# III

# Analysis

**PART III Analysis**

# Gag Proteins

**Robert J. Gorelick and Louis E. Henderson**

*AIDS Vaccine Program, PRI/DynCorp, National Cancer Institute,*
*Frederick Cancer Research and Development Center, Frederick, Maryland 21702*

## INTRODUCTION

All retroviruses utilize a *gag* gene to encode a precursor polyprotein (Gag precursor) that is initially incorporated in the budding particle and ultimately cleaved by a viral protease (encoded by a *pol* gene) into structural proteins of the mature virus collectively referred to as Gag proteins (See Figure 1 for biosynthesis of HIV-1 virion proteins). Gag precursors are self-associating proteins capable of assembling into virion-like particles without the requirement of other viral factors. For most members of the retrovirus family, including lentiviruses, assembly takes place on the cytoplasmic side of the cell membrane and nascent virion-like particles bud through the membrane into extracellular space. During normal viral replication the self-associating Gag precursors selectively bind viral RNA and package two plus strands of the genome inside the particle. They also form associations with viral Env proteins such that the budding particle is encapsulated within a lipid envelope containing viral Env proteins. All retroviruses express viral encoded enzymes from a *pol* gene as Gag-Pol fusion proteins and the self-associating properties of the Gag portion directs assembly of the Gag-Pol fusion protein into the budding particle. Thus, the Gag precursor orchestrates assembly of the nascent virion by forming the core structure and interacting with all other viral components (RNA, Pol proteins, envelope lipid and Env proteins) to facilitate their incorporation into the budding viral particle. The term "Gag precursor" was originally adopted to indicate that the major antigenic protein (capsid antigen, CA; [1]) or "Group antigen" (Gag) was synthesized as a precursor. The term does little to convey the important biological properties of the protein and a more descriptive name such as "Assemblin" might be more appropriate.

## HIV-1 GAG PRECURSOR (ASSEMBLIN)

HIV-1 *gag* genes are generally about 500 codons in length. Figure 2 is a composite diagram indicating posttranslational modifications, proteolytic cleavages and sites for various interactions that are known for the HIV-1 Gag precursor (Assemblin, $Pr55^{gag}$). The upper panel in Figure 2 indicates the posttranslationally modified form of the protein that participates in the assembly of the immature virion particle. During or shortly after assembly, the viral protease (PR; product of the *pol* gene) is activated and the Gag precursor is cleaved (middle panel) into the proteins found in the mature virus (lower panel). The Gag precursor is thought to be organized into domains that roughly correspond to the final proteolytic cleavage products [matrix antigen (MA), capsid antigen (CA), nucleocapsid (NC), and p6] that make up the core of the mature virus.

During or shortly after translation, the initiator Met residue is removed leaving Gly-2 as the new N-terminal residue. Cellular myristoyltransferases catalyze the covalent attachment of a myristoyl group [$CH_3(CH_2)_{12}COO^-$] (Myr) to the amino group of Gly-2 via an amide bond and cellular kinases phosphorylate at least two sites on the protein in the CA domain (one serine and one threonine residue, but the exact sites have not been identified). The protein also binds two zinc ions to form zinc-fingers in the NC domain. As indicated in Figure 2, these posttranslational modifications carry through to the Gag proteins found in the mature virus. Thus MA is myristoylated at the N-terminal end, CA is phosphorylated at two sites and the NC protein contains two zinc fingers.

The myristoyl group and the first 31 N-terminal amino acid residues of the MA domain are necessary for targeting the Gag precursor to the cell membrane [2, 3]. The MA domain of the Gag precursor has also been implicated in binding and assembly of the *env* gene products into mature virus [4, 5]. Presumably the interactions with the *env* gene products involve the cytoplasmic domain of the transmembrane (TM) protein associating with the MA domain, however this has not been directly demonstrated.

The CA domain of the precursor appears to be involved in self-association or Gag/Gag interactions [6]. These interactions have been narrowed down to amino acid residues ~240–430 (see Figure 2, upper panel; [7, 8]). Located within the CA domain is a region called the major homology region (MHR) which is required for efficient viral replication and particle production [9]. The MHR includes residues 285–304 in the HIV-1 (IIIB) CA domain. Amino acid residues, 287/Q, 291/E, 296/Y, and 299/R (in the sequence of IIIB) are completely conserved in the MHR region of all primate lentiviruses.

