Annu. Rev. Biochem. 1998. 67:1–25

HIV-1: Fifteen Proteins and an RNA

Alan D. Frankel
Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94143-0448; e-mail: frankel@cgl.ucsf.edu

John A. T. Young
Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115; jatyoung@warren.med.harvard.edu

KEY WORDS: human immunodeficiency virus, retrovirus, AIDS, viral proteins, viral RNA

ABSTRACT

Human immunodeficiency virus type 1 is a complex retrovirus encoding 15 distinct proteins. Substantial progress has been made toward understanding the function of each protein, and three-dimensional structures of many components, including portions of the RNA genome, have been determined. This review describes the function of each component in the context of the viral life cycle: the Gag and Env structural proteins MA (matrix), CA (capsid), NC (nucleocapsid), p6, SU (surface), and TM (transmembrane); the Pol enzymes PR (protease), RT (reverse transcriptase), and IN (integrase); the gene regulatory proteins Tat and Rev; and the accessory proteins Nef, Vif, Vpr, and Vpu. The review highlights recent biochemical and structural studies that help clarify the mechanisms of viral assembly, infection, and replication.

CONTENTS

0066-4154/98/0701-0001$08.00

INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) has been the subject of intense investigation for 15 years, and a great deal has been learned about how the retrovirus infects cells, replicates, and causes disease. In the past few years, substantial progress has been made toward understanding the detailed biochemical function of each viral component, and in many cases structures have been determined. In this review, we attempt to integrate structural and biochemical information into a view of the HIV-1 particle as a whole, emphasizing key interactions among viral and cellular components during the viral replication cycle. Given this broad scope, we can barely scratch the surface of the relevant literature and have thus chosen a limited number of reviews and recent references that should help guide the reader to more detailed aspects of the viral components. We apologize to all investigators in the field for the arbitrary selection of references, and the reader should recognize that many classic papers have not been cited.

HIV-1 GENES AND THE VIRUS LIFE CYCLE

The HIV-1 genome encodes nine open reading frames (Figure 1). Three of these encode the Gag, Pol, and Env polyproteins, which are subsequently proteolyzed into individual proteins common to all retroviruses. The four Gag proteins, MA (matrix), CA (capsid), NC (nucleocapsid), and p6, and the two Env proteins, SU (surface or gp120) and TM (transmembrane or gp41), are structural components that make up the core of the virion and outer membrane envelope. The three Pol proteins, PR (protease), RT (reverse transcriptase), and IN (integrase), provide essential enzymatic functions and are also encapsulated within the particle. HIV-1 encodes six additional proteins, often called accessory proteins, three of which (Vif, Vpr, and Nef) are found in the viral particle. Two other accessory proteins, Tat and Rev, provide essential gene regulatory functions, and the last protein, Vpu, indirectly assists in assembly of the virion. The retroviral genome is encoded by an ~9-kb RNA, and two genomic-length RNA molecules are also packaged in the particle. Thus, in simplistic terms, HIV-1 may be considered

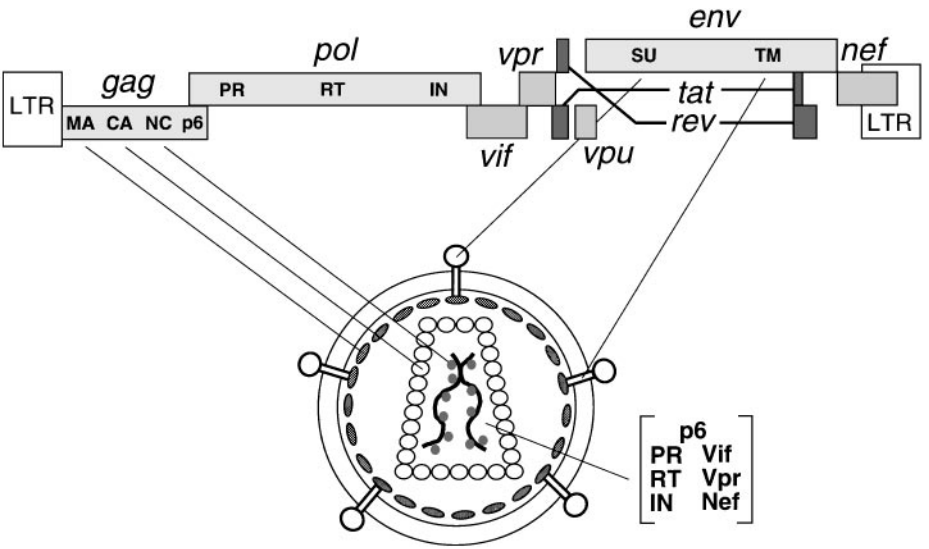

Figure 1 Organization of the HIV-1 genome and virion.

as a molecular entity consisting of 15 proteins and one RNA. We first describe how each component fits into the viral life cycle, and then we review selected structural and biochemical studies to highlight important functional aspects of each protein.

In our view, the HIV-1 replication cycle may be broken into 15 discrete steps, as depicted in Figure 2. We begin the cycle with the viral genome integrated into a host chromosome, and we describe the order of events that lead to expression of the viral gene products, production of virus particles, infection of a new cell, and reintegration of the viral genome. In step 1, viral transcripts are expressed from the promoter located in the 5′ long terminal repeat (LTR), with Tat greatly enhancing the rate of transcription. In step 2, a set of spliced and genomic-length RNAs are transported from the nucleus to the cytoplasm, where they can be translated or packaged. This step is regulated by Rev. In step 3, viral mRNAs are translated in the cytoplasm, and the Gag and Gag-Pol polyproteins become localized to the cell membrane. The Env mRNA is translated at the endoplasmic reticulum (ER). In step 4, the core particle is assembled from the Gag and Gag-Pol polyproteins (later processed to MA, CA, NC, p6, PR, RT, and IN), Vif, Vpr, Nef, and the genomic RNA, and an immature virion begins to bud from the cell surface. To provide SU and TM proteins for the outer membrane coat during budding, the Env polyprotein must

first be released from complexes with CD4 (the cell surface HIV-1 receptor), which is coexpressed with Env in the ER. Vpu assists this process by promoting CD4 degradation, as shown in step 5. Env is then transported to the cell surface (step 6), where again it must be prevented from binding CD4. Nef promotes endocytosis and degradation of surface CD4 (step 7). As the particle buds and is released from the cell surface coated with SU and TM (step 8), the virion undergoes a morphologic change known as maturation (step 9). This step involves proteolytic processing of the Gag and Gag-Pol polyproteins by PR and a less well defined function of Vif. The mature virion is then ready to infect the next cell, which is targeted by interactions between SU and CD4 and CC or CXC chemokine coreceptors (step 10). Following binding, TM undergoes a conformational change that promotes virus-cell membrane fusion, thereby allowing entry of the core into the cell (step 11). The virion core is then uncoated to expose a viral nucleoprotein complex, which contains MA, RT, IN, Vpr, and RNA (step 12). This complex is transported to the nucleus (step 13), where the genomic RNA is reverse transcribed by RT into a partially duplex linear DNA (step 14). IN then catalyzes integration of the viral DNA into a host chromosome and the DNA is repaired (step 15), thereby completing the viral replication cycle. We now describe our current understanding of the biochemistry and structure of each viral component.

VIRAL COMPONENTS

Tat

The HIV-1 promoter is located in the 5′ LTR and contains a number of regulatory elements important for RNA polymerase II transcription. Sites for several cellular transcription factors are located upstream of the start site, including sites for NF-κB, Sp1, and TBP (1). These cellular factors help control the rate of transcription initiation from the integrated provirus (step 1, Figure 2), and their abundance in different cell types or at different times likely determines whether a provirus is quiescent or actively replicating. Despite the importance of these factors, transcription complexes initiated at the HIV-1 promoter are rather inefficient at elongation and require the viral protein Tat to enhance the processivity of transcribing polymerases. Under some conditions, Tat may also enhance the rate of transcription initiation. Tat increases production of viral mRNAs ~100-fold and consequently is essential for viral replication. It is not yet clear which features of the HIV-1 promoter cause initiating transcription complexes to be poorly processive, but experiments in which the TATA box and downstream sequences have been interchanged with different promoters suggest an important role for these regions (1). In the absence of Tat, polymerases generally do not transcribe beyond a few hundred nucleotides, though they do not appear to terminate at specific sites. It is not yet clear how Tat causes transcribing polymerases

