analytical chemistry

pubs.acs.org/ac Article

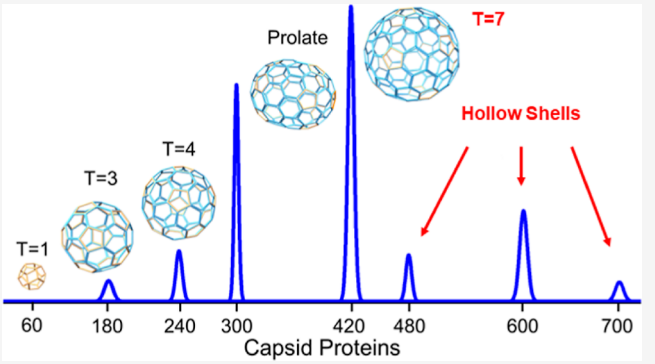
Charge Detection Mass Spectrometry Reveals Favored Structures in the Assembly of Virus-Like Particles: Polymorphism in Norovirus GI.1

Lohra M. Miller, Benjamin E. Draper, Joseph C.-Y. Wang, and Martin F. Jarrold*

Cite This: *Anal. Chem.* 2024, 96, 13150−13157 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The main capsid protein (CP) of norovirus, the leading cause of gastroenteritis, is expected to self-assemble into virus-like particles with the same structure as the wild-type virus, a capsid with 180 CPs in a $T = 3$ icosahedron. Using charge detection mass spectrometry (CD-MS), we find that the norovirus GI.1 variant is structurally promiscuous, forming a wide variety of well-defined structures, some that are icosahedral capsids and others that are not. The structures that are present evolve with time and vary with solution conditions. The presence of icosahedral $T = 3$ and $T = 4$ capsids (240 CPs) under some conditions was confirmed by cryo-electron microscopy (cryo-EM). The cryo-EM studies also confirmed the presence of an unexpected prolate geometry based on an elongated $T = 4$ capsid with 300 CPs. In addition, CD-MS measurements indicate the presence of well-defined peaks with masses corresponding to 420, 480, 600, and 700 CPs. The peak corresponding to 420 CPs is probably due to an icosahedral $T = 7$ capsid, but this could not be confirmed by cryo-EM. It is possible that the $T = 7$ particles are too fragile to survive vitrification. There are no mass peaks associated with the $T = 9$ and $T = 12$ icosahedra with 540 and 720 CPs. The larger structures with 480, 600, and 700 CPs are not icosahedral; however, their measured charges suggest that they are hollow shells. The use of CD-MS to monitor virus-like particles assembly may have important applications in vaccine development and quality control.



INTRODUCTION

Norovirus is a nonenveloped virus that causes acute gastroenteritis. It is the leading cause of foodborne illness in the United States, and responsible for the loss of over $60 billion annually from healthcare expenses and lost productivity worldwide.[1,2] The virus has significant genetic diversity with 10 genogroups and 49 genotypes identified.[1] Genogroups I, II, IV, VIII and IX (GI, GII, GIV, GIV, and GIX) infect humans, though most infections involve GI and GII.[3] There is no vaccine, and due to its genetic diversity and the propensity of norovirus to quickly evolve, a more traditional, cost-effective approach is preferred to the next-generation vaccines used to combat COVID-19.[4]

Virus-like particles (VLPs) are seen as a compromise between protein subunit vaccines and inactivated and live-attenuated vaccines, both of which have drawbacks. Protein subunit vaccines are difficult to deliver and generally require high doses. For inactivated and live attenuated vaccines, incomplete inactivation or attenuated virus reversion can diminish effectiveness.[5] VLP vaccines present the antigen but lack the genomic material required for infection. They rely on the native assembly mechanism to generate biomimetic structures. Viral antigens are usually surface structures and proper assembly of the VLP to mimic the native structure is thought to be important for both immune response and stability.

Many viral capsids have structures that follow the Caspar and Klug (CK) model where repeating subunits (pentamers and hexamers) assemble to create 20 equilateral triangles that are the faces of an icosahedron. The capsid architecture revolves around pentameric vertices with varying number of repeating subunits added as hexamers. The triangulation number identifies allowed icosahedral structures ($T = 1, 3, 4, 7, 9, 12,\cdots$), where $60T$ is the number of capsid proteins (CPs) arranged in 12 pentamers and $10(T-1)$ hexamers.[6] The CK model has been expanded to include prolate extensions of icosahedral symmetry to account for the structures of bacteriophages.[7−11]

Received: April 11, 2024
Revised: July 9, 2024
Accepted: July 15, 2024
Published: July 29, 2024

https://doi.org/10.1021/acs.analchem.4c01913
Anal. Chem. 2024, 96, 13150−13157

For wild-type norovirus, the main capsid protein, VP1, assembles into $T = 3$ icosahedral capsids with 180 CPs. GI.1 and GII.4 are the main genotypes being explored as candidate vaccines. The prototypical GI.1 Norwalk was identified in 1968 and GII.4 is currently one of the dominant genotypes causing outbreaks worldwide. For GI.1 VLPs, $T = 3$ is the only structure identified. In addition to the $T = 3$, $T = 1$ and $T = 4$ structures have been detected for GII.4 VLPs assembled in vitro and a $T = 1$ structure has been found for GI.1 VLPs assembled from truncated VP1.[12−16]

The ability to quickly verify the VLP structure is important for vaccine development, as an incorrect structure may impact the immunological response. A variety of techniques have been used to perform a structural analysis of norovirus VLPs, including electron microscopy, circular dichroism, and gas phase electric mobility measurements.[15−18] Transmission electron microscopy can give insight into the size of the particles while cryoelectron microscopy (cryo-EM) can provide more detailed structural information. However, low-abundance lower-symmetry species can be missed when doing a 3D reconstruction for nonicosahedral species, and more labile structures may not survive vitrification. An orthogonal technique that can quickly identify all species found in a VLP sample would help to advance vaccine development.