A region between residues 178 and ~300 in the CA domain of the Gag precursor has been shown to bind Cyclophilin A [10]. Cyclophilin A is a peptidyl-prolyl cis-trans isomerase thought to play a role in immune regulation and has been found in preparations of purified HIV-1 (Henderson, unpublished). Other protein-protein interactions of the HIV-1 Gag precursor include interactions of the p6 domain with the viral encoded Vpr protein [11, 12].

Gag protein-protein interactions also play a significant role in packaging the viral enzymes encoded by the *pol* gene. The *pol* gene products are expressed as a Gag-Pol fusion protein which results from synthesis initiating as for the Gag precursor and progressing in the *gag* frame to a site just past the NC domain (Gag-Pol frameshift site, Figure 2) where a –1 ribosomal frameshift event occurs followed by synthesis continuing in the *pol* frame [13]. The resulting fusion protein is referred to as the Gag-Pol precursor (Pr160Gag-Pol) and contains the Gag sequences required for association with Gag precursors and incorporation into virion particles. The fusion protein is ultimately cleaved by the activated viral PR to the mature Pol products (PR; reverse transcriptase, RT; and integrase, IN) and the corresponding Gag proteins (MA, CA, p2, NC) and peptides derived from the *pol* reading frame preceding the codon for the N-terminal residue of PR. The "transframe" peptide isolated directly from whole virus [HIV-1 (MN)] has the sequence FLREDLAF [14], where the first two residues (FL) are encoded by the *gag* gene and correspond to the first two residues of the p1 peptide (lower panel, Figure 2) and the remainder of the transframe peptide is encoded by the *pol* gene.

In addition to the various protein-protein interactions involved in the assembly, the Gag precursor also binds and packages the viral genomic RNA. The NC domain of the Gag precursor is directly involved in the nucleic acid binding properties of the protein. The NC domain (Figure 2, upper panel) is rich in basic residues (Lys and Arg) and contains two copies of a zinc binding sequence [$Cys(X)_2Cys(X)_4His(X)_4Cys$] referred to as a retroviral CCHC zinc finger [see Figure 3; 15]. The basic residues confer a nucleic acid binding property to the domain and the CCHC zinc fingers are involved in the specificity for the viral RNA [16–18]. Apparently, during viral assembly, the NC domain of the Gag precursor specifically binds to a site near the 5′ end of the genomic RNA referred to as the Psi site. If either of the two zinc fingers in the NC domain are disrupted by site directed mutagenesis, or the Psi sequence in the RNA is deleted, noninfectious particles are produced with reduced levels of packaged viral genome [16–19]. Mutants with the first finger replaced by the second finger also package reduced levels of viral RNA, but a mutant with the second finger replaced by the first (i.e., NC domain with two identical copies of the first finger) packages normal levels of viral RNA [20]. Thus, even though both zinc fingers are required for specific RNA packaging, they are not functionally equivalent in this regard.

## MATURE GAG PROTEINS

After the Gag precursor assembles into the immature virion, the activated viral protease cleaves the protein into the mature Gag proteins found in the infectious virus (Figure 2). Based on *in vitro* analysis, the order of cleavage is as follows: p2/NC > MA/CA = p1/p6 > CA/p2 > NC/p1. The MA/CA and p1/p6 sites are cleaved approximately 12-fold slower and the CA/p2 cleavage proceeds about 400-fold slower than the initial p2/NC cleavage [21–23]. Following proteolytic cleavage (Figure 2) the mature Gag proteins rearrange to form the core structure of the mature infectious virus. The mature MA protein remains associated with the inner side of the lipid envelope while the CA protein forms the outer shell of the capsid structure which contains the NC protein bound to the genomic RNA. During the infection process, RT and IN together with the MA protein, NC

