Supplemental Table S1

Predicted cathepsin cleavage sites in MN-rgp120

Cathepsin	Cleavage Site	P4	P3	P2	P1		P1'	P2'	P3'	P4'
L										
	45-46	P	V	W	K		E	A	T	T
	96-97	N	M	W	K		N	N	M	V
	183-184	L	L	Y	K		L	D	I	E
	197-198	T	S	Y	R		L	I	S	C
	471-472	E	I	F	R		P	G	G	G
	482-483	D	N	W	R		S	E	L	Y
	487-488	E	L	Y	K		Y	K	V	V
	489-490	Y	K	Y	K		V	V	T	I
S										
	354-355	S	K	L	K		E	Q	F	K
	421-422	C	K	I	K		Q	I	I	N
	440-441	P	P	I	E		G	Q	I	R
D										
	215-216	K	I	S	F		E	P	I	P

Predicted cathepsin L, S and D cleavage sites in MN-rgp120. The location of cathepsin cleavage sites and flanking sequences was predicted using the method of Boyd et al. (8) and cleavage specificity algorithms for cathepsin L and S from Choe et al. (19) and cathepsin D from Scarborough et al. (85). The scissille bond, located between the P1 and P1' residues, and the flanking residues are listed according to the nomenclature of Schechter and Berger (86).

Supplemental Table S2

Observed cathepsin cleavage sites in MN-rgp120

Cathepsin	Site	P4	P3	P2	P1	P1'	P2'	P3'	P4'
L									
	327-328	K	N	I	K	G	T	I	R
	431-432	Q	K	V	G	K	A	M	Y
	435-436	K	A	M	Y	A	P	P	I
S									
	208-209	V	I	T	Q	A	C	P	K
	261-262	P	V	V	S	T	Q	L	L
	322-323	A	F	Y	T	T	K	N	I
	435-436	K	A	M	Y	A	P	P	I
D									
	181-182	Y	A	L	L	Y	K	L	D
	274-275	A	E	E	E	V	V	I	R

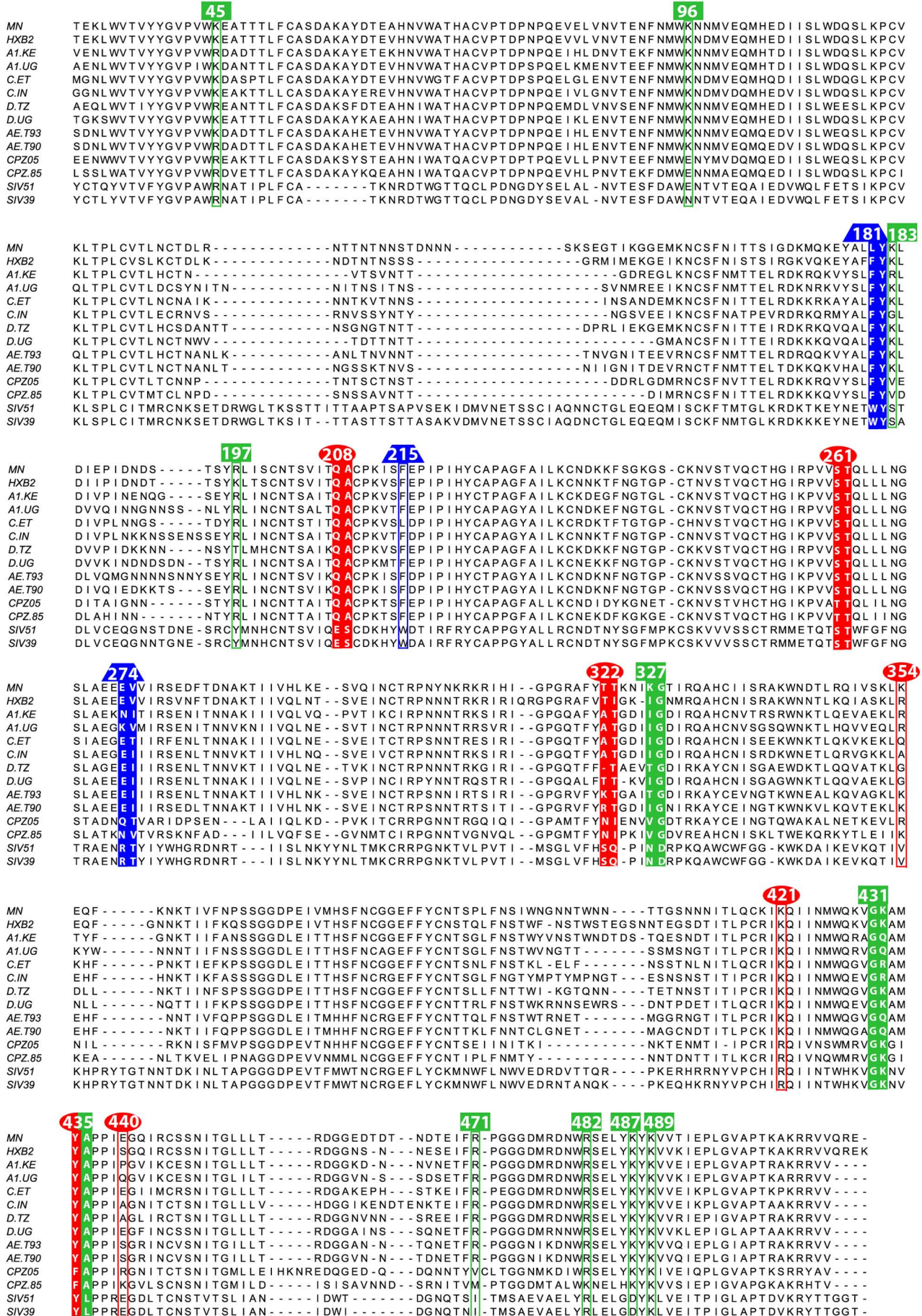**Experimentally determined cathepsin L, S and D cleavage sites in MN-rgp120.** The location of cathepsin cleavage sites and flanking sequences was determined for MN-rgp120 by Edman sequence degradation of peptides recovered after protease digestion (see Material and Methods). The scissille bond, located between the P1 and P1' residues, and the flanking residues are listed according to the nomenclature of Schechter and Berger (86).

