Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimica et Biophysica Acta 1614 (2003) 62–72
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.bba-direct.com

Review

Role of HIV-1 Gag domains in viral assembly

Suzanne Scarlata*, Carol Carter

Department of Physiology and Biophysics, State University of New York at Stony Brook, Stony Brook, NY 11794-8661, USA
Department of Microbiology, State University of New York at Stony Brook, Stony Brook, NY 11794-8661, USA

Received 28 April 2003; accepted 15 May 2003

Abstract

After entry of the human immunodeficiency virus type 1 (HIV-1) into T cells and the subsequent synthesis of viral products, viral proteins and RNA must somehow find each other in the host cells and assemble on the plasma membrane to form the budding viral particle. In this general review of HIV-1 assembly, we present a brief overview of the HIV life cycle and then discuss assembly of the HIV Gag polyprotein on RNA and membrane substrates from a biochemical perspective. The role of the domains of Gag in targeting to the plasma membrane and the role of the cellular host protein cyclophilin are also reviewed.

Keywords: HIV-1; Gag; Viral assembly

1. Introduction

The human immunodeficiency virus (HIV) is the causative agent of acquired immune deficiency syndrome (AIDS). There are two subtypes of HIV, HIV-1 and HIV-2.

Currently, more than 40 million people worldwide are infected with HIV-1. HIV is a member of the retrovirus family (for general background see Refs. [1,2]). These viruses possess a single-stranded RNA genome that is converted to double-stranded DNA shortly after infection of susceptible host cells. The DNA copy of the genome is transported into the nucleus where it is integrated into chromosomal DNA and replicated by host factors. In the case of HIV, the late stages of viral assembly occur on the inner leaflet of the plasma membrane. Virus release occurs by budding, permitting the virus to obtain a host-derived lipid envelope. The compositions of the viral envelope and the plasma membrane are not equivalent, suggesting that virus assembly may occur in nonrandom subdomains of cellular membranes.

The HIV-1 genome is comprised of nine genes (Fig. 1A). Six are unique to the HIV virus; three are common to other retroviruses. The common major genetic domains are 5′-*gag-pol-env*-3′ which are synthesized as a single precursor. Although all nine viral genes are essential for assembly of an infectious particle, the *gag* gene alone can direct the synthesis, transport to the plasma membrane, and assembly of the structural precursor polyprotein Gag, resulting in formation of particles that are morphologically indistinguishable in the electron microscope from immature (noninfectious) virus. The *pol* gene encodes the enzymatic proteins protease (PR), reverse transcriptase (RT), and integrase (IN). Expression of *pol* with *gag* results in PR-catalyzed proteolytic maturation of Gag and formation of particles that are morphologically indistinguishable in the electron microscope from authentic, mature, infectious virus (Fig. 1B). The *env* gene yields a precursor for the envelope glycoproteins gp120 and gp41. The surface (SU) glycoprotein gp120 determines cell tropism by attaching to CD4 receptors and specific co-receptors (members of diverse chemokine receptor molecules) on the surface of cells of lymphatic lineage. This action brings the viral and cellular membrane in sufficient proximity to promote lipid fusion mediated by the transmembrane (TM) glycoprotein gp41. Gene expression is achieved by three distinct mechanisms. *Gag* is translated directly from viral RNA. *Env* is encoded in a spliced mRNA. *Pol* is expressed as a *Gag–Pol* fusion product made by infrequent ribosomal frameshifting. This frameshifting is a subtle form of regulation that ensures overproduction of the structural *gag* gene products relative to the enzymatic *pol* gene proteins.

* Corresponding author. Department of Physiology and Biophysics, State University of New York at Stony Brook, Stony Brook, NY 11794-8661, USA. Tel.: +1-631-444-3071; fax: +1-631-444-3432.
E-mail address: suzanne@dualphy.pnb.sunysb.edu (S. Scarlata).

0005-2736/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0005-2736(03)00163-9

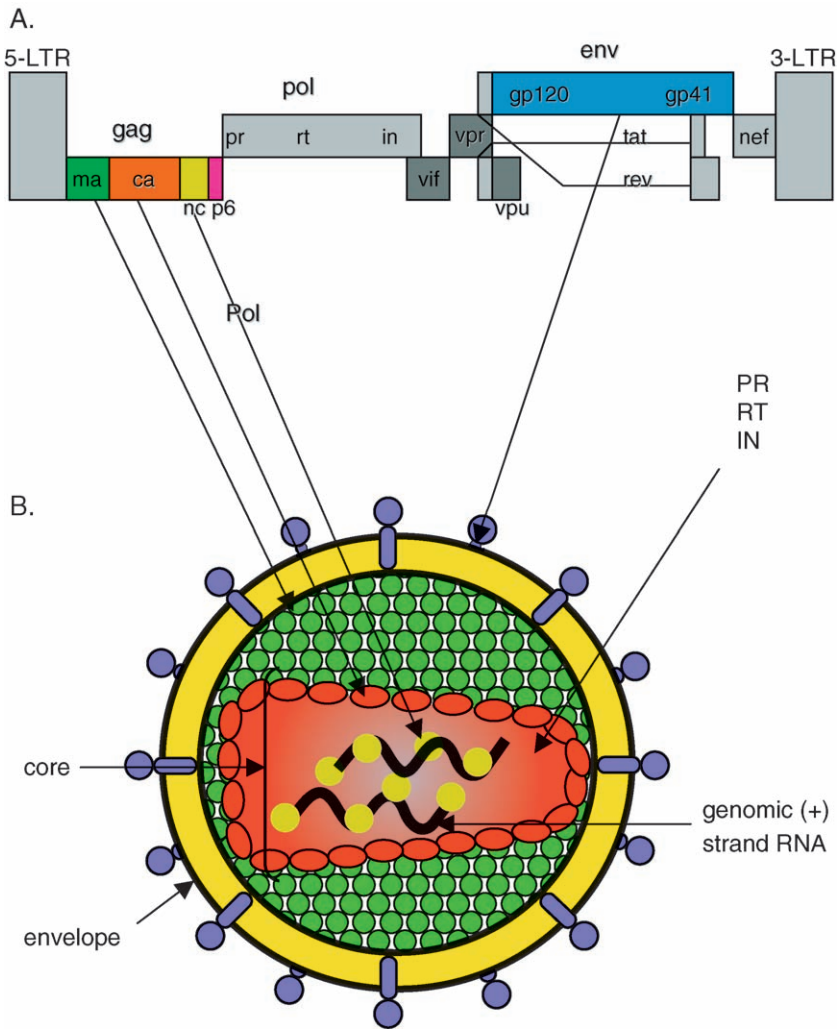

Fig. 1. (A) Organization of the HIV-1 proviral genome. (B) Organization of the HIV-1 mature viral particle.

