

Dissertation

Structural analysis of the *Flaviviridae* non-structural protein complex NS3-NS4A with native mass spectrometry

Dissertation with the aim of achieving the doctoral degree

Doctor rerum naturalium

Dr. rer. nat.

Faculty of Mathematics, Informatics and Natural Sciences

University of Hamburg

and

Leibniz Institute for Experimental Virology

(HPI)

Julia Lockhauserbäumer

Hamburg 2021

Supervisor
Charlotte Uetrecht (HPI)
Julia Kehr (University of Hamburg)

Disputation am 15.12.2021

Wenn es etwas gibt, das den Menschen erhebt, dann
besteht es darin, einen Freund zu haben.
Was ihn aber noch mehr erhebt, besteht darin,
ein Freund zu sein.

(Richard Wagner)

To my children,
Maximilian & Leonard

List of Publications

1. Yan H., Lockhauserbäumer J., Szekeres G.P., Mallagaray A., Creutzmacher R., Taube S., Peters T., Pagel K., Uetrecht C., (2021). *Protein Secondary Structure Affects Glycan Clustering in Native Mass Spectrometry*. MDPI Life 11(6), 554.
2. Anjanappa R., Garcia-Alai M., Kopicki J.-D., Lockhauserbäumer J., Aboelmagd M., Hinrichs J., Nemtanu I.M., Uetrecht C., Zacharias M., Springer S., Meijers R., (2020). *Structures of peptide-free and partially loaded MHC class I molecules reveal mechanisms of peptide selection*. Nat. Commun. 11;11(1):1314.
3. Wegener H., Mallagaray A., Schöne T., Peters T., Lockhauserbäumer J., Yan H., Uetrecht C., Hansman G.S., Taube S., (2017). *Human norovirus GII.4(MI001) P dimer binds fucosylated and sialylated carbohydrates*. Glycobiol. 27(11):1027-1037.
4. Winter I., Lockhauserbäumer J., Lallinger-Kube G., Schobert R., Ersfeld K., Biersack B., (2017). *Anti-trypanosomal activity of cationic N-heterocyclic carbene gold(I) complexes*. MBP 214:112-120.
5. Mahal K., Ahmad A., Schmitt F., Lockhauserbäumer J., et al., (2016). *Improved anticancer and antiparasitic activity of new lawsone Mannich bases*. Eur J Med Chem 126:421-431.
6. von der Heyde A., Lockhauserbäumer J., Uetrecht C., Elleuche S., (2015). *A hydrolase-based reporter system to uncover the protein splicing performance of an archaeal intein*. Appl Microbiol Biotechnol 99(18):7613-24.
7. Mallagaray A. Lockhauserbäumer J., J., Hansman, G.S., Uetrecht, C., Peters T., (2015). *Attachment of Norovirus to Histo Blood Group Antigens: A Cooperative Multistep Process*. Angew Chem Int. Ed. Engl. 5;54(41):12014-9.

Table of Content

I	List of Tables	I
II	List of Figures	II
III	Supplement List of Tables	V
IV	Supplement List of Figures	VII
V	Abbreviations.....	VIII
	Abstract	XI
	Zusammenfassung.....	XIII
1	Introduction	1
1.1	The variety of <i>Flaviviridae</i> with CSFV and HCV as representatives.....	1
1.2	Overview of the viral life-cycle of CSFV and HCV	5
1.3	Genome organization of CSFV and HCV	8
1.4	The multifunctional NS3-NS4A protein complex.....	10
1.5	Studying protein structure based on nano-ESI-MS and Top-Down analyses	14
1.5.1	Quadrupole time-of-flight (Q-TOF) instrument.....	15
1.5.2	Mechanisms of electrospray ionisation (ESI).....	16
1.5.3	TOF detection of analyte ions and data output.....	18
1.5.4	Q-pole precursor selection and MS2 collision induced dissociation	22
1.6	Research goal of the project.....	25
2	Material and Methods	26
2.1	Viral DNA vectors and plasmids	26
2.2	Expression and purification of NS3-NS4A	29
2.3	Determination of protein concentration by UV-VIS and absorption coefficient	30
2.4	Quality control of expressed proteins by SDS-Page.....	31
2.5	Western blot	33
2.6	FRET peptide assays.....	35
2.7	Proteomic analysis	37
2.7.1	Sample preparation for proteomic analysis	37
2.7.2	MS analysis.....	37
2.7.3	Data analysis	38
2.8	Native mass spectrometry	39
2.8.1	Sample preparation	39

2.8.2	Nano-ESI capillaries.....	39
2.8.3	NS3-NS4A sample preparation and ESI measurement.....	39
2.8.4	High Mass <i>Q-ToF 2</i> measurements.....	40
2.8.5	MS data analysis	40
2.9	Software and data banks.....	42
3	Results	43
3.1	Purification of the structurally intact NS3-NS4A proteins.....	45
3.1.1	Proteomics reveals the presence of the HCV NS3-NS4A protein	50
3.1.2	Identification of the HCV NS3-NS4A protein by Western blot.....	58
3.1.3	HCV NS3-protease shows activity when subjected to a FRET substrate	60
3.2	Effect of cysteine bridges on the CSFV protein assemblies.....	65
3.3	Auto-cleavage and proteolytic degradation of the NS3 protein monitored by native MS.....	71
3.3.1	Identification of the NS3-helicase (B3/4(a)) by Top-Down-Sequencing.....	76
3.3.2	Top-Down-Sequencing identifies different NS3-NS4A monomers	78
3.3.3	Native MS reveals the presence of several monomer sub-species in B2	86
3.3.4	The mass of the B1(a) monomer is close to the full-length protein.....	91
3.3.5	Slight differences between the B1 and B2 monomers.....	94
3.4	Protein complex dimerization plays a role for NS3-NS4A	102
3.4.1	MS2 control of the identified batch 2 dimers	106
3.4.2	Low mass ions in MS1 originate from the monomer and helicase and not from the dimer	111
3.4.3	Dimer formation is supported not only by Cys-bridges.....	115
3.5	MS studies of mutated CSFV NS3-NS4A proteins reveal structural differences....	118
3.5.1	The single mutant shows a high amount of unbound NS3-helicase	121
3.5.2	The double mutant shows a high level of auto-proteolytic degradation.....	124
3.6	ESI-MS based NS3-protease activity assay monitored substrate degradation...	130
3.6.1	ESI-TOF instrument conditions	132
3.6.2	Quenching the protease reaction.....	133
3.6.3	The effect of background signals on substrate peaks	135
3.6.4	Background correction and activity of the CSFV NS3-protease	137
4	Discussion	142
4.1	Native MS monitored the oligomerization and the auto-proteolytical activity of the CSFV WT NS3-NS4A complex.....	142

4.2	Top-Down-Sequencing revealed the coexistence of different NS3-protease cleavage products.....	149
4.3	The MS-based assay shows similar results to the published UV-Assay data.....	154
4.4	Possible implications of the results obtained from this thesis for future research..	160
5	References	162
	GHS information.....	182
6	Supplement.....	184
6.1	CSFV aa-sequence further information.....	184
6.2	CSFV protein batches 1 to 4.....	185
6.2.1	Batch 1 NS3-NS4A proteins	185
6.2.2	Batch 2 NS3-NS4A proteins	186
6.3	Top-Down identification of the released NS3-helicase, the NS4A cofactor and various sub-monomers of the WT protein	187
6.4	Stoichiometry of monomers and dimers.....	190
6.4.1	Dimerization of the NS3-NS4A protein	190
6.4.2	Batch 1 and 2 NS3-NS4A WT protein	191
6.4.3	Batch 2 NS3-NS4A WT protein.....	195
6.5	L45A mutations of the CSFV NS3-NS4A protein.....	203
6.6	L45A/Y47A mutation of the CSFV NS3-NS4A protein	205
6.7	MS based activity assay	209
6.8	HCV native MS1 and MS2 addition	210
	Note of Thanks.....	214
	Declaration of Honor.....	215

I List of Tables

Tab. 1 Expressed CSFV and HCV proteins.	26
Tab. 2 Extinction coefficient of CSFV and HCV NS3-NS4A protein at 280 nm.	30
Tab. 3 Recipe for the used sodium dodecyl-sulfate polyacrylamide gel (SDS-Pages).	31
Tab. 4 SDS-Page sample and running buffer, recipe adapted from Sambrook and Russel, 2006 [161].	32
Tab. 5 Used protocol for Western blot analyses.	34
Tab. 6 Used software and data processing tools	42
Tab. 7 Proteomic analyses based on proteolytical cleavage of the NS3-NS4A protein of HCV by trypsin revealed several <i>E. coli</i> contaminations after SEC chromatography.	56
Tab. 8 HCV NS3-NS4A protease activity assay based on the RET S1 FRET substrate. .	63
Tab. 9 Four protein batches of the CSFV _{WT} NS3-NS4A protein reveal the auto-catalytic cleavage.	72
Tab. 10 Main MS2 ions of the 14+ precursor signal of the NS3-helicase at 195 eV.	78
Tab. 11 Batch 2 monomers: overview of the specific MS2 fragment ions with their associated protein amino acid sequence (B2(a)-B2(j) monomer).	90
Tab. 12 Batch 1 MS2 ions that were identified while dissociating the B2(b) monomer. ...	92
Tab. 13 MS1 identifies the lack of C-/N-terminal aa of B1 and 2 monomers based on peak distances.	93
Tab. 14 Comparison of the cleavage products identified in the auto-proteolysis process of the NS3 protease with known cleavage motifs.	101
Tab. 15 Identification of different dimers in batch 1 (B1d) and their assumed missing a-segments.	103
Tab. 16 Identification of four different dimers in B2 by increasing energy in MS1 and MS2 studies.	106
Tab. 17 MS2 ions that were identified while dissociating the NS3-helicase of CSFV L45A/Y47A after three weeks + 4 °C.	121
Tab. 18 Calculations of the activity of the CSFV and HCV NS3-protease.	140
Tab. 19 Cleavage products of the NS3-auto-proteolyse and the corresponding cleavage motifs.	152
Tab. 20 Globally Harmonized System of Classification and Labelling of Chemicals (GHS) used in the present thesis.	182

II List of Figures

Fig. 1 The family Flaviviridae has four genera, which includes <i>Pestivirus</i> with the type classical swine fever virus (CSFV) and Hepacivirus with the human hepatitis C virus (HCV).	1
Fig. 2 World map of official CSFV spread in OIE member countries in 2021 [30].	3
Fig. 3 Schematic of the CSFV and HCV life-cycle.	5
Fig. 4 Organization of the CSFV and HCV genome shows the complete polyprotein and its associated encoding structural and non-proteins, including NS3-NS4A.	8
Fig. 5 Model of the ER membrane bilayer in association with the NS3-NS4A protein complex.	11
Fig. 6 Illustration of the NS3-helicase of CSFV and HCV.	12
Fig. 7 Simplified schematic depiction of the components of the Q-TOF2 nano-ESI-MS instrument and guidance of the analyte ions in positive ion mode.	15
Fig. 8 Electrospray ionisation shows different mechanisms of ion generation.	16
Fig. 9 Scheme of an orthogonal TOF analyser with pusher unit, reflectron and MCP detector.	19
Fig. 10 ESI-MS based mass spectra of the denatured and natively measured protein complex <i>H. pylori</i> urease, adapted from Pinkse <i>et al.</i> 2003 [146] and Heck <i>et al.</i> 2008 [147].	21
Fig. 11 Native ESI-MS and CID workflow and readout, adjusted based on Dülfer <i>et al.</i> 2019, [155].	23
Fig. 12 Scheme of the behaviour of the NS3-helicase/ATPase under native ESI-MS conditions up to Top-Down fragmentation.	24
Fig. 13 Construct of the SCP NS3-NS4A proteins.	26
Fig. 14 CSFV and HCV NS3-NS4A expression plasmids, which were used for protein production.	28
Fig. 15 Workflow for Western blot.	33
Fig. 16 Structure and function of the RET-S1 substrate.	35
Fig. 17 Expression construct and amino acid sequence of the CSFV NS3-NS4A target protein.	46
Fig. 18 Construct and amino acid sequence of the HCV NS3-NS4A target protein.	47
Fig. 19 Protein purification by size-exclusion chromatography (SEC) and control by a 10% SDS gel electrophoresis.	48
Fig. 20 Native ESI spectra of CSFV and HCV non-structural protein complex NS3-NS4A.	49
Fig. 21 The proteomics approach revealed significant peptide spectrum matches (PSMs) for the target NS3-NS4A protein of HCV.	53
Fig. 22 Western blot showed no HCV NS3-NS4A protein complex, which corresponds to the size of 74 kDa.	59
Fig. 23 Simplified FRET assay to prove the activity of the wild-type NS3-NS4A protein of HCV in a complex enzyme mixture.	62
Fig. 24 Native MS data of the CSFV NS3-NS4A oligomers released when protein concentration is 15 μ M and no DTT is added to the protein mixture (-) DTT.	65

Fig. 25 Oligomerization of the CSFV NS3-NS4A protein depends on ion acceleration energy, protein concentration and the accessibility of cysteine residues.....	67
Fig. 26 Oligomerization of the wild-type CSFV NS3-NS4A fusion protein in the presence of low concentration of DTT.	68
Fig. 27 Crystal structure of the dimeric CSFV NS3-NS4A protease.....	69
Fig. 28 Crystal structure of the monomeric NS3-NS4A protein complex shows the internal cleavage sides of NS3 (Leu159/Lys160 and Leu192/Met193).	71
Fig. 29 The NS3-NS4A protein of CSFV - overview of protein degradation within the four different batches.....	73
Fig. 30 Overview of the NS3-NS4A protein species identified with native MS as of this point.....	75
Fig. 31 B4(a) subunit dissociation with MS2 CID.....	76
Fig. 32 Amino acid sequence and associated crystal structure of the released NS3-helicase subunit show the low mass MS2 ions.	77
Fig. 33 B3(b) dissociation revealed the same C-terminal peptide ions compared to B4(a).	79
Fig. 34 Identification of the C -terminal y19 low mass ion for the monomers B2(a)/(b).	80
Fig. 35 B2(a) monomer CID fragmentation revealed a lack of the C-terminal NS4A cofactor.	82
Fig. 36 B2(b) lack of N- and C- terminal amino acids revealed by its specific fragment ion pattern.	83
Fig. 37 First appearing MS2 low mass ions of the identified monomers B2(a) and B2(b) at 100 eV are doubly charged, main ions detected at 195 eV are singly charged.	84
Fig. 38 Crystal structure and related amino acid sequence of the B2(a) and (b) monomers show the location of the main intense MS2 backbone ions.	85
Fig. 39 MS1 analyses of the B2 NS3-NS4A protein identified the main and several sub-monomers at 10 eV.	86
Fig. 40: Non-fragmented precursor in MS2 and MS1 peaks of B2(a) and B2(b) show an overlay of various NS3-NS4A monomeric peak species.....	87
Fig. 41 Specific MS2 fragment ions of all identified B2 monomers (B2(a)-(g)) at 195 eV.....	88
Fig. 42 MS1 analyses of the B1 monomer identified the main and two sub-monomers at 10 eV.....	91
Fig. 43 The ratio between the different batch 1 and 2 monomers shows the hierarchical distribution of the identified protein species.	95
Fig. 44 Based on their intensities, the B2(a) and B2(b) main peaks (17+) show a similar distribution of the associated sub-species at 10 eV, while different activation energies revealed strong changes.	96
Fig. 45 Entire peak width of the B2(a) and B2(b) including sub-monomers showed a contrary distribution at different energies (10 eV to 195 eV).....	98
Fig. 46 The majority of the y19 ion originate from the bigger monomer B2(a).	100
Fig. 47 The monomer B1(a) ((-)aaL752 species) tend to form the homo-dimer B2d(a).	102
Fig. 48 Huge extent of dimerization at the first day of the auto-proteolysis test (dominant batch 1 monomer (-)aaL752 and homo-dimer 2x(-)aaL752).....	104

Fig. 49 Distribution of various dimer species in batch 2 after 1 week of auto-proteolysis.	105
Fig. 50 Control of the authenticity of the B2 dimer peaks at different energies revealed four species.	107
Fig. 51 MS2 dissociation of the 24+ precursor of the two main dimer species B2d(a) and B2d(c) show the same main fragment ions (y19 and y22) as the B2 monomers.	109
Fig. 52 Native MS studies show that an increase of ion activation energy in MS1 mode lead to the dissociation of the monomer and unbound NS3-helicase.	111
Fig. 53 Ratios at different MS1 energies show significant changes between main monomers and main dimers and reveal the dissociation of the unbound NS3-helicase.	112
Fig. 54 The auto-proteolytical activity of the NS3-protease including protein monomers and dimers leads to the release of the NS3-helicase subunit.	115
Fig. 55 Course of the auto-proteolysis process of the NS3 complex: decrease and increase of the dominant NS3 proteins is shown over three weeks.	116
Fig. 56 Native MS shows structural differences between the mutated and WT NS3-NS4A protein.	119
Fig. 57 Crystal structure of the DM L45A/Y47A NS3-helicase (55.8 kDa) revealed the interaction with the complete C-terminal NS4A-cofactor.	120
Fig. 58 L45A exhibits a lower degree of dimerization.	122
Fig. 59 L45A reveals a huge amount of unbound NS3-helicase subunit.	123
Fig. 60 Dynamic structural changes of the L45A/Y47A mutant over time.	125
Fig. 61 The double mutation (DM) L45A/Y47A also demonstrates a high level of auto-proteolytic activity at + 4 °C.	126
Fig. 62 L45A/Y47A shows protease activity even when stored three weeks at + 4 °C.	128
Fig. 63 Schematic of the ESI-MS based NS3-protease activity assay.	131
Fig. 64 ESI-TOF signals of the RET S1 FRET substrate and the CSFV NS3-NS4A _{WT} protease with and without collision energy.	132
Fig. 65 Quenching of NS3-NS4A protease reaction by 1% (v/v) formic acid (FA).	134
Fig. 66 Background noise during ESI-QTOF measurements: raw spectra of two time points indicate the impact on substrate signals.	135
Fig. 67 Background noise during ESI-QTOF measurements: absolute peak intensities of T = 0 to T = 20 min.	136
Fig. 68 Activity of the CSFV _{WT} NS3- protease in presence of the FRET substrate (RET S1 peptide).	138
Fig. 69 Activity of the HCV _{WT} NS3-protease in presence of the RET S1 FRET substrate.	140
Fig. 70 Scheme of the possible course of the auto-cleavage and NS3-NS4A complex degradation on the monomeric level based on PDB 5MZ4.	153

III Supplement List of Tables

S. Tab. 1 MS1 based molecular weight (MW) of the batch 1 (B) monomers.....	185
S. Tab. 2 MS1 based molecular weight (MW) of the batch 1 (B1) dimers.....	185
S. Tab. 3 MS1 based molecular weight (MW) of the batch 2 (B2) main monomers and dimers.	186
S. Tab. 4 MS1 based ratio between B1(a)-(c) monomers (MS1 10 eV).....	191
S. Tab. 5 Ratio and distribution of the batch 1 (B1) and batch 2 (B2) monomers at 10 eV.	191
S. Tab. 6 Comparison of peak intensities and peak areas of B1(a) monomer and B1d(a) dimer show similar ratios for the monomer and dimer at 10 eV in MS1 mode.	192
S. Tab. 7 Dimerization of B1 proteins shows a lack of different <i>C</i> - and <i>N-terminal</i> amino acids based on their specific peak distances in MS1 mode.	193
S. Tab. 8 Batch 2 monomers main peak (17+) and their associated sub-peak distribution.	195
S. Tab. 9 Changes of the peak intensity of batch 2 main monomers and sub-monomers at different acceleration energies in M1 mode (10 eV - 170 eV).	195
S. Tab. 10 Associated charge states to the peak areas of batch 2 main monomers B2(a) and B2(b) at different acceleration energies in M1 mode (10 eV - 195 eV).	196
S. Tab. 11 Changes of the peak areas of batch 2 main monomers B2(a) and B2(b) at different acceleration energies in M1 mode (10 eV - 195 eV).	197
S. Tab. 12 (a) Ratio between main monomer/main dimer of B2(a)/B2d(a) and B2(b)/B2d(c): Applied MS1 energies show significant changes of the distribution of peak areas.	198
S. Tab. 13 (b) Ratio between main monomer/main dimer of B2(a)/B2d(a) and B2(b)/B2d(c): Applied MS1 energies show significant changes of the distribution of peak areas.	199
S. Tab. 14 Identification of four different dimers in B2 by increasing energy in MS1 and MS2 studies.....	201
S. Tab. 15 Course of the auto-proteolytical process of the NS3-protease.....	201
S. Tab. 16 (a) CSFV _{L45A} mutant: Absolute (Abs.) peak areas (PA) of the monomer, dimer and the free NS3-helicase of the NS3-NS4A fusion protein.....	203
S. Tab. 17 (b) CSFV _{L45A} mutant: Relative (Rel.) peak areas (PA) of the monomer and dimer of the fusion protein.....	203
S. Tab. 18 (c) CSFV _{L45A} mutant: Averaged peak areas of charge states of the monomer and dimer.....	204
S. Tab. 19 (d) CSFV _{L45A} mutant: Relative (Rel.) peak areas (PA) of the monomer, dimer and free NS3-helicase.....	204
S. Tab. 20 (e) CSFV _{L45A} mutant: Average of the relative peak areas (PA) of the monomer, dimer and free NS3-helicase.....	204
S. Tab. 21 (a) CSFV _{L45A} /Y47A sample storage test over a four weeks period (1d, 2w and 4w) at + 4 °C.....	206
S. Tab. 22 (b) CSFV _{L45A} /Y47A sample storage test over a four weeks period (1d, 2w and 4w) at + 4 °C.....	206

S. Tab. 23 (a) CSFVL45A/Y47A double mutant (DM) shows protease activity..... 207
 S. Tab. 24 (b) CSFVL45A/Y47A double mutant (DM) shows protease activity. 208
 S. Tab. 25 (c) CSFVL45A/Y47A double mutant (DM) shows protease activity..... 208
 S. Tab. 26 ESI-TOF data: Basis for calculations of CSFV NS3-NS4_{WT} protease activity.
 209
 S. Tab. 27 ESI-TOF data: Basis for calculations of HCV NS3-NS4_{WT} protease activity.
 209

IV Supplement List of Figures

S. Fig. 1 Amino acid sequence of the CSFV _{WT} NS3-NS4A (single chain protease, SCP).	184
S. Fig. 2 Low amount of free NS3-helicase in batch 2 (after 1 week at + 4 °C) was identified in MS1 mode at 10 eV and in MS2 mode while applying different CID energies.....	187
S. Fig. 3 Dissociation of the NS4A-cofactor from the NS3-NS4A complex is shown by CID fragmentation of the 17+ precursor peak of B2(a) in MS2 mode.....	188
S. Fig. 4 Identification of B1(b) and B1(c) MS2 y- and b-ions at 195 eV.	189
S. Fig. 5 Possible location of the NS3 dimerization region at an exposed beta sheet structure (I292TY).....	190
S. Fig. 6 Huge extent of dimerization of batch 1 monomers.	194
S. Fig. 7 Native spectra of the batch 2 NS3-NS4A monomers and dimers reveal strong impact of MS1 energies on the dimer shape.....	197
S. Fig. 8 Changes in the peak intensity and peak width of B2 monomers and dimers based on different ion activation energies 10 to 200 eV.	200
S. Fig. 9 Dimerization of batch 1 and 2 N3-NS4A monomers.....	200
S. Fig. 10 MS1 data show the unbound C-terminal NS4A cofactor segment STAENALL (amino acids aa752-745, 817 Da) within batch 2.....	202
S. Fig. 11 Dissociation of the 55.8 kDa protein of the DM L45A/Y47A revealed the NS3-helicase with complete C-terminal NS4A-cofactor.....	205
S. Fig. 12 Auto-proteolytic activity of the double mutant (DM) L45A/Y47A at + 4 °C.....	207
S. Fig. 13 Activation of the protein complex in the HCV sample with different energies in the MS1 mode shows the dissociation of an approx. 51 kDa protein at 100 eV.....	210
S. Fig. 14 Comparison of 150 eV in MS1modd with MS2 CID (19+ precursors peak) of the protein complex in the HCV sample shows the dissociation of the 51 kDa protein in each case.....	211
S. Fig. 15 Top-down dissociation of the 51 kDa protein at 170 eV (13+ 3900 m / z) shows specific low mass peptides for E. coli IpdA as well as for a sequentially shortened HCV NS3 protein.	211
S. Fig. 16 Top-Down CID fragmentation of the 19+ protein peak (5434 m/z) identified certain low mass peptides of a possibly sequentially shortened HCV NS3-NS4A protein (51 kDa) as well as an E. coli protein (IpdA 51 kDa), which was already identified during the proteomics analysis.....	212
S. Fig. 17 Zoom into the low mass CID fragments of the sequentially shortened HCV NS3 protein.	213

V Abbreviations

aa	amino acid
AAS	Amino acid sequence
CSFV	Classical swine fever virus
Cov.	Coverage
Da	Dalton, unit of mass
DLS	Dynamic light scattering
DM	Double mutant (CSFV NS3-NS4A), protein contains two different mutations
ESI	Electrospray ionisation
FRET	Fluorescence resonance energy transfer
FWHM	full width at half maximum
G	Genogroup, phylogenetic classification of viruses
HCV	Human hepatitis C virus (hHCV)
HDX-MS	hydrogen/deuterium exchange mass spectrometry
IMS-MS	Ion-mobility mass spectrometry
<i>m</i>	mass
MCP	Microchannel plate, ion detector
MRM	multiple reaction monitoring
MS	Mass spectrometry
MS1	native MS analysis
MS2	Tandem MS analysis (MS/MS)
MT	Mutant, mutated protein
MW	Molecular weight
NMR	Nuclear magnetic resonance
NS	Non-structural protein
OA	Open access
OIE	World Organization for Animal Health
ORF	Open reading frames

PMC	PubMed Central (free digital archives of open access full-text articles)
PSMs	Peptide spectrum matches
SAXS	Small-angle x-ray scattering
SM	Single mutant (CSFV NS3-NS4A)
ssRNA	Single-stranded ribonucleic acid
SV	Sapovirus
t	Time
TBE	Tick-borne encephalitis (FSME = Frühsommer-Meningoenzephalitis)
TOF	Time of flight, flight time analyzer
Q	Quadrupole analyzer
q	Charge
Q-TOF2	Quadrupole-time-of-flight-mass spectrometer 2
WT	Wild-type (CSFV NS3-NS4A), without NS4A mutations
z	Number of charges
μM	Micro molar, $\mu\text{mol/L}$

E. coli contaminants (gene names)

arnA	Bifunctional polymyxin resistance protein ArnA
sucA	2-oxoglutarate dehydrogenase component
glnD	Bifunctional uridylyltransferase/uridylyl-removing enzyme
groL	60 kDa chaperonin (Cpn60), alternative name: GroEL protein)
lpdA	Dihydrolipoyl dehydrogenase
sucB	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex
dnaJ	Chaperone protein DnaJ
ssb	Single-stranded DNA-binding protein
rplI	50S ribosomal protein
iclR	Transcriptional repressor IclR
thiD	Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase

Abstract

The focus of the thesis was to develop a suitable native electrospray mass spectrometry (ESI-MS) approach for studying specific enzymes that play an important role in the *Flaviviridae* replication and infection pathway. The so-called non-structural proteins (NS-proteins) are involved in that process and therefore important drug targets. Especially, the multifunctional non-structural protein complex NS3-NS4A is target of diverse research studies, due to the variety of its cellular functions. Among other things, due to its protease function this complex is responsible for the proteolytic cleavage of the polyprotein, which leads to the release and activation of other proteins that play an important role in the viral life-cycle. The NS3-NS4A protein is also involved in the formation of the replicas complex, which is necessary for genome replication. If the interaction between the different NS3-NS4A “subunits” (NS3-protease and helicase/ATPase function + NS4A cofactor) can be blocked by certain active substances, the replication process of the viruses can be stopped and the viruses won’t survive.

As state-of-the-art technology, native ESI-MS provided detailed information on the composition and stability of the NS3-NS4A protein on a molecular level. It is a precise and fast analysis method and is therefore a useful addition to other biochemical methods such as crystallography or SAXS (small-angle x-ray scattering) experiments. The presented MS investigations were mainly performed with the NS3-NS4A fusion enzyme of the classical swine fever virus (CSFV). In addition, some analyzes with the NS3-NS4A protein of the human hepatitis C virus (HCV) were carried out. Of special interest was the analysis of the protease function of the NS3-NS4A complex and the dynamics within the protein, such as stoichiometry, interaction of the three protein “subunits” (NS3-protease, NS3-helicase and NS4A cofactor).

Due to the protease activity, the results of this study highlight several structural differences within the wild-type CSFV protein complex and in comparison to mutated proteins. Using native and Top-Down MS, it was possible to identify different monomeric and dimeric species of the CSFV NS3-NS4A protein as well as to monitor various cleavage products of the intact complex, that were produced over a time range of four weeks by its own NS3-

protease activity. The data suggest that this protease tends to self-degrade under the conditions used after the release of a structurally intact NS3-helicase via cleavage in a minor cleavage site. This auto-proteolytic activity is assumed to be temperature depended.

In addition, the first experiments with two mutated proteins of CSFV, in which an amino acid exchange was carried out in the NS4A-kink region to destabilize the NS3-protease function, could also demonstrate a proteolytic activity of the protease. This seems to be slower compared to the wild-type (WT). Furthermore, the mutations lead to clear structural differences in the protein complex compared to each other and with the WT. The proteins are much more dynamic and unstable than WT proteins.

Finally, a first activity study showed that ESI-MS can be used to detect the protease activity of the NS3-protease of CSFV_{WT} and HCV_{WT}. The MS-based activity study revealed, that the protease within the complete protein complex is able to degrade a served substrate, while this activity can be stopped by acidification. Results are comparable to conventional UV-activity tests such as the fluorescence resonance energy transfer (FRET)-assay.

Overall, the results of this study show that both techniques, native MS and Top-Down, are sensible methods for analysing the NS3-NS4A protein complex.

Zusammenfassung

Im Vordergrund der vorliegenden Dissertation steht die massenspektrometrische (MS) Analyse des Degradations-Prozesses von dem viralen nicht-struktur-Protein (NS-Protein) NS3-NS4A. Bei diesem NS-Protein handelt es sich um einen multifunktionalen Proteinkomplex, der aufgrund seiner vielfältigen zellulären Funktionen im Fokus verschiedener wissenschaftlicher Studien zur Entwicklung von antiviralen Wirkstoffen steht. Dieser Komplex ist unter anderem für die Spaltung des Polyproteins und somit für die Freisetzung weiterer wichtiger Proteine des viralen Lebenszyklus wichtig. Er spielt eine essentielle Rolle bei der Bildung des Replikations-Komplexes.

In der vorliegenden Arbeit wird native Elektrospray-basierte MS (ESI-MS) als weitere Analysemethode für diesen Proteinkomplex verwendet. Durch dieses sensitive und präzise Analyseverfahren, können Ergebnisse anderer Studien, die auf anderen biochemischen Methoden beruhen (z.B. Kristallographie oder *small-angle x-ray scattering*, SAXS) ergänzt werden. In dieser Arbeit wird das NS3-NS4A Fusionsprotein des klassischen Schweinepestvirus (CSFV) und des humanen Hepatitis C Virus (HCV) verglichen.

Das Hauptaugenmerk liegt auf einer detaillierten Analyse des Degradations-Prozesses des gesamten Proteinkomplexes von CSFV, welche durch die NS3-Protease-Funktion durchgeführt wird. Hier wird besonders der Einfluss des stabilisierenden NS4A Cofaktors untersucht.

Zusammengenommen zeigen die erhaltenen Ergebnisse dieser Studie mehrere strukturelle Unterschiede im CSFV-Proteinkomplex. Zunächst konnte mittels MS die Autoprotolysereaktion der NS3-Protease beobachtet werden. Es zeigte sich, dass diese autokatalytische Spaltung innerhalb des Proteins zu einem massiven Abbau des Komplexes führt. Dabei wurde beobachtet, dass die strukturell intakte NS3-Helikase-Untereinheit freigesetzt wird. Mittels Top-Down-Analysen wurden dabei verschiedene Spaltprodukte identifiziert und charakterisiert.

Darüber hinaus konnte in ersten Experimenten mit mutierten Proteinen, die eine labilere Cofaktor-Stabilität simulieren sollen, ebenfalls eine autokatalytische Aktivität des NS3-NS4A Proteinkomplexes nachgewiesen werden. Der Abbau des Proteinkomplexes durch die NS3-Protease ist im Vergleich zum Wildtyp (WT) jedoch verlangsamt. Darüber hinaus führen die Mutationen zu deutlichen strukturellen Unterschieden im Proteinkomplex. Die

Proteine sind wesentlich dynamischer und instabiler als das WT-Protein. Schließlich zeigte eine erste Aktivitätsstudie, dass ESI-MS verwendet werden kann, um die NS3-Proteaseaktivität von CSFV_{WT} und HCV_{WT} unter Verwendung eines FRET-Peptidsubstrats nachzuweisen. Die erhaltenen Ergebnisse sind vergleichbar mit herkömmlichen UV-basierten Aktivitätstests.

1 Introduction

1.1 The variety of *Flaviviridae* with CSFV and HCV as representatives

The family *Flaviviridae* includes small viruses, which have a positive-sense single-strand RNA (ssRNA) genome of 9.0 to 13.0 kilo-bases (kb) [1]. The genome is enclosed by an icosahedral nucleocapsid that is surrounded by a lipid bi-layered envelope (Fig. 1) [2-5]. The family consist of four genera, which include 89 species identified so far [4, 6]. The corresponding genera are *Flavivirus* (with 53 species), *Hepacivirus* (with 14 species), *Pegivirus* (with 11 species) and *Pestivirus* (with 11 species) [4, 7]. The members of *Flaviviridae* are mostly host-specific and infect birds or mammals, which includes important human and animal pathogens [4]. Known vectors are mosquitos or ticks [8]. A very well-known pathogen belonging to the genus *Flavivirus* is the mosquito-borne dengue virus (DENV). These viruses infect an estimated 400 million people each year and spread globally, particularly in the southern hemisphere [9].

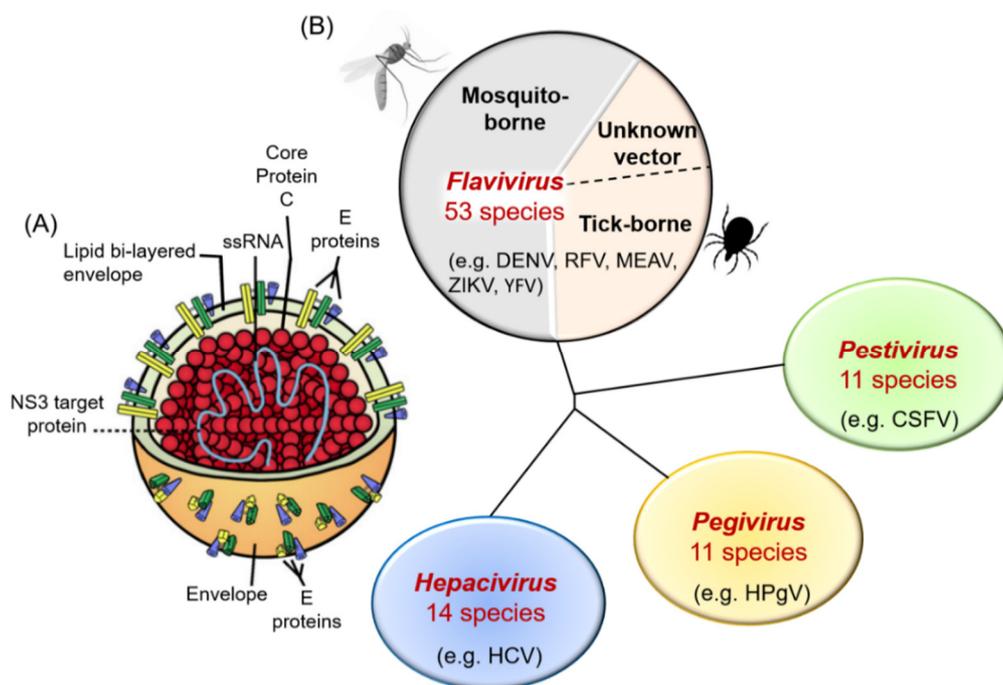


Fig. 1 The family *Flaviviridae* has four genera, which includes *Pestivirus* with the type classical swine fever virus (CSFV) and *Hepacivirus* with the human hepatitis C virus (HCV).

(A) Virion structure of CSFV, adjusted with permission from Elsevier (license terms and conditions, Copyright Clearance Center, Inc. ("CCC")), Beer *et al.* 2007 [10], Copyright 2006. (B) Map of the four different *Flaviviridae* genera.

The introduction of other mosquito-borne *Flaviviruses*, such as the zika virus (ZIKV) into the western hemisphere, has potential for a rapid global spread, with particular regard to climate change [11, 12]. The yellow fever virus (YFV) is the most severe mosquito-borne infection in the tropics, which lead to considerable high amounts of annually deaths (up to 60 000), especially in South America and Africa [13, 14]. Despite the availability of an effective vaccine, the infection poses a serious public health challenge [15].

However, several human diseases in Europe are caused by viruses, which were transmitted by tick bites and lead to neurological diseases. Especially, the tick-borne encephalitis virus (TBEV) is present in more than 20 European countries [16]. Some tick-born infections by *Flaviviruses* are only known for animal diseases, such as the royal farm virus (RFV), deadly for small rodents or the meaban virus (MEAV), which infects seabirds like herring gulls even in Europe [17, 18].

The focus of the present study is the classical swine fever virus (CSFV) and the human hepatitis C virus (HCV). CSFV belongs to the genus *Pestivirus* that contains several important animal pathogens. The virus is the causative agent of classic swine fever, also known as hog cholera [19-21]. This highly contagious and fatal viral disease is of global economic importance. A potential vector of the disease is still unknown [22-24]. It is not believed to be arthropod-born, as is the case with the African swine fever virus (ASFV), which is a tick-born disease [25]. Aerosol transmission of CSFV is believed to play an important role in the route of infection [24, 26]. It is to be expected that this transmission route is mainly influenced by geographical but also by climatological parameters [24], including current climate change.

CSFV was identified for the first time in Ohio in 1833. It is believed that the virus was able to spread across species through mutations in another *Pestivirus* [6, 27]. At the beginning of worldwide intensive pig production in the 20th century, classic swine fever developed into one of the most important pig diseases in industrialized countries. The disease occurs in both, wild and domestic pigs and occasionally in cattle. Swine fever is regularly diagnosed in Africa [28]. Nevertheless, further cases of the disease continue to occur worldwide. Especially in Asia as well as South and Central America and Eastern Europe CSFV is still endemic [6, 29]. In Western Europe, there has been a progressive eradication of CSFV with vaccination. However, the virus is occasionally reintroduced into domestic pigs by imports from abroad, wild boars or by introduction of contaminated pork [6, 27, 28]. Therefor in

Europe, there are regular new outbreaks, which lead to the fact that many pigs are removed from the ecosystem by official order with significant economic losses, respectively [28, 29]. Responsible agency is the Office International Epizooties (OIE, World Organization for Animal Health) in Paris, to which outbreaks of CSFV infections must be reported immediately [21].

Continuous vaccination of domesticated and wild animal populations is therefore an important part of the fight against the spread of the virus. Hitherto, based on the most recent reports from May 2021 by the OIE, most member states, with the exception of Romania, have been advertised as free of CSFV (Fig. 2). However, the re-entry of the virus into pig herds could happen very easily. Since the virus transmission route is not completely understood and new outbreaks occur repeatedly, it cannot be assumed that the pathogen is totally eradicated world-wide.

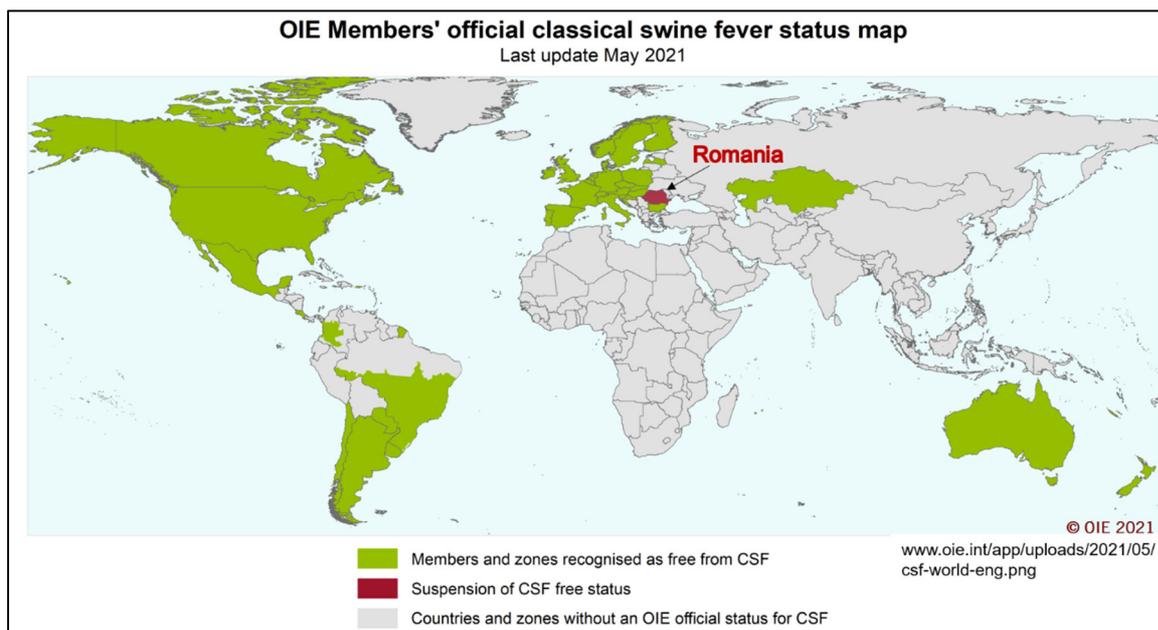


Fig. 2 World map of official CSFV spread in OIE member countries in 2021 [30].

Open access (OA) information and free of license with respect to the OIE's values and principles, World Organization for Animal Health (OIE); (OIE, Copyright 2021).

HCV belongs to the genus *Hepaciviruses* and is the causative agent of hepatitis C, a liver disease in humans [31, 32]. HCV is present in all World Health Organization (WHO) countries [33]. However, the highest degree of spread of the disease is in the European and Eastern Mediterranean Regions. According to WHO information, currently 12 million people are chronically infected in each region [33]. Well known is the transmission route of HCV through blood [34]. Certainly, there must be other transmission routes, especially

arthropod-borne, that have not been in the focus of attention so far [35, 36]. *Culex pipiens* is a mosquito originated from the northern regions of the United States, but in recent times is also wide-spread in European countries [37, 38]. This mosquito shows a high vectorial capacity for transmission of diseases like the West Nile virus (WNV) [39] and, according to new studies, is back in focus as a potential transmitter of HCV [35, 40].

1.2 Overview of the viral life-cycle of CSFV and HCV

The CSFV life-cycle is characterized by interactions with different host cell proteins, where the viruses enter host cells through the endocytic pathway [41]. The fusion between the host cell membrane and the virus envelope is pH-dependent and is triggered by the acidification of endosomes [42]. As published for the entry of the *Pestivirus* member bovine viral diarrhea virus (BVDV), a pH-dependent cell entry suggests that the virus enters by receptor-mediated endocytosis (Fig. 3 (A)) and endosomal fusion [43].

For instance, as described in the first step of the life-cycle (Fig. 3 (A)), heparan sulfate (HS) non-protein receptors were identified as cell attachment factors for CSFV [42, 44, 45]. The entry is cholesterol and dynamin dependent. As shown in the second step of the life-cycle, the entering process of endocytosis is clathrin-mediated like described for other flaviviruses, e.g. DENV where the host cell entry was shown by tracking fluorescently labeled DENV particles in living cells [41, 43, 46].

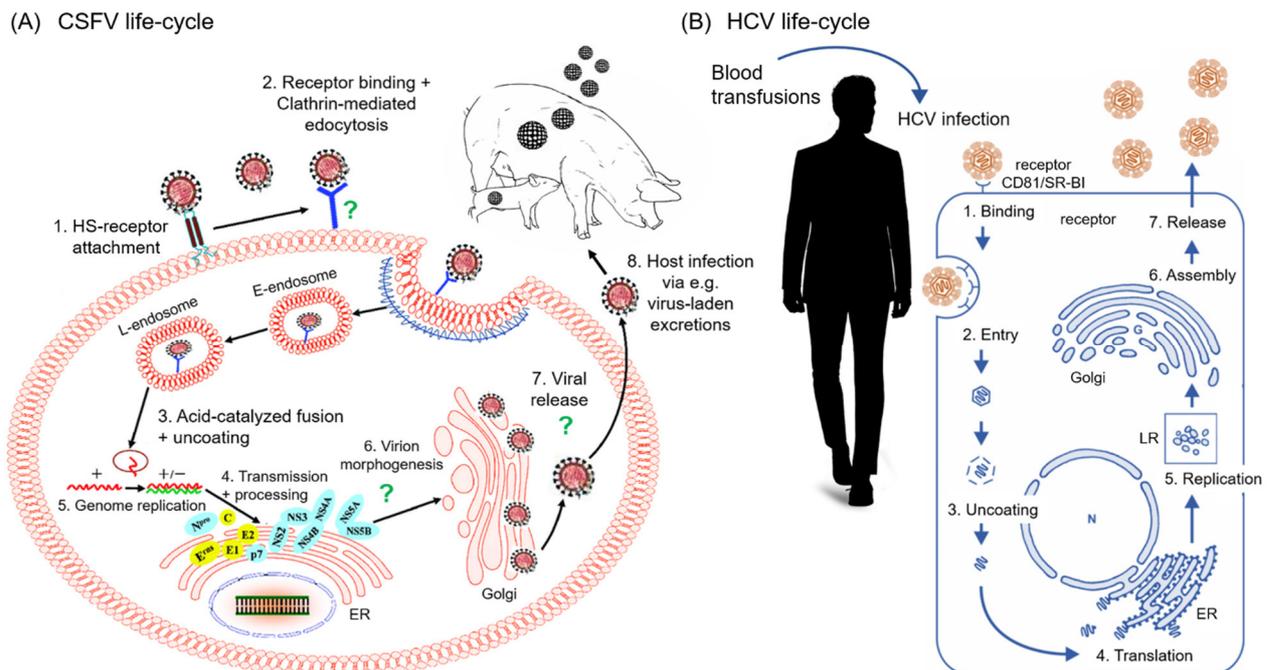


Fig. 3 Schematic of the CSFV and HCV life-cycle.

Figure (A) is adapted based on open access (OA) terms and conditions (licensee MDPI, Basel, Switzerland with terms and conditions of the creative commons attribution license (<http://creativecommons.org/licenses/by/4.0/>), Li *et al.* 2017 [42]. Figure (B) is adapted with permission from Elsevier (license terms and conditions, Copyright Clearance Center, Inc. ("CCC")), Bartenschlager *et al.*, 2004 [47], Copyright 2004) and Chevaliez and Pawlotsky 2006, [31].

Anyway, the viral envelope glycoproteins (Erns, E1, and E2) are involved in the attachment and entering process [42, 45]. The third step in the life-cycle shows, that the CSFV particles are transferred to early and late endosomes (E- and L-endosome) before releasing its ssRNA after viral uncoating [41]. As described for the fourth step, the viral genome then serves as template for the translation of the precursor polyprotein, which is inserted into the endoplasmic reticulum (ER) membrane where it is further processed by host and viral proteases.

Currently, a new study by Zhu *et al.* 2019, demonstrated that CSFV induces ER stress-mediated autophagy to maintain replication. However, the exact underlying mechanisms between ER stress, autophagy, and viral replication remain unclear [48]. The produced polyprotein is cleaved by cellular and viral proteases in order to release the structural proteins (core protein C and the envelope proteins Erns, E1, E2) as well as the non-structural proteins (Npro, p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B) [49-51]. The structural protein region of the polyprotein contains signal peptidase cleavage sites [1, 52, 53]. Processing of the polyprotein is initially mediated by a signal peptidase, which also ensures cleavage at the E1-E2 junction, where E1 and E2 undergo several maturation steps. This is followed by different cis- and trans-cleavages of the non-structural protein regions of the viral polyprotein via the responsible proteases (including NS3) [42, 54].

The cleavage leads to the formation and release of further functional proteins, which are important for the formation of the replication complex and RNA replication on ER membranes, as shown in the fifth step [52]. Here, the NS3-NS4A protein complex is involved in replication [55].

For the sixth step, which is the virion morphogenesis the structural core protein, the capsid proteins and the envelope glycoproteins are known to be important. Additionally, nonstructural (NS) proteins are also involved in the virion formation [52, 56, 57]. Here, the p7 protein and the NS2-3-NS4A complex plays an essential role. However, the formation of virus particles only mediated by NS2-3 and NS4A is controversially discussed [54, 56-58]. It is believed, that *Pestivirus* particles are started to be formed in the ER where assembly is completed in the ER and secretory pathway in the cis-Golgi before their transport through the cell membrane [59, 60]. In the seventh step, the virions are then released from the host cells and are infectious rapidly after budding [60].

Various routes of transmission are known for the discharged virus particle. The virus is excreted in nasal and oral mucosa (saliva) as well as urine and feces, transplacental transmission has also been reported [24, 61, 62].

The life-cycle of HCV (Fig. 3 (B)) is comparable to the CSFV cycle. The envelope glycoproteins E1 and E2 of HCV are necessary for the viral cell entry and membrane fusion process [45, 63]. Various host-specific cell surface molecules are proposed as potential interaction partners that mediate HCV binding [31, 64]. Important here are both, the human tetraspanin CD81 receptor and the human scavenger receptor SR-BI, which play a functional role in the first step of cell entry and in the primary cell attachment process [63, 65, 66]. After entering the target cell, the nucleocapsid is released into the cytoplasm. Similar to the third step in the CSFV life-cycle, this process is also pH and endocytosis-dependent in HCV [64]. Similar to the fourth step of the CSFV life-cycle, the polyprotein expression is associated with the ER. While polyprotein processing of most of the NS proteins are associated with subcellular membranes and the ER [67, 68].

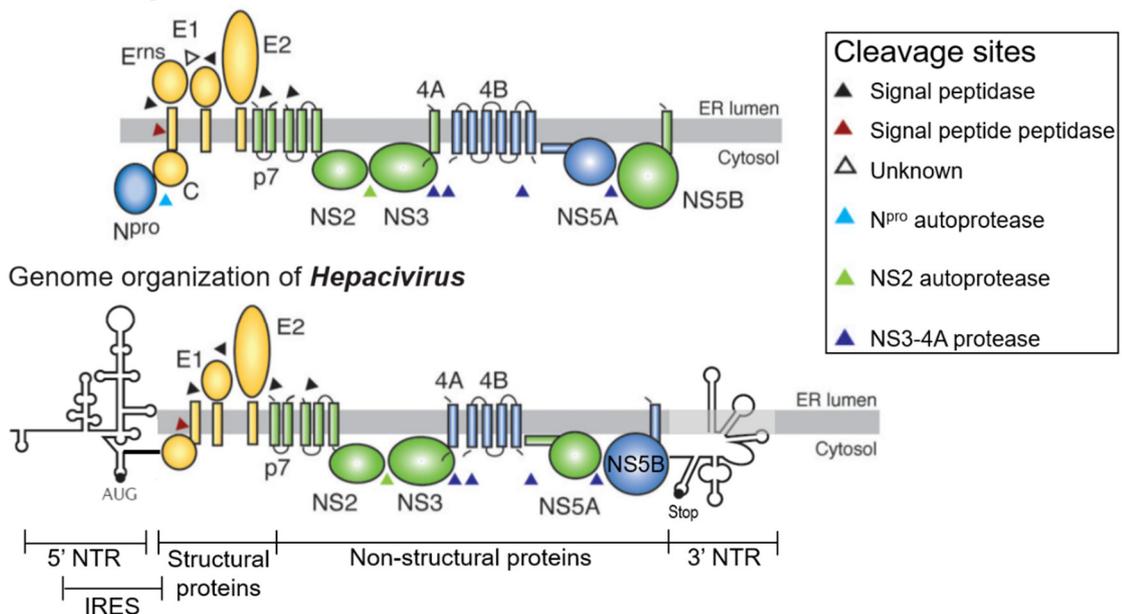
Here shown in the fifth step, the NS proteins and host proteins mediate the formation of the RNA replication complex on lipid raft membranes (LR) through protein-protein interactions [68, 69].

The virion morphogenesis is shown as sixth step of the life-cycle and believed to take place in association with membrane-apposed lipid droplets, at which replication complexes also accumulate [52, 70]. Further, it was also shown that assembly and maturation occur in the ER and post-ER compartments (Golgi), where particles appear to become associated with very low-density lipoproteins (VLDL) during assembly [71]. Virions are infectious immediately after budding [72]. After viral release contaminated blood, which can be transmitted by mosquito bites, is a major source of infection. HCV is transmitted via blood transfusions, intravenous medication or by perinatal transmission, for example [73].

1.3 Genome organization of CSFV and HCV

The organization of the genome is pretty similar between CSFV and HCV, however they differ in their replication strategies [6, 47, 48, 68]. The genome of both viruses consists of a single open reading frame (ORF) [74], which is flanked by non-translated regions (NTRs) at the 5' and 3' ends. A single precursor polyprotein is generated (Fig. 4), where translation is under the control of the internal ribosomal entry site (IRES) [1, 75].

(A) Genome organization of *Pestivirus*



(B)

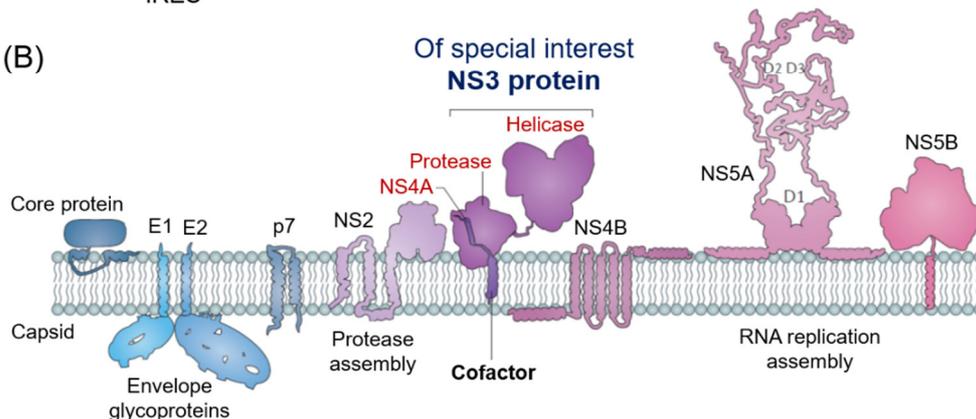


Fig. 4 Organization of the CSFV and HCV genome shows the complete polyprotein and its associated encoding structural and non-proteins, including NS3-NS4A.

(A) Genome organization of *Pestivirus* and *Hepacivirus*. (B) Association of structural and non-structural proteins with the ER membrane. The translation of the ORF is mediated by the internal ribosomal entry site (IRES). Figure (A and B) are adjusted with permission from Springer Nature (license terms and conditions, Springer Nature and Copyright Clearance Center); (A) based on Murray *et al.* 2008, Copyright 2008 and (B) based on Neufeldt *et al.* 2018, Copyright 2018, sources [1, 52].

Starting from 5' NTR to 3' NTR the polyprotein consists of the region, belonging to the structural proteins, including the three envelope glycoproteins (Erns, E1, and E2) and the core protein C [49, 52]. This is followed by the non-structural protein region, starting with p7 and NS2, which support viral particle assembly and in case of NS2 functions as cysteine-like protease that cleaves the junction between NS2 and NS3 [1, 56, 76]. This includes also the N^{pro} auto-proteinase, which is a unique feature among the genus *Pestivirus* and cleaves at the C-terminus of the viral polyprotein [74, 77, 78]. Further followed by the multifunctional NS3 serine-protease/helicase and the NS4A cofactor [55, 79] as well as by NS4B and NS5A. Both proteins may play a role in the viral resistance [80, 81]. Last is the NS5B protein, which has a role in the formation of the replication complex [74, 81].

1.4 The multifunctional NS3-NS4A protein complex

NS3-NS4A is a multifunctional fusion protein with a molecular weight (MW) of approximately 70 kDa to 80 kDa, depending on the viral species. The complex works similarly for CSFV and HCV and is ingenious for viral survival due to its diverse functions. Its activity is required for polyprotein processing as well as for viral RNA replication and virion morphogenesis [2]. The heterodimeric protein complex is formed from the stabilising NS4A cofactor and the NS3 protein. The NS3 protein consist of the NS3 trypsin/chymotrypsin-like serine protease domain in the N-terminal region and the NS3-helicase/ATPase domain in the C-terminal region. The protein is responsible for the cleavage at four sites of the polyprotein [54, 79, 82-88]. This includes the proteolysis of the junctions between NS3/NS4A (self-cleavage), NS4A/NS4B, NS4B/NS5A, and NS5A/ NS5B (Fig. 4, Fig. 5). It was first shown by Grakoui 1993 [84], that all four cleavage sites of HCV have several common features. In both CSFV and HCV, this leads to a specific, preferred amino acid sequence of the cleavage sites and NS3 substrates, where CSFV and HCV have distinct specificity. For example, at the P1 position is a Leu for CSFV and Cys or Thr residues for HCV and Ser or Ala at the P1'. This leads to the following favoured substrate structures for CSFV (X-X-(Leu)↓(Ser/Ala)-X-X) and HCV (X-X-(Cys/Thr)↓(Ser/Ala)-X-X), where (-X-) indicates variable amino acids and (↓) the position of cleavage [55, 84, 89]. Moreover, a potential cleavage product can have additional cleavage sites, which can be identified and cleaved by the NS3-protease. The presence of these minor-cleavage sites was shown for NS4B and NS5A. Here, the first site is located close to the N-terminus of the NS4B protein and the second one, is in the middle of the NS5A protein [90, 91]. In addition, Lamp *et al.* 2013 [92] discovered two further minor internal cleavage sites within the CSFV NS3 protein at position Leu192/Met193 (NS3-protease)↓(NS3-helicase) and Leu159/Lys160. Amino acid substitution of Leu192 compromised viral growth and deletion of Leu159 or Leu192 inhibited viral replication. According to Lamp *et al.* 2013, this is caused by a loss of protease function. However, the exact biological role of the cleavage between protease and helicase remains unclear. The catalytic triad (violet-coloured balls, Fig. 5) of the NS3-protease domain (blue-coloured, Fig. 5), which is responsible for the cleavage of certain products, includes three different amino acids (His, Asp and Ser) [87, 93]. Moreover, in the NS3 proteolytic domain, a zinc ion is located, which is coordinated

by three cysteines and one histidine residue [94]. The NS3 zinc ion also seemed to be important for the functionality of the NS2 auto-protease, where the activity was shown to be zinc-dependent [94, 95]. For HCV it is assumed that zinc plays a catalytic role for the NS2/3 auto-protease complex and has a structural function for the NS3 protease [96].

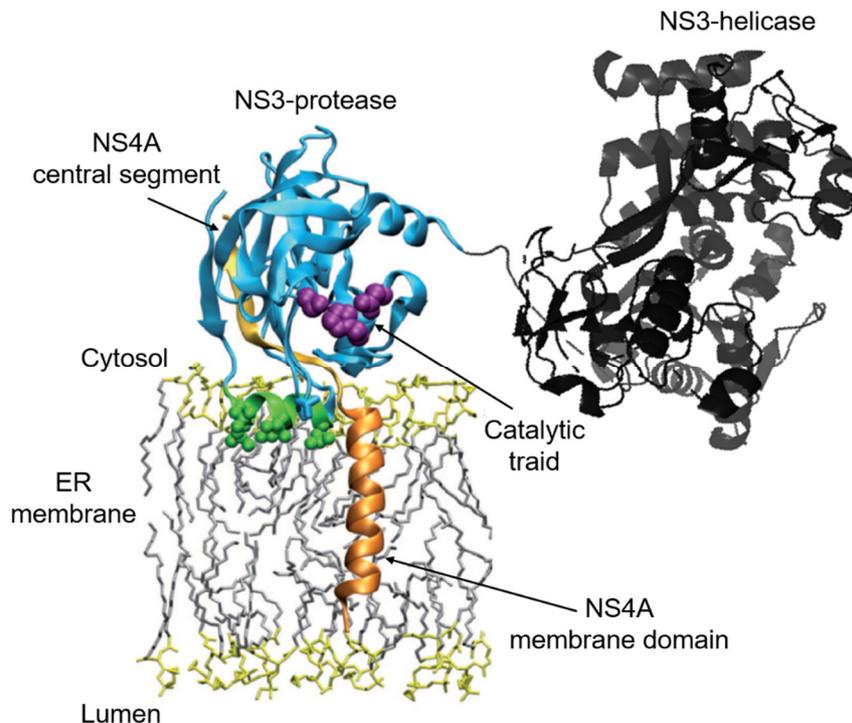


Fig. 5 Model of the ER membrane bilayer in association with the NS3-NS4A protein complex. NS3-protease = blue; side chain atoms of the catalytic triad (CSFV = His69, Asp97, and Ser163, HCV = His57, Asp81, and Ser139) = purple spheres; NS3- helix $\alpha(0)$ = green; the five hydrophobic residues are represented by sticks and balls; NS3-helicase domain = black. NS4A N-terminal transmembrane amino acids (aa1–20) orange; central segment of NS4A (amino acids 21–32) = yellow. Structure is based on Brass *et al.* 2008 [97], PMC open access article (OA) with respect to the conditions of the creative commons attribution license (<http://creativecommons.org/licenses/by/4.0/>); Copyright 2008 by the National Academy of Sciences of the USA), PDB: 1CU1, PDB: 5ZM4 (helicase).

Furthermore, the small cofactor NS4A (MW approx. 8 kDa) is required for several NS3 functions. Its hydrophobic N-terminal region is located within the ER membrane and its hydrophilic C-terminus protrudes into the cytoplasm. In more detail, the N-terminal amino acids (aa) 1 to 20 form the α -helical membrane anchor of the NS3-NS4A protein complex. The helix (orange-coloured, Fig. 5) is embedded into the ER membrane, which influences the orientation of the protein complex on the membrane [97]. The amino acids 21 to 32 form a β -barrel (yellow-coloured, Fig. 5), which lies within the NS3-protease, where it serves as a structure-stabilizing cofactor. This borders on the NS4A kink region, which is

the link to the acidic domain of NS4A [98, 99]. Thus, the N-terminal region of NS4A stabilizes the NS3 protein after polyprotein cleavage. [85, 100, 101].

The part of the NS4A cofactor used in the present thesis comprises the N-terminal amino acids aa21-57 for CSFV or aa21-54 for HCV. For both protein complexes, the constructs are missing the membrane domain to increase the solubility of the complex. Furthermore, the construct used here is based on the x-ray crystal structure of Tortorici *et al.* 2015 [102] and Dubrau *et al.* 2017 [103], whereby the first 8aa (originally N-terminal membrane anchor) of NS4A were fused to the NS3-helicase domain in order to stabilize the structure for crystallization.

Both, the NS4A and the NS3 protein form a non-covalent complex, which influences the protease and the helicase (black-coloured, Fig. 5) function. On the one hand, the interaction with the central region of NS4A is required to enhance the activity of the NS3 serine protease [78, 79, 85, 104, 105]. On the other hand, NS4A acts as a membrane anchor that holds the NS3 protein in position [97]. For HCV and the West Nile virus (WNV) a member of the genus *Flavivirus*, NS4A was found to regulate the ATPase activity of the NS3-helicase [100, 106]. Shiryayev *et al.* 2009 [106] were able to show that when the NS3-helicase, shown in Fig. 6, is linked to the N-terminus of NS4A, the helicase domain conserves energy when

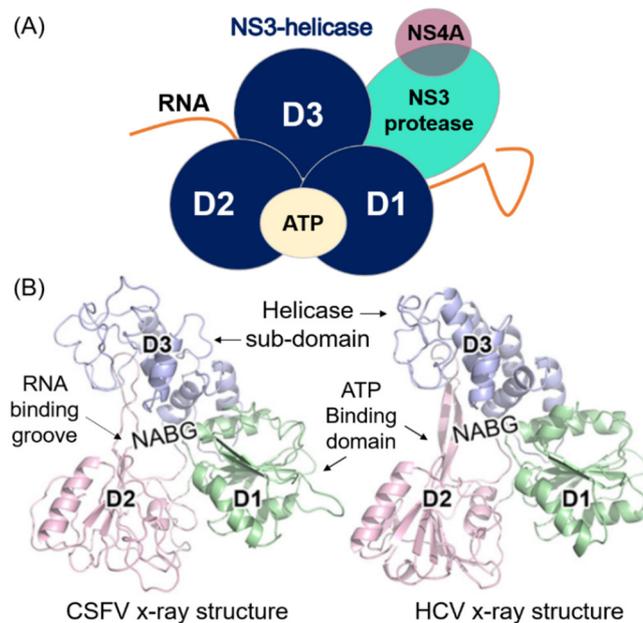


Fig. 6 Illustration of the NS3-helicase of CSFV and HCV.

(A) Schematic organization of the NS3-4A protein complex, adjusted based on Beran *et al.* 2009, [100]. (B) x-ray topology of CSFV and HCV NS3-helicase domains (D1 + D2 = ATPase, D3 = helicase) with the nucleic acid binding groove (NABG), modified based on Tortorici *et al.* 2015 [102] and PMC open access article (OA) terms and conditions (<https://creativecommons.org/licenses/by/3.0/>; Copyright 2015 by the American Society for Microbiology).

the RNA is unwound, which can then be converted into movement. The NS3-helicase is two-third of the NS3 protein (Fig. 6) and it could be shown that its activity is necessary for viral replication [107, 108]. The helicase domain forms three subdomains [109]. The N-terminal subdomain 1 (D1) and the middle subdomain 2 (D2), both with ATPase activity. The C-terminal subdomain 3 (D3) is important for unwinding nucleic acid duplexes and differs in the nucleic acid binding groove (NABG) among *Flaviviridae* [102]. The unwinding of duplex RNA, which is formed when the ssRNA genome is copied, is based on the motor protein function of the helicase domain, where the motor drive is based on conversion of free energy from ATP hydrolysis [102, 107, 110]. However, the exact mechanisms of the helicase, including the mobility of the helicase is still not fully understood. It is believed that specific conformational changes induce protein movement [102, 110].

1.5 Studying protein structure based on nano-ESI-MS and Top-Down analyses

Traditional mass spectrometry (MS) is a method for the analysis, identification and characterization of molecules. This is done on the basis of the mass of their atomic components. The analytes are ionised and converted into the gas phase. In the mass spectrometer, they are separated and detected based on their mass-to-charge ratio (m/z) and characterized based on their isotope pattern [111, 112]. MS can be used in a variety of ways, which is why different government, medical and natural science areas can benefit from this technology. MS-based procedures are widely used in different areas like in forensics in court investigations in the context of drug screenings or warfare agent analyses, in environmental analysis, in food quality control and in research to name a few.

