Characterizing the protein-DNA binding of the HSV-1 terminase by optimizing sample preparation for native mass spectrometry

Dissertation

with the aim of achieving a doctoral degree at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Chemistry

University of Hamburg

by

Alice Frederike Rosa Grün

born in Hamburg

Hamburg, 2025

Thesis reviewer:

Prof. Dr. Charlotte Uetrecht

CSSB Centre for Structural Systems Biology, Deutsches Elektronen-Synchrotron DESY &

Leibniz Institute of Virology (LIV) & University of Lübeck, Notkestraße 85, 22607 Hamburg,

Germany;

Institute of Chemistry and Metabolomics, University of Lübeck, Ratzeburger Allee 160, 23562

Lübeck, Germany.

Prof. Dr. Wolfram Brune

Leibniz Institute of Virology (LIV), Martinistraße 52, 20246 Hamburg, Germany.

Examination board:

Prof. Dr. Kay Grünewald

Centre for Structural Systems Biology, Hamburg, Germany;

Structural Cell Biology of Viruses, Leibniz Institute of Virology (LIV), Hamburg, Germany;

Department of Chemistry, University of Hamburg, Hamburg, Germany

Prof. Dr. Jens Bosse

Hannover Medical School, Institute of Virology, Hanover, Germany Centre for Structural

Systems Biology, Hamburg, Germany;

Cluster of Excellence RESIST (EXC 2155), Hanover Medical School, Hanover, Germany;

Leibniz Institute of Virology (LIV), Hamburg, Germany

Prof. Dr. Maria Riedner

Technology Platform Mass Spectrometry, University of Hamburg, Martin-Luther-King-Platz

6, 20146 Hamburg, Germany

Thesis submission:

10.04.2025

Thesis defense:

13.06.2025

This thesis was prepared from 2020 to 2025 under the supervision of Prof. Dr. Hartmut Schlüter at Universitätsklinikum Hamburg-Eppendorf (UKE) and the University of Hamburg and Prof. Dr. Charlotte Uetrecht at the Leibniz Institute of Virology (LIV)/Centre for Structural Systems Biology (CSSB) and the University of Lübeck in the working group *Dynamics of Viral Structures*, as well as Prof. Dr. Jens Bosse from Hannover Medical School/Centre for Structural Systems Biology (CSSB).

Contents

List of Figures

List of Tables

Publications

Simanjuntak, Y.; Schamoni-Kast, K.; **Grün, A**.; Uetrecht, C.; Scaturro, P., Top-Down and Bottom-Up Proteomics Methods to Study RNA Virus, Biology. Viruses 2021, 13, 668. https://doi.org/10.3390/v13040668

Drees, A.; Ahlers C.; Kehrer, T.; Ehmke, N.; **Grün, A.F.R.**; Uetrecht, C.; Ignatova, Z.; Schumacher, U.; Fischer, M., Automated high-throughput selection of DNA aptamers using a common optical next-generation sequencer, bioRxiv 2024, https://doi.org/10.1101/2024.06.24.600375

Grün, A.F.R.; Said, F.A.; Schamoni-Kast, K.; Damjanovic, T.; Bosse, J.; Uetrecht, C., Fast tracking native mass spectrometry: Skipping over buffer exchange, bioRxiv 2025, https://doi.org/10.1101/2025.02.22.639503

Acronyms

E elution buffer. 102, 104, 106

L lysis buffer. 101–104, 106

WA wash buffer A. 102, 103, 106

WB wash buffer B. 102, 103, 106

W wash buffer. 104

E. coli *Escherichia coli*. 22, 24, 25, 31, 32, 39, 50, 77, 83, 105, 106

K_d dissociation constant. 88–90, 95

ADH alcohol dehydrogenase. 29, 40–42

ATP adenosine triphosphate. 6, 9, 48, 53, 69, 70, 94

AUC area under the curve. 42, 83, 85, 88, 115

BSA bovine serum albumin. 99

bV bed volume. 103

CD circular dichroism. 29, 53, 55, 57, 106

CID collision-induced dissociation. 12, 19–21, 41, 94, 109

cryoEM cryogenic electron microscopy. 47, 65, 66, 69–71, 77, 79, 94

CV collision voltage. 41, 42, 94, 108, 109

cV column volume. 103

DC direct current. 17

DNA deoxyribonucleic acid. 2, 4, 6, 7, 9, 29, 47, 48, 52, 53, 60, 61, 63–66, 68–71, 73–76, 78, 79, 93, 97, 99, 104–106

dsDNA double-stranded deoxyribonucleic acid. 2, 4, 52, 56, 57, 64, 93

EGFP enhanced green fluorescent protein. 27, 83, 87, 88, 90

EM electron microscopy. 77, 78

ESI electrospray ionization. 12–14

FFU fluorescent-forming units. 99, 100

FP fluorescent protein. 27, 28, 30, 81–83, 85–90, 95, 104, 106, 109, 114, 116

FWHM full width at half maximum. 15, 42, 72–74, 76, 78, 114

G4 G-quadruplex. 4, 5, 29, 30, 52, 53, 55–57, 59–61, 63, 64, 93

HCD high-energy collision dissociation. 20, 21, 51, 109

HCMV human cytomegalovirus. 1, 2, 5, 50, 56, 57, 78, 80, 94

HDX hydrogen-deuterium exchange. 63, 94

HEK human embryonic kidney. 22, 101

HSV-1 herpes simplex virus 1. 1, 2, 4, 6, 7, 29, 30, 35, 50–52, 56, 57, 77, 93, 94

HTC heteromeric trimer terminase complex. 6, 33, 35, 38, 40, 45, 47, 48, 51–53, 64, 65, 68–74, 76, 79

HTC$_2$ dimer of the heteromeric trimer terminase complex. 37, 45, 47, 51–53, 68, 69, 71, 72, 74, 94

HTC$_3$ trimer of the heteromeric trimer terminase complex. 45, 47, 51–53, 60, 61, 64, 65, 68–70, 72, 74–76, 79, 93, 94

HTC$_6$ hexamer of the heteromeric trimer terminase complex. 6, 33, 47, 51–53, 69, 70, 94

IM ion mobility. 61, 63

IPTG Isopropyl-β-ᴅ-thiogalactopyranosid. 31, 83, 105

ipTM interface predicted template modeling. 64, 65

KSHV Kaposi's sarcoma-associated herpesvirus. 2, 29, 77, 80, 94

LC-MS liquid chromatography mass spectrometry. 74, 93

mEGFP monomeric enhanced green fluorescent protein. 27, 82, 83, 87–90

MS mass spectrometry. 10, 11, 40, 41, 51, 52, 63, 69, 71, 83, 94

MS^1 single-stage mass spectrometry. 19, 78

MS^2 two-stage mass spectrometry. 12, 16, 19, 20, 51, 78, 93, 94

nESI nano-electrospray ionization. 12–15, 26, 47, 61, 75, 83

nMS native mass spectrometry. 11–14, 22, 26–31, 33, 37, 39, 40, 47, 48, 51, 53, 55–57, 61, 63, 64, 66, 70, 71, 80, 81, 86–88, 93, 94, 114

NPC nuclear pore complex. 2

NTA nitrilotriacetic acid. 102, 103, 106

OSER organized smooth endoplasmatic reticulum. 27, 28, 82, 86, 87

PAE predicted aligned error. 65

PBS phosphate-buffered saline. 99–101

PCR polymerase chain reaction. 97, 104

pTM predicted template modeling. 64, 65

PTMs post-translational modifications. 10, 15, 25, 72, 73, 76, 93

QE-UHMR Q Exactive™ UHMR Hybrid Quadrupole Orbitrap™ mass spectrometer. 20, 21, 41–43, 46, 48, 51, 60, 75, 89, 93, 94, 109, 114

QMA quadrupole mass analyzer. 17–20

RF radio frequency. 17

SDS-PAGE sodium dodecylsulfate polyacrylamide gel electrophoresis. 33, 35, 48, 51, 83

SEC size-exclusion chromatography. 22, 25, 26, 33, 35, 37, 39, 40, 45, 101

SOC super optimal broth with catabolic repressor. 105

ssDNA single-stranded deoxyribonucleic acid. 6, 29, 52, 53, 56, 57, 60, 64, 66, 93

TB Terrific Broth. 105

TMAA trimethylammonium acetate. 55

TOF-MS time-of-flight mass spectrometer. 10, 18, 19

WB western blot. 50

WHO World Health Organization. 1

XL cross-link. 63

Zusammenfassung

Das Herpes Simplex Virus 1 (HSV-1) ist auf den Terminase-Komplex angewiesen. Dieser besteht aus den Proteinen pUL15, pUL28, pUL33 und packt die virale Desoxyribonukleinsäure (DNS) in das Prokapsid während der Replikation. Trotz ihrer Bedeutung als antivirales Ziel sind die Mechanismen, die die Terminase-DNA-Interaktionen steuern, nur unzureichend bekannt. Diese Arbeit untersucht die Stöchiometrie der HSV-1-Terminase und die Rolle des G-Quadruplexes, einer sekundären DNA-Struktur, bei der DNS-Protein-Komplexierung. Mittels nativer Massenspektrometrie wurde die Stöchiometrie des Terminase-Komplexes analysiert und es konnte gezeigt werden, dass virale G-Quadruplex-haltige DNS-Sequenzen nicht spezifisch an die Terminase binden. Während dieser Experimente wurden andere oligomerer Zustände des Terminase-Komplexes beobachtet, als vorherige Veröffentlichungen suggerierten. Anstelle der bisher publizierten Hexamere des heteromeren Trimer-Terminase-Komplexes wurden hauptsächlich Monomere, Dimere und Trimere beobachtet. Aus diesen Ergebnissen wurde ein Modell für ein neues Zwischenprodukt des Terminase-DNA-Komplexes während der Verpackung entworfen (Kapitel 3.3).

Eine große methodische Herausforderung war die geringe Ausbeute und Qualität des gereinigten Terminase-Komplexes für die nMS-Analyse. Um dieses Problem zu lösen, wurde ein schnelles Proteinreinigungsprotokoll entwickelt, das den Proteinverlust minimiert und den zeitintensiven und verlustreichen Pufferaustauschschritt überflüssig macht. Dieser Ansatz, der mit dem Standardprotein Alkoholdehydrogenase (ADH) validiert und anschließend auf die Terminase und verwandte Proteine angewandt wurde, lieferte durchweg qualitativ hochwertige Proben, die für nMS geeignet sind. Die optimierte Methode bietet eine schnelle und effiziente Strategie

für die Untersuchung verschiedenster Proteinkomplexe (Kapitel 3.2).

Neben den Fragestellungen aus der HSV-Virologie wurde das Fast-Track-Protokoll auch eingesetzt, um die Neigung von fluoreszierenden Proteinen zur Dimerisierung zu untersuchen, die häufig in der Mikroskopie verwendet werden. Durch die Analyse ihrer Stöchiometrie mittels nativer Massenspektrometrie lieferte diese Arbeit systematische Einblicke in die Dimerisierung von FP unter nativen Bedingungen. Diese Erkenntnisse erleichtern die Auswahl von FPs für Mikroskopiestudien, um eine zuverlässige Einzelmolekülverfolgung zu ermöglichen (Kapitel 3.5).

In dieser Arbeit werden Virologie, Massenspektrometrie und methodische Innovationen miteinander verknüpft, um das Verständnis des HSV-1-Verpackungs-mechanismus zu verbessern. Hierzu wurden zunächst Arbeitsabläufe bei der Proteinreinigung für native Massenspektrometrie optimiert. Diese wurden außerdem verwendet, um geeignete fluoreszierende Proteine für die Fluoreszenzmikroskopie bezüglich ihrer Dimerisierung zu charakterisieren. Die Ergebnisse unterstreichen die Vielseitigkeit der nativen Massenspektrometrie, insbesondere in Kombination mit dem *fast-track* Protokoll, bei der Untersuchung verschiedener biologischer Systeme und Proteinkomplexe.

Abstract

Herpes simplex virus 1 is dependent on the terminase complex. It consists of the proteins pUL15, pUL28, pUL33 and packs the viral DNA into the procapsid during replication. Despite its potential as an antiviral target, the mechanism of the terminase-DNA interaction is poorly understood. This work investigates the stoichiometry of the herpes simplex virus 1 terminase and the role of the secondary DNA structure G-quadruplex in the DNA-protein complex formation. Using native mass spectrometry, the stoichiometry of the terminase complex was analyzed and it was shown that viral G-quadruplex containing DNA does not bind specifically to the terminase. In these experiments, different oligomeric states of the terminase complex were observed than suggested in previous publications. Instead of hexamers of the heteromeric trimer terminase complex, mainly monomers, dimers and trimers were detected. Based on these results, a model for a new intermediate of the terminase-DNA complex during packaging was designed (**Chapter** 3.3).

A major experimental challenge was the yield and quality of the purified terminase complex being too low for native mass spectrometry analysis. To solve this problem, a fast-track protein purification protocol was developed that minimizes protein loss and eliminates the time-consuming and wasteful buffer exchange step. This approach, validated with the standard protein alcohol dehydrogenase and subsequently applied to terminase and related proteins, consistently yielded high-quality samples suitable for native mass spectrometry. The optimized method provides a rapid and efficient strategy for the analysis of a wide range of protein complexes (**Chapter** 3.2).

In addition to the studies on herpes simplex virus 1, the fast-track protocol was also used to investigate the tendency of fluorescent proteins, frequently used in microscopy,

to dimerize. Through a stoichiometry analysis using native mass spectrometry, this work provided systematic insights into the dimerization of fluorescent proteins under native conditions. These findings facilitate the selection of fluorescent proteins for microscopy studies to ensure reliable single-molecule tracking (**Chapter** 3.5).

In this work, virology, mass spectrometry and method development are combined to improve the understanding of the herpes simplex virus 1 packaging mechanism. First, protein purification workflows were optimized for native mass spectrometry. These workflows were also used to characterize suitable fluorescent proteins for fluorescence microscopy in terms of their dimerization. The results underline the versatility of native mass spectrometry, especially in combination with the fast-track protocol, in the investigation of different biological systems and protein complexes.

1

Introduction

According to the World Health Organization (WHO)[1] "[an] estimated 3.8 billion people under age 50 (64 %) globally [are infected with] herpes simplex virus type 1 (HSV-1) [. . .], the main cause of oral herpes". They also specify that "[most] HSV infections are asymptomatic or unrecognized, but symptoms of herpes include painful blisters or ulcers that can return over time". For immunocompromised or critically ill people, reactivation of herpes simplex virus 1 (HSV-1) is more likely and thus mortality from pneumonia is significantly increased [2–4]. Since the immune system in newborns is not fully developed, they are at risk of contracting the virus, often resulting in fatal outcomes[5]. To combat this virus effectively, the fundamental mechanism of infection and involved viral biomolecules need to be understood. Over the years, multiple research groups from all over the world significantly improved our understanding of the life cycle of HSV-1[6, 7]. To this day, vaccine and drug development is challenging and new potential targets are still studied[8–11]. One drug called Letermovir has been developed to combat human cytomegalovirus (HCMV)[12]. It targets the terminase complex, which is responsible for the packaging of the new procapsid during viral replication[13]. Even if the exact inhibition mechanism of the terminase is still unclear, it could also be a potential drug target in other herpesviruses such as HSV-1. Understanding the structure and detailed mechanism of the termi-

1

nase complex is therefore necessary to progress HSV-1 drug and vaccine development.

1.1.   The herpesvirus DNA packaging motor: Terminase

Herpesvirales is an order of large enveloped double-stranded deoxyribonucleic acid (dsDNA) viruses. They are classified into three families: *Alloherpesviridae*, *Malacoherpesviridae* and *Orthoherpesviridae*. The latter comprises the three human herpesviruses subfamilies: α- (e.g., HSV-1), β- (e.g., HCMV) and γ-*Herpesvirinae* (e.g., Kaposi's sarcoma-associated herpesvirus (KSHV))[6, 14].

1.1.1. The replication cycle of herpesviruses

The herpesvirus virions are composed of an icosahedral capsid encapsulating the deoxyribonucleic acid (DNA), a proteinaceous tegument that surrounds the capsid and a lipid bilayer membrane with embedded glycoproteins[15, 16]. These glycosylated proteins interact with receptor proteins located on the surface of the host cell, which is crucial for the virus to enter the cell (step **a**)[16, 17]. This interaction between the viral glycoproteins and host cell receptors leads to the fusion of the cellular and viral membranes and the release of the capsid into the cytosol (step **b**). Most of the tegument is then shed and the capsid is actively transported to the nuclear pore complex (NPC) via microtubules (step **c**) [16, 18]. Subsequently, the capsid releases the viral DNA into the nucleus through the NPC (step **d**). The viral genome is replicated via a rolling-circle mechanism into a concatemeric dsDNA (step **e**) [16, 19]. At the same time, transcription of structural and nonstructural viral proteins takes place. In nucleus the newly formed procapsid undergoes assembly and packaging, a process involving several viral proteins such as pUL6, pUL15, pUL28, pUL32 and
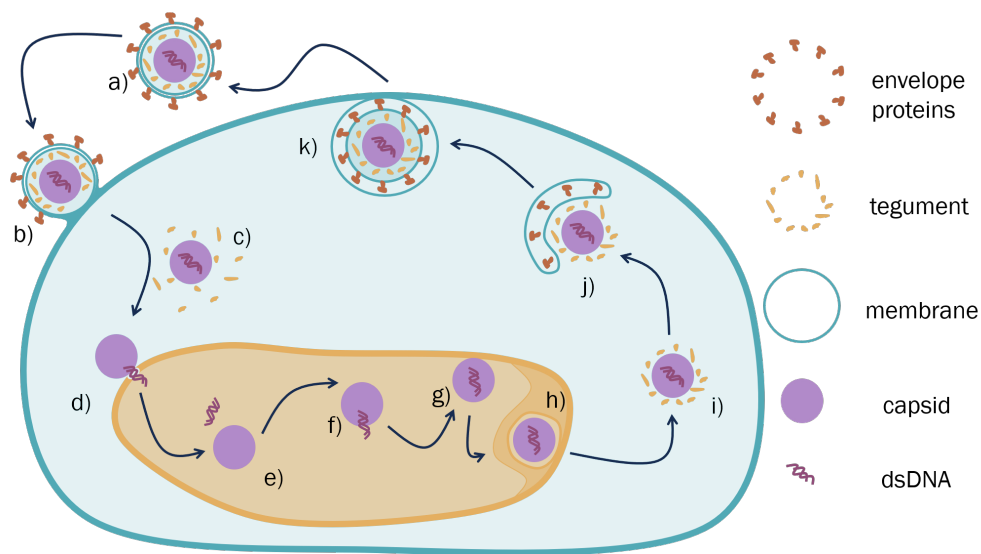
Figure 1.1: Schematic illustration of the replication cycle of herpesviruses: **a)** A fully infectious herpesvirus particle. **b)** Binding of the glycoproteins located on the viral membrane with the host cell receptors leads to the fusion of the cellular and viral membranes. **c)** Shedding of the tegument in the cytosol. **d)** The capsid docks on to the nuclear pore and ejects the viral genome into the nucleus. **e)** Replication of the viral genome and transcription of viral genes. Formation of structural and functional protein complexes, including the procapsid. **f)** Terminase complex packages newly formed dsDNA into the procapsid. **g)** The capsid associates with the nuclear egress complex. **h)** The genome containing capsid exits the nucleus. **i)** Release of the procapsid into the cytosol, where the tegument forms around it. **j)** Formation of the virion by membrane envelopment of the procapsid with tegument. **k)** Release of a mature viral particle by fusion of the vesicle to the membrane.

pUL33 (HSV-1) (step **f**) [15, 20]. After packaging, the capsid leaves the nucleus via an envelopment-deenvelopment step mediated by the nuclear egress complex (step **g**) [21] and is then released back into the cytosol (step **h**) [16, 22], where virion assembly occurs (step **i**) [16, 23]. The tegument is acquired, the capsid is enveloped and finally, the virion exits the cell (steps **j** and **k**) [16, 24].

1.1.2. The terminase complex

The terminase is a protein complex that plays a key role in the packaging and cleavage of viral dsDNA. It is highly conserved among herpesviruses and even shares similarities with those found in bacteriophages[25]. The complex binds to specific sites on the concatemeric viral DNA and translocates it into the procapsid until an entire genome has been packaged. Once this is achieved, the terminase cleaves the DNA to complete the process[26–30].

Conserved *pac* sequences direct DNA cleavage

In all herpesviruses, the terminase protein complex binds the concatemeric DNA at specific regions, the so-called a-sites, located on the individual viral genome termini. An a-site contains two packaging elements: *pac* 1 and *pac* 2[31, 32]. These packaging elements control the cleavage of the concatemeric DNA during procapsid packaging[33]. Although the process is not yet fully understood, it is known that each procapsid ultimately contains only one monomeric subunit of the concatemeric DNA after cleavage. It was found that the terminase cuts at a precise distance from *pac* 1 after packaging[33–35]. Both, *pac* 1 and *pac* 2, contain a T-stretch flanked by GC-rich sequences. In the case of *pac* 1 the flanking G-residues form a G-quadruplex (G4)[36], a DNA secondary structure composed of four guanines forming Hoogsteen-type hydrogen bonds. These can occur whenever four separate G-stretches are in close proximity to each other. The number of layers formed is equal to the number of

guanosines in these stretches. G4s are usually stabilized by cations like potassium, sitting between the four guanosines of each layer[37–39]. It was found that the G4 in *pac* 1 is conserved in all herpesviruses[36]. As another example, telomeres have a G4 at the 3′ end, which in cancer cells inhibits the digestion by telomerase. Due to the bulky structure of G4s, which can be stabilized by ligands and small molecules, the telomerase cannot bind to the 3′ end[40]. Because of this, targeted stabilization of the G4 is another interesting approach in viral drug design, especially considering that these structures are fairly common in all kinds of viruses as well as the human genome[41].

Even though, numerous questions regarding the terminase are still uncovered, an approved antiviral drug called Letermovir, which targets the terminase of another herpesvirus, HCMV, has shown great success. Based on mutations occurring in UL56, the UL15 homologue, the drug has been proposed to specifically target this protein. Inhibition of the terminase results in poorly packaged virions, hindering viral replication[12, 13]. While the exact inhibition mechanism has not yet been unraveled, the terminase seems to be a promising target for antiviral drugs[42, 43].
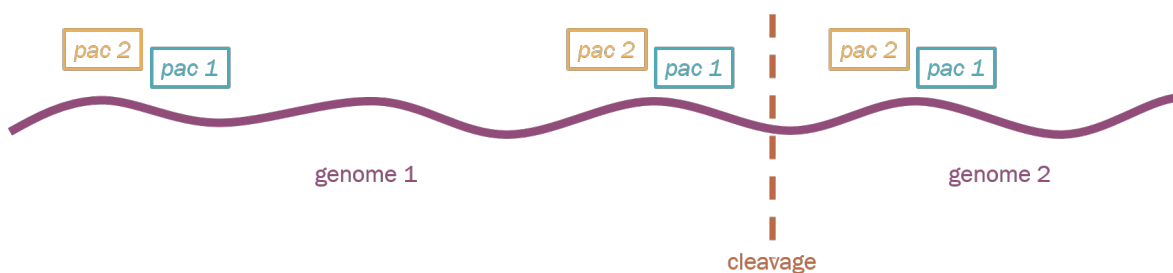
Figure 1.2: Simplified representation of the concatemeric HSV-1 genome. The motifs of *pac* 1 and 2 are marked at the genomes termini and a line symbolizes where the terminase cleaves after packaging.

The composition of the terminase protein complex

During the packaging of viral concatemeric DNA in HSV-1, a full genome is transported into the procapsid. Here, a multitude of different protein complexes including the portal pUL6, the capsid vertex-specific proteins pUL17, pUL25, the terminase (pUL15, pUL28, pUL33) and terminase-associated protein pUL32 are present[27, 44]. The subunit pUL15 functions as an ATPase cutting the DNA between the two *pac* sequences on the genomes' termini (see **Chapter** 1.1.2)[45, 46]. Subunit pUL28 contains a zinc-finger that binds single-stranded deoxyribonucleic acid (ssDNA)[31]. The role of pUL33 might be related to the initial cleavage of the DNA but is not yet fully understood[47]. These three proteins form the heteromeric trimer terminase complex (HTC). The terminase binds to the viral DNA and the portal protein. Then, driven by adenosine triphosphate (ATP) hydrolysis, it pumps the DNA into the procapsid and cuts it at the appropriate length[15]. Even though, the composition of the terminase was known for nearly 20 years[15, 30], its structure remained largely unclear. Only recently, it was shown in a study from Yang *et al.*[29] that the HSV-1 terminase forms a donut-like structure containing six HTCs which will be referred to as hexamer of the heteromeric trimer terminase complex (HTC_6) in the following.

Mechanism of the terminase packaging motor

As seen in **Figure** 1.4 panel **c**, every terminase HTC has an arginine finger. In HSV-1 the arginine finger is located at R346. After binding DNA, a HTC undergoes a conformational change, causing the arginine finger to rotate and bind to the active center close to the ATP binding site of a neighboring HTC. Here, the arginine finger catalyses ATP hydrolysis through an interaction with the γ-phosphate of ATP, stabilizing the transition state. Afterward, the finger returns to its initial position, triggering another conformational change in the ATPase active center and the basic patch. The motif of the basic patch consists of six residues (R291, K292, R306,
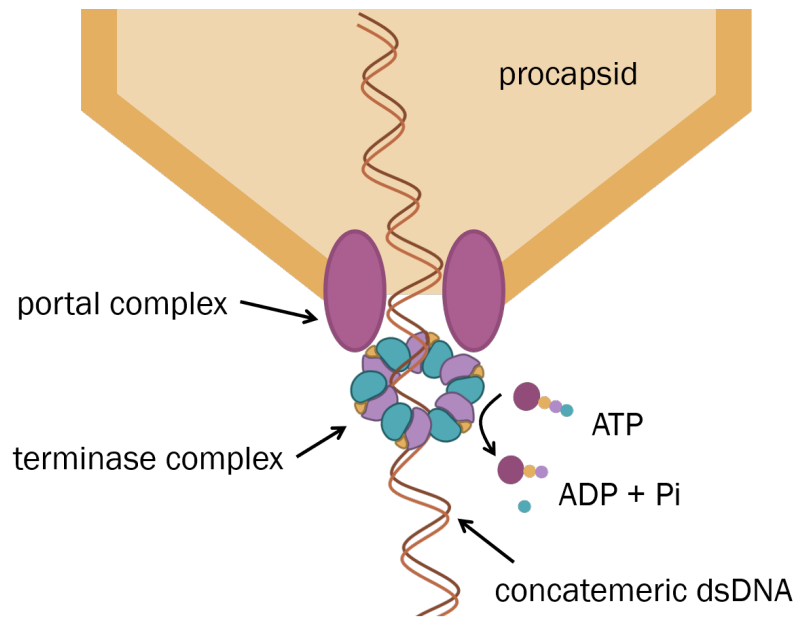
Figure 1.3: Illustration of the DNA packaging process into the empty procapsid. In the depicted moment, the terminase is docked to the portal complex of the procapsid. By ATP hydrolysis, it displaces the dsDNA into the empty procapsid through the portal complex.

R313, K318, R331) and is highly conserved across bacteriophage and herpesvirus terminases[29]. It is located in the inner channel of the donut-like structure and likely binds DNA. The conformational change results in a 15° turn of the basic patch around the strut and when bound to DNA, causes the DNA being pushed upwards in a spiral[29]. This proposed mechanism would fit a revolution model[48]. Before the structure of the terminase (HSV-1) was resolved, most of our understanding of the mechanism and composition of the terminases came from studies that focused on terminases in bacteriophages[49]. Here, a rotation model was initially proposed that explained the packaging mechanism[50]. Further investigation strongly suggested that this model had a major weakness: When the terminase rotates around the DNA it would lead to super coiling, friction and torque. As this would be too energy-costly to resolve after packaging, a new model had to be found. The revolution model, as seen in **Figure** 1.4, does not involve the rotation of the subunits. Here, the DNA binds unidirectionally inside the positively charged center of the donut-shaped

Figure 1.4: Structure and mechanism of the terminase packaging complex. **a)** Structure of the hexameric terminase during translocation of the dsDNA. Subunits pUL28 and pUL33 as well as different important motifs located on pUL15 are marked in different colors. **b)** The same terminase composition as in a) but during cleavage. **c)** Detailed depiction of important motifs inside the terminase complex. **d)** Two of the subunits of the hexameric assembly of the terminase bind DNA. This triggers a conformational change in the first subunit and leads to ATP hydrolysis in the second. This conversely leads to the conformational change of the second subunit and pushes the DNA upwards in a spiral. A chain-reaction occurs and the DNA is displaced into the procapsid. Reprinted and adapted from Yang *et al.*, Protein cell, 2020[29] under CC BY 4.0 license.

terminase. Through ATP-induced conformational changes, the affinity of the protein-DNA complex changes, resulting in a constant formation and breaking of bonds. In the latter case, the DNA is being "flipped" in an upwards motion needed for packaging[51, 52].

1.2. Characterizing protein complexes using mass spectrometry

A powerful tool for the characterization of proteins is mass spectrometry (MS)[53]. It became a staple for proteomics and can be utilized for various applications ranging from analyzing the exact sequence of a protein and its post-translational modifications (PTMs)[54], over mapping binding partner to exact locations on the protein[55], to unraveling composition of protein complexes[56]. In a MS experiment, the detection and quantification is based on the yields and m/z of the ion species which are stable in the gas-phase. For this, the ions are transported into the gas phase and analyzed via time-of-flight mass spectrometer (TOF-MS) (**Figure** 1.10) or Orbitrap (**Chapter** 1.2.5) mass spectrometry[57, 58]. While sample preparation and delivery can vary, the basic components of an MS experiment are always the same. An overview of generic MS experiment is shown in **Figure** 1.5.

The sample is introduced from solution into the gas phase in the form of ion species. After entering the experimental chamber, the ions are separated by their m/z. Ions with different m/z are then analyzed by the mass analyzer.[57, 58].
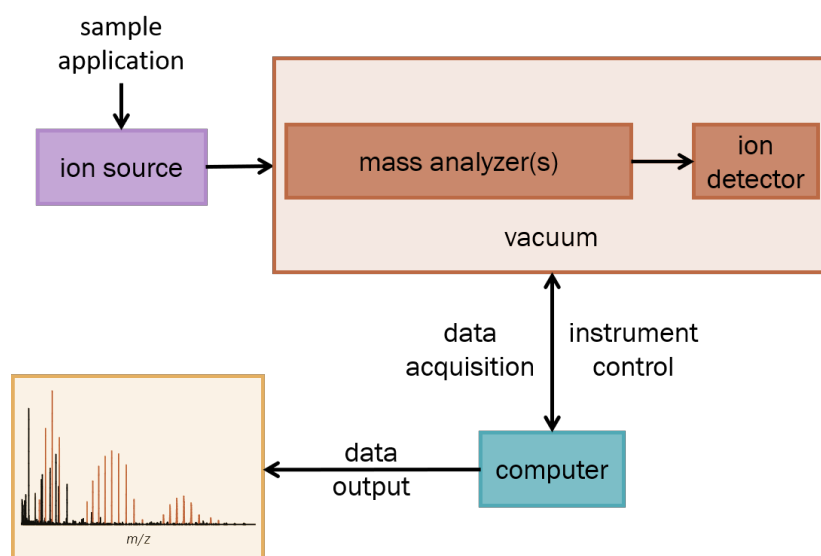
Figure 1.5: Simplified setup of a mass spectrometer and all of its major components, including the ion source, operated at atmospheric pressure as well as at least one mass analyzer and the detector, both of which operate under vacuum. In addition, a computer is required for instrument operation and data collection. The sample is applied to the ion source, where it is ionized and then separated by its m/z in the mass filter. The ions are then analyzed in the mass analyzer and detected.

1.2.1. Mass spectrometry-based proteomics

The field of mass spectrometry expanded from small chemical analytes to big biochemical compounds such as proteins[58–61]. The analysis of proteins regarding their function and structure in an organism is called proteomics. In this field, two major workflows have been established. Bottom-up MS is a method in which the proteins are digested by proteases into peptides before injection into the mass spectrometer. In top-down MS, whole proteins are being injected into the system. Both methods typically use denaturing solvents such as acetonitrile due to prior reverse-phase liquid-chromatography protein/peptide separation. When utilizing solvents with high organic percentages, proteins tend to unfold, leading to a loss of structural information.

In native mass spectrometry (nMS) the term, "native" refers to the biological protein complex structure in solution before it is transferred in to the gas-phase[62]. In a

physiological environment, proteins form characteristic tertiary structures based on their amino acid sequence. From these proteins, quaternary structures can be formed by binding to other proteins via non-covalent bonds. The function of these homogeneous or heterogeneous complexes are highly dependent on their stoichiometry and assembly. Therefore, nMS utilizes non-denaturing volatile buffer surrogates at physiological pH and ionic strength conditions to retain the structural information of these proteins or protein complexes in solution. The native structure of the proteins in the gas-phase is often discussed among scientists, but it has been shown previously that native-like structures can be retained[62, 63]. It is therefore possible to analyze both conformation and composition, as well as the stoichiometry of proteins and protein complexes via nMS. It also allows for the observation of structural dynamics. It is often necessary to conduct a two-step analysis in nMS for the in-depth analysis of a protein complex. Here, usually a two-stage tandem mass spectrometry, also referred to as two-stage mass spectrometry (MS^2), is utilized, which results in the gradual dissociation or even fragmentation of a protein or protein-complex for example via collision-induced dissociation (CID)[64]. A tandem mass spectrometer consists of multiple mass analyzers, which is why it is referred to as MS^n, n referring to the number of mass analyzer. In between, mass analyzer steps often CID is performed in a collision cell (see chapter 1.2.4).

1.2.2. Native mass spectrometry with nano-electrospray ionization

Conventional electrospray ionization (ESI) involves the application of the sample via a continuous flow, which is regulated by a pump. In contrast, nano-electrospray ionization (nESI) utilizes a capillary to contain the sample, enabling the attainment of low flow rates without the necessity of an external pump[65]. Because of the low flow rate and the small diameter of the capillary, a reduced sample volume is necessary for

nMS analysis. As a very soft, ionization technique keeps the non-covalent interactions intact and preserves the three-dimensional structure of the protein[61, 66, 67]. It has been shown that whole viruses can be brought into the gas phase via nESI and still be infectious[68].
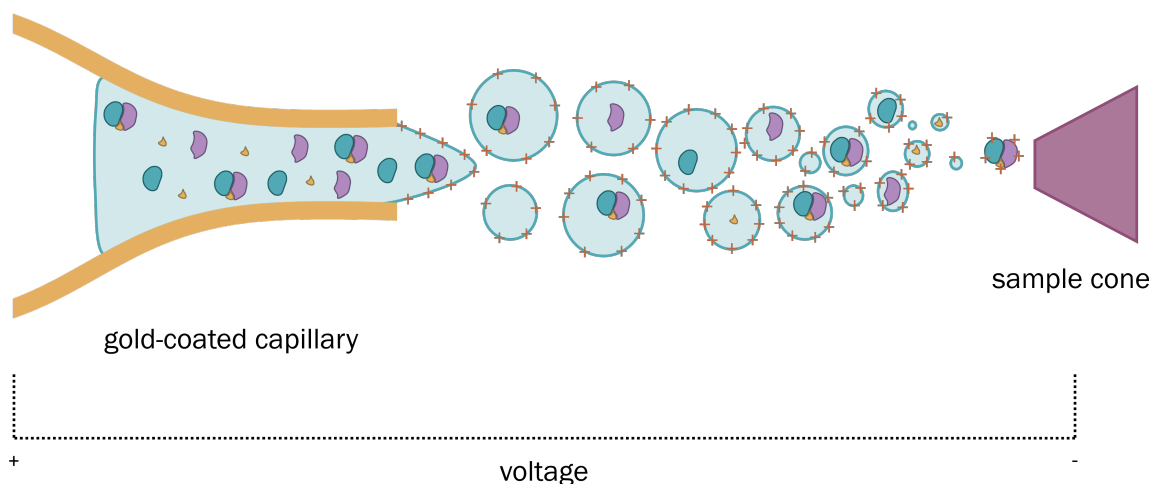
Figure 1.6: Simplified illustration of the ionization of large bio molecules with nESI. The sample is introduced via a gold-coated capillary that is placed in an electric field, resulting in the formation of a *Taylor cone*. Positively charged droplets are emitted. The solvent evaporates, leading to bare gas-phase ions entering the mass spectrometer via the sample cone.

During nESI, the protein sample is filled into a gold-coated capillary. A high voltage is applied in between the capillary and a sample cone at the entrance of the mass spectrometer. In the electric field, a positively charged sample droplet accumulates at the tip of the capillary, forming a Taylor cone[61, 66, 67, 69]. Solvent droplets, are emitted towards the sample cone. The solvent evaporates and causes the droplets to shrink. Here, the surface tension and Coulomb repulsion act as counteracting forces. When a droplet's size falls below the Rayleigh limit, the repulsion outgrows the surface tension, leading to fission of the droplet. This process is referred to as Coulomb fission[57, 70].

The ionization mechanisms during ESI are not fully resolved yet. There are

different models described in the literature. The charged residue model is often referred to for large biomolecules. Here, each molecule is in a separate droplet in the gas phase. During evaporation the charges from the solvent are transferred to the molecule[71]. The ion evaporation model usually describes the process for small analytes, here the ion is ejected out of the droplet over a solvent bridge[69].

For efficient ionization an appropriate solvent has to be used. In nMS, ammonium-based buffer surrogates are often chosen. Most commonly, ammonium acetate, a weak, volatile salt, is utilized. The volatile nature of the salt assures adduct-free proteins after ionization[72]. It is crucial to remove other non-volatile cations like sodium before sample injection, as these result in highly complex spectra that often prove difficult to analyze.

1.2.3. Native protein spectra analysis

A protonated gas-phase protein ion generated by ESI or nESI can exhibit multiple charge states. These states typically differ by a single charge unit and the absence of any particular charge state is uncommon. The distribution of charge states usually follows a Gaussian pattern[73, 74]. The number of protons a protein can carry depends on several factors, including the protein's isoelectric point (pI), the pH of the buffer or solvent and the protein's conformation. Unfolded proteins expose a larger surface area to protons and therefore tend to carry more charges. In contrast, the compact structure of folded proteins limits the number of accessible protonation sites, resulting in lower charge states[75]. Additionally, the charge state can be chemically modulated using supercharging reagents (e.g., sulfolane) to increase protonation, or charge-reducing agents (e.g., imidazole) to lower it[76, 77].

Deconvolution—a computational process used to extract the neutral molecular mass from a complex spectrum—is possible when the mass-to-charge ratio (m/z)

and charge state (z) of at least two peaks from the charge state distribution can be identified. The accuracy of mass determination improves with the number of resolved charge states[78]. However, the experimentally determined mass of a protein often differs from its theoretical mass. This discrepancy is commonly due to PTMs, chemical modifications of specific amino acid residues that occur after protein synthesis[79, 80]. These modifications, which do not change the amino acid sequence itself, include common examples such as phosphorylation and glycosylation[81]. Another contributor to mass increase is adduct formation, often caused by non-volatile salts that persist during nESI[82]. Incomplete desolvation is particularly problematic for large protein complexes, such as virus-like particles with masses in the megadalton range, as residual solvent molecules can remain with the ions. This variability leads to peak broadening and complicates the interpretation of mass spectra[83, 84]. The quality of a mass spectrum is commonly assessed using the full width at half maximum (FWHM), which reflects the width of a peak at half its maximum intensity. A low FWHM indicates high spectral resolution and improved mass accuracy. High-resolution spectra are better at revealing fine structural features, such as those introduced by PTMs, which may be visible as subtle variations within the charge state distribution[85]. In contrast, broad peaks resulting from low resolution or overlapping signals make it difficult to distinguish between closely related species. Therefore, a high FWHM often indicates sample heterogeneity or the presence of multiple binding partners. Due to factors such as incomplete desolvation, adducts and PTMs, the practical (experimental) resolution of a mass spectrum is often much lower than the theoretical resolution of the instrument. A high-quality, charge-state-resolved spectrum is thus defined by well-separated peaks, where adjacent charge states are clearly distinguishable at no less than half of their maximum intensity[86].

The molecular mass of the analyte is usually calculated from the m/z values of

the peaks in the charge state distribution. Even in pure samples, multiple charge state series may appear due to the formation of homogeneous protein complexes. In such cases, tandem mass spectrometry (MS^2) is often used to determine the exact stoichiometry of the complexes[56].
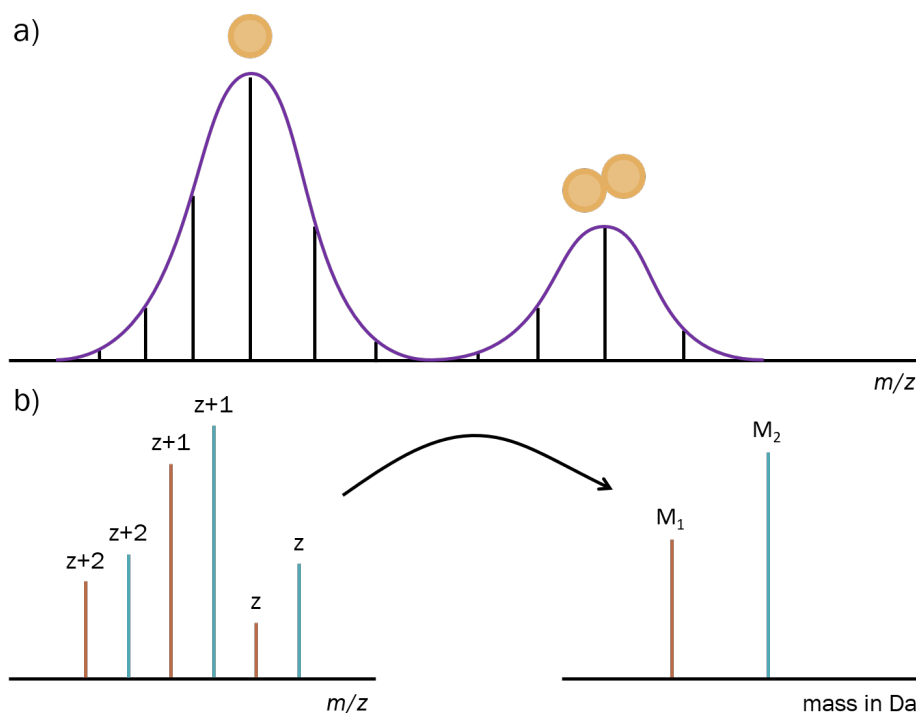
Figure 1.7: Simplified native mass spectra of proteins. **a)** Gaussian charge state distribution of different oligomeric states of the same protein in one spectrum. **b)** Deconvolution of two protein species with different masses with overlapping charge state distributions.

1.2.4. Gas-phase dissociation of protein complexes

For MS^2 the ion is fragmented, which usually occurs in a collision cell. In a collision cell, precursor ions collide with inert gas, often nitrogen, argon, or xenon[57, 87, 88]. During collision, kinetic energy of the inert gas is transferred to the precursor ion and converted into vibrational energy. The excess of inert gas leads to collision of the precursor ion not just with one but multiple of these inert gas atoms/molecules. During every collision, kinetic energy is transferred to the sample molecule and

increases its internal energy. At low internal energies the protein complex unfolds. This can be used to get additional structural information by gradually unfolding the protein [89]. With increasing collision events and therefore higher internal energy, weak non-covalently bound subunits of protein complexes dissociate. The charge is hereby often distributed asymmetrically between the dissociation products. At even higher energies' fragmentation of the precursor ion sets in[90–93].
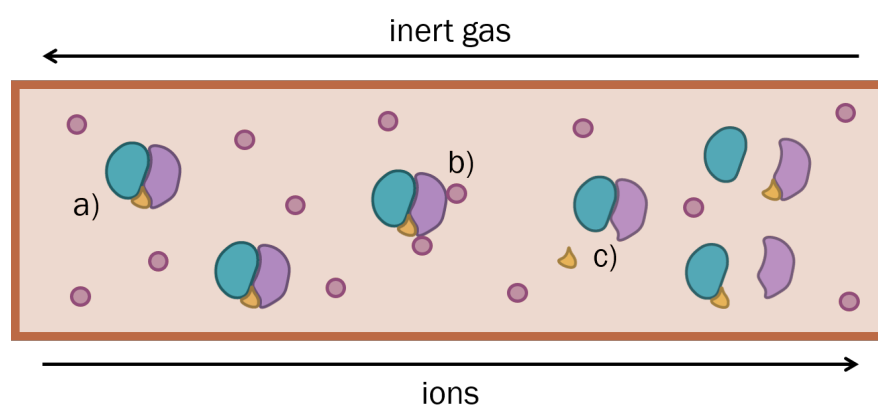
Figure 1.8: General overview of the processes occurring within a collision cell. **a)** Ions are let into the collision chamber from the left whereas inert gas molecules (e.g., Ar, Xe) are entering from the opposite side. **b)** The gas molecules collide with the protein complexes, which leads to an energy transfer. **c)** When the inner energy of the ions gets too high, the protein complex dissociates, releasing a subunit.

1.2.5. Mass analyzer

Quadrupole mass analyzer

A quadrupole mass analyzer (QMA) consists of four metal rods arranged in parallel in a square cross-section[57]. Opposing rods are electrically coupled. Electric radio frequency (RF) potentials with equal amplitudes, but opposite polarity, are applied to both pairs of rods. Additionally, a direct current (DC) voltage is applied between the rod pairs.[87]. Together, these voltages enable the stable movement of ions within a specific m/z range through the QMA. Other ions collide with the metal rods[57, 87].

Ions that have traversed the QMA successfully proceed to the subsequent section of the mass spectrometer via the exit slit[94, 95].
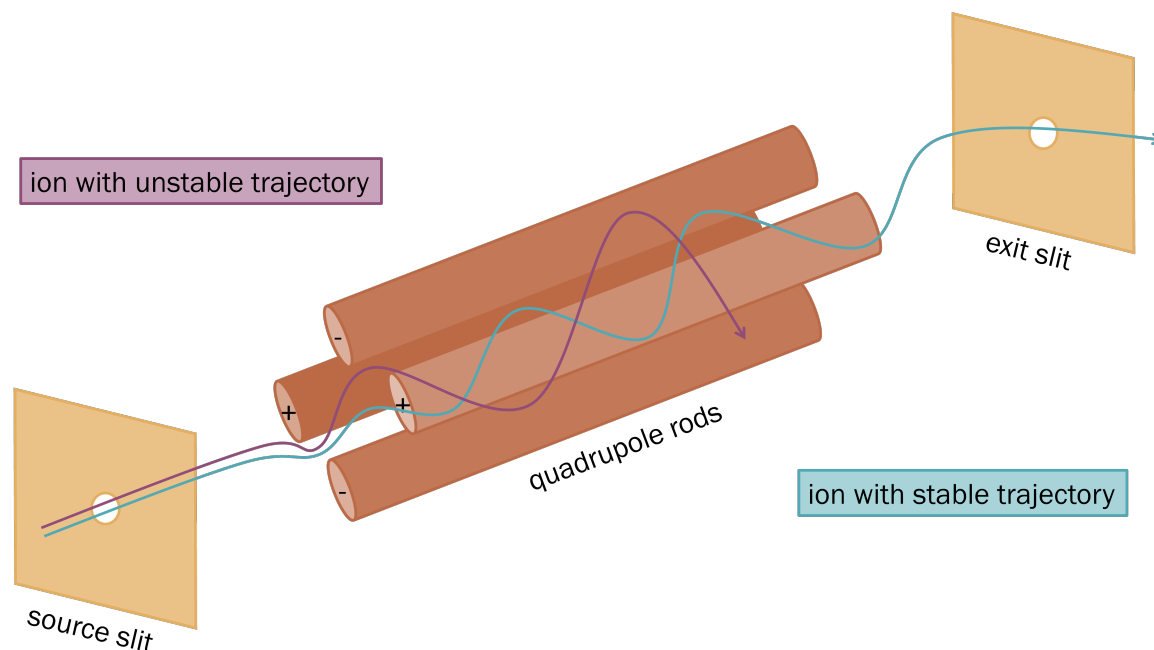
Figure 1.9: Illustration of a QMA in positive mode. Four rods are placed in a square with altering voltages leading to incoming ions to oscillate. Adjustment of applied DC and RF voltages allow for a selection of ions within a chosen m/z window. These are transferred through the QMA in a stable motion.