Retroviruses have common morphological features that reflect similar mechanisms of assembly. Motifs that direct membrane binding, particle release, and high particle density are functionally interchangeable in the Gag structural precursor polyprotein. In HIV-1, each virion contains ~ 2000 copies of Gag. The Gag polyprotein, which is by itself sufficient for formation of immature particles, forms a spherical shell that looks similar for all retroviruses. The structural proteins that comprise mature infectious virus particles, matrix (MA), capsid (CA), spacer p2, nucleocapsid (NC), and spacer p1 and p6, originate from domains in Gag (Fig. 1B). The outermost shell underlying a lipid bilayer, derived from the plasma membrane during budding, is composed of MA. There is an inner core comprised of a shell assembled from the CA protein and inside this core is a ribonucleoprotein containing NC complexed to the diploid genomic RNA and the replicative enzymes. The Gag–Pol polyprotein contains all of the domains in Gag except for p6 in addition to the enzymes PR, RT, and IN, as noted above. Concomitant with, or following the release from infected cells, all retroviruses undergo morphological rearrangements associated with cleavage by viral PR of the precursor polyproteins into mature products. Activation of viral PR and proteolytic processing of the precursor proteins are essential steps in formation of infectious virus and convert the spherical immature particle to particles containing a characteristic conical-shaped capsid. In the mature particle, the host-derived membrane envelope is studded with the surface glycoprotein gp120 (SU) and the TM glycoprotein p41. SU and TM are also formed from precursor proteins (gp160). Proteolytic processing of the Env precursor into mature proteins occurs by cellular proteases.

The life cycle of HIV-1 is shown in Fig. 2 where the virus first enters the host cell and disassembles. RNA is then transcribed to DNA and associates with a pre-integration complex which enters the nucleus. The newly synthesized viral proteins are then trafficked to the plasma membrane where they assemble and bud from the cell to give the mature form of the virus. Since the Gag poly-

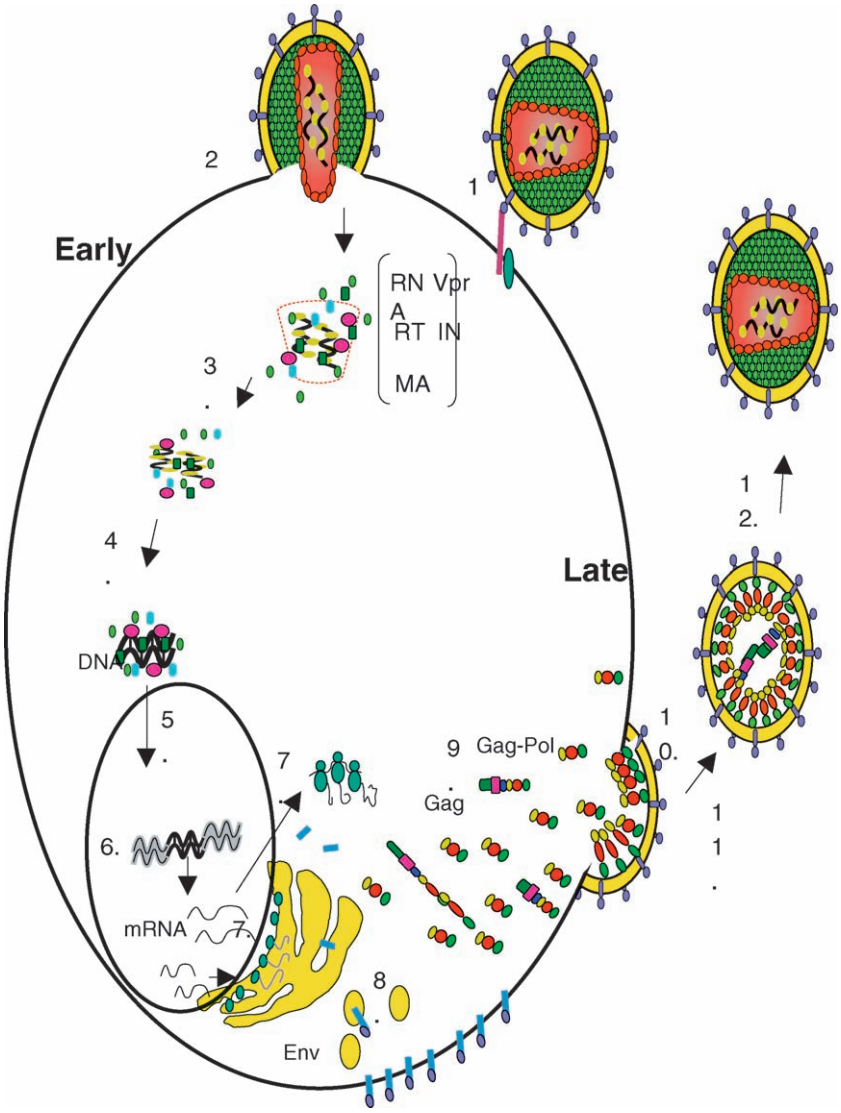

Fig. 2. Life cycle of HIV-1.

protein alone can give assembled particles, we will focus on the role of the individual Gag domains in controlling HIV-1 assembly.

2. Role of the individual Gag domains in viral assembly

2.1. Structure and function of the matrix protein (MA)

2.1.1. Conformation and oligomerization of MA

The MA domain of Gag is the region that forms the final contacts with the plasma membrane of the host cell and interacts with gp120 [3]. In virions, the MA domain is visualized as a thin membrane-bound layer by cryoelectron microscopy [4]. Structural analysis of recombinant MA shows that the membrane-binding face is comprised of a cluster of basic residues in the N- and C-terminal region that are brought into proximity by the three-dimensional fold of the protein [5] (Fig. 2). Genetic studies demonstrate that the regions responsible for membrane binding are located on the N-terminal region of MA which encompasses a cluster of basic residues and a myristyl group [6]. The protein crystallizes as a trimer with four helixes arranged in a circular pattern and one helix extending outward from the other four (Fig. 3). These helices are then capped off by a three-stranded mixed λ-sheet. The membrane-binding face of each subunit is tilted $\sim 30^\circ$ from the trimer interface encompassing helix 2, so it is unlikely that the protein would remain an oligomer with the same subunit contacts after membrane binding. Supporting this idea are fluorescence studies that have shown that even under conditions where MA is a trimer in solution, binding to model membranes results in complete dissociation of the oligomer [7]. Based on these studies and the observation that Gag can still form infectious particles if all of the MA domain was deleted expect for the myristoylation signal, it appears that