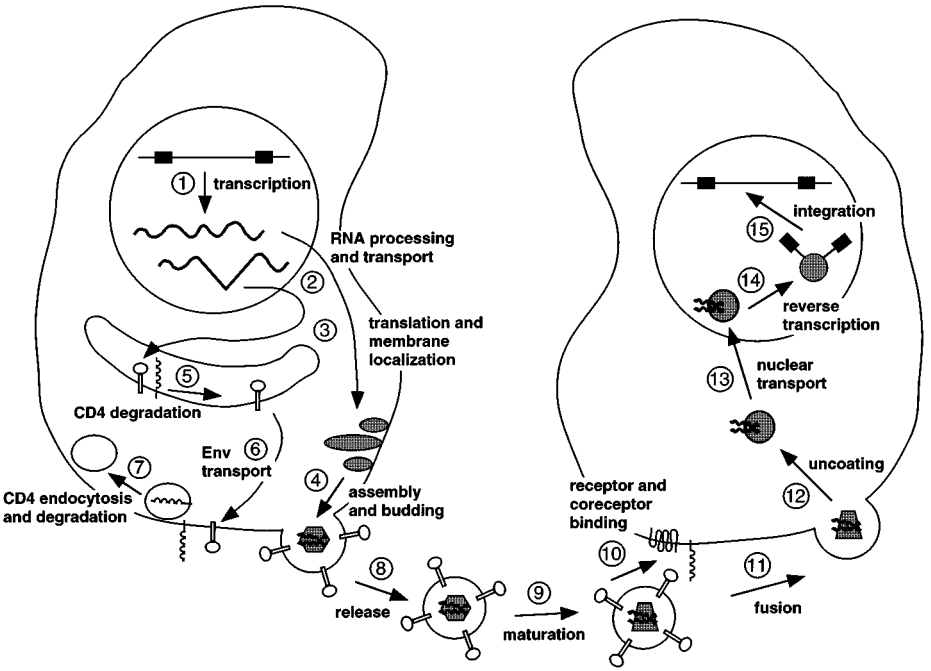

Figure 2 HIV-1 replication cycle. Steps 1–15 are described in the text.

to become sufficiently processive to completely transcribe the ∼9-kb viral genome, but recent experiments suggest that Tat may assemble into transcription complexes and recruit or activate factors that phosphorylate the RNA polymerase II C-terminal domain (CTD), including the general transcription factor TFIIH and other novel kinases (2–6, 6a). These findings support a model in which Tat enhances phosphorylation of the CTD, a process known to occur as RNA polymerase II converts from an initiating to an elongating enzyme.

Unlike typical transcriptional activators, Tat binds not to a DNA site but rather to an RNA hairpin known as TAR (trans-activating response element), located at the 5′ end of the nascent viral transcripts. An arginine-rich domain of Tat helps mediate binding to a three-nucleotide bulge region of TAR, with one arginine residue being primarily responsible for recognition. NMR studies of TAR complexed to arginine (7, 8) show a base-specific contact between the arginine side chain and a guanine in the RNA major groove (Figure 3*A*). The complex is stabilized by additional contacts to the phosphate backbone and by formation of a U-A:U base triple between a bulge nucleotide and a base pair above the bulge (Figure 3*A*). NMR studies of the full-length 86–amino acid Tat protein have suggested that a hydrophobic core region of about 10 amino

acids adopts a defined structure but that the rest of the molecule, including the arginine-rich RNA-binding domain, is relatively disordered (9). It seems likely that Tat requires interactions with cellular proteins in addition to TAR to adopt a stable structure. Aside from proteins of the transcription apparatus, another protein is needed to bind to the loop of the TAR hairpin, apparently helping to stabilize the Tat-TAR interaction (10). Functional data suggest that the loop-binding protein is encoded by human chromosome 12. Several candidates have been identified, but none have yet been definitively shown to be essential for Tat activity.

RNA

The transcript produced from the viral promoter is ∼9 kb long and may be thought of as a large macromolecular component of the virion containing structured subdomains throughout its length. Beginning at the 5′ end, several essential regions have been defined [nucleotide numbers (nts) are approximate and vary among HIV-1 isolates]:

1. The TAR hairpin (nts 1–55) is the Tat-binding site.

2. The primer-binding site (nts 182–199) is important for initiating reverse transcription by annealing to a cellular $tRNA^{Lys}$.

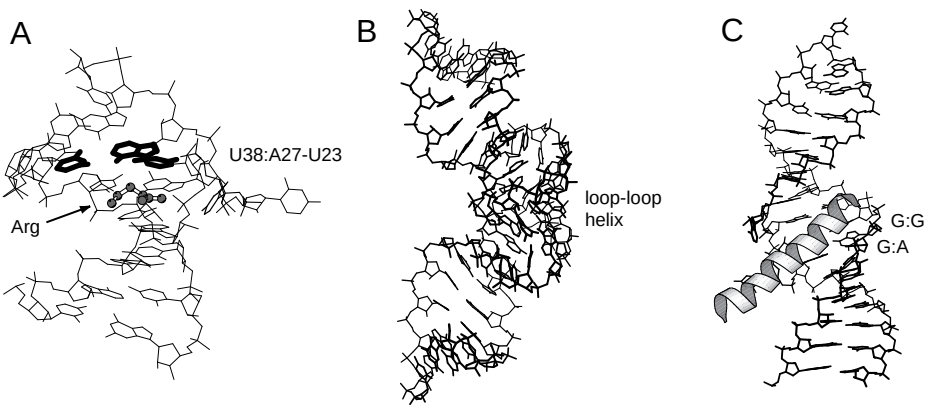

Figure 3 Structures of HIV-1 proteins and RNA. (*A*) TAR-arginine complex (7), (*B*) kissing-loop dimerization hairpins (A Mujeeb, T Parslow, T James, personal communication) (*C*) Rev peptide- RRE IIB complex (22), (*D*) MA trimer (24), (*E*) CA C-terminal dimerization domain (37), (*F*) CA N-terminal domain dimer (39), (*G*) NC (53), (*H*) Nef (73), (*I*) PR-inhibitor complex (133), (*J*) TM_{core} trimer (105), (*K*) RT-nevirapine complex (120), (*L*) IN N-terminal domain dimer (129), (*M*) IN catalytic domain dimer (130), (*N*) IN DNA-binding domain dimer (132). Figures were prepared using MOLSCRIPT (134).

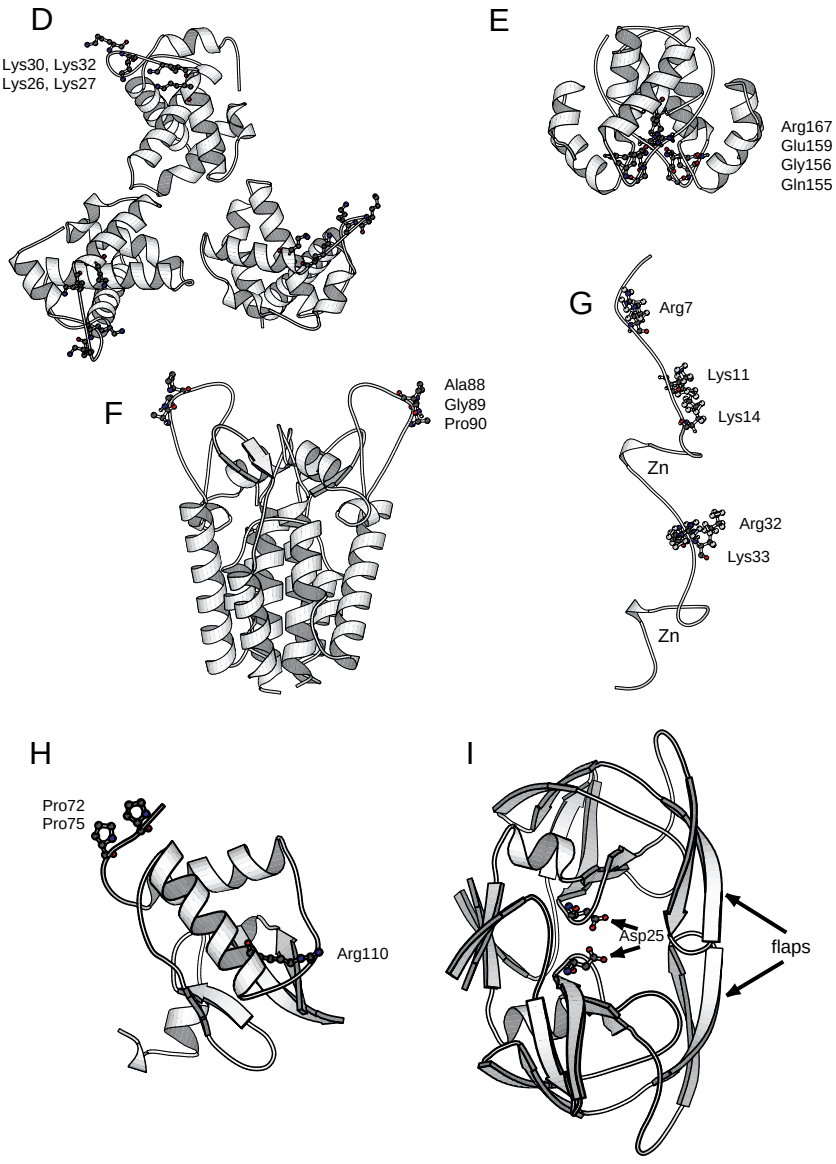

Figure 3 (*Continued*)

