

# VII

## PLV Proteins

### Contents

---

VII-1	Introduction . . . . .	<b>401</b>
VII-2	Sequences . . . . .	<b>402</b>
VII-3	Alignments . . . . .	<b>410</b>
	VII-3.1 Gag . . . . .	410
	VII-3.2 Pol . . . . .	414
	VII-3.3 Vif . . . . .	421
	VII-3.4 Vpx . . . . .	423
	VII-3.5 Vpr . . . . .	424
	VII-3.6 Tat . . . . .	425
	VII-3.7 Rev . . . . .	427
	VII-3.8 Vpu . . . . .	429
	VII-3.9 Env . . . . .	430
	VII-3.10Nef . . . . .	437

---

### VII-1 Introduction

The selection of Primate Lentivirus Protein sequences for the following alignments was based on the sequences in the complete genome alignment as a starting alignment, and complete or nearly complete genes from other isolates were added if they increased the diversity of samples represented. For example, several diverse African green monkey virus isolates have been sequenced only in a region of the *env* gene, and recent Gorilla sequences are only available in the *pol* and *env* genes. When necessary, some of the more common sequences (such as HIV-1 M group) were removed to make room on the Compendium pages for these diverse virus sequences. More complete alignments are available from our web site <http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html> where space limitations are not an issue.

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

## VII-2 Sequences

Sequences included in the PLV protein alignments.

Name	Accession	Proteins	Author	Reference
H1B.FR.83.HXB2	K03455	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	van Beveren, C.P.	(in) Weiss,RL, Teich,N, Varmus,H and Coffin,J(Eds); RNA TUMOR VIRUSES, SECOND EDITION, 2, Vol 2: 1102-1123; Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
H1A1.UG.85.U455	M62320	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Oram, J.D.	ARHR 6(9):1073-1078 (1990)
H1B.US.90.WEAU160	U21135	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Wei, X.	Nature 422(6929):307-312 (2003)
H1C.ET.86.ETH2220	U46016	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Salminen, M.O.	ARHR 12(14):1329-1339 (1996)
H1D.CD.84.84ZR085	U88822	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	J Virol 72(7):5680-5698 (1998)
H1F1.BE.93.VI850	AF077336	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T.	Virology 269(1):95-104 (2000)
H1G.SE.93.SE6165	AF061642	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Carr, J.K.	Virology 247(1):22-31 (1998)
H1H.CF.90.056	AF005496	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	J Virol 72(7):5680-5698 (1998)
H1J.SE.93.SE7887	AF082394	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T.	ARHR 15(3):293-297 (1999)
H1K.CM.96.MP535	AJ249239	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Triques, K.	ARHR 16(2):139-151 (2000)
H101_AE.TH.90.CM240	U54771	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Carr, J.K.	J Virol 70(9):5935-5943 (1996)

Name	Accession	Proteins	Author	Reference
H102_AG.NG.x.IBNG	L39106	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Howard, T.M.	<i>ARHR</i> <b>10</b> (12):1755-1757 (1994)
H103_AB.RU.97.KAL153_2	AF193276	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Liitsola, K.	<i>ARHR</i> <b>16</b> (11):1047-1053 (2000)
H104_cpx.CY.94.CY032	AF049337	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	<i>J Virol</i> <b>72</b> (12):10234-10241 (1998)
H1N.CM.95.YBF30	AJ006022	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Simon, F.	<i>Nat Med</i> <b>4</b> (9):1032-1037 (1998)
H1O.BE.87.ANT70	L20587	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, M.	<i>J Virol</i> <b>68</b> (3):1586-1596 (1994)
H1O.CM.91.MVP5180	L20571	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gurtler, L.G.	<i>J Virol</i> <b>68</b> (3):1581-1585 (1994)
H1P.FR.06.RBF168	GQ328744	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env	Plantier, J.-C.	<i>Nat Med</i> <b>15</b> (8); 871-2 (2009)
CPZ.CM.01.SIVcpzCAM13	AY169968	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Nerrienet, E.	<i>J Virol</i> <b>79</b> (2):1312-9 (2005)
CPZ.CM.05.SIVcpzEK505	DQ373065	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	<i>Science</i> <b>313</b> (5786):523-526 (2006)
CPZ.CM.05.SIVcpzLB7	DQ373064	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	<i>Science</i> <b>313</b> (5786):523-526 (2006)
CPZ.CM.05.SIVcpzMB66	DQ373063	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	<i>Science</i> <b>313</b> (5786):523-526 (2006)
CPZ.CM.05.SIVcpzMT145	DQ373066	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, B.F.	<i>Science</i> <b>313</b> (5786):523-526 (2006)
CPZ.CM.-.SIVcpzMB897	EF535994	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Heuverswyn, F.V.	<i>Virology</i> <b>368</b> (1):155-171 (2007)

Name	Accession	Proteins	Author	Reference
CPZ.CM.-.SIVcpzDP943	EF535993	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Heuverswyn, F.V.	<i>Virology</i> <b>368</b> (1):155-171 (2007)
CPZ.CM.98.CAM3	AF115393	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env	Corbet, S.	<i>J Virol</i> <b>74</b> (1):529-534 (2000)
CPZ.CM.98.CAM5	AJ271369	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Muller-Trutwin, M.C.	<i>J Med Primatol</i> <b>29</b> (3-4); 166-72 (2000)
CPZ.GA.88.GAB1	X52154	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Huet, T.	<i>Nature</i> <b>345</b> (6273):356-359 (1990)
CPZ.GA.88.GAB2	AF382828	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bibollet-Ruche, F.	<i>ARHR</i> <b>20</b> (12):1377-1381 (2004)
CPZ.US.85.CPZUS	AF103818	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F.	<i>Nature</i> <b>397</b> (6718):436-441 (1999)
CPZ.CD.90.ANT	U42720	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, M.M.	<i>Virology</i> <b>221</b> (2):346-350 (1996)
CPZ.TZ.00.TAN1	AF447763	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Santiago, M.L.	<i>J Virol</i> <b>77</b> (3):2233-2242 (2003)
CPZ.TZ.2001.TAN2	EF394357	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	<i>J Virol</i> <b>81</b> (14):7463-7475 (2007)
CPZ.TZ.2002.TAN3_1	DQ374658	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	<i>J Virol</i> <b>81</b> (14):7463-7475 (2007)
GOR.CM.2007.SIVgor2139_287	FJ424866	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	<i>J Virol</i> <b>83</b> (4):1635-1648 (2009)
GOR.CM.2007.SIVgorCP2135con	FJ424863	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	<i>J Virol</i> <b>83</b> (4):1635-1648 (2009)
GOR.CM.2004.SIVgorCP684con	FJ424871	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Takehisa, J.	<i>J Virol</i> <b>83</b> (4):1635-1648 (2009)

Name	Accession	Proteins	Author	Reference
H2A.DE.x.BEN	M30502	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kirchhoff, F.	<i>Virology</i> <b>177</b> (1):305-311 (1990)
H2A.GW.x.ALI	AF082339	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Azevedo-Pereira	Unpublished
H2A.SN.x.ST	M31113	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kumar, P.	<i>J Virol</i> <b>64</b> (2):890-901 (1990)
H2B.CL.x.EHO	U27200	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Rey-Cuille, M.A.	<i>Virology</i> <b>202</b> (1):471-476 (1994)
H2B.GH.86.D205	X61240	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Dietrich, U.	<i>Nature</i> <b>342</b> (6252):948-950 (1989)
H2G.CI.92.ABT96	AF208027	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Brennan, C.A.	<i>ARHR</i> <b>13</b> (5):401-404 (1997)
H2U.FR.96.12034	AY530889	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Damond, F.	<i>ARHR</i> <b>20</b> (6):666-672 (2004)
SMM.SL.92.SL92B	AF334679	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Chen, Z.	<i>J Virol</i> <b>70</b> (6):3617-3627 (1996)
SMM.US.x.H9	M80194	Gag, Pol, Vif, Vpx, Vpr, Tat	Courgnaud, V.	<i>J Virol</i> <b>66</b> (1):414-419 (1992)
SMM.US.x.PGM53	AF077017	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Novembre, F.J.	<i>J Virol</i> <b>72</b> (11):8841-8851 (1998)
SMM.US.x.SIVsmH635F_L3	DQ201172	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kuwata, T.	<i>J Virol</i> <b>80</b> (3):1463-75 (2006)

Name	Accession	Proteins	Author	Reference
MAC.US.x.239	M33262	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kestler, H.	<i>Science</i> <b>248</b> (4959):1109-1112 (1990)
MAC.US.x.251_1A11	M76764	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Marthas, M.L.	<i>J Med Primatol</i> <b>18</b> (3-4):311-9 (1989)
MAC.US.x.251_BK28	M19499	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hirsch, V.	<i>Cell</i> <b>49</b> (3):307-319 (1987)
MAC.US.x.EMBL_3	Y00295	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Franchini, G.	<i>Nature</i> <b>328</b> (6130):539-543 (1987)
STM.US.x.STM	M83293	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Novembre, F.J.	<i>Virology</i> <b>186</b> (2):783-787 (1992)
MNE.US.x.MNE027	U79412	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kimata, J.T.	<i>J Virol</i> <b>72</b> (1):245-256 (1998)
DEB.CM.99.CM40	AY523865	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	<i>J Virol</i> <b>78</b> (14):7748-7762 (2004)
DEB.CM.99.CM5	AY523866	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	<i>J Virol</i> <b>78</b> (14):7748-7762 (2004)
DEN.CD.x.CD1	AJ580407	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Dazza, M.C.	<i>J Virol</i> <b>79</b> (13):8560-8571 (2005)
LST.CD.88.447	AF188114	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	<i>J Virol</i> <b>74</b> (8):3892-3898 (2000)
LST.CD.88.485	AF188115	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	<i>J Virol</i> <b>74</b> (8):3892-3898 (2000)

Name	Accession	Proteins	Author	Reference
LST.CD.88.524	AF188116	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	<i>J Virol</i> <b>74</b> (8):3892-3898 (2000)
LST.KE.x.lho7	AF075269	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	<i>J Virol</i> <b>73</b> (2):1036-1045 (1999)
SUN.GA.98.L14	AF131870	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	<i>J Virol</i> <b>73</b> (9):7734-7744 (1999)
MND-1.GA.x.MNDGB1	M27470	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Tsujimoto, H.	<i>Nature</i> <b>341</b> (6242):539-541 (1989)
OLC.CI.97.97CI12	FM165200	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	<i>J Virol</i> <b>83</b> (1):428-439 (2009)
WRC.CI.97.97CI14	AM745105	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	<i>J Virol</i> <b>83</b> (1):428-439 (2009)
WRC.CI.98.98CI04	AM713177	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	<i>J Virol</i> <b>83</b> (1):428-439 (2009)
WRC.GM.05.Pbt_05GM_X02	AM937062	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Locatelli, S.	<i>Virology</i> <b>376</b> (1):90-100 (2008)
MND-2.CM.98.CM16	AF367411	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Nef	Takehisa, J.	<i>ARHR</i> <b>17</b> (12):1143-1154 (2001)
MND-2.GA.x.M14	AF328295	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Souquiere, S.	<i>J Virol</i> <b>75</b> (15):7086-7096 (2001)
MND-2.x.x.5440	AY159322	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	<i>J Virol</i> <b>77</b> (8):4867-4880 (2003)
DRL.x.x.FAO	AY159321	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J.	<i>J Virol</i> <b>77</b> (8):4867-4880 (2003)

Name	Accession	Proteins	Author	Reference
MON.CM.99.L1	AY340701	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	<i>J Virol</i> <b>77</b> (23):12523-12534 (2003)
MON.NG.x.NG1	AJ549283	Gag, Pol, Vif, Vpr, Tat, Rev, Env	Barlow, K.L.	<i>J Virol</i> <b>77</b> (12):6879-6888 (2003)
GSN.CM.99.CN166	AF468659	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	<i>J Virol</i> <b>76</b> (16):8298-8309 (2002)
GSN.CM.99.CN71	AF468658	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	<i>J Virol</i> <b>76</b> (16):8298-8309 (2002)
TAL.CM.00.266	AY655744	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	<i>Virology</i> <b>349</b> (1):55-65 (2006)
TAL.CM.01.8023	AM182197	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F.	<i>Virology</i> <b>349</b> (1):55-65 (2006)
MUS-1.CM.01.1085	AY340700	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V.	<i>J Virol</i> <b>77</b> (23):12523-12534 (2003)
MUS-1.CM.01.CM1239	EF070330	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	<i>Virology</i> <b>360</b> (2):407-418 (2007)
MUS-2.CM.01.CM1246	EF070329	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	<i>Virology</i> <b>360</b> (2):407-418 (2007)
MUS-2.CM.01.CM2500	EF070331	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, A.F.	<i>Virology</i> <b>360</b> (2):407-418 (2007)
SYK.KE.x.KE51	AY523867	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	<i>J Virol</i> <b>78</b> (14):7748-7762 (2004)
SYK.KE.x.SYK173	L06042	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	<i>J Virol</i> <b>67</b> (3):1517-1528 (1993)
RCM.GA.x.GAB1	AF382829	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F.	<i>J Virol</i> <b>78</b> (14):7748-7762 (2004)



Name	Accession	Proteins	Author	Reference
RCM.NG.x.NG411	AF349680	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Beer, B.E.	<i>J Virol</i> <b>75</b> (24):12014-12027 (2001)
GRV.ET.x.GRI_677	M66437	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Fomsgaard, A.	<i>Virology</i> <b>182</b> (1):397-402 (1991)
SAB.SN.x.SAB1C	U04005	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Jin, M.J.	<i>EMBO J</i> <b>13</b> (12):2935-2947 (1994)
TAN.UG.x.TAN1	U58991	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Soares, M.A.	<i>Virology</i> <b>228</b> (2):394-399 (1997)
VER.DE.x.AGM3	M30931	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Baier, M.	<i>Virology</i> <b>176</b> (1):216-221 (1990)
VER.KE.x.9063	L40990	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, V.M.	<i>J Virol</i> <b>69</b> (2):955-967 (1995)
VER.KE.x.AGM155	M29975	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Johnson, P.R.	<i>J Virol</i> <b>64</b> (3):1086-1092 (1990)
VER.KE.x.TYO1_patent	DJ048201	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Omori, T.	Patent: WO 2007049749-A 13 03-MAY-2007; Dनावेक CORPORATION
COL.CM.x.CGU1	AF301156	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Courgnaud, V.	<i>J Virol</i> <b>75</b> (2):857-866 (2001)