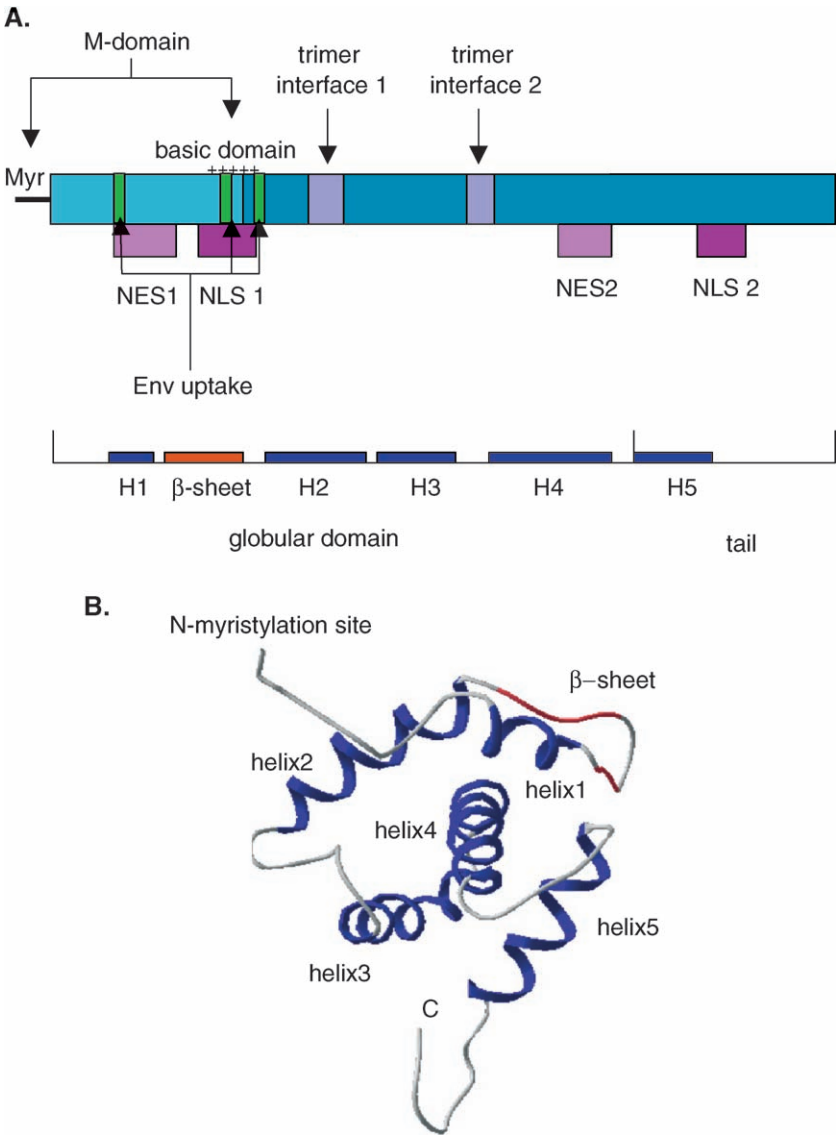

Fig. 3. (A) Organization of MA where NES and NLS refer to nuclear export and nuclear localization signals, respectively, and H1–5 refer to helices 1–5. (B) Ribbon diagram of the matrix protein.

MA contacts may not drive assembly [8]. We note that these observations do not argue against the idea that the MA domain may form trimers in the Gag–RNA complex that precedes membrane binding (see below). Mutations in the N-terminal region typically result in altered membrane binding and, in some cases, assembly and release events are affected [9]. The conclusion from these and other studies is that residues in the N-terminal domain of MA may affect assembly directly by fine-tuning Gag–Gag contacts made during assembly, or indirectly by affecting the structure of upstream domains.

2.1.2. Interaction of matrix with lipid membranes

The membrane binding properties of MA have been characterized both in vitro and in vivo [6,10,11]. Like Gag, purified MA binds most strongly to negatively charged vesicles as would be expected for a protein that contains a highly positively charged face. Even though other Gag domains may participate in membrane association [10], the MA domain of Gag, including its myristoylation signal, appears to be primarily responsible for plasma membrane targeting, since alterations in MA can result in loss of membrane association or the mislocalization of Gag to other cellular compartments [12,13]. A schematic of the MA sequence and structure is presented in Fig. 3.

Recently, many investigators have focused on the targeting of MA to lipid rafts in cells (see Ref. [14]). The description of lipid rafts was based on studies of the composition of the pellet fraction that remains after treatment of cells with a mild detergent, such as Triton X-100, (for review see Ref. [15]). These pellets contain a high composition of cholesterol and sphingomylein. Pellets of similar cholesterol and sphinglomylein ratios also form after detergent treatment in vitro using model membranes of

specific composition, and these have been used to study raft properties. Phase diagrams of these lipid mixtures showed that if lipids that have at least one saturated acyl chain are mixed with at least 33% cholesterol, then cholesterol will aggregate at temperatures above the phase transition of the lipid, usually below 4 °C. If a lipid with saturated chains was incorporated into this system, then the saturated chains could complex with cholesterol to form a 'liquid-ordered' phase which is characterized by high lipid mobility (see Ref. [15]). The term lipid raft was used to indicate that domains of lipids in the liquid-ordered phase coexist or 'float' in the lipids in the more fluid, liquid-disordered phase. Note that lipid rafts differ from caveolae which are ordered domains that contain high amounts of the protein caveolin.

Since the detection of lipid rafts in cells involves disruption and fractionation of the plasma membrane, then it is difficult to determine whether the protein aggregates that also pellet after cell disruption are associated with rafts [16]. Many studies of this type have implicated Gag-raft association (e.g. Refs. [17–21]). However, one could argue that the main components of rafts are uncharged and, since MA prefers negatively charged surfaces (see below), it is likely that the MA domain of Gag would prefer to be localized outside of rafts.

It has been well-established that GPI-anchored proteins target exclusively to rafts and these can be used as markers to define these regions in membranes as can other commercially available lipid fluorescent probes. Fluorescence microscopy studies using these markers show that HIV-1 Gag is targeted to raft domains, supporting studies using separation techniques [22]. It is noteworthy that one factor that has been shown to be responsible for raft localization in other protein systems is the presence of two saturated acyl chains, such as myristoyl and palmitoyl. Studies using model systems indicate that the MA membrane targeting signal, i.e. the basic residues and myristylation, will not result in raft-targeting [23] and thus for HIV-1 to localize to rafts, some other factor such as interaction of MA and envelope proteins [24], or fusion from endosomal membranes, must be responsible.

As expected from its sequence, the positive lobe of MA results in a strong attraction to membranes with negatively charged lipids and in vitro MA shows much weaker affinity towards electrically neutral membranes [25]. Since the inner leaflet of the plasma membrane is enriched in acidic phospholipids, it is possible that these negatively charged lipids could localize proteins with basic domains to the cytoplasmic face of the membrane. However, other cellular regions may contain anionic regions and so the precise roles of both the positive residues and myristoylation are still unclear. Based on in vitro measurements of the membrane energy of unmyristoylated Gag to model membranes the myristoyl moiety should only contribute ~1% of additional the binding energy [25,26]. Gag mutants lacking the myristoylation signal localize to other cellular membranes besides the plasma membrane and so the myristoyl group is thought to serve some targeting function [12,13]. Other myristoylated proteins are found in other cellular compartments and so it is unlikely that this signal serves solely as a plasma membrane targeting signal. It is also noteworthy that other closely related retroviruses do not use this signal and must use other unidentified plasma membrane targeting signals (e.g. Ref. [27]). Since some population of MA must interact with membranes in transient fashion (see below), it has been suggested that while in the context of Gag, the myristoyl group of MA is freely available to directly interact with the lipid bilayer and promote membrane binding. This model, termed the 'myristoyl-switch model', states that while the myristoyl group is exposed in the context of Gag, upon Gag maturation, MA refolds in such a way as to sequester the myristoylate in the protein matrix and thereby decrease its membrane affinity (see Refs. [28–30]). Indirect support for this model comes from numerous in vivo studies (e.g. Refs. [29,31]). At this point, it is unclear whether the myristoyl group is exposed in the context of Gag and can contribute directly to the membrane binding of Gag.

More insight into the possible role of myristoylation comes from studies in transiently transfected COS cells which have shown that myristoylation results in the formation of higher order Gag proteins in the cytosol that presumably are primed to target the plasma membrane [11]. While myristoylation may promote oligomerization of the Gag precursor by providing a hydrophobic patch to stabilize protein–protein associations, some proteolytic and antigenic sites differ when the myristylation group is attached, indicating that interaction of the myristoyl group with the protein core may result in alterations in the tertiary structure as well [32].

