

VII

PLV Proteins

Contents

VII-1	Introduction	395
VII-2	Sequences	396
VII-3	Alignments	402
	VII-3.1 Gag	402
	VII-3.2 Pol	406
	VII-3.3 Vif	413
	VII-3.4 Vpx	415
	VII-3.5 Vpr	416
	VII-3.6 Tat	417
	VII-3.7 Rev	419
	VII-3.8 Vpu	420
	VII-3.9 Env	421
	VII-3.10Nef	428

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 XXX 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”.

VII-2 Sequences

Sequences included in the PLV protein alignments.

Name	Accession	Proteins	Author	Reference
BLU.KE.x.KE31	DQ222474	Nef	Bibollet-Ruche, F	<i>J Virol</i> 78 (14); 7748-62 (2004)
COL.CM.x.CGU1	AF301156	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Courgnaud, V	<i>J Virol</i> 75 (2); 857-66 (2001)
CPZ.CD.90.ANT	U42720	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, MM	<i>Virology</i> 221 (2); 346-50 (1996)
CPZ.CM.01.SIVcpzCAM13	AY169968	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env	Nerrienet, E	<i>J Virol</i> 79 (2); 1312-9 (2005)
CPZ.CM.05.SIVcpzEK505	DQ373065	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, BF	<i>Science</i> 2006 May 25;
CPZ.CM.05.SIVcpzLB7	DQ373064	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, BF	<i>Science</i> 2006 May 25;
CPZ.CM.05.SIVcpzMB66	DQ373063	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, BF	<i>Science</i> 2006 May 25;
CPZ.CM.05.SIVcpzMT145	DQ373066	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Keele, BF	<i>Science</i> 2006 May 25;
CPZ.CM.98.CAM3	AF115393	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Corbet, S	<i>J Virol</i> 74 (1); 529-34 (2000)
CPZ.CM.98.CAM5	AJ271369	Gag, Pol, Vif, Vpr, Tat, Vpu, Env, Nef	Muller-Trutwin, MC	<i>J Med Primatol</i> 29 (3-4); 166-72 (2000)
CPZ.GA.88.GAB1	X52154	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Huet, T	<i>Nature</i> 345 (6273); 356-9 (1990)
CPZ.GA.88.GAB2	AF382828	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F	<i>ARHR</i> 20 (12); 1377-81 (2004)
CPZ.TZ.01.TAN1	AF447763	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Santiago, ML	<i>J Virol</i> 77 (3);2233-2242 (2003)

Name	Accession	Proteins	Author	Reference
CPZ.US.85.CPZUS	AF103818	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F	<i>Nature</i> 397 (6718); 436-41 (1999)
DEB.CM.99.CM40	AY523865	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bibollet-Ruche, F	<i>J Virol</i> 78 (14); 7748-62 (2004)
DEB.CM.99.CM5	AY523866	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Bibollet-Ruche, F	<i>J Virol</i> 78 (14); 7748-62 (2004)
DEN.CD.x.CD1	AJ580407	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Dazza, MC	<i>J Virol</i> 79 (13); 8560-71 (2005)
DRL.x.x.FAO	AY159321	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Hu, J	<i>J Virol</i> 77 (8):4867-4880 (2003)
GRV.ET.x.GRI_677	M66437	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Fomsgaard, A	<i>Virology</i> 182 (1); 397-402 (1991)
GSN.CM.99.CN166	AF468659	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V	<i>J Virol</i> 76 (16); 8298-309 (2002)
GSN.CM.99.CN71	AF468658	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V	<i>J Virol</i> 76 (16); 8298-309 (2002)
H101_AE.TH.90.CM240	U54771	Gag, Pol, Vif, Vpr, Tat, Vpu, Env, Nef	Laukkanen, T	<i>J Virol</i> 70 (9); 5935-43 (1996)
H102_AG.NG.x.IBNG	L39106	Gag, Pol, Vif, Vpr, Tat, Vpu, Env, Nef	Howard, TM	<i>ARHR</i> 10 (12); 1755-7 (1994)
H103_AB.RU.97.KAL153_2	AF193276	Vpu	Liitsola, K	<i>AIDS</i> 12 (14); 1907-19 (1998)
H104_cpx.CY.94.CY032	AF049337	Vpu	Gao, F	<i>J Virol</i> 72 (12); 10234-41 (1998)
H1A1.UG.85.U455	M62320	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Oram, JD	<i>ARHR</i> 6 (9); 1073-8 (1990)
H1B.FR.83.HXB2	K03455	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Wong-Staal, F	<i>Nature</i> 313 (6000); 277-84 (1985)
H1B.US.90.WEAU160	U21135	Vpu	Tozser, J	<i>FEBS Lett</i> 281 (1-2); 77-80 (1991)
H1C.ET.86.ETH2220	U46016	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Salminen, MO	<i>ARHR</i> 12 (14); 1329-39 (1996)

Name	Accession	Proteins	Author	Reference
H1D.CD.84.84ZR085	U88822	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gao, F	<i>J Virol</i> 72 (7); 5680-98 (1998)
H1F1.BE.93.VI850	AF077336	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T	<i>Virology</i> 269 (1); 95-104 (2000)
H1G.SE.93.SE6165	AF061642	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T	<i>Virology</i> 247 (1); 22-31 (1998)
H1H.CF.90.056	AF005496	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Murphy, E	<i>ARHR</i> 9 (10); 997-1006 (1993)
H1J.SE.93.SE7887	AF082394	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Laukkanen, T	<i>ARHR</i> 15 (3); 293-7 (1999)
H1K.CM.96.MP535	AJ249239	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Peeters, M	<i>ARHR</i> 16 (2); 139-51 (2000)
H1N.CM.95.YBF30	AJ006022	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Simon, F	<i>Nat Med</i> 4 (9); 1032-7 (1998)
H1O.BE.87.ANT70	L20587	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Vanden Haesevelde, M	<i>J Virol</i> 68 (3); 1586-96 (1994)
H1O.CM.91.MVP5180	L20571	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Gurtler, LG	<i>J Virol</i> 68 (3); 1581-5 (1994)
H2A.DE.x.BEN	M30502	Pol, Vif, Vpx	Kirchhoff, F	<i>Virology</i> 177 (1); 305-11 (1990)
H2A.GW.x.ALI	AF082339	Gag, Vpx, Vpr, Tat, Rev, Env, Nef	Azevedo-Pereira, JM	Unpublished (1998)
H2A.SN.x.ST	M31113	Pol, Vpx	Kumar, P	<i>J Virol</i> 64 (2); 890-901 (1990)
H2B.CL.x.EHO	U27200	Pol, Vpx, Vpr, Tat, Rev, Nef	Rey-Cuille, MA	<i>Virology</i> 202 (1); 471-6 (1994)
H2B.GH.86.D205	X61240	Gag, Pol, Vif, Vpx, Env	Kreutz, R	<i>ARHR</i> 8 (9); 1619-29 (1992)
H2G.CL.x.ABT96	AF208027	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Brennan, CA	<i>ARHR</i> 13 (5); 401-4 (1997)

Name	Accession	Proteins	Author	Reference
H2U.FR.96.12034	AY530889	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Damond, F	<i>ARHR</i> 20 (6); 666-72 (2004)
LST.CD.88.447	AF188114	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, BE	<i>J Virol</i> 74 (8); 3892-8 (2000)
LST.CD.88.485	AF188115	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, BE	<i>J Virol</i> 74 (8); 3892-8 (2000)
LST.CD.88.524	AF188116	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, BE	<i>J Virol</i> 74 (8); 3892-8 (2000)
LST.KE.x.lho7	AF075269	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, VM	<i>J Virol</i> 73 (2); 1036-45 (1999)
MAC.US.x.239	M33262	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kestler, H	<i>Science</i> 248 (4959); 1109-12 (1990)
MAC.US.x.251_1A11	M76764	Vpx, Rev	Planelles, V	<i>ARHR</i> 7 (11); 889-98 (1991)
MAC.US.x.251_BK28	M19499	Vpx, Rev	Franchini, G	<i>Nature</i> 328 (6130); 539-43 (1987)
MAC.US.x.EMBL_3	Y00295	Pol, Rev	Franchini, G	<i>Nature</i> 328 (6130); 539-543 (1987)
MND_1.GA.x.MNDGB1	M27470	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Tsujimoto, H	<i>Nature</i> 341 (6242); 539-41 (1989)
MND_2.CM.98.CM16	AF367411	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Takehisa, J	<i>ARHR</i> 17 (12); 1143-54 (2001)
MND_2.GA.x.M14	AF328295	Gag, Pol, Vif, Vpx, Vpr, Tat, Env, Nef	Souquiere, S	<i>J Virol</i> 75 (15); 7086-96 (2001)
MND_2.x.x.5440	AY159322	Gag, Pol, Vif, Vpx, Vpr, Tat, Env, Nef	Hu, J	<i>J Virol</i> 77 (8); 4867-4880 (2003)
MNE.US.x.MNE027	U79412	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Kimata, JT	<i>J Virol</i> 72 (1); 245-56 (1998)
MON.CM.99.L1	AY340701	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V	<i>J Virol</i> 77 (23); 12523-12534 (2003)
MON.NG.x.NG1	AJ549283	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Barlow, KL	<i>J Virol</i> 77 (12); 6879-88 (2003)
MUS_1.CM.01.1085	AY340700	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Courgnaud, V	<i>J Virol</i> 77 (23); 12523-12534 (2003)

Name	Accession	Proteins	Author	Reference
MUS_1.CM.01.CM1239	EF070330	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, AF	<i>(er) Virology</i> 360 (2):407-418 (2007)
MUS_2.CM.01.CM1246	EF070329	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, AF	<i>(er) Virology</i> 360 (2):407-418 (2007)
MUS_2.CM.01.CM2500	EF070331	Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Aghokeng, AF	<i>(er) Virology</i> 360 (2):407-418 (2007)
RCM.GA.x.GAB1	AF382829	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Gao, F	<i>Science</i> 300 (5626); 1713 (2003)
RCM.NG.x.NG411	AF349680	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Beer, BE	<i>J Virol</i> 75 (24); 12014-27 (2001)
SAB.SN.x.SAB1C	U04005	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Jin, MJ	<i>EMBO J</i> 13 (12); 2935-47 (1994)
SMM.SL.92.SL92B	AF334679	Pol, Vpx, Rev	Chen, Z	<i>J Virol</i> 70 (6); 3617-27 (1996)
SMM.US.x.H9	M80194	Gag, Vpx, Rev, Env	Courgnaud, V	<i>J Virol</i> 66 (1); 414-9 (1992)
SMM.US.x.PGM53	AF077017	Vif, Vpx, Vpr, Tat, Rev, Nef	Novembre, FJ	<i>J Virol</i> 72 (11); 8841-51 (1998)
SMM.US.x.SIV sm H635FC	DQ201174	Vpx, Rev	Kuwata, T	<i>J Virol</i> 80 (3); 1463-75 (2006)
SMM.US.x.SIV sm H635F_L3	DQ201172	Vpx, Rev	Kuwata, T	<i>J Virol</i> 80 (3); 1463-75 (2006)
SMM.US.x.SIV sm H635SB10	DQ201173	Vpx, Rev	Kuwata, T	<i>J Virol</i> 80 (3); 1463-75 (2006)
STM.US.x.STM	M83293	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env, Nef	Novembre, FJ	<i>Virology</i> 186 (2); 783-7 (1992)
SUN.GA.98.L14	AF131870	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Beer, BE	<i>J Virol</i> 73 (9); 7734-44 (1999)
SUN.GA.x.SOL_36	DQ222476	Nef	Schindler, M	<i>Cell</i> 125 (6); 1055-67 (2006)
SYK.KE.x.KE44	DQ222473	Nef	Bibollet-Ruche, F	<i>J Virol</i> 78 (14); 7748-62 (2004)
SYK.KE.x.KE51	AY523867	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Bibollet-Ruche, F	<i>J Virol</i> 78 (14); 7748-62 (2004)
SYK.KE.x.SYK173	L06042	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, VM	<i>J Virol</i> 67 (3); 1517-28 (1993)
TAL.CM.00.266	AY655744	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F	<i>Virology</i> 349 (1):55-65 (2006)

Name	Accession	Proteins	Author	Reference
TAL.CM.01.8023	AM182197	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Liegeois, F	<i>Virology</i> 349 (1); 55-65 (2006)
TAN.FR.x.B87_14	DQ222475	Nef	Schindler, M	<i>Cell</i> 125 (6); 1055-67 (2006)
TAN.UG.x.TAN1	U58991	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Stivahtis, GL	<i>Virology</i> 228 (2); 394-9 (1997)
VER.DE.x.AGM3	M30931	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Baier, M	<i>Virology</i> 176 (1); 216-21 (1990)
VER.KE.x.9063	L40990	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Hirsch, VM	<i>J Virol</i> 69 (2); 955-67 (1995)
VER.KE.x.AGM155	M29975	Gag, Pol, Vif, Vpr, Tat, Rev, Env, Nef	Johnson, PR	<i>J Virol</i> 64 (3); 1086-92 (1990)
VER.KE.x.TYO1	DJ048201	Gag, Env	Omori, T	Patent: WO 2007049749-A 13 03-MAY-2007; Dनावेक CORPORATION
VER.KE.x.TYO1	X07805	Pol, Vif, Vpr, Tat, Rev, Nef	Fukasawa, M	<i>Nature</i> 333 (6172); 457-61 (1988)

	membrane binding	Gag start, p17 start	nuclear localization	phosphorylation site	nuclear localization	p17 end	p24 start
H1B.FR.83.HXB2	MSGARSASVLSGGELDRWEKIRLRPGGKKYKLVKHIWASRELERFAVNPGLLETSEGRQILGLOPLSQTGSEELRSYNTVATLYCVHQRIEIKDKALDKEIEEQ	-----	-----	-----	-----	-----	-----
H1A1.UG.85.U455	-----	-----	-----	-----	-----	-----	-----
H1C.ET.86.ETH2220	-----	-----	-----	-----	-----	-----	-----
H1D.CD.84.84ZR085	-----	-----	-----	-----	-----	-----	-----
H1F1.BE.93.VI850	-----	-----	-----	-----	-----	-----	-----
H1G.SE.93.SE6165	-----	-----	-----	-----	-----	-----	-----
H1H.CF.90.056	-----	-----	-----	-----	-----	-----	-----
H1J.SE.93.SE7887	-----	-----	-----	-----	-----	-----	-----
H1K.CM.96.MP535	-----	-----	-----	-----	-----	-----	-----
H1O1.AE.TH.90.CM240	-----	-----	-----	-----	-----	-----	-----
H1O2.AG.NG.x.IBNG	-----	-----	-----	-----	-----	-----	-----
H1N.CM.95.YBF30	-----	-----	-----	-----	-----	-----	-----
H1O.BE.87.ANT170	-----	-----	-----	-----	-----	-----	-----
H1O.CM.91.MVP5180	-----	-----	-----	-----	-----	-----	-----
CPZ.CM.05.SIVcpzEK505	-----	-----	-----	-----	-----	-----	-----
CPZ.CM.05.SIVcpzLB7	-----	-----	-----	-----	-----	-----	-----
CPZ.CM.01.SIVcpzCAM13	-----	-----	-----	-----	-----	-----	-----
CPZ.CM.05.SIVcpzMB66	-----	-----	-----	-----	-----	-----	-----
CPZ.CM.05.SIVcpzMT145	-----	-----	-----	-----	-----	-----	-----
CPZ.CD.90.ANT	-----	-----	-----	-----	-----	-----	-----
CPZ.CM.98.CAM3	-----	-----	-----	-----	-----	-----	-----
CPZ.CM.98.CAM5	-----	-----	-----	-----	-----	-----	-----
CPZ.US.85.CPZUS	-----	-----	-----	-----	-----	-----	-----
CPZ.GA.88.GAB1	-----	-----	-----	-----	-----	-----	-----
CPZ.TZ.01.TAN1	-----	-----	-----	-----	-----	-----	-----
CPZ.GA.88.GAB2	-----	-----	-----	-----	-----	-----	-----
H2A.GW.x.ALI	-----	-----	-----	-----	-----	-----	-----
H2B.GH.86.D205	-----	-----	-----	-----	-----	-----	-----
H2G.CL.x.ABT96	-----	-----	-----	-----	-----	-----	-----
H2U.FR.96.12034	-----	-----	-----	-----	-----	-----	-----

	Gag p15 start	p15	p27
MAC.US.x.239	-----	-----	-----
SMM.US.x.H9	-----	-----	-----
STM.US.x.STM	-----	-----	-----
MNE.US.x.MNE027	-----	-----	-----
DRL.x.x.FAO	-----	-----	-----
MND_1.GA.x.MNDGB1	-----	-----	-----
MND_2.GA.x.M14	-----	-----	-----
MND_2.CM.98.CM16	-----	-----	-----
MND_2.x.x.5440	-----	-----	-----
TAL.CM.01.8023	-----	-----	-----
TAL.CM.00.266	-----	-----	-----
TAN.UG.x.TAN1	-----	-----	-----
VER.KE.x.TYO1	-----	-----	-----
VER.DE.x.AGM3	-----	-----	-----
VER.KE.x.9063	-----	-----	-----
VER.KE.x.AGM155	-----	-----	-----
GRV.ET.x.GRL_677	-----	-----	-----
SAB.SN.x.SAB1C	-----	-----	-----
RCM.NG.x.NG411	-----	-----	-----
RCM.GA.x.GAB1	-----	-----	-----
SUN.GA.98.L14	-----	-----	-----
MON.NG.x.NG1	-----	-----	-----
MON.CM.99.L1	-----	-----	-----
MUS_1.CM.01.CM1239	-----	-----	-----
MUS_1.CM.01.1085	-----	-----	-----
MUS_2.CM.01.CM2500	-----	-----	-----
MUS_2.CM.01.CM1246	-----	-----	-----
DEB.CM.99.CM40	-----	-----	-----
DEB.CM.99.CM5	-----	-----	-----
GSN.CM.99.CN166	-----	-----	-----
GSN.CM.99.CN71	-----	-----	-----
DEN.CD.x.CD1	-----	-----	-----
LST.CD.88.447	-----	-----	-----
LST.CD.88.485	-----	-----	-----
LST.KE.x.lho7	-----	-----	-----
LST.CD.88.524	-----	-----	-----
SYK.KE.x.SYK173	-----	-----	-----
SYK.KE.x.KE51	-----	-----	-----
COL.CM.x.CGU1	-----	-----	-----

H1B.FR.83.HXB2
H1A1.UG.85.U455
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H1O1.AE.TH.90.CM240
H1O2.AG.NG.x.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMT145
CPZ.CD.90.ANT
CPZ.CM.98.CAM3
CPZ.CM.98.CAM5
CPZ.US.85.CPZUS
CPZ.GA.88.GAB1
CPZ.TZ.01.TAN1
CPZ.GA.88.GAB2
H2A.GW.x.ALI
H2B.GH.86.D205
H2G.CL.x.AB796
H2U.FR.96.12034
MAC.US.x.239
SMM.US.x.H9
STM.US.x.STM
MNE.I.x.MNE027
DRL.x.x.FAO
MND.1.GA.x.MNDGB1
MND.2.GA.x.M14
MND.2.CM.98.CM16
MND.2.x.x.5440
TAL.CM.01.8023
TAL.CM.00.266
TAN.UG.x.TAN1
VER.KE.x.TYO1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
GRV.ET.x.GRL.677
SAB.SN.x.SAB1C
RCM.NG.x.NG411
RCM.GA.x.GAB1
SUN.GA.98.L14
MON.NG.x.NG1
MON.CM.99.L1
MUS.1.CM.01.CM1239
MUS.1.CM.01.1085
MUS.2.CM.01.CM2500
MUS.2.CM.01.CM1246
DEB.CM.99.CM40
DEB.CM.99.CM5
GSN.CM.99.CM166
GSN.CM.99.CM71
DEN.CD.x.CD1
LST.CD.88.447
LST.CD.88.485
LST.KE.x.lho7
LST.CD.88.524
SYK.KE.x.SYK173
SYK.KE.x.KE51
COL.CM.x.CGU1

Table with 2 columns: CyPA binding and major homology region. The table lists amino acid sequences for various HIV strains, with the first column representing the CyPA binding site and the second column representing the major homology region. The sequences are aligned to show conserved and variable regions. The table ends with a line number 300.

