HMMER Sequences in the Tat Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 484 (1985)
C_UG268A2	HIVUG268A2	L22948	Louwagie,J.J.	JVI69, 263 (1995)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
F_BZ163A	HIV1BZ163A	L22085	Louwagie,J.J.	ARHR 10, 561 (1994)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment and most likely sequence were generated by the HMMER program as described in Part III. For simplicity, only representative types and subtypes are shown. The annotation is based on HIV1s, therefore the user should be cautious about its applicability to other PIV sequences. The BLOCKMAKER program was employed to identify one conserved, gapless block (using either the Motif or Gibbs approach); the block has been indicated with shading. A putative nuclear localization signal (nls) is shown. The Motif "Cobbler" sequence follows this alignment. For further information regarding the Tat protein, see the section by K.-T. Jeang in Part III of this compendium.

```
                                                                                                  \/3'sj
most-likely MEPVDPE....LEPWNHPGSCTSEADAATQELANLGEEILSQLY.RPKTACTNKCYCKKCCYHCQVCFLKKGL     68
U455        ------N...----K----........................Q.-T---S-.----V--W---L-------     43
HXB2R       ------R...----K----........................Q.-------.-------F------IT-A-     43
UG268A                      X--.......................Q.---P--.--FG---S--------N---     30
ELI         -D----N...---------.......................Q.-R-P-N.--H--------P----N---     43
UG269A                                                                            X-      2
NDK         -D----N....--S------......................Q.-R---N.--H------------IT---     43
SE365A      -D----N...---------.......................Q.---P-N-.------------N--IT---     43
UG274A2     SVXI--R...---------.......................Q.-S-P-N.------------L--IT---     43
MAL         -D----N...---------.......................Q.-R-P-N.------------M--IT---     43
BZ163A      --L---N....-D-------......................Q.-T-P--.-----R--F---W--TT---     43
ANT70       -D-----....VP--H----......................Q.-QIP-N-.----R-----Y---VR---     43
MVP5180     -D-----....MP--H----......................K.-Q-P-N-.----R-----Y---T----     43
CPZGAB      -D-I--D...----K----......................Q.-R-V-N-.----A-----IY--T----     43
CPZANT      -D---A-....TP--L--PA......................T.-A-P-N-.----C-----PL--T----     43
ROD         --TPLKAPESS-KSC-E-F-R---Q-V------RQ---------.--LET-N-S----R------M---N---     72
EHOA        --IPLK-QESS-NSSSGHS-S---GV-N--G-D-R---------.--LK--S-T------S----L-------     72
MM251       --TPLR-QENS--SS-ERS--IL----T-P-S------------.--LE--Y-T-----------F-------     72
STM         --TPLK-QESS-RSSSE-S------V---PG---QE----W---.--LEE-C---F---------L--VT---     72
AGM3        -...-KG....EDEQGAYHQ................DLIE--K.A-LKR--------C------L---Q---     48
AGM677      -...-K-....E--..--LL...................QD-H.--LQP--------------EL---Q---     43
SAB1C       -...-Q-QEARPQV-EELQE....................E-H.--LQ--D-T-F--V--F--IL--H--A-     48
SYK         -SST-QI....CQTQRV-P-FLEGTFLEK...............G.-P-P-N.--F--N------L---Q---     52

                \/3'sj
       rev cds ->
                <-  nls  ->            exon \/ exon
most-likely GISYGRKKR..RQRRR.TPQKSQTHQVSLS.KQPISQPRGD.PTGPKESKKK.VERETETDP...........    124
U455        -----K---..KP--G.P--G-KD--TLIP.---LP-SQRV.SA-QE-----.--SKAK--R...........     99
HXB2R       ---------..------.AH-N-----A---.---T------.------X---.---------...........     99
UG268A      ---------..----N.A-PS-ED--NLI-.---LP-----.----EK----.--SK-----...........     86
ELI         ---------..----G.P--GG-A---PIP.---S------.------Q---.--S-A----...........     99
UG269A      ---------..------.S-KGG-A---PIP.---S------.------X---.--SKA-A--...........     58
NDK         ---------..----K.P--GD-A---PIP.E--S--S---.-----KX---.--S-A----...........     99
SE365A      ---------..------.---GN------VP.---S--H---.------X---.--SKA---Q...........     99
UG274A2     ---------..------.P--S--A--DPIP.---S-----N.-----KX---.--SKA-A--...........     99
MAL         ---------..------.P--GN-A--DP-P.E--S--H---H------X---.--SKA-A-Q...........    100
BZ163A      ---------..K--H-.---S--I--DLVP.------A--N.-----K---E.--SKA----...........     99
ANT70       ---------..G..-P.AAASHPD-KDPVP.--SPTI.TKR.KQERQ-EQEEE--KKAGPGGYPRRKGS.CHC    107
MVP5180     ---H-----..-.-PA.AAASYPDNKDPVP.E-SL-H.T-R.KQKRQ-EQE-K--K--GPSGQPCHQDSCNSC    109