The mass distribution could be used to characterize VLP samples and determine what species are present. Heterogeneous and high mass VLP samples are beyond the range of conventional MS. However, they can be analyzed by charge detection mass spectrometry (CD-MS), an emerging approach where the mass of each ion is determined from simultaneous measurement of its mass to charge (m/z) ratio and charge.[19−23] The high resolution and high sensitivity of CD-MS measurements makes it possible to identify and quantify preferred assembly products as well as assembly intermediates. CD-MS is faster and more accurate than methods based on light scattering like mass photometry.[24]

The mechanism by which virus capsids assemble to give perfect icosahedral symmetry remains poorly understood.[25,26] For some capsids, assembly is hierarchical. For example, the pseudo $T = 7$ polyomavirus capsid assembles from pentamers of the capsid protein.[27] For small single-stranded RNA viruses the capsids assemble around their viral RNA which appears to guide assembly using electrostatics[28,29] and packaging signals.[30,31] However, larger double-stranded DNA or RNA viruses invariably require scaffolding proteins or an inner core to template assembly.[32−39] For example, the native structure of bacteriophage P22 is $T = 7$, but in the absence of scaffolding proteins smaller particles, consistent with a $T = 4$ diameter, form in addition to the $T = 7$.[28−30] Similarly, the herpes simplex virus structural coat proteins are organized into a $T = 16$ structure but without scaffolding proteins, a $T = 7$ assembles.[40] Because of the presence of scaffolding proteins in the assembly of larger icosahedral viruses and results from coarse-grained simulations,[35] the prevalent view has been that scaffolding proteins are necessary for the growth of large T number capsids. The experimental findings presented here indicate that there might be exceptions to this conjecture.

EXPERIMENTAL METHODS

CD-MS Measurements. CD-MS is a single particle technique where the m/z and charge are determined simultaneously for each ion. The measurements reported here were made on our updated prototype CD-MS instrument.[41−43] The ions were generated by nanoelectrospray using a Triversa NanoMate (Advion) and enter the instrument through a capillary. The resulting supersonic jet is disrupted in a FUNPET interface designed to thermalize and desolvate the ions. They then pass through two stages of differential pumping that house in turn a hexapole and a segmented quadrupole. Ions that exit the quadrupole are focused into a dual hemispherical deflection energy analyzer which transmits ion kinetic energies centered on the nominal ion energy of 100 eV/charge. Ions that are within the band-pass are focused into an electrostatic linear ion trap (ELIT) based on the cone trap design.[44] Trapped ions oscillate back and forth through a detection cylinder located between the two end-caps that make up the ELIT. A charge sensitive amplifier with a cryogenically cooled input FET[45] is connected to the detection cylinder and measures the induced charge from the oscillating ion. The resulting time domain signal is amplified, digitized, and analyzed by fast Fourier transforms (FFTs) using a series of short time windowed FFTs.[39] Trapping events where ions are not trapped for the full period are discarded. The m/z ratio is obtained from the fundamental frequency and the charge is determined from the magnitudes of the fundamental and second harmonic.[46] The uncertainty in the charge is ∼1 e for a trapping time of 100 ms. Usually around 5000 individual ion measurements are performed for each data set.

Norovirus samples were purchased from The Native Antigen Company. The samples were shipped on dry ice, thawed and aliquoted on arrival, and then stored at −80 °C. Immediately before the CD-MS measurements, samples were thawed, and buffer changed into 200 mM pH 7.0 ammonium acetate solution (ThermoFisher, AM9070G and 10977015) using Micro Bio-Spin $P\overline{6}$ (Bio-Rad, 7326225) size exclusion columns. During the time course studies the samples were stored in their shipping buffer (pH 7.4 phosphate buffered saline) at 4 °C. The samples were electrosprayed at a protein concentration of approximately 0.5 mg/mL for the CD-MS measurements. All mass histograms have a bin size of 59 kDa and masses were determined using Gaussian fits. All measurements were performed at least twice.

Cryo-EM Measurements. For cryo-EM 3-μL of sample solution was placed on a glow-discharged Quantifoil R2/2 holey carbon film-coated copper grid. Excess sample was removed, and the grid was plunged into liquid ethane cooled by liquid nitrogen to embed the VLPs in a thin layer of vitrified ice for EM analysis. The freezing procedure was performed on a Thermo Fisher Scientific Vitrobot Mark IV system. The frozen hydrated grid was then transferred to a 200 kV Thermo Scientific Talos Arctica cryo-EM equipped with Falcon III direct electron detector camera for data collection.

Multiple low-dose movie frames were collected at a pixel size of 0.7 Å with a total dose of 30 $e^-/Å^2$. Motion correction, defocus estimation, reference-free 2D classification, 3D refinement, and 3D reconstruction were done in Relion v3.0.[47] After discarding classes with blurred density, three major classes of particles, namely $T = 3$, $T = 4$, and elongated particle classes, were observed in the 2D classification. To proceed with 3D image reconstruction, particles within $T = 3$ and $T = 4$ classes were selected and processed separately with a scaled pixel size of 2.8 Å. The initial starting model was built *de novo* in Relion. A total of 4156 particles were used to generate a final 3D model of $T = 3$ VLP to 9.7 Å and a total of 1245 particles were used to generate a final model of $T = 4$ VLP to 13.2 Å. The 3-D reconstructions were visualized using UCSF ChimeraX.[48]

https://doi.org/10.1021/acs.analchem.4c01913
Anal. Chem. 2024, 96, 13150−13157

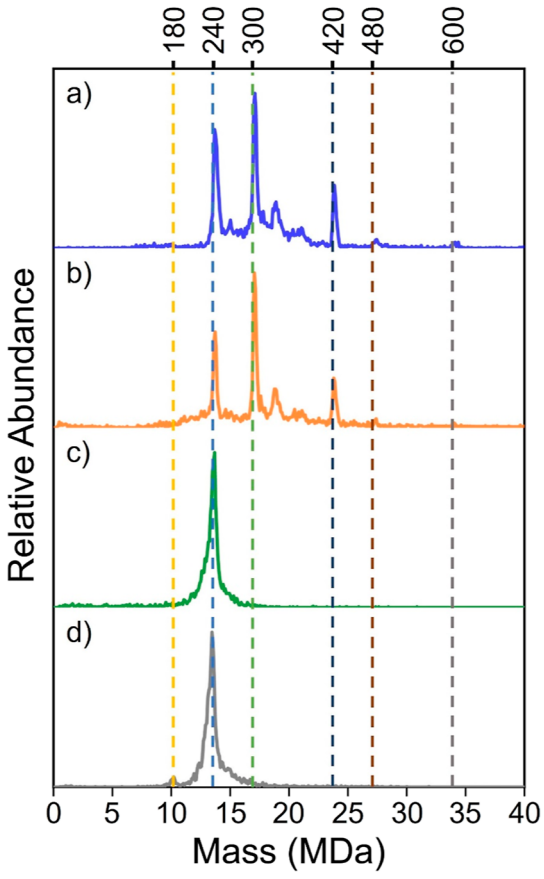


Figure 1. CD-MS mass distributions measured for GI.1 samples. (a) Lot #1 order #1. (b) Lot #1 order #2. (c) Lot #2. (d) Lot #3. Each sample was electrosprayed immediately after buffer exchange into 200 mM ammonium acetate. Dashed vertical lines show the expected masses for particles with 180 CPs (T = 3), 240 CPs (T = 4), 300 CPs, 420 CPs, 480 CPs, and 600 CPs.

