPVFQRL...LQRIREVVRTE.........ISY---GCRWFQEAVQAAWR-RGETL-SAWRD 857
SMM.US.x.SIVsmH635F_L3  GVL--L-V------IY-VQMLA-L----R-VFSPPPAYVQQIPIHK..DQEP-TKEG---EGG--G.G-RSWPWQIEYIHFLIRQLIR-LTWLFSSCRDWLLRIYQI-QPVLQRLSRTLQRVREVIRIE.........ITY---GWSYFQEA-QAWWKVARETL-SAWRD 859
SMM.SL.92.SL92B         GVFL-I-I-L----IY-VQMLS-L----R-VFSSPPSYHQQIHIQR..DQEL-AKEGT--EGGNGG.GYRLWPWQIEYIHFLIRQLIRILTWLYNNLTR-ASRAYQN-QQLCQRLSEISQPIRELVRREAG.........YIR-GWNYFIEACQEAWRSAQE-IVGAWGL 858
STM.US.x.STM            GVYL-I---M--VAIYIMQLLA-L-K--R-VFSSPPSCRQQIPIHK..-QEQ-TKEGT-E-GG--G.GINSWPWQIEYTHFLIRQLVR-LTWLYNNFRACLSRIYQT-HPTFQRISRILQRIREVVRLG.........AAY---GCIWIQEA-QAAWR-AGETL-SAGRD 854
MNE.US.x.MNE027         GVY--V-VIL----IYIVQMLAKL----R-VFSSPPSYFQQTHIRQ..DQAL-TKEGT--DGG-SG.GNSSWPWQIEYIHFLIRQLIR-LTWLFSNCRT-LSRAYQI-QPIFQRFSTTLQRVREVLRTE.........LTY---GWSYFQEAVQVAWRSATETL-GAWGD 854
GRV.ET.x.GRI_677        GFLVIVIIL---FAWVLWGCIRNI----N--P........QIHIHS..SAE--DNGGGQDRGGESSSSKLIRLQEESSTPSRINNWWLN-KSCS--IRTWCYNICLT--IFIRTAVG.......................Y---GL-Q-QEA-TG-AQ-L-R-AR-AWG- 826
VER.KE.x.TYO1_patent    GFLVIV-II---LLYT-YGCIV------V---P.......QIHIHQ.V-KG--DNAD-P-EGG-NSRIKLESW*KDSKSRCMQ-TAWLTRLN......TWLYNSCLT--IQLRKAFQ.......................Y---GLA---TG-QEI-QTL-GVAQNACHQ 835
VER.KE.x.9063           GFLV-V-II---LLYTIY-CIG----------P.......QIHIHP..WKG--GNAG-PEEGGE--ND-SDSWQKGSGTRQKRGSWSKRLTNWLSTFSTWLYNTFLT--IHLRGAWQ.......................YI--GLG---TA-QEARL-V-RFAQNAGHQ 849
VER.KE.x.AGM155         GFLA-I-VI---LLYTLYTCIA----------P.......QIHIHP..WKGQ-DNAG-PEEGGRTGKSKSTH*QKEFGGRDKRTSWCRQLTNWF-TLSIWLHNSCLT--LKLRSAWQ.......................Y---GLG---AA-QEAV-HL-SFARNAAHQ 842
VER.DE.x.AGM3           GFLD-L-II---LLYT-Y-CIA----------P.......QIHIHP..WKGQ-DNA-GP-EGG-KRKN-SEPWQKESGTAEWKSNWCKRLTNWCSISSIWLYNSCLT--VHLRSAFQ.......................YI--GLG---AA-QEAVV-L-RLAQNAGYQ 849
TAN.UG.x.TAN1           GFY--I--ILF-MAWLIWGCIA------F---P.......QINIRL..-REQ-DNAGG-DKDSSSS.-DKSPPSVKES-LPNRGGIQAEERAW-QHLTNWCL-ISSW--RLYQILRRSLTTLLQLLRQE.........CQYI--GW-QF-EG-ARSFE-L-S-AQSASRT 851
SAB.SN.x.SAB1C          A--VIA-IIVA-VLLVIIG-LRKF-K--A---SLPSSH*.QIHIHL..DKGQ-DKE--Q-SGGLNSDSRSYTWQREFL............RHLCH-LITWLRN.LTSWFSTIFSNLH-CLQDIQQRTRQLTA........H-E-GW----AA-A--TV-VVQ-ATSAS-S 839
OLC.CI.97.97CI12        GLLV-LAY---N--LGLVRKIRIM----VS-QLP......FSDTLQ..DKEQ-DNG-G-DS-FFES..TFNEW-CTTECTGTLRVNWRDLLS-LWKGVVRVTR......................................WI.QKSLQHVWRTLKQKRQ......HGIN 814
DRL.x.x.FAO             GFF--MAII----LA-LW-T-G-F----R--.PYIFKDAYRPETTN..YHRQ-DREKG-EQD-GEW.NIRSEPSKPGSSKAWSEE..TVGTWLKESRGYIWLKNVKAV-EYGWSE.........................-QEAGRGIC.TLLQEVAQRIWREGRQLGLS 867
RCM.NG.x.NG411          G-F-IL-II----I-LLW-C-SKL-A--T--LSPPPCYHQQIPIRN..REEQ-IKEG-K--SGEGG.GHKYYNWQRG......YFYILILRPIEILCRR-YQICSNS-SVIYQSLQSIPSRIL-HLRWIGA.........K--DGW--F-GFSSW-AEMARTNAYYTWRG 829
RCM.GA.x.GAB1           GFF--I-II---VAWLLWNCLSNL----R--..SPPSYVQQIHIHN.T-EPQTP-EKR-D-GEEGG..NKYNNWLRE......YCWIQ-IHPLSRIWTQ-SQICRSCSSIIFQSLRWILA....................KI--GW--F-EFSSWFAEMALQNAYYTWRG 822
MND_1.GA.x.MNDGB1       GVFL-L-II----LL-LWNTIS-F----R-VFS.......QDCQQN.LYRK--DNG---SNSLELG.EHNSENL...KEESLNRSLIEDLTSFARE*PIWLWLKNLRAAIEYGF............LAVQEA...IRSLGS-LVSFAAISWR-GYTGWLQ-NRIFT-CRE 853
MND_2.x.x.5440          GLLV-VVII---MLACLW-VLGKF----R--PYVFKGDYLRPH.....NLK--DREGG-EPDLEKQ.NIKSESSRQESRKPWKPE...QVRSWLK-ST-Y-WLKNLQAVIEYGWQE........................-KAAGAALYQVLQGFAQRLWSRGYQL-LSG 863
MND_2.CM.98.CM16        GLL--IVIIV---LACLW-VLGKF----R--PYVFKGDYLRPH.....NLKQ-GKER--EPDSEKQ.STKSDSSKVEFGKPWSKE...QIRDWLKTSRGYVWLKNLQAVIEYGWQE........................-KTAGGKIF-VLQGYAQGLWSRGHQW-LST 839
MND_2.GA.x.M14          G-L-IMVVI----LACLW-TIG-F----R--PYVLKGDYLRPR.....NLRQ-DEER-GELNLEEQ.NIKSESSKKEFGRPWRPE...QIRNWLKESTIY-WLKNLQAVIEYGWQE........................-KAAGARICQVLQSFAQRLWSGGHQL-LSV 840
DEB.CM.99.CM40          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-DTITWAGGELWQWLQGIAEM 842
DEB.CM.99.CM5           VVY-IAA-IL---AMFGVN-GTKLCR-----QTL-PK.........EVEVLQLG-EGDKS--GVKS.-WTTSPR-F-R-----ISQ-LTWL-QT--NCAWLIR-GF---QETIYRLSFAF-QR-RLLGEHLRVV....-GYC--GWK-F-DTIIWAGQEIWEWLQGIAQV 822
DEN.CD.x.CD1            VVYVIA---I---IMFIINVLS-LCR-----LQIPIR.........-QKAANQPEEG-GESGDGNRLKWWQYPR-F-SI--E--SQ-LTW--LT--N-VFLIRTACSRIWDFV..........HRHSLILLG........YF--GL---QTGLRD-GTSAIQQGRATAEV 836
TAL.CM.01.8023          AVIV-AS-IL----MYIINVLKLL----L--P-K-PI-S.......N-E-GV--EEGT---KP--S.-WRTSPQ-FF-I--E--N--LSY--QT--N-VWLSWDLLKR-KATTQQAAQATYAY-RIKFTAAV.......ATI--GFS-IQAAWGDCVT-L-RFTWRW-SG 862
TAL.CM.00.266           AVIV-AS-IL----MYIVNVLKLL----L--P-K-PI-S.......N-D-EA-G-DAT---KQGSY.KWN-SQA-F-SI--E--TG-IRYI-QT-AN-AWLIWDLAKAAKAKSQQVAQATYAF-RLHLTAAAATLQY.......GFS-IQAAWRDGIT-L-RFTWWW-NG 844
MON.CM.99.L1            --IV-AA-I----LMFI-GVIS-LG----L--S-IPI-S.......HA-QPT-D-TGAG--DGSNS.--PAYLK-FFTI--E---N-V-WT-QI-K-SV-VIY--LQRVSQRLPPLLHIRLLQ-WESLRRLLAYCQY.......GI---QAAVT---D-L-RFTIVW--A 841
MON.NG.x.NG1            --IV-AA-----VLMFI-G-IR--G----L--P-...IPTQVPR....E-GA--ETG-G---QG-NRXTVYARGFLTIIWT-CRDLIVWTFQLLRDSG-T-YRSLQRVRSHLIPLLRDLCRQLREASSRLL........AY---GL--XQXACTGAID-L-RFTVIW--A 855
SUN.GA.98.L14           AVL-I-IIIAA--LSFIIQQIY-MC---RV--PSAYVEQDWLQETC.PK-TDK-EE--TEK--I..YINLEQSKKESLPPPWTVDWDEPLRDSL-VT--KWLKAWGIV-AQNIYHLLSFL-HL-TTSFH.HGQRLWQTLRGWLGSHLIQAT-RIRNACR-SRERVSSSQK 886
WRC.CI.97.97CI14        GLLVI-IV-----LACLMNSIRIA-----K-IS-......QNHLNN.LEEGQQDSE-GIEEDYNSIEINDWQYRRES-IVFLNSLLTS---ICQHLLHIVWSLLSCL-WWINDGW........SLLKDI..............SGRAYGAIRKKIAKSRKENTQ--VSAT 844