2.1.3. Role of matrix in virus assembly

As mentioned, in the context of Gag–Gag protein associations, it is unlikely that trimer formation of MA drives assembly per se or that the MA domain is required. However, mutations that prevent trimer formation in solution adversely affect Gag assembly [26,27]. And in processed form, NMR studies show that both MA and MA–CA form trimers in solution while isolated CA is monomeric under similar conditions [33].

The MA domain controls membrane binding and this membrane association, rather than its protein–protein contacts, is expected to be the primary role of MA in assembly. In vitro studies using model membranes and purified proteins and also fractionation studies in COS cells have shown that the membrane binding affinity of MA is weaker than that of the whole Gag protein [6,10,28]. Mutations in the N- or C-terminal domains of MA usually result in impaired binding in cells and, often, second mutations in other globular regions of MA can reverse this effect [34], supporting the idea that specific MA contacts are made during assembly. In model membranes, both mature CA and NC can interact with membranes [10], but it is unlikely that they contribute to membrane binding in vivo since these domains serve to

stabilize Gag oligomers and bind viral RNA (see below). Although this leaves MA as the primary membrane binding motif, in COS cell studies where the MA region was deleted and only the myristoyl signal remained [35,36], and studies where the MA region was replaced by a different viral protein, membrane binding and targeting of Gag were not affected, leading to the idea that trafficking of Gag in vivo is more complex than being driven by simply biophysical interactions between Gag and the plasma membrane (e.g. Ref. [37]). It is also important to note that in vitro assembly of Gag can occur in the absence of membranes [38,39].

2.1.4. Role of MA in non-assembly events

Besides membrane binding, MA may play other roles in the HIV-1 life cycle. MA has been implicated in release from the host cell mediated by the viral protein Vpu [40]. Also, MA appears to aid in the transport of HIV components into the nucleus. During the initial stages of virus infection, a small population of MA dissociates from the membrane, exposing a nuclear localization signal, and becomes incorporated into the preintegration complex [41,42]. Interestingly, one region of this nuclear localization signal lies in the highly basic region of MA that is involved in membrane binding and thus membrane binding must be somehow weakened to allow for MA dissociation (Fig. 3). While certain factors are known to weaken MA–membrane interactions, such as cleavage of the MA domain from Gag and dissociation of the MA trimers to monomers, most likely, other signals such as phosphorylation [43] or strong interaction with the other preintegration proteins occur. Note that another viral protein, Vpr, also contains a seemingly redundant nuclear localization signal.

MA also contains a signal to be exported out of the nucleus [44]. Nuclear export of MA is mediated by a central component in the protein export pathway, Crmp1p. Thus, MA dissociates from the membrane in the early stages of infection and incorporates into the preintegration complex which then enters the nucleus. During this time, the nuclear export signal must somehow be masked either by occlusion from components in the preintegration complex or by a modification such as phosphorylation. The nuclear export signal counteracts the nuclear localization signal to keep Gag in the cytoplasm and out of the nucleus. Since MA plays at least two critical roles in the HIV life cycle, it is not surprising that mutations in this region affect infectivity and viral transmission.

2.2. Structure and function of the capsid domain

2.2.1. Structure and assembly of capsid proteins

Sequentially located in the center of the Gag is the CA domain whose role in the viral life cycle is less understood (Fig. 1). Evidence exists for its essential role in particle assembly [36,37] and in vitro mature capsid alone has the capacity to assemble into particles [25]. However, the manner in which CA is organized in virus particles is not clear but several studies suggest regions of protein–protein interactions in the C-terminal domain or hexamerization in the N-terminal region [45–50]. Morphologically, mature CA forms the distinctive conical core of the virus that encapsulates the viral RNA–protein complex. Virions with abnormal core structures are defective in initiation of reverse transcription and exhibit reduced infectivity [51]. Thus, CA provides structural stability to the virion and also plays a key role in forming the protein–protein contacts required for productive assembly.

Sequentially, CA can be divided into two domains, an N-terminal domain (NTD) and a C-terminal domain (CTD) (Fig. 4). The structure of these separate domains as well as the whole protein has been solved [49,52,53]. In Fig. 5 we show the structure of capsid as solved using an Fab antibody to promote structural integrity and crystallization [54]. The N-terminal domain consists of five long helices forming a stable coiled-coil structure, two short ones, two λ-hairpins, and a Pro-rich loop. Although mutations in the N-terminal domain adversely affect core formation, the first 56 residues can be deleted without adversely affecting particle assembly. Deletions in the Pro-rich region give wild-type morphology but a loss in infectivity, which is thought to be due to loss of the incorporation of the host protein cyclophilin into the virions (see below).

Upon cleavage of CA from MA, the N terminus is thought to refold to form a new λ-hairpin helix stabilized by a buried salt bridge between Pro1 of CA and the carboxyl side chain of Asp^{51} [55]. N-terminal extensions ranging from a few residues to the entire MA domain on CA prevent the conversion from spherical to tubular structures while mutations in the spacer behind the C-terminal domain of CA affect aggregation but do not give regular tubular structures [38,56]. It is this change upon MA cleavage which is thought to be associated with the dramatic transformation to conical structures. Spherical, immature-like and tubular, mature-like CA assemblages can form from mature CA or CA-p2-NC in vitro, and a switch from tubular to spherical forms can be induced by changing the pH from 7.0 to 6.8 [57,58]. It is unlikely that this small change in pH alters the protonation state of the N terminus and more likely that other CA residues are involved. Nevertheless, the proximity of this switch to physiological pH values indicates that the local electronic environment in the virion may be critical in the maturation process.

The C-terminal half of CA contains four conserved helices and a highly conserved sequence among retroviruses known as the major homology region (MHR; 285–304 in HIV-1; see Fig. 4). This region is conserved throughout the retrovirus group [59] and thus offers a novel and stable target for viral vaccines. The role of the MHR sequence has been investigated using mutagenesis and biophysical strategies. Various mutations within the MHR block viral replication at different and distinct stages, such as assembly, maturation, or target cell infection in vivo, indicating that the MHR may play a role in Gag interactions with viral protein, the host