Supplemental Figure S1. Alignment of predicted and observed cathepsin L, S and D cleavage sites

Position markers: 45, 96, 181, 183, 197, 208, 215, 261, 274, 322, 327, 354, 421, 431, 435, 440, 471, 482, 487, 489

```
                45                                        96
MN      TEKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEAHNVWATHACVPTDPNPQEVELVNVTENFNMWKNNMVEQMHEDIISLWDQSLKPCV
HXB2    TEKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCV
A1.KE   VENLWVTVYYGVPVWRDADTTLFCASDAKAYETEKHNVWATHACVPTDPNPQEIHLDNVTEKFNMWKNNMVEQMHTDIISLWDQSLKPCV
A1.UG   AENLWVTVYYGVPIWKDANTTLFCASDAKAYDTEVHNVWATHACVPTDPSPQELKMENVTEEFNMWKNNMVEQMHTDIISLWDQGLKPCV
C.ET    MGNLWVTVYYGVPVWKDASPTLFCASDAKAYDTEAHNVWGTFACVPTDPSPQELGLENVTENFNMWKNDMVEQMHQDIISLWDQGLKPCV
C.IN    GGNLWVTVYYGVPVWKEAKTTLLCASDAKAYEREVHNVWATHACVPTDPNPQEIVLGNVTENFNMWKNDMVDQMHEDVISLWDQSLKPCV
D.TZ    AEQLWVTIYYGVPVWREANTTLFCASDAKSFDTEAHNIWATHACVPTDPNPQEMDLVNVSENFNMWKNNMVEQMHEDIISLWEESLKPCV
D.UG    TGKSWVTVYYGVPVWKEATTTLFCASDAKYKAEAHNIWATHACVPTDPNPQEIKLENVTENFNMWKNNMVEQMHEDIISLWEQSLKPCV
AE.T93  SDNLWVTVYYGVPVWKDADTTLFCASDAKAHETEVHNVWATYACVPTDPNPQEIHLENVTENFNMWKNNMVEQMQEDVISLWEQSLKPCV
AE.T90  SDNLWVTVYYGVPVWKDADTTLFCASDAKAHETEVHNVWATYACVPTDPNPQEIHLENVTENFNMWKNNMVEQMQEDVISLWEQSLKPCV
CPZ05   EENWWVTVYYGVPVWREAKTTLFCASDAKSYSTEAHNIWATQACVPTDPTPQEVLLPNVTEEFNMWENYMVDQMQEDIISLWEQSLKPCV
CPZ.85  LSSLWATVYYGVPVWRDVETTLFCASDAKAYKQEAHNIWATQACVPTDPNPQEVHLPNVTEKFDMWENNMAEQMQEDIISLWDQSLKPCI
SIV51   YCTQYVTVFYGVPAWRNATIPLFCA-------TKNRDTWGTTQCLPDNGDYSELAL-NVTESFDAWENTVTEQAIEDVWQLFETSIKPCV
SIV39   YCTLYVTVFYGVPAWRNATIPLFCA-------TKNRDTWGTTQCLPDNGDYSEVAL-NVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV

                                                                          181 183
