THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : CSV- Chimie et Sciences du Vivant
Spécialité : Biologie Structurale et Nanobiologie
Unité de recherche : BGE - Laboratoire Biosciences et bioingénierie pour la Santé

Vers une caractérisation structurale des particules pseudo-virales par spectrométrie de masse à l'aide de nanorésonateurs

Towards a structural characterization of engineered Virus-Like Particles using nanoresonator-based mass spectrometry

Présentée par :

Vaitson ÇUMAKU

Direction de thèse :

Christophe MASSELON Ingénieur-chercheur CEA, CEA - Centre de Grenoble	Directeur de thèse
Thomas FORTIN Ingénieur-chercheur CEA, CEA - Centre de Grenoble	Co-encadrant de thèse

Rapporteurs :

Delphine MURIAUX
DIRECTRICE DE RECHERCHE, CNRS délégation Occitanie Est
Rodolphe ANTOINE
DIRECTEUR DE RECHERCHE, CNRS délégation Rhône Auvergne

Thèse soutenue publiquement le **6 novembre 2024**, devant le jury composé de :

Christophe MASSELON, DIRECTEUR DE RECHERCHE, CEA centre de Grenoble	Directeur de thèse
Delphine MURIAUX, DIRECTRICE DE RECHERCHE, CNRS délégation Occitanie Est	Rapporteure
Rodolphe ANTOINE, DIRECTEUR DE RECHERCHE, CNRS délégation Rhône Auvergne	Rapporteur
Elisabetta BOERI ERBA, DIRECTRICE DE RECHERCHE, CEA centre de Grenoble	Examinatrice
Victor WEISS, ASSOCIATE PROFESSOR, Technische Universität Wien	Examinateur
Marc JAMIN, PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Président

"**This page was intentionally left blank.**"

Dedikim

Ky disertacion i dedikohet gjyshes sime, **Vjoleta Sino**, e cila u shua më 30 qershor 2022. Ishte ajo që më rriti dhe më edukoi gjatë një pjese të madhe të fëmijërisë sime, deri sa erdha në Francë, dhe ishte një nga njerëzit e parë që desha të takoja sa herë që kthehesha në Gjirokastër. Do të kisha dashur me gjithë shpirt të më shikonte të bëhesha Doktor i shkencave, dhe të ndihej e nderuar për edukimin që më dha. Më mungon shumë dhe do të jetë gjithmonë e paharruar në zemrën time !

"**This page was intentionally left blank.**"

Acknowledgements - Remerciements - Falenderimet

J'aimerais remercier mes directeurs de thèse initiaux, **Sébastien Hentz** et **Christophe Masselon**, de m'avoir accueilli au sein de leur équipe NEMS-MS (Nano-electromechanical Sensors Mass Spectrometry) et de m'avoir permis de mener ce projet de thèse, financé par GRAL (ANR-17-EURE-0003). Je tiens à les remercier particulièrement d'avoir accepté ma candidature pendant une période aussi compliquée que celle de la pandémie de COVID-19, sans même me connaître au préalable.

Même si Sébastien n'a été présent que durant la première année de ma thèse, son expertise approfondie sur les NEMS et ses projections en lien avec la biologie m'ont permis de comprendre rapidement cette technologie, qui était toute nouvelle pour moi. Il m'a également aidé à trouver ma place en tant que spécialiste en biologie au sein de l'équipe NEMS.

Je remercie Christophe de m'avoir montré ce que signifie être un chercheur du XXIe siècle, et particulièrement de mener des recherches dans des domaines interdisciplinaires comme celui de ma thèse. Nos échanges scientifiques ont été enrichissants et m'ont aidé à mieux appréhender le domaine de la spectrométrie de masse native, qui m'était totalement inconnu au début de ma thèse. Ces discussions m'ont permis de prendre pleinement en main ce projet si important pour moi : ma thèse. De plus, elles m'ont aussi donné l'autonomie nécessaire pour rencontrer, échanger, et collaborer avec de nombreuses personnes au cours de mes travaux, probablement l'un des aspects de la recherche que j'ai le plus apprécié.

Je tiens à remercier particulièrement **Thomas Fortin**, qui a eu un impact très significatif sur mon projet de thèse. Dès le début, sa logique mathématique et son esprit rigoureux m'ont donné des perspectives extrêmement enrichissantes, qui complètent mes idées parfois excentriques et débordantes d'imagination. Cela m'a poussé à fournir des efforts pour mieux structurer mes idées et à les organiser de manière plus compréhensibles pour mes collaborateurs. Je souhaite particulièrement le remercier d'avoir pris en charge intégralement mon encadrement pendant la période la plus critique de ma thèse, à savoir la rédaction et la préparation de l'oral durant la majeure partie de ma quatrième année. Son soutien empathique et son suivi attentif à ce moment crucial ont été d'une valeur inestimable jusqu'à ma soutenance. Je le considère comme un bon ami avec qui j'aurai grand plaisir à collaborer à l'avenir.

Je tiens à remercier **Rodolphe Antoine** et **Delphine Muriaux**, qui ont accepté d'être rapporteurs pour ce manuscrit, ainsi qu'**Elisabetta Boeri-Erba, Victor Weiss,** et **Marc Jamin**, qui ont accepté d'évaluer mon travail en tant qu'examinateurs.

Je tiens également à remercier les membres de mon CSI (Comité de Suivi Individuel), **Rob Ruigrok, Sarah Sanglier-Cianférani,** et **Isabelle Michaud-Soret**, qui m'ont énormément aidé et conseillé durant ces quatre années de thèse, tant sur le plan scientifique que sur le plan humain et moral.

Je tiens à remercier tout particulièrement Isabelle, qui m'a apporté un grand soutien, notamment juste avant la soumission de mon manuscrit, en lisant attentivement ma thèse et en me donnant des commentaires détaillés, tant d'un point de vue scientifique que grammatical. Je remercie aussi **Jacques** et **Marion Coves**, pour les discussions sur mon manuscrit et ma soutenance, tout ceci dans un délai très court mais avec une grande pertinence.

***

Je remercie toutes les personnes qui m'ont encadré ou aidé pour la réalisation de certaines expériences :

- **Yohann Couté** et **Anne-Marie Hesse** pour leur aide dans la mise en place des expériences de protéomique *bottom-up* pour les T5-CLPs (*Capsid-Like Particles* du bactériophage T5) au sein de mon laboratoire d'accueil, EdyP (*Exploring the Dynamics of Proteomes*).
- **Julia Novion Ducassou** et **Alexandra Kraut** pour m'avoir aidé à apprendre la protéomique en un temps très court, ainsi que pour la réalisation des analyses de spectroscopie de masse *bottom-up* sur mes différents échantillons et le traitement des données sur Proline, au sein d'EdyP.
- **Kavya Clement** pour m'avoir montré les bases de l'instrument NTA (*Nanoparticle Tracking Analysis*) utilisé dans notre laboratoire.
- **Guy Schoehn** et **Daphna Fenel** pour m'avoir accueilli au sein de la plateforme de microscopie électronique de l'IBS (*Institut de Biologie Structurale*) et pour l'acquisition des images utilisées dans l'analyse par MET (*Microscopie Électronique en Transmission*) avec coloration négative de mes différents échantillons viraux.
- **Valentine Bolcato** et **Christophe Lemonias** (La PTA - *Plateforme Technologique Amont*) pour m'avoir encadré et énormément aidé dans l'apprentissage des mesures avec la MEB (*Microscopie Électronique à Balayage*).

- **Pascal Fender** (IBS) pour m'avoir fourni des échantillons de HAdV (*Human Adenoviruses*), de sérotype 3 et sérotype 5, ainsi que pour les échanges très instructifs après analyses par MET sur ces échantillons.
- **Jean-Luc Pellequer** et **Jean-Marie Teulon** pour m'avoir accueilli au sein de la plateforme d'imagerie par AFM (*Microscopie à Force Atomique*) de l'IBS, et pour leurs nombreuses tentatives d'acquisition des images AFM des T5-CLPs.
- **Doriane Delafosse** (Abiolab à Montbonnot-Saint-Martin) pour m'avoir fourni des échantillons de norovirus murins (MNV) et de coliphage MS2, ainsi que pour les échanges détaillés concernant la préparation de ces échantillons.
- **Stéphane Mazlan** et **Marianne Marchioni** (Izon Science - Lyon) pour leur aide dans la mise en place et la réalisation des expériences de TRPS (*Tunable Resistive Pulse Sensing*) avec mes T5-CLPs.

***

I would also like to thank the many collaborators I had the pleasure of working with during this thesis project:

- First of all, our collaborators at **CEA – LETI (Grenoble)**, namely: **Marc Sansa** and **Wioletta Trzpil**. They were very cooperative and helpful in our discussions regarding the NEMS detectors, providing all the necessary knowledge and technical expertise when required.

- Next, our collaborators at **I2BC (Paris-Saclay)**, namely: **Pascale Boulanger** and **Nicolas Ducrot**. Even though my PhD project did not have an official collaboration with them, the T5 bacteriophage team at I2BC was kind enough to provide some of their samples for my analysis, and I even had the chance to spend a week in their lab to produce my own batch of T5-CLPs. I would like to especially thank Nicolas for providing me with access to his thesis (currently under confidential restriction), which allowed me to better understand the biology of the T5-CLPs.

- I would also like to thank the **CEA – LITEN (Grenoble)** team (collaborated thanks to the VIA-NEMS project), namely: **Sébastien Artous** and **Bastien Pellegrin**. Their assistance in setting up the SMPS measurements for aerosolized nanoparticles and their knowledge on aerosols was very helpful in confirming my hypothesis regarding the propagation of aerosolized viruses.

- Lastly, I would like to thank the **SAWN team at IFW Dresden (Germany)** (collaborated thanks to the AERO-NEMS project), namely: **Andreas Winkler** and **Mehrzad Roudini**. Their expertise in SAWN and their zealous desire to implement this technology in biological projects provided new insights and fruitful discussions about the necessity of having generalist interdisciplinary scientists in today's research landscape.

***

Je veux remercier les membres de l'équipe NEMS-MS avec qui j'ai eu la chance et le plaisir de travailler jour après jour, et avec qui j'ai partagé d'incroyables moments pendant ces quatre années. Et que j'ai le plaisir de les considérer comme mes amis.

Tout d'abord, **Adrien Reynaud**, qui a été mon mentor technique principal jusqu'au début de 2022. Il m'a appris à travailler avec le banc NEMS-MS #1, et ensemble, nous avons passé de longues heures à essayer de faire marcher la machine à nos besoins. Nous avons également mis sur pied le prototype #2 du banc NEMS-MS, souvent en travaillant au rythme de la musique disco jusque tard dans la soirée. Son empathie et sa manière logique de résoudre les problèmes m'ont énormément aidé à me sentir intégré dans une équipe aussi interdisciplinaire.

La deuxième personne avec qui j'ai pris un immense plaisir à travailler est le jeune débrouillard **Louis Dartiguelongue**. Sa positivité et son optimisme m'ont permis de lui confier les principales expériences côté instrumentation. Grâce à lui, j'ai trouvé la concentration nécessaire pour me tourner vers des expériences plus approfondies en biologie structurale. Cela m'a aussi permis d'aller plus loin dans la recherche de futurs collaborateurs et d'échantillons à analyser. Son honnêteté et son grand esprit ont été un véritable rayon de soleil pendant certaines des journées les plus difficiles.

I would also like to thank my young interns, with whom I had the utmost pleasure of guiding through their personal research developments and, hopefully, toward success in their future projects.

Firstly, **Danille Sclafani**, a brilliant young student in chemical engineering. Her perseverance and always-positive spirit, coupled with her readiness to tackle new challenges head-on, gave a significant boost to one of my key projects in just two months. She optimized the NTA instrument in our lab, starting from implementing the syringe pump to producing the first draft of the Python code to retrieve and analyze NTA data. I am certain that her PhD at

Princeton (and now at Caltech) will be a success and that she has a bright scientific career ahead of her.

Secondly, **Wassim Hannouf**, for whom I was an unofficial advisor, but for whom I am grateful to have provided guidance. Although his initial project did not work out, his hardworking nature stood out to me. I deeply appreciated his meticulous assistance in performing repetitive experiments, and always trying to improve the quality control of our previous work with the NEMS-MS. He even sought to go further by trying to establish more solid working protocols. I was delighted to see him excel during his Master 2 thesis defense, where he presented this very work with confidence and success.

***

Je tiens également à remercier l'intégralité du personnel de mon laboratoire d'accueil, EdyP (*Exploring the Dynamics of Proteomes*). Ce fut une expérience très enrichissante de travailler dans une équipe avec une si grande plateforme de protéomique, qui m'a donné l'opportunité d'observer et d'interagir de très près avec ce domaine de recherche en pleine expansion.

Cette expérience m'a permis de rencontrer des personnes incroyables et charmantes, dont je suis profondément reconnaissant qu'elles aient croisé mon chemin. Je tiens à remercier tout particulièrement **Michel Jaquinod** pour ses idées précieuses sur l'interaction entre la biologie et la spectrométrie de masse, ainsi que pour ses plaisanteries incessantes et sa bonne humeur, qui ont toujours apporté une belle atmosphère au laboratoire.

Je tiens également à remercier **Pascal Albanese, Sabine Brugière** et **Alexandra Kraut** pour leur bienveillance ainsi que pour leurs commentaires pertinents lors de mon oral blanc, qui m'ont permis de mieux me préparer à ma soutenance.

Je voudrais également remercier tous les doctorants avec qui j'ai eu le plaisir d'interagir pendant cette période.

- **David**, pour son optimisme et sa détermination à continuer à faire du sport, même si ce n'était pas toujours dans mon programme de thèse. Et aussi de m'avoir aidé à cueillir mes 3 kg de cèpes.

- **Lucas**, pour sa bonne humeur et pour avoir été l'une des rares personnes avec qui j'ai pu avoir de véritables discussions sur le football au laboratoire. Je n'oublierai jamais les montagnes russes émotionnelles vécues en regardant la finale de la Coupe du monde entre la France et l'Argentine.

- **Hassan**, pour avoir été l'une des personnes les plus intelligentes et empathiques que j'ai eu le plaisir de rencontrer au cours de cette thèse. Il était toujours là pour m'écouter et me conseiller lorsque j'en avais besoin. Je suis plus qu'heureux de le considérer comme l'un de mes meilleurs amis.
- **Zakaria**, qui, bien qu'il ne fasse pas partie d'EdyP, a été l'une des rares personnes à rester en contact avec moi durant cette dernière année de doctorat. Je suis particulièrement reconnaissant de le considérer comme un ami, notamment pour son aide précieuse et ses conseils tout au long de notre longue quête vers la nationalité française.
- It was also thanks to EdyP that I was able to meet **Henry**, who is not only an incredibly intelligent and mature person, but also an exceptional mentor and a great friend.

***

Enfin, ce doctorat m'a permis de rencontrer ma belle **Julie**, qui a été mon pilier pendant cette quatrième année de doctorat, sans doute la plus difficile. Notre relation a commencé à un moment où mon avenir était réellement incertain, tant en ce qui concerne la fin de mon contrat de doctorat que la possibilité d'obtenir la nationalité française. Sa volonté de rester à mes côtés pendant tout ce temps, ainsi que son amour et sa compassion constants, ont été comme une gorgée d'eau fraîche au milieu du désert. Grâce à elle, j'ai pu continuer et persévérer pendant une année entière jusqu'à ma soutenance. Je pense que la plupart des gens auraient abandonné dans de telles circonstances, mais elle ne l'a pas fait. C'est ce qui a rendu notre relation encore plus forte, et j'espère qu'elle continuera à l'avenir.

J'aimerais également remercier tous mes amis à Grenoble, que je connais depuis la Licence (**Élise, Nicolas, Manil, Baptiste S.**), qui ont été de merveilleux amis et qui sont restés à mes côtés depuis mon arrivée en France.

Je tiens particulièrement à remercier **Finan Rodinson**, mon meilleur ami, ainsi qu'à cette occasion toute la famille Rodinson (**Maryvonne et Daniel, Eva et Gabriel**), qui a été ma famille d'accueil en France. Ils m'ont toujours donné énormément d'amour et de compassion dès le début, et je leur en serai éternellement reconnaissant.

Je veux également remercier tous mes amis rencontrés pendant le Master, ceux qui n'étaient pas très loin géographiquement (**Luis, Camila et Francesco, Baptiste V.-S., Vincent, Florent, Maxime, Benoît, Jonathan**), avec qui j'ai partagé de nombreux petits cafés ou de

grands verres d'alcool. J'espère que nos relations resteront aussi fortes et authentiques à l'avenir.

And also, to all my other geographically distant but yet very close friends (**Ragnar, Daphné, Chintan, Esther, Thijs, Elaine, Alexandre, Mirabel**), thank you for knowing how to be by my side and for cheering me up in moments of need. Thank you all for being my friends. I am a better person for having crossed life's paths with you all!

***

Do të desha të bëj një falenderim dhe një dedikim të veçantë për profesorin tim të frëngjishtes, Profesor **Vasil Ristani,** i cili u nda nga jeta në korrik të vitit 2020, pak kohë para se të nisja doktoraturën. Prandaj kam menduar se një eulogji e shkurtër për këtë njeri të madh do të ishte vlerësimi im për kujtimin e tij në një farë mënyre.

> Do të thosha me gojën plot dhe me bindjen më të madhe se nuk do të isha këtu ku jam sot, në qoftë se nuk do të kisha pasur fatin e madh të njoh dhe të jem një nga qindra, ose ndoshta mijëra, studentët e Profesor Ristanit. Mund të them me gojën plot që ka qenë pedagogu më i mirë që kam njohur ndonjëherë gjatë të gjithë shkollimit tim në Shqipëri, dhe se ka qenë gjithmonë një frymëzim për vlerat që ishte në gjendje të transmetonte tek nxënësit e tij. Do të kisha dashur që ta takoja për një herë te fundit dhe ta falënderoja nga afër për çdo gjë që bëri të mundur për mua, me anë të dijes dhe udhëheqjes që më ofroi, për të më mundësuar ardhjen time në Francë.

Qoftë dheu sa më i lehtë për këtë njeri grandioz. Repose en paix cher Professeur !

***

Do të desha të falenderoja te gjithë të afërmit e mi, që besuan tek unë gjatë gjithë kësaj kohe. Në veçanëri dajkot e mi, **Bekimin** dhe **Gjergjin**, që erdhën deri në Grenoble një vit më parë (dhjetor 2023) për të më dhënë kurajon dhe besimin e nevojshëm që do t'ia dilja mbarë dhe të shkoja deri në fund të kësaj aventure.

Dhe së fundmi do të desha të falenderoja **familjen time**. Në radhë të parë, prindërit e mi, **Ilirin** dhe **Zhanetën**, që më rritën, më edukuan dhe më dhanë gjithë dashurinë e tyre. Megjithëse kam qënë larg shtëpisë, tanimë prej mbi 9 vjetësh, ata kanë ditur gjithmonë të më këshillojnë dhe të më mbështesin pa ndalim gjatë gjithë kësaj kohe. Kanë qënë ata që më kanë udhëhequr në shumë nga vendimet e mëdha të jetës time. Që nga vendimi për të ardhur në Francë, kryesisht falë frankofilisë së babit tim; tek zgjedhja e Grenoble, për të mos qënë shumë larg nga malet; tek nxitja kur më mungonte motivimi gjatë Licencës; dhe në veçanti gjatë dokotraturës, kur isha duke pyetur veten nëse kryerja e një doktorature ishte zgjedhje më e mirë

për mua gjatë pandemisë të COVID. Në veçanti do të doja të falenderoja mamanë time, që më dëgjonte dhe më kuptonte për çdo problem dhe stres që hasja në jetën e përditshme. Kjo ndihmesë ka qënë gjithashtu e lehtësuar falë motrës time, **Vanesës**, e cila është dhe shoqja ime më e mirë, e cila me humorin dhe shakatë e saj ka ditur gjithmonë të më ngrëjë moralin kur ka qënë e nevojshme. E di që kjo doktoraturë ishte po kaq e gjatë dhe e lodhshme si për mua dhe për ata, por jam i sigurt se pa ata nuk do t'ia kisha dalë dot mbanë t'i shkoja deri në fund. Prandaj kjo fitore është po aq e imja sa edhe e tyre. Dhe gëzohem që i kam bërë krenarë tani që jam bërë doktori i parë, dhe tanimë, edhe shtetasi i parë francez, në familjen Çumaku. Ju dua shumë fort dhe e konsideroj veten një njeri me shumë fat që kam një familje të tillë!

Résumé de la thèse

Les progrès du séquençage du génome ont relégué l'évaluation phénotypique des virus au rang de simple complément du génotypage. À la lumière des récentes avancées dans les approches de spectrométrie de masse (MS) native pour la caractérisation des virus, nous pensons que les mesures de masse peuvent très bien compléter les approches génomiques, à des rangs taxonomiques élevés. Pour démontrer cette approche, nous avons calculé la masse totale des virions des virus respiratoires humains (HuReV), et nous les avons regroupés dans leurs familles respectives, en se basant seulent à leur masse totale. J'ai nommé ce processus de classification taxonomique "phylobarique" ("phûlon" : groupe ; et "baros" : poids). Nous soutenons la nécessité de constituer une base de données expérimentales, qui pourrait être utile en cas de pandémie, pour rapidement classé les virus pathogènes et aider au triage des patients. En outre, la mesure de la masse des virus pourrait ajouter une dimension à l'analyse phylogénétique, d'une manière phénétique (morphologique), permettant d'évaluer la proximité évolutive au-delà des simples séquence homologues. Pour affirmer ce cadre théorique, des efforts supplémentaires ont été consacrés à l'amélioration de l'instrumentation de la spectrométrie de masse à nano-résonateurs (NEMS-MS), une approche expérimentale unique permettant de mesurer des particules virales individuelles, non chargées et aérosolisées. Dans une gamme qui recouvre parfaitement la gamme de masse estimée des HuReV (5-6000 MDa). Ces travaux ont conduit au développement d'un nouveau prototype NEMS-MS, plus compact et plus facile à utiliser. Les travaux ultérieurs se sont concentrés sur la caractérisation de la relation masse-structure des virions en effectuant des analyses structurelles d'un modèle de particule analogue à un virus (VLP) qui est la capside du phage T5, un système modèle bien étudié par NEMS-MS. Les principales comparaisons ont porté sur l'étude des différences de structure entre les T5-VLPs produites par infection ou par recombinaison, et sur la vérification de la présence ou de l'absence des protéines de décoration sur ces VLPs. En outre, j'ai également étudié de nouvelles sources alternatives de particules virales bien étudiées, telles que les adénovirus humains, les norovirus murins et le coliphage MS2. Une approche méthodologique utilisant une variété de techniques de caractérisation de particules uniques en phase liquide et gazeuse a été développée pour valider le lien entre la masse et la structure. L'idée principale était qu'en effectuant des mesures à chaque étape de l'analyse, le sort de l'échantillon est connu à chaque étape, de la préparation de l'échantillon à la nébulisation, jusqu'à ce qu'il atteigne le détecteur de masse. Nanotracking Analysis (NTA) a été utilisée pour les mesures liquides ;

Scanning Mobility Particle Sizer (SMPS) a été utilisé pour les mesures dans les aérosols et la microscopie électronique à balayage (MEB) a servi de méthode de contrôle tout au long de la chaîne. Des mesures supplémentaires en liquide par ME à transmission (TEM par coloration négative), Tunable Resistive Pulse Sensing (TRPS), proteomique ‘bottom-up’ et microscopie à force atomique (AFM) ont également été effectuées en tant que contrôles. Ce pipeline peut être facilement adapté à tous virus d'intérêt futur. Ce travail constitue une analyse détaillée du transfert des virus de la phase liquide à la phase gazeuse, qui pourrait finalement être comparée à l'aérosolisation naturelle des virus respiratoires pathogènes. En conclusion, cette thèse apporte la preuve que la "masse moléculaire" des virions entiers représente une propriété phénotypique virale intrinsèque et distinctive, qui peut être utilisée pour identifier et catégoriser les virus. En outre, l'étude des virus lors de leur transfert de la phase liquide à la phase gazeuse fournit des informations importantes sur la relation entre leur structure native et leur masse moléculaire.

Summary of the thesis

Advances in genome sequencing have relegated phenotypic evaluation of viruses to mere complements of genotyping. In light of recent advances in native mass spectrometry (MS) approaches for virus characterization, we believe that mass measurements may very well complement genomic-sequence-based phylogenetic approaches at high taxonomic ranks. To demonstrate the feasibility of such an approach, we computed the total virion mass of human respiratory viruses (HuReVs). We thus observed that they could be grouped into their respective families based solely on mass information. I termed this process of taxonomic separation "phylobaric classification" (from the ancient Greek 'phûlon': group, tribe; and 'baros': weight). We argue about the need to compile a database of experimental virus masses, which could be helpful in the event of a pandemic, to rapidly determine the class of viral pathogens and help triage patients. Furthermore, measuring virus mass could add a dimension to phylogenetic analysis, in a phenetic (morphological) manner, by allowing evolutionary proximity to be assessed beyond the sequence-based closest relatives. In continuation of this theoretical framework, additional efforts were devoted to improving the instrumentation of nanoelectromechanical sensors mass spectrometry (NEMS-MS), an experimental approach to perform single-virion MS measurement. This unique method can measure uncharged, individual aerosolized viral particles over a range that perfectly overlaps the estimated mass range of HuReVs (5-6000 MDa). This work led to the development of a newer NEMS-MS prototype, more compact, and user friendly. Subsequent work focused on characterizing the mass-to-structure relationship of virions by performing structural analyses of a model Virus Like Particle (VLP) which is the phage T5 capsid, a model system well-studied by NEMS-MS. The main comparisons were in investigating the structure differences between T5-VLPs produced by infection or recombination and checking the presence or absence of the decoration proteins on these capsids. In addition, I also investigated new alternative sources of well-studied viral particles, such as Human Adenoviruses, Murine Noroviruses, and MS2 coliphage. A methodological approach using a variety of liquid and gas-phase single particle characterization techniques was developed to validate the connection between mass and structure. The main idea was that by performing measurements at every step of the analysis, the fate of the sample is known at each step, from sample preparation to nebulisation, to reaching the mass detector. Nanotracking Analysis (NTA) was used for liquid measurements; Scanning Mobility Particle Sizer (SMPS) for aerosol measurements and Scanning Electron Microscopy (SEM) as a control

of deposited particle along the whole pipeline. Additional measurements in liquid by negative staining Transmission Electron Microscopy (TEM), Tuneable Resistive Pulse Sensing (TRPS), bottom-up proteomics and Atomic Force Microscopy (AFM) were also performed as controls. This pipeline can be applied to all viruses of interest, as much as safety restrictions allows. This work constitutes a detailed analysis of viruses transfer from the liquid to the gas phase, which could ultimately be compared with the natural aerosolization of pathogenic respiratory viruses. In conclusion, this thesis presents evidence that the "molecular mass" of whole virions represents an intrinsic and distinctive viral phenotypic property, which can be used to identify and categorize viruses. In addition, investigations of viruses during transfer from the liquid to the gas-phase transfer provide important information in relating their native structure to their total mass.

Quotes

“I would rather have questions that can't be answered, than answers that can't be questioned.”

“Study hard what interests you the most in the most undisciplined, irreverent and original manner possible.”

— Richard P. Feynman

"Un aspect curieux de la théorie de l'évolution est que tout le monde croit la comprendre."

— Jacques Monod

“We are a way for the cosmos to know itself.”

— Carl Sagan

“Eins og blindur maður í orgíu, þurfti ég að þreyfa mig áfram.”

— L. Nielsen

“Abundance of knowledge does not teach men to be wise.”

— Heraclitus

"Dans les champs de l'observation, le hasard ne favorise que les esprits préparés."

— Louis Pasteur

"**This page was intentionally left blank.**"

Thesis outline

Summary of Contents

Dedikim 2

Acknowledgements - Remerciements - Falenderimet 4

Résumé de la thèse 12

Summary of the thesis 14

Quotes 16

Thesis outline 18

Summary of Contents 18

Glossary (in alphabetical order) 24

General Introduction 26

A brief historical background 26

Research Problem and Significance 28

Research Objectives 29

Methodology Overview and Thesis Structure 29

1 Chapter 1 : Proposing the 'phylobaric' classification of human respiratory viruses (HuReVs). .. 31

1.1 Introduction to Chapter 1 31

1.1.1 What is a virus ? 31

1.1.2 Viruses as Macromolecules 42

1.2 The experimental design 44

1.2.1 Proof of Concept - Focus on human respiratory viruses (HuReVs) 44

1.2.2 Virus data 46

1.2.3 Statistical methods 47

1.3 Biophysical perspective 47

1.3.1 Methods: The biophysical perspective 47

1.4 Macromolecular perspective 52

1.4.1 Method: Genome mass calculation 53

1.4.2 Genome mass calculation 53
1.4.3 Method: Calculating the mass of the ‘genome container’ 57
1.4.4 Calculating the total mass of the virion 59
1.4.5 Method Limitations 63
1.5 Conclusions on Chapter 1 67
1.6 Outlook and Perspectives on Chapter 1 68
2 Chapter 2: Methods for virion mass evaluation and the place of NEMS-MS 71
2.1 Introduction of Chapter 2 71
2.2 Overview of total virion mass measurement methods 71
2.2.1 Methods capable to perform virion mass measurements. 73
2.3 The NEMS-MS instrumentation 76
2.3.1 A mass range adapted for virion analysis. 77
2.3.2 Architecture of the first prototype for NEMS-MS 79
2.3.3 Nebulisation 80
2.3.4 Aerodynamic lens 88
2.3.5 NEMS and the Neutral mass measurement 89
2.4 Proof of concept: bacteriophage T5 Viral Like Particles (VLPs) 99
2.5 Conclusions on Chapter 2 101
3 Chapter 3: Developments on the NEMS-MS instrumentation as an Analytical Method 103
3.1 Introduction of Chapter 3 103
3.2 Characterization of the first NEMS-MS bench (prototype #1) 103
3.2.1 Evaluation on the 15 nm and 30 nm GNPs 104
3.2.2 Limitations of the prototype #1 106
3.3 Prototype #2 : the mini-NEMS-MS bench 107
3.3.1 Low pumping requirements 109
3.3.2 Benchmarking the NEMS-MS prototype #2 115
3.3.3 Evaluation the transmission efficiency 116
3.4 NEMS-MS: an analytical method 118
3.4.1 Temporal Influence on Particles Deposition 118

3.4.2 Peculiar results for 20 nm GNPs 119
3.4.3 Particle desolvation on the NEMS resonator 121
3.4.4 Viral particle desolvation issues 124
3.4.5 Overview and perspectives regarding the analytical capabilities of the NEMS-MS bench. 129
3.5 Conclusions for Chapter 3 129
4 Chapter 4: Establishing a structural biology pipeline prior to the NEMS-MS analysis 132
4.1 Introduction of Chapter 4 132
4.1.1 The three states of a virus during the NEMS-MS analysis 133
4.2 Materials: Viral samples 134
4.2.1 Bacteriophage T5 Capsid-Like Particles (T5-CLPs) as models 134
4.2.2 Other viral samples and Sample related limitations. 143
4.3 Summary of the methods used for VLP characterisation 146
4.3.1 The literature on virus characterisation methods 146
4.3.2 State of the art on the T5-CLPs 153
4.4 Design of the in-line analysis prior to the NEMS-MS bench 154
4.4.1 Method related limitations, focus on SEM and AFM. 156
4.5 Conclusions on Chapter 4 159
5 Chapter 5: Optimising the Analytical & Biophysical Methods 160
5.1 Introduction of Chapter 5 160
5.1.1 Preface: The proper view of objects 161
5.2 Negative staining Transmission Electron Microscopy (TEM) 162
5.2.1 Setting the stage, indications from I2BC 162
5.2.2 Principle of TEM 164
5.2.3 Optimisation of the TEM experiments 166
5.2.4 Conclusion on TEM 185
5.3 Nanoparticle Tracking Analysis (NTA) 186
5.3.1 Principle of NTA 186
5.3.2 NTA optimisation 190

5.3.3 Conclusion on NTA 199
5.4 Tunable Resistive Pulse Sensing (TRPS) 200
5.4.2 Conclusion on TRPS 204
5.5 Bottom-up Proteomics, using LC-MS 204
5.5.1 Principle 204
5.5.2 Quantification in Bottom-Up Proteomics 206
5.5.3 Experimental design 211
5.5.4 Filtering the data 213
5.5.5 Conclusion on Proteomics 216
5.6 Optimising nebulisation and measuring in the gas phase 217
5.6.1 Methods for aerosol evaluation 218
5.6.2 Nebulisation optimisation 221
5.7 Discussion and Perspectives on Chapter 5 227
6 Chapter 6: Characterizing T5-CLPs as vaccine candidates 230
6.1 Introduction of Chapter 6 230
6.2 Concentration evaluation 233
6.3 Aging effects on non-decorated phCLPs (Inter-batch comparison) 235
6.3.1 Results by TEM 235
6.3.2 Results by NTA 238
6.3.3 Results by Bottom-up Proteomics 242
6.3.4 Conclusions on the inter-batch comparison 242
6.4 Production method & Decoration status evaluation 243
6.4.1 Results by TEM 244
6.4.2 Results by NTA 245
6.4.3 Results by TRPS 248
6.4.4 Conclusion on the Biophysical characterisation 250
6.4.5 Results by Proteomics 251
6.4.6 Conclusions on the Production method & decoration status 255
6.5 Benchmarking SMPS with the T5-CLPs 256

6.6 Conclusions on Chapter 6 258
6.6.1 A more philosophical view 259
General Conclusion 261
Key findings in the context of the initial PhD objectives 261
Main Contributions 262
Limitations 262
Perspective and Conclusions 263
Summary of my scientific production during my thesis, *and those to come* 265
7 Annexes 267
7.1 Chapter 1: Proposing the 'phylobaric' classification of human respiratory viruses (HuReVs). 267
7.1.1 Definitions and Clarifications 267
7.1.2 Relevant Tables 270
7.1.3 Describing HuReVs as aerosolised particles 273
7.2 Chapter 2: Methods for virion mass evaluation and the place of NEMS-MS 277
7.2.1 Compiling the data for measurements reported in the available literature. 277
7.2.2 NEMS related tables 280
7.3 Chapter 3: Developments on the NEMS-MS instrumentation as an Analytical Method 281
7.3.1 Supplementary (NEMS-MS bench #1) 281
7.3.2 Supplementary (mini NEMS-MS bench - prototype #2) 284
7.4 Chapters 4 & 5: Materials and Methods 285
7.4.1 Materials 285
7.4.2 Methods 292
7.4.3 Supplementary on nebulisation optimisation 305
7.4.4 Comparing the DMA to the IMS, and how to relate it to the virus scale! 309
7.5 Chapter 6: Characterizing T5-CLPs as vaccine candidates 314
7.5.1 Concentration evaluation at I2BC 314
7.5.2 TEM related information. 314
7.5.3 NTA related information. 315

7.5.4 TRPS related information. 317

References 318

Glossary (in alphabetical order)

- AFM: atomic force microscopy.
- Allometry: how the characteristics of living creatures change with size. [link]
- AUC: analytical ultra-centrifugation.
- Baltimore classification : classification based on the process on which viruses synthesize their messenger RNA (mRNA), splitting viruses into 7 groups, as defined by Baltimore [1].
- BNID: Bionumbers Identification number, as reported by Milo et al.[2]
- bp (unit): base pairs; same as nucleic acids; same as nucleotides.
- cc (unit) : cubic centimetre, used as g/cc = g/cm^3.
- CD-MS: charge-detection mass spectrometry.
- COVID-19: Coronavirus disease 2019.
- cryoEM: cryogenic electron microscopy.
- cryoET: cryogenic electron tomography.
- CLP: capsid-like particle.
- Da (unit): daltons, where 1 Da = 1.66054·10^{-27} kg; Megadaltons (MDa) [= 10^6 Da].
- Encapsidated viruses: viruses whose genome is encased within a capsid.[3]
- GEMMA: gas-phase electrophoretic mobility analyser.
- HuReVs : human respiratory viruses (**NB**. term that I invented for a simplified shorthand writing. The idea was to avoid using the term HRV, which already is used for Human rhinoviruses).
- ICTV : International Committee on Taxonomy of Viruses.
- MassP (MP): mass photometry.
- MS: mass spectrometry.
- MW: molecular weight (i.e. molecular mass).
- NEMS-MS: nano-electro-mechanical sensor mass spectrometry.
- nESI QRTOF-MS: nano-electrospray ionization quadrupole reflection time-of-flight Mass Spectrometry.
- nMS: native mass spectrometry.
- PCR: polymerase chain reaction.
- phylobaric: from old Greek, for *phylon*= family, tribe; and *baros* = mass, weight. (**NB**. term that I coined, regarding my observation that viruses of similar mass tend to be in the same viral families.)

- QTEM: quantitative transmission electron microscopy.
- (Q)STEM: (quantitative) scanning transmission electron microscopy.
- SANS: small angle neutron scattering.
- SARS-CoV-2: severe acute respiratory syndrome coronavirus 2, strain of coronavirus that causes COVID-19.
- SAXS: small angle X-ray scattering.
- SEC-SLS: size exclusion chromatography- static light scattering.
- SP-ICP-MS: single particle inductively coupled plasma mass spectrometry.
- Sv = Svedberg unit (10^{-13} s = 100 fs), the unit for measuring sedimentation coefficients.[4]
- T-number: Triangulation number, as defined by Caspar and Klug.[5]
- Torr (unit): pressure unit, where 1 Torr = 0.0013 atm = 133.3 Pa.
- vEM: volume electron microscopy.
- Virion: When not inside an infected cell or in the process of infecting a cell, viruses exist in the form of independent viral particles, or virions, consisting of (i) genetic material, i.e., long molecules of DNA or RNA that encode the structural proteins by which the virus acts; (ii) a protein coat, the capsid, which surrounds and protects the genetic material; and in some cases (iii) an external envelope of lipids. [link]
- VLP: virus-like particles (en français: « particules pseudo-virales »).

General Introduction

My PhD started during the Coronavirus disease 2019 (COVID-19) pandemic (declared on the 11th march 2020 by the World Health Organization-WHO {source}), which is still ongoing, at the moment of the writing of this thesis. The WHO made the assessment that COVID-19 was a pandemic, due to it being quote: "*Deeply concerned both by the alarming levels of spread and severity, and by the alarming levels of inaction*". The reasons for such concern are clearly visible by the drastic death toll it reached in the past four years, i.e. around 7.1 million people (data accessed in 17 September 2024 {source}). This is the reason why I wanted to be part of the acting fraction of the scientific community, in an effort to fight back the spread of this virus, in the short term, and hopefully help prevent any other human respiratory virus pandemics in the future.

A brief historical background

Some of the first measures that was were envisioned by most governments during the COVID-19 pandemic were the non-pharmaceutical interventions (NPIs), such as: Case isolation in the home; Social distancing of entire population.[6] And the testing was primarily performed by two methods {source} : PCR (Polymerase Chain Reaction)[7], the "gold standard" for COVID-19 tests, and immunological (antigen) tests[8]. These tests would not have been possible of course without first having the genome sequence, made possible thanks to NGS (Next-Generation Sequencing) techniques[9], of the severe acute respiratory syndrome associated coronavirus 2 (SARS-CoV-2).[10] But these tests very quickly showed their limitations. For example, PCR tests were quickly shown to fail when used in a surveillance regimen, due to the need a delocalised expert lab staff and equipment, having thus high cost, low frequency and slow turnaround times, resulting in the incapacity to quickly identify and isolate infected people, and the eventual resulting contact tracing.[11] On the other hand, the lateral flow immunoassays (also known as antigen rapid diagnostic tests), had limitations such as them not being efficient in diagnosing active cases, due to them being likely to report false-negatives in the early stages of infection, and they also miss infections among people who are immunocompromised (not able to produce antibodies).[12] Therefore, the possibility of using a rapid, non-targeted and physics-based approach seemed like an appealing choice.

Before PCR[13] and NGS[14] became common place in virus detection and classification, this task was performed only by physical characterization techniques, such as electron microscopy or sedimentation techniques to name a few.[4] These methods served to characterise parameters such as: the genome length and molecular mass, the particle density, the virion size and morphology, and also the total virion mass.[4] These were key parameters for virus classification, and many serve as metadata for phylogenetic evaluations, for what is inaccessible only through NGS techniques.[15,16] However, the interest in viral genome or total virion mass evaluation seems to have fallen behind as a less informative approach.[17] This fall of interest on virion mass evaluation can be attributed to two main reasons: a) the 'How': the methodological complexity of performing mass evaluation; and b) the 'Why': the lack of a clear question that relates virus biology to virion mass.

To answer the first question (the 'How'), the Mass Spectrometry (MS) community has been working painstakingly in the last three decades to put in place techniques that can measure native structures of larger and larger molecular mass.[18] We can highlight especially the development of Charge Detection MS (CDMS) and Nanoelectromechanical sensors MS (NEMS-MS).[18] I will talk more in details about the technical aspects of these methods in Chapter 2, and I will focus hereafter on the applications that these MS techniques as tools for structural biology.

Regarding the second question (the 'Why'), these improvements in native MS have primarily been made in order to render this method potent enough for a multitude of applications in structural biology, such as: proteoform profiling; non-covalent interactions of proteins with small ligands; and characterisation of protein assemblies (e.g. ribosomal or proteasomal particles; membrane-embedded protein complexes; Virus(-like) Particles, also called VLPs).[19] So, let us now have a brief look at the main usage of virus and VLPs in relation to MS.

The utilization of viruses and VLPs was very interesting from the very beginnings for native MS analysis, primarily because of their biophysical stability, allowing to establish whether such large assemblies would be intact in their nebulisation in the gas phase and into the mass spectrometer, and subsequently mass-analysed.[19] The current fields of investigation has been focused on: a) empty capsids and VLPs (which are the most studies samples, primarily thanks to their biosafety profiles being not infectious for humans); b) genome-loaded authentic viruses (especially interesting for viruses used for gene delivery and therapy); and c) more novel

samples such as: viral-like assemblies used for nanoparticle vaccines and synthetic VLP-like protein nanocages.[19]

However, all this work has focused on rather small (< 20 MDa) viruses and VLPs, primarily because of the difficulty to transmit these analytes to the mass analyser and deconvolute the mass spectra for larger mass, therefore making the investigation of larger viruses a very difficult task. In 2018, the NEMS-MS group (using the instrument patented by Sébastien Hentz [CEA-LETI] and Christophe Masselon [CEA-IRIG][20]) was the first to evaluate the mass of a virus (bacteriophage T5) over the 100 MDa range.[21] This charge independent method showed that it had the capacity to have a mass-independent resolution, opening the door to analysing even 'heavier' viruses.

Research Problem and Significance

In this thesis, the problem that I aimed to resolve was to primarily show that analysing the mass of the total virion should be revisited in the light of the scientific maturation of new MS techniques, which are getting better by the day.[19] I wanted to see first if the virus mass falls in a given mass range that was accessible by these methods. Then, make the link between this mass to the structural components making up these viruses. And, of course, this knowledge had to then be used to develop the NEMS-MS instrument to best suit the needs for viral analysis, with the aim of rendering into a robust structural biology technique.

In this thesis work, I demonstrated that mass is an intrinsic structural property of virus biology. Being able to measure it, could provide us with the necessary information to be able to distinguish between different viral families, without any prior knowledge about these viruses. I also developed the blueprint for the analytical pipeline that is needed to examine the macro-structural modification that may occur to these particles, primarily focusing on their drastic passage from the liquid to the gas phase. I hope that this knowledge would also allow future applications such as using virion mass for improving their detection in large settings, such as airports or hospitals.[22]

Research Objectives

This thesis therefore has three main objectives, which I separated as follows, Figure 1.

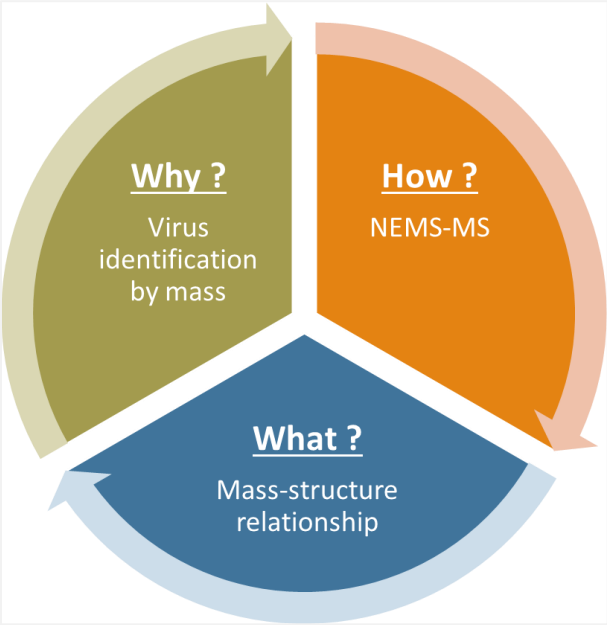


Figure 1. Thesis research objectives.

The 'Why', focused on the theoretical calculations regarding the virus identification by mass. The 'How' focused on applying the NEMS-MS instrument in evaluating the total virion mass, primarily working with our model virion, bacteriophage T5. As for the 'What', it consisted of developing a multitude of structural biology methods around the NEMS-MS instrument, in order to build a mass to structure relationship. With the final goal of building a sustainable pipeline so that any new virus of interest can be analysed with this similar protocol.

Methodology Overview and Thesis Structure

In the first axis of my research, mainly based on the current literature available, I was able to make a bioinformatic evaluation of HuReVs mass, from both the biophysical and biological-based perspective. I supported my first claim by having an in-depth look at human respiratory viruses (HuReVs). These data were compared to mass measurements from several techniques. . Then, I summarised the scientific toolbox currently available to perform virion mass analysis, and defined the mass range at which MS methods can be used to fill in the gap.

In the second axis of my research, I worked on optimising one of these native MS instruments, the NEMS-MS, developed in our laboratory.

In the third axis of my research (see Figure 2), I focused on improving this approach, both from an instrumental and from a biochemical perspective. Then, I optimised the analytical pipeline around the NEMS-MS, using a multitude of structural biology and biophysical techniques. The following figure, Figure 2, provides a concise overview of each chapter's content.

In the first chapter, I will focus on analysing the theoretical calculations of the virion mass in order to evaluate the virus mass window, and I will show how the virus mass can be used as a unique parameter to distinguish viral families from one another. Here I coined the name of this concept as '*phylobaric'*, from old Greek, for *phylon*= family, tribe; and *baros* = mass, weight.

In the second chapter, I will focus on the different methods that can be used to perform virion mass measurements. Which will be followed by the third chapter, regarding the developments that I was able to contribute to with respect to the NEMS-MS method.

In the fourth chapter, I will focus on detailing the bibliography that exists around other structural methods that can be used in collaboration to the NEMS-MS analysis. Which will be followed by the fifth chapter, regarding how I optimised these methods and put them to use in the specific analytical pipeline that I designed. Lastly the sixth chapter analyses the model sample of the bacteriophage T5 capsid like particles (T5-CLPs), with the perspective of establishing some possible critical quality attributes (CQA) for their usage as vaccines in the future.

This Figure 2 summarises the axes of research and the presentation of the work by chapters, as follows:

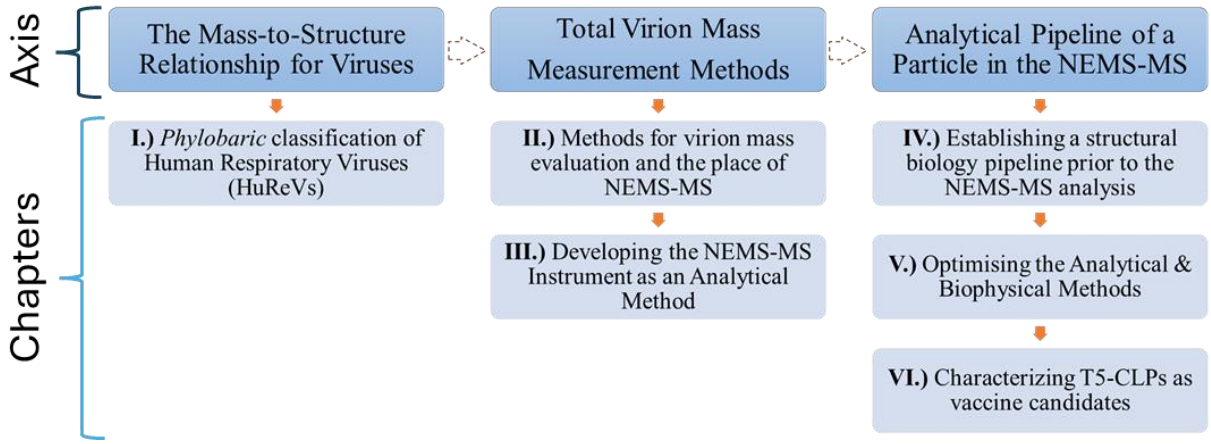


Figure 2. Schematic representation of the structure of the entirety of the document. The research axes are found in the dark shaded blue rectangles. The chapter numbers are shown in the light shaded blue rectangles, noted by the Roman numerals, .

1 Chapter 1 : Proposing the 'phylobaric' classification of human respiratory viruses (HuReVs).

1.1 Introduction to Chapter 1

In this chapter I aimed to make the connection between the mass-to-structure at the biological mesoscale, where viruses lie. And I went further to discuss about how mass can prove to be an innate characteristic of viruses, perhaps even to be used as an evolutionary trait or a biomarker for their detection.

This chapter is based on the following preprint, for which I am the first author:

Vaitson Çumaku, et al. *On the 'phylobaric' classification of human respiratory viruses.*, 09 Oct.2023, PREPRINT (V1) Research Square [https://doi.org/10.21203/rs.3.rs-3395557/v1]

1.1.1 What is a virus ?

The definition of a virus in the Encyclopaedia Britannica is as follows "an infectious agent of small size and simple composition that can multiply only in living cells of animals, plants, or bacteria."[23]. The term originated from the Latin word 'vīrus', which could be translated as 'poison' or 'slimy liquid'. This was primarily inspired by the definition of 'filtrable viruses', meaning particles that are small enough to pass bacterial filters.[24] It has to be noted here that this definition of a virus does not really describe a molecular structure, but the function of an entity. The fact is that viruses do not possess a metabolism of their own, and thus are quintessential parasites, i.e. they depend entirely on the host cell for almost all of their life-sustaining functions. That is why they are considered by many as non-living organisms[25]. Nevertheless they do present well defined macromolecular compositions, and have a 'replication life cycle'[26], as the one shown in Figure 3.[27]

1.1.1.1 The 'birth' of a virus

In Figure 3 the classical virus replication cycle is described using the Human metapneumovirus (HMPV), as was described in the work of Jones et al.[27]

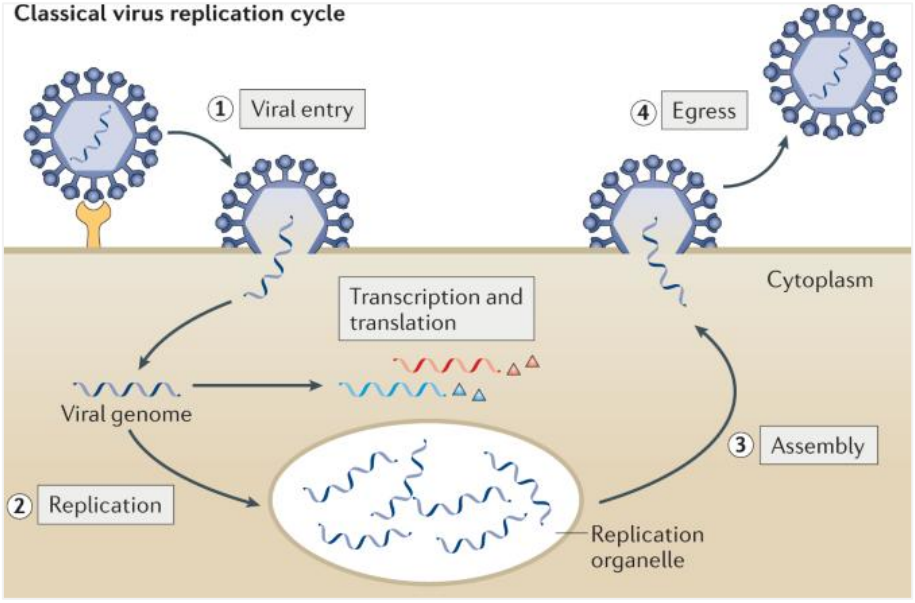


Figure 3. Classical virus replication cycle. Traditional depiction of the viral replication cycle with its series of defined, uniform steps, using human metapneumovirus (HMPV) as an example. HMPV virions bind to a target receptor and fuse with the host cell membrane (step 1). After fusion, viral RNA is replicated in cytoplasmic replication organelles (step 2) and viral proteins are produced and undergo assembly with the viral RNA (step 3). Homogeneous mature virus particles undergo egress through the cellular membrane (step 4). Figure and description were reprinted from Jones et al.[27]

The first step occurs when the HMPV virions bind to their target receptor, which permits their fusion with the host cell membrane, allowing for the viral entry. In the second step, the viral RNA is replicated by the cell's ribosomes, in cytoplasmic replication organelles. This step is different depending how viruses transcribe their genome into mRNA (messenger RNA)[1] (cf. see Figure 7). Next, in the third step, the viral proteins are synthesised and undergo assembly with the viral RNA. Lastly, mature virions undergo egress (i.e. exit, escape) through the cellular membrane, forming homogeneous virus particles. This of course is the simplest representation possible of how we can describe the viral replication life cycle, since factors such as the viral and host heterogeneity[27], their mRNA generation mechanisms [1], or the entry and egress pathways [28], and many more, can make this image much more complex. But since this was not the objective of my thesis, I now focus more on the explaining why not all viruses are the same and how we can differentiate them.

1.1.1.2 Virus structure: they come in different shapes and sizes !

In order for us to properly understand and then investigate virus structure, we first need detail the terminology accepted in the field of virology. Thus we have to talk about 'virions'

when we describe the "infective, extracellular (outside the cell) form of a virus".[23] They can also be referred to as 'virus particles', generally composed of an outer protein shell (named capsid) and an inner core of nucleic acids, i.e. RNA or DNA.[29] They can sometimes also have an additional fatty membrane, which envelopes the genome, or even the capsid, which is named the 'envelope' (see also Figure 5 below). This is created as a result of 'budding', where the newly formed virion collects a portion of the host cell's plasma membrane while pushing out of it.[30] Another structural form in which viruses may present themselves is the one called 'viroids' (i.e. "virus like"), which are "are disease-causing organisms that contain only nucleic acid and have no structural proteins."[23]

It is a generally accepted fact that viruses are very small. Usually they are quoted as being around 20-200 nm in diameter, averaging at about 100 nm.[31] Of course, not all viruses are round and they are not always of the same size. For example, the smallpox virus (Poxviridae family) can approach 400 nm in length, and filoviruses, (e.g. Ebola virus and Marburg virus), have a diameter of around only 80 nm, but extend into long threads of over 1000 nm.[31] A fairly newly discovered class of viruses that infect amoebas, called megavirus, have diameter of around 400 nm, and the pandoraviruses have an elliptical or ovoid structure approaching 1000 nm in length.[31] The following figure (Figure 4) gives a direct size comparison of some known viruses and how they compare in size to prokaryotic (bacterium) and eucaryotic (human cell).

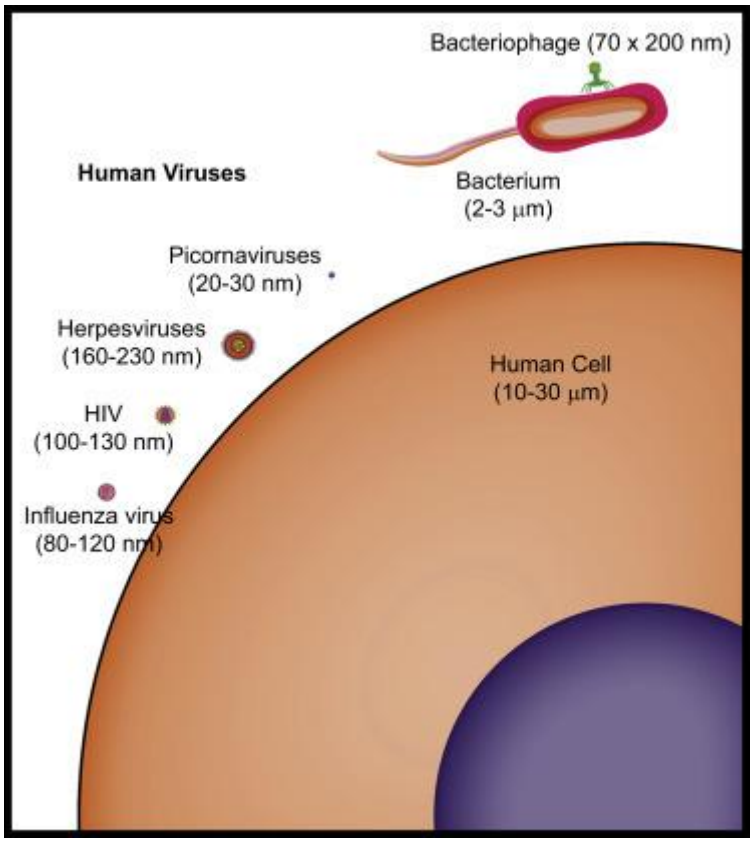


Figure 4. Virus and cell size comparison. Human viruses can vary in size but are generally in the range of 20–200 nm in diameter. In comparison, bacteria are generally 2–3 μM in length, and an average human cell is 10–30 μM. Figure and description were reprinted from Louten.[31]

Now that we have established that viruses have different sizes, let us look a little bit at their molecular conformations as well. The following figure (Figure 5) shows the basic architecture of either a naked encapsidated virus or that of an enveloped virus (lipid bilayer). Encapsidated viruses are viruses whose genome is encased within a capsid.[32]

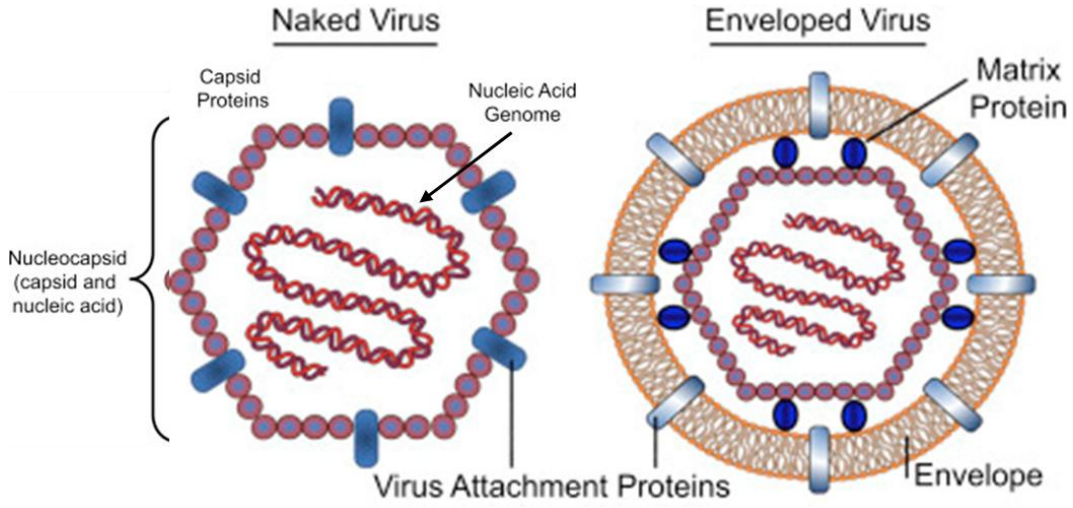


Figure 5. Comparison between a Naked and Enveloped Virion. The naked virus has a very basic architecture. Viral capsid proteins protect the fragile genome, composed of nucleic acid, from the harsh environment. The capsid and nucleic acid together are known as the nucleocapsid. The capsid of an enveloped virion is wrapped with a lipid membrane derived from the cell. Virus attachment proteins located in the capsid or envelope facilitate binding of the virus to its host cell. Figure and description were adapted from Louten.[31]

I would like to make a small parenthesis here, stating that in the rest of this thesis chapter, I decided to rather split them in two slightly different structural groups, that is the encapsidated or the non-encapsidated virions. This separation will be explained in detail in the rest of this chapter, when I will talk about the methodology used to collect the data (cf. section 1.2.2). In the continuation of this work, I will refer only to virions, these being either naturally generated or as artificially generated virus like particles (VLPs).

These different structural conformations are what allow for the diversity of the different shapes that viruses present. As it can be seen from the following figure (Figure 6), if we were to only focus on the viruses that infect humans, we already have a large diversity of shapes and structural conformations.[31]

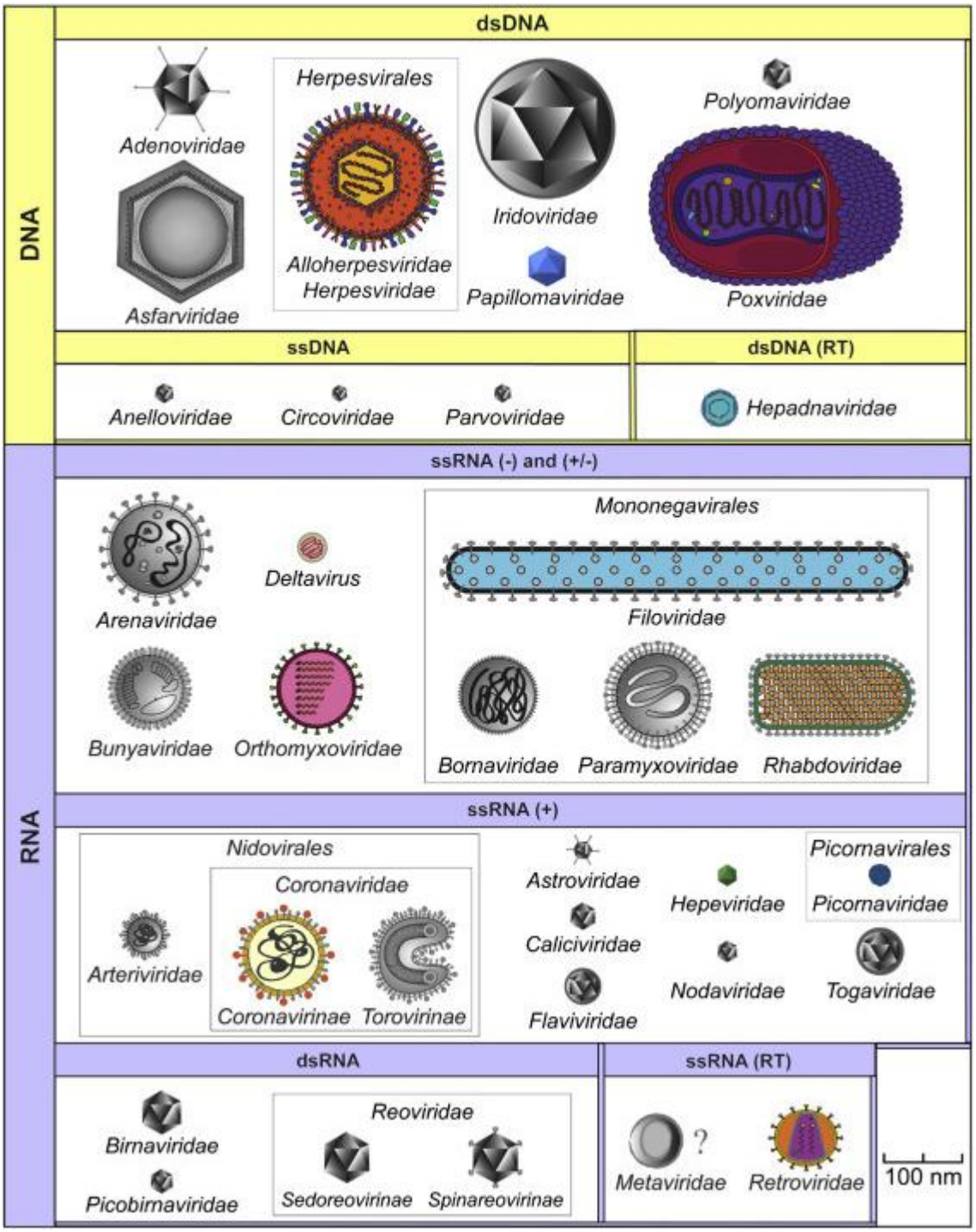


Figure 6. Taxa of viruses that infect vertebrates. Viruses are categorized based upon their type of nucleic acid (DNA viruses in yellow boxes and RNA viruses in blue boxes) and further classified based upon distinguishing characteristics. Note the nucleic acid, size, and architectural differences between viruses of different families. Figure and description were reprinted from Louten.[31]

We can see here that they are generally spherical but can also be bullet shaped (e.g. Rhabdoviridae), cylindrical (e.g. Filoviridae) or brick shaped (e.g. Poxviridae). In addition to that, even the ones that are spherical are not all composed similarly, that is to say that there are viruses that have only the capsid surrounding the genome (e.g. Adenoviruses), they can have only the envelope surrounding the genome (e.g. Arenaviridae), but they can also have both

capsid and envelope (e.g. Retroviridae). In addition, for encapsidated viruses, their capsids usually present an icosahedral (20-faceted) symmetry[5], but it may also be helical (e.g. Filoviridae) or complex (e.g. Arenaviridae). I will detail more about these different conformations in the rest of this chapter, when I will talk about the methodology used to collect the data (cf. section 1.2.2).

Their genome also can be either RNA or DNA and can have different strandedness, i.e. they can be ss (single stranded) or ds (double stranded). Also, they can transcribe their mRNA in the positive sense (5' to 3'), meaning that the mRNA directly translated into viral proteins, or the negative sense (3' to 5'), meaning that a positive-sense mRNA must be produced by an RNA-dependent RNA polymerase from it prior to translation.[33] In addition, some of them can also perform Reverse transcription (RT), meaning that there is a synthesis of DNA from an RNA template, which is driven by RNA-dependent DNA polymerases, also known as reverse transcriptases[34], for example in the case of the infamous HIV-Human Immunodeficiency Virus. These genome structural differences will be discussed when talking about the Baltimore categories (cf. Figure 7).

All these incredible diversity makes them naturally complex and some incredible examples to observe evolution at the nanoscale. But to understand them, we as humans need to have the necessary tools to group and classify them.

1.1.1.3 A brief history on virus classification

I would like to make a small parenthesis here in order to explain the meaning of taxonomy, considered as "the science of categorizing and assigning names (nomenclature) to organisms based on similar characteristics. It is important to note that viruses, since they are not alive, belong to a completely separate system that does not fall under the tree of life."[31] For viruses, it is the International Committee on Taxonomy of Viruses (ICTV) that is charged with the task of developing, refining, and maintains a complete history of virus taxa, back to the release of their first report in 1971.[35]

The ICTV stance had historically been to classify viruses based on "comparison of various characters that describe the virus and can then be used to distinguish one virus from another. Characters can consist of any property or feature of the virus and include the molecular composition of the genome; the structure of the virus capsid and whether or not it is enveloped; the gene expression program used to produce virus proteins; host range; pathogenicity; and sequence similarity."[35] More precisely, they always included examinations based on biochemical test, e.g. susceptibility to inactivation by high temperature, organic solvents, and

low pH, or by more physical evaluations, such as the morphology of the virions and evaluating the type and the quantity of nucleic acids in their genomes.[17]

One of the earliest examples of virus classification, that was the most popular for a long time prior to the genome sequencing revolution, were the so-called Baltimore categories. They were introduced in 1971 by the virologist David Baltimore[1] (Nobel Prize in Physiology or Medicine in 1975), to help the taxonomy of viruses. This approach separates viruses into seven categories, depending on how they transcribe their genome into mRNA, as described below in Figure 7.

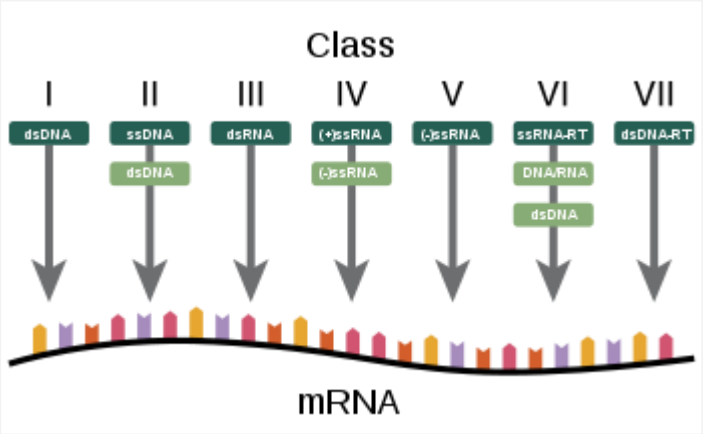


Figure 7. The Baltimore classification in seven categories (by Roman numerals). Legend: (+): positive sense transcription; (-) : negative sense transcription; ds: double-stranded; ss: single-stranded; DNA: Deoxyribonucleic acid; RNA: Ribonucleic acid; RT: Reverse-Transcription. *The figure was reprinted from © Wikipedia.*

This classification separated viruses into 7 categories, numbered by Roman numerals:

I. dsDNA
II. ssDNA
III. dsRNA
IV. positive sense: +ssRNA
V. negative sense: -ssRNA
VI. ssRNA-RT, able to perform Reverse-Transcription
VII. dsDNA-RT

This classification was very important in my work, in order to differentiate the structural composition of viral genomes, grouping them into four clearly distinctive groups, by combing the strandedness (ss/ds) and nucleic acid content (RNA/DNA) information (cf. which will be detailed in section 1.2.2).

The colossal effort put in succeeding the Human Genome Project (1990-2003),[36] had a major impact in the field of virus taxonomy, especially with the development of high-throughput sequencing (HTS) techniques.[17] These last have become increasingly powerful and

prevalent, filling important gaps in our knowledge of virology, and making powerful diagnostic approaches possible.[13] This progress led to reduced interest in methods assessing the physical attributes of viral particles, which came to be considered too technically cumbersome (i.e. long experimental times; expensive equipment; need highly specialised personnel) and less informative than genomics.

Indeed, this appears to have become the main stance of the International Committee on Taxonomy of Viruses (ICTV), which in 2016 made the endorsement of "*taxonomic assignments without prior knowledge of a virus phenotypic properties*".[17] Which was followed by a 2017 consensus statement article from a panel of experts, declaring that: "*viruses that are known only from metagenomic data can, and should be, incorporated into the official classification scheme of the ICTV*".[37] And thus later in 2018, the ICTV stated the following: "*While all characters are important in determining taxonomic relationships, sequence comparisons using both pairwise sequence similarity and phylogenetic relationships have become one of the primary sets of characters used to define and distinguish virus taxa.*"[35]

1.1.1.4 The current consensus view on virus taxonomy

In February of 2023, a consortium of 45 experts in different fields related to virus biology, going from clinicians to bioinformaticians, published a consensus view article called 'Four principles to establish a universal virus taxonomy', by Simmonds et al.[17], summarised as follows:

1. Virus taxonomy should reflect the evolutionary history of viruses.
    - They base this evolutionary history on what is generally known as 'hallmark genes', meaning a set of ancestral orthologous genes, generally of key importance to replication or virion formation.
2. Virus properties may guide assignment of ranks to maximize their utility.
    - The complexity of forming evolutionary ranks and assigning viruses to the classic 15 groups, from realm to species (as seen on the left side of Figure 8), is that these notions are human-made constructs, and viruses do not usually fit perfectly in their 'assigned boxes'.
    - Therefore, additional metagenomic and phenotypic properties are required (bottom of Figure 8), for example information related to the host range, pathological or epidemiological associations, with the only obligation of reaching monophyletic group separation.
3. Taxonomy is but one of many possible means to classify viruses.

- Restricting classification only to evolutionary traits may not be the most optimal solution in all fields relating to viruses, especially regarding clinical or epidemiological properties, e.g. for human immunodeficiency viruses, or more historical and general classifications, e.g. the Baltimore categories (see Figure 7).

4. Taxonomic assignments of viruses inferred from metagenomic sequences require strict sequence quality control.
    - The fact that metagenomic analysis have become much more proficient in producing taxonomic data, nearing two to three orders of magnitude more data than other methods, a strict quality control needs to be applied on the metagenomic data, and guidelines have already been put in place in order to provide the minimum information prior to taxonomic assignment.

Figure 8, reprinted from Simmonds et al.[17], summarises the different ranks needed to achieve virus taxonomy, on the left of the pyramid, and on the right side it shows the different methodologies that are needed to make the necessary assignments and determine virus evolutionary relationships. The relevance of these methodologies is from top to bottom, i.e. from species to realm, starting with nucleotide sequence (NS) similarity, amino acid sequence (AAS) similarity, core genes, and lastly protein fold change and motives. A plethora of bioinformatic approaches have thus been proposed to infer proximity based on phylogenetic trees, and a detailed table of the different tools can be found in the supplementary (Table 13).[17]

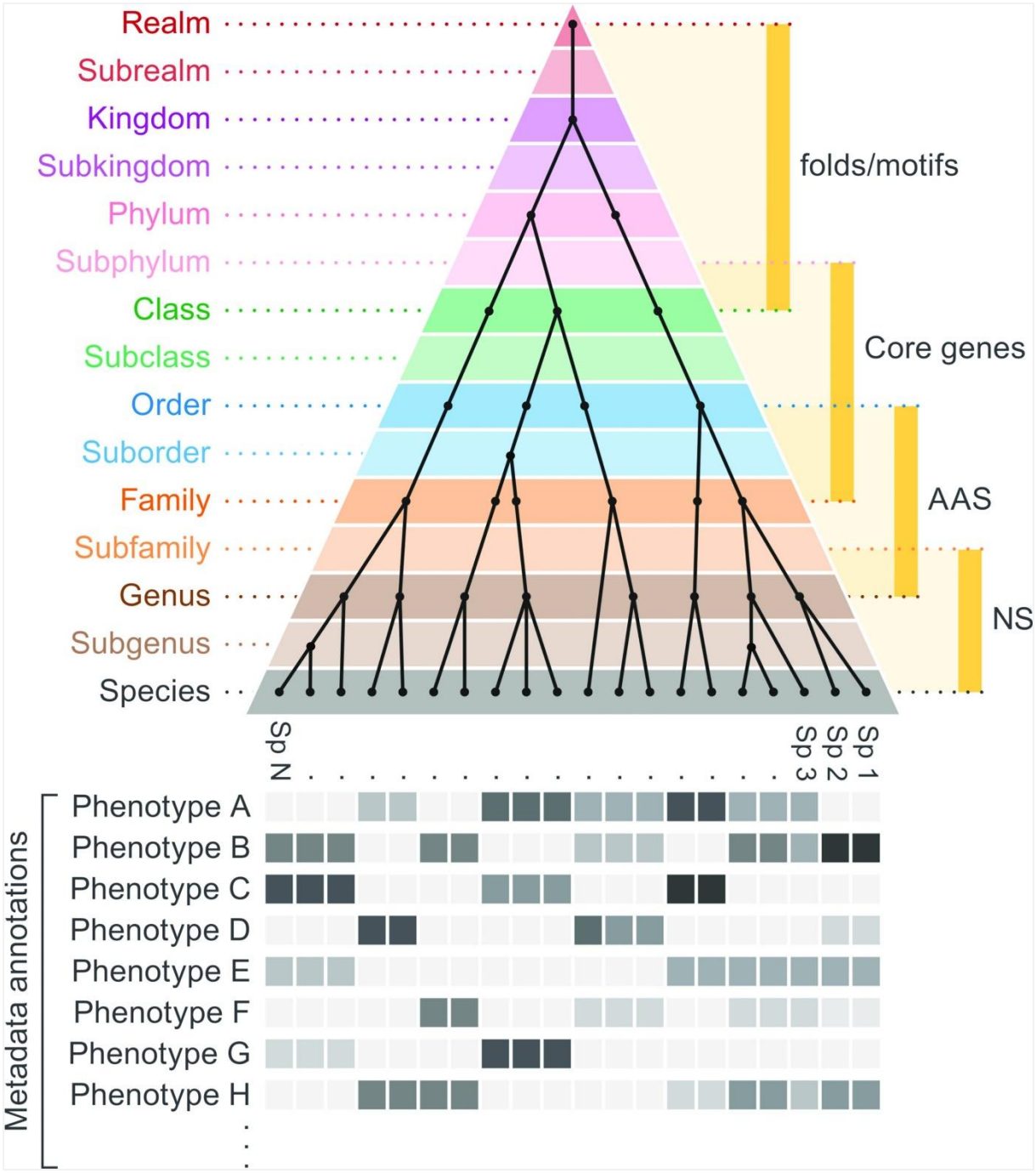


Figure 8. Ranks used in virus taxonomy. Schematic depiction of the 15-rank taxonomic framework used by the ICTV. It includes the methodologies that may be used to determine virus evolutionary relationships and make assignments at each rank. The pyramid shape indicates that the number of taxa increases from the top rank (realm) to the most basal rank (species, Sp.). The names of the 15 ranks are shown on the left of the pyramid, and the methodologies are on the right (AAS, amino acid sequence similarity; NS, nucleotide sequence similarity). The pyramid includes a hypothetical example of the taxonomy of a realm, indicating the number of taxa at each rank (filled circles). The phenotypic properties of classified viruses that may inform rank placements are depicted below the pyramid. The figure and description are taken from the original, Simmonds et al. [17]

For related viruses with very similar genome sequences, *i.e.* going from species to genus or subfamily level, affiliations can readily be inferred by nucleotide alignment. Going further, amino acid sequence alignment allows reaching the genus to order level. At higher taxonomic ranks, however, classification becomes more challenging, particularly for RNA-viruses due to

their high mutation rate.[38] Clustering by sequence similarity of highly conserved core/hallmark genes, allows to family to subphylum range. And to reach the level of kingdom, analysis of optimal protein fold has shown to be a highly evolutionarily conserved attribute of a protein coding sequence, and structural homology.[17]

Interestingly, these approaches often rely on inferences of virion's structural attributes (*e.g.* amino acid sequences, protein folds and motifs), either derived from the available genetic information or even requiring high-resolution structures, especially true the higher the rank and the further we go from the nucleotide sequence. This shows that the structural analysis of the protein content cannot be neglected even after knowing the genome of viruses, thus making phenotypic evaluation a necessity to a certain degree.

Now that we know the life cycle of a virus, how it can structurally present itself and how scientist have tried to classify them, let us now take a step back and come to the fundamentals that relate biophysical properties of viruses and their structure, in order to see the bigger picture of viruses as biological macromolecules.

1.1.2 Viruses as Macromolecules

In 2013, Božič *et al.* described viruses outside their host as "*very complicated macromolecules with a life cycle*".[39] Well before this epithet was coined, in 1998, Horace M. Mazzone compiled the "Handbook of viruses: Mass-molecular weight values and related properties."[4] This monograph brought together prior scientific knowledge on virus mass measurements, fostering the concept of viruses as macromolecules. This concept arises from considerations of viruses' essential components (*i.e.* oligonucleotides, proteins, and lipids), and their thermodynamically controlled self-assembly during replication.[40] This process is particularly evident for encapsidated viruses, where individual viruses could be considered as nucleoprotein complexes with equivalent or quasi-equivalent molecular composition.

Now let us see how mass analysing has revealed itself to be a powerful tool in the structural characterisation of biological entities.

1.1.2.1 Mass as a specific property of biological macromolecules

Mass is a specific property of matter at the atomic and molecular levels, and a molecule's elemental composition can often be deduced from accurate masses and isotopic patterns obtained from Mass Spectrometry (MS) measurements.[41] MS is therefore extensively used to characterize the molecular structures of biomolecules based on prior knowledge of their subunits (amino acids, nucleotides, glycans).[42] This is made possible thanks to the limited

assortment of structures that can actually be synthesized by a biological entity, restricting the theoretical combinatorial complexity of biopolymers related to their sheer size and the diversity of their subunits. In MS-based proteomics, for instance, using mass to identify a peptide amongst all possible combinations, of the 20 naturally occurring amino acids, quickly becomes intractable as sequence length exceeds a few subunits.[43] Yet, by breaking down the peptide of interest into smaller fragments, and taking into account the genome sequence of the organism under study, hypothetical amino acid sequences can be predicted, and accurate mass information used to identify a peptide among the ones potentially produced by the organism.[44]

The discriminatory power of mass further extends to intact proteins as their composition reflect their function. Counterintuitively, this becomes easier as mass increases as intact proteins masses are more specific than those of medium-sized peptides, since they are less common, and they spread over a broader mass range. Veenstra *et al.* thus showed that accurate mass information coupled with single amino acid counts allowed unambiguous protein identification amongst all predicted *E.coli* proteins.[45] Similarly, Lloyd Smith's group showed how to elucidate yeast proteoform families using intact-mass and Lysine-count measurements.[46]

However, at the macroscopic scale, mass becomes largely irrelevant for compound identification, as it reflects the total quantity of matter present in the sample studied, and not its molecular composition.[47] Nonetheless, researchers have used MS and related technologies to determine the stoichiometry and binding affinity of protein complexes up to several millions of atomic mass units or daltons (Da).[18] Therefore, I looked at the total virion mass in order to evaluate if this property of mass remains intrinsic at the 'mesoscale', that is in the range 10^{-8} to 10^{-5} m (10 nm to 10 μm)[48], which is also where viruses lie. A more schematic explanation of the 'mesoscale' can be seen in supplementary (Figure 125).

Reports by Matthews (1975)[49] and Bahr (1980)[50], gave indications that the molecular mass of specific viruses could be a distinctive phenotypic attribute with relevance to their biology. However, the utility of mass as a classification criterion was somewhat moot at a time when it could only be indirectly determined by cumbersome methods that required significant amounts of highly pure sample,[51] such as crystallography, sedimentation, and a variety of light-scattering approaches.[4]

Consequently, it remains to be seen whether the total virion mass could constitute, in itself, a truly distinctive criterion for individual viruses.

1.2 The experimental design

To perform our evaluation, I examined the data available in the literature from two main perspectives: biophysical and macromolecular. I first sought information on physical characteristics, such as size and density, and compared the findings to experimental total mass measurements for virions (left panel of Figure 9). The second sought details was the sum of all the structural components of viruses, by performing two different approaches, depending on whether the virus was encapsidated ('capsid-derived') or not-encapsidated ('genome contribution'), as summarised in the right panel of Figure 9.

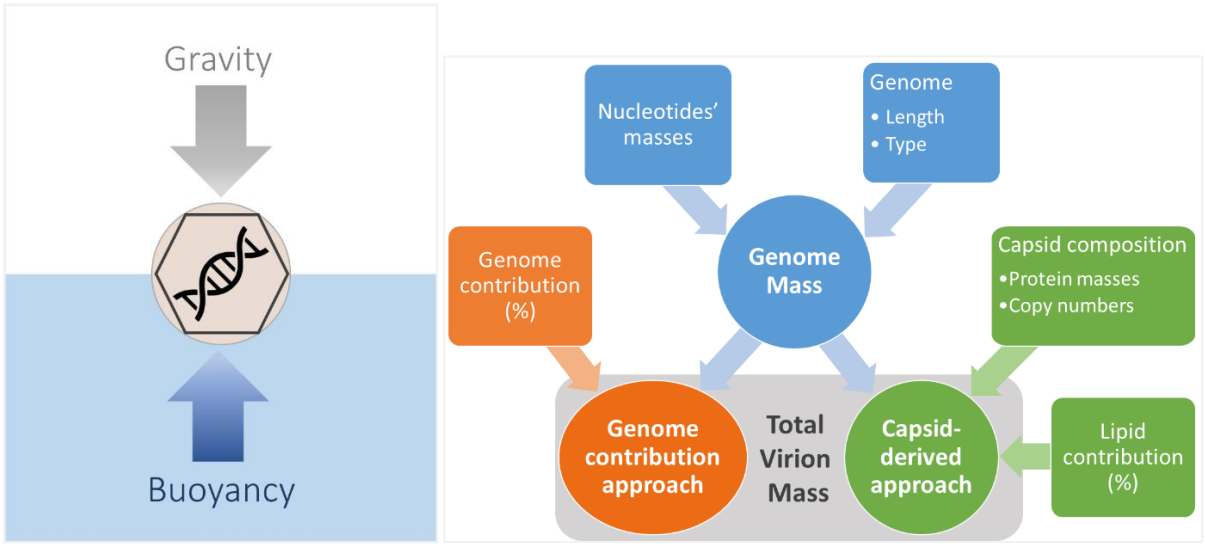


Figure 9. Summary of the two approaches used to calculate total virion mass. Left (Biophysical approach): The balance between gravity and buoyancy for a spherical body (the hexagon, with its DNA double helix, represents a generic virus). Equilibrium is reached when the body, with a given mass, has the same density as the fluid (here in blue). Right (Macromolecular approach): Starting from the genome mass, which can be readily computed for all HuReVs (in blue), we used a capsid-derived approach for encapsidated virions (in green), and a genome-contribution approach for viruses without a capsid (in orange), to determine the total virion mass (in grey). (See SI #1 for more details on Methods)

1.2.1 Proof of Concept - Focus on human respiratory viruses (HuReVs)

The sheer number of viruses on Earth[52] would make any attempt to investigate all their masses intractable. I therefore decided to focus my attention on what I considered the most relevant subgroup: viruses affecting human health.[53] Spurred by the recent pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2),[54] I specifically chose to focus our attention on viruses causing human respiratory diseases, as seen in Figure 10.[55] In this sunburst diagram, I presented the collective plethora of human viruses, separated and classified by their Baltimore category[1], viral Family and Genus, exocentrically ordered. Baltimore classification separates viruses into seven categories, depending on how they

transcribe their genome into mRNA (messenger RNA). For more details see supplementary (Figure 7).

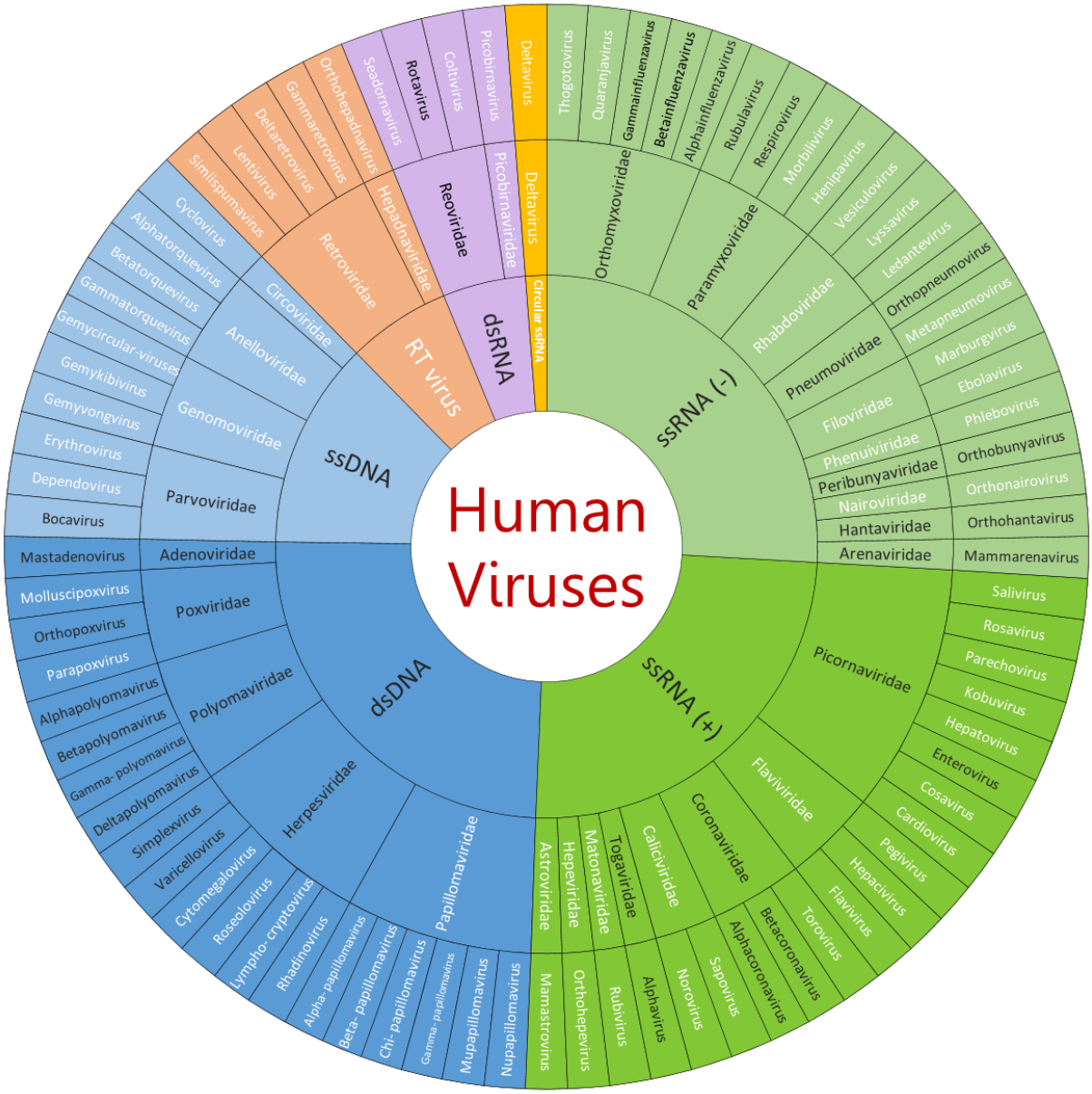


Figure 10. Human viruses and human respiratory viruses (HuReVs), in black letters. Sunburst diagram of viruses infecting Humans, ordered exocentrically by Baltimore category[1], Family and Genus. It was formed by collecting the information found in ViralZone. [link] The information on the respiratory viruses is a combination of the information from ViralZone [link] and the article of Geoghegan et al. [53]

The epidemiological relevance of human respiratory viruses (hereafter HuReVs) is supported by several elements: they are the most frequent causative agents of disease in humans,[56] and have been the target of nearly three quarters of all approved vaccines for humans in the last 20 years.[57] In addition to the prevalence of these viruses, the high number of vaccines developed is linked to their high rate of mutation, which causes vaccines to progressively lose efficacy.[58] From an evolutionary perspective, respiratory viruses cover a significant cross-section of viral families known to cause human disease, as they account for nearly half of all

known vertebrate viral families.[59] Hence, I believe that the conclusions from this study, regarding the analytical relevance of virus mass, could be usefully extrapolated beyond the context of HuReVs.

1.2.2 Virus data

I collected the taxonomic data, genome information, structural attributes as well as pathological features, from the list of 203 human viral pathogens compiled by Geoghegan et al.[53] I then filtered the list by transmission route containing the term "respiratory", to retain a subset of 39 human respiratory viruses (HuReVs). I updated this filtered list with viruses discovered since, using the ViralZone database.[60] Four additional, more recent viruses were thus included, yielding a full list of 43 HuReVs (combined information in supplementary Table 15 and Table 16). Minor discrepancies were identified and corrected by comparison with the NCBI viral genomes database.[61]

Biophysical data: I compiled the experimental data for virus density and volume from various sources,[4,49,50,62–64] and used them to estimate a mean mass for each HuReV family. I also collected actual mass measurements from literature, when available.[4,49,62,64,65] (see supplementary Table 14). I was unable to find experimental mass values for the family *Arenaviridae*, and I found it difficult to provide an estimate on their mass, due to their well-known pleomorphism.[62]

Capsid data: For encapsidated viruses, their Triangulation number (T-number)[5], related to protein copy number, I primarily obtained them from the work of Baker, Olson & Fuller[66], and cross-validated with ViralZone[60]. I combined this information to the list of structural proteins, their sequences and masses, obtained from the Uniprot database[67], which I manually verified. Cleaved proteins were accounted for, and the final protein mass was calculated using the tool Peptide Mass.[68] I collected the data for lipid content (%) based on various references, including BioNumbers (ID-106100).[2,49,64,69] (see supplementary Table 15)

Genome information: I derived the oligonucleotides molecular masses as a function of genome length from the BioNumbers (ID-110193; ID-110968) database.[2] I obtained the empirical data regarding the genome proportions from various literature resources, including Bionumbers (ID-106100).[2,64] To understand the methods that were used to get to these genome content values, I found the work from Green and Piña (1963,1964) to be pioneering this biochemical methodologies.[51,70] *Coronaviridae* were the only HuReVs for which I was unable to find empirical genome contribution values in the literature. (see supplementary Table 16)

1.2.3 Statistical methods

All mass data are expressed in Megadaltons (MDa) [= 10^6 Da], where 1 Da = $1.66054 \cdot 10^{-27}$ kg. Due to the order of magnitudes differences in mass between the various species, the mass axis were converted into their logarithmic base-10 values, denoted as Log Mass. I performed unsupervised hierarchical clustering on non-standardised Log Mass values using the statistical software SAS-JMP (version 17.0.0). More specifically, the agglomerative clustering Ward method,[71] was used to calculate the minimum variance criterion, thereby minimising the total within-cluster variance, and yielding hierarchical groupings. To facilitate interpretation, the number of clusters represented was arbitrarily selected to correspond to the number of HuReV families considered. Dendrograms with geometric spacing were used to represent distance relationships between individual viruses.

1.3 Biophysical perspective

1.3.1 Methods: The biophysical perspective

Mass (m) can be defined as the product of density (ρ) to Volume (Vol)

(1)

$$m = \rho * Vol$$

Assuming perfectly spherical geometry for all HuReVs, the relationship between diameter (d), buoyant density (ρ_b), and the molecular mass of the total virion (M_r^{TV}) can be expressed as:

(2)

$$M_r^{TV} = \rho_b * \frac{\pi * d^3}{6}$$

Most HuReVs present capsids with an icosahedral structure (*i.e.* with 20 facets) [5]. The icosahedron is the "roundest" of all the platonic solids, featuring the greatest volume relative to its surface area. It is therefore widely accepted that these structures can be approximated as spheres.[72,73] The exception were the family *Poxviridae*, considered as 'brick shaped'. Thus, to compute their volume, I used Vol = length * height * width, and replaced the computed Vol value in equation (1), to get the final M_r^{TV}. Similar approaches in calculating the virion volume have been used in prior works.[72,73] In addition, I considered these spheres to have a homogeneous density, which I will detail in the following section.

1.3.1.1 The virus window, of density and sedimentation

When seeking to define viral mass, a primary consideration stems from its simple physical make-up, with two properties that are readily accessible from straightforward experimental observations: density and size. As early as 1966, N.G. Anderson *et al.*[74] observed that viruses' sedimentation rates and buoyant densities fell within a range that they termed the "virus window", as seen in the in Figure 11.

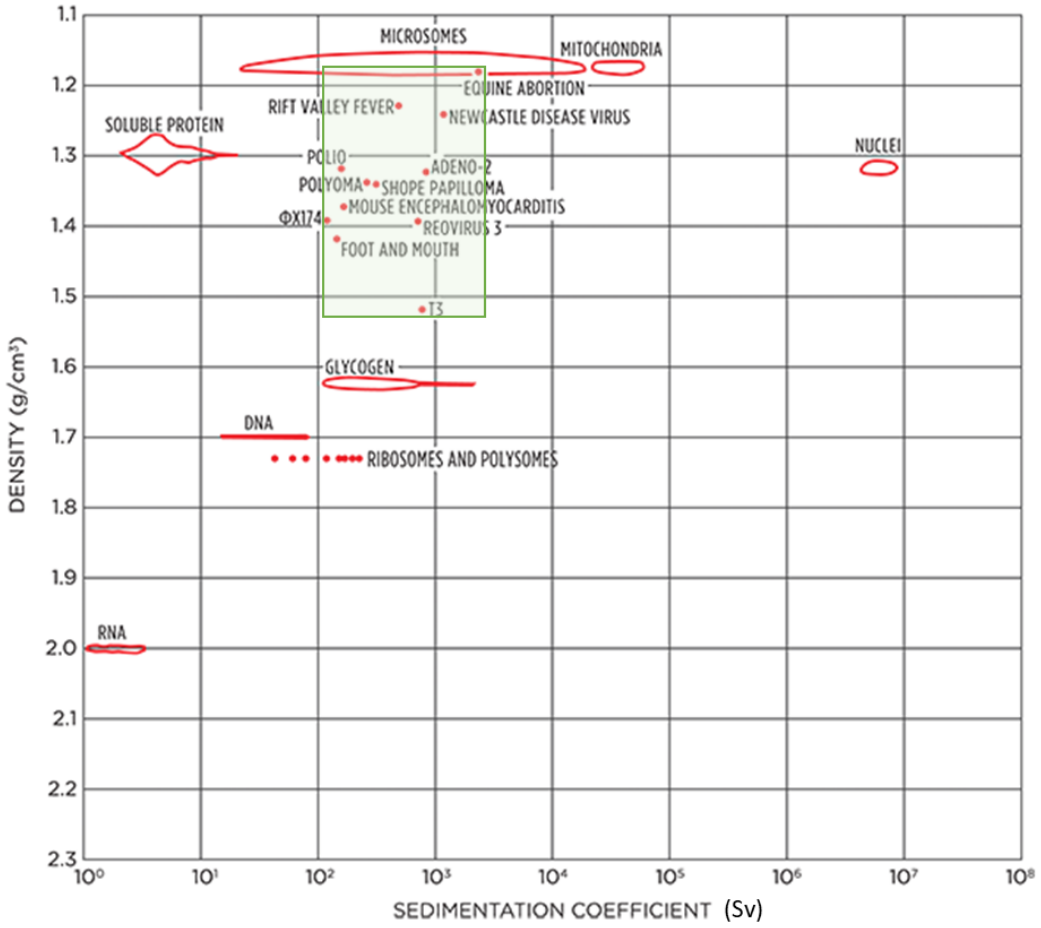


Figure 11. The density virus window. Description from the original work, Anderson et al. (1966) [74], stated: Sedimentation coefficients (Svedberg units) and banding densities (g/cm³) of cell components and viruses. Sources of data are given in the original text. Note that most virus particles fall in an area, here termed the "virus window," essentially free of cell constituents. The graph was adapted from image taken here [link]. The green rectangle was indicated to highlight the region of the virus window, broadly based on the suggestions from Mazzone[4].

This window was largely free of host cell constituents and thus allowed efficient virus purification. Subsequently, Mazzone broadly outlined the limit values of this "virus window", with densities between 1.2 and 1.6 g/cm^3, and sedimentation coefficient values between 100 and 2000 Sv (Svedberg).[4] A highlight of this area (in green) can be seen in Figure 11.

The graph in Figure 12 shows that this principle of virus 'density window' is consistent with my data.

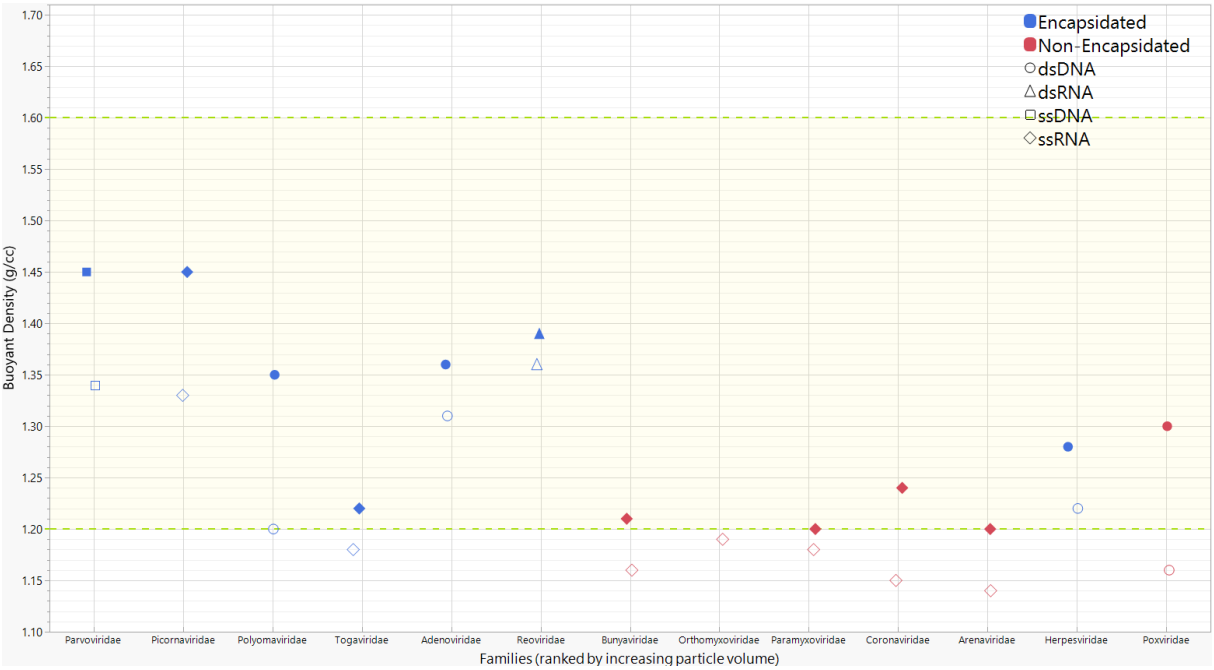


Figure 12. Figure describing the relationship between the buoyant densities (in g/cc: gram/cubic centimetre) as a function of the HuReV families, ranked in increasing particle volume. Legend (top right): Encapsidation status: blue- encapsidated, red- non-encapsidated. Filled: maximum; empty: minimum, density value per family. Genome type: double-stranded DNA [dsDNA] (circle), single-stranded DNA [ssDNA] (square), dsRNA (triangle), ssRNA (rhomboid). The dashed line at 1.2 g/cc and 1.6 g/cc is indicated, referring to the minimum and maximum density value, highlighting the range in yellow, as reported by Mazzone.[4]

In Figure 12, we can observe that the values of density for the viruses analysed by me are found between 1.13-1.45 g/cm^3 (see supplementary Table 14). Thus, the lower boundary in Figure 11 is slightly overestimated to my data, but the density range is small enough for it to be considered as homogeneous. Another interesting remarks that I want to point out on this graph, is that the non-encapsidated viruses (in red) are more often on the lower range of the graph (*i.e.* < 1.3 g/cm^3) and have higher volumes. From the encapsidated viruses (in blue) the two families that present a similar 'low density' are families that present both a capsid and an envelope, *i.e.* *Herpesviridae* and *Togaviridae*. Nevertheless, this does not seem to be compensatory enough to have overlapping families. (see later in Figure 13).

1.3.1.2 Building the "mass ribbon"

Most viruses, including HuReVs, are found in the size range, 20 to 300 nm in diameter.[31] Interestingly, particles with diameters between these two extremes, whatever their composition, tend to remain in the low and middle troposphere for long time periods, meaning a day to a week.[75] One can argue that this is perhaps a result of evolutionary physical constraints, and further experimental work can reveal if this hypothesis holds true in the future. A more detail analysis on this observation can be found in annex, in '

Describing HuReVs as aerosolised particles' (section 7.1.3).

So, we can use this size value, 20 - 300 nm, to reach a particle volume, in order to calculate a particle mass, as I did in Figure 13. There I presented the expected mass of viral particles, in MDa, as a function of particle volume, in nm^3. To assess how well the biophysical model matches actual measurements, I searched the literature for independent estimates of the size, density and mass of HuReV families. From this data, I calculated a 'biophysical' mass, using the average size and density values for each respective family (green symbols) and compared them to reported mass values (blue symbols), as shown in Figure 13. This comparison was possible for most HuReV families.

However, I was unable to find experimental mass values for the *Arenaviridae* family, and their pleomorphism[62] made it difficult to provide an estimate. In addition, I only found a single reported size measurement for *Paramyxoviridae*, and unique mass values for *Togaviridae, Orthomyxoviridae,* and *Coronaviridae,* precluding the definition of size (respectively mass) ranges for these families.

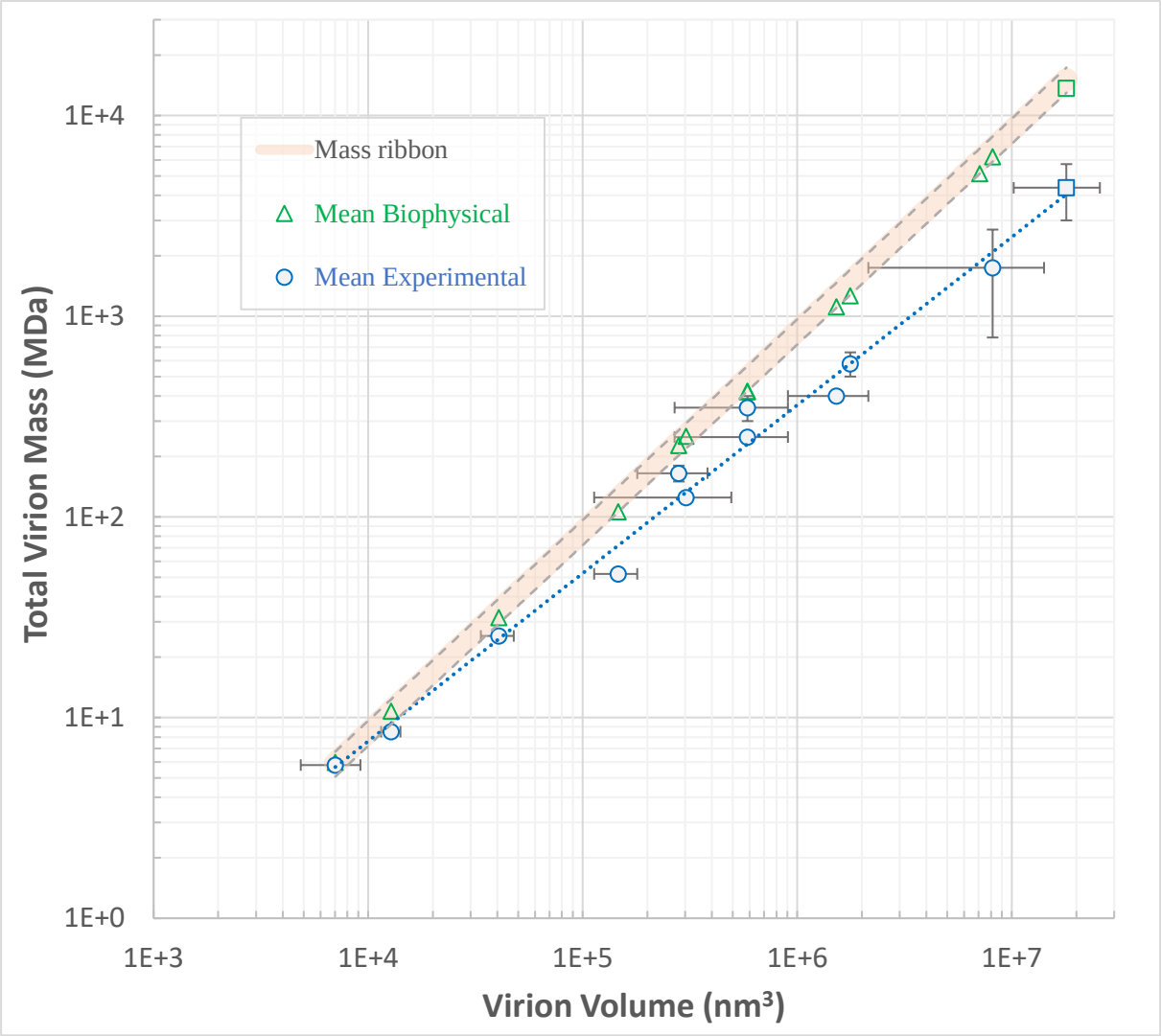


Figure 13. The virus "mass ribbon" and the hypoallometric relationship between average mass and size for virus families. The theoretical 'mass ribbon' values cover the density range [1.2-1.6] g/cm3, as outlined by Mazzone[4] (highlighted in orange). Mean biophysical masses (green symbols) were calculated using size and density data reported in the literature. Mean experimental reported mass values for each family (blue symbols) were extracted from the literature; error bars denote reported range for mass (vertical) and size (horizontal); blue dotted line: power law trend fit to mean values (y = 0.0034 x0.8375; R^2 = 0.9891). Poxviridae, which do not have a spherical geometry, are denoted by squares. Complete data are available in supplementary Table 14.

From this analysis, a first observation is that the range of experimental masses reported for HuReV families also spans three orders of magnitude, from 5.5 to 5721 MDa. However, there are significant differences between the masses expected from the biophysical parameters (green symbols) and the experimental derivations (blue symbols). In fact, the biophysical approach appears to overestimate the mass by an average of ~16.25 %, based on the exponent of the power law fit ($y = 0.0034\ x^{0.8375}$). In addition, experimental masses for virus families of similar size, e.g. *Reoviridae* and *Adenoviridae* (Vol ~$3*10^5$ nm^3); and *Orthomyxoviridae* and *Bunyaviridae* (Vol ~$6*10^5$ nm^3), did not coincide as predicted by the biophysical model.

The ‘density window’ consideration yields a “mass ribbon” (highlighted in orange), which spans a ~3.5 orders of magnitude range ~3 to ~7900 MDa. In comparison, size, often given as particle diameter, only covers a single order of magnitude: 20 to 300 nm. As densities reported for HuReVs appear to be consistent with the ‘density window’ for viruses (see Figure 11), the discrepancies between the computed and experimental masses appear linked to errors in volume estimations from size measurements (see x-error bars in Figure 13), or to the underlying spherical shape approximations. Another possible explanation relates to the necessity of drying the sample to perform mass analysis by electron microscopy based methods.[76,77] However, such errors cannot satisfactorily explain the observed differences.

Importantly, the virion's mass captures its three-dimensional structure and its heterogeneous composition better than a single dimensional parameter or average density values. As a result, all other things being equal, a methodology evaluating viral mass should be more sensitive to structural changes than approaches relying purely on dimensions or mean density. For a fixed number of families, the wider the range of parameters observed, the smaller the number of families per parameter unit. This improves family separation and thus their classification.

These initial observations led us to the conclusion that, although the mass of HuReVs could be a potentially useful classification criterion, a straightforward biophysical perspective did not provide adequate information to test this hypothesis. I therefore delved deeper into the mass of individual HuReVs from a macromolecular perspective.

1.4 Macromolecular perspective

One can image a virus as a ‘nucleoprotein complexes’.[4] This suggests an alternative approach to estimating the total mass of a virion: summing the masses of its molecular constituents, namely the viral genome, its constitutive proteins, and any additional lipid components in its envelope. To compute the mass of HuReVs from a macromolecular perspective, I therefore considered several pieces of information and estimated the relative contributions of each component; finally combining those to determine a total virion mass (see Figure 9).

1.4.1 Method: Genome mass calculation

A genome mass is readily accessible for all viruses knowing their genome type and length in nucleotides (NA: Nucleic Acids). I used the following formula, based on Bionumbers database[2] (ID: 110193), to give an approximate average of the mass:

(3)

$$m_{gen} = (\, m_{NA} \times n_{NA}) + m_{end}$$

With m_{NA} the average mass of single nucleotides or nucleotide pairs for a given genome type, n_{NA} the genome length in nucleotides or nucleotide pairs, and m_{end} the mass of end groups.

1.4.2 Genome mass calculation

The genome mass is the most readily accessible contribution to virion total mass. To compute genome masses, I used the average mass of individual nucleotides or nucleotide pairs for each genome type, the genome length in bases or base pairs (as reported by Geoghegan et al.[53]) and the masses of end groups. Importantly, for viruses featuring segmented genomes (e.g. *Hantavirus*), the sum of all the segments was calculated. Figure 14 (right panel) displays the logarithm of calculated genome mass values (in MDa) of individual HuReV species, ranked in decreasing order, and labelled depending on their genome composition (DNA/RNA) and structure (single or double-stranded).

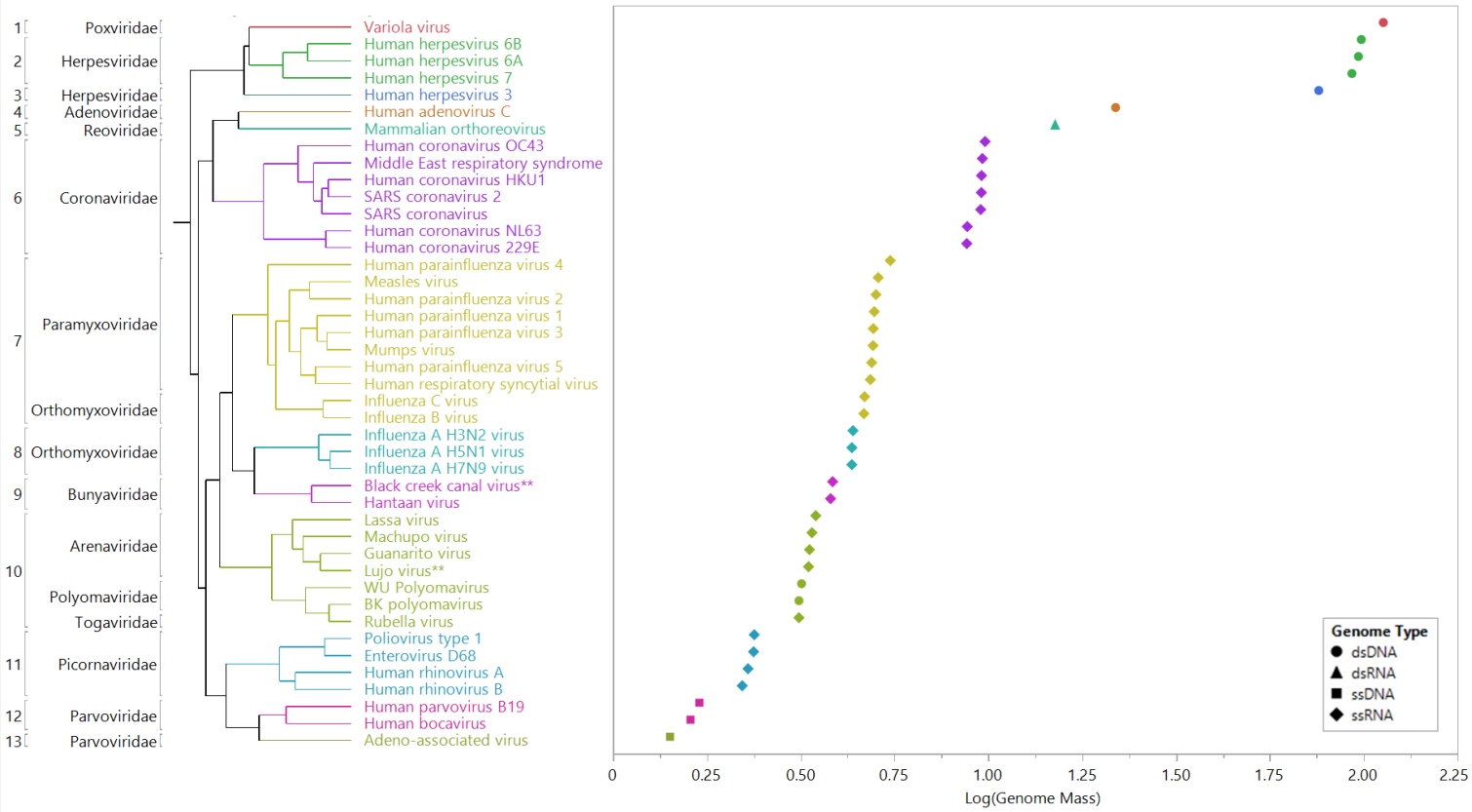


*Figure 14. Repartition of HuReV genome masses. Left: The respective viral families are matched to their corresponding cluster number (n=13). Middle: Dendrogram (geometric spacing) resulting from the unsupervised hierarchical clustering of indidivual viruses based on their log(Genome Mass [MDa]). Right: Individual virus log(Genome Masses [MDa]) ranked in decreasing order and colour-coded by cluster. ** indicate rare viral species, as reported by Geoghegan et al.[53]. The data in this plot are highlighted in blue in the supplementary tables (Table 15 and Table 16).*

Computed HuReV genome masses were observed to spread over a range spanning slightly more than two orders of magnitude (from a few MDa to slightly over a hundred MDa). Unsurprisingly, single-stranded genomes tend to have lower masses than their double-stranded counterparts, apart from *Polyomaviridae*. In addition, whereas RNA viruses group together in an intermediate mass range (~ 2-20 MDa), DNA viruses tend to occupy the extremes. Interestingly, the overall distribution appears discontinuous, suggesting that genome mass may somehow act as a distinctive characteristic. To investigate this possibility, I performed unsupervised hierarchical clustering, the results of which are presented in the dendrogram and by the colour-coding of the different clusters. I arbitrarily defined the number of clusters to correspond to the number of HuReV families considered (n=13) and assessed how well the clusters mapped to the original HuReV families. A perfect match was found for 6 out of 13 HuReV families (*Poxviridae*; *Adenoviridae*; *Reoviridae;* *Coronaviridae*; *Bunyaviridae*; *Picornaviridae*). Other families were either incorrectly split between multiple clusters {*Herpesviridae* [#2, 3], *Orthomyxoviridae* [#7, 8], *Parvoviridae* [#12, 13]} or wrongly assigned

to the same cluster {*Paramyxoviridae* & *Orthomyxoviridae* [#7]; *Arenaviridae*, *Polyomaviridae* & *Togaviridae* [#10]}.

