ASSEMBLY OF HUMAN IMMUNODEFICIENCY VIRUS PRECURSOR GAG PROTEINS*

Doug Huseby, Robin Lid Barklis, Ayna Alfadhli, and Eric Barklis**

Vollum Institute and Department of Microbiology, Oregon Health & Science University, Mail Code L220, 3181 SW Sam Jackson Park Road, Portland, Oregon, 97201-3098

Running title: HIV Gag protein assembly

SUMMARY
To investigate the mechanism by which human immunodeficiency virus (HIV) precursor Gag (PrGag) proteins assemble to form immature virus particles, we examined the \textit{in vitro} assembly of MACANC proteins, composed of the PrGag matrix (MA), capsid (CA), and nucleocapsid (NC) domains. In the absence of other components, MACANC proteins assembled efficiently at physiological temperature, but inefficiently at lower temperatures. However, addition of RNA reduced the temperature sensitivity of assembly reactions. Assembly of MACANC proteins also was affected by pH, since the proteins preferentially formed tubes at pH 6.0, whereas spheres were obtained at pH 8.0. Because neither tubes nor spheres were amenable to analysis of protein-protein contacts, we also examined the membrane-bound assemblies of MACANC proteins. Interestingly, MACANC proteins organized on membranes in tightly packed hexameric rings. The observed hexamer spacing of 79.7Å is consistent with the notion that more PrGag proteins assemble into virions than are needed to provide capsid proteins for mature virus cores. Our data also are consistent with a model for PrGag contacts in immature virions where CA hexamers are tightly packed, where NC domains align beneath CA C-terminal domains (CTDs), and where matrix domains form trimers at the nexus of three neighbor hexamers.

INTRODUCTION

During human immunodeficiency virus 1 (HIV-1)** assembly, the viral precursor Gag (PrGag) protein oligomerizes on cellular membranes, and directs the budding of immature virus particles. Normally, during or after budding, PrGag proteins are processed into the mature Gag
proteins: matrix (MA), capsid (CA), nucleocapsid (NC), and p6 (1). PrGag processing is accompanied by a dramatic change in HIV-1 particle morphology. By conventional thin section electron microscopy (EM), the electron dense protein shell of immature virions appears to reorganize into centrally located, conical or cylindrical mature virus cores (1).

Recently, we demonstrated that the major HIV-1 Gag protein, CA, could assemble in vitro into two distinct arrangements (2). Both of these arrangements showed hexameric rings of CA N-terminal domains (NTDs) linked via C-terminal domain (CTD) contacts, but differed in that one showed tight packing of hexamers, with adjacent NTD rings in apparent contact, whereas the other featured clearly separated NTD rings (2). Interestingly, the two in vitro arrangements appear to have in vivo counterparts. The tightly packed arrangement is similar to that observed in immature virions, where hexamer-to-hexamer spacing is in the range of 65-80Å (2-5). In contrast, the loosely packed arrangement appears to correspond to the organization of capsid proteins in mature cores assembled in vitro and in vivo, with a hexamer ring-to-ring spacing of 95-110Å (6-7). The similarity between our results and those observed in vivo prompted us to propose a model in which viral morphogenesis was accompanied by a shift from a tightly packed hexamer rings to loosely packed rings, and predicted that virions would have an excess of capsid protein than was needed for mature core formation (2). This prediction was borne out by the observation of frequent virions with multiple cores, and the determination that HIV-1 appears to contain on the order of 5000 Gag proteins per particle—higher than previously expected (5).
Despite the apparent agreement between studies on particles and our investigations on HIV-1 capsid proteins in vitro, our CA proteins lacked other PrGag domains that might have an effect on Gag protein oligomerization, particularly as the proteins are organized in immature assemblies. To address this issue, we have undertaken the in vitro analysis of multidomain Gag proteins, carrying matrix, capsid, spacer (SP1) and nucleocapsid domains. Interestingly, these MACANC proteins assembled efficiently in the absence of other components at 37°C, but not at lower temperatures, suggesting a temperature-sensitive assembly switch. We also observed that the proteins preferentially assembled into tube forms at low pH, but into spheres at high pH. Because these tube and sphere forms were not amenable to structural analysis, we additionally examined membrane-bound assemblies of the proteins. Image analysis of membrane-bound protein arrays indicated that the MACANC proteins organized into hexamer rings with a spacing of 79.7Å. Our results indicate that the presence of MA and NC domains is compatible with a tightly packed arrangement of PrGag proteins, and support a model where PrGag processing yields a loose capsid core from a more crowded precursor arrangement.