WRC.CI.98.98CI04        GLL-I-IV--I--LGWLMTSIRIA-----K-IS-......QSHLNN.LEEGL-DRE--AEEDNSTLDV-DWQFKKES-ITLLNSLINF-SAVCQTLLQIVWSLLACLFWWIQHGW........NLLKRI..............SGAAHACIRKKITKSRKETTEK-ISEN 854
WRC.GM.05.Pbt_05GM_X02  VLLVI-LV--I-VLGCLFTSLRIA-----K-VS-......QSHILN.LEEGL-DNE-T-EEDFIAIEE-DWQFKKESLILFWNSLATL-Y-ICLSLLQ-SYSLLSYLWI-IQNGW........SLLKRL..............CGAAYAALQEKIAARRQKRKEK-IYTP 846
LST.CD.88.524           AVL-II-III---LSFIIQS-VKMCR--RV-APSAYVEQDFKWENG.ESQEQ-DKEG-TEKD-ENIYINLEQCKKESFRPPWNVDWNEPLQDSLFVT--KWLKAWGI--LTLVWQFLTWLGHIVILSYH.HGQHLWQTLSRFLVE-AQKTA-WIRNKSRRNRGELSK-NG 887
LST.CD.88.485           AVL-II-IIV----SFIVQNIVKMCR--RV-APSAYVEQDYKWEKG.EN-EQ-DKEG-IEKDKEDIYINLEQCKKESSIPPWNVDWSEPLQDSLFVT--KWIKAFGI--LSLVWQFLSWLGHLVILFFQ.HGQHLWKASSGWMVENAQKIARWSWNKIRRNRRDISAAHA 884
LST.CD.88.447           AVL-II-IIV--V-SFIVQNIVKMCR--RV-APSAYVEQDCKWEKR.ENQEQ-DKEG-IEKDKENIYINLEQYKKESSTPPWNVDWSEPLQDSLFVT--KWIKAFGI--LSLVWQFLSWLGHLVILFFQ.HGQHLRKACSRRVVENAQKIA-WSWNKIRRNRRDISAAHA 885
LST.KE.x.lho7           AVL-ILVII---VLSFIIQN-VKMCR--RV--PSVYIEQDYKWEKE.ENQEQ-DRE--K-ADTETIYINLEQCKKESSRPLWNVDWNEPLQDSL-VT--KWLKEGGI--LSLVWQSLSWL-HL-ILFFQ.NGQRLWQTSSRWMVENAQKIQ-WLREKCRRNRGQLSS--- 882
SYK.KE.x.SYK173         GFYVI-A--L--L-SFSVG-IKNLLG--V-ILQNPT...........Q-RKD-GKPAD-EEGSGDR.EGLNVSTF-RESLRQS-EAGQQLWRTVCSSFRSLIRQLTITW-..............................FIS-GFN---IA-A--GREIRDW--AIWQA 804
SYK.KE.x.KE51           GFFAIAAI-I--VLSFAWG--RNMLG-----LQKPQP--........LYHKA-A-P--GIETDTDN.YKPSVSTF-REFLRQS-AEAQKLWNRVSESCRSLIRGL-IAW-..............................FIS-GVT---EA-I--GREV-HQMVAIWQA 812
COL.CM.x.CGU1           AAYV----I----IMV-IACIRGAFRVKGFQQIGRTNVSS......QIRGQDR-WEQPDNA-EIHR.EQENSSSRFW...............RLMEEWWSSLL-HCSNGILT............VLPQHPPTTS....DKRRSA-.-T-RYL..LVPRGPSLQIL-TLQ- 809
```


PLV Proteins

```
                                                   gp41 end
                                                    Env end
                                 *
H1B.FR.83.HXB2          VIEVV...QGACRAIRHIPRRIRQGLER..ILL*   856
H1A1.UG.85.U455         ---IG...TIG---LN------------..A---   856
H1B.US.90.WEAU160       ---I-...-RT----L------------..A---   860
H1C.ET.86.ETH2220       F--LI...-RIW--FCN----------A..A-Q-   852
H1D.CD.84.84ZR085       I-DI-...RR--K-VL---T--------..A---   857
H1F1.BE.93.VI850        I---L...-R-G--VLN--------A---..A---   833
H1G.SE.93.SE6165        ----A...-R-----LN--T--------..A---   867
H1H.CF.90.056           I-VI-...-R-W---L---------F--..S---   846
H1J.SE.93.SE7887        I--IA...-R-F---L------------..A---   849
H1K.CM.96.MP535         I--IG...-R-F--LL------------..A---   843
H101_AE.TH.90.CM240     ----A...---W---L------------..T---   858
H102_AG.NG.x.IBNG       A--IG...-RVG----N--------F--..A---   855
H103_AB.RU.97.KAL153_2  ---IG...-RF---M-N--------A-K..A-Q-   840
H104_cpx.CY.94.CY032    I--A-...-R-----CN-----------..A---   856
H1N.CM.95.YBF30         I--LA...-RIG-G-L------------..A-I-   853
H1O.BE.87.ANT70         I-AGI...-RIGTG--N-----------..S---   857
H1O.CM.91.MVP5180       I-LGL...-RIGQGFL--------A--..--V-   870
CPZ.CD.90.ANT           ILLAL...-TIV-I--EV---------I..A-N-   859
CPZ.CM.01.SIVcpzCAM13   I-RA-...-IVF-I-GN------------..T---   865
CPZ.CM.05.SIVcpzEK505   I--L-...-RIG-G-L------------..A---   836
CPZ.CM.05.SIVcpzMB66    I--IA...-RFG-G-LN-----------..A---   878
CPZ.CM.98.CAM3          IL--A...-IIG-G-L------------..S---   860
CPZ.GA.88.GAB2          I---A...TRIG-G-L------------..A--E   856
CPZ.CM.98.CAM5          I--F-...-IVG-G-L------------..S---   866
CPZ.CM.05.SIVcpzMT145   IL-AI...TRLG-G-L------------..A---   853
CPZ.TZ.00.TAN1          I--G-...VYIA-G--N---------L..A-N-   872
CPZ.CM.05.SIVcpzLB7     I--I-...-RIG-G-L---------F--..A---   846
CPZ.US.85.CPZUS         I--LT...RRLFLG-I------------..S--X   837
CPZ.GA.88.GAB1          I--AF...-VTL-I--N-----------..A---   855
GSN.CM.99.CN166         -LHAC...RRVW-EFLA--------A-I..LFN-   882
GSN.CM.99.CN71          LVQT-...GRVW-EFLA--------A-I..L-N-   887
MUS_1.CM.01.1085        -LQAG...RRVW-EFLA----V---A-I..L-N-   898
MUS_1.CM.01.CM1239      ILQIG...RRVW-EFLA--------A-I..L-N-   892
MUS_2.CM.01.CM1246      ILQFS...-KIW-KFLA--------A-I..L-N-   883
MUS_2.CM.01.CM2500      -LQ-S...KQL*K-FLA--K----A-I..L-N-   879
                                                    Env end
MAC.US.x.239            LW-TL...RRGG-W-LA----------L..T---   880
H2A.GW.x.ALI            LWGA-...-WVG-R-LAV-------A-I..A---   863
H2A.DE.x.BEN            LW-AA...RRIG-G-LAV-------A-L..A---   861
H2A.SN.x.ST             FWGTL...GQIG-G-LAV-------A-I..A---   861
H2B.GH.86.D205          LW-AL...RR-A---IA----------L..T---   860
H2B.CI.x.EHO            SWG-L...RR-AGE-IA--------A-L..A---   853
H2G.CI.92.ABT96         LWKTL...GRVG-G-LA----------L..T---   885
H2U.FR.96.12034         LW-TL...GRVG-WL-A--------F-L..A---   868
MAC.US.x.251_1A11       L*-AL...RRGG-W-LA----------L..T---   877
MAC.US.x.251_BK28       LW-TL...RRGG-W-LA----------L..T---   881
MAC.US.x.EMBL_3         LW-TL...RRGG-W-LA----------L..T---   882
SMM.US.x.H9             LW-TL...GRVG-G-LA-XX-------L..T---   887
SMM.US.x.PGM53          SW-TL...GRVG-WVLA----------L..A---   886
SMM.US.x.SIVsmH635F_L3  IW-TL...GRVG-G-LA-----V-----L..T---   888
SMM.SL.92.SL92B         IW-TL...GRVG-G-AA----------L..M-N-   887
STM.US.x.STM            LW-TL...GRVG-R-GA----------L..T---   883
MNE.US.x.MNE027         LW-TL...GRVG-W-LA-------E--L..T---   883
GRV.ET.x.GRI_677        LGAI-...RS-Y--VINS---V-----K..V-G-   855