	p24 end_p2 start	p2 end_p7 start	Zn motif	Zn motif	
H1B.FR.83.HXB2	ASQEVKNWMTETLLVQANPDCCKITLKALGPAATLEEMMTACQGVGGPGHKARVLAEMASQVTSNAT	IMMQRFNRNR	KIVKCFNCGKEGHTARNCRAPRK	GCWCKCGKEGHQMKDCT	ERQ 430
H1A1.UG.85.U455	-T-D-----S-R---G-----	QQTS-----GP-----	RII-----L-K-----		424
H1C.ET.86.ETH2220	-T-D-----D-----R-G-S-----	N-TT-----KS-KGPK	RAI-----L-----		427
H1D.CD.84.84ZR085	-D-----G-----S-----	A-SASAA-----KS-KGT-----			431
H1F1.BE.93.VI850	--D-G-D-----TG-----S-----	ANSA-----KS-KG-----	RV-----I-----	R-----	424
H1G.SE.93.SE6165	--D-G-D-----R-QG-S-----	ASGA-AA-----S-KGP-----	RTI-----L-----		431
H1H.CF.90.056	-T-D-----R-QG-SI-----S-----	TN-A-----K-KG-----	F-----I-----	R-----	431
H1J.SE.93.SE7887	-T-D-----D-----SG-----	TN-----DHK-----	R-----Q-I-K-----		429
H1K.CM.96.MP535	-T-----D-----G-S-----S-I-----	PV-----V-K-KGH-----		I-----	425
H101_AE.TH.90.CM240	-T-----S-----TG-----S-----	HAQHA-----HAQ-----	RI-----R-L-Q-----		428
H102_AG.NG.x.IBNG	-T-----S-R-TG-----G-----	QOAN-----V-----G-----	RTI-----L-K-R-----		425
H1N.CM.95.YBF30	-T-----QL-G-----A-----	QOPT-S-----VFA-----KGI-----	PTI-----L-K-RG-----	Q-----KNE-----G-	435
H1O.BE.87.ANT70	-T-----Q-S-G-----V-T-----	ATAQDLKGGYTAVF-----QNPPIRK	GPI-----I-----	Q-----RN-----GK-	435
H1O.CM.91.MVP5180	-T-----S-Q-----E-----V-T-----	ASAQDLKGGYTAVF-----QNP-RK	GPI-----I-K-----R-	Q-----KN-----G-	435
CPZ.CM.05.SIVcpzEK505	-T-----N-R-----G-S-----A-----	T-AQTATS-----VF-----KGI-----	TI-----L-K-----	Q-----RSG-----	435
CPZ.CM.05.SIVcpzLB7	-T-----D-----S-R-----G-S-----A-----	ANAA-----TV-----KGP	RVI-----I-K-R-----	Q-----R-Q-NM-----GK-	430
CPZ.CM.01.SIVcpzCAM13	-T-----T-D-----Q-----G-S-----A-----	TM-QSQ-RT-----D-FF-K-PGATP	RKI-Y-----L-K-R-----	Q-----S-----G-	437
CPZ.CM.05.SIVcpzMB66	-T-----D-----S-R-----G-----GDEP-F-----S-----	AQH-NDA-----KRO-KGPK	R-----I-K-R-----	Q-----RN-NE-----	426
CPZ.CM.05.SIVcpzMT145	--D-----D-I-----N-R-----G-----	NKT-----EIF-ORG-QN-PPN	KI-----L-R-----	Q-----NONN-----G-	422
CPZ.CD.90.ANT	--P-A-----I-----H-----TG-S-----L-----A-----	ASAN-AQGT-----AVFLORGNNG-GGK	RPL-----K-R-----	R-Q-----L-N-PAT-----NTGK-	455
CPZ.CM.98.CAM3	-T-----D-----N-R-----G-----A-----	LQ-PT-----GVFL-----NGKPT	RKI-----L-R-----	Q-----N-ENG-----	432
CPZ.CM.98.CAM5	-T-----D-----S-R-----G-----A-----	Q-S-----NVF-----NGRVN	KI-Y-----V-R-----	Q-----N-E-TG-----G-	431
CPZ.US.85.CPZUS	--P-T-----D-----R-----G-----V-----A-----	C-MK-PS-----SVFL-K-AGKPG	RKI-----L-R-----	Q-----AG-----N-	433
CPZ.GA.88.GAB1	--P-----D-----Q-----G-----S-----	M-Q-OGRA-----DVFF-K-OGAGPK	RKI-----L-K-R-----	R-Q-----G-----G-	440
CPZ.TZ.01.TAN1	--P-T-----D-----H-----QG-----L-----S-KI-----	ASA-AGG-----VN-L-G-KRPL	RGGGRLQ-----V-----	R-Q-----TRN-----NSTG	461
CPZ.GA.88.GAB2	--D-----Q-----R-----G-----A-----	IQOTS-----S-F-AGG-TPP	RR-----I-K-R-----	Q-----QN-----EG-	434
H2A.GW.x.ALI	TDPA-----Q-----LV-G-MNP-----L-----Q-LM-----LKEAMTP-PIP.FAAA-Q-----		RRTI-W-----S-Q-----Q-----	P-L-AN-P-----	429
H2B.GH.86.D205	TDPA-----Q-----I-----LV-G-MNP-----L-----I-----Q-LM-----LKEALTP-PIP.FAAV-QKAGK-----		RGT-T-W-----Q-----Q-----RQ-----	T-I-SK-P-----	428
H2G.CL.x.ABT96	TDPA-----Q-----I-----LV-G-MNP-----L-----X-----Q-LM-----LKEAL-PTALP.FAAA-QKTGG-----		KRSTI-W-----V-Q-----RQ-----	P-I-AK-P-----	434
H2U.FR.96.12034	TDPA-----Q-----I-----LV-G-MNP-----L-----I-----Q-LM-----LKEALQMPPIP.FAAV-QRGGP-----		RRP-R-W-----Q-K-----RQ-----	E-P-N-AS-P-----TK	433
		p27_p2	p2_p8		
MAC.US.x.239	TDAA-----Q-----I-----LV-G-VNP-----L-----Q-LM-----LKEALAPVIP.FAAA-QRGP-----		R-PI-W-----S-Q-----RQ-----	MD-V-AK-P-----D--	431
STM.US.x.H9	TDPA-----Q-----I-----LV-G-XINP-----L-----X-----Q-X-LM-----LKDALTOGPLP.FXAV-QKGG-----		X-IX-X-X-----S-QF-----RX-----	A-V-AK-P-----	432
STM.US.x.STM	-DPS-----R-----I-----LV-G-MNP-----L-----Q-LM-----LKE-FOPDPLP.FAAA-QQG-----		RRT-W-----A-----KQ-KG-----RQ-----	QO-----AK-P-----	432
MNE.US.x.MNE027	TDPA-----Q-----I-----LV-G-MNP-----L-----Q-KLM-----LKEALAPGLL.PFAAA-QKGP-----		R-PI-W-----S-Q-----RQ-----	QM-V-AK-P-----D--	431
DRL.x.x.FAO	TDAA-----Q-----I-----V-G-MNPS-----LL-----M-----REQOAA-----		QNPP-GPP-----RGPRPPP-NPR-P-QF-----L-Q-TK-----	FR-ALD-MLRN-----PK-	438
MND_1.GA.x.MNDGB1	MND-----M-QQH-----E-----Q-RS-KG-----LE-----Q-LM-M-RT-VGQ-----		SNFVQQRGPGQ-----GPVQPTGRKPI-----N-V-F-LG-KAPRRK-----	N-AMD-KAQ-P-----PKPA-----QQQR	430
MND_2.GA.x.M14	TDPO-----T-----Q-----I-----S-G-MNPS-----LL-----KY-QMM-----QKMQS-----		EV-----NSGGPRG-----PPROPNPR-P-F-VL-D-----	F-DT-M-RN-P-----KM	436
MND_2.CM.98.CM16	TDPO-----T-----Q-----I-----S-G-MNPS-----PL-----KY-QMM-----KEAQS-----		AV-----NSGGPRG-----PPROPNPR-P-F-GL-D-IS-----	F-DL-I-RN-P-----KM	436
MND_2.x.x.5440	TDPO-----T-----Q-----I-----AT-G-MNP-----LL-----KY-QMM-----Q-AQAA-----		V-----NSGGPRG-----PPROPNPR-P-F-VL-D-----R-----	F-DP-L-RN-----PKM	436
TAL.CM.01.8023	TD-A-----Q-----E-A-M-Q-----LQ-----Q-S-LM-----AEALKON-QQ.VIA-V-QGGP-----		KGRGGPRRTPPGQIR-Y-----F-I-K-P-Q-PRGPPG-----	S-F-QM-RAAQ-R-----QP-	456
TAL.CM.00.266	TD-A-----Q-----E-A-M-Q-----LQ-----Q-S-LM-----AEALKON-QQ.VIA-V-QGGP-----		RGRGGPRRTPPGQIR-Y-----F-I-K-P-PROKVPPG-----	T-Y-P-IA-Q-S-----SA-	453
TAN.UG.x.TAN1	TP-D-----I-----LV-G-IHP-----L-----KLMV-----Q-MQG-----		VN-V-GAPRGGG-----RGRGPRR-K-QI-IQKD-PRAGPN-----	K-L-P-LA-----R-----SG-	440
VER.KE.x.TYO1	--G-Q-----S-I-----V-G-MHP-----L-----SY-K-M-M-QTMQ-Q-----		N-V-Q-GPKR-----RPLR-Y-----F-MQ-Q-PE-----T-----	K-L-L-LA-----R-----G-	436
VER.DE.x.AGM3	--G-Q-----S-I-----V-G-MHP-----L-----SY-K-M-M-QMMSQ-----		N-Q-GQ-GRP-----RPP-----Y-----F-MQ-Q-PE-----M-----	R-L-P-LA-----R-----G-	441
VER.KE.x.9063	--G-Q-----S-I-----V-G-MHP-----L-----SY-----M-M-QMMSQ-----		N-V-Q-AAGGV-----RORPP-----Y-----F-MQ-Q-PE-----I-----	K-L-L-LA-----R-----G-	439
VER.KE.x.AGM155	--G-Q-----S-I-----V-G-MHP-----L-----SY-K-M-M-QNLQSQ-----		N-V-Q-GG-GRP-----RPPP-----Y-----F-MQ-Q-PE-----I-----	K-L-P-LA-----R-----G-	437
GRV.ET.x.GRL_677	-P-D-----Q-----I-----L-G-MNP-----LI-----Q-KLMV-M-NGO-----		N-V-V-PQKPG-----RGPL-----	F-MQ-E-K-----Q-----	NG-
SAB.SN.x.SAB1C	TDPA-----Q5-----I-----V-G-MNP-----L-----I-AQ-----LM-----TAAFOQQ-V.GN-FV-Q-ARPRGP-----		LGGRPLNPNII-Y-----P-L-F-K-----RQ-----	SPD-----Q-----K-	455
RCM.NG.x.NG411	-DPA-----Q5-----I-----V-G-MNP-----L-----A-----Q-----Q-----		SM-A-Q-GP-KGP-----PKLGGGPRF-----L-Y-----T-----Y-KTS-----	R-E-L-----P-----KE	436
RCM.GA.x.GAB1	-DPA-----Q5-----I-----MV-G-MNPS-----L-----Q-----Q-----QMMQSN-----		Q-----Q-----Q-----	Q-----Q-----Q-----	K-
SUN.GA.98.L14	G-E-----E-K-KM-----LVI-----ETPS-----R-----S-GKI-----		ASAFRQVQ-----RQA-V-QNLPPRSQGRFVIRGGGPR-PMT-----	P-L-NQ-E-K-G-----PPG-----	S-----M-KQAQ-P-----QK-
MON.NG.x.NG1	TD-A-----Q-----L-M-G-S-LQ-----A-S-L-----ATAT-NMP.M-N-V-GRGG-----		XOPRRQXQIR-Y-----F-V-K-T-----T-----	FR-----XS-N-----PNGGQONPRN	115
MON.CM.99.L1	TD-A-----Q-----L-M-G-S-DLQ-----M-S-I-----AGAT-ANMPNMVQARGPPORRG-----		P-----P-----F-M-K-K-QRR-----	K-YN-QP-LA-----P-----QPPKQNG	440
MUS_1.CM.01.CM1239	SD-A-----S-----E-L-SM-G-LQ-----A-S-LM-----ASALRONS-----QLN-V-G-----		ARGKGSQGGPRNPR-Y-----QF-M-S-PK-KTR-----	K-F-R-LA-Q-R-----SEGAKS	436
MUS_1.CM.01.1085	SD-A-----S-----E-L-SM-G-LQ-----T-S-LM-----LASAFKQ-G-----TLN-V-GAKG-----		ARGRNGSQGNRGNPR-Y-----QY-V-D-PK-RN-----K-----	K-F-R-LA-Q-R-----SDNTKA	440
MUS_2.CM.01.CM2500	-D-A-----Q5-----E-LV-M-G-LQ-----I-SKLM-----ASALRON-IE.TIN-V-G-QGPR-----		GGAGGGPRTPR-----QF-I-KD-PK-VR-----	K-F-R-LA-Q-R-----TNSDKGA	445
MUS_2.CM.01.CM1246	-D-A-----Q5-----E-LV-M-G-K-LQ-----IE-M-S-LM-----TINALRON-IN.TIN-V-----		QSPR-----GVMGKRENSTR-Y-----QF-L-D-PK-KST-----	R-F-----LARQ-R-----TDTGKSA	444
DEB.CM.99.CM40	-D-A-----T-MI-----RI-G-QNP-----LH-----QQ-LM-----ASALKE-GS-----LG-V-QRRGPPGS-----		RRRIQ-----QI-LQKD-KR-----T-----	K-F-Q-----IA-N-GQ-----TPR	462
DEB.CM.99.CM5	-D-A-----T-MI-----RI-G-QNP-----LH-----QQ-LM-----ASALKE-GS-----LG-V-QRRGQGGG-----		RRRLR-----QI-VQ-D-KK-----V-----	K-F-----IA-N-GQ-----VPR	459
GSN.CM.99.CN166	-D-A-----Q5-----E-L-MPG-----N-L-----T-S-L-----ATALKGTSS-----YN-V-K-PPGR-----		G-TP-Y-----QF-L-D-PK-KER-----	K-F-RA-FS-P-R-----GTPIKE	442
GSN.CM.99.CN71	-D-A-----QL-----E-L-MPG-----S-L-----T-S-L-----ATALKG.SS-----YN-V-----PPGR-----		G-TP-Y-----F-I-D-PK-KER-----	K-F-----LA-Q-KT-----GTPAKE	441
DEN.CD.x.CD1	-D-S-L-----TS-----Q-E-----NP-----LQ-----Q-----PT-----LASAFKQKGLCHGSRKGAQPPVER-----		RGLR-----QI-L-KQ-RKP-RVTPQG-----	A-F-----L-----A-N-----R-----SNTNQQG-	470
LST.CD.88.447	G-H-----E-K-KM-----RMVI-----EG-----R-----A-G-I-----ATAMQ-QMR-----QN-V-VTPP-----AQ-GRFVRTGGGGPR-PLT-----		P-----M-Q-QE-----	N-SKE-RFAQ-PKP-----KGK	433
LST.CD.88.485	G-H-----E-K-KM-----RMVI-----EG-S-D-K-----A-GKI-----ATAMQ-QMR-----QN-V-VTPP-----EQ-GRFVRTGGGGPR-PLT-----		P-----M-Q-QE-----	N-SKE-RFAQ-PKP-----KGK	433
LST.KE.x.lho7	G-H-----E-K-KM-----RLVI-----EG-S-K-----A-GKI-----ASATQQVR-----QN-V-VTPL-----AQ-GRFVRTGGGGPR-PLT-----		P-----M-Q-QE-----	N-SKE-RFAQ-PKP-----KPKGK	436
LST.CD.88.524	G-H-----E-K-KM-----MVI-----EG-----K-----A-RKI-----ATAMQ-QMK-----HL-V-QTPP-----AQ-GRFVRTGGGGPR-PLT-----		P-----S-KM-Q-QE-----	N-AKD-R-AQ-PKP-----KGK	431
SYK.KE.x.SYK173	-DPS-----G-L-Q-----I-----E-RQ-----MVK-P-----LQ-----L-KLM-----VMAQO-V-----N-V-GPSK-----		RRMI-L-Y-----QI-MQKD-KK-L-A-----	K-FN-----T-LARA-ROPKRNQGGPPVA-	461
SYK.KE.x.KE51	-DPA-----L-Q5-----I-----E-RS-----MVK-P-----LQ-----QY-KL-----VMTQH-L-----G-I-GPRQGSNP-----		RRGPR-----QL-LQKD-PR-K-L-----	K-FN-GT-IARQ-ROPKRQGGNPPP-	462
COL.CM.x.CGU1	-AG-I-A-ANN-I-I-----R-----G-QK-PS-D-LA-----D-----K-----Q-FQER-N-----		MIEV-TA-----QGI-L-L-M-PKPIGGAGRGRGRGGFRGAPRRPVR-FT-NQ-----MQR-PN-----	K-----	435

Table with columns: p7 end, p1 start, p1 end, p6 start, Vpr binding, Vpr binding p6 end, Gag end. Rows include various HIV strains like H1B.FR.83.HXB2, H1A1.UG.85.U455, etc.

Table with columns: p8 p1, p1 p6, PTAP motif, PSAP in HIV-2 B, U, p6 end Gag end. Rows include various HIV strains like MAC.US.x.239, SMM.US.x.H9, etc.

Table with columns for protein names, protease end_p66, p51 RT start, M41L, K70R D67N, and D110 catalytic site. It lists various HIV sequences and their corresponding amino acid positions across these domains.

	polymerase motif	K219Q T215Y
H1B.FR.83.HXB2	DEDFRKYFTAPTIPINNETPGIRYQNVLPQGWKSPAIQSSMTEKLEPFKQNPDIYVYQYMDLLYVGSLEIQHRTKIEELRHQLLRLWGLTTPDKKHQKEPPFLWMGYELHPDKWTVPQIVLPEKDS..	WTVNDIQIKLVGKLNWASQIYPIGKIVRQKCLLLR.G
H1A1.UG.85.U455	--S-----V-----S-----S-H-----A-----A-----S-FI-----Q-----A-----K-----A	440
H1C.ET.86.ETH2220	--G-----T-----P-----PQ-----AP-E-----AP-E-----K-F-----Q-----A-----A-----A	448
H1D.CD.84.84ZR085	K-----I-----I-----EV-----A-----K-----E-----S-T-----E-----K-----A	442
H1F1.BE.93.VI850	-K-K-----V-----C-----MK-----E-----F-----H-----Q-N-----P-----A	441
H1G.SE.93.SE6165	-----V-----R-----AN-EM-----A-----E-----K-F-----Q-D-E-----TH-----A	438
H1H.CF.90.056	-KE-----A-----E-EM-----A-----A-----K-F-----TVK-----N-----K-----A	440
H1J.SE.93.SE7887	Y-----C-----K-ER-E-----E-----R-----K-----E-----K-F-----Q-----ED-----K-----K-----A	439
H1K.CM.96.MP535	-K-----V-----H-----IK-EM-----P-----E-----K-F-----E-----K-----Q-D-----K-----V	441
H101_AE.TH.90.CM240	--S-----I-----IK-EM-----K-----A-----A-----S-F-----R-----E-----A-----K-----A	446
H102_AG.NG.x.IBNG	-K-----V-----A-----TK-E-----A-----G-----K-F-----VE-----A-----IK-----R-----A	441
H1N.CM.95.YBF30	-K-----T-----EKH-E-I-----LA-----EAV-D-----D-----K-F-----R-----K-----I-----A	448
H1O.BE.87.ANT70	-P-----V-----D-----RD-ELE-C-----PLTE-----KR-----L-----E-----YQ-----F-----S-----Q-----N-----V-----I-----Q-----R-----E-----I-----A	437
H10.CM.91.MVP5180	-P-----V-----V-----D-----S-----EVE-----I-----PLAE-----KRV-----L-----E-----YQ-----F-----Q-----D-----EV-----I-----Q-----R-----KE-----I-----A	437
CPZ.CD.90.ANT	-Q-----V-----C-----A-----A-----DKY-AVE-----M-----TA-----EM-----K-----QV-----E-----Q-----K-----K-----P-----D-----TK-----R-----I-----V-----A	441
CPZ.CM.98.CAM3	--N-----V-----I-----T-----Q-H-LI-----LRE-----E-----V-----L-----T-----F-----K-----K-----IK-----A	446
CPZ.CM.01.SIVcpzCAM13	-K-----V-----A-----AK-----TV-----D-----T-----Q-----V-----H-----F-----K-----EV-----L-----I-----IK-----I-----A	440
CPZ.CM.05.SIVcpzEK505	--N-----V-----T-----EAV-R-----S-----T-----F-----K-----T-----Q-----T-----Q-----T-----K-----I-----A	447
CPZ.CM.05.SIVcpzLB7	--N-----V-----R-----K-----E-----N-----V-----A-----K-----F-----Q-----E-----Q-----IK-----I-----R-----V-----A	444
CPZ.CM.05.SIVcpzMB66	--N-----V-----E-----I-----K-----EL-----E-----V-----A-----K-----F-----Q-----E-----I-----C-----R-----K-----I-----A	444
CPZ.CM.98.CAM5	--N-----V-----V-----X-----D-----H-----LI-----K-----E-----V-----L-----T-----F-----Q-----Q-----EI-----S-----K-----IK-----A	442
CPZ.GA.88.GAB2	--N-----V-----V-----D-----QK-----EL-----SLEE-----KRV-----Q-----E-----F-----Q-----Q-----EN-----IK-----I-----A	444
CPZ.GA.88.GAB1	-K-----V-----S-----EK-----T-----D-----K-----V-----K-----F-----IK-----I-----A	439
CPZ.CM.05.SIVcpzMT145	-P-----F-----TKH-E-I-----P-----EE-----KRV-----A-----K-----F-----Q-----K-----QEV-----Q-----K-----I-----A	440
CPZ.TZ.01.TAN1	-P-----L-----N-----K-----F-----D-----H-----EV-----D-----I-----NEEE-----KL-----KK-----T-----F-----E-----Y-----EK-----M-----N-----T-----PEQ-----H-----H-----IK-----I-----A	437
CPZ.US.85.CPZUS	-KE-----V-----D-----H-----VI-----D-----H-----VI-----NLEK-----E-----V-----L-----Y-----T-----F-----E-----Q-----Q-----EI-----K-----I-----A	443
H2A.DE.x.BEN	H-----Q-----L-----AV-----ME-----K-----I-----K-----YT-----RQV-----A-----VILI-----ILIA-----RTGLE-----DKVVLQ-----KEL-----NGL-----FS-----EE-----F-----D-----Q-----C-----W-----T-----KL-----KLO-----Q-----I-----V-----A-----S-----TKH-----R-----I-----A	466
H2A.SN.x.ST	H-----Q-----L-----AE-----K-----I-----K-----S-----YT-----RQV-----A-----ILI-----ILIA-----RTGLE-----DRVVLQ-----KEL-----NGL-----FS-----EE-----F-----D-----YO-----W-----T-----KL-----R-----Q-----Q-----EV-----V-----A-----T-----N-----R-----I-----A	466
H2B.GH.86.D205	--PN-----Q-----L-----V-----AE-----K-----I-----K-----QS-----C-----Y-----R-----V-----D-----A-----S-----VI-----I-----ILIA-----RSDL-----DRVVSQ-----KEL-----NDM-----FS-----EE-----F-----D-----K-----W-----K-----KL-----K-----Q-----EV-----A-----V-----A-----LF-----T-----HI-----I-----A	467
H2C.CL.x.EHO	--P-----Q-----L-----AV-----AE-----K-----L-----K-----YT-----A-----V-----D-----A-----N-----VT-----I-----IL-----A-----RSDLE-----DRVVSQ-----KEL-----NDM-----FS-----EE-----F-----D-----K-----W-----K-----KL-----K-----Q-----EV-----V-----A-----LF-----T-----HI-----I-----A	466
H2G.CL.x.ABT96	--V-----Q-----L-----V-----AE-----K-----I-----K-----HT-----R-----V-----D-----A-----E-----XIXI-----IL-----A-----RDTLE-----DQVVLQ-----KEL-----NKG-----FS-----EE-----F-----D-----YO-----XW-----K-----KL-----K-----X-----ET-----V-----X-----A-----L-----TKH-----X-----I-----A	449
H2U.FR.96.12034	C-----Q-----L-----AE-----K-----I-----K-----T-----YT-----RNL-----A-----E-----A-----I-----ILIA-----RNDLE-----DQVLSQ-----KEL-----NKT-----FS-----EE-----F-----D-----HK-----W-----K-----KL-----K-----E-----R-----V-----V-----A-----V-----TKH-----IK-----A	463