MS in biochemical research concern itself for example with the elucidation of the structure of proteins and protein complexes. Here, native mass spectrometry is used as state-of-the-art technology, where native describes the state of the folded, assembled and functional protein [113, 114]. Proteins take on their intact quaternary structure based on their amino acid sequence composition, their environment and their binding partner [115, 116]. Homogeneous or heterogeneous complexes are often formed by the non-covalent attachment of other proteins. In solution, this only happens under conditions that simulate a corresponding physiological/near native environment.

When focusing on protein complex or assembly dynamics, for some structural biology technologies, such as crystallography or NMR spectroscopy it is near impossible to simulate sufficient physiological conditions. In the area of native electrospray mass spectrometry (native ESI-MS) however, this can be maintained experimentally even in the gas phase based on soft instrument condition and volatile solutions such as ammonium acetate [114, 117-119].

Therefore, native MS is a soft method that offers the possibility of analysing the stoichiometry and topology, the dynamics and the interaction of proteins with one another with low sample consumption. Complex protein mixtures can be resolved due to different protein masses. Further structural information can be obtained by related MS methods such as Top-Down protein sequencing, ion mobility spectrometry (IMMS), H/D exchange or cross-linking [120-123].

1.5.1 Quadrupole time-of-flight (Q-TOF) instrument

The Q-TOF 2 mass spectrometer is composed of following components (Fig. 7): an element for sample supply (coated capillary), a nano-electro spray (ESI) ion source (production of ions from the sample analyte), a hexapole ion guide, a quadrupole (Q-pole) for ion selection, a collision cell for MS2 collision induced dissociation (CID) of selected ions, a time-of-flight mass analyser (TOF), including pusher + reflectron and a microchannel plate detector (MCP). All components are in a vacuum apart from the sample injection unit, which is under atmospheric pressure. Based on voltage differences the sample analytes are first transferred into the gas phase. Neutral molecules are ionised at the ion source and are further guided through the instrument. The ion beam is focused via the Q-pole, the hexapole of the CID cell and different lenses or ion optics within the instrument. Whereby the beam is accelerated in the direction of the pusher, which is within the orthogonal TOF mass analyser. The pusher then pulses a portion of the ion beam in the direction of the reflectron, where the ions are reflected and accelerated to the MCP detector. Based on this, the generated ions are separated according to different m/z values. The separated ions, which reach the detector generate an electrical signal. This is recorded as a chromatogram (scans/time). From adding up the scans, a mass spectrum is finally mapped, which represents the signal intensity as a function of the m/z value [124, 125].

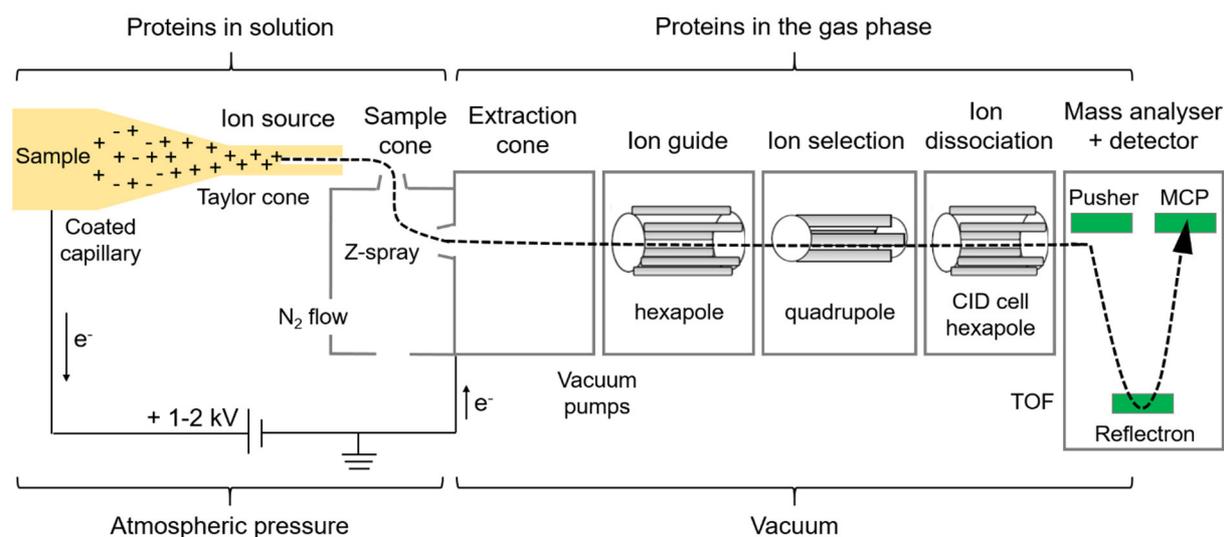


Fig. 7 Simplified schematic depiction of the components of the Q-TOF2 nano-ESI-MS instrument and guidance of the analyte ions in positive ion mode.

1.5.2 Mechanisms of electrospray ionisation (ESI)

Electrospray ionisation is similar to matrix-assisted laser desorption/ionisation (MALDI) a soft method, which allows the investigation of non-covalent proteins under near physiological conditions [119]. The ESI process takes place under atmospheric pressure. This powerful technique was developed by Nobel Prize winner John Fenn in 1989 [126]. Ionisation is suitable for large (in the MDa range), non-volatile molecules and protein complexes, which can easily absorb charges in the gas phase [127].

An electrically conductive capillary (usually with a thin gold coating) contains the sample solution (Fig. 8). This analyte solution is exposed to a high potential of several kV.

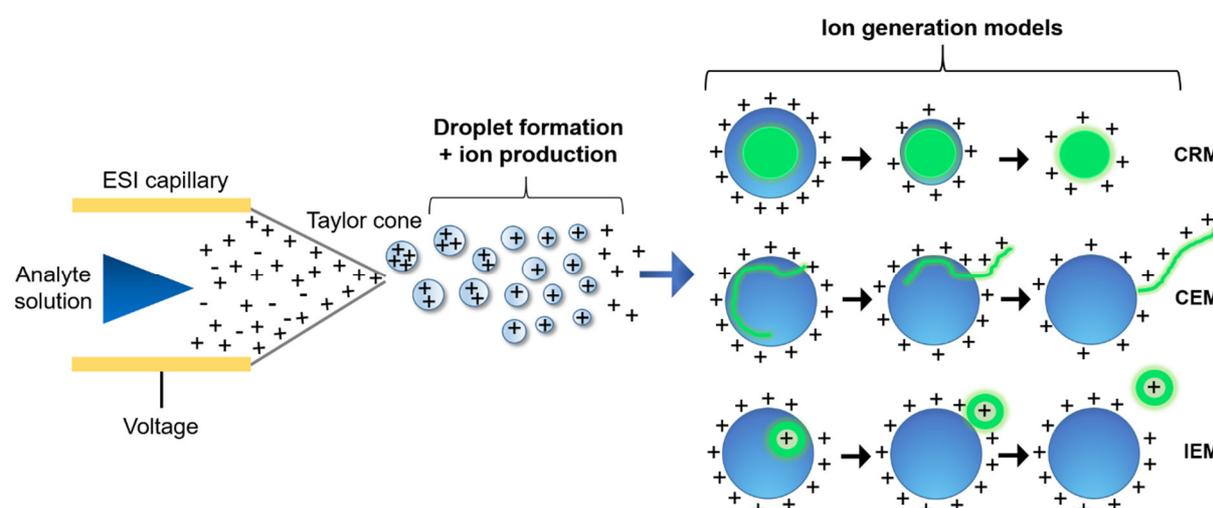


Fig. 8 Electrospray ionisation shows different mechanisms of ion generation.

The three well-known models, the charged residue model (CRM) for large analytes such as globular proteins or protein complexes, the chain ejection model (CEM) for non-polar polymer chains such as denatured proteins and the ion evaporation model (IEM) for small analytes are shown. Figure is based on Konermann *et al.* 2013, [128].

The voltage applied to the capillary causes the sample solution to be drawn towards the tip of the capillary in the form of a cone. This is where the so-called Taylor cone is created. Charged ions come out as droplets surrounded by the analyte solution [128]. Due to Coulomb's repulsion, the droplets drift apart, creating the electrospray aerosol.

The ion current depends on the inner diameter of the capillary. This is very low with nano-ESI capillary needles used in the present work (~ 5 to $10 \mu\text{m}$), which is why the flow rate is extremely low in the range of 10-50 nL/min and less sample material is consumed [129-

131]. With conventional ESI methods, on the other hand, the consumption of analyte solution is much higher (4 to 200 $\mu\text{L}/\text{min}$) based on the higher inner capillary diameter (50-200 μm) [132, 133].

The results presented in this work are based on ESI measurements in positive ion mode. This is based on a positively grounded potential, where the capillary acts as the anode and the mass spectrometer as the cathode. Molecules in the ESI-spray carry positive charges, which are transferred by the solution, for example in form of NH_4^+ and H^+ . Proteins get their positive net-charge from the protonation of basic amino acids and the N-terminus. Therefore, proteins in the unfolded state have significantly higher numbers of charges than in the folded state [128] because more amino acids are exposed to the solvent.

Following the ionisation process the analyte is leaving the Taylor cone, the volatile, aqueous ammonium acetate solution evaporates, whereby the charge density on the droplet surface increases steadily until it bursts and even smaller droplets are formed [128, 134]. This process is controlled by surface tension and Coulomb repulsion and can be explained by the so-called Rayleigh limit charging, which deals with the Rayleigh stability limit [135] for coulomb repulsion of the charged droplet and is reached when the aqueous solution droplets have the same radius (R) as the analyte proteins [134].

The Rayleigh limit [135] can be explained using the following formula (Eq. 1), where (Z_R) is the number of the elementary charge (e), (R) is the radius of the charged analyte droplet, (ϵ_0) the electric field constant and (γ) the surface tension:

$$Z_R = \frac{8 \times \pi}{e} \sqrt{\epsilon_0 \times \gamma \times R^3} \quad \text{Eq. 1}$$

The nano-electrospray generates very fine, charged ion droplets, which can then be analyzed in the mass spectrometer. However, the formation of ions and charge states by ESI strongly depends on the size of the analyte to be ionised as well as on the structure and surface area of the protein [117, 136]. There are different models that explain the processes of gas phase ion formation from charged droplets (Fig. 8).

For globular molecules, such as natively folded proteins or large complexes, it is assumed that the ionisation or the transfer into the gas phase takes place according to the charged residue model (CRM) [127, 137]. Here, tiny analyte drops are generated ($R \approx 1 \text{ nm}$) as a

result of a series of solvent evaporation and Coulomb fission and produce an extremely small charged droplet [134, 138].

The chain ejection model (CEM) is assumed for non-polar polymer chains (e.g. unfolded proteins) [128]. This mechanism is based on the fact, that hydrophobic interactions drive the molecule to the droplet surface. As soon as a chain end penetrates the surface, more charges are transferred to the analyte until it is completely expelled.

For analytes with a low molecular weight, the ion evaporation model (IEM) is assumed [134, 139], where at excess charge the analyte emerges from the droplet. This forms a bridge of solvent, which tears shortly afterwards and releases the analyte.

For all protein ions generated by ESI, the mass to charge distribution can be determined and described as $[M + zH]^{z+}$, where multiple charging ($z \gg 1$) is the usual case [128]. According to the statistical frequency the distribution of the charge states is shown in the mass spectrum.

1.5.3 TOF detection of analyte ions and data output

Mass analysers can separate ions spatially or temporally, depending on the type of analyser. The accuracy and resolution (R) of detectable MS signals strongly depends on the resolving power of the mass analyser. The resolving power is defined as the closest measurable separation between two peaks, which have equal width and height [140]. (R) can be represented in equation (Eq. 2):

$$R = \frac{m/z}{\Delta m/z} \quad \text{Eq. 2}$$

Here, the orthogonal-acceleration TOF mass analyser separates the generated ions according to their mass to charge ratio (m/z) (Fig. 9). In the TOF, the ions are spatially separated in a vacuum along a field-free drift path. The pusher deflects the ion flow by 90°. This means that all incoming ions receive the same acceleration of the TOF. The ions are separated on the basis of the pusher impulse that depends on the m/z value [125, 141].

The pusher then pulses a section of the ion beam orthogonally in the direction of the reflectron [142], where ions of the same type get focused by a constant electrostatic field [143, 144], which correct its direction of flight and increases the resolution. Then, the ions are

reflected toward the MCP detector. Here, all detection events result from single ions [145]. More precisely, this is the secondary measurement of the electron emission efficiency, which is based on the ions hitting the surface of the MCP detector from which electrons are being emitted. The accelerated electrons in turn hit the MCP surface, from which further electrons then emerge. Based on this, a cascade of electrons is formed that creates a current, which is detected as signal and recorded to finally visualize a mass spectrum [125, 145].

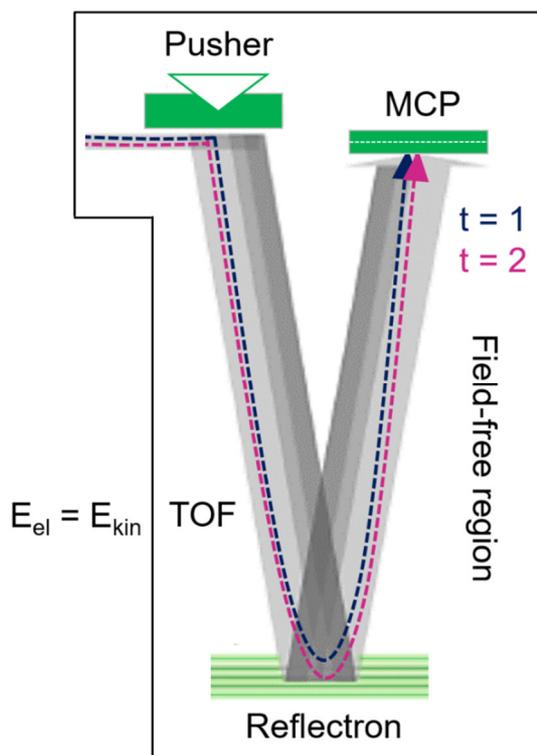


Fig. 9 Scheme of an orthogonal TOF analyser with pusher unit, reflectron and MCP detector.

The ions are bundled and accelerated by the pusher at 90° in the direction of the reflectron and reflected on to the MCP.

The distinct separation of ions in the TOF is based on the fact that lighter ions are reflected more strongly ($t = 1$) than heavier ones ($t = 2$) see Fig. 9. This is also explained by the number (z) of the elementary charges (e), where (q) is the charge of an ion with the mass (m). Thus, the electrical energy (E_{el}) of an ion, which is under the influence of the electrical potential (V) is represented by equation (Eq. 3):

$$E_{el} = qV = ezV \quad \text{Eq. 3}$$

Further, in the TOF the ions are accelerated in the electric field. This represents a conversion of the potential/electric energy (E_{el}) into kinetic energy (E_{kin}), which is described as followed:

$$E_{\text{el}} = E_{\text{kin}} \rightarrow E_{\text{el}} = zeV = \frac{1}{2} mv^2 = E_{\text{kin}} \quad \text{Eq. 4}$$

The m/z ratio of an ion can be calculated by recording the time (t). This depends on the voltage (U) that the ion needs for acceleration in order to travel through the TOF distance (s). This can be derived from the velocity (v) and distance-time law ($v = s/t$) as follows.

The velocity (v) of an ion can be applied to the basis of distance-time law if it is assumed that the ion only experienced an acceleration in the TOF. In addition, the time (t) that is needed by the ion to travel the distance (s) through the TOF is proportional to the square root of its m/z ratio.

This results in the following equations (Eq. 5) for the flight time (t) of the ion at constant speed (v):

$$v = \sqrt{\frac{2ezU}{m}} \rightarrow t = \frac{s}{v} = \frac{s}{\sqrt{\frac{2ezU}{m}}} \rightarrow t = \frac{s}{\sqrt{2eU}} \times \sqrt{m/z} \quad \text{Eq. 5}$$

The MCP signal is translated by the associated software into a mass spectrum, which contains information about the charge of the detected ions. The spectrum displays the intensity of signal peaks in dependence on the associated m/z value.

The intensity is further based on the frequency of ion detection, where the position of each peak indicates the m/z value of a detected ion. The peak with the highest intensity shows the most frequently occurring ion species. This base peak serves as the norm, so that the readout of the spectra is the relative peak intensity. In native MS, mostly folded globular proteins and large complexes are in the focus of research. Therefore, the spectra are much more complex (Fig. 10). In positive ESI-MS1 ion mode, a single protein species has different charge states, based on the varying amount of protons H^+ . Therefore, molecular ions are given as: $[M+zH]^{z+}$. The appearance of different charge states results in a characteristic peak distribution for natively folded proteins. This peak pattern is similar to the Gaussian distribution (Fig. 10 (B)), where the highest peak shows the most frequently detected ion ($z = \text{charge}$). The corresponding neighbouring peaks have a charge difference of $\Delta z = 1$.

The mass (M) of the protein can be calculated from the charge on the basis of the m/z values.

It follows: the higher m/z value of the corresponding peak = $(m/z)_1$, the lower = $(m/z)_2$, where $z_1 = z_2 + 1$. In contrast, proteins in the unfolded state have significantly higher numbers of charges than in the folded state (Fig. 10 (A)).

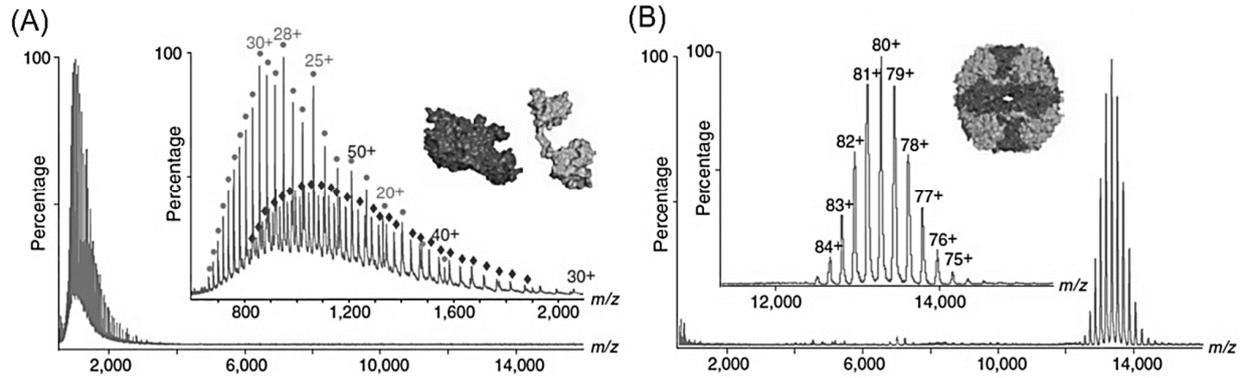


Fig. 10 ESI-MS based mass spectra of the denatured and natively measured protein complex *H. pylori* urease, adapted from Pinkse *et al.* 2003 [146] and Heck *et al.* 2008 [147].

(A) Shows the denatured protein which is split in its two subunits based on the aqueous 50% (vol/vol) acetonitrile containing 0.1 % (vol/vol) formic acid solution. The spectra revealed individual charge distributions from multiply charged monomers (α -unit = 26.6 kDa, light-grey) and (β -unit = 61.7 kDa, dark-grey). (B) Shows the corresponding native mass spectrum of the intact protein complex in an ammonium acetate solution. The spectrum shows multiple ions which originate from multiple charged species of the α 12- β 12 with a measured mass of $1,063.4 \pm 1.0$ kDa. Figure is adapted with permission from John Wiley and Sons (license terms and conditions, Copyright Clearance Center, Pinkse *et al.* 2003 [146], Copyright 2003).

Based on Covey *et al.* 1988 [148] and taking into account the proton mass H^+ (mH), the $(m/z)_1$ and $(m/z)_2$ ratio is given by equation (Eq. 6(a) and (b)). The charge was originally defined as (n), but here replaced by (z). The charge (z_1) and the mass (M) result from the equations (Eq. 6 (a) and (b)) and are shown in Eq. 7(a)-(b):

$$(m/z)_1 = \frac{M + z_1 mH}{z_1} \quad \text{Eq. 6(a)}$$

$$(m/z)_2 = \frac{M + z_2 mH}{z_2} = \frac{M + (z_1 + 1)mH}{z_1 + 1} \quad \text{Eq. 6(b)}$$

$$z_1 = \frac{(m/z)_2 - mH}{(m/z)_2 - (m/z)_1} \quad \text{Eq. 7(a)}$$

$$M = z_1((m/z)_1 - mH) \quad \text{Eq. 7(b)}$$

1.5.4 Q-pole precursor selection and MS2 collision induced dissociation

In MS1 mode, the spectrum of a native folded protein is shown in the typical Gaussian distribution of the product ions as shown before in Fig. 10. The displayed peaks of the ionised proteins can be analysed individually in the quadrupole (Q-pole), by separation of a single charge state (precursor ion) [149, 150]. In MS2 mode the ions get further fragmented in the collision cell (CID cell) in order to obtain further structural information about their composition.

The Q-pole mass analyser acts as a mass filter, which enables the transmission of ions belonging to a selected m/z range. Only these precursor ions can pass through the Q-pole, while other ions are not transmitted [124]. The Q-pole consists of four parallel rod electrodes, which are arranged in a square. The opposing rods are electrically coupled. Each rod is subject to the same electrical potential, which is composed of a direct and a high-frequency alternating current. The applied voltage is periodic and due to a sinusoidal radio frequency (RF) potential an inhomogeneous electrical field is generated resulting in the movement of ions along the rods [125, 151].

The fragmentation of the ion species, which is selected in the Q-pole, is based on collisional induced dissociation (CID) in the collision cell. The fragment ions obtained by CID allows the creation of a complex picture of the stoichiometry of a protein. Dissociation occurs through collision with inert gas atoms, usually argon, xenon or nitrogen. This increases the internal energy of the ion, by means of converting part of the kinetic energy of the ion into vibration energy, which causes the proteins to unfold. As a result, individual amino acid (aa) segments or entire subunits of a non-covalent complex unfold and split out of the complex and the selected precursor ion breaks down [124, 152, 153].

The resulting product ions are shown in the mass spectrum, where the partly unfolded smaller product ions carry more charges and thus appear at lower m/z values, whereas the more folded product ions carry less charges and are located at the higher m/z -range [154]. This does not apply to very small y and b fragment ions of the backbone amino acids, which usually only carry charges of +1 to +3. Due to their low mass and charge, they appear in the very low m/z -range (Fig. 11 (B)-(D)).

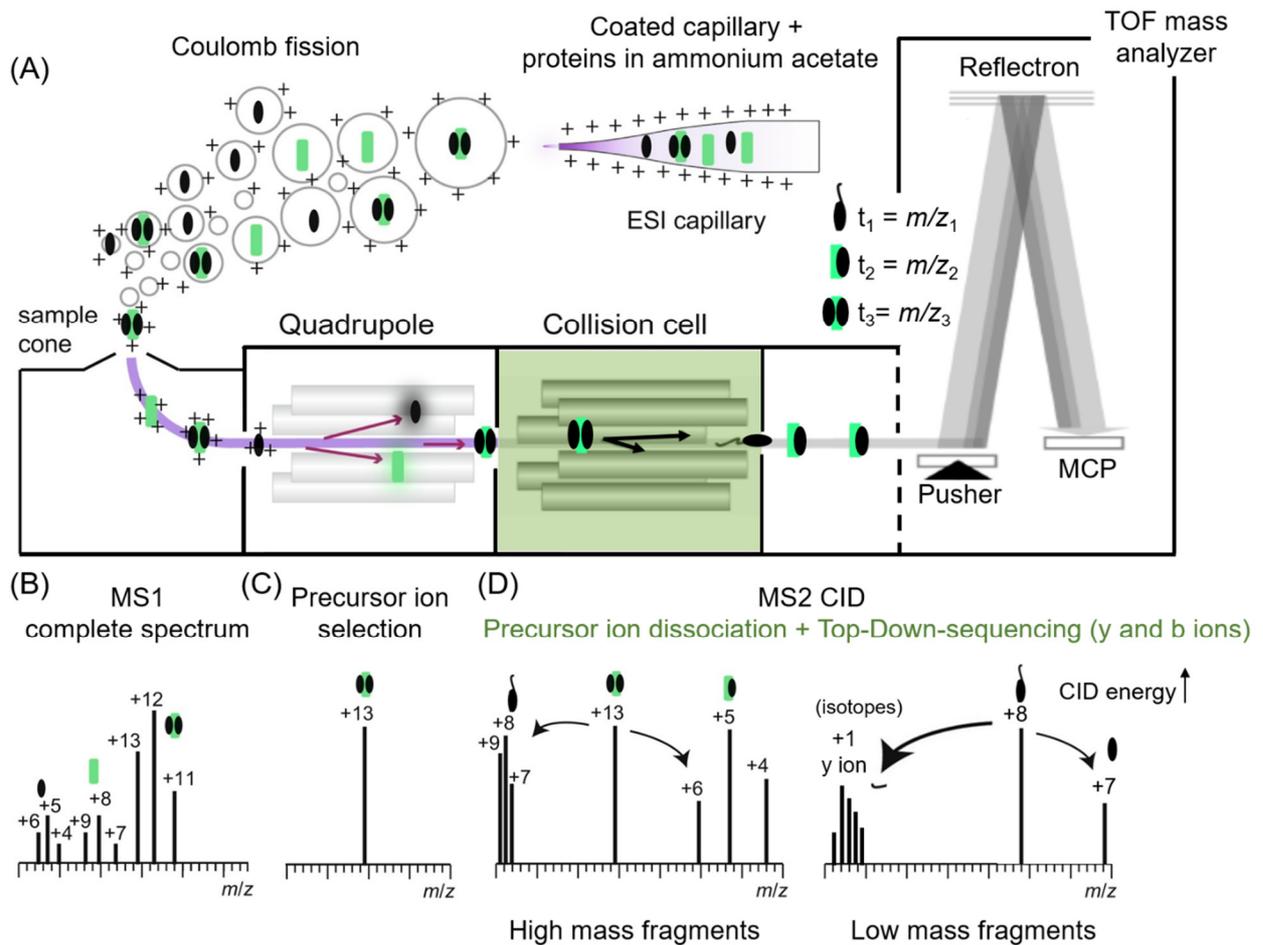


Fig. 11 Native ESI-MS and CID workflow and readout, adjusted based on Dülfer et al. 2019, [155].
 (A) A protein complex with two subunits (black and green) is sprayed from ammonium acetate solution, by the application of an electric potential and enters the mass spectrometer as molecular ions. The modules of the mass spectrometer allow for manipulation and analysis *in vacuo* as depicted in schematic spectra. (B) Full ESI-MS data reveal a Gaussian distribution of the charge states of each mass of the sample proteins. An additional charge distribution indicates a trimeric complex. (C) In order to confirm subunit stoichiometry of the complex, its molecular ions are filtered for in the Q-pole region. (D) Selected ions are subsequently collided with a neutral gas in the collision cell. The complex dissociates into two subunits of characteristic masses (high mass fragments), which confirms 2:1 stoichiometry. The dissociated subunit unfolds and takes a large number of charges with it. It is therefore detected in the lower m/z -range of the spectrum, whereas the remainder of the complex is detected at higher m/z . Further, applying higher voltages (CID energies), backbone amino acids of the dissociated smaller subunit get released and show the distribution of a singly charged y ion. Figure is adapted with permission from Elsevier (license terms and conditions, Copyright Clearance Center, Inc. ("CCC")), Dülfer et al. 2019, [155], Copyright 2019).

In addition, the sequence in which the subunits are released from the complex correlates with their position within the complex. Based on the aa-sequence of the analyte protein and the resulting fragment ion pattern, its composition can be elucidated and C- and N-terminal modifications can be detected (Fig. 12 (A)).

This is particularly relevant for the fragmentation of amino acids of the protein backbone and high CID energies in the context of Top-Down sequencing studies. Where y and b ions

formed by CID fragmentation (Fig. 12 (A), (B)) can provide information about the composition of the aa-sequence.

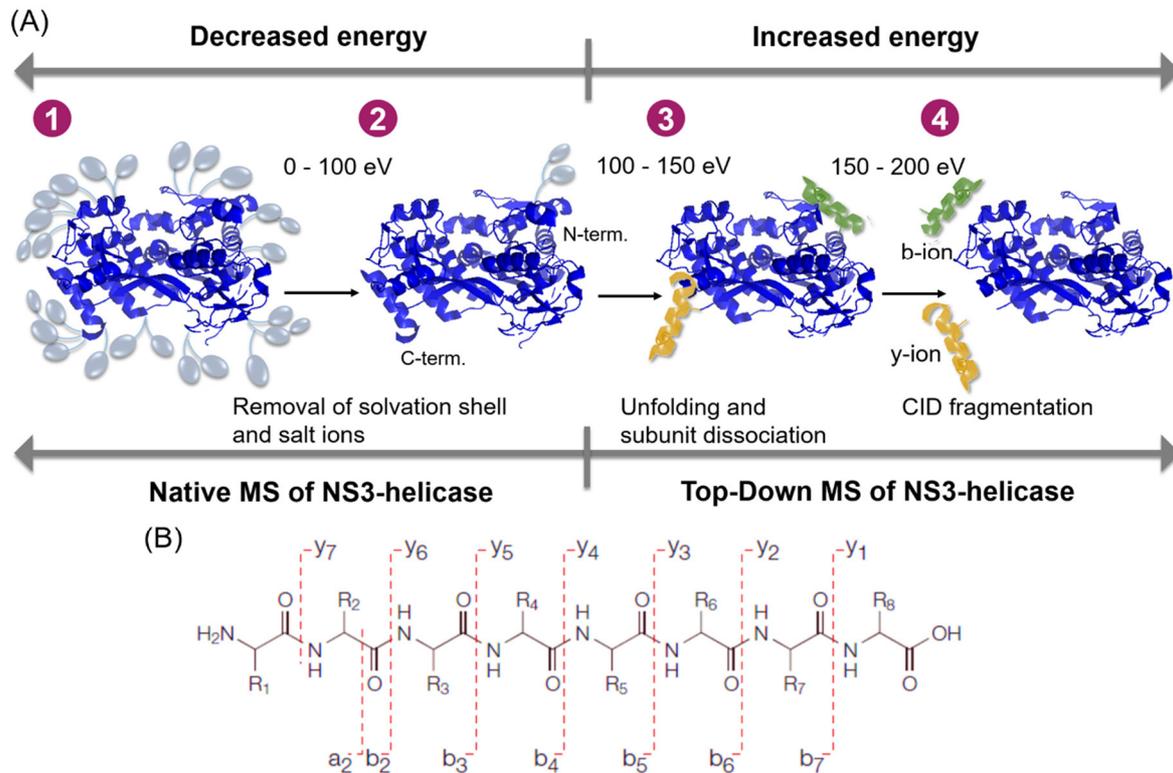


Fig. 12 Scheme of the behaviour of the NS3-helicase/ATPase under native ESI-MS conditions up to Top-Down fragmentation.

(A) The NS3-helicase/ATPase is exposed to increased CID energies. (1) + (2) The removal of solvation shell and salt ions requires up to 100 eV. (3) Partial unfolding of C- (D3 subunit of the helicase) and N-terminal (D1 subunit of the helicase) regions is observed under MS2 and MS1 conditions when increasing the energy up to 150 eV. (4) Dissociation of C-terminal y-ions and N-terminal b-ions from the complex in MS2 CID mode provides information of structural modifications based on the amino acid sequence. X-ray structure is based on PDB: 5ZM4 published by Dubrau *et al.* 2017, [103]. The schematic is designed based on Konijnenberg *et al.* 2015 [122]. (B) Fragmentation pattern of MS2 CID product ions, including mainly y and b ions, adapted from Steen and Mann 2004 [156].

1.6 Research goal of the project

For a better understanding of the replication process of *Flaviviridae*, the function of the non-structural protein (NS-protein) NS3-NS4A is important. Especially, for the development of anti-viral drugs as well as on a basic research level. This protein complex is a central target for drug development due to its variety of cellular functions and is therefore the focus of this thesis. Based on its serine-protease activity and due to several internal cleavage sites the NS3 protein is responsible for the splitting of various polyproteins and for the formation of the replication-complex. Previous data from other research groups employed a variety of different biochemical methods, such as crystallography, NMR-spectroscopy, FRET-assay, SAXS analyzes and other biological approaches [102, 103, 157] to analyze this protein. However, to the best of the author's knowledge there exist no mass spectrometry (MS) data, especially native MS for this protein in the scientific literature.

Based on this the present work has three main aims. First is the development of an electrospray mass spectrometry-based approach (ESI-MS) that should enable the monitoring of the auto-cleavage of the NS3-protease and NS3-helicase in more detail. This process is assumed to be based on the two minor cleavage sites (Leu159/Lys160 and Leu192/Met193) [92]. The second aim is based on the hypothesis, that the NS3-serine-protease has an auto-proteolytic activity. This process is assumed to result in the release of the protease and helicase from the complex [92]. Based on some preliminary MS tests, it is further suspected that the splitting of the complex can result in its complete degradation. As state-of-the-art technology, native ESI-MS allows detailed investigation for both aims. Krichel *et al.* 2021 [158] have already shown that the ESI-MS-based monitoring of the proteolytic cleavage of non-structural proteins of alpha and beta coronaviruses was possible. Analyses are mainly be based on the structural comparison of the wild-type protein of the classical swine fever virus (CSFV) with two mutants and the NS3-NS4 protein of the human hepatitis C virus (HCV). To show the viability of ESI-MS for activity-based studies the development of an activity assay for the WT proteins of CSFV and HCV as proof of concept will be the third aim of this thesis

2 Material and Methods

2.1 Viral DNA vectors and plasmids

The non-structural protein NS3-NS4 was adapted from the classical swine fever virus (CSFV, strain Alfort, GenBank accession number J04358.2) and human hepatitis C virus (hHCV, strain genotype 1b, GenBank accession number KJ564295.1/D50480.1) (Tab. 1). For knowledge transfer about the correct conditions for protein expression and purification the first samples were generated at a collaborators laboratory (University of Lübeck, department of virology and cell biology, laboratory of Prof. Dr. Norbert Tautz and laboratory of Prof. Dr. Lars Redecke). Afterwards, the plasmids, which were used for protein expression in Lübeck were kindly provided by Prof. Dr. Tautz for in-house protein expression and purification. For both proteins, DNA was amplified by PCR from plasmids (Fig. 13, Fig. 14).

Tab. 1 Expressed CSFV and HCV proteins.

Virus	Strain	Expressed pET-SCP proteins	amino acids (aa)	Monoisotopic MW (kDa)
CSFV	Alfort	WT	752	83.521
CSFV	Alfort	Y47A	752	83.429
CSFV	Alfort	L45A/Y47A	752	83.387
HCV	1b	WT	692	74.085

For both, the CSFV and the HCV NS3-NS4A proteins a pET vector system [159] was used for protein expression in *E. coli* cells. Both proteins are constructed as single chain proteases (SCP) with similar features at the N- and C-terminal ends (Fig. 13).

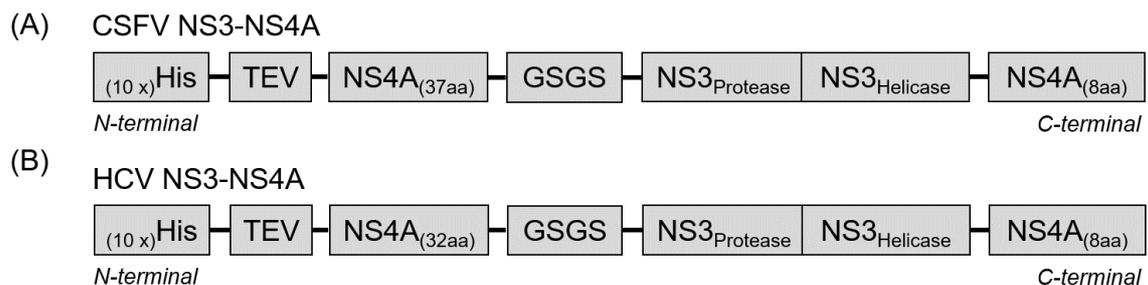


Fig. 13 Construct of the SCP NS3-NS4A proteins.

The amino acid sequence consists of (N-term. to C-term.): on mRNA level the start codon AUG, which codes for methionine and leads to the starting sequence (MAS + (10 x)-His-tag, a TEV protease cleavage site (ENLYFQG), the NS4A(37aa) cofactor aa-sequence, a connection via GSGS-linker, the NS3-protease, the NS3-helicase + eight N-terminal amino acids of NS4A(8aa) (aa are originally part of the membrane anchor).

The pET-11a is a 5676bp plasmid, which was used for expression of CSFV wild-type (WT) and mutated NS3-NS4A proteins (Fig. 14 (A)). The pET-21a is a 5443bp plasmid, which was used for expression of the HCV wild-type (WT) protein (Fig. 14 (B)).

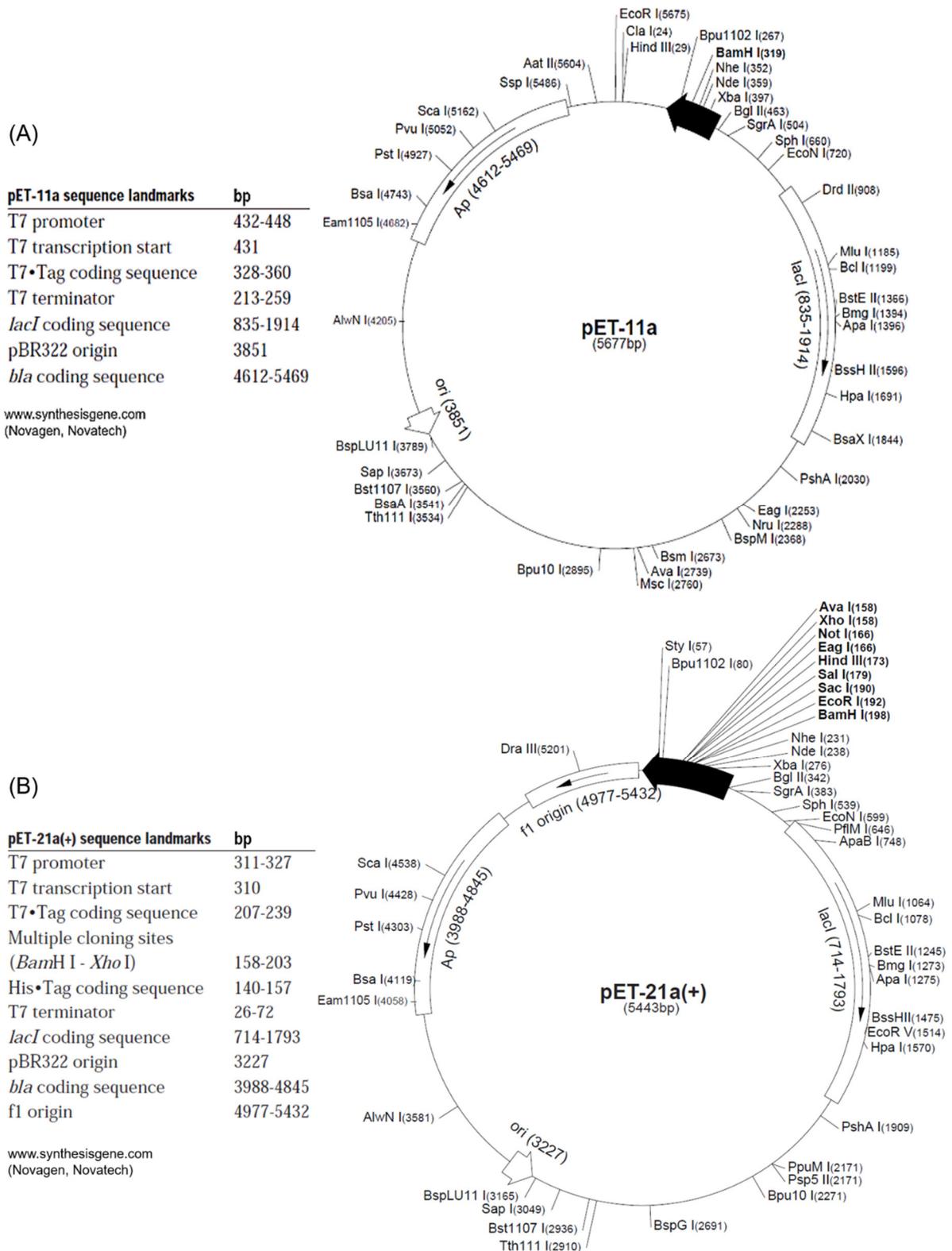


Fig. 14 CSFV and HCV NS3-NS4A expression plasmids, which were used for protein production.

2.2 Expression and purification of NS3-NS4A

CSFV and HCV NS3 protein expression and purification was adjusted based on the protocol of Lamp *et al.* 2013, [92] and Taremi *et al.* 1998, [160]. Precisely, NS3-4A proteins were produced in *E. coli* Rosetta-DE3 pLys competent cells induced with 1.0 mM IPTG at 0.6 OD and further incubation at + 20 °C for 20h overnight. Cells were harvested via centrifugation (Thermo Multifuge 3S-R, Heraeus) at 5500 g for 20 min at + 4 °C, directly followed by a centrifugal washing step with T1 washing buffer (30 mM Na₃PO₄, 400 mM NaCl, pH 7.4) at 3500 x g for 10 min + 4 °C.

The supernatant was discarded, and the cell pellets were stored at - 20 °C. For separation of soluble proteins, thawed cell pellets were resuspend at RT in 1:5 (v/v) T2 buffer (40 mM Na₃PO₄, 300 mM NaCl, pH 7.4 + 20 mM imidazole) (Sigma-Aldrich 99.5%). Further, cells were lysed with lysozyme (1:100) at + 4 °C while shaking for 2h. Additionally, five sonication cycles were performed to further disrupt the cell membranes and release all proteins (1 cycle: 5 sec sonication, 30 sec cooling) (Branson digital sonifier SFX 150) followed by a centrifugation step at + 4 °C and 20000 g for 45 min. Cell lysate was collected and mixed with T2 buffer and then transferred to a Ni²⁺-NTA-gravity-flow-column (Thermo Fisher Scientific, BioRad). The Ni²⁺-NTA beads were equilibrated with 20 column volumes (CV) T2 buffer (+ 20 mM imidazole), then followed by protein binding for 60 min at + 4 °C. Non-specific binding proteins were eluted by washing with 20 CV of T2 buffer (+ 20 mM imidazole) directly followed by 15 CV of T2 buffer (+ 50 mM Imidazole). Target proteins were eluted with T2 buffer (+ 300 mM Imidazole) in up to twelve fractions each with 0.5 mL volume. All fractions were fixed with 4 mM DTT immediately after elution. Subsequently, after affinity purification, protein samples were concentrated via Amicon centrifugal filter devices 10000 MWCO (Amicon, Merck Millipore, 10K) and prepared for purification by size exclusion chromatography (SEC, GE Healthcare, ÄKTApurifier). SEC was performed by running a Superdex- Sepharose 200 10/300 column (GE Healthcare) loaded with 500 µL of pooled and concentrated protein (5 to 10 absorption units per sample). Fractions of the main peak centre were collected for additional experiments or directly flash frozen in liquid nitrogen and stored at - 80 °C. Protein concentration was checked by UV absorption at 280 nm (DeNovix DS11-FX+) before and after SEC. Resulting protein masses are shown in the result part as well as in the supplement.

2.3 Determination of protein concentration by UV-VIS and absorption coefficient

To determine the protein concentration a UV spectrum (UV micro droplet spectroscope, DeNovix DS-11 FX+) at 280 nm (A_{280}) was recorded. The absorption A_{280} was represented as 1 AU~1mg/mL. The calculation of the exact protein concentration is further based on the estimated extinction coefficient (software: Expasy/Protpara, see 2.9) (Tab. 2).

Estimation of protein concentrations is based on the Lambert-Beer equation (Eq. 8) as seen below:

$$A = \varepsilon \times b \times C \quad \text{Eq. 8}$$

C = protein concentration in $\text{mg}\cdot\text{mL}^{-1}$

A = absorbance in 10 mm equivalent

b = pathlength of the sample expressed mm (DeNovix, 0.02 mm)

ε = extinction coefficient in $\text{mol}^{-1}\text{cm}^{-1}$

Samples for MS measurements were diluted with ammonium acetate solution to concentrations ranging from 5 to 30 μM . Determination of the theoretical monoisotopic protein masses is based on the amino acid sequence (aa-sequence) and the online software Expasy/MW/pI (see 2.9). The calculation of the theoretical pI is based on Expasy/Protpara (see 2.9).

Tab. 2 Extinction coefficient of CSFV and HCV NS3-NS4A protein at 280 nm.

Protein NS3-NS4A	Ext. coefficient A_{280} ($\text{mol}^{-1}\text{cm}^{-1}$) reduced cysteines	Ext. coefficient A_{280} ($\text{mol}^{-1}\text{cm}^{-1}$) pairs of cysteines	Theoretical pI
CSFV full-length	69680	70305	7.11
free NS3-helicase	49740	49865	5.83
free NS3-protease+NS4A (non-tagged)	18450	18950	8.99
free NS3-protease+NS4A (tagged)	19940	20440	8.87
HCV full-length	78730	79855	7.13

2.4 Quality control of expressed proteins by SDS-Page

Sodium dodecyl-sulfate polyacrylamide Page (SDS-Page) was used for quality control of NS3 protein expression and Ni²⁺-NTA/SEC purification. Either purchased long life gels were used (Criterion XT Bis-Tris 4 - 12% gels, Bio-Rad) or the gels were manufactured in-house as described in Tab. 3.

Proteins were separated as follows:

The purchased pages were inserted into a gel chamber (Bio-Rad) with XT-MOPS running buffer (4 - 12% gel separation range = 10 - 300 kDa) and 200 V was applied. The ROTI tricolor protein marker (Carl Roth) and the PageRuler prestained protein ladder (Fermentas) were used to indicate mass separation from 10 - 245 kDa.

The in-house gels were prepared in a multiple gel caster (Hoefer) with 5% stacking gel and 10% or 15% separating/resolving gel. Samples were prepared with four times concentrated sample buffer in _mqH₂O as described in Tab. 4. Different dilutions were used for the different fractions/samples (elution fractions = 1:3, flow through = 1:2, crude extract = 1:2, cell pellet = 5:1). For in-house prepared gels, proteins were separated using a mini gel chamber (Hoefer) at constant 25 mA per gel was used.

Depending on the sample composition the gel pockets of each gel were filled with 5-10 μ L sample mixture. Afterwards, to detect the protein bands the gels were first placed in a staining solution with 0.5% (w/v) Coomassie Brilliant Blue R250, 50% (V/V) ethanol and 7% (V/V) acetic acid solution (Coomassie Brilliant Blue R250, AppliChem). They were then incubated in a destaining solution (20% (V/V) methanol, 7% (V/V) acetic acid).

Tab. 3 Recipe for the used sodium dodecyl-sulfate polyacrylamide gel (SDS-Pages).

SDS-Page components	Stacking gel (5%)	Resolving gel (15%)	Resolving gel (10%)
	v/v (%)	v/v (%)	v/v (%)
_m qH ₂ O	60.5	23.0	35.5
Acrylamide (40%)	12.5	50.0	25.0
Tris 1.5 M, pH 8.8/6.8	25.0	25.0	37.5
SDS (10% stock)	1.0	1.0	1.0
APS (10% stock)	1.0	1.0	1.0
TEMED	0.1	0.1	0.1

Tab. 4 SDS-Page sample and running buffer, recipe adapted from Sambrook and Russel, 2006 [161].

Sample buffer components	4x Stock	10 mL solution (mL)	Running buffer components	10x stock 1 L (g)
Tris-HCL (1 M, pH 6.8)	250 mM	2.5	Tris base	30.0
SDS	8%	0.8	Glycine	144.0
Glycerol (100%)	40%	4	SDS	10.0
β -mercaptoethanol (14.3 M)	20%	2		
Bromophenol blue (0.2%)	0.008%	0.4		
mgH_2O	-	0.3	pH	8.2

2.5 Western blot

In addition, unstained SDS gels were used for Western blot analysis. Here the Precision Plus Protein™ Kaleidoscope standard (BioRad) was used as marker to determine protein size. The proteins were transferred to a PVDF membrane (Amersham™ Protran) using a Mini-PROTEAN® Tetra electrophoresis system with Mini Trans-Blot cell (BioRad). The membrane was activated with methanol for 1 min and rinsed with transfer buffer (192 mM glycine, 25 mM TRIS, 20% (v/v) methanol) before preparing the stack. Protein transfer to the membrane was performed on ice (+ 4 °C) at 20 V for 2.5 h. The free binding sites on the membrane were blocked for 1 h with bovine serum albumin (BSA) in tris-buffered saline (TBS-T buffer = 20 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 7.4). The membrane was then incubated for 2 h at room temperature with the primary antibody. This was either the mouse-anti-His (H-3) antibody (Santa Cruz) or mouse-anti-L-CSFV-NS3 antibody (collaborators, Prof. Dr. N. Tautz, University Lübeck). The membrane was washed three times with TBS-T and blocked again followed by the incubation with the second antibody, which was a goat IgG anti-Mouse IgG+IgM (H+L)-HRPO (MinX none Jackson Immuno Reseach) for 1 h at room temperature. Finally, the last washing step with TBS-T was performed as shown in the schematic of Fig. 15 and in Tab. 5.

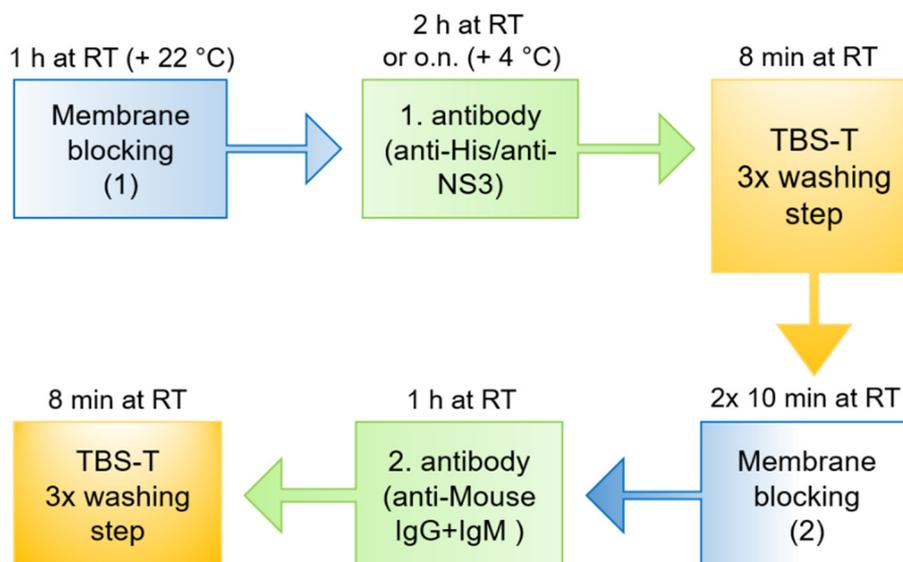


Fig. 15 Workflow for Western blot.

The proteins were detected using the Enhanced Chemiluminescence (ECL™, GE Healthcare) Western blotting system. The membrane was prepared with 500 µL of solution (1) and solution (2). It was further wrapped in transparent film and placed in a developer cartridge. An Amersham Hyperfilm ECL was applied to the membrane and left there for 30 seconds and finally developed in a Structurix M ECO (GE Measurement & Control). Further films were left on the membrane for 1 and 10 min and finally developed.

Tab. 5 Used protocol for Western blot analyses.

Solution	Components	mL
(1) Blocking	BSA in TBS-T (5% w/v)	30
1. antibody	mAnti-His antibody (H-3) or (CODE4)mAnti-L-CSFV-NS3 (1:1500 and 1:10) in PBS + 5% milk + 0.1% Tween-20	5
3x wash	TBS-T	30
(2) Blocking	BSA in TBS-T (0.5% w/v)	10
2. antibody	goat IgG anti-Mouse IgG+IgM (H+L)-HRPO (1:1500) in PBS + 0.1% Tween-20	5
3x wash	TBS-T	35

2.6 FRET peptide assays

A quick and simplified Förster resonance energy transfer (FRET) assay was implemented to check whether an activity of the impure HCV NS3-protease mixture could be detected. Therefore, the contaminated Ni^{2+} -NTA eluates of the NS3-NS4A protein were pooled. For the assay the specific HCV NS3-protease substrate RET-S1 peptide (AnaSpec) was used (MW = 1548.6 Da). RET-S1 is a NS4A/NS4B junction mimic that contains a C-terminal (EDANS) fluorophore and a N-terminal (DABCYL) quencher (Fig. 16) [162].

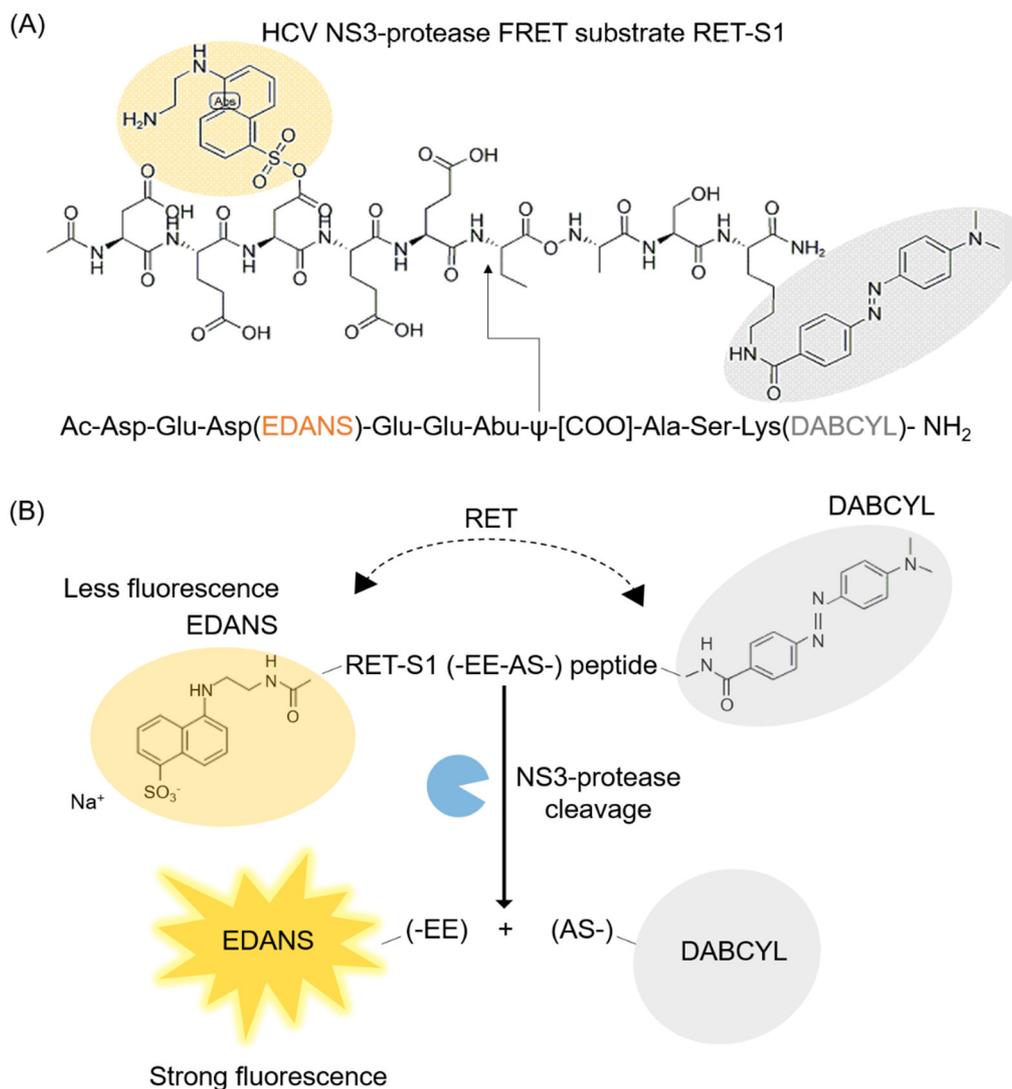


Fig. 16 Structure and function of the RET-S1 substrate.

(A) Sequence of the RET-S1 peptide. The ester bond between Abu (P1) and Ala (P1') is shown [162]. (B) RET-S1 peptide cleavage by the NS3-protease. Within the complete RET-S1 peptide, the EDANS fluorescence donor signal is quenched by DABCYL based on resonance energy transfer (RET). If the active NS3-protease is present, the peptide will be cleaved between Glu and Ala. Upon this process the EDANS fluorophore and its DABCYL quencher get separated from each other, whereby a strong fluorescence signal is detected.

It enables detection of NS3-protease activity with sub-nanomolar concentrations, where the increase in fluorescence can be monitored upon peptide cleavage. Here, the cleavage of RET-S1 is proportional to the enzyme concentration [162, 163].

In the WT NS4A/4B the cleavage site (aa-sequence: DEMEECASHL) is at the residues in P5 and P4 (Glu and Met). NS3 cleaves the substrate between residues Cys (P1) and Ala (P1'). In the used RET-S1 peptide the cysteine is substituted by aminobutyric acid (Abu) [162].

In the present study, the RET-S1 peptide was mainly used to implement the ESI-MS based activity assay, which is introduced in chapter 3.6. The RET-S1 powder must be stored protected from light. Therefore, it was dissolved in DMSO (2 mM) and stored at -20 °C. For the FRET analyses, RET-S1 was diluted into 10 µL fractions in reaction buffer (50 mM HEPES, 15% (v/v) glycerol 1 mM DTT at pH 7.5) to a final concentration of 1 mM.

The FRET assay was performed on a 96-well plate reader (Infinite200, Tecan). Polystyrene 96-well plates for fluorescence applications were used (Flat Bottom Black, Greiner). The excitation wavelength was set to 355 nm (Infinite200 bandwidth = 9 nm), the emission wavelength to 500 nm (Infinite200 bandwidth = 20 nm), number of flashes 15, 40 µs integration time, gain manual to 80. Substrate degradation is monitored by the changes in the RET-S1 fluorescence (fluorescence units, FU) over five time points. To determine the specific NS3-protease kinetic reaction rates 20 µL protease (final concentration 0.045 µM) were mixed with 80 µL RET-S1 substrate (final concentrations of 4.8 µM). The fluorescence was measured for 5 minutes (every 60 seconds).

2.7 Proteomic analysis

Sample preparation and analysis were kindly done by a member of Prof. Schlüter's group at the proteomics core facility at the UKE. The following methods are based on their protocols.

2.7.1 Sample preparation for proteomic analysis

100 μ L (40 μ g) of each virus sample was handed over to the group of Prof. Schlüter. Sample preparation was performed by a member of the group according to established standard operating procedures (SOP). Briefly, cysteines were reduced using 10 mM dithiothreitol (DTT, Sigma) and incubated at + 60 °C for 30 min. After cooling to RT, free cysteines were labeled using 20 mM 2-iodoacetamide (IAA, Sigma) and incubated at + 37 °C in the dark for 30 min. Trypsin was added to the solution in a ratio of 1:50 (Trypsin:Protein, sequencing grade) and incubated at + 37 °C overnight. Peptides were lyophilized using a vacuum centrifuge.

2.7.2 MS analysis

Samples were analyzed on a nano-ultra-pressure-liquid chromatography system (Ultimate 3000 RSLCnano, Thermo Fisher Scientific) coupled to a tandem mass spectrometer (Fusion, Thermo Fisher Scientific) with a nano-spray source.

Peptides were applied and trapped on a reversed phase trap column (2 cm x 180 μ m ID; Acclaim PepMap trap column packed with 3 μ m beads, Thermo Fisher Scientific) and separated on a reversed phase column (25 cm x 75 μ m ID, Acclaim PepMap, 3 μ m beads, Thermo Fisher Scientific). The column temperature was kept at + 45 °C. Peptides were separated using a 60 min gradient starting at 2% increasing to 30% buffer B (0.1% formic acid in acetonitrile (ACN)) a constant flow rate of 250 nL/min. Data were acquired in data dependent mode. Spray voltage was set to 1800 V and the transfer capillary temperature set to + 300 °C. All data were acquired in positive mode using dynamic exclusion for precursor ions of 20 sec.

Fullscan spectra were acquired in the Orbitrap using a resolution of 120000 with a scan range of 400 to 1300 m/z . AGC target was set to 2×10^5 with a maximum injection time of 100 ms. All Fullscan spectra were acquired in profile mode. The top speed method for precursor ion selection was used for fragmentation with a minimum intensity of 1×10^4 . Signals with unassigned, singly charged or with +6 or higher charges were excluded from fragmentation. Ions were isolated using a 1.6 m/z window and fragmented using higher energy collisional dissociation (HCD) with normalized collision energy of 30%. Fragment spectra were acquired in the Iontrap using the rapid scan rate setting with a fixed first mass of 110 m/z . AGC target was set to 5×10^4 with a maximum injection time of 60 ms. Ions were injected for all available parallelizable time.

2.7.3 Data analysis

MS data was analyzed using ProteinDiscoverer 2.4 (Thermo Fisher Scientific). As *E. coli* was used as host for the expression of the viral proteins a background FASTA database from the *E. coli* strain k12 (obtained in March 2020 from UniProt) which was supplemented with the aa-sequences of the target viral proteins.

Spectra were searched against the custom *E. coli* database with the following parameter settings. Carbamidomethylation on cysteines was set as fixed modification and oxidation of methionine, pyroglutamate formation at peptide N-terminal glutamines as well as acetylation on protein N-terminus, methionine-loss protein N-terminus and acetylation after methionine-loss protein N-terminus was set as variable modifications. Trypsin was set as specific enzyme and up to two missed cleavages were allowed. Peptides were filtered to 1% False-discovery rate (FDR). Precursor tolerance was set to 10 ppm and fragment tolerance to 0.5 Da.

2.8 Native mass spectrometry

2.8.1 Sample preparation

To analyse samples by native MS volatile buffers are required, thus salts from protein purification need to be replaced. Therefore, all proteins were buffer exchanged before measuring. For the NS3-NS4A protein of CSFV and HCV a 200 mM and 250 mM ammonium acetate solution at pH 7.2 to 7.4 is appropriate. Further, to prevent artificial protein oligomerization 0.5 to 2 mM of freshly prepared DTT in $m_q\text{H}_2\text{O}$ was added. Buffer exchange was performed at + 4 °C by six rounds of dilution and concentration in centrifugal filter units (Amicon, 10000 MWCO, Merck Millipore).

2.8.2 Nano-ESI capillaries

In-house produced nESI capillaries based on borosilicate glass (1.2 mm outer diameter, 0.68 mm inner diameter, by World Precision Instruments) were used for MS analyses. The capillaries were baked at + 200 °C for 30 min and afterwards pulled in a two-step program using a micropipette puller (P-1000, Sutter instruments) with a squared box filament (2.5×2.5 mm, Sutter instruments). Further, the capillaries were gold-coated using a sputter coater (Q150R, Quorum Technologies, settings: 40 mA, 200 s, tooling factor of 2.3 and end bleed vacuum of 8×10^2 mbar argon).

2.8.3 NS3-NS4A sample preparation and ESI measurement

CSFV WT and mutants as well as the HCV WT proteins were independently buffer exchanged to a final concentrations of approximately 15 μM to 5 μM (monomer concentration (1:4)), respectively. Freshly buffer exchanged protein samples were stored on ice at + 4 °C until MS analysis. A sample volume of 1 to 3 μL was withdrawn by means of a microliter syringe (5 μL , Hamilton) with flexible fused silica tubing (Optronis). Further the sample was loaded into the in-house produced nano-ESI capillaries. Subsequently, the filled capillary was mounted on the nano-ESI source. This sample handling process should only take up to 2 minutes. For a simple sample quality control, spectra were recorded within the first 50 to 100 scans (1 min). For the more precise analyses (e.g. changes of setting or energy) scan recording was extended to 1400 scans per run (20 - 30 min).

2.8.4 High Mass *Q-ToF 2* measurements

Measurements of viral proteins were performed using a *Q-ToF 2* (nESI-ToF) instrument (Waters, UK and MS Vision, the Netherlands) which is modified for high mass experiments [164]. The resolving power of the instrument is 10,000 full width at half maximum (FWHM) [195]. Native mass spectra were recorded in positive ion mode, using a nano-electrospray source (room temperature was + 22 °C to + 25 °C). Instrument voltages and pressures were optimized for the NS3-NS4A protein complexes based on Tahallah *et al.* 2001, [165]. Measurements were performed with applied voltages at the capillary with 1250 to 1450 V and at the cone with 120 to 150 V, optimized for minimal complex dissociation. Argon was used as collision gas for the purpose of protein dissociation. The pressure in the collision cell was adjusted to 1.2 to 1.4×10⁻² mbar argon flow. Gas pressure in the source region was set to 10 mbar. Accelerating voltages for near native conditions were set to 5 to 30 eV within the collision cell.

Pusher time settings as well as MS profile were adapted to the scanned mass range, respectively. Reflectron and ToF voltages were adjusted to 35 and 9.1 kV.

For Top-Down analyses instrument settings were adjusted for identification of low mass *y* and *b* ions. Therefore, the *Q-Pole* settings were adapted to tandem MS measurements (MS2), which were performed in ToF MS/MS mode. The preferred m/z value was set manually according to the precursor peak of interest. Mass pass filters (LMres, HMres) were adjusted to the m/z filter range individually until the desired peak signal was recorded. Depending on the experimental parameters for collision-induced dissociation and Top-Down analyses of the analysed protein energy was increased up to 200 eV for precursor dissociation.

For the calibration of raw MS data a spectrum of 25 mg/mL caesium iodide from the same day was recorded. Data calibration was performed with MassLynx 4.1 software (Waters).

2.8.5 MS data analysis

After MS measurement, the obtained spectra were explored by smoothing (2x5) in Masslynx 4.1 (Waters), followed by analysis of the relative signal intensities. A peak-list was cre-

ated which was then assigned to the target proteins. Further, the peak intensities were transformed into peak areas by the OriginPro 9.0 software. Data were normalized to obtain the total peak area, which was set to 100%. Additionally, as a comparison the same was performed with the signal intensities, whereby the total intensity was 100%. Moreover, the averaged signal (peak intensity or peak area) of the associated charge states was analysed. The values are given in percent in order to be able to compare the corresponding ratios of the peak signals of certain proteins/subunits. The standard deviation was calculated by combining three independent spectra. Spectra that were selected are based on the same conditions and scan ranges. Poorly resolved spectra were not included in the data analysis.