Time-of-flight mass spectrometer

The ions enter the TOF-MS at a 90° angle and are then accelerated by a pusher electrode into a field-free drift tube which has a defined length[57]. The ions pass the tube and hit the multi-channel plate detector directly behind it. Ions with higher m/z are slower and hit the detector later than ions with a lower m/z[57, 96]. Often a reflectron, or ion mirror, is utilized to increase the resolution of the TOF-MS. Incoming ions are deflected back in to the field-free room, increasing the length. Depending on the ion's kinetic energy, they penetrate the reflectron area at different depths, focusing the ion bundles and increasing the spectral resolution[87, 96].
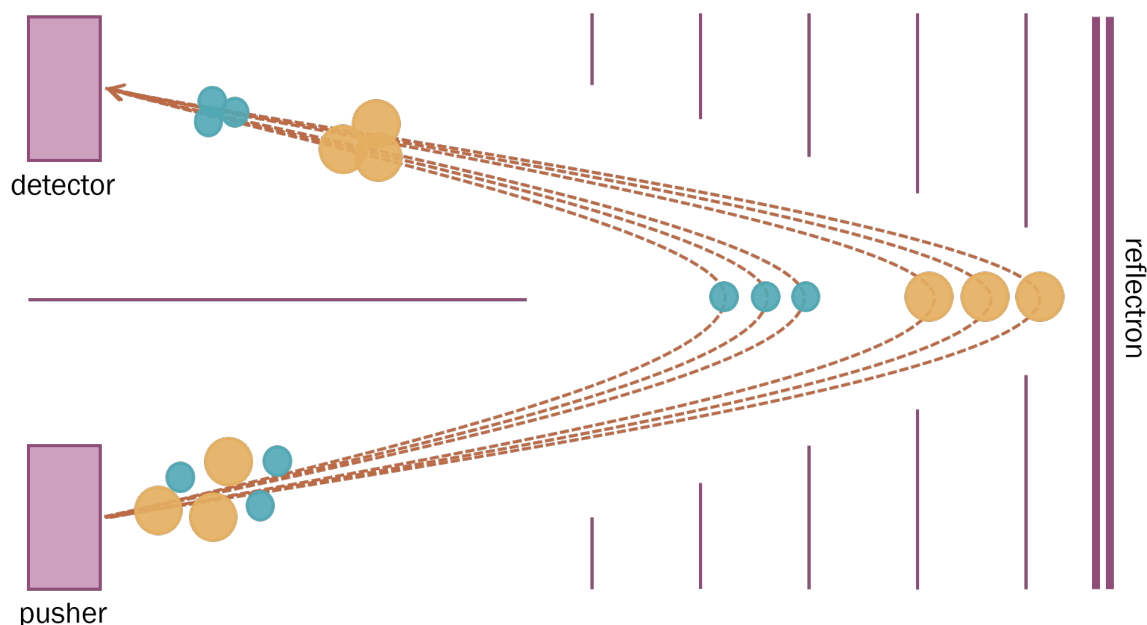
Figure 1.10: Illustration of a reflectron time-of-flight mass spectrometer. Ions are accelerated by the pusher electrode into a field-free space. The higher the initial velocity, the deeper they penetrate the electromagnetic field of the reflectron. Ions with the same m/z but different kinetic energies are therefore focused in to ion bundles.

Q-TOF 2 mass spectrometer

The Q-TOF 2 consists of a QMA and an orthogonal TOF-MS. Every QMA has two modes. By altering the voltages, it can either function as an ion guide or a mass filter. In the first mode all ions are forwarded to the TOF-MS for single-stage mass spectrometry (MS^1) data acquisition. In the second mode, the ions are filtered by their m/z ratio and guided into the collision cell for CID and subsequent MS^2 analysis. This robust system is suitable for large protein complex analysis, as shown in previous studies conducted on viral capsids[97, 98].

The Orbitrap™ mass analyzer of a Q Exactive Orbitrap™ mass spectrometer

Before the ions enter the Orbitrap™, a spindle-shaped mass analyzer[99–102], they are accumulated in the C-trap. Here they are deaccelerated and experience a small axial potential. The accumulated ion bundles then are sent into the mass analyzer

at an 90° angle[103]. The ion packages oscillate axially around the inner electrode.

Even though the equations of motion of all ions have the same amplitude, their

frequencies differ depending on their m/z. The outer electrode is divided into two

parts and detects these oscillation frequencies and a computer unit converts them via

Fourier-Transformation into their respective frequencies, from which their m/z can

be obtained[99–101]. In some setups, a high-energy collision dissociation (HCD) cell

is connected to the C-trap. Before ion bundles enter the mass analyzer, they are let

through the C-trap to undergo CID in the HCD cell. The fragments or dissociation

products then re-enter the C-trap to then get sent to the Orbitrap™[100].
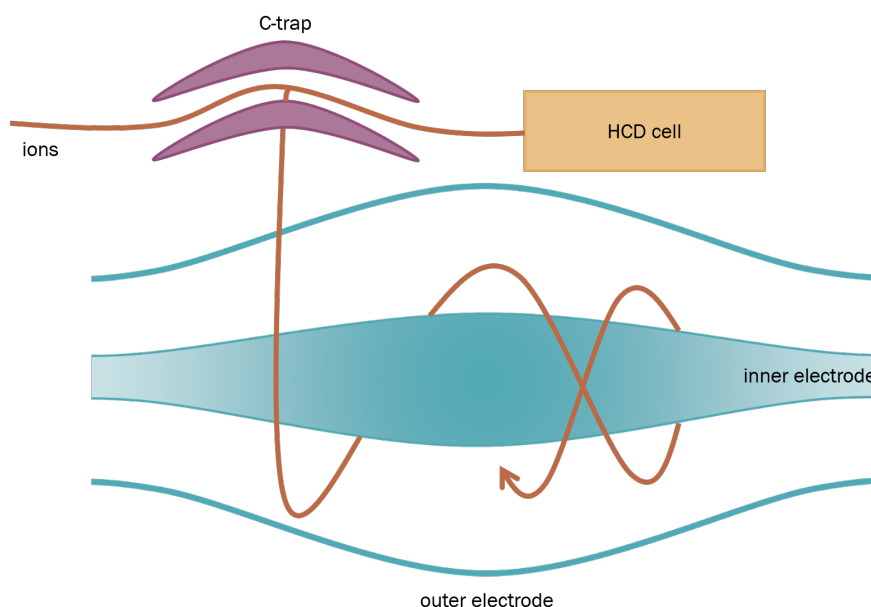
Figure 1.11: Illustration of an Orbitrap™ mass analyzer including C-trap and HCD cell. The C-trap accumulates ions and releases them either into the HCD-cell, for MS^2 experiments, or in an 90° angle into the Orbitrap™. Here, they orbit a spindle-shaped inner electrode. The frequency with which the ion bundles axially oscillate is dependent on the ion's m/z.

Q Exactive™ UHMR Hybrid Quadrupole Orbitrap™ mass spectrometer

The Q Exactive™ UHMR Hybrid Quadrupole Orbitrap™ mass spectrometer (QE-

UHMR) combines a QMA with an Orbitrap mass analyzer. Here, the QMA has

the same operating principle as in the Q-TOF 2. It has a higher mass accuracy and

resolution than the Q-TOF 2. As its name suggests, the QE-UHMR is particularly suitable for analyzing large molecules like protein complexes[104, 105].

Unlike the Q-TOF 2, the QE-UHMR is equipped with in-source CID in addition to the HCD cell[106]. This allows for the removal of small molecule clusters and low-binding adducts, as well as a better desolvation of large proteins. Moreover, the sample cone is often operated at higher temperatures than the cone of the Q-TOF 2[106, 107]. This promotes desolvation and transfer of large molecules into the mass spectrometer. Because of this, the ion source of the QE-UHMR is often referred to as being "hot", whereas, the Q-TOF 2 ion source is labeled "cold".

1.3. Protein sample preparation for native mass spectrometry

During this work, several proteins from a multitude of origins have been expressed. Bacterial (*Escherichia coli* (*E. coli*)), insect (*Spodoptera frugiperda*) and mammalian (human embryonic kidney (HEK)) cells were utilized in the production of the proteins. Every expression system has its advantages and disadvantages. *E. coli* expression typically results in highly over-expressed proteins and is a relatively rapid and a straightforward method for protein production[108]. On the other hand, proteins from human viruses are generally more closely aligned with their native counterparts when expressed in insect cells than *E. coli*[109–111]. They are also exhibiting a closer resemblance in their folding patterns. Mammalian cells produce the proteins closest to natural viral proteins, but yields are often poor. Given these distinctions, the selection of an expression system must be made with careful consideration, tailored to the specific properties of the protein in question[112].

After expressing the proteins, they have to be purified. A typical approach for protein purification involves employing several chromatographic steps[113]. Affinity chromatography is typically the first chromatography step, followed by size-exclusion chromatography (SEC). This 2D approach increases the purity of the protein sample[114, 115]. The commonly used buffers during protein purification contain non-volatile salts which are incompatible with nMS. Protein samples therefore have to be exchanged to ammonium acetate-based buffer surrogates for nMS analysis.
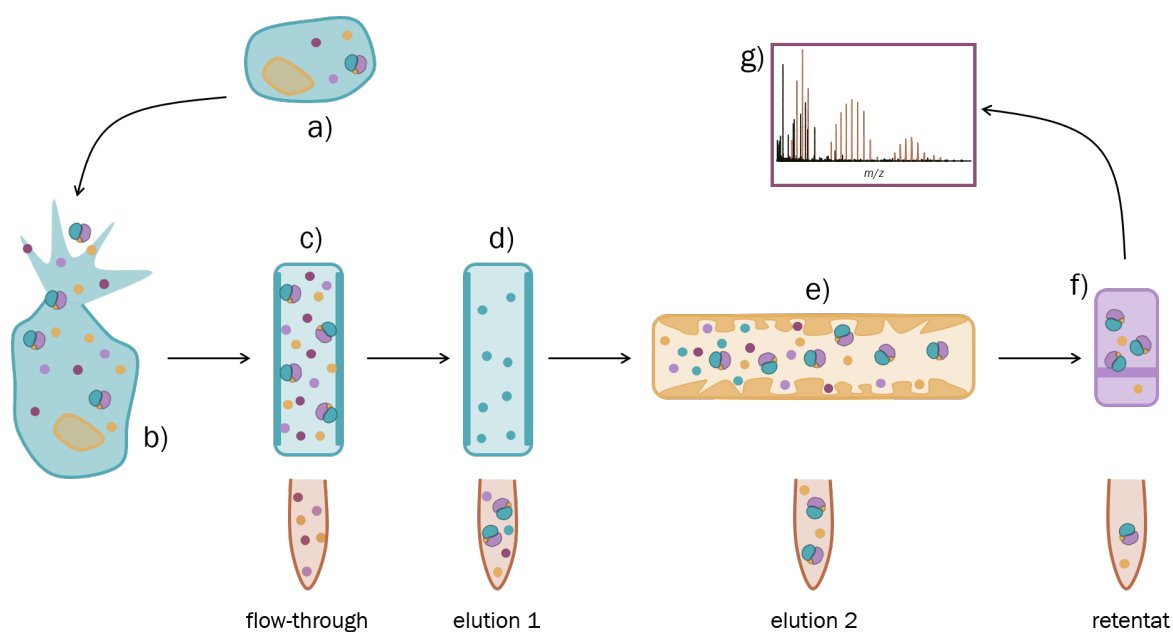
Figure 1.12: Overview of a standard protein sample preparation protocol for nMS. **a)** Protein expression in a suitable cell expression system. **b)** Cell lysis to release expressed proteins. **c)** Affinity chromatography against a specific tag fused to the target protein. **d)** Elution of retained proteins from the matrix of an affinity chromatography column with a matrix-specific eluent. **e)** SEC purification as a secondary purification step to reduce the number of contaminants. **f** Buffer exchange to replace non-volatile salts from the nMS sample. **g)** Acquisition of the nMS spectrum.

1.3.1. Advantages and disadvantages of different expression systems

The most established expression system used in research is *E. coli*. Multiple cell lines have been engineered to express proteins in high quantities in a short time period. The most advantageous aspect of *E. coli* as an expression system is its ease of maintenance and low cost. *E. coli* expression is also a fast process; the transformation of a plasmid containing the desired gene to harvesting the protein of interest is often achieved in significant less time, compared to other expression systems[116, 117]. Additionally, *E. coli* expression cultures are often conducted in high volumes, resulting in exceptional protein yields. On the contrary, proteins can be not folded correctly. This leads to the formation of inclusion bodies, insoluble protein aggregates that are difficult to purify[118, 119].

Contrary to *E. coli*, insect cells usually have to be kept in desired cell densities for multiple days before expression. The cells are temperature and pH sensitive and the media degrades when exposed to light. The eukaryotic insect cells are much less stress resistant than the prokaryotic *E. coli* bacteria. Physical stress like sheering or oxidative stress from reactive oxygen species, which are often degradation products of light sensitive chemical components in the media (e.g., riboflavin or amino acids), often lead to cell death in insect cells[110, 120–122].

Mammalian cells can also be maintained in suspension culture, but most often they are cultivated in dishes. Adherent cells are the gold standard for life cell imagine, but the low cell density in each dish is often not sufficient for the expression of high-protein yields[123]. Often it can be challenging to adapt adherent cells to a suspension culture. Similar to the insect cells, the protein expression is usually performed in much smaller volumes and expression yields are lower than expression yields from *E. coli*. The major advantage of mammalian cells is that human virus

proteins expressed in them are usually much closer to the biological proteins than when they were expressed in insect cells or *E. coli*[124].

By taking all advantages and disadvantages into context, often insect cells are chosen for expression of different biologically interesting proteins because they have a higher protein yield than mammalian cells and their PTMs and folding are closer to the natural protein species than *E. coli* expressed proteins. Mammalian cells, on the other hand, are often more suitable when expressing highly glycosylated protein species[112].

1.3.2. Protein purification via affinity chromatography

Affinity chromatography is a common method for purifying recombinant expressed proteins. It is based on the interaction of a matrix and a protein tag. Two of the most commonly employed tags for this purpose are 6xHis-tags[125] and StrepII-tags[126]. His-tags bind to Ni^{2+}, among other metal ions and are typically eluted with imidazole at concentrations between 100 and 500 mM[127]. This method is cost-effective and commonly employed. The interaction between Ni^{2+} and histidine residues, which can occur naturally in proteins, may result in non-specific binding[128]. Because of this, His-tag affinity chromatography is often followed up with a secondary chromatographic technique. Often this technique is SEC, which increases the purity of the sample and tolerates the solvent parameters used in affinity chromatography. Here, the molecules are traversing the column at different retention times based on the size of the molecule. The larger the molecule, the less it interacts with the surface of the column and the faster it elutes[129].

In contrast, StrepII-tags are capable of mimicking the high-affinity interaction between streptavidin and biotin, which is one of the strongest binding interactions known in nature to date[130]. As biotinylation is an uncommon protein modification, non-

specific binding is significantly reduced, resulting in highly pure protein samples. Elution buffers for StrepII-tagged proteins typically include 50 mM biotin or 2.5 mM desthiobiotin, with the latter exhibiting a slightly lower affinity that facilitates column regeneration[131].

1.3.3. Disadvantages of buffer exchange

The purification of proteins frequently entails the utilization of Tris- or HEPES-based buffers that serve to stabilize proteins, yet simultaneously introduce complications regarding nMS detection. During protein ionization via nESI, the formation of adducts between non-volatile salts such as sodium and the protein ions can greatly increase the complexity of the resulting spectrum or completely mask peaks. Therefore, buffer exchange represents a crucial stage in the preparation of samples for nMS-based structural proteomics. The exchange of buffers is typically conducted using spin filters, spin columns, SEC, or dialysis cartridges[132–138]. However, proteins frequently irreversibly bind to the various matrices or precipitate, resulting in considerable protein loss. Furthermore, buffer exchange is a time-consuming process, which can be problematic when working with labile proteins or protein complexes[132]. Previous studies have made significant efforts to reduce negative effects from non-volatile salts on the mass spectrum by finding new ways to perform buffer exchange or adding additives to disrupt sodium adduct formation[132, 139–141], thereby demonstrating the benefits of avoiding buffer exchange altogether.

1.4. Assessing the dimerization of fluorescent proteins by native mass spectrometry

Studying protein complex stoichiometries is a common practice in nMS. Numerous studies have been conducted on diverse protein complexes regarding their composition, as the stoichiometry of a protein complex is crucial for its functionality. Proteins are often used in different techniques to accomplish certain actions. One example would be to use protein tags for purification. In light microscopy, a well established technique in virology, fluorescent tagged proteins are tracked to study interactions between proteins[142]. Since it is impossible to distinguish a singular protein in a cell by eye, the target protein and a fluorescent protein (FP) are fused[143]. A FP is a protein that absorbs and emits light at different wavelengths[144]. By focusing a light with the specific excitation wavelength onto the cell, these FPs are then excited and can be tracked. Fusing different proteins to different FPs and exciting with multiple lasers at the same time can reveal interactions and pathways of these proteins[21, 145, 146]. There are countless studies on single cell imaging using various FPs like enhanced green fluorescent protein (EGFP) or monomeric enhanced green fluorescent protein (mEGFP)[147]. Choosing a suitable FP is vital for imaging and parameters such as brightness and monomericity should be considered. While the monomeric nature of these proteins is often implied, the accuracy is critical since dimerization could lead to misleading results[148]. In 2012, a publication established a method to determine the oligomerization of FPs called organized smooth endoplasmatic reticulum (OSER)[149]. During the experiment, the FP is fused to CytERM, an endoplasmic reticulum signal anchor membrane protein and expressed in cells. The FP monomers would then interact with each other forming dimers, changing the endoplasmic reticulum's structure from a structured tubular network into smooth whorl structures. This

way, oligomerization can be observed in a cellular environment. As of today, the OSER of a lot of new FPs have not yet been analyzed or have only been discussed in one publication that based their findings on one cell type. The process is fairly time-consuming and not an efficient approach to analyze multiple FPs tendency to form dimers at the same time. To systematically screen potential FP candidates, a fast and efficient *in vitro* test can be performed by using nMS.

2

Aim and objectives

The objective of this thesis was to structurally characterize the HSV-1 terminase with nMS and the *pac* sequences with circular dichroism (CD) spectroscopy. It was the goal to subsequently analyze the importance of the conserved *pac* 1 G4 on the binding of the DNA to the terminase complex.

Even though the terminase proved a valuable anti-viral drug target, the mechanism of how it binds and packages DNA is not fully understood to this day. Multiple groups tried to unravel the mechanisms of this highly complex system and made substantial progress, but progress has been slowed by difficulties in expressing the protein subunits *in vitro*[15]. The difficulties regarding purification of the terminase resulted in a major part of the thesis being the development of a method to successfully express and purify the intact proteins by minimizing the overall protein loss and time spent. The fast track protocol for nMS protein purification was developed and firstly tested on the standard protein alcohol dehydrogenase (ADH). The terminase and the terminase-associated protein pUL32 (along with its KSHV homologue pORF68) were also purified accordingly. From this, the stoichiometries of the terminase complex were determined. Then, ssDNA oligomers of the *pac* sequences were added to the terminase and protein-DNA interactions were studied with nMS. This included

two mutated oligomers of *pac* 1, one where the G4 formation and one where the T-stretch was disrupted. During this thesis, the structure of the HSV-1 terminase was published by another group[29]. Therefore, comparing the findings with nMS became an additional objective for this project.

The fast track protocol was also applied to investigate the dimerization behavior of FPs, which are widely used in light microscopy. While the relevance of such characterization has long been recognized, previous methods were often time-consuming and complex. One established approach aims to quantify dimer formation in *in vivo* conditions to closely reflect biological reality[149]. However, due to the effort involved, this type of analysis has not yet been carried out for all FPs. The fast track protocol coupled with nMS offers a much quicker alternative, allowing researchers to assess dimerization properties efficiently and apply the method to a wide range of proteins in succession.

3

Results and discussion

3.1. The expression of the terminase in insect cells is challenging

The structural and compositional properties of the terminase complex are of paramount importance to its function as a DNase and for its packaging mechanism. At first, recombinant expression of the proteins was necessary to enable its analysis via nMS. Analogous to previously published work, both insect cells[29, 47] and *E. coli*[28] were used. Multiple expression parameters for both expression systems were screened to maximize protein yield and quality. The results of the study are summarized in **Table** 3.1. The first conclusion drawn from the data is that expression in *E. coli* was unsuccessful regardless of the induction method used (Isopropyl-β-ᴅ-thiogalactopyranosid (IPTG) or auto-inducing medium) and therefore not continued. The second conclusion is that insect cells express the terminase proteins in higher yields in smaller suspension culture volumes. In this study, "high yields" are defined as protein concentrations of the target protein that exceed 1 mg/mL, whereas "low yields" are defined as concentrations that are less than 0.5 mg/mL. Insect cells are more sensitive in comparison to *E. coli*[121, 150–152], therefore scaling up the expression volumes is a complex process. Oxygen must be evenly distributed throughout the

culture[121, 150–152], which is usually achieved by having enough liquid surface in contact with air at the correct shaking speed. At higher shaking speeds, formation of bubbles can lead to mechanical disruption of cells, which underlined the importance of optimizing the culture conditions to ensure cell viability. In conclusion, this study demonstrated that a 25 mL suspension culture of insect cells produced the highest protein yield. For higher protein quantities, it was not the volume of the individual expression culture that needs to be increased but the quantity of expression cultures. Harvesting after three days was found to be optimal, as prolonged incubation leads to cell death and subsequent protein degradation[121, 150, 151]. Expression in *E. coli* was unsuccessful regardless of the induction method used.

Table 3.1: Expression of the HSV-1 terminase in different expression systems with different parameters. The term "high yields" is employed to denote protein concentrations greater than 1 mg/mL, whereas "low yields" is used to indicate concentrations less than 0.5 mg/mL. Concentrations below 0.2 mg/mL are designated as "very low yields". For *E. coli* two forms of induction were tested, IPTG and an auto-inducing media. For expression in insect cells the cell suspension volume and harvest time was altered. It is important to note that purification protocols were optimized in parallel with expression optimization.

Expression system	Altered parameter	Protein yield	Comment
E.coli, BL21 DE3	1000 mL suspension volume 0.8 mM IPTG	none	no expression
E.coli, BL21 DE3	1000 mL suspension volume auto-inducing media	none	no expression
insect cells, ExpiSf	25 mL suspension volume 3 dpi harvest	medium	highest protein yield obtained
insect cells, ExpiSf	25 mL suspension volume 3.5 dpi harvest	medium	high amount of dead cells
insect cells, ExpiSf	50 mL suspension volume 3 dpi harvest	medium	N/A
insect cells, ExpiSf	100 mL suspension volume 3 dpi harvest	low	N/A
insect cells, ExpiSf	200 mL suspension volume 3 dpi harvest	low	N/A
insect cells, ExpiSf	500 mL suspension volume 3 dpi harvest	low	N/A
insect cells, ExpiSf	1000 mL suspension volume 3 dpi harvest	very low	high amount of dead cells

Following expression, the protein complex required purification. Initially, a standard protocol for protein sample preparation was employed, utilizing Tris- and HEPES-based buffers supplemented with sodium and magnesium ions to maintain stability. The purification process was executed on an automated ÄKTA pure™ system (ÄKTA pure 25, Cytiva, USA), comprising a two-step chromatography procedure that incorporated both affinity and SEC (see **Chapter** 5.1.7). The buffer system was subsequently exchanged for a 150 mM ammonium acetate solution using an ultra-centrifugal filter (Amicon®, 50 kDa MWCO, Merck). However, during the purification a substantial loss of protein was recorded, rendering the concentration unsuitable for nMS analysis. To overcome this limitation, the number of expression cultures was increased and the resulting lysates were pooled before purification. As shown in **Figure** 3.1, after affinity chromatography, many proteins are detected in the sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). The terminase bands (marked in blue) are slightly darker and therefore distinguishable, but the sample still contained contamination from cellular proteins. This is most likely due to the low amount of terminase proteins causing cellular proteins to bind unspecifically to the resin during affinity chromatography. This approach significantly improved protein yield (approximately 0.2 mg/mL) and ultimately enabled the successful acquisition of the first nMS spectrum of the terminase complex using a Q-TOF 2 mass spectrometer.

As seen in **Figure** 3.2, the heterogeneity of the sample leads to increased complexity and thus to overlapping peaks and an overall low resolution of the spectrum. The masses of these proteins are found mainly between 25 to 140 kDa. The mass of the HTC is approximately 194 kDa (see **Table** 3.2) and the peaks are seen at approximately $m/z = 7500$, whereas peaks from the cellular proteins appear between $m/z = 3000 - 7000$. Even though the HTC was found, no higher complexes like the HTC_6 could be observed, which would have been expected according to previous
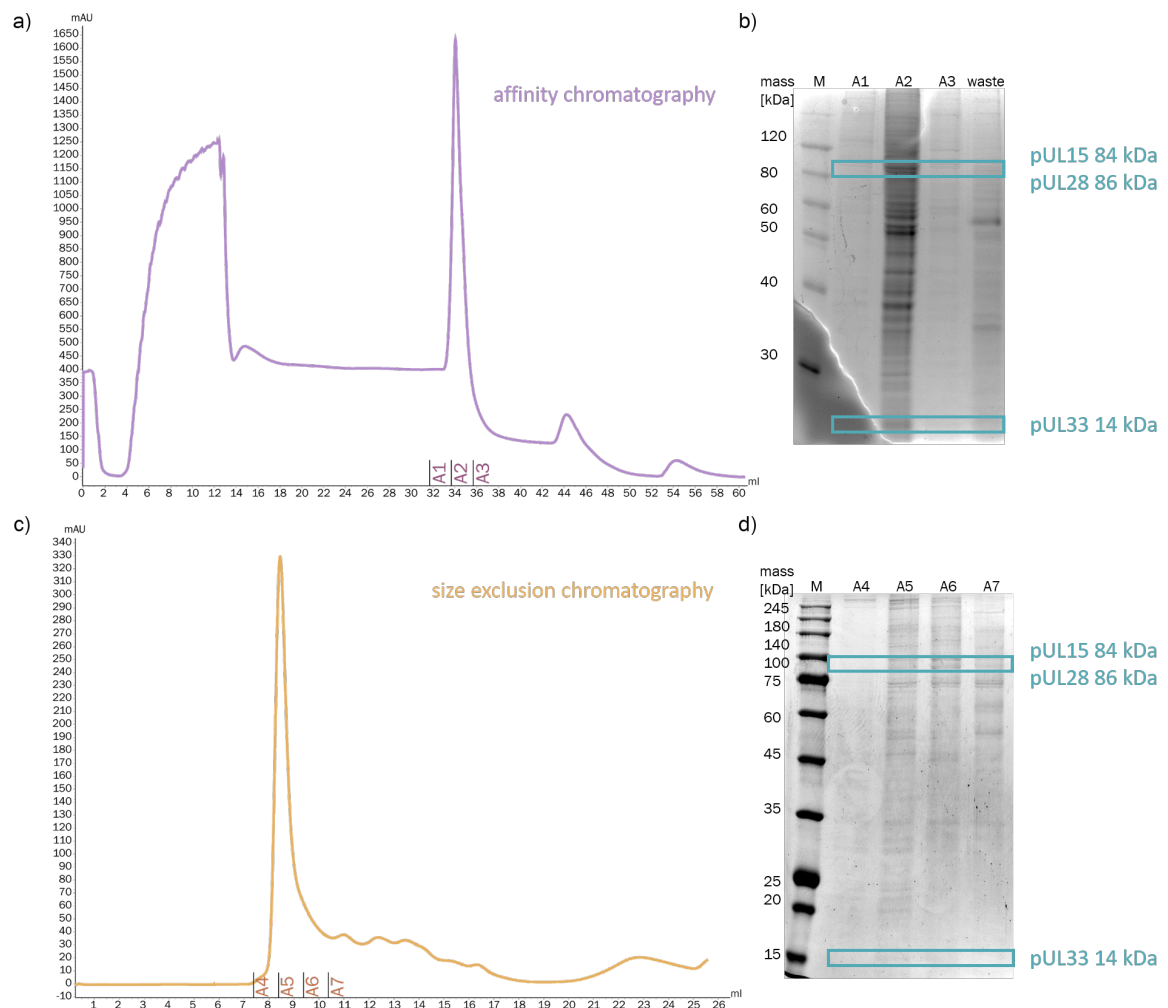
Figure 3.1: Chromatograms and SDS-PAGEs from the terminase purification with ÄKTA™in a Tris-based buffer. In **a)** the chromatogram of the affinity chromatogram with a 1 mL His-Trap HP column (29051021, Cytiva, USA) is depicted. Proteins were detected with a UV detector at 280 nm. Collected fractions are labeled on the bottom. **b)** shows the SDS-PAGE from the collected fractions in a). The terminase proteins are marked in blue. In **c)** the chromatogram of the size exclusion chromatogram with a Superdex™ 200 increase 10-300 GL (28990944, Cytiva, USA) is depicted. Proteins were detected with a UV detector at 280 nm. Collected fractions are labeled on the bottom. **d)** shows the SDS-PAGE from the collected fractions in c). The terminase proteins are marked in blue.

discoveries[29]. The most plausible explanation for this phenomenon is the substantial protein loss that occurs following sample preparation. According to the law of mass action, complexation is concentration depended. Especially larger protein complexes often need sufficient protein concentration to form[153, 154]. This could explain why only HTC can be detected at relatively low concentrations.

Through repeated rounds of HSV-1 terminase expression and purification, it became clear that significant protein loss occurred after affinity purification, as consistently low concentrations were observed following buffer exchange. In **Figure** 3.3 a direct comparison of the signal of His-pUL15 before (**a**) and after SEC (**b**) is shown. In both experiments, the same amount of sample was applied to the gel (10 µL). Bands of pUL15 are clearly visible after affinity chromatography, whereas the western blot after SEC is seemingly empty. Interestingly, the band for His-pUL15 is lower than expected. That is because a truncated version of pUL15, which is about 60 kDa, has been observed to be the most abundant band in SDS-PAGEs. In the literature, a 59 kDa truncated version of pUL15 is described[46]. This fit the band observed in the figure. It has been suggested that this truncated version does not contribute to viral growth[46] and is therefore a mystery for what the role of it is. The 59 kDa band was found consistently throughout the experiments and usually was much more abundant than the 82 kDa band. In a later chapter, this truncated pUL15 will be described in more detail (**Chapter** 3.2.2). After SEC, the sample had to be buffer exchanged through a spin filter, which is usually used to concentrate a sample. Unfortunately, the concentration did not improve and very low protein concentrations (often below 0.2 mg/mL) were recovered. It is well known that proteins can bind non-specifically to matrices and protein loss is to be expected when performing SEC and buffer exchange. In the case of the HSV-1 terminase, protein loss is very substantial, resulting in low protein concentrations that were often too low for proper
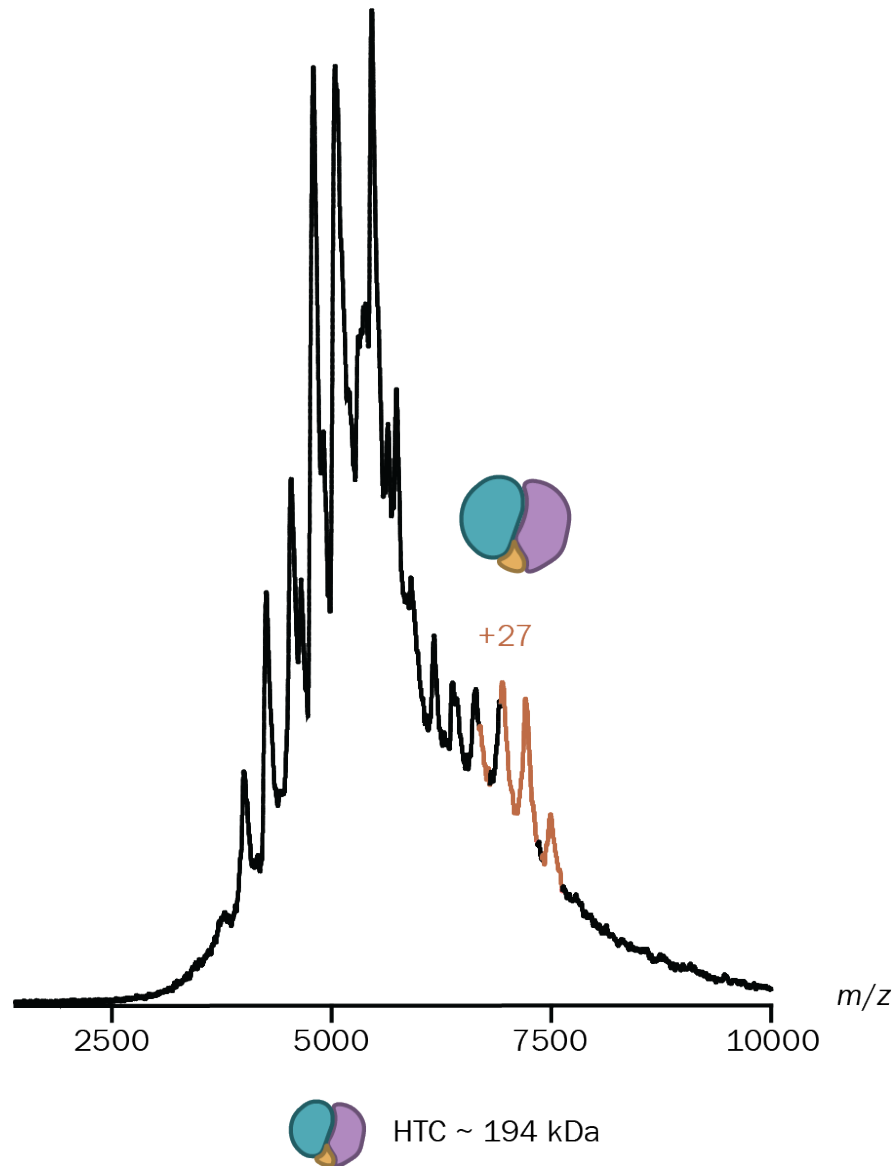
Figure 3.2: Spectrum of the terminase complex (3 µM) after His-tag and SEC purification, followed by buffer exchange into 150 mM ammonium acetate at a pH of 7.5 via ultra filtration. The sample was measured on the Q-TOF 2 at 25 V. The heterogenic trimer complex of the terminase (193 kDa) is labeled. Data were obtained from one biological replicate and one technical replicate.

nMS analysis.
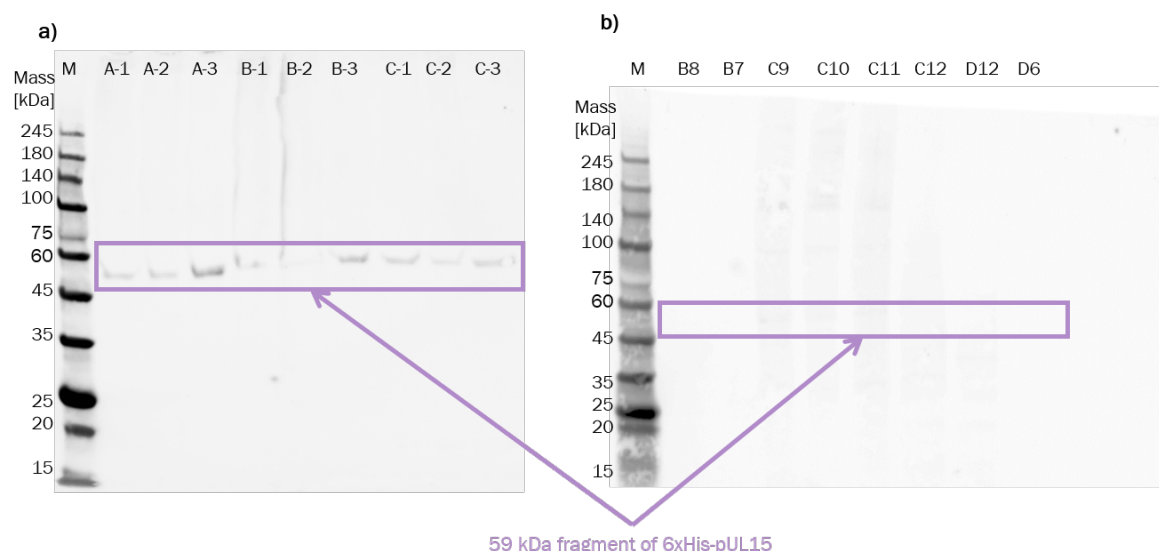
Figure 3.3: Western blots of the same terminase sample before and after SEC with an anti-6xHis antibody. **a)** The first three elutions of a cell suspension harvested 3 dpi (A), 3.5 dpi (B) and 4 dpi (C).**b)** Different elution fractions of sample A associated with peaks observed in the chromatogram. While a signal was detectable after affinity purification, no signal could be observed after SEC, even upon contrast enhancement of the image. For both blots, the same buffers were used and the same amount of sample volume was applied to the gel (10 µL).

To minimize protein loss during automated affinity chromatography and SEC, these steps were omitted. Instead, proteins were purified using affinity chromatography resin on a gravity flow column. After elution, the proteins were buffer exchanged and subsequently analyzed using the Q-TOF 2. As expected, slightly higher protein concentrations (approximately 0.5 mg/mL) were detected by spectroscopy. Due to the omission of SEC, cellular proteins did appear more prominently in the spectra compared to **Figure** 3.2. In contrast to **Figure** 3.2, the dimer of the heteromeric trimer terminase complex (HTC$_2$) is the only terminase complex which was detected. Due to higher cellular protein contamination, peaks are overlapping and therefore not baseline separated. The increased protein concentration allowed for dimer formation, but no other oligomeric states were identified. This suggests that the background signal remains excessively high and the protein concentration insufficient, likely

preventing the detection of low-abundance higher-order oligomeric states.
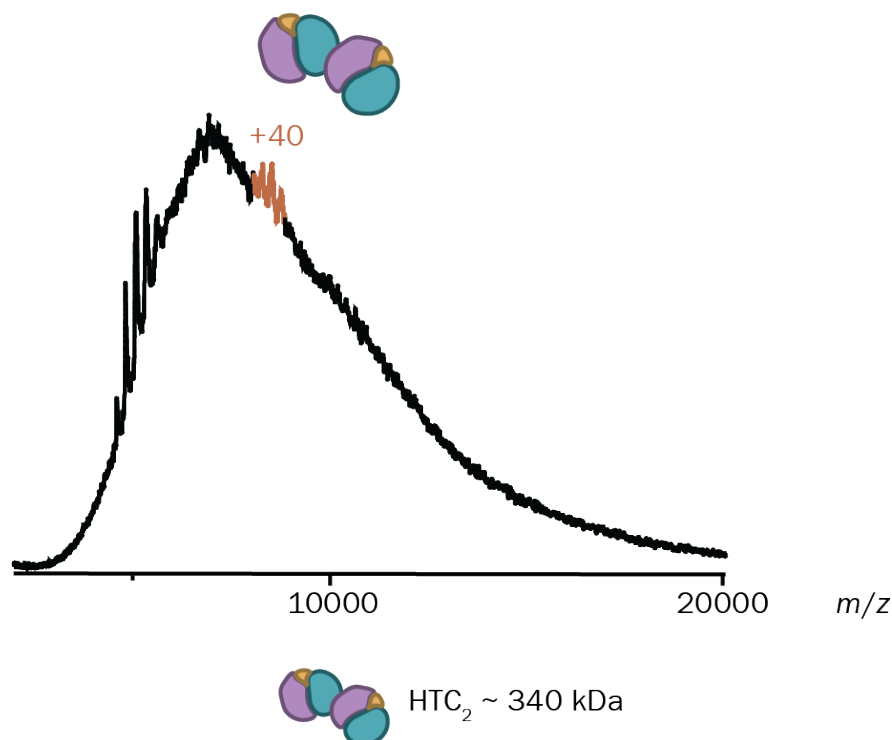
Figure 3.4: Spectrum of the terminase (5 μM) on the Q-TOF 2 in 320 mM ammonium acetate at pH 7.5 at 150 V. The HTC$_2$ (approximately 340 kDa) is labeled. Data were obtained from one biological replicate and two technical replicate.

The spectra in **Figure** 3.2 and 3.4 suggest that protein purification remains suboptimal, likely due to significant protein loss during buffer exchange. This results in elevated contamination signals, which causes peak overlap and prevents the charge state distributions from reaching baseline separation. It is also important to note that three technical replicates could not be obtained due to inconsistencies with the spray and overcrowded spectra. This applies to the measurements shown in **Figure** 3.2 and **Figure** 3.4. Buffer exchange using spin filter or spin columns bears the risk of protein loss, either through precipitation or adsorption of the protein to the matrix or membrane[155, 156]. Since the formation of higher-order protein complexes can be highly dependent on the protein concentration[153, 154], a novel purification strategy had to be developed to study the complexation of the HTC and obtain high-quality

spectra.

Overall, expression screening showed that the HSV-1 terminase complex could not be produced in *E. coli*, regardless of the induction strategy. In contrast, insect cells yielded significantly higher protein levels, especially in small-volume cultures. Scale-up proved challenging due to cell fragility and oxygenation requirements. A standard purification protocol involving affinity and SEC, followed by buffer exchange, resulted in substantial protein loss, likely due to non-specific binding and sample dilution. SEC was found to be particularly detrimental. An optimized purification strategy omitting SEC and using gravity flow columns improved yields and allowed for successful nMS detection of terminase oligomers. However, even this approach showed limited detection of higher-order assemblies, highlighting the sensitivity of complex formation to protein concentration and purification conditions.

3.2.   Fast tracking protein sample preparation for native mass spectrometry

This fast-track protocol was developed with the help of Fatema-Aqila Said, who was a master student during that time. She performed all ADH measurements and prepared **Figure** 3.5 and **Figure** 3.6 under my supervision. Parts of the following chapter are submitted as a preprint:

Grün, A.F.R.; Said, F.A.; Schamoni-Kast, K.; Damjanovic, T.; Bosse, J.; Uetrecht, C., Fast tracking native mass spectrometry: Skipping over buffer exchange, bioRxiv 2025, https://doi.org/10.1101/2025.02.22.639503[157]

The massive loss of the terminase complex during SEC and buffer exchange was the central issue of this study. Due to insufficient protein concentrations, higher-order HTC complexes remained undetectable by nMS. Overcoming this obstacle was crucial and required the development of an alternative sample preparation method. Previous studies have already explored ways to reduce protein loss[158]. The Sharon group, in particular, demonstrated that cell lysis can be performed directly in ammonium acetate, allowing the sample to be sprayed directly into the MS[133, 158–160]. This approach enables the direct analysis of highly over-expressed proteins without further sample processing. However, a major disadvantage of this protocol is that impurities are not removed, which can interfere with measurements, especially when protein concentrations are lower. Since this was the case for the proteins used in this work, modifications to the protocol were necessary. Cell pellets were lysed in ammonium acetate, and the protein was purified via affinity chromatography also using ammonium acetate at physiological pH (7.5-8) and ionic strength (320 to

430 mM). The target protein was then eluted from the matrix using the appropriate amount of the appropriate elution buffer in ammonium acetate. The sample was diluted to a protein concentration of 5 µM or lower and then directly sprayed into the MS. However, elution substances such as imidazole and biotin, used for His-tag and Strep-tag purification respectively, remain in the purified protein sample. Their impact on spectral quality and potential unwanted side effects had to be assessed before applying this method to address biological questions.

3.2.1. Common eluents are suitable for fast-track protein purification

To assess if protein complexes could be identified via nMS in the presence of high concentrations of imidazole, biotin, or desthiobiotin, we first assessed their impact using commercially available, purified, tetrameric ADH. Data was obtained on two mass spectrometers, which are known to have substantially different desolvation/declustering capacity in the source region, with a Q-TOF 2 (**Figure** 3.5) representing "soft conditions" as also encountered on Synapt setups and with the QE-UHMR (**Figure** 3.6) representing a "hot" source. Even under soft conditions on the Q-TOF (**Figure** 3.5), mass spectra with commonly used eluent concentrations such as 20 mM biotin, 2.5 mM desthiobiotin, or 300 mM imidazole it was possible to produce high-quality, charge-state resolved spectra of tetrameric ADH. Good spectra were obtained at medium collision voltages for all tested eluents, allowing for further increase of collision voltage (CV) for CID if needed. At low CV, spectra showed cluster formation for biotin and desthiobiotin, which were easily disrupted at increased CV (see **Figures** A1 to A3). After further declustering of desthiobiotin and biotin, the resulting spectra looked similar to ADH without eluents and charge state distributions were nearly identical. Solely for imidazole, peak broadening was observed compared to ADH spectra without eluent. Even though His-tag purification

required a large amount of imidazole, the CV needed to obtain a useful spectrum was comparable to the other two eluents. As opposed to biotin and desthiobiotin, imidazole is a charge-reducing agent, which leads to peaks at higher m/z[77]. On the Q-TOF 2, higher amounts of eluent resulted in no longer resolvable peaks independent of the CV employed. The upper limit for biotin was 20 mM, which implies that at least a 2.5-fold dilution of the purified sample would be required after elution with the standard eluent concentration of 50 mM. This also applied to imidazole when concentrations higher than 300 mM are required. Strikingly, the upper concentration limit for desthiobiotin was not even reached at 25 mM, far beyond requirements for purification protocols that range from 2.5 mM[131, 161] to 5 mM[162]. Biotin and desthiobiotin also increased the number of detectable charge states compared to pure ADH and when imidazole was present. This can be advantageous when deconvoluting the spectra, since more peaks increase the confidence of the calculated mass. Notably, the QE-UHMR tolerated even higher amounts of eluent and obtaining peak resolution was unproblematic (see **Figure** 3.6). Again, higher than necessary concentrations could be used for desthiobiotin, but now mass spectra with 50 mM biotin or 500 mM imidazole were also resolved (**Figure** 3.6 and A3). The graphs in **Figure** 3.5 **e** and 3.6 **e** show the relation between area under the curve (AUC) and the mean FWHM of the charge states for each eluent. On the Q-TOF 2, the AUCs roughly stay constant (or show a slight increase for biotin) until around 160 V, when the signal breaks down due to tetramer dissociation. However, the FWHM reduced significantly, which was in line with improved declustering, resulting in peak sharpening at elevated CV. On the QE-UHMR, the situation was different. Due to nicely resolved peaks at low CV, the FWHM did not change much for biotin and desthiobiotin, but the transmission improved, resulting in an initial increase of the AUCs. For imidazole, we again observed peak sharpening, which was accompanied by an increase in AUC.