	membrane binding		nuclear localization		phosphorylation site		p17 end_p24 start
	Gag start_p17 start	nuclear localization	nuclear localization	phosphorylation site	phosphorylation site		
H1B_FR.83.HXB2	-----	-----	-----	-----	-----	-----	-----
H1A1_UG.85_UJ455	-----	-----	-----	-----	-----	-----	-----
H1B_US.90.WEAU160	-----	-----	-----	-----	-----	-----	-----
H1C_ET.86.ETH2220	-----	-----	-----	-----	-----	-----	-----
H1D_CD.84.84ZRB885	-----	-----	-----	-----	-----	-----	-----
H1J_BE.93.VJ785	-----	-----	-----	-----	-----	-----	-----
H1G_SE.93.SE6165	-----	-----	-----	-----	-----	-----	-----
H1H_CF.90.056	-----	-----	-----	-----	-----	-----	-----
H1J_SE.93.SE7887	-----	-----	-----	-----	-----	-----	-----
H1K_CM.96.MP535	-----	-----	-----	-----	-----	-----	-----
H101_AE.TH.90.CM240	-----	-----	-----	-----	-----	-----	-----
H102_AG.NG.1E6C	-----	-----	-----	-----	-----	-----	-----
H103_AB.RU.97.KAL153.2	-----	-----	-----	-----	-----	-----	-----
H104_CPX.CY.94.CY032	-----	-----	-----	-----	-----	-----	-----
H1N_CM.95.YBF30	-----	-----	-----	-----	-----	-----	-----
H10_BE.87.ANT76	-----	-----	-----	-----	-----	-----	-----
H10_CM.91.MVP5180	-----	-----	-----	-----	-----	-----	-----
H1P_FR.06.RBF168	-----	-----	-----	-----	-----	-----	-----
CPZ_CM.01.SIVcpzCAM13	-----	-----	-----	-----	-----	-----	-----
CPZ_CM.05.SIVcpzEK505	-----	-----	-----	-----	-----	-----	-----
CPZ_CM.05.SIVcpzLB7	-----	-----	-----	-----	-----	-----	-----
CPZ_CM.05.SIVcpzNB66	-----	-----	-----	-----	-----	-----	-----
CPZ_CM.05.SIVcpzMT145	-----	-----	-----	-----	-----	-----	-----
CPZ_CM.-SIVcpzMB897	-----	-----	-----	-----	-----	-----	-----
CPZ_CM.-SIVcpzDP943	-----	-----	-----	-----	-----	-----	-----
CPZ_CM.98.CAM3	-----	-----	-----	-----	-----	-----	-----
CPZ_CM.98.CAM5	-----	-----	-----	-----	-----	-----	-----
CPZ_GA.88.GAB1	-----	-----	-----	-----	-----	-----	-----
CPZ_GA.88.GAB2	-----	-----	-----	-----	-----	-----	-----
CPZ_US.85.CPZUS	-----	-----	-----	-----	-----	-----	-----
CPZ_CD.90.ANT	-----	-----	-----	-----	-----	-----	-----
CPZ_TZ.00.TAN1	-----	-----	-----	-----	-----	-----	-----
CPZ_TZ.2001.TAN2	-----	-----	-----	-----	-----	-----	-----
CPZ_TZ.2002.TAN3.1	-----	-----	-----	-----	-----	-----	-----
GOR_CM.2007.SIVgor2139.287	-----	-----	-----	-----	-----	-----	-----
GOR_CM.2007.SIVgorCP2135.con	-----	-----	-----	-----	-----	-----	-----
GOR_CM.2004.SIVgorCP684.con	-----	-----	-----	-----	-----	-----	-----
H2A_DE.X.BEN	-----	-----	-----	-----	-----	-----	-----
H2A_RW.X.ALI	-----	-----	-----	-----	-----	-----	-----
H2A_SN.X.ST	-----	-----	-----	-----	-----	-----	-----
H2B_CI.X.HO	-----	-----	-----	-----	-----	-----	-----
H2B_GH.86.D205	-----	-----	-----	-----	-----	-----	-----
H2C_GI.92.ABT96	-----	-----	-----	-----	-----	-----	-----
H2U_FR.98.12833	-----	-----	-----	-----	-----	-----	-----
SMM_SL.92.SL92B	-----	-----	-----	-----	-----	-----	-----
SMM_US.X.H9	-----	-----	-----	-----	-----	-----	-----
SMM_US.X.PGM53	-----	-----	-----	-----	-----	-----	-----
SMM_US.X.SIVsmH635F_L3	-----	-----	-----	-----	-----	-----	-----
MAC_US.X.239	-----	-----	-----	-----	-----	-----	-----
MAC_US.X.251.IA11	-----	-----	-----	-----	-----	-----	-----
MAC_US.X.251.BK28	-----	-----	-----	-----	-----	-----	-----
MAC_US.X.EMBL.3	-----	-----	-----	-----	-----	-----	-----
STM_US.X.STM	-----	-----	-----	-----	-----	-----	-----
MNE_US.X.MNE027	-----	-----	-----	-----	-----	-----	-----
DEB_CM.99.CM40	-----	-----	-----	-----	-----	-----	-----
DEB_CM.99.CM5	-----	-----	-----	-----	-----	-----	-----
DEN_CD.X.CD1	-----	-----	-----	-----	-----	-----	-----
LST_CD.88.447	-----	-----	-----	-----	-----	-----	-----
LST_CD.88.485	-----	-----	-----	-----	-----	-----	-----
LST_CD.88.524	-----	-----	-----	-----	-----	-----	-----
LST_KE.X.lho7	-----	-----	-----	-----	-----	-----	-----
SUN_GA.98.L14	-----	-----	-----	-----	-----	-----	-----
MND.1.GA.X.MNDGB1	-----	-----	-----	-----	-----	-----	-----
OLC_CI.97.97CI12	-----	-----	-----	-----	-----	-----	-----
WRC_CI.97.97CI14	-----	-----	-----	-----	-----	-----	-----
WRC_CI.98.98CI04	-----	-----	-----	-----	-----	-----	-----
WRC_GM.05.Pbt.05GM.X02	-----	-----	-----	-----	-----	-----	-----
MND.2.GA.X.M14	-----	-----	-----	-----	-----	-----	-----
MND.2.X.X.5440	-----	-----	-----	-----	-----	-----	-----
DRL.X.X.FA0	-----	-----	-----	-----	-----	-----	-----
MON_CM.99.X.L1	-----	-----	-----	-----	-----	-----	-----
MON_NG.X.NG11	-----	-----	-----	-----	-----	-----	-----
GSN_CM.99.CN166	-----	-----	-----	-----	-----	-----	-----
GSN_CM.99.CN71	-----	-----	-----	-----	-----	-----	-----
TAL_CM.00.266	-----	-----	-----	-----	-----	-----	-----
TAL_CM.01.8023	-----	-----	-----	-----	-----	-----	-----
MUS.1.CM.01.1085	-----	-----	-----	-----	-----	-----	-----
MUS.1.CM.01.CM1239	-----	-----	-----	-----	-----	-----	-----
MUS.2.CM.01.CM1246	-----	-----	-----	-----	-----	-----	-----
MUS.2.CM.01.CM2500	-----	-----	-----	-----	-----	-----	-----
SYK_KE.X.KE51	-----	-----	-----	-----	-----	-----	-----
SYK_KE.X.SYK173	-----	-----	-----	-----	-----	-----	-----
RCM_GA.X.GAB1	-----	-----	-----	-----	-----	-----	-----
RCM_NG.X.NG411	-----	-----	-----	-----	-----	-----	-----
GRV_ET.X.GRI.677	-----	-----	-----	-----	-----	-----	-----
SAB_SN.X.SAB1C	-----	-----	-----	-----	-----	-----	-----
TAN_UG.X.TAN1	-----	-----	-----	-----	-----	-----	-----
VER_DE.X.AGM3	-----	-----	-----	-----	-----	-----	-----
VER_KE.X.12033	-----	-----	-----	-----	-----	-----	-----
VER_KE.X.AGM155	-----	-----	-----	-----	-----	-----	-----
VER_KE.X.TY01.patent	-----	-----	-----	-----	-----	-----	-----
COL_CM.X.COL	-----	-----	-----	-----	-----	-----	-----



Accession	PLV protein	p24 end_p2 start	p2 end_p7 start	Zn motif	Zn motif	Accession
H1B_FR.83.HXB2		QAQSQEVKNMTE	LLVQNA	DPCKTILKALGPA	ATLEEMMTACQGVGG	PGPKARVLAEAM
H1A1_UG.85.U455		S	R	G		
H1B_US.90.WEAU160		T	D			
H1C_ET.86.ETH2220		T	D			
H1D_CD.84.84ZRB85						
H1F1_BE.93.VT05						
H1G_SE.93.SE6165						
H1H_CF.90.056						
H1J_SE.93.SE7887						
H1K_CM.96.MP535						
H101_AE.TH.90.CH240						
H102_AG.NG.x.IE6C						
H103_AB.RU.97.KAL153_2						
H104_cpx.CY.94.CY032						
H1N_CM.95.YBF30						
H10_BE.87.ANT76						
H10_CM.91.MV95180						
H1P_FR.06.RBF168						
CPZ_CM.01.SIVcpzCAM13						
CPZ_CM.05.SIVcpzEK505						
CPZ_CM.05.SIVcpzLB7						
CPZ_CM.05.SIVcpzMT145						
CPZ_CM.05.SIVcpzMB897						
CPZ_CM.05.SIVcpzDP943						
CPZ_CM.98.CAM3						
CPZ_CM.98.CAM5						
CPZ_GA.88.GAB1						
CPZ_GA.88.GAB2						
CPZ_US.85.CPZUS						
CPZ_CD.90.ANT						
CPZ_TZ.90.TAN1						
CPZ_ZI.2001.TAN2						
CPZ_TZ.2002.TAN3_1						
GOR_CM.2007.SIVgor2139_287						
GOR_CM.2007.SIVgorCP2135con						
GOR_CM.2004.SIVgorCP684con						
H2A_GE.x.BEN						
H2A_GW.x.ALI						
H2A_SN.x.ST						
H2B_CI.x.EHO						
H2B_GH.86.D205						
H2G_CI.92.ABT96						
H2U_FR.96.I2834						
SMM_SL.92.SL92B						
SMM_US.x.H9						
SMM_US.x.PGM53						
SMM_US.x.SIVsmH635F_L3						

Accession	PLV protein	p27_p2	p2_p8	Zn motif	Zn motif	Accession
MAC_US.x.239		-TDA	A-Q	-I	-LV	-G
MAC_US.x.251_1A11		-TDA	A-Q	-I	-LV	-G
MAC_US.x.251_BK28		-TDA	A-Q	-I	-LV	-G
MAC_US.x.EMBL_3		-TDA	A-Q	-I	-LV	-G
STM		-TDA	A-Q	-I	-LV	-G
MNE_US.x.MNE027		-TDA	A-Q	-I	-LV	-G
DEB_CM.99.CM40		-D	A	-T	M	I
DEB_CM.99.CM5		-D	A	-T	M	I
DEN_CD.x.C01		-D	S	-L	T	S
LST_CD.88.447		GG	H	-E	K	KM
LST_CD.88.485		GG	H	-E	K	KM
LST_CD.88.524		GG	H	-E	K	KM
LST_KE.x.lho7		GG	H	-E	K	KM
SUN_GA.98.L14		SG	E	-G	K	KM
MND_1_GA.x.HNDGB1		-G	-M	-O	Q	H
QLC_CI.97.97C112		-G	-M	-O	Q	H
WRC_CI.97.97C114		-G	-M	-O	Q	H
WRC_CI.98.98CI04		-G	-M	-O	Q	H
WRC_GM.05.Pbt_05GM_X02		-G	-M	-O	Q	H
MND_2_CM.98.CM16		-G	-M	-O	Q	H
MND_2_GA.x.M14		-G	-M	-O	Q	H
MND_2.x.x.5440		-G	-M	-O	Q	H
DRL_x.x.FA0		-G	-M	-O	Q	H
MON_CM.99.L1		-G	-M	-O	Q	H
MON_NG.x.NG1		-G	-M	-O	Q	H
GSM_CM.99.GSM166		-G	-M	-O	Q	H
GSM_CM.99.CN71		-G	-M	-O	Q	H
TAL_CM.00.266		-G	-M	-O	Q	H
TAL_CM.01.8023		-G	-M	-O	Q	H
MUS_1_CM.01.1085		-G	-M	-O	Q	H
MUS_1_CM.01.CM1239		-G	-M	-O	Q	H
MUS_1_CM.01.CM1246		-G	-M	-O	Q	H
MUS_2_CM.01.CM2500		-G	-M	-O	Q	H
SYK_KE.x.KE51		-D	A	-Q	S	I
SYK_KE.x.SYK173		-D	A	-Q	S	I
VER_GA.x.GAB1		-D	A	-Q	S	I
VER_NG.x.NG411		-D	A	-Q	S	I
VER_KE.x.AGM155		-D	A	-Q	S	I
VER_KE.x.TY01_patent		-D	A	-Q	S	I
COL_CM.x.CGU1		-P	-A	-G	-I	-A

	p7_end_p1_start	p1_end_p6_start	Vpr_binding	Vpr_binding	Gag_end				
	ERQANFLGKZIP	SYKGR	PGNFLOS	RPE	PTAPPEES	FRSGVE	TTTTPOKQEP	DKEYL	PLTSLRSLFGNDPPSSQ*
H1B.FR.83.HXB2	.....N	.....P	.....P	.....L	.....A-I	.....-GM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1A1.UG.85.U455	.....S	.....Q	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1B.US.90.WEAU160	.....R-L	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1C.ET.86.ETH2220	.....H	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1D.CO.84.84ZRB85	.....N	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1F1.BE.93.VT850	.....N	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1G.SE.93.SE6165	.....S	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1H.CF.90.056	.....S	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1J.SE.93.SE7887	.....H	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1K.CM.96.MP535	.....H	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H101.AE.TH.90.CH240	.....W	.....P	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H102.AG.NG.x.IE8G	.....R	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H103.AB.RU.97.KAL153.2	.....R-M	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H104.cpx.CY.94.CY032	.....E	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1N.CM.95.YBF30	.....GK	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H10.BE.87.ANT76	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1C.CM.91.MVP180	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
H1P.FR.06.RBF168	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.CM.01.SIVcpzCAM13	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.CM.05.SIVcpzEK505	.....GK	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.CM.05.SIVcpzLBT7	.....NN	.....G-I	.....N	.....T	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.CM.05.SIVcpzNB66	.....MOR	.....PR	.....N	.....T	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.CM.05.SIVcpzMT145	.....GD	.....G-S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.CM.05.SIVcpzMB897	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.CM.05.SIVcpzDP943	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.CM.98.CAM3	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.CM.98.CAM5	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.GA.88.GAB1	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QT	.....V-K
CPZ.GA.88.GAB2	.....EG	.....V	.....G-S	.....P	.....L	.....A-I	.....K-M-S-A	.....L-K-R-QH	.....V-K
CPZ.US.85.CPZUS	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
CPZ.CD.90.ANT	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
CPZ.TZ.00.TAN1	.....RN	.....N	.....T	.....T	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
CPZ.TZ.2001.TAN2	.....KNMNSPGV	.....RT	.....LWGC	.....V-N	.....T	.....A-I	.....K-M-S-A	.....L-K-R-QH	.....V-K
CPZ.TZ.2002.TAN3.1	.....KNMNSPGV	.....RT	.....LWGC	.....V-N	.....T	.....A-I	.....K-M-S-A	.....L-K-R-QH	.....V-K
GOR.CM.2007.SIVgor2139.287	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
GOR.CM.2007.SIVgorCP2135.con	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
GOR.CM.2004.SIVgorCP684.con	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
H2A.DE.x.BE9	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
H2A.CW.x.ALI1	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
H2A.SN.x.ST	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
H2B.CI.x.H0	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
H2B.GH.86.D205	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
H2C.CI.92.ABT96	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
H2U.FR.06.G234	.....TKVIG	.....LGPW	.....GKK	.....R	.....PMA	.....QVH	.....GGLM	.....DPA	.....VDLKNYMLGKQ
SMM.SL.92.S192B	.....NQK	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH
SMM.US.x.H9	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
SMM.US.x.PGM53	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K
SMM.US.x.SIVsmH635F.L3	.....G	.....S	.....P	.....L	.....A-I	.....-FM	.....K-M-S-A	.....L-K-R-QH	.....V-K