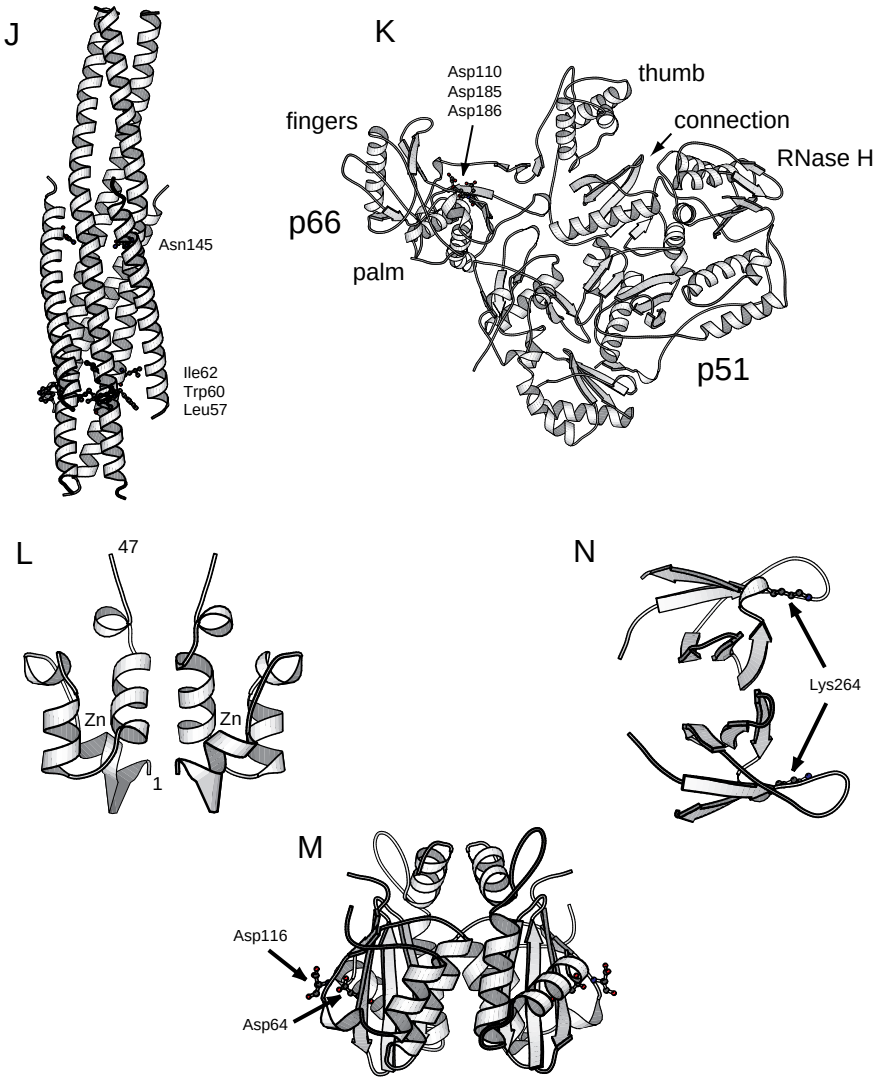

Figure 3 (*Continued*)

3. The packaging signal or ψ (nts 240–350) binds NC and is critical for incorporation of genomic RNA into the virion (11).

4. The dimerization site includes a "kissing loop" hairpin (nts 248–271) that facilitates incorporation of two genomic RNAs into the virion (11–13).

5. The major splice donor site (nt 290) is used to generate all subgenomic spliced mRNAs.

6. The Gag-Pol frameshifting region (nts 1631–1673) comprises a heptanucleotide slippery sequence and RNA hairpin that promote -1 ribosomal frameshifting, thereby translating a fused Gag-Pol polyprotein at a frequency of $\sim$5–10% (14).

7. The Rev response element (RRE) (nts 7362–7596) is the Rev-binding site.

8. Splice acceptor sites are present at several downstream regions of the RNA and allow production of a relatively large number of spliced products (the two major sites are at nts 5358 and 7971).

9. The polyadenylation signal (nts 9205–9210) is used to generate the 3′ end.

The structures of several important RNA elements have been solved by NMR. A kissing-loop complex, which helps mediate dimerization of the genomic RNA, contains two identical RNA hairpins (one from each genomic strand) with complementary loops that form a six–base pair helix. The loop-loop helix lies perpendicular to the helical stems of the hairpins, introducing a slight bend that is stabilized by the stacking of unpaired adenosines (Figure 3*B*) (A Mujeeb, T Parslow, and T James, personal communication). The junctions cause unpairing of the adjacent stems, which may facilitate a transition to a more stable duplex structure. As described in other sections, the structures of TAR-arginine and RRE-Rev peptide complexes have also been determined, and the structure of a ψ-NC complex is expected in the near future.

Rev

When viral mRNAs are first produced, most are doubly spliced and encode the Tat, Rev, and Nef proteins. Later, when other viral components are needed to assemble infectious virions, singly spliced and unspliced transcripts are transported to the cytoplasm (step 3, Figure 2), where they are translated and where genomic RNAs are packaged. Rev is important in this switch because it overcomes the default pathway in which mRNAs are spliced prior to nuclear export and functions by binding to the RRE site located in the *env* coding region. Whether Rev directly enhances the export of unspliced mRNAs or inhibits

splicing has been unclear, but recent studies lend strong support to a role in export (14a). Microinjection experiments in *Xenopus* oocytes have shown that Rev is required to export unspliced RNAs that contain an RRE (15). Rev contains a leucine-rich nuclear export signal (NES) that allows it to shuttle between the nucleus and cytoplasm (16) and that interacts with a nucleoporin-like protein (hRip/Rab) located at the nuclear pore (17, 18, 18a). The interaction with hRip/Rab may be bridged by CRM1, a nuclear export receptor that is important for Rev export (18b). Thus, Rev binding to the RRE is believed to target the attached mRNA to the nuclear export machinery. There is evidence that entry into the splicing pathway may also be important for Rev function because mutating 5′ splice sites on RRE-containing mRNAs eliminates Rev activity but compensatory mutations in U1 snRNA, which binds at 5′ splice sites, can restore activity (19). Furthermore, Rev can directly inhibit splicing by preventing entry of additional snRNPs during the later stages of spliceosome assembly (20). Possible relationships between the splicing and transport pathways and the precise mechanism of Rev function remain to be clarified.

The RRE contains several hairpins and binds several Rev monomers, nucleated by the interaction of a single monomer with a high-affinity site, hairpin IIB (21). Oligomeric binding is important for Rev function, presumably because it increases the concentration of NES sites on a single mRNA. Binding is mediated by an arginine-rich domain that forms an α-helix and specifically recognizes an internal loop in the IIB stem. The structure of a Rev peptide complexed to IIB has been determined by NMR (Figure 3*C*) as well as complexed to an in vitro selected RNA (22, 23). The internal loop contains G:G and G:A base pairs that widen an otherwise narrow major groove. The widened groove allows amino acids on the Rev α-helix to recognize specific features of the site, primarily through hydrogen bonds between three arginines and specific bases and phosphates and between Asn40 and the G:A pair. The structures of the intact 116-residue Rev protein and oligomeric complexes remain to be determined.

MA

MA is the N-terminal component of the Gag polyprotein and is important for targeting Gag and Gag-Pol precursor polyproteins to the plasma membrane prior to viral assembly (step 3, Figure 2). In the mature viral particle, the 132-residue MA protein lines the inner surface of the virion membrane (Figure 1). Two discrete features of MA are involved in membrane targeting: an N-terminal myristate group and basic residues located within the first 50 amino acids. The crystal structure of residues 1 through 104 (24) shows five α-helices capped by a three-stranded mixed β-sheet, with three monomers arranged like a triskelion (Figure 3*D*). The trimeric form is presumed to be biologically relevant because mutation of residues involved in trimerization (residues 42–77) abolishes

viral assembly and because basic residues important for membrane localization (lysines 26, 27, 30, 32) are arranged on the putative membrane-binding surface of the trimer (Figure 3*D*). The MA structure suggests an obvious model for membrane binding that involves the insertion of three myristate groups into the lipid bilayer located directly above the trimer and interactions between basic residues of the membrane-binding surface and phospholipid head groups. However, the N-terminal basic region is not strictly required for the formation of virus particles because noninfectious virus particles that lack MA can be produced if a myristate group is placed directly upstream of CA (25). In addition to targeting Gag and Gag-Pol to the membrane, MA also appears to help incorporate Env glycoproteins with long cytoplasmic tails into viral particles (26, 27). Indeed, the array of threefold symmetric holes located between matrix trimers appears to be large enough to accommodate the long cytoplasmic tails of full-length Env (24, 28).

In addition to its function in viral assembly, MA facilitates infection of nondividing cell types, principally macrophages. Its precise role in viral entry is controversial. Some studies have shown that a subset of phosphorylated MA proteins are associated with viral preintegration complexes and that MA contains a nuclear localization signal (NLS) that interacts with Rch1, a member of the karyopherin-α family, to facilitate rapid nuclear transport (29–32) (step 13, Figure 2). Phosphorylation of Tyr131 was shown to mediate association with IN, thereby linking MA to the preintegration complex (33, 34). Other studies, however, have found no evidence for an MA NLS and suggest that phosphorylation of Tyr131 is not important for macrophage infection (26, 31, 35, 36). Instead, mutation of the putative MA NLS in a macrophage-tropic HIV-1 isolate decreased infectivity in both nondividing and dividing cells and resulted in delayed proteolytic processing of the Gag polyprotein, presumably because the mutations affect association of MA with the membrane (36). Additional studies are needed to clarify the role of MA in infection of nondividing cells.