2.9 Software and data banks

Tab. 6 Used software and data processing tools

Function	Software/tool	Origin/domain	Access/date
Assigning peak series, MW determination	MassLynx V4.1	Waters, UK	Working laptop, evermore
Protein pI/Mw calculator (based aa-sequence)	Expasy/MW/pI	Swiss-Prot/TrEMBL, (https://web.expasy.org/compute_pi/)	July 2021
Peptide MW calculator (based aa-sequence)	CEM	CEM Corporation, US (https://cem.com/de/peptide-calculator)	July 2021
Estimation of extinction coefficient (based aa-sequence)	Expasy/Prot-param	Swiss-Prot/TrEMBL, (https://web.expasy.org/protparam/)	July 2021
CID fragment ion calculator	Proteomics toolkit, <u>Prosight Lite</u>	(http://db.systemsbiology.net:8080/proteomicsToolkit/index.html , http://prosiglitem.northwestern.edu/)	July 2021
Sequence alignment tool	Tcoffee/Expresso	(http://tcoffee.crg.cat/apps/tcoffee/do:expresso)	July 2021
Sequence analysis	Serial Cloner 2.6	(http://serialbasics.free.fr/Serial_Cloner.html)	Working laptop, evermore
Transcription and Translation Tool	Sequence converter tool	(http://bio-model.uah.es/en/lab/cybertory/analysis/trans.htm)	July 2021
Crystal structure visualization & editing	PyMOL V.1.7.4	PyMOL modification based on PDB: 5MZ4, [103])	Private PC, evermore
Processing & assigning of data (curve fitting, peak annotation etc.)	OriginPro 9.0	OriginPro 2016 SR2 (OriginLab)	Private PC, evermore
Processing of data (adjustment of line thickness, peak annotation, magnification of axes, coloration of lines)	Adobe Illustrator CC-2015 & 2020	Adobe Illustrator	Private PC, evermore
Genomic information on CSFV and HCV	GenBank	National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/)	July 2021

3 Results

ESI-MS reveals structural variations, auto-cleavage, degradation and enzymatic activity of the non-structural protein complex NS3-NS4A

Native mass spectrometry was used to analyse the non-structural fusion protein NS3-NS4A (NS-protein) of two different members of the *Flaviviridae* family, the *Pestivirus* and the *Hepacivirus*, respectively. The NS3 protein of both viral genera is a multifunctional heterodimer that contains an N-terminal serine protease and a C-terminal (NTPase)/helicase [79, 166, 167]. The NS3-protease is stabilized by the structurally integrated NS4A cofactor [79, 168]. The interaction of the NS3 protein with the NS4A cofactor results from a non-covalent fusion of both [169].

Native MS investigations were mainly carried out with the enzyme of the classical swine fever virus (CSFV). The analyses are based on the wild-type (WT) NS3-NS4A protein complex, which has its complete enzymatic activity. In contrast, the introduced x-ray crystal structures have a S163A mutation at the C-terminal end of the NS3-protease part. The mutation is located at the catalytic triad region to slow down the protease activity. This protein variant is modified specifically for crystallographic studies [102, 103].

MS studies of the WT protein were supplemented by analysis of CSFV proteins which possess mutations at the NS4A cofactor region, followed by investigations on the WT NS3-NS4A protein of the human hepatitis C virus (HCV). Since the NS3-protease is known to split the linkages of several further NS-polyproteins [55, 89, 92, 93], the main aim of this thesis was to address the question whether the auto-catalytic cleavage within the NS3-NS4A protein can be verified with the help of native MS. Additionally, in earlier studies it is controversially discussed whether the NS3-NS4 protein is only active as a monomer or also in other oligomeric forms. Dimerization seemed to play a role but the exact function and the mechanism have not yet been fully clarified [102, 170-172]. Therefore, native MS was employed to determine which oligomers are the dominant protein forms.

Results indicate that the protein was consistently found in a variable monomer-dimer ratio that apparently is independent of the concentration of the monomer. Whereas the formation of higher ordered oligomers is favored by a higher protein concentration (15 μ M).

This resulted in a further sub-aim to investigate protein oligomerization using MS based stability studies.

In the first part, the NS3-NS4A wild-type target protein was identified based on sodium dodecyl sulfate (SDS) gel-electrophoresis, by native MS, by a proteomics-based approach and Western blot analyses. Monomer and dimer sub-species were identified based on the protein sequence information, by their molecular weight in MS1 studies and by MS2 fragmentation (Top-Down analysis). Structural differences and similarities between the enzymes were analysed. Additionally, different mutated NS3-NS4A proteins of CSFV were compared with a focus on dynamics in structural changes using native MS. Furthermore, an MS-based activity assay was established with the wild-type enzymes of CSFV and HCV as proof of concept.

3.1 Purification of the structurally intact NS3-NS4A proteins

The following chapter shows the extraction and purification of the non-structural (NS) proteins in more detail. The NS-proteins were produced in *E. coli* (Rosetta 2(DE3) pLys). For the expression of the CSFV and HCV NS-proteins bacterial vectors were used (pET11a and pET21a respectively). The plasmids used for both, the CSFV and the HCV proteins, contain a target construct consisting of an N-terminal His-tag followed by an amino acid-sequence (aa-sequence) section of the NS4A cofactor, the NS3-protease and NS3-helicase subunit and a shorter section of the NS4A-cofactor (Fig. 17, Fig. 18). The part of the NS4A cofactor that interacts with the NS3-protease comprises the N-terminal amino acids aa21-57 for CSFV or aa21-54 for HCV. The present constructs are missing the membrane domain to increase the solubility of the proteins. The design of the construct is based on the sequence used by Tortorici *et al.* 2015 [102] and Dubrau *et al.* 2017 [103] for their study on the x-ray crystal structure, in which the first 8aa (originally N-terminal membrane anchor) of NS4A were fused to the NS3-helicase domain in order to stabilize the structure for crystallization.

Some N-terminal amino acids, especially those from the kink region, seem to stabilize the CSFV protein structure [103].

The final C-terminal amino acids of the cofactor serve to stabilize the structure of the protein for crystallization studies [78, 102, 103, 173]. Thus, the monomeric conformation of the fusion protein consists of three different sub-units [2, 54, 87, 103, 157, 174, 175]. Looking more closely into the subunits, the NS3-protease has a mass of approximately 21 kDa, the NS3-helicase has 55 kDa and the NS4A cofactor has 5 kDa.

The individual segments of the fusion protein are labelled using different colours (Fig. 17, Fig. 18). Additionally, based on the protein degradation various arrows in the aa-sequence of the CSFV enzyme show further sub-species of the NS3-4A protein. These sub-species were identified by native MS and Top-Down analyses and are missing certain parts of the aa-sequence. Experiments and data that identified these sub-species are shown in section 3.3.1 and 3.3.2.

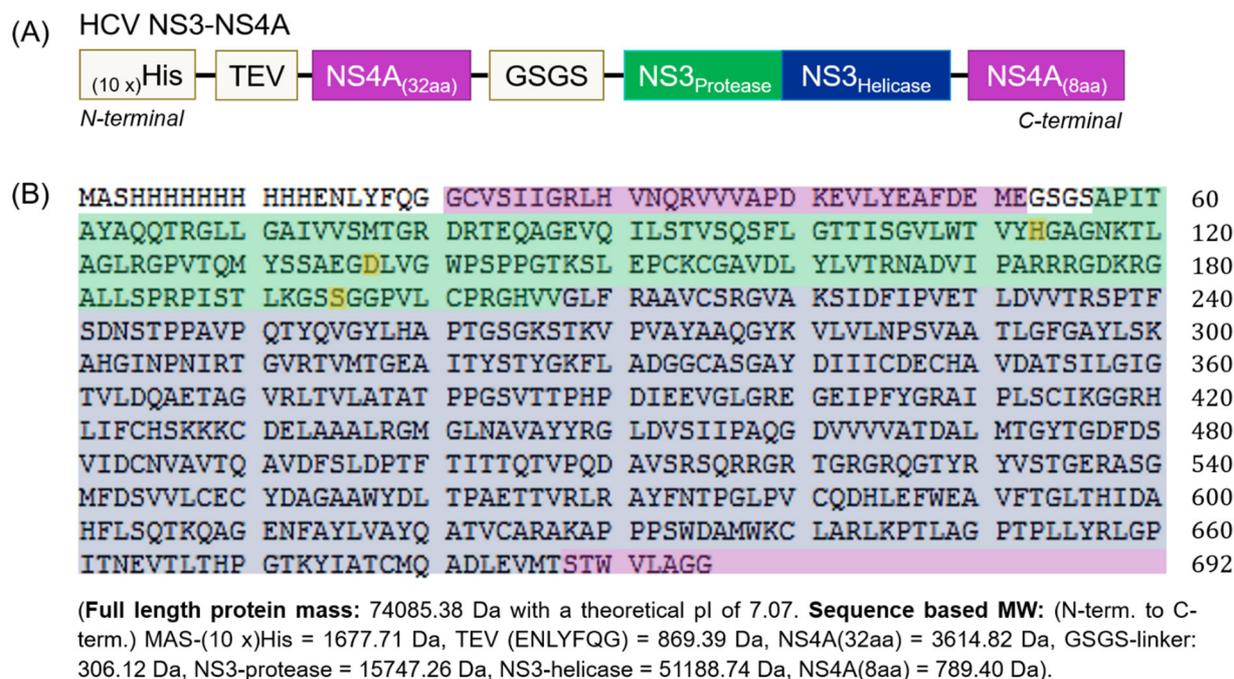


Fig. 18 Construct and amino acid sequence of the HCV NS3-NS4A target protein.

(A) The construct of the protein, (protease (green), helicase (purple), NS4A cofactor (pink), HIS-tag, TEV-interfaces and GS-linker (white)). (B) Complete amino acid (aa)-sequence of the wild-type protein. The different parts of the fusion protein subunits (NS3-protease, helicase) and the NS4A cofactor are coloured as shown in (A). The catalytic triad residues are labelled in orange (His57, Asp81, Ser139, [84]).

After expression of the NS-proteins, SDS gel-electrophoresis was used to determine whether the protein expression was successful (Fig. 19). For this purpose, fractions of the protein pellet (PE) and as a control the protein flow through (FT) and the crude extract (CE) were loaded onto the gel. For samples where the target proteins were present, the following purification steps were carried out: Ni²⁺affinity chromatography with gravitational flow, followed by size exclusion chromatography (SEC) (Fig. 19 (A)/(B)). In order to remove proteins that bind non-specifically to the column, the Ni²⁺-beads were washed in various steps with low concentrations of imidazole. Proteins were collected in storage buffer with 4 mM DTT, to prevent protein aggregation. Purity of the samples were checked by SDS-Page after SEC. SEC fractions (F) from 6 to 10 and 12 mL were collected. Further investigations were carried out with the fractions that showed spectra with least impurities (MS control analysis: CSFV F 8 or 9-10, HCV F 6-7) in Fig. 20 (A)/(B). Surprisingly, in several SEC purifications, a very intense 280 UV signal also occurred in the first eluting fractions. It was hypothesized that these fractions mainly contain co-eluting *E. coli* proteins and artificial higher-ordered assemblies of the NS-proteins. Moreover, for the HCV protein in particular the SDS gels (Fig. 19 (C)) show that, in addition to the suspected band for the

NS3-NS4A target protein (gel band approx. 75 kDa, black arrow), various smaller proteins (in the mass range from approx. 55 kDa to 20 kDa, blue and green arrows) are present in the samples after purification. These could be cleavage products of the NS protein or contaminations which are still present.

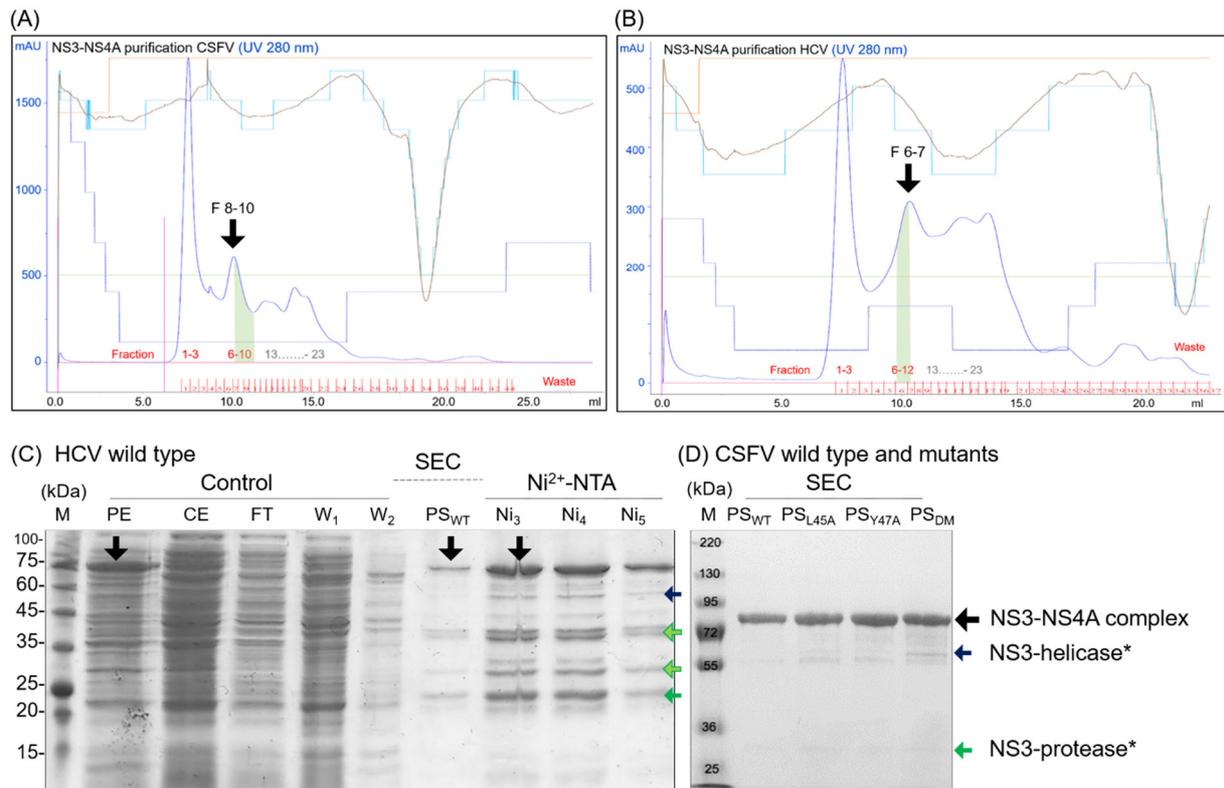


Fig. 19 Protein purification by size-exclusion chromatography (SEC) and control by a 10% SDS gel electrophoresis.

(A) Chromatogram of the NS-protein of CSFV. (B) Chromatogram of the NS-protein of HCV. SEC with Superdex200(10/300) column, UV absorption measured at 280 nm (blue line). Peak used for further analysis marked (black arrows and green bars). (C) SDS gel shows the purity of the NS-protein belonging to HCV after it's corresponding purification steps. The different protein bands indicate: PE = pellet, CE = crude extract, FT = flow through, Ni = fractions 3-5 after Ni²⁺-chromatography that were collected for SEC, PS_{WT} = purified wild-type protein after SEC (fraction 6), M = Roti tricolour protein marker. (D) SDS gel of the purified wild-type (WT) and mutated (DM = double mutant L45A/Y45A, SM = single mutant Y47A, L45A) NS3-NS4A proteins belonging to CSFV after SEC chromatography (M = Fermentas PageRuler prestained protein ladder). For both HCV and CSFV the NS3-NS4A protein bands are indicated by black arrows, blue arrows mark the suspected NS3-helicase, green arrows mark the suspected NS3-protease, light-green arrows mark contaminations, (*) CSFV proteins, that were identified by ESI-MS analyses.

However, these proteins were not identified in the ESI-MS-based control measurements (Fig. 20 (B)). As these various proteins were particularly pronounced in the HCV samples, in addition to the SDS-Gel, further sample quality controls based on bottom-up proteomics and Western blot were carried out with the HCV samples (Fig. 22, Fig. 21). Native MS clearly identified the CSFV wild-type enzyme (Fig. 20 (A)). Briefly, MS data was recorded

at soft acceleration energies and appropriate sample solution conditions in order to obtain the proteins in a native similar state (chapter 2.8, material and methods). Monomer protein concentrations of 3 to 15 μM were tested. At lower protein concentrations, the MS1 signal peaks were too weak and too broad and the signal to noise ratio too high. At higher concentrations there was frequent protein aggregation in the injection capillary, which quickly clogged it and no sufficiently stable nano-spray was achieved, especially when DTT is absent in the MS-sample solution. The best spectra were recorded at protein concentrations of 5 to 10 μM with 1 to 2 mM DTT. Data showed that 10 eV energy and ammonium acetate buffer (200 mM to 250 mM) at pH 7.0 to 7.4 are appropriate for these proteins to preserve their folded conformation and to identify the target protein by their mass to charge ratio. Based on the molecular weight the target proteins were clearly identified with the help of the underlying amino acid sequence (Fig. 17 (B)) and the obtained MS1 data, although there is a small deviation of 146 Da from the monoisotopic mass of the protein in some samples. To explain this mass difference, further tests based on MS2 Top-Down analyses were carried out (chapter 3.3.4). According to the complete tagged protein sequence, the theoretical mass of the full-length CSFV wild-type protein is 83.521 kDa.

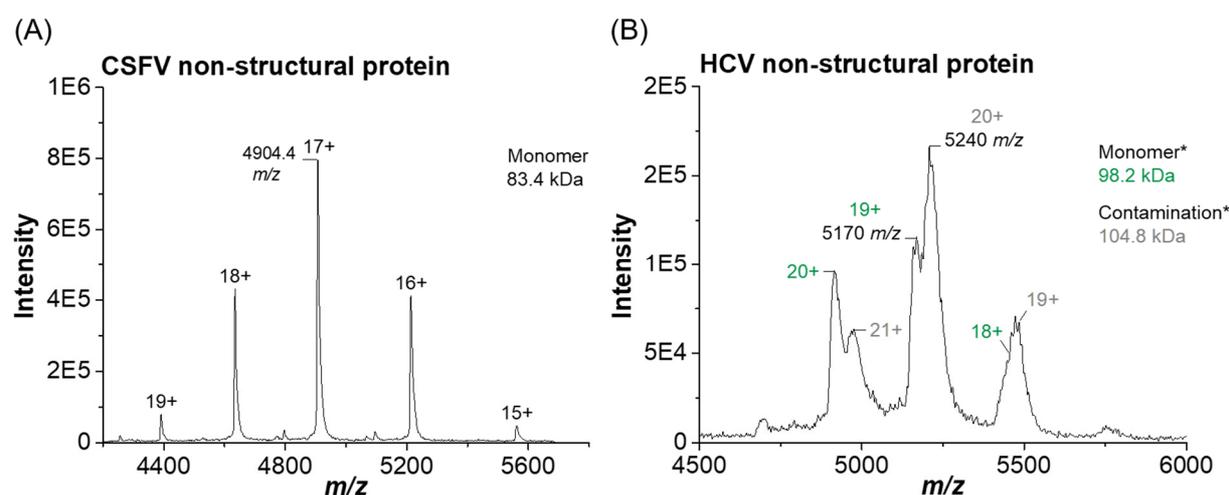


Fig. 20 Native ESI spectra of CSFV and HCV non-structural protein complex NS3-NS4A.

(A) Native MS data of the monomeric wild-type NS-protein of CSFV (83.4 kDa). (B) Native MS data of the supposed monomeric wild-type NS-protein of HCV (*98.2 kDa, green), suspected contamination (*104.8 kDa, grey). Clear identification of the protein was not achieved, mass is too high based on the underlying aa-sequence (should be 74.1 kDa). Conditions: buffer (250 mM ammonium acetate, pH 7.0). Sample concentration: 7 μM (monomer), acceleration energy: 10 eV.

Usually, the identified monomers of the enzymes show a charge state distribution from 19+ to 15+ (Fig. 20). ESI-MS could only provide an ambiguous interpretation of the HCV NS3-

NS4A wild-type protein structure (Fig. 20 (B)). These first results hinted that native MS alone would not be enough to study this protein in more detail.

Although a protein with the expected mass could be clearly identified in the SDS gel after expression and purification (Fig. 19 (C)), the associated monomeric protein mass (74.1 kDa) was not identified in the MS control measurements. A clear identification was not possible in any of the SEC fractions, not even with the help of MS2 analyses. Only in fractions 1-3 and 6-12 an approximately large mass of 98.2 kDa could be measured. Best spectra were recorded for fraction 6-7. Nevertheless, this mass absolutely did not fit the protein sequence (Fig. 18 (B)). Spectra seemed to show another kind of protein which is not related to the expected NS3-NS4A of HCV (see supplement chapter 6.8, S. Fig. 13 to S. Fig. 17). However, the 98.2 kDa protein does not seem to be a contamination, but rather the non-specifically extended/modified NS3-NS4A protein from HCV or even an aggregation of two cleavage products. Therefore, further investigations are necessary to check whether the full-length NS3-NS4A protein is present.

In the following, the importance of the use of native ESI-MS in connection with other orthogonal techniques (proteome analysis, Western blot and FRET test) for the interpretation of these unclear native MS and MS2 spectra (see supplement chapter 6.8) is shown.

3.1.1 Proteomics reveals the presence of the HCV NS3-NS4A protein

The fact that the protein with all its subunits appears to be present in the analysed samples after purification and exchange into ammonium acetate solution, albeit in a previously unknown structural conformation, could be demonstrated using proteomics.

The proteomics approach carried out in this study is to be regarded as a supplementary quality control of the protein purification of the HCV samples, as SDS-Page (Fig. 19) and native ESI-MS (Fig. 20) based controls indicated contaminations. Native ESI-MS in contrast to the SDS-Page control did not allow an identification of the expected NS3-NS4A target protein mass. The proteomics experiment was performed with solution exchanged (200 mM ammonium acetate, pH 7.2) SEC fractions (F1-2, F6-7). To check whether the protein is still present after buffer exchange or whether the protein is lost as a result of this process, proteomic sample preparation was done after buffer exchange to ammonium acetate solution. Another goal of this experiment was to check whether the 98.2 kDa and

104.8 kDa proteins measured in the native state and the proteins identified by SDS-Page (Fig. 20 (B)) could be *E. coli* contaminants. To ensure correct execution the samples were handled according to the protocol in 2.6. This experiment is not designed to provide detailed information about specific cleavages within the NS3-NS4A complex of HCV, nor to assign peptides to specific amino acid segments of the target protein. Since no protein with the expected mass of 74.1 kDa could be detected in the native control measurements, the implementation of this approach is only designed to check whether the NS3 target protein is present after Ni²⁺-NTA in combination with the second purification step via SEC and solvent exchange.

For this study, SEC fractions 1-2 and 6-7 were pooled and proteolyzed with the serine protease trypsin. According to the proteomics protocol (chapter 2.6) cysteines were reduced using DTT. Followed by chemical blocking of the free cysteines before proteolytic cleavage in solution [176]. Trypsin is an endopeptidase with high specificity and cuts the peptide backbone at the carboxyl-side of the basic amino acids R (arginine) and K (lysine) [177]. Usually, treating proteins with trypsin results in peptides with a size of 600 to 1500 Da [129, 178]. Under acidic condition (as used in conventional reverse-phase ESI) the C-terminal R and K peptides are positively charged, which makes them detectable by MS. Peptides are then identified based on their MS2 fragment spectra and matched to a theoretical *in-silico* spectrum from the database. This results in peptide spectrum matches (PSMs) [156, 179]. The amount of PSMs can be used as a proxy for relative quantitation (spectral counting) although it is not as precise as methods that use area under the curve [180]. Since the proteins were expressed in *E. coli* cells, the complete *E. coli* protein (strain k12) UniProt database was used as background. This allows detection of possible contamination caused by *E. coli* in the sample. The total number of identified PSMs is shown in Fig. 21. As a result of this quality control unique peptides belonging to the NS3-NS4A protein complex were clearly identified (Tab. 7). Data show that a large amount of the target protein is in the pooled SEC fractions 1 and 2 (F1-2: 339 PSMs) and to a lesser degree in the fractions 6 and 7 (F6-7: 74 PSMs).

The percentage of the protein sequence that is covered by identified peptides (Tab. 7, Cov. (%)) indicate that not all associated peptides for the NS-protein were found (approx. 53% for F1-2 and 46% for F6-7). Thus, the data only covers a part of the aa-sequence. This could indicate that the NS-protein is incomplete. However, when considering the coverage,

it must be noted that this does not necessarily mean that only the identified parts of the protein are present in the sample. In contrast, a large protein for which only a few peptides have been identified has very low coverage and a small protein with a smaller number of identified peptides has very high coverage, although the large protein is knowingly present in a higher concentration [181, 182]. A protein sequence coverage of 100% is therefore very rare. This is amongst other things the result of the difference in ionisation efficiency of different peptides [183]. To estimate the relative quantity of the protein the PSMs are used in conjunction with the number of identified unique peptides. The higher the PSM value, the more MS2 spectra were assigned to this hit [182].

In accordance with established standards for protein identification “in a mixture”, as is the case here shown by the SDS-Page in Fig. 19, to be specific, proteins should be identified with at least two unique peptides [184]. The PSMs thus provide a quick overview of the relative quantity of the proteins in the sample. However, the table (Tab. 7) also shows proteins that were only identified by one unique peptide as these could also very well be true positive hits (and are already controlled with 1% false discovery rate, FDR) and may give a hint at which contaminants are present in the sample.

To minimize events of false positives, in accordance with established standards for calculating the false discovery rate (FDR) [184] it was set to 1% at peptide and protein level.

The data show that the samples contain various contaminations besides the NS-proteins (Fig. 21). This is mainly based on certain *E. coli* proteins, which were identified by their unique peptides (Tab. 7).

Data show, that besides the target protein there is a larger amount of these *E. coli* contaminants in the first SEC fractions (F1-2), which is also indicated by the high number of associated peptides shown in Tab. 7. This could explain why, despite the larger amount of target protein, the native MS data of these fractions does not provide good spectra. The data are based on the database of the entire *E. coli* genome and a database of typical laboratory contaminants (i.e. human keratin etc.). The results are therefore based on a comprehensive screening of a large number of possible contaminants (Fig. 21).

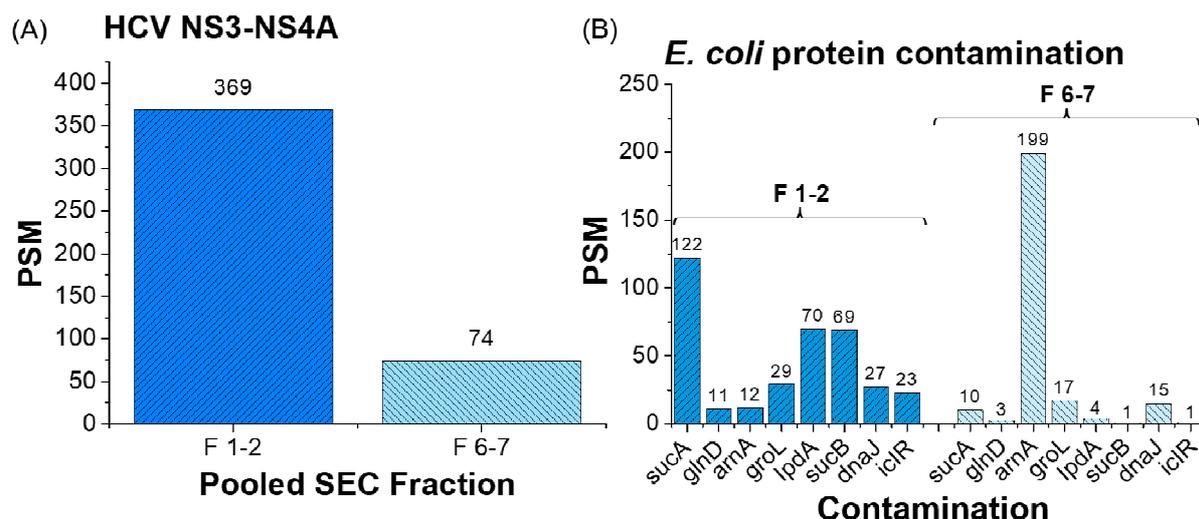


Fig. 21 The proteomics approach revealed significant peptide spectrum matches (PSMs) for the target NS3-NS4A protein of HCV.

(A) Number of PSMs for the target NS3-NS4A protein in pooled fractions of the SEC (1-2 and 6-7). (B) Identified contaminations mainly belong to certain *E. coli* proteins (gene names are indicated).

However, for both samples (F1-2 and F6-7) no contamination was identified that could match the 98.2 kDa protein. Nevertheless, there is a huge amount of the *E. coli* protein ArnA (gene name: *arnA*) in F6-7, which almost corresponds to the expected mass of 74 kDa. This protein mass was also identified in the SDS gel, whereby the question arises as to whether this could also be the ArnA protein of *E. coli* (UniProtKB code: P77398). Fact is that the aa-sequence of ArnA consist of many histidine residues (27xHis). Some of them are structurally exposed (PDB: 1Z7E) [185]. In Ni²⁺-chromatography, proteins which are untagged but consist of various histidine residues can bind non-specifically because His exhibits a strong interaction with immobilized metal ion matrices, including Ni²⁺-beads [186]. The aa-sequence of ArnA contains two pairs of adjacent histidine residues, which means that this protein has a high affinity for the Ni²⁺-matrix [186].

Thus, it is possible, that both the 10×His tagged target protein and the *E. coli* protein co-eluted. It is therefore assumed that both proteins are in the sample after the Ni²⁺-NTA purification, which could also explain why the protein bands at approx. 75 kDa in the SDS gel are significantly more intense. Nevertheless, according to the SDS-Page results after SEC also various bands of smaller proteins were observed (Fig. 19 (C)).

On the one hand, these proteins could be cleavage products of the NS3-NS4A complex, such as the released NS3-helicase domain (51.2 kDa), which would bind non-specifically (10 His-residues within its aa-sequence) to the Ni²⁺-beads due to the missing of a His-tag

or the protease domain (22.2 kDa, including N-term. NS4A, linker, TEV and 10xHis-tag) as shown in the aa-sequence in Fig. 18. Based on the proteomics approach carried out here, however, no quantitative determination of the amount of the free helicase/protease subunit can be carried out, which is why only PSMs and peptides of the entire NS3-NS4A complex could be taken into account in the data analysis.

On the other hand, these proteins could also be contaminants. The presence of several non-target proteins is confirmed by the proteomics results, which clearly show that the HCV sample is still contaminated with *E. coli* proteins despite the two orthogonal purification steps (Ni^{2+} -NTA and SEC) (Fig. 21, Tab. 7).

Based on the proteomics quality control, two proteins in the mass range of 50-60 kDa were identified, which also corresponds to the mass of the possibly released NS3-helicase subunit (51.2 kDa) from the NS3-NS4A protein complex. This could mean that these proteins also elute after Ni^{2+} -NTA purification and that the protein band in the SDS gel originate from one or both of these *E. coli* proteins and is not based on the free NS3-helicase subunit. The first *E. coli* protein in this mass range shows 29 PSMs and is a 60 kDa chaperonin (GroEL), which is encoded by the gene *groL* (UniProtKB - P0A6F5). GroEL assists the folding of recombinant proteins, which are over-expressed in *E. coli* and thus is often observed in such experiments [187]. Its aa-sequence only consists of a single histidine. Usually, it should have been washed off the Ni^{2+} -Column and should not coeluted with the His-tagged NS3-NS4A protein.

Nevertheless, an unknown amount of the protein seems to be able to stick to the Ni^{2+} -beads, which is maybe through co-aggregation/interaction with a His-rich protein and was thus transferred to the SEC after the Ni^{2+} -purification and identified by the proteomics analysis.

The second *E. coli* protein shows 70 PSMs. The protein is encoded by *Ipda* (UniProtKB - P0A9P0) and is a lipoamide dehydrogenase of approx. 51 kDa, which is close to the mass of the NS3-helicase domain. The aa-sequence of that protein contains 13 His-residues. The His-residues are divided over the entire sequence, with C-terminal three His only a few amino acid residues apart from one another. Based on this, coelution with the target protein cannot be ruled out.

The PSM value (F1-2 = 70, F6-7 = 4) of IpdA is lower compared to NS3-NS4A, which could indicate that, due to the missing of a His-tag only a smaller amount of the protein was able to bind non-specifically to the Ni²⁺-NTA.

However, because this protein can theoretically bind to the Ni²⁺-matrix, it is possible that there is an overlay of this protein signal with that of a possibly non-specifically bound NS3-helicase subunit in the SDS gel in the 51 kDa range.

Additionally, as presented in Tab. 7 *sucA* is a gene, that encodes a protein of approx. 105 kDa, which shows huge number of PSMs in fraction 1 and 2 (F1-2 = 122 PSMs) and is much lower in fraction 6 and 7 (F6-7 = 10 PSMs). The *E. coli* protein encoded by *sucA* (UniProtKB - P0AFG3) is a dehydrogenase (OGDH) complex, that contains 34 His-residues within its aa-sequence.

These are distributed over the entire sequence. Some of the His-residues are only one amino acid away from each other, with two His being directly adjacent close to the C-terminus of the protein. It is therefore assumed that this protein coelutes with the NS3-NS4A protein via Ni²⁺-chromatography. Furthermore, it is hypothesised that the 104.8 kDa protein that was identified in the native ESI-MS control measurement (Fig. 20 (B)) is OGDH encoded by *sucA*.

In addition, based on proteomics two proteins with low PSM values (10 to 2 PSMs) were detected in the mass range of 15 to 20 kDa (*rpII* = 15.8 kDa and *ssb* = 18.9 kDa) (Tab. 7). Both have hardly been identified in the SDS gel after protein purification. Based on their very low PSM values, however, it can be assumed that these proteins have less relevance for the contamination of the HCV samples.

Tab. 7 Proteomic analyses based on proteolytical cleavage of the NS3-NS4A protein of HCV by trypsin revealed several *E. coli* contaminations after SEC chromatography.

PSM = peptide spectrum matches, AAS = amino acid sequence (total number of aa in the protein sequence in the library), Cov = coverage, MW = molecular weight. The *E. coli* contaminants are labelled with the gene names of the respective proteins.

SEC F1-2	NS3-NS4A	sucA	glnD	arnA	groL	lpdA
MW (kDa)	73.796	104.996	102.326	74.242	57.293	50.657
PSMs	369	122	11	12	29	70
Unique peptide	31	37	4	10	18	18
AAS	692	933	890	660	548	474
Cov. (%)	52.890	42.122	7.079	20.455	39.051	47.890
SEC F1-2	sucB	dnaJ	iclR	thiD	ssb	rplI
MW (kDa)	43.984	41.074	29.720	28.616	18.963	15.759
PSMs	69	27	23	10	10	9
Unique peptide	20	9	19	5	4	5
AAS	405	376	274	266	178	149
Cov. (%)	45.432	32.447	43.431	25.940	28.090	36.913
SEC F6-7	NS3-NS4A	sucA	glnD	arnA	groL	lpdA
MW (kDa)	73.796	104.996	102.326	74.242	57.293	50.657
PSMs	74	10	3	199	17	4
Unique peptide	29	7	3	39	13	4
AAS	692	933	890	660	548	474
Cov. (%)	45.520	11.040	4.944	57.727	36.131	9.072
SEC F6-7	sucB	dnaJ	iclR	thiD	ssb	rplI
MW (kDa)	43.984	41.074	29.720	28.616	18.963	15.759
PSMs	1	15	1	4	2	4
Unique peptide	1	8	1	2	2	4
AAS	405	376	274	266	178	149
Cov. (%)	2.222	26.064	4.380	9.774	16.292	29.530

Furthermore, the SDS-Pages (Fig. 19 (C), Fig. 22 (A)) show a band between 25 to 30 kDa after Ni²⁺-NTA as well as after SEC, which could be the *E. coli* proteins (iclR = 29.7 kDa and/or thiD = 28.6 kDa) identified by proteomics (Tab. 7). Both *E. coli* proteins are not remarkably high in PSMs (iclR: 1-23 and thiD: 4-10 PSMs) in comparison to the NS-protein (F1-2: 369 and F6-7: 74 PSM). The gene iclR (UniProtKB - P16528) encodes for the transcriptional repressor IclR, which contains 13xHis-residues distributed over the aa-sequence. The *E. coli* protein encoded by thiD (UniProtKB - P76422) is a kinase, which aa-sequence consists of eight His-residues, whereby three of them are located close to one another, including two neighbouring His. Based on this, a coelution of the thiD encoded kinase with the target protein after Ni²⁺-chromatography can be assumed but cannot be ruled out for IclR either.

Overall, data show that the target protein is present in both, F1-2 and F6-7, whereas for the first SEC fractions a high PSM value with a high number of peptides could be assigned to the NS3-NS4A protein. However, these fractions also contain the largest amount of *E. coli* proteins. Especially, *sucA* (105 kDa) and *IpdA* (51 kDa) are shown to be the most dominant contaminants. In contrast to this, the SEC fractions F6-7 show a lower amount of contaminants, but the contamination by *ArnA*, which has a similar mass to the target protein, is very high here. According to the principles of the SEC, large proteins and complexes should elute first and smaller ones later [188].

Since the *sucA* encoded OGDH is a large protein, it cannot migrate through the pores of the SEC matrix and as expected, eluted in the first fractions, while molecules of smaller size should migrate through the pores and thus only elute later. However, significantly smaller proteins compared to OGDH such as the 60 kDa protein encoded by *gorL* and the 51 kDa encoded by *IpdA* and a larger amount of the NS3-NS4A target protein also elute in the first two fractions, which might indicate that these proteins tend to form aggregations.

For a clear identification of the proteins in each band in the SDS gel, a Bottom-up-in-gel digest in addition to the in-solution test could be carried out. Although for the purpose of this experiment the in-solution digestion was completely sufficient as a quality control of the HCV samples, as it revealed specific contaminations as shown by the overall results. As to why these contaminations do not seem too problematic in the CSFV sample a possible explanation would be that a higher percentage of the sample consist of the target protein when compared to the HCV sample. This could be caused by the different plasmids used for expression. Resulting in lower amounts of contaminants being injected for native MS as well as lower amounts of contaminants separated on SDS-Page when compared to the target protein. The data also suggest that the HCV protein does not fold correctly during expression in *E. coli*, at least to some degree. This seemed not to be the case with the NS3-NS4A complex from CSFV, so there is more CSFV protein and fewer chaperones. For further analysis of the HCV samples, a Western blot was carried out (3.1.2). It allowed the identification of the NS3-helicase domain of the HCV protein complex after Ni^{2+} -NTA and within SEC fraction 2 and 3. Here the NS3-NS4A protein was already identified by the proteomics approach (369 PSMs, F1-2) and the 98.2 kDa protein by native MS. Additionally the SEC fraction 11 was analysed (Fig. 22).

3.1.2 Identification of the HCV NS3-NS4A protein by Western blot

To corroborate the results of the proteomics experiment, a Western-Blot of some of the SEC fractions (in - 80 °C storage buffer) was carried out. In addition, a cell pellet fraction was examined (Fig. 22). Two different primary antibodies were used, a mAnti-His and a mAb-NS3 antibody known to detect the NS3 protein of *Pestiviruses*. For the used mAb-NS3 antibody two epitopes in the CSFV NS3 protein (aa364 to aa373 = HPIEEFIAPE) and (aa380 to aa393 = LGSQFLDIAGLKIP) are known. Both are located in the region of the NS3-helicase [189]. Comparable to the first epitope of CSFV a structurally similar epitope is located within the HCV NS3-helicase domain (aa333 to aa342 = HPDIEEVGLG).

Thus, it is possible to detect the HCV NS3 protein with the mAb-NS3 antibody. Interestingly, the results show that the NS3 protein is present (mAb-NS3 antibody) but not at the correct mass (approx. 74 kDa). Instead, a specific protein in the mass range of 50 to 55 kDa (~ 51 kDa) was detected (Fig. 22 (A), Fig. 19 (C), SDS gel blue arrow). Based on its specific interaction with the mAb-NS3 this protein was identified as the NS3-helicase, which is present in the protein mixture after Ni²⁺-NTA and within SEC fraction 2, 3 and 11 as shown in Fig. 22 (B)/(C). Further, no protein with the expected wild-type mass was identified in the Western blot using the mAnti-His antibody. In contrast, two proteins of approx. 51 kDa and one in the range of 15 to 20 kDa (~ 19 kDa), which are supposed to carry a His-tag, were identified. It was hypothesised that one of the proteins in the 51 kDa range is part of the NS3 protein, as it was identified by both antibodies in the Western blot. One possibility would be that it is a His-tagged part of the NS3-NS4A target protein, which is probably the N-terminal section consisting of the NS4A-cofactor, the NS3-protease and part of the NS3-helicase. However, this would mean that the protein would be cleaved at a position in the NS3-helicase, which is located behind the corresponding binding motif of the specific mAb-NS3 antibody (e.g. in the range of amino acid D480), which cannot be excluded because of non-specific cleavage products.

Another hypothesis is that the 51 kDa protein, binds non-specific to the mAnti-His antibody via histidine residues within its aa-sequence. The NS3-helicase domain of the NS3-NS4A protein complex of HCV has a total of 10xHis. Five of them (His545, His541, His528, His333, His246) are structurally exposed and labelled in green in the x-ray structure in Fig. 22 (C). However, due to the proteins being in a denatured state after SDS-Page these

His residues are structurally far apart from each other and would be unlikely to be recognised by the antibody. The data from the SDS gel also indicate that the SEC fraction (1-10) contain smaller proteins and probably *E. coli* GroEL chaperones in the range of 60 kDa (marked by black stars). These could be responsible for the formation of the 98.2 kDa complex. Fraction 11 appears to be free of these proteins.

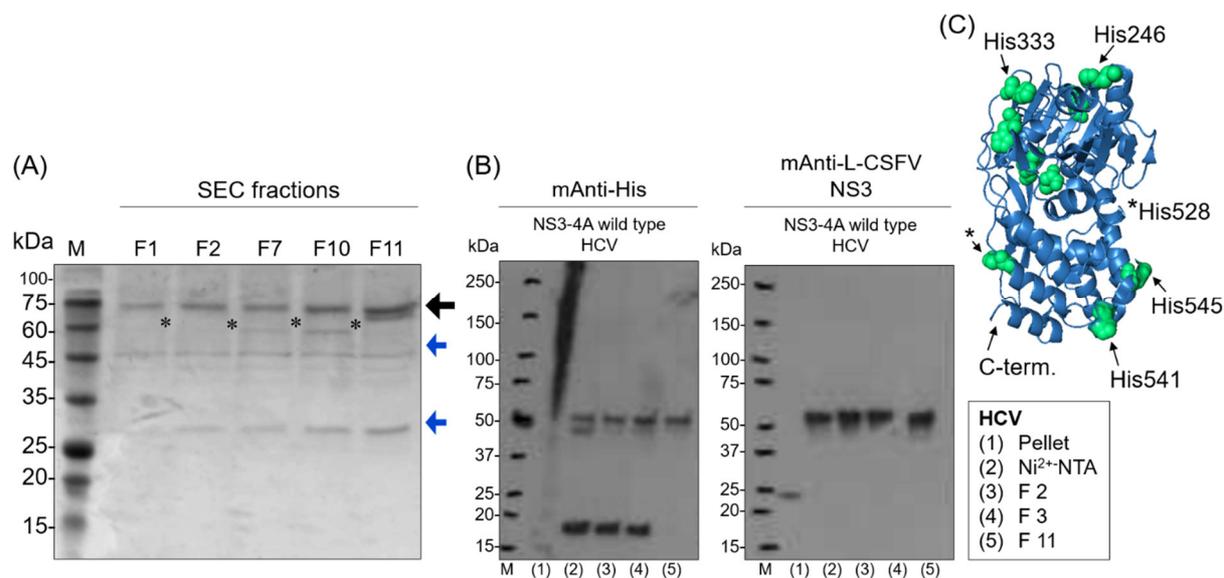


Fig. 22 Western blot showed no HCV NS3-NS4A protein complex, which corresponds to the size of 74 kDa.

(A) SDS-Page as control for SEC fractions of HCV NS3-NS4A. This gel was not the basis for the Western blot. Black arrow = kDa range of the target protein NS3-NS4A, black stars = range of the *E. coli* chaperone GroEL, blue arrows = range of possible cleavage products of NS3-NS4A or *E. coli* protein contaminations. (B) Western blot: exposure time of the primary antibody mAnti-L-CSFV NS3 (code 4 mAb-NS3) = 60 seconds, exposure time of the primary antibody mAnti-His = 10 min. Gel loaded with 10 μ L of the protein (25 μ g of the total protein based on cell lysate), gel percentage = 10%, SDS marker = Roti tricolour protein marker, Western blot marker = Bio-Rad precision plus protein, Kaleidoscope standard. (C) x-ray structure of the NS3-helicase domain of HCV shows 5 structural exposed histidines (green sphere), crystal structure is modified based on Gu and Rice 2016 [190] and PDB: 5E4F. Amino acid labelling is based on the complete NS3 protein aa-sequence.

However, based on the Western blot, the mAnti-His antibody detects another protein in the range of 50-51 kDa (Fig. 22 (C)). This could be the IpdA encoded *E. coli* lipoamide dehydrogenase (approx. 51 kDa), which was identified in the proteomics test (Tab. 7). The protein carries 13 His-residues divided over the entire sequence, which can also interact non-specifically with the mAnti-His antibody.

The 19 kDa protein could be a shorter section of the full-length NS3 protein that does not have a helicase part. However, the question arises as to why the complete NS3-NS4A protein of HCV is not recognized, neither by the mAnti-His nor by the mAb-NS3 antibodies. The obtained results may indicate that the full-length protein cannot be detected because it

is cleaved and therefore only the two cleavage products NS3-helicase and protease/NS4A cofactor could be identified in the Western blot. This could also indicate that the HCV protein is overall more fragile than the CSFV protein under the conditions used.

In summary, the results of the SDS-Page and Western blot analysis did not fit the native MS data. Using trypsin for proteolytic cleave of the protein mixture of certain SEC fractions based on a Bottom-up proteomics approach, peptides which clearly belong to the NS3-NS4A fusion protein could be identified (Fig. 21).

However, Bottom-up proteomics does not provide any information as to whether the NS3-NS4A protein was present as a native protein complex or whether the specific peptides detected originate from its subunits. Nevertheless, this test clarified that the protein was present in a previously unknown state in the samples examined. On the one hand, it would be possible that, particularly in the native MS measurements, different domains/subunits of the “self-cleaved” NS3-NS4A protein (separated/splitted NS3-helicase and NS3-protease + NS4A) form aggregates with one another, resulting in molecular weights of 98.2 kDa and 104.8 kD. On the other hand, if the full-length protein is present, it could also interact with released NS3-cleavage products or even with a smaller *E. coli* contaminant, which carries many accessible His-residues (e.g. thiD kinase). This interaction would lead to a mass of 102 kDa, which is close to the mass of 104 kDa measured with native MS. Although the native ESI-MS-solution was reduced using 2 mM DTT. Therefore, formation of intramolecular Cys-bridges should be improbable.

Nevertheless, as described above, it was possible to identify the NS3-helicase of the NS3 protein of HCV using the Western blot and the mAb-NS3 antibody. A simplified FRET test, which is presented below (3.1.3), was also able to demonstrate the activity of the NS3-protease in the samples.

3.1.3 HCV NS3-protease shows activity when subjected to a FRET substrate

Another sample quality control was carried out based on a simplified FRET-assay. Here, activity of the “purified” and solution exchanged NS3-protease is demonstrated. The previously shown mass spectrometry investigations with the HCV NS3-NS4A protein complex (chapter 3.1, Fig. 20) did not provide any meaningful results comparable to the also investigated CSFV protein. In addition to the difficulties in expressing and purifying the

proteins, where the yield of purified protein seemed to be very low, the ESI-MS based protein analysis of various test-conditions (*150 mM, 200 mM and *250 mM ammonium acetate solution at pH 6.5, 7.2 & 7.4 and *8.0, * = data not shown) show recurrent protein ionisation problems. Impurities in a similar mass range of the NS3 target protein occur frequently, which is also shown by the above-described proteomic quality control (chapter 3.1.1). Thus, in the further course it was decided to check this impure protein mixture with another orthogonal and quick method. The approx. 51 kDa protein which was identified as the helicase subunit released from the NS3 complex by means of Western blot, indicate the presence of parts of the target complex.

The smaller 19 kDa protein, which is suspected to carry a His-tag, could be the NS3-protease in connection with the N-terminal section of the NS4A cofactor. Since, based on the proteomics data presented, it can be assumed that both helicase and protease are contained in the sample, the simplified FRET test is used to check whether an enzymatically active NS3-protease can be detected in the protein mixture after Ni²⁺-NTA purification.

For this purpose, a simple test setup was used.

The test is based on the HCV RET S1 FRET substrate (sequence: Ac-Asp-Glu-Asp(EDANS)-Glu-Glu-Abu-Ψ-[COO]-Ala-Ser-Lys(D-ABCYL)-NH₂) obtained from AnaSpec. The chosen substrate results in a fluorescence signal at 355/510 nm excitation/emission and is able to detect the NS3-protease activity with a sub-nanomolar protein concentration. The obtained results are shown in Tab. 8, Fig. 23. The increase of the fluorescence signal demonstrates the cleavage and degradation of the RET S1 substrate by the active NS3-protease. The release of the related product is monitored (Fig. 23). In addition, the experiment is based only on the analysis of a single concentration of the substrate, which according to Beran and Pyle 2008 [191] is in a state of complete substrate oversaturation. The HCV protein concentration of pooled Ni²⁺-fractions (Fig. 19 (C)) was 0.045 μM (protein mix, determination of the concentration is based on the extinction coefficient of the HCV protein at A280) and 4.8 μM RET-S1 starting concentration (in 100 μL) at time point T = 0 min (Tab. 8).

Based on the conditions used, the situation of the maximum rate of reaction (V_{\max}) is simulated. Its calculation is limited to the rate of reaction $V = [S]/t$, which assumes a linear course of the substrate decrease and product increase under the conditions used. Because

of this simplified method, this assay is termed simplified FRET test; however, data allow the indirect calculation of the decreased substrate concentration (Fig. 23).

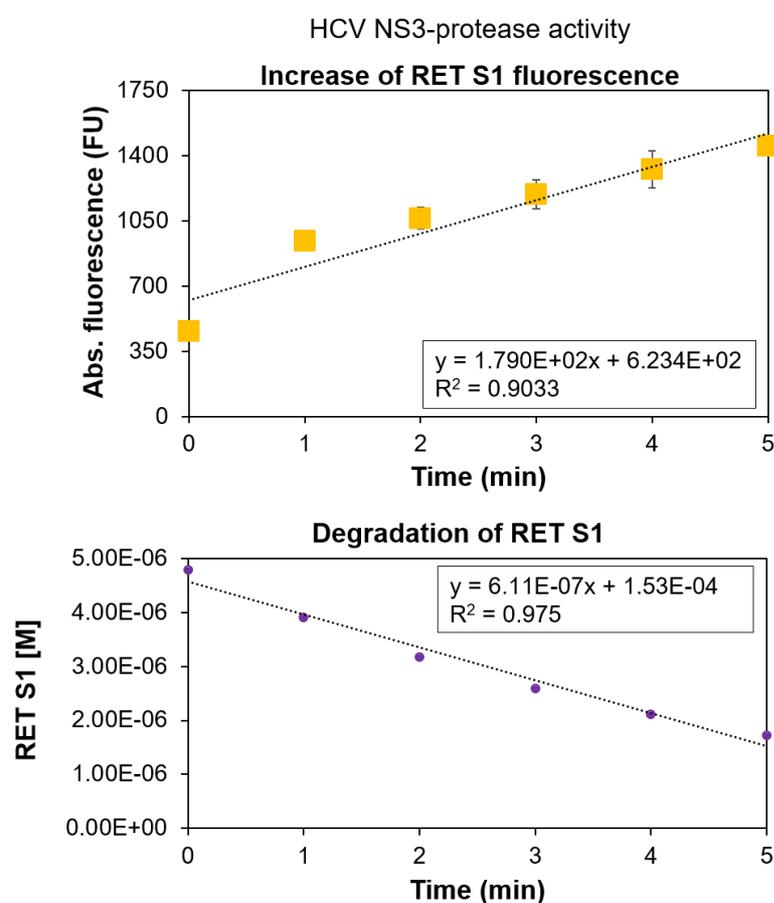


Fig. 23 Simplified FRET assay to prove the activity of the wild-type NS3-NS4A protein of HCV in a complex enzyme mixture.

The increase of fluorescence signal was monitored over five different time points. Above: data show the increase of the absolute fluorescence (FU). Below: data show the decrease of RET S1 substrate concentration. Data are based on a fixed substrate concentration, which simulates oversaturation and the situation of V_{\max} .

The general reaction scheme of enzyme catalyses [192] is described in the following:



The calculation of V_{\max} is described as shown in equation (Eq. 10), where (V) simulates the state V_{\max} . The corresponding equation is based on the slope (m) of the linear regression (Fig. 23), where (V) = (m):

$$m = \frac{\Delta y}{\Delta x} = V = \frac{\Delta y}{\Delta x} = \frac{\Delta[S]}{\Delta t} \quad \text{Eq. 10}$$

For the simplified determination of k_{cat} , the HCV sample is treated as a "pure" protease without impurities. Because only in this way does k_{cat} depend on the total enzyme concentration $[E_T]$ and V_{max} . This leads to a linear relationship as described in the following equation (Eq. 11). However, since it is known that HCV is not a pure sample, the calculations based on HCV only give approximate values for k_{cat} . These are therefore marked with (*) in the following. The calculation of k_{cat} is based on the total enzyme concentration $[E_T]$ and V_{max} as described in Eq. 11:

$$k_{cat} = \frac{V_{max}}{[E]_T} \quad \text{Eq. 11}$$

The data demonstrate that the NS3-protease is present and shows a detectable activity within the impure protein mixture. The velocity of reaction is based on the slope of the plot (Fig. 23) and leads to $V_{max} = 6.11E-07 \text{ M/min} = 1.018 \text{ pmol/s}$. Thus, for the state V_{max} it is possible to calculate k_{cat} based on the assumed enzyme concentration. Using Eq. 11 the value for $*k_{cat} = 0.226 \text{ s}^{-1}$.

Tab. 8 HCV NS3-NS4A protease activity assay based on the RET S1 FRET substrate.

Measurements are based on six different time points of reaction and on triplicate measurements. Substrate concentration at $T = 0 \text{ min}$ was $4.8 \text{ }\mu\text{M}$ RET-S1 starting concentration (in $100 \text{ }\mu\text{L}$). Reaction temperature was $+ 25 \text{ }^\circ\text{C}$. M_1 - M_3 = sum of triplicate intensities, Av = average, SD = standard deviation, $[M]$ = molarity.

Time (min)	M_1 Abs. Fluorescence	M_2 Abs. Fluorescence	M_3 Abs. Fluorescence
0	433	410	520
1	881	979	960
2	1059	1116	1016
3	1229	1236	1110
4	1403	1353	1220
5	1570	1451	1331
(min)	Av_{M1-M3}	SD	$[M]_{\text{Substrate}}$
0	454	47	4.80E-06
1	940	42	3.91E-06
2	1064	41	3.18E-06
3	1192	58	2.59E-06
4	1325	77	2.11E-06
5	1451	98	1.72E-06

Together, the proteomics data, the Western blot results and the FRET data indicate, that the target protein is present. Based on the FRET results the activity of the NS3-protease was shown. Therefore, it was decided to use the presented HCV protein in addition to the

CSFV protein for the development of a proof-of-concept ESI-MS based activity assay, which is shown in more detail in chapter 3.6 but exclude it from further native MS experiments.

However, based on the results presented, more detailed investigations on the HCV protein should be carried out with focus on improving the expression and purification protocol, as well as using Bottom-up approaches to clarify the missing of the target protein in Western blot and in native MS measurements as well as for quality control of contaminants.

In the interest of scope and time of this thesis it was decided that further experiments were conducted using the CSFV NS3 protein complex, which were easily identified by native MS. In order to make a more precise statement about the composition and stability of the CSFV protein, reducing and non-reducing buffer conditions were investigated.

3.2 Effect of cysteine bridges on the CSFV protein assemblies

The CSFV protein could be clearly identified by native MS using the theoretical amino acid based mass of the fusion protein (Fig. 17, Fig. 20 (A)), although a further look into the spectrum indicates that more than two different oligomeric states are present in the sample. Fig. 24 shows a complete spectral overview of the protein complex with its variety of structural forms. These early data provide an indication of the fundamental composition of the CSFV enzyme as well as the influence of solution conditions, protein concentration and instrument settings on protein oligomerization.

In the following, investigations were performed with a high protein concentration (15 μM) and without sample reduction to elucidate the effect of cysteine bridges on protein oligomerization in native MS measurements (Fig. 24).

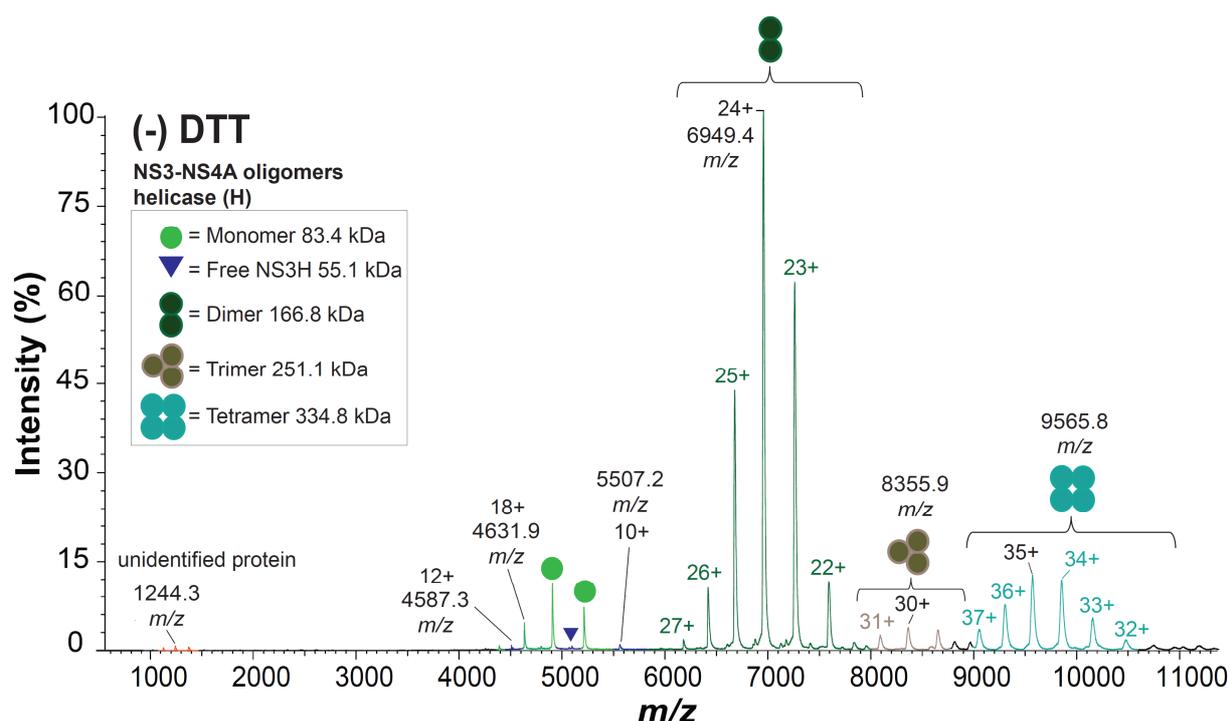


Fig. 24 Native MS data of the CSFV NS3-NS4A oligomers released when protein concentration is 15 μM and no DTT is added to the protein mixture (-) DTT.

Native ESI-MS measurements were performed in MS1 positive ion mode at 25 eV, 250 mM ammonium acetate without DTT. Protein concentration was 15 μM (monomer). Monomer = light-green, dimer = dark-green, unbound NS3-helicase = blue, trimer = brown, tetramer = cyan, unidentified protein = orange.

The monoisotopic protein masses were used to describe the peaks. In addition to the monomer (83 kDa), a dimer with a charge state distribution of 27+ to 22+ and mass of 167 kDa

was found in the higher mass to charge range (6000 - 8000 m/z). Furthermore, in the mass to charge range of about 8000 - 9000 m/z a minor amount of trimer was identified with a mass of 251 kDa (charges from 31+ to 29+).

Moreover, a tetrameric complex with a mass of 334 kDa (charges 37+ to 32+) was detected in the mass to charge range from 9000 to 15000 m/z . Additionally, in the lower mass to charge range (1000 - 1500 m/z) a likely smaller protein was detected that was not clearly associated with the NS-protein (Fig. 24). Considering its charge state distribution, it has a mass of approximately 13 kDa.

Interestingly, the data show that under the conditions used, the most dominant species is the dimer, followed by the tetramer. The monomer and the trimer are present in lesser amounts. Here, the amount of monomer and trimer seems to correlate with the occurrence of the tetramers. The data may imply that a certain rate of oligomerization has a biologically relevant function for the protein complex. It is controversially discussed, but different studies based on other members of the *Flaviviridae* family indicate that the full-length NS3 fusion protein and especially the helicase part of it forms higher ordered oligomers, which is supposed to be necessary for its enzymatic activity [172, 193]. The native MS data presented in Fig. 24 suggest that dimers are the preferred form.

Both, dimers and monomers are also discussed in the literature as the active form, which play an important role in the viral life-cycle [171].

Furthermore, MS data indicate that acceleration energies could influence the formation of oligomers (Fig. 25).

When applying an energy level of 10 eV while measuring, the dimer and tetramer are the dominant protein species (Fig. 25 (B)). When the energy is increased to 25 eV (Fig. 25 (A)) the dimer and tetramer signals decrease, while the monomer and trimer signal increase.

However, since 25 eV are not extremely high energies, this suggest that the trimer as well as a monomer are split of the tetramer indicating a weak interaction while the dimer signal remains the dominant signal. This could be a further indication that the dimer is the preferred and possibly active protein form under the tested conditions.

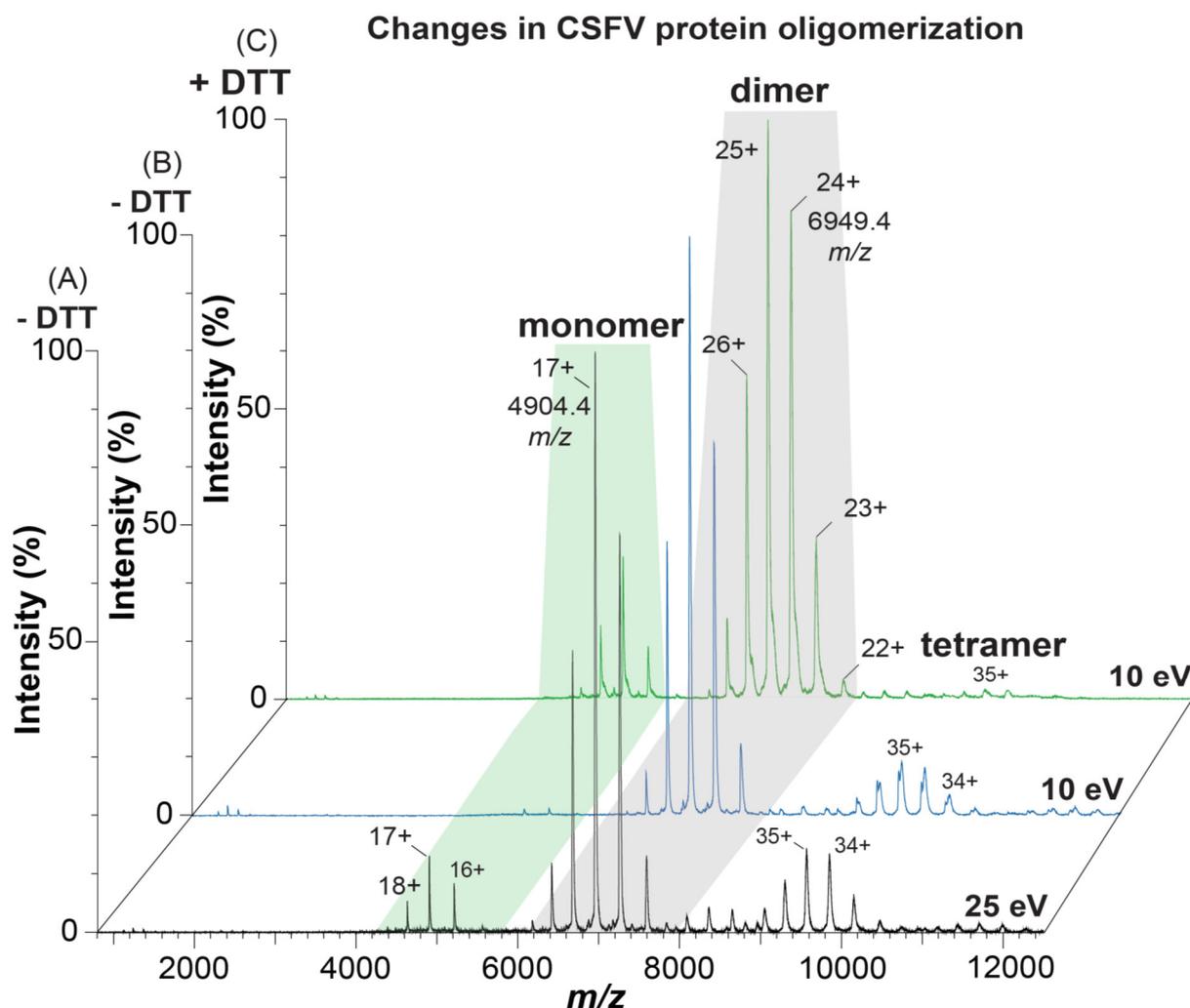


Fig. 25 Oligomerization of the CSFV NS3-NS4A protein depends on ion acceleration energy, protein concentration and the accessibility of cysteine residues.

(A) spectrum at 25 eV, without DTT. Protein concentration = 15 μM (monomer). (B) spectrum at 10 eV, without DTT. Protein concentration = 15 μM (monomer). (C) Native spectrum at 10 eV, 250 mM ammonium acetate and 0.5 mM DTT in the sample buffer. Protein concentration = 10 μM (monomer). Native ESI-MS measurements were performed in positive ion mode.

Additionally, as shown in Fig. 25 (C) reduction of the protein Cys-bridges using DTT influences the formation of oligomers.

In the following, to prove the effect of Cys-bridge formation, the enzyme was reduced using 0.5 mM DTT and a slight decrease of the protein concentration to 10 μM was introduced. The reduction of the enzyme and the decrease of the protein concentration revealed the shrinkage of the trimer and tetramer. The data indicate the protein monomer and dimer conformations as dominant species (Fig. 26).

Thus, the data imply that higher ordered oligomers (trimer and tetramer) are formed due to higher protein concentrations and through the interaction of cysteines within the protein structure.