Gag-Pol TF start		Gag-Pol TF end protease start	
H1B.FR.83.HXB2	FFREDLAF . . . . . LOGKAREFSSQETRAN . . . . . SPTPRRE . . . . . LOVWGRD . . . . . NNSPSEAG . . . . . ADROG . . . . . TVSFNFPVTLWQRPLVTIKIGQGLKEALLDGTADDTVLEMSL . . . . . PGRWPKPMIGGTGGFKVRYQDYLIIIEICGHKAI . . . . . G	129	
H1A1.UG.85.U455	-N . . . . . Q . . . . . E . . . . .	-W . . . . . D . . . . . G . . . . . K . . . . . I . . . . .	-D . . . . . I . . . . .
H1B.US.90.WEAU160	-V . . . . . PK . . . . . T . . . . .	-Q . . . . . N . . . . . A . . . . .	-E . . . . . I . . . . .
H1C.ET.86.ETH2220	-T . . . . . Q . . . . . P . . . . . SPTRESOTRAN . . . . . T . . . . .	-R . . . . . S . . . . .	-T . . . . . L . . . . .
H1D.CO.84.84Z8885	-N . . . . . P . . . . . G . . . . . L . . . . .	-F . . . . . G . . . . . F . . . . .	-T . . . . . L . . . . .
H1F1.BE.93.01059	-Q . . . . . N . . . . . R . . . . . K . . . . . P . . . . .	-E . . . . . R . . . . . T . . . . .	-V . . . . . I . . . . .
H1G.SE.93.5E6165	-N . . . . . Q . . . . . E . . . . . K . . . . . P . . . . . D . . . . . T . . . . .	-C . . . . . K . . . . .	-P . . . . . R . . . . .
H1H.CF.90.056	-N . . . . . Q . . . . . R . . . . . K . . . . . P . . . . . A . . . . . T . . . . .	-C . . . . . K . . . . .	-P . . . . . R . . . . .
H1J.SE.93.5E7887	-N . . . . . Q . . . . . R . . . . . E . . . . . L . . . . . P . . . . .	-S . . . . .	-P . . . . . R . . . . .
H1K.CM.96.MP535	-P . . . . . E . . . . .	-S . . . . .	-P . . . . . R . . . . .
H101.AE.TH.90.CM240	-N . . . . . Q . . . . . E . . . . . G . . . . .	-S . . . . . K . . . . .	-G . . . . . D . . . . .
H102.AG.NG.x.IEBC	-N . . . . . Q . . . . . R . . . . . E . . . . . K . . . . . GT . . . . .	-S . . . . . S . . . . . K . . . . .	-G . . . . . D . . . . .
H103.AB.RU.97.KAL153.2	-N . . . . . Q . . . . . R . . . . . E . . . . . K . . . . . I . . . . .	-S . . . . . S . . . . . K . . . . .	-W . . . . . D . . . . .
H104.cpx.CY.94.CY032	-NV . . . . . Q . . . . . R . . . . . E . . . . . K . . . . . A . . . . .	-S . . . . . P . . . . .	-W . . . . . D . . . . .
H1N.CM.95.YBF30	-E . . . . . VS . . . . .	-R . . . . . E . . . . . T . . . . .	-W . . . . . S . . . . .
H10.BE.87.AN176	-QT . . . . . S . . . . .	-G . . . . . G . . . . .	-O . . . . . L . . . . .
H10.CM.91.MVPS180	-W . . . . . P . . . . .	-G . . . . . G . . . . .	-O . . . . . L . . . . .
H1P.FR.06.RBF168	-K . . . . . S . . . . .	-W . . . . . G . . . . .	-O . . . . . L . . . . .
CPZ.CM.01.SIVcpzCAM13	-G . . . . .	-P . . . . . R . . . . .	-E . . . . .
CPZ.CM.05.SIVcpzEK505	-G . . . . . S . . . . .	-P . . . . . R . . . . .	-E . . . . .
CPZ.CM.05.SIVcpzLB7	-T . . . . .	-P . . . . .	-K . . . . .
CPZ.CM.05.SIVcpzMT145	-KK . . . . . S . . . . .	-G . . . . . E . . . . .	-L . . . . .
CPZ.CM. .SIVcpzMB897	-N . . . . . S . . . . .	-P . . . . . R . . . . .	-E . . . . .
CPZ.CM. .SIVcpzDP943	-R . . . . . VS . . . . .	-V . . . . . E . . . . .	-R . . . . .
CPZ.CM.98.CAM3	-R . . . . . VS . . . . .	-V . . . . . E . . . . .	-R . . . . .
CPZ.CM.98.CAM5	-G . . . . . VS . . . . .	-V . . . . . E . . . . .	-R . . . . .
CPZ.CM.98.CAM1	-G . . . . . VS . . . . .	-V . . . . . E . . . . .	-R . . . . .
CPZ.CM.98.GAB2	-GG . . . . .	-P . . . . . R . . . . .	-E . . . . .
CPZ.US.85.CPZUS	-T . . . . . VP . . . . .	-I . . . . . V . . . . .	-E . . . . . R . . . . .
CPZ.CO.90.ANT1	-TD . . . . . PH . . . . .	-P . . . . .	-K . . . . .
CPZ.TZ.00.TAN1	-TH . . . . . PL . . . . .	-P . . . . .	-K . . . . .
CPZ.TZ.001.TAN2	-AH . . . . . PL . . . . .	-P . . . . .	-K . . . . .
CPZ.TZ.2002.TAN3.1	-TH . . . . . PL . . . . .	-P . . . . .	-K . . . . .
GOR.CM.2007.SIVgor2139.287	-I . . . . . S . . . . .	-R . . . . . G . . . . .	-R . . . . .
GOR.CM.2007.SIVgorCP2135.con	-I . . . . . S . . . . .	-R . . . . . G . . . . .	-R . . . . .
GOR.CM.2004.SIVgorCP684.con	-T . . . . . S . . . . .	-R . . . . . G . . . . .	-R . . . . .
H2A.CE.x.BEN	-V . . . . . GP . . . . .	-K . . . . . E . . . . .	-P . . . . .
H2A.GW.x.ALI1	-V . . . . . GP . . . . .	-K . . . . . E . . . . .	-P . . . . .
H2A.SN.x.SI1	-V . . . . . GP . . . . .	-K . . . . . E . . . . .	-P . . . . .
H2B.CI.x.EH0	-V . . . . . RP . . . . .	-L . . . . .	-K . . . . .
H2B.GH.86.D205	-V . . . . . RL . . . . .	-L . . . . .	-K . . . . .
H2C.CI.92.AB196	-V . . . . . WT . . . . .	-L . . . . .	-K . . . . .
H2U.FR.96.12034	-A . . . . . MP . . . . .	-K . . . . .	-S . . . . .
SMM.SL.92.S192B	-P . . . . . PM . . . . .	-K . . . . .	-S . . . . .
SMM.US.x.H9	-A . . . . . WP . . . . .	-K . . . . .	-S . . . . .
SMM.US.x.PGM53	-A . . . . . WP . . . . .	-K . . . . .	-S . . . . .
SMM.US.x.SIVsmH635F.L3	-V . . . . . WM . . . . .	-K . . . . .	-S . . . . .
MAC.US.x.239	-P . . . . . WS . . . . .	-K . . . . .	-S . . . . .
MAC.US.x.251.1A11	-P . . . . . WS . . . . .	-K . . . . .	-S . . . . .
MAC.US.x.251.BK28	-P . . . . . WS . . . . .	-K . . . . .	-S . . . . .
MAC.US.x.EHBL.3	-P . . . . . WS . . . . .	-K . . . . .	-S . . . . .
STM.US.x.S1M	-L . . . . . WP . . . . .	-K . . . . .	-S . . . . .
MNE.US.x.MNE027	-L . . . . . WP . . . . .	-K . . . . .	-S . . . . .
DEB.CM.99.CM40	-Y . . . . . Y . . . . .	-L . . . . .	-K . . . . .
DEB.CM.99.CM5	-C . . . . . SL . . . . .	-L . . . . .	-K . . . . .
DEN.CO.x.C01	-W . . . . . SW . . . . .	-E . . . . .	-S . . . . .
LST.CO.88.447	-L . . . . . WT . . . . .	-L . . . . .	-K . . . . .
LST.CO.88.485	-L . . . . . WT . . . . .	-L . . . . .	-K . . . . .
LST.CO.88.524	-L . . . . . WT . . . . .	-L . . . . .	-K . . . . .
LST.KE.x.lh07	-L . . . . . WT . . . . .	-L . . . . .	-K . . . . .
SUN.GA.98.L14	-W . . . . . MP . . . . .	-L . . . . .	-K . . . . .
MND.1.GA.x.HNDGB1	-W . . . . . MP . . . . .	-L . . . . .	-K . . . . .
QLC.CI.97.97C114	-W . . . . . MP . . . . .	-L . . . . .	-K . . . . .
WRC.CI.97.97C114	-W . . . . . MP . . . . .	-L . . . . .	-K . . . . .
WRC.CI.98.98C104	-W . . . . . MP . . . . .	-L . . . . .	-K . . . . .
WRC.GM.05.Pbt.05GM.X02	-W . . . . . MP . . . . .	-L . . . . .	-K . . . . .
MND.2.CM.98.CM16	-Y . . . . . PL . . . . .	-L . . . . .	-K . . . . .
MND.2.GA.x.M14	-Y . . . . . PL . . . . .	-L . . . . .	-K . . . . .
MND.2.x.x.5440	-Y . . . . . PL . . . . .	-L . . . . .	-K . . . . .
DRL.x.x.FA0	-Y . . . . . Y . . . . .	-L . . . . .	-K . . . . .
MON.CM.99.L1	-S . . . . . FR . . . . .	-L . . . . .	-K . . . . .
MON.NG.x.NG1	-S . . . . . FR . . . . .	-L . . . . .	-K . . . . .
GSN.CM.99.CM166	-P . . . . . WP . . . . .	-K . . . . .	-S . . . . .
GSN.CM.99.CM77	-P . . . . . WP . . . . .	-K . . . . .	-S . . . . .
TAL.CM.00.266	-V . . . . . FR . . . . .	-L . . . . .	-K . . . . .
TAL.CM.01.8023	-G . . . . . FR . . . . .	-L . . . . .	-K . . . . .
MUS.1.CM.01.1085	-G . . . . . VS . . . . .	-L . . . . .	-K . . . . .
MUS.1.CM.01.CM1239	-G . . . . . VS . . . . .	-L . . . . .	-K . . . . .
MUS.1.CM.01.CM1246	-G . . . . . VS . . . . .	-L . . . . .	-K . . . . .
MUS.2.CM.01.CM2500	-R . . . . . FR . . . . .	-L . . . . .	-K . . . . .
SYK.KE.x.KE51	-R . . . . . V . . . . .	-L . . . . .	-K . . . . .
SYK.KE.x.SYK173	-R . . . . . V . . . . .	-L . . . . .	-K . . . . .
RCM.GA.x.GAB1	-T . . . . . RL . . . . .	-L . . . . .	-K . . . . .
RCM.NG.x.NG411	-C . . . . . SL . . . . .	-L . . . . .	-K . . . . .
GRV.ET.x.GRI.677	-V . . . . . WL . . . . .	-L . . . . .	-K . . . . .
SAB.SN.x.SAB1C	-V . . . . . WL . . . . .	-L . . . . .	-K . . . . .
TAN.UG.x.TAN1	-N . . . . . AN . . . . .	-L . . . . .	-K . . . . .
VER.DE.x.AGM3	-V . . . . . MP . . . . .	-L . . . . .	-K . . . . .
VER.KE.x.9063	-V . . . . . MP . . . . .	-L . . . . .	-K . . . . .
VER.KE.x.AGM155	-V . . . . . MP . . . . .	-L . . . . .	-K . . . . .
VER.KE.x.TY01.patent	-V . . . . . MP . . . . .	-L . . . . .	-K . . . . .
COL.CM.x.CGU1	-G . . . . . GN . . . . .	-L . . . . .	-K . . . . .

	protease end_p66_p51 RT start	M41L	K70R D67N <sup>1</sup>	D110 catalytic site		
H1B_FR.83.HXB2	TVLVGPTVNIIGRNL	TLQIGCTLNFPISPIETV	...PVKLLKPGMDGPKVQWPLTEEKIKALVET	TEMEKEGKISKIGPENPYNTVFATKKKDKSTKWRKLVDFRELNKRTDFWVQLGIPHPAGLKKKKS	...VTVLDVGDAYFSVPLDEDFRKYFTAFTPIS	290
H1A1_UG.85.U455	M		E		S	289
H1B_US.90.WEAU160	M	L	R			290
H1C_ET.86.ETH2220				T	A	297
H1D_CD.84.84ZRB85						290
H1F1_BE.93.VT850						287
H1G_SE.93.SE6165	I	I	M			289
H1H_CF.90.056						288
H1J_SE.93.SE7887	I	I	M	L		290
H1K_CM.96.MP535						295
H101_AE.TH.90.CM240						290
H102_AG.NG.x.IE86						290
H103_AB.RU.97.KAL153_2						290
H104_cpx.CY.94.CY032						290
H1N_CM.95.YBF30						297
H10_BE.87.ANT76						286
H10_CM.91.MVP5180						286
H1P_FR.06.RBF168	X	X	I	L	S	286
CPZ_CM.01.SIVcpzCAM13						289
CPZ_CM.05.SIVcpzEK505						296
CPZ_CM.05.SIVcpzLB7						293
CPZ_CM.05.SIVcpzMT145						289
CPZ_CM.SIVcpzMB897						298
CPZ_CM.SIVcpzDP943						297
CPZ_CM.98.CAM3						295
CPZ_CM.98.CAM5						291
CPZ_GA.88.GAB1						288
CPZ_GA.88.GAB2						293
CPZ_US.85.CPZUS						292
CPZ_CD.90.ANT						290
CPZ_TZ.00.TAN1						286
CPZ_TZ.2001.TAN2						287
CPZ_TZ.2002.TAN3_1						287
GOR_CM.2007.SIVgor2139_287						282
GOR_CM.2007.SIVgorCP2135con						282
GOR_CM.2004.SIVgorCP684con						282
H2A_NE.x.BEN	-IMT	D	I	F	-I	AV
H2A_GW.x.ALI	-IMT	D	I	F	-I	AV
H2A_SN.x.ST	-IMT	D	I	F	-I	AV
H2B_CI.x.EHO	-MT	D	I	F	-I	AV
H2B_GH.86.D205	-IMT	D	I	F	-I	AV
H2G_CI.92.ABT96	-MT	D	I	F	-I	AV
H2U_FR.96.L2034	-IMT	D	I	F	-I	AV
SMM_SL.92.SL92B	-LMT	D	I	F	-I	AV
SMM_US.x.H9	-IMT	D	I	F	-I	AV
SMM_US.x.PGM53	-IMT	D	I	F	-I	AV
SMM_US.x.SIVsmH635F_L3	-IMT	D	I	F	-I	AV
MAC_US.x.239	-IMT	D	I	F	-I	AV
MAC_US.x.251.IA11	-IMT	D	I	F	-I	AV
MAC_US.x.251.BK28	-IMT	D	I	F	-I	AV
MAC_US.x.EMBL_3	-IMT	D	I	F	-I	AV
STM	-IMT	D	I	F	-I	AV
MNE_US.x.MNE627	-IMT	D	I	F	-I	AV
DEB_CM.99.CM40	DL	D	I	L	Y	AV
DEB_CM.99.CM5	D	I	L	Y	AKM	AV
DEN_CD.x.CD1						316
LST_CD.88.447	L	D	V	I	NV	AV
LST_CD.88.485	L	D	V	I	SV	AV
LST_CD.88.524	L	D	V	I	NPL	AV
LST_KE.x.lho7	L	D	V	I	AM	AV
SUN_GA.98.L14						307
MND_1_GA.x.MNDGB1	A	D	DD	V	AV	AV
OLC_CI.97.97C112						299
WRC_CI.97.97C114						314
WRC_CI.98.98CI04						313
WRC_GM.05.Pbt_05GM_X02						305
MND_2_CM.98.CM16	S					302
MND_2_GA.x.M14	A					303
MND_2.x.x.5440	A					304
DRL.x.x.FA0	S					301
MON_CM.99.L1						298
MON_NG.x.NG1	L	X	I	L	V	AV
GSN_CM.99.CN166						298
GSN_CM.99.CN71						294
TAL_CM.00.266						299
TAL_CM.01.8023						300
MUS_1_CM.01.1085						291
MUS_1_CM.01.CM1239						296
MUS_2_CM.01.CM1246						295
MUS_2_CM.01.CM2500						328
SYK_KE.x.KE51	SL					322
SYK_KE.x.SYK173	L					304
RCM_GA.x.GAB1	H					308
RCM_NG.x.NG411	H					308
GRV_ET.x.GRI_677	SI	I	S	I	I	AV
SAB_SN.x.SAB1C	I	E				312
TAN_UG.x.TAN1	N	I	I	V	I	AV
VER_DE.x.AGM3	I	I	S	I	I	AV
VER_KE.x.AGM155	I	A	I	I	I	AV
VER_KE.x.AGM155	I	A	I	I	I	AV
VER_KE.x.TY01 patent	I	L	A	I	I	AV
COL_CM.x.CG11	DI	I	H	I	V	AV





H1B\_FR.83.HXB2
H1A1\_UG.85.U455
H1B\_US.90.WEAU160
H1C\_ET.86.ETH2220
H1D\_CD.84.84ZRO85
H1F1\_BE.93.X1850
H1G\_SE.93.SE6165
H1H\_CF.90.056
H1J\_SE.93.SE7887
H1K\_CM.96.MP535
H101\_AE.TH.90.CM240
H102\_AG.NG.x.IENG
H103\_AB.RU.97.KAL153.2
H104\_cpx.CY.94.CY032
H1N\_CM.95.YBF30
H10\_BE.87.ANT76
H10\_CM.91.MVP5180
H1P\_FR.06.RBF168
CPZ\_CM.01.SIVcpzCAM13
CPZ\_CM.05.SIVcpzEK505
CPZ\_CM.05.SIVcpzLB7
CPZ\_CM.05.SIVcpzMB66
CPZ\_CM.05.SIVcpzMT145
CPZ\_CM.SIVcpzMB897
CPZ\_CM.SIVcpzDP943
CPZ\_CM.98.CAM3
CPZ\_CM.98.CAM5
CPZ\_GA.88.GAB21
CPZ\_GA.88.GAB2
CPZ\_US.85.CPZUS
CPZ\_CD.90.ANT1
CPZ\_TZ.90.TAN1
CPZ\_TZ.2001.TAN2
CPZ\_TZ.2002.TAN3.1
GOR\_CM.2007.SIVgor2139.287
GOR\_CM.2007.SIVgorCP2135con
GOR\_CM.2004.SIVgorCP684con
H2A\_GE.x.BEN
H2A\_GW.x.ALI
H2A\_SN.x.ST
H2B\_CI.x.EHO
H2B\_GH.86.D205
H2C\_C1.92.ABT96
H2U\_FR.96.12034
SMM\_SL.92.SL92B
SMM\_US.x.H9
SMM\_US.x.PGM53
SMM\_US.x.SIVsmH635F\_L3

AEINREILKPEVHYVYDPSKDLIAEIKQKGGQWYIYQEP... FKNLTKGYARMGATNDNVKQLTEAVQKITTESIINGKTP... KFKLPIQKETWETWTFYQWATWIEPEWFVNTPLVLWYQLKLEKPIVGAETFPYVDGAANRETKLKGAGYVNTNRGRQKVVLTLD
D V L D K S V V VS I R A M D A V K I D R S E
L F ND F F F KRGT AV AL R A D A V K I D R S E
M L ND F F F KRGT AV AL R A D A V K I D R S E
R R E V LD N Y KV S V A L RSP L D TD T A D S K I D R S E
R R I R LD N Y KV S V A L RSP L D TD T A D S K I D R S E
K R I SA E V LD H KR S A V AL A G R R R D T M Y I S H T R DK IIS E
T V V ND H RGS R V VA R R R M D S I D S I D S E
T V L D KK S R V VAM R R R M D S I D S I D S E
A V L A V AM C R V D A D TD A Q D S SE
L A G E V V L H K S I V V R VA R V V A DH TD A Q D S SE
R Q Q D WVN GE V DE H T OKAS IR A VI VSO I L VTR AD S I R S M Y EQ K IIK DE
H10\_CM.91.MVP5180 H1P\_FR.06.RBF168 CPZ\_CM.01.SIVcpzCAM13 CPZ\_CM.05.SIVcpzEK505 CPZ\_CM.05.SIVcpzLB7 CPZ\_CM.05.SIVcpzMB66 CPZ\_CM.05.SIVcpzMT145 CPZ\_CM.SIVcpzMB897 CPZ\_CM.SIVcpzDP943 CPZ\_CM.98.CAM3 CPZ\_CM.98.CAM5 CPZ\_GA.88.GAB21 CPZ\_GA.88.GAB2 CPZ\_US.85.CPZUS CPZ\_CD.90.ANT1 CPZ\_TZ.90.TAN1 CPZ\_TZ.2001.TAN2 CPZ\_TZ.2002.TAN3.1 GOR\_CM.2007.SIVgor2139.287 GOR\_CM.2007.SIVgorCP2135con GOR\_CM.2004.SIVgorCP684con H2A\_GE.x.BEN H2A\_GW.x.ALI H2A\_SN.x.ST H2B\_CI.x.EHO H2B\_GH.86.D205 H2C\_C1.92.ABT96 H2U\_FR.96.12034 SMM\_SL.92.SL92B SMM\_US.x.H9 SMM\_US.x.PGM53 SMM\_US.x.SIVsmH635F\_L3

