# I

# Nucleic Acid Alignments and Sequences

Nucleotide sequence alignments were generated using the PIMA program developed by Randy Smith and Temple Smith (*Protein Engineering* **5**:35, 1992). By this approach, a protein pattern is created which then serves as a template for the nucleotide sequence alignment. Hand-editing, when required, was done with the MASE program (D. V. Faulkner and J. Jurka, *TIBS* **13**:321, 1988). For information concerning either of these programs, contact Dr. Randall Smith, Institute of Molecular Genetics, Baylor College of Medicine, Houston, Texas, 713-798-4735.

With few exceptions, only full-length coding sequences were included in these alignments. Tables of information pertaining to each sequence in an alignment are provided for the first time in 1995. The common names given to sequences in alignment and in the accompanying tables were selected on several grounds: for sequences corresponding to samples provided by the NIAID repository, WHO and DAIDS conventions are employed for the names; for other sequences, the name given by the authors of the paper reporting the sequence are usually utilized. Locus names, which no longer appear in alignments but are provided in the tables, will usually correspond to GenBank Locus names, however there is no guarantee for that and in any case EMBL and DDBJ names may differ. The Accession numbers, also provided in the tables, will be universal identifiers. We wish to thank many sequencers for providing data prior to publication.

Mixed case consensus sequences are used as the reference sequences for each alignment. Upper case letters indicate 100% conservation of nucleotide bases in a given position of the alignment, and lower case letters represent bases conserved in at least 50% of the sequences. The symbol "?" indicates no consensus at a position. To reduce the alignment uncertainty with the large HIV-1 set, sequence subtypes (see below and Part III) were constituted and consensus sequences were generated for each. Divergent sequences that could not be easily assigned to any given subtype and mosaic sequences (perhaps the result of recombinational events) are shown at the bottom of each block. Large deletions, if present in only one sequence (*e.g.*, the HIV2nihz *nef* coding region sequence), were also excluded from the consensus sequences.

Following the alignment for each coding sequence, just the consensus sequences over that coding region are re-presented in alignment.

Prior to the 1992 database compendium, HIV-1 sequences had been roughly classified as 'U.S.' and 'African.' In light of many new HIV-1 *gag* and *env* sequences, it became more useful, starting with the 1992 compendium, to categorize HIV-1 sequences into five sequence subtypes, depending upon the coding sequence. HIV-1 subtypes now number eight (A through H). Collectively, these are called group M sequences, as they are significantly distinct from group O sequences. The bases for this classification, discussed at greater length in Part III of the 1994 compendium, are:

i) subtypes are approximately equidistant from one another in *env* ( a "star phylogeny");

ii) the *env* phylogenetic tree is for the most part congruent with *gag* phylogenetic trees;

iii) two or more samples are required to define a sequence subtype.

Subtype naming problems have arisen for several reasons. A small but not insignificant number of viral sequences are hybrid, clustering with one sequence subtype in *gag* and another sequence subtype in *env*, for example; or, to take another example, clustering over different stretches with two or more subtypes in *env*. All subtype E sequences based on *env* have *gag* sequences that align with subtype A sequences; and, moreover, the 3' half of the gp41 cds of E and subtype G *env* sequences align with subtype A sequences (Gao *et al*, J. Virol., in press, 1996). Given the homogeneity of G and E subtype

sequences, these are handled as subtypes and not as hybrids in the following alignments. It remains to be discovered whether these states have arisen from recombination or lack of divergence.

Several analyses of mosaic molecules are presented in Part III. Naming also becomes problematic when highly divergent forms of a given subtype arise: such forms are sometimes designated A', B', F', etc. It is increasingly necessary to have sequence data from both *gag* and *env* coding sequences when a new form or subtype is being claimed.

There are at least five sequence subtypes for the HIV-2s, A through E.

The authority for some of the annotation is limited largely to invariance—the recurrence of patterns such as AATAAA, for example. The reader should be cautious in drawing upon this information. Due to space limitations, only certain sequences are reported as entries in this section. Beginning in 1995, most sequences are cataloged in Part I using just the headers of GenBank entries; to gain the full entry including the sequence itself, users should go to the HIV Sequence Database WWW site (*http://hiv-web.lanl.gov*), the FTP site (also described in Part V of this compendium) or any of the large gene libraries. Sequences used in alignments are provided on the accompanying diskettes, and the alignments themselves are on the WWW site. Sequence entries from previous years are presented in the 1990–1994 compendiums.

## Contents Part I 1995

### PART I HIV and SIV Nucleic Acid Sequences

Introduction . . . . . I-1
Contents(1995) . . . . . I-3

### A. HIV-1 Nucleotide Alignments and Consensus Sequences

Nucleotide alignments of HIV-1 LTR . . . . . I-A-1
Alignments of HIV-1 *LTR* consensus sequences . . . . . I-A-16
Nucleotide alignments of HIV-1 *gag* . . . . . I-A-18
Alignments of HIV-1 *gag* consensus sequences . . . . . I-A-70
Nucleotide alignments of HIV-1 *pol* . . . . . I-A-74
Alignments of HIV-1 *pol* consensus sequences . . . . . I-A-121
Nucleotide alignments of HIV-1 *vif* . . . . . I-A-125
Alignments of HIV-1 *vif* consensus sequences . . . . . I-A-135
Nucleotide alignments of HIV-1 *vpr* . . . . . I-A-137
Alignments of HIV-1 *vpr* consensus sequences . . . . . I-A-143
Nucleotide alignments of HIV-1 *tat* . . . . . I-A-144
Alignments of HIV-1 *tat* consensus sequences . . . . . I-A-151
Nucleotide alignments of HIV-1 *rev* . . . . . I-A-153
Alignments of HIV-1 *rev* consensus sequences . . . . . I-A-160
Nucleotide alignments of HIV-1 *vpu* . . . . . I-A-161
Alignments of HIV-1 *vpu* consensus sequences . . . . . I-A-167
Nucleotide alignments of HIV-1 *env* . . . . . I-A-168
Alignments of HIV-1 *env* consensus sequences . . . . . I-A-358
Nucleotide alignments of HIV-1 *nef* . . . . . I-A-366
Alignments of HIV-1 *nef* consensus sequences . . . . . I-A-390