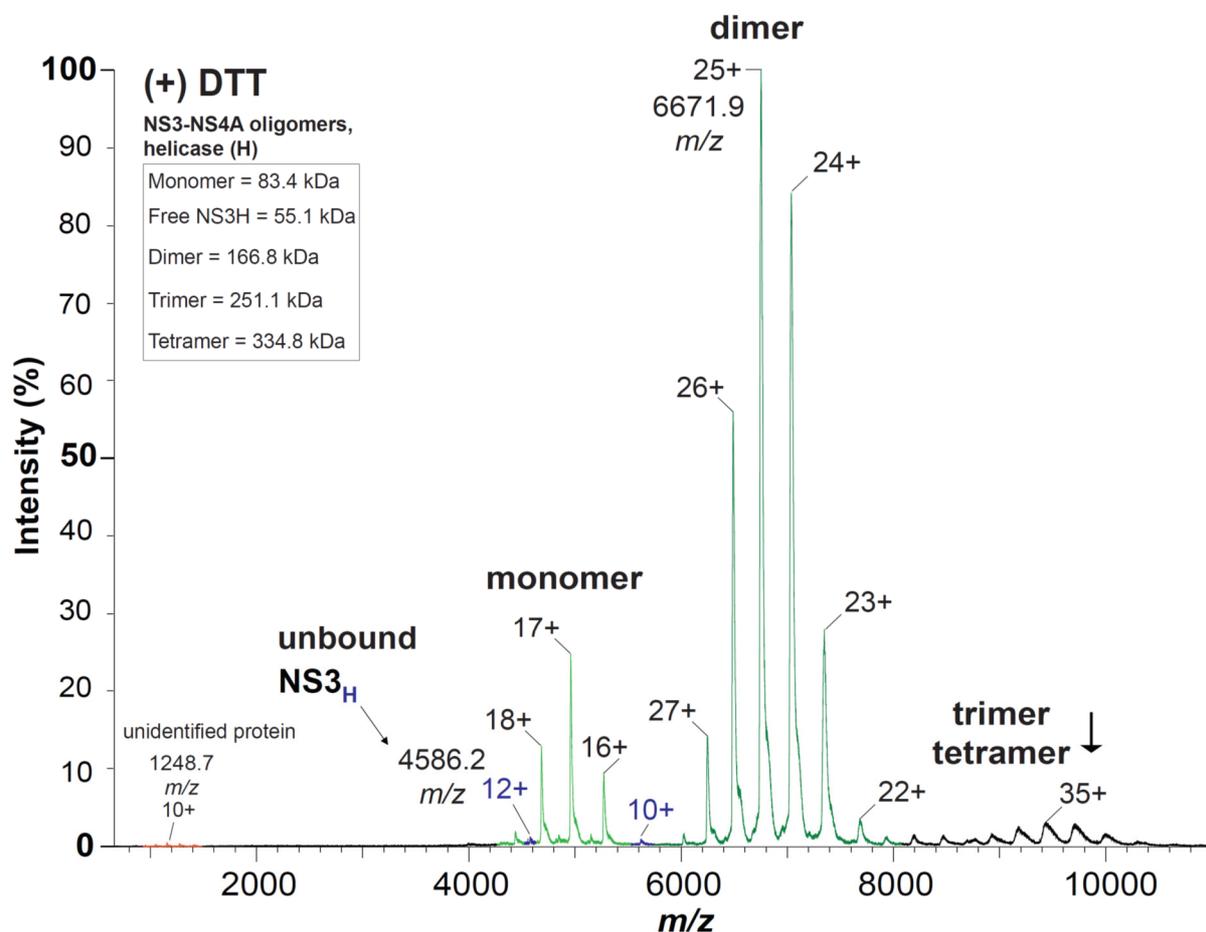


Fig. 26 Oligomerization of the wild-type CSFV NS3-NS4A fusion protein in the presence of low concentration of DTT.

Native ESI-MS measurements were performed in MS1 positive ion mode at 10 eV of collisional energy, in 250 mM ammonium acetate solution with 0.5 mM DTT in the MS-sample buffer. Protein concentration was 10 μ M (monomer). Monomer = light-green, dimer = dark-green, unbound NS3-helicase = blue, low intense oligomers = black, unidentified protein = orange.

In the following, to identify potential locations of the cysteines, the protein crystal structure was analysed based on the published PDB: 5MZ4.

In Fig. 27, the x-ray structure of the CSFV protein complex is shown to contain eleven cysteines. Eight of these cysteine residues are located within the protease part of the protein and three of them are incorporated in the helicase part. A closer look at the serine protease subunit of the enzyme shows that three of the cysteines are located close to the active-site

residues including His69, Asp97, S163(A) [102, 103]. Cysteines offer the possibility of forming *in vivo* and *in vitro* covalent bonds within the tertiary and quaternary structure of proteins. The bonds can occur intermolecular and lead to increased protein aggregation, which could also have an important role for proteins functions [194, 195]. The Cys-residues in the helicase subunit are not as exposed as those of the protease domain. Based on this it is assumed that the trimer and tetramer identified with native MS are formed via Cys-bridges of the protease domain (Fig. 24, Fig. 27).

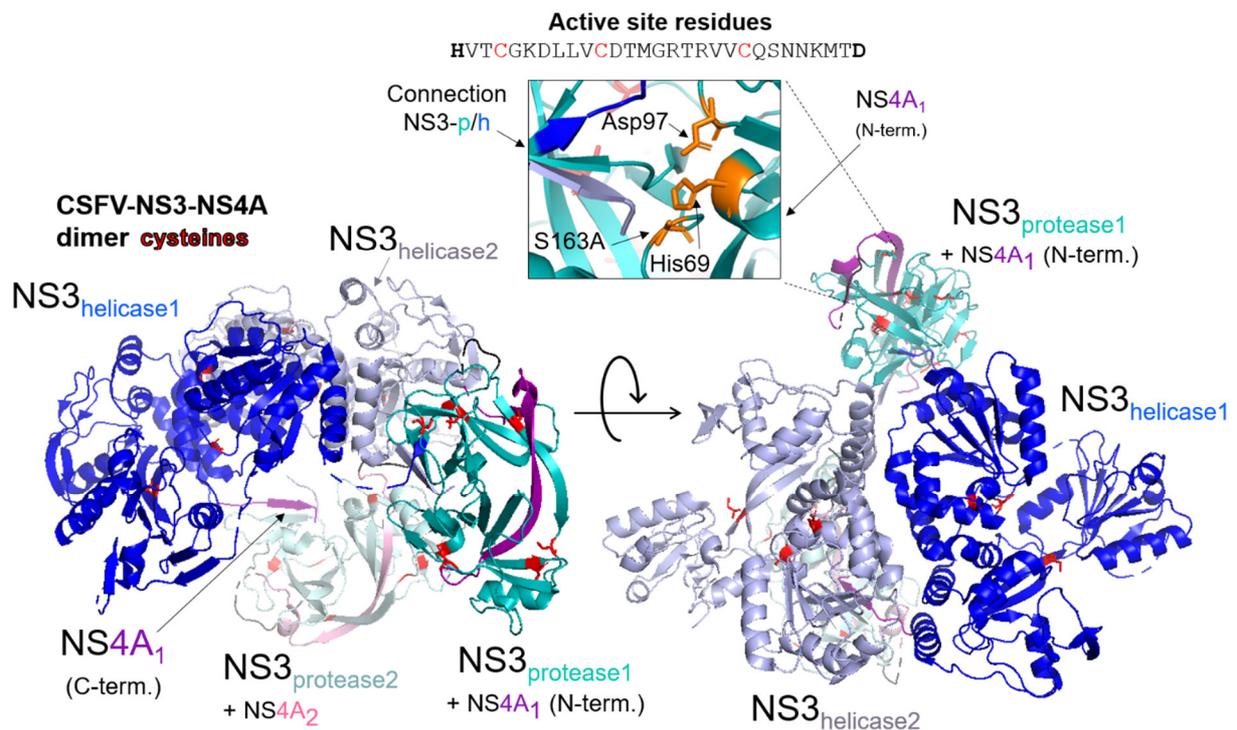


Fig. 27 Crystal structure of the dimeric CSFV NS3-NS4 protease.

Shown are the subunits (N-terminal NS3-protease and C-terminal NS3-helicase with NS4A cofactor) and the cysteines (red coloured) within the structure. The side chains of the active site aa Asp97 + His69 + S163(A) are shown as orange-coloured sticks. (Fig. adjusted according to PDB accession code: 5MZ4, [103]).

ESI-MS enables the investigation of native proteins under physiological conditions. Although, working with purified proteins under laboratory conditions does not mimic *in vivo* conditions, it cannot be excluded that these Cys-bridges observed by native MS occur in nature and form functional higher-ordered oligomers. In nature, the NS3 protein works within the cytosol along the ER membrane. In general, oligomerization of proteins can take place in the ER and in the cytosol. In the case of homo-oligomers, oligomer formation usually takes place from natively folded monomers in the ER and can continue in the Golgi.

In the cytosol, the protein assembly process begins in the ribosome [196, 197]. Thus, the environment of the NS3 protein offers the possibility to form *in vivo* Cys-interactions of neighbouring proteins. Thus, various studies indicate that monomers and dimers are the active forms of the NS3 protein [171, 198, 199]. Since the formation of the higher-order oligomers in the native MS studies, in contrast to the protein dimers, can be influenced from outside, harsher conditions are chosen in the further course which should eliminate higher oligomers [152, 200]. Lower protein concentrations (5 to 8 μM) and higher levels of DTT as an appropriate candidate for reducing this protein was used. Based on this, the “authentic” oligomers (monomer, dimer) are initially retained and can be compared and examined more closely in the following sections.

3.3 Auto-cleavage and proteolytic degradation of the NS3 protein monitored by native MS

In the following, based on the protein monomer level the monitoring of the autolytic cleavage process of the CSFV NS3-NS4A wild-type (WT) protein is in the focus. Various studies indicate that the NS3 protein of the *Flaviviridae* family has an auto-catalytic cleavage activity, identifying several internal cleavage sites [55, 85, 89, 92]. For the CSFV and HCV viruses, this process is also mediated by the NS3-serine protease. In association with the NS4A cofactor, the NS3-protease cleaves the viral polyprotein [55, 88, 89]. This leads to the release of further important proteins of the viral life-cycle of CSFV and HCV, which describes the importance of this protease for the viruses. Without the release of the other proteins from the polyprotein, the viruses are not able to replicate and survive. The protease of CSFV is known to cleave between Leu-residues and Ser, Ala or Asn [55, 89, 201]. Based on the internal NS3-protease cleavage sites (Leu159/Lys160, Leu192/Met193), the link between protease and helicase seems to split based on the NS3-protease activity (Fig. 28).

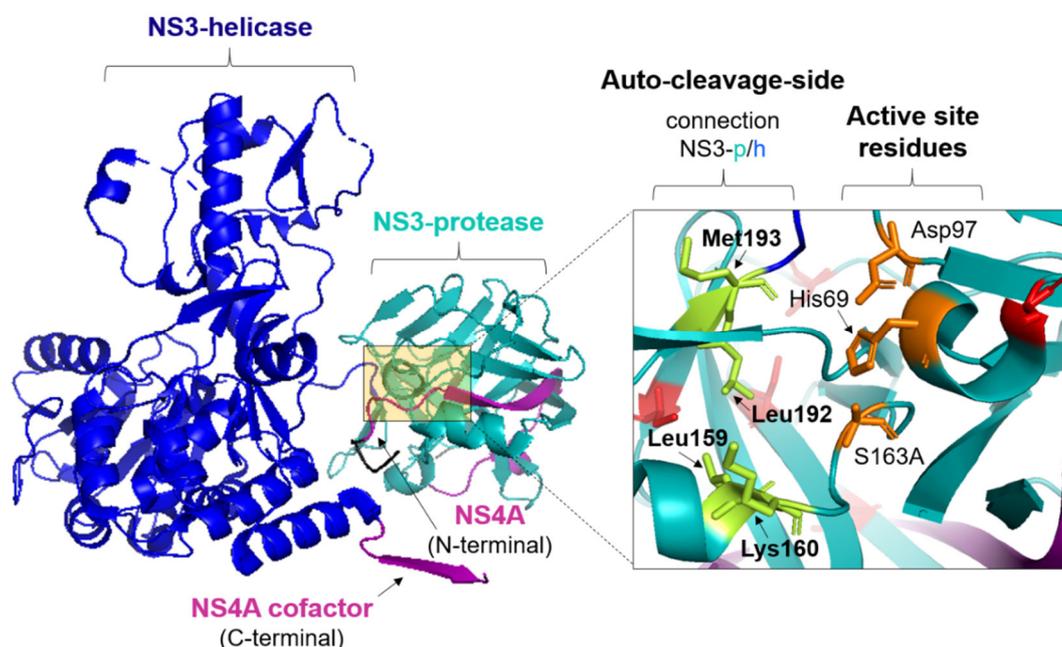


Fig. 28 Crystal structure of the monomeric NS3-NS4A protein complex shows the internal cleavage sides of NS3 (Leu159/Lys160 and Leu192/Met193).

Structure of the full-length protein (based on PDB: 5ZM4, [103]). Subunits are labelled with different colours: NS3-helicase (blue), NS3-protease (cyan), NS4A cofactor (purple). The zoom shows the side chains of two internal cleavage sites (Leu159/Lys160, Leu192/Met193, green-coloured sticks), the active site residues Asp97 + His69 + S163(A) (orange-coloured sticks) and cysteines (red-coloured sticks).

Both cleavage sites were discovered by Lamp and colleagues in 2013 [92], by their investigations of the auto-catalytic cleavage within the NS3 protein of CSFV. The cleavage results in the release of a complete free NS3-helicase, whereby the biological function of this mechanism is still unclear [92, 202]. These two cleavage sites were designated, by Zheng *et al.* 2017, as so-called minor auto-cleavage sites [202]. Moreover, it was shown, by Lamp *et al.* 2013 that the cleavage at Leu192/Met193 lead to the release of a fully active protease that is no longer fused with the helicase domain [92]. It is assumed that *in vivo* this process is necessary for both the NS3-protease and helicase to achieve even higher enzymatic activity showing a synergistic effect.

Based on the results of Lamp *et al.* in 2013 the question arose whether the cleavage between NS3-protease and NS3-helicase can also be followed with native MS and whether a proteolytic activity of the protease subunit can be observed in this context. Therefore, this chapter aims to investigate whether native MS offers another way of confirming the results of Lamp and colleagues. For the MS-based studies, a protein stock solution was stored at + 4 °C. Samples were taken at four different time points and analyzed using native MS and Top-Down MS. Based on the four time points, the individually removed protein solutions were designated as batches 1 to 4 (B1-B4), which are shown in Tab. 9.

Tab. 9 Four protein batches of the CSFV_{WT} NS3-NS4A protein reveal the auto-catalytic cleavage.

The identified monomeric and most dominant products (marked by lowercase letters (a) and (b)) of the auto-catalytic cleavage process at a certain time point of investigation are shown in column „Identified cleavage products“ and named as batches 1 to 4 (B1-B4).

Protein batch	Storage conditions	Identified cleavage products (kDa)	
B1	1 day on ice (+ 4 °C)	B1(a)	83.374
		B2(a)	82.683
B2	1 week fridge (+ 4 °C)	B2/3(b)	80.779
		B3/4(a)	55.061
B3	2 weeks fridge (+ 4 °C)	B2/3(b)	80.779
		B3/4(a)	55.061
B4	3 weeks fridge (+ 4 °C)	B3/4(a)	55.061

After purification, the protein sample was flash frozen by liquid nitrogen within the storage-buffer (50 mM HEPES, 10% Glycerol pH7.5, 4 mM DTT) and stored at - 80 °C. The sample was carefully thawed on ice at + 4 °C before being prepared for native MS measurements.

To achieve a complete reduction of the protein cysteine bridges, the protein stock solution, including B1, B2, B3 and B4 consists of 200 mM ammonium acetate solution with 2 mM DTT. B1 was stored on ice for the time of MS measurements. B2, B3 and B4 were stored in the fridge at + 4 °C for different periods of time (Tab. 2).

MS1 spectra were recorded at 10 eV and 50 eV. Under the conditions used, the results obtained clearly show that there exist structurally different proteins over the course of three weeks in the four batches (Fig. 29 (A)-(D), Fig. 30), indicating an auto-proteolytic activity of the protein complex. Of particular note, the cleavage of the NS3 protein complex resulted in the release of a monomeric NS3-helicase (Fig. 29 (D)).

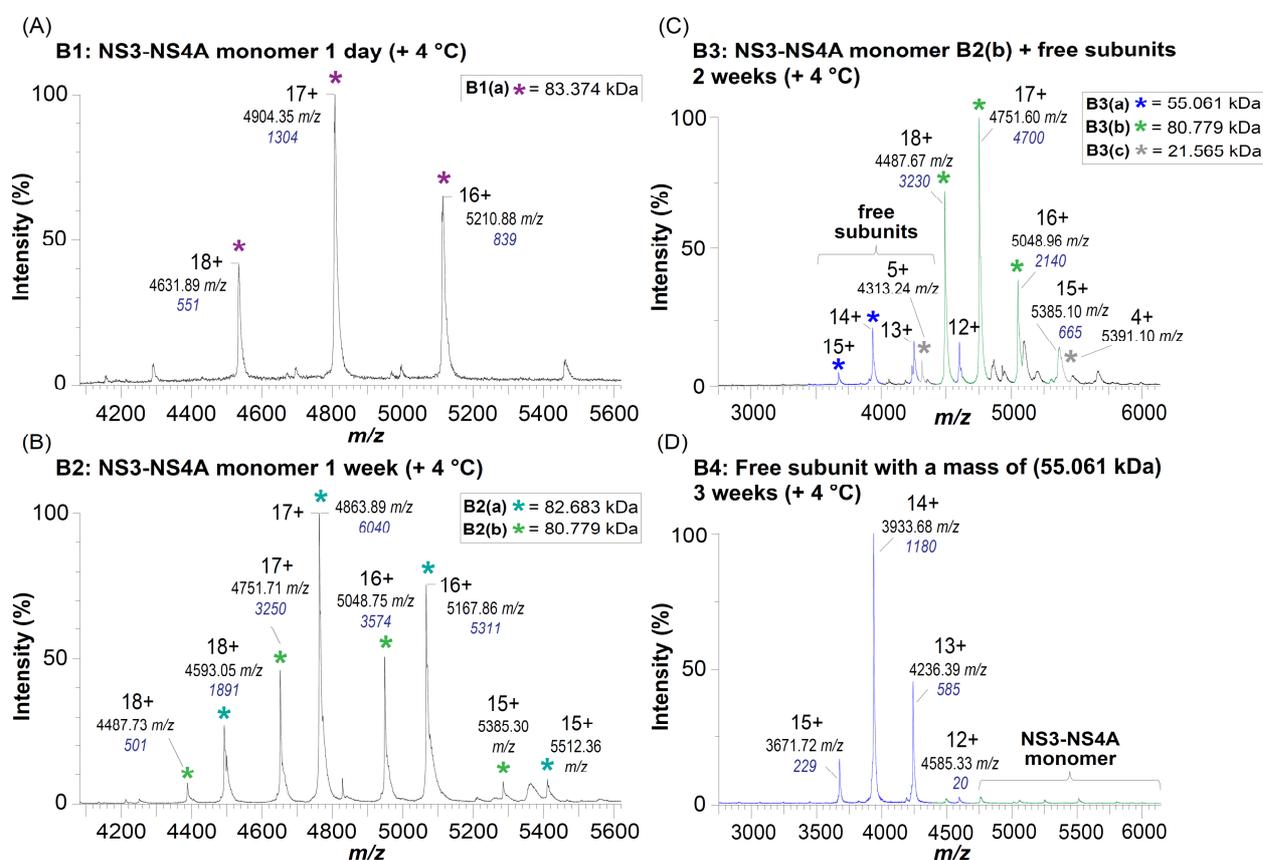


Fig. 29 The NS3-NS4A protein of CSFV - overview of protein degradation within the four different batches.

(A) Batch 1: MS1 identification of the NS3-NS4A monomer species B1(a), protein monomer = 83.374 kDa, (protein storage for 1 day on ice (+ 4 °C) while measuring). (B) Batch 2: MS1 spectrum of two identified monomers B2(a) = 82.683 kDa and B2(b) = 80.779 kDa, (protein storage for 1 week at + 4 °C before measuring). (C) Batch 3: MS1 identification of two presumably unbound subunits B3(a) = 55.061 kDa, B3(c) = 21.566 kDa and the B3(b) monomer, (protein storage for 2 weeks at + 4 °C). (D) Batch 4: spectrum of the unbound protein subunit = 55.061 kDa, (protein storage for 3 weeks at + 4 °C). Peak intensity labelled by blue italic numbers. MS1 energy was set to 10 eV and 50 eV, ammonium acetate 200 mM at pH7.2, 2 mM DTT, protein concentration was 5 to 8 μ M.

B1 shows a main peak and charge distribution (18+ to 16+) that can be assigned to the monomer B1(a) with a mass of 83.374 kDa. In contrast, after one week of auto-proteolysis at + 4 °C B2 clearly shows two main different peak series in MS1 spectra. Both proteins showed a charge state distribution of 18+ to 15+. Thus, due to the m/z values both proteins could be identified as monomers, which were named as B2(a) (82.683 kDa) and B2(b) (80.779 kDa).

Additionally, after two weeks of proteolysis B3 also shows three dominant peak series. Based on their MS1 masses and in line with the aa-sequence B3(a) and B3(c) seemed to be two released subunits of the NS3-NS4A complex, the NS3-helicase (55.061 kDa) and the NS3-protease (21.565 kDa) tied to several N-terminal amino acids of the NS4A cofactor (K₅₃TEGK-GSGS = 0.849 kDa). B3(b) could be the same monomer already identified as B2(b).

Moreover, after three weeks of auto-proteolysis (Fig. 29 (D), B4, Fig. 30 (A)/(B)) all monomers as well as the released NS3-protease seemed to be degraded. Evidently, the degradation process of the NS-protein is dependent on time and storage conditions (Tab. 9). An overview of the crystal structures of the so far identified protein species based on sequence information and MS1 data is shown in Fig. 30.

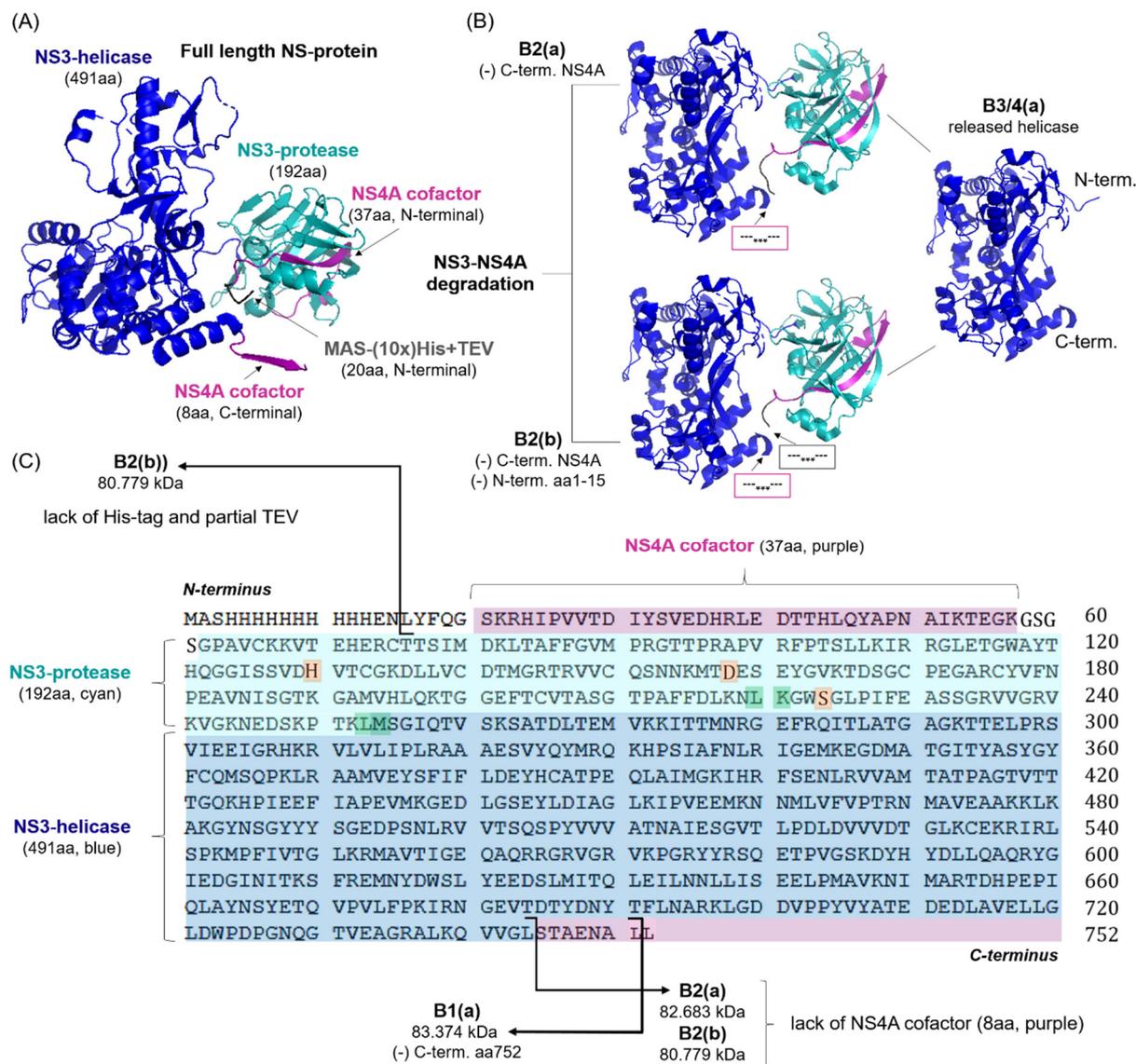


Fig. 30 Overview of the NS3-NS4A protein species identified with native MS as of this point.

(A) Crystal structure of the full-length protein (based on PDB: 5ZM4, [103]). Subunits are labelled with different colours: NS3-helicase (blue), NS3-protease (cyan), NS4A cofactor (purple). (B) Products of the NS3 auto-proteolysis process: monomers B2(a) and B2(b) (lack of aa is labelled by three black stars), released NS3-helicase B3/4(a). (C) Protein aa-sequence of the full-length NS-protein, including tags. Subunits are labelled with similar colours as for the crystal structures. Lack of aa for the identified protein species are marked with black arrows. Location of side chains of two internal cleavage sites (Leu159/Lys160, Leu192/Met193, green coloured) and the position of the active side residues Asp97 + His69 + S163(A) (orange coloured).

The co-existence of the different protein cleavage products needs to be analyzed in more detail by MS2, which is shown in the following sections. B3 and B4 shows the most noticeable changes in the MS1 spectra compared to B1 and B2. Therefore, B4 which was stored for the longest at + 4 °C, to monitor the course of the auto-proteolytic activity, is used as the starting point for further explanations by using Top-Down-sequencing.

3.3.1 Identification of the NS3-helicase (B3/4(a)) by Top-Down-Sequencing

The lack of a pure MS1 monomer signal in B3 and even more in B4 shows that degradation of the NS3-NS4A protein took place. The observed peaks with lower charges can be assigned to a protein (B3/4(a)) of approx. 55 kDa. As can be assumed based on the MS1 data, this small protein could be the NS3-helicase subunit of the full-length protein. In order to ensure that this was not contaminant, a more detailed analysis of the protein was carried out using the MS2 mode and the protein aa-sequence of the NS3-helicase (Fig. 31).

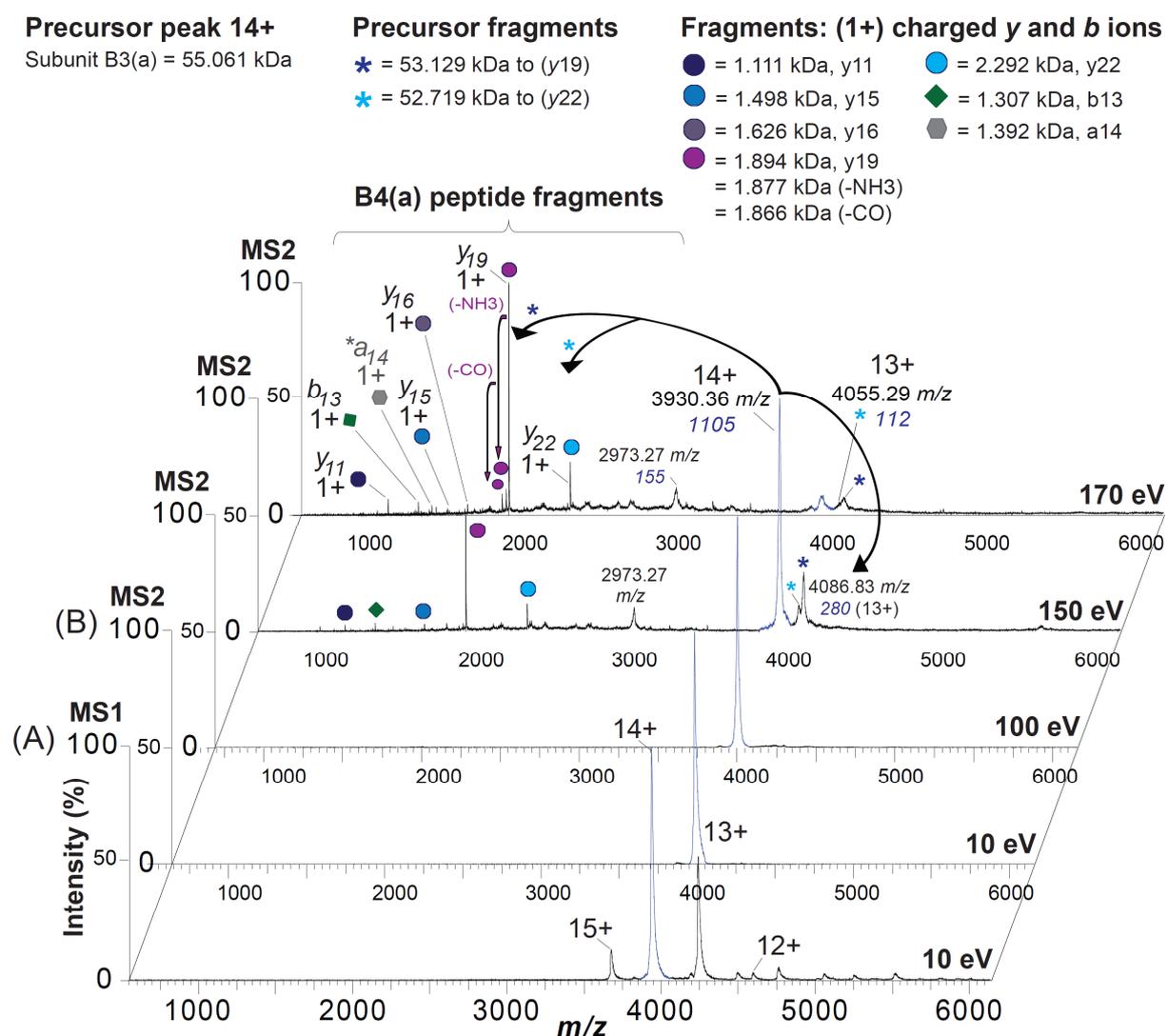


Fig. 31 B4(a) subunit dissociation with MS2 CID.

(A) MS1 spectrum of B4(a) at 10 eV energy. (B) MS2 spectra of the 14+ precursor peak (blue, 3933.68 m/z) with increased collision energy (10 eV to 170 eV). Identified peptide ions are marked with different coloured balls and were identified by their monoisotopic masses in association with the protein aa-sequence. Identified protein fragments are marked with different coloured stars. Peak intensity is shown by blue italic numbers. Conditions: 200 mM ammonium acetate + 2 mM DTT pH 7.2, protein concentration 5 μ M.

The proteomic tool MS/MS fragment ion calculator from the ISB database SSL server (<http://db.systemsbiology.net:8080/proteomicsToolkit/index.html>) was used for the identification of the MS2 fragment ions. MS2 measurements were performed based on the most intense MS1 peak (at 3934 m/z) with a charge of 14+ (blue peak). This precursor ion was dissociated with increasing collision energy by means of CID (10 eV to 170 eV). The spectra show a well defined fragment ion pattern (Fig. 31, coloured balls). Based on the numerous C- and N-terminal fragment ions (y - and b -ions) as well as the parent-ion mass, B3/4(a) can be clearly assigned to the aa-sequence of the NS3-helicase subunit. The obtained product-ions are mainly based on CID energy levels between 150 eV and 170 eV. The fragments y_{19} (purple ball) and y_{22} (light-blue ball), which occur at 150 eV with high intensity, are particularly noteworthy for the clear identification of the NS3-helicase (Fig. 32, Tab. 10).

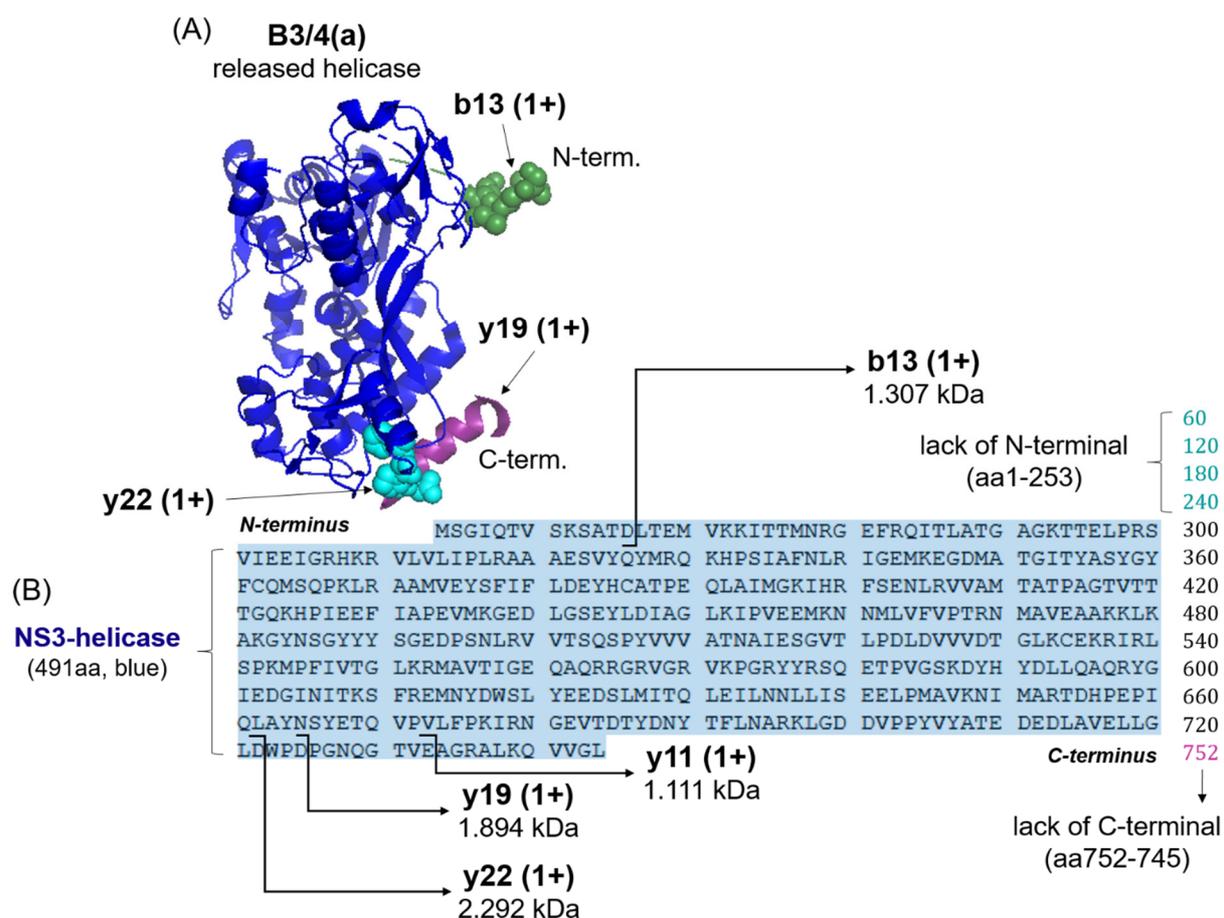


Fig. 32 Amino acid sequence and associated crystal structure of the released NS3-helicase subunit show the low mass MS2 ions.

Helicase sequence and the related crystal structure (based on PDB: 5ZM4, [103]) demonstrates the lack of N-terminal aa1-253 and C-terminal aa752-745. (A) Main intense MS2 y -fragment ions and the specific N-terminal b -ion are labelled in the crystal structure as y_{19} (deep purple, α -helix), y_{22} (light-blue, sphere), b_{13} (green, sphere). (B) In the aa-sequence specific C- and N-terminal ions are labelled by black arrows.

Furthermore, these fragments can also be assigned to the two cleaved protein fragments (53.129 kDa and 52.719 kDa) with a charge of 13+. Consequently, the NS3 protein showed auto-proteolytic activity at + 4 °C, leading to the release of the helicase subunit. This subunit was not further degraded during the observed period of time. This could indicate that the helicase achieves its actual activity through the auto-proteolysis of the NS3-protease and through its release into the environment.

Tab. 10 Main MS2 ions of the 14+ precursor signal of the NS3-helicase at 195 eV.

hProF = high mass protein fragments, lProF = low mass protein fragments, z = charge. MW is based on the protein monoisotopic mass [M], TP = theoretical peptide, aa = amino acid.

hProF (kDa)	z	lProF m/z	TP m/z	z	Ion	aa-sequence (N- to C-term.)	Crystal structure
53.129	13+	1111.87	1111.69	1+	y11	AGRALKQVVGL	α -helix
		1894.23	1894.05	1+	y19	PGNQGTVEAGRA LKQVVGL	α -helix
		1877.11	-	1+	y19 (-NH ₃)	-	
		1866.17	-	1+	y19 (-CO)	-	
52.719	13+	2292.43	2292.21	1+	y22	WPDPGNQGTVE AGRALKQVVGL	(spheres)
		1307.04	1306.63	1+	b13	MSGIQTVSKSATD	(spheres)

3.3.2 Top-Down-Sequencing identifies different NS3-NS4A monomers

Furthermore, the question arises as to how the identified monomers (B2(a), B2(b) and B3(b)) after one and two weeks of auto-proteolysis differ. Structural modifications should be identified by the comparison of C- and N-terminal fragment ions based on the known helicase subunit fragments of B4. Therefore, MS2 analyses with the monomers were performed at different CID energies (Fig. 33, Fig. 34).

Based on the MS1 data (Fig. 29 (A)-(D)), it was assumed that the monomeric species B3(b) is the same as B2(b). To confirm this, the main peak (17+) of B3(b) was selected as precursor and fragmented in MS2 mode. The fragment ions obtained were then compared with those of the helicase (B4) and with those of the monomer B2(b). With an applied dissociation energy of 170 eV it can be clearly seen that the two fragment ions (y19 and y22 (1+)) are identical to the dissociation of the NS3-helicase (Fig. 33, Fig. 32, Fig. 31, Tab. 10). Moreover, there was another fragment ion observed (b20 (1+), 2333 m/z), which can be assigned to the N-terminal end of the B3(b) monomer (Tab. 11, 3.3.3).

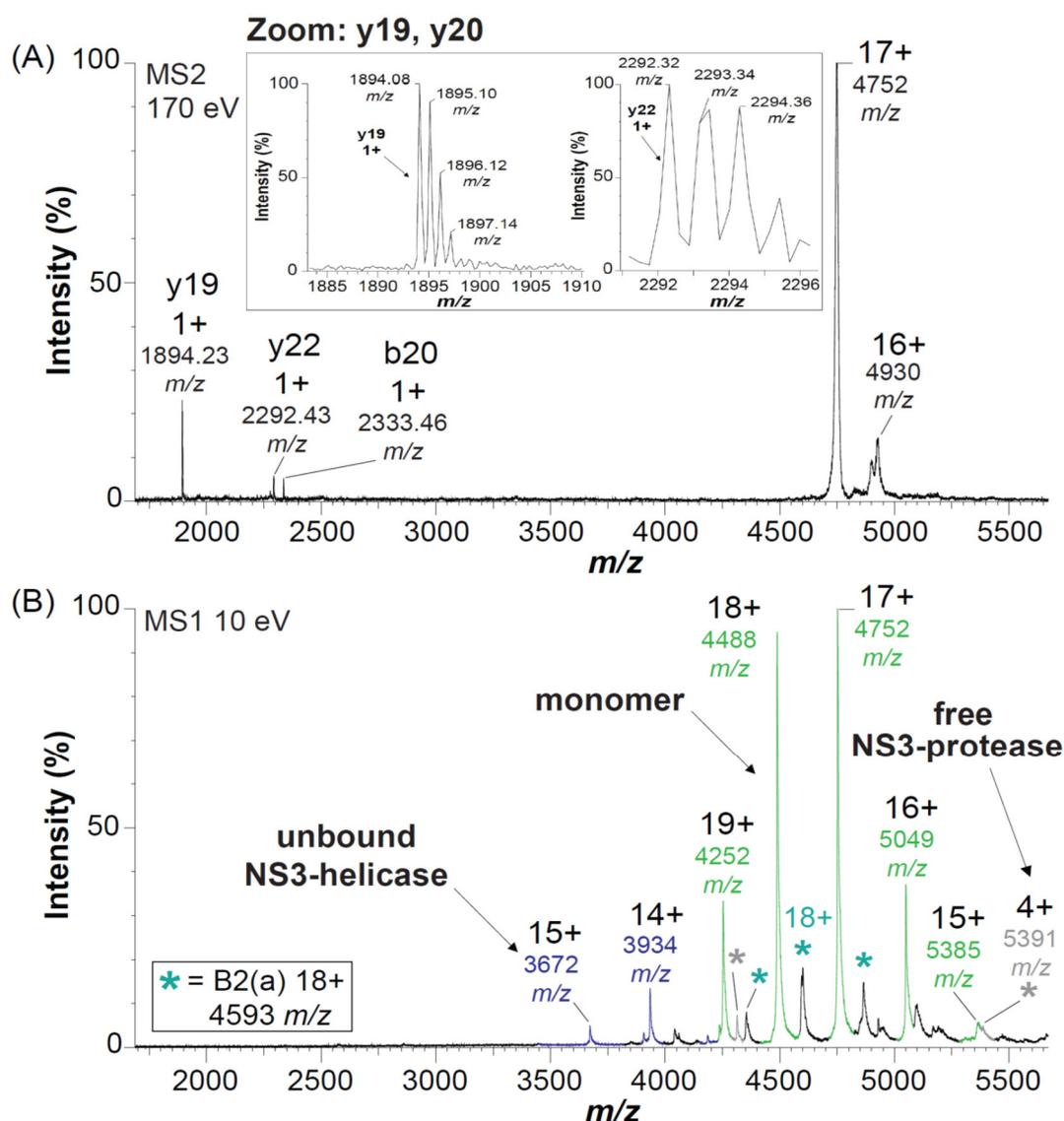


Fig. 33 B3(b) dissociation revealed the same C-terminal peptide ions compared to B4(a). (A) MS2 spectrum of the 17+ precursor of B3(b) at 170 eV. (B) MS1 spectrum of B3(a) NS3-helicase, blue line, B3(c) NS3-protease, grey line and B3(b) monomer, green line at 10 eV. Cyan coloured stars label the B2(a) monomer. Conditions: 2 weeks at + 4 °C, 200 mM ammonium acetate + 2 mM DTT pH 7.2, protein concentration 8 μ M.

In addition, the MS1 spectrum shows the released helicase subunit (blue line) and a smaller amount of suspected NS3-protease B3(c) (grey lines, star) as well as a series of peaks (labelled by cyan stars) that could later be assigned to the B2(a) monomer. Obviously, the signal intensity of these monomer species has decreased significantly in batch 3 after two weeks of auto-proteolysis. In contrast the intensity of unbound helicase has increased considerably. Therefore, it can be assumed that the degradation of the monomer B2(a) triggered the release of the NS3-helicase subunit.

The obtained fragment ion pattern of B3(b) was further compared with the two monomers of batch 2 (B2 (a), B2 (b)) and as a control with B4. In each case, an MS1 spectrum and the associated MS2 spectrum at 150 eV were recorded (Fig. 34).

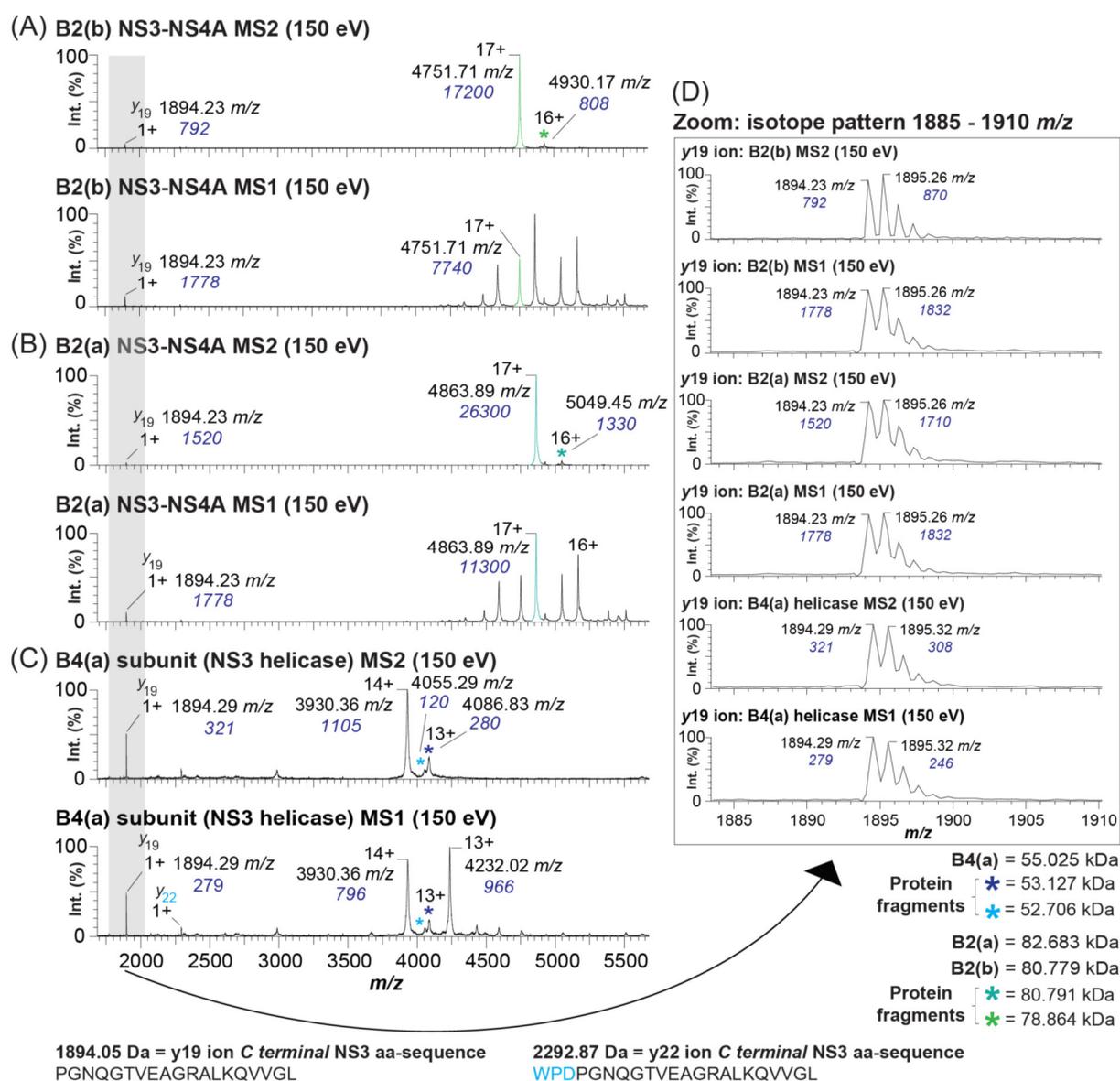


Fig. 34 Identification of the C-terminal y19 low mass ion for the monomers B2(a)/(b).

(A) MS1 and MS2 spectra at 150 eV show B2(b) and the associated fragment y19 ion. (B) MS1 and MS2 spectra of the B2(a) monomer at 150 eV with the associated y19 fragment. (C) MS1 and MS2 control spectra of the helicase subunit B4 at 150 eV. (D) Shows the isotope pattern of the y19 ion. Peak intensity labelled by blue italic numbers. Conditions: (B)/(A) 1 week at + 4 °C, 200 mM ammonium acetate + 2 mM DTT pH 7.2, protein concentration 8 μM. (C) 3 weeks at + 4 °C, 200 mM ammonium acetate + 2 mM DTT pH 7.2, protein concentration 6 μM.

Surprisingly, the y19 and y22 fragments, that clearly identified the NS3-helicase subunit also appeared when dissociating the B2(a) and B2(b) monomers. Together with the monomer

B3(b) and in line with the helicase protein sequence these results lead to the assumption that a certain sequence segment is missing at the C-terminal end of the identified monomers.

This is also confirmed by a comparison with the full length aa-sequence of the NS3-NS4A fusion protein (Fig. 38). Since the helicase aa-sequence has no NS4A cofactor at the C-terminus, the two characteristic y -ions are created here. Consequently, it can be assumed that the C-terminal section of the cofactor (aa752-745 = 8aa = STAENALL, 817 Da) is also missing at the B2(a), B2(b) and B3(b) monomers. A peptide that fits to that mass was identified in MS1 mode at 50 eV (see supplement 6.4.3, S. Fig. 10). Indicating that during the auto-cleavage of the NS3 protein the cofactor is apparently initially split off. The protein mass after this C-terminal loss fits the measured monomer mass of B2(a).

In order to examine the changes and differences at the N-terminal end of the monomers more closely, further MS2 analyses were carried out. The fragment ions obtained were then compared with the aa-sequence of the full length NS3-NS4A protein. Thus, the associated sequences for the individual monomers could be reconstructed.

First, MS2 analyses were performed with the B2(a) monomer (missing aa752-745 at the C-terminus). For a detailed investigation of the protein, the main signal (17+) was selected as precursor peak and dissociated by means of CID. The energy was gradually increased from 10 eV to 195 eV. Various peptide ions (Fig. 35, Tab. 11, 3.3.3) as well as the fragmented N-terminal part of the NS4A cofactor (37aa, 4.2 kDa) originating from the parent ion were detected (supplement chapter 6.3, S. Fig. 3). This indicates that the B2(a) protein is only missing the NS4A cofactor part at the C-terminal end, but has an intact sequence at the N-terminus.

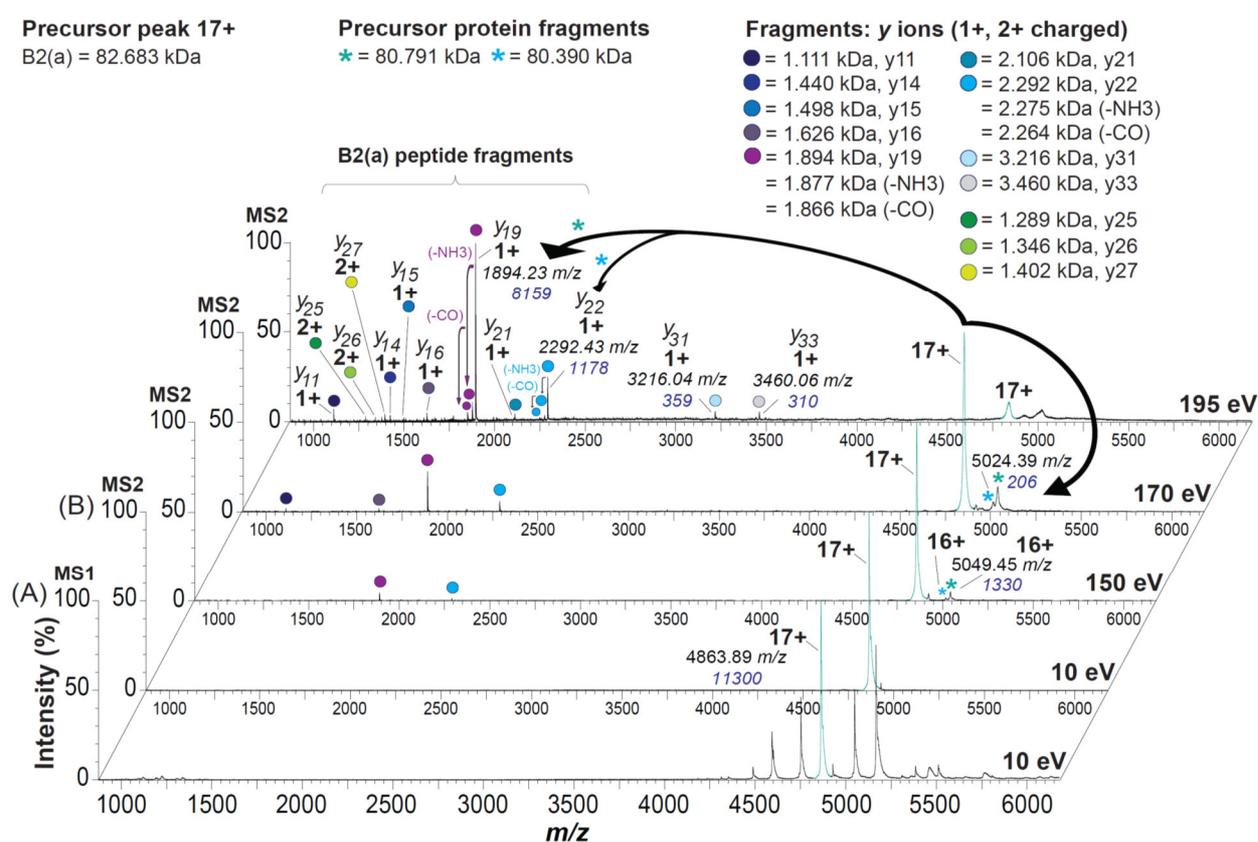


Fig. 35 B2(a) monomer CID fragmentation revealed a lack of the C-terminal NS4A cofactor.

(A) MS1 spectrum of B2(a) at 10 eV energy. (B) MS2 spectra of the 17+ precursor peak (cyan line) with increased collision energies (10 eV to 195 eV). Identified peptide ions are marked with different coloured balls and are based on the monoisotopic masses in association with the aa-sequence. Identified protein fragments are marked with different coloured stars. Peak intensity labelled by blue italic numbers. Conditions: 200 mM ammonium acetate + 2 mM DTT pH 7.2, protein concentration 8 μ M.

37 (A)/(B)). Further ions are the three doubly charged ions y19 (947.74 m/z (2+)), y22 (1146.86 m/z (2+)) and y23 (1204.35 m/z (2+)).

With a CID energy of 195 eV, the main ions formed are singly charged, including the main fragment ions y19 and y22 (Fig.37 (C)/(D)). The biggest difference between both monomers is the introduced b20 ion (Fig. 36, Fig. 37, Fig. 38).

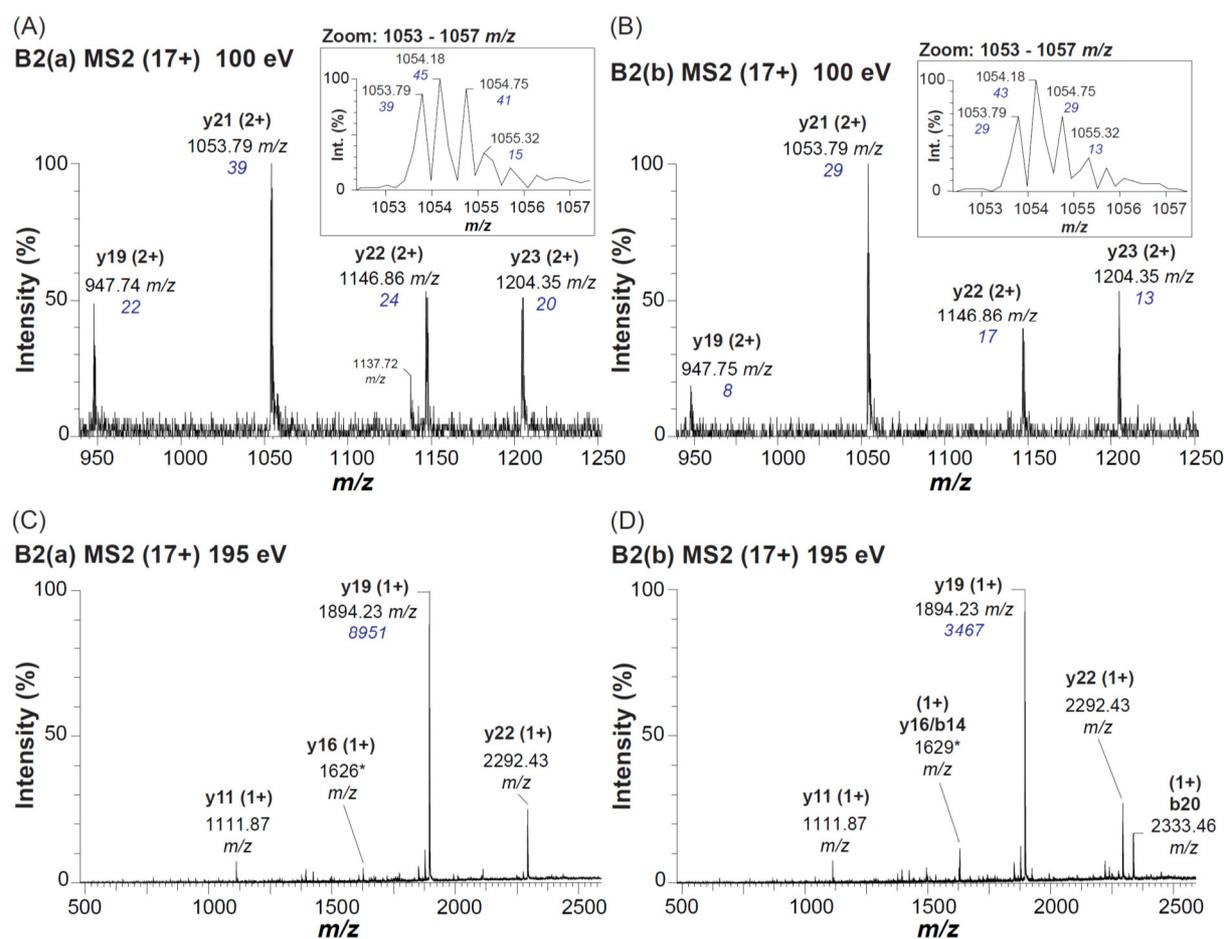


Fig. 37 First appearing MS2 low mass ions of the identified monomers B2(a) and B2(b) at 100 eV are doubly charged, main ions detected at 195 eV are singly charged.

(A) MS2 spectrum of the first intense y-ions of the 17+ precursor peak (B2(a)) at 100 eV. (B) MS2 spectrum of the first intense y-ions of the 17+ precursor peak (B2(b)) at 100 eV. (C) MS2 spectrum at 195 eV of the main y-ions based on 17+ precursor peak (B2(a)). (D) MS2 spectrum at 195 eV of the main y- and b-ions based on 17+ precursor peak (B2(b)). Identified ions are based on the monoisotopic masses in association with the aa-sequence. Peak intensity labelled by blue italic numbers. Conditions: 200 mM ammonium acetate + 2 mM DTT pH 7.2, protein concentration 8 μ M, collision energies 100 - 195 eV.

Since the same N- and C-terminal ions as for B2(b) (including singly charged y19, y22 and b20 ions) were identified in the dissociation of the B3(b) monomer, which was identified after two weeks of auto-proteolysis, it is obvious that both proteins are the same (B2(b) = B3(b), missing aa1-15 and aa752-745). So far, no cleavage site is known for the N-terminally

missing aa1-15. However, the observed cleavage took place between the two amino acids Asn (N) and Leu (L), which form an already identified cleavage motif for the NS3-protease of *Pestivirus*s [55, 89].

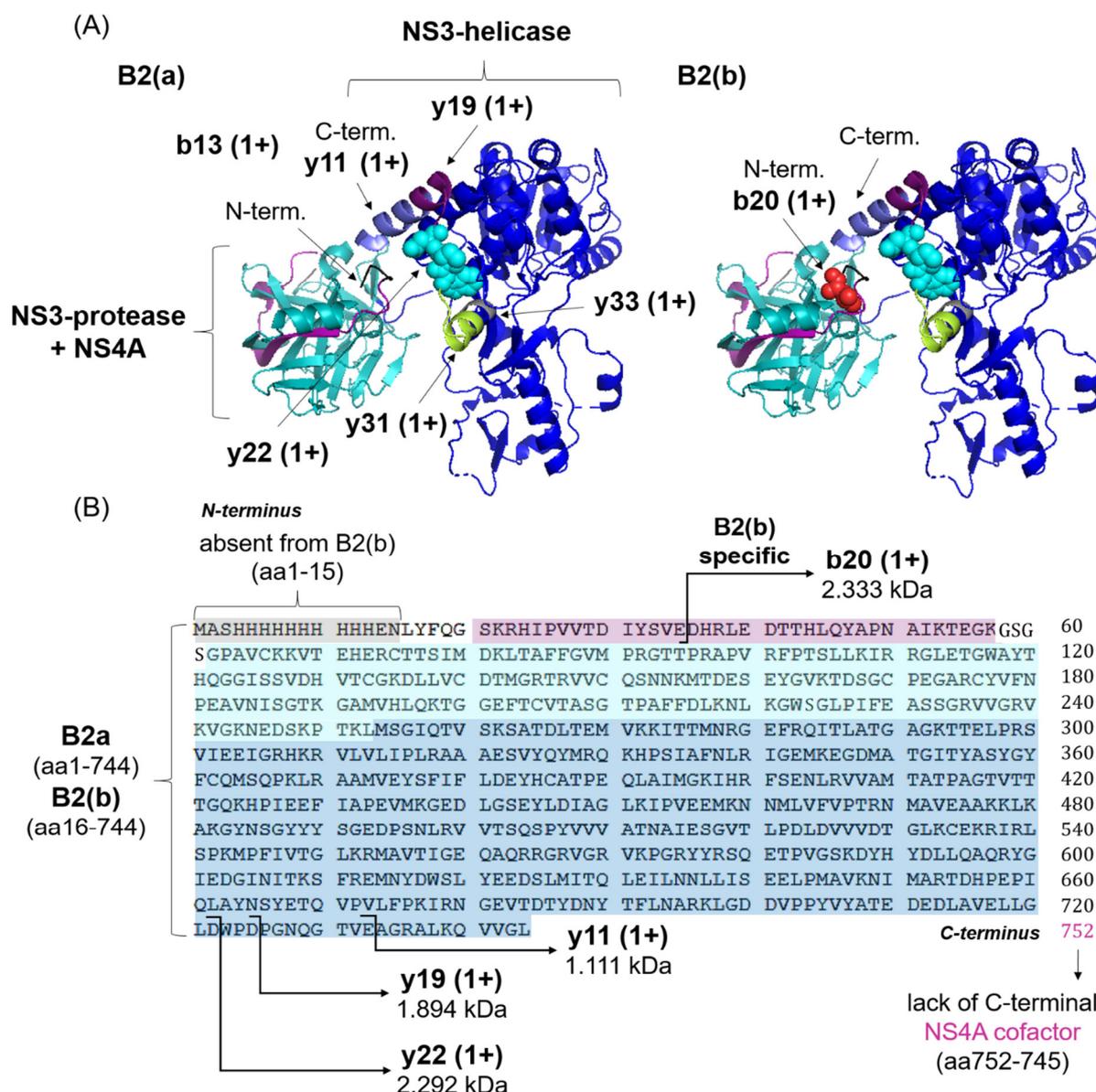


Fig. 38 Crystal structure and related amino acid sequence of the B2(a) and (b) monomers show the location of the main intense MS2 backbone ions.

Reduced aa-sequence and the related crystal structure (based on PDB: 5ZM4, [103]) demonstrates the lack of N-terminal aa1-15 and C-terminal aa752-745 for B2(b) and aa752-745 for B2(a). (A) B2(a) and (b) main intense MS2 y-fragment ions and the specific N-terminal b-ion are labelled in the crystal structure as y11 (violet, α -helix), y19 (deep purple, α -helix), y22 (light-blue, sphere), y31 (green, α -helix), y33 (grey, α -helix-loop) partial b20 (red, sphere). (B) In the aa-sequence specific C- and N-terminal ions are labelled by black arrows.

However, on closer examination of the MS2 data some fragment ions were found that could not be assigned to either of the three main monomer species. Therefore, the selected

17+ precursor peaks were examined in more detail in the next step. In particular, their shape and their m/z ranges were compared.

3.3.3 Native MS reveals the presence of several monomer sub-species in B2

In contrast to the batch 1 monomer a large number of sub-peaks were found in addition to the actual main monomers of batch 2 (Fig. 38 (A)-(B)). Based on their MS1 masses, the aa-sequence and the additionally found and previously unassigned MS2 ions, these peaks could also be identified as monomers of the NS3-NS4A protein. These sub-species were seen most clearly at an energy of 110 eV in MS1 mode. However, they were also identified at lower energies (Fig. 39, Fig. 40 (A), 10 eV). This clearly indicates that these are not artificial monomers that were created by collision with gas molecules in the ESI process.

Batch 2: NS3-NS4A monomer MS1 (10 eV)

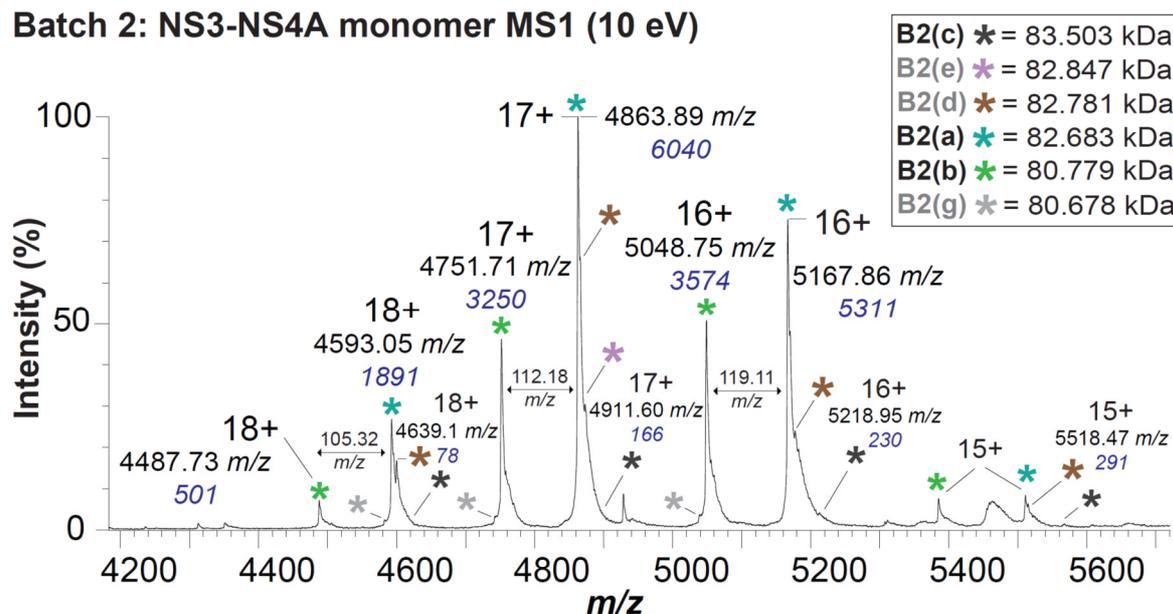


Fig. 39 MS1 analyses of the B2 NS3-NS4A protein identified the main and several sub-monomers at 10 eV.