EXPERIMENTAL PROCEDURES

**Protein purification.** MACANC and MACANCexact proteins were expressed in *E. coli* strain BL21(DE3)/pLysS (Novagen) from the bacterial expression plasmids pET15B-HIVMACANC and pET15B-HIVMACANCexact. Both constructs encode an amino-terminal histidine-tag (histag) of MGSSH HHHHH SSGLV PRGSH MLEDPP fused to the second codon of HIV-1 HXB2 gag. The pET15B-HIVMACANCexact plasmid expresses protein which terminates precisely at
the HIV-1 NC C-terminus; the pET15B-HIVMACANC plasmid encodes the vector-derived residues KIRAA NKARK EAELA AATAEQ at the C-terminus of NC. Bacterial expression and purification of proteins by nickel-chelate chromatography followed methods described previously (2, 8-12). Purified protein fractions were desalted by buffer exchange in Sephadex G25 spin columns in 10 mM sodium phosphate (pH 7.8), 500 mM NaCl, and stored under nitrogen at -80°C. Purified (>90%) proteins at 0.5-1.5 mg/ml, as well as protein purification fractions were evaluated after sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE; 13) by Coomassie blue staining (2, 8-12), and by immunoblot detection. As described previously (13), immunoblot detection employed an anti-HIVCA monoclonal antibody (Hy183; 13) as the primary antibody, an alkaline phosphatase-conjugated antimouse secondary antibody (Promega), and visualization of bands using a color reaction of nitro blue tetrazolium plus 5-bromo-4-chloro-3-indolyl phosphate in 100 mM Tris-hydrochloride [pH 9.5], 100 mM NaCl, 5 mM MgCl₂.

**Pelleting assays.** Incubations of 10 uM MACANC proteins were performed in 20 ul reactions at 23°, 30°, or 37° C for 16 h. Incubation buffers consisted of 375 mM NaCl, 5 mM dithiothreitol, 100 mM Hepes (pH 8.0 or 6.0) in the presence or absence of 0.2 mg/ml *E. coli* rRNA (Roche). Following incubations, reactions were centrifuged 10 min at 4°C at 14,000 x g. Supernatant and pellet fractions were collected for fractionation by SDS-PAGE and visualized by Coomassie blue staining.

**2D crystallization.** 2D crystallization incubations were essentially as described previously (2, 8-12). Briefly, 10 ul drops of protein (0.25-0.75 mg/ml) in buffer (25 mM sodium phosphate [pH
8.3], 500 mM NaCl, 5 mM sodium acetate, 20% glycerol) were incubated beneath 1 ul of 200 ng/ul phosphatidyl choline (Avanti) plus 50 ng/ul nickel-charged 1,2-dioleoyl-sn-glycero-3-[N-((5-amino-1-carboxypentyl)iminodiacetic acid)-succinyl] (DOGS; Avanti) in 1:1 chloroform:hexane. Incubations were performed for 16-24 h at 22-25°C in sealed plastic dishes, humidified with 6.25 mM sodium phosphate [pH 8.3], 125 mM NaCl. Following incubations, monolayers and associated proteins were lifted onto lacey carbon grids (Ted Pella 01883) for 30-60 sec, washed on drops of distilled water for 15-30 sec, wicked, stained with 1.33% uranyl acetate, wicked again, and allowed to dry at least 15 min.

Samples were viewed on a Philips CM120 transmission electron microscope (EM) under low dose conditions. All images were taken at 100 kV, and at defocus values between 200 nm and 1500 nm. Images either were collected on a Gatan 794 CCD multi-scan camera (MSC) or on Kodak SO163 negative films. For processing, CCD images were converted from Gatan Digital Micrograph 3 (DM3) format to eight bit grayscale TIFF format using Gatan software. Negatives were scanned and digitized using a Nikon Coolscan 8000 film and slide scanner in super fine scan (4000 dots per inch) mode with the multi-sampling mode at 8x. Images were saved in TIFF format using Nikon Scan 3.1 software.

**Sphere and tube incubations.** Proteins were incubated at pH 6.0 (for tubes) or pH 8.0 (for spheres) beneath lipid monolayers, and then lifted onto lacey EM grids, prior to staining, EM viewing, and image collection as described above.
**Unit cell statistics.** Untitled 2D crystal images, converted from TIFF to MRC format as described previously (2, 8-12), were boxed using the MRC BOXMRC program (2, 14-15), and Fourier transformed using the 2DFFT function of ICE (2, 16-17). Transform TNF files were viewed as power spectra on SPECTRA (18), hand-indexed, and converted to amplitudes and phases (APH) files, using the LATREF, UNBENDA, and MNBOX functions of SPECTRA (18). Unit cell dimensions were obtained using the SPECTRA file information interface from indexed transforms, and axes lengths were corrected by dividing raw values by sinγ* (2). Spacegroup analysis was performed using ALLSPACE in 3° phase origin search steps (2, 14-15).