MAC\_US.x.239
MAC\_US.x.251.1A11
MAC\_US.x.251.BK28
MAC\_US.x.EMBL\_3
STM\_US.x.STM
MNE\_US.x.MNE027
DEB\_CM.99.CM40
DEB\_CM.99.CM5
DEN\_CD.x.C01
LST\_CD.88.447
LST\_CD.88.485
LST\_CD.88.524
LST\_KE.x.lho7
SUN\_GA.98.L14
MND\_1\_GA.x.MNDGB1
QLC\_CI.97.97C112
WRC\_CI.97.97C114
WRC\_CI.98.98C104
WRC\_GM.05.Pbt.05GM.X02
MND\_2\_CM.98.CM16
MND\_2\_GA.x.M14
MND\_2.x.x.5440
DRL.x.x.FAO
MON\_CM.99.L1
MON\_NG.x.NG1
GSM\_CM.99.CM166
GSM\_CM.99.CM171
TAL\_CM.00.266
TAL\_CM.01.8023
MUS\_1\_CM.01.1085
MUS\_1\_CM.01.CM1239
MUS\_2\_CM.01.CM1246
MUS\_2\_CM.01.CM2500
SYK\_KE.x.KE51
SYK\_KE.x.SYK173
RCM\_GA.x.GAB1
RCM\_NG.x.NG411
GRV\_ET.x.GRI.677
SAB\_SN.x.SABIC
TAN\_UG.x.TAN1
VER\_KE.x.AGM3
VER\_KE.x.AGM4
VER\_KE.x.AGM155
VER\_KE.x.TY01.patent
COL\_CM.x.CG11

E KI SOEPE C QEG P E TVI SODN S K H D I V F KIKNT G RL AHVI GK A QV H VE DV Q D V D D IS R VFN V D E E Y T SC KOS E I D KD KV EQ 637
E KI SOEPE C QEG P E TVI SODN S K H D I V F KIKNT G RL AHVI GK A QV H VE DV Q D V D D IS R VFN V D E E Y T SC KOS E I D KD KV EQ 637
E KI SOEPE C QEG P E TVI SODN S K H D I V F KIKNT G RL AHVI GK A QV H VE DV Q D V D D IS R VFN V D E E Y T SC KOS E I D KD KV EQ 637
E NI SOEPE C QEG P E TVI SODN S K H D I V F KIKNT G RL AHVI GK A QV H VE DV Q D V D D IS R VFN V D E E Y T SC KOS E I D KD KV EQ 637
E KI SOEPE C QEG P E TVI SODN S K H D I V F KIKNT G RL AHVI GK A QV H VE DV Q D V D D IS R VFN V D E E Y T SC KOS E I D KD KV EQ 637
E KI SOEPE C QEG P E TVI SODN S K H D I V F KIKNT G RL AHVI GK A QV H VE DV Q D V D D IS R VFN V D E E Y T SC KOS E I D KD KV EQ 637
R K Q E S KEQEP LS L K G V R PK GI FS DK Y FH AK MY G F RI Q R V E DN HNH A D AIH H E VS PD D Y S EW K S RC EN 631
R K TP Q E S KEQEP S S I R G V R TK GL FS KDK Y FH AK MY G Y V R V E DN HNH D A H Q R E VS PD D Y S ISEW K G KS EN 631
K A Q S S R E P S L EN G T K HL HKS S Y PYO AKVMA GR AL RL T R VS VD DA SDH V L S L RNF N VS DK P Y T Q M ATN K EE 637
EQ QO KLO A T E VVRV OSK II F WR GN SI RA R Q OKA PLOK A GK HV IOV VTR V DO SDH V L IS Q EQE S A E VD Y EKVGT OS KE KE 647
EQ QO KLO A T E VVRV OSK II F WR GN SI RA R Q OKA PLOK A GK HV IOV VTR V DO SDH V L IS Q EQE S A E VD Y EKVGT OS KE KE 643
ET KT Q A E VVRV ONK VI F WR GN SI RA R Q OKS M PLOK A GK HV IOV VTR V DO SDH V L IS Q EQE S A E VD Y EKVGT OS KE KE 644
GO K V KMO A E VVRV ONK VI F WR GN NI RA R Q OKA PLOK V I GK FV IOV VTR V DO SDH V L IS Q EQE I W A I VD Y EKVGT OS KE KE 643
QS Q V Q S S S E D E CRV VK IL F WL GK OV RV R OKRGA E PCO AA L GR FV IOV R I SQ AD C L S K EQE T AT VP D Y EKLE R I OG SR KH EN 637
EQ K A EOE S N E YRF TTG DTG F W K N V RA GKOOT S LMK AGT VGR L P MOI TTR I D H C V IS MERE S SP LE V Y DS M I D F R EYLN 637
E EOE M KEEP OCT L CY YG R W KH GT RFQ KNT K FYO M G IE GK LI OV TRKD A AFDD VH DT SENSEORTFWE VPA MEV Y KS E SSG KH I HIOE 625
WRC CI 97 97C114 WRC CI 98 98C104 WRC GM 05 Pbt 05GM X02 MND 2 CM 98 CM16 MND 2 GA x M14 MND 2 x x 5440 DRL x x FAO MON CM 99 L1 MON NG x NG1 GSM CM 99 CM166 GSM CM 99 CM171 TAL CM 00 266 TAL CM 01 8023 MUS 1 CM 01 1085 MUS 1 CM 01 CM1239 MUS 2 CM 01 CM1246 MUS 2 CM 01 CM2500 SYK KE x KE51 SYK KE x SYK173 RCM GA x GAB1 RCM NG x NG411 GRV ET x GRI 677 SAB SN x SABIC TAN UG x TAN1 VER KE x AGM3 VER KE x AGM4 VER KE x AGM155 VER KE x TY01 patent COL CM x CG11

p51 RT end\_p15 RNase H start

p51 RT end\_p15 RNase H start

H1B.FR.83.HXB2
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZRB85
H1F1.BE.93.IV185
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101.AE.TH.90.CM240
H102.AG.NG.x.IBNC
H103.AB.RU.97.KAL153.2
H104.cpx.CY.94.CY032
H1N.CM.95.YBF30
H10.BE.87.ANTI76
H10.CM.91.MVP5180
H1P.FR.06.RBF168
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMT145
CPZ.CM.05.SIVcpzMB897
CPZ.CM.05.SIVcpzDP943
CPZ.CM.98.CAM3
CPZ.CM.98.CAM5
CPZ.GA.88.GAB1
CPZ.GA.88.GAB2
CPZ.US.85.CPZUS
CPZ.CD.90.ANT1
CPZ.TZ.00.TAN1
G207.01.2001.TAN2
CPZ.TZ.2002.TAN3.1
GOR.CM.2007.SIVgor2139.287
GOR.CM.2007.SIVgorCP2135.con
GOR.CM.2004.SIVgorCP684.con
H2A.DE.x.BEN
H2A.GW.x.ALI1
H2A.SN.x.ST
H2B.CI.x.EHO
H2B.GH.86.D205
H2C.CI.92.ABT96
H2U.FR.96.I.2034
SMM.SL.92.SL92B
SMM.US.x.H9
SMM.US.x.PGM53
SMM.US.x.SIVsmH635F.L3

MAC.US.x.239
MAC.US.x.251.IA11
MAC.US.x.251.BK28
MAC.US.x.EMBL.3
STM.US.x.STM1
MNE.US.x.MNE627
DEB.CM.99.CM40
DEB.CM.99.CM5
DEN.CD.x.CD1
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.x.Lho7
SUN.GA.98.L14
MND.1.GA.x.MNDGB1
QLC.CI.97.QTC112
WRC.CI.97.QTC114
WRC.CI.98.PbC104
WRC.GM.05.PbT05GM.X02
MND.2.CM.98.CM16
MND.2.GA.x.M14
MND.2.x.x.5440
DRL.x.x.FA0
MON.CM.99.L1
MON.NG.x.NG1
GSN.CM.99.CM666
GSN.CM.99.CM71
TAL.CM.00.266
TAL.CM.01.8023
MUS.1.CM.01.1085
MUS.1.CM.01.CM1239
MUS.1.CM.01.CM1246
MUS.2.CM.01.CM2500
SYK.KE.x.KE51
SYK.KE.x.SYK173
RCM.GA.x.GAB1
RCM.NG.x.NG411
GRV.ET.x.GRI.677
SAB.SN.x.SABIC
TAN.UG.x.TAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TY01.patent
COL.CM.x.CGUI

Table of amino acid sequences for HIV-1 P15 Nucleocapsid protein (p15 NC) from various strains. Each row corresponds to a strain (see lists on the left), and columns show the amino acid sequence. Gaps are represented by dashes. The sequence starts with 'TNTQKTELQAIYLALQDSGLVEINVITDSQYALGIITQAPDQSESELVNQIIETQLIKKIKKYLVAWPVAHKIGGGNEQVDKLVSAG' and ends with 'KLVFLDGDID.KAODEHEK.YH5NWRAMASDFNLPVVAKEIVASCDKCLKL.GEAMHGQVDCSPGIWDLCTHLEKGVILV'. The table is organized into two main sections, with the first section ending at position 797 and the second starting at position 801.

Continuation of the amino acid sequence table for HIV-1 P15 Nucleocapsid protein (p15 NC). This section covers positions 801 to 890. It continues with the same strain labels on the left and the corresponding amino acid sequences. The sequence for this section starts at position 801 and ends at position 890. The format is identical to the first section, with dashes indicating gaps in the alignment.



H1B\_FR.83.HXB2 WKGEHAVVIQDNS.DIKVPRRKAIIIRDYGMAGDDCVA...SRQDED .....\*  
 H1A1\_UG.85.U455 .....M.....G.....  
 H1B\_US.90.WEAU160 .....A.....G.....  
 H1C\_ET.86.ETH2220 .....E.....I.....V.....G.....  
 H1D\_CD.84.84ZRO85 .....N.....E.....E.....G.....  
 H1F1\_BE.93.VI850 .....E.....E.....E.....G.....  
 H1G\_SE.93.SE6165 .....E.....E.....E.....G.....  
 H1H\_CF.90.056 .....E.....E.....E.....G.....  
 H1J\_SE.93.SE7887 .....E.....E.....E.....G.....  
 H1K\_CM.96.MP535 .....E.....E.....E.....G.....  
 H101\_AE.TH.90.CM240 .....E.....E.....E.....G.....  
 H102\_AG.NG.x.IENG .....E.....E.....E.....G.....  
 H103\_AB.RU.97.KAL153.2 .....N.....N.....G.....  
 H104\_cpx.CY.94.CY032 .....N.....G.....G.....  
 H1N\_CM.95.YBF30 .....G.....G.....G.....NOEME.....  
 H10\_BE.87.ANT76 .....KG.....E.....T-SM.....G-T-SESVEOPGEIP.....  
 H1O\_CM.91.MVP5180 .....KG.....E.....T-SM.....G-T-SESVEOPGEIP.....  
 H1P\_FR.06.RBF168 .....KG.....E.....T-SM.....DI-R-SESLE.....  
 CPZ\_CM.01.SIVcpzCAM13 .....QG.E.....G.....G.....SODME.....  
 CPZ\_CM.05.SIVcpzEK505 .....G.....G.....G.....SODME.....  
 CPZ\_CM.05.SIVcpzLBT7 .....L-OE.E.....E-AN-L.....D.....N.....  
 CPZ\_CM.05.SIVcpzNB66 .....L-OE.E.....E-AN-L.....D.....N.....  
 CPZ\_CM.05.SIVcpzMT145 .....L-QD.E.....A-S.....S.....  
 CPZ\_CM.-.SIVcpzMB897 .....L-KE.....VE-N-L.....D.....S.....  
 CPZ\_CM.-.SIVcpzDP943 .....K-KE.EV.....GSM.....G.....SONLE.....  
 CPZ\_CM.98.CAM3 .....KERE.EV-I.....SM.....GG.....SOGLE.....  
 CPZ\_CM.98.CAM5 .....KE-EV.....GSM.....G.....N-QNLE.....  
 CPZ\_GA.88.GAB1 .....QG.EL.....N.....N.....  
 CPZ\_GA.88.GAB2 .....G.....KH---V.....L.....GG---NONME.....  
 CPZ\_US.85.CPZUS .....KE-E.EV.....K.....A-SM.....G.....S.....  
 CPZ\_CD.90.ANT1 .....K-OE.E.....KE-R-KIEDR-DL.....G-N.....  
 CPZ\_TZ.00.TAN1 .....KEGE.....KE.....AGGMD.....D-N-T.....  
 CPZ\_TZ.2001.TAN2 .....KEGE.....KE.....AGGMD.....D-N-T.....  
 CPZ\_TZ.2002.TAN3.1 .....KEGE.....KE---V-ANSMD.....D-N-TONME.....  
 GOR\_CM.2007.SIVgor2139.287 .....KG.....DT-G-SESLEQSS.....  
 GOR\_CM.2007.SIVgorCP2135con .....KG.....DT-G-SESLEQSS.....  
 GOR\_CM.2004.SIVgorCP684con .....KG.....DT-G-SESLEQSS.....  
 H2A\_DE.x.BEN .....IVKVG.....GROEL-SSPHLEGAREDDG-HACPCOVPEIQNK.....  
 H2A\_GW.x.ALI .....D-IVKVG.....GROEL-SGPHLEGAREG-VA.....  
 H2A\_SN.x.ST .....D-IVKVG.....II.....K.....GROEM-SGSNLEGAREG-VA.....  
 H2B\_CI.x.EH0 .....I-KVGT.E---I.....N-GGKEL-CSADVEDTMOAR-VAQSN.....  
 H2B\_GH.86.D205 .....I-KVGT.E---H-GGKEL-CSADVEDTMOAR-VAQSN.....  
 H2G\_CI.92.ABT96 .....I-KVGT.E---K-GGKEV-SSTMEDTROTG-VA.....  
 H2U\_FR.96.12034 .....L-KVGT.....I.....K-GGKEL-SGNHLEDTGEAR-VA.....  
 SMM\_SL.92.SL92B .....L-KVGT.E---K-GGKEL-SGNHLEDTGEAR-VA.....  
 SMM\_US.x.H9 .....ILKVG.....K-GGKEL-SGNHLEDTGEAR-VA.....  
 SMM\_US.x.PGM53 .....ILKVG.....K-GGKEL-SGNHLEDTGEAR-VA.....  
 SMM\_US.x.SIVsmH635F\_L3 .....ILKVG.....K-GGKEL-SGNHLEDTGEAR-VA.....

