Solution Properties of Murine Leukemia Virus Gag Protein: Differences from HIV-1 Gag

Siddhartha A.K. Datta1,* Xiaobing Zuo2†, Patrick K. Clark3, Stephen J. Campbell1‡, Yun-Xing Wang2, and Alan Rein1*

HIV Drug Resistance Program1, Structural Biophysics Laboratory2, National Cancer Institute-Frederick., Basic Research Program3, SAIC-Frederick inc., Frederick, MD 21702-1201

Running Title: Solution properties of MLV Gag

†Present address: X-ray Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439

‡Present address: Boyce Thompson Institute for Plant Research, Ithaca, NY 14853

Words in Abstract: 218
Words in Text: 5,635

*Corresponding authors. Mailing address: HIV Drug Resistance Program, National Cancer Institute-Frederick, P.O. Box B, Frederick, MD 21702-1201. Fax: (301) 846-6013. Phone for Siddhartha A.K. Datta: (301) 846-1844. E-mail: dattasi@mail.nih.gov.
Phone for Alan Rein: (301) 846-1361. E-mail: reina@mail.nih.gov.
ABSTRACT

Immature retrovirus particles are assembled from the multi-domain Gag protein. In these particles, the Gag proteins are arranged radially as elongated rods. We have previously characterized the properties of HIV-1 Gag in solution. In the absence of nucleic acid, HIV-1 Gag displays moderately weak inter-protein interactions, existing in monomer-dimer equilibrium. Neutron scattering and hydrodynamic studies suggest that the protein is compact, and biochemical studies indicate that the two ends can approach close in 3 dimensional space, implying the need for a significant conformational change during assembly. We now describe the properties of the Gag protein of Moloney murine leukemia virus (MLV), a gammaretrovirus. We find that this protein is very different from HIV-1 Gag: it has much weaker protein-protein interaction and is predominantly monomeric in solution. This has allowed us to study the protein by small angle x-ray scattering and build a low-resolution molecular envelope for the protein. We find that MLV Gag is extended in solution, with an axial ratio of ~ 7, comparable to its dimensions in immature particles. Mutational analysis suggests that runs of prolines in its MA and p12 domains, and the highly charged stretch at the C-terminus of its CA domain, all contribute to this extended conformation. These differences between MLV Gag and HIV-1 Gag and their implications for retroviral assembly are discussed.
Expression of a single virus-coded protein, termed Gag, in permissive host cells is sufficient for efficient assembly of immature retrovirus-like particles (VLPs) (52). In most retroviruses, Gag molecules are initially targeted to the plasma membrane where they accumulate and organize into a protein lattice, ultimately leading to the formation and budding of immature VLPs. After the immature particle is released from the cell, the Gag proteins within it are cleaved by the viral protease into their constituent domains to form a mature virus. Immature VLPs are roughly spherical, pleomorphic membrane-enclosed structures, with a diameter of approximately 120 nm. Under the membrane exterior several thousand Gag molecules are arranged radially, with their MA domains in contact with the lipid bilayer and their NC domains facing the interior of the particle, presumably interacting with RNA. The Gag proteins in the immature particle are highly extended rods, with a length ~ 20 nm and a width only ~ 2.3 nm (2, 57, 58).

Each Gag protein molecule that participates in the structural lattice of these nearly regular VLPs must interact in a highly specific way with several other Gag molecules. In turn, these interactions must depend upon the overall conformation of the participating monomers. A structural feature of the immature protein lattice, conserved in all retroviruses studied, is the hexagonal arrangement of Gag molecules, mediated largely by interactions between their CA domains (13). The presence of irregular defects, which allows for curvature of the hexagonal lattice to form an enclosed sphere, has been observed in several retroviruses (13, 57).
The similarity in organizational principles of immature retroviral VLPs across genera might suggest similarities in the overall structural properties of Gag proteins. Indeed, there are conserved features between Gag proteins from different genera, despite a low degree of sequence conservation. All retroviral Gag proteins are multi-domain proteins, always containing (from N- to C-terminus) the matrix (MA), capsid (CA), and nucleocapsid (NC) domains (52), though Gag proteins often contain one or more genus-specific domains in addition to the three canonical domains. Structural studies on isolated domains of Gag reveal conservation in three-dimensional shape: the MA domains of HIV-1, equine infectious anemia virus, Rous sarcoma virus, and Moloney murine leukemia virus (MLV) Gag show a remarkable similarity in 3 dimensional shape (8, 21, 22, 32, 34, 45) as do the N-terminal domains (NTDs) of CA from HIV-1 and MLV (18, 35). In contrast to studies on isolated domains, detailed structural information on intact Gag polyproteins has not been forthcoming, presumably due to their larger size and conformational heterogeneity.