This probably was due to the signal preprocessing in the QE-UHMR, which is known to affect peak shape. Ultimately, both instruments produced high-quality spectra with high amounts of eluents, but the QE-UHMR predictably had an advantage at low protein amounts. The fast-track protocol is ideal for the analysis of intact proteins that have to be purified prior nMS analysis, without the need for additional materials and purification steps.
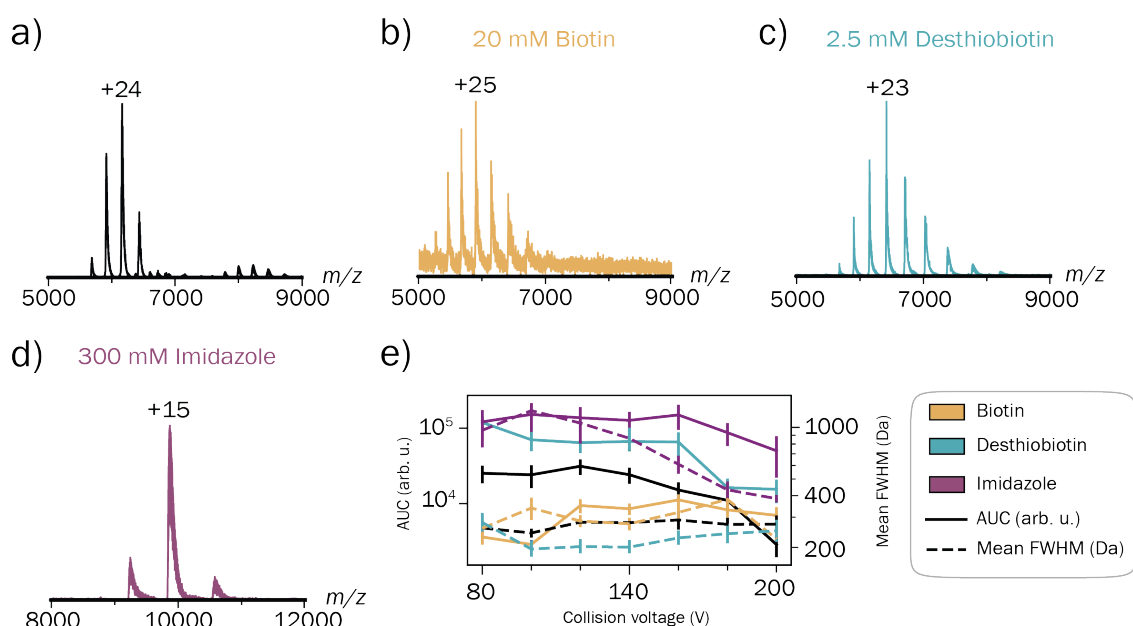
Figure 3.5: Spectra of the tetrameric ADH complex (148 kDa, 5 μM tetramer) in 150 mM ammonium acetate at pH 8 mixed with typical amounts of eluent at optimized collision voltages on the Q-TOF 2. **a)** ADH at 5 V. **b)** ADH with 20 mM biotin at 100 V. **c)** ADH with 2.5 mM desthiobiotin at 80 V. **d)** ADH with 300 mM imidazole at 100 V. **e)** AUC and FWHM are dependent on the CV at the respective eluent concentration. Data were obtained from one biological replicate, with each condition measured in three technical replicates.
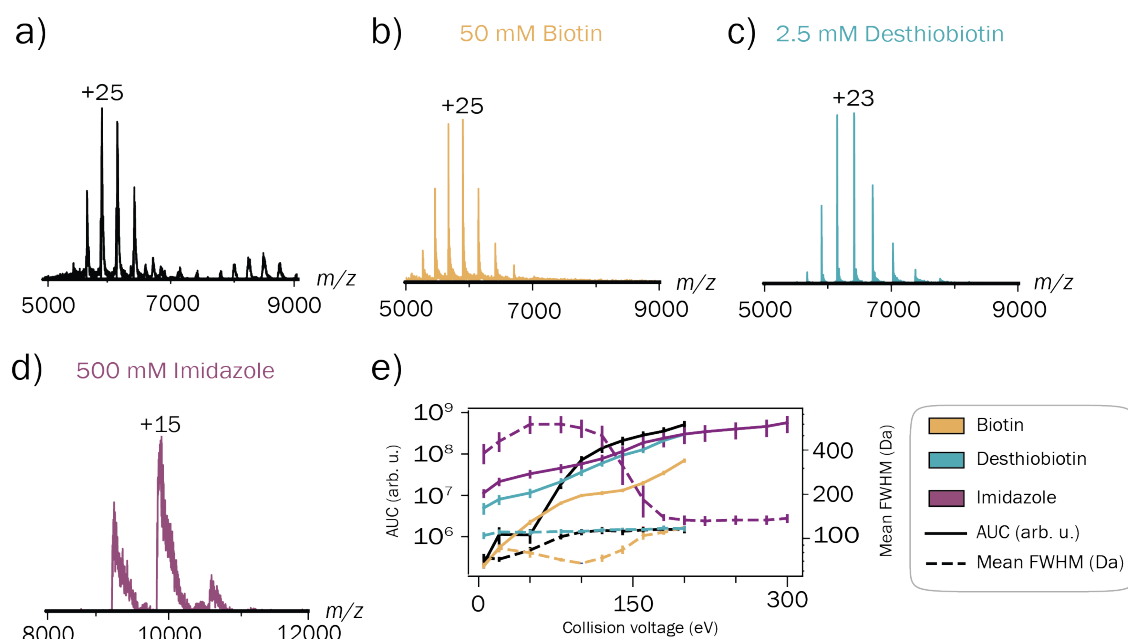
Figure 3.6: Spectra of the tetrameric ADH complex (148 kDa, 2.5 µM tetramer) in 150 mM ammonium acetate at pH 8 mixed with typical amounts of eluent at ideal collision voltages on the Q Exactive UHMR. **a)** ADH at 5 eV. **b)** ADH and 50 mM biotin at 140 eV. **c)** ADH and 2.5 mM desthiobiotin at 100 eV. **d)** ADH and 500 mM imidazole at 50 eV. **e)** AUC and FWHM concerning the CV at the respective eluent concentration. Data were obtained from one biological replicate, with each condition measured in three technical replicates.

3.2.2. Fast tracking the purification of the terminase reveals higher-order complex formation

The fast-track protocol was tailored to meet the demanding requirements of terminase purification, which resulted in higher amounts of protein yields. This is not only due to the elimination of buffer exchange, but also the omission of the SEC purification step. Even with the standard protocol, protein samples remained contaminated with cellular proteins, even after an additional SEC step, there was no concern that removing this step would significantly increase the amount of non-target proteins in the sample. An improved way to reduce the amount of unwanted proteins is to optimize the affinity purification parameters, especially the amount of Ni(II)-matrix used. Since each matrix has specific binding affinity for its respective tag, the amount of bound protein in a given volume of matrix may be calculated. Usually, this can be already found on the data sheet of the purchased resin. By minimizing the amount of resin the tagged proteins that have a higher affinity for the Ni(II)-resin displace unwanted proteins with lower affinity that bind non-specifically due to short histidine stretches. After applying the modified fast-track purification protocol, the sample underwent a dilution process and was subsequently sprayed directly into the Q-TOF 2, yielding the spectra depicted in **Figure** 3.7. The obtained spectra are comparable to the spectrum in reference (**Figure** 3.4). This is because both spectra were obtained from the same cell suspension culture. In contrast to the standard purification protocol, the fast-track protocol led to a higher protein yield of 0.8 mg/mL. The sample was then diluted to the same concentration of 5 µM and injected into the Q-TOF 2. Due to the higher concentration of terminase compared to cellular proteins, the charge state series for HTC, HTC_2 and trimer of the heteromeric trimer terminase complex (HTC_3) were identified. However, the charge state distributions remained

insufficiently baseline separated.
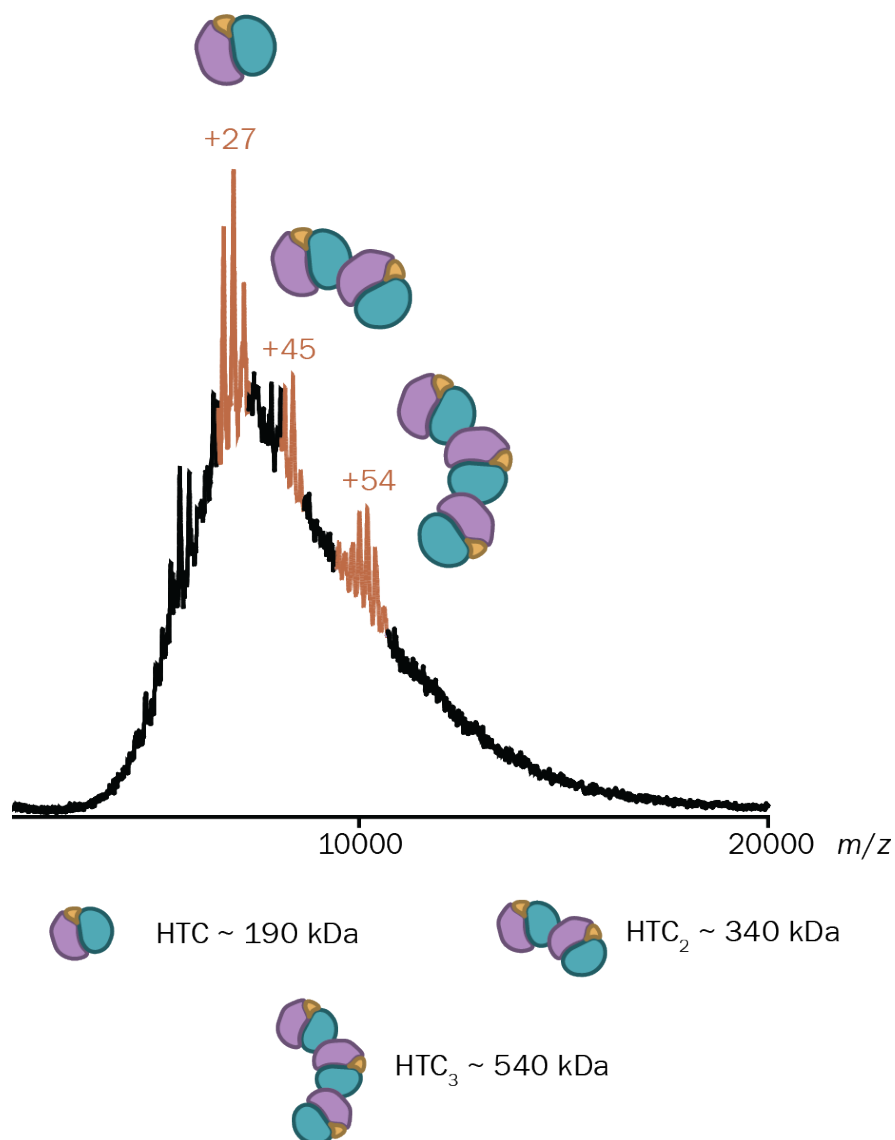
Figure 3.7: Spectrum of the terminase (5 µM) on the Q-TOF 2 in 320 mM ammonium acetate at pH 7.5 at 50 V. The HTC (approximately 190 kDa), HTC$_2$ (approximately 340 kDa) and HTC$_3$ (approximately 540 kDa) are labeled. Data were obtained from one biological replicate and three technical replicates.

As previously described in **Chapter** 3.2.1, the QE-UHMR demonstrates significantly improved transmission efficiency. Given the relatively high mass of the terminase complexes, this likely posed a challenge in earlier experiments. Additionally, the desolvation during ionization can be improved by increasing the temperature of the transfer capillary on the QE-UHMR to 250 °C, which is usually required with

spraying larger protein complexes. These fairly harsh ionization conditions can on one hand help the protein to get into the gas phase, but often lead to dissociation of the protein complex. At 250 °C, the oligomeric states of the terminase were detected, whereas no spectra were obtained at higher temperatures due to signal loss. Lower temperatures did not reveal any higher-order oligomeric states. Finally, a baseline-resolved and charge-state-resolved spectrum of the terminase complex was obtained (**Figure** 3.8). The different complexes are clearly distinguishable, with HTC_2 being the most prominent species in the range of $m/z = 8000 - 10000$. The HTC appears just below $m/z = 8000$, while HTC_3 is the least abundant species, observed around $m/z = 12000$. Similar to **Figure** 3.4, only oligomeric states up to HTC_3 are detectable, which contrasts with the expected predominance of the HTC_6 structure observed in cryogenic electron microscopy (cryoEM)[29]. There may be various reasons for this. Of course, cryoEM and nMS are fundamentally different techniques. In cryoEM, the sample is rapidly frozen down to conserve its structure[163, 164], whereas in nMS the proteins go into the gas-phase. When obtaining a nMS spectrum, protein species with particularly low abundance can fall below a threshold and disappear in the background. With cryoEM, however, even very low abundant species are detectable if the image quality is sufficient. One possible explanation for the absence of HTC_6 in the spectra is the dissociation of the protein complex during nESI, which is why very soft conditions are often employed to mitigate this effect. It was observed that the same species (HTC, HTC_2 and HTC_3) were consistently detected, regardless of variations in experimental protocols and measurement time points (compare **Figure** 3.4 and **Figure** 3.8). Another possibility why the terminase is predominantly present as lower-order complexes could be its dynamic nature. As an enzyme which binds and cleaves DNA, the terminase undergoes constant conformational changes when interacting with specific binding partners. In the case of the HTC, it has

already been proposed that the complex is in a continuous state of flux[28]. The proposed packaging mechanism of the terminase, illustrated in **Figure** 1.4, suggests that a conformational change occurs when DNA binds two HTC[28, 29]. Upon ATP hydrolysis at HTC, a structural shift is triggered at the DNA binding site, causing it to twist upward — generating the characteristic "drilling" motion. The mechanism emphasizes the interaction between two HTC at a time, depicted by a red-marked caricature arm (**Figure** 1.4). This bond appears to be more stable and, therefore, remains intact during ionization.

An important difference between the measurements on the Q-TOF 2 and the QE-UHMR was the quality of the spray. On the Q-TOF 2, only a single spectrum could be obtained for **Figure** 3.2 and just two spectra for **Figure** 3.4, due to the challenges in achieving a stable spray. In contrast, the QE-UHMR allowed for triplicate measurements with relative ease. This suggests that the "hot" source improves desolvation and reduces spectral crowding, thereby facilitating the analysis of the large terminase complex.

The SDS-PAGE and Western blots prepared for QE-UHMR measurements consistently displayed a strong band around 60 kDa—also visible in **Figure** 3.3—along with a weaker band between 25 to 35 kDa. The latter likely corresponds to the truncated pUL15C fragment, which lacks a His-tag and does not associate with the truncated N-terminal, resulting in it ending up in the flow-through (FT) (see **Figure** 3.9 **a**). Since the sample was in ammonium acetate, this band migrated slightly higher than observed in previous experiments (**Figure** 3.3). Interestingly, despite the presence of this strong band on the gel, the nMS spectrum did not reveal any peaks corresponding to the 59 kDa species. This suggests that the truncated 59 kDa variant of pUL15 may still interact with pUL28 and pUL33, forming a terminase complex. A previous study proposed that the truncated 59 kDa version of pUL15 does not participate
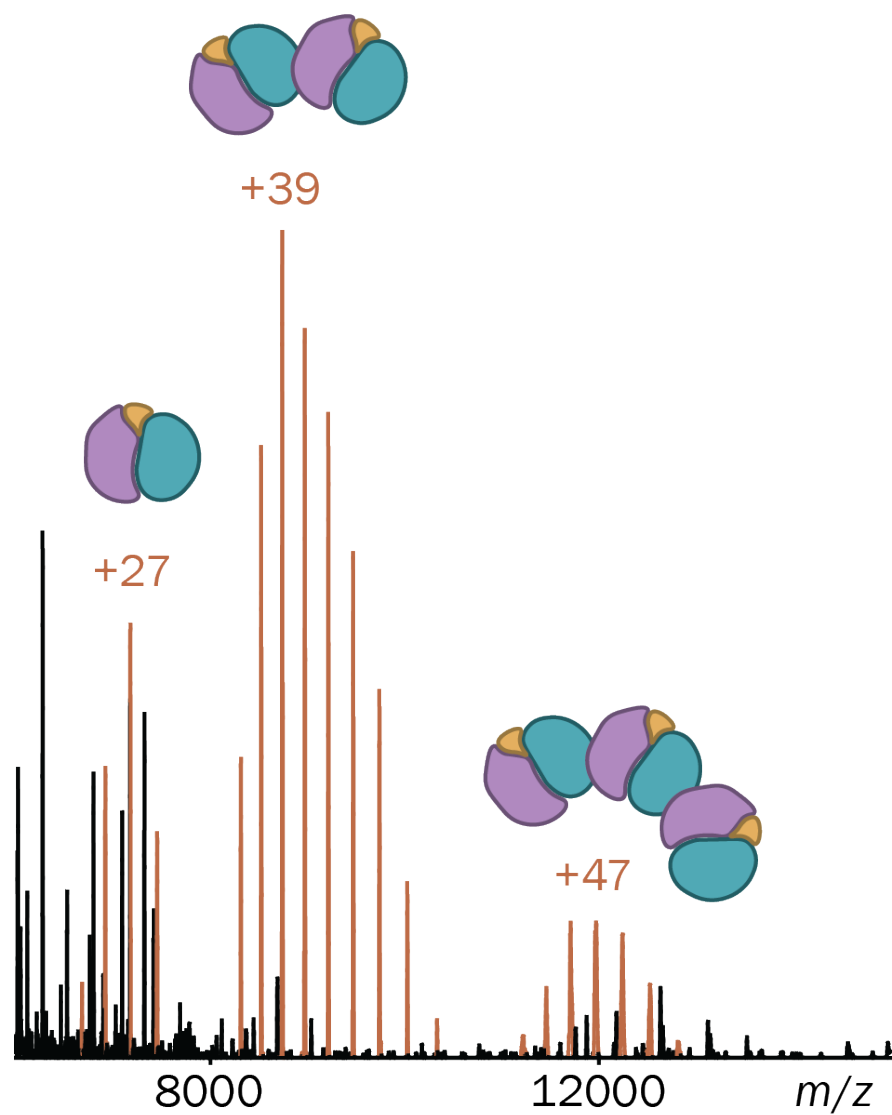
Figure 3.8: Spectrum of the terminase complex (3 µM) in 320 mM ammonium acetate and 35 mM imidazole at pH 7.5 on the Q Exactive UHMR with turned off CV. The HTC (approx. 190 kDa), HTC$_2$ (approx. 340 kDa) and HTC$_3$ (approx. 540 kDa) are labeled. Data were obtained from one biological replicate and three technical replicates.
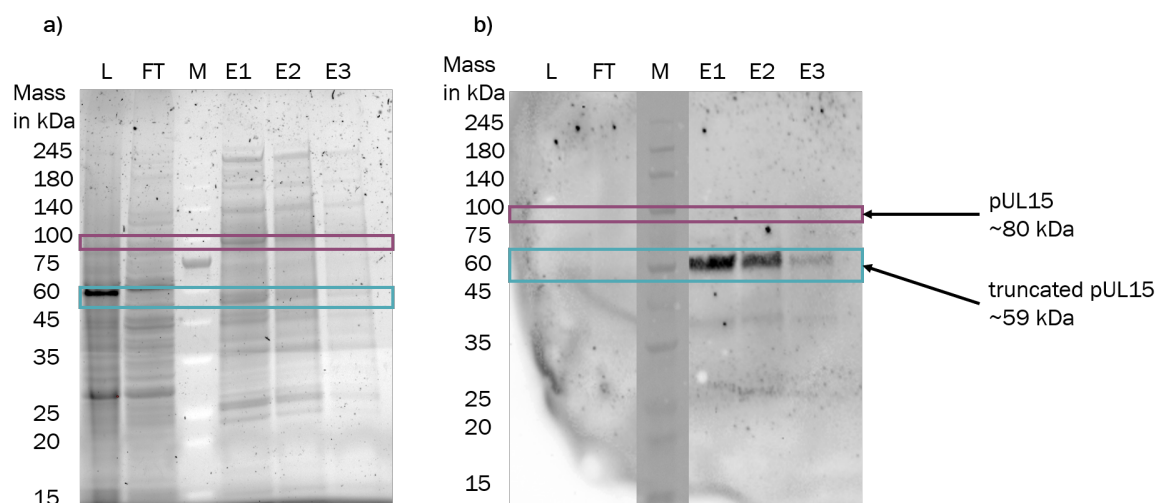
Figure 3.9: Fractions of the fast-track purification of the terminase with L being the lysate, FT being the flow-through, M being the marker molecular weights in kDa and E1-3 being the elutions in their respective order. a) SDS-PAGE, b) western blot (WB) with antibody staining against the 6×-His-tag of His-pUL15.

in terminase complex formation, but the here presented data suggests otherwise (this will be discussed further in **Chapter** 3.3.5)[46]. Notably, this truncated variant has also been observed during infection of Vero cells with wild-type HSV-1, not only in recombinant expression systems such as *E. coli*[165]. This indicates that the truncated pUL15 is not merely an artifact resulting from proteolytic degradation during heterologous expression, but likely occurs naturally. This phenomenon may be explained by the genomic structure of pUL15. The gene is encoded across two exons[166] and the C-terminal 30 kDa fragment can be expressed independently of the full-length protein. This is due to Met^{443}, which is in-frame and located in the second exon[167]. Moreover, the C-terminal fragment has been reported to be dispensable for viral replication[167]. Furthermore, It is noteworthy that Met^{443} is not well conserved among herpesviruses; for instance, HCMV and HHV-6 possess different amino acids at this position[167]. Taken together, the most likely explanation for the frequent detection of the 60 kDa truncated version of pUL15 in this study is that it results from

proteolytic cleavage after expression. However, the precise mechanism behind this truncation remains unclear. To better understand this process, future experiments should aim to identify the interaction partners of both the N-terminal and C-terminal fragments. This could be achieved through co-immunoprecipitation followed by Western blotting and mass spectrometry. Both bottom-up and native MS approaches would be suitable for this analysis.

It is also important to note a significant limitation of the QE-UHMR instrument used in this study. When selecting specific ions for fragmentation in the HCD cell, no dissociation products were observed—likely due to the large size of the precursor ions. A common strategy for validating oligomeric assemblies is to induce dissociation in the collision or HCD cell, allowing for analysis of both the released monomers and the remaining complex[102, 168]. This enables confirmation of stoichiometry and monomer mass. Unfortunately, this approach was not successful for any of the proteins analyzed in this work and remains a task for future investigation.

In summary, implementation of a fast-track purification protocol enabled the efficient isolation of the HSV-1 terminase complex, yielding sufficient quantities for high-resolution nMS analysis. Mass spectra revealed the consistent presence of trimeric (HTC_3), dimeric (HTC_2) and monomeric (HTC) assemblies, but not the expected hexameric form (HTC_6), likely due to dissociation during ionization or the dynamic nature of the complex. A consistently detected approximately 60 kDa band on the SDS-PAGE was attributed to a truncated form of pUL15, which appears to participate in complex formation, despite earlier reports. These findings suggest natural proteolytic cleavage and potential alternative translation from the second exon. Limitations of the QE-UHMR instrument prevented successful dissociation of large precursor ions in the HCD cell, leaving oligomer validation via MS^2 for future experiments.

3.3.   Terminase binding to *pac* sequences

To investigate the interaction between the terminase complexes and DNA, an additional experiment was conducted using *pac* 1, which has previously been reported to bind the complex[36]. Interestingly, even though HSV-1 is a dsDNA virus, it appears that the terminase binds ssDNA[31].Therefore, a fresh terminase sample was mixed with a ssDNA strand containing the G4 motif and subsequently introduced into the MS. The resulting spectrum is shown in **Figure** 3.10. In the spectrum, peaks slightly above $m/z = 12000$ are observed, corresponding to the HTC_3. However, no distinct peaks corresponding to HTC or HTC_2 could be identified. In its unbound state, the terminase complex exists in three distinct stoichiometric forms, as shown in **Figure** 3.8. Although the dimer appears to be the most dominant species, the addition of *pac* 1 shifts the equilibrium primarily toward HTC_3. Meanwhile, HTC and HTC_2 become indiscernible within the noise, thereby precluding further deconvolution. This can be clearly seen when comparing the spectra with and without *pac* 1 (see **Figure** 3.11). This observation is quite surprising, as it suggests the formation of a transitional state in which HTC_3 is stabilized while bound to *pac* 1. Notably, no HTC_6 was detected, indicating that its formation is not limited by DNA stabilization. Instead, DNA appears to stabilize HTC_3 specifically. This finding aligns with the proposed terminase packaging mechanism (**Figure** 1.4)[29]. According to this model, a HTC_2 recruits DNA, which then facilitates the binding of another HTC subunit to both HTC_2 and DNA in a slightly offset manner. This enables the terminase to bind to a position of the DNA situated slightly upstream, thereby facilitating the forward movement of the DNA strand. The mechanism would explain how DNA is displaced into the procapsid. The trimeric complex may then interact with another HTC while releasing the first HTC, with the energy required to break these bonds provided by

ATP hydrolysis. Naturally, this hypothesis remains speculative and would require thorough experimental validation. A possible strategy would be to monitor the dynamics of the packaging process by adding ATP to the sample. This may lead to a loss of the HTC_3 signal—attributable to its short lifetime—or to fluctuating intensities of HTC, HTC_2 and HTC_3.

Additionally, it was necessary to understand the role of the highly conserved G4 of *pac* 1, to get a more in-depth understanding of the terminase-DNA complex that forms during packaging[36]. These secondary DNA structures often hinder the binding of DNA to enzymatic proteins, including the terminase. This regulation occurs through either complexation or destabilization of the G4, leading to the linearization of DNA[41, 169–171]. Since the aim was to assess the role of the G4 structure in protein–DNA binding, it was essential to additionally investigate the conformation of the *pac* sequence using CD spectroscopy. This ensured that *pac* 1 adopts the expected secondary structure in nMS suitable solvent conditions.

Taken together, binding of the *pac* 1 sequence to the terminase complex shifted the oligomeric equilibrium toward the trimeric HTC_3 state, while HTC and HTC_2 became undetectable in the spectra. This finding suggested that ssDNA binding stabilized the HTC_3 conformation, supporting a model in which DNA recruits and aligns terminase subunits for directional genome packaging. No HTC_6 was observed, indicating that higher-order assembly was not DNA-stabilized under the tested conditions.
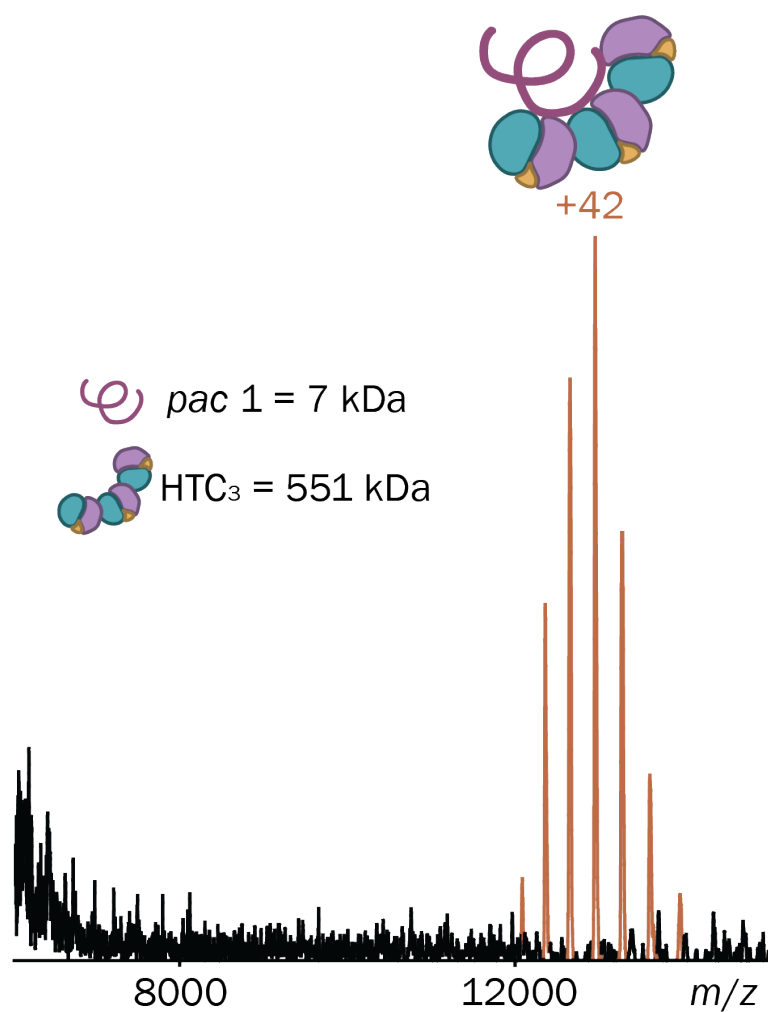
Figure 3.10: Spectrum of 3 µM terminase (HTC) in presence of 9 µM *pac* 1, recorded in 320 mM ammonium acetate and 35 mM imidazole at pH 7.5 on the QE-UHMR. Data were obtained from one biological replicate and three technical replicates.
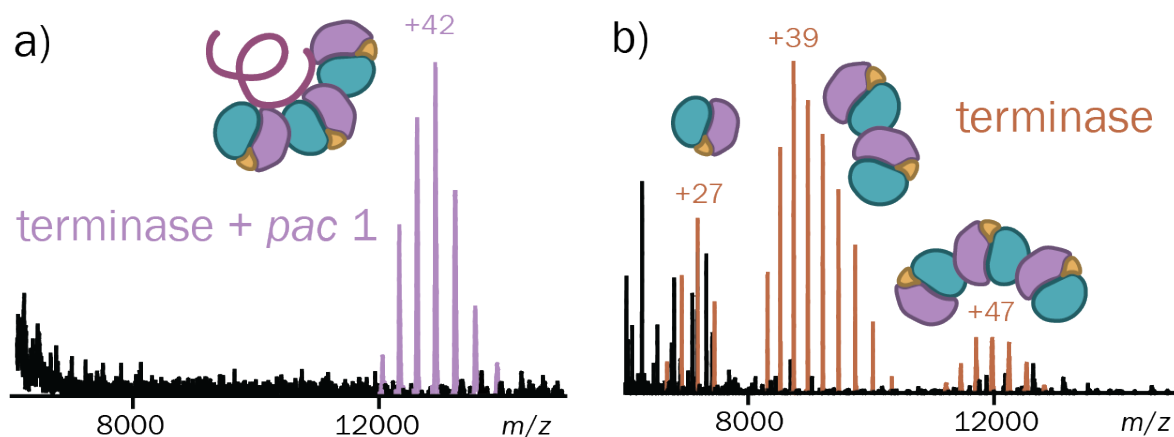
Figure 3.11: Spectra of 3 µM terminase (HTC) in 320 mM ammonium acetate and 35 mM imidazole at pH 7.5, measured in the presence of a) 9 µM *pac* 1 and b) without any added DNA. Data were obtained from one biological replicate, with each condition measured in three technical replicates.

3.3.1. CD spectroscopy reveals G4 formation only in ssDNA strands of *pac* 1

Single-stranded oligomers of *pac* 1 and *pac* 2 were analyzed using CD spectroscopy to assess the stability of the *pac* 1 secondary structure. Previous studies have shown that *pac* 1 contains a conserved G4 motif across all herpesviruses[36]. To evaluate the stability of these structures under nMS-compatible conditions, CD spectra were first recorded at varying K^+ concentrations. Measurements were also performed in an ammonium acetate-based buffer derivative, trimethylammonium acetate (TMAA). G4 structures are typically stabilized by monovalent cations such as K^+, which occupy the central cavity between stacked guanines[38, 39, 172]—the use of potassium is incompatible with nMS. Therefore, the minimum concentration required for G4 stabilization was determined. Additionally, NH_4^+ ions can also stabilize G4 structures, as they fit into the central pocket. However, to avoid unintended stabilization during CD spectroscopy, TMAA was chosen for its bulkier cation, trimethylammonium, which is too large to enter the G4 core[172]. The identity of the cation is crucial, as it can influence both the stability and conformation of the G4 structure. Before

analyzing the *pac* sequences from HSV-1, corresponding sequences from HCMV were tested at different KCl concentrations (see **Figure** 3.12a). All spectra exhibited a positive signal at 210 nm and a maximum at 265 nm, with the sample lacking KCl showing an additional maximum at 280 nm. A minimum was observed at 245 nm across all samples. These features suggest that the ssDNA sequences are capable of forming G4 structures even in the absence of K^+, indicating a relatively stable complex. The intensity of the G4-associated signals increased with higher K^+ concentrations, reaching a plateau at approximately 500 µM. Based on the spectral features—namely, a maximum at 265 nm and a minimum at 245 nm—the *pac* 1 sequence adopts a parallel G-quadruplex conformation[172].

In **Figure** 3.12b, spectra of HCMV *pac* 1 and *pac* 2 ssDNA (both at 1 mM KCl) are compared. The *pac* 2 sample serves as a negative control and does not form a G4. In contrast, the double-stranded form of *pac* 1 (see **Figure** 3.12c) shows a broader peak centered around 270 nm rather than the characteristic 265 nm, resembling the *pac* 2 profile. This suggests that *pac* 1 forms a G4 structure only in its single-stranded form. This observation is unexpected, given that herpesviruses are dsDNA viruses and that the terminase packages the concatemeric dsDNA into the procapsid[27, 31, 44]. The same experiment was repeated for single-stranded HSV-1 *pac* sequences at 1 mM in the absence of potassium ions. As shown in **Figure** 3.13, the *pac* sequence in HSV-1 also adopts a stable parallel G4 conformation, with a minimum at 245 nm and a maximum at 265 nm[172]. Given the high sequence conservation of *pac* among herpesviruses, the behavior of the dsDNA version in HSV-1 is expected to mirror that of HCMV. The double-stranded *pac* 1 sequence of HSV-1 was analyzed in ammonium acetate to simulate nMS buffer conditions. Indeed, **Figure** 3.13 **c** shows a similar profile to the negative control, with a pronounced minimum at around 210 nm, indicating the absence of G4 structures in the double-stranded form. This

absence may be explained by two possibilities: either strand separation occurs during packaging—via a mechanism analogous to transcription—or the G-quadruplex is not stabilized under native conditions. A single-stranded DNA region is likely transiently exposed during packaging, as G4 motifs often function as regulatory elements that modulate binding[170]. This aligns with previous findings that the terminase subunit pUL28 binds specifically to ssDNA, likely due to local denaturation of the concatemeric dsDNA[31]. Furthermore, secondary structure formation within *pac* 1, later identified as G4, has been shown to enhance terminase binding affinity[31, 36]. To ensure compatibility with nMS and to minimize the concentration of non-volatile salts, all nMS experiments were conducted without the addition of K^+ ions.

In summary, CD spectroscopy confirmed that the parallel conformation of the G4 structure within *pac* 1 forms only in its single-stranded form and is stabilized by increasing K^+ concentrations. Both HCMV and HSV-1 *pac* 1 sequences exhibited characteristic spectral features of parallel G4 structures, whereas the double-stranded forms did not. This suggests that the G4 motif is destabilized in the duplex context. The findings align with previous reports that pUL28 preferentially binds ssDNA, implying that transient strand separation during packaging may expose regulatory structures such as G4.
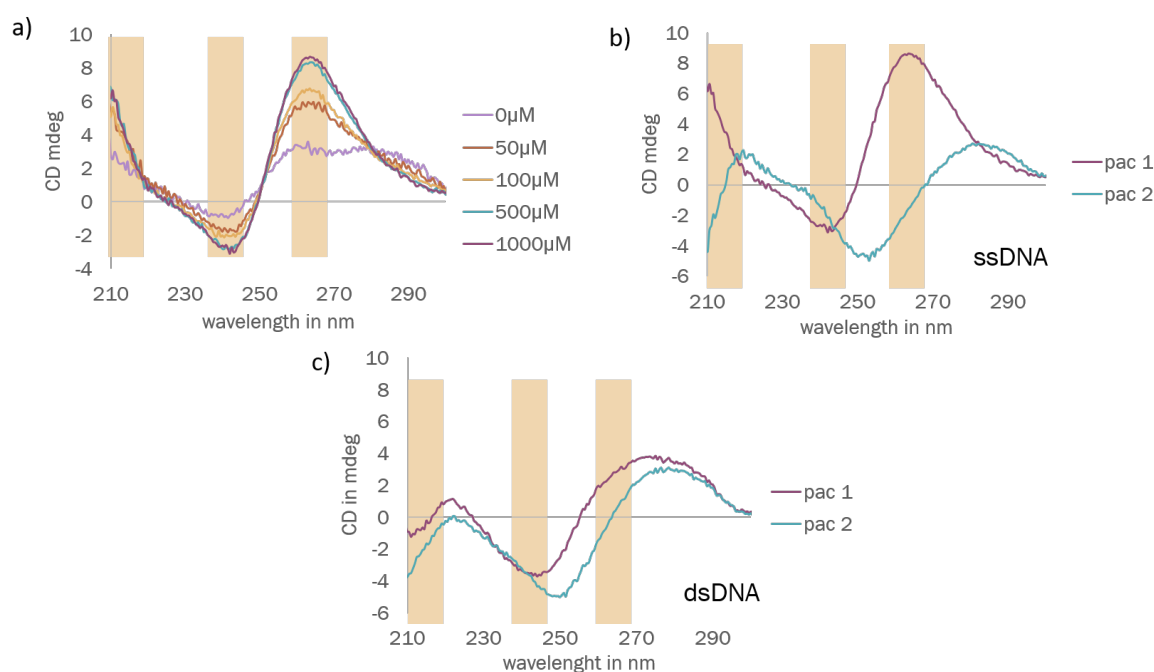
Figure 3.12: Circular dichroism spectroscopy of HCMV *pac* sequences. The yellow areas indicate at which wavelengths the CD spectrum of parallel G4 complexes show minima and maxima[172]. **a)** By increasing the KCl concentration, the G4 structure of the ssDNA *pac* 1 sequence is stabilized to a parallel conformation. When no K$^+$ ions are present, two maxima can be detected, one at 265 nm and one at around 280 nm. Increasing the concentration leads to the loss of the maxima at 280 nm and intensity increases at 265 nm. For all concentrations, a minimum can be identified at 245 nm. This suggests the stabilization of a parallel conformation by K$^+$ ions. **b)** The ssDNA *pac* sequences at 1 mM KCl show a clear difference between *pac* 1 and 2 where *pac* 2 serves as a negative control and *pac* 1 shows a maximum at 265 nm and a minimum at 245 nm that suggests a parallel G4 conformation. **c)** The dsDNA *pac* 1 and 2 sequences at 100 μM of KCl shows no G4 formation for both oligomers. Data were obtained from one biological replicate, with each condition measured in three technical replicates.
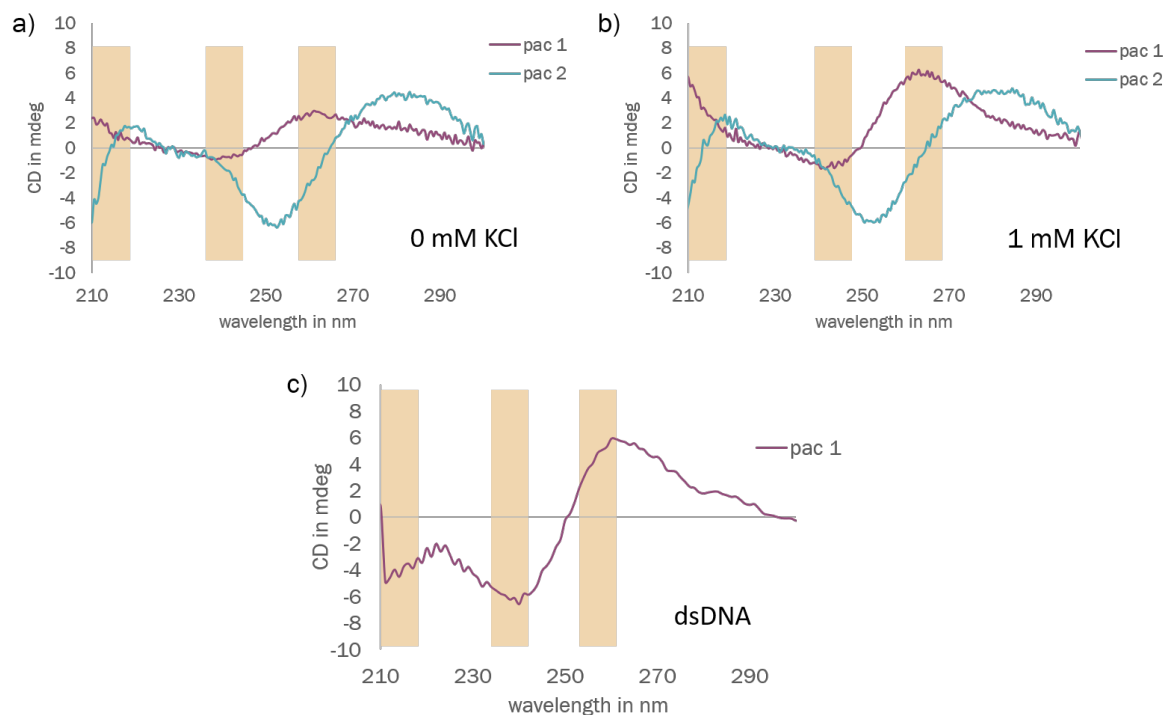
Figure 3.13: Circular dichroism of HSV-1 *pac* 1 and 2 sequences at 0 (**a)**) and 1 mM KCl (**b)**). The yellow regions highlight the spectral ranges where parallel G4 complexes exhibit characteristic minima and maxima[172]. Independent if KCl is present or absent, the graph of *pac* 2 remains unchanged. In contrast, the graph of *pac* 1 in the absence of KCl exhibits maxima at the same wavelengths as the sample containing 1 mM KCl, albeit with lower intensities. Data were obtained from one biological replicate, with each condition measured in three technical replicates. **c)** shows the graph of dsDNA *pac* 1 in 320 mM ammonium acetate. Both the maximum and minimum are slightly shifted toward higher wavelengths and a negative signal is observed around 210 nm. This indicates that the G4 structure is not stabilized in the double-stranded DNA.

3.3.2. The *pac* sequences bind to the terminase independently of their secondary structure

In addition to the previously described ssDNA *pac* 1 and *pac* 2 samples, two additional variants, *pac* 1 mT and *pac* 1 mG, were designed to gain more in-depth insight into the terminase binding mechanism in the context of the G4 structure. In *pac* 1 mG, multiple guanosine residues were replaced with cytosine, thereby disrupting G4 formation (see **Table** A1). In contrast, *pac* 1 mT preserved the G4 structure but involved the substitution of a single thymine within the T-stretch to a guanine. A three-fold molar excess of DNA was added to ensure sufficient binding to the terminase. However, further increasing the concentration of *pac* DNA was not feasible due to residual salts in the samples, which caused spectral crowding. Upon mixing terminase with *pac* 1, *pac* 2, *pac* 1 mT and *pac* 1 mG and analyzing the complexes using QE-UHMR, it became evident that all four *pac* sequences bind to HTC_3 (**Figure** 3.14). The peaks corresponding to the HTC_3-*pac* complexes (slightly above m/z = 12000) aligned almost perfectly across all samples, which was unexpected since initially, only *pac* 1 (purple) was hypothesized to bind. Interestingly, *pac* 2 (blue), though slightly larger than the other sequences (see **Table** A1), appears at the same m/z ratio, contrary to the expected shift towards higher m/z values. This anomaly may be due to dissociation of unbound DNA ends during ionization at high-temperatures (250 °C). The absence of free DNA peaks in the spectra supports this interpretation. Notably, the complexes formed with *pac* 1 and *pac* 1 mG appeared at slightly higher charge states, though still overlapping with other samples. Given the small mass differences between the *pac* variants, this shift likely results from reduced protonation. For *pac* 1, the compact G4 structure may hinder proton access. While a similar behavior was expected for *pac* 1 mT, its most intense peak corresponds to a slightly higher charge state, possibly

due to partial destabilization of the G4 caused by the thymine-to-guanine mutation, leading to a less stable G4 and thus more easily protonated structure during nESI. The behavior of *pac* 1 mG is more complex. Although its dominant peak aligns with those of *pac* 1 mT and *pac* 2, its higher m/z peaks exhibit asymmetry, suggesting the presence of two DNA conformers. One, likely more open and flexible, overlaps with the other samples, while the second, less flexible species appears in a minor, asymmetric distribution. However, this interpretation is challenged by the possible loss of terminal G-stretches critical for G4 formation during ionization. Thus, if the G4 structure forms in solution, it may be stabilized by terminase binding before mass spectrometry analysis. Alternatively, the protonation differences may have another, as yet unidentified, cause. The current data cannot conclusively support any single hypothesis. To clarify these observations, further experiments are necessary. First, nMS of the DNA strands should be conducted—though this was not feasible here due to the presence of non-volatile salts. Future nMS experiments should ideally be coupled with ion mobility (IM), as linear and G4 DNA show distinct IM profiles, enabling conformation analysis in the gas phase[173].