Main monomers B2(a) = 82.683 kDa and B2(b) = 80.779 kDa, sub-monomers of B2(a) = B2(c)-(e), B2(b) sub-species = B2(g)-(j) (protein storage for 1 week at + 4 °C before measuring). The supposed specific protein cleavage products are written in black, the supposed non-specific cleavage products are written in grey. Peak intensity labelled by blue italic numbers. MS1 energy was set to 10 eV, ammonium acetate 200 mM at pH 7.2 + 2 mM DTT, protein concentration was 8 μ M.

Four further sub-species were identified, both under the peak of the B2(a) monomer as well as under the B2(b) monomer (Tab. 11). Monomer species B2(c)-(f) can be assigned to the main B2(a) peak and species B2(g)-(j) can be associated with B2(b) (Fig. 40). In addition

to the 17+ peak in MS1 the 18+ and 15+ peak also show a comparable distribution of the sub-peaks (Fig. 40 (C)/(D)). Based on the associated MS1 mass a peak series with low intensity was identified, which consists of the complete aa-sequence (Fig. 40 (D), B2(c)).

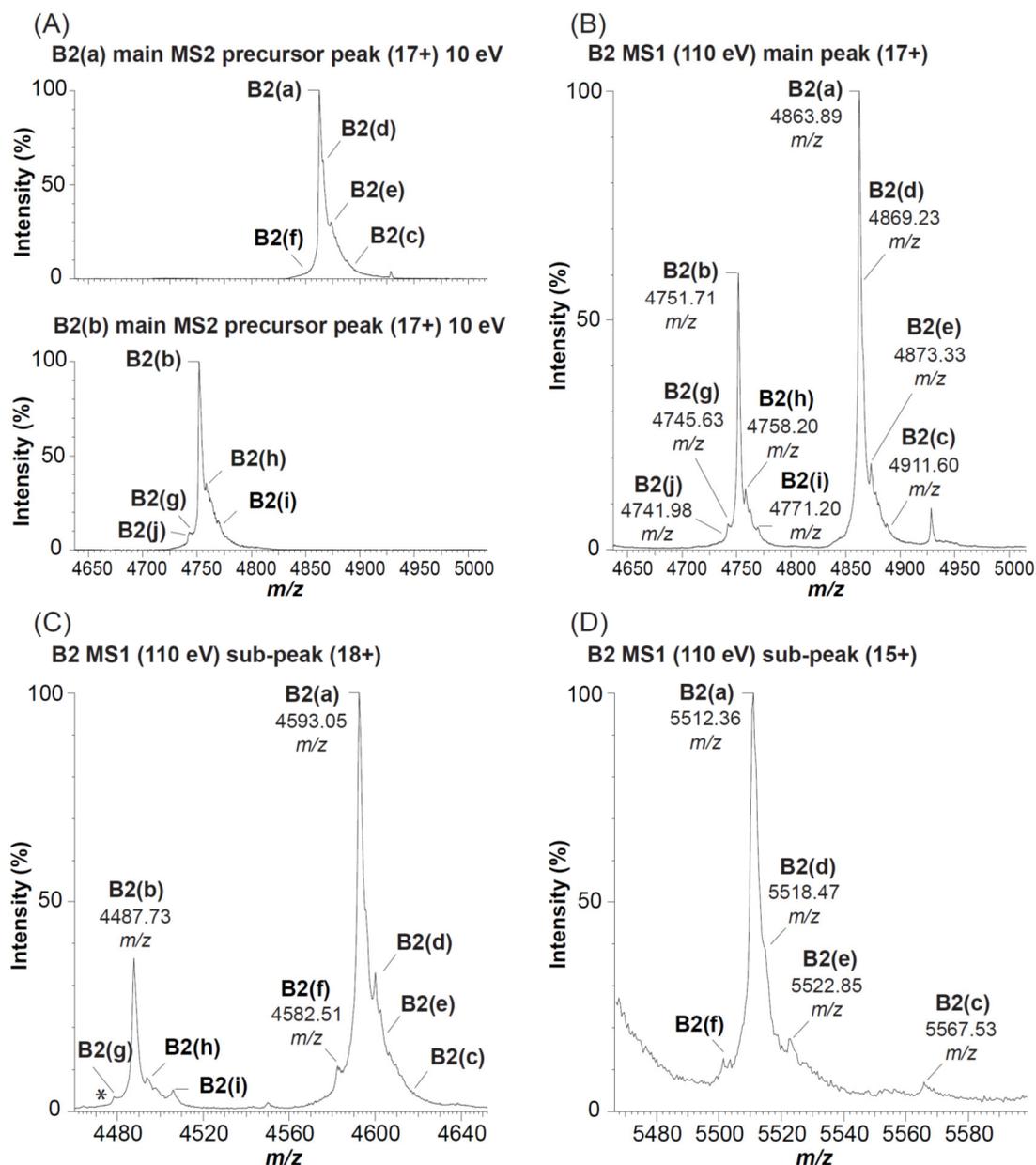


Fig. 40: Non-fragmented precursor in MS2 and MS1 peaks of B2(a) and B2(b) show an overlay of various NS3-NS4A monomeric peak species.

(A) MS2 spectrum of the 17+ precursor peak (above B2(a), below B2(b)). (B) MS1 spectrum of the 17+ peak (B2(a) and B2(b)). (C) MS1 spectrum of the 18+ peak (B2(a) and B2(b)). (D) MS1 spectrum of the 15+ peak (B2(a) and B2(b)). Identified monomer sub-species for B2(a) are labelled B2(c)-(f), B2(b) sub-species are labelled with B2(g)-(j). Conditions: 200 mM ammonium acetate + 2 mM DTT pH 7.2, protein concentration 8 μ M, collision energies 110 eV.

Based on their MS1 masses (Tab. 11) and in comparison with the reconstructed B2(a) and B2(b) aa-sequences (Fig. 38 (B)) specific MS2 ions could be assigned to the sub-monomers.

These ions are mainly singly charged and of significantly lower intensity (Fig. 41). This can be explained by the fact that the associated 17+ precursor peaks are also of lower intensity than those of the two main monomers (B2(a)/(b)).

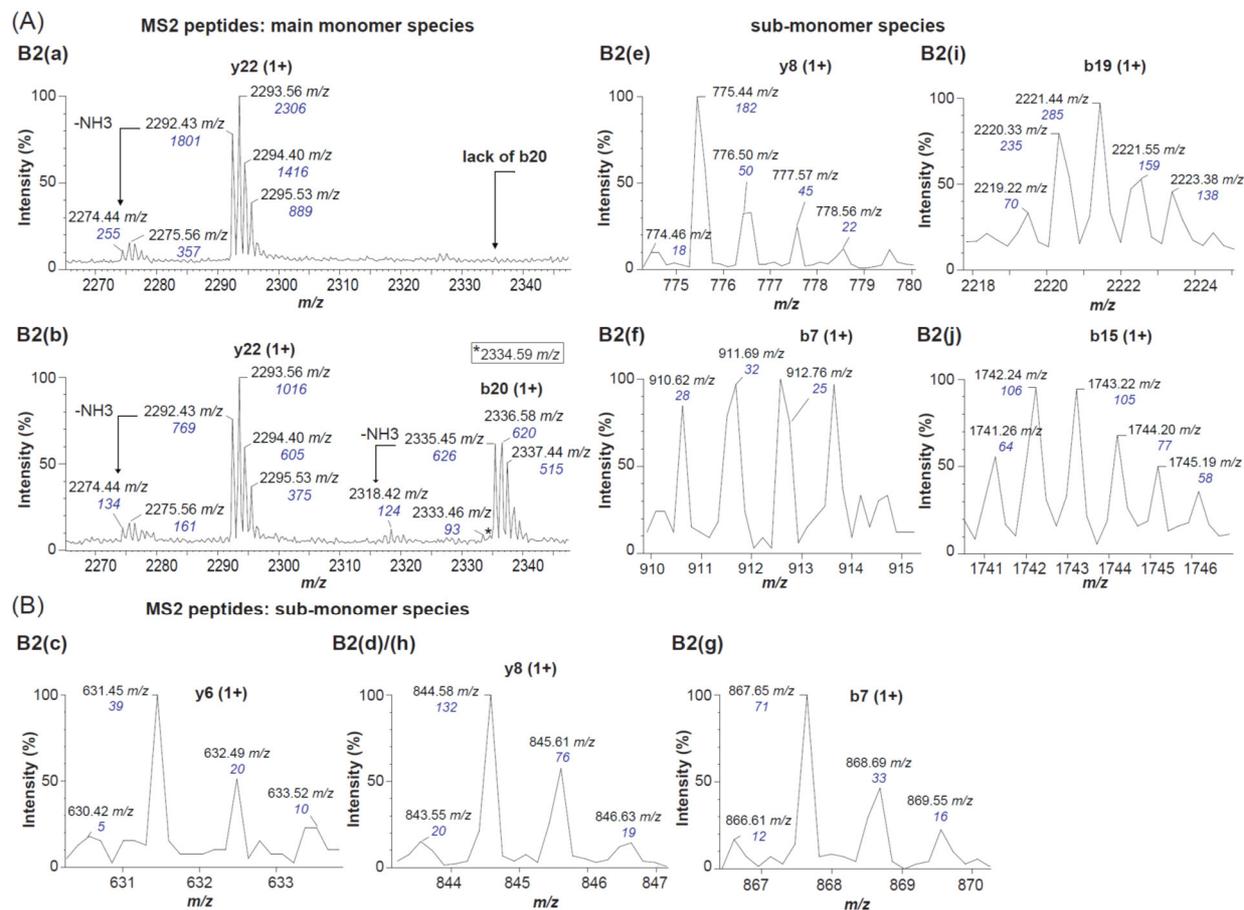


Fig. 41 Specific MS2 fragment ions of all identified B2 monomers (B2(a)-(g)) at 195 eV.

(A) MS2 ions of the main monomer species B2(a) and B2(b). (B) MS2 ions of the identified monomer sub-species B2(c)-(j). The fragment ions (b7 = 910.62 m/z) for B2(f) and (b15 = 1741.26 m/z) for B2(j) show a different distribution of the isotopic peak pattern due to peak overlays. Peak intensity labelled by blue italic numbers. Conditions: 200 mM ammonium acetate + 1 mM DTT pH 7.2, protein concentration 8 μ M, collision energies 195 eV.

The obtained MS2 results above show the isotope peak 2333 m/z (b20 (1+)), which is specific for the N-terminus. This ion is characteristic of the B2(b) monomer (80.779 kDa), which was identified to be identical to the detected B3(b) monomer. Furthermore, this ion is noticeable due to its peak shape. The isotopic distribution shows an overlay with another peak species (2335 m/z (1+)), which appears also as a singly charged b20 ion. This ion can be assigned to the monomer B2(g) (80.678 kDa, missing aa1-16 at the N-term. and aa752-745 at the C-term.) (Tab. 11). It turned out that some monomeric species lack the same C-terminal amino acids, but that they differ at the N-terminus and *vice versa*. Thus, each show

the same CID y-ion pattern but differ in b-ions. In the case of the y21 ion (1053.79 m/z (2+)), this presumably leads to an increase in the peak intensity compared to other fragments. As a result, it can also be assumed that the dominant singly charged ions (Fig. 37, Fig. 41, e.g. y19 and y22) also acquire their increased intensity through the simultaneous dissociation of sub-monomers (B2(c)/(g)) and main-monomers (B2(a)/(b)). In addition, based on the MS1 masses of the sub-monomers and on the fragment ions obtained, it can be assumed that the degradation of the protein species differs. Certain aa-segments seem to degrade preferentially.

C- and N-terminal sequence differences indicate that the N-terminus without His-tag (B2(b)/B3(b), missing aa1-15 at the N-term.) is degraded less quickly than the tagged N-terminus after two weeks of auto-proteolysis at + 4 °C (Fig. 33, B3). Here the previously more dominant B2(a) monomer had clearly lost its intensity. However, this was not observed after one week of auto-proteolysis (Tab. 11) as is clearly shown in Fig. 39. The associated sub-monomers of both main species have lost different sections of the aa-sequence at the C- and N-terminal ends.

Tab. 11 Batch 2 monomers: overview of the specific MS2 fragment ions with their associated protein amino acid sequence (B2(a)-B2(j) monomer).

lProF = low mass MS2 protein fragments, z = charge, MW = monoisotopic mass [M], TP = theoretical peptide m/z , aa = amino acid.

Main species	kDa	Lack aa	lProF	TP	m/z	z	Peptide sequence
B2(a)	82.683	<i>C term.</i> (aa ₇₅₂₋₇₄₅)	y21	1053.57	1053.79	2+	PDPGNQGTVEAG RALKQVVGL
			y22	2292.21	2292.43	1+	WPDPGNQGTVEAG RALKQVVGL
		<i>N term.</i> —	b5	564.23	564.44	1+	MASHH
B2(b)	80.779	<i>C term.</i> (aa ₇₅₂₋₇₄₅)	y21	1053.57	1053.79	2+	PDPGNQGTVEAG RALKQVVGL
			y22	2292.21	2292.43	1+	WPDPGNQGTVEAG RALKQVVGL
		<i>N term.</i> (aa ₁₋₁₅)	b20	2333.23	2333.46	1+	LYFQGSKRHIPVV TDIYSVE
Sub-species	kDa	Lack aa	lProF	TP	m/z	z	Peptide sequence
B2(c)	83.503	<i>C term.</i> —	y6	630.35	630.42	1+	AENALL
		<i>N term.</i> —	b5	564.23	564.44	1+	MASHH
B2(d)	82.781	<i>C term.</i> (aa ₇₅₂₋₇₄₆)	y8	843.53	843.55	1+	LKQVVGLS
		<i>N term.</i> —	b5	564.23	564.44	1+	MASHH
B2(e)	82.845	<i>C term.</i> (aa ₇₅₂₋₇₄₈)	y8	774.44	774.46	1+	QVVGLSTA
		<i>N term.</i> (aa ₁)	b6	707.31	707.38	1+	ASHHHH
B2(f)	82.499	<i>C term.</i> (aa ₇₅₂₋₇₄₅)	y21	1053.57	1053.79	2+	PDPGNQGTVEAG RALKQVVGL
			y22	2292.21	2292.43	1+	WPDPGNQGTVEAG RALKQVVGL
		<i>N term.</i> (aa ₁₋₂)	b7	910.39	910.62	1+	SHHHHHH
B2(g)	80.678	<i>C term.</i> (aa ₇₅₂₋₇₄₅)	y21	1053.57	1053.79	2+	PDPGNQGTVEAG RALKQVVGL
			y22	2292.21	2292.43	1+	WPDPGNQGTVEAG RALKQVVGL
			b7	867.45	867.65	1+	YFQGSKR
		<i>N term.</i> (aa ₁₋₁₆)	b20	2335.45	2335.17	1+	YGQGSKRHIPVVTD IYSVED
B2(h)	80.889	<i>C term.</i> (aa ₇₅₂₋₇₄₆)	y8	843.53	843.55	1+	LKQVVGLS
		<i>N term.</i> (aa ₁₋₁₅)	b20	2333.23	2333.46	1+	LYFQGSKRHIPVV TDIYSVE
B2(i)	81.099	<i>C term.</i> (aa ₇₅₂₋₇₄₇)	y9	944.58	944.67	1+	LKQVVGLST
		<i>N term.</i> (aa ₁₋₁₄)	b19	2219.16	2219.46	1+	NLYFQGSKRHIP VVTDIYS
B2(j)	80.616	<i>C term.</i> (aa ₇₅₂₋₇₄₆)	y8	843.53	843.55	1+	LKQVVGLS
		<i>N term.</i> (aa ₁₋₁₇)	b15	1741.94	1741.26	1+	FQGSKRHIPVVT- DIY

3.3.4 The mass of the B1(a) monomer is close to the full-length protein

Based on the MS1 data, batch 1 (storage = one day + 4 °C, during the MS analysis) shows the smallest structural changes (Fig. 42). However, it was found that B1(a) = 83.374 kDa does not correspond to the exact molecular weight (MW = 83.521 kDa) of the complete aa-sequence. The difference is 146 Da. This is a relatively large variance that can hardly be explained by broad peaks or poor signal intensity. In fact, this rather large difference suggests that an amino acid is missing.

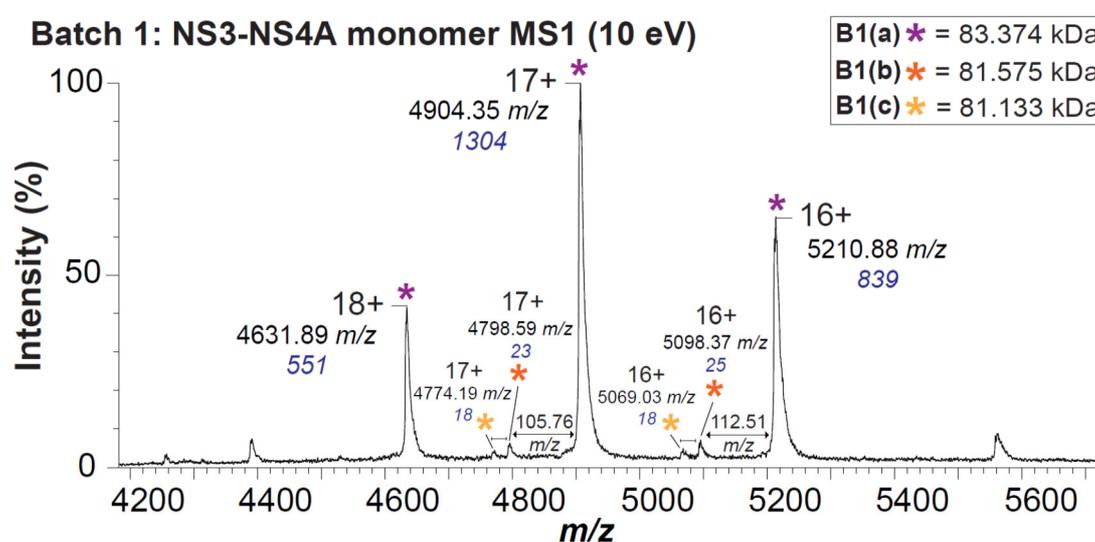


Fig. 42 MS1 analyses of the B1 monomer identified the main and two sub-monomers at 10 eV. Main monomers B1(a) = 83.374 kDa, sub-monomers B2(b) = 81.575 kDa and B2(c) = 81.133 kDa, (protein storage for 1 day at + 4 °C while measuring). Peak intensity labelled by blue italic numbers. MS1 energy was set to 10 eV, ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 7 μ M.

The first N-terminal amino acid (methionine (aa M1) = 149 Da -H₂O) fits to this difference in mass rather well. However, at the C-terminal end there is a leucine (aa L752) = 131 Da, which could also match this modification. Regarding the slightly larger mass difference, it is suggested that there is a lack of H₂O (M-18). Neutral losses of the various monomeric molecular weights were basically observed in the range from M-1 to M-18, which includes NH₃, H₂O. Additionally because of the rather “low” resolving power of the used instrument (10,000 full width at half maximum (FWHM)) [203], in contrast to small peptides, no isotopic resolution of the intact protein, could be measured. As such no monoisotopic mass could be determined which results in mass uncertainties. Based on MS2 information it is assumed that the B1(a) monomer lacks L752 at the C-terminal site. This could already have

happened during protein expression whereas another assumption would be that the protein is very sensitive to temperature fluctuations. Therefore, the missing of this amino acid could be explained by a gradual degradation of the protein by its own serine protease before it is frozen or thawed or when it is stored on ice for several hours. This hypothesis is supported by the fact that batch 2 monomers also show missing amino acids at the C-terminal end. While for example B2(a)/(b) lack the aa752-745, B2(e) only misses the aa752-748 and B2(d) lacks aa752-746 (Tab. 11).

In addition, two further peak series with low intensity were observed in batch 2. These could be assigned to the two monomers B1(b) = 81.575 kDa and B1(c) 81.133 kDa. Both proteins must have a C-terminal deficiency in certain amino acids (Tab. 12 and Tab. 13). Based on the complete aa-sequence and the calculated MS1 masses, B1(b) seemed to miss the aa752-734 at the C-terminal end. B1(c), however, must lack the aa1-4 at the N-terminal end as well as the aa752-734 at the C-terminus. The previous MS2 data indicate that in addition to the already identified B2 monomers, the two species B1(b) and B1(c) are also covered in the spectra. This assumption is confirmed by their corresponding y- and b-ions, which are shown in Tab. 12 and in the supplement section 6.3 (S. Fig. 4).

Tab. 12 Batch 1 MS2 ions that were identified while dissociating the B2(b) monomer.

Ions are based on 17+ precursor peak of B2(b). The amino acid sequence of the y9 fragment ion is marked in blue. TP = theoretical peptide mass (Da), lProF = low mass MS2 protein fragments, z = charge, C-terminal ions = y, N-terminal ions = b, aa = amino acid sequence of the MS2 ions, CID energy = 195 eV.

B1(b)+(c)	TP	<i>m/z</i>	lProF	<i>z</i>	aa-sequence
	916.40	916.67	y9	1+	DPGNQGTVE
	1427.64	1427.38	y13	1+	LDWPDPGNQGTVE
	1484.67	1483.94	y14	1+	GLDWPDPGNQGTVE
	1710.83	1710.22	y16	1+	LLGLDWPDPGNQGTVE
	2009.98	2009.36	y19	1+	AVELLGLDWPDPGNQGTVE
	2238.09	2237.55	y21	1+	DLAVELLGLD- WPDPGNQGTVE
B1(c)	1097.48	1096.81	b8	1+	8xHis
	1753.77	1753.06	b13	1+	9xHis+ENLYF

In principle, the presence of these two other sub-monomers was not expected because the batch 1 sample was very fresh and was only stored for a few hours at + 4 °C throughout the period of MS measurements. However, these two sub-species were present in the sample from the beginning of the first MS measurement. The appearance of these other species, however, reinforces the assumption that the protein is very sensitive to storage conditions

and shows a high rate of auto-proteolysis. Furthermore, non-specific cleavage products may have been formed through contamination with NS3-proteases, which were coeluted during protein purification. However, these two sub-monomers could not clearly be identified by MS1 in batch 2. For one, it could mean that they were quickly degraded. A more reasonable assumption, however, is that they are hidden under the main monomer peak species of B2(b), similar to the other B2 sub-monomers that have already been identified.

Tab. 13 MS1 identifies the lack of C-/N-terminal aa of B1 and 2 monomers based on peak distances. Shown are batch 1 monomers B1(a)/(b)/(c) and as a further example batch 2 monomers B2(a)/(b)/(c)/(d). TM = theoretical monomer mass (kDa), z = positive ion charge, aa = lack of N-/C-terminal amino acids based on MS1 mass and aa-sequence in comparison with peak distances/differences (PD), energy = 10 eV.

B1 relation	TM (kDa)	B1(a) (kDa)	B1(b) (kDa)	B1(c) (kDa)	PD (a:b) m/z	PD (b:c) m/z	
	83.521	83.374	81.575	81.133	105.76 (17+)	112.51 (16+)	
(TM:a), (a:b), (b:c)	–	0.131	1.799	0.442	29.34 (17+)	24.40 (16+)	
Lack aa (kDa)	–	0.131	1.797	2.241 (1.797+ 0.444)			
C-terminal	–	L ₇₅₂	(L ₇₅₂)* + aa ₇₅₁₋₇₃₄ (AGRAL-KQVVGL STAENAL)	(L ₇₅₂)* + aa ₇₅₁₋₇₃₄ (AGRAL-KQVVGL STAENAL)			
N-terminal	–	–	–	aa ₁₋₄ (MASH)	(*It is assumed that all species are lacking the C-terminal aa(L ₇₅₂).		
B2 relation	TM (kDa)	B2(a) (kDa)	B2(b) (kDa)	B2(c) (kDa)	B2(d) (kDa)	PD (a:b) m/z	z
	83.521	82.683	80.779	83.503	82.781	105.32	18+
(TM:a/c/d) (a:b)	–	0.838	1.903*	0.018**	0.740	112.18	17+
						119.11	16+
∑ Peak diff.						1902.86	
Lack aa (kDa)	–	0.818	1.921	–	0.730		
C-terminal	–	aa ₇₅₂₋₇₄₅ (STA-ENALL)	aa ₇₅₂₋₇₄₅ (STA-ENALL)		aa ₇₅₂₋₇₄₆ (TA-ENALL)		
N-terminal	–	–	aa ₁₋₁₅ (MASHHHH HHHHHH EN)		–		

(* 1.903 kDa = deviation of the mass compared to the aa-sequence due to the width of the protein peaks.
** MW neutral loss M-18 = H₂O).

3.3.5 Slight differences between the B1 and B2 monomers

Presented data show, that the monomers of the NS3-NS4A complex occur temporarily and are almost completely degraded over the course of three weeks when stored at + 4 °C. This results in the release of a larger amount of the NS3-helicase subunit. Interestingly, the MS signal of the NS3-protease is always very low. This is presumably an indication that the protease proteolysis itself after release from the protein complex based on its increasing concentration. An overview of the proteins identified during the auto-cleavage process and the corresponding locations of the cleavage are shown in the table below (Tab. 14).

Since most of the sub-monomers were identified in batch 2, this is particularly interesting for further analysis. The presence of the various protein species suggests that the NS3-protease was particularly active in this sample. In order to obtain an overview of the distribution and the relationship between the monomers, the individual peak intensities of the charge states (mostly 18+ to 15+) were averaged and their ratios were compared as shown in Fig. 43 (see S. Tab. 4 in section 6.4 in the supplement). As a comparison to batch 2, batch 1 was used in which the fewest cleavage products of the NS3-protease should be present. As expected, the analysis of the monomers from batch 1 showed that B1(a) (missing aa752) is the dominant species with 82% (Fig. 43, (A)). However, the two sub-monomers B1(b) (missing aa752-734) with 11% and B1(c) (missing aa1-4, aa752-734) with 7% show a significant presence. In accordance with the MS1 spectra (Fig. 42), it can be clearly seen that the NS3-NS4A protein is very fragile, reacts quickly to external conditions and is cleaved and proteolytically active *in vitro* after a short time at + 4 °C.

In contrast, the analysis of the monomers from batch 2 showed a more prominent degradation into various species, which obviously lack different C- and N-terminal amino acids. The overview of the individual monomers showed that B2(a) (missing aa752-745) with 32% and B2(b) (missing aa1-15, aa752-745) with 18% represent the largest percentages, as expected (Fig. 43 (B), S. Tab. 5). The data show that both, batch 1 and batch 2 have dominant species (B1(a) and B2(a)/B2(b)). In B2 the monomers are more distributed across different sub-species (Fig. 43 (B)), whereby B1 shows a lower amount of sub-monomers (Fig. 43 (A)). Data associated with B2 demonstrate the process of the breakdown of the NS3 protein. Due to the occurrence of these many sub-monomers, it could be assumed that the

NS3-protease splits off the individual amino acids at the N- and C- terminus indiscriminately. However, they could also represent an image of the individual monomer levels at a specific point in time.

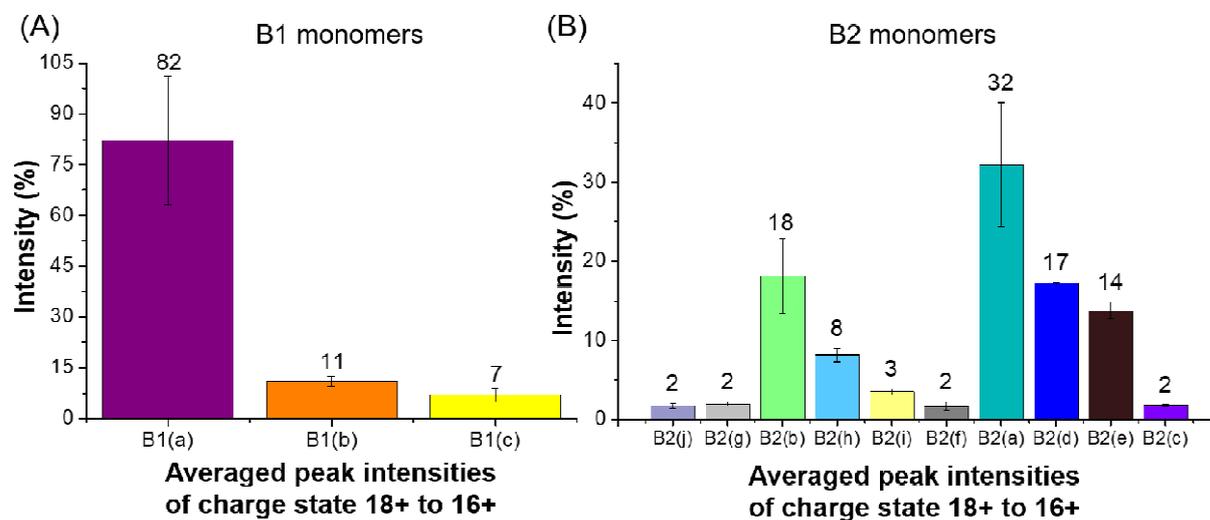


Fig. 43 The ratio between the different batch 1 and 2 monomers shows the hierarchical distribution of the identified protein species.

Ratios between species are labelled as (%) above the bars, see S. Tab. 5. MS1 energy was set to 10 eV, ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 7 μ M B1 and 8 μ M B2. Data are based on averaged peak intensities.

Thus, it appears that the protease would initially attack the NS4A-cofactor at the C-terminal end within a week of storing at + 4 °C. This part is partially split off, which would lead to the formation of B2(e) (missing aa1, aa752-748) with 14% and B2(d) (missing aa752-746) 17%, for example. At the same time, the N-terminal part of the protein is attacked, which then results in other lower intense species such as B2(f) with 2%, B2(i) 4% and B2(h) 8%. In the further course the N-terminus would split off up to aa1-15 and the C-terminus up to aa752-745 (between the already known cleavage motif of Leu and Asn, [89]), whereby B2(b) appears very dominant after one week at + 4 °C.

However, the degradation of the N-terminal end seems to proceed more slowly than the C-terminal end. This could explain that B2(a) is formed, which has lost the C-terminal NS4A cofactor part (aa752-745) but has not yet been attacked at the N-terminus. The two monomers B1(a) and B1(b), which lacked the suspected aaL752 and aa752-734 after a short time, also indicate that the C-terminal end is attacked first (Tab. 11, Tab. 13, Tab. 14). In order to investigate the distribution of the individual sub-monomers in relation to the main

peak, all species with lower mass, which were partially overlaid by peak B2(a) were considered separately from those below the B2(b) peak (Fig. 44). This way, the peak proportions of the sub-species in the main signal can be observed more closely. Thus, it is shown that B2(b) with 54.3% (Fig. 44 (B)) has a larger share of the total peak signal than B2(a) with 48.4% (Fig. 44 (A) and S. Tab. 8).

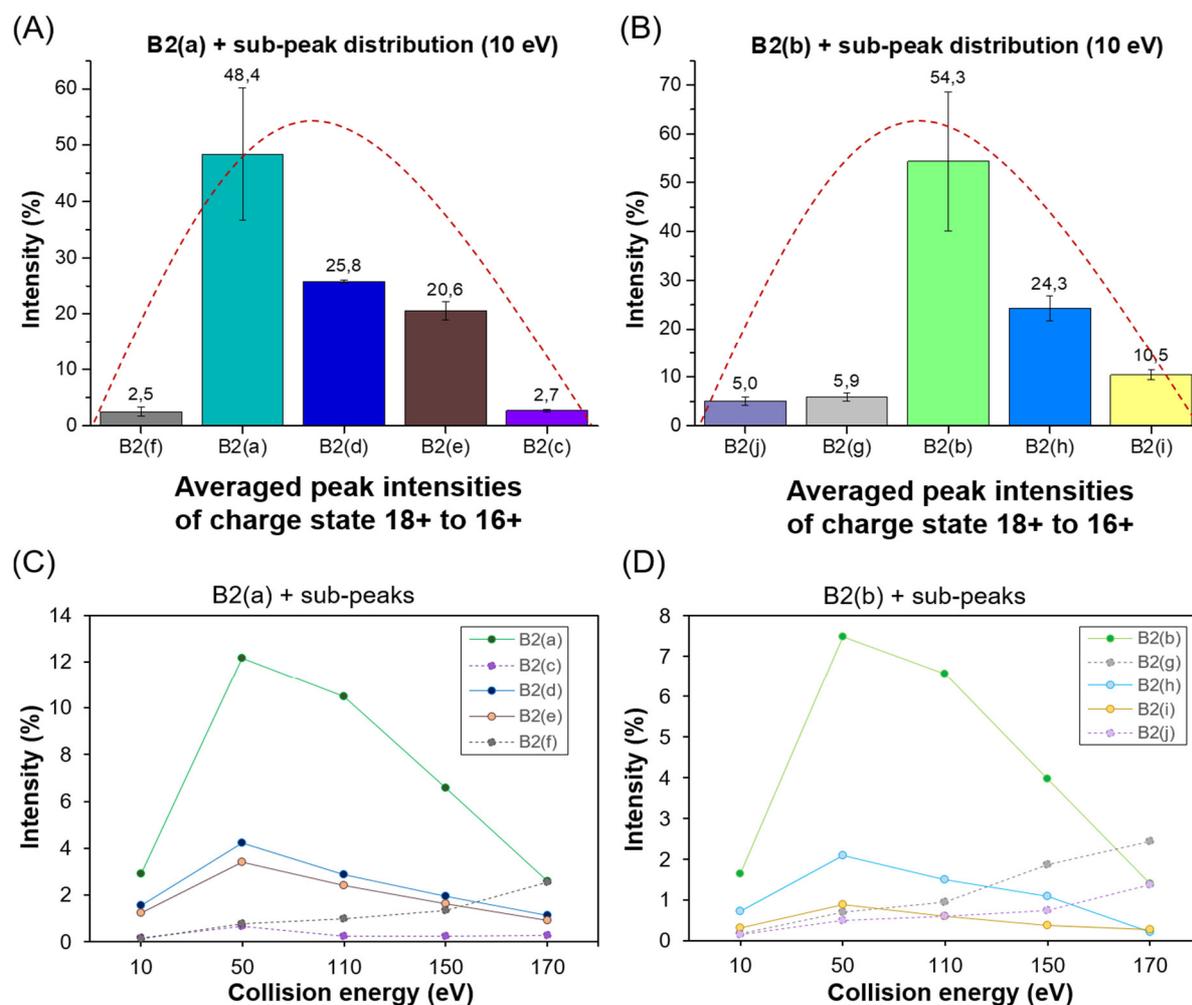


Fig. 44 Based on their intensities, the B2(a) and B2(b) main peaks (17+) show a similar distribution of the associated sub-species at 10 eV, while different activation energies revealed strong changes. Activation energy: (A)+(B) MS measurements were performed at 10 eV. (A) Associated bars with B2(a) (dark-cyan) are labelled with dark colours, (B) associated bars with B2(b) (light-green) are labelled with light colours, red dashed line marks the shape of the overall peak. (C)+(D) MS1 data with applied energies of 10 eV to 170 eV. (C) Associated lines with B2(a) (dark-green) are labelled with dark colours, (D) associated lines with B2(b) (light-green) are labelled with light colours. Starting intensity at 10 eV for B2(a) = 2.9% (SD = 0.7), B2(b) = 1.6% (SD = 0.2), end intensity at 170 eV for B2(a) = 2.6% (SD = 0.9), B2(b) = 1.4% (SD = 0.3), see S. Tab. 8 and S. Tab. 9. Ammonium acetate 200 mM, pH7.2 + 2 mM DTT, protein concentration was 8 μ M.

Here, the peak proportion of the two monomers B2(d) with 25.8% and B2(e) 20.6% clearly comes into play. This in turn could indicate that B2(a) is in the degradation process and therefore starts to decrease faster than B2(b). On the other hand, as a result of this and because of the degradation of further sub-monomers, B2(b) exhibits an increased signal intensity. The fact that the NS3-protease degrades all monomers in direction to B2(b) is also evident from its ascendancy in batch 3. Here B3(b) is identical to B2(b) as shown in Tab 14. The associated MS1 data show that the signal intensity of B2(a) at 10 eV is significantly reduced after three weeks of storage at + 4 °C (Fig. 33). Furthermore, it is noticeable for the batch 2 sub-species that most N-terminal cleaved amino acids were not associated with a specific cleavage motif of the NS3-protease (Tab. 14), except for the cleavage that leads to the formation of B2(b) (Leu, Asn motif). Therefore, the other sub-species may be non-specific cleavage-products.

In order to observe the dynamics of the protein complexes under different activation steps, to get more information about the stability and behaviour of the monomers in addition to the previously presented MS2 analyses, higher ion activation energies were also examined in MS1 mode (Fig. 44 (C)/(D) and S. Tab. 9). For this purpose, voltages of 10 eV to 170 eV were tested. Usually, in native ESI-MS experiments sharp protein peaks can be measured between 10 and 50 eV [204]. Occasionally, solvation shell, detergent or adducts such as ammonium and sodium ions from the sample solution can be attached to the proteins surface, which effects the peak shape [153]. If the signal peaks are too wide, it makes sense to use slightly higher energies for collisional-induced cleaning [153, 154, 205, 206]. However, by applying stronger ion activation energies (up to 200 eV) complexes can split off due to local structural unfolding which can result in dissociation. Thus, an intact conformation of the proteins is no longer guaranteed [207]. Higher collisional activation however, also generates a purer peak signal [153, 208], which in this case led to a more precise identification of the sub-monomers as shown in Fig. 40, Fig. 43 and Fig. 44.

The presented MS data in Fig. 44 (C)/(D) show that the highest signal intensity of all monomers is reached at 50 eV (Fig. 44 C)/(D), solid lines) and the sharpest peak signal is at 110 eV (supplement S. Fig. 8). The intensity as well as the peak width changes and the resolution decreases steeply from 110 to 195 eV (Fig. 44 (C)/(D, S. Fig. 8). This is clearly shown for the main protein species B2(a) (82.7 kDa, dark-green line) and B2(b) (80.8 kDa, light-green line). A slight shift to the lower m/z -range was also observed which results from

a loss of mass (fragmentation). First comparing the effect of ion activation on the peak intensities, B2(a) starts with an initial value of 2.9%, B2(b) begins with an intensity of 1.6%. At 50 eV, B2(a) has reached approx. 12.2% and B2(b) 7.5%. This results in a factor of 0.24 for B2(a) and a slightly lower factor of 0.21 for B2(b). This factor is the result from the comparison of the individual intensities (50 eV to 170 eV) examined in relation to the intensity at 10 eV (e.g. ((%) B2(a) 10 eV)/((%) B2(a) 50 eV) = 0.24). The greater the factor in relation to a series (e.g. B2(a) 10 eV to 170 eV), the lower the increase in intensity within this series. When considering the other energies up to 150 eV, it becomes apparent that the factor of B2(a) is always above that of B2(b). At 50 eV it is 0.24 to 0.21, at 110 eV 0.28 to 0.24 and at 150 eV 0.44 to 0.40. This shows that the two proteins behave slightly different when the same energies are applied, which is also confirmed by investigations based on the averaged peak areas of the different charge states (Fig. 45).

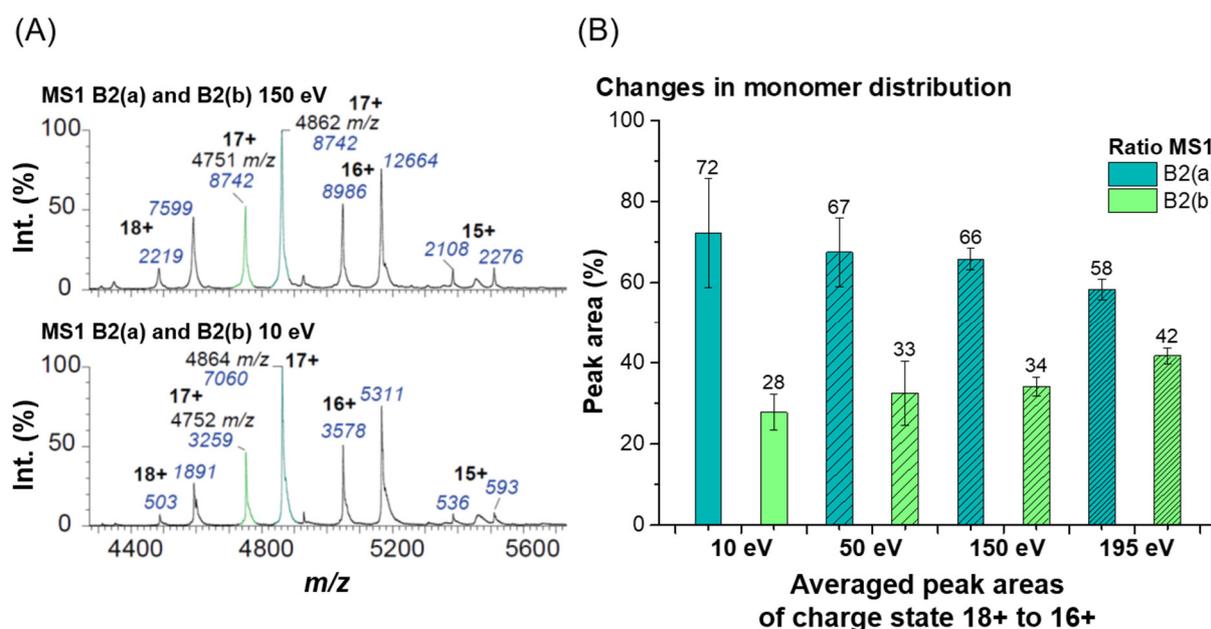


Fig. 45 Entire peak width of the B2(a) and B2(b) including sub-monomers showed a contrary distribution at different energies (10 eV to 195 eV).

(A) Native MS1 spectra recorded at 10 eV and 150 eV. B2(a) labelled by a light-green line, B2(b) labelled by a dark-cyan line. Peak intensities are labelled by blue italic numbers, charge states are indicated. (B) Distribution of the monomers at different energies. Data are based on the averaged peak areas of charge states (18+ to 16+), see S. Tab. 10 and S. Tab. 11. Ratios between species are labelled as area (%) above the bars. MS1 energy was increased from 10 eV to 195 eV. Further conditions: ammonium acetate 200 mM at pH 7.2 + 2 mM DTT, protein concentration was 8 μ M.

B2(a) is activated slightly more by collision with gas molecules than B2(b) (Fig. 45 (B)). The increased activation of B2(a) could be explained by a larger surface area that this monomer

species offers. This indirectly gives information on the protein structure and shape, which is not as accurate as MS2 analyses or ion-mobility (IMS-MS) mass spectrometry investigations. Nevertheless, it shows that the two proteins behave differently under the same conditions. Which is due to their slightly different structure, which was already confirmed by the MS2 results.

Where B2(a) and B2(b) are identical at the C-terminal end (lack of aa752-745) but differ at the N-terminus (B2(b) lacks aa1-15). Correspondingly, B2(a) has a slightly higher mass than B2(b) and is localized in the higher m/z -range for the same charge state. In association with the slightly larger protein surface and the higher m/z -range of B2(a), the exposed 15 N-terminal amino acids on B2(a) lead to a higher degree of vibration resulting in a more labile structure at higher collision energies. With an energy of 170 eV, (a) and (b) are just below their starting intensity at 2.6% and 1.4%, which results in the factor 1.12 for (a) and 1.14 for (b). The sub-monomers B2(d)/(e)/(h)/(i), on the other hand, show a flatter decrease in intensities for the different energies. The starting intensity are B2(d) = 1.6%, B2(e) = 1.2%, B2(h) = 0.7% and B2(i) = 0.3%. At 50 eV (d) has an intensity of 4.2%, (e) 3.4%, (h) 2.1% and (i) 0.9%. For these sub-monomers, the activation of the larger species (d) = 82.8 kD and (e) = 82.9 kDa also results in a higher factor (0.38 and 0.35) compared to the smaller ones (both 0.33) which is in line with B2(a) and B2(b) and is almost observed until 110 eV with intensities of 2.9%, 2.4%, 1.5% and 0.6%. Further, the two smaller monomers (h) (80.9 kDa, lack of aa1-15 and aa752-746) and (i) (81.1 kDa, lack of aa1-14 and aa752-747) differ not as much due to their size and shape (diff. approx. 200 Da) as B2(a) and (b) (diff. approx. 1900 Da). This results in a similar activation for both species at 50 eV but starts to differ at 110 eV (0.47 and 0.5) and is also indicated by the identical factor of 0.33 at 50 eV. When looking at the lower intense monomers (B2 (f)/(g)/(j)) it is noticeable that these increase in intensity with increasing energy up to 170 eV (dashed lines). In fact, this can mainly be explained by the shift of the entire peaks from higher to lower m/z -values. All monomers lose mass due to the collision with gas molecules at higher energies, as a result a better separation and thus a more precise quantification of the proteins is possible. The more dominant monomers are represented by a higher intensity. Since they occur in a larger amount than the lower intense species B2(f)/(g)/(j), it has to be concluded that they collide more often with the inert gas molecules and lose mass more quickly. This leads to peak overlays and gives the smaller front peaks a higher artificial intensity. This peak overlay

could lead to an incorrect interpretation of the data. For this reason, these monomer species are initially disregarded in this analysis. The comparison of the B2(a) and (b) in Fig. 45 is based on the averaged peak area of different charge states. The results have shown that the monomers behave differently at the same MS1 energies (10 eV to 195 eV). Due to the peak overlaps, the other sub-monomers are also included in the calculation, with the standard deviation being quite high, especially at 10 eV. Therefore, the percentage peak area at 10 eV appears rather high in comparison to Fig. 44 (intensities of the individual monomers of the 17+ peak). However, the overall trend is very similar. It is of particular note that B2(a) + sub-monomers tend to lose surface at higher energies whereby B2(b) + sub-monomers increase. This corroborates the assumption that B2(a), due to its larger surface area, collides more frequently with the gas molecules than B2(b) and breaks down in the process. For both proteins (B2(a) and (b)) the observed y19 ion occurs in MS1 and MS2 mode at 150 eV as shown in Fig. 46 (B)/(C). The ratio on MS2 level (based on their peak intensities) between y19 and B2(a) ($B2(a)_{\text{int}}/y19_{\text{int}} = 26300/1520 = 17.3$) is lower than the ration between y19 and B2(b) ($B2(b)_{\text{int}}/y19_{\text{int}} = 17200/792 = 21.7$). It can therefore be assumed that the majority of the y19 ion originate from B2(a) (Fig. 46) and thus B2(a) is more prone to fragmentation than B2(b).

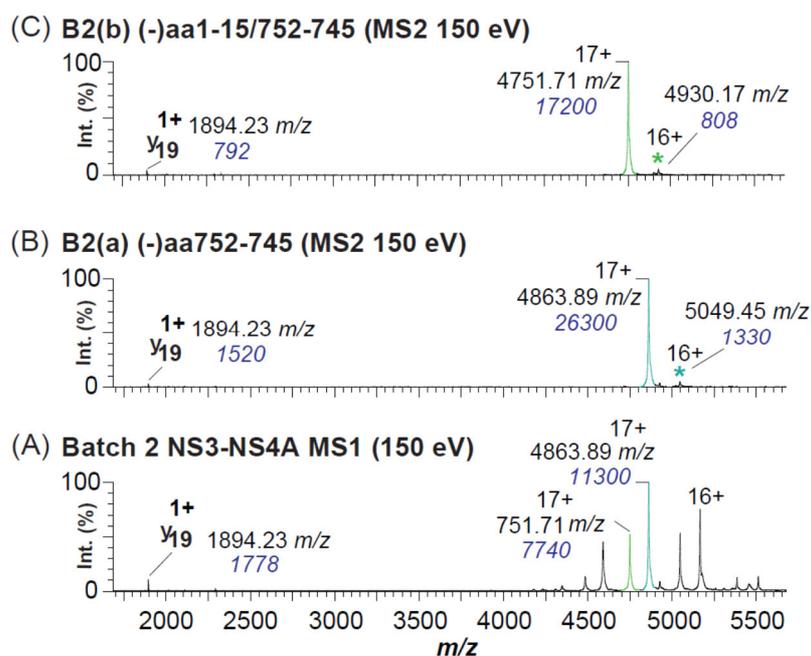


Fig. 46 The majority of the y19 ion originate from the bigger monomer B2(a).

(A) Spectrum shows the two dominant monomers B2(a) and (b) of batch 2 and the associated fragment ion y19 at 150 eV in MS1 mode (one week of auto-proteolysis). (B) MS2 data of the 17+ precursor of B2(a) (150 eV). (C) MS2 data of the 17+ precursor of B2(b) (150 eV). Conditions: 200 mM ammonium acetate + 2 mM DTT pH 7.2, protein concentration 6-8 μM , peak intensities are labelled by blue italic numbers.

Tab. 14 Comparison of the cleavage products identified in the auto-proteolysis process of the NS3 protease with known cleavage motifs.

Shown are the name and species (spec.) of the identified proteins as well as the identified lack of N- and C-terminal amino acids (aa) based on Top-Down sequencing. The location of the cleavage within the protein complex and the corresponding cleavage motifs are shown. Green bars within the table indicate previous identified cleavage sites (e.g. Tautz *et al.* 1997 [55] and Lamp *et al.* 2013 [92]). Missing aa labeled with (**) indicate the possibilities of other cleavages. Cleavage motifs labeled with (*) indicate not known motifs or non-specific cleavages. Proteins marked with (‘) and (“) show identical species within different batches.

Protein name	NS3-spec.	Missing aa N-terminal	Missing aa C-terminal	N-term. location of cleavage	N-term. cleavage motif	C-term. location of cleavage	C-term. cleavage motif
B1(a)	mon.	-	aaL572	-	-	NS4A	L/L*
		-	aa752-751**	-	-	NS4A	A/L
		aaM1**	-	s.-sequence	M/A*	-	-
B1(b)	mon.	-	aa752-734	-	-	hel.	E/A*
B1(c)	mon.	aa1-4	aa752-734	H-tag	H/H*	hel.	E/A*
B2(a)	mon.	-	aa752-745	-	-	NS3/NS4A	L/S
‘B2(b)	mon.	aa1-15	aa752-745	TEV	N/L	NS3/NS4A	L/S
B2(c)	fl.mon.	-	-	-	-	-	-
B2(d)	mon.	-	aa752-746	-	-	NS4A	S/T*
B2(e)	mon.	aaM1	aa752-748	s.-sequence	M/A*	NS4A	A/E*
B2(f)	mon.	aa1-2	aa752-745	s.-sequence	A/S*	NS3/NS4A	L/S
B2(g)	mon.	aa1-16	aa752-745	TEV	L/Y*	NS3/NS4A	L/S
B2(h)	mon.	aa1-15	aa752-746	TEV	N/L	NS4A	S/T*
B2(i)	mon.	aa1-14	aa752-747	TEF	E/N*	NS4A	T/A*
B2(j)	mon.	aa1-17	aa752-746	TEV	Y/F	NS4A	S/T*
‘B3(a)	hel.	aa1-253 Tag-NS4A- NS3-pro.	aa752-745	pro./hel. aa192/193	L/M	NS3/NS4A	L/S
‘B3(b) B2/3(b)	mon.	aa1-15	aa752-745	TEV	N/L	NS3/NS4A	L/S
B3(c)	pro.	aa1-61 Tag-NS4A	aa752-253	GSGS-linker	S/G*	pro./hel. aa192/193	L/M
‘B4(a) B3/4(a)	hel.	aa1-253 Tag-NS4A- NS3-pro.	aa752-745	pro./hel. aa192/193	L/M	NS3/NS4A	L/S

3.4 Protein complex dimerization plays a role for NS3-NS4A

In the following, dimerization of batch 1 (B1) and batch 2 (B2) proteins is in the focus. For batch 1 the main protein dimer was identified based on its charge state distribution from 26+ or 27+ to 22+ and its mass of 166.8 kDa as the homo-dimer of the main monomer of batch 1 (B1(a)). Thus, it is named as B1d(a) dimer and is shown in more detail in Fig. 46.

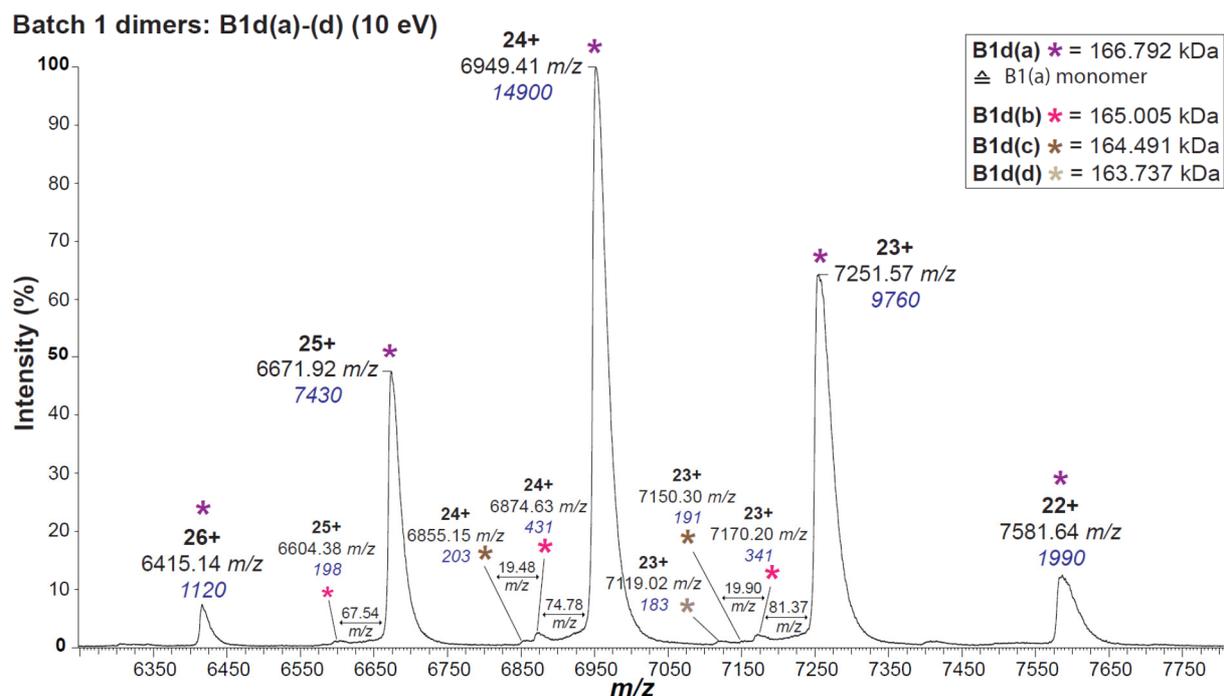


Fig. 46 The monomer B1(a) ((-)-aaL752 species) tend to form the homo-dimer B2d(a).

The most prominent dimer B1d(a) (labelled with a violet-coloured star) shows a distribution of 26+ to 22+. Smaller sub-dimers were identified and labelled with pink- and brown-coloured stars. Measurements were performed in MS1 mode at 10 eV in 200 mM ammonium acetate, pH 7.2 + 2 mM DTT, protein concentration was 7 μ M (monomer). MS measurements were performed at 10 eV.

The 24+ or the 25+ peak mostly show the highest intensity. Based on the good correlation of MS1 mass to theoretical mass, B1d(a) can be identified as the homo-dimer of the monomer B1(a) (83.374 kDa) despite the small mass deviation. It is assumed that both proteins lack the aa(L752), which belongs to the NS4A-cofactor part that is located at the C-terminal end of the aa-sequence. In the absence of sufficient MS2 data, however, it cannot be ruled out that the first amino acid (M1) is missing in the sequence instead of aaL752. Both positions in the sequence do not exactly match a known cleavage motif of the NS3-protease, but in the case of L752 the Leu residue is at least part of known cleavage sites (Tab. 14). Therefore, it is suggested that both the monomer and dimer lack L752, whereby their

masses do not completely match the theoretical aa-sequence (83.521 kDa monomer, 167.042 kDa dimer).

Furthermore, MS data show the co-existence of a small amount of additional sub-dimers named as B1d(b) (165.01 kDa, B2d(c) (164.5 kDa) and B2d(d) (163.7 kDa) (Fig. 46). Their occurrence is similar to the batch 1 monomers, which were introduced before (sub-chapter 3.3.4, Fig. 42). These smaller dimers have much lower peak intensity than the dominant species B1d(a), but show a similar distribution of charges in a different m/z -range close to the main peaks (25+, 24+, 23+) (Fig. 46, pink-, brown- and beige-coloured stars).

Based on the good correlation to the theoretical mass, as described above for the main homo-dimer B1d(a) to simplify the data evaluation, it is hypothesized that the smaller ones are also made up of two identical monomers and thus form homo-dimers. Based on this their hypothetical monomer content was determined and the computationally based theoretical deficiency of amino acids was estimated (S. Tab. 7). For this simplified assumption, the molecular masses of the dimers were halved, which resulted in the theoretical monomer masses (CM) (Tab. 14).

Tab. 14 Identification of different dimers in batch 1 (B1d) and their assumed missing aa-segments. CM = monomer masses are based on calculations of the dimer mass (= 1/2 the dimer), TM = theoretical monomer mass (aa-sequence based), TD = theoretical homo-dimer mass, (*) = fits the real monomer B1(a).

TD (kDa)	TM (kDa)	B1d(a) (kDa)	CM (a) (kDa)	B1d(b) (kDa)	CM (b) (kDa)	B1d(c) (kDa)	CM (c) (kDa)	B1d(d) (kDa)	CM (d) (kDa)
167.042	83.521	166.792	83.396*	164.997	82.499	164.491	82.245	163.795	81.897
CM missing (aa) C-term.		(-)aaL752		(-)aaL752		(-)aaL752		(-)aa752-750	
CM missing (aa) N-term.		-		(-)aa1-7 (MASHHHH)		(-)aa1-9 (MASHHHHHH)		(-)aa1-10 (MASHHHHHHH)	

In comparison with the theoretical monomer mass (TM) and the aa-sequence the first sub-dimer B1d(b) is lacking 1006 Da. It also appears to have the assumed a C-terminal deficient of aa(L752). Furthermore, a specific segment at the N-terminus is probably missing, aa(1-7). In contrast the second sub-dimer B1d(c) presumably shows a lack of the aa(1-9) at the N-terminus but also misses aa(L752) at the C-terminal end. The third sub-dimer B1d(d) has the lowest mass of the identified dimers. Accordingly, most of the C- and N-terminal amino acids of these sub-species are missing. This dimer seemed to lack the aa(752-750) at the C-

terminus and aa(1-10) at the N-terminus (Tab. 14). Possibly these smaller dimers are already non-specific cleavage products of the auto-proteolytic activity of the NS3-protease.

In order to obtain an overall overview of the distribution of monomer and dimer on the first day of the proteolysis test, their ratios were compared to one another. The determination of the ratios is based on the main monomer B1(a) and the main dimer B1d(a). The evaluation of data is based on the different peak areas (Fig. 47, S. Fig. 6, S. Tab. 6). This resulted in a calculated ratio between 13% monomer and 87% homo-dimer.

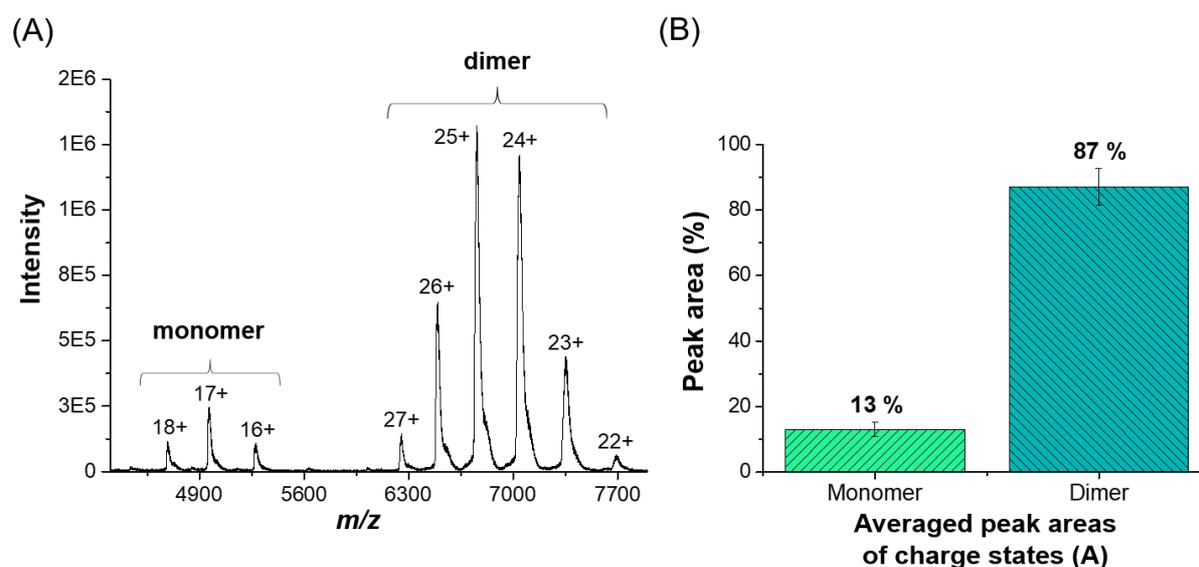


Fig. 47 Huge extent of dimerization at the first day of the auto-proteolysis test (dominant batch 1 monomer (-)aaL752 and homo-dimer 2x(-)aaL752).

(A) Native spectrum shows the distribution of monomer and dimer peaks. (B) Proportion of monomer and dimer, comparison is based on averaged areas of all charge states, S. Tab. 6. Peak areas are normalized to the dimer peak 25+. Measurements were performed in MS1 mode at 10 eV in 200 mM ammonium acetate, pH 7.2 + DTT, protein concentration was 7 μ M (monomer).

In contrast to batch 1, batch 2 data shows the co-existence of two prominent dimers with the same charge state distribution. Their occurrence is similar to that of the two main monomers of batch 2 (see 3.4). These dominant protein dimers were named as B2d(a) 165.6 kDa and B2d(c) 163.9 kDa. Based on their mass to charge distribution, it can already be shown that these are not exclusively dimers of the two main monomers. Instead, two lower intense dimer species (B2d(b) with 165.4 kDa and (d) with 163.9 kDa) were identified by MS1 and MS2 studies as homo-dimers of the dominate monomers (Fig. 48).

The lower intense dimer B2(b) (dark-green star) clearly emerges from the main peak signal (B2d(a) purple-colored star) at slightly higher energies in MS1 mode (approx. 100 eV) but shows a lower level of dimerization compared to the main dimers (a) and (c). Based on the

determination of the theoretical monomer mass (CM), it is suggested that B2d(b) is the homo-dimer of the monomer main monomer B2(a) (missing aa752-745, the C-terminal NS4A cofactor sequence).

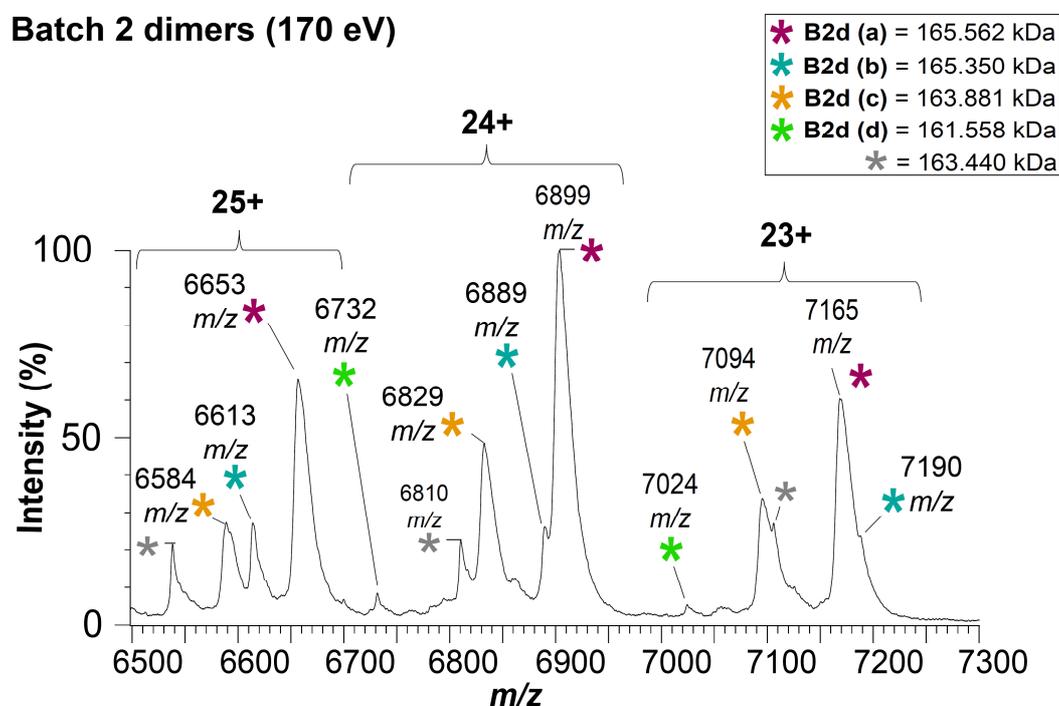


Fig. 48 Distribution of various dimer species in batch 2 after 1 week of auto-proteolysis.

The identified dimers are labelled with different coloured stars. Their MS-based mass is indicated. MS1 analyses were performed at 170 eV. Protein concentration was 8 μ M (monomer). Sample buffer was 200 mM ammonium acetate pH 7.2 and 2 mM DTT. Identified dimers are marked by different coloured stars (grey star = labels a further potential dimer based on the MS1 mass, but no related monomer was identified).

Based on CM a suitable monomer for the more dominant dimer B2d(c) could not be clearly identified by MS1 and MS2. Therefore, one assumption is that the monomer species, which forms B2d(c) is hidden under the other peaks and thus has not yet been identified. Since no corresponding monomer was found, based on the calculated monomer mass it has been assumed that the aa1-12 are missing at the N-terminal end (Tab. 15, S. Tab. 14). This is the best fit with the mass difference to the full aa-sequence of the monomer.

In contrast, based on their masses the most intense dimer B2d(a) could be assigned to the monomer B2(d) (82.7 kDa, missing aa752-746) and the very low intensity dimer B2d(d) was associated with the main monomer B2(b) (80.8 kDa, missing aa1-5 and aa752-745) (Tab. 14 and Tab. 15).

Tab. 15 Identification of four different dimers in B2 by increasing energy in MS1 and MS2 studies. CM = monomer masses are based on calculations of the dimer mass (= 1/2 the dimer), TM = theoretical monomer (aa-sequence based), TD = theoretical dimer, (*) = fits the real monomer B2(a), (***) = fits the real monomer B2(b), (***) = fits the real monomer B2(d) (82.781 kDa). The corresponding monomer to the second dominant dimer B2d(c) 163.881 kDa could not be identified. The calculated monomer mass CM = 81.941 kDa.