This apparent limited capacity to discriminate based on genome mass could explain the fact that, even though genome mass is readily accessible from sources such as the NCBI Viral Genomes database[61], no classification based solely on this characteristic has been proposed. In addition, the genome masses only span a fraction of the ~3 orders of magnitude range obtained for experimental total masses (see Figure 13). This appears logical since viruses are not simply made of genomes, and their 'genome containers', *i.e.* their capsid shells or envelopes, must also be accounted for.

1.4.2.1 Adding the mass of the 'genome container'

Virion structures consist of a genome within a 'genome container', and the two entities are obviously related. In fact, several reports present allometric relationships between the genome and its container. Thus, Matthews et al.[49] presented experimental evidence of the relationship between a viral particle's dry mass and the dry mass of its nucleic acid constituents. More recently, Cui et al.[73] derived a relationship between the genome length and the particle volume, for all types of viruses. This relationship is independent of genome type, 'genome container', or host. In addition, Božič *et al.*[39] showed that the triangulation number (T-number) is proportional to the square root of the capsid radius, which led Chaudhari et al.[72] to conclude that the genome length (L) correlated with the radius of the capsid (R), according to the relationship:

(4)

$$L = kR^{\alpha}$$

where k is a proportionality factor and $\alpha \in [1, 2[$ is an allometric exponent. The respective values of the two terms depend on the nature of the capsid components.

Taking into account that we are using the relationship of L to the volume, thus R^3, we have to find a relationship that is $\delta = 3/\alpha$, and thus has an allometric exponent $\delta \in [1.5, 3[$. So we checked if this expected relationship between the virion volume and the genome length, presented in number of base pairs (bp), remained true for HuReVs, as seen in Figure 15.

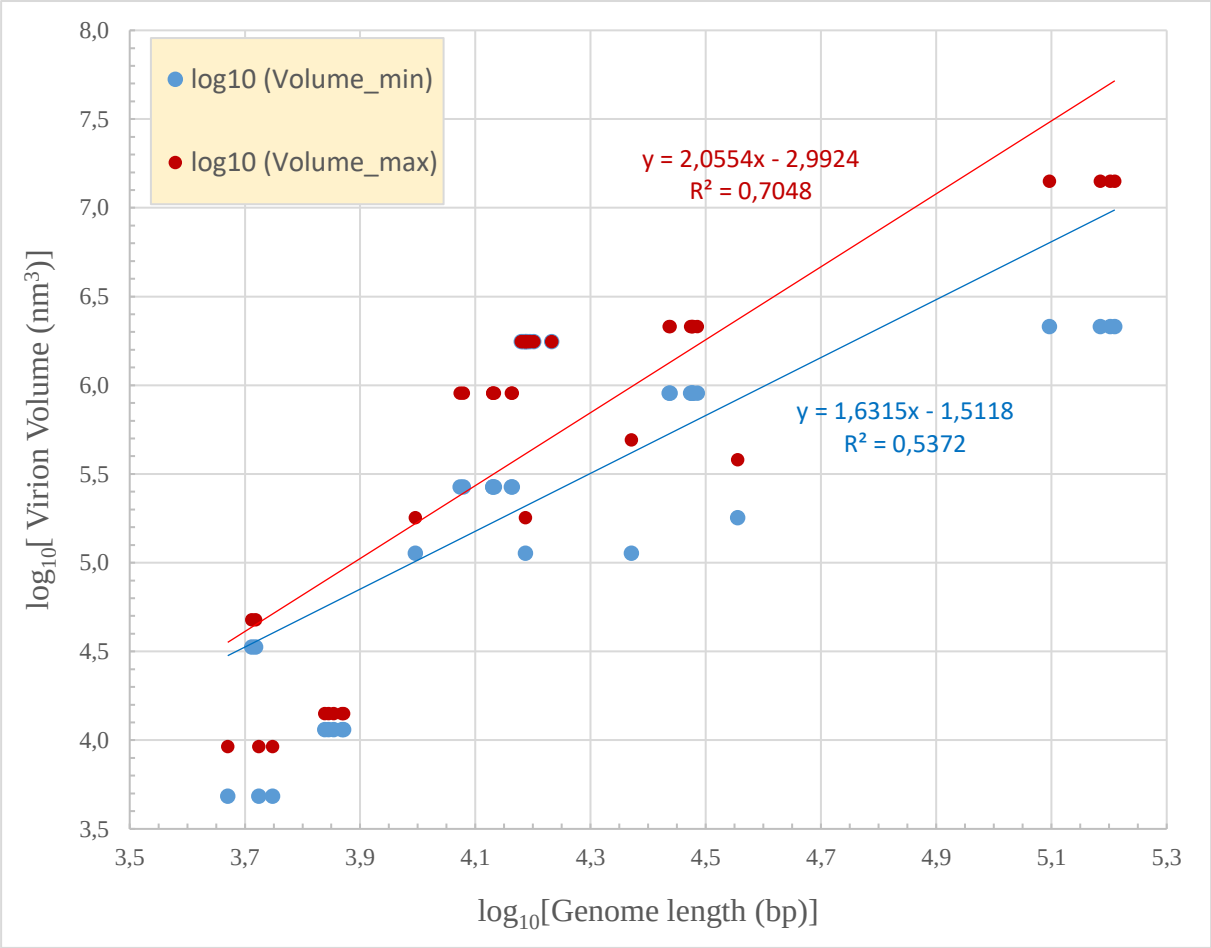


Figure 15. The allometric relationship between the virion volume, in nm^3, versus the genome length, in number of base pairs (bp). The power law trend lines are drawn for the maximum volume (red) and the minimum volume (blue), with their respective equation and R^2 values. The legend is found on the top left corner of the graph.

In Figure 15 we see that the allometric exponent is indeed higher than 1.5, for both the maximum virion volume (red) and the minimum virion volume (blue), at 2.0554 and 1.6315, respectively. Although the reliability of the fitted power law trend line is better for the maxima ($R^2 = 0.7048$) then for the minima ($R^2 = 0.5372$), the overall tendency conforms our hypothesis, and is consistent with the model by Chaudhari et al.[72]

Next, I investigated whether the addition of the mass of the 'genome container' allowed better inter-family separation.

1.4.2.1.1 Proposed model including the genome container

We will consider two viruses, virus 1 and virus 2, and their associated radii (R_1 and R_2), where $R_2 > R_1$. We can estimate several mass differences (Δm) between them using a) the mass of their genomes alone, or b) the mass of the total virion, *i.e.* the sum of the masses of their genomes and their 'genome containers'.

If only genome masses are taken into account, the difference in genome mass (Δm_g) is written as:

(5)

$$\Delta m_g = \rho_L k (R_2^\alpha - R_1^\alpha)$$

where ρ_L is the linear genome density.

Conversely, if we consider the total virion mass, the difference in mass of the total virion (Δm_{TV}) can be written:

(6)

$$\Delta m_{TV} := \rho_L k (R_2^\alpha - R_1^\alpha) + \rho_S 4\pi(R_2^2 - R_1^2)$$

where ρ_S is the surface density of the virion, assumed to be a constant due to the constant capsid thickness determined by Božič et al.[39]

From the relations, seen in equation (5) and (6), we can see that $\Delta m_{TV} > \Delta m_g$ is always true. Moreover, it can be seen that it is independent of the alpha value. Therefore, taking into account capsid masses would increase the mass differences, yielding a potential structural signature allowing virus differentiation.

1.4.3 Method: Calculating the mass of the 'genome container'

Once the genome mass was computed, I set out to estimate the mass of the 'genome containers'. Due to important differences in their structures and compositions, I treated encapsidated and non-encapsidated viruses separately.

1.4.3.1 'Capsid-derived' mass calculation approach

For encapsidated viruses (7/13 families) [31], I derived the total virion mass by adding the genome mass to the calculated mass of the capsid. To compute the capsid mass, I relied on their preserved icosahedral geometry, their T-number, and the mass of their structural proteins. The protein copy count is linked to the T-number, which indicates the number of structural units per face of the icosahedron[66]. A more detailed definition of the T-number is describe in the supplementary section 'Defining the Triangulation (T-) number'. These structural units are composed of defined numbers of Major Capsid Proteins (MCP) and minor Capsid Proteins (mCPs) per facet. Since capsids may have one, or more mCPs, I expressed the capsid mass with the following formula:

(7)

$$m_{cap} = \sum_{i=1}^{N} (m_i \cdot n_i)$$

Where (m_i) = mass of structural protein *i*; (n_i) = copy number of protein *i*, based on the value of the T-number; [$i=1,2,...N$]. This approach has been widely used to obtain theoretical values for comparison with experimental MS data.[21,78]

For viruses featuring an additional lipid component (e.g. *Herpesviridae, Togaviridae)*, I estimated the added mass from the experimentally determined ratio of the lipid content on the virus total mass, denoted as $r_{lipids,}$ *i.e.* lipid ratio (see supplementary Table 15). I thus computed the total virion mass, M_r^{TV}, using the following formula:

(8)

$$M_r^{TV} = \frac{m_{gen} + m_{cap}}{1 - r_{lipids}}$$

I named this approach the 'capsid-derived' method (Figure 9). A detailed diagram of this calculation can be seen in supplementary (Figure 126).

1.4.3.2 'Genome contribution' mass calculation approach

However, 6/13 HuReV families do not possess a capsid, and like many enveloped viruses, their molecular composition is often not precisely known. But encapsidated viruses represent only about 60 % of the total number of virion structures.[31] Contrary to encapsidated viruses, the actual molecular composition of enveloped viruses cannot be easily derived from the copy numbers of their structural protein, since they are not precisely known. To estimate their total mass, M_r^{TV}, I relied instead on the computed genome mass (see section 1.4.1 and Figure 14) and literature reported values of their genome dry weight contribution to the total virion mass, denoted as r_{genome}, *i.e.* genome ratio (see supplementary Table 16). This parameter has historically been an important criterion when describing virions in physicochemical terms, at least up until the 6th report of the ICTV (1995).[69,2] I expressed this approach by the following formula:

(9)

$$M_r^{TV} = \frac{m_{gen}}{r_{genome}}$$

I termed this approach the 'genome contribution' method (Figure 9). A detailed diagram of this calculation is represented in supplementary (Figure 126).

Unexpectedly, despite tremendous recent research efforts, I was unable to find genome contribution data for the *Coronaviridae* family, and therefore had to disregard them in my calculations, leaving me with a total of 12 families. Additional difficulties in acquiring data for this family are explained later (cf. section Difficulty of benchmarking the virus mass ribbon).

1.4.4 Calculating the total mass of the virion

Figure 16 displays the calculated values of total virion mass of individual HuReVs in decreasing order. In this chart I compiled all of the 'capsid-derived' calculation values (filled circles) and the 'genome contribution' calculation values (empty circles) described above.

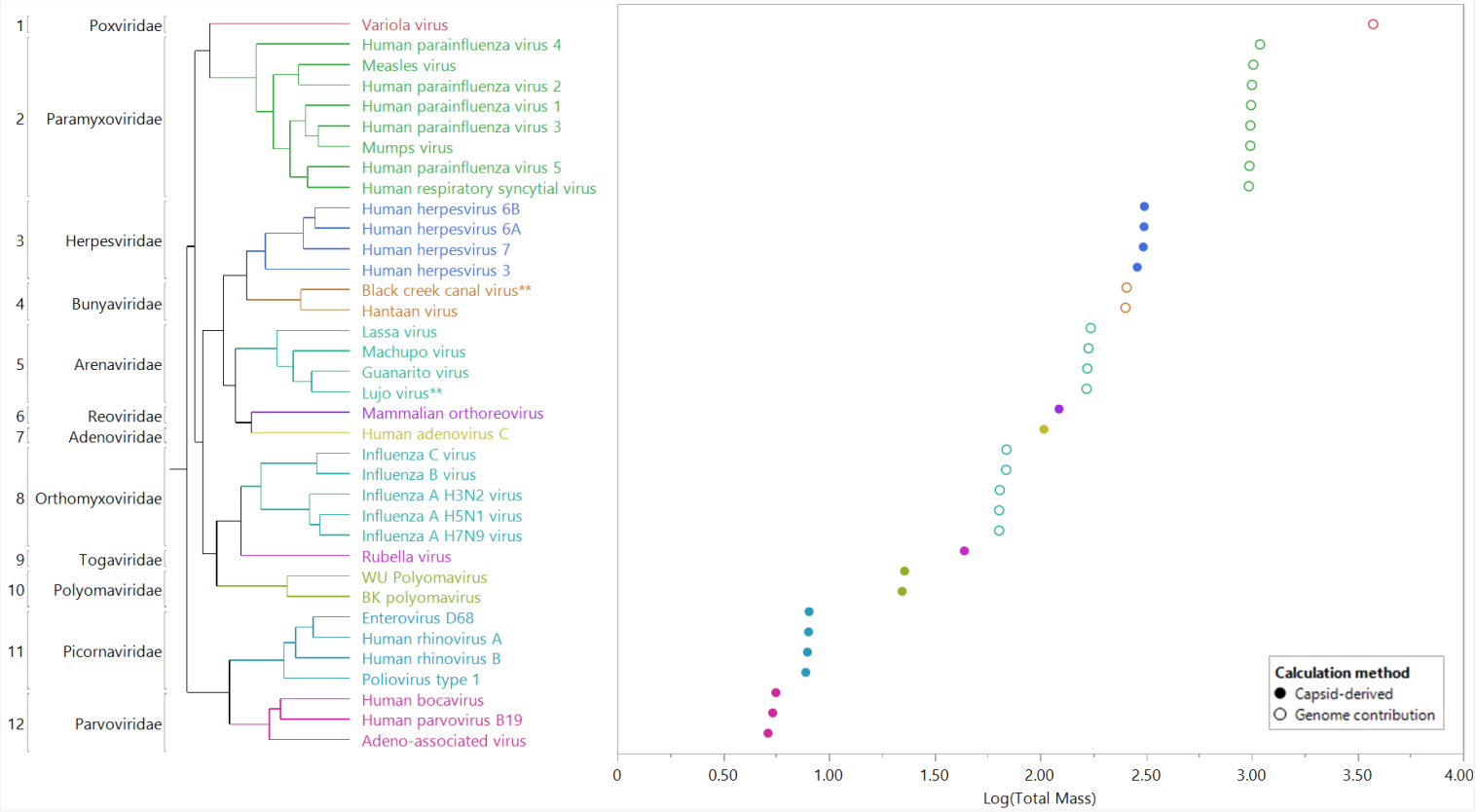


Figure 16. Phylobaric classification of HuReVs by total virion mass. Left: The respective viral families are matched to their corresponding cluster number (n=12). Middle: Dendrogram (geometric spacing) resulting from the unsupervised hierarchical clustering of indidivual viruses based on their log(Total Mass [MDa]). Right: Individual virus log(Total Mass [MDa]) ranked in decreasing order and colour-coded by cluster. (**) indicate rare viral species, as reported by Geoghegan et al.[53]. The data in this plot are highlighted in orange in the tables in Table 15 and Table 16.

The total calculated virion masses of HuReVs covers three orders of magnitude (~ 5 - 5000 MDa), which is comparable to the range of experimental mass values found by biophysical approaches (see Figure 13), and an order of magnitude broader than the masses computed from their genomes alone (see Figure 14). Importantly, and as anticipated based on our mathematical considerations, the discontinuous nature of the mass distribution is more readily noticeable, supporting my hypothesis that this method would enhance classification performance. As above, unsupervised hierarchical clustering of total virion masses was performed, yielding the associated dendrogram and colour-coded clusters. Once again, the number of clusters was set to match the number of families considered (12 in this case). The

clusters obtained matched perfectly (12/12) with the known HuReV families, confirming the potential of total virion mass as an identification criterion.

To give this classification based on molecular mass greater meaning, I coined the term '*phylobaric*', from the ancient Greek '*phûlon*' (tribe, genus, species) and '*báros*' (weight, load).

1.4.4.1 Rates of mass separation

The classification presented in Figure 16 is valid in relation to equation (5), as it demonstrates the better separation between consecutive masses, when the 'container' mass is added to the genome mass. However, the inter-family discrimination in Figure 16 cannot be explained quite so simply based on this added contribution.

To clarify how the various families are distinguished, we need to return to our previous mathematical considerations and examine the prevalence of capsid mass over genome mass. This can be determined by identifying the radius $R_{2,0}$ at which the quadratic term prevails over the subquadratic term in equation (5). This radius verifies:

(10)

$$\rho_L k\, R_{2,0}^{\alpha} \leq \rho_S 4\pi\, R_{2,0}^{2}$$

Finally, radius $R_{2,0}$ satisfies:

(11)

$$R_{2,0} \geq \left(\frac{\rho_L k}{\rho_S 4\pi}\right)^{\frac{1}{2-\alpha}}$$

Thus, for viruses with radius equal or greater than $R_{2,0}$, the quadratic term prevails in the mass of the virus, and the difference between masses is mainly due to the 'genome container' mass (region highlighted in red). The reciprocal is also verified: below this radius, the subquadratic term dominates, and the mass difference is essentially due to the genome mass (region highlighted in green), as we can observe in Figure *17*.

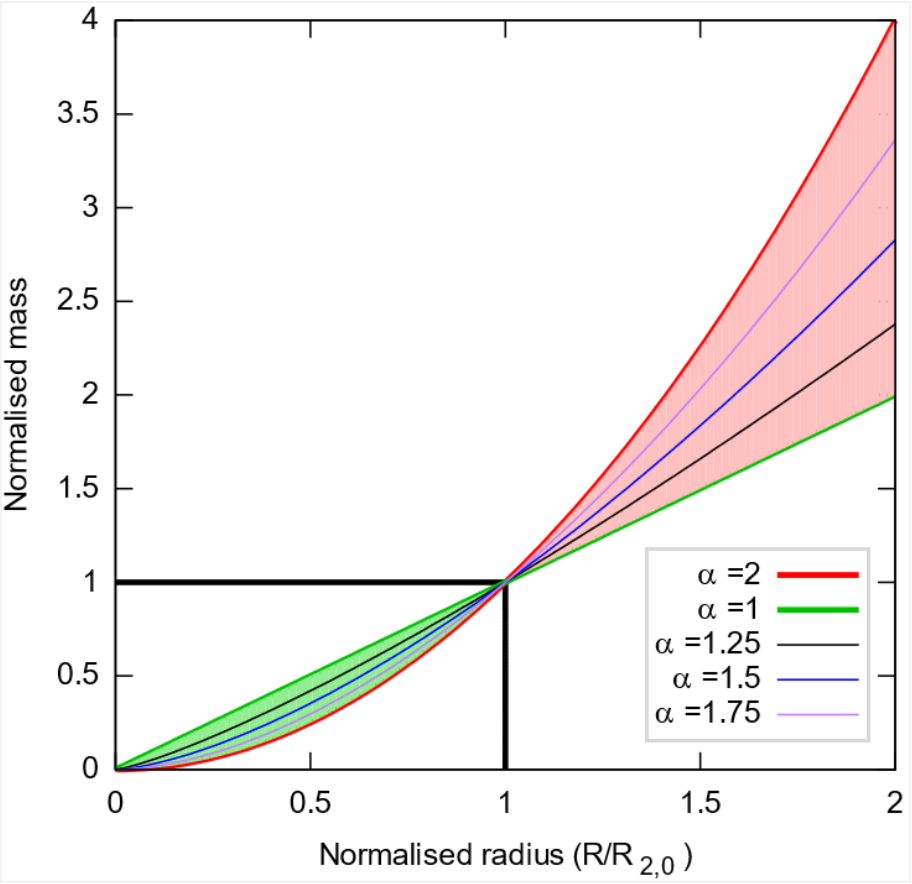


Figure 17. The relationship between the normalised mass and the normalised ratio between the Radius and the Critical Radius ($R/R_{2,0}$). The respective values of α are given (bottom right), from 1 to 2, with a 0.25 step. In green: the linear relationship, representing here the 'genome mass' contribution; in red: the quadratic relationship, representing here the 'genome container' mass contribution; in grey: the value of the allometric exponent (1.25), taken from Chaudhari et al.[72]. The region highlighted in green shows the section where the genome mass alone is more relevant than total virion mass, delimited by the critical radius (solid black lines), after which the region in red shows the section where the 'genome container' mass prevails.

This singularity suggests that $R_{2,0}$ is the critical radius based on which variations in mass between similar viruses (intra-family) can be distinguished from variations in mass between distinct viruses (inter-family). This may explain why clustering by total virion mass (genome + 'container') is more effective than by genome mass alone. The fact that viruses from distinct families appear to separate more efficiently than those within the same family suggests that intra-family mass distances could be explained by the R^{α} term, while inter-family mass distances could relate to the R^2 term.

Thermodynamics and free energy considerations could certainly be invoked to refine this model. To quote Šiber et al.: "*some viruses can be viewed as thermodynamically optimal structures, i.e. they represent minima of the free energy.*"[40] A major breakthrough going toward this logic was performed very recently, where the structural models of a HK97 virion were performed, thanks to the usage of multiple microsecond-long all-atom simulations.[79] There the effect of the packaged genome on capsid structure were characterized, taking into account the internal pressure, electrostatics and diffusion of water, ions and DNA, even revealed the

structural imprints of the capsid onto the genome. But these considerations were put forward only very recently and are beyond the scope of the present discussion.

In order to check how our theory and model compare to the few available mass values from the literature, I decided to compare the data, as shown in Figure 18.

1.4.4.2 Family average masses: cluster versus experimental

The relationship between the mean experimental masses as a function the mean cluster masses is shown in Figure 18.

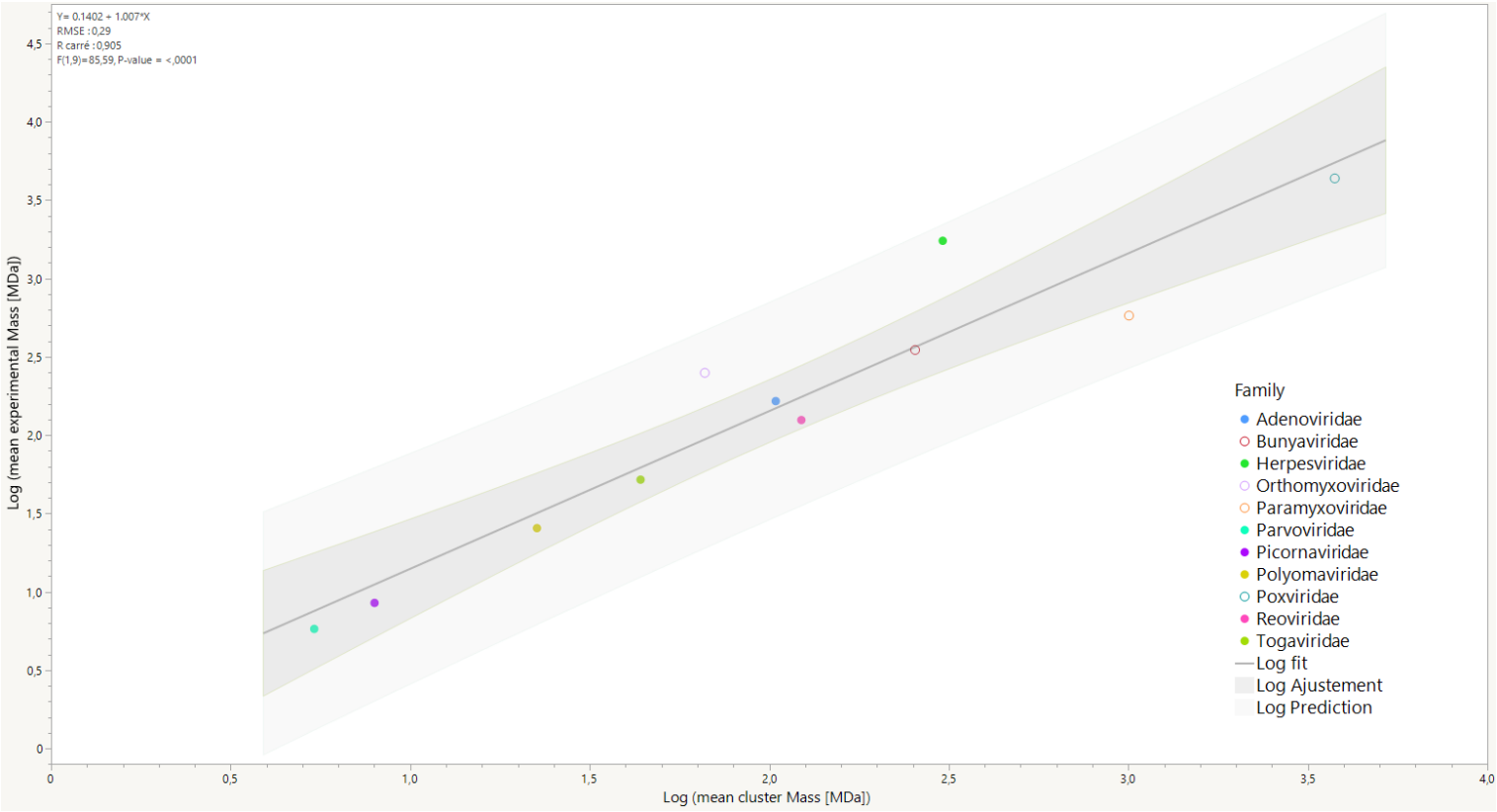


Figure 18. The relationship between the mean experimental masses and the mean cluster masses. Log-log fit with n=11 families (missing Arenaviridae, Coronaviridae). Filled dots: capsid-derived approach; hollow circles: the genome-contribution approach. Dark grey highlight: the confidence interval; the grey blue: prediction interval. The legend for each family is given in the bottom right (in alphabetical order), and the fit equation and relative statistics are given on the top left.

The values for the mean mass of the clusters are slightly underestimated (y = 0.1402 +1.007*x; R^2 =0.905), which is probably due to lack of accounting for non-structural proteins. While the majority of points closely match the trendline and are included in the prediction (light grey shaded area), three families appear to deviate, i.e. *Orthomyxoviridae* (hollow purple circle), *Paramyxoviridae* (hollow orange circle), and *Herpesviridae* (solid green dot).

We know that *Orthomyxoviridae* and *Paramyxoviridae*, present segmented genomes (different genome segments in one viral particle),[53] so I think any error in the genome contribution (e.g. losing one of the segments, during purification for example), could become

very significant in the final mass calculation. Another complexity could be added to the fact that some of these viruses are multipartite (different genome segments in different viral particles), which is still a misunderstood phenomenon in virology.[80] As for *Herpesviridae*, it has already been observed before that the presence of the envelope can nearly half the mass, e.g. for Herpes simplex virus, the mass went from ~780 MDa to ~420 MDa, when the envelope was removed.[81] Importantly the trend is drawn from a limited number of data and more data would needed to validate these observations, especially by using the same measurement techniques.

This part also draws attention also on the method that were used to reach conclusions on the total virion mass and some of the limitations that I met when collecting the data.

1.4.5 Method Limitations

I am aware of certain limitations of the calculation model presented. Hereafter I will discuss of these limitations in the following order: First, the main data limitations relating to the assumption made for the theoretical method of mass calculation; Second, difficulty on relying on the data and avoiding bias, with a focus on the *Coronaviridae* case; Third, limitations for comparing the theoretical model with experimental mass data, and their method dependence.

1.4.5.1 Limitations on data collection

In my model, I did not account for all capsid components (*e.g.* spikes or genome binding proteins), when performing the 'capsid-derived' approach, as precise data were unavailable for many of their protein counts. I reasoned that these components would constitute minor contributions to the total virion mass. Although this appears to hold true for smaller virions, it could be less valid for larger entities, as observed in recent experiments with human adenoviruses.[78] Importantly, the mass spectrum of viral particles may not in fact form a single peak at the total virion mass. If multiple features are represented, they could potentially constitute a viral signature pattern.[82,83] Importantly, I had to rely on a limited body of information to estimate the genomes and total virions mass. Moreover, there was clear partiality in favour of a few virus families in terms of available data, which may be due to them being widely used models (*e.g.* Adenoviruses), or because their low risk-profile makes them easier to analyse in the laboratory (*e.g.* Adeno-Associated Viruses). Additional difficulties were sometimes encountered when comparing cited sources to their original reported values, and these difficulties were further aggravated with the rather large discrepancies encountered when

trying to compare data acquired through diverse methods, as seen for case of the family *Coronaviridae* (see below).

1.4.5.2 Coronaviridae, and the importance of choosing the right data and methods.

The recent COVID pandemic, made us very aware of the presence and the risk that Coronaviruses represent, especially as a pathogen causing human respiratory diseases.[54] As highlighted earlier, I was unable to compute the total mass of *Coronaviridae* virions using the approach I developed, because of the lack of experimental values for 'genome contribution'. This came to us as a surprise and suggests that such measurements are particularly challenging or no longer systematically performed. Therefore, I looked at the literature for experimental measurements in order to make empirical evaluations about the proposed model.

I found that the genome mass was in fairly good agreement with the report of Robb and Bond (~8-9 MDa)[65], to my mass estimate (8.8-9.8 MDa), taken that the genome length of coronaviruses spreads over 22–36 kbp.[84] When it comes to direct total virion mass measurements, the conclusions were harder to draw. Older evaluations, by Hierholzer et al. (1972), using analytical ultracentrifugation (AUC), showed that the intact Human coronavirus OC43 virions were nearly 112 MDa (390 Sv),[85] which seems to have been misreported by Robb and Bond, as being 390 MDa [65], probably referring to the sedimentation coefficient of the first. This evaluation is very close to the value (400 MDa) currently reported by the ICTV 10th report on Coronaviruses.[84] In more recent examination by Degueldre et al. (2021), indirect mass analysis by single particle inductively coupled plasma mass spectroscopy (SP-ICP-MS) on SARS-CoV-2 virions, showed the value to be at 221,5 MDa [86], which is nearly double that reported by AUC.[85] A more direct single particle mass evaluation by Erdogan et al. (2022), on inactivated SARS-CoV-2 virions, using Nanomechanical resonator-based MS (NEMS-MS) following electrospray ionization, estimated the mass to be 454 MDa [87], which they argued that it corresponded well to size and density values reported for SARS-CoV-2 particles. Nevertheless, this seems to be twice the mass measured by SP-ICP-MS[86] and four times that measured by AUC [85], but not so far from that reported by ICTV 10th report.[84]

These difficulties highlight the need for quality control when reporting data in the future, particularly if they are to be included in a well-structured and well-curated total virion mass database. In addition, these characteristics were often determined using outdated methods, sometimes leading to overestimations, or overly imprecise estimations. This case also showed that the consolidation of the data by a variety of methods is needed to make sure of the validity

of each approach, which lead me to look more in depth on the methods used to evaluate the total virion mass.

1.4.5.3 Difficulty of benchmarking the virus mass ribbon

In order to see how previously calculated virion masses (cf. Figure 16), matched experimental measurements, I also compiled the graph displayed in Figure 19. In this graph, the viral families are shown in order of decreasing mass, whose measurement is derived from different methods, labelled according to the underlying physical principle. Scattering methods are divided by the particle source, that is EM (Electron Microscopy); Neutrons; X-rays. Here there is data for 12 families, with the absence of *Arenaviridae*. The totality of the data for Figure 19 can be found in supplementary (Table 17).

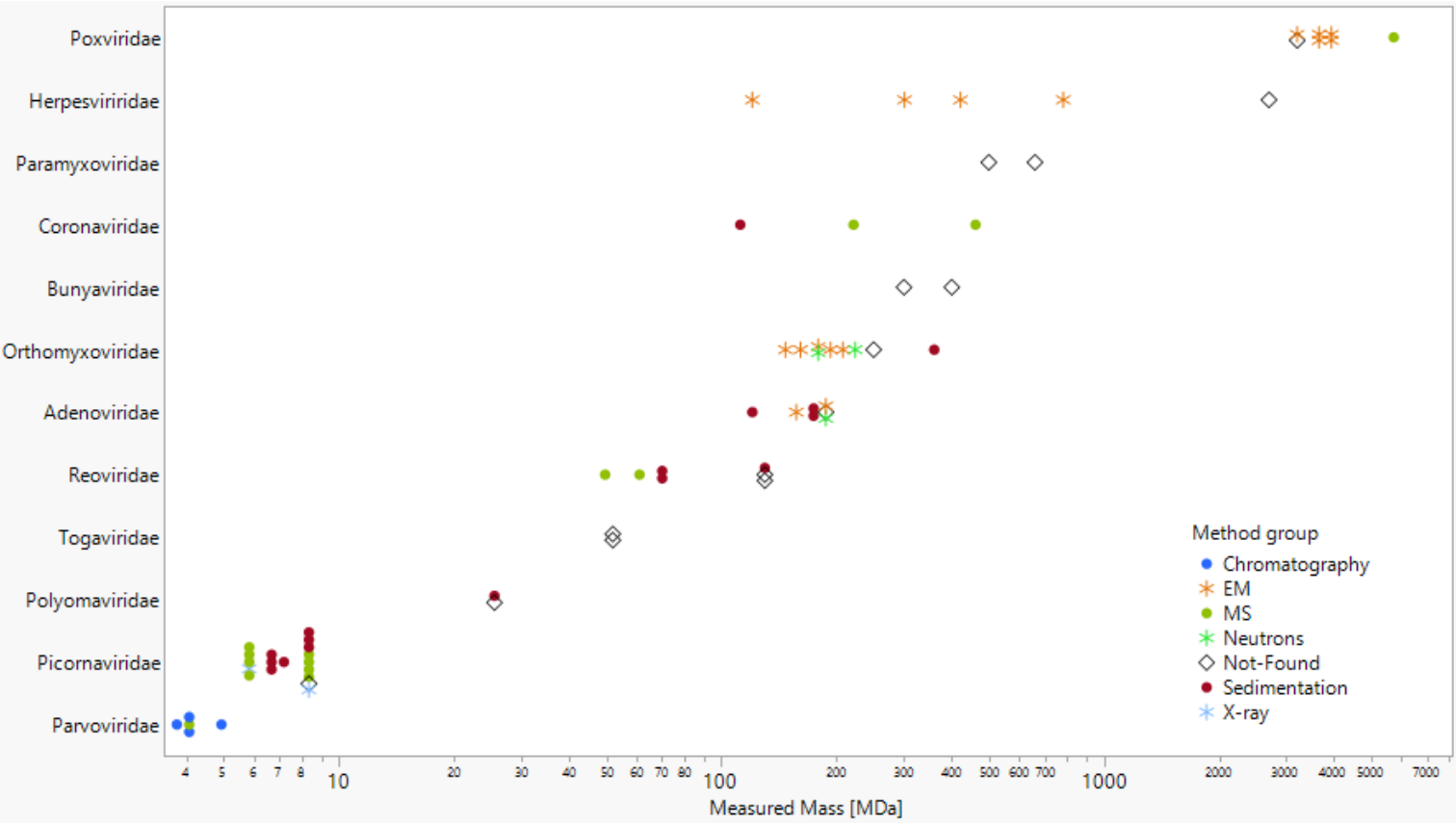


Figure 19. Viral families, ranked in decreasing mass order for maximum measured value, as a function of molecular mass, as determined by various experimental methods (legend found in bottom right). Method's group abbreviations (in alphabetical order): EM: Electron Microscopy; MS: Mass Spectrometry; 'Not Found' (empty rhomboid): not specified in the original article. Scattering methods (highlighted in green in Table 1) are denoted by different coloured asterisks. Data are found in supplementary *Table 17.*

At first sight, we can see that there is an unequal representation of mass measurement values between families. Several of them have tens of measurements, whereas others only have a single one, sometimes with the method being undefined in the original source, marked here as 'Not Found'. This data discrepancy can perhaps be related to the fact that different families

have presented varying degrees of interest to the scientific community over time, especially from the mass measurement perspective.

When comparing the ranking of the families, 11 of them being the same between Figure 16 and Figure 19. We can primarily see a conservation of the ranking for 6/11 families, more precisely for *Poxviridae* (#1) and *Bunyaviridae* (#4, not counting for *Coronaviridae*, on the top half; and *Togaviridae* (#9), *Polyomaviridae* (#10), *Picornaviridae* (#11) and *Parvoviridae* (#12), on the bottom half. Their masses also fall in a very similar range to those theoretically calculated. A switching of places has occurred for *Adenoviridae* with *Reoviridae,* and for *Herpesviridae* with *Paramyxoviridae*. In addition, a significant increase can be seen for *Orthomyxoviridae,* going from ~70 MDa to ~150-400 MDa.

To explain the discrepancy between my model and the experimental data, from a biological perspective, I propose the following explanations. It is known that the *Reoviridae* tend to present a double capsid shell, which could perhaps explain the two mass groups recorded, between ~50-80 MDa and ~150 MDa, which matches to the theoretically calculated value. The *Adenoviridae* seem consistently higher to those of *Reoviridae*, ~200 MDa. Such divergence can be seen even in the recent work from Barnes et al.[78], where they showed the discrepancies between theoretically calculated molecular component sum and their measurement by CDMS (Charge Detection Mass Spectrometry). Regarding the big difference in estimation for *Herpesviridae*, we can notice an underestimation by nearly an order of magnitude (~300 MDa to ~3000 MDa) for the very high mass reported. But the following measurements, all reported by the same article (Lampert et al.[81]), show that the mass changes depending on the whether the virion contains its genome, only the core or with the envelope, the changes become significantly different. Nevertheless, my theoretical estimation is only two-fold lower compared to these new estimations, which is a good thing given the orthogonal approach chosen. This could come perhaps due to the fact that I only accounted for the structural proteins that compose the capsid and added the lipid content measured by empirical methods, which seems to have been an underestimate (reported = 22% versus ~ 46% in Lampert et al.[81]). As for *Paramyxoviridae,* the discrepancy downgrades it from ~1000 MDa to ~600 MDa. Taking into account the biases of the 'genome contribution method', it is evident that an error of just a few percentage points can results in major changes in the final result, which can be also true for *Orthomyxoviridae*. In addition, these viruses have segmented genomes,[53] which can also hinder the precise evaluation of the total genome content, depending on the technical preparation. It can be added that I also noticed an overlap between measurement for several virus families,

namely *Coronaviridae*, *Orthomyxoviridae*, *Adenoviridae,* and *Reoviridae*. For the first two, the large spread in the mass range could also be related to their high lipid content and diverse morphology. Interestingly, this overlap comes in the same range where the spherical approximation calculation of the 'mass ribbon' gets more distant from the experimentally reported data (cf. Figure 13).

Nevertheless, and more importantly, it appears that when considering the maximum values for each family, the discrimination between them remains possible using their molecular mass, as this is how I plotted them in Figure 19. These examples also show that the entire mass range of HuReVs can be covered using single particle mass measurement technologies, and that the order of magnitudes found by calculations matched well to the experimental data, supporting the feasibility of the viral mass database. In addition, these demonstrations concern mostly VLPs, and additional developments may be required to tackle real-life samples.

Despite this wide disparity in measurements, the ranking of families obtained by these simple measurements is found to be +50% by my theoretical model, meaning that this model fits rather well despite these difficulties. However, I believe that the data that does not fit my model has two main possible explanations:

- My model is not perfect, and further improvement can be done to refine it;

- But also, the measurements found in the literature are in themselves intrinsically highly variable.

Here, I pointed out the large discrepancies that occur when comparing data acquired through diverse methods. I concluded that different methods yielded highly varying results, within a given family of viruses, potentially making the inter-family discrimination 'method-dependent'. Such inaccuracies cannot be ignored given the type of precision we seek to provide. This observation also highlights the need to perform virion mass measurements with a single technique, or at least easily comparable approaches, in order to have reliable and representative measurements for every family.

1.5 Conclusions on Chapter 1

In this chapter I argue that individual virus molecular masses is a distinctive characteristics of virus evolution. Based on available data, focusing of human respiratory viruses, I estimated their theoretical masses and observed that they distribute non-contiguously over a well-defined and characteristic mass range, which I named the 'virus mass ribbon'. Based on these findings, I suggested that the total virion mass constitutes a meaningful attribute of

human respiratory viruses' biology, allowing for a new type of phenotypic classification, which I named "phylobaric" (from ancient Greek 'phûlon' = lineage and 'baros' = weight). We also made a simple mathematical model than can help explain the improved inter-family separation of viruses, that can aid in clarifying these data. I believe that this observation could have fundamental implications in virology, and a major relevance to public health, especially considering recent advances in single particle mass characterization techniques.

1.6 Outlook and Perspectives on Chapter 1

Although from our human perspective, viruses appear quite similar in size and mass, the scale of mass difference is somewhat equivalent to the difference between the smallest rodents – the African pygmy mouse (a few grams) – and the largest land animal – the African bush elephant (up to 10 tons). Giving the impression of looking at molecular giants when we look at their masses, like in Figure 20.[88]

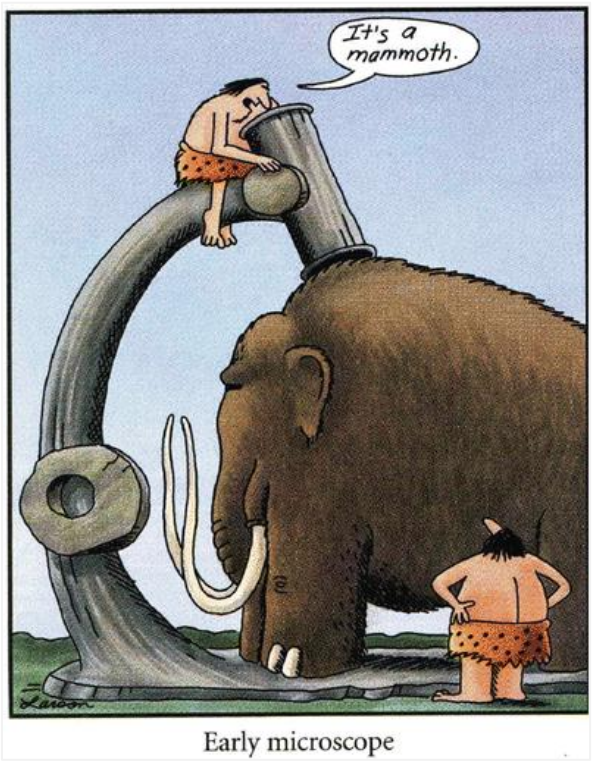


Figure 20 The early microscope. Cartoon from the Larsen's Far Side. Figure was reprinted from Verrecchia and Trombino.(2021).[88]

Over the past two decades, the mass spectrometry community has made major breakthroughs, and modern instrumentations have reached the level of accuracy required for structural virology studies. A remarkable review by Wörner et al.[89] captured the developments in the field for applications in structural virology in detail, and called for efforts to be focused

on developing methods to enhance sensitivity and precision. Several examples of single-particle MS methods application in the biopharmaceutical area attest the compatibility of modern approaches with the multi-MDa virus mass range highlighted in the present work. Such approaches have been applied to characterise gene-therapy vectors, assess vaccine preparations, or study the assembly and disassembly of encapsidated viruses.[90]

Echoing the words of Mazzone[4] in the conclusion of his book, I believe that a virus mass database is needed, and perhaps now more than ever. Some efforts seems to have started to be put forward by the VirusScan initiative, but the current virus database [https://www.viruscanproject.eu/database1/ (only 33 entries); accessed in 4 July 2024] is lacking a significant amount of the available data, even when compared to the data that I was able to collect; with 69 entries (cf. supplementary Table 17). I envisage that a much larger open-access database, combined with future portable MS detectors capable of performing real-time mass measurements,[91] would improve biosafety applications for aerosol-transmissible viruses in locations where there is a high risk of propagation, e.g. hospitals or airport terminals[92,93]. This can also be of critical value in large international events, for example, the ones happening in Europe in the summer of 2024, such as the Olympic games (France) or the European Football cup (Germany).

Initiatives such as the Global Virus Network (GVN)[94] have already shown the impact they can have in times of crisis, by creating efficient task forces to identify and study emerging pathogens. For example, the most recent WHO program called Coronavirus Network (CoViNet), focusing on bringing together surveillance programs and reference laboratories to enhance epidemiological monitoring and laboratory (phenotypic and genotypic) assessment of all potentially pathogenic coronaviruses. Therefore, in this spectrum, the application of new MS methods could facilitate early detection of future potential epidemic outbreaks. Considering the recent COVID-19 pandemic, a rapid non-targeted approach to triage patients that could efficiently determine whether someone was infected by a rhinovirus, by influenza, or by a coronavirus would have been a tremendous help in initial screening, when PCR and antigenic tests were being developed.

My observations suggest that integral virion “molecular mass” should be revisited as an intrinsic and distinctive viral phenotype. In fact, total virion mass might have phyletic implications, when analysing high taxonomic ranks, with the potential to detect evolutionary relationships. A somewhat related approach termed ‘mass-tree phylonumerics’ has been proposed by Downard, who used peptide mass fingerprinting data to construct protein-based

viral phylogenies.[95] That is why I believe that the presented *phylobaric* classification, presented by me in this chapter, is actually an 'emergent property' of viral biology.

I would like to make a small parenthesis regarding emergent properties. In his now very famous 1972 article, named 'More Is Different', Philip W. Anderson [Nobel Prize in Physics (1977)], stated the following: "*The behavior of large and complex aggregates of elementary particles, it turns out, is not to be understood in terms of a simple extrapolation of the properties of a few particles.*"[96] He goes on to give examples such as: Psychology is not applied biology, nor is biology applied chemistry. The latest definition of 'emergence in physics', as : "*An emergent behaviour of a physical system is a qualitative property that can only occur in the limit that the number of microscopic constituents tends to infinity.*"[97] That is why I believe that given the particular status that viruses hold, i.e. as an intermediary of living/non-living things, having biological attributes (such as family formation) derived from physical attributes (such as molecular mass) can be a very interesting case example of their evolutionary existence on Earth.

Since the discovery of the genetic code, two visions of molecular biology have co-existed, an 'informationist' view and a 'structuralist' standpoint.[98] In a February 2024 book review article, named "It's time to admit that genes are not the blueprint for life", Denis Noble wrote that, despite the major breakthrough the development of genomic method has allowed to the entirety of biologic research, focusing on it as the only type of important research in biology is "*oversimplified and out of date*".[99] Even at the cellular scale, in a recent perspective article summarising the book 'The Master Builder'[100], it was stated that the 'gene-centric view of life' is being reconsidered when it comes to determining cell phenotypes, structures and morphologies; in favour of an alternative, cell-centric viewpoint.[101] Likewise, with the concept of *phylobaric* classification of viruses and the development of single-particle mass measurement methods, the time may be ripe for both the phylogenetic perspectives ('informationist' view) and the mass-to-structure ('structuralist' view) to be combined for tackling such fundamental questions surrounding viral biology.

2 Chapter 2: Methods for virion mass evaluation and the place of NEMS-MS

2.1 Introduction of Chapter 2

In the first chapter I presented evidence that the total virion mass is a distinctive property of HuReV families. If our observation can be generalised to other types of viruses, it would simplify the discovery and classification of new and potentially dangerous strains. This would require assembling a library of total virion masses to serve as a reference source for known viruses, and help the discovery of new viruses in the ever-expanding virosphere.[102] To achieve this, numerous techniques have been developed over the years.

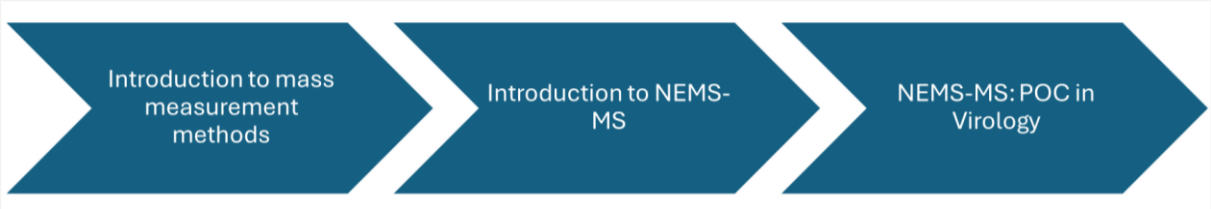


Figure 21. Schematic representation of the structure of Chapter 2.

In this chapter, as indicated in Figure 21, I will first describe from the literature the past and present of mass measurement methods, especially highlighting the development of new approaches in MS (Mass Spectrometry) that have put their scope to virology, and especially onto single particle virion mass evaluation. Then I will demonstrate how these methods have made their contributions in benchmarking the theoretical calculation shown for the virus mass ribbon until now, and their limitations. Next, I focus on the Nanoelectromechanical sensors mass spectrometry (NEMS-MS) instrumentation, with special emphasis on the approach that our group has developed[20], and in its potential to become one of the leading methods in this burgeoning field of research.

2.2 Overview of total virion mass measurement methods

In his 1998 “Handbook of Viruses: Mass-Molecular Weight Values and Related Properties” [4], Mazzone made a comprehensive summary of the available methods at the time. It covers imaging techniques related to scattering or transmission of different electromagnetic

waves (i.e. electrons, X-rays, neutrons, visible light), as well as sedimentation analysis by centrifugation. These approaches have seen great improvements and have had their own challenges over the years, as we will briefly detail in the following paragraph.[66,103]

What has now come to be called 'the resolution revolution'[104] in the field of cryogenic electron microscopy (cryoEM), has been the driving force that has led to it becoming the new 'golden standard' for structural biology, and taking over the once omnipotent X-rays crystallography. The major challenge with this last being the necessity to have the biological entity locked inside a crystal,[105] which is very cumbersome and sometimes proves too challenging, especially for dynamic structures.[106] De Oliveira and Silva wrote that the primary turning point that have made this rapid advancement possible have been: "the vitrification process of EM samples (confining samples into an amorphous phase ice structure); the technological developments of the field emission gun, the automated electron microscope, and the direct electron detector".[103] On top of all these analytical and technical improvement, major advancements in the field of image processing and deep learning algorithms have significantly simplified the workflow of single particle analysis.[107] De Oliveira and Silva also wrote that some major resolution-limiting factors in cryoEM are, quote: "the quality of the electromagnetic lenses, the under focus settings, the low contrast data, the radiation damage of macromolecules, the quality of recording devices, among others"[103]. Analytical ultracentrifugation (AUC) has also seen major instrumental developments in recent years, such as optics integration in order to measure absorbance, interference, and fluorescence, and to be able follow samples in real time as they sediment.[108]

Nevertheless, although imaging techniques may give an ever-increasing resolution for 2-D images for viruses, the third dimension can only be achieved through image reconstruction. In addition, sample structural modifications can arise from sample preparation[109,110] and surface-interface interactions when depositing.[103,111] In order to improve the resolution of the third dimension, techniques such as Atomic Force Microscopy[112,113] or volume Electron Microscopy[114], usually combing cryoEM to cryogenic electron tomography, are emerging as promising alternatives, yet they are unable to directly access the particle mass.

In fact, in Mazzone's time, native Mass Spectrometry (MS)[18] was in its infancy, and single particle ultra-high mass analysis methods, such as charge detection MS (CD-MS),[90] mass photometry,[115] or nanoresonator MS,[21] were mere curiosities, used only in specialized laboratories, or not yet conceptualised. Since then, new instruments have been developed and analytical methods have advanced, meaning that it is now feasible to directly measure high

mass analytes, in the multi-million of Dalton (MDa) range, and more specifically viruses.[89,116] Incidentally, structural integrity is a strong indicator of virus infectivity, but is notoriously difficult to determine from gene sequencing data.[117]

Consequently, effective direct approaches to virus mass measurements are now available that could provide specific information on virus identity, help build useful databases, and contribute to investigations of virion structural integrity.[118]

In the next portion of this chapter, I will highlight the key methods that we were able to find, that have been used in the past or are currently being developed, and that can be used for total virion mass evaluation.

2.2.1 Methods capable to perform virion mass measurements.

Hereafter I will briefly demonstrate the modus operandi and physics that these methods use, and also present some of the advantages and disadvantages, if they are to be used in the near future.

2.2.1.1 Physical principle and modus operandi

In Table 1, I present the different methods, and grouped them in two major categories, primarily relying on the physical principles that they use to reach a mass evaluation. The first group is what we denoted as 'radiation-based methods' (highlighted in green in the column 'Physical principle'), They were ranked by the energy of the radiation used, from lowest to highest, in the order: visible light; neutrons; X-rays; electrons. And the second group are the more 'analytical chemistry methods' (highlighted in orange).

Regarding the 'radiation-based methods', first we have SEC-SLS, combining Size Exclusion Chromatography (SEC) and Static Light Scattering (SLS), does a first chromatographic separation of the analytes based on their hydrodynamic size, and thus volume, which is directly related to the mass (sometimes called molecular weight). The mass is then evaluated by relating the light intensity scattered by the analyte, measured at different angles, to its molecular weight and the gyration radius of the molecule. Mass photometry (MP) is based on interferometric scattering video-microscopy to derive the mass of individual viral particles in solution from the intensity of the scattered light.

Small Angle Neutron Scattering (SANS) and Small Angle X-ray Scattering (SAXS) are two complementary methods in structural biology, and as the names may indicate, are primarily differentiated by their radiation source. Since SANS uses neutrons, they interact with the nuclei of the bulk sample, and it is more sensitive to the lighter elements, such as hydrogen and

deuterium, thus allowing a better contrast variation. On the other hand, SAXS uses X-rays, which interact with atoms' electron clouds, allowing for a more in depth and higher resolution evaluation of the structures of interest. The mass is calculated by the following formula:

$$\text{Intensity } [Q = 4\pi\sin\theta/\lambda] = MW^2$$

where Q: momentum transfer (scattering vector); θ: scattering angle; λ: wavelength; MW: molecular weight. Quantitative Transmission Electron Microscopy (QTEM) and Scanning Transmission Electron Microscopy (STEM) use a focused electron beam to scan across a thin sample. The intensity of the beam is calibrated to a particle with a known mass, that being a virus particle or a nanoparticle, thus providing an indirect mass evaluation.

Analytical ultracentrifugation (AUC) subjects analytes to a high centrifugal force field and their sedimentation rate is measured and related to the particles shape and size, and once the particles reaches a sedimentation equilibrium, which is density dependent, we can measure its displaced volume and thus the molecular mass of the analyte. Charge detection mass spectrometry (CD-MS) relies on the simultaneous determination of the charge (z) and mass to charge ratio (*m/z*) of individual ions, by means of impaction with a surface or from the image current induced on electrodes. Nano-electro-mechanical sensor mass spectrometry (NEMS-MS), where NEMS are oscillating nanostructures, which are sensitive to the mass increment caused by single particles landing onto their active surface, through changes in their vibration frequencies. Gas-phase electrophoretic mobility analyser (GEMMA), also named native nano-electrospray Differential Mobility Analyser (nES-DMA), or commercialised also as Scanning Mobily Particle Sizer (SMPS^{TM}), uses gas flows and electric fields to separate particles as a function of their electrophoretic mobility.

2.2.1.2 Method catalogue

One of the major points that I wanted to make with this table is summarizing the mass range and concentration range in which these techniques work. This is in order to give an idea of the limits and requirements for these approaches, and how their contributions compare to one another. I was able to do this by evaluating a number of articles, over a large timescale, and was able to piece together how these methods have been used to perform virus mass evaluation.

Table 1 details the key information for each method, in relation to its functioning principle, the mass and concentration range, and lastly, their accessibility. Abbreviations on methods used are also described in Glossary (page 24).

*Table 1. Table presenting the differences between different methods currently existing to measure mass. Legend: **Physical principal** classification (highlighted): radiation-based methods (in green); analytical chemistry methods (in orange). Method names can be found in the main text and are summarised in Glossary.*

Method name	Physical Principle	single particle	Mass determination	Mass range	Conc. range	Requirements	Availability	ref
SEC-SLS	Light scattering	No	Indirect	Density dependent	> $4*10^{13}$ (cps/mL)	Known density and refractive index	Commercial	119–121
MassP	Light scattering	Yes	Indirect	$4*10^4$ – $5*10^6$ (Da)	10^{12} –10^{13} (cps/mL)	Calibration needed for proteins & nucleic acids	Commercial	121
SANS	Neutron scattering	No	Indirect	10^5 – 10^9 (Da)	< 1 (mg/mL)	Known concentration; ultra-pure sample	Specialized Research facilities	4,122
SAXS	X-ray scattering	No	Indirect	Wavelength dependent	10 (mg/mL)	Known concentration, partial specific volume, and electron density	Specialized Research facilities	4,123,124
QTEM	Electron transmission	Yes	Indirect	$6*10^6$ – $6*10^{13}$ (Da)	≥ 10^7 (cps/mL)	Standard reference in the exact same conditions as the sample	Commercial	4,125–130
STEM	Electron scattering/ transmission	Yes	Indirect	< $3*10^8$ (Da)	10^2 - 10^3 (cps/mL)	External or internal standard of well-defined mass	Commercial	4,131,76,77,132–135
AUC	Sedimentation	No	Indirect	10^4 – 10^9 (Da)	2 *10^{12} – $5*10^{12}$ (cps/mL)	Known concentration, partial specific volume, and diffusion coefficient.	Commercial	4,136–138
nMS	Mass Spectrometry	No	Direct	< $2*10^7$ (Da)	~ 0.04 (µM capsids)	ESI compatible buffer; highly charged analyte	Commercial	19,89,139
CD-MS	Mass Spectrometry	Yes	Direct	< 10^9 (Da)	~ $3.3*10^8$ (cps/mL)	ESI compatible buffer, highly charged analyte	Research facilities	140–142,121,143
GEMMA (SMPS™)	Mass Spectrometry	Yes	Indirect	< $2.7*10^7$ (Da)	$7*10^{12}$ (cps/mL)	ESI compatible buffer; compound specific calibration standards	Commercial; Research facilities	144–148,89
NEMS-MS	Mass Spectrometry	Yes	Direct	10^6 – 10^9 (Da)	> 10^{11} (cps/mL)	sample aerosolization	Research facilities	21,149,87

If we focus on evaluating single particle mass measurement, the distribution between methods is equilibrated, that is five approaches cannot evaluate single particles and rely on bulk measurement, i.e. SEC-SLS, SANS, SAXS, AUC, native MS (nMS); and the other six techniques can perform single particle measurements, i.e. MassP; QTEM; STEM; CD-MS; GEMMA; NEMS-MS. Most of these techniques are based on indirect evaluation of particle mass, with the exception of only three MS methods, i.e. nMS, CD-MS and NEMS-MS.

Most of the methods in the radiation-based group (in green), i.e. SANS, SAXS, QTEM, STEM, are still used in structural biology, but no longer to measure total virus mass, as this is very time consuming, gives relative mass evaluation and thus it is not enough informative for structural biologists in their current applications. Recently developed commercial light-based methods have already made some proof of concept in validating virus mass, e.g. SEC-SLS and MassP, but are often on the lower mass range of the virus mass ribbon, i.e. 10^6-10^7 Da. AUC has one of the largest mass ranges compared to all the other methods, extending over five orders

of magnitude, i.e. 10^4 – 10^9 Da. Nevertheless, it requires large sample quantities, i.e. over 10^{12} particles/mL, and again total virus mass evaluation has not been one of the goals in its instrumental development.

Lastly, the strongest contenders for the measurement of total viral mass have been the instrumental developments of the mass spectrometry (MS) community, especially pronounced by the variety of CD-MS detectors, and their ever-increasing resolution. Although the *phylobaric* evaluation, I presented in the previous chapter, has not been one of the primary objectives of the MS community, it has coincided nicely with the fact that many of these methods use different virus particles, as simple biological nanoparticle models, in their instrumental development. We can conclude here that a number of these methods cover a large scope of the virus mass ribbon, especially true for CD-MS and NEMS-MS, reaching higher limits of detection of several GDa (10^9 Da).

Finally, and what is of key interest, is that most of these instruments are currently commercially available, but not necessarily used to perform total virion mass measurements, e.g. QTEM, STEM, SMPS and AUC. Whereas techniques such as SANS and SAXS, do require specialized facilities, such as synchrotrons or nuclear reactors, which are very costly to run and maintain. Although most of the current MS techniques are primarily based in research labs, some efforts have been made in commercializing them. This is especially relevant for the CDMS methods, based on a Orbitrap detectors (being commercialized as 'Direct Mass Technology Mode' by ThermoFischer) or the Megadalton solutions technology, based on a electrostatic linear ion trap detectors (being commercialized by Waters).

During my thesis, I had the opportunity to use one of these single-particle mass measurement method, called Nanoelectromechanical sensors mass spectrometry, abbreviated as NEMS-MS, which will be my focus in the rest of this chapter.

2.3 The NEMS-MS instrumentation

Nanoelectromechanical sensors mass spectrometry (NEMS-MS) is a single-particle mass measurement method that has multiple advantages that make it suitable for the measurement of virus and virus-like particles (VLPs).

It is a new technology compared to the other methods presented earlier, and has found applications in the field of nanoparticle [150] and virus [21] analysis. Its main attributes are:

1. Large mass range for high mass analytes,

- from few MDa to several GDa {few ag ($\times 10^{-18}$ g) to several fg ($\times 10^{-15}$ g)}.

2. Charge-independent (neutral) mass measurement.
3. Mass-independent resolution.
4. Low pumping requirements,
    - considerable possibility for instrument downscaling.

Each of these contributions will be detailed in the following paragraphs. One of the major focuses of this work is to show, on top of the aforementioned advantages, how we intend to turn this method into an analytical tool for biological and nanoparticle samples, reaching a high throughput mass analysis and scalable instrumentation.

I would like to highlight here that I got help on writing this part thanks to having prior access to the draft of the review article "Advances in Single Particle Mass Analysis".[151]

2.3.1 A mass range adapted for virion analysis.

The first indication for resonators being used as detectors of mass difference came from the work of Thundat et al. in 1995, where resonating microcantilevers where used to detect the adsorption of mercury vapor.[152] Several great achievements came simultaneously in 2004 for the field of resonator-based MS. First, the measurement of the vaccinia virus, of the *Poxviridae* family, was reported by Gupta et al. on a microresonators with nanoscale thickness, at about 9.5 fg.[153] Next, resonating mechanical cantilever was used in solution, by Ilic et al, to perform the mass measurement of the insect baculovirus, around 3 fg.[154] And in that same year, the first demonstration of doubly-clamped NEMS mass detection was shown by Ekinci et al.[155] This opened the race for researcher to reach the smallest resolution possible by NEMS-MS, with the milestone of detecting a single atom with nano-resonators. In 2006, the mass resolution was already brought down to ~7 zg ($\times 10^{-21}$ g), or 4 kDa.[156] Chaste et al, in 2012, then reached a resolution of 1.7 yg ($\times 10^{-24}$ g), which corresponds to the mass of one proton.[157] It has to be noted that at this point all these measurements were just proof of concept for different NEMS designs to measure mass, but none of them had the capability to perform single particles MS analyses, and deliver an actual 'mass spectrum' regarding the sample.

The first demonstration of a NEMS-based mass spectrometers, was performed in 2009 by Naik et al., and consisted of a modified MS architectures in which the detector was replaced by a nano-resonator.[158] Their demonstration was performed on the evaluation of Bovine Serum Albumin (BSA), at 66 kDa, and of β-amylase, 200 kDa, and used gold nanoparticles as mass calibrants, assuming a 2.15 nm mean radius. This is because the mass evaluation was done by

having the event probability versus the amplitude of the frequency-shift, and then compared the single-molecule to single-nanoparticle adsorption events, to reach to their mass evaluation. This was assuming a perfect distribution of the mass of the gold nanoparticles, which prohibited the validation of the mass measurement accuracy.

In 2012 in the context of a collaboration between CEA-Grenoble and Caltech, Hannay et al.[159] introduced NEMS-based MS as a unique approach to analyse heavy weight species, featuring mass-independent resolution, and allowing the acquisition of mass spectra of neutral analytes, one molecule at a time. They demonstrated the feasibility of the method by analysing, in real time, Immunoglobulin M (IgM) antibody aggregates, reaching a mean mass of ~1 MDa, complexes in the 0.5-2.5 MDa mass range. The sample desorbed by MALDI (Matrix-assisted laser desorption/ionization), was guided towards a single nanoresonator using hexapole ion guides, but presented large sample consumption and analysis time, and had a low capture efficiency.

The NEMS-MS group made the first direct comparison of the neutral mass measurement, based on nano-resonator mass sensing, to a conventional mass spectrometer, as described in the work of Sage et al, in 2015.[160] For this they used a customized time of flight (TOF) instrument, equipped with a metallic cluster source, where the NEMS were put directly at the end of the instrument, and the two evaluations of mass were compared. The sample used was tantalum (Ta) clusters, for which the mass ranged from 470 kDa up to 4.5 MDa, which corresponds to a mean diameter of ~ 9.5 nm. The neutrality of the evaluation was demonstrated by cancelling the ion charge, using deflector plates, when performing the TOF-MS evaluation. The NEMS were still able to generate a mass spectrum, allowing the detection of neutrals species, to which the TOF was innately blind. In the discussion of this article,[160] it was argued that performing neutral measurement can help reach simpler and clearer mass spectra and improve sample identification. They also proposed that this approach would allow the innovation of weakly ionizing or non-ionizing methods for particle nebulisers, improving transfer efficiency and providing softer landing on the detectors. They especially highlighted the usage of surface acoustic wave nebulisation (SAWN), which we will discuss in detail later on. Lastly, they highlighted the potential of combing these neutral nebulisation methods to 'ultradense NEMS arrays', to improve the capture efficiency and limit-of-detection by several orders of magnitude to the existing methods at the time.

That last objective was achieved in the 2018 article by Sage *et al.*, where the NEMS-MS group demonstrated using an array of 20 multiplexed NEMS resonators in the customized

TOF-MS setup, mentioned previously.[161] The same test and comparison to TOF-MS was done as before, showing excellent agreement, for three distinct populations of Ta clusters, with mean mass of ~1000 kDa, ~2150 kDa and ~2420 kDa, which convert to 5.8 nm, 7.4 nm and 7.7 nm, respectively. A major improvement of this system is that the total array response time, measured at 152 ms, has the same mass limit of detection as a single NEMS resonator.

All this developments led the team building the very first prototype of the NEMS-MS as an analytical instrument.[20]

2.3.2 Architecture of the first prototype for NEMS-MS

The NEMS-MS instrument (which we sometimes also refer to as NEMS-MS bench) is described in Figure 22 (taken from the original, Dominguez-Medina et al.[21]), and is the first prototype (prototype #1) that was built by the NEMS-MS group[20]. It can be divided into three main compartments (from left to right in Figure 22):

1) the sampling of the analyte, performed using a nESI (nano-Electrospray Ionisation), or the SAWN (Surface Acoustic Wave Nebulisation) methods.
2) the aerodynamic lenses, that guide single neutral particles to the sensors.
3) the sensors/particle detectors, called Nanoelectromechanical sensors (NEMS).

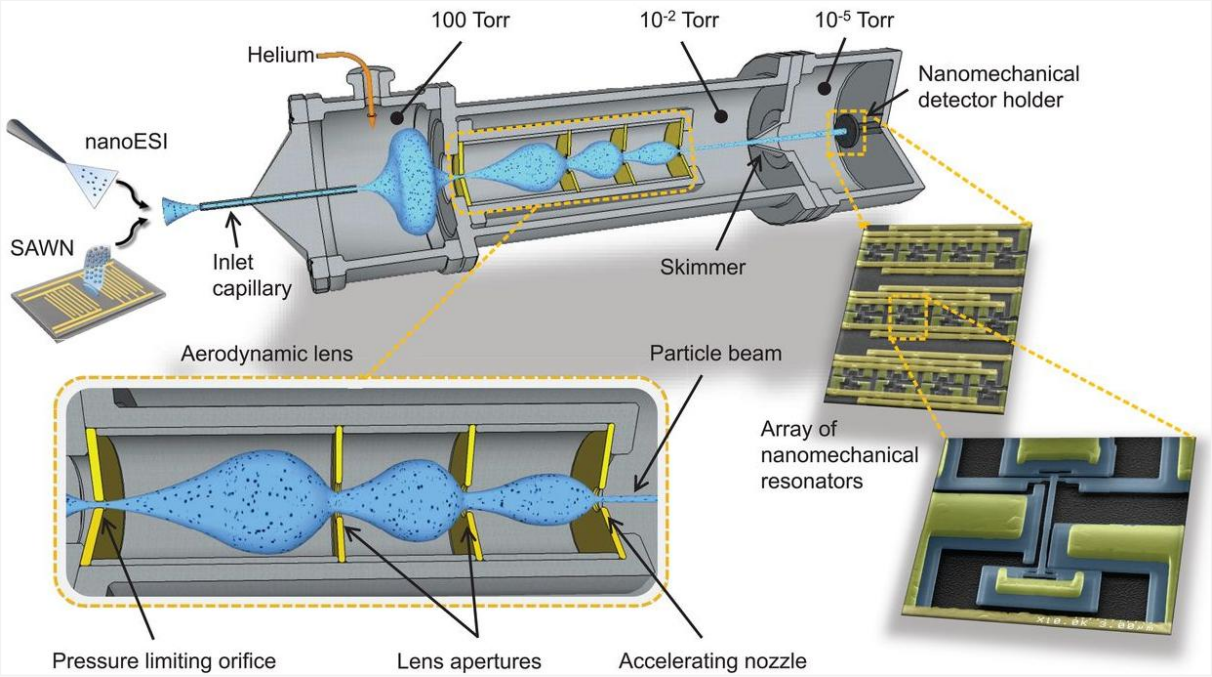


Figure 22. Architecture of the first prototype for NEMS-MS (prototype #1). The setup consists of three chambers with decreasing pressures. Analytes in solution are nebulised by SAWN (Surface Acoustic Wave Nebulisation) or nano-ESI (ElectroSpray Ionisation) and aspirated through a heated metal capillary inlet at atmospheric pressure. An aerodynamic lens focuses the particle stream (shaded blue area), which is then transferred onto an array of frequency-addressed nanomechanical resonators. Left inset: Simplified schematic of the aerodynamic lens (actual design shown in fig. S8). Right inset: SEM image of a portion of an array showing 12 out of 20 resonators and magnified SEM image of a single resonator with its metallic layer

and silicon material, falsely colored in yellow and gray, respectively. Resonator dimensions: 160 nm (thickness), 300 nm (width), and 7 to 10 μm (length). Figure and description were reprinted from the original, Dominguez-Medina et al. [21]

First, the nESI system charges the analytes by producing a spray of nanoscale droplets, that evaporate in the interface.[162] In the case of SAWN, an acoustic wave propagates across a piezoelectric surface, forming a thin mist from the liquid sample.[163] The produced aerosolised particles are then aspirated inside the system, through an inlet capillary tubing, due to the pressure drop induced by a differential pumping system. An aerodynamic lens, inspired from the work of Wang et al.[164], guides the analyte particles to an array of detectors. This is in major contrast to conventional mass spectrometers, in which ion guides are required to enable particle separation and detection. Therefore, the NEMS-MS system allows for the detection of single neutral particles. In continuation, the particle beam that is formed at the exit of the aerodynamic lenses passes through a skimmer in order to reach the NEMS sensors, that will be described in Figure 30 and Figure 36.

Possessing all these tools, this unique NEMS-MS system is able to characterise a large variety of individual particles, with a sub-MDa sensitivity and is theoretically able to reach masses as high as several GDa.[21]

I would like to underline here that the reason why I denoted this instrument as 'NEMS-MS bench prototype #1' is because I contributed to building a second prototype (prototype #2), which we called 'mini-NEMS-MS' bench (due to it being more compact and modular)[165], and which will be detailed primarily in the next chapter of this thesis. In the rest of this chapter, I will detail the state of the art and knowledge which was already present regarding this instrument, prior to my arrival in the group, and indicate how we tried to improve our understanding on this technique when building the 'mini-NEMS-MS'.

2.3.3 Nebulisation

In order to introduce the non-volatile analytes of interest to reach the detectors, i.e. the NEMS array in Figure 30, we need to first nebulise (or atomise) them, meaning the transferring of the analytes from the liquid phase to the gas phase, at atmospheric pressure.

This is detailed the schematic representation introduced in Figure 22.

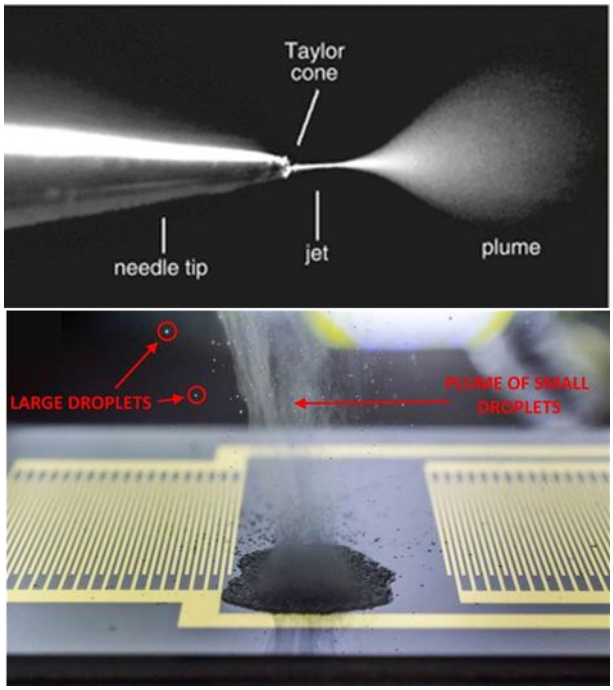


Figure 23. Different nebulisation methods used for sample introduction inside the NEMS-MS bench. (Top) nano Electro Spray Ionisation (nESI) The jet emitted from the Taylor cone is clearly visible and separate from the region of rapid expansion into a plume of microdroplets. Picture reprinted from Gross. [166] *(Bottom) Surface Acoustic Wave Nebulisation (SAWN) chip. Two droplet size regimes can be observed as indicated by the red arrows. Picture reprinted from Kooij et al.*[167]

The primary methods that we used to achieve this goal were: nano Electro Spray Ionisation (nESI), (top of Figure 23), or the Surface Acoustic Wave Nebulisation (SAWN), (bottom of Figure 23). In the following paragraphs will detail how these techniques function, and what are the advantages and disadvantages in relation to our usage.

A very important issue that we have to understand here, and that will be detailed further on later chapters, is the big question :

- what happen to the sample when it goes from the liquid to the gas phase?
- And the later question, what happens to the sample when it lands on a solid surface, in our case that of the NEMS detector ?

Since these are especially ‘unnatural’ conditions for biological entities, a myriad of physicochemical drastic changes occur that we have to account for when we perform MS experiments. These are primarily related to the desolvation of the sample during transition and fact that we ‘loose visual’ of the integrity of the sample once it lands. We can also see that this are rising matters in the community, by the recent usage of native MS related methods in collaboration to electron microscopy methods, especially for single protein particles. We can mention especially the usage of Ion Beam Deposition (IBD) Low-Energy Electron Holography

(LEEH)[168,169], and more recently to the IBD combination, and native MS comparison to cryoEM.[170]

2.3.3.1 nano Electro Spray Ionisation (nESI)

2.3.3.1.1 The fundamentals of ESI

The most classical, and most commonly used, nebulisation method in native MS is the nano Electrospray Ionisation (nESI) system. The development of this nebulisation technique by John Fenn et al, in 1989 [171], which revolutionised the field of MS for applications in biology.[172] Resulting with Fenn earning the prestigious Nobel Prize in Chemistry, in 2002, and him entitling his Nobel award lecture presentation as "Electrospray Wings for Molecular Elephants."[173]

The main working principle of the ESI (picture seen on the top of Figure 23) is that, by charging the analytes, either negatively or positively (preferred mode), and thus producing a voltage difference between the tip of the needle and the entry of the MS instrument (several kV), a spray is formed, containing nanoscale droplets, that quickly evaporate on the interface, reaching the mass spectrometer.[166] Figure 24 gives the schematic depiction of the principle of this method.

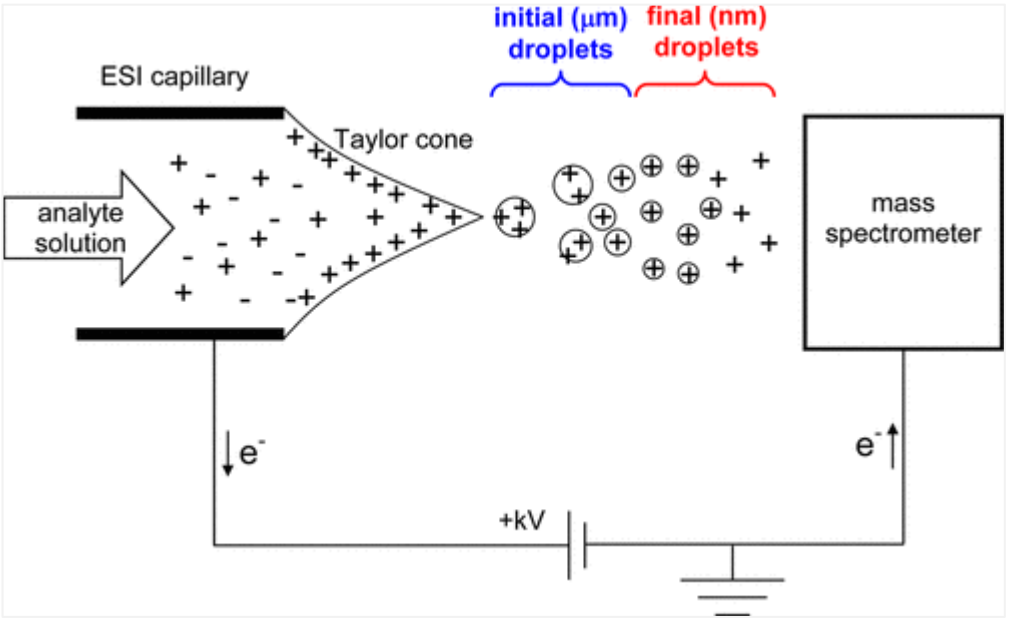


Figure 24. Schematic depiction of an ESI source operated in positive ion mode. Figure and description were reprinted from Konermann, et al.[174]

There are three main hypotheses as to what happens with the analyte during nebulisation by ESI, when the droplet desolvation occurs, leading to the dried analyte (under the red parenthesis in Figure 24). These main hypotheses can be seen presented in Figure 25.

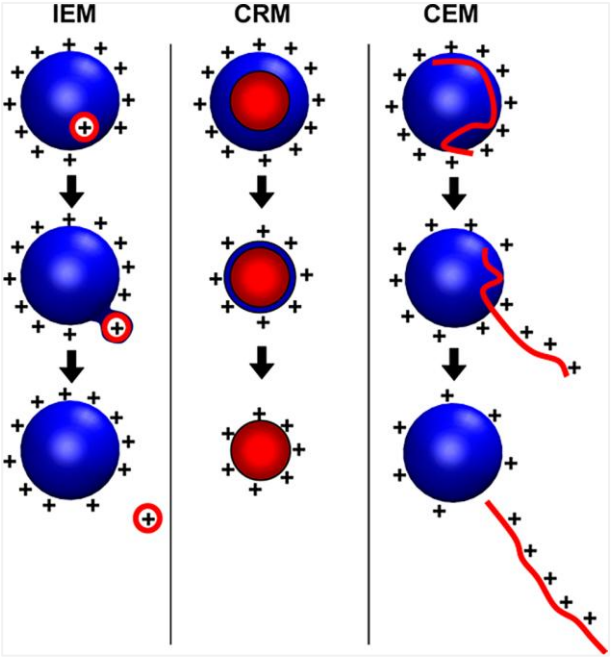


Figure 25. Hypothesis on the main ion release mechanisms for ESI. (a) IEM (ion evaporation model): Small ion ejection from a charged nanodroplet. (b) CRM (charged residue model): Release of a globular protein into the gas phase. (c) CEM (chain ejection model): Ejection of an unfolded protein. Figure and description were reprinted from Konermann, et al.[174]

The first hypothesis, named Ion Evaporation Model (IEM), describes that nebulisation occurs when the small charged molecule is transported as single ion to the surface of the droplet, and when the Coulomb force is high enough for it to be ejected out of the droplet, (left panel of Figure 25). In the middle panel of Figure 25, we see the Charged Residue Model (CRM), describing that as the droplet shrinks down, due to solvent desolvation, the charging on the ion gets closer and closer to the analyte, thus resulting in enveloping its total surface charge, without any solvent remaining on its surface. This is expected to be the case for globular proteins, and most large macromolecules, in the gas phase. In the right panel of Figure 25, the Chain Ejection Model (CEM) is shown, in this case for unfolded proteins, or large polymers. It can be seen that as the charges are distributed on the surface of the polymer, once this polymer breaks through the droplet surface, it is then ejected out of it and takes a portion of the charges with itself.

So now let us see how ESI has played a major role in Native Mass Spectrometry and what makes it so relevant.

2.3.3.1.2 ESI and its usage in native MS

In a comprehensive, and the appropriately named review "Native Mass Spectrometry: What is in the Name?", Leney and Heck[18] described how the evolution of ESI-MS instrumentation has gone hand in hand with the development of biomolecular mass spectrometry. Thus, playing a major role in the field of structural biology, and elucidating many

structure–function relationships, while retaining them in their native state in the gas phase. They presented some of the major milestones of the timeline of native ESI-MS, as seen in Figure 26.[18]

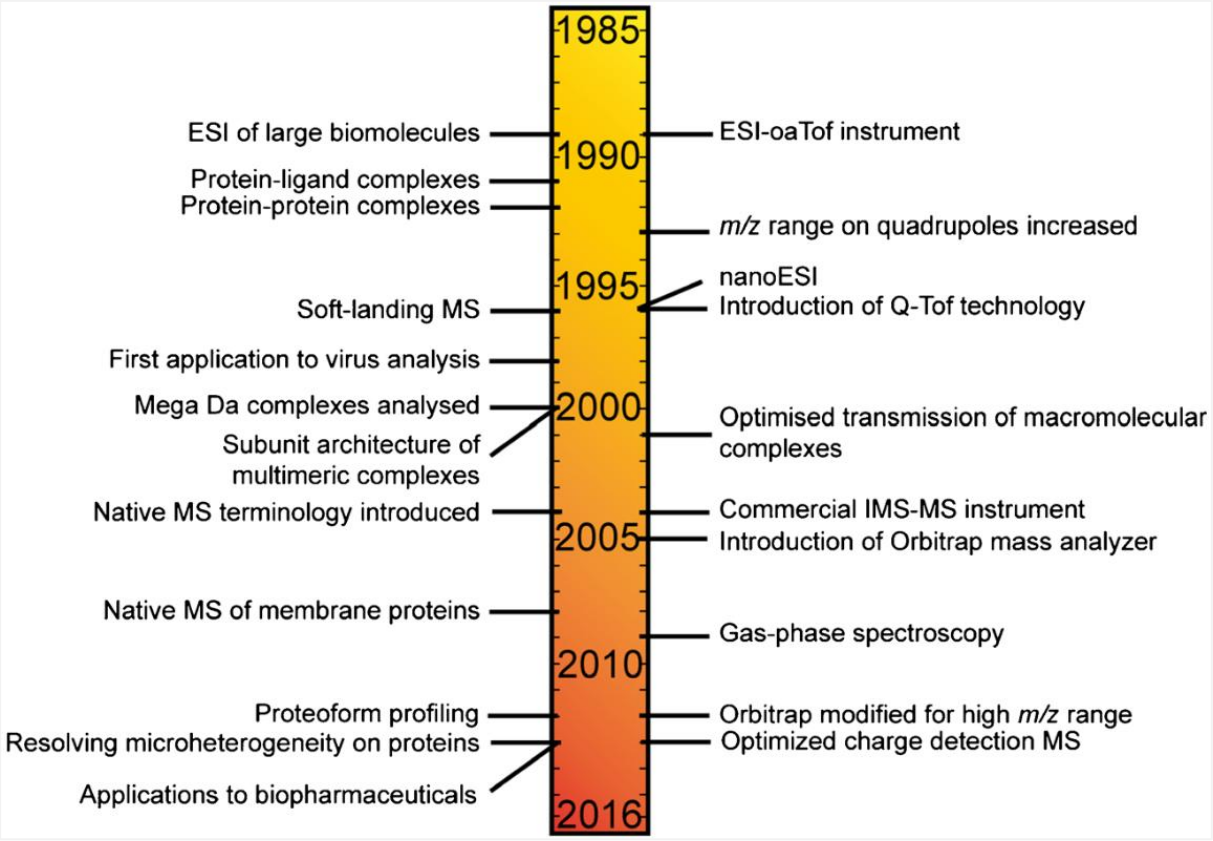


Figure 26. Timeline and origin of native MS, over three decades. On the right: Key instrumental developments; On the left: biological applications. Figure was reprinted from A. C. Leney and A. J. R. Heck.[18]

Some works that are of key importance to be mentioned here are the following. In 1996, soft-landing MS analysis of the tobacco mosaic virus (TMV) were performed by Siuzdak et al.[175], indicating that the usage of ESI on viral samples maintained their quaternary structure and even their biological functions. They extended on this work in 2004,[176] by analysing samples from the rice yellow mottle virus (RYMV), cowpea mosaic virus (CPMV), bacteriophage MS2, and the human adenovirus (HAdV), and observed the structures of these viruses in the gas phase after ESI nebulisation. They noticed some size shrinking for the CPMV particles, giving indications of the desolvation of these particles in the gas phase. A 2024 study also showed that the ESI method could be introduced in cryoEM pipelines, in order to deposit samples in a more efficient manner on the grid, reducing dominant orientation issues, prior to freezing.[177]

But although much development has been made in structural biology thanks to ESI, its understanding still remains difficult at the molecular scale.

2.3.3.1.3 The difficulties of working with ESI

As it was shown in Figure 24, ESI is a fairly simple technique to conceptualise, but a difficult one to properly understand its functioning.

Firstly, at the droplet formation stage, where the Taylor cone ends, and the initial droplets are generated, at the micrometre scale (under the blue parenthesis in Figure 24). It has to be noted here that since there is a difference in potential, there is proton generation from the water molecules, which affects the local pH of these droplets. This is of major importance when choosing the correct solution is which we can perform native MS experiments. In addition, these micrometre droplets are also submitted to a high temperature at the entry of the bench, which fastens the desolvation process on their surface, which again increases the local proton content, reaching nanometric droplet (under the red parenthesis in Figure 24). This results in a build-up of charge around the analyte, until the point where the Coulomb force applied by them is balanced out by the surface tension of the droplets. This results in the so-called 'Rayleigh limit', where the charge density is defined by z_R, indicating the number of elementary charges, denoted e, on the surface, resulting in the following formula:

$$z_R = \frac{8\pi}{e}\sqrt{\epsilon_0 \gamma R^3}$$

where R is the droplet radius, ε_0 is the vacuum permittivity, and γ is the surface tension.[174] This is a critical parameter to understand for ESI and mass evaluation, because it is this elementary charge number (z_R) that is one of the fundamental pillars of mass determination by CDMS methods. All CDMS techniques, independently of the instrumental engineering, rely on the precise determination of this charge number, in order to measure the mass, by subtracting it from their m/z measured values.

One of the major focuses of our work (and since we can perform neutral mass measurements) has thus been to shift away from the usage of the classical Electro Spray Ionisation (ESI) method, to form aerosols at the bench entry. Therefore, we have tried different techniques that do not require ionisation, are less prone to clogging, or to causing sample aggregation, and can use native, not necessarily volatile buffers, to reach sample nebulisation. One such method that has gained more and more usage in the field of MS is the Surface Acoustic Wave Nebulisation (SAWN).

2.3.3.2 *Surface Acoustic Wave Nebulisation (SAWN)*

SAWN (Surface Acoustic Wave Nebulisation) is a technique that can produce aerosols that are independent of charge. With SAWN, a liquid sample deposited on a piezoelectric surface is nebulised as a thin mist upon propagation of an acoustic wave across the surface, (Figure 27 on the right).[163] This method has already being used in MS proteomics [178], and also previous investigations from the group had been tried, but relevant spectra could not be acquired, which was attributed to previously reported aggregation issues.[21]

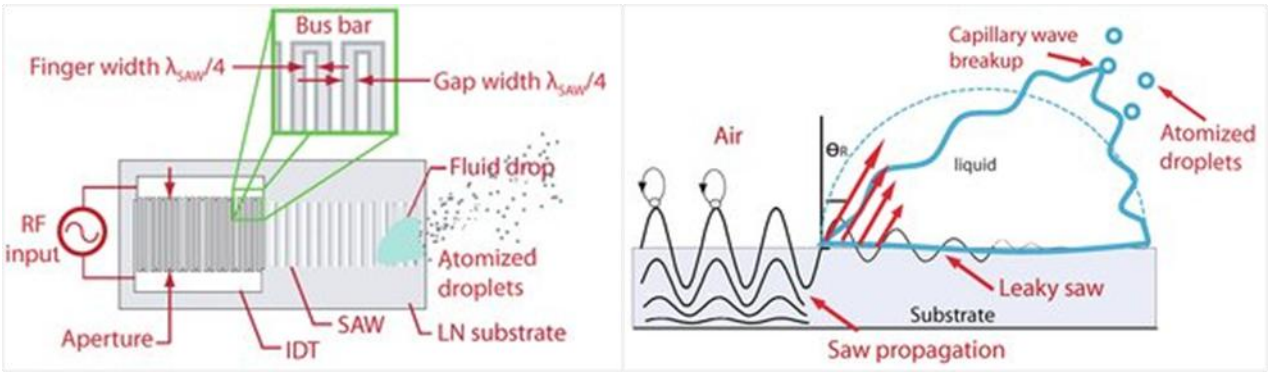


Figure 27. SAWN principle. (Left) Schematic of the SAW device showing the interdigital transducers (IDT) electrodes patterned on the piezoelectric (LiNbO 3) substrate. A radiofrequency (RF) power source supplies the atomizer with a sinusoidal electric signal via the IDT which excites the SAW on the substrate. (Right) Interaction of the SAW with a liquid drop placed on the substrate as shown in the image on the left. The SAW energy leaks into the drop and induces capillary waves on its free surface. Due to the large energy transfer, the interface rapidly destabilizes and results in the atomization of the drop through break-up of the capillary waves. Figure and description were taken from Aisha Qi et al. [163]

That is why we collaborated with colleagues at IFW Dresden (The Leibniz Institute for Solid State and Materials Research, Dresden, Germany), in order to use an improved model of SAWN, with integrated microfluidics, as shown in Figure 28, and reported for optimised MS usage in the work of Kiontke, et al.[179] During my thesis we tried to implement a more compact SAWN system, as the one visualized in Figure 28.

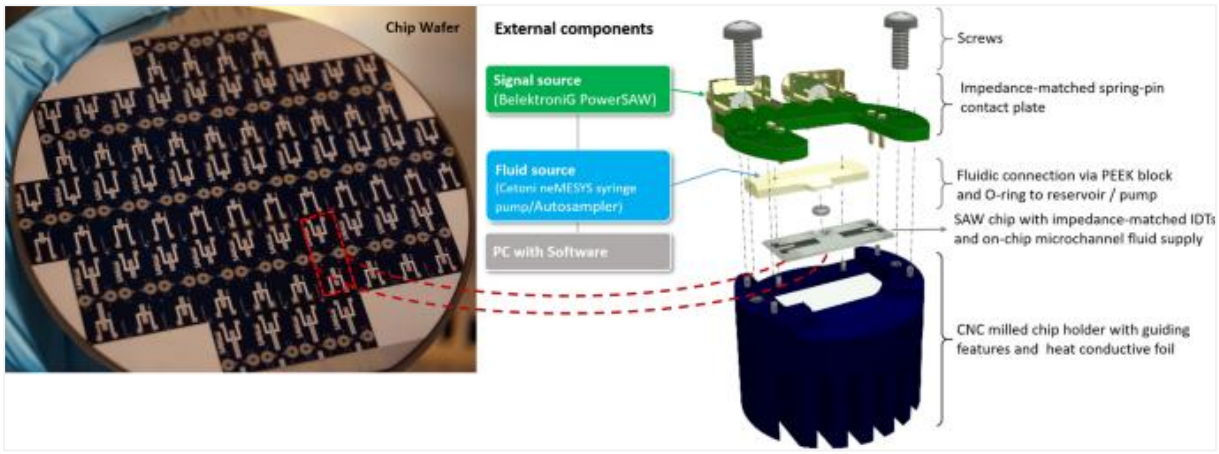


Figure 28. Schematic of the compact SAW aerosol generator and photograph of a wafer with the used SAWN chips (before dicing). Figure and description were taken from Kiontke, et al.[179]

This compact SAW aerosol generator operates without macroscopically moving parts or nozzles (Figure 28 on the right). The SAWN works by disintegrating an acoustically stabilized liquid film, which is in the same order of thickness as the SAW wavelength, thus forming μm-sized droplets. These last can be optimized for a given liquid by adjusting the applied SAW wavelength. For this IDT (Interdigital Transducers) design we have a finger width of λ/4, for a wavelength of 90 μm; using 46 electrode pairs, with a 0.5 mm aperture, matched to 50 Ω impedance by the number of finger electrodes.[179] Each IDT receives an electrical forward power of 1.9 W signal at its high operation frequency, thanks to SMA (SubMiniature version A) connecting cables, from a dual-channel PowerSAW signal source (PSG, BelektroniG GmbH, Germany).[179] This does cause the chip to reach high temperatures after a prolonged usage, nearing 60-70 °C, and newer systems are being developed by our colleagues at IFW Dresden that can cool down and control the temperature of the chip during usage.

In addition, and of critical importance, these devices have the capacity of controlling the flow rate introduced on the SAWN chip, which makes possible to define the nebulised volume at any given time, thus making concentration control much easier compared to our home-made sample introduction mechanism. In addition, the outflow capillary from the sampler is connected to the microchannels via sealing rings.[179] Lastly, is the need to closely investigate, prior to starting the nebulisation, the electrical Radio Frequency (RF) behaviour of the chip, by observing the "resonance curves" of the chip, which are based on its complex scattering parameters in a frequency range around the excitation frequency.

In order to see how the SAWN compares to the ESI, we usually can perform a transmission efficiency test. This test is the ratio of the number of particles on the target to the number of particles in the nebulised solution, accounting for the initial solution concentration, the liquid flow rate and the deposition time.

2.3.3.3 Comparing transmission efficiency

The NEMS-MS group showed in the past [21], that since SAWN produces droplets with lower kinetic energy, it can result in a more efficient uptake into the NEMS-MS bench, when compared to the nESI system, as the values indicate in Table 2.

Table 2. Comparison of transmission efficiency for polystyrene nanoparticles of different size and mass with both SAWN and ESI. Table was adapted from the supplementary information of Dominguez-Medina et al.[21]

Sample	SAWN transmission efficiency (%)	ESI transmission efficiency (%)
46 nm (NIST polystyrene NPs)	5.1	0.4
100 nm (polystyrene NPs)	1.2	0.1

We can notice here that the transmission efficiency for SAWN is more than an order of magnitude higher than that of nESI, shown to be true for two types of polystyrene nanoparticles, 46 nm and 100 nm, which were similar in size to viral particles. In order to reach such an evaluation of transmission efficiency, we need to make sure that as many particles as possible reach the detectors, and the chosen system to perform this task is called an aerodynamic lens, which we will detail in the following subchapter.

2.3.4 Aerodynamic lens

Subsequent research on NEMS-MS focused on simplifying the measurement and the mass spectrometer's architecture, as well as developing more effective methods to focus the analyte on the NEMS mass analyser/detector. A significant milestone was independently proposed by Hentz and Masselon,[180] and by Malvar et al.[181] They proposed NEMS-MS systems devoid of ion guides, allowing them to create prototypes with somewhat relaxed pumping requirements. These systems consisted of series of chambers with decreasing pressures: a nebulisation stage operating at ambient pressure, a heated capillary inlet with a first pressure drop to 10 – 100 mbar, a pressure-limiting orifice that could optionally be followed by a series of focusing orifices (aerodynamic lens), and a resonator chamber operating in the 10^{-3} to 10^{-5} mbar regime. A schematic of the principle of an aerodynamic lens is shown in Figure 29.

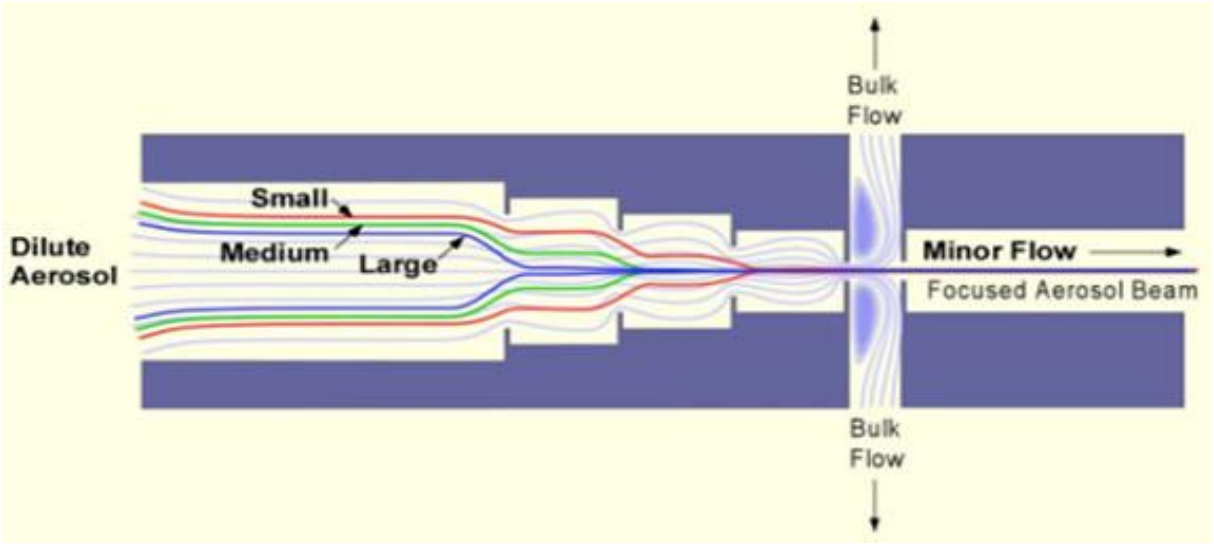


Figure 29. Schematic design of a classical aerodynamic lens with bulk flow skimmer. Reprinted from © Biotechnix

The idea behind an aerodynamic lens is conceptually simple, as we want to go from a dilute aerosol, composed of a particle of varying size, into a focused aerosol beam (going from the left to the right of Figure 29). The pressure gradient formed inside the instrument (indicated by the shrinking rectangles in the centre of Figure 29) forces the particles to follow a laminar

flow (light blue lines), ever closer to the centre. This thus results in the particles being focused, starting from the largest (more sensitive to the drag force), shown in dark blue, to the smallest (more sensitive to Brownian motion), shown by the red line. The presence of a skimmer is needed, to remove any bulk air flow that is of excess and that could induce turbulent flow in the system, resulting in the final minor flow containing all our particles of interest.

However, the prototypes were still characterized by modest particle capture efficiency defined as the ratio of the number of detected over emitted particles. In the work by Malvar et al. [181], where 1 particle per 5^{e}8 was detected, with event rates of ~0.3 particle per minute. In the NEMS-MS group, Dominguez et al.[182] thus reported the detection of 1 viral particle per 2.6^{e}8, with an event rate of ~0.8 to 1.35 particle per minute. This was achieved using a 20 resonators NEMS array, which we will detail in the following subchapter.

2.3.5 NEMS and the Neutral mass measurement

Due to the charge independence in NEMS-MS, it is insensitive to mass to charge convolution generated by similar species with differing charge states, and thus particles could be focused by using the particle inertia, and thus the mass is measured directly. In addition, since the NEMS works as both detector and analyser, it was possible for a simplified architecture to increase in detection area without requiring an increase in a large number of other components. This allowed the team to produce a detector with an array of 20 NEMS detectors, as presented in Figure 30a.[161] This capacity to perform neutral particle mass evaluation is of great importance when it comes to the different methods that we can use to nebulise the particles in the gas phase, as described earlier.

2.3.5.1 The NEMS array

In 2013, Sage *et al.* reported on the development of NEMS arrays, interconnected with a dedicated readout scheme in order to probe each device sequentially.[183] This had the purpose of multiplying the capture area of small, but highly sensitive mass detectors. The idea used was to control the device by reduce the number of electrical connections, maintaining the same input/output port ratio as a single sensor. Meaning that the voltage at input was applied to the whole array, and the output signal was recovered as the sum of individual resonators. The signal deconvolution on output was possible thanks to characteristic "frequency addressing" for each NEMS, seen in Figure 30 **e** and **f**, where variations in the sensor's beam length gave discernible resonance frequencies for each resonator, allowing for individual identification and monitoring. Another major advantage that the authors noted regarding this approach, was that operating

devices within the array did not deteriorate mass resolution.[183] We will discuss more in detail about this contribution in the following section, cf 'Mass-independent resolution.'.

One can imagine NEMS "frequency addressing" is like playing the guitar: the shorter the string, so the closer to the sound hole the finger pinches the strings, the faster they will move, so the higher the pitch will be, and vice versa. Concretely, this means that the longer beams, having higher mass, display lower resonance frequencies.[21,184] So having this in mind, the range of NEMS length, described in Table 3. This design allowed for the final design seen in Figure 30a.

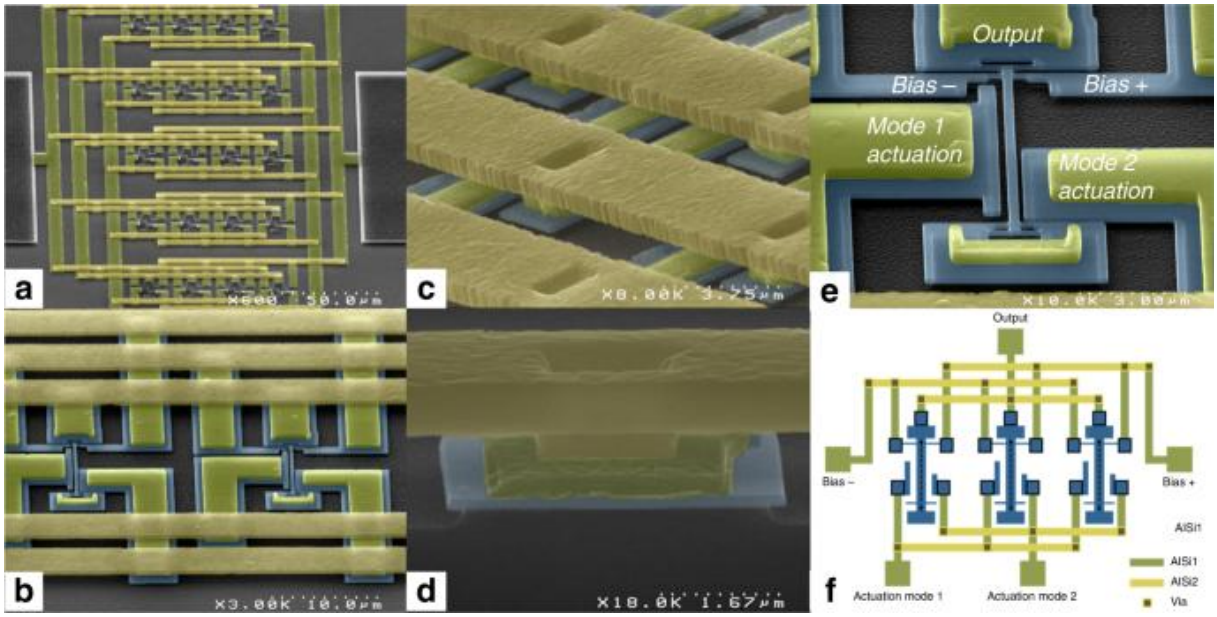


Figure 30. Array of nanomechanical resonators. SEM images of 5×4 NEMS array used for nanomechanical mass spectrometry. Typical horizontal and vertical pitches are 20 and 55 µm respectively. ***a*** *General view of the array,* ***b*** *zoom on two resonators (silicon is false-colored in deep blue), and their metal interconnects (AlSi).* ***c***, ***d*** *zoom-in on interconnects and via. The first metal level is colored in deep yellow, the second one in light yellow.* ***e*** *Typical doubly clamped in-plane resonator used in this study. The beams are designed to resonate around 30 MHz for mode 1 and 80 MHz for mode 2. Typical dimensions for the resonant beam are: 160 nm (thickness), 300 nm (width), and 7.5–10 µm (length). In-plane motion transduction is performed using piezoresistive nanogauges in a bridge configuration to allow background cancellation. Electrodes are specifically patterned for efficient mode 1 and mode 2 actuation. For a resonance frequency f_0, bias voltages at $f_0+\Delta f$ are applied to both nanogauges (with 180° dephasing). Tension/compression in the gauges mix their resistance change to obtain a downmixed differential output voltage at Δf, typically around a few 10's of kHz.* ***f*** *Schematic of the interconnect layout. Each resonator has a unique beam length, hence a unique resonance frequency. Figure and description were taken from the original, reprinted from Sage et al.*[161]

The NEMS sensors used in our team are designed as an array, consisting of 20 [4×5] doubly clamped silicon beams. The advantage of this arrays format is that it increases the area of analysis to a surface of 100 µm×250 µm, seen in Figure 30a.[161] In addition, it has the advantage of having NEMS with different length, and thus resonating at unique frequencies different from one another, which can be found detailed in supplementary (Table 18). This uniqueness is made possible during their production, by Very Large Scale Integration (VLSI)[185], and their frequency difference comes by designing them with different lengths and

anchor gauge distances, while all other parameters are kept the same, as seen in Table 18, and schematically represented in Figure 31.

In parenthesis here, I would like to add that the capacity to reach ever higher masses and increasing the detectors surface area might be even easier in the future, thanks to the flexibility of the type of detectors that the NEMS-MS can use, especially with the development of optomechanical resonators[186], which have already been multiplexed, allowing to cover an even larger detection area.[187] But this portion is out of the scope of my personal contributions, and I will not detail it any further in this thesis.

2.3.5.2 Focusing on a single NEMS resonator

Let us now focus on a single NEMS resonator, as schematised in Figure 31, of the aforementioned array (Figure 30), in order to understand the physics behind it, and most importantly, apprehend how we can use these properties in order to measure particle mass.

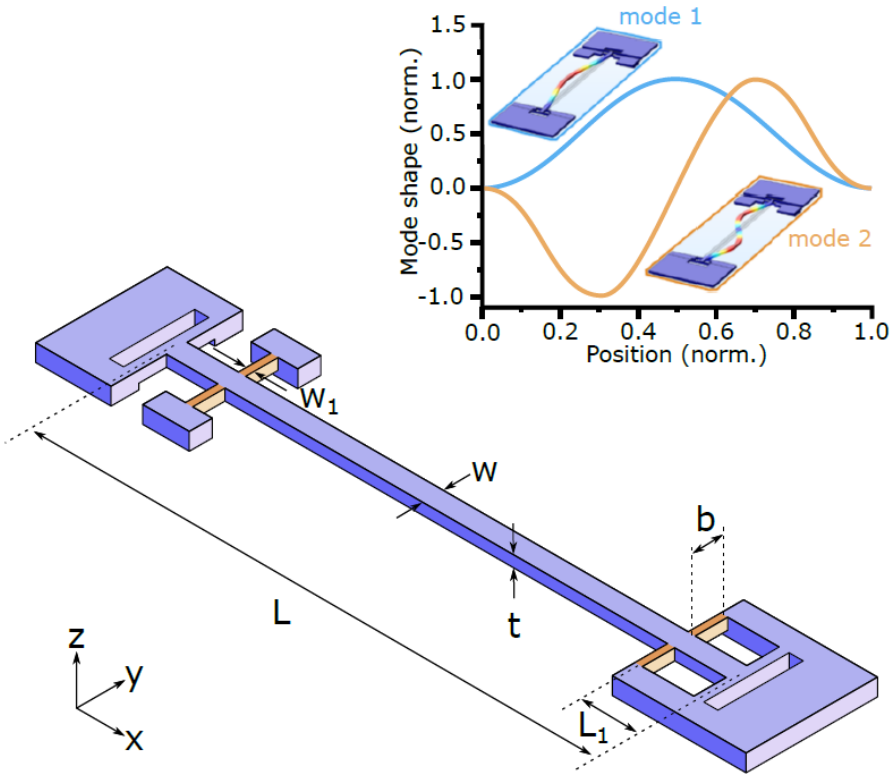


Figure 31. Sketch of the nano-resonator devices discussed in the present work (not to scale). The elongated doubly clamped beam oscillates laterally (in the x/y plane); two piezoresistive nanogauges (in orange, top) placed in a bridge configuration capture these oscillations. The other end of the beam is not electrically contacted but the mechanical design preserves mode symmetry. Inset displays normalized mode shapes used to actuate the beam and derive the particle mass and the position on the sensor. Figure and description were taken from Clement et al [184].

In the inset of Figure 31, we can also see the first two modes in which the beam vibrates, the first in blue and the second in orange. These correspond to the first and second harmonics of the beam, which is presented visually by the relationship between the normalised mode shape

to the normalised position of the beam. The parameters described in Figure 31, are as follows: Beam length (L), Beam width (w), Beam thickness (t), Anchor/gauge distance (L1), Gauge length (b), Gauge width (w1). They take the following values, as in Table 3.

Table 3. Typical dimensions of the nano-resonator devices used in the present study. Ranges pertain to the variable dimensions of individual devices in an array of 20 frequency-addressed resonators. Taken from Clement et al [184].

Beam length (*L*)	Beam width (*w*)	Beam thickness (t)	Anchor/gauge distance (L_1)	Gauge length (*b*)	Gauge width (*w*1)
7.61-10 µm	300 nm	160 nm	305-400 nm	400 nm	80 nm

In has to be noted here that, there are an infinite number of vibration modes, but we only use the first two because they are the most "accessible" with our current NEMS design. Work by the group of Tamayo has shown the interest of integrating a third mode for example in measuring the particle stiffness.[181] In their more recent work, by Sanz-Jiménez et al., they showed that they are able to measure as many as 9 vibration modes, indicating an interesting relationship between bacterial cell mass, and the conservation of their shape and size.[188]

In the work of Clement et al., we also demonstrated that the mass measurement is affected by the stiffness of the landed particle, with a maximal absolute mass correction was ~0.3 MDa in the unlikely case of a particle landing at the edge of both beam and event selection area, but the average value of this effect is ~0.1 MDa.[184] In addition, we also showed that the resonator's mass and residual stress could significantly affect the measurement, which we were able to counter by calibrating the beam. This was performed by relating its frequency to its width and residual stress, and comparing to numerical simulations. The mass distribution was affected by the calibration of the beams masses and the average value increased by 0.9 MDa for a model bacteriophage T5 particle, going from 29.6 MDa to 30.5 MDa. This correction is larger than the NEMS resolution (typically ~0.1 MDa) and should not be neglected when performing mass analysis.[184]

A more detailed explanation on the reasoning on why we need the two first modes is described in the following section.

2.3.5.3 Using two vibration modes

However, when a particle lands in one of these NEMS, collecting information on a single frequency is not enough. This is because, depending on the landing position of the particle on top of the beam, its frequency signature will not be the same. Therefore, we get a system with two unknowns, the mass and the landing position of individual particles. To get

around this issue, we induce the NEMS to vibrate in two different harmonic modes (n = 1, 2), which allows us to resolve the following equation, for these two unknown variables.[184,189]

$$\Delta m = M \frac{\Delta f_n}{f_n} \frac{\alpha_n}{\phi_n^2(x)}$$

Where *Δm*: is the mass of the landed particle; *M* is the total mass of the beam; *Δfn/fn* is the relative change in resonance frequency of the resonator's n^{th} mode of vibration (n = 1, 2); $\varphi_n(x)$ is the mode shape of the beam of the nth mode; and α_n is a constant.[184] Therefore, when the particle reaches any of these resonators, we are able to determine precisely its mass and its landing position.

In addition, as we mentioned earlier, since each NEMS with its unique length has a unique resonance frequency, we can differentiate them from one another, in both modes, as seen in Figure 32.

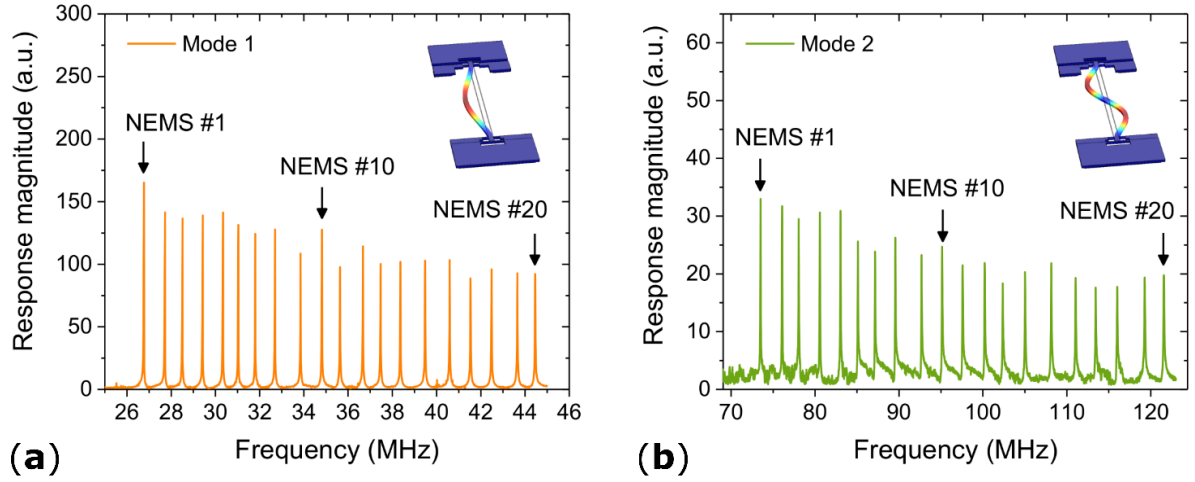


Figure 32. Frequency response of a 20 NEMS array for (a) mode 1; (b) mode 2. Insets of each panel present the mode of vibration obtained using COMSOL simulation and represented with exaggerated amplitudes for the sake of visualization. Each peak corresponds to the resonance of a single NEMS resonator. The length of resonators NEMS #1, NEMS #10, NEMS #20 are 9.04 µm, 8.7 µm and 7.61 µm, respectively. Figure reprinted from Reynaud et al.[165]

In Figure 32 we have the relative amplitude of the NEMS, as a function of their corresponding frequency, in megahertz (MHz), for both modes, in orange for the first (Figure 32a) and in green for the second (Figure 32b). The NEMS #1, 10 and 20, are highlighted in order to better visualise the spreading of the NEMS used.

But of course, the term analytical measurement has an underling term, which is that of measurement resolution. Let us now see how this term is implied when we discuss about mass measurement by NEMS resonators.

2.3.5.4 *Frequency noise*

Before we reach the definition of the resolution in our NEMS, let us first discuss briefly the concept of frequency noise. Every physical measurement is naturally affected by external phenomena that induce uncertainty in the measurement. In the case of the NEMS-MS, these external phenomena are essentially additive Gaussian random noise affecting the two resonance frequencies simultaneously. This means that the system is stochastic, represented by a normal (Gaussian) distribution, (in purple in Figure 33). In a perfect scenario, the main axes of the ellipse are aligned with the x and y axes: This corresponds to the case where noise in the two modes is uncorrelated. In reality, there is correlation between the modes (cf. seen also in panel 1 of Figure 35), in our case it is 'pink noise' (also known as 1/frequency noise)[189], primarily coming from the physics of the NEMS and the electronics that drive them.