RESULTS

Figure 1 shows the CD-MS mass histogram of several norovirus GI.1 VP1 only VLP samples measured immediately after thawing and sample preparation. The dashed vertical lines show the expected masses for 180 CPs (T = 3), 240 CPs (T = 4), 300 CPs, 420 CPs, 480 CPs, and 600 CPs. The expected molecular weights of the GI.1 assemblies were determined from a weighted average of the VP1 mass spectrum measured using a Waters Synapt G2S (see Figure S1 in Supporting Information). The average mass obtained in this way (56.573 kDa) is virtually identical to the sequence mass (56.58 kDa). The CD-MS mass distributions shown in Figure 1 are for three different lots from the same manufacturer. Two orders were obtained from Lot #1 and their CD-MS spectra are shown in Figure 1a,b. They are very similar, and quite different from the CD-MS spectra measured for Lot #2 and Lot #3 (shown in Figure 1c,d, respectively). For Lot #1 order #1, a CD-MS spectrum measured before aliquoting was virtually identical to the spectrum measured after a freeze−thaw cycle.

The only notable population in Figure 1 near the mass of the expected T = 3 particles with 180 CPs (yellow dashed line at 10.18 MDa) is the small peak in Figure 1d. Instead, there are prominent peaks in Figure 1a,b at 13.71 MDa which is 1% larger than the expected mass for the T = 4 particle with 240 copies of VP1 (13.58 MDa). The samples were electrosprayed from a volatile salt solution (native MS). The salt (ammonium acetate in this case) provides the ionic strength needed to retain the integrity of higher-order structures in solution, and because it is volatile, it evaporates during the electrospray process so that salt adducts are minimized and the measured masses are close to the expected values. Typically, the measured masses are slightly (<1%) larger than the expected values.[49,50] A 1% excess mass is within the range expected for electrospray from volatile salt solutions, albeit at the high end. Thus, the peaks at 13.71 MDa in Figure 1a,b are assigned to T = 4 particles.

Beyond the T = 4 capsid, the next largest icosahedron that obeys the Caspar−Klug rules is the T = 7.[5] With 420 CPs, the T = 7 VLP has an expected mass of 23.76 MDa (brown dashed line in Figure 1). The peak at 23.86 MDa in Figure 1a is around 0.5% higher than the expected mass, leading to the assignment of this peak to the T = 7 VLP. The peak at 17.10 MDa lies between the two Caspar−Klug icosahedra, indicating that this peak must belong to a structure that is not an icosahedron. Based on its mass, it contains 300 CPs (see dashed green line in Figure 1). A likely structure, is obtained by elongating the T = 4 capsid along its 5-fold axis by separating the T = 4 end-caps at the midpoint, inserting a ring of 10 hexamers, rotating one of the end-caps by 18° and then recombining the components.[6,7,51] Such a geometry contains 300 proteins and has an expected mass of 16.97 MDa. The peak at 17.10 MDa is 0.8% above the expected mass for this structure. Thus, we assign the peak at 17.10 MDa to a prolate T = 4 capsid with 300 CPs. In addition to the peaks due to T = 4, elongated T = 4, and T = 7, there is a small peak at around 18.97 MDa. There are also small peaks that are close to the lines indicating 480 CPs and 600 CPs that we note here because they become more prominent under different conditions.

Lot #2 and Lot #3, Figure 1c and d respectively, show single prominent peaks near the 240 CPs line. The peaks are relatively broad with both high and low mass tails, indicating heterogeneous distributions containing both incomplete and overgrown T = 4 capsids.

Figure 2 shows the evolution of GI.1 lot #2 over the span of a week. Initially (Figure 2a), the spectrum is similar to Figure 1c with a peak at 13.60 MDa (close to the expected mass for 240 CPs) with low and high mass tails. After storage at 4 °C for 8 h, another peak has appeared close to the mass expected for 180 CPs (T = 3) (Figure 2b). After 1 day (Figure 2c), the peak near 180 CPs is the largest peak in the spectrum and a third peak has appeared near to the expected mass for 300 CPs. There is also a small peak near the expected mass of 420 CPs. The spectrum after 1 day also has a significant heterogeneous population at masses below 180 CPs. After 1 week (Figure 2d) the spectrum shows less heterogeneity, the intensity at low masses is much smaller and the peaks at higher masses are sharper. The mass distribution is similar to that found for Lot #1 (Figure 1a), except for the addition of the peak at 10.16 MDa due to the expected T = 3 structure. Thus, far, the GI.1 VP1 protein has assembled into the following structures: T = 3 (180 CPs), T = 4 (240 CPs), elongated T = 4 (300 CPs), and T = 7 (420 CPs), and there is evidence of even larger structures with small peaks in Figure 2d close to 480 CPs and 600 CPs.

In previous studies of GI.1 VP1 VLPs the T = 3 capsid was the dominant species found.[13,16] T = 1 particles have been observed to assemble for N-terminal truncated VP1 of the GI.1 genotype.[17] However, we did not observe any T = 1 particles,