protein and the genomic RNA are taken into the cell as the "pre-integration complex" [24]. It has been suggested that proviral DNA synthesis takes place within the pre-integration complex. The mature MA protein plays a role in transport of the viral nucleoprotein complex into the nucleus for integration of provirus into the host genome [24, 25]. The MA protein detected in the nucleus is phosphorylated [26], whereas the MA protein in the virus is not. Thus, the phosphorylation of MA occurs during or after the viral infectious event. Within the MA protein there is a nuclear localization signal (NLS) [25] that is similar to other nuclear targeting sequences. The signal is located between residues 14–36 and 107–116 of the HIV-1 p17 MA protein (see Figure 2) and mutants lacking the signal accumulate proviral DNA in the cytoplasm. The NC protein may also play a role in the infection process. Mutants of the NC protein which still package RNA but have alterations within the zinc-finger structures have reduced infectivity [20] suggesting that a role is performed by the NC protein which may be at the level of the preintegration complex. The NC protein has additional functions in the virus life cycle. These include strong nucleic acid annealing and melting activities [27, 28]. It has been suggested that the NC protein is involved in annealing of the primer tRNA to the primer binding site of the viral genome. These activities may also have implications in strand transfer reactions that occur after the first strong stop synthesis of the proviral DNA, to complete synthesis of the provirus for subsequent integration. The NC protein has also been shown to catalyze the formation of the genomic RNA dimer found in virion particles [29]. A recent report showed that HIV-1 with a defective protease had an RNA dimer structure that was different from that found in wild-type HIV-1 [30]. This indicates that dimer formation is probably catalyzed by the mature NC protein and not by the Gag precursor.

---

**References**

[1] J. Leis, D. Baltimore, J.M. Bishop, J. Coffin, E. Fleissner, S.P. Goff, S. Oroszlan, H. Robinson, A.M. Skalka, H.M. Temin, and V. Vogt (1988) Notes: Standardized and simplified nomenclature for proteins common to all retroviruses. J. Virol. 62:1808-1809.

[2] W. Zhou, L.J. Parent, J.W. Wills, and M.D. Resh (1994) Identification of a Membrane-Binding Domain within the Amino-Terminal Region of Human Immunodeficiency Virus Type 1 Gag Protein Which Interacts with Acidic Phospholipids. J. Virol. 68:2556-2569.

[3] M. Bryant and L. Ratner (1990) Myristoylation-dependent replication and assembly of human immunodeficiency virus type 1. Proc. Natl Acad. Sci. USA 87:523-527.

[4] T. Dorfman, F. Mammano, W.A. Heltine, and H.G. Göttlinger (1994) Role of the matrix protein in the virion sociation of the human immunodeficiency virus type 1 envelope glycoprotein. J. Virol. 68:1689-1696.

[5] X. Yuan, X. Yu, T.-H. Lee, and M. Essex (1993) Mutations in the N-terminal region of human immunodeficiency virus type 1 matrix protein block intracellular transport of the Gag precursor. J. Virol. 67:6387-6394.

[6] L.S. Ehrlich, B.E. Agresta, and C.A. Carter (1992) Assembly of recombinant human immunodeficiency virus type-1 capsid protein in vitro. J. Virol. 66:4874-4883.

[7] J. Luban, K.B. Alin, K.L. Bossolt, T. Humaran, and S.P. Goff (1992) Genetic assay for multimerization of retroviral gag polyproteins. J. Virol. 66:5157-5160.

[8] E.K. Franke, H.E.H. Yuan, K.L. Bossolt, S.P. Goff , and J. Luban (1994) Specificity and sequence requirements for interacitons between various retroviral gag proteins. J. Virol. 68:5300-5305.

[9] F. Mammano, A. Ohagen, S. Hoglund, and H.G. Göttlinger (1994) Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. J. Virol. 68:4927-4936.

[10] J. Luban, K.L. Bossolt, E.K. Franke, G.V. Kalpana, and S.P. Goff (1993) Human immunodeficiency virus type 1 gag protein binds to cyclophilins A and B. Cell 73:1067-1078.

[11] C. Lavallée, X. J. Yao, A. Ladha, H. Göttlinger, W.A. Haseltine, and É. Cohen (1994) Requirement of the Pr55gag precursor for incorporation of the vpr product into human immunodeficiency

virus type 1 viral varticles. J. Virol. 68:1926-1934.

[12] W. Paxton, R.I. Connor, and N.R. Landau (1993) Incorporation of vpr into human immunodeficiency virus type 1 virions: Requirements for the p6 region of gag and mutational analysis. J. Virol. 67:7229-7237.

[13] T. Jacks and H.E. Varmus (1985) Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science 230:1237-1242.

[14] L.E. Henderson, unpublished.

[15] J.M. Berg (1986) Potential metal-binding domains in nucleic acid binding proteins. Science 232:485-487.

[16] A. Aldovini and R.A. Young (1990) Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. J. Virol. 64:1920-1926.