## Sequences in the *LTR* Alignment

| COMMON NAME | LOCUS | ACC # | FIRST AUTHOR | REFERENCE |
|---|---|---|---|---|
| **SUBTYPE A:** | | | | |
| U455 | HIVU455 | M62320 | Oram,J.D. | ARHR 6, 1073 (1990) |
| SF1701 | HIVSF1701 | M66534 | Evans,L. | PNAS 85, 2815 (1988) |
| IBNG | HIVIBNG | L39106 | Rasheed,H. | Unpublished (1995) |
| **SUBTYPE B:** | | | | |
| LAI | HIVLAI | K02013 | Wain-Hobson,S. | Cell 40, 9 (1985) |
| HXB2R | HIVHXB2R | K03455 | Starcich,B. | Science 227, 538 (1985) |
| N1TE | HIVN1TE | M64760 | Golub,E.I. | JVI 64, 3654 (1990) |
| MN | HIVMN | M17449 | Gurgo,C. | Virol. 164, 531 (1988) |
| BRVA | HIVBRVA | M21098 | Anand,R. | Virol. 168, 79 (1989) |
| SC | HIVSC | M17450 | Gurgo,C. | Virol. 164, 531 (1988) |
| JH31 | HIVJH31 | M21137 | Komiyama,N. | ARHR 5, 411 (1989) |
| BAL1 | HIVBAL1 | M68893 | Gartner,S. | Science 233: 215 (1986) |
| JRCSF | HIVJRCSF | M38429 | Koyanagi,S. | Science 236: 819 (1987) |
| JRFL | HIVJRFL | M74978 | O'Brien,W.A. | Nature 348, 69 (1990) |
| OYI | HIVOYI | M26727 | Wain-Hobson,S. | AIDS 3: 707 (1989) |
| SF2 | HIVSF2 | K02007 | Sanchez-Pescador,R. | Science 227, 484 (1985) |
| SF2B13 | HIVSF2B13 | L07422 | Cheng-Mayer,C. | JVI 65, 6930 (1991) |
| NY5 | HIVNY5 | K03346 | Willey,R.W. | PNAS 83, 5038 (1986) |
| NL43 | HIVNL43 | M19921 | Buckler,C.E. | JVI 59: 284 (1986) |
| SF162 | HIVSF162 | M38428 | Cheng-Mayer,C. | JVI 64, 4390 (1990) |
| CAM1 | HIVCAM1 | D10112 | McIntosh,A. | Unpublished (1991) |
| CDC41 | HIVCDC41 | M13136 | Desai,S.M. | PNAS 83, 8380 (1986) |
| SF33 | HIVSF33 | M38427 | York-Higgins,D. | JVI 64, 4016 (1990) |
| HAN | HIVHAN | U43141 | Sauermann,U. | ARHR 6, 813 (1990) |
| P896 | HIVP896 | M96155 | Collman,R. | JVI 66, 7517 (1992) |
| D31 | HIVD31 | U43096 | Dietrich,U. | Unpublished (1992) |
| SWB884 | HIVSWB884 | M58206 | Delassus | JVI 65, 225 (1990) |
| RF | HIVRF | M17451 | Starcich,B.R. | Cell 45, 637 (1986) |
| YU2 | HIVYU2 | M93258 | Li,Y. | JVI 65, 3973 (1991) |
| BCSG3C | HIVBCSG3C | L02317 | Ghosh,S.K. | Virol. 194, 858 (1993) |
| E90NEF | HIVE90NEF | U43108 | Reitz,M., et,al. | Unpublished (1993) |
| YU10 | HIVYU10 | M93259 | Li,Y. | JVI 65, 3973 (1991) |
| WEAU160 | HIVWEAU160 | U21135 | Ghosh,S. | Unpublished (1995) |
| 3202A12 | HIV3202A12 | U34603 | Guillon,C. | ARHR 11,1537 (1995) |
| 3202A21 | HIV3202A21 | U34604 | Guillon,C. | ARHR 11,1537 (1995) |
| **SUBTYPE D:** | | | | |
| ELI | HIVELI | K03454 | Alizon,M. | Cell 46, 63 (1986) |
| Z2Z6 | HIVZ2Z6 | M22639 | Theodore,T. | Unpublished (1988) |
| Z6 | HIVZ6 | K03458 | Srinivasan,A. | Gene 52, 71 (1987) |
| NDK | HIVNDK | M27323 | Spire,B. | Gene 81, 275 (1989) |
| CBL41 | HIVCBL41 | M63474 | Goudsmit,J. | AIDS 4, 559 (1990) |
| **SUBTYPE O:** | | | | |
| MVP5180 | HIVMVP5180 | L20571 | Gurtler,L.G. | JVI 68, 1581 (1994) |
| ANT70 | HIVANT70 | M31171 | De Leys,R. | JVI 64, 1207 (1990) |
| ANT70C | HIVANT70C | L20587 | Vanden Haesevelde,M. | JVI 68, 1586 (1994) |
| **SUBTYPE U:** | | | | |
| Z321 | HIVZ321 | M15896 | Srinivasan,A. | ARHR 5, 121 (1989) |
| MAL | HIVMAL | K03456 | Alizon,M. | Cell 46, 63 (1986) |
| CPZ: | | | | |
| CPZGAB | SIVCPZGAB | X52154 | Huet,T. | Nature 345, 356 (1990) |
| CPZANT | SIVCPZANT | U42720 | Vanden Haesevelde,M. | Virol. In Press (1996) |

**HIV1 LTR**

```
     U3 start ->
CONSENSUS-A   TGGATGGGTTAATTtACTCcaaGAaAAGACAAGAAATCCTTGATCTGTGGGTCTAtaACACACAAGGaTt     70
U455          --------------C----AC-----------------------------------C-------------     70
IBNG          ------------------------------------------------------------------------     70
SF170         ---------------------G--G------------------------------C-----------C-A     70

CONSENSUS-B   TGGAaGGGcTAaTttacTCcCaaaaaAGaCAaGAtATcCTTGATcTgTGGatctAccACACACAAGGCTa     70
SF2           ---------------GG-------G---------G-----------------------------------     70
SF2B13        ---------------GG-------G---G-----G-----------------------------------     70
LAI           --------------C--------CG-------------------------------------------     70
HXB2          --------------C--------CG-------------------------------------------     70
N1TE          --------------C--------CG-------------------------------------------     70
NY5           ---------------GG-------G-------------------------------------------     70
NL43          ---------------GG-----------------G-----------------------------------     70
MN            ----T---T----------------G--------C---------------G-------------------     70
BRVA          --------------C--------C--------------------T-----G----T--------------     70
SC            -------------.C-------G-G-----------------------------------------------     69
BAL1          --------------C-----------------------------T-----G-------------------     70
JRCSF         -------------------A--G-----------------------------------------------     70
JRFL          --------------C----A--G---------------------------G-------------------     70
OYI           ----------------------G--------------T------------G----T--------------     70
SF162         --------T-------------------------------------------C----------------     70
CAM1          --------A---------------G------G--------------------------------------     70
CDC41         -----------G-----------C--------------------------G-------------------     70
SF33          -----------G-------------------------------------------------------     70
HAN           --------T------------C------------G---------------G-------------------     70
P896          -----------G--C-------G---------------------------G------------------T     70
SWB884        -----------G--C-------------------G---------------G-------------------     70
RF            ----T------G-G-T------G---------------------------G-T-----------------     70
YU2           --------------C--------C--------------------------G-------------------     70
BCSG3C        ---------------T--------G--------------------------C------------------     70
E90NEF        -----------G--C-T-------GG--------C---------T-A---G-----A-------------     70
YU10          --------------C--------C--------------------------G-------------------     70
3202A12       ------------------------GG-----------------------------------------------     70
3202A21       -------------------------G----------------------------------------------     70
WEAU160       ------------------------------------------------G------------------T     70

CONSENSUS-D   TGGAAGGGCTAATTTGGTC?AAaAa?AGACAAGAgATCCTTGAtCTTTGGGTCTACAACACACAAGGCAT     68
ELI           -------------------C-----G----------------------------------------     70
Z2Z6          -------------------A----GA--------C----------------------------------     70
Z6            -------------------A-----G-----------------C--------------------------     70
NDK           -------------------C--G--A-----------------------------------------     70

CONSENSUS-O   TGGA?GGGTTAATTTACTCCCATAA?AGAGCAGAAATCCTGGATCT?TGG?T?TAT?ACACTCAGGGATT     64
ANT70         ----A--------------------A--------------------T---G-G---A-------------     70
MVP5180       ----T--------------------G--------------------C---A-A---C-------------     70

CONSENSUS-U   TGGATGGG?TA?TTT??TCC???AAAAGACAAGAAATCCTTGA?CTGTGGGTCTA?CA?ACACAAGGCT?     59
Z321          --------C--A---AC---AAG--------------------C-----------T--T----------T     70
MAL           --------T--G---GG---CCA--------------------T-----------C--C----------A     70

CONSENSUS-CPZ TGGAAGGGTTAGTTTACTCCAGGAGAAGACAAGAGATCCTTGACCTCTGGGTCTATCACACACAAGGCTT     70
CPZGAB        ------------------------------------------------------------------------     70
```

```
                                                  AP-1          AP-1
                        /<-        COUP        ->/<-->/<-COUP->/<-->/
CONSENSUS-A   CTTCCCtGATTGGCAGAAtTACACACCAGGGCCAGGgAccAGATtCCC.ACTaACaTTTGGaTGGTGCTt    139
U455          ------------------------------------A-T-----A---.--------------------A    139
IBNG          ------A-----------C--------------------T--------.---G--C-----G--------    139
SF170         ------------------------------------------------.---------------------    139

CONSENSUS-B   cTTCCCtGAtTgGCAGAAcTACACAcCAGGgccagGgatcAGaTatCC?AcTGaccttTGGATGGTGcTt    139
SF2           ------------------T-----------------------------.---------------------    139
SF2B13        ------------------------------------------------.---------------------    139
LAI           -------------------------------------G----------.--------------------A    139
HXB2          -----------A-------------------------------------.--------------------A    139
N1TE          ------------------------------------------------.--------------------A    139
NY5           ------------------------------------------------.---------------------    139
NL43          ------------------------------------------------.---------------------    139
MN            ------------------------------------------------.---------------------    139
BRVA          -------------------------------------G----------.---------------------    139
SC            ------------------------------------------------C----TGT--------------    139
BAL1          --------------------------------------C-----T---.---------------------    139
JRCSF         -------------------------G----A------G------T---.---------------------    139
JRFL          -------------------------------------A-------T---.-T-------------------    139
OYI           T-----------------------------------------------.----TG---------------    139
SF162         ------------------------------------------------.---------------------    139
CAM1          ------------------------------------------------.------A--------------    139
CDC41         ----------------------------------A---C-----TC--.---------------------    139
SF33          -------------------------------------G------T---.---------------------    139
HAN           -------------------------------------G----------.---------------------    139
P896          ------A-----------------------------A-----------.--------------------A    139
SWB884        --------------------------------------C-----GG--.-------CC------------    139
RF            ---------C---------------------------C----------.---------------------    139
YU2           -------------------------------...----CT----GG--.---------------------    136
BCSG3C        ------------------------------------------------.------------------T--    139
E90NEF        T-------------------------------------C------C--.---------------------    139
YU10          --------------------------------------CT----GG--.---------------------    139
3202A12       ------------------------------------------------.---------------------    139
3202A21       ------------------------------------------G-----.---------------------    139
WEAU160       --------------------------------------CT-----C--.----TG---------------    139

CONSENSUS-D   CTTCCCtGATTGGCAgAACTACACACCAGGGCCAGGGATCAGATAtCC.ACTgACCTTTGGATGGTGCTt    137
ELI           ---------------A--------------------------------.---A----------------A    139
Z2Z6          ------------------------------------------------.---------------------    139
Z6            ------------------------------------------------.---------------------    139
NDK           ------C-------------------------------------C--.---------------------    139

CONSENSUS-O   CTTCCCTGATTGGCAG???TACACACC?GGACCAGGA?C?AG?TTCCC.ACTGACATTTGGATGGTTGTT    126
ANT70         ----------------AAC--------A---------A-C--G-----.---------------------    139
MVP5180       ----------------TGT--------G---------C-T--A-----.---------------------    139

CONSENSUS-U   CTTCCCTGATTGGCA?AA?TACACACCAGGGCCAGGGA??AGAT?CCC.ACTG???TT?GGATGGTGCTT    119
Z321          ---------------C--C-------------------CC----A---.----TGT--T-----------    139
MAL           ---------------G--T-------------------TT----T---.----ACC--C-----------    139

CONSENSUS-CPZ CTTCCCTGACTGGCAGAACTACACAACAGGACCAGGAACAAGATTCCC.ACTCTGTTTTGGATGGTGCTT    139
CPZGAB        ------------------------------------------------.---------------------    139
```