1003  
 1003  
 1004  
 1011  
 1004  
 1005  
 1001  
 1003  
 1002  
 1004  
 1009  
 1004  
 1004  
 1003  
 1015  
 1010  
 1011  
 1015  
 1010  
 1007  
 1011  
 1003  
 1003  
 1012  
 1015  
 1013  
 1009  
 1002  
 1011  
 1006  
 1004  
 1000  
 1010  
 1005  
 1003  
 1003  
 1003  
 1043  
 1033  
 1033  
 1036  
 1035  
 1016  
 1039  
 1006  
 1016  
 1016  
 1016

MAC\_US.x.239 .....ILKVG.....K-GGKEV-SSSHMEDTGEAR-VA.....  
 MAC\_US.x.251.1A11 .....ILKVG.....K-GGKEM-SSSHMEDTGEAR-VA.....  
 MAC\_US.x.251.BK28 .....ILKVG.....K-GGKEV-SSSHMEDTGEAR-VA.....  
 MAC\_US.x.EMBL\_3 .....ILKVG.....K-GGKEV-SSSHMEDTGEAR-VA.....  
 STM\_US.x.STM .....ILKVG.....K-GGKEV-SSSHLEDTGEAR-VA.....  
 MNE\_US.x.MNE027 .....ILKVG.....K-GGKEV-SSSHMEDTGEAR-VA.....  
 DEB\_CM.99.CM40 .....CKTEVG.....V-KE---D-DSKVNTONSNE.....  
 DEB\_CM.99.CM5 .....CKTEVG.....V-KE---GVDKTDTONPSE.....  
 DEN\_CD.x.CD1 .....KTOAG-FP.....KP-NAE-SRKGEONKGM-D-ET-I.....  
 LST\_CD.88.447 .....VKEGD-T-FS.....LVK---EGPKDSEGLHNN.....  
 LST\_CD.88.485 .....VKEGD-T-FS.....LVK---EGPKDSEGLDNN.....  
 LST\_CD.88.524 .....VKEGE-N-FSI.....LVK---EGPKDSQSSMDNO.....  
 LST\_KE.x.lho7 .....VKEGE-N-FS.....LVK---E-GPK-SESSLDNN.....  
 SUN\_GA.98.L14 .....KQ-GEN-L.....LVK---GESSVEM-G.....  
 MND\_1.GA.x.MNDGB1 .....LKQYE-E-I---C---K-E-SGKNSO-NLESV.....  
 OLC\_CI.97.97CI12 .....LLKEGD.KYFS-L.....VKP---VDSSANV.....  
 WRC\_CI.97.97CI14 .....L-ETPEG-IT.....L-KVWNGEGMDRSSDKN.....  
 WRC\_CI.98.98CI04 .....L-VETPEG-IT-K.....KVWNGEGMDRSSSNKD.....  
 WRC\_GM.05.Pbt.05GM.X02 .....L-V-TPEG-LTT.....KTWDGKGMDSSTTN.....  
 MND\_2\_CM.98.CM16 .....KEGT-L.....KE---TVD5-PH-EA.....  
 MND\_2\_GA.x.M14 .....KEGT-L.....K---TVD5NPHMEDROETA.....  
 MND\_2.x.x.5440 .....KEGV-L.....K---TVD5NPHMGR.....  
 DRL.x.x.FA0 .....KEGT-L-I.....K---NVD5HTNMES.....  
 MON\_CM.99.L1 .....KTEEG-LT.....KP-TTENVGDDTNOYNLRK-GLAN.....  
 MON\_NG.x.NG1 .....L-KTDQGEVIT.....P-SKEDVGSKPSAHOITREV-GMAD.....  
 RCM\_CM.99.CM166 .....V-TQAGE-IT.....KP-EAKTE-VGG-TITMND.....  
 GSN\_CM.99.CN71 .....V-TQAGE-IT.....KP-EAK-E-VGS-AHTSNDRÉ-GRMAD.....  
 TAL\_CM.00.266 .....LRTDEGEVIT.....KP-QALGNKIDLESSKQDDAEMGRDN.....  
 TAL\_CM.01.8023 .....LRTDEGEVIT.....KP-QA-GNKTDLEGSKQDDAEMGRDN.....  
 MUS\_1\_CM.01.1085 .....R-KTD-GEVIT.....KP-AKEDVGSKSDTGDLRKE-RLDN.....  
 MUS\_1\_CM.01.CM1239 .....KTEEQGVIT.....KP-AKENVGSKNTGDHRKE-GLDN.....  
 MUS\_2\_CM.01.CM1246 .....L-KTKEGE-VT.....K-TRQNV-SEPSVHVYRKE-GLAD.....  
 MUS\_2\_CM.01.CM2500 .....I-KTDQGE-IT.....K-TKENV-SDPNPVDYRKE-GLAD.....  
 SYK\_KE.x.KE51 .....TPDQVIA.....SSD-ERVD5GTHLEITSK-N.....  
 SYK\_KE.x.SYK173 .....V-TEEG-FA.....T-H-ER-D5GSH-ENDPKTD.....  
 RCM\_GA.x.GAB1 .....EETG-L.....KE-RKQVDSSEANLAGROEEN.....  
 RCM\_NG.x.NG411 .....QE-L.....C---K-RKEVRETMMEGROES.....  
 GRV\_ET.x.GRI.677 .....KEGE.....K-ERKTM-SEGSMEGVREANKQMEQSDLQDQE.....  
 SAB\_SN.x.SAB1C .....EOG.EL-TI.....K---L-SQAPLEGNV-TAGEVD.....  
 TAN\_UG.x.TAN1 .....KEGE.EL.....KE-ERKTVG-KTNMEG.....  
 VER\_DE.x.AGM3 .....K.GGVEL-EY.....K-EPRKRMG-ESNLEGAGGA-N.....  
 VER\_KE.x.9063 .....K.EGEL-I.....K-EPRKRMG-EQNMELRGP-NOMAR-NSQILD0.....  
 VER\_KE.x.AGM155 .....LK.EGEL.....K-EPRKTLG-ETHLEGAGGS-HQMAG.DS.....  
 VER\_KE.x.TY01.patent .....LK.DGS-L.....K-EPKQRVNEGDEVEGTRGS-N.....  
 COL\_CM.x.CG1 .....RNSQGLTF-K---V-T-LQYGEDV-SENLLSNGQKAEATVKGM.....

1020  
 1020  
 1016  
 1014  
 1016  
 1016  
 1008  
 1008  
 1032  
 1020  
 1020  
 1021  
 1020  
 1010  
 995  
 1023  
 1022  
 1010  
 1017  
 1012  
 1013  
 1018  
 1021  
 1016  
 1017  
 1020  
 1021  
 1012  
 1012  
 1017  
 1014  
 1037  
 1019  
 1023  
 1041  
 1037  
 1023  
 1034  
 1032  
 1026  
 991





MAC.US.x.239  
 H2A.DE.x.BEN  
 H2A.GW.x.ALI  
 H2A.SN.x.ST  
 H2B.CI.x.EH0  
 H2B.GH.86.D205  
 H2C.CI.92.ABT96  
 H2U.FR.96.12034  
 SMM.SL.92.SL92B  
 SMM.US.x.H9  
 SMM.US.x.PGMS3  
 SMM.US.x.S1VsmH635F\_L3  
 MAC.US.x.251.1A11  
 MAC.US.x.251.BK28  
 MAC.US.x.EMBL\_3  
 STM.US.x.STH  
 MNE.US.x.MNE027  
 MND-2.CM.98.CM16  
 MND-2.GA.x.M14  
 MND-2.x.x.5440  
 DRL.x.x.FA0  
 RCM.GA.x.GAB1  
 RCM.NG.x.NG411

```

Vpx start
MSD..PRERIP..PGNSGEETIGEAFELNRTVEEINREAVNHLPRELIFQVWORSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLEGEHGAGGW.RPGPPPPPPGLA*
-T...-V...-E-I-AL...-R...-A-T...-M--I-T-F-R-T-W-DM-RE-LEDO-----V-
-AN...TV...-D--AL...-R...-O-T...-M--MYT-FM--T--G--P--S-----V-
-AG...-T...-D--A...-R...-I-T...-M--M-I-S-R-T--G--P--S-----V-
...-V...-V...-E-T-L-HL-V...-K-A-RE...-I-T...-M--M-I-FA--G-R--P--S-----
...-V...-D--V...-A-E-T-L...-A-RE...-I-T...-L-M--M-V-YT...-Q--P--S-----
...-X...-E-D--V...-E-XT...-I...-R...-A-T...-L-M--M-V...-T-QK--P--Q-----
-G...-E--V...-S-E--RD--A-Q...-Q-R...-E-T...-VQ-M--M-Q-FR--T-R--S-----I-
-T...-X...-HN--AL-OT-Q...-R-C--V--Y--A--VQ-M--M-Q-FR--T-R--S-----T-----
-XX...-X...-XX-X-D-H-X...-A--X...-R-X...-M--V--T-X...-R...-E...-S-----
...-E...-G--A...-M--E-T...-M...-T-----
...-A-E...-H--D...-R...-PGNX...-S-R-----
...-AE...GA-EI-EGA...VDLNT--E-SL-K--Q--RL-FHP-FL-RL-NACI-H--RHOR-L-A---L-MN-M-T-MQE-P-RSG-P-----MV-
-AE...A-EA-QGA...VGLQ--E-SL-QV--QL-FHP-FL-RL-NTCV-H--RLRRTL-A---L-MH-M-V-MQO-P-RSG-S-----M-
-AE...A-EA-EGA...VGLQ--ETS-L-R--RL-FHP-FL-RL-NTCV-H--RHOR-LD-A---L-MH-MY-MQO-P-RNG-RPR-----M-
-AERQSV-A-AE-MGA...V-LE--Q-SLLR-Q--RL-FHP-FL-RL-NTCV-HY--ALQL-FT-S---L-L--M--FQO-S-QGR-PPPLRPAGDRL---PP-
-AE...A...EVPT-AGEAEFQWLRDMLKVNLEA.RL-FHP-F-RL-RTCV-H--VHOR-LE-AA---L-M---I-OT-SQR...-PNPR-AV-ERITIL-M-
-AE...G--V...EAPT-AGDVEF-PWLHRMLT-VNLEA.RL-FHP-F-RL-RTCV-H--RL-R-LE-AG---L-M---I--QS--SQR...-Q-QAREA-ERIQIL-M-

```

112  
 114  
 113  
 113  
 112  
 112  
 113  
 113  
 113  
 113  
 113  
 66  
 113  
 113  
 100  
 100  
 100  
 114  
 105  
 109

	Vpr start	oligomerization	amphipathic $\alpha$ -helix	Vpr end in HXB2	frameshift in HXB2	H(S/N)RIG motifs	Vpr end	
H1B_FR.83.HXB2	MEQA	PEDQGPQREPHNEWTLELLEELKNEAVRHFPRITWLHGLGHQIHYETG	.....	TWAGVEAIRIRLQQLLFIH#FRIGCRHSRIGVTR	.....	QRRAR_NGASRS	.....	96
H1A1_UG.85.U455		.....	.....	.....	.....	.....	.....	97
H1B_US.90.WEAU160		.....	.....	.....	.....	.....	.....	96
H1C_ET.86.ETH2220		.....	.....	.....	.....	.....	.....	97
H1D_CD.84.84ZRB85		.....	.....	.....	.....	.....	.....	97
H1F1_BE.93.VT89		.....	.....	.....	.....	.....	.....	97
H1G_SE.93.SE6165		.....	.....	.....	.....	.....	.....	97
H1H_CF.90.056		.....	.....	.....	.....	.....	.....	97
H1J_SE.93.SE7887		.....	.....	.....	.....	.....	.....	97
H1K_CM.96.MP535		.....	.....	.....	.....	.....	.....	97
H101_AE_TH.90.CH240		.....	.....	.....	.....	.....	.....	97
H102_AG_NG.X.IBNG	-R	.....	.....	.....	.....	.....	.....	97
H103_AB_RU.97.KAL153.2		.....	.....	.....	.....	.....	.....	97
H104_cpx.CY.94.CY032		.....	.....	.....	.....	.....	.....	97
H1N_CM.95.YBF30	-R	.....	.....	.....	.....	.....	.....	96
H10_BE.87.ANT76		.....	.....	.....	.....	.....	.....	96
H10_CM.91.MP5180		.....	.....	.....	.....	.....	.....	98
H1P_FR.06.RBF168		.....	.....	.....	.....	.....	.....	101
CPZ_CM.01.SIVcpzCAM13	-H	.....	.....	.....	.....	.....	.....	99
CPZ_CM.05.SIVcpzEK505		.....	.....	.....	.....	.....	.....	97
CPZ_CM.05.SIVcpzL87	-V	.....	.....	.....	.....	.....	.....	96
CPZ_CM.05.SIVcpzNB66		.....	.....	.....	.....	.....	.....	97
CPZ_CM.05.SIVcpzMT145		.....	.....	.....	.....	.....	.....	97
CPZ_CM.SIVcpzMB897		.....	.....	.....	.....	.....	.....	97
CPZ_CM.SIVcpzDP943		.....	.....	.....	.....	.....	.....	96
CPZ_TZ.2001.TAN2		.....	.....	.....	.....	.....	.....	84
CPZ_TZ.2002.TAN3.1		.....	.....	.....	.....	.....	.....	84
CPZ_TZ.00.TAN		.....	.....	.....	.....	.....	.....	84
CPZ_CM.98.CAM3		.....	.....	.....	.....	.....	.....	96
CPZ_CM.98.CAM5	-I	.....	.....	.....	.....	.....	.....	96
CPZ_GA.88.GAB1		.....	.....	.....	.....	.....	.....	97
CPZ_GA.88.GAB2		.....	.....	.....	.....	.....	.....	97
CPZ_US.85.CPIJUS	-V	.....	.....	.....	.....	.....	.....	101
CPZ_CD.90.ANT		.....	.....	.....	.....	.....	.....	89
GOR_CM.2007.SIVgor2139.287	-I	.....	.....	.....	.....	.....	.....	99
GOR_CM.2007.SIVgorCP2135.con		.....	.....	.....	.....	.....	.....	99
GOR_CM.2004.SIVgorCP684.con	-I	.....	.....	.....	.....	.....	.....	99
H2A_DE.X.BEM	-TE	PTEFP	.....	.....	.....	.....	.....	105
H2A_GW.X.ALI1	-TE	PTEFP	.....	.....	.....	.....	.....	106
H2A_SN.X.ST	-TE	PTESP	.....	.....	.....	.....	.....	106
H2B_CI.X.EHO	-AE	VPETP	.....	.....	.....	.....	.....	102
H2B_GH.86.D205	-AE	APETP	.....	.....	.....	.....	.....	102
H2C_HI.92.ABT96	-AE	PEP	.....	.....	.....	.....	.....	103
H2U_FR.96.2034	-AE	PEP	.....	.....	.....	.....	.....	98
SMM_US.92.SL92B	-TE	EHA	.....	.....	.....	.....	.....	101
SMM_US.X.H9	-TE	TERP	.....	.....	.....	.....	.....	102
SMM_US.X.PGM53	-TE	TERP	.....	.....	.....	.....	.....	102
SMM_US.X.SIVsmH635F_L3	-TE	AERP	.....	.....	.....	.....	.....	102
MAC_US.X.239	.....	EERP	.....	.....	.....	.....	.....	102
MAC_US.X.251.IA11	.....	EERP	.....	.....	.....	.....	.....	101
MAC_US.X.251.BK28	.....	EERP	.....	.....	.....	.....	.....	98
MAC_US.X.EMBL_3	.....	.....	.....	.....	.....	.....	.....	98
STM_US.X.ST14	.....	THRP	.....	.....	.....	.....	.....	34
MNE_US.X.MNE827	.....	EERP	.....	.....	.....	.....	.....	102
DEN_CD.X.CD1	-A	.....	.....	.....	.....	.....	.....	99
LST_KE.X.lh07	-SR	.....	.....	.....	.....	.....	.....	117
SUN_GA.98.L14	-AS	.....	.....	.....	.....	.....	.....	117
MND_1.GA.X.MNDGB1	.....	.....	.....	.....	.....	.....	.....	105
LST_CD.88.447	-SR	.....	.....	.....	.....	.....	.....	115
LST_CD.88.485	-SR	.....	.....	.....	.....	.....	.....	115
DEB_CM.99.CM40	-R	.....	.....	.....	.....	.....	.....	118
DEB_CM.99.CM5	-R	.....	.....	.....	.....	.....	.....	118
LST_CD.88.524	-SR	.....	.....	.....	.....	.....	.....	115
OLC_CI.97.97C112	-AE	RVP	.....	.....	.....	.....	.....	97
WRC_CI.97.97C114	-OE	CLP	.....	.....	.....	.....	.....	133
WRC_CI.98.98C104	-OE	CLP	.....	.....	.....	.....	.....	133
WRC_GM.05.Pbt.05GM.X02	I+E	CLP	.....	.....	.....	.....	.....	133
MND_2.GA.X.M14	.....	OPP	.....	.....	.....	.....	.....	102
MND_2.X.X.5440	.....	OPP	.....	.....	.....	.....	.....	102
DRL_X.X.FA0	.....	RPP	.....	.....	.....	.....	.....	101
MON_CM.99.L1	-V	.....	.....	.....	.....	.....	.....	121
MON_NG.X.NG1	.....	.....	.....	.....	.....	.....	.....	155
GSN_CM.99.CN71	-R	.....	.....	.....	.....	.....	.....	136
TAL_CM.00.266	-R	.....	.....	.....	.....	.....	.....	116
TAL_CM.01.8023	-R	.....	.....	.....	.....	.....	.....	115
MUS_1_CM.01.1085	-R	.....	.....	.....	.....	.....	.....	136
MUS_1_CM.01.CM1239	-R	.....	.....	.....	.....	.....	.....	136
MUS_2_CM.01.CM1246	-R	.....	.....	.....	.....	.....	.....	136
MUS_2_CM.01.CM2500	-R	.....	.....	.....	.....	.....	.....	136
RCM_GA.X.GAB1	.....	.....	.....	.....	.....	.....	.....	101
RCM_NG.X.NG411	.....	.....	.....	.....	.....	.....	.....	116
SVK_KE.X.KE51	-A	.....	.....	.....	.....	.....	.....	101
SVK_KE.X.SYK173	-AE	.....	.....	.....	.....	.....	.....	114
GRV_ET.X.GRI 677	-AS	.....	.....	.....	.....	.....	.....	119
SAB_SN.X.SAB1	-AS	.....	.....	.....	.....	.....	.....	141
TAN_UG.X.TAN1	-AE	.....	.....	.....	.....	.....	.....	120
VER_DE.X.AGM3	-AS	.....	.....	.....	.....	.....	.....	120
VER_KE.X.9063	-AS	.....	.....	.....	.....	.....	.....	120
VER_KE.X.AGM155	-AS	.....	.....	.....	.....	.....	.....	120
VER_KE.X.TY01_patent	-AS	.....	.....	.....	.....	.....	.....	120
COL_CM.X.CGU1	-SS	.....	.....	.....	.....	.....	.....	93