We have previously analyzed the properties of purified, recombinant unmyristoylated HIV-1 Gag protein. The free protein is soluble, but assembles spontaneously into VLPs upon addition of nucleic acid (3, 4). In the absence of nucleic acid, the protein is in monomer-dimer equilibrium in solution; the dimerization interface was localized to the C-terminal portion (“CTD”) of the CA domain (12). This interface in mature CA protein was previously shown to participate in the assembly of the conical core structure within mature HIV-1 particles (30, 54, 55), and is probably involved in the assembly of Gag proteins into immature HIV-1 particles as well.
Monomeric HIV-1 Gag protein appears to adopt relatively compact conformation(s) in solution (9), or when bound to lipid membranes (10), and biochemical experiments suggest that its N- and C-termini can be quite near each other in 3-dimensional space (12). It is known to contain several unstructured/flexible regions, including the C-terminal regions of MA (32, 53) and CA (19, 55), the SP1 linker (38), and the NC domain (28). Since it is compact in solution but extended in VLPs, assembly must involve a reorganization of the relative positions of its major globular domains.

In the present report, we describe the properties of the recombinant unmyristoylated Gag protein from MLV, a prototypical gammaretrovirus. While immature MLV particles are very similar in overall structure and organization to immature HIV-1 particles (17, 58), we found that monomeric MLV Gag molecules are different in several respects from their HIV-1 counterparts. MLV Gag has a much weaker propensity for inter-protein interactions, and is monomeric under dilute conditions. Moreover, despite being a multi-domain protein, MLV Gag has a rigid and extended structure, with dimensions in solution similar to those in VLPs. Its rodlike shape implies that the linkers between its domains are all relatively rigid. The implications of these observations are discussed.

MATERIALS AND METHODS

Recombinant protein expression and purification:
Beginning with an infectious molecular clone of MLV, expression plasmids for MLV Gag, mutants, and chimeras (Fig. 1a) were generated in pET3xc vectors by standard
molecular cloning protocols and expressed in BL21-CodonPlus-(DE3) RPIL E.coli (Stratagene). HIV-1 Gag protein was purified using phosphocellulose chromatography as described earlier (11,12). Plasmids containing MLV/HIV-1 chimeras were a kind gift from Dr. Michael Emerman, Fred Hutchinson Cancer Research Center. The mutant MLV 18PA has 18 of its proline codons (at positions 92-94, 105-107, 109, 110, 124, 125, 161-163, 169-171, 203, and 204) replaced with alanine codons. Cells expressing proteins were grown to an OD of ~0.8 and induced with 0.4 mM IPTG for 5 hours at 37°C. The purifications of the MLV Gag proteins and chimeras were similar to that of HIV-1 Gag protein (11, 12). Briefly, MLV Gag was fractionated from the cell lysate by addition of saturated ammonium sulfate to 40% saturation. The precipitated proteins were resolubilized in 20 mM Tris HCl pH 7.4, 0.5 M NaCl, 5 mM DTT, and purified by phosphocellulose (PC) affinity chromatography as described for HIV-1 Gag (11, 12). The PC-purified MLV Gag proteins were further purified by size exclusion chromatography (SEC) on a Superose 12 column (GE Health Care), equilibrated in 20 mM Tris HCl, pH 7.4, 0.5 M NaCl, 5 mM DTT, and stored at -80°C after addition of glycerol to 10% v/v. For sedimentation equilibrium (SE) experiments and small angle x-ray scattering (SAXS) measurements, the protein was further purified on an SP sepharose column (GE healthcare), and dialyzed into 20 mM Tris HCl pH 8.0, with 0.5 M NaCl and 1 mM (tris(2-carboxyethyl)phosphine) (TCEP).

In vitro assembly
MLV Gag at 1 mg/ml in 20 mM Tris HCl pH 7.4, 0.5 M NaCl, 5 mM DTT was incubated with 10% w/w yeast tRNA, and subjected to overnight dialysis at 4°C against 20 mM Tris HCl pH 8.0/0.1 M NaCl/5 mM β-mercaptoethanol (BME) (assembly buffer).
Assembled VLPs in the dialysate were visualized by electron microscopy (EM) after negative staining with 2% uranyl acetate.

**Analytical ultracentrifugation**

All sedimentation equilibrium (SE) experiments were carried out at 4 °C in an Optima XL-A analytical centrifuge (Beckman-Coulter Instruments). Protein samples were in 20 mM Tris HCl, pH 7.4/0.5 M NaCl/1 mM TCEP. For SE, 12 mm cells were loaded with 180 µL protein while 3 mm cells were loaded with 50 µL, and centrifuged to equilibrium at 8K, 10K, and 15K rpm. Data was acquired at 280 and 254 nm at radial increments of 0.001 cm and with ten repeats. The software Sedenterp [http://www.bbri.org/RASMB/rasmb.html, (26)] was used to estimate the partial specific volumes of proteins and buffer parameters, viscosity (η) and density (ρ). Data analysis was performed using SEDPHAT [http://www.analyticalultracentrifugation.com/sedphat, (46)] as reviewed in (27). The goodness of fit in SEDPHAT is the “reduced chi-square”, $\chi^2$.