MN      KLTPLCVTLNCTDLR-------------NTTNTNNSTDNNN-----------SKSEGTIKGGEMKNCSFNITTSIGDKMQKEYALLYKL
HXB2    KLTPLCVSLKCTDLK-------------NDTNTNSSS----------------GRMIMEKGEIKNCSFNISTSIRGKVQKEYAFFYKL
A1.KE   KLTPLCVTLHCTN----------------VTSVNTT--------------------GDREGLKNCSFNMTTELRDKRQKVYSLFYRL
A1.UG   QLTPLCVTLDCSYNITN-----------NITNSITNS--------------SVNMREEIKNCSFNMTTELRDKNRKVYSLFYKL
C.ET    KLTPLCVTLNCNAIK------------NNTKVTNNS--------------INSANDEMKNCSFNITTELRDKKRKAYALFYKL
C.IN    KLTPLCVTLECRNVS-----------RNVSSYNTY----------------NGSVEEIKNCSFNATPEVRDRKQRMYALFYGL
D.TZ    KLTPLCVTLHCSDANTT----------NSGNGTNTT-------------DPRLIEKGEMKNCSFNITTEIRDKRKQVQALFYKL
D.UG    KLTPLCVTLNCTNWV-----------TDTTNTT----------------GMANCSFNITTEIRDKKKQVQALFYKL
AE.T93  QLTPLCVTLHCTNANLK-----------ANLTNVNNT-----------TNVGNITEEVRNCSFNMTTELRDRQQKVYALFYKL
AE.T90  KLTPLCVTLNCTNANLT----------NGSSKTNVS-----------NIIGNITDEVRNCTFNMTTLTDKKQKVHALFYKL
CPZ05   KLTPLCVTLTCNNP------------TNTSCTNST--------------DDRLGDMRNCSFNVTTELRDKKRQVYSLFYVE
CPZ.85  KLTPLCVTMTCLNPD------------SNSSAVNTT-------------DIMRNCSFNITTELRDKKKQVYSLFYVD
SIV51   KLSPLCITMRCNKSETDRWGLTKSSTTITTAAPTSAPVSEKIDMVNETSSCIAQNNCTGLEQEQMISCKFTMTGLKRDKTKEYNETWYST
SIV39   KLSPLCITMRCNKSETDRWGLTKSIT--TTASTTSTTASAKVDMVNETSSCIAQDNCTGLEQEQMISCKFNMTGLKRDKKKEYNETWYSA

                   197      208   215                             261
MN      DIEPIDNDS-----TSYRLSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFSGKGS-CKNVSTVQCTHGIRPVVSTQLLLNG
HXB2    DIIPIDNDT-----TSYKLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGP-CTNVSTVQCTHGIRPVVSTQLLLNG
A1.KE   DIVPINENQG----SEYRLINCNTSAITQACPKVSFEPIPIHYCTPAGFAILKCKDEGFNGTGL-CKNVSTVQCTHGIRPVVSTQLLLNG
A1.UG   DVVQINNGNNSS--NLYRLINCNTSALTQACPKVTFEPIPIHYCAPAGYAILKCNDKEFNGTGL-CKNVSTVQCTHGIRPVVSTQLLLNG