https://doi.org/10.1021/acs.analchem.4c01913
Anal. Chem. 2024, 96, 13150−13157

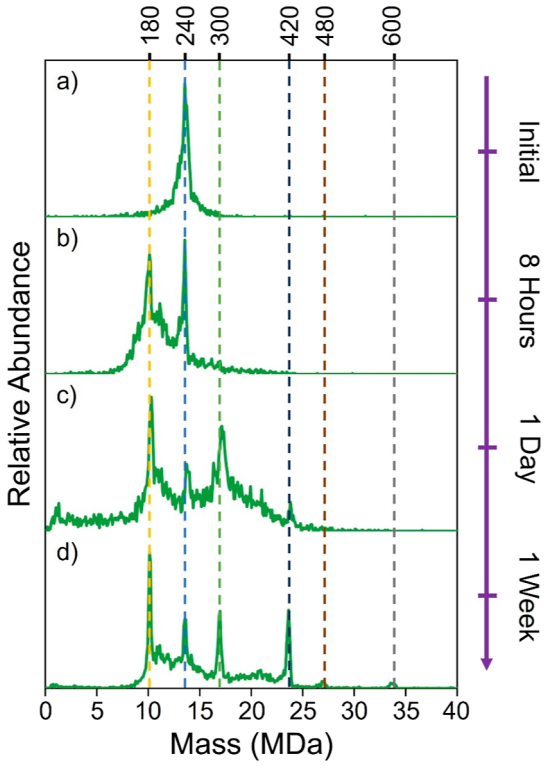


Figure 2. CD-MS mass spectra of GI.1 Lot #2 measured as a function of time stored in its shipping buffer (pH 7.4 phosphate buffered saline) at 4 °C. Dashed lines show expected masses for particles with 180, 240, 300, 420, 480, and 600 CPs. (a) time zero, (b) after 8 h, (c) after 1 day, and (d) after 1 week. Each sample was electrosprayed immediately after buffer exchange into 200 mM ammonium acetate.

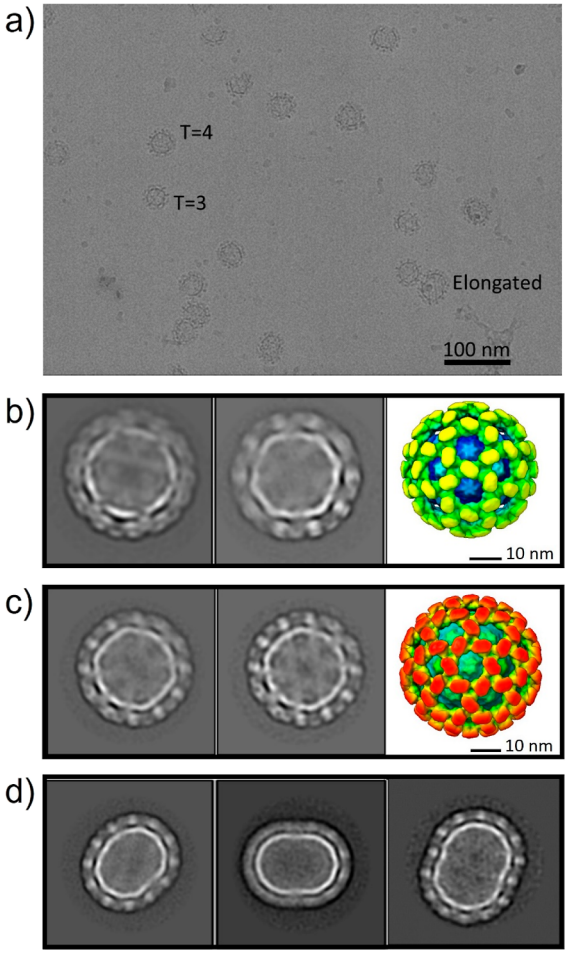


Figure 3. Results of Cryo-EM studies of time-evolved norovirus GI.1 VLPs. (a) Cryo-EM micrograph showing $T = 3$, $T = 4$, and elongated structures. (b) Representative 2D classification of $T = 3$ particles and a 3-D reconstruction at 9.7 Å. (c) Representative 2D classification of $T = 4$ and 3D reconstruction at 13.2 Å and (d) 2-D classification results for the elongated $T = 4$ particles. See text for more information.

and conventional MS measurements for the intact GI.1 VP1 revealed no truncations (Figure S1). Initially, all GI.1 lots lacked a significant abundance of the expected $T = 3$ structure, instead favoring $T = 4$. However, the $T = 3$ structure emerged over time when the sample was stored at 4 °C, along with higher mass structures that have not been observed before for the GI.1 genogroup.

To further investigate the structures of the GI.1 VP1 VLPs, the time-evolved sample from Lot #2 (Figure 2d) was analyzed with Cryo-EM. A representative portion of the Cryo-EM micrograph in Figure 3a shows elongated particles and large and small spherical particles attributed to $T = 3$ and $T = 4$ icosahedra. Upon 2D classification, 4156 small spherical particles led to a $T = 3$ particle that was 3D reconstructed at 9.7 Å resolution. The resulting structure is shown in Figure 3b. 1245 larger spherical particles were 3D reconstructed to give $T = 4$ icosahedral structures at 13.2 Å resolution. The resulting structure is shown in Figure 3c. For the elongated structures, 2D classification led to 2262 elongated particles, representative examples are shown in Figure 3d. They have the size and aspect ratio expected for the extended $T = 4$ capsid with 300 CPs. However, the small number of nonicosahedral particles and their lack of uniformity precluded a 3D reconstruction. Nonicosahedral structures have not previously been reported for norovirus VLP particles of any genotype. The larger $T = 7$ icosahedral structure was not observed in the cryo-EM micrographs, despite the fact that the $T = 7$ particle is as abundant in the CD-MS mass distributions as the $T = 4$ and elongated $T = 4$ (see Figure 2d).

Note that the number of particles classified and used to generate the 3D reconstructions is modest. We did not concentrate the sample because we wanted to image at the same concentration used for the CD-MS measurements. The resolutions achieved (9.7 Å for the $T = 3$ particle and 13.2 Å for the $T = 4$) are sufficient to distinguish the different spatial organizations of the norovirus VLPs and confirm the presence of the $T = 3$ and $T = 4$ particles. In view of the absence of larger particles in the cryo-EM images we spotted some grids and recorded negative stain TEM images. Representative results are shown in Figure S2. While it was possible to identify some of the particles as $T = 3$ and $T = 4$, many of the particles, particularly the larger ones, appeared to have been flattened.