Boundary sedimentation velocity (SV) experiments were performed at 45,000 RPM, and 20°C or 4°C, using either 400 µl (12 mm cells) or 100 µl (3 mm cells) of sample. Data acquired at 280 nm and 0.003 cm radial increments were analyzed with SEDFIT (http://www.analyticalultracentrifugation.com/), using previously described computational methodology (47). Briefly, the SV data were subjected to maximum entropy regularization with the confidence level set to $p = 0.95$, for the most parsimonious distribution of sedimenting species. The results of this computational analysis give $c(s)$ versus $s$ plots where $c(s)$ is the concentration of protein divided by the sedimentation coefficient at the respective $s$ position.
Size exclusion chromatography and light scattering

A Rainin HPXL solvent delivery system connected to a Rainin Dynamax UV-1 detector and a Wyatt systems Dawn EOS static and quasi-elastic light scattering (SLS and QELS respectively) detector was used to study the hydrodynamic properties of the proteins on a Superose 12 (GE Healthcare) column. Solvent viscosity of 0.00936 centipoise and a temperature dependence of $-1.95 \times 10^{-4}$ g/cm s K were assumed. Data collected simultaneously from light detectors 5 through 18, with the exception of detector 13, was used for SLS, while detector 13 was modified for QELS measurements.

The column was calibrated with standards (GE Healthcare) including RNase A ($R_h$ 16.4 Å); chymotrypsin ($R_h$ 20.9 Å); ovalbumin ($R_h$ 30.5 Å); BSA ($R_h$ 35.5 Å); aldolase ($R_h$ 48.1 Å); and catalase ($R_h$ 52.2 Å), for determination of $R_h$ from retention time. Retention time was used to calculate the parameter $(K_D)^{1/3}$, which has a linear relationship to the $R_h$ of the eluting species (48). Sample proteins, except for MLV 18PA and M/HIV, were injected onto the column at concentrations of 2-3 mg/ml, such that the concentrations in the eluting peaks ranged from 0.22-0.35 mg/ml. M/HIV and MLV 18PA were injected at ~0.8 mg/ml, resulting in concentrations of ~0.06 mg/ml in the eluting peaks.

SAXS and modeling

Both small-angle and wide-angle x-ray scattering (SAXS and WAXS) were performed at beamlines 12-ID of Advanced Photon Sources (APS) at Argonne National Laboratory (Chicago site). The wavelength, $\lambda$, of x-ray radiation was set as 0.689 Å and the scattered x-ray photons were recorded with a charge-coupled device X-ray detector. A cylindrical quartz capillary x-ray flow cell with a diameter of 1.5 mm and a wall of 10
µm was used. The x-ray beam, with size of 0.1 × 0.2 mm², was adjusted to pass through the center of the cell. The exposure time was set to 0.4-0.6 seconds to avoid detector saturation and radiation damage. Potential radiation damage was further reduced by maintaining a constant flow of the samples. The range of momentum transfer \( q = \frac{4\pi \sin \theta}{\lambda} \), where \( 2\theta \) is the scattering angle\] of SAXS experiments was 0.006-0.260 Å⁻¹, and that of WAXS was 0.1 – 2.6 Å⁻¹. Twenty images were acquired for each sample or buffer solution. The 2-D scattering images of buffers and samples, azimuthally averaged after solid angle correction, were normalized with the intensity of the incident x-ray beam. The resulting 1-D scattering data sets were averaged before buffer background subtraction. The background subtractions were performed as follows. First, the WAXS profile was obtained by using eq. 1:

\[
I_{\text{solute}}(q) = I_{\text{sample}}(q) - \alpha I_{\text{buffer}}(q),
\]

(1)

The value of \( \alpha \), an adjustable parameter, was tuned to eliminate the scattering from buffer, indicated by the disappearance of the solvent peak at ~2.0 Å⁻¹. The resulting WAXS profile was then used as a guide for the SAXS background subtraction, by tuning the value of \( \alpha \) in SAXS subtraction and overlaying the resulting SAXS profile with the WAXS profile at the overlapping \( q \) range, i.e., 0.1 – 0.26 Å⁻¹ in our experiments.

All samples were measured at three concentrations, ranging from 0.6 to 3.2 mg/ml. The final SAXS profiles were obtained from linear extrapolation of data from these concentrations to that of zero concentration. The total scattering profiles were obtained by piecing the resulting SAXS and WAXS data together in the range of 0.006 – 2.5 Å⁻¹.
The radius of gyration (R_g) was calculated from data at low q values in the range of qR_g < 1.2, using the Guinier approximation (eq. 2),

\[ \ln I(q) = \ln(I(0)) - \frac{R_g^2 q^2}{3} \]  

The scattering intensities at and near q = 0 were extrapolated with the Guinier equation. The Kratky plots, q^2I(q) vs q, Fig 4A, indicate that all molecules under study are in folded form.

The pairwise distance distribution function, PDDF or p(r), in real space was calculated using GNOM (50). To avoid underestimation of the molecular dimension and consequent distortion in low resolution structural reconstruction, the parameter R_max, the upper end of distance r, was manually scanned in the range of 2R_g to 6R_g, and chosen such that the resulting PDDFs gradually approach zero at high r.