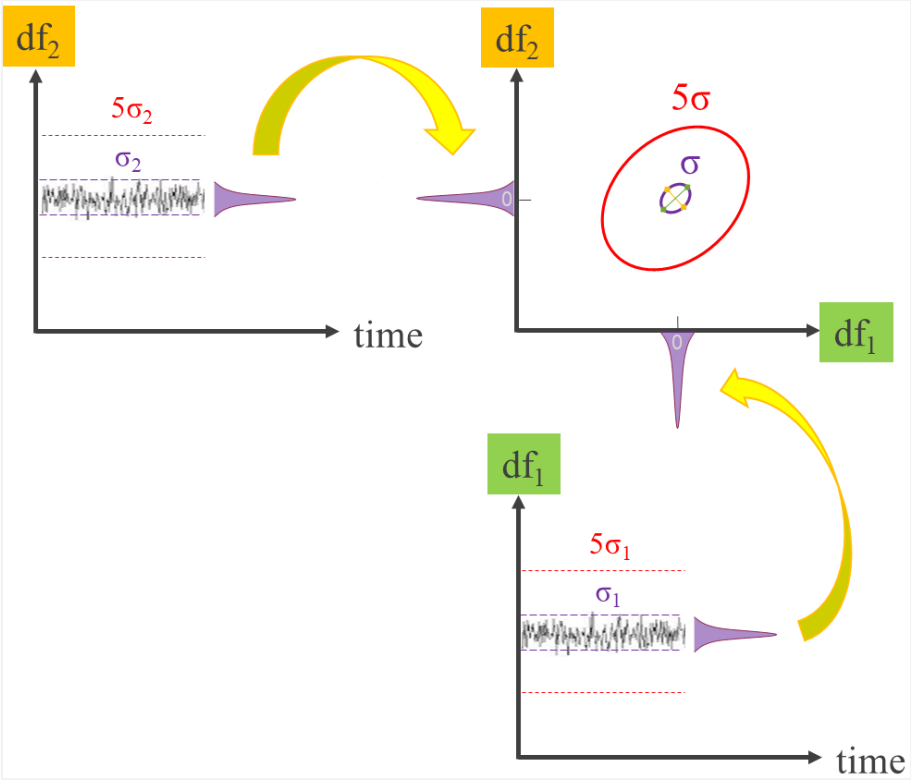


Figure 33. Construction of the noise ellipse (top right). Difference in frequency: mode 1 (in green; bottom); mode 2 (in orange; left). The Gaussian of the signal (in purple).

As a result, the 'true' resonance frequencies are only known to within a certain amount of noise and therefore, by extension, to within a certain level of uncertainty. In the frequency plane, the 'true' resonance frequencies lie within a perimeter around the measured frequencies, defined by this level of uncertainty. The elliptical shape of the uncertainty domain stems from the fact that the main natural NEMS perturbations are Gaussian noises at both frequencies, and that these two-frequency perturbations are therefore characterized by elliptical iso-values. In

fact, estimating the resolution of a device comes down to comparing each of the results given by the device without disturbance, with the same results given by the device with natural disturbances: the delta between the measurements is an indicator of the resolution for each result. These disturbances can take on different values but are always small compared to the undisturbed signal. Thus, we have several amplitudes of disturbance, represented by the iso-value lines (cf. Figure 34). Consequently, we can build a detection threshold region around it (in our case 5σ), defining the LOD (Limit of Detection; in red).

Now let us see how these frequency disturbances are used to calculate the mass resolution.

2.3.5.5 Mass-independent resolution

One of the most important contributions of the 2012 work by Hannay et al.[159] was the evaluation of the dependence of the mass resolution to the landing position of the analyte, on the doubly clamped nanoresonator oscillating in same-plane modes. This is possible since frequency fluctuations in the two modes translate into uncertainties in both mass and position. Therefore, an error model was devised to compute the Joint Probability Density Function of particle mass and position, relative to frequency noise, thereby defining a mass resolution for individual particle landing events.

In Figure 34, thanks to the symmetry by design of the doubly clamped nanoresonators, we simplify the measurement to only half of the normalized portion of the adsorption positions on the length of the beam (denoted as a), and thus from a = [0, 0.5] (solid black lines crossing from the origin in Figure 34). The two resonance frequencies (f_1 & f_2) and their respective frequency shifts (df_1 & df_2) are measured, to reach to the fractional frequency jumps per mode, (df_1/f_1 & df_2/f_2). This allows to draw the 'mass contours' (quasi-elliptical coloured curves in Figure 34), which show the ratio of m_p/M (particle mass over beam mass), for a constant and uniform analyte deposition, evaluated in parts per million (ppm).

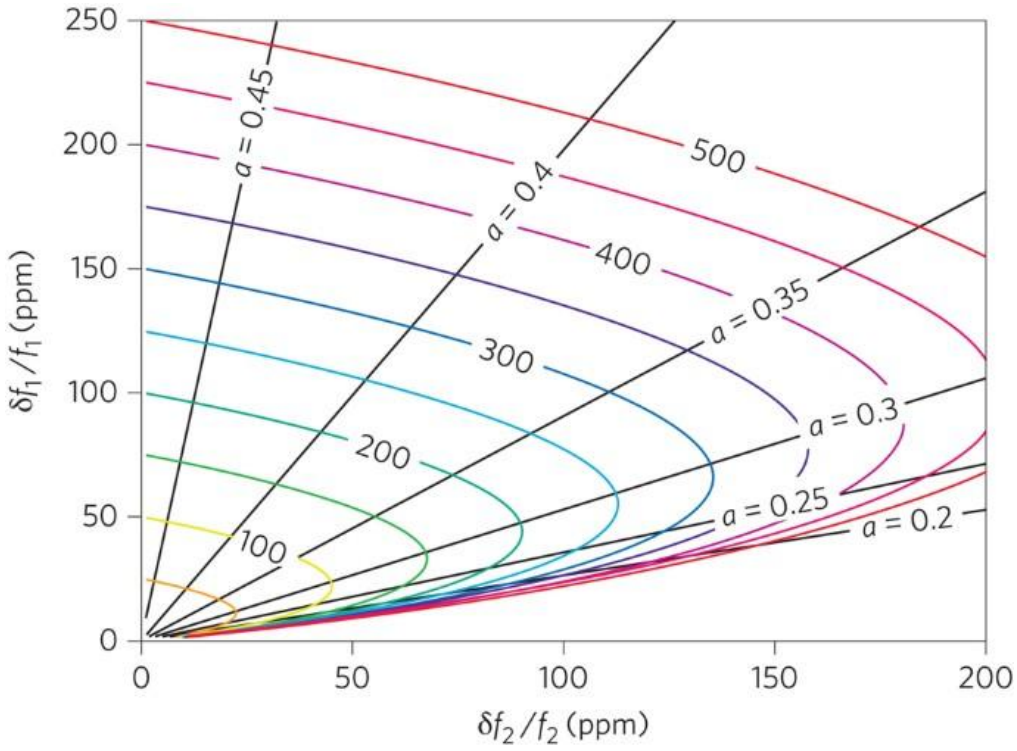


Figure 34. Determining analyte mass and position of adsorption from time correlated two-mode frequency jump data. Universal mass and position contours. The y- and x-axes represent the fractional frequency jumps, |df1/f1| and |df2/f2|, measured for modes 1 and 2, scaled in parts per million (ppm). The parametric curves represent the following. Straight black lines denote constant values of adsorption position, where a is the normalized position on the resonator and a ¼ 0 and 0.5 (both not shown) represent one end of the resonator and the centre of the resonator, respectively. The elliptical curves represent contours for constant adsorbate mass and are labelled in units of mp/M and scaled in ppm. These parametric curves are valid for the first two same-plane modes of a doubly-clamped beam, and assume only Euler–Bernoulli beam theory. Figure and description were taken from Hanay et al.[190]

This figure shows that the area that is the most resolute in the doubly clamped NEMS conformation is the range 0.25 < a < 0.50, indicated by the solid black lines. We can see that the closer the analyte lands to the centre of the beam (meaning at a = 0.5), the more resolved the mass will be, as seen by the ever larger separation of the 'mass contours' lines. Which shows that a major limitation of performing highly resolute mass evaluation is the landing position dependence.

From a more analytical point of view, the mass resolution σ_m depends on the resonator geometry and the frequency fluctuations. It can be defined as:

$$\sigma_m = 2M_{eff}\sigma_{\delta f}$$

Where M_{eff} is the effective mass of the resonator and $\sigma_{\delta f}$ is the frequency fluctuation, measured using the Allan deviation.[191] This parameter is the square root of the Allan variance, which is a type of two-sample variance, which evaluates the frequency stability of oscillators, such as our NEMS for example. Our collaborators at CEA-LETI, who designed the NEMS detectors, we are currently using, have evaluated extensively the type of noise and the physical effects that can affect the resonators frequency, and ultimately the resolution of the mass

measurement, described in detail in the work of Sansa et al.[192] Taking into account landing positions effects, we can define a reliable sensing interval: in practice, only the regions situated in the interval $a = [0.25 - 0.47]$, and by symmetry $a = [0.53 - 0.75]$, along the beam should be taken into account for particle mass determination.[184]

By combining all these information, we get the following workflow for mass resolution evaluation in our system, going from the frequency signal to the final mass resolution, detailed in Figure 35.

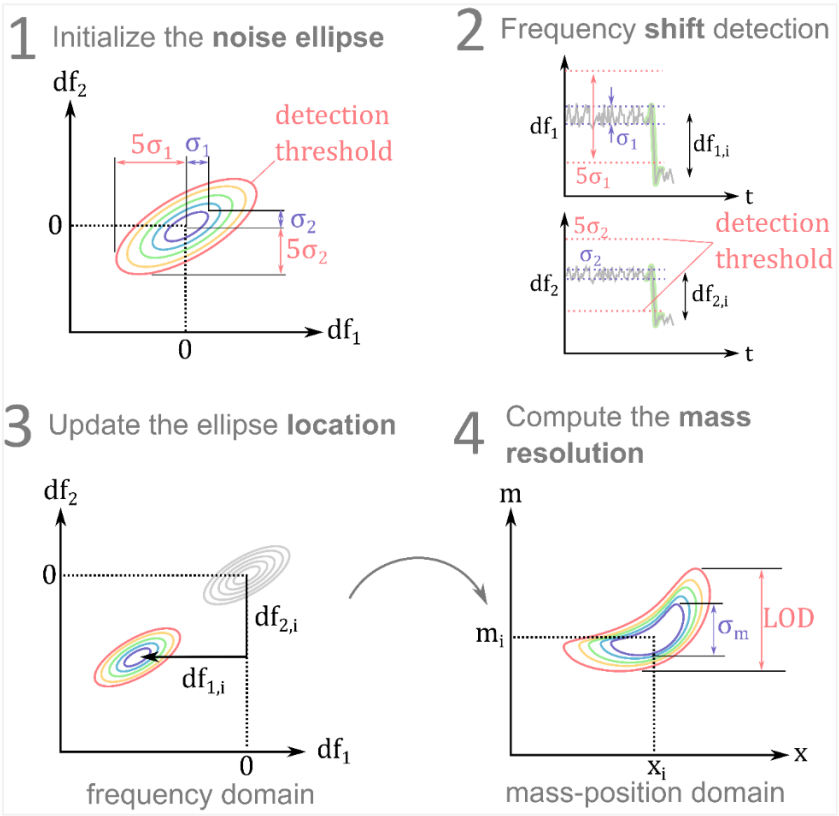


Figure 35. Simplified data process workflow to extract the mass resolution and the limit of detection (LOD) from one frequency jump. (1) The noise is characterized by evaluating σ_1 and σ_2, leading to the noise ellipse. (2) The whole frequency trace is swept to detect the shifts corresponding to landing events and the mass resolution is computed by (3) updating the resonance frequency and (4) extracting the extrema values of the noise ellipsoid in the mass-position domain. Figure and description were taken from the supplementary data of Clement et al. [184]

The uncertainties discussed above (cf. Figure 33) on the resonance frequencies are naturally transferred to the mass and position via the transfer formulae presented above. Since these formulae are non-linear, the ellipses are deformed into more complex continuous curves (part 4 in Figure 35). The mass and position are therefore also only known to within a level of uncertainty. The LOD then simply corresponds to a particular level of uncertainty (generally set by the user) and depends on the characteristics of the noise affecting the resonance frequencies. In addition, it has to be noted that depending on the length of the beam, varying from 7.61-10 µm (cf. Table 3), the NEMS' resolution also slightly varies, that is between 0.12

-0.16 MDa for mode 1, and between 0.14 - 0.18 MDa for mode 2, shown more in detail for each NEMS, in supplementary (Table 19).

In order to collect enough data to build a mass spectra, we need to detect frequency shifts over time, as shown in Figure 36.[184]

2.3.5.6 *Frequency shift over time*

In order to reach an overall mass distribution, the relative frequency shift is recorded for both modes over time, seen in Figure 36.[184]

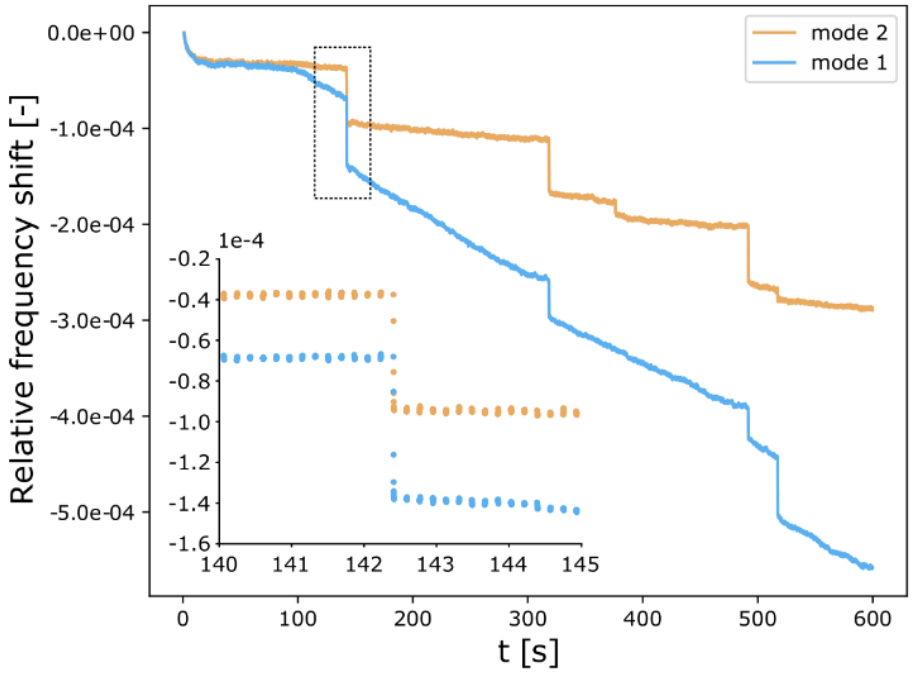


Figure 36. Frequency trace of both modes over time (blue= mode 1; orange= mode 2), after bacteriophage T5 capsids mass measurement. The inset represents the zoom on the dashed rectangle. Figure was reprinted from Clement et al [184].

We can notice that over the course of 600 second, we have several ‘landing events’ on a single NEMS. These shifts are visible in both modes as downward jumps, resulting from the increase in mass of the NEMS itself, after each mass deposition. One such example is seen in the inset.

One thing we can also notice is the seemingly continuous shift in frequency over the entire duration of the experiment, especially visible in the first mode (in blue). This is what we refer as ‘frequency drift’, which can be defined more precisely as a continuous portion of a frequency trace that does not contain a trace discontinuity, that is a ‘landing event’.

I will not detail ‘frequency drifts’ any further in this thesis, but extensive work has been done in our team in order to better understand this phenomenon, which has a significant effect on the final mass evaluation, by focusing on the Total Variance (TV) signal processing method.

More details can be found in the work of Thomas Fortin et al., which I had the honour to contribute as well, initially in a 2021 article[189], and a follow up article, which was very recently published[193].

Let us now see how all these different technicalities come together to perform a structural biology mass measurement, and more specifically that of a virus.

2.4 Proof of concept: bacteriophage T5 Viral Like Particles (VLPs)

All the previously described instrumental developments led to the culminating proof of concept (POC) mass measurement performed by the NEMS-MS group, using T5-Viral Like Particles (VLPs), back in 2018, in the article reported by Dominguez-Medina et al.[21] The main result of this work can be found in Figure 37.

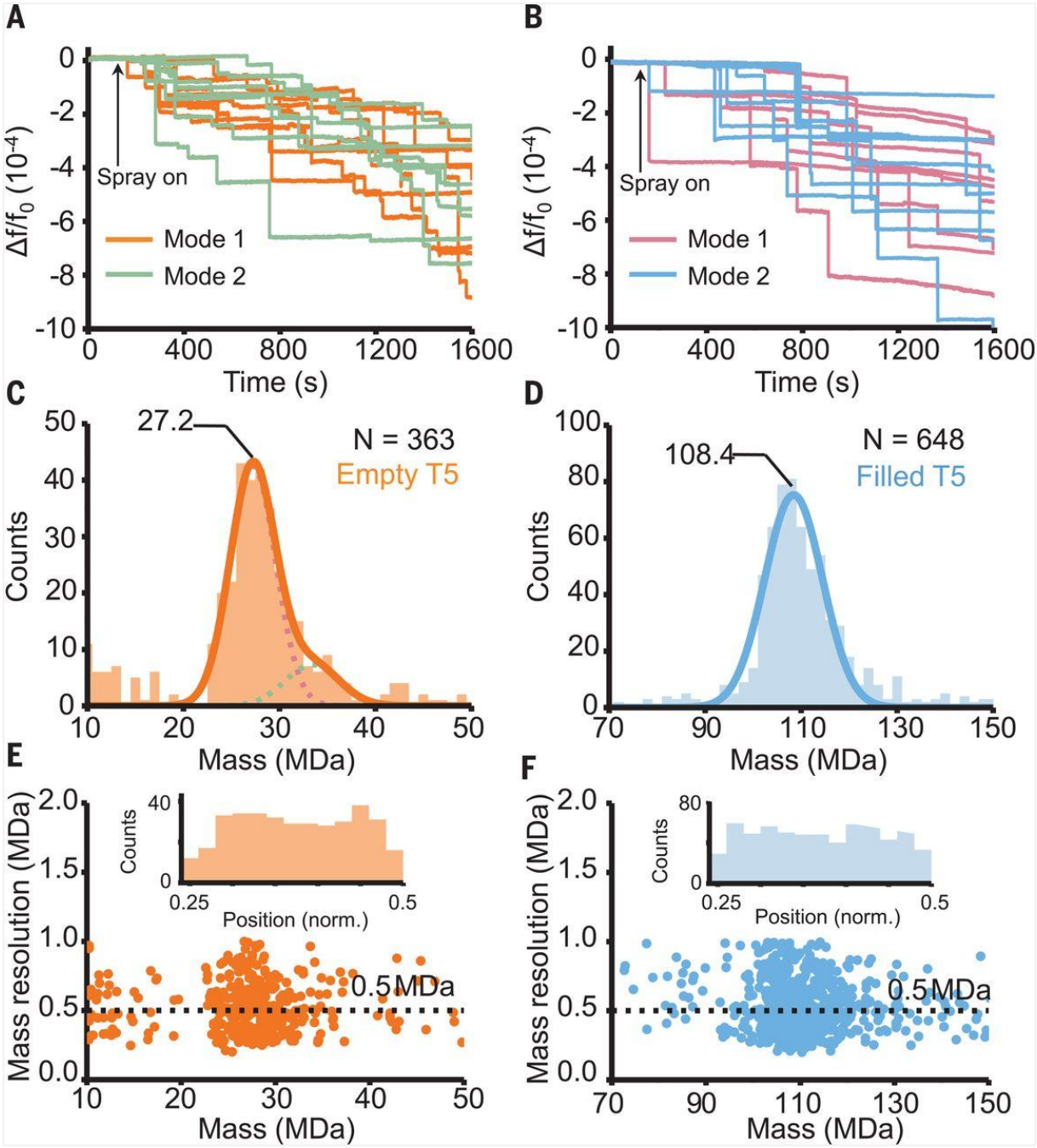


Figure 37. Molecular mass measurement of empty and filled bacteriophage T5 capsids. (A and B) Representative raw frequency traces while spraying empty (A) and filled (B) capsids of bacteriophage T5, shown in fractional frequency change (Δf/f0) with respect to their initial frequency f0. For the sake of readability, the two modes of only eight nanoresonators are shown for 1600 s out of the whole experiment. (C and D) Accumulated mass histograms of 363 empty (C) and 648 filled (D) capsids nebulised using nano-ESI (with 1- and 2-MDa bin size, respectively), fitted to two and single normal distributions, respectively. Mass resolution per particle for empty (E) and filled capsids (F), with their respective particle landing position histograms (insets) in normalized units (beam extremities are at position 0 and centre is at position 0.5). Capsids landing below 0.25 were discarded because of their poor mass resolution. Figure and description were reprinted from the original, Dominguez-Medina et al. [21]

This system demonstrated the characterization of bacteriophage T5 viral capsids devoid of genome (left panels), or containing viral DNA, 121 750 bp long (right panels), in just a few hours. In Figure 37 (A and B), we have the relative frequency shift over time for both modes and for both measurements, that is : in A, mode 1 (in orange) and mode 2 (in green) for empty

capsids; and in B, mode 1 (in pink) and mode 2 (in blue) for filled capsids. Similarly to Figure *36*, we have the landing events obliging the relative frequency to jump downwards, as particles deposit, and this is very obvious when the ESI spray is turned on, as indicated.

For the empty capsids, shown in orange (Figure 37 C), the central mass was at ~27.2 MDa (st. dev. = 2.3 MDa) with the most abundant bin mass at ~26 MDa, which was very close to the expected theoretical mass of 26.02 MDa. A second peak (dotted green distribution) can be seen in panel C, with a mass of ~33.4 MDa (st. dev. = 2.8 MDa). In the original work, this result was attributed to the presence of a remaining short DNA fragment associated to some T5 capsids during purification, and was substantiated by a differential scanning calorimetry experiment, showing two denaturation profiles.[21]

Concerning the genome filled capsid, shown in blue (Figure 37 D), the Gaussian fit gave a mean mass of ~108.4 MDa (st. dev. = 6.0 MDa), and the most abundant bin centre ed at ~107.5 MDa.[21] In this case, this central measured mass was within 2.5 % of the calculated mass, i.e. 105.4 MDa, based on the known protein composition of the capsid and the phage genome sequence.[21]

Lastly, in Figure 37 E & F, the mass resolution for both measurements, and the average value is ~0.5 MDa highlighted by the horizontal dotted line can be seen. We can see that for a mass measurement that is four times higher for the filled capsid compared to the empty capsid, the resolution remains the same, thus demonstrating the mass independent resolution capabilities of the NEMS-MS, especially interesting for higher mass analytes (cf. those calculated in Figure 19). We can also see, in the insets of panels E & F, that the distributions are quite stable over the normalized position range [0.25; 0.5] for all NEMS combined, demonstrating a homogeneous deposition of the particles on top of the resonators.

It has to be highlighted that in 2018, when this work was published, this was the highest mass ever reported by a native MS method to be above 100 MDa, and especially coming from a neutral mass spectrometer.

2.5 Conclusions on Chapter 2

In this chapter I wanted to detail the currently available analytical tools that can serve to measure the mass of whole virions, as summarized in

Table *1*. This showed that the NEMS-MS has its rightful place in this list and could be one of the strongest contenders in helping build a virus mass database in the near future. Primarily because of its large mass range and single particle direct mass evaluation capabilities.

In the rest of the chapter, I detailed how the NEMS-MS method functions, focusing primarily on the one developed by our team, and the main considerations that our team had to make in building this first prototype (prototype #1, seen in Figure 22). Furthermore, I detailed the main reasons why I think that this tool is well adapted to be used as an analytical approach. Starting from the versatility that it provides by having innovative neutral nebulisation methods such as SAWN integrated to it, that have shown to be more efficient in transmitting particles to the NEMS detectors, as demonstrated in Table 2. Allowing for innovative transmission methods such as compacts aerodynamic lenses that help neutral particle transmission. Lastly, and the most important element of the instrument, is the NEMS detectors themselves (Figure 31). It is their capability to be both detectors and analysers, in addition to providing neutral mass measurement, which has allowed for the development of the analytical tool that is our current NEMS-MS instrument. We also meticulously catalogued multiple factors that have to be taken in consideration when performing NEMS-MS measurements, including some particle's extraneous physical properties (size, aspect ratio, stiffness), and the influence of frequency noise and device fabrication defects. We evaluated the magnitude of these effects for a specific implementation of the technology, namely doubly clamped beams oscillating in plane, and for a test case particle: bacteriophage T5 capsid. Overall, a conservative estimate of NEMS-MS mass precision taking into account these effects is ~1 MDa.[149] Lastly its proof of concept was demonstrated in measuring a model virus like particle, the T5 bacteriophage, seen in Figure 37.

In the next chapter I would like to focus on the instrumental developments that we have been able to perform in the last few years, especially focusing of my contributions during my time with the group.

3 Chapter 3: Developments on the NEMS-MS instrumentation as an Analytical Method

3.1 Introduction of Chapter 3

In this chapter, I will focus on our latest work, especially highlighting my contribution to the team. It will focus primarily on how my work has contributed in characterising and improving the analytical capacities of this instrument[149], with a special focus on its usage as a structural biology method, that ultimately led to building a more compact prototype of the first bench, named the 'mini NEMS-MS bench'.[165]



Figure 38. Schematic representation of the structure unfolding of Chapter 3.

Regarding the work on the first NEMS-MS bench prototype (prototype #1), we have focused on mass measurement optimization, by using inert gold nanoparticles in order to benchmark the mass evaluation by the NEMS-MS method. Next, I will talk about the developments that have led to building the second prototype, which we call the 'mini-NEMS-MS bench' (prototype #2), and some of the improvements that we have been able to achieve with this instrument, when compared to the first prototype. Lastly, I will specifically focus on our interest in resolving particle desolvation issues, primarily coming from their nebulisation, transfer from the liquid to the gas phase, and shown how this can complicate virus mass measurements.

3.2 Characterization of the first NEMS-MS bench (prototype #1)

One of the initial goals of the team, was to check whether the NEMS mass spectrometer concept works, and to do this, we used several commercially available Gold Nanoparticles (GNPs) standards. They are one of the most widely used engineered metallic nanoparticles, particularly interesting for biomedical applications, such as sensing, targeted drug delivery,

imaging, photothermal and photodynamic therapy, or a combination of two or three of these applications.[194] These particles are highly interesting for our team as they are commercially available, are well characterised and have standardised parameters. Therefore, theoretical estimates of their mass can be easily deduced from their size and known bulk density. In addition, their masses and concentrations reside in the range of interest of the NEMS-MS, making their use as standards, or sample benchmarks, highly interesting. Sample preparation is also very easy, as a simple (1:1 v/v) dilution with methanol, to facilitate the electrospray nebulisation by nESI, is needed prior to spraying.

In the following sections, I will discuss some of the mass evaluations that we were able to make using these GNPs particles, and how it helped us to acquire robust characterization of the mass measurements with the NEMS-MS bench.

3.2.1 Evaluation on the 15 nm and 30 nm GNPs

This work was primarily reported on the article of Lai et al.[150], in which I contributed by collecting and interpreting the data for the following measurements, but in which I am not a co-author. There we reported the mass evaluation of sphere-like GNPs, having expected dimeters at 15 nm (expected mass = 15–20 MDa) and 30 nm (expected mass = 100–250 MDa).

In Figure 39 we report the particle mass distributions (PMD), Figure 39a, and the converted particle size distribution, Figure 39b, for 15 nm GNPs (in yellow) and for the 30 nm GNPs (in purple). The mass-to-size conversion were performed using the density of gold (19300 kg m^{-3}) and assuming perfect sphericity. The inset in Figure 39a shows the PMDs (as violin plots) of the 15 nm GNPs, conducted under similar operating conditions for four days, repeated on multiple, consecutive or non-consecutive, days, over the timespan of a week. The measurement on day 4 was performed with a different nanoresonator array. Both samples were acquired by BBI^{TM} Solutions (Crumlin, United Kingdom). The measurement were performed with a constant flow rate= 30 µL/hr, and a heated inlet capillary at 175 °C.

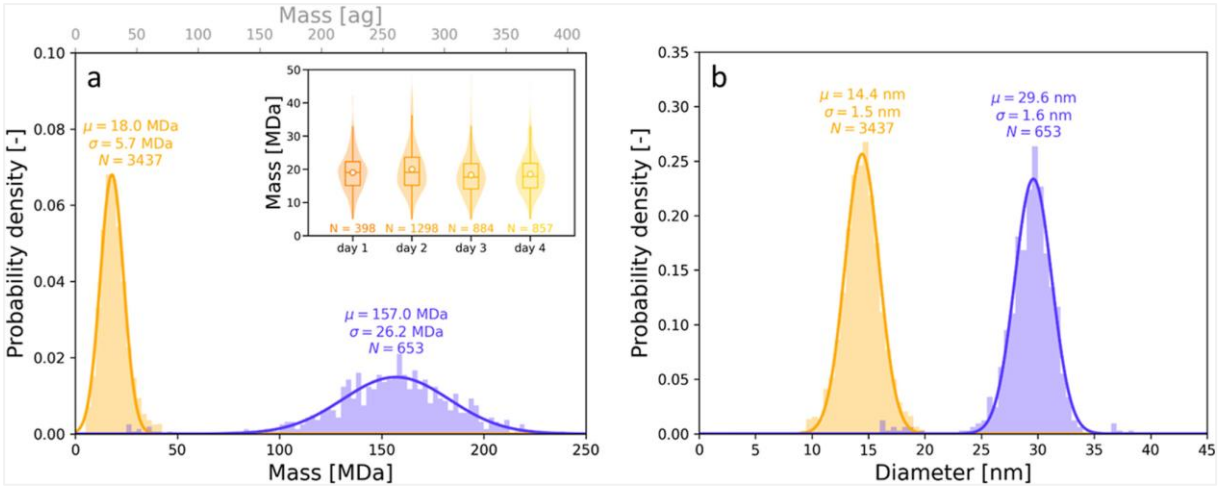


Figure 39. (a) Overlaid particle mass distributions (PMDs) for two commercial GNP (gold nanoparticle) samples. Bottom abscise axis: in megadaltons (MDa); top abscise axis: in attogram (ag). In orange: 15 nm GNPs; in purple: 30 nm GNPs. Inset in (a) shows the boxplots of the results from repeated measurements of the 15 nm gold NP sample on multiple consecutive and non-consecutive days over a week, with two distinct nanoresonator arrays. (b) Overlayed size distributions derived from PMDs, assuming uniform density and perfectly spherical shape. Figure was taken from Lai et al.[150]

The inset in Figure 39a serves to show the stability of the NEMS-MS measurement, independent of day, or the NEMS array used (different on day 4), showing the inter-array comparability. When fitting these four populations of 15 nm GNPs, assuming a Gaussian distribution model (using non-linear least-squares fitting), the calculated variance for the average and standard deviation (between these four populations) did not exceed 1.3 and 0.5 MDa, respectively.

After assembling all the data (of the four days) for 15 nm GNPs in a single PMD, the overall mean mass was 18.0 MDa, [Std Dev = 5.7 MDa] (Figure 39a). This gave a mean diameter of 14.4 nm [Std Dev = 1.5 nm] (Figure 39b). This evaluation is in excellent agreement with the average reported size from the supplier (14.2 nm), thus only 1.4 % larger.

We performed similar mass analyses on 30 nm GNPs, with an expected mass at 151.7 MDa. For these NPs, the average mass determined was 157.0 MDa, [Std Dev = 26.2 MDa] (Figure 39a). This gives an equivalent computed average diameter of 29.6 nm [Std Dev= 1.6 nm] (Figure 39b). Here again, the calculated size is in excellent agreement with the reported size from the supplier, again only being 1.4 % larger.

These measurements proved to show that NEMS-MS is a reliable and suitable method for evaluating intact NPs, and that it can be complementary to refine the results from more classical microscopy methods, such as Scanning Electron Microscopy (SEM) and Transmission Electron Microscopy (TEM).[195] The totality of the discussed data can be found in the supplementary (Table 20).

It is important to highlight that the uncertainty in the nanoparticle size being derived from NEMS measurements varies with density. In this instance, the high density of gold results in a lowering this uncertainty. Our theoretical mass uncertainty, based on the resonator's quality factor and frequency noise level, is ~ 0.5 MDa.[184] The actual mass uncertainty, discussed earlier (inset of Figure 39a), observed over multiple measurements, aligns closely with this theoretical value. These findings demonstrate the stability and reliability of NEMS-MS measurements over several days and across different arrays.

Nevertheless, we realised after the usage of this bench that several limitations prohibited the current instrumental format to become a route analytical method.

3.2.2 Limitations of the prototype #1

One of the major drawbacks of the first bench (prototype #1) was its size and weight, which lie in the range of around 2 meters and around 200 kg, respectively. This made it very difficult to transport and extremely bulky when it came to space it took inside the lab. This large structure also requires large pumps to create the necessary optimal pressure for the NEMS to function (i.e. 10^{-5} Torr), increasing further the weight of the equipment. All these elements make it impossible to perform fieldwork with it. In addition to that, the day-to-day usage and troubleshooting was very complex and delicate, requiring an in-depth understanding of all the elements of the bench from the user, making it not very user friendly to learn and repair.

Furthermore, this large structure also had the intrinsic problem of having a low sample transmission, due to the fact that the distance from the skimmer to the NEMS detectors was equal to 8 cm (cf. Figure 22). This results in a cascade of difficulties, such as:

- a High LOD, thus requiring samples of very high concentration (i.e. 10^{11}-10^{13} particles/mL);
- it takes a long time to obtain enough data to be statistically reliable, given its low deposition rate, resulting in measurements having a long run time
- and a difficulty to align the aerodynamic lens to the NEMS detectors

Based on the previously described literature (cf. detailed in Chapter 2) and our extensive hand-on experience while working with the bench, we hypothesized that several improvements were thus possible, that is:

a) decreased pumping requirements,
b) conserved NEMS resolution in moderate vacuum,
c) superior focusing performances,

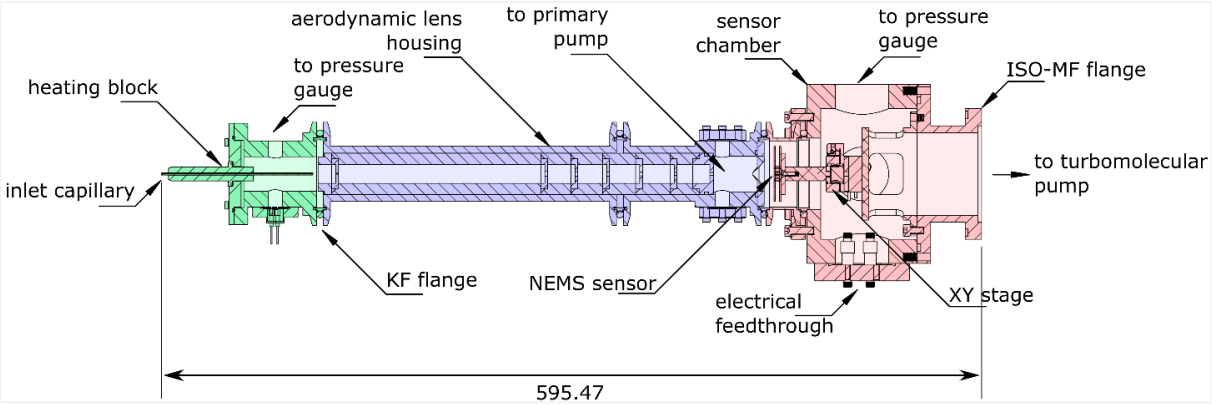
d) improved particle capture efficiency, over prototype #1 (Dominguez-Medina et al., 2018)[21].

Ultimately, we applied these findings to develop a second prototype of the NEMS-MS bench, called the the mini-NEMS-MS bench, which will be described hereafter.

3.3 Prototype #2 : the mini-NEMS-MS bench

As discussed in the previous chapter, the development of the NEMS-MS has relied on progressive improvements on simplifying the measurement and the mass spectrometer's architecture, as well as developing more effective methods to focus the analyte on the NEMS mass analyser/detector.

Ultimately, we applied our findings to develop a new NEMS-MS prototype characterized by decreased pumping requirements, superior focusing performances and improved particle capture efficiency over our previous system (Dominguez-Medina et al., 2018)[21]. In Figure 40, we can visualize the schematic of this new system (top panel) and a real-life photo, as it is currently set-up in our lab (bottom panel), which I helped to build. This work was primarily reported on the article of Adrien Reynaud et al.[165], in which I extensively contributed in building the bench and performed the following experiment in order to characterise it.



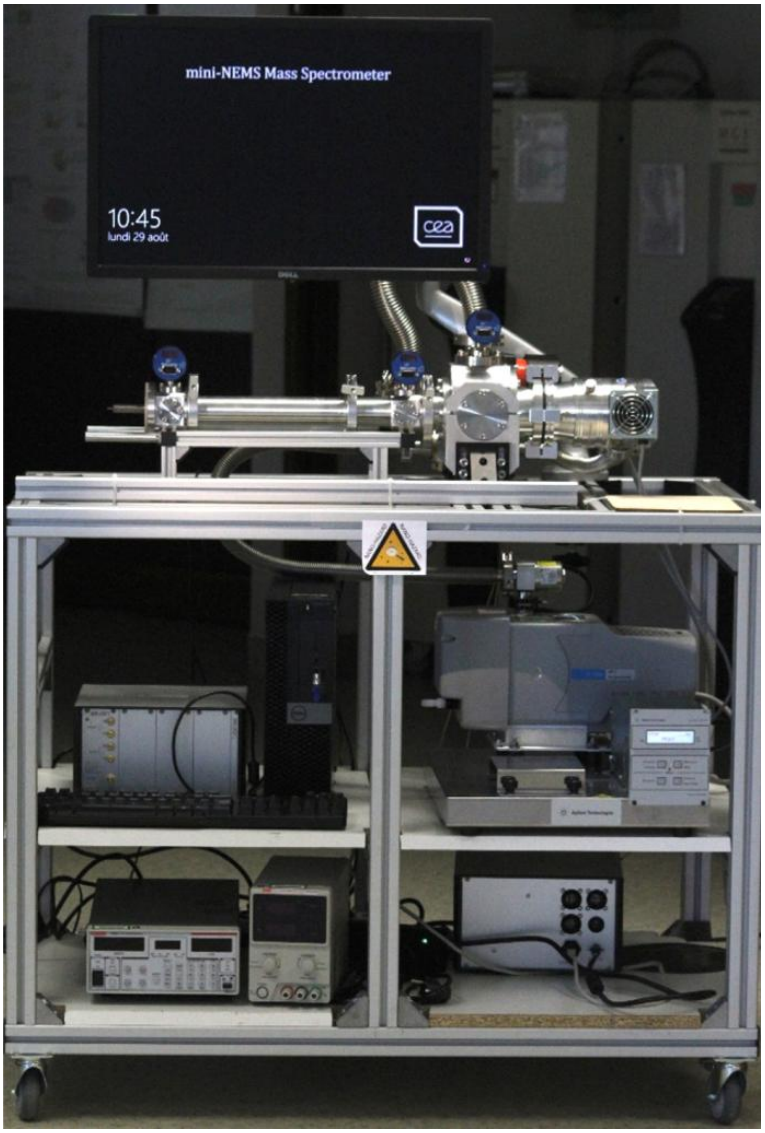


Figure 40. ***Top****: Mechanical compact and modular design of the NEMS-MS prototype. The green parts are the inlet section, the blue parts comprise the aerodynamic lens and the skimmer. The red parts are the sensor chamber, designed to hold the turbo molecular pump and the XY actuated stages. Figure reprinted from Reynaud et al.*[165] ***Bottom****: Photograph of Prototype 2 of the mini-NEMS-MS Instrument, © CEA.*

In the rest of this chapter, I will detail some of the improvements that we were able to make in order to build the instrument (seen in Figure 40), and lastly demonstrate some of the

results that I was able to acquire using model GNPs, as a benchmark for mass measurement and particle transmission efficiency.

3.3.1 Low pumping requirements

Recently, Hannay's group proposed a NEMS-MS system operating under ambient conditions while providing improved particle focusing, reported in Erdogan et al.[196] As a result, they achieved higher capture efficiency detecting 1 particle per 1.85×10^5 for 20 nm gold nanoparticles (GNPs) and 1 particle per 4.97×10^5 for 40 nm GNPs on one device. However, this decreased their NEMS performance and mass resolution.[196,197] This atmospheric evaluation by Hannay's group indicated that we could lower the pumping requirements.

Analysis over a wide pressure range allowed to define the critical architectural design parameters, that is the inlet capillary length and diameter, lens orifice diameters, skimmer diameters. These of course have to be matching NEMS and aerodynamic lens compatible pressure ranges. This model showed that the optimal pressure range, accounting for turbomolecular pumps capacity, steeply degrades below 10^{-3} mbar.

Therefore, the vacuum system and the new pumping strategy was defined to maintain a 0.1 standard liter per minute (SLM) mass flow, and related pressures of 2.7 mbar downstream the lens and 1.3×10^{-2} mbar inside the sensor chamber.[165] This allowed the replacement of the bulky pumping setup in the first prototype[21], i.e. two turbomolecular pumps of speed: 450 l/s, pumping downstream of the aerodynamic lens, and 250 l/s, in the sensor chamber. The replacement was a single smaller turbo-molecular pump (TwisTorr 74FS, Agilent Technologies, Les Ullis, France), pumping downstream the sensor chamber (in pink in Figure 40). A backing primary scroll pump (IDP-3, Agilent Technologies, Les Ullis, France) was also used to pump the chamber downstream of the aerodynamic lens and back up the turbo-molecular pump. One of the main objectives underlying the reduced pumping is to downsize the NEMS-MS bench in order to make it more compact, thus easier to use and to transport.

As mentioned earlier, the work of Hannay showed that it was possible to work with the NEMS in atmospheric pressure, but this method of operating significantly reduced the detector's quality factor.[196] We wanted to investigate if there was a possible compromise between reducing the vacuum needed in our detectors chamber, without significantly losing the resolution of our detectors, as I will discuss below.

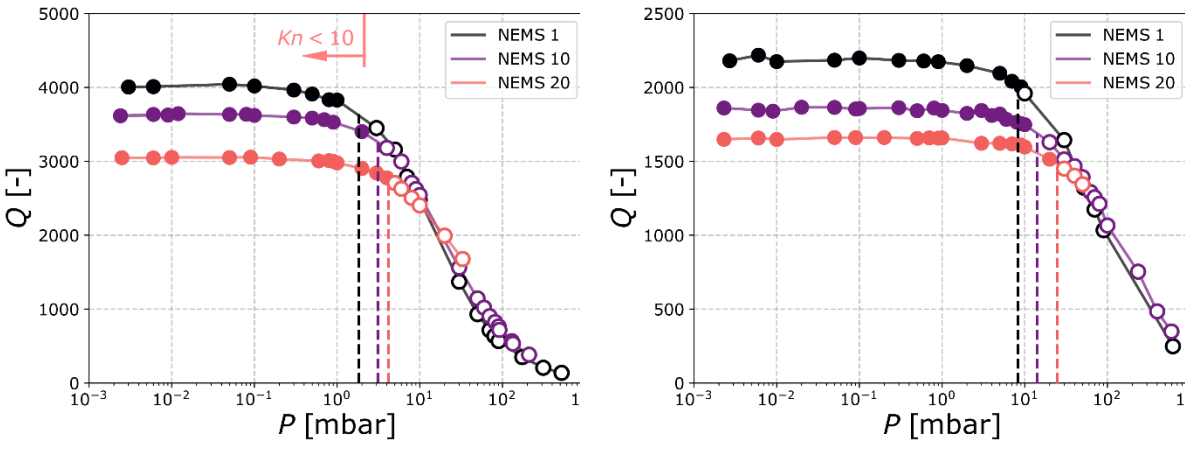
3.3.1.1 Nano-resonators in moderate vacuum

Thus, the first step was to verify that that the NEMS detectors that we use are not affected by the increasing pressure in the chamber (in pink in Figure 40). In order to evaluate this, we investigated the quality factor (Q factor), which captures information about energy (E) dissipation in the system. The following formula, taken from Li et al.[198], describes all the components that interfere in calculating this factor as a function of pressure (P), as follows:

$$Q = \left(\frac{\pi}{2}\right)^{\frac{3}{2}} \rho d f_0 \frac{\sqrt{\frac{R_0 T}{M_0}}}{P} \simeq 93 \left(\frac{d}{L}\right)^2 \frac{\sqrt{E\rho}}{P}$$

Where ρ, d, L and f_0 are the cantilever's density, thickness, length and resonant frequency, respectively; and T is the temperature, the perfect gas constant R_0 = 8,317 J K^{-1} and M_0 = 29 kg mol^{-1} is the molar mass of air.

In the following figure, Figure 41, we plotted this calculated Q factor as a function of the pressure in the chamber, for three different NEMS in our array, varying in length, i.e. NEMS (#1, black): longest; NEMS (#10, purple): median; NEMS (#20, red): shortest. The figure represents the evaluation for the first mode on the left (a), and the second mode on the right (b). More details on these NEMS length, frequency and the Q factor can be found in supplementary (Table 23).



(a) mode 1 **(b) mode 2**

Figure 41. Resonance quality factor of the first (a) and second (b) modes for the NEMS number 1, 10 and 20 when operated in at various pressures. This three NEMS have different lengths and thus different resonance frequencies (see Table 23). The dashed vertical lines indicate pressures at which the quality factor loses 10% of its maximum value. Figure reprinted from Reynaud et al.[165]

Before reaching high pressures, slightly lower than ~1 mbar, Q factors are constant. These are, for resonators #1, #10 and #20, at: 4000, 3600 and 3000, for the first mode; and 2200, 1900, and 1700, for the second mode. The dashed vertical lines represent a loss of 10 %

of the maximum Q factors value at: 2, 3, and 4 mbar, for the first mode; and 10, 20, and 25 mbar, for the second mode, for NEMS #1, #10 and #20, respectively. These results were consistent with those described by Li et al.[198], on the effect of increasing pressure on NEMS.

Regarding these decreasing quality factors, the manifestation of the minor effect of pressure on quality factor can be given by the flow regime. It can be determined by the Knudsen number $Kn = \lambda/L$, which is the ratio between the molecule mean free path λ and a resonator characteristic dimension L. In our case, the width of the beams (300 nm) can be considered to estimate the Knudsen regime. [199] If the mean free path λ is larger than L (when $Kn > 10$), the probability that the beam encounters a gas molecule is small, leading to a null viscous damping. Using the model developed by Trzpil et al.[200], we quantified the contribution of each phenomena, i.e. support-related, thermoelastic and acoustic, over the Q factor. The calculations were done for a 9 µm long, 300 nm wide and 160 nm thick beam. We concluded that in this regime, the damping is mainly caused by thermoelastic effects and energy transfer into the support (clamps). Viscous damping is introduced for a Knudsen number $Kn < 10$, leading to decrease of the quality factor. A vertical line (Figure 41a) was plotted to delimit the relevant Knudsen regimes. For devices with characteristic dimensions in the hundreds of nm, this regime shifts toward higher pressure values, and the operating pressure may be increased without sacrificing performance. Showing that these pressures are compatible with the increased pressure range calculated above, that is in the range 2-5 mbar.

Thus, the next step we investigated was to see if we could improve the focusing of the particles in a particle beam, and concentrate them more precisely towards the NEMS, in order to increase the capture efficiency of our detectors.

3.3.1.2 Aerodynamic lens optimisation

In order to evaluate the functioning of the aerodynamic lens in the desired pressure range described above, a series of theoretical and numerical analyses were performed as will be discussed below.

In Figure 42, we can see the particle trajectories inside the aerodynamic lens, simulated for different particle diameters. The different panels correspond to (a) 500 nm, (b) 250 nm, (c) 100 nm and (d) 25 nm. They all have a particle density of 1.06 g/cm^3, which corresponds to polystyrene (PSL), and as we mentioned previously is very close to that of viral samples, making these lens optimally designed for this usage. They are computed with an exit pressure of 1.33 mbar, compatible with the figure above (Figure 41), at the last chamber on the right,

where the NEMS detectors are at the centre position = 0, at 4 cm away from the skimmer position (last lens on the right; presented as the thicker white line).

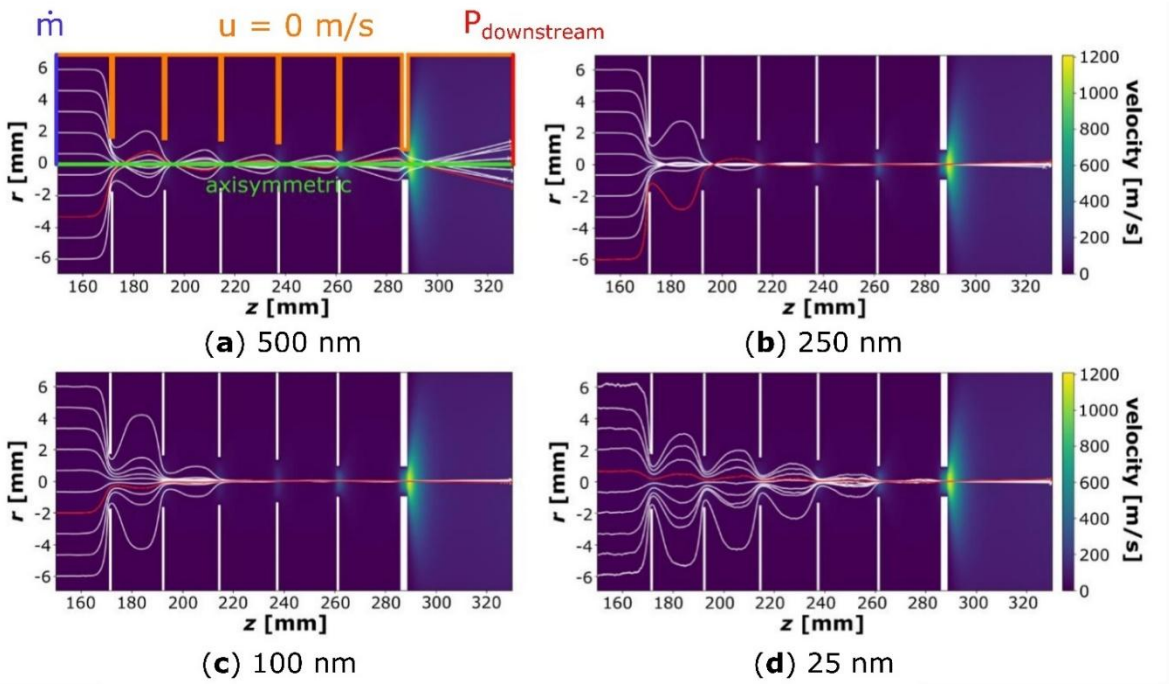


Figure 42. Particle trajectories through the lens, simulated for different particle diameters: (a) 500 nm, (b) 250 nm, (c) 100 nm and (d) 25 nm. Figure reprinted from Reynaud et al.[165]

We can see that from this figure that for a large particle, e.g. 500 nm in diameter, the beam is very dispersed near the detectors. As we go to smaller particles, rather below 250 nm and more optimally at 100 nm, which is an optimal size range for viruses, the beam is much more concentrated towards the NEMS, with very little dispersion. When we go even smaller, e.g. 25 nm particles, which would be very small viruses, the dispersion is maintained but very slightly enlarged.

Based on these results, both the solid angle Ω and the particle beam radius, at 4 cm downstream the aerodynamic lens (written as r_{4cm}) were calculated for different particle diameter d_p. The distance of 4 cm corresponds to the location of the NEMS array in the new mechanical architecture.[165] These results are reported in the right panel of the following figure, Figure 43.

In the left panel of Figure 43, we wanted to experimentally verify the focusing of the aerodynamic lens. I performed this by using a silicon chip as a target and exposing it for the duration of an experiment with PSL-NPs, with an average diameter of 90 nm. Afterwards we observed this deposition, using a binocular loop and performed image processing by the image J software.[201]

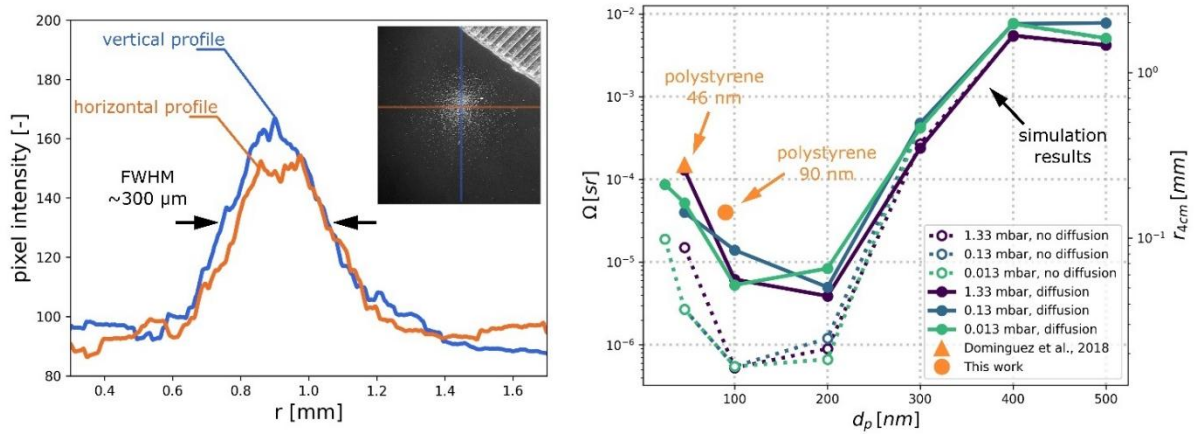


*Figure 43. (left) Pixel intensity profiles of the deposit photograph along two axes, located by the orange and blue lines on the inset. (right) Solid angle **Ω** and particle beam radius at 4 cm downstream the lens (location of the NEMS sensor) computed for several downstream pressures. Orange points represent experimental data and dots are simulation results. Figure taken from Reynaud et al. [165]*

In the left panel of Figure 43, in the inset on the right, we can visually observe the deposition of PSL-NPs, on the silicon chip, and by performing the vertical (blue) and horizontal (orange) profile of this deposition, we estimated the distribution to be at ~300 nm in its full width half maximum (FWHM). This result was consistent with the theoretical and numerical analysis, as I will detail below.

The results reported in in the right panel of Figure 43, show how the solid angle behaves depending on particle diameter, and that the optimal range for the solid angle is at 100 − 200 nm. In addition, when accounting for diffusion, i.e. the Langevin force, we get a more realistic presentation of the data, and we see that the experimental results (orange dots) are much closer to these models (solid lines). Furthermore, taking into account for diffusion seems to be critical below the 300 nm diameter threshold, and after that point there seems to be no difference between the model accounting for diffusion or non-diffusion. We can also see the difference between the work from Dominguez-Medina et al. (orange triangle) [182], and from the Reynaud et al. article (orange dot)[165], in this last there is a 5-fold improvement. This is probably primarily because the novel design (prototype #2) has a smaller lens to detector distance, which has gone from 8 cm to 4 cm. Finally, within the investigated pressure range (1.33×10^{-2} − 1.33 mbar), the model showed that the impact of diffusion remained relatively constant, and a limited influence was caused by pressure on the resulting solid angle. All these data indicate that there seems to be no loss of NEMS performance due to the aerodynamic lens for the examined pressure range.

Another major gain coming from the decrease of bench mass, and from the hand-on experience gained on the prototype #1, was our intention to make the system modular and user-friendly, which I will discuss below.

3.3.1.3 System modularity

Because of the heavy pumps operated on the first-generation prototype, heavy duty vacuum components were used to support the weight and to prevent mechanical vibrations. Large hardware parts made the aerodynamic lens-to-detector distance optimization difficult and constrained the sensing area to deposit area ratio. Moreover, large mechanical parts being complex and expensive to machine, a built-in alignment solution could not be implemented and a port aligner and a XY stage were used to align the lens outlet to the NEMS array. As a result, one of the most time-consuming tasks to perform when replacing a NEMS chip was the alignment process, which was required to keep the sensor at the centre of the particle beam produced by the aerodynamic lens. This operation was facilitated in the new system by designing a native alignment between the inlet capillary, the aerodynamic lens housing and the skimmer. Henceforth, the alignment step now reduces to positioning the sensor within the particle beam area instead of aligning multiple parts (aerodynamic lens and sensor) as in the previous NEMS-MS prototype.

Finally, since the novel NEMS-MS prototype may host various sensor technologies (nano-electromechanical as well as nano-optomechanical devices[186]) the sensor chamber was designed accordingly, guaranteeing an easy and compact alignment solution whatever the type of chip used (in pink in Figure 40 top). The sensor itself was packaged onto a printed circuit board which was mounted on a motorized XY stage (Agilis AG-LS25, Newport, Irvine, California, USA), which could be used to monitor sensor location and scan the particle beam.

This second prototype was engineered considering not only smaller dimensions, but also the modularity of the instrument, allowing future users to easily upgrade the instrument. Therefore, each function has been attributed a different mechanical module, as shown in Figure 40, that is: particle intake (in green), aerosol focusing (in violet) and particle mass measurement (in pink). KF50 flanges ensure the mechanical cohesion of the assembly, resulting in a limiting the diameter of the instrument, making it easier to mount, open and close, compared to the first prototype. Moreover, the overall system was mounted on a rail, keeping the sensor chamber fixed and allowing the inlet chamber and the aerodynamic lens to slide away from the nano-resonator housing to access and replace the sensor's chip.

Therefore, the next logical step was to verify if the aforementioned improvements were efficient. To this end, I proceeded with mass evaluations using this new version of our bench (denoted as prototype #2 in the following), and compared the results to those obtained with prototype #1.

3.3.2 Benchmarking the NEMS-MS prototype #2

Eventually, to assess the new-generation compact and modular NEMS-MS system, I performed the mass analysis of a series of calibrated samples of GNP of sizes 20 nm, 30 nm and 40 nm. The distributions obtained for the three sampled analysed in this study are presented in Figure 44.

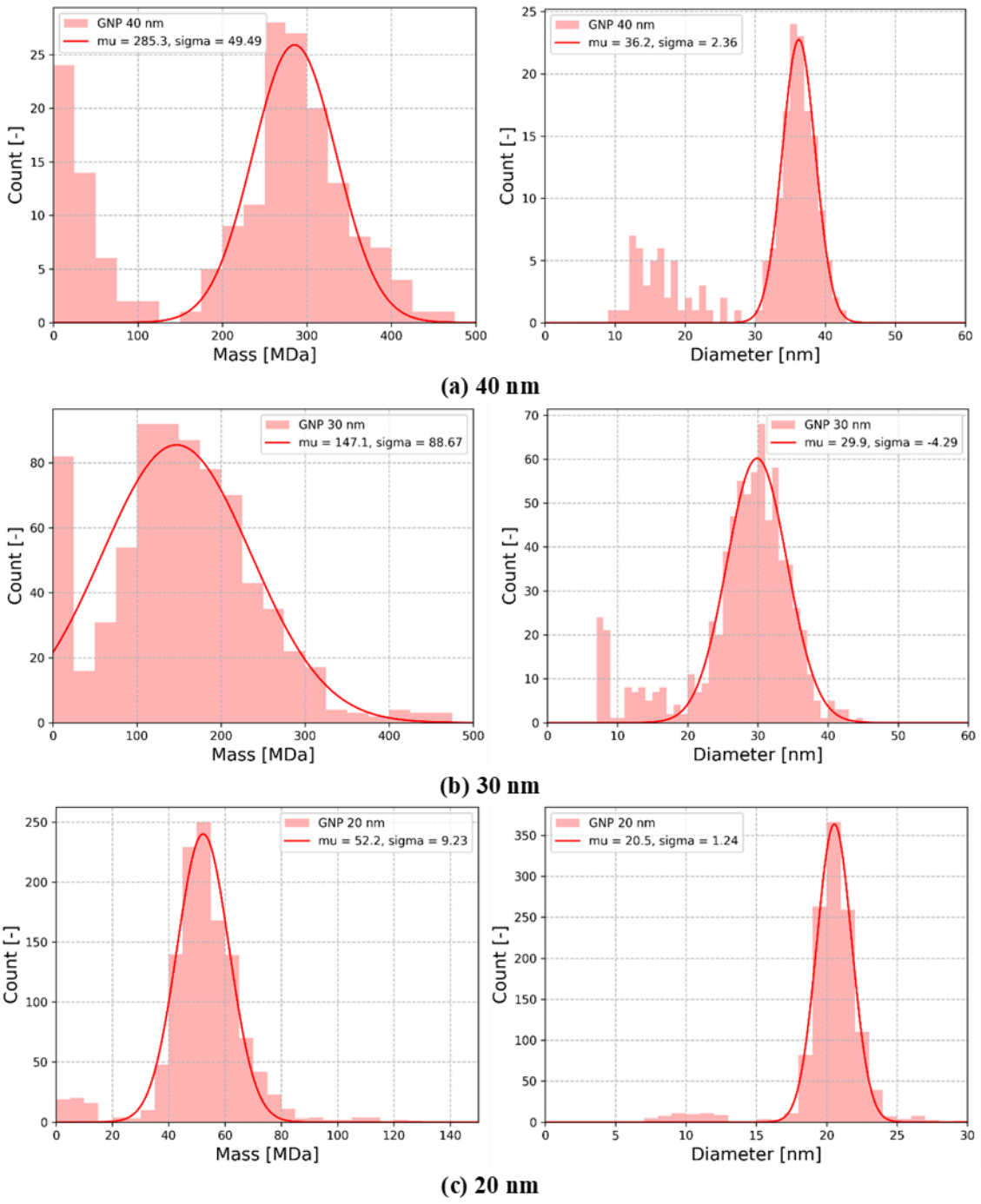


Figure 44. Histogram of the mass measurement performed during 60 minutes of acquisition time retrieved from the frequency shifts for gold NP of diameter a) 40nm b) 30nm c) 20nm. Figure reprinted from Reynaud et al.[165]

Nanoparticle colloidal suspensions were diluted in methanol, 1:1 v/v dilutions, achieving concentrations of 4.50×10^{10} to 3.50×10^{11} NP/mL depending on nanoparticle diameter. The amplitudes of the observed frequency discontinuities can be translated into the mass-position domain, allowing the production of a particle mass distribution (PMD).

These show good agreement with manufacturer's data (diameter measured by SEM), the relative difference in size ranging from 0.3 % and 16 % depending on the sample, and from 8.0 % to 11.9 % in mass. The standard deviation also compared well with manufacturer's data. Moreover, it should be highlighted that even if this comparison allowed quantifying the difference between NEMS-MS mass measurement and SEM size measurement, these techniques do not determine the same metric. Indeed, in order to compute the diameter of a particle, SEM is based on 2D image processing and NEMS-MS performs a mass-to-size conversion. Thus, both techniques assume the sphericity of particles, but the different metrics may induce different error factors. Finally, one might keep in mind that nano-resonator mass measurement has its own uncertainty, as we investigated previously in Clement et al.[149] More details on the statistics related to this mass measurement and the comparison to size evaluations can be found in supplementary (Table 24).

I will now focus on the particle transmission efficiency, as it is a defect of prototype #1 and one of the most important parameters for prototype #2 efficiency. Consequently, I verified the particle transmission from the nebulisation source to the NEMS detectors and evaluated our yield during this process.

3.3.3 Evaluation the transmission efficiency

Next, I contributed in helping to compared the transmission of the particles in the bench, in order to quantify the improvement factor between the first-generation prototype, based on the work of Dominguez-Medina et al.[182], and this new 'mini NEMS-MS system'.[165]

For this, we diluted GNPs in methanol (1:1 v/v) to facilitate nESI nebulisation. We computed the particle capture efficiency, as:

$$capture\ efficiency = \frac{C \times Q \times \Delta t}{N_{detected}/N_{NEMS}}$$

Where C: particle concentration in solution; Q: flow rate; Δt: duration of the NEMS exposure. The ratio between the number of sprayed particles and the number of detected particle ratio was then computed for individual NEMS devices, using the relationship seen above. The capture efficiency was also computed for one single NEMS, since the number of used devices can vary from one experiment to the other, and since our prime objective was to compare

performances of NEMS-MS system generations. The following table, Table 4, present the data for all three GNP samples analysed earlier in Figure 44.

Table 4. Capture efficiency computed for the gold nanoparticles (GNPs). (*) Range computed based on the upper and lower capture efficiency reported by Dominguez-Medina et al., (2018) [182], namely 1 per ~2×10^9 and 1 per ~1×10^{10} sprayed particle, respectively. (**) Capture efficiency expressed as 1 detected particle per N (nebulised particles).

GNP sample	Sprayed concentration [$\times 10^{10}$ NP/ml]	Flow rate [µl/min]	ESI voltage [kV]	Duration [s]	Sprayed particles [$\times 10^9$ NP]	Number of NEMS [-]	Event count [-]	Capture efficiency (**) [-]	Capture efficiency for 1 NEMS (**) [-]	Improvement factor (*) [-]
20 nm (BBI)	35	0.5	2.3	4200	12.3	9	1220	1 per 4.5×10^6	1 per 9.0×10^7	22 – 111
30 nm (Sigma-Aldrich)	8.95	0.55	2.5	3800	3.12	12	739	1 per 2.5×10^6	1 per 5.1×10^7	40 – 198
40 nm (BBI)	4.50	0.5	2.2	4200	1.58	9	183	1 per 3.9×10^6	1 per 7.8×10^7	26 – 129

As we can see from Table 3 the measurements for all particles had an improvement rate from nearly 20 to around 200 times better than those reported in Dominguez-Medina et al.[182] We believe that many of the mechanical optimizations allowed these improvements, such as for example the reduction of the in aerodynamic lens-to-sensor distance, going from 8 cm to 4 cm. Furthermore, the lens housing (in blue in Figure 40 top) and the skimmer are mounted contiguously, in addition to the lens housing being aligned with the rest of the mechanical parts of the system, thus removing any degrees of freedom. This reduces any vibration caused by the turbo pump, in addition to the pumping requirements downsizing, which allowed reducing every dimension of the system.

To compare to Erdogan et al.[196], who achieved a capture efficiency of 1 per 1.85×10^5 particle to 1 per 4.97×10^5 particle for one device, our system still has room from improvement (by a factor of 10-20 times). At this point, it should be stressed that direct particle capture efficiency comparisons are complicated, due to potentially different raw resonance frequency postprocesses. Consequently, measurements performed with their prototype on gold nanoparticles and viral particles exhibited substantial mass dispersion.[196,197]

Therefore, the next step would be to analyse the other factors that could affect the particle deposition on the NEMS and try to optimize the conditions in order to have better reproducibility and more robust data.

3.4 NEMS-MS: an analytical method

In the rest of this chapter, I will detail some of the advancements that I was able to contribute to, in collaboration to the rest of the NEMS-MS group, in order to render this instrument closer to an analytical method and further on as a robust structural biology tool for future usage. It has to be noted here that the following results present a resume of the main results for the NEMS-MS as a whole, therefore a combination of result from both prototypes is presented.

3.4.1 Temporal Influence on Particles Deposition

In order to check the temporal effect on the deposition, I made the following experimental design:

a) First, spray a constant particle concentration (10^{10} NP/mL), with a constant capillary temperature of 100 °C.
b) Then analyse the deposition for four different time steps, at: 45'; 1 h 00'; 1 h 30'; 2 h 00'.

The results are reported in Figure 45, and the image analysis and FWHM calculation was the same as that described earlier (Figure 43).

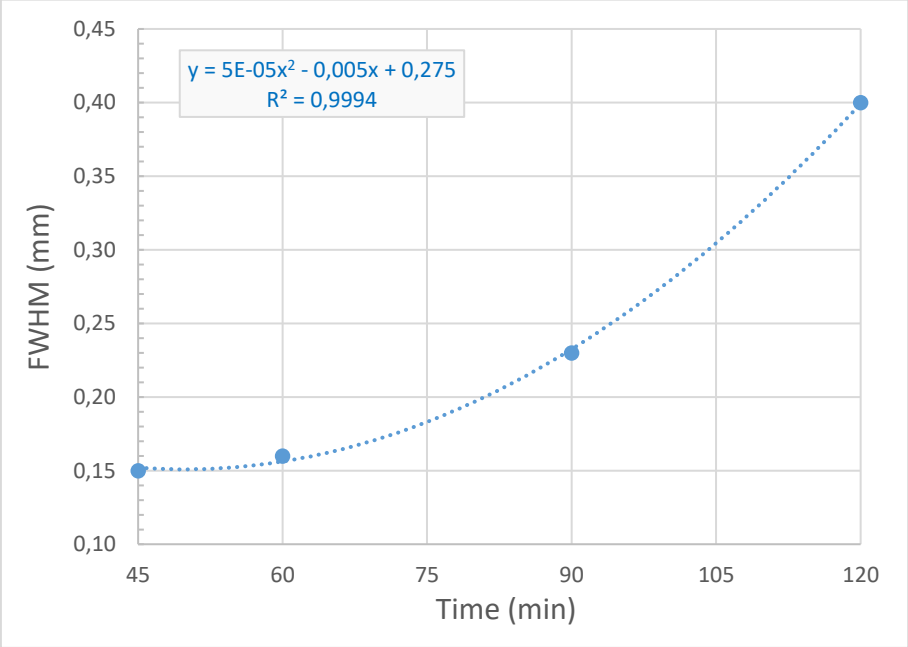


Figure 45. Correlation between Full Width at Half Maximum (FWHM) of a normalized intensity profile and deposition time for four PSL nanoparticle depositions: 45 min, 1 h, 1 h 30 min, 2 h. Dotted line represents the 2nd degree polynomial fit on the data, and the respective equation and R^2 are presented in the top left of the graph.

It is important to note here that the rise in profile width with longer deposition times was not linear, as highlighted by the fit equation in Figure 45. The relationship between FWHM

and time exhibited a quadratic relationship, signifying that the deposition area or surface expands linearly with increasing deposition time.

The observed widening of deposition profiles over time can be attributable to a variety of reasons, one of which being particle aggregation. As deposition develops, particles may have a propensity to aggregate, resulting in broader profiles (one can imagine it like sand pouring down a pipe onto a flat surface). These interactions may also be charge-mediated, especially if the incoming particles carry charges.

Now that we built the bench (Figure 40), and observed that there was a temporal effect on the particle focusing, the next step was to use the NEMS-MS bench to explore the influence of physical parameters on the mass evaluation.

3.4.2 Peculiar results for 20 nm GNPs

In order to have a reliable study of the possible conditions that could affect the mass measurement, Kavya Clement had identified in her thesis (using the bench prototype #1) that several parameters were very critical, namely: the flow rate, the relative humidity; and the capillary temperature.[202] Since I could not control the relative humidity, I kept track of it in a regular manner, and decided to play with the other two parameters, that is the flow rate and the inlet capillary temperature.

Therefore, I decided to perform a first study, spraying particles with ESI, maintaining a constant flow rate of 30 µL/hr, and capillary temperature = 175 °C, since we had previously shown it to be compatible to our experiment.[184] Next, I decided to investigate a series of GNPs having different sizes (15; 20 and 30 nm), and that were prepared in the same solutions, suspended in citrate buffer. The results of this analysis are presented in Table 5. It has to be noted that these data were collected in a series of independent measurements and not in a single experimental run.

Table 5. Table describing the gold nanoparticle (Au-NP) analysis of different sizes, Mass measurement values are reported for: flow rate = 30µL/hr and capillary temperature = 175 °C. St.dev. = standard deviation.

Reported Particle diameter [nm]	**Number of mass events [count]**	**Mass fit**		**Size fit**	
		Mean [nm]	Std. dev [nm]	Mean [nm]	St. dev [nm]
14.2 ± 0.9	251	16.7	3.47	14,2	1,2
18 - 22	**542**	**22.6**	**4.7**	**15,5**	**1,0**
29.2 ± 2.1	636	156.9	26.46	29,5	1,6

First, we can see that the values for the 15 and 30 nm (BBI) are very similar to those discussed earlier (Figure 39). Nevertheless, the 20 nm GNPs (Sigma-Aldrich) sample presented

an unexpected mass, at 22.5 MDa, instead of the theoretically calculated value, at 48.7 MDa. This would represent an average size of 15.5 nm, instead of expected 18-22 nm range, reported by the manufacturer (Table 5).

We can see from Table 5 that the values for the reported 20 nm values (highlighted in red) are not at all in the range expected, and are much closer to a 15 nm diameter in reality, which pushed me to investigate further this sample. Since we are not aware of the methods used by the manufacturer (Sigma-Aldrich/Merck KGaA, Darmstadt, Germany) to determine the diameter, I thought that performing an extensive study to temperature could give us important additional information on the real size of these particles.

3.4.2.1 Temperature dependence and size change of GNPs

The experimental design consisted of spraying particle with ESI, at a flow rate = 30 µL/hr, and performing the mass measurement from 150 to 250 °C, every 25 °C. In Figure 46, I report the density distribution of the mass for the two temperature extremities, i.e. 150 °C (purple) and 250 °C (red), and a central value, i.e. 200 °C (green). The more in-depth analysis can be found in supplementary (Figure 129).

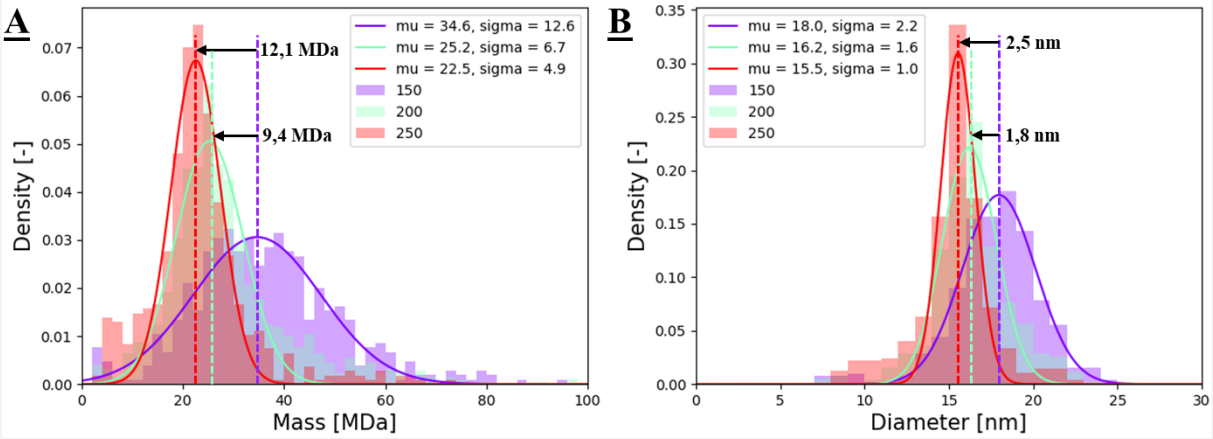


Figure 46. Analysis of the 20 nm Au-NPs as a function of Temperature variation in the inlet capillary, presented here for 150°C (purple), 200 °C (green) and 250 °C (red). The values of mu= mean; sigma= standard deviation, arise from the Gaussian fit applied on the experimental data.

The first thing we see is that the values for size (Figure 46 B), never reach the expected 20 nm size. We can also see that the more the temperature increases, the more the size reduces, and the narrower the peaks get. I noted that there was a very significant change in the mass, around 12.1 MDa, between the two extremities investigated. When converted to size, this change was 2.5 nm (> 10%), which is very significant. I also noted that the standard deviation of the Gaussian fit got smaller and smaller as I increased the temperature, suggesting a constant reduction of the size. This could potentially be related to the characterization of the NPs core

diameter or their hydrodynamic diameter, when performed by other sizing methods such as SEM or DLS.

Our hypothesis is that this could potentially indicate either a continuous desolvation of the buffer surrounding the NPs, or a possible sublimation (solid-to-gas phase change) of the Au-NPs themselves. Although this second seems much less likely at these temperatures, due to the high temperatures that gold requires to sublimate. Reported in the literature, I found that for a given 20 nm GNP, the pre-melting region is from 1000 - 1090 K (~ 727- 817 °C), and the melting point is ~ 1090 K (816.85 °C).[203] Let us now see how we tested and why we think that the first hypothesis seems more probable.

I need to specify here that our experimental conditions to observe the evaporation phenomenon were not optimal when using ESI. That is because this nebulisation method generates primarily submicron droplets, rarely exciding 5-20 μm droplet even after needle optimisation, and in addition, also knowing that smaller droplets carry more charges.[204] That is why we chose to switch to Surface Acoustic Wave Nebulisation (SAWN), because this nebulisation approach produces larger droplet (cf. later to Figure 103), and thus the evaporation phenomenon is more clearly visible. Therefore, the next step was to see if using SAWN would indeed improve our understanding of particle desolvation in the NEMS.

3.4.3 Particle desolvation on the NEMS resonator

In order to investigate the effects of particle desolvation on the NEMS, we generated an aerosol using SAWN, containing GNPs, prior to the mass measurement, while keeping the inlet capillary at a constant temperature. The experimental set-up was simply reduced to having an in-line desiccant, between the box where the aerosol is generated and the NEMS-MS bench. Thus, resulting in a dry aerosol when the desiccant is present (top left of Figure 47), and a hydrated aerosol when it is removed (bottom left of Figure 47).

In Figure 47, we present the mass evaluation of 20 nm GNPs (by BBI^{TM} Solutions). On the left side, it is the normalized frequency versus time; and on the right (only for the hydrated aerosol), we see the normalized landing position on the active segment of the NEMS beam (0.25-0.47), on top (in blue), and on the bottom (in red), the respective mass over time.

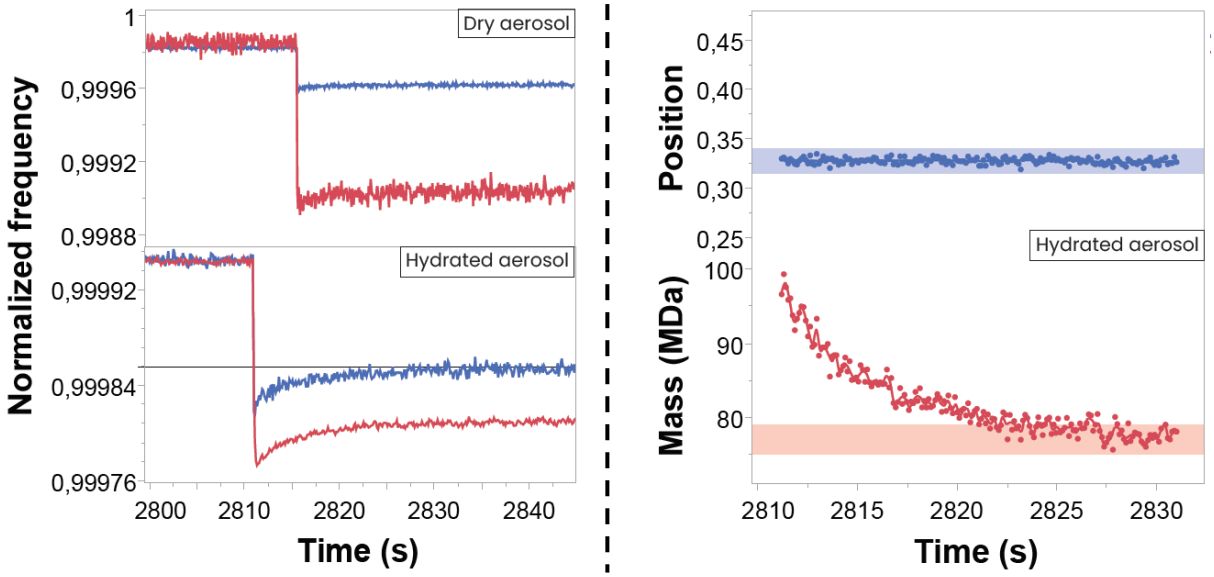


Figure 47. Particle desolvation on the NEMS devices. (Left) The normalised frequency over time if a 20 nm GNPs landing on the NEMS. On top, for a dry aerosol; on the bottom, for a hydrated aerosol. Mode 1: blue; Mode 2: red. (Right) Computed mass and landing position of a single landing event, over time. On top, the landing position (in blue); on the bottom, the mass in MDa, (in red). The shaded area indicate the stability interval.

In Figure 47 for the hydrated aerosol, on the bottom left, we observe a frequency recovery that is significantly less sharp than the one seen for the dry aerosol, top left. The frequency seems to drop and does not recover entirely (highlighted by the solid dark line for mode 1) for nearly 20 seconds, from ~2810 to 2830 seconds (highlighted by the solid dark line for mode 1).

When zooming in on this event for the hydrated aerosol and converting the frequency into its correspond mass and landing position, Figure 47 on the right, we can clearly see that the landing of the particle is stable. Meaning that we are observing the same particle over time, (in blue). While for the mass, the difference is significant, as it goes from around 100 MDa, at 2811 seconds, to about 78 MDa, at 2830 seconds. This can probably be attributed to on-device particle desolvation. It should also be pointed out here that not all frequency jumps have a relaxation time like the one shown here, many behave like in the case of the dry aerosol.

Another interesting thing that we were able to observe was also that the mass decreased as a function of the log $(\text{time})^3$, as observed in Figure 48.

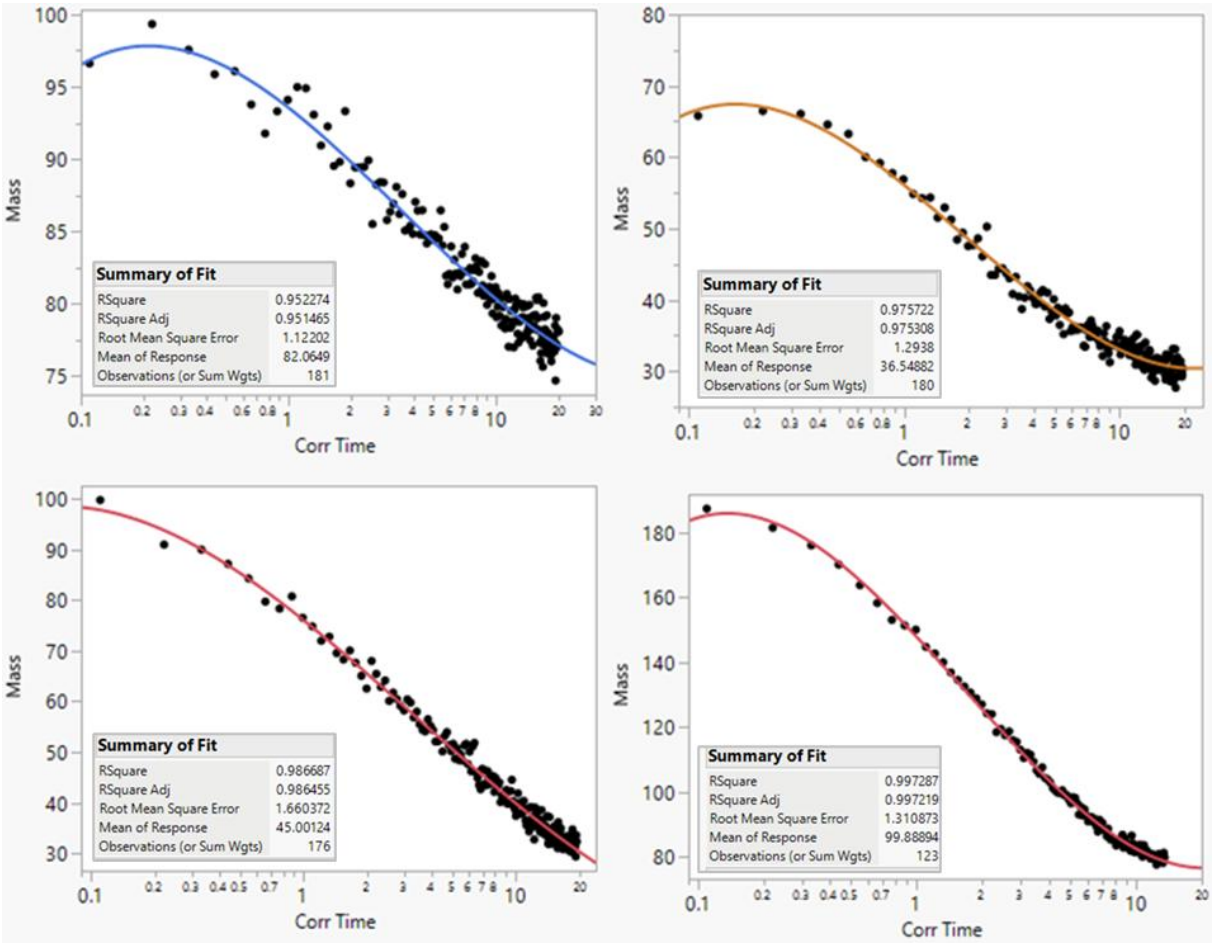


Figure 48. Fitting mass with a third-degree log-polynomic, as a function of corrected time. Data shown for three different NEMS resonators (colour code). The summary for each fit is given on the bottom left of each panel.

Here we can see that the third-degree log polynomial fits well to the data, for different NEMS, and for multiple observations on the same NEMS. This also indicates that these landings are not a singular phenomenon, and that collecting more statistics in the future will be possible. We believe that this is most likely an indication that the evaporation is a volume related effect, which can be of major interest to understanding the particle solvation on the NEMS. The perspective would be that with enough data we can build an analytical model that explain these observations.

Overall, these observations on the NEMS indicate that we need to be aware that to measure the mass, our measure is not only affected by what happens in the front-end of the bench, i.e. the different nebulisation source or the heated capillary, but also at the level of the NEMS detector itself.

All the evaluation that I discussed until now were related to inert gold nanoparticles (GNPs), and not on viruses. As I said earlier these GNPs are easy to spray standards, but also, they were an ideal candidate for assessing the importance of physical parameters on particle

deposition, without adding the complexity of biological characteristics. So let us now see what are the physical effects that can affect viral mass measurements.

3.4.4 Viral particle desolvation issues

As mentioned previously, in the thesis Kavya Clement it was said that the critical parameters were primarily: the flow rate, and the capillary temperature.[202] In our case, a key element for such investigations came as a result of our work in showing that the empty T5-CLPs desolvation was dependent of the flow rate and the voltage applied on the ESI tip.[149]

The following measurement was performed on empty T5-CLPs, using ESI, while keeping a constant inlet capillary temperature of 175 °C, for 5 different flow rates, increasing from 10 to 50 µL/hr (with a 10 µL/hr incremental step), and simultaneously adapting the voltage applied on the ESI tip[149]. Over the course of a 4200 sec (70 min) experiment, the flow rate was changed every 14 minutes in order to have an equal amount of time between flow rates. These results are presented in Figure 49. I contributed to designing the experiment and performing the data analysis of these measurements.

In Figure 49a, we show how the individual T5-CLP mass is affected as the flow rate is increased, in the duration of a single experimental run, and seeing the mass and the flow rate correlation in Figure 49b. In Figure 49c, we see the mass resolution as a function of time. In Figure 49d, we plotted how the PMD was affected at low (10-20 µL/hr) and high (30-50 µL/hr) flow rate.

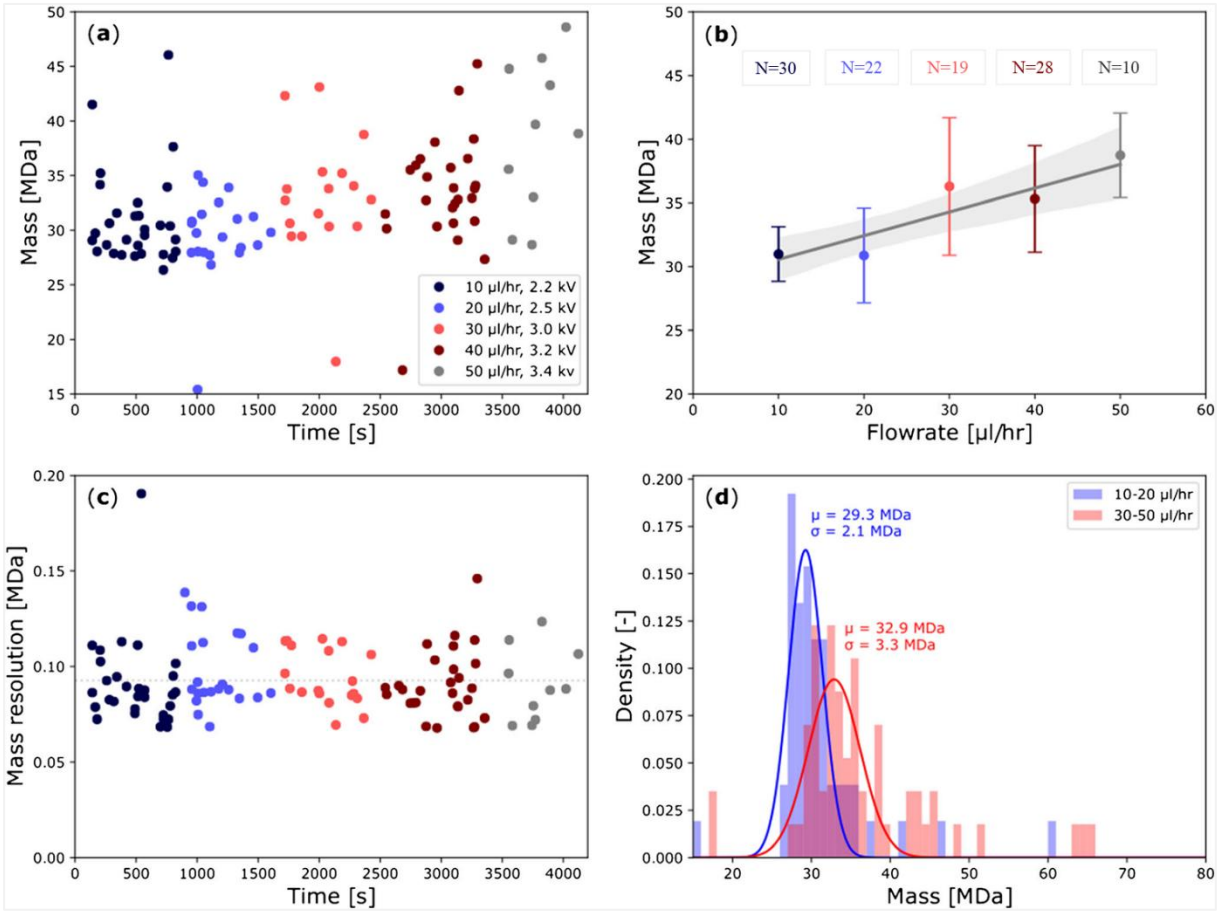


Figure 49. Investigation of the nano-ESI flow rate effects. (a) Mass of individual particles landing on the resonator over time as the flow rate was increased (N=109). (b) Average mass vs. flow rate with linear trend. The data are fitted by the equation 0.18 × x + 28.5. (c) Mass resolution for individual particle landing events over time. The gray dashed line corresponds to the average mass resolution of 93 kDa. (d) Composite spectrum at higher (30–50 µl/h; N= 57; in red) and lower (10–20 µl/h; N= 52; in purple) flow rates showing a mass shift of several MDa. Figure and description were reprinted from Clement et al.[149]

It was observed that a clear upward shift was taking place when the flow rate was increased consistently, from 10 to 50 µL/h (0.17 – 0.83 µL/min), during a single experimental run. This is even more visible in Figure 49b, where the 50 µL/h value does not overlap with the 10 µL/h, nor with the 20 µL/h value. That is why, we summed up the values, into low (10-20 µL/hr; in purple) and high (30-50 µL/hr; in red) flow rates, in Figure 49d. We measured that there was a consistent mass increase from the theoretical mass (~ 26.018 MDa), an increase by ~ 3.3 MDa for the low flow rate, and ~ 6.9 MDa for the high flow rate.

An important element here is that the mass resolution is consistently independent from the flow rates, as it can be highlighted by the average line (dotted grey line) over the course of the whole experiment (Figure 49c). Indicating that the mass measured is only due to the analyte, and not due to the NEMS detector. An additional element that needs to be highlighted here, is that this measurement was performed in pure water, compared to the one reported in Dominguez-Medina, et al. (2018)[21], performed in 12.5 mM Ammonium Acetate, which had a

mass discrepancy ~ 1.2 MDa. More details on this evaluation can be found in supplementary (Table 21 and Table 22).

These results show how much of an important parameter the flow rate is for the mass measurement, in addition to the non-negligible effects that the solvent might have in inducing mass discrepancies. The logical added mass seemed to come from the excess water left on the capsid during their transition from the ESI to the NEMS detectors. This prompted me to try to model what the mass of the capsid would be if it were filled with water (Figure 50).

3.4.4.1.1 Water filled T5 capsid

Therefore, I decided to do some back-of-the-envelope calculations to complement the calculations of the viral mass of the T5 empty capsid if it were to be filled entirely with water, as shown in Figure 50.[184]

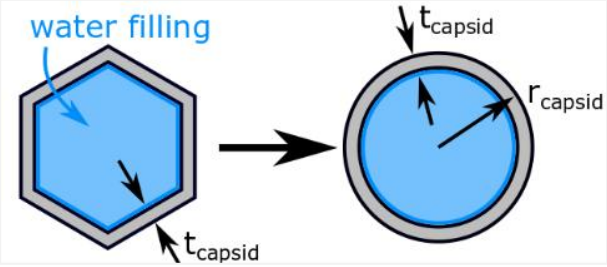


Figure 50. Sketch of a water filled T5 capsid particle. Figure was adapted from Clement et al.[149]

Taken that the average envelope thickness is [1.5, 4.5] *nm*, shown as true for more than 75% of viruses, by Lošdorfer et al.[205] Therefore I took the inner radius of the T5 phage capsid to be R_{in} (T5) = 46 *nm*.[149] Then, I calculated that the inner Volume of the T5 capsid, Vol_{in} (T5) = 407 720.1 nm^3, assuming spherical geometry for the particle.

Assuming that the density of bulk water remains the same at the nanoscale, we know that ρ_{wat} = 1 *g/mL* = 1 g/cm^3. Knowing also that M = V*ρ_{wat}, we can get that the mass of water inside the T5 particles, is $M_{water/in}$ (T5) = 246 273 875.9 *Da*, or *246,3 MDa* (case shown in Figure 50). Compared to what we measured earlier (Figure 49), this mass is almost an order of magnitude higher, meaning that this model is not the one that can explain our results.

Consequently, I looked for other explanations that could account for this mass increase. As a result, I proposed two new models, based on a partial water filling of the capsid. One considers a water layer outside the capsid (Figure 51a) and the other a water layer inside the capsid (Figure 51b). The next paragraph is devoted to a careful study of these models.

3.4.4.1.2 Potential explanation: Water layer thickness

A simple calculation can be done to provide a first estimation of the thickness of a potential water coating of the capsids. The calculation lies on geometric consideration, converting the mass excess into an outer and inner spherical annulus, as shown in Figure 51 (a) and (b), respectively.

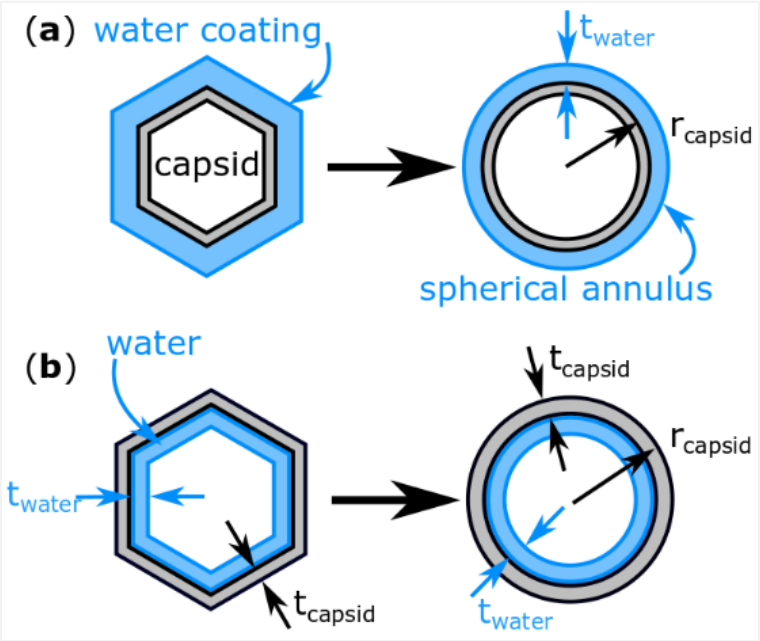


Figure 51. Sketch describing the geometric models used to compute the water layer thickness. Figure was adapted from Clement et al.[149]

In this framework, the thickness of the water coating t_{wat} is given by the formula:

$$t_{wat} = \left(\frac{3}{4\pi}\frac{m_{wat}}{\rho_{wat}} + r_{cap}^3\right)^{\frac{1}{3}} - r_{cap}$$

Where r_{cap} is the outer diameter of the capsid, m_{wat} is the mass of water, estimated as the difference between the theoretical mass m_{theo}, *i.e.* ~26.018 MDa, and the measured average mass m_{exp} (from Figure 49), *i.e.* around 29.3 MDa, and ρ_{wat} is the water density (= 1000 kg/m^3). Knowing that the inner volume Vol_{in} (T5) = 407 720.1 nm^3, as I calculated earlier, we can compute the water thickness t_{wat} by using the formula:

$$t_{wat} = r_{cap} - t_{cap} - \left((r_{cap} - t_{cap})^3 - Vol_{in}\right)^{\frac{1}{3}}$$

Where t_{cap} is the thickness of the capsids, as reported by Lošdorfer et al.[205] The results are reported in the

Table 6.

Table 6. Results of the calculation for the mass measured in different conditions and considering different capsid thicknesses. Table was reprinted from Clement et al.[149]

Variable		Flowrate/Voltage Low	Flowrate/Voltage High	Unit
m_{expm}		29.3	32.9	MDa
t_{cap}		$1.5 - 4.5$	$1.5 - 4.5$	nm
t_{wat}	in	$0.220 - 0.253$	$0.464 - 0.535$	nm
	out	0.204	0.426	nm

We can thus see here that the hypothesis about water being on the outside or the inside of the capsid is consistent with observations. Indeed the low flow rate (voltage) values, *i.e.* about 0.2 nm, and high flow rate, *i.e.* about 0.42-0.54 nm, are on the same order of magnitude as the thickness of water monolayers, *i.e.* 0.25 ± 0.05 nm, as reported by Opitz et al.[206] This would actually correspond to a single monolayer for the low floor rate, and a bilayer of water for the high flow rate case.

This is an important result as it shows that these considerations about water layer thickness were not in our considerations when performing the mass evaluation. One way we can discard the hypothesis of the water layer being in the inside of the capsid (Figure 51 b) is to do a similar measurement with GNPs. Knowing that these particles are solid on the inside, we can only observe an outer layer of water, if it is still physico-chemically similar to that of a virus.

I also checked if the totality of the water can leave the capsid while it is passing inside the bench, and if there is any other explanation that could help us evaluate this added mass of water.

3.4.4.1.3 Drying time evaluation, for a particle filled with water

The article of Andoh et al.[207] performed Molecular Dynamics (MD) simulations on the time it takes for the medium inside a poliovirus to be exchanged from the outer buffer, in order to replace the integrity of the inside of the capsid. In the article, they stated that in their MD simulations it took approximately about $\tau = 25$ μs, to replace all the water molecules inside the capsid (n_w), where $n_w \sim 200\,000$ *H_2O molecules*.

Knowing that $n_w = M_{water/in}/m_w$, and that the mass of a water molecule is $m_w = 18$ *Da*, I calculated that n_w (inside T5) = 13 681 882 *molecules of H_2O*. If we assume that the surface

flux is the same between poliovirus and T5-CLP, that would be a constant Φ (poliovirus) between 3.25 and 5.26 molecules*nm^{-2}*μs^{-1}. This provided me with the **τ (T5)** between 97.8 and 158.32 μs.

For a particle to go from the entry of the bench to the NEMS detectors (in prototype #1, in which the mass measurement shown in Figure 49 was performed), it must cross a Length (L) = 70 *cm*. Given that the particle reaches supersonic speeds inside the system, we can assume that it reaches an average particle speed in the system, **v** = 500 *m/s.* I therefore calculated that the crossing time, that is **t (prototype #1)** = L/v = 1.4 *ms* = 1400 μs.

We can clearly see that **t (prototype #1) >> τ (T5)**, meaning that the crossing time (t) that it takes for a T5 phage particle to pass inside the bench is significantly larger than the time it takes for the particle to release the water content found inside the particle, (τ). As a result, viruses are supposed to be "dry" when they arrive on the NEMS. This result appears to contradict the models of water layer I developed above and does not align with what we observed on Figure 49.

As an alternative explanation to the water layer models, which would be consistent with the previously estimated drying time, I propose that other proteins that we are not aware of are found in the inside of the capsid, thus rendering the theoretical mass calculation of the virus erroneous. One way to check this could be to combine the native mass evaluation with another relative mass evaluation method, such as bottom-up proteomics, for example. I will discuss this relative approach further in the next chapter.

3.4.5 Overview and perspectives regarding the analytical capabilities of the NEMS-MS bench.

These considerations helped me to get a better understanding of the instrument and the difficult points that we may encounter in analysing different samples of interest. We can see that the primary difficulties, especially in understanding the particle desolvation, are critical at the entry of the instrument, especially during the nebulisation phase. In the following chapters, I will also talk about some of the recent improvements that I was able to contribute, in order to optimize the nebulisation source, with a particular focus on the SAWN nebuliser.

3.5 Conclusions for Chapter 3

In this chapter I first focused on the capabilities of the first prototype to perform the mass analysis of different gold nanoparticles (GNPs). As I mentioned earlier, these GNP are

well standardised and characterised by their manufacturers; we can obtain them in large quantities; they are easy to spray; but it was also important to say that they are an ideal candidate for assessing the importance of physical parameters on particle deposition, without adding the complexity of biological characteristics at first.

In addition, the first prototype was stationary and too bulky to transport, and thus not too practical for real life applications. This is an especially important limitation, as I presented in the first chapter, regarding real-time virus diagnostics, since we need an easily transportable instrument, in order to conduct terrain measurements. This was the driving force for us to build a more modular and compact NEMS-MS, which we named the 'mini-NEMS-MS' (Prototype #2), which I helped to build.

Particle capture efficiency being one of the main limitations of NEMS-based MS, it was one of the main motivations underlying a new bench design, and its optimization. Improving it allowed us to perform faster analysis: reducing run time from 10 hours to 60 minutes, in order to build the mass spectra (e.g. Figure 44). Overall, the NEMS-MS technique presents a trade-off between particle capture efficiency and mass resolution: increasing the NEMS exposition to aerosol seems to imply an increase in pressure, which in return affects the mass resolution. In our system, the trade-off to keep a ~0.1 MDa mass resolution[149], seems to be pressure-wise, at 0.5 mbar (Figure 41). However, this threshold might not be suitable for smaller particles, especially below 100 nm in diameter, as the diffusion could cause the divergence of the particle beam downstream the aerodynamic lens.[208] The interest in analysing large sample fractions grows when the amount of available sample volume is small, which is often the case for biological samples, or for broad mass distributions, which require many events to acquire adequate and relevant statistics.

In the rest if this chapter, I focused on the mass measurement of different samples, mainly on GNPs and the model viral particle empty T5 capsid. One of the main tasks, that I had the opportunity to investigate in detail during my thesis, was about the desolvation of samples in their crossing from the liquid to the gas phase. I helped in demonstrating that there are noticeable effects due to bad particle desolvation, performing a series of measurement, by changing parameters such as capillary temperature, flow rate and different nebulisation methods. This proved that desolvation is an important factor for both inert and viral particles. As I had mentioned in the previous chapter, I also helped in estimating that the impact of the frequency noise over the frequency-to-mass calculation to be ~0.1 MDa. Where we argued that a conservative estimate of NEMS-MS mass precision taking into account these effects is ~1

MDa.[149] As we can see by the measurement in Figure 47 and Figure 49, the influence of all these parameters appears largely dominated by that of physico-chemical phenomena occurring during the transfer of viral (or inert particles) from solution into the gas-phase. This will be an important focus on using other methods of evaluation in solution or in gas phase for future analysis, which I will focus on in the following chapters.

All this chapter aims to show that the NEMS-MS can soon become a more common tool for analytical chemistry and structural biology, focusing on virology, of course in cohabitation to other structural biology tools. That is why the third axis of my work was the integration of more conventional structural biology methods in the pipeline, and especially as a pre-requisite for the NEMS-MS analysis. The primary goal of this work was to have a detailed characterization of the samples before the NEMS-MS analysis, and the cautions needed to be taken for viral sample preparation.

4 Chapter 4: Establishing a structural biology pipeline prior to the NEMS-MS analysis

Let us now take a step back, in order to put in perspective, the knowledge acquired during the previous chapters. We currently have found there is a clear interest to characterise total virion mass, where we stated that mass is theoretically a method for differentiating viruses. And then we followed up by describing the different MS methods to achieve such a differentiation. However, on top of these measurement tools, we also need to make sure that mass measurement does not affect the physical integrity of viruses, otherwise it is difficult to test the validity of our theoretical model. And this methodology must of course be independent of the choice of mass-measuring device, so as not to bias the answer.

In the following chapters, I will detail my perspective of how to make sure that we have a real quality control on our mass measurements and how this could render them into even more robust structural biology instruments.

4.1 Introduction of Chapter 4

In this chapter I will focus on the question I had during my time performing MS experiments: “What happens to a viral particle in the NEMS-MS bench ?” This question has a fundamental aspect, regarding ‘what are the physical states that the virus goes through during its passage inside the NEMS-MS bench ?’. One can separate these into three main parts:

(i) from its conservation in solution to
(ii) its transfer into the gas phase, as an aerosol, to
(iii) its deposition on the NEMS-MS detector.

To answer such a large question, one has to design and build an appropriate analytical pipeline in order to observe these transitions, and of course to extract informative results by building the necessary checkpoints all along the experiment.

This thus requires two main aspects: 1) material: finding the right model viral particle to work with; 2) methods: finding the right analytical techniques, in order to reach conclusive results. The following figure (Figure 52) summarises the main point described in this chapter.

Here I will thus detail the choices that I made and explain the optimisations that I put in place in order to have a functional analytical pipeline.

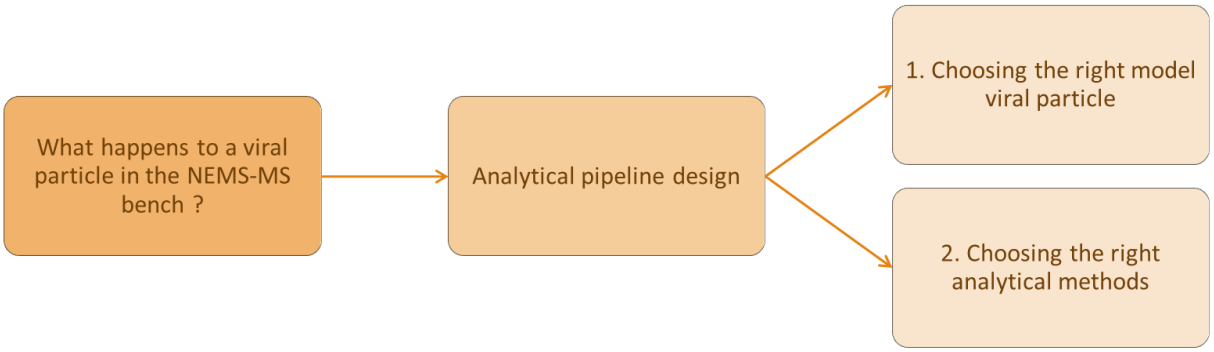


Figure 52. Schematic representation of the structure of the chapter 4.

4.1.1 The three states of a virus during the NEMS-MS analysis

A certain number of modifications may happen to viral particle when it is found in its buffer (part highlighted in blue in Figure *53*). In order to prepare the samples we usually have to dilute the preparation, or even perform a buffer exchange, such as water [184], or more often ammonium acetate, a volatile buffer, which is commonly used for native MS measurements.[139]

A major critical step that follows the solution preparations is the nebulisation step (part highlighted in violet in Figure *53*), where depending on the technique chosen the particle behavior may differ greatly. The most well studied nebulisation method is electrospray ionisation ESI, but we also tried to investigate what would happen with the sample if we used SAWN. This nebulisation step is critical in the way that particles are separated inside droplets and how they are introduced inside the system.

The following step is the transmission of the particles to the detector, where particles inside the aerodynamic lens are expected to reach supersonic speeds, sometimes as high as 500 m/s.[164] These particles then land on the NEMS resonator, allowing us to perform the mass evaluation. We can imagine that structural modifications may happen during impact, which could result in them slightly or significantly differing from their native conformation.

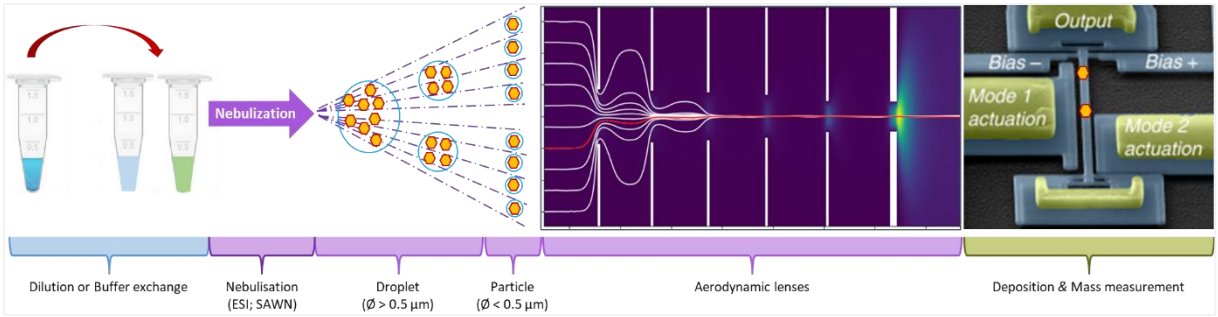


Figure 53. Schematic of the journey of viruses and Virus Like Particles (VLPs) {orange hexagons} during the Nano-Electro-Mechanical Sensor Mass Spectrometry (**NEMS-MS**) experiments. In liquid phase (underlined in blue), i.e. in buffer or in diluent solutions, we evaluate the preparations prior to nebulisation, by checking the stability and purity of the samples. In the gas phase (underlined in violet), first we nebulise using different methods: Electro Spray Ionization (ESI); Surface Acoustic Wave Nebulisation (SAWN). Which produces droplets (diam. > 0.5 µm), and particles (diam. < 0.5 µm). The produced particles are guided through aerodynamic lenses, e.g. the red line, forming a particle beam, which are then deposited on a NEMS detector (underlined in olive green), allowing individual particle mass measurement, e.g. orange hexagons. The image of the aerodynamic lenses was adapted from Reynaud et al.[165], and the image of the NEMS detector was adapted from Sage et al. [161].

Based on the design of the bench (Figure *53*) and the already established literature, which I will detail below, I imagined the following analytical pipeline, with the end goal of understanding how the NEMS-MS instrument can be used as a relevant analytical method for virion characterisation.

This was put in place for bacteriophage T5 Capsid-Like Particles (T5-CLPs), in order to assess the stability of viral particles structure during their passage from the buffer solution to the gas phase, and finally to the mass measurement. I will thus develop the reasoning going from how we chose the sample, to choosing the methods used to characterise it, and their limitation, closing in on the perspectives of future work to be done.

4.2 Materials: Viral samples

4.2.1 Bacteriophage T5 Capsid-Like Particles (T5-CLPs) as models

Our model viral particle in the past has been the capsid head taken from the bacteriophage T5. They were kindly provided by our collaborators at Institute for Integrative Biology of the Cell (I2BC), under the direction of Dr Pascale Boulanger, in the team that specialises on bacteriophage T5 {team bacteriophage T5 site}.

The following biological descriptions are primarily based on the thesis manuscript of Dr Nicolas Ducrot [209], which is currently confidential.

This phage belongs to the order *Caudovirales* and the family *Siphoviridae*. It is a strictly lytic bacteriophage which infects the bacterium *Escherichia coli*. To contextualize here, the American Society of Microbiology states: “There are two types of phages: lytic and temperate.

Strictly lytic phages infect their host cell and cause it to burst, thus killing the bacterium. Temperate, or lysogenic, phages don't kill their bacterial prey outright—they integrate their genome into the host cell, which can eventually lyse the cell, but this does little to immediately thwart bacterial infection." [210]

The phage T5 possesses a linear double-stranded DNA with a genome of 121,752 base pairs, contained in a 90 nm diameter icosahedral capsid with triangulation number T=13. [211,212] Attached to the capsid is a long, non-contractile tail, 250 nm in length, which enables transfer of phage DNA from the capsid to the cytoplasm of the infected host, [213,214] and is involved in recognition and binding to its bacterial receptor FhuA, an *E. coli* outer membrane protein involved in ferrochrome transport.[215] A visual representation of the phage T5 structure can be seen in Figure 54.

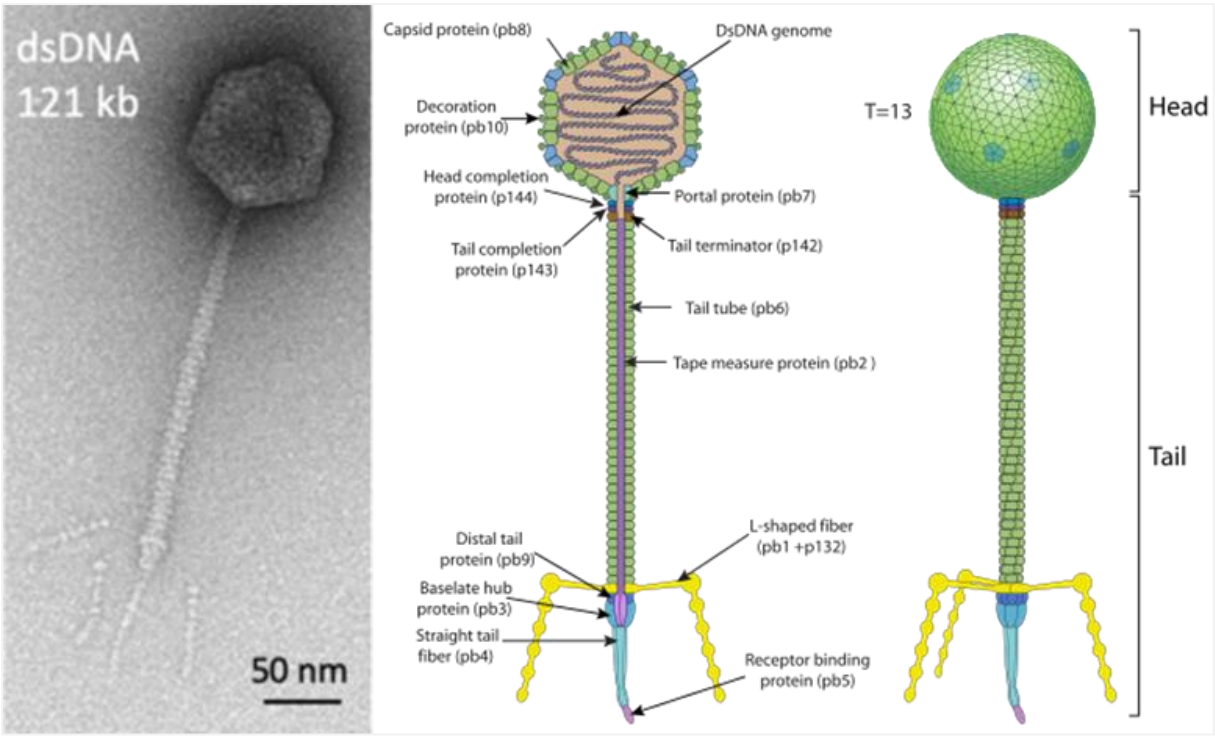


Figure 54. Visualizing the T5 bacteriophage. Left: a negative staining electron microscopy image; right: a schematic representation of the virus and the main proteins found and their location on the particle. © The TEM image (on the left) was taken from the I2BC site [https://www.i2bc.paris-saclay.fr/equipe-bacteriophage-t5/.]. The diagrams (on the right) taken from ViralZone [216][https://viralzone.expasy.org/511].

The T=13 triangulation number implies that the capsid head is an icosahedron, which is composed of 60 asymmetric units made of 13 proteins, amounting to a total of 780 capsid proteins. In the case of T5, an asymmetric subunit is formed by 2 hexamers of the major capsid protein (MCP; corresponding to pb8 in Figure 54) - one in contact with a vertex (formed either by a pentamer of MCP or by the portal protein), and a second in contact with hexamers only, plus an additional copy of MCP from a pentamer, making 13 copies of MCP for each

asymmetric units. [209] Thus, the capsid head of T5 phage is made up of 775 copies of the MCP (pb8), and 12 copies of the portal protein (pb7), where these last assemble to form a dodecameric ring and enable the genome to be encapsidated. The T5 capsid also possesses a decorator protein (pb10), 120 copies of which bind to the capsid surface at the centre of the MCP hexamers, as shown in the 3D reconstruction of the capsid in Figure 55. The assembly of the T5 capsid heads passes through several intermediates, such as prohead I, prohead II, and lastly the mature capsid, details of which are given below in Figure 55.

4.2.1.1 T5-CLP head assembly mechanism

As mentioned previously, the production of T5 capsid heads has a several intermediates, as shown in Figure 55. Here we go from the elementary proteins that compose these capsids, to the final Capsid-Like Particles (CLP).

