

# Conformation and dynamics of the Gag polyprotein of the human immunodeficiency virus 1 studied by NMR spectroscopy

Lalit Deshmukh<sup>a</sup>, Rodolfo Ghirlando<sup>b</sup>, and G. Marius Clore<sup>a,1</sup>

<sup>a</sup>Laboratory of Chemical Physics and <sup>b</sup>Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

Contributed by G. Marius Clore, January 30, 2015 (sent for review January 12, 2015; reviewed by Robert T. Clubb and Michael Summers)

**Assembly and maturation of the human immunodeficiency virus type 1 (HIV-1) are governed by the Gag polyprotein.** Here we study the conformation and dynamics of a large HIV-1 Gag fragment comprising the matrix, capsid, spacer peptide 1 and nucleocapsid domains (referred to as ΔGag) by heteronuclear multidimensional NMR spectroscopy. In solution, ΔGag exists in a dynamic equilibrium between monomeric and dimeric states. In the presence of nucleic acids and at low ionic strength ΔGag assembles into immature virus-like particles. The structured domains of ΔGag (matrix, the N- and C-terminal domains of capsid, and the N- and C-terminal zinc knuckles of nucleocapsid) retain their fold and reorient semi-independently of one another; the linkers connecting the structural domains, including spacer peptide 1 that connects capsid to nucleocapsid, are intrinsically disordered. Structural changes in ΔGag upon proteolytic processing by HIV-1 protease, monitored by NMR in real-time, demonstrate that the conformational transition of the N-terminal 13 residues of capsid from an intrinsically disordered coil to a β-hairpin upon cleavage at the matrix|capsid junction occurs five times faster than cleavage at the capsid|spacer peptide 1 junction. Finally, nucleic acids interact with both nucleocapsid and matrix domains, and proteolytic processing at the spacer peptide 1|nucleocapsid junction by HIV-1 protease is accelerated in the presence of single-stranded DNA.

HIV-1 Gag | interdomain motion | proteolytic processing | residual dipolar couplings | real time NMR

**G**ag, the primary structural polyprotein of human immunodeficiency virus type 1 (HIV-1), mediates all essential events during the assembly and maturation of HIV-1 (1–5). Gag is composed of six organizational units, matrix (MA)–capsid (CA)–spacer peptide 1 (SP1)–nucleocapsid (NC)–spacer peptide 2 (SP2)–p6, with an HIV-1 protease cleavage site at each junction. During HIV-1 assembly, cotranslationally myristoylated Gag traffics from the cytoplasm to the plasma membrane of the host cell and multimerizes to generate spherical immature virions that subsequently bud from the host cell. During the budding process, controlled proteolysis of Gag by HIV-1 protease generates the individual Gag constituents, thereby triggering a cascade of events that eventually leads to the formation of mature, infectious virions containing conical capsids. A better understanding of these late stages of the HIV-1 life cycle has potential therapeutic implications because any alterations in these interconnected events results in the production of defective, noninfectious virions.

Here, we present a heteronuclear NMR study of a recombinant, nonmyristoylated HIV-1 Gag fragment comprising MA–CA–SP1–NC (hereafter referred to as ΔGag) to characterize the interrelationship between the domains and the conformational changes occurring during processing by HIV-1 protease.

## Results and Discussion

**Biophysical Analyses of Recombinant ΔGag.** Nonmyristoylated recombinant ΔGag (HXB2 strain, residues 1–432, ~48.3 kDa per subunit) was purified by ion exchange and size exclusion

chromatography, and analyzed using mass spectrometry and analytical ultracentrifugation (see *SI Materials and Methods* for additional information). In contrast to isolated CA, which assembles into hollow cylinders in the presence of high ionic strength buffer (6), ΔGag requires high ionic strength ( $\geq 300$  mM sodium chloride) to prevent oligomerization and subsequent assembly. Under these conditions, analytical ultracentrifugation (*SI Materials and Methods*) shows that ΔGag exists in a monomer–dimer equilibrium ( $K_{\text{dimer}} = 35 \pm 5 \mu\text{M}$ ), mediated by the C-terminal domain of CA (Table S1). These results are in agreement with chemical cross-linking studies that showed that Gag is predominantly dimeric in the cytoplasm and does not form higher order oligomers until it reaches the plasma membrane (7). Gag trimerization in the presence of inositol phosphate derivatives has been reported previously (8). However, no evidence for the existence of a ΔGag trimer or higher order structures (e.g., hexamer comprising a trimer of dimers) was found by analytical ultracentrifugation either in the presence of inositol hexakisphosphate ( $K_{\text{dimer}} = 35 \pm 5 \mu\text{M}$ ) or nucleic acids ( $K_{\text{dimer}} = 14\text{--}18 \mu\text{M}$ ) (Table S1 and Fig. S1).

In the presence of DNA and at low ionic strength (~100 mM NaCl), ΔGag assembles into spherical virus-like particles, including complete assemblies as well as partially formed shells, that can be visualized by negative stain electron microscopy (Fig. 1). Because recombinant ΔGag is not myristylated, these experiments were carried out in the absence of lipid/membranes, which are present in immature HIV-1 particles. Fully formed ΔGag particles are similar in appearance to authentic immature HIV-1 virions (9) in terms of their dimensions (~80 nm in diameter), curvature and palisade-like shell appearance.

## Significance

The late stages of the HIV-1 life cycle are governed by the Gag polyprotein (also known as group specific antigen). Here we use modern solution NMR methods to probe the structure and conformation of Gag. We show that Gag exists in a dynamic monomer–dimer equilibrium in solution and its structured domains reorient semi-independently of one another. We monitor proteolytic processing of Gag by HIV-1 protease in real-time, and show that the conformational transition at the N-terminus of the capsid domain from a random coil to a β-hairpin upon cleavage at the matrix|capsid junction occurs five times faster than cleavage at the capsid|spacer peptide 1 junction. These results provide insights into the HIV-1 life cycle that may lead to the development new antiretroviral strategies.

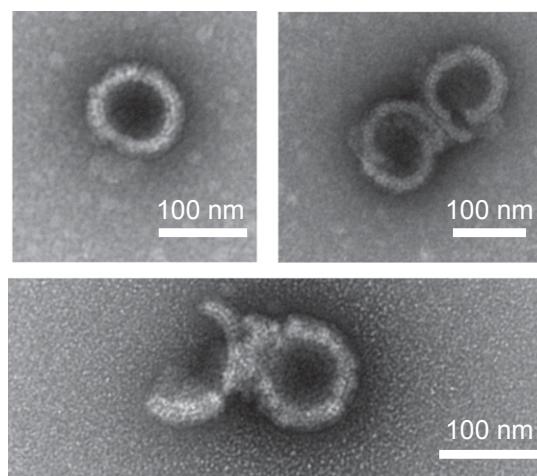
Author contributions: L.D., R.G., and G.M.C. designed research; L.D. and R.G. performed research; L.D., R.G., and G.M.C. analyzed data; and L.D. and G.M.C. wrote the paper.

Reviewers: R.T.C., University of California, Los Angeles; and M.S., Howard Hughes Medical Institute, University of Maryland, Baltimore County.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence should be addressed. Email: mariusc@mail.nih.gov.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1501985112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1501985112/-DCSupplemental).



**Fig. 1.** Negatively stained EM images of  $\Delta$ Gag assemblies. Assembly was initiated by the addition of d(TG)<sub>15</sub> DNA to ~1 mg/mL  $\Delta$ Gag at a protein:DNA molar ratio of 5:1. (See *SI Materials and Methods* for EM sample preparation details.) Samples were negatively stained with 1% uranyl acetate.

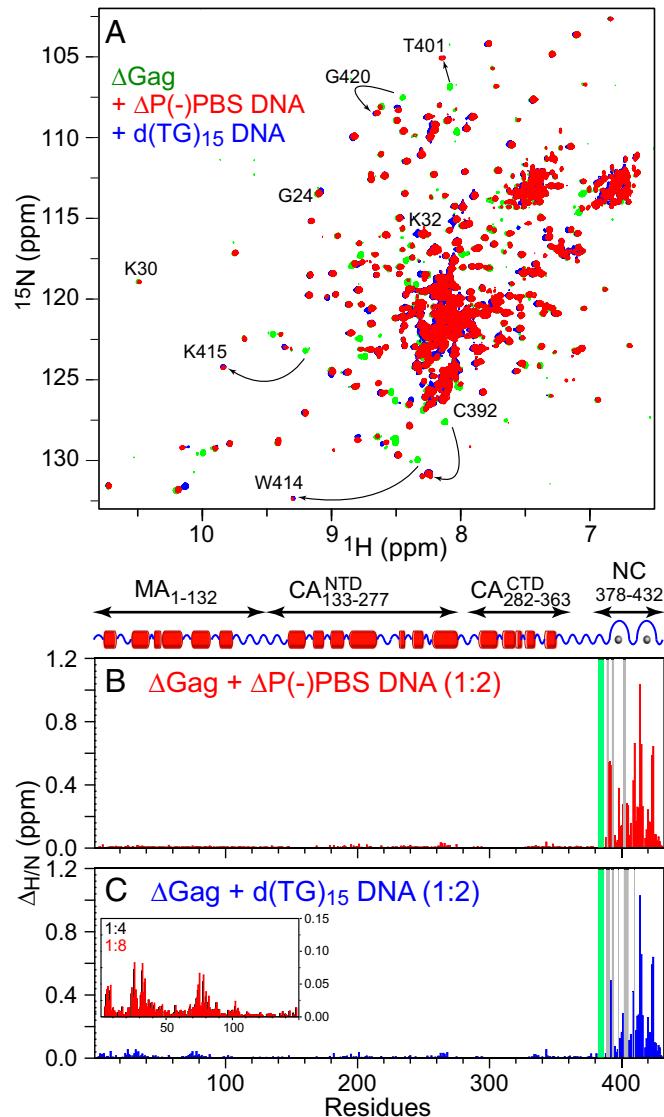
**ΔGag–Nucleic Acids Interactions.** The NC domain is involved in viral genome recognition and possesses RNA/DNA binding ability (4, 10, 11). In addition, MA has also been implicated in interactions with nucleic acids (12, 13). To probe structural and conformational changes that take place in  $\Delta$ Gag upon nucleic acid binding, we made use of  $^1\text{H}_N/^{15}\text{N}$  chemical shift perturbation mapping.