**HIV1 LTR**

```
                                        /<-              NF-AT              ->/
CONSENSUS-A   CAAgCTAGTACCAgT...tGATCCAGc?GaaGTAGAGgAaGccActG?aGGAGAGAaCAACAG.......    197
U455          ---------------...---------T--------------T----G---------------.......    199
IBNG          ---A---------A-...G--------A--G-------------A--A---------------.......    199
SF170         ---------------...--------AG-G-------A-G-A---G-TG-------C------.......    199

CONSENSUS-B   CAAGcTAGTACCAgT...tgAgCcAGagaaggtagaagaggccaatgaAggAGAgAAcaaCag???????    199
SF2           ---------------...--------------------------------------------.......    199
SF2B13        -------------C-...----------C-----------------------------------.......    199
LAI           ---------------...---------T----------------A------------C---.......    199
HXB2          ---------------...-------------T-------A-----CA------------C---.......    199
N1TE          ---------------...-----A---------------A-------------------T-C-.......    199
NY5           ----T----------...--------G-C-----------------------------------.......    199
NL43          ----T----------...---A------C-A--------------AT----------G-----.......    199
MN            ---------------...-------------A--------------A--------------T-.......    199
BRVA          ---------------...-------------A-------------------------------.......    199
SC            ---------------...-A-A---------A-------A------------------------.......    199
BAL1          ---------------...-----------------------------------------T-.......    199
JRCSF         ---------------...---T-----------------------------------------T-.......    199
JRFL          ---------------...---------A--------------------------------T-.......    199
OYI           ----T--------A-...G--T-----CC----------------C------------------.......    199
SF162         ---------------...---T-----TT-T----------------C----------------.......    199
CAM1          ---------------...----------C-----------------A--A--------GC---.......    199
CDC41         -------------C-...----------C----------------C------------------.......    199
SF33          ----T----------...A-----------A--------------------------------.......    199
HAN           ----T----------...---A-----...................-----A--------G---.......    181
P896          ---------------...----------.................------------------AGAGGAC    188
SWB884        ---------------...C---------------------------------------------.......    199
RF            ---------------...---------T----------------C------------------.......    199
YU2           ---------------...-------------A---------------C-------------T-.......    196
BCSG3C        ---------------...---T---------A---------------------A--------.......    199
E90NEF        ---------------...------------A------------------------------A.......    199
YU10          ---------------...-------------A---------------C-------------T-.......    199
3202A12       ---------------...-----A-------A--------------------------------.......    199
3202A21       ---------------...-----A-------A--------------------------------.......    199
WEAU160       ---------------...G--A-----------------A--T--------------------.......    199

CONSENSUS-D   CgAGCTAGTACCAGT...TGATCCAC?GGAGGTAGAAGAgGcCACTGAAgGAGAGacCAACtG.......    196
ELI           ---------------...--------A------------A-A-------------------A-.......    199
Z2Z6          ---------------...--------G-----------------------------------.......    199
Z6            ---------------...--------G-----------------------------------.......    199
NDK           -C-------------...--------A----------------------A-----GA------.......    199

CONSENSUS-O   TAAACT?GTACCAGTGTCAG?AGAAGAGGCAGA?AGACT?GG??ATACA??TGA????GCTA?.......    176
ANT70         ------A-------------A------------A-----A--AG-----TG---GAGG----A.......    202
MVP5180       ------G-------------C------------G-----G--TA-----AA---AGAT----G.......    202

CONSENSUS-U   ?AAG?TAGTACCA?T...???TCCA???GAAGTAGAGGA?GCCAAT??AGGAGAGAACAACTG.......    167
Z321          C---C--------G-...TGA----AGA-----------A------AC----------------.......    199
MAL           T---T--------A-...GAG----GAG-----------G------GA----------------.......    199

CONSENSUS-CPZ CAAACTAGTGCCCCTGACAGA...GGAGCAGGTAGAACAAGCCAATGAAGGAGATAACAACTG.......    199
CPZGAB        ---------------------...--------------------------------------.......    199
```

```
                                                        /<-  NF-AT  ->/
CONSENSUS-A   .....tcTAtTACACCCtATaTGcCAACATGGAaTgGAtGAtGA?GA?AgAGAAgtgtTaATgTGGAagT    260
U455          .....----C-----------------------G-A--------A--G-A--------------------    264
IBNG          .....-T----------C--C--T--------------G--C--C--T---------C-G--C----GA-    264
SF170         .....C--------------------------------------G--A------ACA-------------    264

CONSENSUS-B   ?????ctTgttAcacCCtatgagCCaGcATGGgatgGAtgACccgGAGagagAAGtgtTagtgTGGAggT    264
SF2           .....--------------------T------------G---G------A--------------------    264
SF2B13        .....---------------A----------------------------A----------------AA-    264
LAI           .....-------------G------T------A-----------T----------------A--------    264
HXB2          .....-------------G------T------A----------------------------A--------    264
N1TE          .....-------AG----G------T-------------A----------------A----A--------    264
NY5           .....----------------------------------G---------G-------A----------A--    264
NL43          .....----------------------------------G---------G-------A----------A--    264
MN            .....----------------------------.----------------------------------A--    263
BRVA          .....T--------------------------A------------------------C-GCA--------    264
SC            .....------------G-------T------------G----------------------A--------    264
BAL1          .....-----C---------------T----------------------A-----------CA----A--    264
JRCSF         .....----------------------------A-----C-----A----AG----------------A--    264
JRFL          .....------------------------------A--G----------AG----------A--------    264
OYI           .....---------T-----A----T------------------A----A-----------------A--    264
SF162         .....-------------------------------------------A------A-------------    264
CAM1          .....----------------------------------------A----A------AC--A------A--    264
CDC41         .....----------------------G----------------A----A-----------------A--    264
SF33          .....--------------------T------------G----------A-----------------A--    264
HAN           .....-C-----------GC-----T--------CA--G---A-----------------AA-----A--    246
P896          AACAG----C--------GCT-A---------AG-A--A---T--------C------------------    258
SWB884        .....--------------------T------A-----G----------A-----------------A--    264
RF            .....---------------AT---T------------------A----A-----------------A--    264
YU2           .....--------------------------A----------------------G-----A--------    261
BCSG3C        .....----C-------------------------------------------C---A--------    264
E90NEF        .....A--A---------G------T------------------A----------------------A-    264
YU10          .....----------------------------A----------------------------A--------    264
3202A12       .....---------------------------------C----------------------A--------    264
3202A21       .....---------------------------------C----------------------A--------    264
WEAU160       .....--------------------T-----------------AT----A------A---A------A--    264

CONSENSUS-D   .....CTTGTTACACCCT?T?TGCCAGCAtGGAATGGAGGAC?CGGAGAGA?AAGTGTTAAagTGGAGAT    257
ELI           .....-------------A-A---------------------C--------C----------A-------    264
Z2Z6          .....-------------G-A---------------------A--------G------------------    264
Z6            .....-------------G-G---------------------A--------G------------------    264
NDK           .....-------------A-G--------A------------C--------C---------T--------    264

CONSENSUS-O   .....TCT?CT?CATCCAGC?TGT???CATGGA??TGA?GAT?CACA????GA?ATACT?A??TGG?AGT    222
ANT70         .....---C--G--------A---GCC------TT---A---A----TAAA--A-----G-TG---A---    267
MVP5180       .....---T--A--------T---AAT------GC---G---G----CGGG--G-----A-AA---C---    267

CONSENSUS-U   .....?CTGTTACACCC?AT?AGCCA?CATGGAATGGA?GA?G??GAAAGAGAAGTG?TAA??TGGAAGT    221
Z321          .....C-----------C--G-----G-----------T--T-AT------------T---TG-------    264
MAL           .....T-----------T--T-----A-----------G--C-CA------------C---AA-------    264

CONSENSUS-CPZ .....CCTGTTGCATCCCATATGTCAGCATGGGATGGAAGATGAGGACAAAGAGGTCTTGGTCTGGCGCT    264
CPZGAB        .....--------------------------------------------------------------    264
```