Since HTC$_3$ forms complexes with all *pac* sequences, two explanations are possible: either the terminase binds to a T-stretch unaffected by the single mutation in *pac* 1 mT, or it binds non-specifically to various DNA strands. The latter is supported by the broad peaks of HTC$_3$ in **Figure** 3.8, where no exogenous DNA was added, implying binding to a heterogeneous mix of DNA, potentially including short cellular DNA fragments. Theoretically, it is plausible that the terminase exhibits low sequence specificity when binding DNA, as it must bind the entire viral genome during packaging. In bacteriophages, the large terminase subunit, functionally analogous to pUL15, exhibits a slightly higher affinity for specific sites on the DNA (analogous to *pac* 1 and 2) during cleavage. However, it is also capable of binding to different DNA
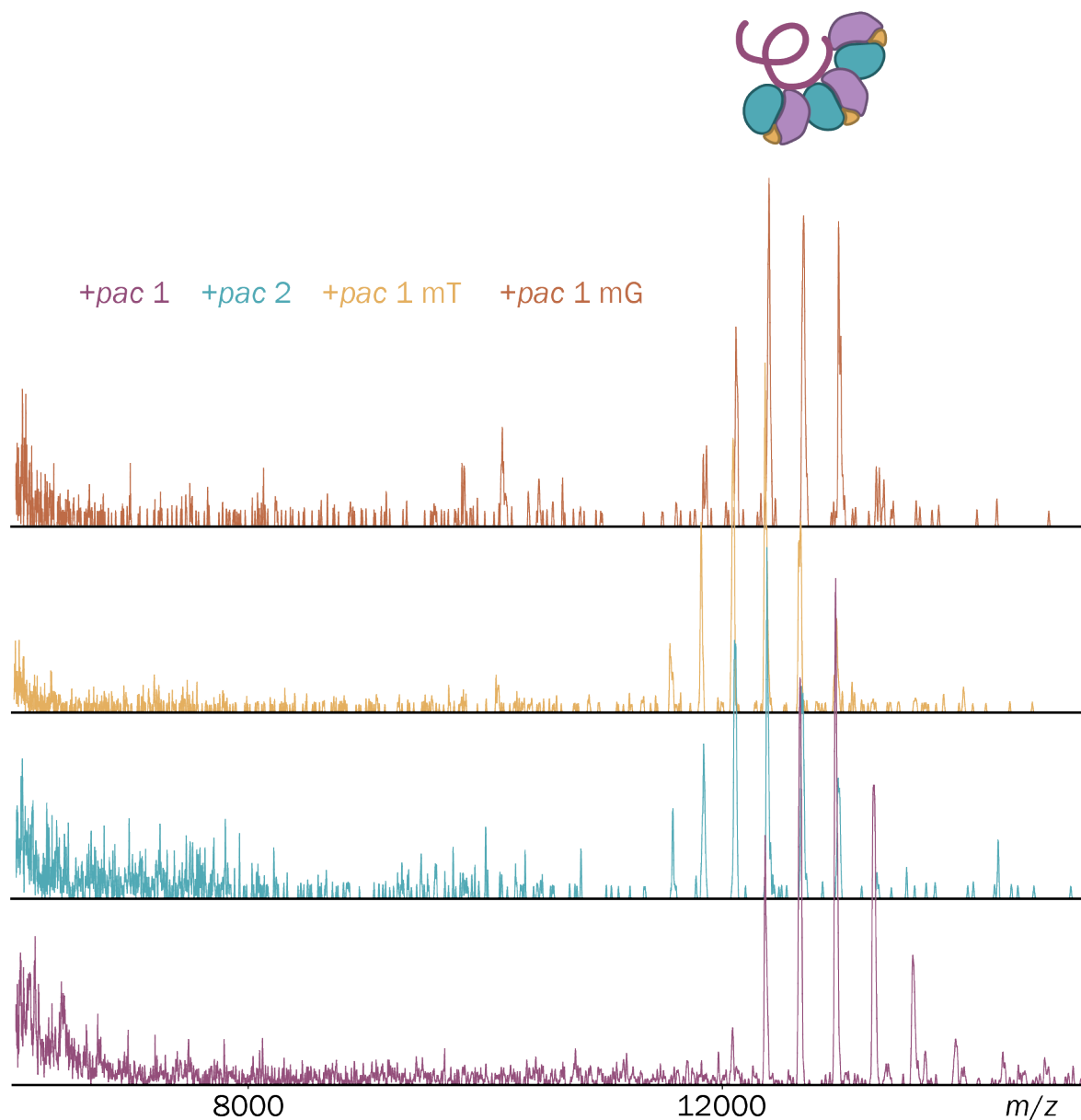
Figure 3.14: Stacked mass spectra of 3 µM terminase complex incubated with 9 µM of various *pac* DNA sequences in 320 mM ammonium acetate and 35 mM imidazole at pH 7.5 on the QE-UHMR. Samples are color-coded as follows: *pac* 1 (purple), *pac* 2 (blue), *pac* 1 mT (yellow) and *pac* 1 mG (orange). While the spectra largely overlap, the *pac* 1 sample displays a slight shift toward higher charge states. Data were obtained from one biological replicate, with each sample measured in three technical replicates.

strands, indicating a degree of sequence flexibility[26].

To test whether G4 stabilization affects binding, this experiment should be repeated with low K^+ concentrations. Prior studies strongly suggest that such concentrations can stabilize G4s without compromising nMS data, whereas NH_4^+ offers weaker stabilization[39, 172]. Proper G4 formation could clarify whether it sterically hinders terminase binding or enhances affinity particularly for *pac* 1, which also functions as a cleavage signal[36]. To investigate differences between linearized and structured *pac* 1, binding experiments should be performed at varying K^+ concentrations using high *pac* 1 levels to outcompete cellular DNA for pUL28 binding. To analyze the conformation of *pac* sequences IM could be performed before nMS analysis. Additionally, cross-link (XL)- or hydrogen-deuterium exchange (HDX)-MS could be used to pinpoint binding sites on *pac* 1. Cross-linking or deuteration would help localize the terminase interaction region via bottom-up MS, offering detailed insights that remain poorly defined in current literature[34–36, 55, 174].

Taken together, all tested *pac* variants, including mutated and structurally distinct sequences, formed stable complexes with HTC_3, suggesting that terminase binding is largely independent of G4 structure or precise sequence context. Minor shifts in charge states point to subtle differences in protonation and conformational flexibility. The data support a model in which terminase exhibits low sequence specificity, potentially allowing interaction with various cellular or viral DNA fragments. Further experiments, including IM and binding assays under varying ionic conditions, are required to clarify the influence of G4 stabilization and to map precise binding sites on *pac* DNA.

3.3.3. The predicted HTC$_3$ structure is stabilized by ssDNA

With the machine learning-based AlphaFold3 tool[175, 176], predicting protein complex structures from sequences has become highly accessible[177, 178]. While we cannot necessarily determine which stoichiometry is the most favorable, we can assess the stability of specific stoichiometries and explore their most probable structural conformations[175, 177–179]. The HTC$_3$ complex presents an intriguing case, as its predicted structure exhibits drastic differences depending on whether it is bound to DNA or not (see **Figure** 3.15 and 3.16). In the absence of DNA, the HTC subunits arrange themselves in a triangular formation (**Figure** 3.15). However, when *pac* 1 is introduced, the formation adopts a half-moon shape (**Figure** 3.16). This is likely due to the small inner channel diameter, which is too small (approximately 10 Å) for DNA interaction in the triangular conformation[28]. In contrast, the half-moon shape appears to facilitate a more stable DNA binding interface because it is more accessible, as it is open. These findings align well with the previously discussed nMS data that the addition of ssDNA stabilizes a HTC$_3$ conformation. This results in the observed DNA-HTC$_3$ peaks seen in **Figure** 3.11. Notably, when dsDNA was included in the structural prediction, the algorithm struggled to incorporate it as a double-stranded molecule. This can be seen in **Figure** 3.16, where the G4-containing strand is shown interacting with the protein complex, while the complementary strand remains unbound. This again aligns with data from the literature that pUL28 binds ssDNA[31]. AlphaFold3 evaluates the confidence of predicted structures based on both the overall protein fold (predicted template modeling (pTM)) and the arrangement of subunits (interface predicted template modeling (ipTM))[180, 181]. The prediction for the triangular conformation of HTC$_3$ shows low confidence scores (pTM = 0.50, ipTM = 0.48), falling below the thresholds for a reliable prediction. In contrast, the alternative half-moon-

shaped conformation of HTC$_3$, which includes DNA as a stabilizing ligand, yields moderately confident scores (pTM = 0.67, ipTM = 0.64), suggesting a more plausible structural model at least at the level of subunit folds. AlphaFold3 evaluates the confidence of predicted structures based on both, the overall protein fold (pTM) and the arrangement of subunits (ipTM)[180, 181]. A prediction with a pTM above 0.5 and an ipTM over 0.8 is considered confident and high-quality. For the triangular structure, both the ipTM and pTM scores are relatively low (0.48 and 0.5, respectively), with the subunit layout falling below the threshold for a reliable prediction. A pTM score below 0.5 and an ipTM score below 0.6 indicates an unreliable prediction. Consequently, the triangular conformation in **Figure** 3.15 is unlikely to be a stable structure, supporting the hypothesis that HTC$_3$ is stabilized by DNA and adopts a half-moon shape, as seen in **Figure** 3.16, instead ($ipTM = 0.64, pTM = 0.67$). In this configuration, the structure of the individual subunits is reliable, whereas the overall arrangement of the HTC remains uncertain[180, 181]. More critically, a closer inspection of the individual ipTM scores in the half-moon conformation reveals that while the averaged inter-subunit interactions are moderately confident, the predicted interactions between protein and DNA consistently score below 0.3. This indicates a clear failure in modeling the protein–DNA interface. Such limitations in current AI-based structure prediction highlight the challenges of accurately representing ligand interactions within large protein complexes[182]. This low confidence of the interaction can also be seen in the predicted aligned error (PAE) plot next to the predicted structure (see **Figure** 3.16). Therefore, it remains essential to validate these models against experimental data, such as cryoEM, to ensure structural reliability. In the literature a triangular shape of the terminase can be found but as mentioned previously, the diameter is not big enough so that DNA can bind[28]. Overall, AlphaFold3 predictions suggest that the HTC$_3$ complex adopts different conformations depending on DNA binding. Without

DNA, a triangular conformation is predicted, but it shows low confidence and lacks a sufficiently large channel for DNA interaction. In contrast, the presence of ssDNA stabilizes a half-moon-shaped structure with higher prediction confidence. Despite this, AlphaFold3 fails to reliably model the protein–DNA interface, highlighting current limitations in AI-based structure prediction. Validation through experimental approaches like cryoEM and nMS remains essential to confirm the physiological relevance of predicted models.
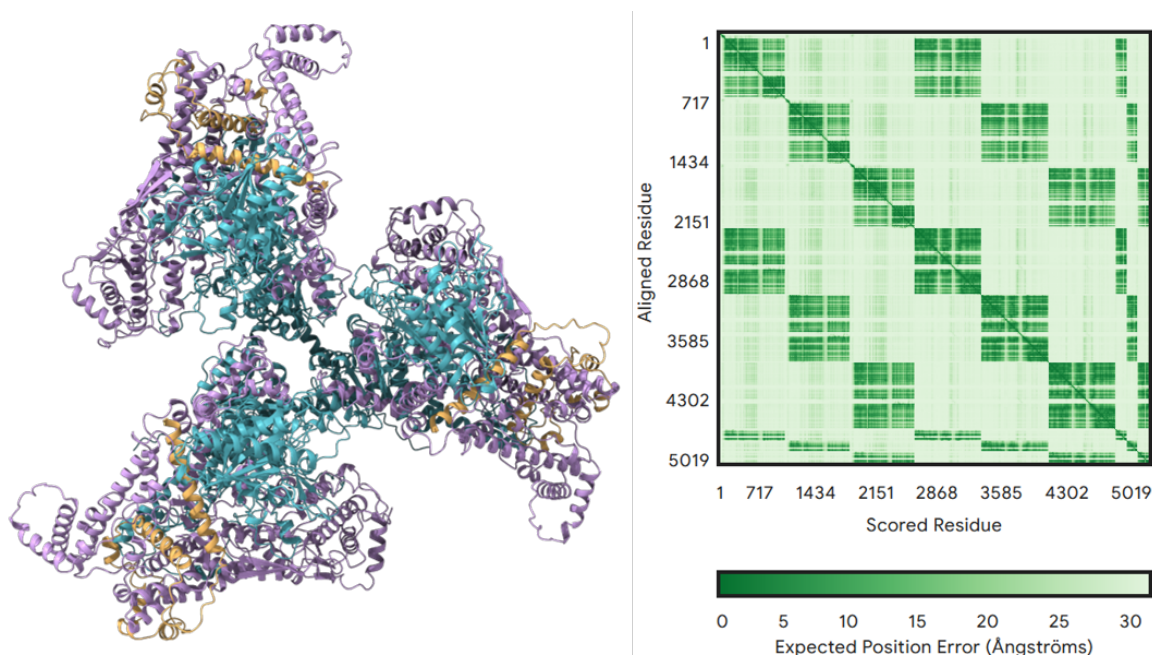
Figure 3.15: Predicted HTC$_3$ terminase complex structure without DNA using AlphaFold3. Three HTC subunits assemble into a triangular architecture with a central constriction (left). On the right, the corresponding PAE plot is depicted. It indicates spatial confidence between residues. Darker green regions denote high-confidence interactions (e.g., stable subunit interfaces). Lighter green regions reflect uncertainty, likely due to weaker inter-subunit contacts. Diagonal blocks with low PAE correspond to rigid intra-subunit domains, while off-diagonal low-PAE areas validate inter-subunit assembly.
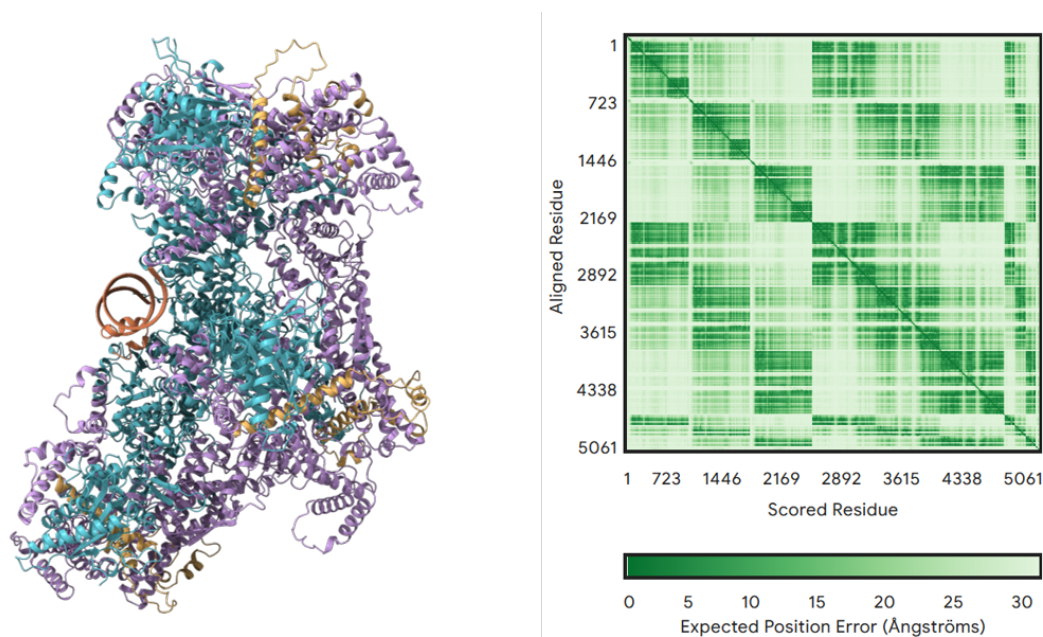
Figure 3.16: Predicted HTC$_3$ terminase complex structure with *pac* 1 DNA using AlphaFold3. Three HTC subunits assemble into a half-moon architecture with a large DNA binding interface that binds to the reverse strand containing the G4 (left). On the right, the corresponding PAE plot is depicted. It indicates spatial confidence between residues. Darker green regions denote high-confidence interactions (e.g., stable subunit interfaces). Lighter green regions reflect uncertainty, likely due to weaker inter-subunit contacts. Diagonal blocks with low PAE correspond to rigid intra-subunit domains, while off-diagonal low-PAE areas validate inter-subunit assembly.

3.3.4. The presence of DNase hinders HTC₃ formation
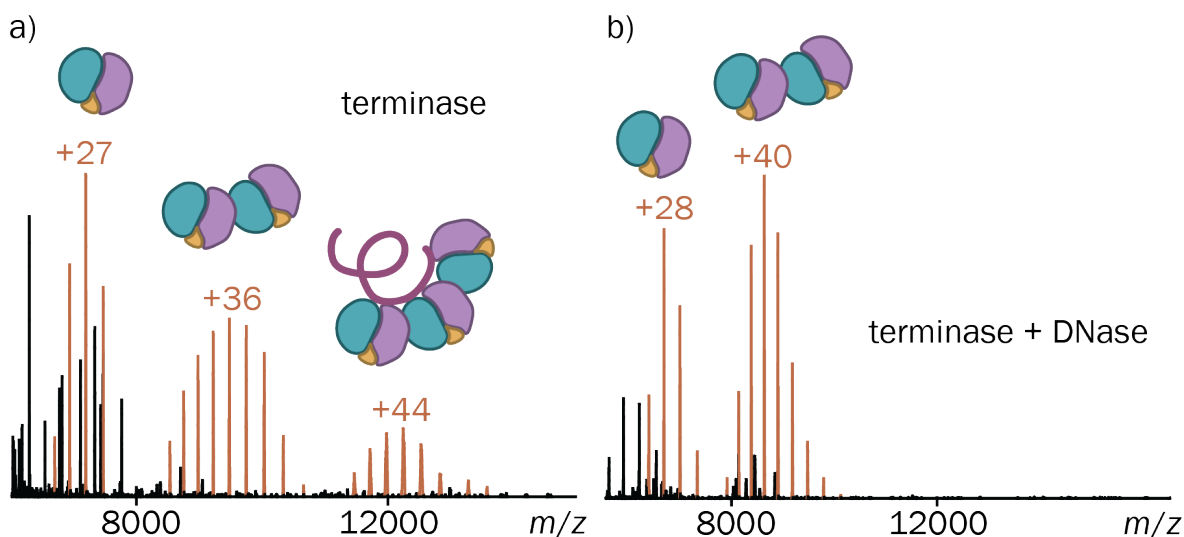
Figure 3.17: Spectra of 3 µM terminase complex in 320 mM ammonium acetate and 35 mM imidazole at pH 7.5. Charge state series were identified for HTC, HTC₂ and HTC₃ which is bound to a *pac* strand. **a)** Shows the sample without DNase and **b)** with 0.5 units of benzonase, incubated for 2 h at 4 °C. Data were obtained from one biological replicate, with each condition measured in three technical replicates.

Peaks of the HTC₃ were detectable even before *pac* DNA was added (see **Figure** 3.8). Its increased formation upon DNA addition suggests that DNA stabilizes the complex. However, this raises the question of how HTC₃ can form in the absence of *pac* DNA. To study the stability of the HTC₃ in the absence of DNA, new samples were prepared. These included terminase with *pac* sequences and a negative control without additional DNA. Directly after mixing the terminase with the *pac* DNA, a DNase (benzonase) was added (for the negative control the DNase was directly added to the terminase sample). In **Figure** 3.17 **a)** the terminase without additional *pac* DNA and without added DNase is shown. Three charge series can be identified and assigned to the three oligomeric states previously discussed (HTC just below $m/z = 8000$, HTC₂ just above $m/z = 8000$ and HTC₃ around $m/z = 12000$). In **b)** a sample containing the terminase and benzonase is shown, here only peaks from HTC

and HTC_2 are seen. Even though peaks from HTC seem to be comparable for both experiments, the peaks from HTC_2 have shifted seemingly to slightly lower m/z. Upon further inspection of the HTC_2 masses it was revealed that without adding benzonase the mass is 340960 ± 13 Da and upon adding benzonase a mass of 341010 ± 33 Da. Considering the measurement error, these values are essentially identical. Since the HTC peaks remain unchanged between spectra and no mass loss is observed for HTC_2, this indicates that no DNA is bound to either complex. The data further suggest that only HTC_3 is directly stabilized by DNA, as it disappears upon DNase treatment. This supports the hypothesis that HTC_3 formation is dependent on the binding of the DNA molecule to HTC_2. Notably, even in the absence of specifically added *pac* DNA, small peaks corresponding to HTC_3 were still detectable (see **Figure** 3.8), as previously mentioned. This finding implies again that the terminase complex interacts with cellular DNA fragments, reinforcing the idea that it does not exhibit strict sequence specificity. Similar behavior has been observed in other DNA-binding herpesvirus proteins, such as ICP8, as described in the literature[183, 184].

The MS data aligns well with the revolution model discussed earlier (**Chapter** 1.1.2) regarding the terminase packaging mechanism. During DNA packaging, the arginine finger penetrates the ATPase active site of the adjacent HTC subunit. Hydrolysis of the bound ATP then induces a conformational "flipping" motion of the arginine finger, pushing the DNA upward in a drilling-like motion. The data presented here strongly support a mechanism in which two and three HTC subunits interact directly at a given time. Considering the cryoEM data[29], it appears that the mechanism engages all six HTC subunits in the packaging process by sequentially binding the next available HTC. Notably, the HTC_6 complex appears to be a pseudo-complex or, at most, weakly bound and therefore not stable in the gas phase. In this context, mass photometry is a particularly suitable method to confirm the presence of a HTC_6

complex in the samples. Given its capacity to facilitate stoichiometric analysis of protein complexes in solution, it has to be performed in future experiments[185]. Based on the data presented here, along with previously published findings, particularly the cryoEM data[29, 51, 52], it can be hypothesized that all six HTC subunits are arranged around the DNA, with only up to three engaging in direct interactions at any given time. In **Figure** 3.18, the proposed mechanism is illustrated. The yellow sticks represent the bonds between the protein and DNA, while the deep purple stick indicates the penetrating arginine finger[29]. Initially, only two subunits bind to the DNA, with the first subunit inserting its arginine finger into the ATP binding site of the second subunit, thereby facilitating ATP hydrolysis. The energy released from this reaction induces a snapping motion, moving the DNA upwards. The disappearance of HTC_3 peaks upon DNase addition could therefore indicate that HTC_3 functions as a transient intermediate when DNA is present. This suggests that HTC_3 forms only at a specific stage during the packaging process and that dynamic assembly and disassembly may play a key role in the packaging mechanism. In a first step to confirm this hypothesis, an nMS experiment could be performed to track the relative intensities of the HTC_3 over time by adding *pac* DNA and ATP. Since no ATP was added during the here presented experiments, HTC_3 formation must occur before the arginine finger retracts from the active site, causing the packaging process to stall in a transitional state where three HTC subunits remain bound. However, based on the current data, it is unclear why the process halts specifically at three subunits and does not result in higher-order complex formation. It is most likely that HTC_6 is inherently unstable and likely disassembles during transport into the gas phase[186], whereas the interactions within HTC_3 are sufficiently strong to survive ionization when stabilized by DNA. Conversely, in rapid-freezing conditions, weak interactions within HTC_6 are preserved and therefore hexameric complex formations were observed [29]. These

findings underscore the importance of employing synergistic analytical methods, such as cryoEM and nMS, to unravel the intricate structural and biochemical mechanisms underlying complex molecular assemblies[187].

In summary, the addition of DNase led to the disappearance of the HTC_3 complex, indicating that its formation was stabilized by DNA binding. This suggests that DNA plays a direct structural role in assembling higher-order terminase complexes. The observed disappearance supports the hypothesis that HTC_3 represents a transient intermediate in the packaging mechanism, involving up to three HTC subunits interacting with DNA simultaneously. These findings align with the revolution model of DNA packaging and suggest a dynamic assembly process. MS and cryoEM together provide complementary insights, but further experiments are required to fully resolve the stoichiometry and stability of the HTC_6 complex.
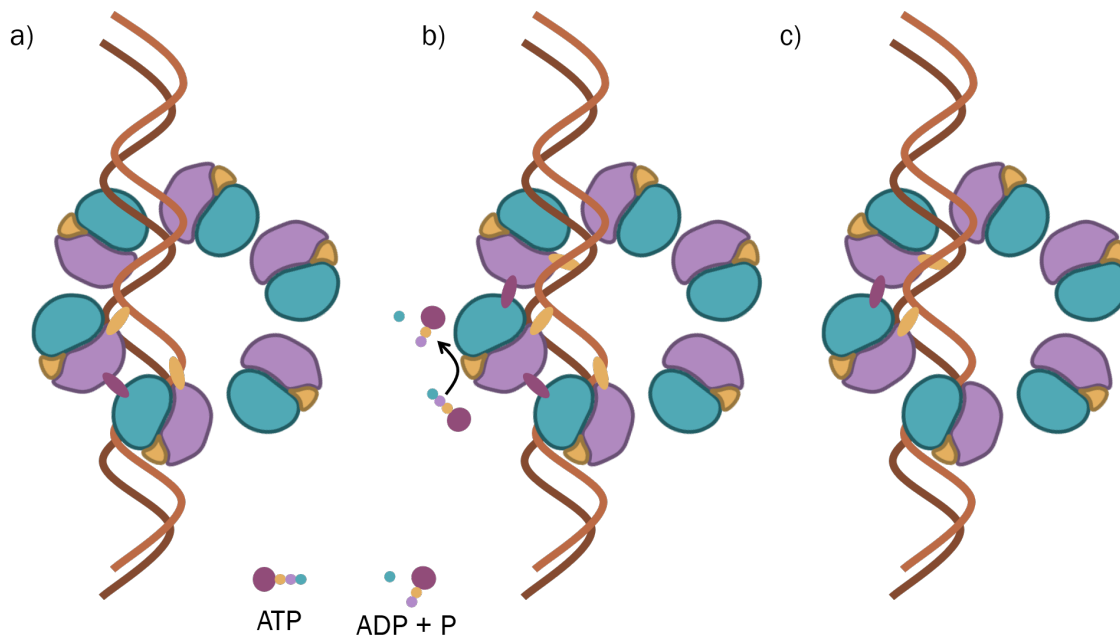
Figure 3.18: Hypothetical mechanism of the terminase packaging mechanism, including the HTC_3 transitional state. a) The first two HCTs bind the DNA. These two subunits form a HTC_2 while the arginine finger of the first HTC penetrates the second HTC. b) The transitional state is formed, where three HTCs (formation of the HTC_3 complex) bind the DNA while ATP is being hydrolyzed. c) Following ATP hydrolysis, retraction of the arginine finger in the first HTC subunit results in its dissociation from the complex, thereby reverting the assembly to the HTC_2 complex.

3.3.5. The terminase complex exhibits lower-than-expected mass due to pUL15 truncation and DNA fragmentation

The theoretical complex masses for different stoichiometries can be readily calculated from the protein sequence using tools such as ProtParam (Expasy, Swiss Institute of Bioinformatics)[188]. However, it is crucial to note that the calculated mass is only accurate if the actual sequence of the plasmid used for protein expression is considered, including tags and enzymatic cleavage sites. Additionally, these calculations do not account for the mass contribution of PTMs. The mass of PTMs can vary significantly, while phosphorylation typically adds a mass of 80 Da[189], glycosylation often results in mass shifts of a several hundred Dalton[190]. Given their inherent heterogeneity, these substances often appear in various glycan compositions, which, contingent upon the spectral resolution, manifest either as broad peaks or multiple charge series[191]. In particular, for larger biomolecules, the calculated masses often appear smaller than the experimentally determined masses. This discrepancy is not only due to the presence of PTMs but could also arise from incomplete desolvation and the formation of adducts during mass spectrometric analysis. This results in inaccurate mass measurements. To mitigate this issue, it is essential to optimize sample preparation and spray conditions.

Table 3.2: Theoretical and experimental masses of the terminase complexes. Experimental masses and standard deviation were obtained from technical triplicates. FWHM was multiplied by the charge of the respective peak and averaged over all peaks and replicates. Data were obtained from one biological replicate and three technical replicates.

Species	Th. M in Da	Exp. M in Da	ΔM in Da	FWHM in Da
HTC	183653	193775 ± 1	+9802 ± 1	191 ± 7
HTC$_2$	367307	340969 ± 1	-26339 ± 1	290 ± 18
HTC$_3$	550961	538600 ± 100	-12400 ± 100	1200 ± 110

The masses calculated in **Table** 3.2 are rather unexpected. While the mass of

the HTC complex appears to be significantly higher than expected (approximately 10 kDa), the masses of the higher-order complexes are surprisingly lower (around 10 to 26 kDa) as seen in **Table** 3.2. Since the measured mass of HTC is approximately 10 kDa higher than the theoretical value, one possible explanation is bound DNA, as pUL15 possesses a designated DNA binding site that, as previously discussed, does not appear to be highly sequence-specific. This is refuted by the DNase experiment (see chapter 3.3.4). Here, the mass of the HTC does not decrease after treatment with the enzyme and the corresponding peaks do not shift to lower m/z, indicating that DNA was not bound to the HTC. As mentioned previously, a mass increase often occurs due to PTMs. Since there is no research on the protein species and PTMs of the terminase subunits, the terminase sequence was fed into the prediction tool MusiteDeep[192–194]. The software predicted several PTMs sites, as shown in **Table** A3. Phosphorylation is essential for some herpesvirus proteins such as the viral proteases pUs2 and pUL13 [80] and is therefore a likely candidate for modification of the terminase. Phosphorylation on one amino acid residue increases the protein mass by 80 Da[195], which means 125 modifications would be necessary for a mass shift of 10 kDa, which seems highly unlikely. Another common modification in herpesvirus proteins found in the prediction is ubiquitylation[196, 197]. The 8.6 kDa protein ubiquitin is fused to a lysine residue and is important for protein stability and localization[197], it also leads to a mass shift much closer to that observed for HTC. In the case of modification of the protein under consideration, multiple peaks for a single charge in the charge series are often detected, given that not all proteins are modified. At lower spectral resolution, these peaks can overlap, resulting in broader peaks with a FWHM. In contrast, the mass spectrum of the HTC does not exhibit this behavior (see **Table** 3.2 and **Figure** 3.8). Another potential modification is glycosylation, yet the observation that the terminase is a highly glycosilated protein

would be unexpected due to the absence of neighboring peaks or fine structure of the charge states detected in the mass spectra [198]. To obtain a more profound comprehension of the modifications of the terminase complex, it is imperative to undertake an in-depth proteoform investigation. A commendable approach would be to employ a conventional bottom-up strategy within a liquid chromatography mass spectrometry (LC-MS) system, as previously outlined in the literature[79].

Interestingly, the higher-order complexes exhibit a decrease rather than an increase in mass, which is unusual, as it typically suggests that certain protein fragments are missing. However, as discussed in a previous chapter (**Chapter** 3.2.2), a truncated 59 kDa version of pUL15 is observed when expressed in insect cells. Notably, the mass difference between the theoretical and experimental HTC_2 is approximately 26 kDa, which is nearly identical to the difference between full-length pUL15 and the 59 kDa truncated fragment of pUL15. This suggests that the HTC_2 complex may form with both intact and truncated versions of pUL15, potentially explaining the absence of peaks corresponding to the 59 kDa fragment.

Here it is important to mention that the narrow peak widths of HTC (191 ± 7 Da) and HTC_2 (290 ± 18 Da), as indicated by their low FWHM values, suggest a relatively homogeneous composition. In contrast, the HTC_3 peaks are significantly broader and its observed mass difference is even smaller than that of HTC_2. The peak broadening could be attributed to bound DNA of various lengths, as described previously. Additionally, the mass difference for HTC_3 is approximately half that of HTC_2, suggesting that, on average, only one out of three subunits is truncated. Additionally, cellular DNA bound to the complex, leads to the high FWHM and increased mass.

The masses of the HTC_3 complex were then calculated when bound to the *pac* sequences. Although *pac* 2 is considerably longer than the *pac* 1 sequences, all *pac*

Table 3.3: Mass increase of the HTC$_3$ when bound to *pac* sequences compared to the unbound state. FWHM was multiplied by the charge and averaged over all peaks and replicates. Data were obtained from one biological replicate and three technical replicates.

Species	Exp. M in Da	M increase in Da	FWHM in Da
HTC$_3$ + *pac* 1	541700 ± 200	+3200 ± 200	900 ± 200
HTC$_3$ + *pac* 2	541800 ± 100	+3200 ± 100	1000 ± 200
HTC$_3$ + *pac* 1 mT	541800 ± 200	+3200 ± 200	1100 ± 200
HTC$_3$ + *pac* 1 mG	542200 ± 400	+3700 ± 400	1000 ± 300

variants increased the overall mass of the complex by approximately the same amount. The theoretical masses of the DNA (see **Table** A1) range between 7 and 10 kDa, approximately two to three times higher than the observed increase. As previously discussed, the high temperatures of the QE-UHMR source are believed to cause fragmentation of unprotected or unbound DNA ends, leading to their dissociation from the complex. This would explain why the observed mass increase is only around 3 kDa, regardless of the specific *pac* sequence. While there is no direct reference to unbound DNA fragmentation in nESI found in the literature, it is well established that increased internal energy can induce ion fragmentation[199]. Given that an in-source temperature of 250 °C was used, the unprotected regions of the DNA, not shielded by the large terminase complex, may have undergone fragmentation. The experimentally derived mass shifts suggest that only an approximately 11 b DNA strand remains bound to the terminase complex. This can be calculated from the masses of the individual nucleotides which are around 300 Da when in a DNA strand[200]. It would be expected to see the DNA fragments in the low m/z of the mass spectrum. However, these fragments were not detected, likely due to the instrument's high mass detection parameters. Consequently, numerous low-mass species may have been lost during detection. In a subsequent experiment, it will be necessary to scan the entire m/z range to locate these fragments. Despite the addition of *pac* DNA at a threefold molar excess relative to the terminase, which due to its proposed higher binding affinity

to the protein complex[36] would be expected to replace cellular bound DNA[26] and thereby affect the FWHM values, no significant changes of the FWHM are observed compared to the data in **Table** 3.2. This suggests that the amount of *pac* DNA in the samples may not have been sufficient to displace all cellular DNA fragments. In future experiments, the concentration of *pac* DNA should be increased to check if it has an effect on the FWHM of the peaks from the HTC$_3$-DNA complex.

In conclusion, the experimentally determined masses of the terminase complexes deviate from theoretical predictions, with HTC appearing heavier and higher-order species lighter than expected. The observed mass increase for HTC is unlikely to result from DNA binding, but may be attributed to PTMs. In contrast, the mass deficits of higher-order complexes are consistent with truncated forms of pUL15, as previously reported. Binding of *pac* DNA to HTC$_3$ resulted in modest mass increases, likely due to fragmentation of unprotected DNA ends during ionization. No significant change in FWHM was observed upon *pac* addition, suggesting incomplete displacement of cellular DNA. Further optimization of experimental conditions and a proteoform-level analysis will be needed to fully resolve the contribution of PTMs and DNA to the observed mass shifts.

3.4. Fast-track purification of terminase-associated proteins from various expression systems

Parts of the following chapter are submitted as a preprint:

Grün, A.F.R.; Said, F.A.; Schamoni-Kast, K.; Damjanovic, T.; Bosse, J.; Uetrecht, C., Fast tracking native mass spectrometry: Skipping over buffer exchange, bioRxiv 2025, https://doi.org/10.1101/2025.02.22.639503[157]

The proteins pUL32 and pORF68 are homologous and associated with the viral packaging process of HSV-1 and KSHV, respectively[201, 202]. Both proteins have recently been analyzed using electron microscopy (EM)[203]. The study provides strong evidence that both homologues assemble into homo-pentameric structures[203]. While pORF68 expression was exclusively carried out in mammalian cells, the expression of pUL32 was evaluated in both insect cells and *E. coli*. Both proteins were purified with the fast-track protocol to demonstrate that proteins from different expression systems can be effectively purified by the fast-track protocol and result in high-quality spectra. The goal was to determine the stoichiometry of both protein complexes, compare them to the cryoEM data, and assess potential binding between pUL32 and the terminase. To investigate potential binding, pUL32 was mixed with the terminase after purification.

Since these proteins originate from human viruses, their expression in mammalian cells, particularly in human cells, is expected to produce protein variants that closely resemble their natural counterparts. As shown in **Figure** 3.19, the two homologues exhibit distinct spectral profiles. In **Figure** 3.19 **a**, the spectrum of pORF68 displays a charge-state-resolved peaks between $m/z = 10000 - 11000$, which can be assigned to

Table 3.4: Expression of pUL32 in *E. coli* and insect cells with different parameters. The term "high yields" is employed to denote protein concentrations greater than 1 mg/mL, whereas "low yields" is used to indicate concentrations less than 0.5 mg/mL. Concentrations below 0.2 mg/mL are designated as "very low yields". For *E. coli* two different constructs were employed. The first had a N-terminal Strep-tag and the second had a N-terminal His- and SUMO-tag. For insect cells only one parameter set was tested.

Expression system	Altered parameter	Protein yield	Comments
E. coli, BL21 DE3	1 L suspension volume 0.4 mM IPTG	none	inclusion bodies
E. coli, BL21 DE3	1 L suspension volume SUMO-tag, 0.4 mM IPTG	none	inclusion bodies
insect cells, ExpiSf	25 mL suspension volume 3 dpi harvest	high	baculovirus is unstable

a 549 kDa decameric pORF68 complex which fits to the expected mass of 550 kDa when taking the high FWHM into context (see **Table** 3.5). This observation is plausible, as electron microscopy (EM) studies have identified decameric structures in addition to the expected pentameric form, likely representing two stacked pentameric rings[203]. Since the sample was only measured once in a MS^1 experiment, this cannot be confirmed by the present data. In future experiments, a MS^2 experiment has to be utilized to confirm the presence of two stacked pentameric rings. In addition, a denaturing solvent (e.g., acetonitrile) could be mixed in to the sample at low concentrations. By slightly unfolding the protein complex, a more profound understanding of the complex composition could be gained. Furthermore, it is noteworthy that, similar to the terminase complex, the associated proteins pORF68 and pUL32 appear as donut-shaped structures with a highly positively charged inner channel and a zinc-finger in the literature[203, 204]. More recently, a direct interaction between pUL52 — the HCMV homologue of pUL32 and pORF68 — and the terminase complex has been experimentally demonstrated[205], suggesting a close functional association between these complexes during DNA packaging. Given that both pUL52 and the terminase are likely to bind the same DNA strand, it can be inferred that their DNA-binding properties must be similar. This may also account for the broad FWHM

observed in peaks associated with the decameric pORF68 complex. Comparable to HTC$_3$, pORF68 may also bind to cellular DNA fragments, a hypothesis that should be further validated by repeating the DNase experiment with pORF68.

In contrast, pUL32 forms a trimer as seen in **Figure** 3.19 **b**, instead of a pentamer which would have been expected from the cryoEM data[203]. As illustrated in the figure, the peaks are situated between $m/z = 7000 - 7500$ (purple), marginally above the peaks of the terminase HTC (orange). The terminase peaks are analogous to the peaks in **Figure** 3.8 and demonstrate no mass shift, which would be anticipated if pUL32 were to bind to it. This is most likely due to the trimer formation of pUL32, which might be an artifact from expressing the protein in insect cells instead of mammalian cells[203]. Furthermore, the masses of the trimer are slightly lower than anticipated, as observed in, **Figure** 3.19 which could be attributed to fragmentation or truncation, akin to the case observed in pUL15. It is noteworthy that no truncated versions of pUL32 have been documented in the literature up to this point. Due to the unexpected stoichiometry detected for pUL32, it is highly likely that protein interaction between the terminase complex and pUL32 is hindered due to the wrong folding of pUL32. This would explain the lack of pUL32-terminase peaks, which would be expected based on recent data found in the literature[205]. In future experiments, it will be essential to express pUL32 in mammalian cells, analogous to previous approaches with pORF68. This would allow for the assessment of whether the observed trimer formation is a result of misfolding. Subsequently, a follow-up experiment could be performed by reconstituting a mixture of pUL32, the terminase complex and the *pac* DNA. Properly folded pUL32 could then be analyzed for its ability to interact with the terminase and with DNA, thereby providing further insight into its functional role during viral DNA packaging.

For both proteins, charge-state-resolved spectra were obtained in the presence

Table 3.5: The protein complexes of pORF68 (KSHV) and pUL32 (HSV-1) were detected on the QExactive UHMR Orbitrap. FWHM and measured masses were averaged and standard error is given. Data were obtained from one biological replicate each. Technical triplicates were recorded for pUL32. For pORF68, only one spectrum was recorded.

Protein	Th. Mass (in Da)	Measured Mass (in Da)	FWHM (in Da)
pUL32	202340	201440 ± 14	130 ± 25
pORF68	550965	549180 ± 34	1700 ± 519

of desthiobiotin, enabling the determination of the stoichiometries of the respective complexes. These results demonstrate that the fast-track protocol can be successfully applied to proteins expressed in mammalian cells, in addition to those produced in insect cells, to generate high-quality nMS data. This strategy could significantly broaden the applicability of nMS to protein complexes that have previously posed challenges in terms of in vitro expression and purification. This includes various terminase homologues from other herpesviruses, such as HCMV and KSHV. Although the pUL32 homologue is known to play a key role in viral replication and is functionally associated with the terminase complex[205], no direct interaction was detected in the present data. This lack of interaction is most likely due to pUL32 misfolding under the experimental conditions used.
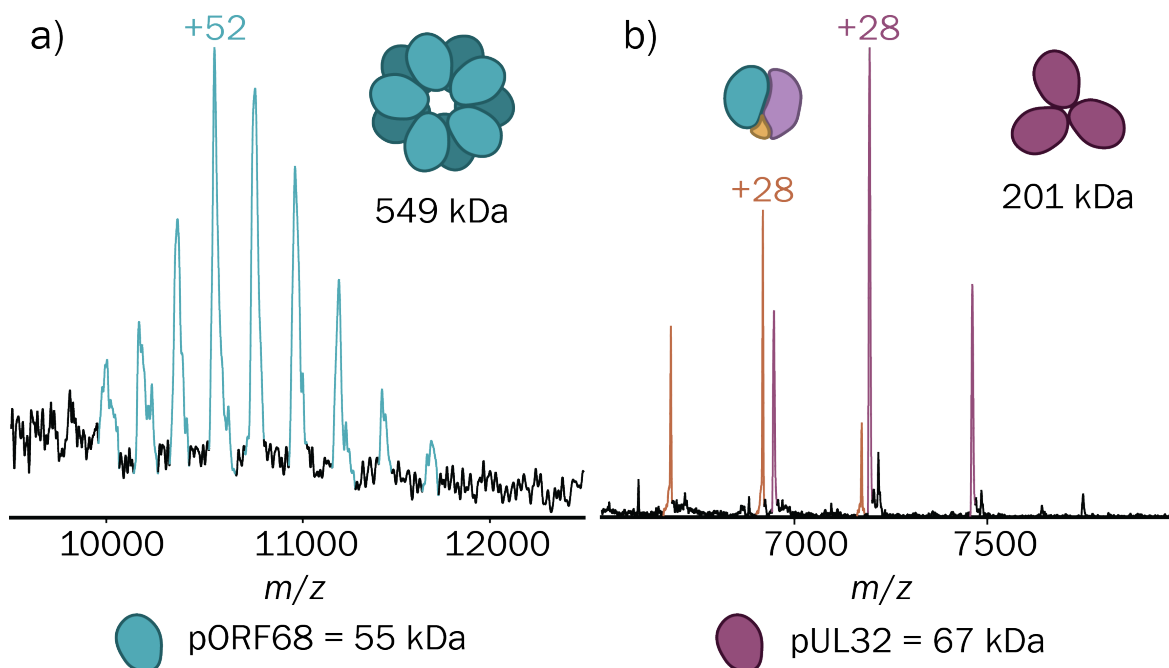
Figure 3.19: Spectra of terminase associated proteins from different herpes viruses at a concentration of 5 µM on the QE-UHMR. **a)** pORF68 (KSHV) expressed in mammalian cells and purified via Twin-Strep with 2.5 mM desthiobiotin at 250 eV. **b)** Strep-pUL32 (HSV-1) expressed in insect cells mixed with 1 µM terminase hetero trimer, and 2.5 mM desthiobiotin at 250 eV.

3.5. Fast-track nMS enables dimerization bias analysis of fluorescent proteins

The analysis of the FP spectra was performed with the assistance of a custom written Python script to facilitate nMS data analysis, developed by Janine-Denise Kopicki, which helped calculate the AUC for the peaks of the monomer and dimer charge state series.

The fast-track purification protocol is a highly versatile approach that enables rapid screening of protein oligomerization states, facilitating efficient analysis of complex formation. In addition to their biological significance, protein stoichiometries play a crucial role in protein function[206]. This is particularly relevant for FPs used in

fluorescence live-cell imaging[207, 208], where accurate characterization of protein complexes is essential. To track a protein inside a cell, an FP is fused to said protein. By shining a light on the cell with an FP-specific wavelength, these proteins can be made visible. Over time, the movement, aggregation, or complex formation of the protein can be detected. A major drawback of this technique is interruption of these interactions by the oligomerization of the FP tag[209, 210]. To tackle this issue, specifically monomeric proteins like mEGFP were designed for fluorescence microcopy. Even though dimerization of these "monomeric" FPs has been acknowledged, it is often overlooked in discussions regarding the validity of live-cell imaging data[148]. This is largely due to a lack of methods for large-scale analysis of such complex formations. Additionally, determining the local concentration of tagged viral proteins in an individual cell remains challenging. The complexation of FPs is highly dependent on local concentration[153, 154] and can therefore interfere with fluorescence microscopy experiments. The formation of aggregates by the FPs results in the generation of strong signals in the image, potentially giving rise to the erroneous impression of an accumulation of the attached protein at a particular location in the image[211]. These challenges highlight the need to assess the monomeric state of FPs. To address this, the OSER method[149] was developed, allowing the determination of dimer percentages in various FPs. However, due to the complexity and lengthy execution time of the experiment, many newly developed FPs remain uncharacterized. To overcome this limitation, a fast and reliable method for determining dimer content based on the fast-track protocol was developed.

3.5.1. The dimerization of FPs can be characterized via nMS

Protein concentrations can reach relatively high levels at specific subcellular locations, promoting interactions between FPs and often resulting in dimer formation. To

determine the concentration threshold at which FP dimerization occurs, various FP variants were analyzed at different concentrations. These proteins were first expressed in *E. coli* and subsequently purified using the fast-track protocol. To verify successful expression and purification, samples were analyzed by SDS-PAGE. The gels shown in **Figure** 3.20 display clear bands around the 35 kDa marker for most FPs, as expected. However, expression of EGFP, mEGFP and mTurquoise2 yielded significantly lower amounts of protein, indicating suboptimal expression conditions. Notably, a prominent band is visible in the mScarlet3 lanes at approximately 60 kDa, corresponding to the heat shock protein GroEL[212]. Additionally, faint bands below the expected FP bands suggest partial protein degradation. As all constructs were expressed using the same protocol, it is likely that the purification process was not fully optimized. Protein degradation during recombinant expression can often be mitigated by lowering the IPTG concentration and reducing the expression temperature. Instead of rapid expression at elevated temperatures, prolonged expression at lower temperatures allows for improved protein folding and stability[213, 214].

All proteins were measured on the MS in four different concentrations, ranging from 1 to 20 µM. For each concentration, the monomer-to-dimer ratio was calculated based on the average area under the curve (AUC) of all peaks corresponding to each species. Despite most FPs being classified as monomeric, dimer formation was observed in all variants at higher concentrations. It is important to note that at elevated protein concentrations, artificial oligomerization can occur in the gas phase during nESI[215, 216]. This effect is typically detectable at concentrations ≥ 10 µM[217].