VER.KE.x.TYO1_patent    IWLAC...RS-Y-N-VNS---V-----E..--N-   864
VER.KE.x.9063           IWLAC...RSTY-H-ISS---V-----E..--N-   878
VER.KE.x.AGM155         IWLAC...RS-Y---INS---V-----E..V-N-   871
VER.DE.x.AGM3           IWLAC...RS-Y---INS---V-----G..--N-   878
TAN.UG.x.TAN1           LWNAC...RS-Y---LEH---M--E---..WFN-   880
SAB.SN.x.SAB1C          -RHAC...RSIV--VIAH---M--E---..WFN-   868
OLC.CI.97.97CI12        LQQKK...-RRGSPLQKAQEQLGREGAE.TTGV-   844
DRL.x.x.FAO             SARAL...RALAQEVAA--------A-V..LFN-   896
RCM.NG.x.NG411          LCA-C...RDPAGWPATLC------F--..F-N-   858
RCM.GA.x.GAB1           LCA-A...RDFAGWPAMVC----------..LCN-   851
MND_1.GA.x.MNDGB1       A-IAA.....GTC-W--L-----SA---..P-N-   879
MND_2.x.x.5440          -RGAA....AFG-G-WN--------A-A..L-N-   891
MND_2.CM.98.CM16        AACFR....AIA-G-IN--------A-V..L-N-   867
MND_2.GA.x.M14          IRGAA....AIG---GN--------A-V..L-N-   868
DEB.CM.99.CM40          ALQGL...VR-GG--LRV-A-----A-L..--N-   871
DEB.CM.99.CM5           A-QGL...VWGG-NLLA--A-----A-L..L-N-   851
DEN.CD.x.CD1            -LAAL...TR-A-EVVA----------I..V-N-   865
TAL.CM.01.8023          -LAAA...RW-VQ--AA----T-----Y..C-A-   891
TAL.CM.00.266           -LAAA...RW-GE--AA----T-----Y..CFT-   873
MON.CM.99.L1            LLHAG...GRLW---VA--------A-I..F-N-   870
MON.NG.x.NG1            -LRLG...GRLW-GLVAV-------A-I..L-N-   884
SUN.GA.98.L14           ARSRTFSLGRKW-PKWNRT-GS-IPS-TTETT--   920
WRC.CI.97.97CI14        GSRKT...AASS-IRMLAT-LCASFDKCRRSFSE   875
WRC.CI.98.98CI04        SNRET...T-SS-VWKL-T-LCSSLNKSRRR-F-   885
WRC.GM.05.Pbt_05GM_X02  GSRET...SRSSSIRGITA-LCTNLNPCRRGQQ-   877
LST.CD.88.524           TNIRL...GQNR-RW.RFRF-SGLPS-TTETT--   917
LST.CD.88.485           RDIRL...GQKK-WRFRFRG-SGFPS-TTETA--   915
LST.CD.88.447           RNLRL...GQKK-WRFRFRG-SGFPS-TTETA--   916
LST.KE.x.lho7           KNIQL...GKKK-WRLRFGG-SGISS-ATETA--   913
SYK.KE.x.SYK173         IYAAT...RRVVE-VAAL---L-----I..Y-N-   833
SYK.KE.x.KE51           LLAYA...RRVAENVAAL---L-----I..Y-N-   841
COL.CM.x.CGU1           WLRSA...ARGW-RAPEYL-GWIYDRPQ..GPA-   838
```

```
                        myristoylation
                         Nef start                                                                                    acidic cluster poly-P helix  phos  (phos)          HXB2 premature Nef end
H1B.FR.83.HXB2          MGGKWSKSSVIGWPTVRERMRR............AEPAADRVGAASRDLEKHGAITSS.......................NTAATNAACAWLEAQ.......EEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDLWIYHTQGYFPD*QNY 126
H1A1.UG.85.U455         -------K-RVE--E--K---E............TPA--KG---V-Q--D-Y--V---......................--SS---S-------.......--GD-----R-------------F---F--------D------K--E-----V-----F---W--- 127
H1B.US.90.WEAU160       ---I---R-GS---AI----K-..............-----EG---V----A---------......................-E..N--D-V--K--.......-D-------R-------------H-----#----------Y--K---------V-----F---W--- 124
H1C.ET.86.ETH2220       ---TM--C-PV---AI---I--............-A---EG--------D-Y--L---......................--P-N-PD-------E......---------R-------------F---L------------Y-KK--E-----V-N---F---W--- 128
H1D.CD.84.84ZR085       ---------IV---AI---I-K....TDPRERRRP-----G---V-----R--------......................---T-------V---E......-D-------R-------------------RK-------VY--K---------V-----F---W--- 136
H1F1.BE.93.VI850        ---------IV---A-G----Q............TPT--EG---V----DRR-------......................--RT--PDL------......----------R----V-----------------------Y-KK-G-T----V---------W--- 127
H1G.SE.93.SE6165        ---------IV---E----I-N............TPT--EG---V-Q--DR--------......................----N-PD-------E......-DS-----R---------F-G-F---F--------D---Y-KK--E-----V-N-------W--- 128
H1H.CF.90.056           --------RMG--S-I------............---V-EG---V----DRR--V-IN......................---S--RDA------E......DG-------R-----------G-F------------D---Y-KQ---------V-N-------W--- 128
H1J.SE.93.SE7887        --N-----......--Q-------...........APA----G---V-Q--A---------......................-------D-------.......T--------K--I--------G-----F--------D---Y-KK--E-----VHN-------W--- 124
H1K.CM.96.MP535         ---------IV---AI-------....ARPAADRVGTQ----G---V-Q--AR---V---......................--SHN-PD--------......----------R-------------F--GF---------D---Y-KK--E-----V-----F---W--- 136
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-KK--E-----V-N---F---W--- 125
H102_AG.NG.x.IBNG       ---------IV---K-MK---Q............TPT--TG-----Q--DR--------......................---Q--PD--------......-D-N-----R-----------G---------------D---Y-KK--E-----V-N---F---W--- 127
H103_AB.RU.97.KAL153_2  ---------IV---Q----I--...........APA---RG--PV-Q--D-Y--V---......................----N--D--------......----------R-----------G-F------------D---Y-KK--E-----V---------W-.. 126
H104_cpx.CY.94.CY032    ---------IV---EI------ARAEPERMRRAQ-----AG---V-Q--D------IN......................-------PDKT-----E......----------R---------F-G-L------------D---Y-KK--E-----V-----F---WD-- 140
H1N.CM.95.YBF30         --KI-----LV---EI---------......QTQEP-VEP-VGA----Q--ANR----IR......................--RDN-ESI------E......----------R-------I---Q-F---F---D-------VW-RK---------M-----IL--WH-- 133
H1O.BE.87.ANT70         --NALR-GKFE--AA----------......TRTFPES--C-PG--QI--E-AAR-G-P--......................H-PQN---L-F--SH......Q---------A-----------G-F---F--------------Y-HK-AE-----V-N---F---W--- 133
H1O.CM.91.MVP5180       --NA----KFA--SE--D-------......SSSDPQQ-C-PG---V--E-ATR-G-S--......................H-PQN---L-F-DSH......KD-D-----R---------F---F---F--------D---Y-HK-AE-----------F---W-C- 132