Molecular envelopes (also known as bead models) were obtained using the program DAMMIF (16), a fast version of DAMMIN (51). In DAMMIF, a spherical space with a radius of R_max, read from the PDDF result, is initially filled with multiple-phase dummy atoms. To avoid distortion caused by possible under-estimation of D_max, DAMMIF can automatically adjust the value of R_max during the calculation. At each step, the envelope evolves by randomly phasing a dummy-atom in (as a part of the molecule) or out (as a part of the solvent). A simulated annealing algorithm drives the envelope evolution, by reducing the discrepancy between the experimental and calculated scattering curves during the annealing process. Thirty-two DAMMIF calculations were performed for each molecule, running in the “slow” mode with default setting of DAMMIF. The resulting structural models were subjected to averaging and the
normalized spatial discrepancy (NSD) values between each pair of models were computed, using DAMAVER (41). The model with lowest average NSD with respect to the rest of models was chosen as the reference model. The remaining models were superimposed onto the reference model using SUPCOMB (24) except that possible outliers identified by NSD criteria were discarded. The dummy atoms of these superimposed models were remapped onto a densely packed grid of atoms with each grid point marked with its occupancy factor. The grids with non-zero occupancies were chosen to generate a final consensus model with the volume equal to the average excluded volume of all the models. Scattering data in a $q$ range of $0-0.22 ~\text{Å}^{-1}$, which reflect the global shape without significant undesired influence from the internal structure, were used in DAMMIF calculations. The goodness of fit, $R_f$ values (51)(http://www.embl-hamburg.de/biosaxs/manual_dammin.html), as defined in the following:

$$R_f^2 = \sum_q \left( \frac{[(sc* I_{\text{mod}}(q) - I_{\text{exp}}(q))^2]}{\sum_q [I_{\text{exp}}(q) * q^2]} \right)^2$$

were in the range of 0.004 – 0.006, indicating good match between the experimental scattering data and the calculated ones for individual models. In eq 3, $sc$ is a scaling factor which makes the experimental data and a given back-calculated profile have the best match in intensities. The average NSD for those bead models were 0.70 – 0.95, which indicate good convergences in both individual DAMMIF fits and overall bead model ensembles for each sample.

The averaged molecular envelopes/bead models were further smoothed using the program Situs (56) (version 2.4) and the smoothed molecular envelope images shown in Fig. 4 were produced using program UCSF Chimera (42) (version 1 build 2540).
RESULTS

In Vitro assembly of MLV Gag

The proteins used in our studies, including MLV Gag, WM HIV-1 Gag (a monomeric mutant of HIV-1 Gag (12)), and a series of related proteins, are depicted in Fig. 1A. They were purified from *E. coli* lysates as described in Materials and Methods. Their purity is indicated by the Coomassie brilliant blue-stained SDS-PAGE gel of several of the proteins, shown in Fig. 1B.

MLV Gag was tested for its ability to assemble into VLPs in a defined *in vitro* system. As in prior studies with HIV-1 Gag, the protein was dialyzed at 1 mg/ml into assembly buffer in the presence of 100 g/ml of yeast tRNA at 4°C. Under these assembly conditions, roughly spherical VLPs were formed, as visualized by negative stain EM (Fig. 1C) and transmission EM of thin sections (not shown). These VLPs had a median diameter of 95 nm with a standard deviation of 4.5 nm (Fig. 1D), similar to that of immature VLPs assembled in cells (58). The VLPs have a somewhat irregular appearance, and thin sections of these VLPs made after high speed centrifugation often show “collapsed” structures (not shown), suggesting that the VLPs are fragile.

Oligomerization of MLV Gag in Solution

Recombinant HIV-1 Gag protein is in monomer-dimer equilibrium in solution, with a $K_d \approx 0.5 \times 10^{-5}$ M (12). We tested for oligomerization of MLV Gag in several ways. MLV Gag at concentrations ranging from 0.15 mg/ml (i.e., 2.5 µM) to 2 mg/ml (33 µM) was subjected to SV analysis. Boundary sedimentation measurements (not shown)
were performed to determine the weight-average sedimentation coefficient of MLV Gag and the data analyzed by the Sedfit program. Fig. 2A shows the c(S) vs. S resulting from this analysis. The profiles show a predominant single peak with an S value of 2.85 at all three concentrations, at 20°C. In contrast, the peak of the weight-average S value for HIV-1 dp6, which dimerizes with a K_d of ~ 5 µM, shifts from 2.8 to 4.2 as the concentration is increased from 0.5 µM to 60 µM (Fig. 2B). The weight-average S value for HIV-1 Gag at intermediate concentrations of 2 µM and 20 µM was 3.2 and 3.8 respectively (not shown). The invariance of the MLV Gag S value across the concentration range tested suggests that the oligomeric composition of MLV Gag is not significantly altered under these conditions. The S value of 2.83 (S_{20,w} 3.15) for MLV Gag, whose molecular mass is 6.0 x 10^4 Da, represents an f/f_0 of 1.85 and a R_h of ~ 47 Å (Table 1), suggesting that the molecule is highly asymmetric or extended. MLV dp12 and CANC were also subjected to SV analysis (data not shown). MLV dp12 had an S value of 2.73 (S_{20,w} 3.04) and an f/f_0 of 1.62, while CANC had an S value of 2.31 (S_{20,w} 2.57) and an f/f_0 1.64, indicating that these portions of MLV Gag also are extended and/or asymmetric.

