Detection of Late Intermediates in Virus Capsid Assembly by Charge Detection Mass Spectrometry

Elizabeth E. Pierson, David Z. Keifer, Lisa Selzer, Lye Siang Lee, Nathan C. Contino, Joseph C.-Y. Wang, Adam Zlotnick, and Martin F. Jarrold

INTRODUCTION

Virus capsids containing hundreds of subunits have evolved to assemble with high fidelity in a short period of time. In many cases, capsid formation is sufficiently robust that, given the right conditions, purified capsid proteins spontaneously assemble into icosahedral virus-like particles. Despite a large effort, the underlying principles of capsid assembly are far from fully understood. The geometric control shown in forming a virus capsid far exceeds our ability to control assembly in man-made chemical systems. Thus, understanding virus assembly at a molecular level will not only increase our knowledge of a process of great biological and medical importance, but it will also help to develop new self-assembly strategies for materials science.

Capids assembly reactions necessarily have a single starting point of bulk subunit and a single ending point of complete capsid. The number of possible intermediates grows combinatorially from both ends of the assembly landscape. However, the concentrations of almost all the intermediates are exceedingly low during assembly. A small number of key intermediates, which may lead to kinetic bottlenecks, are anticipated at both ends of the assembly reaction, due to nucleation at the beginning and capsid closure at the end. The early intermediates are expected to be small fragments of the capsid while the late ones are expected to be almost complete capsids that are missing just a few subunits. Recent experiments have begun to identify early intermediates, however, whether late-assembly bottlenecks exist has remained unknown because it has not been possible to detect and identify them.

The assembly of hepatitis B virus (HBV) is of particular interest because it is a devastating pathogen and because it is an attractive target for the development of new assembly-directed antiviral molecules. Around 360 million people suffer from chronic HBV infection, which kills around 600,000 people annually. HBV self-assembly is usually studied with the core protein assembly domain, a 149-residue protein, Cq149. The building block for the HBV capsid is the Cq149 dimer. Under the right conditions, the dimers spontaneously assemble to yield icosahedral capsids with 90 dimers (in a T = 3 structure) and 120 dimers (in T = 4). Icosahedral capsids are usually described using a triangulation number (T), which gives the number of monomers (60T) in the capsid. The 60T monomers are arranged into 12 pentamers and 10(T−1) hexamers. The interdimer association energy for HBV is only about −15 kJ/mol at physiological temperature and ionic strength, but this supports formation of globally stable capsids because Cq149 dimers are tetravalent. A stronger association energy, which can be contrived by high salt assembly, results in kinetic trapping of intermediates on the assembly pathway because the high frequency of nucleation events and minimal dissociation of partial capsids deplete free subunits. In this work we analyze the trapped intermediates by charge detection mass spectrometry (CDMS) and cryo-electron microscopy (cryo-EM). These two techniques are complementary: CDMS provides information about the

ABSTRACT: The assembly of hundreds of identical proteins into an icosahedral virus capsid is a remarkable feat of molecular engineering. How this occurs is poorly understood. Key intermediates have been anticipated at the end of the assembly reaction, but it has not been possible to detect them. In this work we have used charge detection mass spectrometry to identify trapped intermediates from late in the assembly of the hepatitis B virus T = 4 capsid, a complex of 120 protein dimers. Prominent intermediates are found with 104/105, 110/111, and 117/118 dimers. Cryo-EM observations indicate the intermediates are incomplete capsids and, hence, on the assembly pathway. On the basis of their stability and kinetic accessibility we have proposed plausible structures. The prominent trapped intermediate with 104 dimers is attributed to an icosahedron missing two neighboring facets, the 111-dimer species is assigned to an icosahedron missing a single facet, and the intermediate with 117 dimers is assigned to a capsid missing a ring of three dimers in the center of a facet.
number of dimers in the trapped intermediates, and class averages from cryo-EM provide information about their structures.

Over the past few years, mass spectrometry (MS) has emerged as a powerful tool to investigate the composition of biological assemblies, including viruses. Stockley, Ashcroft, and co-workers have used MS to probe the early steps in the assembly of the MS2 bacteriophage capsid, and Heck and collaborators have investigated the early stages of HBV capsid assembly. In conventional MS the mass-to-charge ratio ($m/z$) spectrum is measured, and for large multiply charged ions the charge must be deduced from the charge state ensemble to determine the mass. This approach starts to become problematic for masses around 100 kDa. Though the $m/z$ peaks are usually still resolved in this range, they broaden and shift due to salt addition and complex formation. The $m/z$ peaks often do not shift by the same amount, making it difficult to assign the charge state. In some cases, the charge state resolution can be improved by purification, or the mass can be determined because the analyte identity is known. However, as the mass and heterogeneity increase, the problems are compounded until it is no longer possible to determine the mass from the $m/z$ spectrum.