**3D reconstructions.** Tilt series images of membrane-bound protein arrays were collected on negatives at tilt angles of -55°, -45°, -30°, -15°, 0°, +15°, +30°, +45°, and +50° and scanned as described above to produce TIFF files. From these TIFF files, thirty separate crystalline patches were tracked through the entire range of tilt angles, and 512 x 512 pixel areas for each angle of each of the thirty crystalline patches were windowed using Gatan Digital Micrograph (DM) software. The first steps of preparing zero degree reference images were performed using DM software: the 512 x 512 pixel areas were Fourier transformed, hand-indexed from power spectra, masked, low pass filtered to include only the three lowest resolution reflections (1,0; 1,1; 2,0), back-transformed, and windowed to obtain 128 x 128 pixel (262.4 x 262.4 Å) raw reference images. Raw zero degree reference images, along with all 512 x 512 pixel areas from the thirty tilt series were converted from TIFF format to MRC format and then to SPIDER SPI format (19), and subsequent processing steps were performed using the SPIDER suite (19) of programs.
To generate final zero degree reference images, raw 128 x 128 pixel images were thresholded with the SPIDER TH command so that pixels with grayscale values below the image average were raised to the image average. After this, images were padded to 512 x 512 pixels using the PD command and the original average grayscale value as background to create the final zero tilt reference images. At this point, references were used in cross-correlations (CC command) versus their original 512 x 512 pixel zero tilt images to generate cross correlation maps. Peaks in cross-correlation maps were hand-picked using the WEB interface (19), after rotating the maps by 180° using the SPIDER RT 90 command, to account for the fact that SPIDER designates the origin as the image lower right corner, whereas the WEB origin is at the upper left. From 20-100 hand-picked cross-correlation peak locations per zero tilt image, corresponding 148 x 148 pixel windows in the original images were cut and averaged using the WV command. Final 128 x 128 pixel zero degree averages, centered on hexameric protein-free centers, were cut using the WV command and an image center location, hand-picked from the 180° rotated image on WEB. After obtaining 30 zero degree averages, the steps described above were used to generate averages for each tilt angle in each tilt series. During this process, thresholded and padded zero degree averages were used as references in -15° and +15° cross-correlation searches, and output averages were used recursively in obtaining higher tilt averages.

Of the original thirty tilt series, 226 image averages from a total of 8455 windows were used for 3D reconstructions using SPIDER suite (19). Average pixel intensity values for each image average were normalized and inverted (to account for negative staining), and rotated +223° to align the tilt axes with the image y axes. Each set of tilt series averages then was used to create a 3D volume by back-projection using the BP 3D command and input phi, theta, psi.
Eulerian angles of 0.0, tilt angle, 0.0. Determination of the relative rotations between tilt series was accomplished using the AP RA command on 2D projections from threefold symmetrized, calculated tilt series volumes. The output of these operations included rotation angles which corresponded to the reverse of the psi angles needed for final reconstructions. Thus, the 226 tilt series image averages were employed as input for final 3D reconstructions using the SPIDER 3D BP command. The quality of 3D reconstructions was assessed by splitting data sets in half, calculating 3D volumes for each half data set, and comparison of the two half data sets using the RF 3 command. The resolution of the reconstruction was estimated to be at 27.5Å, based on the resolution at which a phase residual value of 45° was reached (2, 8-12, 19). Final filtered volumes were viewed using WEB (19) or a VTK-based ISO_VIEW interface, and were converted to TIFF or JPEG format for presentation.

Rotational correlations. Zero tilt average images from tilt series were compared to rotated versions of themselves, by use of the SPIDER CC command. Maximum peak values (1.0 is defined as a perfect match) were recorded and plotted for comparisons made at one degree intervals.

RESULTS

In vitro assembly of MACANC proteins
To investigate the influences that the HIV-1 Gag MA and NC domains might have on the assembly of CA proteins, histidine-tagged (his-tagged) MACANC proteins were expressed in bacteria (see Experimental Procedures). Two proteins, MACANC and MACANCexact, were made. These differed by the presence of 21 vector-derived codons at the C-terminus of MACANC, but because the additional encoded residues did not appear to have an effect in our studies, we will refer to them interchangeably as MACANC proteins. As shown in Figure 1, both of the MACANC proteins carried N-terminal his-tags in place of the PrGag myristate (20) group, and neither encoded the PrGag p6 domain, which does not appear to play a structural role, but does hamper purification of bacterially expressed proteins (21-22). Following previous methods (2, 8-12; Experimental Procedures), the proteins were purified using nickel-chelate chromatography to concentrations of 0.5-1.5 mg/ml, and estimated purities of >90% (Figure 2, lane I).