As a more rigorous test for weak self-association, SE analysis of MLV Gag in solution was performed, and the data fitted to different models using SEDPHAT (46). MLV Gag at concentrations of 2, 6 and 20 µM was centrifuged to equilibrium at 8,000, 10,000 and 15,000 rpm. The SE profiles obtained were first modeled assuming the presence of a single non-interacting species, with a MW of 60 kDa. This resulted in a fit with a global reduced chi-square \( \chi^2 = 1.823 \), However, allowing the MW to float while assuming the presence of a single species resulted in a predicted MW of 68 kDa and a
better fit to the data ($\chi^2 = 1.346$). This indicated that the MLV Gag protein might have a propensity for weak association. Since HIV-1 Gag protein has previously been shown to dimerize (12), the self-association of MLV Gag was initially modeled as a monomer-dimer equilibrium. All nine profiles were analyzed to obtain a global fit with a monomer-dimer association model, assuming the monomer mass to be 60,000 Da. The top panel in Figure 2C shows only three profiles (20 µM sample) of nine (for clarity). The continuous line through each data set is the non-linear regression fit to the data, yielding a $K_a$ of $2.45 \times 10^3$ M$^{-1}$ for dimeric association of the monomers, with a global reduced chi-square $\chi^2 = 1.32$ for the fit. The bottom panel shows a plot of the respective residuals for each fit shown above. It is evident that the residuals are all close to zero and mostly randomly distributed. The determined association constant corresponds to a $K_d$ of ~400 µM. The same data was also subjected to a monomer-trimer model for self association (not shown). Here the best fit obtained was for a $K_a$ of $3.63 \times 10^7$ M$^{-2}$, with a global reduced chi-square $\chi^2 = 1.37$. Such analysis was performed with SE data obtained using three separately purified batches of protein, and we have obtained $K_d$ values for monomer-dimer equilibrium ranging from ~200-500 µM (data not shown). While the very weak interactions observed with MLV Gag do not allow us to determine the strength or mode of interaction unambiguously, these results demonstrate that MLV Gag does self-associate, but far more weakly than HIV-1 Gag.

Conformation of MLV Gag in Solution

We characterized MLV Gag (~60 kDa) by size exclusion chromatography (SEC). It elutes (red profile) from Superose 12 more rapidly than WM HIV-1 Gag (~ 50 kDa, pink profile), Fig. 3A. In fact, it elutes at a position comparable to that of a dimer of BSA.
(~130 kDa, black profile) (a shallow peak before the BSA monomer). The elution position of MLV Gag was constant over injection concentrations ranging from 4 to 25 µM (data not shown). MLV dp12 (~ 52 kDa, blue profile) also elutes well before BSA (~67 kDa), while MLV CANC (~37 kDa, green profile) elutes slightly before monomeric WM HIV-1 Gag (~50 kDa). The high ionic strength of the buffer precludes non-specific interactions with the column matrix. The eluate was simultaneously monitored by both SLS, to determine the molecular mass of the eluting material, and QELS, as a measure of the Rₙ of the eluting species. The molar mass for MLV Gag measured by SLS (Fig. 3) was 6.07 x 10⁴ Da, while those for MLV dp12 and MLV CANC were 5.4 x 10⁴ Da and 4.0 x 10⁴ Da, respectively, confirming that the proteins were predominantly monomeric under these conditions. Under identical conditions the molar mass of BSA was estimated to be 6.47 x 10⁴ Da, while that of WM HIV-1 Gag was 4.89 x 10⁴ Da. The unusual elution behaviors of MLV Gag and derived proteins on the size exclusion column were therefore a result of their conformations and not their oligomeric status. The QELS data gave Rₙ values of 48, 42, and 37 Å for MLV, dp12 and CANC proteins, respectively, while that for WM HIV-1 Gag was 38 Å (Table 1).

As a completely independent approach to determining the conformation of these proteins, we performed SAXS analysis. SAXS data for MLV Gag and CANC were first analyzed to generate Kratky plots, q²I(q) vs. q (Fig 4A), as described in Materials and Methods. The shape of the Kratky plot is an indicator of the “unfolded-ness” of a protein. In the case of multidomain proteins, the plot is also indicative of how domains are extended or packed and has been used as an indicator of global protein conformation. Specifically, folded proteins have a characteristic peak at low q, followed by a drop in the
curve at high $q$, while unfolded proteins do not have the characteristic peak and increase monotonically with $q$ (43). As Gag is a multidomain protein, the Kratky plots are well suited for characterization of the global domain organization. Fig 4A indicates that while both MLV Gag and CANC are folded, MLV Gag has a more extended conformation with some disorder.

The pairwise distance distribution function (PDDF or $p(r)$) plots, i.e., histograms of distances between scattering atoms in the protein, give an indication of the shape and a measure of the dimensions of the protein. The $p(r)$ functions for MLV Gag and CANC proteins calculated using GNOM (Fig. 4B) are asymmetric, with a tail at high $r$ values. This is a feature of elongated molecules; the $p(r)$ of globular proteins is symmetrical about the peak. The $D_{max}$, derived from the distance $r$ at which $p(r)$ drops to zero, is indicative of the longest dimension in the molecule. This was estimated to be $\sim 210 \text{ Å}$ ($\pm 10 \text{ Å}$) for MLV Gag and $150 \text{ Å}$ ($\pm 10 \text{ Å}$) for MLV CANC. The radius of gyration, $R_g$, determined by SAXS for MLV Gag (57.8 Å), and that of MLV CANC (41.1 Å), were both higher than that of WM HIV-1 Gag (38 Å, not shown), or for WM HIV-1 Gag previously determined by small angle neutron scattering as $\sim 34 \text{ Å}$ (9, 10).