For all FPs, no dimer formation was detected at an initial protein concentration of 1 µM, as shown in **Figure** 3.21, most likely due to low ion yields of the dimer species resulting in signal intensities below the detection threshold. As expected, the proportion of detected dimers increases with protein concentration for all FPs.
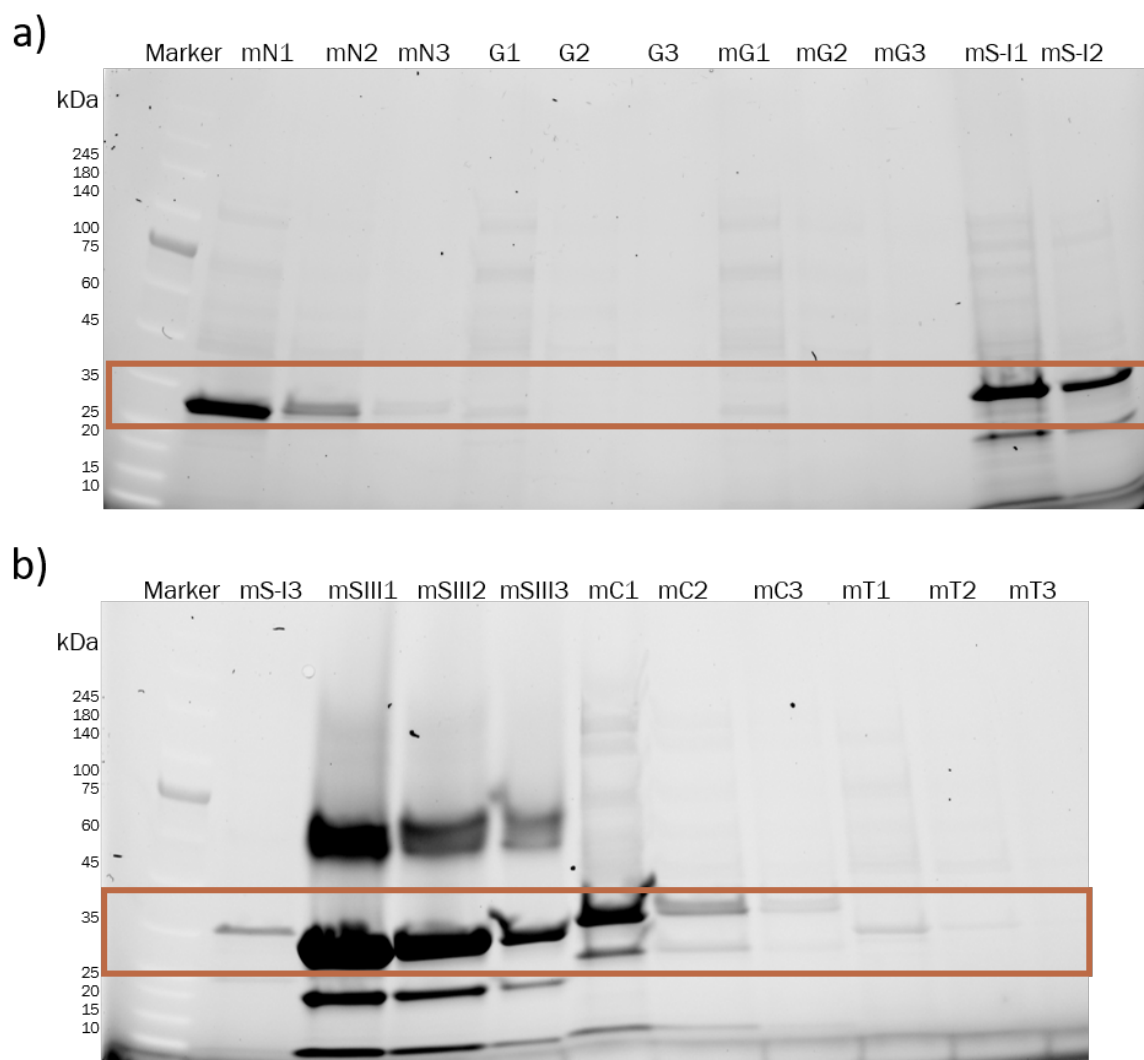
Figure 3.20: SDS-PAGE analysis of purified fluorescent proteins (FPs) expressed in E. coli. The first three Ni-NTA gravity column eluates for each FP were loaded (15 µL per lane). All target proteins migrate near the expected molecular weight of 30 to 35 kDa. The orange border marks the bands. Prominent bands indicate successful overexpression and purification for most FPs, though reduced yields were observed for EGFP (G), mEGFP (mG) and mTurquoise2 (mT). A prominent contaminating band (approximately 57 kDa) in the mScarlet3 (mSIII) lanes corresponds to the E. coli heat shock protein GroEL, likely resulting from overexpression stress. Lower molecular weight bands across samples suggest partial protein degradation. Sample abbreviations: mN = mNeongreen; G = EGFP; mG = mEGFP; mS-I = mScarlet-I; mSIII = mScarlet3; mC = mCherry; mT = mTurquoise2.
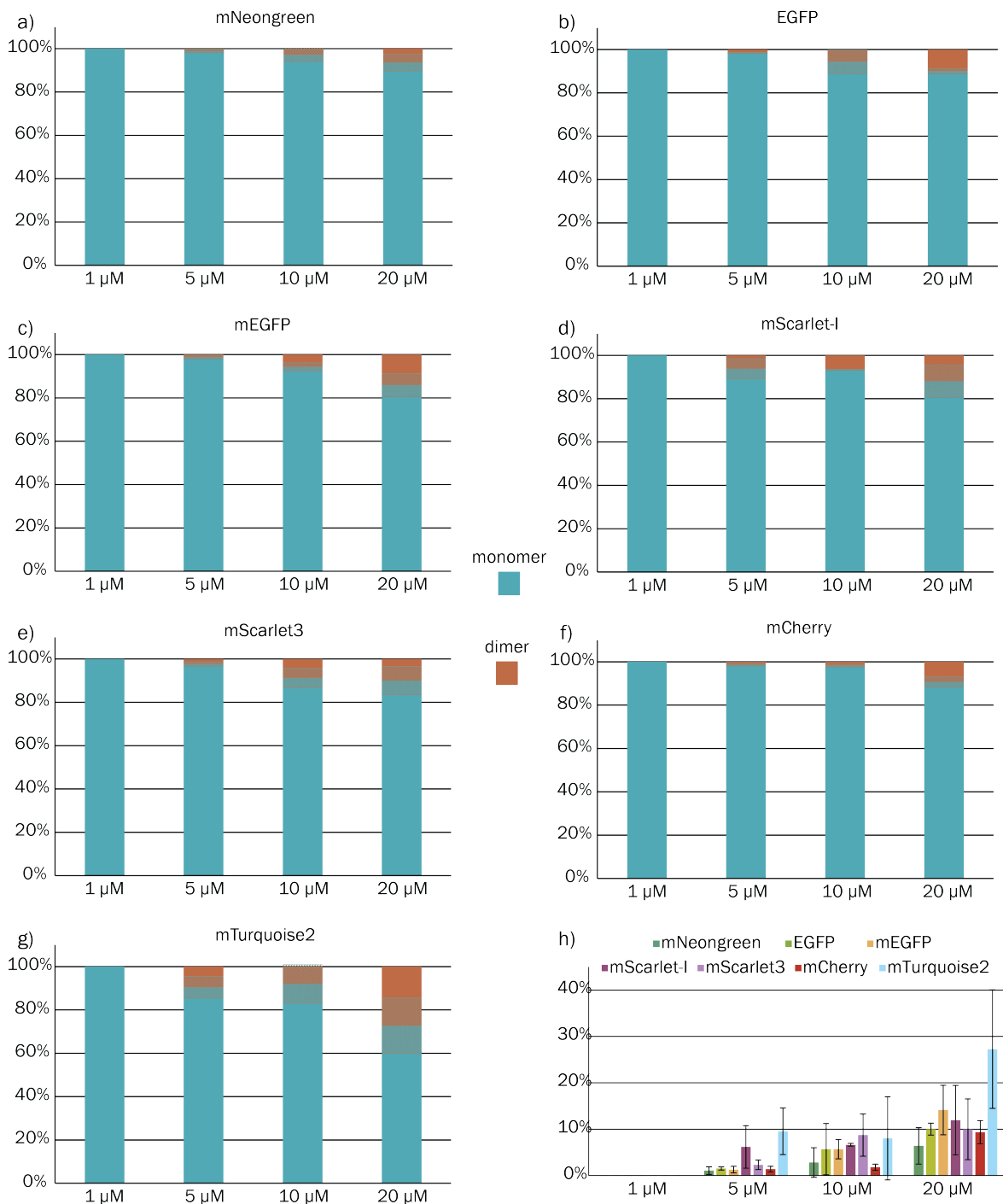
Figure 3.21: Bar graphs in panels **a)**–**g)** illustrate the dimer-to-monomer ratios of different FPs at varying initial protein concentrations. Blue bars represent monomeric species, orange bars represent dimers and the checkered areas indicate the associated error. Ratios were calculated based on the average AUC values of all monomer and dimer peaks. Panel **h)** summarizes the dimer fractions across all FPs and concentrations for comparative analysis. Data were obtained from one biological replicate, with each condition measured in triplicate (three technical replicates).

As mentioned, at higher concentrations (≥ 10) proteins tend to oligomerize in the gas phase. This would explain the relatively high amount of dimer detected for all FPs at 20 µM. Most FPs exhibit similar behavior, displaying comparable dimer-to-monomer ratios. However, two proteins deviate notably from this trend. The first is mCherry (see **Figure** 3.21 **f**), which exhibits a significantly lower dimer percentage than the other FPs up to 20 µM. For instance, at 10 µM, only 2 % dimer was detected. Additionally, mCherry consistently shows the lowest error, indicating good spray behavior and high reproducibility in the measurements. Conversely, mTurquoise2 behaves in the opposite manner. Not only does it exhibit the largest error margin, but it also shows the highest average dimer percentage (most notably at 5 µM and 20 µM). However, due to the high variability, this result must be interpreted with caution. When accounting for the error margin, the bar in **Figure** 3.21 **h** for concentrations 10 µM and 20 µM is comparable to those of most other FPs. The dimer detected at 5 µM, on the other hand, still is significantly higher for mTurquoise2 compared to most other FPs.

As previously mentioned in **Chapter** 1.4, multiple publications have determined the OSER values for most FPs discussed in this thesis. These values are summarized in **Table** 3.6 and can be directly compared to the data derived from nMS.

Table 3.6: Comparing the calculated dimerization percentages for all measured protein concentrations of the FP to OSER data derived from the literature of different studies.

FP	OSER	Calculated monomer %			
		1 µM	5 µM	10 µM	20 µM
mNeongreen	90.4 ± 2.1[218, 219]	100 ± 0	99 ± 1	97 ± 3	94 ± 4
EGFP	76.5 ± 6.9[149, 218–220]	100 ± 0	98 ± 0	94 ± 6	90 ± 1
mEGFP	96.8 ± 0.5–98.1 ± 1.6[218, 219, 221, 222]	100 ± 0	99 ± 1	94 ± 2	86 ± 5
mScarlet-I	76.0[223]	100 ± 0	94 ± 5	93 ± 0	88 ± 7
mScarlet3	N/A	100 ± 0	98 ± 1	91 ± 5	90 ± 7
mCherry	69.0–95.0 ± 0.8[219, 223–225]	100 ± 0	99 ± 1	98 ± 1	91 ± 3
mTurquoise2	88.0–93.8 ± 1.0[219, 226]	100 ± 0	90 ± 5	92 ± 9	73 ± 13

Some publications did not report error margins for their determined OSER percentages, suggesting a lack of replicates in their measurements. Despite this limitation, there remains a notable discrepancy in the reported OSER values for mCherry and mTurquoise2. Before directly comparing these data sets, it is important to acknowledge that, by design, the concentration of FPs in an OSER experiment depends on cellular factors and the expression of CytERM, which cannot be easily determined. Given that dimerization is highly concentration-dependent, the extent of dimer formation in an actual live-cell imaging experiment can vary significantly. The OSER approach provides insight into the dimerization behavior of FPs in a cellular system that is as close to physiological conditions as possible, while still allowing for comparative analysis between different FPs. In contrast, the nMS approach is primarily used for general screening to assess the dimerization tendency of FPs. When combined with expected intracellular protein concentrations, this method can help predict the likelihood of dimer formation in a given experimental setup. One particular aspect that warrants attention is the comparison between EGFP and mEGFP. These proteins are nearly identical, with mEGFP derived from EGFP through the A206K mutation. This mutation was analyzed using analytical ultracentrifugation and deemed sufficient to render mEGFP monomeric, the OSER experiments suggest a similar result. However, when analyzed via nMS, the difference in monomer content between EGFP and mEGFP appears significantly smaller than expected. This raises whether mEGFP is truly as monomeric as assumed, or if the A206K mutation does not have as strong an effect on dimerization as previously thought. Another possibility is that EGFP dimers dissociate in the gas phase, leading to low detected dimer intensities. Given that dimer content increases at higher concentrations, the most likely explanation is that different concentrations were used in prior literature. In the previously referenced study, analytical ultracentrifugation was performed at protein

concentrations ranging from 50 to 700 μM[227]. It is also important to note that the expression of both EGFP and mEGFP was suboptimal (see **Figure** 3.20). Protein yields were low and the bands corresponding to the FPs at approximately 35 kDa were only marginally more intense than the surrounding contamination bands. The mass spectra obtained for both EGFP and mEGFP also exhibited significant contamination, which hindered the detection of dimer species, particularly in the case of EGFP (see **Figure** A5). As a result of these challenges, including poor spectral resolution and overlapping signals, the calculated monomer-to-dimer ratios for EGFP may be prone to error and should be validated in future experiments.

Taken together, FP dimerization was assessed across varying concentrations using nMS. While most FPs behaved as monomers at low concentrations, increasing dimerization was observed at higher concentrations. mCherry showed the lowest dimer tendency and highest reproducibility, while mTurquoise2 exhibited the strongest dimerization bias but with high variability. Comparisons with OSER data generally supported the nMS findings, though discrepancies, particularly for mEGFP and EGFP, highlight the influence of concentration, gas-phase effects and protein purity on dimer detection.

3.5.2. K_d determination of FPs reveals differences in their dimerization bias.

The dissociation constant (K_d) for each protein for all four concentrations was calculated from AUCs and presented in **Table** 3.7 and **Figure** 3.22. This is possible with nMS due to the law of mass action[228, 229].

In **Table** 3.7, the K_d values derived from nMS analysis for all FPs discussed in this chapter are summarized. Most values fall within the triple-digit micromolar range, with the notable exception of mTurquoise2, which exhibited the lowest K_d at

Table 3.7: Averaged K_d values from protein concentrations from 5 to 20 µM. The error was calculated based on the Gaussian error propagation. Data were obtained from one biological replicate, with each condition measured in three technical replicates.

FP	K_d in μM
mNeongreen	359 ± 126
EGFP	218 ± 68
mEGFP	321 ± 249
mScarlet-I	125 ± 40
mScarlet3	122 ± 26
mCherry	306 ± 52
mTurquoise2	56 ± 24

$56 \pm 24\ \mu M$, suggesting a strong tendency toward dimerization. Generally, the lower the K_d, the higher the proportion of bound dimers[230]. In contrast, the highest K_d values were observed for mNeongreen, mEGFP and mCherry, aligning well with the OSER assay results presented in **Table** 3.6.

At an initial protein concentration of 1 µM, all FPs yielded a calculated K_d of 0 µM, which is likely an artifact caused by the noise suppression algorithm used in QE-UHMR. This algorithm tends to eliminate peaks corresponding to low-abundance species, resulting in undetectable dimer signals at low concentrations. Theoretically, K_ds should remain independent of overall protein concentration, as it reflects the equilibrium between bound and unbound species (see **Chapter** 5.6.2). However, as shown in **Figure** 3.22, minor variations in calculated K_d values were observed, likely due to protein oligomerization in the gas phase[215–217]. This effect was particularly evident in the decrease in K_d at 20 µM for all FPs (see **Figure** 3.22). Another important consideration is the variability in the calculated error margins, which were especially high for mNeongreen and mEGFP, as seen in **Figure** 3.22. For mEGFP at 5 µM, measurements showed substantial variance, potentially due to sample impurities or the noise suppression algorithm. At this concentration, the dimer peaks were of low intensity and likely fell below the detection threshold in some replicates.

Gaussian error propagation amplified these variations, resulting in large uncertainties in the calculated K_d. At 10 µM, mEGFP, mScarlet-I and mScarlet3 showed good reproducibility, with mScarlet-I exhibiting the highest K_d among the three.

Overall, all calculated K_d values fall within the range typical for weak protein–protein interactions[231], supporting the predominantly monomeric nature of these FPs. The literature reports a K_d of 0.11 mM for EGFP[227], which is consistent with the value obtained at 10 µM in this study. However, no comparable data were found for the other FPs.

In summary, K_d values for various fluorescent proteins were determined using native mass spectrometry, revealing significant differences in dimerization propensities. Most FPs exhibited weak interactions in the triple-digit micromolar range, with mTurquoise2 showing the strongest dimerization tendency while mCherry remained mostly monomeric. While noise suppression and spectral artifacts affected low-concentration measurements, especially for mEGFP and mNeongreen, overall trends aligned well with previous OSER assay data.
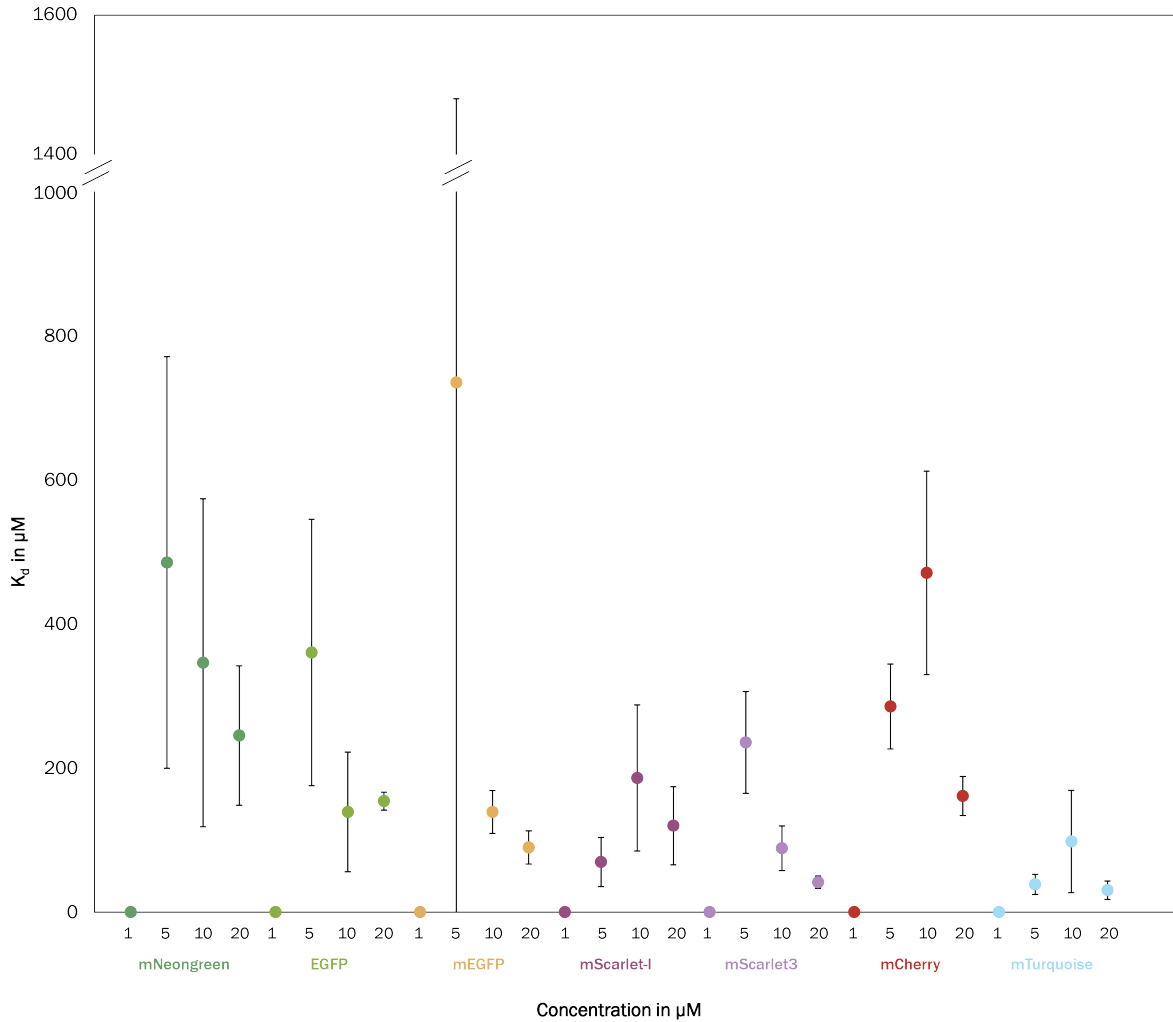
Figure 3.22: Dissociation constants (K_D) of FPs determined by native mass spectrometry across four different protein concentrations (1 μM, 5 μM, 10 μM, 20 μM). Each dot represents an individual replicate measurement and error bars indicate the standard deviation between replicates. Colors correspond to different FP variants. The K_D values reflect the monomer–dimer equilibrium for each FP, with lower values indicating stronger dimerization propensity. Note that exceptionally large error bars in some cases arise, especially for mEGFP (5 μM), where variation between measurements is amplified during Gaussian error propagation. Data were obtained from one biological replicate, with each condition measured in three technical replicates.

4

Conclusion and outlook

Understanding the composition and structure of viral replication proteins is crucial for antiviral drug development[8, 42, 43]. This thesis characterized the stoichiometry and DNA binding of the HSV-1 terminase complex using nMS, supporting a proposed revolution model of the packaging mechanism[29, 52]. It also presented the first nMS-based analysis of the HSV-1 terminase, offering insights into the dynamics of ssDNA-dependent HTC_3 formation and disassembly. The terminase was found to bind ssDNA, not dsDNA and showed no specific interaction with *pac* 1 (or its mutants) or *pac* 2. As the G4 in *pac* 1 forms only in ssDNA, it likely serves as a packaging termination signal rather than a binding site for the terminase.

As discussed in **Chapter** 3.2.2, oligomeric state assignments were based solely on intact complexes, as MS^2 could not be performed on the QE-UHMR. Future experiments must include MS^2 to validate these findings. Since PTMs have not yet been analyzed for the terminase, a bottom-up LC-MS approach would allow exact subunit identification and characterization of PTMs[79]. This approach requires infecting mammalian cells with HSV-1, followed by direct or gel-based digestion[232]. Lower-than-expected masses suggest incorporation of a truncated pUL15 variant. The truncated pUL15 subunit, expressed from two exons[166], may result from proteolysis

or natural expression, as similar truncations were observed in wild-type HSV-1-infected cells[165]. Literature indicates that the 30 kDa C-terminal fragment does not inhibit packaging[167]. The 60 kDa N-terminal fragment likely forms complexes, as suggested by the data presented in this thesis. Immunoprecipitation, western blot and follow-up bottom-up MS or nMS could clarify its role.

Future experiments should explore the packaging mechanism further (**Chapter** 3.3.4). Time-resolved nMS using ATP could monitor transitions between HTC_2, HTC_3 and intermediate states (**Figure** 3.18). Mass photometry, which preserves solution-phase interactions, may allow detection of higher-order assemblies like HTC_6. Structural methods such as HDX and cryoEM can complement this by elucidating the DNA-bound complex architecture. Investigating terminase interaction with correctly folded pUL32 is also essential, as current nMS data show no complex formation (see **Chapter** 3.4), even though pUL32 may be important for portal binding or DNA cleavage[203]. Additionally, Baculovirus constructs for the KSHV and HCMV terminase complexes have been generated. Although purification was previously unsuccessful, the fast-track protocol developed may now facilitate expression and purification. Given the high conservation of terminase architecture across herpesviruses[27], trimeric structures similar to HSV-1 are expected.

The fast-track protocol (see **Chapter** 3.2) for nMS outperformed conventional methods, yielding higher protein concentrations compared to conventional methods (up to 0.8 mg/mL vs. 0.2 mg/mL) and maintaining spectral quality despite the presence of imidazole, biotin and desthiobiotin. Combining Strep-tag purification with in-source CID on the QE-UHMR enabled robust characterization and left headroom for MS^2 at higher CVs.

Furthermore, a method for assessing FP dimerization at varying concentrations was established (see **Chapter** 3.5). All tested FPs showed significant concentration-dependent dimerization and their K_ds were calculated. To refine this method, defining physiologically relevant concentration ranges is necessary, as intracellular protein levels vary. This would allow experiments to better reflect actual dimerization biases, aiding in optimal FP selection and reducing tag-induced artifacts in microscopy.

In conclusion, this thesis advances our understanding of the HSV-1 terminase's composition, DNA-binding behavior and structural dynamics, while introducing an effective native MS sample preparation protocol. These contributions not only support future studies of viral DNA packaging but also provide tools and perspectives for investigating herpesvirus assembly more broadly.

<div style="text-align: right">

5

</div>

<div style="text-align: right">

Methods

</div>

5.1. HSV-1 terminase preparation for native mass spectrometry

5.1.1. Cloning of HSV-1 UL15, UL28 and UL33

The genomes of UL15/28/33 were ordered in pfastBac1 vectors with Amp resistance from GeneArt (Thermo Fisher, USA). All plasmids were transformed into XL10-Gold competent cells and afterward purified with the NucleoSpin® Plasmid (NoLid, Macherey-Nagel) Miniprep Kit. UL28 and UL33 were to be cloned into a pfastBacDual vector, for that, the inserts were first amplified via polymerase chain reaction (PCR), for each reaction 33 µL of Milli-Q water were mixed on ice with 10 µL of 5x HF buffer (F518L, Thermo Fisher, Germany), 1 µL of 20 ng/µL plasmid DNA, 5 µL of 10 µM mixture of forward and reverse primer, 1 µL of dNTP mix (R0191, Thermo Fisher, Germany) and 1 µL of the Phusion™-HF polymerase (F530S, Thermo Fisher, Germany). The PCR program was as followed:

The reactions were then applied to a 1 % agarose gel and relevant bands were cut out and purified via the NucleoSpin Gel and PCR Clean-up Kit (740609.50, Macherey-Nagel, Germany). The inserts were then digested with BamHI (R3104S, New England

Table 5.1: PCR parameter for UL28 and UL33 insert amplification

T [°C]	t [s]	cycles
98	30	1
98	10	30
55	10	30
72	90	30
72	600	1
10	∞	

Biolabs, USA) and HindIII (R0136S, New England Biolabs, USA) at 39 °C for 1 h. They were then again run on a gel and purified. Ligation was performed in two steps, at first UL33 was cloned into the pfastBacDual vector. Here, 12 µL of the MightyMix (6023, Takara Bio, USA) were added to a mixture of 2 µL pfastBacDual vector and 10 µL UL33 insert. The mixture was incubated for 30 min at 16 °C. Afterward, the plasmid was again transformed in XL10-Gold cells and verified by Sanger sequencing. Analogous to UL33, UL28 was then cloned into the pfastBacDual plasmid, which from now on will be referred to as Dual.

5.1.2. Baculovirus generation of HSV-1 UL15 and Dual

Now the baculoviruses had to be generated for UL15 and Dual. Here, 47 ng of the Dual and 49 ng of the UL15 plasmid were pipetted into 50 µL of competent DH10Bac cells. The transformation was stopped after 4 h at 37 °C and the cells were spread onto agar plates containing 120 µL of 20 mg/mL X-Gal in DMF and 40 µL of 100 mM IPTG in water as well as the antibiotics' tetracycline, kanamycin and gentamicin. These were then incubated at 37 °C. 48 h later, the plates were checked for white colonies, which were spread onto new plates, repeating the process. positive (white) colonies were transferred to fresh medium and grown overnight. These were then purified via Midipreps (K210005, Plasmid DNA Midiprep Kit, Thermo Fisher, Germany) and stored at 4 °C in aliquots in appropriate amounts.

25 mL of the insect cells ExpiSf™ (A3767806, ExpiSf™ Expression Kit, Gibco, Thermo Fisher, Germany) were diluted to a cell density of $2.5 \cdot 10^6$ cells/mL in 25 mL and put on a shaker with a speed of 125 rpm at 27 °C. Meanwhile, 30 µL of ExpiFectamine™ was diluted in 1 mL OptiMEM (51985034, Thermo Fisher, Germany) reduced serum medium. 5 min later, 12.5 µg of bacmid DNA is added and the mixture is gently mixed. Another 5 min later, the transfer mixture was pipetted on to the cells and put back on to the shaker. Approximately, 72 h later, the supernatant was harvested and stored at −80 °C as the P0 virus stock. New cells were then infected with the P0 stock and harvested five days later. This was then again performed to get P2 viral stocks that were stored at −80 °C.

5.1.3. Determination of FFU for P2 viral stocks by light microscopy titer assay

The fluorescent-forming units (FFU) for the P2 stocks of His-UL15 and Dual were determined by light microscopy. On day one $0.2 \cdot 10^6$ cells/well were seeded into a 24-well plate (83.3922, Sarstedt, Germany) and infected with P0/P1 stock in a dilution series (10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}). They were then mixed by slowly shaking the plate and afterward incubated at 27 °C for 16 h without shaking. On day two, the supernatant was carefully discarded and the cells were cautiously washed with phosphate-buffered saline (PBS) twice. To reduce non-specific binding, 0.3 g of bovine serum albumin (BSA) in 10 mL PBS (300 µL per well), was added to the wells and incubated for 30 min. The cells were then again washed with PBS twice and 100 µL of diluted gp64 antibody (a baculovirus envelope-specific antibody; 14-6995-82, Invitrogen, Thermo Fisher, Germany) was added for 30 min. The cells were again washed twice and 100 µL of diluted Hoechst and Alexa 647 (Hoechst: stains DNA;

Alexa 647: secondary antibody with fluorescent marker) were added and it was again incubated for 1 h. The well-plate was then put onto a wide-field microscope (Leica DMi8) and infected cells were counted.

5.1.4. Calculation of the ffu and MOI for the baculovirus titer assay

Calculation of the baculovirus titer from counted FFU:

$$\frac{N_{ffu(perwell)}}{0.2\text{mL} * dilution factor} = titer[\frac{ffu}{\text{mL}}] \tag{5.1}$$

Calculating the number of FFU for specific multiplicity of infection (MOI):

$$N_{ffu} = Ncells * MOI \tag{5.2}$$

Virus volume needed for infection:

$$V_{baculovirus} = \frac{N_{ffu}}{titer} \tag{5.3}$$

5.1.5. Expression of terminase and pUL32 in ExpiSf™ cells

Bacmid and Baculovirus generation from the pFastBac1 plasmid containing UL32 (GeneArt, Thermo Fisher, USA) was performed analogous to UL15 in chapter 5.1.1. Expression was performed according to the ExpiSf™ Expression System User Guide (Gibco, Thermo Fisher). Expression was performed in two 25 mL suspension cultures for the terminase and 200 mL for pUL32 and the cells were infected with P2 viral stock at an MOI of 3. The cells were harvested after three days at 300 xg for 5 min and washed once with PBS. The pellets were then stored at −80 °C.

5.1.6. Expression of pORF68 in HEK 293 suspension cells

The protein pORF68 was expressed in a Freestyle™ 293 expression system (Gibco, Thermo Fisher), the HEK 293 cells are kept in 100 mLsuspension cultures. For transfection, 115 µg ORF68 plasmid (pcDNA4/TO-2xStrep-ORF68 was a gift from Britt Glaunsinger (Addgene plasmid #162625;http://n2t.net/addgene:162625; RRID:Addgene_162625)) and 10 µg (expression control) were complexed by the transfection reagent 293fectin™. Cells were harvested 48 hpi at 300 xg for5 min and washed once with PBS. The pellets were then stored at −80 °C.

5.1.7. Purification of terminase complex via affinity and size exclusion chromatography

The pellet was thawed on ice and resuspended in ice-cold lysis buffer (**L**) for 15 min. Afterward, a Dounce homogenizer was used to break the cell walls and release the proteins. The suspension was then centrifuged twice at 12 000 xg 20 min. The supernatant was then filtered through a 0.22 µm filter. Sample purification was then performed on an ÄKTA system for affinity chromatography and SEC. At first, the sample was run over a 1 mL His-Trap HP column (29051021, Cytiva, USA), then it was purified further on a Superdex™ 200 increase 10-300 GL (28990944, Cytiva, USA) column. In between runs, the sample was concentrated on a spin filter (Amicon®Ultra-Centrifugal Filter, 10 kDa MWCO, Merck KGaA, Darmstadt, Germany) to a sample size of approximately 500 µL. The elution was then rapidly frozen down to −80 °C. Elution A6 from the SEC run was then buffer exchanged five times (Amicon®Ultra-Centrifugal Filter, 100 kDa MWCO, Merck KGaA, Darmstadt, Germany) in to 150 mM ammonium acetate buffer surrogate at pH 7.5.

5.1.8. Purification of terminase complex via affinity without size exclusion chromatography

The pellet was thawed at 4 °C in 10 mL **L** on a rolling shaker. Meanwhile, 1 mL of a 50 % Ni-nitrilotriacetic acid (NTA) suspension was pipetted into an empty gravity column and equilibrated with 3x3 cV **L**. The lysate was then sonicated twice for 5 min (1 s ON, 1.5 s OFF, 60 % and afterward centrifuged for 45 min at 40 000 xg. The lysate was then applied to the gravity column and incubated on a rolling shaker at 4 °C for 1 h. The flow-through was then collected and the beads washed with wash buffer A (**WA**) for 10 cV. Afterward, the column was washed with wash buffer B (**WB**) until the concentration of the solution reached zero. The terminase was then eluted with elution buffer (**E**) in 500 µL steps until no more protein was eluted. It was then buffer exchanged with a 100 kDa cut-off spin filter (Amicon®Ultra-Centrifugal Filter, 100 kDa MWCO, Merck KGaA, Darmstadt, Germany) and applied to the Q-TOF 2 mass spectrometer.

5.2.   Fast-track nMS protocol

5.2.1. Preparation of ADH for the fast-track protocol experiments

6 mg of ADH (37 kDa, Alcohol Dehydrogenase from *Saccharomyces cerevisiae*, Sigma Aldrich) was dissolved in 100 µL Milli-Q water. A 50 µL aliquot was applied to a Biospin column (Biospin mini columns, 6000 MWCO, Bio-Rad) pre-equilibrated with 150 mM ammonium acetate (99.99 % purity, Sigma-Aldrich) at pH 8. The column was centrifuged at 1 000 rcf at 4 °C. ADH in ammonium acetate was eluted and diluted to final concentrations of 7.5 µM and 2.5 µM. Aliquots from this preparation were then spiked with varying concentrations of imidazole, biotin and desthiobiotin.

5.2.2. HSV-1 terminase purification with fast-track protocol for QE-UHMR

3 mL of the pellet were lysed in 6 mL **L** on ice. They were then sonicated as previously described and centrifuged for 10 min at 20 000 xg. This time only 0.5 mL of the Ni-NTA suspension was prepared. Before adding the lysate to the resin, it was filtered through 0.22 µm. Flow-through was again collected and the resin was then washed with 10 bV of **WA**. A bed volume (bV) refers to the volume that is equal to the volume of the resin, whereas a column volume (cV) refers to the total volume of the column. 3 bV of **WB** were then added to remove non-specific proteins. Elution was performed with 5 bV into five separate reaction tubes. Protein concentration was then determined by UV-spectroscopy (DeNovix).

5.2.3. Desalting of *pac* samples and preparation of terminase samples with *pac*

All four oligomeres were ordered from IDT (Integrated DNA Technologies, Inc., USA) in powder form and resuspended in Milli-Q water. They were then desalted with a pre-equilibrated Biospin P6 spin column in 300 mM ammonium acetate, pH 8. They were then diluted to a total concentration of 10 µM. The terminase with *pac* samples were prepared by mixing both samples in a ratio of 1:9 (terminase : *pac*) for the total concentration of 3 µM terminase and 9 µM *pac*. For the DNase samples 0.5 µL of benzonase (Thermo Fisher) was added to the samples and incubated on 4 °C for 2 h.

5.2.4. Purification of pUL32 and pORF68 via fast-track protocol

The cloning and expression of both proteins were analogous to the terminase. Both proteins are purified with a Strep-tag system. To the frozen cell pellet, three times the volume **L** was added and the cells were lysed on ice via sonication twice for 5 min (1 s ON, 1.5 s OFF, 60 %). Pre-packaged Strep-Tactin Sepharose (1mL, IBA) columns were washed with 2 cV wash buffer (**W**). Meanwhile, the lysate was centrifuged at 20 000 xg for 5 min at 4 °C and 100 L Biolock was added. The lysate was then added on to the column and the flow-through was collected. Afterward, the column was washed five times with **W**. Then, six times 0.3 cV of **E** were added and the elutions were collected in 1.5 mL tubes. The protein concentration was calculated via UV-spectroscopy (DeNovix).

5.3. Protein expression and purification of the fluorescent proteins

5.3.1. Cloning and expression of fluorescent proteins in *E. coli*

Cloning of FPs into pET21a(+) Vector

The genome of all FPs were taken from in lab fusion proteins with said fluorescent protein and cloned via In-Fusion Snap Assembly (Takara Bio Inc.) into pET21a(+) vectors with N-terminal 6xHis-tag. Here, approximately 50 ng of the PCR DNA fragment containing the His-tag and the FP were added to a mix containing 2 µL of the In-Fusion Master Mix, 50 ng of the linearized vector and water, to a total volume of 10 µL. The reaction was then incubated for 15 min at 50 °C and then placed on ice. The DNA was then transformed into DH5α (Thermo Fisher, USA) competent cells

overnight. The next day, a colony was picked and expanded in 5 mL Terrific Broth (TB) medium with ampicillin. The DNA was then purified with the NucleoSpin® Plasmid (NoLid, Macherey-Nagel) Miniprep Kit.

Transformation of the FPs into BL21 DE3 cells

BL21 DE3 cells (25 µL) were thawed on ice for 10 min before 1 mL of the pET-vector containing the target DNA was pipetted on to the cells and carefully mixed. After chilling, the cells on ice for 30 min a heat shock was performed. Here, the cells were placed in a water bath (42 °C) for 45 s. Afterward, cells were put back on ice for 2 min. For the incubation 475 µL of super optimal broth with catabolic repressor (SOC) media was added to the cells and they were transferred to a heated shaker (37 °C) for 1 h. After the time has passed 50 µL of the *E. coli* suspension was spread on to an agar plate containing the appropriate resistance marker. On the following day, a colony was picked and transferred into an culture tube containing 5 mL TB media and the same resistance marker as previously used for the plates. The tube was then shaken overnight at a temperature of 30 °C.

Expression of the FPs

1 mL of the overnight culture was pipetted into an 500 mL Erlenmeyer flask containing 100 mL TB media and the antibiotic and shook at (37 °C) until an OD_{600} of 0.8 was reached. At this point, 80 µL IPTG for a total concentration of 0.5 M was added. Proteins were expressed at 30 °C for 2.5 h at 250 rpm. Cells were harvested by spinning them down into 50 mL tubes at 3 300 rcf, 4 °C for 10 min. Pellets were then stored at −20 °C.

5.3.2. Purification of fluorescent proteins with fast-track protocol

Lysis of *E. coli* cells containing FPs

About 3 mL of frozen pellet were thawed in 6 mL of ice-cold **L**. It was resuspended by carefully pipetting until the solution was homogeneous. Lysis was performed on ice via sonification (ON 3 s, OFF 15 s, TOTAL ON 30 s, 70 %) and followed up by centrifuging for 40 min at 20.000 rcf in 1.5 mL tubes. The clear supernatant was filtered through a 0.22 µL filter and loaded on to a gravity flow column containing pre equilibrated Ni-NTA beads.

FP purification via Ni-NTA gravity column

After adding the supernatant to the column, it was sealed and incubated with the Ni-NTA beads for 1 h on a roll shaker. Afterward, the column was put back into a vertical position and flow-through was collected. The beads were then washed with ten times bed volume of **WA**, followed by three times bed volume **WB**. The proteins were eluted with five times **E** and collected as 500 µL fractions in suitable tubes. After elution the beads were cleaned and regenerated.

5.4. Circular Dichroism of *pac* sequences

The DNA oligomers arrived in powdered form and were solved with RNase-free water to a concentration of 20 µM. Samples were prepared at room temperature and then heated up to 90 °C for 5 min. Afterward, they were cooled down again to room temperature for 2.5 h and stored at 4 °C. After about 72 h, the samples were then applied to the CD spectrometer. A high-precision cell (1 mm, Hellma Analytics) was

filled with 300 µL of a sample and inserted into the circular dichroism spectrometer (Chirascan, Applied Photophysics, UK). With the Pro-Data Chirascan software, the following parameters were adjusted:

Table 5.2: Parameters of the CD experiments for the *pac* sequence secondary structure analysis

wavelength	200-320 nm
wavelength steps	0.5 nm
bandwidth	1.0
temperature	20 °C
time-per-point	0.5 s
replicates	3

5.5. Native mass spectrometry

To transfer the sample into the mass spectrometer, nESI capillaries were pulled on a micropipette puller (P-1000, Sutter Instruments). A two-step method to heat a squared-box filament (2.5 mm x 2.5 mm) was used for pulling the borosilicate capillaries (1.2 mm outer and 0.68 mm inner diameter, World Precision Instruments). The glass capillaries were then gold-coated by a sputter coater (CCU-010, Safematic, $5.0 \cdot 10^{-2}$ mbar, 30.0 mA, 120 s, three runs to vacuum limit $3.0 \cdot 10^{-2}$ mbar argon).

5.5.1. Instrument settings

Two instruments were used for mass spectrometry, the first was a Q-TOF 2 (Waters, Manchester, UK and MS Vision, Almere, the Netherlands) that was modified for high masses[97] and mass calibrated with CsI ($25 \, \mathrm{mg \, mL^{-1}}$). The second was a QExactive UHMR Orbitrap (Thermo Fisher, USA).

Q-TOF 2 parameters of the ADH measurements with different eluents

ADH measurements were analyzed in triplicates. The capillary voltage was kept at 1.35 kV and the cone voltage at 150 V. The cone temperature was kept at 80 °C. The collision energy was ramped from 10 to 200 V. Argon was used as a collision gas in the collision cell. The pusher time was set as 190 µs the acquisition window was 100 to 30,000 m/z. Gas pressure in the collision cell was fixed at $1.8 \cdot 10^{-2}$ mbar for both measurements.

Q Exactive UHMR parameters for the ADH measurements with different eluents

Positive ion mode was used and the CV was kept between 5 to 300 eV in the HCD cell. Trapping gas (nitrogen) pressure was set to 7. Detector optimization was set to "low m/z". The Ion transfer m/z settings were adjusted as follows:

Table 5.3: QExactive UHMR Orbitrap settings for ADH measurements.

Parameter	
spray voltage (V)	1.35
capillary temp.	80
Inj. Fl. RF Ampl.	300 V
Bent. Fl. RF Ampl.	940
Trans MP	900
HCD-cell RF Ampl.	900
C-Trap RF Ampl.	2950

Q-TOF 2 parameters for HSV-1 terminase measurements

For the measurements of the non fast-track purified terminase, the sample was measured at a capillary voltage of 1.30 kV and the cone voltage at 130 V. The Cone temperature was kept at 80 °C. The collision energy was 25 V and argon was used as a collision gas. The pusher time was set as 120 µs the acquisition window was 100 to 20,000 m/z. The gas pressure in the collision cell was fixed at $1.5 \cdot 10^{-2}$ mbar.

The fast-track purified terminase was measured with a capillary voltage of 1.5 kV

and cone voltage of 150 V. The cone temperature was kept at 80 °C and for CID argon gas was used with a collision voltage of 50 V with a gas pressure of $1.5 \cdot 10^{-2}$ mbar. Pusher time was set as 120 μs and the acquisition window was kept between 100 and 50000 m/z.

Q Exactive UHMR parameters for HSV-1 terminase measurements

For the terminase measurements on the QE-UHMR the CV in the HCD cell was kept off. The trapping gas (nitrogen) pressure was set to 5. The acquisition window was fixed to 500-30000 m/z. The cone temperature was kept at 250 °C. More parameters are displayed in **Figure** 5.1.

Q Exactive UHMR parameters for pUL32 measurements

The pUL32 measurements were conducted on the QE-UHMR. The CV was kept at 250 eV. The mass range was between $m/z = 500 - 30000$. The cone temperature was kept at 50 °C. The trapping gas (nitrogen) pressure was set to 7. More parameters are displayed in **Figure** 5.2.

Q Exactive UHMR parameters for pORF68 measurements

The pORF68 measurements were conducted on the QE-UHMR. The CV was kept at 300 eV. The mass range was between $m/z = 500 - 30000$. The cone temperature was kept at 50 °C. The trapping gas (nitrogen) pressure was set to 7. More parameters are displayed in **Figure** 5.3.

