VII

Primate Lentivirus Proteins

Primate Lentivirus Protein Alignments

As has been noted in the PLV Complete Genome alignment (section IV), the phylogenetic relationships between genes, and thus their corresponding protein sequences, is not completely clear. See the tree and discussion in section IV.

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. 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* gene. 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/hiv-db/ALIGN_CURRENT/ALIGN-INDEX.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.”

Table 1: Table of sequences in the Primate Lentivirus (PLV) complete genome alignments

Name	Accession	Region	Author	Reference
SAB.SN.SAB1C	U04005	ENV GAG NEF POL REV TAT VIF VPR	Jin, MJ	*EMBO J* **13**(12):2935–47 (1994)
HIV1-B.FR.83.HXB2	K03455	ENV GAG NEF POL REV TAT VIF VPR VPU	Kantor, R	PLoS Med 2(4):e112 (2005)
HIV1-B.US.WEAU160	U21135	ENV GAG NEF POL REV TAT VIF VPR VPU	Wei, X	*Nature* **422**(6929):307–312 (2003)
HIV1-N.CM.YBF30	AJ006022	ENV GAG NEF POL REV TAT VIF VPR VPU	Kantor, R	PLoS Med 2(4):e112 (2005)
HIV1-O.BE.ANT70	L20587	ENV GAG NEF POL REV TAT VIF VPR VPU	Kantor, R	PLoS Med 2(4):e112 (2005)
HIV1-O.CM.MVP5180	L20571	ENV GAG NEF POL REV TAT VIF VPR VPU	Gurtler, LG	*J Virol* **68**(3):1581–5 (1994)
CPZ.CD.90.ANT	U42720	ENV GAG NEF POL REV TAT VIF VPR VPU	Vanden Haesevelde MM	*Virology* **221**(2):346–50 (1996)
CPZ.CM.01.SIVcpzCAM13	AY169968	ENV GAG POL REV TAT VIF VPR VPU	Nerrienet, E	*J Virol* **79**(2):1312–9 (2005)
CPZ.CM.05.SIVcpzEK505	DQ373065	ENV GAG NEF POL REV TAT VIF VPR VPU	Keele, BF	*Science* **313**(5786):523–6 (2006)
CPZ.CM.05.SIVcpzLB7	DQ373064	ENV GAG NEF POL REV TAT VIF VPR VPU	Keele, BF	*Science* **313**(5786):523–6 (2006)
CPZ.CM.05.SIVcpzMB66	DQ373063	ENV GAG NEF POL REV TAT VIF VPR VPU	Keele, BF	*Science* **313**(5786):523–6 (2006)
CPZ.CM.05.SIVcpzMT145	DQ373066	ENV GAG NEF POL REV TAT VIF VPR VPU	Keele, BF	*Science* **313**(5786):523–6 (2006)
CPZ.CM.98.CAM3	AF115393	ENV GAG NEF POL REV TAT VIF VPR VPU	Corbet, S	*J Virol* **74**(1):529–34 (2000)
CPZ.CM.98.CAM5	AJ271369	ENV GAG NEF POL REV TAT VIF VPR VPU	Muller-Trutwin, M	*J Med Primatol* **29**(3-4):166–72 (2000)
CPZ.GA.88.GAB1	X52154	ENV GAG NEF POL REV TAT VIF VPR VPU	Huet, T	*Nature* **345**(6273):356–9 (1990)
CPZ.GA.88.GAB2	AF382828	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	*ARHR* **20**(12):1377–81 (2004)
CPZ.TZ.01.TAN1	AF447763	ENV GAG NEF POL REV TAT VIF VPR VPU	Santiago, ML	*J Virol* **77**(3):2233–2242 (2003)
CPZ.US.85.CPZUS	AF103818	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	*Nature* **397**(6718):436–41 (1999)
GOR.CM.06.CP684	AM296489	POL	Van Heuverswyn, F	Nature 444(7116):164 (2006)
GSN.CM.99.CN166	AF468659	ENV GAG NEF POL REV TAT VIF VPR VPU	Courgnaud, V	*J Virol* **76**(16):8298–309 (2002)
GSN.CM.99.CN71	AF468658	ENV GAG NEF POL REV TAT VIF VPR VPU	Courgnaud, V	*J Virol* **76**(16):8298–309 (2002)
HIV2-A.DE.BEN	M30502	ENV GAG NEF POL REV TAT VIF VPR VPX	Kirchhoff, F	*Virology* **177**(1):305–11 (1990)
HIV2-A.GW.ALI	AF082339	ENV GAG NEF POL REV TAT VIF VPR VPX	Azevedo-Pereira, J	Unpublished (1998)
HIV2-A.SN.ST	M31113	ENV GAG NEF POL REV TAT VIF VPR VPX	Kumar, P	*J Virol* **64**(2):890–901 (1990)
HIV2-B.CI.EHO	U27200	ENV GAG NEF POL REV TAT VIF VPR VPX	Rey-Cuille, MA	*Virology* **202**(1):471–6 (1994)
HIV2-B.GH.86.D205	X61240	ENV GAG NEF POL REV TAT VIF VPR VPX	Kreutz, R	*ARHR* **8**(9):1619–29 (1992)
HIV2-G.CI.ABT96	AF208027	ENV GAG NEF POL REV TAT VIF VPR VPX	Brennan, CA	*ARHR* **13**(5):401–4 (1997)
HIV2-U.FR.96.12034	AY530889	ENV GAG NEF POL REV TAT VIF VPR VPX	Damond, F	*ARHR* **20**(6):666–72 (2004)

MAC.US.239	M33262	ENV GAG NEF POL REV TAT VIF VPR VPX	Kestler, H	*Science* **248**(4959):1109–12 (1990)
MAC.US.251_1A11	M76764	ENV GAG NEF POL REV TAT VIF VPR VPX	Planelles, V	*ARHR* **7**(11):889–98 (1991)
MAC.US.251_BK28	M19499	ENV GAG NEF POL REV TAT VIF VPR VPX	Franchini, G	*Nature* **328**(6130):539–43 (1987)
MAC.US.EMBL_3	Y00295	ENV GAG NEF POL REV VIF	Franchini, G	*Nature* **328**(6130):539–543 (1987)
SMM.SL.92.SL92B	AF334679	ENV GAG NEF POL REV TAT VIF VPR VPX	Chen, Z	*J Virol* **70**(6):3617–27 (1996)
SMM.US.H9	M80194	ENV GAG NEF POL REV TAT VIF VPR VPX	Fultz, PN	*PNAS USA* **83**(14):5286–90 (1986)
SMM.US.PGM53	AF077017	ENV GAG NEF POL REV TAT VIF VPR VPX	Novembre, FJ	*J Virol* **72**(11):8841–51 (1998)
SMM.US.SIVsmH635FC	DQ201174	ENV GAG NEF POL REV TAT VIF VPR VPX	Kuwata, T	*J Virol* **80**(3):1463–75 (2006)
SMM.US.SIVsmH635F_L3	DQ201172	GAG NEF POL REV TAT VIF VPR VPX	Kuwata, T	*J Virol* **80**(3):1463–75 (2006)
SMM.US.SIVsmH635SB10	DQ201173	GAG NEF POL REV TAT VIF VPR VPX	Kuwata, T	*J Virol* **80**(3):1463–75 (2006)
MNE.US.MNE027	U79412	ENV GAG NEF POL REV TAT VIF VPR VPX	Kimata, JT	*J Virol* **72**(1):245–56 (1998)
STM.US.STM	M83293	ENV GAG NEF POL REV TAT VIF VPR VPX	Novembre, FJ	*Virology* **186**(2):783–7 (1992)
DEB.CM.99.CM40	AY523865	ENV GAG NEF POL REV TAT VIF VPR VPU	Bibollet-Ruche, F	*J Virol* **78**(14):7748–62 (2004)
DEB.CM.99.CM5	AY523866	ENV GAG NEF POL REV TAT VIF VPR VPU	Bibollet-Ruche, F	*J Virol* **78**(14):7748–62 (2004)
RCM.GA.GAB1	AF382829	ENV GAG NEF POL REV TAT VIF VPR VPX	Gao, F	*Science* **300**(5626): 1713 (2003)
RCM.NG.NG411	AF349680	ENV GAG NEF POL REV TAT VIF VPR VPX	Beer, BE	*J Virol* **75**(24):12014–27 (2001)
DRL.x.FAO	AY159321	ENV GAG NEF POL REV TAT VIF VPR VPX	Hu, J	*J Virol* **77**(8):4867–4880 (2003)
MND-1.GA.MNDGB1	M27470	ENV GAG NEF POL TAT VIF VPR	Tsujimoto, H	*Nature* **341**(6242):539–41 (1989)
MND-2.CM.98.CM16	AF367411	ENV GAG NEF POL REV TAT VIF VPR VPX	Takehisa, J	*ARHR* **17**(12):1143–54 (2001)
MND-2.GA.M14	AF328295	ENV GAG NEF POL REV TAT VIF VPR VPX	Souquiere, S	*J Virol* **75**(15):7086–96 (2001)
MND-2.-.5440	AY159322	ENV GAG NEF POL REV TAT VIF VPR VPX	Hu, J	*J Virol* **77**(8):4867–4880 (2003)
GRV.ET.GRI_677	M66437	ENV GAG NEF POL REV TAT VIF VPR	Fomsgaard, A	*Virology* **182**(1):397–402 (1991)
TAN.FR.B87-14	DQ222475	NEF	Schindler, M	*Cell* **125**(6):1055–67 (2006)
TAN.UG.TAN1	U58991	ENV GAG NEF POL REV TAT VIF VPR	Stivahtis, GL	*Virology* **228**(2):394–9 (1997)
VER.DE.AGM3	M30931	ENV GAG NEF POL REV TAT VIF VPR	Baier, M	*Virology* **176**(1):216–21 (1990)
VER.KE.9063	L40990	ENV GAG NEF POL REV TAT VIF VPR	Hirsch, VM	*J Virol* **69**(2):955–67 (1995)
VER.KE.AGM155	M29975	ENV GAG NEF POL REV TAT VIF VPR	Johnson, PR	*J Virol* **64**(3):1086–92 (1990)
VER.KE.TYO1	X07805	ENV GAG NEF POL REV TAT VIF VPR	Fukasawa, M	*Nature* **333**(6172):457–61 (1988)
BKM.CD.AY518534	AY518534	POL	Takemura, T	*J Gen Virol* **86**:(7) 1967–71 (2005)
TAL.CM.00.266	AY655744	ENV GAG NEF POL REV TAT VIF VPR	Liegeois, F	*Virology* **349**(1):55–65 (2006)

TAL.CM.01.8023	AM182197	ENV GAG NEF POL REV TAT VIF VPR	Liegeois, F	*Virology* **349**(1):55–65 (2006)
TAL.US.1511	AY655742	POL	Liegeois, F	*Virology* **349**(1):55–65 (2006)
TAL.US.1867	AY655740	POL	Liegeois, F	*Virology* **349**(1):55–65 (2006)
TAL.US.416	AY655741	POL	Liegeois, F	*Virology* **349**(1):55–65 (2006)
TAL.US.742	AY655743	POL	Liegeois, F	*Virology* **349**(1):55–65 (2006)
MON.CM.99.L1	AY340701	ENV GAG NEF POL REV TAT VIF VPR VPU	Courgnaud, V	*J Virol* **77**(23):12523–34 (2003)
MON.NG.NG1	AJ549283	ENV GAG NEF POL REV TAT VIF VPR VPU	Barlow, KL	*J Virol* **77**(12):6879–88 (2003)
MUS-1.CM.01.1085	AY340700	ENV GAG NEF POL REV TAT VIF VPR VPU	Courgnaud, V	*J Virol* **77**(23):12523–12534 (2003)
LST.CD.88.447	AF188114	ENV GAG NEF POL REV TAT VIF VPR	Beer, BE	*J Virol* **74**(8):3892–8 (2000)
LST.CD.88.485	AF188115	ENV GAG NEF POL REV TAT VIF VPR	Beer, BE	*J Virol* **74**(8):3892–8 (2000)
LST.CD.88.524	AF188116	ENV GAG NEF POL REV TAT VIF VPR	Beer, BE	*J Virol* **74**(8):3892–8 (2000)
LST.KE.lho7	AF075269	ENV GAG NEF POL REV TAT VIF VPR	Hirsch, VM	*J Virol* **73**(2):1036–45 (1999)
SUN.GA.98.SOL-36	DQ222476	NEF	Schindler, M	*Cell* **125**(6):1055–67 (2006)
SUN.GA.L14	AF131870	ENV GAG NEF POL REV TAT VIF VPR	Beer, BE	*J Virol* **73**(9):7734–44 (1999)
BLU.KE.KE31	DQ222474	NEF	Bibollet-Ruche, F	*J Virol* **78**(14):7748–62 (2004)
SYK.KE.KE44	DQ222473	NEF	Bibollet-Ruche, F	*J Virol* **78**(14):7748–62 (2004)
SYK.KE.KE51	AY523867	ENV GAG NEF POL REV TAT VIF VPR	Bibollet-Ruche, F	*J Virol* **78**(14):7748–62 (2004)
SYK.KE.SYK173	L06042	ENV GAG NEF POL REV TAT VIF VPR	Hirsch, VM	*J Virol* **67**(3):1517–28 (1993)
COL.CM.CGU1	AF301156	ENV GAG NEF POL REV TAT VIF VPR	Courgnaud, V	*J Virol* **75**(2):857–66 (2001)

```
                           / Gag p17 Matrix start
SAB.SN.x.SAB1C            MGASNSV.LSGRKLDAFESVRLRPNGKKKYKLRHLVWASKELDRFSLSANLLETKEGVVKILSVLLPLVPTGSENLIALFNLCCVLACIHAEIKVKDTEEAKAKVKEEVPAEMTESATATSSG....QTK.....   125
H1B.FR.83.HXB2            ---RA--.---GE--RW-KI----G-------K-I----R--E--AVNPG----S--CRQ--GQ-Q-SLQ----E-RS-Y-TVAT-Y-V-QR-EI---K--LD-IE--Q....NK-KKKAQQA....AAD.....   121
H1B.US.90.WEAU160         ---RA--.---GE--RW-KI----G-N-----K-I----R--E--AVNPG----S--CRQ--GQ-Q-SLQ----E-RS-Y-TVA--Y-V-QK-E----K--LD-IE--Q....NK-KKKAQQA....AAD.....   121
H1N.CM.95.YBF30           ---RA--.-T-G---QW--IY---G-----RMK------R--E--ACNPG-MD-AD-CA-L-NQ-E-ALK----E-RS-Y-ALA--Y-V-SR-QIHN-Q--LD-I--KQ.EQHKPEPKNPEA-....AAA.....   124
H1O.BE.87.ANT70           ----A--.-T-S----W-QI--K-GS----R-K------R--E--ACNPE----A--NE-L-QQ-E-ALK---DS-QS-W-AIV--W-V-NRY-IG--QQ-IQ-L--VM....GSRKS-DAAK....EDT.....   121
H1O.CM.91.MVP5180         ---RA--.-T-S----W-RI----GS--A-R-K------R--E-YACNPG----A--TEQL-QQ-E-ALK----D-KS-W-AIA--W-V-NRFDIR--QQ-IQ-L--VM....ASRKS-EAAK....EET.....   121
CPZ.CD.90.ANT             ---GA--.-R-E---TW--I----G-----MIK-----RS--Q--A--SS----S--CE-AIHQ-S-SIEIR-PEI-S---TI---W-V-KGE-I----Q-VKT--MK-....MQTQAE-G-SQTASRGMLLRLL   130
CPZ.CM.01.SIVcpzCAM13     ---RA--.-T-G---RW-K-----G-R---MMK------R--E--ACNPG-M-SA--CL-L-QQ-E-ALK----G-RS---TLA--W-V-SN-E-E--QK-LEQL-KHH....G-QVRKNENTENNESRQ.....   125
CPZ.CM.05.SIVcpzEK505     ---RA--.-T-G---QW-KIY---G-----MMK------R--E--ACNPG-MD-A--CAQL-RQ-E-ALK----G-RS---TLA--Y-V-NN---QN-Q--LE-LR-KM....KAEQKEPEPEQ...AAG.....   122
CPZ.CM.05.SIVcpzLB7       ---RA-I.-T-G---DW-K-----G-R-R-MMK------R--E--ACNPG-M--T--CEQLITQ-E-ALK------KS---TIV--W-V-QR-AI---K--LDTLT.......AVRQKQQKEKE...AAA.....   119
CPZ.CM.05.SIVcpzMB66      ---RA--.-T-G---RW-KIY---G-----MMK------R--E--ACNPS-M--T--CKQL-QQ-E-ALK----G-RS---TIV--W-V-QG-P----K--LDQLQ.......EAQQKG.KQEV...AAA.....   118
CPZ.CM.05.SIVcpzMT145     ---RA-I.-T-GR---W-KI----G-----MMK------R--E--ACNPG-M-NA--CE-L-KQ-E-ALK----G-RS---TLA--W-V-KR-E-E--QQ-LQ-L-.......AAVNPEQGE............   114
CPZ.CM.98.CAM3            ---RA--.-T-GR---W-KI----G-----MMK------R-----ACNPG-M--A--CEQL-RQ-E-ALK----G-RS---TLA--W-V-KR-S-E--QQ-LT-L--A-.....-A-REQEVAQ....PQQ.....   121
CPZ.CM.98.CAM5            ---RA--.-T-GR---W-KI--X-G-----MMK------R--E--ACNPG-M--A--CENL-KQ-E-ALR----G-RS---TLA--W-V-KR-Q-E--QQ-LS-L--T-....AAQE.QEVAQ....PQQ.....   120
CPZ.GA.88.GAB1            ---RA--.-T-G---RW-K-----G-R-R-MMK------R--E--ACDPG-M-S---CT-L-QQ-E-ALK----G-RS---TLA--W---SD-T-E--QK-LEQL-RHH....G-QQSK-E-N....SGSREGGA   126
CPZ.GA.88.GAB2            ---RA-I.-T-GR---W-KI----G-R-R-MMK-I----R--E--ACDPG-M-----CYQL-KQ-E-ALK----G-RS---TLA--W-V--R-E-E--QQ-LQ-L--F-....QKETATAQEA....AA-.....   121
CPZ.TZ.01.TAN1            ---RA--.-R-D---TW--I--KSR-R---LIK-----GS--Q--AMNPG-M-NV--CW--ILQ-Q-S-DI--PEI-S---TI---Y-V--GER-Q-----VKI--MKL....-VQKNNSTAT..SSGQR....Q   124
CPZ.US.85.CPZUS           ---RA--.-T-GR---W-KI----G-----MMK------R--E--ACNPG-M--AD-CLQL-KQ-E-ALK----G-RS---TLA--W-V-SRVT-E--QQ-LV-L--V......VQLQKQQE.....EKEQ....   119
GSN.CM.99.CN166           ---RH-AM-T-T---RY-K-----K-----MIK---------E-YA--DA---NQ--CRR--EIVF--E-----A-KS--GIVS--Y----D-Q-E---Q--QQ-RIRCHL-E-KGKEKQNKD...........   123
GSN.CM.99.CN71            ---RH-AM-T-T---RY-K-S---R-----AIK---------E-YA--GS---S---CQQV-E-TS--E-N---S-KS-YGIVS--Y----D-W-E---Q--KQ-QIRCHLVG-KEKETQNKD...........   123
H2A.DE.x.BEN              ---R---.-R-K-A-EL-K-----G-----R-K-I---AN---K-G-AES---S---CQ---R--D----------KS---TV--IW-L---E---------KLA.QRHLVAE.............TGT.....   114
H2A.GW.x.ALI              ---R---.-R---A-EL-RI----G-----Q-K-I---AN-----G-AES---S---CQR--K--E----------KS---TV--VW-V---E------G--QII.QRHL-AE.............IET.....   114
H2A.SN.x.ST               ---R---.-R-K-A-EL-KI----G-----R-K-I---AN-----G-AES---S---CQ---T--D----------KS---TV--IW-----E-A-------Q--.QRHLVAE.............TKT.....   114
H2B.CI.x.EHO              ---RG--.---K-T-EL-K-----G---R-M-K-I---VN--E--G-AESR-GS---CR--RK--G----------KS-Y-TV--IF-L---E---------KIA.QRHL-AD.............TE-.....   114
H2B.GH.86.D205            ---RG--.---K-T-EL-K-----G-----M-K-V---VN-----G-AES---S---CQ---K--A----------KS---IV--IF-L---E---------KIA.QRHL-AD.............TE-.....   114
H2G.CI.x.ABT96            ---RS--.---K-A-EL-K-----G-----M-K-V---AN-----G-AES---S---CQ------A----------KS-Y-TV--IW-L---Q---H-----QV-.QRHLVVE.............TGT.....   114
H2U.FR.96.12034           ---R-X-.---K-V-EL-K-----------M-K-V---AN-----G-AES---A---CQ------A----------KS-Y-TV--IW-L---E---H-----QI-.QRHL-VE......................   111
MAC.US.x.239              --VR---.---K-A-EL-KI----------M-K-V---AN-----G-AES---N---CQ------A----------KS-Y-TV--IW-----E---H-----QI-.QRHLVVE.............TGT.....   114
MAC.US.x.251_1A11         ---R---.---K-A-EL-KI----------M-K-V---AN-----G-AES---N---CQ------A----------KS-Y-TV--IW-----E---H-----QI-.QRHLVVE.............TGT.....   114
MAC.US.x.251_BK28         ---R---.---K-A-EL-KI----G-----M-K-V---AN-----G-AES---N---CQ------A----------KS-Y-TV--IW-----E---H-----QI-.QRHLVVE.............TGT.....   114
MAC.US.x.EMBL_3           ---R-A-.---K-A-EL-KI----G-----M-K-V---AN-----G-AES---N---CQ------A----------KS-Y-TV--IW-----E---H-----QI-.QRHLVVE.............TGT.....   114
SMM.SL.92.SL92B           ---RG--.---K-A-EL-K-----G-R---M-K-II--AR-----GSAES---S---CQR--A--A--M-------KS--STV--VW-L---M---------KT-.QSHLVVE.............SGT.....   114
SMM.US.x.H9               --VR---.---K-A-EL-KI----G-----M-K-I---AN-----G-AES---N---CQ------A----------KS-Y-TVR--W-----E---H-----QI-.QRHLVVE.............TGT.....   114
SMM.US.x.PGM53            ---R---.---K-A-EL-KI----------M-K-V---AN-----G-AES-------CQ--I---A----A-----KS-Y-TV--IW-----E---H-----QI-.QRHLVVE.............TGT.....   114
SMM.US.x.SIVsmH635FC      ---R---.---K-T-EL-KI----------M-K-V---AN-----G-AES--DN---CQ------A----------KS-Y-TI--IW-----E---H-----QI-.QRHLVVE.............TGT.....   114
SMM.US.x.SIVsmH635F_L3    ---R---.---K-T-EL-KI----------M-K-V---AN-----G-AES--DN---CQ------A----------KS-Y-TI--IW-----E---H-----QI-.QRHLVVE.............TGT.....   114
SMM.US.x.SIVsmH635SB10    ---R---.---K-T-EL-KI----------M-K-V---AN-----G-AES--DN---CQ------A----------KS-Y-TI--IW-----E---H-----QI-.QRHLVVE.............TGT.....   114
MNE.US.x.MNE027           ---R---.---K-A-EL-KI----G-----M-K-V---AN-----G-ADS---N---CQ----A-A----------KS-Y-TV--IW-----E---H-----QI-.QRHLVVE.............TGT.....   114
STM.US.x.STM              ---RS--.---K-A-EL-K-----G-----M-K-V---AN-----G-AES---S---CQ--IT--E----------KS---TV--IW-----E---H-----QV-.KRHLVVE.............TGT.....   114
DEB.CM.99.CM40            ---GR-AP-T-AA--R--K-----K-----MIK-----G------G-A-A---N---CQE--TR-Q--EA------KS-YGIV--IWAC--NLS-D------KA-YRRIKQVE.............QEEELEMAL   122
DEB.CM.99.CM5             ---GR-AP-T-AA--R--KI----K-----LIK------R-----G-AGA-------CQE--TR-Q--EAS---G-KS-YGIV--VWAC--NLS-E------KA-ARRIKQAE.............QEEELEMAL   122
RCM.GA.x.GAB1             ---RA-L.---K----W-------G-----M-K-----C-K-NK-G--DH---AT-CE---G-----------G-KS--------W-V-K-V--------V-----CCHLVEKAEN-TEKEK....GAT.....   125
RCM.NG.x.NG411            ---RA-I.---K----W-Q-----G-----M-K-----CR--E--G--DT---SA--IQ--IG-I---------G-KS--------W-V-K-V--------V-H--KQCHLVDKDENAGEQEK....GAT.....   125
DRL.x.x.FAO               ----A-G.-R-E------K-----S-------K----VA------G-HER---SQ--CQ---G--F------------S---T---VW-V-SKV--T-----VQ---QHCHLVDKNENA-SKNENGETA-SSGR..   132
MND_1.GA.x.MNDGB1         --NG--A.-L-TD--K--KI--KRG---C-R-K--C-CKG-----G--DK----QQ-CE-----CW--YDQ--D--K--VGTV--V-----G-EI-S-QD-LK-L-VIT.........................   108
MND_2.CM.98.CM16          ----A-G.-R-E---EL-KI----S-----Q-K-II-V-------G-HEK---S---CE------F-----------S-Y-T--CVW-V--KE--T------E---QKLHLVAEKENA-SEKEQRA.IVT.....   128
MND_2.GA.x.M14            ----A-G.-R-E---EL-TI----S-----Q-K-II-V-------G-HEK----Q--CE------F---------L--Y-T--CIW-V--KV--T------K--EQCFHLAAKGESA-SEKEEKA.TAT.....   128
MND_2.x.x.5440            ----A-G.-R-E---EL-KI----S-----Q-K-VI-V-------G-HEK---SQ--CE------F-----------S-Y-T--CIW-V--KV--T------E---QRYHLVVERENA-SEEEKGA.TAT.....   128
GRV.ET.x.GRI_677          --GGH-A.----S--T--KI----------QIK--I--G--ME--G-HEK-------CQ--IE--T--E-----G-K--------IW-----Q--------VVT--QHYHLVDKNEKA-KKKNE...T-A.....   126
TAN.UG.x.TAN1             ---GH-A.----N--T--KI----------Q-K--I--G-KME--G-HEK-------CQ--IE--S--E-----G-KS-------IW-----Q--------VVL--QRCHLVEK-KTA-AP--....GQQ.....   125
VER.DE.x.AGM3             ---AT-A.-NR-Q--K--HI----T-----QIK--I--G--ME--G-HER---SE--CK--IE--Y--E-----G-KS----V---F-V-KDKE-------V-I-RQCCHLVEK-RNAERNTT....E-S.....   125
VER.KE.x.9063             ---AT-A.-K--Q--Q--KI----------QIK--I--G--ME--G-HDR---ED-CK--IE--F--E-----G-KS----V---Y---R-Q--------V-T-RQCCHLVEK-KRATEP-S....GE-.....   125
VER.KE.x.AGM155           ---AT-A.-NR-Q--E--HI----------QIK--I--G-KM---G-HEK----E--CK--IE--S--E-----GMKS-Y--V---L-V-Q-K--------L-I-RQCCHLVDK-KTAV-PP-....GQQ.....   125
VER.KE.x.TYO1             ---AT-A.-NR-Q--Q--KI----------QIK--I--G--ME--G-HER----E--CKR-IE--Y--E-----G-KS----V---Y-L-K-Q--------V-T-RQHCHLVEK-KSATET-S....GQ-.....   125
TAL.CM.00.266             ---RH-AM-T--A--KY-K-----K-----MVK-I------ME--G--DA-------C---IEAIYM-E-N---GIKS--GIV---Y----G-DIE---Q---E-RKRCHLGEKKET-EEEK.............   122
TAL.CM.01.8023            ---RH-AM-T--A--KY-K-----K-----EV--V------ME--G--DA-------C---IEAIS--E-N---GIKS--GIV---Y----G-EIE---Q---E--KRCHPVEKKEESKGTTE....-GA.....   126
MON.CM.99.L1              ---RH-AM---T---KY-K-----R-----LIK-I---A------G--DS----QD-CK---E-I---Q-----SIKS--GIAS--Y----G-EIE---Q--QQ--IRCHLAG.....................   113
MON.NG.x.NG1              ...........................................................................................................................................     0
MUS_1.CM.01.1085          ---RH-AM-T-N---RY-K-----K-----LIK-II---R-----G--DS---SRD-CQ--IE-I---E-A---S-KS--GIVS--W-----VH-E------KQIRARCHL-EKKGE.................   117
LST.CD.88.447             --SG---.--RQIEGE-C-I----DS--T-QK------TR-----G-G-H----AD-CK---G-CW--YS---K--K--VGTV---C-C-LG---A--Q--MQ-I-I-PTQ-ERQ...................   114
LST.CD.88.485             --SG---.--RQIEGE-C-I----GS--T-QK--I---T------G-G-H----AD-CKR--G-CW--YS---K--K--VGTV---C-C-LG--IAN-A--MQ-I-I-PTQ-DRQ...................   114
LST.CD.88.524             --SG---.--RQIEKD-CN-----GS--T-QK--VD--T------G-G------A--CK---G-RW--Y----K--K--VGTV--IC-C-LG-R-N--Q--IN---I-QNKPV.....................   112
LST.KE.x.lho7             --SG---.--RQIEKD-C------GS--T-QK--VE--T------G-GSQ----A--CK-----CW--YA---K--K--VGTV--IC-C-LGVRIS--Q--IS---I-PAP-AAG...................   114
SUN.GA.98.L14             ---G--T.VDREVVRS--R-A-K-G---T-QA--V---G------G-NKE--R-V--CQ-----CW--YAS--K--K--VGTV-----C-QG-P----Q--LK--RL-PAKNEE....................   113
SYK.KE.x.KE51             ---GG-AI-T-----KY-KI----K---R-LV------G------G--DQ---S---CEQ-IKTI---EKH-----KS--GITA-VWAV---KD-E---Q--Q----ACNW-DE-TV-SSGQK....ENS.....   126
SYK.KE.x.SYK173           ---AG-AI-T--E--RY-KI----K---R-LV------K------G--DQ-M-S---CE---T-----EAN-----KS--GIIS-VWAV--KKE-E---Q--Q----ACNWKDDPP--SGGQS....ENS.....   126
COL.CM.x.CGU1             --NEQGL.-GKKT-EDLQK---KKGK-GC--IK-VR-MCT-VS-CV-IFE--KSAT--AQ--EKVT---D----V-RS-YG.-S-CY-L-RKWNIE--Q--EK--E-AYKKQAMIEMA.................   116
```

```
                                                       Gag p17 Matrix \ / p24 Capsid                                                       Cyclophilin A binding site |
SAB.SN.x.SAB1C          ..ELQAKKKNEPTV.........TPS.GGS...RNY.PIVSVN.NQWVHQPLSPRTLNAWVKVIEEKKFSAEVVPMFSALAEGAIPYDINQMLNAVGEHQGALQIVKDVINEEAADWDLRHP.P.PQQPPAQGVL   241
H1B.FR.83.HXB2          ..TG...................HSN.QV-...Q--.---QNIQG-M---AI-----------V---A--P--I------S---T-Q-L-T---T--G--A-M-ML-ET------E--RV--.V.HAG-I-P-QM   228
H1B.US.90.WEAU160       ..TE...................NRN.QV-...Q--.---QNLQG-M---A----------------A--P--I------S---T-Q-L-T---T--G--A-M-ML-ET------E--RL--.V.HAG-I-P-QM   228
H1N.CM.95.YBF30         ..ATD....................-.NI-...---.-L-QTAQG-M-----T--------------A--P--I---M--S---T-S-L-T---T--G--A-M-ML-E----------RT--.V.-VG-LPP-Q-   230
H1O.BE.87.ANT70         ..S.....................AR.QAG...Q--.----NAQG-M---AI----------AV---A-NP-II---M--S----S----T----I-G------VL-E------VE--RT--.-.-VG-LPP-QI   226
H1O.CM.91.MVP5180       ..S.....................-R.QT-...Q--.---TNAQG-M---AI----------AV---A-NP-II---M--S---V-----T----I-G------VL-E-------E--RT--.-.AMG-LPP-QI   226
CPZ.CD.90.ANT           LLNK-..WCQR.............HLS-EG...---.--IVDAGGIAR----T---------CV---N-NP--I------S---T-H-L-T------D----M-VL-E-------E--RL--.T.HAG-VQA-Q-   244
CPZ.CM.01.SIVcpzCAM13   ..GAS-...................-S-T-...G--.---QNAQG-M--T-I-----------V---A-NP--I------S---L-L-V-T------G----M-VLRE-------E--RM--.V.HAG-I-P-Q-   233
CPZ.CM.05.SIVcpzEK505   ..AAA-..................-ESSI-...---.-L-QNAQG-M----------------V---A-NP--I---M--S---T-Q-L-T---T--G--A-M-ML-E-------E--RG--.V.HMG-IPP-QV   231
CPZ.CM.05.SIVcpzLB7     ..AAAR..................DNDS--...---.---TNAQG-P---A------------V---A--P--I---A------T-Q-L-T---T--G--A-M--L-E------QE--RL--.I.HMG-I-P-QM   228
CPZ.CM.05.SIVcpzMB66    ..TADG...................TSTV-...--F.---ANAQG-M----I-----------V---A--P--I---M--S---T-Q-L-T---T--G--A-M-ML-E-------E--RL--.V.HAG-VPP-QM   226
CPZ.CM.05.SIVcpzMT145   .........................TPSA-...A-F.-V-RNAQG-LT-N-I----------AV---A-NP--I---M--S---T-Q-L-T---TI-G--A-M--L-E---D---E--RT--.T.HAG-I-P-QM   218
CPZ.CM.98.CAM3          ..QQ-...................D-.AV-...---.-V-QNAQG-L----M---------------N-NP--I---M--S---T-Q-V-T----I-G----M-VL-E-------E--RT--.V.HAG-V-P-Q-   228
CPZ.CM.98.CAM5          ..QQ-....-..............S-..I-...---.---QNAQG-MI--AM---------------N-NP--I---M--S---T-Q-V-T------G----M-VL-E-------E--RL--.V.HAG-I-P-Q-   227
CPZ.GA.88.GAB1          SQGAS-....S.............AG..I-...G--.-L-QNAQG-M---AI-----------V---A--P--I------S---L-Q-V-T------G----M-VL-E-------E--RL--.T.HAG-I-P-Q-   236
CPZ.GA.88.GAB2          ..QSE..................SNTTEI-...---.-V-QNAQG-MI--AM-----S----AV---A--P--I---M--S---T-Q-V-T----I-G----M-VL-E-------E--RL--.L.HAG-V-P-QM   230
CPZ.TZ.01.TAN1          NAGEKEETVPPSGNTGNTGRATE---.-.-...-L-.-VITDAQGVAR---I---------R-----G-NP--I------S---T---L-S---------A-M-ML-E-------E--RA--.A.HAG-QQA-M-   251
CPZ.US.85.CPZUS         ..QQ-....-ASG...........SN.I--...S--.-VIQNAQG-M---AM----------AV---A-NP--I---M--S---T-Q-V-T----I-G----M-VL-EA-------*-RT--.L.HAG-I-P-QM   230
GSN.CM.99.CN166         ..................PPGAAGGQAVV-...Q--.-VIRNAQG-YQ--A-NS-I-KT--SIV-----AP-T-AL-Q--S--C----M------I-DY---V--I------Q--E---L--.V.-AA-QPVAG-   234
GSN.CM.99.CN71          ..................PPGAAGGQAVV-...Q--.---RNAQG-FQ--A-NS-V-KT--SIV-----AP-T-AL-Q--T--C----M--------DY---V--I------Q--E---L--.Q.-AA-QPVAG-   234
H2A.DE.x.BEN            ..AEKMPNTSR--A.........P--.-KR...G--.-VQQAG.GNY--V------------LV-----G-----G-Q--S--CT---------C--D--A-M--IREI---------SQ--.I.-.G-LPA-Q-   229
H2A.GW.x.ALI            ..AEKMPSTSR--A.........P--.EQG...G-F.-VQQ-A.GNYT-V-------D----LV-----G-----G-Q--S--CT---------C--D--A-M--IREI---------VA--.I.-.G-LPA-Q-   229
H2A.SN.x.ST             ..TEKMPSTSR--A.........P--.-NG...G-F.-VQQ-A.GNYT-V------------LV-----G-----G-Q--S--CT---------C--D--A-M--IREI---------AQ--.I.-.G-LPA-Q-   229
H2B.CI.x.EHO            ..MPAMS-PSK.............-T.SR....LA-.-VQQIA.GNYS-L------------LV-----G-----G-Q--S--CT---------C-----A-M--IREI---------QQ--.S.-.G-MPA-Q-   224
H2B.GH.86.D205          ..MPATN-PTA............P--...G...G--.-VQQLA.GNY--L------------LV-----G-----G-Q--S--CT---------C-----A-M--IREI---------QQ--.S.-.G-MPA-Q-   224
H2G.CI.x.ABT96          ..TEKVPATSR-IA..........-PS-RG...G--.-VQQ-G.GNY--L------------LV-----G-----G-Q--S--CL---------C-----A-M--IREI------X---Q--.Q.Q.G--PA-QX   229
H2U.FR.96.12034         ..TET-E-MPATSRPTX......P--G-R....G--.-VQQIG.GNYA-L------------LV-----G-----G-Q--S--CT---------C-----A-M--IREI--D------TQ--.N.Q.G--PA-Q-   229
MAC.US.x.239            ..TETMP-TSR--A.........PS-.-RG...G--.-VQQIG.GNY--L------------L------G-----G-Q--S--CT---------C--D--A-M--IR-I----------Q--.Q.-A.-.Q--Q-   228
MAC.US.x.251_1A11       ..AEIMP-TSR--A.........PS-.-RG...G--.-VQQIG.GNY--L------------L------G-----G-Q--S--CT---------C--D--A-M--IR-I----------Q--.Q.-A.-.Q--Q-   228
MAC.US.x.251_BK28       ..AETMP-TSR--A.........PS-.-RG...G--.-VQQIG.GNY--L------------L------G-----G-Q--S--CT---------C--D--A-M--IR-I----------Q--.Q.-A.-.Q--Q-   228
MAC.US.x.EMBL_3         ..AETMP-TSR--A.........PS-.-RG...G--.-VQQIG.GNY--L------------L------G-----G-Q--S--CT---------C--D--A-M--IR-I----------Q--.Q.-A.-.Q--Q-   228
SMM.SL.92.SL92B         ..AEKLPAQSR--A.........P--...G...G--.-VQQ-G.-NY--T------------LV-----G-----G-Q--S--CT---------C-----A-M--IREI----------Q--RG.Q-PAQPA-G-   229
SMM.US.x.H9             ..ADKMPATSR--A.........P--.-RG...G--.-VQQ-G.GNYT-L------------LV-----G-----G-Q--S--CT-X-------C-----A-M--IREI----------Q-X.Q.-.G-IPA-Q-   229
SMM.US.x.PGM53          ..ADKMPITSR--A.........P--.-RG...G--.-VQQIG.GNYT-L------------LV-----G-----G-Q--S--CT---------C-----A-M--IREI----------Q--.Q.-.G-IPA-Q-   229
SMM.US.x.SIVsmH635FC    ..ADKMPATSR--A.........P--.-RG...G--.-VQQ-G.GNY--L------------LV-----G-----G-Q--S--CT---------C-----A-M--IREI----------Q--.Q.-.G-LPA-Q-   229
SMM.US.x.SIVsmH635F_L3  ..ADKMPATSR--A.........P--.-RG...G--.-VQQ-G.GNY--L------------LV-----G-----G-Q--S--CT---------C-----A-M--IREI----------Q--.Q.-.G-LPA-Q-   229
SMM.US.x.SIVsmH635SB10  ..ADKMPATSR--A.........P--.-RG...G--.-VQQ-G.GNY--L------------LV-----G-----G-Q--S--CT---------C-----A-M--IREI----------Q--.Q.-.G-LPA-Q-   229
MNE.US.x.MNE027         ..AETMP-TSR--A.........PS-.-KR...G--.-VQQIG.GNYT-L------------L------G-----G-Q--S--C----------C-----A-M--IREI----------Q--.Q.QA.-.Q--Q-   228
STM.US.x.STM            ..ANKMPATSR--A.........P--.-RG...G--.-VQQ-G.GNY--L------------LV-----G-----G-Q--S--CT------L--C-----A-M--I-EI---------MQ--.Q.-PG-LPA-Q-   230
DEB.CM.99.CM40          IQRQKE-EVEQKQQKQQEQRQQEPQAAPQTTPSV--.--LRQG.Q-F--M-----IVKT-INAV------P-I--L-QV----CT-----GL---I-DL---M--I---------E---Q--.Q.Q-.-.P--Q-   251
DEB.CM.99.CM5           IQKKAE-E.L-KKQK...TQAAAADTQQKTNQPA-F.--LRQG.Q-Y--M-----IVKT-INAV------P-I--L-QV----CT-----GL---I-DL---M--I---------E---Q--.Q.Q-.-.P--Q-   247
RCM.GA.x.GAB1           ..........................APP-GQRG--.--ITI-.Q-PE-N-I-----------V---------A------S--C-----------I-----------E----------A---.V.-.G-IPA-Q-   229
RCM.NG.x.NG411          ..VT...................SSG.QR....G--.---TI-.Q-PE---I-----------V-----G----------S--C----V------I-----------E---D----------.N.-.G-LPA-Q-   229
DRL.x.x.FAO             .............................-...---.--QV--.Q-A---AI-------------------------L--S--C-S---------I-D----------I---------R---.Q.VGP.LP----   230
MND_1.GA.x.MNDGB1       ...................RKEE.KQEDE-...K-F.-VQRDAAG-YQYT-I---IIQT---TV----WKP--I-L----T----SH-L-I------D----M-VL------Q--E---T--.Q.Q-PAQPG-G-   217
MND_2.CM.98.CM16        ......PSGR...................-...K--.--QII-.QTP---GI----------C-------P-I----I--S--CL---L-G----I-D-------------------N----.Q.V.G-MP----   230
MND_2.GA.x.M14          ......PSGR...................-...K--.--QII-.QTP---GI----------C-------P-I----I--S--CL---L-G----I--------------------------.Q.V.G-LP----   230
MND_2.x.x.5440          ......PAVR...................-...K--.--QVI-.QTP---GI----------C-------P-I----I--S--C----L-G----I-D------------------------.-.V.G-MP----   230
GRV.ET.x.GRI_677        .......................P-..--ES.----.-V-NQ-.-A----------------CV---RWG-------Q--S--CLS--V-----VI-D-------L-E-------E--RT-R.-.-AG-LPA-Q-   230
TAN.UG.x.TAN1           ..QNYNTAA..............---.-RH...G--.-V-QQ-.-----T------------TV---R-G--I----Q--S--CLS--------VI-D----M--I-E---D---Q--IT--.-.-AG-LPA-Q-   236
VER.DE.x.AGM3           ..SG-K-NDKGV--..........-P.----..Q-F.-AQQQG.-A-I-V------------AV-----G--I----Q--S--CT---------VL-D---------EI------Q--IA--.-.-AG-LPA-Q-   240
VER.KE.x.9063           ..KSNRETTA.............P-G.-.-...Q-F.-AQQQG.-A---V------------AV-----G--I----Q--S--CT---------VL-D---------EI------Q--VT--.-.-AG-LPP-Q-   236
VER.KE.x.AGM155         ..KNNTGGTA.............--G.-.-...Q-F.-AQQQG.-A---V------------AV-----G--I----Q--S--CT---------VL-D---------EI------Q--VT--.-.-AG-LPA-Q-   236
VER.KE.x.TYO1           ..KNDKGIAA.............P-G..C-...Q-F.-AQQQG.-A---V------------AV-----G--I----Q--S--CT---------VL-D---------EI------Q--VT--.L.-AG-LPA-Q-   236
TAL.CM.00.266           ......................VRKATVT-SGQTGNY--IRNAQQ-YQ--A----L-KT-IAT------AP-T-AL-------C------------------M--I--I---Q-SE--QM--.Q..IG-LPA-Q-   232
TAL.CM.01.8023          ......................VRSAIVT-SGHSGNY--IRNAQQ-YQ--A----L-KT-IAT------AP-T-AL-------C------------------M--I--I---Q--E--QM--.Q..IG-LPA-Q-   236
MON.CM.99.L1            ..-QGEQ-AAAAAA.........P-T.-.GVPSG--.-V-RTQGGGFQ--AVE--L-KT--Q-------AP---AL-Q--S--M-------L---I-D-------I---------N---V--.QP--.-QPNAG-   232
MON.NG.x.NG1            .......................................................................................................................................     0
MUS_1.CM.01.1085        .......................KE-.QAQAASA--.-VIRNAQG-FQ--A-NA-I-KT--SIV-----A---A---Q-------------L---I------I--I------Q--E---LN-.QPQ-.-QPNAG-   225
LST.CD.88.447           ...........................PK-...G--.-L-RE-.QR---T------IQT---IV-DRGWKP-T-A-----T-K-L-D-L-V----I-D----M--I--H-V--G-E--RK--.Q.Q-PAQPG-G-   215
LST.CD.88.485           ...........................QK-...G--.-LIRE-.QR---T------IQT---IV-DRGWKP-T-A-----T-K-L-D-L-V----I-D--V-M--I--H-V--G-E--RT--.Q.Q-PAQPG-G-   215
LST.CD.88.524           ...........................PE-...K-F.-LIRE-.QR---T------IQT---IV-DRGWKP-T-A-----T-K-L-E-L-V----I-D----M--I--H-A--G-E--RQ--.Q.Q-PAQPG-G-   213
LST.KE.x.lho7           .......................KKQ.QTG...G--.-LIRE-.QR---T------IQT---IV-DRGWKP-T-A-----T-K-L-D-L-V----I-D----M--I--H-V--G-E--RQ--.Q.Q-PAQPG-G-   218
SUN.GA.98.L14           .........................A.QA-...Q-F.-VQREG.QNYI--------VQT---IV---GWKP-T-A--A--MT----E---V----I------M-MI--H-V--S-E--RM--.Q.Q-PAQPG-G-   215
SYK.KE.x.KE51           ........NDTA-SSGREGKMQLPAAMPS-GGSG--.-LIRNPQ---I-VGVNT---KT--EAVNS---D-SL--L-QI-T--F----L-D----I-D----M-VI-------GSE---Q--.Q.---QQPVAG-   250
SYK.KE.x.SYK173         ..SQNMASETSSGQKVVQQEKQKAAT.PPPR..G--.-LLRNPQ---I-TGVPV---KT--EAVNS---D-SI--L-Q--T--F----L-G------D----M-VI--I----G-A---E--.Q.----AP-AG-   253
COL.CM.x.CGU1           ...................SKEEEKAKKEAE..KLDM---TGP.QGP----------G----CV-.GGIAPSLA---L-YST---A--M-L---ILDT---FP-VL--E--KK-EEY--L--.V..---QQ--A-   225
```

```
SAB.SN.x.SAB1C          RDPQGSDIAGTTSTIQEQIEWTTRA....QNAVNVGNIYKGWIILGLQKCVKMYNPVNILDIKQGPKEPFKDYVDRFYKALRAEQTDPAVKNWMTQSLLIQNANPDCKTVLKGLGMNPTLEEMLTACQGIGGAQH  372
H1B.FR.83.HXB2          -E-R----------L----G-M-N.....NPPIP--E---R------N-I-R--S-TS----R-------R--------T-----ASQE------ET--V---------I--A--PAA-----M-----V--PG-  358
H1B.US.90.WEAU160       -E-R----------L----A-M-N.....NPPIP--E---R------N-I-R--S--S----R-------R--------T-----ATQE------ET--V---------I--A--PGA---D-M-----V--PG-  358
H1N.CM.95.YBF30         ---R----------LA--VA-M-A.....NPP-P--D--RR--V---NRI-R--S--S--E---------R--------T-----ATQE------ET--V--------QL--A--PGA-----M-----V--PA-  360
H1O.BE.87.ANT70         -E-T----------Q----H----P....NQPIP--D--RK--V---N-M----S--S------------R--------T-----ATQE------ET--V--------QI--S--PGA-----MV----V--PT-  357
H1O.CM.91.MVP5180       -E-T----------Q----I----G....A-SIP--D--RK--V---N-M----S--S----R-------R--------T-----ATQE------ET--V--S-----QI--A--PEA-----MV----V--PT-  357
CPZ.CD.90.ANT           -E-T----------V---MQ-MSTP..QQNGG-P--D---R---M--N-V-RX-S--S--E---------R--------TI----ASQP--A---ET-----------HI--A--TGAS----------V--PA-  377
CPZ.CM.01.SIVcpzCAM13   -E-R----------L----G-A-A.....NPPIP--D--RR-V----NRV----C--S-------------------TM----ASQE--T---DT--V--------QI--A--PGAS----M-----V--PA-  363
CPZ.CM.05.SIVcpzEK505   -E-R----------LA--VA-M-A.....NPP-P--D--RR--V---N-I-R--S-AS---------T-R--------T-----ATQE------ET--V--------NI-RA--PGAS----M-----V--PA-  361
CPZ.CM.05.SIVcpzLB7     -E-R----------V----G-M-S.....NPPIP--E---R---M--N-V-R--S--S------------R--------T-----ATQE------DT--V--------SI-RA--PGAS----M-----V--PA-  358
CPZ.CM.05.SIVcpzMB66    -E-R--------------VG-M-S.....NPPIP--D---R------N-I----C--S---------S-R--------T-----ATQE------DT--V--------SI-RA--PGA---GDEP-F--V--PS-  356
CPZ.CM.05.SIVcpzMT145   -E-R---------SLE--VG-M-A.....NPPIP--D--RR--V---N-I----S--S------------R--------T-----ASQD------DT-----------NI-RA--PGA-----M-----V--PS-  348
CPZ.CM.98.CAM3          -EAR----------LV---A-M-A.....NPP-P--E--RR-VV---N-V-R--C--S------------R--------T-----ATQE------DT--V--------NI-RA--PGA-----M-----V--PA-  358
CPZ.CM.98.CAM5          -EAR----------L---VA-M-A.....NPPIP--E--RR-VV---N-V----C--S------------R--------T-----ATQE------DT--V--------SI-RA--PGA-----M-----V--PA-  357
CPZ.GA.88.GAB1          -E-R----------L----G---A.....NPPIP--DV-RR-V----N-V-R--C--S----R-------R--------T-----ASQE------DT--V--------QI--A--PGA-----M-----V--PS-  366
CPZ.GA.88.GAB2          -E-R----------L---VG-M-S.....NPP-P--E--RR-VV---N-V-R--C--S------------R--------V-----ASQD------ET--V--------QI-RA--PGA-----M-----V--PA-  360
CPZ.TZ.01.TAN1          -E-T-A--------L---VL-M-TPQA..-GG-P--D---R------N-L-R--S--S------------R--------TI----ASQP--T---ET--V--------HI--A--QGA-----------V--PS-  384
CPZ.US.85.CPZUS         -E-R----------L---VG-M-A.....NPPIP--D--RR-VV---N-V----C--G------------R--------T-----AS-E--T---ET--V---------I-RA--PGA-----V-----V--PA-  360
GSN.CM.99.CN166         ---S-A----V---PN-----I--Q....NQP---SD--RK-------R-----------------------E-----F-C-----A-Q-----------V-----E--LI--AMPG.AN----------V--PT-  364
GSN.CM.99.CN71          ---S-A----V---PN-----I--Q....N-P---AD--RK-------R-----------G----------E-----F-C-----A-Q--------L--V-----E--LI--AMPG.AS----------V--PT-  364
H2A.DE.x.BEN            ---R----------VD---Q-MY-P....--P-P-----RR--QI------RK---T-------------QS-------S----------------T-----------L-------------------V--PGQ  360
H2A.GW.x.ALI            -E-R----------VE---Q-MF-P....R-P-P-----RR--QI------R----T-------------QS-------S----------------T--V--------L-------------------V--PGQ  360
H2A.SN.x.ST             -E-R----------VE---Q-MF-P....--P-P--S--RR--QI------R----T-------------QS-------S----------------T--V--------L------I-------------V--PGQ  360
H2B.CI.x.EHO            -E-R----------VE---Q-MY-P....--P-P-----RR--Q-------R----T-------------QS-------S----------------T-----------L-----------------------PGQ  355
H2B.GH.86.D205          ---R----------VE---Q-MY--....--P-P-----RR--Q-------R----T-------------QS-------S----------------T-----------L-----------------------PGQ  355
H2G.CI.x.ABT96          -E-T-----------E---Q--H-Q....--PIP---X-RR--X-------R-X--T----V--------QS-------SX---------------T-----------L-------------------X--PGQ  360
H2U.FR.96.12034         -E-R----------VE---Q-MY-P....--PIP-----RR--Q-------R----T-------------QS-------S----------------T-----------L-----------------------PGQ  360
MAC.US.x.239            -E-S---------SVD---Q-MY-Q....--PIP-----RR--Q-------R----T----V--------QS-------S-------A--------T-----------L------V-------------V--PGQ  359
MAC.US.x.251_1A11       -E-S---------SVD---Q-MY-Q....--PIP-----RR--Q-------R----T----V--------QS-------S-------A--------T-----------L------V-------------V--PGQ  359
MAC.US.x.251_BK28       -E-S---------SVD---Q-MY-Q....--PIP-----RR--Q-------R----T----V--------QS-------S-------A--------T-----------L------V-------------V--PGQ  359
MAC.US.x.EMBL_3         -E-S---------SVD---Q-MY-Q....--PIP-----RR--Q-R-----R----I----V--R-----QS-------S-------A--------T-----------L------V-------------V--PGQ  359
SMM.SL.92.SL92B         -E-S----------PS-----MY--....--P-P--D--RR--Q-------R----T----V--------QS-------S----------------T-----------L-------------------V--PGQ  360
SMM.US.x.H9             -E-X----------XD---Q-MX-Q....--PIP-----XR--Q-------R----T----V------S-QS-------S---------X------T-----------L-----XI-------------VX-PGQ  360
SMM.US.x.PGM53          ---R----------VE---Q-MY-Q....--PIP-----RR--Q-------R----T----V--------QS-------S----------------T-----------L-------------------V--PGQ  360
SMM.US.x.SIVsmH635FC    -E-R----------VE---Q-MY-Q....--PIP-----RR--Q-------R----T----V--------QS-------S----------------T-----------L-----------------------PGQ  360
SMM.US.x.SIVsmH635F_L3  -E-R----------VE---Q-MY-Q....--PIP-----RR--Q-------R----T----V--------QS-------S----------------T-----------L-----------------------PGQ  360
SMM.US.x.SIVsmH635SB10  -E-R----------VE---Q-MY-Q....--PIP-----RR--Q-------R----T----V--------QS-------S----------------T-----------L-----------------------PGQ  360
MNE.US.x.MNE027         -E-S----------VD---Q-MY-Q....--PIP-----RR--Q-------------T-----R--------QS-------S----------------T-----------L-------------------V--PGQ  359
STM.US.x.STM            -E-S---------SPE---Q-MY-Q....--PIP-----RR--Q-------R---------V------T-QS-------S-----A--S------RT-----------L-------------------V--PGQ  361
DEB.CM.99.CM40          -E-S-A-----N--VE---A-M--PAGQG-GPID--Q--RR-V-----R-------T----V-------------------T-----A-Q-------TT-M--------RII-----Q--------H----V--P-Q  386
DEB.CM.99.CM5           -E-S-A-----N--VE---A-M--PAGQGAGPID--Q--RR-V-----R-------T----V-------------------T-----A-Q-------TT-M--------RII-----Q--------H----V--P-Q  382
RCM.GA.x.GAB1           -E-T---------S-A---A-----....N-PI----L-RN---V----W------------------S---------------A--------------------M---------S-----------V--P--  360
RCM.NG.x.NG411          -E-T---------N-A---A-----....N-PIA-----RN--V------------------------------------------A-------------------------------------------A--P--  360
DRL.x.x.FAO             -N-S---------S-E----------....-DSI---A--RQ-VV----R--T---------V-------------------------H--A--------T-----------VI---------S-----L----V--PG-  361
MND_1.GA.x.MNDGB1       -T-S----------VE--LA-MNMQ....---I---T---S-----MNRL--SHC-IS-T-VR-----A---------NVM----ASGE--M--Q-H---E----E--QI-RS--KGA------E----V--P--  348
MND_2.CM.98.CM16        -N-T---------S-E--------E....-D-----G---Q--V-------S-----------------------------------R---Q--T----T-----------SI---------S----PL----V--PKY  361
MND_2.GA.x.M14          -N-T---------S-E--------Q....-EQ----A---Q--V-------S-----------------------------------R---Q--T----T-----------SI---------S-----L----V--PKY  361
MND_2.x.x.5440          -N-T-N-------S-E--------Q....-DQ----G---Q--V-------S-----------------------------------R---Q--T----T-----------AT--------------L----V--PKY  361
GRV.ET.x.GRI_677        ---T---------S--------FN-....NPRID--AQ-RK-V------V-Q----QKV---R-------Q--------------APQD-------T-----------LI--------------I----V--P--  361
TAN.UG.x.TAN1           ---R---------SVA------FN-....NPR-D--R--R--V-------------ISV---R--A------------Q-------PQD------ET-----------L------IH-------------V--PG-  367
VER.DE.x.AGM3           ---R----------V---L--IYT-....NPR-D--A--RR-----------------SV---R-----A-----------I----ASGE--Q---E------------VI------H-------------V--PSY  371
VER.KE.x.9063           ---R----------V---L--IYT-....NPR-D--A--RR-----------------SV---R-----------------I----ASGE--Q---E------------VI------H-------------V--PSY  367
VER.KE.x.AGM155         ---G----------V---L--IYT-....NPR-D--A--RR-----------------SV---R-----------------I----ASGE--Q---E------------VI------H-------------V--PSY  367
VER.KE.x.TYO1           ---R---------SV---L--IYT-....NPR-D--A--RR-----------------SV---R-----------------I----ASGE--Q---E------------VI------H-------------V--PSY  367
TAL.CM.00.266           -E-A----------PA--V------....--P-D--A---R-V-M---R-----------V----R-------------RC-------Q--------T--V-----E--AI--AM-PGA------Q----V--P--  363
TAL.CM.01.8023          -E-A----------PA--V-----Q....--P-D--S---R-V-I---R-----------V----R----E------RC-------Q--------T--V-----E--AI--AM-PQA------Q----V--P--  367
MON.CM.99.L1            G--T-A----VS--P-Q----I---....N-P-Q-SD--RK-V-----R-------------------------------C-------Q--------T--V--------LI--AM.PGAS--D--Q----V--PM-  362
MON.NG.x.NG1            .......................................................................................................-----LI--AM.PGAS-----Q----V--PA-   31
MUS_1.CM.01.1085        -N-T------VS--P------I---....N-PI--AE---K-V-M---R-------------------------------F-C-----S-Q-------ST--V-----E--LI--SM.PGA------Q----V--PT-  355
LST.CD.88.447           -T-NAT-V--V---VE--LA---AD.....TP-D--K---E-V-QAME-V-RIHQ--SVI-----------E-A---F------GGSHE--E--KEKM-V-------RM-I-A--EGA-----MR----V--PA-  345
LST.CD.88.485           -T-NAT-V--V---VE--LA---AD.....TP-D--K---E-V-QAME-V-RIHQ--SVM-----------E-A---F------GGSHE--E--KEKM-V-------RM-I-A--EGAS--D-MK----V--PA-  345
LST.CD.88.524           -T-NAT-V--I---VE--QA---AD.....TPID--K---E-V-QAME-V-RIHQ--SVM-------------A---F------GGSHE--E--KEKM-V--------M-I-A--EGA-----MK----V--PA-  343
LST.KE.x.lho7           -T-NAT-V--V---VE--LA---AD.....TP-D--K---E-V-QAME-V-RIHQ--SVM-----------E-A---F------GGSHE--E--KEKM-V-------RL-I-A--EGAS----MK----V--PA-  348
SUN.GA.98.L14           -T-N-T----V---VE---Q-IYGL..NGATR-A-QD-----V-E-MERV-RLHQTTSVVE-R-------R--T---F------SGSEE--E--KEKM-V--------L-I-A--ET-S----MR----V--PS-  348
SYK.KE.x.KE51           ---SA------------------I--Q....N-PIQ--Q--RQ----------QV---------R-----T---------HC-----A-------L-----------E-RSI--AMVK.-------Q----V--P-Y  380
SYK.KE.x.SYK173         ---SA---------S-A-----I--Q....N-P-Q--E--RR----------QV----S-----------I---------HC-----A--S--G-L--T--------E-RQI--AMVK.-------Q----V--PL-  383
COL.CM.x.CGU1           -Q-TA---T-N--SVA--VA-GE........PI..A---R---VQS-E-VIQIAR-SSV---R--S--D--S------S-----PAAGEI-A--ANN----H------RI----QK.-S--D--A----V--PD-  349
```

```
             Gag p24 Capsid \/ p2            p2 \/ p7 NC                                                                                              p7 \/ p1
SAB.SN.x.SAB1C         KARLMAEAMTAAFQQ..QTV..GNIFVQQGARPRGPLGGRGRPL......N....P.NIKCYNCGKPGHLARFC.KAPRRQ....................GCWKCGSPDHQMKDC..Q........KQVNFLGF   461
H1B.FR.83.HXB2         ---VL----SQVTNS..A-.....-MM-R-NFRNQ...............R....K.IV--F----E--T--N-.R---KK....................------KEG------..T.......ER-A----K   436
H1B.US.90.WEAU160      ---VL----SQVTNT..A-.....MMM-R-NFRSP...............R....K.T---F----E--I--N-.R---KK....................------QEG------..T.......ER-A----K   436
H1N.CM.95.YBF30        ---VL----SQVQ-P..T-S....V-A-R-NFKGI...............R....K.P---F----E-----N-.-----G....................------QEG------..KNE.....GR-A----K   441
H1O.BE.87.ANT70        ---VL----AT-Q-D..LKGGYTAV-M-R-QN-IR...............K....G.T---F----E--I--N-.R---KK....................------QEG------..RN......G--A----K   441
H1O.CM.91.MVP5180      --KIL----AS-Q-D..LKGGYTAV-M-R-QN-NR...............K....G.P---F----E--I-KN-.R---KR....................------QEG------..KN......GR-A----K   441
CPZ.CD.90.ANT          ---VL----AS-NNAQGT.....AV-L-R-NGN--........G......K....R.PL--F----E--T--N-.-----K....................---R--QEG--L-N-..PAT....NTGK-----K   461
CPZ.CM.01.SIVcpzCAM13  ---VL-----MVQS-A.R-...DI.-F-K-PGATP...............R....R.K--------E-----N-.-----K....................------QEG------..S.......GR------K   443
CPZ.CM.05.SIVcpzEK505  ---VL-----Q-QTA.......TSV-M-R-NFKGI...............R....K.T---F----E-----N-.----KK....................------QEG------..RSG.....ER-A----K   441
CPZ.CM.05.SIVcpzLB7    ---VL----SQ-NAA.......TV-MM-R-NFKGP...............K....R.V---F----E--I--N-.-----K....................------QEG-R-Q--..NM......G-.-----K   436
CPZ.CM.05.SIVcpzMB66   ---VL----SQ-QHS..........NDAKRQFKGP...............K....R.IV--F----E--I--N-.-----K....................------QEG---RN-..TNE......R-A----K   432
CPZ.CM.05.SIVcpzMT145  ---VL----SQVNKT........E--M-R-GQN-P..............P-....K.K---F----E-----N-.R----K....................------QEG------..NQNN....GR.I----K   428
CPZ.CM.98.CAM3         ---VL----SQLQNP..T.....GV-L-R-NNGKP...............T....R.K---F----E-----N-.R----K....................------QEG----N-..PAG.....ER-A----K   438
CPZ.CM.98.CAM5         ---VL----SQVQNS..S.....-V-M-R-NNG-V...............-....K.K--------E--V--N-.R----K....................------QEG-N--E-..TTG.....GR-A----K   437
CPZ.GA.88.GAB1         ---VL----SMVQN-GRA.....DV-F-K-QGAGP...............K....R.K---F----E-----N-.-----K....................---R--QEG------..T.......GR------K   446
CPZ.GA.88.GAB2         ---VL----SQIQ-T.......SS--M-RAGGRTP...............P....R.RV--F----E--I--N-.-----K....................------KEG------..-N.....EGR------G   440
CPZ.TZ.01.TAN1         --KIL----AS-TAGG.V......NML-G-K--PL...............K..K.G.QLQ-F----V--T--N-.R---KK....................---R--QEG------..TTRN...NSTG-----K   467
CPZ.US.85.CPZUS        ---VL----CQMKNP..S.....SV-L-K-NAGKP...............G....R.K---F----E-----N-.R----K....................------QEG------..TAG.....NR-A----K   440
GSN.CM.99.CN166        -S--L----AT-LKG..TSS...YNM--K..............GPPG...R..QGK.TP------QF-----D-.PK-KER....................K-F---RAG-FS-P-R.T....GTPI-ELY--DL   448
GSN.CM.99.CN71         -S--L----AT-LKG...SS...YNM--R..............GPPG...R..QGK.TP-------F--I--D-.PK-KER....................K-F---KAG-LARQ-K.T....GTPA-EL---DL   447
H2A.DE.x.BEN           --------LKE-MGP..SPI..PFAAA--.....................R....K.A-RYW----E--S--Q-.R-----....................------K-G-I-AN-..P.......ER-AG---L   435
H2A.GW.x.ALI           --------LKE-MTP..API..PFAAA--.....................R....R.T---W----E--S--Q-.R---K-....................------K-G-L-AN-..P.......ER-AG---L   435
H2A.SN.x.ST            --------LKE-MAP..API..PFAAA--.....................R....R.T---W----E--S--Q-.R-----....................------KAG-I-AK-..P.......ER-AG---L   435
H2B.CI.x.EHO           --------LKE-LTP..S-N..PFAAA-PR-G..................K....R.TVT-W----A--T--Q-.------....................------QQG-I-SK-..P.......ER-AG----   433
H2B.GH.86.D205         --------LKE-LTP..API..PFAA---K-G...........K......R....G.TVT-W----Q--T--Q-.R-----....................------KTG-I-SK-..P.......ER-AG---L   434
H2G.CI.x.ABT96         --------LKE-LNP..TAL..PFAAA--KTG...........G......K....RST---W----E--TV-Q-.R-----....................------K-G-I-AK-..P.......ER-AG----   440
H2U.FR.96.12034        --------LKE-L-P..MPI..PFAA---RGG-.................R....R.PVR-W----E--T--Q-.------....................-----EK-G-N-AS-..P.......ETK-G---L   439
MAC.US.x.239           --------LKE-LAP..VPI..PFAAA--RGP..................R....K.P---W----E--S--Q-.R-----....................------KM--V-AK-..P.......DR-AG---L   437
MAC.US.x.251_1A11      --------LKE-LAP..VPI..PFAAA--RGP..................R....K.P---W----E--S--Q-.R-----....................------KM--V-AK-..P.......DR-AG---L   437
MAC.US.x.251_BK28      --------LKE-LAP..VPI..PFAAA-KRGP..................R....K.P---W----E--S--Q-.R-----....................------KM--V-AK-..P.......DR-AG---L   437
MAC.US.x.EMBL_3        --------LKE-LRP..VPT..PFAAA--RGP..................R....K.P---W----E--S--Q-.R-----....................R-----KM--V-AK-..P.......DR-AG---L   437