Although the overall size of the  $\Delta$ Gag dimer is large for solution NMR studies (~100 kDa), excellent spectral quality was obtained by perdeuteration in combination with transverse relaxation optimized spectroscopy (TROSY) (14), and low protein concentrations ( $\leq 0.4$  mM in subunits) and high salt (300 mM NaCl) to avoid protein aggregation (Fig. 2). As for isolated CA (15) and the smaller CA-SP1-NC Gag fragment (16),  $^1\text{H}_N/^{15}\text{N}$  cross-peaks for residues within and adjacent to the hinge region connecting the N-terminal (CA<sup>NTD</sup>) and C-terminal (CA<sup>CTD</sup>) domains of CA (residues 277–286) and at or near the CA dimerization interface (residues 300–325) are not observed in  $\Delta$ Gag due to monomer–dimer exchange on a time scale that is intermediate on the chemical shift scale.

Addition of a 2:1 (DNA to protein) molar equivalent of two different DNA oligonucleotides as substitutes for viral genomic RNA [a linear single-stranded 30mer d(TG)<sub>15</sub> or a 14mer hairpin  $\Delta$ P(–)PBS, representing the DNA (–) primer binding site (4)] results in significant  $^1\text{H}_N/^{15}\text{N}$  chemical shift perturbations within NC due to the formation of a 1:1  $\Delta$ Gag/DNA complex (16) (Fig. 2). No chemical shift perturbations are observed for CA, SP1, and the linker connecting MA and CA. Further increases in DNA results in concentration dependent perturbations in three distinct regions of MA (residues 4–8, 24–37, and 73–82) indicative of a secondary weak interaction (Fig. 2C, *Inset*), confirming previous NMR studies on free MA (17). Among these, the N-terminal MA motif ( $^1\text{M GARASVL}^8$ ) is involved in myristylation and plays a key role in plasma membrane interactions; the conserved “highly basic region” ( $^{18}\text{KIRLRPGGGKKYKLKH}^{33}$ ) interacts with anionic lipid headgroups of the plasma membrane; the  $^{69}\text{QTGSEE}^{74}$  region has been implicated in MA trimerization (18); and residues  $^{76}\text{RSLY}^{79}$  are involved in interactions with phosphatidylinositol 4,5-bisphosphate, a factor that regulates Gag localization to the plasma membrane (19). The overall magnitude of the chemical shift perturbations in the MA domain upon addition of DNA, however, is much smaller than that within NC, even at high DNA concentrations, indicative of the strong and weak nature of nucleic acid interactions with NC and MA, respectively. The exact role of nucleic acid (including tRNA) binding to MA is

unclear but it has been suggested that it serves to negatively regulate plasma membrane interactions by preventing nonspecific binding between residues of the “highly basic region” and acidic lipids and by inhibiting myristate-dependent hydrophobic interactions (12, 13, 20).

**ΔGag Conformation and Dynamics.** Residual dipolar couplings (RDCs) in weak alignment media provide information on bond vector orientations and overall molecular shape (21). Backbone ( $^1\text{D}_{\text{NH}}$ ) RDCs, measured in PEG/hexanol (22), were used to probe domain reorientation of  $\Delta$ Gag in the absence of nucleic



**Fig. 2.** Interaction of  $\Delta$ Gag with DNA. (A) Overlay of the  $^1\text{H}-^{15}\text{N}$  TROSY correlation spectra of  $^2\text{H}/^{15}\text{N}$ -labeled  $\Delta$ Gag in the absence (green) and presence of  $\Delta$ P(–)PBS (red) and 5' d(TG)<sub>15</sub> (blue) DNA. Some isolated cross-peaks from the MA and NC domains that undergo chemical shift changes upon addition of DNA are labeled. (B and C)  $^1\text{H}_N/^{15}\text{N}$  chemical shift perturbation profiles upon addition of  $\Delta$ P(–)PBS (red) (B) and 5'd(TG)<sub>15</sub> (blue) (C) DNA. Experimental conditions: protein:DNA molar ratio of 1:2 at a  $\Delta$ Gag concentration of 50  $\mu$ M per subunit; 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM ZnCl<sub>2</sub>, 1 mM TCEP, 30 °C. Semitransparent gray bars indicate NC residues (387, 389–390, 393–394, and 402–403) that exhibit line broadening in the presence of DNA; and green bars, NC residues (383–386) that could not be assigned unambiguously in the absence of DNA. (C, *Inset*) The  $^1\text{H}_N/^{15}\text{N}$  chemical shift perturbations observed for the MA domain upon addition of higher DNA concentrations.

acids. Analysis of the RDC data by singular value decomposition (SVD) was carried out using the crystal structures of MA [Protein Data Bank (PDB) ID 4JMU; ref. 23] and CA (PDB ID 3NTE; ref. 24) and an NMR structure of NC (PDB ID 2M3Z; ref. 25) as templates.

The five folded domains of  $\Delta$ Gag (i.e., MA, CA<sup>NTD</sup>, CA<sup>CTD</sup>, and the N- and C-terminal zinc knuckles of NC) display excellent agreement between observed and calculated RDC values with overall RDC *R*-factors (26) of less than ~26% within regions of secondary structure (Fig. 3), indicating that their folds are essentially unchanged relative to their isolated counterparts.

The magnitude of the principal component of the alignment tensor ( $D_a^{\text{NH}}$ ) and rhombicity ( $\eta$ ) for the individual domains of  $\Delta$ Gag differ significantly from one another (Table S2), well outside any differences that could be accounted for by measurement and/or coordinate errors (27), indicating that the five structural units of  $\Delta$ Gag behave as rigid bodies that reorient semi-independently of one another in solution (i.e., wobble independently in cones with semiangles >30°) (28), and that any interactions between the domains, if present, are highly transient. (Thus  $D_a^{\text{NH}}$  for the CA<sup>NTD</sup> is 40–60% larger than for MA, depending on the fraction of dimer present;  $D_a^{\text{NH}}$  for the CA<sup>CTD</sup> and CA<sup>NTD</sup> are of opposite sign with  $\eta$  axially symmetric for the former but semirhombic for the latter;

$D_a^{\text{NH}}$  for the N-terminal zinc knuckle of NC and the CA<sup>CTD</sup>, both of which are axially symmetric, are of opposite sign; and finally  $D_a^{\text{NH}}$  is about 3.5 times smaller for the C-terminal zinc knuckle of NC relative to the N-terminal one with the former being almost fully rhombic and the latter axially symmetric). These findings are consistent with a recent cryo-EM study of immature HIV-1 particles (29) where no density was seen for the MA and NC domains, presumably due to the high flexibility of the interdomain linkers. Moreover, the values of  $D_a^{\text{NH}}$  and  $\eta$  for MA and the zinc knuckles of NC are independent of the concentration of  $\Delta$ Gag and hence on the ratio of monomer to dimer present in solution, indicating that their alignment is effectively independent of dimerization status (Table S2). This observation, however, is not true of the two domains of CA, where a systematic increase in  $D_a^{\text{NH}}$  is observed for CA<sup>NTD</sup>, and changes in both  $D_a^{\text{NH}}$  and  $\eta$  are observed for CA<sup>CTD</sup>, the site of dimerization. These effects can be accounted for by the long linkers separating MA from CA<sup>NTD</sup> and CA<sup>CTD</sup> from NC. The effect of dimerization on the alignment tensor of CA<sup>NTD</sup> arises from the shorter linker connecting the two domains of CA, such that the dimerization status of CA<sup>CTD</sup> still influences the alignment of CA<sup>NTD</sup>, despite the fact that no intersubunit interactions are observed between CA<sup>NTD</sup> domains in the CA dimer (15).

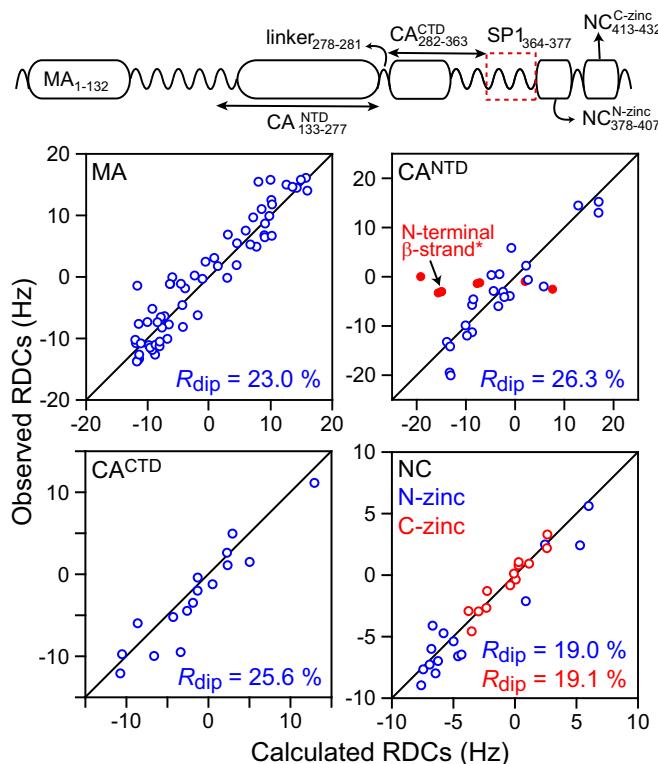
A previous hydrodynamic and neutron scattering study of a monomeric variant of Gag, obtained by a double mutation (W316A and M317A) at the dimerization interface, claimed that MA and NC are close to one another in three-dimensional space (8). The current RDC data, however, are inconsistent with this proposal because the values of  $D_a^{\text{NH}}$  and  $\eta$  for MA and the two zinc knuckles of NC are very different from one another (Table S2).

Technical difficulties precluded us from measuring RDCs on  $\Delta$ Gag samples in the presence of nucleic acid owing to significant interactions between the  $\Delta$ Gag/DNA complex and various alignment media (SI Materials and Methods). However, RDC measurements on isolated NC and the smaller CA-SP1-NC construct indicate that both zinc knuckles of NC lock on to DNA and tumble as a single unit (16). Because the chemical shifts in NC upon binding nucleic acids are the same for  $\Delta$ Gag and the smaller fragments, it seems likely that the same is true of  $\Delta$ Gag.