H1B\_FR.83.HXB2
H1A1\_UG.85.U455
H1B\_US.90.WEAU160
H1C\_ET.86.ETH2220
H1D\_CD.84.84Z885
H1F1\_BE.93.VT919
H1G\_SE.93.S6E165
H1H\_CF.90.056
H1J\_SE.93.SE7887
H1K\_UM.96.MP535
H101\_AE.TH.90.CM240
H102\_AG.NG.X.IE8G
H103\_AB.RU.97.KAL153.2
H104\_CPX.CY.94.CY032
H1N\_CM.95.YBF30
H1O\_BE.87.ANT76
H1P\_FR.06.RBF168
GOR\_CM.2004.SIVgorCP684con
GOR\_CM.2007.SIVgor2139.287
GOR\_CM.2007.SIVgorCP2135con
CPZ\_XD.90.ANT
CPZ\_CM.SIVcpzDP943
CPZ\_CM.SIVcpzMB897
CPZ\_CM.01.SIVcpzCM13
CPZ\_CM.05.SIVcpzEK595
CPZ\_CM.05.SIVcpzLB7
CPZ\_CM.05.SIVcpzNB66
CPZ\_CM.05.SIVcpzMT145
CPZ\_CM.98.CAM3
CPZ\_CM.98.CAM5
CPZ\_GA.88.GAB1
CPZ\_GA.98.GAB2
CPZ\_TZ.00.TAN1
CPZ\_TZ.2001.TAN2
CPZ\_TZ.2002.TAN3.1
CPZ\_US.85.CPZUS
H2A\_IE.X.BEN
H2A\_GW.X.ALI
H2A\_SN.X.ST
H2B\_CI.X.EHO
H2B\_GH.86.D205
H2G\_CI.92.AB896
H2U\_FR.96.12934
MNE\_US.X.MNE827
SMM\_SL.92.SL92B
SMM\_US.X.H9
SMM\_US.X.PGM53
SMM\_US.X.SIVsmh635F\_L3
STM\_US.X.STM

Table with columns: Tat start, disulfide bonding, NLS, exon 1 end, exon 2 start, Tat end. Rows include protein sequences for various HIV strains and domains like ME, TPLKPEGLSMPYNEP, etc.

MAC\_US.X.239
MAC\_US.X.251.IA11
MAC\_US.X.251.BK28
MAC\_US.X.EMBL.3
DEB\_CM.99.CM40
DEB\_CM.99.CM5
DEN\_CD.X.CD1
DRL.X.X.FAO
GRV\_ET.X.GRI.677
GSN\_CM.99.CM166
GSN\_CM.99.CM71
COL\_CM.X.CG1
LST\_CD.88.447
LST\_CD.88.585
LST\_CD.88.524
LST\_KE.X.lho7
MND.1.GA.X.MNDG1
MND.2.CM.98.CM16
MND.2.GA.X.M14
MND.2.X.X.5440
MON\_CM.99.L1
MON\_NG.X.NG1
MUS.1.CM.01.1085
MUS.2.CM.01.CM1239
MUS.2.CM.01.CM1246
MUS.2.CM.01.CM2500
OLC\_CI.97.97C12
RCM\_GA.X.GAB1
RCM\_NG.X.NG411
SAB\_SN.X.SAB1C
SUN\_GA.98.L14
SYK\_KE.X.KE51
SYK\_KE.X.SYK173
TAL\_CM.00.266
TAL\_CM.01.8923
TAN\_UG.X.TAN1
VER\_DE.X.AGM3
VER\_KE.X.9063
VER\_KE.X.AGM155
VER\_KE.X.9063.patent
WRC\_CI.97.97C14
WRC\_CI.97.97C14
WRC\_GM.05.Pbt.05GM.X02

Table with columns: Tat start, disulfide bonding, NLS, exon 1 end, exon 2 start, Tat end. Rows include protein sequences for various HIV strains and domains like MAC\_US.X.239, MAC\_US.X.251.IA11, etc.

H1B.FR.83.HXB2	.....*	100
H1A1.UG.85.U455	.....*	102
H1B.US.90.WEAU160	.....*	102
H1C.ET.86.ETH2220	.....*	101
H1D.CD.84.84ZRO85	.....*	87
H1F1.BE.93.H1F50	.....*	102
H1G.SE.93.SE6165	.....*	102
H1H.CF.90.056	.....*	100
H1J.SE.93.SE7887	.....*	102
H1K.CM.96.MP535	.....*	102
H101.AE.TH.90.CM240	.....*	102
H102.AG.NG.x.IENG	.....*	101
H103.AB.RU.97.KAL153_2	.....*	87
H104.cpx.CY.94.CY032	.....*	103
H1N.CM.95.YBF30	.....*	102
H10.BE.87.ANT760	.....*	115
H10.CM.91.MVP5180	.....*	116
H1P.FR.06.RBF168	.....*	72
GOR.CM.2004.SIVgorCP684con	.....*	83
GOR.CM.2007.SIVgor2139_287	.....*	83
GOR.CM.2007.SIVgorCP2135con	.....*	83
CPZ.CD.90.ANT	.....*	74
CPZ.CM.x.SIVcpzDP943	.....*	103
CPZ.CM.x.SIVcpzMB897	.....*	101
CPZ.CM.01.SIVcpzCAM13	.....*	100
CPZ.CM.05.SIVcpzEK505	.....*	94
CPZ.CM.05.SIVcpzLB7	.....*	100
CPZ.CM.05.SIVcpzNB66	.....*	102
CPZ.CM.05.SIVcpzMT145	.....*	102
CPZ.CM.98.CAM3	.....*	103
CPZ.CM.98.CAM5	.....*	108
CPZ.GA.88.GAB1	.....*	101
CPZ.GA.88.GAB2	.....*	100
CPZ.TZ.00.TAN1	GYTRPFKTS <del>SGSGSACKH</del>	137
CPZ.TZ.2001.TAN2	GSTRPVETSSGSGRSCKH	137
CPZ.TZ.2002.TAN3_1	GSTRPVKTS <del>SGSGQSKR</del>	137
CPZ.US.85.CPZUS	.....*	103
H2A.DE.x.BEN	.....*	131
H2A.GW.x.ALI	.....*	137
H2A.SN.x.ST	.....*	131
H2B.CI.x.EHO	.....*	139
H2B.GH.86.D205	.....*	129
H2G.CI.92.ABT96	.....*	113
H2U.FR.96.12034	.....*	112
MNE.US.x.MNE027	.....*	131
SMM.SL.92.SL92B	.....*	127
SMM.US.x.H9	.....*	129
SMM.US.x.PGM53	.....*	131
SMM.US.x.SIVsmH635F_L3	.....*	129
STM.US.x.STM	.....*	129
MAC.US.x.239	.....*	131
MAC.US.x.251_1A11	.....*	131
MAC.US.x.251_BK28	.....*	131
MAC.US.x.EMBL_3	.....*	117
DEB.CM.99.CM40	.....*	116
DEB.CM.99.CM5	.....*	102
DEN.CD.x.CD1	.....*	100
DRL.x.x.FA0	.....*	125
GRV.ET.x.GRI_677	.....*	95
GSN.CM.99.CN166	.....*	93
GSN.CM.99.CN71	.....*	120
COL.CM.x.CGU1	.....*	80
LST.CD.88.447	.....*	101
LST.CD.88.485	.....*	101
LST.CD.88.524	.....*	112
LST.KE.x.lho7	.....*	112
MND-1.GA.x.MNDGB1	LIEDLTSFARE.....*	147
MND-2.CM.98.CM16	.....*	123
MND-2.GA.x.M14	.....*	88
MND-2.x.x.5440	.....*	111
MON.CM.99.L1	.....*	105
MON.NG.x.NG1	.....*	127
MUS-1.CM.01.1085	.....*	99
MUS-1.CM.01.CM1239	.....*	102
MUS-2.CM.01.CM1246	.....*	109
MUS-2.CM.01.CM2500	.....*	102
OLC.CI.97.97CI12	.....*	108
RCM.GA.x.GAB1	.....*	106
RCM.NG.x.NG411	.....*	116
SAB.SN.x.SAB1C	.....*	74
SUN.GA.98.L14	.....*	103
SYK.KE.x.KE51	.....*	99
SYK.KE.x.SYK173	.....*	111
TAL.CM.00.266	.....*	111
TAL.CM.01.8023	TVTRP.....*	128
TAN.UG.x.TAN1	.....*	102
VER.DE.x.AGM3	.....*	120
VER.KE.x.9063	.....*	120
VER.KE.x.AGM155	.....*	120
VER.KE.x.TY01_patent	.....*	101
WRC.CI.97.97CI14	.....*	126
WRC.CI.98.98CI04	.....*	133
WRC.GM.05.Pbt_05GM_X02	.....*	133

	Rev start	exon 1 end	exon 2 start	NLS	Leu-rich effector domain	Rev end	
H1B_FR_83_HXB2	MAGRS	GDSEDEILTRIVTLKLLQYS	NPPNP	PEGTRQARRRNRREORORIHSTSERLITGLYLGR	SAEVPVLPQLPLERLTDLCNECDGTSGTG	GVGSPQLL VESPTVLESGTKE	
H1A1_UG_85_UJ455	R	NP-D-LKA-I	C	R-S-K-A	D-L-L-SDC	P	I-R-S-S
H1B_US_90_WEAU160	D-LK-I	I	I	Y-T	A-Y	G-L	SNF
H1C_ET_86_ETH2220	LKA-I	I	I	Y-T	A-Y	G-L	SNF
H1D_CD_84_84ZR085	D-LTA-K	I	I	Y-S	A-Y	G-L	SNF
H1F1_BE_93_M1295	T-LKA-KC	I	I	Y-K	A-Y	G-L	SNF
H1G_SE_93_SE6165	ST-A-KA-I	I	I	Y-P	A-Y	G-L	SNF
H1H_CF_90_056	A-T-LQVCKI	I	I	C-E-T	A-RE	TSC	PP
H1J_SE_93_SE7887	DO-LLA	I	I	Y-K-N-S	A-N-D	PSSC	P
H1K_CM_96_MP535	R-P-Q-LT-T	I-E	I	Y-S	A-N-D	P-SAC	P
H1L_AE_TH_90_CM240	ST-L-A-I	I	I	F-S-S-T	A-S	S-C	T
H1M2_AG_NG_x_IBNG	A-L-A-I	I	I	Y-P	A-S	S-C	T
H1O3_AB_RU_97_KAL153_2	LK-I	I	I	F	H	Q	S
H104_cpx_CY_94_CY032	NI-D-FKAA	A-I	I	Y-N-T	A-K-L	L-A	A
H1N_CM_95_YBF30	VN-L-A-V	I	I	Y-SK	A	R-L	A
H1O_BE_87_ANT176	E-DO-LQAQI	I	I	Q-S-R-S-N-K	R-A	VD-LAA-V	VVHG
H1P_US_95_CP25180	E-DO-LQAQI	I	I	C-T-A-N-K	R-A	VD-LAT-A	VVHG
H1P_FR_06_RBF168	EDDLR-T-I	I-I	I	W-E-RGSS-NST-K	S	DO-AG	ASR
CPZ_CM_01_SIVcpzCAM13	EHD-AR-LQA-KI	I-E	I	Y-S	A	RE-N	AAH
CPZ_CM_05_SIVcpzEK505	VN-L-AI-V	I	I	Y-DS	K	A	RAL
CPZ_CM_05_SIVcpzLB7	R-L-A-I	I	I	F-H	S	A	H
CPZ_CM_05_SIVcpzMB66	G-N-L-A-I	I	I	Y-S	S	A	H
CPZ_CM_05_SIVcpzMT145	E-G-N-L-A-I	I	I	Y-QS-R	K	A	N-L
CPZ_CM_SIVcpzMB897	R-L-A-I	I	I	F-S	P	K	SAL
CPZ_CM_SIVcpzDP943	E-A-TR-LOV-I	I-D	I	Y-D-NK	A	H	L
CPZ_TZ_2001_TAN2	EE-AN-LO-I	I-D	I	Y-S-AA-S-T	Q	H	VDALAN
CPZ_TZ_2002_TAN3_1	EE-AN-LO-I	I-D	I	Y-S-GA-S-A	R	L	VD-AN
CPZ_CM_98_CAM1	EE-AN-LO-I	I-D	I	Y-S-GA-S-T	Q	H	VDALAN
CPZ_CM_98_CAM3	E-D-AC-LQA-I	I-V-S	I	Y-D-NK	A	L	S
CPZ_CM_98_CAM5	E-E-TR-LQA-I	I-D	I	Y-D-NK	A	E	GO-V
CPZ_GA_88_GAB1	EPDAR-LQA-KI	I	I	Y-S	K	A	K
CPZ_GA_88_GAB2	E-R-L-A-KI	I-D	I	Y-S-NR-S	A	CE	Q
CPZ_GA_88_CP2US	G-A-L-A-I	I-A	I	F-NST	KN	NE	Q
CPZ_CD_90_ANT	EELGEG-LKA-KI	I-R	I	Y-K-A-S-A	K	KK-D	VEGLAA
GOR_CM_2007_SIVgor2139_287	DEDLO-LTKI	I-R	I	W-E-GPS-NST	KG	DO-AG	VSH-LG
GOR_CM_2007_SIVgor2135con	DEDLO-LTKI	I-R	I	C-E-GPS-NST	KG	DO-AG	AAH
GOR_CM_2004_SIVgorCP684con	DEDLO-LTKI	I-R	I	W-E-GPS-NST	KG	DO-AG	ASH
H2A_DE_x_BEN	SE-A-DE-G-QGLK	LR-H-T	I	Y-Q-GP-ASO	R	QWLRLVALANLCAVDPDP	TDQ
H2A_GW_x_ALI	TE-A-GE-D-KL	R-H-T	I	Y-Q-GP-ANQ	Q	WG-V	VALAD
H2A_SN_x_ST	NE-A-EE-R-KL	R-H-T	I	Y-Q-GP-ASO	Q	L	VALAD
H2B_CI_x_EHO	NA-EDR-QKEL	LH-H-T	I	Y-Q-GP-ASO	K	Q	LALAD
H2B_GH_86_D205	TA-EDG-QKEL	LH-H-T	I	Y-Q-GP-ASO	K	R	LALAD
H2C_CI_92_ABT96	SNL-EE-RKR	HF-H-T	I	Y-Q-GP-KANQ	K	K	WL-LALAD
H2I_FR_96_I2834	SNP-E-GW	T-GP-ANQ	I	Y-Q-GP-ANQ	Q	H	LALAD
SMM_SL_92_S192B	ONP-E-R-RL	H-H-T	I	Y-A-GP-ASO	KR	WI	VALAD
SMM_US_x_PGM53	SSS-AGE-R-RL	HF-H-T	I	Y-Q-GP-ANQ	R	Q	WL-LALAD
SMM_US_x_SIVsmH635F_L3	SST-E-RKR	LF-H-T	I	Y-Q-GP-ANQ	R	Q	WL-LALAD