SMM.SL.92.SL92B        ---------KD-LTGSLVAAQFRGAAKG--....................-..K.-.I-R-F----T--S--Q-.R----K....................------EEGRIQAN-..P.......NQKAG---L   441
SMM.US.x.H9            -X------LKD-LT-..GPL..PFXA---KGQ..................X....K.I-X-X--X-E--S--QF.R----X....................------KAG-V-AK-..P.......ER-AG---L   438
SMM.US.x.PGM53         --------LKE-LTP..GQL..PFAA---KGP..................R....K.T---W----E--S-KQ-.R-----....................------KTG-V-AK-..P.......ER-AG---L   438
SMM.US.x.SIVsmH635FC   --------LKE-LRP..DQL..PFAA---KGQ..................R....R.T---W----E--S--Q-.R-----....................---G--KTG-V-AK-..P.......ER-AG----   438
SMM.US.x.SIVsmH635F_L3 --------LKE-LRP..DQL..PFAA---KGQ..................R....R.T---W----E--S--Q-.R-----....................---G--KTG-V-AK-..P.......ER-AG----   438
SMM.US.x.SIVsmH635SB10 --------LKE-LRP..DQL..PFAA---KGQ..................R....R.T---W----E--S--Q-.R-----....................---G--KTG-V-AK-..P.......ER-AG----   438
MNE.US.x.MNE027        --K-----LKE-LAP..GPL..PFAAA--KGP..................R....K.P---W----E--S--Q-.R-----....................------QMG-V-AK-..P.......DR-AG----   437
STM.US.x.STM           --------LKEV--P..DPL..PFAAA--QG...................R....R.TV--W----A--T-KQ-.-G----....................------QQG---AK-..P.......ER--G----   438
DEB.CM.99.CM40         ---------AS-LKE.AGSLG...M---RR.............GPPG.......SRRQ-R-F---QI---QKD-.-R-.-K...................TK-F---QEG-IA-N-.G-.......TPRA----N   468
DEB.CM.99.CM5          ---------AS-LKE.AGSLG...M---RGGQ...........GGGP........RRQLR-F---QI--VQ-D-.-K-.-K...................VK-F---KEG-IA-N-.G-.......VPRA----N   465
RCM.GA.x.GAB1          ---VL----QMMQSN.IMAQQSA-..RGPPR...................RSGGN-.-LR---------IS-Y-.-----K....................----------LL---..T........--I----R   441
RCM.NG.x.NG411         ---VL----QMVQSNIMAQQGGP.RKGPPKLG...........GG.....P....R.FL-------T--T--Y-.-TS-KK....................---R--EEG-L----..P........-E-----N   442
DRL.x.x.FAO            ---V-----REQQAA.IMMQQN..PPRGPPRG--.........GPPP........K.-PR-P---QF--TL-Q-.TK--KK....................--FR--AL--MLRN-..PK........------N   444
MND_1.GA.x.MNDGB1      -------M-RTVVG-.S-NFVQQRGPQRGPV-..........QPTG....R....K..PI-F--N-E--V---F.-----K....................---N--AM---KAQ-..PK...PAQQQR-----Y   436
MND_2.CM.98.CM16       --QM-----KE-QSA.VMMQNS-GPPRGPPR............QPP....R......--R-P----F--GL-D-.IS--KK....................--F---DLG-I-RN-..PK........M-----N   442
MND_2.GA.x.M14         --QM-----QKMQSE.VMMQNS-GPPRGPPR............QPP....R......-PR-P---RY--VL-D-.RL--KK....................--F---DTG-M-RN-..PK........M-----N   442
MND_2.x.x.5440         --QM-----QQ-QAA.VMMQNS-GPPRGPPR............QPP....R......-PR-P----F--VL-D-.R---KR....................--F---D-G-L-RN-..PK........M-----N   442
GRV.ET.x.GRI_677       --K--V-M-SNGQNM.V-V...-PQKKGP.....................R....G.PL--F----F--MQ-E-.----QI....................K-F---KIG-MA---..K.......NG-A----Y   436
TAN.UG.x.TAN1          --K--V---QQMQGV.NMVQ..-APRGGR-RG..................R....G.PPR-FK--QI--IQKD-.PRAGPN....................K-L---K-G-LA---..R.......SG-A----R   446
VER.DE.x.AGM3          --KV---M-QNMQS-.........NMM---GQRGR-..............R....-.PV-------F--MQ-Q-.PE--KM....................R-L---K-G-LA---..R........G------Y   447
VER.KE.x.9063          ---V---M-QNMQT-.........NM-----AGGVRRQ............R....-.PL-------F--MQ-Q-.TE--KI....................K-L---KLG-LA---..R........G------Y   445
VER.KE.x.AGM155        --KV---M-QNLQS-.........NM----GGRGR-..............R....-.PP-------F--MQ-Q-.PE--KI....................K-L---K-G-LA---..R........G------Y   443
VER.KE.x.TYO1          --KV---M-QTMQN-.........NM----GPK-Q...............R....-.PLR------F--MQ-Q-.PE--KT....................K-L---KLG-LA---..R........G------Y   442
TAL.CM.00.266          -S-------AE-LK-..........NT--VI.AMVQQ-QG-PRGRGG..PRRNPPG.Q-R------F--I-KN-.P--P--...............KVPPGT-Y---K-G-IA-Q-..SS.......A-A----K   459
TAL.CM.01.8023         -S-------AE-LK-..........NT--VI.AMV.Q-QG-PKGRGG..PRRTPPG.Q-R------F--I-KN-.P--QKP...............RGPPGS-F---QMG-RAAQ-..RQ.......P-A----K   462
MON.CM.99.L1           -S-IL----AG-IANMPMNMVQARGPP-RKGQ..........................P--F----F--M-KN-.---Q-R....................K-YN--Q-G-LA---..PQ..PPKQN-G-----N   446
MON.NG.x.NG1           -S--L----AT-I.NSNMPM....NM--GRGG.................XQPRRQGXQ-R------F--V-KN-.T---KT....................--FR--KEG-XS-N-.PNGGQN.QPRNR-----N   121
MUS_1.CM.01.1085       -S------LAS--.KQAG-L....NM--GAKGA--RGN-SQG.NRGN...........PR-----QY--V--D-.PK-.-N...................KK-F---REG-LARQ-..R....SDNT-A-----D   446
LST.CD.88.447          -G-IL----AT-M-S..-MR...QNM--.VTP--NAQ-RFV-TGGGG..PR....K.PLT-F-------T--M-.RQ--QE....................---N---KE-RFAQ-..PK.....PKGK-----Y   439
LST.CD.88.485          -GKIL----AT-M-S..-MR...QNM--.VTP--NEQ-RFV-TGGGG..PR....K.PLT-F-------T--M-.RQ--QE....................---N---KE-RFAQ-..PK.....PKGK-----Y   439
LST.CD.88.524          -RKIL----AT-M-G..-MK...HLM--.QTP--NAQ-RFV-TGGGG..PR....R.PLT-F-------S-KM-.RQ--QE....................---N--AK--R-AQ-..PK.....PKGK-----Y   437
LST.KE.x.lho7          -GKIL----AS-I--..-VR...QNM--.VTPL-NAQ-RFV-TGGGG..PR....K.PLT-F-------T--M-.RQ--QE....................---N---KE-RFAQ-..PK.....PKGK-----Y   442
SUN.GA.98.L14          -GKIL----AS--R-..-VG..RQAM---NLP--NSQ-RFV-IGGGG..PR....K.PMT-F---------NQ-.RE-KKG.................PPGS-----KMG-KQAQ-..PQ.......G------Y   446
SYK.KE.x.KE51          --K-L----VMTQHS..L......GMI-GPRQG..............SNPR....RGPTR-F---QL---QKD-.PR-KKL....................K-FN--GTG-IARQ-RQPRKGQGNPPP-A----K   468
SYK.KE.x.SYK173        --K------VM-Q-S..V......NM--GPSKG.................R....S.M------QI--MQKD-.-K-LKA....................K-FN--KTG-LARA-RQPKRNQGPPVA-A----K   467
COL.CM.x.CGU1          --KVL----QQFQ-E.......RTNMIEVKTA...........................--F--QGI-----M-P-R-IGGAGRGRGRGRGGFRGAPRRPVR-FT-NQEG-MQR--..P.......N--A----.   440
```

```
                                                            PTAP motif                                                                               Vpr binding site
                                    p1 \/ p6                  |- -|                                                                                         |-  -|
SAB.SN.x.SAB1C         GPW......GRGK.PRNFPLT...........SIR....PTAPP................MERDYSRPEENWYADRPPTRGPGPDDPATALLKQYAVQGKRQKQQWQNHSPQQSPYEEAYSSLRSLFGED...Q    554
H1B.FR.83.HXB2         I.-PS....YK-R.-G--LQ....SRP..E.........-----..........................EESFRSGVET................TTPP.......QKQEPIDKELYPLT-------N-PSS-    500
H1B.US.90.WEAU160      I.-SS....QK-R.-G---Q....SRL..E.........-----..........................EESFRFREET................TTPS.......QKQEPIDKELYPLT--K----N-PSS-    500
H1N.CM.95.YBF30        S.-SP....FK-R.-G---Q..TTTRK..E.........-----..........................LESYGFQEEKSTQG..................-EMQE-QERTENSLYPPLT-------N-PSS-    512
H1O.BE.87.ANT70        Y.-PP....-GTR.-G-YVQ....RPA..H.........-S---...............................MEEEVKGQ......................ENQEQKGGPNELYPFA--K----T-...-    498
H1O.CM.91.MVP5180      Y.-PP....-GTR.-G-YVQ....KQV..S.........-S---...............................MEEAVKEQ......................ENQSQKGDQEELYPFA--K----T-...-    498
CPZ.CD.90.ANT          P.TPTWW...GCR.-G--VQ....KEEVVE.........-----..........................IE..IYQEE........................HKRTQKGLKGEEELPPSY--K----K-...-    522
CPZ.CM.01.SIVcpzCAM13  -.-PSR....S-R.-G--VQ....NRT..E.........-----..........................AESYGYQ........................E-EKN-EEQEEGNRLYPPLT--K----S-PSS-    506
CPZ.CM.05.SIVcpzEK505  V.-PLS....K-R.-G---Q..TTTRK..E.........-----..........................IESYGYQEEKTTQGT...................EREEKEKTE-SLYPPLT--K----S-PSL-    511
CPZ.CM.05.SIVcpzLB7    L.-PSS....Q-R.-G---Q....SRP..E.........-----..........................IESFGFQ-KTT....................---EDREEDKTENKLYPPLT--K----N-PLS-    503
CPZ.CM.05.SIVcpzMB66   L.-PSS....K-R.-G--LQ....KRP..E.........-----..........................IESFGFREEEPA....................QKEEVKSEETENKLYPQLT--K----N-PLS-    499
CPZ.CM.05.SIVcpzMT145  N.-PPW....G-R.-G-YVQ....GRK..E.........-----..........................MESFGFQEE........................KKVKEEVREE-KPPSPLT--K----N-PSS-    491
CPZ.CM.98.CAM3         -.-SPW.SG-SKR.-G--LQ....GRK..E.........-----..........................LESYGYQEAT-.....................-R-EEEREKEKE-LYPPLT--K----S-PSSE    507
CPZ.CM.98.CAM5         D.-SPW.SGRSKR.-G---Q....GGK..E.........-----..........................MESYGFQE-TM.........................EKKKEEKE-LYPQLT--K----S-PSLE    502
CPZ.GA.88.GAB1         -.-PSR.SGR....-G--VQ....NRT..E.........-----..........................IESYGYQEEE.........................KSQEKKEGESSLYPPT--K----S-PSS-    508
CPZ.GA.88.GAB2         -.-PSR.SGR....-G---Q....RKV..E.........-----..........................PETFGYAEQTQT-....................KSQKGEESTTEKESPL-P--K----S-PSS-    507
CPZ.TZ.01.TAN1         R..TPLW...GCR.-G--VQNTPEKGKAQEQETAQTPVV-----....................................................................LEMTMKGGF--K-I--S-...-    524
CPZ.US.85.CPZUS        H.-SPSWSG-SKR.-G--L.....ENR.KE.........-----.............IEDFGYQEETVTQEKQGX.................................GKG...AL.XQLT--K----S-PSSE    507
GSN.CM.99.CN166        -RGR......-RA.-----V-FLL...............----A.............................................EEEEPQQLNLSTPVGEP-SKGQKEESRNTL-P--S------...-    512
GSN.CM.99.CN71         -RG.......-RA.-----V-SLL...............-S--A.............................................EEVEPQQLNLSPPVGEP--KEQKAEKRNTL-P--S------...-    510
H2A.DE.x.BEN           --R.......GK-.-----V-...QAP....QGLI....-----..........................ADP-AELLE-YMQQGRKQREQRERPYKEVTEDLLHLEQRETPHREET-DLLH-N----K-...-    521
H2A.GW.x.ALI           ---.......GK-.-----V-...RVP....QGLT....-----..........................AEP-ADLLEQYMQQGRKQREQRERPYKEVTEDLLHLEQGETPHKEVT-DLLH-N----K-...-    521
H2A.SN.x.ST            ---.......GK-.-----VA...QIP....QGLT....-----..........................IDPVEDLLEKYMQQGKRQREQRERPYKEVTEDFL-LEKQETPCRETT-DLLH-N----K-...-    521
H2B.CI.x.EHO           ---.......GK-.-----VQ...APQ....GIV.....-S---........MNPAFGMTPQGAIPSAPPADP-EEMLKNYMQLGKKQKE.................-RERPYKEVT-DLLH-N------...-    519
H2B.GH.86.D205         ---.......GK-.-----M-...QVP....QGVT....-S---........MNPAEGMTPRGATPSAPPADP-VEMLKSYMQMGRQQRE.................SRERPYKEVT-DLLH-N------...-    521
H2G.CI.x.ABT96         ---.......GK-.-----M-...QVP....QGLT....-S---..........................MDP-VDLLKNYMXLGRKQKEQ.................RNKPYKEVTXXLLH-S----D-...-    509
H2U.FR.96.12034        ---.......EK-.-----MA...QVP....QGLT....-----............AEPAVDLLTPTAPPADP-VDLLKSYMQQGKKQKE.................-RERPYKEVT-DLLH-N----N-...-    522
MAC.US.x.239           ---.......GK-.-----MA...QVH....QGLM....-----..........................EDP-VDLLKNYMQLGKQQREKQRE.............SREKPYKEVT-DLLH-N----G-...-    510
MAC.US.x.251_1A11      ---.......GK-.-----MA...QVH....QGLT....-----..........................EDP-VDLLKNYMQLGKQQREKQRE.............SREKPYKEMT-DLLH-N----G-...-    510
MAC.US.x.251_BK28      ---.......GK-.-----MA...QVH....QGLT....-----..........................EDP-VDLLKNYMQLGKQQRE.................SREKPYKEVT-DLLH-N----G-...-    506
MAC.US.x.EMBL_3        ---.......GK-.-----MA...QVH....QGLT....-----..........................EDP-VDLLKNYMQLGKQQRE.................SREKPYKEVT-DLLH-N----G-...-    506
SMM.SL.92.SL92B        ---.......GK-.-----MQ.........TT-LT....-S---...........................DP-A-IVKEYLEKAQREKT.................RR-RPYKEVT-DLLH-N------...-    507
SMM.US.x.H9            ---.......GK-.-----MA...QMP....QGLT....---X-..........................EDP-VDLLKNYMKVGRRQRE.................-RERPYKEVT-DLLH-N------...-    507
SMM.US.x.PGM53         ---.......GK-.-----MA...QMP....QGLT....-----..........................EDP-VDLLRNYMKMGRKQRE.................-RERPYKEVT-DLLH-N----G-...-    507
SMM.US.x.SIVsmH635FC   ---.......GK-.-----MA...QMP....QGLT....-----..........................EDP-VDLLKNYMKMGRKQRE.................-RERPYKEVT-DLLH-N------...-    507
SMM.US.x.SIVsmH635F_L3 ---.......GK-.-----MA...QMP....QGLT....-----..........................EDP-VDLLKNYMKMGRKQRE.................-RERPYKEVT-DLLH-N------...-    507
SMM.US.x.SIVsmH635SB10 ---.......GK-.-----MA...QMP....QGLT....-----..........................EDP-VDLLKNYMKMGRKQRE.................-RERPYKEVT-DLLH-N------...-    507
MNE.US.x.MNE027        ---.......GK-.-----MA...QMH....QGLT....-----..........................EDP-VDLLKNYMQLGKQQRE.................SKRKPYKEVA-DLLH-N------...-    506
STM.US.x.STM           ---.......GK-.-H---MA...QIP....QGLT....-----..........................EDP-ADLLRSYMQLGKKQRE.................SRKTPYKEVT-DLLH-N------...-    507
DEB.CM.99.CM40         T--AS......RR.----LEG...IGK..............YI-........................................EST-DQVYSTTR-PVQG-ENNP-REETTRRGKYPDLMG-K------...-    535
DEB.CM.99.CM5          A--AS......RR.----LEG...IGK..............YI-........................................EST-DQVYSTTR-PVQGERNAP-EDTTTRRGKYPDLMG-K------...-    532
RCM.GA.x.GAB1          L--G......Q--.-------...........-LT....-S--G......................MESNYDP-EEMLKNYLRRAGEQKRQQR-............EESKKREGA-Q--LT--N----S-QL.-    518
RCM.NG.x.NG411         A--K......Q--.---L---...........-L-....----G..................-EEGWNPQYDP-EEMLRKYLALGRQHKQEQRK............E-KEKVGRA--D-L---N----S-...L    521
DRL.x.x.FAO            T--......-S--.-----QA...........-LT....-S---....................LPGYLQEDP-E-MLQKYMEQGAQQKRQQR-Q..............QQKRG------N--S----S-QL.-    521
MND_1.GA.x.MNDGB1      ---......-PS-.-G-Y-AQ...........EVT....-----.........................LEEKPLQKTLSTYQKLGRGL..................RQKMKEEKR--DFH--ST--Q--Q...    502
MND_2.CM.98.CM16       T--......-S--.-----AM...........PLT....-----......................M-GMEDP-E-MLLDYMKKGQQQK-..............ESKQEKKERG---A--N--S----T-QL.-    517
MND_2.GA.x.M14         A--......-S--.-----AM...........PLT....-S---......................M-GMEDP-EKMLLDYMRKGQQQR-.............AKETKQEKDKG---A--N--S----T-QL.-    518
MND_2.x.x.5440         T--......-S--.-----AM...........PLT....-S---......................M-GLEDP-EKMLLDYMKKGQQQR-.............AAGAQKEEKKG---A--N--S----T-QL.-    518
GRV.ET.x.GRI_677       -H-......-GA-.----VQY..........RGDTVGLE-----......................METAYDP-KKLLQQYAEKGQRLREE...............REQTRK-KEK-VEDV--S----G-...-    513
TAN.UG.x.TAN1          M-TW......GT-.----LEQ...GGA...........V-----...........MPAHGFPTGSPAAGAYDP-RKLLEQYAKKG-QLR......................K-KEK-LEDY--S----G-...-    523
VER.DE.x.AGM3          -R-......MGA-.-----AA...........TLG..VE-----.........................PPSPY-PAKKLLQQYA-KGKQ-RE-RKKPPA............VN-DWTEGY--N------...-    521
VER.KE.x.9063          -R-......TGT-.----LAA...........THG..VE-S---.........................PPTPY-PAKKLLQQYAEKGKQ-RE-GKRTPP............TN-DWAEGY--N------...-    519
VER.KE.x.AGM155        -R-......MGT-.-----AA...........TLG..AE-S---......................P-NNSTPY-PAKKLLQQYAEKGKQMRN-NRNPPA............NN-DWNEGY--N------...-    520
VER.KE.x.TYO1          -R-......MGA-.-----AA...........TLG..AE-S---......................P-SGTTPY-PAKKLLQQYAEKGKQ-RE-KRNPPA............MN-DWTEGY--N------...-    519
TAL.CM.00.266          SFLGKIP..PVRG.-----VVQ.................-S---...........IEPLTASWQEGTTTWTPP-K................................KPQETTTTR-SL-P--A----D-...-    527
TAL.CM.01.8023         AFLGKIP..PVKG.-----VVQ.................-----...........IEPLTASWQEGTTTWTPPTK...............................DAAA-K-QTR-SL-P--A----D-...-    531
MON.CM.99.L1           PFG......PKKG.-----..............LT.SVQ-S--T........................LPAGT-EPVNLNQEIGS.....................PKTGTPKETRRDL-P--A------...-    510
MON.NG.x.NG1           PFGP.....-KRG.-----........VTSLQ.......----AEPTXGXYKVNFSPPVGPPETPAPTAP..........................................KETRKDL-P--A----D-QSE-    192
MUS_1.CM.01.1085       PFG.....LQKRR.----..............PQT.SLL-----...............................................EQEGELVNLSPAV...GPGEGKQDRKSL-P--S---QD-QSD-    509
LST.CD.88.447          ---.......-SGP-G----M...GGN........AGVV-S---......................MERSPTQ-E-ALETYRNLGQQLR..................RQQQIPQKCVDEPCLSFFF-PD-...-    512
LST.CD.88.485          ---.......-NGP-G----M...GGN........AGVV-S---......................MERNSTK-E-ALETYRNLGQQLR..................KQQQVPQKCVDEPCLSFFF-PD-...-    512
LST.CD.88.524          ---.......-NGPSG---IM...GGA........AGVI-S---.....................MEKVPPTR-E-AIETYRNLGQQLR..................KQQQVPQKCVDEPCLNFFF-PD-...-    511
LST.KE.x.lho7          ---.......SS-P-G-Y--L...GGA........AGRI-S---......................MESAPTK-E-ALETYRTLGQQLK.................RQQQQVPQKCVDEPCLNS.--PD-...-    515
SUN.GA.98.L14          ---......N--P.-G---......VM............-S---................L-DLTLGNRMTPPQSKAERALETYRLLGQG-RA-Q...............KRKERG-CQEPC-NM--P--...P    521
SYK.KE.x.KE51          -WG......N-KP.-A---........VMRET.......-----...............AEDWPWQQQTWGN--SPQQKHIAMATPLQPQSSPKTVKDL.LVPGKPSEKKEEMKEKGPL-P--Q----D-...-    561
SYK.KE.x.SYK173        -WGV.....S-RP.-A---........VRSE........-S---..............LEDIE-GPWLTWSAQMSQQAQAKAQNSPSKKPPT........NREVLSPKE-SGKEETKSL-P--S------...-    554
COL.CM.x.CGU1          ........................ATRGVELQT......AIF--........................KMSKD...................................LP-RREKG-SL-P--K----D-...-    482
```

```
                       / Pol reading frame (-1 from Gag) p15                                                               Gag-Pol TF \/ Pol Protease
SAB.SN.x.SAB1C         FRVWPLG.....QRET.QEFPSD....LHQ...TNSSPNG.TGLQQ....AGGKLVCRQTSDQRTRARRSSNSPVKAVCCSGETAETAVA.......KPLATTEPLRGGLQLPQVSLWRRPMKTVYIEGQ  105
H1B.FR.83.HXB2         --ED.-AF....LQGK.AREF-S....EQT...RAN.......SPT....RRELQ-WGRDN.........................NSPS......EAGADRQGTV..SFNF---T--Q--LV-IK-G--   73
H1B.US.90.WEAU160      --ED.-VF....PKGK.AREF-S....EQT...RTN.......SPT....RRELQ-QGRDN.........................NSLS......EAGANRQGAV..SFNF--IT--Q--LV-IK----   73
H1N.CM.95.YBF30        --EE.-VS....LQRE.TRKLPP....DNN...KERAH.....SPA....TRELW-SGG.....................EEH-G-GDAG......EPGEDRELSVP.TFNF--IT--Q--VI--K-GKE   80
H1O.BE.87.ANT70        --QI.-AS....GGHE.ARQLCA....ETS...-PI.......SPT...................................DGGGSEGTG......ESGTERGPERALSVC---IP--D--IV-ARVG-H   69
H1O.CM.91.MVP5180      --EV.-AS....GGHE.ARQLCA....ETS...VPI.......SPT.................................NG-GSEG-RES......ESEGGSGRAVP..IC---IP--D--IV-AKVG-H   69
CPZ.CD.90.ANT          --ET.DPHVV.GVQ...TRELCA....EGGSS.GAN.......SST....HRDL..SGGAQED.................................SEGSQGGGGTT.S-VF-EIP--Q---ME-L-Q--   73
CPZ.CM.01.SIVcpzCAM13  --EG.-AF....PQRE.ARQLCT....EQD...RAN.......SPT....SRELW-P-E...........................GEEP......RGTGRGEQTIP.SSNF--IT--Q--LIP-R----   72
CPZ.CM.05.SIVcpzEK505  --EG.-AS....LQRE.ARKFPP....DNNKE.RAN.......SPS....NRELW-S......................GGEDHTGDREG......R-GEDRELSVP.T-NF--IT--Q--IL--K-G-E   79
CPZ.CM.05.SIVcpzLB7    --ET.-AF....LPGK.ARKL-S....EQT...RAH.......SPA....NRELW-SAK.......................NHP--GGQ......GRGQDREQAVP.S-NF--I---Q--IV-IKVG-Y   76
CPZ.CM.05.SIVcpzMB66   --ET.-AF....-QGK.AREFPS....EET...RTN.......SST....NRELR-QGGGTCP.......................EGGS......EERGDREQAVS.SANF--I---Q--VV--R----   76
CPZ.CM.05.SIVcpzMT145  --KK.-AS....LGGE.ARELCA....RQE...GTN.......SST....Y-ILRIS..........................GG-ESQG......GGEGGEQAAV..SFNF--IT--Q--LIP-K----   72
CPZ.CM.98.CAM3         --ER.-VSVE.RGIKE.TRELPA....GQE...GTH.......SSA....PRELW-SGSNR........................KETGG......GEGEGKGT-VS.SVN---IT--Q--LLG-K----   78
CPZ.CM.98.CAM5         --EG.-VSVE.REIKK.TREL-A....GRE...GTH.......SST....H-ELW-SGNNN............................G......E-EGGEGT-VS.SIN---IT--Q--LLG-KV---   74
CPZ.GA.88.GAB1         --ER.-AFPQ.RE....ARQLCA....EQN...RTN.......GPT....DRELW-PGGR...........................EE.......PGEERGR-QSI.STN---IT--Q--LIP-KV---   71
CPZ.GA.88.GAB2         --GG.-AF....PERE.AREL-P....KEG...GTN.......SST....SRDLR---ADT........................DRKES......EGGGEYNRKGEPS-SF--IT--Q--LLP-RVA--   76
CPZ.TZ.01.TAN1         --ET..HPLV.GVQ...TRELCA....E-PR..ERE......................................GSGAGD-TD-SGANCP...........--GDDDERRV----N--Q---M--KVQ--   69
CPZ.US.85.CPZUS        --ET.-VPIVERGIKE.TRELP.....GK-..EGAH..SSTNRGFRVSRRDS...............................DTGETGK............GKGAL.SANF--IT--Q--LLRIKVA--   75
GOR.CM.06.CP684        ..................................................................................................................................    0
GSN.CM.99.CN166        --P--GA.....PPSPQELSSNFPAANSPG...G..................-GRTPAAQ-L...........DTSRG.........TP........EQGTEGGEQEHS-S--VI---G----E-S-G--   76
GSN.CM.99.CN71         --P--GA......TSPQELSSNFPAAISPG...G..................-GGTPAAQPV...........SAGGG.........-P........EQG-EGRKEEHS-S--VI---G---RE-SMG--   75
H2A.DE.x.BEN           ---G-T-.....KEAS.-LPRDP....SPS...GADTN.....STS.......GRSSSG-VGEIYA--EKAEGAEGETIQR-DGGLA-PR.......AERD-SQRGDR--AA--F---K--VV-A---D-   98
H2A.GW.x.ALI           --A--M-.....KEAS.-LPRNP....SSA...GINTN.....STP.......SRASSGPAGAVYA-GEKAKRAEREAIQR-DGGL--PR.......AGRD--QRGDR--AA--F---K--VV-A-----   98
H2A.SN.x.ST            ---G-M-.....KEAP.-FPCGP....NPA...GADTN.....STP.......DRPS-GPTREVHA--EKAERAEREAIQRSDRGLP-AR.......ETRD-MQRDDR--AA--F---K--VV-AHV---   98
H2B.CI.x.EHO           ---R---.....KEAS.-FPRPG....TPG...DSAI......CAP......DEPSI-HDTSGCDSICTPCR-SRGDAKELHA-R-E-EG.......EQRE-LQGGDR-FAA--F------VVKAT----   98
H2B.GH.86.D205         ---RT--.....KEAS.-LPHDP....SAS...GSDTI.....CTP......DEPSRGHDTSGGDTICAPCR-SSGDAEKLHADG--TER.......E-RE-LQGGDR-FAA--F------VVKAC----   99
H2G.CI.x.ABT96         ----T--.....KEAS.-LPHDP....SPS...GSD-I.....STT.......DGPSSRPTEELHAXGEEAKGA........................EKQ-LQGGDRXXAA--L------VV-A---S-   81
H2U.FR.96.12034        --A----.....KKAS.-LPHGP....SPA...GTDPN.....SPS.......SRTSSGSA-PDSPSSGP-SG...SAKELHA-GQ.ETK......GEQRE-IQGGD--FAA--F---K--IV-A-----   95
MAC.US.x.239           --P-SM-.....KEAP.-FPHGS....SAS...GADAN.....CSP........................RGPSCGSAKELHAVGQA-ER...KAER-QREALQGGDR-FAA--F------VV-AH----   85
MAC.US.x.251_1A11      --P-SM-.....KEAP.-FPHGS....SAS...GADAN.....CSP........................RGPSCGSAKELHAVGQA-ER...KTER-QREALQGDDR-FAA--F------VV-AH----   85
MAC.US.x.251_BK28      --P-SM-.....KEAP.-FPHGS....SAS...GADAN.....CSP........................RGPSCGSAKELHAVGQA-ER.......-QREALQGGDR-FAA--F------VV-AH----   81
MAC.US.x.EMBL_3        --P-SM-.....KEAP.-FPHGS....SAS...GADAN.....CSP........................RGPSCGSAKELHAVGQA-ER.......-QREALQGGDR-FAA--F------VV-AH----   81
SMM.SL.92.SL92B        --P--M-.....KEAS.-LPHA-....NII...DTV-TPRS..SSP...............................DSQGVSRES-EGKD......-EEQALQGGD--FAA--F--R---VVRAL----   78
SMM.US.x.H9            --A--M-.....KEAP.-FPHGP....DAS...GADTN.....CSX........................RG-SCGSTEELH-GGQK-EG.......EQRE-LQGGN--FAA--F------VV-A---E-   81
SMM.US.x.PGM53         --A--M-.....KEAP.-FPHGP....DAS...GADAN.....CSPR........................G-SCGSTEELH-DGQK-EG.......EQRE-LQGGD--FAA--F------VVAA---E-   81
SMM.US.x.SIVsmH635FC   -----M-.....KEAP.-FPHGP....DAS...GADTN.....CSP........................RG-SCGSTEELH-DGQK-EG.......EQRE-LQGGDR-FAA--F------VV-A---E-   81
SMM.US.x.SIVsmH635F_L3 -----M-.....KEAP.-FPHGP....DAS...GADTN.....CSP........................RG-SCGSTEELH-DGQK-EG.......EQRE-LQGGDR-FAA--F------VV-A---E-   81
SMM.US.x.SIVsmH635SB10 -----M-.....KEAP.-FPHGP....DAS...GADTN.....CSP........................RG-SCGSTEELH-DGQK-EG.......EQRE-LQGGDR-FAA--F------VV-A---E-   81
MNE.US.x.MNE027        --L--M-.....KEAP.-FPHGP....NAS...GADTN.....CSP........................RGPSCGSAKKLHAVGQ--ER.......-QKEALQGGGR-FAA--F------VV-AH----   81
STM.US.x.STM           --L--M-.....KEAP.-LPHGP....NTS...GADAN.....CPS........................RRPSCGSAEKLHAAGQE-ER.......EQED-LQGGDR-FAA--F------VV-AH----   81
DEB.CM.99.CM40         --EYS--LQETQELSRRNWEVHPGEHSRSGILDYQGTST-AGEQPPE...................................................GGNNE-REVP-FNGAEI-------T--H---R   79
DEB.CM.99.CM5          --ECS--LPETQEFFGRDWEIHPREHSRSGLFNNQGTCARGKECSPG...................................................GHNDQ-REVPRFNGAEI-------TA--V--K   79
RCM.GA.x.GAB1          --ET---S....GEAE.ELSFDF....-DS....LC-RD-EQL.......................RPCRRDAKELSEEGR-TKET-EAG.......REQEERGSIS-S-N---FA--K--TTIAQ----   87
RCM.NG.x.NG411         --ECS-E.....AG--.E-PSL-....FLE.....TY......CSR...........ERGRMESPVRSSRRDAQEIPSIGETAQTGAEER.......EQGESG-S---CVV---F---K--VTW-E----   91
DRL.x.x.FAO            --EYS--.Q...RETQ.EFSSGL................ADPISSS..........IARIPPGGSGREDAPKVHGAGSSAEET-ET-T............-EEG----------F--RL--TTI-EV---   87
MND_1.GA.x.MNDGB1      -------S....LQTG.ELSGTR....GDS.....................................NS-TIRGETSA....-NS-HLSEIRE..RAQAEDEGGEE---FSF-EY--S---IEE-SVD-V   78
MND_2.CM.98.CM16       --EY---.Q...WQTQ.ELPGDA................IDPNGPS....................NA-DGRPSRENAVRLH-ERA--ESGKQ....TGKEREGSI--S-----F---N--TTV-E----   85
MND_2.GA.x.M14         --ECS--.Q...WQAQ.ELSCHA................TDPISTP....................DA-DGG-SRED-IRLY-KGA-TEGSKGD...QAGEGQGS---S-----F---N--TTV-E----   86
MND_2.x.x.5440         --EY---.Q...WQAQ.ELPCHA................IDPISTP....................DA-VGRPGR-DAVRLY-ERA-TEGSSRS...PERGKEGS---S-----F---N--TTV-E----   86
GRV.ET.x.GRI_677       -------.....RS--.KK-CAI....QRR...........................HSW-GTNSPPNGN-LRSS-EAPPAVCREG--PER....GERTDKE--GE-S-GCFLELP-------R-I---T   89
TAN.UG.x.TAN1          --ENANLG....NKAA.-F-RAR....RSR.SDGSPDA.....CARIP.........HRFASCGS..I-P-QETAG---Q......KG.........RSIEKA-GEGT-GLFLEFP------SKI--Q--   89
VER.DE.x.AGM3          -----VD.....GS--.-K-SRR....YSW...GGAN.....CAP.....STES.............I-PCKEAPA-I-RQ--AV-GTKE.......-TTSSESR-DR-IFFEL.P-----I-------V   86
VER.KE.x.9063          ----TVD.....GDK-.EKLSRC....HSW...SGTE.....RAP.....STDT.............I-P-KEAPA---RE-K-V-GTRE.......ENTTNKSR-G-RIFFEL.P-----I---I---V   86
VER.KE.x.AGM155        -----VD.....GDK-.KK-SRS....HSW...GGTK.....CAP.....STEQ-YT..........L-P-KEAPA---RER--N-KSEQ.......--PSEQSR-ER-IFFEL.P-----IR-CI-G-T   89
VER.KE.x.TYO1          ----TVD.....GGK-.EK-SRR....YSW...SGTE.....CAS.....STERHHP..........I-P-KEAPA-I-RER--T-G-KE.......ESTGNESG-DR-IFFEL.P-----I-------V   89
BKM.CD.02.AY518534     ..................................................................................................................................    0
TAL.CM.00.266          --EVFFR.....ENP-.GQRTEE....................................F-SGA-IGTANRA-HSIMAGGDYHVDP-SQE........APRNNHNKGES-S-FGQ------IVK-L----   80
TAL.CM.01.8023         --EGFFR.....EDPP.GQGTEE....................................LPSGA-HRTADRA-HSIMAGGDYHMDPSDQG.......CRSTQATDEGES-S-VSQ------IVK-L---R   81
TAL.US.x.1511          ..................................................................................................................................    0
TAL.US.x.1867          ..................................................................................................................................    0
TAL.US.x.416           ..................................................................................................................................    0
TAL.US.x.742           ..................................................................................................................................    0
MON.CM.99.L1           --ESFRS.QEGAKELP.SHLCAAQRTDSAS........................................R..DCG-S.Q-Q.PRDWFP........EDGDPKGDEA-S-S-LSI---G--IRS-L-G--   76
MON.NG.x.NG1           --ESFWPR....-EGA.K-LS-NLLAA............................NSPSGTNPXGXQGELLT-SGTTRDP.................SADSPEGDEE-S-S-LSIP--G--VR--E-A--   80
MUS_1.CM.01.1085       --GSLRPP....EK-A.E-L-P-LSLA..............................NCSARTGGGIGQ-LS-SGSRGG.......................ETGQEES-SF-VIP-SG--VRP-E----   72
LST.CD.88.447          --L-T-E.....EWSS.RKLSF-....GGE...CR--T.......................ISSPNGEEPNT-RESLR.NIQ-PRT--EE.......ATA.DPTEMC-RAMPEFLL-S---IENILVN--   85
LST.CD.88.485          --L-T-E.....EWSS.RKLSF-....GGE...CR--T.......................ISSPNGKELNK-RESLR.NIQ-PRT-TEE.......ATA.GSTEMC-RAMPEFLL-S---IENILVNR-   85
LST.CD.88.524          ----S-E.....EWTF.RQLSNH....GRG...CR-DS......................FSATHGESA-NK-REGNR.DIQ-PRT-TEE.......ATT.SSTEVC-RALPEFLL-S---IEK-LVN--   86
LST.KE.x.lho7          --I-S-E.....-QTS.RQLS-T....GRS...CRENS.......................ISSPNGEC-NK-RESTG.DLQDFRT--EK.......AAA-SPTEMC-RALPEFTL.S---IESILVN--   85
SUN.GA.98.L14          ------E.....-GAA.R-LS-H....AF.............CSA.....SR-SDIGE........QNDTTPVKGRESSGDLQ-VR-GPQ.......STAEEEGKGGMSGAMLEYA-S---LEK-F-N--   87
SYK.KE.x.KE51          --ERVGE.....-ETP.SK-SC-..........EGDTN....SPT.....S--LAMAATDMGKLCFTPTEAH-NGHPATATVLSKDSQGPLGPWETFREERRD-GEGTS-S--SI-------MEINV--T  105
SYK.KE.x.SYK173        --ERMG-....L-EAP.SQL--E..........ERAIC....SPS.....G-HRRWAMAHMVSTDEPTST-EGTELSLEETPHQQGS-L-.......EGEQWEGGDQES-S-S-L-------IE-DV--D   99
COL.CM.x.CGU1          --.........................GN-............................................GGGATD-HLPP-DEQRP........ATQEGEGGES-S-LE---R---IVI-EV---   53
```

```
                           Asp25 catalytic site
                             |
                                                                                                        Protease \/ Pol p51 RT
SAB.SN.x.SAB1C         KVTALLDTGADDSVIQGIEL....GDNWKPRIIGGIGGCINVKAYHNQEVKIEDKTCK.....ATILVGETPVNIIGRNVLAQLGVTLNLTQREIEPI...KVHLKPGQDGPRIRQWPLSKEKIEALKAI   223
H1B.FR.83.HXB2         LKE---------T-LEEMS-....PGR---KM------F-K-RQ-DQILIE-CGHKAI.....G-V---P---------L-T-I-C---FPISP--TV...P-K----M---KVK----TE---K--VE-   191
H1B.US.90.WEAU160      LKE---------T-LEDMN-....PGK---KM------F-K-RQ-DQVPIE-CGHKAI.....G-V---P---------L-T-I-C---FPISP--TV...P-K----M---KVK----TE---K--VE-   191
H1N.CM.95.YBF30        VRE---------T--EELQ-....EGK---KM------F-K-RQ-D-IT-D-QGRKAV.....G-V---P---------L-T-I-C---FPISP--TV...P-K----M---KVK----TT------RE-   198
H1O.BE.87.ANT70        LCEV--------T-LNN-Q-....EGK---KM------F-K--E-D-VT-E--GREVQ.....G-V---P---------I-TG--C---FPISP-A-V...P-K----M---KVK------------T--   187
H1O.CM.91.MVP5180      LCE---------T-LNN-Q-....EGR-T-KM------F-K--E-N-VT-EVQG-EVQ.....G-V---P-----L---I-TG--C---FPISP-A-V...P-K----M---KVK-----R------T--   187
CPZ.CD.90.ANT          -CQ---------T-VE--H-....QG----KT------F-S-QQ-NKVPIQ-G-R-VL.....--V-L-PN----------CL--C---FPISKV-TV...P-K--E-M----VK-------------E-   191
CPZ.CM.01.SIVcpzCAM13  LCE---------T---D-Q-....QGI---KM------F-K--Q-D-VSIE--G-KVQ.....G-V---P---------I-T-I-C--VFPISS--TV...P-K----M-----K-----A---K--TE-   190
CPZ.CM.05.SIVcpzEK505  IKE---------T--EE-Q-....EGK---KM------F-K--Q-D-VIIE-QG-KAV.....G-V---P---------F-T-I-C---FPISP--T-...P-K----M----VK----TE---K--TE-   197
CPZ.CM.05.SIVcpzLB7    LKG---------T-LEDL--....-GR---KM------F-K--Q-D-IPIE-GG-KAL.....G-V---P---------I-T-I-C---FPISP--TV...P-Q----M---KVK----TE---K--TE-   194
CPZ.CM.05.SIVcpzMB66   LKE---------T-LEE---....-GR---KM------F-K-RQ-D-VTID-CG-RAV.....G-V---P---------I-T-I-C---FPISP--TV...P-S----M----VK----TE---R--TE-   194
CPZ.CM.05.SIVcpzMT145  LIE---------T--EDLN-....PGK-S-KM------F-K--QFERVNIE--G-KIT.....S-V---P-----------TK--C---FPISP-DTV...--F----M---KVK------------TE-   190
CPZ.CM.98.CAM3         LIE---------T--DNVQ-....TGK-R-KM------F-K--Q-D-IAIE--GRKTT.....G-V---P---------I-T-I-C---FPISP--TV...P-K----M----VK----TA---K--TE-   196
CPZ.CM.98.CAM5         LIE---------T--DNVQ-....TGK-R--M------F-K--Q-D-IVIE--GRKTT.....G-V---P---------I-T-I-C---FPISPV-TV...P-K----M----VK----TA---K--TE-   192
CPZ.GA.88.GAB1         LCE---------T--ER-Q-....QGL---KM------F-K--QFD-VHIE--GRKVV.....G-V---P---------I-T---C--VFPISS--TV...P-K----M---KVK-----A---K--TE-   189
CPZ.GA.88.GAB2         LCE---------T-LDN-Q-....EGK-R-KL------F-K--Q-D-V-IE--G-RVE.....S-V---P-----------T---C---FPISP-T-V...P-K----T----VK------------TE-   194
CPZ.TZ.01.TAN1         VCQ-----------FCN-K-....KGQ-T-KT------FVP-SE-Y-IP-Q-GN-EVR.....--V------I------I-K---C---FPISP--VV...--Q--E-M---KVK------------TE-   187
CPZ.US.85.CPZUS        I-E---------T-LDN-QI....EGT-R-KMM-----F-K--Q-DHVNIE--GRKAQ.....GSV---P---------I-T-I-C---FPISS-KTV...P-R----M----VK----TA---K--TE-   193
GOR.CM.06.CP684        ..................................................................................................................................     0
GSN.CM.99.CN166        V-SM--------TIV-DSAI..ELDTP-T-KMV-----L-Q-REHKHV--IFNE-RI-.....--V---P-----L---C-SK------MV-QK----...E----A-KE--K-K-----R------TQ-   196
GSN.CM.99.CN71         V-SM--------TIV-ESAI..ELDTP-T-KTV-----L-Q--EHRHV--IFNE-RI-.....--V---P-----L---C-SK------MV-QKL--V...E----S-KE--K-K------------TQ-   195
H2A.DE.x.BEN           P-EV---------IVA----....---YT-K-V-----F--T-E-K-V-I-VLN-RVR.....---MT-D--I--F---I-TA--MS---PVAK----...--T----K----LK----T--------E-   216
H2A.GW.x.ALI           P-EV---------IVA----....-S-YT-K-V-----F--T-E-EDV-I-VLN-RV-.....---MT-D--I--F---I-TA--MS---PVAK----...E-R----K---KL-----T--------E-   216
H2A.SN.x.ST            P-EV---------IVA-V--....-S-YS-K-V-----F--T-E-K-V-IRVLN-RVR.....---MT-D--I--F---I-TA--MS---PVAK----...-IM----K---KL-----T--------E-   216
H2B.CI.x.EHO           S-EV---------IVA----....-S-YT-K-V-----F--TNE-K-V-IEVVG-RVR.....--VMT-D--I--F---I-NS--M---FPVAR---V...--Q---EK---K-----------L---E-   216
H2B.GH.86.D205         S-EV-----V---IVA----....-S-YT-K-V-----F--T-E-KDV-IEVVG-RVR.....---MT-D--I--F---I-NT--M---FPVAKV--V...--E----K---K-------R---L---E-   217
H2G.CI.x.ABT96         P-EV---------IVA----....-S-YS-K-V-----F--T-E-K-V-I-VVG-RIQ.....--VMT-D--I--F---I-VK--MS--FPIAKV--V...--R----M---K-----------Q--RE-   199
H2U.FR.96.12034        P-KV--N------IVA--K-....EPYYT-KTV-E-E-F--T-KF--VNI-VLG--IR.....G--MT-D--I--F---L-TE--MS--FPIGKEALV...IGG----K---KL---------L----E-   213
MAC.US.x.239           P-EV---------IVT----....-PHYT-K-V-----F--T-E-K-V-IEVLG-RI-.....G--MT-D--I--F---L-TA--MS--FPIAKV--V...--A----K---KLK---------V--RE-   203
MAC.US.x.251_1A11      P-EV---------IVT----....-PHYT-K-V-----F--T-E-K-V-IEVLG-RIR.....G--MT-D--I--F---L-TA--MS---PIAKV--V...--A----KV--KLK---------V--RE-   203
MAC.US.x.251_BK28      P-EV---------IVT----....-PHYT-K-V-----F--T-E-K-VKIEVLG-RI-.....G--MT-D--I--F---L-TA--MS---PIAKV--V...--T----KV--KLK---------V--RE-   199
MAC.US.x.EMBL_3        P-EV---------IVT----....-PHYT-K-V-----F--T-E-K-V-IEVLG-RI-.....R--MT-D--I--F---L-TA--MS---PIAKV--V...--A----KV--KLK---------V--RE-   199
SMM.SL.92.SL92B        PLEV---------IVA----....-PGYT---V-----F--T-E-RTVKIEVLG-VIR.....E-LMT-D--I--F---I--A---S--YPVAKV-YT...--K--E-M----LK---------Q--TE-   196
SMM.US.x.H9            PIEV---------IVA----....-P-YT-K-V-----F--T-E-KDVKI-VLG-VI-.....G--MT-D--I--F---L-TAM-MS---PIAKVX--...--T----K---KL----------I--RE-   199
SMM.US.x.PGM53         P-EV---------IVA----....-P-YT-K-V-----F--T-E-KDVKI-VLG-VI-.....G--MT-D--I--F---L-TAM-MS---PVAKV---...--T----K---KL----------V--RE-   199
SMM.US.x.SIVsmH635FC   P-EV---------IVT----....-P-YT-K-V-----F--T-E-KDVKI-VLG-VI-.....G--MT-D--I--F---L-TAM-MS--FPIAKV---...--T----KE--KV----------I--RE-   199
SMM.US.x.SIVsmH635F_L3 P-EV---------IVT----....-P-YT-K-V-----F--T-E-KDVKI-VLG-VI-.....G--MT-D--I--F---L-TAM-MS--FPIAKV---...--T----KE--KV----------I--RE-   199
SMM.US.x.SIVsmH635SB10 P-EV---------IVT----....-P-YT-K-V-----F--T-E-KDVKI-VLG-VI-.....G--MT-D--I--F---L-TAM-MS--FPIAKV---...--T----KE--KV----------I--RE-   199
MNE.US.x.MNE027        PAEV---------IVA----....-PHYT-K-V-----F--T-E-K-V-IEVLG-RI-.....G--MT-D--I--F---L-TA--MS--FPIAKV--V...--T----K---KL----------V--RE-   199
STM.US.x.STM           P-EV---------IVA----....-LQYT-KVV-----F--T-EFK-VNIEVLG-KI-.....G--MT-D--I--F---L-TA--MS--FPVAKV--V...--T----K---K-K-------------E-   199
DEB.CM.99.CM40         --QV-I------TI-SEKDI..DLEAP-T-KTV--L--F----C-PGI-IAMAN-VAV.....SDL---D--I--L---Y--KM--SI-FPVLDYT-V...-L..-E-A---K-K-----R-------D-   197
DEB.CM.99.CM5          R-QV-I------TI-SEKD-..DLDAP-T-KTV--L--F----C-PGI-I-MA--VAV.....SDV---D--I--L---Y--KM--SI-FPVLDFT-V...-L..-E-A-----K-----R-------E-   197
RCM.GA.x.GAB1          --EV--------T--E----....-ND-T-K-------Y----Q-K-C-IE-AG-RTH.....-HV---P-----------KK--A---FPISQ--T-...--E--S------VK------------TE-   205
RCM.NG.x.NG411         --E--------GT---NLD-....-E----K-------Y----QFNHCKIT-AG--TH.....-HV---P-----------SK--AS--FPISKA-T-...--E---------VK------------TE-   209
DRL.x.x.FAO            --E---------T----L--....TGK---Q-------T---RQ-F-C-I--AG-ITH.....-SV---P-----------CK--A---FPVSQV--V...--T----M---K-K-------------E-   205
MND_1.GA.x.MNDGB1      TIR---------TIFNERNI..KLKG--Q-K-------NLR--Q-D-VY-E-RG-GTF.....G-V-I-P--ID-----IMEK--GK-I-A-LSDKIP.IT--K----V-----K---------VG-QK-   200
MND_2.CM.98.CM16       --E---------T--KDLD-....KG----Q-------S----QFF-CK-T-AG--TH.....-SV---P-----V-----KK--C---FPVSKV-TV...--T----T---K-K---------L--QE-   203
MND_2.GA.x.M14         --E---------T--KDVD-....TG----Q-------A-R--Q-F-CKITVAG-STH.....-SV---P-----------KK--C---FPISK--TV...--T----M-----K---------L--QT-   204
MND_2.x.x.5440         --E---------T--KDL--....TG------------A-R-RQ-F-CK-E-AG-ITH.....-AV---P-----------KK--C---FPISK--TV...--T----T-----K---------L--QE-   204
GRV.ET.x.GRI_677       P-Q---------TI--EKD-HFPPHKP-RSKVV-----G-H--E-QGVQ-QL---IIT.....GS--I-S--I------I---A-MK-VMGVLSSQIE.ET--Q--E-K---KLK-----R------TE-   213
TAN.UG.x.TAN1          P-QV--------TI-KEEDIHL..EGR-R-KVV-----G-S--Q-D-IL-N-DG-ETR.....GNV-I-P--I--V---I--P--AR-IMG-LNEKIP.IT--Q----MS--KLK-----E---K--TE-   211
VER.DE.x.AGM3          PIR---------TI-KEAD-..QLSGT---K-------GL---E-SDR--RL---ILR.....G---I-S--I------I--PA-AK-VMG-LSEQIP.ITP-K--E-AR--FLK---------K--QE-   208
VER.KE.x.9063          P-KV--------TI-KETDI..QLTGP-R-KV------GL---E-S-V-------ILR.....G-V-L-A--I------L--PA-AR-VMG-LSEQIP.ITP-K--D-AR--H-K---------K--TE-   208
VER.KE.x.AGM155        A-K---------TI-KDTD-..QLRGS-R-K-V-----GL---E-D-V--QL---ILR.....G-V-I-A--I------F---A-AK-VMG-LSQTIP.ITP-R--E-AR---LK---------I--QE-   211
VER.KE.x.TYO1          PIK---------TI-KEND-..QLSGP-R-K-------GL---E-NDR-------ILR.....G---L-A--I------L--PAVPR-VMG-LSEKIP.VTP-K--E-AR--CV-------------QE-   211
BKM.CD.02.AY518534     ..................................................................................................................................     0
TAL.CM.00.266          E-EV--------TI--VG-I..ELEGVPHQKTV--LA-Y-Q-QS-SAI-IQW-N-RAR.....D-V-I-P--I--L---F--KFS----MVSSD-PIT...--K--E-KEP-K-K----TR----G-QI-   200
TAL.CM.01.8023         E-EV--------TI--EGDI..ELEGVPHQKTV--LA-Y-Q-QS-TGV-IEW-H-RAR.....D-V-I-P--I--L---F--KF-----MVST--PV-...--K--E-KEP-K-K----T-----G-Q--   201
TAL.US.x.1511          ..................................................................................................................................     0
TAL.US.x.1867          ........----TI--EG-I..NLDGVPHQKTV--LA-Y-Q-QS-SAV-IQW---RAR.....D-V-I-P--I--L---F--KF-----MISTD-PIT...--K--E-KEP---K----TRK---G-Q--   112
TAL.US.x.416           ..................................................................................................................................     0
TAL.US.x.742           ..................................................................................................................................     0
MON.CM.99.L1           --SM--------TI--DQHV..NL-E--T-KVV-----M-E--Q-KAIQ-IF--REVW.....--V---P--I--L------KM-----MAAGGDLQPPSLE-T--A-KE--KLK-----R------TQ-   199
MON.NG.x.NG1           S-SM----X---TIV-XQDI..NL-E--T-KMV-----L-Q--QFKGVT-SFQG--VR.....X-V-L-X--I--L-----SRM-I---MX-GIAQSP.SLT-T----K---KLK------------TT-   202
MUS_1.CM.01.1085       V-SM--------TI--DAHI..HLEGP-N-KSV-----N-S--E-RGVQ-CFNG--IR.....--V---P--I--M---C-SKF-I---MI-GKL---...--T--E-AK--MVK----TQ---R--ED-   192
LST.CD.88.447          P-S---------TIFSENSV..RIEGPYT--VV-----Q-R--E-KDVFIE-AG--TV.....G-V-L-P---D-V---I-TNV-AR-V-S-LSDKIP.IT--K----C----VK----------G-Q--   207
LST.CD.88.485          P-S-----R---TIFSENSV..KIEGPYT--VV-----Q-R--E-KDVFIE-AG--TV.....G-V-L-P---D-V---I-TSV-AK-V-A-LSDKIP.IT--R----C----VK----------G-Q--   207
LST.CD.88.524          E-----------TIFSENSV..RIEGPYT--VV-----Q-K--EFRDVFIQ-AG--TV.....G-V-L-P---D-V---I-NP--AR-V-A-LSEKIP.IT-------C----VK----------G-Q--   208
LST.KE.x.lho7          P-S---------TIFSENSV..RIEGPYT--VV-----Q-R--E-RDVFIE-AG--TV.....G-V-L-P---D-V---I-TAM-AK-I-A-LSDKIP.IT--S----C----VK----------G-Q--   207
SUN.GA.98.L14          P-----------TIVSEDSV..SIEGH----VV-----Q-R--E-ADVF-E-R---AT.....G-V---P---D-----I-SA--GR-V-ATLSEKIP.IT--K--E-AA--KVK-----R----G-QK-   209
SYK.KE.x.KE51          ---M--------TI-REQDI..QLHQP-T-K-V--L--N-R-RQ-NQIRFQ--RPDGRIKEVEGSL---P-----L---I-SK--AK-VMIAQKL--M...--Q----KEL-KLK----TR--L---R--   230
SYK.KE.x.SYK173        L-QM-V------TI-REEDI..QLHQP-S-K-V--L--N-T-RQ-R-IRFTVVKPSG-RKQVEG-L---P-----L---I-TK---K-VMV-TAL--V...--S---DKEL--LK-----V--L------   224
COL.CM.x.CGU1          -CQ---------T-FKEEDV..QLTGA-TAVQ-Q----A-R--Q-RDKIL-MGG-EYR.....GD--I-H--I--V---M-KEGKLVVAQLSDR-PVT...--T--E-M---KVK----------G-QK-   173
```

```
                          |-  drug resistance sites   -|                                      catalytic site
                         M41L                A62V D67N K70R                                       Asp110
                          |                    |    |  |                                             |
SAB.SN.x.SAB1C         CEDLEKQGHLERIGPENPYNTPVFAIRKKDKTQWRILMDFRQLNKSTQDFQEVQLGIPHPAGLQQREQIT......VLDIGDAYFSCPLDPDFQKYTAFTIPSVNNREPGIRYQYKVLPQGWKGSPTIFQ   347
H1B.FR.83.HXB2         -TEM--E-KISK--------------K---S-K--K-V---E---R----W------------KKKKSV-......---V------V---E--R---------I--ET-------N----------A---   315
H1B.US.90.WEAU160      -TEM--E-KISK--------------K---S-K--K-V---E---R----W---------S--KKKKSV-......---V------V---E--R---------I--ET-------N----------A---   315
H1N.CM.95.YBF30        -TEM--E-KIS-----------I---K---S-K--K-V---E---R----W------------K-KKSV-......---V----------K--R---------I--ET-------N----------A---   322
H1O.BE.87.ANT70        -QEM-QE-KIS-----------I---K---G-K--K-V---E---R--E-W---------G--K-KQSV-......---V-------------R------------ET-------N----------A---   311
H1O.CM.91.MVP5180      -QEM-QE-KIS-----------I---K---S-K--K-V---E---R----W---------G--K--QSV-......---V-------------R------------ET--V----N----------A---   311
CPZ.CD.90.ANT          -DK--AENKIS----D------I---K---TSK--K-V---E---R----W-I----------K-KKSV-......---V------I---Q--R------------ET-----C-N----------A---   315
CPZ.CM.01.SIVcpzCAM13  -QEM--E-KISK----------I---K---SSK--K-V---E---R----W------------KEKKSV-......---V----------K--R------------ET-------N----------A---   314
CPZ.CM.05.SIVcpzEK505  -TEM--E-KIS-----------I---K---S-K--K-V---E---R----W------------KKKKSV-......---V----------EN-R------------ET-------N----------A---   321
CPZ.CM.05.SIVcpzLB7    -TEM-QE-KIS---------------K---S-K--K-V---E---R----W------------KKKRSV-......---V----------EN-R---------I--ET-------N----------A---   318
CPZ.CM.05.SIVcpzMB66   -TEM--E-KIS-V---------I---K---S-K--K-V---E---R----W------------K-KKSV-......---V----------EN-R------------ET-------N----------A---   318
CPZ.CM.05.SIVcpzMT145  -KEM--E-KISK----------I---K---S-K--K-V---E---R--E-W------------KKKKSV-......---V-------------R-F-------I--ET-------N----------S---   314
CPZ.CM.98.CAM3         -QEM--E-KIS-----------I---K---S-K--K-V---E---R----W------------KKKKSV-......---V----------EN-R------------ET-------N----------A---   320
CPZ.CM.98.CAM5         -QEM--E-KISK----------I---K---S-K--K-V---E---R--E-W------------KKKKSV-......---V----------EN-R------------ET--V----N--X-------A---   316
CPZ.GA.88.GAB1         -QEM--E-KISK----------I---K---S-K--K-V---E---R----W------------KKKKSV-......---V----------K--R---------I--ET--V----N----------S---   313
CPZ.GA.88.GAB2         -REM--E-KIS-----------I-------S-K--K-V---E---R----W------------KKKKSV-......-P-V----------EN-R------------ET-------N----------A---   318
CPZ.TZ.01.TAN1         -KT---E-KISAV---------I---K---TSK--K-V---E---R----W-L----------RK-NMV-......---V------I------R---------L--NT--K-F--N----------A---   311
CPZ.US.85.CPZUS        -QEM--E-KIT-V---------I---K---S-K--K-V---E---R----W------------KKKRSV-......---V----------KE-R------------ET-------N----------A---   317
GOR.CM.06.CP684        ..................................................................................................................................     0
GSN.CM.99.CN166        TQEMLRL-Q--K--------S-----K----S---M-I---K--EA---LA-----------IPHPAGLEQKEHVTII-MK---Y-I--YE--R-----SV-----QA-AE---F--------A------   326
GSN.CM.99.CN71         TQEML-L-Q--K--------S-----K----S---M-I---K--EA---LA-----------......LEQKEHVTII-MK---Y-V--YEE-R-----SV-----QT-AE---F--------A------   319
H2A.DE.x.BEN           --KM--E-Q--EAP-T------T---K----NK--M-I---E---V----T-I----------AKKKR-S......I--V------I--HE--RQ-----L-A---M---K--I------------A---   340
H2A.GW.x.ALI           --KT-RE-Q--EAP-T------T---K----NK--M-I---E---V----T-I----------AKKRR--......---V------I--HES-RQ-----L-----A---K--I------------A---   340
H2A.SN.x.ST            --KM-RE-Q--EAP-T------T---K----NK--M-I---E---V----T-I----------AKKKR--......---V------I--HE--RQ-----L--I--A---K--I---S--------A---   340
H2B.CI.x.EHO           --KM--E-Q--EAP-T----S-T---K----NK--M-I---E---V--E-T------------ASKKR--......---V------V------RQ-----L-A---A---K--L------------A---   340
H2B.GH.86.D205         --KM--E-Q--EAP-T------T---K----NK--M-I---E---V----T--NX-------XAEKRR--......-I-V------I----N-RQ-----L-----A---K--I-----------QS-C-   341
H2G.CI.x.ABT96         -DKM-QE-Q--EAP-T------T-T-K----NK--M-I---E--RV----T------------AEKRR--......----------I---V--RQ-----L-----A---K--I------------A---   323
H2U.FR.96.12034        --KM--E-Q--EAP-T------T---K----NK--M-I---E--RV----S------------AEKKR--......---V------I--CE--RQ-----L--I--A---K--I----------------   337
MAC.US.x.239           --KM--D-Q--EAP-T------T---K----NK--M-I---E--RV----T------------AK-KR--......----------I---EE-RQ-----L-----A---K--I------------A---   327
MAC.US.x.251_1A11      --KM--D-Q--EAP-T------T---K----NK--M-I---E--RV----T------------AK-KR--......----------I---EE-RQ-----L-----A---K--I------------A---   327
MAC.US.x.251_BK28      --KM--D-Q--EAP-T------T---K----NK--M-I---E--RV----T------------AK-KR--......----------I---EE-RQ-----L-----A---K--I------------A---   323
MAC.US.x.EMBL_3        --KM--D-Q--EAP-T------T---K----NK--M-IH--E--RV--ELYRSPIR-------AK-KR--......----------I---EE-RQ-----L-----A---K--I------------A---   323
SMM.SL.92.SL92B        --KM--E-Q---AP-T------T---K--N-DK--M-I---E---M--E-T------------KEM-R--......----------V----E-RQ-----------Q---K--I------------A---   320
SMM.US.x.H9            --KM--D-Q--XAP-T------T---KXX-GNK--M-I---E---V----T---------X--AK-RR--......---V------I---EE-RQ-----L-----A---K--I------------A--X   323
SMM.US.x.PGM53         --KM--D-Q--EAP-T------T---K----NK--M-I---E--RV----T------------AK-RR--......---V------I---EE-RQ-----L-----A---K--I------------A---   323
SMM.US.x.SIVsmH635FC   --KM--D-Q--EAP-T------T---K----NK--M-I---E---V----T------------AK-RR--......---V------I---EE-RQ-----L-----A---K--I------------A---   323
SMM.US.x.SIVsmH635F_L3 --KM--D-Q--EAP-T------T---K----NK--M-I---E---V----T------------AK-RR--......---V------I---EE-RQ-----L-----A---K--I------------A---   323
SMM.US.x.SIVsmH635SB10 --KM--D-Q--EAP-T------T---K----NK--M-I---E---V----T------------AK-RR--......---V------I---EE-RQ-----L-----A---K--I------------A---   323