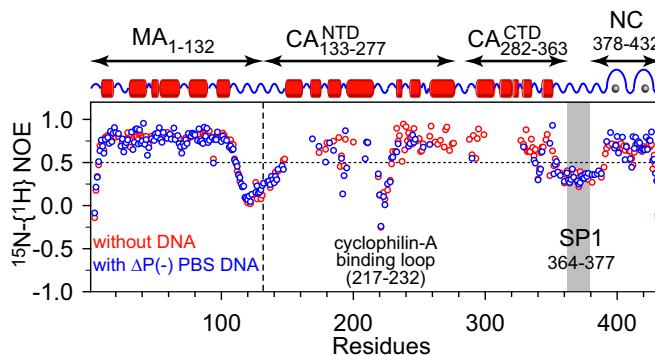
Although the overall folds of the structured domains are preserved in intact  $\Delta$ Gag, there are some important local differences with respect to their isolated counterparts. Upon cleavage at the MA/CA junction, the first 13 N-terminal residues of CA (residues 133–145) fold into a  $\beta$ -hairpin (30), disruption of which leads to noninfectious virions with aberrant morphologies (31–33). The  ${}^1D_{\text{NH}}$  RDC values for these hairpin residues calculated from the X-ray coordinates of CA<sup>NTD</sup> (24), using the alignment tensor obtained from the SVD fits for the remaining CA<sup>NTD</sup>, however, do not match with their experimentally observed counterparts in  $\Delta$ Gag (Fig. 3, filled-in red circles Upper Right). Further, the experimental  ${}^1D_{\text{NH}}$  RDC values for the hairpin residues are close to zero, indicating that they are unstructured in  $\Delta$ Gag.

The linker regions in  $\Delta$ Gag exhibit low  ${}^{15}\text{N}-{}^1\text{H}$  heteronuclear NOE values, close to zero  ${}^1D_{\text{NH}}$  RDC values, and nearly random coil chemical shifts (Fig. 4, Table S3, and Fig. S2), indicative of intrinsic disorder and high mobility such that the ordered structural domains are like beads on a string. These regions include the large stretch of residues connecting MA to CA (residues 110–148), the C-terminal tail of CA (residues 353–363) and the SP1 region that connects CA to the N-terminal zinc-knuckle of NC (residues 364–377), and the linker (residues 408–412, especially residues 411 and 412) that connects the N- and C-terminal zinc knuckles of NC. Overall, these flexible, solvent exposed regions are primarily responsible for extreme sensitivity of Gag to proteolysis and conformational heterogeneity that precludes crystallization.

Critical among the disordered segments of Gag is the SP1 region (residues 364–377) that links CA and NC. With the



**Fig. 3.** Backbone RDC analysis of  $\Delta$ Gag. (Upper) A schematic representation of the overall domain organization of  $\Delta$ Gag. (Lower) SVD analysis showing agreement of the experimental backbone amide ( ${}^1D_{\text{NH}}$ ) RDCs acquired in 5% PEG-hexanol with those calculated from the coordinates of isolated MA (PDB ID 4JMU; ref. 23), the CA<sup>NTD</sup> and CA<sup>CTD</sup> domains of CA (PDB ID 3NTE; ref. 24), and the N- and C-terminal zinc knuckles of NC (PDB ID 2M3Z; ref. 25). The RDC *R*-factor,  $R_{\text{dip}}$ , is given by  $\langle (D_{\text{obs}} - D_{\text{calc}})^2 \rangle / (2 \langle D_{\text{obs}}^2 \rangle)^{1/2}$ , where  $D_{\text{obs}}$  and  $D_{\text{calc}}$  are the observed and calculated  ${}^1D_{\text{NH}}$  RDC values, respectively (26). Only residues in secondary structure elements are used for the SVD fits. The lack of agreement between the experimental RDCs for the first 13 N-terminal residues of CA (residues 133–145) in intact  $\Delta$ Gag and those predicted from the crystal structure of isolated CA (24) using the alignment tensor obtained from the SVD fits to the rest of CA<sup>NTD</sup> is depicted by the red filled-in circles in the upper right-hand panel.



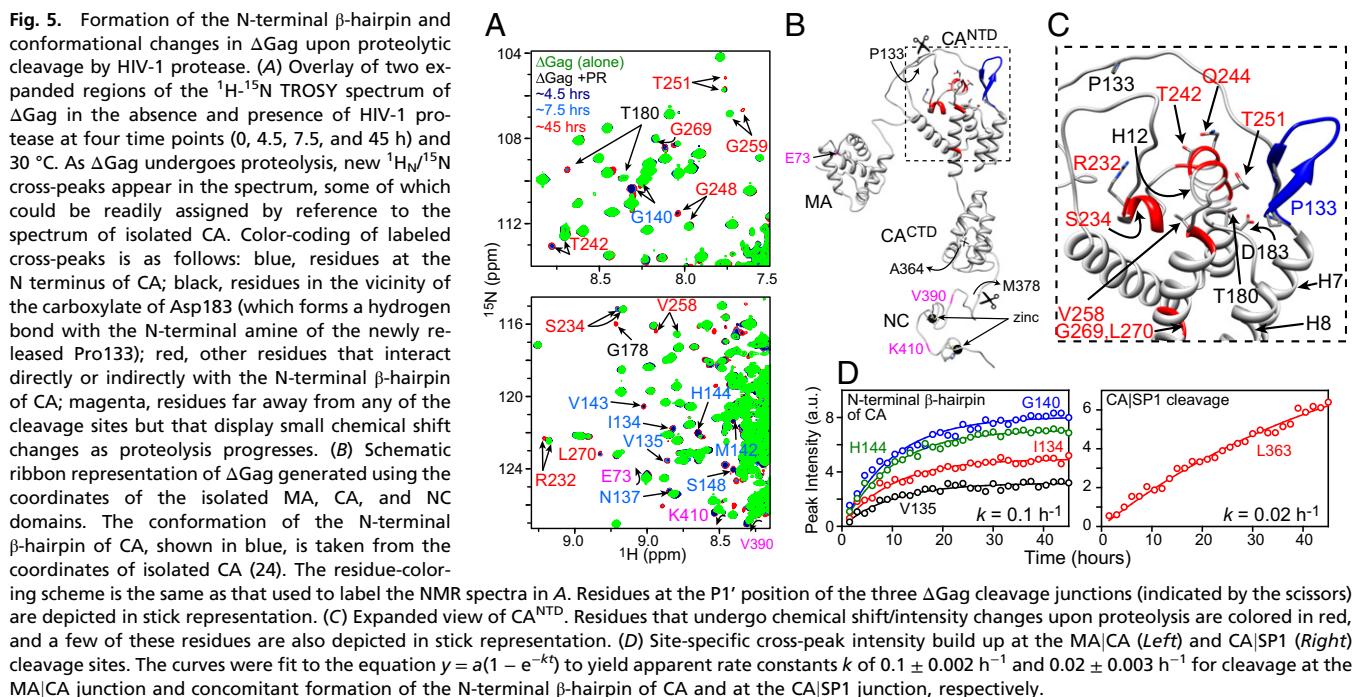
**Fig. 4.**  $^{15}\text{N}-\{^1\text{H}\}$  heteronuclear NOE data for  $\Delta\text{Gag}$  in the absence and presence of  $\Delta\text{P}(-)\text{PBS}$  DNA. Secondary structure elements and domain organization are indicated above the panel. The semitransparent gray rectangle marks the location of SP1 (residues 364–377). The MA|CA proteolytic cleavage site is indicated by a vertical dashed line. The experimental data were recorded at 30 °C and a spectrometer  $^1\text{H}$  frequency of 800 MHz.

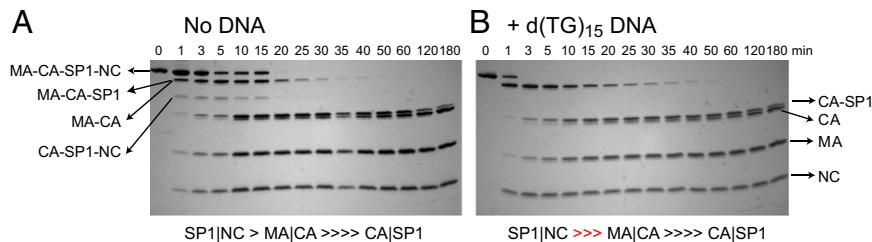
exception of Met367, the backbone  $^{13}\text{Ca}$ ,  $^{13}\text{C}\beta$  and  $^{13}\text{C}'$  chemical shifts of all SP1 residues are close to their random coil values (Fig. S2). For Met367, however, the  $^{13}\text{Ca}$  and  $^{13}\text{C}\beta$  chemical shifts are downfield- and upfield-shifted respectively, suggesting a possible kink in the polypeptide chain at this location. In the smaller CA-SP1-NC construct (16), derived from strain pLN4-3, residues 368–372 of SP1 were not visible in the  $^1\text{H}-^{15}\text{N}$  TROSY spectrum. In contrast, barring Ser368, the  $^1\text{H}_\text{N}/^{15}\text{N}$  cross-peaks for all residues of SP1 are observed in the spectrum of  $\Delta\text{Gag}$ , derived from strain HXB2. This finding can be attributed to the nature of residue 373, proline versus serine in strains pLN4-3 and HXB2, respectively (Los Alamos database; ref. 34), resulting in a slower motional time scale and subsequent resonance line broadening for the former.

It has been proposed that SP1 has helical propensity (9), and low-resolution cryo-electron microscopy has suggested that SP1 forms an intermolecular six-helix bundle in immature HIV-1

capsid (29, 35), although the resolution is insufficient to permit any definitive conclusions in this regard. An unstructured SP1, on the other hand, is supported by solution NMR studies on  $\Delta\text{Gag}$  (this work) and the smaller CA-SP1-NC (16) and CA<sup>CTD</sup>-SP1-NC (36) fragments, and by solid-state NMR studies on tubular assemblies formed by the shorter CA-SP1 construct (37). Understanding the conformational transition of SP1 from an unstructured motif in monomeric/dimeric  $\Delta\text{Gag}$  to a helical conformation in an assembled Gag lattice, if this does indeed occur, will require atomic resolution structures of immature Gag assemblies. It is interesting to note that a six helix bundle of SP1 has been proposed as the binding site for a new class of anti-HIV drugs (Berivimat and related compounds) known as maturation inhibitors (38). No indication for any interaction between Berivimat (0.3 mM) and  $\Delta\text{Gag}$  (protein:drug molar ratio of 1:6) was observed by NMR under conditions where  $\Delta\text{Gag}$  exists in a monomer/dimer equilibrium (*SI Materials and Methods*). The poor solubility of Berivimat precluded the use of higher concentrations that may have pushed the equilibrium in favor of a potential complex; however, the existence of an intrinsically disordered SP1 may explain these negative results *in vitro*.