For monitoring the assembly properties of the MACANC proteins, we employed pelleting assays in which purified proteins were incubated under varying conditions. Following incubations, centrifugation was used to separate pelletable assembled products (P) from monomers and small oligomers in the supernatant (S), and protein levels in each fraction were visualized by staining, after electrophoresis. Results from these studies demonstrated a marked temperature-dependence on the assembly process. As illustrated in Figure 3, lanes 1-2, during 23°C incubations, very little MACANC protein assembled into a pelletable form either at pH 6.0 (Panel A) or pH 8.0 (Panel B). In contrast, incubations performed at 30°C and 37°C yielded increasing amounts of assembled protein (lanes 5-6, 9-10).
Because NC domains on PrGag proteins have been postulated to facilitate assembly after binding to RNA (23-25), we repeated pelleting assays in the presence of RNA (Figure 3). As illustrated in lanes 3-4, 7-8, 11-12, while assembly efficiency still appeared affected by temperature at pH 6.0 (Panel A), RNA addition modulated this effect. The influence of RNA was even more pronounced at pH 8.0 (Figure 3, Panel B, lanes 3-4, 7-8, 11-12). These results suggest that putative NC-RNA interactions decrease the temperature-sensitivity of MACANC assembly reactions, an observation that is discussed below (see Discussion section).

What are the morphologies of the structures assembled by MACANC proteins? To examine this issue, assembly products were lifted onto electron microscope (EM) grids, negatively stained, and viewed by transmission EM. Of note was the fact that pH 8.0 MACANC incubations yielded spherical particles with diameters on the order of 100 nm. In some cases, these spheres clustered into paracrystalline arrays (Figure 4). In contrast, pH 6.0 incubation of MACANC proteins gave large tubes, which, at low magnification, appeared as dark rods of up to 750 nm in length, against the grid substrate (Figure 5, Panels A, B). At higher magnification, the MACANC rods appeared to be 60-70 nm wide cylinders with hollow centers of 20-25 nm (Panels C, D). Our observation of different particle morphologies at different pHs is in agreement with the observations of others, and is consistent with the notion that pH regulates the HIV-1 Gag protein assembly pathway, at least in vitro (22, 26). Additionally, the cylinder wall widths in Figure 5, Panels C and D (20-25 nm) are comparable to the distances observed for the MA plus CA plus NC domains observed in radial density profiles of immature HIV-1 virus-like particles (27-28), suggesting a similar MA-CA-NC radial arrangement in our tubes. Despite these correlations, the MACANC sphere and tube forms were not amenable to higher resolution
analysis of how the proteins associate in higher order assemblies. Consequently, we opted to examine membrane-bound assemblies of MACANC proteins.

**MACANC membrane-bound arrays**

As noted above, MACANC spheres and tubes were not amenable to image analysis techniques that might help clarify how protein units associate to form higher order assemblies. This was because the MACANC tubes were not regular enough to give clear diffraction patterns, and the spheres were not compatible with icosahedral reconstruction methods. As an alternative approach, we examined the assembly of MACANC proteins on membrane monolayers, as we have previously (2, 8-12). To do so, the N-terminally his-tagged MACANC proteins were incubated beneath nickel-chelating lipid monolayers, lifted along with the membranes onto EM grids, and imaged by transmission EM. Images of such membranes showed darkly staining protein patches on lightly staining lipid layers (Figure 6A). Although the MACANC arrays were only 50-200 nm wide, Fourier transforms of boxed arrays yielded calculated diffraction patterns with hexagonal sets of reflections. (Figure 6B). The lowest resolution 1,0 reflections were observed at 1/69 Å, corresponding to a real space hexagonal unit cell spacing of 79.7 Å, consistent with the tightly packed arrangement observed for HIV-1 CA proteins. Indeed, backtransformation of filtered reflections gave a 2D projection image (Figure 6C) of proteins (in
white) viewed perpendicular to membranes, surrounding protein-free holes spaced at 8 nm distances.