	DD catalytic site	
MAC.US.x.239	--E-----Q-----L-----V-----AE-----K-----I-----K-----YT-----RHV-----A-----VTLV-----ILIA-----RTDLE-----DRVVLQ-----SKEL-----NSI-----FS-----EE-----F-----D-----Q-----W-----T-----KL-----K-----E-----QRET-----V-----A-----TKH-----R-----I-----A	453
MAC.US.x.EMBL_3	--E-----Q-----L-----V-----AE-----K-----I-----K-----YT-----RHV-----A-----VTLV-----ILIA-----RTDLE-----DRVVLQ-----KEL-----NSI-----FS-----EE-----F-----D-----Q-----W-----T-----KL-----K-----E-----QRET-----V-----A-----TKH-----R-----I-----A	449
SMM.SL.92.SL92B	--PE-----Q-----V-----QE-----K-----I-----K-----AT-----RQV-----A-----VLLV-----LI-----NRGLTE-----DKMVTQ-----DM-----NML-----FS-----ED-----F-----D-----L-----Q-----W-----Y-----K-----KL-----K-----E-----ER-----A-----TKN-----MI-----I-----A	446
STM.US.x.STM	--G-----Q-----L-----V-----AE-----K-----I-----K-----YT-----RN-----RA-----VTLI-----ILIA-----RTDLE-----DRVVLQ-----KEL-----NML-----FS-----EE-----F-----D-----Q-----W-----T-----KL-----K-----E-----QR-----V-----A-----TKH-----I-----A	449
MNE.US.x.MNE027	--E-----Q-----L-----V-----AE-----K-----I-----K-----HT-----RN-----A-----VTLV-----IL-----A-----RDTLE-----DRVVLQ-----KEL-----NSI-----FS-----EE-----F-----D-----Q-----W-----T-----KL-----K-----E-----QET-----V-----A-----TKH-----R-----I-----A	449
DRL.x.x.FAO	-P-----Y-----P-----V-----V-----QG-----K-----M-----C-----GTVASL-----V-----NH-----TVQL-----F-----YTAAE-----EKA-----V-----AL-----MT-----N-----E-----E-----Y-----H-----K-----KIEKVO-----LAEQ-----P-----E-----A-----L-----TK-----I-----A	455
MND_1.GA.x.MNDGB1	-S-----V-----V-----QG-----ATAD-----KT-----KEEY-----EVL-----F-----SASE-----NKMNVR-----E-----RF-----E-----F-----E-----V-----Q-----K-----Q-----ET-----S-----TK-----I-----A	454
MND_2.x.x.M14	-V-----V-----V-----QG-----ATAD-----KA-----KEKH-----EVL-----F-----SASE-----SKM-----V-----E-----QF-----E-----F-----E-----V-----K-----K-----Q-----E-----S-----TKE-----I-----A	454
MND_2.CM.98.CM16	-V-----V-----V-----QG-----ATAD-----Q-----ERH-----V-----F-----RVAPLE-----SRM-----Q-----D-----F-----E-----F-----E-----I-----K-----KVQ-----EE-----S-----TKE-----R-----I-----A	453
TAL.CM.00.266	-P-----S-----Q-----L-----DR-----R-----FK-----T-----HTSDQL-----RK-----DKY-----QVAF-----L-----NESLEE-----RIVG-----AL-----NKV-----IQ-----EA-----F-----DK-----IK-----L-----Y-----K-----E-----D-----L-----T-----EL-----I-----C-----S-----TKA-----RY-----V-----A	450
TAL.CM.01.8023	-P-----AR-----L-----DR-----R-----FK-----T-----YTSDDL-----RK-----EKY-----QVAF-----LI-----ESLPE-----RIVG-----AL-----NKV-----IQ-----EA-----F-----DK-----IK-----L-----Y-----E-----K-----N-----E-----OT-----L-----S-----TKA-----C-----V-----A	451
TAN.UG.x.TAN1	--PE-----P-----V-----QG-----T-----F-----C-----T-----HTAQRV-----KEIG-----KVQ-----LE-----I-----MMI-----PEPE-----DE-----V-----V-----KL-----LR-----FE-----V-----K-----R-----YD-----K-----Y-----SH-----SLNK-----EI-----QEH-----V-----L-----TKA-----A-----Q	461
VER.DE.x.AGM3	--PE-----V-----V-----QG-----T-----F-----C-----T-----NTAAS-----EIK-----ELKPLT-----V-----W-----QEDVY-----DRLV-----Q-----MK-----SA-----E-----V-----V-----K-----YE-----K-----W-----H-----QISS-----E-----ED-----EE-----R-----A-----L-----LRTKN-----I-----K-----A	458
VER.KE.x.9063	--PE-----V-----V-----QG-----V-----F-----C-----T-----NTAA-----KIKELKPLT-----V-----W-----QEDVY-----DKLV-----Q-----VK-----KT-----E-----V-----V-----K-----YE-----K-----W-----Q-----QLNS-----N-----E-----EQ-----A-----L-----LRTK-----I-----R-----I-----K-----A	458
VER.KE.x.AGM155	--PE-----V-----TV-----QG-----F-----C-----T-----NTAS-----EIK-----ELKQIT-----V-----W-----QEGP-----DQVLT-----NR-----QE-----E-----E-----V-----R-----E-----K-----W-----H-----KL-----S-----E-----EK-----EQ-----L-----L-----LRTKNI-----I-----K-----A	461
VER.KE.x.TYO1	--PN-----TV-----QG-----F-----C-----T-----NTAAS-----EIKRNL-----ALT-----V-----W-----QENHT-----DKLV-----Q-----TK-----QA-----E-----E-----M-----YE-----K-----W-----H-----ELSR-----Q-----E-----E-----A-----L-----L-----T-----I-----IT-----GK-----A	461
GRV.ET.x.GRI_677	CKE-----V-----TG-----F-----C-----T-----NTAAN-----EIKRHT-----GLE-----V-----WLA-----HDETR-----NQVDIV-----KM-----EK-----E-----V-----R-----WE-----K-----N-----INK-----E-----PLEG-----E-----K-----V-----V-----TKHT-----AM-----K-----A	464
SAB.SN.x.SAB1C	-P-----Q-----V-----RE-----K-----K-----T-----TTAN-----QE-----QK-----VD-----MLIA-----RPAKE-----LVMVQQ-----DY-----ET-----FK-----E-----F-----D-----Y-----Y-----K-----QL-----E-----T-----REE-----T-----TKH-----R-----I-----A	473
RCM.NG.x.NG411	-Y-----P-----LA-----QG-----E-----V-----V-----TTAN-----E-----V-----F-----RTAHE-----SQM-----KR-----E-----T-----FE-----F-----DK-----E-----V-----E-----K-----L-----EE-----S-----TKE-----I-----A	459
RCM.GA.x.GAB1	-Y-----P-----AV-----QG-----V-----V-----A-----AN-----Q-----EE-----VI-----F-----RTKLE-----DKM-----KQ-----D-----F-----FE-----F-----DK-----Y-----KS-----E-----K-----EE-----S-----LRTKE-----I-----A	455
SUN.GA.98.L14	-P-----Y-----V-----Q-----A-----K-----V-----C-----GTVASL-----K-----R-----Y-----EVQL-----LI-----Y-----KKK-----EIVVKQ-----L-----ME-----N-----E-----E-----Y-----G-----YK-----I-----R-----EIEK-----K-----PLEEEE-----P-----I-----VI-----L-----E-----LRTKE-----I-----A	460
MON.CM.99.L1	-W-----P-----S-----L-----AE-----AK-----FK-----TV-----HTAAELPOEIT-----KY-----EVTLLI-----LI-----MHSLPE-----KIVGVI-----AT-----K-----IO-----PE-----FY-----DQ-----LQ-----SH-----KIAK-----E-----DQEQ-----E-----I-----A-----S-----TKN-----SI-----A	449
MON.NG.x.NG1	--N-----N-----S-----V-----AE-----AV-----FR-----X-----T-----Y-----ASELXKXI-----RY-----EVTLLI-----LI-----HSLPE-----KIVTAT-----TX-----SEK-----IQ-----IE-----Y-----DKY-----LQ-----LQ-----TH-----K-----E-----Q-----Q-----G-----V-----A-----TKH-----XCIX-----X-----A	452
MUS_1.CM.01.CM1239	--PA-----SL-----QV-----QG-----AR-----FK-----T-----YTTAA-----L-----QEI-----EA-----TLI-----LI-----NR-----SG-----RVVAQI-----NM-----KL-----IQ-----QE-----Y-----DY-----VQ-----LQ-----RG-----KI-----VT-----QOEN-----A-----A-----V-----TK-----CI-----V-----A	442
MUS_1.CM.01.1085	--P-----P-----SL-----QV-----LG-----AR-----FK-----T-----FTAA-----L-----EIK-----A-----VTLI-----LI-----NR-----LSS-----KVVAAQI-----NM-----KL-----IQ-----QE-----F-----EDY-----VQ-----LQ-----RG-----KI-----VA-----DQEV-----V-----V-----A-----TK-----CI-----V-----A	442
MUS_2.CM.01.CM2500	-Y-----E-----P-----SV-----V-----LG-----AK-----FK-----T-----ATTA-----L-----QI-----E-----Q-----AL-----V-----LI-----R-----E-----RLVKKV-----DL-----TSK-----IQ-----EE-----QV-----VE-----LQ-----GG-----RIR-----VE-----DOET-----E-----V-----A-----S-----TKH-----CI-----V-----A	446
MUS_2.CM.01.CM1246	--KE-----P-----SV-----AV-----MG-----AK-----FK-----T-----ATTA-----L-----QV-----D-----Q-----AL-----V-----LI-----R-----E-----RLVKKI-----DL-----SSK-----IQ-----EE-----PDY-----VE-----LQ-----KG-----RIK-----VE-----DQ-----T-----E-----A-----V-----S-----TKH-----RCI-----I-----A	447
DEB.CM.99.CM40	--PT-----A-----V-----V-----IQ-----E-----EFR-----SA-----A-----VGRQ-----QI-----EK-----ELI-----V-----L-----RKD-----LQ-----VAQ-----F-----ER-----R-----PE-----Y-----ED-----H-----K-----RLM-----VT-----DEEE-----HK-----Q-----S-----TKH-----RAI-----VP-----A	447
DEB.CM.99.CM5	--PQ-----AR-----V-----IQ-----E-----EFR-----SA-----ATVGRQ-----KL-----KEYK-----LI-----V-----L-----NDKVS-----LEQVQ-----KF-----EK-----R-----PE-----Y-----ED-----YH-----K-----HLL-----E-----EEE-----HK-----Q-----S-----TKH-----AT-----AR-----A	447
GSN.CM.99.CN166	Y-----E-----SV-----V-----Q-----A-----AE-----FK-----A-----T-----NIVASL-----ROI-----D-----E-----V-----LI-----LI-----RKLE-----QVVKI-----ML-----TSYNIO-----EA-----PDY-----VK-----L-----KG-----RL-----VE-----DQ-----T-----V-----H-----RTK-----CI-----A-----A	452
GSN.CM.99.CN71	-Y-----E-----SV-----V-----Q-----AE-----FK-----A-----T-----ATVASL-----YOI-----D-----E-----V-----I-----LI-----RKLA-----QVVKI-----ML-----TSYNIO-----EA-----PDY-----VR-----L-----KG-----RL-----VE-----DQ-----T-----V-----H-----RTK-----RCI-----A-----A	445
DEN.CD.x.CD1	-P-----P-----AQ-----V-----G-----Q-----IFK-----V-----T-----ATAD-----M-----SQ-----AH-----VTLI-----L-----A-----RSLE-----KMKV-----DF-----AV-----FE-----A-----F-----ET-----IQ-----KH-----QI-----K-----D-----LPEDGQ-----A-----V-----L-----S-----TK-----SI-----V-----A	469
LST.CD.88.447	-P-----Y-----Q-----V-----QE-----R-----R-----C-----ATVAGL-----SE-----L-----L-----TL-----LI-----R-----DDKKG-----DQAVK-----DL-----VT-----N-----K-----E-----F-----A-----YH-----V-----R-----EIEKTK-----DI-----LOKT-----N-----V-----V-----A-----L-----D-----RTKE-----I-----V-----A	459
LST.CD.88.447	-P-----Y-----Q-----V-----QE-----R-----R-----C-----ATVAGL-----SE-----L-----L-----TL-----LI-----R-----DDKKG-----DQAVK-----EL-----MT-----N-----E-----E-----F-----A-----YH-----V-----R-----EIEKVK-----DI-----LOKT-----N-----V-----V-----A-----L-----D-----RTKE-----I-----V-----A	459
LST.CD.88.524	-P-----Y-----Q-----V-----L-----QE-----K-----R-----C-----ETVAGL-----AE-----L-----L-----LL-----LI-----R-----KEG-----SQAVK-----EL-----IT-----N-----E-----E-----F-----DK-----YH-----V-----R-----EIERVK-----I-----TRKE-----E-----V-----V-----A-----L-----D-----RTKE-----I-----V-----A	460
LST.KE.x.lho7	-PNY-----K-----L-----QE-----K-----R-----C-----TVAGL-----SE-----L-----L-----MI-----FI-----R-----RKG-----DQAVK-----EL-----MT-----N-----E-----E-----F-----A-----YH-----V-----R-----EIEK-----K-----M-----LTKT-----Q-----V-----V-----A-----L-----D-----RTKE-----I-----V-----A	459
SYK.KE.x.SYK173	-KE-----P-----P-----V-----AS-----E-----FT-----T-----T-----T-----INO-----Q-----KYS-----LTLI-----LI-----T-----RSEKA-----QEIYQVIVTA-----KV-----FKV-----KE-----W-----DOY-----MOL-----T-----V-----QL-----K-----E-----NI-----DE-----I-----Q-----L-----I-----V-----S-----TK-----CI-----V-----A	474
SYK.KE.x.KE51	--E-----Q-----V-----SS-----E-----FK-----T-----ATVNO-----Q-----I-----EK-----S-----I-----V-----L-----NRSEK-----GOIVS-----IVKS-----AV-----FSI-----PE-----W-----DKF-----LQ-----L-----T-----SL-----KVQ-----ITES-----P-----EL-----II-----V-----CKA-----S-----CI-----A	480
COL.CM.x.CGU1	-P-----YQ-----V-----LA-----T-----V-----K-----T-----GTVAR-----RQQL-----Q-----A-----L-----P-----LEE-----RAVQ-----DA-----YY-----E-----E-----Y-----AD-----YK-----KQ-----KL-----KVKI-----Q-----K-----V-----A-----L-----L-----RTKEIS-----CTK-----K-----A	422

H1B.FR.83.HXB2
H1A1.UG.85.U455
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.V1850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H1O1.AE.TH.90.CM240
H1O2.AG.NG.x.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.90.ANT
CPZ.CM.98.CAM3
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.98.CAM5
CPZ.GA.88.GAB2
CPZ.GA.88.GAB1
CPZ.CM.05.SIVcpzMT145
CPZ.TZ.01.TAN1
CPZ.US.85.CPZUS
H2A.DE.x.BEN
H2A.SN.x.ST
H2B.GH.86.D205
H2B.CL.x.EHO
H2G.CL.x.ABT96
H2U.FR.96.12034

MAC.US.x.239
MAC.US.x.EMB_L_3
SMM.SL.92.SL92B
STM.US.x.STM
MNE.US.x.MNE027
DRL.x.x.FAO
MND_1.GA.x.MNDGB1
MND_2.GA.x.M14
MND_2.x.x.5440
MND_2.CM.98.CM16
TAL.CM.00.266
TAL.CM.x.18023
TAN.UG.x.TAN1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TYO1
GRV.ET.x.GRI_677
SAB.SN.x.SAB1C
RCM.NG.x.NG411
RCM.GA.x.GAB1
SUN.GA.98.L14
MON.CM.99.L1
MON.NG.x.NG1
MUS_1.CM.01.CM1239
MUS_1.CM.01.1085
MUS_2.CM.01.CM2500
MUS_2.CM.01.CM1246
DEB.CM.99.CM40
DEB.CM.99.CM5
GSN.CM.99.CN166
GSN.CM.99.CN71
DEN.CD.x.CD1
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.x.lho7
SYK.KE.x.SYK173
SYK.KE.x.KE51
COL.CM.x.CGU1

ALTEVIPLTEEALELAENREILKPEVHGYYDPSKDLIAETIQKGGQWYQIQYQEP.FKNLKTGYKARMRGAHTNDVQKLEAVQKITTESIIVIGKTK.PKFKLPIQKETWETWTEYQWATWIPEWEFVNTPLVKLWYQLEKEPIVGAETFFYDGAANRETKLGA
---DIVT---D-----V-----D-----K-S-----V-S-----I-----R-----A-M-----D-A-----609
---DIVT---F-----ND---F-F-----F-KRGT---AV---AL-----R-----A-D-----A-V-----I-----610
---M-----L-----A-----I-----R-----I-----D-----I-----K-----610
---DIV---A---K---R-----D-----N-----KV-S-----AL-----RS-----L-----D-----D-----T-----A-----D-----S-----K-----610
---DIVS---A---M-----R-----E---V---LD-----Y-----KRG5-----V---A-----R-----I-----D-----R-----T-----P-----Y-----606
---DI---K-----R---I-----R-----K-T---I-----S-----I-----R-----I-----H-----T-----A-----Y---I-----608
---DIV---R-----K-----SA-E---V---LD-----H-----KR-S-----A-V---AL---A-----R-----R-----D-----M-----S-----T-----607
---DIV---A-----T-----V---V---D-----H-----RGS-----R---V---VA-----R-----R-----M-----D-----S-----614
---DIVA-----T---V---L---D-----KK-S-----V---VAM-----R-----R-----M-----D-----S-----I-----609
---VNF-----L-----G-E-V-----L-H-----K-S-----I---V-V-R-VA-----R---V---V---A---DH-----T---S-----616
S---V---SR-----E---R---Q---Q-D---WVN---GE-----E-H---T-QKAS---IR---A-VI---VSO---I---L---VTR-----AD-----S---I---R---S---M---Y-----605
S---V---SK-----E---K---Q-D---WVS---H-E---V---DE-H---QKAS---IR---A-V---VSO---A---L---R---VTR-----A-----S---I---T---EDVY-----N-----605
S---DRVQM-R-----E-KQ---QKQIE-Y-Q-GLP-K-T---S-----NE-G-L-A---PT-T---E-R---AGV---GL---EV---Q---TR---DA---SD-----IR---N-LAD---PE-----NSQ-----609
---DIVTM-Q-M-E-X---XD-----E-E---X---NS---F-Q-H---Q-S---IR---A-V---A-----R---V---V---S-----D-X-----T---P-----K-----614
R---D---V---P-----VST---N-D-E-V---E-----F-Q-HR---Q-ST---IR---A-VI---AV-----R-R-V---S---A-----I-----S---TD---PT-D-Y-----K-----608
---VT---R-----A---N-D-E-----DL-H---K-ST---IR---V---VAL-----R---V---V-----D-----T---S---Y-----615
NS---VTF-R-----N---A---E---V---S---Q---Q6S-----A-VM---AI---I---R---A---A---D-----Y---SH---F---Q---P-----612
---VTF-T-----A---E---V---F-Q-Y-----S-----V---VAL-----V---R-R-----A---D-----Y-----QD---P-----612
---VAM-Q---M-E---D-----E-E---V---NS---F-Q-H---Q-S---IR---A-V---A-----R---V---V---S-----D-----P---P-----K-----610
---IVEM-R-----E-----T---E-E---V---NN---F-T-GR---Q-A-P---IR---A-V-P-VSP---IL-TV-----R-V-R-V---A---D-----R-----Q---Y-----RR---612
K---D---V---P-----VST---D-E-E---NC---F---H---Q-S---IR---A---A-----R---V---S---A---A-----I-----S---T---PTTD---Y-----T-----607
G---VA---P-----Q-----A-E-E-----F---H---GKQ-A---IR---A-T---A-----R---V---M---A---AD---V---I-----T---E---Y-----R-----608
G---PVEM-R-----E-KQ---K-Q-A---KLP-Q-A-----E-G---KSP-T---EIR-AGLI---GN---I---TV---L---VS---SQ-D---V---V---I---IR---N-LSD-PE-----DS-K-R-605
---VNF-H---M-E-----E-V---V---RS---F-R-H---Q-S---IR---V---V-----R---V---V---S-----D-----K-T-A-E-----Q---K-----611
T---EVQW-L-A-A-E-KI---SQEQE-Y-QEE-E-E-T---SQGH---K-H-E---I---V---KIKNT---G-RL-AHV---GK-AL---RI---H-VERD---Q-DN---V---D-S---R-TFN-VGD-P-----T-SC-QS-E---633
T---EVQW-L-A-A-E-KI---SQEQE-C-QEE-E-E-TV---DDDN---K-H-GG---I---V---KVKNT---G-RL-AHV---GK-AL---RI---H-VERD---Q-DN---V---D-S---R-TFN-VGD-P-----T-SC-QS-E---633
T---EVQW-L-A-A-Q-KI---EQEQE-S-KERP-E-TV-NLAN---K-H-GN---V---V---KVKNT---G-RL-AHV---GK-AL---EI---V-H-VERD---D-D-V---D-S---IR-A-N-V-D-LE-R-Y-T-SC-TS-E---634
T---EVQW-L-A-A-FQ-KI---EQEQE-S-KERP-E-TV-NLAN---K-H-GD---I---V---KVKNT---G-RL-AHV---GK-AL---EI---M-H-VERD---D-D-V---D-S---IR-A-N-V-D-LE-V-Y-T-SC-KAS-E---633
T---EVQW-L-A-A-Q-KI---XQEQE-A-KEXEPK-TV-NLDN---KVH-GN---RI---V---KIKNT---G-RL-AHV---X-GK-AL---L-F-H-VERD---Q-D-D-V---D-D-S---H-R-V-N-V---LEQT-Y-T-SC-KTS-X---616
T---EVQW-M-A-A-F---KI---SQEQE-A-REEN-E-TVL-NQDN---K-H-GD---RT---V---F-KVKNT---G-RL-AHV---GK-A-----I---M-H-VER-V-Q---D---V---VLD-D-S---R-A-N-V---LL-E-Y-T-SC-AS-Q---629

p51 RT end_p15 RNase H start
T---EVQW-M-A-A-YE-KI---SQEQE-C-QEG-P-E-TVI-SQDN-S-K-H-D---I---V---F-KIKNT---G-RL-AHV---GK-A---QV---H-VE-DV-Q---D---V---D-IS---R-VFN-V-D-E-E-Y-T-SC-QS-E---620
T---AVQW-M-A-A-YE-NI---SQEQE-C-QEG-P-E-TVI-SQDN-S-K-H-D---I---V---F-KIKNT---G-RL-AHV---GK-A---QV---H-VERD-V-Q---D---V---D-IS---R-VFN-V-D-EVE-Y-T-SC-QS-E---616
T---EVQW-L-A-A---I---NQEQE-R-REDEP-E-TVL-NQDN-S-K-H-GD---RI---V---F-KIKNT---G-RL-AHV---GK-A---QI---F-H-VER-V-DQ---D---D-IS---IR-VFN-V-D-EKE-VY-I-SC-NS-E---613
T---EVQW-M-A-A-Y---KI---SQEQE-Q-RED-P-E-TVV-DQDN-S-K-H-D---I---V---F-KIKNT---G-RL-AHV---GK-A---QI---H-VER-V-Q-AD---V---D-S---R-VFN-V-D-LE-T---SC-QS-E---616
T---EVQW-M-A-A-YE-KI---SQEQE-C-QEG-P-E-TVI-NQDN-S-K-H-D---I---V---F-KIKNT---G-RL-AHV---GK-A---QV---H-VE-DV-Q---D---V---D-IS---R-VFN-V-D-E-E-Y---SC-QS-E---616
P-D-EVEW-R---YE-KL-I-QM---Q-E-P-K-K-L-N-S---E-DD-N-P---KVNKS---MRM-AGL---AK-AL---RL---I-Y-VER-V-Q-Q---V---D---IS---H-I---N-LSD-P-E-VY---VS-E---623
NI---KVTM---R---YE-Q-K---A-EQE-S---N-E-YVRF---TTG-DISF-WK-GN---V-RA---GKQKT---S---LKM-AG-T-VGR---FV---MQI-TTR-I-D-H---C---V---IS---M-ERE-S-SP---LE-V-Y---DS-M---618
P-D-VIEW-R---YE-KL-VQ-E---Q-G-P-M-KV-LT---S---E-E-N-P-V---TKH---ELRV-AGL---AK-C---EL---Y-LER-V-DQ-HD---D---IS---IR---N-L---P-EDVY---NS-E---622
P-DKIVEW-R---YE-KL-VQ-E---Q-E-P-M-KV-LT---S---E-E-N-P-V---TKN---ELRV-AGL---VAK-CL---L---Y-LER-V-DQ-HD---V---D-IS---IR---N-L-D-P-EDVY---NS-E---622
P-D-KVEW-R---YE-KL-VQ-E---Q-E-P-M-KV-LT---S---E-E-N-P-V---TKN---ELRV-AGL---AK-AL---RL---I-Y-ER-V-DQ-P-----S-H-IG---N-LR-VQ-EDVY---NS-E---621
G-L-EVQ-S-A---A-E-T---QEQEG-E-LT-LAE---G-M-K-GK---RL---F-KQ-T-S-PYQ---IG-M---GK---F-QV---V-R-VV-E-Q---D---H-CS---TIVPIH---R-N-VS-E---D-Y---H-S-E---617
G-L-EVT---A---A-A-E-T-S---KEGEP-E-LT-LAE---G-MVK-GK---RL---F-KQ-T-S-PYQ---VG-M---GK-A---F-QV---I-R-VV-E-Q---S---CS-L-TIVPIH---R-N-VS-E---D---H-S-E---618
D-A-VQW-Q-V-FRQ-Q---EQQ-A---EQ-S---S-R-T-LDD---G-FK-G---V---TKSKNT---EFRV-AGM-RVCK-A-TF---L-IMEV-V-Q-Q-SD---V-L-I-IS---I---N-V-D-PK-KVY-C---C-NS-E---628
N-L-TVTW---A-Y---K---TEQE-T-K-GRPIR-AV-LEG---S-FK-G---QV-V---TKQKNT---EFRV-AGL---LCK-L---EL-VLE---ER-V-Q-AD---VS---D---S---T---PKEDVY---C-NS-E---625
N-L-EVQW---A-Y---K---TEQE-T-A-G-P-R-AV-LEG---S-FK-G---I---V---KQKNT---ELRT-AGL---CK-AL---RL---I-E-VER-V-Q-AD---VS---SV---T---P-EDVY---C-NS-E---625
N-LD-VIEW-P---A-YE-K---TEQE-T-A-E-P-R-AV-L-D---S-FK-G---I---V---F-KQKAT---ELRV-AGV---GK-AL---QL-T-E-VERD-Q-AD---VS---D-SV---T---T---P-EDVY---C-QS-E---628
N-L-LVAV-P---A-Y-A---TEQE-T-K-GIPR-AV-LEG---S-FK-G---QV-V---TKQKNT---ELRT-AGL---CK-AL---IL-VLE---ER-V-Q-AD---VS---D-S---L---T---P-EDVY---C-NS-E---627
N-L-E-VW---A-YKN-QG-VQ-TQE-T---L-E-TV---E---FT---G-AV-V-R---KQ-ET---LRT-AHL---CK-ALT---RL---RVQ-V-D-K-DM-QD---VS---S---L---S-V---K-EDVY---SKV---631
P---IVQW---E---Q---RQKQ-Q---ALP-R-KVL-L-D-G---PE-N-I---V---KIKT---ELRM-AGL---GK---QI---IME-VER-L-Q-SD---V---M-S---Q-IR---K-V-D-P-EAVY---NS-E---641
---DDKVEF-R---YE-KAV---KI---N-E-P---KV-LEQ-E-G-MFS-G---R---M-SA-F-KT-N-I-SYQ-FAD-LT-VGR-Q-T-VPD-RI-V-Q-DA-MN---TI-AIH-H-LRQ-T-VS-LED-V-Y---TS---610
---D-KVEM-K---I-YE-KM---KL---EK-P-V-N-LEG-S---E-S-G-P---KQKT---EIRM-AGL---GR---A---RL-T-R-ER-V-D-RSQ-V---D---S---IR-G-N-V-D-P-E-VY---NS-I---622
P-A-KVKM---RE-YQS-Q-V-Q-S-S-E-D-E-CRV-VK-IL-F-WL-G---KQV-RV-R-QKGA-E-PCQ-AA-L---AK---FV---IQV---R-I-SQ-AD---C---L---S-K-EQE-T-AT-VP-D-Y---EKLE-R---626
S---DTVV-SDL-QA-----EST-G---NEP-KV-LTSLAEN-G-RFF-K---SV-----F-KI-ST-S-TYQ-AD-LA-LGK-A-T-RL---I-R-VV-Q-DA-ADN-IN-V-DI-A-Y-S-H-LRQ-T-VQD-E-P-Y---TS-E---616
---DIVE-SDL-QA-----SQST-G---KL-VY-TSLXE---G-XYF---F-KIKN---I-R-VN-Q-D-ADN-V---I-P-HS-H-LRQ-T-VQX-EH-P-Y---AS-E---619
S---EVO-Q-Q-Q-Q-Q-Q-A-F-EQP-V-VSL-DS-G-NFL-NK---GI-S---F-KV-SV---SYQ-SD-IAR-GR-L---PPE-VRI-VI-Q-DQ---H-VS---DI-A-H-TH-LRQ-FT-VP-LSE-P-Y---HKVS---610
S---EVO-Q-Q-Q-Q-Q-Q-Q-Q-Q-Q-A-E-EQP-V-VSL-DS-G-NFL-SK---GV-R---F-KV-SV---SYQ-AD-MA-GR-L---PPD-VRI-VI-Q-DQ-IAH-VS---TI-A-H-TH-L-Q-FT-VP-LKE-P-Y---VS---610
---TVE-Q-Q-Q-Q-Q-Q-K-A---E-P-L-VSL-EQ-G-TFT-K---NM---F-KV-T-I-TYQ-A-LA-AAR-L-C---PD-CRI-VV-Q-DH-A-S-T-V-I-A-H-H-LRQ-FT-VP-LEA-P-Y---NS---R---614
G---LVE-S-Q-Q-Q-Q-Q-Q-EAG-A---E-P-VL-VSL-EQ-G-TFT-DR-NM-R---F-KI-T-S-PYQ-A-L-SRASK-AL-C---PD-CRI-VV-Q-DN-ADS-T---DI-A-H-Y-LRQ-FT-VP-ET-P-Y---NS---R---615
G---DPVT-S-QA-R-K---Q-E-S---KEQEP---LS-L-K---G-V-R-PK---GI---FS-DK---Y---FH---AK-MY---G---Y-FRI-Q-R-VV-E-DN-HNH---A-D-AIH-H---E-VS---PD-D-Y---S---614
G---DKVS-S-QA-R-K-TP-QE-S---KEDEP-S---S-I-R---G-V-R-TK---GL---FSKDK-T-Y-FH---AK-MY---G---Y-FRI-Q-R-VV-E-DN-HNH---A-D-AIH-H---E-VS---PD-D-Y---S---M---610
G---DQVE---QI-----QSSE-G---AE-P-VV-TSL-EQ-G-MFS-G---R---M-SA-F-KT-N-I-SYQ-FAD-LT-VGR-Q-T-VPD-RI-V-Q-DA-MN---TI-AIH-H-LRQ-T-VS-LED-V-Y---TS---624
G---QVE---QI-----QSSE-S---AE-P-VV-TSL-EQ-G-MFS-DS---M-RS-F-KT-T-I-SYQ-AD-MT-VGR-Q-T-VPD-RI-V-Q-D-MN---TI-A-H-H-LRQ-T-VS-LEN-V-Y---TS---613
D-G-T-S---G-T-A-K-AQ---SIS-S-R-E-D-E-CRV-VK-IL-F-WL-G---HL---HKDK-S-Y-PYQ-AKQMA-GR-AL---RVL-T-R-VS-VQ-DA-S---VN-D-M-S---LRNF-N-VS---DK-P-Y-T---Q-Q-R---636
---D-VTW-DA-LE-YEQ-Q---KLQ-A---T-E-VVRV-QSKK-II-F-WR-GN---SI-RA-R-Q-QKS---PLQK-A---GK---HV---IQV-VTR-V-DQ-SDH-V---L-IS-Q-EQE-S-A---E-V-D-Y---EKVG-T---626
---E-VTW-DA-LE-YEQ-Q---KLQ-A---T-E-VVRV-QNKK-VI-F-WR-GN---SI-RA-R-Q-QKA---PLQK-A---GK---HV---IQV-VTR-V-DQ-SDH-V---L-IS-Q-EQE-S-A---E-V-D-Y---EKVG-T---626
P-EQLVTW---LE-YET-KT---K-Q-A---E-VVRV-QNKK-VI-F-WR-GN---SI-RA-R-Q-QKS-M-PLQ-A---GK---FV---IQV-VTR-V---N-INF-V---DL-IS-Q-EQE-S-A---E-V-D-Y---E-VG-T---627
P-E-I-NW---LE-YGQ-K-V-KM-Q-A---E-E-VRV-QNKK-II-F-WR-GN---SI-RA-R-Q-QKA---PLQK-V-I---GK---FV---IQV-VTR-V-H-SDH-V---DL-IS-Q-EQE-IW-A---I-V-D-Y---EKVG-T---626
Q---LT---A-A-E---Q---EQA-S---K-P-E-H-T-L-SQ-G-M-K-QKQPP-I---T-KTFA-S-YQS-AQLLN-GIQ-LWY-V---T-H-VKR-E-K---D---V-LK-IS---RMY-N-VP---PE-V---DS-T-N---643
---DIVE---A-M-----VEQT-S-Q-E-P-E-H-S-L-Q-G-I-K-G-QE-P-I---TGKAYAT-H-YQA-AQLMN-GTQ-LLY---I-E-H-VKR-E-K---DH---L-VKC-H-I-RMY-N-VS-V-E-----NS-E-Q---648
H-L-EVQ-SA---E-QD-KQ-I-E-Q---I-HE-IWVD-FRL-K-G---G-AVC-K---HGT-RR---HNTGKNQ-F-SMQE-ASVI---GR-A-T---V---M-V-AKR-D-Q-SD---SA---I---SSSYV---VMN-VL---LKE-P-YWT-GCS-KLGA---589