To further explore the dynamic nature of the structures found for the GI.1 VP1 VLPs we incubated samples of Lot #1. Figure 4 shows representative results. Figure 4a is a control, measured before incubation. The peak in Figure 4a due to the $T = 4$ icosahedron (240 CPs) is reduced compared to Figure 1a,b. This change just reflects the time evolution of this sample. Incubation at 37 °C for 30 min led to the spectrum shown in Figure 4b. The distribution of structures found after incubation at 37 °C is shifted to higher mass. The largest peak

https://doi.org/10.1021/acs.analchem.4c01913
Anal. Chem. 2024, 96, 13150−13157

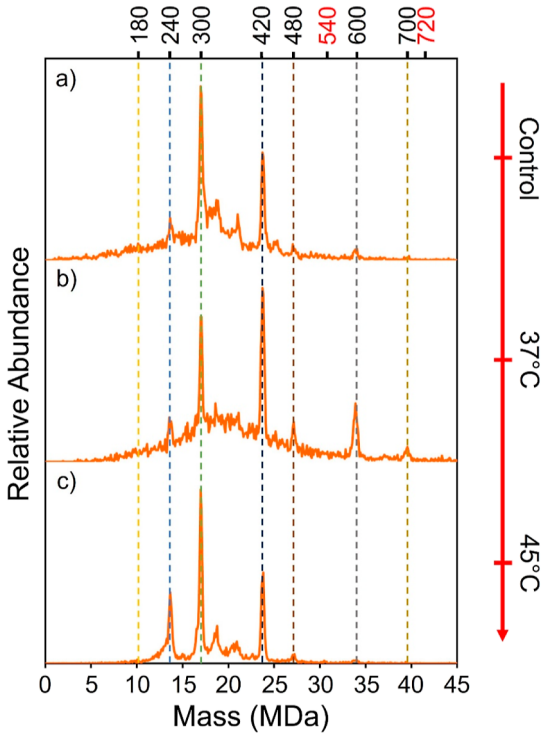


Figure 4. Incubation of GI.1 VP1 VLPs. (a) Control, (b) after incubation at 37 °C, and (c) after incubation at 45 °C. After incubation, the samples were buffer exchanged into 200 mM ammonium acetate, and then electrosprayed. Dashed vertical lines show the number of CPs associated with expected structures (see text).

in the spectrum at 23.74 MDa is now due to the T = 7 icosahedron (420 CPs). There are significant peaks at 27.18 MDa near 480 CPs and at 33.94 MDa near 600 CPs. There is even a small peak at 39.56 MDa near 700 CPs. Incubation at a slightly higher temperature, 45 °C, does not lead to larger structures. Instead, the structures found after incubation at 45 °C are similar to those found for the control, except for the larger peak near 240 CPs (T = 4).

It is noteworthy that the peaks near to 480 CPs, 600 CPs, and 700 CPs after incubation at 37 °C do not correspond to Caspar−Klug icosahedra. Beyond T = 7, the next Caspar−Klug icosahedra are T = 9 and T = 12 with 540 CPs and 720 CPs, respectively. These values are indicated by the red numbers at the top of Figure 4. Since the higher mass peaks corresponding to 480, 600, and 700 CPs do not correspond to Caspar−Klug icosahedra, there is a question about their structures. For example, it is possible that the higher mass structures are not capsids (i.e., empty on the inside), but more compact structures? We used the charge to explore this possibility.

Large electrosprayed ions are expected to be generated by the charge residue mechanism[52,53] where the analyte remains in an electrospray droplet as the droplet evaporates, leaving the ion with the charge on the droplet just before it disappears. The charge on the droplet is, in turn, expected to follow the Rayleigh charge limit for a water droplet which is determined from a balance between electrostatic repulsion and surface tension

$$q_R = 8\pi(\varepsilon_0 \gamma r^3)^{1/2} \quad (1)$$

where q_R is the Rayleigh limit charge for a droplet of radius r, ε_0 is the permittivity of free space, γ is the surface tension. Thus, larger ions are expected to be more highly charged, and the measured charge can be used to provide crude structural information.[54]

Figure 5b shows a 2D charge versus mass intensity plot for GI.1 VP1 VLPs after incubation at 37 °C. In this plot, higher

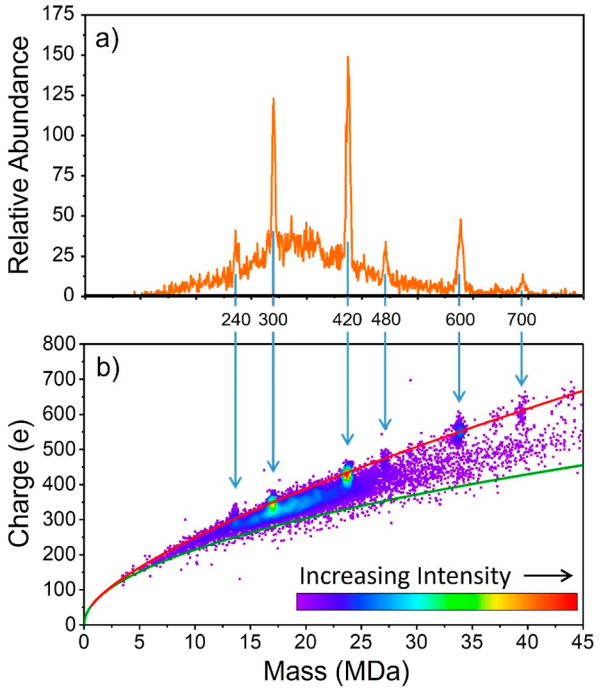


Figure 5. Correlation between charge and mass for GI.1 VP1 VLPs after incubation at 37 °C. (a) shows the mass histogram, and (b) shows the 2D charge vs mass intensity plot where warmer colors indicate higher intensity as shown by the scale in the inset. The blue arrows show the connection between the peaks in the mass distribution and the features in the charge versus mass intensity plot. The green line shows the charge expected for a solid spherical structure and the red line shows the charge expected for a spherical shell (see text for details).

intensities are indicated by warmer colors (as indicated in the scale in the inset). Such a plot can reveal correlations between charge and mass. Figure 5a shows the mass distribution and blue arrows show the correlations between the features in the mass distribution and the high intensity regions in the charge versus mass intensity plot. The green line in Figure 5 was calculated assuming a solid spherical structure with a typical globular protein density of 1.3 g/cm^3. Most of the intensity in Figure 5b lies above the green line. The intensity below the line probably results because not all ions are charged at the Rayleigh limit. The red line shows the charge calculated using eq 1 assuming that the ions are spherical shells with a wall thickness of 6 nm. The ions in Figure 5b associated with the prominent mass peaks in Figure 5a form clusters that have higher charges than ions between the peaks. These more highly charged clusters of ions fall on the red line suggesting that the larger structures with 480 CPs, 600 CPs, and 700 CPs are hollow shells.