	TD (kDa)	TM (kDa)	B2d(a) (kDa)	B2d(a) CM (kDa)	B2d(b) (kDa)	B2d(b) CM (kDa)
Missing (aa) C-term.	167.042	83.521	165.562	82.781***	165.350	82.675*
Missing (aa) N-term.	–	–	–	–	–	aa ₇₅₂₋₇₄₅ (STA- ENALL)
				aa ₁₋₆ (MASHHH)	–	–
	B2d(c) (kDa)	B2d(c) CM (kDa)	B2d(d) (kDa)	B2d(d) CM (kDa)		
Missing (aa) C-term.	163.881	81.941 kDa	161.558	80.779**		
Missing (aa) N-term.	–	–	–	aa ₇₅₂₋₇₄₅ (STA- ENALL)		
	–	aa ₁₋₁₂ (MASHHH HHHHHH)	–	aa ₁₋₁₅ (MASHHH HHHHHH- HEN)		

3.4.1 MS2 control of the identified batch 2 dimers

As shown by MS1 (Fig. 48), the dimer signals of the proteins from batch 2 are less resolved compared to the signals of the corresponding monomers. Thus, deviations in the calculation of the masses between monomer and the corresponding homo-dimer occur in particular in the case of highly charged and broad peaks [153, 154]. MS1 data at 170 eV clearly show, that the signal of the main B2d(a) dimer (see 24+ peak, Fig. 48) is overlapped by the less abundant homo-dimer B2d(b).

However, this less pronounced dimer is important, since it was identified by MS1 as the corresponding dimer for the main monomer B2(a). To confirm its authenticity at softer energies, the 24+ peak was dissociated by MS2, while simultaneously monitoring the changes of the peak shape in MS1 (Fig. 49).

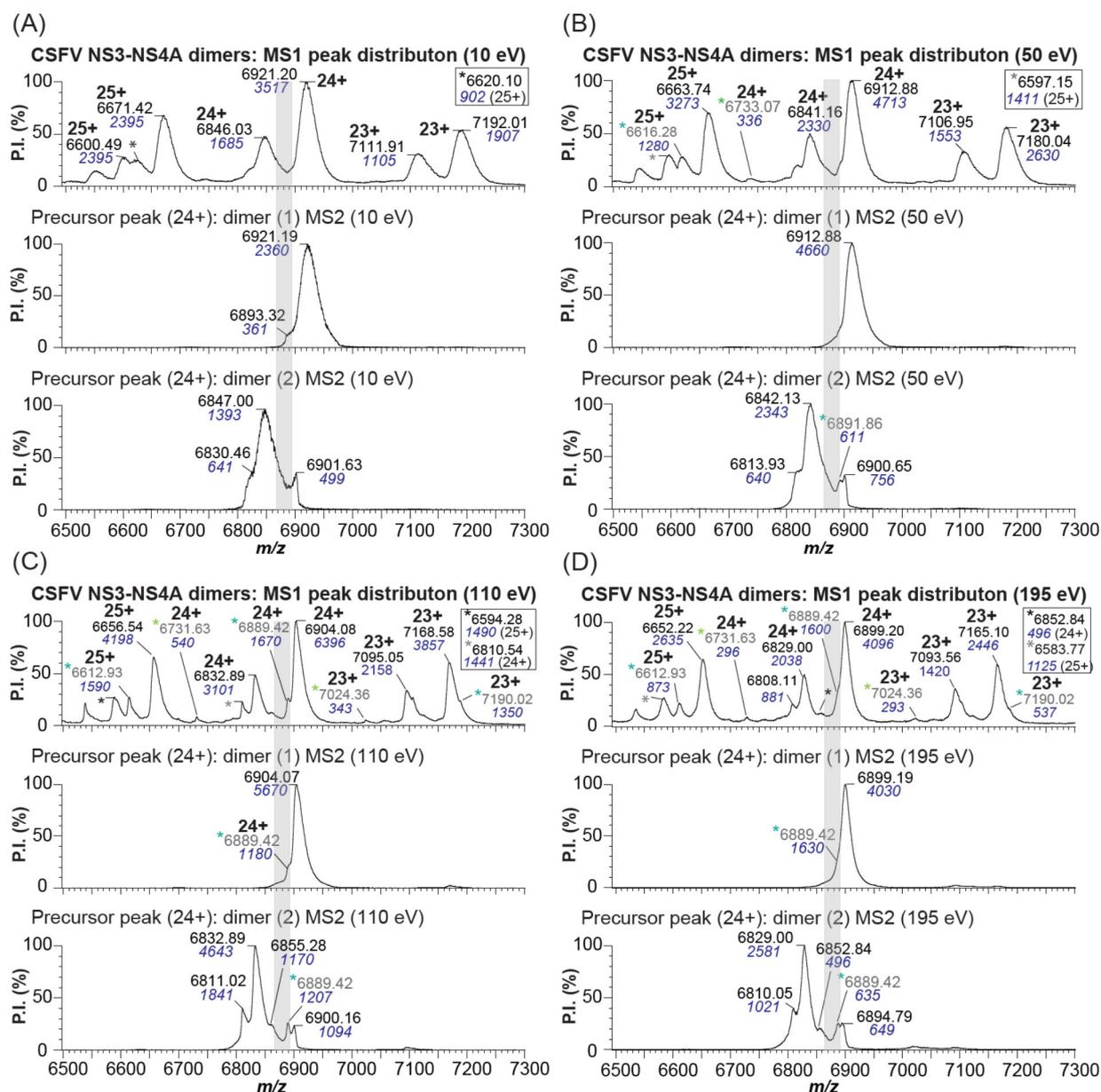


Fig. 49 Control of the authenticity of the B2 dimer peaks at different energies revealed four species. MS2 precursor = 24+ was, protein concentration = 8 μ M monomer in 200 mM ammonium acetate, pH 7.4, 2 mM DTT. (A) 10 eV MS1 and MS2 spectra show broad dimer peaks. (B) 50 eV MS1 and MS2 spectra show first additional peaks. (C) 110 eV MS1 and MS2 spectra shows the sharpest peaks. (D) 195 eV MS1 and MS2 spectra show the decrease in peak intensity and widening peaks. Grey bar = region of the location of the B2(b) dimer 165.350 kDa, (*) = B2d(b), (*) = B2d(b).

The peak shape at four different energy levels (10 eV, 50 eV, 110 eV, 195 eV) in MS1 mode was compared with the same energies in MS2-CID. The character of the peak form changed significantly with increasing energy. While the MS data at 10 eV still have very broad peaks in both MS1 and MS2 mode, the first additional dimer peaks appear at 50 eV (Fig. 49 (A)/(B)). At 110 eV the peaks are sharpest and further dimeric sub-species can be clearly recognized (Fig. 49 (C)). In addition, the previously identified B2d(b) peak in Fig. 48 and

Fig. 50 can now be clearly assigned as dimer of the B2(a) monomer, due to its occurrence at softer conditions. Although higher energies from 50 eV are not necessarily counted as a native MS measurement [153, 154, 206], a clear identification of further dimer species was not possible with softer energies. Moreover, other dimer peak series could be assigned to their associated monomers when applying 50 to 110 eV which were first identified in MS1 at 170 eV (Fig. 48).

However, associated monomers could not be found for all dimers. Thus, it seems reasonable to assume that not all monomers dimerize to the same extent. This is in line with the associated MS1 and MS2 data of the two main monomers B2(a) and B2(b). Both species dimerize less readily (Fig. 49, marked with green stars) compared to the sub-monomer B2(d), which was assigned to the highest intense dimer B2d(a).

To estimate the differences between the two main dimer species B2(a) and (c) in more detail MS2 analyses were performed (Fig. 50).

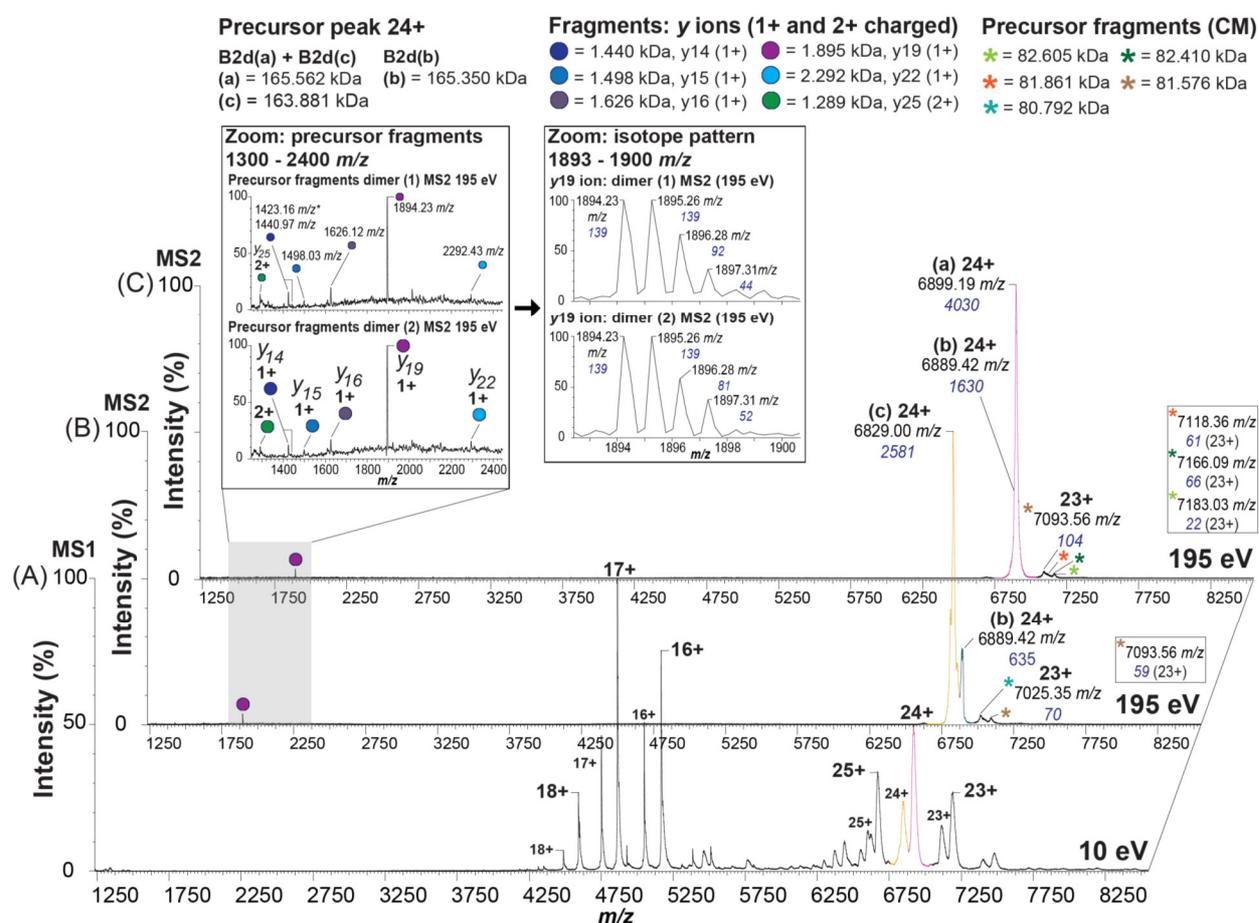


Fig. 50 MS2 dissociation of the 24+ precursor of the two main dimer species B2d(a) and B2d(c) show the same main fragment ions (y19 and y22) as the B2 monomers.

(A) Collision induced dissociation (CID) energy 10 eV. (B) CID energy 150 eV. (C) CID energy 195 eV. B2d(a) dimer (pink line), B2d(c) dimer (yellow line), B2d(b) dimer (dark-cyan line). Analyses were performed at energies between 10 eV to 195 eV. Peak intensity is labelled by blue italic numbers. Protein concentration was 8 μ M (monomer). Sample buffer was 200 mM ammonium acetate pH 7.2 and 2 mM DTT.

In contrast to the B2 monomers, both precursor peaks (24+) were hardly fragmented. Usually, one would expect monomers to be expelled from the dimer complex. Since this was not the case, it would be possible that Cys-bridges or Van-der-Waals forces prevent it. The dissociation of the two peaks only begins at high collision energies. The 24+ precursor peak of the dimer characterize a CID stable protein structure, which is to be expected in principle, since these large complexes experience less activation with gas molecules and can distribute more internal energy. However, the formation of these dimers can not only be based on cysteines, as the protein was fixed with 4 mM DTT immediately after SEC elution and

further reduced with DTT during buffer exchange into ammonium acetate solution. Interestingly, the increase in the CID energy (170 eV - 195 eV) led to the same C-terminal low mass fragment ions as it is the case for the MS2 data shown for the B2(a) and B2(b) monomers ($y_{19} = 1.894$ kDa, $y_{22} = 2.292$ kDa). MS2 data show that increased CID energies only fragments the exposed structures (N- and C-terminus) of the dimeric complexes.

A closer look into the low m/z -range also shows that there are less intense fragment ions, which could already be assigned to the B2(a) monomer (y_{14} to y_{16} , 1+ charged). However, the high mass precursor fragments obtained could not be clearly assigned to the smaller fragment ions.

As shown in the following, in contrast, the fragments with lower mass were also identified when higher MS1 energies were used.

3.4.2 Low mass ions in MS1 originate from the monomer and helicase and not from the dimer

In the following, the overall analyzes of the stability of the batch 2 monomers, dimers and helicase were carried out by MS1. Therefore, the enzyme was measured with different energy levels for ion activation. The ion activation energy was increased stepwise, comparable to the monomer studies (Fig. 44 and Fig. 45, shown in sub-section 3.4.1), from 10 eV to 195 eV (Fig. 51, Fig. 52).

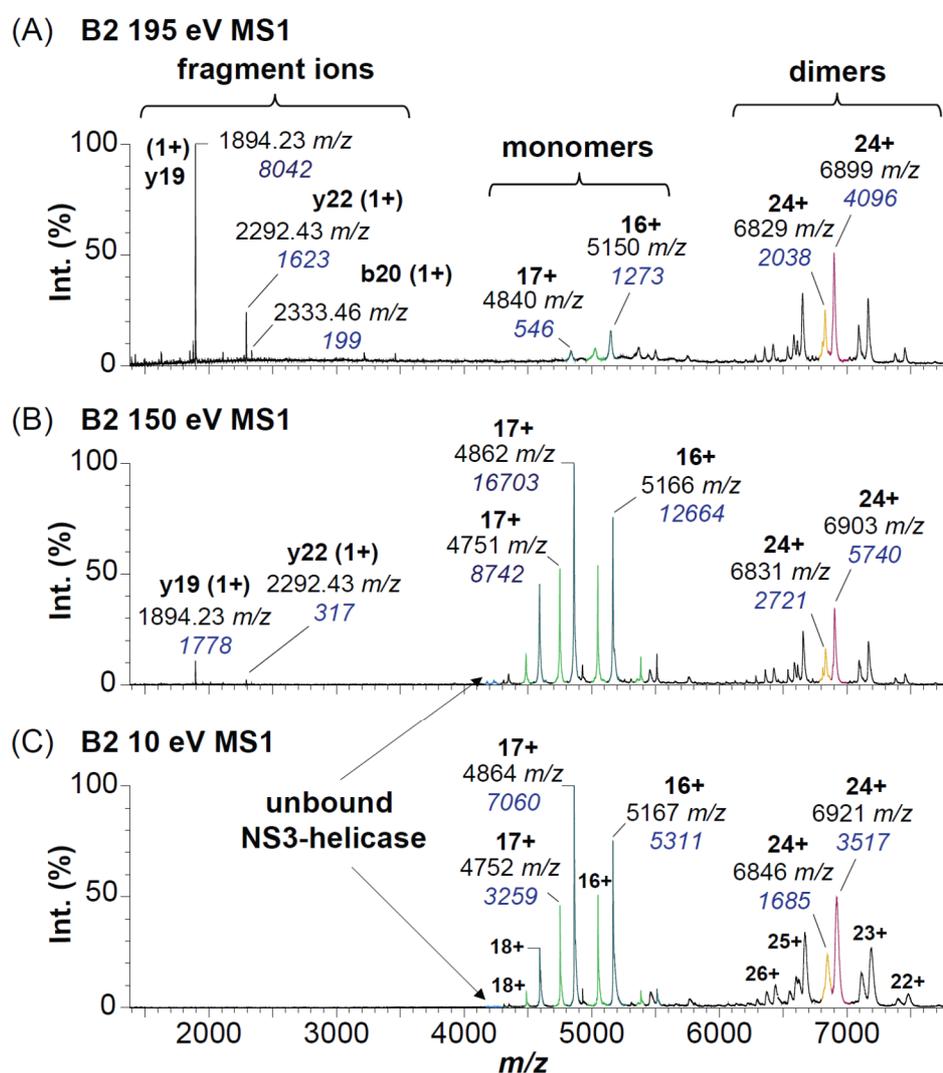


Fig. 51 Native MS studies show that an increase of ion activation energy in MS1 mode lead to the dissociation of the monomer and unbound NS3-helicase.

(A) Ion activation energy 10 eV. (B) Ion activation energy 150 eV. (C) Ion activation energy 195 eV. B2(a) monomer (dark-cyan line), B2(b) monomer (light-green line), B2d(a) dimer (pink line), B2d(c) dimer (yellow line), unbound helicase (blue line). Analyses were performed at energies between 10 eV to 195 eV. Peak intensity is labelled by blue italic numbers. Batch 2 protein concentration was 8 μ M (monomer). Sample buffer was 200 mM ammonium acetate pH 7.2 and 2 mM DTT.

The recorded spectra and the comparison of their respective peak areas show that both NS3 assemblies are very stable at higher energies but starts to degrade at 150 eV (Fig. 52).

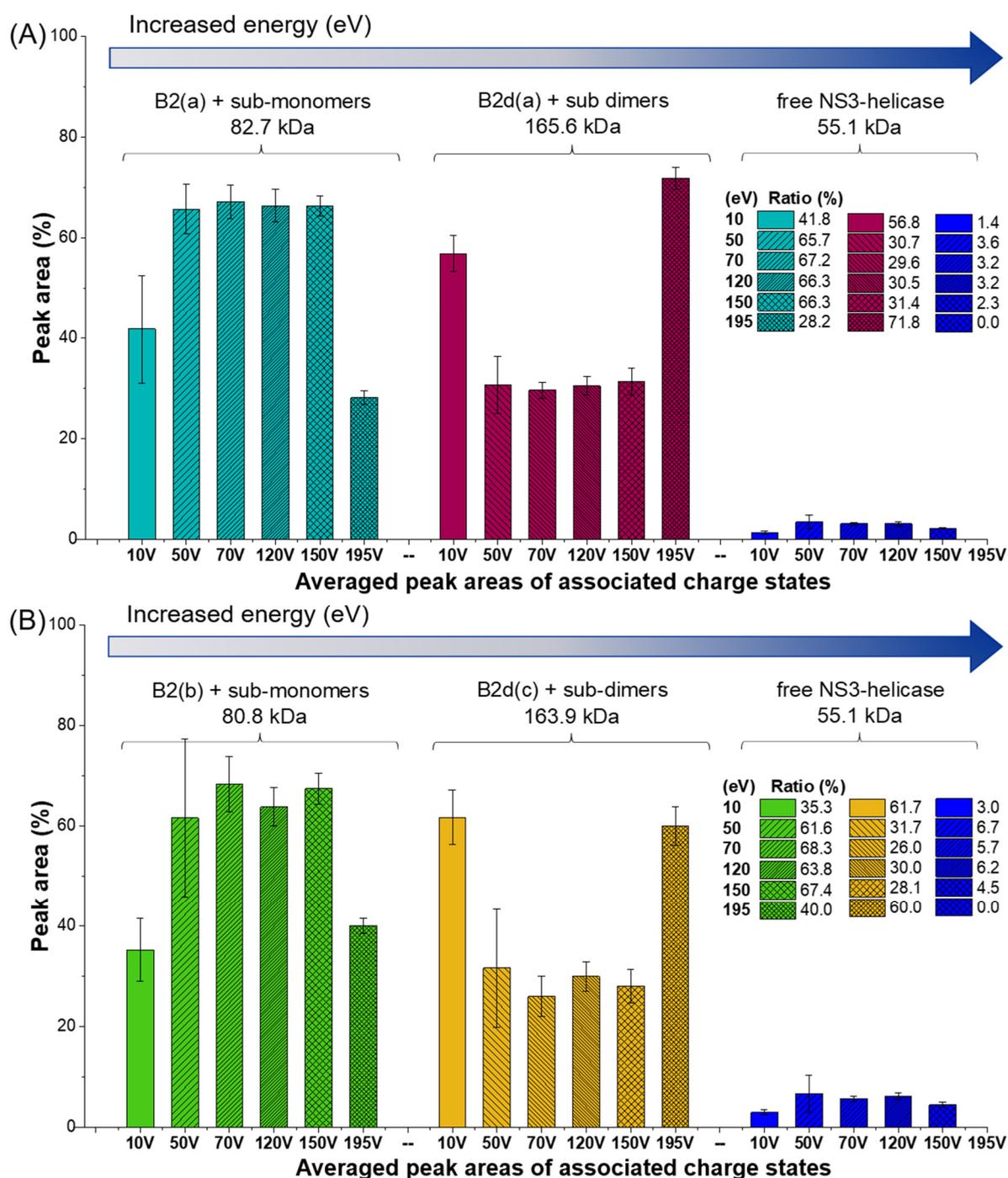


Fig. 52 Ratios at different MS1 energies show significant changes between main monomers and main dimers and reveal the dissociation of the unbound NS3-helicase.

(A) Averaged peak areas of charge states for the monomers: (B2(a) = 82.683 kDa) + sub-species B2(c),(d),(e),(f) in cyan; dimer: (B2d(a) = 165.562 kDa) + sub-species, in dark-green and free helicase part in dark-blue. (B) Averaged peak areas of charge states for the monomer: (B2(b) = 80.779 kDa) + sub-species B2(g),(h),(i),(j) in light-green; dimer: (B2d(c) = 163.881 kDa) + sub-species, in yellow and free helicase part in dark-blue. Associated charge states: monomer (18+ to 15+), dimer (26+ to 22+), helicase (14+ to 13+). Protein concentration = 8 μ M (monomer). Sample buffer with 200 mM ammonium acetate pH 7.2 and 2 mM DTT, energy: 10 eV to 195 eV (V = eV).

The comparison is based on the main monomers B2(a), (b) and their sub-monomers as well as the main dimers B2d(a), (c) and their respective sub-species and a small amount of the unbound NS3-helicase. Initially, the proportion of both monomers increases significantly from 10 eV to 50 eV and each makes up over 60% of the area. This holds up to 150 eV and then drops radically to 28% (B2 (a)) and 40% (b) at 195 eV (Fig. 52). The dimers behave in opposite ways. At 10 eV, their area makes up around 60% of the total area. At higher energies, it decreases to approx. 30% in relation to the extremely increasing surface area of the monomers. However, as soon as the area of the monomers diminishes, the area of the dimers increases proportionally. The MS2 data as well as the associated MS1 spectra confirm that the dimers are much more stable at higher energies (Fig. 50, Fig. 51).

However, the higher the energy used, the more the ratio of monomer to dimer changes (Fig. 52). Furthermore, at energies between 120 eV and 150 eV, C-terminal fragment ions (primarily y19 and y22 (+1)) were observed (S. Fig. 7). Their intensity increases even more at 150 eV (Fig. 51 (B)).

If the energy is increased further to 195 eV, additional C- and N-terminal fragments are detected. At this point the monomers dissociate faster, which is evident from their loss of intensity and peak area (Fig. 52, Fig. 51 (A), blue italic numbers). Noticeable is the appearance of the b20 ion, which already identified the B2(b) monomer in MS2 studies.

The dimers, on the other hand, hardly show any reduction in intensity at higher energies. However, it is obvious that they show a shift in their mass, which is associated with fragmentation events. If looking at the main peaks, from 10 eV (B2d(a) = 6921 m/z (24+)) to 195 eV ((a) = 6899 m/z (24+), pink line) there is a difference of 22 m/z . This corresponds to a mass loss of approx. 528 Da. From 10 eV to 150 eV the loss of mass is a bit lower (432 Da). For the second main monomer B2d(c) (yellow line) there is a slight difference. From 10 eV to 195 eV the loss is 408 Da and from 10 eV to 150 eV 360 Da. In contrast, the m/z -shift of monomers is much lower. From 10 eV to 150 eV it is 2 m/z = 34 Da for B2(a) (dark-cyan line) and for B2(b) (light-green line) it is 1 m/z = 17 Da, which could be the loss of NH₃. On the other hand, the monomer shift also increases up to an energy of 195 eV. For example, for the main monomer B2(a) there is a shift of 24 m/z = 408 Da (4864 m/z - 4840 m/z , (17+)). In the case of the monomers, this higher mass difference

can no longer be attributed exclusively to a simple loss of small ions. In line with the significant decrease in intensity, it can be assumed that larger fragments (y19 and y22 ions) were cleaved off at the C-terminal end.

For the dimer, however, this has not been clearly demonstrated. Due to their higher surface area, the dimers show more solvent exposed basic residues [114, 138], which could pick up more charges. Assumable, more solvent residuals from the sample solution are attached to them. While, the competition with ammonium acetate can prevent the formation and protein binding of other salt adducts [209, 210]. Thus, the measured mass of such large complexes is higher than expected by the corresponding aa-sequences of their components, which means that they have an artificially higher mass at low energies (10 eV) [153]. In contrast to the smaller monomer, the mass loss at high energies applied to the NS3-NS4A dimers is therefore more likely caused by the loss of NH_3 .

Furthermore, a small amount of unbound NS3-helicase subunit was found, which makes up about 1.4% to 3.0% of the total peak areas. The free helicase subunit was only detected at energies from 10 eV to 150 eV. At 195 eV, an associated MS1 signal could no longer be detected. Hence it is suggested that the subunit is destroyed at high MS1 energies. Thereby, it could also increase the signal of the y19 and y22 ions. Thus, these fragment ions presumably originate from both the monomer and the helicase subunit. When comparing monomer B2(a) and dimer B2d(a), it is noticeable that the intensity of the monomer decreases significantly with increasing energy up to 195 eV, comparable to the helicase subunit (Fig. 51 (A), Fig. 52). The dimer B2d(a) behaves contrary to the monomer and shows a pronounced signal at 195 eV in relation to the monomer. The same can be observed for the ratio of B2(b) monomer compared to the B2d(c) dimer. In addition, greater intensity fluctuations occur with increasing energy for both (Fig. 52 (B)).

3.4.3 Dimer formation is supported not only by Cys-bridges

The reduction of the enzyme using DTT revealed two oligomeric states, the protein monomer and dimer. However, it is possible that certain Cys-bridges have been preserved, which is assumed based on the MS2 data shown above. These presumably support the formation of the dimers. On closer inspection of the MS data, however, it becomes apparent that this cannot be the main factor for the pronounced dimerization (Fig. 53).

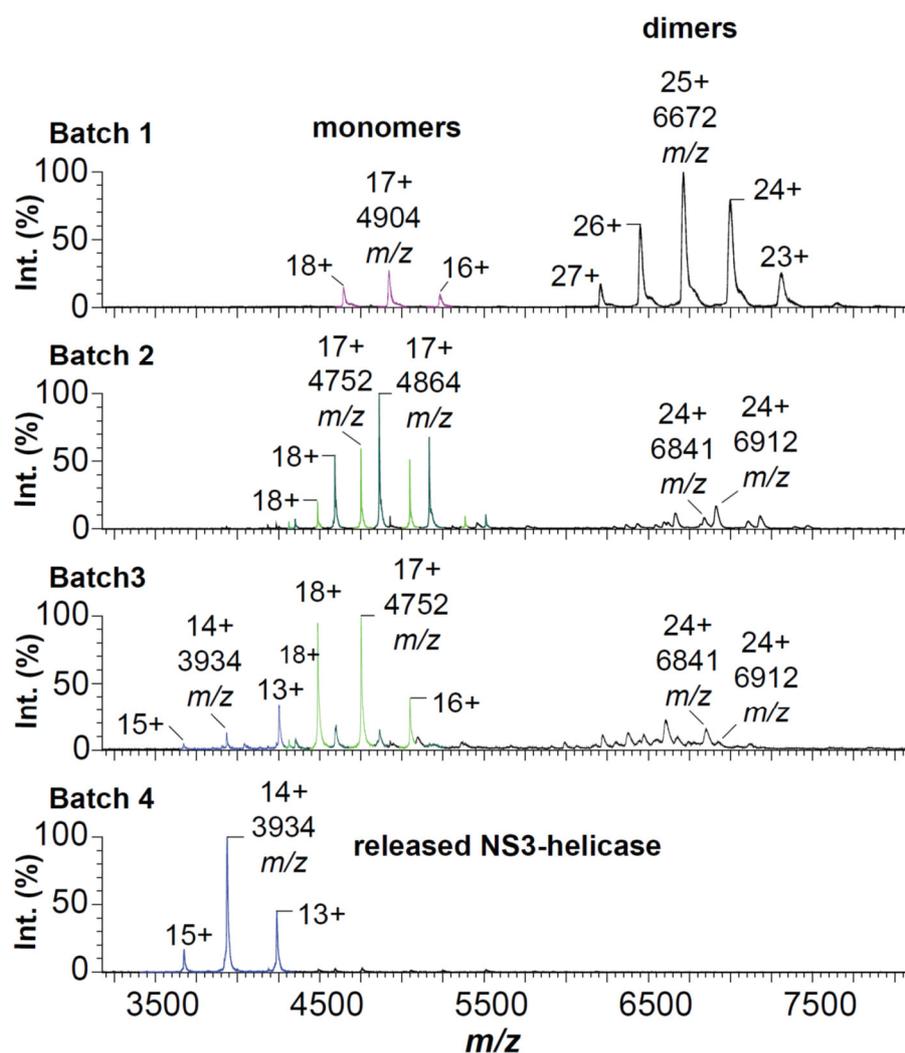


Fig. 53 The auto-proteolytical activity of the NS3-protease including protein monomers and dimers leads to the release of the NS3-helicase subunit.

(Batch 1): MS1 identification of the NS3-NS4A monomer species B1(a), protein monomer = 83.374 kDa, purple line and dimer B1d(a) = 166.8 kDa, (protein storage for 1 day on ice while measuring). (Batch 2): spectrum of two identified monomers B2(a) = 82.683 kDa, dark-cyan line, B2(b) = 80.779 kDa, light-green line and the dimers B2d(a) = 165.9 kDa and B2d(c) = 164.2 kDa, (protein storage for 1 week at + 4 °C before measuring). (Batch 3): unbound NS3-helicase subunit B3(a) = 55.1 kDa, blue line and the B2/3(b), light-green and the B2/3(a), dark-cyan line, (protein storage for 2 weeks at + 4 °C). (Batch 4): spectrum of the released helicase B4(a)/B3(a) = 55.1 kDa, blue line, (protein storage for 3 weeks at + 4 °C). Peak intensity labelled by blue italic numbers. MS1 energy: 10 eV and 50 eV, ammonium acetate 200 mM + DTT at pH7.2, protein concentration: 6 to 8 μ M.

In Fig. 53 the data impressively show that the dimer, which is initially more dominant than the monomer (batch 1), is completely degraded over time. In batch 2, the dimer content is decreased, whereby the bigger dimer is more obviously present (6912 m/z , 24+). After a longer proteolysis time at + 4 °C (batch 3) the bigger dimer is more easily degraded. Whereas the small one (6841 m/z , 24+) remains longer. Along with the release of the NS3-helicase, both the monomer and dimer are completely proteolytically degraded after three weeks at + 4 °C. These results also confirm that the measured dimers must be homo-dimers of the already identified monomers. It is already noticeable that the dimer formation between batch 1 and batch 2 is different. In batch 1 there is only one major dimer, whereas in batch 2 two main species were identified. (Fig. 53, Fig. 54, S. Tab. 15). MS data clearly show that the release of the helicase goes hand in hand with the loss of complex dimerization, while the monomers initially remain stable. Since hardly any helicase released was detected in batch 1, a high level of dimer formation can be observed here. This indicates that batch 1 exhibits the lowest amount of auto-proteolysis as can be seen in (Fig. 54).

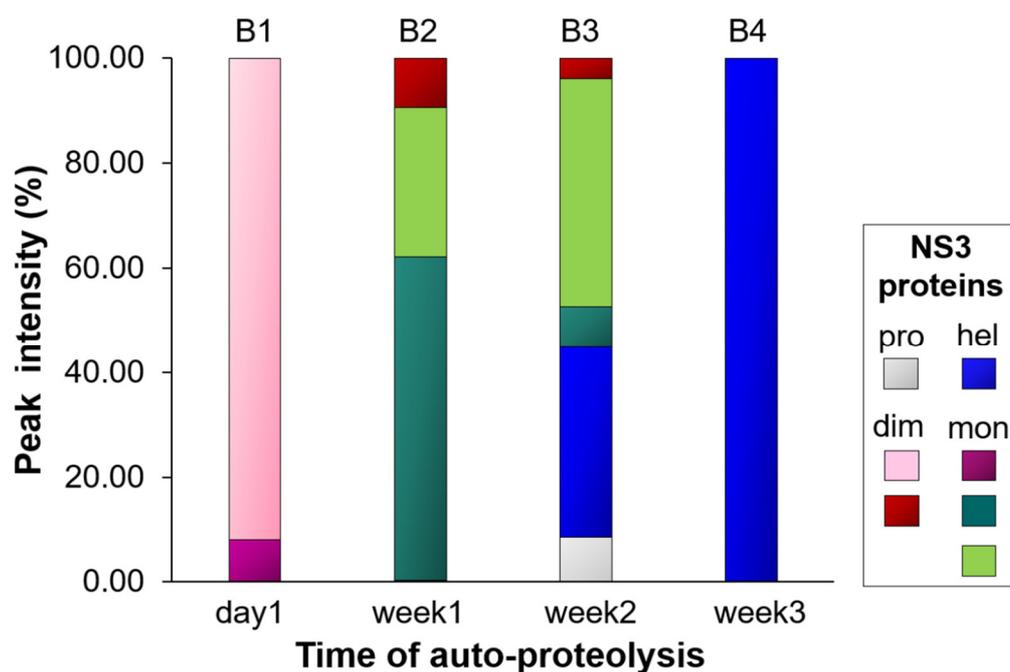


Fig. 54 Course of the auto-proteolysis process of the NS3 complex: decrease and increase of the dominant NS3 proteins is shown over three weeks.

(B1) Distribution of the main protein species of batch 1: B1(a) monomer and B1d(a) dimer (marked in violet 8% and pink 92%). (B2) batch 2: small amount of released helicase, B2(a)+B2(b) monomer and B2d(a) dimer (marked in blue 0.4%, dark-green 61.7%, light-green 28.5% and red 9.4%). (B3) batch 3: small amount of released protease, B3(a) released helicase, B2/3(a)+B32/(b) monomer and B2/3d(a) dimer (marked in grey 8.5%, blue 36.3 %, dark-green 7.8%, light-green 43.4% and red, 4.0%). (B4) batch 4: B4/3(a) released helicase 100%. Abbreviation: pro = released NS3-protease, hel = released NS3-helicase, dim = complex dimer, mon = complex monomer. Data are based on associated peak intensities. Protein concentration: 6 to 8 μM (monomer). Conditions: MS1 mode at 10 eV, 200 mM ammonium acetate, pH 7.2 + DTT.

The data plot of the auto-proteolysis reaction over the course of three weeks (Fig. 54) show on the first day of the analysis (batch 1) that there are still approx. 92% dimer in addition to 8% monomer, in batch 2 it is only 9% dimer after one week of auto-proteolysis and after two weeks (batch 3) only 4% dimer remain. It is also noticeable that an increase in the smaller monomer (B2(b) = B3(b), missing aa1-15, aa752-745) can be observed, whereas the larger monomer (B2(a) = B3(a) missing aa752-745) is also degraded comparable to the dimer. These results therefore also support the assumption that the NS3-protease proteolytically degrades the proteins present in the direction of the smaller monomer species that misses aa1-15 and aa752-745. Accordingly, the data also show that the released helicase must first originate from both the dimer and the larger monomer that only misses aa752-745. After three weeks (batch 4) the NS3-NS4A protein complex is finally almost completely broken down. In addition to the released NS3-helicase, there are hardly any other proteins present. It can be assumed that the helicase subunit is a major supporting factor for the dimerization of the intact protein complex. A possible course for the degradation process would be that the auto-proteolytic cleavage starts first between the protease and helicase subunit in the homo-dimer leading to a rapid decrease in dimer quantity. This would result in an increase of free protease leading to a more rapid cleavage of the monomer starting between helicase subunit and NS4A-Cofactor followed by a cleavage of helicase and protease subunit.

For example, this is shown in the case of the batch 2 and 3 proteins, where degradation has progressed significantly. A multitude of different monomeric subspecies has formed here as shown in the chapter above. The pattern of dimer formation behaves accordingly, creating different dimers. These results also reinforce the assumption that dimer formation plays an important role in the functionality of the NS3-NS4A enzyme complex.

However, no dimers of the released helicase were observed. Thus, it is possible that dimer formation does not play an important functional role in the released subunit itself.

However, several studies have suggested dimerization of the NS3-helicase domain of HCV and CSFV under experimental conditions, but so far the real *in vivo* function is unclear [102, 172, 193]. It is assumed to be beneficial for the enzymatic activity to build a further active protein conformation. In contrast to SAXS data from other studies [102], the obtained native MS data could indicate that the NS3-NS4A enzyme based on the complete aa-sequence also dimerizes independently *in vitro* under “near-native” experimental conditions.

3.5 MS studies of mutated CSFV NS3-NS4A proteins reveal structural differences

The following part of the thesis presents the results of the native MS analyses of two mutated NS3-NS4A proteins of CSFV (Fig. 55). Although this study provides a good preliminary characterization of the examined mutants, more data is needed to fully characterize them, which is well beyond the scope of this thesis. The data demonstrates diverse structural dynamics and variations compared to the wild-type (WT) protein. Based on several studies on the HCV NS3 protein it is known to require the interaction with the NS4A cofactor for complete folding and enzyme activity as well as for stabilizing by anchoring the complex to cellular membranes [47, 85, 86, 104, 173]. Due to the structural similarities between HCV and CSFV NS4A, both cofactors are assumed to have the same impact on the NS3 protein function [55, 89].

As labelled in the aa-sequence in Fig. 17 (B) (yellow rectangle, chapter 3.1), both mutations are located close to the N-terminus of the NS3 protein, which is associated with the kink-region of the NS4A cofactor. It was already shown that mutations on this surface region of the NS4A protein play an important regulatory role for the protein-protein interaction between NS3 and NS4A cofactor. It could be shown that the contact between NS3 and the NS4A-kink region is essential for viral RNA replication. Whereas a more open conformation (with only little interaction) of the complex seems to be an important prerequisite for the formation of individual infectious virus particles (virions) [103].

Thus, Dubrau *et al.* 2017 [103] assumed that the kink-mutations should lead to a more exposed NS4A cofactor and thus to a more open protein conformation. Based on the assumptions made by Dubrau and colleges, the mutations were introduced into the WT aa-sequence. In comparison with the WT enzyme native MS data of the mutated proteins demonstrated various dynamics which could be related to the more flexible N-terminal NS4A cofactor.

The amino acids L45 and Y47 were exchanged by A45 and A47. For the first mutation, Leu is replaced by A and is named single mutant (SM) L45A. The second mutation, through the exchange of two amino acids, is termed double mutant (DM) L45A/Y47A (Fig. 55 (B)/(C)).

As can be seen in Fig. 55, both mutated NS3 proteins on the first day of the MS analysis show major differences compared to the wild-type protein (WT), which was introduced in chapter 3.1 to 3.4. The mutants appear to be structurally very labile (Fig. 55 (B)/(C)).

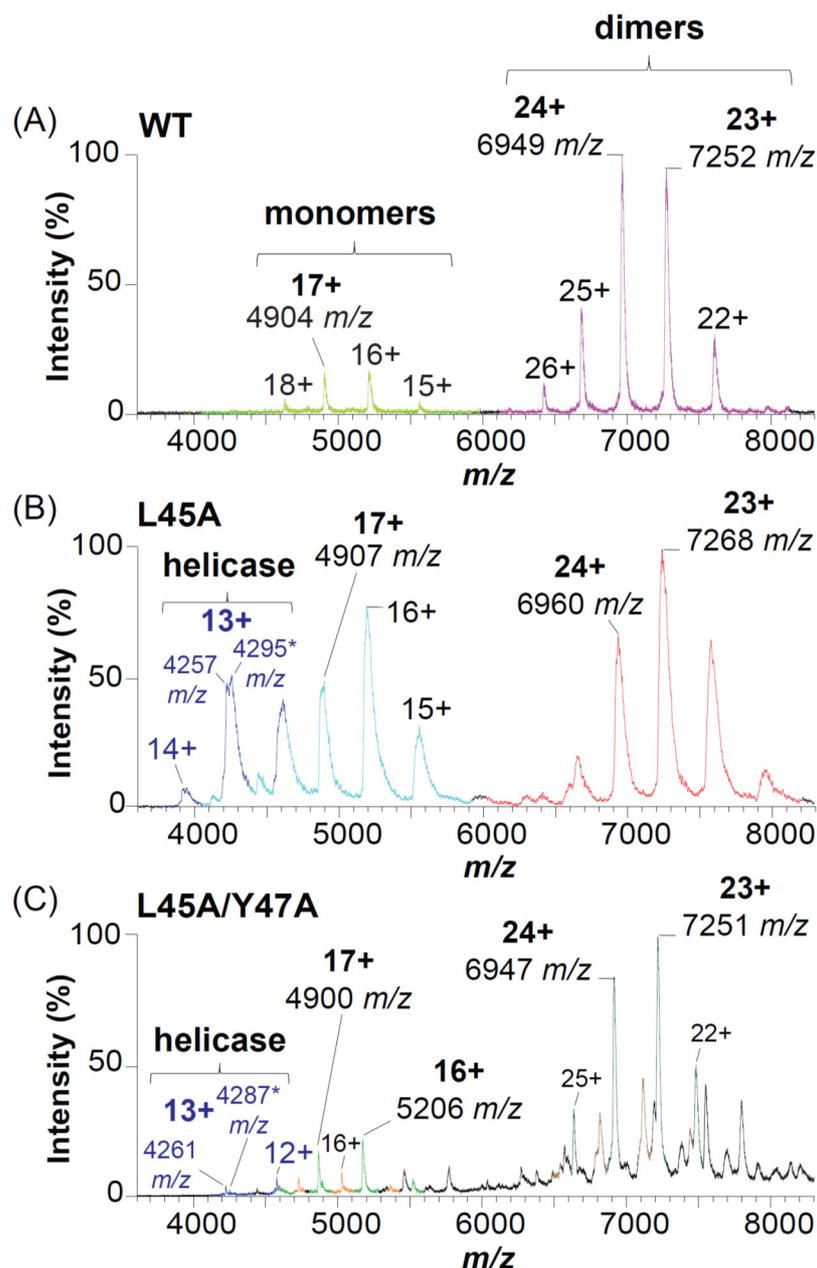


Fig. 55 Native MS shows structural differences between the mutated and WT NS3-NS4A protein. (A) Wild-type (WT) protein, dominant monomer B1(a) = 83.374 kDa (light-green-coloured line), dominant dimer B1d(a) = 166.792 kDa (pink-coloured line). (B) Single mutant L45A, identified monomer SM(a) = 83.419 kDa (light-blue line), dimer SMd(a) = 167.118 kDa* (red line), NS3-helicase = 55.3 kDa and 55.8 kDa* (blue-coloured line). (C) Double mutant L45A/Y47A, main intense monomers DM(a) = 83.302 kDa (green-coloured line), DM(b) = 81.382 kDa (orange-coloured line), main intense dimers DMd(a) = 166.751 kDa (dark-green-coloured line), DMd(b) = 164.390 kDa (brown-coloured line). Free NS3-helicase = 55.3 kDa and 55.8 kDa* (blue-coloured line). Protein concentration = 5 to 7 μ M (monomer) in 200 mM ammonium acetate + DTT, pH 7.4, measurements: positive ion mode at 10 eV energy.

L45A and L45A/Y47A already have a higher amount of unbound helicase subunit that also shows double peaks with slightly higher masses compared to the WT protein. For the L45A protein the relatively wide and less resolved peaks hinder the calculation of the exact mass. This could also indicate that a small amount of NS4A is still attached to the C-terminal end of the protein or that different internal cleavage sites were involved compared to the WT protein. Based on sufficient MS2 data (Fig. 56, Tab. 16 and S. Fig. 11) three dominant singly charged y -ions were identified ($y_{19} = 1911.12$ Da, $y_{27} = 2692.72$ Da, $y_{30} = 3090.56$), which showed the 55.8 kDa* protein as NS3-helicase in association with the complete C-terminal NS4A cofactor. The double mutant shows two different dominant monomers DM(a) and DM(b) as well as a smaller proportion of free NS3-helicase (Fig. 59).

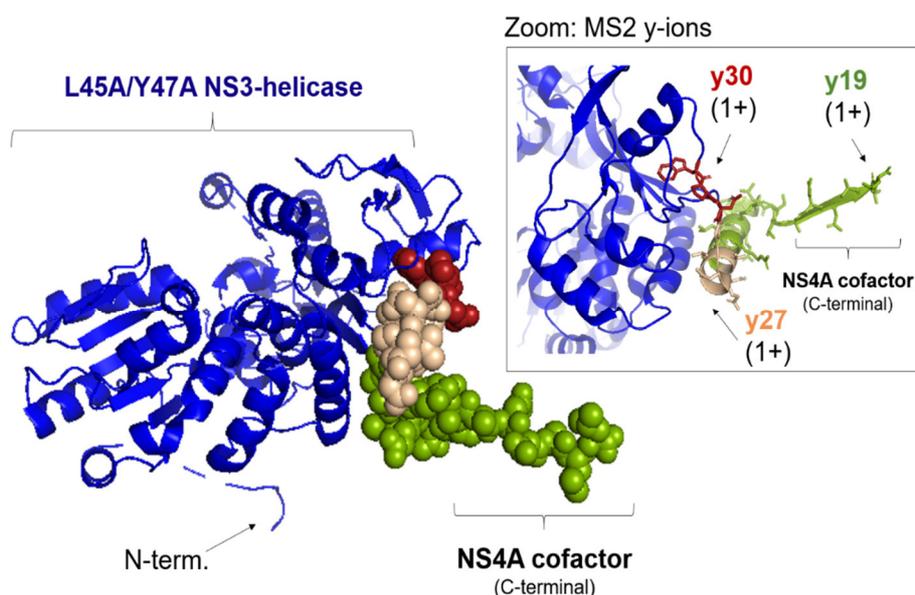


Fig. 56 Crystal structure of the DM L45A/Y47A NS3-helicase (55.8 kDa) revealed the interaction with the complete C-terminal NS4A-cofactor.

MS2 y -ions are labelled with different colours (y_{19} = green sticks/sphere, y_{27} = sand-coloured sticks/sphere, y_{30} = ruby-coloured sticks/sphere). Structure is designed based on PDB: 5ZM4.

The identified monomer species dimerize, which indicates that the helicase is involved in that process similar to the WT protein. The double mutant also forms two main dimers (DMd(a) and DMd(b)), which indicate similar distribution to the WT dimers that were identified in batch 2. Both mutants, L45A and L45A/Y47A are characterized by distinct peak and charge state distributions in MS1 investigations.

Tab. 16 MS2 ions that were identified while dissociating the NS3-helicase of CSFV L45A/Y47A after three weeks + 4 °C.

Ions are based on 13+ precursor peak (4287 m/z). TP = theoretical peptide (m/z), z = charge, C-terminal ions = y, aa = amino acid sequence of the MS2 ions, CID energy = 195 eV. The purple-coloured amino acids of the sequence mark the C-terminal part of the NS4A cofactor (NS4A(8aa)), which was attached to the NS3-helicase subunit to stabilize the protein complex.

Ion	z	m/z	TP	Peptide aa-sequence
y19	1+	1911.12	1911.10	AGRALKQVVGLSTAENALL
y27	1+	2692.72	2693.46	PGNQGTVEAGRALKQVVGLSTAENALL
y30	1+	3090.56	3091.62	WPDPGNQGTVEAGRALKQVVGLSTAENALL

3.5.1 The single mutant shows a high amount of unbound NS3-helicase

The L45A mutation is located at the N-terminal part of the NS4A cofactor at the beginning of the α -helical-loop (Fig. 57). Due to the exchange of the non-polar amino acid L (131 Da) by A (89 Da), the theoretical mass (TM) of the protein changes (TM WT = 83.521 kDa to TM SM = 83.479 kDa, monoisotopic mass), resulting in the mutant being slightly larger than the wild-type.

Three different dominant protein forms have been identified for the single mutant (SM) as shown in Fig. 57. This includes the fusion protein monomer SM(a), which is labelled with light-blue charge states (19+ to 15+) and has a mass of 83.419 kDa (Fig. 57 (A)). The identified protein dimer SMd(a) has a mass of approx. 167.118 kDa that is associated with a charge state distribution of 25+ to 21+ and labelled in red (Fig. 57 (A)). Finally, an unbound/released unit of the fusion protein could be identified as the NS3-helicase. It has a mass of approx. 55.3 kDa to 55.8 kDa and a charge state distribution of 14+ to 12+, which is labelled in dark-blue (Fig. 57 (A)). The location of the amino acid exchange (L45A) is illustrated in the crystal structure in Fig. 57 (B).

In the following, the focus will be on the comparison of the ratio of monomer and dimer as shown in Fig. 57 (C)/(D). The L45A mutant shows a dimerization of 64% compared to the monomer 36% (Fig. 57 (D)). The degree of dimerization is lower compared to the WT protein from the first day of the auto-proteolysis test (86%), the percentage of monomer is doubled in the mutant compared to the WT (14%), as shown in sub-chapter 3.4.2 and in Fig. 47. This changes when comparing the monomer B2(a) from batch 2 and the main

dimer B2d(a) with this mutant. At 10 eV approx. 42% monomer and 57% dimer were observed (Fig. 52), which is of a similar order of magnitude as for the SM proteins. This suggests that freshly thawed SM proteins are in a similar state of proteolysis on the first day as the WT protein after one week of proteolysis test. The lower dimer formation could also be an indication that the dimerization is influenced not only by the helicase but also by the NS4A cofactor and thus by a structurally stable protease. Consequently, the exchange of amino acids in the NS4A kink region seems to limit dimer formation.

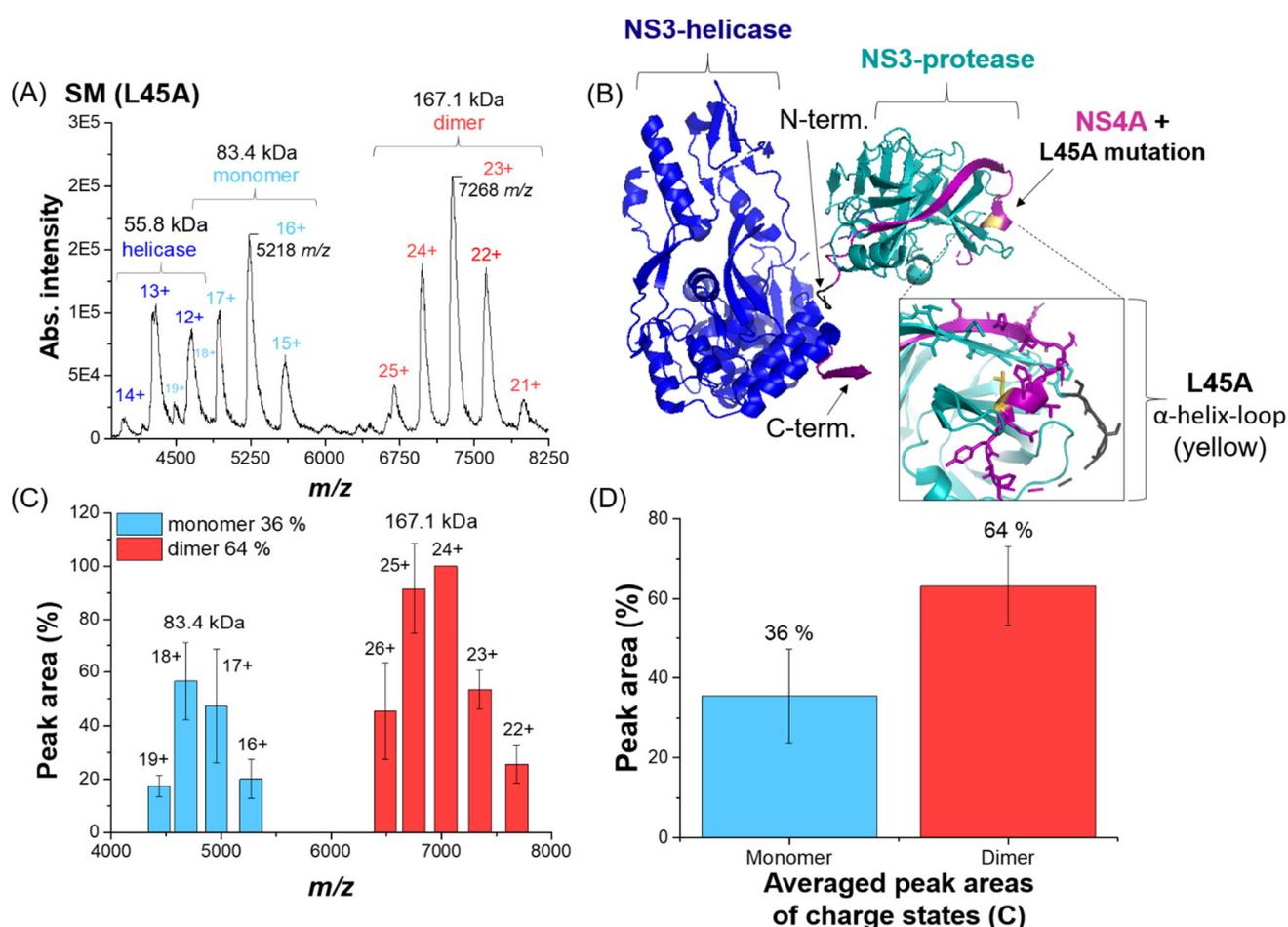


Fig. 57 L45A exhibits a lower degree of dimerization.

(A) Native spectrum of the L45A mutant. (B) Crystal structure of the protein monomer based on PDB: 5MZ4, [103]. Location of the mutation (α -helix-loop) at the NS4A cofactor (purple) is marked in yellow, NS3-protease (cyan), GSGS-linker (grey sticks), NS3-helicase (blue). The zoom shows the aa-backbone (L45 = yellow, GSGS-linker = grey). (C) Distribution of charge states normalised to the 24+ peak of the protein dimer (red), monomer (light-blue). (D) Determination of associated peak areas for the monomer and dimer. Calculations are based on triplicated measurements and on averaged peak areas. Data are based on S. Tab. 17 and S. Tab. 18. Protein concentration = 7 μ M (monomer) in 200 mM ammonium acetate, pH 7.4 + 0.5 mM DTT, measurements: positive ion mode at 10 eV energy.

It is possible that this protein region not only plays an important role in the complex interaction on monomer level, but also promotes the formation of dimers. It could therefore be assumed that the exchange of similar amino acids (Leu to Ala) leads to a more flexible N-terminal cofactor.

In the following, the free helicase subunit was examined in more detail and included into ratio determinations. The entire distribution of peak areas, including helicase, monomer and dimer of the fusion protein was set to 100%. It turned out that the proportion of free helicase is quite high at 48%. In contrast, the wild-type only started to show free helicase in batch 2 after a longer time of storing at + 4 °C. The monomer (19%) has about half the amount of dimer (33%) as shown in Fig. 58.

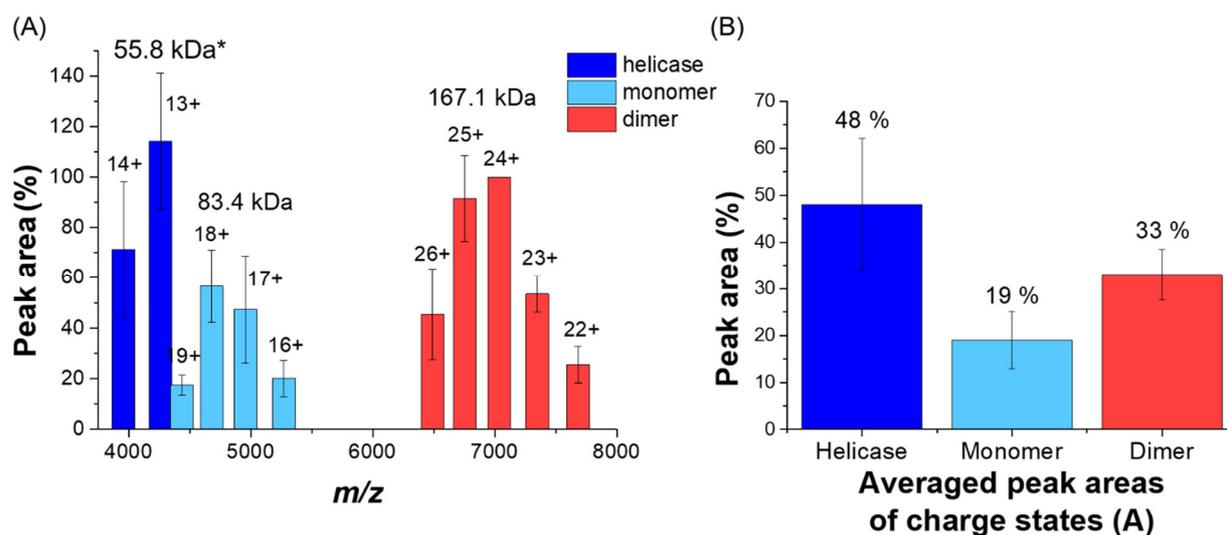


Fig. 58 L45A reveals a huge amount of unbound NS3-helicase subunit.

(A) Distribution of charge states normalised to the 24+ peak of the NS3-4A fusion protein dimer (red), monomer (light-blue), free NS3-helicase (blue). (B) Determination of associated peak areas for the monomer and dimer. Calculations are based on triplicated measurements and on averaged peak areas. Data are based on S. Tab. 19 and S. Tab. 20. Protein concentration = 7 μ M (monomer) in 200 mM ammonium acetate, pH 7.4 + 0.5 mM DTI, measurements: positive ion mode at 10 eV energy.

The increased amount of released helicase subunit is remarkable. The MS data confirm that the NS4A cofactor must play a very important role in the overall stability of the protein complex.

3.5.2 The double mutant shows a high level of auto-proteolytic degradation

The additional Y47A mutation is located close to the above-described single mutation L45A. The exchange of the two amino acids (Leu to Ala and Tyr to Ala) leads to a higher mass shift of the protein complex. Compared to the wild-type protein, which has a theoretical mass of 83.521 kDa, the theoretical aa-sequence-based monomer mass of the double mutant (DM) is 83.387 kDa. Additionally, the replacement of the polar Tyr with the non-polar Ala should have a stronger impact on the protein stability, which was also possible to monitor by native MS in this thesis. Due to its aromatic ring system, tyrosine has a stabilizing effect on the protein structure [211-213], as it can delocalize additional electrons, especially at higher CID energies [214-216].

Thus, aromatic rings stabilize the native protein structure through hydrogen bonds and electrostatic interaction better than amino acids without an aromatic ring system. The crystal structure of the NS-protein complex also showed that histidine and proline (His44, P49), which are located in the aa-sequence of NS4A are structurally in close contact with tyrosine (Y47) (Fig. 61). The interaction of the two aromatic amino acids strengthens the protein structure of the cofactor and ensures stability of it. As it was assumed, the kink-mutations seemed to lead to an increased flexibility of the NS4 cofactor, which could lead to a decrease of the NS3-protease activity. The native MS data show a faster dissociation of the dimer compared to the WT proteins (Fig. 59). One hypothesis is that the mutation could result in a more open protein conformation and different protein structures. On the other hand, due to the mutation, only the peptide binding might be locally weaker, without structural changes occurring in the entire protein complex. However, this cannot be clarified on the basis of the native MS data. For further analysis of such influences on the protein structure, an HDX-MS based approach would be useful. The presence of DM(b) (81.382 kDa) and its huge mass difference to the theoretical monomer mass (83.387 kDa) of over 2000 Da is remarkable but due to the lack of sufficient MS2 data it is currently not fully understood. This monomer is possibly a cleavage product of the NS3-protease, which is based on non-specifically cleavage at the C-terminus within the helicase subunit. However, this is contradicted by the fact that the sample contains free, structurally intact helicase. Possibly there is rather a deficit of a few C-terminal (aa752-750 315 Da, ALL) and several N-terminal amino acids (aa1-13 before 1700 Da, which corresponds to the start sequence MAS +

10xHis). The two dominant monomer peak series were named as DM(a) (83.3 kDa, light-green line) and DM(b) (81.4 kDa, orange line). The associated dimers DMd(a) (166.8 kDa, dark-green line) and DMd(b) (163.2 kDa, brown line) were also identified based on MS analyses (Fig. 59).

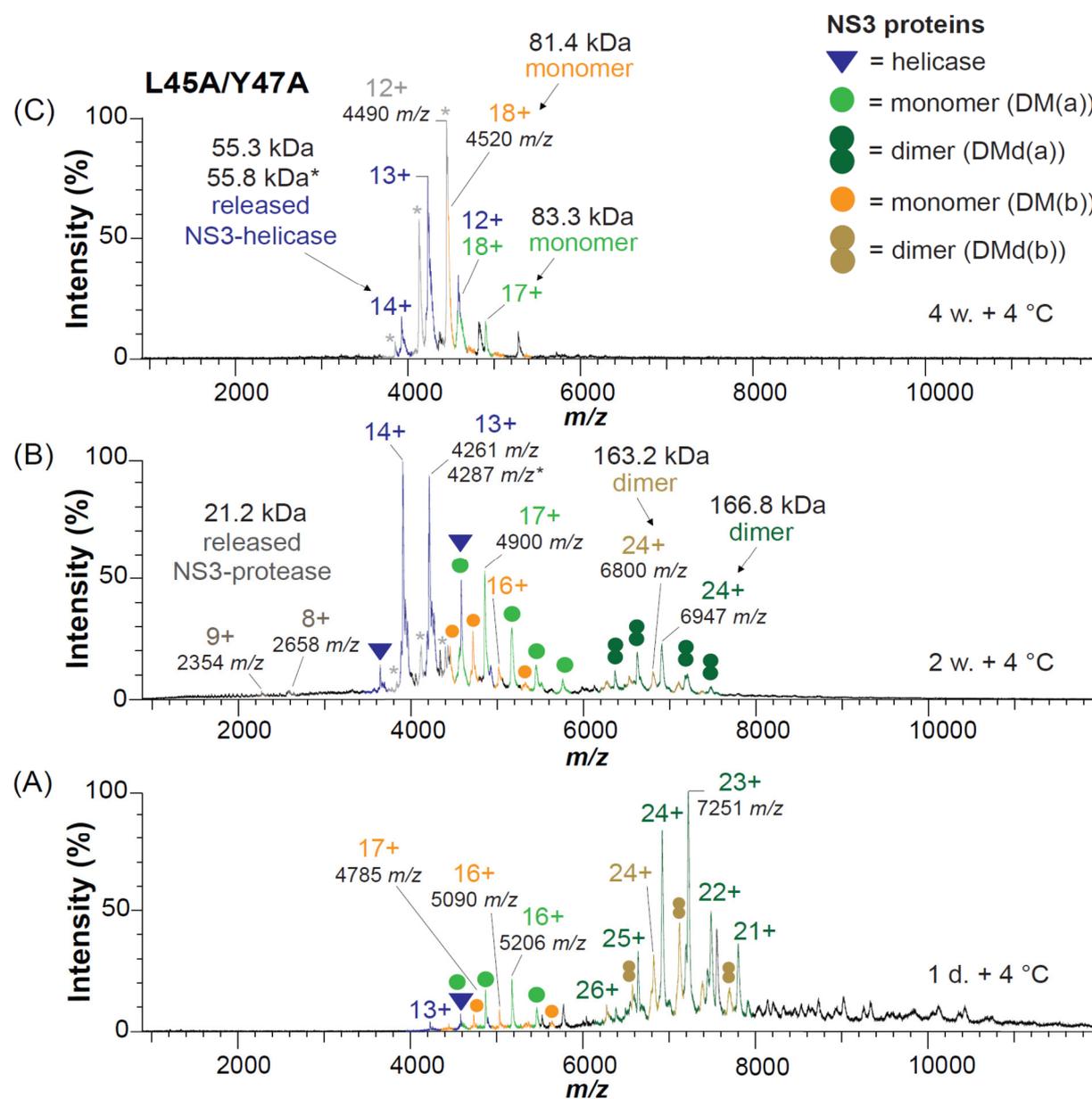


Fig. 59 Dynamic structural changes of the L45A/Y47A mutant over time.

Protein storage test at + 4 °C. (A) Native spectrum of the fresh protein (1 day). (B) Spectrum after 2 weeks of storage at + 4 °C. (C) Spectrum after 4 weeks at + 4 °C. The main monomer species DM(a) = 83.298 kDa, light-green line and DM(b) = 81.382 kDa, orange line are shown. Associated dimers DMd(a) = 166.8 kDa, dark-green line and DMd(b) = 163.2 kDa, brown line are labelled. NS3-helicase (blue line), un-identified proteins (black lines, grey star (*)) possibly 53.9 kDa). Protein concentration = 8 μ M, 200 mM ammonium acetate, pH 7.4 + 0.5 mM DTT, measurements: positive ion mode at 10 eV energy.

With the wild-type protein, the formation of different monomer and dimer species was observed depending on storage for a longer time at + 4 °C. However, data suggest that these NS3 proteins are not structurally stable enough for long time storage at + 4 °C (Fig. 59). All proteins seemed to degrade over time. Since they behave so dynamically, it was of interest to check whether the degradation process also takes place within a similar time range with the L45A/Y47A mutant. This test was performed based on a simple storage study where the double mutated protein was stored for 4 weeks at + 4 °C (Fig. 59). The data show several similarities to the WT and the SM proteins. Though, different monomer and dimer species were identified on the first day of the MS measurement, which suggests that this protein complex appears to be more unstable overall than the WT and SM variants. This could be a result of the amino acid substitution from the stabilizing Tyr to Ala. As a result, the released NS3-helicase subunit (blue lines) and even a small amount of a 21 kDa protein was found in the range of 2000 to 3000 m/z (9+ and 10+), which is assumed to be the released NS3-protease (interacting with the last N-terminal amino acids of NS4A +GSGS-linker) (Fig. 59 (B)). It was observed by native MS, that the DM protein complex starts to degrade over time (Fig. 60), similar to the WT protein.