**The N-Terminal  $\beta$ -Hairpin of CA.** It has been suggested that the N-terminal  $\beta$ -hairpin of CA destabilizes the immature lattice and facilitates the formation of mature capsid (39, 40). To explore the formation of the N-terminal  $\beta$ -hairpin of CA in real-time and to probe other conformational changes that take place in  $\Delta\text{Gag}$  upon ordered proteolytic cleavage by HIV-1 protease,  $^1\text{H}-^{15}\text{N}$  TROSY spectra of  $\Delta\text{Gag}$  were recorded at a series of time points in the presence of a minute amount of HIV-1 protease (Fig. 5) under conditions that do not support Gag assembly into hexamers or higher order assemblies. In intact  $\Delta\text{Gag}$ , the N-terminal residues of CA (residues 133–145) are intrinsically disordered and exhibit narrow chemical shift dispersion (see above and Fig. S2). As proteolysis of  $\Delta\text{Gag}$  proceeds, new  $^1\text{H}_\text{N}/^{15}\text{N}$  cross-peaks appear in the spectrum and grow in intensity with time (Fig. 5A). A few of these can be readily assigned by reference to the spectrum of isolated CA. Among these Ile134, Val135, Asn137, Gly140, Met142, Val143, His144, and Gln145 are directly





**Fig. 6.** Cleavage of  $\Delta$ Gag by HIV-1 protease in the absence (*A*) and presence (*B*) of  $d(TG)_{15}$  DNA.  $\Delta$ Gag was incubated with HIV-1 protease (molar ratio ~50:1, pH 6.5) with and without  $d(TG)_{15}$  DNA (protein:DNA molar ratio of 1:2) for 3 h. at room temperature. Aliquots were taken at regular time intervals and visualized by Page Blue staining (18% Tris-glycine gel). The cleavage products of the reaction, MA-CA-SP1, MA-CA, MA, CA-SP1, CA, and NC, were also verified by liquid chromatography-mass spectrometry. Buffer conditions were 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM  $ZnCl_2$  and 1 mM TCEP.

involved in the formation of the  $\beta$ -hairpin (residues 133–145) and exhibit completely different chemical shifts (labeled in blue in Fig. 5*A*) compared with their unstructured counterparts. Judging by the similarity of the  $^{1}H_N/^{15}N$  chemical shifts for the hairpin residues in isolated CA and those generated upon proteolytic processing of  $\Delta$ Gag, we conclude that the newly formed hairpin is relatively mobile in solution similar to that in CA (15) and CA-SP1-NC (16). The  $\beta$ -hairpin is primarily stabilized by a hydrogen bond between the N-terminal amine of the highly conserved Pro133 and the carboxylate of Asp183 (30). Although the resulting chemical shift changes in the  $^{1}H_N/^{15}N$  cross-peak of Asp183 could not be identified due to cross-peak overlap, neighboring residues, in particular Thr180 and Gly178, have well-resolved  $^{1}H_N/^{15}N$  cross-peaks, and exhibit chemical shift/intensity changes upon  $\beta$ -hairpin formation due to changes in their local environment (labeled in black in Fig. 5*A*). A few other residues within CA<sup>NTD</sup> (Arg232, Ser234, Asp235, Ile236, Thr242, Gln244, Glu245, Ile247, Gly248, Thr251, Val258, Gly269, and Leu270) also show chemical shift/intensity changes (depicted in red in Fig. 5*A–C*). Among these, residues located in helix 12 (residues 242–251, corresponding to helix 6 in isolated CA) directly interact with the  $\beta$ -hairpin (Fig. 5*C*). Although several other regions within CA<sup>NTD</sup> display chemical shift/intensity changes, these are likely due to subtle side-chain repacking as the RDC data (Fig. 3 and Table S2) indicate that any conformational changes in the rest of CA<sup>NTD</sup> are negligible.

$\Delta$ Gag also undergoes cleavage at the CA|SP1 and SP1|NC junctions. Among the products of CA|SP1 cleavage, Leu363 within CA<sup>CTD</sup> is easily identified in the  $^{1}H-^{15}N$  TROSY spectrum. The build-up of the Leu363  $^{1}H_N/^{15}N$  cross-peak, however, is much slower than that of the N-terminal  $\beta$ -hairpin residues of CA<sup>NTD</sup> (Fig. 5*D*).  $^{1}H_N/^{15}N$  cross-peaks belonging to the SP1|NC cleavage products could not be identified.

The site-specific build-up of cross-peak intensities upon proteolytic processing by HIV-1 protease can be fit to a simple rate equation yielding apparent rate constants of 0.1 and 0.02  $h^{-1}$  for MA|CA and CA|SP1 cleavage, respectively. Thus, formation of the  $\beta$ -hairpin upon  $\Delta$ Gag proteolysis is five times faster in solution than that of cleavage at the CA|SP1 site. These data contrast with those reported in an earlier hydrogen/deuterium exchange study on immature and mature virus-like HIV-1 particles using mass spectrometry, which claimed that the  $\beta$ -hairpin does not form immediately upon MA|CA cleavage but instead requires proteolytic processing at the downstream CA|SP1 junction (41). However, given that the  $\beta$ -hairpin is very distant from the CA|SP1 junction, there does not appear to be any structural or physical basis whereby cleavage at the CA|SP1 junction could influence the rate of  $\beta$ -hairpin formation upon MA|CA cleavage.

**ΔGag-Protease Interactions.** The five cleavage sites targeted by HIV-1 protease differ considerably in their primary sequence, making it difficult to draw conclusions regarding the sequence/

conformational preferences of HIV-1 protease. However, the presence of numerous amino acids of varying properties in and around the cleavage sites suggests a possible explanation. It has been predicted that HIV-1 protease interacts with an extended peptide segment, requiring 6–8 residues to constitute an effective substrate (42). These predictions are supported by crystallographic studies in which synthetic peptides mimicking Gag substrates are bound to HIV-1 protease in an extended  $\beta$ -strand conformation, suggesting that the protease recognizes the shape of the substrate rather than its primary sequence (43). This “substrate envelope” hypothesis predicts that interactions between substrate side chains and the protease vary among the cleavage sites, resulting in a unique processing rate for each site (43). At neutral pH, the proposed order of Gag proteolysis is SP1|NC > SP2|p6 ~ MA|CA > CA|SP1 ~ NC|SP2 (44) with a ~400 fold difference in cleavage rates at the SP1|NC and CA|SP1 sites. To further explore  $\Delta$ Gag-protease interactions,  $\Delta$ Gag was incubated with HIV-1 protease (at a 50:1  $\Delta$ Gag:protease molar ratio), the cleavage products monitored using SDS/PAGE (Fig. 6), and bands identified by liquid chromatography-mass spectrometry. Based on band intensities,  $\Delta$ Gag in the absence of nucleic acids is cleaved in the following order: SP1|NC > MA|CA > CA|SP1. Addition of  $d(TG)_{15}$  DNA to  $\Delta$ Gag greatly accelerates the cleavage rate at the SP1|NC site, but has no affect on the MA|CA and CA|SP1 cleavage rates. Because the presence of DNA per se does not affect the catalytic activity of the protease and chemical shift perturbation data (Fig. 2) provide no evidence for any interaction between SP1 and  $d(TG)_{15}$  DNA, we suggest that proteolysis at the SP1|NC junction is accelerated due to changes in its availability for enzymatic processing, making substrate accessibility rather than substrate conformation a limiting factor in Gag-protease interactions.

**Concluding Remarks.** In summary, HIV-1  $\Delta$ Gag exists in a dynamic monomer-dimer equilibrium with no evidence for a  $\Delta$ Gag trimer, both in the absence and presence of nucleic acids and inositol phosphate derivatives.  $\Delta$ Gag is a physiologically relevant system because it assembles into virus-like particles that are similar to immature HIV-1 virions. Both NC and MA domains of  $\Delta$ Gag are involved in interactions with nucleic acids, with the latter being much weaker than the former. All five structured domains of  $\Delta$ Gag (MA, CA<sup>NTD</sup>, CA<sup>CTD</sup>, and the N- and C-terminal zinc knuckles of NC) are folded in solution and reorient semi-independently of one another in the absence of nucleic acids. The stretches of residues that connect the structured domains of  $\Delta$ Gag, including residues within and surrounding the MA|CA junction and the SP1 region that connects the CA and NC domains, are highly dynamic and intrinsically disordered in solution such that the ordered structured domains effectively behave like beads on a string. Formation of the N-terminal  $\beta$ -hairpin of CA upon proteolytic cleavage at the MA|CA junction by HIV-1 protease can be readily monitored in real-time and was found to be five times faster than cleavage at the

CA|SP1 junction. Additionally, although nucleic acids do not interact with SP1, cleavage at the SP1|NC junction by HIV-1 protease is significantly accelerated in the presence of nucleic acids, suggesting that substrate accessibility rather than conformation constitutes the rate limiting step for Gag proteolysis.

## Materials and Methods

**Protein Expression, Purification, and Characterization.** The constructs of HIV-1 ΔGag and protease were subcloned into pET-11a vectors and expressed in BL21-CodonPlus (DE3)-RIPL competent cells (Agilent Technologies). Full details on cloning, expression, isotope ( $^2\text{H}/^{15}\text{N}/^{13}\text{C}$  and  $^2\text{H}/^{15}\text{N}$ ) labeling, purification, and analytical ultracentrifugation experiments used to determine the equilibrium dimerization constants ( $K_{\text{dimer}}$ ) are provided in the *SI Materials and Methods*. Samples for NMR ( $^2\text{H}/^{15}\text{N}/^{13}\text{C}$  and  $^2\text{H}/^{15}\text{N}$  labeled) were prepared in a buffer containing 20 mM sodium phosphate, pH 6.5, 300 mM sodium chloride, 0.1 mM ZnCl<sub>2</sub> and 1 mM Tris(2-carboxyethyl)phosphine (TCEP).