Bead models of the proteins, derived from the SAXS data using DAMMIF (as described in detail in Materials and Methods), are shown in Fig. 4C and D. It is striking to see that ab-initio modeling of SAXS data yields a rod-shaped model for MLV Gag, with a long axis of the molecular envelope estimated at $\sim 201 \text{ Å}$. Similarly, the SAXS data show that CANC is evidently elongated as a rod $\sim 150 \text{ Å}$ long. The width of the bead model at different regions for MLV Gag ranges from 35-45 Å, with a narrow dimension of 22 Å. These dimensions are common to both models, as are the presence of
“kinks” or bends along the length. The tentative superimposition of molecular envelopes of MLV Gag and CANC with the known crystallographic structures of MLV MA (45) and the NTD of MLV CA (35), is displayed in Fig. 5.

Structural Features of MLV Gag

We examined the sequence of MLV Gag in an effort to identify possible features contributing to its extended conformation. We noted that both the C-terminal half of the MA domain and the N-terminal half of the p12 domain are very proline-rich, and in fact contain several short runs of prolines (Fig. 6A). It seemed possible that these runs might form short polyproline helices or other such extended structures, and thus contribute to the rigidity of the protein. To test this possibility, we mutated 18 proline codons (boxed in Fig. 6A) to alanine codons. The resulting mutant Gag protein, MLV 18PA, has a molecular mass of 60,389 Da, very similar to that of MLV Gag (60,858 Da). However, its elution from an SEC column is clearly delayed (Fig. 6B), indicating a decrease in $R_h$ from 54.6 to 52.1 Å (Fig. 6B, Table 1). These data suggest that the runs of prolines in the C-terminal half of MA and the N-terminal half of p12 are at least in part responsible for the rigidity of the MLV Gag protein.

Another striking feature of the MLV Gag sequence is an extraordinary run of charged residues near the C-terminus of the CA domain, termed the “electric wire” or “charged assembly helix motif” (6). Cheslock et al. (6) presented genetic evidence suggesting that this region forms an $\alpha$-helix during assembly. They also showed that deletion of 22 residues (outlined in Fig. 6A) is compatible with proper assembly, and indeed with infectivity, of the virus. We determined the $R_h$ of MLV proteins lacking these 22 residues (designated “EW22D” in Cheslock et al.). EW22D MLV Gag elutes slightly
after the WT but before MLV 18PA (Fig. 6B). We also tested the effect of the EW22D deletion in the context of dp12 and CANC proteins (QELS data not shown). As indicated in Table 1, both of these changes significantly decrease the $R_h$ of these smaller proteins, as well as that of full-length Gag. Therefore, the charged stretch in the C-terminal domain of CA also contributes to the extended structure of MLV Gag.

In light of the large differences in molecular dimensions between MLV Gag and HIV-1 Gag (Fig. 3, Table 1), it was of interest to determine the contribution of individual domains to these differences. Emerman et al. have previously characterized viruses in which the MA domains of HIV-1 and MLV Gags have been exchanged (14, 15). These chimeric Gag proteins were purified and analyzed by SEC. As shown in Table 1, replacement of the MA domain of either Gag with the heterologous MA domain did not significantly affect the $R_h$ of either parent protein.

DISCUSSION

Gag proteins of different retroviruses assemble into very similar VLPs in vivo (2, 3, 13). Since the information content for the assembly of VLPs exists within the Gag polyprotein (5), a comparison of the properties of Gag from viruses from different genera might give insight into common underlying mechanisms of assembly.

We have previously characterized the solution properties of unmyristoylated HIV-1 Gag protein. Briefly, it contains a dimer interface in the CTD of its CA domain; the monomer-dimer equilibrium in solution has a $K_d$ of $\sim 5 \times 10^{-6}$ M (12). Further, monomeric HIV-1 Gag appears to adopt relatively compact conformation(s) in solution (9). It assembles into small, thin-walled VLPs (4) while presumably still in a compact
conformation. An extension in length to ~20 nm can be observed *in vitro* when both ends
are supplied with their preferred ligands, i.e., inositol phosphates or lipids, and nucleic
acid (3, 10).

We now describe the properties of the unmyristoylated MLV Gag protein. Remarkably, it is different from HIV-1 Gag protein in each of these respects. Thus, its
tendency to oligomerize in solution is barely detectable: we estimate that the affinity of
monomers for each other is at least 40-fold weaker than in the case of HIV-1 (Fig. 2).
Further, it exhibits a rod-like conformation in solution, apparently similar to that of the
Gag monomers in immature VLPs; this conformation was evident from SV data (Fig.
2A); SEC (Fig. 3); and SAXS (Fig. 4). This extended conformation presumably enables it
to assemble into full-size VLPs upon addition of nucleic acid alone (Fig. 1C and 1D), and
without a major change in its overall dimensions.

MLV and HIV-1 Gag are similar to each other with respect to the secondary and
tertiary structures of the MA, NTD, and CTD regions (18, 20, 22, 32, 35, 45); however,
the linkers between these regions are apparently unstructured/flexible in HIV-1 Gag but
rigid in MLV Gag. The fact that exchanging MA domains between the two proteins has a
negligible effect on Rs (Table 1) is consistent with this assumption.