CA

CA is the second component of the Gag polyprotein and forms the core of the virus particle, with ~2000 molecules per virion (Figure 1). The C-terminal domain (residues 152–231) functions primarily in assembly (step 4, Figure 2) and is important for CA dimerization and Gag oligomerization (37). Although mutations in the N-terminal domain (residues 1–151) do not prevent assembly or budding, the domain is important for infectivity, apparently by participating in viral uncoating (step 12, Figure 2) through its association with a putative cellular chaperone, cyclophilin A (CypA) (38).

Structures of the C-terminal domain, N-terminal domain, and N-terminal domain complexed to CypA have been solved by crystallography and NMR

(37, 39–41). The C-terminal domain is composed of an extended strand followed by four α-helices, with an extensive dimer interface (37) (Figure 3*E*). The major homology region (MHR), a 20–amino acid sequence that is one of the most highly conserved within all retroviral Gag proteins, adopts a compact fold in which the four most conserved residues (Gln155, Gly156, Glu159, and Arg167) form a stabilizing hydrogen-bonding network (Figure 3*E*). Additional hydrophobic residues from the MHR contribute to the hydrophobic core. The MHR is essential for particle assembly and may have a role in incorporation of Gag-Pol precursors through interactions with Gag (42), though not all mutants show this phenotype (43). Biochemical experiments also suggest a possible role for the MHR in membrane affinity, perhaps reflecting exposure of hydrophobic residues (44). The structures of two dimeric forms of the N-terminal domain, one complexed to an antibody fragment and the other complexed to CypA, show the same monomeric CA structure but different subunit interfaces (39, 40). The CA-CA interfaces observed in the CypA complex are blocked in the antibody complex, but given that the C-terminal domain is largely responsible for dimerization (the N-terminal domain is monomeric at mM concentrations), it still remains to be determined whether the observed N-terminal domain interfaces represent true subunit interactions. The most extensive dimer interface from the CypA complex is shown in Figure 3*F*. The CA subunits are also seen to arrange in strips within the crystal, consistent with a plausible packing arrangement in the virion core (39). Residues from an extended region of CA interact with CypA, with Ala88, Gly89, and Pro90 buried in the CypA active site groove (Figure 3*F*). A short spacer peptide located between CA and NC—p2—may also influence CypA incorporation into the virion (45).

NC

NC is the third component of the Gag polyprotein and coats the genomic RNA inside the virion core (Figure 1). The primary function of NC is to bind specifically to the packaging signal and deliver full-length viral RNAs into the assembling virion (step 4, Figure 2). The packaging signal, ψ, is not completely defined but is probably composed of three RNA hairpins located around the major splice donor site (11, 13), the first of which contains the kissing loop involved in RNA dimerization. Studies with a chimeric Gag containing NC from HIV-1 and the remainder of Gag from Moloney murine leukemia virus (Mo-MLV) demonstrate that genomic HIV-1 RNA is preferentially packaged but that additional downstream sequences, which result in packaging of spliced RNAs, may contribute (46). NC is a basic protein that also binds single-stranded nucleic acids nonspecifically, leading to coating of the genomic RNA that presumably protects it from nucleases and compacts it within the core. Nonspecific binding also provides chaperone-like functions that enhance other nucleic

acid–dependent steps in the life cycle; for example, by promoting annealing of the tRNA primer, melting of RNA secondary structures, or DNA strand exchange reactions during reverse transcription (47–49) or by stimulating integration (50).

NC is 55 residues long and contains two zinc finger domains (of the CCHC type) flanked by basic amino acids. The specific NC-ψ interaction requires intact fingers as well as several basic amino acids (51, 52). The residues that make specific versus nonspecific contacts are not yet well defined. The structure of NC has been determined by NMR (53, 54) and shows two well-ordered zinc domains with a relatively flexible linker in the absence of RNA (Figure 3*G*). Basic residues shown to be particularly important for in vitro RNA binding (Arg7, Arg32, and Lys33) and viral replication (Lys11 and Lys14) are indicated, though mutation of Arg32 or Lys33 seems to have little effect on RNA packaging in vivo (52). Disulfide-substituted benzamide compounds specifically remove zinc from the NC domains and inhibit viral replication (55), providing additional evidence of the importance of these structures.

p6

p6 comprises the C-terminal 51 amino acids of Gag and is important for incorporation of Vpr during viral assembly (step 4, Figure 2). Residues 32–39 and three hydrophobic residues within a highly conserved sequence motif (Leu41-X42-Ser43-Leu44-Phe45-Gly46) are important for Vpr binding (56–58). In Vpr, a predicted α-helical structure located near its N-terminus contains amino acids responsible for p6 binding (59). p6 also helps mediate efficient particle release (step 8, Figure 2), and a region of four amino acids (Pro7-Thr8-Ala9-Pro10) has been implicated in this function (60).

Vpu

Newly synthesized Env glycoproteins (gp160), which are later cleaved into SU (gp120) and TM (gp41), are sometimes held in the endoplasmic reticulum through interactions with newly synthesized CD4 molecules. Vpu promotes degradation of CD4 in these complexes, thus allowing Env transport to the cell surface for assembly into viral particles (steps 5 and 6, Figure 2). Vpu is an 81-residue oligomeric integral membrane protein with an N-terminal 24-residue hydrophobic membrane-spanning domain and a C-terminal cytoplasmic tail (59, 61). Amino acids important for receptor binding and degradation have been mapped to the C-terminal region of Vpu and to putative α-helices in the cytoplasmic tail of CD4 (62). Coimmunoprecipitation experiments have shown that Vpu associates with wild-type CD4 or with recombinant proteins containing the CD4 cytoplasmic tail, but it is not yet known if the interaction is direct or indirect (63). These complexes are probably relevant to CD4 degradation

because there is a direct correlation between the extent of Vpu association and their relative levels of degradation (63).

The effect of Vpu on CD4 degradation appears to be regulated by posttranslational modification. Vpu is phosphorylated on Ser52 and Ser56 by a casein kinase-2-related protein, and mutation of these positions decreases the levels of CD4 degradation (59). The mechanism of degradation is not clear but may involve the cytoplasmic proteasome, because Vpu-mediated degradation can be blocked by proteasome inhibitors such as lactacystin (64). Vpu can also downregulate cell surface expression of MHC class I proteins, which may protect infected cells from recognition and killing by cytotoxic T lymphocytes (65).

In addition to its role in CD4 degradation, Vpu can also stimulate virion release, and it has been proposed to be an ion channel (61). In Vpu mutant viruses, significantly increased numbers of particles either remain associated with the cell surface or are localized to intracellular membranes (59, 61). In contrast to Vpu-mediated CD4 degradation, its effect on particle release requires the hydrophobic N-terminal domain and is not influenced by serine phosphorylation (59, 61). The mechanism appears to be relatively nonspecific in that Vpu can also promote the release of heterologous retroviral particles (59, 61).

Nef

Nef is a 206–amino acid, N-terminally myristoylated protein that, like Vpu, reduces the levels of cellular CD4. Nef facilitates the routing of CD4 from the cell surface and golgi apparatus to lysosomes, resulting in receptor degradation and preventing inappropriate interactions with Env, as for Vpu (step 7, Figure 2) (66). A dileucine-based sorting signal located in the cytoplasmic tail of CD4 is essential for Nef-mediated downregulation and is presumed to interact with Nef (66). Nef has been proposed to serve as a direct bridge between CD4 and the cellular endocytic machinery by interacting with β-COP and adaptins, which link proteins in the golgi and plasma membrane to clathrin-coated pits (66). By downregulating CD4, Nef may enhance Env incorporation into virions, promote particle release, and possibly affect $CD4^+$ T-cell signaling pathways (66). As with Vpu, Nef can also downregulate expression of MHC class I molecules, which may help protect infected cells from killing by cytotoxic T cells (67).

Nef mutant viruses also exhibit decreased rates of viral DNA synthesis following infection (68). This defect can be overcome if Nef is supplied in *trans* in virus-producing cells but not in target cells, suggesting possible roles in virus assembly, maturation, or entry. Such roles are consistent with the observation that $\sim$70 Nef molecules are incorporated per virion; these virion-associated proteins are cleaved by PR at residue 57 to generate a soluble C-terminal fragment (68). The mechanism of Nef incorporation has not been defined but is probably relatively nonspecific, because Nef can also be incorporated into Mo-MLV particles (69).