MNE.US.x.MNE027        --KM--D-Q--EAP-T------T---K----NK--M-I---E---V----T------------AK-KR--......---V------I---EE-RQ-----L-----A---K--I------------A---   323
STM.US.x.STM           --KM--D-Q--EAP-T------T---K----NK--M-I---E---V----T-I----------AK-RR--......---V------I---EG-RQ-----L-----A---K--I------------A---   323
DEB.CM.99.CM40         LKPMIEA-QIVPAA-T----S---V-K----SK--M-I-L-A---A--EVW---T-----SA-P-M----......---LA---Y-IA---T-A------V-----IQ--E--EFR------SA--A---   321
DEB.CM.99.CM5          IRPMVEA-QIVPTE-T---SS---V-K----NK--M-I-L-A---A---VW---T-----S--P-M----......---LA---Y-I----Q-AR-----------TQ--E--EFR------SA--A---   321
RCM.GA.x.GAB1          -NAM--E-KISK----------I-C-K---S-K--K-V---E---R----F---------G--K-C-R--......-----------L-YEP-R--------A---QG--V----N----------A---   329
RCM.NG.x.NG411         -NAM--E-KIS-----------I-C-K---S-K--K-V---E---R----F---------G--K-CR---......-------------YEP-R-------LAI--QG-------N----------A---   333
DRL.x.x.FAO            --QM--E-QIS-------------C-K---G-K--K-V-------A----F---------G--KAC----......---V-------------R------------QG-------N----------A---   329
MND_1.GA.x.MNDGB1      -DR--EE-KIS-VD-G-N----I---K----NE--K-I---E---L----H-L---------IKKCKR--......----------I-----YRP-----V-----QA--K--M-N----------C---   324
MND_2.CM.98.CM16       -SQM--E-QISK------------C-K---G-S--K-V-------V----F---------G--K-C----......--------------E--R------------QG-------N----------A---   327
MND_2.GA.x.M14         -DEM--E-KIS-------------C-K---G-S--K-V------RV----F---------G--K-C----......I-------------ES-R------------QG-------N----------A---   328
MND_2.x.x.5440         -NQM--E-KIS-V-----------C-K---G-S--K-V-------V----F------T--G--K-C----......--------------V--R------------QG-------N----------A---   328
GRV.ET.x.GRI_677       -KQM-EE-K-S---G-----------K--------M-V---E---A----F-------------KKK---......-I------Y-I--CKE-R------------TG------FNC-------------   337
TAN.UG.x.TAN1          TQA---E-KISKV-G---------C----------M-I---A---A----F------------KRKK---......--------Y-I----E-RP-------Q---VG------FNC-------------   335
VER.DE.x.AGM3          -DQ---E-KISK--G--A------C-K----S---M-V---E---A----F---------S-FEKMTE--......--------Y-I----E-R------------QG--T---FNC-------------   332
VER.KE.x.9063          --Q---E-KISK--G--A----I-C----------M-V---E---V----F-------------KMK---......--------Y-I----E-R------------QG--V---FNC-------------   332
VER.KE.x.AGM155        -KT--EE-K-S-V-GD-A------C------S---M-V---E---A----F------------KKMK---......II-V----Y-I----E-R--------T---EG------FNC-------------   335
VER.KE.x.TYO1          -SQ--QE-KIS-V-G--A----I-C-K----S---M-V---E---A----F------------RKMR---......---V----Y-I----N-R--------T---QG------FNC-------------   335
BKM.CD.02.AY518534     .............................................................X-EK-KNVS......IV-MK---Y-V----S-SP-----------EQ--V--VF---------------    63
TAL.CM.00.266          ID-MV-A-Q--EV-------S--------------M-I---A--QA----A---Q-----S--P-MK---......---MK---Y-I------SQ--------L--DR--R---F---------------   324
TAL.CM.01.8023         ID-MV-A-Q--E--------S--------------M-I---A--QA----A---Q-----S--P-KK---......---MK---Y-I------AR--------L--DR--R---F---------------   325
TAL.US.x.1511          ..................................................................................................................................     0
TAL.US.x.1867          ID-MV-A-Q--E--------S--------------M-I---A--QA----A---Q-----S--P-MK---......---MK---Y-I----E-S---------L--DR--R---F---------------   236
TAL.US.x.416           ..................................................................................................................................     0
TAL.US.x.742           ..................................................................................................................................     0
MON.CM.99.L1           TQEML-L-Q--PTE-N----S-C---K----SK--M-I---E---A----F------------EKMDHV-......IV-MK--FY-I--WEP-R-----S---L--A--AK---F------------V--   323
MON.NG.x.NG1           XREMIEL-Q--EAE-T----S-C----------K-M-I---E---A----F-------------NMAN--......IV-MK--FY-I---EN-R-----S------A--AV---FR----X---------   326
MUS_1.CM.01.1085       TQQML-V-QI-E--------S-C--------SK--M-I------EA--ELA------------IEH-HV-......IV--K--FY-V------R-----SL-Q---LG-AR---F---------------   316
LST.CD.88.447          -DR---E-KISKADAG------I---K----NE--K-I------AL----H-----------IKKCK-V-......-V--------I-----YRQ-----V--I--Q---R----N----------C---   331
LST.CD.88.485          -DR---E-KISKADAG------I---K----NE--K-I------AL---LH-----------IKKCK---......-V--------I-----YRQ-----V--I--Q---R----N----------C---   331
LST.CD.88.524          -DR---E-KISPVD-G------I---K----NE--K-I---K--EL----H-L---------IKKCK---......-V--------I-----YRQ--------L--Q---K----N----------C---   332
LST.KE.x.lho7          -DR---E-KISPVD-G------I---K----NE--K-I---K--EL----H-L---------IKKCK---......-V--------I----NYR---------L--Q---K----N----------C---   331
SUN.GA.98.L14          -DR--AE-KIS-AD-G------I---K----NE--K-I---V--GM----H-L----------KKCK---......-V-V------I-----YR------------QA--K--V-N----------C---   333
SYK.KE.x.KE51          VS-MLEK-Q--KAS-T--------V-Q----LK--MP----K--EA----W---------G--STKR.M-......---LK---YTV---EE-RQ-----V--I--SS--E---F---------------   353
SYK.KE.x.SYK173        V--ML-A-Q--KAS-T--------V------KK--M-I---K--EA----F---------G--KRQK.L-......II-LK---Y-V---KE-RP-----V--I--AS--E---FT--------------   347
COL.CM.x.CGU1          --R--AE-KI-KAELG------I-C------NE--K-I---E---R----M---------G--IEKTH--......----------I-----Y-------------LA--T--V----------------   297
```

```
                                    polymerase motif
                                         |- -|                        drug resistance
                                        catalytic site                  T215Y  K219Q
                                        Asp185||Asp186                      |  |
SAB.SN.x.SAB1C         TTANKILQEFRQKNPDVDIYQYMDDMLIASDRPKAEHLVMVQQLRDYLETWGFKTPEKKFQKDPPYLWMGYELYPKKWQLQEITLPEREE..WTVNDIQKLVGKLNWASQIYTGIKTKHLCRLIRG.ARP   474
H1B.FR.83.HXB2         SSMT---EP--KQ---IV-------LYVG--LEIGQ-RTKIEE--QH-LR--LT--D--H--E--F-------H-D--TV-P-V---KDS..-------------------P---VRQ--K-L--.TKA   442
H1B.US.90.WEAU160      SSMT---EP--KQ---IV-------LYVG--LEIGQ-RTKIEE--QH-LR---T--D--H-----F-------H-D--TV-P-K---K-S..-------------------A---V-Q--K-L--.TKA   442
H1N.CM.95.YBF30        S-MT---EP--E-H-EII-------LYVG--LEL-Q-REA-ED---H-LK---T--D--H--E--F-------H-D--TV-P-K---KDV..-------------------P--RV-Q--K----.TKA   449
H1O.BE.87.ANT70        SSMT---DP--RD--ELE-C-----LYVG--L-LT--RKRIEL--EH-YQ---T--D--H--E--F-------H-D--TV-S-Q--NKDV..---------I---------Q--RVRE--K----.TKS   438
H1O.CM.91.MVP5180      SSMT---DP--KS--E-E----I--LYVG--L-L---RKR-EL--EH-YQ---T--D--H--E--F-------H-D--TV-P-Q--DK-V..-------------------Q--RV-E--K----.TKS   438
CPZ.CD.90.ANT          ASMT---AP--D-Y-A-E-------LYVG--MEITA-RE-IEK--QH-QV--LE--D--H--E--FQ------H-D--TV-K-K---PDD..-------------------P-----Q-------.VKS   442
CPZ.CM.01.SIVcpzCAM13  SSMT---EP--A----ITV------LYVG--LDITQ-RQK-EE--QH-LH---T--D--H--E--F-------H-D--TV-P-K---K-V..-----L---I---------P---I-Q--K----.TKR   441
CPZ.CM.05.SIVcpzEK505  S-MT---EP--KN--ELV-------LYVG--LEITQ-REA-ER--SH-L----T--D--H--E--F-------H-D--TV-T-Q---KDT..-------Q-----------P---V-Q--K----.-KA   448
CPZ.CM.05.SIVcpzLB7    SSMT---EP--RQ---IV-------LYVG--LKIEQ-RNK-EE--AH-LK---T--D--H--E--F-------H-D--TV-P-Q---K-S..-------------------P---I-QI---L--.VKN   445
CPZ.CM.05.SIVcpzMB66   SSMT---EP--KQ--EII-------LYVG--LKIEL-REK-EE--AH-LK---T--D--H--E--F-------H-D--TV-P-Q---K-S..---------I------C---P--RV-Q--K----.TKA   445
CPZ.CM.05.SIVcpzMT145  SSMT---EP--T-H-EII-------LYVG--L-IE--RKR-EE--AH-LK---T--D--H--E--F-------H-D--TV-Q-K---Q-V..-------------------P---I-Q--K----.-KG   441
CPZ.CM.98.CAM3         SSMI---TP---QH--LI-------LYVG--LELR--REK-EL--QH-L----T--D--H--E--F-------H-D--TV-P-Q--QK-T..-------------------P---V-Q--K--K-.TKA   447
CPZ.CM.98.CAM5         SSMT---DP--KQH--LI-------LYVG--LEIGK-REK-EL--QH-L----T--D--H--E--F-------H-D--TV-P-Q--QK-I..-------------------S---V-Q--K--K-.TKA   443
CPZ.GA.88.GAB1         SSMT---EP--E----IT-------LYVG--LEIDQ-RKK-EE--QH-LK---T--D--H--E--F-------H-D--TV-P-Q---K-V..---------I---------P---I-Q--K----.TKK   440
CPZ.GA.88.GAB2         SSMT-V-DP------ELV-------LYVG--LSLE--RKR-E---EH-LR---T--D--H--E--F-------H-D--TV-P-Q--QK-N..-------------------P---I-Q--K----.-KA   445
CPZ.TZ.01.TAN1         SSMT---DP--KEH-----------LY-G--LNEE--RKLIKK--QH-L---LE--D--Y-EK--FM------H-N--TV-N-----P-Q..----H--------------H-----E--K----.VKG   438
CPZ.US.85.CPZUS        SSMT---DP--KQH---I-------LYVG--LNLEK-REK-EL--Q--L----T--D--H-EE--F-------H-D--TV-P-Q--QK-I..-------------------P---I-Q--K--K-.-KA   444
GOR.CM.06.CP684        .....................................................................X---H-D--TV-P-K---K-S..-------------------P--RV-E--K----.TKS    58
GSN.CM.99.CN166        NIVASL-RQI-DQE---VLI-----L--G---KLG--RQV-HKI-NL-TSYNIQ---A-H-P-Y-VK-L----H--G-R--PVE--DQDT..----------------V---HP--R--Q--KC---.-KG   453
GSN.CM.99.CN71         A-VASL-YQI-DQE---V-I-----L--G---KL---RQV-HKI-NL-TSYNIQ---A-H-Q-Y-VR-L----H--G-R--PVE--DQDI..----------------V---HP--R--Q---C---.-KG   446
H2A.DE.x.BEN           Y-MRQV-EP--KA----ILI-----I------TGL--DKV-L--KEL-NGL--S--DE-------FQ---C--W-T--K--KLQ--QKDI..-------------V----A---S--------------.KMT   467
H2A.GW.x.ALI           H-MRQ--EP--KA-Q--ILI-----I------TDL--DKV-L--KEL-NGL--S--DE--------K----G-W-T--K--K-Q--QK-V..---------H--V----A---P--------------.KMT   467
H2A.SN.x.ST            Y-MRQV-EP--KA---IILI-----I------TDL--DRV-L--KEL-NGL--S--DE--------Q------W-T--K--R-Q--QK-V..-------------V----A---P----RN--------.KMT   467
H2B.CI.x.EHO           Y-MA-V-DP--KA-N--T-I-----I-V----SDL--DRV-S--KEL-NNM--S---E-------FK------W----K--K-Q---K-V..-------------V----A-LFP----R-I-K----.KMT   467
H2B.GH.86.D205         YSMR-V-DP--KA-S--I-I-----I------SDL--DRV-S--KEL-NDM--S---E-------FK------W----K--K-Q---K-V..----A--------V----A-LFP----R-I-K----.KMT   468
H2G.CI.x.ABT96         H-MR-V-DP--KA-E-XIXI-----I-V----TDL--DQV-L--KEL-NKX--S--DE--------Q-----XW----K--K-N-X-K-T..-------------V-X--A-L-P-----N-XK----.XMT   450
H2U.FR.96.12034        Y-MRNL-EP--KA-E-AV-I-----I------NDL--DQL-S--KEL-NKT--S--DE-------HK------W----K--K-E----DV..-------------V----A-V-P--------K--K-.KMT   464
MAC.US.x.239           Y-MRHV-EP--KA----TLV-----I------TDL--DRV-L-SKEL-NSI--S---E-------FQ------W-T--K--K-E--Q--T..-------------V----A---P--------------.KMT   454
MAC.US.x.251_1A11      Y-MRHV-EP--KA----TLV-----I------TNL--DRV-L--KEL-NSI--S---E-------FQ------W-T--K--K-E--Q--T..-------------V----A---P--------------.KMT   454
MAC.US.x.251_BK28      Y-MRHV-EP--KA----TLV-----I------TDL--DRV-L--KEL-NSI--S---E-------FQ------W-T--K--K-E--Q--T..-------------V----A---P--------------.KMT   450
MAC.US.x.EMBL_3        Y-MRHV-EP--KA----TLV-----I------TDL--DRV-L--KEL-NSI--S---E-------FQ------W-T--K--K-E--Q--T..-------------V----A---P--------------.KMT   450
SMM.SL.92.SL92B        A-MRQV-EP--KA----LLV-----L--G-N-GLT--DK--T----M-NNL--S---D----N--LQ----L------K--K-E---K-R..-------------V----A---P-----N--KM---.KMT   447
SMM.US.x.H9            H-XRNV-EP--KA----TLI-----I------TDL--DRV-L--KEL-NSI--S---E-------FQ------W-T--K--K-E--Q--T..-----N-------V----A---P--------------.KMT   450
SMM.US.x.PGM53         H-MRNV-GP--KA----TLI-----I------TDL--DRV-L--KEL-NSI--S---E-------FQ------W-T--K--K-E--Q--T..-------------V----A---P---------K----.KMT   450
SMM.US.x.SIVsmH635FC   Y-MRNV-EP--KA----TLI-----I------TDL--DRV-L--KEL-NGI--S---E-------FQ------W-T--K--K-E--Q--T..-------------V----A---P--------------.KMT   450
SMM.US.x.SIVsmH635F_L3 Y-MRNV-EP--KA----TLI-----I------TDL--DRV-L--KEL-NGI--S---E-------FQ------W-T--K--K-E--Q--T..-------------V----A---P--------------.KMT   450
SMM.US.x.SIVsmH635SB10 Y-MRNV-EP--KA----TLI-----I------TDL--DRV-L--KEL-NGI--S---E-------FQ------W-T--K--K-E--Q--T..-------------V----A---P--------------.KMT   450
MNE.US.x.MNE027        H-MRNV-EP--KA----TLV-----I-V----TDL--DRV-L--KEL-NSI--S---E-------FQ------W-T--K--K-E--QK-T..-------------V----A---P--------------.KMT   450
STM.US.x.STM           Y-MRN--EP--RA----TLI-----I------TDL--DRV-L--KEL-NNL--S---E-------FQ------W-T--K--K-E--Q-DV..-------------V----A---P---------K----.KMA   450
DEB.CM.99.CM40         ASVGRQ--I--E---ELI-V-----L-VG--LR-DQ--KK-A---QF-LER-LR--PE-Y-E---FH------H----R-MPV---DE--..---HK------Q--------S---------A---.VPG   448
DEB.CM.99.CM5          A-VGRQ-KL-KE-Y--LI-V-----L-VG-NLD-VS--EQ-K---KF-LEK-LR--PE-Y-E----H------H----H-LP-E---E--..---HK------Q--------S--------KA---.VPG   448
RCM.GA.x.GAB1          AS------P--EE----I-------LFVG---T-L--DK-IK----H-LF---E--D----DK----------H--S-TV---K---K--..--------------------S-LR--E--K----.-KA   456
RCM.NG.x.NG411         -------EP--KE----V-------LFVG---TAH--SQ-IKR--EH-L----E--D----DK--FE----V-H-D--TV---K-L-K--..--------------------S-----E--K----.VKA   460
DRL.x.x.FAO            A--D----P--E-----I-------LFVG-----HL-EQ-IKE--TH-KF--LE--D----DN---E----Q-H----TV---R--DK-I..-------------V--------S-----E--K----.-K-   456
MND_1.GA.x.MNDGB1      G-VASL-EV--KNH-T-QL------LFVG--YTAE--EKAIVE--AL-M--NLE-----Y--E--FH------H-D--KIEKVQ---LA-Q.P---E------------A-L-P-----Q--K----.GLN   452
MND_2.CM.98.CM16       A--D----P--ERH---V-------LFVG---VAP--SR-I-E---H-LF--LE--D-----E--FE----I-H----TV-KVQ---K--..--------------------S-----E-------.-K-   454
MND_2.GA.x.M14         A--D---KT-KEEY-E-L-------LFVG--LSAS--NK--NR--EH-RF--LE--D-----E--FE----V-H-Q--TV-K-Q---K-T..--------------------S-----E--K----.-K-   455
MND_2.x.x.5440         A--D---KA-KE-H-E-L-------LFVG--LSAS--SK-IVE--EH-QF--LE--D-----E--FE----V-H----TV-K-Q---K-S..--------------------S-----E--K----.---   455
GRV.ET.x.GRI_677       N--AN--E-IKRHT-GLE-V-----LWL---HDETR-NQQ-DIV-KM-LEK-LE--D--V-RE--WE----K-H-N--TINK-E--PL-G.E----K---V--V---------P------T-AML--.KKN   465
TAN.UG.x.TAN1          H--QRV-K-IGK-VQ-LE-I-------W-G--L-EK--DEK-EEV-KL-LLR--E--D--V--R---D----K---SH-S-NK-EI--Q-H..------------V------L-P-----A--K-L--.-QD   462
VER.DE.x.AGM3          N--AS--E-IKKELKPLT-V-----LWVG-QEDEYT-DRL-E---MK-SA--LE--D--V--K---E----K-W-H---ISS-E-EDK--..-------R---------A-L-P-LR--N--K----.KKN   459
VER.KE.x.9063          N--A---EKIKEELKPLT-V-----LWVG-QEDEYT-DKL-E---VK-K---LE--D--V--K---E----K-W-Q----NS-N-E-K-Q..-----------------A-L-P-LR--QI------.KKN   459
VER.KE.x.AGM155        N--S---E-IKKELKQLT-V-----LWVG-QEEGPK-DQL--T--NR-QE--LE-----V-RE--FE----K-W-H--K--S-E-EKK-Q..-----L-----------A-L-P-LR--NI-K-L--.KKN   462
VER.KE.x.TYO1          N--AS--E-IKRNL-ALT-V-----LWVG-QENEHT-DKL-E---TK-QA--LE-----M--E---E----K-W-H--E-SR-Q-E-KD-..-----------------A-L-P-L--R.I-K--T-GKKN   462
BKM.CD.02.AY518534     N--AQ--KKV-E-YQ--T-I----EL--G---TND--YKI-TEI-QD-MGY-IA--QE-Y-RGH-VE-L--T-H-TT-TINP-E---DT-..M---EL---I--------V-Q-------AKCL--.TGS   190
TAL.CM.00.266          H-SDQL-RK--D-Y-Q-ALF-----L-VG-NESLE--RRI-GE--AL-NKV-IQ---A-Y-DK--IK-L------D--KV-P-E--DKLT..----EL---I------C---S-----A---YL--.VKG   451
TAL.CM.01.8023         Y-SDQL-RK--E-Y-Q-ALF-----L--G--ESLEQ-RRI-GE--AL-NKV-IQ---A---DK--IK-L------E--KV-N-E---KQT..-----L--------------S-----A--KCL--.VKG   452
TAL.US.x.1511          ..................................................................................................................................     0
TAL.US.x.1867          N--DHL-KK--E-Y-Q-ALF-----L-LG-NESLEA-RRI-GE--GL-NKV-IQ---A---EK--IK-L------D--KV-A-E--DK-Q..-----L------------L-S-----A--KCL--.VKG   363
TAL.US.x.416           ..................................................................................................................................     0
TAL.US.x.742           ..................................................................................................................................     0
MON.CM.99.L1           H--AELP--I-K-Y-E-TLI-----L--G-NHSLP--RKI-GVI-AT-LRK-IQ--PE---D-Y-VQ-L----H-SH-KIAK-E--DQ-Q..----E----I------A---S-----N--KS---.--S   450
MON.NG.x.NG1           YS-SELXKXI-R-Y-E-TLI-----L--G--HSLP--RKI-TAI-TX-SEK-IQ--IE-Y-DKY-LQ-L----H-TH-TV-K-E---QDQ..----G-----------A---P--------XC-X-.XKA   453
MUS_1.CM.01.1085       F--A-L-E-IKKA----TLI-----L--G-N-ELSS-RKV-A-I-NM-LKL-IQ--QE---E-Y-VQ-L----H-RG-KI-PVA--QQ-V..-----V----------A-A-P-----Q--KC---.VKS   443
LST.CD.88.447          A-VAGL-S---KL---IIL------LF-G---D-KG-DQA-KE---L-V--NL--------AE---H----V-H-DR-EIEKTK--DIDLQKT---N------V----A-L-E--R--E--K----.VKA   460
LST.CD.88.485          A-VAGL-S---KL---IIL------L--G---D-KG-DQA-KE--EL-M--NLE-------AE---H----V-H-DR-EIEKVK--DIDLQKT---N------V----A-L-E--R--E--K----.VKA   460
LST.CD.88.524          E-VSGL-A---KL---ILL------LF-G---E-EG-SQA-KE--EL-I--NLE-------DK---H----V-H-DR-EIERVK---IDTRKT---E------V----A-L-D--R--E--K----.VK-   461
LST.KE.x.lho7          G-VAGL-S---KL---MI-------LF-G---ERKG-DQA-KE--EL-M--NLE-------AE---H----V-H-DR-EIEK-K---MDLTKT---Q------V----A-L-D--R--E--K----.VK-   460
SUN.GA.98.L14          G-VASL-EK--RQY-E-QL------L--G--YE-KK-EEI-K---QL-ME-NLE-----Y-GE---K----I-H-DR-EIEK-K--PL--E.P-------I--VI-----L-E-LR--E--K----.NK-   461
SYK.KE.x.KE51          A-V-Q---PI-E--S-II-V-----L-VG-N-SEK--GQI-SEIVKS-LAV--SI-PE-W-DKF-LQ-L--T-H-D--S--KVQ---IT-S.P---EL--II-V--------P---C-A-SKC---.-KA   481
SYK.KE.x.SYK173        S-I-Q---P--K-YS-LTLI-----L--GT--SEKA-QEI---IVTA-LKV---V-KE-W-DQY-MQ-L--T-H-D-----K-E--NIDDE.I---QL---I-V--------S-----E--KC---.TKQ   475
COL.CM.x.CGU1          G-VAR--EP--KRGQL.Q-A-----LYLG--L-LE--RRA--E---A-LYY-LE-----Y-A----K------H--Q-K--KVKI--QDK..-----------A---L--L-P--R--EISKCTK-.KKH   423
```

```
SAB.SN.x.SAB1C          LTEIVQWTEEAELELEENRQILRQKQQGQYYDPALPLRAKVLKLGDGQWGYQIYQPE.NKILKVGKYAKIKTAHTNELRMLAGLVQKIGKESIVIWGQI.PIMELPVERELWEQWWSDYWQVTWIPEWEM   602
H1B.FR.83.HXB2          ---VIPL--------A---E--KEPVH-V----SKD-I-EIQ-Q-Q---T-----EP.F-N--T----RMRG----DVKQ-TEA----TT-------KT.-KFK--IQK-T--T--TE---A-------F   570
H1B.US.90.WEAU160       ----IPI--------A---E--KEPVH-V----SKD-I-ELQ-Q-Q---T-----EP.F-N--T----RVRG----DVKQ-TEA----TT-------KT.-KFK--IQK-T--T--TE---A-------F   570
H1N.CM.95.YBF30         ---V-NF--------A---E--KEPLH-V----GKE-V-EIQ-Q-Q---T-----EL.H-N--T-----MRS----DIKQ-VEV-R-VAT-------KT.-KFR---QK-V--A--T-H--A-------F   577
H1O.BE.87.ANT70         ---V-PLSR----------ER-K-PVH-V--Q-DKD-WVNIQ-Q-GE--T-----E-.H-N--T---TRQ-AS---DI-Q--EVI--VSQ---I---KL.-KFK---T--T--T--A----A-------F   566
H1O.CM.91.MVP5180       ---V-PLSK----------EK-KEPVH-V--Q-DKD-WVSIQ-H-E---T--V--D-.H-N--T----RQ-AS---DI-Q--EV---VSQ-A-----KL.-KFR---T--T--T--AE---A-------F   566
CPZ.CD.90.ANT           --DR--M-R---------K---Q--IE-Y--Q-G---K-TIQ-Q-S---T-----N-.G-L--A----RPTGT----V-Q---V-----L-------EV.-KFQ--IT--T-DA-------A-------F   570
CPZ.CM.01.SIVcpzCAM13   --DV-PL-P------A---E-VSTPVH-V--N-DKE-I-E-Q-Q-E---T---F-EQ.HRN--T----RQRST---DI-Q--EVI---AV-------KT.-RFR---QK-S--T--AE---A-------F   569
CPZ.CM.05.SIVcpzEK505   ---V-TL-R------A---E--KEPVH-A--N-DKE-I-EIQ-Q-Q---T-----DL.H-N--T-----MRST---DI-Q-TEV---VAL-------KT.-KFR---QK-V--T--TE---A----D--F   576
CPZ.CM.05.SIVcpzLB7     S--V-TF-R------A---E--KNPVH-V-----KD-I-E-Q-Q-QS--T-----EQ.F-N--T----RQGS----DVKQ--EVM---AI-------K-.-KFR--IQK-A--A--T----A-------Y   573
CPZ.CM.05.SIVcpzMB66    ---V-TF-T------A---E--KEPVH-A----SKE-I-EIQ-Q-Q---T---F-EQ.Y-N--T----RMRS----DVKQ-TEV---VAL-------KV.-RFR--IQK-T--A--T----A-------Y   573
CPZ.CM.05.SIVcpzMT145   ---V-AL-P------A--QE--KEPVH-A----EKE-I-EIQ-Q-Q---T---F-EP.H-N--T---G-QRA----DI-Q--ET----AT-------KT.-KFR---QK-M--A--A----A-------F   569
CPZ.CM.98.CAM3          --D--TM-Q---M-----XE--XDPVH-V----EKE-I-EXQ-Q-NS--T---F-EQ.H-N--T----RQRS----DI-Q--EV----AT-------KT.-KFR---QK-V--T---E---A----D-XF   575
CPZ.CM.98.CAM5          ---V-AM-Q---M------E--KDPVH-V----EKE-I-E-Q-Q-NS--T---F-EQ.H-N--T----RQRS----DI-Q--EV----AT-------KT.-KFR---QK-V--T---E---A----D--F   571
CPZ.GA.88.GAB1          --DV-PL-P------A---E-VSTPVH-V----DKE-I-EIQ-Q-NC--T---F-EP.H-N--T----RQRS----DI-Q--EA----AT-------KT.-KFR---QK-S--A--AE---A-------F   568
CPZ.GA.88.GAB2          -----EM-R----------E--KEPVH-T----EKE-I-E-Q-QNN---T---F-ET.GRN--T----RQRA-P--DI-Q--EV-P-VSP---IL--TV.-KFK---Q--V--A--T----A-------F   573
CPZ.TZ.01.TAN1          ---P-EM-R---------K---KE-V--A----K---Q-AIQ-Q-Q---T-----E-.G-N--T-----SPGT----I-Q----I----N---I---IV.-KFL---SK-T-S---T-------V----F   566
CPZ.US.85.CPZUS         ---V-NF-H---M------E--KEPVH-V----EKE-V-E-Q-Q-RS--T---F-ER.H-N--T----RQRS----DI-Q-VEV----AT-------KV.-KFK-L-QK-V--T---E---A----D--F   572
GOR.CM.06.CP684         ---V-TLSK----------E--KEPVH-V--Q-EKE-IVN-Q-QEE---T---F-D-.H-N--T----RQRAT---DI-Q--EVI--VSQ-------KL.-KFR--IC-ST-DAF-T----AS------F   186
GSN.CM.99.CN166         --DQ-EL----QI--A---E--K-SSE-G---AEK--VVEITS--EQ----MFS-EG..-M--SA-F--TRN--I-SYQQF-DALT-V-R--Q-T--KVPDKFRI--VK-Q-DA--MN---A----TI-A   581
GSN.CM.99.CN71          ---Q-EL----QI--A---E--K-SSE-S---AEK--VVEITS--EQ----MFS-DS..-M-RS--F--TR---I-SYQQ--DAMT-V-R--Q-T--KVPDKFRI--VK-Q-DT--MN---A----TI-A   574
H2A.DE.x.BEN            ---E-----L--A-----KI--S-E-E-Y--QEEKE-E-TIQ-SQGH--T-K-H-E-..------------NT---GV-L--QV-------AL----R-.-KFH------T-----DN----------DF   594
H2A.GW.x.ALI            ---G-----L--A------I--S-E-E-H--QEEKE-E-T-Q-DQ-N--T-K-H-G-..-----E----M-NT---GV-L--QV-------AL----R-.-RFH------T-----D---------D-DF   594
H2A.SN.x.ST             ---E-----L--A-----KI--S-E-E-C--QEEKE-E-T-Q-DQ-N--T-K-H-GG..----------V-NT---GV-L--QV-------AL----R-.-KFH-----DT-----DN---------D-DF   594
H2B.CI.x.EHO            ---E-----L--A-FQ--KI--E-E-E-S--KEGV--E-T-Q-NLAN--T-K-H-GD..----------V-NT---GV-L--HV-------AL----E-.-MFH------T-D---T-----------DF   594
H2B.GH.86.D205          ---E-----L--A--Q--KI--E-E-E-S--KERV--E-T-Q-NLAN--T-K-H-GN..-V--------V-NT---GV-L--HV-------AL----E-.-VFH------T-D---T-----------DF   595
H2G.CI.x.ABT96          ---E-----L--A--Q--KI--X-X-E-A--KEXE-XE-T-Q-NL-N--T-KVH-GN..R------------NT---GV-L--HV--X----AL----KL.-FFH-----DT-----T----------D-DF   577
H2U.FR.96.12034         ---E-----M--A-FA--KI--S-E-E-A--REEEN-E-T---NQ-N--T-K-H-GD..RT-----F--V-NT---GV-L--HV-------A-----K-.-MFH------V-----T-------VLD-DF   591
MAC.US.x.239            ---E-----M--A-Y---KI--S-E-E-C--QEGK--E-T-I-SQ-N--S-K-H-ED..-------F----NT---GV-L--HVI------A------V.-KFH----KDV-----T-----------DF   581
MAC.US.x.251_1A11       ---E-----M--A-Y---KI--S-E-E-C--QEGK--E-T-I-NQ-N--S-K-H-E..D-------F----NT---GV-L--HVI------A------V.-KFH----KDV-----T-----------NF   581
MAC.US.x.251_BK28       ---E-----M--A-Y---KI--S-E-E-C--QEGK--E-T-I-SQ-N--S-K-H-ED..-------F----NT---GV-L--HVI------A------V.-KFH-----DV-----T-----------DF   577
MAC.US.x.EMBL_3         ---A-----M--A-Y---NI--S-E-E-C--QEGK--E-T-I-SQ-N--T-K-H-ED..-----R-F----NT---GV-L--HVI------A---V--V.-KFH-----DV-----T-----------DF   577
SMM.SL.92.SL92B         ---E-----L--A--A---I--N-E-E-R--REDE--E-T---NQ-N--S-K-H-GD..R------F----NT---GI-L--NV--------L----KT.-FFH------V-D---T----A------DF   574
SMM.US.x.H9             ---E-----M--A-Y---KI-PS-EPE-C--QEGK--E-P-I-SQ-N--S-K-H-ED..-------F----NT---GVSL--HV-------A-I----V.-RFH----X-I-----T-----------DF   577
SMM.US.x.PGM53          ---E-----M--A-Y---KI--S-E-E-C--QEGK--E-T-I-SQ-N--S-K-H-ED..-------F----NT---GV-L--HV-E-----A------V.-KFH------I-----T-----------DF   577
SMM.US.x.SIVsmH635FC    ---E-----M--A-Y---KI--S-E-E-C--QEGK-IE-T-I-SQ-N--S-K-H-ED..-V-----F--V-NT---GV-L--HV-------AL----EV.-KFH------I-----T---------D-DF   577
SMM.US.x.SIVsmH635F_L3  ---E-----M--A-Y---KI--S-E-E-C--QEGK-IE-T-I-SQ-N--S-K-H-ED..-V-----F--V-NT---GV-L--HV-------AL----EV.-KFH------I-----T---------D-DF   577
SMM.US.x.SIVsmH635SB10  ---E-----M--A-Y---KI--S-E-E-C--QEGK-IE-T-I-SQ-N--S-K-H-ED..-V-----F--V-NT---GV-L--HV-------AL----EV.-KFH------I-----T---------D-DF   577
MNE.US.x.MNE027         ---E-----M--A-Y---KI--S-E-E-C--QEGK--E-T-I-NQ-N--S-K-H-ED..-------F----NT---GV-L--HVI------A------V.-KFH----KDV-----T-----------DF   577
STM.US.x.STM            ---E-----M--A-YA--KI--S-E-E----REDK--E-T-V-DQ-N--S-K-H-ED..-------F----NT---GV-L--HVI------A-------.-KFH------V-----A-----------DF   577
DEB.CM.99.CM40          --DP-TLS---QA--R--KE--K-EVS-V--KEQE--I-ELS---K-----V-R--K..G---T--FSRD-G--Y-DFHQ--KAMY---T----F--R-.-QFR---VK-E-DN--HNH--AA---D--A   575
DEB.CM.99.CM5           --DK-SLS---QA--R--KETPK-EVS-V--KEDE--ISEIS-I-R-----V-R-TK..GL--T--FS-D-GT-Y-DFHQ--KAMY---T----Y--KV.-KFR---VK-E-DN--HNH--A----D--A   575
RCM.GA.x.GAB1           -D-K-EM-K---I-Y---KM--KE-LH-V---EKK--V-NIQ--EG---S---E-ES.G-P--T-----Q-------I----------A--A-----RL.-TFR--I---V-D*-R-Q---------D--F   584
RCM.NG.x.NG411          -DDK-EF-R-----Y---KAV-KE-IH-V--N-EK--I---Q--EG---S---E-GD.S-P--T-----Q-------I----------C--A-----KL.-KF---I---V------------S------F   588
DRL.x.x.FAO             -D-E-E--R-----Y---KL-IKEQMH-V--Q-EK--K--IQ---N---S---E-DD.--P--T-----V-NS---DM-----------A--AL----RL.--FY------V-----QE---------D--F   584
MND_1.GA.x.MNDGB1       I--K-TM----R--Y-Q-KE--AEE-E-S----NKE-YVRFQ-TTG-DISF-WK-GN..-V-RA---G-Q----S-D-MK---AT--V-R-------FV.-K-QI-TT--I--D--HE---C-----V-F   579
MND_2.CM.98.CM16        -D-K-E--R-----Y---KL-VQEEVH-V--Q-EK--M---Q--TQ---S---E-ED.--P-------RT-N-------V--------A--AL----KL.-KFY--I---V-D---PE---A-------F   582
MND_2.GA.x.M14          -D-V-E--R-----Y---KL-VQEEVH-V--Q-GK--M---Q--TQ---S---E-E-.--P-------RT-H-------V--------A--C-----EL.-KFY--L---V-D---H----A-------F   583
MND_2.x.x.5440          -DK--E--R-----Y---KL-VQEEVH-V--Q-EK--M---Q--TQ---S---E-ED.--P-------RT-N-------V-------VA--CL----KL.-KFY--L---V-D---H-----------DF   583
GRV.ET.x.GRI_677        -L-EIV------A-YKN-QG-VQET-E-T----LKE-I-T-Q-Q-E---T--FT-EG..AV----R---QRET---D--T--H-----C--ALT---RL.-RVQ---DKKT-DM--Q-----S------F   592
TAN.UG.x.TAN1           --AV-----Q--V-FRQ-QE--KEE---A----SKS---TIT--D-------FK-EG..-V--T---T-S-NT----F-V---M--RVC--A-TF--KL.----V-----V-------------L--I-F   589
VER.DE.x.AGM3           -L-T-T------A-YA--KE--KTE-E-T--K-GR-I--A-Q--EG---S--FK-E..GQV------T-Q-NT----F-V-------LC---L----EL.-VL---I---V-----A-----S---D--F   586
VER.KE.x.9063           -L-E--------A-YA--KE--KTE-E-T--A-GK----A-Q--E----S--FK-E..G---------Q-NT------T--------C--AL----EL.--F-------V-----A-----S------F   586
VER.KE.x.AGM155         -LDV-E--P---A-Y---KE--KTE-E-T--A-EK----A-Q-------S--FK-E..G-------F--Q-AT------V---V-------AL-----L.-TF------DT-----A-----S-----DF   589
VER.KE.x.TYO1           -L-L-A--P---A-YA--AE--KTE-E-T--K-GI-I--A-Q--EG---S--FK-E..GQV------T-Q-NT------T--------C--AL----IL.-VL---I---V-----A-----S-----DF   589
BKM.CD.02.AY518534      -V-M-TLSK---A--AI-KEV-AK-AE-T---NSK--I-EITN--KQ----IVR-E-KGIP--T--F--VRSV----YDQ-VQAMA---R-A-TV--RL.-KFRI--VK-Q-DA--HEH--ES---DI-A   319
TAL.CM.00.266           -L-E--LS----A--A---E--AEETH-T--QEGE--E-ELT--AE-----M-K-GK..RL--T--F--QR---S-PYQQ-I-AM----------F---V.-VFRI--VK-E-----TEH--CS---TIVP   578
TAL.CM.01.8023          -L-E-TL-----A--A---A--AEETH-S--KEGE--E-ELT--AE-----MVK-GK..RL--T--F--QR---S-PYQQ-V-AM------A--F---V.--FRV--VK-E------E---CS-L-TIVP   579
TAL.US.x.1511           ..............................................................................................................................     0
TAL.US.x.1867           -LDE-TLS----A--A------Q-EVH-S---KEK--E-ELT--AE-----M-K-G..H-V--T--F-RQR---S-PYQQ-T-AM----R--L-Y---L.-TFRI--VK-E--A---SH--CS-V-TI-P   490
TAL.US.x.416            ..............................................................................................................................     0
TAL.US.x.742            ..............................................................................................................................     0
MON.CM.99.L1            --DT-VLSDL-QA--A---E--KEEST-G----NE--KVELTS-AEN----RFF-EK..SV--T--F---RST-S-TYQQ--DALA-L---A--T--RL.--FR---VK-Q-DA--A-N--IN-V-DI-A   577
MON.NG.x.NG1            --D--ELSDL-QA--A---E--S-QST-G----Q-D-KVEITS-XE-----XYF-E..XRV--T--F----N--S-TYQQ--DA-T-L---AL----KL.--FR---NK-Q-DT--A-N--------I-P   580
MUS_1.CM.01.1085        ---E--L----QA--A--QE--Q-TV--A----EQ--V-EIVS---S----NFT-S..KGV-RT--F--VRSV---SYQQ--DAMA---R--L----KPPDKVRI--IK-Q-D---IAH---S---TI-A   571
LST.CD.88.447           -D-V-T--DA-LE-Y-Q-Q---KE-L--A----TKE-VVR-QQSKK-IITF-WR-GN..S--RA-R-QRQ-S----P-QK--EA-------------HV.-KIQV--T--V-D-----H--------L-F   587
LST.CD.88.485           -E-V-T--DA-LE-Y-Q-Q---KE-L--A----TKE-VVR-QQNKK-VITF-WR-GN..S--RA-R-QRQ-A----P-QK--EA-------------HV.-KIQV--T--V-D-----H--------L-F   587
LST.CD.88.524           -EQL-T-----LE-Y-T-KT--KE-V--A----SKE-VVR-QQNKK-VITF-WR-GN..A--RA-R-QRQ-S--M-P-QQ--EA-------------FV.-KIQV--T--T--N--INF-------DL-F   588
LST.KE.x.lho7           -E--IN-----LE-YGQ-KEV-KE-M--A----EKE-IVR-QQNKK-IITF-WR-GN..N--RA-R-QRQ-A----P-QK-VEAI------------FV.-KIQV--T--V--H----H-------DL-F   587
SUN.GA.98.L14           -A-K-KM----RE-YQS-QEV-QESVS-S--E-DKE-ICR-Q-VKQ-ILTF-WL-G..KQV-R--R-Q-RGA--E-PCQQ--AAL----R-------FV.-KIQV-IQ--I-S---A----C-----L-F   588
SYK.KE.x.KE51           --D--EL-----A-MA---E--KVE-T-S--Q-EK--E-HIS---QQ----I-K-GPQE-P-IT--TG-AYAT-H-DYQA--Q-MN---IQALW---K-.-EFH---K--E--K--T-H--A--L--VKC   610
SYK.KE.x.SYK173         ---VLTL--A--A---------KEE-A-S----KK--E-HIT---SQ----M-K-EQKGPP-IT--T--TFA--S-DYQS--Q-LN---IQ-LWY--KV.-TFH---K--E--K--T----A--V--VKF   604
COL.CM.x.CGU1           -L-E--LSA---E--QD-K--IKEEV--V--I-HEDIWVDIFR--K-----AVC-K..HGT-RR--HNTG-NQ-F-SMQE--SVI----R-A--T--KV.-K-KV-AK--D----------SA----I-F   550
```

```
                              p51 RT \/ Pol p15 RNAse H
SAB.SN.x.SAB1C       VSTPQLIRLWYKLVKDPIPGEAVYYVDGAANRNSKEGKAGYLTDRGDQKVVALENTTNQKAELEAILLALRDSGSKVNIITDSQYAMGIIAGEPTESDNNIVQQIIEELIKKEAVYIAWVPAHKGVGGNE  732
H1B.FR.83.HXB2       -N--P-VK---Q-E-E--V-AETF--------ET-L-----V-N--R----T-TD-----T--Q--Y---Q---LE---V------L---QAQ-DQ-ESEL-N----Q-----K--L--------I----  700
H1B.US.90.WEAU160    -N--P-VK---Q-E-E--V-AETF-------KET-L-----V-N--R-R--S-TD-----T--Q--H---Q---LE---V------L---QAQ-DQ-ESEL-S----Q-----K--L--------I----  700
H1N.CM.95.YBF30      -N--P-VK---Q-ETE--S-AETF--------ET-L----FV----R----SIAD---------Q---M--QE--RD---V----------HSQ-DK-ESEL-S----------R--LS-------I----  707
H1O.BE.87.ANT70      ----P--K---R-ESE--M-AET---------ET-L-----V-EQ-K--IIK-DE---------M-------Q--KET---V------L-V-SSQ--Q-ESP---------T---Q--LT-------I----  696
H1O.CM.91.MVP5180    ----P--K---Q-ETE--V-AETF---------T-L-----V-EQ-K-NIIK--E---------M-V-I--Q--KEQ---V-----VL---SSQ--Q--SP---------T---R--LT-------I----  696
CPZ.CD.90.ANT        -N--P------N-LA----EAETF---------QL-----V----RSR-KH-QK----Q---Q---M--E--TGP---V------L-VLQ-T-DQ-ESPL-EE--QK---R-QI-LS-------I----  700
CPZ.CM.01.SIVcpzCAM13 IN--P-VK---S-ET----TADT---------ET-K-----I--K-K--IIS--------Q---G-LY---Q--EQQ---V-----VL---QSQ-DH-ESEL-N----------DK--LS-------I----  699
CPZ.CM.05.SIVcpzEK505 -N--P-VK---Q-ETE--S-AET---------ET-L----FV----R---TSISE----Q---Q-V-M--Q-A-QE---V-----VL---HSQ-DK-ESEL-N----------RI-LS-------I----  706
CPZ.CM.05.SIVcpzLB7  -N--SHVKF--Q-EQE----AETF--------ET-L-----V----K--IIP-TE---------Q--Q---Q---PE---V------L--LQAQ-DQ--SEL-N----N-----K--LT-------I----  703
CPZ.CM.05.SIVcpzMB66 -N--P-VK---Q-EQ-----AETF--------ET-L-----V--K-R--IIS-TE---------Q--Q---Q--EVE---V------L---Q-Q-DT-ESE--N----------K--LS-------I----  703
CPZ.CM.05.SIVcpzMT145 IN--P-VK---Q-ETE--E-AET---------ET-R-----V-N--R--IHS--D---------Q-V-M--Q---QE--VV-----VL--LQSH-DH-ESD--N----------KI-LS-----R-I----  699
CPZ.CM.98.CAM3       -N--P-VK---Q-ETE----AETF--------ET-K-----V----R--I-N------------T-VY---Q-AEHI---V-----VL---HSQ-DQ-ESEL-NL---------KS-LS------------  705
CPZ.CM.98.CAM5       -N--P-VK---Q-EPE----AETF--------ET-K-----I----R--IIN------------T-VY---Q-AEHI---V-----VL---HSQ-DQ-ESEL-GS---------KS-LS-------I----  701
CPZ.GA.88.GAB1       IN--P-VK---S-ETE---TTDT---------ET-T-----V--K-K--IIS--------Q---K-L----Q--DQQ---V-----VL---QSQ-DH-ESEL-N----------KI-LS-------I----  698
CPZ.GA.88.GAB2       -R--P-VK---Q-E-E--Q-AET---------ETRR-----VINN-K---IS--E-----------LRI--Q---QE--VV-----VL---NSQ-DQ-ESE--N----------K--LT-------I----  703
CPZ.TZ.01.TAN1       IN--P------N-LS----EAETF--------D--K-R---V-N--RYRSKD-------Q---W-VD---K---AQ---V-----V--VLQ-L-DQ--SP--E---QK-TQ-T-I-L--------I----  696
CPZ.US.85.CPZUS      -N--P-VK-----ETEA-E-AETF------Q-ET-K-----V----R--IIT------------T-VY---K--ENT--VV-G---VL---HSQ-DQ-ESEL-N----------KS--S-------I----  702
GOR.CM.06.CP684      I---P--K---Q-E-E--T-AETF--------E--Q-----V----R--IRK--G-----------V----EE--KN---V-----VL----AS-DQ--SPL--K----MTT--K--L--------I----  316
GSN.CM.99.CN166      IH--H-L-Q--T--SE-LEDAVT---------T--L-----V-NT-KHRAIE--E----Q---H-V----KEGPP-M-LV-----V---LQSQ-EV-TSAL-E---Q--L----I-LS-------I---T  711
GSN.CM.99.CN71       -H--H-L-Q--T--SE-LENAVT---------T--L-----V-NT-KYRAIE--G----Q---Q-V----KEGPPRM-LV-----V---LQSQ-EV-TSPL-E---Q--LG-----LS-------I---T  704
H2A.DE.x.BEN         ----P-V--TFN--G-----AETF-T--SC--Q--------V----KD--KV--Q----Q----VFRM--A---P-----V-----V---V--Q----E-R--N-----M-------V--------I---Q  724
H2A.GW.x.ALI         ----P-V--AFN------L-AETF-T--PC--Q--------I----RD--KV--Q----Q-----FA--VT---P-A---V-----V---V--Q----E-R--N-----M-----I-V--------I---Q  724
H2A.SN.x.ST          I---P-V--VFN------L-AETF-T--SC-KQ-R------I----RD--RL--Q----Q-----FAM-VT---P-A---V-----V---V--Q----ESK--N-----M-----I-V--------I---Q  724
H2B.CI.x.EHO         ----P----A-N-----LE-VET--T--SC-KA--------V----KD--KP--Q----Q-----FA---Q---PQ----V-----V---V-AQ---TESP--RE----M-----KI-VG-------L---Q  724
H2B.GH.86.D205       ----P----A-N-----LE-RET--T--SC--T--------V----KD--KV--Q----Q-----FA---T--EPQ----V-----V-----AQ---TESP--AK----M-------VG-------L---Q  725
H2G.CI.x.ABT96       ----H-V--V-N---E-LEQTET--T--SC-KT--X-----V----KD--XV--Q----QX----FAM--Q---P-----X-----V-X---XQ---TESPL-NK----X-----XL-VG-------I---Q  707
H2U.FR.96.12034      ----P-V--A-N---E-LL.-ET--T--SC--A-Q------V----KDR-KP--Q----Q-----FAM--E---P-----V-----I------Q---TESPL-SK------R-----VG-------I---Q  720
MAC.US.x.239         I---P-V--VFN------E--ET--T--SC-KQ--------I----KD--KV--Q----Q-----F-M--T---P-A---V-----V-----T-C----ESRL-N-----M---SEI-V--------I---Q  711
MAC.US.x.251_1A11    I---P-V--VFN------E--ET--T--SC-KQ--------I----KD--KV--Q----Q-----F-M--T---P-A---V-----V-----T-C----ESRL-N---K-M---SEI-V--------I---Q  711
MAC.US.x.251_BK28    I---P-V--VFN------E--ET--T--SC-KQ--------I----KD--KV--Q----Q-----F-M--T---P-T---V-----V-----T-C----ESRL-N-----M---SEI-V--------I---Q  707
MAC.US.x.EMBL_3      I---P-V--VFN------EV-ET--T--SC-KQ--------I----KDI-KV-.T----Q------YHGIE---P-R---VEL-VCY-NNNRF----ESRL-N-----M--.VR--V------LE-I---Q  705
SMM.SL.92.SL92B      I---P----VFN------EK-E---I--SC-----------V----KE--LP--QA---Q--Q-L----K--P----VV-----VLN--T-Q-S---SD--A----Q-VQ------G--------I---N  704
SMM.US.x.H9          ----P-V--VFN---E--Q-AETF---RSC--Q-R------V----RG-TKL--Q----Q-----FY---A---P-A---V-----V---V--Q----ESRL-N-----KM-----I-V-------XI-E-Q  707
SMM.US.x.PGM53       ----P-V--VFN---E--Q-AETF----SC--Q-R------V----RD-AKP--Q----Q-----FY---A---P-A---V-----V---V--Q----ESRL-N-----M-----I-V--------I---Q  707
SMM.US.x.SIVsmH635FC ----P-V--VFN---E--Q-AETF----SC--Q-R------V----RD-AKL--Q----Q-----FY---A---P-A---V-----V---V--Q----ESRL-N-----M-----I-V--------I---Q  707
SMM.US.x.SIVsmH635F_L3 ----P-V--VFN---E--Q-AETF----SC--Q-R------V----RD-AKL--Q----Q-----FY---A---P-A---V-----V---V--Q----ESRL-N-----M-----I-V--------I---Q  707
SMM.US.x.SIVsmH635SB10 ----P-V--VFN---E--Q-AETF----SC--Q-R------V----RD-AKL--Q----Q-----FY---A---P-A---V-----V---V--Q----ESRL-N-----M-----I-V--------I---Q  707
MNE.US.x.MNE027      I---P-V--VFN------E--ET-----SC-KQ--------I----KD--RV--Q----Q-----F-M--T---P-A---V-----V-----T-C----ESRL-N-----M---TEI-V--------I---Q  707
STM.US.x.STM         ----P-V--VFN---E-LE-TETF----SC--Q--------I----KN--K---Q----Q-----FAM--A---P-A---VV----V-----T-Q----ESKL-N-----M-----I-V--------I---Q  707
DEB.CM.99.CM40       IH--H-VK---E--SE---DADT---------E--L-----V-EW-K-S-KC-----------------EEGP--M--V------L---LEH-S-TEHK--EKV-QA-QG--QI-L--------I---D  705
DEB.CM.99.CM5        -H----V----E--SE---DADT---------E--M-----ISEW-K-G-KS-----------------EEGP--M--V------L---LLEH---TEHE--EK--QA-QE--Q--LN-------I---D  705
RCM.GA.x.GAB1        ----P---G-N---------E------------I-----V-N--KE--KE--E---------V----K---P----V----VY--LEAQ-DT--SGL-TE--NQM-G-----LS-------I----  714
RCM.NG.x.NG411       ----P------N--------E------------L-----V----KE--KE----------G-V----K---KR---V---RHVF--L--S-DT-ESPL------QM-G--E--LS-------I----  718
DRL.x.x.FAO          I---H--K---N-LS-----E----------V--------V-S--KE--I-I-E---------Q------K---P----V----------SSS-EI---P--N----QM-S------LN-------I----  714
MND_1.GA.x.MNDGB1    I---M-E-E--S-SPE-LE-VET---------D--M-----I----F-R-EEYL-----QT--H-VK---E----Y---V-----VV--L-SR---T-HP--KE---LMKG--KI-LS-L-----I----  709
MND_2.CM.98.CM16     ----H--G---N-LRE-V---D-------------------V-A-NKSR-I-----------------KM--Q---P----V----------LSAA-DQ---P--RE---LM-H-----L--------I----  712
MND_2.GA.x.M14       I---P------N-L-E-----D-------------------Y-A--KS--I------------------K---PRA---V------L--LSAT-DR---P--RE---NQM-A-----L--------I----  713
MND_2.x.x.5440       I---P------N-L-------D-------------------Y-S--KS-----E--------T--KM--Q---PR---V------L--LSAA-DQ---P--RE---LM-G--EI-LS-------I----  713
GRV.ET.x.GRI_677     ----L-VK---S---E--K--D---------SKVT-L------SE--KSRIRE-------Q---T-VKM--E---EN---V-----V-N-LTAC-Q--NSPL-E---QA-M--RQ--LQ-------I---T  722
TAN.UG.x.TAN1        I---P--K---N-------KAK---C---C-----L----F--NT-ESQ-IE--D----Q--Q-V----KH-K-----V-----V---LTAQ--Q--SVL-N-----M-N-D-I-LT-------I---Q  719
VER.DE.x.AGM3        ----P-VK---T-T-E---K-D--------C----R------I-QY-K-R-EK-------Q---M--KM--E---PN---V----------LTAQ--Q--SPLIE---ALMVQ-HQI-LQ----D--I----  716
VER.KE.x.9063        --V-P-VK---T-T-E-----D--------C------------I-QY-K-R-EK--Q----Q---V--KM--E---E----V----------LTAQ--Q--SPL-E---ALMV--R-I-LQ-------I----  716
VER.KE.x.AGM155      --V-P-VT---T-T-E-----D--------C--Q--------I-QQ-K-R-QQ-------Q---T--KM--E---P----V----------LTAQ--Q--SPL-E---AQMVQ---I-LQ-------I----  719
VER.KE.x.TYO1        ----P-LK---T-T-E---K-D----.--C------------ISQY-K-R-ET-------Q-K-T--KM--E---PN---V----------LTAQ--Q--SPL-E---ALM-Q-QQI-LQ-------I----  718
BKM.CD.02.AY518534   -H--H-L-S-FT--SE---DA-T--------K---L-----V--T-YS---H--G-----------K---Q--------V-----VL---SAQ--S--HP--ED--QL-LG-DQ---T---G---I---V  449
TAL.CM.00.266        IH--P-V----N--QE--E-ADT-------H-E--------V-AT-KEH-IK--K-----------V----K---P----V-----VY--L--S---T--K-IED--QL-LS-----L--L-----I----  708
TAL.CM.01.8023       IH--P-V----N--SE--E-AET-------H-E--------V-TS-KEH--G-S-----------V----KH--P----V-----VY--LV-N---T--K--ED--QQ-LG------LT-L-----I----  709
TAL.US.x.1511        ........................................................................................................................--  2
TAL.US.x.1867        IH--P-V----S--QE--E-AET-------S-D--------V--S-KE--IG-SD--------Q-M----KE--P----V-----VY--LI-N--DTT-Q--EE--QQ-LS-----L--L-----I----  620
TAL.US.x.416         ......................................................................................................................I----  5
TAL.US.x.742         ...............................................................................................................QRYRS-----I----  15
MON.CM.99.L1         -YS-H-L-Q--T--Q---E-APT---------T--------V-NF-AR-A-T--Q----------V------GPPQM--------VL--L-SC-EI-ESP--EA--Q--L--DQ-FLS------I----  707
MON.NG.x.NG1         -HS-H-L-Q--T--QX--EHAPT---------A--------V-ST-QR--IP--Q----------V----QNGPASM--V-----V---L-SK-EV-ESP--EK--QQ-LG-KQ--LS------I----  710
MUS_1.CM.01.1085     -H-TH-LKQ-FT--PE-LKEAPT---------V--L-----V-NT-KE--IS--S----------V----KEGPPRM-LV-----VL--V-SQ-Q--TSPL-E---Q--LS---I-LS------I----  701
LST.CD.88.447        I-----EQE--S-EAE--E-VDT-------EKVG-T-----V-QS-KE--KE-TD----Q-----V--------K-R---V-----VIKVLSQR---TEHPV-K-----CK--DQ--LG-------I---Q  717
LST.CD.88.485        I-----EQE--S-EAE--E-VDT-------EKVG-T-----V-QS-KE--KE-TD----Q-----V----K--K-R---V-----V-KVLSQR---TEHPV-K-----CK--DQ--LG-------I---Q  717
LST.CD.88.524        I-----EQE--S-EAE--E-TDT-------E-VG-T-----I-QS-KE-IKE-QD----Q-----V----Q--K-T---V-----V-KVLSQR---TEHP--KN--K--Q--D---LG-------I---Q  718
LST.KE.x.lho7        I-----EQE--IWEAE--I-VDT-------EKVG-T-----I-QS-KE--KE-TD----Q-----V-M--Q--N-----V-----V-K-LSQR---TEHP--KD---QCKQ-DQ--LG-------I---Q  717
SUN.GA.98.L14        ----K-EQE--T-ATE-V--.DT-------EKLE-R-----I-QG-RSR-KK-------Q------KM--E--R-S---V------LRLLSKR---T-SEL-KE-V-LIRQ-DQ--LG-------I---Q  717
SYK.KE.x.KE51        -H--I-V-WY-N--SE-V.EAETF-----------Q---I----K-TLKR-----------V-M--Q---PE---V------L--LMNC--S-EHPV-E--MQKAME--KI--T-------I---Q  739
SYK.KE.x.SYK173      I---P-V-WY-N--PE---EAVTF--------D--T-N---VASD-T-R-QY--Q----Q----GL-M--Q--KD----VV----SY--LMTC--NTEHP--E---Q-A-----I-VT-------I----  734
COL.CM.x.CGU1        --SSYVVK-VWN--LE-LKEAPT-WT--GCS-KLGA----WINSK-EEE--N-HEGS--Q---TGV----KHGPKRM-LV---L--L---T-Q-YDQLPSSRGE--QAGMA---IHV--C-----I----  680
```

```
                              Pol p15 RNAse \/ Pol p31 Integrase
SAB.SN.x.SAB1C              EIDKLVSQG.IR..QVLFLDRIE.EAQEEHDK.YHANWRSMQQEFGLPAIVAKEIVAACPKCQIKGESVHGQVDASPGVWQMDCTHLEGKIIIVAVHVASGFIEAEVIPQETGKATAHFLLKLASRWPIT   857
H1B.FR.83.HXB2              QV-----A-.--..K-----G-D.K--D--E-.--S---A-ASD-N--PV-------S-D---L---AM-----C---I--L--------V-L--------Y-------A---QE--Y------G---VK   825
H1B.US.90.WEAU160           QV-----S-.--..K-----G-D.K-----E-.--S---A-ASD-N--PV-------S-D---L---AM-----C---I--L----------L--------Y-------A---QE--Y-I----G---VK   825
H1N.CM.95.YBF30             QV-----S-.--..KI----G--.K---D--R.--S--KA-ASD-N--P--------S-D---L---AM----NC-------L----------L--------YL------A---QE--Y-I----G---VK   832
H1O.BE.87.ANT70             K------KD.--..R----EG-D.Q---D-E-.--S--KALAS-----PV-----I-S----H----AI-----C--E---I----M---------------------A---QE--Y------A---VK   821
H1O.CM.91.MVP5180           K------KD.--..R----EG-D.Q---D-E-.--S---ALASD----P------I-S----H----AT-----Y--EI--------------------D--------A---QE--Y------A---VK   821
CPZ.CD.90.ANT               QV-------.--..-----EG-D.K---D---.--S----LAD-YN--P------I-Q-D--HV---AR-----C---I--V---------V--------S---------MAD----S--Y-----------VK   825
CPZ.CM.01.SIVcpzCAM13       QV-----S-.--..K-----G-D.K-----ER.--S--KALAG--N--P--------N-D---V---AL-----C---T------------------I---Y-------A---QE--Y-I---------VK   824
CPZ.CM.05.SIVcpzEK505       Q------T-.--..K-----G-D.K-----ER.--S--KA-ASD-N--P--------S-D---L---AI---INC-------L----------L--------YL------A---QE--Y-I----G---VK   831
CPZ.CM.05.SIVcpzLB7         QV-----S-.--..K-----G-D.R-----E-.--N---A-ASD-N--P--------S-D---L---AM-----C---I--L--------V-L--------YM------A---QE--Y-I---------VK   828
CPZ.CM.05.SIVcpzMB66        Q------S-.--..K-----G-D.K-----E-.--N---A-ASD-N--P--------N-D---L---AI-----C---I--L----------L--------YM------A---QE--Y-I----G---VK   828
CPZ.CM.05.SIVcpzMT145       QV-----S-.--..K-----G--.K-----ER.--S--KA-A-D-N--P--------Q-D---L---AM-----C---I--I-------------------Y-------A---QE--Y-V----G---VK   824
CPZ.CM.98.CAM3              QV-----S-X--..K-----G-DXKHK---EEX--N---A-ASD-NI-H-----L--R-D---L---AM-----C---I--L--------V-L--------YL------A---QE--Y-I----G---VK   833
CPZ.CM.98.CAM5              QV-----S-.--..K-----G-D.K-----E-.--N--KA-ASD-N--PV-------S-D---L---AM-----C---I--L----------L--------YL------A---QE--Y-I----G---VK   826
CPZ.GA.88.GAB1              QV-----A-.--..K-----G-D.R-----ER.--S--KA-ASD-N--P--------H-D---V---AM-----C---I--V--------V----------Y-------A---QE--Y------G---VK   823
CPZ.GA.88.GAB2              QV-----S-.--..K-----G-D.K---D-ER.--S---A-AD--Q--PV-------T-D---L---AI-----C---I--L----------L--------YL------A---QE--Y-I----G---VK   828
CPZ.TZ.01.TAN1              -V-----KN.--..KI----G-N.----D---.--S--KALAD-YN--PV-----I-Q----H----AI-----Y--EI--I--------V------------------E---RE--Y-I----G---VK   821
CPZ.US.85.CPZUS             Q------S-.--..K-----G-D.K---D---.--S--TA-ASD-N--PV-------S-D---P---AI-----C---I--L-------------------YL------A---QE--Y-I----G---VK   827
GOR.CM.06.CP684             -------KD.--..R----EG-D.Q---D-E-.--S---ALASD----PV-----INS----HV---AR-----C---I--L----------L-----------------A---QE--Y------A---VK   441
GSN.CM.99.CN166             -V-----K-.--..-----EN--.P-V-D-E-.--S--KYLRDQYKI--LL-----NK-S---VH--PK----N-EL-M-----------V-L--------YVW-RI------RQ--LK--E--AT--V-   836
GSN.CM.99.CN71              -V-----R-.--..----MEN--.P-V-D-E-.--S--KYLRDQYKI--LL-----NK-S----H--PK----N-EL-I-----------V-L--------YVW-RI------RQ--LK--E--AT--V-   829
H2A.DE.x.BEN                -V-H-----.--..-----EK--.P-----E-.--SIIKELTHK--I-LL--RQ--NS-AQ--Q---AI----N-EI------Y--------------------------S-RQ--L-----------   849
H2A.GW.x.ALI                -V-H-----.--..-----EK--.P-----E-.--S-VKELSHK---NL--RQ--NT-AQ--Q---AI----N-EL-T-----------V--I------------------S-RQ--L-----------   849
H2A.SN.x.ST                 -V-H-----.--..-----EK--.P-----E-.--S-VKELSHK---KL--RQ--NT-TQ--Q---AI----N-EL-T----------------------------------S-RQ--L-----------   849
H2B.CI.x.EHO                -V-H-----.--..-I---EK--.P-----E-.--N-VKELVHK--I-QL--RQ--NS-D---Q---AI----NSEL-T-----------V-----------------------RQ--L-----------   849
H2B.GH.86.D205              -V-H-----.--..-----EK--.P-----E-.--G-VKELVHK--I-QL---Q--NS-D---Q---AI----N-DL-T-----------------------------------RQ--L-----------   850
H2G.CI.x.ABT96              -V-X-----.--..-----XK--.P-----EX.F-S-VKELTHK--I-QL---Q--NSYHS--Q---AI----N-EL-T-----------V-----------V-----------RQ--L------G-----   832
H2U.FR.96.12034             -V-H-----.--..-----EK--.P-----E-.--S-VKEL-HK--I-QL---Q--NT-HV--Q---AM----NTEV-T-----------VV----------------------RQ--L------G-----   845
MAC.US.x.239                ---H-----.--..-----EK--.P-------.--S-VKELVFK----R---RQ--DT-D--HQ---AI---ANSDL-T-----------------------------------RQ--L------G-----   836
MAC.US.x.251_1A11           -V-H-----.--..-----EK--.P-------.--S-VKELVFK----R---RQ--DT-D--HQ---AI----NSDL-T-----------------------------------RQ--L------G-----   836
MAC.US.x.251_BK28           ---H-----.--..-----EK--.P-------.--S-VKELVFK----R---RQ--DT-D--HQ---AI----NSDL-T------------V----------------------RQ--L------G-----   832
MAC.US.x.EMBL_3             --GP-----.F-..-----EK--.P-------.--S-VKELVFK----R---RQ--DT-D--HQ---AI----NSDL-T------------V----------------------RQHY.-------G---YL   829
SMM.SL.92.SL92B             -V-R-----.--..-----ES--.P---D---.--S-VKELA-KYNI-QL---Q--N--N---Q---AI---TN-EV-T-----------V-----------------R---RQ--L----I------K   829
SMM.US.x.H9                 -V-H----E.--..-----EK--.P-----E-.--S-VKKLVFK-X--RL---Q--DT-D--HL---XI----NXXL-T-------------------G---------X---RQ--L-----------   832
SMM.US.x.PGM53              -V-H-----.--..-----EK--.P-----E-.--S-VKELVFK--I-RL---Q--DT-H--HQ---AI----N-EL-T-----------------------------------RQ--L-----------   832
SMM.US.x.SIVsmH635FC        -V-H-----.--..-----EK--.P-----E-.--S-VKELVFK--I-RL---Q--DT-DR-HQ---AI----N-EL-T-----------------------------------RQ--L------G-----   832
SMM.US.x.SIVsmH635F_L3      -V-H-----.--..-----EK--.P-----E-.--S-VKELVFK--I-RL---Q--DT-DR-HQ---AI----N-EL-T-----------------------------------RQ--L------G-----   832
SMM.US.x.SIVsmH635SB10      -V-H-----.--..-----EK--.P-----E-.--S-VKELVFK--I-RL---Q--DT-DR-HQ---AI----N-EL-T-----------------------------------RQ--L------G-----   832