Compilation of data from multiple crystalline patches supported the results shown in Figure 6. Our average membrane-bound MACANC unit cell had edges of $a=b=79.7 \pm 2.0\text{Å}$ and a reciprocal space $\gamma^*$ angle of 60.9° (Table 1). Consistent with the hexagonal appearance of the 2D projection (Figure 6C), phase comparisons of Fourier transform reflections gave low phase residuals when unit cells were assumed to be p1, p2, p3 or p6 symmetry (Table 1). Based on the observation that MACANC crystalline patches obeyed p3 or p6 symmetry, we modified a single particle reconstruction approach to obtain a 3D model of membrane-bound MACANC from low dose tilt series of the crystalline patches. To do so, crystal unit cell centers were identified in untilted images by cross-correlation with Fourier filtered projections such as that in Figure 6C. The highest peaks in cross-correlation maps such as those shown in Figure 7 (top panel, 0° map) were used to pick real space windows to sum in image averages (Figure 7, bottom panel). For each series, after obtaining a zero tilt image average (Figure 7, bottom panel, center image), the average was used in successive cross-correlation, windowing, and averaging steps to obtain image averages for tilt angles in the range of $-45^\circ$ to $+45^\circ$ (Figure 7).

With the zero-tilt averages obtained from our 30 tilt series, it was of interest to examine whether the MACANC proteins assembled around protein-free holes in a trigonal (p3) or hexagonal (p6) fashion. For this analysis, individual zero tilt averages from different tilt series were compared with rotated versions of themselves by cross-correlation. The rationale for this approach was that for a trigonal arrangement, one would predict a set of three roughly equivalent
cross-correlation peaks at 0°, 120°, and 240°, whereas for a hexagonal pattern, six equal peaks spaced at 60° intervals were predicted. As shown in Figure 8, the latter prediction was observed, supporting the interpretation that MACANC proteins align on membranes in a hexagonal fashion.

To generate a 3D model of membrane-bound MACANC proteins, tilt image averages from the 30 tilt series were used to build a 3D volume by back-projection (see Experimental Procedures). The quality of the reconstruction was evaluated by splitting the data set in half, calculating 3D volumes for each half set, and comparing the two half set volumes. Using this methodology, we estimate our 3D reconstruction to be at 27.5Å resolution. Not surprisingly, surface renderings of the 3D reconstruction viewed perpendicular to the membrane from either membrane side up (Figure 9a) or membrane side down (Figure 9b) look similar to the 2D projection shown in Figure 6c: central protein-free holes are surrounded by interconnected dumbbell units (highlighted in white in panel B) that we interpret to be dimers of capsid NTDs.

At a lower contour level, a view of the bottom surface of the volume shows putative CA CTD connections as peanut shapes extending radially from hole centers (Figure 10a). Successive images of the volume tilted gradually to the reader (Figure 10b-d) show additional views of the protein ring arrangement. In the final panel (10d), the view is parallel to the membrane, with the membrane side up. Although the knobs at the bottom of the panel could represent some portion of NC density, neither NC nor MA domains are resolved in the reconstruction. Despite this, the arrangements of putative capsid NTDs and CTDs, and the relatively tight packing of the hexamer rings are similar to the tight packed capsid pattern that we have observed previously (2). These
results demonstrate that MA and NC domains do not interfere with HIV-1 CA tight packing preferences, and support the notion that immature virions feature closely associated capsid hexamer rings.

DISCUSSION

Although much now is known concerning the three dimensional structures of the individual HIV-1 Gag protein structures (29-37), considerably less is understood about how the Gag proteins or PrGag protein domains associate in immature and mature HIV virions. Our recent demonstration that HIV-1 capsid proteins can assemble onto membranes in either tightly packed or loosely packed hexagonal arrangements (2) substantiated observations concerning capsid contacts in alternative in vivo and in vitro assemblies (3-7), and suggested a model for viral morphogenesis (2). However, the contributions of matrix and nucleocapsid domains were not tested in our studies, and thus we decided to examine the interactions of MACANC proteins in vitro. It should be noted that our proteins have excluded the C-terminal PrGag SP2 and p6 domains. Previous work suggested that the in vitro assembly properties of p6-deleted and partially pure full-length PrGag proteins were similar (21), and p6 domain deletions appeared to have no effects on radial density profiles of immature virions (28). However, it has been observed in fluorescence studies that full-length HIV-1 PrGag proteins were not induced to oligomerize by the addition of tRNA (43). Thus, although that study employed a different RNA and lower concentrations of protein (200 nM) than we did, our results might not be completely comparable to results with full-length PrGag proteins.
One consistent (38) observation with our SP2-p6-truncated MACANC proteins is that they preferentially assembled into higher order structures at 37°C versus 23°C or 30°C (Figure 3). One interpretation of this result is that the increased temperature may be necessary for the proteins to overcome an energy barrier between assembly-impaired and assembly-efficient conformations. If this is the case, then at least two models might explain how RNA (Figure 3) might facilitate assembly. In one case, MACANC binding to RNA might substitute for the temperature effect by lowering the energy barrier for the conformation change. Alternatively, a two step process might be envisioned, where MACANC-RNA binding might increase the levels of an assembly intermediate. Either way, it is interesting that in the absence of RNA and membranes, MACANC proteins assembled tubes at pH 6.0, and spheres at pH 8.0, as was observed previously (22). Since HIV-1 CA proteins can assemble tubes in either a loose packing arrangement (6), or in a tight packing arrangement (Huseby and Barklis, unpublished observations), we do not believe that MACANC tube or sphere morphologies are dictated by loose versus tight hexamer packing arrangements. Rather, we speculate that the tube and sphere morphologies employ the same type of hexamer packing, but differ in the frequency and location of pentamer placement (39-40).