Nef contains a consensus SH3 domain binding sequence (PXXP) that mediates binding to several Src-family proteins (e.g. Src, Lyn, Hck, Lck, Fyn), thereby regulating their tyrosine kinase activities (70–72). These interactions appear to be important for enhancing viral infectivity but not for downregulating CD4 (72). It is not yet clear which SH3-containing proteins are relevant for Nef function. The crystal structure of a Nef-SH3 complex (73) shows that the PXXP motif is in a left-handed polyproline type II helix and interacts directly with the SH3 domain (Figure 3*H*). Two residues that define the motif, Pro72 and Pro75, are important for enhancing viral replication and pack against hydrophobic residues of the SH3 domain (73). The central core of Nef comprises two antiparallel α-helices packed against a layer of four antiparallel β strands (Figure 3*H*) (73, 74). A hydrophobic crevice, which is presumably a ligand-binding site, is located between the two helices and is close to Arg110. Arg110 has been defined as an important residue for association with NAK, a Nef-associated serine/threonine kinase related to a p21 kinase (PAK) (75). PAKs are known to bind the p21 Rho-like GTP-binding proteins Rac-1 and CDC42hs, suggesting possible mechanisms by which Nef can interfere with both endocytosis and T-cell signaling (76). However, mutations that disrupt the Nef-NAK complex do not affect Nef-mediated CD4 downregulation (72). Nef has also been reported to bind other cellular proteins, including p53, MAP kinase, and TEase-II (70, 72), but the significance of these interactions remains to be determined.

PR

As the core virion is assembled to include the Gag and Gag-Pol polyproteins, the Vif, Vpr, and Nef proteins, and the genomic RNA, and as the membrane coat containing SU and TM surrounds the particle, the virus buds from the membrane surface and is released (steps 4 and 8, Figure 2). The immature particles formed are noninfectious. The Gag and Gag-Pol polyproteins must be cleaved by PR, and conformational rearrangements must occur within the particle, to produce mature infectious viruses (step 9, Figure 2). Some of these "maturation" events may occur simultaneously with assembly and budding (77); the precise timing is not clear. PR cleaves at several polyprotein sites to produce the final MA, CA, NC, and p6 proteins from Gag and PR, RT, and IN proteins from Pol. The final stoichiometries are determined largely by the amount of Gag-Pol produced by ribosomal frameshifting and incorporated into the virion ($\sim$5–10% of Gag). Because assembly and maturation must be highly coordinated, factors that influence PR activity can have dramatic effects on virus production. PR functions as a dimer and is part of Pol, so PR activity initially depends on the concentration of Gag-Pol and the rate of autoprocessing, which may be influenced by adjacent p6 sequences (78). Cleavage efficiencies can vary substantially among sites, thereby influencing the order of appearance of

different processed proteins (79). The p2 spacer peptide located between CA and NC may also help control relative cleavage rates and infectivity (80, 81), and processing of NC and p6 may be further influenced by RNA binding to NC (82). Overexpression of PR can lead to aberrant rates of processing and decreased infectivity (83).

PR has been a prime target for drug design, and crystal structures of many PR-inhibitor complexes have been solved, including peptidomimetic and nonpeptide inhibitors (84). The enzyme active site is formed at the dimer interface, with each 99-residue monomer contributing a catalytically essential aspartic acid (Asp25) (Figure 3*I*). The active site resembles that of other aspartyl proteases and contains a conserved triad sequence, Asp-Thr-Gly. The PR dimer contains flexible flaps (Figure 3*I*) that close down on the active site upon substrate binding. Amino acid side chains surrounding the cleavage site bind within hydrophobic pockets of PR, helping to explain some of the rate differences observed between different sites. Several PR inhibitors are in wide clinical use, and mutants resistant to multiple inhibitors have been observed (85, 86). Resistance mutations are located both within the inhibitor binding pocket and at distant sites, and some mutants show increased catalytic activities (87). An alternative approach to inhibitor design involves the use of inactive subunits that act as dominant negative inhibitors (88).

Vif

Vif is a 192-residue protein that is important for the production of highly infectious mature virions. Vif mutant viruses show markedly reduced levels of viral DNA synthesis and produce highly unstable replication intermediates (59, 89), suggesting that Vif functions before or during DNA synthesis. It is intriguing that Vif mutants show defects in infectivity only when produced in certain cell types, designated nonpermissive or semipermissive, but not when produced in permissive cells. It is possible that permissive cells produce a factor or factors that compensate for a lack of Vif or that expression of Vif in permissive cells blocks an inhibitor of viral infectivity (59). Vif activity may be regulated by posttranslational modification because mutation of one of three serine phosphorylation sites (Ser144) causes a defect in viral infectivity (90).

Compared with mature wild-type virions, Vif mutant viruses have similar protein and RNA contents but grossly altered core structures, suggesting that Vif may play a role in viral assembly and/or maturation (59) (steps 4 and 9, Figure 2). Consistent with this role, the infectivity defect can be complemented by supplying Vif in *trans* in virus-producing cells but not in target, nonpermissive cells (59), as also seen with Nef. It has been estimated that 7 to 100 molecules of Vif are packaged into the virion (91–94), suggesting that Vif may function directly within the particle. Incorporation of Vif is probably

nonspecific because there is no apparent requirement for any viral protein or RNA and, like Nef, Vif can be incorporated into Mo-MLV particles (91).

SU

Viral entry is initiated by the binding of the SU glycoprotein, located on the viral membrane surface (Figure 1), to specific cell surface receptors (step 10, Figure 2). The major receptor for HIV-1 is CD4, an immunoglobulin (Ig)-like protein expressed on the surface of a subset of T cells and primary macrophages. The 515-residue SU protein binds CD4 with high affinity (K_d ~4 nM), and amino acids important for binding have been mapped primarily to four separate conserved regions of SU and to the C′-C″ ridge of CD4, which protrudes from the first Ig-like extracellular domain (95). Structural details of the interactions are not yet known.

The SU-CD4 interaction is not sufficient for HIV-1 entry. Instead, a group of chemokine receptors (a family of seven transmembrane G-coupled proteins) that mobilize intracellular calcium and induce leukocyte chemotaxis serve as essential viral coreceptors (96). There are two major classes of HIV-1: those that are macrophage (M)-tropic and non–syncytium inducing (NSI) and those that are T-cell (T)-tropic and syncytium inducing (SI). CXCR4/fusin was the first coreceptor identified; it permits entry of T-tropic but not M-tropic viruses. CCR5 is a major coreceptor for M-tropic but not T-tropic viruses. Other molecules, including CCR3, CCR2b, Bonzo/STRL33, and BOB/GPR15, serve as coreceptors for some HIV-1 isolates (96). The physiological ligands for CXCR4, CCR5, and CCR3 (SDF-1, RANTES/MIP-1α/MIP-1β, and eotaxin, respectively) are each able to inhibit viral entry by competing with the cognate coreceptor (97–100). Some ligand derivatives have been described that block infection without activating chemokine signaling pathways and may represent a novel class of HIV-1 therapeutics (100).

Binding of CD4 to SU appears to cause structural changes in Env that facilitate coreceptor binding and subsequent viral entry (100). The variable V3 loop of SU is an important determinant of viral tropism. It becomes exposed upon CD4 binding and presumably interacts with the cognate coreceptor (100). However, the V3 loop is probably not the sole determinant of coreceptor specificity, because HIV-1 isolates that use the same coreceptor can have highly variable V3 sequences (99, 101, 102). Determinants for virus specificity are located in each of the extracellular regions of the coreceptors, and the signaling functions of these receptors apparently are not important for viral infection (100).

TM

The primary function of TM, a 345–amino acid protein located in the viral membrane (Figure 1), is to mediate fusion between the viral and cellular membranes

following receptor binding (step 11, Figure 2). An N-terminal hydrophobic glycine-rich "fusion" peptide has been predicted to initiate fusion, and a transmembrane region is important both for fusion and for anchoring Env in the viral membrane (103). Two crystal structures of the core region (TM_{core}) have been reported (104, 105). In the larger of the two structures (shown in Figure 3*J*), residues 30–79 and 113–154 of TM were fused to a 31-residue trimeric coiled-coil from GCN4 in place of the N-terminal fusion peptide (105). TM_{core} lacks residues 80–112 of TM.

TM_{core} forms a trimer containing a central parallel α-helical coiled-coil (residues 1–77) and an outer antiparallel α-helical layer (residues 117–154) (105). The structure of TM_{core} probably does not represent the native TM structure but rather a structure formed during the fusion reaction, as suggested by the following. First, mutations at the interface between the outer and central helical layers (including Ile62, Figure 3*J*) specifically block membrane fusion. Second, TM_{core} is extremely thermostable, a feature predicted of the fusion-active protein and not the native protein (104, 105). Third, the structures of TM_{core} and a low pH fusion-active form of the influenza virus HA_2 protein are strikingly similar (104, 105). Fourth, the estimated distance between the C-terminus of TM_{core} and the viral lipid bilayer cannot be spanned by the 18 C-terminal extracellular residues missing from the structure. However, the distance is consistent with a conformation in which the fusion peptides and transmembrane regions are located at the same end of the central rod structure when viral and cell membranes are brought together (Figure 3*J*) (105). The structure helps to explain how two peptides known to inhibit fusion may act. A peptide from the C-terminus may bind to the central trimer, disrupting the structure of the N-terminal region, whereas a peptide from the N-terminus may either compete with folding of the central trimer or bind to the C-terminal region and prevent association with the central core (104, 105). Emerging rules for the design of coiled-coils may aid in the development of new fusion inhibitors.