**HIV1 LTR**

```
                                /<- IL2/NRE ->/
                              <-     NRF     ->
CONSENSUS-A   TTGACAGtagcCTGGCATTaAaACACAgAGCt?aaGAGcTGCATCCGGAGTtCTAcAAA....GACTG..    323
U455          ---------C----------------------T-T--------------------T---....-----..    328
IBNG          ----------A--------C-G-----C----AG-----------------A-------....-----..    328
SF170         -------CC----------------------CC-----A--------------------....-----..    328

CONSENSUS-B   tTGACAGccgcCTaGcATTtCatCAcatgGCCCGAGAgctgcATCCGGAGTacTacAA?..??gACtg??    327
SF2           --------AAA----------------------------------------------A....-----..    328
SF2B13        --------AAG----------------------------------------------A....-----..    328
LAI           --------------------------G---------------------------T---G....A----..    328
HXB2          -------------------------------------------------------T---G....A----..    328
N1TE          --------------------------G---------------------------T---G....A---TGC    330
NY5           ---------T-----------G-------------------------------------A....-----..    328
NL43          ---------T-----------G-------------------------------------A....-----..    328
MN            C--------A------------G--TTAT-----------------------------G....A----..    327
BRVA          ------------------------------------------------------------G....A----..    328
SC            ------A------------------------------T--------------------A....-----T.    329
BAL1          --------A----------C------G----------A                                    302
JRCSF         --------AAG--------G------G-------------------------------G....-----..    328
JRFL          --------AAG---------------G-------------------------------G....-----..    328
OYI           --------------------G----------------G-------------------A....-----..    328
SF162         -------------C-G------------------------CG----------------A..AA-----..    330
CAM1          --------------------------------------AA-----------T------A....-----..    328
CDC41         ---------T-----------G-------------------------------------A....-----..    328
SF33          ---------A-----------G-------------------------------------A....-----C    329
HAN           ---------AT----------------A-------------------------------A....-----..    310
P896          -------------------C------G---------------------------T---G....A----A.    323
SWB884        --------A-------------C----A------------------------T-----A....-----..    328
RF            -------------C------------G-C---------AA------------------A....-----..    328
YU2           --------------------------G-------------------------------G....A----..    325
BCSG3C        ------------------------------------------------------------G....A--C-..    328
E90NEF        -------------------C------G-T-----------------------------G....A---A..    328
YU10          --------------------------G-------------------------------G....A----..    328
3202A12       -------T-----------------------------------------------------G....-----..    328
3202A21       -------T-----------------------------------------------------G....-----..    328
WEAU160       --------AAA---------------G-A-------------------------TT--A....-----..    328

CONSENSUS-D   TTAACAGCAGACTAGCAtTTGA?CACAAGGCCCGAGAG?TGCATCCGGAGTTCTACAAA....gACTG..    319
ELI           ----------------------G---------------A--------------------....A----..    328
Z2Z6          ----------------------A---------------C--------------------....-----..    328
Z6            ----------------------A---------------A--------------------....-----..    328
NDK           -----------------C----G---------------C--------------------....-----..    328

CONSENSUS-O   TTGATAGATC??TAGGC???AC?CAT?T?GC??TG??AA??CACCCAGAGCTCTTCC???A???ACT?..    269
ANT70         ----------TC-----AAC--C---G-T--TA--AT--CT----------------AG.-AGG---A..    334
MVP5180       ----------AT-----TTA--A---A-A--CC--CA--AG----------------CCA-GTA---G..    335

CONSENSUS-U   TTGACAGC??CCTAGCAC?AA?ACAC??AGCC?GAGA????CATCCGGAGT?CTACAAA....GACT?..    272
Z321          --------TC--------G--A----CT----C----GATG----------T-------....----A..    328
MAL           --------AG--------T--G----AG----A----ACAA----------A-------....----G..    328

CONSENSUS-CPZ TTGACAGCAGGCTGGCCTTAAGACATATTGCCAGAGAACAACATCCGGAGTACTACAAA....GACTG..    328
CPZGAB        ------------------------------------------------------------....-----..    328
```

```
                                                                  /<-NF-kB ->/
CONSENSUS-A   .............????????CTGACACAgAAGTTGCTGACA................GgGACTTTCC    354
U455          .............AAAGACTG---------------------.................----------    367
IBNG          ......................---------------------................----------    359
SF170         ......................--------A------------................-A--------    359

CONSENSUS-B   ??????????????????????cTGACatcgagcTttCT.aCaa................GGGACTTTCC    358
SF2           ......................-----------------.----................----------    359
SF2B13        ......................------------------.---G...............----------    359
LAI           ......................--------------G--.----................----------    359
HXB2          ......................--------------G--.----................----------    359
N1TE          ATCCGGAGTACTTCAAGAACTG--------------G--.----................----------    383
NY5           ......................-----------------.----................----------    359
NL43          ......................-----------------.----................----------    359
MN            ......................-------------A---.----................----------    358
BRVA          ......................-------------A---.G--.................----------    359
SC            ......................----------TG-G---.----................----------    360
JRCSF         ......................------C----------.----................----------    359
JRFL          ......................-----------------.----................----------    359
OYI           ......................------------------.---G...............----------    359
SF162         ......................--------                                            338
CAM1          ......................-----------------.----................----------    359
CDC41         ......................------C-T-T------.----................----------    359
HAN           ......................------C----T-----.----................----------    341
P896          ..................ACTG------CT-----A---.----................----------    358
SWB884        ......................-----------T-----.--G-................----------    359
RF            ......................-----------T-----.----................----------    359
YU2           ......................A----C-----------.----................----------    356
BCSG3C        ......................A--------C-------.----................----------    359
E90NEF        ................GAACTG-----------A-----.----................----------    365
YU10          ......................A----------------.----................----------    359
3202A12       ......................-----------------.----................----------    359
3202A21       ......................-----------------.----................----------    359
WEAU160       ......................-----------T-----.----................----------    359

CONSENSUS-D   ......................cTGACAcCgAGtTTtCT.ACAA................GGGACTTTCC    350
ELI           ......................A----------C-----.----................----------    359
Z2Z6          ......................--------A--------.----................----------    359
Z6            ......................--------A--------.----................----------    359
NDK           ......................------T-------C--.----................----------    359
CBL41                                ----------------.----................----------     29

CONSENSUS-O   ...............??????????????????????????????..........???CTG?GG?ACTTTCC    281
ANT70         ...............AAAACTGCTGACCTGAAGATTGCTGACA..............---T--A-------    376
MVP5180       ..........................................................ACA---C--G-------    352

CONSENSUS-U   ...............???????CTGACACAGAAGTTGCTGACA?................GGGACTTTCC    303
Z321          ...............AAGACTG---------------------.................----------    366
MAL           ......................---------------------G................----------    360

CONSENSUS-CPZ .........ACTTTAGA.....CTGGCGCA.....TGCGCACAAGAACTGCTGACTCTGCGGGACTTTCC    379
CPZGAB        .........--------....--------.....----------------------------------    379
```

