

Figure 1S. HIV-1 consensus proteome where variable residues are masked with the "." symbol

>ENV

MRV.....N.....W.WG.....C.....LWVTVYYGVPVW..A.TTLFCASDAKAY..E.HNVWATHACVPTDPNPQE
..L.NV.FNMWKN.MV.QMH.DIISLWDQSLKPCVKLTPLCVTL.....
NCSFN.TT...DK.....ALFY..D.V.....YRLI.CNTS.ITQACPKV.F.PIPIHYCAPAGFAILKC...F
NGTG.C.NVSTVQCTHGI.PVVSTQLLNGSLA.....IRS.N.T.N.K.IIV.L...V.I.C.RP..N.....PG..
.Y.....G.IR.A.C.....W...L.....L.....I.F.....GGD.EI..H.FNC.GEFFYCNT..LFN.....
.....I.L.CRIKQIIN.WQ.VG.AMYAPPI.G.I.C.SNITGLLLTRDGG.....
..FRP.GGDMRDNRSELYKYKV.V.I.PLG.APT.AKRRVV.RE..AVG..GA.FLGFLGAAGSTMGAAS.TLTVQAR.LLSGIVQ
QQ.NLLRAIEAQQHLL.LTVWGIKQLQARVLA.ERYL.DQQLLGIWGCSSGK.ICTT.VPWN.SWS.....IW.NMTW..W..EI
.NYT..IY.LIE.SQ.QQE.NE..LL.LD.W..LW.WF.IT.WLWYIKIFIMIVGGLIGLRI.F.VLS.VNRVRQGYSPLSFQT..
P.....DRP..IEE.GGE.DR.RS.RLV.GFL.L.W.DLR.LCLFSYH.LRD..LI..R.VE.LG.....GWE.LKY...L.
.YW..ELK.SA..L...AIAVA..TDR.IE..QR..RA...IP.RIRQG.E.AL.

>TAT

MEPVDP.LEPW.HPGSQP.T.C..CYCK.CC.HC..CF..KGLGISYGRKKR..RR..P.....HQ....KQP.....TG..
E.KK.V.....

>VPR

MEQAPEDQGPOREP.NEW.LELLEELK.EAVRHFP.R.WLH.LGQ.IY.TYGDTW.GVEAIR.LQQLLF.HFRIGC.HSRIGI...
.RR.RNG..RS

>REV

MAGRSG..DE.LL...R.IK.LYQSNP.P...GTRQARRNRRRRWR.RQ.QI..IS.RIL...LGR..EPVPLQLPP.ERL.L...
E..G..GT.....E....L.....

>VPU/VPX

M.....L.VA..I..I.VW.I..IE.....Q.KI..L..RI.ERAEDSGNES.GD.EEL.....G.....

>NEF

MGGKWSK...GW...RERM.....AA.GVGA.S.DL...GA.TSSNT...NA.CAWLEAQE...E.VGFPVRPQVPL
RPMYK.A.DLSHFLKEKGGLEGLI.S..RQDILDWVYHTQGYFPDWQNYTPGPG.R.PLTFGWCFKLPV.P..VEE
AN.GENN.LLHP...HGM.D.E.EVL.W.FDS.LA..H.ARE.HPEYYK.C

>POL

FFRE.LAF.QGEAR.F..EQ.....R.L.....E.G.....G.....FPQITLWQRPLV..KIGGQ
..EALLDTGADDTVLE...LPG.WKPKMIGGIGGFIVRQYDQI.IEICG.KAIGTVLVGPTPVNIIGRN.LTQIGCTLNFPISPI
ETVPVKLKPGMDGPKVKQWPLTEEKIKAL..IC.EMEKEGKISKIGPENPYNTPVFAIKKDKSTKWRKLVDFRELNKRTQDFWEVQ

LGIPHPAGLKKKSVTVLVDGDAYFSVPLD..FRKYTAFTIPS.NNETPGIRYQYNVLPQGWKGSFAIFQ.SMTKILEPFR..NP.
.VIYQYMDLDLYVGSLEIGQHR.KIEELR.HLL.WGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPI.LP.K.SWTVNDIQKLVGK
LNWASQIY.GIKVKQLCKLLRG.KALTD.V.IT.EAELELAENREILK.PVHGVYYP.KDL.AE.QKQG..QWYQIYQEP.KNL
KTGKYA..R.AHTNDVKQLTE.VQKI..ESIVIWGK.PKF.LPIQKETWE.WW.EYWQATWIPEWFEVNTPLVWKLWYQLE.EPI.
GAETFYVDGAANRETK.GKAGYVTDGRQK.V.L..TTNQKTEL.AI.LALQDSG.EVNIVTDSQYALGIIQAQPD.SESE.V.QI
IE.LI.KE.VYL.WVPAHKGIGGNEQVDKLV.S.GIRKVLFLDGIDKAEHE.YH.NWRAMASDFNLPPIVAKEIVASCDCQKLG
EAMHGQVDCSPGIWQLDCTHLEGK.ILVAVHVASGYIEAEVIPAETGQETAYF.LKLAGRWPVK.IHTDNG.NFTS..VKAACWWA
.I.QEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDIIATDIQTKELQKQI.K
IQNFRVYRDRDP.WKGPALLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYQKQ MAG.DCVA.RQDED

>VIF

MENRWQVMIVWQVDRMRI.TW.SLVKHHMY.S.KA..W.YRHYES.HPR.SSEVHIPLG.A.LV..TYWGL.TGER.WHLG.GVS
IEWR..Y.TQVDP.LADQLIH..YFDCF..SAIR.AILG..V.P.C.Y.AGHNKVGSQYLAL.AL..P...KPPLPS
V.KL.EDRWKPKQT.G.RG.HTMNGH

>GAG

MGARASVL.GG.LD.WEKIRLRPGGK.Y..KHLVWASRELERFA.NP.LLET.EGC.QI..QLQP.L.TGSEEL.SL.NTVATLY
CVH..I.V.DTKEAL.K.EE.Q.....K.....SONYPIVQN.QGQMVHQ..SPRTLNAWVKVIEEKAFSPE
VIPMF.ALSEGATPQDLN.MLN.VGGHQAAMQMLKDTINEEAAEWDR.HP..AGP..PGQREPRGSDIAGTTSTLQEQI.WMT.N
PPIPVG.IYKRWIILGLNKIVRMYS.SILDIRQGPKEPFRDYVDRFFKTLRAEQATQ.VKNWMTETLLVQANANPDCK.ILKALG.
GATLEEMMTACQGVGGP.HKARVLAEMSQ.....IMMQR.NF.....KCFNCGKEGH.ARNCRAPRKKGCWKCG.EGHQM
KDCTERQANFLGKIWPS.KGRPGNF.Q.R..PTAPP.E.....EE.....KQ.....PL.SLKSFLG.DP...