Vpr

Following fusion and entry, the virus is "uncoated" in the cytoplasm (a poorly defined process; see step 12, Figure 2) and nucleoprotein complexes are rapidly transported to the host cell nucleus, mediated by the 96–amino acid Vpr protein (59) (step 13, Figure 2). The components of the transported complexes are not completely defined but certainly include RT, IN, and MA (106) and probably the genomic RNA and partially reverse-transcribed DNA. Vpr is especially important for nuclear localization in nondividing cells, such as macrophages, because it contains an NLS that directs transport even in the absence of mitotic nuclear envelope breakdown (59). Vpr does not contain a canonical karyophilic NLS but instead contains two important putative N-terminal amphipathic α-helices (107). This unusual NLS localizes Vpr to the nuclear pores rather than to

the interior of the nucleus and does not use an importin-dependent pathway (32, 107). Vpr is incorporated into viral particles through an interaction with p6 and may later become associated with the nucleoprotein complexes through an interaction with the C-terminal region of MA (108).

In addition to its nuclear uptake function, Vpr can also induce G2 cell cycle arrest prior to nuclear envelope breakdown and chromosome condensation, and sustained expression can reportedly kill T cells by apoptosis (107). Vpr acts before dephosphorylation of the $p34^{cdc2}$ cyclin-dependent kinase by CDC25, which is required to initiate mitosis (107). Although G2 arrest occurs with Vpr proteins from different primate lentiviruses, it is not known how the activity contributes to viral replication (107). Amino acids important for G2 arrest are located in the C-terminal region of Vpr, and cellular proteins have been identified that bind Vpr, including the 65-kDa regulatory subunit of protein phosphatase 2A (PP2A), a serine/threonine phosphatase that regulates the transition from G2 to mitosis (107). In addition to roles in nuclear localization and cell cycle arrest, Vpr can also influence mutation rates during viral DNA synthesis (109) and has been proposed to form an ion channel (61).

RT

Before the viral genome can be integrated into the host chromosome, it must first be reverse transcribed into duplex DNA (step 14, Figure 2). RT catalyzes both RNA-dependent and DNA-dependent DNA polymerization reactions and contains an RNase H domain that cleaves the RNA portion of RNA-DNA hybrids generated during the reaction. Reverse transcription initiates from the 3′ end of a $tRNA_3^{Lys}$ primer annealed to the primer binding site near the 5′ end of the genomic RNA. RT can use other tRNAs if complementary binding sites are provided, but reverse transcription is most efficient with $tRNA_3^{Lys}$ (110). $tRNA_3^{Lys}$ is incorporated into virions during assembly and is often extended by several nucleotides inside the particle (110, 111). The remainder of the reaction probably occurs after uncoating in the cytoplasm. The kinetic properties of RT during the initiation and elongation phases of the reaction are quite different, becoming highly processive during elongation, and posttranscriptional modifications of $tRNA_3^{Lys}$ enhance the formation of initiation complexes (112). These kinetic transitions are reminiscent of those observed in transcription complexes, with $tRNA_3^{Lys}$ performing a role analogous to σ factor (112). Following tRNA-primed initiation, reverse transcription involves two DNA strand transfer reactions that are catalyzed by RT and are important for priming the synthesis of both minus and plus strands (see references 113 and 114 for details of the mechanism).

RT has also been a major target for drug design, and crystal structures of unliganded RT, an RT-DNA complex, and RT-inhibitor complexes have been solved (115–120). RT is a heterodimer containing a 560-residue subunit (p66)

and a 440-residue subunit (p51) both derived from the Pol polyprotein. Each subunit contains a polymerase domain composed of four subdomains called fingers, palm, thumb, and connection, and p66 contains an additional RNase H domain (Figure 3*K*). Even though their amino acid sequences are identical, the polymerase subdomains are arranged differently in the two subunits, with p66 forming a large active-site cleft and p51 forming an inactive closed structure (121). The p66 polymerase active site contains a catalytic triad (Asp110, Asp185, and Asp186) conserved in many polymerases (Figure 3*K*), and the 3′OH group of the primer strand in an RT-DNA complex is positioned close to the active site for nucleophilic attack on the incoming nucleoside triphosphate (118). The DNA in this complex has primer and template strands clamped between the palm, thumb, and fingers subdomains of p66 and is bent. Portions of the DNA near the active site adopt an A-form geometry expected of RNA-DNA hybrids or RNA duplexes bound during reverse transcription. The positioning of the DNA near the RNase H domain does not explain how RNA-DNA hybrids are cleaved. Cross-linking experiments suggest that the 5′ part of the $tRNA_3^{Lys}$ primer may contact both the RT dimer interface and a C-terminal region of p66 during initiation (122).

Two classes of RT inhibitors are in clinical use: nucleoside analogs such as AZT and ddI that are presumed to bind to the polymerase active site and non-nucleoside inhibitors such as nevirapine. The structures of several non-nucleoside inhibitor-RT complexes show a common hydrophobic binding site near to, but distinct from, the polymerase active site that rearranges to fit the particular drug and lock RT into an inactive conformation (119, 120). Mutations that confer resistance to nucleoside or non-nucleoside inhibitors map to different parts of RT, including regions in and around the active site and DNA-binding cleft, suggesting that some mutations directly alter the drug-binding site while others have more indirect effects (120, 123). Structures of the unliganded RT show substantial variability in the positioning of the p66 thumb subdomain (115–117), indicating that large-scale conformational rearrangements occur upon nucleic acid or drug binding. Such conformational changes may be important during reverse transcription; for example, to allow translocation of RT along the nucleic acid or to correctly position the RNase H and polymerase active sites.

IN

Following reverse transcription, IN catalyzes a series of reactions to integrate the viral genome into a host chromosome (step 15, Figure 2). In the first step, IN removes two 3′ nucleotides from each strand of the linear viral DNA, leaving overhanging CA_{OH} ends (113). The CA dinucleotide is found at the ends of many retrotransposons, and mutation of these nucleotides substantially reduces the efficiency of 3′-end processing. In the second step, the processed 3′ ends

are covalently joined to the 5′ ends of the target DNA. In the third step, which probably involves additional cellular enzymes, unpaired nucleotides at the viral 5′ ends are removed and the ends are joined to the target site 3′ ends, generating an integrated provirus flanked by five base-pair direct repeats of the target site DNA. The viral substrate used for integration is a linear DNA molecule containing a complete minus strand and a discontinuous plus strand, which is presumably completed by cellular enzymes following integration (124). In vitro systems using viral preintegration complexes, or purified IN with short oligonucleotides, have helped define important nucleotides near the viral DNA ends, important features of the target DNA, and critical amino acids of IN (113). The enzymatic mechanism involves two sequential transesterification reactions and requires no exogenous energy source, but an appropriate metal cofactor (either Mn^{2+} or Mg^{2+}) is needed (113).

Integration can occur at many target sites within the genome. In vitro studies have indicated a preference for kinked or distorted DNA, such as that found in nucleosomes, but it is not yet clear how these target sites relate to those used in vivo (125). It has been suggested that interactions with other DNA-binding proteins might target IN to specific sites, and a yeast two-hybrid screen has identified a human Snf5-related protein, Ini1, as a possible partner (126). Ini1 influences the efficiency of integration, but its effect on target site selection in vivo is unknown. Another cellular factor, HMG I(Y), is associated with preintegration complexes and plays a crucial role in integration (127).

IN is active as an oligomer, probably a tetramer (128), and the 288-residue monomer can be divided into three domains whose structures have been determined. The N-terminal domain (residues 1–55) contains a zinc-binding site (coordinated by two histidines and two cysteines) and forms a dimer with a largely hydrophobic interface, as shown by NMR (Figure 3*L*) (129). Each monomer contains a helix-turn-helix structure very similar to those found in DNA-binding proteins and exists in two closely related conformational states. The catalytic domain (residues 50–212) contains a D,D(35),E motif. This motif is conserved among integrases, is crucial for the processing and joining reactions, and is proposed to bind the active site metal ion (128). The isolated catalytic domain cannot perform processing or joining reactions but can perform an apparent reverse reaction, termed disintegration, indicating that it contains the catalytic site for polynucleotidyl transfer. The crystal structure of the catalytic domain shows a dimeric structure, with each monomer containing a five-stranded β-sheet and six α-helices similar to other polynucleotidyl transfer enzymes (Figure 3*M*) (130). Asp64 and Asp116 of the D,D(35),E motif are clearly seen in the structure, but Glu152 is located on a disordered loop. The two active sites in the dimer are too far apart to permit five base pair staggered cleavage of the target DNA, suggesting either that a very large conformational change occurs during catalysis or, more likely, that IN functions as a tetramer

or other oligomeric form during some steps of the reaction (128, 129). The C-terminal domain (residues 220–270) has nonspecific DNA-binding activity and forms a dimer of parallel monomers, as shown by NMR (Figure 3*N*). The structure of each monomer consists of a five-stranded β barrel strikingly similar to a SH3 domain, with a saddle-shaped groove that might accommodate double-stranded DNA and containing Lys264, an important DNA-binding residue (Figure 3*N*) (131, 132). The relative orientations of the three IN domains remain to be established (128, 129).