MNE.US.x.MNE027             ---H-----.--..-----EK--.P-------.--S-VKELVFK---RL---Q--DT-D--HQ---AI----NSDL-T-----------------------------------RQ--L-----------   832
STM.US.x.STM                -V-H-----.--..-----EK--.P-----E-.--S-VKELVFK--I-RL---Q--DT-D--HQ---AI----N-EL-T-----------------------------------RQ--L----------V-   832
DEB.CM.99.CM40              QV-----K-.--..RI---E-V-.----D-E-.--G--KQLRD-YQ--TLI--Q-ITQ-----TR--PT-------M-I---------N-----------Y--TKIL-A----E--L---Q-GA---V-   830
DEB.CM.99.CM5               QV-----R-.--..RI---E-V-.----D-E-.--G--KQLRD--Q--TLI--Q-I-Q---------PM-------M----I------N-V-----------M-TK-LSA----E--L---Q-GA----K   830
RCM.GA.x.GAB1               -V-----K-.--..------G--.K-----E-.--N---ALAED-QI-Q--------Q-----V---AI---------T---------------------Y-------A----E---------A---VR   839
RCM.NG.x.NG411              QV-------.--..------G-G.K-----E-.--S---A-AE--QI-Q--------Q-----V---A----------T---------------------Y-------A----E---------A---VK   843
DRL.x.x.FAO                 -E-----R-.--..------NM-.K------L.--N----LA------G--------Q------H--PI---------T---------------------H-------A---EK--Y------G---VS   839
MND_1.GA.x.MNDGB1           Q------S-.--..K----QN--.P-----E-.--S-EAQLREK-H---L---Q--QS-S--CHH--PIK--T---L----I------NQ-------------MK----TA----K--E------AQ---S   834
MND_2.CM.98.CM16            QV-----R-.V-..-----EG-D.K-------.--N---ALA-D-CI-N--------Q-----T---PI-----T---T-------M---V--A------RYL------T----E---------G---VK   837
MND_2.GA.x.M14              Q------K-.--..-----EG-D.K-------.--N----LS---SI-P--------Q-----V---PI------A--T-----------V-----------YL------A----E---------G---VK   838
MND_2.x.x.5440              Q------K-.--..-----EG-D.K-------.--N---ALS-D-N--P--------Q-----T---P-----N-D--T----------------------YL------A----E---------G---VK   838
GRV.ET.x.GRI_677            -------K-.--..-I-------.---DD-A-.--N-----V------N----------------R--PK-------IET-----------V-------------------R----E--------LA---VK   847
TAN.UG.x.TAN1               ---------.--..-------M-.----S---.--T--QFIRDA--I--L---------------R--PI-------V---------------------------L-----AR----E--K----II------   844
VER.DE.x.AGM3               ---------.M-..KI---EK--.-------ER.--N---NLADTY---Q--------M---------P-------------------V------------------R----E--K----IL------   841
VER.KE.x.9063               ---------.--..R---IG---.-------R.--S--KNLADT----Q--------M--NAK----PI-----------------V----V-----------------R---RE--K----IL-----V   841
VER.KE.x.AGM155             -------K-.V-..RI--IG---.-------R.--S---NLADT----Q--------M-----V---PI-----------------I----V-----------------R----E--K----IIG-----   844
VER.KE.x.TYO1               -------K-.--..R----EK--.-----K-ER.--N--KNLADTY---Q--------M---------P----------T---------K-VV-----------------R----E--K----IL------   843
BKM.CD.02.AY518534          -V-R---KD.--..-I-L-EGVT.P---D-E-.--S--KYLREK-HI--LI----INQ--G--Q---PM----NTEI-T-------M---VV----NT---YM--Q-L-D-K--S--L---Q-CA-----   574
TAL.CM.00.266               QV-------.--..KI--VEQ-P.------ER.--N---DLKAR-K--T----A-IE------VQ--PKT--NN-AV-T----------Q--C--------Y--TKIL-R---RE--L---QV------S   833
TAL.CM.01.8023              QV-------.--..RI--VEQ-P.------ER.--N--KDLKAR-K--T----A-IE------VQ--PKT--TN-AV-T----------QV-C--------Y--TKIL-R---RE--L---QV------E   834
TAL.US.x.1511               QV-------.--..R---MEH-P.------ER.--N---DLKSR-K--T----A-IE------VH--PRT--NN-EV-T-------M--QV-C--------Y--TKIL-R---RE--L---QI------R   127
TAL.US.x.1867               QV-------.--..R---MEH-P.------ER.--N---DLKAR-K--T----A-IE------VH--PRT--NN-EV-T-------M--QV-C--------Y--TKIL-R---RE--L---QI------R   745
TAL.US.x.416                QV-------.--..R---MEH-P.------ER.--N---DLKSR-K--T----A-IE------VH--PKT--NN-EV-T-------M--QV-C--------Y--TKIL-R---RE--L---QV------R   130
TAL.US.x.742                QV-------.--..R---MEH-P.-------R.--N---DLKSR-K--T----A-IE------VH--PKT--NN-EV-T-------M--QV-C--------Y--TKIL-R---RE--L---QV------R   140
MON.CM.99.L1                -V-R--G--.--..----MENL-.P-R-D-E-.--S--KYLRDTYHI-TLL-----NH-H---NH--PKS---N-EV------------S--L--------YVW-KIL-R----C-GIA--E--AM--V-   832
MON.NG.x.NG1                -X----X--.--..----MEN--.X-R-D-E-.--S--KYLRDTY-I-TLL-----NH----HTQ--PK----XSDI-I-------Y--S--L--------YVW-KIL-R----E-GIA--E--AT---H   835
MUS_1.CM.01.1085            -V----NS-.--..----MEQ--.P-K-D-E-.--S--KYLRDKYQI--LL-----NH-----TH--PKT---N-EI----------------L--------YMW-KIL-R---RQ--LG--E---L--A-   826
LST.CD.88.447               -V-H---K-.--QK-----GK--.P----YS-.F-N-AKDLEEK-HI-PM---Q--NS-N---T---AIT----T---I--I----M--QV--N--------MV-----D----T--------C----VQ   844
LST.CD.88.485               -V-H---K-.--QI-----EK--.P-----S-.F-N-AKDLEEK-H--PM---Q--NS-N---T---AIT----T---I--I----M--QV--N-I---C--MV-----D----T--N-----C----VQ   844
LST.CD.88.524               ---Q--IK-.--TK--M--EK--.P-V---G-.F-N-AKDLEEKY-I-PM---Q--NE-A---Q---AIT----V-V-I--L-------Q--MN---M----MV-----D----T--N-----C----VK   845
LST.KE.x.lho7               -V-H---K-.--QK--M--EK--.P-V---S-.F-N-AKDLEEK-N--PM---Q--ND-AN--K---AIT----V-V-I--L-------QV--N--------MV-----D----T-S------C----VK   844
SUN.GA.98.L14               ---Q-----.--KR--M-IEK--.P-V---G-.F-N-AA-L-EM-DI-LV---Q--NE-AQ--Q---AIT------V-I--I----M-E-V----------YMV---L-N-Q--T--TW----CAM--VK   844
SYK.KE.x.KE51               -V-Q---K-.--..KI----N-P.K-----E-.--T-VEYIR---H--RQ---A--EM------R--PK-----TDL-T-----------V-C---NT---YT-TKILKR---QE-GL---QI-A----K   864
SYK.KE.x.SYK173             AV-----K-.--..KI---E--P.Q---D-ER.--S-MEYLR---H--RQ---A-IQQ-----NR--PK-----VDIYN-------E---V-C---NT---Y--TKILKR---DE--L--MQI------K   859
COL.CM.x.CGU1               L--QK-..-.V-..--MWI-K--.A-E-D-Q-.F-S-VQYLKEQ----TV-----WER-SE--N--QA----L-Y-Y-L--L----E---V-L-----CTL-CW-TILKR---EE-GRA-I----Q-EVR   803
```

```
SAB.SN.x.SAB1C        QL.HTDNGTNFTSQQVAAICWWGKIEHTFGVPYNPQSQGVVESMNKQLKEIIGQIRDDAERLETAVIMAVHIHNFKRKGGIGGYSAAER.LINIIHTELETKTLQQKI..SKIQNFRVYYREGRDPVWKG   983
H1B.FR.83.HXB2        TI.-----S---GAT-R-A---AG-KQE--I---------------E--K----V--Q--H-K---Q---F---------------G--.IVD--A-DIQ--E--KQ-..T----------DS-N-L---   951
H1B.US.90.WEAU160     TI.-----S----TT-K-A---AG-KQE--I---------I-----E--K----V--Q--H-K---Q---F---------------G--.I-D--A-DIQ-QQ--KQ-..T----------DS---L---   951
H1N.CM.95.YBF30       VI.-----S----AT-K-A---AN-KQE--I--------A------E--K-------Q--H-K---Q---F-------------T-G--.I-D--A-DIQ-TN--TQ-..L-V--------DS---I---   958
H1O.BE.87.ANT70       VI.-----P----TTMK-A---AN-Q-E--I-----------A---E--S--Q-V--Q--H-R---Q---FV------------T-G--.I-D-LASQIQ-TE--KQ-..L-XHK------DS---I---   947
H1O.CM.91.MVP5180     VI.-----P----AAMK-A---TG-Q-E--I-----------A---E--S--Q-V--Q--H-K---Q---FV------------T-G--.--D-LASQIQ-TE--KQ-..L--N-------DS---I---   947
CPZ.CD.90.ANT         TI.-----A----AA-K-A---AN-QQE--I------------------Q-------Q--Q-K---V-----------------TPGQ-.ILD-LA-DIQ-TQ--NQ-..L---Q---H--DS-------   951
CPZ.CM.01.SIVcpzCAM13 V-.----------NA-K-A---AN-KQK--I---------------E--K--E-V--Q--H-K---Q---FS------------T-G--.I-D--A-DIQ--E--KQ-..L---K------DS---I---   950
CPZ.CM.05.SIVcpzEK505 VI.-----S----AT-K-A---AN-QQE--I--------A------E--K-------Q--H-K---Q---F-------------T-G--.I-D--A-DIQ-TK--TQ-..L-V--------DS-E-T---   957
CPZ.CM.05.SIVcpzLB7   II.-----S----TA-K-A---AQ-QQE--I---------------E--K----V--Q--H-K---Q---F---------------G--.I-D--A-DIQ-TE--KQ-..--V-K------DR---I---   954
CPZ.CM.05.SIVcpzMB66  VI.-----S----ST-K-A---AG-QQE--I---------------E--K-------Q--H-K---Q---F---------------G--.I-D-LA-DIQ-TK---Q-..-N--K------DS---I---   954
CPZ.CM.05.SIVcpzMT145 TI.-----P----NA-K-A---AG-RQE--I---------------E--K--E-V--Q--H-K---Q---F---------------G--.I-D--A-DIQ--E--KQ-..L---Q------DSK--I---   950
CPZ.CM.98.CAM3        VI.-----P----NT-K-A---AG-QQE--I---------------E--K------EQ--H-R---Q---F-------------T-G--.I-D--A-DIQ-TN--KQ-..L-V--------DS---I---   959
CPZ.CM.98.CAM5        VI.-----P----NT-K-A---AG-QQE--I---------------E--K-------Q--Q-K---Q---F-------------T-G--.I-D--A-DIQ-TN--KQ-..L-V--------DS---I---   952
CPZ.GA.88.GAB1        TI.-----P----AA-K-A---AD-KQE--I------------L--E--K----V--Q--H-K---Q---F-------------T-G--.I-D--A-DIQ-SE--KQ-..L-V-K------DS---I---   949
CPZ.GA.88.GAB2        KI.----------AA-K-A---AG-KQE--I---------------E--K----V--Q--H-K---Q---F-------------T-G--.IVDMLA-D-Q-QE--KQ-..L---K------DS---I---   954
CPZ.TZ.01.TAN1        KI.-----P----TA-K-A---AQ-Q-E--I------------------Q--E-V--Q--Q-R-------Y-----------E-T-G--.-LD-LT-NIQ--Q--KQ-..L-V--------DA---I---   947
CPZ.US.85.CPZUS       VI.-----S----ST-R-A---AG-QQE--I---------------E--K-------Q--H-K---Q---F-------------T-G--.I-D--AS--Q-DL--KQ-..L-V--------DS---I---   953
GOR.CM.06.CP684       VI.-----P----AT-K-A---LNVT-E--I---------------E--K--Q-V--Q--H-K---Q---FV-.........................................................   513
GSN.CM.99.CN166       H-.-----P--V-KELE-A---AD-Q--T-------------N-------T-TK--EEVTY-----AQ-CY-----------DMCPT--.IV-M-------QH-NTQN..--F-K------Q-AN-L-Q-   962
GSN.CM.99.CN71        H-.-----P--I-KELE-A---AN-Q-ST-------------N-------T-QK---EVTY-----AQ-CY-----------DMCPT--.-V-M-------QH-NTQS..--F-K------Q-AN-H-Q-   955
H2A.DE.x.BEN          H-.-----P-----E-KMVA--VG--QS--------------A--HH--NQ-SR--EQ-NTI--I-L----CM----R----DMTP---.---M-T--Q-IQF--R-N..-NFK--Q--------QL---   975
H2A.GW.x.ALI          H-.---S-V-----E-KMVA--VG--QS--------------A--HH--NQ-SR--EQ-NTV--I-L----CM----R----DMTP---.---M-S--Q-IQF--T-N..L-FK--P--------QL---   975
H2A.SN.x.ST           H-.-----A-----E-KMVA--IG--QS--------------A--HH--NQ-SR--EQ-NTV--I-L----CM----R----DMTP---.---MVTA-Q-IQF--A-N..--L------F-----QL---   975
H2B.CI.x.EHO          H-.-----A-----D-KMAA--IG--Q--------E------A--HH--NQ-DR---Q-VSI--V-L--T-CM----R----DMTP---.IV-M-T--Q-IQF--T-N..L-F------------QL---   975
H2B.GH.86.D205        H-.-----A----PS-KMVA--VG--Q---------------A--HH--NQ-DRL--Q-VSI--V-L--T-CM----R----DMTP---.-V-M-T--Q-IQFF-A-N..L-F---Q--------QL---   976
H2G.CI.x.ABT96        H-.-----A-----E-KMVA--VG--QA-------------------HH--KQ-DK---Q-NSI--I-L----CM----R----DMTP---.I--M-T--Q-IQFQ-T-N..--FK-----------QL---   958
H2U.FR.96.12034       H-.-----A-----E-KMVA--VG------------------A--HH--NQ-ER--EQ-NTM--I-L--A-CM----R----DMTP---.I--M-S--Q-IQF--S-N..--FK-----------QL---   971
MAC.US.x.239          H-.-----A--A--E-KMVA--AG------------------A--HH--NQ-DR--EQ-NSV--I-L----CM----R----DMTP---.---M-T--Q-IQFQ-S-N..--FK-----------QL---   962
MAC.US.x.251_1A11     H-.-----A--A--E-KMVA--AG------------------A--HH--NQ-DR--EQ-NSV--I-L----SM----R--T-DMTP---.---M-T--Q-IQFQ-S-N..--FK-----------QL---   962
MAC.US.x.251_BK28     H-.-----A--A--E-KMVA--AG------------------A--HH--NQ-DR--EQ-NSV--I-L----CM----R----DMTP---.---M-T--Q-IQFQ-S-N..--FK-----------QL---   958
MAC.US.x.EMBL_3       HIYTHS--A--A--E-KMVT--AG--AHLW------------A--HH--NQ-DR--EQ-NSV--I-L----CM----R----DMTP---.---M-T--Q-IQFQ-S-N..--FK-----------QL---   956
SMM.SL.92.SL92B       H-.-----A-----E-KMVA--LGV-QS--------------A-DLH--KN-DK--EQ--SV--L-L--A-CM----R----DMTP---.IV-M-T-----QY-NSQN..--F------------QL---   955
SMM.US.x.H9           H-.-----A-----E-KMVA--AX--Q--X--------X---A--HH--TQ-DR--EQ-NSI--I-L--I-CI----R----DMTP---.-V-M-T--Q-IQFQ-S-N..--FK-----------QL---   958
SMM.US.x.PGM53        H-.-----A-----E-KMVA--AG--Q---------------A--HH--TQ-DR--EQ-NSI--I-L----CM----R----DMTP---.-V-M-T--Q-IQFQ-S-N..--FK-----------QL---   958
SMM.US.x.SIVsmH635FC  H-.-----A-----E-KMVA--AG--Q---------------A--HH--TQ-DR--EQ-NSV--I-L----CM----R----DMTP---.-V-M-T--Q-IQFQ-S-N..--FK-----------QL-R-   958
SMM.US.x.SIVsmH635F_L3 H-.-----A-----E-KMVA--AG--Q---------------A--HH--TQ-DR--EQ-NSV--I-L----CM----R----DMTP---.-V-M-T--Q-IQFQ-S-N..--FK-----------QL-R-   958
SMM.US.x.SIVsmH635SB10 H-.-----A-----E-KMVA--AG--Q---------------A--HH--TQ-DR--EQ-NSV--I-L----CM----R----DMTP---.-V-M-T--Q-IQFQ-S-N..--FK-----------QL-R-   958
MNE.US.x.MNE027       H-.-----A-----E-KMVA--AG------------------A--HH--NQ-DR--EQ-NSM--I-L----CM----R----DMTP---.-L-M-T--Q-IQFQ-S-N..--FK-----------QL---   958
STM.US.x.STM          H-.-----A-----E-KMVA--AG--Q---------------A--HH--TQ-DR---Q-NTV--V-L----CM---KR--L-DMTP---.-V-M-T--Q-IQFQ-S-N..--FK-----------QL---   958
DEB.CM.99.CM40        -I.-----P--I--P-K-A---LG---T--------------K-RI--DT-K-------------A--TY-I----R----DM-PI--.-V-M-T--I----------..--LLG-K------A--T---   956
DEB.CM.99.CM5         -I.-----P--I--P-K-A---LG-D--T-------------K-RV---T-K---E---------A--TF-I----R----DMA-I--.-V-M-T-D---Q-------..--LLGYK------A--T---   956
RCM.GA.x.GAB1         K-.-----A----AA-Q-V---AQ---A---------------------I--E-V-EQ--K-----Q---LV----------------G--.I-D--A-D-A-NK--NQ-..---------------QL-R-   965
RCM.NG.x.NG411        H-.----------AA-Q-D---AQ-----------------------H---I-------Q--K-----Q---L-----------------G--.I-D--ASD-Q-TK--NQ-..----------F-----QQ---   969
DRL.x.x.FAO           H-.-----P----EK--TV---A-----T------------------N---K-------Q--K-----Q---L-----------E------.IVD--ASD-L-SK---N-..L--------------QL---   965
MND_1.GA.x.MNDGB1     K-.-----P-----E-ETM---LG------I-----------NK--Y---L-EK--E-CKE-K---A--TF-----QR--L--MT-G--.IV-M-N----YQYQ-NQ-..--NL--K--F-----QL---   960
MND_2.CM.98.CM16      H-.-----P----EK--TV---AQ----T-I---------I-AK-HH--Q----V--Q--K-----Q---L------------E--PG--IVDI-ATDL-T--LQHXX-..Q--------------QQ---   964
MND_2.GA.x.M14        H-.-----P--V-EK--TV---AQ----T-I-----------AK-HH--V--E-V--Q--------Q---L------------E--PG--.IVD--T-DIL-TK---N-..---------------EQL---   964
MND_2.x.x.5440        H-.-----P--V-EK--TV---AQ----T-------------AK-HH--K--E-V--Q--------Q---L------------E--PG--.IVD--T-DIL-TK-R-N-..-------Q--------QQ---   964
GRV.ET.x.GRI_677      H-.-----P-----N---V----N----T-I--------S-----R------S-----C-------Q--T-------------I-S---.-V-MLT-Q--LN---NQ-..Q--L--K-------------   973
TAN.UG.x.TAN1         K-.-----P--V--E-QT-----QV---T-I--------S-----R----T-EK--E-CAF-----L--C--------------MTP---.---M-T-Q--IQHI-TQQ..Q--S--K-------E-----   970
VER.DE.x.AGM3         --.-----P-----E---M---------T----------SI-----------K----CQYT----L--C--------------LTP---.---M-T-Q--LQH--T--..Q--L----------------   967
VER.KE.x.9063         --.-----P-----E-------------T----------S------------K----CQYT----L--C--------------LTP---.---M-T-Q--LQ---T--..Q--L----------------   967
VER.KE.x.AGM155       H-.-----P-----E---M-----V---T----------SI-----------K----CQYT----L--C--------------LT----.---M-T-Q--IN---T--..Q--L----------------   970
VER.KE.x.TYO1         --.-----P-----E-------------T-I--------SI-----------K----CQYT-A--L--CIL------------QTS---.-----T-Q--IQH--T--..Q--L----------------   969
BKM.CD.02.AY518534    HI.-----P--I--EY--AV--TG----T----------I--N------ST-KK---EVQY-----AQ-CF-L
TAL.CM.00.266         H-.-----P--V-AEMQ-MV--L----SA----------A--NK-----KT-T----EVQY-S---AQ-TF-L-Y--R--L-DMCP--A.I--M-Y---Q-TQ--NQ-..QNFSD-K----K-AN-L-Q-   959
TAL.CM.01.8023        H-.-----P--V-AEMQ-TA--L-----T----------S--NK-----KT-Q----EVQY-S---AQ-TF-L----R--L-DMCP--A.---M-Y---Q-T---NQ-..HNFSD-K----K-AN-L-Q-   960
TAL.US.x.1511         H-.-----P--V-AEMQ-MA--L-----T----------A--NK-----KT-T----EVQY-S---GQ-TF-LX..........................................................   200
TAL.US.x.1867         H-.-----P--V-AEMQ-MA--L-----T----------A--NK-----KT-T----EVQY-S---GQ-TF.............................................................   815
TAL.US.x.416          H-.-----P--V-AEMQ-MA--L-----T-------------NK-----KT-T----EVQY-S---GQ-TF.............................................................   200
TAL.US.x.742          H-.-----P--V-AEMQ-MT--L-----T-------------NK-----KT-T----EVQY-S---GQ-TFX............................................................   211
MON.CM.99.L1          -I.-----P-----EFE-AA--AN-H--T-------------N--R----T-KK---EV-Y-P--LAQ-LF-L----------DMAPVD-.F--M------LQ-SNNQ-..T-FSK----F-T-A--Q---   958
MON.NG.x.NG1          HV.-----P-----EF-XAA--AN---ST-------------N----I-TTL-K---EVQF-S--LAQTLY-L----X----DI-PI--.-V-M---D--LQHSNNQ-..T-FSK----F-T-A--S---   961
MUS_1.CM.01.1085      -I.-----A--I-AEFE-A---AS----T-------------NK------A-HK--EEVTY-----AQ--F-L---K---V-DL-N---.I--ML-----LQH---QK..Y-F--------T-S--S---   952
LST.CD.88.447         -V.-----P--V-KE-Q-VI--LG----T-I-----------AK--A--Q--ERV-E--QQ-K---L--T------QR--L--L-P---.F--M-NAD---QY--KLN..---LK-K----Q----Q---   970
LST.CD.88.485         -V.-----P--V-KE-Q-VT--LG---ST-I-----------AK--V--Q--ERV-E--QQ-K---L----------QR--L--L-P---.F--M-NAD---QY--KLN..---LK-K----Q----Q---   970
LST.CD.88.524         -I.-----P--V-KD-Q-VT--LS----T-I-----------AK--V--Q--ERV-E--QQ-K---L----------QR--L--L-P---.Y--M-NAD---QYI-KLN..---LK-K----Q----Q-R-   971
LST.KE.x.lho7         -I.-----P--V-KE-Q-VT--IG----T-I-----------AK--V--S--ERV-E--QQ-K---L----------QR--L--LTP---.F--M-NA----QY--KLN..---LK-K----Q----Q---   970
SUN.GA.98.L14         -I.-----P--I-KD-E-V---LG-Q--T-I-----------AK--V--Q--SR--E--QE-K---L--L-------QR--L--MT----.F--M-NAD---QY--KIN..---LK-K---------Q---   970
SYK.KE.x.KE51         HI.-----P--I-DAF--A---AGV---T-I---------I--NK-R----T-N---EEV-------A--TF-L----------TTP-D-.Y--MLY---...Q--NTH.TQ-FS--K----H-T.SD-Q-   987
SYK.KE.x.SYK173       -I.-----P--V-DKFK-A---CG----T-I---------I---K-RY---A-S-----VTH-Q---A--TF-L----------I-PG--.Y--MLY---...Q---NTTSP-FS-------Q-K.NE---   983
COL.CM.x.CGU1         -V.-----P--V--HFK-AV--LG-A--T-H-----------QR--DV-RK-KKMK-Q--T--SK-A---YAL-------L--K-PW--QVERA-IE.-D-QN-.T-LQNQ-FK--KA-WK-H.TGE-Q-   929
```

```
                                                                 Pol p31 Integrase end \
SAB.SN.x.SAB1C          PAKLIWKGEGAVVIQEQG.ELKTIPRRKAKIIKDYG.K..ALDSQAPLEGNGRTAG...EVD......    1038
H1B.FR.83.HXB2          ----L----------DNS.DI-VV--------R---KQ.M-G-DCVASRQDED...............    1002
H1B.US.90.WEAU160       ----L----------DNS.DI-VV--------R---KQ.M-G-DCVASRQDED...............    1002
H1N.CM.95.YBF30         ----L----------DN-.DI-VV--------R---KQ.M-G-GCVASGQDENQEME...........    1013
H1O.BE.87.ANT70         --Q-L----------DK-.DI-VV--------RE--KQ.M-GTDSMASGQTESESVEQPSEIP.....    1008
H1O.CM.91.MVP5180       --Q-L----------DK-.DI-VV--------R---KQ.M-GTDSMANRQTESESMEQPGEIP.....    1008
CPZ.CD.90.ANT           --Q-L---------KD-E.-I-VV---------E-RK-.IEDRDDLAGRQ-ED...............    1002
CPZ.CM.01.SIVcpzCAM13   --T-L----------D--.-I-VV--------R---KQ.M-G-DCVASRQDED...............    1001
CPZ.CM.05.SIVcpzEK505   ----L----------DN-.DI-VV--------R---KQ.M-G-GCVASGQDESQDME...........    1012
CPZ.CM.05.SIVcpzLB7     ----L--------L-D-E.-I-VV--------R---KQ.MEGANCLADRQDEN...............    1005
CPZ.CM.05.SIVcpzMB66    ----L--------L-D-E.-I-VV--------R---KQ.M-G-DCVASRQDENQNME...........    1009
CPZ.CM.05.SIVcpzMT145   ----L----------D-D.-I-VV--------R---KQ.M-GADSVASRQDES...............    1001
CPZ.CM.98.CAM3          --R-L---------K-RE.-V-V---------R---KQ.M-G-DSMAGGQDESQGLE...........    1014
CPZ.CM.98.CAM5          --R-L---------K-NE.-V-VV--------R---KQ.M-G-GSMAGRQ-EDQNLE...........    1007
CPZ.GA.88.GAB1          --T-L----------D--.---VV--------R---KQ.M-G-DCVASRQ-ED...............    1000
CPZ.GA.88.GAB2          --S-L----------DN-.DI-VV---------H--KQ.V-G-DCLAGGQDENQN.............    1007
CPZ.TZ.01.TAN1          --R-L---------K-GE.DI-VV---------E--KQ.M-GAGGMDDRQ-ET...............     998
CPZ.US.85.CPZUS         ----L---------K-NE.-V-VV------------KQ.M-GADSMAGRQDES...............    1004
GOR.CM.06.CP684         ....................................................................     513
GSN.CM.99.CN166         --V-L--------V-T-AG-II-V---------P--EA.KTE-VGGATHT-ND...............    1014
GSN.CM.99.CN71          --V-L--------V-T-AG-II-V---------P--EA.K-E-VGSAAHTSNDREEGRMAD.......    1015
H2A.DE.x.BEN            -GE-L-------IVKVGT.DI-VV--------R---GR.QE---SPH---ARED.-EMACPCQVPEIX    1040
H2A.GW.x.ALI            -GE-L---D---IVKVGT.DI-VV--------R---GR.QE---GPH---ARED.-EVA.........    1031
H2A.SN.x.ST             -GE-L---D---IVKVGA.DI-I-------------GR.QEM--GSN---ARED.-EVA.........    1031
H2B.CI.x.EHO            -GD-L-------I-KVGT.-I-V---------RN--GG.KE--CS-DV-DTMQA.REVAQ......SN    1034
H2B.GH.86.D205          -GE-L-------I-KVGT.-I-VV--------RH--GG.KG--CS-DM-DTRQA.REMAQ......SD    1035
H2G.CI.x.ABT96          -GE-L-------I-KVGT.-I-VV------------GG.KEV--STNM-DTRQ-.-EVA.........    1014
H2U.FR.96.12034         --D-L-------L-KVGT.DI-V---------R---GG.KE---SSN---AEKV.REMAL.PDQTPXV    1035
MAC.US.x.239            -GE-L-------ILKVGT.DI-VV------------GG.KEV--SSHM-DT-EA.REVA.........    1018
MAC.US.x.251_1A11       -GE-L-------ILKVGT.DI-VV------------GG.KEM--SSHM-DT-EA.REVA.........    1018
MAC.US.x.251_BK28       -GE-L-------ILKVGT.DI-VV------------GG.KEV--SSHM-DT-EA.REVA.........    1014
MAC.US.x.EMBL_3         -GE-L-------ILKVGT.DI-VV------------GG.KEV--SSHM-DT-EA.REVA.........    1012
SMM.SL.92.SL92B         --E-L---------KVGT.-I-VV------------GR.QEMG-S-SM-DQ.................    1004
SMM.US.x.H9             -GE-L-------ILKVGT.-I-VV------------GG.KE---GSH--DT-EA.REVA.........    1014
SMM.US.x.PGM53          -GE-L-------ILKVGT.-I-VV------------GG.KE---GPH--DT-EAREVA..........    1014
SMM.US.x.SIVsmH635FC    -GE-L-------ILKVGT.-I-VV------------GG.KE---GSH--DT-EA.REVA.........    1014
SMM.US.x.SIVsmH635F_L3  -GE-L-------ILKVGT.-I-VV------------GG.KE---GSH--DT-EA.REVA.........    1014
SMM.US.x.SIVsmH635SB10  -GE-L-------ILKVGT.-I-VV------------GG.KE---GSH--DT-EA.REVA.........    1014
MNE.US.x.MNE027         -GE-L-------ILKVGT.DI-VV------------GG.KEV--SSHM-DT-EAREVA..........    1014
STM.US.x.STM            -GE-L-------IVKVGT.DI-VV------------GG.KEV--GSH--DTREA.-EVA.........    1014
DEB.CM.99.CM40          --T-L--------CKTEVGDI-VV------V--E--KD.MDSKVNTQNSNE.................    1006
DEB.CM.99.CM5           --T-L--------CKTEVGDI-VV------V-----KG.VDSKTDTQNPSE.................    1006
RCM.GA.x.GAB1           ----------------ETGD--VV---------E--R-.DVDSEANLAGRQEEN..............    1018
RCM.NG.x.NG411          ------E--------D-E.D--VV----C-------R-.EVDRETNMEGRQEES..............    1021
DRL.x.x.FAO             --E-V---------K-GT.D--V-------------KN.VDSHTNMES....................    1011
MND_1.GA.x.MNDGB1       -GI-L--------LKY-E.-I-IV----C-------ES.G.KN--VN--SV.................    1008
MND_2.CM.98.CM16        --E-----------K-GT.D--VV--------R---KT.VDS-PNVEA....................    1010
MND_2.GA.x.M14          --E-----------K-GT.D--VV--------R---KT.VDSNPHMEDRQETA...............    1015
MND_2.x.x.5440          --E-----------K-GV.D--VV------------KT.VDSNPHMGR....................    1010
GRV.ET.x.GRI_677        --R-----------K-GE.DI-VV------------ER.KTM--EGSM--VREANKQMEGDSDLQDQE    1039
TAN.UG.x.TAN1           --T-L---------K-GE.---VV---------E--ER.KTVGDKTNM--..................    1018
VER.DE.x.AGM3           -GQ-----------KGGV.---EY-----------EPR.KRMGDESN---A-GADN............    1021
VER.KE.x.9063           --Q-----------K-GE.---IV-----------EPR.KRMGDEGNM--LRGPDNQMAR.NSQILDD    1032
VER.KE.x.AGM155         --R----------LK-GE.---VV-----------EPR.KT-GDETH---A-GSDHQMAG.-S.....    1030
VER.KE.x.TYO1           --Q----------LKDGS.D--VV-----------EP-.QRVGNEGDV--TRGSDN............    1023
BKM.CD.02.AY518534
TAL.CM.00.266           --H----------LRTDEG-VI-V---------P--QA.LGNKIDLESSKEQDAEMGRDN........    1018
TAL.CM.01.8023          --H-V--------LRTDEG-VI-V---------P--QA.MGNKTDLEGSKEQDAEMGRDN........    1019
TAL.US.x.1511           ....................................................................     200
TAL.US.x.1867           ....................................................................     815
TAL.US.x.416            ....................................................................     200
TAL.US.x.742            ....................................................................     211
MON.CM.99.L1            --R-L---------KTEEGDIL-V---------P--TT.ENVGGDTNQYNLRKQD-LAN.........    1016
MON.NG.x.NG1            --H-L------L--KTDQG-VI-V--------RP--S-.EDVG-KPSAHQIREVD-MAD.........    1019
MUS_1.CM.01.1085        --V-L-R-------KTDSG-VI-V---------P--A-.EDVG-KSDTGDLRKED....RL-N.....    1010
LST.CD.88.447           --Q-L--------VK-GD.TIFSV------LV----E..GPK--EGG-HN-.................    1018
LST.CD.88.485           --QIL--------VK-GD.TIFSV------LV----E..GPK--EGS-DN-.................    1018
LST.CD.88.524           --Q-L--------VK-GE.NIFS-------LV----E..GPK---SSMDNQ.................    1019
LST.KE.x.lho7           --Q-L--------VK-GE.NIFSV------LV----E..GPK--ESS-DN-.................    1018
SUN.GA.98.L14           ----L---------KQGE.NILVV------LV---G......ESSSV-MV-.................    1015
SYK.KE.x.KE51           --S-L----------TPDQQVIAV--------SSD-E...RV--GTH--ITSKSN.............    1039
SYK.KE.x.SYK173         --R-L--------V-TEEGDIFAV--------T-H-E...RM--GSHV-NDPK-D.............    1035
COL.CM.x.CGU1           -GE-V---------RNSQGT-FVK----V--TRLQYGEDVGSENLLSNGQKEAETVKGMD........     989
```

```
                       / Vif start
SAB.SN.x.SAB1C         MEK...HWIVRPLWKVTGGQQERWT.................................SLVKYHMHVSKQCVHWRYTPHTKIRWNW.YSYQEWVIPLK..DGALIKVTNYWH.LTPEKGWLETYA.....    85
H1B.FR.83.HXB2         --NR.WQVMIVWQVDRMRIR...................................TWK----H--Y--GKARG-.FYR-HYESPHPR.ISS-VH---G..-AR.LVI-T--GLH-G-RD-HLGQG.....    84
H1B.US.90.WEAU160      --NR.WQVMIVWQVDRMRIR...................................TWK----H--YI--KAKG-.SYR-HYESTHPR.ISS-VH---G..E-K-.VI-T--GLH-G-RD-HLGQG.....    84
H1N.CM.95.YBF30        --NR.WQVM-VWQVDRMKIRKWNS...................................---H--Y---KAKG-.-YR-HYETHHPK.ISS-VH--VG..QAR.LVTVT--GLT-G-QS-HLGHG.....    84
H1O.BE.87.ANT70        --NR.WQVLIVWQVDRQKVKAWNS...LVKYH................................KYR-RKTEN-.WYR-HYESR-PR.VSSSVY--VG..VAH.VV--T--GLMPG-RDEHLGHG.....    84
H1O.CM.91.MVP5180      --NR.WQVLIVWQIDRQKVKAWNS...LVKYH................................KYM--KAAN-.RYR-HYESR-PK.VSSAVY--VA..EAD.-V--T--GLMPG-REEHLGHG.....    84
CPZ.CD.90.ANT          I--R.*KIEMIWQVDRMRINIWKS...................................---H-IWET-VLKP-.KYR-HYENDHPK.KGE-VH---PTL-KK-.V--VF-GLQCG-RP-HLGHG.....    86
CPZ.CM.01.SIVcpzCAM13  --NK.WQVMIVWQVDRMRIKAWNS...................................------YKT-KAKD-.FYR-HYEHP-PR.VSS-VH--F-.DNSK-.II-T--ALT-G-RA-HLGHG.....    85
CPZ.CM.05.SIVcpzEK505  --NR.WQVM-VWQVDRMRVRTWNS...................................---H--Y---KARG-.FYR-HYETD-PK.ISS-IH---G..EAK-.VIVT--GLMPG-RP-HLGHG.....    84
CPZ.CM.05.SIVcpzLB7    --NR.WKVLIVWQIDRMRIKAWNS...................................-----IY--RKARG-.FYR-HYESE-PK.VSS-IH---G..EAK-.VIVT--GLTPG-RD-HLGHG.....    84
CPZ.CM.05.SIVcpzMB66   --NR.WQVMIVWQVDRMRIKTWNS...................................---H--Y--RKARN-.FYR-HYESN-PK.ISS-VH---G..EAK-.VI-T--GLT-G-RA-HLGHG.....    84
CPZ.CM.05.SIVcpzMT145  --NR.WQVQIVWQVDRMRVKTWNS...................................------YR--KAKN-.FYR-HYDGT-PR.MASQIY--IN..QAK-.I--T--GLMPG-RD-HLGHG.....    84
CPZ.CM.98.CAM3         --NR.WLVMIVWQVDRMRVKAWNS...................................---C--YKT-KAKD-.FYR-HYESSHPK.VSS-IH--IE..KAK-.V--T--GLMPGQHA-HLGHG.....    84
CPZ.CM.98.CAM5         --NR.WQVM-VWQVDRMRIKTWNS...................................---T--YRT-KAKK-.FYR-HYESD-PK.VSS-VH---D..TAK-.V--T--GLT-GQHA-HLGHG.....    84
CPZ.GA.88.GAB1         --NR.WQVMIVWQVDRMRIKTWNS...................................-----IYR--KARG-.FYR-HYDHP-PK.VAS-IH--FRD.YSK-.I--T--ALSPG-RA-HLGHG.....    85
CPZ.GA.88.GAB2         --SR.WQVMIVW-VDRMRIRT...................................WN----F--YK--KARG-.FYR-HYDSP-PK.VSS-VH--IR..-AKI.V-CT--NLMPG-QP-HLGHG.....    84
CPZ.TZ.01.TAN1         --NR.WQVQ-VWMIDRMRLRTWTS...................................---H-IFTT-C-KD-.KYR-HYETDTPK.RAG-IH---T.ERSK-.V-LH--GLACG-RP-HLGHG.....    85
CPZ.US.85.CPZUS        --NR.WQVLIVWQVDRMRVKAWNS...................................------YR--KAKG-.FYR-HYEST-PR.VNS-VH--ID..KAKI.VIIT--GLK-G-QA-HLGHG.....    84
GSN.CM.99.CN166        --RQ.RQKMWVVQPIQIMTERKVDWLLR..ATK................................-HIWSGKTPFVYV-HYQLQHQRFTQNKIRLAMDT.RKVG.EEVEATYIEITILWDTTSHG.....    88
GSN.CM.99.CN71         --RQ.RQKMWVVQPIQVMTERKVEWLIR..ATK................................-HIWSGKTPFVYI-HYQLQHQRFTQNKIRLA-DT.RQVG.EEVEATYIEITILWDTTSHG.....    88
H2A.DE.x.BEN           --ED.RN---V-T-R-P-R.M-K-HA.LVKYLKY................................RT-DLEEV--V--H-VG-A-.WTCSRVIF--Q..GKSH.LEIQAYWN--------SSH-.....    86
H2A.GW.x.ALI           --EG.KS---V-T-R-P-R.M-K-HS.LVKYLKY................................RT-DLEKVC-V--H-VG-A-.WTCSRVIF--Q..GRSH.LEIQAYWN--------SS--.....    86
H2A.SN.x.ST            --EG.KR--AV-T-R-P-R.M---HS.LIKYLKY................................RTGDLEKVC-V--H-VG-A-.WTCSRVIF---..GESH.LEIQAYWN--------SS-S.....    86
H2B.CI.x.EHO           --EE.KN--AV-T-RIPCR.L---HS.LIKYLKY................................RT-DLQQVS-V--H-VG-A-.WTCSRVIF---..E--H.LEVQGYWN----R-F-SS--.....    86
H2B.GH.86.D205         --EE.KD---V-T-RIP-R.L---HS.LIKYLKY................................RTGELQQVS-V--H-VG-A-.WTCSRIIF--N..K--W.LEVQGYWN----R-F-SS--.....    86
H2G.CI.x.ABT96         --EE.KR--AV-T-RIP-R.L-K-HS.LIKFLKF................................KT-DLQKAV-V--H-VG-A-.WTCSRVIF--T..KE-H.LEIQGYWN--------SQ--.....    86
H2U.FR.96.12034        --EE.KN---V-T-R-PRR.L---HS.LIKHLXY................................NT-ELQKVC-V--H-VG-A-.WTCSRVIF---..K-SH.LEVQAYWN--------SS--.....    86
MAC.US.x.239           --EE.KR--AV-T-RIPER.L---HS.LIKYLKY................................KT-DLQKVC-V--F-VG-A-.WTCSRVIF--Q..E-SH.LEVQGYWH--------S---.....    86
MAC.US.x.251_1A11      --EE.KR--AV-T-RIPER.L---HS.LIKYLKY................................KT-DLQKVC-V--F-VG-A-.WTCSRVIF--Q..E-SH.LEVQGYWH--------S---.....    86
MAC.US.x.251_BK28      --EE.KR--AV-T-RIPER.L---HS.LIKYLKY................................KT-DLQKVC-V--F-VG-A-.WTCSRVIF--Q..E-SH.LEVQGYWH----R---S---.....    86
MAC.US.x.EMBL_3        --EE.KR--AV-T-RIPER.L---HS.LIKYLKY................................KT-DLQKVC-V--F-VG-A-.WTCSRVIF--Q..E-SH.LEVQGYWH----R--PS---.....    86
SMM.SL.92.SL92B        --EG.KR-VAV-V-RISRRIV---HS.CIKFHKY................................KTRELEKAC-V--H-VG-A-.-TASRVIF--E..E-SH.LEVQVYWN--------SS--.....    87
SMM.US.x.H9            --EE.KN---V-T-RIP-R.L-K-HS.LIKHLKY................................NT-DLQKAC-V--H-VG-A-.WTCSRVIF--R..-E-H.LEVQGYWN----X---S---.....    86
SMM.US.x.PGM53         --EE.KN---V-T--IP-R.L-K-HS.LIKYLKF................................RT-DLQKAC-V--H-VG-A-.WTCSRVIF--R..-ESH.LEVQGYWN-A------S---.....    86
SMM.US.x.SIVsmH635FC   --EE.KN---V-T-RIPER.L---HS.LIKHLKY................................NT-DLQMAC-V--H-VG-A-.WTCSRVIF--R..-KTH.LEVQGYWN--------S-H-.....    86
SMM.US.x.SIVsmH635F_L3 --EE.KN---V-T-RIPER.L---HS.LIKHLKY................................NT-DLQMAC-V--H-VG-A-.WTCSRVIF--R..-KTH.LEVQGYWN--------S-H-.....    86
SMM.US.x.SIVsmH635SB10 --EE.KN---V-T-RIPER.L---HS.LIKHLKY................................NT-DLQMAC-V--H-VG-A-.WTCSRVIF--R..-KTH.LEVQGYWN--------S-H-.....    86
MNE.US.x.MNE027        --EE.KR--AV-T-RIPER.L---HS.LIKYLKY................................KT-DLQKVC-V--H-VG-A-.WTCSRVIF--Q..EESQ.LEVQGYWN----R---S---.....    86
STM.US.x.STM           --EE.KR---V-T-RIP-R.L---HS.LIKHLKY................................NT-ELSKAC-V--H-VG-A-.WTCSRVIF--Q..GE-H.LEVQGYWN--------SE--.....    86
DEB.CM.99.CM40         ---T.WIVRLTHKIPMNKLDRWCSAIKREKYKTK...................................ELDTV.QWI-HFELR-SY.YT-TKIVFPIT..KEGG.-VVELLWCLTPEKGWL........    81
DEB.CM.99.CM5          ---E.WIVRLTHKIP-NKLDKWCSAIKRGKYKTK...................................ELQAI.KWI-HYELRRAF.YT-TKIVFPVT..EEGK.-VVEIFWSLAPEKGWM........    81
RCM.GA.x.GAB1          -AEK..M-----I-R-DRRKI-................................QWHSLVKYHQYKGK-AAKE-E-V--F-VP-G-.W-HS-VH---E..E-SK.LKITTYWN--V-----G--G.....    87
RCM.NG.x.NG411         -AEK..R--E--I---DRRKV-Q-HS.LVKYHQY................................KG-KEAKDWEYVPHFKVPWG.WWSRSE-HIPL..GNNTKIKVTTYWS--I-----G--G.....    87
DRL.x.x.FAO            --.K..M---I-T-R-ERYMV-................................RWNSLCKYHKHKGE-YLER-H-A--FQCS.G-.WTHSQ-T--F-..-KSK.I-VTVLWN---D----SK--I....    86
MND_1.GA.x.MNDGB1      --.R.VER---LT---SSQRI-K-HW.LVRRQMA.....................WAT.ANNEEGCWWLYPHFMA-NEWYTCS.KV............................................    60
MND_2.CM.98.CM16       --.K..Q---I-T--LDR................................HAVERWNSLVKYHKYKGE-HLDK-DLKA-FQCS.G-.WTHSQKY---E..EDEQ.IIITILWN--------S---.....    85
MND_2.GA.x.M14         --.K..Q---I-T-RIDRR-L-................................KWNSLVKYHKYKEE-HLDK-ELFH-FQCS.G-.WTHSQKI--F-..--SK.IIITALWN--------SQ--.....    85
MND_2.x.x.5440         --.K..Q--AI-T-EDR...............................QDVKLRDWNSIVKYHKYKGE-HLEN-QLY--FQCS.G-.WTHSQKK--F-..--SK.IIITVLWN--------SQ--.....    86
GRV.ET.x.GRI_677       --.REKQ----VV-R-SER-IS--RG.IVTYKIRN..................................KQLP-E-RH-WQVQ.WQFWT-SQFI---S..KDDY.IEV-IY-N----R---SSHG.....    85
TAN.UG.x.TAN1          --.REKL-VT-LT-R-S...................................GEHIDKWKGIVKY-MRNRLQD-T-LM-YQCG-A-.-TCSRFL---G..GEGK-V-DC---.----Q---S---.....    85
VER.DE.x.AGM3          -NQE.KE-VM-VT---PEELITK-Q.GIVRYWMR..................................TRKLD-K-RM-YQ-T-A-.-TMSRYE---G..QHGS-H-DL---.--------S---.....    85
VER.KE.x.9063          -NQE.KE-VM-VT---S-DLITK-Q.GIVKYWMT..................................-RNLK-Q-MM-YQ-N-A-.-TMCRYI--IG..LEGE-QIDC---.----R---S---.....    85
VER.KE.x.AGM155        -SQE.K--VM-LT---QEEVITK-Q.GIVRYWMN..................................-RNLK-E-KM-YQ-T-A-.-TMSRY----P..GSGE-H-DI---.-A-KQ---S---.....    85
VER.KE.x.TYO1          -NPN.KE-VM-VT---P-DLITK-Q.GIVRYWMR..................................QRNLK-N-YM-YQ-T-A-.-TMSRY---IG..KHGE-C-DL---.----Q---S---.....    85
TAL.CM.00.266          -D-L.WVTRLTWKVPRNRM-RWVG-I...................................RF-I-KT-YLQKLIWVH-FQL-HAS.-TQAK-IL--EEK-TRVEYL-ITQYWCLAPDHGYTSDWD....    89
TAL.CM.01.8023         -D-L.WVTRLTWKVPRNRM-RWVG-I...................................R---YKT-YLQKLIWVH-FQL-MAS.FTQAK-IL--EEKENRVEYI-ITQYWCLAPDKGYTSDWD....    89
MON.CM.99.L1           -APR.KM-V-T-TNIISESKMDWLIRCTKWHILT...................................GSAPFV-VH-YQLHNQR.F-QNKIKL--DLGITQEGESWATYLE.ITIYWD-TNVGPSALS    92
MON.NG.x.NG1           -APR.KM-V-S-VHIRSERWM-WLIRATKYNIYT...................................GK-PFV-VH-YQLTHQR.F-QCKIKL--EMGITQXXEQWATYXEI-ILWDXTNVGPASLSP    93
MUS_1.CM.01.1085       -APR.KM-V-S-IQETSERKIDWIIRATKYYILS...................................GRAAFT-VH-YQLYHQR.F-QSKIR---TQNTRESGA-ESTYIE--VLWDVTNVGPASCSK    93
LST.CD.88.447          --R..DQR---VAYIT-R....................................SLVEKFLDIRRAHKETQND-VSLHG-GVG-EF.-T-NNI----E..T-T-.V-R.CYAH-AAGR--ISQW-.....    82
LST.CD.88.485          --R..DQR---VA-IT-R....................................NLIEKFLDIRRAHKETQ-D-VSLHG-GVG-EF.-T-NKI----E..I-T-.V-R.CYAH-AAGR--ISQW-.....    82
LST.CD.88.524          --R..DQR---VA-ITNR....................................SLVEKFLDIKRNH-ESKND-VSLHG-GVG-EF.-T-NKLL---D..T-T-.V-R.IYAH-AAGR--ISQW-.....    82
LST.KE.x.lho7          --R..DQR---VA-IT-R....................................SLIEKFLDIVRLNTEAKEN-VGLHG-GVG-EF.-T-HKL----E..S-T-.V-R.IYAH-AAGR--ISQW-.....    82
SUN.GA.98.L14          --E.....NP-Q-RWWDE................REWEDRRQYKIVRIVWLIDRIAVEKFLDLRRMHRETKDD-VSMYG-GTG-E-.-T-NKII--VT..Y-TV.V-R.IYGH---A--LINQWG.....    99
SYK.KE.x.KE51          ---...E---VLT--LHPN-ID.................................KMQHIL-V-KYK---LEKAT-KH-YQ-T-E-.WTRAQ-E--VG..Q-V-..IIKFY-N------L-R-EG.....    84
SYK.KE.x.SYK173        ---...E---V-T-RM-PR-ID.................................RLQHII-T-KYK--ELEKAT-KH-YQ-E-Q-.-T-CQ-T--VG..--TI..WITFY-N-A--R---HMQG.....    84
COL.CM.x.CGU1          ---M.WAVRICYPMGKRKPKPL-EWI........................................KKQFPECQY-YLGKTESPE-.WTEGKIEFY-G.SAIK-..SII-LGT..LTPDRY.H.......    75
```

```
                                                                 phosphorylation site             |-    involved in inhibition of APOBEC 3G     -|
                                                                            Thr96              Cys114                      Cys 133
                                                                              |                   |                          |
SAB.SN.x.SAB1C          ......................................T....GIGYSKGEWFTELDPWTADHI.IHWS.YFPCFTDRA.....VQQAIRGEKYLW.CKHQVGHQPT...GQVPSLQYLALRVYT.....   157
H1B.FR.83.HXB2          ......................................VS.IEWRKKRYS.TQVDPE..L--QL.--LY.--D--S-S-.....IRK-LL-HIVSPR-EY-A--N......K-G-------AALI.....   154
H1B.US.90.WEAU160       ......................................VS.IEWRKQRYS.TQVDP-..L--QL.--LY.H-D--SES-.....IRN--L-HLVIPR-EY-A--N......K-G-------TALI.....   154
H1N.CM.95.YBF30         ......................................VS.IEWRLRKYK.TQVDPE..M--KL.--LH.--D---AS-.....IR--VL-RPV-PR-EYPA--K......--GT------TAWV.....   154
H1O.BE.87.ANT70         ......................................VS.IEWRYKKYK.TQIDPE..---RM.--LH.--T---AS-.....-RK--L-QRV-TK-EYPT--S......--GT--L----AVV.....   154
H1O.CM.91.MVP5180       ......................................VS.IEWQYKEYK.TQIDPE..---RM.--LH.--T---ES-.....IRK--L-QRV-TK-EYLA--S......--GT--F---KAVV.....   154
CPZ.CD.90.ANT           ......................................VS.IEWRCGKYI.TQVDPE..---QM.--QY.-----S-Q-.....-R---L--RI-TY-HYKK--S......--GT-----FCKILEFR..   159
CPZ.CM.01.SIVcpzCAM13   ......................................VS.IQWRMG-YV.TQVDPY..V--KL.--SQ.--D--A-S-.....IRR--L-HIVEPR-AYRE--R......--GT--F---KALV.....   155
CPZ.CM.05.SIVcpzEK505   ......................................VS.IEWRQGIYR.TQIDPE..L--KL.--LY.--D---AS-.....IR--VL-RPVIPK-EYPA--K......--G-------IAWV.....   154
CPZ.CM.05.SIVcpzLB7     ......................................VS.IEWKQKKYS.TQIDPE..L---M.--LY.--N--AES-.....IRK--L-RVVSPR-EFPA--K......K-GT-----ISALL.....   154
CPZ.CM.05.SIVcpzMB66    ......................................VS.IEWRQR-YR.TQIDP-..L--QL.--LY.--D--AES-.....IRK--L-QIVSPR-EYPT--N......K-G-------STLI.....   154
CPZ.CM.05.SIVcpzMT145   ......................................VA.IEWRERGYS.TQVDPE..---RL.--LH.--D---AT-.....IRN--LEKAV-PK-EFPE--S......K-G----W--KQLV.....   154
CPZ.CM.98.CAM3          ......................................VS.IEWRCG-YS.TQVDPE..M--KL.--LQ.--D--AAS-.....IRK-VL-KQVYPK-EYPA---......--G--------AWV.....   154
CPZ.CM.98.CAM5          ......................................VS.IEWRCG-YK.TQVDPE..---RL.--LH.--D--AAS-.....IRK-VL-KQVYPK-EYPT---......-XG--------AWV.....   154
CPZ.GA.88.GAB1          ......................................VS.IQWRLG-YV.TQVDPF..---RL.--SQ.--D--AET-.....IRR--L-QLVAPR-EYKE--R......--G---F---KALI.....   155
CPZ.GA.88.GAB2          ......................................VS.IEWRQRTYR.TQIDPQ..---RL.--CY.--D--AAS-.....-RR--L--IVTKV-EFPR--P......--Q------IVQVV.....   154
CPZ.TZ.01.TAN1          ......................................IG.LEWRQGKYS.TQIDPE..---QL.--TR.--T--AAG-.....-R---L--RI-TF-HF-S--R......--GT--F--F-KVVESQ..   158
CPZ.US.85.CPZUS         ......................................VS.IEWRQGTYK.TQIDPE..M--KL.--LY.--D---AS-.....IRK-VL-RPVIPK-EYPT--N......K-G---F---KALV.....   154
GSN.CM.99.CN166         ..................PASLSRSTYYQQAVIIEWVYNRSKL-AREGDII-YSTNLTPGVAMQI--GK.-----QNED.....IRR-----QI-GN-E-PEA-L......--GTP-T-EQLAFFAYVKH   188
GSN.CM.99.CN71          ..................PASLSRSTYYQQAVTIEWVYNRSKL-ARETDIT-YSTNLTPDVAMQI--TR.-----QNED.....-RR-----QL-GY-D-PEA-P......--GTPHT-ERLAFFAYVRY   188
H2A.DE.x.BEN            ..........................................VRLTWYTEKFW-DVT-DC--IL.--ST.--S---AGE....-RR------L-SC-NYPQA-KA....---------V-VQQ....   158
H2A.GW.x.ALI            ..........................................VR-TWYTEKFW-DVT-DC--SL.--GT.--S---AGE....-RR------L-SC-NYPQA-KS....------F---V-VQQ....   158
H2A.SN.x.ST             ..........................................VRLTWYTEKFW-DVT-DC--SL.--ST.--S---AGE....-RR------L-SC-NYPQA-KY....------F---V-VQQ....   158
H2B.CI.x.EHO            ..........................................VRLTWYERSFY-DVT-DV--RL.L-G-.--SS--ANE....-RR------I-SH-NYPSA-TG....------F----VQEGK..   160
H2B.GH.86.D205          ..........................................VRLTWYERNFY-DVT-DV--QL.L-G-.--S--SANE....-RR------I-SY-NYPSA-EG....------F----VQEGK..   160
H2G.CI.x.ABT96          ..........................................VRLTWYTRKFY-DVT-E---QL.L-G-.--D---AGE....-RR-----QI-SC-NYPTA-KR....------F---Q-VQKG...   159
H2U.FR.96.12034         ..........................................VRLTWYTRGFW-DVT-DC--QL.L-G-.--S---AGE....-RR-----QL-SC-NFPSA-RQ....------F----ALQ.....   157
MAC.US.x.239            ..........................................VR-TWYSKNFW-DVT-NY--IL.L-ST.------AGE....-RR-----QL-SC-RFPRA-KY....----------K-VS.....   157
MAC.US.x.251_1A11       ..........................................VRVTWYSKNFW-DVT-NY--IL.L-ST.------AGE....-RR-----QL-SC-RFPRA-KY....----------K-VS.....   157
MAC.US.x.251_BK28       ..........................................VR-TWYSRNFW-DVT-DY--IL.L-ST.------AGE....-RR-----QL-SC--FPRA-RY....----------K-VS.....   157
MAC.US.x.EMBL_3         ..........................................VR-TWYSRDLLDRCNTRLCR-F.SCIA.LISL--AGE....-RR-----QL-SC--FPRA-RY....----------K-VS.....   157
SMM.SL.92.SL92B         ..........................................VR-TWYSEKFW-DVT-DV--QL.A-ST.-----AAH-.....-R------QV-SY-GYA-A-HS.....S-Q---L---K-VLQ....   159
SMM.US.x.H9             ..........................................VR-TWYSRNFW-DAT-DY--TL.L-GT.-----SEGE....-RR------L-SC--FPKA-KN....----------T-VS.....   157
SMM.US.x.PGM53          ..........................................VR-TWYSRKFW-DVT-DY--TL.L-GT.-----SEGE....-RR------L-SC--FPKA-KN....----------T-VS.....   157
SMM.US.x.SIVsmH635FC    ..........................................VR-TWYSRNFW-DVT-DC--TL.L-ST.-----SEGE....--R------L-SC--FPKA-KN....----------T-VS.....   157
SMM.US.x.SIVsmH635F_L3  ..........................................VR-TWYSRNFW-DVT-DC--TL.L-ST.-----SEGE....--R------L-SC--FPKA-KN....----------T-VS.....   157
SMM.US.x.SIVsmH635SB10  ..........................................VR-TWYSRNFW-DVT-DC--TL.L-ST.-----SEGE....--R------L-SC--FPKA-KN....----------T-VS.....   157
MNE.US.x.MNE027         ..........................................VR-TWYSRNFW-DVT-DY--IL.L-ST.------AGE....-RR-----QL-SC-RFPRA-KN....------------VS.....   157
STM.US.x.STM            ..........................................VR-TWYTRNFWSDVT-DC--QL.L-GT.------AGE....-RR------L-SC-RFTKA-KN....----------K-VE.....   157
DEB.CM.99.CM40          .......................................PTMAVAIAWETKTWKTELTPDLADHL--LR.-----SQQ-.....-R------IL-QQ-QQYHRGP...EGKGP-----I----AVK....   159
DEB.CM.99.CM5           .......................................PSMAVGISWEQTGWQTELTPDLADHL--LR.-----SQQ-.....-T--LK---L-QQ-QQYHRGP...EGKGP---------TAVK....   159
RCM.GA.x.GAB1           ..........................................V--LWI--DYV-DVF-----SL.--KI.----------.....IR-SLL---V-V.-AF-G--RD.....--GT--F--IQAWAK.SQL   161
RCM.NG.x.NG411          ......................................VGIAYIDNRCDPCY--DI--TV--KL.V-NQ.-------K-.....IR---L--QV-L.-GFEK--RD.....--G------IKAWARQQLK   166
DRL.x.x.FAO             ...........................................T-EHIQEKFC-LI--T---R-.--LE.-------Q-.....IRGVLL-YRIQA.-YFPR--KG.....--G-------QAHIKYLK.   159
MND_1.GA.x.MNDGB1       ..................................................VIIINR-IRLIVRSYW-LQIEVG-LSTY-VSIEA-VRPPPF--EW..-TEITPEVA.....DHLIHLHFYDCFMDSAVM.   132
MND_2.CM.98.CM16        ..........................................MT-E-KFTNYY-HI--Q---RM.---K.-L----EQ-.....IR--LL-KRLTV.-YFHW--KG.....K-G-------LS--AYCN.   159
MND_2.GA.x.M14          ..........................................ITAE-N--DYY-HI--V---RM.---E.------AA-.....-RKVLY--RIVA.-YSPW--KG.....--GT--L----A-IKFCR.   159
MND_2.x.x.5440          ..........................................FTVE-Q-ENY--YI--V---R-.--GE.-------Q-.....IRK-LF--RLVA.-YFPW--RG.....--GT--F---QA-LRGRK.   160
GRV.ET.x.GRI_677        ......................................VGL.SYYHQKG..YK--V--G---RM.--LY.--N------.....I--------T-.-TFKE--KG.....--Q---L---VA--.....   156
TAN.UG.x.TAN1           ......................................VAI.SFENWQN.TYK--VT-DV---M.--CH.-----------.....I-------SF--.-TYKE--VAENHW---R---F---T---DFLR.   166
VER.DE.x.AGM3           ......................................EGI.QYLSNRDPWYR-----A---SL.--TH.--T---E--.....IRK-LL-QRFTF.-QFPE--KK-...-----------LA.......   158
VER.KE.x.9063           ......................................IGI.QYL-AVG.NFR-----A---GM.--LH.--S---E--.....I-K-----RFVF.-QFPE--KT-...-K---------LA.......   157
VER.KE.x.AGM155         ......................................VGI.QYVSLVNDKYR-----N---SM.--CH.--T------.....I---L--NRFIF.-QFPG--KL-...-----------LA.......   158
VER.KE.x.TYO1           ......................................VGI.QYVSNLESKYR-----A---S-.--GH.--N--KE--.....I---L--HRFVF.-QFPE--KS-...-----------LA.......   158
TAL.CM.00.266           ..................................TISIEWSVSD-VRDEKVYA--VT-NV---L.--TK.------AS-.....-R------VL-AV-GVRHKGRD.....TP----F---TAAVR....   169
TAL.CM.01.8023          ..................................TISIEWSVSDVTLDAKVYS---T-EV---L.--TK.------AS-.....-R------VL-AV-EVRHKGRD.....SP----F---K-AVR....   169
MON.CM.99.L1            PTVYNKQAYTIEWVYWMQ.KVER..VPLDPVWNPNPR..DREV.RTWYATHL.-...-DL-NQ-.V-TH.--A--QQLD.....-RR------L-GQ-Q-LAT-Y-.....K-GTPLT-EKLAFL.....   195
MON.NG.x.NG1            SVKWKQAYIIEWVYW.RQPRVPQQEGVVDRVWFPSPRRPDREIDRVWYATYL.-...-DV--Q-.--TR.--S--QASD.....-RR-----AX-GN-H-IQT-YNKVGTPPTLEKLAFLACFKYGKCTT   211
MUS_1.CM.01.1085        STWWKQSFILEWVYL.......................RRHPQDREQDYIWYS-Y-S-TL-MR-.--TQ.--S--YSQD.....IPRV---QHT-GD-E-PDA-TKVGPPPSLQV-AL---KK.......   186
LST.CD.88.447           ..........................................VS-EWIY-NYQ--I--IL-NQM.--CK.--N--ST-D.....IRK-LW---IMAF-AMPT--KG.....C-L---F-C--QLQ.....   153
LST.CD.88.485           ..........................................VS-EWIY-NYQ--I--IL--QM.--CK.--N--ST-D.....IRK-LW---IMAI-AMPT--KG.....C-L---F-C--QLQ.....   153
LST.CD.88.524           ..........................................VS-EWIY-NYQ--V--IL--QM.L-VK.--D-WVH-D.....IVR------I-HS--MPT--KG.....C-L-----C--QLQ.....   153
LST.KE.x.lho7           ..........................................VS-EWVYRDYQ--I--IL--QM.--YK.--N--IQ-D.....IRR-LL--RITMC-Q.PL--KG.....C-L---FIC--QLQ.....   152
SUN.GA.98.L14           ......................................C...SMEWIYNSYQ--I--LV--QM.--FK.--D-W-S-C.....IRR-ML---I-HE-RN--A-KG.....L-L---F-C---L......   169
SYK.KE.x.KE51           ......................................VGLIWAHPSGWS.TEVTPQ..--.DALV.H-Q-----S...DRA.----L----LTSH-WNF.-KE.....--L-------QK-LSK...   156
SYK.KE.x.SYK173         ......................................I.RIQYQWNQWN.TDLTPA..V--RL.--NF.------A--.....-N--V---LLTSH-WTP..-TD.....----------Q--LK....   154
COL.CM.x.CGU1           .....................................RGPPV.SYELQI-K-E-D-SWQQ-VGY.M-II.RGK-VRIEE...EARK----LPWNP.-DF----L......G-LK--E-C-LALREK.ER   153
```

```
                                                                                                       Vif end \
                                                                                                          |Thr188 phosphorylated
SAB.SN.x.SAB1C         ...........NGLRRVAPTSRRGSSQGSPQESQRRDTRM.ARNMGFAQRAVRRMAPRHVTGPQFRGPVPLPKESPFPSLVEYCGRTSH................................          234
H1B.FR.83.HXB2         .................................................TPKKI.K-PLPSVTKLTEDRWNKPQKTKGHRGSHTMN.................................GH          192
H1B.US.90.WEAU160      .................................................TPKKI.K-PLPSVKKLTEDRWNKPQKTKGHRGSHTMN.................................GH          192