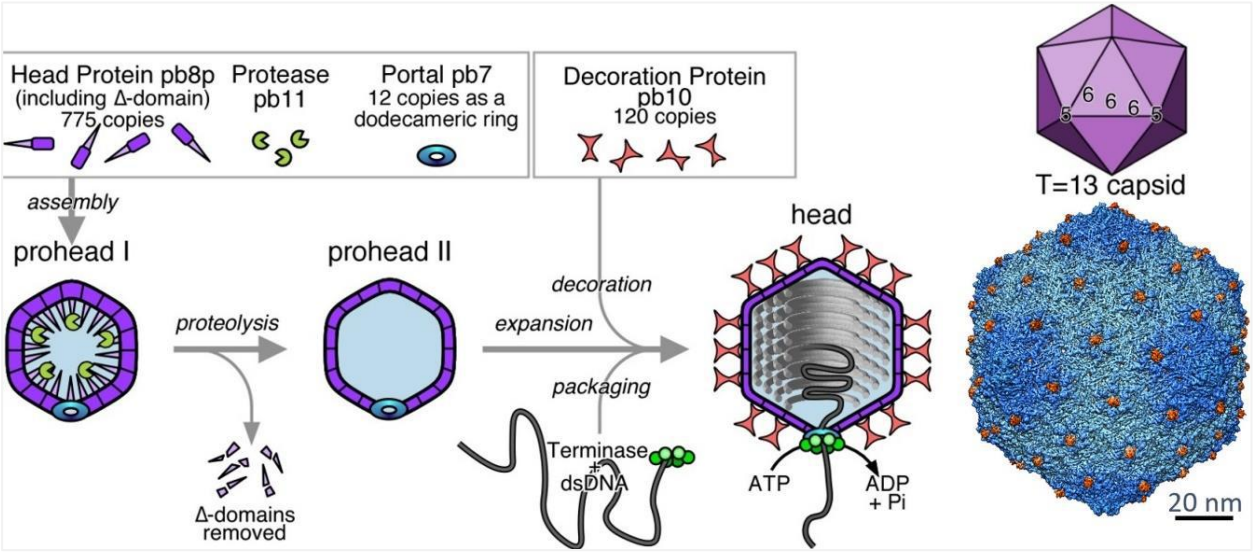


Figure 55. Phage T5 capsid assembly diagram. The precursor form of pb8, pb8p, comprises a scaffolding domain (domain-Δ, amino acids 1 to 159) represented by a white triangle, and a mature domain (pb8m, amino acids 160 to 458) represented by a purple rectangle. It is co-assembled with a dodecameric ring of the pb7 portal protein, represented by a blue ring, to form prohead I. The maturation protease, pb11, shown in green, cleaves the Δ-domains contained in prohead I to form prohead II, ready to receive the phagic genome. Encapsidation is mediated by the terminase complex and induces expansion of prohead II, followed by decoration of the capsid surface by the decorator protein, pb10, shown in red. On the right is an iso-surface of the T5 phage capsid obtained by cryo-electron microscopy, showing T=13 icosahedral symmetry. Figure taken from Ducrot.[209], and adapted from Huet et al.[211]

Initially, pb8 is produced in its precursor form, pb8p (50.8 kDa), which contains a scaffolding domain (the Δ-domain), consisting of the first 159 N-terminal amino acids. This Δ-domain is necessary for the assembly of the initial prohead I form, in which pb8 is organized into 120 hexamers on the faces, and as pentamers in 11 of the 12 vertices of the icosahedron. The twelfth vertex is formed by a dodecameric ring of the portal protein (pb7). pb7 assembles into a dodecameric ring to form the channel for genome encapsidation during capsid assembly and subsequent ejection during infection. The assembly sequence of prohead I and the

regulatory elements enabling insertion of the portal complex at only one of the capsid vertices are not known. This first stage forms the intermediate known as prohead I, within which are found the Δ-domains of all 775 copies of pb8p.[209]

In a second step, these Δ-domains are cleaved and digested by the T5 maturation protease, pb11 (26 kDa), and ejected from the capsid. This cleavage by pb11, during the maturation of prohead I, yields the mature form of pb8, pb8m (32.8 kDa). It is also known that pb11 interacts with pb7 by cleaving the first 10 N-terminal amino acids of the latter. Although the chronology of this proteolytic maturation is not known, it is suspected to be involved in the assembly of an initiator complex.[212] During its action, pb11 self-proteolyses at both ends, leaving only a few copies of a proteolyzed form (15.3 kDa) found in mature capsids, but the role of these proteolytic maturations is not known. This second intermediate capsid is called prohead II and is capable of receiving the phage genome. Finally, the genome is encapsidated in prohead II, which then forms the mature capsid head.

4.2.1.2 Production of phage-derived Capsid-Like Particles (phCLPs)

It is possible to produce DNA-free T5 capsids using the mutant phage T5amN5Δdec, where am: amber mutation and Δdec: without the decoration gene. This amber mutation (UAG stop codon) is in the T5AmN5Δdec phage terminase gene and is therefore a conditional mutation, that takes effect only in non-suppressive strains (e.g. *E. coli F*). This mutation renders them unable to encapsidate their genome.[217] Infection of *E. coli F* with this phage leads to the accumulation of prohead II in the bacterium, which is released into the medium after lysis of the infected bacterium. Infection of *E. coli* CR63 with phage T5AmN5Δdec results in phage amplification. The prohead II produced can then be purified and expanded in vitro to obtain mature T5 capsids containing no DNA (Figure 56C,D). This is the pathway for infection/phage-derived Capsid-Like Particles - CLPs, resulting in what we will note as phCLPs.

4.2.1.3 Production of recombinantly-derived Capsid-Like Particles (recCLPs)

Several studies have been carried out by the I2BC for more than a decade in order to explore the possibility of producing proheads I and proheads II from the elementary proteins forming the capsid, either in vitro from purified proteins, or by overexpressing in *E. coli* the pb7, pb10, pb11, and pb8 genes carried by a plasmid. Initial production trials of prohead II, including the decoration protein pb10 suggested that capsids could be assembled recombinantly.[211,218] The four genes pb7, pb10, pb11 and pb8 had been cloned as they are organized, in sequence, within the same operon in the viral genome. This expression system

was optimized and rendered highly efficient during the thesis of Nicolas Ducrot [209], resulting in the production of recombinant CLPs (recCLPs) (Figure 56A) in similar yields to those of phCLPs.

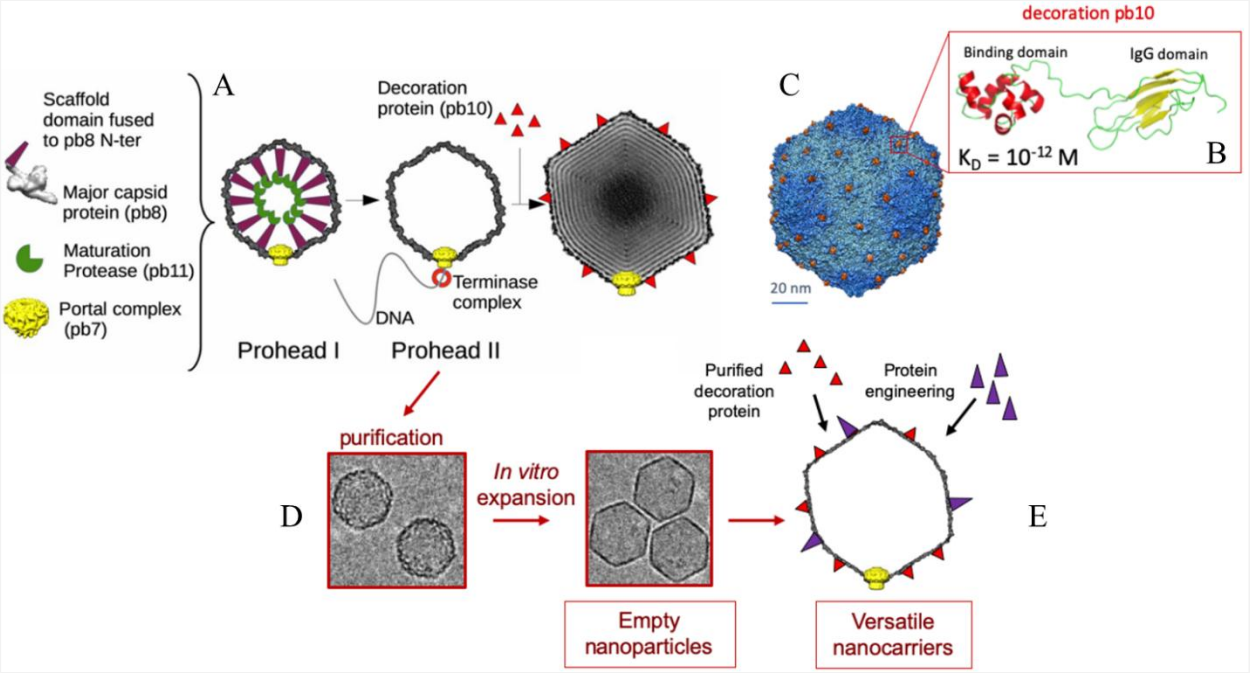


Figure 56. Production of empty T5 capsids for use as a modular platform for protein presentation. Figure was taken from the I2BC site [here], adapted from [209,217,219,220].

The details of this work are resumed by the thesis of Dr Nicolas Ducrot, which is currently confidential[209], and their protocol for recCLP production is patented (PCT/FR2020/051628).[221]

The next critical step, what makes these CLP suitable for vaccine usage is the decoration protein attached on the capsids, which I will discuss on the next paragraph.

4.2.1.4 Particle decoration

Although the decoration protein pb10 was supposed unlikely to be involved in capsid assembly, since it binds to the mature expanded capsid, it could not be ruled out that translation-regulating elements were present in its sequence and essential for capsid assembly. That is because during the infectious cycle, and once its assembly is complete, the T5 capsid is decorated by pb10, 120 copies of which bind to the capsid surface, each at the centre of one of the pb8 hexamers. The pb10 protein consists of two domains linked by a flexible linker i) the N-terminal domain, (~9 kDa), involved in interacting with the capsid; and ii) the C-terminal domain (also ~9 kDa), which is a solvent-exposed Immunoglobulin-like (Ig-like) domain.[219] The modular nature of pb10 was demonstrated in the thesis of Emeline Vernhes, resulting from

the its two-domain organization.[219,222] There she showed that the C-terminal domain can be effectively modified by replacing it with a heterologous protein, or by grafting this protein onto it, without affecting the affinity of the N-terminal domain for its capsid binding site (Figure 56B,E). The pb10-capsid binding occurs solely by a molecular recognition mechanism. The particularity of this protein is that the pb10-capsid interaction affinity is characterized by a dissociation constant (K_D) of ~10^{-12} M (pM).[219] This is a very high affinity, much higher than the high-affinity antibodies (typically K_D <1 nM)[223], which makes pb10 binding to the capsid quasi-irreversible.

Decoration can be replicated in vitro by incubating undecorated expanded capsids with purified pb10 protein. Under these conditions capsid decoration is complete when capsids are incubated with pb10 at a 1:1 ratio, equivalent to one pb10 protein per binding site, i.e. 120 copies of pb10 per capsid. It was observed that if capsids are incubated with insufficient pb10, two populations will be observed: either fully decorated capsids on which all 120 sites are occupied by the decoration protein, or undecorated capsids on which no sites are occupied. The fact that partially decorated capsids cannot be demonstrated suggests a highly cooperative process, and decoration is quasi-irreversible given the affinity of the N-terminal domain of pb10 for its binding site ($K_D \approx 10^{-12}$ M).[219] This can be more graphically visualised in Figure *57*, where we have the horizontal separation by the production method: Inf (by infection), or phage derived, (phCLP), or Rec (by recombination, thus recCLP. And the vertical separation is the: absence (-) or presence (+) of decoration.

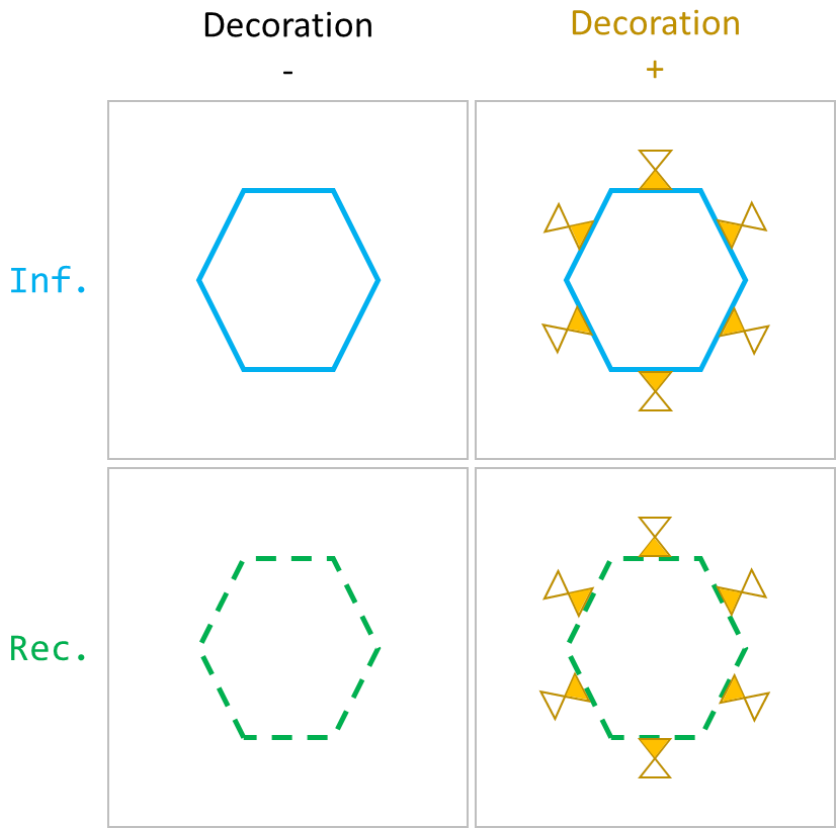


Figure 57. The different type of T5-CLPs analysed. The production method (in horizontal): Inf (by infection, or phage derived, phCLP), in solid blue lines; Rec (by recombination), in dashed green lines. Decoration: absent (-) or present (+), is presented by the yellow hourglass figures. The solid triangle indicates the N-terminal that has a high affinity for the capsid and is the conserved part.

From a larger biophysical perspective (i.e. at the scale of the whole particle), the expansion step increases capsid diameter from 70 nm, for prohead II, to 90 nm, for the expanded mature capsid, doubling the internal volume of the capsid, and enabling it to contain the entire genome. Expansion also stabilizes the capsid: mature capsids can withstand temperatures of up to 96 °C, whereas prohead II is only stable up to 77 °C.[217] Compression of the genome inside the capsid generates an internal pressure of several MPa and concentrates the DNA to ~500 mg/mL.[224]

All these aforementioned properties make the capsid of bacteriophage T5 a very interesting case study as a model VLP. We can state here that it is a large and highly stable particle, very suitable for our applications in native mass spectrometry, but not only that. In addition, it is also a well characterized VLP and fairly simple biological nanoparticle to work with. Now let us see how knowing more about the biology of these particles can help us better understand this model VLP, but also how it can be used as a future vaccine particle.

In the following, I will focus on the methods usually used to characterize VLPs or even inert nanoparticles that serve as simpler models. I will particularly discuss how I tried to implement a robust analytical pipeline for the analysis of these VLPs, which are important for

future vaccine applications. Additionally, I will explore an inter-method approach to improve particle characterization prior to performing NEMS-MS analysis.

4.2.1.5 Why CLPs are interesting for the mass measurement by NEMS-MS?

In order to understand really how mass can be used as a relevant analytical method for these CLPs, the following table, Table 7, allows seeing the mass differences of its macromolecular components. More details on their calculation can be found in the annexes, in section 'production of T5 CLPS' (page 288).

Table 7. Description of the analyses of the T5-CLPs, from a macromolecular point of view. Legend of the highlights: blue (capsid related); yellow (decoration related); orange (genome related). The table was built based on references [21,184,211].

Analyte	Accession in Uniprot	One copy (Da)	copy number	Total mass (Da)
pb7 protein	PORTL_BPT5	43 897	12	526 764
pb8 protein	CAPSD_BPT5	32 892	775	25 491 300
pb10 protein	DECO_BPT5	18 300	120	2 196 000
Ovalbumine protein	OVAL_CHICK	43 000	120	5 160 000
pb114 protein	PRO_BPT5	19 210	12	230 520
Genome (dsDNA)	NC_005859 (**NCBI-Id**)	650	121750	79 137 500

Which gives the final table, Table *8*, containing all the possible samples that can be generated and their expected theoretical mass, as follows:

Table 8. The combinatory possibilities of the possible T5-CLPs and their respective total expected mass, based on sample composition. Are as follows: a) Capsid produced by infection (phage derived) or recombination {inf / rec}; b) Genome content filled or empty {fill / empt}; c) Decoration protein pb10 is present: {deco+} / absent: {deco-}; d) Ovalbumin is present: {Ova+} or absent: {Ova-}. The colour coding is the same as that in Figure 57.

Sample number	Sample composition	Total mass sum (Da)
1	inf_fill_deco+_Ova+	112 742 084
2	inf_fill_deco+_Ova-	107 582 084
3	inf_fill_deco-_Ova-	105 386 084
4	**inf_empt_deco+_Ova+**	**33 374 064**
5	**rec_empt_deco+_Ova+**	**33 374 064**
6	inf_empt_deco+_Ova-	28 214 064
7	rec_empt_deco+_Ova-	28 214 064

8	**inf_empt_deco-_Ova-**	**26 018 064**
9	**rec_empt_deco-_Ova-**	**26 018 064**

We expect that the NEMS-MS approach should enable easy assessment of CLP integrity and antigen loading, two key requirements for future pharmacological developments, for example in CQA (Critical Quality Attributes). Our NEMS-MS approach, has already shown its proof of concept in measuring the mass of sample #3 and #8 in the work of Dominguez-Medina et al.[21], and we showed the desolvation properties of sample #8 in the work of Clement et al.[184]

4.2.1.6 Naming of the T5-CLP samples

I would like to make a parenthesis, in order to explain the naming of the samples for the rest of this manuscript. During my thesis, I acquired three batches of sample #8, and the other samples (in bold in Table *8*), were acquired in November 2022. This inspired the naming and the classification as seen below, Figure 58.

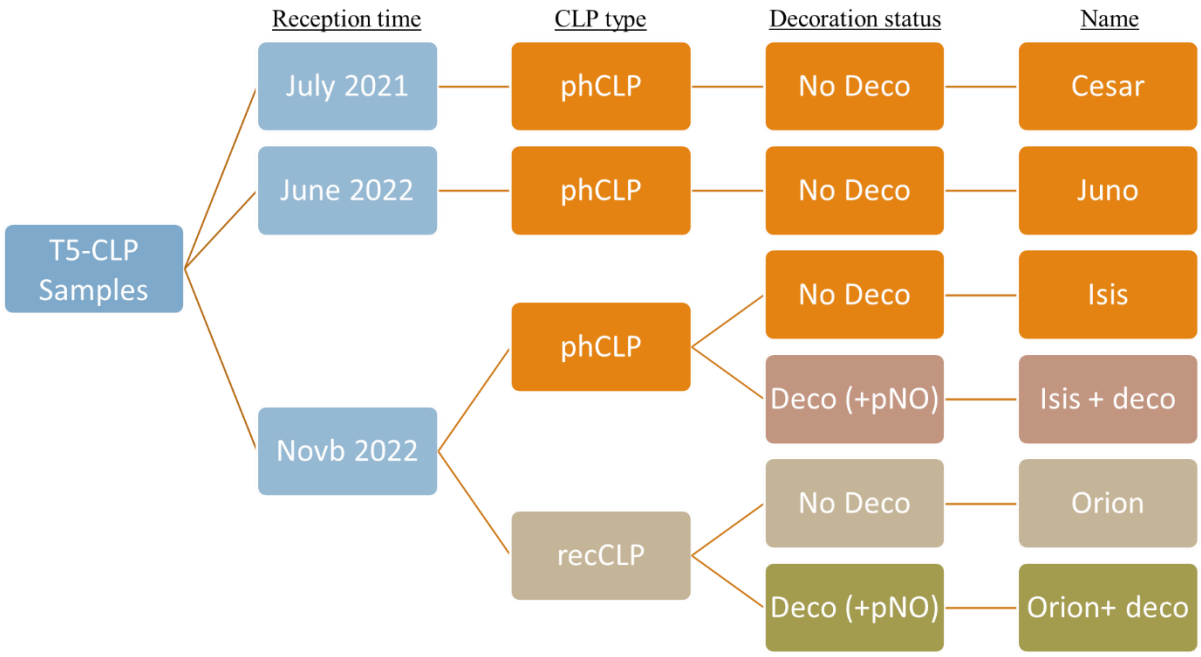


Figure 58. Schematic of the samples and their naming. Legend (seen on top of each column, read from left to right). Reception time: the month and year when the sample was received. CLP type: the capsid method of production {phage derived (phCLP) or by recombination (recCLP)}. Decoration status: if the decoration {with pb10- N-terminal and Ovalbumine (pNO)} was present [deco (+pNO)] or absent (No deco). Name: inspired from the month of the year we got the samples, from Roman history and mythology.

Therefore, the relationship between the sample naming (and their numbers in bold in Table *8*) and the rest of this work are as follows:

- #4 : phCLP with deco: Isis + deco

- #5 : recCLP with deco: Orion + deco
- #8 : phCLP without deco: Cesar; Juno; Isis
- #9 : recCLP with deco: Orion

This nomenclature will be respected for the rest of the manuscript, and I will refer to it when talking about the analysis of the T5-CLPs, especially in Chapter 5.

But this was not the only samples that I had the opportunity to work with during my thesis. In order to increase the catalogue of available samples and to open the possibly for future collaborations, I managed to acquire several other samples, which I will briefly detail below.

4.2.2 Other viral samples and Sample related limitations.

The search for other viruses and virus like particles was one of the main goals for my last year of PhD. The main reason was to first increase the possible viruses that have been measured with the NEMS-MS instrument, allowing a larger variety of possible samples to work with, and thus more opportunities to not be limited due to sample acquisition. In addition, I wanted to add my own experiential measurement to the theoretical results presented in Chapter 1, regarding the mass of Human Respiratory Viruses (HuReVs), using the NEMS-MS approach. Lastly, due to my difficulties of acquiring samples from our collaborators in Paris at I2BC, primarily because of the COVID pandemic, I realized that having more local collaborators could help in having faster sample allocation, especially in times of dire necessity.

My first intuition was thus to look into purchasing commercially available viral vectors and vaccines.

4.2.2.1 Industrial samples

Regarding the industrial samples, some commonly and extensively analysed samples by charge detection mass spectrometry (CDMS) methods are the adeno associated viruses (AAV), especially focused on quantifying the genome content inside these viral vectors.[225,226] Unfortunately, these viruses are in the lower part of the dynamic range of our NEMS-MS system, i.e. ~ 3-5 MDa, which would make the mass analysis very complicated with the existing system.

Another potential lead was to acquire commercial vaccines and trying to evaluate the mass of the particles. This has already been shown to work well with CDMS, more specifically for the following vaccines: IPOL (Sanofi Pasteur SA), which is an inactivated poliomyelitis whole virus vaccine; RotaTeq (Merck), a live-attenuated Rotavirus; and Gardasil-9 (Merck),

recombinant vaccine made from nine strains of the human papillomavirus's (HPV-VLPs).[142] Unfortunately for us, their low particles per volume concentration (e.g. 5-80 $\times 10^6$ infectious units/mL in the RotaTeq vaccine)[142] was much lower than the limit of detection of our first instrument (i.e. 1 viral particle per 2.6×10^8 particles)[21]. The new NEMS-MS instrument having a 200-fold improved limit of detection (i.e. 1 per 2.5×10^6 particles)[165] can perhaps be used in the future to measure such samples. In addition, it was difficult due to logistic problems to acquire these samples in our lab, in a secure and conform manner.

For this reason, I decided to contact a number of Grenoble-based labs, in order to initiate future collaboration and be able to acquire samples rapidly and efficiently.

4.2.2.2 Local collaboration

Even though I had the opportunity to contact a number of local labs that were very willing to collaborate with me in this initiative, it proved to be very difficult to have satisfactory samples for a multitude of reasons, as follows.

The primary reason was the difficulty of finding pure enough samples that we could work with in order to benchmark our analysis. For example, we got two samples from Abiolab-Asposan (Montbonnot-Saint-Martin, 38330 France), the MS2 bacteriophage and the murine norovirus serotype 5 (MNV-5), which were retrieved directly from their hosts, no purification steps, thus containing a multitude of cytosol proteins and other contaminants from the host cells. Both of these samples have been well studied by native MS, resulting in a good reference point for comparison to NEMS-MS.[227,228] Some TEM images that we were able to acquire for these samples are shown in the supplementary materials, respectively Figure 130 for MNV-5 and Figure 131 for MS2, better detailing the difficulties that we encountered with each of these respective samples.

The second main reason that this sample acquisition initiative proved difficult was the administrative and security level difficulties. I was able to contact two labs in Grenoble that could produce human adenoviruses- HAdV (Pascal Fender Team, Institut de Biologie Structurale IBS) and human/monkey chimeric lentiviruses (PAVAL lab - Pathogénèse et vaccination lentivirales). The difficulties here were largely because since these viruses could be potentially infectious to humans, it immediately made them not adapted to the security level that we could provide in our lab. This caused us to perform a variety of crosslinking and inactivation steps, which proved detrimental to the structural stability of the forrticles, thus rendering them useless to further analysis. One such example is shown in the supplementary data (Figure 132), regarding the HAdV-5 analysis by TEM. I have opted here not to include

more data regarding these samples, in order to maintain cohesion and conciseness in our descriptions.

Therefore, I will mainly focus on the rest of this thesis on the methods and results that we acquired for the T5-CLPs.

4.3 Summary of the methods used for VLP characterisation

The following is a catalogue of the different methods that are commonly used for virus analysis. First, I will do an overview of some of the literature, and then focus on the methods already used on T5. Lastly, I will discuss about the methods I chose and why.

4.3.1 The literature on virus characterisation methods

In the review of Nooraei et al. (2021)[229], regarding the Preparation, immunogenicity and role of Virus-like particles (VLPs) as nano-vaccines and drug nanocarriers, a detailed examination of the different methods that can be used to analyse VLPs in the context of future human usage was made. Three main categories of analytical characterisation (Figure 59) were described: **A)** Biochemical; **B)** Biophysical; **C)** Biological.

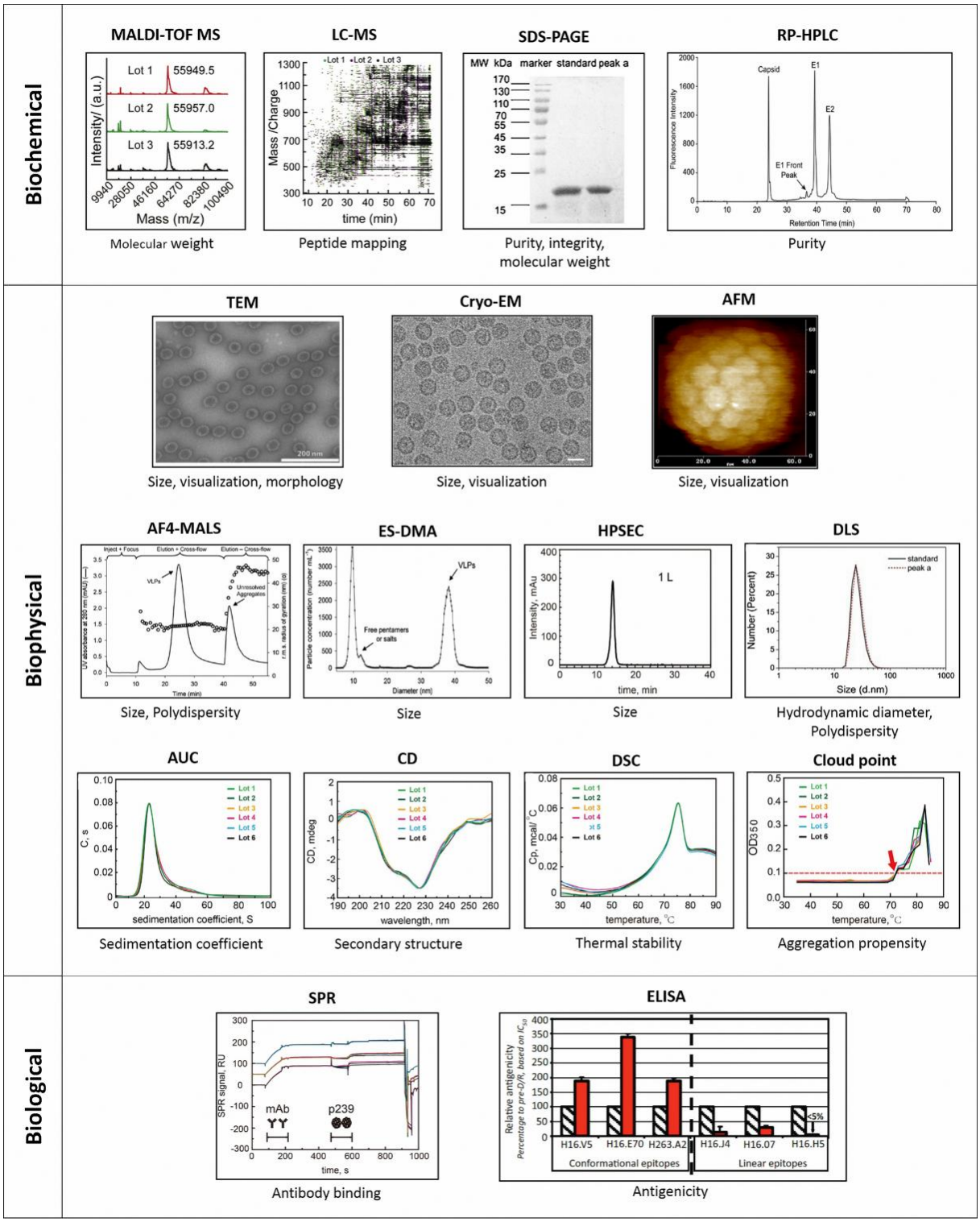


Figure 59. Analytical techniques for characterization of VLPs. A series of analytical tools has been used for biochemical, biophysical and biological characterization of VLPs. Biochemical method: Matrix assisted laser desorption ionization time of flight-mass spectrometry (MALDI-TOF MS), liquid chromatography–mass spectrometry (LC–MS), sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and reverse phase-high performance liquid chromatography (RP-HPLC). Biophysical method: Transmission electron microscopy (TEM), cryo-electron microscopy (Cryo-EM), atomic force microscopy (AFM), asymmetric flow field-flow fractionation coupled with multiple-angle light scattering (AF4-MALS), electrospray differential mobility analysis (ES-DMA) and high-performance size exclusion chromatography (HPSEC), dynamic light scattering (DLS), analytical ultracentrifugation (AUC), circular dichroism (CD), differential scanning calorimetry (DSC), Cloud point. Biological characterization: Surface plasmon resonance (SRP), enzyme-linked immunosorbent assay (ELISA). The figure and description are taken from original, Nooraei et al.[229]

First, regarding the Biochemical methods, they stated the following methods, as ordered in (Figure 59): (**A.1**) Matrix assisted laser desorption ionization time of flight-mass spectrometry (MALDI-TOF MS) is a soft ionization technique, used for analysing biomolecules like proteins, peptides, and oligonucleotides. The sample is co-crystallized with an energy-absorbing matrix on a metal plate. A laser pulse then desorbs and ionizes the sample, and the ions are accelerated through a flight tube. The time of flight inside this tube is measured to determine the mass-to-charge ratio, generating a mass spectrum. The primary use of the technique is to get the molecular mass of proteins in the sample. (**A.2**) Liquid chromatography–mass spectrometry (LC–MS) is an analytical technique that combines liquid chromatography, for separating the analytes of a mixture, with mass spectrometry for identifying the separated analytes. It allows identification and structural characterization of molecules, such as peptides and lipids, based on their mass-to-charge ratios. The primary use of the technique is to get the peptide mapping of the proteins in the sample. (**A.3**) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is a technique used to separate proteins based on their electrophoretic mobility, which is determined by their molecular weight. Proteins are denatured by beta mercaptoethanol (βME), which cuts disulfide bridges, and SDS negatively charges the proteins, that could be then considered as a negatively charged denatured proteins. Then they are separated, by applying an electric field through a polyacrylamide gel matrix, and thus will migrate differently as function of the mass of the subunits of the protein. The primary use of the technique is to verify the purity, integrity and get the molecular mass of the proteins in the sample. (**A.4**) Reverse phase-high performance liquid chromatography (RP-HPLC) is a chromatographic technique used to separate molecules based on their hydrophobicity. Here stationary phase is non-polar, while the mobile phase is polar, therefore resulting in the more non-polar molecules interacting more strongly with the stationary phase and thus eluting later. The primary use of the technique is to verify the purity of the sample.

Secondly, there are the Biophysical methods, which are more numerous than the biochemical and biological methods combined, and also provide the largest number of characterisations, and they are as follows, ordered as in Figure 59: (**B.1**) Transmission electron microscopy (TEM) uses a beam of electrons transmitted through a thin specimen to form an image. It allows visualization of fine details and structures of biomolecules like proteins at near-atomic resolution. The primary use of the technique is to visually verify the purity of the sample, and also get physical information about the particles size and morphology. (**B.2**) Cryo-electron microscopy (Cryo-EM) involves imaging frozen-hydrated samples at cryogenic temperatures using an electron microscope. It enables structural determination of biomolecules like proteins

and nucleic acids in their native state. The primary use of the technique is to visually observe the sample, and also get the particles size. (**B.3**) Atomic force microscopy (AFM) scans the sample, using a sharp probe (with a tip point as small as few atoms in width) over a sample surface to generate a topographic map, based on the interaction of forces between the probe and the surface. It can visualize the 3D structure of biomolecules at near-atomic resolution, and acquire their size and morphology. (**B.4**) Asymmetric flow field-flow fractionation coupled with multiple-angle laser light scattering (AF4-MALLS-RI) separates and characterizes biomolecules based on their diffusion coefficients and molar masses, without shear degradation. Using the Refractometer (RI) to obtain the protein concentration, together with the light scattering information, thanks to the Raleigh equation a calculated mass can be deduced. The primary use of the technique is to acquire the particle size and polydispersity. (**B.5**) Electrospray differential mobility analysis (ES-DMA) measures the gas-phase mobility and size of ions, generated by electrospray ionization, often in order to determine biomolecular structure and dynamics in the gas phase. This technique can also be found coupled to condensation particle counter (CPC), which quantifies the number particles in an aerosol per unit volume. This combined DMA+CPC instrument can be found commercially, trademarked under the name Scanning Mobility Particle Sizer (SMPS™) spectrometers, which I will detail later in this chapter. (**B.6**) High performance size exclusion chromatography (HPSEC) separates biomolecules based on their hydrodynamic volume by passing them through a porous matrix. The primary use of the technique is to evaluate virus size in solution. (**B.7**) Dynamic light scattering (DLS) determines the size distribution of biomolecules in solution by measuring fluctuations in scattered light intensity caused by Brownian motion. This is a bulk measurement, that provide information about the overall hydrodynamic size of particles, and their polydispersity. (**B.8**) Analytical ultracentrifugation (AUC) measures the sedimentation and diffusion of viruses in a centrifugal field to determine their molar masses and shapes. (**B.9**) Circular dichroism (CD) probes the structural features of biomolecules by measuring the differential absorption of left and right circularly polarized light. The primary use of the technique is to evaluate the secondary protein structure. (**B.10**) Differential scanning calorimetry (DSC) measures the heat absorbed or released by biomolecules as a function of temperature to study their thermodynamic stability. The primary use of the technique is to measure the melting temperature of biomolecules when they are denatured, informing on their thermal stability. (**B.11**) Cloud point measure is the temperature above which a biomolecular solution becomes turbid due to aggregation or phase separation.

Lastly, there are the biological methods, as ordered in Figure 59: (**C.1**) Surface plasmon resonance (SPR) is a label-free technique used to study biomolecular interactions in real-time. One molecule (ligand) is immobilized on a sensor chip, and binding of another molecule (analyte) to the ligand is measured by detecting changes in the refractive index at the sensor surface. SPR provides real-time binding data without labels, often used for evaluation of antibody binding. (**C.2**) Enzyme-linked immunosorbent assay (ELISA) is a widely used analytical technique that uses antibodies and colour change to identify and quantify substances like peptides, proteins, antibodies, and hormones. ELISA is well suited for high-throughput screening and quantification of analytes in complex samples, and testing antigenicity.

As mentioned earlier, the biophysical methods allow for the largest number of observations of viral particles. In addition, my main interest was to look at viruses at a single particles level, as this is what the NEMS-MS acquires as well. That is why, Bruinsma et al. (2021)[230] propose in this review a list of single-particle techniques, primarily with the purpose of providing insight into the mechanics of viral assembly (Figure 60).

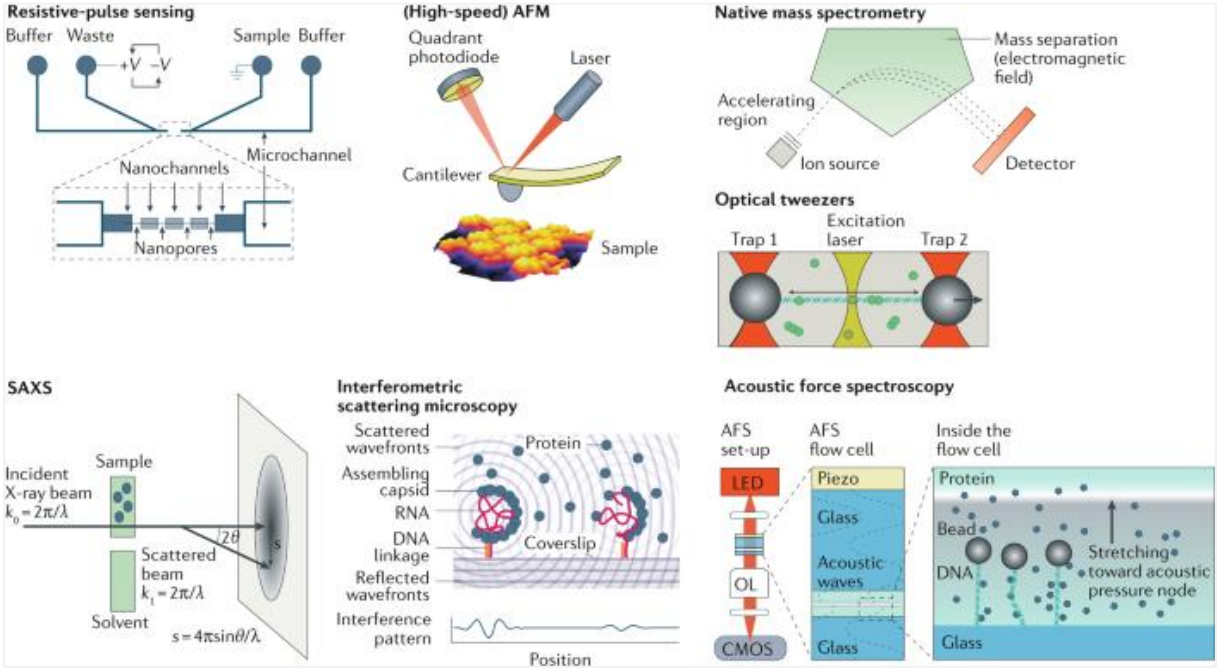


Figure 60. Experimental techniques described in the review by Bruinsma et al. [230]. Legend: resistive-pulse sensing (RPS), atomic force microscopy (AFM), small angle X-ray scattering (SAXS), acoustic force spectroscopy (AFS), light-emitting diode (LED), objective lens (OL), complementary metal–oxide–semiconductor (CMOS) camera.

Resistive-pulse sensing (RPS) is a technique that measures changes in ionic current across a nanopore as individual molecules or particles pass through it. It allows characterizing the size, concentration, and surface charge of biomolecules like proteins and nucleic acids. High-speed atomic force microscopy (HS-AFM) enables direct visualization of biomolecular processes at high spatiotemporal resolution by rapidly scanning a probe over the sample surface.

This has the advantage of imaging dynamics of single molecules and studying interactions in physiological conditions. Native mass spectrometry, including charge-detection mass spectrometry (CDMS), allows studying intact protein complexes and assemblies in the gas phase by gently transferring them from solution to the mass spectrometer. It provides information on stoichiometry, topology, and dynamics. Optical tweezers use a highly focused laser beam to trap and manipulate microscopic dielectric particles like biomolecules with high precision. They enable measuring piconewton (pN) forces and studying mechanochemical processes at the single-molecule level. Small Angle X-ray Scattering (SAXS) is a solution technique that provides low-resolution structural information on the size, shape, and assembly state of biomolecules and complexes in a label-free manner. Interferometric scattering mass spectrometry, exemplified by Mass Photometry (MP), correlates the mass and physical properties of individual biomolecules in solution by combining interferometric tracking with mass spectrometry. A notable improvement to MP, especially for virus analysis was recently published, named RYtov MIcroscopy for Nanoparticles Identification (RYMINI). This optical approach merges holographic label-free 3D tracking with high-sensitivity quantitative phase imaging, into a compact optical setup that can be used in highly demanding environments, such as level 3 (L3) biological laboratories.[231] Acoustic force spectroscopy (AFS) applies acoustic radiation forces to manipulate and characterize biomolecules in solution, such as proteins and nucleic acids, based on their compressibility and density. It enables label-free analysis of structure, folding, and interactions.

From the world of inert nanoparticle analysis, the comparative power of in solution methods, primarily comparing the light scattering based technologies to other new methods was described in the perspective article of Maguire et al. (2018)[232], as shown in Figure 61.

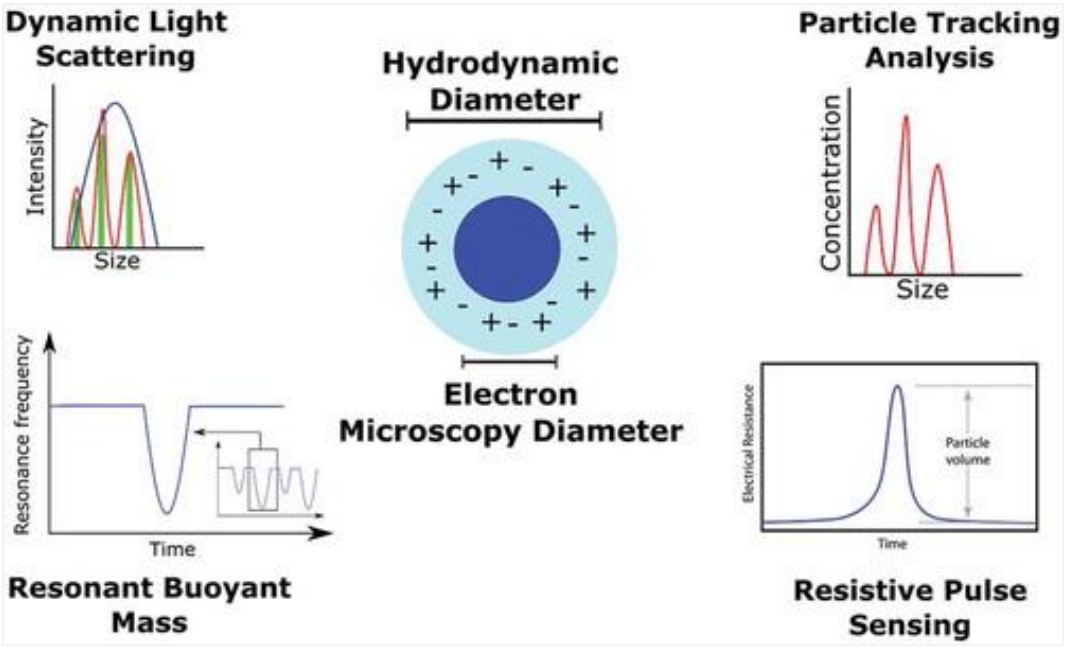


Figure 61. Comparison of different methods for sizing. Here the evaluation is compared between dynamic light scattering (DLS); particle tracking analysis (PTA); resonant buoyant mass; Resistive Pulse Sensing (RPS). In the centre the difference between the hydrodynamic diameter and the electron microscopy diameter is recalled. Graphical abstract taken from Maguire et al. [232]

Dynamic light scattering (DLS) and particle/nanoparticle tracking analysis (PTA/NTA) are light scattering techniques used to determine the size distribution of particles in solution based on their Brownian motion. PTA/NTA directly tracks the Brownian motion of individual particles using a microscope and camera. The diffusion coefficient and hydrodynamic diameter are calculated for each tracked particle, providing higher resolution for polydisperse samples compared to DLS. In addition, NTA can also provide concentration measurements, by counting particles for a defined volume, in the focal plane of the microscope, and extrapolating this number to the total particle count. As it was described earlier (Figure 59), DLS provides information on diffusion coefficient and particle's hydrodynamic diameter, by measuring the fluctuations in the intensity of scattered light over time. However, DLS is biased towards larger particles due to the intensity being proportional to the particle diameter in power 6 (d^6). Lastly, resonant buoyant mass is a technique that measures the buoyant mass of individual nanoparticles in solution using a suspended microchannel resonator. It provides high-resolution mass and buoyant mass distributions, complementing light scattering techniques. Resistive Pulse Sensing (RPS) was already described in Figure 60.

4.3.2 State of the art on the T5-CLPs

The T5-CLPs, especially the batches from sample #6 and #8

Table *8*(Table *8*), have the advantage of having been well characterised in prior works by the NEMS-MS group and our primary collaborators at I2BC. The following is a bibliographic resume of these analyses.

From the group of I2BC, the thesis work of Dr Oliver Preux[233], focusing on the assembly and maturation, and decoration properties of bacteriophage T5 capsids, and the thesis work of Dr Emeline Vernhes[222], which focused primarily on the decoration protein (pb10). They analysed these CLPs with a myriad of methods, some of which already mentioned above, such as: MALDI-TOF, SDS-PAGE, SEC-MALS, DSC, DLS, CD, AUC, SAXS, negative staining TEM, cryo-EM, and SPR. Other methods that were used, but not listed above, were also: X-ray diffraction imaging, Fluorescence-based Thermal Shift Assay (FTSA), Nuclear Magnetic Resonance (NMR), Western blot. Nicolas Ducrot directly compared recCLPs with phCLPs, especially regarding their morphology, using different electron microscopy (EM) methods, such as negative staining and cryoEM, which allowed for the 3D reconstruction of the whole particle. The affinity of the decoration protein for these new capsids was determined by Flow Induced Dispersion Analysis (FIDA), and compared with that of phCLPs, priorly analysed by surface plasmon resonance (SPR). Their stability was measured by differential scanning calorimetry (DSC) and several biochemical approaches, such as Electrophoretic Mobility Shift Assay (EMSA); Western Blot; Dot Blot; LC-MS proteomics.[209]

From the NEMS-MS group, the thesis work of Dr Kavya Clement[202] was focused on the biochemical evaluation by LC-MS and SDS-PAGE. One of the major contributions of this work was the examination of the copy count of protease pb11, and the evaluation of the absolute mass of the capsid using proteomics. In addition, the biophysical evaluation with Nanotracking analysis (NTA) and NEMS-MS, allowed for checking the state of the CLPs prior to performing mass evaluations.

So now that we have seen the methods that are available for single particle virus analysis, and also seen the state of the art on the T5-CLPs in the past, let us now see how I integrated these information into building the analytical pipeline described below.

4.4 Design of the in-line analysis prior to the NEMS-MS bench

Based on all the aforementioned literature, I was able to propose the following experimental design, in order to assess the stability of VLPs structure during its passage from the buffer solution to the gas phase, and finally to the mass measurement. This experimental design will serve as an evaluation pipeline, containing several checkpoints, by appropriate orthogonal methods. A schematic representation can be found below, in Figure *62*.

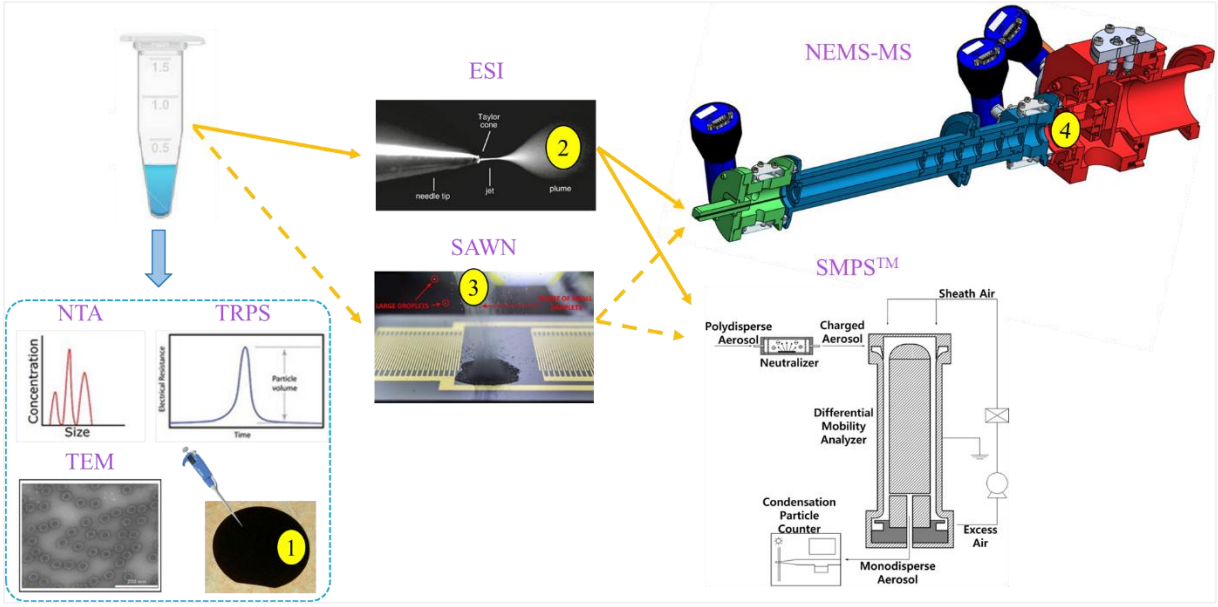


Figure 62. Schematic of the analytical pipeline of the Biophysical analysis of T5 capsids, prior to the Nano-Electro-Mechanical Sensor Mass Spectrometry (**NEMS-MS**) instrument. In liquid phase (blue dashed square), we use: **NTA** (Nano-Tracking Analysis) or **TRPS** (Tunable Resistive Pulse Sensing); and also negative staining **TEM** (Transmission Electron Microscopy). For deposition on silicon wafer (yellow circle #1): **SEM** (Scanning Electron Microscopy) or **AFM** (Atomic Force Microscopy). In the gas phase, first we nebulise using different methods: Electro Spray Ionization (ESI; solid orange arrows); Surface Acoustic Wave Nebulisation (SAWN; dashed orange arrows). Then, we measure the particle electrodynamic diameter, using Scanning Mobility Particle Sizer (**SMPS™**), and check on deposition (#2 & #3). The produced particles are guided through the **NEMS-MS** bench and can also be deposited on a NEMS detector or silicon wafer (#4). Images were taken from: NTA & TRPS ([232]); TEM ([229]); ESI ([166]); SAWN ([167]); SMPS™ ([234]).

First, we evaluate the stability of VLPs in the solution phase, i.e. in buffer or in a diluent solution, before nebulisation, using NTA (Nano Tracking Analysis) and TRPS (Tuneable Resistive Pulse Sensing). This has to be compared to negative staining TEM (Transmission Electron Microscopy), given that is the method of reference for virus analysis. This sample can also be deposited on a silicon wafer (denoted by the yellow circle #1), to further be analysed by adapted methods such as SEM (Scanning Electron Microscopy) or AFM (Atomic Force Microscopy). This is interesting also as it can be directly compared to TEM, since it is also a deposition method.

The next step, after we have nebulised with different methods, was to determine the efficiency of the different particle aerosolization strategies and the effect that they may have on

the particles structure. That is why we analysed them with SMPS™ (Scanning Mobility Particle Sizer). We also thought of combining this method to APS™ (Aerodynamic Particle Sizer) in order to evaluate the generated droplet's size and concentration in aerosol. This serves to evaluate all the step of the virus in aerosol, as presented in Figure *53*. SMPS™ can also be used to perform 'inline analysis' when performing mass measurements, by splitting the nebulised particles into: one part going to the NEMS-MS bench; and the other to SMPS. This requires some dedicated instrumental modifications, depending on the nebulisation methods used.

Finally, I imagined examining particles after landing on the NEMS resonator, thus measuring the mass, and then check possible structural alterations, using SEM or AFM (wafer #4). This can also have the major advantage of knowing exactly what particle we measured the mass of, since we know their exact location deposition on the NEMS, thus having both mass and conformation for individual particles.

Additionally, the group had shown the way to get a relative mass estimation of the total particle mass, using bottom-up proteomics (biochemical method), which I will describe in greater detail below.

All these techniques are resumed in the following table, Table 9, where the information on the possible measurements that they can provide us with is given, and the way that these collaborations were foreseen.

Table 9. Table describing the different analytical methods envisioned to be used during this work, the measurements and the possible analysis that can be performed. Abbreviations (in order of appearance): NTA (Nano Tracking Analysis); TRPS (Tunable Resistive Pulse Sensing); negative staining TEM (Transmission Electron Microscopy); SMPS (Scanning Mobility Particle Sizer); Aerodynamic Particle Sizer (APS); SEM (Scanning Electron Microscopy); AFM (Atomic Force Microscopy).

Methods	Milieu	Measurement possible	Service provided
Negative staining TEM	Solution; Deposition	• Particle Size (visual) • Particle Morphology	Scientific platform/ Collaboration
NTA™	Solution	• Particle Size (hydrodynamic diameter) • Particle Concentration (in solution)	Owned by our group
TRPS™	Solution	• Particle Size (electrodynamics diameter) • Particle Concentration (in solution) • Zeta potential	Commercial service provider

Bottom-Up Proteomics	Solution	• Peptide Abundance (Intensity)	Scientific platform/ Collaboration
SMPS™	Aerosol	• Particle Size (electrodynamics diameter) • Particle Concentration (in aerosol)	Scientific collaboration
APS™	Aerosol	• Droplet size (aerodynamic diameter) • Droplet Concentration (in aerosol)	Scientific collaboration
SEM	Deposition	• Particle Size (visual) • Particle Morphology • deposition check	Scientific platform
AFM	Solution; Deposition	• Particle Size (in 3D) • Particle Morphology • Stiffness • deposition check	Scientific platform/ Collaboration

All these methods were combined in the purpose of tracking the structural stability of the particles all along the NEMS-MS bench, particularly in the idea of using the particle size as a proxy to its total mass conservation. The usage of proteomics to have an approximated result to the total mass has the validity of reaching an empirical estimation of the total mass, which is more realistic basing the calculation entirely on the literature derived data. Hereafter I will detail these tools and see how I adapted, or tried to adapt, in order to perform our analysis on the samples of interest.

4.4.1 Method related limitations, focus on SEM and AFM.

Before I go any further, I would like to state the limitations that I encounter during my thesis in order to put in place the in-line analysis described in Figure *62*. My analysis by both SEM and AFM proved to be hard to develop and optimize, and even harder to reach conclusive results, as I will describe hereafter.

I initially imagined using SEM or AFM as ‘checkpoint’ methods. This means that we can use it to control how the sample looks in solution (wafer #1), thus as a control, and then observe if we see changes on the morphology and size all along the bench, *i.e.* after we nebulise by ESI (wafer #2), or with SAWN (wafer #3), and lastly after the aerodynamic lenses, i.e. on

the NEMS themselves (or a wafer, #4).). To this end, I suggest to use an undoped silicon chip as a reference substrate (since the NEMS detectors are made of the same material) in which we can collect the sample at the previously mentioned checkpoints. The intention was to combine SEM and AFM for 'at line analysis', in order to reach cross method validation, with the intention of reaching an 'online analysis' with further instrumental development, if possible. In addition, AFM can be performed in liquid conditions, providing an additional third dimension to the particle deposition, in liquid or in dried/deposition conditions.

4.4.1.1 Problems with SEM

First, the SEM was not conclusive because direct observation of the viral particle deposition on silicon was not possible. This is mainly because the measurement is done under high vacuum (at least 10^{-4} Torr, usually 10^{-6} Torr or higher, as reported by instrument manufacturers), and thus, any traces of potential water inside the particles would result in their explosion. Therefore, a certain heating is required prior to evaluating the samples, which in our case was around 60-65 °C, for a duration of 5 minutes. This temperature is compatible for these CLPs, as their thermal stability has been already demonstrated using Differential scanning calorimetry (DSC) (supplementary Figure 135).[21] Moreover, after collecting the samples at the different steps (shown in yellow in Figure *62*), additional stages of metallic deposition are required to increase the resolution, and to find the most adapted metallic deposition technique. This is done in order to have sufficient contrast to make morphological observations, but they also need to maintain a state that could still be considered as native.

Prior work by our team used a deposition of ~2 nm layer of TiN (Titanium nitride) using atomic layer deposition (ALD), which is a subgroup of the greater Chemical vapour deposition (CVD) methods (supplementary Figure 134).[21] The issue with ALD is the high temperatures used to prepare such as deposition, e.g., in the case of TiN (Titanium nitride), the precursors need to be heated at a minimum temperatures of around 350 °C.[235] This of course would heavily degrade the native structure of the capsids, thus beating the purpose of using this technique as a native method for sample analysis.

Therefore, as an alternative approach I decided to try two classical physical vapour deposition (PVD) methods, sputtering [236] (supplementary *Figure 139*) and E-Vap, (short for Evaporation deposition) [237] (supplementary Figure 140). This last requires extreme vacuum conditions (~$7*10^{-6}$ Pa) which can also be detrimental to the sample. Our depositions of ~5 nm platinum (Pt), showed that E-Vap was most adapted for further morphological analysis, but the energy of the beam needed to be very low (2-3 kV), and the exposure time also highly reduced,

nearly 2 to 5 seconds, compared to usual characterizations of non-biological samples. Further exposure or higher beam intensity caused bleaching of the samples (supplementary *Figure 137*), which is a well-known problem when performing electron microscopy experiments on biological samples.[238]

Therefore, for further analysis with SEM, the user has to be highly qualified and to have a very exact protocol on analysing depositions, especially if they are to be performed directly on the NEMS detectors. Lastly, a need for adapted software would be required to best analyse the images, as a human-dependent analysis is not a long-term viable solution in this case, since it is a very time consuming and error prone approach. Perhaps image analysis combined with machine learning algorithms could eventually improve the data analysis for this objective, but this was out of the scope of this work.

4.4.1.2 Problems with AFM

In parallel, the AFM was not conclusive because it seemed that the surface of pure undoped silicon was not adapted to attach efficiently the T5-CLPs, resulting in a low rate of particle count and difficulty in performing stiffness evaluations (supplementary *Figure 133*). This proved to be a challenge even for the position on MICA[239], which shows that the surface of these particles is not as easy to work with, thus requiring extensive surface functionalization. In addition, evaluations on trying to observe the decoration proteins on the particles using AFM were not successful, as it was not sufficiently contrasted compared to the particle itself. In the interest of coherence and brevity in our descriptions, we have chosen to omit the data in this section.

My recommendation for integrating SEM or AFM in future analysis to the NEMS-MS bench would be that these initial results should be used to build upon in order to construct the in-line analysis imagined in Figure *62*.

If for example in SEM, the E-Vap deposition is optimized to use after each mass examination, this could be then compared to the solution measurement, and also to evaluate each nebulisation methods, not only on the structure of the particles, but also for example in the concentration generated after each nebulisation method. The counting and image analysis need also to be optimized properly, if this pipeline is to be integrated into the NEMS-MS routine. This would result in both performing the mass evaluation and having a visual confirmation of the structure of the particles. This is actually the power of having NEMS mass detector, as the SEM visualization is very much adapted for such a future potential application.

A similar approach could be imagined with AFM, but this would require much more hands-on knowledge from the user, if it is to be included an in-line method. Nevertheless, AFM can be used separately, simply in solution, in order to acquire physical information on the particles and their confirmation depending on the different biological preparations, such as different production methods (phCLP/recCLP); different decoration proteins; or if the capsid contain or not a genome inside of them.

4.5 Conclusions on Chapter 4

The analytical pipeline (Figure 62), that I designed was primarily focused on analysing the sample from its passage in the different stages of the NEMS-MS bench. This analysis consisted of analysing the sample in liquid gas and solid phase, meaning in its transition from solution to aerosol until its deposition on the NEMS detectors. These analyses also have the attribute of being useful to characterize T5-CLPs in themselves, since they are envisioned as vaccine particles, assisting in identifying important quality control attributes for future examination.

As mentioned earlier, analysing the deposition of T5-CLPs on the detector proves highly complex due to the physical constraints that the particle was put under. Nevertheless, I was able to make some progress in optimizing these steps, especially when using SEM (scanning electron microscopy), and I hope that future work will be conducted in order to reproduce results and improve on them, by potentially integrating adapted machine learning software for image analysis, and hopefully allowing to make in-line particle imaging inside the bench.

That is why I focused on the more ‘classical analytical methods’ for virus analysis, primarily on those based for in-solution evaluation. In the next chapter, I will detail the optimizations that I was able to put in place during my thesis, to these methods and how I combined them in order to have a cross-examination of the T5-CLP samples.

5 Chapter 5: Optimising the Analytical & Biophysical Methods

5.1 Introduction of Chapter 5

When working with biological samples, a series of constraints arise that an experimentalist should be aware of, and if possible, should be able to control.

In my case, I worked with viral particles in solution, in order to perform mass spectrometry. Therefore, my questions when working with them can most often be summarised as follows:

- Biochemical:
  - Is my sample pure enough, i.e. are there any contaminants ?
  - Do I have enough sample to do experiments with, i.e. concentration related problems ?
  - Is this the right buffer/solution to work with ?
- Biophysical
  - Is this sample in the expected size range ?
  - Can I nebulise it ?
  - Can I detect it in NEMS-MS ?

Depending on the question, the answer can be more or less intuitive, and often times our initial hypothesis is proven wrong. However, since we do not have nanometric eyes, we need to rely on a variety of methods to go on and have a look, with all the pros and cons that these methods may bring.

Therefore, the following chapter describes the different methods that I used and optimised in the previous chapter, adapted to analyse the T5-CLP samples. Even though many of these methods are available by several scientific platforms, by groups that we collaborated with, or by commercial service providers, I needed to adapt these tools in order for them to fit best our needs and the scientific questions I was inquiring for.

5.1.1 Preface: The proper view of objects

In the picture below, Figure 63, we can see the shadow of a human, on the left, and the shadow on the right would be that of 'a mechanical human' being built. Which in our case would be the analogy for the intact virus (left structure), or for us the example of broken viruses or VLPs (right structure). The man in construction, on the right, has various human features, such as its right arm, whereas the rest of its body is purely mechanical. For viruses, this could be imaged as being their tails, their spikes, lacking their genome or having damaged capsid shells.



Figure 63. Sculpture named 'Nasty Pieces of Work', by artists Tim Noble and Sue Webster. Dated: September 2008. Content: two wooden stepladders, discarded wood, broken tools, light projector. [Link to source].

Nevertheless, in reality, these are both 'just junk', as the list of objects that make them up may let one think, containing things such as discarded wood and broken tools. But an important element of the equipment needed for this effect, as the artist point out, is a light projector. Meaning that it is all a matter of well-placed lighting and shadowing. This is the same as all methods that allow us to observe an object in two dimensions; we are simply observing their projections. Of course, by placing the light anywhere else in the room, the shadow would not be the same, so an orthogonal viewpoint would be of great value. However, in the 3D world that would not be enough, so weighing or touching them would allow for an even more in-depth knowledge about these objects.

This seems obvious for an object at our scale, and it should not be too different for viruses, the difference being that they are at the nanoscale. That is why one has to come with a large toolbox of techniques, when examining viruses, to get as close to the reality of things as possible.

5.2 Negative staining Transmission Electron Microscopy (TEM)

The most classical method for biophysical analysis of viruses has always been TEM, often used in the form of negative staining TEM. But the "one-size-fits-all" approach is shown to not work most times, and thus one has to adapt this tool to one's needs.

5.2.1 Setting the stage, indications from I2BC

The following is a brief summary of the results and conclusions that were communicated to me by our colleagues at I2BC.

5.2.1.1 Cryo-EM results

Our colleagues at I2BC in Paris had indicated that they were conducting cryo-EM (cryogenic electron microscopy) on their modified recombinant CLPs (recCLPs). This helped them reconstruct the 3D images of these CLPs (personal communication, based on Nicolas Ducrot's thesis manuscript).[209] One such example can be seen in the image below (Figure 64).

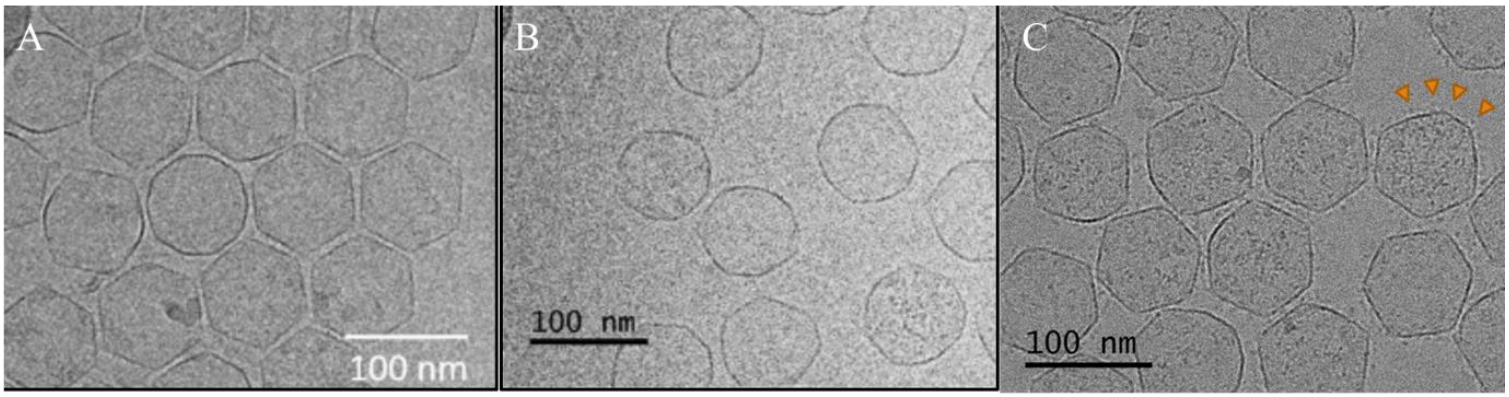


Figure 64. Cryogenic electron microscopy analysis of different purified empty T5 capsids. A) phage-derived CLPs (from infection), without decoration; B) Recombinant CLPs (from plasmid), without decoration; C) Recombinant CLPs (from plasmid), Orange triangles are pNO decoration (pb10 N-terminal combined to Ovalbumin). Figure reprinted from the thesis of Nicolas Ducrot.[209]

It compares the "classically produced", phage derived CLPs (in Figure 64A) versus two types of recombinant CLPs: non-decorated (in Figure 64B); and decorated by pNO (in Figure 64C). This decoration was highlighted by orange triangles pointing at visible decorations in the third picture.

However, cryo-EM observations of these first recombinant capsids reveal that some of them are irregularly assembled. Indeed, particle diameters vary, and the expected icosahedral shape appears mostly distorted when compared with the regular shape of capsids produced by the phage. He was able to quantify this ratio between irregular/regular capsids as being ~18 % for phCLPs, and ~ 43 % for recCLPs.[209]

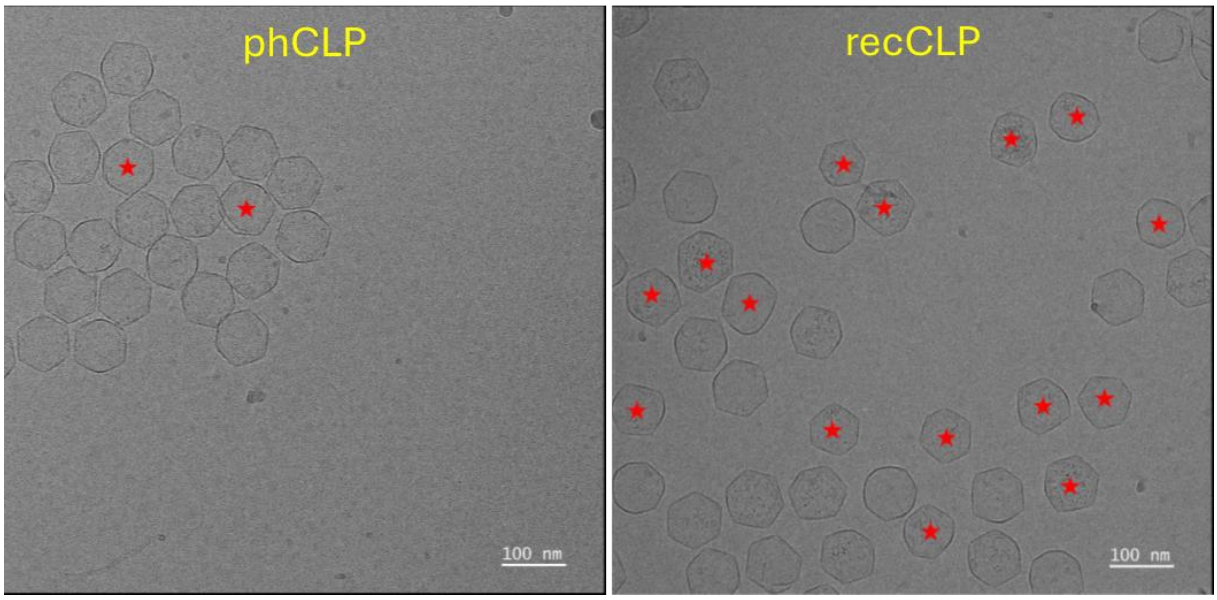


Figure 65. Selected cryo-ME images of capsids representative of those used to estimate the proportion of regular recombinant capsids. Irregular capsids are indicated by a red star. Figure reprinted from the thesis of Nicolas Ducrot.[209]

Nicolas argued that this irregularity in recombinant capsids could be explained by a lack of assembly regulation. In fact, the recombinant production system is reduced to capsid protein genes alone and this does not include the various regulatory signals that could be found upstream or downstream of these sequences. These regulatory signals could, for example, be used to adjust the stoichiometry of capsid proteins during their production to respect the ratio of protein copies: 775 for pb8, 12 for pb7, ~10-15 copies for pb11.[202,212]

5.2.1.2 Negative staining TEM results

In addition, they had also imaged these same particles using a significantly less costly type of electron microscopy, when compared to cryoEM, which is negative staining TEM, as shown in Figure 66.

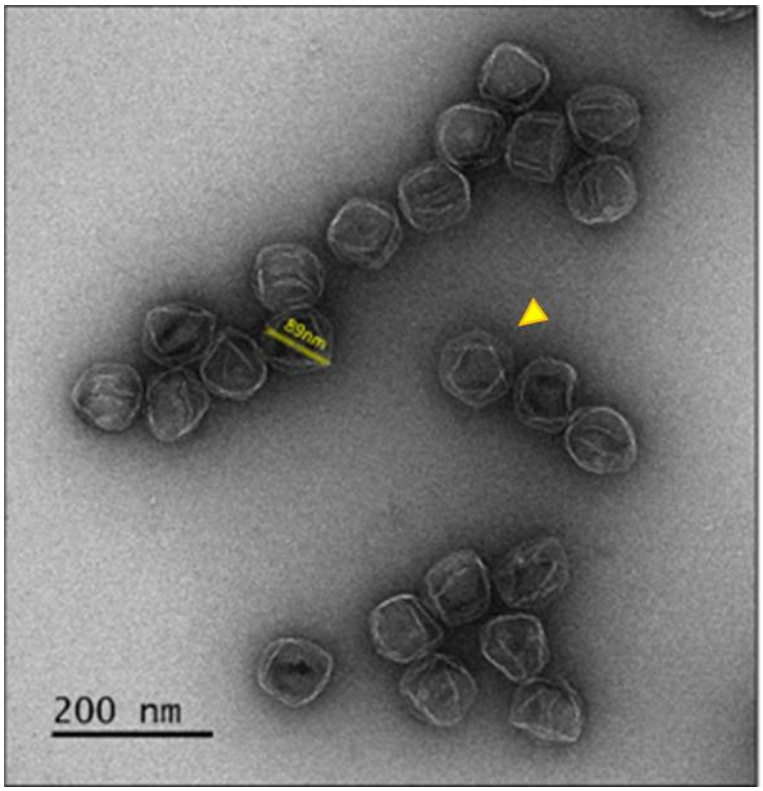


Figure 66. Negative staining transmission electron microscopy analysis of purified capsids. CLPs have an average diameter of 90nm after in vitro expansion and no deteriorated capsids could be observed in this sample. The yellow line show the measurement for one example of a regular capsid, and the yellow triangle highlights a regular looking CLP. © I2BC, Nicolas Ducrot's thesis manuscript.

They had used a negative staining by UAc (Uranyl Acetate), which is a classic colorant for this method when analysing viruses. The known issue with such a colorant is that it has a pH~4, and our collaborator had said that capsid often tend to look, quote: "shrivelled up, like a deflated ball that's been jumped on".

Therefore, since the EM platform in Grenoble could perform such experiments, we decided it was worth to have a look at our sample batches using this method.

5.2.2 Principle of TEM

The technique of Negative staining TEM works by the presence of heavy metal ions in the stain scatter electrons, thus creating contrast in the transmitted image. The stain surrounds the sample, enhancing contrast, by providing a dark background against which the sample is visualized. That is because electrons that pass through the unstained sample regions are scattered less, thus appearing in a lighter colour in the image, while the heavy metal-stained areas appear darker, due to the increase in electron scattering. It is this contrast difference that allows for detailed visualization of sample morphology and refined visualisation of structural features.

The main elements that made me integrate this method in my analysis and comparison to other methods can be resumed as follows. Firstly, it allows for visualization of samples in their native, hydrated state. There are no additional steps required to fix the sample (of course taking into account the effect that some colorants might have), no dehydration is required, and there is no for embedding in resin, or plunging the sample in cryogenic conditions for example. This is of course related to the relatively quick and straightforward sample preparation method, as described earlier, making it suitable for rapid visualization and screening of a variety of biological samples. Last, but not least, it also provides high contrast and extremely high resolution, nearing atomic scale depending on the stain used. This makes it particularly useful for visualizing small and low-contrast biological specimens, and even reach visualizing individual molecules, and macromolecular complexes.

Below is a brief schematic explanation of the key steps on the how to produce a sample for TEM, Figure 67, and found in Annex (7.4.2.1, page 292) 'Results on the other viral samples' is the protocol that was used in my thesis, based on the so called Valentine procedure.[240]

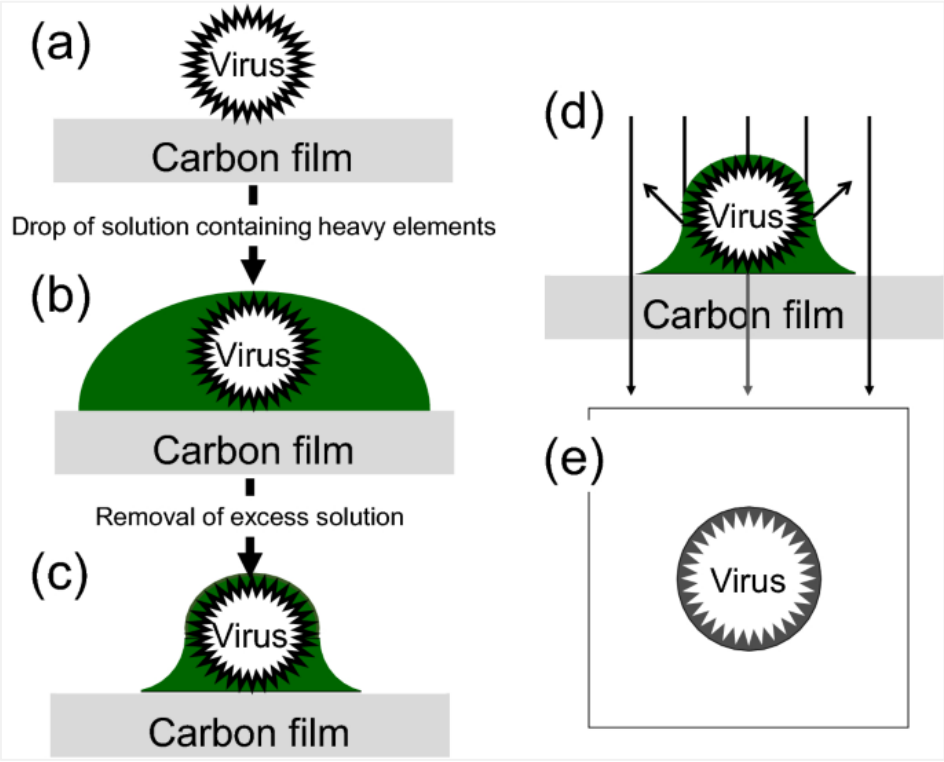


Figure 67. Negative staining *TEM* method. (a) Virions are attached to the carbon support film. (b) A solution containing heavy metals (negative-staining reagent) is dropped onto the film. (c) Excess solution is removed, and the sample is dried. (d) Transmission electron microscopy (TEM) of heavy-metal-coated virions produces (e) a reverse-contrast image of the virus. Figure and Description (from original), Sahiro et al.[241]

The sample is dropped on the carbon film, or MICA, on top of which we then inject a diluted solution of salts containing heavy elements such as uranium, tungsten or molybdate.

The liquid excess is removed in order to have remaining only the solution surrounding the virus. And then, after irradiation by the electron beam, the transmitted image is observed using reverse contrast, revealing the virus structure.

5.2.3 Optimisation of the TEM experiments

In a first evaluation, I performed a qualitative analysis of the samples, by visually looking at the samples and trying to see what the best conditions for our analysis could be. This is a common approach when using this method, as it provides rapid conclusions on the samples, especially on their shape and conformation. It goes without saying that there is no statistics here, and thus no quantitative conclusion on the sample size, and there is also a clear selection bias of images of interest when looking at our samples.

5.2.3.1 Qualitative examination

Therefore, my first experiment was to check if the batches that I had already received were similar to one another or not.

5.2.3.1.1 Inter-batch comparison (Cesar versus Juno)

The first examination performed, was to compare two different batches, received at different times, of the exact same sample production. Therefore, I compared Cesar (July 2021) and Juno (June 2022), both phCLP-deco samples, shown in Figure 68.

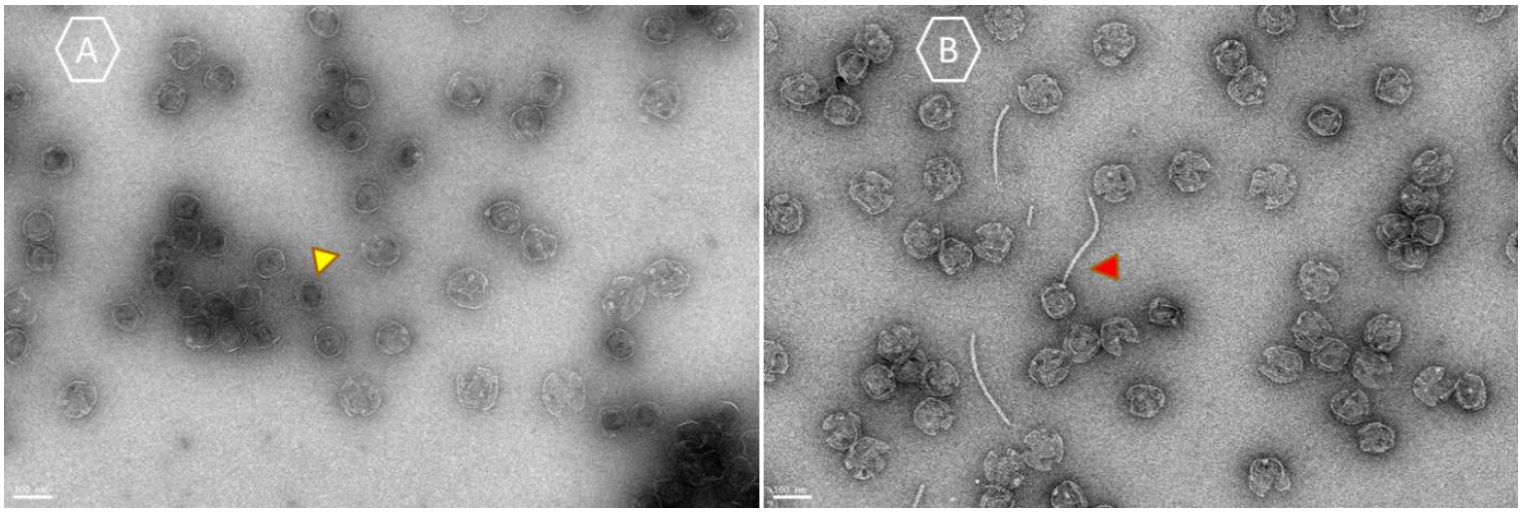


Figure 68. Negative staining TEM analysis of comparing of the different batches of phCLP-deco, A) Cesar and B) Juno. Coloration by UAc (Uranyl Acetate, at 2 % v/v). The yellow triangle points at one example of a regular capsid and the red triangle in figure B points at the presence of a seeming intact phage, still attached to its tail. Other tails can be spotted in the background of the image.

The darker shade in the image shows an accumulation of the colorant, thus making it more opaque to transmitting the electron through. This may be due to the samples being itself

more opaque, having more material, e.g. containing genome in its insides or having a compilation of several capsids being on top of one another. When the second case happens, we can often see the outlining of the capsid, and thus conclude an overlap of several particles. In the case of Cesar, it was probably a mix of the two, as darker areas were very much present in many of the images we took, and the presence of the genome cannot be excluded. For Juno, the presence of contamination by phages was evident, by the high presence of phage tails in all the images taken for this sample, one example is shown in Figure 68B, being pointed by a red triangle. In addition, we could clearly see that the capsids were highly broken and extremely deteriorated, even when compared to Cesar. Therefore, it can be seen that there is a high variability even when looking at the same sample, between different batches.

However, I was aware that these CLPs are highly affected by pH, as reported in the PhD thesis by Emeline Vernhes [222], and it is well known that uranyl salt have a better resolution in negative staining TEM, with a grain size of 4 - 5 Å, but their low pH can be detrimental to some samples.[242] Therefore, in order to observe that, I pursued an examination of the same sample in different negative stains, which have a more neutral pH value.

5.2.3.1.2 The pH problem of the negative stains

In order to see if the broken capsid were indeed coming from the biological preparation of the sample, or the pH of the negative stained used, we decided to compare the same sample in different colorants. Therefore, to test this hypothesis, we tested three different stains: UAc (Uranyl Acetate), as done previously and known to have a pH~4; and to compare it with two other stains with 'neutral' pH (pH ~7), which are SST (Sodium silicotungstate) and AmMo (Ammonium molybdate). The tungstate and molybdate stains have a lower resolution and a larger grain size of 8 - 9 Å, but have a neutral pH.[242] These results can be seen in Figure 69.

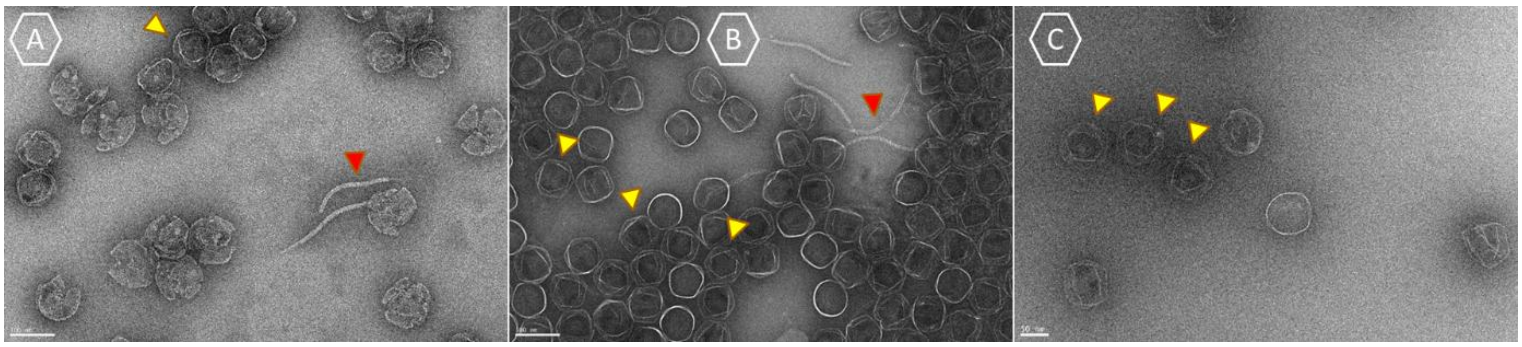


Figure 69. Negative staining TEM analysis of the same sample (phCLP-deco), named Juno, in different negative stains, at 2 % v/v. A) UAc (Uranyl Acetate) (pH=4), Mag= 23000x; B) SST (Sodium silicotungstate) (pH=7), Mag= 23000x; C) AmMo (Ammonium molybdate) (pH=7), diluted (1/50), Mag= 30000x. The yellow triangles point at examples of regular capsids, and the red triangles in figure B and C point at the presence of tails.

These results showed that indeed, for both neutral stains (SST and AmMo) the capsids behaved more closely as they were expected, presenting more regular shapes and resembling more to the shapes seen in cryoEM, as seen in Figure 64. This was highlighted by the yellow triangles in the figure above (Figure 69), pointing at some regular looking capsids.

So now the obvious next step was to look if the decoration was present or not, and if we could easily visualise it with TEM.

5.2.3.1.3 Looking for the decoration protein

The next step, after we established that the negative staining TEM was a good quality control tool for our analysis, we wanted to check how visible the decoration of the capsids would be with such a technique (Figure *70*).

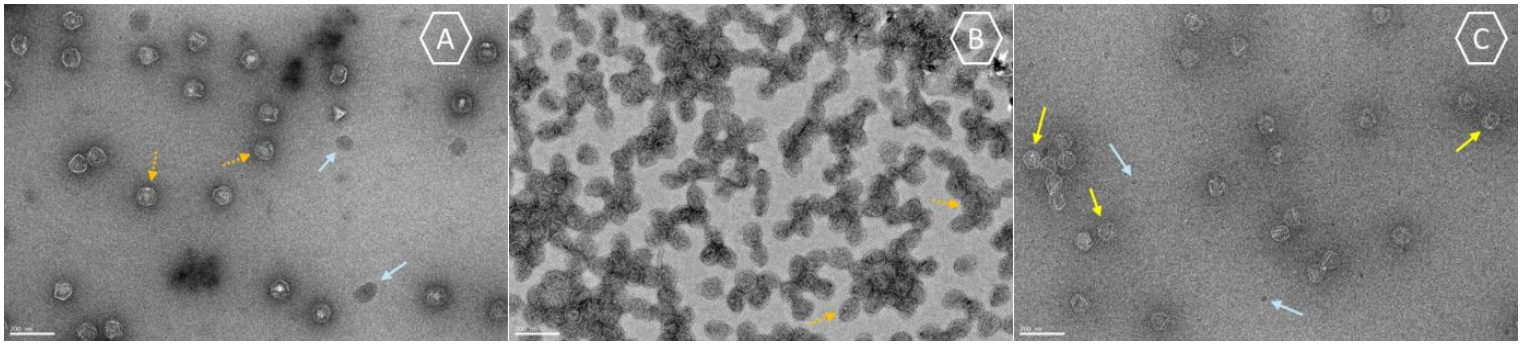


Figure 70. Negative staining TEM analysis of Isis (phCLP+deco) samples, in different negative stains, at 2 % v/v. A) UAc (Uranyl Acetate) (pH=4), diluted (1/50); B) SST (Sodium silicotungstate) (pH=7); C) AmMo (Ammonium molybdate) (pH=7), diluted (1/50). All performed at Mag= 11000x, scale bar = 200 nm. Orange dashed arrows: potentially decorated phCLPs; yellow: well decorated phCLPs; blue arrows: other particles (not phCLPs).

It is clear that the least adapted stain is SST (Figure 70 B) since it presents a very high background noise of the decoration protein, seen as a white colour covering the particles. When comparing between UAc (Figure 70 A) and AmMo (Figure 70 C) we can notice that the particles are much more visible for UAc, since it improves the image resolution, but the decoration proteins are much better visualized with AmMo, as highlighted by the yellow arrows. Based on these qualitative observations, AmMo coloration seems most adapted to visualise decoration protein (pNO), although not having the best contrast.

Another interesting thing seen for these two stains, as noted by the blue arrows, is the presence of other smaller particles being visible in all of our images, indicating the presence of possible contamination or particle aggregation. This can perhaps be due to the decoration protein aggregating together, when these stains are used, and it is something that is apparently not dependent on the stain's pH.

5.2.3.1.4 Comparing production and decoration

After visually and qualitatively concluding that the coloration with AmMo seemed the most satisfactory compromise, to be able to simultaneously observe regular capsids and their decoration, I decided use this stain to compare the phage-derived CLPs (Isis, Figure 71 A & B) versus the recombinant CLPs (Orion, Figure 71 C & D).

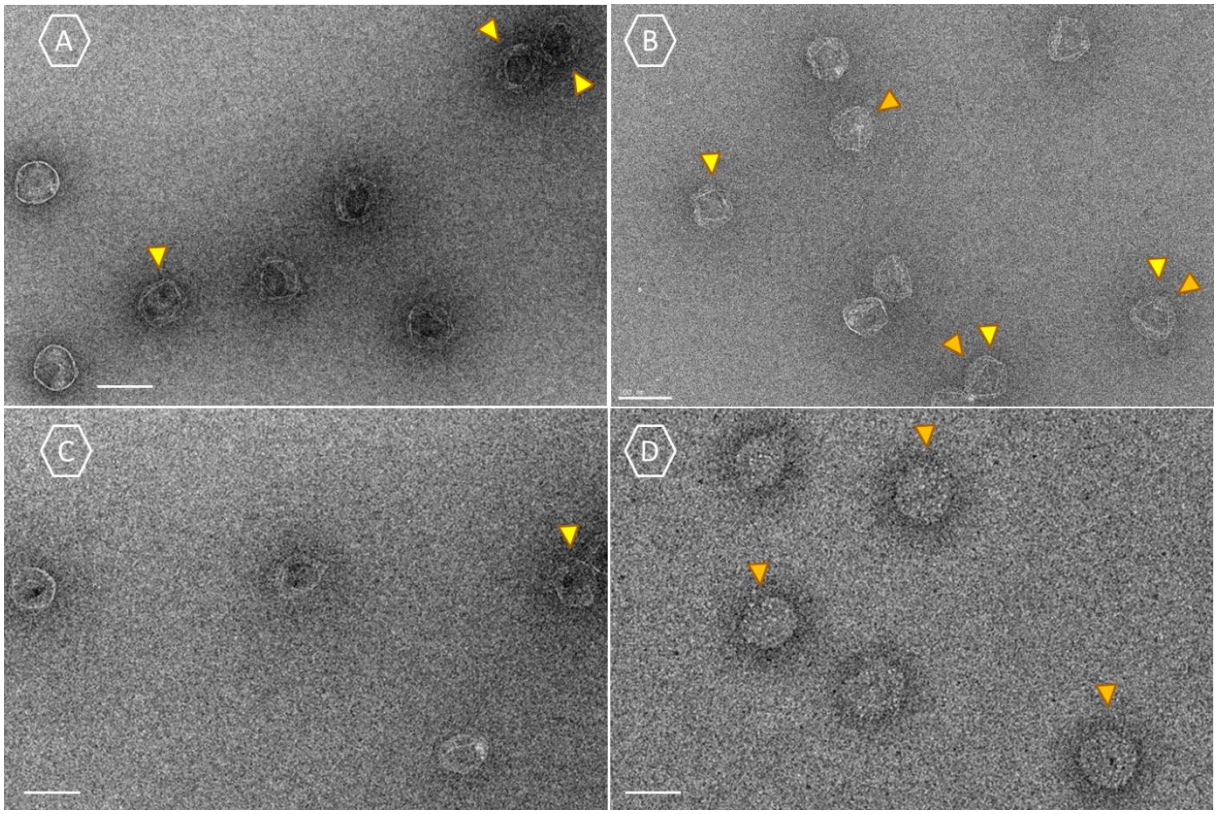


Figure 71. Negative staining TEM analysis comparing decorated and non-decorated CLPs, produced by different methods, phage-derived (Isis) and recombinant (Orion). A) Isis – deco, Mag= 11000x; B) Isis + deco, Mag= 23000x; C) Orion – deco, Mag= 11000x; D) Orion + deco, Mag= 11000x. All preparation were done with AmMo (Ammonium molybdate) at 2% v/v (pH=7), and samples were diluted at 1/50 to avoid particles overlapping. All images were adapted in order to be at a similar scale, with the white bar at the bottom = 100 nm. Yellow triangles indicate regular looking capsids, whereas the orange triangles indicate the capsid where noticeable decoration is visible.

It can be seen in Figure 71, looking at the number of the yellow triangles, that the amount of regular capsids seems to be more dominant in in the phCLPs (Isis; A & B), and that the recCLP (Orion; C & D) seem to be more round and amorphous. This result is similar to that observed by cryoEM by Nicolas Ducrot, shown in Figure 65.[209] Nevertheless, their structural integrity seems to be maintained, and they do not appear broken. The background of the images where the pNO decoration protein was injected in the sample, also seems to be slightly opaquer than for the images where this protein was lacking. This would make sense, as a mixture of the stain with this protein could increase the background noise of the image.

5.2.3.1.5 Conclusions on the qualitative results

By looking at these results and the previous images in the qualitative way, I drew two main types of conclusions.

First, the sample related ones, which would be as follows:

a) **Juno** (phCLP) is highly contaminated by phages (tailed capsids). {Figure 68 and Figure 69}

b) **Isis** (phCLP) seems to have more well-defined and regular capsids compared to **Orion** (recCLPs). {Figure 71}

c) **Orion** (recCLPs) seem to be more 'swollen' compared to **Isis** (phCLP). {Figure 71}

Secondly would be those related to the method, mainly related to the negative stain used. Therefore, I concluded the following:

d) The low pH coloration (UAc, pH=4) seems to induce broken capsids compared to coloration by SST and AmMo, both at pH=7. {Figure 69}

e) AmMo (pH=7) coloration seems most adapted to visualise decoration protein (pNO), although not having the best contrast. {Figure *70*}

These experiments indicated that the usage of negative staining TEM, to observe the shape and decoration of these CLPs, can be used as a good first qualitative control tool for any batch examination, prior to performing mass measurements. But of course, there is also a clear selection bias on choosing the images of interest, when we visually examine our samples. That is why I wanted to go further by analysing all the TEM images, of the different batches, in all the different conditions discussed in the qualitative manner

The main drawback of the qualitative analysis being that there is no statistics to be discussed and thus no significant conclusions to be made. In addition, my interest was to evaluate the sample particle size, and from there to relate it to the particle mass. This of course requires a meticulous examination of the particle, in order to have a representative population size. In addition, having the maximum amount of data on this part would thus help to streamline the analysis by TEM, especially important for future samples by our group.

5.2.3.2 Quantitative examination

The following analysis was performed with the objective of examining the size and shape of T5-CLPs, in order to render it as quantitative as possible. Considering for a variety of

variables such as: the production method, the decoration status, the negative stain used, and the overall stability of the structures when performing the TEM analysis.

5.2.3.2.1 Methodology: Data analysis by ImageJ

I manually performed the following analysis, using ImageJ (Fiji).[201] The image acquisition of these particles is the same as that seen later on (Figure 72). I analysed nearly 170 TEM images, and I manually measured more than 3300 individual virus particles (see Table 10). This represented a considerable amount of manual work and time spend on my side in order to acquire these data, making this approach as a routine evaluation very unlikely.

ImageJ software analyses TEM images in two steps. First, individual particles are identified by manually tracing their contour. Then, based on these contours, it provides the main geometrical features of each identified particle. To this end, ImageJ always assumes the selected particles are elliptical and consequently provides the following information for each particle: its major and minor axes, denoted a and b respectively; its perimeter, denoted as $Perim$, which is calculated as the sum of the drawn segments used to delimit the selected particle; and its surface area, denoted as $Area$.

By definition, we always have $a \geq b$ and from there, we define the 'aspect ratio' r, as:

$$r = \frac{a}{b}.$$

From the major and minor axes of an ellipse, we can approximate its perimeter, knowing that there is no analytical formula. To this end, we used one of the simplest formulas, which is:

$$Perim_{ellipse} = 2\pi \left(\frac{a^2 + b^2}{2} \right)^{\frac{1}{2}}$$

This last definition implies there can be a discrepancy between the value of perimeter $Perim$ as estimated with ImageJ and the value of $Perim_{ellipse}$ (also denoted as $Perim_{calc}$ or referenced as $Calculated\ Perimeter$ in the following). This discrepancy can naturally be used to discriminate between particles that are truly ellipsoidal and particles that lost their natural spatial configuration. Indeed, $Perim$ value is changing depending on how we draw the lines surrounding these particles. To better visualize this point, let us briefly look at how I acquired the size and shape of these particles, using Image J (Figure 72).

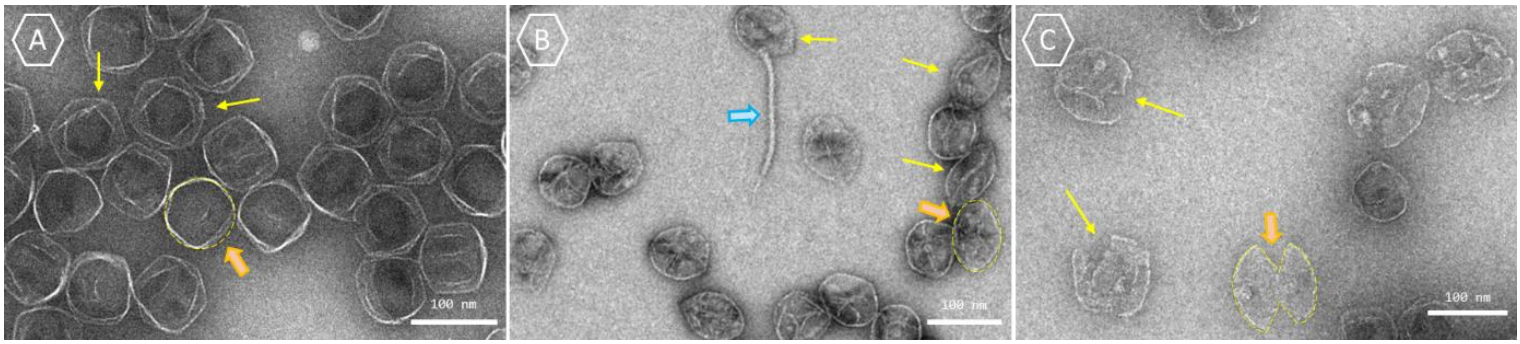


Figure 72. Selected negative staining TEM images of three types of particle shapes: A) spherical and polygonal; B) elliptical; C) broken. The orange arrows highlight the main example, showing the line that was drawn around the particle, and the yellow arrows give two additional examples. The blue arrow highlights an example of a phages tail.

In these selected images we can see that broken particle (case C), although still presenting major and minor axis that could be similar to regular capsid (case A), their perimeter is much different, since it has indentations that cannot be neglected. In the case of elliptical shapes (case B), the difficulty becomes harder as the axes get elongated, which the aspect ratio parameter is more appropriate to work with.

I would like to highlight here that the notion of 'mean diameter', taken between the average of the major and the minor axis, did not sound appropriate when talking about not perfectly spherical particles. Therefore, I will use what is known as Equivalent spherical diameter (ESD), which is defined as the diameter of a circle with an equal aggregate sectional area.[243] Its formula is:

$$ESD = 2\sqrt{\frac{Area}{\pi}}$$

This is a classic method of particle size calculation in metrology.[244]

5.2.3.2.2 The raw data analysis

The following plot (Figure 73) shows the relationship between the ESD, as a function of the sample name, and of the presence or absence of decoration.

The yellow highlighted band is the size range commonly expected for these viruses, lying between 75 and 90 nm range, as reported in the literature.[245] This will be the reference range for all the figures referring to the T5-CLPs size.

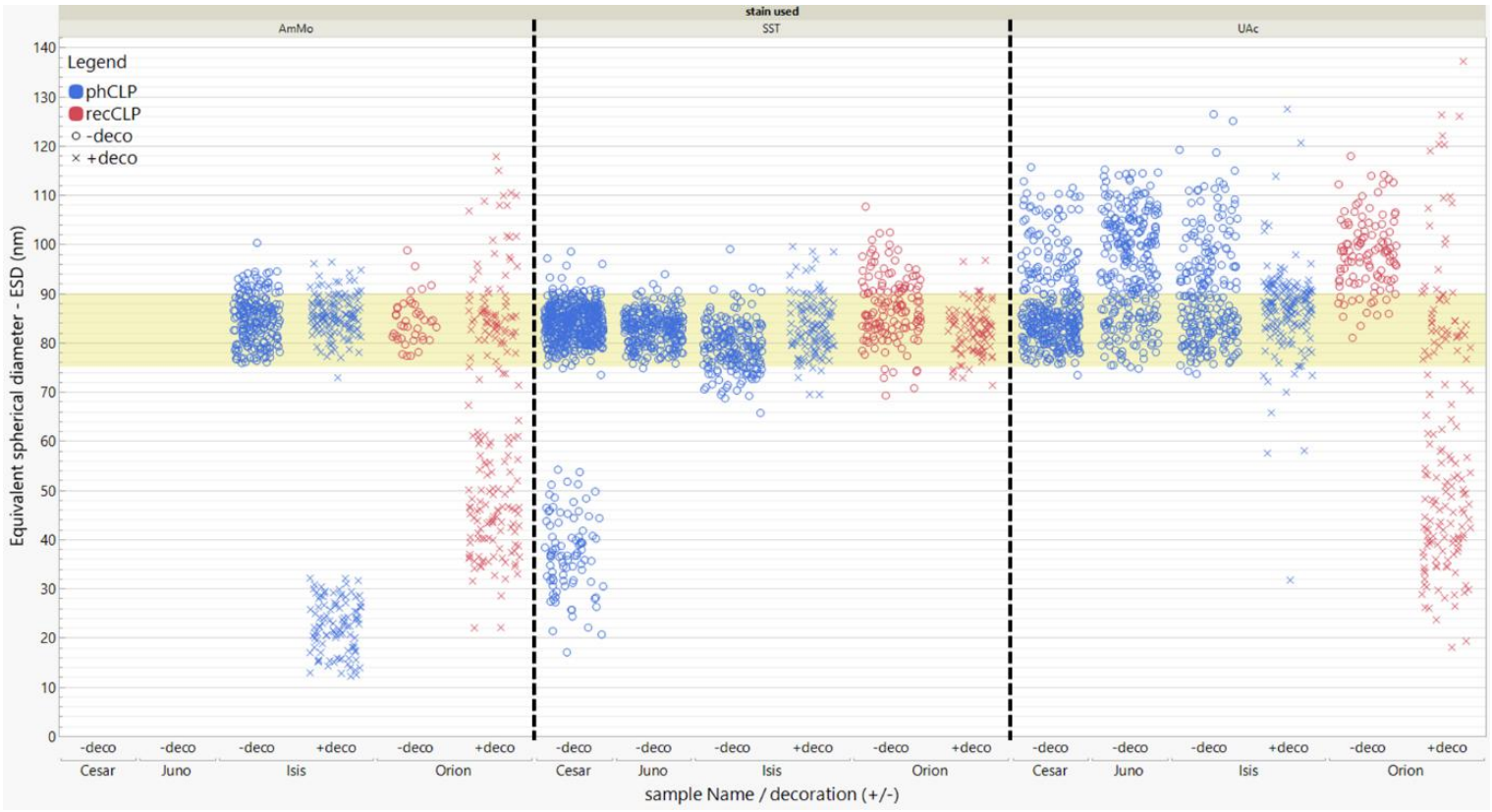


Figure 73. Dot plots showing the relationship between the Equivalent spherical diameter (ESD), in nm, as a function of the sample name. The figure is split into three panels (left to right), depending on the negative stain used: AmMo (first); SST (second); UAc (third). Sample type: phCLP (blue) and recCLP (red); and decoration: presence (+deco; crosses) or absence (-deco; circles), shown in the legend (top left). Samples are: Cesar, Juno & Isis (phCLP), Orion (recCLP). The yellow highlighted band is the size range 75-90 nm.

We can see from the previous plot (Figure *73*) that the diameter across the different stains is changing noticeably. The distribution for AmMo seems to present a second smaller peak for the decorated capsids, at around 10 to 30 nm for Isis, and at around 30 to 60 for Orion.

Using the stain SST seems to be the most stable colorant for measuring the size, as we see that the capsids are stable varying in the expected range, between 75 to 90 nm. Nevertheless, there is a smaller distribution again, but this time only for Cesar, in the 20 to 60 nm range. We can see that the three replicates between batches are not exactly the same, i.e. for Cesar, Juno and Isis (non-decorated). The decoration does seem to have an effect on Isis, making it larger in size, but the opposite effect seems to be visible for Orion. We can also see that the decoration of the capsids seems to stabilize them for this stain.

Whereas for UAc the distributions seem to be much more variable all across the samples, indicating that the stain might have a noticeable effect on the structure. We can observe this quite noticeably between the three similar batches observed, i.e. for Cesar, Juno and Isis (non-decorated), showing high variability. We can also see that the centre of mass of the Isis+deco distribution is slightly higher than the one without decoration. The distribution

for Orion-deco seems to be again higher than that of the decorated one, but this last is stretched out in all of the size range between 20 and 200 nm.

Let us now see (Table 10), the statistics I collected for these different samples, and how they are separated between the three stains.

Table 10. Table capturing the particle count for sample and stain used. The condition in bold represent the minority condition, for both sample type and decoration. The total particle sum of all the measurements is presented in orange.

sample Name	sample type (phCLP/ **recCLP**)	decoration (-deco/ **+deco**)	Nb (AmMo)	Nb (SST)	Nb (UAc)	Nb Sum (per sample)
Cesar	phCLP	-deco	/	568	346	914
Juno	phCLP	-deco	/	261	261	522
Isis	phCLP	-deco	218	202	224	644
		+deco	245	120	165	530
Orion	**recCLP**	-deco	40	136	111	287
		+deco	188	96	161	445
		Nb Sum (per stain)	691	1383	1268	3342

These data show that if we are to take a limit of statistical significance (usually at +500 counted particles), there is only one such evaluation that satisfies this particle count quota, i.e. Cesar (in SST). Most of the samples are at a semi-quantitative level, i.e. less than the aforementioned limit, with a median evaluation at ~195 particles per sample per stain.

These semi-quantitative results confirm the qualitative analysis I made earlier, as far as the integrity of the capsid structure is concerned. The most visible case being the larger size of particle when using UAc staining.

This raised the question regarding the effects of the stain on the virus structure, and most importantly, how can we be sure that this is indeed a problem of the stain or the presence of decoration. The following data analyses were performed in order to try to elucidate this.

5.2.3.2.3 Particle size problem: Stain or Decoration ?

In a first moment, I will investigate the stain impact on the particle size and shape. In order to make sure that this is indeed a stain related problem, as I already suspected earlier (Figure 69), I decided to plot the measured perimeter versus the calculated perimeter and obtained Figure 74. The idea here is to separate the capsids that are perfectly elliptical from the others.

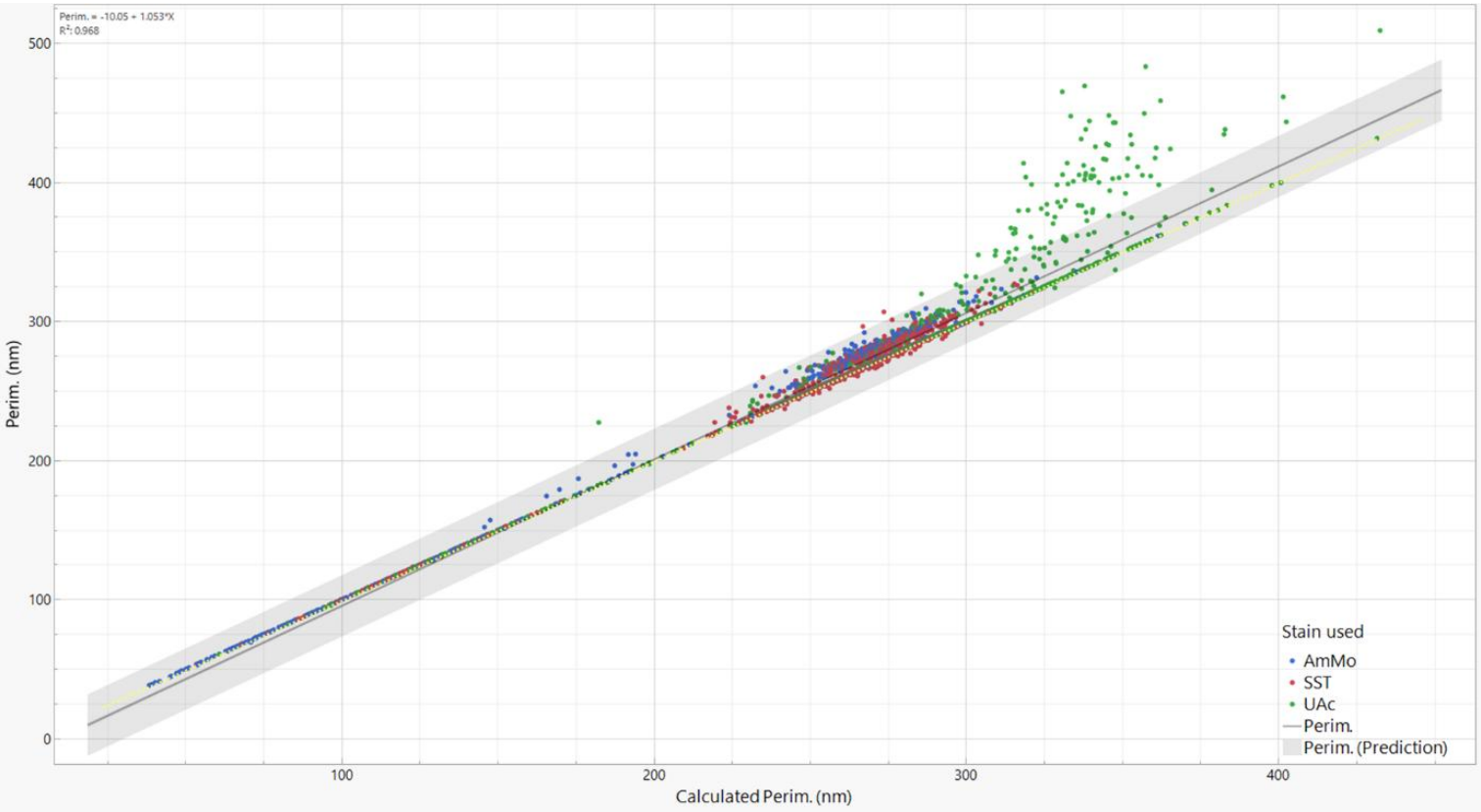


Figure 74. The relationship between the measured particle perimeter and the calculated particle perimeter, depending on the stain used. Legend (bottom right): AmMo (blue); SST (red); UAc (green). The linear fit is shown, with is CI 90% shaded interval, and the corresponding statistics of the fit are indicated in the top left of the graph. The Ideal case of the perfect ellipse is shown by the yellow dashed line.

We can see in Figure 74 that the relationship between *Perim* versus *Calculated Perimeter* seems to have an upwards trend, especially for UAc. It seems to break viruses more than the other stains. As we saw earlier the SST data seem to be much closer to one another, centring between 200 and 300 nm, and the AmMo data present the majority under the 200 nm line. We can nevertheless see that there are particles in the 'perfect ellipse line' (highlighted by the yellow dashed line), meaning that some particles most likely present a 'rounder' appearance, thus we need to find out how to select them from the other points appearing as not round. In addition, these particles are not the majority of this population, and thus we need to understand if the rest of these particles are all broken capsids. This of course serves the purpose of choosing the regular capsids from the irregular ones.

I made an additional investigation of the same data, but here I took into account the decoration status of the particles, Figure 75.

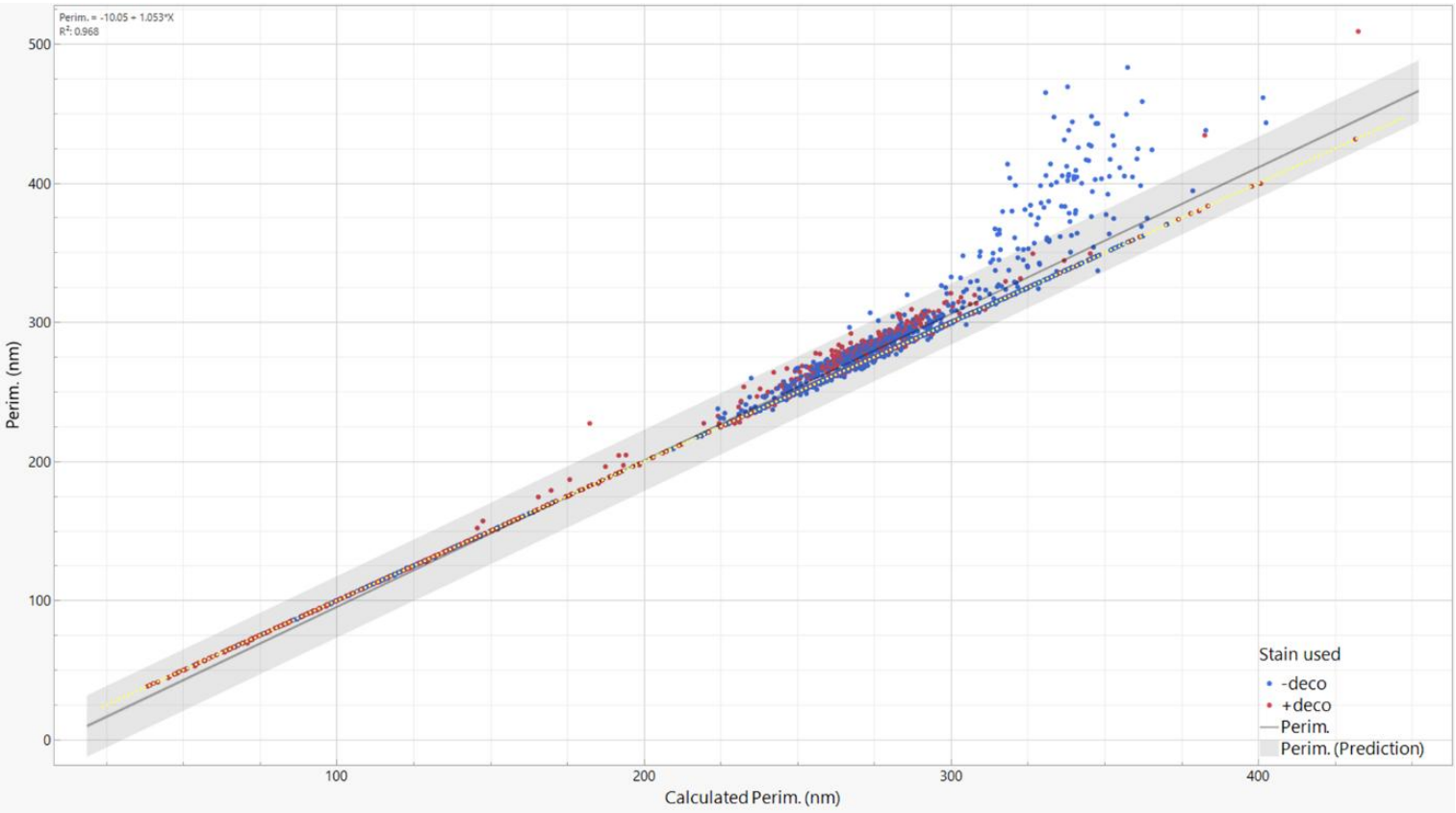


Figure 75. Relationship between the measured particle perimeter and the calculated particle perimeter, depending on the presence (+deco; red) or absence (-deco; blue) *of decoration. Legend (bottom right).* The Ideal case of the perfect ellipse is shown by the yellow dashed line.

We can see in Figure 75 that the non-decorated particles constitute the majority of the particles not following the main trend line of the data, and also that the decorated particles are predominant for those below the 200 nm line. Similarly, a mixed population can be seen in the 200 to 300 nm range showing a parallel group appearing to the main trendline. As seen earlier, the UAc affected the non-decorated viruses most. These results demonstrate that carrying out semi-quantitative evaluations have significant advantages over simple qualitative analysis.

To further examine these factors, more statistical analyses should be performed on additional measurement points. One also has to be very aware of the large manual effort that this task represents. Nevertheless, despite the manual effort required, more data will enable for better statistics, allowing us to separate multiple parameters more effectively. Specifically, powerful statistical methods like Principal Component Analysis (PCA) could be used to evaluate the influence of each variable on particle structure (i.e.. size and shape) in multiple dimensions. This could help determine if the observed effects are primarily stain-related or due to the particle decoration, and if they are consistent to the data that I was able to collect.