Table with 2 columns: Accession ID and Sequence. Accession IDs include H1B.FR.83.HXB2, H1A1.UG.85.U455, H1C.ET.86.ETH2220, H1D.CD.84.84ZR085, H1F1.BE.93.VI850, H1G.SE.93.SE6165, H1H.CF90.056, H1J.SE.93.SE7887, H1K.CM.96.MP535, H101.AE.TH.90.CM240, H102.AG.NG.x.IBNG, H1N.CM.95.YBF30, H1O.BE.87.ANT70, H1O.CM.91.MVP5180, CPZ.CD.90.ANT, CPZ.CM.98.CAM3, CPZ.CM.01.SIVcpzCAM13, CPZ.CM.05.SIVcpzEK505, CPZ.CM.05.SIVcpzLB7, CPZ.CM.05.SIVcpzMB66, CPZ.CM.98.CAM5, CPZ.GA.88.GAB2, CPZ.GA.88.GAB1, CPZ.CM.05.SIVcpzMT145, CPZ.TZ.01.TAN1, CPZ.US.85.CPZUS, H2A.DE.x.BEN, H2A.SN.x.ST, H2B.GH.86.D205, H2B.CL.x.EHO, H2G.CL.x.ABT96, H2U.FR.96.12034, MAC.US.x.239, MAC.US.x.EMBL_3, SMM.SL.92.SL92B, STM.US.x.STM, MNE.US.x.MNE027, DRL.x.x.FAO, MND_1.GA.x.MNDGB1, MND_2.GA.x.M14, MND_2.x.x.5440, MND_2.CM.98.CM16, TAL.CM.00.266, TAL.CM.01.8023, TAN.UG.x.TAN1, VER.DE.x.AGM3, VER.KE.x.9063, VER.KE.x.AGM155, VER.KE.x.TYO1, GRV.ET.x.GRI_677, SAB.SN.x.SABIC, RCM.NG.x.NG411, RCM.GA.x.GAB1, SUN.GA.98.L14, MON.CM.99.L1, MON.NG.x.NG1, MUS_1.CM.01.CM1239, MUS_1.CM.01.1085, MUS_2.CM.01.CM2500, MUS_2.CM.01.CM1246, DEB.CM.99.CM40, DEB.CM.99.CM5, GSN.CM.99.CM166, GSN.CM.99.CM71, DEN.CD.x.CD1, LST.CD.88.447, LST.CD.88.485, LST.CD.88.524, LST.KE.x.lho7, SYK.KE.x.SYK173, SYK.KE.x.KE51, COL.CM.x.CGU1, WQLDCTHLEGGKVLVAVHVASGYIEAEVIPAETGQETAYFLLLKLAGRPVKTI...HTDNGSNFTGATVRAACWVWAGIKQEFIPYNPQSGVSVSMNKLKIKIGQVRDQAEHLKTAVMQAVFIHNRKFRKGGIGGYSAGERIVDIIATDIQTKELQKQI...TKIQNFRVYYR

H1B.FR.83.HXB2
 H1A1.UG.85.U455
 H1C.ET.86.ETH2220
 H1D.CD.84.84ZR085
 H1F1.BE.93.VI850
 H1G.SE.93.SE6165
 H1H.CF.90.056
 H1J.SE.93.SE7887
 H1K.CM.96.MP535
 H101_AE.TH.90.CM240
 H102_AG.NG.x.IBNG
 H1N.CM.95.YBF30
 H1O.BE.87.ANT70
 H10.CM.91.MVP5180
 CPZ.CD.90.ANT
 CPZ.CM.98.CAM3
 CPZ.CM.01.SIVcpzCAM13
 CPZ.CM.05.SIVcpzEK505
 CPZ.CM.05.SIVcpzLB7
 CPZ.CM.05.SIVcpzMB66
 CPZ.CM.98.CAM5
 CPZ.GA.88.GAB2
 CPZ.GA.88.GAB1
 CPZ.CM.05.SIVcpzMT145
 CPZ.TZ.01.TAN1
 CPZ.US.85.CPZUS
 H2A.DE.x.BEN
 H2A.SN.x.ST
 H2B.GH.86.D205
 H2B.CL.x.EHO
 H2G.CL.x.ABT96
 H2U.FR.96.12034

MAC.US.x.239
 MAC.US.x.EMBL_3
 SMM.SL.92.SL92B
 STM.US.x.STM
 MNE.US.x.MNE027
 DRL.x.x.FAO
 MND_1.GA.x.MNDGB1
 MND_2.GA.x.M14
 MND_2.x.x.5440
 MND_2.CM.98.CM16
 TAL.CM.00.266
 TAL.CM.01.8023
 TAN.UG.x.TAN1
 VER.DE.x.AGM3
 VER.KE.x.9063
 VER.KE.x.AGM155
 VER.KE.x.TYO1
 GRV.ET.x.GRI_677
 SAB.SN.x.SAB1C
 RCM.NG.x.NG411
 RCM.GA.x.GAB1
 SUN.GA.98.L14
 MON.CM.99.L1
 MON.NG.x.NG1
 MUS_1.CM.01.CM1239
 MUS_1.CM.01.1085
 MUS_2.CM.01.CM2500
 MUS_2.CM.01.CM1246
 DEB.CM.99.CM40
 DEB.CM.99.CM5
 GSN.CM.99.CN166
 GSN.CM.99.CN71
 DEN.CD.x.CD1
 LST.CD.88.447
 LST.CD.88.485
 LST.CD.88.524
 LST.KE.x.lho7
 SYK.KE.x.SYK173
 SYK.KE.x.KE51
 COL.CM.x.CGU1

p31 Integrase end
 Pol end

DSRNPLWKGPALKLWKGEAVVIQDNS.DIKVVPRRKAIIIRDYQKMGAGDDCVASRQDED.....*
 --D-I-----M-G-----
 --D-I-----A-G-----
 --D-I-----
 --V-----E-I-----V-G-----
 --D-V-----N-E-----G-----
 --D-I-----E-E-----
 --D-I-----P-E-----G-----
 --E-I-----E-----G-----
 --D-I-----G-----
 --D-I-----G-----
 --D-I-----G-----G-----NOEME.....
 --D-I-----Q-KG-----E-T-SM-G-T-SESVEQPSEIP.....
 --D-I-----Q-KG-----T-SM-N-T-SESMEQPGEIP.....
 --D-V-----Q-K-QE-E-----KE-R-KIEDR-DL-G-N-----
 --D-I-----R-KERE.EV-I-----SM-GG-----SQGLE.....
 --D-I-----T-QG-E-----
 --E-T-----G-----G-----SODME.....
 -R-D-I-----L-QE-E-----E-AN-L-D-----N-----
 --D-I-----L-QE-E-----NONME.....
 --D-I-----R-KE-E.EV-----GSM-G-N-QNLE.....
 --D-I-----S-G-----KH-V-----L-GG-----NQN.....
 --D-I-----T-QG.EL-----N-----
 -KD-I-----QD-E-----A-S-----S-----
 -A-D-I-----R-KEGE-----KE-AGGMDD-N-T-----
 --D-I-----R-KE-E.EV-----K-A-SM-G-----S-----
 EG-DQ-----GE-----IVKVG.T-----GRQEL-SSPHLEGAREDDGEMACPCQVPEIX..
 EG-DQ-----D-IVKVG.A-----II-K-GRQEM-SGSNLEGAREDDGEVA.....
 EG-DQ-----GE-----I-KVGT.E-----H-GGKGL-CSADMEDTQROAREMAQSD.....
 EG-DQ-----GD-----I-KVGT.E-----I-N-GGKEL-CSADVETMQAREVAQSN.....
 EG-DQ-----GE-----I-KVGT.E-----K-GGKEV-SSTNMDTROTGEVA.....
 EG-DQ-----D-L-KVGT-----I-GGKEL-SSNLEGA-KVREMA.LPDQTPXV.X

p31 Integrase end
 Pol end

EG-DQ-----GE-----ILKVG.T-----K-GGKEV-SSSHMEDTGEAREVA.....
 EG-DQ-----GE-----ILKVG.T-----K-GGKEV-SSSHMEDTGEAREVA.....
 EG-DQ-----E-----KVGT.E-----K-GRQEMGSSASMEDQ.....
 EG-DQ-----GE-----IVKVG.T-----K-GGKEV-SGSHLEDTREAQEA.....
 EG-DQ-----GE-----ILKVG.T-----K-GGKEV-SSSHMEDTGEAREVA.....
 EG-DQ-----E-V-----KEGT.-L-I-----K-KNV-SHTNMES.....
 EG-DQ-----GI-----LKYQE.E-I-----C-K-ESGKNSQ-NLESV.....
 EG-EQ-----E-I-----KEGT.-L-----KTV-SNPHMEDRQETA.....
 EG-DQ-----E-I-----KEGV.-L-----KTV-SNPHMGR.....
 EG-DQ-----E-I-----KEGT.-L-----KTV-SDPNVEA.....
 KGA---Q---H-I-----LRTDEGEVIT-----KP-QALGNKIDLE-SKEQ-AEMGRDN.....
 KGA---Q---H-V-----LRTDEGEVIT-----KP-QA-GNKTDLEGSKEQ-AEMGRDN.....
 EG-E-V-----T-----KEGE.EL-----KE-ERKTVG-KTNMEG.....
 EG-D-V-----GQ-I-----KGGV.EL-EY-----K-EPRKRMG-ESNLEGAGGADN.....
 EG-D-V-----Q-I-----KEGE.EL-I-----K-EPRKRMG-EGNMEGLRGPDNQMARNQSILDD..
 EG-D-V-----R-I-----LKEGE.EL-----K-EPRKTLG-ETHLEGAGGSDHQMAGDS.....
 EG-D-V-----Q-I-----LK-G.-L-----K-EPKQRVNEGQVEGTRGSDN.....
 EG-D-V-----R-I-----KEGE-----K-ERKTM-SEGSMEGVREANKQMEGDSLDLQDQE..
 EG-D-V-----I-----EQG.EL-TI-----K-L-SQAPLEGNRTAGEVD.....
 EG-DQ-----I-E-----QE.-L-----C-K-R.KEV-RETNMEGRQEEES.....
 EG-DQ-R-----I-----EETG-L-----KE-R.KDV-SEANLAGRQEEEN.....
 EG-D-Q-----QKQE.N-L-----LVK--GESSVEM-G.....
 TGAD-Q-----R-----KTEEG-LT-----KP-T.TENVGGDTNQYNLRKQDGLAN.....
 TGAD-S-----H-----L-KTDQGEVIT-----P-S.KEDVGSKP-AHQIREVDGMAD.....
 TGSD-S-----A-----VKTEQGGVIT-----KP-A.KENVGSKSNTG-HRKEGDLN.....
 TGSD-S-----V-----R-----KTD-GEVIT-----KP-A.KEDVGSKSDTG-LRKEGDLN.....
 RGA-S-Q-----Q-----I-KTDQGE-IT-----K-T.KENV-SDPNPVYDRKEDGLAD.....
 RGA-Y-LR-R-----L-KTKEGE-VT-----K-T.RQNV-SEPD5VHRKEDGLAD.....
 EGAD-T-----T-----CKTEVG-----V-KE-----K.D.M-SKVNT-NSNE.....
 EGAD-T-----T-----CKTEVG-----V-K-----K.G.V-SKTD-T-NPSE.....
 QGA---Q---V-----V-TQAGE-IT-----KP-EAKTE-VGG-THTNN.....
 QGA-H-Q---V-----V-TQAGE-IT-----KP-EAK-E-VGS-AHTSN-REEGRMAD.....
 EGA-Q-Q---E-----KTQAG-FP-----KP-NAESRKGQONKGMDS-TDI.....
 QG-D-Q-----Q-----VKEGD.T-FS-----LVK--E-.GPK-SEGGLHNN.....
 QG-D-Q-----QI-----VKEGD.T-FS-----LVK--E-.GPK-SEG-LDNN.....
 QG-D-Q-R-----Q-----VKEGE.N-FSI-----LVK--E-.GPK-SQS-MDNO.....
 QG-D-Q-----Q-----VKEGE.N-FS-----LVK--E-.GPK-SES-LDNN.....
 QGK.NE---R-----V-TEEG-FA-----T-H-E.....RM-SGSHVENDPKTD.....
 HGT.SD-Q---S-----TPDQQVIA-----SSD-E.....RV-SGTHLEITSKSN.....
 EHT.GE-Q---GE-V-----RNSQGLTF-K---V---T-LQYGEDV-SENLL-NGQKAEYKGM.D.....

1002
 1002
 1010
 1004
 1004
 1000
 1002
 1001
 1003
 1008
 1003
 1014
 1009
 1009
 1003
 1015
 1002
 1013
 1006
 1010
 1008
 1007
 1001
 1002
 999
 1005
 1040
 1032
 1036
 1035
 1015
 1036

1019
 1013
 1005
 1015
 1015
 1012
 1009
 1016
 1011
 1011
 1019
 1020
 1019
 1022
 1033
 1031
 1024
 1040
 1039
 1022
 1018
 1016
 1017
 1020
 1011
 1011
 1013
 1016
 1007
 1007
 1015
 1016
 1031
 1019
 1019
 1020
 1019
 1036
 1040
 990

Table with columns for sequence identifiers (e.g., H1B.FR.83.HXB2), sequence alignments (e.g., MEN.....RWQVMIVWQVDRMR), and accession numbers (e.g., 104). Includes a 'Vif start' marker.

Table with columns for sequence identifiers (e.g., MAC.US.x.239), sequence alignments (e.g., --EEK.....IAVPT-RIPER), and accession numbers (e.g., 106). Includes a 'Vif start' marker.

H1B.FR.83.HXB2
 H1A1.UG.85.U455
 H1C.ET.86.ETH2220
 H1D.CD.84.84ZR085
 H1F1.BE.93.VI850
 H1G.SE.93.SE6165
 H1H.CF.90.056
 H1J.SE.93.SE7887
 H1K.CM.96.MP535
 H1O1.AE.TH.90.CM240
 H1O2.AG.NG.x.IBNG
 H1N.CM.95.YBF30
 H1O.BE.87.ANT70
 H1O.CM.91.MVP5180
 CPZ.CM.05.SIVcpzEK505
 CPZ.CM.05.SIVcpzLB7
 CPZ.CM.01.SIVcpzCAM13
 CPZ.CM.05.SIVcpzMB66
 CPZ.CM.05.SIVcpzMT145
 CPZ.CD.90.ANT
 CPZ.CM.98.CAM3
 CPZ.CM.98.CAM5
 CPZ.GA.88.GAB1
 CPZ.GA.88.GAB2
 CPZ.TZ.01.TAN1
 CPZ.US.85.CPZUS
 H2A.DE.x.BEN
 H2B.GH.86.D205
 H2G.CL.x.ABT96
 H2U.FR.96.12034

MAC.US.x.239
 SMM.US.x.PGM53
 STM.US.x.STM
 MNE.US.x.MNE027
 DRL.x.x.FAO
 MND_1.GA.x.MNDGB1
 MND_2.x.x.5440
 MND_2.GA.x.M14
 MND_2.CM.98.CM16
 TAL.CM.01.8023
 TAL.CM.00.266
 TAN.UG.x.TAN1
 VER.DE.x.AGM3
 VER.KE.x.9063
 VER.KE.x.AGM155
 VER.KE.x.TYO1
 GRV.ET.x.GRL.677
 SAB.SN.x.SAB1C
 RCM.NG.x.NG411
 RCM.GA.x.GAB1
 SUN.GA.98.L14
 MON.CM.99.L1
 MON.NG.x.NG1
 MUS_1.CM.01.1085
 MUS_1.CM.01.CM1239
 MUS_2.CM.01.CM2500
 MUS_2.CM.01.CM1246
 DEN.CD.x.CD1
 DEB.CM.99.CM40
 DEB.CM.99.CM5
 GSN.CM.99.CN166
 GSN.CM.99.CN71
 SYK.KE.x.SYK173
 SYK.KE.x.KE51
 LST.CD.88.447
 LST.CD.88.485
 LST.CD.88.524
 LST.KE.x.lho7
 COL.CM.x.CGU1

QLIHLVYDFCFSDSA.IRKALLGHVSPRCEYQAGHN.....KVGSLQYLALAAI.T.....TPKKIKPPLPSVTKLTEDRWNKPKTKGHRGSHTMNGH.*
 H--H--E--R-I-Q-R--K--V-----TRA--K-----R-----R-L-R-----192
 H--MH--AE--I-YR--D-----T-----K-A-A--S-V-K-----R-R-N-----193
 --M--A--I-----T-----A-R-----RQ--R-----193
 --I-----E--I--RI--N-----T-----A-E-T--Q-V-----E-R-----193
 H--H--I--Q-----Q-----KV-V-----S-RSR-----E-A-----ENP-----193
 --MH--E--I-RV-R--N-P--K-----Q-T-----T-V-----A-----R-V-----R-----193
 --MC-----I-Q--D-----T-----R-RR-----Q-V-----T-E-----193
 --I-----E--V--I--R--C-----T--V-----A-RRP--V--K-V-----R--Q-----193
 K--Q--G--I-QV-RR--PS-----K-T-----R-R-----E-I-----RGHREN-P-----191
 --N-----I-E-R-----N-V-----A-T-T-----R-A-----KE--R-----RP-----193
 K--H--TA--Q-V-RP-L--P--K-----Q-T-----T-WV-----GA-R-----EH--MQ--NPI-----193
 RM--H-T-TA--V-I-QR-LTK--PT-S-----Q-T-L--R-VV-----KASRS-----Q-----HLRTRDQLK-PS-----193
 RM--H-T-TE--I-QR-LTK--L-S-----Q-T-F--K-VV-----KV-RN-----QR-----W-IRDQL--S-----193
 K-----TA--Q-V-RP-I-K--P--K-----Q-----I-WV-----GVQ-RR-----RH-I-D-----T-L-----193
 HM--N-AE--I-RV--FP--K-----T--IS--L-----QROPTR--S-A-----S-----I-S-----193
 K--SQ--A--R-I--E--A-RE-R-----Q-T-F--K-V-----KE-RLR--A-----T-R--PEN--S-----194
 --AE--I-Q-----PT-----ST-----KS-PRR--R-A-----NS--RD-K-N-I-----193
 R--H--TAT--N-I-EKA-L-K--FPE-S-----W-KQ-V-----KT-NR--A-A-----R-RD-QE-PMSS-C-----193
 --M-Q--P--Q-V-Q-I-ERILTY-H-KK-S-----Q-T-----FCKILEFR-----GY--GPRRQF--LSI-----RRMR--ENO*-S-C-----197
 K--Q--AA--V-KQ-Y-K--P--Q-----Q-----R-WV-----RVG-K-----RH--IRV-QE-----193
 R--H--AA--V-KQ-Y-K--PT-Q-----QX-----R-WV-----GVQ-KT--IX-A-----Y--R--QEN--X-----193
 R--SQ--AET--R-I-QL-A--KE-R-----Q--F--K-----SERRR--A-----H-R--V-QENL-R-----194
 R--C--AA--V-R-I-E-TKV-FPR-P-----Q-Q-----IVQVV-----KT--C-----I-A-----RDOQNL--S-----193
 --TR--T-AAG--V-Q-I-ERILTF-HF-S-R-----Q-T-F--FRKVVESQ-----DKQ--GPRR-----HRT-T-R-EN--LS-C-----199
 K-----TA--V-RP-I-K--PT-----F--K-V-----GQS-RR-----C-----QPEN--S-----193
 I--ST-S--TAGE-V-R-IR-EKLLSC-N-PQA-KA-----Q-P-----VVVQQ-----NDRPQKGTARKQWRDHRGLRVAREDHRSLSKGGGSE-SA-RAHFPGVAKVLEILA.....216
 --L-GS--S--ANE-V-R-IR-EKILSY-N-PSA-EG-----Q-P--F--RVQEQK-----NG.SOGESATRKQRRNRSRIRLARKNNRRAQQGSG-FA-RTYFPLGAEVLGILA.....217
 --L-GS--S--TAGE-V-R-IR-EQILSC-N-PTA-KR-----Q-P--F--QVQKG-----HG.SKGESXTRKQRRGDRNRGIRMARKNSXRDQDSSQSFTQGAYFPLGAKVLGILA.....216
 --L-GS--S--TAGE-V-R-IR-EQLLSC-NFPSA-RQ-----Q-P--F--R-Q-----DGRSKRENPTRKFRRRNCRRGFLQARKNCEYVQGSSTSSQGTYPFPLGAKVLAILA.....215