[17] Gorelick, R.J., S.M. Nigida, Jr., J.W. Bess, Jr., L.O. Arthur, L.E. Henderson, and A. Rein (1990) Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. J. Virol. 64:3207-3211.

[18] T. Dorfman, J. Luban, S.P. Goff, W.A. Haseltine, and H.G. Göttlinger (1993) Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. J. Virol. 67:6159-6169.

[19] R. Carroll, J.-T. Lin, E.J. Dacquel, J.D. Mosca, D.S. Burke, and D.C. St. Louis (1994) A human immunodeficiency virus type 1 (HIV-1)-based retroviral vector system utilizing stable HIV-1 packaging cell lines. J. Virol. 68:6047-6051.

[20] Gorelick, R. J., Chabot, D. J., Rein, A., Henderson, L. E., and Arthur, L. O. (1993) The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein ar not functionally equivalent. J. Virol. 67:4027-4036.

[21] Erickson-Viitanen, S., J. Manfredi, P. Viitanen, D.E. Tribe, R. Tritch, C.A. Hutchison III, D.D. Loeb, and R. Swanstrom (1989) Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease. AIDS Res. and Hum. Retroviruses 5:577-591.

[22] Krausslich, H.G., R.H. Ingraham, M.T. Skoog, E. Wimmer, P.V. Pallai, and C.A. Carter (1989) Activity of purified biosynthetic proteinase of human innumodeficiency virus on natural substrates and synthetic peptides. Proc. Natl. Acad. Sci. USA 86:807-811.

[23] Krausslich, H.G., G.H. Schneider, G. Zybarth, C.A. Carter, and E. Wimmer (1989) Processing of in vitro synthesized gag precursor proteins of human immunodeficiency virus (HIV) type 1 by HIV proteinase generated in Escherichia coli. J. Virol. 62:4393-4397.

[24] M.I. Bukrinsky, N. Sharova, T.L. McDonald, T. Pushkarskaya, W.G. Tarpley, and M Stevenson (1993) Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. Proc. Natl. Acad. Sci. USA 90:6125-6130.

[25] M.I. Burkrinsky, S. Haggerty, M.P. Dempsey, N. Sharova, A. Adzhubei, L. Spitz, P. Lewis, D. Goldfarb, M. Emerman, and M. Stevenson (1993) A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nature 365:666-669.

[26] M. Stevenson, personal communication.

[27] F. Dib-Hajj, R. Kawa, and D.P. Giedroc (1993) Retroviral nucleocapsid proteins possess potent nucleic acid strand renaturation activity. Protein Sci. 2:231-243.

[28] R. Khan and D.P. Giedroc (1992) Recombinant human immunodeficiency virus type 1 nucleocapsid (NCp7) protein unwinds tRNA. J. Biol. Chem. 267:6689-6695.

[29] J.-L. Darlix, C. Gabus, M.-T. Nugeyre, F. Clavel, and F. Barré-Sinoussi (1990) Cis elements and trans acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. J. Mol. Biol. 216:689-699.

[30] W. Fu, R.J. Gorelick, and A. Rein (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. J. Virol 68:5013-5018.

Sorry, figure not currently available on line.

**Figure 1. Biosynthesis of HIV-1 Virion Proteins**. Steps in the biosynthesis of proteins found in mature infectious HIV-1 virion particles are outlined in a schematic diagram. Proteins are indicated in color and designated according to convention. Naturally occurring proteolytic cleavage sites are indicated by red arrows and some posttranslational modifications are indicated (glycosylation sites in gp120 and gp41 are omitted).

Sorry, figure not currently available on line.

**Figure 2. Biosynthetic Processing of HIV-1 Gag Proteins**. The biosynthetic processing of the HIV-1 Gag precursor (Assemblin) and its natural proteolytic cleavage products is indicated in the schematic diagram. The upper panel shows the posttranslationally modified Gag precursor that participates in assembly and budding of the immature viral particle. The protein is depicted as organized into domains that correspond to final cleavage products, matrix antigen (MA), capsid antigen (CA), nucleocapsid (NC) protein and p6 (lower panel) found in the mature virus. The Gag precursor and MA are myristoylated (Myr-) on the N-terminal amino group of glycine (Gly-2). The precursor and CA are phosphorylated at two sites (one serine and one threonine residue). The NC domain of the Gag precursor and the NC protein bind two zinc ions to form zinc-fingers. The MA protein is not phosphorylated in the virus but is phosphorylated during the infectious process (indicated in the lower panel). The middle panel indicates the sites for proteolytic attack by the viral protease during the normal maturation process. The lower panel indicates the mature Gag proteins and also the products of partial proteolysis of p1, p2, and p6 found in the infectious virus.