1. Freed EO (1998) HIV-1 gag proteins: Diverse functions in the virus life cycle. *Virology* 251(1):1–15.
2. Sundquist WI, Kräusslich HG (2012) HIV-1 assembly, budding, and maturation. *Cold Spring Harbor Perspect Med* 2(7):a006924.
3. Ganser-Pornillos BK, Yeager M, Pornillos O (2012) Assembly and architecture of HIV. *Adv Exp Med Biol* 726:441–465.
4. Lu K, Heng X, Summers MF (2011) Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol* 410(4):609–633.
5. Bell NM, Lever AM (2013) HIV Gag polyprotein: Processing and early viral particle assembly. *Trends Microbiol* 21(3):136–144.
6. Ganser-Pornillos BK, von Schwedler UK, Stray KM, Aiken C, Sundquist WI (2004) Assembly properties of the human immunodeficiency virus type 1 CA protein. *J Virol* 78(5):2545–2552.
7. Kutluay SB, Bieniasz PD (2010) Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLoS Pathog* 6(11):e1001200.
8. Datta SA, et al. (2007) Conformation of the HIV-1 Gag protein in solution. *J Mol Biol* 365(3):812–824.
9. Datta SA, et al. (2011) HIV-1 Gag extension: Conformational changes require simultaneous interaction with membrane and nucleic acid. *J Mol Biol* 406(2):205–214.
10. Fisher RJ, et al. (1998) Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J Virol* 72(3):1902–1909.
11. Mirambeau G, Lyonnais S, Goretzky RJ (2010) Features, processing states, and heterogeneous protein interactions in the modulation of the retroviral nucleocapsid protein function. *RNA Biol* 7(6):724–734.
12. Chukkapalli V, Oh SJ, Ono A (2010) Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proc Natl Acad Sci USA* 107(4):1600–1605.
13. Alfaadhi A, Barklis E (2014) The roles of lipids and nucleic acids in HIV-1 assembly. *Front Microbiol* 5:253.
14. Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94(23):12366–12371.
15. Deshmukh L, et al. (2013) Structure and dynamics of full-length HIV-1 capsid protein in solution. *J Am Chem Soc* 135(43):16133–16147.
16. Deshmukh L, Ghirlando R, Clore GM (2014) Investigation of the structure and dynamics of the capsid-spacer peptide 1-nucleocapsid fragment of the HIV-1 gag polyprotein by solution NMR spectroscopy. *Angew Chem Int Ed Engl* 53(4):1025–1028.
17. Cai M, Huang Y, Craigie R, Clore GM (2010) Structural basis of the association of HIV-1 matrix protein with DNA. *PLoS ONE* 5(12):e15675.
18. Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI (1996) Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: Implications for membrane association and assembly. *Proc Natl Acad Sci USA* 93(7):3099–3104.
19. Saad JS, et al. (2006) Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proc Natl Acad Sci USA* 103(30):11364–11369.
20. Kutluay SB, et al. (2014) Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell* 159(5):1096–1109.
21. Bax A, Grishaev A (2005) Weak alignment NMR: A hawk-eyed view of biomolecular structure. *Curr Opin Struct Biol* 15(5):563–570.
22. Rückert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* 122(32):7793–7797.
23. LaPlante SR, et al. (2014) Enantiomeric atropisomers inhibit HCV polymerase and/or HIV matrix: Characterizing hindered bond rotations and target selectivity. *J Med Chem* 57(5):1944–1951.
24. Du S, et al. (2011) Structure of the HIV-1 full-length capsid protein in a conformationally trapped unassembled state induced by small-molecule binding. *J Mol Biol* 406(3):371–386.
25. Goudreau N, et al. (2013) Discovery and structural characterization of a new inhibitor series of HIV-1 nucleocapsid function: NMR solution structure determination of a ternary complex involving a 2:1 inhibitor/NC stoichiometry. *J Mol Biol* 425(11):1982–1998.
26. Clore GM, Garrett DS (1999) R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures. *J Am Chem Soc* 121:9008–9012.
27. Zweckstetter M, Bax A (2002) Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *J Biomol NMR* 23(2):127–137.
28. Braddock DT, Cai M, Baber JL, Huang Y, Clore GM (2001) Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. *J Am Chem Soc* 123(35):8634–8635.
29. Schur FK, et al. (2015) Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature* 517(7535):505–508.
30. Gitti RK, et al. (1996) Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science* 273(5272):231–235.
31. von Schwedler UK, et al. (1998) Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *EMBO J* 17(6):1555–1568.
32. Tang C, Ndassa Y, Summers MF (2002) Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nat Struct Biol* 9(7):537–543.
33. Kelly BN, et al. (2006) Implications for viral capsid assembly from crystal structures of HIV-1 Gag(1–278) and CA(N)(133–278). *Biochemistry* 45(38):11257–11266.
34. Ratner L, et al. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 313(6000):277–284.
35. Bharat TA, et al. (2012) Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy. *Nature* 487(7407):385–389.
36. Newman JL, Butcher EW, Patel DT, Mikhaylenko Y, Summers MF (2004) Flexibility in the P2 domain of the HIV-1 Gag polyprotein. *Protein Sci* 13(8):2101–2107.
37. Han Y, et al. (2013) Magic angle spinning NMR reveals sequence-dependent structural plasticity, dynamics, and the spacer peptide 1 conformation in HIV-1 capsid protein assemblies. *J Am Chem Soc* 135(47):17793–17803.
38. Nguyen AT, et al. (2011) The prototype HIV-1 maturation inhibitor, bevirimat, binds to the CA-SP1 cleavage site in immature Gag particles. *Retrovirology* 8:101.
39. Fitzton T, et al. (2000) Proline residues in the HIV-1 NH2-terminal capsid domain: Structure determinants for proper core assembly and subsequent steps of early replication. *Virology* 268(2):294–307.
40. Abdurrahman S, Youssefi M, Höglund S, Vahlne A (2007) Characterization of the invariable residue 51 mutations of human immunodeficiency virus type 1 capsid protein on *in vitro* CA assembly and infectivity. *Retrovirology* 4:69.
41. Bush DL, et al. (2014) Higher-order structure of the Rous sarcoma virus SP assembly domain. *J Virol* 88(10):5617–5629.
42. Chou KC (1996) Prediction of human immunodeficiency virus protease cleavage sites in proteins. *Anal Biochem* 233(1):1–14.
43. Ozen A, Haliloglu T, Schiffer CA (2011) Dynamics of preferential substrate recognition in HIV-1 protease: Redefining the substrate envelope. *J Mol Biol* 410(4):726–744.
44. Tritch RJ, Cheng YE, Yin FH, Erickson-Viitanen S (1991) Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 gag polyprotein. *J Virol* 65(2):922–930.
45. Clore GM, Gronenborn AM (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol* 16(1):22–34.
46. Fitzkee NC, Bax A (2010) Facile measurement of  $^1\text{H}/^{15}\text{N}$  residual dipolar couplings in larger perdeuterated proteins. *J Biomol NMR* 48(2):65–70.
47. Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Mag Res Sp* 48(1):47–62.

# Supporting Information

Deshmukh et al. 10.1073/pnas.1501985112

## SI Materials and Methods

**Materials.** The 14mer DNA (–) Primer binding site  $\Delta P(-)$ PBS (5'-dTGCCTGTTGGC) (1) and the alternating TG motif 30mer, 5'-d(TG)<sub>15</sub>, were purchased from Integrated DNA Technologies, Inc. Oligonucleotides were dissolved in deionized water, and dialyzed overnight (Spectra/Pro DispoDialyzer, 500 Da cutoff) in a buffer containing 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM ZnCl<sub>2</sub>, and 1 mM (Tris(2-carboxyethyl)phosphine) (TCEP). The HIV-1 protease inhibitor Darunavir and the maturation inhibitor, Bevirimat Dimeglumine, were obtained from the NIH AIDS Reagent Program. Inositol hexakisphosphate was purchased from Calbiochem (catalog no. 407125).

**Protein Expression and Purification.** HIV-1  $\Delta$ Gag (MA-CA-SP1-NC, residues 1–432, strain HXB2), and HIV-1 protease (residues 1–99 with point mutations Q7K, L33I, L63I, C67A, and C95A, to prevent auto-proteolysis and thiol oxidation, strain HXB2) were subcloned in pET-11a vectors and expressed in BL21-CodonPlus (DE3)-RIPL competent cells (Agilent Technologies).  $\Delta$ Gag, either at natural isotopic abundance or uniformly <sup>2</sup>H/<sup>15</sup>N/<sup>13</sup>C- or <sup>2</sup>H/<sup>15</sup>N-labeled, was expressed at 19 °C using our published protocol (2). Briefly, cells were grown at 37 °C in 1 L Luria-Bertani (LB) medium or minimal M9 medium containing 0.3 g/L <sup>2</sup>H/<sup>15</sup>N/<sup>13</sup>C Isogro (Sigma-Aldrich), <sup>2</sup>H<sub>2</sub>O, 1 g/L <sup>15</sup>NH<sub>4</sub>Cl and 3 g/L <sup>2</sup>H<sub>7</sub>/<sup>13</sup>C<sub>6</sub>-D-glucose for <sup>2</sup>H/<sup>15</sup>N/<sup>13</sup>C-labeling; and 0.3 g/L <sup>2</sup>H/<sup>15</sup>N Isogro (Sigma-Aldrich), <sup>2</sup>H<sub>2</sub>O, 1 g/L <sup>15</sup>NH<sub>4</sub>Cl and 3 g/L <sup>2</sup>H<sub>7</sub>/<sup>12</sup>C<sub>6</sub>-D-glucose for <sup>2</sup>H/<sup>15</sup>N-labeling. 30 min. before induction, the temperature was reduced to 19 °C. Cells, induced with 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) at an optical density of A<sub>600</sub> ~0.8, were harvested 24 h later. For HIV-1 protease expression, cells were grown at 37 °C in 1 L LB medium, and were harvested 3 h after induction with IPTG.