CONCLUDING REMARKS

The past decade has seen remarkable progress in elucidating the structures and functions of many of the HIV-1 proteins. Achieving such a sophisticated level of understanding of this complex retrovirus in such a short period of time is a true testament to the collaborative efforts of the scientific community. With this basic framework in hand, it should now be possible to probe each viral component in greater detail and to focus attention on the remaining gaps in our knowledge of HIV-1 biology, including issues pertaining to virus-host interactions and pathogenesis. At the molecular level, the structures of several viral proteins and RNAs and the domain arrangements in proteins such as CA and IN remain to be solved, and interaction surfaces between viral factors and other viral or cellular partners remain to be mapped. Many fundamental questions are unanswered: How do the Gag and Gag-Pol proteins, RNA, SU, and TM interact to form the virus particle? How do Vif, Nef, cyclophilin A, Vpr, and the Gag proteins contribute to viral uncoating, nuclear transport, and other early steps of replication? How are the activities of the viral enzymes PR, RT, and IN regulated at the appropriate steps in the replication cycle? What cellular factors are required for the functions of Tat, Rev, and other viral proteins? How do viral factors take advantage of existing cellular mechanisms, including transcription, RNA processing, protein synthesis and degradation, protein and RNA transport, and membrane trafficking? It is hoped that answers to these and other questions will lead to the discovery of new classes of effective HIV-1 therapeutics.

ACKNOWLEDGMENTS

We thank our many colleagues who have contributed structure coordinates to the Brookhaven Protein Database, and we thank Marius Clore, Andrea Dessen, Theresa Gamble, Angela Gronenborn, Tom James, Peter Kim, John Kuriyan, Anwer Mujeeb, Tris Parslow, Michael Summers, Wes Sundquist, and Don Wiley for contributing coordinates of recent structures. We also thank Ron Fouchier and Michael Malim for sharing unpublished results and our many colleagues who have made substantial intellectual contributions to our thoughts

on HIV, particularly members of our laboratories and Mark Feinberg and Raul Andino. Work in our laboratories has been supported by NIH grants AI29135, GM47478, and GM39589 (ADF) and CA62000 and CA70810 (JATY).

Literature Cited

1. Jones KA, Peterlin BM. 1994. *Annu. Rev. Biochem.* 63:717–43
2. Zhou QA, Sharp PA. 1996. *Science* 274:605–10
3. Parada CA, Roeder RG. 1996. *Nature* 384:375–78
4. Yang XZ, Herrmann CH, Rice AP. 1996. *J. Virol.* 70:4576–84
5. Garcia-Martinez LF, Mavankal G, Neveu JM, Lane WS, Ivanov D, Gaynor RB. 1997. *EMBO J.* 16:2836–50
6. Cujec TP, Cho H, Maldonado E, Meyer J, Reinberg D, Peterlin BM. 1997. *Mol. Cell. Biol.* 17:1817–23

6a. Jones KA. 1997. *Genes Dev.* 11:2593–99

7. Puglisi JD, Tan RY, Calnan BJ, Frankel AD, Williamson JR. 1992. *Science* 257: 76–80
8. Aboul-ela F, Karn J, Varani G. 1995. *J. Mol. Biol.* 253:313–32
9. Bayer P, Kraft M, Ejchart A, Westendorp M, Frank R, Rosch P. 1995. *J. Mol. Biol.* 247:529–35
10. Alonso A, Cujec TP, Peterlin BM. 1994. *J. Virol.* 68:6505–13
11. Clever JL, Parslow TG. 1997. *J. Virol.* 71:3407–14
12. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. 1996. *Proc. Natl. Acad. Sci. USA* 93:5572–77
13. Laughrea M, Jette L, Mak J, Kleiman L, Liang C, Wainberg MA. 1997. *J. Virol.* 71:3397–406
14. Cassan M, Delaunay N, Vaquero C, Rousset JP. 1994. *J. Virol.* 68:1501–8

14a. Hope TJ. 1997. *Chem. Biol.* 4:335–44

15. Fischer U, Huber J, Boelens WC, Mattaj IW, Luhrmann R. 1995. *Cell* 82:475–83
16. Meyer BE, Malim MH. 1994. *Genes Dev.* 8:1538–47
17. Fritz CC, Zapp ML, Green MR. 1995. *Nature* 376:530–33
18. Bogerd HP, Fridell RA, Madore S, Cullen BR. 1995. *Cell* 82:485–94

18a. Stutz F, Neville M, Rosbash M. 1995. *Cell* 82:495–506

18b. Ullman KS, Powers MA, Forbes DJ. 1997. *Cell* 90:967–70

19. Lu XB, Heimer J, Rekosh D, Hammarskjold ML. 1990. *Proc. Natl. Acad. Sci. USA* 87:7598–602
20. Kjems J, Sharp PA. 1993. *J. Virol.* 67: 4769–76
21. Zemmel RW, Kelley AC, Karn J, Butler PJ. 1996. *J. Mol. Biol.* 258:763–77
22. Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, et al. 1996. *Science* 273:1547–51
23. Ye XM, Gorin A, Ellington AD, Patel DJ. 1996. *Nat. Struct. Biol.* 3:1026–33
24. Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI. 1996. *Proc. Natl. Acad. Sci. USA* 93:3099–104
25. Lee PP, Linial ML. 1994. *J. Virol.* 68:6644–54
26. Freed EO, Englund G, Martin MA. 1995. *J. Virol.* 69:3949–54
27. Mammano F, Kondo E, Sodroski J, Bukovsky A, Gottlinger HG. 1995. *J. Virol.* 69:3824–30
28. Massiah MA, Worthylake D, Christensen AM, Sundquist WI, Hill CP, Summers MF. 1996. *Protein Sci.* 5:2391–98
29. Gallay P, Swingler S, Song JP, Bushman F, Trono D. 1995. *Cell* 83:569–76
30. Bukrinsky MI, Haggerty S, Dempsey MP, Sharova N, Adzhubel A, et al. 1993. *Nature* 365:666–69
31. Bukrinskaya AG, Ghorpade A, Heinzinger NK, Smithgall TE, Lewis RE, Steverson M. 1996. *Proc. Natl. Acad. Sci. USA* 93:367–71
32. Gallay P, Stitt V, Mundy C, Oettinger M, Trono D. 1996. *J. Virol.* 70:1027–32
33. Gallay P, Swingler S, Aiken C, Trono D. 1995. *Cell* 80:379–88
34. Trono D, Gallay P. 1997. *Cell* 88:173–74
35. Freed EO, Englund G, Maldarelli F, Martin MA. 1997. *Cell* 88:171–73
36. Fouchier RAM, Meyer BE, Simon JHM, Fischer U, Malim MH. 1997. *EMBO J.* 16:4531–39