https://doi.org/10.1021/acs.analchem.4c01913
Anal. Chem. 2024, 96, 13150−13157

DISCUSSION

For the Norovirus GI.1 genotype, only the $T = 3$ icosahedron has been observed in previous studies, though $T = 1$ particles have been found for truncated VP1. In contrast, CD-MS measurements show prominent peaks corresponding to 180 CPs, 240 CPs, 300 CPs, 420 CPs, 480 CPs, 600 CPs, and 700 CPs. The presence of $T = 3$ and $T = 4$ icosahedra with 180 CPs and 240 CPs was confirmed by cryo-EM studies. As noted above, 300 CPs does not correspond to a Caspar−Klug icosahedron. It does, however, correspond to a $T = 4$ capsid elongated along its 5-fold axis by the addition of a ring of 10 hexagons and there is precedent for such a structure.[52−54] Cryo-EM measurements support this assignment. The next largest peak corresponds to 420 CPs. We assigned this feature to a $T = 7$ capsid that was not observed in the cryo-EM studies. A possible alternative explanation for the peak at 420 CPs is that it results from the addition of two more rings of 10 hexamers to the elongated $T = 4$ structure with 300 CPs. The argument against this assignment is that there is no peak at 360 CPs that would result from the addition of one more ring of 10 hexamers to the elongated $T = 4$. Thus, the $T = 7$ assignment remains the most likely explanation.

The next highest mass peak is at 480 CPs, 60 CPs above the peak at 420 CPs. This raises the question as to whether the 480 peak could be due to an elongated $T = 7$ structure, with a band of 10 hexagons. However, the $T = 7$ structure is expected to elongate through the addition of units of 5 hexamers.[8] Thus, the peak at 480 CPs would correspond to an elongated $T = 7$ particle with the addition of two groups of 5 hexamers. There is no peak at 450 CPs corresponding to the addition of one group of 5 hexamers. So, the explanation that the peak at 480 CPs is due to the addition of two groups of five seems unlikely.

For the 600 CPs peak, one explanation that comes to mind is a $T = 3$ particle nested inside a $T = 7$. Such a structure would have 180 CPs + 420 Cps = 600 CPs. However, the resulting nested capsid would have an outside diameter similar to the $T = 7$ capsid without an internalized $T = 3$ particle. So, the nested capsid would be expected to have a charge similar to the $T = 7$ particle with 420 CPs. The results in Figure 5b show that the charge on the 600 CP particle is significantly larger than on the 420 CP particle, and consistent with a hollow structure with a single shell.

The failure to detect particles larger than the elongated $T = 4$ (with 300 CPs) in the cryo-EM studies is puzzling. The peak in Figure 2d assigned to the $T = 7$ (420 CPs) is the same size as the peaks due to the $T = 4$ (240 CPs) and the elongated $T = 4$ (300 CPs). Furthermore, the number of particles found in the 2D classification of the cryo-EM images roughly correlates with the relative intensities of the $T = 3$, $T = 4$, and elongated $T = 4$ in the CD-MS mass distributions (Figure 2d). Particles were manually picked from the cryo-EM micrographs to eliminate the possibility that larger particles were missed by the computer program. Since all the particles are thought to be hollow single-wall shells, it is likely that the larger particles are more fragile and more easily deformed.

We hypothesize that the larger particles were destroyed during the sample preparation step. In cryo-EM specimen preparation, the plunge freezing step can result in the loss of up to 90% of the material to the filter papers. In addition, fragile particles can be deformed at the air−water interface or disassembled due to local concentration changes.[55] If the surface properties of the particles change, they may be more easily picked up by the filter paper. In our cryo-EM micrographs, we observed wedges of particles (Figure 3a, top), indicating potential breakdown of fragile particles. Moreover, in the negative stain TEM micrographs (Figure S2), we observed many particles larger than $T = 4$, but they appeared "flattened" on the grids. Taken together, these analyses indicate that the larger particles are fragile species that may not survive plunge freezing but were detected by CD-MS.

In light of the preceding discussion, we cannot suggest plausible structures for particles with 480, 600, and 700 CPs, except to say that they all appear to be hollow shells. The fact that they are prominent features in the CD-MS mass distributions indicates that there must be significant free energy minima associated with their structures. And the fact that they are formed indicates that there is a low energy assembly pathway. The assembly of large virus capsids is usually templated or requires the assistance of scaffolding proteins.[28−35] In the absence of the scaffolding protein assembly is usually much slower and more error prone. Even a virus as small as adeno-associated virus which has a $T = 1$ capsid with 60 CPs employs an assembly activating protein to aid assembly in vivo.[56] The building block for norovirus capsid is the dimer,[11,57] thus assembly of the larger particles requires hundreds of CP dimers to come together at the correct location and in the correct orientation. This challenge is reminiscent of protein folding and Levinthal's paradox that a random search of all possible configurations would take much too long.[58] The corollary is that there must be well-defined assembly pathways that can find the low energy geometries without encountering traps. This may explain the absence of $T = 9$ and $T = 12$ icosahedra, and the preference for large nonicosahedral structures.

Polymorphism is not uncommon for virus-like particles.[59,60] Removing the scaffolding protein or template that aids the assembly of the wild-type virus opens up assembly into different morphologies. For example, for Qβ, a single stranded RNA bacteriophage, assembly of the CPs around the genomic RNA is mediated by packaging signals to form a pseudo icosahedral $T = 3$ capsid.[61,62] For Qβ VLPs, in addition to the icosahedral $T = 3$ capsid with 180 CPs, prolate and oblate geometries have been observed in cryo-EM studies.[63] The prolate form (with 210 CPs) is made by adding 5 hexamers to the $T = 3$ capsid and the oblate form (with 150 CPs) by removing 5 hexamers. A small prolate form (132 CPs) was also found, and a $T = 1$ particle (60 CPs) has been detected.[63] What is remarkable about the results presented here for the norovirus GI.1 VLPs is the large number of structures and their broad size range.

Another remarkable feature of the results presented here is the dynamic nature of the structures that are present. A small change in conditions can lead to a substantial change in the nature of the structures present. This dynamic behavior indicates that the structures have similar free energies, but there must also be low energy assembly and disassembly pathways so that the structures can be reached. There may be other structures that are deep potential wells, that are not accessible. An important question is whether or not one structure must almost completely disassemble to convert into another. Or is it possible to convert, for example, from a $T = 4$ capsid to an extended $T = 4$ or even a $T = 7$ by adding additional subunits and then undergoing a structural rearrangement. The extent of the rearrangement that is required makes this pathway seem less likely. However, complete disassembly

https://doi.org/10.1021/acs.analchem.4c01913
Anal. Chem. 2024, 96, 13150−13157

followed by reassembly along a different pathway is also long and torturous. Subunit addition and rearrangement provides a way of reaching large structures without an assembly chaperone.