Q Exactive UHMR parameters for fluorescent proteins measurements

For the FP measurements, the CV in the HCD cell was at 10 eV. The trapping gas (nitrogen) pressure was set to 7. The acquisition window was fixed to 500-15000 m/z. More parameters are displayed in **Figure** 5.4.

```
20240912AFRG04
Segment: 65536
```

=== Tune Data: ===:	
Spray Voltage (+):	1500.00
Spray Voltage (-):	1500.00
Capillary Temperature (+ or +-):	250.00
Capillary Temperature (-):	250.00
Sheath Gas (+ or +-):	0.00
Sheath Gas (-):	0.00
Aux Gas (+ or +-):	0.00
Aux Gas (-):	0.00
Spare Gas (+ or +-):	0.00
Spare Gas (-):	0.00
Max Spray Current (+):	50.00
Max Spray Current (-):	50.00
Probe Heater Temp. (+ or +-):	350.00
Probe Heater Temp. (-):	350.00
S-Lens RF Level:	200.00
Ion Source:	NSI
IST desolvation voltage (+):	-150.00
IST desolvation voltage (-):	150.00
=== Calibration Data: ===:	
Mass Cal. (+) age (d):	0.00
Mass Cal. (-) age (d):	0.00
Isolation Cal. (+) age (d):	0.0
Isolation Cal. (-) age (d):	0.0
Amplifier Gain:	800.000

De-Inject Voltage (-):	0.0
De-Measure Voltage (+):	400.0
De-Measure Voltage (-):	410.0
CE-Inject Voltage (+):	3700.0
CE-Inject Voltage (-):	3700.0
CE-Inject Volt (+) HM:	3200.0
CE-Inject Volt (-) HM:	3250.0
LowSNQuanLevel:	1
=== Instrument Mode: ===:	
UHMR Mode Setting:	On
=== Configuration Data: ===:	
Preamp Protect Mode:	0
TMP Sweep Mode:	0
Temp Sensor Mode:	0
=== Identification: ===:	
SW Version:	2.13-213012/2.13.0.3162
Model:	Q Exactive UHMR
Instrument:	Exactive S... #SN08065L
Hardware Id:	0030591C004B
=== Direct Mass Data: ===:	
Detector m/z Optimization:	High m/z
Ion Transfer Target:	High m/z
Trapping gas pressure setting:	5.0
Averaging:	0
AGC mode:	fixed
Direct Mass enabled:	No

Figure 5.1: MS parameter for terminase measurements on the QE-UHMR.

```
20231009AFRG05
Segment: 65536
=== Tune Data: ===:
Spray Voltage (+):                      1450.00
Spray Voltage (-):                      1500.00
Capillary Temperature (+ or +-):        50.00
Capillary Temperature (-):              250.00
Sheath Gas (+ or +-):                   0.00
Sheath Gas (-):                         0.00
Aux Gas (+ or +-):                      0.00
Aux Gas (-):                            0.00
Spare Gas (+ or +-):                    0.00
Spare Gas (-):                          0.00
Max Spray Current (+):                  50.00
Max Spray Current (-):                  50.00
Probe Heater Temp. (+ or +-):           350.00
Probe Heater Temp. (-):                 350.00
S-Lens RF Level:                        200.00
Ion Source:                             NSI
IST desolvation voltage (+):            -150.00
IST desolvation voltage (-):            150.00
=== Calibration Data: ===:
Mass Cal. (+) age (d):                  0.00
Mass Cal. (-) age (d):                  0.00
Isolation Cal. (+) age (d):             0.0
Isolation Cal. (-) age (d):             0.0
Amplifier Gain:                         800.000
De-Inject Voltage (-):                  0.0
De-Measure Voltage (+):                 400.0
De-Measure Voltage (-):                 410.0
CE-Inject Voltage (+):                  3700.0
CE-Inject Voltage (-):                  3700.0
CE-Inject Volt (+) HM:                  3200.0
CE-Inject Volt (-) HM:                  3250.0
LowSNQuanLevel:                         1
=== Instrument Mode: ===:
UHMR Mode Setting:                      On
=== Configuration Data: ===:
Preamp Protect Mode:                    0
TMP Sweep Mode:                         0
Temp Sensor Mode:                       0
=== Identification: ===:
SW Version:                             2.13-213012/2.13.0.3162
Model:                                  Q Exactive UHMR
Instrument:                             Exactive S... #SN08065L
Hardware Id:                            0030591C004B
=== Direct Mass Data: ===:
Detector m/z Optimization:              Low m/z
Ion Transfer Target:                    High m/z
Trapping gas pressure setting:          7.0
Averaging:                              0
AGC mode:                               fixed
Direct Mass enabled:                    No
```

Figure 5.2: MS parameter for pUL32 measurements on the QE-UHMR.

```
20231009AFRG07
Segment: 65536
=== Tune Data: ===:
Spray Voltage (+):                    1450.00
Spray Voltage (-):                    1500.00
Capillary Temperature (+ or +-):      50.00
Capillary Temperature (-):            250.00
Sheath Gas (+ or +-):                 0.00
Sheath Gas (-):                       0.00
Aux Gas (+ or +-):                    0.00
Aux Gas (-):                          0.00
Spare Gas (+ or +-):                  0.00
Spare Gas (-):                        0.00
Max Spray Current (+):                50.00
Max Spray Current (-):                50.00
Probe Heater Temp. (+ or +-):         350.00
Probe Heater Temp. (-):               350.00
S-Lens RF Level:                      200.00
Ion Source:                           NSI
IST desolvation voltage (+):          -150.00
IST desolvation voltage (-):          150.00
=== Calibration Data: ===:
Mass Cal. (+) age (d):                0.00
Mass Cal. (-) age (d):                0.00
Isolation Cal. (+) age (d):           0.0
Isolation Cal. (-) age (d):           0.0
Amplifier Gain:                       800.000
De-Inject Voltage (-):                0.0
De-Measure Voltage (+):               400.0
De-Measure Voltage (-):               410.0
CE-Inject Voltage (+):                3700.0
CE-Inject Voltage (-):                3700.0
CE-Inject Volt (+) HM:                3200.0
CE-Inject Volt (-) HM:                3250.0
LowSNQuanLevel:                       1
=== Instrument Mode: ===:
UHMR Mode Setting:                    On
=== Configuration Data: ===:
Preamp Protect Mode:                  0
TMP Sweep Mode:                       0
Temp Sensor Mode:                     0
=== Identification: ===:
SW Version:                           2.13-213012/2.13.0.3162
Model:                                Q Exactive UHMR
Instrument:                           Exactive S... #SN08065L
Hardware Id:                          0030591C004B
=== Direct Mass Data: ===:
Detector m/z Optimization:            High m/z
Ion Transfer Target:                  High m/z
Trapping gas pressure setting:        7.0
Averaging:                            0
AGC mode:                             fixed
Direct Mass enabled:                  No
```

Figure 5.3: MS parameter for pORF68 analysis on the QE-UHMR

```
20240213AFRG15
Segment: 65536
=== Tune Data: ===:
Spray Voltage (+):                    1150.00
Spray Voltage (-):                    1500.00
Capillary Temperature (+ or +-):      50.00
Capillary Temperature (-):            250.00
Sheath Gas (+ or +-):                 0.00
Sheath Gas (-):                       0.00
Aux Gas (+ or +-):                    0.00
Aux Gas (-):                          0.00
Spare Gas (+ or +-):                  0.00
Spare Gas (-):                        0.00
Max Spray Current (+):                50.00
Max Spray Current (-):                50.00
Probe Heater Temp. (+ or +-):         350.00
Probe Heater Temp. (-):               350.00
S-Lens RF Level:                      200.00
Ion Source:                           NSI
IST desolvation voltage (+):          -150.00
IST desolvation voltage (-):          150.00
=== Calibration Data: ===:
Mass Cal. (+) age (d):                0.00
Mass Cal. (-) age (d):                0.00
Isolation Cal. (+) age (d):           0.0
Isolation Cal. (-) age (d):           0.0
Amplifier Gain:                       800.000
De-Inject Voltage (-):                0.0
De-Measure Voltage (+):               400.0
De-Measure Voltage (-):               410.0
CE-Inject Voltage (+):                3700.0
CE-Inject Voltage (-):                3700.0
CE-Inject Volt (+) HM:                3200.0
CE-Inject Volt (-) HM:                3250.0
LowSNQuanLevel:                       1
=== Instrument Mode: ===:
UHMR Mode Setting:                    On
=== Configuration Data: ===:
Preamp Protect Mode:                  0
TMP Sweep Mode:                       0
Temp Sensor Mode:                     0
=== Identification: ===:
SW Version:                           2.13-213012/2.13.0.3162
Model:                                Q Exactive UHMR
Instrument:                           Exactive S... #SN08065L
Hardware Id:                          0030591C004B
=== Direct Mass Data: ===:
Detector m/z Optimization:            Low m/z
Ion Transfer Target:                  High m/z
Trapping gas pressure setting:        7.0
Averaging:                            0
AGC mode:                             fixed
Direct Mass enabled:                  No
```

Figure 5.4: MS parameter for all FP measurements on the QE-UHMR.

5.6.   Data Analysis

5.6.1. Mass spectra analysis

Mass spectra from the Q-TOF 2 instrument were analyzed using MassLynx™ (version 4.1, Waters, Manchester, UK), while spectra from the QE-UHMR were processed with Xcalibur (version 4.2.47, Thermo Fisher). A high-quality or charge state-resolved spectrum is defined by clear peak separation, where adjacent peaks are distinguishable at no less than half of their maximum height. Spectra were then exported to Adobe Illustrator (version 19.5.0.84, Adobe) for figure preparation. Peak detection and FWHM determination were carried out with a custom Python script and with mMass (ver.5.5.0)[233]. Peak detection and deconvolution for the nMS spectra was performed with mMass (ver.5.5.0)[233], UniDec (ver. 6.0.3)[234] and Massign[235]. The FWHM were multiplied with the charge for normalization and error from replicates were derived from Gaussian error propagation.

5.6.2. Calculation of the dissociation constant K_d of the FPs

Defining the dissociation constant with $[A]$ being the concentration of substance A, same goes for B. $[AB]$ is the concentration for the complex AB:

$$K_d = \frac{[A][B]}{[AB]} \tag{5.4}$$

Applying for the FPs where the dimer is $B = A$:

$$K_d = \frac{[FP]^2}{[FP_2]} \tag{5.5}$$

The initial concentration $[FP]_0$ is composed of the monomer concentration $[FP]$ and dimer concentration $[FP_2]$.

$$[FP]_0 = [FP] + 2 \cdot [FP_2] \tag{5.6}$$

The AUC of the species (e.g., monomer) is obtained from the spectra and is proportional to the concentration.

$$[FP] \propto AUC_{FP} \tag{5.7}$$

The monomer concentration can therefore be calculated as followed.

$$[FP] = [FP]_0 \cdot \frac{AUC_{FP}}{AUC_{FP} + 2 \cdot AUC_{FP_2}} \tag{5.8}$$

The dimer concentration is calculated analogous.

$$[FP_2] = [FP]_0 \cdot \frac{AUC_{FP_2}}{AUC_{FP} + 2 \cdot AUC_{FP_2}} \tag{5.9}$$

These are then inserted into the K_d equation.

$$K_d = \frac{\left([FP]_0 \frac{AUC_{FP}}{AUC_{FP} + 2 \cdot AUC_{FP_2}} \right)^2}{[FP]_0 \frac{AUC_{FP_2}}{AUC_{FP} + 2 \cdot AUC_{FP_2}}} \tag{5.10}$$

Since the AUC values were derived from triplicates, the average was used and a Gaussian error propagation was performed.

5.6.3. Data analysis of the FPs with python

To deconvolute the data, first the raw data had to be converted in to text format. This was performed by a python script (ver. 3.8) written by **Janine-Denise Kopicki** (see

?? A1). The amount of data after deconvolution due to the number of samples at different conditions was considerable. To handle overlapping charge states in the FP dimerization spectra, the following Python script, developed in Spyder (ver. 5.5.1, python ver. 3.12.4), was used to estimate adjusted intensities for each species:

```python
import pandas as pd

def calculate_adjusted_intensity_sums(file_path, sheet_name,
    output_path):
    # Load data from the specified sheet
    data = pd.read_excel(file_path, sheet_name=sheet_name)


    # Prepare a list to store results
    results = []


    # Constants
    OVERLAPPING_MONOMER_CHARGE = 8
    OVERLAPPING_DIMER_CHARGE = OVERLAPPING_MONOMER_CHARGE * 2


    # Group by concentration and replicate
    for (concentration, replicate), group in data.groupby(['c in uM
        ', 'no.']):
        monomer_peaks = group[group['species'] == 'monomer']
        dimer_peaks = group[group['species'] == 'dimer']


        # Identify peaks
        overlapping_monomer = monomer_peaks[monomer_peaks['z'] ==
```

```python
                    OVERLAPPING_MONOMER_CHARGE]
21          overlapping_dimer = dimer_peaks[dimer_peaks['z'] ==
                OVERLAPPING_DIMER_CHARGE]
22          neighbor_monomer = monomer_peaks[monomer_peaks['z'] == (
                OVERLAPPING_MONOMER_CHARGE + 1)]
23          neighbor_dimer = dimer_peaks[dimer_peaks['z'] == (
                OVERLAPPING_DIMER_CHARGE - 1)]
24
25          # Initialize adjusted intensity sums
26          adjusted_monomer_intensity_sum = 0
27          adjusted_dimer_intensity_sum = 0
28
29          # Add non-overlapping intensitys to the sums
30          adjusted_monomer_intensity_sum += monomer_peaks[
                monomer_peaks['z'] != OVERLAPPING_MONOMER_CHARGE]['
                intensity'].sum()
31          adjusted_dimer_intensity_sum += dimer_peaks[dimer_peaks['z'
                ] != OVERLAPPING_DIMER_CHARGE]['intensity'].sum()
32
33          # Handle overlapping peak
34          if not overlapping_monomer.empty and not overlapping_dimer.
                empty:
35              # Extract intensity values
36              intensity_overlap = overlapping_monomer['intensity'].
                    iloc[0]
37              intensity_neighbor_dimer = neighbor_dimer['intensity'].
                    iloc[0] if not neighbor_dimer.empty else None
38              intensity_neighbor_monomer = neighbor_monomer['
```

```python
                        intensity'].iloc[0] if not neighbor_monomer.empty
                        else None


            if intensity_neighbor_dimer is not None and
                intensity_neighbor_monomer is not None:
                # Both neighbors exist: calculate contributions
                dimer_contrib = (intensity_neighbor_dimer / (
                    intensity_neighbor_dimer +
                    intensity_neighbor_monomer)) * intensity_overlap
                monomer_contrib = (intensity_neighbor_monomer / (
                    intensity_neighbor_dimer +
                    intensity_neighbor_monomer)) * intensity_overlap
            elif intensity_neighbor_dimer is not None:
                # Only Dimer neighbor exists: 100% to Dimer
                dimer_contrib = intensity_overlap
                monomer_contrib = 0
            elif intensity_neighbor_monomer is not None:
                # Only Monomer neighbor exists: 100% to Monomer
                dimer_contrib = 0
                monomer_contrib = intensity_overlap
            else:
                # No neighbors exist: cannot calculate
                    contributions
                dimer_contrib = None
                monomer_contrib = None


            # Add overlapping contributions to the sums if they
                exist
```

```python
58              if dimer_contrib is not None:
59                  adjusted_dimer_intensity_sum += dimer_contrib
60              if monomer_contrib is not None:
61                  adjusted_monomer_intensity_sum += monomer_contrib
62
63              # Append to results
64              results.append({
65                  "Concentration (uM)": concentration,
66                  "Replicate": replicate,
67                  "Overlapping Peak intensity": intensity_overlap,
68                  "Neighbor Monomer intensity":
                        intensity_neighbor_monomer,
69                  "Neighbor Dimer intensity":
                        intensity_neighbor_dimer,
70                  "Dimer Contribution": dimer_contrib,
71                  "Monomer Contribution": monomer_contrib,
72                  "Adjusted Monomer intensity Sum":
                        adjusted_monomer_intensity_sum,
73                  "Adjusted Dimer intensity Sum":
                        adjusted_dimer_intensity_sum,
74                  "Status": "Calculation successful" if dimer_contrib
                         is not None and monomer_contrib is not None
                        else "Missing neighbor(s)"
75              })
76          else:
77              # If overlapping peaks are missing
78              results.append({
79                  "Concentration (uM)": concentration,
```

```python
                    "Replicate": replicate,
                    "Overlapping Peak int": overlapping_monomer['
                        intensity'].iloc[0] if not overlapping_monomer.
                        empty else None,
                    "Neighbor Monomer intensity": neighbor_monomer['
                        intensity'].iloc[0] if not neighbor_monomer.
                        empty else None,
                    "Neighbor Dimer intensity": neighbor_dimer['
                        intensity'].iloc[0] if not neighbor_dimer.empty
                        else None,
                    "Dimer Contribution": None,
                    "Monomer Contribution": None,
                    "Adjusted Monomer intensity Sum":
                        adjusted_monomer_intensity_sum,
                    "Adjusted Dimer intensity Sum":
                        adjusted_dimer_intensity_sum,
                    "Status": "Missing overlapping peak(s)"
                })

    # Convert results to DataFrame
    results_df = pd.DataFrame(results)


    # Save results to Excel
    results_df.to_excel(output_path, index=False, sheet_name=f"
        Results_{sheet_name}")
    print(f"Results saved to {output_path}")


# Example usage
```

```python
99  file_path = r"C:\Users\gruena\Desktop\FP_dataanalysis\
        dimerization_fp.xlsx"  # Path to the input file
100 sheet_name = "mCherry"  # Sheet to process
101 output_path = r"C:\Users\gruena\Desktop\FP_dataanalysis\
        mCherry_intensity.xlsx"  # Path to the output file
102 calculate_adjusted_intensity_sums(file_path, sheet_name,
        output_path)
```

Listing 5.1: Python script for intensity distribution of overlapping peaks

5.7. Materials and buffer

Table 5.4: Protein purification buffer for Strep-Tag purification

Fast-track pUL32 and pORF68	Lysis buffer	ammonium acetate pH 8
		1 mM DTT
		0.1 % CHAPS
		EDTA-free protease inhibitor cocktail
	Wash buffer	400 mM ammonium acetate pH 8
		1 mM DTT
		0.1 % CHAPS
	Elution buffer	400 mM ammonium acetate pH 8
		1 mM DTT
		0.1 % CHAPS
		2.5 mM desthiobiotin
	Regeneration buffer R	100 mM Tris-Cl
		150 mM NaCl
		1 mM EDTA
		1 mM HABA

Table 5.5: Protein purification buffer for His-Tag purification

Non fast-track terminase	Lysis buffer	20 mM HEPES pH 8
		400 mM NaCl
		1.5 mM MgCl2
		1 mM DTT
		EDTA-free protease inhibitor cocktail
	Wash buffer	20 mM HEPES pH 8
		400 mM NaCl
		1.5 mM MgCl2
		1 mM DTT
	Elution buffer	20 mM HEPES pH 8
		400 mM NaCl
		1.5 mM MgCl2
		1 mM DTT
		500 mM imidazole
	Buffer exchange buffer	150 mM ammonium acetate pH 8
Non fast-track, non SEC terminase	Lysis buffer	20 mM HEPES pH 7.5
		300 mM NaCl
	Wash buffer A	20 mM HEPES pH 7.5
		500 mM NaCl
	Wash buffer B	20 mM HEPES pH 7.5
		500 mM NaCl
		20 mM imidazole
	Elution buffer	20 mM HEPES pH 7.5
		500 mM NaCl
		200 mM imidazole
	Buffer exchange buffer	320 mM ammonium acetate pH 7.5
Fast-track terminase UHMR	Lysis buffer	1 M ammonium acetate pH 7.5
		1 mM DTT
		EDTA-free protease inhibitor cocktail
	Wash buffer A	320 mM ammonium acetate pH 7.5
		1 mM DTT
	Wash buffer B	320 mM ammonium acetate pH 7.5
		1 mM DTT
		35 mM imidazole
	Elution buffer	320 mM ammonium acetate pH 7.5
		350 mM imidazole
Fast-track FP UHMR	Lysis buffer	1 M ammonium acetate pH 7.5
		1 mM DTT
		EDTA-free protease inhibitor cocktail
	Wash buffer A	350 mM ammonium acetate pH 7.5
		1 mM DTT
	Wash buffer B	350 mM ammonium acetate pH 7.5
		1 mM DTT
		35 mM imidazole
	Elution buffer	350 mM ammonium acetate pH 7.5
		350 mM imidazole

References

(1) Herpes simplex virus - WHO. `https://www.who.int/news-room/fact-sheets/detail/herpes-simplex-virus`. (accessed on 2025-01-12).

(2) Jellinge, M. E.; Hansen, F.; Coia, J. E.; Song, Z. *Journal of Intensive Care Medicine* **2021**, *36*, 1398–1402.

(3) Meyer, A. et al. *Critical Care* **2021**, *25*, 417.

(4) Pérez-Pedrero Sánchez-Belmonte, M. J.; Sánchez-Casado, M.; Moran Gallego, F. J.; Piza Pinilla, R.; Gomez Hernando, C.; Paredes Borrachero, I. *Medicina Clínica (English Edition)* **2023**, *160*, 66–70.

(5) Kabani, N.; Kimberlin, D. W. *NeoReviews* **2018**, *19*, e89–e96.

(6) Roizman, B., *The Herpesviruses*; Springer Science & Business Media: 2013.

(7) Bai, L.; Xu, J.; Zeng, L.; Zhang, L.; Zhou, F. *Molecular Biomedicine* **2024**, *5*, 35.

(8) Holwerda, B. C. *Antiviral Research* **1997**, *35*, 1–21.

(9) Andrei, G.; De Clercq, E.; Snoeck, R. *Infectious Disorders - Drug TargetsDisorders)* **2009**, *9*, 201–222.

(10) Yang, L.; Yang, Q.; Wang, M.; Jia, R.; Chen, S.; Zhu, D.; Liu, M.; Wu, Y.; Zhao, X.; Zhang, S.; Liu, Y.; Yu, Y.; Zhang, L.; Chen, X.; Cheng, A. *Viruses* **2019**, *11*, 219.

(11) Field, H. J.; Vere Hodge, R. A. *British Medical Bulletin* **2013**, *106*, 213–249.

(12) Saullo, J. L.; Miller, R. A. *Annual Review of Medicine* **2023**, *74*, 89–105.

(13) Ljungman, P.; Schmitt, M.; Marty, F. M.; Maertens, J.; Chemaly, R. F.; Kartsonis, N. A.; Butterton, J. R.; Wan, H.; Teal, V. L.; Sarratt, K.; Murata, Y.; Leavitt, R. Y.; Badshah, C. *Clinical Infectious Diseases* **2020**, *70*, 1525–1533.

(14)   Current ICTV Taxonomy Release | ICTV. https://ictv.global/taxonomy. (accessed on 2025-12-23).

(15)   Heming, J. D.; Huffman, J. B.; Jones, L. M.; Homa, F. L. *Journal of Virology* **2014**, *88*, 225–236.

(16)   Zeev-Ben-Mordehai, T.; Hagen, C.; Grünewald, K. *Current Opinion in Virology* **2014**, *5*, 42–49.

(17)   Connolly, S. A.; Jardetzky, T. S.; Longnecker, R. *Nature Reviews Microbiology* **2021**, *19*, 110–121.

(18)   Döhner, K.; Wolfstein, A.; Prank, U.; Echeverri, C.; Dujardin, D.; Vallee, R.; Sodeik, B. *Molecular Biology of the Cell* **2002**, *13*, ed. by Goldstein, L. S., 2795–2809.

(19)   Lehman, I.; Boehmer, P. E. *Journal of Biological Chemistry* **1999**, *274*, 28059–28062.

(20)   Homa, F. L.; Brown, J. C. *Reviews in Medical Virology* **1997**, *7*, 107–122.

(21)   Pražák, V.; Mironova, Y.; Vasishtan, D.; Hagen, C.; Laugks, U.; Jensen, Y.; Sanders, S.; Heumann, J. M.; Bosse, J. B.; Klupp, B. G.; Mettenleiter, T. C.; Grange, M.; Grünewald, K. *Nature Microbiology* **2024**, *9*, 1842–1855.

(22)   Roller, R. J.; Baines, J. D. In *Cell Biology of Herpes Viruses*, Osterrieder, K., Ed.; Springer International Publishing: Cham, 2017, pp 143–169.

(23)   Mettenleiter, T. C.; Klupp, B. G.; Granzow, H. *Current Opinion in Microbiology* **2006**, *9*, 423–429.

(24)   Owen, D. J.; Crump, C. M.; Graham, S. C. *Viruses* **2015**, *7*, 5084–5114.

(25)   Rixon, F. J.; Schmid, M. F. *Current Opinion in Virology* **2014**, *5*, 105–110.

(26) Hilbert, B. J.; Hayes, J. A.; Stone, N. P.; Xu, R.-G.; Kelch, B. A. *Nucleic Acids Research* **2017**, *45*, 3591–3605.

(27) Iwaisako, Y.; Fujimuro, M. *Biological and Pharmaceutical Bulletin* **2024**, *47*, 912–916.

(28) Selvarajan Sigamani, S.; Zhao, H.; Kamau, Y. N.; Baines, J. D.; Tang, L. *Journal of Virology* **2013**, *87*, 7140–7148.

(29) Yang, Y.; Yang, P.; Wang, N.; Chen, Z.; Su, D.; Zhou, Z. H.; Rao, Z.; Wang, X. *Protein & Cell* **2020**, *11*, 339–351.

(30) Yang, K.; Baines, J. D. *Journal of Virology* **2006**, *80*, 5733–5739.

(31) Adelman, K.; Salmon, B.; Baines, J. D. *Proceedings of the National Academy of Sciences* **2001**, *98*, 3086–3091.

(32) Davison, A. J. *Annals of the New York Academy of Sciences* **2011**, *1230*, DOI: `10.1111/j.1749-6632.2011.06358.x`.

(33) Tong, L.; Stow, N. D. *Journal of Virology* **2010**, *84*, 321–329.

(34) McVoy, M. A.; Nixon, D. E.; Adler, S. P.; Mocarski, E. S. *Journal of Virology* **1998**, *72*, 48–56.

(35) Deiss, L. P.; Chou, J.; Frenkel, N. *Journal of Virology* **1986**, *59*, 605–618.

(36) Biswas, B.; Kumari, P.; Vivekanandan, P. *ACS infectious diseases* **2018**, *4*, 744–751.

(37) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. *Nucleic Acids Research* **2008**, *36*, 5482–5515.

(38) Monsen, R. C.; Trent, J. O.; Chaires, J. B. *Accounts of Chemical Research* **2022**, *55*, 3242–3252.

(39)   Marchand, A.; Gabelica, V. *Journal of the American Society for Mass Spectrometry* **2014**, *25*, 1146–1154.

(40)   Figueiredo, J.; Mergny, J.-L.; Cruz, C. *Life Sciences* **2024**, *340*, 122481.

(41)   Teng, F.-Y.; Jiang, Z.-Z.; Guo, M.; Tan, X.-Z.; Chen, F.; Xi, X.-G.; Xu, Y. *Cellular and Molecular Life Sciences* **2021**, *78*, 6557–6583.

(42)   Melendez, D. P.; Razonable, R. R. *Infection and Drug Resistance* **2015**, *8*, 269–277.

(43)   Ligat, G.; Cazal, R.; Hantz, S.; Alain, S. *FEMS Microbiology Reviews* **2018**, *42*, 137–145.

(44)   Heming, J. D.; Conway, J. F.; Homa, F. L. *Advances in anatomy, embryology, and cell biology* **2017**, *223*, 119–142.

(45)   Baines, J. D.; Cunningham, C.; Nalwanga, D.; Davison, A. *Journal of Virology* **1997**, *71*, 2666–2673.

(46)   Yu, D.; Sheaffer, A. K.; Tenney, D. J.; Weller, S. K. *Journal of Virology* **1997**, *71*, 2656–2665.

(47)   Beilstein, F.; Higgs, M. R.; Stow, N. D. *Journal of Virology* **2009**, *83*, 8938–8945.

(48)   Guo, P. *Protein & Cell* **2020**, *11*, 311–315.

(49)   Rao, V. B.; Feiss, M. *Annual Review of Genetics* **2008**, *42*, 647–681.

(50)   Hendrix, R. W. *Proceedings of the National Academy of Sciences* **1978**, *75*, 4779–4783.

(51)   Zhao, Z.; Khisamutdinov, E.; Schwartz, C.; Guo, P. *ACS Nano* **2013**, *7*, 4082–4092.

(52)   Schwartz, C.; De Donatis, G. M.; Zhang, H.; Fang, H.; Guo, P. *Virology* **2013**, *443*, 28–39.

(53) Simanjuntak, Y.; Schamoni-Kast, K.; Grün, A.; Uetrecht, C.; Scaturro, P. *Viruses* **2021**, *13*, 668.

(54) Smith, L. M.; Agar, J. N.; Chamot-Rooke, J.; Danis, P. O.; Ge, Y.; Loo, J. A.; Paša-Tolić, L.; Tsybin, Y. O.; Kelleher, N. L.; The Consortium for Top-Down Proteomics *Science Advances* **2021**, *7*, eabk0734.

(55) O'Reilly, F. J.; Rappsilber, J. *Nature Structural & Molecular Biology* **2018**, *25*, 1000–1008.

(56) Heck, A. J. R. *Nature Methods* **2008**, *5*, 927–933.

(57) Gross, J. H., *Mass Spectrometry*; Springer International Publishing: Cham, 2017.

(58) Glish, G. L.; Vachet, R. W. *Nature Reviews Drug Discovery* **2003**, *2*, 140–150.

(59) Maher, S.; Jjunju, F. P. M.; Taylor, S. *Reviews of Modern Physics* **2015**, *87*, 113–135.

(60) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198–207.

(61) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64–71.

(62) Leney, A. C.; Heck, A. J. R. *Journal of the American Society for Mass Spectrometry* **2017**, *28*, 5–13.

(63) Bakhtiari, M.; Konermann, L. *The Journal of Physical Chemistry B* **2019**, *123*, 1784–1796.

(64) Webb, I. K. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2022**, *1870*, 140732.

(65) Schmidt, A.; Karas, M.; Dülcks, T. *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 492–500.

(66) Yamashita, M.; Fenn, J. B. *The Journal of Physical Chemistry* **1984**, *88*, 4451–4459.

(67) Nguyen, S.; Fenn, J. B. *Proceedings of the National Academy of Sciences* **2007**, *104*, 1111–1117.

(68) Hogan, C. J.; Kettleson, E. M.; Ramaswami, B.; Chen, D.-R.; Biswas, P. *Analytical Chemistry* **2006**, *78*, 844–852.

(69) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. *Analytical Chemistry* **2013**, *85*, 2–9.

(70) Gomez, A. In *Experimental Heat Transfer, Fluid Mechanics and Thermodynamics 1993*, Kelleher, M. D., Sreenivasan, K. R., Shah, R. K., Joshi, Y., Eds.; Elsevier Series in Thermal and Fluid Sciences; Elsevier: Amsterdam, 1993, pp 270–282.

(71) Wilm, M. *Molecular & Cellular Proteomics : MCP* **2011**, *10*, M111.009407.

(72) Konermann, L. *Journal of the American Society for Mass Spectrometry* **2017**, *28*, 1827–1835.

(73) Kafader, J. O.; Melani, R. D.; Schachner, L. F.; Ives, A. N.; Patrie, S. M.; Kelleher, N. L.; Compton, P. D. *Journal of the American Society for Mass Spectrometry* **2020**, *31*, 574–581.

(74) Susa, A. C.; Xia, Z.; Tang, H. Y. H.; Tainer, J. A.; Williams, E. R. *Journal of the American Society for Mass Spectrometry* **2017**, *28*, 332–340.

(75) Zhang, H.; Cui, W.; Gross, M. L.; Blankenship, R. E. *FEBS Letters* **2013**, *587*, 1012–1020.

(76) A. Abaye, D.; A. Agbo, I.; V. Nielsen, B. *RSC Advances* **2021**, *11*, 20355–20369.

(77) Townsend, J. A.; Keener, J. E.; Miller, Z. M.; Prell, J. S.; Marty, M. T. *Analytical Chemistry* **2019**, *91*, 14765–14772.

(78)   Rolland, A. D.; Prell, J. S. *Chemical Reviews* **2022**, *122*, 7909–7951.

(79)   Chen, W.; Ding, Z.; Zang, Y.; Liu, X. *Journal of Proteome Research* **2023**, *22*, 3178–3189.

(80)   Zhou, T. et al. *Veterinary Research* **2022**, *53*, 93.

(81)   Reid, D. J.; Thibert, S.; Zhou, M. *Protein Science* **2023**, *32*, e4612.

(82)   Kruve, A.; Kaupmees, K. *Journal of the American Society for Mass Spectrometry* **2017**, *28*, 887–894.

(83)   Snijder, J.; Rose, R. J.; Veesler, D.; Johnson, J. E.; Heck, A. J. R. *Angewandte Chemie International Edition* **2013**, *52*, 4020–4023.

(84)   Weiss, V. U.; Pogan, R.; Zoratto, S.; Bond, K. M.; Boulanger, P.; Jarrold, M. F.; Lyktey, N.; Pahl, D.; Puffler, N.; Schelhaas, M.; Selivanovitch, E.; Uetrecht, C.; Allmaier, G. *Analytical and Bioanalytical Chemistry* **2019**, *411*, 5951–5962.

(85)   Čaval, T.; Buettner, A.; Haberger, M.; Reusch, D.; Heck, A. J. *Journal of the American Society for Mass Spectrometry* **2021**, *32*, 2099–2104.

(86)   Lössl, P.; Snijder, J.; Heck, A. J. R. *Journal of the American Society for Mass Spectrometry* **2014**, *25*, 906–917.

(87)   Greaves, J.; Roboz, J., *Mass Spectrometry for the Novice*; Taylor & Francis Group, LLC: 2014.

(88)   Lorenzen, K.; Versluis, C.; van Duijn, E.; van den Heuvel, R. H. H.; Heck, A. J. R. *International Journal of Mass Spectrometry* **2007**, *268*, 198–206.

(89)   Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T. *Current Opinion in Chemical Biology* **2018**, *42*, 93–100.

(90) Pagel, K.; Hyung, S.-J.; Ruotolo, B. T.; Robinson, C. V. *Analytical Chemistry* **2010**, *82*, 5363–5372.

(91) Hayes, R. N.; Gross, M. L. In *Methods in Enzymology*; Mass Spectrometry, Vol. 193; Academic Press: 1990, pp 237–263.

(92) Mitchell Wells, J.; McLuckey, S. A. In *Methods in Enzymology*; Elsevier: 2005; Vol. 402, pp 148–185.

(93) Beardsley, R. L.; Jones, C. M.; Galhena, A. S.; Wysocki, V. H. *Analytical Chemistry* **2009**, *81*, 1347–1356.

(94) Paul, W. *Angewandte Chemie International Edition in English* **1990**, *29*, 739–748.

(95) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *Journal of Mass Spectrometry* **2001**, *36*, 849–865.

(96) Boesl, U. *Mass Spectrometry Reviews* **2017**, *36*, 86–109.

(97) Van Den Heuvel, R. H. H.; Van Duijn, E.; Mazon, H.; Synowsky, S. A.; Lorenzen, K.; Versluis, C.; Brouns, S. J. J.; Langridge, D.; Van Der Oost, J.; Hoyes, J.; Heck, A. J. R. *Analytical Chemistry* **2006**, *78*, 7473–7483.

(98) Pogan, R.; Schneider, C.; Reimer, R.; Hansman, G.; Uetrecht, C. *Journal of Physics: Condensed Matter* **2018**, *30*, 064006.

(99) Makarov, A. *Analytical Chemistry* **2000**, *72*, 1156–1162.

(100) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *Journal of Mass Spectrometry* **2005**, *40*, 430–443.

(101) Zubarev, R. A.; Makarov, A. *Analytical Chemistry* **2013**, *85*, 5288–5296.

(102)   Tamara, S.; Den Boer, M. A.; Heck, A. J. R. *Chemical Reviews* **2022**, *122*, 7269–7326.

(103)   Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Analytical Chemistry* **2006**, *78*, 2113–2120.

(104)   L. Fort, K.; Waterbeemd, M. v. d.; Boll, D.; Reinhardt-Szyba, M.; E. Belov, M.; Sasaki, E.; Zschoche, R.; Hilvert, D.; A. Makarov, A.; R. Heck, A. J. *Analyst* **2018**, *143*, 100–105.

(105)   Eliuk, S.; Makarov, A. *Annual Review of Analytical Chemistry* **2015**, *8*, 61–80.

(106)   Scientific, T., *Exactive Series Operating Manual*; Thermo Fisher: 2017.

(107)   Limited, M. U., *Q-Tof 2 User's Guide*; Micromass; Vol. 1.

(108)   Chen, R. *Biotechnology Advances* **2012**, *30*, 1102–1107.

(109)   Jarvis, D. L. In *The Baculoviruses*, Miller, L. K., Ed.; Springer US: Boston, MA, 1997, pp 389–431.

(110)   Luckow, V. A. *Current Opinion in Biotechnology* **1993**, *4*, 564–572.

(111)   Kost, T. A.; Condreay, J. P.; Jarvis, D. L. *Nature Biotechnology* **2005**, *23*, 567–575.

(112)   Nettleship, J. E.; Assenberg, R.; Diprose, J. M.; Rahman-Huq, N.; Owens, R. J. *Journal of Structural Biology* **2010**, *172*, 55–65.

(113)   Labrou, N. E. In *Protein Downstream Processing*, Labrou, N. E., Ed.; Humana Press: Totowa, NJ, 2014; Vol. 1129, pp 3–10.

(114)   Arora, S.; Saxena, V.; Ayyar, B. V. *Methods* **2017**, *116*, 84–94.

(115) Urh, M.; Simpson, D.; Zhao, K. In *Methods in Enzymology*, Burgess, R. R.,
       Deutscher, M. P., Eds.; Guide to Protein Purification, 2nd Edition, Vol. 463;
       Academic Press: 2009, pp 417–438.

(116) Francis, D. M.; Page, R. *Current Protocols in Protein Science* **2010**, *61*, DOI:
       `10.1002/0471140864.ps0524s61`.

(117) Hanahan, D. *Journal of Molecular Biology* **1983**, *166*, 557–580.

(118) Kaur, J.; Kumar, A.; Kaur, J. *International Journal of Biological Macromolecules*
       2018, *106*, 803–822.

(119) Fahnert, B.; Lilie, H.; Neubauer, P. In *Physiological Stress Responses in Bioprocesses:*
       -/-; Springer: Berlin, Heidelberg, 2004, pp 93–142.

(120) Chalmers, J. J. *Cytotechnology* **1996**, *20*, 163–171.

(121) King, L., *The Baculovirus Expression System: A laboratory guide*, Google-Books-ID:
       4HbwCAAAQBAJ; Springer Science & Business Media: 2012.

(122) King, L. A.; Hitchman, R.; Possee, R. D. In *Baculovirus and Insect Cell Expression*
       Protocols, Murhammer, D. W., Ed.; Humana Press: Totowa, NJ, 2007, pp 77–93.

(123) Malm, M.; Saghaleyni, R.; Lundqvist, M.; Giudici, M.; Chotteau, V.; Field,
       R.; Varley, P. G.; Hatton, D.; Grassi, L.; Svensson, T.; Nielsen, J.; Rockberg, J.
       Scientific Reports **2020**, *10*, 18996.

(124) Longo, P. A.; Kavran, J. M.; Kim, M.-S.; Leahy, D. J. In *Methods in Enzymology*,
       Lorsch, J., Ed.; Laboratory Methods in Enzymology: DNA, Vol. 529; Academic
       Press: 2013, pp 227–240.

(125) Zhao, X.; Li, G.; Liang, S. *Journal of Analytical Methods in Chemistry* **2013**, *2013*,
       581093.

(126) Skerra, A.; Schmidt, T. G. M. *Biomolecular Engineering* **1999**, *16*, 79–86.

(127) Spriestersbach, A.; Kubicek, J.; Schäfer, F.; Block, H.; Maertens, B. In *Methods in Enzymology*, Lorsch, J. R., Ed.; Laboratory Methods in Enzymology: Protein Part D, Vol. 559; Academic Press: 2015, pp 1–15.

(128) Block, H.; Maertens, B.; Spriestersbach, A.; Brinker, N.; Kubicek, J.; Fabis, R.; Labahn, J.; Schäfer, F. In *Methods in Enzymology*, Burgess, R. R., Deutscher, M. P., Eds.; Guide to Protein Purification, 2nd Edition, Vol. 463; Academic Press: 2009, pp 439–473.

(129) Barth, H. G.; Boyes, B. E.; Jackson, C. *Analytical Chemistry* **1996**, *68*, 445–466.

(130) Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. *Science* **1989**, *243*, 85–88.

(131) Schmidt, T. G.; Skerra, A. *Nature Protocols* **2007**, *2*, 1528–1535.

(132) Rapid Online Buffer Exchange: A Method for Screening of Proteins, Protein Complexes, and Cell Lysates by Native Mass Spectrometry. `https://chemrxiv.org/engage/chemrxiv/article-details/60c7453dee301c4b17c7929a`.

(133) VanAernum, Z. L.; Busch, F.; Jones, B. J.; Jia, M.; Chen, Z.; Boyken, S. E.; Sahasrabuddhe, A.; Baker, D.; Wysocki, V. H. *Nature Protocols* **2020**, *15*, 1132–1157.

(134) Van Dyck, J. F.; Burns, J. R.; Le Huray, K. I. P.; Konijnenberg, A.; Howorka, S.; Sobott, F. *Nature Communications* **2022**, *13*, 3610.

(135) Hernández, H.; Robinson, C. V. *Nature Protocols* **2007**, *2*, 715–726.

(136) Prodanov, M.; Garrido, I.; Vacas, V.; Lebrón-Aguilar, R.; Dueñas, M.; Gómez-Cordovés, C.; Bartolomé, B. *Analytica Chimica Acta* **2008**, *609*, 241–251.

(137) Donnelly, D. P. et al. *Nature Methods* **2019**, *16*, 587–594.

(138) Yin, H.; Killeen, K.; Brennen, R.; Sobek, D.; Werlich, M.; Van De Goor, T. *Analytical Chemistry* **2005**, *77*, 527–533.

(139) Flick, T. G.; Cassou, C. A.; Chang, T. M.; Williams, E. R. *Analytical Chemistry* **2012**, *84*, 7511–7517.

(140) Susa, A. C.; Xia, Z.; Williams, E. R. *Angewandte Chemie International Edition* **2017**, *56*, 7912–7915.

(141) Saikusa, K.; Kato, D.; Nagadoi, A.; Kurumizaka, H.; Akashi, S. *Journal of the American Society for Mass Spectrometry* **2020**, *31*, 711–718.

(142) Sanders, S.; Jensen, Y.; Reimer, R.; Bosse, J. B. In *Advances in Virus Research*, Finke, S., Ushakov, D., Eds.; Imaging in Virus Research, Vol. 116; Academic Press: 2023, pp 45–88.

(143) Straight, A. F. In *Methods in Cell Biology*; Digital Microscopy, 3rd Edition, Vol. 81; Academic Press: 2007, pp 93–113.

(144) Niwa, H.; Inouye, S.; Hirano, T.; Matsuno, T.; Kojima, S.; Kubota, M.; Ohashi, M.; Tsuji, F. I. *Proceedings of the National Academy of Sciences* **1996**, *93*, 13617–13622.

(145) Sample, V.; Newman, R. H.; Zhang, J. *Chemical Society Reviews* **2009**, *38*, 2852–2864.

(146) De Oliveira, A. P.; Glauser, D. L.; Laimbacher, A. S.; Strasser, R.; Schraner, E. M.; Wild, P.; Ziegler, U.; Breakefield, X. O.; Ackermann, M.; Fraefel, C. *Journal of Virology* **2008**, *82*, 4974–4990.

(147) Cormack, B. P.; Valdivia, R. H.; Falkow, S. *Gene* **1996**, *173*, 33–38.

(148)   Hogue, I. B.; Bosse, J. B.; Engel, E. A.; Scherer, J.; Hu, J.-R.; Del Rio, T.; Enquist, L. W. *Viruses* **2015**, *7*, 5933–5961.

(149)   Costantini, L. M.; Fossati, M.; Francolini, M.; Snapp, E. L. *Traffic (Copenhagen, Denmark)* **2012**, *13*, 643–649.

(150)   Schmitz, S., *Der Experimentator: Zellkultur*, 4. Aufl. 2020; Experimentator; Springer Berlin Heidelberg: Berlin, Heidelberg, 2020.

(151)   *Insect cell culture engineering*; Goosen, M. F. A., Daugulis, A. J., Faulkner, P., Eds.; Bioprocess technology v. 17; CRC Press: Place of publication not identified, 2019.

(152)   Vlak, J. M.; Gooijer, C. D.; Tramper, J.; Miltenburger, H. G., *Insect Cell Culture: Fundamental and Applied Aspects*; Current Applications of Cell Culture Engineering 2; Kluwer Academic Publishers: Dordrecht, 2002.

(153)   Effect of coat-protein concentration on the self-assembly of bacteriophage MS2 capsids around RNA. `https://arxiv.org/abs/2307.04171`.

(154)   Blancas-Mejía, L. M.; Misra, P.; Ramirez-Alvarado, M. *Biochemistry* **2017**, *56*, 757–766.

(155)   Evans, D. R.; Romero, J. K.; Westoby, M. In *Methods in Enzymology*; Elsevier: 2009; Vol. 463, pp 97–120.

(156)   Pohl, T. In *Methods in Enzymology*, Deutscher, M. P., Ed.; Guide to Protein Purification, Vol. 182; Academic Press: 1990, pp 68–83.

(157)   Fast tracking native mass spectrometry: Skipping over buffer exchange. `https://www.biorxiv.org/content/10.1101/2025.02.22.639503v1`. (accessed on 2025-12-23).

(158)   Deslignière, E.; Ley, M.; Bourguet, M.; Ehkirch, A.; Botzanowski, T.; Erb, S.; Hernandez-Alba, O.; Cianférani, S. *International Journal of Mass Spectrometry* **2021**, *461*, 116502.

(159)   Vimer, S.; Ben-Nissan, G.; Sharon, M. *Nature Protocols* **2020**, *15*, 236–265.

(160)   Busch, F.; VanAernum, Z. L.; Lai, S. M.; Gopalan, V.; Wysocki, V. H. *Biochemistry* **2021**, *60*, 1876–1884.

(161)   Skerra, A.; Schmidt, T. G. M. In *Methods in Enzymology*; Applications of Chimeric Genes and Hybrid Proteins Part A: Gene Expression and Protein Purification, Vol. 326; Academic Press: 2000, pp 271–304.

(162)   Schmidt, T. G. M.; Batz, L.; Bonet, L.; Carl, U.; Holzapfel, G.; Kiem, K.; Matulewicz, K.; Niermeier, D.; Schuchardt, I.; Stanar, K. *Protein Expression and Purification* **2013**, *92*, 54–61.

(163)   Dobro, M. J.; Melanson, L. A.; Jensen, G. J.; McDowall, A. W. In *Methods in Enzymology*, Jensen, G. J., Ed.; Cryo-EM Part A Sample Preparation and Data Collection, Vol. 481; Academic Press: 2010, pp 63–82.

(164)   Earl, L. A.; Falconieri, V.; Milne, J. L.; Subramaniam, S. *Current Opinion in Structural Biology* **2017**, *46*, 71–78.

(165)   Salmon, B.; Nalwanga, D.; Fan, Y.; Baines, J. D. *Journal of Virology* **1999**, *73*, 8338–8348.

(166)   Dolan, A.; Arbuckle, M.; McGeoch, D. J. *Virus Research* **1991**, *20*, 97–104.

(167)   Yu, D.; Weller, S. K. *Virology* **1998**, *243*, 32–44.

(168) Dyachenko, A.; Wang, G.; Belov, M.; Makarov, A.; De Jong, R. N.; Van Den Bremer, E. T. J.; Parren, P. W. H. I.; Heck, A. J. R. *Analytical Chemistry* **2015**, *87*, 6095–6102.

(169) Shu, H.; Zhang, R.; Xiao, K.; Yang, J.; Sun, X. *Biomolecules* **2022**, *12*, 648.

(170) Oganesian, L.; Bryan, T. M. *BioEssays* **2007**, *29*, 155–165.

(171) Onel, B.; Lin, C.; Yang, D. *Science China Chemistry* **2014**, *57*, 1605–1614.

(172) Ghosh, A.; Largy, E.; Gabelica, V. *Nucleic Acids Research* **2021**, *49*, 2333–2345.

(173) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. *Nature Chemistry* **2014**, *6*, 281–294.

(174) Narang, D.; Lento, C.; J. Wilson, D. *Biomedicines* **2020**, *8*, 224.

(175) Jumper, J. et al. *Nature* **2021**, *596*, 583–589.

(176) Abramson, J. et al. *Nature* **2024**, *630*, 493–500.

(177) Bertoline, L. M. F.; Lima, A. N.; Krieger, J. E.; Teixeira, S. K. *Frontiers in Bioinformatics* **2023**, *3*, DOI: `10.3389/fbinf.2023.1120370`.

(178) Jones, D. T.; Thornton, J. M. *Nature Methods* **2022**, *19*, 15–20.

(179) Wayment-Steele, H. K.; Ojoawo, A.; Otten, R.; Apitz, J. M.; Pitsawong, W.; Hömberger, M.; Ovchinnikov, S.; Colwell, L.; Kern, D. *Nature* **2024**, *625*, 832–839.

(180) Zhang, Y.; Skolnick, J. *Proteins: Structure, Function, and Bioinformatics* **2004**, *57*, 702–710.

(181) Xu, J.; Zhang, Y. *Bioinformatics* **2010**, *26*, 889–895.

(182) Desai, D.; Kantliwala, S. V.; Vybhavi, J.; Ravi, R.; Patel, H.; Patel, J. *Cureus* **2024**, *16*, e63646.

(183) Bayliss, G. J.; Marsden, H. S.; Hay, J. *Virology* **1975**, *68*, 124–134.

(184) Taylor, T. J.; Knipe, D. M. *Journal of Virology* **2004**, *78*, 5856–5866.

(185) Asor, R.; Kukura, P. *Current Opinion in Chemical Biology* **2022**, *68*, 102132.

(186) Jecklin, M. C.; Touboul, D.; Bovet, C.; Wortmann, A.; Zenobi, R. *Journal of the American Society for Mass Spectrometry* **2008**, *19*, 332–343.

(187) Klykov, O.; Kopylov, M.; Carragher, B.; Heck, A. J.; Noble, A. J.; Scheltema, R. A. *Molecular Cell* **2022**, *82*, 285–303.

(188) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. In *The Proteomics Protocols Handbook*; Humana Press: 2005.

(189) Daly, L. A.; Clarke, C. J.; Po, A.; Oswald, S. O.; Eyers, C. E. *Chemical Communications (Cambridge, England)*, *59*, 11484–11499.

(190) Yang, Y.; Barendregt, A.; Kamerling, J. P.; Heck, A. J. R. *Analytical Chemistry* **2013**, *85*, 12037–12045.