**Figure 3. HIV-$1_{MN}$ Nucleocapsid (NC) Protein.** The upper panel shows the amino acid sequence of the NC protein and indicates the distribution of charged residues. Amino acid side chains that coordinate with zinc ions are indicated as well as aromatic residues that are required for specific RNA binding. The panel also lists fundamental molecular data for the protein. The lower panel is a cartoon to indicate the zinc-finger structures formed by binding metal ions.

## HIV-1 Gag Consensus Sequences

```
CONSENSUS.A mGARaSvLsggkLDawekIrLRPgGkKkYrlKHlvwAsreLerFaLnPslLeT?egcqqimeQlqsalkTg?e     71
CONSENSUS.B -----------e--r-------------k---i-----------v--g---ts---R--lg---Ps-q--s-     73
CONSENSUS.C ------i-r-----?------r-----h-Mi----------------pg----s---k--ikq--P--Q--T-     72
CONSENSUS.D --------------?---------g----?---i---------f-l--G----s---k--ig---P-iqt-s-     71
CONSENSUS.F ------------------------------?-------------??-g----s---rk-Ig---pS-Q--S-     70
CONSENSUS.G --------------?------------?--?-----------?-----G----T------??---P?-Q--T-     66
CONSENSUS.H -----------?---------------?--------------------?--?-----L-?I----P-----T-     67
CONSENSUS.O ---?----T-S------?---?--S--?---------------?-C--?----A--?E?LLQ--EP-----S?     64

CONSENSUS.A ElkSLfNtvatLyCvHqrIdvkDtKeA..ldkiEei.qnkskqk..tqq...aaA???????.........Tg    120
CONSENSUS.B --r--y-------------e-------..------E.-----k-..a--.??--ad............???t-    123
CONSENSUS.C --r--?----------??-e-r------..-----e-E.---?Q--??----..ak?-D..............?-    117
CONSENSUS.D e----?----------e--e--d-----..-e-m--E.-----k-.?a--???-t-D..............tr    120
CONSENSUS.F --r--?---?v--f----vE?--------..-?-L--E.----q--..------...?--d..............K-    115
CONSENSUS.G -?---?-??-?-?-?----?----k--..-eEV-Ka.-kn-Q--..------...???-..............?k-    104
CONSENSUS.H --Q----LL-?---------?-------..-?-??-?.--??Q??..------...-T?D..............K?    105
CONSENSUS.O ?-?--W-AI?V-W---N-??I?--QQ-..IQ-LK-VM.?SR-SA..?AA....KE-.............-S    104

                      p17 \/ p24
CONSENSUS.A nss?????....kvSqNYPIVQNaQgQm?hQ?lSPrTLnAwVKviEekaFspEVIPmFsaLSEGATpQdLNmM    182
CONSENSUS.B ---.........q---------nlq---V--ai-----------v---------------s---------T-    187
CONSENSUS.C ............-----y-----L----v--ai----------------f--e?ip-ft-lsegat-q-lntm    177
CONSENSUS.D ---.........Q----------L----V--ai---------k-----------i----------------t-    184
CONSENSUS.F ............----------l----V--ai---------------------------se--------T-    175
CONSENSUS.G ---.........Q---------n-----v--pis----------v-------------------------t-    168
CONSENSUS.H ???.........-?-------------?V--AI-----------V-------------------------A-    164
CONSENSUS.O .?R.........Q??------?------V--AI----------AV-----N--I----M------??Y-I-T-    161

CONSENSUS.A LNiVgGHQAAMQMLKdtINeEAAewDR?HPVhAgPippgQmREPrGSDIAGtTStlqEqigwmT...sNPPiP    251
CONSENSUS.B --T------------e----------l--v-----a-------------------------...n-----    257
CONSENSUS.C lntv-ghqaamqmlk--in-eaa--drlhpv-a-pva--q--e----------------a---...?-----    246
CONSENSUS.D --Tv---q-------E--n--------l--v-----A----------------------?---...------    253
CONSENSUS.F --T------------------------L--v?--------?------------------q---...----v-    243
CONSENSUS.G l-T----------l-------------I--pQ--------I-d----------------R---...------    238