C.ET    DIVPLNNGS-----TDYRLINCNTSTITQACPKVSLDPIPIHYCAPAGYAILKCRDKTFTGTGP-CHNVSTVQCTHGIKPVVSTQLLLNG
C.IN    DIVPLNKKNSSENSSEYRLINCNTSAITQACPKVTFDPIPIHYCAPAGYAILKCNNKTFNGTGP-CNNVSTVQCTHGIKPVVSTQLLLNG
D.TZ    DVVPIDKKNN----NSYTLMHCNTSAIKQACPKVSFEPIPIHYCAPAGFAILKCKDKNFNGTGP-CKKVSTVQCTHGIRPVVSTQLLLNG
D.UG    DVVKINDNDSDN--TSYRLINCNTSAITQACPKMTFEPIPIHYCAPAGFAILKCNEKKFNGTGP-CKNVSTVQCTHGIKPVVSTQLLLNG
AE.T93  DLVQMGNNNSNNYSEYRLINCNTSVIKQACPKISFDPIPIHYCTPAGYAILKCNDKNFNGTGP-CKNVSSVQCTHGIKPVVSTQLLLNG
AE.T90  DIVQIEDKKTS---SEYRLINCNTSVIKQACPKISFDPIPIHYCTPAGYAILKCNDKNFNGTGP-CKNVSTVQCTHGIKPVVSTQLLLNG
CPZ05   DITAIGNN------STYRLINCNTTAITQACPKTSFEPIPIHYCAPAGFALLKCNDIDYKGNET-CKNVSTVHCTHGIKPVATTQLILNG
CPZ.85  DLAHINN-------NTYRLINCNTTAITQACPKTSFEPIPIHYCAPPGFAILKCNEKDFKGKGE-CKNVSTVQCTHGIKPVVTTQLIING
SIV51   DLVCEQGNSTDNE-SRCYMNHCNTSVIQESCDKHYWDTIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFGFNG
SIV39   DLVCEQGNNTGNE-SRCIMNHCNTSVIQESCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFGFNG

           274                               322 327              354
MN      SLAEEEVVIRSEDFTDNAKTIIVHLKE--SVQINCTRPNYNKRKRIHI--GPGRAFYTTKNIKGTIRQAHCIISRAKWNDTLRQIVKSKL
HXB2    SLAEEEVVIRSVNFTDNAKTIIVQLNT--SVEINCTRPNNNTRKRIRIQRGPGRAFVTIGK-IGNMRQAHCNISRAKWNNTLKQIASKLR
A1.KE   SLAEKNITIRSENITNNAKIIIVQLVQ-PVTIKCIRPNNNTRKSIRI--GPGQAFYATGDIIGDIRQAHCNVTRSRWNKTLQEVAEKLR
A1.UG   SLAEGKVMIRSEINNNVKNIIVQLNE--SVTINCTRPNNNTRSVRI--GPGQTFYATGDIIGDIRQAHCNVSGSQWNKTLHQVVEGLR
C.ET    SIAEGETIIRFENLTNNAKIIIVQLNE--SVEITCTRPSNNTRESIRI--GPGQTFYATGDIIGDIRQAHCNISEEKWNKTLQKVKEKLQ
C.IN    SLAEGEIIIRSENLTNNVKTIIVHLNQ--SVEIVCTRPNNNTRKSIRI--GPGQTFYATGDIIGDIRQAHCNISRDKWNETLQRVGKKLA
D.TZ    SLAEEEIIIRSENLTNNVKTIIVQLNE--TVKINCTRPNNNTRKGIRI--GPGQTFF-TAEVTGDIRKAYCNISGAEWDKTLQQVATKLG
D.UG    SLAEEEIIIRSENLTNNAKIIIVQLNE--SVPINCIRPYNNTRQSTRI--GPGQALFTT-KVIGDIRQAHCNISGAGWNKTLQQVAEKLG