H1N.CM.95.YBF30        .................................................G-KK-.K-PLPSVTKLTEDRWNEHQKMQGHRGNPIMN.................................GH          192
H1O.BE.87.ANT70        .................................................K-RS-.K-PLPSVQKLTEDRWNKHLRIRDQ-KSPSMN.................................GH          192
H1O.CM.91.MVP5180      .................................................KVK-N.K-PLPSVQRLTEDRWNKPWKIRDQ-GSHSMN.................................GH          192
CPZ.CD.90.ANT          .................................................GYPKGPRRQFPSLSILTEDRWNKPRRMRGHRENQ*MS.................................GC          198
CPZ.CM.01.SIVcpzCAM13  .................................................KEK-L.R-PLPSVAKLTEDRWNTPQRTKGHPENHTMS.................................GH          193
CPZ.CM.05.SIVcpzEK505  .................................................GVQK-.R-PLPSVTKLTEDRWNRHQKIKDHRGSHTTN.................................GL          192
CPZ.CM.05.SIVcpzLB7    .................................................QRQPT.R-PLPSVSKLAEDRWNKSQKTKGHRGSHIMS.................................GH          192
CPZ.CM.05.SIVcpzMB66   .................................................KSKP-.R-PLPSVRKLAEDRWNNSQKTRDHKGNHIMN.................................GH          192
CPZ.CM.05.SIVcpzMT145  .................................................KTKKN.R-PLPSVAKLAEDRWNKPQRTRDHQESPMSS.................................GC          192
CPZ.CM.98.CAM3         .................................................-VGKK.K-PLPSVTKLTEDRWNRHQKIRVHQESHTMN.................................GH          192
CPZ.CM.98.CAM5         .................................................GVQKK.T-PLPSIXKLAEDRWNKYQKTRGHQENHTMN.................................XH          192
CPZ.GA.88.GAB1         .................................................SER-H.R-PLPSVAKLTEDRWNKHQRTKVHQENLTRN.................................GH          193
CPZ.GA.88.GAB2         .................................................KTKKK.K-PLPSVTILAEDRWNKPQKTRDQQENLTMS.................................GH          192
CPZ.TZ.01.TAN1         ................................................DKQPKGPRRPLPSVTKLTEDRWNKHRTTTGRRENHTLS.................................GC          198
CPZ.US.85.CPZUS        .................................................GQSK-.R-PLPSVTKLTEDRWNKCQKTKGQPENHTMS.................................GH          192
GSN.CM.99.CN166        .............................................YGKNPTQSSY-LAQPK-AHNH-YRSKRSVGNKRGGR-TL.LKGRAPWNLA..........................          237
GSN.CM.99.CN71         .............................................YGKNPTKSPSTLDQPK-AHNH-SCSKRHVGNKRRGRKTL.LKGRAPWNLA..........................          237
H2A.DE.x.BEN           .............................................ND.RPQRKGT-RKQWRRDHW--LRVAREDHRSLKQGGSEPS.APRAHFPGVAKVLEILA.................          215
H2A.GW.x.ALI           .............................................NG.KPQRNSTTRKQWRRDYR--LRVARQD-RGLKQRGGESP.APGAHFPGVAKVLEILA.................          215
H2A.SN.x.ST            .............................................NG.RPQRDNTTRKQWRRNYR--LRVARQDGRSHKQRGSEPP.APRAYFPGVAKVLEILA.................          215
H2B.CI.x.EHO           .............................................DG..SQGESTTRKQRRRNSR--IRMARDNIRTSQQSSSQSL.AQGTYFPGLAEVLGILA.................          216
H2B.GH.86.D205         .............................................NG..SQGESATRKQRRRNSR-SIRLAR-NNNRAQQGSGQPF.APRTYFPGLAEVLGILA.................          216
H2G.CI.x.ABT96         .............................................HG..SKGESXTRKQRR-DNR--IRMAR-N-XRDQQDSSQSF.TQGAYFPGLAKVLGILA.................          215
H2U.FR.96.12034        .............................................DGR.SKRENPTRKFRRRNCR--FQLAR-NCERYQQGSSATS.-QGTYFPGLAKVLAILA.................          214
MAC.US.x.239           .............................................DVR.SQGENPTWKQWRRDNR--LRMAKQN-RGDKQRGGKPP.TKGANFPGLAKVLGILA.................          214
MAC.US.x.251_1A11      .............................................DVR.SQGENPTWKQWRRDNR--LRMAKQN-RGDKQRGGKPP.TKGANFPGLAKVLGILA.................          214
MAC.US.x.251_BK28      .............................................DVR.SQGENPTWKQWRRDNR--LRMAKQN-RGDKQRGSKPP.TKGADFPGLAKVLGILA.................          214
MAC.US.x.EMBL_3        .............................................DVR.SQGENPTWKQWRRDNR--LRMAKQN-RGDKQRGGKPP.TKGADFPGLAKVLGILA.................          214
SMM.SL.92.SL92B        .............................................NDR.PKGKNPTRKQWRRNNR--LRMVTQY-GG-ESNSSTTP.TKRVNFPGLEKVLGILGR...............          217
SMM.US.x.H9            .............................................HVX.SQGEXPTWKQWRRNXR--XRLATQNXRRNKQGSSESF.AXGANFPGLAKXLGILA.................          214
SMM.US.x.PGM53         .............................................HVR.SQREDPTWKQWRRNNR--IRMAKQN-RRNKQGSSESL.AEGANFPGLAKVLGILA.................          214
SMM.US.x.SIVsmH635FC   .............................................HVR.SQREDPTWKQWR-NNR--FRMAKQN-RRNKQGSSKSP.AEGANFPGLAKVLGILA.................          214
SMM.US.x.SIVsmH635F_L3 .............................................HVR.SQREDPTWKQWR-NNR--FRMAKQN-RRNKQGSSKSP.AEGANFPGLAKVLGILA.................          214
SMM.US.x.SIVsmH635SB10 .............................................HVR.SQREDPTWKQWR-NNR--FRMAKQN-RRNKQGSSKSP.AEGANFPGLAKVLGILA.................          214
MNE.US.x.MNE027        .............................................YVR.SQRENPTWKQWRRDNR-SLRMAKQN-RGDKQRGSKPP.TKGADFPGLAKVLGILA.................          214
STM.US.x.STM           .............................................HVR.SQRENT-RKQWRRGNRGSIRVATQNGRGHKPRGSKPS.TEGTDFPGLAKVLGILA.................          214
DEB.CM.99.CM40         ................................................HGK-PTLT-TTFHLKNNTHDKARVAAGHAR-KRGSPKTF.-QRRTIWSLESLCRLTRPT.RLDGRTSMGCIGH          230
DEB.CM.99.CM5          ................................................YGK-STLT-TTFHIKNSTNDKASTTT-HAG-KRGSFKTL.-KRRVMGSLESLCRPSCPT.RLDRRTSMGGVSN          230
RCM.GA.x.GAB1          DR...........................................YGR....KSPR.GPHW-WRS-V-ALATGHARKGQ-GSQVTLS-.RVHFPSVAHLCGTLA.................          216
RCM.NG.x.NG411         .............................................DGR.GQRKSARGP-WGWRCRVC-LASQNANRSQLRS.QVALS-.RVHFPSVAYLCGTLA.................          222
DRL.x.x.FAO            .............................................HGRKTVSGESSSGANGSRRG-VRRMATE--LKNQPGGSITLPP.RVPLPSLEHLHGALP.................          217
MND_1.GA.x.MNDGB1      .............................................KAIR..GEEVLKVCRFPAGHKAQGV-SLQFLCLRVIYGPEER..................................          172
MND_2.CM.98.CM16       .............................................NGRR....GPR.DPRRSRRGGSKYLVG-I-GENQSGGKITLPP.RVPFPSLERMHRTLA.................          212
MND_2.GA.x.M14         .............................................NGRK....STR.GTTG-RRSGTRTVAG-IIGTSEQRGSITLPP.RVPFPSLEHLCRTLA.................          212
MND_2.x.x.5440         .............................................HGRE....GTR.GARRSRRG-TGAMAGNVTGENQPGGPVTLPP.RVPFPSLEHMCRTLA.................          213
GRV.ET.x.GRI_677       .............................................NGIRKRSK-TFT-MAGNLGS-QGAMGRMATRHAQGSKRRSQKALWNEHANPSMELLCRGGKET.............          219
TAN.UG.x.TAN1          .............................................NGRRKRFQGKKT-M-RNLGSQ..QGAVGRMIKRHGSRTQSGSTTPF.WERTPLPSMEL.LSGRRGKEWGTNDRKGL          238
VER.DE.x.AGM3          .........HQ----Q...............R---SK-GG.T-----E-G--G---K--ARRY-SGSQDAFWARA-V--ME.LLSGGGRK...ESHSHARKG.L.................          232
VER.KE.x.9063          .........HQ--I-E...............R---KQA-R.S-DL-SK--TMG---KG--RRS-SGSQITFWQRA-L---E.LLSGGRRE...TGSPNDGEG.L.................          231
VER.KE.x.AGM155        .........HQ----K...............R---GE--R.T--L-SQ-G--G---Q-YGRRN-Q-SQTAFWPRT-I--ME.LLSGGRGE...TGKTHSGKG.I.................          232
VER.KE.x.TYO1          .........HQ----E...............R-K-GK--R.S--L-SK-G--GQ--K-Y--RS-PG-EAAFWERT-V--ME.LLSGGRRK...TWYSHDGKG.LQIL..............          235
TAL.CM.00.266          ...................................................YGKV-TQSSSTADLQTHSNTKASARKAHERSSRRNAYTL...HKRRGHWSLESLC.......RIASRP..          227
TAL.CM.01.8023         ...................................................HGKV-TQSSSANNIKAHTNTKSSIRKTHERSKGRDAHTL...HKRRGHWSLESLC.......RVTSRA..          227
MON.CM.99.L1           .............................................ALV.-YGATPTKSSAPLDLPD-SHRKTGPTGRHLGAKRRSPKTLLKGRAPWHLGTGNGAAC...............          255
MON.NG.x.NG1           EPSSTLDLP......................................RGSHRA-RSAGRHM--KRG--XA-L-GRAPWH-GS.....................................GH          257
MUS_1.CM.01.1085       ...........H-.....KLAPQPX................Q-PGTPEWSQPPCARHKKPC-K..........................................................          217
LST.CD.88.447          .............................................HVQNK-AKESRKTS..RGLWATKGAMGSMASRHHGGNKR-GQTPFPRNNFGPSLRILCQQYRVRGESLHEIHYHSK          227
LST.CD.88.485          .............................................HVQSK-AKESRKTS..RGLWTTKRAMGSMVSRHHGGNKG-GQTPFPRSHFGPSLGILCQQYRVRGGSLHEIHYHSK          227
LST.CD.88.524          .............................................HVQTKTAAQPRKAP..RGLWSSKRAMGSMASRHNGGNQGRGKTPFPRSNFGPSLGILCQQYRVRRGGLHEIHHF.S          226
LST.KE.x.lho7          .............................................HVQAE-AKKSRETT..RGLWTTKGAMGSVVSRHHGGNKR-GKTPFPRSNVGPSLAILCRQYRVRRGGLHEIHHPSQ          226
SUN.GA.98.L14          ...........H-Q.................QERASR-AT..-RR-ATE--ICT--G-YHGRN-G-SGKAF-SSNSC.PSLAILC.......GLYRVRKGRV.YESNHFAQ..........          241
SYK.KE.x.KE51          ...........D-TGF.........L-SL-AAARGTTILHS..KKCQLG-GRSNKCHCQSRTGSP-SMQAFYSGRNI.WSL-SML-GGSRSRRNTHHGLD.....................          233
SYK.KE.x.SYK173        ...........D-GGF.........L-SL-ACARNTMVLHS..KKCRVDP.K-DQCHCKGRTGSD-SIQAFYSSRNI.WSL-SILKRRGR.D.............................          221
COL.CM.x.CGU1          THQK.........................................MRVLKGTSHGCRTR..............................................................          171
```

```
                                                                               Vif CDS \
H2A.DE.x.BEN           MTDPRERVPPGNSGEETIGE..AFEWLERTIEALNREAVNHL.PRELIFQVWQRSWRYWHDEQGMSASYTKYRYLCLMQKAIFTHFKRGCTCWG...ED.MGREGLEDQGPPPPPPPGLV                113
H2A.GW.x.ALI           -AN---T-------------..-----D--V-----------.-----------------------Q-------------MY---MK----L-...GG.H-PG-W.RS-----------                112
H2A.SN.x.ST            -AG---TI------------..-----D--V--I--------.-----------------------I-------------M-I-S------L-...GG.H-PG-W.RS-----------                112
H2B.CI.x.EHO           -.----------------V--..------T-L-H---V-----.----------K--A--RE-----I-------------M-I--AK--G-LR...-G.H-PG-W.RS----------A                111
H2B.GH.86.D205         -.-----------D---V--..--A------TE---V-----.--------------A--RE-----I--------L-----M-V-YTK--R-LQ...-G.H-PG-W.RS----------A                111
H2G.CI.x.ABT96         -.-----X--E--D------..---------XTEI--I-----.---------R---E------------------L-----M-V-C-K----LQ...KG.H-PG-W.R-----------A                111
H2U.FR.96.12034        -G-----I-----E---V--..--S-----VRDI--A--Q--.--------------Q--R------E----------I---L-M-C-K--Q-L-...-G.H-PG-W.RS----------I                112
MAC.US.x.239           -S-----I------------..-----N--V-EI--------.-------------E---------P--V-------I---L-M-C-K--R-L-...-G.H-AG-W.RP----------A                112
MAC.US.x.251_1A11      -S-----I------------..-----N--V-EI--------.-------------E---------P--V-------I---L-M-C-K--R-L-...-G.H-AG-W.RP----------A                112
MAC.US.x.251_BK28      -S-----I------------..-----N--V-EI--------.-------------E---------Q--V-----------L-M-C-K--R-L-...-G.H-AG-W.RP----------A                112
SMM.SL.92.SL92B        -------I------------..-----HN-V----QT--Q--.---------R-C-E--V----Y-P--A----VQ-----M-Q--RK----R-...-G.HSQG-W.RT----------A                112
SMM.US.x.H9            -XX---XI--------XX--..X-D--H--X-EI--A-----.X--------R-X-E-----M---V----X-----I---L-M-C-K--R-L-...GE.H-AG-W.RP----------A                112
SMM.US.x.PGM53         -S-----I------------..-----N--V-EI--A-----.---------R---E--R--M---E----------I---L-V-C----R-L-...-E.H-AG-W.RS----------A                112
SMM.US.x.SIVsmH635FC   -S-----I----------E-..-----N--V-GI--A-----.-------------E-----M---E----------I---L-M-C-K--R-L-...-G.H-AG-W.RT----------A                112
SMM.US.x.SIVsmH635F_L3 -S-----I----------E-..-----N--V-GI--A-----.-------------E-----M---E----------I---L-M-C-K--R-L-...-G.H-AG-W.RT----------A                112
SMM.US.x.SIVsmH635SB10 -S-----I----------E-..-----N--V-GI--A-----.-------------E-----M---E----------I---L-M-C-K--R-L-...-G.H-AG-W.RT----------A                112
MNE.US.x.MNE027        -S-----I------------..-----N--V-EI--------.-------------E---------P--V-------I---L-M-C-K--R-L-...-G.H-AG-W.RP----------A                112
STM.US.x.STM           -S-----I--------A-E-..-----H--V-DI--------.---------R---E--------PG--V-------I---L-M-S----R-L-...-G.H-AG-W.RP----------A                112
RCM.GA.x.GAB1          -AE.....RAPE.VPTGA--AEFQP--RDML-KV-L--RL-FH-.-F--RL-RTCVEH---VHQR-LE-AA----L-----L-I-CQT--SQR.......H-.PNPRAV-ERITIL--M.                104
RCM.NG.x.NG411         -AE..G-ERVPE.APTGA-DVEFAP--H-MLTEV-L--RL-FH-.-F--RL-RTCVEH---RL-R-LE-AG----L-----L-I-CQS--SQR.......H-QGQAREA-ERIQIL--M.                108
DRL.x.x.FAO            -AERQSVERAPA.EPMGA--VELE---Q-SLLRI-Q--RL-FH-.-FL-RL-NTCMEHY--ALQL-FT-S-----L-L---M-M--QQ--S-LQGRHPPPLRPA-D.RL.--------...                113
MND_2.CM.98.CM16       -AE.....GAPE.IP-GA--VDLNT----SL-KI-Q--RL-FH-.-FL-RL-NACIEH---RHQR-L--A-----L--N--M---MQQE-P-RS....G.HP.....R.-...------M-                 99
MND_2.GA.x.M14         -AE.....RAPE.APQGA--VGLEQ----SL-QV----QL-FH-.-FL-RL-NTCVEH---RLRRTLN-A-----L-----M-V-MQQ--P-RS....G.HS.....R.-...------A                 99
MND_2.x.x.5440         -AE.....RAPE.AP-GA--VGLEQ---TSL-RI----RL-FH-.-FL-RL-NTCVEH---RHQR-LD-A-----L--H--MY--MQQ--P-RN....G.RP.....R.-...-----MA                 99
```

```
                                                                                                                frameshift in HXB2 |
SAB.SN.x.SAB1C         MASGGWLPPVGGDPPKDPPKNPREEIPGW..LETWDLPREPFDEWLRDMLQDLNSEAQCHF.PRNLLFRLWWNIVEEPAI...DHGQTRLEGW.YKYCRILQKALFVH.MKGRCCK...PKT...HPAYG   116
H1B.FR.83.HXB2         -EQA...........................P-DQGPQ---HN--TLEL-EE-KN--VR--.--IW-HG-GQH-Y-TYGD........TWA-V.EAII----QL--I-XFRIG-RHSR.IGV.....TR.    86
H1B.US.90.WEAU160      -EQA...........................P-DQGPQ---YN--TLEL-EE-K---VR--.--VW-HS-GQH-Y-TYGD........TWT-V.EAII----QL--I-.FRIG-RHSR.IGI.....TP.    85
H1N.CM.95.YBF30        -ERA...........................P-DAGPQ---YN--ALEL-EE-KN--VR--.--IW-HG-GQH-YNTYGD........TW--V.EAII----QL--I-.YRIG-QHSR.IGI.....TP.    85
H1O.BE.87.ANT70        -EQA...........................P-NQGPAK---N--ALEL-EE-KA--VR--.--PW-HA-GQY-Y-TYGD........TWV-V.MAII----QL--T-.YRIG-QHSR.IGI.....NP.    85
H1O.CM.91.MVP5180      -EQA...........................--NQGPA----N--TLEL-EE-KE--VR--.--PW-QACGQY-Y-TYGD........TW--V.MAII----QL--T-.YRIG-QHSR.IGI.....LP.    85
CPZ.CD.90.ANT          -EQA...........................PQDEGPQ---MN---LET-EEIKN--VR--.-QPT-QH-GNWVYANYGD........SWR-V.QVLIIL------T-.YRHG-AHSR.IGH.....GP.    85
CPZ.CM.01.SIVcpzCAM13  -EHA...........................P-DQGP----YN--ALEI-E--KN--VK-Y.--SW-HQ-GQH-YTTYGD........TW--V.EAII----QL----.FRVG-HHSR.IGI.....IP.    85
CPZ.CM.05.SIVcpzEK505  -EQA...........................P-DQGPQ---YN--ALEL-E--KN--LR--.--PW-HG-GQYFYNTYGD........TW--V.EAII-T--QL--I-.YRIG-QHSR.IGI.....TP.    85
CPZ.CM.05.SIVcpzLB7    -EQV...........................P-DQGPQ---YN--ALEV-EE-KQ--VR--.--PW-HQ-GQY-YDTYGD........TW--V.QAII----QL--I-.YRIG-QHSR.IGI.....TP.    85
CPZ.CM.05.SIVcpzMB66   -EQL...........................P-DQGPQ---YN--TLEI-EE-KR--VR--.--DW-HQ-GQH-YTTYGD........TWA-V.EAII----QL--I-.YRIG-HHSR.IGI.....NI.    85
CPZ.CM.05.SIVcpzMT145  -EQA...........................P-DQGPS---YEQ--L-T-EEIKQ--VR--.--PW-HA-GEH-Y-TYGD........SWA-V.QAII----QL----.FRIG-HHSR.IGI.....NP.    85
CPZ.CM.98.CAM3         -EQA...........................P-DQGP----YN--ALEL-EE-KN--VR--.--AW-HG-GQF-YDTYGD........TWA-V.EAII----QL--I-.FRLG-QHSR.IGI.....IP.    85
CPZ.CM.98.CAM5         -EQI...........................P-DQGP----YN--XLEI-EE-KT--LR--.--PW-H--GQY-Y-TYGD........TWI-V.EAII----HL----.FRLG-QHSR.IGI.....LX.    85
CPZ.GA.88.GAB1         -EQA...........................P-DQGP----YQ--ALET-EE-KN--VR--.--PW-HQ-GQF-YDTYGD........TWV-V.EAII----HL--I-.FRLG-QHSR.IGI.....LP.    85
CPZ.GA.88.GAB2         -EQA...........................P-DQGPAG--YN--ALEI-E--KN--VR--.--PW-HQ-GQL-YNNYGD........TW--V.EAII----HL----.FRIG-QHSR.IGI.....ER.    85
CPZ.TZ.01.TAN1         -EQA...........................PNDNGPQ---YT---L-I-EEIKQ--VK--.--PI-QGVGNWVFTIYGD........SW--V.QELIK---R---T-.YRHG-IHS.RIGS........    83
CPZ.US.85.CPZUS        -EQV...........................P-DQGPA---YN--AL-V-EE-KE--LR--.--PW-QG-GQY-YSTYGD........TW--V.EAII----QL--I-.YRIG-NHSR.IGISLARRTP.    90
GSN.CM.99.CN166        -ER....................IPPSHPMP..WLSRRVPTTMAIAQNA-WEI-E--EK--.S-DE-RGI-HDVT-L--D.PDWTVDQAAIACAID-I-.V-TL--R-.F-DG-FHRVN.LS....GR-P    99
GSN.CM.99.CN71         -ER....................TPPSHPLP..WISRRVPTTMAVAQSAMWEI-E--EK--.S-EE-RGI-HDVT-L--D.PDWTVDQAAIACAID-V-RV-TL--R-.FRDG-FHRYNRIV....RR-P   101
H2A.DE.x.BEN           -......................T-A-TEFPP-DGTPR-DLGSD-VIET-REIKE--LR--D--L.-IA-GYY-HNRHGD.........TLEGARELIKT--R-----.FRAG-NRSR.IG.........    87
H2A.GW.x.ALI           -......................T-A-TEFPPAGMGPHQGAR---VIEV-REIKE--LR--D--M.-IA-GGY-YTRHGD.........TLERARELINA--R---M-.FRAG-GRSR.VGQTRGR..NP    94
H2A.SN.x.ST            -......................T-A-TESPP-DRTP----G---VIET-REIK*--LK--D--L.-IT-GNY-YARHGD.........TLEGARGLI----R--LL-.FRAG-GRSR.IGQPRGR..NP    94
H2B.CI.x.EHO           -......................A-AVPEIPP-DKNPQ---WEQ-VV-V-EEIKQ--LK--D--L.-TA-GNF-YNRHGN.........TLEGAGELIKL--R---L-.FR-G-QHSR.IGQPGGG..NP    94
H2B.GH.86.D205         -......................A-AAPEIPP-NENPQ---WE--IGEI-EEIKQ--LK--D--L.-TA-GNF-YSRHGD.........TLAGAGELIK---R---L-.FRAG-QHSR.IGQSGGG..NP    94
H2G.CI.x.ABT96         -......................A-EIPP...-DEAPQ---W---VVEV-EEIKE--LK--D--L.-TA-GNY-YDRHGD.........TLEGAGELIK---R---L-.FR-D-RCSR.IGGNAGG..NP    91
H2U.FR.96.12034        -......................A....EIPP-DGAPQ---G---V-EI-EE-KE--VR--D--L.-TA-GNY-YNRHGD.........TLEGAGELIK---R-I-L-.FR-G-RHSR.IGSSRGG..NP    90
MAC.US.x.239           -..........................EERPP-NEGPQ---W---VVEV-EE-KE--LK--D--L.-TA-GNH-YNRHGD.........TLEGAGELI----R---M-.FR-G-IHSR.IGQPGGG..NP    90
MAC.US.x.251_1A11      -..........................EERPP-NEGP*---WG--VVEV-EE-KEG-LK--D--L.-TA-GNH-YNRHGD.........TLEGAGELI----R---M-.FR-G-IHSR.IGQPGGR..NP    90
MAC.US.x.251_BK28      -..........................EERPP-NEGPQ---W---VVEV-EE-KE--LK--D--L.-TA-GNH-YNRHGD.........TLEGAGELI----R---M-.FR-G-NHSR.IGQPGGG..NP    90
SMM.SL.92.SL92B        -..........................E.HAP-DETN----W---I--V-EE-KE--LK--D--L.-TA-GNYVYDTYGD.........TIEGAGEIIK-------L-.FRHG-THSR.IGQPRGP..NP    89
SMM.US.x.H9            -..........................TERPP-DEAPQ---W---XVEV-EEXKE--LNX-D--L.-XA-GNY-YDRHGD.........TLEGAGELI----G---I-.FR-G-RHSR.IGQSGGX..NP    90
SMM.US.x.PGM53         -..........................TERPP-DEAPQ---W---VVKV-EEVKE--LK--D--L.-TA-GNY-YDRHGD.........TLEGAGELI----R---I-.FR-G-AHSR.IGQSRGG..NP    90
SMM.US.x.SIVsmH635FC   -......................A-R-PEDEAPQREPWD-WVV-V-EEI....KE--LK--D--L.-TA-GNY-YDRHGD.........TLEGAGELI----R---I-.FRSG-AHSR.IGQSRGG..NP    90
SMM.US.x.SIVsmH635F_L3 -......................A-R-PEDEAPQREPWD-WVV-V-EEI....KE--LK--D--L.-TA-GNY-YDRHGD.........TLEGAGELI----R---I-.FRSG-AHSR.IGQSRGG..NP    90
SMM.US.x.SIVsmH635SB10 -......................A-R-PEDEAPQREPWD-WVV-V-EEI....KE--LK--D--L.-TA-GNY-YDRHGD.........TLEGAGELI----R---I-.FRSG-AHSR.IGQSRGG..NP    90
MNE.US.x.MNE027        -..........................EERPP-DEGPQ---W---VVEV-EE-KE--LK--D--L.-TA-GNY-YDRHGD.........TLEGAGELIK---R---M-.FR-G-NHSR.IGQSGGG..NP    90
STM.US.x.STM           -..........................THRPP-DEGPQ---W---VVEV-EEIKQ--LR--D--L.-SA-GNY-YDRHGD.........TLEGAGELIK---R---M-.FR-G-RHSR.IGQPGGG..NP    90
DEB.CM.99.CM40         -E....................-YPPSHP..PHFTSRTIPMTRLG-QQAM----E--LK--.T-EE--GV-NHC-DL--Q.PDWT-EQAWAASVID-IK-V--M-WL-.LREA-FHREREA-RGYPNIRP   105
DEB.CM.99.CM5          -E....................-YPPSHP..PHFTSRTVPMTRLA-QQAM----E--LK--.T-EE-WGV-NHC-DL--Q.PDWT-EQAWAASVID-IK-V-RM-WL-.LREA-FHREREA-RRYPNIRP   105
RCM.GA.x.GAB1          -EL...........................PP-DEGPQ---Y----M-T-IE-QE--KK--TYAL.-TQIGDYVY-QHGD.........SIEGVQAMI-L--R---L-.FRNG-AGSR.IG-SRG..SNP    89
RCM.NG.x.NG411         -EM...........................LP-DEEPQ---Y----M-T-IEIQE--KK--T-EL.-TQEGNY-Y-QHGD.........SLEGVKAMITL-N----L-.FRHG-EGSR.IGGARGG..NP    89
DRL.x.x.FAO            -ER...........................PP-DEGP----Y----V-T-VEIQE--LK--DR-L.-HNVGSWVY-RYGD.........SLEGVHQLIIL--R---L-.FRYG-QESR.LG.PAGR.WNP    89
MND_1.GA.x.MNDGB1      -GQKR...............DEQVS.......-DQGP----YNQ--A-TMEEIKE--RK--.-.LIILNAVSEYCV.......QNTGSEE-AC.E-FITLMNR-IW--.LAQG-DGTFRE..RR.PQLPP    94
MND_2.CM.98.CM16       -EQ...........................PP-DEAPQ---YN---I-T-AEVQE--LK--DR-L.-HAVGSW-YKTYGD.........TLEGVRQLIS------SM-.YRYG-RESR.IGQEGGG.C-P    90
MND_2.GA.x.M14         -EQ...........................PP-DEAPQ---YN---I-T-AEIQE--LK--DR-L.-HAVGSW-Y-TYGD.........TLEGVQ-LITL--R---L-.FRHG-RESR.IGQEGGK.YNP    90
MND_2.x.x.5440         -EQ...........................PP-DEAPQ-G-YNK--VGT-AEIQE--LK--DR-L.-HAVGSWVY-QQGD.........TLEGVQQLIS---R---L-.FRYG-RGSR.IGQEGGR.YNP    90
GRV.ET.x.GRI_677       ----RDPREPLPGW.................--I---D---W----Q---R---E--RR--.GM-M-I-V-N.YCV-EGRR....HN-PWNEIG---Y--V--SM---.FRCG-RRRG.-FSPYEERRN-   105
TAN.UG.x.TAN1          --E-R...............DS--RR---LEIWDLS..---W--*----VAE--Q---R--.G-E---QV-N.FCQ-EGER...N-APMI-RA.-R-Y-LV-------.FRCG-RRRT.-FEPYEERRN-   105
VER.DE.x.AGM3          ----R...............D---AR--E..--I---S---W--------E-I-Q--KM--.G-E---QV-N.YCQ-EGER...NRTPM--RA.---YKLV-------.FRCG-RRRQ.-FE.....P-.    99
VER.KE.x.9063          ----R...............G---NR--E..V-I---N---W--------E---Q--KL--.G-E---QV-N.YCQ-EGER...R-APMT-RA.-R-Y-LV-------.FRCG-RRRQ.-FE.....P-.    99
VER.KE.x.AGM155        ----R...............D----R--G..--I---S---W-------VEEI-N--KL--.G-E--YQV-N.YCQ-EGER...Q-RPIA-RA.---Y-LV-------.FRCG-RRRQ.-FE.....P-.    99
VER.KE.x.TYO1          ----R...............D---AR--E..V-I---S---W-K----------Q--RL--.G-E---QV-N.YCQ-EGER...--TPMM-RA.---Y-LV-------.FRCG-RRRQ.-FE.....P-.    99
TAL.CM.00.266          -ERL......................-PSHP-PLTSRLIPTPRPQ-ERLMREVRD-MLI--.T-DEVIG.V--HCV-LPADPDWT-EQAWAASVIDFA-KGNQM-LL-.FQQG-YHEQQRR-PHYPNIRP   105
TAL.CM.01.8023         -ERL......................-PSHP-PITSRLIPTPRAQ-ERLMREVRD-MLI--.T-EEVIG.V--HCV-LPAEPDWT-EQAWAASVIDFA-KGNEM--L-.FQQG-FHEQQRR-AHYPNIRP   105
MON.CM.99.L1           -V.....................................PIERQA-QAAIWE--E--LK--.S-EE-RG.I-EQ-T-LPADPAWNADQAWAACAID-T-WV-TI-YR-.YREG-YHRYAEQIRRYPVLRP    90
MON.NG.x.NG1           ....................................XVVPIERAA-QAAIW---E--LX--.SKEE-RGI-DQVT-L--DPG.WNADQAWAAXAID-T-RV-TXI-R-.-RDG-YHRHARVXRRYPXLRP    91
MUS_1.CM.01.1085       -.E....SLPPSH-.................APWHSRVVPTTMRQAQEAMWEV-E--EK--.S-EE-RGI-NEVT-L--DPD.WTVDQAAIACAID-V-RV-TLV-R-.FRDG-YHRYNRVMRRYPTIRP   105
LST.CD.88.447          -SRTR...............QQ-SPEKPP...-DFGP----WGQ--A-TMEEIKE--RR--.-.ETILAQVSEYCV.......SSTGSEEKAC.M-FIT-VNR---I-.LP.T-PSYA.AAR..AGT-HP    97
LST.CD.88.485          -SRAR...............QQ-SPERPP...-DFGP----WGQ--A-TMEEIKE--RR--.-.EAILAQVSEYCV.......SSTGSEE-AC.MRFIT-VNR---I-.LP.T-PSYA.ATR..AGT-HP    97
LST.CD.88.524          -SRQR...............QQ-SPERPP...-DFGP----WAQ--A-TMEEIKE--RR--.-.EAILAQVSEYCV.......SSTGSEE-AC.MRFITLVNR--YL-.LP.Q-PSYA.AAR..AGT-HP    97
LST.KE.x.lho7          -SRQR...............QQ-SPERPP...-DFGP----WGQ--A-TMEEIKE--RR--.-.EVMLAQVSQYCV.......-STGSEE-AC.M-FITLVNR---L-.LP.T-PSYA.AAR..AGT-HP    97
SUN.GA.98.L14          ---.................RREPVEQPP...-DEGP-----AQ--A-TMEEIRD--ER--.-AAIVA.QVLQYCA.......-STGSEK--C.MRAITL-NR---I-.LP.M-PSYIRIRSGSGNVNPR    98
SYK.KE.x.KE51          --QAF.........FRVYQRQ-G-PLFYIP...RNVNWD-AGVISATAKAEQVAQ--CK--.TAEEI-.GV--QCL-VEAGPEETPTMAWIRCMLDMH-A-NFM--.EHFAAG-PQRTK.YA.....RH.   108
SYK.KE.x.SYK173        --EAF.........FNPSQHVQGTPWFFIP...RNVELTP.NVINVTVKAELVVT--SK--.TPQEIY.GV--QSLNEEA.GT-SPTMAW-RTMLDMV-A-NLM--E-.FAAG-PQRT.RYA.....RH.   106
COL.CM.x.CGU1          -R.................XPE-DFT..................--Q-QVENITR--RR--PLTHISVIM-..GCTRR....HW.-RPW.TAG.LA-V--MEC--YL-.VHAGQ-RFP.STQNP......    78
```

```
SAB.SN.x.SAB1C         PGAGGPPPGLGGASGGAASAAPGL..........        140
H1B.FR.83.HXB2         .....QRRARN---RS..................         97
H1B.US.90.WEAU160      .....QRRARN---RS..................         96
H1N.CM.95.YBF30        .....QRRRN.-T-RS..................         95
H1O.BE.87.ANT70        ....RGRGRRN-S-RS..................         97
H1O.CM.91.MVP5180      SNTR-RGRRN.-S-RS..................        100
CPZ.CD.90.ANT          ............RRR...................         88
CPZ.CM.01.SIVcpzCAM13  .....QRRRAN--NRP..................         96
CPZ.CM.05.SIVcpzEK505  .....Q.RRRN---RS..................         95
CPZ.CM.05.SIVcpzLB7    .....QRRRRD--NRS..................         96
CPZ.CM.05.SIVcpzMB66   ...NQQRRRRN--NRS..................         98
CPZ.CM.05.SIVcpzMT145  .....QRRRRD-TNRA..................         96
CPZ.CM.98.CAM3         .....GRRRN.---RS..................         95
CPZ.CM.98.CAM5         .....ERRRN-TSRX...................         95
CPZ.GA.88.GAB1         ....QRRRSN-SNRS...................         96
CPZ.GA.88.GAB2         .....QRRRRRDG-YRS.................         97
CPZ.TZ.01.TAN1         ..................................         83
CPZ.US.85.CPZUS        .....QGRRN.-S-RS..................        100
GSN.CM.99.CN166        AIRPSRGTAPPDSNSVSHADPEQPRRPSRYRMDE        133
GSN.CM.99.CN71         VIRPLRGTAPPDSNSVPHADPEQPRRPSRYRMDE        135
H2A.DE.x.BEN           ..................................         87
H2A.GW.x.ALI           LS-IPT-RNMQ.......................        105
H2A.SN.x.ST            LS-IPT-R-MR.......................        105
H2B.CI.x.EHO           LS-IP-S...........................        101
H2B.GH.86.D205         LSTIP--...........................        101
H2G.CI.x.ABT96         LSTIP-SR-VF.......................        102
H2U.FR.96.12034        LTTVP-S...........................         97
MAC.US.x.239           LS-IP-SRSML.......................        101
MAC.US.x.251_1A11      LS-IP-SRSML.......................        101
MAC.US.x.251_BK28      LSTIP-S...........................         97
SMM.SL.92.SL92B        L-SIPSARDVL.......................        100
SMM.US.x.H9            LSTIP-XR-VL.......................        101
SMM.US.x.PGM53         LSTIP-SR-VL.......................        101
SMM.US.x.SIVsmH635FC   LSTIP-SRAML.......................        101
SMM.US.x.SIVsmH635F_L3 LSTIP-SRAML.......................        101
SMM.US.x.SIVsmH635SB10 LSTIP-SRAML.......................        101
MNE.US.x.MNE027        LSTIP-SRSML.......................        101
STM.US.x.STM           LATIP-TR-VL.......................        101
DEB.CM.99.CM40         LTGRNREVRD-E......................        117
DEB.CM.99.CM5          LTGRNREVRD-E......................        117
RCM.GA.x.GAB1          LRSIPQTRNIM.......................        100
RCM.NG.x.NG411         LRSILHSRNIL.......................        100
DRL.x.x.FAO            FRTFPRANQQL.......................        100
MND_1.GA.x.MNDGB1      S-FRPRGDR-........................        104
MND_2.CM.98.CM16       LRSFPRSDNPL.......................        101
MND_2.GA.x.M14         LRSFPR-NNPL.......................        101
MND_2.x.x.5440         LRSFPS-NNPL.......................        101
GRV.ET.x.GRI_677       Q-G-A---PP-L-.....................        118
TAN.UG.x.TAN1          V-G-RDGREPPPGLA...................        120
VER.DE.x.AGM3          .....EERRD-QGG-R-GRVP---D.........        119
VER.KE.x.9063          .....EERRD-QGG-R-NR-P---D.........        119
VER.KE.x.AGM155        .....EERRN-QGG-RPGRVP---D.........        119
VER.KE.x.TYO1          .....EERRD-QGG-R-NRVP---E.........        119
TAL.CM.00.266          LSGR-ERTMQ........................        115
TAL.CM.01.8023         LSGRR.NTMQ........................        114
MON.CM.99.L1           MRGTA-G-TSSVPQADPDNPRRPSRYRMDE....        120
MON.NG.x.NG1           LRGTA-G-TSSVPHADPERPLHPSRYHHDE....        121
MUS_1.CM.01.1085       LRGTQ---SNSTPNADPVPPLSPSRYRMDE....        135
LST.CD.88.447          TE-APR-TKRDIQR-RY.................        114
LST.CD.88.485          TE-APK-TERDIQR-RY.................        114
LST.CD.88.524          TE-APR-T-RDIQR-HY.................        114
LST.KE.x.lho7          TE-APR-NEGDIQR-SD.................        114
SUN.GA.98.L14          -PTRPR-GQGDIRR-LQD................        116
SYK.KE.x.KE51          .....RGYPHPS......................        115
SYK.KE.x.SYK173        .....RGYPHPS......................        113
COL.CM.x.CGU1          ..................................         78
```

```
                                                                          Cys-rich domain   Lys acetylated
                       / Tat start                                      *   * *  **  *  *   |                             exon 1 \/ exon2
SAB.SN.x.SAB1C         MD.QEQEAR..........PQVWEELQEELH.............RPLQACDNTCFCKVCCFHCILCFHKKALGIRYYVPRPR..RASKK.......ISHNQVSLHN*SISTWTRDSQTKKKSKAAVG...   94
H1B.FR.83.HXB2         -EPVDP............................RLEPWKHPGSQ-KT--.TN-Y--K-----QV--IT-----S-GRKKR-..QRRRAHQN...SQT-QASLSKQ.PT-QPRG-PTGP-E*-KK-ERET   95
H1B.US.90.WEAU160      -EPVDH............................RLEPWKHPGSQ-KT--.TN-Y--R-----QV--MT-G---S-GRKKR-..QRRRSPQ...NSQT-QDSLSKQ.PT-QPRG-PTGP-E--KK-ERET   95
H1N.CM.95.YBF30        -EPVDP............................RLEPWNHPGSQ-KT--.-N-Y--R--Y--LY--T--G---S-GRKKRS..QRRRTPQ...SSK--QDLIPEQ.PL-QQQG-QTGQ--Q-E-LESKT   95
H1O.BE.87.ANT70        --PVDP............................EVPPWHHPGSQ-QIP-.-N-Y--R--Y--YV--VR-G---S-GRKKRG..-PAAA.S....HPD-KDPVPKQ.-PTITK-KQERQEEQEEE-EKKA   93
H1O.CM.91.MVP5180      --PVDP............................EMPPWHHPGSK-QTP-.-N-Y--R--Y--YV--T--G---SHGRKKR-..-PAAAAS....YPDNKDPVPEQ.-L-HTG-KQKRQEEQEKK-EKET   94
CPZ.CD.90.ANT          --PVDA............................ETPPWLHPPAT-ATP-.-N-Y--C--Y--P---T--G---S-GRK-RAR.-NRRTTAE..SSENNQDPVSK*.-LPKTS-IQSSQ--*EKK-EEKT   97
CPZ.CM.05.SIVcpzEK505  -EPVDP............................ELEPWNHPGSQ-KT--.-N-H-----Y--VY--T--G---S-GRKKRS..QRRRTPQS...NK--QDPLPKQ.PL-QRLG-QTGQ-EQ-KTLES*T   95
CPZ.CM.05.SIVcpzLB7    -EPIDP............................SLDPWNHPGSQ-KTP-.-N----K--Y--PF--T--G---S-GRKKW-..SRRRTSKS...SPHYQDPISKQ.PFPPSRG--LSPEEQ-KK-ESKA   95
CPZ.CM.05.SIVcpzMB66   -EPIDP............................NLEPWNHPGSQ-KT--.-N-Y--Q--Y--Q---T--G---S-GRRKRK..QRRRTSES...SQN-QDPVPKQ.PL-QPGGIETGQ----KE-ESQT   95
CPZ.CM.05.SIVcpzMT145  -EPI-P............................DLEPWNHPGSK-RTE-.-K----K--Y------T--G---S-GRKKRG..-RKPASSN...QD.NKDPLRKQ.PL-QPI--QTGSE-Q-KE-ERTS   94
CPZ.CM.98.CAM3         -EPVDP............................DLEPWLHPGSQ-RT--.-N-Y--M------V--E--G---S-GRKKR-..GRRRLSAA..SSTHNQGSLQQQ.PL-RQQGNQTGQEEQ-KTLESQA   96
CPZ.CM.98.CAM5         -EPVDX............................NLEPWLHPGSQ-KT--.-N-Y--C-----MV--E--G---S-GRKKR-..GRRHSLAS..SSTHNQDSIRQQ.PL-RQQGNQTGEEEQ-KT-ESQA   96
CPZ.GA.88.GAB1         --PIDP............................DLEPWKHPGSQ-RTV-.-N-Y--A--Y---Y--T--G---S-GRKKRTT.-RRTAPAG...SKNNQDSIPKQ.PL-QSRGNKEGSE--TKE-ASKT   96
CPZ.GA.88.GAB2         --............................PIDPSLEPWNHPGSQ-KT--.-S-Y--R--Y--M---TS-G---S-GRKKRG..-RRTALKG...SK-NQDPLRQQ.PL-QQQG-QKCQ-E-EEE-ESKT   95
CPZ.TZ.01.TAN1         -N............................PIDPQVAPWEHPGAA-ETP-.TN-Y--K-----PV--T------S-GRK-RG..-K-AV.HS....TNNQDPVRQQ.-LPKRS-IQNSQEE-QEE-EAET   93
CPZ.US.85.CPZUS        --PVDP............................NIEPWKQPGSQ-KT--.-N-Y--K-----VV--T--G---S-GRKKRGR.-RRTTAKS...SADNQDPLRKQ.PF-QQLGNQTGQ-EQ-K-LEKQT   96
GSN.CM.99.CN166        -PILSSLGDPQDIAWMNRD-EECLLP...........NWQQPGAA-ATP-S.A-Y--K-AL--Q---LR-G--LSLHGRSRKRK-PKATEPD.PSVS-NQDSA*VQ.-L-HKQWNAVRQTQP-K-LARKA  116
GSN.CM.99.CN71         -LILSSPEDLQDTAWMNRD-AECLLP...........KWQQPGAA-NTP-S.A-Y--R-A---Q---LR-G--LSLHGRSRKRK-SEASEPA.PSVPTNQDPIQIQ.-LPHSKWNAIRQTEQEKT-ARKA  116
H2A.DE.x.BEN           -ETPLKAPESSLKPYNEPSSCTS-RDVTAQELAKQGEELLAQLH---EP-T-K-Y--R-S---Q---S--G---S-ERKGR-..-RTPRK.....TKTPSPSAPDK.----R-G---PT-EQ-KTSEATV  122
H2A.GW.x.ALI           -ETPLK-PGSSLMPYNEPSSCTS-QDVAVQELAKQGEEILSQLY---ET-N---Y--E--Y--Q---LN-G---W-DRKGR-..-R-P-K.....IKA-SSSASDK.----R--N--PEE-Q-KTLETTL  122
H2A.SN.x.ST            -ETPLKAPEGSLGSYNEPSSCTS-QDAAAQGLVSPGDEILYQLYQ--E----K-Y--K--Y--QM--LN-G---W-ERKGR-..-RTP-K.....TKA-SSSASDK.----R-GN--PE--Q-KTLETAL  122
H2B.CI.x.EHO           -EIPLK-QESSLNSSSGHSSSTS-GVANTQGLDNRGEEILSQLY---K--S---Y--K-SY--Q---L--G---C-ERS-K-..SSKRAK.......TTTSSAPNE.-L-AR-G---PT--Q-KE-ETTR  120
H2B.GH.86.D205         -EIPL--QESSLKSSNEPSSSTS-PVVNTQGLDNQGEEILSQLY---K------Y--K--Y--Q---L--G---C-DRS-R-..S-KRAK.......TTAPSAPDK.-L-AR-G---PT--Q-KE-ETTG  120
H2G.CI.x.ABT96         -ETPSR-QENSLKSCRELSSCTF--TVDAPGLEGTQAEILYQLY---E--S-K-Y--K--Y--Q---LN-G--VC-EQ--R-.RTPK-TK.......ANSFSASNE.---AR--NX-P---*-EK-ETKV  121
H2U.FR.96.12034        -EIPLK-QESSLKSCKEQSSSTS-GDVGTQGSAAAGEEILSQLY---E--S-K-Y--K--Y--Q---L--G---C-EQ--R-..SPK-VK.......AYSSAPSNK.-L-NG--N--P-E-*-K-LETAV  120
MAC.US.x.239           -ETPLR-QENSLESSNERSSCIS-ADASTPESANLGEEILSQLY---E--Y---Y--K--Y--QF--L--G---C-EQS-K-..-RTP-K.....AKANTSSASNK.P--NR--HC-PE-AK-ET-EKAV  122
MAC.US.x.251_1A11      -ETPLR-QENSLESSSGHSSCTS-AAASTPELANLGEEILSQLY---E--Y---Y--K--Y--QF--L--G---C-EQS-K-..-RTP-K.....AKANTSSASNK.P--NR--HC-PE-AK-ET-EKAV  122
MAC.US.x.251_BK28      -ETPLR-QENSLESSNERSSCIL-ADATTPESANLGEEILSQLY---E--Y---Y--K--Y--QF--L--G---C-EQS-K-..-RTP-K.....AKANTSSASNK.L-PNR--HC-PE-AK-ET-EKAV  122
SMM.SL.92.SL92B        --........................APIQGLDSQGDQIPWDLY---ET-Y-S----K-----Q---LR-G---T-AK--R-..VKK-AK.......ANPFDTSNQ.---SG--NC-P--EQ-EK-ETEV   96
SMM.US.x.H9            -ETPLK-QESSLESSREHSSSIS-VDADTPESASLEXEILSQLY-XVE--Y-KXY--K--Y--QH--X--G---C-EQQ-R-..TPK-TK.......ANTFSASDK.-LXXRA-NC-P--EK-ET-EXEV  120
SMM.US.x.PGM53         -ETPLK-QENSLGSSNERSSYTS--DVLTPESAKVEEEILCQLY---E--S---Y--K--Y--Q---L--G---C-EQQ-R-..-RTP-K.....AKANTFSAPNE.-L--RA-N--P--EK-ET-ETEV  122
SMM.US.x.SIVsmH635FC   -ETPLK-QENSLESYREHSSSIS-VDVPTPESANLEEEILSQLY---EP-Y-K-Y--R--Y--QH--L--G---C-EQH-R-..TPK-TK.......TNPLPASNK.-L--R--NR-P--EK-EK-ETEV  120
SMM.US.x.SIVsmH635F_L3 -ETPLK-QENSLESYREHSSSIS-VDVPTPESANLEEEILSQLY---EP-Y-K-Y--R--Y--QH--L--G---C-EQH-R-..TPK-TK.......TNPLPASNK.-L--R--NR-P--EK-EK-ETEV  120
SMM.US.x.SIVsmH635SB10 -ETPLK-QENSLESYREHSSSIS-VDVPTPESANLEEEILSQLY---EP-Y-K-Y--R--Y--QH--L--G---C-EQH-R-..TPK-TK.......TNPLPASNK.-L--R--NR-P--EK-EK-ETEV  120
MNE.US.x.MNE027        -ETPLR-QENSLKSSNGRSSCTS-AAATT-ESANLEEEILSQLY---E--Y---Y--K--Y--QF--L--G---C-EQS-R-..-RTP-K.....VKANTSSASNK.P--DR-KHC-P--EQ-ET-ETAV  122
STM.US.x.STM           -ETPLK-QESSLRSSSEPSSCTS-AVAATPGLANQEEEILWQLY---EE-C-K----K--Y--Q---VT-G---T-ERS-R-..VKK-AK.......TYPISASNK.-L--RA-N--P--EQ-KE-ETEV  120
DEB.CM.99.CM40         -ESIDP..........................FEERANTPHP..........A-M--A-IL--Q---IQ-G---S-GPR-RKRKCRREYTPAAESNQPDQDFVPEQ.PLPQRR-KCFDSQ-KE-E-EETA   93
DEB.CM.99.CM5          -EEIDP..........................FKETANTPHP..........N-M--A-TM--Q---MQ-G---Y-GSR-RKRR-G-SFAANDKDHQEFIPEQ....PLPQRRWKCFNSEEKETK-EKAS   90
RCM.GA.x.GAB1          --.......................V-GVGLEHPEEVILYDPFRKRETS-.---Y--K--Y--Q---LQ-G---N-ASRAR-RRSKEENK......ADKFP-PNQ..----TRGNRKLQE-KEKT-EKKV   98
RCM.NG.x.NG411         --.......................VKGAGLEEREEEILLDPFFTRETS-.-S-Y--R-VY--Q---LQ-----N-ASR-R-..-GRPGNKK.KNQ-NPFDT-EQ..-L-AIG-N--S--EK-E--EKAV  101
DRL.x.x.FAO            --.......................ARKVDLDQQDAGTHFEPFQGRTNN-.---W--Y-AY--Q---LQ-G-----N-S---..KRTATVV..SRVSETGDTPVAR.RPT-TDSQTEK-E---TEASGTS  101
MND_1.GA.x.MNDGB1      -E.............PSGKEDHNCPP-DSGQEEIDYKQLLEEYYQ-----E-K-W--K-----M---Q--G-----H-Y-K-....VPGT..NKKIPGSGEEAIRR.A-DLSFHRTASRTYTANGQTTEK  110
MND_2.CM.98.CM16       --......................AGKAVSDKKEGDVTPYDPFRDRTTP-.D--Y--R--Y--Q---LQ----VH-HAY-N-..-SRQRLL..EKISEDSAAPVGR.V-ALFLMS-RETISVPTTSNSQA  102
MND_2.GA.x.M14         --......................AGKA--DKKEANITPLDPFQGRTTP-.-K-Y--K--Y--QC--LQ-----H-H-Y-T-..-TR-RLL..GKISQDSEAAVGR.V-ALFLMS*RETISDL-TSDSQT  102
MND_2.x.x.5440         --......................VGEVASDKKEEDITHFDPFRARTTP-.-K-------Y--Q---LQ----VH-H-Y-Q-..-PRRGLL..EKVP-NS-AA-GR.D-ALFLMS-RETISGPTTSNGQT  102
GRV.ET.x.GRI_677       --..KE-EP..............HP-LQD--R.............---P-T-K-Y--K--Y--E---LQ-G--V--H-S-K-..-KTSTQ.......DNQDPIRQQ.D-ILSP-SIS-VQRNGQTTEEGK   91
TAN.UG.x.TAN1          -E..SEGDG............MA-S-LQD--R.............--TP-T-K----R--Y--QV--LQ-G---T-H-S-I-..-PK-NH......SN-QNLVSQQ.VTFLCLLR-ISAWGGNSQTTQEE   94
VER.DE.x.AGM3          --KG-D-QGAYHQDLI........................EQLKA--KR-T-K-Y--C--Y--Q---LQ-G--VT-HA--I-..-KKIAP.......LDRFPEQKQ.----RG-----TQ-GQEK-ETSA   96
VER.KE.x.9063          --KG-E-RTVLHQDLI........................RQYKK--TT-R-K----K--Y--Q---LQ-G--VT-HA--T-..-KKSVQ.......PNRLSQQDQ.----RG--G-ATQE-QKK-ERET   96
VER.KE.x.AGM155        --KG-EDQDVSHQDLI........................KQYRK--ET-T-K----K--Y--QF--LR-G---T-HAF-T-..-KKIAS.......ADRIPVPQQ.---IRG-----TQE-QKK-EEQA   96
VER.KE.x.TYO1          --KG-A-QIVSHQDLS........................EDYQK---T-K-K----K--Y--Q---LQ-G--VT-HA--T-..-KKIRS.......LNLAPLQ-Q.----KWGRDGQTTPTSQEKVETT   96
TAL.CM.00.266          -..SSK-EL..............RTTPISDPFQEEGRGP..........-NK.-Y--K--Y--Q---LQ-G-----VPD-RFRK-RRSTPSSNQEDKDPVSKQ....PL-RA-GIQKRQEETLQK-ESKA   99
TAL.CM.01.8023         -..SSK-EL..............PTTPTYDPFPAGGTP...........-NK.-Y--K--Y--Q---LQ-G-----VPD-RIRK-RRSTSSSNQKDKDPVSKQ....PL-RA-GNQECQ--KVQK-ESPT   98
MON.CM.99.L1           -....M-PV.............DPD-PK-Q-PPATPRTP..........-T-.---R------QI--LT-G---S-GRK-K-RRAT-PVPGLSSSKNPARKQ*...X-L-HP-----RPTEQAQ--ATAA   99
MON.NG.x.NG1           -..NSAXIV.............NPA-PP-Q-PSATPLTP..........-N-.----I-AL-XQV--LQ-G---S-GRG-K-RRGPRHPTIGDRSYQDNPG-K....-LPKSLGNQERQ--QEK--ESKA  100
MUS_1.CM.01.1085       --.PSV-EL......................PKEQRPG.....AA-ATP-TS.----K-A---V---Q------S-GRRFRKRS--DSS.....LPSHSDP-QIK..-L-DRR--QKRQ-EQ-ET-EAEA   94
LST.CD.88.447          -..-Q--QE..............QLTRQKQ-QDPLKETYKEAVTK--S--N-K-C--K---------QQ-G-------H-R-..KRVAQVF......EDK-DPVDG.EPR-ARQRRRDRERQRKHLHQLR  105
LST.CD.88.485          -..-Q--QE..............QLTRQKQ-QNPLKEIYKEAVTK--SV---K-C--R---------QQ-G-------H-R-..-RVTQVF......EDK-DPVDG.EPR-ARQRRRDRERQRRHLHQLR  105
LST.CD.88.524          -..-QP-QE..............QHT-QKQ-LDQLEEIYKEAITD---S-Q-K-H------------QQ---------H-T-..-RVTYIL......ETE-DPVDR.EPR-ARQRRRDRERQRKHLHQLR  105
LST.KE.x.lho7          -..-QP-QE..............QPI-QKQ-QDLMKETYNEAVTKA----Q---C--K---------QQ-G--LH---H-K-..-RVTSIL......ETQ-DPVDR.EPR-ARQRRRERSRYRDYLHQLR  105
SUN.GA.98.L14          -ST-GHQQDQDQGKGTLEEAYKTN......................-E----K-W-RR-----Q---LQ-G---H---Y-G-..QHQSIS......KDKQEAVRR..KPAPNPQTKKK--RQRKKESTST   98
SYK.KE.x.KE51          -STEDKVRETQGIPTSFL...................EGTFLSNG-QTP-.SR-Y--K--Y--YR--LQ-G---T-VRE-K-RHARTTA.....EDFAAHSSTTR..-R-PLPSTTRPQQGQ-KE.....   98
SYK.KE.x.SYK173        -SSTD-ICQTQRVPPSFL...................EGTFLEKG-PTP-.-K----N--Y--Q---LQ-G---T-AR--K-..A-RSI.......SEDDSAPTG..TLPRAG-TQANPQTK-K--ETEK   99
COL.CM.x.CGU1          -SLSLY.....................................AK--E--NDK-W--A-----Q---LS-G-----..KKNK...SATQ...............TQ.TNVSSQIRG-DREWEQPDNAEEI   72
```

```
SAB.SN.x.SAB1C         .........................................................................................................................      94
H1B.FR.83.HXB2         ETDPFD...................................................................................................................     101
H1B.US.90.WEAU160      ETDPED...................................................................................................................     101
H1N.CM.95.YBF30        EADPCD...................................................................................................................     101
H1O.BE.87.ANT70        GPGGYPRRKGSCHCCTRTSEQ....................................................................................................     114
H1O.CM.91.MVP5180      GPSGQPCHQDSCNSCTRISGQ....................................................................................................     115
CPZ.CD.90.ANT          GSGGGPCSTGSSHSCGWT.......................................................................................................     115
CPZ.CM.05.SIVcpzEK505  TADPCA...................................................................................................................     101
CPZ.CM.05.SIVcpzLB7    QTDQ.....................................................................................................................      99
CPZ.CM.05.SIVcpzMB66   TSDQFA...................................................................................................................     101
CPZ.CM.05.SIVcpzMT145  EADQSDW..................................................................................................................     101
CPZ.CM.98.CAM3         KADPFA...................................................................................................................     102
CPZ.CM.98.CAM5         EADPRDWSAGS..............................................................................................................     107
CPZ.GA.88.GAB1         EADQ.....................................................................................................................     100
CPZ.GA.88.GAB2         EADL.....................................................................................................................      99
CPZ.TZ.01.TAN1         TSGGRPRQQDSSVSSGRTSGTSSSGYTRPFKTSSGSSGSACKH..............................................................................     136
CPZ.US.85.CPZUS        TSDQRD...................................................................................................................     102
GSN.CM.99.CN166        RSGP.....................................................................................................................     120
GSN.CM.99.CN71         TSGPSDISQDSRVSCGTT.......................................................................................................     134
H2A.DE.x.BEN           VTTCGLGQ.................................................................................................................     130
H2A.GW.x.ALI           GTDCGPGRSHIYIS...........................................................................................................     136
H2A.SN.x.ST            ETIGGPGR.................................................................................................................     130
H2B.CI.x.EHO           ATDLGPGRSNTSTSRFAN.......................................................................................................     138
H2B.GH.86.D205         TTDPGPGR.................................................................................................................     128
H2G.CI.x.ABT96         ATDLGLGR.................................................................................................................     129
H2U.FR.96.12034        ATDLGLGR.................................................................................................................     128
MAC.US.x.239           ATAPGLGR.................................................................................................................     130
MAC.US.x.251_1A11      ATAPGLGR.................................................................................................................     130
MAC.US.x.251_BK28      ATAPGLGR.................................................................................................................     130
SMM.SL.92.SL92B        DTDCGLGR.................................................................................................................     104
SMM.US.x.H9            ATDLGLXR.................................................................................................................     128
SMM.US.x.PGM53         ATDLGLGR.................................................................................................................     130
SMM.US.x.SIVsmH635FC   AADLGLGR.................................................................................................................     128
SMM.US.x.SIVsmH635F_L3 AADLGLGR.................................................................................................................     128
SMM.US.x.SIVsmH635SB10 AADLGLGR.................................................................................................................     128
MNE.US.x.MNE027        ATVPGLGR.................................................................................................................     130
STM.US.x.STM           ESTPGLGK.................................................................................................................     128
DEB.CM.99.CM40         GPGGQHCREDSSVSSGRISNNC...................................................................................................     115
DEB.CM.99.CM5          RAGGRHRPEDS..............................................................................................................     101
RCM.GA.x.GAB1          ATSTTIG..................................................................................................................     105
RCM.NG.x.NG411         DTNITTGREDIFIS...........................................................................................................     115
DRL.x.x.FAO            DQNPPSQDPPKLGAKKQ........................................................................................................     118
MND_1.GA.x.MNDGB1      KKATA....................................................................................................................     115
MND_2.CM.98.CM16       KKEKKSPTQKNRVPNQTAPK.....................................................................................................     122
MND_2.GA.x.M14         KKEKESST.................................................................................................................     110
MND_2.x.x.5440         EKEEKSQT.................................................................................................................     110
GRV.ET.x.GRI_677       TEVEKAAAAN...............................................................................................................     101
TAN.UG.x.TAN1          KTKIPAAAETSRRPQ..........................................................................................................     109