CPZ.CD.90.ANT           --SA---IKWV-A...-QAI-K...........IH-TNP-DI-PCGNE-ASR--L---......................TIGTEKDVITYS-DH......T--GT----R----M----E-L-----X--------DX--XXLK-AA---M-MFN---I---W--- 125
CPZ.CM.01.SIVcpzCAM13   --AQ-----LV---EI-K-IKE............-PI--EG--EV-K----Y-----R......................H-PS--QTL---DEMKH.....H---------R-----------G-F------RG-------VY--K--E-----V-----F---W--- 129
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-TK--E-----V-----I---W--- 127
CPZ.CM.05.SIVcpzMB66    ---------LV---E--D-L--............TQT--EG--PV----AR------R......................--SQ--ETL----EV......QN-------R---------F-G-F---------------VY-KK--E-----V-----F---W--- 127
CPZ.CM.98.CAM3          --NR-----LV---D----LSE............LNLL-PG--PV-Q---A---V-TR......................H-PQN-QTL----EMQE....HN-------R---------F---F-------K--------Y-IXXXXXX-HGY-----F---W--- 129
CPZ.GA.88.GAB2          --N------IV---Q----L--............TQE--EG--EV-Q--ARR-----R......................H-PQ--QTL----EM......-Q-D-----R-------------F---------------VY-K---E-----V---------WN-- 127
CPZ.CM.98.CAM5          --N------LV--AA--Q-I-Q............TP---SG--PV-Q---AR----TR......................-H.TKYQTL----EMEN....H---------R-----------Q-F---------------VY-TK--E-----V-----I---W--- 128
CPZ.CM.05.SIVcpzMT145   --A------LV---EI---IK-............-PT---G--EV-Q-IASR--V-IR......................--PQ--ETL----EM......QD-------K---------F-G-F----------------W-RK--E-----V---------W--- 127
CPZ.TZ.00.TAN1          --NIFGRWP..-ARKAI-DL..............HNTSSEP--Q--Q--QNK-GL-TN......................TLGTSADVLEYSADH......T--------R-A--M----E-L-I---W--------D--FF-PK-AA---T-M-N---V---W--- 123
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-KK--E----GV-----I---W--- 127
CPZ.US.85.CPZUS         --N------IV---E--N-L-Q..........TQTTA--EG--PV-Q--AE-----TR......................--PQN-QTL---DEMTN....H-S------R-----------Q-F--GF-----------VY-R---E-----V-----I---W--- 131
CPZ.GA.88.GAB1          --T------LV---E--R-I-E............-PT--EG--EV-K---R------R......................--PE--QTL----EM......DN-------R----T--------F---------------VY-R---E-----V-----F---W--- 127
GSN.CM.99.CN166         --S-N--QQQQESS-ALLSS-GTGQRPYFTL...VDEYGENFWLSPDASD-GRRYYLT......................EEPKPKRGS-..........--Y-PSC--R-R----DP---VM--------------MFYCED-HQK-EQYA-LEW-LV-GWLSF 131
GSN.CM.99.CN71          --S-N--QQSQESS-ALLSSHGTGQSPSFRL...LDEYGENSWLSPDASDRGRRYYLT......................EESPSRQNCI..........DY-PSC--R------DP---LM--------------AMFYCED-HQK-ESYC-YEW-IV-GWLQW 130
MUS_1.CM.01.1085        --S-S--QPAQQSLISSPPSPGTGRKQYFKL...V-EYGENSWLSPDASGRGRRYSLT......................EGASKRPVI..........E-H-.RTC--R-R--I-DP---LM--M------------MFV-TE-LQK-ET-A-VEW-II-GWLM- 130
MUS_1.CM.01.CM1239      --S-N--QQEQQPL-SSPSSPGTGQSPYFRL...VDEYGENSWLSPDASDRGRKYSLT......................EGS-ERPLI..........E-H-.RSC--R-R----DP---MMI----Y---------MFYCEE-HQK-ETYT-VDW-IV-GWLQF 130
MUS_2.CM.01.CM1246      --S-N--PQSQERS-PLRSSHGTGPNRYFSL...VKKYGENSWLSPDASARGRKYSLT......................ENKGRRQII..........E-H-.CSC--R-K--I-DVN--LMI----Y--K-R----MFVCID-LQK-ETYA-VKSKIV-RSLI- 130
MUS_2.CM.01.CM2500      --S-N--KESQQPLMGSPSGPQTGQTPYFKL...VNNYKKHSWLSPNASARGRKYSLT......................KGS-KKAII..........KKHK.RSC--R--H-IQNP---LIIN------------MFY-EE-HQK-ETYA-VEW-LI-GWLM- 130
                         Nef start      R17Y mutation                                                 premature stop in Mac239                                                         max HIV-1 similarity
MAC.US.x.239            ---AI-MRRSRPSGDL-Q-LL-ARGETYGRL...LGEVE-GYSQSPGG-D-GLSSL-C..EGQKYNQGQYMNTPWRNPAEEREKL-YRKQNMDDI.DE*.....DDDL--VS-R-K----T-S--L-I-M---I--------IYY-A--HR---IYLEKEE-II--W-D- 158
H2A.GW.x.ALI            --ASG--KRSGPLQGL---LLQTPGETCGGQ...CSGSGGGYSQSQGGSGRGQKLP-C..EGQRYQQGDFMNTPWRTPATEREKELYKQQNMDDV.DLDD....DDSL--VS---R-Q--T----L---M--LI--R-----MFY-E--HR---IYLEKEE-II--W--- 160
H2A.DE.x.BEN            --ASG--KLSKHSRGL---LL-ARGDGYGKQ...RDASGGEYSQFQEESGREQNSP-C..EGQQYQQGEYMNSPWRNPATERQKDLYRQQNMDDV.DSD.....DDDLI-V----R--R-E----L-I-M---I------Q-MFY-R--HR---IYLEKEE-II--W--- 159
H2A.SN.x.ST             --ASG--KRSEPSRGL---LLQTPGEASGGH...WDKLGGEYLQSQEGSGRGQKSP-C..EGRRYQQGDFMNTPWRAPA.EGEKGSYKQQNMDDV.DSD.....DDDL--V----R----E---RL-R-M--LI--------YY-D--RRV--IYLEKEE-IIG-W--- 158
H2B.GH.86.D205          --SAG--KRSERQQGL--KLL-VPERPYGRL...SGERREQSSRSPGESD-DLNSP-C...................EGQNARG-EGGGQQDAD-.SDE.....DD-VGAICK--I---------L---M---I--Q----MYY-E--HR---TYFENEE-IVSGW--- 142
H2B.CI.x.EHO            --SAG--KQSKQQ-GL---LL-ARRGPRGES...SGERQE-SLQYPGGSD-GLNSP-C...................DGQKTLG-EGGGKQDSD..EDD.....-DN---VR-R-G-------F-L---M--------E---IFY-E--HK---TYLENEE-IVSGW--- 141
H2G.CI.92.ABT96         --SAG--KRAPQQLGL-K-LLQARGEPYGKL...W-GLEEGYSQYREESG-GLSSL-C..EGQQYTQGQFMNTPWRNPATERAKL-YRXQNNDDV.DSD.....DNDL--V--Y-K----V-S--L-I-M---I--------IYY-E--HR---IYMEKE--II--W--- 159
H2U.FR.96.12034         ---SI--KHSKRSGKL---LLAARGENYGRL...WGGLE-GFELYQGGSG-DLSSRFCEAQGQGYSEGEFMNTPWRTPATGKDKLQYKQQNMDDV.DEE.....DNDL--VA-R-R-A--T----L---M---I------D-IFY---HR---IYLENEE-II--W--- 161
MAC.US.x.251_1A11       ---TI-MRRSRSTGDL-Q-LL-ARGETYERL...LGEVE-GSSQSLGE-D-GLSSL-C..EGQKYNQEQYMNTPWRNPAEEREKL-YRKQNMDDI.DEE.....DDDL--VS-R-K----T-S--L-I-M---I--------IYY-A--HR---IYLEKEE-II--W-D- 159
MAC.US.x.251_BK28       ---AI-MRRSKPAGDL-QKLL-ARGETYGRL...LGEVE-GSSQSLGG-G-GLSSR-C..EGQKYNQGQYMNTPWRNPAEEKEKL-YRKQNMDDI.DEE.....DDDL--VS-R-K----A----L-I-M---I--------IYY-A--HR---MYLEKEE-II--W-D- 159