In the case of  $\Delta$ Gag, cells were resuspended in a lysis buffer containing 20 mM Tris, pH 8.0, 500 mM NaCl, 0.1 mM ZnCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol (BME), and 1 cOmplete Protease Inhibitor Mixture tablet (Roche Applied Science).  $\Delta$ Gag was purified by a combination of ion exchange and size exclusion chromatography. The cell lysate was loaded onto a HiPrep 16/10 Q FF column (GE Healthcare) with a 0.5–1 M NaCl gradient in a buffer containing 20 mM Tris, pH 8.0, 500 mM NaCl, 0.1 mM ZnCl<sub>2</sub> and 5 mM BME. Relevant flow-through fractions were diluted in a buffer containing 20 mM Tris, pH 8.0, 0.1 mM ZnCl<sub>2</sub> and 5 mM BME (1:1 dilution) and loaded onto a HiLoad 16/10 SP Sepharose HP column (GE Healthcare) with a 0–1 M NaCl gradient. The eluted protein was concentrated (Amicon ultra-15, 30 kDa cutoff) and loaded onto a HiLoad 26/60 Superdex 200 column (GE Healthcare) pre-equilibrated with 20 mM Tris, pH 8, 500 mM NaCl, 0.1 mM ZnCl<sub>2</sub>, and 5 mM BME. Relevant fractions were diluted in a buffer containing 20 mM Tris, pH 8.0, 0.1 mM ZnCl<sub>2</sub> and 5 mM BME (1:1 dilution), and  $\Delta$ Gag was further purified using a Mono S 10/100 GL column (GE Healthcare) with a 0–1 M NaCl gradient. This purification scheme resulted in a yield of ~10 mg/L (<sup>2</sup>H/<sup>15</sup>N labeled  $\Delta$ Gag).

HIV-1 protease samples (at natural isotopic abundance) were purified as described previously (2). Briefly, cells were resuspended in a buffer containing 50 mM Tris, pH 8.0 (buffer A), sonicated and centrifuged. Inclusion bodies were first washed with buffer A, then with buffer A containing 0.1% Triton X-100, 1 M NaCl and 1 M urea, and finally with buffer A alone. Purified inclusion bodies were then resuspended in 8 M urea and purified using a combination of anion/cation exchange chromatography under denaturing conditions. The eluted protein from a HiPrep

16/10 SP FF column (GE Healthcare) was refolded in buffer containing 20 mM sodium phosphate, pH 6.5, 50 mM NaCl, 5 mM BME, and 10% glycerol.

All constructs were verified by DNA sequencing and mass spectrometry (using an Agilent 1100 LC/MS system equipped with an Agilent Zorbax 300SB-C3 column coupled to a quadrupole mass analyzer). Note that the initiator methionine of  $\Delta$ Gag is lost during bacterial recombinant expression (expected mass: 48,183 Da, experimental mass: 48,184.86 Da).

**Analytical Ultracentrifugation.** Solutions of  $\Delta$ Gag in 20 mM sodium phosphate (pH 6.5), 300 mM NaCl, 0.1 mM ZnCl<sub>2</sub> and 1 mM TCEP at nominal (monomer) concentrations of 31, 15.0, 7.5 and 3.8  $\mu$ M were prepared by dilution of a stock solution (0.36–0.46 mM). High concentration samples were loaded into 3 mm 2-channel epon centerpiece cells (100  $\mu$ L), whereas low concentration samples were loaded into 12 mm 2-channel epon centerpiece cells (400  $\mu$ L). Sedimentation velocity experiments were conducted at 50,000 rpm and 30 °C on a Beckman Coulter ProteomeLab XL-I analytical ultracentrifuge using both absorbance (280 nm) and Rayleigh interference (655 nm) optical detection systems. Time-corrected data (3) were initially analyzed in SEDFIT 14.4f (4) in terms a continuous  $c(s)$  distribution covering an  $s$  range of 0.0–6.0 S with a resolution of 120 and a confidence level (F-ratio) of 0.68. Excellent fits were obtained and  $c(s)$  profiles were consistent with a reversible monomer-dimer self-association. Data were exported into SEDPHAT 12.01beta (5) for further analysis. The solution density ( $\rho$ ) and viscosity ( $\eta$ ) were measured experimentally at 20 °C on an Anton Paar DMA 5000 density meter and an Anton Paar AMVn rolling ball viscometer, respectively, and corrected for temperature. Partial specific volumes ( $v$ ) of  $\Delta$ Gag at 20 and 30 °C were calculated based on the amino acid composition using SEDNTERP 1.09 (6). Global absorbance and Rayleigh interference data were analyzed in terms of a reversible monomer-dimer self-association by direct Lamm equation modeling (7) in SEDPHAT 12.01beta to obtain the equilibrium constant. Excellent fits were obtained with r.m.s.d. values ranging from 0.004 to 0.012 fringes or 0.003–0.005 absorbance units. Errors in the best-fit equilibrium constants were determined using the method of F-statistics. The protein extinction coefficient at 280 nm and interference signal increment (8) was calculated based on the amino acid composition in SEDNTERP 1.09 and SEDFIT 14.4f, respectively.

To investigate the effects of nucleic acids, samples of  $\Delta$ Gag were treated with two equivalents of  $\Delta P(-)$ PBS or d(TG)<sub>15</sub> DNA (2). Mixtures with  $\Delta P(-)$ PBS DNA were studied at loading concentrations of 29, 14.0, 7.0 and 3.5  $\mu$ M, whereas mixtures with d(TG)<sub>15</sub> DNA were studied at concentrations of 27, 13.1, 6.6 and 3.2  $\mu$ M. Interference data were collected and analyzed as above, except that the self-associating species is now the 1:1 DNA: $\Delta$ Gag complex. Excess free DNA was accounted for in the model, for which we assumed a partial specific volume of 0.55 cm<sup>3</sup>g<sup>-1</sup> and a signal increment  $dn/dc$  of 0.185 cm<sup>3</sup>g<sup>-1</sup>. Excellent fits were obtained with r.m.s.d. values ranging from 0.005 to 0.011 fringes. Similar experiments were carried out in the presence of 20 equivalents of inositol hexakisphosphate (IP6). Data were collected at loading concentrations of 31, 24, 8.6, and 4.4  $\mu$ M and excellent fits were obtained to a monomer-dimer self-association Lamm equation model with r.m.s.d. of 0.005–0.008 fringes and 0.003–0.004 absorbance units.

**Transmission Electron Microscopy (TEM).** ΔGag samples, ~1 mg/mL in 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM ZnCl<sub>2</sub> and 1 mM TCEP, were mixed with d(TG)<sub>15</sub> DNA (protein:DNA molar ratio ~5:1) and subsequently dialyzed into 20 mM Tris, pH 8, 100 mM NaCl, 1 mM EDTA for 2 h at 4 °C. After dialysis, samples were centrifuged in an Eppendorf (Model 5415R) bench-top centrifuge for 5 min at 15,700 × g, and resuspended in the same volume of the dialysis buffer. Samples (3 μL) were applied to the TEM grids (400-mesh formvar and carbon coated copper, Electron Microscopy Sciences, catalog no. FCF400-Cu). 1 min after deposition, the sample solution was wicked with filter paper, followed by a quick wash with 3 μL of water and addition of 3–5 drops of 1% (wt/vol) aqueous uranyl acetate solution. The uranyl acetate was wicked immediately with filter paper and grids were air dried at room temperature. TEM was carried out using a JEOL JEM 1200EX transmission electron microscope (accelerating voltage 80 keV) equipped with an AMT XR-60 digital camera.

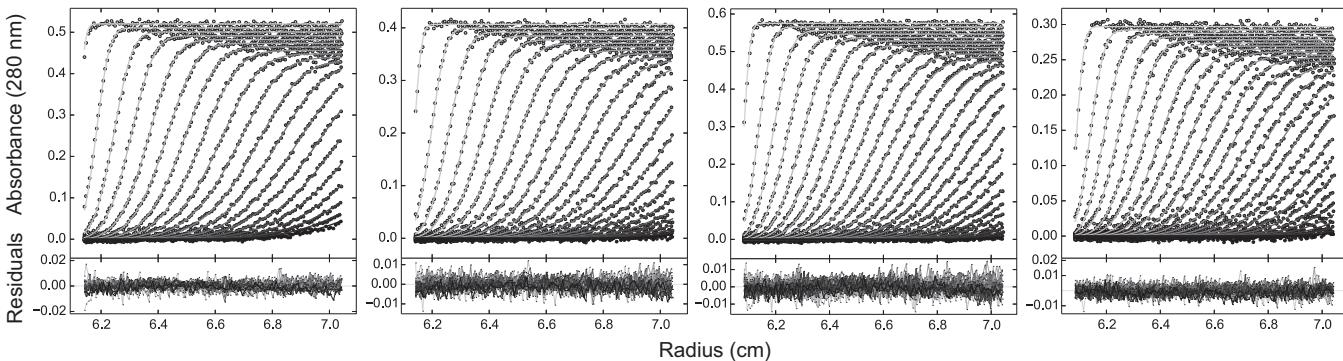
**NMR Sample Preparation.** All heteronuclear NMR experiments were performed on uniformly <sup>15</sup>N/<sup>13</sup>C/<sup>2</sup>H or <sup>15</sup>N/<sup>2</sup>H labeled (as appropriate for the particular NMR experiment) ΔGag samples prepared in buffer containing 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM ZnCl<sub>2</sub>, 93% H<sub>2</sub>O/7% D<sub>2</sub>O (vol/vol), and 1 mM TCEP. To negate the effect of high salt and to achieve higher signal-to-noise ratios and lower proton pulse widths, shaped NMR tubes were used for data acquisition (Bruker Biospin, catalog no. Z106898). Aligned samples were prepared using 5% PEG-hexanol (9) in the absence of nucleic acids. Backbone amide (<sup>1</sup>D<sub>NH</sub>) RDC data were measured on 0.2, 0.15, and 0.1 mM (in subunits) ΔGag samples. Aligned samples could not be made in the presence of nucleic acids, as the ΔGag-nucleic acid complex interacted with various alignment media (including PEG/hexanol, lipid bicelles, and bacteriophage pf1). For heteronuclear <sup>15</sup>N-{<sup>1</sup>H} NOE measurements, a protein concentration of ~0.3–0.35 mM in subunits was used. For NMR titration experiments between ΔGag and DNA, the concentration of ΔGag was 50 μM in subunits. For chemical shift perturbation experiments between Bevirimat and ΔGag, Bevirimat Dimeglumine

was dissolved in deuterated DMSO and added to a 50 μM ΔGag NMR sample (at a protein:drug molar ratio of 1:6). No changes in chemical shifts were observed for the ΔGag + Bevirimat samples, although samples did show signs of precipitation. For real-time NMR measurements to monitor the formation of the N-terminal β-hairpin of capsid upon cleavage by HIV-1 protease, a 0.2 mM <sup>15</sup>N/<sup>2</sup>H-labeled ΔGag sample was mixed with 0.033 μM HIV-1 protease, and a series of <sup>1</sup>H-<sup>15</sup>N TROSY correlation spectra were recorded at intervals of 1.5 h.