(191) Füssl, F.; Criscuolo, A.; Cook, K.; Scheffler, K.; Bones, J. *Journal of Proteome Research* **2019**, *18*, 3689–3702.

(192) Wang, D.; Liang, Y.; Xu, D. *Bioinformatics (Oxford, England)* **2019**, *35*, 2386–2394.

(193) Wang, D.; Zeng, S.; Xu, C.; Qiu, W.; Liang, Y.; Joshi, T.; Xu, D. *Bioinformatics (Oxford, England)* **2017**, *33*, 3909–3916.

(194) Wang, D.; Liu, D.; Yuchi, J.; He, F.; Jiang, Y.; Cai, S.; Li, J.; Xu, D. *Nucleic Acids Research* **2020**, *48*, W140–W146.

(195) Mann, M.; Ong, S.-E.; Grønborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. *Trends in Biotechnology* **2002**, *20*, 261–268.

(196) Huffmaster, N. J.; Sollars, P. J.; Richards, A. L.; Pickard, G. E.; Smith, G. A. *Proceedings of the National Academy of Sciences* **2015**, *112*, 12818–12823.

(197) Bell, C.; Desjardins, M.; Thibault, P.; Radtke, K. *Journal of Proteome Research* **2013**, *12*, 1820–1829.

(198) Wu, D.; Struwe, W. B. In *Glycoprotein Analysis*, Struwe, W. B., Ed.; Royal Society of Chemistry: 2024, pp 260–278.

(199) Gabelica, V.; Pauw, E. D. *Mass Spectrometry Reviews* **2005**, *24*, 566–587.

(200) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. *Nucleic Acids Research* **2025**, *53*, D1516–D1525.

(201) Albright, B. S.; Kosinski, A.; Szczepaniak, R.; Cook, E. A.; Stow, N. D.; Conway, J. F.; Weller, S. K. *Journal of Virology* **2015**, *89*, ed. by Sandri-Goldin, R. M., 443–453.

(202) Gardner, M. R.; Glaunsinger, B. A. *Journal of Virology* **2018**, *92*, ed. by Sandri-Goldin, R. M., e00840–18.

(203) Didychuk, A. L.; Gates, S. N.; Gardner, M. R.; Strong, L. M.; Martin, A.; Glaunsinger, B. A. *eLife* **2021**, *10*, ed. by Carter, A. P.; Wolberger, C.; Graham, S. C., e62261.

(204) Muller, C.; Alain, S.; Gourin, C.; Baumert, T. F.; Ligat, G.; Hantz, S. *Viruses* **2021**, *13*, 1638.

(205) Harmening, S.; Bogdanow, B.; Wagner, K.; Liu, F.; Messerle, M.; Borst, E. M. *Journal of Virology* **2025**, ed. by Cliffe, A. R., e02201–24.

(206) Darby, N. J.; Creighton, T. E., *Protein structure*, Repr; In focus; IRL Pr. at Oxford Univ. Pr: Oxford, 1995.

(207) Sanders, S.; Jensen, Y.; Reimer, R.; Bosse, J. B. In *Advances in Virus Research*; Elsevier: 2023; Vol. 116, pp 45–88.

(208) Ettinger, A.; Wittmann, T. In *Methods in Cell Biology*; Elsevier: 2014; Vol. 123, pp 77–94.

(209) Chudakov, D. M.; Matz, M. V.; Lukyanov, S.; Lukyanov, K. A. *Physiological Reviews* **2010**, *90*, 1103–1163.

(210) Wiedenmann, J.; Oswald, F.; Nienhaus, G. U. *IUBMB Life* **2009**, *61*, 1029–1042.

(211) Day, R. N.; Davidson, M. W. *Chemical Society Reviews* **2009**, *38*, 2887–2921.

(212) Mallis, C. S.; Zheng, X.; Qiu, X.; McCabe, J. W.; Shirzadeh, M.; Lyu, J.; Laganowsky, A.; Russell, D. H. *International Journal of Mass Spectrometry* **2020**, *458*, 116451.

(213) Gottesman, S. *Annual Review of Genetics* **1996**, *30*, 465–506.

(214) Rosano, G. L.; Ceccarelli, E. A. *Frontiers in Microbiology* **2014**, *5*, DOI: `10.3389/fmicb.2014.00172`.

(215) Muneeruddin, K.; Thomas, J. J.; Salinas, P. A.; Kaltashov, I. A. *Analytical Chemistry* **2014**, *86*, 10692–10699.

(216) Wang, W.; Kitova, E. N.; Klassen, J. S. *Analytical Chemistry* **2005**, *77*, 3060–3071.

(217) Lu, Y.; Liu, H.; Saer, R.; Zhang, H.; Meyer, C.; Li, V.; Shi, L.; King, J. D.; Gross, M. L.; Blankenship, R. E. *Biochemistry* **2017**, *56*, 160–166.

(218) Shaner, N. C.; Lambert, G. G.; Chammas, A.; Ni, Y.; Cranfill, P. J.; Baird, M. A.; Sell, B. R.; Allen, J. R.; Day, R. N.; Israelsson, M.; Davidson, M. W.; Wang, J. *Nature Methods* **2013**, *10*, 407–409.

(219) Cranfill, P. J.; Sell, B. R.; Baird, M. A.; Allen, J. R.; Lavagnino, Z.; de Gruiter, H. M.; Kremers, G.-J.; Davidson, M. W.; Ustione, A.; Piston, D. W. *Nature Methods* **2016**, *13*, 557–562.

(220) Hoi, H.; Howe, E. S.; Ding, Y.; Zhang, W.; Baird, M. A.; Sell, B. R.; Allen, J. R.; Davidson, M. W.; Campbell, R. E. *Chemistry & Biology* **2013**, *20*, 1296–1304.

(221) Paez-Segala, M. G.; Sun, M. G.; Shtengel, G.; Viswanathan, S.; Baird, M. A.; Macklin, J. J.; Patel, R.; Allen, J. R.; Howe, E. S.; Piszczek, G.; Hess, H. F.; Davidson, M. W.; Wang, Y.; Looger, L. L. *Nature Methods* **2015**, *12*, 215–218.

(222) Bajar, B. T. et al. *Nature Methods* **2016**, *13*, 993–996.

(223) Bindels, D. S.; Haarbosch, L.; van Weeren, L.; Postma, M.; Wiese, K. E.; Mastop, M.; Aumonier, S.; Gotthard, G.; Royant, A.; Hink, M. A.; Gadella, T. W. J. *Nature Methods* **2017**, *14*, 53–56.

(224) Manna, P.; Hung, S.-T.; Mukherjee, S.; Friis, P.; Simpson, D. M.; Lo, M. N.; Palmer, A. E.; Jimenez, R. *Integrative Biology* **2018**, *10*, 516–526.

(225) Costantini, L. M.; Baloban, M.; Markwardt, M. L.; Rizzo, M. A.; Guo, F.; Verkhusha, V. V.; Snapp, E. L. *Nature Communications* **2015**, *6*, 7670.

(226) Meiresonne, N. Y.; Consoli, E.; Mertens, L. M.; Chertkova, A. O.; Goedhart, J.; Den Blaauwen, T. *Molecular Microbiology* **2019**, *111*, 1025–1038.

(227)   Zacharias, D. A.; Violin, J. D.; Newton, A. C.; Tsien, R. Y. *Science (New York, N.Y.)* **2002**, *296*, 913–916.

(228)   Kopicki, J.-D.; Saikia, A.; Niebling, S.; Günther, C.; Anjanappa, R.; Garcia-Alai, M.; Springer, S.; Uetrecht, C. *Communications Biology* **2022**, *5*, 488.

(229)   Zhang, S.; Van Pelt, C. K.; Wilson, D. B. *Analytical Chemistry* **2003**, *75*, 3010–3018.

(230)   Ugo, P.; Marafini, P.; Meneghello, M., *Bioanalytical Chemistry: From Biomolecular Recognition to Nanobiosensing*; De Gruyter: 2021.

(231)   Smith, M. C.; Gestwicki, J. E. *Expert Reviews in Molecular Medicine* **2012**, *14*, e16.

(232)   Granvogl, B.; Plöscher, M.; Eichacker, L. A. *Analytical and Bioanalytical Chemistry* **2007**, *389*, 991–1002.

(233)   Strohalm, M.; Kavan, D.; Novák, P.; Volný, M.; Havlíček, V. *Analytical Chemistry* **2010**, *82*, 4648–4651.

(234)   Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V. *Analytical Chemistry* **2015**, *87*, 4370–4376.

(235)   Morgner, N.; Robinson, C. V. *Analytical Chemistry* **2012**, *84*, 2939–2948.

Table A1: Characteristics of the HSV-1 *pac* sequences.

species	ssDNA sequence	Theoretical mass in Da
pac 1	GGGGGGTGTGTTTTGGGGGGGG	7088.96
pac 2	CCGCCGCCCGCCTTTTTT GCGCGCGCGCGCGC	9730.78
pac 1 mG	GCGGCGTGTGTTTTGGCGGCGG	6928.84
pac 1 mT	GGGGGGTGTGTTGTGGGGGGGG	7113.98

Table A2: Theoretical masses and extinction coefficient of the terminase complexes.

Species	Theoretical mass in Da
His-TEV-UL15	83886.64
Ul28	85635.13
Ul33	14451.37
HTC	183973.14
HTC_2	367946.28
HTC_3	551919.42
HTC_6	1103838.84
Ext. Co. of HTC	$114\,290\,\text{L} \cdot \text{mol}^{-}1 \cdot \text{cm}^{-}1$

Table A3: Predicted PTMs sites on the sequence of the terminase complex by MusiteDeep[192–194].

ID	Position	Residue	PTMscores	Cutoff=0.5
terminase	2	S	Phosphoserine: 0.106; O-linked glycosylation: 0.114	None
terminase	3	Y	Phosphotyrosine: 0.208	None
terminase	4	Y	Phosphotyrosine: 0.215	None
terminase	14	Y	Phosphotyrosine: 0.241	None
terminase	17	P	Hydroxyproline: 0.093	None

Table A3 – continued from previous page

ID	Position	Residue	PTMscores	Cutoff=0.5
terminase	18	T	Phosphothreonine: 0.082; O-linked glycosylation: 0.174	None
terminase	19	T	Phosphothreonine: 0.067; O-linked glycosylation: 0.106	None
terminase	42	K	Ubiquitination: 0.312; SUMOylation: 0.068; N6-acetyllysine: 0.141; Methyllysine: 0.041; Hydroxylysine: 0.104	None
terminase	47	K	Ubiquitination: 0.487; SUMOylation: 0.071; N6-acetyllysine: 0.142; Methyllysine: 0.077; Hydroxylysine: 0.056	None
terminase	75	T	Phosphothreonine: 0.854; O-linked glycosylation: 0.517	Phosphothreonine: 0.854, O-linked glycosylation: 0.517

Table A3 – continued from previous page

ID	Position	Residue	PTMscores	Cutoff=0.5
terminase	77	K	Ubiquitination: 0.4; SUMOylation: 0.039; N6-acetyllysine: 0.213; Methyllysine: 0.1; Hydroxylysine: 0.022	None
terminase	191	K	Ubiquitination: 0.355; SUMOylation: 0.054; N6-acetyllysine: 0.725; Methyllysine: 0.038; Hydroxylysine: 0.032	N6-acetyllysine: 0.725
terminase	341	K	Ubiquitination: 0.277; SUMOylation: 0.901; N6-acetyllysine: 0.184; Methyllysine: 0.082; Hydroxylysine: 0.242	SUMOylation: 0.901
terminase	542	S	Phosphoserine: 0.352; O-linked glycosylation: 0.613	O-linked glycosylation: 0.613
terminase	672	K	Ubiquitination: 0.235; SUMOylation: 0.04; N6-acetyllysine: 0.593; Methyllysine: 0.088; Hydroxylysine: 0.04	N6-acetyllysine: 0.593

Table A3 – continued from previous page

ID	Position	Residue	PTMscores	Cutoff=0.5
terminase	676	S	Phosphoserine: 0.479; O-linked glycosylation: 0.09	None
terminase	800	R	Methylarginine: 0.828	Methylarginine: 0.828
terminase	807	K	Ubiquitination: 0.631; SUMOylation: 0.044; N6-acetyllysine: 0.389; Methyllysine: 0.279; Hydroxylysine: 0.272	Ubiquitination: 0.631
terminase	1009	S	Phosphoserine: 0.679; O-linked glycosylation: 0.095	Phosphoserine: 0.679
terminase	1037	S	Phosphoserine: 0.897; O-linked glycosylation: 0.486	Phosphoserine: 0.897
terminase	1091	N	N-linked glycosylation: 0.897	N-linked glycosylation: 0.897
terminase	1165	T	Phosphothreonine: 0.686; O-linked glycosylation: 0.099	Phosphothreonine: 0.686
terminase	1269	R	Methylarginine: 0.034	None

Table A3 – continued from previous page

ID	Position	Residue	PTMscores	Cutoff=0.5
terminase	1270	S	Phosphoserine: 0.67; O-linked glycosylation: 0.057	Phosphoserine: 0.67
terminase	1367	N	N-linked glycosylation: 0.034	None
terminase	1400	K	Ubiquitination: 0.367; SUMOylation: 0.805; N6-acetyllysine: 0.181; Methyllysine: 0.106; Hydroxylysine: 0.086	SUMOylation: 0.805
terminase	1555	T	Phosphothreonine: 0.71; O-linked glycosylation: 0.095	Phosphothreonine: 0.71
terminase	1620	S	Phosphoserine: 0.794; O-linked glycosylation: 0.099	Phosphoserine: 0.794
terminase	1640	T	Phosphothreonine: 0.829; O-linked glycosylation: 0.456	Phosphothreonine: 0.829
terminase	1672	R	Methylarginine: 0.068	None
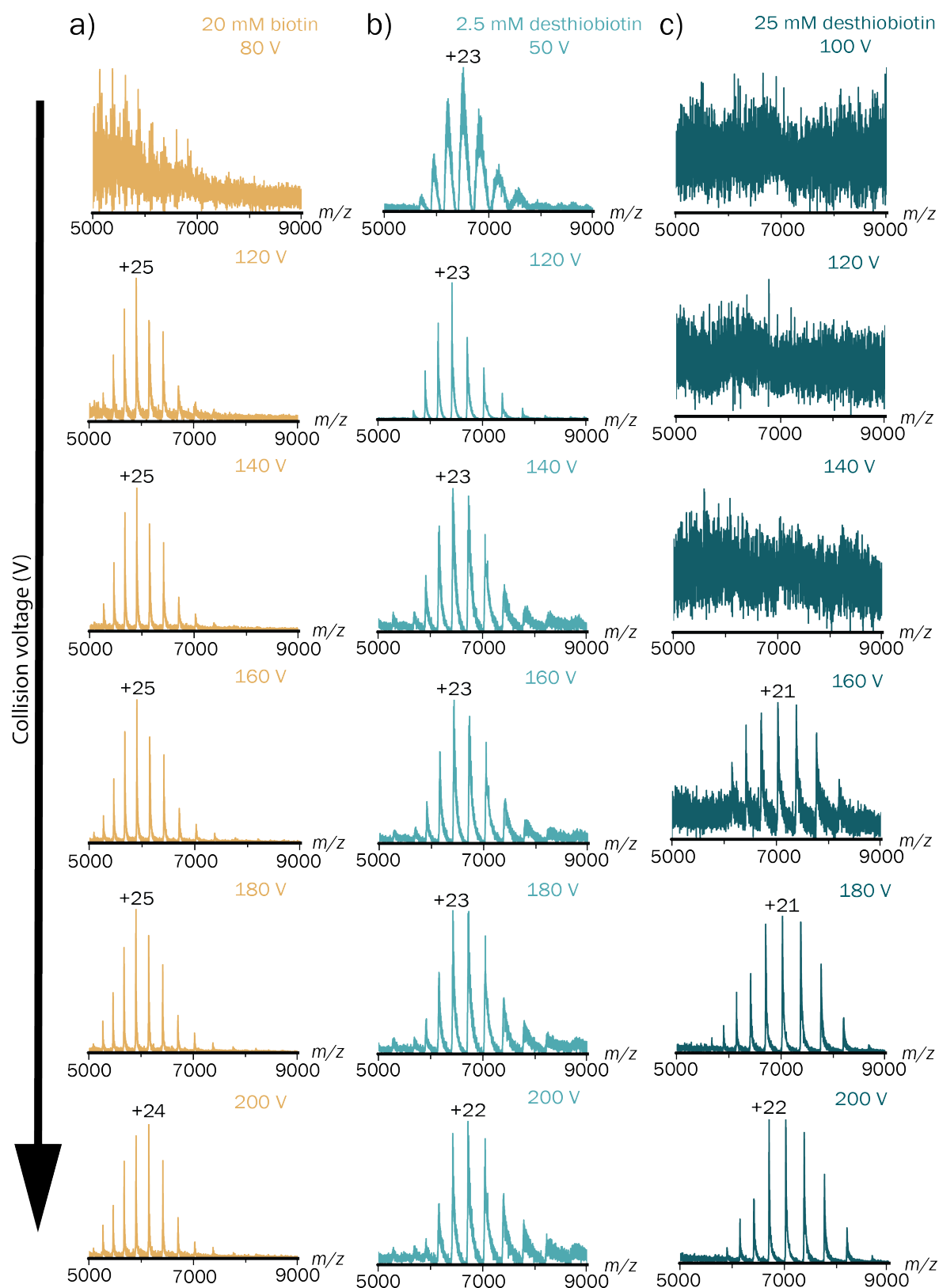
Figure A1: Spectra of the tetrameric ADH complex (148 kDa, 2.5 μM) in 150 mM ammonium acetate at pH 8 spiked with typical amounts of eluent at different collision voltages, increasing from top to bottom, on the Q-TOF 2. a) ADH and 20 mM biotin at 80-200 V. b) ADH and 2.5 mM desthiobiotin at 50-200 V. c) ADH and 25 mM desthiobiotin at 100-200 V
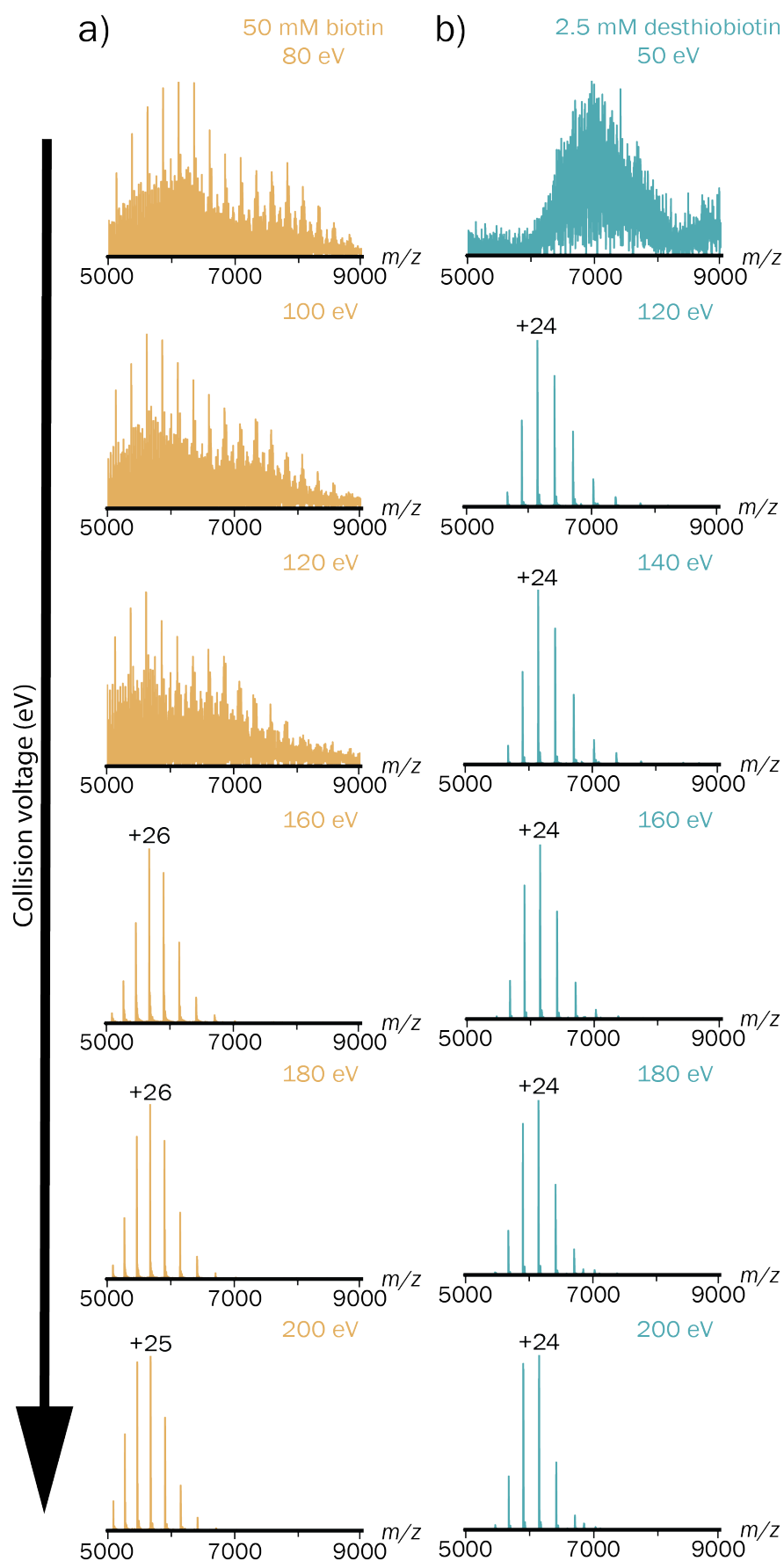
Figure A2: Spectra of the tetrameric ADH complex (148 kDa, 2.5 μM) in 150 mM ammonium acetate at pH 8 spiked with typical amounts of eluent at different collision voltages, increasing from top to bottom, on the QE-UHMR 2. a) ADH and 50 mM biotin at 80-200 eV. b) ADH and 2.5 mM desthiobiotin at 50-200 eV.
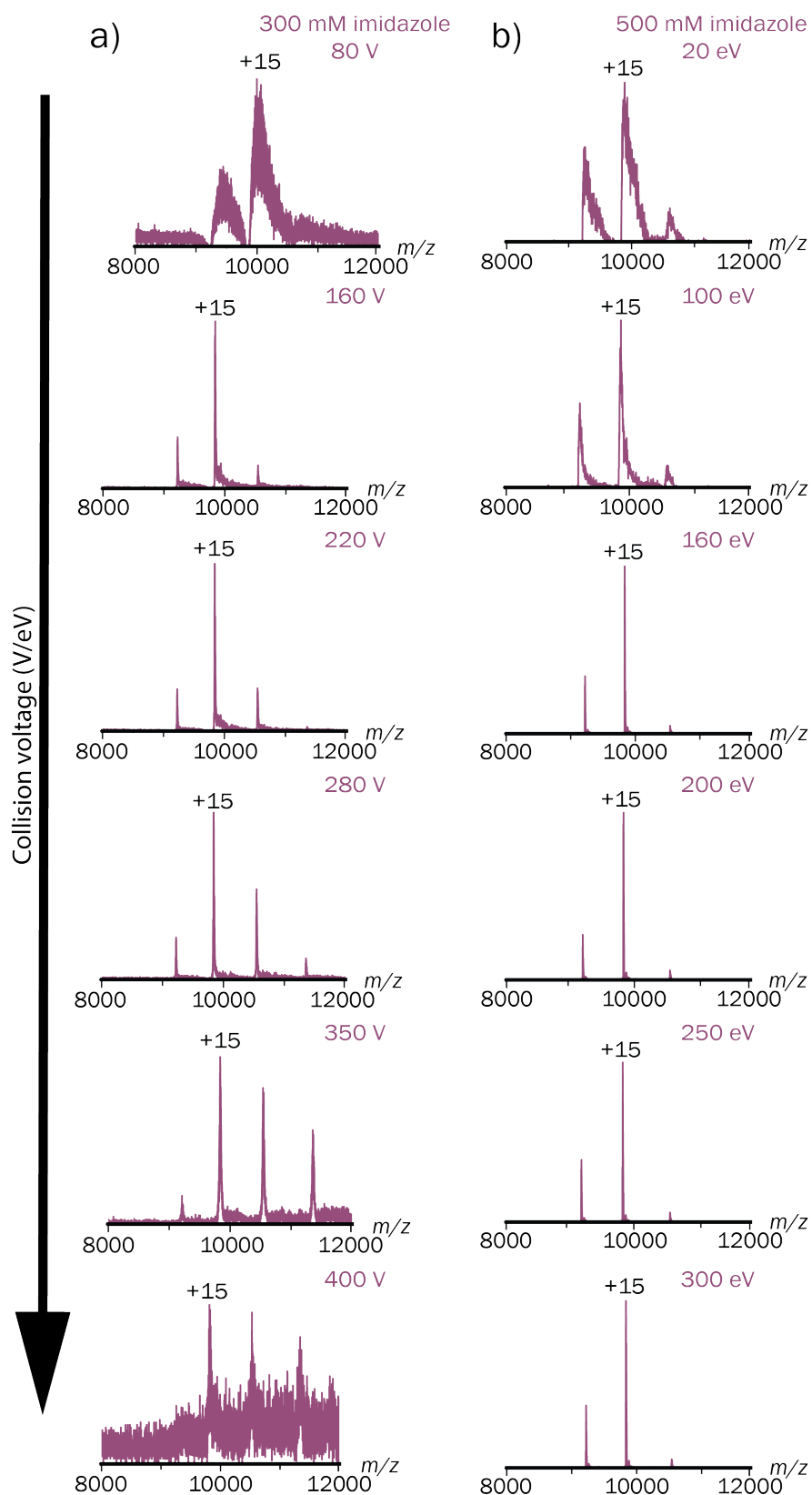
Figure A3: Spectra of the tetrameric ADH complex (148 kDa, 2.5 µM) in 150 mM ammonium acetate at pH 8 spiked with typical amounts of eluent at different collision voltages, increasing from top to bottom, analyzed on two different mass spectrometers. a) ADH and 300 mM imidazole at 80-400 V on the Q-TOF 2. b) ADH and 500 mM imidazole at 20-300 eV on the Q Exactive UHMR.
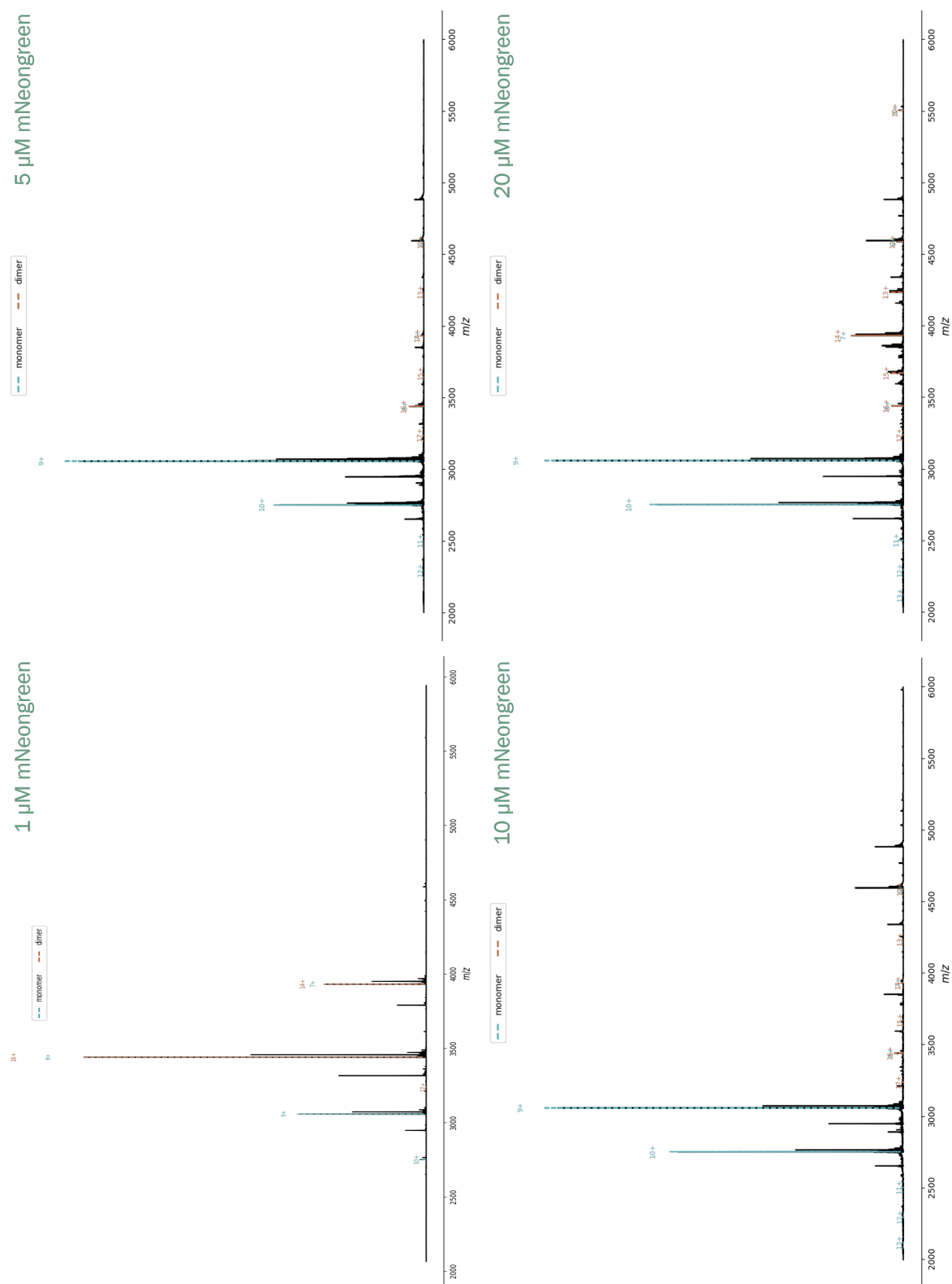
Figure A4: Spectra of mNeongreen at four different monomer concentrations (1, 5, 10, and 20 μM) in 350 mM ammonium acetate, and 1 mM DTT with less than 1 mM imidazole at pH 7.5 on the QE-UHMR. Monomer peaks are marked in blue and dimer peaks in orange. Peak finding was performed on a custom python script (see listing A1) from Janine-Denise Kopicki.
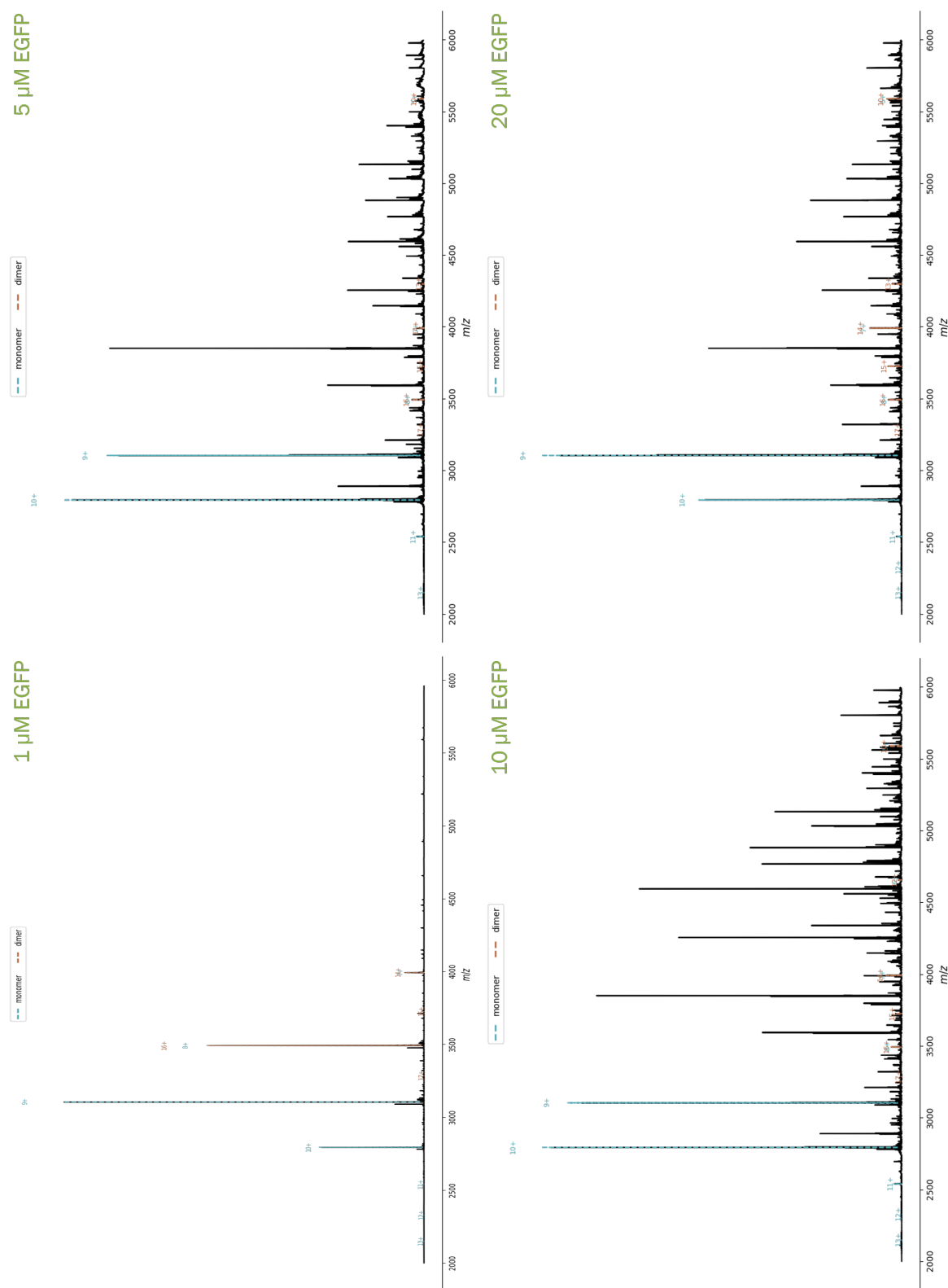
Figure A5: Spectra of EGFP at four different monomer concentrations (1, 5, 10, and 20 µM) in 350 mM ammonium acetate, and 1 mM DTT with 350 mM imidazole (20 µM monomer concentration), 175 mM imidazole (10 µM monomer concentration), 87.5 mM imidazole (10 µM monomer concentration), and 17.5 mM (10 µM monomer concentration) imidazole at pH 7.5 on the QE-UHMR. Monomer peaks are marked in blue and dimer peaks in orange. Peak finding was performed on a custom python script (see listing A1) from Janine-Denise Kopicki.
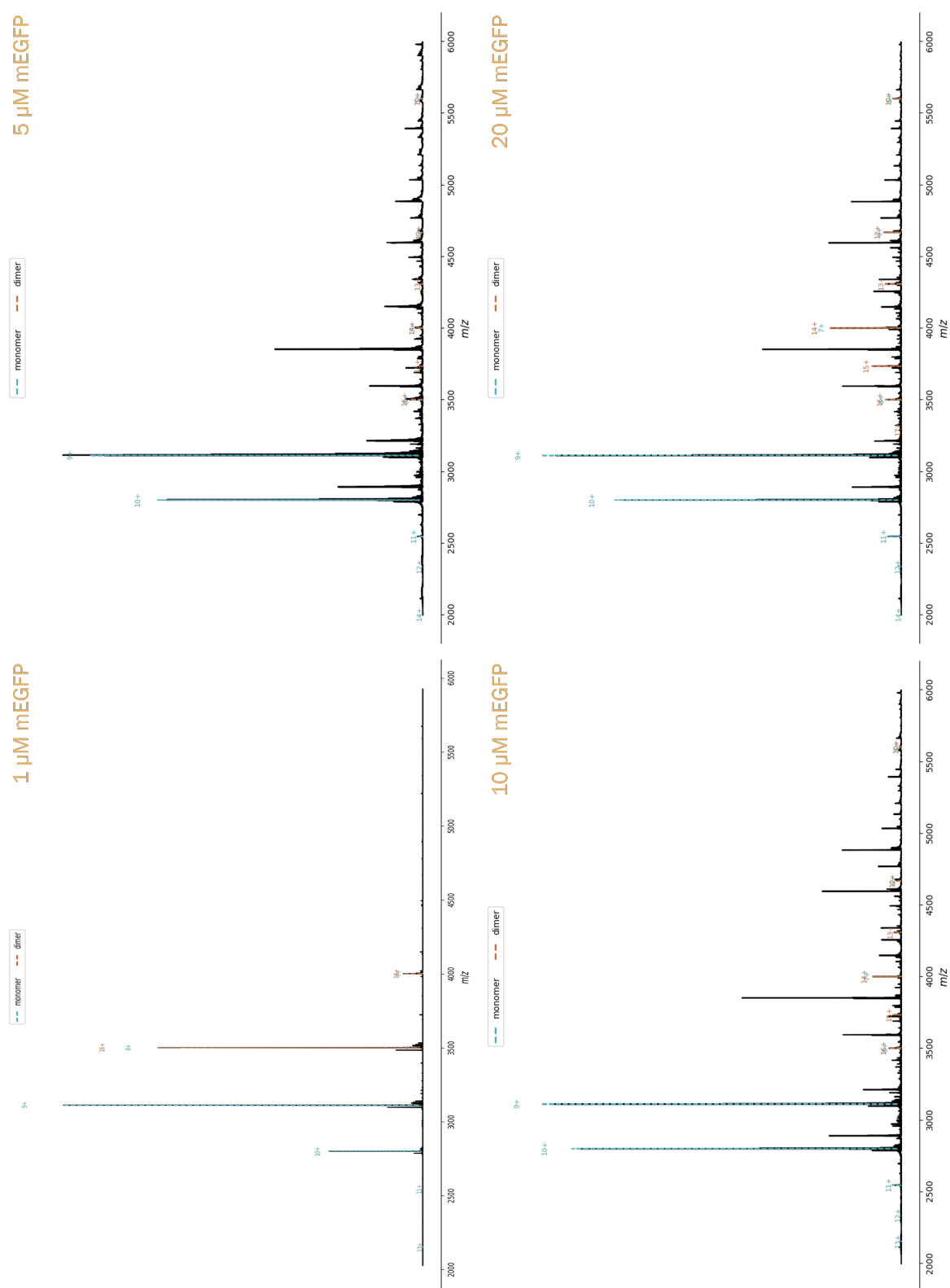
Figure A6: Spectra of mEGFP at four different monomer concentrations (1, 5, 10, and 20 μM) in 350 mM ammonium acetate, and 1 mM DTT with 350 mM imidazole (20 μM monomer concentration), 175 mM imidazole (10 μM monomer concentration), 87.5 mM imidazole (5 μM monomer concentration), and 17.5 mM (1 μM monomer concentration) imidazole at pH 7.5 on the QE-UHMR. Monomer peaks are marked in blue and dimer peaks in orange. Peak finding was performed on a custom python script (see listing A1) from Janine-Denise Kopicki.
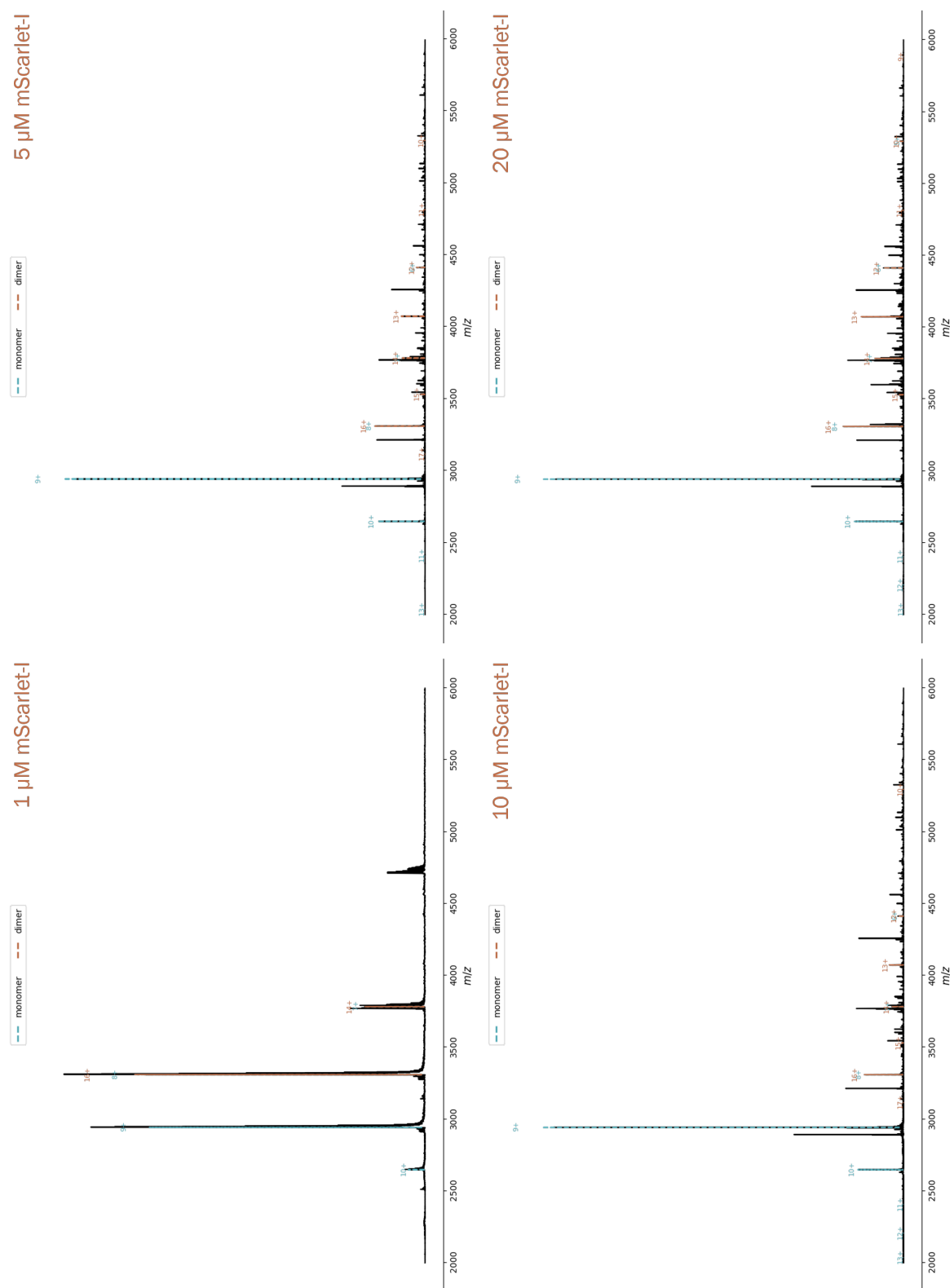
Figure A7: Spectra of mScarlet-I at four different monomer concentrations (1, 5, 10, and 20 µM) in 350 mM ammonium acetate, and 1 mM DTT at pH 7.5 on the QE-UHMR. Monomer peaks are marked in blue and dimer peaks in orange. Peak finding was performed on a custom python script (see listing A1) from Janine-Denise Kopicki.
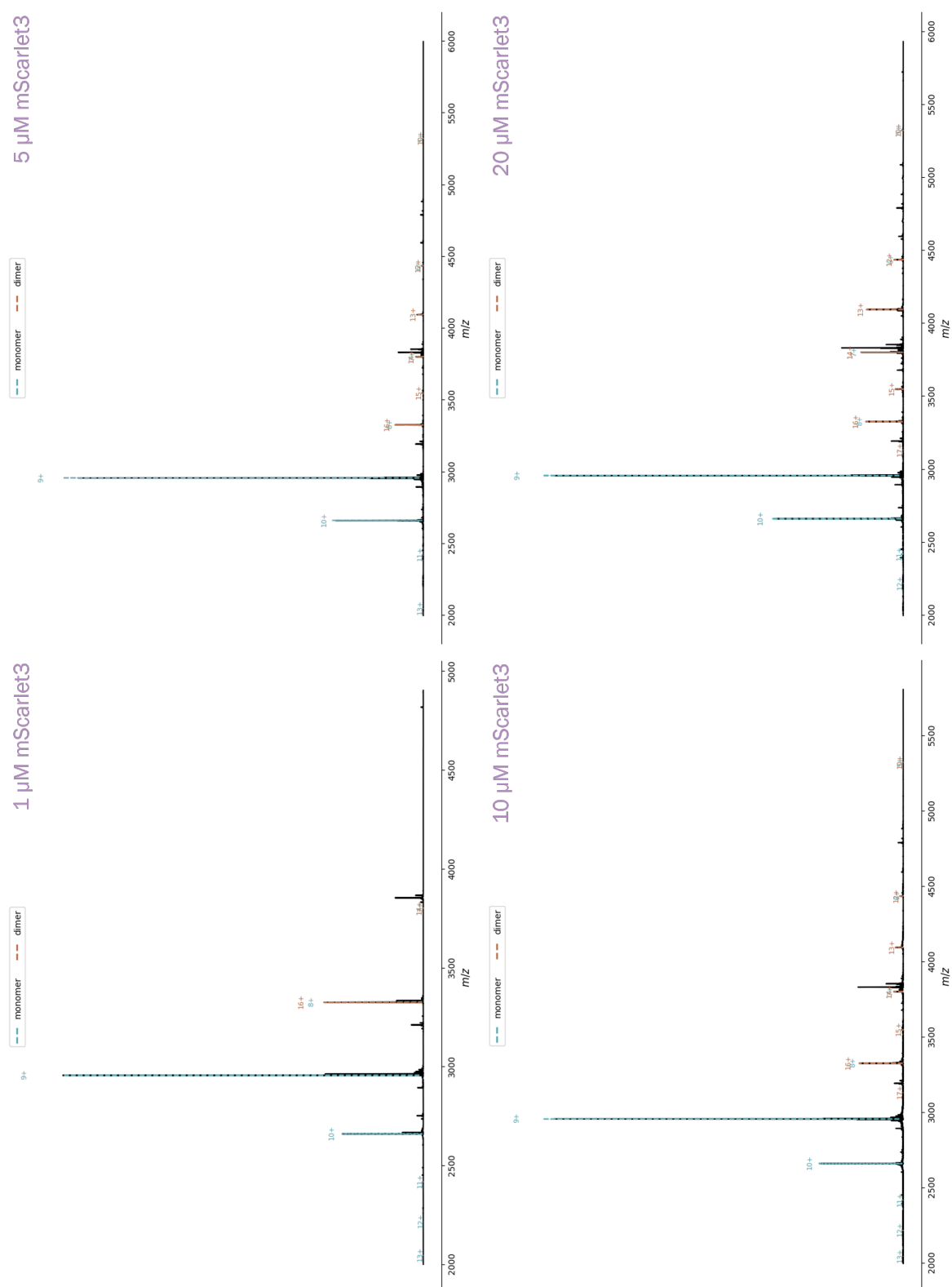
Figure A8: Spectra of mScarlet3 at four different monomer concentrations (1, 5, 10, and 20 µM) in 350 mM ammonium acetate, and 1 mM DTT with less than 1 mM imidazole at pH 7.5 on the QE-UHMR. Monomer peaks are marked in blue and dimer peaks in orange. Peak finding was performed on a custom python script (see listing A1) from Janine-Denise Kopicki.
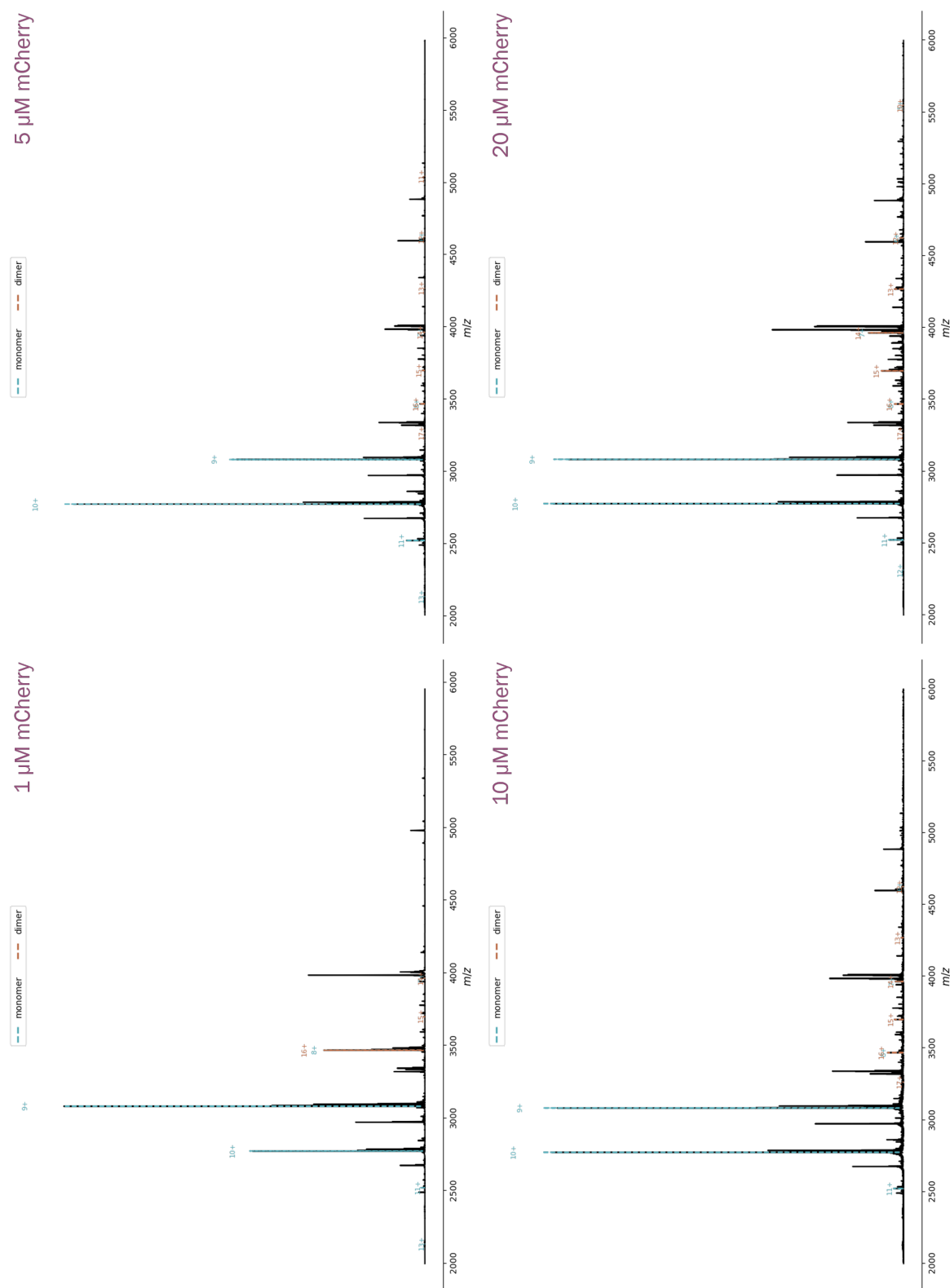
Figure A9: Spectra of mCherry at four different monomer concentrations (1, 5, 10, and 20 μM) in 350 mM ammonium acetate, and 1 mM DTT with 175 mM imidazole (20 μM monomer concentration), 87.5 mM imidazole (10 μM monomer concentration), 43.8 mM imidazole (5 μM monomer concentration), and 8.8 mM (1 μM monomer concentration) imidazole at pH 7.5 on the QE-UHMR. Monomer peaks are marked in blue and dimer peaks in orange. Peak finding was performed on a custom python script (see listing A1) from Janine-Denise Kopicki.
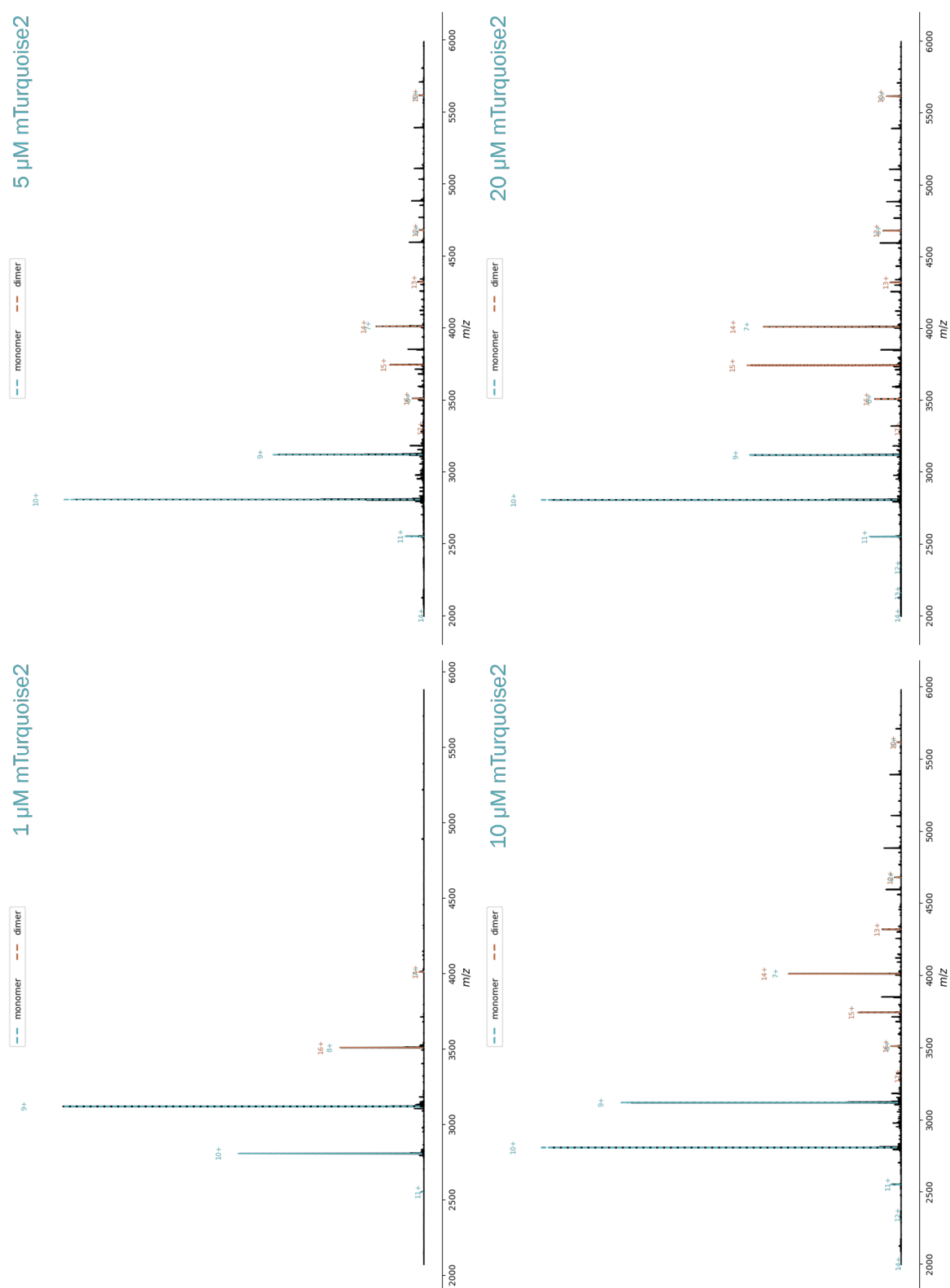
Figure A10: Spectra of mCherry at four different monomer concentrations (1, 5, 10, and 20 µM) in 350 mM ammonium acetate, and 1 mM DTT with 116.7 mM imidazole (20 µM monomer concentration), 58.3 mM imidazole (10 µM monomer concentration), 29.2 mM imidazole (5 µM monomer concentration), and 5.8 mM (1 µM monomer concentration) imidazole at pH 7.5 on the QE-UHMR. Monomer peaks are marked in blue and dimer peaks in orange. Peak finding was performed on a custom python script (see listing A1) from Janine-Denise Kopicki.