What might be responsible for the extended conformation of MLV Gag?
Examination of its sequence reveals two notable features not present in HIV-1 Gag (Fig.
6A). First, MLV Gag is quite proline-rich: 11.9% of its residues are prolines, compared
with 5% in HIV-1 dp6, and 5% in the protein database (33). Further, 28% of the amino
acids between positions 90 and 214 (i.e., the C-terminal region of MA and p12) are
proline residues. They are often found as short runs of 2-4 consecutive prolines; such
runs frequently have an extended conformation in known protein structures (49). In fact, we found that a change of 18 prolines in MA and p12 to alanines, which in themselves favor helical structures, results in a detectable decrease in the $R_h$ of the protein (Fig. 6B, Table 1).

It has previously been reported that the p12 domain of MLV Gag (which has no obvious counterpart in HIV-1 Gag) lacks a unique structure (25); it seems likely that the short runs of prolines can assume an ensemble of structures, but that these are predominantly extended. We also noted that deletion of p12 (84 aa) causes a drop of $R_h$ from ~50 Å to 43 Å, while a further deletion of MA (130 aa) to generate CANC reduces the $R_h$ from 43 Å to 39 Å (Table 1). The larger change in $R_h$ due to a deletion of p12 would suggest that this domain, even if disordered, might adopt relatively extended conformation(s) keeping the globular domains of MA and CA far apart.

Second, the C-terminal end of the CA domain contains a remarkable run of ~33 charged residues. This region, called the “charged assembly helix motif” or “electric wire”, carries no net charge; it is critical for proper VLP assembly (6). Like MLV Gag, MLV CANC displays unusual mobility in SEC: although its MW is only 37.4 kDa, it elutes slightly before monomeric HIV-1 Gag (50.2 kDa) (Fig. 3). The P(r) plot from the SAXS data (Fig. 4B), as well as the $R_g$ of ~41 Å from SAXS and $R_h$ of ~39 Å from SEC and QELS (Table 1) for CANC, all suggest that it also has an extended structure. Since retroviral NC domains are very small and largely unstructured (28), it seemed possible that the “electric wire” region might be responsible in part for the extended structure of MLV CANC and Gag proteins. We found that deletion of 22 of these residues in either MLV Gag, dp12 or CANC caused a detectable decrease in the $R_h$ of the proteins (Figs.
6A and 6B, Table 1). This was particularly apparent in context of the smaller CANC
construct.

What is the biological significance of these features? Interestingly, several
mutants in which individual proline tracts in p12 are replaced with alanine tracts retain
infectivity (29, 59), as does the electric-wire deletion that we studied (6). It thus appears
that the rigidity and/or length of the monomer can be compromised somewhat without
loss of function \textit{in vivo}. In addition, several observations suggest that VLP assembly may
be more dependent upon Gag-RNA interaction in MLV than in HIV-1. First, deletion of
the NC domain, the primary RNA-binding domain of Gag, modestly reduces VLP
assembly \textit{in vivo} in HIV-1 (39) but virtually eliminates it in MLV (36). Second, digestion
of detergent-stripped VLPs with \textit{RNase} disrupts MLV VLPs (37) but not HIV-1 VLPs
(3). It seems plausible that the stronger interaction between free HIV-1 Gag molecules,
relative to MLV Gag molecules, renders its lattice structure less dependent upon nucleic
acid-binding for stability.

On the other hand, HIV-1 MA is known to bind RNA, and it has been suggested
that this is responsible for the ability of HIV-1 Gag to assemble with no NC domain (1,
31, 40, 44). Several reports show that in HIV-1 Gag, the MA domain participates along
with the NC domain, in interactions with RNA (7, 23). It will be of interest to learn
whether similar phenomena are observed in MLV Gag, in which the MA and NC
domains are probably always physically distant from each other.

In conclusion, the “building blocks” for HIV-1 and MLV virus particles are
surprisingly different from each other. Unlike HIV-1 Gag, MLV Gag in solution has
dimensions comparable to that in VLPs; the protein would undergo very little, if any,
change in global conformation during the process of assembly. This is a striking example of the contrasts in fundamental properties between members of different retroviral genera.
ACKNOWLEDGMENTS

The authors thank Demetria Harvin for technical assistance and Richard Frederickson for help with illustrations. SD gratefully thanks Ritu Kanwar for many fruitful scientific discussions. This work was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, and in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN26120080001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.
Table 1. Rₜ Values for Proteins in This Study (Å)

<table>
<thead>
<tr>
<th>Method</th>
<th>MLV</th>
<th>H/MLV</th>
<th>MLV 18PA</th>
<th>EW22D dp12</th>
<th>EW22D dp12</th>
<th>CANC</th>
<th>EW22D CANC</th>
<th>WM</th>
<th>M/HIV</th>
<th>BSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEC</td>
<td>54.6</td>
<td>54</td>
<td>52.1</td>
<td>53.3</td>
<td>45.5</td>
<td>44.3</td>
<td>39.2</td>
<td>37.2</td>
<td>36</td>
<td>38.3</td>
</tr>
<tr>
<td>QELS</td>
<td>48</td>
<td>47</td>
<td>N.D.</td>
<td>47</td>
<td>42</td>
<td>39</td>
<td>37</td>
<td>34</td>
<td>38*</td>
<td>N.D.</td>
</tr>
<tr>
<td>SV</td>
<td>47</td>
<td>N.D.</td>
<td>N.D.</td>
<td>41.3</td>
<td>N.D.</td>
<td>36.4</td>
<td>N.D.</td>
<td>41*</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
</tbody>
</table>

* - from (9)


Atomic resolution structure of Moloney murine leukemia virus matrix protein and its 
relationship to other retroviral matrix proteins. Structure (Camb) 10:1627-36.