**NMR Spectroscopy.** All heteronuclear NMR experiments were carried out at 30 °C on Bruker 500 and 800 MHz spectrometers equipped with z-gradient triple resonance cryoprobes. Spectra were processed using NMRPipe (10) and analyzed using the CCPN software suite (11). Sequential <sup>1</sup>H, <sup>15</sup>N, and <sup>13</sup>C backbone resonance assignments were performed using conventional TROSY-based through-bond 3D triple resonance experiments (12). Backbone amide (<sup>1</sup>D<sub>NH</sub>) RDCs were measured on <sup>2</sup>H/<sup>15</sup>N-labeled proteins using the TROSY-based ARTSY technique (13) and analyzed with the program Xplor-NIH (14). Heteronuclear <sup>15</sup>N-{<sup>1</sup>H} NOE measurements were carried out on uniformly <sup>2</sup>H/<sup>15</sup>N-labeled ΔGag at a <sup>1</sup>H frequency of 800 MHz. The <sup>15</sup>N-{<sup>1</sup>H} NOE and reference spectra were recorded with a 10-s saturation time for the NOE measurement and equivalent recovery time for the reference measurement in an interleaved manner, each preceded by an additional 1-s recovery time.

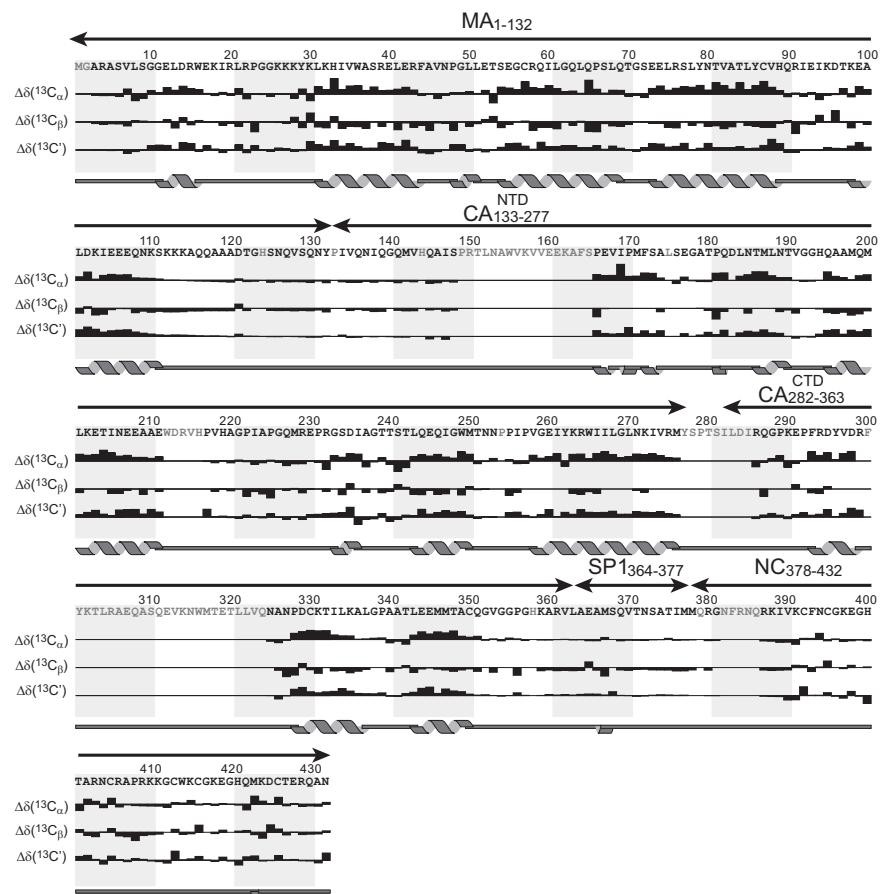
**ΔGag-Protease Interaction.** 50 μM (in subunits) ΔGag was mixed with 1 μM (in subunits) active HIV-1 protease in a buffer containing 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM ZnCl<sub>2</sub>, and 1 mM TCEP. Aliquots (5 μL each) were taken at regular time intervals, mixed with 1× SDS loading solution (Quality Biological, catalog no. 351-082-661) and loaded onto a SDS/PAGE gel (18% Tris-glycine gel, Life technologies, catalog no. EC65055). Aliquots were also collected for LC-MS analysis, and the reaction was terminated immediately upon sample collection by addition of the HIV-1 protease inhibitor Darunavir.

1. Bourgibot S, et al. (2008) How the HIV-1 nucleocapsid protein binds and destabilizes the (-) primer binding site during reverse transcription. *J Mol Biol* 383(5):1112–1128.
2. Deshmukh L, Ghirlando R, Clore GM (2014) Investigation of the structure and dynamics of the capsid-spacer peptide 1-nucleocapsid fragment of the HIV-1 gag polyprotein by solution NMR spectroscopy. *Angew Chem Int Ed Engl* 53(4):1025–1028.
3. Zhao H, et al. (2013) Recorded scan times can limit the accuracy of sedimentation coefficients in analytical ultracentrifugation. *Anal Biochem* 437(1):104–108.
4. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78(3):1606–1619.
5. Schuck P (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal Biochem* 320(1):104–124.
6. Cole JL, Lary JV, Moody TP, Laue TM (2008) Analytical ultracentrifugation: Sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol* 84:143–179.
7. Brautigam CA (2011) Using Lamm-Equation modeling of sedimentation velocity data to determine the kinetic and thermodynamic properties of macromolecular interactions. *Methods* 54(1):4–15.
8. Zhao H, Brown PH, Schuck P (2011) On the distribution of protein refractive index increments. *Biophys J* 100(9):2309–2317.
9. Rückert M, Otting G (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J Am Chem Soc* 122(32):7793–7797.
10. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
11. Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59(4):687–696.
12. Clore GM, Gronenborn AM (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol* 16(1):22–34.
13. Fitzkee NC, Bax A (2010) Facile measurement of <sup>1</sup>H-<sup>15</sup>N residual dipolar couplings in larger perdeuterated proteins. *J Biomol NMR* 48(2):65–70.
14. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160(1):65–73.



**Fig. S1.** Sedimentation velocity data for  $\Delta$ Gag collected at 50,000 rpm and 30 °C in the presence of 20 molar equivalents of inositol hexakisphosphate (IP6). Absorbance data (280 nm) collected at loading concentrations (in subunits) of 30, 24, 8.5, and 4.4  $\mu$ M (shown from left to right) were analyzed in terms of a reversible monomer–dimer self-association using Lamnn equation modeling. Every third data point and every third scan are shown together with the global best-fit (solid line) and corresponding residuals (bottom panel). Samples at 30 and 24  $\mu$ M were loaded in 3 mm path-length cells, whereas samples at 8.5 and 4.4  $\mu$ M were loaded in 12 mm path-length cells. Data were plotted using GUSSI (obtained from [biophysics.swmed.edu/MBR/software.html](http://biophysics.swmed.edu/MBR/software.html)). The combined interference and absorbance sedimentation data were best modeled in terms of a reversible monomer–dimer equilibrium. Introduction of other higher order species (e.g., trimer, hexamer) always leads to an increase in  $\chi^2$ . To determine the maximum measurable amounts of trimer or hexamer, data were also modeled in terms of reversible monomer–dimer–trimer or monomer–dimer–hexamer reversible equilibria. As the best-fit monomer–dimer model leads to sedimentation coefficients that scale as the 2/3 power of the mass (as expected for globular proteins having identical frictional ratios), sedimentation coefficients for the trimer and hexamer were assumed to scale in a similar manner. Relative values of the global reduced  $\chi^2$  were monitored using a cutoff that corresponds to a confidence level of 0.683, as described in equation 5 of Johnson (1). Values of the molar masses and sedimentation coefficients were fixed and data were refined (i.e., float monomer–dimer  $K_{ass}^{12}$  and lifetime of complexes) with different fixed values of the monomer–trimer ( $K_{ass}^{13}$ ) or monomer–hexamer ( $K_{ass}^{16}$ ) equilibrium constants until the  $\chi^2$  cutoff was reached, resulting in a statistically valid upper limit for  $K_{ass}^{13}$  or  $K_{ass}^{16}$ . In this manner an upper bound of  $2.46 \times 10^7 \text{ M}^{-2}$  is obtained for  $K_{ass}^{13}$  resulting in a maximum concentration of 0.10  $\mu$ M trimer at the highest loading concentration of 30  $\mu$ M; similarly, a maximum concentration of 22 nM hexamer was determined with an upper bound of  $1.26 \times 10^{21} \text{ M}^{-5}$  for  $K_{ass}^{16}$ . Any possible contributions from trimer and hexamer formation are thus very small (<1% in terms of subunit concentrations) in the context of the monomer–dimer equilibrium. As noted in the text, simple modeling in terms of a monomer–trimer equilibrium led to statistically worse fits and poor residuals.

1. Johnson ML (1992) Why, when, and how biochemists should use least squares. *Anal Biochem* 206(2):215–225.



**Fig. S2.** Summary of NMR-derived secondary structure backbone chemical shift ( $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$ , and  $^{13}\text{C}'$ ) indicators for  $\Delta\text{Gag}$  in the absence nucleic acids. Missing/unassigned residues are shown in gray. The location of secondary structure elements is derived from the backbone chemical shifts using the program Talos-N (1).

1. Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56(3):227–241.