CONCLUSIONS

Prior to this study, no structures larger than the expected $T = 3$ icosahedron had been reported for the Norovirus GI.1 Norwalk genotype. We observed CD-MS mass peaks corresponding to 180, 240, 300, 420, 480, 600, and 700 CPs. For particles with 180, 240, and 300 CPs, assignments to $T = 3$, $T = 4$, and elongated $T = 4$ structures were confirmed by cryo-EM measurements. However, larger structures were not observed in the cryo-EM micrographs, possibly because they were destroyed during sample preparation. It is likely that the peak due to 420 CPs is due to a $T = 7$ icosahedron. The larger species with 480, 600, and 700 CPs are hollow single-wall shells, but they are not icosahedra.

The results presented here show that CD-MS can be used to quickly elucidate the VLP structures present in solution. Accurate mass measurements can also reveal whether the particles are missing a subunit, information that is not available from cryo-EM measurements due to averaging. On the other hand, combining information from CD-MS and cryo-EM leads to a more comprehensive picture of the structures that are present. From a practical perspective, the ability to quickly identify what VLP structures are present makes CD-MS a useful tool for vaccine development and quality control during production. Information about the structures is important for VLP vaccines because the structure may determine how the antigen presents.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.4c01913.

m/z spectrum for the GI.1 VP1 protein, negative stain TEM images of time-evolved norovirus GI.1 VLPs (PDF)

AUTHOR INFORMATION

Corresponding Author

Martin F. Jarrold − *Chemistry Department, Indiana University, Bloomington, Indiana 47405, United States;* orcid.org/0000-0001-7084-176X; Email: mfj@iu.edu

Authors

Lohra M. Miller − *Chemistry Department, Indiana University, Bloomington, Indiana 47405, United States*

Benjamin E. Draper − *Megadalton Solutions Inc, Bloomington, Indiana 47401, United States;* orcid.org/0000-0002-9727-2509

Joseph C.-Y. Wang − *Department of Microbiology and Immunology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, United States;* orcid.org/0000-0001-6580-3531

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.4c01913

Notes

The authors declare the following competing financial interest(s): Two of the authors (BED and MFJ) are shareholders in Megadalton Solutions, a company that is engaged in commercializing CDMS. MFJ is a consultant for Waters.

ACKNOWLEDGMENTS

We are grateful to the NSF IUCRC Center for Bioanalytic Metrology (CBM) for financial support provided under Grant Number IIP-1916645, and for valuable discussions with CBM industry partners and staff. The intact mass measurement for the GI.1 VP1 protein was performed using resources of the Biological Mass Spectrometry Facility at Indiana University. We thank Jonathan V Trinidad for assistance.

REFERENCES

(1) Ludwig-Begall, L. F.; Mauroy, A.; Thiry, E. *Viruses* **2021**, *13*, 1541.

(2) Center for Disease Control and Prevention. Burden of Norovirus Illness in the US 2021, https://www.cdc.gov/norovirus/trends-outbreaks/burden-US.html (accessed Feb 02, 2022).

(3) Chan, M. C. W.; Kwan, H. S.; Chan, P. K. S. Chapter 4 - Structure and Genotypes of Noroviruses. *The Norovirus*; Chan, P. K. S., Ed.; Academic Press, 2017; pp 51−63.

(4) Graham, B. S.; Gilman, M. S. A.; McLellan, J. S. *Annu. Rev. Med.* **2019**, *70*, 91−104.

(5) Bull, J. J. *Virus Evol.* **2015**, *1*, vev005.

(6) Caspar, D. L.; Klug, A. *Cold Spring Harbor Symp. Quant. Biol.* **1962**, *27*, 1−24.

(7) Moody, M. F. *Virology* **1965**, *26*, 567−576.

(8) Moody, M. F. *J. Mol. Biol.* **1999**, *293*, 401−433.

(9) Luque, A.; Zandi, R.; Reguera, D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5323−5328.

(10) Luque, A.; Reguera, D. *Biophys. J.* **2010**, *98*, 2993−3003.

(11) Llorente, J. M. G.; Hernández-Rojas, J.; Bretón, J. *Soft Matter* **2014**, *10*, 3560−3569.

(12) Pogan, R.; Dülfer, J.; Uetrecht, C. *Curr. Opin. Virol.* **2018**, *31*, 59−65.

(13) Devant, J. M.; Hofhaus, G.; Bhella, D.; Hansman, G. S. *Antiviral Res.* **2019**, *168*, 175−182.

(14) Jung, J.; Grant, T.; Thomas, D. R.; Diehnelt, C. W.; Grigorieff, N.; Joshua-Tor, L. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 12828−12832.

(15) Devant, J. M.; Hansman, G. S. *Virology* **2021**, *553*, 23−34.

(16) Weiss, V. U.; Pogan, R.; Zoratto, S.; Bond, K. M.; Boulanger, P.; Jarrold, M. F.; Lyktey, N.; Pahl, D.; Puffler, N.; Schelhaas, M.; Selivanovitch, E.; Uetrecht, C.; Allmaier, G. *Anal. Bioanal. Chem.* **2019**, *411*, 5951−5962.

(17) Shoemaker, G. K.; van Duijn, E.; Crawford, S. E.; Uetrecht, C.; Baclayon, M.; Roos, W. H.; Wuite, G. J.; Estes, M. K.; Prasad, B. V. V.; Heck, A. J. *Mol. Cell. Proteomics* **2010**, *9*, 1742−1751.

(18) Pogan, R.; Weiss, V. U.; Bond, K.; Dülfer, J.; Krisp, C.; Lyktey, N.; Müller-Guhl, J.; Zoratto, S.; Allmaier, G.; Jarrold, M. F.; Muñoz-Fontela, C.; Schlüter, H.; Uetrecht, C. *Vaccines* **2020**, *9*, 8.

(19) Keifer, D. Z.; Pierson, E. E.; Jarrold, M. F. *Analyst* **2017**, *142*, 1654−1671.

(20) Jarrold, M. F. *Chem. Rev.* **2022**, *122*, 7415−7441.

(21) Hanozin, E.; Harper, C. C.; McPartlan, M. S.; Williams, E. R. *ACS Cent. Sci.* **2023**, *9*, 1611−1622.

(22) Harper, C. C.; Miller, Z. M.; McPartlan, M. S.; Jordan, J. S.; Pedder, R. E.; Williams, E. R. *ACS Nano* **2023**, *17*, 7765−7774.