If HIV-1 MACANC tubes or spheres were of high enough regularity, it might be possible to extract unit cell information from their Fourier transforms (2, 5-7). In lieu of this, we opted to investigate the assembly properties of the his-tagged proteins on nickel-chelating membrane monolayers. As demonstrated in Table 1, and Figures 6, 9, and 10, the membrane-bound proteins associated to form hexamer (Figure 8; Table 1) rings around protein-free holes, which may
accommodate HIV envelope protein cytoplasmic tails in immature virions. The hole-to-hole distance that we observed is in the range of that observed for tightly packed hexamers in immature virus forms (2-5), and below that observed for mature virus cores assembled in vivo and in vitro (6-7). Thus, the presence of MA and NC domains on MACANC proteins is wholly compatible with a tightly packed pattern. These and other results support a model where immature HIV-1 virions incorporate about 5000 (5) PrGag proteins in a tightly packed arrangement, and that only a fraction of the PrGag capsid domains contribute to the formation of mature cores, composed of loosely packed hexamers. Although the excess CA proteins may assemble an additional virus core within a mature virion (7), the fate of the unutilized CA is unknown. Also unknown are the precise contacts made by MA and NC domains, which are not resolved well, presumably due to their greater freedom of motion within arrays, and/or the often observed loss of resolution perpendicular to the membrane plane (2, 10-12, 15). Because the assembly function of NC domains can be replaced by protein dimerization domains (41-42), it seems probable that NC domains group in pairs in PrGag assemblies. In contrast, matrix domains have shown a tendency to trimerize (29-30), suggesting that they will associate as trimers within PrGag arrays. A model to accommodate these observations and our current results is illustrated in Figure 11. As shown, CA NTDs group in hexamer rings using asymmetric, side-by-side contacts, and are tightly packed by virtue of putative symmetric contacts between NTDs of adjacent hexamers. Above the NTD hexamers are MA trimers, at the point where three hexamers come together. Aligned beneath putative NTD dimers are CTD dimers, which interconnect hexamer rings, and NC dimers, presumably associated with the viral HIV-1 RNA. It will be of interest to test this and other models of HIV assembly in vitro, and in vivo.
REFERENCES


28. Wilk T., Gross, I., Gowen, B., Rutten, T., de Haas, F., Welker, R., Krausslich, H.-G.,


33. Gamble, T., Yoo, S., Vajdos, F., von Schwedler, U., Worthylake, D., Wang, H.,


FOOTNOTES

* This work was supported by a National Institute of General Medical Sciences Grant (NIGMS GM060170) to EB, and an American Foundation for AIDS Research Post-doctoral Fellowship Grant (amfAR 106523-35-RFNT) to AA.

** To whom correspondence should be addressed: TEL: 503-494-8098; FAX: 503-494-6862; Email: barklis@ohsu.edu.

*** The abbreviations used are: HIV, human immunodeficiency virus; PrGag, precursor Gag protein; MA, matrix; CA, capsid; NC, nucleocapsid; EM, electron microscopy; NTD, capsid N-terminal domain; CTD, capsid C-terminal domain; SP1, spacer 1; SP2, spacer 2; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; DOGS, 1,2-dioleoyl-sn-glycero-3-[N-((5-amino-1-coarboxypentyl) iminodiacetic acid)-succinyl]; his, histidine.

TABLE AND FIGURE LEGENDS

**Table 1. Unit cell statistics.** Unit cell statistics were obtained from images of untilted membrane-bound crystalline arrays. The real space a axis length was calculated from Fourier
transform a* axes, and corrected for gamma* angles. Statistics for space group fitting were obtained from the program ALLSPACE, which compares the phases of Fourier transform reflections for internal consistency with 2D space groups, and outputs phase residuals for consistency comparisons. Using this algorithm, a perfect fit gives a phase residual of 0°, while a random fit yields a 90° phase residual. Note that because internal phase residual comparisons are not relevant with the primitive (p1) space group, the phase residual for p1 is on the basis of signal-to-noise ratios of observed amplitudes. N indicates the number of transforms which contributed to the averages and standard deviations. Note that space group fitting statistics were obtained from transforms in which a minimum of seven comparisons could be made.