Next, I looked at whether I was dealing with ellipses, which I will call here as the "roundness hypotheses". With this hypothesis I made the consideration that all ellipses are intact viral particles.

5.2.3.2.4 Correlating the perimeter ratio to the aspect ratio

My next step was to examine the 'roundness hypotheses', and to do that, I needed to separate the effects of the parameters a and b, the major and minor axes respectively. The perimeter (examined in Figure 74 & Figure 75), as a relation dependent on the sum $(a + b)$, is insufficient to solely select ellipses. Consequently, I decided to plot the perimeter ratio $(Perim/Perim_{calc})$, versus the aspect ratio $(r = a/b)$. The result is shown on Figure 76.

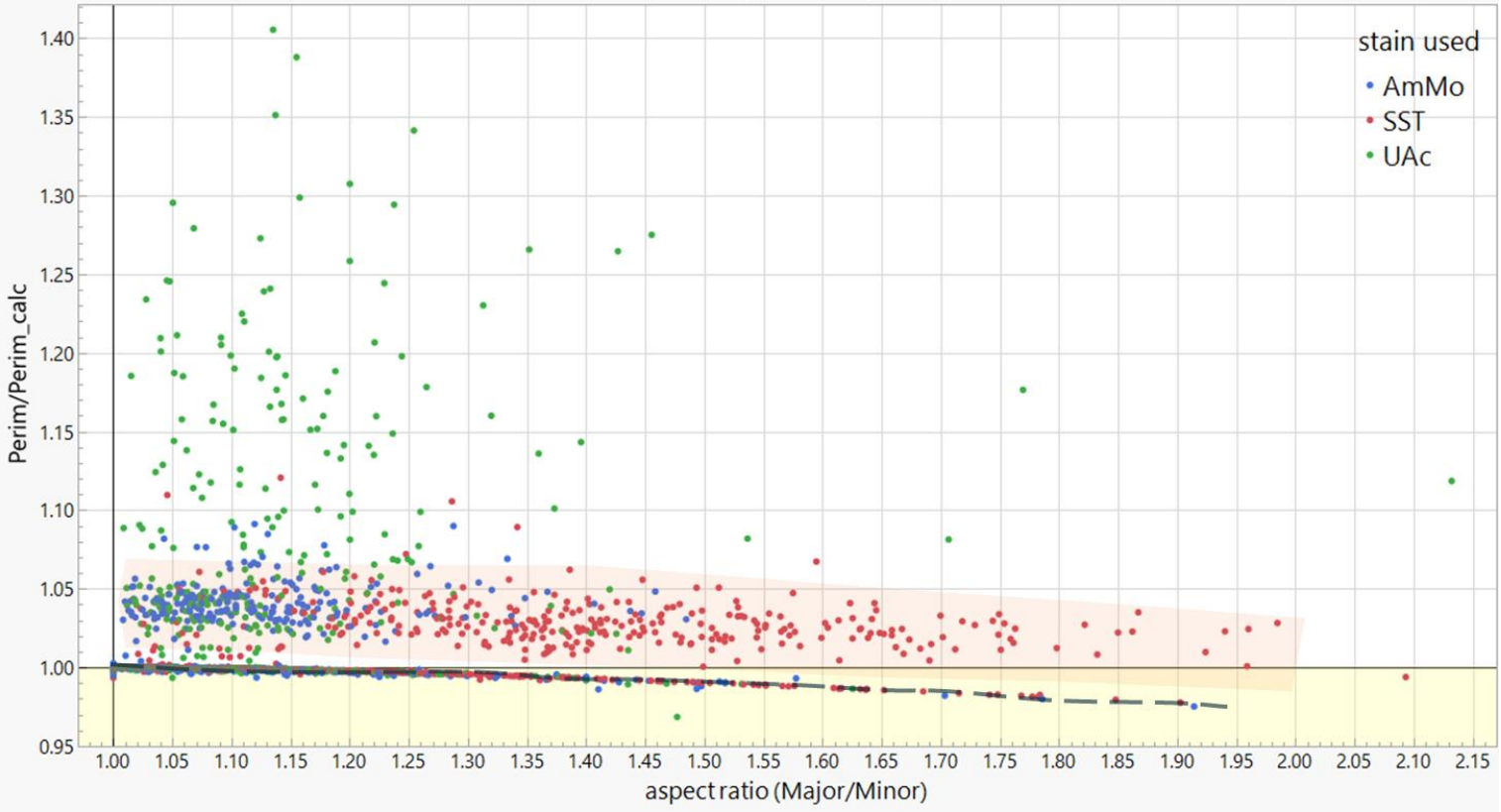


Figure 76. Relationship between the ratio of the measured particle perimeter over the calculated particle perimeter (Perim/Perim._calc) as a function of the aspect ratio (Major/Minor). Legend (top right): AmMo (blue); SST (red); UAc (green). The solid black lines serve to show the threshold where both axes are equal to 1. The region (Perim/Perim._calc)<1 is highlighted in yellow.

The simplest way to understand Figure 76 is that the further up the point go in the perimeter ratio $(Perim/Perim.calc)$ axis, the more the particles are broken down, and the more we go up the aspect ratio axis, the more the particles get elongated. A perfect ellipse would lie at any point in the black solid line defined by $Perim/Perim.calc = 1$. In addition, we can see that there are rarely any points above 2 for the aspect ratio, indicating that we rarely have

two particles directly joined to one another end to end, thus showing that the seen values are less likely to be aggregation of two particles, and more likely to be broken particles.

We can see in Figure 76 that there is a trend line appearing (represented as a black dash line), starting from the [1,1] point, and drifting below the line $Perim/Perim.calc = 1$ (highlighted in yellow). This I believe to be the part of the capsids that represent the regular polygons, which I will detail below.

A larger band (highlighted in orange) can be seen starting at around 1.02 to 1.07 for the perimeter ratio $Perim/Perim.calc$, and again extending towards the threshold line $Perim/Perim.calc = 1$. Interestingly the points that go above the $r = 1.3$ line are the ones coming primarily from SST (red). Perhaps indicating that this stain induces a larger portion of the capsids to become more elongated in their major axis.

On the other hand, the AmMo particles (blue) are mainly concentrated between the 1 to 1.2 values for r, and in the range 1.02 to 1.05 for the perimeter ratio $Perim/Perim.calc$. Meaning that this stain seems to slightly deform the capsids without inducing a major damaging of their structures.

Here we see again that the UAc (green) particles are the ones going above the 1.07 value for the perimeter ratio $Perim/Perim.calc$. Meaning most likely that the particle breaking is indeed most likely when using this stain.

As mentioned earlier, I wanted to explain the trend seen by the black dashed line. That is why I decided to look at a theoretical approach in order to extract the shape of these particles from the data.

5.2.3.2.5 The theory: the roundness of the capsids and when it breaks !

Of course, it is easier to imagine viruses as perfectly round spheres when we want to develop simple mathematical models, but this does not perfectly match reality. We know that Caspar and Klug developed the T-number for viruses that present capsids, in order to explain their icosahedral (20 faceted) shapes[5], highlighting the complex geometry of these particles.

It has to be pointed out that, in what follows, I will focus on the case where capsids are initially modelled as spheres. This enables to develop simple arguments and mathematical equations. The more general elliptical case can be easily deduced from the spherical case, by applying to the latter a dilation factor.

In order to understand why the black dashed line in Figure 76 is found below the ($Perim/Perim_{calc}$) ratio, we can use a simple trigonometrical trick.

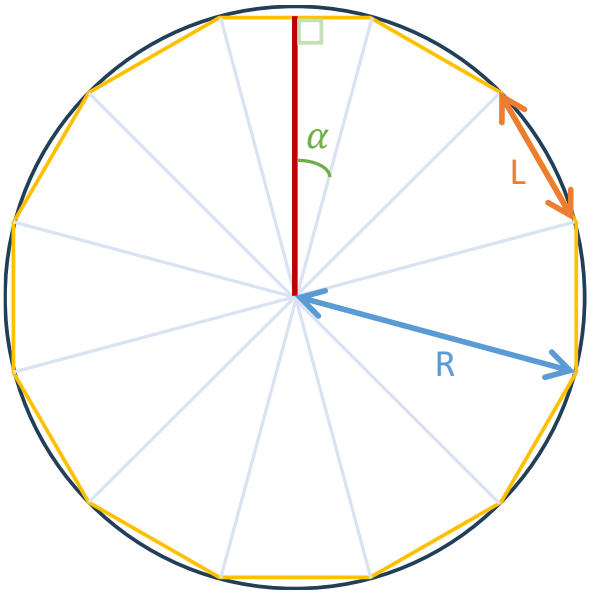


Figure 77. The circle englobing a regular polygon, example here is the dodecahedron. The variables of interest are indicated as follows: R is for radius (blue double arrow); L is for length (orange double arrow); alpha (in green) is the angle between the radius and the orthogonal line to L (red line).

We can thus write that the perimeter of a circle is:

$$Perim_{circle} = 2\pi R$$

And the perimeter of the polygon is defined by the number of its sides (n) as follows:

$$Perim_{polygon} = nL$$

If we were to draw an orthogonal line from the centre of the circle to one of the sides of the Polygon, shown in dark red, we would form an angle denoted as alpha (α). Thus, this angle would be:

$$\sin\alpha = \frac{L/2}{R} = \frac{L}{2R}$$

But the angle α can also be written as:

$$\alpha = \frac{\pi}{n}$$

Thus, we have L = 2Rsin (π/n), and we can rewrite the equation for $Perim_{polygon}$ as follows:

$$Perim_{polygon} = 2\mathrm{R}n \sin\left(\frac{\pi}{n}\right)$$

Therefore, we have that the ratio between the circle perimeters versus the Polygon perimeter can be written as follows:

$$\frac{Perim_{circle}}{Perim_{polygon}} = \frac{2R\pi}{2\mathrm{R}n \sin\left(\frac{\pi}{n}\right)}$$

Since:

$$\frac{\pi}{n} \geq \sin\left(\frac{\pi}{n}\right)$$

Then:

$$\frac{Perim_{circle}}{Perim_{polygon}} \geq 1$$

Thus, we can conclude that:

$$Perim_{circle} \geq Perim_{polygon}$$

This shows that the polygon being englobed by the circle will always be smaller than the circle starting from the same base. This is what we wanted to demonstrate from Figure 76. We believe that this is the most probable explanation for the black dashed line (Figure 76) appearing smaller than the perimeter ratio $Perim/Perim_{calc} = 1$ threshold.

The next step was to check if we could use the information derived from my image analysis, in order to get an analytical separation, which could easily help us define when we are looking at regular particles or not.

Hence, the next logical step was to build an algorithm that could do this selection for us, between spherical, elliptical, regular polygons and separate them from the broken capsids, simply based on the data acquired by the image analysis.

5.2.3.2.6 Algorithm to select the shape of a virus

Following the arguments developed above, I devised an algorithm to distinguish the different capsid types (polygons; spherical; elliptical; broken) as follows:

Algorithm to select the shape of a virus:

If ($r \leq \alpha$) & $\frac{\text{Perim}}{\text{Perim}_{\text{calc}}} \leq \beta$ then:

A virus is 'spherical'

If $\frac{\text{Perim}}{\text{Perim}_{\text{calc}}} \leq 1$:

A virus is "polygonal" in a sphere

Else if ($r > \alpha$) & $\frac{\text{Perim}}{\text{Perim}_{\text{calc}}} \leq \beta$ then:

A virus is 'elliptical'

If $\frac{\text{Perim}}{\text{Perim}_{\text{calc}}} \leq 1$:

A virus is "polygonal" in an ellipse

Else:

A virus has a 'broken capsid'

I defined the value of α for the aspect ratio (shown by the dashed olive line in Figure 78), by looking in detail at my data near $r = 1$ (in the range [0.999, 1.010] in Figure 76).

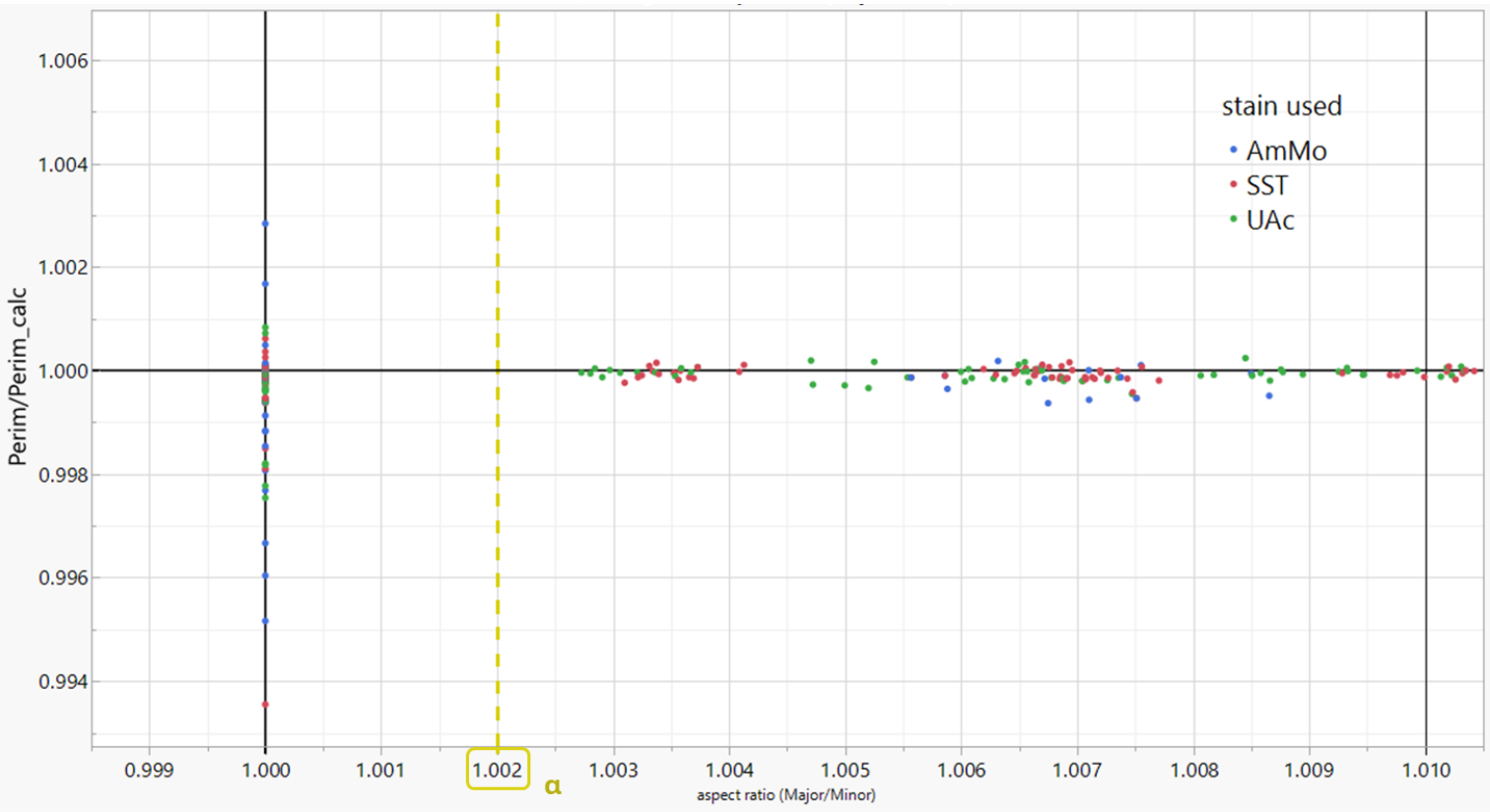


Figure 78. Choosing the alpha value (dashed olive line) in the aspect ratio range. Legend (top right): AmMo (blue); SST (red); UAc (green). The solid black lines serve to show the threshold where both axes are equal to 1. The line at aspect ratio =1.01 was also highlighted in order to have a better understanding of the zooming on the axis, when compared to Figure 76.

Recalling that a capsid with an aspect ratio r equal to 1 is spherical, a value of α close to 1 may be a good threshold to distinguish spheres from ellipses. Since there are almost no capsids whose aspect ratio lies within the range 1 to 1.0025, setting α value to 1.002 seems then reasonable. Since ellipses are identified as spheres of radius b when their aspect ratios lie within the range $[1,\alpha]$, their geometrical features should also be updated accordingly. In particular, their perimeters, initially obtained with the formula $Perim_{ellipse}$, boil down to the perimeters of spheres of radius b, i.e. $2\pi b$. Consequently, the ratio $Perim_{ellipse}/2\pi b$ can be seen as a threshold to discriminate between the perimeters of ellipses and spheres. This ratio is denoted β and is as follows:

$$\beta = \left(\frac{\alpha^2 + 1}{2}\right)^{\frac{1}{2}}$$

It is noteworthy that I proposed a value for the threshold α based on my current observations: this value may be modified with new available data.

Based on this algorithm, Figure 79 shows the proportions of spherical, polygonal, elliptical, and broken capsids in the overall capsid population.

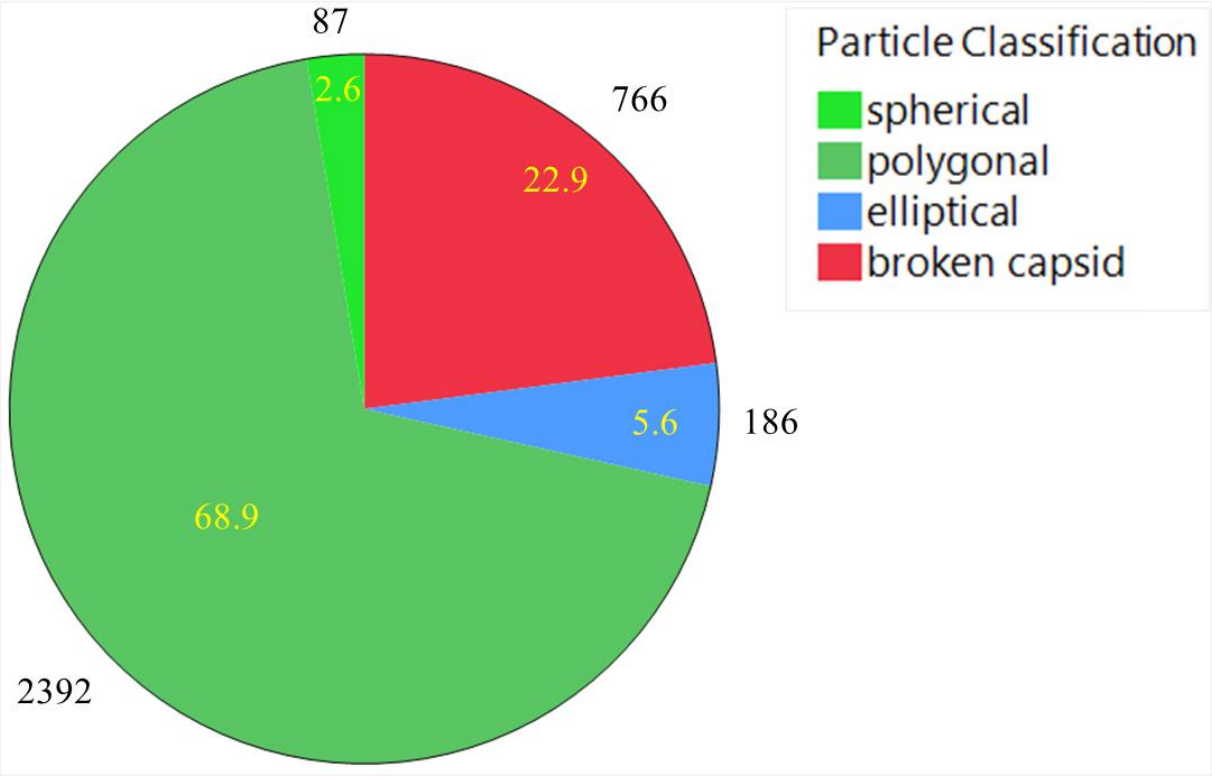


Figure 79. Pie chart representing the respective portion of different particles in all samples combines, depending on the following classification: spherical (light green), polygonal (dark green), elliptical (blue), broken (red). Numbers: percentage of the total count (in yellow, inside the pie chart); particle count (in black; outside the pie chart).

We can see here that the majority of the particles, ~ 69 %, present a polygonal shape. And the second largest fraction are the broken capsids, at ~ 23 %. What is left is separated between the elliptical and spherical capsids, at ~6 % and ~3 %, respectively. This would mean that the overall proportion of regular capsids (~77 %) is more than three times that of broken capsids, indicating that a major part of the population maintains its native structure.

Nevertheless, a slightly different picture dresses before us if we look at these same proportions but separating for the different stains and for the different samples, Figure *80*.

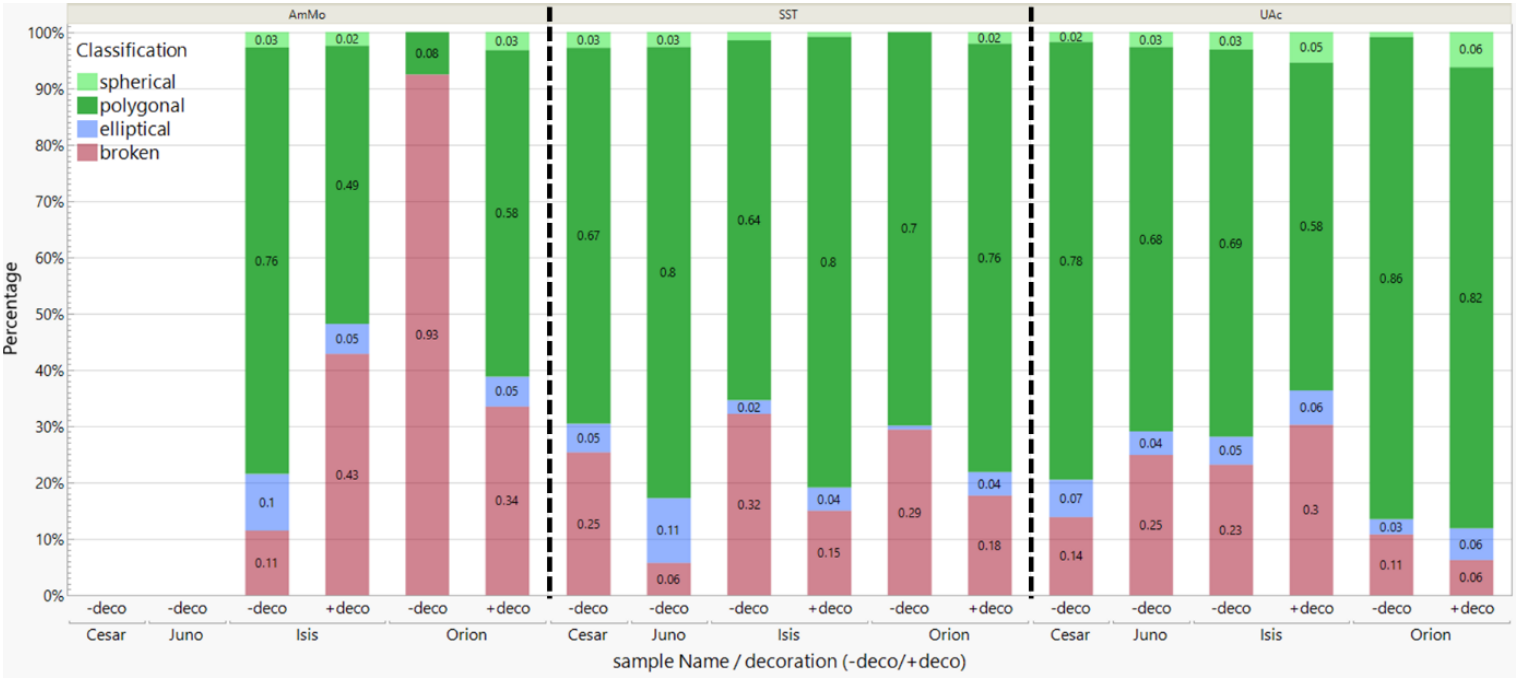


Figure 80. Portion of the capsids (in %), depending on their conformation, as a function of the sample name and on the presence (+deco) or absence (-deco) of decoration. Legend (top left): spherical (light green), polygonal (dark green), elliptical (blue), broken (red). The figure is split into three panels (left to right), depending on the negative stain used: AmMo (left); SST (middle); UAc (right).

It can be seen from Figure *80*, that the proportion of broken capsids is much larger for AmMo varying from 11 to 93 %, compared to SST and UAc. When using SST this percentage is less important, varying between 6 and 32 %. And it is between 6 and 30 % for UAc. This large discrepancy between AmMo and the other two stains is also probably accentuated by the fact that it has less the data points, especially true for Orion -deco (Table 10).

It can be also seen that the percentage of spherical particles is the minority, varying slightly between samples, with the largest proportions being for UAc, reaching a maximum of 6 %. Thus, the majority of samples seen across stains is polygonal, varies between 64 and 80 % for SST, and between 58 and 86 % for UAc. It can be added that for SST the proportion of broken capsids decreases when the decoration is present, compared to the non-decorated ones.

One can conclude from this graph that SST and UAc are very similar to one another and that the worst stain (i.e. causing more broken capsids) is actually AmMo. This is very interesting as it contradicts what the qualitative observations would have stated, as discussed earlier. Highlighting again the need to perform quantitative analysis in such a thorough manner and making it a necessity in order to have a better grasp of the results, with the risk of overlooking potentially important data.

With this new algorithm, I went back to the raw data shown on Figure 73 and verified if the capsid distribution would change, when accounting for the different capsid types (spherical, polygonal, elliptical, or broken).

5.2.3.2.7 Filtering the TEM data

The corresponding results are shown on Figure *81* and represent the distribution of the capsid conformation, depending on the stain used.

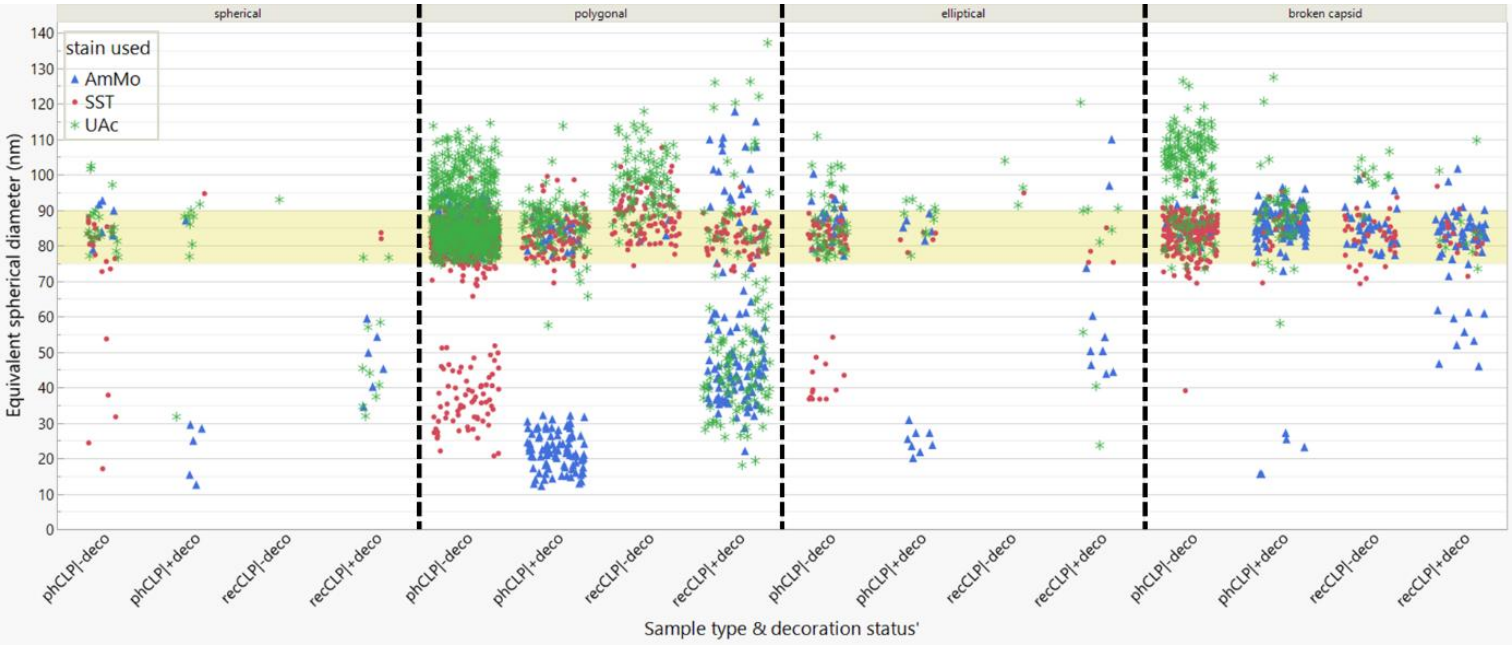


Figure 81. Dot plots showing the relationship between the ESD, in nm, as a function of the sample type (phCLP; recCLP) and decoration status (-deco; +deco). The figure is split into four panels (left to right), depending on the conformation: spherical (first), polygonal (second), elliptical (third), broken capsid (forth). Legend (top right) depending on the negative stain used: AmMo (blue triangle); SST (red dot); UAc (green asterisk). The yellow highlighted band is the size range 75-90 nm.

The first thing we notice is that the separation by size between regular or broken capsids is not that different, meaning that a simple classification by size is not enough to distinguish between stable structures or broken ones. We can also notice that this is primarily true for the AmMo and SST coloration. The UAc does seem to be the one that is dragging the size up for these capsids. So perhaps possible improvements can be done in the future to the filter to have a better cut-off to the filter, and thus recover even more regular capsids, perhaps lying in the orange shaded area in Figure 76. That is why having such an analytical approach on characterizing these particles is critical in separating native structures from damaged ones. The non-decorated capsids seem to be much closer to the expected value (yellow band). We can see also that the variability for the decorated capsids is much larger compared to the non-decorated ones.

It looks like the presence of decoration causes a higher number of outliers in these distributions, sometimes pulling them on the lower end, which seems to be especially true for AmMo (blue), when decoration is present. UAc (green) is the one always overestimating non-decorated particles, compared to the expected value. However, when decoration is present, UAc does not have that same 'oversizing effect' on the particles, perhaps indicating the stability of decoration on the capsid's structure. It seems that the staining by SST (red) is the most stable one, always remaining close to the central range, without much dispersion, being that with or without the presence of decoration. Thus, one can conclude that the staining by SST (red) is the most appropriate one when analysing these samples.

5.2.4 Conclusion on TEM

My examinations on this chapter regarding the optimization of negative staining TEM resulted in several improvements compared to the current state-of-the-art for our group, and hopefully to the larger scientific community.

My first contribution was to evaluate that several commonly used negative stains, such as UAc (Uranyl Acetate) play a major role in changing capsid structures, which I attributed mainly to the fact that its pH is acidic (pH=4). This was supported when compared to more neutral pH stains, such as SST (Sodium silicotungstate) and AmMo (Ammonium molybdate), both at pH=7. I also evaluated that the best negative stain to visualize decoration proteins attached to the capsids was AmMo.

Next, I quantitatively evaluated the sizes of the viral particles for all the different samples, in different conditions, in order to streamline and have data-driven comparison in order to be integrated in the analytical pipeline described in Figure *62*. This work helped build an analytical model, in order to distinguish and separate regular viruses from any irregular ones. This was possible by making the approximation that the rounder the capsids, the more regular they are. This allowed us to build an algorithm that can evaluate the number of capsids that are broken or shaped elliptically, allowing an effective filter that can be useful in future research. A major improvement to this work would be if this data can be used to feed a machine learning algorithm that can analyse future images containing viruses, and thus distinguish them for statistical analysis, effectively measuring their shape and structural conformation.

5.3 Nanoparticle Tracking Analysis (NTA)

Nanoparticle Tracking Analysis (NTA) allows the determination of the diameter of nanoparticles in suspension based on their Brownian motion. In this method, individual nanoparticle locations are tracked over time; the nanoparticle's mean squared displacement versus lag time is related to the nanoparticle's hydrodynamic diameter through the Stokes-Einstein equation. Nanoparticle suspension concentration is additionally determined by counting the average nanoparticles in a known volume in the field of view.

5.3.1 Principle of NTA

The diffusive motion of particles under Brownian motion is resisted by viscous drag. The balance of these effects is modelled by the Stokes-Einstein Equation (see below). This equation predicts that larger particles move slower than smaller particles under the same conditions. NTA exploits this dependency to determine particle size. In this method, a laser beam is passed through the particle suspension and light scattered by the particles is recorded by a video-microscope (Figure 82 a). The particle size cannot be determined directly from the microscope image since it is a visualization of the particle-scattered light and not of a resolved image.[246] However, the centres of the scattered spots indicate the particle locations, which are tracked over time to calculate the two-dimensional mean squared displacement. This results in using these particle tracks to infer the particle diffusivity (Figure 82 b). The diffusion coefficient is used to calculate the particle diameter, and the particle counts per frame are determined to calculate the sample concentration.

$$\frac{\overline{(x,y)^2}}{4t} = D_{x,y} = \frac{k_B T}{3\pi\eta d}$$

The Stokes-Einstein equation relating a particle's diffusion coefficient to its hydrodynamic diameter. $D_{x,y}$: diffusion coefficient; k_B: Boltzmann's constant; T: temperature; η: solvent viscosity; d: sphere-equivalent hydrodynamic diameter.

The particle size determined by the Stokes-Einstein equation is the sphere equivalent hydrodynamic diameter, which is the diameter that surrounds an ideal spherical particle, diffusing through the solvent. This ideal particle has the same diffusion rate as the particle being measured (e.g. Figure 82 d). This model assumes that the particle is a hard sphere moving under Brownian motion in a solution of specified temperature and viscosity. The model further assumes that the particle moves slowly enough that inertia can be neglected.

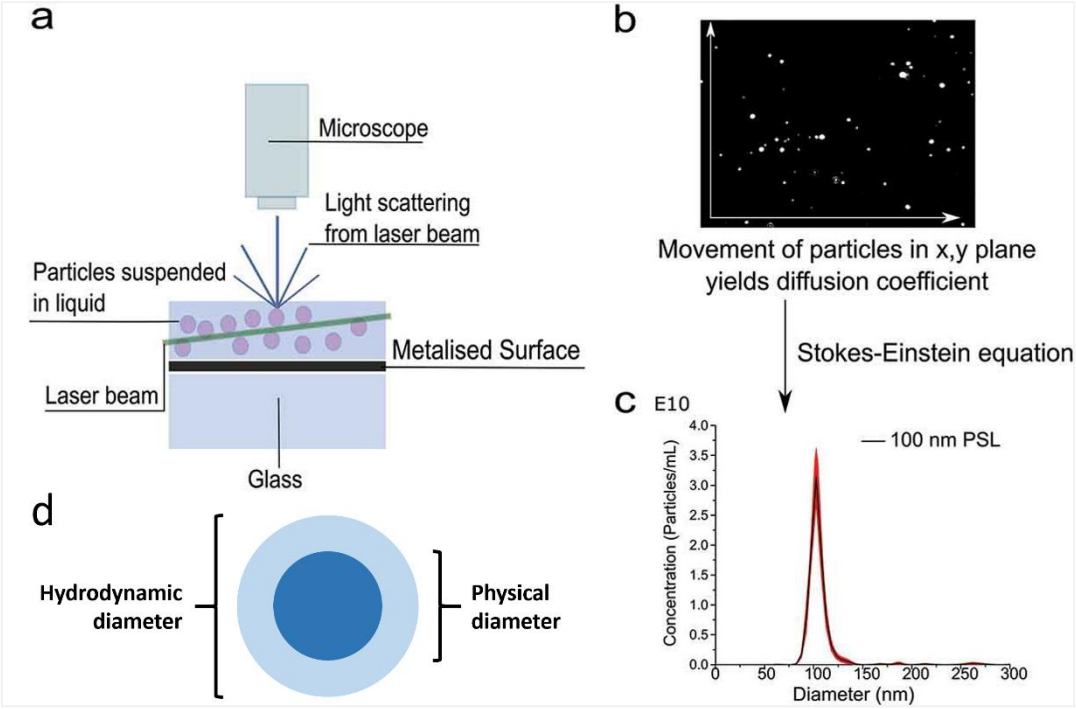


Figure 82. NTA set-up schematic (a) Suspended particles scatter light from a laser beam, which is recorded by a video-microscope. (b) Software analysis of the videos locates and tracks the centre of each particle over time to determine particle count and diffusion coefficients. (c) The diffusion coefficient and particle count are used to calculate the particle hydrodynamic diameter and concentration. 100 nm polystyrene latex nanoparticles were measured. (d) The hydrodynamic diameter inferred from the Stokes Einstein equation is larger than the physical diameter due to the inclusion of the ionic double layer. The figure was adapted from Maguire et al.[247]

Gold nanoparticles as small as 10 nm can be sized by this method, however biological particles must be at least 35 nm to be accurately sized, since they do not scatter as much light and cannot be visualized as easily. Brownian motion limits the upper bound to ~ 1000-2000 nm. Measurable suspension concentrations range from ~10^7 to 5 x 10^9 particles/mL. At the lower limit, an insufficient number of particles tracked leads to poor statistics, while at the upper limit, particles are too close together causing particle paths to cross creating distinct particle tracks of insufficient length and number.[246]

5.3.1.1 Measurement Quality Control criteria

Quantitative quality control metrics are imperative to ensure measurements are reliable and are within easily determined measurement ranges. Malvern, which is the company that has developed the NTA technology, has provided some parameters, although the values are not explicitly listed in their manuals. Some parameters control for noise in the measurement, while others for screening enough particles to create a robust size distribution.

Particles per frame (particles/frame) indicates whether the sample concentration measured is appropriate for NTA. Malvern recommends that particles/frame remain between 20 and 100. Counts below 20 indicate too few particles for statistical analysis. Samples with

counts over 100 are too densely packed in frame, causing particle trajectories to cross, so the software cannot gather sufficient information on each particle.

Ideally, each particle tracked has a singular centre, therefore the ideal ratio between centres per frame (centres/frame) and particles per frame (particles/frame) would be equal to one. Deviation from this one-to-one pairing indicates measurement noise. NTA software identifies centres and particles based on the detection threshold (DT) selected. This DT parameter is something that the user inputs at the beginning of each measurement, but it is very user dependent, and it can heavily affect the measurement. The main reasons being that this parameter is fixed for the entirety of the measurement, and thus not flexible in case there needs to be any troubleshooting during the experimental run. It is recommended that the ratio of centres/frame to particles/frame be < 1.2.[246] In our coded output, we include [(centres/particles) – 1], which represents the ratio of centres/frame over particles/frame minus 1, which should thus remain <0.2 for valid measurements.

Finally, the total number of particles tracked in a video reveals the sample size for determining the particle distribution. When flowing samples, the sample volume recorded is not restricted to the volume in the video frame, as it is for stationary samples. Thus, the total particles included in the distribution becomes a better indicator of quality, than the average particles per frame as a lower bound. Comfort et al. suggests a minimum of 150 particles tracked during a sample.[248] We decided to put this cut-off at 500 particles, in order to make sure that we always had sufficient data for relevant statistics.

Comparing X-Drift and Y-Drift may be helpful in determining whether unknown outside forces are influencing the sample. For stationary samples, both parameters should be close to zero, while for flowing samples, X-Drift will gain a significant positive value. An exact use of this as a screening parameter has not been determined, but could be useful in troubleshooting the system, especially for clogging or for the presence of bubbles.

5.3.1.2 Summary Statistics

Measurement results can be quickly determined using summary statistics. For particle size, the mode size is mostly used since this is the particle size occurring at the highest concentration. If no dilution factor is input when making a measurement, the initial sample concentration can be calculated by multiplying the output concentration by the dilution factor. Sample concentration is based on the average particles per frame count and a multiplier representing the acquisition volume.

Displaying the lowest 10 percentile (D10), median (D50), and the highest 10 percentile (D90) aids in understanding spread within a measurement. Therefore, previous work has suggested that a good indicator of polydispersity, is the so-called Span, calculated as (D90-D10/D50).[249] As this value approaches zero, the distribution typically is more monodisperse and uniform. This should not be used as the sole indicator of polydispersity in the sample. Another such parameter of interest we observed is the so called relative standard deviation (RSD), which we calculated as (St.Dev./Mean). This value is usually presented as a percentage, and same as Span, the closer the value is to zero, the more monodisperse the sample is.

5.3.1.3 NTA vs DLS, and the need for FTLA

By default, the NTA particle distribution data is output in their reports as Finite Track Length Analysis (FTLA) data. This method takes the amount of data present on each particle tracked into account when building the most probable particle size distribution. FTLA's purpose is to yields much more distinct peaks than raw data, and helps to remove outliers caused by insufficient track lengths on particles.[250] This is summarised in the following figure, Figure 83, taken from Maguire et al.[247], the same analysis was performed by NTA and DLS (dynamic light scattering) on a mixture of PSL-NPs, containing particles of 3 different sizes, i.e. 100, 200 and 380 nm. The graph A (on left) shows the data in a linear scale, and the graph B (on the right) shows the original DLS data, presented in a log scale.

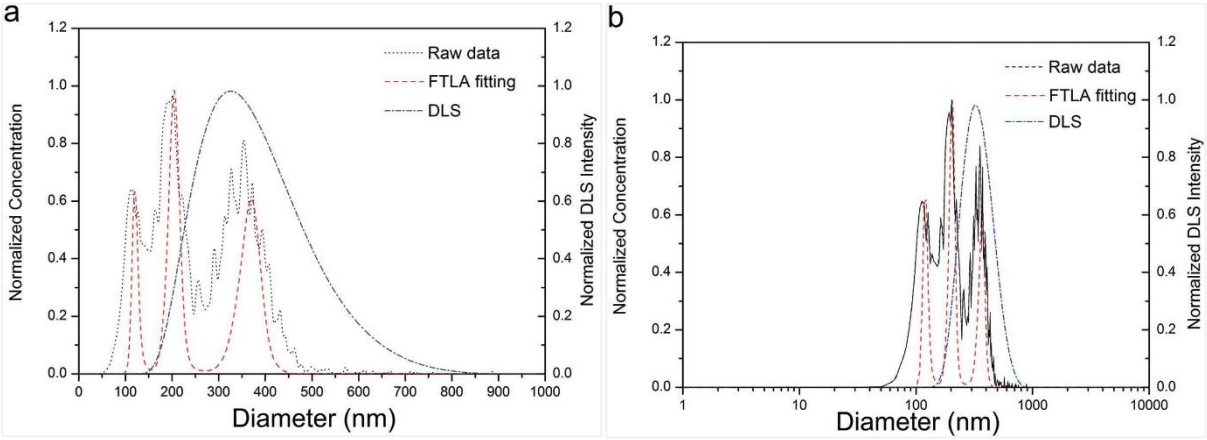


Figure 83. Analysis of a 100, 200 and 380 nm PSL NP mixture using NTA and DLS. Black dotted line represents raw unprocessed NTA data, red dashed line represents NTA data following FTLA being applied, and blue line represents the distribution of the same sample as analysed by DLS. (a) Transformation of the DLS data from a log to linear scale, (b) original data. The figure was taken from Maguire et al.[247]

In Figure 83, we see three different distributions fittings: the raw data (black dotted line), the FTLA (red dashed line) and the DLS (blue line). We can see that the FTLA fitting matches very well to the expected particle sizes, and yet preserving the information of the raw data,

allowing clear peak separation for the 100 and 200 nm peaks. On the other hand, the DLS peak only fits around the 380 nm peak. This is because the DLS performs a bulk measurement, and thus the larger particles will draw the mean size up, and of course we can only see one distribution, thus any separation between species is impossible.

5.3.2 NTA optimisation

The following is a series of experiments and analysis I performed, with a variety of samples and checking a number of parameters, in order to get consistent reliable results, and improve the accuracy of the data to our needs.

5.3.2.1 Checking the lower detection limit: with Gold-NPs

Since the NTA is highly dependent on the refractive index of the particles analysed, I tested it with GNPs, which were already used for the NEMS-MS measurements, shown in the previous chapter (Figure 46). Here I tested different parameters such as the camera level and the dilution factor, shown in Figure 85, for 20 nm GNPs.

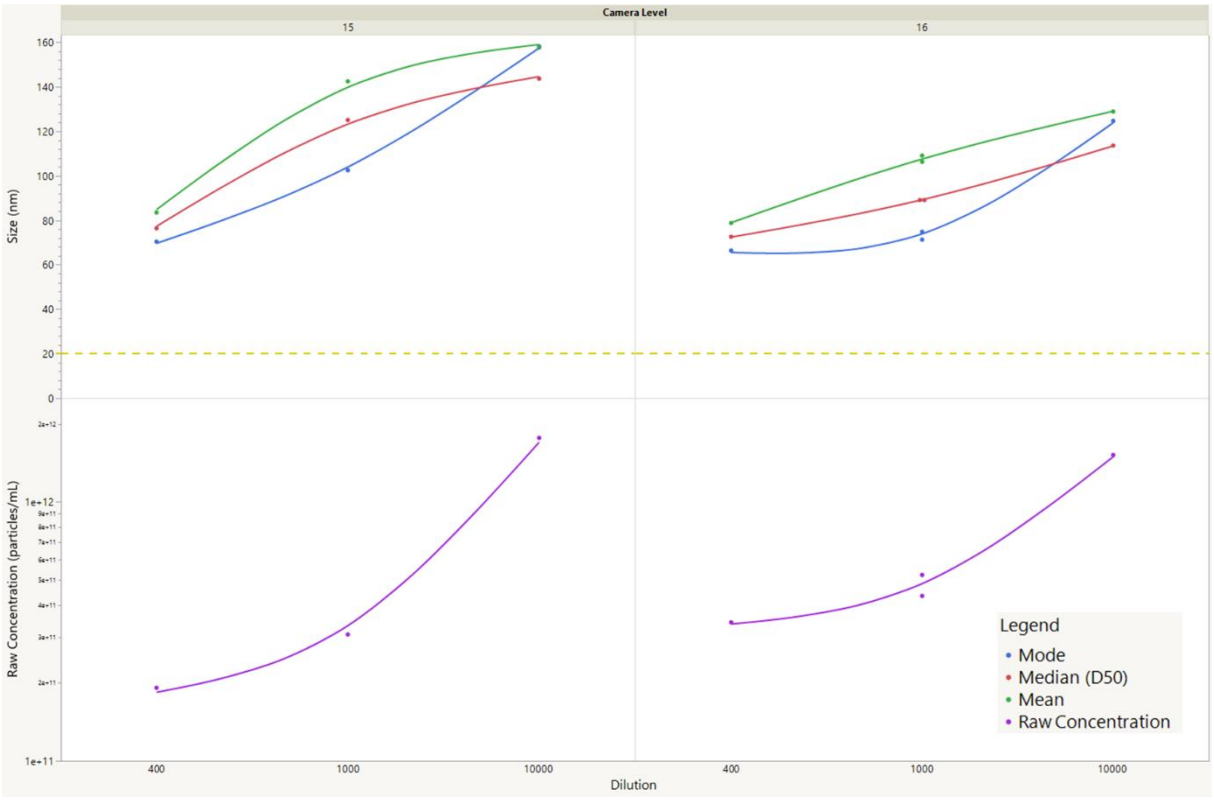


Figure 84. NTA measurements of the size (nm) and particle concentration (particles/mL) of 20 nm gold NPs (GNPs), as a function of dilution factor. The colour coding corresponds to the different parameters of size: blue= mode; red= median; green=mean. Left panels CL=15; right panel, CL=16. The horizontal dashed lines are placed at 20 nm, to indicate the theoretical size expected, as denoted by the manufacturer. Legend on the bottom right of the graph. Each point is an average of n=10 measurement.

In Figure 85, the first thing we notice is that none of the size values even comes closer to the expected size value (yellow dashed line). This is independent of the dilution factor or the camera level that we use. The size increases drastically when we go further up in dilution factor. This drastic size increase can be perhaps due to an aggregation of particles in solution, which can happen to GNPs in water at such high dilutions. In addition, it can be that the instrument is incapable of detecting such small particles, and they are below the detection threshold described by the manufacturer. Similarly, the same upward trend can be seen about the concentration values, increasing by an order of magnitude for camera level = 15. The concentration values seem to be more stable for the 400 and 1000 dilution factor, for both camera levels, indicating that perhaps the particle count per frame is in the optimal range for these values compared to the 10 000-fold dilution.

Therefore, for further technical optimization we decided to go to the standard particles provided by NTA manufacturer (Malvern), made of polystyrene (PSL-NPs), which varied in size from 100 to 200 and 400 nm. These are of great interest as the viruses of interest are also in the similar range, so we expect them to be behaving in the same way in regard to the particle Brownian motion. In addition, they have a closer refractive index to that of viruses, due to their density being similar.

I would like to highlight here that our group does not have the NTA provided 'sample pushing pump', which is calibrated for this instrument's usage. Therefore, one of the topics that I worked on was to implement a classical syringe pump on our system. Therefore, I used these PSL-NPs to evaluate the effect of the flow rate and observe if and how the size measurement would be affected by this new implementation.

5.3.2.2 Evaluating the effects of the flow rate on the sample size and concentration

The first step that I evaluated was in running the three samples, that is PSL-NP = 100 to 200 and 400 nm (Figure 85), for different flow rates.

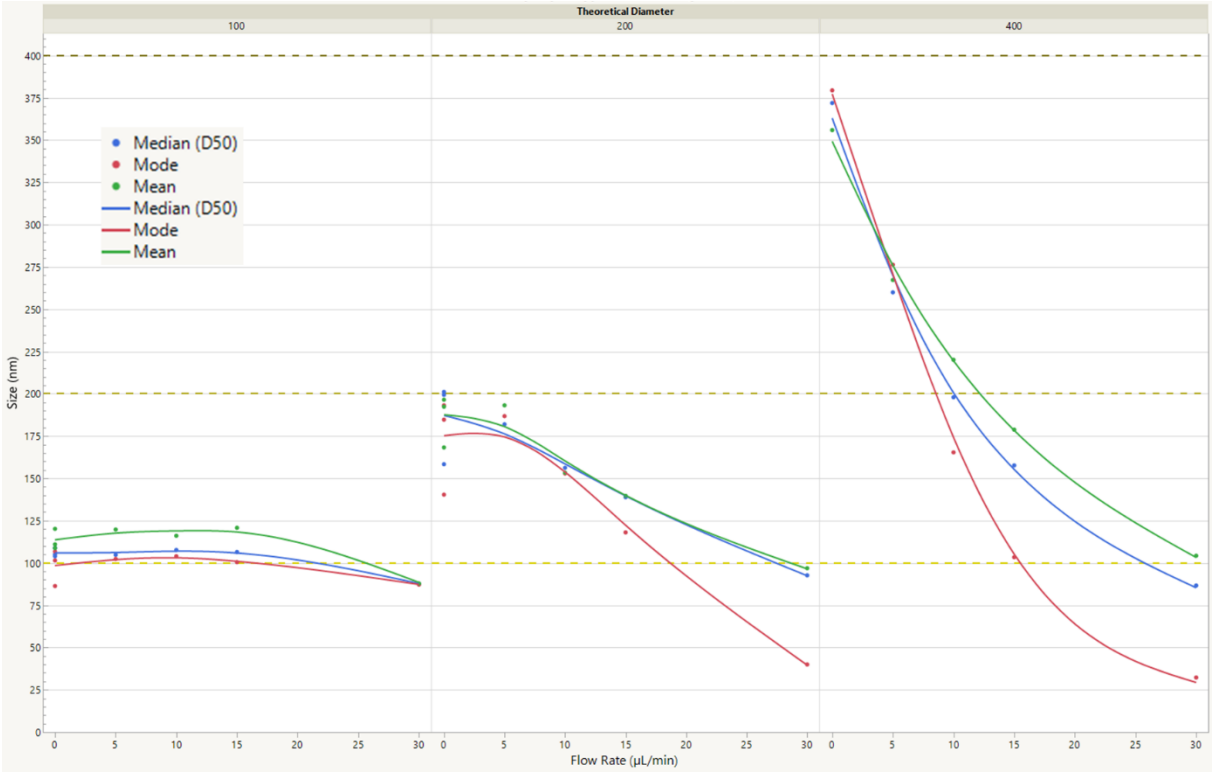


Figure 85. NTA measurements of the size (nm), as a function of flow rate (µL/min). The colour coding correspond to the different parameters of size: blue=median; red=mode; green=mean. The horizontal dashed lines are placed at 100, 200 and 400 nm, to indicate the theoretical size expected, as denoted by the manufacturer. Each point is an average of n=10 measurements.

We can see from Figure 85 that the size is more stable for static flow rate, being the closest to the expected size for all three samples. But when the flow rate is increased, only the 100 nm sample is stable all along the flow rates, with a slight decrease at 30 µL/min. Whereas the 200 nm sample is decreasing rapidly after the 5 µL/min point, going as down as 30 nm for the mode when the flow rate is 30 µL/min, signifying a 5-fold decrease in size. The most drastic changes are seen for the 400 nm sample, where the size goes from an average mode size of nearly 375 nm for static measurement, to nearly 25 nm for the 30 µL/min, signifying a 15-fold decrease in size. This indicates that the flow rate has a drastic effect on larger particles and that static measurement is highly suggested in order to account for these changes.

After looking at the size data now we look at how the concentration is affected by these different flow rates. As one would expect this shouldn't be something that causes any differences, since we are simply moving the particles faster, but moving the same volume, thus the concentration should remain the same. This data is plotted in the following figure, Figure 86.

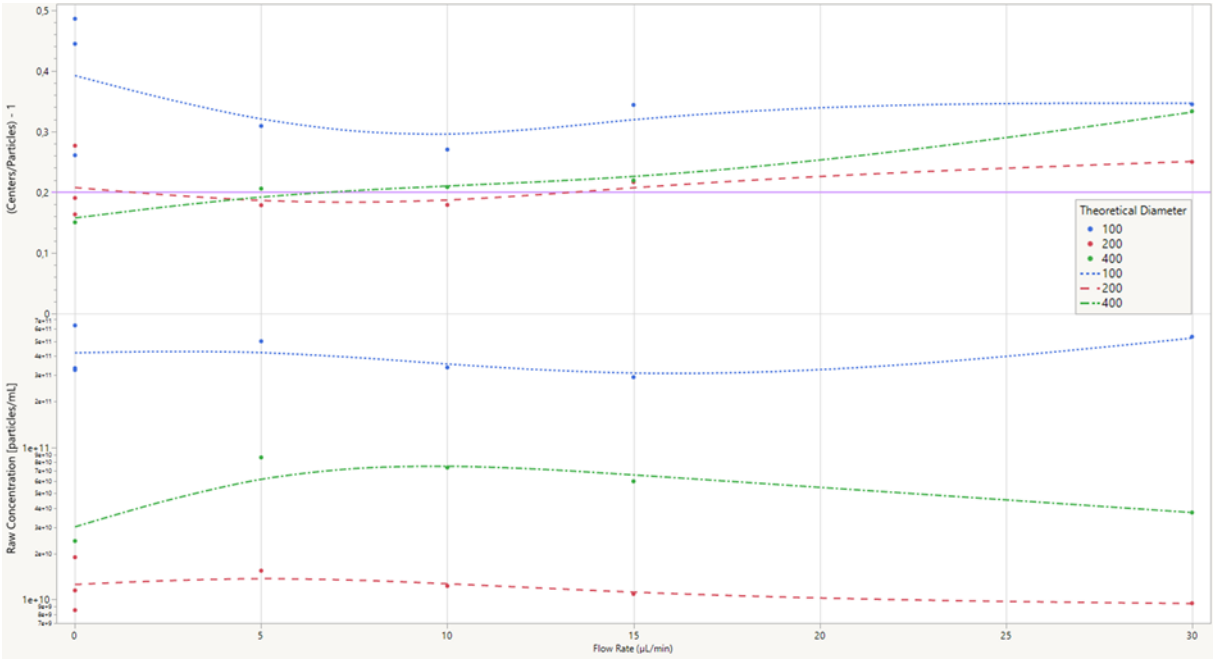


Figure 86. NTA measurements of the ratio (centres/frame)/(particles/frame) minus one, written as [(centres/particles) -1], and raw particle concentration (particles/mL) as a function of flow rate (µL/min). The colour coding corresponds to particles of different expected size: blue=100 nm; red= 200 nm; green= 400 nm. Dilution factor = 100x (for both 200 & 400 nm particles); 5000x (for the 100nm particles). Each point is an average of n=10 measurements.

We can see from Figure 86, that the sample of 100 nm, in blue, is entirely over the 0.2 threshold, indicating that we are possibly overestimating all of our measurements, although these were conducted for the suggested dilution from the manufacturer. The 200 nm, in red, and the 400 nm, in green, measurements seem to be more stable, but tending towards values above the 0.2 threshold when we go in the upper flow rates, especially at 30 µL/min. Concerning the concentrations seen in the lower panel, we see that the values tend to be much more stable indicating that the concentration for the 100 nm changes very slightly, between 3 to 5e11 particles/mL. Concerning the 400 nm sample, the concentration is mostly in the range between 5e10 to 1e11 particles/mL. Lastly, for sample 200 nm, the concentration is in the range 8e9 to 1.5e10 particles/mL.

Thus, as expected the sample concentration is not affected drastically by the increasing flow rate, as much as the size determination was. Unfortunately, only the dilution factor suggested by the manufacturer was given, and not the original concentration of the particles.

5.3.2.3 Effects of the video length and DT on the statistics

Therefore, we were now ready to analyse the particles of interest. In the following analysis I investigated if the video duration had any effect on the particle concentration

evaluation or if the detection threshold would have an effect as it had been already reported in the literature.[251]

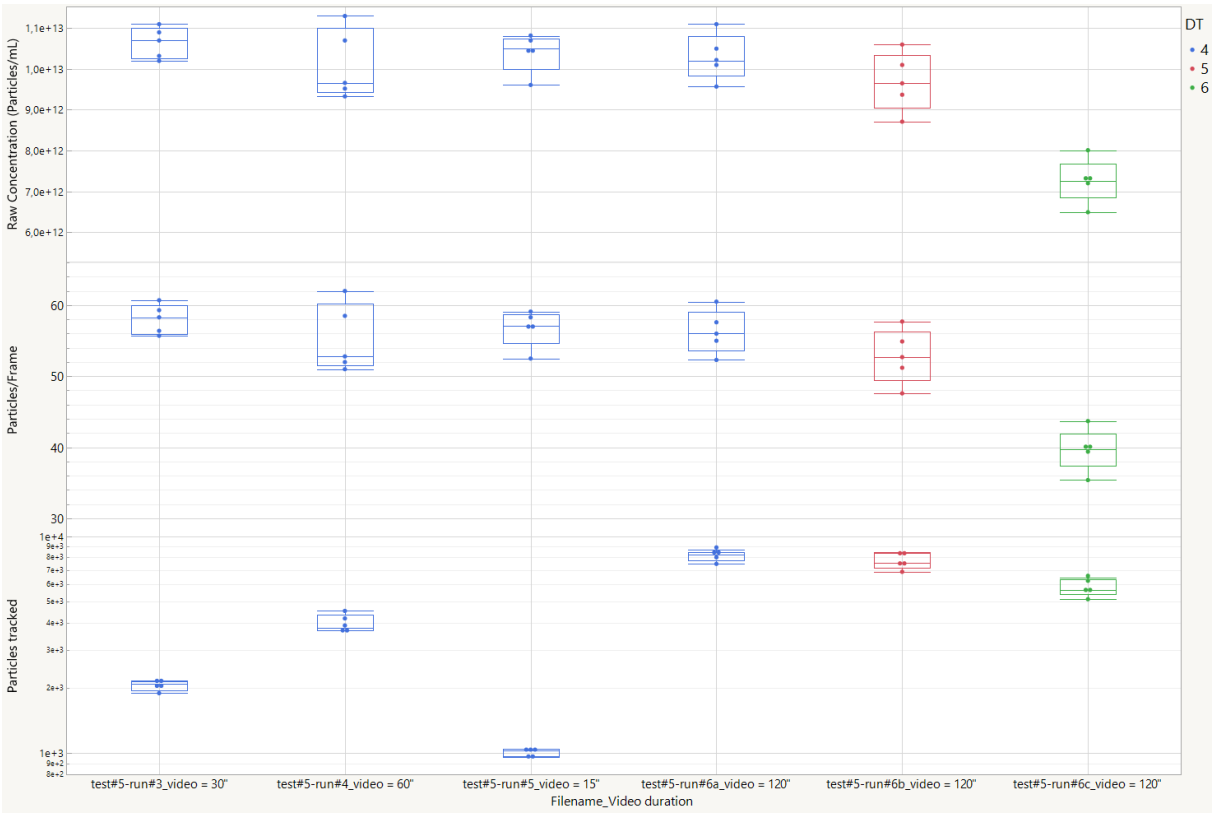


Figure 87. Boxplot distribution of the NTA measurements of sample Cesar, separated into three horizontal panels: 1st: for the raw concentration (particles/mL), 2nd: particles per frame, and 3rd: particles tracked in total. The horizontal axis indicates the different runs that were performed, with their corresponding file name, given as a test number and a run number, and separated by the underscore (_) is the length of the video acquired. Different detection thresholds (DT) are indicated by the colors: blue=4, red=5, green=6. Legend found on the top right of the graph.

It can be seen that the particle total count is linear in time with a linearity factor of 1. We see no effects due to the order of the measurements, given that these runs are not in specific increasing or decreasing order, regarding the video length. However, most importantly we see that the particles per frame is almost identical, varying between 52-60 particles/frame between the four runs, shown in blue. Moreover, the same thing can be said for the particle concentration, only slightly varying between 9.5^{e}12 to 1.1^{e}13 particles/mL.

We can see that the detection level threshold is a much more important parameter as we see when we compare between the post processing of round 6, that is in a DT= 4 (blue), DT= 5 (red), DT= 6 (green). We can see that between DT equal to 5 and 4 the distinction is rather minor, although decrease his scene for the central values in all three panels. This is much more visible for DT= 6 where we see that the distributions are not even overlapping anymore, without changing any of the experimental conditions. Thus it has gone from nearly 56 particles/frame on average for DT= 4, to around 40 particles/frame on average for DT= 6.

From these results, we can conclude that the video length does not have any effect on the particle concentration calculation. Whereas the detection threshold (DT) value has very noticeable effects on the concentration and users have to be very cautious how they integrate these values in the instrument during their experimental setup.

5.3.2.4 The importance of choosing the right 'buffer': PBS versus Water

Now that I examined the instrumental parameters, I focused on regarding the biochemical parameters that could affect the measurement. The idea was to compare water, as a volatile solution, to PBS, a non-volatile salt solution, in order to have a comparison of the samples prior to performing mass measurements. My first test, shown in Figure 88, was to assess the stability of T5-CLP samples diluted in their physiological conservation buffer, PBS-1x (Phosphate Buffer Saline, in blue), and compare with dilutions in pure deionized water (in red).

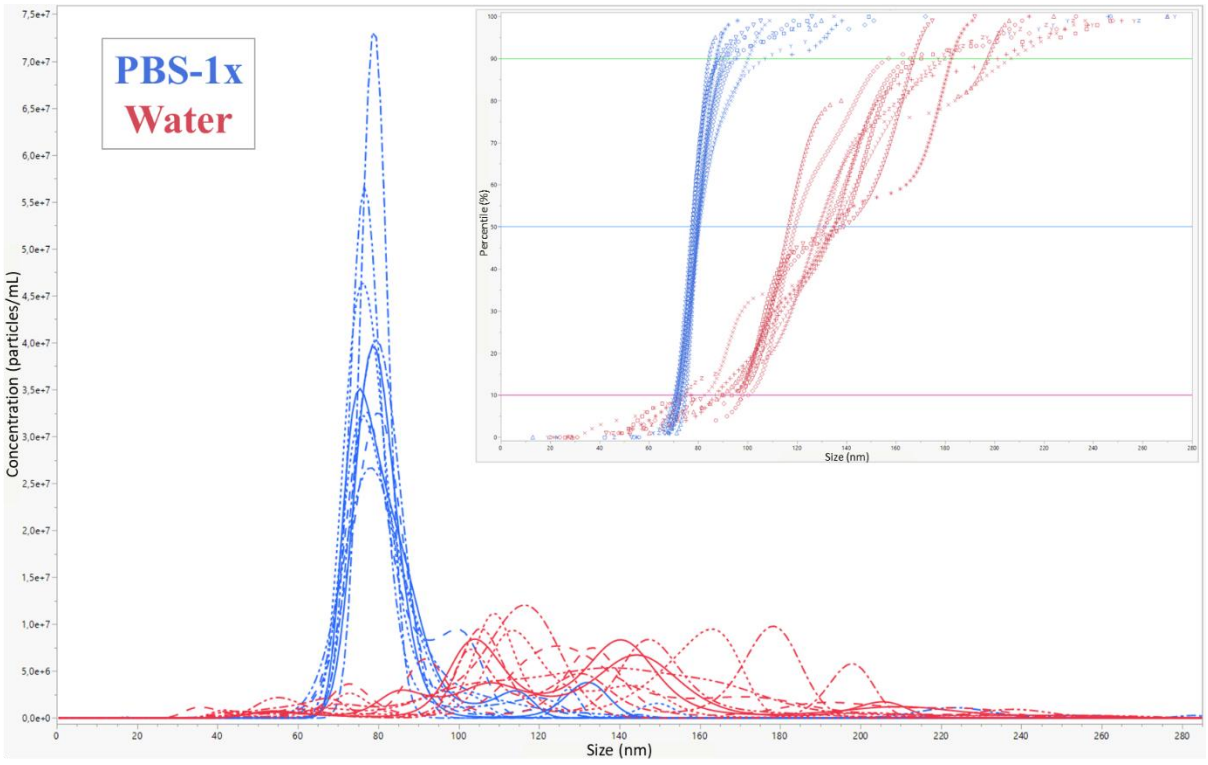


Figure 88. NTA measurement, of Juno (phCLPs -deco) in PBS-1x [blue] and Deionized Water [red], for dilution factor = 70000x. Particle concentration measured (particles/mL) as a function of the size (nm). ***Inset (top right)****: percentile as a function of size. Horizontal lines: D-10% [pink], Median [light blue], and D-90% [green]. Each measurement: n=10; duration of acquisition = 30"/video; flow rate =10 µL/min. Legend: top left of the figure.*

A clear increase of particle size in water compared to PBS-1x was observed, having a respective median size range from 120 to 140 nm for water, and between 75 to 80 nm for PBS, as can be seen in the inset of Figure 88. Furthermore, the histograms for PBS clearly overlap between the 10 measurements, whereas for water the peaks are not at all overlapping and even

presenting multimodal distributions, clearly visible by the cascade curves seen in the inset of Figure 88.

I suggest that these distributions behave in this manner for water, due to possible particle aggregation, indicated by their higher size and lower concentration (depending on the peak height) compared to PBS. In addition, we cannot exclude that particles in water may be subject to swelling due to an osmotic pressure introduced into them when passing from my high salt solution to water. Furthermore, the fact that larger particles are noticed does not exclude the fact that phages might be present, as some PBS measurements do show several other peaks, ranging from 100 to 130 nm. In addition, as seen for TEM this phage presence is clearly present for the sample Juno (cf. Figure 68), contributing to the multimodal distributions. These results showed that there was indeed instability when using water as a measurement buffer for further NTA analysis, thus resulting in all the rest of our measurements being performed in PBS-1x. But it also serves to highlight that when performing mass measurement, we need to be careful when analysing the data, knowing that solutions like deionized water can be detrimental to the particle's initial structure.

In order account for this already known dispersity, the NTA uses a software tool to evaluate the number of species in solution, separating them by size, called FTLA, which I will discuss further hereafter.

5.3.2.5 To FTLA or not to FTLA, this is the question …

The next step that we noticed was that the stability between measurements due to the Finite Track Length Adjustment (FTLA) optimization was not as adapted to our sample as suggested by the literature. The differences between the raw data and the data derived from the FTLA can probably be explained by the principles underlying the operation of the FTLA. Without going into detail, the article describing the FTLA[250], generalises to all forms of distribution an approach described in an earlier article.[252] However, the latter article only mathematically demonstrates the validity of the FTLA approach for one and only one type of distribution, the Poisson distribution.[252] The generalisation proposed by the FTLA is therefore not based on any rigorous demonstration and its area of validity is undoubtedly very limited. It is therefore more reasonable to consider only the raw data.

We can see in the following figure, Figure 89, that for the same measurement we have very different distributions when we look at the raw data (left of Figure 89), and of the FTLA data distributions, provided by the software, (right of Figure 89).

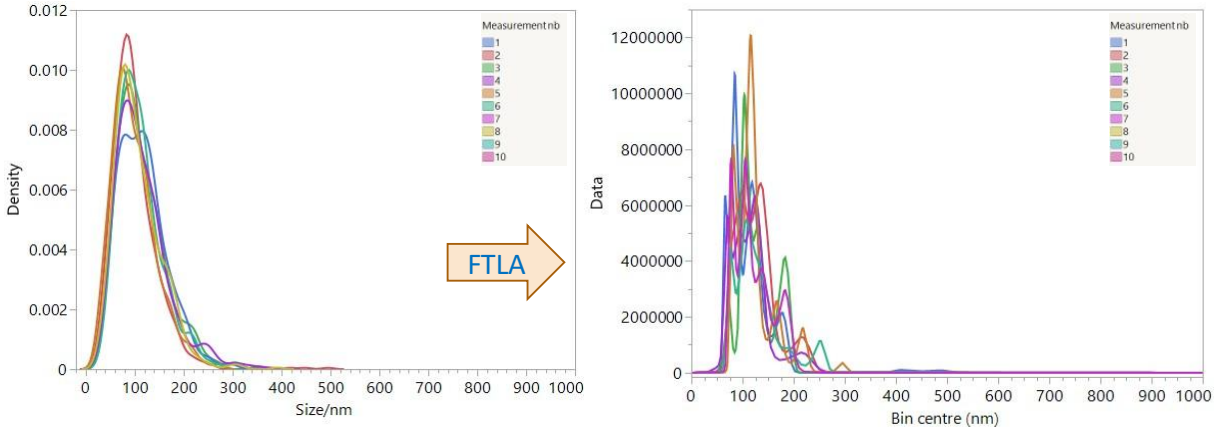


Figure 89. Comparing data from the raw file (left) versus after the Finite Track Length Adjustment (FTLA) data post processing (right). The data are plotted as the density function versus the size in nanometres, on the left, and as a function of the bin centre nanometres, on the right. The same 10 measurements are presented in both figures, with the corresponding colour coding on the top right of each graph.

We see that the peaks are overlapping somewhat for the FTLA data, but many other smaller peaks are seen in the range between 150 and 300 nm, which one would interpret as possible aggregation, but without being able to retrieve the information from this data in a useful manner. We can see from the raw data distribution that the peaks are much more consistent between the different runs, and that the distribution is much more dispersed compared to that seen after the FTLA data processing. Therefore, we need to find a better way to plot the data than using the instrument's default FTLA analysis.

Therefore, I plotted the raw data and evaluated all the models available in the JMP-SAS statistical software, which checks between 11 different models, one of them being the normal (Gaussian) distribution. It compares parameters such as AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion). I was able to evaluate that the model that best fit to our data was the natural log (Lognormal) fitting. A visual example of this fit is shown by the example for Cesar (phCLP) below (Figure 90).

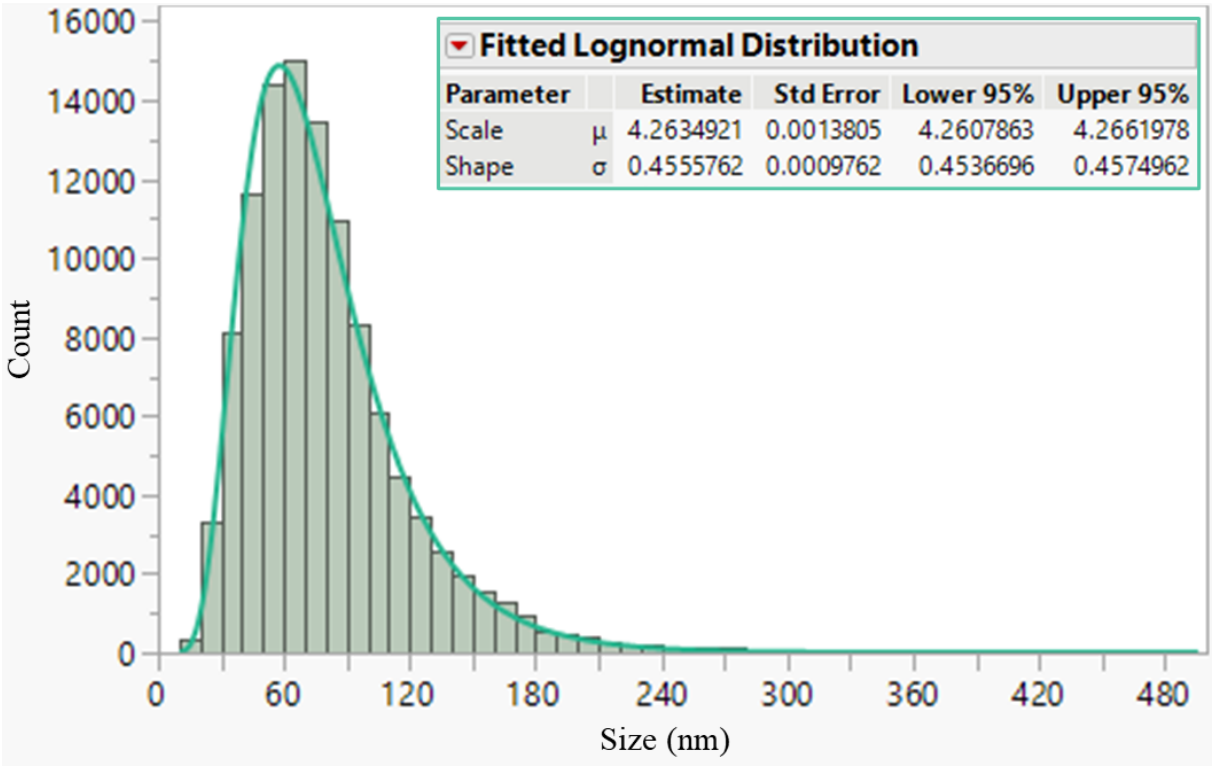


Figure 90. Lognormal fit on NTA data. Example sample Cesar (phCLP, non-decorated). N= 108905 particles counted. Bin size = 10 nm. The fit parameters are presented on the top right corner of the graph.

A lognormal distribution arises in observations where the underlying variable is the product of many independent positive random variables, meaning that it is grows or decays in a multiplicative, exponential fashion rather than an additive one. This leads to a skewed, right-tailed distribution of the particle size values. An important thing to highlight here is that the parameters μ (mean, scale in Figure 90) and σ (standard deviation, shape in Figure 90) are not the mean and standard deviation of the original particle size. They represent the mean and standard deviation of ln(particle size), which is normally distributed. This means that the geometric mean and the median represent better the central tendency, rather than the arithmetic mean. Now that we know what the properties of this fit are let us look at what this physically means for the particle size of these viruses.

It is very well documented that many particle size distributions are best fitted by lognormal distribution. This has been shown to be the case for example for : soils produced largely by glaciation in the higher latitudes on Earth [253]; non-spherical particles on hopper discharge characteristics [254]; particle properties in atmospheric aerosols [255]; optical properties of colloidal plasmonic nanoparticles [256]; in addition to gold [257,258] and silver [259] nanoparticles. This phenomenon has never been explicitly reported for virus' particle size, as far as we know.

For example, this is the best model to plot experimental data in aerosol science, as shown SMPS™ measurements (cf. later on, Figure 104). This is because the particle sensors in the nm

size range measure total number and mass concentrations (N, M), and by assuming a log-norma1 size distribution, the range of possible distribution parameters can be bound to (d_{gN}, σ_N), which are the geometric number-median diameter and the geometric standard deviation of the distribution, respectively.[255] The simplest way to explain this model would be that particle size cannot be negative; therefore a symmetrical distribution (e.g. the Gaussian one) cannot be applied. Irani and Callis, in 1963, derived it for nanoparticles from a kinetic based approach, by assuming that times for particle growth and destruction are normally distributed.[260] The particle growth rate (i.e. negative for shrinking particles) is considered to be linearly dependent on particle size, for an unrestricted minimum or maximum size. An additional explanation can be the presence of non-spherical particles (especially ellipsoids or cylinders) in the overall population, which would drag the distribution (i.e. extended tail) and thus increase the probability of particles appearing larger.[254]

5.3.3 Conclusion on NTA

In this chapter, I described how we optimized and adapted the NTA instrument for our purpose, in addition to putting it forward as a potentially important structural biology method.

First, I demonstrated that the lower detection limit proposed by the manufacturer (i.e. 10 nm) was not adapted for evaluating the size of standard gold nanoparticles (in my case 20 nm), which are useful for us as mass calibrants (Figure 84). This means that in the future this method either requires further optimisation in order to examine these particles, or there are indeed physico-chemical problems that can arise for these particles when we dilute them in water for the NTA examination, that make the particle appear bigger.

Then I improved the statistics of this measurement by integrating the syringe pump to our analytical pipeline, compared to the static hand pushed method that was used before. The previous approach led to selection bias and was largely user dependent, which can significantly false the experimental observations. I also helped in the building of a python code, in order to have more control on better filtering and selecting the data for further statistical examination, based on the evaluation of the quality control parameters established by Malvern.

Then I evaluated some of the most visible contributes of the syringe pump integration, such as those related to the flow rate and the detection limits, especially in term bias on particle size (Figure 85) and concentration evaluation (Figure 86). This examination showed that the standard particles that were larger than 100 nm were the most affected by the flow rate increase, when it came to their particle size evaluation but not their concentration. This is probably due

to the NTA software not being able to cope with the flow rate increases above a certain level, and the larger the particle size, the lower this threshold gets. The simplest solution would be to go back to a manual push for such particles, but this is rarely a problematic limit if we focus on viruses and VLPs. I also conducted further instrumental evaluations involving examinations on the video acquisition length or the Detection threshold-DT value (Figure 87). This work showed that the video length had no effects whatsoever on the particle concentration. Whereas the DT showed to have a significant effect on the concentration calculation, and one has to be extremely careful in examining several of these DT value for any new sample, in order to make sure that we have the correct compromise between the right concentration evaluation and a low noise level.

Next, I focused on the biochemical optimizations such as those related to using the most adapted buffer for this analysis, which was found to be the conservation buffer PBS-1x (Figure 88). Whereas pure water deteriorated the samples heavily closing particle size increase and aggregation.

Lastly, I evaluated that the Finite Track Length Adjustment (FTLA) optimization used by default in the NTA software, resulted in deteriorating the data analysis for us, as it resulted in generating multimodal distributions, when there were none (Figure 89). We statistically proved that the best fit for our data was the log-normal distribution, and highlighted on the potential implications that this could reveal about these particles (Figure 90). It is important to highlight here that such a distribution has never been explicitly described in the literature when talking about particle size distribution for viruses, as far as we are aware of, although it is a very common fit for size evaluation in other similar fields, especially in the nano and micro particle size analysis.

5.4 Tunable Resistive Pulse Sensing (TRPS)

5.4.1.1 Principle

Tunable Resistive Pulse Sensing (TRPS) is solution based analytical technique used for the characterization of nanoparticles, macromolecules, and particles in a wide range of sizes and concentrations. This technic relies on the measurement of impendence, based on the Coulter principle. Meaning that it relies on the difference in resistivity that a given particle causes when it passes through a nanopore, lowering the baseline current since it partially blocks the nanopore, as in Figure 91.

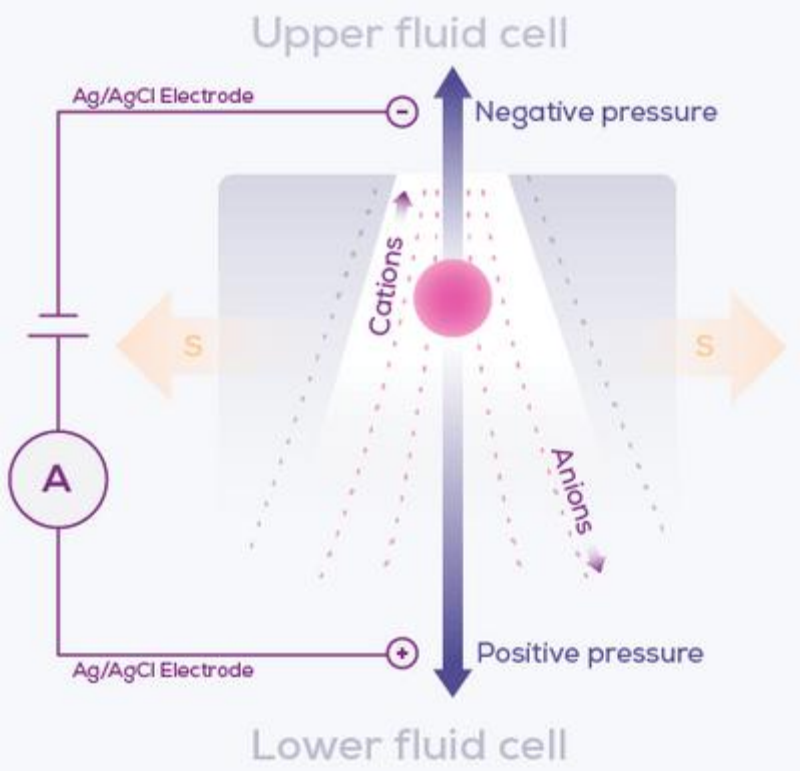


Figure 91. The TRPS principle. The upper fluid cell contains the sample and the electrolyte of interest, e.g. PBS, and in the lower fluid cell the exact same electrolyte of interest is used. Voltage is applied across the nanopore, creating an electrical field. In orange the S (Stretch), represent the force applied on the nanopore, in order to induce the necessary deformation to make possible the aperture for the sample to pass through. A difference in pressure is applied, when performing concentration measurements. Figure was taken from here, © IZON.

5.4.1.2 Concentration evaluation

In order to perform the concentration measurement, as seen in Figure 92, a difference in pressure is applied, ranging from 0-2000 Pa, when performing concentration measurements. This has a similar usage to a syringe pump, in order to push the sample through this pore. And the variation on the pressure, allowing three different pressures, during the same experiment, with the same sample is done to have a sort of in-built triplicate measurement performed on the sample. The measurement of the concentration is done by comparing to a solution of calibrated nanoparticles, with pre-determined size and concentration. This solution is submitted to the same identical conditions as the sample of interest in order to reach the proper calibration.

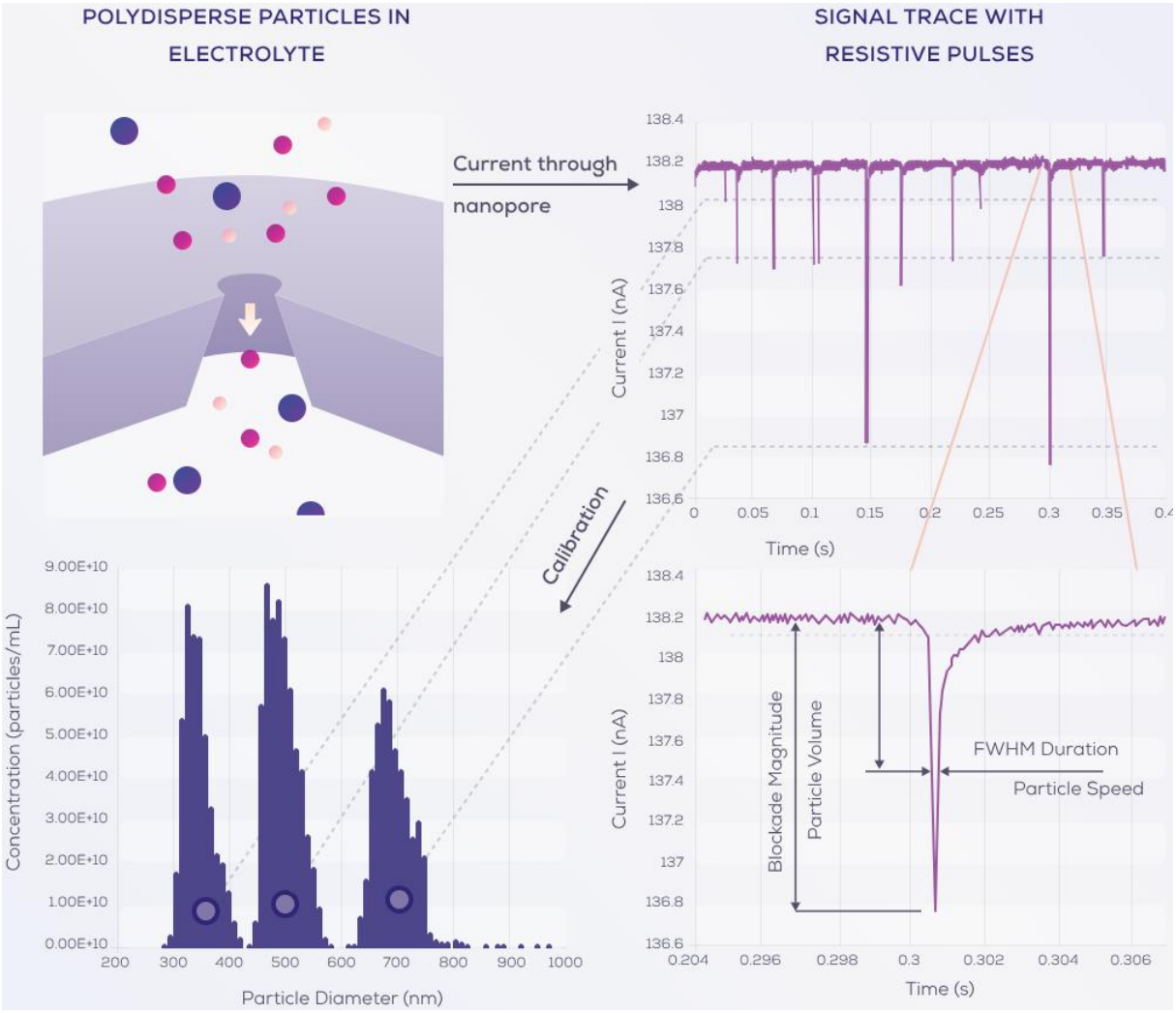


Figure 92. Key features of tuneable resistive pulse sensing (TRPS) showing nanoparticles passing through a nanopore (top left), signal trace with resistive pulses (top right), high-resolution particle size distribution (bottom left), and a breakdown of how each individual blockade is analysed (bottom right). The following image and description were taken from here, © IZON.

5.4.1.3 Zeta-potential evaluation

For the Zeta-potential measurement, the pressure is not applied, and the particles are left to pass the pore without a pressure gradient in order to measure their innate charge. This in term provides information on the particle's stability and surface properties. Thus, the surface charge is generated from the combined net charge of extrinsic moieties on the particle's surface.

I was interested in this method since it could be used as a method to separate particles, not only by the size, but also by the macromolecular content that they may have. I envisioned a similar approach by trying to see if the charge distribution would be different if the particles were decorated, or not, by pb10. The idea is that my work would serve to build a first signature for the different types of particles of T5, separation by production method and decoration. And perhaps in future experiments, the possibility would be that the decoration protein could be added kinetically over a given time range, similarly to what Nicolas Ducrot did in his thesis

using FIDA-Bio.[209] Unfortunately, this proved to be difficult to achieve for our samples, as I will detail below.

5.4.1.3.1 Excluding the Z-potential data

I compared the mean value and the standard deviation between the concentration-based sizing method versus the Zeta potential sizing method, as shown in the following, Table 11.

Table 11. The comparison of concentration-based sizing method versus the Zeta potential sizing methods for the phCLP sample non-decorated (Isis -deco).

sample label	Method	Mean (nm)	Std. Dev. (nm)	RSD (%)
Isis-deco (2000x)	Concentration	93	8.4	9.0
Isis-deco (1000x)	Concentration	94	8.8	9.4
Isis-deco	Zeta-potential	77	38.8	50.4

We can see that the distribution of particle size obtained with the Zeta-potential method is much larger than the distribution of particle size obtained with the concentration-based method. This nearly fourfold discrepancy between standard deviations thus makes the particle size evaluation using Zeta-potential very unreliable. Moreover, the number of counted particles obtained with the Zeta-potential method is insufficient to obtain reliable statistics, compared to the number obtained with the concentration-based methods (e.g. the Zeta-potential method counted 51 particles for Isis-deco, whereas concentration-based methods counted more than 500 particles).

These discrepancies can perhaps be due to the physics that is used to perform these two types of measurements. For example, in the case of the concentration-based method the particles are quantified by having a pressure applied to them in order to cross the pore. Whereas in the case of the Zeta potential measurement, the particles cross the pore only when a current is applied to the solution. One can imagine that a possible deformation of the particles upon pore-crossing is due to the applied pressure. This would make them appear larger in size, due to the fact that a unilateral force is applied on one side of them, thus forcing the geometry of the particles. Whereas in the case of the Zeta potential measurement the dependency on charge is indeed a problem, due to the fact that no neutral or positive particles can be measured, making the Zeta potential measurements biased toward negatively charged particles, therefore, we do not size all particles present. It could be that we only size broken capsids, displaying the negatively charged interior of the capsid, resulting in only the smaller particles being sized.

I decided therefore to exclude the measurements by the Zeta potential method, and only focus on the measurements performed by the concentration-based method, since it had consistent and statistically significant results.

5.4.2 Conclusion on TRPS

My goal in integrating the TRPS method to my analytical pipeline was to have another new analysis in-solution, complementary to my analysis by NTA. Given that both methods allow for size and concentration measurements I wanted to have an examination by both these methods as it was never done before by our team, in addition to the TRPS method never being used on the T5-CLPs particles prior to this work. A major breakthrough that I wanted to implement was the evaluation of the Z-potential, which unfortunately proved to be very difficult to examine properly on these first tests.

I do believe that further analysis by this method on other batches can provide relevant information if properly optimized to our needs, as I was able to do with the other methods explained earlier. This would be of major importance for examining these particles as vaccine candidates, since viral surface charges have shown to be of major importance when it comes to cell infection, and virus-host cell interaction.

5.5 Bottom-up Proteomics, using LC-MS

In order to be able to have a preliminary relative evaluation of the CLP mass, I decided to perform bottom-up proteomics as a structural tool. This is of course directly related to our initial goal of relating the structure to mass and comparing these evaluations to examinations by NEMS-MS in the future.

5.5.1 Principle

Bottom-Up proteomics is an analytical method that can identify, quantify and characterize proteins starting from an initial biological sample of interest. The usual workflow, can be seen in Figure 93, as it was described by Shuken.[42]

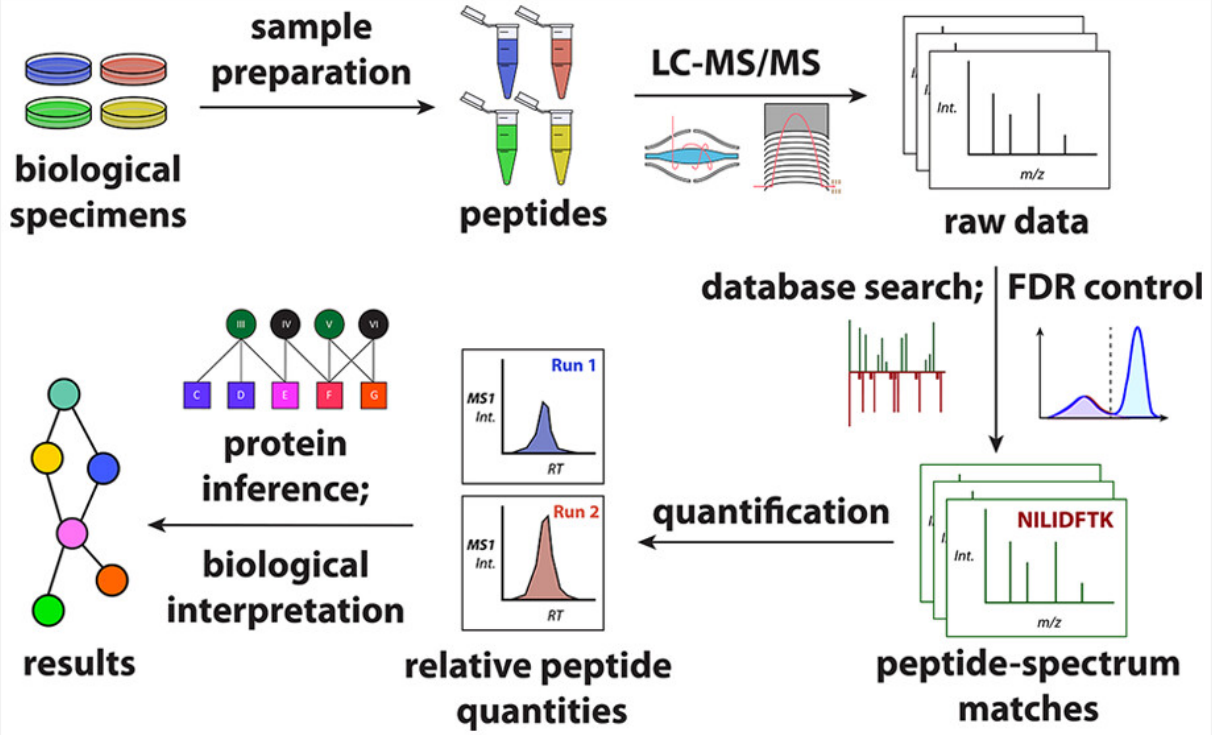


Figure 93. The classical proteomics by Mass Spectrometry workflow. Reprinted from Shuken.[42]

First, the sample is prepared by breaking it down into peptides using enzymatic digestion, usually Trypsin, which prefers cutting at the Lysines (K) of proteins. The sample is then transferred to the mass spectrometer (MS), using a liquid chromatography (LC) column. The LC separates the peptides through a column, giving the dimension of the retention time (RT), and then they are analysed by mass spectrometry (MS), giving the dimension of the intensity of the signal and the m/z (mass to charge ratio). The RT dimension is based on peptide hydrophobicity, meaning the gradient makes the most hydrophobic peptides leave the column last. Peptides of interest can be re-analysed more thoroughly, which gives the annotation of the first Spectra (MS1) and the second Spectra (MS2), also known as MS/MS.

Next the raw data is generated, and it has to be matched to what the existing database is, in order to allow for the peptide spectrum matches (PSM). PSM are defined as the relationship between the intensity and the m/z, and it has to go through the so-called false discovery rate (FDR) control. FDR checks for a certain threshold above which we can accept that the results are statistically significant in order to filter out any false positives or not significant results.[261] We can then proceed to perform peptide quantifications. That is done by analysing several runs and comparing them individually to one another, shown here as the MS1 intensity versus the RT. Here the Spectra are aligned based on the RT dimension, in order to make sure we are analysing the same peptides and thus be able to normalize the data between

the different runs. Lastly, based on our knowledge on the biological sample we can look at the relevant proteins, and make conclusive results about our sample.

The high throughput and high detail of the data is one of the major reasons why MS based techniques have found their place in proteomics, allowing to precisely determine their mass and sequence.

5.5.2 Quantification in Bottom-Up Proteomics

Given that Bottom-Up Proteomics is not an absolute quantification method, one has to be able to quantify the results, comparing and normalising them to a pre-existing known quantity. In the work from Bantscheff, et al.[262], they described the main workflows by which MS proteomic quantification is achieved, which can be seen in Figure 94. We can separate them into the ones that are dependent on labelling or label-free methods, as we will describe below.

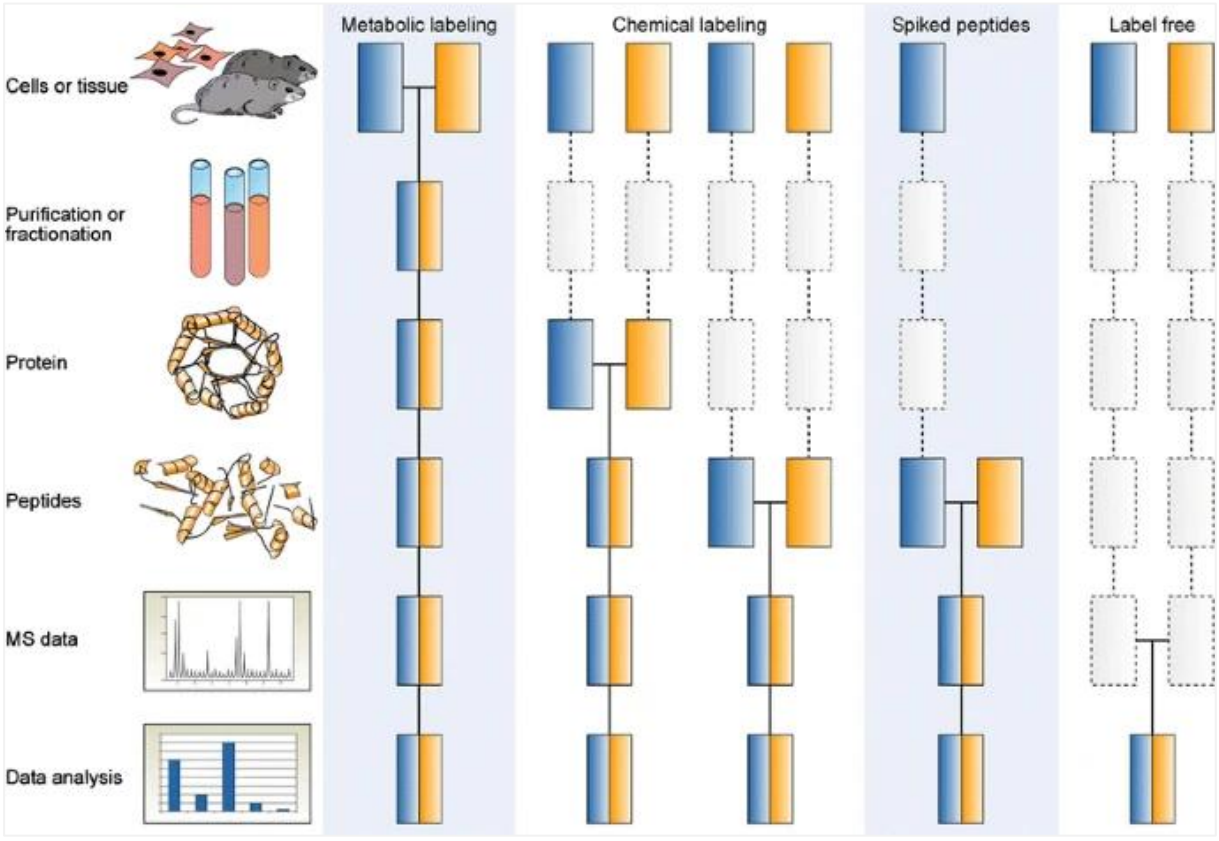


Figure 94. Common quantitative mass spectrometry workflows. Boxes in blue and yellow represent two experimental conditions. Horizontal lines indicate when samples are combined. Dashed lines indicate points at which experimental variation and thus quantification errors can occur. Figure and description were reprinted from Bantscheff, et al.[262]

This separation is based on the 'stable isotope dilution theory', which states that: "a stable isotope-labelled peptide is chemically identical to its native counterpart, and therefore the two peptides also behave identically during chromatographic and/or mass spectrometric analysis. Given that a mass spectrometer can recognize the mass difference between the labelled

and unlabelled forms of a peptide, quantification is achieved by comparing their respective signal intensities."[262] This labelling can occur at different stages in the workflow (Figure 94), that is (i) metabolically, thus at the organism level; (ii) chemically, either at the protein level, or (iii) enzymatically, at the peptide level; or (iv) it can be introduced as spiked peptides, in the form of pre-labelled synthetic peptides. Lastly (v), is the label-free method, which we will detail further in the next paragraph.

The aforementioned labelling comes in the form of stable isotopes, usually those of ^{14}N ^{15}N, ^{13}C, ^{18}O. Which allows for the separation into 'absolute' or 'relative' quantitative proteomics, as in Figure 95. Absolute quantification determines the precise copy number of a specific protein in a sample, using purified protein standards or synthetic peptides for calibration. Relative quantification methods compare the abundance of proteins between different samples or experimental conditions, often relative to a reference sample or control. It provides semi-quantitative data, usually expressed as fold changes, ratios, or percentage differences. One can thus imagine that the absolute quantification method is much more resource-intensive and most suitable for targeted analysis of specific proteins or biomarkers, whereas relative quantification is mostly used in comparative proteomics studies to identify differentially expressed proteins and decode biological pathways.

5.5.2.1.1 Label-free quantification in proteomics

The label-free methods aim at comparing two or more experiments by: (1) using the number of acquired spectra matching to a peptide or a protein as an indicator for their respective amounts in a given sample, or (2) comparing the direct mass spectrometric signal intensity for any given peptide.[262] This can be visualized as 'peptide counting' (for 1) and 'ion intensity' (for 2) in Figure 95.

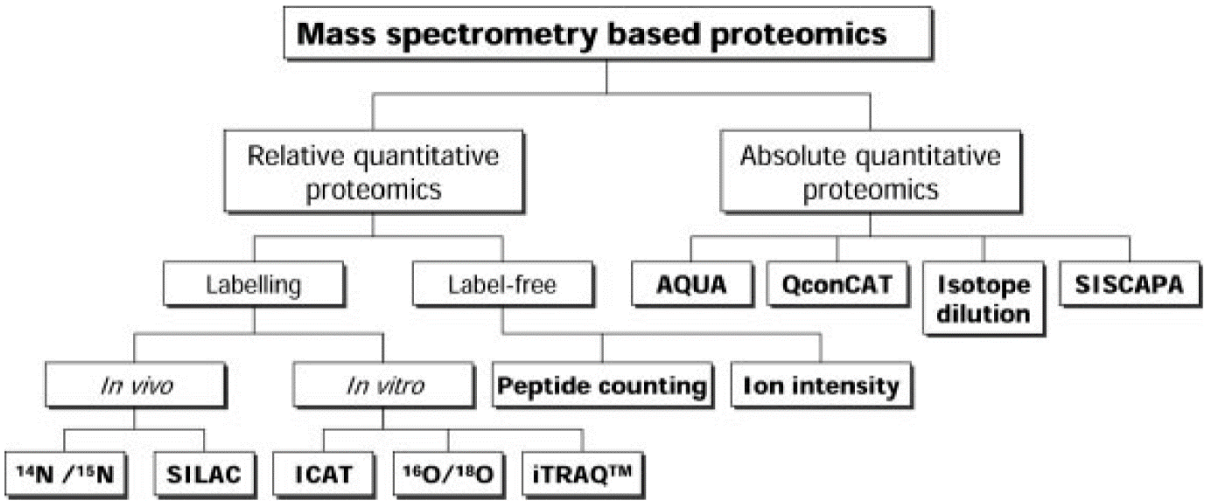


Figure 95. Hierarchy of methods developed for MS-based quantitative proteomics. Abbreviations (in alphabetical order): AQUA: protein Absolute QUAntification; ICAT: isotope-coded affinity tag; iTRAQ: isobaric tags for relative and absolute quantitation; MRM: multiple reaction monitoring; QconCAT: absolute protein quantification uses labeled quantification concatemer; SILAC: stable isotope labelling with amino acids in cell culture; SISCAPA: Stable Isotope Standards and Capture by Anti-Peptide Antibodies. Figure was taken from Lau, et al.[263]

The thesis work of Kavya Clement [202], helped the NEMS-MS group to put in place a method for validation the copy number for the protease protein (pb11) in T5 samples. In Chapter 3 of her thesis, two main methods were compared, the 'absolute quantification method', using AQUA (protein Absolute QUAntification) peptides, and the 'semi-quantitative method', normalizing the known copy count of the MCP (Major Capsid Protein), {Uniport accession number: CAPSD_BPT5}. The copy count for the MCP is equal to 775, due to the conformation of the capsid shells, related to its T-number, in this case T=13. This normalisation method was inspired by the previously described 'proteomic ruler' approach, used for eukaryotic cells, described by Wisniewski et al.[264] In their work, the MS signal for a given protein was normalized to the MS signal of histones, which are directly related to the amount of DNA in the cells and thus to the number of analysed cells.

A more in-depth analysis making the comparison between the different normalisation methods for Label-free quantification will be described in the following paragraph.

5.5.2.1.2 Comparison between the different methods for Label-free quantification

Label-free quantification in Bottom-Up Proteomics is highly versatile but comes with challenges. This can be in the form of variations in sample composition, instrument sensitivity, and data analysis complexity. Normalization, quality control, and robust statistical analysis are crucial to obtaining accurate and reliable quantitative results. In order to choose the quantification method that best suits our specific experimental design and goals, we need to balance several factors such as sensitivity, accuracy, and coverage of quantified proteins.

Although there may be "Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics"[265], the following is a non-exhaustive summary of some of the main methods to achieve label-free quantification in proteomics, and how they compare to one another.

5.5.2.1.2.1 Spectral counting *versus* Intensity-Based Absolute Quantification (iBAQ)

Spectral counting is one of the simplest label-free quantification methods, as it involves counting the number of MS/MS spectra, attributed to each protein in a sample. The underlying hypothesis is that the proteins that are most abundant in the sample will have the higher number of identified spectra.

One of the most used label-free peak intensities based quantification method is Intensity-Based Absolute Quantification (iBAQ). It normalises the protein intensity by the number of theoretically observable peptides for each protein. The underling hypothesis being that higher iBAQ values are considered more abundant.[266]

In the work of Arike et al. (2012) [267] they compared iBAQ (in blue) to some commonly used Spectral counting methods, such as absolute protein expression (APEX, in green) and exponentially modified protein abundance index (emPAI, in yellow), as it can be seen in Figure 96, where two replicates are compared to one another.

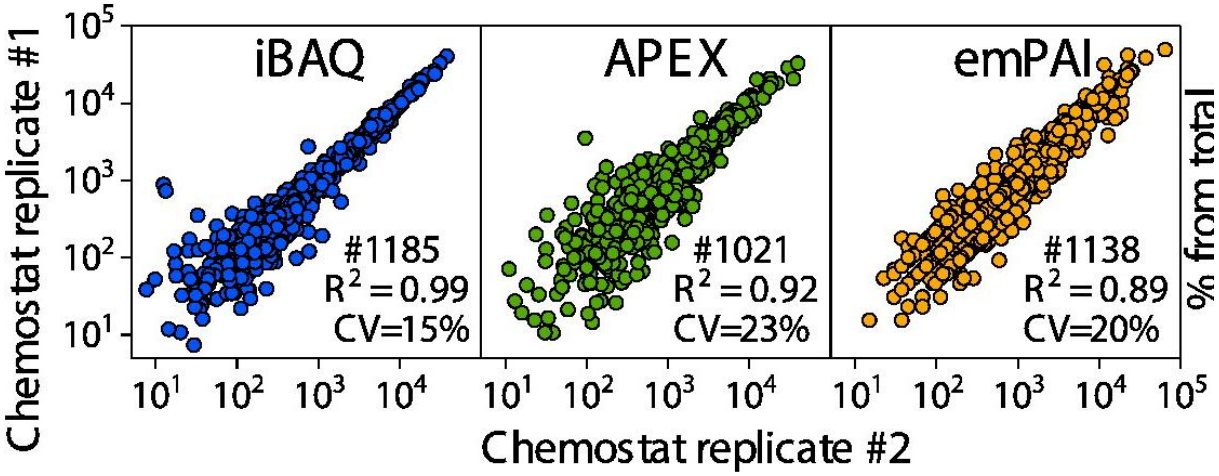


Figure 96. Correlation of replicate chemostat experiments. All data are in logarithmic scale, # — number of quantified proteins; R^2 — squared Pearson product moment correlation; CV — correlation of variation in percentage. Figure and description were taken from the original, Arike et al. [267] The abbreviations are as follows: Intensity-Based Absolute Quantification (iBAQ); absolute protein expression (APEX); exponentially modified protein abundance index (emPAI).

In Figure 96, it can be seen that quantification based on peak intensity, represented by iBAQ, was superior to the spectral counting methods, such as APEX and emPAI. This can be seen by the fact that both the R^2, representing the squared Pearson product moment correlation, and the CV (correlation of variation) were significantly better for iBAQ to the spectral counting

methods, respectively R^2 = 0.99 (iBAQ) versus 0.92 (APEX) and 0.89 (emPAI); and CV = 15 % (iBAQ) versus 23 % (APEX) and 20 % (emPAI).

This justifies the reasoning that peak intensity-based quantification methods are better than spectral counting ones. A known limitation of spectral counting is that it is not necessarily suitable for quantifying low-abundance proteins, and thus it does not provide precise protein concentration values. [267]

5.5.2.1.2.2 Comparing intensity-based quantification methods - iBAQ *versus* Top-3 *versus* LFQ (TPA)

In the work of Al Shweiki et al. (2017)[268], the authors evaluated and quantified the differences and variability between biological difference, such as inter-sample variability, and technical variability, such the one coming from sample preparation and handling and the mass spectrometry analysis. They also quantified the data using three different methods, which were the Top3; LFQ and iBAQ, as seen in Figure 97. The 'Top 3' method, also called the 'average-of-3-most-abundant-peptides', as the name suggest uses the average of the intensity of the three most abundant peptides to normalize the data, [269] contrary to iBAC, it does not need 'identifiable peptides adjustment'. Whereas the Label-free quantification (LFQ) is just another name for the total protein approach (TPA), previously described by Wisniewski et al. [264] by the so-called 'proteomic ruler' approach.

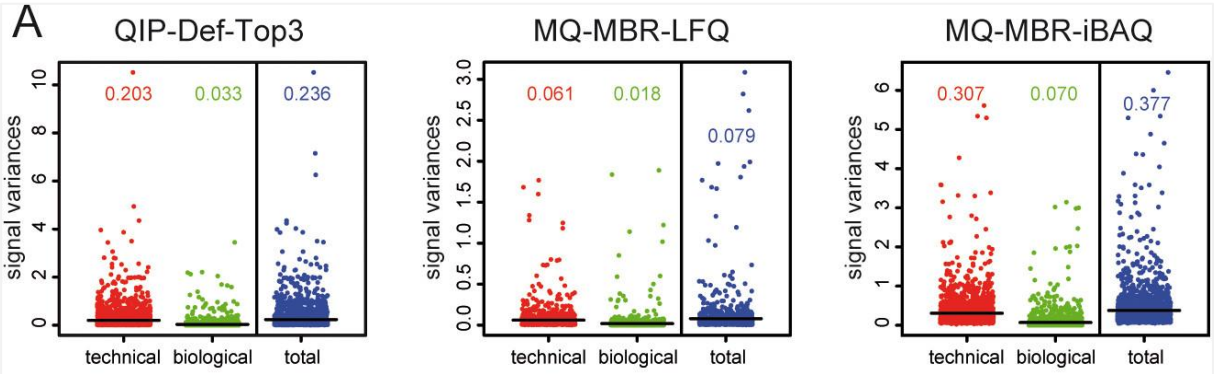


Figure 97. Analysis of the contributions of the shotgun proteomics technology (technical variance, in red) and the plants (biological variance, in green) to the total variance of protein abundance, in blue. Total variance of the abundance of every protein measured in all three repeated analyses of every plant in the nested experimental design and quantified by three software parameter sets (QIP-def-Top3, MQ-MBR-LFQ, and MQ-MBR-iBAQ) was decomposed into technical and biological variance using a linear mixed model (LMM) for each of the software parameter sets. Mean variances for all proteins are indicated by horizontal bars and by numbers. The abbreviations are as follows: Progenesis QIP (QIP); Def: Default; MaxQuant LFQ (MQ); Match Between Runs (MBR); Label-free quantification (LFQ); Intensity-Based Absolute Quantification (iBAQ). Figure and description were taken from the original, Al Shweiki et al. [268]

It can be seen from Figure 97 that the total variance (in blue) for LFQ (middle panel) was the lowest, at 0.079, compared to both Top3 (left panel, at 0.236) and iBAQ (right panel, at 0.377). This total variability is measured as the sum of the biological variance (in green) and the technical variance (in red), per each panel. This study also showed that quote: "the technical

component was considerably higher than the biological inter-sample variability, suggesting an effect on the degree and validity of reported biological changes in protein abundance." [268]

Therefore, one can conclude that iBAQ is the least reliable method when compared to the other commonly used peak intensity-based quantification methods. This justifies our usage of the TPA/LFQ (proteomic ruler approach [264]) as the most adapted tool to reach reliable relative label-free quantification for the rest of our analysis.

5.5.3 Experimental design

In order to design the most adapted proteomics experiments to compare the results for all our samples, we needed to perform a simultaneous analysis that accounted for the different parameters that might interfere in the final result.

For that, we first needed to foresee that the decoration count would be highly affected by decoration protein that was injected into the sample to start with. That is why, in order to remove the excess of the decoration protein, we need to perform dialysis, using a Micro Float-A-Lyzer® Device (Repligen), with a 100 kDa Molecular weight cut-off, to remove any excess of floating proteins. In addition, to have a realistic result, taking into account the potential preparation biases, both decorated and non-decorated samples needed to be subjected to dialysis. The following figure, Figure 98, gives the schematic separation of the samples prepared and analysed.

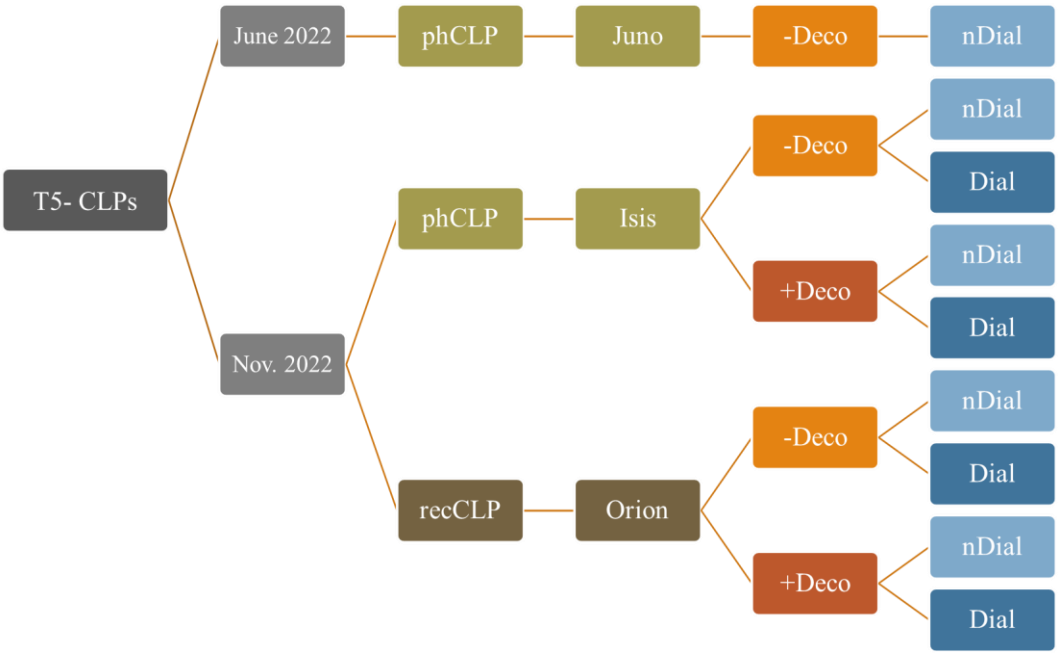


Figure 98. The description of the bottom-up proteomics workflow for T5-CLPs. We have four separation levels (from left to right): 1st: reception date; 2nd : production type, phage-derived (phCLPs; light green) or recombinant (recCLPs; dark green), defining the name of the sample; 3rd: decoration status, present (+Deco; dark orange) or absent (-Deco; light orange); 4th: dialysis status, performed (Dial; dark blue) or not (nDial; light blue).

We thus have four separation levels from left to right. First is the reception date; second is the production type, *i.e.* either phage-derived (phCLPs) or recombinant (recCLPs); third is the decoration status, *i.e.* present (+Deco) or lacking (-Deco); forth is the dialysis status, *i.e.* if performed (Dial) or not (nDial). Juno and Isis_-Deco & nDial serve as biological replicates, coming from two batches produced at different times. The last cell, *i.e.* Orion_+Deco & non-Dial, was unfortunately lost during experimentation, and for which there will be no results shown from here on. For all boxes indicated, we did three technical replicates, therefore totalling 24 samples analysed, since the Orion_+Deco & nDial was not included.

In order to be able to quantitively distinguish between of the aforementioned samples, we need to calculate the total mass of the capsids, which will be detailed in the following paragraph.