MAC.US.x.239
 SMM.US.x.PGM53
 STM.US.x.STM
 MNE.US.x.MNE027
 DRL.x.x.FAO
 MND_1.GA.x.MNDGB1
 MND_2.x.x.5440
 MND_2.GA.x.M14
 MND_2.CM.98.CM16
 TAL.CM.01.8023
 TAL.CM.00.266
 TAN.UG.x.TAN1
 VER.DE.x.AGM3
 VER.KE.x.9063
 VER.KE.x.AGM155
 VER.KE.x.TYO1
 GRV.ET.x.GRL.677
 SAB.SN.x.SAB1C
 RCM.NG.x.NG411
 RCM.GA.x.GAB1
 SUN.GA.98.L14
 MON.CM.99.L1
 MON.NG.x.NG1
 MUS_1.CM.01.1085
 MUS_1.CM.01.CM1239
 MUS_2.CM.01.CM2500
 MUS_2.CM.01.CM1246
 DEN.CD.x.CD1
 DEB.CM.99.CM40
 DEB.CM.99.CM5
 GSN.CM.99.CN166
 GSN.CM.99.CN71
 SYK.KE.x.SYK173
 SYK.KE.x.KE51
 LST.CD.88.447
 LST.CD.88.485
 LST.CD.88.524
 LST.KE.x.lho7
 COL.CM.x.CGU1

I-L-ST-P--TAGE-V-R-IR-EQLLSC-RFPRA-KY.....Q-P-----KVVS.....DVRSQGENPTWKQWRDRNRRLRMAKQNSRGDKQRGG--TKGANFPGLAKVLGILA.....215
 T-L-GT-P--EGE-V-R-IR-EKLLSC-KFPKA-KN.....Q-P-----TVVS.....HYRSQREDPTWKQWRDRNRRLRMAKQNSRRNKQGSSESLAEGANFPGLAKVLGILA.....215
 --L-GT-P--TAGE-V-R-IR-EKLLSC-RFTKA-KN.....Q-P-----KVVE.....HYRSORENTARKQWRDRNRGRSIRVATONGRHG-RGS--STEGTDFPGLAKVLGILA.....215
 I-L-ST-P--TAGE-V-R-IR-EQLLSC-RFPRA-KN.....Q-P-----RVVS.....HYRSORENTWKQWRDRNRRLRMAKQNSRGDKQRGG--TKGADFPGLAKVLGILA.....215
 RI--E-P--T-Q--GV--YRIQA-YFPR-KG.....Q-----Q-H-KYLKH.....GRKTVSGESSGANGSRGVRRRMATEESLKNQ--GGSITLP-R-PLPSLEHLHGALP.....218
 H--HFY--M--VM--IR-EE-LKV-RFP--RQ.....G-L--F-C-RVYQPEER.....VATQSSADLQTHSNTKASARKAHERSRRNAYTLHKRGHWSLESLECRIASRP.....173
 RI--GE--P--T-Q--F--ERLVA-YFPW--KG.....Q-T--F--Q-YLRGRKH.....GREGTRGARRSRRTGTAMAGNVGENQ--GGVPTLP-R-PFPSLEHMCRTLA.....214
 RM--WE--P--TAA--V--Y-ERIVA-YSPW--KG.....Q-T--L--R-Y-KFCRN.....GRKSTRGTTGRRSRTVAGKIIGTSEQGSIITLP-R-PFPSLEHMCRTLA.....213
 RM--WK-LP--TEQ--Q--KRLTV-YFHW--KG.....G-----LSYTAYCNN.....GRRGPRDRRSRRGGSKYLVGKISGENQSGKITLP-R-PFPSLE-MHRTLA.....213
 H--TK--P--TA--V-Q-IR-EVLLAV-VRHKGRD.....SPP--F--KVAVRHGK.....VATQSSANNIKAHNTKSSIRKTHERSKGRDAHTLHKRGHWSLESLECRVTSRA.....228
 H--TK--P--TA--V-Q-IR-EVLLAV-VRHKGRD.....TPP--F--T-AVRYGK.....VATQSSADLQTHSNTKASARKAHERSRRNAYTLHKRGHWSLESLECRIASRP.....228
 HM--CH--P--T-R--Q-IR-ESFLW--T-KE--VAENH--WGQ--R--F--TVYTDFLR.....NGRRKRFQGGKTRMVRNLGSQQGAVGRMIKRHSRTQSGSTT-FWERTPLPSMELLSGRRKEWGTDRDKGL.....239
 S--TH--T--TER--QRFTF--QFPE--KKT.....GQ-P-----L-HQNGLRQ.....RSQSKTGGTRNMGFQEGAVGRMAKRRHARRYQSGSQDAFWARAPVPSMELLSGGRKES-SHARKL.....233
 GM--H--S--TER--Q--IR-ERFV--QFPE--KTT.....G-P-----L-HQNGLRE.....RSQRQARRSRDLGSKQRTMGRMAKHVRRSQQSGQITFWQAPLPSLELLSGGRRET-SPNDGGL.....232
 SM--CH--T--T-R--Q-IR-ERFV--QFPE--KTT.....G-P-----L-HQNGLRK.....RSQRGTTRTRNLGSQQGAVGRMAQRYGRNQRQSATAFW-RTPIPSMELLSGGRGET-KTH--GKI.....233
 SI--GH--N--KER--Q-IR-ERFV--QFPE--KST.....GQ-P-----L-HQNGLRE.....RSKRGKTRSRNLGSQQGAVGMAKRYVTRSQ-GGEEAUFWERTPVPMSMELLSGRRKTWYSHDGKGLQIL.....236
 RM--N--T--R--Q-IR-EKYTW--TFKE--KG.....Q-Q--L--V-YT.....NGIRKRSKRTFRMAGNLGSRQGMATRHAAQSKRRSQ-ALWNEHANPSMELLCRGG-ET.....220
 HI--WS--P--T-R--VQ-IR-EKYLW--KH-V--OPT.....GQ-P-----RVYTNGLRRVAPTSRRGSSQGSQESQRDRTRMARNMGAQAVRMRMAPRHVTGPQFRGPV-LPKESPFPSPVLEYCGRTSH.....235
 K-V-NQ--P--T-K--Q--I--EQ-LL--GFKE--RD.....Q-----IK-WARQQLK.....DGRGQRKSARGPHWGRCVCLPASQANMRSQSLRSQVALSSR-HFPSVAYLCGTLA.....223
 S--KT--P--T-R--Q5--EK-LV--AF-G--RD.....Q-T--F--IQ-WAK.....SQLDRYGRKSPRGPHWGRSRVPLATG.....ARKQGLGSQVTLSSRVHFPSSVAHLCGTLA.....217
 --M--FK--WTSRC--R-M--EKILHE-RN-VA-K.....GL-L--F-C-RV--HG.....QGERASRTATRRRATERAICTMAGRYHGRNRGSRG-AFPS-NSCPSLAICGLYVRVKG-VYESNHFAQ.....242
 --IV--TH--A--QQLD-V-R-IR-EKLLGQ-OH-L-T-YP.....T-PLT-EKL-FL-ALVR.....YGATPTKSAPKLDLPGSHRKTGPTGRHLGAKRRSP-TL-KGRAPWHLGTG-GAAC.....256
 --I--TR--S--QA-D-V-R-IR-EAXLGN-HHIQT--YMKVG.....TPPT-EK--FL-CFK.....YGKCTTEPSTLDLPRGSHARRSAGRHMGPKRGGPXAL-KGRAPWHLGSGH.....258
 RI--TQ--S--YSQD--PRVIR-QHTLGD--HPDA-T-KVG.....PPP--V--L--KK.....HGKLAQPXQRPPTPEWSPQPCARHKKPCGK.....218
 KI--THH--S--YSQD--PRVIR-QPLGD--HPGA-T-KVG.....PPPT--V--LT-KK.....HGKSTTQPSTSMALQSGPNYHAAGPTGHVGPKRGRG-TL-FQRRAPWNL.....236
 I--T-H-S--QEQD--IR--R-ERILG--OH-TA-LPKYK.....T-PP--EK--FF-Y-K.....OGDGTAKSPSPMAKPPGNTNNHGAGTKRHVGS-RRGRKALLQGRAPV-LGPSHGAPSRP.....248
 --I--TH--S--QQRD--QR-IR-EQLLRK--HIKT--YPKYV.....T-PL--EK--FF-Y-K.....HTNGASAKSPSLMAKSNANNYSGSASRPVGT-RRGRKTLRQAPWNLG.....239
 ET--KDH--S--TRQ--V-Q--R-EV-LST-ARKKA-YHRTGPREPP--IV-Q-A-R.....YGVVTITPTPTSI5FKNGRNHKAATGSSMGYHRGGKTLQQRDHMMMSLLVIT-APT-DRRTSHGSRHRF-KNHAER-MEAL.....259
 H--R--P--Q-0-Q-V-IR-E-LLOQ-QOYHRGPEGK.....GPP--I--RVAVK.....HGKVPTLPTTFHLLKNTHDKARVAAGHARSKRGS-FTHQRR-INSLESCLRTRPRLD-RTS-GCIGH.....231
 --R--P--Q-0-0-VTQ--K-EKLLQ-QOYHRGPEGK.....GPP-----RTAVK.....YGVKSTLPTTFHLLKNTHDKASTTSHAGSKRGS-F-TLHKRRWMSLESLECR-SCPTRLDRRTS-G-VSN.....231
 --I--GK--P--QNE--R-IR-EQILGN--HPEA-L.....Q--TP-T-EQL-FFAVVKH.....YGNKPTQSSYALAQPKGAHNHGYSKRKSVGNKRRGGRETL-KGRAPWNL.....238
 --R--TR--P--QNE--V-R-IR-EQLLGY-DHPEA-P.....Q--TPHT-ERL-FFAVVRY.....YGNKPTQSSYALAQPKGAHNHGYSKRKSVGNKRRGGRETL-KGRAPWNL.....238
 --N--F--P--TAR--VNO-VR-ELLTSH-WTP.....TD.....Q-P-----QVYLK.....DGGGFLQSLPACARNTMVLHSSKCRVDPKRDQ-CHKRGTSDRS-QAFYS-RNIWSLESIL-RRGD.....222
 A-V-SQ--P--R--VQ-IR-EKLTSH-WNF--KE.....Q-L-----QKYLK.....DGTGFLQSLPAAARGTTLHSSKQCLGRGRSNKCHQCSRTGS-RSMQAFYSGRNIWSLESMLRGGSRNRNTH-GLD.....234
 --M--CK--N--TRD-----W-EKIMAF-AMPT--KG.....C-L--F-C-RQ-OHVON.....KAAKESRKTSGRLWATKMGAMGSMASRHGGNKRRGGT-FPRNFGPSLRLCQQYVRR-GSLHEIHYHSK.....228
 --M--CK--N--TRD-----W-EKIMAF-AMPT--KG.....C-L--F-C-RQ-OHVOS.....KAAKESRKTSGRLWATKMGAMGSMVSRHHGGNKGGGT-FPR-HFGPSLGLCQQYVRR-GSLHEIHYHSK.....228
 --ML--VK-----WVHRD--VR-IR-EKILHS-KMPT--KG.....C-L--C-RQ-OHVOT.....KTAAPQPKAPRGLWSSKRAMGSMASRHGGNKRRGGT-FPR-NFGPSLGLCQQYVRRRGLHEIHYHSK.....227
 --M--YK--N--IQRD--R--ERITMC-Q-PL--KG.....C-L--FIC-RQ-OHVOA.....EAAKSRRETRGLWTTKMGAMGSMVSRHHGGNKRRGGT-FPR-NFGPSLGLCQQYVRRRGLHEIHYHPSQ.....227
 GYM-IIRGK-VRIEEEA--IR-LPWN--DF-V-L.....G-LK--E-C-L--RE.....ERTHQMR-L-G-SHGCRTR.....192

```

MAC.US.x.239          Vpx start                                  Vpx end
H2A.GW.x.ALI        MSDP R R I P P G N S G E E T I G E . . A F E W L N R T V E E I N R E A V N H L . P R E L I F Q V W Q R S W E Y W H D E Q G M S P S Y V K Y R Y L C L I Q K A L F M H C K K G C R C L G . . . E G . H G A G G W . R P G P P P P P P P G L A *
H2A.DE.x.BEN        -AN--TV-----D--AL-----R-----Q--T-----M--MYT-FM--T-----G--P--S-----V-
H2A.SN.x.ST         -T--V-----E--I-AL-----R-----A--T-----M--I-T-F-R--T-W-. . . D.M-RE-LEDQ-----V-
H2B.GH.86.D205     -AG--T-----D--A-----R-----I--T-----M--M-I-S-R--T-----G--P--S-----V-
H2B.CL.x.EHO        -V--D--V-. . . A-E-IT-L-V-----A-RE--I--T-----L-M--M-V-YT-----Q-. . . P--S-----
H2G.CL.x.ABT96      -V--V-----ET-L-HL-V-----K-A-RE--I--T-----M--M-I-FA--G--R-. . . P--S-----X
H2U.FR.96.12034     -X-E-D-----E--XT--I-----R-----A--T-----L-M--M-V-----T-Q..K..P--Q-----
MAC.US.x.251_1A11    -G-----E--V-. . . S-E--RD--A--Q-. . . Q--R-----E--T-----Q-. . . P--S-----IX
MAC.US.x.251_BK28    -Q-----M-----
SMM.US.x.S1VsmH635FC -E-. . . G-A-----M--E--T-----T-----
SMM.US.x.S1VsmH635SB10 -E-. . . G-A-----M--E--T-----T-----
SMM.US.x.S1VsmH635F_L3 -E-. . . G-A-----M--E--T-----T-----
SMM.US.x.H9         -XX--X-----XX--X-D-H-X-A-. . . X-. . . R-X-----M--V--T-X-----GE. . . T-----X
SMM.SL.92.SL92B     -T-----HN--AL-QT-Q-. . . R-C--V--Y--A--VQ-M--M-Q-FR--T-R-. . . SQ--T-----
SMM.US.x.PGM53      -A-E-. . . H--D-----R-----R--M--E--T-----V--R-----E-. . . S-----
STM.US.x.STM        -R-----PG-----S-R-----
MNE.US.x.MNE027     -AE. . . G-ERVPE.APTGA-DVEFAP--H-MLT-V-L--RL-FH-. . . F-RL-RTCV-H--RL-R-LE-AG--L-M----I--QS--SQR. . . Q-QAREA-ERIQIL--M-.
RCM.NG.x.NG411     -AE. . . . RAPE.VPTGA--AEFQP--RDML-KV-L--RL-FH-. . . F-RL-RTCV-H--VHOR-LE-AA--L-M----I--QT--SQR. . . . PNPRAV-ERITIL--M-.
RCM.GA.x.GAB1      -AERQSVVERAPA.EPMGA--VELE--Q-SLLR--Q--RL-FH-. . . FL-RL-NTCM-HY--ALQL-FT-S--L-L--M--FQQ-S--QGRHPPPLRPA-D.-L-. . .
DRL.x.x.FAO        -AE. . . . RAPE.AP-GA--VGLEQ--ETSL-R--RL-FH-. . . FL-RL-NTCV-H--RHQR-LD-A--L-MH--MYT-MQQ--P-RN. . . . RP. . . . M-.
MND_2.x.x.5440     -AE. . . . RAPE.APQA--VGLEQ--E-SL-QV--QL-FH-. . . FL-RL-NTCV-H--RLRRTLN-A--L-M--M-V-MQQ--P-RS. . . . S. . . .
MND_2.GA.x.M14     -AE. . . . GAPE.IP-GA--VDLNT--E-SL-K-Q--RL-FH-. . . FL-RL-NACI-H--RHQR-L--A--L-MN--M-T-MQOE-P-RS. . . . .P. . . . . MV-
MND_2.CM.98.CM16

```

Vpr start	oligomerization	amphipathic α -helix	Vpr end in HXB2 frameshift in HXB2 H(S/N)RIG motifs	Vpr end	
H1B.FR.83.HXB2	MEQA	PEDGGQREPHNEWTLELLEELKNEAVRHFPRILWHLGQHQHYETGYD	TWAGVEAIRILQQLLFTHXFRIGCRHSRIGVTR	QRRRANGASRS	97
H1A1.UG.85.U455	YA *A	H-D	Q	IIP	96
H1C.ET.86.ETH2220	Y -A	Q	P -N -Y	S -L -T -M	97
H1D.CD.84.84ZR085	A -Y	S	S -Y	I -P	97
H1F1.BE.93.VI850	G -Y	A -I	P -N	E	97
H1G.SE.93.SE6165	Y -A	E	L	N	97
H1H.CF.90.056	Y -A	I	V -Q	N	97
H1J.SE.93.SE7887	H -YH	P -S	Y -S	E	97
H1K.CM.96.MP535	N -I	R	P -N	T	97
H1O1.AE.TH.90.CM240	E -Y	H	P -Q	Y -NN	96
H1O2.AG.NG.x.IBNG	A -S	H	P	Q	97
H1N.CM.95.YBF30	N -A	F -A	A	P -A -Y	96
H1O.BE.87.ANT70	L -N	A -F	E	P -QAC -Y	96
H1O.CM.91.MVP5180	Y -A	D	L	P -Q	101
CPZ.CM.05.SIVcpzEK505	Y -A	V	Q	Y -D	96
CPZ.CM.05.SIVcpzLB7	Y -A	V	Q	Y -D	96
CPZ.CM.01.SIVcpzCAM13	P -Y	A -I -D	K -Y -S	Q	97
CPZ.CM.05.SIVcpzMB66	Y -I	R	D -Q	T	97
CPZ.CM.05.SIVcpzMT145	S -YEQ -L -DT	I -Q	P -A -E	S	99
CPZ.CD.90.ANT	Q -E	M -L -T -I	QPT -QH -NMV -AN	S -R -QVL -IL -KA -T -Y -H -A	89
CPZ.CM.98.CAM3	P -Y	A	F -D	L -Q	96
CPZ.CM.98.CAM5	P -Y	X -I	T -L	P -R -Y	96
CPZ.GA.88.GAB1	P -YQ -A -T		P -Q -F -D	V -H -V -L -Q	97
CPZ.GA.88.GAB2	AG -Y -A -I -D		P -Q -L -NN	E -H -V -Q	98
CPZ.TZ.01.TANI	N -N	Y -T -L -DI	I -Q -K	PI -Q -V -NMVFTI	84
CPZ.US.85.CPZUS	A -Y	A -DV	E -L	P -Q -Y -S	101
H2A.GW.x.ALI	PTFEP -AGM -HQGAR	-VI -V -R -I -E -L	-DPRM -IA -GY -TRH	-LERAREL -NA -RA -M -A -GR -V -Q -GR	106
H2B.CL.x.EHO	VPEIP -KN	-WEQ -VVDV	-I -Q -LK -DPRL -TA -NF -NRH -N	-LE -AGEL -KL -RA -L -G -Q	102
H2G.CL.x.ABT96	AEEIP -EA	-WD -VV -V -I -E -LK	-DPRL -TA -NY -DRH	-LE -AGEL -K -RA -L -G -C	103
H2U.FR.96.12034	AEIP -GA	-GD -VR -I -E	-DPRL -TA -NY -NRH	-LE -AGEL -K -RAI -L -G	98
MAC.US.x.239	EERP -NE	-WD -VV -V -E -LK	-DPRL -TA -N -NRH	-LE -AGEL -RA -M -G -I	102
SMM.US.x.PGM53	P -EA	-WD -VVKV -V -E -LK	-DPRL -TA -NY -DRH	-LE -AGEL -RA -M -G -A	102
STM.US.x.STM	P -E	-WD -VV -V -I -Q -L	-DPRL -SA -NY -DRH	-LE -AGEL -K -RA -M -G	102
MNE.US.x.MNE027	P -E	-WD -VV -V -E -LK	-DPRL -TA -NY -DRH	-LE -AGEL -K -RA -M -G -N	102
DRL.x.x.FAO	P -E	-P -YD -LVD -V -IQE	-LK -D -RL -N -MV -SR	SLE -HQL -IL -RA -L -Y -Q	101
MND_1.GA.x.MNDGB1	G -KR	DEQVS	P -Y -Q -LADTM -I -E -RK	-L -I -NAVSEYCVNT -S	105
MND_2.x.x.5440	P -EA	-G -Y -K -LVGT -A -IQE	-LK -D -RL -AV -SW -Q	-LE -QQL -S -RA -L -Y -G	102
MND_2.GA.x.M14	P -EA	-Y -L -LIDT -A -IQE	-LK -D -RL -AV -SW	-LE -QKL -TL -RA -L -H -E	102
MND_2.CM.98.CM16	P -EA	-Y -L -LIDT -A -VQE	-LK -D -RL -AV -SW -K	-LE -RQL -S -KA -SM -Y -Y -E	102
TAL.CM.01.8023	RL	PPSHPLPITSRLIPTPRPQLER -MR -VRD -MLI -T -EEVI -VMN -CV -LPAEPD	WTGEOAWA -S -IDFA -KGNEM -L -L -QQ -F -EQRR -AHYPNIRPLSGRR -NTMQ		115
TAL.CM.00.266	RL	PPSHPLPITSRLIPTPRPQLER -MR -VRD -MLI -T -DEVI -VMN -CV -LPA -PD	WTGEOAWA -S -IDFA -KGN -M -LL -L -QQ -Y -EQRR -PHYNIRPLSGR -ERTMQ		116
TAN.UG.x.TAN1	ASGR	DSRERRPGWL -IWDLS -WD -LRDMVA -NQ -Q -G -EL -FOVMNFCO -EGER	NGAPMIERAYRY -LV -KA -V -C -RRPFPEYEERRNGGGGRDREPPGGLA		120
VER.DE.x.AGM3	ASGR	DPREARPGEL -IWDLS -WD -LRDM -DINQ -KM -G -EL -FOVMNYCO -EGER	NRTPMLEAYRY -LV -KA -V -C -RRPFPEYEERRDGGGGGRAGRVPPGLD		120
VER.KE.x.9063	ASGR	GPRENRPGEV -IWDLN -WD -LRDM -D -NQ -KL -G -EL -FOVMNYCO -EGER	RGAPMIERAYRY -LV -KA -V -C -RRPFPEYEERRDGGGGGRANRAPPGLD		120
VER.KE.x.AGM155	ASGR	DPREERPGEL -IWDLS -WD -LRDMV -IN -KL -G -EL -YQVMNYCO -EGER	QGRPIAERAYRY -LV -KA -V -C -RRPFPEYEERRNGGGGRGRVPPGLD		120
VER.KE.x.TYO1	ASGR	DPREARPGEV -IWDLS -WD -LRDM -D -NQ -RL -G -EL -FOVMNYCO -EGER	HGTPMMEAYRY -LV -KA -V -C -RRPFPEYEERRDGGGGGRANRVPPGLE		120
GRV.ET.x.GRL_677	ASGR	DPREPLPGWL -IWDLD -WD -LQDM -RD -NE -R -G -MNM -IRVMNYCV -EGRR	HTPWNIYGYKY -V -KSM -V -C -RRGFSPYEERRNGGGG -APPPGGLA		119
SAB.SN.x.SAB1C	ASGGWL	PPVGGDPPKPPNPREIEPGWL -TWDL -FD -LRDM -D -NS -QC -N -L -FR -WM -V -EPAI	DHGQTRLEGW -YKYC -KA -V -MKGR -CKPKTHPAYGAGG		141
RCM.NG.x.NG411	M	YD -LMDT -I -IQE -KK -TPEL -TQE -NY -QH	SLE -K -M -TL -NKA -L -H -EG -GA -GG		101
RCM.GA.x.GAB1	L	EE -YD -LMDT -I -QE -KK -TYEL -TQI -DYV -QH	SIE -Q -M -L -RA -L -N -AG -TS -G		101
SUN.GA.98.L14	AS	RREPVEQP -E -P -FAQ -LADTM -IRD -E -AAIVA -QVLQYCA	DSTGSEKGC -MRA -TL -NRA -L -LPM -PSYIRIRSGSG		117
MON.CM.99.L1	V	PIERQAQAIW -NE -LX -SKEE -R -IWOQVT -LPA -PA -WNADOAWA -CAIDYT -WV -TI -YR	Y -E -Y -RYAEQI -RYPVLRPMRGTA -PGTSSVPQADPNRPRPSRYRME		121
MON.NG.x.NG1	SL	XVPIERAAQAIW -NE -LX -SKEE -R -IWOQVT -LPA -PG -WNADOAWA -XAIQVT -RV -TXI -R -M -D -Y -RHAR -X -RYPXLRPLRGT -APGTTSSVPHADPERLHPSRYHDE			122
MUS_1.CM.01.1085	RV	PPSHPAWHSRVVPTTMRQAQ -AMW -VNE -EK -S -EE -R -IWEVNT -LPA -PD -WTVDOAAI -CAIDYV -RV -T -V -R -D -Y -RYNR -M -RYPTIRPLRGTQPP -NSTPNADVPPLPSRYRME			136
MUS_1.CM.01.CM1239	H	PPSHR -PWSHRVPTTMOQAQAMW -NE -EK -S -EE -R -IWDVNT -LPA -PN -WTVDOAAI -CAIDY -RT -T -R -Y -E -Y -RYSNTI -RYPNIRPLRGTQ -PP -NSMPNADPTPLPSRYRME			136
MUS_2.CM.01.CM2500	RV	PPSHPLWLSR -VPTTMOQAQAMW -NE -EK -S -DE -R -IWDVNT -LPA -PN -WTVDOAAI -CAIDY -WV -G -R -R -Y -Q -Y -RYSQ -I -RYPQVRLRGTQPP -NSTPNADVPPLPSRYRME			136
MUS_2.CM.01.CM1246	RV	PPSHPLWLSR -TPTTMOQAQAMW -NE -EK -S -DK -R -IWDVNT -LPA -PD -WSDVQAAITCAIDY -KSV -T -R -Y -D -Y -YAQTI -RYPNIRPLRGTQPP -NSMPNADPTPLPSRYRME			136
DEB.CM.99.CM40	RV	PPSHR -HFTSRTIPMTRGLQQAQMD -NE -LK -T -EE -F -VMN -CVDLPAQPD -WTGEOAWA -S -IDY -K -V -KM -WL -L -EA -F -RREA -RYPNIRPLT			118
DEB.CM.99.CM5	RV	PPSHR -HFTSRTIPMTRGLQQAQMD -NE -LK -T -EE -F -VMN -CVDLPAQPD -WTGEOAWA -S -IDY -K -V -RM -WL -L -EA -F -RREA -RYPNIRPLT			118
DEB.CM.99.CM166	RV	PPSHR -HFTSRTIPMTRGLQQAQMD -NE -EK -S -DE -R -IWDVNT -LPA -PD -WTVDOAAI -CAIDY -RV -T -R -R -D -Y -RYNR -M -RYPTIRPLRGTQPP -NSTPNADVPPLPSRYRME			134
DEB.CM.99.CM971	RV	PPSHR -HFTSRTIPMTRGLQQAQMD -NE -EK -S -DE -R -IWDVNT -LPA -PD -WTVDOAAI -CAIDY -RV -T -R -R -D -Y -RYNR -M -RYPTIRPLRGTQPP -NSTPNADVPPLPSRYRME			136
DEN.CD.x.CD1	RV	PPSHR -ALVSRMA -TTQRQLQ -AVWDITE -RK -SKEEIT -IWD -CWSLPAHP -WTGEOAWA -A -IDF -N -M -KEIWK -Y -V -F -REARV -HYPNIRPLRPREP			114
LST.CD.88.447	SRTR	QORSPEKP -F -P -WGO -LADTM -I -E -R -ETI -AQVSEYCV	SSTGSEEK -C -MKF -T -VNRA -L -LPT -PSYAAARAGTAHPTAEP -PTKRDIQRGY		115
LST.CD.88.485	SR -R	QORSPEKP -F -P -WGO -LADTM -I -E -R -EAI -AQVSEYCV	SSTGSEEE -C -MRF -T -VNRA -L -LPT -PSYAAARAGTAHPTAEP -PTKRDIQRGY		115
LST.CD.88.524	SRQR	QORSPEKP -F -P -WGO -LADTM -I -E -R -EAI -AQVSEYCV	SSTGSEEE -C -MRF -TLVNRA -YL -LPQ -PSYAAARAGTAHPTAEP -PTGRDIQRGHY		115
LST.KE.x.lho7	SRQR	QORSPEKP -F -P -WGO -LADTM -I -E -R -EVM -AQV -YCV	SSTGSEEE -C -MKF -TLVNRA -L -LPT -PSYAAARAGTAHPTAEP -PNEGDIQRGSD		115
SYK.KE.x.SYK173	AE -FFNPSQH	VQGTWPFIPRNVELTP -NVZINVTYKA -LVLT -SK -TPQEIY -VMQSLNEEAG	TDSPMTAWERTMLDMV -A -NLM -E -AA -PORTYARHRGYPHPS		114
SYK.KE.x.KE51	A -FFRYQR	QPGELFYIPRNVD -AGVISATAKA -QVAQ -C -TAEIF -VMQCL -VEAG	PEETPTMAWIRCLMDH -A -NFM -E -AA -PORTYARHRGYPHPS		116
COL.CM.x.CGU1	RXPREDFT	L -LQDV -NITR -R -LTHISVIMGCTRR	HW -GRP -T -L -YV -MECA -YL -VHA -QCRFP -STQNP		79