AE.T93  SLAEEEIIIRSENLTNNAKTIIVHLNK--SVEINCTRPSNNIRTSIRI--GPGRVFYKTGAITGDIRKAYCEVNGTKWNEALKQVAGKLK
AE.T90  SLAEEEIIIRSENLTNNAKTIIVHLNK--SVEINCTRPSNNTRTSITI--GPGRVFYRTGDIIGDIRKAYCEINGTKWNKVLKQVTEKLK
CPZ05   STADNQTVARIDPSEN---LAIIQLKD--PVKITCRRPGNNTRGQIQI--GPAMTFYNIENVCDTRKAYCEINGTQWAKALNETKEVLR
CPZ.85  SLATKNVTVRSKNFAD---IILVQFSE--GVNMTCIRPGNNTVGNVQL--GPGMTFYNIPKIVCDVREAHCNISKLTWEKQRKYTLEIIK
SIV51   TRAENRTYIYWHGRDNRT---IISLNKYYNLTMKCRRPGNKTVLPVTI--MSGLVFHSQ-PINDRPKQAWCWFGG-KWKDAIKEVKQTIV
SIV39   TRAENRTYIYWHGRDNRT---IISLNKYYNLTMKCRRPGNKTVLPVTI--MSGLVFHSQ-PINDRPKQAWCWFGG-KWKDAIKEVKQTIV

                                                              421        431
MN      EQF------KNKTIVFNPSGGGDPEIVMHSFNCGGEFFYCNTSPLFNSIWNGNNTWNN----TTGSNNNITLQCKIKQIINMWQKVGKAM
HXB2    EQF-----GNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTEGSNNTEGSDTITLPCRIKQIINMWQKVGKAM
A1.KE   TYF------GNKTIIFANSSGGDLEITTHSFNCGGEFFYCNTSGLFNSTWYVNSTWNDTDS-TQESNDTITLPCRIKQIINMWQRAGQAM
A1.UG   KYW------NNNTIIFNSSSGGDLEITTHSFNCAGEFFYCNTSGLFNSTWNVNGTT-------SSMSNQTITLPCRIKQIINMWQVGQAM
C.ET    KHF------PNKTIEFKPSSGGDLEITTHSFNCGGEFFYCNTSNLFNSTKL-ELF-----NSSTNLNITLQCRIKQIINMWQGVGRAM
C.IN    EHF------HNKTIKFASSSGGDLEITTHSFNCRGEFFYCNTSGLFNGTYMPTYMPNGT---ESNSNSTITIPCRIKQIINMWQEVGRAM
D.TZ    DLL------NNKTIINFSPSSGGDPEITTHSFNCGGEFFYCNTSLLFNTTWI-KGTQNN---TETNNSTITIPCRIKQIINMWQGVGKAM
D.UG    NLL------NQTTIIFKPSSGGDPEITTHSFNCGGEFFYCNTTRLFNSTWKRNNSEWRS---DNTPDETITLQCRIKQIINMWQEVGKAM
AE.T93  EHF------NNTIVFQPPSGGDLEITMHHFNCRGEFFYCNTTQLFNSTWTRNET------MGGRNGTITLPCRIKQIINMWQGVGQAM
AE.T90  EHF------NKTIIFQPPSGGDLEITTHSFNCRGEFFYCNTTQLFNTTCLGNET-------NSTNLNITLPCKIKQIINMWQGVGQAM
CPZ05   NIL-------RKNISFMVPSGGDPEVTNHHFNCGGEFFYCNTSEIINITKI----------NKTENMTI-PCRIRQIVNSWMRVGKGI