5.5.3.1 The total mass Absolute Quantitation method (Abs_{total})

First, ($c\#_j$), is the normalised non-filtered copy count, for a given protein j, calculated as:

$$c\#_j = \frac{I_j}{I_{MCP}} * c\#_{MCP}$$

Where I_j : Recorded Abundance (Intensity value) of a given identified protein j; I_{MCP} is the Abundance (Intensity) of the MCP, {accession: CAPSD_BPT5}; $c\#_{MCP}$, is the expected copy count for the MCP, here = 775.

Next, we filter by removing all protein count $c\#_j < 1$. One can imagine that any value below one copy is primarily due to the sensitivity of the instrument and most likely not because it is actually present in the sample. A similar approach was also recently used in the work of Scheuch et al. [270], where they analysed the prohead protease (gp245) of *Salmonella* myovirus SPN3US (capsid T-number = 27). There they made a clear distinction between quote: "'true virion proteins' (e.g., a protein that is only present in one copy per virion) versus non-virion proteins, proteins that are not a true part of the final virion structure." [270]

Finally, we calculate the total mass of the capsids as follows:

$$\boldsymbol{Abs_{total}} = \sum MW_j * c\#_j$$

Making sure that the MW_j, the molecular weight (mass) of the protein j, is that of the final state, meaning that we are taking into account the cleavage and processing of the protein.

5.5.4 Filtering the data

5.5.4.1 Overall proteomic profile

In the following figure, Figure *99*, we compare the abundance, or the signal intensity of each of the proteins found in our samples, separating them between phage derived capsids (phCLPs), represented by Isis (left), or the recombinant derived capsids (recCLPs), represented by Orion (right). We also separate them by the absence (top) or presence (bottom) of the modified decoration protein, written as pb10Nter-ova in the figure. The interest of this figure is to show the number of proteins that we observe in our samples and have a preliminary comparison between phage / recombinant derived, and when the decoration protein is added.

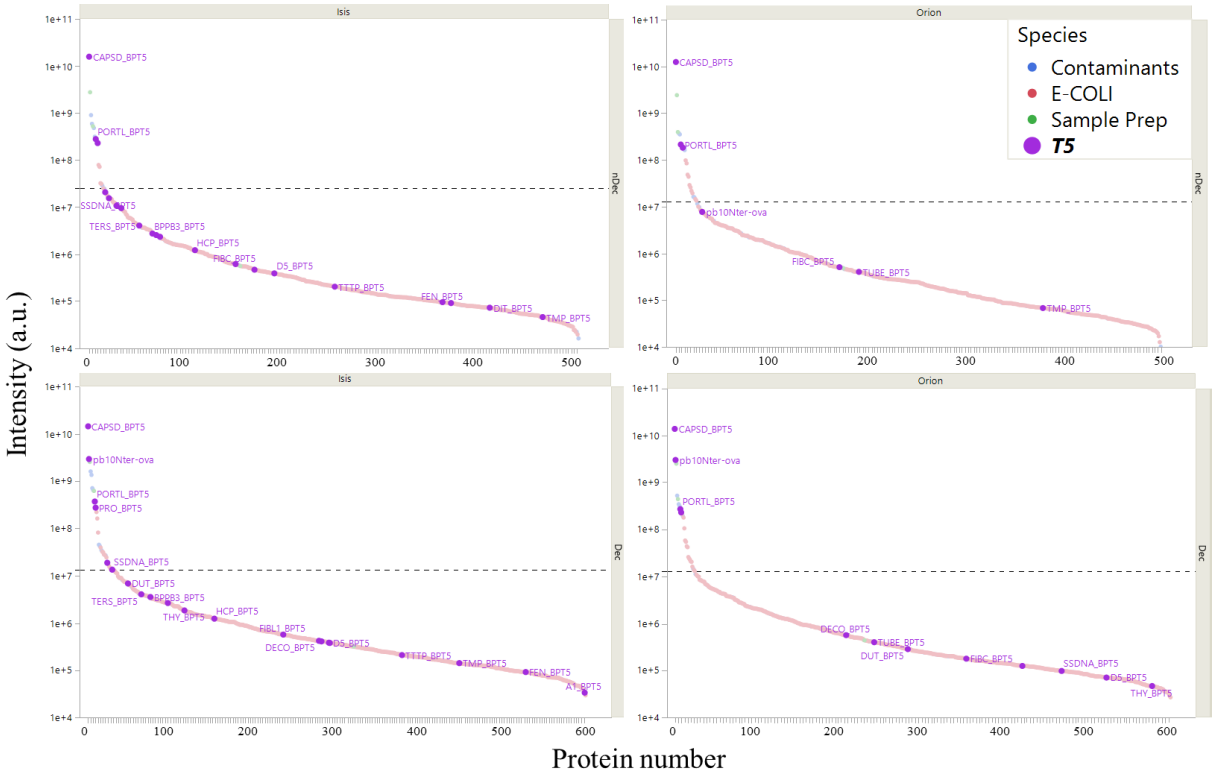


Figure 99. Overall proteomic profile of the intensity, in arbitrary unit, as a function of the proteins numbered, for the different samples, phCLP (Isis; left panels) *versus* recCLP (Orion; right panels), and non-decorated (nDec; top panels) *versus* decorated (Dec; bottom panels). The legend (top right) shows the different proteins species: contaminants, from technical preparation (blue); *E.coli*, from host organism (red); Sample prep, proteins used in the preparation protocol (green); T5, proteins found in the T5-CLPs (violet). The dashed horizontal lines indicated the threshold at which the protein copy count drops < 1, relative to the signal intensity of the Major capsid protein (CAPSD_BPT5) value, always assumed as = 775 copies.

We can see from Figure *99* that the T5 proteins (in violet), are not the majority of the proteins counted, since the host proteins *E. coli* (in red) would be the most numerous in copy count number. It can be also seen that the proteomic representation of the phCLPs is larger than the recCLPs. This is a reassuring result since we know that the phCLPs will naturally have more proteins, and that the recCLPs should only have the proteins that have been designed to be put there. In addition, we can also see that when we put the threshold relative to the Major capsid

protein – MCP (CAPSD_BPT5) value, the signal of most proteins is < 1 protein copy count, indicated by the horizontal black dashed line. This indicates that when we calculate the total sum of the capsid, the majority of proteins will not be accounted for, and will not play a major role in the final result.

As for the decorated samples (bottom panel), we can see that the signal is almost the same for both samples and that it is in a significant amount, close to that of the MCP value. We can also add here that the natural decoration protein, indicated by the nomenclature DECO_BPT5, is only present for the decorated samples and at a similar value for both Isis and Orion. This is most likely due to the signal coming from the artificially introduced decoration protein, making it come out of the noise. It can most likely be attributed to the fact that the artificially produced protein has the same N-terminal portion of the natural decoration protein, explaining the low signal proportionate to the pb10Nter-ova protein.

Therefore, the next step was to look at relevant protein markers that could be useful to indicate how the production method of the CLPs, that is either phCLP or recCLP, affect their biology.

5.5.4.2 Looking for relevant protein markers

The idea here is to check that we have comparable batches between two identically produced samples (i.e. non-decorated Juno and Isis), comparing them to the recombinant sample (Orion), and also compare the effects that the decoration might have on the CLPs. In the following figure, Figure 100, we can see six key proteins, for which when we combine the information collected from their signal; we can actually make relevant biological conclusions, that indicate how we can use bottom-up proteomics in order to reach a better structural understanding of our capsids of interest.

The sample analysed are divided into phage-derived capsids (phCLPs), represented by Juno and Isis, or the recombinant derived capsids (recCLPs), represented by Orion. We separate them also by their decoration status that is non-decorated (nDec) or decorated (Dec).



Figure 100. Key marker proteins, shown as a function of the intensity to the sample name and decoration status. Legend (label on top of each panel): CAPSD_BPT5: Major capsid protein (blue); PRO_BPT5: Prohead protease (orange); PORTL_BPT5: Portal protein (violet); *pb10Nter-ova:* modified decoration protein (green)*;* TUBE_BPT5 : Tail tube protein pb6 (aquamarine); HCP_BPT5: Head completion protein (red).

First, we can estimate the stability of the measurement by observing the signal coming from the major capsid protein (CAPSD_BPT5; in blue) and the Portal protein (PORTL_BPT5; in violet). We can see here that the signal is stable for all five samples, which indicates that the measurement has been performed in comparable quantities between samples, thus no major bias due to the sample preparation. Next, we see that the relative quantification of the Prohead protease (PRO_BPT5; in orange) is comparable to the Portal protein. This last is expected to be at 12 copies for all CLPs, indicating that the Prohead protease protein is in a similar count, being between 12 and 16 copies. This is an important result as it matches in perfect agreement to previous results observed by our group, using bottom up proteomics, already shown in the thesis work of Kavya Clement.[202] We can also see that this protease is present in all capsids, independent of them being produced by phage or recombinant, indicating its importance as a potentially important structural protein, even after the capsid has been fully expanded.

The major structural difference between phCLP to recCLP, can be observed by looking at the Tail tube protein pb6 (TUBE_BPT5; in aquamarine) and the Head completion protein (HCP_BPT5; in red). The Tube protein (pb6) polymerizes to form the tail tube of the phage, and is involved in viral genome delivery during ejection.[213,214] It is reported to be the main component of the tail tube, which is made of 40 stacked pb6 trimers.[214,212] As for the HCP, it closes the capsid once the viral DNA has been packaged.[212] We can see from these proteins that

the signal is much higher for the phCLP samples compared to the recCLP samples. This is explicitly seen by the complete nonexistence of any signal when looking at the HCP protein for the recCLP samples. Interestingly enough, the tube protein is not exactly behaving the same as the HCP protein. We can see that the signal going from Juno to Isis is almost an order of magnitude lower, respectively 1e8 for Juno and 1e7 for Isis. Still significantly higher compared to the values for Orion, that is around 5e5. This indicates that the phage presence can be relatively quantified by the presence of the tube protein, and the fact that for Juno this number is nearly 10 times higher than Isis, it is a good indicator of phage contamination. This observation is complementary to what was seen during the TEM analysis, as we showed earlier (Figure 68). Lastly, the decoration of the particles is clearly visible and quantifiable, when observing the modified decoration protein signal (pb10Nter-ova; in green). It can be seen that the signal increases by slightly more than two orders of magnitude, going from 1e7 to above 1e9. The fact that we still observe a high value of intensity for the non-decorated captions can be due to the fact that if there are natural phages present, they create a high baseline for the decoration signal. Signal, which can be detected by the fact that we are perhaps observing peptides from the N-terminal region of the decoration protein.

5.5.5 Conclusion on Proteomics

In this chapter I described how we decided to use relative quantification in bottom-up proteomics to reach total mass evaluation of T5-CLPs and explain how the absolute intensity quantification method, based on the 'proteomic ruler approach', is the most adapted technique to reach this goal. Then I described how we can perform this calculation in order to reach the total mass of these particles. Later I described how the overall proteomic profile of T5 is presented in our sample, and what type of filters we need to apply in order to or reach conclusive results, highlighting the necessary proteomic biomarkers for future analysis.

The proteomics pipeline described in Figure 98 was designed to reply to the following questions:

1. How close or different are the preparations of T5-CLPs, when they are produced from a phage infection (phCLPs) or by recombinant (recCLPs) ?
2. Are the samples decorated in the appropriate/expected copy number ?
3. Is our data consistent with previous analysis on the same type of samples, i.e. for phCLPs?
    - Primarily concerning the copy number evaluated for the protease protein (pb11), from the work reported in the thesis of Kavya Clement.[202]

4. How comparable are the different batches of the same sample, prepared at different time intervals, especially regarding their ageing?
5. Is the total mass estimated by the relative abundance close to the theoretically expected one ?
    - This was inspired by previous works, especially Dominguez-Medina et al. (on T5 capsids),[21] and Barnes et al. (on Human Adenoviruses),[78] providing back-of-the-envelope calculation of the mass of the capsids, prior to total virion mass analysis, but without comparing them to empirical data from proteomic analysis.

These questions will drive the reasoning and help the understanding of the results in the following chapter, with regard tp proteomic analysis.

5.6 Optimising nebulisation and measuring in the gas phase

Let us come back to the analytical pipeline design (Figure 62), where I mentioned that particle nebulisation could be generated by two methods, that is ESI or SAWN, and then can be analysed by the NEMS-MS bench, or by an aerosol technique. One of the main goals of our group in the last few years has been to go further than using the classical ESI for native MS approach and try to integrate alternative nebulisation methods that are more adapted to our bench, especially regarding neutral nebulisation mechanisms. That is why one of the research areas that we focused on was to characterize and optimize aerosols produced by SAWN chips, designed and manufactured by our collaborators at the Leibniz Institute for Solid State and Materials Research (IFW Dresden).

The main idea was to check how well the SAWN approach worked, by evaluation the particle and droplet size evaluation for a range of concentrations that were in optimal range for both the Aerodynamic Particle Sizer (APS™) instrument, in the micron range; and the Scanning Mobility Particle Sizer (SMPS™) instrument, in the nanometric range. The experiments in this subproject were primarily conducted by the engineer Louis Dartiguelongue, with whom I closely collaborated, in the goal of optimising the pipeline in order to achieve virus measurements. Below, I will talk more in detail about these methods and their applications.

5.6.1 Methods for aerosol evaluation

5.6.1.1 *Principle of the Aerodynamic Particle Sizer (APS), micron range*

The Aerodynamic Particle Sizer (APS™) is an instrument used for the measurement and characterization of airborne particles and droplets, at the micron scale, manufactured by TSI. It measures size distribution of airborne particles, based uses the principle of inertia, thus resulting in the evaluation of an aerodynamic diameter. In addition to that, it can also measure a discrete droplet concentration, by counting and binning the particles by defined size groups. The following figure, Figure 101, shows the working principle of the method. This description was based on that provided from the page of The University of Manchester.

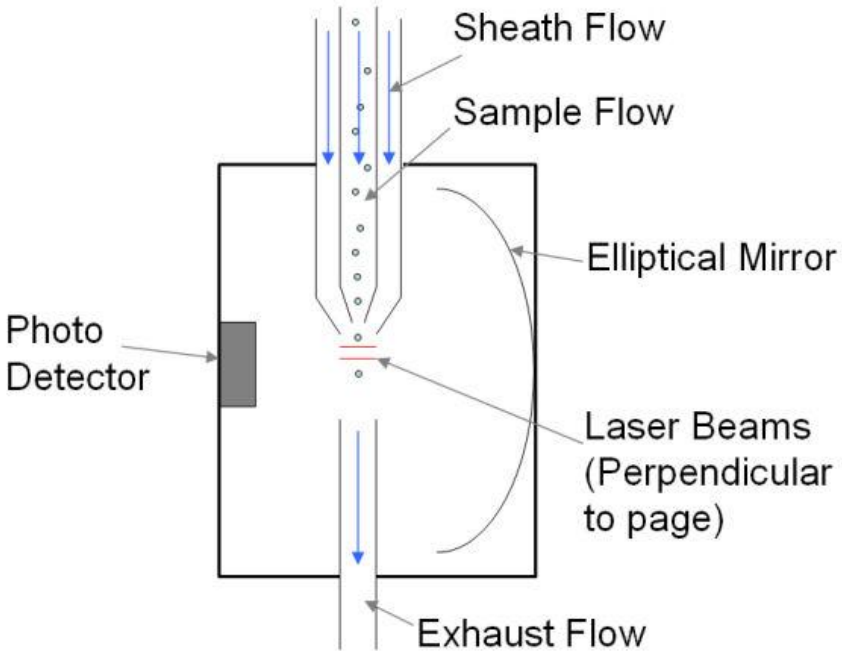


Figure 101. Schematic of the Aerodynamic Particle Sizer - APS (model 3321) detector block, manufactured by TSI. The main components of the instrument are labelled and indicated by arrows. ***Credit****: Figure taken from The University of Manchester.*

As shown in Figure 101, the sample is introduced inside a measurement chamber thanks to an airflow, in parallel to a sheath flow, as shown on the top of the figure. Then they are both constricted through a nozzle, which accelerated the airflow, resulting in a better separation of the particles from one another. This happens since the particles of different sizes have different inertia, assuming that they have the same unit density. This of course is dependent on the surface area that these particles present to the airflow pushing them, resulting on the particles of larger mass/size having a slower inertia, when smaller particles are faster. This can thus be concluded by making the relationship between particle size to their velocity exciting the nozzle, for a

known and constant airflow, resulting in the evaluation of their aerodynamic diameter, by assuming particle sphericity.

In order to be able to measure the precise particle velocity, the instrument makes use of the elements shown in the measurement chamber, in Figure 101. Once the particles are ejected from the nozzle, they pass through two parallel laser beams, about 200 µm apart from one another, producing two pulses of scattered light, and are then ejected out of the chamber in the exhaust flow. The scattered light is reflected from the elliptical mirror (right of Figure 101) onto a photodetector (left of Figure 101). For the particle to be valid, it has to obligatorily pass through both beams, as this allows calculating its velocity, and thus the particle's aerodynamic diameter, by measuring the time relapsed between the two pulses appearing. An additional particle size evaluation is possible with this method, permitted by the evaluation of the amplitude of the peaks from the baseline, generated by the optical scattering, thus named as an optical sizer. The size ranges suggested by the manufacturer are about 0.5-20 µm for the aerodynamic sizing, and about 0.3-0.5 µm for the optical sizing. The resolution is defined by the 52 bins for aerodynamic sizing, and an additional bin for the optical sizer. Measurement frequency is at the order of 10 Hz for a single size distribution.

No let us see how the Scanning Mobility Particle Sizer (SMPS™) can perform particle measurements at the nanometric range.

5.6.1.2 Principle of the Scanning Mobility Particle Sizer (SMPS), nanometre range

The Scanning Mobility Particle Sizer (SMPS™) is a commercial instrument produced by TSI. [link] It is also known by a myriad of other names, as pointed out in the work of Steinberger et al. [271] such as nano electrospray gas-phase electrophoretic mobility molecular analysis (nES GEMMA); nES Differential Mobility Analyser (nES DMA), LiquiScan ES, Macro Ion Mobility Spectrometer (MacroIMS). This method has the capacity to measure particle size in aerosols, relying on their electrical mobility, using a Differential Mobility Analysers (DMA), and their concentration by a Condensation Particle Counters (CPC). A schematic presentation and a more detailed description of these components can be seen in Figure 102.

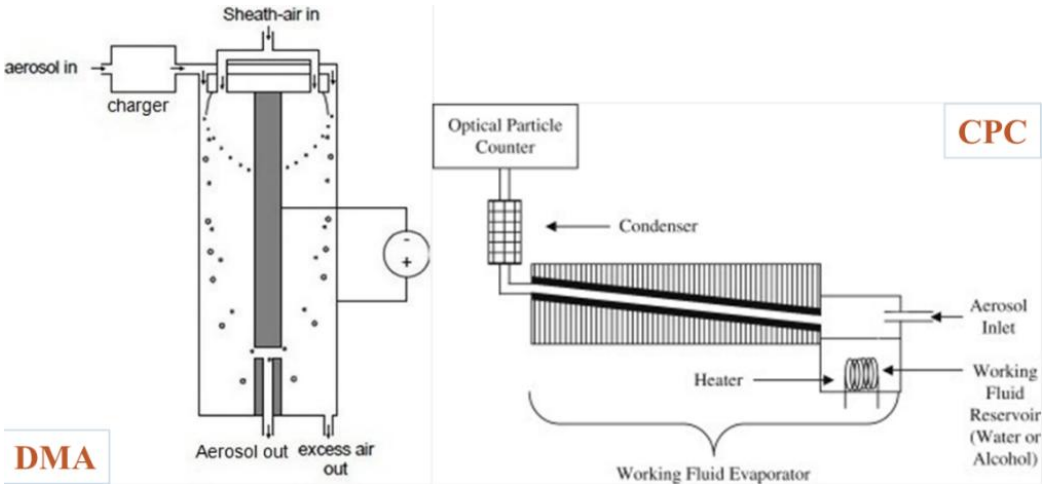


*Figure 102. Schematic illustration of Differential Mobility Analyser (DMA) (**left**) and Condensation Particle Counter (CPC) (**right**).* {**Left**} *DMA. The aerosol entering the instrument is charged and particles subjected to an electric field. A gap in the bottom cylinder collects particles of a selected size by varying the voltage on the central electrode. Adapted from [272], and [273]. The aerosol exiting the DMA (aerosol out) is directed to a CPC (aerosol inlet).* {**Right**} *A working fluid condenses on the aerosol, and particles are detected by optical means. Figure from [274].*

The aerosol entering the DMA is charged, usually by a radioactive source, to have a Boltzmann distribution of charges on the particle's surface. From there, the aerosol, due to the effect of a sheath of air is guided through a vertical column that has a defined length and applies a well-known voltage difference. This difference permits the separation of particles by charge, and evidently by size as they are correlated by the particles surface area. It allows for a separation and analysis of aerosols in the range of few nanometres to nearly half a micron.[273] The particles then go through a slit, which lets the aerosol back out of the DMA, seen on the bottom of the left figure, in Figure 102.

From here, the SMPS combines the DMA to the CPC, as the released aerosol is inserted at the entry of the later, seen on the right side, of the right figure, in Figure 102. Here, the idea is to introduce the aerosolised particles in a chamber containing a supersaturated vapour, in our case it was butanol and not water, as shown in Figure 102. Then by applying a difference in temperature, this vapour will condense around the aerosol, increasing their diameter to about 10 µm.[273] Taken that now the particles have become optically visible by a laser source, we can estimate their quantity, by counting the number of times they cross the light source. The LOD (Limit of detection) of the lowest size detectable here is typically around 5-10 nm.[273] This is of course dependent on the residence time that the particle spends inside the condenser tube, in order to make it visible by the laser.

It has also the advantage of being a real-time method, allowing immediate examination of the sample of interest, and combination to other methods at the same time. Other research

groups have also used nano-electrospray differential mobility analyser (nES-DMA), in measuring the concentration, and even the mass of viruses and VLPs.[144,275]

5.6.1.3 Side note about the unit: dN/dlogDp

One important thing to be remembered when we look at the data from TSI particle sizers, such as Aerodynamic Particle Sizer™ (APS™) or Scanning Mobility Particle Sizer™ (SMPS™), is their concentration is usually plotted in the unit (dN/dlogDp). This is because these instruments have equally spaced size channels, on the log scale. Therefore, we have 64 channels per decade resolution, i.e. dN/dlogDp = 1/64, for SMPS, and 32 channels per decade resolution, i.e. dN/dlogDp = 1/32, for APS, as described in the Application Note PR-001 by TSI.[276]

5.6.2 Nebulisation optimisation

The idea is to see what the most optimal SAWN generated aerosol would be, in order to best replicate a virus aerosol, prior to the mass evaluation by the NEMS-MS. The sample of choice was a PolyStyrene Latex Nanoparticles (PSL-NPs), with an expected median size of 100 nm, and an expected solution concentration of ~ 1e14 NP/mL. This choice was made as this sample has the particularity of being of an appropriate size comparable to many of the HuReVs, and have a comparable density to viruses (1041-1063 kg m^{-3})[277]. In addition, the high concentration makes it possible to perform a large number of experiments, with little sample discarding.

5.6.2.1 Aerodynamic Particle Sizer (APS™) measurement

The first experiment consisted of evaluating the droplet containing the PSL-NPs diluted in water, for 5 different concentrations, from 1e8 to 1e12 NP/mL. Due to technical difficulties, the 1e9 NP/mL was not possible. Therefore, in Figure 103 we show the Aerodynamic Particle Sizer (APS™) measurement for these preparations. The absolute droplet concentration per droplet size regarding these measurements is presented in supplementary Figure 143.

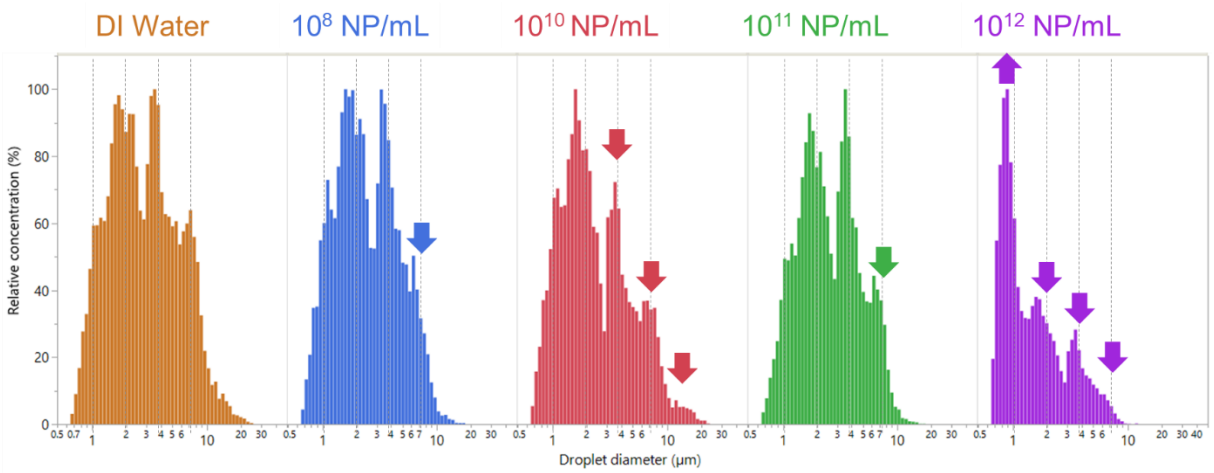


Figure 103. APS measurement of PSL-NPs in different concentrations, generated by SAWN. The figure shows the relative concentration (in %)(normalized to maxima), as a function of droplet aerodynamic diameter (in µm), shown in the range 0.5-40 µm, for 5 different solutions. Legend (values found on top of the graphs): deionised (DI) water: orange; the different PSL-NPs concentrations values (in solution) are presented in increasing order, from left to right: 1e8 NP/mL (blue), 1e10 NP/mL (red), 1e11 NP/mL (green), 1e12 NP/mL (purple). The vertical dashed lines serve as references for size, respectively at 1, 2, 3.5 and 7 µm. The arrows serve to draw attention to the main peaks of interest, detailed further in the main text. The distributions were normalised by the maximum, for an average between n=2 runs.

It can be seen in Figure 103 that for DI water (negative control), in orange, we already have the presence of several peaks, more precisely at 1 µm, 2 µm, 3.5 µm and 7 µm (indicated by the vertical dashed line at these values). The distribution in blue (1e8 NP/mL) and in green (1e11 NP/mL) are very similar to that of DI water. When looking at the intermediate distribution, at 1e10 NP/mL (red), the behaviour is very different. We can see that there is an additional peak at around 15 µm (arrow), which was absent before. This changes entirely when we look at the 1e12 NP/mL (purple) distribution. Here we see that all the previously pronounced peaks are now almost entirely absent, whereas the peak at 1 µm (indicated by the upward arrow) is much more pronounced and dominates the distribution.

This most likely indicates that the NP are largely present at this droplet size, getting out of the noise of the water, or that their concentration in these amounts allows for better nebulisation by SAWN of this range of droplets.

But this was of course at the droplet, micron scale, and in order to observe the particles themselves, we need to go at the nanometre range, possible thanks to Scanning Mobility Particle Sizer (SMPS™).

5.6.2.2 Scanning Mobility Particle Sizer (SMPS™) measurement

In order to observe the PSL-NPs directly in the gas phase, we performed the same dilution experiments to those shown in Figure 103. Therefore, in Figure 104 we show the SMPS measurement for the same preparations.

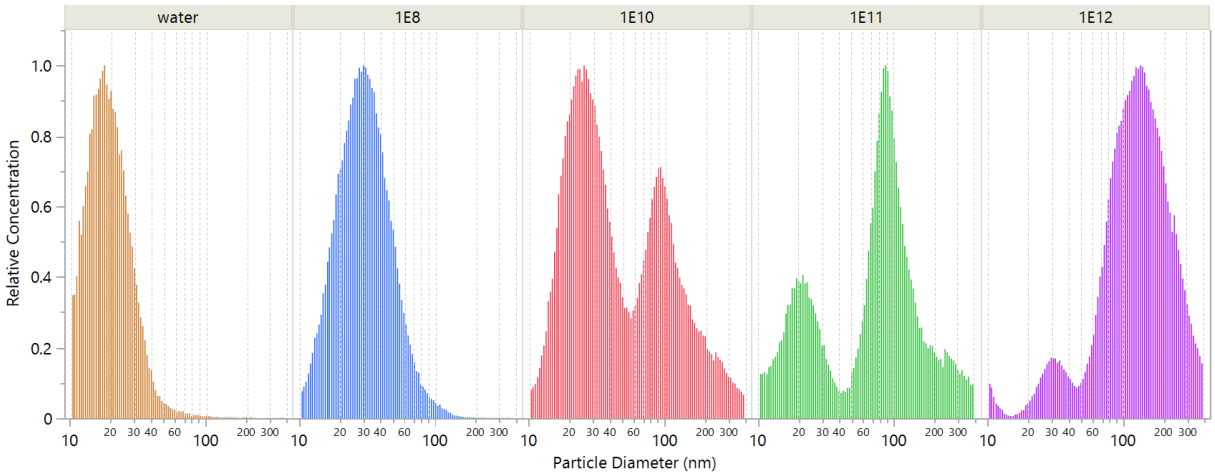


Figure 104. SMPS measurement of PSL-NPs in different concentrations. Relative concentration in aerosol (normalized to maxima) as a function of electrodynamic particle size (nm), in the range 10-400 nm, presented in logarithmic scale. Legend, coloured as for Figure 103, (values found on top of the graphs): deionised (DI) water: orange; the different PSL-NPs concentrations values (in solution) are presented in increasing order, from left to right: 1e8 NP/mL (blue), 1e10 NP/mL (red), 1e11 NP/mL (green), 1e12 NP/mL (purple). The arrows serve to draw attention to the main peaks of interest, i.e. left: for the water peak; right: 100 nm for the expected PSL-NP peak. The vertical dashed lines serve as references for size, representing the logarithmic scale examined.

It can be seen in Figure 104, that for DI water (orange), we have the presence of a major peak, at ~ 18 nm. Similarly, the distribution in blue (1e8 NP/mL) presents a similar trend, with the water peak slightly shifting to the right, at ~30 nm, and a complete absence of the 100 nm expected peak for the PSL-NPs distribution. The 1e10 NP/mL (red) distribution present a very clear intermediate state, where we have the coexistence of two peaks, one at ~25 nm and the other at ~90 nm, very close to the expected value. This may suggest an instability for the PSL-NPs at this dilution, as it was similarly already observed to behave in an unexpected manner in the APS measurements, shown in Figure 103. The green (1e11 NP/mL) distribution shows the clear dominance of the PSL-NPs peak, at ~90 nm, and the reduction of the water peak. Going even further, at the 1e12 NP/mL (purple) distribution, we see that the water peak is largely reduced to the second peak, and we see that the peak has shifted to the right compared to the 1e11 NP/mL (green) distribution, culminating at ~120 nm. This is largely bigger than the 100 nm expected size, suggesting possible particle aggregation.

We believe that the most logical explanation for these particles to have an apparent effective diameter as such is because the particle aggregates appear as a single particle, generated by the sum of their individual volumes. A more detailed explanation for this reasoning can be found in the supplementary Figure 141.

Next, we wanted to relate the two measurements of concentration by the different methods, thus the APS measurements for droplet size (Figure 103) and the SMPS measurements for PSL-NP size (Figure 104).

5.6.2.3 Relating APS to SMPS

In order to relate the two measurements, we did a simple approximation that the concentration of the NPs is equal to the ratio of the Number of NPs measured for the volume of the droplet, this last is directly related to the droplet aerodynamic diameter. The results, presented in supplementary Figure 142, show the theoretical relation between particles per droplet, as a function of the expected droplet aerodynamic diameter.

In order to be able to have the most optimal nebulisation, meaning that: if we are > 1 particle/droplet threshold, we get particle aggregation; if we are < 1 particle/droplet, we get inefficient nebulisation. Thus, we can clearly conclude from this relationship that the critical droplet size for the optimal threshold of 1 particle/droplet, for each concentration, is:

- 1e8 NP/mL = between 20-35 µm;
- 1e10 NP/mL = between 4.5-7 µm;
- 1e11 NP/mL = between 2-3.5 µm;
- 1e12 NP/mL = between 1-1.5 µm.

This approximation allows us to reach the next step, which is the calculation of the number of generated particles for a given droplet size.

5.6.2.4 Particle generation number versus droplet size

In order to establish this relationship, we need to combine the information from the Absolute droplet concentration per droplet size (supplementary Figure 143) and the theoretical relation between particles per droplet (supplementary Figure 142). This allowed us to build the following graph, in Figure 105, comparing the number of generated particles as a function of the droplet diameter.

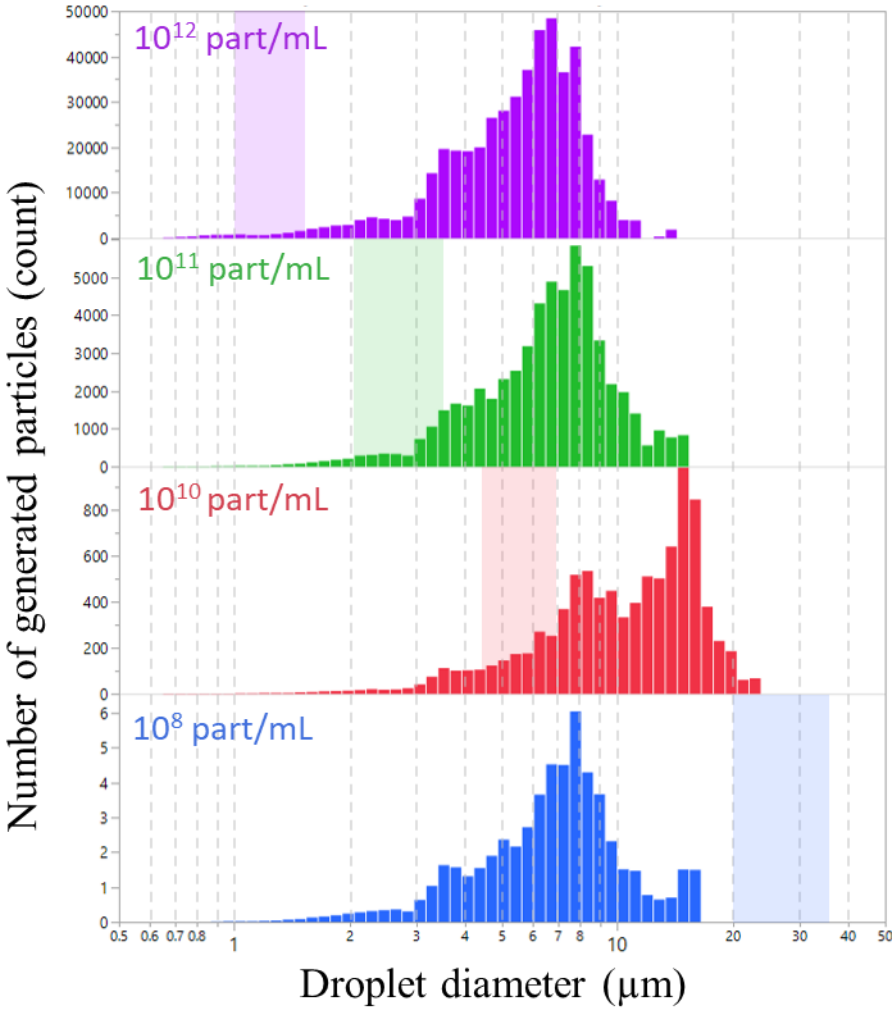


Figure 105. Number of particles generated by SAWN, as a function of droplet aerodynamic diameter (µm). Legend (same as for Figure 103), the different PSL-NPs concentrations values (in solution) are presented in decreasing order, from top to bottom: 1e12 NP/mL (purple), 1e11 NP/mL (green), 1e10 NP/mL (red), 1e8 NP/mL (blue). The highlighted coloured vertical bands represent the range where the 1 particle per droplet (part./drop) is found for each concentration.

In this figure, Figure 105, we can see that as expected the number of particles generated is largely different between the different concentrations and behaves linearly in relation to the expected solution concentrations. In addition, we see that they have a main peak at ~7-10 µm, except for the 1e10 NP/mL (red) distribution, whose main peak is at ~18 µm.

The shaded area in the 1e12 NP/mL (purple) distribution is much smaller than the maximum peak of the distribution, indicating that we are getting a small number of the main bulk of droplets generated, in addition to probably causing aggregation. For the 1e11 NP/mL (green) distribution, we see that the shaded area is closer to the main peak, indicating a more optimal nebulisation. Whereas, for the 1e10 NP/mL (red) distribution, although the shaded area is closer to the centre of the other distributions, it is much further from the main peak, at ~18 µm. In addition, the main peak is shifted towards the 10-20 µm range, which can be linked to the instability of the concentration, as mentioned previously. Whereas for the 1e8 NP/mL (blue)

distribution the shaded area, i.e. the range 20-35 µm, is past the main peak, indicating that we are generating droplets far below the optimal concentration, and in very low numbers.

From these observations, we can see that the most adapted concentration would be between the 1e10 - 1e11 NP/mL, avoiding instability and having the maximum particle per droplet concentration possible.

5.6.2.5 Conclusions

The experimental design described in this section aimed at observing what was the most optimal droplet size range produced by our SAWN set-up and investigated the most optimal particle concentrations in solution prior to nebulisation by SAWN. We concluded that the SAWN generated a main peak the droplet size at ~7-8 µm (see Figure 105), which is compatible to what was reported by our collaborators.[278]

In addition, we were able to determine that the most optimal concentration where the particles were stable and in high enough numbers was in the range 1^e10 - 1^e11 particles/mL. We can see that actually the most compatible region where the two conditions, maximum particle number and the 1 particle/droplet efficiency threshold, are met would be somewhere lower than 1^e10 particles/mL, and closer to 1^e9 particles/mL. However, unfortunately, this concentration seems to be unstable and there is an intermediate range where the particles are still in the water noise detection region. One possible solution for this would be to build a different SAWN with a higher frequency, thus resulting in droplet generation of smaller size, in order to approach to the lower regions, ~1 µm, where higher concentration solutions can be recovered.

5.7 Discussion and Perspectives on Chapter 5

In this chapter, I was able to implement, or help in implementing, a number of methods that were either never previously used by our team, or had not been enough optimised, for application on evaluating virus structures. The following figure (Figure 106) gives a summary of the main point described in this chapter.

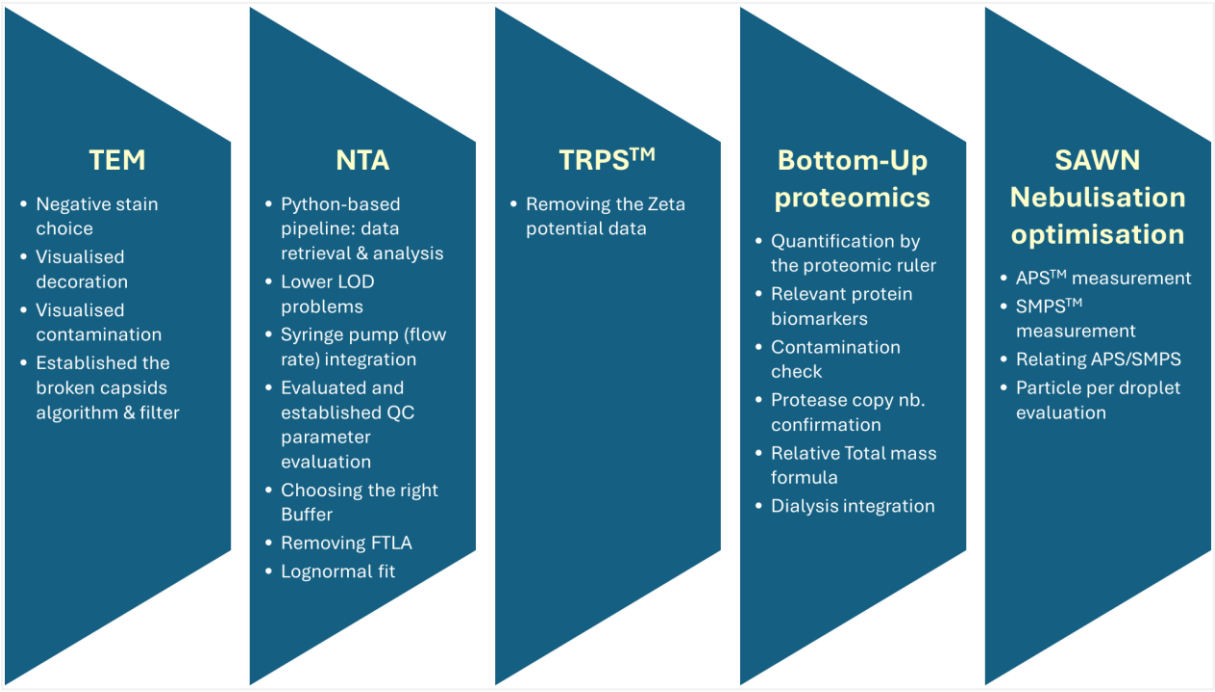


Figure 106. Schematic representation of the structure of Chapter 5.

The classical method in virology has often been transmission electron microscopy (TEM), historically negative staining TEM, and more recently, cryoEM has taken its lead in this field. By using negative staining TEM, and analysing the collected data from the different images, I was able to help in devising and algorithm to select the particles when they present irregular shapes, and that's allow for filtering Broken capsids. I believe that this method can be of an extended use to my current usage, as it allows for a simple analytical approach two selecting particles of interest and could again be even better improved by integrating adapted machine learning software for image analysis, Is this currently the norm for cryoEM for example.

My work on NTA highlighted the need to be very cautious when examining samples with commercial methods. Several parameters that we as users are supposed to take for granted in technical protocols are not always very well understood, and sometimes can prove challenging to optimize. We can now examine for example the lower detection limit of the particles; the effects of the detection threshold and camera level on the measurement; and the

routine integration of the quality control criteria (such as: the particle count per frame; the relationship between centres per frame and particles per frame; and the minimum particles needed to be counted when we perform the measurement). I was also able to establish a robust pipeline for NTA measurement in our laboratory, that can be easily integrated for other viruses in the future. We also optimized the data extraction and data analysis software, by integrating a Python code that can do this work in a very simplified manner. This gave us more liberty and removed some of the obscurity of understanding the NTA measurement and adapting it for our everyday experimental usage. Lastly, this work also allowed observing for the first time that the T5-CLPs present a lognormal distribution with this data. This is also the first time, as far as we are aware, that this has ever been observed for a viral sample, although it is a very common model for describing standard inert nanoparticles.

I also took the initiative to integrate and compare for the first time these samples by the TRPS method, first as another in-solution method and also because it was an independent evaluation to NTA. This proved to be non-conclusive for the Z-potential evaluation and I decided to disregard the data, until further optimisation is performed.

The work performed on SMPS allowed us to optimize the nebulisation by the SAWN method. Allowing for one of the first evaluations of the effects of particle concentration on this nebulisation technique. This allowed us to find the optimal range in which particles need to be nebulised in order to have one particle per droplet content, which is the most optimal range for mass evaluation by the NEMS-MS.

Lastly, on top of all the biophysical methods presented above, I integrated the measurement by bottom-up proteomics in order to have an orthogonal view of the particle mass, prior to the NEMS-MS. This is the first time that this method has been used consistently to reach relative total mass for the T5-CLPs, instead of performing a simple theoretical calculation of these particles. This technique also has others important attributes, such as evaluating the contamination by phages, and verify the protein count reported from the literature, especially confirming what was previously described by our group for protease pb11.[202]

All these techniques were resumed previously in

Table 9, where the information on the possible measurements is given, and how these collaborations were foreseen. The following, Table 12, is a revised version, containing only the techniques and measurements that I was able to optimise and put to use during my thesis, which will be discussed in the following chapter.

Table 12. Table describing the different analytical methods used during this work, the measurements and the analysis performed. Abbreviations (in order of appearance): NTA™ (Nano Tracking Analysis); TRPS™ (Tunable Resistive Pulse Sensing); TEM (Transmission Electron Microscopy); SMPS™ (Scanning Mobility Particle Sizer).

Methods	Milieu	Measurement possible	Service provided
Negative staining TEM	Solution; Deposition	• Particle Size (visual) + algorithm • Particle Morphology	Scientific platform/ Collaboration
NTA™	Solution	• Particle Size (hydrodynamic diameter) • Particle Concentration (in solution)	Owned by our group
TRPS™	Solution	• Particle Size (electrodynamics diameter) • Particle Concentration (in solution)	Commercial service provider
Bottom-Up Proteomics	Solution	• Peptide Abundance (Intensity)	Scientific platform/ Collaboration
SMPS™	Aerosol	• Particle Size (electrodynamics diameter) • Particle Concentration (in aerosol)	Scientific collaboration

In conclusion of this chapter, and in preamble to the next chapter, we can state that the current pipeline is to: start primarily with the negative staining transmission electron microscopy (TEM). Then comparing the data between the nano tracking analysis (NTA) and the tuneable resistive pulse sensing (TRPS) methods, in order to have orthogonal in-solution evaluation. Lastly, the pre-evaluation by bottom-up proteomics, serves as a more empirical comparison to single particle MS, rather to a simply theoretical one.

The following chapter with thus focus on the results and the conclusive structural analysis, describing the T5-CLPs and the main methods described in Chapter 4, that is TEM, NTA, TRPS, and Bottom-up proteomics.

6 Chapter 6: Characterizing T5-CLPs as vaccine candidates

6.1 Introduction of Chapter 6

Vaccines work by priming the immune system to recognize and destroy pathogens or infected cells. They must prompt the presentation of pathogen-specific antigens (Ag) by Ag-presenting cells (APC) thereby promoting specific humoral (mediated by macromolecules, primarily antibodies) and cellular (mediated by phagocytes, and independent of antibodies) responses. Among the diversity of platforms used to develop vaccines, nanoparticle-based vaccines have had a growing attention as a means for targeted Ag delivery [279]. They self-assemble into bio-compatible repetitive structures, triggering potent immune responses, which offers ways to circumvent the need for adjuvants[280]. Currently most Virus-Like Particles (VLP) are derived from human pathogens,[281] where some famous examples include Hepatitis B virus - HBV[282], Human Papilloma Virus - HPV[283], and a variety of Adenovirus derived vaccines,[284–286]. But viruses infecting bacteria, or bacteriophages, have recently been proposed as new ways of potentially safer, more stable, and versatile Ag nano-carriers [287].

The modularity of pb10, combined with the possibility of producing T5-CLPs, has made it possible to use the empty T5 capsid as a nanoparticle, by exposing on its surface chimeric proteins composed of the pb10 capsid-binding domain fused to a peptide or protein of interest. The pNO protein consists of the fusion of the N-terminal domain of pb10 (pN) with ovalbumin (O), a model antigen. This protein was used to decorate empty capsids and obtain CLPs exposing ovalbumin on their surface, which were then used to immunize mice.[288] Immunization of mice with ovalbumin multimerized on the surface of CLPs induces a strong and durable immune response against ovalbumin. This response is obtained without the addition of exogenous adjuvant, thus simplifying the vaccine formulation and lowering the risk of additional potential immune side effects.[289]

To summarize, T5-CLPs present a good model VLP and a great platform for vaccine production. This platform allows for the fabrication of large, genome-free, non-infectious, stable particles, that are easily purifiable, are generated as isolated VLPs, and can be yielded in large quantities. In addition, they are accompanied by a decoration protein that binds quasi-

irreversibly to its surface. Moreover, this last is easily modifiable, enabling peptides, domains or entire proteins of interest to be exposed at the capsid surface, without recourse to bioconjugation or chemical bypass processes. Recent biotechnological developments by the I2BC team allowed for the development of a 'new breed' of viral particles, recombinantly produced (recCLPs). These last were developed to be a simpler, cleaner and faster solution to producing CLPs, and were extensively characterized and compared to the native forms (phCLPs) by Nicolas Ducrot during his thesis.[209] One schematic representation of the possibilities of the decoration of T5-CLPs for vaccine usage, by expressing different antigens in its surface, can be found in Figure 107.

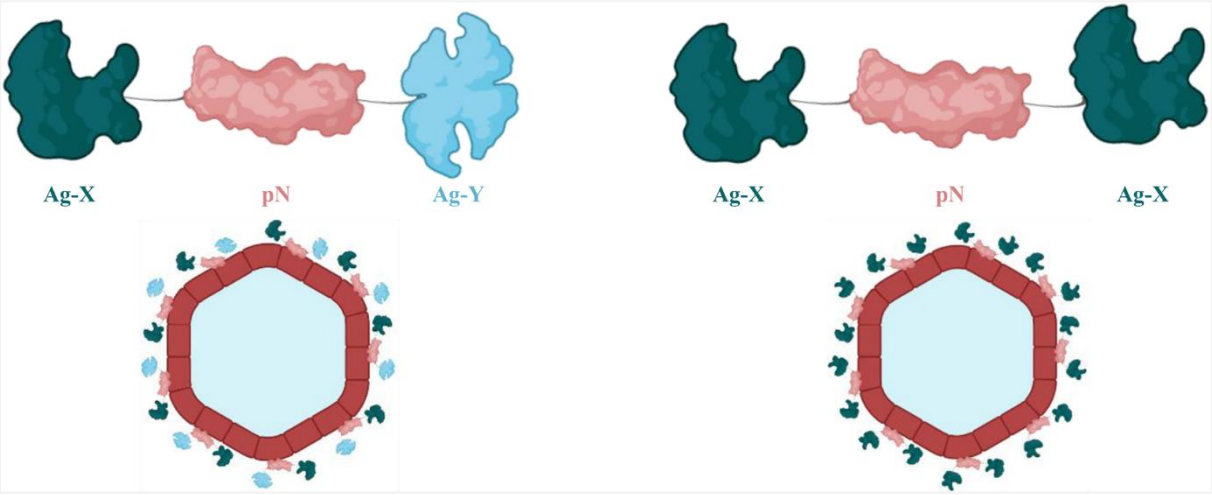


Figure 107. Examples of possibilities for decorating the T5 capsid with fusion proteins containing N- and C-terminal sequences of the pN domain. Left: decoration by 120 copies of Antigen X (Ag-X) and 120 copies of Antigen Y (Ag-Y); Right: decoration by 240 copies of Ag-X. Figure adapted from the thesis of Nicolas Ducrot.[209]

As we saw in the previous chapter, I put in place and optimised a pipeline to perform the most compatible analysis with the NEMS-MS workflow. The following chapter describes the main results from the different analytical tools that we used to examine our samples with, in order to have an improved structural information on these samples to the current state of the art. This research also serves to characterise the recombinant samples with a multitude of new methods, thus helping our collaborators at I2BC in better understanding the biology of these particles.

In the rest of this chapter I will discuss several biological questions that I was able to evaluate with the discussed analytical pipeline, as shown in Figure 108.

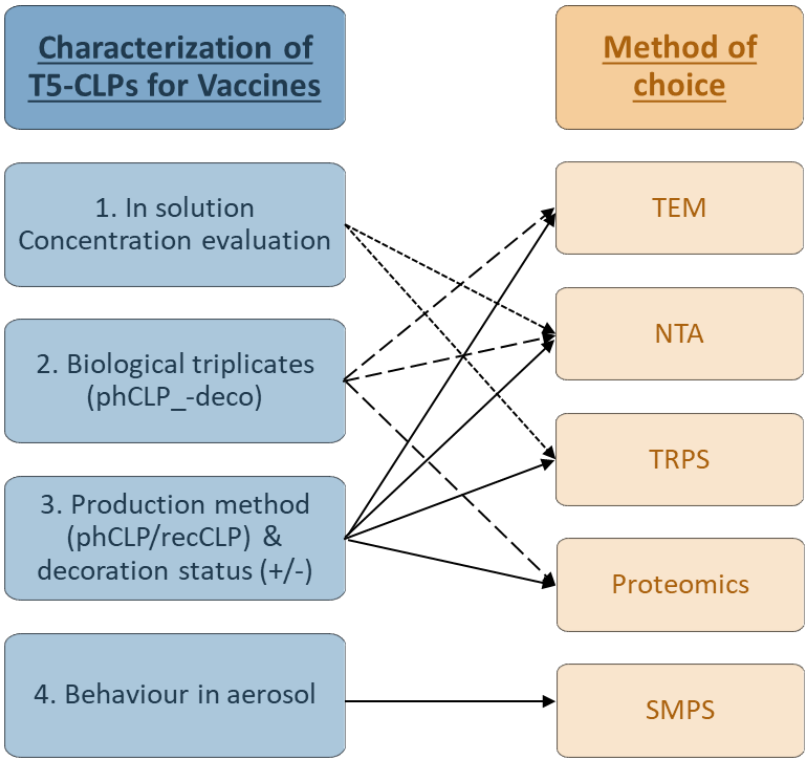


Figure 108. Schematic representation of the structure of chapter 5. Summary of the different examinations evaluated in this chapter (in order of appearance) and the different methods that were used to characterise them.

I will start by discussing the results about concentration evaluation, comparing the data between the nano tracking analysis (NTA) methodology, for which I optimised the workflow in our lab, to the commercial tuneable resistive pulse sensing (TRPS) method, and to the reported data by our collaborators at I2BC, that produced the CLP samples.

Then I will focus on comparing between three samples of non-decorated phCLPs, which has been the model particle for our lab. The primary comparison will rely in analysing samples acquired during different moments of my PhD, in order to evaluate sample aging and trying to identify if the structural confirmation is maintained throughout. This is a very important quality control attributes when producing vaccines, and is in line with the work conducted by our collaborators at I2BC.

Next, I will proceed to compare between two samples produced in the same time, in order to neglect the aging effects, and thus focus on comparing them based on their production methods, i.e. between phCLP and recCLP, and also take into account their decoration status, i.e. between non decorated (-deco) and decorated (+deco) particles.

In addition, apart from the negative staining TEM analysis, also conducted at I2BC[209], this is the first time that the recCLP particles have ever been examined by all the other methods presented here. Furthermore, it has to be highlighted here that the pre-evaluation by bottom-up proteomics, prior to the NEMS-MS evaluation, serves as a 'ground truth' approach to having

empirical observations of viral samples by mass, compared to 'theoretical estimates' approach, which is currently the norm in virus native MS analysis. Lastly, I will close on the perspective of examination of these particles in aerosol, using the scanning mobility particle sizer (SMPS™) spectrometer, which is also the first time that these particles have been examined, after being nebulised by SAWN.

6.2 Concentration evaluation

The first examination performed was the evaluation of the concentration of particle per unit volume by NTA, which I optimized for this usage in our lab, and compare it two orthogonal methods, to the reported data from I2BC, and to a commercial equipment (TRPS), as reported below in Figure 109.

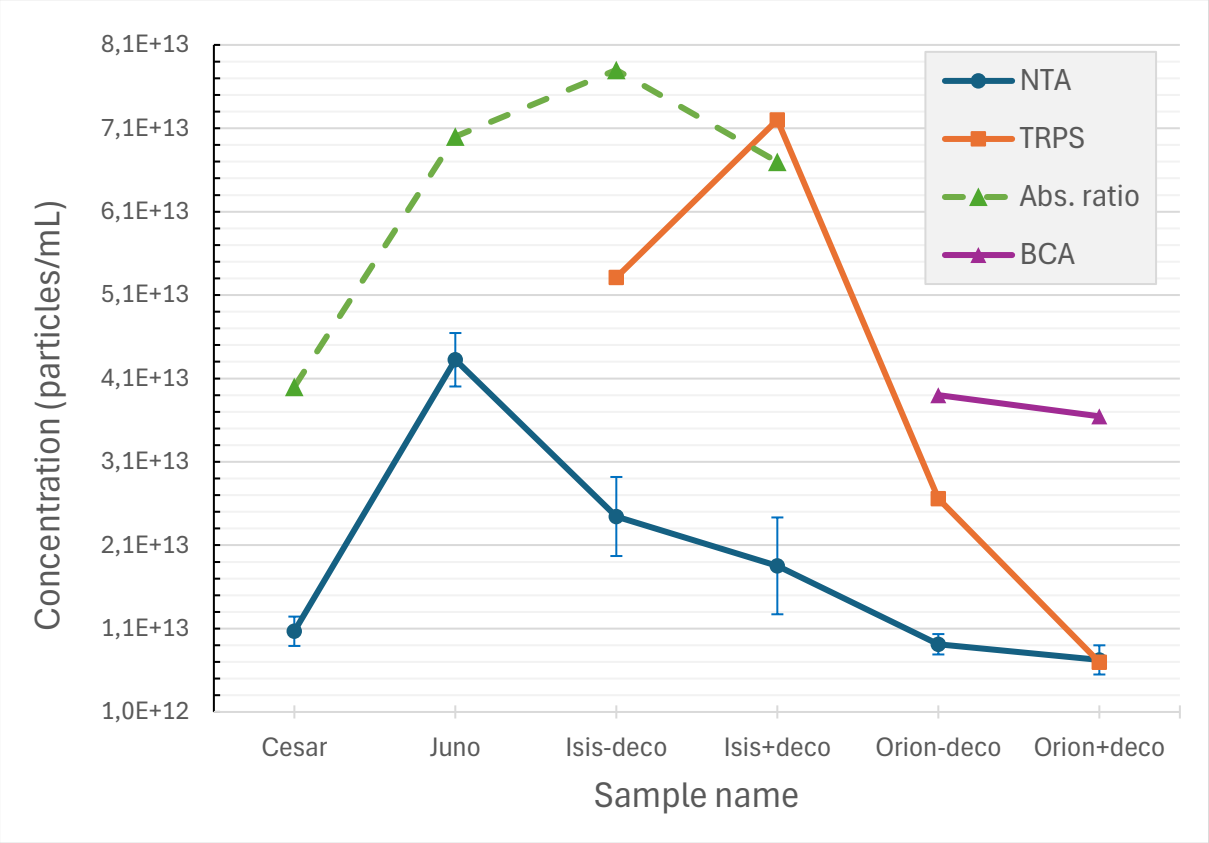


Figure 109. Concentration evaluation by different methods. Legend (top right): NTA, in blue circles; TRPS, in orange triangles; the reported data by I2BC, in green squares. Error bars present the standard deviation, where applicable. Samples are: Cesar, Juno & Isis (phCLP), Orion (recCLP); -deco: non-Decorated, +deco: Decorated.

The I2BC evaluations are not the same between phCLP (in green) and recCLPs (in purple). The previous method for concentration evaluation, applied to phCLPs, was the one developed by Olivier Preux during his thesis, named the (A_{260}/A_{280}) absorbance ratio.[233] Nicolas Ducrot, in his thesis, argued that this method was not well suited for recCLP, arguing that it

underestimated the real concentrations.[209] He used the BCA (BiCinchoninic acid Assay)[290] approach, which he argued that it overcomes the diffusion factor constraint associated with capsid shape. When comparing the Abs ratio method to BCA, he found that the first underestimated concentrations by ~50 %.[209] More details on this analysis can be found in supplementary. I believe that this overestimation is due to the relative nature of these evaluations, and the fact that it is based only on measuring the protein concentration in solution. This of course generates a major bias, given that any floating proteins that might be left in the preparation can be counted in the final particle concentration. That is why I believe that this data is not robust enough to show the real particle concentration in solution.

In addition, these are single TRPS measurement, and thus we have no idea of the possible error bars of these measurements, and also if the tendencies observed are of any major significance. Given this information, we can question the result by TRPS, and ask that further examination for these samples should be acquired in order to conclude. To summarise we can only confidently rely on the NTA results as the most controlled and optimised method for concentration evaluations for these particles in the future, with the evaluation pipeline I established.

Therefore, for NTA we can see that the concentration varies in the range between $0.5^{e}13$ particles/mL and $4.7^{e}13$ particles/mL.

We can observe that for the inter-batch comparison, i.e. between all three of the phCLP non-decorated particles, the variations are significant between the same preparations, doubling from Cesar *to* (Isis- deco) *to* Juno. This is important information in regard to the production of these batches, meaning that one would want to have the highest particle production for a vaccine preparation, which would be Juno in this case.

Next, we compare the samples differently produced (phCLP/ recCLP), before and after their decoration. I measured that Orion (recCLP) is in a lower concentration, lying in the range $0.5\text{-}1^{e}13$ particles/mL, whereas Isis (phCLP) is found in the range $1.3\text{-}2.9^{e}13$ particles/mL. Nevertheless, no significant change is seen between decorated versus non-decorated CLPs, both for Isis or Orion, since the error bars overlap. The main results here is that recCLP are all in lower concentrations than phCLP, for the observed batches.

Let us now focus on the comparison between the different batches of non-decorated phCLPs, that is for Cesar, Juno & Isis.

6.3 Aging effects on non-decorated phCLPs (Inter-batch comparison)

The following analysis will be presented grouped by the method used, thus primarily I will start with the results from TEM, followed by the comparison by NTA, and lastly show the Bottom-up proteomics (cf. as shown in the methodology in Figure 108).

This type of analysis has never been done before by our group at this level of detail, and especially not using different structural methods simultaneously. In addition, given that Cesar, Juno & Isis, were produced in July 2021, June 2022, and November 2022, respectively, this work has the added value of laying the groundwork for examining the stability of the samples over large timescales, i.e. their aging over several years, which is especially important for vaccine preparations.

6.3.1 Results by TEM

6.3.1.1 Before filtering

First, I examined the samples by TEM, comparing Cesar, Juno & Isis (-deco), as in Figure *110*. Here I showed the data only for SST (Sodium silicotungstate) which was found to be the most suitable stain for TEM analysis on these samples and it will thus be the only stain used for any future TEM analysis (cf. Chapter 5). The size presented is the Equivalent spherical diameter (ESD). All the data shown here was collected in November-December 2022.

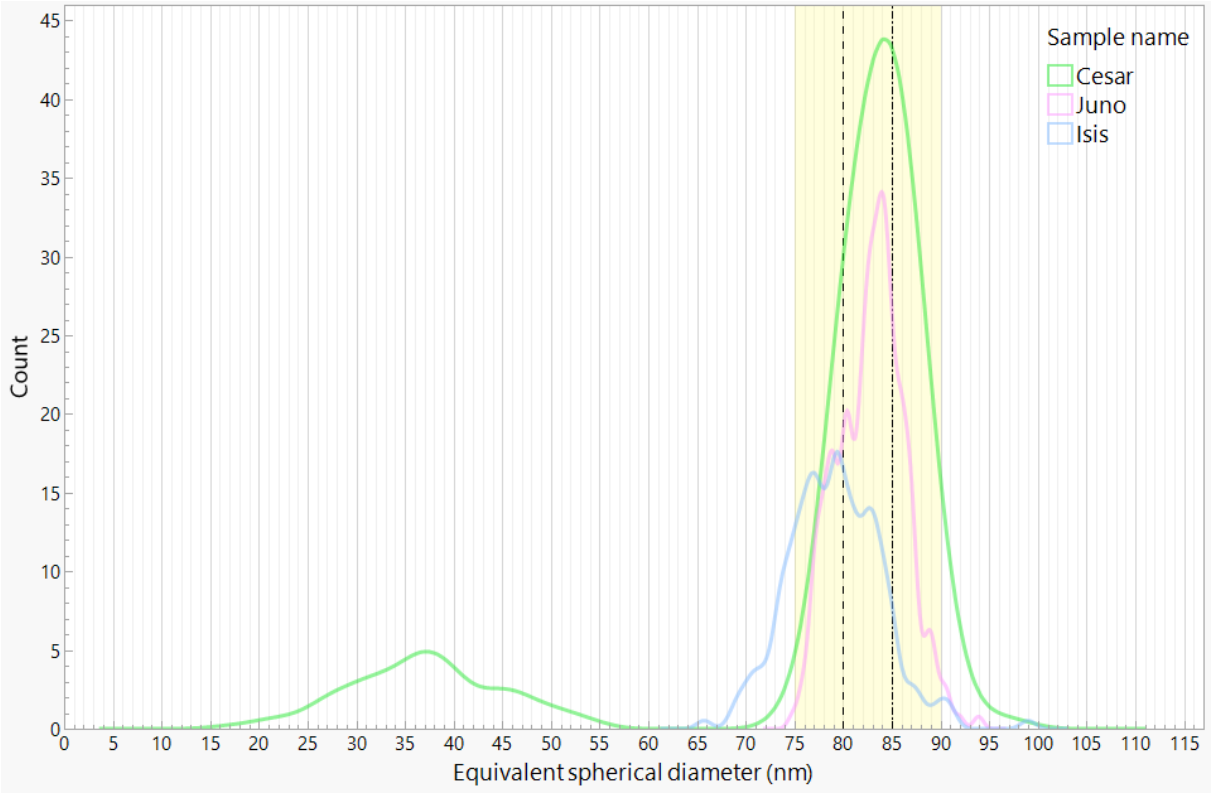


Figure 110. TEM size distribution measurement for T5-CLPs, comparing biological triplicates (phCLP; non-Decorated). Non-filtered data. Legend (top right), sample names: Cesar (green; N= 568), Juno (violet; N= 261), Isis (blue; N= 202). Line smoothness factor = (-0.303). The yellow highlighted band is the expected size range 75-90 nm. The dashed vertical lines at 80 & 85 nm serve as reference. Bin size = 1 nm.

We can see that the main part of Cesar, Juno and Isis distributions is found in the expected size range (highlighted in yellow). Cesar overlaps nicely with the centre of mass of the distribution of Juno and Isis. This suggests that the triplicates are similar to one another in the majority of their distributions. We also not that Cesar presents a second smaller distribution, centring at 36 (± 7.9) nm, which is most likely a sign of sample degradation in time, thus forming smaller particles from aggregated floating proteins.

Looking at Juno and Isis samples, it appears that their size distributions have many peaks. These peaks are presumably artefacts due to sub-sampling problems. Indeed, Cesar distribution was obtained with almost 600 measured particles whereas Juno and Isis distributions are obtained with 250 and 200 measured particles respectively. Usually, for monomodal distributions, 500 data points are reasonable to begin obtaining reliable statistics: this is why Cesar distribution appears well defined and smooth. In the opposite, Juno and Isis distributions have not enough data points to be well resolved: this is why peaks appear in their profile. Consequently, trying to find explanation for these peaks is irrelevant. Previous observations suggest that collecting more statistics when performing TEM analysis will enhance Juno and Isis distribution profiles and thus, help to draw definitive conclusion about sample similarity.

But the previous graph does not account for the broken capsids, which can be identified with the algorithm described in Chapter #5.

6.3.1.2 *After filtering*

In the following graph (Figure 111), I separated the broken capsids (top panel) from the main distribution, in order to focus only on the regular shaped capsids, i.e. spherical and elliptical (bottom panel). I chose to only show it for Cesar, since it was the only sample that has enough statistics from Figure *110*.

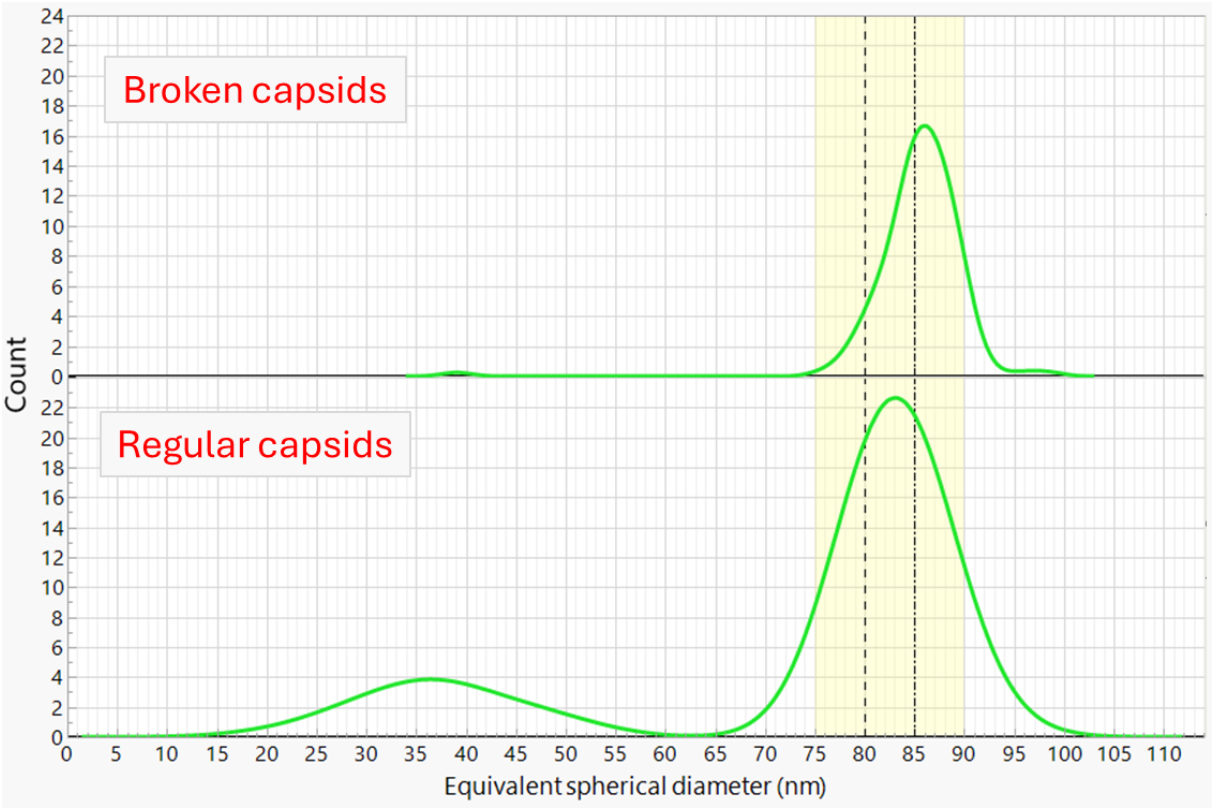


Figure 111. TEM size distribution measurement for T5-CLPs, comparing Cesar (phCLP; non-Decorated). Filtered data: broken capsids (top panel); regular capsids {i.e. spherical, polygonal & elliptical} (bottom panel). The yellow highlighted band is the expected size range 75-90 nm. The dashed vertical lines at 80 & 85 nm serve as reference. Bin size = 1 nm.

We can see again from Figure 111 that the maxima of these distributions still fall within the expected size range (highlighted in yellow). For Cesar, we can also see that the main peak shifts for the broken capsids, going from 83 to 86 nm, although the lower distribution, between 20 to 50 nm, is actually more visible for regular capsid particles.

6.3.1.3 *Conclusions on TEM*

If the distributions of Cesar, Juno and Isis seem to be globally similar, drawing definitive conclusion for stronger similarities is difficult, due to a lack of statistics. To reach an objective, of statistical significance, I think that one has to count at least 900-1000 capsids in total, to

account for the large portion of broken capsids. I would say that at the current stage, the only relevant observations would be for Cesar, and the corresponding size shift when comparing regular versus broken capsids. The observed shift seems reasonable, given that intuitively more broken capsids would have a larger perimeter, thus larger size.

The relevance of the applied filter comes from the fact that, after performing the filtering of the broken capsids, explicit dominant peaks emerge for these distributions. It has to be highlighted as well that the algorithm used here is a beta version, meaning that further analytical or empirical-based improvements can result in better particle selection and thus, better filtering of the data. One way to do this would be to manually select a large number of particles and train a machine learning algorithm in order to select automatically through images in future samples.

In addition, further examinations of batches produced simultaneously and examined by the proposed TEM methodology, can perhaps allow in distinguishing if these size changes are indeed of critical structural value. Or perhaps, if these observed tendencies, are indeed coming from sample aging, this can be relevant to future applications in distinguishing degraded particles, and thus service as quality control attribute for vaccine production. This hypothesis can be tested in the future by performing regular checkups of freshly received samples, for example every couple of months, in order to see if this observed peak shifting to the right (i.e. larger particles) is indeed due to sample aging.

Given this lack of statistics coming from the TEM, I decided to perform measurements by NTA. This is an in-solution method, which allows to have more statistics and does not have the bias of selecting particle visually and neither of selecting the broken capsids out of the size evaluation, as the one that I developed using TEM.

6.3.2 Results by NTA

The second step was the examination by NTA. As mentioned in Chapter 4, the particle size distribution is log-normal, and I should represent the data accordingly. However, since a variable following a log-normal distribution equivalently means that the logarithm of the same variable follows a normal distribution, I will present these results in this formalism. It is indeed easier and more common to analyse data coming from a normal distribution than from a log-normal distribution. Thus, in the following, I will only consider the logarithm of the particle size data.

With this in mind, I obtained Figure 106 that represents the data distribution for three samples (Cesar, Juno, Isis-deco). It is worth noting that number of particles used to obtain

Figure 106, goes from ~47’000 to 125’000, meaning that the results obtained are statistically significant.

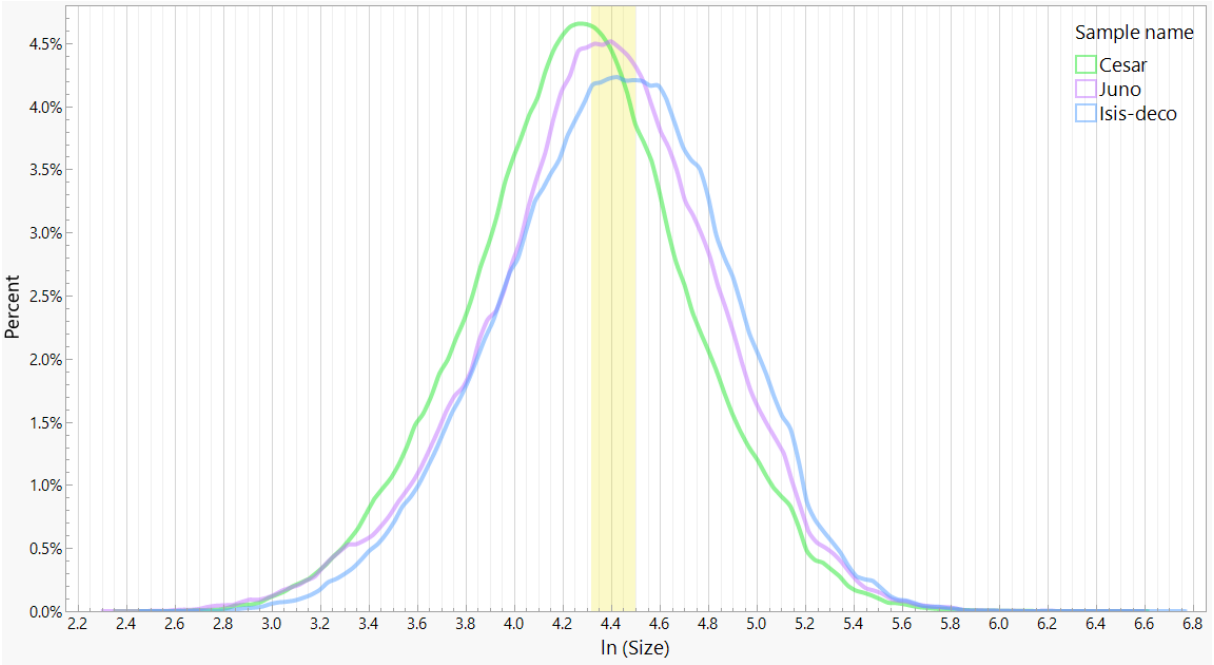


Figure 112. Histogram of the lognormal of the hydrodynamic diameter, measured by NTA. Samples are: Cesar, Juno & Isis (phCLP)*; non-Decorated. The yellow highlighted band is the expected size range 75-90 nm. Legend: (top right) sample names:* Cesar (green), Juno (violet), Isis (blue).

First, we observe that the three curves resemble normal distributions, as expected. We also notice that the main peaks of the distributions are found below the expected range (highlighted in yellow, corresponding to 75-90 nm); and that these peaks are offset from one another. This last observation is unexpected, and it is interesting to verify if these offsets are significant or not. To assess this, I could have used specific statistical tests (such as the Kolmogorov-Smirnov test). However, since the studied data follow normal distributions, I will only verify if the mean and standard deviation can be considered similar. This is done by computing an error interval for the mean and standard deviation of each sample distribution and verifying if the obtained intervals overlap.

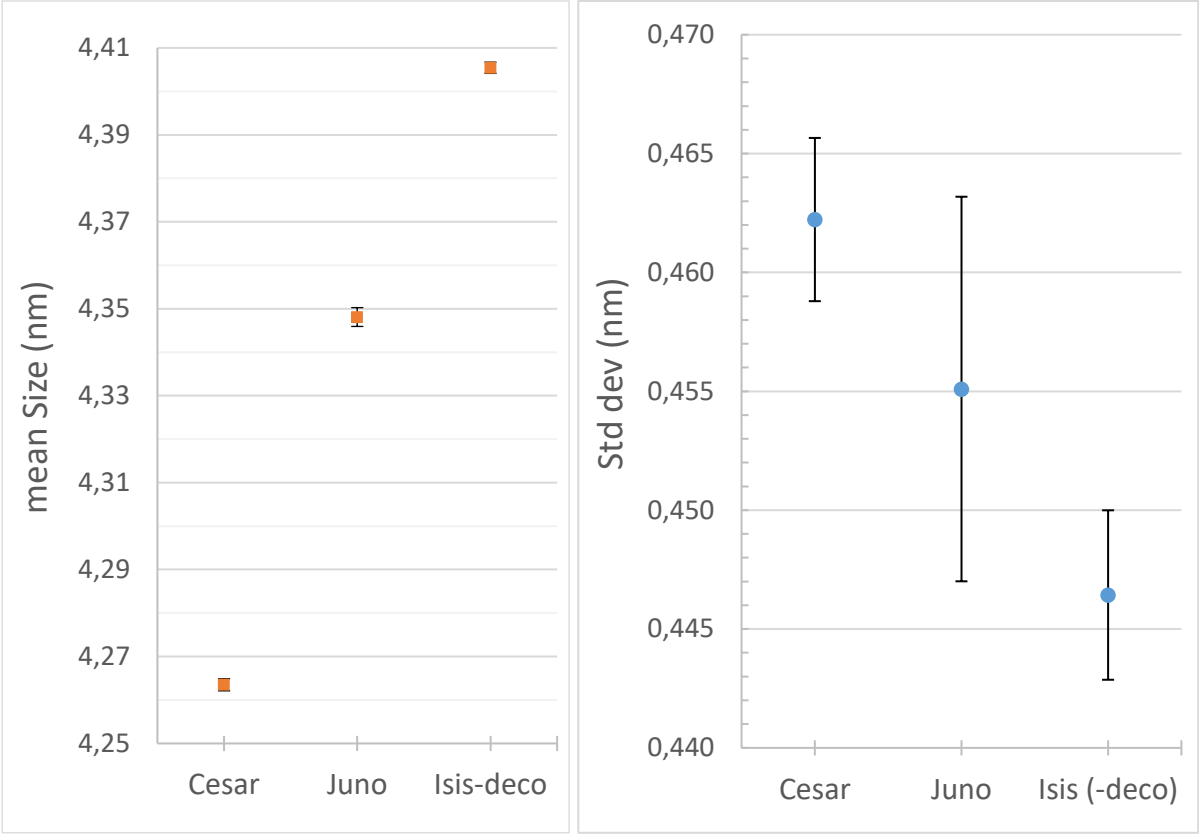


Figure 113. Central statistics of the phCLP (-deco) triplicates. Mean (orange square), and its standard error of the mean (SEM); Standard deviation (blue circle), and its standard error. We recall that the samples are ordered according to their date of production (Cesar being the oldest).

When comparing the samples, we first observe that the older a sample is, the lower its corresponding mean. Then, it appears that the error intervals for the mean (in orange, Figure 113 left) do not overlap for all three samples, indicating that the means, and consequently the samples, are significantly different from one another. If I were to order the samples according to the means of their distributions, the order would be as follows:

Cesar < Juno < Isis-deco

Similarly, we observe that the older a sample is, the higher its corresponding standard deviation. Additionally, the error intervals for the standard deviation of Cesar and Isis-deco distributions (blue, Figure 113 right) do not overlap, indicating that their standard deviations are significantly different. The situation is less clear for Juno since its error interval overlaps with those of the other two samples. Statistically the data count for Juno is the lowest, 2.3 to 2.7 times compared to the other two samples and probably explains the overlap. This most definitely contributes to it having the largest error to the other two samples. From a biological perspective, it can also probably be due to the large presence of contaminant phages in the sample, indicated by the TEM images (Figure 69).

My hypothesis is that the older the sample, the smaller the particles and consequently the smaller the distribution means. This may be due to the degradation of particles over time, with these particles losing some of their integrity and structural proteins, thus forming smaller particle conformations. It also seems logical that older sample distributions possess a larger standard deviation, due to increased degradation with time. This hypothesis can perhaps be tested in the future by performing regular checkups of freshly received samples (e.g. every couple of months), in order to check if this observed distribution enlargement and peak shift to the left (i.e. smaller particles) is indeed due to sample aging, or perhaps to something that we are still not aware of. This is an important conclusion as it indicates that this observation by NTA can be a quality control criterion, that is easy to implement for our lab, but also for future routine use in vaccine production.

An interesting evaluation to check the 'smaller particle' production hypothesis would be to follow how the diffusion coefficient of these particle changes over time. This information is extracted by NTA, in order to calculate the particle size, but I did not have the time to properly investigate this property during my thesis. I believe that this measurement can be useful to be checked more in detail in future analysis, knowing that when particles become smaller, they fluctuate at a higher frequency, thus having a faster diffusion speed. In addition to this correlating the particle counts for the same experimental conditions (i.e. dilution; buffer), for the same sample, followed in time, could indicate that particle count has increased in time (due to the increasing smaller particles being formed).

Another interesting evaluation for this hypothesis would be to perform the comparative measurement with DLS, since it is a population-based measurement. This will give us an evaluation if the particles get degraded over time (i.e. giving a transition from larger and fewer particles, to smaller and more numerous particles) will result most likely result in: faster fluctuations in the scattered light intensity; increased influence of number fluctuations (i.e. the number of particles entering and leaving the scattering volume), and a shift in the optimal concentration range to higher values.[291]

Now that we examined the biophysical sizing methods, let us look at the biochemical examination by bottom-up proteomics in order to get an evaluation of the protein content in these samples and how they compare between them.

6.3.3 Results by Bottom-up Proteomics

The last step of comparing these triplicates was the examination by Proteomics. Here we were not able to analyse Cesar, since it was considered to be too old for the proteomics platform in our laboratory. So, we only compared between Juno and Isis (in Figure 114).

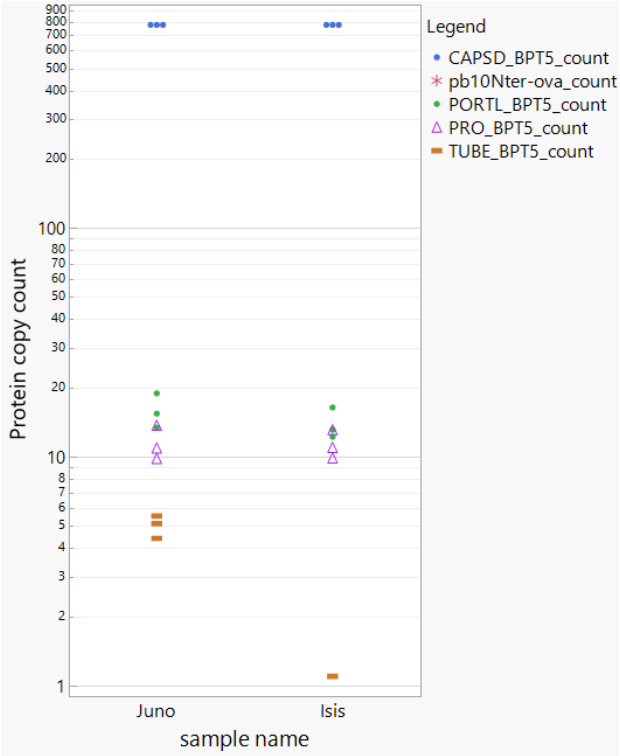


Figure 114. Protein copy count as a function of sample. Samples are: Juno & Isis (phCLP), non-Decorated. Legend for the proteins (top right): CAPSD_BPT5: Major capsid protein (blue dot); pb10Nter-ova: modified decoration protein (red asterisk); PORTL_BPT5: Portal protein (green dot); PRO_BPT5: Prohead protease (violet triangle); TUBE_BPT5 : Tail tube protein pb6 (orange stick).

We can see that copy counts are very similar between the two samples. The Protease (PRO_BPT5) and Portal protein (PORTL_BPT5) have similar copy count, in the range 10-20 copies. This results is similar to what was reported before in the thesis work of Kavya Clement.[202] Indicating that the Protease most probably plays a structural role in T5 capsids, even after it has removed the scaffolding proteins, which is something of major interest for CLP production. The Tube protein presents a higher quantity by 3 to 5-fold for Juno compared to Isis, which is interesting, as it was also observed by TEM images that contamination by tails was present for Juno (Figure 68).

6.3.4 Conclusions on the inter-batch comparison

Checking by the biophysical methods, we can see that their central tendency mainly fall in the expected size range (75-90 nm, highlighted by the yellow band) for TEM and that the particles overall distribution overlap with each other in their majority, indicating that we are

indeed measuring similar particles (Figure *110*). Whereas when using NTA, I saw that the centres of the distributions were significantly different form one another (Figure 113), suggesting major structural differences between these samples. I observed that in the case of NTA (Figure 112), it was the newer samples that had the larger sizes, which seemed counter intuitive.

Hence, I believe that the NTA results are heavily influenced by the presence of broken capsids or even pieces of broken capsids. This could draw down the average size seen for this population, resulting in the order inversion seen between these two methods. Of course, the next logical step would be to check every newly received batch still behave similarly, visualise them by TEM, and observe if they consistently present a larger size in NTA, when compared to the current samples in the lab. Also, regular follow-up of these samples for example every couple of months, by both these methods could indicate if this is indeed an aging induced effect, or something else that is actually structurally different between these samples, which would have important implications on the way these samples are produced.

When looking at the bottom-up proteomics data we see that the samples are very similar to one another, presenting almost identical copy numbers, meaning that we are indeed observing similar samples, at the molecular level. These results also contributed to evaluating the ratio of the contamination by phages for Juno, based on the quantity of tail proteins in solution, which had already been visually confirmed by TEM analysis. This orthogonal verification highlights the need to combine these structural methods with one another when examining these biological samples, reinforcing the idea for the need of the proposes analytical pipeline (cf. Figure 62).

Now that I established that particle aging most likely has structural implications on non-decorated phCLP samples, let us focus on examining particles that were produced in the same time, using both production methods (phCLP/recCLP), and also checking how the presence or absence of the decoration protein can affect their structural confirmation.

6.4 Production method & Decoration status evaluation

In Nicolas Ducrot thesis, he observed that recCLP are bigger than phCLPs, suggesting that recCLP are already expanded during purification on a glycerol gradient.[209] The vast majority of these capsids have a diameter of ~90 nm, and a morphology corresponding to that of expanded phCLPs.

In this subchapter I will discuss the different particles generated by these two production methods (phCLP/recCLP) and combine this to the effects that the decoration by pNO (pb10 N-

terminal combined to ovalbumin) might have on the structures as a whole. It is important to highlight here that all the following samples were acquired in the same time period (November 2022). This means that any aging effects described in the previous portion of this chapter should be therefore neglectable. I would like to point out that the legend for the following figures will be as follows:

- Production method:
  - phCLP: blue;
  - recCLP: red;
- Decoration status:
  - (-deco): light solid lines (as in light weight);
  - (+deco): dark dashed lines).

The first examination is that by the classical negative staining TEM method as shown below.

6.4.1 Results by TEM

In Figure 115, we can see the comparison of the different T5-CLP samples for Isis (phCLP), and Orion (recCLP), for TEM measurements using the SST staining. The data was acquired in the same day and under the same experimental conditions.

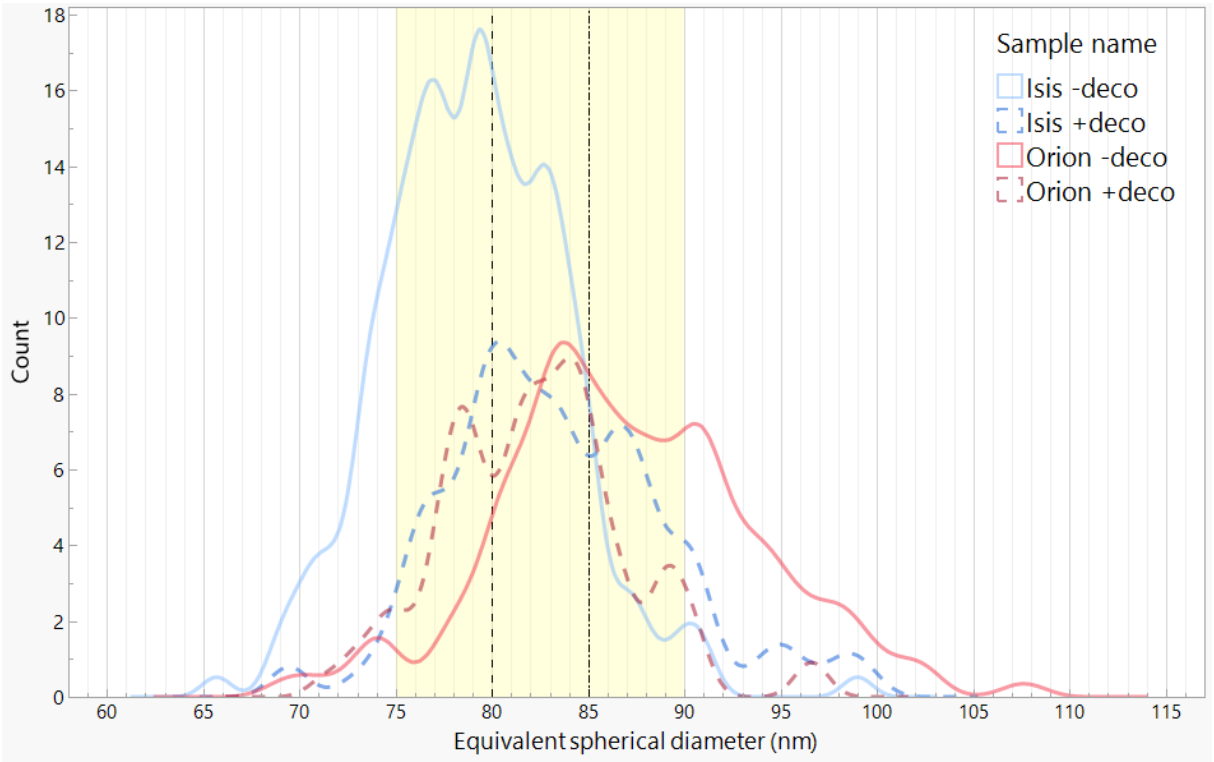


Figure 115. TEM size measurement for T5-CLPs, comparing production method and decoration. Non-filtered data. Samples are (top right): Isis (phCLP; blue), Orion (recCLP; red); -deco: non-Decorated (light solid lines), +deco: Decorated (dark dashed lines). Statistics: Isis-deco (N= 202); Isis+deco (N= 120); Orion -deco (N= 136); Orion +deco (N= 96). Line smoothness factor

= (-0.303). The yellow highlighted band is the expected size range 75-90 nm. The dashed vertical lines at 80 & 85 nm serve as reference. Bin size = 1 nm.

We can see that the main part of these distributions are mainly found in the expected range (highlighted in yellow). Unfortunately, we have even less statistics of what was compared in Figure 111, so we have to be very careful when comparing these results. Regarding the contribution of broken capsids, I also performed the before and after the TEM filter evaluation (results are not shown) and the results are the same. This filtering reduces even more the statistics, meaning that the results are most likely even more erroneous and probably misleading.

We can nevertheless see that the main tendency seems to be that recombinantly produced particles (recCLPs) are larger in size than phage derived particles (phCLP). This result is confirmed by what was observed by our colleagues at I2BC, and demonstrated in the results reported in the thesis of Nicolas Ducrot thesis, using the cryoEM method.[209]

Let us now see what happens to the sample if we analyse it with a solution method, which has much more statistics, and does not have a filter on separating between broken or regular capsids.

6.4.2 Results by NTA

As was done for the examination of the phCLP triplicates, the second step to examining the production method and decoration in these particles was the examination by NTA (Figure 116). Similar to the previous section, I will present the distribution of the logarithm of particle sizes, to facilitate analyses. Observe that in this case, the number of counted particles goes from ~100’000 to 150’000 (supplementary Table 29), which indicates that the following results are statistically significant, contrary to the TEM results.

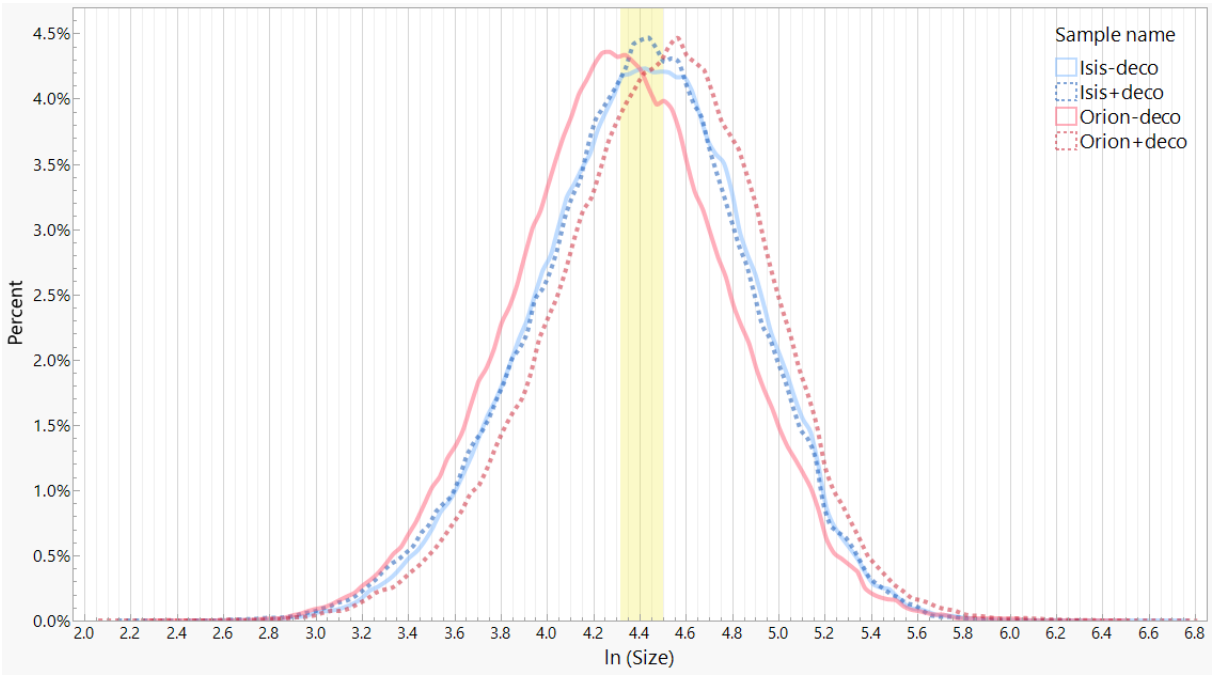


Figure 116. NTA measurement of the particle size as a function of the samples. Samples are: Isis (phCLP; blue), Orion (recCLP; red); -deco: non-Decorated (light solid lines), +deco: Decorated (dark dashed lines). The yellow highlighted band is the expected size range 75-90 nm. Legend (top right) shows the different samples. (top left) The mean, the standard deviation and the CI= 95% for each respective sample.

Once again, we first observe that the obtained distributions are very similar to normal distributions, as expected. We also observe that the peaks of the different distributions are offset from one another. This is clearly unexpected and needs a closer examination to be validated. To this end, I computed the error intervals for both the mean and standard deviation of each sample distribution (Figure 111) and determined the similarity or dissimilarity of samples based on whether the error intervals overlap.

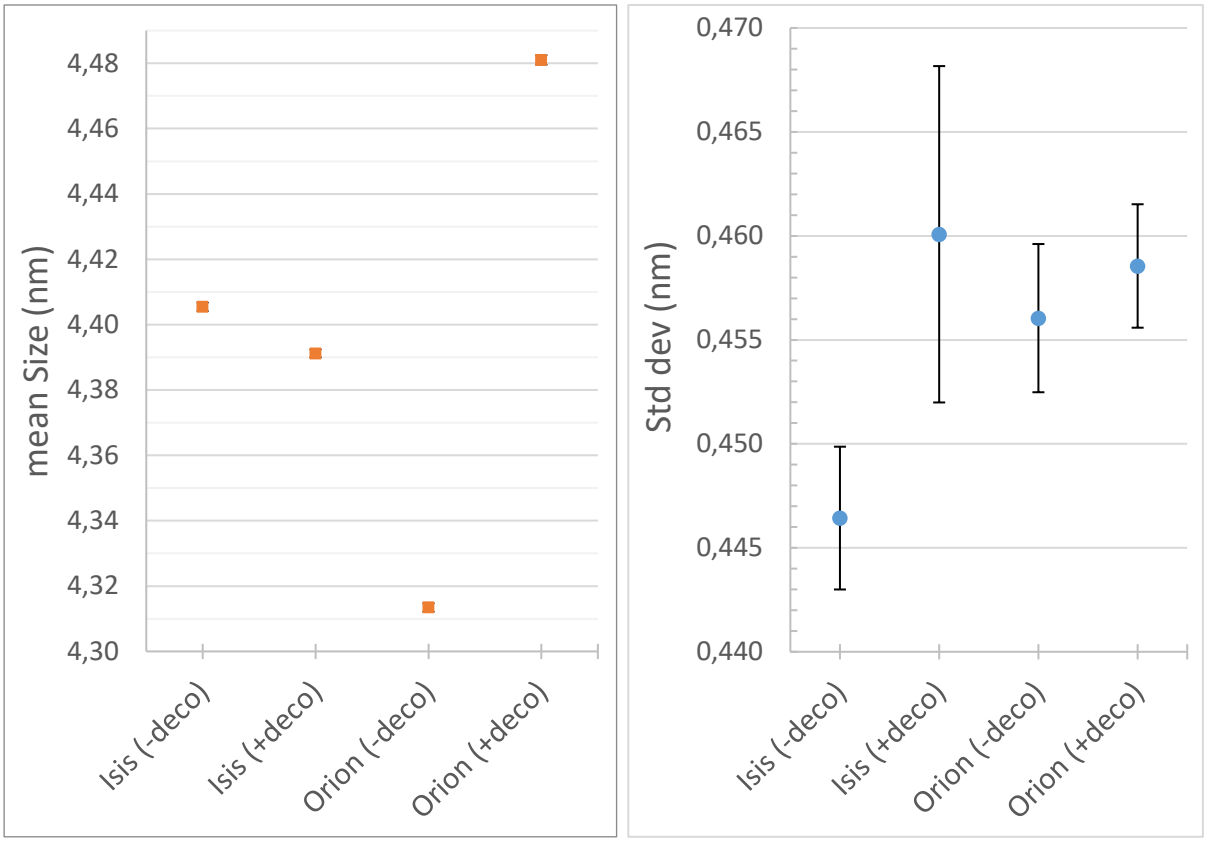


Figure 117. Central statistics of the samples comparing production method and decoration. Samples are: Isis (phCLP), Orion (recCLP); -deco: non-Decorated, +deco: Decorated. Mean (orange square), and its SEM; Standard deviation (blue circle), and its standard error.

Looking at the means (in orange, Figure 117 left), three conclusions can be drawn. First, the Isis and Orion batches are clearly different, as their respective means cannot be considered similar. Second, the means of decorated and non-decorated Isis distributions are slightly different, indicating that the presence or absence of decoration only has a slight effect on the Isis distribution. Finally, I conclude that the means of Orion distributions are different from one another. If I were to order the samples according to the means of their distributions, the order would be as follows:

(Orion-deco) < (Isis+deco) < (Isis-deco) < (Orion+deco)

The picture changes when observing the standard deviations and their error intervals (in blue, Figure 117 right). The three batches (Isis-deco, Orion-deco and Orion+deco) have normal distributions with similar standard deviations. In contrast, the Isis-deco batch has a distribution with a smaller standard deviation, which can be also considered different from those of the other batches. Thus, the Isis-deco and Isis+deco distributions have different standard deviations, and the Orion-deco and Orion+deco distributions have the same standard deviations. Here the odd one out is actually Isis-deco, as the other samples are in the same range as Juno (in Figure 113).

This can indicate that the Isis has a structural behaviour different to the others, which is very interesting as this has always been our sample of reference.

As a conclusion, I would say that the decorated and non-decorated Isis samples are quite similar due to the similarity of their distribution means but have slight differences because of their distribution standard deviations. This may be due to differences in the fabrication process. In contrast, the decorated and non-decorated Orion samples are different, due to the difference in their distribution means. As their distribution standard deviations is nevertheless similar, I deduce that the process of fabrication has no other impact for Orion.

The similarity of the two Isis samples seemed rather intuitive, but this analysis indicated that the decoration significantly increases the structure of Orion. This result can also be read as decoration not working (not being present) in Isis but being present only in Orion. This result would be impossible, given that we have already observed these particles with TEM as having decoration on them (cf. Figure 71). Highlighting again the need to have several techniques when observing these samples. If these tendencies are shown to be true, then it would mean that in reality the decoration protein does not interact similarly to the phage-derived capsids (phCLP), which are the closest to the native structure of the virus, compared to the recombinant ones (recCLP). This would mean that the recombinant production methods is not optimized yet to mimic precisely the structure of phages, inducing major structural modifications on the CLPs when they interact with the decoration protein. It could also show that perhaps the decoration protein does not simply attach on the capsids, but also structurally modifies its binding domain, at the centre of hexagons of Major capsid protein (CAPSD_BPT5). These tendencies need to be confirmed with further analysis from other similar CLP batches, again comparing the production method and the decoration protein attachment, in order to make sure that the observed results are indeed of biological relevance.

Next, I decided to evaluate the samples using the TRPS method, in order to have an in-solution orthogonal measurement, and on top of that, it was independently performed by IZON (IZON Science Europe, 69008, Lyon 8eme, FRANCE). These results will be shown in the next paragraph.

6.4.3 Results by TRPS

Since I was unable to access the raw data, I only processed the filtered data from Izon, which represented a single size evaluation for every 1 nm size bin. Each filtered dataset had

only around a hundred data points, which is not enough for a consistent statistical analysis (such as realizing a Kolmogorov-Smirnov test).

I decided to perform a qualitative evaluation, using an empirical cumulative function for size evaluation, as shown in Figure 118. Since the cumulative analysis is a sum-based method of data analysis, it minimizes the effect of outliers or missing values, and thus flattens possible errors that can arise from this initial reduced database.

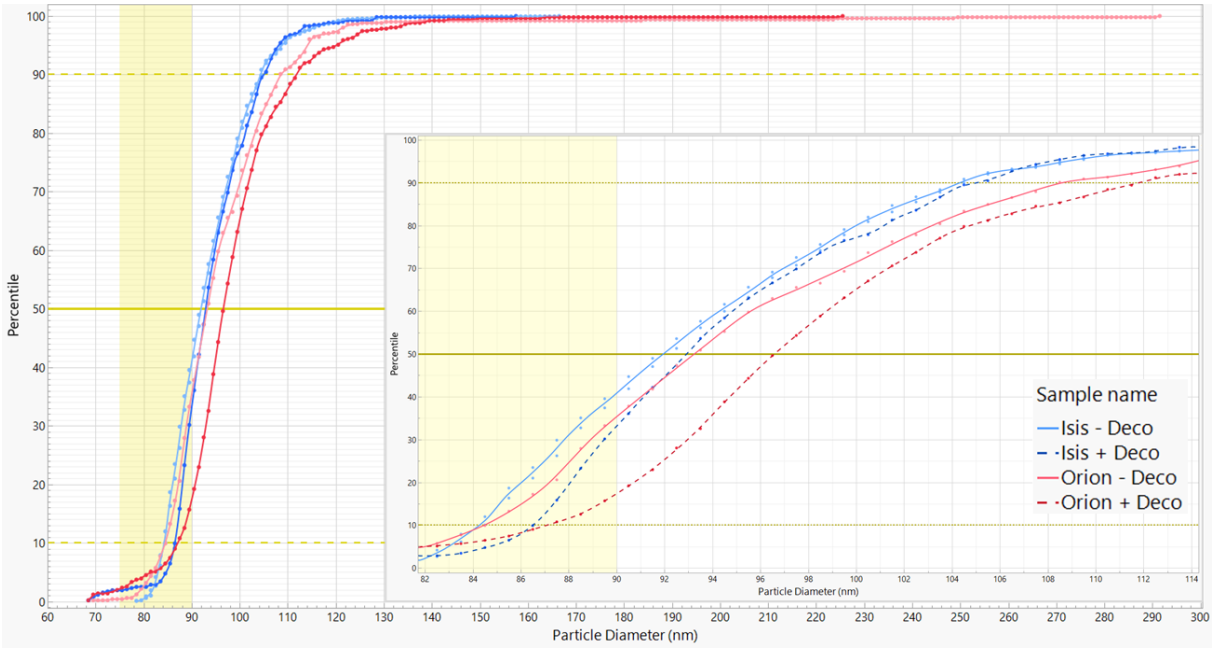


Figure 118. TRPS measurement presented as of the cumulative particle size data, for every 1 nm bin. Inset: zoom in on the size range D10 to D90 (dashed olive lines), D50 point (solid olive line). Samples are: Isis (phCLP; blue), Orion (recCLP; red); -deco: non-Decorated (light solid lines), +deco: Decorated (dark dashed lines).

We can see that the medians (D50 point) of all the samples are found above the expected range (highlighted in yellow). One can imagine that a possible deformation of the particles upon pore-crossing is due to the applied pressure. This would make them appear larger in size, because of a unilateral force applied on one side of them, thus forcing the geometry of the particles.

When we zoom in on the range from D10 to D90 (inset of Figure 118), we can see that the decoration seems to cause a size increase for both Isis and Orion (at their median value). In the case of Isis, the decoration seemed to have a very minor effect, whereas the size increase is much more pronounced for Orion. If I were to order the samples according to the medians of their distributions, the order would be as follows:

(Isis-deco) <= (Isis+deco) ≤ (Orion-deco) < (Orion+deco)

This result corresponds to intuition, where the decorated particle is larger than the non-decorated one, for both samples. We can also see in the inset of Figure 118 that both decorated

(+deco) particles seem larger than their non-decorated (-deco) counterparts in the D10 to D50 range and seem to join to one another on the D50 to D90 range. This could perhaps suggest that smaller particles from when decoration is present, and the upper pert of these distributions is similar between decorated or non-decorated particles.

The presentation in the inset of Figure 118 also helps to better visualize the Span (D90-D10/ D50) for these samples. Which is as follows: Isis-deco (= 0.22); Isis+deco (= 0.20); Orion-deco (= 0.26); Orion+deco (= 0.26). This shows indeed that the phCLP samples have a narrower, more uniform particle size distribution, meaning that they are more likely to be similar between each other during production, when compared to recCLP.

Overall, we can conclude that the structural changes seem to be minor for Isis, but much more significant for Orion. Indicating again that the structure of phCLPs is more stable and regular than that of recCLPs. This would again show that the recombinant production methods is not optimized yet, when compared to the structure of phages, thus inducing major structural modifications on the recCLPs, when they interact with the decoration protein.

More statistics are needed to verify these tendencies, as unfortunately I did not have access to the raw data of these analyses, in order to do more statistics, e.g. analysing the standard error.

6.4.4 Conclusion on the Biophysical characterisation

We can also see that the tendency of TRPS is similar to that observed by NTA, i.e. major structural effects on the confirmation of Orion, but almost no significant observable difference for Isis, although the size evaluated by TRPS are higher. This is a very reassuring result as it shows that the in-solution behaviour of the sample is the same, independent of the method.

And it can even indicate that the results by TEM have to be read very carefully, as although they are claimed to be those of a native ‘golden standard’ method, they implicate having the sample positioned on a metal grid, which can actually modify the confirmation of the particles. One can imagine surface interface interactions or even flattening the structure of the capsid, perhaps due to its less rigid nature, which can be interesting to evaluate with AFM in the future.

Now that I examined and compared the different biophysical sizing methods, let us look again at the biochemical examination by bottom-up proteomics, and examine how these samples compare between them and what is their relative total mass.

6.4.5 Results by Proteomics

We can convert the signal (abundance/intensity) to a respective protein copy count in order to quantify the number of proteins that compose our samples, as we see in the following Figure 119.

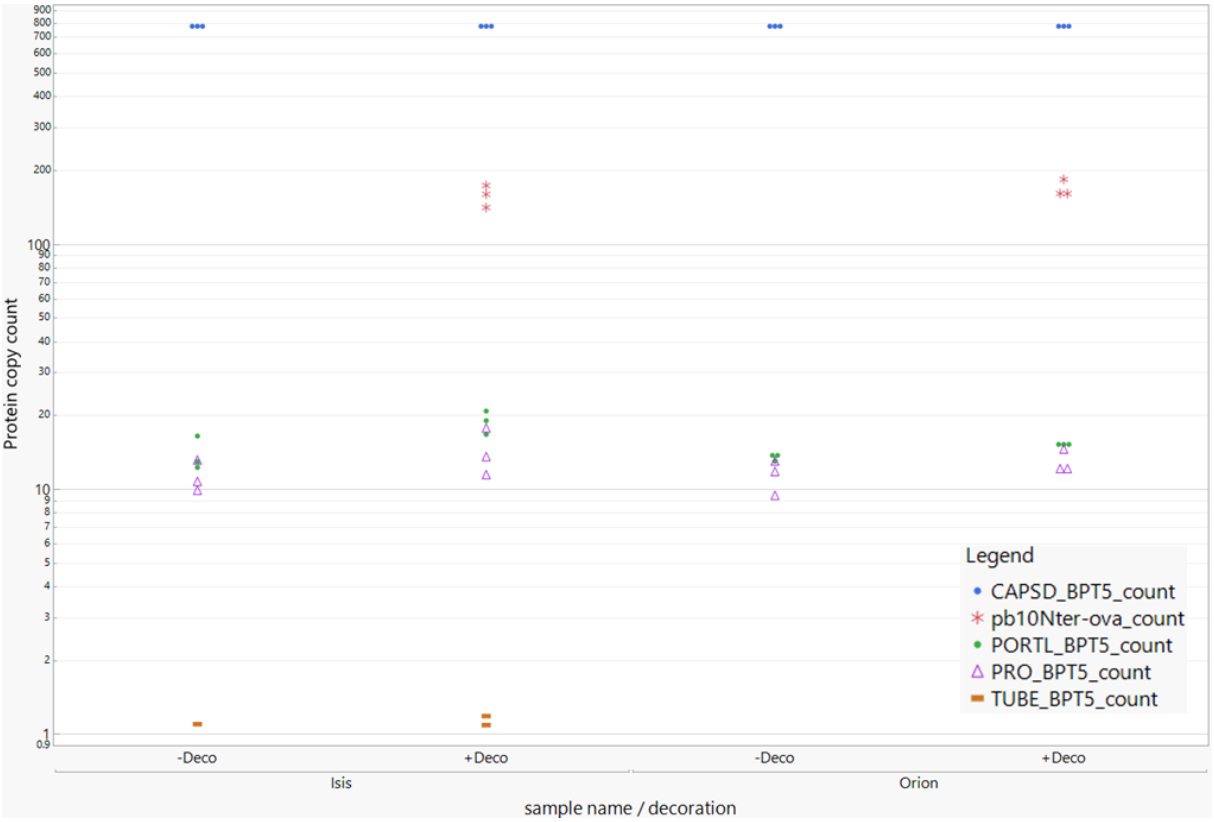


Figure 119. Protein copy count as a function of sample. Samples are: Isis (phCLP), Orion (recCLP); -Deco: non-Decorated, +Deco: Decorated. *Legend for the proteins (bottom right): CAPSD_BPT5: Major capsid protein (blue dot); pb10Nter-ova: modified decoration protein (red asterisk); PORTL_BPT5: Portal protein (green dot); PRO_BPT5: Prohead protease (violet triangle); TUBE_BPT5 : Tail tube protein pb6 (orange stick).*

We can see in Figure 119 that the number of major capsid proteins (CAPSD_BPT5) is always 775, since we want to use this protein as our reference to quantify the other proteins in the sample. We can also see that the data for all technical triplicates is quite reproducible, for all the samples, indicating that there was little technical error during sample preparation. Next, we can see clearly that the decoration protein (pb10Nter-ova) is in the expected range, primarily between 100 and 200 copies, which is close to the expected number (120 copies). Furthermore, we can see that the portal protein (PORTL_BPT5) and the protease proteins (PRO_BPT5) are found in all the samples analysed and are similar in terms of copy number, which is around 10 to 20 copies. Lastly, we can see that the tube protein (orange sticks) is only present for the phCLP samples, and not for recCLP. This indicates that the presence of this protein is a marker of phages being in the CLP preparation, again serving as a contamination checking method.

We need to highlight here that the fact that the protease (PRO_BPT5) is found in all of the samples analysed is a very important result. The known role for this serine protease is that it is involved in the capsid assembly and maturation. It serves to cleave the major capsid protein (CAPSD_BPT5), the decoration protein (DECO_BPT5), the portal protein (PORTL_BPT5), in order to yield mature procapsids competent for DNA packaging. It has also been reported to act as a trigger for assembly of the capsid protein.[212] The fact that we observe this protein in all of the capsids (also seen earlier in Figure 114), in both phCLPs (Juno & Isis) and in the recCLPs, even after the maturation step has concluded. This is an important result, as it indicates that actually the protease might be a structural protein, since it is found in the capsid even at late stages of its assembly, when it is supposed to have finished its role at removing the scaffolding proteins. This could potentially indicate that this protein has another role (currently unknown) inside the capsid after its initial one, raising potentially interesting biological questions, regarding this protein and its interaction with the other proteins inside the capsid.

So now, having these copy counts, we can make the sum of all the proteins in the sample, in order to get the total mass for each capsid.

6.4.5.1 Relative total mass of CLPs

In the following Figure 120 we have the sum of all the proteins found, shown in Figure 119, for each of the samples, and compared to the expected theoretical mass (orange line), calculated by adding only the Major capsid protein (CAPSD_BPT5) and Portal protein (PORTL_BPT5), as was done before by the team in Dominguez et al [21].

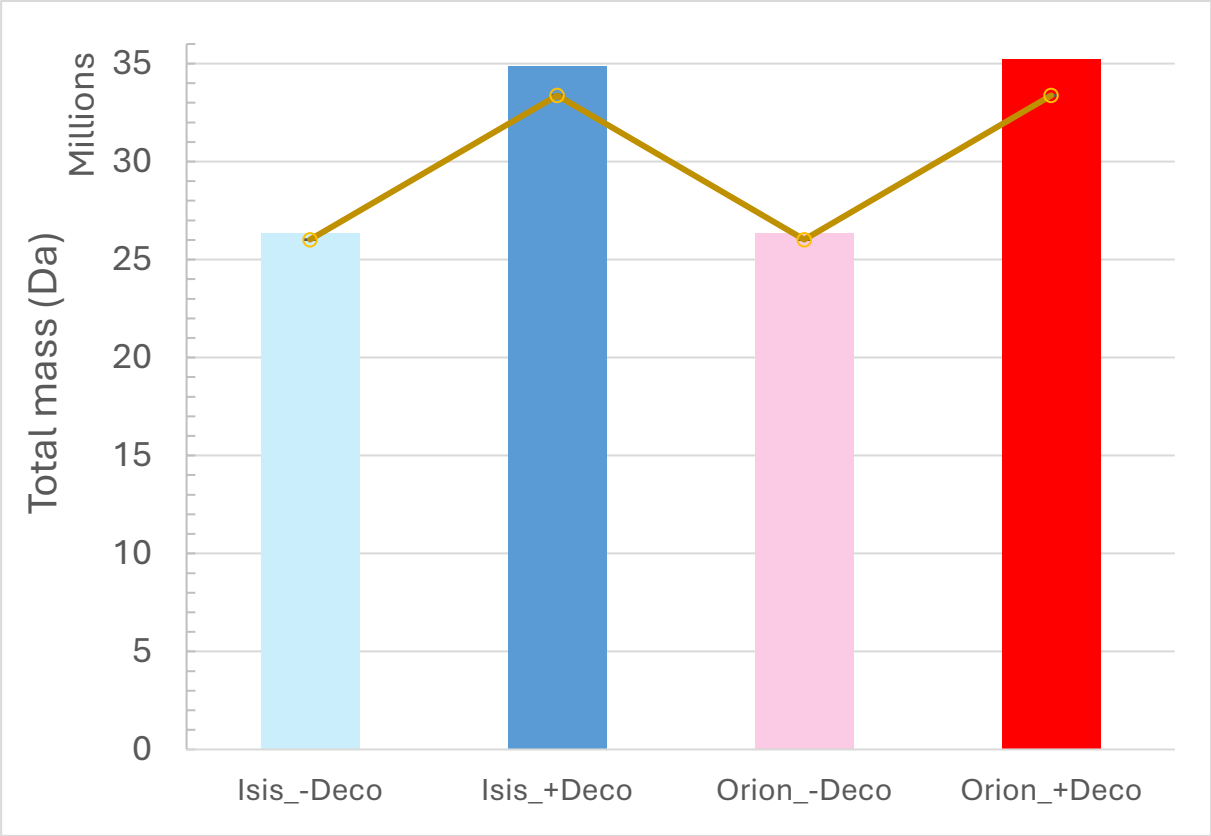


Figure 120. Overall view of the average total mass of CLPs (in MDa) as a function of all the samples analysed. Samples are: Isis (phCLP; blue), Orion (recCLP; red); -Deco: non-Decorated (light shade), +Deco: Decorated (dark shade). In blue: the experimentally based total sum mass calculation (n=3). In orange: the theoretically calculated sum.