**Figure 1. HIV Gag proteins.** The HIV-1 Gag precursor protein, Pr55Gag is an N-terminally myristoylated precursor protein composed of the matrix (MA), capsid (CA) nucleocapsid (NC), and p6 domains, and the spacer peptides SP1 and SP2. For bacterial expression and purification, the MACANC and MACANCexact proteins carry an N-terminal histidine tag fused to the second codon of HIV-1 HXB2 gag, and are truncated either exactly at the NC C-terminus (MACANCexact) or after 21 additional vector-derived residues (MACANC). Because the additional vector-encoded sequences did not appear to have an effect in our studies, we refer to them collectively as MACANC proteins. As shown, both proteins carry the HIV-1 MA, CA, SP1, and NC domains.

**Figure 2. Protein purification.** Nickel chelate chromatography column fractions of the bacterially expressed MACANC protein were fractionated by SDS-PAGE, and visualized by Coomassie blue staining. Column fractions consisted of starting bacterial lysate material (A);
column loading buffer washes (B,C); 10 mM imidazole buffer washes (D,E); and elutions (E-J). Final elutions yielded a protein with a mobility similar to the 52.2 kDa marker in the protein standards lane (J), and which was detected by anti-HIV capsid antibodies in immunoblots performed in parallel.

**Figure 3. Assembly properties of his-MACANC proteins.** Incubations containing 10 μM MACANC proteins were performed at the indicated pHs and temperatures in the absence or presence of 0.2 mg/ml *E. coli* rRNA. After incubations, supernatant (S) and pellet (P) fractions were separated by centrifugation. Samples were subjected to SDS-PAGE, and were visualized by Coomassie blue staining. Arrowheads indicate the full-length MACANC proteins.

**Figure 4. Morphology of MACANC spheres.** Spheres assembled during pH 8.0 incubations of MACANC were lifted onto lacey EM grids, negatively stained with uranyl acetate, and imaged by transmission electron microscopy. As compared with the 200 nm size standard, sphere diameters appeared to be approximately 100 nm.

**Figure 5. Morphology of MACANC tubes.** After protein incubation at pH 6.0, tubes of up to 750 nm in length were observed by EM at low magnification as dark rods on grey carbon lace-lipid monolayer substrates (panels A, B). At higher magnification (panels C, D), the tubes appeared as hollow cylinders of 60-70 nm diameter, with white (protein) walls, and central pores containing accumulated uranyl acetate negative stain (dark). Size standards are as indicated.
**Figure 6. Membrane-bound assemblies of MACANC proteins.** When incubated beneath membrane monolayers containing nickel-chelating lipids, MACANC proteins assembled two dimensional (2D) crystalline patches which appear as dark areas on negatively stained membranes (panel A). Fourier transformation of crystalline patch areas yielded diffraction patterns (power spectra) with a set of six reflections at $1/69\AA$, one of which is boxed in panel B. After indexing and masking hexagonal reflections from diffraction patterns, backtransformation yielded Fourier filtered, real space images as shown in panel C, where protein areas are light, and protein-free zones are dark. Note that for a hexagonal unit cell, unit cell dimensions of approximately 8 nm, indicative of the spacing between the large protein-free holes, correspond to the $1/69\AA$ 1,0 reflections observed in diffraction patterns.

**Figure 7. Projection images of membrane-bound MACANC proteins.** Crystalline patches of membrane-bound MACANC proteins were identified as white peaks in cross-correlation maps (top panel) as described in the Experimental Procedures section. Major peaks in cross-correlation maps were used for picking windows of crystalline patches for real space averaging to generate a panel of projection image averages (lower panel) from micrographs taken at different degrees of tilt. Note that protein areas appear white in the image averages, and that the size bar and tilt axis for this tilt series (#19) are as indicated.

**Figure 8. Rotational correlation of membrane-bound proteins.** Zero tilt image averages from tilt series #13 (A), #17 (B), and #30 (C) were compared with rotated versions of themselves to examine unit cell symmetry. Plots show maximum peak values obtained in cross-correlations.
performed with unrotated images versus the same images rotated at the indicated angles. Note that perfect matching corresponds to a peak value of 1.0.

**Figure 9.** Protein assemblies viewed from below and above membrane monolayers. Three dimensional reconstruction of membrane-bound MACANC assemblies to a resolution of 27.5Å was performed as described in the Experimental Methods section. Panel A shows the arrangement of proteins viewed perpendicular to the membrane from below, while panel B shows the arrangement viewed from above. Proteins appear gray against a black background, and a putative symmetric capsid dimer is outlined in white. The volumes were depicted using WEB, at a contour level of approximately 1.1 standard deviations.