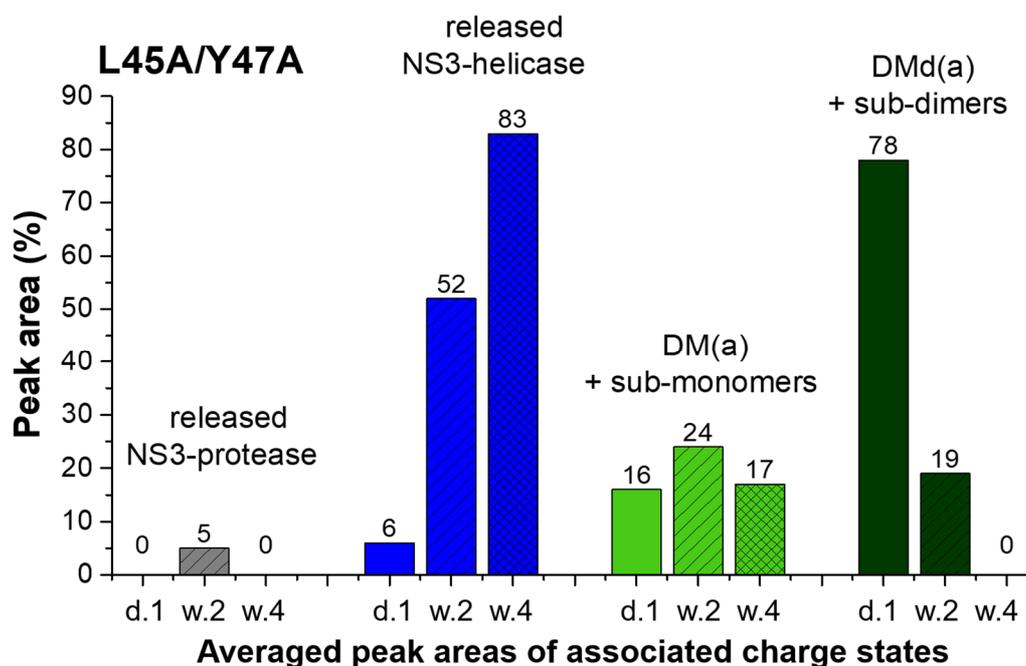


Fig. 60 The double mutation (DM) L45A/Y47A also demonstrates a high level of auto-proteolytic activity at + 4 °C.

The numbers on the bar tips are percentages (%). Data are based on averaged peak areas. Associated charge states: monomer (18+ to 16+), dimer (25+ to 22+), helicase (14+ to 12+), protease (9+, 8+). Protein concentration = 8 μ M, 200 mM ammonium acetate, pH 7.4 + 0.5 mM DTT, measurements: positive ion mode at 10 eV energy. Data are based on S. Tab. 21 and S. Tab. 22.

The entire distribution of peak areas for three different time points (1 day to 4 weeks), including free/released NS3-protease, released NS3-helicase as well as the monomer (DM(a)) and dimers of the fusion protein (DMd(a)) was compared and shown in Fig. 60. The peak signals were evaluated as area under the curve including overlaid sub-peak signals. Data are based on averaged peak areas of the associated charge states in order to obtain an overview of the signal distribution. The ratio associated with the free NS3-helicase for the first time point (1 day) is 6%, the monomer is at 16%, the dimer at 78% and the free NS3-protease is not present yet (Fig. 60, Fig. 59 Fig (A)). For the second time point (2 weeks) the amount of free helicase increased significantly to 52%, comparable to the monomer which also increased up to 24%.

In contrast, the amount of protein dimer decreased to 19%. However, a small amount of released NS3-protease was observed (5%) (Fig. 60, Fig. 59 Fig. (B)). Interestingly, the situation shifted for the last time point (4 weeks) in the direction of the released helicase subunit. The amount of free NS3-helicase is now at 83%. Whereas the DM(a) monomer is only 17% and both the protein dimer and the higher charged species (9+ and 8+) of the released protease subunit are no longer present (Fig. 60, Fig. 59 Fig (C)).

Due to the appearance of the smaller 21 kDa protein after two weeks, which is presumably the free NS3-protease, a protein sample that was three weeks stored at + 4 °C is particularly interesting and will be examined more closely. However, since the identified protein peaks are of very low intensity and resolution, it was not possible to identify additional meaningful MS2 ions. Nevertheless, the amount in which the 21 kDa protein occurs in comparison to the DM(a) monomer as well as to the free helicase subunit was examined more closely (Fig. 61). The data excludes the protein dimers but shows that the overall distribution of the peak areas of NS3-helicase and DM(a) monomer are slightly different compared to the WT and the L45A mutant. The overall distribution of the charge states, including the DM(a) monomer (green), the released NS3-helicase (blue) and the free NS3-protease (grey) is shown in Fig. 61 (A). The location of the mutation in the NS4A-kink region is shown in the crystal structure in Fig. 61 (B). The calculations are based on the normalized peak areas (Fig. 61 (C)). The data indicate that, the degradation of the NS3-NS4A complex always seems to go hand in hand with the release of the helicase subunit. As shown in Fig. 61 (D),

the DM protein shows a huge amount of unbound helicase (79%) and only a small amount of monomer (14%).

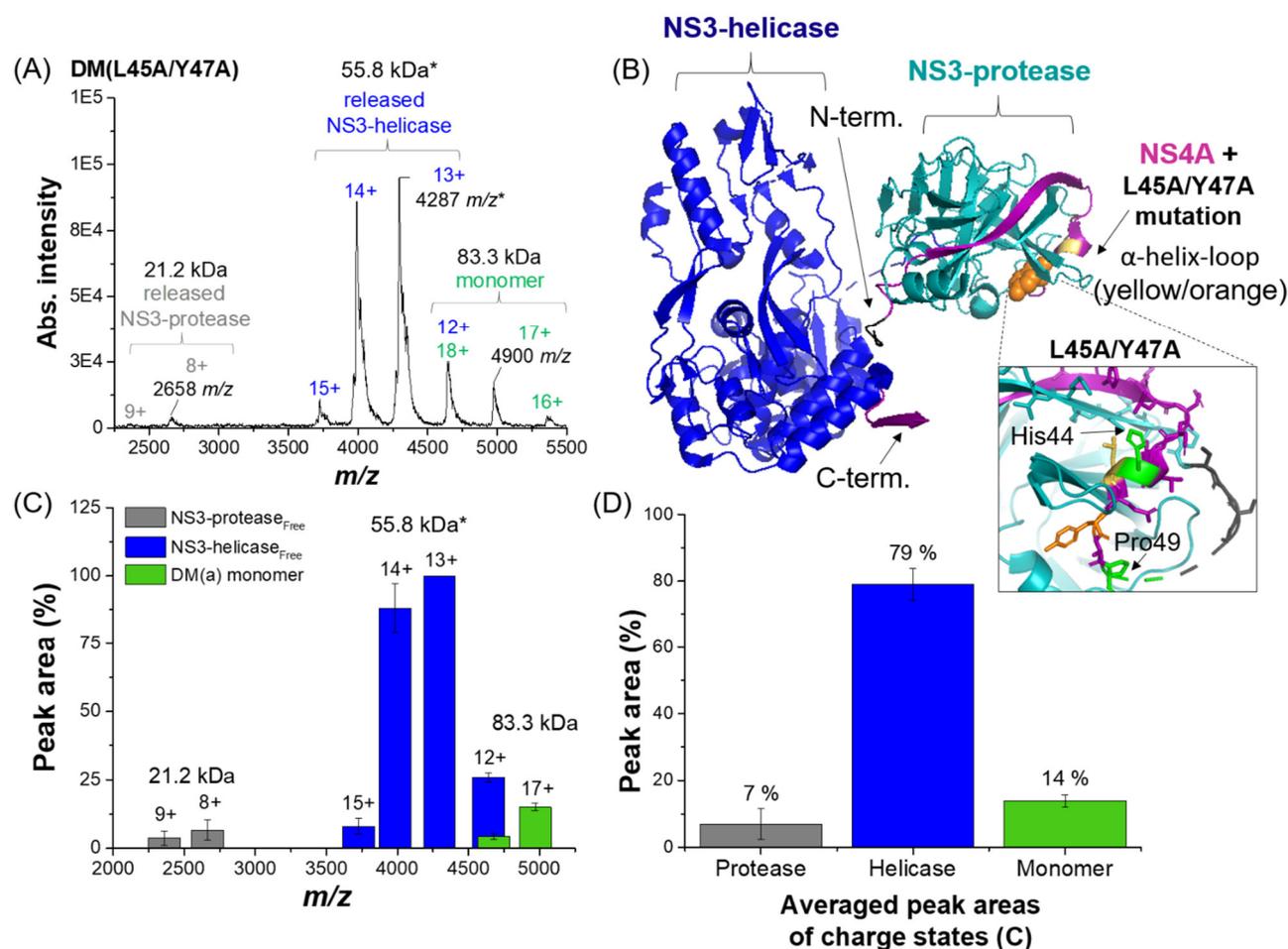


Fig. 61 L45A/Y47A shows protease activity even when stored three weeks at + 4 °C.

(A) Native spectrum of the L45A/Y47A DM mutant. (B) Crystal structure of the protein monomer based on PDB: 5MZ4, [103]. Location of the mutation at the NS4A cofactor (purple) is marked as orange and yellow sticks, NS3-protease (cyan), GSGS-linker (grey sticks), NS3-helicase (blue). (C) Distribution of charge states normalised to the 13+ peak of the helicase (blue), monomer (green), released protease (grey). (D) Determination of associated peak areas for the DM(a), released helicase and protease. Calculations are based on triplicated measurements and on averaged peak areas. Data are shown in S. Tab. 23, S. Tab. 24 and S. Tab. 25. Protein concentration = 7 μ M (monomer) in 200 mM ammonium acetate, pH 7.4 + 0.5 mM DTT, measurements: positive ion mode at 10 eV energy.

The released NS3-protease was not identified in the SM samples. This is assumed to be based on the much lower resolved spectra and the overall lower resolution of the protein peaks. Interestingly, the 21 kDa protein was found in the DM after 2 and 3 weeks and accounts for approx. 7% of the total peak area (Fig. 61). The data also confirm the auto-proteolytic activity of the NS3-protease, which appears to be active at + 4 °C regardless of the mutations in the N-terminal NS4A cofactor region. So far, however, it is not yet clear

whether the mutations increase the protease activity or rather reduce it or if they have no effect at all. However, the L45A/Y47A mutation of the NS4A cofactor apparently seemed to lead to changes in the dimerization of the protein complex and indicate a much faster formation of different monomers compared to WT and SM as shown in the overview of Fig. 55. Both mutated proteins appear to be structurally more unstable than WT. In accordance with native MS data, SDS gels of SM and DM show the free protease and helicase (chapter 3.1 Fig. 19 and chapter 3.5 Fig. 55).

In contrast to the WT sample, after four weeks of storage at + 4 °C, MS results have shown that various protein species are still present in the sample and the NS3-helicase has not yet been fully released. Therefore, it is assumed that the auto-cleavage between NS3-helicase/protease is faster at the beginning but progressed more slowly, when a certain amount of protease, which is not stabilized by NS4A is released from the complex. This is in line with the assumption, that the mutation in the kink region of the cofactor influence the interaction between the protease and NS4A. The NS4A cofactor is suggested to be important for complete protease activity [103]. According to MS results, as long as the protease is in complex with the helicase, the kink mutation does not seem to hinder the cleavage of helicase and protease. If the protease is then released from the complex it appears to be less active, which could be caused by the disturbed interaction with the NS4A cofactor. In contrast to WT, which constantly releases more active protease and completely separates the helicase and protease, this is not observed for the mutated proteins. In addition, in contrast to the WT sample, the C-terminal part of the NS4A cofactor in the DM protein is not completely cleaved off, which is a further indication that the kink mutations influence the activity of the protease.

To be able to check the differences in enzyme activity between WT and mutated proteins by MS in future experiments, a simple ESI-MS-based activity test was developed. This should be seen as proof of concept of whether mass spectrometry can represent an alternative analysis method to conventional fluorescent measurements.

3.6 ESI-MS based NS3-protease activity assay monitored substrate degradation

Substrates with fluorophores do not simulate the natural environment in which the enzymes are naturally proteolytically active. MS-based assays, on the other hand, can work with unmodified substrates, which means that this method enables the investigation of enzyme activities under more natural conditions than is the case with UV-based approaches. This part of the thesis demonstrates the possibilities of native MS as a complementary technique for the investigation of enzyme activities. Investigations can be seen as proof of concept and are at a preliminary stage for activity studies of the NS3-NS4A wild-type (WT) protease. The CSFV enzyme was used as a model system to develop this assay.

For the development of this test, however, the already known HCV NS3-NS4A protease FRET substrate (RET S1 peptide, mass = 1548.6 Da) was used in order to be able to better compare and discuss the ESI-MS-based results with the literature. The RET S1 peptide was introduced in chapter 2.5.

A stock solution of 250 μg substrate (1 mM) was prepared. RET S1 interacts with the active site of the protease, which leads to its degradation. Degradation of the substrate is monitored by a decrease of the substrate signal over time, as described in Fig. 62. According to Beran and Pyle 2008 [191] for calculations of the maximum rate of reaction, the activity test was performed with a high concentration of the RET S1 substrate to achieve oversaturation of the protease system. The protease concentration $[E]$ was 9.75 μmol and substrate concentration $[S]$ was 2.02 nmol (factor of oversaturation $[E:S] = 0.005$). Based on the selected concentrations and the huge oversaturation the calculation are assumed to be in the range of the maximum rate of the reaction.

Before the results obtained in this experiment will be presented in detail, the developed protocol and the procedure for carrying out the MS-based activity test will be described in sub-chapters 3.6.1 to 3.6.3.

The way of calculation as well as the results of the estimation of the maximum rate of reaction (V_{max}) and the associated turnover number (k_{cat}) are presented in sub-chapter 3.6.4. Here the obtained data for the CSFV NS3-protease are compared with the results obtained for the not entirely pure HCV NS3-protease. The process of substrate degradation is shown in the following schematic (Fig. 62).

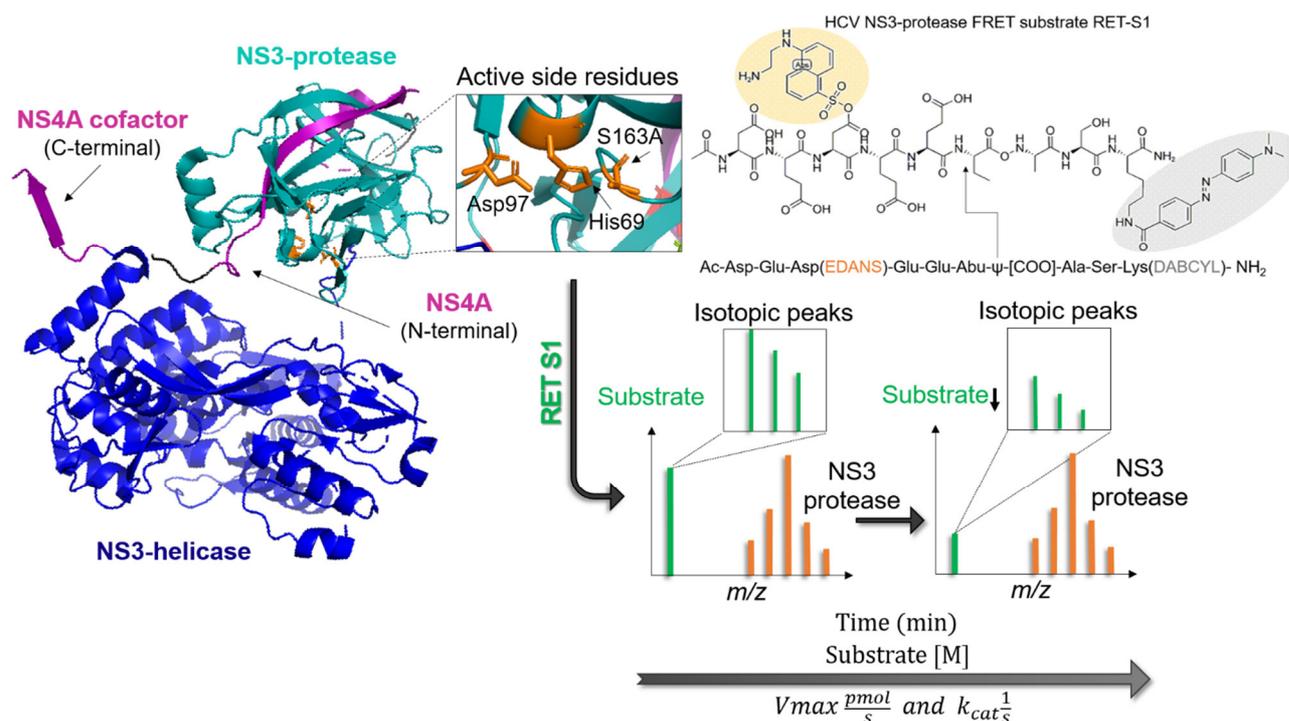


Fig. 62 Schematic of the ESI-MS based NS3-protease activity assay.

The cartoon of the CSFV NS3-NS4A_{WT} fusion protein (protease = cyan, helicase = blue, cofactor = purple; adapted from PDB 5MZ4) is shown with its active site residues (Asp97, His69). The NS3-NS4A cleavage site is marked in yellow and orange. The structural formula of the NS3-NS4A protease FRET substrate (RET S1, mass = 1548.6 Da) is displayed above. Protease (dark-blue) activity is indicated by a decrease of the substrate signal (green) over time (demonstrated by the colours of the arrow from dark- to light-grey). Calculation of the maximum rate of reaction (V_{\max}) and the turnover number (k_{cat}) is possible.

3.6.1 ESI-TOF instrument conditions

For the assay development, several things need to be considered when using an ESI-TOF instrument for the NS3-NS4A activity analyses. Measurements were performed at very soft ion activation energies (0 to 5 eV).

Best substrate peak signals were observed without introduction of collision gas in the Mass-Lynx software (settings: collision cell pressure = 1.5×10^{-4} mbar, Pirani pressure (G2) = 2.8 mbar, TOF penning pressure (G3) = 4.3×10^{-7} mbar, cone voltage = 135 to 150 V). A slight increase in collision voltage demonstrated the fragile peptide structure of the substrate and the limit in Q-TOF 2* instrument sensitivity for this assay (Fig. 63).

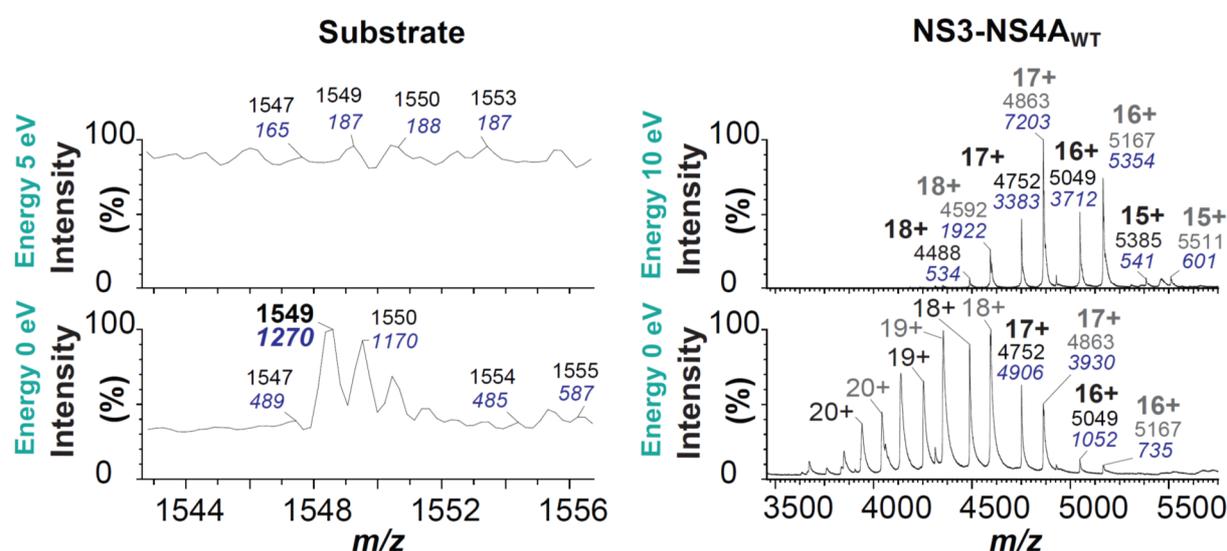


Fig. 63 ESI-TOF signals of the RET S1 FRET substrate and the CSFV NS3-NS4A_{WT} protease with and without collision energy.

The Substrate signal of the RET S1 FRET peptide (1548.6 Da, 1549 m/z , 1+ charged) is marked in bold. Black peak numbers indicate the mass to charge ratios (m/z), blue italic numbers indicate the absolute peak intensity at the specific m/z ratios. Peak and charge state distribution of the NS3-NS4A wild-type (WT) CSFV protease are shown (charges: 20+ to 15+). Black charge state numbers indicate the protein mass is 80.8 kDa and grey charge state numbers = 82.6 kDa. Collision energy was set from 10 eV to 0 eV (0 eV = acceleration energy is deactivated).

The left diagram shows the substrate signal with a mass to charge range from 1544 m/z to 1556 m/z . The spectrum below indicates the substrate peak distribution with deactivated collision gas (energy = 0 eV). The monoisotopic substrate signal can be found at 1549 m/z . The spectrum above shows the same mass to charge range but the voltage was set to 5 eV. Data clearly show that this slight increase in collision energy results in a complete lack of the substrate peak.

As a spectrum control, the right diagram shows the peak distribution of the CSFV wild-type protease with comparable energy conditions to the substrate. The mass to charge range was $3500\ m/z$ to $5500\ m/z$. The spectrum above was measured at 10 eV collision energy, as usual with these proteins. The spectrum below was recorded at 0 eV. Results only show a shift in the peak distribution to the lower mass to charge range but there is no impact on the intensity of the signal comparable with the substrate. Thus, the protease secondary structure is intact comparable to the native MS studies which were shown in the chapter before.

Additionally, different sample preparation steps need to be regulated. Based on the instrument setup (direct infusion with no LC), manual sample loading and capillary opening is necessary. Hence, this is the most time-consuming step, substrate degradation reaction must be quenched at the time points of interest for the activity assay to ensure reproducibility. Moreover, quality of raw data output varies greatly. Measurement inaccuracy strongly depends on the quality of sample capillaries, pipetting errors, buffer conditions (i.e. salt) and instrument background noise. The different steps that must be considered are listed in more detail in the following. Illustration of peak and charge state distribution is based on triplicate measurements.

3.6.2 Quenching the protease reaction

The activity of the NS3-NS4A protease can be quenched with 1% (v/v) of formic acid (FA) which is demonstrated in Fig. 64. The illustration shows the monoisotopic substrate signal (bold numbers: $1549\ m/z$) and the isotopic peak distribution (bold numbers: $1550\ m/z$ first isotope, regular numbers: $1551\ m/z$, $1552\ m/z$ second and third isotope) at three different steps of reaction. In the further analyses, only the monoisotopic peak is considered, as it shows the sharpest signal.

The first spectrum below shows the result of immediate quenching of the enzymatic reaction with 1% FA after addition of the protease to the substrate solution (**T0 = 0 min**). This was followed by immediate mass spectrometric analyses and led to a stable and clearly visible substrate signal (Fig. 64). The second spectrum (**+FA***) was recorded as described for T0. A direct reaction quenching by FA after addition of the NS3-NS4A protease was used but in contrast, the MS measurements were started after 20 min of substrate and protease incubation. The obtained spectra showed similar substrate peak signals comparable to T0.

Slight fluctuations in the intensities (blue, italic numbers) are due to measurement inaccuracy and sample handling as described above. The third spectrum (-FA**) shows the substrate peak distribution without FA quenching. Mass spectrometry investigations were performed after 10 min of substrate and protease incubation (Fig. 64).

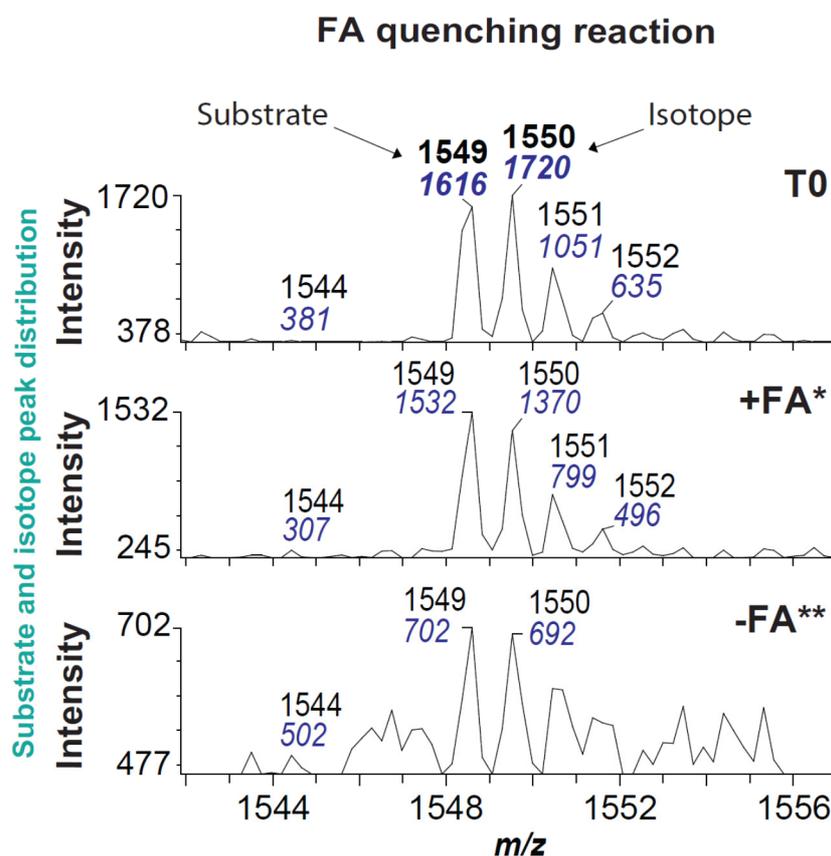


Fig. 64 Quenching of NS3-NS4A protease reaction by 1% (v/v) formic acid (FA).

Shown are substrate signals of the RET S1 FRET peptide (1548.6 Da) at three different points of protease reaction using an ESI-QTOF instrument (protease: 9.75 pmol, substrate: 2.02 nmol). Marked is the main substrate peak at 1549 m/z (1+ charged) and the first isotopic peak at m/z . Black peak numbers indicate the mass to charge ratios, blue numbers indicate the absolute peak intensity at the specific m/z ratios. Collision gas was deactivated in the MassLynx software during measurement. Spectra were fitted to their individual noise level in MassLynx. T0 = direct reaction quenching by FA after addition of protease and MS measurement, +FA* = direct reaction quenching by FA after addition of protease, MS measurement after 20 min sample incubation on ice, -FA** = sample without reaction quenching, MS measurement after 10 min sample incubation on ice.

These data demonstrate a strong decrease of the substrate signal due to the protease activity compared to the two steps which were described before. Here, it is also necessary to note the high background noise signals. These interfering conditions should be kept in mind when looking into the data as described in the following (3.6.3). In conclusion, obtained

data show that FA efficiently quenches the activity of the CSFV protease and prevents substrate degradation at different time points.

3.6.3 The effect of background signals on substrate peaks

This subsection presents an overview of the influence of noise signals on the specific substrate peak. Shown are different time points of enzymatic reaction on the ESI substrate signal (absolute peak intensities of RET S1). The presented results are not background corrected. In the first panel, raw spectra of two different time points of CSFV protease reaction are shown (Fig. 65). The highest concentration of the substrate is at time point $T = 0$ min. At this point the substrate peak ($1549 m/z$) is most intense. The lowest accurately measurable concentration is at $T = 20$ min.

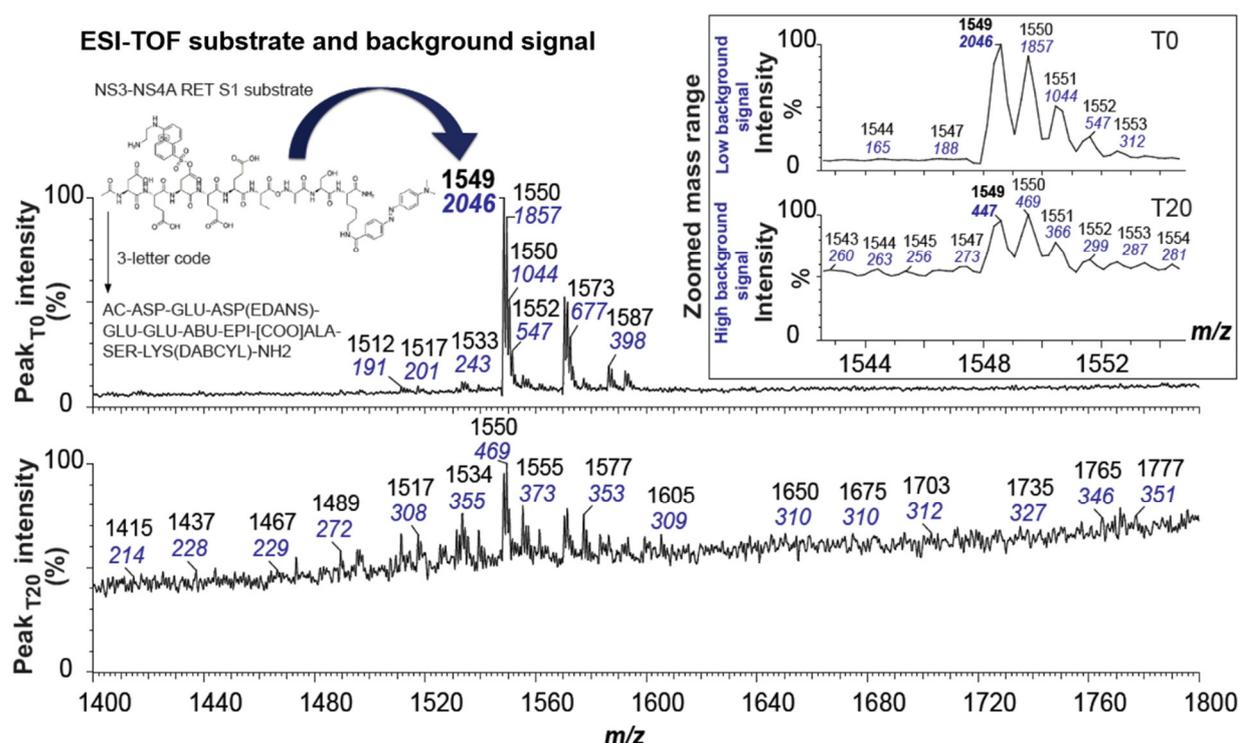


Fig. 65 Background noise during ESI-QTOF measurements: raw spectra of two time points indicate the impact on substrate signals.

Two different points of protease reaction (T_0 and T_{20} min) are shown as an example. Enzymatic reaction was performed with protease (9.75 pmol) and substrate (2.02 nmol). Main substrate peak is at $1549 m/z$ (1+ charged) followed by different isotopic peaks (1550 - 1553 m/z). Black peak numbers indicate the mass to charge ratios, blue numbers indicate the absolute peak intensity at the specific m/z ratios. Collision gas was deactivated during measurement.

As shown in Fig. 65 after 20 min of protease activity the background signals of the ESI-TOF instrument revealed an impact on the CSFV activity studies. The substrate signals are incorrectly too high because the baseline is no longer matched to the very low RET S1 concentration. This led to incorrect interpretations of the intensity values and must be rectified.

To clearly demonstrate this effect, calculations of three different measurements for each, the monoisotopic substrate peak, the first isotope and a set background noise signal were compared in Fig. 66. The data plot shows six different time points (from 0 to 20 min) of the CSFV protease reaction. The graph shows the processed raw data without background correction.

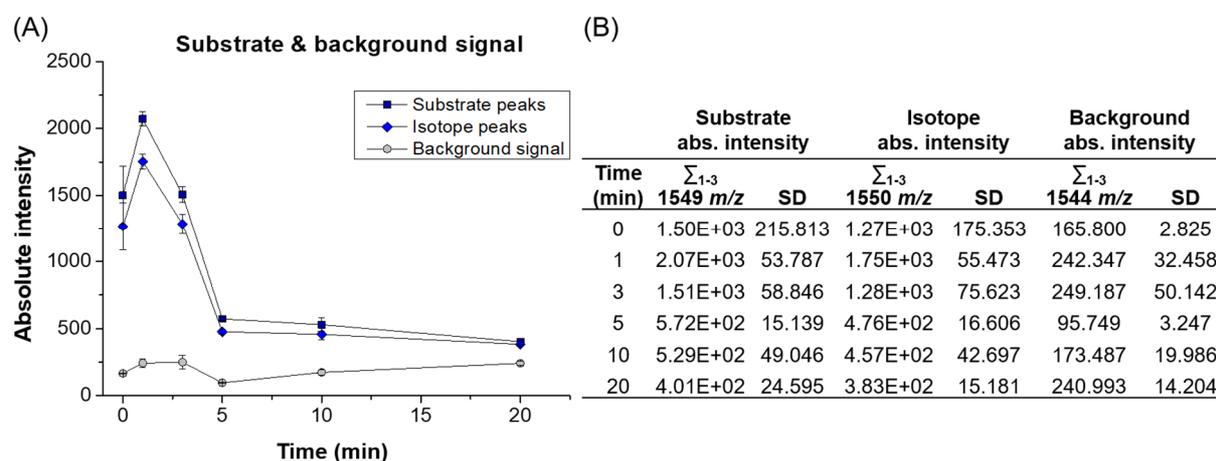


Fig. 66 Background noise during ESI-QTOF measurements: absolute peak intensities of T = 0 to T = 20 min.

(A) Comparison of the absolute peak intensity of the substrate signal with the first isotopic peak and the background/noise signal of the Q-TOF2 instrument. (B) Shown are triplicate measurements (\sum_{1-3}) of the main substrate signal (1549 m/z ; dark-blue), the 1. isotope (1550 m/z ; light-blue) and the background signal (1544 m/z ; grey) with their associated standard deviation (SD). Six different time points of the protease reaction are indicated.

The noise signals do not match the baseline. As mentioned above, this hinders the accurate calculation of the maximum rate of reaction (V_{max}) and its associated turnover number (k_{cat}). Therefore, a correction of the background signals is necessary. This is achieved by determining a signal-to-noise ratio (S/N).

3.6.4 Background correction and activity of the CSFV NS3-protease

This part of the thesis presents the results of the estimated V_{\max} as well as k_{cat} for the CSFV NS3-protease and provides a comparison of the protease activity from HCV. As an example, the calculations are carried out in more detail exclusively with the CSFV protease.

This was shown by a decrease in the specific substrate peak (1549 m/z , charge 1+) as well as by the reduction of the first isotopic peak (1550 m/z). Summarized data for both virus proteins are shown in Tab. 17. The determination of the signal-to-noise ratio (S/N) is based on peak intensity data from the substrate and the highest non-specific signal close to the substrate peak (1544 m/z), which was set as the noise signal. In order to achieve better data accuracy, two different substrate peaks (monoisotopic substrate signal = 1549 m/z and the first isotopic peak = 1550 m/z as a control, charge 1+) were used for data evaluation.

To follow the determination of the S/N ratios, important values for the calculation of CSFV protease activity as well as the uncorrected raw intensity data are presented (S. Tab. 26 and S. Tab. 27). In the following, the calculations of the enzyme activities are shown based on the CSFV protease. Fig. 67 demonstrates the degradation of the RET S1 substrate by this protease. Fig. 68 shows the substrate degradation by using the not entirely pure HCV protease. The obtained results are displayed in a linear plot diagram. The linear equation with its R^2 is shown in the graph.

In Fig. 67 the top panel shows the decrease in the monoisotopic substrate signal. The panel below shows the recorded data for the isotopic peak signal. Both data plots clearly indicate the degradation of RET S1 by a decrease in the substrate concentration over time (six different time points = T0, T1, T3, T5, T10, T20 min).

The calculation of V_{\max} and k_{cat} is similar as described for the simplified FRET-test in subchapter 3.1.3, whereby V_{\max} is equal to the slope (m) of a linear regression as shown in equation (Eq. 12):

$$m = V_{\max} = \frac{\Delta y}{\Delta x} = \frac{\Delta[S]}{\Delta t} \quad \text{Eq. 12}$$

The number of times, that the CSFV protease converts RET S1 substrate to product per second is shown by k_{cat} . The number of k_{cat} depends on the total enzyme concentration

$[E_T]$ and V_{max} , when assuming that CSFV is the pure protease. This leads to a linear relationship as described in equation (Eq. 13):

$$k_{cat} = \frac{V_{max}}{[E]_T} \quad \text{Eq. 13}$$

The results of the MS-based activity test show that the NS3-protease of CSFV has a detectable activity, where the velocity of reaction is based on the slope of the plot shown in Fig. 67.

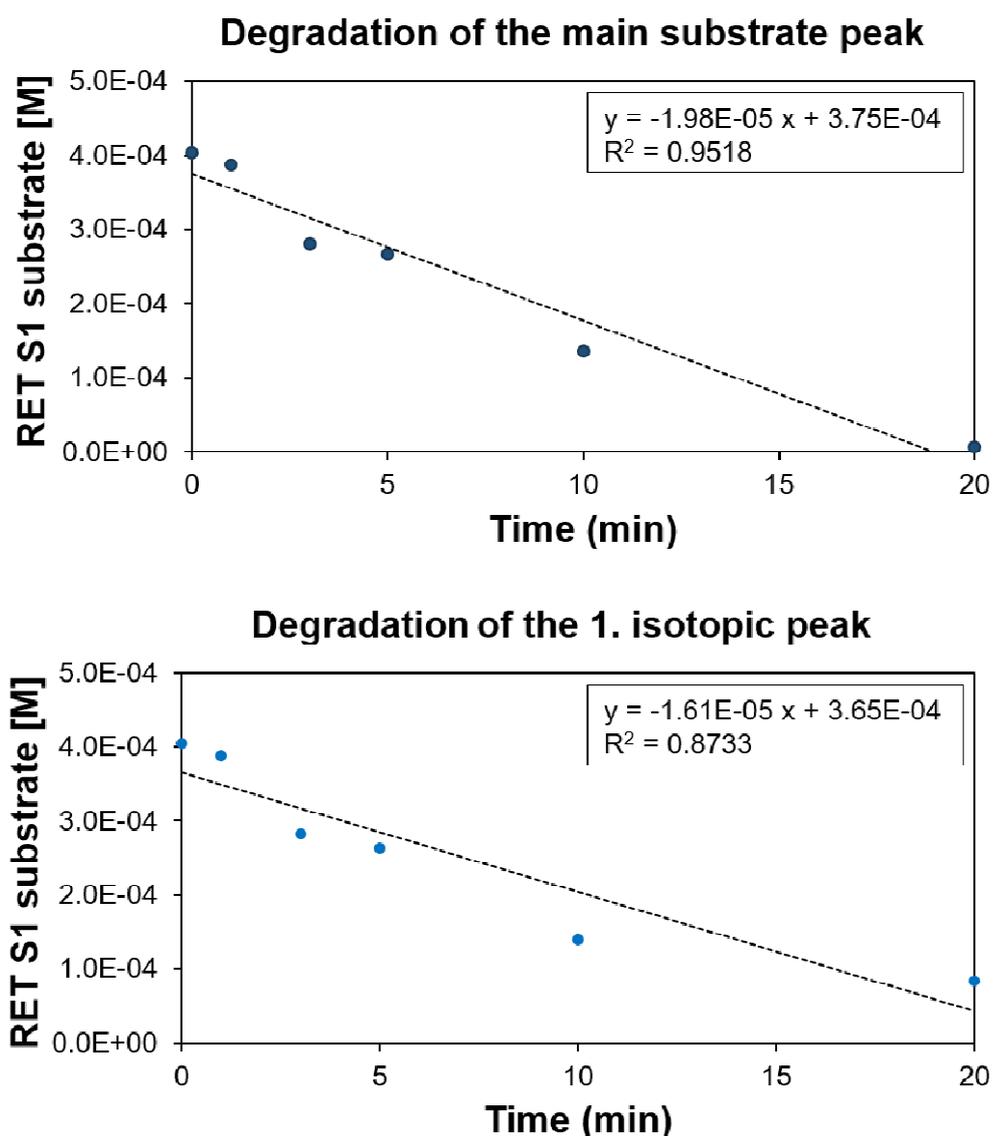


Fig. 67 Activity of the CSFV_{WT} NS3- protease in presence of the FRET substrate (RET S1 peptide). Shown is the degradation of the RET S1 substrate (monoisotopic peak above, first isotopic peak below) over 6 different time points, dotted line indicates the linear fit, small inserts in the top right corner show the equation of the linear fit as well as the goodness of fit R^2 .

At the time point $T = 0$ min the starting concentration of the substrate was 2.02 nmol ($4.04\text{E-}04$ M). The end concentration at $T = 20$ min was 0.03 nmol ($5.97\text{E-}06$ M). The underlying enzyme degradation rate is calculated as V_{\max} , where the goodness of the linear fit of the underlying parameters is described as R^2 which is at 0.952 for the monoisotopic signal and at 0.873 for the isotope.

The velocity of the reaction at the given substrate concentration $[S]$ was calculated based on Eq. 5 and leads to a $V_{\max} = 1.98\text{E-}05$ M/min $\hat{=} 1.650$ pmol/s. For the state V_{\max} the calculated k_{cat} is based on the enzyme concentration used and estimated based on Eq. 13, which leads to $k_{\text{cat}} = 0.170$ s $^{-1}$. An overview of the calculated activity values for the CSFV protease, including the monoisotopic substrate signal (main substrate peak) as well as the first isotopic substrate peak is shown in Tab. 17.

Furthermore, comparable to the CSFV protease, the enzymatic activity of the not entirely pure HCV NS3-NS4A_{WT} protease is analysed by ESI-MS (Fig. 68 and Tab. 17). The top panel of the linear plot diagram shows the degradation of the monoisotopic substrate signal. The panel below shows the degradation of the isotope peak. The enzymatic reaction was observed over five different time points ($T_0, T_1, T_5, T_{15}, T_{20}$ min).

The starting concentration of the substrate at time point $T = 0$ min was similar to the CSFV protease ($4.04\text{E-}04$ M). The final concentration at $T = 20$ min was 0.501 nmol ($1.00\text{E-}04$ M) (Tab. 17). Similar to the FRET-test for the determination of k_{cat} , the HCV sample is treated as a "pure" protease without impurities to simulate a linear relationship between k_{cat} , the total enzyme concentration $[E_T]$ and V_{\max} . Since it is known that HCV is not a pure sample, k_{cat} numbers are marked with (*).

The results indicate that the end concentration at $T = 20$ min was higher compared to the end concentration of the CSFV protease. The linear equation for HCV demonstrates a goodness of $R^2 = 0.739$ (main substrate signal) and $R^2 = 0.743$ (first isotope). This indicates higher measurement inaccuracies, especially for this protein, which can probably be ascribed to the impure sample and the lower concentration of intact protease complex. Therefore, the calculated values for HCV are only an approximation.

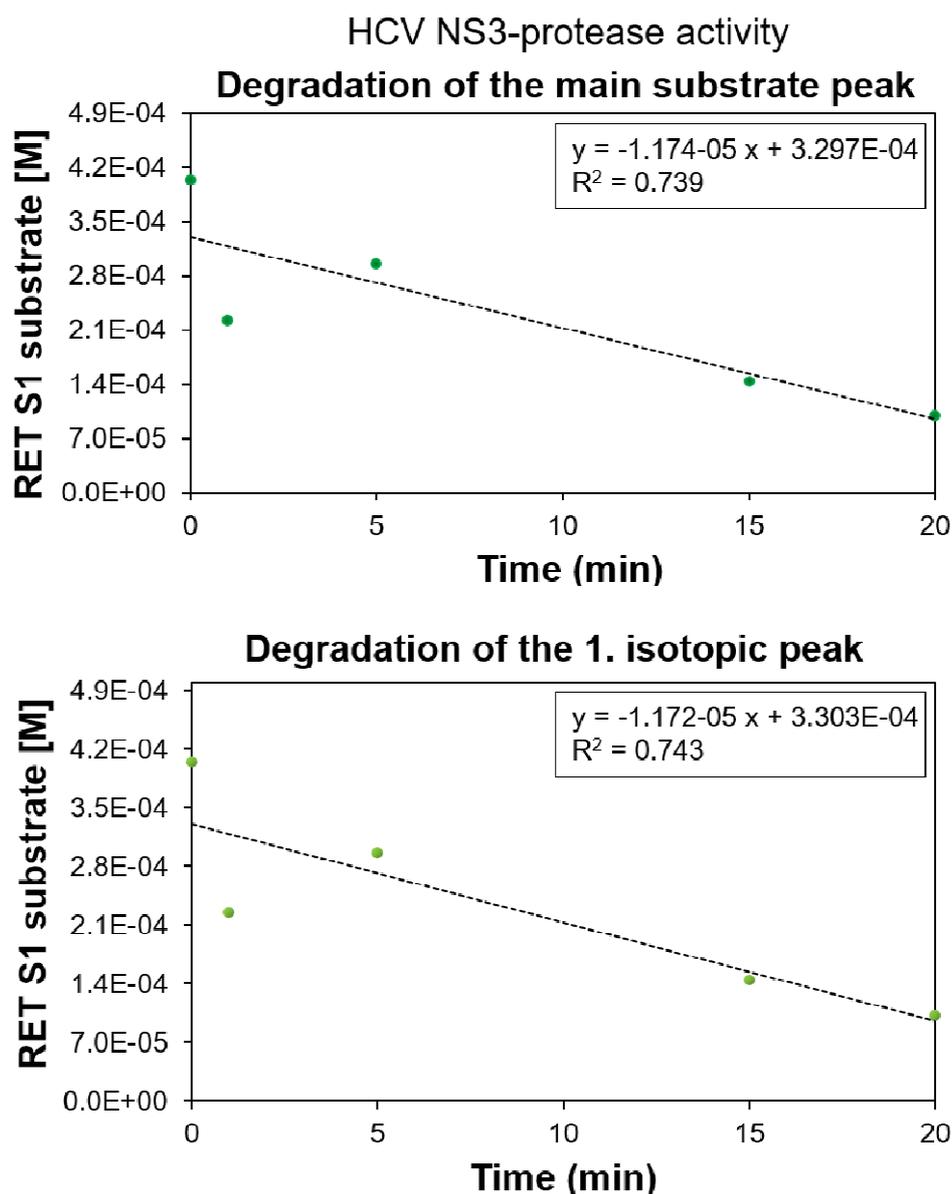


Fig. 68 Activity of the HCV_{WT} NS3-protease in presence of the RET S1 FRET substrate. Shown is the degradation of the RET S1 substrate (monoisotopic peak above, first isotopic peak below) over 5 different time points, dotted line indicates the linear fit, small inserts in the top right corner show the equation of the linear fit as well as the goodness of fit R^2 .

Tab. 17 Calculations of the activity of the CSFV and HCV NS3-protease.

Shown is the velocity of substrate degradation (V_{\max}) and the catalytic rate constant (turnover number) (k_{cat}). Concentrations MS-based activity test: CSFV = 1.95 μM and HCV = 0.65 μM + 0.4 mM RET-S1 in 5 μL , Concentrations FRET-based activity test: HCV_{Ni} FRET = 0.045 μM + 4.8 μM RET-S1 in 100 μL). Due to impurities of the HCV sample, k_{cat} numbers are marked with (*).

	CSFV RET S1		HCV RET S1		FRET RET S1
	turnover		turnover		turnover
	main peak	1. isotope	main peak	1. isotope	HCV _{Ni}
V_{\max} (pmol/s)	1.650	1.342	0.978	0.977	1.018
k_{cat} (1/s)	0.170	0.140	0.303*	0.303*	0.226*

In conclusion, the MS results demonstrate measurable enzymatic activity for both virus proteins. Although the HCV NS3-protease was not completely pure after Ni²⁺-NTA and SEC chromatography, it showed an approximated activity for the ESI-MS-based approach as well as for the simplified FRET-test (Tab. 17).

Moreover, the data show that the sample purity does not appear to have a significant influence on the results obtained for the protease activity. The V_{\max} values obtained are of a similar order of magnitude. Thus, this test also enables the determination of enzyme activities in impure samples or protein mixtures. However, the limit of the test could be reached when the enzyme concentration is too low. Further studies are necessary to determine this limit for different substrates.

In addition, with minor adjustments to the voltages and pressures (e.g. cone voltage, source pressure) in the MS instrument, this method enables not only the determination of the enzyme activity. Simultaneously monitoring of the resulting product as well as the monitoring of the protein and potential changes in its structure, while substrate conversion is possible. This is not possible with UV-based assays (e.g. FRET test).

All things considered, this preliminary data indicate that ESI-MS offers an alternative option for performing activity analysis with the NS3-protease.

4 Discussion

4.1 Native MS monitored the oligomerization and the auto-proteolytical activity of the CSFV WT NS3-NS4A complex

Nano-ESI-MS was used to analyse structural properties of the multifunctional fusion protein NS3-NS4A of CSFV and HCV. The NS3 protein is a heterodimer, which contains three interacting subunits, the N-terminal serine protease with its stabilizing cofactor NS4A and a C-terminal (ATPase)/helicase [79, 166-169]. The analysed NS3-NS4A wild-type (WT) protein complex of CSFV has a S163A mutation at the C-terminal end of the NS3-protease part, which is located at the catalytic triad region to slow down the protease activity. In order to be comparable with other studies, this variant was also used in the present work [102, 103]. Earlier studies controversially discussed whether the NS3-NS4 protein is active as a monomer or also in other oligomeric forms. Dimer formation originating from the helicase subunit seemed to play a role, but the exact function and the mechanism have not yet been fully clarified [102, 170-172]. However, oligomerization of helicases seem not to be uncommon. For other viruses, helicase dimerization has also been described (i.e. herpes simplex helicase (HSV (UL5 dimers)) as well as the formation of large hexameric helicases by the simian virus 40 (SV40) [217-219].

Under non-reducing conditions and higher protein concentrations (15 μM), the obtained native MS data clearly show a variety of full-length NS3 oligomers, including the monomer, dimer, trimer and tetramer (Fig. 24). This is in line with an early publication of Levin and Patel 1999 [172]. They used biochemical kinetic studies to show that oligomerization of HCV NS3-helicase stimulates the ATPase activity of the enzyme. Levin and Patel used protein-protein cross-linking and unwinding assays with an ATPase deficient mutant. Based on the cross-linking experiment they found that the oligomeric state of the NS3-helicase was not static. Rather it seemed that the helicase starts to dimerize in the presence of DNA, but also forms higher oligomers. They concluded that the protein functions as an oligomer to unwind DNA and that the C-terminal NS3-helicase domain is required for an active oligomeric form.

Tackett *et al.* 2005 [199] have published a DNA-unwinding assay. Here they analyzed, in the presence of a protein trap, the first cycle of unwinding by the full-length NS3 protein.

Based on their results, they hypothesized that optimal DNA unwinding required multiple molecules of NS3 which bind to the single-stranded DNA portion of the substrate. They also found that NS3 interacts heavily with itself *in vitro*.

Sikora *et al.* 2008 [193], showed that the HCV NS3 complex forms oligomers. Based on dynamic light scattering (DLS) they could show that the full-length NS3 protein exists as an oligomer, whereas the NS3-helicase domain alone exists as a monomer in solution.

This is in line with the study of Preugschat *et al.* 2000 [198], where they used steady-state unwinding assays with mixed NS3-helicase mutants (lacking strand-separating activity) and the wild-type protein but could not observe oligomerization. Thus, they concluded that the functional form of the NS3-helicase is a monomer.

The results of the present work are not completely in line with all studies above. The variety of oligomers under non-reducing conditions at high protein concentrations demonstrates that the full-length NS3 protein tend to form higher ordered protein species under the conditions used. The biggest measured oligomer was a protein tetramer (334 kDa) and the smallest was a monomer (83 kDa). The appearance of the higher ordered forms was not only dependent on protein concentration. Data show that a small amount of DTT (0.5 mM) and changes of ion acceleration energy hinders higher-ordered oligomer formation (Fig. 25). The obtained results indicate that the protein oligomerization is causing by inter-protein disulfide bond formation. Cysteines offer the possibility of forming *in vivo* and *in vitro* covalent bonds within the tertiary and quaternary structure of the proteins. This could have a biological function, which is unknown yet. However, oligomerization could also occur during the protein expression in *E. coli* cells. The bonds can occur intermolecular and lead to increased protein aggregation [194, 195]. The reduction of the NS3 enzyme using DTT in the present study revealed mainly two dominant oligomeric conformations, the protein monomer and dimer. Since the protein dimer is very prominent, it is assumed that the formation of dimers play an important role as an active enzymatic form. This hypothesis is also in line with publications of Tortorici *et al.* 2015 [102] and Khu *et al.* 2001 [171], where investigations based on SAXS analyzes, x-ray crystallography and analytical gel-filtration identified protein dimerization. Especially, Khu *et al.* 2001 showed that the HCV wild-type NS3-helicases tend to build dimers in the presence of an oligonucleotide, whereas the mutated proteins remain monomeric. In more detail, they found that the inserted mutations in conserved motifs for ATP-binding, ATPase and helicase activities (target motifs from N-

to C-terminal: G₂₀₇SGKST, D₂₉₀ECH, T₃₂₂AT were mutated to AAGKST, DECA and AAA) are not responsible for protein dimerization. To investigate the extend of oligomerization they generated promising mutants with mutations that cluster around aa 200 to 300 (T266A, Y267S, M288T) and tend to minimize the dimerization to a very low level. Based on an *in vitro* assay using double-stranded DNA oligonucleotides as a substrate they could show that the WT NS3-helicase due to its dimerization unwound the substrate very fast, while the mutants show a low rate in comparison. Therefore, they concluded that the monomeric NS3 protein possesses low helicase activity, but the dimer is the more active enzyme form. Based on this and the performed native MS study in the present work, it is obvious that the dimerization of the analyzed CSFV NS3 protein is mainly formed by the helicase domain of the protein complex. The MS data clearly showed that the release of the helicase during the proteolysis process is associated with the loss of dimers and an increase in monomer (Fig. 53, Fig. 54). Thus, the dimer formation could include amino acids which are located in the helicase domain in the range between aa260-300. Khu *et al.* 2001 showed that the mutation in the NS3 protein at threonine and tyrosine (aa T266 and Y267) reduces protein dimerization (HCV aa-sequence location I₂₆₅TYASYG). A similar amino acid sequence was found for the CSFV NS3 protein, which is located at a more exposed β -sheet region (aa I₂₉₂TYSTYG) close to the α -helical tyrosine (Y297) (S. Fig. 5). Dimerization based on edge-to-edge interaction of β -sheets is a prevalent mechanism of protein interaction [220, 221]. Since a reduction of the dimers by DTT was not possible, the dimerization of the full-length NS3 protein by a β -sheet interaction is a plausible explanation. When looking at further native MS data, however, it quickly becomes clear that the released NS3-helicase (55 kDa) does not form any dimers in the gas phase (Fig12). Since the nano-ESI process guarantees a very gentle ionisation of the protein, which does not change the tertiary and quaternary structure of the protein [118, 126, 222], it is assumed that in solution, before the injection into the mass spectrometer no dimerization of the helicase has taken place. This is corroborated by the observations of Sikora *et al.* 2008 [193]. The question arises how and why the helicase subunit is released from the NS3 protein complex. Since the NS3-protease is known to split the linkages of several NS-polyproteins [55, 89, 92, 93], the auto-catalytic cleavage mechanism of this enzyme comes into focus. Impressively, the proteolytic cleavage within the NS3-NS4A protein has been monitored by native MS. Results show that the cleavage within the NS3 protein runs hand in hand with the complete

degradation of the NS3-protease and its interacting cofactor NS4A but led to the release of a native like monomeric NS3-helicase. The conformation of the helicase is assumed to be folded based on the MS1 charge state distribution (15+ to 12+). With partial or complete unfolding of the protein, the MS1 spectrum would show a significantly broader charge distribution similar to a denatured protein spectrum [146]. Previous studies revealed that the association between NS3-protease and NS4A cofactor is important for the downstream cleavage of the polyproteins [55, 88, 89]. The protease of CSFV is known to cleave between a Leu-residue and either a Ser, Ala or Asn [55, 89]. Lamp *et al.* 2013 [92] showed that based on the internal NS3-protease cleavage sites (Leu159/Lys160, Leu192/Met193) the connection between protease and helicase can be split (Fig. 28), which leads to a complete free NS3-helicase [92, 202]. This could be confirmed by native MS data in this work. However, both mutated NS3-proteins show major differences compared to the wild-type protein (WT). The mutants appear to be structurally very labile. L45A and L45A/Y47A show a higher amount of unbound helicase subunit from the start of the analysis, which also shows double peak tips (meaning overlapping peaks) with slightly higher masses compared to the WT protein. For the L45A protein, relatively wide and less resolved peaks hinder the exact determination of different protein species. However, based on the MS2 ion fragmentation pattern it has been shown that the identified cleaved NS3-helicases exists in their complete aa-sequence.

Therefore, it is assumed that the auto-catalytic cleavage within the NS3 protein happened in the Leu192/Met193 region. In line with Lamp *et al.* 2013 [92] using native MS it was also possible to monitor the release of a fully active NS3-protease (see wild-type B3(c) and double mutant (DM)), which results in a rapid self-degradation and the dismantling of the NS4A cofactor, shown for the WT and the DM protein. Since the dominant MS1 peaks of the NS3-protease in batch 3 (B3) show a lower amount of positive charges (5+ and 4+) than those of the DM (9+ and 10+), it is assumed that the NS3-protease in B3 is structurally more compact (closely folded) than the identified protease in DM. This can be caused by the mutation in the kink region of the NS4A cofactor. It is assumed that the exchange of these amino acids hinders the interaction of protease and its cofactor [103]. The cofactor appears to be structurally more exposed based on this mutation, which means that the complex consisting of NS4A and protease carries more positive charges (charge state 9+ and 8+ in contrast to WT, which shows 5+ and 4+). Indeed, the more exposed NS4A

cofactors seems to influence the enzymatic activity of the protein. First MS data from the storage test at + 4 °C show that the NS3-protease is active. This is comparable to the WT protein, characterized by the breakdown of the NS protein complex and the release of the NS3-helicase.

In contrast to the WT, where this process is completed after three weeks, the L45A/Y47A mutant requires more than four weeks for this. Interestingly, it could be observed during this process that the helicase at the C-terminal end initially interacts with the complete NS4A cofactor. This was demonstrated on the one hand by the higher mass of 55.8 kDa compared to the WT (55.06 kDa), and on the other hand by the MS2 fragment ions (y19, y27, y30). However, this part of the cofactor seems to be degraded over time by the NS3-protease, which is shown by the appearance of overlapping peaks with the 55.8 kDa signal. A protein with a mass of 55.3 kDa was measured. Presumably, based on low intense MS2 fragment ions (y16, y18, y23) this protein is missing four C-terminal amino acids of the NS4A cofactor (aa752-aa749, NALL, see S. Fig. 11).

However, this hypothesis needs to be supported by further MS2 data. Nevertheless, the previous results of the present thesis impressively demonstrate and confirm that the NS3-serine protease of CSFV has an auto-catalytic as well as a proteolytic activity. This process ultimately has been shown to result in its own degradation and the release of a structurally intact helicase. The biological benefit of this mechanism has not yet been clarified. It can be assumed that further enzymatic functions arise or can be strengthened by this process. According to Lamp *et al.* 2013 [92] one possible function of the released NS3-helicase could be a negative regulation of the replication complex formation. In their study, they showed by sequence alignments of *Pestiviruses*, that both internal NS3 cleavage sites are well conserved among the different *Pestivirus* species [92]. Lamp *et al.* 2013 [92] also showed that, CSFV mutants with deletions of leucine residues (Leu159 or Leu192) were non-replicative and also no infectious viruses were identified. They assumed that this is due to a loss of the protease function. Further, they speculated that the auto-catalytic cleavage of the NS3 protein and the release of an active protease and a potentially active helicase demonstrates a primordial situation in the evolution. In this scenario, both proteins were unrelated due to their location on different polypeptides.

Consequently, this also resulted in their different functions. In the course of evolution, the condensation of the two different enzymatic functions within a single protein turned out

to be a sensible further development, which seemed to result in better regulation of both functions. Furthermore, within the members of the *Flaviviridae* family the NS3 serine proteases differ in their aa-sequence and substrate specificity. The protease of HCV prefers Cys/X (Ala or Ser) [84, 223] motives at P1 and the members of *Pestiviruses* cleave at Leu/Ser, Leu/Ala, and Leu/Asn sites [88, 92, 104].

However, the internal processing and proteolytical activity of the NS3 protein has also been demonstrated for HCV by Shoji *et al.* 1995 [224], Shoji *et al.* 1999 [225] and Grakoui *et al.* 1993 [84], as well as for the dengue virus type 2 by Teo and Wright 1997 [226] and Leung *et al.* 2001 [227]. In contrast to the presented MS results and to the study published by Lamp *et al.* 2013, they found internal cleavage sites for these viruses were located within the NS3-helicase domain. Especially, Shoji *et al.* 1999 [225] showed that the NS3-serine-protease was not responsible for this cleavage. Instead, they prefer the hypothesis that the internal cleavage is mediated by cellular proteases. They concluded that the functional role of the internal cleavage within the NS3 protein is not known so far. However, they suggest that the cleavage within the NS3 protein is a common phenomenon among the members of the *Flaviviridae* family and is important for their life cycle. Based on this they assumed that the identified internal cleavage mechanism consequently inactivates the RNA-dependent RNA helicase, which regulates the mechanism of viral RNA synthesis.

Another study by Beran and Pyle 2008 [191], in contrast showed based on an activity study that the NS3-protease activity is enhanced by the presence of the NS3-helicase domain. Accordingly, the complex consisting of both protein subunits seems to be important for the cleavage of existing substrate molecules.

In conclusion, it can be said that despite many years of intensive research the multifaceted function and underlying mechanisms of the NS3-protease and helicase have not yet been fully understood. Nevertheless, this work shows that native MS is a very useful tool to study these proteins. Especially for the analysis of the NS3-protease activity, protein subunit interaction and protein degradation. Through intensive research in the past decades, it was possible to develop antiviral active ingredients, which however, show various deficiencies, especially for the *Pestivirus* group. A potentially new generation of possible vaccines is based on a compound composed of a polymeric shell encapsulating a highly conserved non-structural protein (NS3), which has already been shown to be a target of the T-cell response

[228, 229]. The design has been introduced by Riitho *et al.* 2017 [230] and 2020 [231] based on the bovine viral diarrhoea virus (BVDV). In the development and analysis of the mechanism of action of potential active ingredients, native MS can play an important role in the future.

4.2 Top-Down-Sequencing revealed the coexistence of different NS3-protease cleavage products

Based on the gas phase protein assembly of batch 1 (B1) and batch 2 (B2) NS3-NS4A proteins the obtained Top-Down data for the selected precursor ions (17+ and 24+ charged peak) revealed the coexistence of various monomeric and dimeric sub-species, which were identified as cleavage products of the NS3-protease. Three dominant fusion protein complexes have been identified by their distinct MS1 masses (monomer level: B1(a) 83.374 kDa, B2(a) 82.683 kDa, B2(b) 80.779 kDa), dimer level: B1d(a) 166.792 kDa, B2d(b) 165.350 kDa and B2d(d) 161.558 kDa (Fig. 34, Fig. 48, Fig. 51). The main B2 oligomers show the same C-terminal ion fragmentation pattern compared to the NS3-helicase. The Top-Down results clearly indicate the missing of the C-terminal NS4A cofactor sequence. The lack of a C-terminal amino acid (Leu752) could already be observed for B1. However, the loss of this amino acid could not be clearly demonstrated due to a lack of suitable MS2 data. Based on the aa-sequence and the measured MS1 mass of the protein, the lack of this aa seems to be reasonable. However, the lack of aaL752 does not correspond to a complete cleavage motif of the NS3-protease. This would be in the range of aa750/751 (Ala / Leu). Thus, the loss of aaL752 may be based on a non-specific cleavage process. Hence, the results notably monitor the beginning of the proteolytical degradation process of the NS3-protease which, over time consequently lead to the release of the NS3-helicase subunit. These findings also support other studies, which show that the protease of CSFV cleaves between Leu-residues and Ser, Ala or Asn [55, 89]. The last C-terminal amino acid of the NS3-helicase subunit is Leu744 and the adjacent amino acid of the NS4A cofactor is Ser745. Impressively, in accordance with Tautz *et al.* 1997 [55], these native MS and Top-Down results (based on the occurrence of the specific $y_{19} = 1894.23 \text{ m/z}$ and $y_{22} = 2292.43 \text{ m/z}$ fragment ions, Fig. 34) clearly reveal the splitting between the motif Leu/Ser, which corresponds to the cleavage site at the C-terminus of NS3 protein (Leu2272 P1 position and Ser2273, P1' position) of the complete polyprotein. Interestingly, further MS2 y-ions were identified that can be assigned to other cleavage products which are not yet missing in the complete C-terminal segment of the NS4A cofactor. This mainly includes the identified monomers (B1(a), B2(d), B2(e), B2(h), B2(i) and B2(j) and associated dimers e.g. B2d(a) dimer (165.562 kDa associated with B2(d) monomer), which

are shown in Fig. 40, Fig. 41, Fig. 48 and Tab. 11 in sub-chapter 3.3.3. For example, the loss of the C-terminal amino acids aa752 to aa746 was assigned to the sub-monomers B2(d), (h) and (j). The absence of the C-terminal amino acids aa752 to aa747 was assigned to the sub-monomer B2(i) and aa752 to aa748 to B2(e). Due to the missing of the natural cutting pattern (Leu + Ser/Ala/Asn) of the NS3-protease, these identified sub-species could indicate other minor cleavage sites. In view of their similar chemical properties [232] and based on the present results the NS3-protease cuts between aliphatic/neutral amino acids even in absence of leucine. Unexpectedly, when looking at the aa752 to aa748, which indicate a cleavage between Ala747 and Glu748 it could be assumed that the NS3-protease can also cut between aliphatic/neutral and aliphatic/acetic amino acids.

Summarized by Di Cera 2009 [233] the primary specificity for trypsin-like proteases with the catalytic residues His, Asn and Ser includes side chains of Ala, Glu, Phe, Gly, Lys, Gln, Arg, Trp, Tyr. Moreover, a glutamic acid specific serine protease has also been described by Nienaber *et al.* 1993 [234]. Therefore, it can be speculated that the MS results presented show the gradual proteolytic degradation of the C-terminal end of the NS3 protein. This starts between Ala747 and Glu748 within the NS4 cofactor and results in the cleavage of the complete C-terminal sequence of the NS4A cofactor (Leu744 and Ser745) at position P1/P1'. However, there is another explanation for the distribution of the individual cleavage products (B2(a), (b), (d) and (e)). It becomes apparent that the dominant proteins B2(a) and B2(b) show the complete cleavage of the NS4A cofactor at position P1/P1' (NS3/NS4A).