Here we have used charge detection mass spectrometry (CDMS), a single-particle technique where the mass is directly determined from $m/z$ and $z$ measurements for each ion. With this approach, masses can be measured for heterogeneous mixtures of large ions that resist analysis by conventional MS. CDMS was first used to measure the masses of micrometer-sized metal particles in 1960. The application of CDMS to electrospray ions was described by Fuerstenau and Benner in 1995, and the use of CDMS to measure the masses of intact viruses was reported in 2001. While groundbreaking, the mass resolution achieved in these early studies was rather poor (the peak for rice yellow mottle virus with a mass of $\sim 6.5$ MDa had a full width at half-maximum of $\sim 12$ MDa). Consequently, the technique received little attention. In recent papers, we have described a number of key improvements to CDMS that have led to large gains in the mass resolution and the limit of detection. Without these improvements, the measurements reported here would not have been possible.

**METHODS**

**Charge Detection Mass Spectrometry.** Mass spectra were measured using a home-built charge detection mass spectrometer that is described in detail elsewhere. Briefly, electrosprayed ions enter the instrument through a heated capillary. They are separated from the ambient gas flow by three differentially pumped regions incorporating an RF ion funnel, an RF hexapole, and an RF quadrupole. Ions are extracted from the quadrupole, accelerated through a 100 V potential, and focused into the entrance of a dual hemispherical energy analyzer. The energy-selected ions are then focused into a modified cone trap that contains the charge detection cylinder. Ions oscillate in the trap, passing back and forth through the charge detector, for 129 ms. A charge-sensitive preamplifier incorporating a liquid nitrogen-cooled FET detects the image charge. The amplified signal is digitized and transmitted to a computer where the signals are analyzed using a fast Fourier transform (FFT). The fundamental frequency from the FFT is used to derive the $m/z$ of the ion and the magnitude of the FFT is used to obtain its charge ($z$). The masses determined for each ion are then binned to yield the mass spectrum. Results for ions trapped for less than the full 129 ms are discarded. Figure S1 in Supporting Information (SI) shows an example of the charge detection signal and its FFT. Though a rare event, some ions lose charge while trapped. The loss of charge usually results in a sudden change in the ion’s oscillation frequency. In these cases, the ions are rejected by the data analysis program.

The measured masses are systematically overestimated by around 1% due to the presence of adducts (the addition of small molecules and counterions to the virus capsid) and imperfect calibration of the $m/z$ and $z$ measurements. The relationship between the oscillation frequency of an ion in the trap and the $m/z$ is determined from SIMION simulations and the charge is calibrated by introducing test charges, ranging from approximately 1500–10000 elementary charges ($e$), through a capacitor. Though the calibration is performed with high charges, we have checked that it is valid down to charges as low as 7 $e$ by measuring the $m/z$ spectrum of cytochrome $c$ and ubiquitin with CDMS. In this $m/z$ range ($<1500$ Da/$e$), we were able to resolve individual charge states. This allows us to compare the charge deduced from the $m/z$ spectrum (i.e., $m/z$ charge) to the image charge (i.e., measured charge). For the $+7$–$18$ charge states, the $m/z$ charge and the measured charge were in agreement. These results will be published elsewhere. To correct for the presence of adducts and small inaccuracies in $m/z$ and $z$, the measured masses were calibrated to the $T = 3$ and $T = 4$ peaks of HBV using the sequence mass for the Cp149 monomer. The presence of two internal standards helps ensure the accuracy of the mass scale.

The mass resolution is limited by the uncertainties in the $m/z$ and $z$ measurements. With the 129 ms trapping time used here, the absolute root-mean-square (rms) deviation of the measured charge from its true value is expected to be 1.2 $e$. The kinetic energy distribution of the ions entering the trap limits the $m/z$ resolution. The hemispherical energy analyzer was operated under low resolution conditions in these experiments, where the relative rms deviation of the $m/z$ is around 0.010. The relative error in $m/z$ is constant for all measurements; the absolute error in charge is constant for a given trapping time. Combining the contributions from the charge and the $m/z$, the relative rms deviation of the mass for the experiments reported here is expected to be around 0.013–0.014. Because the ions are highly charged (>100 $e$) and they are trapped for a long time, the chance of a false positive is essentially zero.