**HIV1 LTR**

```
                                  /<- NF-kB ->/      Sp-1       Sp-1     Sp-1
CONSENSUS-A   ..........aCT........GGGGACTTTCCg?.GGGAGGTGTGGTTTGGGcGGAGTTGGGGAGTGGCT    404
U455          ...........---........------------A.------------------------------------    418
IBNG          ...........---........------------..-----------------A-----------------    409
SF170         ..........G--........-----------AG.------------------------------------    410

CONSENSUS-B   ..........gcT........GGGGacTTTCcaG.GG?aggcgtGgccTGGGCgGgACtGGGGAGTGGCg    408
SF2           ...........---........-------------.--.--------------------------------    409
SF2B13        ...........---........-------------.--.--------------------------------    409
LAI           ...........---........-------------.--.--------------------------------    409
HXB2          ...........---........-------------.--.--------------------------------    409
N1TE          ..........-T-........-------------.--.--------------------------------    433
NY5           ...........---........-------------.--.---T----------------------------    409
NL43          ...........---........-------------.--.---T----------------------------    409
MN            ...........---........-------------.--.---T----------------C-----------    408
BRVA          ...........---........-------------.--.----.----------------------------    408
SC            ...........---...........----CA-------.--.-----C---.----------------------    409
JRCSF         ...........---........-------------.--.--------------------------------    409
JRFL          ..........A--........-------------.--.--------------------------------    409
OYI           ...........---........-------------.--.--------------------C-----------    409
CAM1          ...........---........-------------.--G---T----------------------------    410
CDC41         ...........---........----------A--.--.--.-A--------------------------    408
HAN           ...........---........-------------.--.-------A-----------------------T    391
P896          ...........---........-------------.--.---T-----------C-A--------------    408
SWB884        ...........---........-------------.--.-A------------------------------    409
RF            ...........---........-------------.--.---T----------------------------    409
YU2           ...........---........-------------.--.-A------------------------------    406
BCSG3C        ...........---........-------------.--.GA------------------C-----------    409
E90NEF        ...........---........-------------.--.GA------------------C-----------    415
YU10          ...........---........-------------.--.-A------------------------------    409
3202A12       ...........---........-------------.--G--------------------------------    410
3202A21       ...........---........-------------.--G--------------------------------    410
WEAU160       ..........A--........-----------G-.--.-------AT-----------A-----------    409

CONSENSUS-D   ..........GCT........GGGGACTTTCCaG.GG.AGGcGTggactGGGCGGGACtgGGgAGTGGCT    400
ELI           ...........---........-------------.--.--------------------------------    409
Z2Z6          ...........---........-----------G-.--.--------------------------------    409
Z6            ...........---........-------------.--.--------------------C-----------    409
NDK           ...........---........-------------.--.--------C------------T----------    409
CBL41         ...........---........-------------.--.---A--AACGG-------------.-------     78

CONSENSUS-O   A?????GACTGCTGACACTGCGGGGACTTTCCAG??G??GG???GA?A?GGGGCGGTTCGGGGAGTGGCT    337
ANT70         -GCAAA----------------------------TG-GA--...--C-G---------------------    443
MVP5180       -.....------------------------------.C-TG--AGG--T-A---------------------    416

CONSENSUS-U   ..........GCT........GGGGACTTTCCAG.GGGAGG?GT??C?TGGGCGGG???GGGGAGTGG?T    346
Z321          ...........---........-------------.------T--GG-C--------GTT---------T-    417
MAL           ...........---........-------------.------C--AA-T--------ACC---------C-    411

CONSENSUS-CPZ A...................AGGGACTTTCCGGGGAGAGGTGGTCGGGAGGCGGGGTTGGGGTGTGGCT    429
CPZGAB        -...................--------------------------------------------------    429
```

```
                                                                  -LBP-1 -  -LBP-1
                                    signal ->                     mRNA ->  -UBP-1- /<- TAR CORE
CONSENSUS-A   ...........AaCCCTCAGAtgcTGCATATAAGCAGCTGCTTtTCGCCTGTACTGGGTCTCTCTTGtTA    463
U455          ...........-G-------------------------------------------------------    477
IBNG          ...........----------CCG-------------------C-----------------------C--    468
SF170         ...........-------------------------------------------------------    469

CONSENSUS-B   ??????????agCCCTCAGAtgctgCATATAAGCAgctGCTTTttGCCTGTACTGGgTCTCTCTgGTTa    467
SF2           ...........T.-------------------------------------------------------    467
SF2B13        ...........T-------------------------------------------------------    468
LAI           ...........-------------------------------------------------------    468
HXB2          ...........-----------C-------------------------------------------    468
N1TE          ...........-------------------------------------------------------    492
NY5           ...........-------------------------------------------------------    468
NL43          ...........--------------A----------------------------------------    468
MN            ...........-----------CG--------------------C-------------------------    467
BRVA          ...........---------------------------------C-------------------------    467
SC            ...........-------------------------CG-------------------------------    468
JRCSF         ...........-------------------------------------------------------    468
JRFL          ...........-------------------------------------------------------    468
OYI           ...........-------------------------------------------                453
SF162                                                       -------------------------    362
CAM1          ...........-------------------------------------------------------    469
CDC41         ...........------------.--------------------G-------------------------    466
HAN           GGGGAGTGGCG---------------------------C------------------C------------    461
P896          ...........-------------------------------------------------------    467
D31                                                                 -------------G     14
SWB884        ...........-------------------------------------------------------    468
RF            ...........-------------------------------------------------------    468
YU2           ...........---------------------------------.-------------------------    464
BCSG3C        ...........---------------------------------.-------------------------    467
E90NEF        ...........----------CCTG-------------------C-------------------------    474
YU10          ...........---------------------------------.-------------------------    467
3202A12       ...........-------------------------------------------------------A----    469
3202A21       ...........-------------------------------------------------------A----    469
WEAU160       ...........-------------------------------------------------------    468

CONSENSUS-D   ...........AACCCTCAGATGCTGCATATAAGCAGCTGCTTTTtGCCTGTACTGGGTCTCTCTGGTTA    459
ELI           ...........-------------------------------------------------------    468
Z2Z6          ...........-------------------------------------------------------    468
Z6            ...........-------------------------------------------------------    468
NDK           ...........----------------------------------C------------------------    468
CBL41         ...........-------------------------------------------------------    137

CONSENSUS-O   ...........AACCCTCAGA?GCTGCATATAAGCAGC?GCTTTC?GCTTGTACCGGGTCT??GTTAGAG    391
ANT70         ...........----------A----------------C------T---------------CG-------    502
MVP5180       ...........----------T----------------T------C---------------TA-------    475

CONSENSUS-U   ...........AACCCTCAGATGCTGCATA?AAGCAGC?GCTTTTCGC?TGTACTGGGTCTCTCT?GTTA    401
Z321          ...........-------------------A-------C---------T----------------A----    476
MAL           ...........-------------------T-------T---------C----------------T----    470

CONSENSUS-CPZ T..........TGCCCTCAGA.GCTGCATATAAGCAGCTGCTTTCTGCCTGTACTGGGTCTCTTTCACTG    488
CPZGAB        -..........----------.-------------------------------------------------    488
```