37. Gamble TR, Yoo S, Vajdos FF, Von Schwedler UK, Worthylake DK, et al. 1997. *Science.* In press
38. Luban J. 1996. *Cell* 87:1157–59
39. Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, et al. 1996. *Cell* 87:1285–94
40. Momany C, Kovari LC, Prongay AJ, Keller W, Gitti RK, et al. 1996. *Nat. Struct. Biol.* 3:763–70
41. Gitti RK, Lee BM, Walker J, Summers MF, Yoo S, Sundquist WI. 1996. *Science* 273:231–35
42. Srinivasakumar N, Hammarskjold ML, Rekosh D. 1995. *J. Virol.* 69:6106–14
43. Mammano F, Ohagen A, Hoglund S, Gottlinger HG. 1994. *J. Virol.* 68:4927–36
44. Ebbets-Reed D, Scarlata S, Carter CA. 1996. *Biochemistry* 35:14268–75
45. Dorfman T, Gottlinger HG. 1996. *J. Virol.* 70:5751–57
46. Berkowitz RD, Ohagen A, Hoglund S, Goff SP. 1995. *J. Virol.* 69:6445–56
47. Huang Y, Khorchid A, Wang J, Parniak MA, Darlix JL, et al. 1997. *J. Virol.* 71:4378–84
48. Guo JH, Henderson LE, Bess J, Kane B, Levin JG. 1997. *J. Virol.* 71:5178–88
49. Cameron CE, Ghosh M, Le Grice SF, Benkovic SJ. 1997. *Proc. Natl. Acad. Sci. USA* 94:6700–5
50. Carteau S, Batson SC, Poljak L, Mouscadet J-F, deRocquigny H, et al. 1997. *J. Virol.* 71:6225–29
51. Schmalzbauer E, Strack B, Dannull J, Guehmann S, Moelling K. 1996. *J. Virol.* 70:771–77
52. Poon DT, Wu J, Aldovini A. 1996. *J. Virol.* 70:6607–16
53. Summers MF, Henderson LE, Chance MR, Bess JW Jr, South TL, et al. 1992. *Protein Sci.* 1:563–74
54. Morellet N, Jullian N, De Rocquigny H, Maigret B, Darlix JL, Roques BP. 1992. *EMBO J.* 11:3059–65
55. Rice WG, Supko JG, Malspeis L, Buckheit RW Jr, Clanton D, et al. 1995. *Science* 270:1194–97
56. Kondo E, Gottlinger HG. 1996. *J. Virol.* 70:159–64
57. Lu YL, Bennett RP, Wills JW, Gorelick R, Ratner L. 1995. *J. Virol.* 69:6873–79
58. Checroune F, Yao XJ, Gottlinger HG, Bergeron D, Cohen EA. 1995. *J. AIDS Hum. Retrovirol.* 10:1–7
59. Cohen EA, Subbramanian RA, Gottlinger HG. 1996. *Curr. Top. Microbiol. Immunol.* 214:219–35
60. Huang M, Orenstein JM, Martin MA, Freed EO. 1995. *J. Virol.* 69:6810–18
61. Lamb RA, Pinto LH. 1997. *Virology* 229:1–11
62. Tiganos E, Yao XJ, Friborg J, Daniel N, Cohen EA. 1997. *J. Virol.* 71:4452–60
63. Bour S, Schubert U, Strebel K. 1995. *J. Virol.* 69:1510–20
64. Fujita K, Omura S, Silver J. 1997. *J. Gen. Virol.* 78:619–25
65. Kerkau T, Bacik I, Bennink JR, Yewdell JW, Hunig T, et al. 1997. *J. Exp. Med.* 185:1295–305
66. Mangasarian A, Trono D. 1997. *Res. Virol.* 148:30–33
67. Le Gall S, Heard JM, Schwartz O. 1997. *Res. Virol.* 148:43–47
68. Guatelli JC. 1997. *Res. Virol.* 148:34–37
69. Bukovsky AA, Dorfman T, Weimann A, Gottlinger HG. 1997. *J. Virol.* 71:1013–18
70. Greenway A, McPhee D. 1997. *Res. Virol.* 148:58–64
71. Moarefi I, LaFevre-Bernt M, Sicheri F, Huse M, Lee CH, et al. 1997. *Nature* 385:650–53
72. Benichou S, Liu LX, Erdtmann L, Selig L, Benarous R. 1997. *Res. Virol.* 148:71–73
73. Lee CH, Saksela K, Mirza UA, Chait BT, Kuriyan J. 1996. *Cell* 85:931–42
74. Grzesiek S, Bax A, Hu J-S, Kaufman J, Palmer I, et al. 1997. *Protein Sci.* 6:1248–63
75. Sawai ET, Cheng-Mayer C, Luciw PA. 1997. *Res. Virol.* 148:47–52
76. Cullen BR. 1996. *Curr. Biol.* 6:1557–59
77. Kaplan AH, Manchester M, Swanstrom R. 1994. *J. Virol.* 68:6782–86
78. Zybarth G, Carter C. 1995. *J. Virol.* 69:3878–84
79. Dunn BM, Gustchina A, Wlodawer A, Kay J. 1994. *Methods Enzymol.* 241:254–78
80. Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, et al. 1994. *J. Virol.* 68:8017–27
81. Krausslich HG, Facke M, Heuser AM, Konvalinka J, Zentgraf H. 1995. *J. Virol.* 69:3407–19
82. Sheng N, Pettit SC, Tritch RJ, Ozturk DH, Rayner MM, et al. 1997. *J. Virol.* 71:5723–32
83. Luukkonen BG, Fenyo EM, Schwartz S. 1995. *Virology* 206:854–65
84. Wlodawer A, Erickson JW. 1993. *Annu. Rev. Biochem.* 62:543–85
85. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, et al. 1995. *Nature* 374:569–71
86. Ridky T, Leis J. 1995. *J. Biol. Chem.* 270:29621–23

87. Schock HB, Garsky VM, Kuo LC. 1996. *J. Biol. Chem.* 271:31957–63
88. McPhee F, Good AC, Kuntz ID, Craik CS. 1996. *Proc. Natl. Acad. Sci.* USA 93:11466–81
89. Simon JH, Malim MH. 1996. *J. Virol.* 70:5297–305
90. Yang X, Goncalves J, Gabuzda D. 1996. *J. Biol. Chem.* 271:10121–29
91. Camaur D, Trono D. 1996. *J. Virol.* 70:6106–11
92. Fouchier RA, Simon JH, Jaffe AB, Malim MH. 1996. *J. Virol.* 70:8263–69
93. Karczewski MK, Strebel K. 1996. *J. Virol.* 70:494–507
94. Liu HM, Wu XY, Newman M, Shaw GM, Hahn BH, Kappes JC. 1995. *J. Virol.* 69:7630–38
95. Luciw PA. 1996. In *Fundamental Virology,* ed. BM Fields, DM Knipe, PM Howley, pp. 845–916. Philadelphia: Lippincott-Raven. 3rd ed.
96. Clapham PR, Weiss RA. 1997. *Nature* 388:230–31
97. Bleul CC, Farzan M, Choe H, Parolin C, Clark-Lewis I, et al. 1996. *Nature* 382:829–33
98. Oberlin E, Amara A, Bachelerie F, Bessia C, Virelizier JL, et al. 1996. *Nature* 382:833–35
99. Choe H, Farzan M, Sun Y, Sullivan N, Rollins B, et al. 1996. *Cell* 85:1135–48
100. Clapham PR. 1997. *Trends Cell Biol.* 7:264–68
101. Cocchi F, DeVico AL, Garzino-Demo A, Cara A, Gallo RC, Lusso P. 1996. *Nat. Med.* 2:1244–47
102. Oravecz T, Pall M, Norcross MA. 1996. *J. Immunol.* 157:1329–32
103. Hernandez LD, Hoffman LR, Wolfsberg TG, White JM. 1996. *Annu. Rev. Cell Dev. Biol.* 12:627–61
104. Chan DC, Fass D, Berger JM, Kim PS. 1997. *Cell* 89:263–73
105. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. 1997. *Nature* 387:426–30
106. Miller MD, Farnet CM, Bushman FD. 1997. *J. Virol.* 71:5382–90
107. Emerman M. 1996. *Curr. Biol.* 6:1096–1103
108. Sato A, Yoshimoto J, Isaka Y, Miki S, Suyama A, et al. 1996. *Virology* 220:208–12
109. Mansky LM. 1996. *Virology* 222:391–400
110. Oude Essink BB, Das AT, Berkhout B. 1996. *J. Mol. Biol.* 264:243–54
111. Huang Y, Wang J, Shalom A, Li Z, Khorchid A, et al. 1997. *J. Virol.* 71:726–28
112. Lanchy JM, Ehresmann C, Le Grice SF, Ehresmann B, Marquet R. 1996. *EMBO J.* 15:7178–87
113. Katz RA, Skalka AM. 1994. *Annu. Rev. Biochem.* 63:133–73
114. Peliska JA, Benkovic SJ. 1992. *Science* 258:1112–18
115. Rodgers DW, Gamblin SJ, Harris BA, Ray S, Culp JS, et al. 1995. *Proc. Natl. Acad. Sci. USA* 92:1222–26
116. Esnouf R, Ren J, Ross C, Jones Y, Stammers D, Stuart D. 1995. *Nat. Struct. Biol.* 2:303–8
117. Hsiou Y, Ding J, Das K, Clark AD Jr, Hughes SH, Arnold E. 1996. *Structure* 4:853–60
118. Jacobo-Molina A, Ding JP, Nanni RG, Clark AD Jr, Lu XD, et al. 1993. *Proc. Natl. Acad. Sci. USA* 90:6320–24
119. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA. 1992. *Science* 256:1783–90
120. Ren JS, Esnouf R, Garman E, Somers D, Ross C, et al. 1995. *Nat. Struct. Biol.* 2:293–302
121. Wang J, Smerdon SJ, Jager J, Kohlstaedt LA, Rice PA, et al. 1994. *Proc. Natl. Acad. Sci. USA* 91:7242–46
122. Mishima Y, Steitz JA. 1995. *EMBO J.* 14:2679–87
123. Tantillo C, Ding J, Jacobo-Molina A, Nanni RG, Boyer PL, et al. 1994. *J. Mol. Biol.* 243:369–87
124. Miller MD, Wang B, Bushman FD. 1995. *J. Virol.* 69:3938–44
125. Miller MD, Bor YC, Bushman F. 1995. *Curr. Biol.* 5:1047–56
126. Miller MD, Bushman FD. 1995. *Curr. Biol.* 5:368–70
127. Farnet CM, Bushman FD. 1997. *Cell* 88:483–92
128. Rice P, Craigie R, Davies DR. 1996. *Curr. Opin. Struct. Biol.* 6:76–83
129. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM. 1997. *Nat. Struct. Biol.* 4:567–77
130. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. 1994. *Science* 266:1981–86
131. Eijkelenboom AP, Lutzke RA, Boelens R, Plasterk RH, Kaptein R, Hard K. 1995. *Nat. Struct. Biol.* 2:807–10
132. Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, et al. 1995. *Biochemistry* 34:9826–33
133. Fitzgerald PMD, McKeever BM, VanMiddlesworth JF, Springer JP, Heimbach JC, et al. 1990. *J. Biol. Chem.* 264:14209–19
134. Kraulis P. 1991. *J. Appl. Crystallogr.* 24:924–50