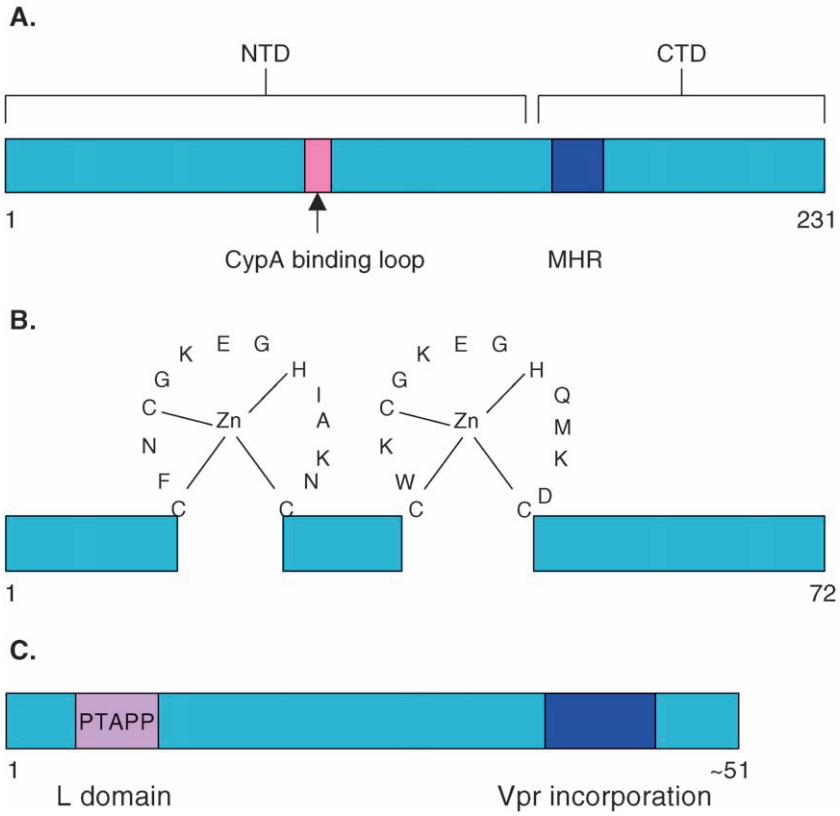

Fig. 4. Linear representation of the HIV Gag proteins CA (A), NC (B) and p6 (C).

membrane, or viral RNA [60]. Direct evidence that the MHR plays a role in membrane binding comes from biophysical studies showing that purified MHR-deleted Gag has a reduced binding affinity to model membranes as compared to wild-type Gag [53,61,62]. Indeed, deletion of the MHR results in abnormal Gag processing and we have shown that MHR-deleted Gag no longer has the ability to assemble on membrane or RNA substrates in vivo or in vitro, arguing that the critical contacts needed for Gag oligomerization reside in this region [63]. These results suggest that the MHR determines the Gag conformation required for productive protein–protein and protein–membrane interactions during assembly. Extending these studies to the assembly of wild-type Gag, RNA may act to seed the formation of Gag oligomers and the contacts in the CA domain may stabilize the aggregates.

2.2.2. The role of the host cell protein cyclophilin in the HIV life cycle

There are several host proteins that become incorporated into HIV virions but only one of these, cyclophilin A, has been shown to enhance viral infectivity[64–68]. Cyclophilin A (CypA) is a cytosolic protein that catalyzes the cis–trans isomerization of proline residues and is the target of the immunosuppressive drug cyclosporin [69]. Cyclophilin A is recruited into the virions through interactions with a proline-rich stretch in the N-terminal region of the capsid domain [70]. Only the retroviruses HIV and chimpanzee-specific SIV require cyclophilin A, and cyclophilin A dependency can be transferred to other viruses by transferring the cyclophilin A binding site. The Gag binding site for cyclophilin A, Gly^{89}–Pro^{90} in capsid (Fig. 4), exists in both the cis and trans forms and recent studies show that cyclophilin does catalyze Gly^{89}–Pro^{90} isomerization of capsid, suggesting a role for cyclophilin in maturation or disassembly [50,71]. The loop containing Pro^{90} in the capsid domain binds directly to hydrophobic binding pocket of cyclophilin [70] although other CA association sites may exist [72]. Conditions that disrupt cyclophilin A–CA association, such as the addition of cyclosporin, mutation of the association site, or depletion of the host cells cyclophilin A, lead to less infectious viruses that show reduced replication [73]. This loss in infectivity suggests that cyclophilin A plays some vital role in the virus's life cycle. Arguing against a role of cyclophilin A in late events is the observation that disruption of Gag–cyclophilin interaction still allows for assembly and budding to give particles that have the proper number of copies of Gag. The ratio of 10 copies of CA to 1 copy of cyclophilin in virions is supported by fluorescence binding studies, showing that cyclophilin will only bind to an aggregated form of immature CA and not a dissociated one [74]. This result suggests that the primary binding site for cyclophilin may lie across several subunits in a CA oligomer.

Fig. 5. Structure of capsid from Ref. [54]. This structure is of a protein with an amino-terminal extension containing His_6 tag. The amino-terminal extension implies that this structure to some extent simulates capsid in the context of Gag. The green region represents the MHR. NTD is the amino terminal domain and CTD is the carboxy terminal domain.

Our laboratories have found that the presence of cyclophilin A does not affect the oligomerization or dissociation of a form of CA that mimics immature conformation [74]. Since cyclophilin A does not affect the later assembly stages of the virus, it may assist in disassembly or uncoating of CA from the viral genome in the early stages of infection [64–66,75]. Indeed, cyclophilin binds differently to Gag and mature capsid, supporting the suggested role in post-assembly events [50]. It appears that the proline isomerase activity of cyclophilin is important for replication, and thus the isomerase activity of cyclophilin may be needed to act on some host protein for proper uncoating or, alternately, the cyclophilin protein may promote uncoating in a nonenzymatic manner.

It has also been observed that inhibitors of cyclophilin reduce the efficiency of viral attachment to target cells [76,77]. There are cyclophilin B receptors on the surface of T cells, and it is possible that cyclophilin A moves to the exterior surface of the virion and binds to this receptor aiding in docking and entry. However, our laboratory found that both cyclophilin A, capsid, and the complex are incapable of moving across or fusing to a model membrane surface, and so how this process would work is unclear (M. BonHomme, C. Carter and S. Scarlata, unpublished results). We have found that the binding affinity of cyclophilin A to capsid decreases dramatically as CA matures which could allow cyclophilin to dissociate from its binding site in the mature virion and move to other locations in the virus.

While cyclophilin has been implicated in promoting uncoating of CA, phosphorylation may also play a role [78]. CA has been shown to have three phosphorylation sites in vitro and two cellular kinases have been found in virions [79]. One of these kinases, MAPK ERK2, does not phosphorylate CA in vitro and so the second, unidentified kinase may be responsible.

2.3. Role of the nucleocapsid domain in the assembly of HIV particles

2.3.1. The importance of RNA binding to NC in Gag oligomerization

The NC domain of Gag is a highly basic region whose role is to encapsulate and protect viral RNA. Two copies of RNA are contained in each particle and the NC domain binds to these RNA through its zinc fingers (Fig. 4). The two fingers have been shown to be the minimal element needed for RNA binding. Proper encapsidation of the HIV genome has been found to involve an ~110-nucleotide segment known as the Psi-site, which contains four stem loops which are required for genome packing. Two of these stem loops bind to NC while one is thought to stabilize the structure of the site [80].

It is RNA binding that aids in localizing and concentrating Gag monomers to promote Gag assembly. While Gag does not show a strong binding preference for RNA over membranes in vitro, COS cell studies indicate that RNA binding of Gag precedes membrane binding, most likely due to a closer proximity between the two during assembly and in vitro studies support this pathway [63]. It is reasonable to assume that the localization of Gag monomers on RNA promotes Gag–Gag interactions between the CA and MA domains. The role of RNA in assembly is so vital that HIV will recruit other cellular RNAs if viral RNA is not available [81]. Upon Gag processing, NC promotes the refolding of genomic RNA to a thermostable dimer [82,83].

The region of nucleocapsid that binds RNA is the conserved I or interaction domain containing the zinc fingers [84,85]. This domain, which was first identified in RSV, is present in two copies in HIV and is required to form particles of proper density [86,87]. Interestingly, the zinc fingers are not necessary for I domain function and a string of basic residues can function as an I domain in RSV. However, charge alone cannot completely mimic I domain function and the observation that some of these I mutants can be rescued by subsequent mutagenesis leads to the simple model that the two I domains work in conjunction to correctly package RNA [88].