**HIV1 LTR**

```
                                TAR CORE ->/
                 -LBP-1-            UBP-2 ->/                        signal ->
CONSENSUS-A   GACCAGAT??GAGcCTGgGAgCTCTCTGGCTAGCGagGG.AACCCACT.GCTTAAGCCTCAATAAAGCTT    529
U455          ---------.C-----------------------------.---------.----------------------    544
IBNG          --------CT-------------------------GA--.---------.----------------------    536
SF170         --------.A---T---A--A------------------.---------.----------------------    536

CONSENSUS-B   GACCAGATctGAGccTGGGAGCTCTCTGGCTagCTAGGg.AACCCACT?GCTTAAGCCTCAATAAAGCTT    535
SF2           --------------------------------A------.---------.----------------------    535
SF2B13        ----------------------------------------.---------.----------------------    536
LAI           --------T-----------------------A------.---------.----------------------    536
HXB2          --------------------------------A------.---------.----------------------    536
N1TE          --------------------------------A------.---------.----------------------    560
NY5           ----------------------------------------.---------.----------------------    536
NL43          --------------------------------A------.---------.----------------------    536
MN            --------------------------------A------.---------.----------------------    535
BRVA          -----------------------                                                    490
SC            ----------                                                                 478
JH31                            --------------A------.---------.----------------------     50
JRCSF         ----------------------------------------.---------.----------------------    536
JRFL          ---------------------------                                                495
SF162         ---------A------------------------------.---------G----------------------    431
CAM1          -------------------------------GA------.---------.----------------------    537
CDC41         -------------TT-------------------------.---------.----------------------    534
HAN           ----------------------------------------.---------.----------------------    529
P896          ----------------------------------------.---------.----------------------    535
D31           ----------------------------------------.---------.----------------------     82
SWB884        ----------------------------------------.---------.----------------------    536
RF            --------------T-------------------------.---------.----------------------    536
YU2           --------------------------------------A.---------.----------------------    532
BCSG3C        ----------------------------------------.---------.----------------------    535
E90NEF        ----------------------------------------.---------.--------                529
YU10          --------------------------------------A.---------.----------------------    535
3202A12       -------------------------------GA------.---------.----------------------    537
3202A21       -------------------------------GA------.---------.----------------------    537
WEAU160       ----------------------------------------.---------.----------------------    536

CONSENSUS-D   GACCAGATTTGAGCCTGaGAGCTCTctGGCTAgcTAGGG.AACCCACT.GCtTAAGCCTCAATAAAGCTT    527
ELI           -----------------G----------------------.---------.----------------------    536
Z2Z6          ----------------------------------------.---------.----------------------    536
Z6            ----------------------------------------.---------.--G------------------    536
NDK           -----------------G-------..-----AT-----.---------.----------------------    534
CBL41         ----------------------------------------.----                              180

CONSENSUS-O   GACCAGGTCTGAGCCCGGGAGCTCCCTGGCCT.CTAGCTGAACCCGCT.GCTTAACGCTCAATAAAGCTT    459
ANT70         --------------------------------.---------------.----------------------    570
MVP5180       --------------------------------.---------------.----------------------    543

CONSENSUS-U   GACCAG?T??GAGCCCGGGAGCTCTCTGGCTAGC.AAGG.AACCCACT.GCTTAAGCCTCAATAAAGCTT    465
Z321          ------A-TT--------------                                                   500
MAL           ------G-C.------------------------.----.---------.----------------------    536

CONSENSUS-CPZ GACCAGATTTGAGCCTGGGAGCTCTCTA.CCAGTGAAGG.AACCCACT.GCTTAAGCCTCAATAAAGCTT    555
CPZGAB        ----------------------------.----------.---------.----------------------    555
```

```
                    <- mRNA
CONSENSUS-A   GCCTTGAG?TGCTT?AAGTA.GTGTGTGCCCGTCTG?TT?T?TGACTCTGGTAACTAGAGATCCCTCAGA    593
U455          --------G-----G-----.---------------.--A-T----------------------------    612
IBNG          --------.-----C-----.---------------A--G-G----------------------------    604

CONSENSUS-B   GCCTTGA.GTGCTtcAAGTA.GTGTGTGCCCgTCTG.TTgTGTGACTCTGGTAaCTAGAGATCCCTCAgA    602
SF2           -------.------------.---------------.---------------------------------    602
SF2B13        -------.------------.---------------.---------------------------------    603
LAI           -------.------------.---------------.---------------------------------    603
HXB2          -------.------------.---------------.---------------------------------    603
NY5           -------.-----A------.---------------.---------------------------------    603
NL43          -------.-----CA-----.---------------.---------------------------------    603
MN            -------.------------.---------------.--A-------------G----------------    602
JH31          -------.------------.---------------.---------------------------------    117
JRCSF         -------.------------.---------------.---------------------------------    603
OYI                           ----.---------------.---------------------------------    505
SF162         --                                                                        433
CAM1          -------.------------.----------A----.---------------------------------    604
CDC41         -------.------T-----.---------------.---------------------------------    601
HAN           -------.------------.---------------.----------------G----------------    596
P896          -------.------------.---------------.----------------G----------------    602
D31           -------.------T-----.---------------.---------------------------------    149
SWB884        -------.------------.-------                                            562
RF            -------.------------.---------------.---------------------------------    603
YU2           -------.------T-----.---------------.---------------------------------    599
BCSG3C        -------.------------.---------------.----------------G--------------T-    602
YU10          -------.------T-----.---------------.---------------------------------    602
3202A12       -------.------------.---------------.---------------------------------    604
3202A21       -------.------A-----.---------------.---------------------------------    604
WEAU160       -------.------------.---------------.---------------------------------    603

CONSENSUS-D   GCCTTGA.GtGCTTCAAGTA.GTGTGTGCCCGTCTG.TTGTGTGACTCTGGTAACTAGAGATCCCTCAGA    594
ELI           -------.------------.---------------.---------------------------------    603
Z2Z6          -------.------------.---------------.---------------------------------    603
Z6            -------.------------.---------------.---------------------------------    603
NDK           -------.-C----------.---------------.---------------------------------    601

CONSENSUS-O   GCCTTGA.GTG...AGAAGCAGTGTGTGCTCATCTG.TTCA...ACCCTGGTGTCTAGAGATCCCTCAGA    521
ANT70         -------.---...----------------------.----...---------------------------    632
MVP5180       -------.---...----------------------.----...---------------------------    605

CONSENSUS-U   GCCTTGA.GTGCCTCAAGCA.GTGTGTGCCCATCTG.TTGTGTGACTCTGGTAACTAGAGATCCCTCAGA    532
MAL           -------.------------.---------------.---------------------------------    603

CONSENSUS-CPZ GCCTTGA.GTGTATTTGAGCAGTGTGTGCCTATCTA.TTCAGACACTCTGGTAACTAGAGATCCCTCAGA    623
CPZGAB        -------.----------------------------.---------------------------------    623
CPZANT                                            .                                      0
```

**HIV1 LTR**

```
                                                <- LTR
                                                  <- Lys-tRNA pbs
CONSENSUS-A   CCACT?TAGACTGTGT..AAAAATCTCTAGCAGTGGCGCCCGAACAGG????????????????GACTCG    644
U455          -----A----------..------------------------------.................------    664
IBNG          -----C----------..------------------------------ACTTGACGGTAATAGG------    672

CONSENSUS-B   cCcttttAGtcAGtgtgG?AAAATCTCTAGCAGTGGCGCCCGAACAGG................gACttg    655
SF2           ------------------A-----------------------------.................---GC-    656
SF2B13        --------------A---.--------------                                        635
LAI           ------------------.------------------------------.....................------    656
HXB2          ------------------.------------------------------.................---C--    656
NY5           ------------------.------------------------------.....................------    656
NL43          ------------------.------------------------------.....................------    656
MN            T--------G--------.------------------------------.....................------    655
JH31          --A----------C----.------------------------------.................---CG-    170
JRCSF         ------------------.------------------------------.................---CG-    656
OYI           ------------------.------------------------------.................---C-T    558
CAM1          ------------------.------------------------------.....................------    657
CDC41         ------------------A-----------------------------.....................------    654
HAN           T-----------------.------------------------------.....................------    649
P896          --A-CC----T-----A-.------------------------------.................---CG-    655
D31           ------------------.------------------------------.....................------    202
RF            --TC--------------.------------------------------.................---CA-    656
YU2           ------------------.------------------------------.....................------    652
BCSG3C        ------------------.------------------------------.................---CG-    655
YU10          ------------------.------------------------------.....................------    655
3202A12       ------------------.------------------------------.....................------    657
3202A21       ---------------A--.------------------------------.................---GC-    657
WEAU160       --A---------------.------------------------------.................---CC-    656

CONSENSUS-D   ?cCcTTTAGTCaGAGTGg?AAAATCTCTAGCAGTGGCGCCCGAACAGG................GaCCTG    646
ELI           C-----------------.------------------------------.....................------    656
Z2Z6          C-----------------.------------------------------.....................------    656
Z6            TA----------------.------------------------------................-T----    656
NDK           T--G-------C------.A-----------------------------.....................------    654

CONSENSUS-O   TCACT.TAGACTGAAGCAGAAAATCTCTAGCAGTGGCGCCCGAACAGG................GACTTG    574
ANT70         -----.------------------------------------------.....................------    685
MVP5180       -----.--------------------------                                          636

CONSENSUS-U   CCACTCTAGACGGTGT..AAAAATCTCTAGCAGTGGCGCCCGAACAGG................GACTTT    584
MAL           ----------------..------------------------------................------    655

CONSENSUS-CPZ TTAAATTATAGTCAA.GG.AAAATCTCTAGCAGTGGCGCCCGAACAGG................GAC?TG    674
CPZGAB        ---------------.--.------------------------------.................---C--    675
CPZANT                                                        --.................---T--      8
```

```
                                                                /<- psi locus
CONSENSUS-A   AAAGCGAAAG......................TTCCAGAGAAG?....TCTCTCGACGCA?.GGACTCG    684
U455          ----------......................-----------T....------------..-------    705
IBNG          ----------......................-----------A....------------A.-------    714

CONSENSUS-B   AaAgcGAAAG?aa...................aaCCaGAGgagc....TCTCTCGACGCA..GGACTCG    698
SF2           ----------T-G..................------------....------------..-------    700
LAI           ----------GG-..................------------....------------..-------    700
HXB2          ----------GG-..................---------...-....------------..-------    697
NY5           -G--------T--..................-G---------A....------------..-------    700
NL43          ----------T--..................-G---------A....------------..-------    700
MN            ----------A--..................---------...-....------------..-------    696
JH31          ----------AG-..................-----------A....------------..-------    214
JRCSF         ----------AG-..................-----------A....------------..-------    700
JRFL                                                     ....     -------..-------    509
OYI           ----T-----TGG..................------------....------------..-------    602
CAM1          ---A------T-G..................----G-------....------------..-------    701
CDC41         ----------TG-..................G-------A---....------------..-------    698
HAN           ----------T-G..................-----------A....------------..-------    693
P896          ----------AG-..................-----------A....------------..-------    699
D31           ----T-----T--..................G----------....------------..-------    246
RF            ----------T-G..................-----------A....------------..-------    700
YU2           ----------G--..................------------....------------..-------    696
BCSG3C        ----------AG-..................------------....------------..-------    699
YU10          ----------G--..................------------....------------..-------    699
3202A12       ----------T-G..................------------....------------..-------    701
3202A21       ----------AG-..................------------....------------..-------    701
WEAU160       ----------GG-..................------------....------------..-------    700

CONSENSUS-D   AAAGCGAAAGTAG...................AACCAGAGaAGc....TCTCTCGACGCA..GGACTCG    690
ELI           -------------..................--------G---....------------..-------    700
Z2Z6          -------------..................------------....------------..-------    700
Z6            -------------..................------------....------------..-------    700
NDK           -------------..................-----------A....------------..-------    698

CONSENSUS-O   AA.GTGAAAGTGA...................AACCAGGGAAGA.AAACCTC.CGACGCAACGGGCTCG    621
ANT70         --.----------..................------------.-------.----------------    732

CONSENSUS-U   AAAGTGAAAGTAACAGGGACTCGAAAGCGGAAGTTCCAGAGAAGT....TCTCTCGACGCA..GGACTCG    648
MAL           ----------------------------------------------....------------..-------    719

CONSENSUS-CPZ A.?G?G??A???A...................?????G??????....??????GACGC??.GGACTCG    692
CPZGAB        -.A-T-AA-GTA-..................CCGGA-GCTGAA....TCTCCC-----AC.-------    719
CPZANT        -.G-A-CT-CAG-..................AACCT-T.................-----CA.-------     41
```