Table with columns: Tat start, disulfide bonding, NLS, exon 1 end, exon 2 start, Tat end. Rows include sequences for HIV-1 proteins like H1B.FR.83.HXB2, H1A1.UG.85.U455, etc.

Table with columns: Tat start, exon 1 end, exon 2 start, Tat end. Rows include sequences for HIV-1 proteins like MAC.US.x.239, SMM.US.x.PGM53, STM.US.x.STM, etc.

H1B.FR.83.HXB2*	100
H1A1.UG.85.U455	102
H1C.ET.86.ETH2220	102
H1D.CD.84.84ZR085	103
H1F1.BE.93.VI850	102
H1G.SE.93.SE6165	102
H1H.CF.90.056	101
H1J.SE.93.SE7887	102
H1K.CM.96.MP535	102
H101_AE.TH.90.CM240	102
H102_AG.NG.x.IBNG	102
H1N.CM.95.YBF30	102
H1O.BE.87.ANT70	115
H1O.CM.91.MVP5180	116
CPZ.CM.05.SIVcpzEK505	94
CPZ.CM.05.SIVcpzLB7	100
CPZ.CM.01.SIVcpzCAM13	100
CPZ.CM.05.SIVcpzMB66	102
CPZ.CM.05.SIVcpzMT145	102
CPZ.CD.90.ANT	114
CPZ.CM.98.CAM3	103
CPZ.CM.98.CAM5	108
CPZ.GA.88.GAB1	101
CPZ.GA.88.GAB2	100
CPZ.TZ.01.TAN1	SSGX.....	119
CPZ.US.85.CPZUS	103
H2A.GW.x.ALI	137
H2B.CL.x.EHO	139
H2G.CL.x.ABT96	129
H2U.FR.96.12034	128
MAC.US.x.239	131
SMM.US.x.PGM53	131
STM.US.x.STM	129
MNE.US.x.MNE027	131
DRL.x.x.FAO	Q.....	125
MND_1.GA.x.MNDGB1	116
MND_2.x.x.5440	100
MND_2.GA.x.M14	100
MND_2.CM.98.CM16	112
TAL.CM.01.8023	SATVTRP...	128
TAL.CM.00.266	111
TAN.UG.x.TAN1	102
VER.DE.x.AGM3	D.....	120
VER.KE.x.9063	A.....	120
VER.KE.x.AGM155	N.....	120
VER.KE.x.TYO1	114
GRV.ET.x.GRL_677	95
SAB.SN.x.SAB1C	94
RCM.NG.x.NG411	116
RCM.GA.x.GAB1	106
SUN.GA.98.L14	T.....	122
MON.CM.99.L1	104
MON.NG.x.NG1	112
MUS_1.CM.01.1085	99
MUS_1.CM.01.CM1239	114
MUS_2.CM.01.CM1246	KTSTT....	119
MUS_2.CM.01.CM2500	112
DEB.CM.99.CM40	GRISNNC..	116
DEB.CM.99.CM5	102
GSN.CM.99.CN71	DSRVSCGTT-	143
GSN.CM.99.CN166	159
DEN.CD.x.CD1	IREDS.....	97
LST.CD.88.447	101
LST.CD.88.485	101
LST.CD.88.524	112
LST.KE.x.lho7	112
SYK.KE.x.KE51	96
SYK.KE.x.SYK173	111
COL.CM.x.CGU1	WWS.....	85

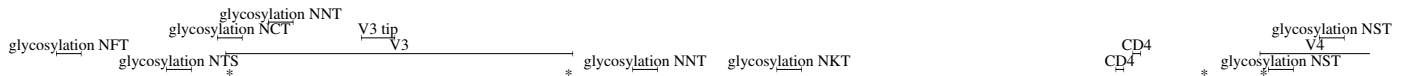
Table with columns: Rev start, exon 1 end, exon 2 start, NLS, Leu-rich effector domain, Rev end. Rows list various HIV sequences and their corresponding amino acid alignments across these domains.

H1B.FR.83.HXB2
H1A1.UG.85.U455
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101.AE.TH.90.CM240
H102.AG.NG.x.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H10.CM.91.MVP5180
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMT145
CPZ.CD.90.ANT
CPZ.CM.98.CAM3
CPZ.CM.98.CAM5
CPZ.US.85.CPZU5
CPZ.GA.88.GAB1
CPZ.GA.88.GAB2
CPZ.TZ.01.TAN1
H2A.GW.x.ALI
H2B.GH.86.D205
H2G.CL.x.AB796
H2U.FR.96.12034

Table with columns: Env start, signal peptide gp120 start, glycosylation NVT, glycosylation NDT, glycosylation NSS, and a numerical column. Rows list amino acid sequences for various HIV strains.

MAC.US.x.239
SMM.US.x.H9
STM.US.x.STM
MNE.US.x.M0E027
DRL.x.x.FAO
MND_1.GA.x.MNDGB1
MND_2.GA.x.M14
MND_2.CM.98.CM16
MND_2.x.x.5440
TAL.CM.01.8023
TAL.CM.00.266
TAN.UG.x.TAN1
VER.KE.x.TYO1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
GRV.ET.x.GRL_677
SAB.SN.x.SABIC
RCM.NG.x.NG411
RCM.GA.x.GAB1
SUN.GA.98.L14
MON.CM.99.L1
MON.NG.x.NG1
MUS_1.CM.01.1085
MUS_1.CM.01.CM1239
MUS_2.CM.01.CM2500
MUS_2.CM.01.CM1246
DEB.CM.99.CM40
DEB.CM.99.CM5
GSN.CM.99.CN166
GSN.CM.99.CN71
DEN.CD.x.CD1
LST.CD.88.447
LST.CD.88.485
LST.KE.x.lho7
LST.CD.88.524
SYK.KE.x.SYK173
SYK.KE.x.KE51
COL.CM.x.CGU1

Table with columns: Env start, signal peptide gp120 start, glycosylation NVT, glycosylation NDT, glycosylation NSS, and a numerical column. Rows list amino acid sequences for various HIV strains.



H1B.FR.83.HXB2
H1A1.UG.85.U455
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.S66165
H1H.CF.90.05616
H1J.SE.93.SE7887
H1K.CM.96.MP535
H1O1.AE.TH.90.CM240
H1O2.AG.NG.x.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMT145
CPZ.CD.90.ANT
CPZ.CM.98.CAM3
CPZ.CM.98.CAM5
CPZ.US.85.CPZUS
CPZ.GA.88.GAB1
CPZ.GA.88.GAB2
CPZ.TZ.01.TANI
H2A.GW.x.ALI
H2B.GH.86.D205
H2G.CL.x.AB796
H2U.FR.96.12034

MAC.US.x.239
SMM.US.x.H9
STM.US.x.STM
MNE.US.x.MNE027
DRL.x.x.FAO
MND_1.GA.x.MNDGB1
MND_2.GA.x.M14
MND_2.CM.98.CM16
MND_2.x.x.5440
TAL.CM.01.8123
TAL.CM.00.266
TAN.UG.x.TAN1
VER.KE.x.TYO1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
GRV.ET.x.GRL_677
SAB.SN.x.SABIC
RCM.NG.x.NG411
RCM.GA.x.GAB1
SUN.GA.98.L14
MON.CM.99.L1
MON.NG.x.NG1
MUS_1.CM.01.1085
MUS_1.CM.01.CM1239
MUS_2.CM.01.CM2500
MUS_2.CM.01.CM1246
DEB.CM.99.CM40
DEB.CM.99.CM5
GSM.CM.99.CN166
GSM.CM.99.CN71
DEN.CD.x.CD1
LST.CD.88.447
LST.CD.88.485
LST.KE.x.lbo7
LST.CD.88.524
SYK.KE.x.SYK173
SYK.KE.x.KE51
COL.CM.x.CG1

glycosylation NNT

	V4 glycosylation NNT	CD4 CD4	glycosylation NIT	glycosylation NES V5	fusion peptide gp120 end gp41 start			
H1B.FR.83.HXB2	NSTWSTEGSNNT	*EGSDTITLPR	CRIKQIIMNMQVKGKAMYPAPPISG	QTRCSSNITGLLLTRDGGNSNNESEIFRPGGMDRDNWRESLYKYKVVKIEPLGVPATKAKRRVVQREKRAVGIGALFLGFLGAAGSTMGAA534
H1A1.UG.85.U455	G-MSNDM-P-GQR-QQI-V-ETNTKN-TK
H1C.ET.86.ETH2220	ELFN-STNLQG-REI-M-RAKEPHSTKE
H1D.CD.84.84ZR085	I-GH-GLNDTIIENN-TQNDTAN-TQNDTE
H1F1.BE.93.VI850	-NGTVG-RTS-AN-T-NESN1-TE-N-KE
H1G.SE.93.SE6165	R-NS-EN-TKVR-RQN-M-E-N-SINNNTNTS
H1H.CF.90.056	MH-NY-SMDTKGNENV-RR-RQN-N-VII-E
H1J.SE.93.SE7887	KNSIEATNDTSXATI-KVRRT-Q-IAN-T-TR
H1K.CM.96.MP535	-G-IAS-NMINTHN-TQ
H101.AE.TH.90.CM240	GNETMAGCNDTKGA-QR-N-VIVNTDN-TNIK
H102.AG.NG.x.IBNG	--NSTANHTGSNDTQV-QQI-DSTN-TR
H1N.CM.95.YBF30	-E-GEFRV-L-TRGIRVLN-TV-EYS-PDTKET-VY-S
H1O.BE.87.ANT70	SCNGT-CSVS-VSQNGNGKLR-VVRSIRGQSGLKNLT-MMI-QM-NTWNSS-NNVT
H1O.CM.91.MVP5180	-C-K-GCQETKGSN-TNKG-IKLR-LVRSM-GESRIPNLT-HMI-QL-QPWNSTGENTL-V
CPZ.CM.05.SIVcpzEK505NTTGN-QR-V-L-TRGIKP-N-LII-DYTK-GT-KYIIV-T
CPZ.CM.05.SIVcpzLB7	RNKTLLETLHNNLR-V-L-TRGIFRN-N-TI-EKHT-TDNTI
CPZ.CM.01.SIVcpzCAM13ES-PFHENM-IR-V-S-MRRGIPH-T-N-LIDHW
CPZ.CM.05.SIVcpzMB66IFNETKDDN-IR-VRLR-RGIRT-N-II-FAQQ-TDRM-K-AM-T-V-E-N
CPZ.CM.05.SIVcpzMT145INKTENMTIIR-V-S-MRGIFRN-T-TMEIKHNREDOGEDODONNTYVCLT
CPZ.CD.90.ANTGMLITNGALIAHV-H-GI-SGI-LA-RRRNVS-T-SIMEGQIY
CPZ.CM.98.CAM3NNTIKGDNITFFR-V-S-MRGIRV-S-T-VII-ETGH-INNSITNITLY-T
CPZ.CM.98.CAM5	TCDNQMNCTR-NTNDTMSDNS-LISR-V-S-MRGIR-N-T-IEAG-KWEN
CPZ.US.85.CPZUSN-DNMTIKR-V-Q-MRGIFKVLS-NMI-DISAVNNDRSRITVM-T
CPZ.GA.88.GAB1GIIIR-VSS-MRRGIRN-T-NS-TPVTN-SGNLT
CPZ.GA.88.GAB2	DTSEYIK-R-V-S-MRGIFRRT-T-N-TEVQNGTGNTEVLS
CPZ.TZ.01.TAN1NRTNHTHTLVALRV-H-GI-SGV-L-RRTVK-HIM-AEKD-NNSYTPQFSAVVE
H2A.GW.x.ALI	KTGQEQHNYAI-H-ANV-L-REELH-N-TV-SIANI-TDGNQTNITFSAEVAELY-L
H2B.GH.86.D205	RLNTTRRNYAH-RT-HNI-L-REELS-N-TV-SIANIENSN-TTNISFSAEVAELY-L
H2G.CL.x.ABT96	RNMTLLKPORQRNNYVH-RT-HNV-L-RELH-N-TV-SIANIEENRSN
H2U.FR.96.12034	R-MEGQTSRRHKRNYVH-R-VT-HRVN-L-REDLT-N-TV-SMIANI-RI-K

	glycosylation NNT	CD4	glycosylation NIT	glycosylation NES	fusion peptide			
MAC.US.x.239	RN-ANQKPK-QHKRNYV*H-RT-HNV-L-REDLT-N-TV-SIANI-WIDGNQTNITFSAEVAELY-L
SMM.US.x.H9	IKNGRWTSONQKRYQKNYVH-XT-HNVXL-REDLT-N-TV-SIAEI-WI-G
STM.US.x.STM	R-SEMRDWNKNKQOKRNYVH-R-VT-HNV-L-RQDLT-N-TV-SSIANI-WT-N
MNE.US.x.MNE027	KNLTGTTQKPOQHKRNYVH-RT-HNV-L-REDLT-N-TV-SIANI-WIDG
DRL.x.x.FA0	RNSMVWYPYASH-RDDARKI-M-VFNNETDQV-EMFFEQKLETEETGSDNSYO
MND_1.GA.x.MNDGB1	IMRAWNDPNEKKWYPYASQ-RVDDMQ-RKI-L-TFNNHTHRV-EMFFEMKQDISNETKMQ-KFLPPSETSNOQFVAVGAH
MND_2.GA.x.M14	QNSMVWYPYMSK-RVDDH-RKI-M-VFNNHTNDV-EMFFEVQLVGKRYL-KFVQDEQNRFTAVGAH
MND_2.CM.98.CM16	QSSVWYPWMSN-RDDH-KI-M-VFNNETNDV-EMFFEVQKTEEGYI-KFVQDEQNRFTAVGAH
MND_2.x.x.5440	KQSSVWYPMYRNR-VGDH-KI-M-VFNNETNDV-EMFFEVQKTEEGGYI-KFIPQDWIQNYTAVGAHL-VD-I-F
TAL.CM.01.8023	TYNYTAKGVLHOEKIGRQEHYSDRTKFFMSQLR-M-RS-EL-L-REHLS-KV-IVDINHXYGN
TAL.CM.00.266	EYNYTSOQKLGHQVIGKKPRSSDRTKFFMSQLR-M-RSRS-L-REHLQ-KVDINHHRGSD-TVEPSSQI
TAN.UG.x.TAN1	RTVDQDH-PCSSKKKGHPAGPCVQRTYVAH-RSVD-YTLS-KTREHLQ-R-TVMTVELNYSK
VER.KE.x.TYO1	RTVDQDH-PCSSKKKGHPAGPCVQRTYVAH-RSVD-YTLS-KTREHLQ-R-TVMTVELNYSK
VER.DE.x.AGM3	Q-VDANH-QCSNETKGNAGPCCVQRTYVAH-RSVD-YTIS-KVREHLE-T-TVMTVELNYSKYRTNVTLSPOIETG
VER.KE.x.9063	LTDVDH-HCKNAGKGRSPGPCVQRTYVAH-RSVD-YTIS-KTREHLQ-R-TVMTVELNYSKYRTNVTLSPOIETG
VER.KE.x.AGM155	RTEDAH-GTNRCTDKKPGGPCVQRTYVAH-RVD-YT-S-KVREHLE-N-SV-AYVLI-YN-KSGP-NVTLSPOV-SI
GRV.ET.x.GRL_677	K-VDPPH-CAKNNTKPCWQRTYVH-RV-D-YTLS-KTREHLE-N-TAYVAINYSK
SAB.SN.x.SAB1C	KTEGNKNARQAMFVIT-MVVD-YT-SRKT-T-RPDALNATV-YADI-YTDN
RCM.NG.x.NG411	ETVGTTNKRKAPFVIT-MV-D-YT-SRKT-T-RPDALAQVSYADI-Y
RCM.GA.x.GAB1	TTSNKWYPWANH-SVDDAS-KI-L-TFNNRANRV-EAWFLERVEDW
SUN.GA.98.L14	-K-DGISPFVNNKPNNTYHGGWA-T-RLVTOGY-S-SI-LLRKHQV-TA-I
MON.CM.99.L1	XKNNKETFEFTFNVTNASTAYNGEWT-LRMVTQGYIS-SX-LRQHQV-TA
MON.NG.x.NG1	TEINESNIHVTYKYHQDQMWR-FVQ-GY-SSI-L-RKLVEIDGAMYRD
MUS_1.CM.01.1085	ITQINESNIHNVTKYHQDQMWR-FVQ-GY-SSI-L-RRHQV-TVIDGAMYRD
MUS_1.CM.01.CM1239	TRQIHNSNIDVTYKYHQDQMWR-FVQ-GYTSSI-L-RQHQ-Q-TVIDGAMYRD
MUS_2.CM.01.CM2500	TREINDSNIHVTYKYHQDQMWR-FVQ-GY-SSI-L-RKHQV-THVIDGAMYRD
MUS_2.CM.01.CM1246	QNNTRNSWKWLMLV-TYT-QHI-L-KEELKH-SAFVDF-HY-G
DEB.CM.99.CM40	MNKTNIIWNNTLMK-NLV-T-FQHV-L-KEELKH-SVSA
DEB.CM.99.CM5	GNASKNTINS-NIGITSKYKGQWA-K-RFVTOGY-S-SI-L-RQH-N-T
GSN.CM.99.CN166	ITAVDDTNIVTSKYKGQWA-K-RFVTOGY-S-SI-L-RQH-N-TIDGAMYRDS-NMTPSA
GSN.CM.99.CN71	PVNNSNIINNQTYMN-LV-THT-MDI-L-AEEH-T-TVS-IIVDT
DEN.CD.x.CD1	KNSDKWYPWANH-SVDDAS-KI-L-TFNNRTHRV-EMFPQMEKFEPEENLGGNS
LST.CD.88.447	KNSDKWYPWANH-SVDDAT-IKI-L-TFNNRTHRV-EMFPQMEKFEPE
LST.KE.x.lho7	KNSDKWYPWANH-SVDDAT-IKI-L-TFNNRTHRV-EMFPQMEKFEPE
LST.CD.88.524	KNSDKWYPWANH-SVDDAT-IKI-L-TFNNRTHRV-EMFPQMEKFEPE
SYK.KE.x.SYK173	GNASKNNYASNLRS-A-RDRY-R-LI-L-TAH-KK-TV-AV
SYK.KE.x.KE51	GYANPSNYAKNLYS-A-RDR-Y-GY-T-L-L-TEH-KK-TV-AV
COL.CM.x.CGU1PVMNT-RKLVS-VTHARILG-PGHLQNWEKQPIAFMGITIEGDNDGNCAYPAAPNFKHALSTL

Table with columns for protein names (e.g., H1B.FR.83.HXB2, H1A1.UG.85.U455) and amino acid sequences. Headers include glycosylation sites: glycosylation NAS, glycosylation NKT, glycosylation NHT, glycosylation NYT.