(23) Jarrold, M. F. *J. Am. Chem. Soc.* **2024**, *146*, 5749−5758.

(24) Parikh, R. A.; Draper, B. E.; Jarrold, M. F. *Anal. Chem.* **2024**, *96*, 3062−3069.

(25) Katen, S.; Zlotnick, A. *Methods Enzymol.* **2009**, *455*, 395−417.

(26) Perlmutter, J. D.; Hagan, M. F. *Annu. Rev. Phys. Chem.* **2015**, *66*, 217−239.

(27) Salunke, D. M.; Caspar, D. L.; Garcea, R. L. *Cell* **1986**, *46*, 895−904.

(28) Belyi, V. A.; Muthukumar, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17174−17178.
(29) Perlmutter, J. D.; Qiao, C.; Hagan, M. F. *Elife* **2013**, *2*, No. e00632.
(30) Basnak, G.; Morton, V. L.; Rolfsson, O.; Stonehouse, N. J.; Ashcroft, A. E.; Stockley, P. G. *J. Mol. Biol.* **2010**, *395*, 924−936.
(31) Perlmutter, J. D.; Hagan, M. F. *J. Mol. Biol.* **2015**, *427*, 2451−2467.
(32) Earnshaw, W.; King, J. *J. Mol. Biol.* **1978**, *126*, 721−747.
(33) Thuman-Commike, P. A.; Greene, B.; Malinski, J. A.; King, J.; Chiu, W. *Biophys. J.* **1998**, *74*, 559−568.
(34) Thuman-Commike, P. A.; Greene, B.; Malinski, J. A.; Burbea, M.; McGough, A.; Chiu, W.; Prevelige, P. E. *Biophys. J.* **1999**, *76*, 3267−3277.
(35) Dokland, T.; Bernal, R. A.; Burch, A.; Pletnev, S.; Fane, B. A.; Rossmann, M. G. *J. Mol. Biol.* **1999**, *288*, 595−608.
(36) Zlotnick, A.; Suhanovsky, M. M.; Teschke, C. M. *Virology* **2012**, *428*, 64−69.
(37) Sikkema, F. D.; Comellas-Aragonès, M.; Fokkink, R. G.; Verduin, B. J. M.; Cornelissen, J. J. L. M.; Nolte, R. J. M. *Org. Biomol. Chem.* **2007**, *5*, 54−57.
(38) Sun, J.; DuFort, C.; Daniel, M.-C.; Murali, A.; Chen, C.; Gopinath, K.; Stein, B.; De, M.; Rotello, V. M.; Holzenburg, A.; Kao, C. C.; Dragnea, B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1354−1359.
(39) Li, S.; Roy, P.; Travesset, A.; Zandi, R. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 10971−10976.
(40) Saad, A.; Zhou, Z. H.; Jakana, J.; Chiu, W.; Rixon, F. J. *J. Virol.* **1999**, *73*, 6821−6830.
(41) Contino, N. C.; Jarrold, M. F. *Int. J. Mass Spectrom.* **2013**, *345*−*347*, 153−159.
(42) Draper, B. E.; Anthony, S. N.; Jarrold, M. F. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 2160−2172.
(43) Draper, B. E.; Jarrold, M. F. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 898−904.
(44) Schmidt, H. T.; Cederquist, H.; Jensen, J.; Fardi, A. *Nucl. Instrum. Methods Phys. Res., Sect. B* **2001**, *173*, 523−527.
(45) Contino, N. C.; Pierson, E. E.; Keifer, D. Z.; Jarrold, M. F. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 101−108.
(46) Keifer, D. Z.; Shinholt, D. L.; Jarrold, M. F. *Anal. Chem.* **2015**, *87*, 10330−10337.
(47) Scheres, S. H. W. *J. Struct. Biol.* **2012**, *180*, 519−530.
(48) Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. *Protein Sci.* **2018**, *27*, 14−25.
(49) Nettleton, E. J.; Sunde, M.; Lai, Z.; Kelly, J. W.; Dobson, C. M.; Robinson, C. V. *J. Mol. Biol.* **1998**, *281*, 553−564.
(50) McKay, A. R.; Ruotolo, B. T.; Ilag, L. L.; Robinson, C. V. *J. Am. Chem. Soc.* **2006**, *128*, 11433−11442.
(51) Pierson, E. E.; Keifer, D. Z.; Kukreja, A. A.; Wang, J. C.-Y.; Zlotnick, A.; Jarrold, M. F. *J. Mol. Biol.* **2016**, *428*, 292−300.
(52) Fernandez de la Mora, J. *Anal. Chim. Acta* **2000**, *406*, 93−104.
(53) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. *Anal. Chem.* **2013**, *85*, 2−9.
(54) Keifer, D. Z.; Motwani, T.; Teschke, C. M.; Jarrold, M. F. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1028−1036.
(55) Glaeser, R. M. *Curr. Opin. Colloid Interface Sci.* **2018**, *34*, 1−8.
(56) Sonntag, F.; Köther, K.; Schmidt, K.; Weghofer, M.; Raupp, C.; Nieto, K.; Kuck, A.; Gerlach, B.; Böttcher, B.; Müller, O. J.; Lux, K.; Hörer, M.; Kleinschmidt, J. A. *J. Virol.* **2011**, *85*, 12686−12697.
(57) Tubiana, T.; Boulard, Y.; Bressanelli, S. *PLoS One* **2017**, *12*, No. e0182056.
(58) Dykeman, E. C.; Stockley, P. G.; Twarock, R. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 5361−5366.
(59) Zeltins, A. *Mol. Biotechnol.* **2013**, *53*, 92−107.
(60) Lie, F.; Szyszka, T. N.; Lau, Y. H. *J. Mater. Chem. B* **2023**, *11*, 6516−6526.
(61) Cui, Z.; Gorzelnik, K. V.; Chang, J. Y.; Langlais, C.; Jakana, J.; Young, R.; Zhang, J. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 11697−11702.
(62) Chang, J. Y.; Gorzelnik, K. V.; Thongchol, J.; Zhang, J. *Viruses* **2022**, *14*, 225.
(63) Shaw, V.; Sungsuwan, S.; McFallBoegeman, H.; Huang, X.; Jin, X. *bioRxiv* **2022**, 01.23.477406.

https://doi.org/10.1021/acs.analchem.4c01913
Anal. Chem. 2024, 96, 13150−13157