MAC.US.x.EMBL_3         ---AI-MRRSKPAGDL-QKLL-ARGETYGRL...LGEVE-GSSQSLGG-G-DLSSR-C..KGQKYNQGQYMNTPWRNPAEEKEKL-YRKQNMDDI.DEE.....DDDL--VS-R-K----A----L-I-M---I--------IYY-A--HR---MYLE...-II--W-D- 156
SMM.US.x.H9             ---VT--KQRXAGGNLX--LLQARGETYGRL...W-GLEGEYSQSXDASG-GLSSL-C..EPQKYCEGQFMNTPWRNPATEGAKLGYRQQNMDDV.DNE.....D..L--C--S-R--V-I----L-I-M---I--------VYY-I--HK---XYLEXEE-II--W--- 157
SMM.US.x.PGM53          --AAG--KQSRQRGGLG-KLLQARGETHGKL...W-GLE-GYSQSRGE-GRDWNLH-F..EGQGYSEGQFMNTPWRNPAREREKLKYRQQNMDDV.DDD.....DD-LI-VS-H-K----A-S--L-I-M---I--------IYY-E--HR---IYLEKEE-IT--W--- 159
SMM.US.x.SIVsmH635F_L3  ---AI--KQHRRGGKL---LL-ARGETYGRL...W-GLEEGYSQSLGASG-GLSSL-C..EPQKYSEGQYMNTPWRNPTAEKAKLGYRQQNMDDV.DDE.....DDDL--VS-H-K----A----L-I-M---I--------IYYNEK-HR---MYMEKEE-II--W--- 159
SMM.SL.92.SL92B         ---IT-SKPARKHGDL-KKLLSARGDSYGRL...WDGLEEGSQQSQGGSDRALNSCLT.RDGYKYNVGEFMNTPWRNPATERQKQQYRQQ-MDDI.DED.....DD-LA-VV-W-K----A-S--L-I-M---I------D-IYY-E--HR---IYLEKEE-II--W--- 160
STM.US.x.STM            --ASG--KQRKQHGEL---LL-ARGETYGKL...L-GLGEGS-PSQGASD-GLNSH-C..EPQRYNEGQFMNTPWKNPAAESAKLEYRQQNMDDV.DEE.....DDNL--VA-H-R----E----L-I-----I-S------IYY-E--HR---MYLEKEE-IV--W--- 159
MNE.US.x.MNE027         ---AT--KRSKSPGDL-Q-LL-ARGETYGRL...W-GLE-GYSQSLGGSD-SLSLL-C..EGQKYNQGQFMNTPWKNPAGEREKL-YRKQNMDDI.DEE.....DDGL--V--R-R----A----L---M---I--------IYY-E--HK---IYLEKEE-IV--W-D- 159
GRV.ET.x.GRI_677        --SSN--RQQQ-LLKLW.-GL-GKPGADWVLL..SD-LIGQSSTVQEECG-ALKKSWG.........KGKMTP.....DGRRLQEGDTF.DEW.DDD.....---.-----Q-R----Q----L---F----S----D-IYY-E--EK--N-YALNEW-IID-W-A- 143
VER.KE.x.TYO1_patent    --SQN--PAHKKYSKLWQALHKTHVTRYGLL...-D-LIGTSSTVQEECD-ALRKSLI......RKQNGNMTE.....EGRRLQEGDTW.EEW.SDD.....---.-----R-R----Q----L---F---------D-IYY-D--NK--N-YALNEW-IID-WNAW 146
VER.KE.x.9063           --LGN--PQHKKHVSLWHALHKTQVTKYGLL...-D-LIGTSS-VQEEYD-GLRKSLI......RKRNGNMTP.....DGRRLQEGDEW.DEW.SDD.....-D-.-----K-R----Q----L---F---------D-IYY-E--NR--N-YALNEW-IID-WNAW 146
VER.KE.x.AGM155         --LGS--PQHKKQL-IWRALHATRHTRYGLL...-D-LIGQSSTLQEECD-GLRKSLI......RKRNGNMTP.....EGRRLQDGDQW.DEW.SDE.....-D-.-----R-R----QI---L---F---------D-IYY-D--NK--N-YALNEW-IID-WNAW 146
VER.DE.x.AGM3           --LGN--PQHKKQLSLWHALHKTRATRYGLL...-D-LIGQSSTLQEECD-ALKESLI......RKRNGKMTP.....EGRKLQEGDKW.DEW.SDE.....-D-.-----R-R----Q----L---F---------D-IYY-D--NQ--N-YALNEW-IID-WNAW 146
TAN.UG.x.TAN1           ---SN--REQQ-LLRLWRAL-KAPVVRYGML...-D-LIGQSSNIQEECD-NWNGG-T........RRGKSTP.....EGRKL-ADDTW.DDW.EPE.....---.-----R-R----Q----L---F-----------IYW-PK-EQ--N-YALNEW-IID-W-A- 144
SAB.SN.x.SAB1C          ----S--QQQRHSLWLWSKL-QAPVIQYDML...-D-LLGQSSHIQEECA-SLRDGLI......RQGDSSRTE.....EGVKMKHQGRQ.PSWYDED.....---.-----R-CL---A----L-I-FG----------IYY-E--KK----YALNEW-IVDGW--- 147
OLC.CI.97.97CI12        --SIC-RRNRE-EVLY-RLRSNWGEKEQKQL..............ESEDQHQTNLAI...................................ESLMTGDTDTDDLE----K----V-QA---TLIE------------MYW-P--ENLIY-YALNCW-IIGNF-AT 118
DRL.x.x.FAO             ---QN--KREEAYARFYK-L--GYGAKGGNLDYHQLEPSEPLLRKWQPFPGGLDKAQRSSSTS............SLDTTHMQ--AAEKDVT-PTEEEQPSEE---------C--R---EP---DL--F---------K-IWW-L--ET----YAQNEW-FIKGW-S- 157
RCM.NG.x.NG411          ----S--A-AA-LRRW-GL-LTTPGEDYARFAETLQDGQP-CAEG-GRASRDFLTR..........................GGFTIETQQSVDAIDWYEDTDDTL-----R----I---S--L-I------------IYF-E--RR----YMENEH-II--W--- 142
RCM.GA.x.GAB1           ----S--N-AA-LLRW-FK-LTTPGEGYVRWHETLLDGQPWCAEG-GRASRDFV-R..........................GGIT-ETQASIDDIDWYEDTDDTL-----K--------S--L-I-M----------IYW-I---R---MYLENEH-II--W--- 142
MND_1.GA.x.MNDGB1       --SSQ--KRSEA-VRYSSAL-QLVGGPVTPDGYKQIESSQGAEKQ-LLRGRAYGTY-EGLDKVQNDPLTK.....................DEKLDLTQQDPE---------CR--S--VPS--DLI-F---I------G-IYY-R--EE----YAENEW-FE-GW-Q- 148
MND_2.x.x.5440          ----S--QREQHYIKYYKAL--GYGAEGTNLDYQVLEEQQPLAG--GTSLAASDKVLRPYSTE.....................RDLRHQQDVTLPSEKEQPSD---------Y-TQ-VPEA---DLI-M------------IWW-K--EE----YAQNEW-LI-NW--- 149
MND_2.CM.98.CM16        ----S--QQEEKYLKYYKA---GYGAEGTNGDYQQLH-SEPLLG-LSTSQEEFDREQKSSSTD...............................----T----Y--C-V-EP---DL--MP-----------IW--K--EE----YAQNEW-FI-TW-S- 129
MND_2.GA.x.M14          ----S--QREQ-YAKYYKAL--GYGAEGTNLDYQLLEEQRLLEGPLGTSLGASDKELKSCSTK..........................VDDEQK--R----Y-TQ-V-EA---NLI-M------------IWF-R--EE--N-YAQNEW-FI--W-E- 133
DEB.CM.99.CM40          ----N--IPSR-QVGSFGSGS-GLLRWRYRDLSEQ-EQFSECLLE-DREQSSSSTE..........................FWGSPRREIKCTNKQQQDLDQT-A-----R---SI-DP---LMI-Y-----------DVFY-A--HA--E-HAQNEW-II-GWLQ- 142
DEB.CM.99.CM5           ----N--TPSF-QGKKSGSGSKGLLRWRYKDLSGEAETFSPFLQE-DKEQSCYSTE........................EPLYADPHREIKCT-RAQECYD--HGI--A-R-R--I-EP---LMI-Y--Y--------DIFY-A--HA--EIHAH-EW-II-GWLK- 144
DEN.CD.x.CD1            --S-N--QASGI-GLAPSNKEGRQRKWYWRLLPEQQEKWWPSPDG-DRAL-LSSTR.........LLEHDEDEETFVDAEEIIQSSPALDPQHPRIHQYVVDHDD--------R-RR--HAP---DMI-M----------QDIFY-P--HA----YAQNEH-FITGW-T- 161
TAL.CM.01.8023          --SVK--LLGETASQRLQGSPGAGRLVYWRLRDGQCRQLQ-SRGERGKGLNIALLD..................AGWQASDANWMGK-LSEGEVWYEKEQEENDM-L-----R-------CN--LGI-F-F---------K-IFYNN--HA--N-YAHNEW-IL--W-S- 151
TAL.CM.00.266           --SVK--LLGETASQRLQGSHGGGRMVYWRLRDGQVKRSQLSPG-RGKDLNIALLD..................EEWQATGADYMGK-LTQGEVWYEKEQEENDM-L-----R-K----TC---LGI-F-F---------K-IFYNK--HA--N-YAHNEW-IL--W--- 151
MON.CM.99.L1            --S-N--QQSPLSS-PLLGSQSSGRMRYFMLEDDYGEQSWLSPD--DRER-YSLTE..............................GRNGKQRRQPLDDDDDDDG--C--R-R----DP-W-LMM----Y-------GEMFYCED-HRKIEQYA-LEW-LI-GWLQ- 137