**HIV1 LTR**

```
                                                          \/ major 5' sj
CONSENSUS-A   GCTTGCTGAGGTGC..?CACAGCAAGAGGCGAGAG....CGGCGACTGGTGAGTACGCC.??AAATTTT?    743
U455          --------------..A------------------....--------------------.TA-------T    768
IBNG          --------------..G------------------....--------------------...-------.    774

CONSENSUS-B   GCTTGCTgaAGCGC??GCaCggCAAGAGGCGAGGGG...cGGCGactGGTGAGTACgCC?aaaaAtTtT?    761
SF2           --------------..----A---------------...--------------------....------T    761
LAI           --------------..--------------------...A-------------------.---------.    763
HXB2          --------------CC--------------------...--------------------.---------.    762
NY5           --------------..--------------------...--------------------.---------.    763
NL43          --------------..--------------------...--------------------.---------.    763
MN            --------------..--------------------...--------------------.-------C-T    760
JH31          --------------..----A---------------...-----G--------------.---------.    277
JRCSF         --------------..----A---------------...--------------------..G-------.    762
JRFL          --------------..--------------------...--------------------.---------.    572
OYI           -------T.-----..--------------------...--------------------.---------T    665
CAM1          --------------..----A---------------...--------------------.-TTT-----.    764
CDC41         --------------..--------------------...--------------------.---------.    761
HAN           --------------..--------------------...--------------------..--------.    755
P896          --------------..----A---------------...--------------------...-------T    761
D31           --------------..----A---------------...--------------------.---------.    309
RF            --------------..--G-----------------...------AC------------.G--------.    763
YU2           --------------..--------------------...-------------------A---------T    761
BCSG3C        --------------..--------------------...--------------------.---------.    762
YU10          --------------..--------------------...-------------------A---------T    764
3202A12       --------------..--------------------...--------------------..----C---.    763
3202A21       --------------..-----C--------------...-----G-----------C--.---------T    765
WEAU160       --------------..--------------------...--------------------..G-------.    762

CONSENSUS-D   gCTTgCTGAAGCGC..GCACgGCAAGAGGCGAGGGG...CAGCGACtGGTGAGTACGCT.?aAAATTTTt    753
ELI           --------------..--------------------...--------------------..---------    763
Z2Z6          --------------..--------------------...-------C------------.A--------.    763
Z6            T---T---                                                                  708
NDK           --------------..----A---------------...--------------------..G--------    761

CONSENSUS-O   GCTTAGCGGAGTGC..ACCCGCTAAGAGGCGAGAGGAACTCACAGAGGGTGAGT..........AATTTT    679
ANT70         --------------..--------------------------------------..........------    790

CONSENSUS-U   GCTTGCTGAGGTGC..ACACAGCAAGAGGCGAGAG....CGGCGACTGGTGAGTACGCC....AATTTTT    708
MAL           --------------..------------------....--------------------....-------    779

CONSENSUS-CPZ G??TGCTG??G?GC..?C??AGCA??AGGCGAG.G....CGG??A??GGTGAGTAC??????????TTTT    731
CPZGAB        -CT-----A.-T--..A-AC----AG-------.-....---CG-CT---------GC.....AAT----    776
CPZANT        -TC-----CA-C--..G-GT----GA-------.-....---..-AG---------CTTCATTTTG----    102
```

```
                    -> psi locus (continues 34 nucleotides into gag cds) ->
                                        gag cds ->
CONSENSUS-A   ?.GACTAGCGGAG......GCTAGAAGGAGAGA?A                                  769
U455          T.-----------......--------------A-                                  796
IBNG          ..-----------......--------------G                                   800

CONSENSUS-B   ?.GACTAGCGGAG......GCTAGAAGGAGAGAGA?                                 788
SF2           ..-----------......----------------G                                 789
LAI           ..-----------......---------------                                   789
HXB2          ..-----------......---------------                                   788
NY5           ..-----------......---------------                                   789
NL43          ..-----------......---------------                                   789
MN            ..-----------......---------------                                   786
JH31          ..-----------......---------------                                   303
JRCSF         ..-----------......---------------                                   788
JRFL          ..-----------......---------------                                   598
OYI           ..-----------......----------------                                  692
CAM1          ..-----------......---------------                                   790
CDC41         ..-----------......---------------                                   787
HAN           ..-----------......----------------                                  782
P896          T.-----------......---------------                                   788
D31           ..-----------......----------------T                                 337
RF            ..-----------......---------------                                   789
YU2           ..-----------......---------------                                   787
BCSG3C        ..-----------......----------------                                  789
YU10          ..-----------......---------------                                   790
3202A12       ..-----------......---------------                                   789
3202A21       ..-----------......---------------                                   791
WEAU160       ..-----------......---------------                                   788

CONSENSUS-D   ..GACTAGCGGAG......GctAGAAGGAGAGAGA                                  780
ELI           ..-----------......---------------                                   789
Z2Z6          ..-----------......---------------                                   789
NDK           ..-----------......-..-------------                                  786

CONSENSUS-O   ...GCTGGCGGTGGCCAGACCTAGGGGAAGGGCGAAGTCTCTAGGGGAGG.AAG               729
ANT70         ...-------------------------------------------------.---              840

CONSENSUS-U   ..GACTAGCGGAG......GCTAGAAGGAGAGAG                                   734
MAL           ..-----------......---------------                                   805

CONSENSUS-CPZ ???GCTAG???TG???A??CCTAG?GGAAGG?CGAAGTCTCTAGG???AG?AAG               768
CPZGAB        ...-----CGG--GCT-GA-----A------G-------------GGG--.---               826
CPZANT        GTT-----...--AGC-AT-----G------T-------------AAC--G---               153
```