VER.DE.x.AGM3          RTAPSLGRKNLAQQSGRATGASD..................................................................................................     119
VER.KE.x.9063          TTAQILGRKDLERDKREAVGANA..................................................................................................     119
VER.KE.x.AGM155        KANLRISRKNLGDETRGPVGAGN..................................................................................................     119
VER.KE.x.TYO1          AGSN.....................................................................................................................     100
TAL.CM.00.266          ATSGTDRKQDS..............................................................................................................     110
TAL.CM.01.8023         GAGGEHRLKDSLPLSGRTSTASSATVTRP............................................................................................     127
MON.CM.99.L1           TPDRQHI..................................................................................................................     106
MON.NG.x.NG1           GTGXQSMREDF..............................................................................................................     111
MUS_1.CM.01.1085       ETDP.....................................................................................................................      98
LST.CD.88.447          TVQERIFHSTVERGLERAFTRLTVCDSPEVDQGLWNSSPQPCVAVPFVARPFSDPFLPTWATSSESLQQTGGGKRSEDCELELEQDQKEQKRHISSSCKKPQTGTKEEMEIQIQRKKWVSQ     226
LST.CD.88.485          TVQERIFHSTVERGLERAFTGLTVCDSPEVDQGIWNSSPQPCVAVPFVARPLSDPFLPTWATSLESQQRMDGGKRSEDCALELEQDQKEQKRHISSSCERHQTGTKEKMEIQIQREKWVSQ     226
LST.CD.88.524          AVQERIFQATVERGLERAFTGLTVCDSPEVAEGVGNTAPHPRVAIPHLAGTYSDPFLPPWATSLANPQQIPGGECSENCELDQEQEQKEQRRAIKN.........................     201
LST.KE.x.lho7          AVQERIFQATVERGLERAFTRLAVSDSPEVAQGRGNTPPITSVAEPQLAVAFVDPFLPKWATPLADQQQMDGGKRSEDSELAQGEMQKEQRTVIEH.........................     201
SUN.GA.98.L14          .........................................................................................................................      98
SYK.KE.x.KE51          .........................................................................................................................      98
SYK.KE.x.SYK173        DSTSLPSAENL..............................................................................................................     110
COL.CM.x.CGU1          HREQENSSSRFWRLMEEWWSSLLRHCSNGILTVLPQHPPTTSDKRRSAWQTLRYLLVPRGPSLQILETLQRWLRSAARGWRRAPEYLRGWIYDRPQGPA......................     171
```

```
                                              exon 1 \/ exon2
SAB.SN.x.SAB1C          MSLGQEE..LLRR.......FRIIKFLYTTNPYPPGQGT............ARQRRRARQRWAKQRQQVIHLAERILETPVSQIDHLAQEFDQLVLDNLQQPPSLPPGHPTENQTANSSS..............   99
H1B.FR.83.HXB2          -AGRSGD..SDEELIR...TV-L--L--QS--P-NPE--............RQA--NR-R--RERQR-IHSIS----G-YLGRSAEPVPLQLPPLERLTLDCNEDCGTSG-Q.......GVGSPQILVESPTVL  110
H1B.US.90.WEAU160       -AGRSGD..SDEDLLK...TV-L--I--QS--P-SPE--............RQA--NR-R--RERQR-IRKISGW--N-YLGRPTEPVPLPLPPLDRLTLDCKEDCGTSG-Q.......GVGSSQILLESPAVL  110
H1N.CM.95.YBF30         -AGRSGV..NDEELLR...AV-V--I--QS----NSK--............RQA--NR-R--RARQR-IRAIS----SSCLGGPPEPVDLPLPPLDRLTLDTEEDSGTPG--S-QGTATTE.............  104
H1O.BE.87.ANT70         -AGRS-D..DQLLQ.A....IQ---I--QS--Q-SPR-S............RNA-KNR-R--RRRQA--DT--A-V-A-V-HGPQNNNIVDLPPLEQLSIRD-EGDQLSEAWTVDPRAEDN.............  102
H1O.CM.91.MVP5180       -AGRS--..DQQLLQA....IQ---I--QS--C-TPA-S............RNA-KNR-R--RRRQA--DS--T---A-V-HGSQDNNLVDLPPLEQLNIRD-EADRLPG-GTVDPGTKDN.............  103
CPZ.CD.90.ANT           -AGRE-LEGTDE.QLLK..AVK---I--QS----KPA-S............RAA--NK-R--K-RQD--EG--A-V-R-L-GGP-EHNPVDLPDLQNLSLE-LGS--TPLQNYFQTVQP-GEQSQHPLDNNQN.  118
CPZ.CM.05.SIVcpzEK505   -AGRSGV..NDEELLRA...I-V--I--QS----NDS--............RQA-KNR-R--RARQR-IRA-S----SSC-GGLQEPSTLPLPPLDRLSLN-EEDLGTSE--HPQGTATT..............  103
CPZ.CM.05.SIVcpzLB7     -AGRSGD..RDEELLRA...V----I--QS--F-HPE--............RSA--NR-R--RARHR-ISEI-G---S-CLERPEGPVPLPLPPIDRLSLNSEEDIGDCG-IHSEGTETGVGSPTISAGNSTVL  117
CPZ.CM.05.SIVcpzMB66    -AG-SGN..SDEELLRA...V----I--QS----SPE-S............RQA--NR-R--RARQR-INS-S----S-CLERPTGPVSLPLPPIERLTLDSAEDIGTGG-DPPQGTETGTGSPNTPEGHSTIL  117
CPZ.CM.05.SIVcpzMT145   -AGRS-G..DENLLRAI...---RI--ES----SQS--............R-A--NRKR--RERQR-INQIG--V-ASCLGRSSEPP-LPLPPIERLCISDLETTGDTG-IQPQGSTTNRYSPTPTERDPAVL  116
CPZ.CM.98.CAM3          -AGRS-GDDD..ACLLQ..AV----V--SS----DNK--............RQA--NR-R--RARQR-IHS-S----SSYLGRSPEPPALLLPPIERLTIDTGEDIGKPG--P-QGTATAEELQQIPVGSDNLL  118
CPZ.CM.98.CAM5          -AGRS-GDED..TRLLQ..AV----I--DS----DNK--............RQA--NR-R--RARQR-IHEIGQ-V-SSCLGRSSEPPDLLLPPIDRLTLN-VEDSANSG--H-QGTATVEGL...........  107
CPZ.GA.88.GAB1          -AGRS-PQDDA.-LLQA...VK---I--QS----SPE--............RKA--NR-R--RARQK-ISEISG-V-A-YLGRPPKPGDLELPELDKLSL-CVETTQDVG-S-TSQPQTATGETVPAGGNYSIL  118
CPZ.GA.88.GAB2          -AGRSGDEDE..-LLRA...VKA--I--DS----SNR--............RSA--NR-R--RARQR-ICEISQ--FSSLLGRHPEPDDLPLPPLERLSLG-LEPAGTIG-TGPERTETTQGASEPAERDRTIL  117
CPZ.TZ.01.TAN1          -AGRE-D..A..NLLYT...V----I--DS----S-A-S............RTA--NR-R--RQRQH--DA--S---QYRLGGPQEPPHLDIPDLSKLHLD-LDQ-ASTE-GDNQLGTQPSNSA..........  105
CPZ.US.85.CPZUS         -AGRSGGD.ADEQLLRA...V----I--AS--F-NNS--............RQA--NR-R--KNRQR-INEISQ---S-CLGRSSEPPDLPLPPIGGLTINSEEDSTNPG--H-QGTATVE.............  105
GSN.CM.99.CN166         -ADHARGNDQKLQNLIL..AC-L--T-HRS----TSS--............RSA--NR-R--RARQG--REISN----SL-GRPEEPGDLDLPDLGQLSLDS-WDREVPVGTAAES-PE-NPASETAQGSSSQS  120
GSN.CM.99.CN71          -ADPANGRDQKLQNLLL..AC-L--T--KS----TAS--............RSA--NKKR--RARQR--RQISH----SL-GRPEESGNLDLPDLGHLSLDS-EDRQVPVGAAAEGVPT-ETAGGSGSQSSSLY  120
H2A.DE.x.BEN            --ERAD-EG..LQGK.....L-LLRL-HQ-----Q-P--............-S---NR-R-RRR-WLRLVA--NKLCAV-DPPT-SPLDRAI-HLQRLTI-ELPD--TDLP-SNSNQGLAET............  103
H2A.GW.x.ALI            -TERAG-ED..LQRK.....L-L-RL-HQ-----Q-P--............-N---NR-R--RQRWG-IVA--D--FTF-DPPASSPLDRAV-HLQGLTI-DLPD--TDLP-SSESADNNQGLAET........  107
H2A.SN.x.ST             -NERA--EE..L-RK.....L-L-RL-HQ-----Q-P--............-S---NR-R--KQRWR-LVA--DK-YTF-DPPA-SPLEQTI-HLQGLTI-ELPD--TNLP-SSESID--QRLAEI........  107
H2B.CI.x.EHO            -NARERD....LQKG.....L-LLHL-HQ-----Q-P--............-S---NR-R--KQRGL-ILA--D--HPL-D-PTEGPLDLAI-RLQNLIIKDLPN--TSTPTA-ASTCIPPIWDQLVPRSNPSS  113
H2B.GH.86.D205          -TAREGD....LQKE.....L-LLHL-HQ-----Q-P--............-S---NR-R--KRRGL-ILA--D--HSFSTAPAEEPLNLAI-RLQNLTVEDLPN--PNLNQSP-TQAPGCVPPVWDQLVPRSA  113
H2G.CI.x.ABT96          --NLE--E...L-KR.....L-L-H--HQ-----Q-P-X............-N---NR-RK-KQRWL-ILA--D--YSF-XPPAREPLDLAI-QLQSLTI-DXPDS-TAVPK-LQDPP-N..............  100
H2U.FR.96.12034         --NPE-G....VQRR.....L-L-HL-HQ-----T-P--............-N---NR-R--RQRWL-ILA--DK-YTF-DPPT-SPVDLAI-QLQGLAI-DLPE--AVVP-H-RTSAP...............   98
MAC.US.x.239            --NHER-EE..L-KR.....L-L-HL-HQ-----T-P--............-N---QRKR--RRRW--LLA--D--YSF-DPPT-TPLDLAI-QLQNLAIESIPD--TNTP-ALCDPTEDSRSPQD........  107
MAC.US.x.251_1A11       --SHER-EE..L-KR.....L-L-HL-HQ-----T-P--............-N---QR-R--RRRW--LLA--D--YSF-DPPT-TPLDLAI-QLQNLAIESIPD--TNTP-ALCDPTKDSRSPQD........  107
MAC.US.x.251_BK28       --SHER-EE..L-KR.....L-L-HL-HQ--S--T-P--............-N---QR-R--RRRW--LLA--D--YSF-DPPT-TPLDLAI-QLQNLAIESIPD--TNTP-ALCDPTKGSRSPQD........  107
SMM.SL.92.SL92B         -QNPE--....L-RR.....L-L-HLIHQ-----A-P--............-S---NR-R--KRRWI-IVA--D--YSF-DPPA-TDIDLAI-QFDSLSI-SLPE--TTVP-TLRDQ-AD..............   99
SMM.US.x.H9             --SNE--....L-RR.....L-L-H--HQ----XX-P--............-N----R-R--XQRW--ILAX-D--YSF-DPPA-TPLDLAI-QLQGLAIEXLPN--ASAP-PLKDAAE-P.............  100
SMM.US.x.PGM53          --SSAG-EE..L-RR.....L-L-H--HQ-----Q-P--............-N----R-R--RQRW--ILA--D--YSF-DPPA-SPLDLAV-QLQ-LSLENIPD--TSVP-AL-E-PR-RQD...........  104
SMM.US.x.SIVsmH635FC    --STE--....L-KR.....L-L-H--HQ-----Q-P--............-N----R-R--RQRW--ILA--D--YSF-NPPT-TPLDLAI-QLQGLAIEDLPD--TSAP-TLKDAAK-S.............  100
SMM.US.x.SIVsmH635F_L3  --STE--....L-KR.....L-L-H--HQ-----Q-P--............-N----R-R--RQRW--ILA--D--YSF-NPPT-TPLDLAI-QLQGLAIEDLPD--TSAP-TLKDAAK-S.............  100
SMM.US.x.SIVsmH635SB10  --STE--....L-KR.....L-L-H--HQ-----Q-P--............-N----R-R--RQRW--ILA--D--YSF-NPPT-TPLDLAI-QLQGLAIEDLPD--TSAP-TLKDAAK-S.............  100
MNE.US.x.MNE027         --SHA--EE..L-RR.....L-L-HL-HQ-----T-PS-............-N---NR-R--RQRW--FLA--D-VYSF-DPPT-TPLDLAI-QLQNLAIESIPD--TNIP-ILHDPTE-PRSPQD........  107
STM.US.x.STM            --DQE--....L-KR.....L-L-Q--HQ-----Q-P--............-N---NR-R--RQRWN-LLA--N--YSF-DPPASTPLDLAV-QLQGLSI-DLPD--PNLPKDLQDTAEN..............   99
DEB.CM.99.CM40          -AHAGGRGSAEENTRQLLKVISL--I--QS----K-G-S............-ST---R--K-RRRQG--DSI------QSHLGGSPTT-DVALPDLSQLHLAD...............................   90
DEB.CM.99.CM5           -AHAGGRGDADQALLRMIRIIKSLYQS...----K-G-S............-ST---R--K-RRRQE--DDI-Q----AHLGRHLTASDVALPNLEELRLAD................................   87
RCM.GA.x.GAB1           -LP--D-ED-RKKIR......L-NFLYLIN.---QHG--............-NS--KK-R--RRRW---QQ------LDSTDPPVEQDLDAAIAD-QK--LNNLPE-PVDFS......................   93
RCM.NG.x.NG411          -L--E--EADQEIRRR....I-L-HLIHLS----QSG--............-N----K-RQ-RRRWT-ILQ-----FLY-DTPP-RDPLQEALSN-QQ-TLSDLPE-PVNPFSNPSSFAVDRS...........  107
DRL.x.x.FAO             --A-P-..REPPPWF..QEYL-LVTR-WQD.DQL-QT..............----K-R-A-QRRVEH-IRT-QA---QSLERRNSRDLV-GIERIHLAEERES-S.............................   86
MND_1.GA.x.MNDGB1       --T-NV....YQEL..IRRYLVVV-K--EG.LSTCLFTGLPAEPIPQT...------RK-QLRTR-A-LRE-EG---KQILDRGPDQLCQGVTNLALAEKSES-N.............................   95
MND_2.CM.98.CM16        -PTETGD..-D-DFW..KKYQK-VQR-WEG.LS-SSLCLQGRLSPSPQPQT----K-R-A-LR-TEY-IRQ-QS--W-ALEQRANQRLA-DFKGLRVAEESTS-N.............................  100
MND_2.GA.x.M14          --TEPG-..-G-DFW..EKYH--V-Q-WEG.LS-SSLCLKGRLSPTSQPQT--R-K-R-AQLRRTEH--RE-Q---WQ-LEARANKELV-GVNNLHLAEESAG-N.............................  100
MND_2.x.x.5440          --TDRGD..-DEGFW..RKYQA-V-Q-WEG.IS-SSLCLQGRLSPAPQPQT------R-A-LR-TEH--RE-QT--*-ALEARASQELA-KVDALHLAEEP-S-N.............................  100
GRV.ET.x.GRI_677        ----K--..KQA........LK---T--GSRI*SSPPDPYPQFSGT.....-----------R--Q--IDKI-G-V-N-FED-QLVAQLQEL--ENKD-VLQHLPD-P-IHQDSSGIPAVWAPATPRGSNRACS  119
TAN.UG.x.TAN1           ----S-D..-R-II......QI-RILYHSR.LLSFVSSDQYPPGEGT....------R---FQQ-QR--AA-S---FIAQQRRDPSGGESLAAAFDQLVLDNQQ-VIETLPDPPQEPHD-SSTA..........  111
VER.DE.x.AGM3           -P--S--..RRLLR.......L-AFLNKNN.----VE--............--------R--RQAQE-LRA-----WHSR-EEQLVQ-IDQLV-DQQH-AIQQLPD-PSSS.......................   89
VER.KE.x.9063           -P--P--..RKVFS.......L-AYLNR-N.----VE--............--------R--RERQR-LRF-----WNETKERQLEQTLDQLA-DVQH-AIQHLPD-PNSS.......................   89
VER.KE.x.AGM155         -P--P--..RRLLR.......L-AFLYRSN.---SVE--............--------R--KNRQK-IYA-----WG-RQEDQLVQ-IDQLV-DTQH-VTQQLPD-PSQA.......................   89
VER.KE.x.TYO1           -P--P--..RRFVR.......L-WLLYS-N.----SGEG...........T--------R--RQ-QD-IRV-V--LQ-QVYAVDRLADEAQHLAIQQLPDP-H-A.............................   84
TAL.CM.00.266           ---TGGSGRGEDPLLAAIRRI--LYQSN..----EQ--S............RSA---RYR--KARQL--EQI-S----HYLGRS-RSD-IHLPDLSQLSLADLGPCESCESQEP-GSA-HIRLPTTPPDSSSSD  119
TAL.CM.01.8023          ---TG-SGKGEDPLLAAIRRI--LYQSN..----EQ---............RSA---RYR--KARQE--ENI-S---CHYLGGPQQPP-LQLPDLEKLSLAELGP-EAPESDHP-GSA-HIRLPEN........  111
MON.CM.99.L1            -AGNGRDEERRLLSLALAAVRILQESSEX..---IPR--............-NA--NR-R--RR-QL-IASIS---FYHYLGRSQEPCPLDIPDLERLSISDLPD--ESVP-AA-PPAHTPAPTVGKP......  114
MON.NG.x.NG1            -AGAERGAAVQDILRLAIGAI---RESN..---SPS--.............RSA--NR-R--RARQE--XS-C---FNHYLDRLPRPNSLDLPAIEGLRLDDLPE--ESAQPPDPSSPRPVPTASGSQ......  113
MUS_1.CM.01.1085        -AGVS-NDRE-I-A.....CRA-QILYKSN.---TAD--............RSA--NR-R--KRKQR-IREIS----LSLLGRPPQPSDLELPDLNKLSLH-LVATSESSPPDTEGVNKIVVHAQGTD......  110
LST.CD.88.447           --T-G-KE..SP-Y......LK-S-I-W-E.-RT.................------D-E-QR-HLH-LRTVQ---FHST-ERGLER-FTRLTVCDSPEVDQGLWNSSPQPCVAVPFVARPFSDPFLPTWATSS  108
LST.CD.88.485           --T-G--E..-P-Y......LK-S-I-W-E.-RT.................------D-E-QRRHLH-LRTVQ---FHST-ERGLER-FTGLTVCDSPEVDQGIWNSSPQPCVAVPFVARPLSDPFLPTWATSL  108
LST.CD.88.524           --T-P-GE..-PTY......LKLSRI-W-E.-RT.................------D-E-QR-HLH-LRAVQ---FQAT-ERGLER-FTGLTVCDSPEVAEGVGNTAPHPRVAIPHLAGTYSDPFLPPWATSL  108
LST.KE.x.lho7           --T-NGDE..-P-Y......L-LSRI-W-E.-RT.................------E-S-YRDYLH-LRAVQ---FQAT-ERGLER-FTRLAVSDSPEVAQGRGNTPPI-SVAEPQLAVAFVDPFLPKWATPL  108
SUN.GA.98.L14           --T-DD..SINQYL......RISKRLYEG..NL-QTHRQ..................---R-D-ER-NLH-LRAVQ---FA-TLDSRLGR-F-RLSVSDSSQVAESLGNSPSTKHLPP-KFLVAPTYDFLPSWATPL  109
SYK.KE.x.KE51           --GRER-DTQEQLL.....RTLLRIAQQLE.AAA-YPLP.........QGPS-A---N-N-YRQLQA-RLYVQQ--F--IARRSAETLEQSFGELQITD...................................   84
SYK.KE.x.SYK173         -PDQGS-..Q-AVF.......LRMIAHLQE.---GPE-P............RQT---R-RQ-RQR-T-RLY-QQ--F-AIFGSRTAALEDSL-QLQISD...................................   77
COL.CM.x.CGU1           I..RRIR.....VLHK......HRLMCLHRSVPR--V--............------NT--AR--Q--ILAIDG-VV-QL-ASLQQWDPYCPASTSA-NLRQEEECLADSSVSIGTQG-KPSDS..........   99
```

```
SAB.SN.x.SAB1C         .........................................................   99
H1B.FR.83.HXB2         ESGTKE...................................................  116
H1B.US.90.WEAU160      EPGTKE...................................................  116
H1N.CM.95.YBF30        .........................................................  104
H1O.BE.87.ANT70        .........................................................  102
H1O.CM.91.MVP5180      .........................................................  103
CPZ.CD.90.ANT          .........................................................  118
CPZ.CM.05.SIVcpzEK505  .........................................................  103
CPZ.CM.05.SIVcpzLB7    GKGAKN...................................................  123
CPZ.CM.05.SIVcpzMB66   GTGAKN...................................................  123
CPZ.CM.05.SIVcpzMT145  DNRNQE...................................................  122
CPZ.CM.98.CAM3         GKGATN...................................................  124
CPZ.CM.98.CAM5         .........................................................  107
CPZ.GA.88.GAB1         GKGAKN...................................................  124
CPZ.GA.88.GAB2         AARTKN...................................................  123
CPZ.TZ.01.TAN1         .........................................................  105
CPZ.US.85.CPZUS        .........................................................  105
GSN.CM.99.CN166        DSLH.....................................................  124
GSN.CM.99.CN71         .........................................................  120
H2A.DE.x.BEN           .........................................................  103
H2A.GW.x.ALI           .........................................................  107
H2A.SN.x.ST            .........................................................  107
H2B.CI.x.EHO           SQGCGRDSCERGEDLVGSPQESGRRDHCNTQEDQTRG....................  150
H2B.GH.86.D205         PSGSKGYGRNSCECRRDLMGGSQESGESNHRDPQENQTRT.................  153
H2G.CI.x.ABT96         .........................................................  100
H2U.FR.96.12034        .........................................................   98
MAC.US.x.239           .........................................................  107
MAC.US.x.251_1A11      .........................................................  107
MAC.US.x.251_BK28      .........................................................  107
SMM.SL.92.SL92B        .........................................................   99
SMM.US.x.H9            .........................................................  100
SMM.US.x.PGM53         .........................................................  104
SMM.US.x.SIVsmH635FC   .........................................................  100
SMM.US.x.SIVsmH635F_L3 .........................................................  100
SMM.US.x.SIVsmH635SB10 .........................................................  100
MNE.US.x.MNE027        .........................................................  107
STM.US.x.STM           .........................................................   99
DEB.CM.99.CM40         .........................................................   90
DEB.CM.99.CM5          .........................................................   87
RCM.GA.x.GAB1          .........................................................   93
RCM.NG.x.NG411         .........................................................  107
DRL.x.x.FAO            .........................................................   86
MND_1.GA.x.MNDGB1      .........................................................   95
MND_2.CM.98.CM16       .........................................................  100
MND_2.GA.x.M14         .........................................................  100
MND_2.x.x.5440         .........................................................  100
GRV.ET.x.GRI_677       SSGEGCEGSLGQTGCYCPIRLSGSHQQSKKSAARP......................  154
TAN.UG.x.TAN1          .........................................................  111
VER.DE.x.AGM3          .........................................................   89
VER.KE.x.9063          .........................................................   89
VER.KE.x.AGM155        .........................................................   89
VER.KE.x.TYO1          .........................................................   84
TAL.CM.00.266          SPVWVQ...................................................  125
TAL.CM.01.8023         .........................................................  111
MON.CM.99.L1           .........................................................  114
MON.NG.x.NG1           .........................................................  113
MUS_1.CM.01.1085       .........................................................  110
LST.CD.88.447          ESLQQTGGGKRSEDCELELEQDQKEQKRHISSSCKKPQTGTKEEMEIQIQRKKWVSQ  165
LST.CD.88.485          ESQQRMDGGKRSEDCALELEQDQKEQKRHISSSCERHQTGTKEKMEIQIQREKWVSQ  165
LST.CD.88.524          ANPQQIPGGECSENCELDQEQEQKEQRRAIKN.........................  140
LST.KE.x.lho7          ADQQQMDGGKRSEDSELAQGEMQKEQRTVIEH.........................  140
SUN.GA.98.L14          ADPQRLAGFAPYSGYEQDQERVQNQQGESIIVSEGKK....................  146
SYK.KE.x.KE51          .........................................................   84
SYK.KE.x.SYK173        .........................................................   77
COL.CM.x.CGU1          .........................................................   99
```

```
                      |-     transmembrane domain      -||-                                 cytoplasmic domain                             -|
                                                                                                        / Env CDS start
                                                                                                 phos   phos
                                                                               |- alpha helix -|   |    | |-  alpha helix   -|
H1B.FR.83.HXB2        T....QP.IPIVAIVALVVAIIIA..IVVWSIVIIE..Y..RKIL....RQRKIDRLIDRLIE.......RAEDSGNESEGEISALV......EMGVEMGHHAPWDVD.......DL              82
H1B.US.90.WEAU160     M....-S.LQ-L--------G---.--------L--.-.----....-----------IRD.......-----------DQEE-SA.......L---------I-.......--              81
H1N.CM.95.YBF30       M....LS.LGFI-LG-A-SIA.....VI--ALLYR-..-..K--K....L-E--KHIRQ-IR-.......-E-------D-DAEW-DG.......DE-WLVTLLSSSKL..DQGNWV              83
H1O.BE.87.ANT70       MHH..RD.LLAII-ISALLF-N....VIL-GFILRK..-..LEQKEQDRKE-E-LERLR-IR-.......IRD--DY--N--EEQE-MD......L-LSHGFDNPMFEP........              85
H1O.CM.91.MVP5180     MHQ..EN.LLALIALSALCL-NV....LI-LFNLRI..-..LVQRKQDR-EQE-LERLR-IK-.......IRD--DY--NE-EQQE-ME......LIHSHGF-NPMFE........-              85
CPZ.CD.90.ANT         ........LLMTN.IFEYAFLAFS..--L-I-C-PILYK..LYKIYK..Q-QIDNKRNQ-I--.VLSRRL.SI--AI-EDE-ADTYYL.......-SGFANPVYREG-........E              85
CPZ.CM.01.SIVcpzCAM13 M....IL..IALGCL-IALILLN...-FI-RNLWRL..C..KQNK....LEVE-EN-AL--T-.......---------D-GEEEKL..QRLLHDHDN-HGF-NPLF-........I              83
CPZ.CM.05.SIVcpzEK505 M....LL.LIKLGFIG-AIETL..IV----A--YRI..-..-EVK....VEE--SQ-RQ-IRD.......---------D-DAEE-ANL.....LPPDRIDQDN-...........V              79
CPZ.CM.05.SIVcpzLB7   M....D..LIELGLIG--IEL-..IV----LKAYQL..-..KENI....--KA-NK--E-IR-.......-G---------DMDE-....HAILRSGDPELVLIDN..........              79
CPZ.CM.05.SIVcpzMB66  M....D..-VQQVGLLV-LI-ELVIV--I-VK-YKL..C..KEDR....--K------A-IR-.......---------D-DTEE-QDLITEGDNLMHI-IRDNRNN..........              86
CPZ.CM.05.SIVcpzMT145 M....-L.EIVLI-LFIA...LMLVA-FA-IAAYK-........YKKLQQV-R-E--Q--IRS.......---------D-DEIL--EELMQVHQHQVNPDWMDRILFW........              86
CPZ.CM.98.CAM3        M....LT.WEQIGLIGIGIE----..--A-G-AFK-..W..--GKEEEN-AQ--RT--E-I-S.......-Q-------D--DQRQL......DREIHVYGFDNPMF-W........              85
CPZ.CM.98.CAM5        M....LI...WEQ-GLIALG-ELIIV----G-AYKK..W..-E-KEERN-TQR-YN-AE-IRI.......-Q-------N--EE-QL......DNLLHANGFDNPMF-W.......P              86
CPZ.GA.88.GAB1        M....TL.LVGLVLILVGLIAWNIC...I-GYI-KWG.-..-RYK.RHRLETE-E--NLI-R-.......---------N--EEERLEQLI..HNYNHNN-F-NPMF-........-              88
CPZ.GA.88.GAB2        M....LS..MW---GLIGIGTLLVIN----G--G-SV.-..KRWK.RHKEEQR-ID--IKIT-.......---------D--DKETLA......TLLHNNGFDNPMFEDR......I              88
CPZ.TZ.01.TAN1        MIKIVVGSVSTNV-GI-CILL-LIGGGL...LIG-GIRRELERERQHQRVLERL.ARRLSIDS.......GV-E....D-EFNWNNFDPHNYN.....PRDW..............I              83
CPZ.US.85.CPZUS       M.....LN.WFEIGLIALGIEG-LVV-IWGLVARL...W..-Q-KIKENT-QE-QN-LE-IRI.......-E-------D--EEETLA......KLLSSLELDNPRI.........V              84
GSN.CM.99.CN166       M......HPAA-WWWGAAIITF-YLCVALLALYLAWDKWVKG-PKPTQVAVIRLIE..............DE----..........IY..DDASSELTGFNGF-NPGFE........V              78
GSN.CM.99.CN71        M......SAAALWWWGAA-ITF-YFCLAIFALYLAWDKWIKG-PKIP.VAVIRLVE..............DD-E--..........IF..EDASSEPNA.YGF-NPGFE........V              76
DEB.CM.99.CM40        M.....RL-QLLLLILGLIV-GTQGKQKKQQYLTVYYGVPVWVDAKVDLFCTANSSESGWAVTACLPHALVR-EV...PMPNVTQNF............NAFDN............              84
DEB.CM.99.CM5         M.........LSLLIIIASLTLVGAAPTKQQW-TVYYGVPVWHNATVGLFCTANSSESGWAVTACLPHAMVR-QV...PMPNV-QYF............DSFSN............              80
MON.CM.99.L1          M......N..YWWSLVAITYSL-LIALP-AAWAWWRYYKITKRFKRIDQEIQRLIQIHE-...........-RH---..VDT-SESEQHEETHGFVNPVF..NDDFGEW........V              87
MON.NG.x.NG1          M......N..TYW-A-IS-WCV-VLLAPCTLYLLYQSYKEHKRVSRFEQDFQRLVQXY............QEX---..Y-D-DEDHNSFDNPLFDDG......DP-QW........V              81
MUS_1.CM.01.1085      M......N..YWYLA-VI-TG-YFVIAIFAFVLAYQRWCKPK-VEVSVI-LLEE................GDG---..IFEDAEDDMAESEHHAFAN.......PAFE........Q              76
```

```
                                                  signal peptide \/ gp120
                                                  *                                ^^^       *                   *                  ^^^
SAB.SN.x.SAB1C        S....MKLLTVL...........LWLSGCWSLVWLV...........QY.VTVFYGIPVWKNSS...VQAFCK.......TPNTNLWASTNCIP.DDEPEGT.IAEVPIPNITEKFDAWKNRNPLVGQAE   91
H1B.FR.83.HXB2        M....RVKEKYQHLW........R-GWRWGTMLLGMLM..ICSATEKLW.---Y--V----EAT...TTL--ASDAKAYD-EVH-V--THA-V-.T-PNPQEVVLVNVTE-FNMW......K-DM-E-MH  105
H1B.US.90.WEAU160     M....RVKGIRKNYQ........H.-WKWGIMLLGILM..ICSAAENLW.---Y--V----EAT...TTL--ASDAKAYD-EVH-V--THA-V-.T-PNPQEVVL-NVTE-FNMW......K-NM-E-MH  104
H1N.CM.95.YBF30       M....KVMGMQSGWM........G.MKSG-L-FY-LVSLIKVIGSEQHW.---Y--V---REAE...TTL--ASDAKAHS-EAH-I--TQA-V-.T-PNPQEVLLPNVTEKFNMW......E-KMAD-MQ  106
H1O.BE.87.ANT70       M....E-RNKK-WTL........Y.-AM..A-ITPC.....LSL.RQL-.A--YA-V---EDAT...PVL--ASDANLTS-EKH-I---QA-V-.T-PTPYEYPLHNVTDDFNIW......K-YM-E-MQ   98
H1O.CM.91.MVP5180     M....K-NNRKSWSL........Y.IAM..A-LIPC.....LSYSKQL-.A--YS-V---EEAA...PVL--ASDANLTS-EQH-I---QA-V-.T-PNPHEFPLGNVTD-FDIW......K-YM-D-MH   99
CPZ.CD.90.ANT         ......XMRKPIHII...WG...-A-LI.QFIEKGT.........NED-.------V---R-AT...PTL--ATNASMTS-EVH-V--T-S-V-.I-PDPI.VVRLN.TS...VW-......-AYKNYMV   92
CPZ.CM.01.SIVcpzCAM13 M....KV..MEERKRS......FNGFYMTMTICMALLIPCLT...SDEW.---Y--V---REVN...TVL--ASDAKAHS-EAH-I--THA-V-.T-PNPQEVVLHNVTEDFNMWN......-QM-E-MQ  104
CPZ.CM.05.SIVcpzEK505 M....KV..-EMQKNW..LIC.C-LIGLIKIIG............SELW.---Y--V---RDAE...TVL--ASDAKAHS-EAH-I--TQA-V-.T-PNPQEVLIPNVTERFDMWK......-NM-D-MQ   98
CPZ.CM.05.SIVcpzLB7   M....RV..KGTWTS.....CMPYYVV-IL--CS-IISSTKN...ENLW.---Y--V----EAE...ATL--ASDAKAQKAEAH-I---QA-V-.T-PDPKAIELSNVTE-FNMWT......-AM-N-MQ  105
CPZ.CM.05.SIVcpzMB66  ......METQRNCRTL..SLKEIILCTLVLGIIGIIKCE......DNMW.---Y--V---READ...TTL--ASDAKAQNPEVH-V---QA-VS.TNPNPEEIELTNVTEYFNAWE......-NM-E-MH  105
CPZ.CM.05.SIVcpzMT145 ......MGMKYYWWRN..SCRSISIKLILIGWIASCFGE......ENWW.---Y--V---REAK...TTL--ASDAKSYS-EAH-I--TQA-V-.T-PTPQEVLLPNVTEEFNMWE......-YM-D-MQ  105
CPZ.CM.98.CAM3        M....KVMEKIKGSL........IEKYMFMG-IIPCLTGS.....DQLW.A--Y--V----EAD...TTL--ASEASALNKEAH-I---QA-V-.T-PNPQEVQIPNVTE-FNMW......N-TMAE-MN  102
CPZ.CM.98.CAM5        M....KAMERKKLNW........IICYMLMG-ITPCLTG......HEWW.A--Y--V----EAE...TTL--ASEASALNKEAH-I---QA-V-.V-PNPQEVYMPNVSESFNMW......K-NMAE-MQ  101
CPZ.GA.88.GAB1        ......MKVMEKKKRD..WNSLSIITIITII-LTPCLT.......SELW.---Y--V---HDAD...PVL--ASDAKAHS-EAH-I--TQA-V-.T-PSPQEVFLPNV-ESFNMW......K-NM-D-MH  104
CPZ.GA.88.GAB2        M....KVMGKTRRRWQ.......PYCII.MA-IIPCLKTE....SDQWW.---Y--V----EAT...TPL--AANATALKEEPH-I--TQA-V-.T-PSPEEV-LINVTEEFNVW......D-AM-E-MQ  103
CPZ.TZ.01.TAN1        ......MKNLI........G...IT.LIL.IITI-GI.......GFST-YT-----V----EAQ...PTL--ASDADITSRDKH-I--TH--V-.L-PNPYEVTLANVSIRFNME......E-YM-QEMK   94
CPZ.US.85.CPZUS       ......MKVMEKKKRL..WLS....YCLLS--IIPGL........SSLW.A--Y--V---RDVE...TTL--ASDAKAYKQEAH-I--TQA-V-.T-PNPQEVHLPNVTEKFDMW......E-NMAE-MQ   99
GSN.CM.99.CN166       M......RK--ESM.........MM-AVSLLALMALLI..PGLKC.ENW.T--Y--V---RDAT...PPL--ASDPDIASNEPG-I-I--A-L-.S-PSPAEVPL.NITEKFNIY......K-YM-DEVR  100
GSN.CM.99.CN71        M......MRK-GFL.........KM-VVSLT.LMALLI..PGSRSKGNW.T-----V---RDAK...PPL--ASDADITSNEPG-I-I--A-L-.S-PSPSEVPL.NITEEFNIY......K-YM-DEVR  100
H2A.DE.x.BEN          M......EPGRNQL.........FVVILLT-ACLVY.....CS....--.--------A---A-...IPL--A.......-K-RDT-GTIQ-L-.-NDDYQE.-ILNVTEAFDAWN......-TVTE--V   87
H2A.GW.x.ALI          M......MSSRNQL.........-VTILLA-ACLVY.....CK....--.------V-A---A-...IPL--A.......-K-RDT-GTIQ-L-.-NDDYQE.--LNVTEAFDAWD......-TVTE--V   87
H2A.SN.x.ST           M.......CGRNQL.........FVA-LLA-ACLIY.....CV....--.------V---R-A-...IPL--A.......-K-RDT-GTIQ-L-.-NDDYQE.--LNVTEAFDAWN......-TVTE--V   86
H2B.CI.x.EHO          M......AHVNNYL.........-VTLLLI-IYGYM.....GK....NF.--------A---A-...IPL--A.......-R-RDT-GTVQ-L-.-NDDYTE.-QLNITEAFDAWD......-TVTD--T   87
H2B.GH.86.D205        M......AYFSSRL........PIA-LLIG.ISGF-.....CK....--.--------A-R-AT...-PLI-A.......-T-RDT-GTVQ-L-.-NGDYTE.-RLNITEAFDAWD......-TVTQ--V   87
H2G.CI.x.ABT96        M......AYLGNQL.........-IATXLL-TYGFY.....CT....K-.--------A-R-A-...-PL--A.......-K-RDT-GT-Q-L-.-NDDYSE.L-LNVTEAFDAWD......-TVTE--V   87
H2U.FR.96.12034       M......ACLGNQL.........-I.ALLCLNACVIY....CT....--.--------A-R-AT...IPL--A.......-N-RDT-GT-Q-L-.-NGDYSE.L-LNVTEAFDAWN......-TVTE--V   87
MAC.US.x.239          M......GCLGNQL.........-IAILLL-VYGIY.....CT....L-.------V-A-R-AT...IPL--A.......-K-RDT-GT-Q-L-.-NGDYSE.V-LNVTESFDAWN......-TVTE--I   87
MAC.US.x.251_1A11     M......GCLGNQL.........-IAILLL-VYGIY.....CT....L-.------V-A-R-AT...IPL--A.......-K-RDT-GT-Q-L-.-NGDYSE.L-LNVTESFDAWE......-TVTE--I   87
MAC.US.x.251_BK28     M......GCLGNQL.........-IAILLL-VYGIY.....CT....--.------V-A-R-AT...IPL--A.......-K-RDT-GT-Q-L-.-NGDYSE.L-LNVTESFDAWE......-TVTE--I   87
MAC.US.x.EMBL_3       M......GCLGNQL.........-IAILLL-VYGIY.....CT....--.------V-A-R-AT...IPL--A.......-K-RDT-GT-Q-L-.-NGDYSE.L-LNVTESFDAWE......-TVTE--I   87
SMM.SL.92.SL92B       M......ACPG-HL.........-I.DILFLS-LGTW....CA....--.--I-----A-R-AT...IPL--A.......-Q-RDT-GTVQ-L-.-NGDYSE.L-LNVTEAFDAWD......-TVTE--I   87
SMM.US.x.H9           M......GCLGNQL.........-IALSLX-ASGIY.....CV....--.------V-A-R-AT...-PL--A.......-R-RDT-GT-Q-L-.-NGDYSE.L-INVTEAFDAWD......-TVTE--I   87
SMM.US.x.PGM53        M......GCLGNQL.........-I.AILLLSAYGIY....CT....--.------V-A---AT...IPL--A.......-K-RDT-GT-Q-L-.-NDDYSE.L-INVTERFDAWN......-TVTE--I   87
SMM.US.x.SIVsmH635FC  M......GCLGNQL.........-IALLLV-VLEIC.....CV....--.------V-A---AT...IPL--A.......-R-RDT-GT-Q-L-.-NDDYSE.L-VNITEAFDAWN......-TVTE--I   87
MNE.US.x.MNE027       M......GCLGNQL..........LIAILFLSAYGIY....CI....--.------V-A-R-AT...IPL--V.......-R-RDT-GT-Q-L-.-NDDYSE.L-LNITESFDAWE......-TVTE--I   87
STM.US.x.STM          M......ACPGNQL..........LIAILLLSAC-TY....CT....--.------V-A-R-AT...IPL--A.......-K-RDT-GT-Q-L-.-NGDYSE.L-INVTEAFDAWD......-TVTE--I   87
DEB.CM.99.CM40        M......PNVTQNF............................................................................................................-AFDNPI-   16
DEB.CM.99.CM5         M......PNVSQYF............................................................................................................DSFSNVM-   16
RCM.GA.x.GAB1         M......DK..............KLVIVLIVVIGIIL..VQGSQK.P--.I-----V---R--T...-PM--V.......-D--QS-GTL----.EGGISPE.VSINVSERFDAWN.....NSLYE.--K   88
RCM.NG.x.NG411        M......A..K-IS...........IGIIIG-LGII......IVGEK--.------V------T...-PM--A.......-D-AQS-GTL----.EGGISPE.VPVNVSEKFDAWE......-S-YE--K   86
DRL.x.x.FAO           M..........IRHI...IIGLCI-GFLGISIGK.............HW.------T-K-RPA-...THLI-A.......-D-HSF-VT-S---.SLLHYEETEINNIEE-F-VPM........TENEVI   84
MND_1.GA.x.MNDGB1     MSTGNVYQELIRRYLVVVKKLYEGKYEVSR-FSYTMFSLLVG.IIGK--.------V----EAK...THLI-A.......-D-SS--VT-----.SLPDYDEVEIPDIKE-F-GLI........REN-IV  109
MND_2.CM.98.CM16      M..........-RY......IILGIIV-LGL.G.........N....-W.---Y--T-K-HKAE...THL--A.......-E-DS--VT-S-V-SLLHY-EQP-PNITWNXX-GPM........EENEVV   81
MND_2.GA.x.M14        M..........-RH.....LITAIVII-IGI.G.........K....-W.---Y--T-K-HPAR...THL--A.......-D-NS--VT-S-V-.SLLHYEEQLIPNITE-F-VPI........QENEVV   81
MND_2.x.x.5440        M..........-RY.....LRYIVLGIIVSVI-GE............-W.---Y--T-K-HKAR...THL--A.......-D-NSF-VT-S-V-.SLLHYEEQHIPNITE-F-GPI........EENEIV   83
GRV.ET.x.GRI_677      M......GRLLIKI.........-IIAIGI-IGIGN.......LY.....-------------T...-----M.......-----M--T-----.--HDNTE.VPLNITEAFEAWD......----K---   85
TAN.UG.x.TAN1         M......GPLRGKG.........VL-VILGLSLIGL.......LYGT--.I-------------...-----M.......--------T-----.--HDYTE.VQLNVSEKFEAWK-R....----A---   91
VER.DE.x.AGM3         M....KLT-LIG...........IL-I-IGVVLNTR.......Q...-W.------V-------...-----M.......--T-R---T--S--.--HDYTE.VPLNITEPFEAWA-R....----A--G   88
VER.KE.x.9063         M....KP--GIV...........-IAII.GISS.-L.......SER.-W.--------------...-----M.......--T-R---T-----.--HDYTE.VPLNITEPFEAWA-R....----A--G   88
VER.KE.x.AGM155       M....T-F-GIF...........IV-GI.GIGI.GI.......STKQ-W.I-----V-------...-----M.......--T-R---T-----.--HDYTE.VPLNITEPFEAWA-R....----A--G   89
VER.KE.x.TYO1         M....RYTIIT-...........GIIVI.GIGI.VL.......SK..-W.I-------------...-----M.......--T-S---T-----.--HDYTE.VPLNITEPFEAWG-R....---IA--A   87
TAL.CM.00.266         M....KY-AVI...............FCLAY-.......GRFI.IGN--.--IY--V-----AT...PPL--A.......-MARDQ-STSQ-V-.S-PQNLE.VPVNLTEYFDVY......K-YM-DHIA   84
TAL.CM.01.8023        M....KC-QVL................CLAIC-IW-ANIGKSLTSAE-QFL--Y--V-----AT...PPL--A.......-MARDQ-STSQ-V-.S-PQNLE.VPVNLTEYFDVY......K-YM-DHIA   92
MON.CM.99.L1          M....R-.................LMALSTLCLMMT......LVNGSNW.T--Y--V----PAT...PPL--ASDPNYGSKEAG-N-LASS-L-.T-PTPQSLYLNITEEFNAYQ.......-YM-EEMV   91
MON.NG.x.NG1          M....RMRMKTI.............IVLTILCLMMGTLTSGSXSGNKSW.------V--Y-PAT...PPL--ASDPXYGSKEAG-N-LASS-L-.S-PVPRALPLNITESFNAYN.......-YM-EEMV  101
MUS_1.CM.01.1085      M....PRMTWLRA..........STTLLLTL-LSSS......IIWGR-W.------V----DAT...PPL--ASDANIANREPG-V-IT-A-L-.T-PAPQEVQLNISGEKFDIYK......-YM-D-MT   99
LST.CD.88.447         M......NCPG-IL.......LLII-RVQG...........EKTKK.--.------V-K-EEAV...-PLI-A.......SA-NS--VT-S---.-LQSYAQ.-PIHN-SLNFT-E.......IKDN-II   85
LST.CD.88.485         M......DCLK-LL.......FL-F-GAQG...........EAKK..--.------V-K-EDAV...-PLI-A.......SA-NS--VT-S-L-.-LQSYAQ.-PIHN-SLNFTQE.......IKDN-II   84
LST.CD.88.524         M......ACPGNIL.........-L-LCLGLTAG.......EKEK..--.------V-K-EDAV...-PLI-A.......SA-NS--VT-S-L-.-LQSYAQ.-PIHN-TLNFTQE.......IKDN-II   86
LST.KE.x.lho7         M.......ACPGLG.........IL-LLLGII..W......GK....--.------V-N-DDNVS..-PLI-A.......SA--S--VT-S-L-.-LQSYAE.VPIYN-SENFTIP.......VKDN-VI   83
SUN.GA.98.L14         M....RC..................PEILIGFSL-L......GSIAI--.------T-K-EPAV...-PLI-A.......SA-NS--VT-S-L-.-LQTYAEVPITGLEE-F--GISN....-QI-Q--W   86
SYK.KE.x.KE51         M....KLFNIISVV...........-GIVL--GIVA........RLE--.------V-N-EDAQ...-PM--A.......--H-GG--TK--V-..SSEQIEVRVNISGEYF-AWNSS....HGIRQ-LL   90
SYK.KE.x.SYK173       M....AAFR-YIVC..............LF--IS-G......FMEKQ--.--------H-EDAY...APL--T.......-SHKGG--TK--V-SA-QI-VRVNITG..EYFPAWNSS....HMIRQ-IL   89
COL.CM.x.CGU1         M......-RYLFIN...........-LVLGIVL..........SN..KW.---YQ-V-A-EEADVNDQ-F--F.......SSSPEIQQVLG-L-PPPGKPVEQNMPNVTEAFDLF......K-SFS-EV.   86
```

```
                                       |-                                V1 loop                                  -|-    V2 loop
                                       *      *    *                                             ^^^              *                  ^^           ^
SAB.SN.x.SAB1C        SNIHL....LFESTLKPCVKLSPMCIKMNCYRLEGGAATTTSPSTSTARPEVVS..................VGFNDSVIEQEMEKEQAMN.CS...FAMAGY..RRDVKKNY...........STV.W   180
H1B.FR.83.HXB2        ED-IS....-WDQS-------T-L-VSLK-T.....DLKNDTNTN-SSGR......................MIMEKGE.........IK-.--...-NISTS..I-GKVQKEYAFFYKL.DIIPIDND   187
H1B.US.90.WEAU160     ED-IS....-WDQS-------T-L-VTL--TNVNVTNLKNETNTN-SSGG..............................-.K--EGEMK-.--...-NVTTL..I-NKR-TEYALFYKL.DVMPIDHD   191
H1N.CM.95.YBF30       ED-IS....-W-QS-------T-L-VT-L-NDSY-EERNN-N..................................MTTR-PDIGYK-MK-.--...-NATTE..LT-K--QVYSLFYVE.DVVPINA.   189
H1O.BE.87.ANT70       ED-IS....-WDQS-----QMTFL-VQ-E-TNIA-TTNE...............................................NLMKK.-E...-NVTTV..IK-K-EKKQALFYVS.DLMELNET   169
H1O.CM.91.MVP5180     ED-IS....-W-QS----E-MTFL-VQ---VD-QTNKTG.........................................LLNETINEMR-.--...-NVTTV..LT-K-EQKQALFYVS.DLSKVNDS   176
CPZ.CD.90.ANT         ESMTEDMXQ--QQSH------T--------TGYN-TPT-P..--T-ST........................-TPKTTTPIVDGM-..LQE.-N...-NQSTG..FK-K-QKMKAIFYKG.DLMKCQDN   186
CPZ.CM.01.SIVcpzCAM13 ED-SS....-WDQS-------T-L-VT-K-SNVTRQNS-SN-N...........................NNPTLKNNTSGKNETGVLQ-KN-T...-NTTTE..L--K--QVYSLFYVD.DLQ-LGS.   195
CPZ.CM.05.SIVcpzEK505 ED-IS....-W-QS-------T-L-VTLS-SSWRSVNNS..........................................VNQTNHVQ-QN--...-NVTTE..L--K--QVYSLFYMG.DIIPLDT.   174
CPZ.CM.05.SIVcpzLB7   QDVIS....-WDQS----I--T-L-VTLH-SIPKFDNSS...........................................INSSNET-KN--...-NTTTE..L--K--QTYSLFYVD.DLTQINK.   180
CPZ.CM.05.SIVcpzMB66  ED-VN....-WDQSV------I-L-VTL--SLFKCIKEN.......................GNTTNCTVQISTG---TANNITVGTID-YN--...-NATTE..L--R--QVYSLFYRQ.DLEPLEG.   200
CPZ.CM.05.SIVcpzMT145 ED-IS....-W-QS-------T-L-VTLT-NNPTNTSC-.........................................NSTDDRLGD-RN--...-NVTTE..L--K-RQVYSLFYVE.DITAIGN.   182
CPZ.CM.98.CAM3        TD-IS....-WDQS-------T-I-VT-E-RKVTFNSTSNRNKTSTMTT..........................-SPN...-KXDSTVK-.-T...-N-TTE..V--KE-KVYSLFYVD.DLVLIDND   191
CPZ.CM.98.CAM5        ED-SS....-WDQS-------T-I-VT-Q-TT-KNIST-N-N.--ELKN...................................I--IMK-.-T...-N-TTE..V--K--EVYSLFYVD.DIVPIGGN   183
CPZ.GA.88.GAB1        ED-IS....-WDQS-------T-L-VTLQ-SKANFS.......QAKNLT........................NQTSSPPL-.......MK-.--...-NVTTE..L--K--QVYSLFYVE.DVVNLGNE   185
CPZ.GA.88.GAB2        ED-TS....-WDQS-R-----N-L-VQLT-TSINATGNE-NATG...............................-GIEK..DGLAQDMR-.-T...-NTTTE..L--K-QQIYSLFWKN.DLVG-NN.   187
CPZ.TZ.01.TAN1        ED-LS....--QQSF------T-F----T-.TMTNTTNK-...........................LNSATTTLTPT-NLSSIPNYEVY-.--...-NQTTE..F--K--QIYSLFYRE.DIVKEDGN   184
CPZ.US.85.CPZUS       ED-IS....-WDQS----I--T-L-VT-T-LNP.............DSN.......................SSAV-TTD.I......MR-.--...-NITTE..L--K--QVYSLFYVD.DLAHINN.   173
GSN.CM.99.CN166       DDMVS....--NQA-------T---VG-Y-NLTNTSTSEP-TTPRPPNVSTTT...................QWGSWGGENGTGQP..LY-.--...-NQTTE..F--Q--QMYSLFWRE.DIMEETHG   197
GSN.CM.99.CN71        DDMVS....--NQA-R-----T---VR-K-KLPNTSTT--P-TT-A-TS-SSPTPTP................WG-WGG.NGTGQP..IY-.--...-NQTTE..F--K-RQMYSLFWKD.DIMRAQDG   199
H2A.DE.x.BEN          EDVWH....---TSI------T-L-VA---S-.VQ-NT--PN-R--SSTTSRPPTS...................AA-I-NETSNCIENNT.-A...GLGYEE..MMQCEF-MKGLEQDKKRRYKDT.-   185
H2A.GW.x.ALI          EDVWR....---TSI------T-L--A-K-SN.ISTES------PGSTL..........................KP.L-NESDPCIK-D-.-P..RGLGDEE..MVNCRF-MTGLQRDKPKQYNET.-   178
H2A.SN.x.ST           EDVWS....---TSI------T-L-VA-R-NS.TTAKN--ST-.-T-TTA.........................-TTIG.ENSSCIRTD-.-T...GLGEEE..MV-CQF-MTGLERDKKKLYNET.-   176
H2B.CI.x.EHO          KDVWS....---TSI------T-L-VT-K-NKTWSS-SKE-TT-SASL-SSTQ........................TLLNEDSKCI-NDS.-A...GIGLEE..MI-CQFKMTGLKRDESKQYKDT.-   181
H2B.GH.86.D205        DDVWR....---TSI------T-L-VA---SKT-TNPGNAS-TT-TKPTTTS......................RGLKT-NETDPCIKNDS.-T...GLGEEE..IMQCNFSMTGLRRDELKQYKDT.-   183
H2G.CI.x.ABT96        EDVWN....---TST------T-L--A-R-NKT-TNKWGL-GTT-V-PATVPTTKMVTA..................ELVNSTSQCLMYD-.-T...GIQSES..MVGC-F-MTGLKRDQRKEYNET.-   187
H2U.FR.96.12034       EDVWS....---TSI------T-L--A-R-SKT-TNRWGL-GTAAT-TS-AMQESS.....................K--NETEPCIRNNS.--...GLEQEP..LVSCRF-MTGLKRDKKQEYKET.-   184
MAC.US.x.239          EDVWQ....---TSI--------L--T-R-NKS-TDRWGL-KS..I-TTASTT-TTA................SAKVDMVNETSSCIAQD-.-T...GLEQEQ..MISC-F-MTGLKRDKKKEYNET.-   187
MAC.US.x.251_1A11     EDVWQ....---TSI------T-L--T-R-NKS-TDRWGL-KS--T.TTAPKAM...................SEKINM-NETSSCIVHD-.-T...GLEQEQ..MIGC-F-MTGLKRDKKKEYNET.-   186
MAC.US.x.251_BK28     EDVWQ....---TSI--------L--T-R-NKS-TDRWGL-KS--TITTAAPT-APV................SEKIDMVNETSSCIAQN-.-T...GLEQEQ..MISC-FTMTGLKRDKTKEYNET.-   189
MAC.US.x.EMBL_3       EDVWQ....---TSI--------L--T-R-NKS-TDRWGL-.........KSSTTITTAAPTSA.......P-SEKIDMVNETSSCIAQN-.-T...GLEQEQ..MISC-FTMTGLKRDKTKEYNET.-   189
SMM.SL.92.SL92B       EDVWN....---TSI------T-L--A-K-NKN-TDRWGL-RAA-T-SS-TTT-PLTAASPSGE.............EIVNDT-SCTKNN-.--...GIEQEP..MIGCQF-MTGLKRDQKRQYNET.-   192
SMM.US.x.H9           EDVWN....---TSI------T-L--T-R-NKS-XXRWGL-GXPAP-TTQTSTTPPSPIIA..............K..-VNDSDPCIRSN-.-T...GLEQEP..MVSC-F-MTGLKRDKKREYNET.-   189
SMM.US.x.PGM53        EDVWN....---TSI------T-L--T-R-NKT-TDRWGL-KNEIP-TTTTTTTKSPKA.............EAITAK--NESDPCISNN-.-T...GLEQEP..MVSCRF-MTGLKKDKKKEYNET.-   192
SMM.US.x.SIVsmH635FC  EDVWN....---TSI------T-L--A-R-NKT-TDRWGL-GRAET-TTAKSTTSTTTTT..............VTPK--NEGDSCIKDNS.-A...GLEQEP..MIGC-F-MTGLKRDKKIEYNET.-   191
MNE.US.x.MNE027       EDVWH....---TSI------T-L--T-K-NKS-TDKWGL-KS--T--A-TTKTTTT.................KEIE-VNENSTCVNRD-.-T...GLEQEP..MISC-F-MTGLKRDKKREYNET.-   188
STM.US.x.STM          EDVWN....---TSI------T-L--A-R-NKN-TDKWGL-GKTVT-VT-TAAAAAT..................KPELVNETSSCVSNN-.-T...GLEEES..LVGC-F-MTGLKRDKKREYNET.-   187
DEB.CM.99.CM40        EQLWQDMTS-YKQSF------T-Y-VS-Q-IKTSTNPTP-NTST-T-TIATTT....................KTTTDWSGENITMT-YW-.--...-NVS-P..Y--K-EKS.SAVWLEDDIQWA...   115
DEB.CM.99.CM5         QQLWDDMTS-YRQSF------T-Y-VR---........SSTTT-T--ATTTI.....................KTTTDWSGENKTVT-YW-.-T...-NVT-P..Y--K-EMQ.SAVWLEDDIQWE...   106
RCM.GA.x.GAB1         D-VWNL....YD-------R---L--T---SAIN-SWDG......IPTSAPPTTTKT.................TTQRT-GV-K-CTAGNET-E...EVQDAD..VMSCEFAVAGLKRDEKHKYNDT.-   184
RCM.NG.x.NG411        --VWH....-YD-------R-T----T-Y-TA-N-SWDG-ATQ-PT-TTRATTRTT.......................K-IGKDCYANQTE--...SVPDAD..VMNCEFTVAGLKRDEKMKVNDT.-   182
DRL.x.x.FAO           KQAWGALSSMIDAV------IN-Y-V--E-SD-...KIVNAAST-P-PTT.................VATTVNTSVA-TTTANDTL-I-V-.NT...EPVMET..N-VC-Y-VTGLCRDCKEEITEN.F   186
MND_1.GA.x.MNDGB1     YQAWHAMGSMLDTI------IN-Y-V--Q-QET-NVS---AK-I-TPTTTST-ASST.........................-IYLDVDK-.NT...EEKVER..NHVCRY-ITGLCRDSKEEIV-N.F   206
MND_2.CM.98.CM16      MQAWGAISSMIDAV-------T-Y-V--Q-TKGKDL-TPIPTT--T-TTTTT.........................KTVANTT-LDIDT-NTE....TTTQQ..N-VC-F-TTGLCRDCKLEIEEN.F   178
MND_2.GA.x.M14        RQAWGAISSMIDAV-------T-Y-V--Q-TKG-NSSTS--T-MPT-TT-TTP....................AVTGLEVTVQN........NE....STTQE..N-VC-F-TTGLCRDCKIEIKES.F   175
MND_2.x.x.5440        TQAWGAISSMIDAV-----R-T-Y-V--K-TEGQNETEQA-AKT-TPVPTTTTPSTTT................SSSTNKTTTPVLVVEKQNNE....TTTQQ..N-VC-F-TTGLCRDCKLEIEEN.F   189
GRV.ET.x.GRI_677      -----....---Q-MR-------I----S-VE-N-T-T-KATTTAT-TMT.........TPC.....................QNCSTEQIEGE....MAEEP..ASNCTFAIAGYQRDVKKNY-MT.-   173
TAN.UG.x.TAN1         -----....------------T--------TK-TST-P-SST-TS-STT..........DPCPNTDE................SSCNATLVTN-....MDYEN..SSICSFAMAGYRRDVKKKYNST.-   183
VER.DE.x.AGM3         -----....---Q----------L----S-VE-NSSEP---PK--TASTTNITASTTT..LPCVQNKTSTV.LESC-ETI--K-LNE-P-S-.-T...------..V--Q--K-...........-V-.-   193
VER.KE.x.9063         -----....---Q----------L----S-VE-NSSRRNP-PT-A--TKA.PKTG....DPCIKSTNNNVNLQPC-A-L--E-L-E-A-S-.-T...------..I--Q----...........-V-.-   190
VER.KE.x.AGM155       -----....---Q----------L------VE-K-S-TS-PAT--TAGTK.........LPCVRNKTDS.NLQSC--TI--K--ND-A-S-.-T...------..I--Q----...........-V-.-   186
VER.KE.x.TYO1         -----....---Q-M--------L------VE-NSTRERA-T-T-TPKSTG........LPCVGPTSGE.NLQSC-A-I--R---D-P-S-.-T...------..V--Q----Y..........-V-.-   186
TAL.CM.00.266         EDMDS....--LTATR--T--T-I-VR-K-EDITKNLTISANTT-PAPT.......................AMS-SNMPWWGDNTTEPVYN-TWNMTGAFRD..K-YTYFSHWYLADLMKEPGNGS..   182
TAL.CM.01.8023        DDMDA....--LTAT---T--T---VR-A-TDITKNLTG-VA-P-TNSTTNS-TS.................KRT--TNVWWGNNATEPVYN-TWNMTGGFRD..KKYTYFSHWYLADLMKEPGNGN..   196
MON.CM.99.L1          EDMKS....--SQA-------T---VR-L-VEVNTVSNAS-T-AP--PT-.........................WGNWG.GN.GTG-PVYN--FNQTTEFRD..KKKQMYSLFWKEDIMKEEG-NG..   185
MON.NG.x.NG1          ED-KS....--SQA-------T---VR-R-QXINESTTVXPTTTXX...................ATTP......WG-WG.GN.GTG-PLYN--FNQTTEFRD..KXKQMYSLFXKEDIMQDKDN-T..   195
MUS_1.CM.01.1085      DDVRS....I-NQA-------T---V--K-AKVTNYTRS.....................TGGNSTGTTAPPTTPWG-WGNGNDSAG-P-FN--FNQTTEFRD..KKRQMYS..LFFVDDLMRD-SE..   197
LST.CD.88.447         QQAWSAMTSMVDTIM-----VN-Y-VR-K-EGEGI-TTKA-KTK-.STT-.........MPCFITEKTTTTTKRSTTGGP-....LIEDI-.-IKWVNETTEI..N-ECRY-VTGLCRDCRTEIKQS.F   196
LST.CD.88.485         QQAWSAMTSMVDTIM-----IN-Y-VR-K-TGGGI-PK--P--IGK-TT-..........MPCFIAETTTTTTK-NTTMES....LIEDT-.-TRWLNETTEI..N-EC-Y-VTGLCRDCKEEIKQS.F   195
LST.CD.88.524         QQAWGAMTSMVDAIM---T-IN-Y-VR-K-TGEGL-KKR--TTT-P-TT-.........MPCFITESTTTTTKSSGTEQPS....LIEDT-.-TRWLNETTEI..N-ECSY-VTGLCRDCKEEIKQN.F   198
LST.KE.x.lho7         QQAWSAMNAMVD-IM-----IN-Y-VR-Q-GEVTKTPT--PK....-TTQ..........MPCFINEQV.....TVKNPGNET.RL-EDL-.-TRGLNETTER..NAECQY-VTGLCRDCRTEIKQS.F   188
SUN.GA.98.L14         QAMTS....MVDAIM-----IN-Y-V--K-PPKPTTPSS...........NSTVKSSCDYWTTTTAKTTTQTTSSTS-TASTTTPMPLDW-.-T.DTENI-ES..NKVC-Y-VTGLCRDCKTEVEQN.F   195
SYK.KE.x.KE51         QDMSD....--LQADR----------RTL-VEDKDSSSG....................NLTTTISPTTTTNKAATPPAWWG.DNSTETRYN--FNMTGGFKD..KKHQ.Y-...AFFYKSDIMREE.G   187
SYK.KE.x.SYK173       EDMSA....--LQANR-----A----R-L-TLDNSP-TS..........................TPTTSPPTTPPNETWWGDNST-PRF-.--................F-LTGGFKDKKQQYRAF.F   170
COL.CM.x.CGU1         ...WIITQTTL-QR-R--A--TAY-AP-I-TKVN.............RTENGT-TVAPTT......TNNSASDWDES.....NWKEYPWY-.-R..MNST-FL..LK-R-ELELGFSVEDLTVLGN..K   181
```

```
                                                   V2 loop    -|
                                  *                       *    *^^^    *           *           *            * ^^^   *     ^^^         ^^^                   ^^
SAB.SN.x.SAB1C         DDQEVVCEE.GREKSN....AT...HT.VGCYMIHCNTSVIKEACDKTYWDTFRLRYCAPAGYALLRCADTDYSGHKA..CRNVTVSACTRLINTTVSTGIGINGSYVA.NRTEIW......QKNGNSND   292
H1B.FR.83.HXB2         T...............T...........SYKLTS------TQ--P-VSFEPIPIH------F-I-K-NNKTFN-TGP..-T--STVQ--HG-RPV---QLLL---LAE.EEVV-R.......SVNFTDN   280
H1B.US.90.WEAU160      N...............T...........SYTL-N-KS-T-TQ--P-VSFEPIPIH------F-I-K-N-KKFN-KGP..-K--STVQ--HG-RPV---QLLL---LAE.EDIV-R.......SENFTDN   284
H1N.CM.95.YBF30        ..............YNK...........TYRL-N---TAVTQ--P--SFEPIPIH----P-F-IMK-NEGNF--NGS..-T--STVQ--HG-KPVI--QLIL---LNT.DGIV-R...........NDS   279
H1O.BE.87.ANT70        SSTN.........-T-SK..........MYTLTN--STT-TQ--P-VSFEPIPIH--------IFK-NS-EFN-TGT..---I--VT--HG-RP----QLIL--TLSK.GKIRMM.......AKDILEG   269
H1O.CM.91.MVP5180      NAVN...........GT...........TYMLTN--STI--Q--P-VSFEPIPIH----T---IFK-N---FN-TGL..-H-IS-VT--HG-KP----QLIL--TLSR.EKIR-M.......GKNITES   273
CPZ.CD.90.ANT          NETN........................CY-LW----TT-TQS-E-STFEPIPIH--------I---E-E-FT-VGM..-K--S-VH--HG-SPM-A-WLLL--T-QTNTSVVMN......GRKN..ES   281
CPZ.CM.01.SIVcpzCAM13  ........................GTG.DTYT--N---TA-TQ--P-VSFEPIPIH------F-I-K-N-VNF--KGK..----STVH--HG-KPV-T-QLI----LAT.ENITVR.......V-NA-KN   290
CPZ.CM.05.SIVcpzEK505  ...................NNSSGNN.SQYRL-N---TAVTQ--P-ISFEPIPIY----P-F-IIK-N-Q-FN-TGE..-N--STVQ--HG-KPVI--QLIL---LAT.SNIV-R.......N-SKD..   271
CPZ.CM.05.SIVcpzLB7    .......................T-.ESYKL-N---TA-TQ--P-ISFEPIPIH------F-----NNKT-N-SGP..-N--STVH--HG-KPVI--QLIL---LAE.EKIM-R.......YKA.D..   271
CPZ.CM.05.SIVcpzMB66   .................NKPPEGDKN.ALYRLYN---TAMTQ--S-VSLEPIPIH------F---K-N-KNFT-IGQ..-K--STVH--HG-RPV---QFLL--TLEEKVTVLDR.......NVS-DM-   302
CPZ.CM.05.SIVcpzMT145  ........................N.STYRL-N---TA-TQ--P--SFEPIPIH------F---K-N-I--K-NET..-K--STVH--HG-KPVAT-QLIL---TAD.-Q-VAR.......IDPSE..   273
CPZ.CM.98.CAM3         T-..........................TYRL-N---TA-TQ--P--SFEPIPIH---TP---IMK-NMPNFN-TGTGR-N-ISTVQ--HG-RPV-T-QLIL---VAE.-K-IAR...........RN   281
CPZ.CM.98.CAM5         KSND........................TYRL-N---TTVTQ--P--SFEPIPIH----P-F-IMK-N-QKFN-TGP..-T-ISTVQ--HG-RPV-T-QVIL---LAEEGPAVAR.......SEDL.RH   278
CPZ.GA.88.GAB1         NN..........................TYRI-N---TA-TQ--P--SFEPIPIH------F-I-K-N-K-F--KGK..-T--STVH--HG-KPV-T-QLL----LAE.GNITVR.......VENK-KN   278
CPZ.GA.88.GAB2         ............................TFRL-N---TA-TQ--P--SFEPIPIH------F---K-N-K--P-KGK..-K--STVH--HG-KP--T-QLIL---LAE.EEEVVF....RTKNMTAPGL   281
CPZ.TZ.01.TAN1         ................NN..........SY-LHN------TQE---STFEPIPI-------F---K-R-QNFT-KGQ..-S--S-VH--HG-YPMIA-ALHL---LEE.EE-KAY........FV-TSV   276
CPZ.US.85.CPZUS        .N..........................TYRL-N---TA-TQ--P--SFEPIPIH----P-F-I-K-NEK-FK-KGE..-K--STVQ--HG-KPV-T-QLI----LAT.KNVTVR........S.K.-F   262