We can see in Figure 120 that the theoretical mass is very close to that measured one for the non-decorated CLPs, indicating that the assumption of summing up the proteins found in each capsid is in a good range in respect to the expected mass. We can also see that the theoretical total mass is always slightly underestimated compared to the measured mass. I think this is primarily due to the fact that we did not expect to have protease as abundantly in the sample, in addition to some of the tube proteins being there.

For the decorated samples, we can also see that the evaluation of the total mass was largely higher, overestimating the theoretical mass by several MDa. I believe that this is due to the copy count being larger than the expected 120 copies the decoration protein. Nevertheless, we see that this estimation is reasonable in terms of scale, as we have an increase of ~8-9 MDa, which we expected to have, due to the added 120 decoration protein copies. This shows that it can be an appropriate approach using this relative quantification method, in order to have a better understanding of the structural composition of our CLPs, then to simply make back of the envelope calculations based on literature data. Of course, having an initial theoretical model helps to give a first estimate and an initial range on where to look.

In order to check if the offset of 8 MDa was due to the added decoration protein, we chose to do a further experiment, by performing dialysis on the samples. This was done to see if the added mass was coming from the additional copies of the decoration protein are indeed fixed on the capsid or not.

6.4.5.2 Dialysed or not

We can see in Figure 121 that the samples were compared before and after dialysis for both the non-decorated (panel A) and the decorated (panel B) sample of Isis (phCLP).

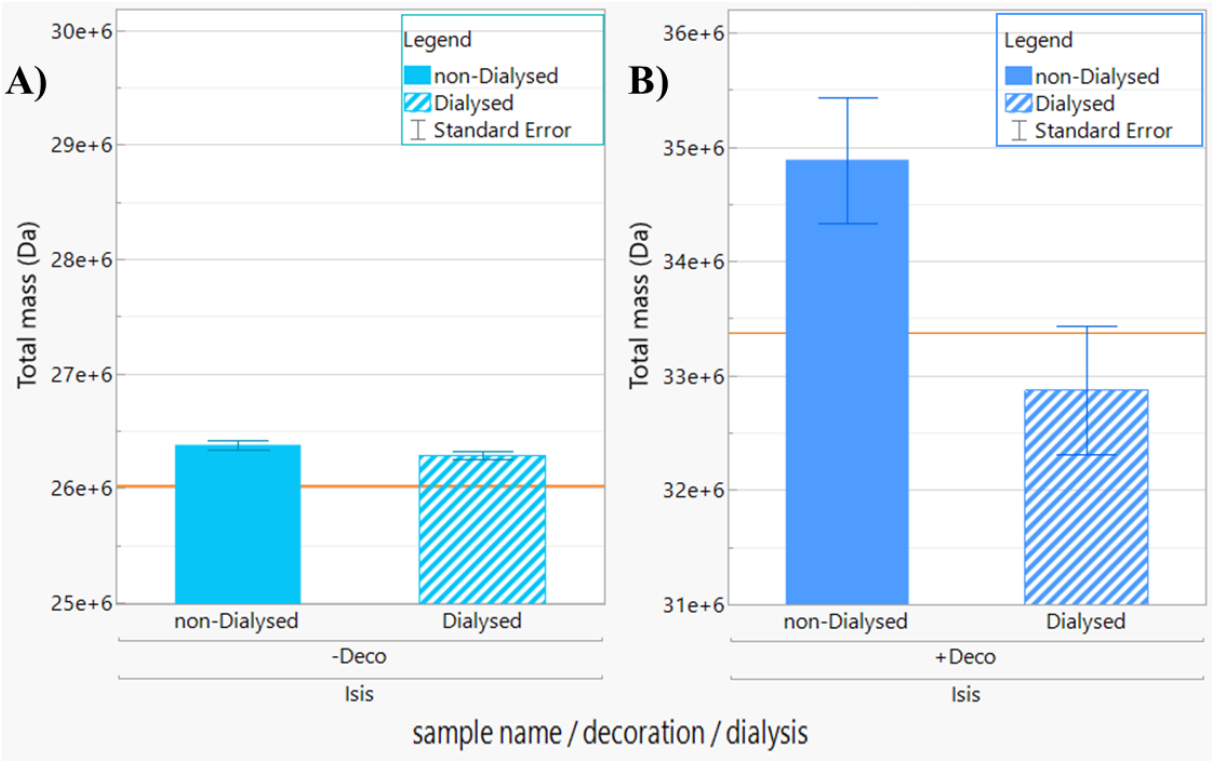


Figure 121. Comparison of effects of dialysis on the average total mass of CLPs (in MDa), for Isis (phCLP). Comparing -Deco: non-Decorated (A panel, aquamarine), and +Deco: Decorated (B panel, dark blue); non-Dialysed (solid), Dialysed (dashed). Error bar represents the standard error of the measurement (n=3). The orange horizontal line represents the expected theoretical mass for both (-Deco, in A) and (+Deco, in B). Scaling on the ordinate axis is 5 MDa for both, as reference.

We can see here that for Isis-deco (Figure 121 A), the mass remains similar and is always higher than the theoretical value, between the non-dialyzed and dialyzed triplicates. So, it shows again that it is probably because I did not account for a high number of proteases in the final stages of the assembled capsid (Figure 119), which does not leave the capsid even after dialysis.

The difference is much more noticeable for Isis +deco (Figure 121 B), where the dialysis sample is ~ 2 MDa smaller. In addition, we see that the mass has gone below the theoretical expected value, but it overlaps the expected value, with its error bars. This could mean that we actually have over-dialyzed our sample, and thus have removed some of the copies that are

attached to these CLPs. But I believe that the most likely explanation, given the high K_D that the protein has to the CLPs[219], is that the copies that were present were actually not attached to these capsid in all the available 120 sites, as these sites might not be available all the time. This can come due to the high presence of irregular particles, as we can imagine that the available sites for the decoration protein docking do not properly form on irregular particles.

Nevertheless, this is an important result showing that a simple dialysis after adding the decoration protein could actually improve the results of the mass measurement, since it reduces the noise that is coming from the sample preparation. And can thus be an added step in future sample preparation procedures before mass evaluation. Further analysis for dialysing the recCLP samples were not successful due to technical issues, but it is an envisioned experiment in the near future.

6.4.6 Conclusions on the Production method & decoration status

In this section of chapter 6, I described the differences between samples that are produced from phage derived capsids (phCLP) and recombinantly derived capsids (recCLP). In addition to that I compared these samples when the decoration by the pNO protein is present (+deco) or absent (-deco), and how this could modify the structural confirmation of these capsids. The batches compared here were all produced in the same time, removing the possible effects that could rise from aging, as was described in the previous section of this chapter, for triplicates of phCLP.

In order to achieve this, I evaluated the samples using three biophysical methods, first the evaluation by the 'golden standard' of sizing methods in virology, which is negative staining TEM; next evaluation by NTA, optimized and performed by myself; and lastly an independent evaluation by the in-solution orthogonal method TRPS. This examination allowed us to see that due to its little statistics and selective particle bias, the usage of TEM was not conclusive, when compared to NTA and TRPS.

These last two approaches showed indeed that there were major structural differences between the samples, primarily concluding that the phCLP tend to have a more compact and smaller confirmation than their recCLP counterpart. In addition, I saw that the size differences were minor for the phCLP particles when decorated, whereas they were very significant when it came to the decoration of recCLP particles. This information is very relevant in showing that the produced recCLP particles, although visually appearing to have many apparent similarities to the native phCLP, they do in reality behave quite differently from one another. This of course

points to the necessity of further optimization of the biotechnological production of these samples, and the continuous and rigorous evaluation by the analytical pipeline established by me during this work (cf. Figure 62).

Regarding the biochemical evaluation by bottom-up proteomics, I was able to verify that the samples are indeed very different when comparing their production methods and when the decoration protein is added. It was clear that the production method has many remnants on the proteins visible in the sample especially the presence of two proteins was an important marker of the phCLP production, even serving as a robust contamination evaluation method. In regard to the decoration method, I clearly saw the difference after the decoration protein was introduced in the sample, and noticed it to be at the expected range, but over the expected copy number (120 copies). I performed an additional dialysis step in order to remove any excess proteins that were found in the sample, which resulted in the copy number to fall much closer to the expected 120 copies. This also allowed me to make an evaluation of the relative total mass of the capsids, which was very close to the theoretically calculated mass range. This presents the idea that the model has limits and needs to be refined, and indicates that in the end, the experimental evaluation should have the final say. This is an important result because it shows that this approach can be on orthogonal check of the protein count of phages, prior to evaluation by native MS methods, such as the NEMS-MS bench.

6.5 Benchmarking SMPS with the T5-CLPs

Of course, when mentioning the NEMS-MS bench, we have to come back to the critical step of nebulising the viruses. So let us come back to the design of the analytical pipeline (Figure 62), where I envisioned that the viruses' size in aerosol can be evaluated by SMPS™, in parallel to reaching the mass detector. As described in Chapter 5, we were able to use the SMPS™ in order to optimize the nebulisation by the SAWN method and characterize it, using PSL-NPs. In this way we were able to establish that the most optimal range for particle generation, where the highest probability of finding one particle per droplet, using our SAWN system, was between 1^e10 - 1^e11 NP/mL.

Therefore, I decided to perform the experiment starting from an initial concentration of T5-CLPs at 5^e10 particles/mL, diluted in water, result shown in the following figure, Figure 122. Using water as a diluent here was a necessity, as PBS is detrimental to the SAWN chips, due to its high salt content (e.g. Dulbecco's PBS solution has 137 mM of NaCl).

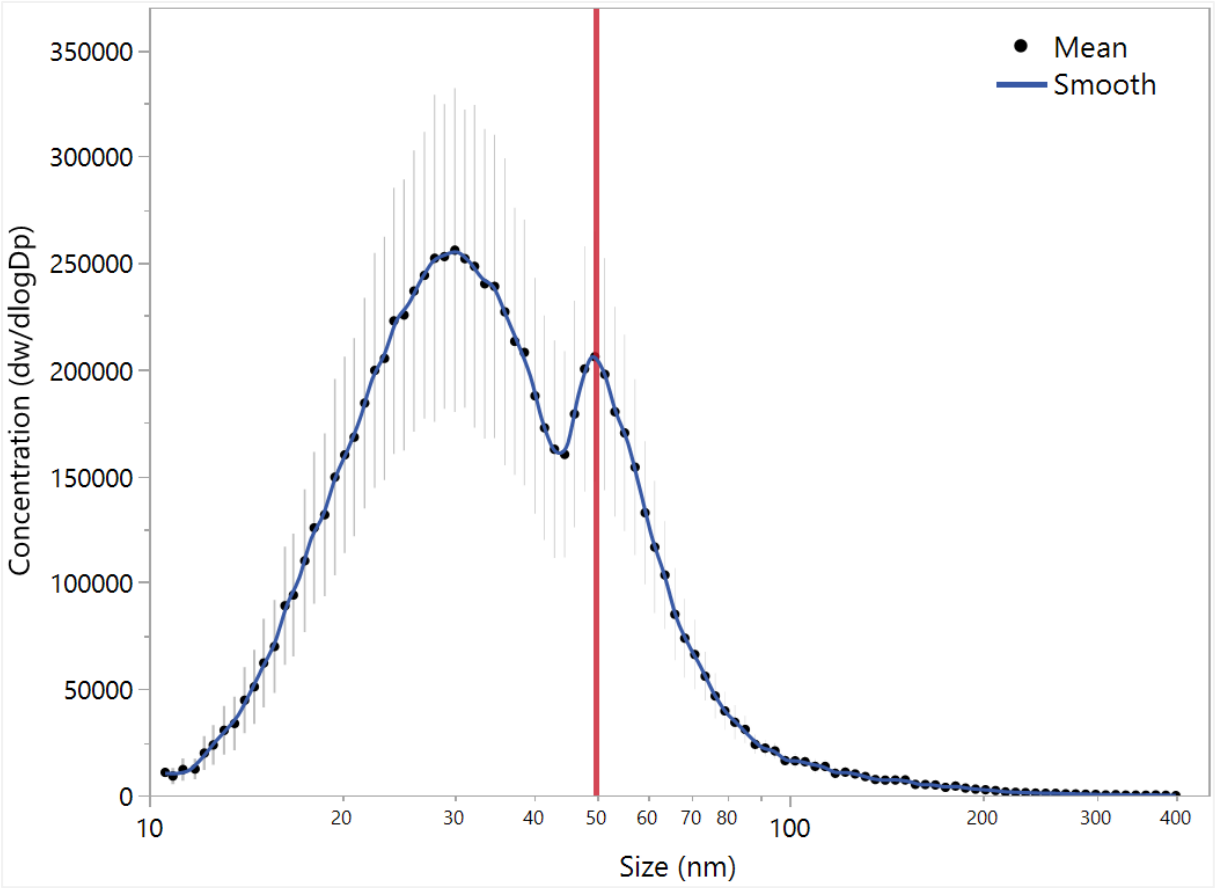


Figure 122. SMPS measurement, showing the normalised particle concentration (dw/dlogDp), generated by SAWN, as a function of particle electrodynamic diameter (nm).Legend: mean values (black dots) and the standard deviation (n=5 consecutive runs, shown in grey); the smoothened tend line is shown in blue; and the expected size for the T5 CLP is shown by the vertical red line.

We can see in Figure 122 that the distribution still presents a main peak at ~30 nm, similar to what was seen prior for PSL-NP concentrations, in Figure 104, which we attributed to the water peak. Nevertheless, we saw a clear smaller peak with a maximum at ~50 nm, which we attributed to the T5-CLPs, although this is much smaller than any other peak that I was able to evaluate for these particles. I believe that this might be due to the surface charge effects that the particle presents when passing through the SMPS instrument. Perhaps indicating that not enough charges are present, or perhaps are not equally distributed on the particle surface, to properly evaluate the surface to volume ratio of these particle inside the detector, thus resulting in a false size evaluation.

We can also see that when measuring our T5-CLPs model, the water noise, although smaller than the particle size, was higher in concentration than the detected viral particle, measured at ~50 nm. Indicating that any particle generated in that particle size range will be hidden by the water noise. This type of size overlapping is very problematic for particle nebulisation for mass measurement and would inundate the signal of the analyte of interest.

In the future, we can envision using aerosol impactors with a low cut off filter, prior to the mass measurement, which would allow to recover the lower size droplets containing most of the water particles, and thus only let the analyte of interest reach the NEMS detectors.

6.6 Conclusions on Chapter 6

In order to summarize the conclusions of this chapter I can start by saying that as expected the particles do indeed present differences between samples, when they are prepared by the phage derived method (phCLP) or the recombinant derived method (recCLP), and in addition to that, the fact that the decoration protein (pb10Nter-Ova) is added on top of these CLPs, also does have an effect on their structure.

Regarding the aging of the samples, I observed that the results by NTA (Figure 112) were in disagreement with TEM (Figure 110), when it comes to the overall behaviour of the samples. Contrary to TEM, where older samples appeared larger, for NTA, it was the newer samples that had the larger sizes, which seemed counter intuitive. I argued that due to the low statistics and selection bias of the TEM method, the NTA was more reliable in giving the true nature of these samples.

A similar comparison was made depending on the production, and yet again the evaluation by the ‘golden standard’ for particle sizing in virology, negative staining TEM, seemed less reliable due to the low statistics. That is why I chose to combine the NTA method to the TRPS, since they are both in solution methods and this last was performed independently. These two approaches showed indeed that there were major structural differences between the samples, primarily concluding that the phCLP tend to have a more compact and smaller confirmation than recCLP. In addition, I saw that the size differences were minor for the phCLP particles when decorated, whereas they were very significant when it came to the decoration of recCLP particles. This information is of major relevance since it shows that the produced recCLP particles are actually very different from their native phCLP counterpart. This of course points to the necessity of further optimization of the biotechnological production of the recCLP samples, and the continuous and rigorous evaluation by the established analytical pipeline that I put in place during my thesis (cf. Figure 62).

I would like to emphasize that NTA, due to its robust statistical power, demonstrates the necessity of continuously using this method for monitoring any samples prior to NEMS-MS. It ensures that the structural conformation of particles within the same sample (intra-sample) is preserved under different experimental conditions, while confirming that observed differences

between different samples (inter-sample) are genuine and not artifacts. This distinction between intra-sample and inter-sample variability should be considered in NEMS-MS evaluations to minimize bias in the interpretation of structural differences based on mass measurements.

Lastly, I decided to go forward in evaluating the mass at a molecular level, for which I decided to use bottom-up proteomics. With this method I performed a relative total mass evaluation, by using the major capsid protein as reference, with a known copy number of 775. In this way, I achieved a mass evaluation that was very much close to the expected theoretical number, and to the prior evaluations performed by our team on similar batches of non-decorated phCLPs.[21,184] I also observed that performing a simple step of dialysis prior to evaluating the mass by proteomics decreased the variance of the measurement and approached the mass to the theoretically expected evaluation. This additional step can be foreseen perhaps in future mass measurements by NEMS-MS, in order to reduce the uncertainties of the total mass evaluation, and make sure to only recover the most basic structural proteins these are found in and on the analysed VLPs.

It is important to highlight here the perspective of using the NEMS-MS bench to evaluate these particles, since now we know that the particles are very different from one another, from both a biophysical and a biochemical point of view, they should indeed present a very different mass when measured by the bench, which would only further confirm the need for these orthogonal observations to be put in place in the near future.

6.6.1 A more philosophical view

As scientists in the quest of ‘the truth’ when we embark on a given scientific endeavour, we always try to be aware and avoid any pitfalls, and to overcome any challenges that can arise in our path towards this quest. This can be philosophically imagined as the Epistemological, (epistḗmē 'knowledge' + logy), study of the nature, in the search for natural truths and knowledge. From this philosophical perspective one can associate this quest to the analogy of Plato’s cave, indicated in the following drawing, Figure 123.



Figure 123. Plato's allegory of the cave. Taken from the article on the The Guernsey Magazine (1882).

In Plato's allegory of the cave, in his work Republic (514a–520a, Book VII), Plato describes how people that have spent their lives in a cave, chained since their childhood and obliged to face a blank wall. Onto the wall, shadows projected from a fire behind them, parade in front of them, and the prisoners give these shadows names. These shadows have now become their reality, but they are in no way an accurate representation of the real world. Plato argued that the philosopher's duty is to get out of the Cave and go towards the sun, were the reality of the world is displayed in front of the eyes of the beholder. In this way it is the duty of the philosopher to bring people towards enlightenment and out of the shadow world of their beliefs.

The reason why this analogy can be of relevance in this work is that by my desire to look at different methods, each method projecting their different shadows, to which we gave names to, that can make us tend to lose the reality of things. No one measurement is good enough on its own, and they all have their biases (shadows) depending on the physical principle they use. Therefore, our path to enlightenment should be to strive in getting towards the sunlight, i.e. the reality of nature, without stumbling in the way. I hope that this work helps in bringing our understanding of these viruses and these techniques at the nanoscale world, closer to the reality of things, always striving for the sunlight.

General Conclusion

Let us come back to the initial question that was that pronounced at the introduction of this thesis, which was to establish a link between the mass and structure for viruses, and therefore establish mass as an intrinsic property of virus biology (cf. Figure 1).

Key findings in the context of the initial PhD objectives

Firstly, I provided an initial proof of concept on HuReVs using data analysis, showing that viruses fall within a precise mass range that is currently easily measurable by native MS methods. Additionally, I demonstrated that each virus family has a unique mass, and I gathered available data from the literature on human respiratory viruses (HuReVs) to establish a proof of concept, indicating that total virion mass could be an intrinsic feature of viral evolutionary biology. I also contributed to building a simple mathematical model that explains why this mass correlates with our observations and compared these findings with total virion mass measurements from the literature. I hypothesized that this is perhaps an emergent property of viruses evolving under specific physical and biochemical conditions. Furthermore, I demonstrated how our instrument and other contemporary or older approaches meet the necessary requirements to help build a virus mass database. I highlighted key attributes essential to achieve this goal, such as the capacity for single-particle measurements, a reasonable limit of detection, and the accessibility of these instruments in research labs and the commercial sector.

Secondly, from a methodological and instrumental perspective, I contributed to the development of an improved version (prototype #2) of the initial NEMS-MS method. And demonstrated how we optimized the NEMS-MS instrument for in-field use by reducing its size and weight and improving its limit of detection. Additionally, I conducted several analytical examinations to optimize mass characterization with this procedure.

Thirdly, I concentrated on improving the characterizations available with this instrument, aiming to establish a clear mass-to-structure relationship for its use as a robust structural biology method. The knowledge gained from working with the NEMS-MS instrument enabled me to design a multi-method analytical investigation to observe samples at different stages of the instrument and in various environments, with a particular focus on the

transition from liquid to gas phase (solution to aerosol). To achieve this, by implementing an analytical pipeline incorporating several analytical and structural biology methods, allowing evaluation of the instrument from multiple angles at critical stages. I employed this pipeline to investigate model particles (T5-CLPs), with the intention of applying this exploratory program to future vaccine development and critical quality attribute (CQA) evaluation.

Main Contributions

I believe that my work has served to put in place a first milestone regarding the mass-to-structure relationships for viruses. I believe that by using the phylobaric classification, shown here for human respiratory viruses (HuReVs), may provide a more general perspective on the evolution of all viruses. This opens a new frontier on virus research, and especially the usage of native mass spectrometry to characterize total virion mass, and helps build the envisioned virus mass database, that can be used for their detection in the future.

From a larger applicative point of view, knowing the virus mass can be used as an important information in their detection in large facilities. Starting from building a 'virus mass database', which could help simplify the implementation of non-pharmaceutical interventions (NPIs) in case of dangerous epidemics or even pandemics.

Lastly, I also believe that the analytical pipeline that I proposed, surrounding the NEMS-MS instrument, will contribute to render this MS methods into a commonly used structural biology approach, ranging from virus detection to vaccine development, making it a common procedure for structural biology in such applications.

Limitations

Of course, every work has its limitations and difficulties. I have the desire to see the phylobaric classification proposed by this work to be generalized to all viruses, and not be limited only to the HuReVs analysed here. I also would have wanted to examine experimentally some, or even all, of the HuReVs mentioned here using the NEMS-MS instrument, in order to render the proposed mass to structure relationships even more explicit. Unfortunately, this was extremely difficult due to the biosafety limitations, heavily impacting the sample supply. This last was one of the most critical elements during my thesis, which was especially aggravated by the COVID pandemic, and the large geographical distance from our virologist collaborators. In addition, some NEMS-MS instrumental limitations made it difficult to perform systemic

evaluation of the samples of interest, and often required extensive periods of optimization and troubleshooting. Other difficulties that are important to be mentioned is the instrument availability time and scheduling for all the other methods and the important optimization that was required to tailor them to my specific needs, which was explicitly highlighted in Chapter 5. Developing a portable or benchtop NEMS-MS that can be used in P2-P3 labs, which are the ones traditionally used to study viruses, will be one way of improving the system and taking care some of these bottlenecks.

Perspective and Conclusions

The future of the work proposed during my thesis can be directed in several perspectives. The first would be to collect more data from the literature in order to build larger theoretical databases, to have blueprint of the total virion mass for the experimental virus mass database. This of course should be enriched by experimental measurements combining comparable methods such as NEMS-MS or CDMS, or other high resolution native single particle MS methods. This database should be in close collaboration with the genomic databases, and hopefully to get implemented to the ICTV classification, and perhaps be integrated into WHO application protocols. Of course, the MS method development and optimization should be one of the major priorities in order to fulfil this database, hoping to have lighter, more resolutive and less costly instruments for terrain usage. This should come in parallel to using native MS in combination with other structural biology methods, in order to develop and clearly establish the mass to structure relationship proposed here. And perhaps reach even further, for example finding intrinsic functional and evolutive properties of viruses that can be related to their total virion mass.

In order to reach these biological goals, there needs to be a large development of technical means, primarily in the sample collection and sample detection departments. Improvements in sample collection needs to be done in order to adapt the high-volume air samplers, currently in use in combination to PCR[292], to achieve a combined usage with native MS instruments. This can also be achieved at the lab scale by adapting nebulisation methods, e.g. the SAWN used in our lab[278], to reach the necessary detection limit for aerosolized viral particles, e.g. in the case of influenza, the infectious dose is in the range of 2-3 × 10^3 viral particles per liter of air.[293] From the sample detection perspective, regarding the NEMS-MS, the integration of more efficient and faster detectors, such as the latest generation of optomechanical systems[294], or optimizing the particle guiding system, such as the recent

orbitrap-NEMS hybrid instrument[295], and reaching even more compact terrain adapted instruments will be the prime technical objectives for the future of this method.

Lastly, I am convinced that the analytical pipeline that has been set up here has a very promising future. Indeed, despite the multitude and the seniority of techniques that can provide in solution virus size evaluation (discussed in detail in my third chapter), the number of in-solution techniques that can provide a realistic total virion mass evaluation is scarce. Therefore, making the use of native MS techniques indispensable, and thus the understanding of the viral structural modifications during the solution-to-gas-phase transition of utmost importance. Right now there has been two main solutions to this question, at the total virion scale: using size as a proxy to mass (e.g. using GEMMA [144]) or having a comparison of mass between a solution-based method and gas phase native MS method (e.g. using MassP and native MS [225]). I believe that a larger integration of the well-known solution-based biophysical techniques, which are well accepted and quite familiar to the structural biology community, to native MS instruments, is a way to bring these last to a broader use, and thus render them into indispensable tools to the structural biologist's toolbox. That was therefore the intent of the analytical pipeline detailed during the last chapters of my thesis.

I believe that the total virion mass is an intrinsic and emergent property of the virus evolution, which should be investigated further in detail by novel native mass spectrometry methods, in order to establish the mass to structure relationship, and to prove the phylobaric classification I proposed in my thesis.

Summary of my scientific production during my thesis, *and those to come*

The following is a summary of all the articles to which I was able to contribute and for which I am a co-author. They are presented in chronological order of publication (from the newest to the oldest), with the currently unpublished works in italic and in blue.

2025 **V. Çumaku**, T. Fortin, S. Hentz and C. Masselon. "*On the 'phylobaric' classification of human respiratory viruses".* (Manuscript to be resubmitted) {in text reference number: [296]}

https://doi.org/10.21203/rs.3.rs-3395557/v1

- In this article I contributed by: emitting the phylobaric hypothesis; designing the data collection; collecting, calculating and interpreting the data; helping in building the mathematical model; writing and discussing the article drafts.

2024 Fortin T., Reynaud A., Jeudy S., Lai S.H., **Çumaku V.**, Abergel, C., Hentz, S., Masselon, C. *"Modeling nanoelectromechanical resonator signals for experimental mass measurements : a Total Variation formulation".* IEEE Access. vol. 12, pp. 187919-187946, 2024. {in text reference number: [193]}

https://doi.org/10.1109/access.2024.3506155

- In this article I contributed by: collecting the experimental data; discussing and reviewing the article drafts.

2023 Reynaud A., Trzpil W., Dartiguelongue L., **Çumaku V.**, Fortin T., Sansa M., Hentz S., Masselon C. "*Compact and modular system architecture for a nano-resonator-mass spectrometer".* Frontiers in Chemistry. 2023 Sep 7;11:1238674. {in text reference number: [165]}

https://doi.org/10.3389/fchem.2023.1238674

- In this article I contributed by: building the instrument; collecting and interpreting the data; discussing and reviewing the article drafts.

2021 Clement K., Reynaud A., Defoort M., Vysotskyi B., Fortin T., Lai S.H., **Çumaku V.**, Dominguez-Medina S., Hentz S., Masselon C. "*Requirements and attributes of nano-resonator mass spectrometry for the analysis of intact viral particles".* Analytical and Bioanalytical Chemistry. 2021

Dec;413:7147-56. {in text reference number: [184]}
https://doi.org/10.1007/s00216-021-03511-4

- In this article I contributed by: designing the data collection; interpreting the data; helping in building the mathematical model; writing, discussing and reviewing the article drafts.

2021 Fortin T., Vysotskyi B., Defoort M., Reynaud A., Lai S.H., Dominguez-Medina S., Clement K., **Çumaku V.**, Hentz S., Masselon C. "*A Nonlinear Model for Nano-Electro Mechanical Mass Sensing Signals Processing".* IEEE Sensors Journal. 2021 Aug 9; 21(19):21852-61. {in text reference number: [189]}
https://doi.org/10.1109/JSEN.2021.3103713

- In this article I contributed by: discussing, reviewing and correcting the article drafts.

Another work in which contributed by collecting and interpreting data, but in which I could not be a co-author, since these data were not included in the final published article, is the following:

2022 Lai, Szu-Hsueh, et al. "*Characterizing Nanoparticle Mass Distributions Using Charge-Independent Nanoresonator Mass Spectrometry*." The Journal of Physical Chemistry C. 126.49 (2022): 20946-20953. {in text reference number: [150]}
https://doi.org/10.1021/acs.jpcc.2c06675

7 Annexes

7.1 Chapter 1: Proposing the 'phylobaric' classification of human respiratory viruses (HuReVs).

7.1.1 Definitions and Clarifications

7.1.1.1 Defining the Triangulation (T-) number

These symmetries were first defined by the work of Caspar and Klug[5], and are classified by the their T-number. The formula for the T-number calculation is as follows:

$$T = f^2 * P = f^2 * (h^2 + h*k + k^2)$$

where **h** and **k** are nonnegative integers with no common factor and represent identical pattern forming the face; **f** is the number of times this pattern is repeated. **h** is the number of units in straight line toward the next pentagon, **k** is the number of units shifted in either side to reach the next pentagon.[5,60] A visual presentation of this concept can be seen in Figure 124.

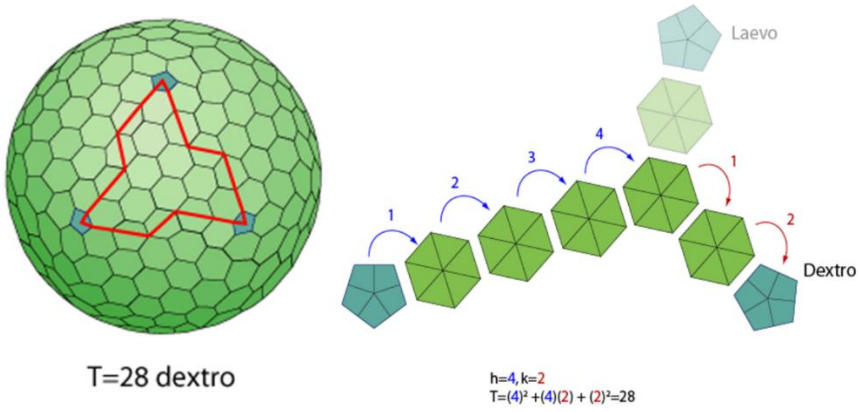


Figure 124. Example of the Triangulation (T-) number and how it is determined, based on the Caspar and Klug [5] method. The figure was reprinted from ©ViralZone

This of course has been a long studied property of viruses, where we can highlight the work from Caspar and Klug [5], in 1962, which laid the basis for the definition of the capsid T-number (T: Triangulation). But this property is still under study concerning matters such as why scaffolding proteins are needed for large viruses [297], or that even this symmetry is not entirely perfect.[298,299] This property and their size evaluation have been characterized by a variety of crystallography and electron microscopy techniques.[66,103,107]

7.1.1.2 Viruses fit in the 'mesoscale'.

The following figure, Figure 125, was adapted from Sear et al.[48], in order to fit the viruses in the original picture (shown by the red lines).

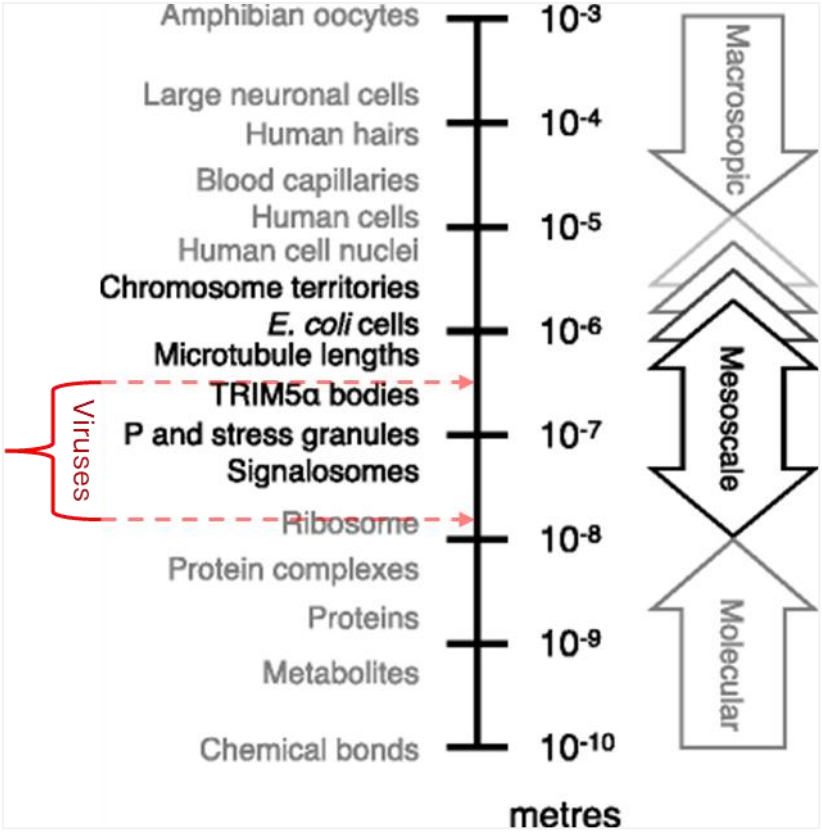


Figure 125. Mesoscale lengths in context, and how Viruses fit in it. They defined mesoscales as length-scales larger than discrete molecular complexes yet remaining intracellular. These limits are bounded below by ribosomal diameters (25 nm) as the molecular scale, and above by typical model cell diameters. For E. coli, this is approximately 1 μm, while for human cells it is of order 10 μm. Unusually large cell diameters such as neurons (100 μm) and amphibian oocytes (1 mm) mean the upper limit will depend on the cell. The figure and description were adapted from the original, Sear et al.[48], in order to fit the viruses in the figure.

The original figure was detailed and put in context during the workshop: "Life at the mesoscale: The self-organised cytoplasm and nucleoplasm", July 2014, at CECAM (Centre Européen de Calcule Atomique et Moléculaire), Lausanne, Switzerland.[48] In which they gave a definition of the mesoscale as being between around 10 nm to 10 μm.

7.1.1.3 Diagram - Data collection for the macromolecular perspective

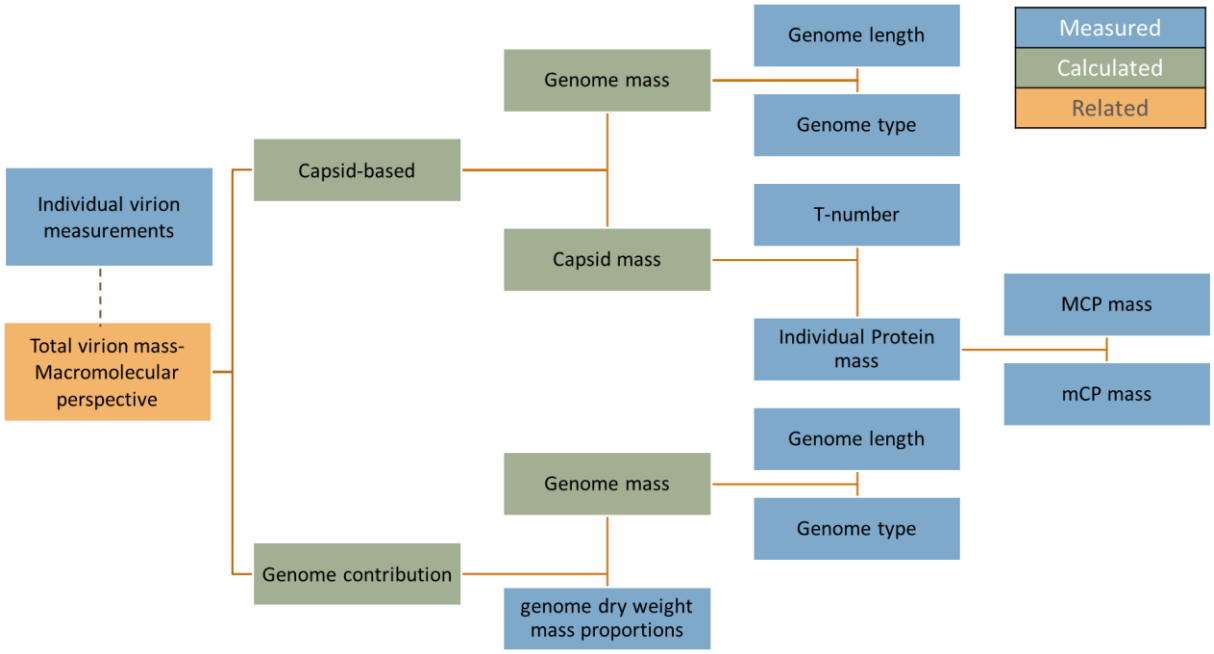


Figure 126. Diagram resuming the 'capsid-based' and the 'genome contribution' approaches in order to get the total virion mass evaluations. In blue: values reported by measurements based on the literature available; in green: calculated values by our work; in orange: related values in order to compare to the literature reported total virion mass measurements (dotted line).

7.1.1.4 *Methodologies used for virus classification at different taxonomic ranks*

Table 13. Examples of methodologies used for virus classification at different taxonomic ranks. The following table is reprinted from the original, Simmonds et al. [17] The reference numbers refer to the ones found in the original work.*

Method	Principle	Rank range	Ref
DEmARC	Analysis of distributions of pairwise evolutionary distances between nucleotide or amino acid sequences	Suborder, Family, Subfamily, Genus, Subgenus, Species	[38]
GRAViTy	Virus relationships from composite Jaccard distances between HMM profiles and genome organizational models	Order, Family, (Genus)	[46]
HSF	Identification of structural equivalence and calculation of structural distances for structure-based phylogenetics	Realm, Kingdom, Phylum, Class, Order, Family, Genus	[58,68]
PASC	Analysis of pairwise nucleotide sequence distance distributions	Genus, Species	[37]
PhageClouds	Graph database of phage genomic sequences and intergenomic distances	Subfamily, Genus, Species	[69]
SDT	Pairwise nucleotide sequence alignment and identity calculation	Species	[70]
vConTACT2	Whole-genome gene sharing profiles integrating hierarchical clustering and confidence scores	Order, Family, Subfamily, Genus	[44,45]
VICTOR	Phylogenomic method optimized to ICTV classification that reports both sequence identity- and gene content-based phylogenies along with a suggested classification; works with either nucleotide or amino acid datasets	Family, Subfamily, Genus, Species	[42]
ViPTree	Virus relationships from genomic distances based translated nucleotide scores using tBLASTx	Family, Subfamily, Genus	[49]
VirClust	Hierarchical clustering based on core protein analysis	Order, Family, Subfamily	[71]
VIRIDIC	Calculates intergenomic similarities between pairs of viral genomes based on BLASTN alignments	Family, Genus, Species	[39]

*Discussions of tools dedicated to general reconstruction of phylogeny based on multiple sequence alignments are beyond the scope of this paper (for more information about this subject, see [72,73]). A more extensive list of virus bioinformatics tools including tools for virus taxonomy can be found at https://evirusbioinfc.notion.site/evirusbioinfc/18e21bc49827484b8a2f84463cb40b8d?v=92e7eb6703be4720abf17a901bc9a947.

https://doi.org/10.1371/journal.pbio.3001922.t001

7.1.2 Relevant Tables

7.1.2.1 *Table – 'Biophysical' mass calculation method.*

Table 14. Table presenting the difference of mass between viral families, presented in alphabetical order, calculated by the biophysical method. Values were calculated as shown before, see section 1.3.1. N/A = Non-Applicable values. ***In bold*** *are the range values measured (horizontal: diameter converted to volume; vertical: estimated mass), as plotted in Figure 13. Highlighted in green, is the mean calculated value by the biophysical method, and highlighted in blue, is the mean experimental mass, as plotted in Figure 13. Abbreviations are as follows: Dia.: diameter, except for Poxviridae, where it is the volume, presented as (length * width * height); Dens.: Buoyant Density; Range $M_{biophy.}$: Range biophysical Mass; Mean M_{biophy}: Mean biophysical mass; M_{expm}: Experimentally measured Mass; Mean M_{expm}: Mean Experimental Mass; Ref: literature references.*

Virus Family	Capsid (Yes/ No)	**Dia. [nm]**	Dens. [g/cm^3]	Range $M_{biophy.}$ [MDa]	Mean $M_{biophy.}$ [MDa]	**Range $M_{expm.}$ [MDa]**	Mean $M_{expm.}$ [MDa]	Ref
Adenoviridae	Yes	**70 - 90**	1,31 - 1,36	141,68 - 312,61	227,15	**150 - 180**	165,00	49,62
Arenaviridae	No	**50 - 300**	1,14 - 1,2	44,93 - 10216,08	5130,51	**N/A**	N/A	2,64,69, 62
Bunyaviridae	No	**80 - 120**	1,16 - 1,21	187,27 - 659,28	423,28	**300 - 400**	350,00	2,64,69
Coronaviridae	No	**120 - 160**	1,15 - 1,24	626,59 - 1601,48	1114,04	**400**	400,00	2,64,69, 65
Herpesviridae	Yes	**160 - 300**	1,22 - 1,28	1575,65 - 10897,15	6236,40	**782,9 - 2700**	1741,45	49,62
Orthomyxoviridae	No	**80 - 120**	1,19	192,11 - 648,38	420,25	**250**	250,00	2,64,69, 49

Paramyxoviridae	No	**150**	1,18 - 1,2	1255,73 - 1277,01	1266,37	**500 - 660**	580,00	2,64,69,49
Parvoviridae	Yes	**21 - 26**	1,34 - 1,45	3,91 - 8,04	5,98	**5,4 - 6,2**	5,80	49,62
Picornaviridae	Yes	**28 - 30**	1,33- 1,45	9,21 - 12,34	10,78	**8 - 9**	8,50	2,64,69,49
Polyomaviridae	Yes	**40 - 45**	1,2 - 1,35	24,22 - 38,79	31,51	**25 - 26**	25,50	49,62
Poxviridae	No	**(270 *270 * 140) – (140-260 * 220 * 450)**	1,16 - 1,3	7129,42 - 20150,82	13640,12	**3000 - 5721**	4360,50	2,64,69,62,49,50,63
Reoviridae	Yes	**60 - 98**	1,36 - 1,39	92,63 - 412,51	252,57	**120 - 130**	125,00	2,64,69,49
Togaviridae	Yes	**60 - 70**	1,18 - 1,22	80,37 - 131,95	106,16	**52**	52	2,64,69,62,49

7.1.2.2 Table - 'Capsid-derived' mass calculation method.

Table 15. The calculation of the mass of individual human respiratory viruses by the 'capsid based' mass calculation method. Values were calculated as seen in section 1.4.3.1, from references seen in section 1.2.2. The viral families are presented in alphabetical order. Highlighted in blue: the calculations described in the section 1.4.1. Highlighted in green: the calculations of the capsids and nucleo-capsids, based on their T-number. The letters in the T-number refer to d:dextro (i.e. right) and l:laevo (i.e. left).[66,67] Column colours refer to the Figure 1, and in bold, the log mass value was plotted in Figure 4, in the main article.

Family	Genus	Species	Genome Type	Genome Length (nt)	Genome calc. mass (MDa)	T-number	(Nucleo) Capsid mass (MDa)	Lipid content (%)	**Total virion mass (MDa)**
Adenoviridae	Mastadenovirus	Human adenovirus C	dsDNA	35 937	21,83	25	82,39	0	**104,21**
Herpesviridae	Roseolovirus	Human herpesvirus 6B	dsDNA	162 114	98,47	16	144,27	22	**311,21**
Herpesviridae	Roseolovirus	Human herpesvirus 6A	dsDNA	159 322	96,77	16	144,98	22	**309,93**
Herpesviridae	Roseolovirus	Human herpesvirus 7	dsDNA	153080	92,98	16	146,81	22	**307,43**
Herpesviridae	Varicellovirus	Human herpesvirus 3	dsDNA	124884	75,85	16	148,78	22	**287,99**
Parvoviridae	Erythroparvovirus	Human parvovirus B19	ssDNA	5 596	1,70	1	3,73	0	**5,43**
Parvoviridae	Bocaparvovirus	Human bocavirus	ssDNA	5 299	1,61	1	4,01	0	**5,62**
Parvoviridae	Dependoparvovirus	Adeno-associated virus	ssDNA	4 679	1,42	1	3,73	0	**5,16**
Picornaviridae	Enterovirus	Poliovirus type 1	ssRNA	7440	2,38	1	5,41	0	**7,79**
Picornaviridae	Enterovirus	Enterovirus D68	ssRNA	7 390	2,37	1	5,70	0	**8,07**

Picornaviridae	Enterovirus	Human rhinovirus A	ssRNA	7 152	2,29	1	5,74	0	**8,03**
Picornaviridae	Enterovirus	Human rhinovirus B	ssRNA	6 900	2,21	1	5,71	0	**7,93**
Polyomaviridae	Polyomavirus	WU Polyomavirus	dsDNA	5 228	3,18	7𝒹	19,66	0	**22,83**
Polyomaviridae	Polyomavirus	BK polyomavirus	dsDNA	5153	3,13	7𝒹	19,12	0	**22,25**
Reoviridae	Orthoreovirus	Mammalian orthoreovirus	dsRNA	23 500	15,06	13ℓ	107,77	0	**122,84**
Togaviridae	Rubivirus	Rubella virus	ssRNA	9 757	3,13	4	27,54	30	**43,81**

7.1.2.3 Table - 'Genome contribution' mass calculation method.

Table 16. The calculation of the mass of individual human respiratory viruses was done using the 'genome contribution' mass calculation method. Values were calculated as seen in section 1.4.3.2, from references seen in section 1.2.2. The viral families are presented in alphabetical order. Species (**): (< 10) reported cases of human-to-human transmission, as reported by Geoghegan et al..[53] Values highlighted in blue are from the calculations described in the section 1.4.1. N/A= Non-Applicable value. Column colours refer to the Figure 1, and in bold, the log mass value was plotted in Figure 4, in the main article.

Family	Genus	Species	Genome Type	Genome Length (nt)	Genome mass (MDa)	Genome contribution (%)	Lipid content (%)	**Total virion mass (MDa)**
Arenaviridae	Arenavirus	Lujo virus**	ssRNA	10352	3,32	2,00	20	**165,90**
Arenaviridae	Arenavirus	Guanarito virus	ssRNA	10424	3,34	2,00	20	**167,05**
Arenaviridae	Arenavirus	Machupo virus	ssRNA	10562	3,39	2,00	20	**169,26**
Arenaviridae	Arenavirus	Lassa virus	ssRNA	10824	3,47	2,00	20	**173,46**
Bunyavirales	Orthohantavirus	Hantaan virus	ssRNA	11 845	3,80	1,00-2,0	20-30	**253,10**
Bunyaviridae	Hantavirus	Black creek canal virus**	ssRNA	12000	3,85	1,00-2,00	20-30	**256,41**
Coronaviridae	Alphacoronaviruses	Human coronavirus 229E	ssRNA	27317	8,76	N/A	yes	**N/A**
Coronaviridae	Betacoronavirus	Human coronavirus NL63	ssRNA	27410	8,79	N/A	yes	**N/A**
Coronaviridae	Betacoronavirus	SARS coronavirus	ssRNA	29751	9,54	N/A	yes	**N/A**
Coronaviridae	Betacoronavirus	SARS coronavirus 2	ssRNA	29903	9,58	N/A	yes	**N/A**
Coronaviridae	Betacoronavirus	Human coronavirus HKU1	ssRNA	29934	9,59	N/A	yes	**N/A**
Coronaviridae	Betacoronavirus	Middle East respiratory syndrome	ssRNA	30063	9,64	N/A	yes	**N/A**
Coronaviridae	Betacoronavirus	Human coronavirus OC43	ssRNA	30578	9,80	N/A	yes	**N/A**
Orthomyxoviridae	Influenzavirus A	Influenza A H5N1 virus	ssRNA	13500	4,33	6,76	48,00	**64,01**

Orthomyxoviridae	Influenzavirus A	Influenza A H7N9 virus	ssRNA	13500	4,33	6,76	48,00	**64,01**
Orthomyxoviridae	Influenzavirus A	Influenza A H3N2 virus	ssRNA	13588	4,36	6,76	48,00	**64,42**
Orthomyxoviridae	Influenzavirus B	Influenza B virus	ssRNA	14548	4,66	6,76	48,00	**68,98**
Orthomyxoviridae	Influenzavirus C	Influenza C virus	ssRNA	14600	4,68	6,76	48,00	**69,22**
Paramyxoviridae	Pneumovirus	Human respiratory syncytial virus	ssRNA	15128	4,85	0,50	20-25	**969,74**
Paramyxoviridae	Rubulavirus	Human parainfluenza virus 5	ssRNA	15246	4,89	0,50	20-25	**977,30**
Paramyxoviridae	Rubulavirus	Mumps virus	ssRNA	15384	4,93	0,50	20-25	**986,15**
Paramyxoviridae	Respirovirus	Human parainfluenza virus 3	ssRNA	15405	4,94	0,50	20-25	**987,49**
Paramyxoviridae	Respirovirus	Human parainfluenza virus 1	ssRNA	15504	4,97	0,50	20-25	**993,84**
Paramyxoviridae	Rubulavirus	Human parainfluenza virus 2	ssRNA	15654	5,02	0,50	20-25	**1003,45**
Paramyxoviridae	Morbillivirus	Measles virus	ssRNA	15894	5,09	0,50	20-25	**1018,84**
Paramyxoviridae	Rubulavirus	Human parainfluenza virus 4	ssRNA	17091	5,48	0,50	20-25	**1095,56**
Poxviridae	Orthopoxvirus	Variola virus	dsDNA	185578	112,72	3	4	**3757,34**

7.1.3 Describing HuReVs as aerosolised particles

Like any aerosolized particles, HuReVs' propensity to remain airborne over long periods and distances while maintaining the ability to deposit deep within the host respiratory tract is directly linked to their dimensions.[55,300] Small particles (diameter ≤ ~20 nm) tend to closely mimic carrier gas flows, which may preclude their deposition. In contrast, larger particles (diameter >> ~300 nm) undergo impaction, sedimentation, and interception. As a result, diffusion and convection become less significant in their dissemination.[300] Interestingly, particles with diameters between these two extremes, whatever their composition, tend to remain in the low and middle troposphere for long periods.[75] The following figure, Figure 127, describes the forces and physical constrains that an aerosol faces depending on its size, presented as its radius, and relating this to its residence time, in second (left y axis) and in days (right y axis).

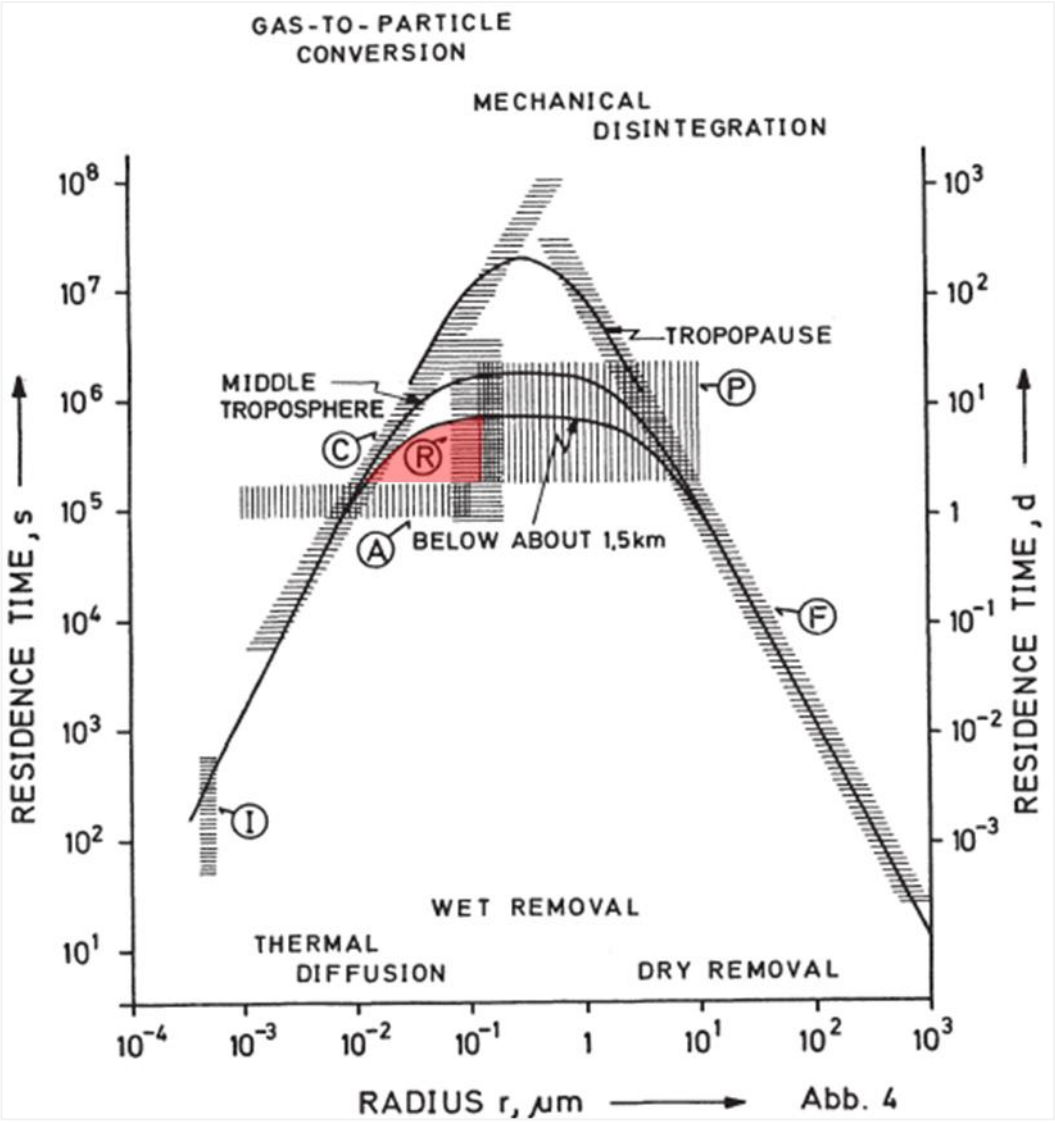


Figure 127. Residence time of tropospheric aerosols as function of particle size. The hatched areas represent published estimates: I - small ions, A - Aitken particles, C - coagulation (thermal diffusion), R - radioactivity, P - precipitation, F – sedimentation according to Jaenicke (1978). The curves represent various regions in the troposphere according to Jaenicke (1980). They are not in disagreement with Pruppacher et al (1978). The major production mechanisms and removal processes are indicated in the most important particle size ranges. Figure and description were adapted from: R. Jaenicke (1982) [75]. The area highlighted in red was added by me in order to show the expected area where HuReVs can be found.

This figure oversees all types of aerosol formation, and based on experimental data, defines areas that correlate the size of the aerosol to the part of the troposphere that it can be found on. The most obvious regions to observe are C - coagulation (thermal diffusion) and F – sedimentation (dry removal), outlining the tropopause. The upper bound of the middle troposphere is defined by the presence of P – precipitation. The split between the middle and lower (below about 1.5 km) region is defined by the presence of A - Aitken particles [301]. The Aitken nuclei, found in the range $10 < D < 80$ nm, arise from ambient-temperature gas-to-particle conversion, in which hot, supersaturated vapours form, and subsequently undergo condensation.[302]

HuReVs fit within this intermediate size-range, of around 20 to 300 nm,[31] area highlighted in red in Figure 127 and Figure 128. Which allows us to make the claim that if uninterrupted, these viruses can stay in the low to middle troposphere from roughly one day to a week. The effects of R- Radiation can be seen as an additional limiting factor, but this is hard to determine precisely, since it is an exposure dependent evaluation. An explanation of the mechanism of these particle formation can be found in Figure 128.

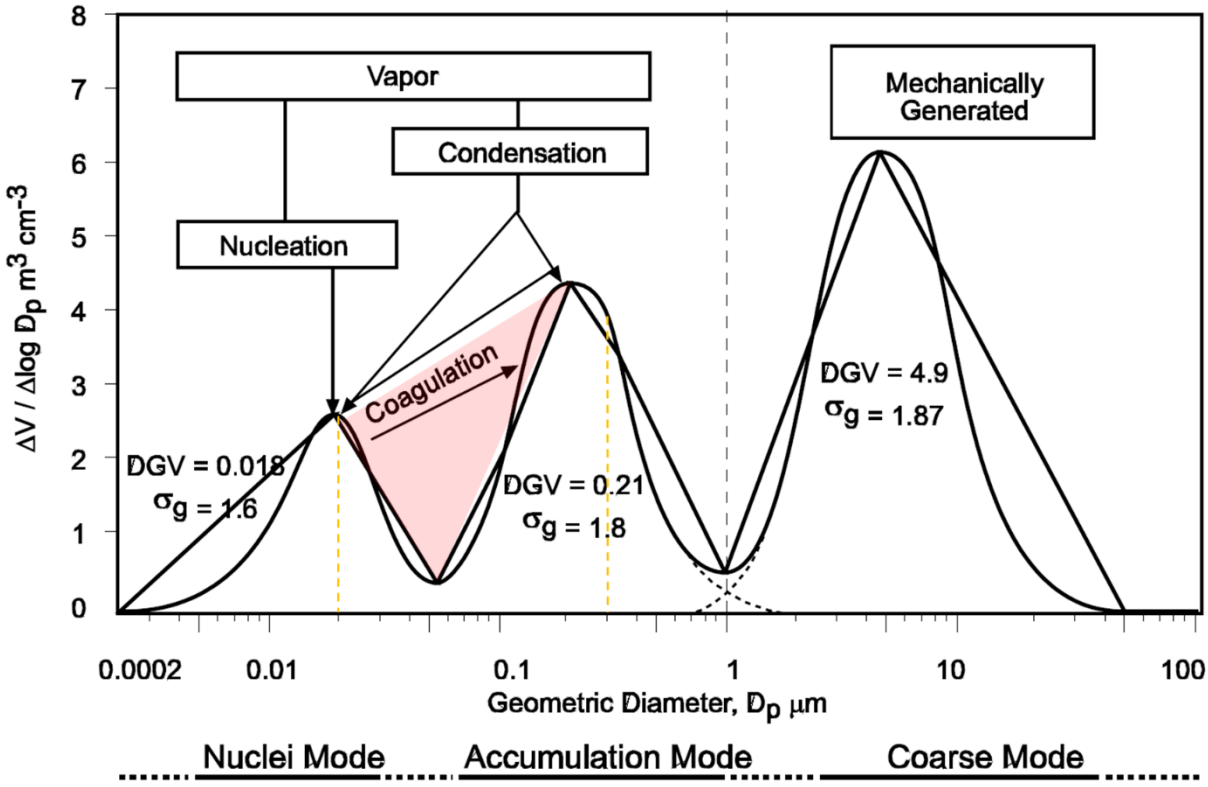


Figure 128. Typical ambient air particulate matter size distributions. Figure adapted from the Air Pollution Training Institute (APTI) document, January 2000. Original source: United States Environmental Protection Agency, Criteria Document, 1997. Legend: DGV: geometric volume mean diameter; σ_g: geometric standard deviation. Look in the text for the definition of the term Volume distribution [dV/dlog(Dp)]. The area highlighted in red was added by me in order to show the expected area where HuReVs can be found. The orange dashed line delimit the diameter region 20-300 nm.

Figure 17 shows how the volume distribution is related to the geometric diameter of the particles in atmosphere. From left to right we go from the atomic region of particles when they start nucleating, reaching a maximum of 0.018 µm = 18 nm for the first peak. We can also see that the second peak is in the region of accumulation and has a maxima at 0.21 µm = 210 nm. Between the first and the second region we have the formation of particles by condensation, as an intermediate region, which coincidentally corresponding to where most viruses particles are found, similarly to Figure 127 (highlighted in red). Above this region, around 1µm, the sedimentation forces make the particles go down, and the third peak, at 4.9 µm, is mechanically generated (e.g. sand particles).

We can thus conclude that this size range, i.e. 20-300 nm, is perhaps a result of evolutionary physical constraints, and perhaps showing the interaction of biophysical constraints at the macroscale. This hypothesis would require a more fundamental analysis, by observing the aforementioned multitude of parameters that affect viruses in aerosols, but that was out of the initial the scope of this work.

7.1.3.1 Note: Definition of the term Volume distribution [dV/dlog(Dp)]

The following definition was taken from the work of Erkki Lamminen[303] :

Volume distribution gives the total volume of all particles in each size range. It is formed by multiplying the current distribution (C_i), by the conversion vector, and by a vector formed from the volumes of spheres having diameter equal to midpoint values (D_i) of each stage.

$$[dV/dlog(D_p)] = [C_i\ ] * [(1/X_i(D_i))] * [1/6 * \pi * D_i^3] * [(1/dlog(D_p))]$$

7.2 Chapter 2: Methods for virion mass evaluation and the place of NEMS-MS

7.2.1 Compiling the data for measurements reported in the available literature.

Table 17 compiles the data for measurements values reported in the available literature, using different measurement methods. Method's group abbreviations (in alphabetical order): EM: Electron Microscopy; MS: Mass Spectrometry. N.A.: Non-Applicable; Not Found: not specified in the original article cited. Abbreviations on methods used are described in Glossary.

Table 17. Viral families, ranked in alphabetical order, as determined by various experimental methods. Data is plotted in Figure 19.

Count	Family	Genus	Virus names	Mass ± St.dev (Mass range) [MDa]	Method used	Method group	Ref
1	Adenoviridae	Mastadenovirus	Adenovirus-12 (Nucleocapsid)	189,096	QTEM	EM	[304]
2	Adenoviridae	Mastadenovirus	adenovirus type 2	185,000	neutron diffraction	Neutrons	[76]
3	Adenoviridae	Mastadenovirus	Adenovirus	183,000	N.A.	Not Found	[49]
4	Adenoviridae	Mastadenovirus	adenovirus	175,245	Sedimentation	Sedimentation	[304]
5	Adenoviridae	Mastadenovirus	adenovirus type 2	172,932	Sedimentation	Sedimentation	[305]
6	Adenoviridae	Mastadenovirus	adenovirus type 5	157,000 ± 10	STEM	EM	[76]
7	Adenoviridae	Mastadenovirus	adenovirus 2	120,440	Sedimentation	Sedimentation	[74]
8	Bunyaviridae	N.A.	(maximum-value)	400,00	N.A.	Not-Found	[64]
9	Bunyaviridae	N.A.	(minimum-value)	300,00	N.A.	Not-Found	[64]
10	Coronaviridae	Betacoronavirus	Sars-Cov 2	221,500	SP-ICP-MS	MS	[86]
11	Coronaviridae	Betacoronavirus	strain OC 43 human coronavirus	112,000 ± 5	Sedimentation	Sedimentation	[85]
12	Coronaviridae	Betacoronavirus	Sars-Cov 2	461,500 (454 - 469)	NEMS-MS	MS	[87]
13	Herpesviriridae	Simplexvirus	Herpes simplex virus (HSV)	2700,000	N.A.	Not Found	[49]
14	Herpesviriridae	Simplexvirus	Herpes simplex virus (HSV)_enveloped nucleocapsid	782,878	QTEM	EM	[81]
15	Herpesviriridae	Simplexvirus	Herpes simplex virus (HSV)_full naked capsid	421,55	QTEM	EM	[81]

16	Herpesviriridae	Simplexvirus	Herpes simplex virus (HSV)_empty naked capsid	301,11	QTEM	EM	[81]
17	Herpesviriridae	Simplexvirus	Herpes simplex virus (HSV)_core	120,44	QTEM	EM	[81]
18	Orthomyxoviridae	N.A.	influenza virus	360,000	Sedimentation	Sedimentation	[4]
19	Orthomyxoviridae	N.A.	Myxovirus (influenza)	250,000	N.A.	Not Found	[49]
20	Orthomyxoviridae	Betainfluenza virus	B/Hong Kong	215,000	SANS	Neutrons	[122]
21	Orthomyxoviridae	Betainfluenza virus	B/Singapore/222/79	182,000 ± 19	STEM	EM	[306]
22	Orthomyxoviridae	Betainfluenza virus	B/Hong Kong/8/73	178,000 ± 22	STEM	EM	[306]
23	Orthomyxoviridae	Influenzavirus A	Influenza A-PR 8	198,130	QTEM	EM	[304]
24	Orthomyxoviridae	N.A.	influenza	180,000	SANS	Neutrons	[4]
25	Orthomyxoviridae	Influenzavirus A	X49/H3N2 (a recombinant strain of A/England/864/75/H3 N2 and A/PR/8/34/HIN1)	161,000 ± 17	STEM	EM	[306]
26	Orthomyxoviridae	Influenzavirus A	influenza A/WSN	147,000	TEM	EM	[306]
27	Paramyxoviridae	Paramyxovirus	Newcastle Disease (NDV - Avian Paramyxovirus-1)	660,000	N.A.	Not Found	[49]
28	Paramyxoviridae	N.A.	N.A.	500+	N.A.	Not Found	[64]
29	Parvoviridae	Dependoparvovirus	AAV8 Capsid (recombinant)	3,819± 16	CD-MS	MS	[307]
30	Parvoviridae	Dependoparvovirus	AAV9 Capsid (recombinant) (full particle)	4,930±0.01	SEC	Chromatography	[121]
31	Parvoviridae	Dependoparvovirus	AAV9 Capsid (recombinant) (empty particle)	3,770±0.01	SEC	Chromatography	[121]
32	Parvoviridae	Bocaparvovirus	human bocavirus (HBoV)	> 4,000	SEC	Chromatography	[308]
33	Parvoviridae	Dependoparvovirus	AAV8 Capsid (recombinant) (empty particle)	3,800± 0,03	SEC	Chromatography	[121,309]
34	Picornaviridae	Enterovirus	POLIO >> Mahoney strain	8,500	crystallography	X-ray	[310]
35	Picornaviridae	Enterovirus	Entero	8,500	Sedimentation	Sedimentation	[49]

36	Picornaviridae	Enterovirus	rhinovirus 1A	8,400	Sedimentation	Sedimentation	[4]
37	Picornaviridae	N.A.	Human rhino	8,400	N.A.	Not Found	[49]
38	Picornaviridae	Enterovirus	Human rhino	8,400	Sedimentation	Sedimentation	[49]
39	Picornaviridae	Enterovirus	poliovirus strain Type 1 (Mahoney) (filled particle)	8,250	CD-MS	MS	[142]
40	Picornaviridae	Enterovirus	poliovirus weighted average (filled particle)	8,230	CD-MS	MS	[142]
41	Picornaviridae	Enterovirus	poliovirus strain Type 3 (Saukett) (filled particle)	8,220	CD-MS	MS	[142]
42	Picornaviridae	Enterovirus	poliovirus strain Type 2 (MEF1) (filled particle)	8,210	CD-MS	MS	[142]
43	Picornaviridae	Enterovirus	Rhinovirus	7,100	Sedimentation	Sedimentation	[4]
44	Picornaviridae	Enterovirus	POLIO >> Mahoney	6,800	Sedimentation	Sedimentation	[4]
45	Picornaviridae	Enterovirus	POLIO >> MEF 1	6,800	Sedimentation	Sedimentation	[4]
46	Picornaviridae	Enterovirus	POLIO >> Saukett	6,400	Sedimentation	Sedimentation	[4]
47	Picornaviridae	Enterovirus	poliovirus strain Type 1 (Mahoney) (empty particle)	5,860	CD-MS	MS	[142]
48	Picornaviridae	Enterovirus	poliovirus weighted average (empty particle)	5,840	CD-MS	MS	[142]
49	Picornaviridae	Enterovirus	poliovirus strain Type 3 (Saukett) (empty particle)	5,830	CD-MS	MS	[142]
50	Picornaviridae	Enterovirus	poliovirus strain Type 2 (MEF1) (empty particle)	5,810	CD-MS	MS	[142]
51	Picornaviridae	Enterovirus	POLIO >> type 1/type 2 > TOTAL	5,800	crystallography	X-ray	[4]
52	Polyomaviridae	N.A.	Polyoma virus	26,000	N.A.	Not Found	[49]
53	Polyomaviridae	N.A.	polyomavirus	25,000	Sedimentation	Sedimentation	[4]
54	Poxviridae	Orthopoxvirus	vaccinia virus	5721,000	NEMS-MS	MS	[63]
55	Poxviridae	Orthopoxvirus	vaccinia strains >> NIH	3900,000	QTEM	EM	[4]
56	Poxviridae	Orthopoxvirus	vaccinia strains >> Lea (7N)	3700,000	QTEM	EM	[4]
57	Poxviridae	Orthopoxvirus	vaccinia strains >> WR	3600,000	QTEM	EM	[4]
58	Poxviridae	Orthopoxvirus	vaccinia strains >> Hamburg	3200,000	QTEM	EM	[4]
59	Poxviridae	N.A.	Poxvirus (vaccinia)	3200,000	N.A.	Not Found	[49]
60	Poxviridae	Orthopoxvirus	vaccinia virus	3914,000 (3.011 - 4.818)	QTEM	EM	[50]
61	Reoviridae	N.A.	Reovirus (dry)	130,000	N.A.	Not Found	[49]

62	Reoviridae	N.A.	Reovirus (hydrated)	130,000	N.A.	Not Found	[49]
63	Reoviridae	Orthoreovirus	Reovirus type 3	70,000	Sedimentation	Sedimentation	[49]
64	Reoviridae	Orthoreovirus	reovirus	70,000	Sedimentation	Sedimentation	[4]
65	Reoviridae	Rotavirus	(5 x strains) human-bovine rotavirus_DLPs	61,060± 0,58	CD-MS	MS	[142]
66	Reoviridae	Rotavirus	(5 x) human-bovine rotavirus strains_ empty DLPs	49,640	CD-MS	MS	[142]
67	Reoviridae	Orthoreovirus	Reovirus type 3	129,500± 5 (127-131)	Sedimentation	Sedimentation	[49]
68	Togaviridae	Rubella	Rubella virus (RUBV)	52,000	N.A.	Not Found	[62]
69	Togaviridae	N.A.	N.A.	52,000	N.A.	Not Found	[64]

7.2.2 NEMS related tables

Table 18. Lengths of every resonator beam of the array, with the NEMS index presented in blue. Table reprinted from Reynaud et al.[165]

NEMS index [-]	**NEMS length** [µm]	**NEMS index** [-]	**NEMS length** [µm]	**NEMS index** [-]	**NEMS length** [µm]	**NEMS index** [-]	**NEMS length** [µm]
1	10	6	9.235	11	8.58	16	8.01
2	9.835	7	9.095	12	8.46	17	7.905
3	9.68	8	8.96	13	8.345	18	7.805
4	9.53	9	8.83	14	8.23	19	7.705
5	9.38	10	8.705	15	8.12	20	7.610

Table 19. Mass resolution of every NEMS of the array, index presented in blue. The mass resolution is represented as (mode#1 / mode #2). Table reprinted from Reynaud et al.[165]

NEMS index [-]	**Mass resolution** [MDa]	**NEMS index** [-]	**Mass resolution** [MDa]	**NEMS index** [-]	**Mass resolution** [MDa]	**NEMS index** [-]	**Mass resolution** [MDa]
1	0.16/0.18	6	0.15/0.16	11	0.14/0.15	16	0.13/0.14
2	0.16/0.17	7	0.14/0.16	12	0.14/0.15	17	0.13/0.14
3	0.16/0.17	8	0.14/0.16	13	0.13/0.15	18	0.13/0.14
4	0.15/0.17	9	0.14/0.16	14	0.13/0.15	19	0.12/0.14
5	0.15/0.17	10	0.14/0.15	15	0.13/0.14	20	0.12/0.14

7.3 Chapter 3: Developments on the NEMS-MS instrumentation as an Analytical Method

7.3.1 Supplementary (NEMS-MS bench #1)

7.3.1.1 The GNPs mass evaluations (15 nm & 30 nm)

The mass-to-size conversion, using the density of gold (19300 kg m^{-3}) and assuming perfect sphericity, we built the following table, Table 20.

Table 20. NEMS-MS-derived diameter and mass compared to supplier's specifications (BBI). All mass-to-size conversion were done using the density of gold (19300 kg m^{-3}), and assuming perfect sphericity. TEM: Transmission Electron Microscopy. Table was reprinted from Lai et al.[150]

BBI specifications (TEM)		**NEMS-MS**		**relative differences**	
measured mean diameter ± SD (nm)	**computed mean mass (MDa)**	**computed mean diameter ± SD (nm)**	**measured mean mass (MDa)**	**Diameter (%)**	**Mass (%)**
14.2 ± 0.9	17.4	14.4 ± 1.5	18.0	1.4	2.9
29.2 ± 2.1	151.7	29.6 ± 1.6	157.0	1.4	3.5

7.3.1.2 More temperature points (20 nm GNPs)

Therefore, I decided to investigate further for more intermediate temperature values and see if any trends emerged. The investigated temperatures were: from 140 °C to 250 °C, for ten different values, seen in Figure 129.

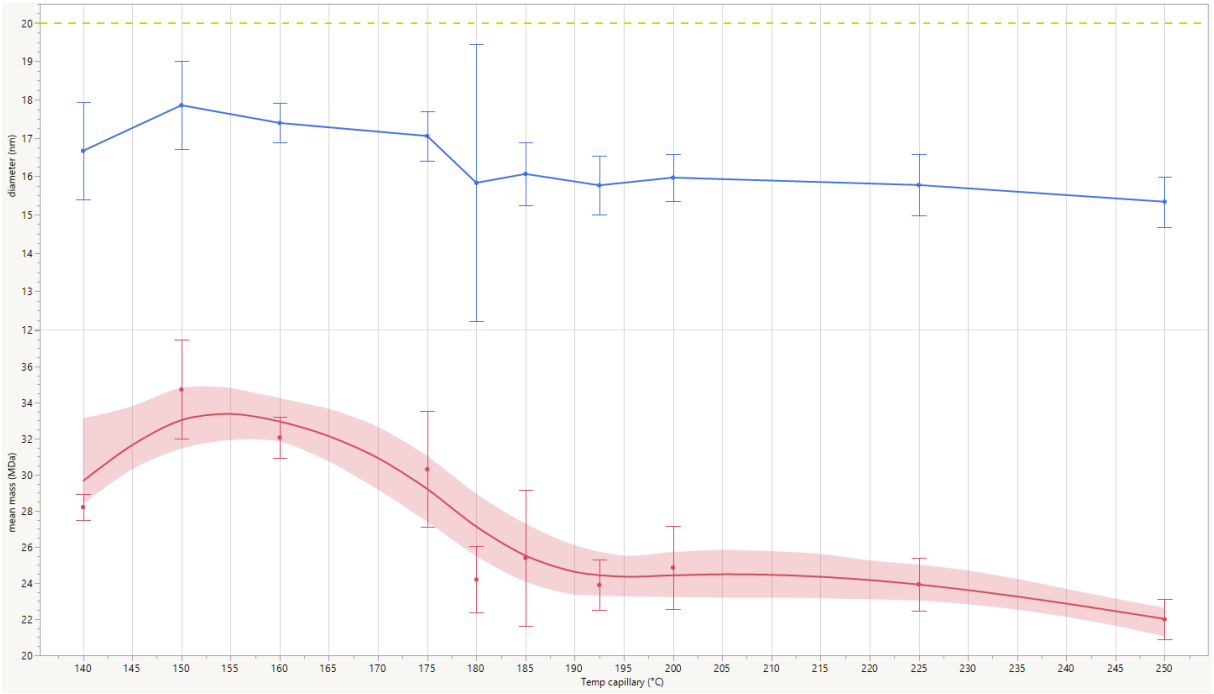


Figure 129. Size and mass relationship to temperature evaluation. Dots are the average value and the standard deviations are presented as vertical bars. The expected size is presented in dashed olive-green colour at 20 nm. Size fit: moving mean; Mass fit: Spline (confidence interval =90 %).

The first thing that we can see is that for none of the temperature values examined in the diameter of the particles (solid blue line) graph never reach the expected 20 nm supposed value, marked in dashed olive-green on the top of the figure. In addition, we can clearly see a significant difference in mass (solid red line) between the points in the lower temperature range, e.g. 150 °C, and the higher temperature range, e.g. 180°C. Furthermore, the standard deviation gets narrower as the temperature gets higher, indicating material removal, that being either the buffer or the particle itself losing atoms form its surface. We can also see that the point at 180 °C presents a fairly narrow distribution in terms of mass, but a much larger one in terms of size, which can be due the fact that its uncertainty is significantly increased when cubed in the mass to size conversion.

In order to check the effects of desolvation, we also tried to evaluate a point at an even lower temperature, i.e. at 140 °C. This proved to hard to execute as the event rate was low and the particle mass achieved was lower than that of 150 °C, seemingly contradicting the desolvation hypothesis. More recent work indicates that a potential reason for this might be that the aerodynamic lens particle guiding is affected by the temperature of the capillary, given that it affects the temperature of the air entering the bench. Thus, one can imagine that the mass evaluation performed here is of a subgroup of the real population, found on the extremities of the lens deposition. This of course would require its own experimental design by precisely

moving the NEMS to the centre of a particle deposition, as a function of the changing temperature, but this was off the scope of the work presented here.

7.3.1.3 Bacteriophage T5 mass spectra: statistics

The following tables refer to Figure 49, in order to detail the necessary statistics. The experimental statistics are reported for every flow rate in Table 21.

Table 21. Experimental parameters and data statistics for every flowrate. Table was reprinted from Clement et al.[149]

Time range (*sec*)	**Flow rate (μl/hr)**	**ESI voltage (*kV*)**	**Number of events (*count*)**	**Average mass (*MDa*)**
0 – 840	10	2.2	30	30.98
840 – 1680	20	2.5	22	30.87
1680 – 2520	30	3.0	19	36.29
2520 – 3360	40	3.2	28	35.32
3360 – 4200	50	3.4	10	38.73

A non-linear least squares fit is performed with a Gaussian function. This results in the optimization of the Gaussian curve parameters (average mass and standard deviation) that are reported in the Table 22. The library SciPy was used in this case.

Table 22. Data statistics depending on the flowrate. Table was reprinted from Clement et al.[149]

Data process	**Metrics**	**Low flow rate**	**High flow rate**
Statistics	Average mass	30.9	35.7
	Standard deviation	5.8	14.3
Least squares	Average mass	29.3	32.7
	Standard deviation	2.1	2.9
Number of events		52	47

7.3.2 Supplementary (mini NEMS-MS bench - prototype #2)

7.3.2.1 NEMS quality factor

Table 23. Length, resonance frequency and quality factors for 3 nano resonators. Table was reprinted from Reynaud et al.[165]

NEMS ID	Length [µm]	Mode 1		Mode 2	
		f [MHz]	Q [-]	f [MHz]	Q [-]
1	9.04	26.75	4000	73.47	2200
10	8.07	34.80	3600	95.10	1900
20	7.61	44.45	3000	122.00	1700

7.3.2.2 The GNPs mass evaluations (20 nm, 30 nm & 40 nm)

Table 24. Gaussian fit parameters (mean and dispersion) compared with expected sample size, according to manufacturer, in parentheses. N/A: Non applicable. Table was reprinted from Reynaud et al.[165]

Sample (manufacturer)	**Mass**			**Diameter**		
	NEMS-MS $\bar{m}$ (σ_m) [MDa]	**SEM** $\bar{m}$ [MDa]	**Relative difference** [%]	**NEMS-MS** $\overline{d_p}$ (σ_{d_p}) [nm]	**SEM** $\overline{d_p}$ (σ_{d_p}) [nm]	**Relative difference** [%]
20 nm (BBI)	52.2 (9.2)	46	11.9	20.5 (1.2)	19.6 (< 1.6)	4.4
30 nm (Sigma-Aldrich)	147.1 (88.7)	164	11.5	29.9 (4.3)	30 (2)	0.3
40 nm (BBI)	285.3 (49.5)	308	8.0	36.2 (2.4)	37 – 42 (N/A)	2.2 – 16

7.4 Chapters 4 & 5: Materials and Methods

7.4.1 Materials

7.4.1.1 T5-CLPs

7.4.1.1.1 Protocol - The production of T5 phCLPs

In order to perform the production of the T5-CLPs, we relied on the experience of our collaborator Dr Pascale Boulanger, at the I2BC lab (Institute for Integrative Biology of the Cell, 91190 Gif-Sur-Yvette, FRANCE). [link] This objective was met with a number of delays; the major ones being due to the two confinement periods caused by the COVID-19 sanitary crisis during the beginning of my thesis, in 2020-2021 period. Nevertheless, I had the possibility to produce my own first batch of phage-derived CLPs (phCLPs), which I will describe hereafter.

The following is a description of the procedure that I was able to assist on in order to perform the growth of the empty non-decorated phage derived CLPs, later named as Cesar (Figure *58*).

The bacterial strain *Escherichia coli F* was grown (at 37 °C) in LB (Lysogeny Broth) medium, complemented with 1 mM of $MgCl_2$ and 1 mM of $CaCl_2$. This bacterial strain is genetically modified to over-express the receptor FhuA, thus facilitating the cell lysis.[215] Salts were also added in order to improve bacterial lysis. A small proportion of this saturated batch was reset in a fresh LB medium from an OD_{600} (Optical Density at 600 nm wavelength) ≈ 0.1 and left to grow until they reach an OD_{600} ≈ 0.5, yielding a bacterial concentration of ~ $5*10^8$ bacteria/mL. At which point, they were infected with a multiplicity of infection (MOI) ≈ 5, with the mutant bacteriophage *T5amN5-TerL*. An MOI = 5 means that the phage/bacteria ratio was = 5:1. This *T5* mutant had been rendered incapable to perform genome packaging properly, thus resulting in the production of empty capsids.[211] After cell lysis occurred, following replication of the bacteriophage, the OD dropped to OD_{600} ≈ 0.1. In order to pellet out the bacterial cell debris and inhibit the expansion of the phage pro-capsids at this stage, 0.5 M of NaCl was added to the culture before centrifugation (8 000 rpm). To incubate and recover the solution of T5 capsids, an overnight precipitation, at 4 °C, in 7.5 % PEG 6000 (polyethylene glycol), was performed. The solution was then centrifuged and the pellet resuspended in phage buffer {10 mM of Tris buffer (pH 7.4); 100 mM of NaCl; 1 mM of $CaCl_2$; 1 mM of $MgCl_2$}.

The first purification step consisted in injecting the resuspended solution on top of a glycerol gradient (10 → 40 %), which also contained 100 mM of NaCl, followed by ultra-

centrifugation (21 000 rpm, 2 hr.). To test for expanded capsids, small portions of the previously extracted fractions were separated overnight by agarose (1.5 %) gel electrophoresis. The revelation of the fragment bands was done using Coomassie blue G-250 (BioRad®). This allowed us to keep all fractions of interest, meaning the ones that presented expanded capsids, and also the not yet expanded capsids.

The second purification step consisted in separating the selected fractions on a HiTrap® Q anion exchange chromatography column (GE Healthcare®) [Column Volume (CV) = 5 mL]. It was equilibrated with 10 CVs of 'injection' buffer **A** {25 mM Tris (pH 8.0); 200 mM NaCl}, and then with 5 CVs of 'elution' buffer **B** {25 mM Tris (pH 8.0); 1 M NaCl}. The pH of both these buffers was greater than the isoelectric point (pI) of the *T5* pro-capsids, thus allowing the molecule's anions to attach to the resin of the column. The previously collected fractions were diluted in buffer **A**, to avoid clogging by glycerol and to reduce the NaCl concentration. Then the elution buffer **B** was injected with a linearly increasing NaCl concertation, for 10 CVs. The last chromatography step was repeated in order to recover a maximum of the sample from the column. To test the efficiency of these purifications, an agarose (1.5 %) gel electrophoresis was performed on designated samples of interest, after considering the real-time chromatogram recordings.

A third purification step was performed on the resulting samples using the detergent Lauryldimethylamine oxide (LDAO), to digest lipopolysaccharides (LPS) from the bacterial cell walls. If not removed, the LPS can cause toxic choc if the capsids are to be used in vaccine preparations for instance.[311] A new injection buffer, similar to buffer **A**, containing [LDAO] = 0.1 %, was prepared, called buffer **A***. We analysed the samples of interest using the following injection program:

a) Injection with buffer **A***;

b) Wash with buffer **A***, for 10 CVs;

c) Wash with buffer **A**, for 10 CVs;

d) Elution with buffer **B** : gradient 0-50 %, for 5 CVs; then 50-100 %, for 4 CVs.

The last chromatography step was repeated in order to recover a maximum of the sample from the column. To test the efficiency of these purifications, an agarose (1.5 %) gel electrophoresis was performed on designated samples of interest, after considering the real-time chromatogram recordings. The resulting high A_{260}/A_{280} ratio data showed that a potential phage contamination was present. (A_{260}: Absorbance of genome, at 260 nm; A_{280}: Absorbance of proteins, at 280 nm).

To quantify the concentration of active/infectious phages in the pooled samples, a series of base 10 dilutions from the stock solution were performed using the phage buffer. A small fraction of each of these dilutions was then applied on Petri dishes that contained a mixture of LB-Agar and a pre-culture of the *E.coli Cr63* strain, grown overnight, at 42 °C. This bacterial strain possess a genetic mutation which supresses the phage mutation *amN5-TerL*, thus allowing the mutant T5 phages to perform their full viral cycle.[312] In order to preserve the samples, and to remove the NaCl, we dialysed them overnight, at 4 °C, against 25 mM of HEPES [*4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid*] buffer (pH 7.01). After the dialysis, capsid expansion was tested by agarose (1.2 %) gel electrophoresis. The agarose electrophoresis showed that all pools presented expanded capsids, but this technique does not allow to resolve between the infectious phages and the expanded pro-capsids.

In addition, the LB-Agar dishes previously prepared indicated that infectious phages were present in our sample at concentrations up to ~10^5 phages/µL, meaning that our sample needed further purification. Hence, the last purification step was to inject all the pooled batches into a gradient of CsCl salt, followed by ultra-centrifugation (25 000 rpm, 4 °C, 2 hr.). We know from prior experiments that the empty pro-capsid have a density ~ 1.3-1.35 g/cm^3, and the infectious phages have a density ~ 1.5 g/cm^3. The empty and expanded capsids were thus recovered from the CsCl solution through dialysis in 25 mM HEPES buffer. Analysis by (NanoDrop®) indicated an A_{260}/A_{280} ratio ~ 0.83-0.89, indicating a good purification and a clear reduction of viral genome. Expanded empty capsids were than stored in phage buffer, at 4 °C, at a final concentration = 0.39 – 1.54 mg/mL (~1 - 5 *10^{13} capsids/mL). The concentration estimates were based on the thesis work of Dr Oliver Preux[233], accounting for a possible 20 % estimation error.

7.4.1.1.2 Calculation of the mass of the empty T5-CLP

Using the upper and lower bounds per atomic weights of the atoms contained in each protein, as described in **Error! Not a valid bookmark self-reference.**, we can make the following estimates:

Table 25. The atomic element that compose the T5 proteins and the respective values of their Standard atomic weight. The reference values for the Lower and Upper bound for the Standard Atomic Weight of elements were taken from the official site of NIST [313], which made reference to data from [314–316].

Atomic element	Standard atomic weight (lower bound)	Standard atomic weight (upper bound)
C	12,00960	12,01160
H	1,00784	1,00811
N	14,00643	14,00728
O	15,99903	15,99977
S	32,05900	32,07600

According to the information given from the work of Huet et al. [211], and that the bacteriophage T5 has a Triangulation number T=13, we can make the following calculation:

- Subunits for pb8 (Processed head protein) = 120***6** + 11***5** = 775;
  - Where in bold are indicated (5= Pentamers; 6= Hexamers)
- Subunits for pb7 (Processed portal protein) = 12

We could also add the following table in order to show the changes of the amino acid sequences that were taken into account in order to reach the final formulations as described in Table 26. The two proteins used in our calculations are highlighted in blue in Table 26.

Table 26. The calculated atomic molecular mass formula for all the proteins composing the capsid for T5. The amino acid sequence of the original protein is presented (in black) and the cleaved region is shown in red. The final processed protein is shown in blue.

Components	**Amino Acid Sequence**	**Molecular formula**
pb8 (Major capsid protein) =	MTIDINKLKEELGLGDLAKSLEGLTAAQKAQEAERMRKEQEEKELARMNDLVSKAVGEDRKRLEEALELVKSLDEKSKKSNELFAQTVEKQQETIVGLQDEIKSLLTAREGRSFVGDSVAKALYGTQENFEDEVEKLVLLSYVMEKGVFETEHGQRHLKAVNQSSSVEVSSESYETIFSQRIIRDLQKELVVGALFEELPMSSKILTMLVEPDAGKATWVAASTYGTDTTTGEEVKGALKEIHFSTYKLAAKSFITDETEEDAIFSLLPLLRKR	C_{2247} H_{3641} N_{609} O_{713} S_9

Cleaved scaffold + pb8 (Processed)	LIEAHAVSIEEAFMTGDGSGKPKGLLTLASEDSAKVVTEAKADGS VLVTAKTISKLRRKLGRHGLKLSKLVLIVSMDAYYDLLEDEEWQD VAQVGNDSVKLQGQVGRIYGLPVVVSEYFPAKANSAEFAVIVYKD NFVMPRQRAVTVERERQAGKQRDAYYVTQRVNLQRYFANGVVS GTYAAS	C_{781} H_{1294} N_{218} O_{259} S_4
pb7 (Portal protein) = Pro-peptide + pb7 (Processed)	MGFKSWITEKLNPGQRIIRDMEPVSHRTNRKPFTTGQAYSKIEILNR TANMVIDSAAECSYTVGDKYNIVTYANGVKTKTLDTLLNVRPNPF MDISTFRRLVVTDLLFEGCAYIYWDGTSLYHVPAALMQVEADAN KFIKKFIFNNQINYRVDEIIFIKDNSYVCGTNSQISGQSRVATVIDSLE KRSKMLNFKEKFLDNGTVIGLILETDEILNKKLRERKQEELQLDYN PSTGQSSVLILDGGMKAKPYSQISSFKDLDFKEDIEGFNKSICLAFG VPQVLLDGGNNANIRPNIELFYYMTIIPMLNKLTSSLTFFFGYKITP NTKEVAALTPDKEAEAKHLTSLVNNGIITGNEARSELNLEPLDDEQ MNKIRIPANVAGSATGVSGQEGGRPKGSTEGD	C_{2005} H_{3184} N_{536} O_{615} S_{14} C_{1948} H_{3099} N_{523} O_{601} S_{13}

This gives a final result of the mass of the bacteriophage T5 empty capsid in the range [26 016 146 ; 26 019 564] Daltons, with an average of 26 017 855 Da, as shown in Table 27.

Table 27. The calculated mass for each of the proteins that compose the bacteriophage T5 capsid. The values of the atomic weights used to calculate the masses in Daltons for the Lower and Upper bounds are described in Using the upper and lower bounds per atomic weights of the atoms contained in each protein, as described in ***Error! Not a valid bookmark self-reference.****, we can make the following estimates:*

Table 25.

Components	Molecular formula	Calculated mass per protein [Da]		Number of subunits	Total mass [Da]	
		min	max		min	max
pb8 (Processed head protein)	C_{1466} H_{2349} N_{391} O_{455} S_5	32889,86	32894,18	775	**25489640**	**25492988**
pb7 (Processed portal protein)	C_{1948} H_{3099} N_{523} O_{601} S_{13}	43875,54	43881,39	12	**526 507**	**526 577**

7.4.1.2 Results on the other viral samples

7.4.1.2.1 Murine Norovirus serotype 5 (MNV-5)

Murine Noroviruses genogroup V (MNV-5) are a model virus, and it is comparable to Human Norovirus.[228] The following shows our analysis by TEM on these samples (Figure 130).

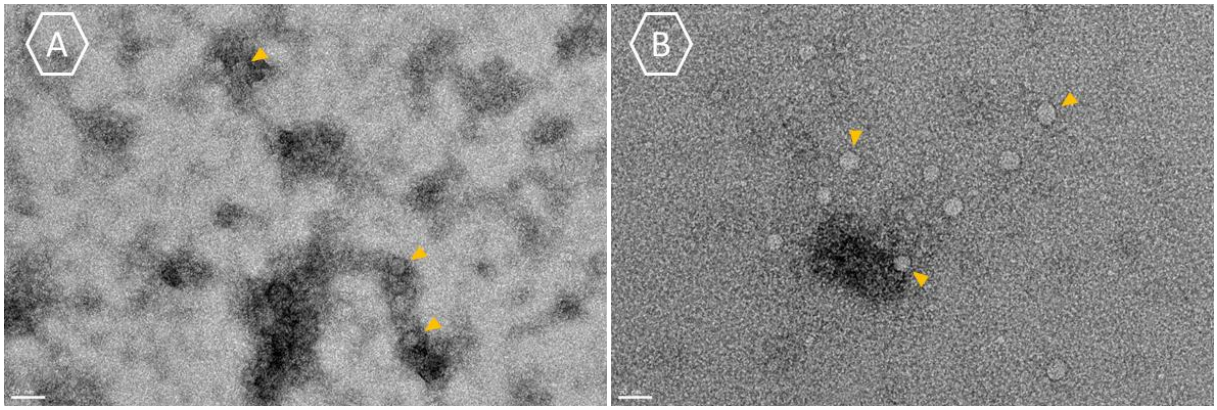


Figure 130. Negative staining TEM images of Murine Noroviruses genogroup V (MNV-5). A) Measured in UAc (1 % v/v in distilled water, pH= 4); B) Measured in SST (1 % v/v in distilled water, pH= 7). Mag =30000X. The scale bar is 50 nm.

The sample seems to be intact and in the expected size, compared to literature.[228] The background (unpurified proteins) is very high when using UAc, but improves when using SST. The capsids are visible but do not appear to be very concentrated. Particles seem to be noticeably different in size and it may be due to the different T-number particles present in the sample. Purification is needed to observe the subunits.

7.4.1.2.2 MS2-bacteriophage

MS-2 bacteriophage is a model viral particle in many usages in biology and instrumental development.[317,318] The following is my analysis by TEM on the samples (Figure 131).

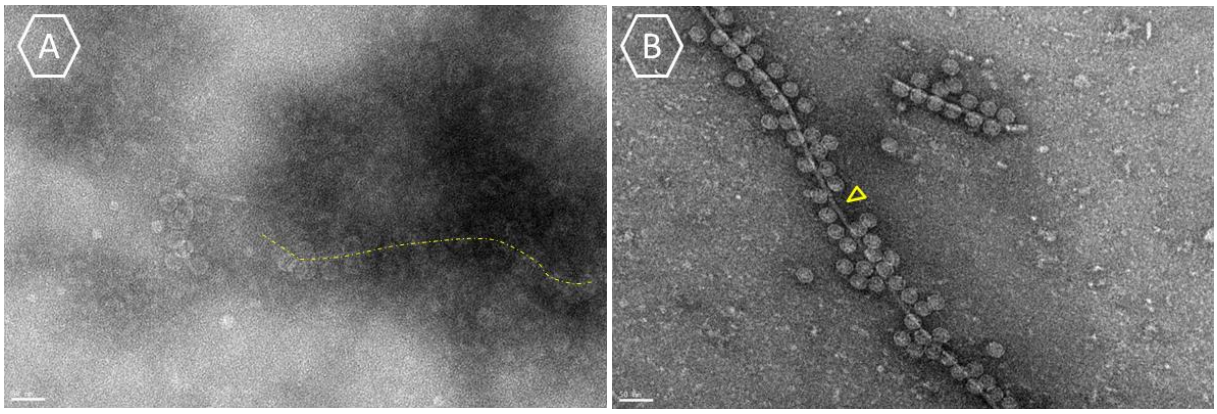


Figure 131. Negative staining TEM images of MS2-bacteriophage. A) Measured in UAc (1% v/v in distilled water, pH= 4); B) Measured in SST (1 % v/v in distilled water, pH= 7). Mag =30000X. The scale bar is 50 nm.

The background appears more charged with uranyl acetate than with SST. The capsids are visible and well concentrated. Phage capsids appear to be more structured in uranyl acetate, although the background is more concentrated. The elongated structure (highlighted in yellow) is suspected to be a pili of the host bacteria, *E.coli*.

7.4.1.2.3 Human Adenoviruses (HAdV)

Human Adenoviruses (HAdV) are model human viruses in many usages in biology and instrumental development.[319,78] The following shows my analysis by TEM on the sample HAdV-5, concerning the serotype 5 (Figure 132). Here I compared them in their native form (images A&B) and after having been inactivated with 2 % of Paraformal Aldehyde (PFA) (images C&D).

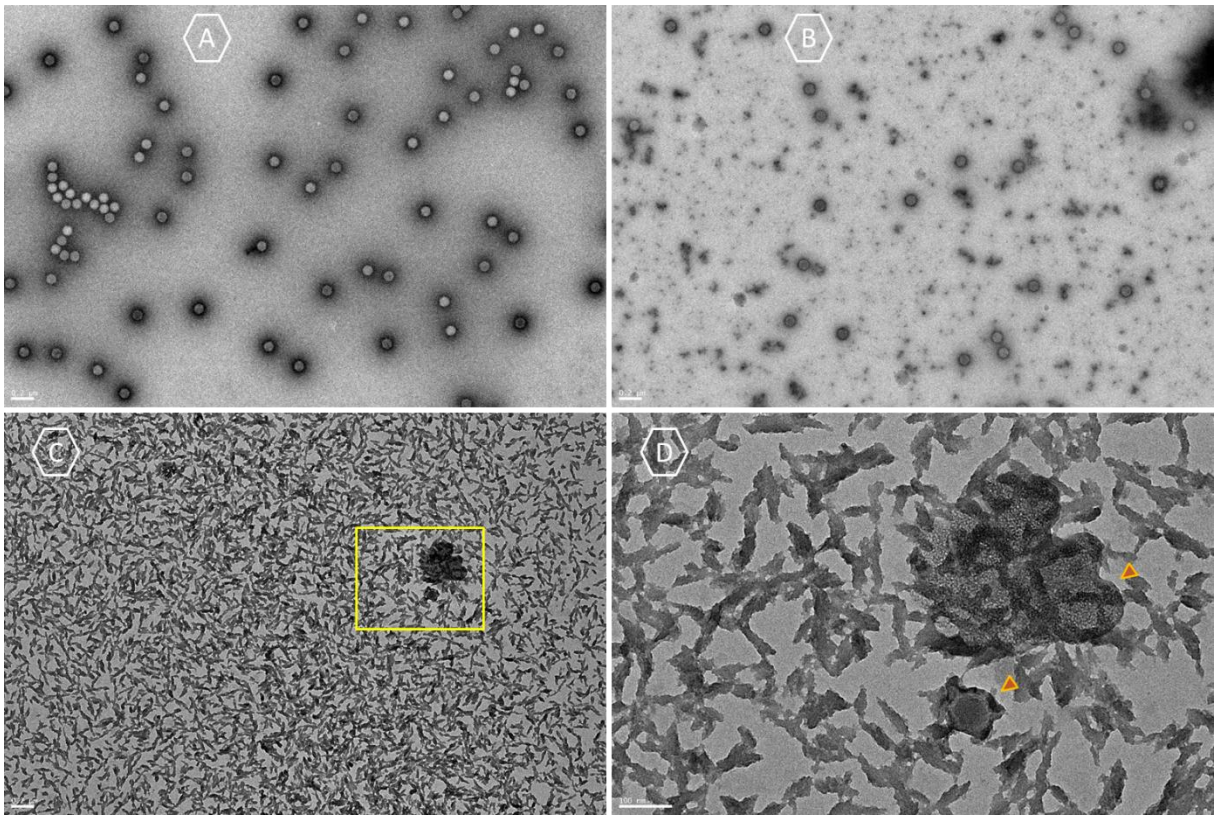


Figure 132. Negative staining TEM images of HAdV5 (2 % v/v UAc coloration), in its native form (images A&B) and after having been cross-linked with 2% PFA (images C&D). The sample in B is 10-x diluted compared to A. Image D, is the zoomed-in area (Mag = 23000x), highlighted in yellow in image C. All other images are at a Mag = 4800x.

WE can see from these images that native viruses seemed stable right after being inactivated (cross-linked) by PFA. Inactivated viruses seemed to be more round (less icosahedral) and seemed to get more attached to one another, forming virus 'dimers' (possibly due to crosslinking with PFA). Three weeks later, when testing for the cross-linking scale (comparing

not diluted & 10x diluted preparations), I noticed that the inactivated sample was completely degraded. In conclusions, it appears that the effect of the PFA crosslinking seem to be that the viral sample is getting degraded with time, and we seem to have lost the conformity of the particles.

7.4.2 Methods

7.4.2.1 Protocol for viral sample preparation for TEM

The following is a protocol of the so called Valentine procedure[240], also named usually as Negative Stain-Mica-carbon Flotation Technique (MFT) is presented as follows. All the discussed image acquisition and analysis was done with the assistance of the EM technician, madame Daphna Fenel [CEA/IRIG/IBS/Plateforme de Microscopie Electronique].

We adsorb the sample of interest using MICA sheets {reference}, which are already coated with a monolayer of carbon on top, graphene layer. This is done by injecting the original sample in between the MICA sheet and the carbon layer. After that we prepare the negative staining agent. Usually, a classical stain for the measurement is Uranyl Acetate (UAc), at a 1-3 % solution, dissolved in distilled water (pH = 4.2 - 4.5). "These solutions also act as a fixative for viruses, but it is preferred to other stains as it has a high electron density and provides improved image contrast. This allows for finer grained images, that are particularly useful for smaller particulate specimens. It has to be underlined here that this stain has a low pH, and it is not recommended for use with specimens that are unstable in acid conditions. Also, the stain precipitates at physiological pH and in the presence of high salt concentration."{reference}

In this work, we also used more neutral, physiological pH (pH ~ 7) negative-staining reagents, such as SST: Sodium Silicotungstate, AmMo: Ammonium Molybdate. The goal here was to observe if the pH would have any visible effects on the sample or if it was independent. In the next step, we plunge the MICA holder sheet, with the carbon facing down, in the container with UAc, sliding it through the surface of the liquid, in order to detach the carbon, which contains the sample of interest at this stage. Using a a 400-mesh copper grid (one square = 10 x 10 µm), we place it on top of the floating carbon layer and make sure it attaches the floating carbon content. The grid's hydrophilic nature helps the sample adhere to the surface. This is then retrieved by an absorbent paper, to get the copper grid out of the UAc solution, which we let dry for some time, before going to the TEM instrument. The grid is then placed on the top of the special prolonged holder, which is then loaded into the instrument's vacuum chamber. The images were taken under low dose conditions (<10 $e^-/Å^2$) with defocus values

between 1.2 and 2.5 µm on a Tecnai 12 LaB6 electron microscope at 120 kV accelerating voltage using CCD Camera Gatan Orius 1000.

7.4.2.2 AFM analysis

7.4.2.2.1 Design of experiments

As far as the design of the experiment is concerned, my main interest lied in assessing the rigidity, topology and size of the different viral particles. In short, my final objective is to compare four samples prepared in different ways:

A. In terms of their production: phCLP vs recCLP.
B. In terms of their decoration: presence vs absence.
C. In solution; Before (after nebulisation) and after the aerodynamic lenses (inside the NEMS-MS bench), as detailed in Figure *62* in the main text.

Thanks to our collaboration with the AFM platform at IBS, I was able to acquire some preliminary results, as follows.

7.4.2.2.2 Preliminary results

The following images (Figure 133) were acquired in liquid, for the sample Juno, deposited directly on an undoped silicon wafer.

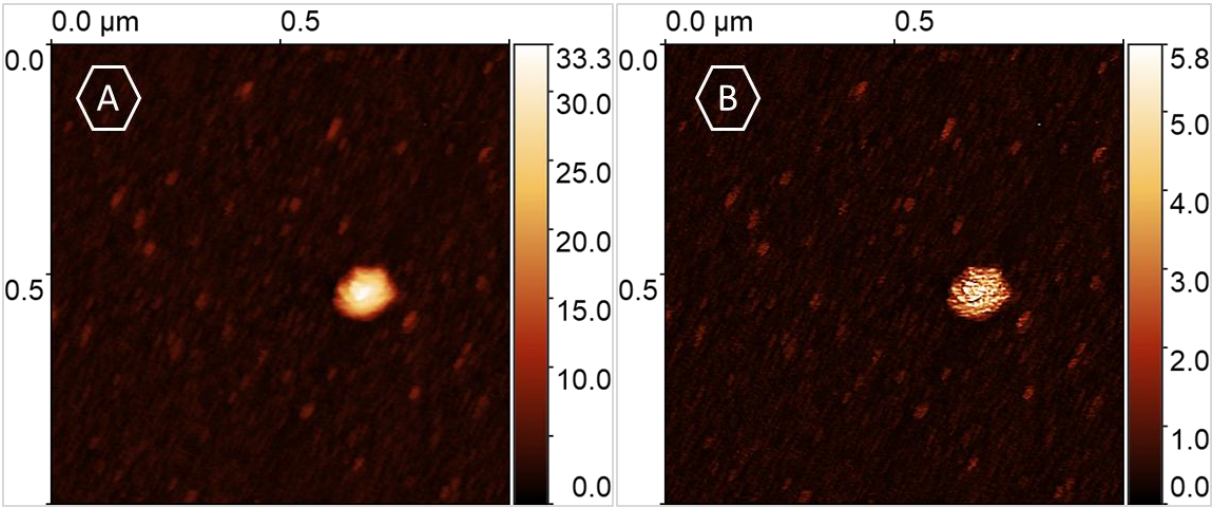


Figure 133. False-colored AFM micrographs of phCLP-deco, named Juno. This image was obtained in liquid and on a silicon wafer. A) Image coloured by gradient, indicated in the right of the figure, normalised to the maximum height; B) image processing of image A to show the sub-units of this CLP, with the appropriate colour gradient indicated in the right of the figure. Images are presented as a (1 µm x 1µm) picture.

We can also see that there are quite a few pieces and bits of particles in the background of the image. Thanks to this image I was able to calculate the volume, assuming it was an ellipse

(height =33.3 nm; width = 93.9 nm; length = 111.5 nm) and compare it to a sphere of dia = 90 nm.

- Vol (ellipse) = $4\pi*a*b*c = 4\pi*16.65*46.95*55.75= 547\ 651\ nm^3$
- Vol (sphere) = $(4\pi/3) * R^3 = (4\pi/3) * 45^3 = 381\ 703\ nm^3$

The 2-fold difference between the calculated sphere and the measured ellipse (ovoid) shape is sufficiently close that we can assume that we are looking at one of the phCLP particles. It can be noticed that the height of the particle does not exceed 35 nm, meaning that the particle has been flattened significantly from its expected spherical shape, thus having a degraded structure, which could explain the two-fold difference between the two calculations.

In perspective, the approach would be the same as with the other methods, primarily trying to reach an inter batch comparison, followed by looking at the virus subunits, that is a comparison between decorated or non-decorated capsids; and of course, the comparison before and after the aerodynamic lenses.

7.4.2.3 SEM analysis and preliminary results

The idea to use SEM as a check point method came for several reasons. Primarily it is a method that is very user friendly and fairly easy to learn even for beginners. In addition to that, the acquisition by SEM of the T5 CLP had already been shown in the past from our group, in the work of Dominguez-Medina, et al.[21], as in Figure 134. Here the particles are visualised when landed on a NEMS detector.

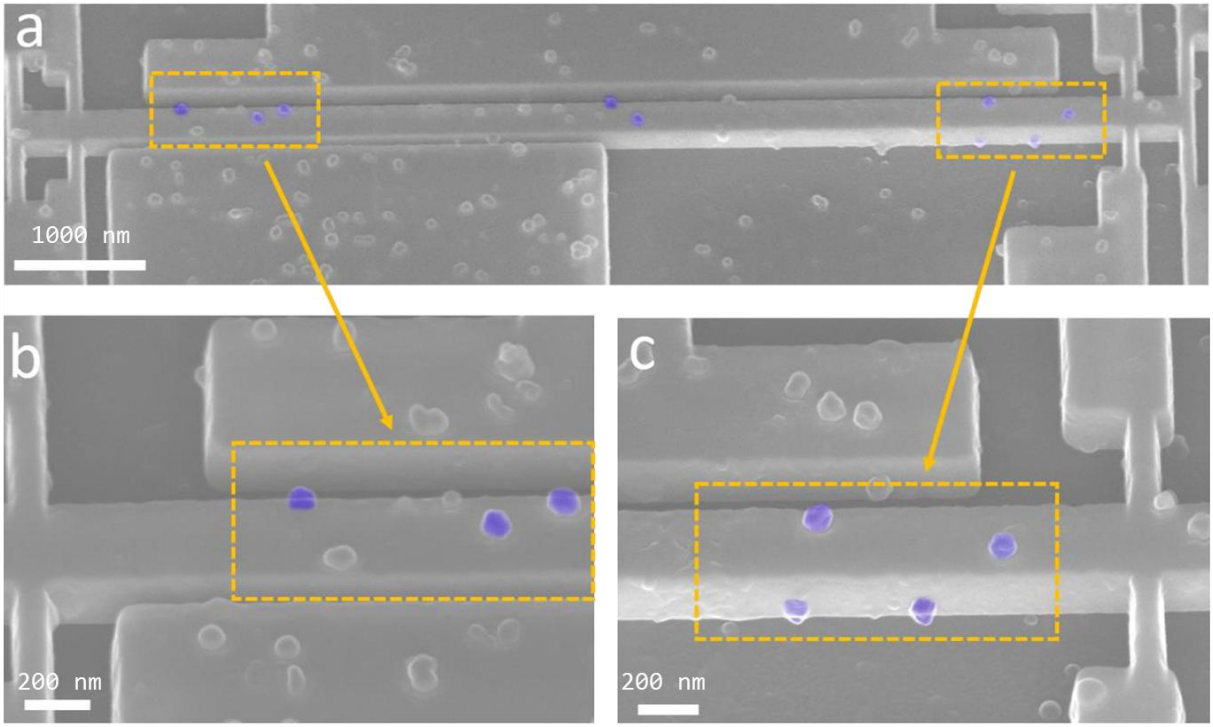


Figure 134. False-colored SEM micrographs of nine T5 filled capsids on a NEMS beam. a) View of the full beam. Scale bar: 1 μm. Panels b), c) Magnified views of seven T5 capsids on different regions of the beam. Scale bar: 200 nm. The other smaller particles observed were deposited before the experiment for aligning purposes and were not considered in the mass histograms in the main text. The size of the capsids are in agreement with their expected dimensions based on previously published negatively-stained electron microscopy images (Zivanovic et al.[320]) and the DLS measurements (~90 nm), despite their exposure to MeOH (methanol) prior to spraying, the high vacuum of the chamber, and the super-sonic speed before landing on the NEMS. A ~2 nm layer of TiN (Titanium nitride) was deposited on the sample using atomic layer deposition (ALD) to enhance the contrast of the capsids under the electron beam. © Figure and description are adapted from the Supplementary information, Fig. S26. [page 36], of the work of Dominguez-Medina, et al.[21]

The technique to visual the samples in this work was that of Atomic Layer Deposition (ALD), which is a subgroup of the greater Chemical vapour deposition (CVD) methods, and has found ever-increasing usage in the field of semiconductor production and has helped its downscaling thanks to its atomic precision. [321,322] This technique allows a sequential deposition of material, in consecutive thin layers of material. This is usually achieved by mixing two precursor, or reactants, in a self-limiting manner, where the two are never present simultaneously in the chamber. [323] The issue in our case with such a method is the high temperatures used to prepare such as deposition. For example, in the case of TiN (Titanium nitride), used for the images in Figure 134, the precursors need to heated at a minimum temperatures of around 350°C.[235]

This of course would heavily degrade the native structure of the capsids, thus beating the purpose of using this technique as a native method for sample analysis. Therefore, an alternative approach needed to be found.

7.4.2.3.1 Experimental design

In order to bypass the problems faced with ALD, alternative methods needed to be found in order to analyse the samples by SEM, in the least steps possible and in the most native conditions possible.

So the first step was to check at what temperature could these samples be heated us in order for them to still be considered as stable. This was demonstrated before by our colleagues, where they measured the thermal stability of the CLPs, using Differential scanning calorimetry (DSC), as seen in Figure 135.

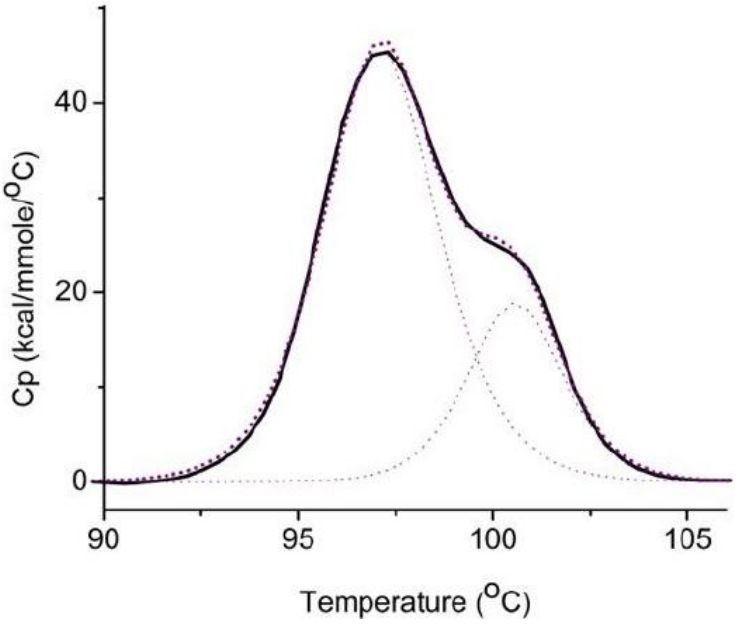


Figure 135. Differential scanning calorimetry profile of empty T5 capsids dialyzed in 25mM phosphate buffer (pH 7.0). Experiments were performed in a VP-DSC Malvern microcalorimeter. The temperature was increased from 25°C to 120°C at a rate of 90°C hr-1. The denaturation profile displays two transitions at 97.1 and 100.6°C as shown by the experimental and fitted curves (solid and dotted lines respectively). DSC analysis of 8 different capsid preparations (concentration of 0.50 mg protein/mL) showed that the surface of the second peak varies from a few percent to 50% of the total surface. This observation suggested that the double peak could come from a contaminant of high thermal stability or from two empty capsid populations present in the sample. © Figure and description are taken as were from the Supplementary information, Fig. S20. [page 30], of the work of Dominguez-Medina, et al.[21]

We can see that the degradation only starts to begin at around 90 °C, and thus any temperature below that would be safe to use.

Next, the common problem with the most deposition methods is that they work at high vacuum, as is the case for the SEM measurement itself. Therefore, assuming that we deposit enough material on top of the capsid, then there is a risk that the capsid becomes a hermetic capsids, meaning that the deposit blocks the transition of fluid from inside to outside, and can result in an explosion of the capsids. In order to make sure theoretically, we estimated that any deposition at the atomic level, or of few nm, would be enough to observe the capsids. And addition to that, since the main orifice in the capsid is expected to be about 10-12 nm in size, there wouldn't be any risks of explosion. This is based on the measurement of the place where the tail is attached to the capsid head, which we relied on visual observation from the work of Huet, et al. [220]

And thus, I decided to proceed with the following pre-treatment for all samples measured:

a) Putting a volume of 5 µL of sample on a undoped silicon ship (Si-chip). All the following measurements were only performed for Juno, empty non-decorated phCLP. The Si-chip is undoped in order to reproduce as best as possible the NEMS that we currently use, thus serving as a good direct comparison, if we were to directly perform measurements on the NEMS.
b) Sample was in its original buffer, as acquired from collaborator, at PBS-1x.
c) Heating the Si-chip to 60-65 °C, for t= 5 min (to remove all excess liquid possible), in order to avoid explosion when in vacuum.