2.3.2. Do inositol phosphates participate in assembly?

In addition to membranes and RNA, there is good evidence that inositol phosophates may play a role in HIV assembly [89]. Inositol phosphates can be generated in the cell by various signaling proteins as well as simple housekeeping proteins. Inositol phosphates are rapidly generated

and modified in the cell and so their precise concentration in cells is very dynamic and difficult to quantify. However, it is not surprising that the high concentration of basic residues in the MA and NC domains would bind these small, anionic molecules and this binding may aid in the condensation of Gag molecules during assembly.

2.3.3. Binding of NC to VPR

Another important function of NC in terms of assembly is to incorporate cellular Vpr into virions along with the p6 domain [46,90,91]. As mentioned above, Vpr participates with MA in the preintegration complex and can induce cell cycle arrest in the G2 phase [92]. Vpr seems to associate to Gag through its C-terminal 16 residues and possibly the zinc fingers of NC [93]. It is possible that the Vpr–NC complex may work to interact with proteins involved in cell division.

2.4. Role of p6 in the late stages of the HIV life cycle

This small segment is at the sequential end of Gag and contains the late (L) domain which is required for the pinching off of the newly assembled virion from the host membrane (Fig. 4).

Within the p6 region of HIV-1 is a proline-rich motif PTAP(P), which appears to play critical roles in exocytosis of the assembled particle [94]. Our laboratory discovered that this region directly interacts with Tsg101, which is a host component of the cellular endocytosis machinery [95], and two other groups have found that budding and maturation is dependent on *tsg101* gene expression [96–98]. Tsg101 is an orthologue of the yeast vacuolar protein sorting protein 23 (Vps23) and an inactive homologue of Ub conjugating (E2) enzymes (see Ref. [99]). E2 enzymes transfer Ub to E3 enzymes. The PTAP(P) motif in HIV-1 Gag recruits a multi-component complex containing proteins involved in the trafficking machinery. This was elegantly demonstrated by Garrus et al. [96], who used small, interfering RNA (siRNA) to deplete cells of Tsg101 and block viral budding from the membrane. These investigators then linked the requirement for Tsg101 to the endosomal trafficking machinery by showing that dominant-negative mutants of Vps4, an ATPase required for intracellular trafficking of Tsg101, prevents virus release in an L-domain-dependent manner.

Protein trafficking through ubiquitin has been explored in other systems (see Ref. [100]). Ubiquitin (Ub)-dependent proteolysis is an important regulatory mechanism involved in diverse cellular functions such as cell cycle control, signal transduction, and regulation of membrane channels (see Ref. [100]). In the ubiquitinylation pathway, the Ub activation enzyme (E1) activates Ub by hydrolyzing ATP to form a high energy bond with Ub. Ub is then transferred to a Ub conjugation enzyme (E2, UbC) by formation of a thioester linkage between the C-terminal Gly of Ub and the SH group of the active site Cys in E2. Interactions between E2 and substrate proteins result in transfer of Ub to the epsilon-amino groups of lysine side chains on the substrate by catalyzing the formation of an isopeptide bond. In some cases, this process is mediated by a Ub protein ligase (E3). The E3 enzyme also catalyzes transfer of additional Ub molecules to form a polyubiquitin chain. It is believed that a single E1 transfers Ub to ~ 30 human E2 enzymes and that each E2 transfers to several E3s. The E3s can be substrate-specific or can recognize several substrates via similar but not identical motifs. Certain substrates can be targeted by several E3s, probably through distinct recognition motifs. Tsg101 has a functional link to Ub and Ub has been found either linked to lysine residues in p6 and/or free in purified virus particles of HIV-1 [79], suggesting that Ub plays a role in Gag trafficking. Additionally, residues in the NC region may also assist in this process [102]. Since at least four Ub are needed for efficient recognition for the proteosome [103] then polyUb Gag would be designed for degradation. These recent discoveries of the ability of HIV-1 Gag to exploit the cellular endocytotic machinery are expected to be explored in the upcoming years (see Refs. [101,104–107]).

3. Working model of HIV assembly

From the studies describe above, we can propose the following model for HIV-1 Gag assembly. Gag is first synthesized on free polysomes in the cytosol. It is probable that Gag then binds to intracellular vesicles entering the endosomal trafficking system. During or shortly after this time, Gag is exposed to newly synthesized viral RNA. While this interaction may occur in the cytosol, there is a finite probability that it occurs in the nucleus and whether the Gag–RNA complex exits through the nuclear export signal on MA is not clear. Binding to RNA through the I domain on NC localizes Gag onto the RNA strand, thereby promoting Gag–Gag association. This oligomerization and subsequent condensation of Gag occur through several contact points leaving Gag vulnerable to mutations that disrupt efficient protein–protein interactions. Gag oligomerization most likely occurs concurrently with RNA condensation and it is possible that highly charged free inositiol phosphosphates may aid in this condensation process. When Gag enters the endosomal trafficking machinery, the MA region becomes bound to endosomal lipids. The final delivery of Gag to the plasma membrane most likely involves the fusion of the endosomal membrane with the plasma membrane. This process would negate the state of Gag oligomerization before lipid binding, supporting cell-based studies indicating that tight Gag complexes are not required for complexation [108] even though in vitro studies show a weaker binding of the complex as opposed to the monomer [10]. The presence of RNA bound to the zinc fingers of NC may assist in appropriately positioning Gag in the correct orientation to the membrane surface thereby promoting interactions between the CA domain although,

since assembly will occur if the zinc fingers are deleted, this interaction is not required [109]. The N-terminal region of MA will contact with the membrane surface, and also contact with envelope proteins that are localized in lipid rafts which have most likely also undergone oligomerization [110]. As more Gag–RNA is delivered to the membrane, the particle buds outward from the cell and this budding maybe aided by the lipid composition of the underlying surface. The final interaction between p6 and some exocytotic machinery allows viral release in an energy-dependent fashion [111]. Shortly after or concurrently, maturation is triggered by some unknown process to yield infectious particles.

4. Concluding remarks

Even with the wealth of information about the assembly of HIV-1 in host cells, many questions as to how the various viral components find their way to the plasma membrane and interact to form productive complexes remain. While much can be gathered from numerous mutagenesis studies of the Gag protein and its individual domains, the challenge now is finding the specific host components, whether in the trafficking machinery or in the nucleus, that these domains must interact with to form productive complexes that can mature into infectious particles.

Acknowledgements

The authors are grateful to the generous support of the NIH GM58271 and to Marjorie BonHomme and Paxton Provitera for critically reading the manuscript and providing Fig. 5. The authors are grateful to Lynn VerPlank and Eva Guttwein for Figs. 1–4. The authors would also like to acknowledge the many background citations that were too numerous to specifically cite in the text, but nevertheless contributed to progress in this field.