CONSENSUS.H --?---------?-------------------------------------------------A---...?--?--    229
CONSENSUS.O --AI----G-L-V--EV-----?----T--P??--L----I---T----------Q----?-T-..R????--    224

CONSENSUS.A VGdIYkrwIiLGLNKIVRMYSPvSILDirQgPKEPFrdYVdrFfKtLRAeqAtQeVKnwMTeTLLvQNANPDC    324
CONSENSUS.B --e-------l----------t----------p----v---Y--------s------m---l--q-a----    330
CONSENSUS.C ----------------------------k-------------------------D------d----------    319
CONSENSUS.D -ge-------l------r---------------e---------Y--------s-d-----------------    326
CONSENSUS.F --e--------------------------------p----------------------g---D----------    316
CONSENSUS.G --e-------------------------------------------------?--?---D----------    309
CONSENSUS.H ----------------------------?----------------?--------?--?---D----------    298
CONSENSUS.O -----RK--V-----M-K----------?--------------Y-----------------------?----    295

                                    p24 \/  \/  'p2'       \/ p7
CONSENSUS.A KsILraLg?gAtLeEMMTacQgVggPgHKArvLAEAmSqvq??n...???iMmQrGnf.?Gqkr?iKCFNCGk    385
CONSENSUS.B -T--K---Pa-------t----------------------..t-..?sat--------.rn-rKtv-------    397
CONSENSUS.C -T------P--s-------------s-----------ann--.......-m---s--.K-p--iv-------    384
CONSENSUS.D kt--K---P?---------------s-----------a?.t-..s-ta-m---g--.K-prki--------    392
CONSENSUS.F -T--K-l-P-------------------------a---a..T-..?-a--m--ks--.K--R-iv-------    381
CONSENSUS.G -T--?---P-----------------?-----------A..SG..?A-A?---K??-.K-P??-----?---    366
CONSENSUS.H -?--?------SI-------------?---?---------.-T-..?-A-?---K---.K--R-I?-------    356
CONSENSUS.O -Q--K?--P?-------V--------T---??-----A?A-QDLKGGYTAVF----QN.P?R-G---------    360
```

```
                                    pol cds ->
                                         p7 \/     'p1'      \/ p6
CONSENSUS.A EGHlArNCrAPrKkGCwKCgkEGHQmKdCT.?e.rQANFlgkiwpSsKG.RPgNFpQsRp.............    441
CONSENSUS.B e--i-k------k----k----------t.--?-----------h--.-p---l----?????????..??    453
CONSENSUS.C ---i--n------------?---------..-.-----------s?--.-----L----eptap????????    443
CONSENSUS.D ---i-k------k----k-----------..-.-----------h--.r----l----.............    448
CONSENSUS.F e--i-kn-------------r---------..-.-----------n--.r----L----.............    437
CONSENSUS.G -------------?------?---------..-.-------?----H--.-----L-?-?.............    417
CONSENSUS.H ---?---------------?----?----..-.----------???--.-----L----.............    406
CONSENSUS.O ---I-?-------?------Q--------?..NG?-------Y--PGGT.----YV-???.............    410

                        \/ (minor)               \/ (minor)       / p6 terminus (80%)
CONSENSUS.A .EPtAPpAE?f?gmgeeit.s?...?pkqeqkd..??ke??ppl?SlKSLFGNDpLSQ                   483
CONSENSUS.B ?------ees-.rf---t-tPs???-q---pi-.....-lY?--as-rs---n--s-q                   500
CONSENSUS.C ?e---p---S-.rF.--t-.Pa....----p--??--?---?--ts-k-------                     482
CONSENSUS.D .--------eS-.-F-----.Ps....q------??----ly.--as-k-------ls-                 494
CONSENSUS.F .---------s-.-F?----.PS....-------....egly---a----                          474
CONSENSUS.G .--?-----sl.-f----?.?S....----Pr-....--LY--------                           451
CONSENSUS.H .---------S-.-F---M-.P-....----??-....?-..?-------                          436
CONSENSUS.O .?-S---M-......-?VK.?Q....EN-?--G..--?-LY.-FA-------T-Q                    443
```