MON.NG.x.NG1            --S-X-XQPAQEPLMPSHGSQSSGQTRSFAWEDDYGEDSWLSPD--DRGRRFSLTE.........................GR-HRSNRRPTTVVDDADEDQDLP-C--R-M----DP-W-IMM------------DK-FYCXDXHQK-EQYA-VMW-LV-GWLQF 141
WRC.CI.97.97CI14        ---VF--T--EEHMERYA-RLQRAGRRIRRQRYQQLA-ERQQQ-VESGCLLPDSAQALIN.............................AE--SVKEDSSD-D-G----------I-QP-F-LL-----LF-------DIFW-A--EE-VN-YAHNEW-LIDGW-A- 138
WRC.CI.98.98CI04        ---IF--E--EQRMHAYA-RLQRAGRKQRRKRYQKIATERQQV-VESGNLLPDSAQALIR.............................AE-DSSKEDSS.DD-------H----I-QA-F-LL-----LF-------YIFW-P--EE-VN-YAFNEW-LIEGW-A- 137
WRC.GM.05.Pbt_05GM_X02  ---VF--D--EQRMLRYR-RLLRVGRKERRRRYTRLA-ERQAE--ASGELLPDSAQTLIR.............................AE-DSSKEDSS.DDDAI---R------QA-F-IL-----LF-------DIFW-P--EE-VN-YA--EW-IVEGW-A- 137
SUN.GA.98.L14           --NAFGRP-EV--VRTLF-L-AGSGTRAEPAGREYHRLRRQEVE..............................PLVS-ENGGPNGI-Q---E---R--R--CKP---QLI-----I---------WY-RT-EE----YAENEW-FITGW-D- 123
LST.CD.88.524           --NIFG-P-ADS-WR-LRKL-AGSGTRAEGTEESYQKLMEQTLG..............................WDKTEEDG.DSDS------R--R--CSS---TLI-----I---------YW----N---T-YCENEW-LIG-FM-- 119
LST.CD.88.485           --NIFG-PAAD--WKTLR-L-AGAGTRSEGTEETYQQLMRETSD..............................WDKRKDGDSDS-G------R--R--CSP---TLI-----I---------YW--------C-YCENEW-LIG-FMT- 120
LST.CD.88.447           --NIFG-PAAD--WKTLR-L-AGAGTRSEGTEETYQQLMQETSD..............................WDKRRDGDSDS--------R--R--CSP---TLI-----I---------YW--------C-YCENEW-LIG-FMT- 120
LST.KE.x.lho7           --NAFGRPAAD--WKTLR-F-AGSGRNAEGTEDSYRALIEKTSN..............................W-KRRDGDSDS--------K--R--CSP---TLI-----I---------YW--------I-YCENEW-LIG-FM-- 120
SYK.KE.x.SYK173         --STS---QQLRSEGKYAIGW-LFGKQYTPLPDELSRPLQPCRGGFDKAWRSTLTE..................PIDPHGPDRDWGHSGGQKFSPGDIVQDEGDTGL-----C--T---TL---L-I-----I-N----Q-MNYCEK-DE--H-YLQNEH-II.-RI-- 150
SYK.KE.x.KE51           --SQS--KQPSR-DEKWRTRWWPFGKPYSPMPDELLRMSQPYHEDFDRGWRSTLTE....................PILDPKRDFDSGGKKWN-GDICHDEGD-DL-----L-------L---L-I-----I-R----Q-MFFC-K-HE--Q-YLKNEH-VI.-DIT- 147
COL.CM.x.CGU1           --SLL-CLNIRQQ-QT-GGVLGRLFGIYWYPGVQ-FRFLRHFRGGYDLRRVA-VERLNICEDGYMTVPKVLPN.........CPCPLCYRHPDVYYHNPEPEDV-DV-A-CY-MSRL-V-QA-I-LL--I-C----------I-RTPE-D-LIEQYA-IEW-CLKGWLE- 161
```

```
                                                                                                                  normal Nef end
H1B.FR.83.HXB2          TPGPG.......VRYPLTFGWCYKLVPVE.PDKIEEA.NKGENTSLLHPVSLHGMDD...PEREVLEWRFDSRLAFHH.............VARELHPEYFKNC.............................*   205
H1A1.UG.85.U455         -----.......I---------------D.-AEV---.TG---N-----ICQ--V--...E-K---M-K---T--LK-.............R-Y-----FY-D..............................-   206
H1B.US.90.WEAU160       -----.......T----C----F-------.-E-V---.-E---N-----M-------...H-K---M-K---K-----.............-----------D-............................-   204
H1C.ET.86.ETH2220       -----.......----------F-----D.-SEV--I.-E---NC----A-----E-...ED----K-K---H--RR-.............M--------Y-D-............................-   208
H1D.CD.84.84ZR085       -----.......I---------FE----D.-EVV-K-.TE--DNC----ICQ---E-...E-K---V---N-----E-.............K-K-KY-------............................-   216
H1F1.BE.93.VI850        -----.......I-----L---F-----D.-EEV-K-.-E---NC----M-Q---E-...ED----R-K---S--LR-.............I---R---FYQD-............................-   206
H1G.SE.93.SE6165        -----.......T-F-------F----MD.-AEV---.-----N-----ICQ---E-...ED----V-----S--RR-.............I--------Y-D-............................-   208
H1H.CF.90.056           -----.......E-F-------F-----N.-QEV-Q-.-E---N-----M-----E-...DG---M-K------LT-.............L--VK----.-D-............................-   207
H1J.SE.93.SE7887        -----.......I---------------D.-SEV---.-E---NC----ICQ--IE-...E----Q-K---S--RR-.............I------FY-D-............................-   204
H1K.CM.96.MP535         -----.......I---------------D.-AEV--T.TE--DNC----INQ---E-...EH--I-M-K---S--RR-.............-------D-Y-D-............................-   216
H101_AE.TH.90.CM240     -----.......I-F--C----F-----D.QREV--D.-----NC----M-Q--IE-...E-----M-K---A--RK-.............----Q----Y-D-............................-   205
H102_AG.NG.x.IBNG       -----.......T-F-------F----MD.-AEV---.-E---N-----ICQ---E-...DD----I--------R-.............T--------Y-D-............................-   207
H103_AB.RU.97.KAL153_2  .--IR........F--------------D.-AEV---.TE---N-----ICQ------...E-K---M-K------LT-.............R-------FY-D-............................-   203
H104_cpx.CY.94.CY032    -----.......E-F--C----F-----D.-QEV---.TE----C----I-Q---E-...E-----K-K------YK-.............--------FY-D-............................-   220
H1N.CM.95.YBF30         -----.......I---V-----F----LS.AEEV---.-E-D-NA----ICQ--A--...DHK---V-----S--RR-.............--------FY---............................-   213
H1O.BE.87.ANT70         -----.......T-F------LF-----S.EEEA-RLG-TC-RAN---ACA--FE-...THK-I-M-K--RS-GNT-.............--MIT---L-QKD............................-   214
H1O.CM.91.MVP5180       -----.......P-F------LF-----S.AEEA-RLG-TN-DA-----ACN--AE-...AHG-I-K-Q--RS-GLT-.............I-LQK---L-PK.............................-   212
CPZ.CD.90.ANT           -EE--.......-----CR---F-------.SPPD......D-RNI----ACT--DG-...-HK-I-R-E--AS-MRR-.............I---R-----RD............................-   199