The MS data of the FP measurements were deconvoluted with a python script written by **Janine-Denise Kopicki**:

```python
# -*- coding: utf-8 -*-
"""
Created on Thu Sep 12 12:52:18 2024

@author: kopickij
"""


import os  # Importing the os module
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from Bio.SeqUtils.ProtParam import ProteinAnalysis
from scipy.signal import find_peaks
from scipy.optimize import curve_fit
import tkinter as tk
import tkinter.ttk as ttk
from tkinter import filedialog, messagebox, colorchooser
from matplotlib.backends.backend_tkagg import FigureCanvasTkAgg
import matplotlib.colors as mcolors
from matplotlib.backends.backend_tkagg import NavigationToolbar2Tk
from sklearn.metrics import r2_score


def get_mass_from_sequence(sequence):
    analyzed_seq = ProteinAnalysis(sequence)
```

```python
27     return analyzed_seq.molecular_weight()

28

29

30 def calculate_mz(mass, charge):
31     return (mass + charge * 1.007276466812) / charge

32

33

34 def process_spectrum(file_path):
35     try:
36         df = pd.read_csv(file_path, delimiter='\t', comment='#',
             skiprows=(7))
37         if 'Mass' not in df.columns or 'Intensity' not in df.
             columns:
38             print("Required columns 'Mass' or 'Intensity' not found
                 in the file.")
39             return None, None
40         return df['Mass'].values, df['Intensity'].values
41     except Exception as e:
42         print(f"An error occurred while reading the spectrum file:
             {e}")
43         return None, None

44

45

46 def match_peaks(mz_values, spectrum_mz, spectrum_intensity, window
    =0.5):
47     matched_peaks = []
48     for charge, mz in mz_values:
49         peaks, _ = find_peaks(spectrum_intensity, height=0)
```

```python
50          best_peak = None
51          best_stddev = float('inf')
52          for peak in peaks:
53              if abs(spectrum_mz[peak] - mz) <= window:
54                  # Calculate the standard deviation of the
                        recalculated mass
55                  recalculated_mass = (spectrum_mz[peak] * charge) -
                        (charge * 1.007276466812)
56                  stddev = np.std(recalculated_mass - mz)
57                  if stddev < best_stddev:
58                      best_stddev = stddev
59                      best_peak = (mz, charge, spectrum_mz[peak],
                            spectrum_intensity[peak])
60          if best_peak is not None:
61              matched_peaks.append(best_peak)
62      return matched_peaks


def gaussian(x, amp, cen, wid):
    return amp * np.exp(-(x-cen)**2 / (2*wid**2))


def fit_gaussian_to_peak(spec_mz, spec_int, mz, mz_window):
    indices = (spec_mz >= mz - mz_window) & (spec_mz <= mz +
        mz_window)
    if np.sum(indices) < 3:  # Ensure there are at least 3 points
        to fit
        print(f"Insufficient data points to fit Gaussian for peak
```

```python
                at m/z {mz}.")
        return None, None  # Return None for both values


    subset_mz = spec_mz[indices]
    subset_int = spec_int[indices]


    try:
        popt, _ = curve_fit(
            gaussian, subset_mz, subset_int,
            p0=[np.max(subset_int), mz, np.std(subset_mz)]
        )
        fitted_curve = gaussian(subset_mz, *popt)
        r_squared = r2_score(subset_int, fitted_curve)
        return popt, r_squared
    except RuntimeError:
        print(f"Could not fit Gaussian for peak at m/z {mz}.")
        return None, None  # Return None for both values if fitting
            fails
    except Exception as e:
        print(f"Unexpected error during fitting: {e}")
        return None, None


def filter_peaks(peaks, r_squared_threshold=0.9, max_charge_state
    =50):
    # Step 1: Filter out peaks that don't meet the
        r_squared_threshold or exceed the max_charge_state
    filtered_peaks = [p for p in peaks if p[-1] >=
```

```python
            r_squared_threshold and p[1] <= max_charge_state]

    # Step 2: Return all filtered peaks without looking for
        consecutive sequences
    return filtered_peaks


def integrate_peak(peak_mz, spectrum_mz, spectrum_intensity,
    use_gaussian_fit=True, r_squared_threshold=0.8):
    window = 0.5
    indices = (spectrum_mz >= peak_mz - window) & (spectrum_mz <=
        peak_mz + window)

    # Initialize area_real to 0 to handle cases where no data
        points are found
    area_real = 0

    if not np.any(indices):
        print(f"No data points found within the window for peak at
            m/z {peak_mz}.")
        return area_real, 0, None  # Return 0 for area_gaussian and
            None for r_squared if no data is found

    # Calculate the real area using the trapezoidal rule
    area_real = np.trapz(spectrum_intensity[indices], spectrum_mz[
        indices])

    if use_gaussian_fit:
```

```python
        popt, r_squared = fit_gaussian_to_peak(spectrum_mz,
            spectrum_intensity, peak_mz, window)
        if popt is not None and r_squared is not None and r_squared
            >= r_squared_threshold:
            amp, cen, wid = popt
            gaussian_indices = (spectrum_mz >= cen - 3*wid) & (
                spectrum_mz <= cen + 3*wid)
            if np.sum(gaussian_indices) < 3:
                print(f"Insufficient data points to calculate
                    Gaussian area for peak at m/z {cen}.")
                return area_real, 0, r_squared  # Return 0 for
                    area_gaussian but keep area_real and r_squared
            area_gaussian = np.trapz(spectrum_intensity[
                gaussian_indices], spectrum_mz[gaussian_indices])
            return area_real, area_gaussian, r_squared  # Include
                this peak in the results


    # If Gaussian fitting fails or is not used, return real area
        with 0 for Gaussian area and None for r_squared
    return area_real, 0, None



def plot_spectrum_in_gui(ax, spectrum_mz, spectrum_intensity):
    ax.clear()
    ax.plot(spectrum_mz, spectrum_intensity, 'k-')
    ax.set_xlabel("$m/z$", fontsize=12, style='italic')
    ax.get_yaxis().set_visible(False)
    ax.spines['top'].set_visible(False)
```

```python
137      ax.spines['right'].set_visible(False)

138      ax.spines['left'].set_visible(False)

139      ax.figure.canvas.draw()

140

141

142  def plot_results_in_gui(ax, spectrum_mz, spectrum_intensity,
         matched_peaks, species_colors, check_buttons,
         r_squared_threshold=0.9):

143      label_positions = []  # To track the positions of labels

144      ax.clear()

145      ax.plot(spectrum_mz, spectrum_intensity, 'k-', label='spectrum'
             )

146

147      handles = []

148      labels = []

149

150      for name, color in species_colors.items():

151          if check_buttons[name].var.get():  # Check if the species
                 is selected via the checkbox

152              peaks = matched_peaks.get(name, [])

153              filtered_peaks = filter_peaks(peaks,
                     r_squared_threshold=r_squared_threshold)

154

155              for peak in filtered_peaks:

156                  mz, charge, peak_mz, peak_intensity = peak[:4]

157                  popt, r_squared = fit_gaussian_to_peak(spectrum_mz,
                         spectrum_intensity, peak_mz, 0.5)

158                  if popt is not None:
```

```
159        amp, cen, wid = popt
160        x_vals = np.linspace(cen - 3*wid, cen + 3*wid,
               100)
161        gaussian_curve = gaussian(x_vals, amp, cen, wid
               )
162        ax.plot(x_vals, gaussian_curve, '--', color=
               color)
163
164        # Adjust the vertical position of charge labels
165        y_offset = 0.1 * peak_intensity
166
167        # Check for overlaps with previous labels
168        for lp_mz, lp_y in label_positions:
169            if abs(peak_mz - lp_mz) < 0.1:  # Adjust
                   threshold as needed
170                y_offset += 0.1 * peak_intensity  #
                       Increment offset if overlapping
171
172        ax.text(peak_mz, peak_intensity + y_offset, f"{
               charge}+", ha='center', va='bottom',
               fontsize=8, color=color)
173        label_positions.append((peak_mz, peak_intensity
               + y_offset))
174
175    # Add species to legend
176    if filtered_peaks:
177        handles.append(plt.Line2D([0], [0], color=color, lw
               =2, linestyle='--'))
```

```python
178                 labels.append(name)

179

180     ax.set_xlabel("$m/z$", fontsize=12, style='italic')

181     ax.get_yaxis().set_visible(False)

182     ax.spines['top'].set_visible(False)

183     ax.spines['right'].set_visible(False)

184     ax.spines['left'].set_visible(False)

185

186     # Custom legend

187     ax.legend(handles=handles, labels=labels, loc='upper center',
            bbox_to_anchor=(0.5, 1.1), ncol=len(handles))

188     ax.figure.canvas.draw()

189

190

191 class CheckButton:

192     def __init__(self, parent, label, command):

193         self.var = tk.IntVar(value=1)

194         self.checkbutton = tk.Checkbutton(parent, text=label,
                variable=self.var, command=command)

195         self.checkbutton.pack(anchor='w')

196

197     def deselect(self):

198         self.var.set(0)

199

200     def select(self):

201         self.var.set(1)

202

203
```

```python
class MassSpecApp:
    def __init__(self, root):
        self.root = root
        self.root.title("openJDK mass spec analysis V01.02")
        self.root.geometry("1200x800")
        self.root.state('zoomed')

        self.file_path = None
        self.spectrum_mz = None
        self.spectrum_intensity = None
        self.matched_peaks = None
        self.species_colors = None
        self.summary_data = None
        self.check_buttons = {}  # Initialize this here to avoid
            errors

        # Layout configuration
        root.grid_rowconfigure(9, weight=1)
        root.grid_columnconfigure(2, weight=1)

        # File selection
        self.selected_file_label = tk.Label(root, text="")
        self.selected_file_label.grid(row=0, column=0, columnspan
            =2, sticky='ew', padx=10, pady=1)
        self.file_label = tk.Label(root, text="select spectrum file
            : ")
        self.file_label.grid(row=0, column=0, sticky='w', padx=10,
            pady=1)
```

```python
228        self.file_button = tk.Button(root, text="browse", command=
               self.browse_file)
229        self.file_button.grid(row=0, column=0, columnspan=2, sticky
               ='e', padx=10, pady=1)
230
231        # Mass range
232        self.min_mz_label = tk.Label(root, text="min m/z")
233        self.min_mz_label.grid(row=1, column=0, sticky='w', padx
               =10, pady=1)
234        self.min_mz_entry = tk.Entry(root)
235        self.min_mz_entry.grid(row=1, column=0, sticky='e', padx
               =10, pady=1)
236
237        self.max_mz_label = tk.Label(root, text="max m/z")
238        self.max_mz_label.grid(row=1, column=1, sticky='w', padx
               =10, pady=1)
239        self.max_mz_entry = tk.Entry(root)
240        self.max_mz_entry.grid(row=1, column=1, sticky='e', padx
               =10, pady=1)
241
242        # Plot spectrum button
243        self.plot_spectrum_button = tk.Button(root, text="plot
               spectrum", command=self.plot_spectrum)
244        self.plot_spectrum_button.grid(row=2, column=0, columnspan
               =2, sticky='ew', padx=10, pady=1)
245
246        # Maximum charge state
247        self.max_charge_state_label = tk.Label(root, text="max z")
```

```python
        self.max_charge_state_label.grid(row=3, column=0, sticky='w
            ', padx=10, pady=1)
        self.max_charge_state_entry = tk.Entry(root)
        self.max_charge_state_entry.grid(row=3, column=0, sticky='e
            ', padx=10, pady=1)

        # Window
        self.window_label = tk.Label(root, text="window")
        self.window_label.grid(row=3, column=1, sticky='w', padx
            =10, pady=1)
        self.window_entry = tk.Entry(root)
        self.window_entry.grid(row=3, column=1, sticky='e', padx
            =10, pady=1)

        # Species input
        self.species_frame = tk.Frame(root)
        self.species_frame.grid(row=4, column=0, columnspan=2,
            sticky='ew', padx=10, pady=1)

        self.species_label = tk.Label(self.species_frame, text="
            species name")
        self.species_label.grid(row=0, column=0, padx=5)

        self.mass_label = tk.Label(self.species_frame, text="mass")
        self.mass_label.grid(row=0, column=1, padx=5)

        self.sequence_label = tk.Label(self.species_frame, text="
            sequence")
```

```
269        self.sequence_label.grid(row=0, column=2, padx=5)

270

271        self.color_label = tk.Label(self.species_frame, text="color
               ")

272        self.color_label.grid(row=0, column=3, padx=5)

273

274        self.species_entries = []

275        self.mass_entries = []

276        self.sequence_entries = []

277        self.color_buttons = []

278        self.color_values = []

279

280        self.add_species_row()

281

282        self.add_species_button = tk.Button(root, text="add species
               ", command=self.add_species_row)

283        self.add_species_button.grid(row=5, column=0, sticky='ew',
               padx=10, pady=1)

284

285        self.remove_species_button = tk.Button(root, text="remove
               selected species", command=self.remove_species_row)

286        self.remove_species_button.grid(row=5, column=1, sticky='ew
               ', padx=10, pady=1)

287

288        # Start analysis button

289        self.start_button = tk.Button(root, text="start analysis",
               command=self.start_analysis)

290        self.start_button.grid(row=6, column=0, sticky='ew', padx
```

```
              =10)
291
292          # Export buttons
293          self.export_button = tk.Button(root, text="export results",
                 command=self.export_results)
294          self.export_button.grid(row=6, column=1, sticky='ew', padx
                =10)
295
296          # Plot and table frames
297          self.plot_frame = tk.Frame(root)
298          self.plot_frame.grid(row=0, column=2, rowspan=7, sticky='
                nsew', padx=5, pady=5)
299
300          self.checkbox_frame = tk.Frame(root)
301          self.checkbox_frame.grid(row=0, column=3, rowspan=7, sticky
                ='nsew', padx=5, pady=5)
302
303          self.table_frame = tk.Frame(root)
304          self.table_frame.grid(row=7, column=0, columnspan=3, sticky
                ='nsew', padx=5, pady=5)
305
306          # Create matplotlib figure and axes for plot
307          self.fig, self.ax = plt.subplots(figsize=(11.69, 8.27))  #
                Larger size, horizontal layout
308          self.canvas = FigureCanvasTkAgg(self.fig, master=self.
                plot_frame)
309          self.canvas.get_tk_widget().pack(fill=tk.BOTH, expand=True)
310
```

```python
311         # Create the toolbar and associate it with the canvas
312         toolbar = NavigationToolbar2Tk(self.canvas, self.plot_frame
            )
313         toolbar.update()
314         self.canvas.get_tk_widget().pack(side=tk.TOP, fill=tk.BOTH,
            expand=1)
315
316         # Create table for displaying results
317         self.table = None
318         self.check_buttons = {}
319
320
321     def browse_file(self):
322         self.file_path = filedialog.askopenfilename(title="select
            spectrum file")
323         if self.file_path:
324             self.selected_file_label.config(text=os.path.basename(
                self.file_path))
325
326
327     def add_species_row(self):
328         row = len(self.species_entries) + 1
329
330         species_entry = tk.Entry(self.species_frame)
331         species_entry.grid(row=row, column=0, padx=5)
332         self.species_entries.append(species_entry)
333
334         mass_entry = tk.Entry(self.species_frame)
```

```python
        mass_entry.grid(row=row, column=1, padx=5)
        self.mass_entries.append(mass_entry)


        sequence_entry = tk.Entry(self.species_frame)
        sequence_entry.grid(row=row, column=2, padx=5)
        self.sequence_entries.append(sequence_entry)


        color_button = tk.Button(self.species_frame, text="choose
            color", command=lambda: self.choose_color(row-1))
        color_button.grid(row=row, column=3, padx=5)
        self.color_buttons.append(color_button)


        self.color_values.append(None)


    def remove_species_row(self):
        selected_indices = [i for i, entry in enumerate(self.
            species_entries) if entry.get() == self.species_entries
            [-1].get()]
        for index in reversed(selected_indices):
            self.species_entries[index].grid_forget()
            self.mass_entries[index].grid_forget()
            self.sequence_entries[index].grid_forget()
            self.color_buttons[index].grid_forget()


            del self.species_entries[index]
            del self.mass_entries[index]
            del self.sequence_entries[index]
            del self.color_buttons[index]
```

```
360            del self.color_values[index]


363    def choose_color(self, index):
364        color = colorchooser.askcolor(title="choose color")[1]
365        if color:
366            self.color_values[index] = color
367            self.color_buttons[index].config(bg=color)
368        else:
369            self.color_values[index] = plt.cm.tab10(index % 10)
370            self.color_buttons[index].config(bg=mcolors.to_hex(self
                   .color_values[index]))


373    def plot_spectrum(self):
374        if not self.file_path or not self.min_mz_entry.get() or not
                   self.max_mz_entry.get():
375            messagebox.showerror("Error", "Please provide the
                   spectrum file and mass range.")
376            return


378        min_mz = float(self.min_mz_entry.get())
379        max_mz = float(self.max_mz_entry.get())


381        self.spectrum_mz, self.spectrum_intensity =
                   process_spectrum(self.file_path)
382        if self.spectrum_mz is None or self.spectrum_intensity is
                   None:
```

```python
            messagebox.showerror("Error", "Failed to load spectrum
                data.")
            return


        # Filter spectrum by the specified mass range
        valid_indices = (self.spectrum_mz >= min_mz) & (self.
            spectrum_mz <= max_mz)
        self.spectrum_mz = self.spectrum_mz[valid_indices]
        self.spectrum_intensity = self.spectrum_intensity[
            valid_indices]


        plot_spectrum_in_gui(self.ax, self.spectrum_mz, self.
            spectrum_intensity)



    def start_analysis(self):
        # Clear previous results
        self.matched_peaks = None
        self.summary_data = None  # Corrected from summary_df to
            summary_data
        self.summary_df = None  # Ensure this is reset

        # Clear the plot
        self.ax.clear()
        self.canvas.draw()

        # Clear the table if it exists
        if self.table:
```

```python
406              self.table.destroy()
407              self.table = None
408
409          # Clear the check buttons
410          for widget in self.checkbox_frame.winfo_children():
411              widget.destroy()
412
413          # Proceed with the normal analysis process
414          if not self.file_path or not self.min_mz_entry.get() or not
                 self.max_mz_entry.get() or not self.species_entries:
415              messagebox.showerror("Error", "Please provide all
                     required inputs.")
416              return
417
418          min_mz = float(self.min_mz_entry.get())
419          max_mz = float(self.max_mz_entry.get())
420
421          # Get the user-defined peak window size, with a default of
                 0.5 if not provided
422          try:
423              window_size = float(self.window_entry.get()) if self.
                     window_entry.get() else 0.5
424          except ValueError:
425              window_size = 0.5
426
427          # Get the user-defined maximum charge state, with a default
                 of 50 if not provided
428          try:
```

```python
            max_charge_state = int(self.max_charge_state_entry.get
                ()) if hasattr(self, 'max_charge_state_entry') and
                self.max_charge_state_entry.get() else 50
        except ValueError:
            max_charge_state = 50

        self.spectrum_mz, self.spectrum_intensity =
            process_spectrum(self.file_path)
        if self.spectrum_mz is None or self.spectrum_intensity is
            None:
            messagebox.showerror("Error", "Failed to load spectrum
                data.")
            return

        # Filter spectrum by the specified mass range
        valid_indices = (self.spectrum_mz >= min_mz) & (self.
            spectrum_mz <= max_mz)
        self.spectrum_mz = self.spectrum_mz[valid_indices]
        self.spectrum_intensity = self.spectrum_intensity[
            valid_indices]

        self.species_colors = {}
        species = []

        for i, species_entry in enumerate(self.species_entries):
            name = species_entry.get()
            mass = self.mass_entries[i].get()
            sequence = self.sequence_entries[i].get()
```

```python
450            color = self.color_values[i]
451
452            if not name or not (mass or sequence):
453                messagebox.showerror("Error", "Please provide all
                       details for each species.")
454                return
455
456            if sequence:
457                mass = get_mass_from_sequence(sequence)
458                self.mass_entries[i].delete(0, tk.END)
459                self.mass_entries[i].insert(0, str(mass))
460
461            species.append((name, float(mass)))
462            if not color:
463                color = plt.cm.tab10(i % 10)
464                self.color_buttons[i].config(bg=mcolors.to_hex(
                       color))
465            self.species_colors[name] = color
466
467            # Create check button for the species and add it to the
                   frame and dictionary
468            self.check_buttons[name] = CheckButton(self.
                   checkbox_frame, name, self.update_plot)
469
470        self.matched_peaks = {
471            name: match_peaks(
472                [(charge, calculate_mz(mass, charge)) for charge in
                       range(1, max_charge_state + 1)],
```

```
                self.spectrum_mz,
                self.spectrum_intensity,
                window=window_size
            ) for name, mass in species
        }

        summary_data = []
        for name, peaks in self.matched_peaks.items():
            filtered_peaks = filter_peaks(peaks,
                r_squared_threshold=0.9, max_charge_state=
                max_charge_state)  # Apply filtering
            species_intensity_sum = 0
            species_area_sum_real = 0
            species_area_sum_gaussian = 0
            species_masses = []
            for mz, charge, peak_mz, peak_intensity in
                filtered_peaks:
                area_real, area_gaussian, r_squared =
                    integrate_peak(peak_mz, self.spectrum_mz, self.
                    spectrum_intensity, use_gaussian_fit=True)
                if r_squared is not None:  # Ensure r_squared is
                    valid before adding to the summary
                    mass_recalculated = peak_mz * charge - charge *
                        1.007276466812
                    summary_data.append([name, mz, charge, peak_mz,
                        peak_intensity, area_real, area_gaussian,
                        r_squared, mass_recalculated])
                    species_intensity_sum += peak_intensity
```

```python
                    species_area_sum_real += area_real
                    species_area_sum_gaussian += area_gaussian
                    species_masses.append(mass_recalculated)

            avg_mass = np.mean(species_masses) if species_masses
                else 0
            std_mass = np.std(species_masses) if species_masses
                else 0
            summary_data.append([name, 'summed', '', '',
                species_intensity_sum, species_area_sum_real,
                species_area_sum_gaussian, None, ''])
            summary_data.append([name, 'average mass', '', '', '',
                '', '', '', avg_mass])
            summary_data.append([name, 'std deviation', '', '', '',
                '', '', '', std_mass])

        self.summary_data = summary_data  # Store the filtered data
            to be used in show_table

        # Convert summary_data to a DataFrame for exporting
        self.summary_df = pd.DataFrame(self.summary_data, columns=[
            "species", "m/z (calc)", "z", "m/z (exp)", "intensity",
            "area (exp)", "area (Gaussian)", "r-Squared", "m (exp)"
            ])

        self.show_table()  # Display the filtered data in the table
        self.update_plot()
```

```python
    def update_plot(self):
        plot_results_in_gui(self.ax, self.spectrum_mz, self.
            spectrum_intensity, self.matched_peaks, self.
            species_colors, self.check_buttons)


    def show_table(self):
        if not self.summary_data:
            return

        if self.table:
            self.table.destroy()

        self.table = ttk.Treeview(self.table_frame, columns=("
            Species", "mz", "Charge", "Peak m/z", "Intensity", "Area
             (Real)", "Area (Gaussian)", "R-Squared", "Recalculated
            Mass"), show="headings")
        self.table.pack(fill="both", expand=True)

        for col in self.table["columns"]:
            self.table.heading(col, text=col)

        for row in self.summary_data:
            self.table.insert("", "end", values=row)


    def export_results(self):
```

```
533         if self.summary_df is not None and not self.summary_df.
              empty:
534             # Get the original file name and directory
535             original_file_name = os.path.basename(self.file_path).
                  split('.')[0]  # Remove file extension
536             original_directory = os.path.dirname(self.file_path)
537
538             # Get m/z range and create formatted name with the
                  range
539             min_mz = self.min_mz_entry.get()
540             max_mz = self.max_mz_entry.get()
541             range_suffix = f"{min_mz}-{max_mz}"
542
543             # Construct the base file name with range in the
                  original file directory
544             base_file_name = os.path.join(original_directory, f"{
                  original_file_name}_{range_suffix}")
545
546             # Save summary DataFrame as Excel and CSV
547             self.summary_df.to_excel(f"{base_file_name}.xlsx",
                  index=False)
548
549             # Save the plot as PNG and EPS with the same base name
550             self.fig.savefig(f"{base_file_name}.png", transparent=
                  True, bbox_inches='tight')
551             self.fig.savefig(f"{base_file_name}.eps", transparent=
                  True, bbox_inches='tight')
552             self.fig.savefig(f"{base_file_name}.svg", transparent=
```

```
                     True, bbox_inches='tight')
553
554          messagebox.showinfo("Success", f"Results saved in {
                 original_directory}")
555      else:
556          messagebox.showerror("Error", "No results to export.
                 Please run the analysis first.")
557

558

559 if __name__ == "__main__":
560     root = tk.Tk()
561     app = MassSpecApp(root)
562     root.mainloop()
```

Listing A1: Python script for the deconvolution of native MS raw data

Table A4: Overview of all substances used in this work. They are listed along with their hazard pictograms as well as hazard and precautionary statements according to the Globally Harmonized System of Classification and Labelling of Chemicals (GHS). The information was retrieved from safety data sheets provided by respective manufacturers.

Chemical (CAS)	Hazard Pictogram	Hazard Statements	Precautionary Statements	Vendor
2-[4-(2-Hydroxyethyl)piperazin-1-yl]ethane-1-sulfonic acid, HEPES, (7365-45-9)		H315, H319, H335	P261, P264, P270, P271, P280, P301+P312, P302+P352, P304+P312, P304+P340, P305+P351+P338, P312, P321, P322, P332+P313, P337+P313, P362, P363, P403+P233, P405, P501	Thermo Fisher Scientific
2-mercaptoethanol >=99%, p.a., (60-24-2)		H301+H331, H310, H315, H317, H318, H373, H410	P273, P280, P302+P352, P304+P340, P305+P351+P338, P310	Carl Roth
3-Dimethyl[3-(3α,7α,12α-trihydroxy-5β-cholan-24-amido)propyl]-azaniumylpropane-1-sulfonate, CHAPS >_98%, (75621-03-3)		H315, H319, H335	P261, P280, P302+P352, P304+P340, P305+P351+P338, P312	Carl Roth
5-Bromo-4-chloro-1H-indol-3-yl β-D-galactopyranoside, X-gal, (7240-90-6)		H302 + H312 + H332	P301 + P330 + P331, P312, P280, P302+P352, P304+P340	Carl Roth
			Continued on the next page.	

acetic acid, (64-19-7)		H226, H314	P210, P280, P303+P361+P353, P305+P351+P338+P310	Fluka
acetonitrile, (75-05-8)		H225, H302, H312, H319, H332	P210, P280, P301+P312, P302+P352, P304+P340, P305+P351+P338	Sigma Aldrich
acrylamide, (79-06-1)		H301, H312+H332, H315, H319, H317, H372, H350, H340, H361	P301+P310, P302+P352, P304+P340, P305+P351+P338, P333+P313, P260, P202	Sigma Aldrich
agarose (9012-36-6)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Carl Roth
alcohol dehydrogenase, ADH, (9031-72-5)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Sigma Aldrich
ammonium acetate, (631-61-8)		H303, H316, H320, H333	P281, P335	Fluka
ammonium hydroxide, (1336-21-6)		H314, H331, H410	P260,P273, P280, P303+P361+P353, P304+P340+P311, P305+P351+P338+P310	Thermo Fisher Scientific
				Continued on the next page.

ampicillin, (69-53-4)		H315, H317, H319, H334, H335	P261, P264, P264+265, P271, P272, P280, P284, P302+P352, P304+P340, P305+P351+P338, P319, P321, P332+317, P333+P313, P337+P317, P342+P316, P362+P364, P403+P233, P403+P233, P405+P501	Thermo Fisher Scientific
argon, (7440-37-1)		H280, H281	P403+P410	Linde
benzonase nuclease®, (9025-65-4)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Merck
biotin, (58-85-5)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			IBA Lifesciences
Blauer Jonas	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			German Research Products
cesium iodide, (7789-17-5)		H315, H317, H319, H335, H410	P280, P305+P351+P338	Sigma Aldrich
				Continued on the next page.

chloroform, (67-66-3)		H302, H315, H319, H331, H336, H351, H361d, H372	P201, P202, P235, P260, P264, P270, P271, P280, P281, P301+P330+P331, P302+P352, P304+P340, P305+P351+P338, P308+P313, P310, P311, P314, P332+P313, P337+P313, P362, P403+P233, P405, P501	Thermo Fisher Scientific
cOmplete™ EDTA-free Protease Inhibitor Cocktail		H314 Causes severe skin burns and eye damage.	P260, P280, P301+P330+P331, P303+P361+P353, P304+P340+P310	Roche
crystal violet, (548-62-9)		H302, H318, H351, H410	P273, P280, P305+P351+P338, P501	Thermo Fisher Scientific
desthiobiotin, (533-48-2)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			IBA Lifesciences
dimethyl sulfoxide, (67-68-5)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Thermo Fisher Scientific
Continued on the next page.				

dithiothreitol, DTT, (3483-12-3)		H302, H315, H319, H335	P261, P264, P264+P265, P270, P271, P280, P301+P371, P302+P352, P304+P340, P305+P351+P338, P319, P321, P332+P317, P362+P364, P403+P233, P405, P501	Sigma Aldrich
ethanol, (64-17-5)		H225, H319, H360D	P210, P233, P240, P241, P242, P305+P351+P338	Thermo Fisher Scientific
formaldehyde, (50-00-0)		H301+H311+H331, H314, H317, H335, H341, H350, H370	P201, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338, P308+P310	Sigma Aldrich
formic acid, (64-18-6)		H226, H290, H302, H314, H331	P210, P280, P303+P361+P353, P304+P340, P305+P351+P338, P310	Thermo Fisher Scientific

gentamicin, (1403-66-3)		H317, H334, H360, H360D, H372, H410	P203, P233, P260, P264, P270, P271, P272, P273, P280, P284, P302, P304, P316, P317, P318, P319, P321, P333, P340, P342, P352, P362, P364, P391, P403, P405, P501	Thermo Fisher Scientific
glycerol, (56-81-5)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Merck
hydrochloric acid, (7647-01-0)		H314, H331	P260, P261, P264, P271, P280, P301+P330+P331, P302+P361+P354, P304+P340, P305+P354+P338, P316, P321, P363, P403+P233, P405, P501	Thermo Fisher Scientific
imidazole, (288-32-4)		H302, H314, H360D	P263, P270, P280, P301+P310, P305+P351+P338, P308+P313	Thermo Fisher Scientific
kanamycin, (59-01-8)		H360D	P201, P202, P280, P308+P313, P405, P501	Thermo Fisher Scientific
magnesium dichloride, (7786-30-3)		H319, H335		Thermo Fisher Scientific

Continued on the next page.

methanol, (67-56-1)		H225, H301, H311, H331, H370	P210, P233, P235, P240, P241, P242, P243, P260, P264, P270, P271, P280, P301+P330+P331, P302+P352, P303+P361+P353, P304+P340, P305+P351+P338, P307+P311, P310, P311, P312, P337+P313, P361, P363, P370+P378, P403+P233, P405, P501	Thermo Fisher Scientific
N,N-dimethylformamide, DMF, (68-12-2)		H226, H312, H319, H332, H360	P280, P305+P351+P338, P308+P313	Thermo Fisher Scientific
nitrogen, (17778-88-0/7727-37-9(N2))		H280, H281	P282, P336+P317, P403, P410+P403	Linde
phenol, (108-95-2)		H301, H311, H314, H331, H341, H373	P261, P280, P301+P310, P305+P351+P338, P310	Thermo Fisher Scientific
phosohate buffered saline	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Sigma-Aldrich
potassium chloride, KCl, (7447-40-7)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Carl Roth

propan-2-ol, (67-63-0)		H225, H302, H319, H336	P210, P261, P305+P351+P338	Thermo Fisher Scientific
propan-2-yl 1-thio-β-D-galactopyranoside, IPTG, (367-93-1)		H350	P201, P202, P280, P308+313, P405, P501	Thermo Fisher Scientific
PureCube 100 Ni-INDIGO Agarose resin	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Cube Biotech
RNase-free water	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Qiagen
sodium chloride, NaCl, (7647-14-5)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Carl Roth
sodium dodecyl sulfate, SDS, (151-21-3)		H228, H302+H332, H315, H318, H335, H412	P210, P273, P280, P301+P312, P304+P340+P312, P305+P351+P338	Carl Roth
sodium hydroxide, (1310-73-2)		H290, H314	P234, P260, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338	Sigma Aldrich
SYBR™Safe DNA Gel Stain, (1030826-36-8)		H227	P210, P280, P370+P378, P403, P501	Invitrogen
tetracycline, (60-54-8)		H302		Thermo Fisher Scientific
				Continued on the next page.

trifluoroacetic acid, (76-05-1)		H314, H332, H412	P260, P261, P264, P271, P273, P280, P301+P330+P331, P303+P361+P353, P304+P312, P304+P340, P305+P35+P338, P310, P312, P321, P363, P405, P501	Thermo Fisher Scientific
trimethylammonium acetate, TMAA, (6850-27-7)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Santa Cruz Biotechnology Inc.
2-Amino-2-(hydroxymethyl)propane-1,3-diol hydrochloride, Tris-HCl, (1185-53-1)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Carl Roth
Triton™X-100, (9002-93-1)		H302, H315, H318, H410	P264, P273, P280, P301+P312, P302+P352, P305+P351+P338	Sigma Aldrich
trypan blue, (72-57-1)		H350	P201, P202, P280, P308+P313, P405, P501	Merck
trypsin, (9002-07-7)		H315, H319, H334, H335	P261, P264, P271, P280, P302+P352, P305+P351+P338	Promega
Tween 20, (9005-64-5)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Thermo Fisher Scientific
urea, (57-13-6)	Not a hazardous substance or mixture according to Regulation (EC) No 1272/2008.			Thermo Fisher Scientific
xenon		H280	P403	Linde

<div align="right">Continued on the next page.</div>

zinc dichloride, (7646-85-7)		H302, H314, H410	P273, P280, P301+330+331, P305+P351+P338, P308+P310	Merck

Acknowledgement

These might've been the best and worst years of my life (so far). . . . I started my PhD during the beginning of the COVID-19 pandemic and a lot has happened since then. It started with me sitting alone in an office having to wear PPF2 masks every time I went outside and ended in a huge office surrounded by some of the best people I have ever met. Who would've thought? Not me!

Let's start from a little earlier, during my masters we had to do *Vertiefungspraktika* and accidentally, I did my first one in a mass spectrometry group at the University of Leipzig. It was just simple MS[1] experiments with small metabolites, but I really enjoyed it and it emboldened my flame of ambition for the field of mass spectrometry. A year later, I did another *Vertiefungspraktikum* in environmental chemistry using GC-MS and LC-MS. Honestly, at that point I knew that mass spectrometry was what I wanted to do, just not in Leipzig. So I went to Hamburg to do my master thesis at Hartmut Schlüters lab — proteomics with LCMS. I might be one of the few ones in Charlottes working group that actually came for the mass spectrometry and not the structural biology. As a chemist, it was really hard in the beginning understanding what others would consider basic biochemistry. Four and a half years later, I am teaching younger students how to do basic lab work and I guess that's how it is supposed to be.

I really want to thank **Hartmut** for helping me get a footing in Hamburg and also telling me about Charlottes work group and I want to thank **Charlotte** and **Jens** for taking me in and providing me the guidance I needed without taking away my creative scientific freedom. Also, thanks for not giving up on me when things weren't going as planned.

The biggest thanks go to my fellow PhD and Postdoc students working both in Jens and Charlottes lab. I changed offices so often and worked with so many people I can't possibly thank all of them, so to keep it short. I really need to thank **Enrico** for helping me with experiments, especially at the end of my PhD and **Tim** for helping me, especially at the beginning of my PhD and just answering all of my stupid questions. Thank you, **Jonas**, for helping my cells survive while I was sick or on holidays also thanks for being the second brain child from Jens and Charlotte sitting in a room with us when I needed to show them my shitty data again. Thank you, **Saskia**, for letting me rant about cloning and also giving me great tips, you are the actual cloning genius of the group. Thanks also to the rest of the Bosse group for being there while I was trying for months to clone my bacmids (blue-white-staining changed me).

Thank you, **Janine**, for comforting me when literally nothing worked and I was crying my tears out and also thank you for being the best roomie. Thank you, also for converting me from a cringe "flexitarian" to a proper giga chad vegan and for guiding me through my PhD and helping me with basically everything. Thank you, **Ronja**, for including me in stuff in the beginning and also just being fun to be around. Thank you, **Boris**, for letting me aliquot competent *E.coli* cells in the 4 °C room as the first-ever lab work in the group. . . but also for showing me around and welcoming me so nicely. Thank you, Kira, for ranting and working in the lab with me and at your place (nice flat btw.), also, please keep whipping out some "great" dad jokes. Thank you, **Lennart**, for being one of the unexpected funniest people I met and also for joining me in that empty office. Thank you, **Lars**, for not flaming us too hard in League of Legends even though we deserved it. Also, thank you for being, on one hand, a funny not taking stuff too serious guy, but also listening and genuinely trying to support me when I was at my lowest. Thank you, **Marius**, for being the

funniest person I know, I will truly miss the before 9 a.m. conversations that were often interrupted by the most unhinged Jürgen comments imaginable. Also, thank you for your unshakable values that you stand by and defend, thanks to you, my father is finally proud of me (he always wanted me to return to Linux). Thank you, **Jürgen**, for reteaching me basic cell culture etiquette and telling us the wildest stories. Thank you, **Simon**, for sitting there. . . In row 3. . . coding. . . . and also helping me set up Anaconda, I wouldn't be able to use OpenJDK without you. Also, thank you for being the only other person actually using LaTeX. Thanks, **Aqila**, (definitely not Aquila) for reaching out to me and wanting to do your master thesis with me. I hope I was able to teach you a few things. Thanks to everybody else I didn't mention by name for being there. I really love you guys and I will miss working with you.

Ich möchte mich auch bei meinen Eltern bedanken, die mich seit meiner Geburt unterstützen und für mich da sind. Auch wenn mein Vater lieber etwas über Physik von mir gelernt hätte, kann ich ihm noch das eine oder andere über Proteine und Massenspektrometrie beibringen. Danke Truong für deine Liebe und dein Vertrauen, danke, dass du immer für mich da bist. Einen besonderen Dank widme ich Norra und Fizz, da ich mich mental so gut wie lange nicht mehr gut fühle. Danke fürs Rummümmeln, Rumhoppeln und Rumliegen.

"Foul Tarnished, in search of the Elden Ring. Emboldened by the flame of ambition." — Margit the Fell Omen, **2022**, Elden Ring

Declaration of authorship |

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben. Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Hamburg, den _____ Unterschrift: _____