**Table S1. Equilibrium dimerization constants for HIV-1  $\Delta\text{Gag}$  in the absence and presence of nucleic acids and inositol hexakisphosphate (IP6) determined by analytical ultracentrifugation at 30 °C**

Construct	$K_{\text{dimer}} (\mu\text{M})^*$
$\Delta\text{Gag}_{1-432}$	$35 \pm 5$
+ $\Delta\text{P}(-)$ PBS DNA	$18 \pm 3$
+ $d(\text{TG})_{15}$ DNA	$14 \pm 2$
+ IP6	$35 \pm 5$

\*Buffer conditions: 20 mM sodium phosphate (pH 6.5), 300 mM NaCl, 0.1 mM ZnCl<sub>2</sub> and 1 mM TCEP. The molar ratios of  $\Delta\text{Gag}$ :DNA and  $\Delta\text{Gag}$ :IP6 were 1:2 and 1:20, respectively.

**Table S2. Magnitude of RDC alignment tensors for the structural domains of ΔGag**

Concentration of ΔGag (in subunits)

Domain*	100 μM			150 μM			200 μM		
	$D_a^{NH}$ (Hz)	$\eta^\dagger$	$R_{dip}$ (%)	$D_a^{NH}$ (Hz)	$\eta^\dagger$	$R_{dip}$ (%)	$D_a^{NH}$ (Hz)	$\eta^\dagger$	$R_{dip}$ (%)
MA (63/66/61)	7.8	0.35	24.3	8.7	0.32	21.9	8.0	0.33	23.0
CA <sup>NTD</sup> <sub>133–277</sub> (29/27/23)	10.9	0.36	26.4	12.1	0.20	26.5	12.8	0.34	26.3
CA <sup>CTD</sup> <sub>282–363</sub> (17/19/16) <sup>†</sup>	-5.9	0.09	24.7	-7.5	0.47	23.5	-8.0	0.42	25.6
NC <sup>N-Zinc</sup> <sub>378–407</sub> (15/15/15)	7.0	0.07	19.0	6.8	0.10	17.8	7.1	0.07	19.0
NC <sup>C-Zinc</sup> <sub>413–432</sub> (13/13/13)	2.0	0.42	16.5	2.4	0.40	19.8	2.2	0.59	19.1

\*The numbers in brackets refer to the number of RDCs at the three ΔGag concentrations. All data were recorded in the absence of nucleic acids. Only RDCs in secondary structure elements are included in the SVD analysis.

<sup>†</sup>The maximum value of the rhombicity  $\eta$  is 2/3. When  $\eta = 0$ , the tensor is axially symmetric.

<sup>‡</sup>For the C-terminal dimerization domain of CA, CA<sup>CTD</sup>, the values of  $D_a^{NH}$  and  $\eta$  reported are the apparent values obtained by fitting against the coordinates of a single subunit, because the CA<sup>CTD</sup> exists in a dynamic monomer-dimer equilibrium (Table S1).

**Table S3.** Backbone amide ( $^1D_{NH}$ ) RDCs and  $^{15}N\{-^1H\}$  Heteronuclear NOE values for the C-terminal tail of MA (residues 108–132), the N terminus of CA (residues 133–148), the C-terminal tail of CA (residues 353–363), and SP1 (residues 364–377) measured on ΔGag

Residues	No DNA		With $\Delta P(-)PBS$
	$^1D_{NH}$ (Hz)*	$^{15}N\{-^1H\}$ NOE	DNA*
C-terminal tail of MA (residues 108–132)			
E107	15.5 ± 0.53	0.76 ± 0.03	0.69 ± 0.06
Q108	-	0.71 ± 0.03	0.65 ± 0.05
N109	9.5 ± 0.45	0.71 ± 0.04	0.66 ± 0.06
K110	-	0.54 ± 0.02	0.53 ± 0.04
S111	10.1 ± 0.26	0.49 ± 0.03	0.50 ± 0.05
K112	5.6 ± 0.18	0.40 ± 0.03	0.45 ± 0.04
K113	5.1 ± 0.17	0.36 ± 0.02	0.35 ± 0.02
K114	4.2 ± 0.13	0.29 ± 0.02	0.33 ± 0.02
Q116	1.1 ± 0.12	0.22 ± 0.01	0.22 ± 0.02
Q117	1.1 ± 0.13	-	0.23 ± 0.02
A118	0.6 ± 0.14	0.09 ± 0.01	0.10 ± 0.02
A119	-0.6 ± 0.13	0.07 ± 0.01	0.08 ± 0.02
A120	-0.5 ± 0.11	0.03 ± 0.01	0.07 ± 0.02
D121	-0.2 ± 0.08	0.11 ± 0.01	0.11 ± 0.01
T122	-1.0 ± 0.12	0.02 ± 0.02	0.05 ± 0.02
G123	0.0 ± 0.14	0.13 ± 0.02	0.13 ± 0.03
N126	-1.8 ± 0.39	0.27 ± 0.06	0.15 ± 0.07
Q127	-1.5 ± 0.11	0.08 ± 0.02	0.23 ± 0.02
V128	-2.0 ± 0.08	0.16 ± 0.01	0.14 ± 0.01
S129	-1.2 ± 0.12	0.15 ± 0.02	0.15 ± 0.02
Q130	-2.2 ± 0.15	0.08 ± 0.02	0.18 ± 0.03
N131	0.0 ± 0.45	0.12 ± 0.09	0.23 ± 0.09
Y132	-3.8 ± 0.18	0.26 ± 0.02	0.26 ± 0.02
N terminus of CA (residues 133–148)			
V135	-3.0 ± 0.10	-	-
Q136	-3.0 ± 0.41	0.26 ± 0.02	0.28 ± 0.03
N137	-3.3 ± 0.27	0.21 ± 0.04	0.24 ± 0.04
I138	-1.4 ± 0.23	0.32 ± 0.02	0.34 ± 0.02
Q139	-2.5 ± 0.25	0.30 ± 0.03	0.30 ± 0.03
G140	-1.2 ± 0.71	0.27 ± 0.03	0.26 ± 0.03
Q141	-1.0 ± 0.16	0.39 ± 0.02	0.34 ± 0.02
M142	-	0.31 ± 0.02	0.30 ± 0.02
V143	0.0 ± 0.13	0.40 ± 0.02	0.42 ± 0.03
A146	-	0.43 ± 0.06	0.46 ± 0.08
I147	-	0.53 ± 0.10	0.42 ± 0.13
S148	-	0.54 ± 0.12	-
C-terminal tail of CA (residues 353–363)			
V353	-0.8 ± 0.79	0.54 ± 0.06	0.48 ± 0.14
G354	-0.3 ± 0.72	0.48 ± 0.11	0.51 ± 0.17
G355	-0.9 ± 0.21	0.38 ± 0.02	0.39 ± 0.04
G357	-1.9 ± 0.45	0.38 ± 0.12	0.35 ± 0.09
A360	-	0.26 ± 0.03	0.28 ± 0.05
R361	-3.8 ± 0.57	0.36 ± 0.05	-
V362	-0.2 ± 0.22	0.40 ± 0.03	0.38 ± 0.04
L363	-0.6 ± 0.41	0.30 ± 0.07	0.29 ± 0.06
SP1 (residues 364–377)			
A364	0.6 ± 0.15	0.26 ± 0.02	0.27 ± 0.02
E365	0.5 ± 0.24	0.30 ± 0.04	0.30 ± 0.05
A366	-1.1 ± 0.39	-	-
M367	2.2 ± 0.14	0.38 ± 0.01	0.33 ± 0.02
Q369	0.4 ± 0.14	0.38 ± 0.02	-
V370	2.0 ± 0.19	0.27 ± 0.04	0.28 ± 0.03
T371	-0.2 ± 0.20	0.22 ± 0.04	0.25 ± 0.03
N372	-1.3 ± 0.28	0.39 ± 0.06	0.27 ± 0.05
S373	0.5 ± 0.36	-	0.34 ± 0.06

**Table S3.** Cont.

Residues	No DNA		With $\Delta P(-)PBS$
	${}^1D_{NH}$ (Hz)*	${}^{15}N\{-{}^1H\}$ NOE	DNA*
A374	1.2 ± 0.14	0.36 ± 0.04	0.31 ± 0.03
T375	1.4 ± 0.14	0.33 ± 0.03	0.32 ± 0.03
I376	1.3 ± 0.16	0.34 ± 0.03	0.37 ± 0.03
M377	1.1 ± 0.20	0.31 ± 0.02	0.34 ± 0.02

All RDCs (measured in 5% PEG-hexanol bicelles) and  ${}^{15}N\{-{}^1H\}$  NOE data were acquired at protein concentrations of 0.2 and ~0.3–0.35 mM, respectively, in subunits. Cross-peaks for residues 124, 125, 134, 145, 358, 359, and 368 were not observed due to line broadening. Residues 133, 149, and 356 are prolines. Residues in bold, corresponding to the C-terminal tail of MA (107–122), the N terminus of CA (residues 135–143), the C-terminal tail of CA (residues 357–363), and the N terminus of SP1 (residues 364–367) were previously predicted to form  $\alpha$ -helices (1–3). The first 13 N-terminal residues of CA (residues 133–145) form a  $\beta$ -strand in the isolated N-terminal domain of CA (4, 5). However, judging from the low  ${}^{15}N\{-{}^1H\}$  NOE values, close to zero backbone amide ( ${}^1D_{NH}$ ) RDCs, and minimal deviations from random coil backbone chemical shifts (see Fig. S2), it can be concluded unambiguously that these regions in  $\Delta$ Gag are intrinsically disordered in solution.

\*Protein:DNA molar ratio of 1:2.

1. Tang C, Ndassa Y, Summers MF (2002) Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nature Struct Biol* 9(7):537–543.
2. Accola MA, Höglund S, Göttinger HG (1998) A putative  $\alpha$ -helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *J Virol* 72(3):2072–2078.
3. Datta SAK et al. (2011) On the role of the SP1 domain in HIV-1 particle assembly: A molecular switch? *J Virol* 85(9):4111–4121.
4. Gitti RK, Lee BM, Walker J, Summers MF, Wundquist WI (1996) Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science* 273(5272):231–235.
5. Gamble TR et al. (1997) Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* 87(7):1285–1294.