CPZ.CM.01.SIVcpzCAM13   -----.......E---PC----F----LT.TEQV-A-.-E-D-NC----ICQ---E-...ESK---I--------LR-.............I---K----Y-D............................-   208
CPZ.CM.05.SIVcpzEK505   -----.......----------F----LS.-EEV---.-Q-D-NV----MCQ---E-...-DK---V-----S--RV-.............R-------FYQ--............................-   207
CPZ.CM.05.SIVcpzMB66    -----.......I---------F----L-.-EEV-R-.-E-D-NI----ICQ--QE-...EA----V-T------LK-.............R--------Y-D-............................-   207
CPZ.CM.98.CAM3          -----.......------Y-G-F----LT.-EEV-Q-.-E-D-NI----ICQ---E-...AHG---V--------LR-.............I---Q----YQ-.............................-   208
CPZ.GA.88.GAB2          -----.......------Y---F----S.-EAV---.-----DNL----LCT--FE-...E-K---I-K---Q--LR-.............L---K----YRD.............................-   206
CPZ.CM.98.CAM5          -----.......----------F----LT.-EEV-K-.-E-DTNI----ICQ---E-...EHG---V--------LK-.............L---Q----YGD...................HNPAPAQ-   214
CPZ.CM.05.SIVcpzMT145   -----.......----------F----LP.-EEV-K-.-E---NC----MCQ---E-...E-----V-T---K--LR-.............L---K----YRD.............................-   206
CPZ.TZ.00.TAN1          -----.......I----CR--LF-----D.....PP.EDD-KNI----ACS--TT-...-DG-T-I-----S--RR-.............I---RY-----..............................-   196
CPZ.CM.05.SIVcpzLB7     -----.......I---I---G-F----LT.-EEV---.-----NL----ICQ---E-...EDK---I-KY--Q--LR-.............I-----S--Y---............................-   207
CPZ.US.85.CPZUS         -----.......------Y---F----LT.EEEV-Q-.-----TNI---MCQ---E-...EHG---I-Q--TE--RR-.............R-K-------R-.............................-   210
CPZ.GA.88.GAB1          -T---.......T-F--C----F----LT.EEQV-Q-.-E-D-NC----ICQ---E-...EDK---V--------LR-.............I---Q----Y-D.............................-   206
GSN.CM.99.CN166         -----.......T---TIP-F-IC-R--A..TTEDSE.PGDDEYL-T--AYQGRSE-...QHK-F-VFS-C-K--IKSG.............IQLDQLQQEERKMRLTANRFL....................-   219
GSN.CM.99.CN71          -----.......I---TMP-F-WC-R--A..MTEDSE.PGDDQYL-N--AYQGQQE-...HH--I-VFS-C----LKSG.............WQMNQLQQEERKKRLTANRFL....................-   218
MUS_1.CM.01.1085        -D---.......T---KIP-FLWC-RE-A..MTEGSE.EGDDQYL-D--AFQGREE-...HHKQF-VFS-C----LKSG.............LQLDQMQQEERKKRLATNHIL....................-   218
MUS_1.CM.01.CM1239      -E---.......T---TLP-FLWC-RE-A..ITEDSE.EGD-EYL-T-AAYQGKEE-...-HKQF-VFS-C----MKSG.............RQLDQIQQEERKRRLTANRIL....................-   218
MUS_2.CM.01.CM1246      -D---.......T---RQP-FLWC-R--A..MTE-SE.PGDDQYL-T--AYVGRDE-...HHK-F-VFS-C----LKSG.............PQLNQIQQKERKRRLTANRIL....................-   218
MUS_2.CM.01.CM2500      -E---.......-----MP-VLFC-R--H..ITENSE.LGDDEHL-C--AFIGREE-...-HK-F-VFS-C----LKSG.............RQLNQMQQEERKRRLAANRILSQSP................-   222
                                 max HIV-1 similarity                                                                              normal Nef end
MAC.US.x.239            -S---.......I---K----LW-----N..VSD-AQ.ED.-EHY-M--AQTSQW--...-WG---A-K--PT--YTYE...AYVRYPEEFGSKSGLS-EEVRRRLTARGLL..NMADKKETR.........-   263
H2A.GW.x.ALI            -H---.......I---MF---LW-----D..VPQ-GE.DT.-THC-----QTSRH--...THG-T-V----PK--HDYK...AFILHPEEFGYKSGL--DEWKARLKARGIP..FSKNRNS..........-   263
H2A.DE.x.BEN            -H---.......----MY---LW---S--..LSQ-AE.ED.-ANC-V--AQTSRH--...EHG-T-V-Q---M--YNYK...AFTLYPEEFGHKSGL--KEWKAKLKARGIP..YSE..............-   258
H2A.SN.x.ST             -H---.......----KF---LW-----D..VPQ-GD.DS.-THC-V--AQTSRF--...-HG-T-V----PT---SYE...AFIRYPEEFGYKSGL--DEWKARLKARGIP..FS...............-   256
H2B.GH.86.D205          -H---.......I---KY---LW------..VPAATR.EEE-THC-M--AQISSW--...IHG-T-I-Q---L--YDYV...AFNRFPEEFGYQSGL--EEWKARLKARGIP..TD...............-   241
H2B.CI.x.EHO            -H---.......----KF---LW---IN..MIA-PE.DE.-THC-V--AQTSAW--...-HE-T-V-Q---L--YDYV...AFSRFPEEFGYQSGM--KEWKAKLRARGIP..TE...............-   239
H2G.CI.92.ABT96         -A---.......I---X----LW----T..VSD-AQ.ED.-THC-V--AQTSPW--...--T---A-K--PT--YDYR...AFILHPEEFRWKSGL--AVWKEKLKQRGLP..IE...............-   257
H2U.FR.96.12034         -S---.......T-W-MC---LW--E--D..ANDKAQ.ED.-RLY-VGSAQTSCEE-...HWG-A-V-K---S--YSYQ...AFIKCPEEFGSKSGLS-EEVKRRLTARGLP..VKNC.............-   261
MAC.US.x.251_1A11       -S---.......I---K----LW-----N..VSD-AQ.ED.-EHY-M--AQTSQW--...-WG--PA-K--PT--YTYE...AYVRYPEEFGSKSGLS-EEVRRRLTARGLL..NMADKKETR.........-   264
MAC.US.x.251_BK28       -S---.......I---K----LW-----N..VSD-AQ.ED.-RHY-MQ-AQTSKW--...-WG---A-K--PT--YTYE...AYARYPEE#GSKSGLS-EEVRRRLTARGLL..NMADKRETR.........-   263
MAC.US.x.EMBL_3         -S---.......I---K----LW-----N..VSD-AQ.ED.-RHY-MQ-AQTSKW--...-W#---T-K--PT--YTYE...AYVRYPEEFGSKSGLS-EEVRRRLTARGLL..NMADKRETR.........-   260
SMM.US.x.H9             -A---.......I---MF---LW--X--N..VSD-AQ.ED.-THC-M--AQTSQW--...-WG---A-K--PX--YXYK...AFVEHPEEFGSQSGLSKEEVQRRLTARGLL..KMADKKKTS.........-   262
SMM.US.x.PGM53          -S---.......I---MF---LW-----S..ASD-AQ.ED.-THY-V--AQISQW--...-WG---A-K---Q--YRYE...AFIRHPEEFGSKSGLS-EEVKRRLTARGLL..KMADKKETS.........-   264
SMM.US.x.SIVsmH635F_L3  -L---.......T---MY---LW-----D..VSD-AQ.ED.-THC-V--AQT-QW--...-WG---A-K--PE--YSYK...AFIKYPEEFGSKSGLS-EEVKRRLTARGIY..KMADKRETS.........-   264
SMM.SL.92.SL92B         -S---.......I---KF---LWQ----D..VSD-AN.-D.-THC-V--AQTYQHS-...-WGK--A-K-NPH--YTYE...AFVRHPEEFGWKSGLT-EEVERRLANKPKPQKKMADKKETS.........-   267
STM.US.x.STM            -A---.......I---KQ---LW-----D..MSN-AQ.EDDGTHY-V--AQT-QW--...-WG---V-K--PL--HTYE...AFVRHPEEFGSKSGL-KEEVERRLTARGLL..KMADKKETS.........-   265