GSN.CM.99.CN166        ...............-QS...........-Y-IRN----Y-TQK-V-SSFQPVPIH----P---M-K-N-VNFT-VGT..-Y--SAVT--HG-QPL-A-WLHL--T-QP.GN.NTR.....IMM--MK-E   292
GSN.CM.99.CN71         ...............-DT...........DY-I-N----YVTQK-V--SFQPVPIH----P-F-M-K-N-ANFT-VGK..-S--SAVT--HG-QPL-A-WLHL--T-QP.GN.NTR.....VMM--KK-E   294
H2A.DE.x.BEN           YLED---DNTT.....AG..........T---R-----I---S---H---AM-F-----P-F-----N--N---FEP.K-TK-VAAS---MME-QT--WF-F--TRAE.---Y-Y........WHGRDN   289
H2A.GW.x.ALI           YSKD----PFNTTTN.QT..........R---N-------T-S---H---AI-F-----P-------D-IN---FAP.N-SK-VAAT---MME-QT--WF-F--TRAE.---Y-Y........WHGRDN   286
H2A.SN.x.ST            YSKD----SNDTK-E..K..........T---N-------T-S---H----M-F-----P-F-----N--N---FEP.N-SK-VAAT---MME-QT--WF-F--TRAE.---Y-Y........WHGRDN   283
H2B.CI.x.EHO           YK-DL---KGT-SN..ES..........K--IKT----I-Q-S---H---SL-F-----P-F-----N--K---FMP.N-SK-V--L.Y-MME-QT--WF-F--TRAE.---Y-Y........WHGKDN   287
H2B.GH.86.D205         YSEDLE-NNTRKYT...S..........R--IRT---TI-Q-S---H---SL-F-----P-FF----N--N---FMP.N-SK-VA-S---MME-QS--WF-F--TRAE.---Y-Y........WHEKDN   289
H2G.CI.x.ABT96         YS-DLQX-QSNKSEN.ES..........R---R-------Q-S---H---AI-F-----P-F-----N--N---FAP.N-SK-V--T---MME-QT--WF-F--TRAE.---Y-Y........WHGRDN   295
H2U.FR.96.12034        YSHDI---QNTS-S..ES..........K---N-------Q-S---H---AI-F-----P-------N--N---FMP.N-SK-V--T---MRG-QT--WF-F--TRSE.---Y-Y........WHGRDN   291
MAC.US.x.239           YSADL---QGN.NTG-ES..........R---N-------Q-S---H---AI-F-----P-------N--N---FMP.K-SK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHGRDN   295
MAC.US.x.251_1A11      YSADL---QGN.STG-ES..........R---N-------Q-S---H---AI-F-----P-------N--N---FMP.K-SK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHGRDN   294
MAC.US.x.251_BK28      YSTDL---QGN.STD-ES..........R---N-------Q-S---H----I-F-----P-------N--N---FMP.K-SK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHGRDN   297
MAC.US.x.EMBL_3        YSTDL---QGNSTDN.ES..........R---N-------Q-S---H----I-F-----P-------N--N---FMP.K-SK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHGRDN   297
SMM.SL.92.SL92B        YSRDL---QG-....-ESS.........R---N-------Q-S---H---AI-Y-----P-------N--N---FAP.N-SK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHGGSN   298
SMM.US.x.H9            YSRDL---QNSNGNETDS..........K---N-------Q-S---H---AI-X-----P-------N--N---FAP.N-TK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHXRSN   298
SMM.US.x.PGM53         YSRDI---QSTTSRN.ES..........K---N-------Q-S---H---AI-F-----P-------N--N---FAP.N-SK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHSRSN   300
SMM.US.x.SIVsmH635FC   YSRDLI--QPA.NG-.ES..........K---Q-------Q-S---H---AI-F-----P-------N-SN---FAP.K-SK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHGNSN   298
MNE.US.x.MNE027        YSADL---QGN.STEDES..........R---N-------Q-S---H---AI-F-----P-------N--K---FMP.N-SK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHSKDN   296
STM.US.x.STM           YSSDLI--QN..VTGEES..........R---R-------Q-S---H---AL-F-----P-------N--N-T-FAP.N-SK-V--S---MME-QT--WF-F--TRAE.---Y-Y........WHGRDN   294
DEB.CM.99.CM40         ..................DNKDGSGN.RTG--K---D---TQS-ETSRFKP-KI--------G----D-KNFN-TGL..-N---AV---N--H-MA--WVQF---DEERAEELHI..IRKEV-GEVQ-G   221
DEB.CM.99.CM5          ...............TENQTNTNKSI.RHG--N---N---TQS-ETSRFKP-KI--------G----D-RNFN-TGL..-T---AV---N--H-MA--WMQF---DEERAEDIHI..IRK...G.LE-K   211
RCM.GA.x.GAB1          YSRDLW--KET.NST-STKK.........K-FVR----TS-QQF-EPK--EP--------P-F---V-K-KN-T-FDT..-V---ATS--HM-----AS-F-L---INV.-E-W-Y........QRRQSN   293
RCM.NG.x.NG411         YSRDLW-KAGNNNT...R...........Q-FIR----TS-QQF-EPK--EP--------P---I-K-L-AN-T-FDT..-Q---ATS--HM-------SF-L---ISE.---W-Y........QRKQSN   287
DRL.x.x.FAO            RYDD---.KC-N-TESPESNNCTNKKR..T--VTS--ST--TQD-N-AST-MMKF-L---P--I----NEKLNISS-...-K-I-AVQ---HMPA-I-SHF-F--TKHGQDELIET.....RG-PDFL-H   305
MND_1.GA.x.MNDGB1      RGDD-K-.....-NN..............T---N---E--NT-D-Q-GLLIRCI-G.-V-P--VM--YNEKLNNNKL...-S-ISAVQ--QHLVA---SFF-F--TMHKEGELIPI.DDKYRGPEEFHQR   312
MND_2.CM.98.CM16       RYED-T-...TNGTN-TY...........S---TQ------TQD-N-AST-EIKF-L---P--V----LERLNVSK-...-T-I-AVQ--QPLPA-T--MC-SS-TKHDY-ELIQTNT..KKG-EEFHDH   289
MND_2.GA.x.M14         RYED-T-...TKKENDTE.......TE..E---T-------TQD-N-AST-KMTF-L---P--I----REKLNKTKL...-A--SAVQ--DPMPA-I--MF-F--TKHN.YDEL-L.....VNPQKEFHDH   285
MND_2.x.x.5440         RYED-T-...TKLNKTGSATNSTEPEY..E---TS--AT--TQDRN-AST-RMTF-L---P-FV--K-NEKLNKTKL...-G--SAVQ--APLPA-I--MF-F--TKHDYDELIQTNP..RKG-DEFHDH   309
GRV.ET.x.GRI_677       Y---L--NNKTGSEKGSK...........D-------D-------------L-V-----------K-N-K--R-FAP.K-K--S-VH--------IT----L---RSE.------........--G--D--   284
TAN.UG.x.TAN1          Y---L---KENNTT.GTR...........--------D-------E-----L---------F-I-K-K--N-T-FGV..----S-VS--G-M-----SAF-----QAE.------........--H-V--N   292
VER.DE.x.AGM3          N-A-IM-KK.-..NN..........SN.RE-------D-------------EL---------F---K-N-Y--A-F-T.N-S--S-VH--N------T--LLL----SE.---Q--........--HRV---   301
VER.KE.x.9063          N-A-IY-KN.SSSSNS.........S-.KE-------D-------E-----EL---------F---K-H-H--N-Y-Q.N-S--S-VH--G-M-------LLL----SE.---Q--........--HRVN--   301
VER.KE.x.AGM155        N-A-IF-KR.ST.SH-.........G-.KE-------D-------------EL-------------K-N-W--A-F-P.E-S--S-VH--T-M----T--LLL----SE.---Q--........--H-V---   296
VER.KE.x.TYO1          N-A-IY-KN.KT..NS.........TS.KE-------D-------------QL-------------K-N-E--N-Y-Q.N-S--S-VH--G-M----T--LLL----HE.---Q--........--HRVN-N   295
TAL.CM.00.266          .............................FY-ATG--V---AQ--E--HYQA-PIQ------F--IK-N-SPWT-RGK..-H-ISAVH--DP-T-IAA-WLIL--TKEN.HTQV-H........QAPK-E   272
TAL.CM.01.8023         ...........................D.SYY-ATG--L---AQ--E--HYQA-PIQ------F--IK-N-NPWT-RGK..-H-ISAVH--DP-T-IAA-WLVL--TKEN.QTQI-H........QAPK-E   288
MON.CM.99.L1           ............................SHY-ILN----Y-TQ--E-SNYEPVP-H----P-------D-PAFT-QGS..-S--SAVT--HA-QPI-A-WFQL-STGN-.PN-TVM.......M-KQK-E   277
MON.NG.x.NG1           ...........................D.DYY-LLN---XYXTQ--E-SNYEPIPIH----P---I---D-PXFT-QNN..-X--SAVT--HG-MPI---WFQL-ATSAT.-X-TVM.......M--KR-E   288
MUS_1.CM.01.1085       SEND.........................SY-ITN----Y-TQ--E-SSFQPVPIQ-------S--K-N-ANFT-QGE..-N---ARH--HG-LPL-A-WLQL--T-LEG-N-AVM.......M--DK-E   293
LST.CD.88.447          RYD-TT-DKN-..................T---T---D-I-QQD-N-GVISNAYF-L------M--K-NEKLNFSAN...-T-I-ATP--DYMISS--SFF-F--TNHT.DDEL-PLTP.KKM.GNLLEA   302
LST.CD.88.485          RYD-TT-DKN-..................T---T---D-I-QQD-N-GVISNAYF-L------M--K-NEKLNFSAN...-T-I-ATP--DYMIS---SFF-F--TNHT.EDKL-PLT..PKKMGNLLDA   301
LST.CD.88.524          RYD-TT-N-N-..................T---T---D-I-LQD-N-GVMSNAYF-L------M--K-NEKLNFSAN...-T-I-ATP--DYMISS--SFF-F--TNHT.EDEL-PLT..PKKMGDLNGA   304
LST.KE.x.lho7          RYDD-T-SGERENR...............T---T---D-I-TQD-N-GVMQNAYF-L------M----NEQLNFSK-...-E-I-ATP--GYMLSS--SFF-F--TNHT.RDEL-PLT..PNKMEDLNGA   297
SUN.GA.98.L14          R-T--T-....NGND..............T---T---D-I-TQD-H-GIMQNAYF-L------M----DEKLNATK-...-K-I-ATP--NYMTS---SFF-F--TRHK.EDEL-PIN...NKVGQGAEG   300
SYK.KE.x.KE51          NES..........................YY-LQ------MRN--E-QTFRP-PIQ----P--S--K-N--EFE-DAE..-Q---AVS--HPF--LA--WFQL--T-K-KDKVRFI.......RQKDK-E   282
SYK.KE.x.SYK173        YKDDLMK--GN...............S.SYY-LL-------SA--E-QTFQP-PIQ----P--S--K-N--NFE-DDV..-T---AVS--QEF--LA--WFQL--T-K-.KDKVRFIK....--.DKNES   276
COL.CM.x.CGU1          N-SNS...........I...........RAT-KD-ANYTVTQV--M-IV-PV-TGF--AP--M----D-KKWD-TG-..-N---AVS--HEF-I--MSHVLV-A-KEL.SDWAKD.....REGVWKNDS   280
```

```
                                      |-            V3 loop                  -|
                      ^         ^^^  *    ^^^                                 *                             ^                             ^^
SAB.SN.x.SAB1C        SVIIRLNRYFNLTIRCRRPGNKTVLPVTIMA.GLVF.HS......QKYNTRLKQAWCWF.GGNWRGAWKEVKETIVRLPPKKYSGTN.............................DTNKIFLQRQWGDP   384
H1B.FR.83.HXB2        AKT-IVQLNTSVE-N-T--N-N-RKRIR-QR.-PGRAFVTI....G-I.GNMR--H-NISRAK-NNTL-QIASKLREQFGNNKTII...................................-K-SSG---   369
H1B.US.90.WEAU160     AKN-IVQLNVSIE-N-T--N-N-RKKI-L...-PGRVLYTT....GEIIGDIRR-H-NLSRTS-NNTL-QIV-KLREIKQF-NKTI..................................V-K-SSG---   373
H1N.CM.95.YBF30       HSNLLVQWNETVP-N-T----N-GGQ-Q-...-PAMTFYNI....E-IVGDIR--Y-NVSKEL-EPM-NRTR-E-KKILG-NN..................................ITFRARE-NE--L   368
H1O.BE.87.ANT70       GKN-IVTLNST-NMT-E--QI.DIQEMR-...-PMAWY-MG..IGGTAGNSSRA-Y-KYNATD-GKIL-QTA-RYLE-VNNTG.................................SI-MT-NHSSG--L   360
H1O.CM.91.MVP5180     AKN-IVTLNTPINMT-I-E-IAE-QDIYT...-PMRWR-MTLKRSNNTSP-SRV-Y-TYNKTV-EN-LQQTAIRYLN-VNQTEN.................................VTI--SRTSG--A   367
CPZ.CD.90.ANT         VLVRFGKEFE----T-I----R--R.NLQI..-PGMTFYNV....EIATGDTRK-F-TVNKTL-EQ-RNKTEHVLAEHWK-VDNK--...............................A-TIWTF-D---   373
CPZ.CM.01.SIVcpzCAM13 THDWIVQLSTAVNLT-K-V--N.TRGKVQI..-PGMTFYNM....DHIFGDTRK-F-ELN-TT-NETLQK-R-SLIKEIKANAN--Y...............................-IT-EPSSG---   382
CPZ.CM.05.SIVcpzEK505 ..TLLVQLNESIP-N-T-----.TRGQVQI..-PGMTFYNI....ENIIGDTR--Y-EVNRTWEQIWNTTKQII-NN........RK...............................-IT-IPNPG--L   353
CPZ.CM.05.SIVcpzLB7   ..KVLVQLNTSIH-N-T-V----IKG.IPL..-PGQLFYGT....ETVVGDTRE-H-EINQTH-YKILNQ--RELTTVFNESNK..................................TVH-ANSSG---   358
CPZ.CM.05.SIVcpzMB66  ..T-IVKLNETVRLN-T-T--N-IKG.IPI..-PSQIFYGI....ETVIGDTR--F-QLNKTV-TNTF-K-RHALNETYKG.-L-NE...............................TIT-GPSTG--L   391
CPZ.CM.05.SIVcpzMT145 .NLAIIQLKDPVK-T------N.TRGQIQI..-PAMTFYNI....ENVVGDTRK-Y-EIN-TQ-AK-LN-T--VLRNILR-.....................................-IS-MVPSG---   358
CPZ.CM.98.CAM3        GYXFLXXFQXXVS-N-T----X.SRGQIQI..-PGMTFYNI....ENIVGDTRR-Y-QINRTV-DERLN-TGQALRE-FTNLTQVNFTVS...................................PG---   369
CPZ.CM.98.CAM5        KI--IQ.FKEGIN-I-M----N.SRGQLQI..-PGVSFYNI....ENIVGDPR--H-NVSKAQ-MKKLN-TALA-AKYDKSIAKV-FTPN...................................PG---   365
CPZ.GA.88.GAB1        TDVWIVQLVEAVSLN-H----N.TRGEVQI..-PGMTFYNI....ENVVGDTRS-Y-KIN-TT-NRTVE---KALATSSNRTAANITLNRA..................................SG---   367
CPZ.GA.88.GAB2        -DTVIVQLKRAIP-N-S----N.TGRAINL..SPGTTFFNT....EALIGNPRK-S-HLN-TL-NNILNRI-QK-KNSTTWHRGDIT..................................-TKHPG---   370
CPZ.TZ.01.TAN1        NTPLLVKFNVSINLT-E-T--N-RGQ-Q-...-PGMTFYNI....ENVVGDTRK-Y-SVNATT-YRNLDWAMAA-NTTMRARNETV................................QQT-QWQRD---   367
CPZ.US.85.CPZUS       AD--LVQFSEGVNMT-I----N--GN-QL...-PGMTFYNI....P-IVGDVRE-H-NISKLT-EKQR-YTL-I-KKEANLTKVELIPN....................................AG---   349
GSN.CM.99.CN166       -IV-GFGEDYH-NLT-I------IRNLQ-...-AGMTFYSQ....VIVGGNTRK-Y-RLDPDR-NR-IR-AMKAMNEHWEN-TG..............................RN--QIRWTSEPK--L   385
GSN.CM.99.CN71        -IV-GFGEDYQ--LT-I------LKNLQ-...-AGMTFYSQ....IIVGGDTRK-Y-KLNHTK-DI-IRQAMKAMKNHWE-INND-T...........................PNK-QIRWTSEPK--L   390
H2A.DE.x.BEN          RT--S--K-Y---M--K---------I-L-S.----.--......-PI---PR----R-.--R--E-MQ---Q-L-QH-RY-GINDT.............................GKINFTKPGAGS--   381
H2A.GW.x.ALI          RT--S--KHY---MH-K-------V-I-L-S.--I-.--......-PI-K-PR------.K-E--K-MQ-----L-KH-RY-GTNDT.............................NQINFTKPGRGS-A   378
H2A.SN.x.ST           RT--S--KFY---VH-K-------V-I-L-S.----.--......-PI-R-PR------.K-E-KE-M----L-LAKH-RY-GTNDT.............................EKIRFIAPGERS--   375
H2B.CI.x.EHO          RT--S--S-Y---MH-K-----M-V-IRTVS.-IL-.--......-PI-K-P-------.K---TE-IQ------KNH-RYSGTTNI.............................SQIRLAEHARSS--   379
H2B.GH.86.D205        RT--S--T-Y--S-H-K-------V-IRTVS.--L-.--......-PI-K-PR------.K---TE-I----R--IKH-RY-GGAK-.IT..........................SVKLVSEHGKGS--   383
H2G.CI.x.ABT96        RT--S--K-Y----X-----------I---S.----.--......-PI-E-PR------.--K-KE-MQ-----V-KH-SY-GTNDT.............................KXITFTTPGEGS--   387
H2U.FR.96.12034       RT--S--K-Y---MY---------------S.----.--......-AI-NNP-------.----K--IQ------ANH-RYHGTKNI.............................SQINLAEPAG.--S   382
MAC.US.x.239          RT--S--K-Y---MK---------------S.----.--......-PI-D-P-------.--K-KD-I----Q---KH-RYTGTNNT.............................-KINLTAPGG.---   386
MAC.US.x.251_1A11     RT--S--K-Y----K---------------S.Q---.--......-PI-D-P-------.R-K-KD-I----Q---KH-RYTGTNNT.............................-KINLTAPGG.---   385
MAC.US.x.251_BK28     RT--S--K-Y---MK---------------S.----.--......-PI-D-P-------.--K-KD-I----Q---KH-RYTGTNNT.............................-KINLTAPGG.---   388
MAC.US.x.EMBL_3       RT--S--K-Y---MK---------------S.----.--......-PLTD-P-------.--K-KD-I----Q---KH-RYTGTNNT.............................-KINLTAPGG.---   388
SMM.SL.92.SL92B       RT--S--K-Y----K---------------S.----.--......-PI-E-PR------.--K--E-MQ---K---KH-RYTGTNDT.............................RKINLTAPGG.---   389
SMM.US.x.H9           RT--S--K-X---M----------------S.----.--......-PI-E-P-------.--E-KE-IR---XXL-KH-RY-GTNXX.............................XQI-LTAPGG.---   389
SMM.US.x.PGM53        RT--S--K-Y--------------------S.--I-.--......-PI-E-P-------.--S--E-IQ-------KH-RYTGTNET.............................KKINLTAPGG.---   391
SMM.US.x.SIVsmH635FC  RT--S--K-Y---MK---------------S.----.--......-TI-K-P-------.----SE-IQ-----L-KH-RYTGTNDT.............................RKINLTAPAG.-N-   389
MNE.US.x.MNE027       RT--S--K-N---MK---------------S.----.--......-PI-E-P-----R-.E---KE-I----Q---KH-RYTGTNNT.............................-KINLTAP-G.---   387
STM.US.x.STM          RT--S--K-Y---MS---------------S.----.--......-PI-E-P-------.--E----I------L-KH-RYTGTNDT.............................AKIR-VAPGG.---   385
DEB.CM.99.CM40        -IT--VPAKY---LT-V------YRAIHMAT.--S-......YTTFIQRL-I-R-H-RL.N-S-AN-T--MRQK-LEIFG-ANRTN-.................................LTIHYPK--R   310
DEB.CM.99.CM5         TITV-IPPKY--SLD-V------YRA-HMAT.--S-......YTTFIPRL-I-R-H-RM.R---T--V--IRSRLKEIY...NVTK..................................-TVHYPG--R   296
RCM.GA.x.GAB1         RTV-G--SFY--SVT----S-R--KGISLAT.-VFI.-.......LRVEK-P-G---R-.E---TD-------RVKTTKGYRGTSNT.............................DKIKIRTVYG.--D   383
RCM.NG.x.NG411        RTV-G---H----VT----S-R--KGISLAT.-V.-.I-......LRVQK-P-G---R-.D-D-K--------A-NKTKGY-GTRQP.............................ENIT-RSVYG.--D   377
DRL.x.x.FAO           KYVF-V-KKW--K-I---K--RSIISTPSAS.--L-Y-G......LEPGKN--RGM-K-.E-Q-GK-LGSLAKELKSVSGSIWRN-T.NSTCISLKGKKIQDRNVTGCML......KQL-VDNYTTK--L   420
MND_1.GA.x.MNDGB1     KFVYKVPGKYG-K-E-H-K--RS-VSTPSAT.--L-Y-G......LEPGKN--KGM-T-.K-R-GL-LWSLAKELNK-NDSI.KV...NQTCKNFTSTGEENKQNTDKQKEFAKCIK-L--DNYTTS--R   430
MND_2.CM.98.CM16      KYVY-VDEKY--QVV---K--RS-ISTPSAT.--I-YSG......LEPGKN--KGM-QL.K-Q-GI-MHDLAIELRKIDSSIWRNVT.KGYCKGLKRK....ENRTGCAL......K-I-VSDYTTK-E-   400
MND_2.GA.x.M14        KYVY-V-KKWG-KVV-D-K--RSIVS-PSAS.--I-Y-G......LEPGRN--KGM-RL.I-Q-GR-LNALSRELKKVNASI-R-F-ASGPCRGINNR..GVVNKTGCAL......K-IEVSNYTTI-E-   399
MND_2.x.x.5440        KYVY-VDKKWG-QV----K--RSIISTPSST.--L-Y-G......LEPGKN-RKGK-QL.E-K-GQ-LHSLSLELRKINDSI-KDNH.NMTCKSSNNK....KNTTGCHL......K-IS-SESTVK-E-   420
GRV.ET.x.GRI_677      T---K--KFY---V-----------------.----.--......------------H-.Q-D-K------R-EVKKVKNLTEVSIE..............................NIHLRRIW..---   373
TAN.UG.x.TAN1         ----K--KHYK-K-V----------------.----.--......-Q---K-R----H-.Q-D-K---R--RK---E--KE--R---.............................N-RQ-W-S------   384
VER.DE.x.AGM3         --LVLF-KHY---VT-K--------------.----.--......-R-----R----H-.Q-----------NE--K--KDR-Q---.............................--EE-Y---LF---   393
VER.KE.x.9063         ---VLF-K-Y---VT-K--------------.----.--......-----K-R----H-.Q-R--E-------E-IK--KDR-R---.............................--E--Y----F---   393
VER.KE.x.AGM155       --L-L--KHY---VT-K--------------.----.--......-------R----H-.Q---K------Q-E--K--KER-Q---.............................----------F---   388
VER.KE.x.TYO1         T-L-LF-KHY--SVT----------------.----.--......----MK-R----H-.E-------R---QK--E--KDR-K---.............................N-EH-Y--------   387
TAL.CM.00.266         -A--L-GKEH----T-I----R-IKDLQ-A-.--M-.--.....QIIAGKN-NR-Y-KL.E---TK-MESAN-E-AKRYKEHINN-D.............................KV-ITWTAVRR--L   365
TAL.CM.01.8023        -A--L-GKEH-I--Q-I----R-IKDLQ-A-.--M-.--.....QIIAGKD--R-Y-KI.E---TR-MED-N-K-AKAYKDHINSSS............................GTAQIRWQGVRK--L   382
MON.CM.99.L1          -IVV--AKHLHVN-T-I------IRNLQ-G-.-MT-Y.-.....QLIVGGNTRK-Y-KVNKTQ-ET-LQA-H-AVKTEWE--NN---VTTIS.............................WRFQPQ--K   371
MON.NG.x.NG1          -IV---XKTLSVN-T-I------LRNLQ-G-.-MT-Y.-.....QIIVGGNXRK-Y-QVNKTQ-DA-LHAAHXA-XXEWES-NNNVTVANIT.............................WRYQPK--N   382
MUS_1.CM.01.1085      -IG-KFGEN-RVNLT-I------IRNLQ-G-.-MT-Y.-.....QLIVDGDTRR-Y-KINKTQ-DP-LRQAMLAMRKHWHGRLQKQ-HSSINETKIKIR......................WRSEPK--Q   394
LST.CD.88.447         KFVYKVAGKWG-I---I-K--RSEVSTISST.--L-YYG......LEHGS--RL-Q-K-.E-Q-GRMFHNLGNMLK--NATAMNY-H.GN.CIDNT.........KPCGRQ...LKGLPIANMT-RG-NL   410
LST.CD.88.485         KFVYKVSGKWG-I---I-K--RSEVSTISST.--L-YYG......LEHGS--RL-Q-K-.E-Q-GRMFNNLGKMLQK-NATAMNY--.GQ.CPEGK.........QPCGRH...LKGLPIANMT-RGEGL   409
LST.CD.88.524         KFVYKVAGKWG-I---I-K--RSEVSTISST.--L-YYG......LEHGS--RL-Q-K-.E-Q-GRMFNNLGRMLKK-NETAMNYKE.GN.CSTPG.........KPCGRQ...LKGLPIANMT-RGV-L   412
LST.KE.x.lho7         KFVYKVAGKWG-I---I-K--RSEVSTISST.-YL-YYG......LEHGS--RL-Q-K-.E-Q-GRMFNNLGKMLKE-NAEAMNY-EGTGTCDSKK.........TTCGRK...LKGLPIANMT-HGA-L   407
SUN.GA.98.L14         EYVWKVAAKWG-V-Q-I-K--RSQVSTISST.--L-YYG......LEPGSK-RL-Q-K-.V-Q-GR-FATLGKLLRQVE-LANMSNG....CTFDNIT..KTCRFTNGTDF...KKMIKFKPMEKHGA-A   413
SYK.KE.x.KE51         ----LVPEKLG-QLV-E----ESIKNIQLA-.-FFLPVIQG...RL-TGKAA-R-F-KV.K-E-GKFFEQ-HNESIKVWKNVT-TS...................................WRSQPQ--L   372
SYK.KE.x.SYK173       .---LVPEALR-Q-I-E----ESIKNIQLA-.-YFLPVIQG...KL-TGRDA-R-F-RV.T---TEFF-Q-H-QATKT.W-NVTN..................................TTWRSQPG--L   365
COL.CM.x.CGU1         GT-EYYWFPKDIALG-I-R--SSHRNLNTAN.-AK-YYE......LIPYSKGIYGR-Q-..VPMT-QN-KN-TQFAIEIK-NLTAWLERISRKNI......................-ITPRNGNRTS--   379
```

```
                                     |-               V4 loop                  -|                                           |-    V5 region
                              *      *                          *  ^^^  *       *                             *^^^                     ^^^^^^
SAB.SN.x.SAB1C        ESEFFFFNCQGEFFYCKMDWFLNYLNNKSVDP......DHNNCAKNNTKPCWQRTYVPCHIRQVVNDWYTLSKKTYAPPREG...HLECNSTATALYVELNYNSK.........NRTNVTLSPQIESI.   494
H1B.FR.83.HXB2        -IVTHS---G------NSTQLF-STWFN-TWST..EGSNNTEGSDT.........ITL--R-K-II-M-QKVG-AM----IS-...QIR-S-NI-G-LLTRDGGN......SNNESEIFRPGGGDMRDN.   478
H1B.US.90.WEAU160     -IVMHS---G------NSTQLF-STWHANGTWK..NTEGAD--...........ITL--R-K-II-R-QEVG-AM----I--...QIR-L-NI-G-LLTRDGG-S.....EENQTEIFRPGGGNMKDN.   480
H1N.CM.95.YBF30       -VTHLM---R------NTSKLF-EELLNETGEP.....................ITL--R---I--L-TRVG-GI----IR-...V-N-T-NI-G-VL-YSGGP.......DTKETIVYPSGGNMVNL.   465
H1O.BE.87.ANT70       -VTHLH---H------NTAKMF--TF.SCNGTT..CSVSNV..SQG--G.......TL--KL----RS-IRGQSGL----IK-...N-T-M-NI-GMILQMDNTWNS.SNNNVTFRPIGGDMKDIWRTEL   474
H1O.CM.91.MVP5180     -VSHLH---H------NTSGMF--TFINCTKSG..CQEIKGS-ETNK-G.......TI--KL--L-RS-MKGESRI----IP-...N-T-H-NI-GMILQ-DQPW......NSTGEN-LRPVGGDMKD-.   478
CPZ.CD.90.ANT         -VKVHW---------DITPWF-ATYT..GNLI..TNGAL................IAH-R-K-I--H-GIV--GI-LA--R-...NVS-T-SI-GIML.EGQIY........NETVKVSPAARVA.DQ.   469
CPZ.CM.01.SIVcpzCAM13 -IANHM---G------DTRKMF-..........ESEPFHE..............NMTI--R---I--S-MRVGRGI----IP-...-IT---LI-G-ILTR.DHV.......NNT-N-FRPIGGDMKN-.   476
CPZ.CM.05.SIVcpzEK505 -VTNLMI--G------NTSQLFTNQ-..........GNTTG..............NITLQ-R---I--L-TRVG-GI----IK-...PIN-L-NI-GIILD........YTKSGTEKY-IYPTGGDMTNL.   448
CPZ.CM.05.SIVcpzLB7   -VANLH---G------NTSALF-DTL..LRNKTLLNETLHN.............NLTL--R---I--L-LRVG-GIF---IR-...NIR----I-G-IL-.........KHTNTD-I-FRPIGGDMTD-.   460
CPZ.CM.05.SIVcpzMB66  -VTNLHLI-G------NTSILF-TSI.......IFNETK-D.............NITI--R---I-RL-QRVGRGIFL--IR-...TIN-I-NI-GILFA.....QQ.KTDRMNKSAMFTPVGGEMRNN.   491
CPZ.CM.05.SIVcpzMT145 -VTNHH---G------NTSEII-ITK.........INKTEN.............M-II--R---I--S-MRVG-GIF---IR-...NIT-T-NI-GMLL-IHK-REDQGEDQDQN-TYVCLTGGNMKD-.   462
CPZ.CM.98.CAM3        -VTNMM---G------NTTTLF--TWKNNNITK..GDNTT................FF--R---I--S-MRVG-GI----IR-...VIS-T-NV-GIIL-TGHGI.....NNSIT-I-LYPTGGNMVDL.   472
CPZ.CM.98.CAM5        -ITNMM---G------NTTPLF-KF-YTCDNQN..NCTRNNT-DTNSDNSTL....TIS-R---I--S-MRVG-GI----IR-...NIT-I-NI-G-LL-AGGKWE....N...NTI-VYPSGGEMRDL.   478
CPZ.GA.88.GAB1        -VTHHM---G------NTSQIF..TD-.....I..TNGII................IL--R---I-SS-MRVGRGI----IR-...NIT---NI-G-LLTSDTPVT....N.NSG-L-FRPTGGNMKD-.   463
CPZ.GA.88.GAB2        -VVN-M---G------NTSRLI..TC-.....S..SDTSE...............YIL--K------S-MRVG-GIF----R-...TIT----I-G-LL-VQN........GTGN-TEVYLSGGDMRD-.   464
CPZ.TZ.01.TAN1        -VTS-W----------NLTNWT-TWTA.NRTNN..THGTL................VA--RL--I--H-GIV--GV-L---R-...TVK-H-NI-G-IMTAEKDNN..........NSYTPQFSAVVEDY   465
CPZ.US.85.CPZUS       -VVNMML--G------NTIPLF-MTY-....NT..DNTTI................TLK-R---I--Q-MRVG-GIF---IK-...V-S---NI-GMILDISISAV...NN.DSR-I-VMPTGGDMTAL.   449
GSN.CM.99.CN166       -VQTHW-Q--------NLSILFQLN--.TINSS...NIGN.......ITSKYKGQWLA-K---F-TQ-GYV--SI-L---Q-...-IN-T-NI-G-LIDGAMYES.........SINMTPSA.DVADA.   489
GSN.CM.99.CN71        -VQTHW-Q--------NLSVLFQFN--ITAVDD..TNINN.......VTSKYKGQWMA-R---F-TQ-GYV-RSI-L---Q-...-IN-T-NI-G-LIDGAMYGH.........SINMTPSA.DVADA.   496
H2A.DE.x.BEN          -VA-MWT--R---L--N-T----WVED-NQTRR.....................N-..---K-II-T-HKVG-NV-L------...E-A-E--V-SIIANIDIDKN........RTH--I-F-AEVAEL.   475
H2A.GW.x.ALI          -VVYMWT--R---LH-N-T----WVE--TGQEQ....................HN-A----K-II-I-HKAG-NV-L------...E-T----V-S-IANIDTDG..........-Q--I-F-AEVAEL.   473
H2A.SN.x.ST           -VAYMWT--R---L--N-T----WVE-RTNQTQ....................HN------K-II-T-HKVG-NV-L------...Q-T----V-SIIANIDGGE..........-Q--I-F-AEVAEL.   470
H2B.CI.x.EHO          -VRYMWT--R---L--N-TF---WVE-RTGLKR.....................N-AS-----I--T-HKIGRNV-L------...E-S----V-S-IANIDWID-.........-L--I-V-AEVSEL.   474
H2B.GH.86.D205        -TTYMWT--R---L--N-T----WVE--TNTTR....................-N-A-----II-T-HKVG-NI-L------...E-S----V-S-IANI-SD..........NST--ISV-AEVSEL.   478
H2G.CI.x.ABT96        -VK-MWT--R---L----T----WIEDRNMTLL.............KPQERQR-N--------II-T-HKVG-NV-L------...D-T----V-S-IANIEE-RSN........-H--IIF-AEVAEL.   491
H2U.FR.96.12034       -VK-MWT--R---L----T----WVE-R-MEGQ.............TSRERHK-N---------I-T-HKVGRNV-L------...D-T----V-SMIANIDRIN-.........-Q--ISM-AEVAEL.   485
MAC.US.x.239          -VT-MWT--R---L----N----WVEDRNTAN.............QKP-EQHK-N--------II-T-HKVG-NV-L------...D-T----V-S-IANIDWIDG.........-Q--I-M-AEVAEL.   489
MAC.US.x.251_1A11     -VT-MWT--R---L----N----WVEDRNTAN.............QKP-EQHK-N--------II-T-HKVG-NV-L------...D-T-I--V-S-IANIDWTDG.........-Q--I-M-AEVAEL.   488
MAC.US.x.251_BK28     -VT-MWT--R---L----N----WVEDRD-TT.............QRP-ERHR-N--------II-T-HKVG-NV-L------...D-T----V-S-IANIDWTDG.........-Q-SI-M-AEVAEL.   491
MAC.US.x.EMBL_3       -VT-MWT--R---L----N----WVEDRD-TTQ.............RP-ERHR-N--------II-T-HKVG-NV-L------...D-T----V-S-IANIDWTDG.........-Q-SI-M-AEVAEL.   491
SMM.SL.92.SL92B       -VT-MWT--R---L----N----WVEDRNTSSP.........RWTTQTK-EQHK-N--------II-T-HRVG-NV-L------...D-T----V-S-IANIDWIDN.........-E--I-M-AEVAEL.   497
SMM.US.x.H9           -VT-MWT--R---L----N----WVEXIKNGS..........RWTSQ-Q-ERY-KN------X-II-T-HKVG-NVXL------...D-T----V-S-IA-IDWING.........-K--I-M-AEVAEL.   496
SMM.US.x.PGM53        -VT-MWT--R---L----N----WVEDRNTSGS.........LWTQQTKADEKK-N--------II-T-HRVG-NV-L------...D-T----V-S-IADIDWIDG.........-Q--I-M-AEVAEL.   499
SMM.US.x.SIVsmH635FC  -VT-MWT--R---L----N----WVEDRDQNSN.........RWKQQKKSEQQK-N--------II-T-HKVG-NV-L------...D-T----V-S-IA-IDWINN.........-E--I-M-AEVAEL.   497
MNE.US.x.MNE027       -VT-MWT--R---L----N----WVED-NLTG............TTQKPQEQHK-N--------II-T-HKVG-NV-L------...D-T----V-S-IANIDWIDG.........-Q--I-M-AEVAEL.   492
STM.US.x.STM          -VT-MWT--R---L----N----WIE-R-TSE.........MRDWNK-K-EQQK-N---------I-T-HKVG-NV-L---Q-...D-T----V-SIIANIDWTNN.........-E--I-A-AEVAEL.   493
DEB.CM.99.CM40        -VQSVW-Q-H------NISKA-DL-................LLQNNTRNSTWSDKWLM--R-N-L--T---VGQHI-L--K--...E-K-S-HIS-FVFDVDHYNG.........SITL-P-AD-RAV.   410
DEB.CM.99.CM5         -VQNTW-Q-H------NVSKAF..N................LLMN-T-NITWSNN-LM--K-N-L--T-FKVGQHV-L--K--...E-K-H-SVS--VF-VAHYGQ..........NM-I-P-SD-KAV.   394
RCM.GA.x.GAB1         -ARY-WL--N---L---LN----L---ET-G................T-NEKRKAPF---ITKMI------V-R-V-T---PD...A-K-SAQVSY-LADID-.........INDSE--I---AEVGDY.   484
RCM.NG.x.NG411        -AKC-WL--D---L----N---------TEGNK.................NEAR-AMF---ITKM-------V-R-V-T---PD...A-R--A-V-Y-LADID-TDN.........MT------AEVGD-.   478
DRL.x.x.FAO           A--NLMML-G--Y-F-NVSKIWRTW--RTSNVW....................YP-AS---K-IID--ARVGR-I-M--VS-FNNEIR-TQDV-EMFF-IQKLETE.ETGDSN.YQIKFIPQDEVQNQ.   526
MND_1.GA.x.MNDGB1     AA-MMMMT----M-F-NVTRIMRAW-DPNEKKW....................YP-AS-Q---I-D--MQVGR-I-L--TS-FNN-IR-THRV-EM-F-MQKIDSN.ET....KMQIKFLPPSETSNQ.   533
MND_2.CM.98.CM16      GA-TIMLL-G--Y-F-NWTKIWRTW-DQNSSVW....................YPWMS-N---IID--HKVG--I-M--VS-FNNEIR-SNDV-EMFF-VQKTEEG........YIIKFVPQDEVQNR.   500
MND_2.GA.x.M14        GA-TIMIL-G--Y-F-NWTRIWKTW--QNSNVW....................YP-MS-K---I-D--HKVGR-I-M--VS-FNN-IR-TNDV-EMFF-VQLVEG.......KRYLIKFLPQDEVQNQ.   500
MND_2.x.x.5440        GA-TIMLL-G--Y-F-NWTKIWKAW-S-QSSVW....................YP-MSRN---I-G--HKVG--I-M--VS-FNNEIR-TNDV-EMFF-VQKTDDD......NGYIIKFIPQDW-QNQ.   522
GRV.ET.x.GRI_677      --AN-W---------------I-----RTE-AE...GTNRTCDKG-PGPG--V-----A-----------V---V-------..------SV-----AID--N-S........GPI-------VR--.   488
TAN.UG.x.TAN1         -AANIWL---------TP---V-W---E-NSGRNVDVEGNNCTTG..KD-R-YK---------SI------------------..----T--V-SMM-S-D----..........E------TANL-N-.   499
VER.DE.x.AGM3         -AANLW--------------------RT---D..HNPCNGTKGKGKAPG--A-----A----S-I-------R----------..--Q-T--V-GMS--------.........-----------T-.   508
VER.KE.x.9063         -AANLW--------------------Q---AN..HNQCSNETKKG-APG--V-----A----S-I-----I---V--------..----T--V-GMT------NN.........Y----------G-.   508
VER.KE.x.AGM155       -AANLW--------------------LT--AD..HNHCKN-AGKGRSPG--V-----A----S-I-----I------------..--Q-T--V-GMT------NQ..........-----------T-.   503
VER.KE.x.TYO1         -ASNLW---------------------TW-AY..HNFCSSKK.KGHAPG--V-----AY---S-I--S---------------..--Q-R--V-GMT--------..........-------------.   501
TAL.CM.00.266         -VQL-WHA--------NISSMFLNTSYEYNYTSNQGKLHQGVIGK.KPRSSDRTKF-MS-QL--MI-R-SRVE-LM-L------...--Q-K-NI-G-L-DI-HHRGS.......SDI.T-EP-S--RDS.   482
TAL.CM.01.8023        -VSL-WHA--------NISSMFMKT-YTYNYTAEKGVLHQEKIGRQEHRYSDRTKFFMS-QL--MI-R-SKVE-LM-L------...--S-K-NV-GIL-DI-HYKGN........DIIS-EP-S--RDA.   500
MON.CM.99.L1          -VQTHW----------NVSALFINRRTNKT-G.......ISPFDVN-KPNTTYHGGWLA-T---M-TQ-GYV--SI-L---K-...-VQ-T-NI---LITGELYQN..........NVTLVP-A-VSDS.   480
MON.NG.x.NG1          -VQTHW----------NLSRLFXSSDPXKNKN....ETFESTFNVST-ASTAYNGEWLT-RL--L-TQ-GYI--SX-L---Q-...-VQ-T-NI---L-DGELYGS..........TVTLIP-ARVSDA.   494
MUS_1.CM.01.1085      -VQTHW---G------NLSVLFQFV-HTEINE..........SNIHTV-TKYHKDQWMV-R---F--Q-GYV--SI-L---K-...LVR-E-NI-G-LIDGAMYRD..........SI-M-P-ANVLDA.   500
LST.CD.88.447         AT-MLMHT-GE-M-F-NVTRIFQEW---NS-KW....................YPWAN---KS-ID--ASVG--I-L--TS-FNNRIR-THRV-EMWFQMEKFEPE.ENLGGNLS.VRFLPPSWETNQ.   516
LST.CD.88.485         AT-MLMHT-GE-M-F-NITRI-QEW---NS-KW....................YPWAN---KS-ID--A-VG--I-I--TS-FNNRIR-THRV-EMWF--EKWEPE.ENLGGNLS.VRFLPPSWETNQ.   515
LST.CD.88.524         AT-MLMHT-GE-M-F-NVTRIFQEW---NS-KW....................YPWAN---KSIID--A-VG--I-L--TS-FNNRIR-THRV-EMWF-MEKWEPH.EDLGGNLS.IKFLPPSWETNQ.   518
LST.KE.x.lho7         AT-MLMHT-GE-M-F-NVTRIFQEW---NS-KW....................YPWAN---KSIID--A-IG--I-L--TS-FNNRIR-THRV-EMFF-MEKWEPH.EDLGGNLS.IKFLPPSWETNQ.   513
SUN.GA.98.L14         AT-MLMMT-GE-M-F-NLTRIFKVW-DTTSNKW....................YPWAN---KS-ID--ASVG--I-L--TS-FNNRIR-ANRV-EAWFT-ERVEDW.KVNGSNISVVAFQPPTNTLNQ.   520
SYK.KE.x.KE51         -VRTHW-Q-G------NVSKIFANVS-GYAN-.....SNYA...........KNL-LS-A---II-Y-GYVT-LM-L--T--...-IK-T-NI--VLTDIE-YPN.........SQL-FAPTANV-DV.   472
SYK.KE.x.SYK173       -VRTHW-Q-G------NVSKLFANIT-GNASKN.....NYA...........SNLRLS-A---II---RYVR-LI-L--TA-...-IK-T-NV--VLTDIE-YPG.........STL-F-PTANV-DV.   465
COL.CM.x.CGU1         -AT-T-VI-HRL----NASSLWKHDS...........................PVMN-T--KL--S-V-HARIL-G--PG-...--Q--WEKQPVIAFMGTIEG...DNDGNGC.AYPAAPNFKHAL.   472
```

```
                                                gp120 \ / gp41
SAB.SN.x.SAB1C         WANELGDY.KLVEIKPIGFAPTKVRRYTG......PERQKR.VPFVLG..FLGFLGAAGAAMGAAATALTVQSQQLLAGILQQQKNLL.AAVEQQQQMLKLTIWGVKNLNARVTALEKYLEDQARLNIWG   613
H1B.FR.83.HXB2         -RS--YK-.-V-K-E-L-V----AK-RVVQ.......-E--A.VG.I-AL---------ST----SMT----AR---S--V---N---.R-I-A--HL-Q--V--I-Q-Q--IL-V-R--K--QL-G---   597
H1B.US.90.WEAU160      -RS--YK-.-V-K-E-L-V----AK-RVVQ.......-E--A.VGM--AM---------ST----SMT----ARL--S--V---N---.R-I-A--HLFE--V--I-Q-Q---L-V-R--K--QL-G---   600
H1N.CM.95.YBF30        -RQ--YK-.-V-S-E---V--G-AK-R-VS.......-E--AAFG.--AL---------ST----SIT----ART--S--V---NI--.R-I-A--HL-Q-S---I-Q-Q-K-L-I-R--R--QI-SL--   585
H1O.BE.87.ANT70        FNYKVVRV.-PFSVA-TRI-RPVISTR-H........-E--A.VG.--ML---V-S---ST------T-A--THT--K--V---D---.R-IQA---L-R-SX--IRQ-R--LL---TL-QN-QL-SL--   592
H1O.CM.91.MVP5180      -RTK-YN-.-V-Q---FSV----MS-PIINIHT..-H-E--A.VG.--ML---V-S---ST----------RTHSV-K--V---D---.R-IQA--HL-R-SV--IRQ-R--LQ---TLIQN-Q---L--   602
CPZ.CD.90.ANT          -RA--SR-.QV---X-LSV---TXK-PEIKQHS....----GI.G.I-LF---L-S---ST----SI---A-TRN-XH--V---A---.Q-I-T--HL-Q-SV----Q-Q--ML-V----R--QL-SL--   591
CPZ.CM.01.SIVcpzCAM13  -RS--YK-.-V-R-E-LSV----AK-.HTVG....ER----A.A-G--AL---------ST-----VT----AR---S--V---N---.R-I-A--HL-Q--V----Q-Q--LL-V-R--Q--QI-GL--   598
CPZ.CM.05.SIVcpzEK505  -RQ--YK-.-V-S-E---V--G-AK-H-VT.......----A.A-G--AL---------ST----SIT----ARK--S--V---N---.R-I-A--HL-Q-SV--I-Q-Q---L-I-R--R--QI-GL--   568
CPZ.CM.05.SIVcpzLB7    -RS--YN-.-I-K-E-L-V---RA--R-VD.......-E--A..AG--VF---------ST----SLT----AR---T--V---S---.R-I-A--HL-Q-SV--I-Q-Q---L-I-R--K--QL-G---   579
CPZ.CM.05.SIVcpzMB66   -RS--YK-.-V-R-E-L-V----AK-R-VH.......-E--A.AVG--AL---------ST----SLT----AR---S--V---S---.R-I-A--HL-Q-SV--I-Q-Q---L-V-R--K--QL-GL--   611
CPZ.CM.05.SIVcpzMT145  -RS--YK-.-I---Q-L-V----S---AVEKQHH...-E--A..LG--AL---------ST----SVV----AR---T--V---N---.R-I-A--HL-Q-SV--I-Q-Q---L-V-R--R--QL-GL--   585
CPZ.CM.98.CAM3         -RL--HK-.-V-S-E---V--S-AK-H-VS.......-E--AAFG.--AL---------ST----SVV----AR---S--V---N---.R-I-A--HL-Q-SV--I-Q-Q---L-V-R--R--QI-GL--   592
CPZ.CM.98.CAM5         -RL--NR-.-V-S-E---V----AK-H-VA.......-E--AAFG.--AL---------ST----SVV----ARH--S--V---N---.R-I-A--HL-Q-SV--I-Q-Q---L-V-R--K--QI-SL--   598
CPZ.GA.88.GAB1         -RS--YK-.-V-R-E-LSV----A--H-VARQ...KD----AAFG.--AL---------ST-----VT----AR---S--V---N---.K-I-A--HL-Q-S-----Q-Q--LL-V-R--Q--QI-GL--   587
CPZ.GA.88.GAB2         -RS--YK-.-I-K-E-L-V----AK---VAKA...LD-S--AAFG.--AL---------ST----SVM----ARN--S--V---N---.R-I-A--HL-Q-SV--I-Q-Q---L-V-R--K--QV-AL--   588
CPZ.TZ.01.TAN1         -KV--AR-.-V---Q-LSV--RPGK-PEIKANH..T.-SR-D-.G.I-LL-----S---ST----SI---A-ARG--S--V---Q---.Q-I-A--HL-Q-SV--I-Q-Q--ML-V---IR--QL-SL--   588
CPZ.US.85.CPZUS        -K---HK-.-V-S-E---V--G-AK-H-VK.......-E--AAFG.--AL---------ST----SVV----AR---S--V---N---.R-I-A--HL-Q-SV--I-Q-Q---L-V-R--K--QI-GL--   569
GSN.CM.99.CN166        -NY--SR-.-V---D-LSM---PAK-KEHP......AVE--A.LS.--IS--T--S---TT---------E--RS-----M---E---.R---A--SL-QPSV--I-Q-QT-LSS-----R--TI-QA--   609
GSN.CM.99.CN71         -KY--SR-.-V---D-LSL---SAQ-RPHPGV....HVK--A.IS.--IS--T--S----T------------RS-----V---E---.R---A--SL-Q-SV--I-Q-Q--LSS-----R--TI-QA--   618
H2A.DE.x.BEN           YRL-----.--I--T-------DQ---SS.T....-V-N--G-F.---.-----AT--S----RSLT-SA--RT-----V---QQ--.DV-KR--E--R--V--T---Q-----I----KH--Q--S--   595
H2A.GW.x.ALI           YRL-----.-----T-------SE---SS.T....-R-N--G-F.---.-----AT--S---T--LT-SA--RT-----V---QQ--.DV-KR--E--R--V--T---Q-----I----K------S--   593
H2A.SN.x.ST            YRL-----.--I-VT-------P-K--SS.A....-V-N--G-F.---.-----TT--------SLT-SA--RT-----V---QQ--.DV-KR--E--R--V--T---Q-----I----K---Q--S--   590
H2B.CI.x.EHO           YKL-----.-----T-------SIK--SS.V....TP-N--G-L.---.-----AT--S-----SLT-SA--RT-----V---QQ-V.DV-KR--EL-R--V--T---Q-----I----K---Q--S--   594
H2B.GH.86.D205         YRL-----.-----T-------D----SS.V....KP-N--G-M.---.-----AM--S----TSLT-SA--RT-----V---QQPV.DV-KR--EL-R--V--T---Q-----I----K---Q--S--   598
H2G.CI.x.ABT96         YRL-----.-----T-------N-K--SS.V....TPKN--G-F.---.-----AM--S---T-SLT-SA--RT-----V---QQ--.DV-KR--E--R--V--T---QT----I----K------S--   611
H2U.FR.96.12034        YRL-----.-----R-------P-K--SS.V....TP-N--G-F.---.-----AT--S-----SLT-SA--RT-----V---QQ--.DV-KR--E--R--V--T---QT----I----K---S--A--   605
MAC.US.x.239           YRL-----.-----T---L---D-K---TGG....TS-N--G-F.---.-----AT--S-----SLT--A--RT-----V---QQ--.DV-KR--EL-R--V--T---QT----I----K---Q--A--   610
MAC.US.x.251_1A11      YRL-----.-----T---L---N-----TGG....TS-N--G-F.---.-----AT--S-----SLT--A--RT-----V---QL--.DV-KR--EL-R--V--T---QT----I----K---Q--A--   609
MAC.US.x.251_BK28      YRL-----.-----T---L---D-K---TGG....TS-N--G-F.---.-----AT--S-----SLT--A--RT-----V---QQ--.DV-KR--EL-R--V--T---QT----I----K---Q--A--   612
MAC.US.x.EMBL_3        YRL-----.-----T---L---D-K---TGG....TS-N--G-F.---.-----AT--S-----SFR--A--RT-----V---QQ--GVVKR--ELLRLTV.--T---QT----I--------Q--A--   612
SMM.SL.92.SL92B        YRL-----.-----T---M---H-K---T.S....TSKN--G-F.---.-----AT--S-----SLT--A--RT-----V---QQ--.DV-KR--EL-R--V--T---QT----I----K---Q--S--   617
SMM.US.x.H9            YRL-----.-----T-------S-K---TTG....AS-N--G-F.---.-----AT--S-----SLT-SA--RT-----V---QQ--.DV-KR--EL-R--V--T---QT----I----K---XX-S--   617
SMM.US.x.PGM53         YRL-----.-----T---L---S-----TGA.....S-N--G-F.---.-----AT--S-----SVTRSA--RT-----V---QQ--.DV-KR--EL-R--V--T---QT----I---RK---Q--S--   619
SMM.US.x.SIVsmH635FC   YRL-----.-----T---L---N-----TTG....AS-N--G-F.---.-----AT--S-----SLT-SA--RT-----V---QQ--DVVKR-HE.L-R--V--T---QT----I----K---Q--S--   618
MNE.US.x.MNE027        YRL-----.-----T---L---N-K---TGG....TP-N--G-F.---.-----AT--S-----SLT--A--RT-----V---QQ--.DV-KR--EL-R--V--T---QT----I----K---Q--A--   613
STM.US.x.STM           YRL-----.-----T---L---N-K---T.S....TS-T--G-F.---.-----AT--S-----SLT--A--RT--T--V---QQ--.DV-KR--EL-R--V--T---QT----I----K---Q--S--   613
DEB.CM.99.CM40         -RAD-FK-.-II-V-------SA----E-.P...ESV-H--AAGIAF-..LVA--ST--------S-------RS--S--V---QE--.K---AHG-L-T--A---R---T-L--I----K---K--E--   532
DEB.CM.99.CM5          -RAD-YK-.-II-V----M---SI---E-.P...ESA----AAGIAF-..LVA--ST--------S-------RS--S--V---QE--.K---AHGHL-S--A---R---T-L--I----K--SK--E--   516
RCM.GA.x.GAB1          --A---R-.-AI--R-------EIK--QT........K---.--L---.-----S---T-------------RH--------------.DI-KR--NL----V--T---Q-----I----A--SL--TF-   601
RCM.NG.x.NG411         --A---R-.-V---M---Y---N----ET........K---VPL.---.-----S---T-------------RH-----------K--.DI-----EL----V--T---Q-----I----A--SL--TF-   595
DRL.x.x.FAO            FTAVGAH-.---KVE-------D-H--NLPG.....H----GA.V---..I--L-SL--S--CSV---M-----A--T-MVE---Q--.RL-----EL-----------Q--L----E-IG---M-SL--   646
MND_1.GA.x.MNDGB1      FVAYGAH-.---K-M---I---D-K-H-LPE....HHKE--GAV.I--..I--L-SL--S---SVSV-------S-VT--VE---Q--.KLI---SEL----------QT-L-S--N-IK---L-SQ--   654
MND_2.CM.98.CM16       FTAVGAH-.Q--KVD-------E-A--HLPE.....A----GAV.L--..MF-L--L--ST--SV-V-------A--N--VE---V--.SLID-HSEL----------Q--L---D-VA-----SM--   620
MND_2.GA.x.M14         FTAVGAH-.---KVD-------D-H--HLPD.....AK---GAV.L--..M--L--L--S---SV-V-------A--N--VE---I--.SLID-HSEL----------Q--L---D-VA--S--AV--   620
MND_2.x.x.5440         YTAVGAH-.---KVD-------DIH-HHLPN.....T----GAV.L--..M--L--L--S---SV-V-------T--N--VE---V--.SLID-HSEL----------QV-L---E-VA--S--SV--   642
GRV.ET.x.GRI_677       --Y-----.-----T-------D------PT......-E--VPF.---.---------T-------T-----RH--------------.--------L---------------------------S--   607
TAN.UG.x.TAN1          --Y---R-.--I--E-------EI---V-PT......-E--VPF.---.--------------T-----------------------.-------------------------------T---L--   618
VER.DE.x.AGM3          --A---R-.-----T-------E------.G....HD-T--.------.---------T-------------H--------------.----A--------------------------------A--   628
VER.KE.x.9063          --A---R-.-----T-------N------.G....QD-S--.------.---------T-------------H-----M--------.----A--------------------------------V--   628
VER.KE.x.AGM155        --A---R-.-----T-------E------.G....Q----.------.---------T-------------H--------------.---GA-------------------------A------A--   623
VER.KE.x.TYO1          --A---R-.-----T-------E------.G....H----.------.---------T------SS-----RH--------------.----A--------------------------------S--   621
TAL.CM.00.266          -RAAVSR-.-VI--V-F-Y---NI---D-P.....ANK---AA-LA--..-I---ST--T----V---------RS--S--V---EH--.R-I-H--HL-Q--V--I------L-----------K--S--   603
TAL.CM.01.8023         -RAAVAR-.-VI--T-------NI---D-P.....ATKE--AA-LA--..-I---ST--T----V---------RS--S--V---EH--.R-I-H--HL-Q--V--I------L---------------S--   621
MON.CM.99.L1           -RS--SR-.-V---D-LSM---TAQ-R--V......H-E--..AIT--MA-----ST--GT------------RS-----V---E---.R--TA--SL-Q--V----Q-Q--L--V--FIK--TL--A--   600
MON.NG.x.NG1           -RS--SK-.-V---D-LSV---AAQ-RH-L......M-ER-..AIT--MA-----ST--GT------------RS-----V---E---.R--TA--SL-Q-SV--I-Q-Q--L--V--FIK--TL--S--   614
MUS_1.CM.01.1085       -RM-PKKVQ-V---D-LSM---E-Q-RPVP......HVHA-..AIS--IT-----S---GT------------RS-----V---A---.R---A--HL-Q-SV--I-Q-Q--L-----FIK---L--L--   621
LST.CD.88.447          FVATGAH-.--IRLR-------D-H--AP......SSK---AA-LA--..A--L-S---T---LVS-I----A-VVIQ-------Q--.VL--K--EL-R---------Q--L--I-E--K--TL-AS--   636
LST.CD.88.485          FVATGAH-.--IR-R-------AEHQ-AP......SSK---AA-LA--..A--L-S---T---LVS-I----A-VVIQ-------Q--.VL--K--EL-R---------Q--L--I-E--K---L-AS--   635
LST.CD.88.524          FVATGAH-.--IRLR-------SEH--AP......RG--T-AA-LA--..A--L-S---T---LVS-I----A-VVIQ-------Q--.VL--K--EL-R---------Q--L--I-E--K---L-AS--   638
LST.KE.x.lho7          FVA-GSK-.--IKLN-------DEH--AP......RG--T-AA-LA--..A--L-S---T---LVS-I----A-AV-Q-------Q--.VL--K--EL-R---------Q--L----E-VKH--L-AS--   633
SUN.GA.98.L14          FVSTGAH-.---R-R-------DEH--AP.......RKE--AA-VA--..A-AL-S---T---LVS-I----A-AV-Q-------Q--.VL--K--EL-R---------Q--L----E-VQ--SL-AS--   639
SYK.KE.x.KE51          -RAD-FN-.--IR---------SQ---ELP......TK---AA-LA--..---L-S---T---S------L---T-----V---QK--.E---A--HL-G--V---------L----T--R---IMSN--   592
SYK.KE.x.SYK173        -RAD-FN-.--IQ---------DQ---ELP.....NT-E--AA-LA--..---L-S---T---G------L---T-----V---QK--.E---A--HL-G--V---------L----T--R---I-SN--   586
COL.CM.x.CGU1          STL---R-.---KMRTTTYV--DIK-SVNVNWH..HG----GIF...AFSI-AL-SG------S-SV---I-A-S-NGRASASSNRM-LKL--T-SAL-Q--V------QV--ATI-G---E--K-ASI-   596
```

```
                      *     *        ^^              ^    ^^^
SAB.SN.x.SAB1C        CAFRQVCHTTVLWKYNN..............TPDWENMTWQEWERQIEKYEANISRILEQAHEQEQKNLDSYQKLVSWSD.FWSWFDLTKWFGWMKIAIMVIAGIIVARVLLVIIGILRKFRKGYAPLS   727
H1B.FR.83.HXB2        -SGKLI-T-A-P-NASWSNKS.........LEQI-NHT--M--D-E-NN-TSL-HSLI-ESQN-QE--EQELLE-DK-AS.L-N--NI-N-LWYI-LF--IVG-LVGL-IVFAVLS-VNRV-Q--S-.L   715
H1B.US.90.WEAU160     -SGKLI-T---P-NASWSNRS.........QDYI-N----M----E-NN-TGL-YNLI-ESQN-QE--EQELLE-DK-AS.L-T---ISN-LWYI--F--IVG-L-GL-IVFTVLS-VNRV-Q--S-.L   718
H1N.CM.95.YBF30       -SGKTI-Y---P-NETWSNNTS........YDTI-N-L---Q-DEKVRN-SGV-FGLI---Q--QNT-EK-LLE-DQ-DS.L----GI---LWYI-----IV---VGI-IISIV-T-IARV-Q--S-.L   704
H1O.BE.87.ANT70       -KGKL--Y-S-K-NRTWIG...........NESI-DTL-----D---SNISST-YEEIQK-QV-QEQ-EKKLLE-DE-AS.I-N-L-I---LWYI----IIVGALVGV--IMIVLN-VKNI-Q--Q-.L   708
H1O.CM.91.MVP5180     -KGKLI-Y-S-K-NTSWSG........RYNDDSI-D-L---Q-DQH-NNVSSI-YDEIQA-QD-QE--VKALLE-DE-AS.L-N---I---LWYI----IIVGAL-GI--IMI-LNLVKNI-Q--Q-.L   721
CPZ.CD.90.ANT         --DKVT-----P-NNSWVNFTQTC.AKNSSDIQCI---------D-LVQNSTGQ-YN--QI----QER-KKELYE-DK--S.L-N---I-Q-LWYI--F--IVGA-VGL-I---LVSC---V-Q--H-.L   718
CPZ.CM.01.SIVcpzCAM13 -SGKSI-Y---P-NKTWSG.........KSMSDI-N-L---Q-DKL-TN-TGT-FGL--E-QS-QE--EKDLLE-DQ-AS.L-N---I-N-LWYI--FL-AVG---GL-IIMSVVSVI-RV-Q--S-.L   716
CPZ.CM.05.SIVcpzEK505 -SGKS--Y-N-P-NTTWSN........NNSYDTI-G-----N-DE-VRN-SGV-FGL----Q--QSI-EK-LLE-DQ--S.L-N---I---LWYI--F---V---VGI-IISI-MSMVARV-Q--S-.L   687
CPZ.CM.05.SIVcpzLB7   -SGKLI-T-S-P-NRTWSN.........KTYNEI-D----M--D-EVRN-TEI-YGLI---QD-QEN-EKKLLE-DH-TS.L-N---ISH-LWYI--F--I-G-L--C-IIFAVLA-VNRV-Q--S-.L   697
CPZ.CM.05.SIVcpzMB66  -SGKLI-T-S-P-NTTWTN.........KSYDDI-Y----MQ-DKEVSN-TDV-YNL--K-QT-QEN-EKELLE-DK-AS.L-N---I-S-LWYI--F-IIVG-L-GL-IVFALLS-VNRV-Q--S-.L   729
CPZ.CM.05.SIVcpzMT145 -TGKTI-P-A-R-NKTWGNI........SDYQVI-N-Y---Q-D-EVNN-TGL-YTL--E-NT-QE--EKELLE-D--AN.L-----I-N-LWYI-MFLI-VG---GL-ICFA-GSLINRV----S-.L   704
CPZ.CM.98.CAM3        -SGKAI-Y---P-NNTWSAN........TSFDEI-N-L---D-DKRVKN-SGV-FSLI---Q--QNT-EK-LLE-DQ--S.L-N---I-R-LWYI-LF--IV--LVGI-IVGA--SFVA-V-Q--S-.L   711
CPZ.CM.98.CAM5        -SGKAI-Y---P-NATWSAN........TSYDEI-N-L---D-DKKVKN-SGV-FSLI---Q--QNT-EKDLLE-DQ--S.L----NI-Q-LWYI--FLI-V--L-GF-LIGIVMSVIA-V-Q--S-.L   717
CPZ.GA.88.GAB1        -SGKA--Y---P-NNSWPG.S.......NSTDDI-G-L---Q-DKLVSN-TGK-FGL--E-QS-QE--ERDLLE-DQ-AS.L-N---I---LWYI--FL-AVG---GL-IIMTVFSVV-RV-Q--S-.L   706
CPZ.GA.88.GAB2        -SGKT--YS--P-NTSWNSNK........SYEDI-S-L---Q-DKLV-NHTGT-FQL-QR----QNS-EKELLE-DQ--S.L-N---I-N-LWYI-MF-ILVG-V-GL-IVIGVVN-I-RS-Q--S-.L   707
CPZ.TZ.01.TAN1        --NKL---SS-P-NLTWAEDSTKCNHSDAKYYDCI-N-L-----D-LV-NSTGT-YSL--K-QT-QE--KQELLE-DK--S.L-D---I-Q-LWYI----IIV--LVGL-I-MF-VNVVKQV-Q--T-.L   716
CPZ.US.85.CPZUS       -SGKTI-Y---P-NDTWSNN........LSYDAI-G-L-----D-KVRN-SGT-FSLI---Q--QNT-EK-LLE-DQ--S.L-N---I-N-LWYI--FLI-V-SLVGI-IVG--FSLVA-V-Q--S-.L   688
GSN.CM.99.CN166       --N-PI---I-P-NTSWANGS..........L---------K-SMLV-NDTYT-QQL----NQ-QAS--NELM--SK-DS.L-----ISD-QRYI--FVI-V-AL-AL-IVMF-LNM--RI-Q--S-.L   726
GSN.CM.99.CN71        --NQPI---I-P-NDSWAKNS..........-----H------SKL--NDTYT-QQL--N-NH-QS--MNDLL--SK-DS.L-----ISN-LWYI--F---V-ALVAL-IIMFVLNM--RV-Q--S-.L   735
H2A.DE.x.BEN          -----------P-VNDS.............LS---K--------K-VRYL-----QS--E-QI-QE--MYEL---N--DI.LGN-----S-VKYIQYGVHIVV---AL-IAIYVVQL-SR-----R-VF   710
H2A.GW.x.ALI          -----------P-VN-S.............LK---D--------Q-VRYL-----EQ--R-QI-QE--TYEL---N--DV.-TN-L---A-VKYIQYGVYI-V--VAL-IVIYVVQM-SRL----R-VF   708
H2A.SN.x.ST           -----------P-VNDT.............L----N--------QR-RNL-----ES----QI-QE--MYEL---N--DV.-GN-----S-IKYIQYGVYIVV----L-IVIYVVQM-SRL----R-VF   705
H2B.CI.x.EHO          -----------P-VNES.............LK---N-----Q----VRFLD---TKL--E-QI-QE--MYEL---NQ-DI.-SN---F-S-MAYIRLGLYIVI--V-L-IAIY--QM-ARL----R-VF   709
H2B.GH.86.D205        -----------P-PNET.............L--N-N-----Q--K-VHFLD---TAL--E-QI-QE--MYEL--IN--DV.-GN-----S-IKYIHLGLYIV--LV-L-IVVY-VQM-ARL----R-VF   713
H2G.CI.x.ABT96        -----------P-DALGANKT.........LE-Q-N--------K--NFL-D--T-L--E-QI-QE--MYEL---N--DV.-GN-----S-VKYVYLGLY-V--V--L-IVIYVVQ--GRL-Q--R-VX   730
H2U.FR.96.12034       -----------P-INDT.............L--N-D--------EKVNYL-E--TQL--A-QI-QE--MYEL---NN-DI.-GN-----S-VKYVYLGLY-V----IL-IVIYVVQL-GSL----R-VF   720
MAC.US.x.239          -----------P-PNAS.............L--K-N-E-------KVDFL-E--TAL--E-QI-QE--MYEL---N--DV.-GN----AS-IKYIQYGVYIVV-V-LL-IVIY-VQM-A-L-Q--R-VF   725
MAC.US.x.251_1A11     -----------P-PNAS.............L--K-N-E-------KVDFL-E--TAL--E-QI-QE--MYEL---N--DV.-GN----AS-IKYIQYGVYIVV-V-LL-IVIY-VQM-A-L-Q--R-VF   724
MAC.US.x.251_BK28     -----------P-PNAS.............L---N-D-------KVDFL-E--TAL--E-QI-QE--MYEL---N--DV.-GN----AS-IKYIQYG-Y-VV-V-LL-IVIY-VQM-A-L-Q--R-VF   727
MAC.US.x.EMBL_3       -----------P-PNAS.............L---N-D-------KVDFL-E--TAL--E-QI-QE--MYEL---N--DV.-GN----AS-IKYIQYG-Y-VV-V-LL-IVIY-VQM-A-L-Q--R-VF   727
SMM.SL.92.SL92B       -------------PNDS.............LV---N--------KKV-FL----TQM--E-RL-QE--MYEL---N--DV.-GN-----S-VRYIQYGVFLVI--VLL-IVIYVVQM-SRL-Q--R-VF   732
SMM.US.x.H9           ---------X-P-PNDT.............L--X-N--X-----K-VNFL----TXX--E-QI-QEX-MYEL---NXXDX.XGN-X---X-IKYIQYGVLIVL-V-GL-IVIYVVQM-ARL-Q--R-VF   732
SMM.US.x.PGM53        -----------P-PNAS.............LV-N-N----------VDDL----TQA--E-QI-QE--MYEL---N--DI.-GN-----S-IKYIQYGVLIVL-VVGL-IVIYVVQMIARL-Q--R-VF   734
SMM.US.x.SIVsmH635FC  -----------P-PNDS.............LV-N-D--------GKVDFL----TQL--E-QI-QE--MYEL---N--DI.-GN-----S-IRYIQYGVLIVL-VVGL-IVIYVVQM-ARL-Q--R-VF   733
MNE.US.x.MNE027       -----------P-PNAS.............L--N-N-E-------KVDFL-E--TAL--E-QI-QE--MYEL---N--DV.-GN----AS-IRYIQYGVYIVV-V-LL-IVIY-VQM-A-L-Q--R-VF   728
STM.US.x.STM          -----------P-PNDS.............LV---N---------KVDFL----TQL--E-QV-QE--MYEL---N--DV.-GN-----S-VRYIQYGVYLVI-LVML--AIY-MQL-ARL----R-VF   728
DEB.CM.99.CM40        ---K-I-----P-NNSL.............ED---D--------MKVAN-TDEWEGA-QR-Q--QER-VHAL-S-QD-DS.L-N----SR--W-IRLVVYI--AL-LL-IAMFGVN-GS-LCR--S-.L   646
DEB.CM.99.CM5         ---K-I-----P-NHTW.............GE---N---------KVAN-TDEWEGA-QR-Q--QER-VHAL-S-TD-DS.L-N----SR--W-IRLVVYI--AL-LL-IAMFGVN-GT-LCR--S-.L   630
RCM.GA.x.GAB1         --W------V-P-TF..............NK--E-QKES-LQ---N-SYL---TIA-QE-QD-HE--VHELE--SN-G-.AF--LN-DW-MQYI--GFFIVI---GL--AWLLWNC-SNL-Q--R-.L   714
RCM.NG.x.NG411        --W--------E-I-SQ..............--E-NKQ--L----N-SRL-G---VA-QD-Q--HER-VHDLE--N--G-.ML--LNMDW-LKYIR-G-FI-L---GL-IIFLLWSCVS-L-A--T--L   709
DRL.x.x.FAO           -S-A-----N-V-PNES.............V--N-TSE--M--QKRVDSISN--TLD-QK-Y------IFEL---GDLTS.WAN---F-W-SKYI--GFFIVMA--GL-I-AALWSTVGR--Q--R--P   761
MND_1.GA.x.MNDGB1     -SWA-----S-E-TNTS.............I--N-TSE--K---TRTDYLQQ--TEM-K--YDR--R-TYEL---GDLTS.WA----F-W-VQYL-WGVFLVL---GL-I--ALWNTISR--Q--R-.V   768
MND_2.CM.98.CM16      -S-A-----H-P-PNDS.............I--N-TSE--L--DKRVTALTD-MTVN-QK-Y-L----IYELE--GD-TS.WA----F-W-LKYV--GLLIVIV---L-I-ACLWSV-G---Q--R--P   735
MND_2.GA.x.M14        -S-S-----N-P-PNES.............I--N-TSE--L--D-RVTAITN-MTID-QR-Y-L----MYEL---GDLTS.WA------W-LKYV--G-LI-MVV-GL-I-ACLWSTIGR--Q--R--P   735
MND_2.x.x.5440        -S-S-----S-K-PN-S.............IV-N-TSE--L--D-RVNSIVT-MTID-QR-Y-L--R-IFEL---GDLNFHGLTG----W-LKYV--GLL-VVV--GL-M-ACLWSV-G---Q--R--P   758
GRV.ET.x.GRI_677      --WK-------P-----..............--K-D----L------NAL-G--TQL--E-QN--S----L----DD--G.-----S-ST-L-YV--GFL--VI-LGL-FAW-LW-CI-NI-Q--N--.   720
TAN.UG.x.TAN1         ---K-------P-TF--..............----D--------S--TAL-G---TT-VK-Y------M-T----GD-TS.W-NI--VSS--W-I-WGFYIVI-L-LF-MAWL-W-CIARV-Q--F---   732
VER.DE.x.AGM3         --WK-------P-QW--.............R---N----L------SYL-G--TTQ--E-RA--E----A----S----.------FS--LNIL--GFLDVL---GL-L-YTVYSCIARV-Q--S-.L   742
VER.KE.x.9063         --WK-------P-QWQ-.............M--N-Q----L------GEL-G--TEQ-VK-R---E----A--R-T---N.------FS--LNIL--GFL-VV---GL-L-YT-YSCIGRV-Q--S-.L   742
VER.KE.x.AGM155       --WK-------P-TW--..............--E-N----L---K---GL-G--TKQ----R---E----A----SD--S.------FS--LNIL--GFLAVI-V-GL-L-YTLYTCIARV-Q--S-.L   736