CPZ.85  KEA------NLTKVELIPNAGGDPEVVNMMLNCGGEFFYCNTIPLFNMTY----------NNTDNTTITLKCRIRQIVNQWMRVGKGI
SIV51   KHPRYTGTNNTDKINLTAPGGGDPEVTFMWTNCRGEFLYCKMNWFLNWVEDRDVTTQR----PKERHRRNYVPCHIRQIINTWHKVGKNV
SIV39   KHPRYTGTNNTDKINLTAPGGGDPEVTFMWTNCRGEFLYCKMNWFLNWVEDRNTANQK----PKEQHKRNYVPCHIRQIINTWHKVGKNV

        435  440                    471            482 487 489
MN      YAPPIEGQIRCSSNITGLLLT-----RDGGEDTDT--NDTEIFR-PGGGDMRDNWRSELYKYKVVTIEPLGVAPTKAKRRVVQRE-
HXB2    YAPPISGQIRCSSNITGLLLT-----RDGGNS-N---NESEIFR-PGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQREK
A1.KE   YAPPIPGVIKCESNITGLLLT-----RDGGKD-N---NVNETFR-PGGGDMRDNWRSELYKYKVVEIEPLGVAPTRAKRRVV----
A1.UG   YAPPIQGVIKCESNITGLILT-----RDGGVN-S---SDSETFR-PGGGDMRDNWRSELYKYKVVKIEPLGVAPTKARRRVV----
C.ET    YAPPIEGIMCRSNITGLLLT-----RDGAKEPH--STKEIFR-PEGGDMRDNWRSELYKYKVVKIPLGVAPTKPKRRVV----
C.IN    YAPPIAGNITCTSNITGLLLV-----HDGGIKENDTENKTEIFR-PGGGDMRDNWRSELYKYKVVEIKPLGVAPTRAKRRVV----
D.TZ    YAPPIAGLIRCTSNITGLLLT-----RDGGNVNN---SREEIFR-PGGGDMRDNWRSELYKYKVVRIEPIGVAPTRAKRRVV----
D.UG    YAPPIEGFINCSSNITGLLLT-----RDGGAINS---SQNETFR-PGGGDMRNNWRSELYKYKVLEPIGLAPTAAKRRVV----
AE.T93  YAPPISGRINCVSNITGILLT-----RDGGAN-N---TQNETFR-PGGGNIKDNWRSELYKYKVVQIEPLGIAPTRAKRRVV----
AE.T90  YAPPISGRINCVSNITGILLT-----RDGGVN-N---TDNETFR-PGGGNIKDNWRSELYKYKVVQIEPLGIAPTRAKRRVV----
CPZ05   FAPPIRGNITCTSNITGMLLE-IHKNREDQGED-Q---PGGGVCLTGGNMKDIWRSELYKYKVVIEQPLGVAPTKSRRYAV----
CPZ.85  FAPPIKGVLSCNSNITGMILD-----ISISAVNND--SRNITVM-PTGGDMTALWKNELHKYKVVSIEPIGVAPGKAKRHTV----
SIV51   YLPPREGDLTCNSTVTSLIAN------IDWT-------DGNQTSI-TMSAEVAELYRLELGDYKLVEITPIGLAPTDVKRYTTGGT-
SIV39   YLPPREGDLTCNSTVTSLIAN------IDWI-------DGNQTNI-TMSAEVAELYRLELGDYKLVEITPIGLAPTDVKRYTTGGT-
```

Supplemental Figure S1

Alignment of predicted and observed cathepsin L, S and D cleavage sites.