CPZGAB      ----------.TT---TA-AG-KNN-D-IP.---L--S--N.KE-SEK-T-E.-ASK--A-Q...........    100
CPZANT      -------R-AR-N--T.-AES-ENN-DPV-.-SLPKTS-IQ.SSQK-XE-.-.--EK-GSGG...........    100
ROD         --C-E--G-..-..--.--K-TK--PSPTP.DKS--TRT--.SQPT-KQ--T.--ATV---T...........    126
EHOA        --C-E-S....-..K-.SSKRAK-TTS-APNNRSL-ART--.SQPT-KQ--E.--TTRA--LGPGRSN....T    132
MM251       --C-EQS....-K---.--K-AKANTS-A-.NKL-PN.-TR.HCQ-EKA--ET--KAVA-A-...........    126
STM         --T-E-S....-..--.VKK-AK-YPI-A-NNRSL-TRARN.SQPK--Q--E.--T-V-ST-...........    125
AGM3        -VT-HA.P-..I..--.KKIAPLD.RFPEQ.--S--TRGR-.SQTTQKGQE-.--TSAR-A-SLGRKNLAQQS    111
AGM677      -VR-HVS....-..K-.RKTST-DN-DPIR.Q-S--TVQRN.GQTTE-G-TE.--                       88
SAB1C       --R-YV.P-..P..--.ASK-ISHN----H.N                                               73
SYK         --T-A-.P-..K..-A.ARSI-ED..D-AP.TGTLPR.A-R.TQANPQT---A--T-KDSTSLP.........    105
```

HIV1 TAT

```
most-likely ....G.DR                                                    127
U455        ....F..A                                                    101
HXB2R       ....F.-X                                                    102
UG268A      ....X.-X                                                     89
ELI         ....XDCX                                                    103
UG269A      ....F.-X                                                     61
NDK         ....FDWX                                                    103
SE365A      ....F.-W                                                    102
UG274A2     ....F.-W                                                    102
MAL         ....FDWX                                                    104
BZ163A      ....XDXX                                                    103
ANT70       CTRTS.EQ                                                    114
MVP5180     TRIS-.QX                                                    116
CPZGAB      ....X.-X                                                    103
CPZANT      ....-.PC                                                    103
ROD         ....-PG-                                                    130
EHOA        STSRFANX                                                    140
MM251       ....-LG-                                                    130
STM         ....-LGK                                                    129
AGM3        GRAT-.AS                                                    118
SYK         ..SAE.NL                                                    110
```

```
      **COBBLER sequence from MOTIF**
>TAT   UG274A2, with embedded consensus blocks
svxidprlepwnhpgsqpstpCNKCYCKKCCYHCQLCFLKKGLGISYGRKKRRqrrrppq
ssqahqdpipkqpssqprgnptgpkkxkkkveskaeadpfdw
```

```
                                  intramolecular    3'sj   3'sj
                                 disulfide bonding   \/     \/
                                 |              | rev cds ->/<- nls ->/
CONSENSUS-A   M?PVDPnLEPWnHPGSqPtTaCskCYCK?CCwHCqlCFLnKGLGISYGrKKR..r?RRgtPQs?kDhQnp     64
CONSENSUS-B   -e----r----k------k---tn----k--f---v--tt------------..-Q--ra--dSqt--vs     68
CONSENSUS-C   ------?-----------K---t-----k-sY--lV--qt------------..-q--sa-?-SE-----     65
CONSENSUS-D   -d----------------?-p-N--h--K--Y---v--it------------..-Q--rp--ggQa--?-     66
CONSENSUS-F   -EL-----D-----------P-T-----R--F---W--TT------------..KQ-HR----SQI--DL     68
CONSENSUS-O   -D----E?P--H----?-Q?P-NN----R--Y--YV--??------?-----...-???AAA--P-?KD-     55
CONSENSUS-U   -D----K-----------K---T-----K--Y--PV----------------..-P--RS--NSE-----     68
CONSENSUS-CPZ -D-?-????--?--???-?-?-NN-------Y--??--TK----------?-???---T????S?NN-D?     45

            exon \/ exon
CONSENSUS-A   ipKQplPqtqg??ptgpkESkKkVeSKteTDrf?$                                       95
CONSENSUS-B   Ls---?s-pr-D.------------rE----P?d?                                       99
CONSENSUS-C   -s-------r-d.----E-------------p-D-                                       98
CONSENSUS-D   -----SS-pR-d-------?-------A---p-Dw$                                      99
CONSENSUS-F   V----IS-AR-N.-----?---E----A??-P?--$                                      96
CONSENSUS-O   V-?-S???-?RK.Q?RQE-QE??--K??GP?G?P????SC??CTR?S?Q$                        83
CONSENSUS-U   ----S--H--RV.S---E----E----A-----D-                                      101
CONSENSUS-CPZ ??-??-?????-.?????K??-?-??--?????-?                                       52
```