	Rev start	exon 1 end	exon 2 start	Rev end
MAC_US_x_239	SNHE	RE-RKRL	H-H-T	I
MAC_US_x_251_I A11	SSHE	RE-RKRL	H-H-T	I
MAC_US_x_251_BK28	SSHE	RE-RKRL	H-H-T	I
MAC_US_x_EMBL_3	SSHE	RE-RKRL	H-H-T	I
STM_US_x_STH	SDD	E-RKRL	QF-H-T	I
WME_SSHIA	SNA	E-RKRL	QF-H-T	I
RCM_GA_x_GAB1	LPGODE	D-RKKI	NF-LI	I
RCM_NG_x_NG411	LLGEEEE	A-Q-IR	RI-H-I	I
SAB_SN_x_SAB1C	SLGQ	L-RF	F-TT	I
GRV_ET_x_GRI_677	SLGK	D-R-ITQI	RT-H	I
VER_DE_x_AGM3	PLG	D-LL	AF-NKN	I
VER_KE_x_9063	PLGP	R-KVFS	AY-NRT	I
VER_KE_x_AGM155	PLGP	R-LL	AF-R	I
VER_KE_x_TY01_patent	PLGP	R-LL	W-ST	I
DEB_CM_99_C040	HAGGR	SAE-NTRQLK	V-S-IKILYOS	I
DEB_CM_99_L1	HAGGR	A-QA	L-M-RIKILYOS	I
MON_CM_99_L5	NGR	RRLLSALA	AVRILQESRRILSSFFSD	I
MON_NG_x_NG1	AERG	AAVDIILRLA	GAIRITRESRLLSPLSD	I
GSN_CM_99_CN166	DHARGN	QK-QNLLIACR	IKTLHRSRLTSLPAD	I
GSN_CM_99_CN71	DPANGR	QK-QNLLIACR	IKTLHRSRLTSLPAD	I
MUS_1_CM_01_1885	A-E	NQVILL	ACRATIQYKSRILSSPSSD	I
MUS_1_CM_01_CM1239	A-E	NQVILL	ICRIRILYQSRLSFPST	I
MUS_2_CM_01_CM1246	A-G	NQVILL	FCRIRILYQSRLSPLSD	I
MUS_2_CM_01_CM2500	TGR	DIEELKACQ	IKIRILYQSRLSFPSSST	I
TAL_CM_90_266	SLTGES	GRGEDP	LALAA-RRIRILYQSRLTSSV	I
TAL_CM_90_8023	SLTGES	GRGEDP	LALAA-RRIRILYQSRLTSSV	I
DEN_CD_x_CD1	ERDO	EL-KAVR	IKKLYOS	I
SYK_KE_x_KE51	S-ER	EDTQELL	TLLRIA-OLEAA	I
SYK_KE_x_SVK173	P	QGGSEALV	LRMIAHLO	I
COL_CM_x_C0U1	TNAGVRP	VAFVYSYAFQK	GWSA-IRIRIVLRHKH	I
LST_CD_98_447	STGG	E-K-SP	YLKIS-T-WGLO	I
LST_CD_98_085	LST	G	P-YLKIS-T-WGLO	I
LST_CD_98_524	STGP	EG	PTYL-SRI-WGLO	I
LST_KE_x_lho7	STGN	GD	P-YL-SRI-WGLO	I
SUN_GA_98_L14	STG	DSINOYL	IS-R-EGLA	I
MND_1_GA_x_HNDG81	STG	WYQ	LIR-YLWV-K-EGSLCTFTG	I
OLC_CI_97_97C12	CLC	SHFO		I
WRC_CI_97_97C14	#VLVLLVCIAVRFAFCK	G-GFLIVL	TARKLSQI00G-L-GLL	I
WRC_CI_98_98C104	#VDVINVLAVSFASCK	G-GFLIVPTAR	L-VQOK-Q-GLL	I
WRC_GM_05_Pbt_05GM_X02	#ADVKNVSHVNYVCK	G-GSLLIHTAR	L-ISO-T-Q-GLL	I
MND_2_CM_98_CN16	MPTETGDRDPRDFFKYYKQI	VRKWEGL	PSSLCLQGRSLS	I
MND_2_GA_x_444	NSTEPGL	RDFW-KYRIRVKI	WEGL-PSSLCLQGRSLS	I
MND_2_x_x_4440	MSTD	GOLDPEFMRKYQAI	IKVQWLEGI-PSSLCLQGRSLS	I
DRL_x_x_FAO	MS	PEREPFMRKYQAI	VTRVLWQGL-PSSLYL*GRITPRDD	I

116  
124  
117  
108  
117  
117  
108  
117  
117  
117  
124  
124  
124  
117  
118  
105  
103  
104  
105  
124  
104  
124  
124  
124  
123  
123  
125  
106  
106  
106  
106  
125  
108  
125  
124  
106  
119  
105  
105  
105  
104  
108  
108  
151  
154  
91  
99  
100  
105  
101

.....\*

H1B\_FR.83.HXB2  
H1A1\_UG.85.U455  
H1B\_US.90.WEAU160  
H1C\_ET.86.ETH2220  
H1D\_CD.84.84ZRO85  
H1F1\_BE.93.VI850  
H1G\_SE.93.SE6165  
H1H\_CF.90.056  
H1J\_SE.93.SE7887  
H1K\_CM.96.MP535  
H101\_AE.TH.90.CM240  
H102\_AG.NG.x.IENG  
H103\_AB.RU.97.KAL153\_2  
H104\_cpx.CY.94.CY032  
H1N\_CM.95.YBF30  
H10\_BE.87.ANT76  
H10\_CM.91.MVP5180  
H1P\_FR.06.RBF168  
CPZ\_CM.01.SIVcpzCAM13  
CPZ\_CM.05.SIVcpzEK505  
CPZ\_CM.05.SIVcpzLB7  
CPZ\_CM.05.SIVcpzMB66  
CPZ\_CM.05.SIVcpzMT145  
CPZ\_CM.-.SIVcpzMB897  
CPZ\_CM.-.SIVcpzDP943  
CPZ\_TZ.2001.TAN2  
CPZ\_TZ.2002.TAN3\_1  
CPZ\_TZ.00.TAN1  
CPZ\_CM.98.CAM3  
CPZ\_CM.98.CAM5  
CPZ\_GA.88.GAB1  
CPZ\_GA.88.GAB2  
CPZ\_US.85.CPZUS  
CPZ\_CD.90.ANT  
GOR\_CM.2007.SIVgor2139\_287  
GOR\_CM.2007.SIVgorCP2135con  
GOR\_CM.2004.SIVgorCP684con  
H2A\_DE.x.BEN  
H2A\_GW.x.ALI  
H2A\_SN.x.ST  
H2B\_CI.x.EHO  
H2B\_GH.86.D205  
H2G\_CI.92.ABT96  
H2U\_FR.96.12034  
SMM\_SL.92.SL92B  
SMM\_US.x.PGM53  
SMM\_US.x.SIVsmH635F\_L3

.....

MAC\_US.x.239  
MAC\_US.x.251\_1A11  
MAC\_US.x.251\_BK28  
MAC\_US.x.EMBL\_3  
STM\_US.x.STM  
MNE\_US.x.MNE027  
RCM\_GA.x.GAB1  
RCM\_NG.x.NG411  
SAB\_SN.x.SAB1C  
GRV\_ET.x.GRI\_677  
TAN\_UG.x.TAN1  
VER\_DE.x.AGM3  
VER\_KE.x.9063  
VER\_KE.x.AGM155  
VER\_KE.x.TY01\_patent  
DEB\_CM.99.CM40  
DEB\_CM.99.CM5  
MON\_CM.99.L1  
MON\_NG.x.NG1  
GSN\_CM.99.CN166  
GSN\_CM.99.CN71  
MUS\_1\_CM.01.1085  
MUS\_1\_CM.01.CM1239  
MUS\_2\_CM.01.CM1246  
MUS\_2\_CM.01.CM2500  
TAL\_CM.00.266  
TAL\_CM.01.8023  
DEN\_CD.x.CD1  
SYK\_KE.x.KE51  
SYK\_KE.x.SYK173  
COL\_CM.x.CGU1  
LST\_CD.88.447  
LST\_CD.88.485  
LST\_CD.88.524  
LST\_KE.x.lho7  
SUN\_GA.98.L14  
MND\_1\_GA.x.MNDGB1  
OLC\_CI.97.97C112  
WRC\_CI.97.97C114  
WRC\_CI.98.98CI04  
WRC\_GM.05.Pbt\_05GM\_X02  
MND\_2\_CM.98.CM16  
MND\_2\_GA.x.M14  
MND\_2\_x.x.5440  
DRL\_x.x.FA0

.....

VGSPQESGRDHCNTQEDOTRG  
MGGSQESGESNHRDPOENQTRT

.....

RLSGSHQSQSKSAARP

.....

DS  
EOKRHTSSCKKPTGTKEEMETIQRKWWSQ  
EOKRHTSSCERHOTGTREKMEIQIREKWSQ  
EORRAIKN  
EORTVIEH  
QQGESIIVSEGKK

.....

SSKQ

.....

KOKQOHGNYCOTLHKP

.....

	transmembrane domain	phos cytoplasmic domain		Vpu end
	Vpu start	$\alpha$ -helix	$\alpha$ -helix	
H1B.FR.83.HXB2	TPP.....IPVIVAIVALVVAIIIVVWSIVITVEY.....	.....RKILRQRKIDRLIDRLTERAEDSGNESEGEISALVEMGVEMGHAPWDV.DDL.....*		82
H1A1.UG.85.U455	MT.....LE-W-TG-I-L-L-----T-G-##.....	.....K-L-K-K-----LN-IR-----	.....D-DTEE-SLL-----NYDLGVD.NN.....	81
H1B.US.90.WEAU160	M-S.....LQ-L-----G-----L-----	.....LQ-L-----G-----L-----	.....DQEE-SAL-----I-----	82
H1C.ET.86.ETH2220	MVOLLAKVYRIV-----FT-L-----T-AY-----	.....L-----R-----K-TR-----	.....D-DTEE-ST--D-NLRL--N-----	87
H1D.CD.84.84ZRB85	M-S.....LQ-L-----L-L-----T-F-----	.....R-K-----W-----IR-E-----	.....DKEE-STL-----D-----	82
H1F1.BE.93.VT850	MSY.....LLAIG-A-E-L-F-----T-Y-----	.....K-LV-----NK-YK-IR-----	.....DAEE-AAL-G-PFI-G-I-NN-----	82
H1G.SE.93.SE6165	M-S.....LV-L--G-I-F-A-----T-F-----	.....E-RK-KR-GK-L-IR-----	.....D-DTEE-TL-----DFD--VG--N-----	82
H1H.CF.90.056	MYI.....LGLGIGALV-TF-A-VI-T-Y-----	.....K-LV-K-----E-IG-----	.....D-DTEE-SKL-M-----LNLGY--A-----	81
H1J.SE.93.SE7887	MI.....LQ-A-----FI--FL-GM-T-Y-----	.....K-L-----K-----IR-----	.....D-DTEE-ADL--R-P-DL-N--N-----	82
H1K.CM.96.MP535	MVS.....LA-SIVALV-AL-LA-I-T-Y-----	.....LVK-KR-NW-----IR-----	.....DAEE-ADI-G-L-LILGNI--N-----	81
H101.AE.TH.90.CH240	MT.....LE-S--G-I-L-L-----T-A-V-----	.....K-----VK-IR-----	.....D-TDF-AKL-----DFD--VG--N-----	81
H102.AG.NG.x.IENG	W.....LT-T--G---F-A-----Y-----	.....RK-K-----L-IR-----	.....D-DTEE-STL-M--YEYIL-N--N-----	82
H103.AB.RU.97.KAL153_2	M.Q.....SLA-A-----VG-----G-F-----	.....L-----L-----IR-----	.....DOE-----M-----LV--A-----	79
H104.cpx.CY.94.CY032	MLF.....WE-W--G--L-V-----TL-F-----	.....K-LR--R--S-YN-IR-----	.....D-DAEE-STL--G--NFD--VG--N-----	82
H1N.CM.95.YBF30	ML.....SLGFI--GA-VS-VI-ALLYR-----	.....K-KL-E-KHIRO-IR-E-----	.....D-DAEW-DGDEEWLVTLLSSSKL-QGNWV-----	84
H10.BE.87.ANT70	MHH.....RDLL--IITSALLF-NVIL-GFI-LRKY.....	.....LEQKEODRKE-E-LERLR-IR-IRD-DY-N-----	.....EEOE-MDL-LSHGFDNPFEP-----	86
H10.CM.91.MVP5180	MHQ.....ENLL-LI--SALCL-NVLI-LFN-LRIY.....	.....LVQRKQDR-EQE-LERLR-IR-IRD-DY-N-----	.....EEOE-MELIISHGF-NPMFEL-----	86
H1P.FR.06.RBF168	MH.....RDEAVLITAG-LL-CI--GK-LLFVL-----	.....ERERD-FVORLV-WR-Q-G-Y-N-----	.....NE-EFEOLRELGNLGFGNIL-----	79
CPZ.CM.01.SIVcpzCAM13	MIL.....-ALGCLAIAL-LLN-FI-RNL--W....	.....RLCK...ONKLEVE-EN-AL-T-----	.....D-GEEELQR.LLHD-DNMMGFANP-FDI-----	84
CPZ.CM.05.SIVcpzEK505	MLL.....LIKLGFIG-AIETL-V--A--YRIY.....	.....EVKVEE--SO-RO-IRD-----	.....D-DAEE-----LANLPPDRI-ODNWV-----	80
CPZ.CM.05.SIVcpzLB7	MDL.....IELGLIG--IEL-V--LKA--YOLY.....	.....KENI--KA-NK--E-IR-G-----	.....D-DMDE-----LHAILRSGDPELVLDN-----	80
CPZ.CM.05.SIVcpzNB66	MDI.....VOQVGLL-V-IIEFLV-V-I-VK--Y....	.....KLCK...EDR-K-----A-IR-----	.....D-DTEE-QDILT-GDNLHMIGIR-NRNN-----	87
CPZ.CM.05.SIVcpzMT145	M-L.....E--L-ILFIALMLV--FA-IAA--Y....	.....KEYK-LQOV-R-E-Q-IRS-----	.....D-DEIL-----ELMQVHO-QVNPDM-RILFW-----	87
CPZ.CM.--SIVcpzMB897	MEIFI.....LGLIGIVIELV-A-VVWLKAYEC.Y....	.....KALKR--QE-R-Q--IR-----	.....D-DTEE-EAILIPEDR-VLVAIRGY-----	83
CPZ.CM.--SIVcpzDP943	MLT.....WEQIGLI--GIEG--T--G-AF-TW....	.....RRRRIEED-AQ-ID-LE-IRA-QD-----	.....D-DOROLMAYGFDNPMFD-----	82
CPZ.TZ.2001.TAN2	MVKLV.....VGS-LTNVIGIFC-LL-LIGGGLL-ITIIKREIERERORHILERLI-RLSIDSGVE.....	.....D-ELNWNFPDPHNYNPRD-I-----	.....D-ELNWNFPDPHNYNPRD-I-----	84
CPZ.TZ.2002.TAN3_1	MVKIV.....VGS-LTNVIGAF-C-LL-LIGGGLL-TAFVRELERERORHIVIERLV-RLSIDSGID.....	.....D-ELNWNFPDPHNYNPRD-I-----	.....D-ELNWNFPDPHNYNPRD-I-----	84
CPZ.TZ.00.TAN1	MIK.....IVGVSVMVIGILC-LLLILIGG...GLITG...TGIRRELERERORHIVIERLRLSI...VE...	.....D-EFNWNFPDPHNYNPRD-I-----	.....D-EFNWNFPDPHNYNPRD-I-----	84
CPZ.CM.98.CAM3	MLT.....WEQIGLIGIGIE-----A-G-A--FKEW...RKGKEEN-AQ--RT--E-I-S-Q-----	.....D-DOROLDREIHYVGFDPNPMFDW-----	.....D-DOROLDREIHYVGFDPNPMFDW-----	86
CPZ.CM.98.CAM5	MLI.....WEQIGLI--GIEL-V--G-A--YKWW...REIKKEEN-TOR-YN-AE-IRI-Q-----	.....N-EE-QLDNLHANGFDNPMFDW-----	.....N-EE-QLDNLHANGFDNPMFDW-----	87
CPZ.GA.88.GAB1	MTL.....LVGLVILVGL--AWN-CI-GYI--KWGY...RRYKRHRLETE--E-NLI-R-----	.....N-EEERL-QLIHNYN--NNHFANPMFDL-----	.....N-EEERL-QLIHNYN--NNHFANPMFDL-----	89
CPZ.GA.88.GAB2	MLS.....MWAIGLIGITLLV-N--G--GISVY...KRWKRH-EQRIT--LI-IKIT-----	.....D-DKETLATLLHNGFDNPMFE-RI-----	.....D-DKETLATLLHNGFDNPMFE-RI-----	89
CPZ.US.05.CPZUS	MLN.....WFE-GL-ALGI-EG-LVVI--GL--ARLW...ROIKT-ENT-QE-QN-LE-IRI-E-----	.....D-EETLAKLSSLELDNPRIV-----	.....D-EETLAKLSSLELDNPRIV-----	85
CPZ.CD.90.ANT	MTN.....IFEYAF-L-FSTVLWI-C-P--ILYKL...YKIV...KQOQDKNRN--I-EV-SR--SIDSAIEE.....	.....D-EADTYYLGSGFANPVYREGDE-----	.....D-EADTYYLGSGFANPVYREGDE-----	84
GOR.CM.2007.SIVgor2139_287	MH.....RD.IIVIIIGITLLAVT-IWLK-ALALY.....	.....L-DRRER-FF--ERL-SNKEDEGYESN-E-AAE-M--N-L-FDFNLH-----	.....L-DRRER-FF--ERL-SNKEDEGYESN-E-AAE-M--N-L-FDFNLH-----	80
GOR.CM.2007.SIVgorCP2135con	MH.....RD.IIVIIIGITLLA-T-IWLK-ALALY.....	.....L-DRRER-FFN--ERL-STKEDEGYESN-E-AAE-M-R-N-L-FDFNLH-----	.....L-DRRER-FFN--ERL-STKEDEGYESN-E-AAE-M-R-N-L-FDFNLH-----	80
GOR.CM.2004.SIVgorCP684con	MH.....RD.ILVIIIGITLLAVT-ISWLK-ALALY.....	.....L-DRRER-FF--ERL-SNKEDEGYESN-E-AAE-M-R-N-L-FDFNLH-----	.....L-DRRER-FF--ERL-SNKEDEGYESN-E-AAE-M-R-N-L-FDFNLH-----	80
DEN.CD.x.CD1	MLK.....-KLGPIEY-CLVF-V-ITW-AA-GVGY.....	.....LAVRAYKSYREE-RYIRLRWSIDSGYESSQEDP-----	.....LAVRAYKSYREE-RYIRLRWSIDSGYESSQEDP-----	65
MON.CM.99.L1	MNY.....WMSLVAITYS-ILIALPVAAWAW-W--RYY...KITK-FKRIDQE-Q--OIER-RH-----	.....VDTSEFS-QHE--THGFVNPVFN--FGEWV-----	.....VDTSEFS-QHE--THGFVNPVFN--FGEWV-----	88
GSN.CM.99.CN166	MSA.....AA-WMWGAIIIF-YLC-VALL--ALY...LAWD-WVKGKPKPTQVAVIRLIEDEE.DSGIYDDASELT-F.NGF-NPGFEV.....	.....LAWD-WVKGKPKPTQVAVIRLIEDEE.DSGIYDDASELT-F.NGF-NPGFEV.....	.....LAWD-WVKGKPKPTQVAVIRLIEDEE.DSGIYDDASELT-F.NGF-NPGFEV.....	79
GSN.CM.99.CN71	MSA.....AALWVGAA-ITF-YLC-LA-F--ALY...LAWD-WVKGKPKPTQVAVIRLIEDEE.DSGIYDDASELT-F.NGF-NPGFEV.....	.....LAWD-WVKGKPKPTQVAVIRLIEDEE.DSGIYDDASELT-F.NGF-NPGFEV.....	.....LAWD-WVKGKPKPTQVAVIRLIEDEE.DSGIYDDASELT-F.NGF-NPGFEV.....	77
MUS-1.CM.01.1085	MNY.....WYLAA-IVTGIVYV-A-FAF-----VLAY.....	.....ORWCKPK--VEVSVIRLLE-GD-DSGIF-DAEDDMAES-HHAF-NPAFEQ-----	.....ORWCKPK--VEVSVIRLLE-GD-DSGIF-DAEDDMAES-HHAF-NPAFEQ-----	77
MUS-1.CM.01.CM1239	MNY.....WYLAALVTGIYV-ALFAF-----VLAY.....	.....ORWC-PK--VEVSVIRLLE-GDSGIF-DAEDFN-DGHAF-NPAFEQ-----	.....ORWC-PK--VEVSVIRLLE-GDSGIF-DAEDFN-DGHAF-NPAFEQ-----	76
MUS-2.CM.01.CM1246	MNY.....WYLGAAIVT-IVYV-ALVAF-----VLAY.....	.....ORWCQPK-PPTVEVNVRLLE-GDSGIF-DACDG-DEDSHRAF-NPSFEP-----	.....ORWCQPK-PPTVEVNVRLLE-GDSGIF-DACDG-DEDSHRAF-NPSFEP-----	78
MUS-2.CM.01.CM2500	MNW.....WVFAAAVVT-IVYV-ALVAF-----VLAY.....	.....ORWCQPK-PPTVEVNVIRLLE-GDTSDFGIF-DAEDGSDORH-FLN-AFEL-----	.....ORWCQPK-PPTVEVNVIRLLE-GDTSDFGIF-DAEDGSDORH-FLN-AFEL-----	78