MNE.US.x.MNE027         -S---.......P---K----LW-----N..VSD-AQ.EG.-ENY----AQTSQW--...-WG---V-K--PT--YTYE...AYIRYPEEFGSKSGLS-EEVRRRLTARGLL..KMADKKETS.........-   264
GRV.ET.x.GRI_677        S----.......I---RV--F-F-----D..LHE-AR.-C.-RHC-M--AQMGEDP-GI.DHG---V-K--PK--VEY...........RPDMFK..DMHEHAKR..........................-   224
VER.KE.x.TYO1_patent    SK---.......I-F-KC--F-F-----D..LHE-AQ.TC.-RHC-V--AQMGEDP-GI.SHG-I-V-K--PM--IQY...........DPNREYFTDMHGLVKRK.......................-   230
VER.KE.x.9063           ST---.......----RC--F-F------..MHE-AE.TC.-RHY-A--AQIKEDP-GI.SHG-T-V-K--PM--VQY...........DPNRQYFEDMHAIVKRK.......................-   230
VER.KE.x.AGM155         SK---.......I---RC--F-F-----A..LHE-AE.TC.-RHC-V--AQ--EDP-GI.NHG-I-A-K--PM--VQY...........DPSREYFTDLYSTVGTGN......................-   231
VER.DE.x.AGM3           SE---.......I---RC--F-F-----D..LHE-AE.TC.-RHC-V--AQVREDP-GI.NHG---V-K--PM--VQY...........DPNRKYLTDMHDLGKRK.......................-   230
TAN.UG.x.TAN1           S----.......T-K-RC--F-FE----D..VSQ-AQ.D..-RHC----AQIEWES-...-WK-T-V-K-NPL--VQY...........NPD..SFKDMHGLVKRK.......................-   223
SAB.SN.x.SAB1C          -D---.......T---KC----F-----D..LSE-AK.-S.--HC----AQVAYE--...AWK-T-V-K--PL--VDY.VAWRLHPEQVPSA.QG..................................-   227
OLC.CI.97.97CI12        EEVEE.....GTL-R-K-P--LWD-RR--.-LPETLEERG.............................................................................................   152
DRL.x.x.FAO             -K---.......I---KM--FLF------..VDEDLV.-Q.QCNR--NSSQMGPI--...--G-R-M-Q--PG--YTF.QACIHHPEEFGHVKSIKSVE-PG......CCKRKWWQFKPTTEGCHGDNLQKC-   269
RCM.NG.x.NG411          -----.......I---TM---LWQ----D..VSD-AT.EN.-RHC----AETS-----...-WG---A-K-KPL--VDY.AGYRLHPEFFGER.KNKTK...............................-   225
RCM.GA.x.GAB1           -----.......I---TL---LWQ----D..VSD-AR.ED.-EH-----AETS--E-...-WG---A-K-NPM--VDY.IGYRLHPEFFGERKNKTQ................................-   225
MND_1.GA.x.MNDGB1       -T---.......T---K---FLF--E--S.RAIGD-Y.AA..-NH---SSQ-CPQE-...--G-T-M-SGTLI-PMTL.QH...............................................-   215
MND_2.x.x.5440          -K---.......I---K---FLF-----A..VSPPLE.ED.-CNR--NSSQMGIQE-...--G-R-M-K---G--YTF.YAPIIRPEEYKCV.TSLSYEA-K-EEKPDCCKRKWWQF...............-   251
MND_2.CM.98.CM16        -D---.......I---K--RFLF--C--A..VPPDQE.-N.-CNK--QSSQ-GIQEE..#-WG-R-V-K---G--YTF.YAPIKRPGEFKHV.QSIS-IA-A-EHKPECCKRKWWQF...............-   231
MND_2.GA.x.M14          -S---.......I---KR--FLF--I--Q..VPPDQE.-Q.-CNR--NSSQ-GIQEE..#-WG-R-M-K--PE--YTF.YAPIKCPKEYRHV.TSLSYTA-QE..KSDCCKRKWWQF...............-   233
DEB.CM.99.CM40          -K---.......----IY--FLF------..IADPDY.END-RNI---DAHQGM-E-...-H--R-V-K---T--YCY.............K-GH.AEQR..EHTRRC.MFPKRK................-   228
DEB.CM.99.CM5           -E---.......P---TY--FLF------..IADPDY.END-RNI---DAHQGQAE-...-HK-R-V-K---S--YCY.............K-GHEAHKP..EHNRRC.MFPKRK................-   231
DEN.CD.x.CD1            -D---.......I----E--FLW----T..IEE-YD.--.-ENC---DRYEGQQA-...-W--T-V----PE--WCY.............K.AGHKKLQTSQHNNYCQRAKPRK................-   249
TAL.CM.01.8023          -E---.......T----C--ILW--C---..IHEDDT.ED.GH.L-M--AYDGQQE-...-WG-A-V-V--EK--YTP....................GAKMAEYDRLERKKKELLAPPQTAS..S......-   241
TAL.CM.00.266           -E---.......T----C--ILW--C---..IHEDDT.ED.GH.L-P--AYDGQAE-...-WG-S-V-V--EK--YTP....................GAKMAEWDRLEREKRMLLAPPQTASS........-   241
MON.CM.99.L1            -E---.......----TMP-FLWC-R--A..TTEDSE.EGD-DFL-T--AYQGR-E-...-H-QF-VFS-C-K--VKS....................GRQLAQLQQEERKKRLAANRIL............-   225
MON.NG.x.NG1            -X---.......----TIP-FLWC-R--A..IQEDSE.DGDXEFL-T--AYQGRDE-...-HKQF-VXS-C-K-GVKS....................GPQLDELQQEERKRRLTANRIL............-   229
WRC.CI.97.97CI14        -E---.......I-W-K-----W-----HFSEESDNTDPYY-RNK----ACA.EPE-...-WG-H-V-K--PK--CDFVAGRLPGQSIATGKE-I-SL......Y.K.NKKQCK..................-   234
WRC.CI.98.98CI04        -E---.......I-W-K-----W------FSEESNDSSQYY-KNK----GCQ.ESE-...-WG-H-V-K--P---CDFVASRLPGHIIAQGED-I-TL......H.KEKLKRHK.................-   234
WRC.GM.05.Pbt_05GM_X02  -D---.......----K-----W-----HFSEESPNSDQGF-KNL----AVA.ESE-...-WK-Y-V-K--P---CDFVAGRLADGQIATGIETL..Y......E.M.KKRNNK.................-   231
SUN.GA.98.L14           -K---.......----KA---LW--A--T..IDEDRD.PNHPCQA---SSQQGVNE-...-WG-R-I-T--PT--YDF.RAIQKHPEEFKHVTSLQWE......V...........................-   209
LST.CD.88.524           -D---.......T----A---LWQ-E--A..CDEFKD.PSD-RQC---SSQ-GVLE-...-WG-R-I-H-NPM--VDY.VALRKQP.................ASVQATAMRFNCERK.............-   207
LST.CD.88.485           -D---.......T--------LWQ-E--A..CDEYTD.PSDASQC---SSQ-GVQE-...-WG-R-I-H-SPM--VDF.VALRKQP.................KSIQASSLSLNCKRK.............-   208
LST.CD.88.447           -D---.......T-F--A---LWQ-E--A..CDEYTD.PSDYRQC---SSQ-GVQE-...-WG-R-I-H-NPM--VDY.VALRKQP.................KSIQASAIGL..QRK.............-   206
LST.KE.x.lho7           -D---.......T--------LWQ-E--A..CDEYKD.PSD-TQC---SSQ-GVLE-...-WG-R-I-H-NPM--VDF.IALKKQP.................AKIQNTAFAFDCKRK.............-   208
SYK.KE.x.SYK173         -S---.......T----I---LWE---N-..IEGCLE.YE.-H-L----A-GQ-SSSMGE-HV-LQPPPGYTPGWEMA.RLQLERQTGKPQE...LQSALSKN..IS....................-   241
SYK.KE.x.KE51           -S---.......T----I---LWE-A-N-..IEGYLV.DE.-D-LMM--AAGV-ASED..-H--N-M-N-NPH--YTPGWEMARQQLERQTGKR.....................................-   229
COL.CM.x.CGU1           EDEL-EDGALKED-K--VA--LW---YI-..QLG-Y-YSYDLSLLSVT......................................................SRKKKKPQQVAIEMVD......-   224
```