References

[1] J.M. Coffin, S.H. Hughes, H.E. Varmus, Retroviruses, Cold Spring Harbor Laboratory Press, Woodbury, NY, 1997.
[2] A.D. Frankel, J.A. Young, Ann. Rev. Biochem. 67 (1998) 1–25.
[3] H.R. Gelderbloom, AIDS 5 (1991) 617–638.
[4] T. Wilk, I. Gross, B.E. Gowen, T. Ruttne, F. de Haas, R. Welker, H.G. Krausslich, P. Boulanger, S.D. Fuller, J. Virol. 75 (2001) 759–771.
[5] C. Hill, D. Worthylake, D.P. Bancroft, A.M. Christensen, W.I. Sundquist, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 3099–3104.
[6] W. Zhou, L. Parent, J. Wills, M. Resh, J. Virol. 68 (1994) 2556–2569.
[7] S. Scarlata, L.S. Ehrlich, C.A. Carter, J. Mol. Biol. 277 (1998) 161–169.
[8] C.-T. Wang, Y.-C. Chou, C.-C. Chaing, J. Virol. 74 (2000) 3418–3422.
[9] E.O. Freed, J.M. Orenstein, A.J. Buckler-White, M.A. Martin, J. Virol. 68 (1994) 5311–5320.
[10] L. Ehrlich, S. Fong, S. Scarlata, G. Zybarth, C. Carter, Biochemistry 35 (1996) 3933–3943.
[11] L. Hermida-Matsumoto, M. Resh, J. Virol. 74 (2000) 8670.
[12] M. Bryant, L. Ratner, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 523–527.
[13] P. Spearman, J.J. Wang, N.V. Heyden, L. Ratner, J. Virol. 68 (1994) 3232–3242.
[14] S. Campbell, S.M. Crowe, J. Mak, J. Clin. Virol. 22 (2001) 217–227.
[15] D.A. Brown, E. London, Annu. Rev. Cell Dev. Biol. 14 (1998) 111–136.
[16] L. Ding, A. Derdowski, J.J. Wang, P. Spearman, J. Virol. 77 (2003) 1916–1926.
[17] I. Russo, M. Mixon, B. Chen, P.S. Kim, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 13523–13535.
[18] Y.H. Zheng, A. Plemenitas, T. Linnenmann, O.T. Fackler, P.M. Peterlin, Curr. Biol. 11 (2001) 875–879.
[19] K. Holm, K. Weclewicz, R. Hewson, M. Suomalainn, J. Virol. 77 (2003) 4805–4817.
[20] A. Ono, E.O. Freed, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 13925–13930.
[21] O. Linwasser, M. Resch, J. Virol. 75 (2001) 7513–7524.
[22] D. Nguyen, J. Hildreth, J. Virol. 74 (2000) 3264–3272.
[23] J.B. McCabe, L.G. Berthiaume, Mol. Biol. Cell 12 (2001) 3601–3617.
[24] Y.H. Zheng, A. Plemenitas, T. Linnenmann, O.T. Fackler, B.M. Peterlin, Curr. Biol. 11 (2001) 875–879.
[25] L.S. Ehrlich, B. Agresta, C.A. Carter, J. Virol. 66 (1992) 4874–4883.
[26] R. Peitzch, S. McLaughlin, Biochemistry 32 (1993) 10436–10443.
[27] P. Provitera, F. Bouamr, D. Murray, C. Carter, S. Scarlata, J. Mol. Biol. 296 (2000) 887–898.
[28] W. Zhou, M.D. Resh, J. Virol. 70 (1996) 8540–8548.
[29] P. Spearman, R. Horton, L. Ratner, I. Kuli-Zade, J. Virol. 71 (1997) 6582–6592.
[30] L. Hermida-Matsumoto, M.D. Resh, J. Virol. 73 (1997) 1902–1908.
[31] A. Ono, E.O. Freed, J. Virol. 73 (1999) 4136–4144.
[32] F. Bouamr, S. Scarlata, C. Carter, Biochemistry 42 (2003) 6408–6417.
[33] Y. Morikawa, W.-H. Zhang, D.J. Hockley, M.V. Nermut, I.M. Jones, J. Virol. 72 (1998) 7659–7663.
[34] R. Kiernan, A. Ono, E.O. Freed, J. Virol. 73 (1999) 4728–4737.
[35] P.P. Lee, M.L. Linial, J. Virol. 68 (1994) 6644–6654.
[36] C.-T. Wang, Y. Zhang, J. McDermott, E. Barklis, J. Virol. 67 (1993) 7067–7076.
[37] M.A. Accola, B. Strack, H.G. Gottlinger, J. Virol. 74 (2000) 5395–5402.
[38] I. Gross, H. Hohenberg, C. Huckhagel, H.-G. Krausslich, J. Virol. 72 (1998) 4798–4810.
[39] S. Campbell, A. Rein, J. Virol. 73 (1999) 2270–2278.
[40] A. Deona, P. Spearman, L. Ratner, Virology 269 (2000) 305–312.
[41] M.I. Bukrinsky, S. Haggerty, M.P. Dempsey, N. Sharova, A. Adzhubei, L. Spitz, P. Lewis, D. Goldfarb, M. Emerman, M. Stevenson, Nature 365 (1993) 666–669.
[42] P. Gallay, S. Swingler, J. Song, F. Bushman, D. Trono, Cell 83 (1995) 569–576.
[43] A.G. Bukrinskaya, A. Ghorpade, N.K. Heinzinger, T.E. Smithgall, R.E. Lewis, M. Stevenson, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 367–371.
[44] S. Dupont, N. Sharova, C. DeHoratius, C.-M. Virbasius, X. Zhu, A. Burinskaya, M. Stevenson, M. Green, Nature 402 (1999) 681–685.
[45] J. Lanman, T.T. Lam, S. Barns, M. Sakalian, M.R. Emmett, A.G. Marshall, P.E. Prevelige, J. Mol. Biol. 325 (2003) 759–772.
[46] W. Paxton, R.I. Conner, N.R. Landau, J. Virol. 67 (1993) 7229–7237.