The following are some of the results acquired using different sample preparation methods.

7.4.2.3.2 No metal deposition

The first and most logical approach, when trying to evaluate the native state of the samples, was the evaluation with any type of metallic deposition. Since the samples are very low in conductivity, a compromise needs to be found about the Electron High Tension (EHT) value when performing these measurements. For these measurements, the values were usually between 2-3 kV, but this is an empirical and sample dependent value, as seen from some discussions in scientific forums. [link]

A first examination of the sample was to check the classical modes of visualisation allowed by the SEM. The most classical ones being the SE2 (Secondary Electron) and the InLense (in lense) modes. The information acquired is different between the two, as SE2 allows for high topological contrast and examination, and the detector is located at the side of the microscope chamber; and the InLense mode allows for improve detection, and the detector is found in the beam path of the microscope. [link] Both modes can be coupled in order to have an improved and complementary visualisation of the sample, as can be seen in the following figure, Figure 136.

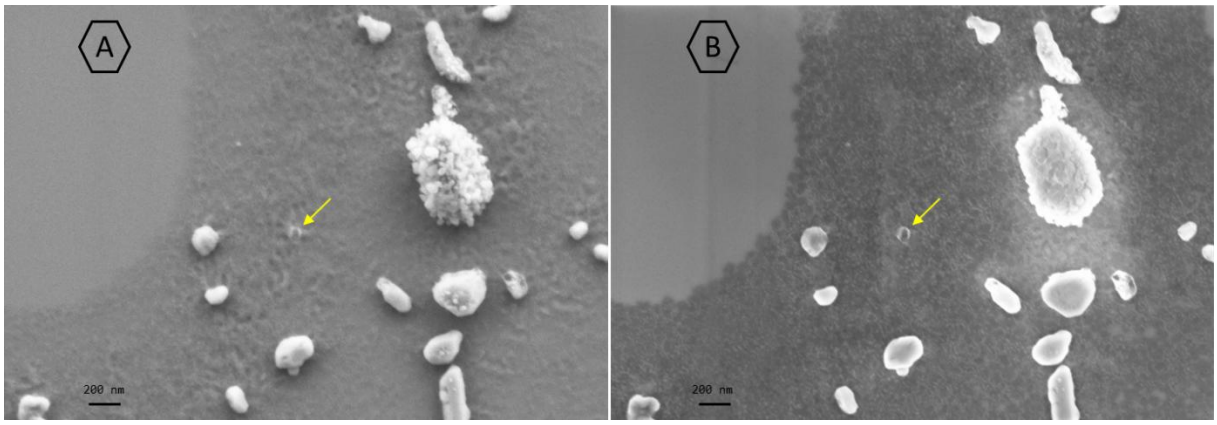


Figure 136. SEM micrographs of empty non-decorated phCLPs, after no metalisation. A) The secondary electrons (SE2) view; B) the In lense detector (InLense) view. The yellow arrow in the centre shows the presence of a single capsid particle. Both performed at an EHT = 2 kV. Scale bar = 200 nm.

We can see from the image above the same image in SE2 mode (Figure 136 A) and in the InLense mode (Figure 136 B). We can see that on the right side of the image, in Figure 136 A, the large particle, nearly 1µm in size, could lead to thing that it may be a bacteria, even looking as if it is being attacked by phages. However, in Figure 136 B, that same particle looks very bright, indicating a large presence of conducting material. This is similar to the other very bright particles in part B, which indicates that these are most likely salt aggregates clustered together. The particle in the centre, indicated by the yellow arrow, was measured at a diameter of around 89.9 nm, which is very compatible to the expected size for T5-CLPs. Therefore, the next logical step was to check how we could image some of these samples.

Electron beam bleaching problem

Imaging the virus particles with this approach proved to be complicated, primarily due to the low conductivity of the viral samples. An additional conductivity issue also arises, as the samples were put onto undoped Si-chip, thus making the electron flow in the material much harder, and thus the imaging by SEM that much more difficult to acquire. This of course result in the need to get closer to the sample and thus the energy deposited on them is higher.

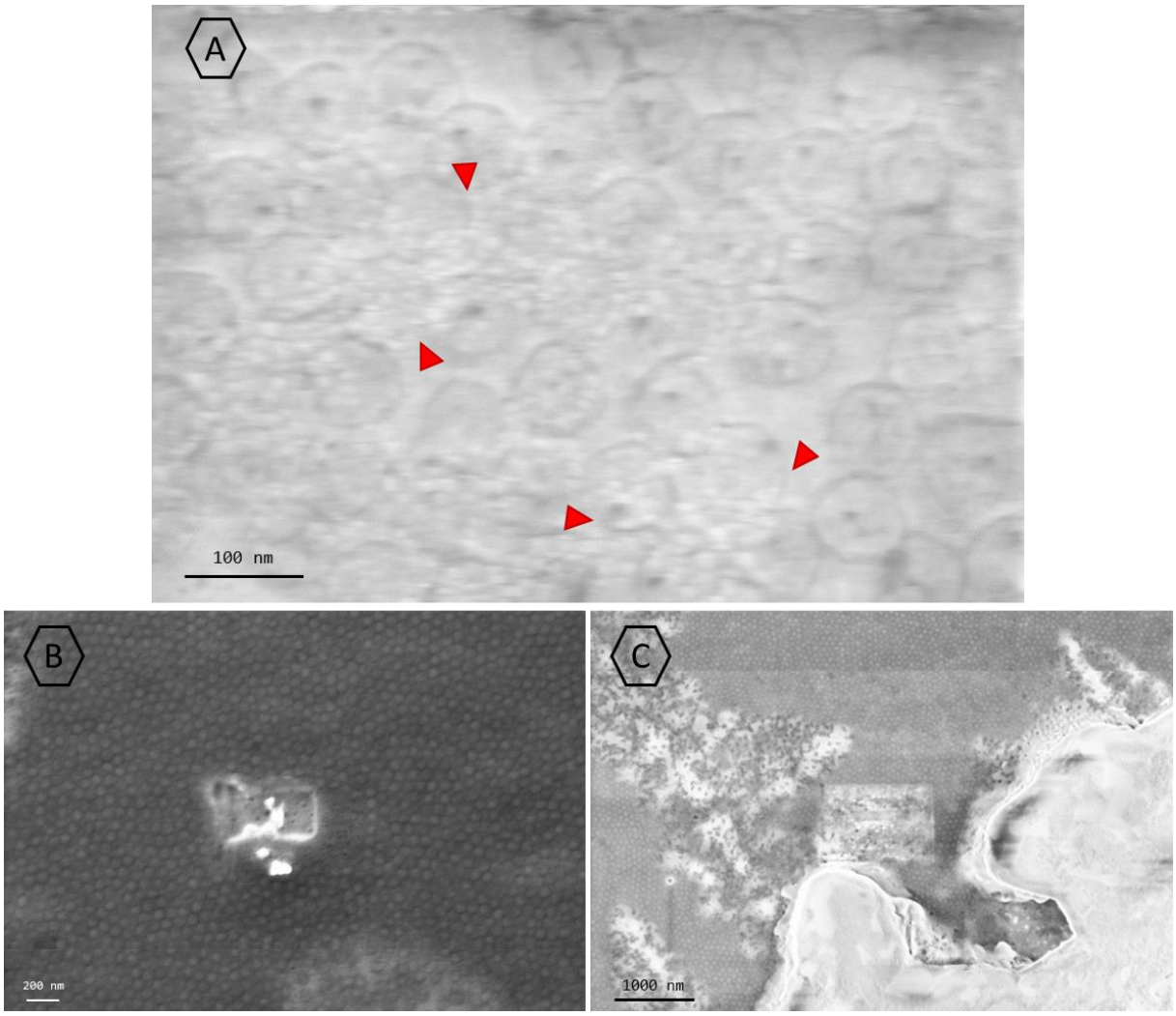


Figure 137. SEM micrographs of empty non-decorated phCLPs, highlighting the damage done by the electron beam. A) Mag=152450x. Red triangles point at visible damage due to e-beam exposure, seeming as brighter spots. Scale bar = 100 nm. B) Image dezoomed, at Mag=313900x. Scale bar = 200 nm. C) Image dezoomed, at Mag=15480x. Scale bar = 1 µm. All images were acquired using the InLense detector, at EHT = 2kV.

This of course results in what is known in microscopy as 'bleaching', or here more specifically as burning of the sample, as seen in Figure 137, and especially visible in the centre of image B and C. We can see very large salt clusters on the bottom half, centre and right, of image C. Some CLPs can be spotted in the top of this image. In image B, we can more clearly the CLPs forming sort of a carpet, covering the entire surface of the area examined. However, when we take a zoomed in image, image A, we can clearly see the CLPs, being at a size of around 70-80 nm, as other methods have confirmed. What we can also clearly see here is that there is extensive bleaching visible, as is pointed out by the red tringles in this image. This is visible as brighter white spots. My explanation for this phenomenon was that this could be spots where nucleate boiling is occurring.[324] Although this is a very complex thermodynamic phenomena, and especially hard to experimentally evaluate, it is demonstrated to occur when

the temperature at the surface is higher than the critical temperature where the vapour-liquid is reached, i.e. boiling temperature, but the heat flux is below the apex, also called the critical heat flux. It is after that point that the liquid will start 'sizzling'. I confirmed that this hypothesis is indeed plausible by personal communication with Christopher Russo (MRC Laboratory of Molecular Biology, Cambridge, UK), who is an expert in cryoEM instrumentation.

Thus, it was concluded that the samples would be very hard to visualise directly as they were, and a metal coating to help the conductivity and visualisation was needed.

7.4.2.3.3 Sputtering deposition

Therefore, I decided to try another way of performing deposition in order to visual the sample, called metal sputtering. Contrary to ALD, which is chemical vapour deposition (CVD) method, sputtering is in the group of physical vapour deposition (PVD) methods.[236] A thin film of metal, in our case we tried platinum (Pt), is deposited on top of a sample of interest by sputtering. This physical concept occurs when a target plate, here Pt, is bombarded by a ions or atoms, and the ejected materials is deposited on top of a target, here the Si-chip containing the T5-CLPs.[325,326]

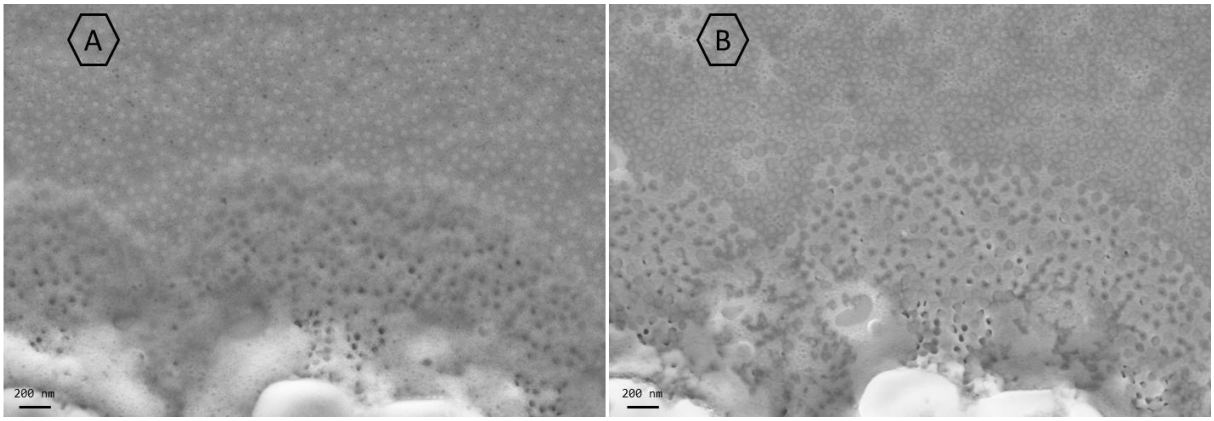


Figure 138. SEM micrographs of empty non-decorated phCLPs, after metal sputtering deposition of Platinum (Pt), comparing the secondary electrons (SE2) view (in A), to the InLense view (in B). Both performed at an EHT = 3 kV. Scale bar, bottom left = 200 nm.

As done before, I compared the SE2 to the InLense modes, to observe the difference in topology and any additional details that could be seen from when we did not metallise. We can see that compared to before, the particles appeared slightly brighter in both modes. They seem well separated from one another, showing no visible overlap between them, primarily on the top of the image. What we can clearly see as well is the presence of the salt on the bottom of the image, appearing very bright. And in the intermediate region, most likely the retraction of the liquid towards the bottom, forming crystals, and the part where some particles were still

surrounded by buffer and the others that are not (on top). We can also see that the size of the particles is consistently between 70-80 nm in these images. So, the next step is to go and delve in more details in these images.

Electron beam bleaching problem

In order to better understand the bleaching effect seen earlier (Figure 137), here we were actually able to acquire some better images (Figure 139), due to the improved exposition time.

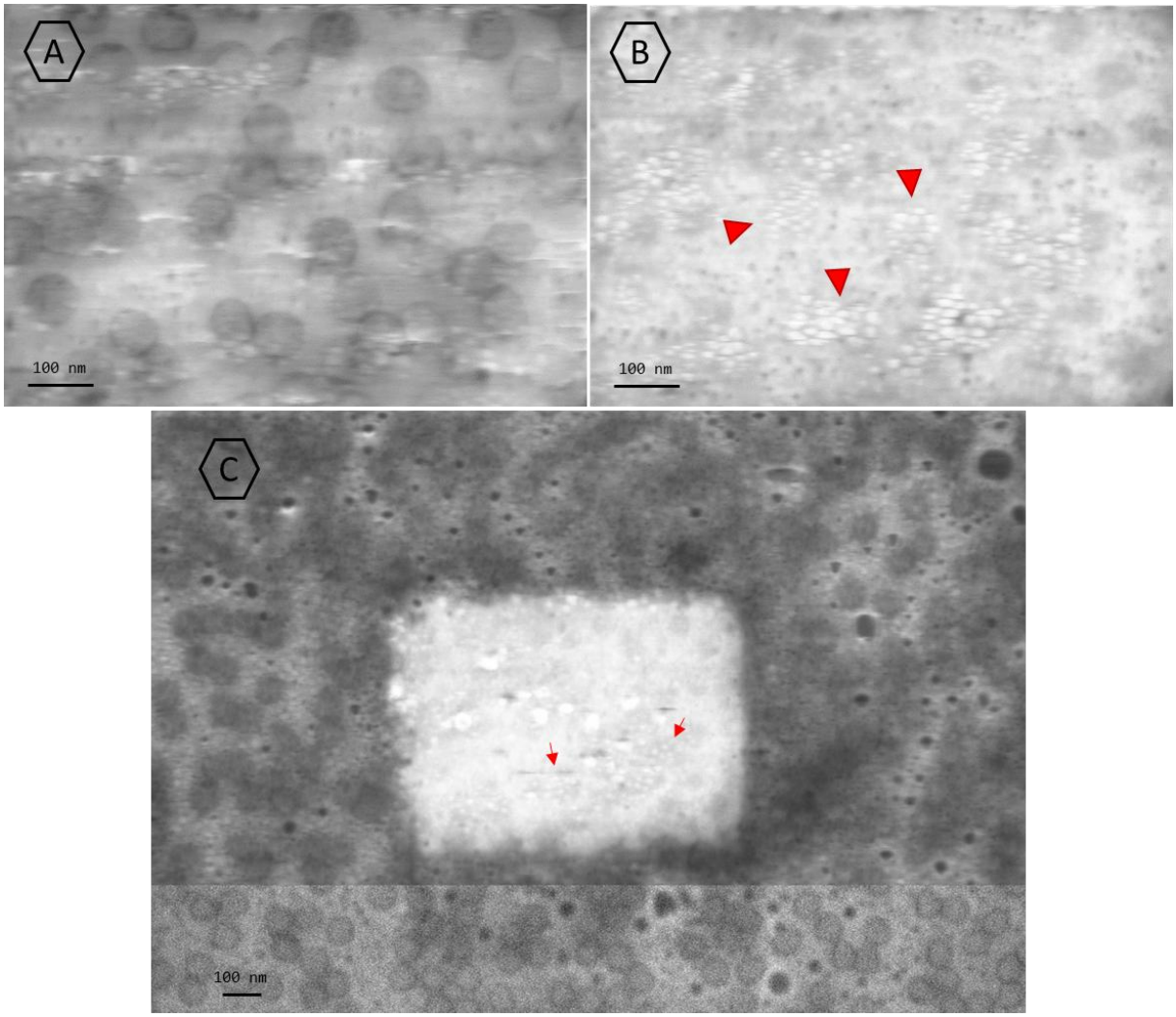


Figure 139. SEM micrographs of empty non-decorated phCLPs, after metal sputtering deposition of Platinum (Pt), comparing e-bean damage. A) Image at Mag=130000x. B) The exact same image as image A, but indicating bleaching, after electron e-beam exposure (t~10 sec). Red triangles point at visible damage due to e-beam exposure, seeming as brighter spots. C) Image dezoomed, at Mag=47910x. Red arrows point at visible damage, seen as brighter spots. Scale bar, bottom left = 100 nm. All images were acquired using the InLense detector, at EHT = 2kV.

In Figure 139 A, again after an exposure of just a few seconds, we cannot reproduce the image, as seen in image Figure 139 B. We see again the presence of whiter spots inside the capsids, pointed by the red triangles, probably indicating nucleate boiling, as explained earlier. The same exact spots can be seen when we de zoom the image, red arrows, in Figure 139 C. In

addition, we can again clearly see the burning of the sample after exposure to the electron beam, at the centre of the image C. Quick testing by X-ray diffraction (XRD) analysis showed no peak for Pt for this sample (data not shown here). So perhaps it could also be that the deposition by sputtering did not work well and that the sample was perhaps not metalized at all, or very little, being below the XRD detection limit.[327] This could explain the very similar degree of bleaching observed between the images in Figure 139, to those seen in Figure 137.

And thus, I decided to try another method for metal deposition, which I hoped could give more resolute results.

7.4.2.3.4 Deposition with the EVAP method

The next method tried, was the one called E-Vap, which is short for Evaporation deposition. This is also a PVD thin-film method, similar to sputtering. This method realises its deposition by having the material that is used for coating, here again we used Pt, heated up to form a vapour. This is done under extreme vacuum, here at $7*10^{-6}$ Pa, and as a results the particles of Pt formed will travel, without any collisions toward the sample of interest, here the Si-chip containing the T5-CLPs.[237] Here the evaporated particles sublimate, forming solids on top of the sample, and thus coating it. Of course, it is assumed that since the sample is not heated up the temperature is room temperature. This process is very short, few minutes, and we left it running in order to get a layer of ~5 nm on top of the sample.

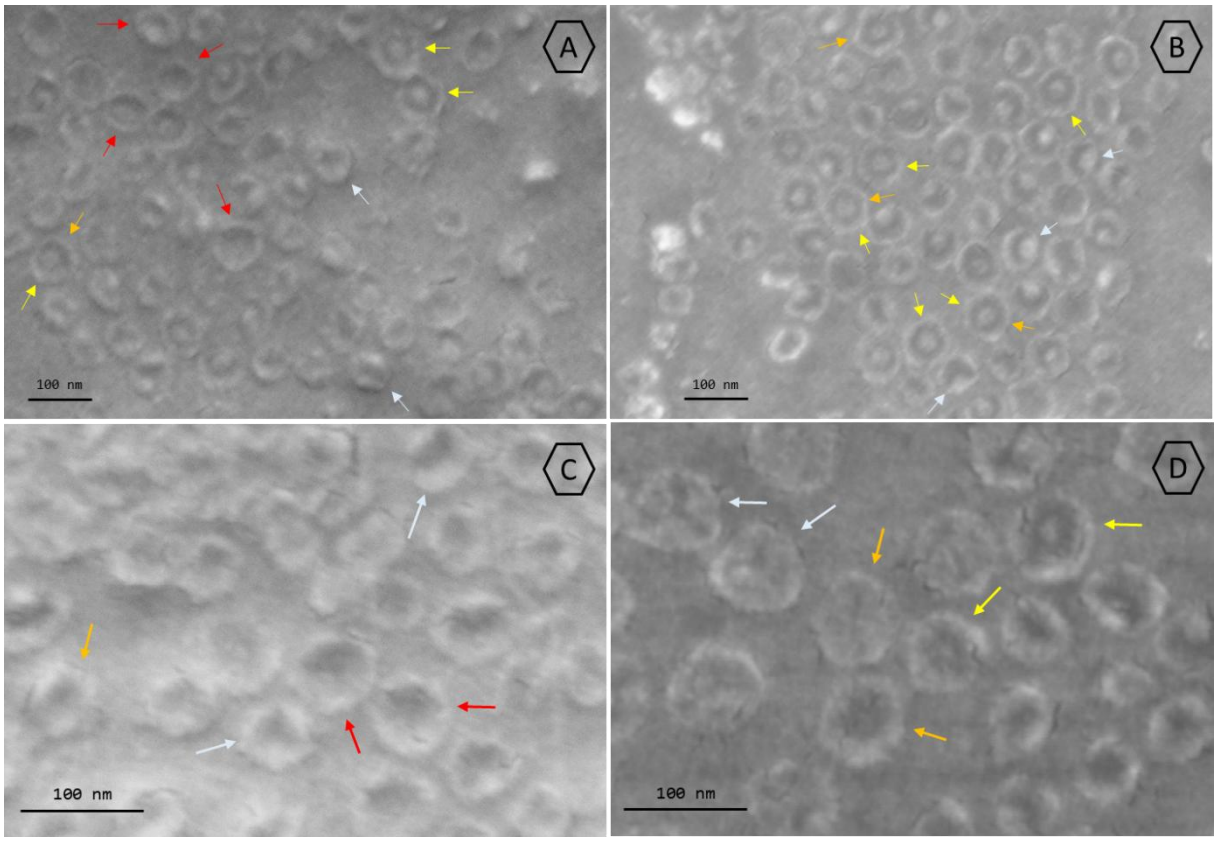


Figure 140. SEM micrographs of empty non-decorated phCLPs. after metal deposition of Platinum (Pt) using the evap method, comparing: The SE2 images on the left (A & C): to the InLense images on the right (B &D). Images were adapted to have the same magnification, on top (A & B), EHT = 3 kV [Mag= 117870x]; and on the bottom (C & D), EHT = 2 kV [Mag=221660x]. Arrow colour coding are as follows: (yellow) capsids that present a centre point; (red) capsids shaped like biconcave disks; (orange) capsids presenting regular, pentagonal shapes; (light blue) crescent shaped or tripartite shaped capsids. Scale bar = 100 nm.

Yellow arrows show capsids that present a centre point, perhaps indicating capsids that are have one of their pentamers pointing up, and thus receiving. Red arrows look like capsids shaped like biconcave disks, similar to red blood cells. Orange arrows capsids presenting regular, pentagonal shapes. Light blue arrows present crescent shaped or tripartite shaped capsids. The tripartite particle, top left of Figure 140D, reminded me of the sepals of an orchid. This could be interesting to note as it can perhaps be related to the fact that these capsids adapt different conformations when they lack their native tails, resulting a different sizes and morphologies, which could explain the differences in size of these CLPs, as observed by TEM for example.

7.4.2.4 TRPS - Protocol used

Tunable resistive pulse sensing (TRPS) was used to characterise T5 Capsid-Like Particles (CLPs) based on their size and concentration by using the Exoid (Izon Science, New Zealand). CLPs were diluted 1000x and 2000x and 35 µL of diluted working samples (in 2x

PBS) were measured using NP80 or NP100 nanopores (Izon Science, New Zealand). The measurements were then calibrated with calibration beads CPC100 (100 nm mean diameter). Data was acquired using the Exoid Control Suite and analysed using Izon Data Suite (Izon Science, New Zealand).

The measurement is done through a unique nanopore, on a stretchable plastic membrane. The solution's conductivity needs to be carefully chosen to ensure optimal particle detection and sizing. The measurement cell is protected by a Faraday cage to remove parasite electrical noise from the environment. The size of this nanopore will define the resolution of the measurement, and also the users skills needed to make these measurement in the most reliable way possible. One can imagine it as a camera slit being opened to a desired size by applying a well-defined force on the material. The data is analysed by an adapted software integrated to the instrument that combines the information from pressure, charge and stretch in order to reach a quality-controlled data analysis. The versatility of the instrument comes from the fact that a variety of nanopore sizes can be used in the instrument, making it reach a size determination of 40 nm to 11 µm. Another advantage is that it is a single-particle, real-time analysis, so as for NTA, so one can check when there are obstructions or bubbles passing on the nanopore and correct as necessary. Same for the concentration check, if the noticed concentration is too high, dilutions will be applied or if too low then one can increase the pressure applied in order to correct for the particle flux passing the nanopore.

7.4.2.5 Proteomics - Protocol used

Samples were treated using the isT kit from Preomics to generate peptides. [link]. The resulting peptides were analysed by online nanoliquid chromatography coupled to MS/MS (UltiMate 3000 RSLCnano and Q-Exactive HF; Thermo Fisher Scientific) using a 90-min gradient : from 2% ACN (acetonitrile), 0.1% Formic Acid to a gradient of 75 % ACN, 0.1% Formic Acid. For this purpose, the peptides were sampled on a precolumn (300 µm × 5 mm PepMap C18; Thermo Fisher Scientific) and separated in a 75 µm × 250 mm C18 column (AUR3-25075C18, 1.7 µm, IonOpticks). The MS and MS/MS data were acquired by Xcalibur (Thermo Fisher Scientific).

Peptides and proteins were identified by Mascot (version 2.8.0; Matrix Science) through concomitant searches against the UniProt databases (Viruses, downloaded January 2023 and an Escherichia Coli, strain K12 database, downloaded 22/05/2022). Proteins from the T5 virus were extracted using the following taxonomy: Taxon: 2695836, Mnemonic: BPT5. Also two

homemade databases were used: one containing the sequences of classical contaminant proteins found in proteomic analyses (human keratins, trypsin, etc.) the other containing the sequence for the pb10-Ovalbumin chimera protein. Trypsin/P was chosen as the enzyme, and two missed cleavages were allowed. Precursor and fragment mass error tolerances were set respectively at 10 ppm and 20 ppm. Peptide modifications allowed during the search were as follows: Carbamidomethyl (C; fixed), Acetyl (Protein N-term; variable) and Oxidation (M,P; variable).

The Proline software [328] was used for the compilation, grouping, and filtering of the results (conservation of rank 1 peptides, peptide length ≥6 amino acids, FDR of peptide-spectrum-match identifications <1% as calculated on peptide-spectrum-match scores using Benjamini-Hochberg estimation, and minimum of one specific peptide per protein group). Proline was then used to perform a MS1 quantification of the identified protein groups based on razor and specific peptides.

7.4.3 Supplementary on nebulisation optimisation

7.4.3.1 Shape dependence and aggregation, as analysed by the SMPS

We did some simple back of the envelope calculation on what the simplest explanation for the particle aggregation and size increase in the SMPS measurement could be, seen in Figure 104. If we assume that the real particle size is at ~90 nm, we can draw the following curves, for the cumulative or aggregate case, seen in the following figure, Figure 141.

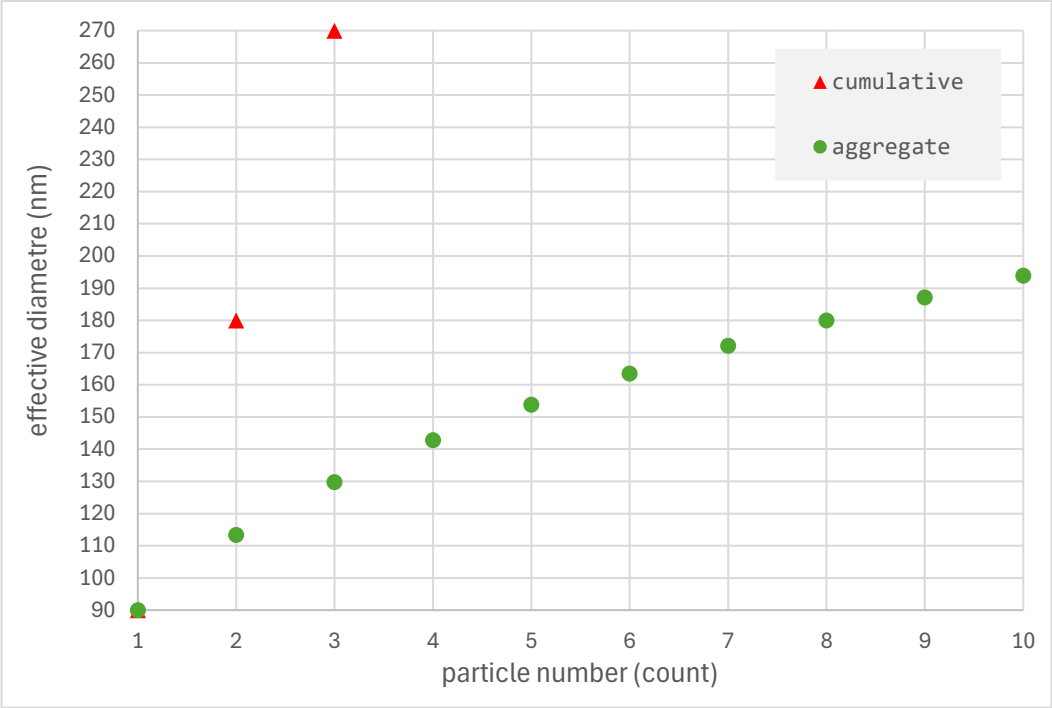


Figure 141. Comparing the effective diameters for different conformations of particle conjugation, inside the SMPS. Legend: cumulative dimeter, when we just add up the particles to one another (red triangle); aggregate: we add up the particle's volumes and convert it into an effective diameter (green circle). All calculations were performed assuming PSL-NPs with a perfect diameter of 90 nm, as it was seen in Figure 104.

As we can see in Figure 141, the most unlikely case is that of 'cumulative effective dimeter' (red triangle), when we just add up the particles to one another only by their length, as that would mean that we would have a power law behaviour on the size of the observed body for every added particle, which is not possible. Perhaps a most adapted explanation for this particle measurement is to add up the particle volumes, and then convert this sum into an 'aggregate effective dimeter' (green circle) in Figure 141. This seems more realistic given that the surface of the spheres is directly related to their volume, although some considerations such as porosity and effective surface exposed to charges have to be accounted for when we make these approximations.

Another interesting, but harder to evaluate observation, would be the consideration of the measurement of the gyration radius (Rg) by the SMPS. This is already an thoroughly investigated topic in polymer science, and even in structural biology for polymeric macromolecules.[329] This concept is deduced by the concept of inertia, where the moment of inertia equals the mass of the body times the square of the Rg. The primary difficulty in such an approximation lies in the fact that the topology of the body analysed is of major importance. Thus, to be able to precisely deduce the respective Rg we have to have prior knowledge of the topology, which would probably require collecting and imaging the aggregates, which actually makes this work very exhaustive, and is out of the scope of the initial analysis.

7.4.3.2 *The theoretical PSL-NP count per droplet*

Next we wanted to see how to relate the two measurements of concentration by the different methods, thus the APS measurements for droplet size (Figure 103) and the SMPS measurements for PSL-NP size (Figure 104).

Thus, we did a simple approximation that since the number of nanoparticles (N_p) is related to the concentration of the NPs (C_p) and the volume of the liquid (V_l) as follows:

$$N_p = C_p * V_l$$

Assumption that the particles are uniformly distributed throughout the solution, and that the droplet contains a representative sample of the solution, we can imagine a (V_l) composed of a number of droplets (N_d) and a droplet volume (V_d), thus:

$$V_l = N_d * V_d$$

Therefore, resulting in the following formula:

$$\frac{N_p}{N_d} = C_p * V_d$$

And since the volume of the droplet (V_d) is directly related to the droplet diameter, we can present the results as in Figure 142, showing the relation between particles per droplet, as a function of the expected droplet aerodynamic diameter.

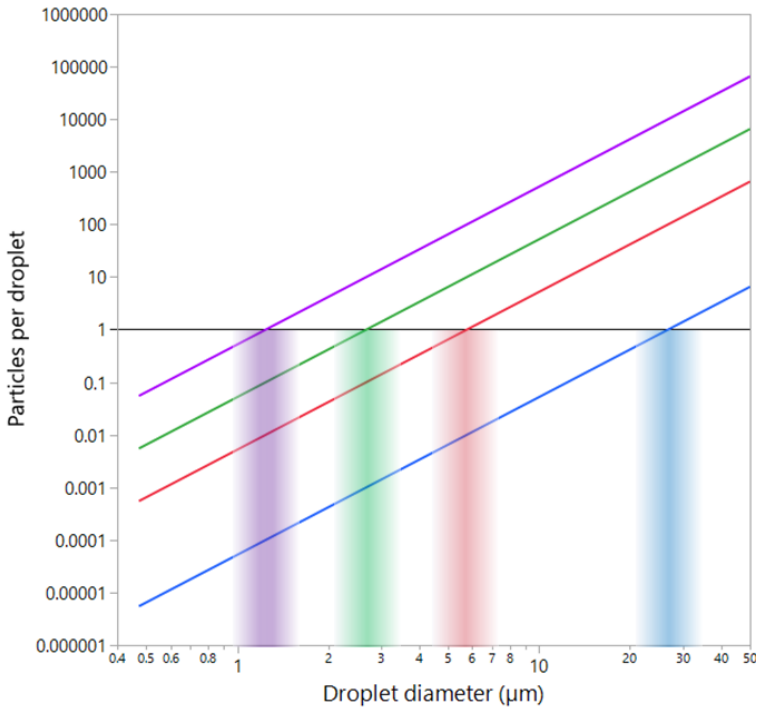


Figure 142. Theoretical relation between particles per droplet, for different concentration of PSL-NPs, as a function of the expected droplet aerodynamic diameter (µm), shown in the range 0.5-50 µm, in a logarithmic scale.. The solid horizontal line at 1 particle/droplet indicates the optimum point for nebulisation efficiency (over = particle aggregation; under= inefficient). The coloured diagonal solid lines represent the converted known particle concentration in solution, into a particle per droplet concentration. The highlighted coloured vertical bands correspond to the intersection of the diagonal lines to the 1

particle/droplet limit. Legend (coloured as for Figure 103), from right to left: 1e8 NP/mL (blue), 1e10 NP/mL (red), 1e11 NP/mL (green), 1e12 NP/mL (purple).

Thus, we can clearly conclude from this graph that the critical droplet size for the optimal threshold of 1 particle/droplet, for each concentration, is as follows: 1e8 NP/mL = between 20-35 µm (blue); for 1e10 NP/mL = between 4.5-7 µm (red); for 1e11 NP/mL = between 2-3.5 µm (green); and lastly, for 1e12 NP/mL = between 1-1.5 µm (purple). This graph also tells us about the type of droplets we would need to generate in order to be able to have the most optimal nebulisation. If we are over the 1 particle/droplet threshold, we get particle aggregation; if we are below this threshold, we get inefficient nebulisation.

This allows us to reach the next step, which is the calculation of the number of generated particles for a given droplet size.

7.4.3.3 Absolute droplet concentration per droplet size

The first step in reaching the particle generation number versus droplet size relation is to have the absolute droplet concentration per droplet size, which is the same as for Figure 103. Thus, in Figure 143 we have the APS measurement for the different PSL-NPs concentrations as a function of the droplet size.

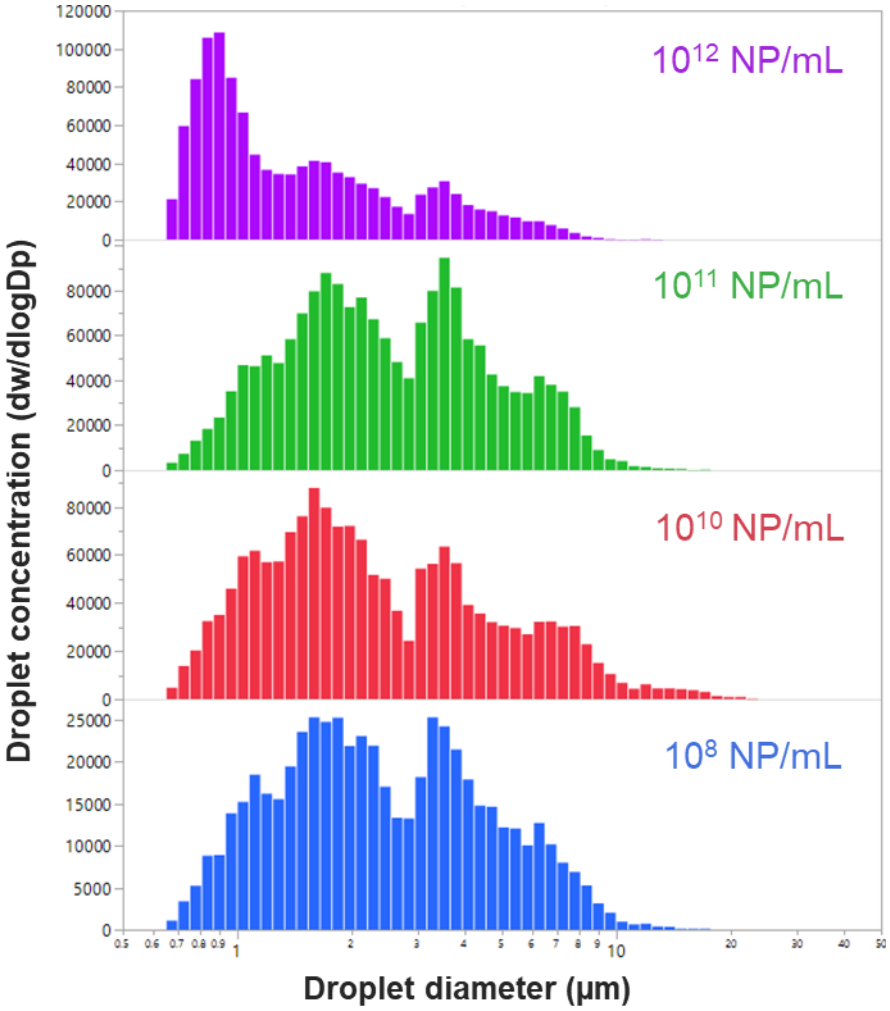


Figure 143. APS measurement, showing the normalised droplet concentration (dw/dlogDp), generated by SAWN, as a function of droplet aerodynamic diameter (µm). Legend (same as for Figure 103), the different PSL-NPs concentrations values (in solution) are presented in decreasing order, from top to bottom: 1e12 NP/mL (purple), 1e11 NP/mL (green), 1e10 NP/mL (red), 1e8 NP/mL (blue).

What we can see from Figure 143 is that for the 1e12 NP/mL (purple) distribution, we have a dominant peak at around 0.8-0.9 µm, with smaller bumps at around 1.7-2 µm and 3-4 µm. This first large peak is absent for all the other distributions. Meaning that that this is most The 1e11 NP/mL (green) and 1e10 NP/mL (red) distributions, seem to be quite similar, around 25 % lower than that of the purple distribution, although their shapes are slightly different. The main difference being that red distribution (1e10 NP/mL) represents a longer tail, extending at the 10-20 µm range, again pointing to an unstable state for this concentration. Lastly, we can see that the blue distribution (1e8 NP/mL) is much lower compared to the other distributions, nearly 3-fold.

7.4.4 Comparing the DMA to the IMS, and how to relate it to the virus scale!

The IMS (Ion Mobility Spectrometry) is a technique that has gained significant track in the field of MS in recent years. It allows to give an additional dimension to the mass/charge

evaluation, by deducing the size of ions, and from there reaching the currently accepted value called CCS (Collisional Cross-Section). I would like to thank the course of Valerie Gabelica (IMS short course at the IMSC 2022), for the introduction of this method from the very fundamental concepts to the practical usages.

7.4.4.1 Shared principle between IMS and DMA

Whatever the measurement methods studied here, they are all based on the same physical principle: an ion is sent into a continuous medium (gas, liquid, etc.) which is more or less "dense" in relation to the characteristic dimensions of the ion (the notion of "fluid density" will be explained below). Which is accelerated by a constant electric field, denoted as **E**, (gravity is assumed to be negligible in all cases). The motion of this ion under these conditions is asymptotically equal (for long times) to a motion at uniform velocity v (depending on the situation, different reasoning leads to this result). The electric mobility of the ion, denoted K, is then defined as :

$$K := \frac{\mathrm{v}_*}{\mathbf{E}}$$

The 'constant' K is the electrical mobility of the ion in question (or, equivalently, its ionic mobility) and has the unit $m^2\ V\ s^{-1}$. While the value of K can be measured experimentally, (it is sufficient to measure v and then divide by the value of **E** to obtain it), linking this parameter to the characteristic quantities of the particle requires modelling the behaviour of the ion in its fluid medium as a function of the physical quantities of interest. Explanation that will be given in the following paragraphs.

7.4.4.2 Introduction to some relevant concepts in fluid mechanics

Before going into the details of the calculations and concepts, it is necessary to introduce a few physical quantities of interest for what follows.

We start with the mean free path of a fluid particle in the fluid itself. This distance, denoted λ, represents the average distance that a fluid particle travels in its medium without colliding with another particle of the same fluid. Intuitively, the greater the mean path, the less "dense" the fluid. We then introduce the Knudsen's number, Kn, which is the ratio between the mean free path of a particle and a physical length representative of the system under study, generally noted L :

$$\mathrm{K_n} = \frac{\lambda}{L}$$

If we are studying the physics of an ion immersed in a fluid, is therefore the mean free path of the constituent elements of the fluid and L is the characteristic distance of the ion itself. As the ion itself is modelled as a sphere, L corresponds to its diameter.

Knudsen's number can be used to define three different ways of modelling the interaction of the ion under study with the fluid in which it is immersed:

1. if K_n is close to 0 (in practice < 0.1), the system is said to be in a continuous regime, (*i.e.* for DMA)
2. if K_n is close to 1 (in practice between 0.1 and 10), we speak of an intermediate regime
3. if K_n is large (in practice > 10), we speak of a free molecular regime, (*i.e.* for IMS).

We interpret the two extreme cases which are conceptually the most accessible. The case of the continuous regime is easy to conceive: the immersed ion is so large in relation to the characteristic dimensions of the fluid that it cannot see the molecules of the fluid individually, but nevertheless feels the collective effects (or, equivalently, the average effects). The carrier fluid is then assimilated to a continuous medium. The case of the molecular regime is just as easy to understand: the ion is so small that it 'sees' the individual molecules that make up the fluid. The fluid is seen as a (very large) collection of hard spheres separated by large empty spaces, in other words a discrete medium.

7.4.4.3 Immersed particle in continuous regime

In such a situation, the behaviour of the ion is commonly modelled using an ordinary differential equation, constructed using the concepts of Newton's second law:

$$\frac{d}{dt}\mathrm{v} = mne\mathbf{E} - \alpha\mathrm{v}$$

where m is the mass of the ion, n is the signed charge number of the ion, e is the elementary charge of an electron, E is the electric field, α is a strictly positive deceleration coefficient and v is the velocity of the ion.

The drag force experienced by a particle in such a medium (in our case, it's the force $-\alpha$v) can often be modelled using Stokes' law:

$$\alpha = 3\pi\mu D$$

where μ is the dynamic viscosity of the liquid (in Pa s) and D the diameter of the ion modelled as a sphere.

The differential equation as presented therefore admits a finite limiting velocity value v for long times (which would go to infinity from a purely mathematical point of view) :

$$v_* = \frac{ne}{\alpha}\mathbf{E}$$

And by introducing the expression of α from the previous equation:

$$v_* = \frac{ne}{3\pi\mu D}\mathbf{E}$$

Given the definition of K given in the introduction, we find for the electric mobility of the ion in continuous regime to be as follows:

$$K = \frac{ne}{3\pi\mu D}$$

This expression is exactly the one used in the SMPS manual (apart from the Cunningham factor) to calculate the ion's electrical mobility (Appendix B, page B-10).[276] There are two things to note about the equation given in this manual:

1. firstly, dynamic viscosity is a variable that does depend on temperature according to the referenced law. However, if the temperature at which the experiment is carried out is identical to the reference temperature, the effective viscosity of the fluid is identical to its reference viscosity ... which is documented everywhere.
2. in the case of "dense" media, the Cunningham factor is reduced to exactly 1, because the Knudsen factor is very close to zero or zero.

Note 1. To extend this result to other regimes, in particular intermediate regimes, we modify the expression of K by the Cunningham coefficient (C_n) according to the following formula:

$$K = \frac{neC_n}{3\pi\mu D}$$

where the values of α, βand γ are tabulated as required:

$$C_n = 1 + \frac{2K_n}{D}(\alpha + \beta e^{-\frac{\gamma}{K_n}})$$

We can see that this expression tends towards 1 when the Knudsen tends towards 0 (i.e. in purely continuous regime).

7.4.4.4 Particle immersed in free molecular regime

In such a situation, the limiting speed of the ion's displacement for long times has to be calculated differently. To do this, we propose a special model: we assume that the ion loses all the speed it has acquired thanks to the electric field E each time it encounters a particle in the

fluid. It is also assumed that the collisions between this ion and the fluid particles are sufficiently numerous and regular for the resulting average friction force to eventually (over long periods of time) cancel out the electric force generated by the E field. Under these conditions, the ion does indeed eventually reach a stationary mean velocity v, aligned with the electric field E, and we place:

$$K := \frac{\mathrm{v}_*}{\mathbf{E}}$$

It is clear that K is inversely proportional to the number of particles in the fluid medium: the greater the number of particles in the fluid, the greater the shock, the more the ion is slowed down and the lower its final velocity. Similarly, the greater the thermal motion of the particles in the fluid, the greater the probability that the ion will fall within the range of a fluid particle close to its trajectory, and the more it will slow down. These considerations, together with more detailed calculations that take into account the angle of collision between the ion and any fluid particle, give us :

$$K := \frac{3}{16}\sqrt{\frac{2\pi}{m_r k_B T}}\frac{ne}{N\Omega_D}$$

where T is the temperature of the fluid, k_B is Boltzmann's constant, N is the number of fluid particles in the system, D is the effective area of collision between the ion and the fluid particles and

$$m_r := \frac{m + m_p}{m m_p}$$

Where m_r is the reduced mass of the system with m_p the mass of a particle of the carrier fluid.

To calculate the diameter of the ion as a function of the previous mobility K, the mass of the ion must be expressed as a function of its specific gravity and its volume. The molecular regime model assimilates the various elements encountered with spheres; it is then very easy to the size of the ion under study, usually reported as a CCS (Collisional Cross-Section).

7.5 Chapter 6: Characterizing T5-CLPs as vaccine candidates

7.5.1 Concentration evaluation at I2BC

We know that the I2BC evaluation is indirect, and based on relative protein content, extrapolated to reach a particle count, and on top of that the calculation is not the same between phCLP (in green) and recCLPs (in purple). The previous method for concentration evaluation, applied to phCLPs, was the one developed by Olivier Preux during his thesis.[233] By this method, the CLP count is determined by simply measuring and performing the ratio between the absorbance at 280 nm (A_{280}), which serves for the protein content evaluation, and 260 nm (A_{260}), for nucleic acid evaluation, and then applying a correction formula that takes into account a diffusion factor, related to capsid size and geometry.

In his thesis, Nicolas Ducrot argued that this method was not well suited for recCLP, arguing that it underestimated the real concentrations.[209] He used the BCA (BiCinchoninic acid Assay), a colorimetric protein assay based on the ability of proteins to reduce Cu^{2+} ion to Cu^{+} ion, in an alkaline medium, such as PBS buffer. Bicinchoninic acid forms a purple complex with the Cu^{+} ion, with absorbance (at 562nm) proportional to the protein concentration of the sample assayed.[290] He argued that this assay, carried out under denaturing conditions, overcomes the diffusion factor constraint associated with capsid shape, thus enabling reliable concentration measurements. Comparison of the concentration of different batches of these two methods, for three batches of recCLPs, showed that concentration estimation by the A_{260}/A_{280} measurement gives underestimated concentrations of ~50 % compared to BCA.[209] This overestimation points to an increase in concentration by 2 to 3-fold, when compared to NTA for example. We believe that this overestimation is due to the relative nature of these evaluations, and the fact that it is based only on measuring the protein concentration in solution. This of course generates a major bias, given that any floating proteins that might be left in the preparation can be counted in the final particle concentration. That is why we believe that this data is not robust enough to show the real particle concentration in solution.

7.5.2 TEM related information.

7.5.2.1 Relevant statistics

The following table, Table 28, gives the main statistics concerning the size measurements by TEM, in SST (Sodium silicotungstate).

Table 28. Size report concerning the TEM measurements in SST. Samples are: Isis (phCLP), Orion (recCLP); non-Decorated: -deco, Decorated: +deco. Abbreviations: Std Dev: standard deviation; Geo. Mean: Geometric mean; the D10, D50 (median) and D90: the 10th, 50th and 90th percentile; RSD: Relative standard deviation.

Sample Name	Partc. count	Mode (nm)	Mean (nm)	Std Dev (nm)	Min (nm)	D10 (nm)	D50 (nm)	D90 (nm)	Max (nm)	Span	RSD (%)
Isis -deco	202	75	80	5	66	74	80	85	99	0.14	6.0
Isis +deco	120	77	84	6	70	77	83	91	100	0.17	7.0
Orion -deco	136	81	88	7	71	80	87	96	108	0.19	7.5
Orion +deco	96	75	83	5	71	77	83	89	98	0.15	5.8

7.5.3 NTA related information.

7.5.3.1 Relevant statistics

The following table, Table 29, gives the main statistics concerning the concentration measurements by NTA.

Table 29. Concentration report concerning the NTA measurements. Samples are: Cesar, Juno & Isis (phCLP), Orion (recCLP); non-Decorated: -deco, Decorated: +deco. *Abbreviations: Conc. (partc./ mL): concentration (particles per mL); Std Dev: standard deviation of the mean.*

Sample name	Video count	Mean (Conc. (partc./ mL))	Std Dev (Conc. (partc.s /mL))	Mean (partc. tracked /video)	Std Dev (partc. tracked /video)	Min (partc. tracked /video)	Max (partc. tracked /video)	Mean (partc./ Frame)	Std Dev (partc./ Frame)	Min (partc./ Frame)	Max (partc./ Frame)	Total partc. count
Cesar	54	1.07 E+13	1.76 E+12	3273	2670	957	8684	48	11	29	63	108905
Juno	35	4.32 E+13	3.21 E+12	1118	329	753	1631	33	12	20	55	47371
Isis -deco	16	2.44 E+13	4.74 E+12	2232	687	1059	3387	69	18	36	98	126736
Isis +deco	27	1.85 E+13	5.81 E+12	1824	658	779	2855	52	17	26	79	148667
Orion -deco	36	9.12 E+12	1.22 E+12	2366	603	1271	3271	71	19	43	100	127286
Orion +deco	14	7.24 E+12	1.75 E+12	1122	244	757	1538	40	10	28	58	98638

The following table, Table 30, gives the main statistics concerning the size measurements by NTA.

Table 30. Size report concerning the NTA measurements. Samples are: Cesar, Juno & Isis (phCLP), Orion (recCLP); non-Decorated: -deco, Decorated: +deco. *Abbreviations: Std Dev: standard deviation; Geo. Mean: Geometric mean; the D10, D50 (median) and D90: the 10th, 50th and 90th percentile; RSD: Relative standard deviation.*

Sample name	Video count	Total partc. count	Mode (nm)	Mean (nm)	Std Dev (nm)	Geo. Mean (nm)	Geo. Std Dev	Min (nm)	D10 (nm)	D50 (nm)	D90 (nm)	Max (nm)	Span average	RSD average (%)
Cesar	44	108905	105	79	38	71	1.58	12	39	71	127	685	1.23	47.9
Juno	40	47371	101	86	41	77	1.60	12	42	79	139	658	1.23	47.5
Isis -deco	67	126736	101	91	42	82	1.58	13	45	83	147	803	1.23	46.6
Isis +deco	90	148667	110	90	42	81	1.59	10	43	82	145	504	1.24	46.7
Orion -deco	58	127286	106	83	41	75	1.59	11	41	75	136	616	1.27	48.9
Orion +deco	65	98638	114	98	46	88	1.59	9	48	90	156	830	1.20	47.4

7.5.3.2 Comparing different concentration methods

The following table and explanation were taken from the original, Anderson et al.[330], comparing the different methods that are used to quantify the bacteriophage number. These are the plaque assay (PA), NanoSight Limited (NS)-based (the same as NTA) and quantitative real-time polymerase chain reaction (QPCR)-based methods.

Table 31. Coefficients of variance (CVs) for the plaque assay (PA), NanoSight Limited (NS)-based and quantitative real-time polymerase chain reaction (QPCR)-based methods. [a]The outlier excluded from further analysis. n = 390. For the pA, ecML-117 and List-36 phages: 30 specimens per expected titer per phage = 180. For the YpP-G phage: 10 specimens per expected titer = 30. For all other assays and phages: 10 specimens per phage per expected titer = 180. Total data points for all specimens: 390. The table and explanation were taken from the original, Anderson et al.[330]

Method	Phage	Expected titer (PFU/mL)			Phage-specific CVS		Overall CVS	
		1×10^8	1×10^9	1×10^{10}	Range	Average	Range	Average
PA	ECML-117	0.11–0.23	0.12–0.14	0.12–0.24	0.11–0.24	0.15		
	List-36	0.14–0.26	0.13–0.23	0.06–0.08	0.06–0.26	0.14	0.06–0.26	0.15
	YpP-G	0.22	0.12	0.06	0.06–0.22	0.13		
NS	ECML-117	0.22	0.28	0.30	0.22–0.30	0.27		
	List-36	0.27	0.86[a]	0.27	0.27	0.27	0.22–0.35	0.28
	YpP-G	0.30	0.28	0.35	0.28–0.35	0.31		
QPCR	ECML-117	0.09	0.09	0.08	0.08–0.09	0.09		
	List-36	0.20	0.11	0.19	0.11–0.20	0.17	0.08–0.20	0.13
	YpP-G	0.15	0.09	0.15	0.09–0.15	0.13		

7.5.4 TRPS related information.

7.5.4.1 Relevant statistics

The following table, Table 32, gives the main statistics concerning the concentration based sizing method by TRPS.

Table 32. Concentration report concerning the TRPS measurements. Samples are: Isis (phCLP), Orion (recCLP); non-Decorated: -deco, Decorated: +deco. *Abbreviations: Std Dev: standard deviation of the mean; the D10, D50 (median) and D90: the 10th, 50th and 90th percentile; RSD: Relative standard deviation; Dil. Fact. : dilution factor; Raw Conc. (partc./ mL): raw initial concentration (particles per mL).*

Sample Name	Mode (nm)	Mean (nm)	Std Dev (nm)	Min (nm)	D10 (nm)	D50 (nm)	D90 (nm)	Max (nm)	Span	RSD (%)	Dil. fact.	Raw Conc. (partc./ mL)
Isis -Deco (1000)	89	94	8.8	80	85	93	106	166	0.23	9.4	1000	5.15 E+13
Isis -Deco (2000)	86	93	8.4	78	84	92	104	136	0.22	9.0	2000	5.47 E+13
Isis +Deco	90	95	8.5	70	87	93	106	159	0.20	8.9	2000	7.20 E+13
Orion -Deco	87	97	16.3	73	85	94	109	293	0.26	16.8	2000	2.66 E+13
Orion +Deco	97	99	12.5	70	88	97	112	225	0.25	12.6	1000	6.96 E+12

References

1. Baltimore, D. Expression of animal virus genomes. *Bacteriol. Rev.* **35**, 235–241 (1971).
2. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**, D750–D753 (2010).
3. Roos, W. H., Ivanovska, I. L., Evilevitch, A. & Wuite, G. J. L. Viral capsids: Mechanical characteristics, genome packaging and delivery mechanisms. *Cell. Mol. Life Sci.* **64**, 1484 (2007).
4. Mazzone, H. M. *CRC Handbook of Viruses: Mass-Molecular Weight Values and Related Properties*. (CRC Press, 1998).
5. Caspar, D. L. D. & Klug, A. Physical Principles in the Construction of Regular Viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
6. Ferguson, N. M. *Non-Pharmaceutical Interventions to Reduce COVID-19 Mortality and Healthcare Demand*. 1–20 https://imperial-college-london.relayto.com/e/non-pharmaceutical-interventions-to-reduce-covid-19-mortality-and-healthcare-demand-z0fx4eftbefl7 (2020).
7. Ward, S., Lindsley, A., Courter, J. & Assa'ad, A. Clinical testing for COVID-19. *J. Allergy Clin. Immunol.* **146**, 23–34 (2020).
8. García-Fiñana, M. & Buchan, I. E. Rapid antigen testing in COVID-19 responses. *Science* **372**, 571–572 (2021).
9. Marston, D. A. *et al.* Next generation sequencing of viral RNA genomes. *BMC Genomics* **14**, 444 (2013).
10. Marra, M. A. *et al.* The Genome Sequence of the SARS-Associated Coronavirus. *Science* **300**, 1399–1404 (2003).
11. Mina, M. J., Parker, R. & Larremore, D. B. Rethinking Covid-19 Test Sensitivity — A Strategy for Containment. *N. Engl. J. Med.* **383**, e120 (2020).
12. Abbasi, J. The Promise and Peril of Antibody Testing for COVID-19. *JAMA* **323**, 1881–1883 (2020).
13. Elnifro, E. M., Ashshi, A. M., Cooper, R. J. & Klapper, P. E. Multiplex PCR: Optimization and Application in Diagnostic Virology. *Clin. Microbiol. Rev.* **13**, 559–570 (2000).

14. Yang, A. *et al.* Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA. *Front. Bioeng. Biotechnol.* **8**, 1032 (2020).
15. Simmonds, P. 2015. Methods for virus classification and the challenge of incorporating metagenomic sequence data. *J. Gen. Virol.* **96**, 1193–1206 (2015).
16. Simmonds, P. & Aiewsakun, P. Virus classification – where do you draw the line? *Arch. Virol.* **163**, 2037–2046 (2018).
17. Simmonds, P. *et al.* Four principles to establish a universal virus taxonomy. *PLOS Biol.* **21**, e3001922 (2023).
18. Leney, A. C. & Heck, A. J. R. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* **28**, 5–13 (2017).
19. Tamara, S., den Boer, M. A. & Heck, A. J. R. High-Resolution Native Mass Spectrometry. *Chem. Rev.* (2021) doi:10.1021/acs.chemrev.1c00212.
20. Hentz, S. & Masselon, C. Device for determining the mass of a particle in suspension or in solution in a fluid. (2016).
21. Dominguez-Medina, S. *et al.* Neutral mass spectrometry of virus capsids above 100 megadaltons with nanomechanical resonators. *Science* **362**, 918–922 (2018).
22. Sattar, S. A. & Ijaz, M. K. Spread of viral infections by aerosols. *Crit. Rev. Environ. Control* **17**, 89–131 (1987).
23. Virus | Definition, Structure, & Facts | Britannica. *Encyclopaedia Britannica* https://www.britannica.com/science/virus (2023).
24. Simon, C. E. The Filterable Viruses and their Nature. *Sci. Mon.* **23**, 407–413 (1926).
25. Microbiology Society. Are viruses alive? https://microbiologysociety.org/publication/past-issues/what-is-life/article/are-viruses-alive-what-is-life.html (2016).
26. Ryu, W.-S. Virus Life Cycle. *Mol. Virol. Hum. Pathog. Viruses* 31–45 (2017) doi:10.1016/B978-0-12-800838-6.00003-5.
27. Jones, J. E., Le Sage, V. & Lakdawala, S. S. Viral and host heterogeneity and their effects on the viral life cycle. *Nat. Rev. Microbiol.* **19**, 272–282 (2021).

28. Molenberghs, F., Bogers, J. J. & De Vos, W. H. Confined no more: Viral mechanisms of nuclear entry and egress. *Int. J. Biochem. Cell Biol.* **129**, 105875 (2020).
29. Virion | Capsid, Envelope & Nucleic Acid | Britannica. https://www.britannica.com/science/virion (2023).
30. Microbiology Society. Viruses | What is microbiology? https://microbiologysociety.org/why-microbiology-matters/what-is-microbiology/viruses.html.
31. Louten, J. Chapter 2 - Virus Structure and Classification. in *Essential Human Virology* (ed. Louten, J.) 19–29 (Academic Press, Boston, 2016). doi:10.1016/B978-0-12-800947-5.00002-8.
32. Oxford Dictionary of Biochemistry and Molecular Biology. encapsidate. *Oxford Reference* https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095750633 doi:10.1093/oi/authority.20110803095750633.
33. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).
34. Baltimore, D. Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature* **226**, 1209–1211 (1970).
35. Lefkowitz, E. J. *et al.* Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **46**, D708–D717 (2018).
36. Gibbs, R. A. The Human Genome Project changed everything. *Nat. Rev. Genet.* **21**, 575–576 (2020).
37. Simmonds, P. *et al.* Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **15**, 161–168 (2017).
38. Pompei, S., Loreto, V. & Tria, F. Phylogenetic Properties of RNA Viruses. *PLoS ONE* **7**, e44849 (2012).
39. Lošdorfer Božič, A., Šiber, A. & Podgornik, R. Statistical analysis of sizes and shapes of virus capsids and their resulting elastic properties. *J. Biol. Phys.* **39**, 215–228 (2013).
40. Šiber, A., Božič, A. L. & Podgornik, R. Energies and pressures in viruses: contribution of nonspecific electrostatic interactions. *Phys Chem Chem Phys* **14**, 3746–3765 (2012).
41. Doucette, A. A. & Chisholm, R. A. Molecular-Formula Determination through Accurate-Mass Analysis: A Forensic Investigation. *J. Chem. Educ.* **96**, 1458–1464 (2019).

42. Shuken, S. R. An Introduction to Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **22**, 2151–2171 (2023).
43. He, F., Emmett, M. R., Håkansson, K., Hendrickson, C. L. & Marshall, A. G. Theoretical and Experimental Prospects for Protein Identification Based Solely on Accurate Mass Measurement. *J. Proteome Res.* **3**, 61–67 (2004).
44. Zubarev, R. & Mann, M. On the Proper Use of Mass Accuracy in Proteomics *. *Mol. Cell. Proteomics* **6**, 377–381 (2007).
45. Veenstra, T. D., Martinović, S., Anderson, G. A., Paša-Tolić, L. & Smith, R. D. Proteome analysis using selective incorporation of isotopically labeled amino acids. *J. Am. Soc. Mass Spectrom.* **11**, 78–82 (2000).
46. Shortreed, M. R. *et al.* Elucidating Proteoform Families from Proteoform Intact-Mass and Lysine-Count Measurements. *J. Proteome Res.* **15**, 1213–1221 (2016).
47. Hoskins, L. M. Mass as Quantity of Matter. *Science* **42**, 340–341 (1915).
48. Sear, R. P., Pagonabarraga, I. & Flaus, A. Life at the mesoscale: the self-organised cytoplasm and nucleoplasm. *BMC Biophys.* **8**, 4 (2015).
49. Matthews, R. E. F. A Classification of Virus Groups Based on the Size of the Particle in Relation to Genome Size. *J. Gen. Virol.* **27**, 135–149 (1975).
50. Bahr, G. F., Foster, W. D., Peters, D. & Zeitler, E. H. Variability of dry mass as a fundamental biological property demonstrated for the case of Vaccinia virions. *Biophys. J.* **29**, 305–314 (1980).
51. Green, M. & Piña, M. Biochemical studies on adenovirus multiplication, VI. Properties of Highly Purified Tumorigenic Human Adenoviruses and their DNA's. *Proc. Natl. Acad. Sci. U. S. A.* **51**, (1964).
52. Microbiology by numbers. *Nat. Rev. Microbiol.* **9**, 628–628 (2011).
53. Geoghegan, J. L., Senior, A. M., Di Giallonardo, F. & Holmes, E. C. Virological factors that increase the transmissibility of emerging human viruses. *Proc. Natl. Acad. Sci.* **113**, 4170–4175 (2016).
54. Ciotti, M. *et al.* The COVID-19 pandemic. *Crit. Rev. Clin. Lab. Sci.* **57**, 365–388 (2020).
55. Wang, C. C. *et al.* Airborne transmission of respiratory viruses. *Science* **373**, eabd9149 (2021).

56. Boncristiani, H. F., Criado, M. F. & Arruda, E. Respiratory Viruses. in *Encyclopedia of Microbiology (Third Edition)* (ed. Schaechter, M.) 500–518 (Academic Press, Oxford, 2009). doi:10.1016/B978-012373944-5.00314-X.

57. Damodharan, K., Arumugam, G. S., Ganesan, S., Doble, M. & Thennarasu, S. A comprehensive overview of vaccines developed for pandemic viral pathogens over the past two decades including those in clinical trials for the current novel SARS-CoV-2. *RSC Adv.* **11**, 20006–20035 (2021).

58. Morens, D. M., Taubenberger, J. K. & Fauci, A. S. Rethinking next-generation vaccines for coronaviruses, influenzaviruses, and other respiratory viruses. *Cell Host Microbe* **31**, 146–157 (2023).

59. Fauquet, C. M. & Fargette, D. International Committee on Taxonomy of Viruses and the 3,142 unassigned species. *Virol. J.* **2**, 64 (2005).

60. Hulo, C. *et al.* ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* **39**, D576–D582 (2011).

61. Brister, J. R., Ako-adjei, D., Bao, Y. & Blinkova, O. NCBI Viral Genomes Resource. *Nucleic Acids Res.* **43**, D571–D577 (2015).

62. *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. (Academic Press, London ; Waltham, MA, 2012).

63. Gupta, A., Akin, D. & Bashir, R. Single virus particle mass detection using microresonators with nanoscale thickness. *Appl. Phys. Lett.* **84**, 1976–1978 (2004).

64. Cann, A. J. RNA viruses. A practical approach. *RNA Viruses Pract. Approach* (2000).

65. Robb, J. A. & Bond, C. W. Coronaviridae. in *Comprehensive Virology: Newly Characterized Vertebrate Viruses* (eds. Fraenkel-Conrat, H. & Wagner, R. R.) 193–247 (Springer US, Boston, MA, 1979). doi:10.1007/978-1-4684-3563-4_3.

66. Baker, T. S., Olson, N. H. & Fuller, S. D. Adding the Third Dimension to Virus Life Cycles: Three-Dimensional Reconstruction of Icosahedral Viruses from Cryo-Electron Micrographs. *Microbiol. Mol. Biol. Rev.* **63**, 61 (1999).

67. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).

68. Wilkins, M. R. *et al.* Detailed peptide characterization using PEPTIDEMASS – a World-Wide-Web-accessible tool. *ELECTROPHORESIS* **18**, 403–408 (1997).

69. Murphy, F. A. *et al.* The International Committee on Taxonomy of Viruses. in *Virus Taxonomy: Classification and Nomenclature of Viruses Sixth Report of the International Committee on Taxonomy of Viruses* (eds. Murphy, F. A. et al.) 509–529 (Springer, Vienna, 1995). doi:10.1007/978-3-7091-6607-9_3.

70. Green, M. & Piña, M. Biochemical studies on adenovirus multiplication: IV. Isolation, purification, and chemical analysis of adenovirus. *Virology* **20**, 199–207 (1963).

71. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).

72. Chaudhari, H. V., Inamdar, M. M. & Kondabagil, K. Scaling relation between genome length and particle size of viruses provides insights into viral life history. *iScience* **24**, 102452 (2021).

73. Cui, J., Schlub, T. E. & Holmes, E. C. An Allometric Relationship between the Genome Length and Virion Volume of Viruses. *J. Virol.* **88**, 6403–6410 (2014).

74. Anderson, N. G. *The Development of Zonal Centrifuges and Ancillary Systems for Tissue Fractionation and Analysis*. (U.S. Department of Health, Education, and Welfare, Public Health Service, National Cancer Institute, 1966).

75. Jaenicke, R. Physical Aspects of the Atmospheric Aerosol. in *Chemistry of the Unpolluted and Polluted Troposphere* (eds. Georgii, H. W. & Jaeschke, W.) 341–373 (Springer Netherlands, Dordrecht, 1982). doi:10.1007/978-94-009-7918-5_14.

76. Ruigrok, R. W. H., Nermut, M. V. & Andree, P. J. The molecular mass of adenovirus type 5 as determined by means of scanning transmission electron microscopy (STEM). *J. Virol. Methods* **9**, 69–78 (1984).

77. Thomas, D., Schultz, P., Steven, A. C. & Wall, J. S. Mass analysis of biological macromolecular complexes by STEM. *Biol. Cell* **80**, 181–192 (1994).

78. Barnes, L. F., Draper, B. E. & Jarrold, M. F. Analysis of Recombinant Adenovirus Vectors by Ion Trap Charge Detection Mass Spectrometry: Accurate Molecular Weight Measurements beyond 150 MDa. *Anal. Chem.* **94**, 1543–1551 (2022).

79. Coshic, K., Maffeo, C., Winogradoff, D. & Aksimentiev, A. The structure and physical properties of a packaged bacteriophage particle. *Nature* 1–10 (2024) doi:10.1038/s41586-024-07150-4.
80. Michalakis, Y. & Blanc, S. The Curious Strategy of Multipartite Viruses. *Annu. Rev. Virol.* **7**, 203–218 (2020).
81. Lampert, F., Bahr, G. F. & Rabson, A. S. Herpes Simplex Virus: Dry Mass. *Science* **166**, 1163–1165 (1969).
82. Wörner, T. P. *et al.* Adeno-associated virus capsid assembly is divergent and stochastic. *Nat. Commun.* **12**, 1642 (2021).
83. Arnaud, C. H. The emerging world of single-molecule mass measurements. *Chemical & Engineering News* (2023).
84. Woo, P. C. Y. *et al.* ICTV Virus Taxonomy Profile: Coronaviridae 2023: This article is part of the ICTV Virus Taxonomy Profiles collection. *J. Gen. Virol.* **104**, (2023).
85. Hierholzer, J. C., Palmer, E. L., Whitfield, S. G., Kaye, H. S. & Dowdle, W. R. Protein composition of coronavirus OC 43. *Virology* **48**, 516–527 (1972).
86. Degueldre, C. Single virus inductively coupled plasma mass spectroscopy analysis: a comprehensive study. *Talanta* 122211 (2021) doi:10.1016/j.talanta.2021.122211.
87. Erdogan, R. T. *et al.* Atmospheric Pressure Mass Spectrometry of Single Viruses and Nanoparticles by Nanoelectromechanical Systems. *ACS Nano* (2022) doi:10.1021/acsnano.1c08423.
88. Verrecchia, E. P. & Trombino, L. Observation of Soils: From the Field to the Microscope. in *A Visual Atlas for Soil Micromorphologists* (eds. Verrecchia, E. P. & Trombino, L.) 1–17 (Springer International Publishing, Cham, 2021). doi:10.1007/978-3-030-67806-7_1.
89. Wörner, T. P., Shamorkina, T. M., Snijder, J. & Heck, A. J. R. Mass Spectrometry-Based Structural Virology. *Anal. Chem.* **93**, 620–640 (2021).
90. Jarrold, M. F. Applications of Charge Detection Mass Spectrometry in Molecular Biology and Biotechnology. *Chem. Rev.* **122**, 7415–7441 (2022).
91. Yang, Z. *et al.* Review and prospect on portable mass spectrometer for recent applications. *Vacuum* **199**, 110889 (2022).

92. Brandt, M. P., Jäger, W., Epple, S., Haferkamp, A. & Schröder, A. SARS-CoV-2 outbreak in medical employees in a large urologic department: spread, containment and outcome. *Am. J. Infect. Control* (2021) doi:10.1016/j.ajic.2021.02.011.

93. Zhao, Y., Feng, Y. & Ma, L. Impacts of human movement and ventilation mode on the indoor environment, droplet evaporation, and aerosol transmission risk at airport terminals. *Build. Environ.* **224**, 109527 (2022).

94. Mann, A. New organization pledges scientific expertise for viral outbreaks. *Nat. Med.* **17**, 394–394 (2011).

95. Downard, K. M. Sequence-Free Phylogenetics with Mass Spectrometry. *Mass Spectrom. Rev.* **41**, 3–14 (2022).

96. Anderson, P. W. More Is Different. *Science* **177**, 393–396 (1972).

97. Kivelson, S. & Kivelson, S. A. Defining emergence in physics. *Npj Quantum Mater.* **1**, 1–2 (2016).

98. Stent, G. S. That Was the Molecular Biology That Was. *Science* **160**, 390–395 (1968).

99. Noble, D. It's time to admit that genes are not the blueprint for life. *Nature* **626**, 254–255 (2024).

100. Martinez Arias, A. *The Master Builder*. (Basic Books, 2022).

101. Horwitz, R. The new cell theory. *Science* **381**, 488–488 (2023).

102. Suttle, C. The viriosphere: the greatest biological diversity on Earth and driver of global processes. *Environ. Microbiol.* **7**, 481–482 (2005).

103. de Oliveira, G. A. & Silva, J. L. Cryo-EM to visualize the structural organization of viruses. *Curr. Opin. Virol.* **49**, 86–91 (2021).

104. Kühlbrandt, W. The Resolution Revolution. *Science* **343**, 1443–1444 (2014).

105. Derewenda, Z. S. Rational Protein Crystallization by Mutational Surface Engineering. *Structure* **12**, 529–535 (2004).

106. Chao, F.-A. & Byrd, R. A. Protein dynamics revealed by NMR relaxation methods. *Emerg. Top. Life Sci.* **2**, 93–105 (2018).

107. Vilas, J. L., Carazo, J. M. & Sorzano, C. O. S. Emerging Themes in CryoEM─Single Particle Analysis Image Processing. *Chem. Rev.* (2022) doi:10.1021/acs.chemrev.1c00850.

108. Cole, J. L., Lary, J. W., P. Moody, T. & Laue, T. M. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. in *Methods in Cell Biology* vol. 84 143–179 (Elsevier, 2008).

109. Scarff, C. A., Fuller, M. J. G., Thompson, R. F. & Iadanza, M. G. Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems. *JoVE J. Vis. Exp.* e57199 (2018) doi:10.3791/57199.

110. Petrova, I. D., Zaitsev, B. N. & Taranov, O. S. Concentration of viruses and electron microscopy. *Vavilov J. Genet. Breed.* **22**, 276–283 (2020).

111. Y, Z. *et al.* Single-particle cryo-EM: alternative schemes to improve dose efficiency. *J. Synchrotron Radiat.* **28**, 1343–1356 (2021).

112. Lyonnais, S. *et al.* Atomic force microscopy analysis of native infectious and inactivated SARS-CoV-2 virions. *Sci. Rep.* **11**, 11885 (2021).

113. Gourdelier, M. *et al.* Optimized production and fluorescent labeling of SARS-CoV-2 virus-like particles. *Sci. Rep.* **12**, 14651 (2022).

114. Collinson, L. M. *et al.* Volume EM: a quiet revolution takes shape. *Nat. Methods* 1–6 (2023) doi:10.1038/s41592-023-01861-8.

115. Young, G. *et al.* Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423–427 (2018).

116. Milewska, A. *et al.* Mass Spectrometry in Virological Sciences. *Mass Spectrom. Rev.* **39**, 499–522 (2020).

117. Leifels, M. *et al.* Capsid integrity quantitative PCR to determine virus infectivity in environmental and food applications – A systematic review. *Water Res. X* **11**, 100080 (2021).

118. Troxell, B. *et al.* Partial genome content within rAAVs impacts performance in a cell assay-dependent manner. *Mol. Ther. - Methods Clin. Dev.* **30**, 288–302 (2023).

119. Chen, M. Quantifying Viral Vector Attributes with Light Scattering. *Viral Vectors* 18 (2020).

120. McIntosh, N. L. *et al.* Comprehensive characterization and quantification of adeno associated vectors by size exclusion chromatography and multi angle light scattering. *Sci. Rep.* **11**, 3012 (2021).

121. Gimpel, A. L. *et al.* Analytical methods for process and product characterization of recombinant adeno-associated virus-based gene therapies. *Mol. Ther. - Methods Clin. Dev.* **20**, 740–754 (2021).

122. Cusack, S. Neutron Scattering Studies of Virus Structure. in *Neutrons in Biology* (ed. Schoenborn, B. P.) 173–188 (Springer US, Boston, MA, 1984). doi:10.1007/978-1-4899-0375-4_10.

123. Gräwert, T. W. & Svergun, D. I. Structural Modeling Using Solution Small-Angle X-ray Scattering (SAXS). *J. Mol. Biol.* **432**, 3078–3092 (2020).

124. Khaykelson, D. & Raviv, U. Studying viruses using solution X-ray scattering. *Biophys. Rev.* **12**, 41–48 (2020).

125. Bahr, G. F., Engler, W. F. & Mazzone, H. M. Determination of the Mass of Viruses by Quantitative Electron Microscopy. *Q. Rev. Biophys.* **9**, 459–489 (1976).

126. Biel, S. S. & Gelderblom, H. R. Diagnostic electron microscopy is still a timely and rewarding method. *J. Clin. Virol.* **13**, 105–119 (1999).

127. Mühlfeld, C. *et al.* Visualization and quantitative analysis of nanoparticles in the respiratory tract by transmission electron microscopy. *Part. Fibre Toxicol.* **4**, 11 (2007).

128. Goldsmith, C. S. & Miller, S. E. Modern Uses of Electron Microscopy for Detection of Viruses. *Clin. Microbiol. Rev.* **22**, 552–563 (2009).

129. Laue, M. Chapter 1 - Electron Microscopy of Viruses. in *Methods in Cell Biology* (ed. Müller-Reichert, T.) vol. 96 1–20 (Academic Press, 2010).

130. Malenovska, H. Virus quantitation by transmission electron microscopy, TCID50, and the role of timing virus harvesting: A case study of three animal viruses. *J. Virol. Methods* **191**, 136–140 (2013).

131. Freeman, R. & Leonard, K. R. Comparative mass measurement of biological macromolecules by scanning transmission electron microscopy. *J. Microsc.* **122**, 275–286 (1981).

132. Vogt, V. M. & Simon, M. N. Mass Determination of Rous Sarcoma Virus Virions by Scanning Transmission Electron Microscopy. *J. Virol.* **73**, 7050–7055 (1999).

133. Müller, S. A. & Engel, A. Structure and mass analysis by scanning transmission electron microscopy. *Micron* **32**, 21–31 (2001).

134. Goldsbury, C. *et al.* Amyloid structure and assembly: Insights from scanning transmission electron microscopy. *J. Struct. Biol.* **173**, 1–13 (2011).

135. Burlaud-Gaillard, J. *et al.* Correlative Scanning-Transmission Electron Microscopy Reveals that a Chimeric Flavivirus Is Released as Individual Particles in Secretory Vesicles. *PLOS ONE* **9**, e93573 (2014).

136. Schuster, T. M. & Toedt, J. M. New revolutions in the evolution of analytical ultracentrifugation. *Curr. Opin. Struct. Biol.* **6**, 650–658 (1996).

137. Lebowitz, J., Lewis, M. S. & Schuck, P. Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Sci.* **11**, 2067–2079 (2002).

138. Robert, M.-A. *et al.* Manufacturing of recombinant adeno-associated viruses using mammalian expression platforms. *Biotechnol. J.* **12**, 1600193 (2017).

139. Heck, A. J. R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **5**, 927–933 (2008).

140. Keifer, D. Z., Pierson, E. E. & Jarrold, M. F. Charge detection mass spectrometry: weighing heavier things. *The Analyst* **142**, 1654–1671 (2017).

141. Barnes, L. F., Draper, B. E., Chen, Y.-T., Powers, T. W. & Jarrold, M. F. Quantitative Analysis of Genome Packaging in Recombinant AAV Vectors by Charge Detection Mass Spectrometry. *Mol. Ther. - Methods Clin. Dev.* (2021) doi:10.1016/j.omtm.2021.08.002.

142. Miller, L. M., Bond, K. M., Draper, B. E. & Jarrold, M. F. Characterization of Classical Vaccines by Charge Detection Mass Spectrometry. *Anal. Chem.* (2021) doi:10.1021/acs.analchem.1c01893.

143. Kondylis, P., Schlicksup, C. J., Zlotnick, A. & Jacobson, S. C. Analytical Techniques to Characterize the Structure, Properties, and Assembly of Virus Capsids. *Anal. Chem.* **91**, 622–636 (2019).

144. Weiss, V. U. *et al.* Virus-like particle size and molecular weight/mass determination applying gas-phase electrophoresis (native nES GEMMA). *Anal. Bioanal. Chem.* **411**, 5951–5962 (2019).

145. Weiss, V. U., Zoratto, S., Marchetti-Deschmann, M. & Allmaier, G. Virus-like particle (VLP) molecular weight determination via gas-phase electrophoresis (nES DMA instrumentation): a comparison to native nESI MS. 1.

146. Bacher, G. *et al.* Charge-reduced nano electrospray ionization combined with differential mobility analysis of peptides, proteins, glycoproteins, noncovalent protein complexes and viruses. *J. Mass Spectrom.* **36**, 1038–1052 (2001).

147. Thomas, J. J., Bothner, B., Traina, J., Benner, W. H. & Siuzdak, G. Electrospray ion mobility spectrometry of intact viruses. *Spectroscopy* **18**, 31–36 (2004).

148. Kapellios, E. A. *et al.* Using nanoelectrospray ion mobility spectrometry (GEMMA) to determine the size and relative molecular mass of proteins and protein assemblies: a comparison with MALLS and QELS. *Anal. Bioanal. Chem.* **399**, 2421–2433 (2011).

149. Clement, K. *et al.* Requirements and attributes of nano-resonator mass spectrometry for the analysis of intact viral particles. *Anal. Bioanal. Chem.* **413**, 7147–7156 (2021).

150. Lai, S.-H. *et al.* Characterizing Nanoparticle Mass Distributions Using Charge-Independent Nanoresonator Mass Spectrometry. *J. Phys. Chem. C* 8 (2023).

151. Lai, S.-H., Maclot, S., Antoine, R. & Masselon, C. D. Advances in Single Particle Mass Analysis. *Mass Spectrom. Rev.* **n/a**, (2024).

152. Thundat, T., Wachter, E. A., Sharp, S. L. & Warmack, R. J. Detection of Mercury Vapor Using Resonating Microcantilevers. *Appl. Phys. Lett.* **66**, 1695–1697 (1995).

153. Gupta, A., Akin, D. & Bashir, R. Single virus particle mass detection using microresonators with nanoscale thickness. *Appl. Phys. Lett.* **84**, 1976–1978 (2004).

154. Ilic, B., Yang, Y. & Craighead, H. G. Virus detection using nanoelectromechanical devices. *Appl. Phys. Lett.* **85**, 2604–2606 (2004).

155. Ekinci, K. L., Huang, X. M. H. & Roukes, M. L. Ultrasensitive nanoelectromechanical mass detection. *Appl. Phys. Lett.* **84**, 4469–4471 (2004).

156. Yang, Y. T., Callegari, C., Feng, X. L., Ekinci, K. L. & Roukes, M. L. Zeptogram-Scale Nanomechanical Mass Sensing. *Nano Lett.* **6**, 583–586 (2006).

157. Chaste, J. *et al.* A nanomechanical mass sensor with yoctogram resolution. *Nat. Nanotechnol.* **7**, 301–304 (2012).

158. Naik, A. K., Hanay, M. S., Hiebert, W. K., Feng, X. L. & Roukes, M. L. Towards single-molecule nanomechanical mass spectrometry. *Nat. Nanotechnol.* **4**, 445–450 (2009).

159. Hanay, M. S. *et al.* Single-protein nanomechanical mass spectrometry in real time. *Nat. Nanotechnol.* **7**, 602–608 (2012).

160. Sage, E. *et al.* Neutral Particle Mass Spectrometry with Nanomechanical Systems. *Nat. Commun.* **6**, 6482 (2015).

161. Sage, E. *et al.* Single-particle mass spectrometry with arrays of frequency-addressed nanomechanical resonators. *Nat. Commun.* **9**, 3283 (2018).

162. Wilm, M. Principles of Electrospray Ionization. *Moll. Cell. Proteomics* **10**, 8 (2011).

163. Qi, A. *et al.* Miniature inhalation therapy platform using surface acoustic wave microfluidic atomization. *Lab. Chip* **9**, 2184–2193 (2009).

164. Wang, X., Kruis, F. E. & McMurry, P. H. Aerodynamic Focusing of Nanoparticles: I. Guidelines for Designing Aerodynamic Lenses for Nanoparticles. *Aerosol Sci. Technol.* **39**, 611–623 (2005).

165. Reynaud, A. *et al.* Compact and modular system architecture for a nano-resonator-mass spectrometer. *Front. Chem.* **11**, (2023).

166. Gross, J. H. Electrospray Ionization. in *Mass Spectrometry: A Textbook* (ed. Gross, J. H.) 721–778 (Springer International Publishing, Cham, 2017). doi:10.1007/978-3-319-54398-7_12.

167. Kooij, S., Astefanei, A., Corthals, G. L. & Bonn, D. Size distributions of droplets produced by ultrasonic nebulizers. *Sci. Rep.* **9**, 6128 (2019).

168. Ochner, H. *et al.* Low-energy electron holography imaging of conformational variability of single-antibody molecules from electrospray ion beam deposition. *Proc. Natl. Acad. Sci.* **118**, e2112651118 (2021).

169. Ochner, H. *et al.* Imaging conformations of holo- and apo-transferrin on the single-molecule level by low-energy electron holography. *Sci. Rep.* **13**, 10241 (2023).

170. Esser, T. K. *et al.* Cryo-EM of soft-landed β-galactosidase: Gas-phase and native structures are remarkably similar. *Sci. Adv.* **10**, eadl4628 (2024).

171. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **246**, 64–71 (1989).

172. Prabhu, G. R. D., Williams, E. R., Wilm, M. & Urban, P. L. Mass spectrometry using electrospray ionization. *Nat. Rev. Methods Primer* **3**, 1–22 (2023).

173. Fenn, J. The Nobel Prize in Chemistry 2002. *NobelPrize.org* https://www.nobelprize.org/prizes/chemistry/2002/fenn/lecture/ (2002).

174. Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **85**, 2–9 (2013).

175. Siuzdak, G. *et al.* Mass spectrometry and viral analysis. *Chem. Biol.* **3**, 45–48 (1996).

176. Bothner, B. & Siuzdak, G. Electrospray Ionization of a Whole Virus: Analyzing Mass, Structure, and Viability. *ChemBioChem* **5**, 258–260 (2004).

177. Yang, Z. *et al.* Electrospray-assisted cryo-EM sample preparation to mitigate interfacial effects. *Nat. Methods* 1–10 (2024) doi:10.1038/s41592-024-02247-0.

178. Heron, S. R., Wilson, R., Shaffer, S. A., Goodlett, D. R. & Cooper, J. M. Surface Acoustic Wave Nebulization of Peptides As a Microfluidic Interface for Mass Spectrometry. *Anal. Chem.* **82**, 3985–3989 (2010).

179. Kiontke, A. *et al.* Surface acoustic wave nebulization improves compound selectivity of low-temperature plasma ionization for mass spectrometry. *Sci. Rep.* **11**, 2948 (2021).

180. Hentz, S. & Masselon, C. Device for determining the mass of a particle in suspension or in solution in a fluid. (2016).

181. Malvar, O. *et al.* Mass and stiffness spectrometry of nanoparticles and whole intact bacteria by multimode nanomechanical resonators. *Nat. Commun.* **7**, 13452 (2016).

182. Dominguez-Medina, S. *et al.* Neutral mass spectrometry of virus capsids above 100 megadaltons with nanomechanical resonators. *Science* **362**, 918–922 (2018).

183. Sage, E. *et al.* Frequency-addressed NEMS arrays for mass and gas sensing applications. *2013 Transducers Eurosensors XXVII 17th Int. Conf. Solid-State Sens. Actuators Microsyst. TRANSDUCERS EUROSENSORS 2013* 665–668 (2013) doi:10.1109/Transducers.2013.6626854.

184. Clement, K. *et al.* Requirements and attributes of nano-resonator mass spectrometry for the analysis of intact viral particles. *Anal. Bioanal. Chem.* (2021) doi:10.1007/s00216-021-03511-4.

185. Mile, E. *et al.* In-plane nanoelectromechanical resonators based on silicon nanowire piezoresistive detection. *Nanotechnology* **21**, 165504 (2010).

186. Sansa, M. *et al.* Optomechanical mass spectrometry. *Nat. Commun.* **11**, 3781 (2020).

187. Lamberti, F.-R. *et al.* Real-Time Sensing with Multiplexed Optomechanical Resonators. *Nano Lett.* **22**, 1866–1873 (2022).

188. Sanz-Jiménez, A. *et al.* High-throughput determination of dry mass of single bacterial cells by ultrathin membrane resonators. *Commun. Biol.* **5**, 1227 (2022).

189. Fortin, T. *et al.* A nonlinear model for nano-electro mechanical mass sensing signals processing. *IEEE Sens. J.* 1–1 (2021) doi:10.1109/JSEN.2021.3103713.

190. Hanay, M. S. *et al.* Single-protein nanomechanical mass spectrometry in real time. *Nat. Nanotechnol.* **7**, 602–608 (2012).

191. Allan, D. W. Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **34**, 647–654 (1987).

192. Sansa, M. *et al.* Frequency fluctuations in silicon nanoresonators. *Nat. Nanotechnol.* **11**, 552–558 (2016).

193. Fortin, T. *et al.* Modeling Nanoelectromechanical Resonator Signals for Experimental Mass Measurements: A Total Variation Formulation. *IEEE Access* **12**, 187919–187946 (2024).

194. Elahi, N., Kamali, M. & Baghersad, M. H. Recent biomedical applications of gold nanoparticles: A review. *Talanta* **184**, 537–556 (2018).

195. Mourdikoudis, S., M. Pallares, R. & K. Thanh, N. T. Characterization techniques for nanoparticles: comparison and complementarity upon studying nanoparticle properties. *Nanoscale* **10**, 12871–12934 (2018).

196. Erdogan, R. T. *et al.* Atmospheric Pressure Mass Spectrometry of Single Viruses and Nanoparticles by Nanoelectromechanical Systems. *ACS Nano* **16**, 3821–3833 (2022).

197. Kaynak, B. E., Alkhaled, M., Kartal, E., Yanik, C. & Hanay, M. S. Atmospheric-Pressure Mass Spectrometry by Single-Mode Nanoelectromechanical Systems. *Nano Lett.* (2023) doi:10.1021/acs.nanolett.3c02343.

198. Li, M., Tang, H. X. & Roukes, M. L. Ultra-sensitive NEMS-based cantilevers for sensing, scanned probe and very high-frequency applications. *Nat. Nanotechnol.* **2**, 114–120 (2007).

199. Gavan, K. B., van der Heijden, J., van der Drift, E. W. J. M. & van der Zant, H. S. J. Effect of pressure on the Q factor and the resonance frequency of SiN microcantilevers. in *2009 4th IEEE International Conference on Nano/Micro Engineered and Molecular Systems* 380–384 (2009). doi:10.1109/NEMS.2009.5068600.

200. Trzpil, W. *et al.* Analytic Optimization of Cantilevers for Photoacoustic Gas Sensor with Capacitive Transduction. *Sensors* **21**, 1489 (2021).

201. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518–529 (2015).

202. Clement, K. Multi-level characterization of bacteriophage T5 capsids and capsid-like particles using mass spectrometry. (Université Grenoble Alpes [2020-....], 2021).

203. Chen, J. *et al.* Heating-Rate and Particle-Size Effects on Melting Process of Au Nanoparticles. *J. Phys. Chem. C* **124**, 7414–7420 (2020).

204. Nemes, P., Marginean, I. & Vertes, A. Spraying Mode Effect on Droplet Formation and Ion Chemistry in Electrosprays. *Anal. Chem.* **79**, 3105–3116 (2007).

205. Lošdorfer Božič, A., Šiber, A. & Podgornik, R. How simple can a model of an empty viral capsid be? Charge distributions in viral capsids. *J. Biol. Phys.* **38**, 657–671 (2012).

206. Opitz, A., Scherge, M., Ahmed, S. I.-U. & Schaefer, J. A. A comparative investigation of thickness measurements of ultra-thin water films by scanning probe techniques. *J. Appl. Phys.* **101**, 064310 (2007).

207. Andoh, Y. *et al.* All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution. *J. Chem. Phys.* **141**, 165101 (2014).

208. Wang, X. & McMurry, P. H. An experimental study of nanoparticle focusing with aerodynamic lenses. *Int. J. Mass Spectrom.* **258**, 30–36 (2006).

209. Ducrot, N. Utilisation de la capside du bactériophage T5 comme nanoparticule d'intérêt thérapeutique. (université Paris-Saclay, 2023).

210. Barron, M. Phage Therapy: Past, Present and Future. *ASM.org* https://asm.org:443/Articles/2022/August/Phage-Therapy-Past,-Present-and-Future (2022).

211. Huet, A., Conway, J. F., Letellier, L. & Boulanger, P. In Vitro Assembly of the T=13 Procapsid of Bacteriophage T5 with Its Scaffolding Domain. *J. Virol.* **84**, 9350–9358 (2010).

212. Zivanovic, Y. *et al.* Insights into Bacteriophage T5 Structure from Analysis of Its Morphogenesis Genes and Protein Components. *J. Virol.* **88**, 1162–1174 (2014).

213. Arnaud, C.-A. *et al.* Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. *Nat. Commun.* **8**, 1953 (2017).

214. Linares, R. *et al.* Structural basis of bacteriophage T5 infection trigger and E. coli cell wall perforation. *Sci. Adv.* **9**, eade9674 (2023).

215. Braun, V. FhuA (TonA), the Career of a Protein. *J. Bacteriol.* **191**, 3431–3436 (2009).

216. Le Mercier, P., Hulo, C., Masson, P. & de Castro, E. [ViralZone : digitalizing for knowledge sharing]. *Virol. Montrouge Fr.* **24**, 437–440 (2020).

217. Preux, O. *et al.* A Two-State Cooperative Expansion Converts the Procapsid Shell of Bacteriophage T5 into a Highly Stable Capsid Isomorphous to the Final Virion Head. *J. Mol. Biol.* **425**, 1999–2014 (2013).

218. Huet, A., Duda, R. L., Hendrix, R. W., Boulanger, P. & Conway, J. F. Correct Assembly of the Bacteriophage T5 Procapsid Requires Both the Maturation Protease and the Portal Complex. *J. Mol. Biol.* **428**, 165–181 (2016).

219. Vernhes, E. *et al.* High affinity anchoring of the decoration protein pb10 onto the bacteriophage T5 capsid. *Sci. Rep.* **7**, 41662 (2017).

220. Huet, A., Duda, R. L., Boulanger, P. & Conway, J. F. Capsid expansion of bacteriophage T5 revealed by high resolution cryoelectron microscopy. *Proc. Natl. Acad. Sci.* **116**, 21037–21046 (2019).

221. Boulanger, P., Vernhes, E., Ducrot, N., Larbi Cherif, L. & Benihoud, K. Production and Functionalization of Nanoparticles Derived from Phage T5 and Therapeutic Uses. (2021).

222. Vernhes, E. Maturation de la capside du bactériophage T5 : étude structurale et fonctionnelle de la protéine de décoration pb10. (Université Paris-Saclay (ComUE), 2016).

223. Andersson, K., Björkelund, H. & Malmqvist, M. Antibody-antigen interactions: What is the required time to equilibrium? *Nat. Preced.* 1–1 (2010) doi:10.1038/npre.2010.5218.1.

224. *Viral Molecular Machines*. vol. 726 (Springer US, Boston, MA, 2012).

225. Ebberink, E. H. T. M., Ruisinger, A., Nuebel, M., Thomann, M. & Heck, A. J. R. Assessing production variability in empty and filled adeno-associated viruses by single molecule mass analyses. *Mol. Ther. Methods Clin. Dev.* **27**, 491 (2022).

226. Barnes, L. F. *et al.* Analysis of AAV-Extracted DNA by Charge Detection Mass Spectrometry Reveals Genome Truncations. *Anal. Chem.* **95**, 4310–4316 (2023).

227. Wick, C. H. *et al.* Mass Spectrometry and Integrated Virus Detection System Characterization of MS2 Bacteriophage. *Toxicol. Mech. Methods* **17**, 241–254 (2007).

228. Pogan, R., Dülfer, J. & Uetrecht, C. Norovirus assembly and stability. *Curr. Opin. Virol.* **31**, 59–65 (2018).

229. Nooraei, S. *et al.* Virus-like particles: preparation, immunogenicity and their roles as nanovaccines and drug nanocarriers. *J. Nanobiotechnology* **19**, 59 (2021).

230. Bruinsma, R. F., Wuite, G. J. L. & Roos, W. H. Physics of viral dynamics. *Nat. Rev. Phys.* **3**, 76–91 (2021).

231. Nguyen, M.-C. *et al.* Label-Free Single Nanoparticle Identification and Characterization in Demanding Environment, Including Infectious Emergent Virus. *Small* **20**, 2304564 (2024).

232. Maguire, C. M., Rösslein, M., Wick, P. & Prina-Mello, A. Characterisation of particles in solution – a perspective on light scattering and comparative technologies. *Sci. Technol. Adv. Mater.* **19**, 732–745 (2018).

233. Preux, O. Assemblage et maturation de la capside du bactériophage T5 : analyse des processus d'expansion et de décoration. (Université Paris Sud - Paris XI, 2013).

234. Kim, K., Woo, D., Park, J.-H., Doo, S.-G. & Kim, T. Formation of Silicon Nanoparticles Using SiH4 Pyrolysis at Atmospheric- and Low-Pressure. *J. Korean Phys. Soc.* **54**, 1021–1026 (2009).

235. Kuo, C.-H. *et al.* Low-Resistivity Titanium Nitride Thin Films Fabricated by Atomic Layer Deposition with TiCl4 and Metal–Organic Precursors in Horizontal Vias. *ACS Appl. Electron. Mater.* **5**, 4094–4102 (2023).

236. Behrisch, R. & Eckstein, W. *Sputtering by Particle Bombardment*. vol. 110 (Springer, Berlin, Heidelberg, 2007).

237. Mattox, D. M. *The Foundations of Vacuum Coating Technology*. (Springer, Berlin, Heidelberg, 2003). doi:10.1007/978-3-662-10329-6.

238. Egerton, R. F., Li, P. & Malac, M. Radiation damage in the TEM and SEM. *Micron* **35**, 399–409 (2004).

239. Costa, L. T. *et al.* Chemical treatment of mica for atomic force microscopy can affect biological sample conformation. *Biophys. Chem.* **109**, 63–71 (2004).

240. Valentine, R. C., Shapiro, B. M. & Stadtman, E. R. Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from Escherichia coli. *Biochemistry* **7**, 2143–2152 (1968).

241. Sahiro, K. *et al.* Preyssler-type phosphotungstate is a new family of negative-staining reagents for the TEM observation of viruses. *Sci. Rep.* **12**, 7554 (2022).

242. Scarff, C. A., Fuller, M. J. G., Thompson, R. F. & Iadaza, M. G. Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems. *J. Vis. Exp. JoVE* 57199 (2018) doi:10.3791/57199.

243. Sun, L. Chapter 13 - Asphalt mix homogeneity. in *Structural Behavior of Asphalt Pavements* (ed. Sun, L.) 821–921 (Butterworth-Heinemann, 2016). doi:10.1016/B978-0-12-849908-5.00013-4.

244. Coleman, V. A., Jämting, Å. K., Catchpoole, H. J., Roy, M. & Herrmann, J. Nanoparticles and metrology: a comparison of methods for the determination of particle size distributions. in *Instrumentation, Metrology, and Standards for Nanomanufacturing, Optics, and Semiconductors V* vol. 8105 13–19 (SPIE, 2011).

245. Sausset, R. *et al.* Comparison of interferometric light microscopy with nanoparticle tracking analysis for the study of extracellular vesicles and bacteriophages. *J. Extracell. Biol.* **2**, e75 (2023).

246. Audfray, A. NTA Basics Understanding the Technique - How to Run a Measurement. (2020).

247. Maguire, C. M., Rösslein, M., Wick, P. & Prina-Mello, A. Characterisation of particles in solution – a perspective on light scattering and comparative technologies. *Sci. Technol. Adv. Mater.* **19**, 732–745 (2018).

248. Comfort, N. *et al.* Nanoparticle Tracking Analysis for the Quantification and Size Determination of Extracellular Vesicles. *J. Vis. Exp.* 62447 (2021) doi:10.3791/62447.

249. Kestens, V., Bozatzidis, V., De Temmerman, P.-J., Ramaye, Y. & Roebben, G. Validation of a particle tracking analysis method for the size determination of nano- and microparticles. *J. Nanoparticle Res.* **19**, 271 (2017).

250. Walker, J. G. Improved nano-particle tracking analysis. *Meas. Sci. Technol.* **23**, 065605 (2012).

251. Parsons, M. E. M. *et al.* A Protocol for Improved Precision and Increased Confidence in Nanoparticle Tracking Analysis Concentration Measurements between 50 and 120 nm in Biological Fluids. *Front. Cardiovasc. Med.* **4**, (2017).

252. Shepp, L. A. & Vardi, Y. Maximum Likelihood Reconstruction for Emission Tomography. *IEEE Trans. Med. Imaging* **1**, 113–122 (1982).

253. Dexter, A. R. & Tanner, D. W. Packing Densities of Mixtures of Spheres with Log-normal Size Distributions. *Nat. Phys. Sci.* **238**, 31–32 (1972).

254. Zhao, Y. & Chew, J. W. Effect of lognormal particle size distributions of non-spherical particles on hopper discharge characteristics. *Chem. Eng. Res. Des.* **163**, 230–240 (2020).

255. Heintzenberg, J. Properties of the Log-Normal Particle Size Distribution. *Aerosol Sci. Technol.* **21**, 46–48 (1994).

256. Borah, R. & Verbruggen, S. W. Effect of size distribution, skewness and roughness on the optical properties of colloidal plasmonic nanoparticles. *Colloids Surf. Physicochem. Eng. Asp.* **640**, 128521 (2022).

257. Song, N. W., Park, K. M., Lee, I.-H. & Huh, H. Uncertainty estimation of nanoparticle size distribution from a finite number of data obtained by microscopic analysis. *Metrologia* **46**, 480 (2009).

258. Bienert, R., Emmerling, F. & Thünemann, A. F. The size distribution of 'gold standard' nanoparticles. *Anal. Bioanal. Chem.* **395**, 1651–1660 (2009).

259. B. Akbari, M. Pirhadi Tavandashti & M. Zandrahimi. PARTICLE SIZE CHARACTERIZATION OF NANOPARTICLES – A PRACTICALAPPROACH. *Iran. J. Mater. Sci. Eng.* **8**, 48–56 (2011).

260. Irani, R. R. & Callis, C. F. *Particle Size: Measurement, Interpretation, and Application*. (Wiley, New York, 1963).

261. Etourneau, L., Varoquaux, N. & Burger, T. Unveiling the Links Between Peptide Identification and Differential Analysis FDR Controls by Means of a Practical Introduction to Knockoff Filters. in *Statistical Analysis of Proteomic Data: Methods and Tools* (ed. Burger, T.) 1–24 (Springer US, New York, NY, 2023). doi:10.1007/978-1-0716-1967-4_1.

262. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**, 1017–1031 (2007).

263. Lau, K. W., Jones, A. R., Swainston, N., Siepen, J. A. & Hubbard, S. J. Capture and analysis of quantitative proteomic data. *PROTEOMICS* **7**, 2787–2799 (2007).

264. Wisniewski, J., Hein, M., Cox, J. & Mann, M. A 'Proteomic Ruler' for Protein Copy Number and Concentration Estimation without Spike-in Standards. *Mol. Cell. Proteomics* **13.12**, 3497–3506 (2014).

265. Blein-Nicolas, M. & Zivy, M. Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1864**, 883–895 (2016).

266. Schwanhäusser, B. *et al.* Correction: Corrigendum: Global quantification of mammalian gene expression control. *Nature* **495**, 126–127 (2013).

267. Arike, L. *et al.* Comparison and applications of label-free absolute proteome quantification methods on *Escherichia coli*. *J. Proteomics* **75**, 5437–5448 (2012).

268. Al Shweiki, M. R. *et al.* Assessment of Label-Free Quantification in Discovery Proteomics and Impact of Technological Factors and Natural Variability of Protein Abundance. *J. Proteome Res.* **16**, 1410–1424 (2017).

269. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C. & Geromanos, S. J. Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel ms Acquisition * S. *Mol. Cell. Proteomics* **5**, 144–156 (2006).

270. Scheuch, A. *et al.* Mass Spectral Analyses of Salmonella Myovirus SPN3US Reveal Conserved and Divergent Themes in Proteolytic Maturation of Large Icosahedral Capsids. *Viruses* **15**, 723 (2023).

271. Steinberger, S. *et al.* Targeting the Structural Integrity of Extracellular Vesicles via Nano Electrospray Gas-Phase Electrophoretic Mobility Molecular Analysis (nES GEMMA). *Membranes* **12**, 872 (2022).

272. Obaidullah, M., Bram, S. & Ruyck, J. An Overview of PM Formation Mechanisms from Residential Biomass Combustion and Instruments Using in PM Measurements. **12**, 50 (2018).

273. Wallace, L. Ultrafine Particles: A Review. (2009).

274. Welker, R. W. Chapter 4 - Size Analysis and Identification of Particles. in *Developments in Surface Contamination and Cleaning* (eds. Kohli, R. & Mittal, K. L.) 179–213 (William Andrew Publishing, Oxford, 2012). doi:10.1016/B978-1-4377-7883-0.00004-3.

275. Hogan, C. J., Kettleson, E. M., Ramaswami, B., Chen, D.-R. & Biswas, P. Charge Reduced Electrospray Size Spectrometry of Mega- and Gigadalton Complexes: Whole Viruses and Virus Fragments. *Anal. Chem.* **78**, 844–852 (2006).

276. General Scanning Mobility Particle Sizer (SMPS) 3938. *TSI.com* https://tsi.com/Products/Particle-Sizers/Scanning-Mobility-Particle-Sizer-Spectrometers/General-Scanning-Mobility-Particle-Sizer-(SMPS)-3938.

277. Tadjiki, S. *et al.* Measurement of the Density of Engineered Silver Nanoparticles Using Centrifugal FFF-TEM and Single Particle ICP-MS. *Anal. Chem.* **89**, 6056–6064 (2017).

278. Roudini, M., Patel, B. & Winkler, A. Developments for SAW-based aerosol generation: Miniaturized, cost-efficient, mass-producible, and reproducible systems. *Aerosol Sci. Technol.* **0**, 1–12 (2024).

279. Pati, R., Shevtsov, M. & Sonawane, A. Nanoparticle Vaccines Against Infectious Diseases. *Front. Immunol.* **9**, (2018).

280. Mohsen, M. O., Gomes, A. C., Vogel, M. & Bachmann, M. F. Interaction of Viral Capsid-Derived Virus-Like Particles (VLPs) with the Innate Immune System. *Vaccines* **6**, 37 (2018).

281. Roldão, A., Mellado, M. C. M., Castilho, L. R., Carrondo, M. J. & Alves, P. M. Virus-like particles in vaccine development. *Expert Rev. Vaccines* **9**, 1149–1176 (2010).

282. Mulder, A. M. *et al.* Toolbox for Non-Intrusive Structural and Functional Analysis of Recombinant VLP Based Vaccines: A Case Study with Hepatitis B Vaccine. *PLoS ONE* **7**, e33235 (2012).

283. Wang, J. W. & Roden, R. B. Virus-like particles for the prevention of human papillomavirus-associated malignancies. *Expert Rev. Vaccines* **12**, 129–141 (2013).

284. Mennechet, F. J. D. *et al.* A review of 65 years of human adenovirus seroprevalence. *Expert Rev. Vaccines* **18**, 597–613 (2019).

285. Chang, J. Adenovirus Vectors: Excellent Tools for Vaccine Development. *Immune Netw.* **21**, (2021).

286. Afkhami, S., Kang, A., Jeyanathan, V., Xing, Z. & Jeyanathan, M. Adenoviral-vectored next-generation respiratory mucosal vaccines against COVID-19. *Curr. Opin. Virol.* 101334 (2023) doi:10.1016/j.coviro.2023.101334.

287. Bao, Q. *et al.* Phage-based vaccines. *Adv. Drug Deliv. Rev.* **145**, 40–56 (2019).

288. Vernhes, E. *et al.* Antigen self-anchoring onto bacteriophage T5 capsid-like particles for vaccine design. *Npj Vaccines* **9**, 1–11 (2024).

289. Bolhassani, A., Talebi, S. & Anvar, A. Endogenous and Exogenous Natural Adjuvants for Vaccine Development. *Mini-Rev. Med. Chem.* **17**, 1442–1456 (2017).

290. Smith, P. K. *et al.* Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76–85 (1985).

291. Stetefeld, J., McKenna, S. A. & Patel, T. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.* **8**, 409–427 (2016).

292. Alonso, C. *et al.* Comparison of two size-differentiating air samplers for detecting airborne swine viruses under experimental conditions. *Aerosol Sci. Technol.* **51**, 198–205 (2017).

293. Nikitin, N., Petrova, E., Trifonova, E. & Karpova, O. Influenza Virus Aerosols in the Air and Their Infectiousness. *Adv. Virol.* **2014**, 1–6 (2014).

294. Furcatte, T. *et al.* Optimizing Optomechanical Resonators for Ultra-High-Frequency Timing Applications. in *2024 Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS (DTIP)* 1–6 (2024). doi:10.1109/DTIP62575.2024.10613067.

295. Neumann, A. P. *et al.* A Hybrid Orbitrap-Nanoelectromechanical Systems Approach for the Analysis of Individual, Intact Proteins in Real Time. *Angew. Chem. Int. Ed.* **63**, e202317064 (2024).

296. Çumaku, V., Fortin, T., Hentz, S. & Masselon, C. On the 'phylobaric' classification of human respiratory viruses. Preprint at https://doi.org/10.21203/rs.3.rs-3395557/v1 (2023).

297. Li, S., Roy, P., Travesset, A. & Zandi, R. Why large icosahedral viruses need scaffolding proteins. *Proc. Natl. Acad. Sci.* **115**, 10971–10976 (2018).

298. Wörner, T. P. *et al. Adeno-Associated Virus Capsid Assembly Is Divergent and Stochastic*. http://biorxiv.org/lookup/doi/10.1101/2020.10.09.332619 (2020) doi:10.1101/2020.10.09.332619.

299. Jose, J. & Hafenstein, S. L. Asymmetry in icosahedral viruses. *Curr. Opin. Virol.* **54**, 101230 (2022).

300. Bui, V. K. H., Moon, J.-Y., Chae, M., Park, D. & Lee, Y.-C. Prediction of Aerosol Deposition in the Human Respiratory Tract via Computational Models: A Review with Recent Updates. *Atmosphere* **11**, 137 (2020).

301. Jaenicke, R. The role of organic material in atmospheric aerosols. *Pure Appl. Geophys.* **116**, 283–292 (1978).

302. Finlayson-Pitts, B. J. & Pitts, J. N. CHAPTER 9 - Particles in the Troposphere. in *Chemistry of the Upper and Lower Atmosphere* (eds. Finlayson-Pitts, B. J. & Pitts, J. N.) 349–435 (Academic Press, San Diego, 2000). doi:10.1016/B978-012257060-5/50011-3.

303. Lamminen, E. Accurate measurement of nanoparticle charge, number and size with the ELPI+^{TM} instrument. *J. Phys. Conf. Ser.* **304**, 012064 (2011).

304. Lampert, F., Bahr, G. F. & Rabson, A. S. Dry mass of human viruses. *Naturwissenschaften* **56**, 573–574 (1969).

305. Piña, M. & Green, M. Biochemical studies on adenovirus multiplication. IX. Chemical and base composition analysis of 28 human adenoviruses. *Proc. Natl. Acad. Sci. U. S. A.* **54**, 547–551 (1965).

306. Ruigrok, R. W. H., Andree, P. J., Hooft Van Huysduynen, R. A. M. & Mellema, J. E. Characterization of Three Highly Purified Influenza Virus Strains by Electron Microscopy. *J. Gen. Virol.* **65**, 799–802 (1984).

307. Pierson, E. E., Keifer, David. Z., Asokan, A. & Jarrold, M. F. Resolving Adeno-Associated Viral Particle Diversity With Charge Detection Mass Spectrometry. *Anal. Chem.* **88**, 6718–6725 (2016).

308. Cecchini, S. *et al.* Evidence of Prior Exposure to Human Bocavirus as Determined by a Retrospective Serological Study of 404 Serum Samples from Adults in the United States. *Clin. Vaccine Immunol.* **16**, 597–604 (2009).

309. Stephan M. Koza, & Weibin Chen. Size-Exclusion Chromatography Analysis of Adeno-Associated Virus (AAV) Preparations Using a 450 Å Diol-Bonded BEH Column and Fluorescence Detection. *www.waters.com* http://www.waters.com/waters/library.htm?lid=135047399.

310. Hogle, J. M., Chow, M. & Filman, D. J. Three-dimensional structure of poliovirus at 2.9 A resolution. *Science* **229**, 1358–1365 (1985).

311. Farhana, A. & Khan, Y. S. Biochemistry, Lipopolysaccharide. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2023).

312. Coli Genetic Stock Center. Strain - CR63. https://cgsc.biology.yale.edu/Strain.php?ID=8270.

313. curtis.suplee@nist.gov. Atomic Weights and Isotopic Compositions with Relative Atomic Masses. *NIST* https://www.nist.gov/pml/atomic-weights-and-isotopic-compositions-relative-atomic-masses (2009).

314. Berglund, M. & Wieser, M. E. Isotopic compositions of the elements 2009 (IUPAC Technical Report). *Pure Appl. Chem.* **83**, 397–410 (2011).

315. Wang, M. *et al.* The Ame2012 atomic mass evaluation. *Chin. Phys. C* **36**, 1603–2014 (2012).

316. Possolo, A., van der Veen, A. M. H., Meija, J. & Hibbert, D. B. Interpreting and propagating the uncertainty of the standard atomic weights (IUPAC Technical Report). *Pure Appl. Chem.* **90**, 395–424 (2018).

317. Valegård, K., Unge, T., Montelius, I., Strandberg, B. & Fiers, W. Purification, crystallization and preliminary X-ray data of the bacteriophage MS2. *J. Mol. Biol.* **190**, 587–591 (1986).

318. Tito, M. A., Tars, K., Valegard, K., Hajdu, J. & Robinson, C. V. Electrospray Time-of-Flight Mass Spectrometry of the Intact MS2 Virus Capsid. *J. Am. Chem. Soc.* **122**, 3550–3551 (2000).

319. Szolajska, E. *et al.* The Structural Basis for the Integrity of Adenovirus Ad3 Dodecahedron. *PLoS ONE* **7**, e46075 (2012).

320. Zivanovic, Y. *et al.* Insights into Bacteriophage T5 Structure from Analysis of Its Morphogenesis Genes and Protein Components. *J. Virol.* **88**, 1162–1174 (2014).

321. Puurunen, R. L. A Short History of Atomic Layer Deposition: Tuomo Suntola's Atomic Layer Epitaxy. *Chem. Vap. Depos.* **20**, 332–344 (2014).

322. Ahvenniemi, E. *et al.* Review Article: Recommended reading list of early publications on atomic layer deposition—Outcome of the "Virtual Project on the History of ALD". *J. Vac. Sci. Technol. Vac. Surf. Films* **35**, 010801 (2017).

323. Oviroh, P. O., Akbarzadeh, R., Pan, D., Coetzee, R. A. M. & Jen, T.-C. New development of atomic layer deposition: processes, methods and applications. *Sci. Technol. Adv. Mater.* **20**, 465–496 (2019).

324. Welty, J. *Fundamentals of Momentum, Heat and Mass Transfer, 5th Edition*. (John Wiley & Sons, Inc., United States of America, 2008).

325. Sproul, W. D., Christie, D. J. & Carter, D. C. Control of reactive sputtering processes. *Thin Solid Films* **491**, 1–17 (2005).

326. Gregoire, J. M., Lobovsky, M. B., Heinz, M. F., DiSalvo, F. J. & van Dover, R. B. Resputtering phenomena and determination of composition in codeposited films. *Phys. Rev. B* **76**, 195437 (2007).

327. Raja, P. B., Munusamy, K. R., Perumal, V. & Ibrahim, M. N. M. 5 - Characterization of nanomaterial used in nanobioremediation. in *Nano-Bioremediation : Fundamentals and Applications* (eds. Iqbal, H. M. N., Bilal, M. & Nguyen, T. A.) 57–83 (Elsevier, 2022). doi:10.1016/B978-0-12-823962-9.00037-4.

328. Bouyssié, D. *et al.* Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **36**, 3148–3155 (2020).

329. Fixman, M. Radius of Gyration of Polymer Chains. *J. Chem. Phys.* **36**, 306–310 (2004).

330. Anderson, B. *et al.* Enumeration of bacteriophage particles. *Bacteriophage* **1**, 86–93 (2011).