## HIV1 LTR CONSENSUS

```
     U3 start ->
CONSENSUS-A   TGGATGGGTTAATTtACTCcaaGAaAAGACAAGAAATCCTTGATCTGTGGGTCTAtaACACACAAGGaTt     70
CONSENSUS-B   ----a---c-----------C-a-----------t---------------a----cc----------C-a     70
CONSENSUS-D   ----A---C------GG--?--a--?--------g-----------T--------C-----------CA-     68
CONSENSUS-O   ----?---------------C-T--?---GC---------G-----?---?-?---?----T--G-----     64
CONSENSUS-U   --------?--?---??---???--------------------?-----------?C-?--------C-?     59
CONSENSUS-CPZ ----A------G---------G--G---------G--------C--C---------C----------C--     70

                                                   AP-1          AP-1
                          /<-        COUP        ->/<-->/<-COUP->/<-->/
CONSENSUS-A   CTTCCCtGATTGGCAGAAtTACACACCAGGGCCAGGgAccAGATtCCC.ACTaACaTTTGGaTGGTGCTt    139
CONSENSUS-B   ------------------c-------------------t-----at--?---G--c--------------    139
CONSENSUS-D   ------------------C-------------------T-----At--.---g--C--------------    137
CONSENSUS-O   ----------------???--------?--A-----A?-?--?-----.---G-------------TG--    126
CONSENSUS-U   ---------------?--?--------------------??----?---.---G???--?-----------    119
CONSENSUS-CPZ ---------C--------C------A----A-----A--A--------.---CTGT--------------    139

                                        /<-             NF-AT              ->/
CONSENSUS-A   CAAgCTAGTACCAgT...tGATCCAGc?GaaGTAGAGgAaGccActG?aGGAGAGAaCAACAG.......    197
CONSENSUS-B   ---------------..---g----aga-g-----a--g----a--a---------------???????    199
CONSENSUS-D   -g-------------..-------C?G--G-----A--g-------A--------c----t-.......    196
CONSENSUS-O   T--A--?--------GTCA-?AGA--AG-C--A?AGACT?-G??A-AC-??T--????GCT-?.......    176
CONSENSUS-U   ?---?--------?-...???----??------------?----A-?--------------T-.......    167
CONSENSUS-CPZ ---A-----G--CC-GACA--...G-AGC-G-----AC------A--A------T------T-.......    199

                                                          /<-  NF-AT  ->/
CONSENSUS-A   .....tcTAtTACACCCtATaTGcCAACATGGAaTgGAtGAtGA?GA?AgAGAAgtgtTaATgTGGAagT    260
CONSENSUS-B   ?????ct-g-----------ga----G-----g--------Cccg--G------------g------g--    264
CONSENSUS-D   .....CT-G---------?-?-----G-----------G--C?CG--G---?---------a-----GA-    257
CONSENSUS-O   .....---?C-?--T--AGC?--T???------??T--?---?CAC--???--?A-AC-?-??---?---    222
CONSENSUS-U   .....?--G--------?--?A----?-----------?--?-?---A---------?---??-------    221
CONSENSUS-CPZ .....C--G--G--T--C-----T--G-----G-----A-----G--C-A---G--C--GG-C---CGC-    264

                             /<- IL2/NRE ->/
                          <-     NRF     ->
CONSENSUS-A   TTGACAGtagcCTGGCATTaAaACACAgAGCt?aaGAGcTGCATCCGGAGTtCTAcAAA....GACTG..    323
CONSENSUS-B   -------cc----a-----tC-t----tg--CCG----------------a------?..??-----??    327
CONSENSUS-D   --A----C--A--A-----TG-?----AG--CCG----?--------------------...-----..    319
CONSENSUS-O   ----T--ATC??-A-GC???-C?--T?T?--?-TG??AA??--C--A---C---T-C???A???---?..    269
CONSENSUS-U   -------C??---A---C?--?----??---C-G---????----------?-------...----?..    272
CONSENSUS-CPZ -------C--G-----C----G---T-TT--CAG---A-AA----------A-------...------..    328

                                                                          /<-NF-kB ->/
CONSENSUS-A   ..............????????CTGACACAgAAGTTGCTGACA.................GgGACTTTCC    354
CONSENSUS-B   ??????????????--------------tc--gc--t--.---a................----------    358
CONSENSUS-D   ......................-------C--Gt--t--.---A................----------    350
CONSENSUS-O   ..............-------??????????????????????..........???CTG?--?-------    281
CONSENSUS-U   ..............-----------------------------?..............----------    303
CONSENSUS-CPZ .........ACTTTAGA.......---G-G--.......---GC---AGAACTGCTGACTCTGC----------    379

                                   /<- NF-kB ->/      Sp-1       Sp-1     Sp-1
CONSENSUS-A   ..........aCT........GGGGACTTTCCg?.GGGAGGTGTGGTTTGGGcGGAGTTGGGGAGTGGCT    404
CONSENSUS-B   ..........g--........-----------aG.--?---c----cc-------gAC-----------g    408
CONSENSUS-D   ..........G--........-----------aG.--.---c----ac-------GAC------------    400
CONSENSUS-O   A?????GACTG--GACACTGC-----------AG??-??--???-A?A?---GC-GT-C-----------    337
CONSENSUS-U   ..........G--........------------AG.------?--??C?-------G???---------?-    346
CONSENSUS-CPZ A....................A-----------GG-A------GTCGGGA-----G-------T------    429

                                                          -LBP-1 -  -LBP-1
                                signal ->                 mRNA ->  -UBP-1- /<- TAR CORE
CONSENSUS-A   ...........AaCCCTCAGAtgcTGCATATAAGCAGCTGCTTtTCGCCTGTACTGGGTCTCTCTTGtTA    463
CONSENSUS-B   ??????????-g--------------------------------t-------------------g----    467
CONSENSUS-D   ...........-----------------------------------t-------------------G----    459
CONSENSUS-O   ..........----------?----------------?-----C?--T-----C------??G--AGAG    391
CONSENSUS-U   ..........------------------?-------?---------?----------------?----    401
CONSENSUS-CPZ T..........TG--------.----------------------CT-----------------T-CAC-G    488

                              TAR CORE ->/
                -LBP-1-          UBP-2 ->/                        signal ->
CONSENSUS-A   GACCAGAT??GAGcCTGgGAgCTCTCTGGCTAGCGagGG.AACCCACT.GCTTAAGCCTCAATAAAGCTT    529
CONSENSUS-B   --------ct-----------------------T----.--------?---------------------    535
CONSENSUS-D   --------TT-------a----------------T----.--------.---------------------    527
CONSENSUS-O   ------G-CT-----C--------C-----CT.-T--CTG-----G--.------CG-------------    459
CONSENSUS-U   ------?--------C------------------.-A--.--------.---------------------    465
CONSENSUS-CPZ --------TT-----------------A.-C--T--A--.--------.---------------------    555

                    <- mRNA
CONSENSUS-A   GCCTTGAG?TGCTT?AAGTA.GTGTGTGCCCGTCTG?TT?T?TGACTCTGGTAACTAGAGATCCCTCAGA    593
CONSENSUS-B   -------.G-----c-----.----------------.--g-G----------------------------    602
CONSENSUS-D   -------.G-----C-----.----------------.--G-G----------------------------    594
CONSENSUS-O   -------.G--...AG-AGCA--------T-A----.--CA...--C-----GT----------------    521
CONSENSUS-U   -------.G---C-C---C-.----------A----.--G-G----------------------------    532
CONSENSUS-CPZ -------.G--TA-TTGAGCA---------TA---A.--CAGAC--------------------------    623
```

```
                                                <- LTR
                                                  <- Lys-tRNA pbs
CONSENSUS-A   CCACT?TAGACTGTGT..AAAAATCTCTAGCAGTGGCGCCCGAACAGG?????????????????GACTCG    644
CONSENSUS-B   --ct-t---t-A----gG?-----------------------------.................----t-    655
CONSENSUS-D   ?-C--T---T-a-A--Gg?-----------------------------..................---CT-    646
CONSENSUS-O   T----.-------AAGCAG-----------------------------..................----T-    574
CONSENSUS-U   -----C-----G----..-----------------------------..................----TT    584
CONSENSUS-CPZ TT-AAT--T-G-CAA.GG.-----------------------------................---?T-    674

                                                                        /<- psi locus
CONSENSUS-A   AAAGCGAAAG.......................TTCCAGAGAAG?....TCTCTCGACGCA?.GGACTCG    684
CONSENSUS-B   ----------?aa....................aa------g--c....------------..-------    698
CONSENSUS-D   ----------TAG....................AA---------c....------------..-------    690
CONSENSUS-O   --.-T-----TGA....................AA----G----A.AAAC---.-------AC--G----    621
CONSENSUS-U   ----T-----TAACAGGGACTCGAAAGCGGAAG-----------T....------------..-------    648
CONSENSUS-CPZ -.?-?-??-???A....................?????-?????-....??????-----?-.-------    692

                                        - - - psi locus - - -
                                                                  \/ major 5' sj
CONSENSUS-A   GCTTGCTGAGGTGC..?CACAGCAAGAGGCGAGAG....CGGCGACTGGTGAGTACGCC.??AAATTTT?    743
CONSENSUS-B   ---------A-C--??G---g------------G-G...--------------------?aa--------    761
CONSENSUS-D   ---------A-C--..G---g------------G-G...-A-----------------T.-a-------t    753
CONSENSUS-O   ----AGC-GA----..A-C-GCT------------GAACTCA-AGAG-------..........-A---T    679
CONSENSUS-U   --------------..A------------------....---------------------...------T    708
CONSENSUS-CPZ -??-----??-?--..--??----??-------.-....---??-??---------????--????---T    731

                     -> psi locus (continues 34 nucleotides into gag cds) ->
                                          gag cds ->
CONSENSUS-A   ?.GACTAGCGGAG......GCTAGAAGGAGAGA?A                                       769
CONSENSUS-B   -.-----------......--------------G-?                                      788
CONSENSUS-D   ..-----------......--------------G-                                       780
CONSENSUS-O   ...G--G----T-GCCAGAC----GG-A--G-CG-AGTCTCTAGGGGAGG.AAG                    729
CONSENSUS-U   ..-----------......--------------G                                        734
CONSENSUS-CPZ -??G----???T-???A??C----?G-A--G?CG-AGTCTCTAGG???AG?AAG                    768
```