VER.KE.x.TYO1         --WK-------E-PWT-.............R----Q----L------ADL-S--TGQ-VK-R---E----A----T----.------FS--LNIL-MGFL--V---GL-L-YTVY-CIVRV-Q--V-.L   735
TAL.CM.00.266         --WK-I-Y-S-P-NKT..........WSNYTD-Q-Q--------MKVDNHTGL--QL-QE-Q--QER-VHEL---ND-DS.L----N-SA--R-LR--VI-V-SL-LL-IVMY-VNV-KLL-Q--L--P   721
TAL.CM.01.8023        --WK-I-Y-S-P-NKT..........WTNSTN---Q--------KLVDNASDT-TVL-QE-Q--QER-VHEL---ND-DS.L----N-SA--R-LR--VI-V-SL-LL-IVMY--NV-KLL-Q--L--P   739
MON.CM.99.L1          --NKA------P-NNS..........WAKGHF-E-D-----Q-SELVDNDTMT-QQL--A-Q--QG--QHELM-PGQ-DF.L-N---IS--LWYI--F-I-V-AL-GL-I-MF-L-VISRLGQ--SL--   718
MON.NG.x.NG1          --N-AX---Q---NNT..........WAKGHF-E-D-----Q-SMLVDNDT-L-QXL--E-Q--QG--AHELM--GQ-DW.L-N---IS--LWYI--F-I-V-ALVGL---MF-L--I-RVGQ--SL--   732
MUS_1.CM.01.1085      --N--I---R-P-NDS..........WANHTQ-G-------Q-S-LVDNDTTT-QEL--L-QR-QEE-QHKL---LE-DS.L-E---IS--LWYI--FC--V--LVLF-LVMFVL----RV-Q--S---   739
LST.CD.88.447         -QWK-----N-E-N--..............I--N-TRD--I--D--VGVL-----TL-QE-YTT-LE-RNAFK--QEFN..--N-L-ILS--QYI-Y-VLI-I----L--VSF-VQNIV-MCR--RV.L   748
LST.CD.88.485         -QWK-I---N-E-N--..............I--N-TRD--I--D-RVGVL-----TL-QE-YTT-LE-RNAFK--QEFN..----L-ILS--QYI-Y-VLI-I----L-IVSF-VQNIV-MCR--RV.L   747
LST.CD.88.524         -QWK-----N-P-N--..............V--N-TRD--I-----VGSL---TTL-QE-YTT-LE-RNNFK--QDFN..----M---T--QYI-Y-VLI-I---IL-I-SF--QSVV-MCR--RV.L   750
LST.KE.x.lho7         -QWK-----N-E-T--..............I--N-TKD--R---SKVAI-DK--TSL-QE-YTT-LE-QNKFK--QEFN..----L-ISH--TYV-Y-VLI-LV--GL---SF--QNVV-MCR--RV.L   745
SUN.GA.98.L14         -QWK-----N-P-N--..............I--N-TKD--M--D--VKM-DD--TAL-QE-YVT-LE-QNKFKQ-QEFN..----L--SQ--LYI-Y-VLI-GI--A--I-SF--QQIYRMCQ--RV.L   751
SYK.KE.x.KE51         ---K-I---A-T-QQA.......CGNNSRCP--Q------HT----VDNLTDH-DNL-RE-Q--QE--VHDLT--QE-DS.L------S---QYL--GFFA--A-VIL---SFAW--V-NMLG--S--L   713
SYK.KE.x.SYK173       ---K-I---A-T-EKACGNNSN......FCPK-Q-K----HR--QEVDNLTDH-DGL-RE-Q--QER-VHDLT--QE-DS.L------S---FYL--GFY--GALVLL-LVSFSV--IKNLLG--V-IL   708
COL.CM.x.CGU1         --NM-I-R-I-P-NKTW............GEEDP-Q----KQ-HERVRN-TDI-EAD-VE-YDLQEE-EKKLAE-GD-TN.WF-G-G-FNI-KYVLY-AY-VG-L-GL-IIM-V-ACI-GAFRVKGF.Q   711
```

```
                                                                          *        ^^^                                     *
SAB.SN.x.SAB1C          SLPSSH*...QIHIHLDK.GQPDKEEEQGSGGLNSDSRSYTWQREFLRHLCHRLITWLRNLTSWFSTIFSNLHR.........................CLQD.IQ...........QRTRQL...TA   811
H1B.FR.83.HXB2          -FQTHLPT..PRGPDR.....-EGI--E-GERDRDR-IRLVNGSLA-I.......WDDLRSLCL--YHRLRDLLLIVTRIVELLG.................RR.GW...........EALKYW....W   797
H1B.US.90.WEAU160       -FQTHLPA..PRGPDR.....-EGI--E-GERDRDR-GRLVDGFLT-I.......WVDLRSLCL-LYHRLIDLLLIAKRIVELLG.................RR.GW...........EALKYC....W   800
H1N.CM.95.YBF30         --QTLIPT..ARGPDR.....-EET-GGVGEQDRGR-VRLVSGFSA-V.......WEDLRNLLI-LYHRLTDSLLILRRTLELLG.........QSLSRG--L.LN...........ELRTH-....W   793
H1O.BE.87.ANT70         --QIPNHH..-EEAGT.....-GRTGGG-GEEGRPRWIPSPQGFLP-LYTDL-T-ILWTYHLLSNLASGIQKVISYLRLGLWILG.........QKIINV.............CRI.......C....-   797
H1O.CM.91.MVP5180       --QIPVPH..RQEAET.....-GRTG-E-GE-DRPKWTALPPGFLQQL.......YTDLRTIILWTYHLLSNLISGIRRLIDYLG.........LGLWILG-K.TI.....EACRL.......C....G   810
CPZ.CD.90.ANT           -FQ..IPTQN-QDPEQ.....-EEIR-E-GRKDRIRW-ALQHGFFA-L.......WVDLTSIIQWIYQICRTCLLNLWAVLQHL.........................CRITFRLCNHLE....N   801
CPZ.CM.01.SIVcpzCAM13   --QTLIPA..-REPGT.....LGETGGE-GEPGSGR-VRLVTGCWP-I.......WDDLRNLVIW-YRSLTSLACIVWRQLEKLG.........QLLI.....NILR........LIQEKL-LLR.....   805
CPZ.CM.05.SIVcpzEK505   --QTLIPT..TRGPDR.....-ERT--DAGELD-GR-VRLVSGFLA-A.......WEDFRNLLL-LYHRLTDCLSILRRTLELLR.........Q........NIHK.....GLQLLNELRIY-....W   776
CPZ.CM.05.SIVcpzLB7     -FQTLSPI..PRGLAQ.....-GGT--E-GEQGTDR-VRLLDGFLA-A.......WNDLRDLCL-LYHRLTDLVLIARRTLEIVG.........Q........YTLK.....GLRLVWEALLY-....Q   786
CPZ.CM.05.SIVcpzMB66    -FQTLIPA..RRDRDR.....-EEI--G-GEPD-VR-IRLVSGFLA-A.......WNDLRDLCL-LYHRLRDLLLIVLRTLELVG.........Q........TLLK.....GLRRGREALIH-....R   818
CPZ.CM.05.SIVcpzMT145   --QTLIPA..NQGPDG.....LGET-KG-GENVRGR-IRLVSGFLP-V.......WEDLRNLLS-LYHQLRDCASLIWRLLEILG.........Q........YSLR.....GVQQIGTLLH--....R   793
CPZ.CM.98.CAM3          --QTLIPT..TREPDR.....-GGT--DVGEPGKGR-IRLVSGFLA-I.......WEDLRNLLL-CYHRLRDLLLILGRTLENLG.........QSLNKG--Q.LR...........NFS-Y-....W   800
CPZ.CM.98.CAM5          --QTLIPT..TREPDR.....RGGT--D-GEPGRGR-TRLVSGFLA-V.......WEDLRNLLI-FYHRLIDLLSILWRTVQILG.........QSINKG--Q.LK...........DFSTY-....K   806
CPZ.GA.88.GAB1          --QTLIPV..-REQGR.....LGEID-G-GEQDR-R-VRLVEGCLP-I.......WDDLRNLGIW-YQSLTSLACNVWRQLKTLG.............HLILH.SL...........RLL-ERLCLLG   795
CPZ.GA.88.GAB2          -FQTLIPA..TGGPEV.....-EGIGGG-GEQDRGR-VRLVSGFLA-C.......WDDIRNLTI-LYRLLRDCLSVLWSLLEQLG.........QQVLRG-RL.LR...........ELLS--....K   796
CPZ.TZ.01.TAN1          FSQIPTQA..EQDPEQ.....-GGIAGG-G-RD-IRWTPSPAGFFSIV.......WEDLRNLLIWIYQTFQNFIWILWISLQALK.........QGIISLA.HSLV.....IVHRTIIVGVRQI....   811
CPZ.US.85.CPZUS         --QTLFPT..TREPDR.....-EGT--GAGKTD-VR-TRLVSGFLA-V.......WEDLRNLLI-LYHRLEDLLLILRRTVQILG.........QNINKG-.Q.LL..........NELRARC....W   777
GSN.CM.99.CN166         -PQILIPQAVERGP.......--AT--GVGAQGKVR-VR-LTGFSS-L.......WDDLRNLVIWIYQILASLAWTLRGIGK.F.LWEQLQKAIQKAIQQAR-L.RE...........VAA-......V   821
GSN.CM.99.CN71          -PQILTPQQVERDP.......--GTRKD-GAQG-VR-VR-LTGFSS-L.......WDDLRSLVIWIYQILATLVWTLQRIGK.F.LW....EQLQKAFQQAR-L.GE...........VAA-......-   826
H2A.DE.x.BEN            -S-PGYLQQIH--K.DRG.Q.-AN-GTEEDV-GD-GY..........DLWPWPI.NYVQF-IHLLTRLLIG-YNICRDLLSKNSP........TRRLISQS-TA.-R...........DWL-LK....-   801
H2A.GW.x.ALI            -S-PGYIQQIH--K.DQE.Q.-TRG-TEEDV-D-VGD..........-LWPWPI.AY-HF-IHLLARLLIG-YSICRDLLSRISP....ILQPIFRSLQRA-TT.-R...........DWL-LK....-   803
H2A.SN.x.ST             -S-PAYFQQIH--K.DRE.Q.-AR--TEEDV-NSVGD..........NWWPWPI.RYIHF-IRQLIRLLNR-YNICRDLLSRSFQ....TLQLISQSLRRA-TA.VR...........DWL-FN....T   800
H2B.CI.x.EHO            -S-P-YTQQIP-RK.DRG.Q.-AN--TEEG--N-EGY..........-SWPWQI.EYIHFPIRQLRDLLIW-YSGCRTLLSKTFQ...............T--P.VL...........-PL-LP....P   793
H2B.GH.86.D205          -S-P-YTQQIP-RK.DRG.Q.-AN--TEEG--NDG-Y..........-SWPWQI.EYIHF-LRQLRNLLIW-YNGCRTLLLKTFQ............ILHQIST.NL...........-PL-LP....V   800
H2G.CI.x.ABT96          -S-P-YVQQIH-RK.DQE.X.-T---IE-ES-NKGGY..........-SWPWQI.EYIHF-XRQLGNLLTW-YSNCRALLSRIXQ...............T--P.LFQR.ISRTLQAIREHL-LE....-   825
H2U.FR.96.12034         -SHP-YVEQIP-QR.DQE.Q.-T-G-IEEGA-DSGGY..........-SWPWQI.RYIHF-IHQLIRLLTW-YSSCRDLLSRICQ....SLQPLFQS........-R...........E-LHLE....I   808
MAC.US.x.239            -S-P-YFQQTH-QQ.DPA.L.-TR-GKERD--EGGGN..........SSWPWQI.EYIHF-IRQLIRLLTW-FSNCRTLLSRVYQ....ILQPILQRLSAT--R.-R...........EVL-TE....L   820
MAC.US.x.251_1A11       -S-P-YFQ*TH-QQ.DQA.L.-TR-GKE-D--EGGGN..........SSWPWQI.EYIHF-IRQLIRLLTW-FSNCRTLLSRVYQ....ILQPILQRLSAT-RR.-R...........EVL-TE....L   819
MAC.US.x.251_BK28       -S-P-YFQ*THTQQ.DPA.L.-TR-GKE-D--EGGGN..........SSWPWQI.EYIHF-IRQLIRLLTW-FSNCRTLLSRAYQ....ILQPILQRLSAT-RR.VR...........EVL-TE....L   822
MAC.US.x.EMBL_3         -S-P-YFQXTHTQQ.DPA.L.-TR-GKE-D--EGGGN..........SSWPWQI.EYIHF-IRQLIRLLTW-FSNCRTLLSRAYQ............ILQPILQ.RLSATLRR....IREVL-TE....L   822
SMM.SL.92.SL92B         -S-P-YHQQIH-QR.DQE.L.-A--GTE-E--NGGGY..........-LWPWQI.EYIHF-IRQLIR-LTW-YNNLTRLASRAYQ....NLQQLCQRLSEIS-P.-R...........ELV-RE....-   827
SMM.US.x.H9             -S-PAXVXQIP-XT.GQE.L.-T--G-E-D--XRGGN..........-SWPXQI.EYIHF-IRQLIRLLTW-FSSCRDWLLRXCQ....ILQPVLQSLSRT--R.AR...........EVI-VE....I   827
SMM.US.x.PGM53          -S-PVYVQQIP--.KGQE.Q.-T--G-E-D--DRGGN..........RSWPWQ-EYIHF-IRQLIRLLTW-FNNCRTYLLRIYQILQPVFQRLLQRIREVVRTE-S...........YLQY.......   831
SMM.US.x.SIVsmH635FC    -P-PAYVQQIP--K.DQE.P.-T--G-E-E--DRGG-..........-SWPWQI.EYIHF-IRQLIRLLTW-FSSCRDWLLRIYQ............ILQPVLQ.RLSRTLQR....VREVI-IE....I   828
MNE.US.x.MNE027         -S-P-YFQQTH-RQ.DQA.L.-T--GTE-D--DSGGN..........SSWPWQI.EYIHF-IRQLIRLLTW-FSNCRTLLSRAYQ....ILQPIFQRFSTT--R.VR...........EVL-TE....L   823
STM.US.x.STM            -S-P-CRQQIP--K.GQE.Q.-T--GTEEG--DRGGI..........NSWPWQI.EYTHF-IRQLVRLLTW-YNNFRACLSRIYQ....TLHPTFQRISRI--R.-R...........EVV-LG....-   823
DEB.CM.99.CM40          QTLT..PK..EEEVLR.....LAE-GGRSG-DGRARWTALPRGFFS-I.......WEDLQQLLTWLYQTFRNCIWLIRRGFELLQ..EGIYRLSFALWQRLRLLGER.............F-V-....G   740
DEB.CM.99.CM5           QTLT..PK..EVEVLQ.....LGG-GDKSGE-VK-RWTTSPRGFLR-I.......WDDISQLLTWLYQTLRNCAWLIRRGFELLQ..ETIYRLSFAFWQRLRLLGEH.............L-VV....W   724
RCM.GA.x.GAB1           -..PPSYVQQIHIHNTGEPQT-GEKR-D-GEEGGNKYNNWLREYCWIQLIHPLSRI-TQLSQICR-CSSIIFQSLRWILAKIQYG................WQ.........................   799
RCM.NG.x.NG411          -P-PCYHQQIP-R.NREE.Q.-I--G-K-GS-EGGGH...........KYYNWQRGYFYI-ILRPIE-LCRRLYQICSNSLSVIYQSLQSIPSRILLHLRWIGAKL-............DGWQE......   807
DRL.x.x.FAO             Y.IFKDAYR.PETTNY.H.R---R-KGEEQDRGEW.NIRSEPS...KPGSSKAWSEETVGTWLKE-RGYIW-KNVKAV........................-EYGWS........ELQEAGRGIC.....   847
MND_1.GA.x.MNDGB1       FSQDC......QQNLY.R.KR--NG--ESNSLELG.EHNSENL...KEESLN-SLIEDLTSFARE..................................................................   821
MND_2.CM.98.CM16        Y.....VFKGDYLRPHNL.K--G--R-EEPDSEKQ.-TKSDSS...KVEFGKPWSKEQIRDWLKT-RGYVW-KNLQAV......................IEYGW-............E.LKTA.....   814
MND_2.GA.x.M14          Y.....VLKGDYLRPRNL.R---E-R-GELNLEEQ.NIKSESS...KKEFGRPWRPEQIRNWLKE---YIW-KNLQAV......................IEYGW-............E.LKAA.....   814
MND_2.x.x.5440          Y.....VFKGDYLRPHNL.KR--R-GGEEPDLEKQ.NIKSESS...RQESRKPWKPEQVRSWLKR--LYIW-KNLQAV......................IEYGW-............E.LKAA.....   837
GRV.ET.x.GRI_677        P.......QIH--.SSAE.R.-.DNGGGQDR-GE-S-..........SK-IRLQEESSTPSRINNWWLNFKSCSLRIRTWCYNICLTLLIFIRTAVGYLQYGLQQL-............EAATG......   811
TAN.UG.x.TAN1           P.......QIN-R.LGRE.Q.-.DNAGGEDKDSS-SR..........DKSPPSVKES-LPNRGGIQAEERAWRQHLTNWCLTISSWLLRLYQILRRSLTTL--LLR-............ECQYI......   823
VER.DE.x.AGM3           -PQIHIHPW.KGQPDNAE.-.-GEGGDKRKNSSEPWQKESGTAEWKSNWCKRLTNWCSISSIWLYNSCLTL-VH.................................................L-SA...FQ   819
VER.KE.x.9063           -PQIHIHPW.KGRPGNAG.E.-EEGG-RDNDSSD-WQKGSGTRQKRGSWSKRLTNWLSTFS-WLYN-FLTL-IH.................................................L-GA...WQ   819
VER.KE.x.AGM155         -PQIHIHPW.KGQPDNAG.E.-EEGGRT-KSKSTH*QKEFGGRDKRTSWCRQLTNWF-TLSIWLHNSCLTL-LK.................................................L-SA...WQ   813
VER.KE.x.TYO1           -PQIHIHQVGKGRPDNAD.E.-GEGGDNSRIK-E-W*KDSKSRCMQ-TAWLT--N......-WLYNSCLTL-IQ.................................................L-KA...FQ   807
TAL.CM.00.266           F.....KTPIPSNRDPEA.PGG-AT-GGKQ-SYKWNRSQAGFLSIIWED-TGLIRYIYQT-ANLAWL-WDLAKAAKAKSQQVAQATYAFLRLHLTAAAATLQYGFSE.......................   822
TAL.CM.01.8023          F.....KTPIPSNREPGV.PEEEGT-GGKPDRSRWRTSPQGFFAIIWED-NSL-SYSYQT-RNLVWLSWDL-K-LKATTQQAAQATYAYLRIKFTAAVATIQYGFSE.......................   840
MON.CM.99.L1            -.....QIPIPS-AGQPT.PDGTGAGGGDGSNSR-PAYLKGFFTIIWED-RNLVLWTYQI-KDSVLV-YRI-Q-VSQRLPPLLHIRLLQLWESLRRLLAY-QYGIQE.......................   819
MON.NG.x.NG1            P.....QIPT-VPREPGA.PEETG-GGGEQ-RNRXTVYARGFLTIIWTDCRDLIVWTFQL-RDSGL--YRS-Q-VRSHLIPLLRDLCRQLREASSRLLAYLQYGLQE.......................   833
MUS_1.CM.01.1085        P.....QILIRPQTGPEA.PEGTEGDGGSGSRDR-VRYLSGFFSL-WDD-RNLV-WSYQT-I-SA--LWWL-Q-AVHLIQKGSIRLWFTLRALIEPLREALQRALAC.......................   840
LST.CD.88.447           APSAYVEQDCKWEKRENQ.E-----G-IEKDKE-I.YINLEQYKKESSTPPWNVDWSEPLQD-L-V-LLKWIKAFGILLLSLVWQFLSWLGHLVILFFQHGQHL.RK...........ACS-RV....V   859
LST.CD.88.485           APSAYVEQDYKWEKGENP.E-----G-IEKDKEDI.YINLEQCKKESSIPPWNVDWSEPLQD-L-V-LLKWIKAFGILLLSLVWQFLSWLGHLVILFFQHGQHL.WK...........ASSGWM....V   858
LST.CD.88.524           APSAYVEQDFKWENGESQ.E-----G-TEKDRE-I.YINLEQCKKESFRPPWNVDWNEPLQD-L-V-LLKW-KAWGILLLTLVWQFLTWLGHIVILSYHHGQHL.W-...........TLS-F-....V   861
LST.KE.x.lho7           -PSVYIEQDYKWEKEENQ.E---R---K-ADTETI.YINLEQCKKESSRPLWNVDWNEPLQD-LLV-LLKW-KEGGILLLSLVWQSLSWLWHLLILFFQNGQRL.W-...........TSS-WM....V   856
SUN.GA.98.L14           -PSAYVEQDWLQE...TC.PK-TDK--EEETEKERIYINLEQSKKESLPPPWTVDWDEPLRD-LLV-LLKW-KAWGIVLAQNIYHLLSFLWHLLTTSFHHGQRL.W-...........TLRGW-....G   860
SYK.KE.x.KE51           QK-QPPTLYHKAPAGPEE.-IETDTDNYKPSVSTFSREFLRQSLAEAQK-WN-VSESC-S-IRGLVIAWGFISYGVTELKEAAISLGREVAHQMVAIWQAL-AYARR.......................   819
SYK.KE.x.SYK173         QN-TQ........GRK-P.-K-AD---GSGDREGLNVSTFSRESLRQSLEAGQQLWRTVCSSFRSLIRQLTITW.........................................................   773
COL.CM.x.CGU1           QIGRTNVSSQIRGQDREW.E---NA--IHREQE--S--FWRLME-WWS..............................SLLRHCSNGIL.TV...........LPQHPP.TTSD   781
```

```
                                                                                   *
SAB.SN.x.SAB1C         HLEYGW.QELKAAAASLTVAVVQAATSASDSVRHACR....SIVRAVIAHPRRMRQELERWFN.        869
H1B.FR.83.HXB2         N-LQY-S----NS-V--LN-TAI-VAEGT-R-IEVVQ....GAC--IRHI---I--G---ILL.        856
H1B.US.90.WEAU160      N-LQY-S----NS-V--LN-TAI-VAEGT-R-IEIVQ....RTC--ILHI---I--G---ALL.        859
H1N.CM.95.YBF30        GILAY-GK--RDS-I--LNTTAIVVAEGT-RIIELAQ....R-G-GILHI---I--G---ALI.        852
H1O.BE.87.ANT70        AVTQY-L---QNS-T--LDTLAV-VANWT-GIIAGIQ....R-GTGIRNI---I--G---SLL.        856
H1O.CM.91.MVP5180      AVMQY-L----NS-TN-LDTIAVSVANWT-GIILGLQ....R-GQGFLHI---I--GA--ILV.        869
CPZ.CD.90.ANT          N-STL-.TIIRTEIIKNIDRLAIWVGEKT--ILL-LQ....T---IIREV---I--G--IAL-.        859
CPZ.CM.01.SIVcpzCAM13  GIIQY-GR--RTS-T--LD-TAI-VGEGT-RIIR-VQ....IVF-IIGNI---I--G---TLL.        864
CPZ.CM.05.SIVcpzEK505  GIIAY-GR---IS-IN-LDTTAV-VAEGT-RIIELVQ....R-G-GILHI---I--G---ALL.        835
CPZ.CM.05.SIVcpzLB7    GILQY-GK---TS-T--LDTTAI-VAEGT-RIIEIVQ....R-G-GILHI---I--GF--ALL.        845
CPZ.CM.05.SIVcpzMB66   GILQY-G----TS-I--LDTTAI-VAEGT-RIIEIAQ....RFG-GILNI---I--G---ALL.        877
CPZ.CM.05.SIVcpzMT145  GTLQY-TT-I-NS-I--LDTTAI-VGEGT-RILE-IT....RLG-GILHI---I--G---ALL.        852
CPZ.CM.98.CAM3         GVITY-GR--QTS-I--LD-TAI-VAEGT-RILEVAQ....I-G-GILHI---I--G---SLL.        859
CPZ.CM.98.CAM5         GVLTY-GR-IQIS-T--LDTTAI-VANGT-RIIEFVQ....IVG-GILHI---I--G---SLL.        865
CPZ.GA.88.GAB1         GIIQY-GK---IS-I--LD-TAI-VAEGT-RIIE-FQ....VTL-IIRNI---I--G---ALL.        854
CPZ.GA.88.GAB2         GIGQY-L---RTS-I--LDTTAIIVAERT-TIIEVAT....R-G-GILHI---I--G---ALLE        856
CPZ.TZ.01.TAN1         IEWSSNTYASLRVLLIQAIDRLANF-GWWTDLIIEGVV...Y-A-GIRNI---I--G--LAL-.        871
CPZ.US.85.CPZUS        GVIAY-AR---VS-T--LDTTAI-VAEGT-RIIELT-....RLFLGI-HI---I--G---SLL.        836
GSN.CM.99.CN166        IAYISYGIQELQ---TGILDSLAIF-WNWTEAVLHACR...RVW-EFL-I---I--GA-IL--.        881
GSN.CM.99.CN71         AAYISYGIQELQ--V-GILDSLAVF-WNWTEPLVQTVG...RVW-EFL-I---I--GA-ILL-.        886
H2A.DE.x.BEN           AQLQYGCEWIQE-FQAFARTTRETLAG-WGWLWE-A-....R-G-GIL-V---I--GA-LALL.        860
H2A.GW.x.ALI           AYLQYGCEWIQE-FRAFARIARETL-NTWRDLWG-VQ....WVG-RIL-V---I--GA-IALL.        862
H2A.SN.x.ST            AYLQYGGEWIQE-FRAFAR-TGETL-N-WRGFWGTLG....Q-G-GIL-V---I--GA-IALL.        859
H2B.CI.x.EHO           AYLRYGISWFQE-IQAAAR-AGETLA--ARTSWGVL-....RAAGEI--I---I--GA-LALL.        852
H2B.GH.86.D205         AYLQYGISWFQE-LRAAAR-TGETLA--GETLWE-L-....RAA--I--I---I--G--LTLL.        859
H2G.CI.x.ABT96         AYFSYGFRW-QE-CTAA-R-AQETL--TWRALWKTLG....RVG-GIL-I---I--G--LTLL.        884
H2U.FR.96.12034        AYLQYGW-YF-E-FQAFGK-ARETLSRTGRELWETLG....RVG-WLR-I---I--GF-LALL.        867
MAC.US.x.239           TYLQYGWSYFHE-VQAVWRSATETLAG-WGDLWETL-....RGG-WIL-I---I--G--LTLL.        879
MAC.US.x.251_1A11      TYLQYGWNYFHE-VQVDWRSATETLAG-WGDL*E-L-....RGG-WIL-I---I--G--LTLL.        878
MAC.US.x.251_BK28      TYLQYGWSYFHE-VQAGWRSATETLAG-WRDLWETL-....RGG-WIL-I---I--G--LTLL.        881
MAC.US.x.EMBL_3        TYLQYGWSYFHE-VQAGWRSATETLAG-WGDLWETL-....RGG-WIL-I---I--G--LTLL.        881
SMM.SL.92.SL92B        GYIRYGWNYFIE-CQEAWRSAQE-IVG-WGLIWETLG....RVG-GIA-I---I--G--LML-.        886
SMM.US.x.H9            TYLQYGWRY-QE--XXWWKFXRETLA--WRDLWETLG....RVG-GIL-IXX-I--G--LTLL.        886
SMM.US.x.PGM53         .....GCRWFQE-VQAAWR-RGETLA--WRDSWETLG....RVG-W-L-I---I--G--LALL.        885
SMM.US.x.SIVsmH635FC   TYLQYGWSYFQE--QAWWKVARETLA--WRDIWETLG....RVG-GIL-I---V--G--LTLL.        887
MNE.US.x.MNE027        TYLQYGWSYFQE-VQVAWRSATETLAG-WGDLWETLG....RVG-WIL-I---I-----LTLL.        882
STM.US.x.STM           AYLQYGCIWIQE--QAAWR-AGETLA--GRDLWETLG....RVG-RIG-I---I--G--LTLL.        882
DEB.CM.99.CM40         GYLQYGWK-F-DTITWAGGELW-WLQGIAEMALQGLV....RAGG-ILRV-A-I--GA-LIL-.        799
DEB.CM.99.CM5          GYCQYGWK-F-DTIIWAGQEIWEWLQGIAQVAIQGLV....WGG-NLL-I-A-I--GA-LLL-.        783
RCM.GA.x.GAB1          ........-F-EFSSWFAEMAL-N-YYTWRGLCAVA-....DFAGWPAMVC--I--G---LC-.        850
RCM.NG.x.NG411         .FKGFSSWLAEM-RTNAYYTWRGLCAVCR-P............AGWPATLC--I--GF--FL-.        857
DRL.x.x.FAO            ...........TLLQEVAQRIWREGRQLGL-SAR-L-....ALAQE-A-I---I--GA-VL--.        895
MND_1.GA.x.MNDGB1      ................................................................        821
MND_2.CM.98.CM16       .....GGKIF-VLQGYAQGLWSRGHQWGLSTAACFR.....A-A-GI-NI---I--GA-VLL-.        867
MND_2.GA.x.M14         .....GARICQVLQSFAQRLWSGGHQLGLSVI-G-A.....A-G--IGNI---I--GA-VLL-.        867
MND_2.x.x.5440         .....GAALYQVLQGFAQRLWSRGYQLGLSG--G-A.....AFG-GIWNI---I--GA-ALL-.        890
GRV.ET.x.GRI_677       .-AQALARAARE-WGR-.............GAIVR......-AY----NS---V--G--KVLG.        854
TAN.UG.x.TAN1          .QYGWQQFKEG--RSFEAL-SAAQSA-RTLWNACR......-AY--ILE--------------.        879
VER.DE.x.AGM3          YIQ--L.G------QEAV--LARL-QN-GYQIWL---....-AY--I-NS---V--G--GIL-.        877
VER.KE.x.9063          YIQ--L.G---T--QEARL--ARF-QN-GHQIWL---....-TY-HI-SS---V--G--EIL-.        877
VER.KE.x.AGM155        Y-Q--L.G------QEAVNHLASF-RN-AHQIWL---....-AY--I-NS---V--G--EVL-.        871
VER.KE.x.TYO1          Y-Q--L.A---TG-QEILQTLAGV-QN-CHQIWL---....-AY-NIVNS---V--G--EIL-.        865
TAL.CM.00.266          .........IQ--WRDGIT-LARFTWWWTNG-LA-A-....WAGE-IA-I---T--G--YC-T.        872
TAL.CM.01.8023         .........IQ--WGDCVT-LARFTWRWTSG-LA-A-....WA-Q-IA-I---T--G--YCLA.        890
MON.CM.99.L1           .........-Q--VT--LD-LARFTIVWT-ALL--GG....RLW--IV-I---I--GA-IFL-.        869
MON.NG.x.NG1           .........XQX-CTGAID-LARFTVIWT-A-LRLGG....RLW-GLV-V---I--GA-ILL-.        883
MUS_1.CM.01.1085       ..LQHGI---QT-STAVLDQLATFTWNWTET-LQ-G-....RVW-EFL-I---V--GA-ILL-.        897
LST.CD.88.447          ENAQKIASWSWNKIRRNRRDISA-...HARNL-LGQK....KRW-FRFRGRSGFPS-TTETAL.        915
LST.CD.88.485          ENAQKIARWSWNKIRRNRRDISA-...HARDI-LGQK....KRW-FRFRGRSGFPS-TTETAL.        914
LST.CD.88.524          ESAQKTASWIRNKSRRNRGELSKT...NGTNI-LGQNR...RRW-FRFR..SGLPS-TTETTL.        916
LST.KE.x.lho7          ENAQKIQSW-REKCRRNRGQLSST...DRKNIQLGKK....KRW-LRFGGRSGISS-ATETAL.        912
SUN.GA.98.L14          SHLIQATSRIRN-CRTSRER-SSSQKAR-RTFSLGRK....WRPKWNRTRGS-IPS-TTETTL.        919
SYK.KE.x.KE51          .........................................VAEN-A-L---L--G--IYL-.        840
SYK.KE.x.SYK173        GFISYGFN---I-----GREIRDWVAAIWQAIYA-T-....RV-E--A-L---L--G--IYL-.        832
COL.CM.x.CGU1          KRRSA-.-T-RYL...-VPRGPSLQILETLQRWLRSAA...RGW-RAPEYL-GWIYDRPQGPA.        837
```

```
SAB.SN.x.SAB1C        MGGKSSKQQQRHSLWLWSKLRQAPVIQYDMLAD.PLLGQSSHIQEECAKSLRDGLI..............RQGDSSRTE.EGVKMKHQGR......QPSWYDEDEEE..VGF.PVRPCLPLRAMTYKLA   104
H1B.FR.83.HXB2        ----W--SSVIGWPTVRERM-R...........A...EPAADRVGA-SRDLEKHGAIT.SSNTAATNAA.........................CAWLEAQE---.---.--T-QV---P----A-    84
H1B.US.90.WEAU160     ---IW--RSGSGWPAIRERMKR...........A...EPAAEGVGAVSRDLAKHGAIT.SSNE..NNAD.........................CVWLKAQ.-D-E.---.----QV---P----A-    82
H1N.CM.95.YBF30       --KIW--SSLVGWPEIRERM-R.......QTQEPA...VEPAVGAGA-SQDLANRGAIT.IRNTRDNNES.........................IAWLEAQE---E.---.----QV---PI---Q-    90
H1O.BE.87.ANT70       --NALR-GKFEGWAAVRERM-R.......TRTFPE..SEPCAPGVGQISRELAARGGIP.SSHTPQNNAA.........................LAFLESHQ---.---.--A-QV---P----G-    90
H1O.CM.91.MVP5180     --NAW--SKFAGWSEVRDRM-R.......SSS-PQ...-PCAPGVGAVSRELATRGGIS.SSHTPQNNAA.........................LAFLDSHKD-D..---.----QV---P--F-A-    89
CPZ.CD.90.ANT         --SAW--IKWVGA...RQAI-K...........I..HETNPDDIGPCGNELASRGALT.SSTI.GTEKD.........................VITYSEDHT--.GT--.----QV-M-P--E---    82
CPZ.CM.05.SIVcpzEK505 --KVW--SSLVGWPQVIDRM-R..............Q-DPAEGVGAVSQDLANRGAIT.IRNTKENNQA.........................LAWLEEQK--A..---.--C-QV---P----E-    84
CPZ.CM.05.SIVcpzLB7   --S-W--SSLVGWPEVRRRI--..............ETVAAEGVGP-SQDLARHGAIT.SSNTPQTNET.........................IAWLEEQK---..IW-.--K-QV---P--F-E-    84
CPZ.CM.05.SIVcpzMB66  ----W--SSLVGWPEVRDR--R..............T-TAAEGVGPVSRDLARHGAIT.SRNTSQTNET.........................LAWLEEVQN--..---.----QV---P--F-G-    84
CPZ.CM.05.SIVcpzMT145 --A-W--SSLVGWPEIRERIKR..............APTAADGVGEVSQDIASRGAVT.IRNTPQTNET.........................LAWLEEMQD--..---.--K-QV---P--F-G-    84
CPZ.CM.98.CAM3        --NRW--SSLVGWPDVRER-SE...........L...NLLAPGVGPVSQDLEAHGAVT.TRHTPQNNQT.........................LAWLEEMQ-HNEE---.----QV---P--F-A-    86
CPZ.CM.98.CAM5        --N-W--SSLVGWAAVRQRI--...........T...PPAASGVGPVSQDLEARGAIT.TRNH.TKYQT.........................LAWLEEMENH-EE---.----QV---P----Q-    85
CPZ.GA.88.GAB1        --T-W--SSLVGWPEVRRRI-E...........A...PTAAEGVGEVSKDLERHGAIT.SRNTPETNQT.........................LAWLEEM-N-..E---.----QV-T-P----A-    84
CPZ.GA.88.GAB2        --N-W--SSIVGWPQVRER--R.........T.....-EAAEGVGEVSQDLARRGAIT.SRHTPQTNQT.........................LAWLEEMEQ-D..---.----QV---P----A-    84
CPZ.TZ.01.TAN1        --NIFGR..WPGARKAIED-.............H...NT-SEPVGQ-SQDLQNKGGLT.TNTL.GTSAD.........................VLEYSADHT--E.---.----AV-M-P--E---    80
CPZ.US.85.CPZUS       --N-W--SSIVGWPEVRNR-...........R.QTQTTAAAEGVGPVSQDLAEHGAIT.TRNTPQNNQT.........................LAWLDEMTNH-SE---.----QV---P----Q-    88
GSN.CM.99.CN166       --S-N-----QE-STALLSS-GTGQRP-FT-V-.EYGENFWLSPDASD-GR-YY-TE..EPKPKRGSLEEYEPSC..........................----RV---DP---VM    88
GSN.CM.99.CN71        --S-N----SQE-STALLSSHGTGQSPSFR-L-.EYGEN-WLSPDASDRGR-YY-TE..ESPSRQNCID.YEPSC..........................----QV---DP----M    87
H2A.DE.x.BEN          --ASG--KLSK--RG-RER-LR-RGDG-GKQR-.ASG-EY-QF---SGREQNSPSCE..GQQYQQGEYMNSPWRNPAT.-RQKDLYRQ-........NMDDV-S-DDD.LI-V.--T-RV-R-E------   116
H2A.GW.x.ALI          --ASG--KRSGPLQG-RER-L-T-GETCGGQCS.GSG-GY-QS-GGSGRGQKLPSCE..GQRYQQGDFMNTPWRTPAT.-R-KELY-Q-........NMDDV-L-DDDSL--V.S-T-RVQ--T------   117
H2A.SN.x.ST           --ASG--KRSEP-RG-RER-L-T-GEASGGHW-.K-G-EYLQS--GSGRGQKSPSCE..GRRYQQGDFMNTPWRAPAE.GE.KGSY-Q-........NMDDV-S-DDD.L--V.--T-RV---E---R--   115
H2B.CI.x.EHO          --SAG--K-SKQQPG-RER-LR-RRGPRGESSG.ERQER-LQYPGGSD-G-NSPSCD.................GQK.TLGAEGGGK-..........DSDED--DNE.--V.R---GV---P--F---    98
H2B.GH.86.D205        --SAG--KRSERQQG-RE--LRV-ERP-GR-SG.ERRE---RSPG-SD-D-NSPSCE.................GQN.ARGAEGGGQ-..........DADES--DDE.--AICKT-IV---P------    99
H2G.CI.x.ABT96        --SAG--KRAPQQ-G-RKR-L--RGEP-GK-WE.G-EEGY-QYR--SG-G-SSLSCE..GQQYTQGQFMNTPWRNPAT.-RAKLAYRX-.........NNDDVDSDDNDL--V.--Y-KV---V-S----   116
H2U.FR.96.12034       ---SI--KHSKR-GK-RER-LA-RGEN-GR-WG.G-EDGFELY-GGSG-D-SSRFCEAQGQGYSEGEFMNTPWRTPAT.GKDKLQY-Q-........NMDDV--EDND.L--V.A---RVA--T------   118
MAC.US.x.239          ---AI-MRRS-P-GD-RQR-LR-RGET-GR-LG.EVEDGY-QSPGGLD-G-SSLSCE..GQKYNQGQYMNTPWRNPAE.-R-KLAYRK-........NMDDI--*DDD.L--V.S---KV---T-S----   116
MAC.US.x.251_1A11     ---TI-MRRS-STGD-RQR-LR-RGET-ER-LG.EVEDG--QSLG-LD-G-SSLSCE..GQKYNQEQYMNTPWRNPAE.-R-KLAYRK-........NMDDI--EDDD.L--V.S---KV---T-S----   116
MAC.US.x.251_BK28     ---AI-MRRSKPAGD-RQ--LR-RGET-GR-LG.EVEDG--QSLGGLG-G-SSRSCE..GQKYNQGQYMNTPWRNPAE.-K-KLAYRK-........NMDDI--EDDD.L--V.S---KV----------   116
MAC.US.x.EMBL_3       ---AI-MRRSKPAGD-RQ--LR-RGET-GR-LG.EVEDG--QSLGGLG-D-SSRSCK..GQKYNQGQYMNTPWRNPAE.-K-KLAYRK-........NMDDI--EDDD.L--V.S---KV----------   116
SMM.SL.92.SL92B       ---IT-SKPA-KHGD-RK--LS-RGDS-GR-W-.G-EEG-QQS-GGSDRA-NSC-TR.DGYKYNVGEFMNTPWRNPAT.-RQKQQYRQ-........CMDDI---DD-.LA-V.V-W-KV-----S----   117
SMM.US.x.H9           ---VT--K-RXAGGN-XER-L--RGET-GR-WE.G-E-EY-QSXDASG-G-SSLSCE..PQKYCEGQFMNTPWRNPAT.-GAKLGYRQ-........NMDDV-NED...L--C.--S-RV-V-I------   114
SMM.US.x.PGM53        --AAG--K-S-QRGG-GE--L--RGETHGK-WE.G-EDGY-QSRG-LGRDWNLHSFE..GQGYSEGQFMNTPWRNPA-.-R-KL-YRQ-........NMDDV-D-DD-.LI-V.S-H-KV-----S----   116
SMM.US.x.SIVsmH635FC  ---AI--K-H-RGGK-RER-LR-RGET-GR-WE.G-EEGY-QSLGASG-G-SSLSCE..PQKYSEGQYMNTPWRNPTA.-KAKLGYRQ-........NMDDV-DEDDD.L--V.S-H-KV----------   116
SMM.US.x.SIVsmH635F_L3 ---AI--K-H-RGGK-RER-LR-RGET-GR-WE.G-EEGY-QSLGASG-G-SSLSCE..PQKYSEGQYMNTPWRNPTA.-KAKLGYRQ-........NMDDV-DEDDD.L--V.S-H-KV----------   116
SMM.US.x.SIVsmH635SB10 ---AI--K-H-RGGK-RER-LR-RGET-GR-WE.G-EEGY-QSLGASG-G-SSLSCE..PQKYSEGQYMNTPWRNPTA.-KAKLGYRQ-........NMDDV-DEDDD.L--V.S-H-KV----------   116
MNE.US.x.MNE027       ---AT--KRSKSPGD-RQR-LR-RGET-GR-WE.G-EDGY-QSLGGSD---SLLSCE..GQKYNQGQFMNTPWKNPAG.-R-KLAYRK-........NMDDI--EDDG.L--V.----RV----------   116
STM.US.x.STM          --ASG--K-RKQHGE-RER-LR-RGET-GK-LE.G-GEG-GPS-GASD-G-NSHSCE..PQRYNEGQFMNTPWKNPAA.-SAKLEYRQ-........NMDDV--EDDN.L--V.A-H-RV---E------   116
DEB.CM.99.CM40        ----N--IPS-GQVGSFGSGSRGLLRWRYRDLS.EQEE-F-ECLL-SDREQSSSSTE....FWGSPRREIKCT.......................NKQQQ-L-QT-A.---.----QVSI-DP----M    99
DEB.CM.99.CM5         ----N--TPSFGQGKKSGSGSKGLLRWRYKDLS.GEAETF-PFLQ-SD-EQSCYSTE..EPLY.ADPHREIKCT.......................-RAQECY---HG.I--.A---RV-I-EP----M   101
RCM.GA.x.GAB1         -------NSAAGL-RWRF-MLTT-GEG-VRWHE.T--DGQPWCA-GSGRAS--FV-R..GGITAETQASIDDIDWYED-DDTL.........................---.--K-QV---P-S----    99
RCM.NG.x.NG411        SAAGLRRWRGLMLTTPGEDYARFAET.........-QDGQPRCA-GSGRAS--F-TR..GGFTIETQQSVDAIDWYED-DDTL.........................---.----QV-I-P-S----    91
DRL.x.x.FAO           ---QN--KREEAYARFYKR--RGYGAKGGN-DYHQ-EPSE.PLLRKWQPFPGGLDKA.QRSSSTSSLDTTH.MQAAAAEKDVTLPT..........EEEQPS-E---E.---.--C-QR---EP---DL   114
MND_1.GA.x.MNDGB1     --SSQ--KRSEAWVRYS-A---LVGGPVTPDGYKQIESSQ.GAEKQSLLRG-AYGTYSEGL..........DKVQNDPLTKDE-L..........DLTQQ-PE---E.---.--CRQVS--VPS--DL   105
MND_2.CM.98.CM16      ---------EEKY-KYYKAM-RGYGAEGTNGDYQQ-HASE..PLLGALSTSQEEFDREQKSSST.............................DE---..T--.--Y-QC-V-EP---DL    86
MND_2.GA.x.M14        --------REQGYAKYYKA--RGYGAEGTN-DYQL-EE-R..LL-GPLGTSLGASDKELKSCST..............................KVD--QK--..R--.--Y-TQ-V-EA---NL    90
MND_2.x.x.5440        --------REQ-YIKYYKA--RGYGAEGTN-DYQV-EE-Q..PLAGASGTSLAASDKVLRPYST........ER-LRHQQDVTLPS..........EKEQPSDE---.---.--Y-TQ-VPEA---DL   106
GRV.ET.x.GRI_677      --SSN--R--QGL-K--RG--GK-GADWVL-S-PLIGQS-TVQE-CGKALKKSWGKGKMTPDGRRLQ.........................EGDTFDEW-D----.---.--Q-RV---Q------   100
TAN.FR.x.B87_14       ---SN--KE-EGL-KM-RLM-K---VR------PLI...GTSSSIQEECDKNWSGGLTKGKGKMTPEGRKLTND-TF....................DE-D--S--..---.----QV---Q------   102
TAN.UG.x.TAN1         ---SN--RE-QGL-R--RA--K---VR-G----...PLIGQSSNIQEECDKNWNGGSTRRGKST.......................PE--KLAADDTWDDWEPE---.---.----RV---Q------   101
VER.DE.x.AGM3         --LGN--P-HKKQ-S--HA-HKTRATR-GL---.--I----TL----D-A-KES--...............-KRNGKM-.P--R-LQEGD......KWDEWSDE-D-..---.----RV---Q------   103
VER.KE.x.9063         --LGN--P-HKKHVS--HA-HKTQ-TK-GL---.--I-T--AV---YD-G--KS--...............-KRNGNM-.PD-RRLQEGD......EWDEWSD--D-..---.--K-RV---Q------   103
VER.KE.x.AGM155       --LG---P-HKKQ-TI-RA-HATRHTR-GL---.--I----TL----D-G--KS--...............-KRNGNM-.P--RRLQDGD......-WDEWSDE-D-..---.----RV---QI-----   103
VER.KE.x.TYO1         --SQN--PAHKKYSK--QA-HKTH-TR-GL---.--I-T--TV----D-A--KS--...............-KQNGNM-.E-ERRLQEGD......TWEEWSD----..---.----RV---Q------   103
TAL.CM.00.266         --SVK--LLGETASQRLQGSHGGGRMV-WR-R-.GQVKR-QLSPGARG-D-NIA-LDEE.WQATGADY.MGKALTQGEVWY................EKEQE-NDM-.L---.----KV---TC----G   108
TAL.CM.01.8023        --SVK--LLGETASQRLQGSPG-GRLV-WR-R-.GQCR-LQRSRG-RG-G-NIA-LDAG.WQASDANW.MGKALSEGEVWY................EKEQE-NDM-.L---.----QV---PCN---G   108
MON.CM.99.L1          --S-N----SPL-STPLLGSQSSGRMR-F--E-.DYGE--WLSPDASDRERKYS-TEGRNGKQRRQP............................LDD-D-DDG..--C.----RV---DP-W--M    94
MON.NG.x.NG1          --S-X-X-PAQEP-MPSHGSQSSGQTRSFAWE-.DYGED-WLSPDASDRGR-FS-TEGRNHRSNRRPTTVVDDA-EDQDLPGC..........................----MV---DP-W-IM    98
MUS_1.CM.01.1085      --S-----PAQQ--ISSPPSPGTGRK--FK-VE.EYGEN-WLSPDASGRGR-YS-T.......EGASKRPVIEEH.ERTC..........................----RV-I-DP----M    87
LST.CD.88.447         --NIFG-PAADG...W-KT--........................................RLRAGAGTRSEGTE-TYQQL-QETSDWDKR.RDGDSDS-E-....---.----QR--CSP---TL    77
LST.CD.88.485         --NIFG-PAADG...W-KT--........................................RLRAGAGTRSEGTE-TYQQL-RETSDWDKR.KDGDSDS-G-....---.----QR--CSP---TL    77
LST.CD.88.524         --NIFG-PSADS...W-RV--........................................KLRAGSGTRAEGTE-SYQKL-EQTLGWDKTEEDGDSDS..-....---.----QR--CSS---TL    76
LST.KE.x.lho7         --NAFGRPAADG...W-KT--........................................RFRAGSGRNAEGTEDSYRALIEKTSNWAKR.RDGDSDS-E-....---.--K-QR--CSP---TL    77
SUN.GA.x.SOL_36       --NAFG-PSEVGWVRTLF---AGSG.......TRAEP.................AGRAYHRIRGETEPLRSNPD.GG................LEELE-T-....---.----QR--TQP---NL    79
SUN.GA.98.L14         --NAFGRPSEVGWVRTLFR--AGSGTRAEPAGREYHRLRR...............................QEVEPLVSAENGGPNGIEQ--..-E-.----QR--CKP---QL    80
BLU.KE.x.KE31         --ST---N--CR-ESPYGTRWWRRAK--TP-P-EL-KPSRPSHGGFDKAWRSTLTEPVDPHGPDRDWEHSGG-KW-PGD.VVHDEGDTGL...................---.--C-QI---PL-----   108
SYK.KE.x.KE44         --SVK-RM--SAWEDRLETGWWKRRGK-TPFP-AL-RASLPSRGGFDKAWRSTLTEPIDPHGPDRDWGHSGG-KW-PGD.MVYDEGDTGL...................---.--C-QT---PL-----   108
SYK.KE.x.KE51         --SQ---K-PSRWDEK-RTRWWPFGKP-SPMP-.E--RM-QPYH-DFDRGW-ST-TEPILDPKR...DFDSGGKKWNAGDICHDEGDED..................L---.--L-QV---PL-----   105
SYK.KE.x.SYK173       --ST---S--LR-EGKYAIGWRLFGK--TP-P-.E-SRPLQPCRGGFD-AW-ST-TEPIDPHGPDRDWGHSGGQKF-PGDIVQDEGDTG..................L---.--C-QT---TL-----   108
COL.CM.x.CGU1         --SLL-CLNI-QQPQTRGGVLGRLFGI...YWYPGVQAFRFLRHFRGGYD--RVAGVERLNICEDGYMTVPKVLP.....NCPCPLCYRHPDVYYHN-EPE-VEDV-..A-C.Y-MSR--V-QA-I--L   118
```

```
SAB.SN.x.SAB1C         IDFGHFLKEKGGLEGIYYSERRKKILDLYALNEWGIVDGWQNYTDGPG.......TRYPKCFGWCFKLVPVDLSEEAKNSEN..HC..LLHPAQVAYEDD....AW.KETLVWKFDPLLAVDY.VAWRLH   217
H1B.FR.83.HXB2         V-LS-----------LIH-Q--QD----WIYHTQ-YFPD*----P---.......V---LT----Y-----EPDKIEEANKGENTS..----VSLHGM--....PE.R-V-E-R--SR--FHH.......   193
H1B.US.90.WEAU160      H-LS--X--------LI--QK-QD----WVYHTQ-FFPD-----P---.......----L-----------EPEKVEEAN-GENNS..----MSLHGM--....HE.--V-M----SK--FHH.......   191
H1N.CM.95.YBF30        F-LSF---D------LVW-RK-QD----WMYHTQ--LPD-H---P---.......I---VT---------LSAE-VEEAN-GDNNA..----ICQHGA--....DH.--V---R--SS--RRH.......   199
H1O.BE.87.ANT70        F-LSF----------LI--HK-AE----WVY-TQ-FFPD-----P---.......--F-LT---L------SEE-AERLGNTCERA.N-----CAHGFE-....TH.--I-M----RS-GNTH.......   200
H1O.CM.91.MVP5180      F-LSF--------D-LI--HK-AE----WIYHTQ-FFPD--C--P---.......P-F-LT---L------SAE-AERLGNTNEDA.S-----CNHGAE-....-H.G-I-K-Q--RS-GLTH.......   199
CPZ.CD.90.ANT          V-LSX--------DXLIXXLK-AA---MWMF-TQ--FPD-----EE--.......V---L-R---------ESPPD.....DERNI..-----CTHGDG-....PH.--I-R-E--AS-MRRH.......   186
CPZ.CM.05.SIVcpzEK505  F-LSF--R-------LVW-TK-QE----WVYHTQ--FPD-----P---.......V---LT---------LSPE-VEEANQGDNNV..----MCQHGME-....PD.--V---R--SS--RVH.......   193
CPZ.CM.05.SIVcpzLB7    F-L-Y----------LV--KK-QE----GVYHTQ--FPD-----P---.......I---IT--G------LTPE-VEEANKGENNL..----ICQHGME-....ED.--V-I--Y-SQ--LRH.......   193
CPZ.CM.05.SIVcpzMB66   F-LS-----------LV--KK-QE----WVYHTQ-FFPD-----P---.......I---LT---------LEPE-VERAN-GDNNI..----ICQHGQE-....EA.R-V---T--SR--LKH.......   193
CPZ.CM.05.SIVcpzMT145  F-LS-----------LIW-RK-QE----WVYHTQ-YFPD-----P---.......V---LT---------LPPE-VE-AN-GENN-..----MCQHGME-....EE.R-V---T--SK--LRH.......   193
CPZ.CM.98.CAM3         F-LS----K------LI--IXXXXXX-HGYYHTQ-FFPD-----P---.......V---LTY-G------LTPE-VEQAN-GDNNI..----ICQHGME-....-H.G-V---R--SR--LRH.......   195
CPZ.CM.98.CAM5         F-LS-----------LV--TK-QE----WVYHTQ--FPD-----P---.......V---LT---------LTPE-VE-AN-GDTNI..----ICQHGME-....EH.G-V---R--SR--LKH.......   194
CPZ.GA.88.GAB1         F-LS-----------LV--R--QE----WVYHTQ-FFPD-----T---.......--F-L----------LTEEQVEQAN-GDNN-..----ICQHGME-....ED.--V---R--SR--LRH.......   193
CPZ.GA.88.GAB2         F-LS-----------LV--K--QE----WVYHTQ-YFPD-N---P---.......V---LTY---------SPEAVEEANKGEDNL..----LCTHGFE-....EE.--V-I----SQ--LRH.......   193
CPZ.TZ.01.TAN1         --LSW--------D-LFF-PK-AA---TWMY-TQ-VFPD-----P---.......I---L-R--L-------PP-D.....DEKNI..-----CSHGTT-....PD.G---I-R--SS--RRH.......   184
CPZ.US.85.CPZUS        F-L-F----------LV--R--QE----WVYHTQ--FPD-----P---.......V---LTY--------LTEE-VEQANKGETNI..----MCQHGME-....EH.G-V-I-Q--TE--RRH.......   197
GSN.CM.99.CN166        V-LS-----------MF-C-D-HQK-EQ--YL---L-P--LSF-P---.......----TIP-F-IC-R--ATT-DSEPGDD.EYL..-T---YQGRSE-....QH.--F--FS-CSK--IKSGIQLDQL   203
GSN.CM.99.CN71         V-LS----------AMF-C-D-HQK-ES-CYY-----P--LQW-P---.......I---TMP-F-WC-R--AMT-DSEPGDD.QYL..-N---YQGQQE-....HH.R-I--FS-CSR--LKSGWQMNQL   202
H2A.DE.x.BEN           --MS--I------Q-MF--R--HR---I-LEK-E--IPD-----H---.......V---MY---LW---S-E--Q--EED-A..N-..-V----TSRH--....EH.G-----Q--SM--YN-.K-FT-Y   229
H2A.GW.x.ALI           V-MS-LI--R-----MF-----HR---I-LEK-E--IPD-----H---.......I---MF---LW------VPQ-GEDT-T..--..----V-TSRH--....TH.G-----R---K--H--.K-FI--   230
H2A.SN.x.ST            R-MS-LI--------L---D--RRV--I-LEK-E--IGD-----H---.......V----F---LW------VPQ-GDD--T..--..-V----TSRF--....PH.G-----R---T--FS-.E-FIRY   228
H2B.CI.x.EHO           V-MS-------E----F-----H----T-LE--E---S------H---.......V----F---LW----INMIA-PEDE-T..--..-V----TSAW--....PH.E-----Q--S---Y--.--FSRF   211
H2B.GH.86.D205         V-MS--I--Q-----M------HR---T-FE--E---S------H---.......I----Y---LW-----EVPAATREE-ET.--..-M----ISSW--....IH.G---I-Q--S---Y--.--FNRF   213
H2G.CI.x.ABT96         --MS--I---------------HR---I-MEK-Q--IPD-----A---.......I---XT---LW-----TV-D--QED-T..--..-V----TSPW--....PE.T-V-A-----T--Y--.R-FI--   229
H2U.FR.96.12034        V-MS--I------D--F--Q--HR---I-LE--E--IPD-----S---.......--W-M----LW--E---ANDK-QED-R..LY..-VGS--TSC-E-....H-.G-A------SS--YS-.Q-FIKC   231
MAC.US.x.239           --MS--I------------A--HR---I-LEK-E--IPD--D--S---.......I----T---LW-----NV-D--QED-E..-Y..-M----TSQW--....P-.G-V-A-----T--YT-.E-YVRY   229
MAC.US.x.251_1A11      --MS--I------------A--HR---I-LEK-E--IPD--D--S---.......I----T---LW-----NV-D--QED-E..-Y..-M----TSQW--....P-.G-VPA-----T--YT-.E-YVRY   229
MAC.US.x.251_BK28      --MS--I------------A--HR---M-LEK-E--IPD--D--S---.......I----T---LW-----NV-D--QED-R..-Y..-MQ---TSKW--....P-.G-V-A-----T--YT-.E-YARY   229
MAC.US.x.EMBL_3        --MS--I------------A--HR---M-LE...--IPD--D--S---.......I----T---LW-----NV-D--QED-R..-Y..-MQ---TSKW--....P-.X-V-T-----T--YT-.E-YVRY   226
SMM.SL.92.SL92B        --MS--I------D--------HR---I-LEK-E--IPD-----S---.......I----F---LWQ-----V-D--N-D-T..--..-V----TYQHS-....P-.GKV-A---N-H--YT-.E-FVR-   230
SMM.US.x.H9            --MS--I--------V---I--H----X-LEX-E--IPD-----A---.......I---MF---LW--X--NV-D--QED-T..--..-M----TSQW--....P-.G-V-A-----X--YX-.K-FVE-   227
SMM.US.x.PGM53         --MS--I---------------HR---I-LEK-E--TPD-----S---.......I---MF---LW-----SA-D--QED-T..-Y..-V----ISQW--....P-.G-V-A----SQ--YR-.E-FIR-   229
SMM.US.x.SIVsmH635FC   --MS--I-----------N-K-HR---M-MEK-E--IPD-----L---.......----MY---LW------V-D--QED-T..--..-V----THQW--....P-.G-V-A-----E--YS-.K-FIKY   229
SMM.US.x.SIVsmH635F_L3 --MS--I-----------N-K-HR---M-MEK-E--IPD-----L---.......----MY---LW------V-D--QED-T..--..-V----THQW--....P-.G-V-A-----E--YS-.K-FIKY   229
SMM.US.x.SIVsmH635SB10 --MS--I-----------N-K-HR---M-MEK-E--IPD-----L---.......----MY---LW------V-D--QED-T..--..-V----THQW--....P-.G-V-A-----E--YS-.K-FIKY   229
MNE.US.x.MNE027        V-MS--I---------------H----I-LEK-E---PD--D--S---.......P----T---LW-----NV-D--QEG-E..NY..------TSQW--....P-.G-V-------T--YT-.E-YIRY   229
STM.US.x.STM           --LS--I-S-------------HR---M-LEK-E---PD-----A---.......I----Q---LW------M-N--QEDDGT.-Y..-V----THQW--....P-.G-V----------HT-.E-FVR-   230
DEB.CM.99.CM40         --YS----------DVF--A--HA--E-H-Q-----IP--LQ--K---.......V---IY--FL------EIADPDYENDE..RN.I---D-HQGMME-PH.....R-R------ST--YC-.K-GHAE   213
DEB.CM.99.CM5          --YS-Y--------D-F--A--HA--EIH-HH----IP--LK--E---.......P---TY--FL------EIADPDYENDE..RN.I---D-HQGQAE-PH.....--R------SS--YC-.K-GHEA   215
RCM.GA.x.GAB1          --MS-------------W-I--QR---M-LE--H--IPD-----P---.......I---TL---LWQ-----V-D--RED-EHS....-----ETSGME-....P-.G-V-A---N-M-----.IGY---   212
RCM.NG.x.NG411         --LS-------------F----RR-----ME--H--IPD-----P---.......I---TM---LWQ-----V-D--TEN-R..--..-----ETSGM--....P-.G-V-A---K-------.AGY---   204
DRL.x.x.FAO            V--S---------K--WW-L--ET------Q----FIK---S--K---.......I----M--FL------EVD-DLV-QQCNR....--NSS-MGPI--....PE.G-R-M-Q---G--YTFQAC.IH-   227
MND_1.GA.x.MNDGB1      ---S--I------G-----R--EE------E----FEP---Q--T---.......-----T--FL---E--SRAIGDEYAA-N.-...---SS-LCPQE-....PE.G---M-SGTLI-PMTL.....Q-   214
MND_2.CM.98.CM16       V-MP------------WH-K--EE------Q----FIPT--S------.......I----T-RFL---C--AVPPDQE-N-CNK....--QSS-LGIQEE....SLXG-R------SG--YTF.Y-PIKR   200
MND_2.GA.x.M14         --MS------------WF-R--EE--N---Q----FIPD--E--S---.......I----R--FL---I--QVPPDQE-Q-CNR....--NSS-LGIQEE....SMXG-R-M-----E--YTF.Y-PIKC   204
MND_2.x.x.5440         --MS------------WW-K--EE------Q----LIPN-----K---.......I----T--FL------AV-PPLEED-CNR....--NSS-MGIQE-....PE.G-R-M----SG--YTF.Y-PIIR   219
GRV.ET.x.GRI_677       V--S----S----D--------E---N---------I-D--A-SP---.......I---RV--F---------H---R-C-R..--..-M----MGEDP-..GIDH.G-V-------K---E-.RPDMFK   215
TAN.FR.x.B87_14        V--S-------------W-P--EQ--N---------I-D--V-SP---.......I-K-----F--E----R---A-ATN-R..--..------TN-M--....P-.G-------N-----Q-NPDCFKD   216
TAN.UG.x.TAN1          V--S-------------W-PK-EQ--N---------I-D--A-SP---.......--K-R---F--E-----V-Q--QD.-R..--..------IEW-S-....P-.--------N-----Q-.NPDSFK   213
VER.DE.x.AGM3          V--S---------D-----D--NQ--N---------I-D-NAWSE---.......I---R---F---------H---ETC-R..--..-V-----REDP-..GINH.G-V-------M---Q-.DPN-KY   218
VER.KE.x.9063          V--S---------D--------NR--N---------I-D-NAWST---.......V---R---F-------EMH---ETC-R..-Y..-A----IKEDP-..GISH.G---------M---Q-.DPN-QY   218
VER.KE.x.AGM155        V--S---------D-----D--N---N---------I-D-NAWSK---.......I---R---F-------A-H---ETC-R..--..-V----LHEDP-..GINH.G-I-A-----M---Q-.DPS-EY   218
VER.KE.x.TYO1          V--S---------D-----D--N---N---------I-D-NAWSK---.......I-F-----F---------H---QTC-R..--..-V----MGEDP-..GISH.G-I-------M--IQ-.DPN-EY   218
TAL.CM.00.266          ---SF--------K--F-NK--HA--N---H-----LPD-----E---.......----L---ILW--C--EIH-DDTEDGH.....L-P---YDGQAE-....P-.G-S---V--EK--YTPGAKMAEW   221
TAL.CM.01.8023         ---SF--------K--F-NN--HA--N---H-----LPD--S--E---.......----L---ILW--C--EIH-DDTEDGH.....L-M---YDGQQE-....P-.G-A---V--EK--YTPGAKMAEY   221
MON.CM.99.L1           M-LS-Y-------GEMF-C-D-HRKIEQ--YL---LIP--LQ--E---.......V---TMP-FLWC-R--ATT-DSEEGDEDF...L-T---YQGRME-....PH.RQF--FS-CSK---KSGRQLAQL   209
MON.NG.x.NG1           M-LS---------DKLF-CXDXHQK-EQ--YVM--L-P--LQF-X---.......V---TIP-FLWC-R--AIQ-DSEDGDXEF...L-T---YQGRDE-....PH.-QF--XS-CSK-G-KSGPQLDEL   213
MUS_1.CM.01.1085       V-MS-----------MFV-TE-LQK-ETW-YV----IP--LM------.......-----IP-FLWC-RE-AMT-GSEEGDDQY...L-D---FQGR-E-....HH.-QF--FS-CSR--LKSGLQLDQM   202
LST.CD.88.447          --LS--I--------L-W-Q--QD--C--CE----LIGDFMT------.......--F-LA---LWQ-E--ACD-YTDP-DYR.Q-..---SS-LGVQE-....P-.G-R-I-H-N-M-----.--L-KQ   191
LST.CD.88.485          --LS--I--------L-W-Q--QD--C--CE----LIGDFMT------.......----LT---LWQ-E--ACD-YTDP-DAS.Q-..---SS-LGVQE-....P-.G-R-I-H-S-M----F.--L-KQ   191
LST.CD.88.524          --LS--I--------L-W-Q--ND--T--CE----LIGDFM-------.......----LA---LWQ-E--ACD-FKDP-DER.Q-..---SS-LGVLE-....P-.G-R-I-H-N-M-----.--L-KQ   190
LST.KE.x.lho7          --LS--I--------L-W-Q--QD--I--CE----LIGDFM-------.......----LT---LWQ-E--ACD-YKDP-DET.Q-..---SS-LGVLE-....P-.G-R-I-H-N-M----F.I-LKKQ   191
SUN.GA.x.SOL_36        --HSF-I--------LW--KT-EE-I-I--E-----IP---A--S---.......P----T---LW--M--TID-DRGPNHP...-QA---SS-QGVNE-....S-.G-R-I-T--SS--Y-FKAIQKHP   194
SUN.GA.98.L14          --LS--I--------LW--RT-EE------E----FIT---D--K---.......V----A---LW--A--TID-DRDPN..HPCQAL-HSSQ-GVN--.....P-.G-R-I-T---T--Y-F.R-IQK-   194
BLU.KE.x.KE31          V-LS----N----Q--N-DA--DE--H--LK--H-VI-RINYT.G---.......I-F-LL---LWE---N-IEGYLTDE-DTL....-----SGKGAEED...PH.G-N-M-N-N-H--YTPGWEMARL   222
SYK.KE.x.KE44          --LS-LI------Q-MNFDS--DE--H--LK--H-VI-RINYT.S---.......----LI---LWE-A-N-IEGYLSDE-DTL....-----AGKGASED...IH.G-N-I-N-NSH--YTPGWELARR   222
SYK.KE.x.KE51          --LS--I-R----Q-MFFCQK-HE--Q--LK--H-VI-D.IT--S---.......----LI---LWE-A-NEIEGYLVDE-DT....LMM---AGVGASE...DPH.R-N-M-N-N-H--YTPGWEMARQ   219
SYK.KE.x.SYK173        --LS--I-N----Q-MN-C-K-DE--H--LQ--H--I.DRI---S---.......----LI---LWE---NEIEGCLEYE-HT....L-----SGQGSSS.MGEPH.V-LQPPPGYTPGW.E....MARL   219
COL.CM.x.CGU1          V-ISC-----------IRTPE-DDLIEQ--YI---CLK--LE-E-EL-EDGALKED-K-LVA--LW-...........................................--YIEQLGEYAYS..YDLSL   203
```

```
SAB.SN.x.SAB1C              PEQ.VPSA.QG.....................................     226
H1B.FR.83.HXB2              ...........................VARELHPEY.FKNC.......     206
H1B.US.90.WEAU160           ...........................VARELHPEY.FKDC.......     204
H1N.CM.95.YBF30             ...........................VARELHPEF.YKNC.......     212
H1O.BE.87.ANT70             ...........................VAMITHPEL.FQKD.......     213
H1O.CM.91.MVP5180           ...........................IALQKHPEL.FPK........     211
CPZ.CD.90.ANT               ...........................IARERHPEY.FRD........     198
CPZ.CM.05.SIVcpzEK505       ...........................RARELHPEF.YQNC.......     206
CPZ.CM.05.SIVcpzLB7         ...........................IARELHSEY.YKNC.......     206
CPZ.CM.05.SIVcpzMB66        ...........................RARELHPEY.YKDC.......     206
CPZ.CM.05.SIVcpzMT145       ...........................LAREKHPEY.YRD........     205
CPZ.CM.98.CAM3              ...........................IAREQHPEY.YQN........     207
CPZ.CM.98.CAM5              ...........................LAREQHPEY.YGDHNPAPAQ.     213
CPZ.GA.88.GAB1              ...........................IAREQHPEY.YKD........     205
CPZ.GA.88.GAB2              ...........................LAREKHPEY.YRD........     205
CPZ.TZ.01.TAN1              ...........................IARERYPEY.FK.........     195
CPZ.US.85.CPZUS             ...........................RAKELHPEY.FRN........     209
GSN.CM.99.CN166             QQE.E.....................RKMRLTANRFL...........     218
GSN.CM.99.CN71              QQE.E.....................RKKRLTANRFL...........     217
H2A.DE.x.BEN                --E.FGHK.S-LPEKEW.........KAKLKARG.IPYSE........     257
H2A.GW.x.ALI                --E.FGYK.S-LPEDEW.........KARLKARG.IPFSKNRNS....     262
H2A.SN.x.ST                 --E.FGYK.S-LPEDEW.........KARLKARG.IPFS.........     255
H2B.CI.x.EHO                --E.FGYQ.S-MPEKEW.........KAKLRARGIPTE..........     238
H2B.GH.86.D205              --E.FGYQ.S-LPEEEW.........KARLKARG.IPTD.........     240
H2G.CI.x.ABT96              --E.FRWK.S-LPEAVW.........KEKLKQRGLPIE..........     256
H2U.FR.96.12034             --E.FG-K.S-LSEEEV.........KRRLTARGLPVKNC........     260
MAC.US.x.239                --E.FG-K.S-LSEEEV.........RRRLTARGLLN..MADKKETR.     263
MAC.US.x.251_1A11           --E.FG-K.S-LSEEEV.........RRRLTARGLLN..MADKKETR.     263
MAC.US.x.251_BK28           --E.XG-K.S-LSEEEV.........RRRLTARGLLN..MADKRETR.     263
MAC.US.x.EMBL_3             --E.FG-K.S-LSEEEV.........RRRLTARGLLN..MADKRETR.     260
SMM.SL.92.SL92B             --E.FGWK.S-LTEEEV.........ERRLANKPKPQKKMADKKETS.     266
SMM.US.x.H9                 --E.FG-Q.S-LSKEEV.........QRRLTARGLLK..MADKKKTS.     261
SMM.US.x.PGM53              --E.FG-K.S-LSEEEV.........KRRLTARGLLK..MADKKETS.     263
SMM.US.x.SIVsmH635FC        --E.FG-K.S-LSEEEV.........KRRLTARGIYK..MADKRETS.     263
SMM.US.x.SIVsmH635F_L3      --E.FG-K.S-LSEEEV.........KRRLTARGIYK..MADKRETS.     263
SMM.US.x.SIVsmH635SB10      --E.FG-K.S-LSEEEV.........KRRLTARGIYK..MADKRETS.     263
MNE.US.x.MNE027             --E.FG-K.S-LSEEEV.........RRRLTARGLLK..MADKKETS.     263
STM.US.x.STM                --E.FG-K.S-LPKEEV.........ERRLTARGLLK..MADKKETS.     264
DEB.CM.99.CM40              QRE.HTRRCMFP..............KRK...................     227
DEB.CM.99.CM5               HKPEHNRRCMFP..............KRK...................     230
RCM.GA.x.GAB1               --F.FGER..................KNKTQ.................     224
RCM.NG.x.NG411              --F.FGER..................KNKTK.................     216
DRL.x.x.FAO                 --E.FGHVKSIKSVEEPG......CCKRKWWQFKPTTEGCHGDNLQKC     268
MND_1.GA.x.MNDGB1           ................................................     214
MND_2.CM.98.CM16            -GE.FKHV.-SISHIAYAKEHKPECCKRKWWQF...............     231
MND_2.GA.x.M14              -KE.YRHV.TSLSYTAYQE..KSDCCKRKWWQF...............     233
MND_2.x.x.5440              --E.YKCV.TSLSYEAYKKEEKPDCCKRKWWQF...............     250
GRV.ET.x.GRI_677            DMH.EHA...................KR....................     223
TAN.FR.x.B87_14             MHG.LV....................KRK...................     224
TAN.UG.x.TAN1               DMH..GLV..................KRK...................     222
VER.DE.x.AGM3               LTD.MHDL.G................KRK...................     229
VER.KE.x.9063               F-D.MHAI.V................KRK...................     229
VER.KE.x.AGM155             FTD.LY-T.V-TGN..................................     230
VER.KE.x.TYO1               FTD.MHGL.V................KRK...................     229
TAL.CM.00.266               DRLE.....................REKRMLLAPPQTASS........     240
TAL.CM.01.8023              DRLE.....................RKKKELLAPPQTASS........     240
MON.CM.99.L1                QQE..E...................RKKRLAANRIL............     224
MON.NG.x.NG1                QQE......................ERKRRLTANRIL...........     228
MUS_1.CM.01.1085            QQE......................ERKKRLATNHIL...........     217
LST.CD.88.447               -KS.IQASAI-L..............QRK...................     205
LST.CD.88.485               -KS.IQASSLSL............NCKRK...................     207
LST.CD.88.524               -AS.-QATAMRF............NCERK...................     206
LST.KE.x.lho7               -AK.IQNTAFAF............DCKRK...................     207
SUN.GA.x.SOL_36             E-F.KHVRSLQWEAD.................................     208
SUN.GA.98.L14               --E.FKHVTSLQWEV.................................     208
BLU.KE.x.KE31               QLE.RQTGKPQELKSALK..............................     239
SYK.KE.x.KE44               QLE.AQTGKPQTVKQALTGKGS..........................     243
SYK.KE.x.KE51               QLE.RQTG..................KR*.L.TAETTGYIS.......     239
SYK.KE.x.SYK173             QLE.RQTG..................KPQELQSALSK.NIS.......     240
COL.CM.x.CGU1               LSV.....................TSRKKKKPQQVAIEMVD.......     223
```