**Sucrose Gradient Centrifugation and Cryo-EM Sample Preparation.** Assembly of 40 $\mu$M (1.3 mg/mL) Cp149 dimer in 50 mM HEPES pH 7.5, was induced by addition of NaCl to 1 M final concentration. The resulting solution was allowed to equilibrate for up to 48 h at room temperature. A 200 $\mu$L aliquot of the assembly reaction was loaded onto a continuous 10–40% (w/v) sucrose gradient, containing 300 mM NaCl in 50 mM HEPES and centrifuged for 5 h at 200000 g. For the cryo-electron microscopy images presented here, the protein layer above the $T = 4$ particle layer, including intermediate and $T = 3$ particles, was removed from the gradient, dialyzed into 300 mM NH$_4$OAc and concentrated to a final concentration of 3.7 mg/mL protein.

**Electron Microscopy.** For cryo-EM specimen preparation, 4 $\mu$L of purified HBV sample from the sucrose gradient was applied to a glow-discharged carbon-coated 300 Cu mesh grid. The sample was left on the carbon film for 25 s, blotted for 4 s, and vitrified by plunge-freezing into liquid ethane using an FEI 2000F Duo-stator. The grid was then picked up and plunge-frozen into liquid ethane using an FEI 2000F Duo-stator.
Vitrobot. The frozen-hydrated samples were visualized using a JEOL-3200 FS electron microscope (EM) equipped with a Gatan 626 cryotransfer system. The EM was operated at 300 kV with an in-column zero-loss energy filter using a slit width of 20 eV to improve image contrast. Grids were viewed under low-dose condition (15–20 e−/Å²) at a nominal magnification of 40000x (corresponding to 3.26 Å/pixel). Images were recorded on a Gatan 4k × 4k CCD camera at a defocus range of 2.0 to 4.0 µm.

**Image Processing.** Particles were semi-manually selected using a 200 × 200 pixel box, centered using EMAN2,42 and then normalized and filtered using Xmipp.43,44 A pool of 5681 particles was subjected to 2-D multireference alignment and classification based on the iterative hierarchical clustering method using CL2D from Xmipp.45 Details about the clustering schemes are given below.

**RESULTS**

To trap intermediates, HBV capsids were assembled under aggressive conditions. Figure 1 shows mass distributions measured by CDMS for HBV capsids assembled for 24 h in 1 M NaCl and then dialyzed into 300 mM NH4OAc. The mass spectrum is a histogram obtained by binning the measured masses for 5893 individual ions with 20 kDa bins. After binning, the histograms were smoothed with a Savitsky-Golay algorithm. The red line shows an expanded view of the region between the T = 3 and T = 4 peaks.

![Figure 1. Mass spectrum measured by CDMS for HBV capsids assembled under aggressive conditions (1 M NaCl) and dialyzed into 300 mM NH4OAc. The mass spectrum is a histogram obtained by binning the measured masses for 5893 individual ions with 20 kDa bins. After binning, the histograms were smoothed with a five point Savitsky-Golay algorithm. The red line shows an expanded view of the region between the T = 3 and T = 4 peaks.](image1)

The blue histogram in Figure 2 shows the optimized Gaussian intensities of the intermediates that provide the best fit to the measured spectrum. The mass scale is in terms of the number of dimers.

![Figure 2. Analysis of the CDMS mass spectrum in Figure 1. The points are the measured mass spectrum and the red curve is a least-squares fit to the points using the model described in the text. The blue histogram shows the intensities of the intermediates that provide the best fit to the measured spectrum. The mass scale is in terms of the number of dimers.](image2)
In addition to the main features discussed above, there is a small signal ~10 dimer units above the $T = 4$ peak that may result from off-pathway assembly into metastable, non-icosahedral structures beyond the $T = 4$ capsid. These species were not detected in the cryo-EM studies discussed below (their abundance is quite small, and they probably have low symmetry, and thus, their absence is not surprising).

HBV capsids were further investigated by cryo-EM, which allows visualization of particles undistorted by staining and drying artifacts. Images were analyzed for evidence of incomplete or irregular particles by hierarchical class averaging. To minimize confusion of authentic $T = 4$ particles with intermediates during EM analysis, the sample was harvested from a sucrose gradient fraction including the $T = 3$ band and extending to, but not including, the $T = 4$ band (red brace in Figure 3a). This way, all intermediate species leading up to the $T = 4$ particles are imaged, with a minimal number of obscuring defects. Scale bar, 50 nm. (c) Selected class averages show a $T = 3$ average, a $T = 4$ average, and three averages with defects, respectively. The red dashed lines indicate regions of weak density on the capsid. Scale bar = 10 nm. The averages in panel (c) correspond to classes 3, 7, 20, 8, and 1, respectively, in Figure S2 of SI.