**Figure 10.** Organization of membrane-bound MACANC proteins. Three dimensional reconstruction of membrane-bound MACANC assemblies to a resolution of 27.5Å was performed as described in the Experimental Procedures section. Panel A is oriented so that the membrane is perpendicular to the viewer, with the membrane side away from the viewer. Panel B is tilted 30° toward the viewer; panel C is tilted 60° toward the viewer; and panel D is tilted 90° toward the viewer such that the view is parallel to the membrane, with the membrane side to the top of the page. Proteins appear as gray or white against a black background, and are depicted using WEB at a contour level of approximately 0.88 standard deviations.

**Figure 11.** Model for PrGag contacts. Shown is a model for how HIV-1 MACANC proteins associate at membranes in immature virions. In panel A, domains of PrGag proteins in hexameric arrays are viewed at increasing distances away from the membrane. In the top section,
an MA trimer unit is outlined. In the capsid NTD panel, a hexamer ring of NTD units is circled, in which asymmetric side-by-side contacts are employed. Also circled is a symmetric, head-to-head NTD dimer, and similar dimer units are outlined in the capsid CTD and NC sections. At the bottom is the composite of the MA, NTD, CTD and NC layers. In panel B, a subset of contacts are depicted, as viewed nearly parallel to the membrane. As shown, NTDs form hexamer rings that are tightly associated via symmetric contacts, and interconnected via symmetric CTD dimers. NC domains are proposed to align beneath CTDs, while MA domains are modeled as trimers at the nexus of three neighbor hexamer rings. Note that for clarity, only two or three CTDs, and NC and MA domains are depicted.
### Table 1. Unit cell statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>a axis length</td>
<td>79.7 ± 2.0Å</td>
<td>13</td>
</tr>
<tr>
<td>long-to-short axis ratio</td>
<td>1.02 ± 0.02</td>
<td>21</td>
</tr>
<tr>
<td>gamma* angle</td>
<td>60.9 ± 2.6°</td>
<td>21</td>
</tr>
<tr>
<td>p1 fit phase residual</td>
<td>31.8 ± 7.7°</td>
<td>20</td>
</tr>
<tr>
<td>p1 fit comparisons</td>
<td>12.6 ± 3.3</td>
<td>20</td>
</tr>
<tr>
<td>p2 fit phase residual</td>
<td>26.2 ± 9.9°</td>
<td>12</td>
</tr>
<tr>
<td>p2 fit comparisons</td>
<td>7.6 ± 0.9</td>
<td>12</td>
</tr>
<tr>
<td>p3 fit phase residual</td>
<td>15.7 ± 9.8°</td>
<td>11</td>
</tr>
<tr>
<td>p3 fit comparisons</td>
<td>10.4 ± 2.7</td>
<td>11</td>
</tr>
<tr>
<td>p6 fit phase residual</td>
<td>18.1 ± 8.4°</td>
<td>22</td>
</tr>
<tr>
<td>p6 fit comparisons</td>
<td>21.0 ± 9.3</td>
<td>22</td>
</tr>
</tbody>
</table>

Unit cell statistics were obtained from images of untilted membrane-bound crystalline arrays. The real space a axis length was calculated from Fourier transform a* axes, and corrected for gamma* angles. Statistics for space group fitting were obtained from the program ALLSPACE, which compares the phases of Fourier transform reflections for internal consistency with 2D space groups, and outputs phase residuals for consistency comparisons. Using this algorithm, a perfect fit gives a phase residual of 0°, while a random fit yields a 90° phase residual. Note that because internal phase residual comparisons are not relevant with the primitive (p1) space group, the phase residual for p1 is on the basis of signal-to-noise ratios of observed amplitudes. N indicates the number of transforms which contributed to the averages and standard deviations. Note that space group fitting statistics were obtained from transforms in which a minimum of seven comparisons could be made.
<table>
<thead>
<tr>
<th></th>
<th>23°C</th>
<th></th>
<th>30°C</th>
<th></th>
<th>37°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>-RNA</td>
<td>S</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+RNA</td>
<td>S</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 3**

**A**

- pH 6.0
- 1 2 3 4 5 6 7 8 9 10 11 12

**B**

- pH 8.0
-  S  P  S  P  S  P  S  P  S  P  S  P

**Fig. 3**
cross correlation maps

tilt angle image averages

Fig. 7
Fig. 8
Fig. 11b

- **MA trimers**
- **symmetric NTD dimer contacts**
- **asymmetric NTD hexamer contacts**
- **NC dimers**
- **symmetric CTD dimers**