Table with columns for protein names (e.g., MAC.US.x.239, SMM.US.x.H9) and amino acid sequences. Headers include glycosylation sites: glycosylation NAS, glycosylation NKT, glycosylation NHT, glycosylation NYT.

H1B.FR.83.HXB2
H1A1.UG.85.U455
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H1O1.AE.TH.90.CM240
H1O2.AG.NG.x.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.01.SIVcpzCAM13
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMT145
CPZ.CD.90.ANT
CPZ.CM.98.CAM3
CPZ.CM.98.CAM5
CPZ.US.85.CPZUS
CPZ.GA.88.GAB1
CPZ.GA.88.GAB2
CPZ.TZ.01.TAN1
H2A.GW.x.ALI
H2B.GH.86.D205
H2G.CL.x.AB796
H2U.FR.96.12034
MAC.US.x.239
SMM.US.x.H9
STM.US.x.STM
MNE.US.x.MNOE027
DRL.x.x.FA0
MND_1.GA.x.MNDGB1
MND_2.GA.x.M14
MND_2.CM.98.M16
MND_2.x.x.5440
TAL.CM.01.8023
TAL.CM.00.266
TAN.UG.x.TAN1
VER.KE.x.TYO1
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
GRV.ET.x.GRL_677
SAB.SN.x.SABIC
RCM.NG.x.NG411
RCM.GA.x.GAB1
SUN.GA.98.L14
MON.CM.99.L1
MON.NG.x.NG1
MUS_1.CM.01.1085
MUS_1.CM.01.CM1239
MUS_2.CM.01.CM2500
MUS_2.CM.01.CM1246
DEB.CM.99.CM40
DEB.CM.99.CM5
GSN.CM.99.CM166
GSN.CM.99.CN71
DEN.CD.x.CD1
LST.CD.88.447
LST.CD.88.485
LST.CE.x.lho7
LST.CD.88.524
SYK.KE.x.SYK173
SYK.KE.x.KE51
COL.CM.x.CGU1

glycosylation NGS
VGLRIVFAVLSIVNRVQGYSPLSFQTHLPT
I---T---I---I---E-LG-GR---OGK---E-LG-PR---
I---I---I---LI-H---LG---QG---F-IF---
I---A---L-A---Q-G-T---FS---N---E-I-AA-T---
I---K---L-LI-S---G---OGK---V-T-F-A---N---RH-FI-AA---DRGL---L---G-TR---I-F-T-V---I---L
I---LTHH---Q-E---G-OG-G---V-S-F-P---SI---A-T---RSSLKGL---L-G-L-GR---I-DTV---NW---A-843
I---LV-N---T-G-Q---V-F-PV-V---S---RL---V-T---R---L---G---ID---T---GI-VI---822
I---LI-N---TEA---G---OG-T---F-A---N---N---FV-AA-T-GT---L---I---V-VW-G---I---T---I-IA
I---A---V---LI---S-A---Q-KN---V-S-F-A---N---RQ-N-I---L-RGL---G---L---V---I---T---G---I-IG
I---I---PSHH---QKE---G-OG---V-S-F-A---T-AA-T---HSSLKGL---G---L---G-L-G---I---D---A-GW---A-834
I---T---LTHH---Q-E---R-G-Q-K---V-S-F-A---I-AA-T---HNCLKGL---L-G---L---IS-V---IN---TT-V-NW---A-IG
I---I---ISI-IT-IA---L-LI---A---ET-GGV-Q-G---V-S-F-V-E---N-LI-L---T-S---LR-TL---QSLSRGL---QLLN-LRTHL---GI-A-GK-RD-I---T---V---I-LA
--V-VIMI-N-KNI---Q---L-IPNH---QEEAGT-GRTGGG---EG-P---W-PSPO-F-P-LYT---TII-WT---L-SN-ASGQKVISY-RLGLWLIG---QKIINVCRIC---AAVT---L---Q---T---DTL-V---NW---GI-AGI
I-I-VIMII-NL-KNI---Q---L-IPV-H---RQEAET-GRTG---G-P---KWT-A-PP-F-QQLYT---TII-WT---L-SN-ISGIR-LIDY-LGLWLIG---QKTI-CRCLC---GAVM---L---TN-DTI-VS-NW-GI-LGL
--I-ISIIM-A---L-LI---T---RT-DA-L-NG---V-S-F-A-E-F-N-L-L---T-C-S-LR-TL---RQNIHKL---QLLN-LRI-L---GIIA-GR---I-IN-DT-V---I-L-816
IVC-I---A---LS-I---LAQ-G-T---OGT---V-LD-F-A-N-D---L---T-V-AR-TL-IV-OYTLKGL---LV---L-L---QGI---GK---T---T---DT---I-I-822
I---IMS-V-VIR---L-LI-A---Q-E-GTLGETGG---PGSG---V-T-CWP---N-VI-W-RS-TS-AC-L-W-QL-K-QLLINIL---LIQ-K-TLL---RGII---GR-RT-T---D---G---I-RA-841
I---L---LI-A---R-DR---E---G---P-NV---S-F-A-N-D---L---T-L-V-OYTLKGL---R---IHL---RGI---G---T---I---DT---I-IA-854
I---C-I-G-LI---K---L-LI-A---N---R---LLET-KG---NV-G---S-F-P-V-E---N-LS-L-Q---CASLW-LI-I-OYSLRGV---QOIGTL-HQL---RGT---TT-I---I---DT---G---IL-IA
--L-LVLV-CLRK---H---I-QN---QOD-EQ-E-R---RK-I---WRA-OH-FF-L-V---T-IQIWI-QIC-TC-NLWAVLQ-CH---ITFRCLNHLE-N-STL-TIIRTEIKNIIDRL-W-G-K-SILLAL
--I---G-II-F-AK---L-LI---T-E---G-T-DV-PGKG---S-F---E---N-L-C---LW-GL-TN---OSLNKGL---QQLRNFSL-R---GVIT-GR-QT-I---D---I-839
I-F-LIGI-M-VIAK---L-LI---T-E---RG-T-D---PG-G---T-S-F-V-E---N-LI-F---T---L-G-T-QI---OSINKGL---QQLKDFST-L---KGV-T-GR-TQT-T---DT---N---I-F-833
I---I---GVIF-L-AK---L-LF---T-E---T-GA-KT-NV---T-S-F-V-E---N-LI-L---E---L-R-T-QI---ONINKGL---QLLN-LRARC---GVIA-AR---V-T---DT---I-LT-813
I---IMT-F-V-R---L-LI-V---Q-EQG-LGE-D-G-Q-S---V-E-C-P---N-GW-QS-TS-ACN-W-QLK-T---HLTHSL---LLR-R-CLL---GGII---GK---T---I---D---I-AF-831
I---IG-WN-IR-S---L-LI-A---TG-EV---GGG-Q-G---V-S-F-C---I-N-TI-L-R-L-C-S-VLWSL-Q-QQVLRGL---LLR-L-SQL---KGTG---L-L-RT-I---D---I-R-TI---A-832
--LMFIVNV-KQ---T-FS-IPTQA---EOD-EQ-G-AGG-G-NI---WTPSPA-FSIV-E---N-LIWI-QTFQNIW-LWLSLQA-KQGISLAHSLVIVH-TIIVGVQRI---I-EWSSNTAYS-RVLLIQADIRL-NFTWG-V-LI-G-848
-A---IY-VQMLS-L-K-R-VFSSPPYQIQIHIHK---DQEQ-TRG-T-EDVG-NV---GDRLWPMPIAY-HFLI-HLLAR-LIGLYSICRD-LRSISPI-OPFIRLSLQRLAITIRDWLRLK-A---Y---GCEWQEAARAFARIARETLNTWRDWA-
-V---YIYVQMLA-L-K-R-VFSSPPY-QOIPIRK---DQEQ-ANE-T-E-GNDGD---DYSWPMQIYIHFLLI-RQLRN-LIWFACR-L-LTKTFQ---LTHOISTNLQPLRLPVA---Y---GISWFOEALRAAR-GETL-SAGETLW-AL
IV---IY-VQ-LG-L-R-VXSSPPSYQOIQIHIRK---DQEQ-TKE-I-ESGNKG---GYRWSWPMQIYIHFLLI-RQLGN-LTWLYSNCRA-LRSIXQT-OPLFQIRISRTLOAIAREHLR-L-A---YFS-GFRW-QEACTAATR-AQETLSTWRALMKT
II---IY-VQLLGLS-L-K-R-VFSSPPSYVEIQPIR---DQEQ-TKG-I-E-AG-SG---GYRWSWPMQIRYIHFLLI-HQLIR-LTWLYSNCRD-LRSICQ-QPLF---Y---QSIRE-LHL-LI---Y---GW-YF-EAQAQFGK-ARETSLTRGRELW-TL
864

glycosylation NAT
I---T---I---I---E-LG-GR---OGK---E-LG-PR---
I---I---I---LI-H---LG---QG---F-IF---
I---A---L-A---Q-G-T---FS---N---E-I-AA-T---
I---K---L-LI-S---G---OGK---V-T-F-A---N---RH-FI-AA---DRGL---L---G-TR---I-F-T-V---I---L
I---LTHH---Q-E---G-OG-G---V-S-F-P---SI---A-T---RSSLKGL---L-G-L-GR---I-DTV---NW---A-843
I---LV-N---T-G-Q---V-F-PV-V---S---RL---V-T---R---L---G---ID---T---GI-VI---822
I---LI-N---TEA---G---OG-T---F-A---N---N---FV-AA-T-GT---L---I---V-VW-G---I---T---I-IA
I---A---V---LI---S-A---Q-KN---V-S-F-A---N---RQ-N-I---L-RGL---G---L---V---I---T---G---I-IG
I---I---PSHH---QKE---G-OG---V-S-F-A---T-AA-T---HSSLKGL---G---L---G-L-G---I---D---A-GW---A-834
I---T---LTHH---Q-E---R-G-Q-K---V-S-F-A---I-AA-T---HNCLKGL---L-G---L---IS-V---IN---TT-V-NW---A-IG
I---I---ISI-IT-IA---L-LI---A---ET-GGV-Q-G---V-S-F-V-E---N-LI-L---T-S---LR-TL---QSLSRGL---QLLN-LRTHL---GI-A-GK-RD-I---T---V---I-LA
--V-VIMI-N-KNI---Q---L-IPNH---QEEAGT-GRTGGG---EG-P---W-PSPO-F-P-LYT---TII-WT---L-SN-ASGQKVISY-RLGLWLIG---QKIINVCRIC---AAVT---L---Q---T---DTL-V---NW---GI-AGI
I-I-VIMII-NL-KNI---Q---L-IPV-H---RQEAET-GRTG---G-P---KWT-A-PP-F-QQLYT---TII-WT---L-SN-ISGIR-LIDY-LGLWLIG---QKTI-CRCLC---GAVM---L---TN-DTI-VS-NW-GI-LGL
--I-ISIIM-A---L-LI---T---RT-DA-L-NG---V-S-F-A-E-F-N-L-L---T-C-S-LR-TL---RQNIHKL---QLLN-LRI-L---GIIA-GR---I-IN-DT-V---I-L-816
IVC-I---A---LS-I---LAQ-G-T---OGT---V-LD-F-A-N-D---L---T-V-AR-TL-IV-OYTLKGL---LV---L-L---QGI---GK---T---T---DT---I-I-822
I---IMS-V-VIR---L-LI-A---Q-E-GTLGETGG---PGSG---V-T-CWP---N-VI-W-RS-TS-AC-L-W-QL-K-QLLINIL---LIQ-K-TLL---RGII---GR-RT-T---D---G---I-RA-841
I---L---LI-A---R-DR---E---G---P-NV---S-F-A-N-D---L---T-L-V-OYTLKGL---R---IHL---RGI---G---T---I---DT---I-IA-854
I---C-I-G-LI---K---L-LI-A---N---R---LLET-KG---NV-G---S-F-P-V-E---N-LS-L-Q---CASLW-LI-I-OYSLRGV---QOIGTL-HQL---RGT---TT-I---I---DT---G---IL-IA
--L-LVLV-CLRK---H---I-QN---QOD-EQ-E-R---RK-I---WRA-OH-FF-L-V---T-IQIWI-QIC-TC-NLWAVLQ-CH---ITFRCLNHLE-N-STL-TIIRTEIKNIIDRL-W-G-K-SILLAL
--I---G-II-F-AK---L-LI---T-E---G-T-DV-PGKG---S-F---E---N-L-C---LW-GL-TN---OSLNKGL---QQLRNFSL-R---GVIT-GR-QT-I---D---I-839
I-F-LIGI-M-VIAK---L-LI---T-E---RG-T-D---PG-G---T-S-F-V-E---N-LI-F---T---L-G-T-QI---OSINKGL---QQLKDFST-L---KGV-T-GR-TQT-T---DT---N---I-F-833
I---I---GVIF-L-AK---L-LF---T-E---T-GA-KT-NV---T-S-F-V-E---N-LI-L---E---L-R-T-QI---ONINKGL---QLLN-LRARC---GVIA-AR---V-T---DT---I-LT-813
I---IMT-F-V-R---L-LI-V---Q-EQG-LGE-D-G-Q-S---V-E-C-P---N-GW-QS-TS-ACN-W-QLK-T---HLTHSL---LLR-R-CLL---GGII---GK---T---I---D---I-AF-831
I---IG-WN-IR-S---L-LI-A---TG-EV---GGG-Q-G---V-S-F-C---I-N-TI-L-R-L-C-S-VLWSL-Q-QQVLRGL---LLR-L-SQL---KGTG---L-L-RT-I---D---I-R-TI---A-832
--LMFIVNV-KQ---T-FS-IPTQA---EOD-EQ-G-AGG-G-NI---WTPSPA-FSIV-E---N-LIWI-QTFQNIW-LWLSLQA-KQGISLAHSLVIVH-TIIVGVQRI---I-EWSSNTAYS-RVLLIQADIRL-NFTWG-V-LI-G-848
-A---IY-VQMLS-L-K-R-VFSSPPYQIQIHIHK---DQEQ-TRG-T-EDVG-NV---GDRLWPMPIAY-HFLI-HLLAR-LIGLYSICRD-LRSISPI-OPFIRLSLQRLAITIRDWLRLK-A---Y---GCEWQEAARAFARIARETLNTWRDWA-
-V---YIYVQMLA-L-K-R-VFSSPPY-QOIPIRK---DQEQ-ANE-T-E-GNDGD---DYSWPMQIYIHFLLI-RQLRN-LIWFACR-L-LTKTFQ---LTHOISTNLQPLRLPVA---Y---GISWFOEALRAAR-GETL-SAGETLW-AL
IV---IY-VQ-LG-L-R-VXSSPPSYQOIQIHIRK---DQEQ-TKE-I-ESGNKG---GYRWSWPMQIYIHFLLI-RQLGN-LTWLYSNCRA-LRSIXQT-OPLFQIRISRTLOAIAREHLR-L-A---YFS-GFRW-QEACTAATR-AQETLSTWRALMKT
II---IY-VQLLGLS-L-K-R-VFSSPPSYVEIQPIR---DQEQ-TKG-I-E-AG-SG---GYRWSWPMQIRYIHFLLI-HQLIR-LTWLYSNCRD-LRSICQ-QPLF---Y---QSIRE-LHL-LI---Y---GW-YF-EAQAQFGK-ARETSLTRGRELW-TL
864

	gp41 end		
	Env end		
H1B.FR.83.HXB2	* QGACRAIRHIPRRIRQGLERILL*		856
H1A1.UG.85.U455	-TIG---LN-----A--		856
H1C.ET.86.ETH2220	-RIW--FCN-----AA-Q		852
H1D.CD.84.84ZR085	RR--K-VL---T-----A--		857
H1F1.BE.93.VI850	-R-G--VLN-----A-A--		833
H1G.SE.93.SE6165	-R----LN--T-----A--		867
H1H.CF.90.056	-R-W--L-----F--S--		846
H1J.SE.93.SE7887	-R-F--L-----A--		849
H1K.CM.96.MP535	-R-F--LL-----A--		843
H101_AE.TH.90.CM240	--W--L-----T--		858
H102_AG.NG.x.IBNG	-RVG---N-----F-A--		855
H1N.CM.95.YBF30	-RIG-G-L-----A-I-		853
H1O.BE.87.ANT70	-RIGTG--N-----S--		857
H1O.CM.91.MVP5180	-RIGQGFL-----A---V-		870
CPZ.CM.05.SIVcpzEK505	-RIG-G-L-----A--		836
CPZ.CM.05.SIVcpzLB7	-RIG-G-L-----F-A--		846
CPZ.CM.01.SIVcpzCAM13	-IVF-I-GN-----T--		865
CPZ.CM.05.SIVcpzMB66	-RFG-G-LN-----A--		878
CPZ.CM.05.SIVcpzMT145	TRLG-G-L-----A--		853
CPZ.CD.90.ANT	-TIV-I--EV-----IA-N-		859
CPZ.CM.98.CAM3	-IIG-G-L-----S--		857
CPZ.CM.98.CAM5	-IVG-G-L-----S--		863
CPZ.US.85.CPZUS	RRLFLG-I-----S--		837
CPZ.GA.88.GAB1	-VTL-I--N-----A--		855
CPZ.GA.88.GAB2	TRIG-G-L-----A-E		856
CPZ.TZ.01.TAN1	VYIA-G--N-----LA-N-		872
H2A.GW.x.ALI	-WVG-R-LAV-----A-IA--		863
H2B.GH.86.D205	RR-A--IA-----LT--		860
H2G.CL.x.ABT96	GRVG-G-LA-----LT--		885
H2U.FR.96.12034	GRVG-WL-A-----F-LA--		868
	Env end		
MAC.US.x.239	RRGG-W-LA-----LT--		880
SMM.US.x.H9	GRVG-G-LA-XX-----LT--		887
STM.US.x.STM	GRVG-R-GA-----LT--		883
MNE.US.x.MNE027	GRVG-W-LA-----E-LT--		883
DRL.x.x.FAO	RALAQEVA-----A-VLFN-		896
MND_1_GA.x.MNDGB1	...GTC-W-L-----SA--P-N		880
MND_2_GA.x.M14	.AIG--GN-----A-VL-N-		868
MND_2_CM.98.CM16	.AIA-G-IN-----A-VL-N-		868
MND_2.x.x.5440	.AFG-G-WN-----A-AL-N-		891
TAL_CM.01.8023	RW-VQ--AA----T-----YC-A-		891
TAL_CM.00.266	RW-GE--AA----T-----YCFT-		873
TAN.UG.x.TAN1	RS-Y---LEH---M--E---WFN-		880
VER.KE.x.TYO1	RS-Y-N-VNS--V-----E--N-		864
VER.DE.x.AGM3	RS-Y---INS--V-----G--N-		878
VER.KE.x.9063	RSTY-H-ISS--V-----E--N-		878
VER.KE.x.AGM155	RS-Y---INS--V-----EV-N-		871
GRV.ET.x.GRL_677	RS-Y--VINS--V-----KV-G-		855
SAB.SN.x.SAB1C	RSIV--VIAH--M--E---WFN-		868
RCM.NG.x.NG411	RDPAGWPATLC-----F--F-N-		858
RCM.GA.x.GAB1	RDFAGWPAMVC-----LCN-		851
SUN.GA.98.L14	KWRPKWN-TRGS--PSETTETT--		920
MON.CM.99.L1	GRLW--VA-----A-IF-N-		870
MON.NG.x.NG1	GRLW-GLVAV-----A-IL-N-		882
MUS_1_CM.01.1085	RRVW-EFLA---V--A-IL-N-		898
MUS_1_CM.01.CM1239	RRVW-EFLA-----A-IL-N-		892
MUS_2_CM.01.CM2500	KQL+K-FLA--K-----A-IL-N-		879
MUS_2_CM.01.CM1246	-KIW-KFLA-----A-IL-N-		880
DEB.CM.99.CM40	VR-GG--LRV-A-----A-L-N-		800
DEB.CM.99.CM5	VWGG-NLLA--A-----A-LL-N-		784
GSN.CM.99.CN166	RRVW-EFLA-----A-ILFN-		882
GSN.CM.99.CN71	GRVW-EFLA-----A-IL-N-		887
DEN.CD.x.CD1	TR-A-EVVA-----IV-N-		865
LST.CD.88.447	KKRW-FRFRGRSGFPSETTETA--		916
LST.CD.88.485	KKRW-FRFRGRSGFPSETTETA--		915
LST.KE.x.lho7	KKRW-LRFGRSG--SSEATTETA--		913
LST.CD.88.524	NR-R-WRFRFRSGLPSETTETT--		917
SYK.KE.x.SYK173	RRVVE-VAAL---L-----IY-N-		833
SYK.KE.x.KE51	RRVAENVAAL---L-----IY-N-		841
COL.CM.x.CGU1	ARGW-RAPEYL-GWIYDRPQGA-		838