[47] P. Gamble, S. Yoo, F. Vajvos, U. von Schwedler, D. Worthphylake, H. Wang, J. McCutcheon, W. Sundquist, C. Hill, Science 278 (1997) 849–853.
[48] S. Li, C. Hill, W. Sundquist, J. Finch, Nature 407 (2000) 409–413.
[49] C. Momany, L. Kovari, A. Prongay, W. Keller, R. Gitti, B. Lee, A. Gorbalenya, L. Tong, J. McClure, L. Ehrlich, M. Summers, C. Carter, M. Rossman, Nat. Struct. Biol. 3 (1996) 763–770.
[50] L. Dietrich, L. Ehrlich, T. LaGrassa, D. Ebbets-Reed, C. Carter, J. Virol. 75 (2001) 4721–4733.
[51] S. Tang, T. Murakami, B. Agresta, S. Campbell, E.O. Freed, J. Levin, J. Virol. 75 (2001) 9357–9366.
[52] R. Gitti, B. Lee, J. Walker, M.K. Summers, S. Yoo, W. Sundquist, Science 273 (1996) 231–235.
[53] T. Gamble, S. Yoo, F. Vajdos, U. von Schwedler, D. Worthylake, H. Wang, J. McCutcheon, W. Sundquist, C. Hill, Science 278 (1997) 849–853.
[54] C. Berthet-Colominas, S. Monaco, A. Novelli, G. Sibai, F. Mallet, S. Cusak, EMBO J. 18 (1999) 1124–1136.
[55] C. Tang, Y. Ndassa, M.F. Summers, Nat. Struct. Biol. 9 (2002) 537–543.
[56] H.G. Krausslich, et al., J. Virol. 69 (1995) 3407–3415.
[57] L. Ehrlich, T. Liu, S. Scarlata, B. Chu, C. Carter, Biophys. J. 81 (2001) 586–594.
[58] I. Gross, H. Hohenberg, T. Wilk, K. Wiegers, M. Grattinger, B. Muller, S.D. Fuller, H.G. Krausslich, EMBO J. 9 (2000) 103–113.
[59] J. Wills, R. Craven, AIDS 5 (1991) 639–654.
[60] F. Mammano, A. Ohagen, S. Hoglund, H. Gottlinger, J. Virol. 68 (1994) 4927–4936.
[61] N. Srinivasakanan, et al., J. Virol. 69 (1995) 6106–6612.
[62] D. Ebbets-Reed, S. Scarlata, C. Carter, Biochemistry 35 (1996) 14268–14275.
[63] P. Provitera, A. Goff, A. Harenberg, F. Bouamr, C. Carter, S. Scarlata, Biochemistry 40 (2001) 5565–5572.
[64] N.A. Kootsra, C. Monk, N. Tonnu, N.R. Landau, I. Verma, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 1298–1303.
[65] G. Towers, M. Bock, S. Martin, Y. Teakeuchi, J. Stoye, O. Danos, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 12295–12299.
[66] G. Towers, N. Collins, Y. Takeuchi, J. Virol. 76 (2002) 2548–2550.
[67] E.K. Franke, H.E. Yuan, J. Luban, Nature 372 (1994) 359–362.
[68] M. Thaili, A. Bukovsky, E. Kondo, B. Rosenwirth, C. Walsh, J. Sodroski, H.G. Gottlinger, Nature 372 (1994) 363–365.
[69] R. Handschumacher, M.W. Harding, J. Rice, R.J. Drugge, Science 226 (1984) 544–547.
[70] D. Braaten, H. Ansari, J. Luban, J. Virol. 71 (1997) 2107–2113.
[71] D. Bosco, E. Eisenmesser, S. Pochapsky, W. Sundquist, D. Kern, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 5247–5252.
[72] M. Endrich, P. Gehrig, H. Gehring, J. Biol. Chem. 274 (1999) 5326–5332 (274).
[73] D. Braaten, E. Franke, J. Luban, J. Virol. 70 (1996) 3551–3560.
[74] M. BonHomme, S. Wong, C. Carter, S. Scarlata, Biophys. Chem. (2003) (in press).
[75] A. Saphire, M. Bobardt, P. Gallay, J. Virol. 76 (2002) 2255–2262.
[76] A. Saphire, M. Bobardt, P. Gallay, EMBO J. 18 (1999) 6771–6785.
[77] B. Sherry, G. Zybarth, M. Alfano, L. Dubrobsky, R. Mitchell, D. Rich, P. Ulrich, R. Bucala, A. Cerami, M. Bukrinsky, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 1758–1763.
[78] C. Cartier, P. Sivard, C. Tranchatt, D. Decimo, C. Desgranges, V. Boyer, J. Biol. Chem. 274 (1999) 19434–19440.
[79] C. Cartier, M. Deckert, C. Grangeasse, R. Trauger, F. Jensen, A. Bernard, A. Cozzane, C. Desgranges, V. Boyer, J. Virol. 71 (1997) 4832–4837.
[80] G.K. Amarasinghe, J. Zhou, M. Miskimon, K.J. Chancellor, J.A. McDonald, A.G. Matthews, R.R. Miller, M.D. Rouse, M.K. Summers, J. Mol. Biol. 314 (2001) 961–970.
[81] A. Khorchid, R. Halwani, M.A. Wainberg, L. Kleiman, J. Virol. 76 (2002) 4131–4137.
[82] S.J. Kim, M.J. Kim, J.H. Lee, J.C. You, S. Jeong, Biochem. Biophys. Res. Commun. 291 (2002) 925–931.
[83] W. Fu, et al., J. Virol. 68 (1994) 5013.
[84] Y.-X. Feng, et al., Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 7577.
[85] D. Muriaux, et al., J. Biol. Chem. 271 (1996) 33686.
[86] L. Parent, R. Bennett, R. Craven, T. Nelle, N. Krishna, J. Bowzard, C. Wilson, B. Puffer, R. Montelaro, J. Wills, J. Virol. 69 (1995) 5455–5460.
[87] L. Garnier, B. Bowzard, J.W. Will, AIDS 12 (1998) S5–S16.
[88] A. Cimarelli, S. Sandin, S. Hoglund, J. Luban, J. Virol. 74 (2000) 4273–4283.
[89] S. Campbell, R. Fisher, E. Towler, A. Fox, H. Isaaq, T. Wolfe, L. Phillips, A. Rein, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10875–10879.
[90] Y.L. Lu, et al., J. Virol. 69 (1995) 6873–6879.
[91] E. Kondo, H.G. Gottlinger, J. Virol. 70 (1996) 1959–1964.
[92] N. Heinzinger, M. Bukinsky, S. Haggerty, A. Ragland, V. Kewalrammi, M.-A. Lee, H. Gendelman, L. Ratner, M. Stevenson, M. Emerman, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 7311–7315.
[93] H. deRocquigny, P. Petitjean, V. Tanchou, D. Decimo, L. Drouot, T. Delaunay, J.-L. Darlix, B. Roques, J. Biol. Chem. 272 (1997) 30753–30759.
[94] H.G. Gottlinger, T. Dorfman, J. Sodroski, W. Haseltine, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 3195–3199.
[95] L. VerPlank, F. Bouamr, A. LaGrassa, B.A. Kikonyogo, J. Leis, C. Carter, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 7724–7729.
[96] J.E. Garrus, et al., Cell 107 (2001) 55–65.
[97] J. Martin-Serrano, et al., Nat. Med. 7 (2001) 1313–1319.
[98] J. Martin-Serrano, T. Zang, P.D. Bieniasz, J. Virol. 77 (2003) 4794–4804.
[99] S. Lemmon, L. Traub, Curr. Opin. Cell Biol. 12 (2000) 457–466.
[100] J. Gruenberg, Nat. Rev., Mol. Biol. 2 (2001) 721–730.
[101] B. Strack, A. Calistri, M.A. Accola, G. Palu, H.G. Gottlinger, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 13063–13068.
[102] B. Strack, A. Calistri, H.G. Gottlinger, J. Virol. 76 (2002) 5472–5479.
[103] J. Thrower, L. Hoffman, M. Reschsteiner, C. Pickart, EMBO J. 19 (2000) 94–102.
[104] E.O. Freed, Trends Microbiol. 11 (2003) 56–59.
[105] O. Pronillos, J.E. Garrus, W. Sundquist, Trends Cell Biol. 12 (2002) 569–579.
[106] C. Carter, Trends Microbiol. 10 (2002) 203–205.
[107] E.O. Freed, J. Virol. 76 (2002) 4679–4687.
[108] A. Ono, D. Demirov, E.O. Freed, J. Virol. 74 (2000) 5142–5150.
[109] A. Aldorini, R. Young, J. Virol. 64 (1990) 1920–1926.
[110] S.-F. Lee, C.-T. Wnag, J. Liang, S.-L. Hong, C.-C. Huang, S.S.-L. Chen, J. Biol. Chem. 275 (2000) 15809–15819.
[111] M. Tritel, M. Resh, J. Virol. 74 (2001) 5473–5481.