Reference sequences for gp120 were obtained from the Los Alamos HIV sequence database (www.hiv.lanl.gov). Envelope sequences from clades A, C, D, and E (crf A/E) as well as reference sequences for chimpanzee isolates of HIV (HIV$_{cpz}$) and SIV were aligned with the sequences of the prototypic clade B MN and HXB2 strains of HIV using the MAFFT sequence alignment program (52). Numbering is provided with reference to the MN strain of HIV (71). The sequences shown begin with the mature amino terminus of gp120. Predicted and observed cathepsin cleavage sites are indicated by open bars and closed bars respectively. Cathepsin L cleavage sites are indicated in green, cathepsin S cleavage sites are indicated in red, and cathepsin D cleavage sites are indicated in blue. The full names of the aligned sequences are as follows: MN (MN-rgp120); HXB2; A1.KE (A1.KE.94.Q23_17); A1.UG (A.UG.92.92UG037); C.ET (CET.86.ETH2220); C.IN (C.IN.93.93IN101); D.TZ (D.TZ.01.A280); D.UG (D.UG.94.94UG114); AE.T93 (AE.TH.93.93TH051); AE.T90 (AE.TH.90.CM240); CPZ05 (CPZ.CM.05.SIVcpzMT145); CPZ.85 (CPZ.US.85.CPZUS); SIV51 (SIV.US.MAC251); SIV39 (SIV.US.MAC239).

Location of Cathepsin Cleavage Sites and MHC Class I and MHC Class II T Cell Epitopes on HIV gp120
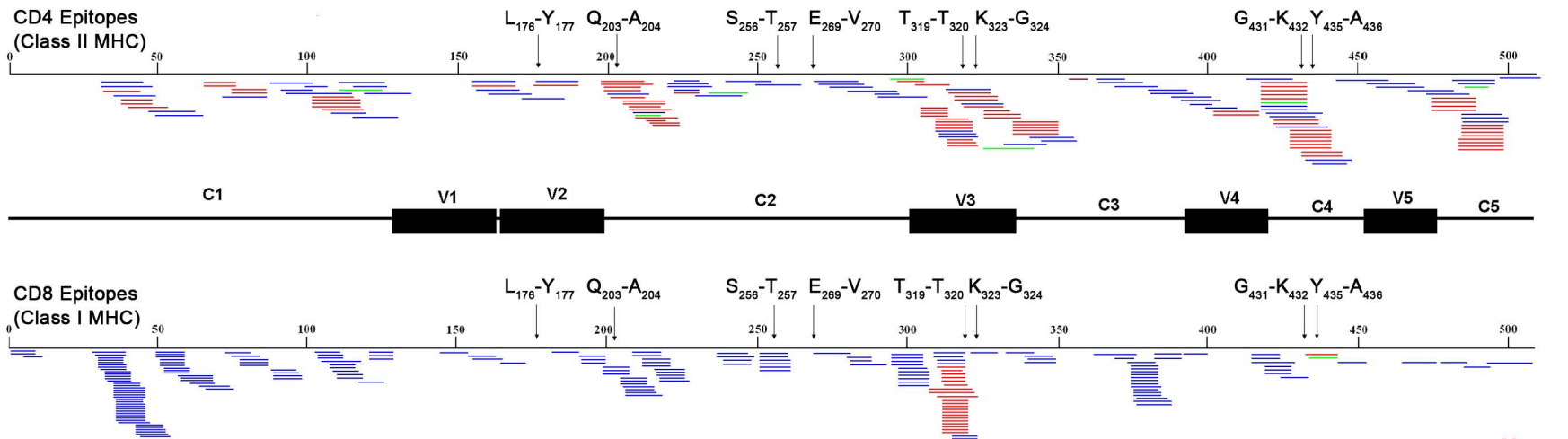
Supplemental Figure S2. Location of cathepsin cleavage sites and MHC class I and class II T cell epitopes on HIV gp120.
T cell epitopes were transcribed from the Los Alamos HIV Sequence database (33) and include data from Zhan et al. (104).
The locations of specific epitopes are provided with reference to the HXB2 sequence for gp120. The locations of cathepsin cleavage
sites are indicated by black arrows with the corresponding amino acid position indicated. Each MHC class I and class II epitope
is indicated by a horizontal line. Human epitopes are indicated by blue lines, mouse epitopes by red lines, and epitopes from other species by green lines.