## HIV-2/SIV Gag Consensus Sequences

```
CONSENSUS.A MGArNSVLRGKKADELEkiRLRPgGKKkYrLKHiVWAANeLDrFGLaESLLESKEGCQkIltVLdPlVPTGSE     73
CONSENSUS.B ----?---S---T-----V--------?-?---??--v---d------?--e-----h----v-a--v-----     67
CONSENSUS.C                                    -------A-------T----------E--------     36
CONSENSUS.D --a-n---S--k-----------------m---v-----------------en--------s--a--------     73
CONSENSUS.E                                    -R-------------N----HR-----E--M-----     36

                                                                          p16 \/ p28
CONSENSUS.A NLkSLfNTvCViWCiHAEEKvKDTEeAk?ivq.RHLvAETgTaEKMPntSRPTAPpSgkggNfPVQq?gG?NY    142
CONSENSUS.B -------t----y-L--ee---------K-A-.s--?v...dt--m-a-s?pt-p---..r-Y----vA-.--    131
CONSENSUS.C -----Y--T-----L---Q---H---RNEV-E.---A---KN-----A-------S--GR--Y----VA-.--    107
CONSENSUS.D -----y---c------------h-----Q---.----v-----e?--k----------r---Y-------.--    142
CONSENSUS.E -------A---VY-L--AV---------KH--QH----GGK-T--L-PQ---------G---Y----I-N.--    108

CONSENSUS.A tHvPLSPrTLNAWVKLvEeKKFGAEVVPGFQALSEGCTPYDINQmLNCVGDHQAAMQIIREIINeeAAdWD?q    214
CONSENSUS.B v-l--------------------ae-----------c-------------e----mqi-------------Q-    204
CONSENSUS.C V-H------------------------------T----------L--------------------------EH    180
CONSENSUS.D v-L-------------i----------p---------t-y----------e--------re----------l-    215
CONSENSUS.E V-S-----------------------------------------------E--------------------V-    181

CONSENSUS.A HPIP.GPLPAGQLRdPRGSDIAGTTSTVeEQIqWMfRpqNPvPVGnIYRRWiQIGLQKCVRmYNPTNiLDikQ    286
CONSENSUS.B --?-.--l----------------------q----y-a-n-------------l----------p-------q    275
CONSENSUS.C --V-.------------------------------Y-A---------------L----------------V--    252
CONSENSUS.D -pQp??-q??----e-s---------tvd------y-Q---I------r----L-----------tn---v--    284
CONSENSUS.E --RG..QP--QG----S----------PA---E--Y-NP--I---D-------L----------------V--    252

CONSENSUS.A GPKEpFQSYVDRFYKSLRAEQTDpAVKNWMTQTLL?QnaNPDCKLVLKGLGmNPTLEEMLTaCQGvGGPgQKA    358
CONSENSUS.B -------s--------------------------I------------------------------I-------    348
CONSENSUS.C --------                                                                    260
CONSENSUS.D ---------------------t--av-----q-l-I--------------g?--------------g-----a    356
CONSENSUS.E ----S---                                                                    260

          p28 \/       p2       \/ p8                                    pol cds ->
CONSENSUS.A RLMAEALKEam?PaPIPFAAAQQ....r?aikcWNCGKEGHSArQcrAPRRQGCWKCGK?GHiManCP?RQAG    423
CONSENSUS.B ----------LT--pi----a-qkAGk-gTVT------?--T----------------kq----SK--E----    420
CONSENSUS.D --------e-lap?pl--a-a-q?Gp?-kp---w--g-e--s-----a---q------kmg-v--K-----a-    425

        p8 \/     p1     \/ p6                                    \/ HIV2s only
CONSENSUS.A FLGlGpwGKkpRNFPvaq?pqGL.................tPTAPP?DPavdLLEkYMQQG?rQRE.Q...R    471
CONSENSUS.B ---------------mt-v---vtPSAPPMnPAeGMTPrGA--S---A----EM-Ks--qm-rqqre????s-    489
CONSENSUS.D -----------r---M---h---?????????.........-p--p-ed------kn--ql-kq---????sr    475

CONSENSUS.A eRPYKEVTEDlLhlEqgetPhRe?tTEDLLHLNSLFGkDQ                                   510
CONSENSUS.B -------.....................................E--                            511
CONSENSUS.D -k----v.....................----l-------?--                               496
```