Table with columns for glycosylation sites (V1, V2, V3) and amino acid sequences. Includes sub-headers like glycosylation NIS, glycosylation NCS, glycosylation NDT, glycosylation NTS, glycosylation NGT, glycosylation NKT, glycosylation NVS, glycosylation NVT, glycosylation NWT, glycosylation NXX, glycosylation NYX, glycosylation NYZ, glycosylation NZX, glycosylation NZY, glycosylation NZZ, glycosylation NZA, glycosylation NZB, glycosylation NZC, glycosylation NZD, glycosylation NZE, glycosylation NZF, glycosylation NZG, glycosylation NZH, glycosylation NZI, glycosylation NZJ, glycosylation NZK, glycosylation NZL, glycosylation NZM, glycosylation NZN, glycosylation NZO, glycosylation NZP, glycosylation NZQ, glycosylation NZR, glycosylation NZS, glycosylation NZT, glycosylation NZU, glycosylation NZV, glycosylation NZW, glycosylation NZX, glycosylation NZY, glycosylation NZZ.

Table with columns for glycosylation NGS, glycosylation NNT, glycosylation NCT, glycosylation NTS, glycosylation NNT, glycosylation NKT, glycosylation NKT, CD4, and V4. It lists various protein sequences and their corresponding glycosylation sites.

Table with columns for glycosylation NGS, glycosylation NNT, glycosylation NCT, glycosylation NTS, glycosylation NNT, glycosylation NKT, glycosylation NKT, CD4, and V4. It lists various protein sequences and their corresponding glycosylation sites, including HIV sequences.





H1B\_FR.83.HXB2
H1A1\_UG.85.U455
H1B\_US.90.WEAU160
H1C\_ET.86.ETH2220
H1D\_CD.84.84ZP885
H1F1\_BE.93.VK194
H1G\_SE.93.SE6165
H1H\_CF.90.056
H1J\_SE.93.SE7887
H1K\_CM.96.MP535
H1O1\_AE.TH.90.CM240
H1O2\_AG.NG.186C
H1O3\_AB.RU.97.KAL153\_2
H1O4\_cpx.CY.94.CY032
H1N\_CM.95.YBF30
H1O\_BE.87.ANT76
H1O\_CM.91.MVP5180
H1P\_FR.06.RBF168
CPZ\_CM.01.SIVcpzCAM13
CPZ\_CM.05.SIVcpzEK505
CPZ\_CM.05.SIVcpzLBT
CPZ\_CM.05.SIVcpzMB66
CPZ\_CM.05.SIVcpzMT145
CPZ\_CM.05.SIVcpzMB897
CPZ\_CM.05.SIVcpzDP943
CPZ\_CD.90.ANT
CPZ\_TZ.2001.TAN2
CPZ\_TZ.2002.TAN3\_1
CPZ\_TZ.00.TAN1
CPZ\_GA.88.GAB2
CPZ\_GA.88.GAB1
CPZ\_US.85.CPZUS
GOR\_CM.2007.SIVgor2139\_287
GOR\_CM.2007.SIVgorCP2135con
GOR\_CM.2004.SIVgorCP684con
MON\_CM.99.L1
MON\_NG.x.NG1
GSI\_CM.99.CM166
CPZ\_CM.98.CAM3
CPZ\_CM.98.CAM5
TAL\_CM.00.266
TAL\_CM.01.8023
GSI\_CM.99.CN71
MUS\_1\_CM.01.T1895
MUS\_1\_CM.01.CM1239
MUS\_2\_CM.01.CM1246
MND\_1\_GA.x.MNDG61
WRC\_C1.97.97C114
WRC\_C1.98.98C104
WRC\_GM.05.Pb7O5GM\_X02
OLC\_C1.97.97C112
MUS\_2\_CM.01.CM2500
DEB\_CM.99.CM40
DEB\_CM.99.CM5
DEB\_CM.99.CM3
H2A.DE.x.BEN
H2A.GW.x.ALI
H2A.SN.x.ST
H2B.CI.x.EH0
H2B.GH.86.D205
H2G.CI.92.ABT96
H2U.FR.96.12034
SHM\_SL.92.SL92B
SHM\_US.x.PGM53
SHM\_US.x.SIVsmh635F\_L3

MAC\_US.x.239
MAC\_US.x.251.IA11
MAC\_US.x.251.BK28
MAC\_US.x.EMB1\_3
STM\_US.x.STM
MNE\_US.x.MNE027
LST\_CD.88.447
LST\_CD.88.485
LST\_CD.88.524
LST\_KE.x.lh07
SUN\_GA.98.L14
MND\_2\_GA.x.M14
MND\_2\_x.x.5440
DRL\_x.x.FA0
SYK\_KE.x.KE51
SYK\_KE.x.SYK173
RCM\_GA.x.GAB1
RCM\_NG.x.NG411
GRV\_ET.x.GRI\_677
SAB\_SNI.x.SAB1C
TAN\_UG.x.TAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TY01.patent
COL\_CM.x.CG11

Table with columns for glycosylation NAS, glycosylation NHT, and glycosylation NYT. Rows correspond to the protein entries listed on the left, showing amino acid sequences and their glycosylation sites.

Table with columns for glycosylation NAS, glycosylation NHT, and glycosylation NYT. Rows correspond to the protein entries listed on the left, showing amino acid sequences and their glycosylation sites.



```

gp41 end
Env end
      *
H1B.FR.83.HXB2      IEVV...QGACRAIRHIPRRIROGLER...ILL.....*
H1A1.UG.85.U455     --IG...TIG--LN-----A-----
H1B.US.90.WEAU160   --I...RT--L-----A-----
H1C.ET.86.ETH2220   --LI...RIW--FCN-----A-A-Q
H1D.CD.84.84ZRB85   --DI...RR--K-VL--T-----A-----
H1F1.BE.93.VI850    --L...R-G-VLN-----A-----
H1G.SE.93.SE6165     --A...R---LN-T-----A-----
H1H.CF.90.056       --VI...R-W--L-----F-S-----
H1J.SE.93.SE7887     --IA...R-F--L-----A-----
H1K.CM.96.MP535     --IG...R-F--LL-----A-----
H101.AE.TH.90.CM240 --A...W--L-----T-----
H102.AG.NG.x.IENG    --A...RVG--N-----F-A-----
H103.AB.RU.97.KAL153 2 --IG...RF--M-N-----A-K-A-Q
H104.cpx.CY.94.CY032 --A...R---CN-----A-----
H1N.CM.95.YBF30     --LA...RIG-G-L-----A-I-----
H10.BE.87.ANT76     --AGI...RIGTG-N-----A-S-----
H10.CM.91.MVP5180    --LGL...RIGGFL-----A-S-V-----
H1P.FR.06.RBF168     --I-G...QIG-GFLN-----I-S-----
CPZ.CM.01.SIVcpzCAM13 --RA...IVF-I-GN-----T-----
CPZ.CM.05.SIVcpzEK505 --L...RIG-G-L-----A-----
CPZ.CM.05.SIVcpzLB7  --I...RIG-G-L-----F-A-----
CPZ.CM.05.SIVcpzMB66 --IA...RFV-G-LN-----A-----
CPZ.CM.05.SIVcpzMT145 --L-AI...TRLG-G-L-----A-----
CPZ.CM.--SIVcpzMB897 --IA...RVG-G-L-----A-----
CPZ.CM.--SIVcpzDP943 --...RIG-G-L-----S-----
CPZ.CD.90.ANT        LLAL...TIV-I-EV-----I-A-N-----
CPZ.TZ.2001.TAN2     --GI...TVIG-G-N-----I-A-N-----
CPZ.TZ.2002.TAN3_1   --G...VYA-G-N-----L-A-N-----
CPZ.TZ.00.TAN1       --G...VYA-G-N-----L-A-N-----
CPZ.GA.88.GAB2       --A...TRIG-G-L-----A-E-----
CPZ.GA.88.GAB1       --AF...VTL-I-N-----A-----
CPZ.US.85.CP2US      --LT...RRLFLG-I-----S-----
GOR.CM.2007.SIVgor2139 287 --AIA...RIG-G-LN-----S-----
GOR.CM.2007.SIVgorCP2135con --AIA...RIG-G-LN-----S-----
GOR.CM.2004.SIVgorCP684con --AIA...RIG-G-LN-----S-----
MON.CM.99.L1         LHAG...GRLW--VA-----A-I-F-N-----
MON.NG.x.NG1        LRLG...GRLW-GLVAV-----A-I-L-N-----
G5N.CM.99.CM166     LHAC...RRVW-EFLA-----A-I-LFN-----
CPZ.CM.98.CAM3      L-A...IIG-G-L-----S-----
CPZ.CM.98.CAM5      --F...IVG-G-L-----S-----
TAL.CM.00.266       LAAA...RW-GE-AA--T--Y-CFT-----
TAL.CM.01.8023      LAAA...RW-VQ-AA--T--Y-C-A-----
G5N.CM.99.CM71      VOT...GRVW-EFLA-----A-I-L-N-----
MUS-1.CM.01.1085    LOAG...RRVW-EFLA-----A-I-L-N-----
MUS-1.CM.01.CM1239  LOIG...RRVW-EFLA-----A-I-L-N-----
MUS-2.CM.01.CM1246  LOFS...KIW-KFLA-----A-I-L-N-----
MND-1.GA.x.MNDGB1   --IAA...GTC-W-L--SA--P-N-----
WRC.CI.97.97CI14    KTAA...SSRI-MLATRLCASFDKCR--SFSERRQ**
WRC.CI.98.98CI04    ETTG...SSRVWKLITRLCSLWMSR--R-F-----
WRC.GM.05.Pbt_05GM_X02 ETSR...SSSI-G-TARLCTNLNPCR--GOO-----
OLC.CI.97.97CI12    QOKK...RRGSPLOKAQEQLGREGAE.TTGV-----
MUS-2.CM.01.CM2500  LQ-S...KOL*K-FLA-K--A-I-L-N-----
DEB.CM.99.CM40     LOGL...VR-GG-LRV-A--A-L-N-----
DEB.CM.99.CM5      --OGL...VMGG-NLLA-A--A-L-L-N-----
DEN.CD.x.CD1       LAVAL...R-A-EVA-----I-V-N-----
H2A.DE.x.BEN        W-AA...RRIG-G-LAV--A-L-A-----
H2A.GW.x.ALI        WGA...WVG-R-LAV--A-I-A-----
H2A.SN.x.ST         WGTL...GOIG-G-LAV--A-I-A-----
H2B.CI.x.EH0        WG-L...RR-AGE-IA--A-L-A-----
H2B.GH.86.D205     W-AL...RR-A--IA-----L-T-----
H2G.CI.92.ABT96     WKTL...GRVG-G-LA-----L-T-----
H2U.FR.96.12034     W-TL...GRVG-WL-A-----F-L-A-----
SMM.SL.92.SL92B     W-TL...GRVG-G-AA-----L-M-N-----
SMM.US.x.PGM53      W-TL...GRVG-WVA-----L-A-----
SMM.US.x.SIVsmH635F_L3 W-TL...GRVG-G-LA-----V-----L-T-----
      Env end
MAC.US.x.239        W-TL...RRGG-W-LA-----L-T-----
MAC.US.x.251.IA11   *-AL...RRGG-W-LA-----L-T-----
MAC.US.x.251.BK28   W-TL...RRGG-W-LA-----L-T-----
MAC.US.x.EMBL_3     W-TL...RRGG-W-LA-----L-T-----
STM.US.x.STM        W-TL...GRVG-R-GA-----L-T-----
MNE.US.x.MNE027     W-TL...GRVG-W-LA--E--L-T-----
LST.CD.88.447       NLRL...GQKK-WRFRFRG-SGFPS-TTETA-----
LST.CD.88.485       D1RL...GQKK-WRFRFRG-SGFPS-TTETA-----
LST.CD.88.524       N1RL...GQNR-RW-RFRF-SGLPS-TTETT-----
LST.KE.x.lho7       NIQL...GKKK-WRLRFGG-SGIS-ATETA-----
SUN.GA.98.L14       RSRTFSLGRKW-PKWNRT-GS-IPS-TTETT-----
MND-2.GA.x.M14      RGAA...AIG--GN-----A-V-L-N-----
MND-2.x.x.5440      RGAA...AFG-G-WN-----A-A-L-N-----
DRL.x.x.FA0         ARAL...RALAGEVA-----A-V-LFN-----
SYK.KE.x.RE51       LAYA...RVAENVAAL--L-----I-Y-N-----
SYK.KE.x.SYK173     YAAT...RRVVE-VAAL--L-----I-Y-N-----
RCM.GA.x.GAB1       CA-A...RDFAGWPAMVC-----LCN-----
RCM.NG.x.NG411      CA-C...RDPAGWPATLC-----F-F-N-----
GRV.ET.x.GRI_677    GAI...RS-Y--VINS--V--K-V-G-----
SAB.SN.x.SAB1C      RHAC...RSIV-VIAH--M-E--WFN-----
TAN.UG.x.TAN1       WNAC...RS-Y--LEH--M-E--WFN-----
VER.DE.x.AGM3       WLAC...RS-Y--INS--V--G--N-----
VER.KE.x.9063       WLAC...RSTY-H-ISS--V--E--N-----
VER.KE.x.AGM155     WLAC...RS-Y--INS--V--E--V-N-----
VER.KE.x.TY01_patent WLAC...RS-Y-N-VNS--V--E--N-----
COL.CM.x.CG1        LRSA...ARGW-RAPEYL-GWIYDRPQ.GPA-----

```

856  
856  
860  
852  
857  
853  
867  
846  
849  
843  
858  
855  
840  
856  
853  
851  
870  
879  
865  
836  
846  
878  
853  
842  
850  
859  
877  
895  
872  
857  
855  
857  
855  
850  
859  
870  
884  
882  
860  
866  
873  
891  
887  
898  
892  
883  
879  
879  
885  
877  
844  
879  
871  
851  
865  
861  
863  
861  
853  
860  
885  
868  
887  
886  
888  
  
880  
877  
881  
882  
883  
883  
916  
915  
91  
913  
920  
868  
891  
896  
84  
833  
851  
858  
855  
868  
880  
878  
878  
871  
864  
838