Figure 3. Characterization of the HBV capsids. (a) HBV capsids separated on a sucrose gradient showing the extracted fraction (red brace) used for cryo-EM analysis. (b) A representative cryo-EM micrograph of a frozen-hydrated HBV specimen. Black arrowhead, $T = 3$ capsid. White arrowhead, $T = 4$ capsid. Red arrowhead, capsid with defects. Scale bar, 50 nm. (c) Selected class averages show a $T = 3$ average, a $T = 4$ average, and three averages with defects, respectively. The masses of the partial capsids along with the missing density indicate that high-molecular weight intermediates can be trapped by assembly in high salt conditions where Cp149-Cp149 association is strong; a strong association energy is predicted by theory to trap intermediates because the high frequency of nucleation events depletes free subunits that are required to complete a nascent particle.29 The prominent trapped intermediates observed in the CDMS spectra must persist because these species are resistant to dimer addition and loss. The most likely explanation for this behavior is that they have a lower free energy than their neighbors. Kinetic trapping is expected to trap intermediates on the assembly pathway. The cryo-EM measurements support this view. If we had observed aberrant structures in the cryo-EM class averages, they would probably be off the assembly pathway. However, we found incomplete $T = 4$ capsids, which indicates that they are on-pathway. Many paths and intermediates are possible; it is surprising we see so few intermediates persist.

DISCUSSION

The CDMS and cryo-EM results presented here all indicate that high-molecular weight intermediates can be trapped by assembly in high salt conditions where Cp149-Cp149 association is strong; a strong association energy is predicted by theory to trap intermediates because the high frequency of nucleation events depletes free subunits that are required to complete a nascent particle.29 The prominent trapped intermediates observed in the CDMS spectra must persist because these species are resistant to dimer addition and loss. The most likely explanation for this behavior is that they have a lower free energy than their neighbors. Kinetic trapping is expected to trap intermediates on the assembly pathway. The cryo-EM measurements support this view. If we had observed aberrant structures in the cryo-EM class averages, they would probably be off the assembly pathway. However, we found incomplete $T = 4$ capsids, which indicates that they are on-pathway. Many paths and intermediates are possible; it is surprising we see so few intermediates persist.
The most prominent trapped intermediates for the \( T = 4 \) capsid appear to be associated with missing icosahedral facets. The intermediates with 104, 111, and 117 dimers are consistent with one assembly path with multiple metastable pause points. However, the 110-mer, missing a pentamer of dimers and its surrounding five dimers, appears to be on a different path.

Finally, trapped intermediates were not observed for the \( T = 3 \) capsid in either the mass spectra or the cryo-EM measurements. The different geometries of the \( T = 3 \) and \( T = 4 \) capsids result in differences in the species that satisfy the stability rules. However, we have not been able to deduce an unequivocal explanation for the lack of \( T = 3 \) intermediates on the basis of the currently available literature.

**CONCLUSIONS**

Late intermediates for the assembly of the HBV \( T = 4 \) capsid are kinetically trapped by assembly under high salt conditions. Intermediates have been detected and identified by charge detection mass spectrometry. Prominent intermediates were found with 104/105, 110/111, and 117/118 dimers. Cryo-EM measurements indicate that the intermediates are incomplete \( T = 4 \) capsids. The detection of incomplete capsids rather than aberrant structures suggests that they are on-path intermediates. From their stability and kinetic accessibility we have proposed plausible structures. The prominent intermediate with 104 dimers is assigned to an icosahedron missing a two neighboring facets. The intermediate with 111 dimers is attributed to an icosahedron missing a single facet, and the 117-dimer species is assigned to a capsid missing a ring of three dimers in the center of a facet. Though assembly conditions used in our experiments are not physiological, we detected intermediates that are kinetically accessible, thermodynamically stable and that may help define the paths of HBV assembly.

**ASSOCIATED CONTENT**

Six additional figures. This material is available free of charge via the Internet at http://pubs.acs.org.

**AUTHOR INFORMATION**

Corresponding Authors

azlotnic@indiana.edu (A.Z.)
mfj@indiana.edu (M.F.J.)

**Notes**

The authors declare the following competing financial interest(s): The authors except A.Z. declare no competing financial interests. A.Z. is associated with a company that is developing anti-viral compounds.

**ACKNOWLEDGMENTS**

We gratefully acknowledge the support of the NSF through Award Number 0832651 to M.F.J. and the NIH through Award Number R01-AI077688 to A.Z. and a subcontract to A.Z. from Award Number 0832651 to M.F.J. and the NIH through Award Number R01-GM100071 to Stephen Jacobson. E.E.P. and D.Z.K. were supported by Indiana University Quantitative and Chemical Biology Training Program Fellowships. Electron microscopy was performed in the cryo-TEM facility at IU.

**REFERENCES**