Schuck, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity 

ratios of proteins in impure systems by use of gel filtration and density gradient 
centrifugation. Application to crude preparations of sulfite and hydroxylamine 

Protein Sci 8:587-95.


Svergun, D. I. 1999. Restoring low resolution structure of biological macromolecules 
from solution scattering using simulated annealing. Biophys J 76:2879-86.


Tang, C., Y. Ndassa, and M. F. Summers. 2002. Structure of the N-terminal 283-

Functional surfaces of the human immunodeficiency virus type 1 capsid protein. J Virol 
77:5439-50.


FIGURE LEGENDS

Figure 1. (A) Schematic representation and molecular weights of Gag proteins studied in this work, showing domain organization. (B) SDS PAGE gel of purified Gag proteins, stained with Coomassie brilliant blue. WM HIV-1 Gag, a monomeric mutant of HIV-1Gag. (C) Negative stain EM of VLPs assembled from MLV Gag \textit{in vitro}. Scale bar, 100nm. (D) Histogram of the size distribution of \textit{in vitro} assembled MLV Gag VLPs.

Figure 2. Oligomeric properties of MLV Gag protein. (A) $c(s)$ analysis of boundary sedimentation data for MLV Gag protein at concentrations between 2.5 and 33 uM. (B) $c(s)$ analysis for boundary sedimentation data for HIV-1 dp6 protein at concentrations ranging from 0.5 to 60 uM. (C) SE analysis of Gag oligomerization. Solutions of Gag, at 2, 6 and 20 μM, were centrifuged to equilibrium at 8,000 (triangle), 10,000 (circle), and 15,000 rpm (diamond). The data-points were globally fit to a model of dimeric association with a $K_a$ of $2.45 \times 10^3$ M$^{-1}$ and a global reduced chi-square $\chi^2 = 1.32$. For clarity, the figure shows only the 3 A$_{280}$ profiles and residuals for the 20 uM protein sample. The best fit rms error is 0.0055.

Figure 3. Size exclusion chromatography and SLS of MLV and HIV-1 Gag proteins. MLV (red), dp12 (blue) CANC (green) and WM HIV-1 Gag (pink) Gag solutions, as well as BSA (black) at 1.5 mg/ml, were chromatographed on a Superose 12 column.
Elution of protein was simultaneously monitored by $A_{280}$, giving rise to the elution profiles, and by SLS, yielding the data-points above the elution profiles. The points indicate the molar mass of the protein in each fraction. These masses in turn were averaged over the breadth of each peak, producing the roughly horizontal lines above the elution profiles.

Figure 4. SAXS analysis of MLV Gag and CANC proteins. (A). Kratky plots for MLV Gag (yellow) and CANC (Blue). (B). P(r) analysis of SAXS data for MLV Gag (Red) and CANC (Blue), (C). Bead model, shown in molecular surface mode, generated from SAXS data using DAMMIF, as described in detail in Materials and Methods, for MLV Gag (D). Bead model for CANC. The data indicate that both proteins are rod-shaped, with maximum dimensions of ~200 Å and 150 Å respectively. The Normalized Spatial Discrepancy (NSD) scores of both the models are less than 0.7, suggesting a good consistency among all models and an excellent fit to experimental SAXS data.

Fig 5. Molecular envelope in mesh of MLV Gag generated by molecular modeling of SAXS data, as described in Materials and Methods. A tentative superposition of the derived molecular envelopes for MLV Gag (green) and MLV CANC (gold), with crystallographic structures for MA (1MN8) and NTD of CA (3BP9).

Figure 6. Hydrodynamic characterization of mutant MLV Gag proteins. (A) Primary amino acid sequence of MLV Gag. Domains are colored as follows: MA (red), p12 (green), CA (purple), NC (black). The proline residues in MA and p12 domains changed
to alanine in MLV P18A Gag are boxed. The 22 amino acid residues deleted in the “electric wire” region of CA to generate E22D are outlined. (B) SEC and DLS of mutant MLV Gag proteins. MLV Gag (red), EW22D (blue), MLV 18PA (green), and BSA (black) were chromatographed on Superose 12, and Rh estimated by in-line DLS. (C) Retention volume used to estimate Rh for proteins, using proteins with known Rh as calibrants(48).
A

MLV (60,858 Da)

| MA | p12 | CA | NC |

MLV dp12 (52,162 Da)

| MA | CA | NC |

MLV CANC (57,397 Da)

| CA | NC |

HIV dp6 (50,169 Da)

| MA | CA | p2 | NC | p1 |

H/MLV (60,940 Da)

| MA | p12 | CA | NC |

M/HIV dp6 (50,087 Da)

| MA | CA | p2 | NC | p1 |

B

C

D

No. of Particles

Outer Diameter (nm)