H1B.FR.83.HXB2	MGGKWSKSSVIGWPTVRERMR.	AEPAADRVGASDRLEKHGAITSSNTAATNA.	CWLAEQEE. VGFVPVTPQVPLRPMYKA	acidic cluster	poly-P helix	phos	HXB2 premature Nef end	131
H1A1.UG.85.U455	---K-RVE-E-K---	ETPA-KG-V-Q-D-Y-V-SS-S	---GD---	R	F	F	D	132
H1C.ET.86.ETH220	---TM-C-PV-PI-AI-I---	A-EK-V-Q-D-Y-L-P-N-PD	---E---	R	P	N	D	133
H1D.CD.84.84ZR085	---IV-AI-I-K---	TDRPERRR	---G-V---	R	R	---	T	141
H1F1.BE.93.VI850	---IV-A-G---	TPT-EG-V-DRR-RT-PD	---L---	R	V	---	---	132
H1G.SE.93.SE6165	---IV-E-I-N---	TPT-EG-V-Q-DR--N-PD	---DSE---	R	F	G	F	133
H1H.CF.90.056	---RMG-S-I---	V-EG-V-DRR-V-IN-S-RD	---A---	R	---	D	---	133
H1J.SE.93.SE7887	---N---	APA-G-V-Q-A----	---T---	K	I	---	---	129
H1K.CM.96.MP535	---IV-AI---	ARPAADRVTG-Q-V-Q-AR-V-SHN-PD	---F-GF---	R	---	---	---	141
H1O1.AE.TH.90.CM240	---S-IV-Q-KIK---	QTP-TEG-V-Q-D-ID-D	---V-R-D---	M	---	G	F	130
H1O2.AG.NG.x.IBNG	---IV-K-MK---	QTP-TG-Q-DR-Q-PD	---D-N---	R	---	G	---	132
H1N.CM.95.YBF30	---KI-LV-EI---	QTQEP-VEP-VGA-Q-ANR-IR-RDN-ES	---I---	E	---	I	---	138
H1O.BE.87.ANT170	---NALR-GKFE-AA---	TRTFPES-C-PG-QI-E-AAR-G-P-H-PQN	---L-F-SHQ---	A	---	G	F	138
H1O.CM.91.MVP5180	---NA-KFA-SE-D---	SSSDPQ-C-PG-V-E-ATR-G-S-H-PQN	---L-F-DSHKD-D---	R	---	F	F	137
CPZ.CM.05.SIVcpzEK505	---KV-LV-Q-ID---	QQDP-EG-V-Q-ANR-IR-KEN-Q	---L-E-K-A---	C	---	E	F	132
CPZ.CM.05.SIVcpzLB7	---S-LV-E-R-I-Q---	ETV-EG-P-Q-AR-PQ-ET	---I-E-K-A-IW-K---	C	---	F	E	132
CPZ.CM.05.SIVcpzMB66	---LV-E-D-L---	TQ-EG-PV-AR-R-SQ-ET	---L-EVON---	R	---	F	G	132
CPZ.CM.05.SIVcpzMT145	---A-LV-EI-IC-	PT-EG-V-Q-IASR-V-IR-PQ-ET	---L-EMOD---	K	---	F	G	132
CPZ.CD.90.ANT	---SA--TKW-A---	QH-TM-DI-PCVGE-ASR-L-TI-G-EKD	---	R	---	M	---	130
CPZ.CM.98.CAM3	---NR-LV-D-LSE---	LNLPP-PG-VQ-A-V-TRH-PQN-QT	---	L	---	EM	HNEE	134
CPZ.CM.98.CAM5	---N-LV-AA-Q-I-Q---	PT-SG-PV-Q-AR-TR-H.TKYQT	---	L	---	EM	NH-EE	133
CPZ.GA.88.GAB1	---T-LV-E-R-I-E---	TP-EG-EV-K-R-R-PE-QT	---	L	---	EM	D-N	132
CPZ.GA.88.GAB2	---N-IV-Q-LG---	TOE-EG-EV-Q-ARR-RH-PQ-QT	---	L	---	EM	Q-E	132
CPZ.TZ.01.TANI	---NIFGR-WP-ARAI-DL---	HMTSSEP-Q-Q-QNQK-GL-TNLT-G-S-D	---	L	---	V	LEYS-DHT	128
CPZ.US.85.CPZUS	---N-IV-E-N-L---	RQTOQA-EG-PV-Q-AE-TR-PQN-QT	---	L	---	---	DEMTHN-SE	136
H2A.GW.x.ALI	---ASG-KRSPLOGL-	LLQTPGETGGQCSGSGGYSQSG-SGRGQKLPSC	---	G	---	G	---	165
H2B.CL.x.EHO	---SAG-KQSKQ-GL-	LL-ARRGRPESSERQERLSQVPG-SDKGLNSPSCD	---	S	---	D	---	166
H2G.CL.x.ABT96	---SAG-KRAPQGL-L-K	LQARGPEYKGLWEGLEEGYSQV-EESGKGLSSLSCE	---	G	---	G	---	164
H2U.FR.96.12034	---SI-KHSKRSGL-	LLAARGENYGRWLGGLEDGFELYQ-SGDKLSSRFCEAQ	---	G	---	G	---	166
Nef start R17Y mutation premature stop in Mac239 max HIV-1 similarity								
MAC.US.x.239	---AI-MRSPSPDGL-Q-L	ARGETYGRLLGEVEDYGSQSPG-LDKGLSSLSCE	---	G	---	G	---	163
SMM.US.x.PBM53	---AAG-KQSRQGG-L	KLLQARGETHGKLEWLEDEGYSQS-GELGRWNLHSFE	---	G	---	G	---	164
STM.US.x.STM	---ASG-KQRKHGEL-	LL-ARGETYKLLLEGLGEGSQSASDKGLNSHSC	---	P	---	Q	---	164
MNE.US.x.MNE027	---AT-KRSKSPDGL-Q-L	ARGETYGRLLGEVEDYGSQSLG-SDKLSLSLSCE	---	G	---	G	---	164
DRL.x.x.FAO	---QN-KREEAYARFYK-L	QVGAKGNDLHYQLPESSFLPKRLLVPPFGGLD-K-QR-SSTSSLDTHM0AAAEKDVLP	---	E	---	E	---	162
MND_1.GA.x.MNDGB1	---SQ-KRSEA-VRYSAL-	GLVGGPVTDPGYKLETSQAQKSL-LI-GRAYTYSGL	---	D	---	K	---	163
MND_2.x.x.5440	---S-QREHYKYKAL-	GYGAEGTNLDYOVL-QR-PLA-SGTS-AASDKVLRPST	---	E	---	Y	---	154
MND_2.GA.x.M14	---S-QREY-YAKYKAL-	GYGAEGTNLDYOVL-QR-LLE-PLGTS-GASDKELK-CST	---	K	---	V	---	138
MND_2.CM.98.CM16	---S-QQEEKYKYYKA-	GYGAEGTNLDYOVL-PLL-L-TSQ-EFDREQ-SST	---	T	---	Y	---	134
TAL.CM.01.8023	---SVK-LLGETASQR	LQSGPGAGRLVYWRRLRDGQKROLRS-GERGKGLNIALLDAG-WOASDA-W	---	M	---	G	---	156
TAL.CM.00.266	---SVK-LLGETASQR	LQSGHGGGRMYWRRLRDGVQRGLSPGARGKDNLIALDEE-WOATGADY	---	M	---	G	---	156
TAN.UG.x.TAN1	---SN-REOQ-LLRLWRAL-	KAPVVRYGMLAD-PLTG0SSNTEEECDKNNMG-STRRGKST	---	P	---	G	---	149
TAN.FR.x.B87_14	---SN-KEQE-LLKMWRL-	KAPVVRYDMLADPL.LGTSSNTEEECDKNNWGG	---	L	---	T	---	150
VER.DE.x.AGM3	---LGN-PQHKKQLSLWALH	KTRATRYGLLADPLIGTSSNTEEECDKALK-SLI	---	R	---	R	---	151
VER.KE.x.9063	---LGN-PQHKKVSLWALH	KTVTRYGLLADPLIGTSS-VQEEYDKLRKLSLI	---	R	---	K	---	151
VER.KE.x.AGM155	---LGS-PQHKKQL-IWRAL	HATRTRYGLLADPLIGTSSNTEEECDKLRKLSLI	---	R	---	R	---	151
VER.KE.x.TYO1	---SQ-N-PAHKKYSLWALH	KTHVTRYGLLADPLIGTSSVQEECDKLRKLSLI	---	R	---	K	---	151
GRV.ET.x.GRL_677	---SSN-RQQQ-LLKLWRGL-	GKPGADWVLLSDPLIGTSSVQEECDKLRKLSW	---	K	---	G	---	148
SAB.SN.x.SABM1	---S-QQQRHSLWLSKL-	QAPVIQYDMLADPLIGTSSHTQECC-KSLRDGLI	---	R	---	Q	---	152
RCM.NG.x.NG411	---SAAGLRRWRGLMLT-PG-DYA-FAET---	LQDGQPRCAE-SGRASRDFLT	---	R	---	R	---	139
RCM.GA.x.GAB1	---S-N-AA-LLRW-FK-L	LTPGEGVYRHETLDDGQPWCAE-SGRASRDFVT	---	R	---	G	---	147
SUN.GA.98.L14	---NAFGRP-EV-VRTL	F-L-AGSGTRAEPAGREYHRLRRQSEVPL-VE	---	E	---	R	---	128
SUN.GA.x.SOL_36	---NAFG-P-EV-VRTL	FKL-AGSGTRAEP	---	R	---	R	---	127
BLU.KE.x.KE31	---STN-NOQCRSESPY	GTRWRRAKQYTLPDPELLKSPRSRSHG-FDKAWRSTLTPEVDPHGRDRDEHESGGQKWS	---	C	---	I	---	155
MON.NG.x.NG1	---S-X-XQPAEQLMPSHG	SOSSGQTRSFAMEDDYGEDSVLSPDASDRGRFSLTEGRNHRSNRRPTTV	---	P	---	C	---	146
MON.CM.99.L1	---S-N-QQ5PLS-PLL	GS05SGRMRYFMLEDDEYQWLSPOASDRGRKYSLTTEGRNKRORRP	---	L	---	D	---	142
MUS_1.CM.01.1085	---S-S-0PAQQLS	ISSPSPSGTGRKQYFKLVEEYGENSMLSPDASDRGRYSLTEGA-KRPVI	---	H	---	R	---	135
MUS_1.CM.01.CM1239	---S-N-QQEQOPL-	SSPSPGQSPYFRVLVDPEYGENSMLSPDASDRGRKYSLTTEGSNERPL	---	H	---	R	---	135
MUS_2.CM.01.CM2500	---S-N-KE	SOQLMGSPPGQTGPYFVKLVNNYKHSMLSPNAS-RGRKYSLTTEGSNKAKII	---	K	---	H	---	135
MUS_2.CM.01.CM1246	---S-N-PQSOERS-	PLRSSHTGPNRYFSLVKYGENSMLSPDAS-RGRKYSLTTEGNKRQII	---	H	---	C	---	135
DEB.CM.99.CM40	---N-TPSR-OVGS	FGSGS-GLLRWYRDLSEQEOEFSECLLESDRQSSSSTE	---	F	---	M	---	147
DEB.CM.99.CM5	---N-TPSF-QGKSGSGSK	GLLWRKYDLSGEAETPFPQESDKEQCSYSTE	---	E	---	P	---	149
GSN.CM.99.CM166	---S-N-QQQEES-	ALLSS-TG0RQPYFTLVDEYGENMFLSPDASDKGRYRYL	---	D	---	P	---	136
GSN.CM.99.CN71	---S-N-QQSQESS-	ALLSSHTGQSPFRLLDPEYGENSMLSPDASDRGRYRYL	---	E	---	P	---	135
DEN.CD.x.CD1	---S-N-QASGI-	GLAPSNKEGRQRKRYWRLLPEQEKWMPSP-SDRALKLSSTRLL	---	E	---	H	---	166
LST.CD.88.447	---NIFG-PAAD---	WVKTL-RLR-GAGTRSEGETEYQQLMQETSDWDRKRDDGSOS	---	E	---	R	---	125
LST.CD.88.485	---NIFG-PAAD---	WVKTL-RLR-GAGTRSEGETEYQQLMQETSDWDRKRDDGSOS	---	E	---	R	---	125
LST.CD.88.524	---NIFG-P-ADS---	WWRVL-KLRAGSNTAEGTEESYQKLMEQTGLWDKTEEDGSDOS	---	E	---	R	---	124
LST.KEx.lho7	---NAFGPAAD---	WVKTL-RFR-GSGRMAEGETSRYALTEKTSNNAKRDDGSOS	---	E	---	K	---	125
SYK.KE.x.KE44	---SVK-RMQQA-EDR-	L-TGWMKRKRYTTPDALLRSLPS-G-FDKAWRSTLTPEIDPHGRDR	---	W	---	G	---	155
SYK.KE.x.SYK173	---STKS--QQLSEKGYA	IGW-LFKQYTPLEPDLRSLQPC-G-FDKAWRSTLTPEIDPHGRDR	---	W	---	G	---	155
SYK.KE.x.KE51	---SQS-QQPSR-	DEKWRTRWPFKGPYSMPDELLRMSOPYHEDFRGRWRTLTPEIDPKRDF	---	S	---	G	---	152
COL.CM.x.CGU1	---SLL-CLNIROQ-QT-	GGVGLRFLGIYWPVQAFRLRHF-G-YDL-RVAGVERLNICEDGYMT	---	V	---	P	---	166

H1B.FR.83.HXB2
H1A1.UG.85.U455
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H1O1.AE.TH.90.CM240
H1O2.AG.NG.x.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CM.05.SIVcpzEK505
CPZ.CM.05.SIVcpzLB7
CPZ.CM.05.SIVcpzMB66
CPZ.CM.05.SIVcpzMT145
CPZ.CD.90.ANT
CPZ.CM.98.CAM3
CPZ.CM.98.CAM5
CPZ.GA.88.GAB1
CPZ.GA.88.GAB2
CPZ.TZ.01.TAN1
CPZ.US.85.CPZUS
H2A.GW.x.ALI
H2B.CLX.EHO
H2G.CLX.ABT96
H2U.FR.96.12034

MAC.US.x.239
SMM.US.x.PGM53
STM.US.x.STM
MNE.US.x.MNE027
DRL.x.x.FAO
MND_1.GA.x.MNDGB1
MND_2.x.x.5440
MND_2.GA.x.M14
MND_2.CM.98.CM16
TAL.CM.01.8023
TAL.CM.00.266
TAN.UG.x.TAN1
TAN.FR.x.B87_14
VER.DE.x.AGM3
VER.KE.x.9063
VER.KE.x.AGM155
VER.KE.x.TYO1
GRV.ET.x.GRL_677
SAB.SN.x.SAB1C
RCM.NG.x.NG411
RCM.GA.x.GAB1
SUN.GA.98.L14
SUN.GA.x.SOL_36
BLU.KE.x.KE31
MON.NG.x.NG1
MON.CM.99.L1
MUS_1.CM.01.1085
MUS_1.CM.01.CM1239
MUS_2.CM.01.CM2500
MUS_2.CM.01.CM1246
DEB.CM.99.CM40
DEB.CM.99.CM5
GSN.CM.99.CN166
GSN.CM.99.CN71
DEN.CD.x.CD1
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.x.lho7
SYK.KE.x.KE44
SYK.KE.x.SYK173
SYK.KE.x.KE51
COL.CM.x.CGU1

VRYP1LTFGWGCKLVPVEPKIEEANKGENTS...LLHPVSLHGMD...PE.REVLEWRFSRLAFHH.VARELHPEY..FKNC...
I-----D-AEV--TG--N-----ICQ--V-----E--K--M-K--T--LK--R-Y---F--Y-D
-----F--D-SEV--I-E--NC-----A-----ED---K-K--H--RR--M-----Y-D
-----FE--D-EVV-K-TE--DNC-----ICQ--E-----E--K--V--N-----E--K-K-KY
I-----L--F--D-EEV-K-E--NC-----M-Q--E-----ED---R-K--S--LR--I--R--F--YQD
T-F-----MD-AEV-----N-----ICQ--E-----ED---V--S--RR--I-----Y-D
E-F-----F--N-QEV-Q-E--N-----M-----E--DG---M-K-----LT--L-VK-----D
I-----D-SEV--E--NC-----ICQ--IE-----E--Q-K--S--RR--I-----F--Y-D
I-----D-AEV--TTE--DNC-----INQ--E-----EH--I-M-K--S--RR-----D--Y-D
I-F--C--F--DQREV--D-----NC-----M-Q--IE-----E--M-K--A--RK-----Q-----Y-D
T-F-----F--MD-AEV--E--N-----ICQ--E-----ED---I-----R--T-----Y-D
I--V-----F--LSAEV--E--D-NA-----ICQ--A-----DH-K--V--S--RR-----F--Y
T-F-----LF-----SEEEA-RLGNTCERA.N-----ACA--FE-----TH-K-I-M-K--RS-GNT--MIT--L--QKD
P-F-----LF-----SAEEA-RLGNTNEDA.S-----ACN--AE-----AH-G-I-K-Q--RS-GLT--I-LQK--L--PK
-----F--LS-EEV--Q-D-NV-----MCQ--E-----D-K--V-----S--RV-RAR-----F--YQ
I--I--G-F--LT-EEV-----NL-----ICQ--E-----ED-K--I-KY--Q--LR-IAR--S--Y
I-----F--L--EEV-R--E-D-NI-----ICQ--QE-----EA--V-T-----LK-RAR-----Y-D
-----F--LP-EEV-K-E--NC-----MCQ--E-----E--V-T--K--LR-LAR--K-----YRD
-----CR--F-----SPDD...D-RNI-----ACT--DG-----H-K-I-R-E--AS-MRR--I--R-----RD
-----Y-G-F--LT-EEV-Q-E-D-NI-----ICQ--E-----AH-G--V-----LR--I--Q-----YQN
-----F--LT-EEV-K-E-DTNI-----ICQ--E-----EH-G--V-----LK--L--Q-----YGDHNPAPAQ
T-F--C--F--LTEEQV-Q-E-D-NC-----ICQ--E-----ED-K--V-----LR--I--Q-----YKD
-----Y--F--S-EAV-----D-NL-----LCT--FE-----E--K--I-K--Q--LR--L--K-----YRD
I--CR--LF-----D-PED...DEK...I--ACS--TT-----D-G-T-I-----S--RR--I--RY-----FK
-----Y--F--LTEEV-Q-----TNI-----MCQ--E-----EH-G--I-Q--TE--RR--R-K-----FRN
I--MF--LW-----DVPQEG-DTET..HC-----QTSRH...G-T-V-----PK--HDY--K-FI-----E-FGYK.SGLPEDEW...KARLKARG.IPFSKNRNS...
-----KF--LW-----NMIAEP-DEET..HC--V--AQTSAW-----H-E-T-V-Q--L--YDY--F--FRF--E-FGYK.SGMPEKEW...KAKLRARGIPTE...
I--X--LW-----TVSDEAQDET..HC--V--AQTSPW-----T--A-K--PT--YDY--R-FI-----E-FRWK.SGLPEAVW...KEKLRQGLPIE...
T-W-MC--LW--E--DANDKAQEDER..LY--VGSQAQTSCEE...HM-G-A-V-K--S--YSY-Q-FIKC--E-FGSK.SGLSEEEV...KRRLTARGLPKVNC...

max HIV-1 similarity

I--K--LW-----NVSDEAQDEE..HY..M--AQTSQW-----W-G--A-K--PT--YTY..E-YVRY--E-FGSK.SGLSEEEV...RRRLTARGLLN..MADKKETR..
I--MF--LW-----SASDEAQEDT..HY..V--AQISQW-----W-G--A-K--Q--YRY..E-FIR--E-FGSK.SGLSEEEV...KRRLTARGLLK..MADKKEETS..
I--KQ--LW-----DMSNEAQEDD.T.HY..V--AQT-QW-----W-G--V-K--PL--HTY..E-FVR--E-FGSK.SGLPKEEV...ERRLTARGLLK..MADKKEETS..
P--K--LW-----NVSDEAQEGEE..NY-----AQTSQW-----W-G--V-K--PT--YTY..E-YIRY--E-FGSK.SGLSEEEV...RRRLTARGLLK..MADKKEETS..
I--KM--FLF-----V-EDLVNQCNR...N--SSQMGPI-----G-R-M-Q--PG--YTFQAC.IH--E-FGHVKS.IKSVEEPG...CCKRKKWQFKPTTEGCHDNLQKC-
T--K--FLF--E--SRAIGD-YAANN.H...S--SQ-CPOE-----G-T-M-SGTLI-PMTL...Q
I--K--FLF-----AVSPLL-EDENR...N--SSQMGIQE-----G-R-M-K--G--YTF.Y-PIIR--E-YKCV.TSLSYEAYKKEEPCDCKRKKWQF
I--KR--FLF--I--QVPPDQ-NOECNR...N--SSQ-GIQE-----SMXG-R-M-K--PE--YTF.Y-PIK--KE.YRHV.TSLSYAYQOE..KSDCCKRKKWQF
I--K--RFLF--C--AVPPDQ-N-ECNK...N--SSQ-GIQE-----SLXG-R-V-K--G--YTF.Y-PIK--GE.FKHV.QSISHIAYAKEHPECCCKRKKWQF
T--C--ILW--C--IHEDDTEGDH...L-M--AYDGOQE-----W-G-A-V-V--EK--YTPGAKMAEYDRLE...RKKKELLAPPOTASS
T--C--ILW--C--IHEDDTEGDH...L-P--AYDGOQE-----W-G-S-V-V--EK--YTPGAKMAEWRDLE...REKRMLLAPPOTASS
T-K-RC--F-FE-----DVSQEAQD.ER..HC-----AQTEVES-----W-K-T-V-K-NPL--VQY.NPDSFKDMH..GLV...KRK
I-K-KC--F-FE-----RLSEAAAT-ER..HC-----AQTNV-----W-G-T-V-K-NPL--VQYNPDCFKDMHG.LV...KRK
I--RC--F-F-----DLHEEA-TCER..HC--V--AQVREDP-GI..NH.G--V-K--PM--VQY.DPNRKYLD.MHDL.G...KRK
I--RC--F-F-----MHHEA-TCER..HY..A--AQIKEDP-GI..SH.G-T-V-K--PM--VQY.DPNRQYF-D.MHAI.V...KRK
I--RC--F-F-----ALHEEA-TCER..HC--V--AQ--EDP-GI..NH.G-I-A-K--PM--VQY.DPSREYFTD.LYST.VGTGN...KRK
I-F-KC--F-F-----DLHEEAQTCER..HC--V--AQMGEDP-GI..SH.G-I-V-K--PM--IQY.DPNREYFTD.MHGL.V...KRK
I--RV--F-F-----DLHEEARNCER..HC..M--AQMGEDP...GIDH.G--V-K--PK--VEY.RPDMFKDMH.EHA...KR
T--KC--F-----DLSEEAKNSEN..HC-----AQVAYE-----AW-K-T-V-K--PL--VDY--WR---Q.VPSA.QG...KNTK
I--TM--LWQ-----DVSDEATE-ER..HC-----AETS-----W-G--A-K-KPL--VDY.AGYR---F.FGER...KNTK
I--TL--LWQ-----DVSDEAREDEEHS...N--AETS--E-----W-G--A-K-NPM--VDY.IGYR---F.FGER...KNTQ
-----KA--LW--A--TI-EDRDP...HPCQA--SSQGVNE-----W-G-R-I-T--PT--YDF.R-IQK--E-FKHVTSLQWEV
P--K--LW--M--TI-EDRGP-HP...COA--SSQGVNE-----SW-G-R-I-T--S--YDFKAIQHP-E-F.KHVRSLQWEAD
I-F--L--LWE-----NDIEGYLDEEDTL...L--A-GK-AEED...H-G-N-M-N-NPH--YTPGWEMARLOLE.ROTKPKQLKKSALK
-----TIP--FLWC-R--AIQEDS-DGDX-F..L-T--AYQGRDE-----H.KQF-VXS-C-K-GVKSQGLDDELQOE...ERKRRLTANRIL
-----TMP--FLWC-R--ATTEDS-EGDEDF..L-T--AYQGR-E-----H-QF-VFS-C-K-VKSGRQLAQLOQE...ERKKRLAANRIL
T--KIP--FLWC-RE-AMTEGS-EGDDQY..L-D--AFQGR-E-----HH.KQF-VFS-C--LKSGLQLDOMQOE...ERKKRLATNHIL
T--TLP--FLWC-RE-AITEDS-EGDE-Y..L-T--AAYQKREE-----H.KQF-VFS-C--MKSQRQLDQIQOE...ERKRRLTANRIL
-----MP-VLFC-R--HITENS-LGDD-H..L-C--AFITGEE-----H.K-F-VFS-C--LKSQRQLNMQOE...ERKRRLAANRIL.SQSP
T--ROP--FLWC-R--AMTEES-PGDDQY..L-T--AYVGRDE-----HH.K-F-VFS-C--LKSQGLNLIQOQ...ERKRRLTANRIL
-----IY--FLF-----IADPDYE-DE..RN.I--DAHOGQ-E-PH...R-V-K--T--YCY.K-GHAEQRE.HTRR-MFP...KRK
P--TY--FLF-----IADPDYE-DE..RN.I--DAHOGQ-E-PH...K-R-V-K--S--YCY.K-GHEAHKPEHNR-MFP...KRK
T--TIP--F-IC-R--ATTEDS-PGDD-Y.L.-T--AYQGRSE...QH.K-F-VFS-C-K--IKSGIQLDQLOQE...ERKMRLTANRFL
I--TMP--F-WC-R--AMTEDS-PGDD-QYL..L-N--AYQGOE-----HH.-I-VFS-C--LKSQGLNLIQOQE...ERKKRLTANRFL
I--E--FLW-----TIEEYDNKEE..NC-----DRYEQOQA-----W--T-V--PE-WCYKAGHKLQTSOHNNY-ORA...KPRK
T-F--A--LWQ-E--AC-EYTDPSDYR.QC...SSQ-GVQE-----W-G-R-I-H-NPM--VDY--LRKQ-KS.IQASAIGL...QRK
T-----LWQ-E--AC-EYTDPSDAS.QC...SSQ-GVQE-----W-G-R-I-H-NPM--VDY--LRKQ-KS.IQASAIGL...NCKRK
T--A--LWQ-E--AC-EFKDPSDER.QC...SSQ-GVLE-----W-G-R-I-H-NPM--VDY--LRKQ-AS.VOATAMRF...NCKRK
T-----LWQ-E--AC-EYKDPSET.QC...SSQ-GVLE-----W-G-R-I-H-NPM--VDY--L-LKKQ-AK.IQNTAFAP...DCKRK
T--I--LWE-A-N-IEGYLSDDEEDTL...L--AAGK-ASED...I.H-G-N-I-N-N-H--YTPGWELARRQLE.AQTGKPTQ...VKQALTKGGS
T--I--LWE--N-IEGCL-YEEHT...L--A-GQ-SSS.MGE-H.V-LQPPPGYTPGW.E...MARLOLE.ROTG...KPOLQSALS.K.NIS
T--I--LWE-A-N-IEGYLVEEDT...LMM--AAGV-ASE...D-H--N-M-N-NPH--YTPGWEMARQLE.ROTG...KR--L.TAETTYIS
-----E-GALKEDRKL-AG-LW.KLVYIEQ.LGEYAYSID.LSLLSVTS...RKKKKPQQAIVEMVD

normal Nef end

normal Nef end

205
206
208
216
206
208
207
204
214
205
207
213
214
212
207
207
207
206
199
208
214
206
206
196
210
263
239
257
261