Thus, based on the MS data the NS3-protease cleaves preferentially between L744 and Ser745. The other MS-based identified cleavage products occur in a lower ratio, which indicates that the cleavage at these positions was significantly less preferred by the NS3-protease. Therefore, it is rather assumed, that the presence of the identified sub-species, which are not based on the specific cleavage at position P1/P1' demonstrates, that the NS3-protease cleaves at non-specific sites if enough enzyme is present or given enough reaction time. In case of the present NS3 protein complex, the more time passes the more NS3-protease is released, which might favour non-specific cleavage. Thus, the distribution of these identified protein sub-species does not represent the gradual C-terminal degradation of the protein. It is more likely that the NS3-protease cleaves first directly at position P1/P1' in the majority of the NS3-NS4A fusion proteins. Furthermore, based on the Top-Down

analyses, different N-termini were assigned to the individually identified NS3 protein species (Tab. 11 in sub-chapter 3.3.3). A particular significant N-terminal difference has been shown for the two main monomers B2(a) and B2(b). These two NS3 protein species differ significantly in their MS1 mass. The MS2 fragmentation showed that this difference is due to the missing of 15 amino acids for B2(b) at the N-terminus. Data showed that this is based on a cleavage between Asn15 and Leu16. These findings fit to other studies, which showed that the NS3-protease of CSFV cleaves between Leu-residues and small non-charged amino acids [55, 89], although here the motif is reversed. Based on the above assumption, the occurrence of these N-terminally different protein species could actually indicate the gradual N-terminal degradation of the NS3 protein, which starts with B2(e) (lacking aa1, Met1/Ala2) and B2(f) (lacking aa1-2, Ala2/Ser3). This continues on to the cleavage between the identified internal cleavage site of the protease and helicase at residues Leu192/Met193 (B2(b), (h)). However, this theory must assume that the protein sample already contains a certain amount of free NS3-protease that catalyzes this step-by-step process. This could possibly have been co-eluted during the protein purification process. On the other hand, a protease from another full-length protein complex could also be responsible for this cleavage. Especially with a head-to-tail interaction (e.g. in the dimer) as shown by Tortorici *et al.* 2015 [102].

In the case of the single mutant (SM) and double mutant (DM) in particular, the unbound NS3-helicase was identified in the MS analyzes immediately after thawing -80 °C and preparing the fresh samples (chapter 3.5 Fig. 57, Fig. 58, Fig. 59). SM showed a huge amount of free helicase from the beginning, while DM showed a massive release of the helicase after two weeks (sub-chapter 3.5.2 Fig. 60 and Fig. 61) comparable to the wild-type protein (WT) (chapter 3.3 Fig. 29). The DM sample also contained a certain amount of free NS3-protease (shown in sub-chapter 3.5.2 by Fig. 60). The early release of the helicase in case of SM and DM could have been forced by the mutation in the NS4A-kink region, because this was not observed for the fresh WT samples. Apparently, mutations in this region result in a more labile protein complex which splits more easily between the protease/helicase at position Leu192/Met193. This could be a further indication that the first cleavage takes place between helicase/protease for the mutated proteins as well as for the WT, whereby both protein subunits are released from the complex. In accordance with native MS data, SDS gels of SM and DM also initially indicate that free protease and helicase is present in

the fresh sample from the start (chapter 3.1 Fig. 19 and chapter 3.5 Fig. 55). This is a further indication of the fragility of these proteins. However, in order to be able to make unambiguous statements about the influence of the kink mutations on the structure and composition of the NS3 protein complex, additional HDX-based experiments or IMMS investigations were useful. For the WT sample, the free helicase subunit and the unbound protease were first identified by MS after one to two weeks (Fig. 29, Fig. 30). Based on the native MS data and the SDS-Pages (chapter 3.1 Fig. 19, Fig. 20 and chapter 3.3. Fig. 29), it is assumed that initially only the NS3-NS4A protein complex was present without free protease or helicase. Based on these results, it is assumed that the first cleavage for a certain amount of the total NS3 protein complex takes place between the protease/helicase at position Leu192 and Met193, which was first identified by Lamp *et al.* 2013 [92]. This then releases the NS3-protease, which ensures the non-specific N-terminal and C-terminal degradation of the structurally intact NS3 proteins that have been retained. In addition, the free NS3-protease forces the cleavage between helicase/protease and the cleavage between NS3/NS4A at the C-terminal end of the protein complex, which leads to the rapid release of large amounts of the structurally intact NS3-helicase subunit, without the NS4A cofactor at the C-terminus (Tab. 18, Fig. 69).

Tab. 18 Cleavage products of the NS3-auto-proteolyse and the corresponding cleavage motifs.

Name and species (spec.) of the identified proteins as well as the lack of N- and C-terminal amino acids (aa) is shown. The location of the cleavage within the protein complex and the corresponding cleavage motifs based on (Lamp *et al.* 2013 [92] and Tautz *et al.* 1997 [55]) are indicated. Missing aa labeled with (**) mark the possibilities of other cleavages. Cleavage motifs labeled with (*) indicate not known motifs or non-specific cleavages. Proteins marked with (°) and (°) show identical species within different batches.

Protein name	NS3-spec.	Missing aa N-terminal	Missing aa C-terminal	N-term. location of cleavage	N-term. cleavage motif	C-term. location of cleavage	C-term. cleavage motif
B1(a)	mon.	-	aaL572	-	-	NS4A	L/L*
		-	aa752-751**	-	-	NS4A	A/L
		aaM1**	-	s.-sequence	M/A*	-	-
B2(a)	mon.	-	aa752-745	-	-	NS3/NS4A	L/S
°B2(b)	mon.	aa1-15	aa752-745	TEV	N/L	NS3/NS4A	L/S
°B3(a)	hel.	aa1-253 Tag-NS4A- NS3-pro.	aa752-745	pro./hel. aa192/193	L/M	NS3/NS4A	L/S
°B3(b) B2/3(b)	mon.	aa1-15	aa752-745	TEV	N/L	NS3/NS4A	L/S
B3(c)	pro.	aa1-61 Tag-NS4A	aa752-253	GSGS-linker	S/G*	pro./hel. aa192/193	L/M
°B4(a) B3/4(a)	hel.	aa1-253 Tag-NS4A- NS3-pro.	aa752-745	pro./hel. aa192/193	L/M	NS3/NS4A	L/S

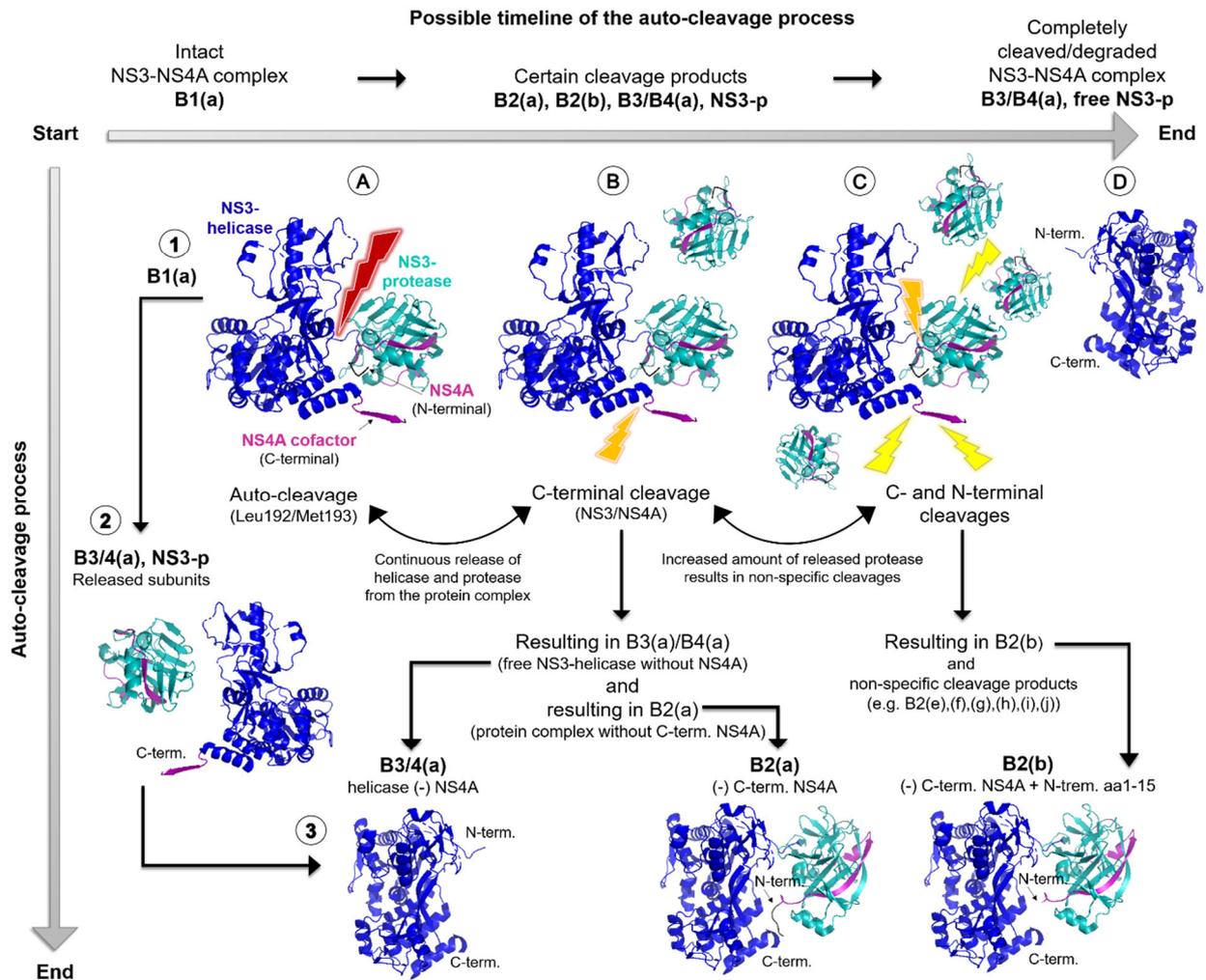


Fig. 69 Scheme of the possible course of the auto-cleavage and NS3-NS4A complex degradation on the monomeric level based on PDB 5MZ4.

On the monomer level the process of NS3-NS4 auto-cleavage starts with (1) = B1(a), which is first cleaved at position L192/M193 between protease/helicase (labelled by a red arrow) and ends with (2) = free NS3-protease (NS3-p) and (3) = B3/B4(a). The different steps inbetween are labelled with capital letters (A)-(D) and show the specific cleavages (B)-(C) = (B2(a) (missing aa752-745), B2(b) (missing aa1-15, aa752-745) (orange arrows) and non-specific cleavages (yellow arrows), ending with (D) = B3/B4(a).

4.3 The MS-based assay shows similar results to the published UV-Assay data

In accordance with the previously evaluated native MS data, the auto-proteolytic activity of the NS3 serine protease has already been demonstrated. In addition, an MS-based activity test was developed, which enables the determination of the specific enzymatic activity of the NS3-protease within the NS3 protein complex. The data generated here is not a complete determination of enzyme kinetic according to Michaelis Menten, but rather a simplified assay as proof of concept for the technique. In this study only the V_{\max} , i.e. the complete substrate oversaturation of the NS3-protease is considered and the rate/velocity at which the substrate decreases. The maximum saturation for this specific substrate is known from the literature for the NS3-protease [191]. The Michaelis-Menten equation of 1st order reactions, which assumes a linear course of the reaction, was used exclusively to calculate the rate of decrease. The velocity of the reaction for both enzymes was monitored in this study based on the substrate decrease. Due to the promising results that have been obtained, the versatile range of applications of native MS as state-of-the-art technology is demonstrated. One possible application being drug screens to monitor effectiveness and turnover which could supplement other screening assays such as x-ray used by Günther *et al.* 2021 [235].

Already in 1998, Zechel *et al.* [236] showed, that ESI-MS can be used to analyse the kinetics of the *Bacillus circulans* xylanase reaction by monitoring a transient enzyme intermediate based on its mass. Ge *et al.* 2001 [237] also showed, based on the model system glutathione S-transferase from porcine liver, that ESI-MS offers the possibility to determine the activity of enzymes. They found that their obtained kinetic parameters (K_m , V_{\max}), agreed with those obtained from traditional UV-vis spectroscopy. Especially, in accordance with Ge *et al.* 2001, the MS-based activity assay is of interest when the selected substrate has no incorporated chromophores or fluorophores. Specifically, natural occurring substrates in particular do not have this type of label [237]. Accordingly, substrates with fluorophores do not correspond to the native environment under which the enzymes are proteolytically active in nature. Since MS-based assays work with unmodified substrates, this method enables the investigation of enzyme activities under more natural conditions than is the case with UV-based approaches. The presented investigations should be seen as proof of concept and are

at a preliminary stage for activity studies of the full-length NS3-NS4A wild-type (WT) protein based on a native MS approach. Both, the CSFV and HCV enzyme were used as a model system to develop this assay. The study is based on the HCV NS3-NS4A protease FRET substrate RET-S1, which is a NS4A/NS4B junction mimic that contains fluorophore residues (RET S1, mass = 1548.6 Da). In the present study, the oversaturation of the NS3-protease (factor $[E]/[S] = 0.005$, based on 0.00975 nmol enzyme $[E]$, 2.02 nmol RET-S1 $[S]$) is in a comparable order of magnitude that has already been presented by Beran and Pyle 2008 [191] for the wild-type HCV NS3-NS4A protein (factor $[E]/[S] = 0.008$, based on 40 nM $[E]$, 5 μ M $[S]$). In their investigations, they used RET-S1 to analyse the role of the HCV NS3-helicase domain to the NS3-NS4A protease activity by monitoring the cleavage on a fluorescence spectrophotometer. They found that the interaction of both proteins (protease and helicase) enhances the enzymatic activity for substrate proteolysis. Thus, they have calculated the velocity (v) of the reaction with 17 to 0.11 pmol/s, whereby the lowest substrate turnover (0.11 pmol/s) is based on the NS3-protease + NS4A but without interaction with the helicase.

Impressively, in line with the results of Beran and Pyle 2008, the velocity of substrate degradation estimated in the MS-approach and the FRET-test for the analysed CSFV and HCV NS3-proteases are in very good agreement (CSFV MS-based assay: $V_{\max} = 1.650$ pmol/s, HCV MS-based assay: $V_{\max} = 0.978$ pmol/s, HCV simplified FRET approach: $V_{\max} = 1.018$ pmol/s).

Further Beran and Pyle 2008 calculated the turnover numbers (k_{cat}) which indicates enzymatic activity of 0.09 s⁻¹ and 3.15 s⁻¹. This is also in line with the calculated turnover number of the CSFV NS3-NS4A protein (MS assay: $k_{\text{cat}} = 0.170$ s⁻¹) and HCV NS3-NS4A protein (MS assay: $k_{\text{cat}} = 0.303$ s⁻¹ and simplified FRET approach: $k_{\text{cat}} = 0.226$ s⁻¹, shown in subchapter 3.6.4, Tab. 17).

However, although the results of the MS-based study are in the same order of magnitude as those published by Beran and Pyle in 2008, there are deviations in the calculated velocity values. Especially in comparison with the intact full-length enzyme from Beran and Pyle (NS3-protease-helicase-NS4A) there are differences (Beran and Pyle: $V_{\max} = 17$ pmol/s and $k_{\text{cat}} = 3.15$ s⁻¹).

The MS-assay-based values are lower in each case. This can mainly be explained by the fact that for both, CSFV and HCV the full-length proteins were also used comparable to Beran

and Pyle however, in contrast to their calculated velocity values, the values of the MS-study are based on the His-tagged enzyme. Where the tag is located close to the protease function at the N-terminal end.

It was shown, by Beran and Pyle in a proteolysis velocity assay of RET-S1, that the His-tagged enzyme is still active but with decreased velocity of the substrate turnover [191]. On the other hand, the physiological environment of the enzyme might play an important role in its function. For example, the pH value of the solution, the ionic strength of the solution and the ambient temperature can also influence the reaction rate of the substrate turnover. For example, another member of the *Flaviviridae* family, the dengue virus (DENV), has been shown to have a temperature dependency of the NS2-NS3-protease. In the study published by Rothan *et al.* in 2012 [238], it was shown, that DENV has the ability to propagate at different temperatures. The authors analysed the activity of the enzyme at temperatures of + 28 °C, as found in the mosquito vector, at + 37.5 °C, as it is the normal case in the human body (situation directly after the pathogen has been transmitted), and at + 40 °C, as is the case with highly infected patients. The enzyme was able to cleave a fluorogenic peptide substrate at every temperature tested. Though, they found that the protease has highest activity at the temperature of mosquitoes (+ 28 °C) and lowest was shown for humans with severe dengue fever (+ 40 °C).

A study by Leung *et al.* 2001 [227] showed, that the NS3-protease of the dengue 2 (DENV 2) virus had an optimal pH of 9.2 for proteolytic processing. An increase in the ionic strength had an impact on substrate processing. The protease was found to be susceptible to inhibition by salt at higher ionic strengths than 50 mM Tris.

In contrast, for the HCV NS3-4A protease it is reported, that it functions under conditions of higher salt (>150 mM NaCl) and pH (7.5 to 8.0) [101]. That was confirmed by Beran and Pyle in 2008, where they showed, that the proteolysis velocities of the wild-type full-length HCV protein was robust at different pH ranges (6.5 and 8.0) and various ionic strengths (30 to 200 mM NaCl) [191].

In addition, it was shown in an earlier study by Beran *et al.* in 2007 [239], that the NS3-helicase is also robust at NaCl concentrations of up to 100 mM and can even extract RNA at higher salt concentrations. From this it became evident that the NS3-protease and helicase can work in the same range of ionic strength and pH conditions (conditions up to >150 mM NaCl). Therefore, the authors concluded that NS3-NS4A is able to function

under different conditions and during different stages of viral replication. They assumed, that the NS3-protease uses the NS3-helicase subunit as a motor after the auto-cleavage of the polyprotein to move along the polyprotein and to scan and cut the subsequent peptide cleavage sites [191].

Moreover, another study by Mao *et al.* 2008 [240] showed, that NS3-protease activity is very sensitive to salt concentrations, when its concentration is too low ($<1\text{nM}$). Full protease activity was not reached even at the physiological ionic strength of 150 mM NaCl [240].

In this study, the MS assay was carried out close to the physiological conditions (200 mM ammonium acetate solution), which still showed acceptable levels of clustering of solvent salt ions and provided good ionisation of the sample. Further, measurements were carried out in a pH-neutral environment at room temperature ($+ 22\text{ }^\circ\text{C}$) until the quenching of the reaction with FA (on ice at low pH). It is assumed, that the activity of the NS3-protease is slowed down, especially at very low and higher temperatures [238]. This can also be assumed on the basis of the data from the proteolysis and auto-cleavage experiment (chapter 3.3), which was carried out at $+ 4\text{ }^\circ\text{C}$, over several weeks instead of at $+ 22\text{ }^\circ\text{C}$. Furthermore, data generated in this thesis show that after addition of FA (to 1% final volume) the activity of the enzyme is stopped (sub-chapter 3.6.2, Fig. 64).

In another study, Lin *et al.* 2004 [163] also used RET-S1 (junction mimic of NS4A/NS4B) as a substrate for their studies on the activity of the NS3-protease of HCV. In the context with a highly selective antiviral agent (ciluprevir (BILN 2061)), they published a turnover number for the untreated NS3-NS4A protein of the HCV 1b wild-type of $k_{\text{cat}} = 1.0\text{ s}^{-1}$. Taremi *et al.* 1998 [160] analysed a peptide substrate, which mimics the junction between 5A/5B for their investigations on the NS3-NS4A protein complex. They calculated a turnover number of 0.166 s^{-1} . Both are of a comparable magnitude to the MS data described here.

The velocity of substrate degradation that was determined for the HCV enzyme in the MS-based study nearly identical to the velocity calculated for the simplified FRET-assay (MS-assay: $V_{\text{max}} = 0.978\text{ pmol/s}$, FRET-approach: $V_{\text{max}} = 1.018\text{ pmol/s}$), see chapter 3.6.4, Tab. 17. Since the two HCV results are of a comparable order of magnitude, it can be assumed that the protein purification by the single Ni^{2+} -NTA (FRET data) and the double Ni^{2+} -NTA (MS data) had no significant influence on the activity of the protease.

Nevertheless, the results show, that the HCV protease has a slightly lower activity in RET-S1 turnover, than the CSFV NS3-protease ($V_{\max} = 1.650$ pmol/s). Usually, it should be assumed that HCV NS3-NS4A proteins show a faster substrate conversion, since the RET-S1 substrate is a peptide that simulates the cleavage site between NSA/NS4B specifically for the HCV NS3-protease.

Hypothetically, should the HCV protein show at least the same activity as the CSFV protein it can be assumed that the “pure” concentration of the HCV protein should be around 3/4 (0.49 μM) of the total concentration that was assumed for the impure HCV sample (0.65 μM). This calculation is based on the protein concentration of CSFV used and the MS-based activity data obtained for both proteins (CSFV and HCV).

RET-S1 does not exactly correspond to the preferred amino acid sequence of the CSFV NS3-protease. This could also be a further explanation for the lower substrate turnover of the CSFV NS3-NS4A protein used for the MS-approach, compared to Beran and Pyle, 2008, who used the untagged HCV NS3-NS4 protein. The NS3-protease of HCV prefers scissile bonds being located between Cys/Thr residues at position P1 and Ala or Ser at P1' [84, 223], whereas the NS3-protease of CSFV prefers to cleave between Leu-residues at P1 and small non-charged amino acids (Ser, Ala, Asn) [55, 89].

Apart from factors such as ionic strength and pH of the solution and tags, which should be monitored precisely for future MS-based activity studies, other factors might possibly have an influence on the proteolytic activity of the NS3 protease.

The comparably low activity of the HCV protein in the present study could be further caused by the fact, that the HCV sample is impure. In addition to a suspected *E. coli* contamination, the prolonged NS3-NS4A HCV complex (98 kDa, see MS1 data, Fig. 20) and a larger amount of released NS3-protease or truncated NS3-NS4A (see Western blot, Fig. 22) is present in the sample, which may show different proteolytic activities.

Thus, data indicate when comparing the CSFV and HCV proteins that sample purity as well as sample concentration influences the calculations and hinders the accurate determination of the substrate degradation velocity. Possibly this also had a reducing effect on the specific protease activity.

Furthermore, for the presented MS-based activity assay, it was not possible to calculate the parameter K_m (Michaelis-Menten constant). The presented data were not based on titration

of different substrate concentrations, but describe the case of substrate saturation at V_{\max} , according to Beran and Pyle 2008 [191]. Therefore, this is to be regarded as a preliminary stage for more detailed experiments. Based on the maximum concentration determined here, further titration experiments can be carried out with lower amounts of substrate in the future.

Overall, preliminary data indicate that MS offers an alternative option for performing activity analysis with the NS3-NS4A protease. While one could contest that fluorescence assays might be more sensitive than MS-approaches still (which can be alleviated when using state of the art triple quadrupole assays), there is still the fact that using MS one could use the “native” substrate, meaning without the need to chemically alter it for detection purpose which might lead to altered rates of reaction.

4.4 Possible implications of the results obtained from this thesis for future research

Since the NS3-NS4A protein complex examined in this thesis is one of the most important enzyme in the viral life cycle of the *Flaviviridae* family, it is the focus of numerous studies on viral drug development.

Seeing as the HCV sample was contaminated with host cell proteins in the same m/z -range as the target protein and thus could not be analysed via native MS, another approach might provide a chance for better characterization. As described by Donnelly *et al.* 2019 [204] an additional orthogonal LC separation using online reverse-phase (i.e. with C4-column) coupled to the MS might be able to separate the target protein from the contaminants. This separation is based on the difference in their physico-chemical properties (Van-der-Waals force) resulting in different retention times. This however, cannot be done under native conditions as elution would need to be done using organic solvents (i.e. acetonitrile). For Top-Down measurements and studying its autoproteolytic activity this approach is still a viable option.

The ESI-MS-based studies on the auto-cleavage process of the NS3-NS4 protein complex of CSFV first showed that native MS can be used as a complementary method to substantiate the results of other studies (see Lamp *et al.* 2013 [92], auto-cleavage study of the NS3-protease and helicase; Tautz *et al.* 1997 [54], P1/P1' cleavage of NS3-helicase/NS4A cofactor). In addition, this native method enables the analysis of NS3 proteins under almost physiological conditions, whereby only small amounts of sample are required for detailed investigations. In addition, various environmental conditions, such as temperature changes, pH influences and different levels of ionic strength, can be carefully examined. With native MS this can take place on the one hand at the NS3 protein level and on the other hand at the virus particle level. The analyses carried out in this work are based on a temperature of + 4 °C. A study published by Rothan *et al.* in 2012 [238] with the dengue virus (DENV) NS2-NS3-protease has shown, that the activity of the enzyme is highest at temperatures of + 28 °C. In order to be closer to the natural working temperature of the NS3-NS4 complex, tests similar to those of Rothan *et al.* 2012 could be carried out based on MS at + 37 °C host temperature. Should the cleavage of helicase and protease be a biologically relevant

process for viral replication then this auto-cleavage will be faster at higher temperatures compared to the results obtained in the present study.

In addition, the biological relevance of this cleavage between NS3-helicase and protease could be analysed with native MS in further experiments. At first, comparative studies with NS3-NS4A proteins of other members of the *Flaviviridae* family could be carried out. A closer analysis of the functionality of the released helicase could also be in focus of future investigations and its RNA binding behaviour could be examined with MS.

In addition, the MS-activity test developed in this study was able to show that it is possible to monitor the activity of the NS3-protease using ESI-MS, showing comparable results to the studies by Beran and Pyle in 2008 [191]. Based on this, the test can be further developed for future experiments on the activity of the NS3-protease in the protein complex and in its released form as well as in interaction with and without the NS4A cofactor. The implementation of various self-produced or purchased substrates with sequences of different lengths and without fluorophores could be examined in detail. In this context, the influence of mutations in the NS4A cofactor (e.g. the kink region) or in the catalytic triad of the protease could also be analysed. Since it is assumed that the interaction of the NS3-protease with the NS4A cofactor is important for the enzymatic activity of the protease, mutations in this region should have an effect on it. Native MS is able to monitor this change. Furthermore, based on the MS-activity test comparative studies with NS3-proteases of other *Flaviviridae* members can be carried out. In addition to the determination of the enzyme activity, simultaneously this method further allows the product and protein monitoring and is able to show possible changes in protein structure during substrate conversion. This is not possible with UV-based assays. The test is also of particular interest for analysing the influence of protease or helicase inhibitors on the functionality and activity of the enzymes. This assay should be even more powerful when combined with targeted analysis (multiple reaction monitoring, MRM) on these specific substrates using triple quadrupole instruments. Thus this assay could be used for the testing of active substances and drugs (e.g. drug screenings) in the future.

5 References

1. Neufeldt C.J., Cortese M., Acosta E.G., Bartenschlager R. Rewiring cellular networks by members of the Flaviviridae family. *Nat Rev Microbiol.* 2018;16(3):125-42. Epub 2018/02/13. doi: 10.1038/nrmicro.2017.170. PubMed PMID: 29430005; PubMed Central PMCID: PMC7097628.
2. Lindenbach B., Thiel H.-J., Rice C.M. Flaviviridae: The Viruses and Their Replication. In: Knipe DM, Howley, P.M. , editor. *Fields Virology.* 5 ed: Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2007.
3. Schweitzer B.K., Chapman N.M., Iwen P.C. Overview of the Flaviviridae With an Emphasis on the Japanese Encephalitis Group Viruses. *Laboratory Medicine.* 2009;40(8):493-9. doi: 10.1309/lm5yws85njpcwesw.
4. Simmonds P., Becher P., Bukh J., Gould E.A., Meyers G., Monath T., *et al.* ICTV Virus Taxonomy Profile: Flaviviridae. *J Gen Virol.* 2017;98(1):2-3. Epub 2017/02/22. doi: 10.1099/jgv.0.000672. PubMed PMID: 28218572; PubMed Central PMCID: PMC5370391.
5. Petersen L.R., Roehrig J.T. West Nile virus: a reemerging global pathogen. *Emerg Infect Dis.* 2001;7(4):611-4. Epub 2001/10/05. doi: 10.3201/eid0704.010401. PubMed PMID: 11585520; PubMed Central PMCID: PMC2631751.
6. MacLachlan N.J.a.D., E. J. Flaviviridae. *Fenner's Veterinary Virology.* 2017;5:525-45. doi: 10.1016/b978-0-12-800946-8.00029-5.
7. Rico-Hesse R., Avsic-Zupanc T., Blitvich B., Bukh J., Cao-Lormeau V.-M., Imrie A., *et al.* Flaviviridae ICTV Report: Flaviviridae: Knowles N. J. and Siddell S. G.; 2017 [updated February 2019; cited 2021 July 2021].
8. Pierson T.C., Diamond M.S. The continued threat of emerging flaviviruses. *Nat Microbiol.* 2020;5(6):796-812. Epub 2020/05/06. doi: 10.1038/s41564-020-0714-0. PubMed PMID: 32367055; PubMed Central PMCID: PMC7696730.
9. Bhatt S., Gething P.W., Brady O.J., Messina J.P., Farlow A.W., Moyes C.L., *et al.* The global distribution and burden of dengue. *Nature.* 2013;496(7446):504-7. Epub 2013/04/09. doi: 10.1038/nature12060. PubMed PMID: 23563266; PubMed Central PMCID: PMC3651993.
10. Beer M., Reimann I., Hoffmann B., Depner K. Novel marker vaccines against classical swine fever. *Vaccine.* 2007;25(30):5665-70. Epub 2007/01/24. doi: 10.1016/j.vaccine.2006.12.036. PubMed PMID: 17239502.
11. Pierson T.C., Diamond M.S. The emergence of Zika virus and its new clinical syndromes. *Nature.* 2018;560(7720):573-81. Epub 2018/08/31. doi: 10.1038/s41586-018-0446-y. PubMed PMID: 30158602.
12. Asad H., Carpenter D.O. Effects of climate change on the spread of zika virus: a public health threat. *Rev Environ Health.* 2018;33(1):31-42. Epub 2018/03/04. doi: 10.1515/reveh-2017-0042. PubMed PMID: 29500926.
13. Garske T., Van Kerkhove M.D., Yactayo S., Ronveaux O., Lewis R.F., Staples J.E., *et al.* Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLoS Med.* 2014;11(5):e1001638. Epub 2014/05/08. doi: 10.1371/journal.pmed.1001638. PubMed PMID: 24800812; PubMed Central PMCID: PMC4011853.

14. Paules C.I., Fauci A.S. Yellow Fever - Once Again on the Radar Screen in the Americas. *N Engl J Med.* 2017;376(15):1397-9. Epub 2017/03/09. doi: 10.1056/NEJMp1702172. PubMed PMID: 28273000.
15. Faria N.R., Kraemer M.U.G., Hill S.C., Goes de Jesus J., Aguiar R.S., Iani F.C.M., *et al.* Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science.* 2018;361(6405):894-9. Epub 2018/08/25. doi: 10.1126/science.aat7115. PubMed PMID: 30139911; PubMed Central PMCID: PMC6874500.
16. Charrel R.N., Attoui H., Butenko A.M., Clegg J.C., Deubel V., Frolova T.V., *et al.* Tick-borne virus diseases of human interest in Europe. *Clin Microbiol Infect.* 2004;10(12):1040-55. Epub 2004/12/21. doi: 10.1111/j.1469-0691.2004.01022.x. PubMed PMID: 15606630.
17. Shi J., Hu Z., Deng F., Shen S. Tick-Borne Viruses. *Virol Sin.* 2018;33(1):21-43. Epub 2018/03/15. doi: 10.1007/s12250-018-0019-0. PubMed PMID: 29536246; PubMed Central PMCID: PMC6866268.
18. Williams R.E., Casals J., Moussa M.I., Hoogstraal H. Royal farm virus: a new tickborne group B agent related to the RSSE complex. *Am J Trop Med Hyg.* 1972;21(5):582-6. Epub 1972/09/01. doi: 10.4269/ajtmh.1972.21.582. PubMed PMID: 4628609.
19. Westaway E.G., Brinton M.A., Gaidamovich S., Horzinek M.C., Igarashi A., Kaariainen L., *et al.* Togaviridae. *Intervirology.* 1985;24(3):125-39. Epub 1985/01/01. doi: 10.1159/000149632. PubMed PMID: 2999027.
20. Thiel H.J., Stark R., Weiland E., Rumenapf T., Meyers G. Hog cholera virus: molecular composition of virions from a pestivirus. *Journal of virology.* 1991;65(9):4705-12. Epub 1991/09/01. doi: 10.1128/JVI.65.9.4705-4712.1991. PubMed PMID: 1870198; PubMed Central PMCID: PMC68926.
21. Moennig V., Greiser-Wilke I. Classical Swine Fever Virus. *Encyclopedia of Virology*2008. p. 525-32.
22. Kaden V., Lange E., Steyer H., Bruer W., Langner C.H. Role of birds in transmission of classical swine fever virus. *J Vet Med B Infect Dis Vet Public Health.* 2003;50(7):357-9. Epub 2003/10/11. doi: 10.1046/j.1439-0450.2003.00670.x. PubMed PMID: 14535936.
23. Elbers A.R., Stegeman J.A., de Jong M.C. Factors associated with the introduction of classical swine fever virus into pig herds in the central area of the 1997/98 epidemic in The Netherlands. *Vet Rec.* 2001;149(13):377-82. Epub 2001/10/17. doi: 10.1136/vr.149.13.377. PubMed PMID: 11601514.
24. Ribbens S., Dewulf J., Koenen F., Laevens H., de Kruif A. Transmission of classical swine fever. A review. *Vet Q.* 2004;26(4):146-55. Epub 2005/01/25. doi: 10.1080/01652176.2004.9695177. PubMed PMID: 15663211.
25. Schulz K., Staubach C., Blome S. African and classical swine fever: similarities, differences and epidemiological consequences. *Vet Res.* 2017;48(1):84. Epub 2017/12/01. doi: 10.1186/s13567-017-0490-x. PubMed PMID: 29183365; PubMed Central PMCID: PMC686370.
26. Weesendorp E., Backer J., Loeffen W. Quantification of different classical swine fever virus transmission routes within a single compartment. *Vet Microbiol.* 2014;174(3-4):353-61. Epub 2014/12/04. doi: 10.1016/j.vetmic.2014.10.022. PubMed PMID: 25465177.

27. Edwards S., Fukusho A., Lefèvre P.-C., Lipowski A., Pejsak Z., Roehle P., *et al.* Classical swine fever: the global situation. *Veterinary Microbiology*. 2000;73(2-3):103-19. doi: 10.1016/s0378-1135(00)00138-3.
28. Paton D.J., Greiser-Wilke I. Classical swine fever – an update. *Research in Veterinary Science*. 2003;75(3):169-78. doi: 10.1016/s0034-5288(03)00076-6.
29. Fan J., Liao Y., Zhang M., Liu C., Li Z., Li Y., *et al.* Anti-Classical Swine Fever Virus Strategies. *Microorganisms*. 2021;9(4). Epub 2021/05/01. doi: 10.3390/microorganisms9040761. PubMed PMID: 33917361; PubMed Central PMCID: PMC8067343.
30. OIE. Classical swine fever disease <https://www.oie.int/en/disease/classical-swine-fever/>; OIE; 2021 [cited 2021 July].
31. Chevaliez S., Pawlotsky, J.-M. HCV Genome and Life Cycle. In: Tan S, editor. *Hepatitis C Viruses: Genomes and Molecular Biology*; Taylor & Francis; 2006.
32. Saeed M., Billerbeck E., Rice C.M. HCV Molecular Virology and Animal Models. 2019;31:29-68. doi: 10.1007/7355_2018_51.
33. WHO. Hepatitis C <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>; WHO; 2021 [cited 2021 July].
34. Conry-Cantilena C., VanRaden M., Gibble J., Melpolder J., Shakil A.O., Viladomiu L., *et al.* Routes of infection, viremia, and liver disease in blood donors found to have hepatitis C virus infection. *N Engl J Med*. 1996;334(26):1691-6. Epub 1996/06/27. doi: 10.1056/NEJM199606273342602. PubMed PMID: 8637513.
35. El-Kholy S.E., El-Husseiny I.M., Meshrif W.S., El-Azm A.A., Salem M.L. Does the mosquito *Culex pipiens* represent a potential vector of hepatitis C virus? *Med Vet Entomol*. 2018;32(2):155-61. Epub 2017/12/15. doi: 10.1111/mve.12288. PubMed PMID: 29239006.
36. Farajollahi A., Fonseca D.M., Kramer L.D., Marm Kilpatrick A. "Bird biting" mosquitoes and human disease: a review of the role of *Culex pipiens* complex mosquitoes in epidemiology. *Infect Genet Evol*. 2011;11(7):1577-85. Epub 2011/08/31. doi: 10.1016/j.meegid.2011.08.013. PubMed PMID: 21875691; PubMed Central PMCID: PMC3190018.
37. Barr A.R. The distribution of *Culex p. pipiens* and *C.P. quinquefasciatus* in North America. *Am J Trop Med Hyg*. 1957;6(1):153-65. Epub 1957/01/01. doi: 10.4269/ajtmh.1957.6.153. PubMed PMID: 13403135.
38. ECDC. *Culex pipiens* - Factsheet for experts <https://www.ecdc.europa.eu/en/all-topics-z/disease-vectors/facts/mosquito-factsheets/culex-pipiens-factsheet-experts>; European Centre for Disease Prevention and Control; 2020 [updated June 2020; cited 2021 July].
39. Goddard L.B., Roth A.E., Reisen W.K., Scott T.W. Vector competence of California mosquitoes for West Nile virus. *Emerg Infect Dis*. 2002;8(12):1385-91. Epub 2002/12/25. doi: 10.3201/eid0812.020536. PubMed PMID: 12498652; PubMed Central PMCID: PMC2738502.
40. Diaz-Badillo A., Bolling B.G., Perez-Ramirez G., Moore C.G., Martinez-Munoz J.P., Padilla-Viveros A.A., *et al.* The distribution of potential West Nile virus vectors, *Culex pipiens pipiens* and *Culex pipiens quinquefasciatus* (Diptera: Culicidae), in Mexico City. *Parasit Vectors*. 2011;4:70. Epub 2011/05/11. doi: 10.1186/1756-3305-4-70. PubMed PMID: 21554725; PubMed Central PMCID: PMC3117809.

41. Shi B.J., Liu C.C., Zhou J., Wang S.Q., Gao Z.C., Zhang X.M., *et al.* Entry of Classical Swine Fever Virus into PK-15 Cells via a pH-, Dynamin-, and Cholesterol-Dependent, Clathrin-Mediated Endocytic Pathway That Requires Rab5 and Rab7. *Journal of virology*. 2016;90(20):9194-208. Epub 2016/08/05. doi: 10.1128/JVI.00688-16. PubMed PMID: 27489278; PubMed Central PMCID: PMC5044825.
42. Li S., Wang J., Yang Q., Naveed Anwar M., Yu S., Qiu H.J. Complex Virus-Host Interactions Involved in the Regulation of Classical Swine Fever Virus Replication: A Minireview. *Viruses*. 2017;9(7). Epub 2017/07/06. doi: 10.3390/v9070171. PubMed PMID: 28678154; PubMed Central PMCID: PMC5537663.
43. Lecot S., Belouzard S., Dubuisson J., Rouille Y. Bovine viral diarrhoea virus entry is dependent on clathrin-mediated endocytosis. *Journal of virology*. 2005;79(16):10826-9. Epub 2005/07/30. doi: 10.1128/JVI.79.16.10826-10829.2005. PubMed PMID: 16051874; PubMed Central PMCID: PMC1182683.
44. Chen J., He W.R., Shen L., Dong H., Yu J., Wang X., *et al.* The laminin receptor is a cellular attachment receptor for classical Swine Fever virus. *Journal of virology*. 2015;89(9):4894-906. Epub 2015/02/20. doi: 10.1128/JVI.00019-15. PubMed PMID: 25694590; PubMed Central PMCID: PMC4403484.
45. Smit J.M., Moesker B., Rodenhuis-Zybert I., Wilschut J. Flavivirus cell entry and membrane fusion. *Viruses*. 2011;3(2):160-71. Epub 2011/11/04. doi: 10.3390/v3020160. PubMed PMID: 22049308; PubMed Central PMCID: PMC3206597.
46. van der Schaar H.M., Rust M.J., Chen C., van der Ende-Metselaar H., Wilschut J., Zhuang X., *et al.* Dissecting the cell entry pathway of dengue virus by single-particle tracking in living cells. *PLoS Pathog*. 2008;4(12):e1000244. Epub 2008/12/20. doi: 10.1371/journal.ppat.1000244. PubMed PMID: 19096510; PubMed Central PMCID: PMC2592694.
47. Bartenschlager R., Frese M., Pietschmann T. Novel Insights into Hepatitis C Virus Replication and Persistence. *Advances in virus research*. 2004;63:71–180.
48. Zhu E., Chen W., Qin Y., Ma S., Fan S., Wu K., *et al.* Classical Swine Fever Virus Infection Induces Endoplasmic Reticulum Stress-Mediated Autophagy to Sustain Viral Replication in vivo and in vitro. *Front Microbiol*. 2019;10:2545. Epub 2019/12/05. doi: 10.3389/fmicb.2019.02545. PubMed PMID: 31798542; PubMed Central PMCID: PMC6861840.
49. Rumenapf T., Unger G., Strauss J.H., Thiel H.J. Processing of the envelope glycoproteins of pestiviruses. *Journal of virology*. 1993;67(6):3288-94. Epub 1993/06/01. doi: 10.1128/JVI.67.6.3288-3294.1993. PubMed PMID: 8388499; PubMed Central PMCID: PMC237670.
50. Meyers G., Thiel H.J. Molecular characterization of pestiviruses. *Adv Virus Res*. 1996;47:53-118. doi: 10.1016/s0065-3527(08)60734-4.
51. Risager P.C., Fahnoe U., Gullberg M., Rasmussen T.B., Belsham G.J. Analysis of classical swine fever virus RNA replication determinants using replicons. *J Gen Virol*. 2013;94(Pt 8):1739-48. Epub 2013/04/13. doi: 10.1099/vir.0.052688-0. PubMed PMID: 23580431.
52. Murray C.L., Jones C.T., Rice C.M. Architects of assembly: roles of Flaviviridae non-structural proteins in virion morphogenesis. *Nat Rev Microbiol*. 2008;6(9):699-708.

- Epub 2008/07/01. doi: 10.1038/nrmicro1928. PubMed PMID: 18587411; PubMed Central PMCID: PMCPMC2764292.
53. Bintintan I., Meyers G. A new type of signal peptidase cleavage site identified in an RNA virus polyprotein. *The Journal of biological chemistry*. 2010;285(12):8572-84. Epub 2010/01/23. doi: 10.1074/jbc.M109.083394. PubMed PMID: 20093364; PubMed Central PMCID: PMCPMC2838279.
 54. Moulin H.R., Seuberlich T., Bauhofer O., Bennett L.C., Tratschin J.-D., Hofmann M.A., *et al.* Nonstructural proteins NS2-3 and NS4A of classical swine fever virus: essential features for infectious particle formation. *Virology*. 2007;365(2):376-89.
 55. Tautz N., Elbers K., Stoll D., Meyers G., Thiel H.J. Serine protease of pestiviruses: determination of cleavage sites. *Journal of virology*. 1997;71(7):5415-22. Epub 1997/07/01. doi: 10.1128/JVI.71.7.5415-5422.1997. PubMed PMID: 9188613; PubMed Central PMCID: PMC191781.
 56. Agapov E.V., Murray C.L., Frolov I., Qu L., Myers T.M., Rice C.M. Uncleaved NS2-3 is required for production of infectious bovine viral diarrhoea virus. *Journal of virology*. 2004;78(5):2414-25.
 57. Dubrau D., Schwindt S., Klemens O., Bischoff H., Tautz N. Determination of Critical Requirements for Classical Swine Fever Virus NS2-3-Independent Virion Formation. *Journal of virology*. 2019;93(18). Epub 2019/07/12. doi: 10.1128/JVI.00679-19. PubMed PMID: 31292243; PubMed Central PMCID: PMC6714812.
 58. Klemens O., Dubrau D., Tautz N. Characterization of the Determinants of NS2-3-Independent Virion Morphogenesis of Pestiviruses. *Journal of virology*. 2015;89(22):11668-80. Epub 2015/09/12. doi: 10.1128/JVI.01646-15. PubMed PMID: 26355097; PubMed Central PMCID: PMC4645674.
 59. Gray E.W., Nettleton P.F. The ultrastructure of cell cultures infected with border disease and bovine virus diarrhoea viruses. *J Gen Virol*. 1987;68 (Pt 9):2339-46. Epub 1987/09/01. doi: 10.1099/0022-1317-68-9-2339. PubMed PMID: 2821174.
 60. Macovei A., Zitzmann N., Lazar C., Dwek R.A., Branza-Nichita N. Brefeldin A inhibits pestivirus release from infected cells, without affecting its assembly and infectivity. *Biochem Biophys Res Commun*. 2006;346(3):1083-90. Epub 2006/06/20. doi: 10.1016/j.bbrc.2006.06.023. PubMed PMID: 16782064.
 61. de Smit A.J., Bouma A., de Kluijver E.P., Terpstra C., Moormann R.J. Prevention of transplacental transmission of moderate-virulent classical swine fever virus after single or double vaccination with an E2 subunit vaccine. *Vet Q*. 2000;22(3):150-3. Epub 2000/08/22. doi: 10.1080/01652176.2000.9695045. PubMed PMID: 10952445.
 62. Van Oirschot J.T., Terpstra C. A congenital persistent swine fever infection. I. Clinical and virological observations. *Veterinary Microbiology*. 1977;2(2):121-32. doi: 10.1016/0378-1135(77)90003-7.
 63. Lindenbach B.D., Evans M.J., Syder A.J., Wolk B., Tellinghuisen T.L., Liu C.C., *et al.* Complete replication of hepatitis C virus in cell culture. *Science*. 2005;309(5734):623-6. Epub 2005/06/11. doi: 10.1126/science.1114016. PubMed PMID: 15947137.
 64. Bartosch B., Vitelli A., Granier C., Goujon C., Dubuisson J., Pascale S., *et al.* Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. *The Journal of biological chemistry*.

- 2003;278(43):41624-30. Epub 2003/08/13. doi: 10.1074/jbc.M305289200. PubMed PMID: 12913001.
65. Scarselli E., Ansuini H., Cerino R., Roccasecca R.M., Acali S., Filocamo G., *et al.* The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J.* 2002;21(19):5017-25. Epub 2002/10/03. doi: 10.1093/emboj/cdf529. PubMed PMID: 12356718; PubMed Central PMCID: PMC129051.
 66. Pileri P., Uematsu Y., Campagnoli S., Galli G., Falugi F., Petracca R., *et al.* Binding of hepatitis C virus to CD81. *Science.* 1998;282(5390):938-41. Epub 1998/10/30. doi: 10.1126/science.282.5390.938. PubMed PMID: 9794763.
 67. Mottola G., Cardinali G., Ceccacci A., Trozzi C., Bartholomew L., Torrisi M.R., *et al.* Hepatitis C virus nonstructural proteins are localized in a modified endoplasmic reticulum of cells expressing viral subgenomic replicons. *Virology.* 2002;293(1):31-43. Epub 2002/02/21. doi: 10.1006/viro.2001.1229. PubMed PMID: 11853397.
 68. Gao L., Aizaki H., He J.W., Lai M.M. Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. *Journal of virology.* 2004;78(7):3480-8. Epub 2004/03/16. doi: 10.1128/jvi.78.7.3480-3488.2004. PubMed PMID: 15016871; PubMed Central PMCID: PMC16871042.
 69. Shi S.T., Lee K.J., Aizaki H., Hwang S.B., Lai M.M. Hepatitis C virus RNA replication occurs on a detergent-resistant membrane that cofractionates with caveolin-2. *Journal of virology.* 2003;77(7):4160-8. Epub 2003/03/14. doi: 10.1128/jvi.77.7.4160-4168.2003. PubMed PMID: 12634374; PubMed Central PMCID: PMC150636.
 70. Miyanari Y., Atsuzawa K., Usuda N., Watashi K., Hishiki T., Zayas M., *et al.* The lipid droplet is an important organelle for hepatitis C virus production. *Nat Cell Biol.* 2007;9(9):1089-97. Epub 2007/08/28. doi: 10.1038/ncb1631. PubMed PMID: 17721513.
 71. Gastaminza P., Cheng G., Wieland S., Zhong J., Liao W., Chisari F.V. Cellular determinants of hepatitis C virus assembly, maturation, degradation, and secretion. *Journal of virology.* 2008;82(5):2120-9. Epub 2007/12/14. doi: 10.1128/JVI.02053-07. PubMed PMID: 18077707; PubMed Central PMCID: PMC2258938.
 72. Gastaminza P., Kapadia S.B., Chisari F.V. Differential biophysical properties of infectious intracellular and secreted hepatitis C virus particles. *Journal of virology.* 2006;80(22):11074-81. Epub 2006/09/08. doi: 10.1128/JVI.01150-06. PubMed PMID: 16956946; PubMed Central PMCID: PMC1642172.
 73. Zumaeta Villena E. Transmission routes of hepatitis C virus infection. *Annals of Hepatology.* 2006;5:S12-S4. doi: 10.1016/s1665-2681(19)31961-1.
 74. Grassmann C.W., Isken O., Behrens S.E. Assignment of the multifunctional NS3 protein of bovine viral diarrhea virus during RNA replication: an in vivo and in vitro study. *Journal of virology.* 1999;73(11):9196-205. Epub 1999/10/09. doi: 10.1128/JVI.73.11.9196-9205.1999. PubMed PMID: 10516027; PubMed Central PMCID: PMC112953.
 75. Fletcher S.P., Jackson R.J. Pestivirus internal ribosome entry site (IRES) structure and function: elements in the 5' untranslated region important for IRES function. *Journal of virology.* 2002;76(10):5024-33. Epub 2002/04/23. doi: 10.1128/jvi.76.10.5024-5033.2002. PubMed PMID: 11967318; PubMed Central PMCID: PMC136163.

76. Thibeault D., Maurice R., Pilote L., Lamarre D., Pause A. In vitro characterization of a purified NS2/3 protease variant of hepatitis C virus. *The Journal of biological chemistry*. 2001;276(49):46678-84. Epub 2001/10/10. doi: 10.1074/jbc.M108266200. PubMed PMID: 11591719.
77. Tratschin J.D., Moser C., Ruggli N., Hofmann M.A. Classical swine fever virus leader proteinase Npro is not required for viral replication in cell culture. *Journal of virology*. 1998;72(9):7681-4. Epub 1998/08/08. doi: 10.1128/JVI.72.9.7681-7684.1998. PubMed PMID: 9696875; PubMed Central PMCID: PMCPMC110041.
78. Wiskerchen M., Collett M.S. Pestivirus gene expression: Protein p80 of bovine viral diarrhea virus is a proteinase involved in polyprotein processing. *Virology*. 1991;184(1):341-50. doi: 10.1016/0042-6822(91)90850-b.
79. Tautz N., Kaiser A., Thiel H.J. NS3 serine protease of bovine viral diarrhea virus: characterization of active site residues, NS4A cofactor domain, and protease-cofactor interactions. *Virology*. 2000;273(2):351-63. Epub 2000/08/01. doi: 10.1006/viro.2000.0425. PubMed PMID: 10915606.
80. Gale M.J., Jr., Korth M.J., Tang N.M., Tan S.L., Hopkins D.A., Dever T.E., *et al.* Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology*. 1997;230(2):217-27. Epub 1997/04/14. doi: 10.1006/viro.1997.8493. PubMed PMID: 9143277.
81. Grassmann C.W., Isken O., Tautz N., Behrens S.E. Genetic analysis of the pestivirus nonstructural coding region: defects in the NS5A unit can be complemented in trans. *Journal of virology*. 2001;75(17):7791-802. Epub 2001/08/03. doi: 10.1128/jvi.75.17.7791-7802.2001. PubMed PMID: 11483722; PubMed Central PMCID: PMCPMC115021.
82. Zhang H., Cao H., Wu Z., Cui Y. A review of molecular characterization of classical swine fever virus (CSFV). *J Vet Med*. 2011;66:89-95.
83. Lin C. HCV NS3-4A Serine Protease. In: Tan S, editor. *Hepatitis C Viruses: Genomes and Molecular Biology*: Taylor & Francis 2006.
84. Grakoui A., McCourt D.W., Wychowski C., Feinstone S.M., Rice C.M. Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. *Journal of virology*. 1993;67(5):2832-43. Epub 1993/05/01. doi: 10.1128/JVI.67.5.2832-2843.1993. PubMed PMID: 8386278; PubMed Central PMCID: PMCPMC237608.
85. Failla C., Tomei L., De Francesco R. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *Journal of virology*. 1994;68(6):3753-60. Epub 1994/06/01. doi: 10.1128/JVI.68.6.3753-3760.1994. PubMed PMID: 8189513; PubMed Central PMCID: PMCPMC236880.
86. Lin C., Thomson J.A., Rice C.M. A central region in the hepatitis C virus NS4A protein allows formation of an active NS3-NS4A serine proteinase complex in vivo and in vitro. *Journal of virology*. 1995;69(7):4373-80. Epub 1995/07/01. doi: 10.1128/JVI.69.7.4373-4380.1995. PubMed PMID: 7769699; PubMed Central PMCID: PMCPMC189178.
87. Gorbalenya A.E., Donchenko A.P., Koonin E.V., Blinov V.M. N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. *Nucleic Acids Res*. 1989;17(10):3889-97. Epub 1989/05/25. doi: 10.1093/nar/17.10.3889. PubMed PMID: 2543956; PubMed Central PMCID: PMCPMC317867.

88. Becher P., Orlich M., Thiel H.J. Complete genomic sequence of border disease virus, a pestivirus from sheep. *Journal of virology*. 1998;72(6):5165-73. doi: 10.1128/JVI.72.6.5165-5173.1998. PubMed PMID: 9573288; PubMed Central PMCID: PMC110089.
89. Xu J., Mendez E., Caron P.R., Lin C., Murcko M.A., Collett M.S., *et al.* Bovine viral diarrhea virus NS3 serine proteinase: polyprotein cleavage sites, cofactor requirements, and molecular model of an enzyme essential for pestivirus replication. *Journal of virology*. 1997;71(7):5312-22. doi: 10.1128/JVI.71.7.5312-5322.1997. PubMed PMID: 9188600; PubMed Central PMCID: PMC191768.
90. Kolykhalov A.A., Agapov E.V., Rice C.M. Specificity of the hepatitis C virus NS3 serine protease: effects of substitutions at the 3/4A, 4A/4B, 4B/5A, and 5A/5B cleavage sites on polyprotein processing. *Journal of virology*. 1994;68(11):7525-33. Epub 1994/11/01. doi: 10.1128/JVI.68.11.7525-7533.1994. PubMed PMID: 7933136; PubMed Central PMCID: PMC191768.
91. Markland W., Petrillo R.A., Fitzgibbon M., Fox T., McCarrick R., McQuaid T., *et al.* Purification and characterization of the NS3 serine protease domain of hepatitis C virus expressed in *Saccharomyces cerevisiae*. *J Gen Virol*. 1997;78 (Pt 1):39-43. Epub 1997/01/01. doi: 10.1099/0022-1317-78-1-39. PubMed PMID: 9010283.
92. Lamp B., Riedel C., Wentz E., Tortorici M.A., Rumenapf T. Autocatalytic cleavage within classical swine fever virus NS3 leads to a functional separation of protease and helicase. *Journal of virology*. 2013;87(21):11872-83. Epub 2013/08/30. doi: 10.1128/JVI.00754-13. PubMed PMID: 23986594; PubMed Central PMCID: PMC3807365.
93. Bazan J.F., Fletterick R.J. Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. *Virology*. 1989;171(2):637-9. doi: 10.1016/0042-6822(89)90639-9.
94. Urbani A., Bazzo R., Nardi M.C., Cicero D.O., De Francesco R., Steinkuhler C., *et al.* The metal binding site of the hepatitis C virus NS3 protease. A spectroscopic investigation. *The Journal of biological chemistry*. 1998;273(30):18760-9. Epub 1998/07/21. doi: 10.1074/jbc.273.30.18760. PubMed PMID: 9668049.
95. Pieroni L., Santolini E., Fipaldini C., Pacini L., Migliaccio G., La Monica N. In vitro study of the NS2-3 protease of hepatitis C virus. *Journal of virology*. 1997;71(9):6373-80. Epub 1997/09/01. doi: 10.1128/JVI.71.9.6373-6380.1997. PubMed PMID: 9261354; PubMed Central PMCID: PMC191910.
96. Vega S., Neira J.L., Marcuello C., Lostao A., Abian O., Velazquez-Campoy A. NS3 protease from hepatitis C virus: biophysical studies on an intrinsically disordered protein domain. *Int J Mol Sci*. 2013;14(7):13282-306. Epub 2013/06/28. doi: 10.3390/ijms140713282. PubMed PMID: 23803659; PubMed Central PMCID: PMC3742187.
97. Brass V., Berke J.M., Montserret R., Blum H.E., Penin F., Moradpour D. Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex. *Proc Natl Acad Sci U S A*. 2008;105(38):14545-50. Epub 2008/09/19. doi: 10.1073/pnas.0807298105. PubMed PMID: 18799730; PubMed Central PMCID: PMC2567209.
98. Kim J.L., Morgenstern K.A., Lin C., Fox T., Dwyer M.D., Landro J.A., *et al.* Crystal Structure of the Hepatitis C Virus NS3 Protease Domain Complexed with a Synthetic NS4A Cofactor Peptide. *Cell*. 1996;87(2):343-55. doi: 10.1016/s0092-8674(00)81351-3.

99. Lindenbach B.D., Pragai B.M., Montserret R., Beran R.K., Pyle A.M., Penin F., *et al.* The C terminus of hepatitis C virus NS4A encodes an electrostatic switch that regulates NS5A hyperphosphorylation and viral replication. *Journal of virology*. 2007;81(17):8905-18. Epub 2007/06/22. doi: 10.1128/JVI.00937-07. PubMed PMID: 17581983; PubMed Central PMCID: PMCPMC1951449.
100. Beran R.K., Lindenbach B.D., Pyle A.M. The NS4A protein of hepatitis C virus promotes RNA-coupled ATP hydrolysis by the NS3 helicase. *Journal of virology*. 2009;83(7):3268-75. Epub 2009/01/21. doi: 10.1128/JVI.01849-08. PubMed PMID: 19153239; PubMed Central PMCID: PMCPMC2655585.
101. Gallinari P., Paolini C., Brennan D., Nardi C., Steinkuhler C., De Francesco R. Modulation of hepatitis C virus NS3 protease and helicase activities through the interaction with NS4A. *Biochemistry*. 1999;38(17):5620-32. Epub 1999/04/29. doi: 10.1021/bi982892+. PubMed PMID: 10220351.
102. Tortorici M.A., Duquerroy S., Kwok J., Vonnrhein C., Perez J., Lamp B., *et al.* X-ray structure of the pestivirus NS3 helicase and its conformation in solution. *Journal of virology*. 2015;89(8):4356-71. Epub 2015/02/06. doi: 10.1128/JVI.03165-14. PubMed PMID: 25653438; PubMed Central PMCID: PMCPMC4442355.
103. Dubrau D., Tortorici M.A., Rey F.A., Tautz N. A positive-strand RNA virus uses alternative protein-protein interactions within a viral protease/cofactor complex to switch between RNA replication and virion morphogenesis. *PLoS Pathog.* 2017;13(2):e1006134. Epub 2017/02/06. doi: 10.1371/journal.ppat.1006134. PubMed PMID: 28151973; PubMed Central PMCID: PMC5308820.
104. Bartenschlager R., Ahlborn-Laake L., Mous J., Jacobsen H. Kinetic and structural analyses of hepatitis C virus polyprotein processing. *Journal of virology*. 1994;68(8):5045-55. Epub 1994/08/01. doi: 10.1128/JVI.68.8.5045-5055.1994. PubMed PMID: 8035505; PubMed Central PMCID: PMCPMC236447.
105. Bartenschlager R., Lohmann V., Wilkinson T., Koch J.O. Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation. *Journal of virology*. 1995;69(12):7519-28. Epub 1995/12/01. doi: 10.1128/JVI.69.12.7519-7528.1995. PubMed PMID: 7494258; PubMed Central PMCID: PMCPMC189690.
106. Shiryaev S.A., Chernov A.V., Aleshin A.E., Shiryaeva T.N., Strongin A.Y. NS4A regulates the ATPase activity of the NS3 helicase: a novel cofactor role of the non-structural protein NS4A from West Nile virus. *J Gen Virol*. 2009;90(Pt 9):2081-5. Epub 2009/05/29. doi: 10.1099/vir.0.012864-0. PubMed PMID: 19474250; PubMed Central PMCID: PMCPMC2887571.
107. Frick D.N. HCV Helicase: Structure, Function, and Inhibition. In: Tan S, editor. *Hepatitis C Viruses: Genomes and Molecular Biology*: Taylor & Francis; 2006.
108. Kolykhalov A.A., Mihalik K., Feinstone S.M., Rice C.M. Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo. *Journal of virology*. 2000;74(4):2046-51. Epub 2000/01/22. doi: 10.1128/jvi.74.4.2046-2051.2000. PubMed PMID: 10644379; PubMed Central PMCID: PMCPMC111684.
109. Warrenner P., Collett M.S. Pestivirus NS3 (p80) protein possesses RNA helicase activity. *Journal of virology*. 1995;69(3):1720-6. Epub 1995/03/01. doi: 10.1128/JVI.69.3.1720-1726.1995. PubMed PMID: 7853509; PubMed Central PMCID: PMCPMC188775.

110. Mastrangelo E., Bolognesi M., Milani M. Flaviviral helicase: insights into the mechanism of action of a motor protein. *Biochem Biophys Res Commun.* 2012;417(1):84-7. Epub 2011/12/06. doi: 10.1016/j.bbrc.2011.11.060. PubMed PMID: 22138238.
111. Chakraborty P., Pradeep T. The emerging interface of mass spectrometry with materials. *NPG Asia Materials.* 2019;11(1). doi: 10.1038/s41427-019-0149-3.
112. Aston F.W. *Mass spectra and isotopes*: Edward Arnold London; 1942.
113. Dinner A.R., Šali A., Smith L.J., Dobson C.M., Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in Biochemical Sciences.* 2000;25(7):331-9. doi: 10.1016/s0968-0004(00)01610-8.
114. Leney A.C., Heck A.J. Native Mass Spectrometry: What is in the Name? *J Am Soc Mass Spectrom.* 2017;28(1):5-13. doi: 10.1007/s13361-016-1545-3. PubMed PMID: 27909974; PubMed Central PMCID: PMC5174146.
115. Ellis R.J., Hartl F.U. Principles of protein folding in the cellular environment. *Current Opinion in Structural Biology.* 1999;9(1):102-10. doi: 10.1016/s0959-440x(99)80013-x.
116. Erijman A., Rosenthal E., Shifman J.M. How structure defines affinity in protein-protein interactions. *PLoS One.* 2014;9(10):e110085. Epub 2014/10/21. doi: 10.1371/journal.pone.0110085. PubMed PMID: 25329579; PubMed Central PMCID: PMC4199723.
117. Fitzgerald M.C., Chernushevich I., Standing K.G., Whitman C.P., Kent S.B. Probing the oligomeric structure of an enzyme by electrospray ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci U S A.* 1996;93(14):6851-6. Epub 1996/07/09. doi: 10.1073/pnas.93.14.6851. PubMed PMID: 8692908; PubMed Central PMCID: PMC38896.
118. Kebarle P., Verkerk U.H. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev.* 2009;28(6):898-917. Epub 2009/06/25. doi: 10.1002/mas.20247. PubMed PMID: 19551695.
119. Schwartz B.L., Gale D.C., Smith R.D. Noncovalent interactions observed using electrospray ionization. *Methods Mol Biol.* 1996;61:115-27. Epub 1996/01/01. doi: 10.1385/0-89603-345-7:115. PubMed PMID: 8930869.
120. Benesch J.L., Ruotolo B.T., Sobott F., Wildgoose J., Gilbert A., Bateman R., *et al.* Quadrupole-time-of-flight mass spectrometer modified for higher-energy dissociation reduces protein assemblies to peptide fragments. *Anal Chem.* 2009;81(3):1270-4. doi: 10.1021/ac801950u. PubMed PMID: 19105602.
121. Rathore D., Dodds E.D. Collision-induced release, ion mobility separation, and amino acid sequence analysis of subunits from mass-selected noncovalent protein complexes. *J Am Soc Mass Spectrom.* 2014;25(9):1600-9. Epub 2014/07/09. doi: 10.1007/s13361-014-0946-4. PubMed PMID: 25001382.
122. Konijnenberg A., Bannwarth L., Yilmaz D., Kocer A., Venien-Bryan C., Sobott F. Top-down mass spectrometry of intact membrane protein complexes reveals oligomeric state and sequence information in a single experiment. *Protein Sci.* 2015;24(8):1292-300. doi: 10.1002/pro.2703. PubMed PMID: 25970171; PubMed Central PMCID: PMC4534180.
123. Lossl P., van de Waterbeemd M., Heck A.J. The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J.* 2016;35(24):2634-57. Epub 2016/11/01. doi: 10.15252/embj.201694818. PubMed PMID: 27797822; PubMed Central PMCID: PMC5167345.

124. Chernushevich I.V., Loboda A.V., Thomson B.A. An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom.* 2001;36(8):849-65. Epub 2001/08/28. doi: 10.1002/jms.207. PubMed PMID: 11523084.
125. Greaves J., Roboz J. *Mass Spectrometry for the Novice*. Boca Raton F, editor: Taylor & Francis Group, LLC; 2014.
126. Fenn J.B., Mann M., Meng C.K., Wong S.F., Whitehouse C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* 1989;246(4926):64-71. Epub 1989/10/06. doi: 10.1126/science.2675315. PubMed PMID: 2675315.
127. Nguyen S., Fenn J.B. Gas-phase ions of solute species from charged droplets of solutions. *Proc Natl Acad Sci U S A.* 2007;104(4):1111-7. Epub 2007/01/11. doi: 10.1073/pnas.0609969104. PubMed PMID: 17213314; PubMed Central PMCID: PMC1783130.
128. Konermann L., Ahadi E., Rodriguez A.D., Vahidi S. Unraveling the mechanism of electrospray ionization. *Anal Chem.* 2013;85(1):2-9. Epub 2012/11/09. doi: 10.1021/ac302789c. PubMed PMID: 23134552.
129. Wilm M., Shevchenko A., Houthaeve T., Breit S., Schweigerer L., Fotsis T., *et al.* Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature.* 1996;379(6564):466-9. doi: 10.1038/379466a0. PubMed PMID: 8559255.
130. Wang W., Bajic S., John B., Emerson D.R. Numerical Simulation of Ion Transport in a Nano-Electrospray Ion Source at Atmospheric Pressure. *J Am Soc Mass Spectrom.* 2018;29(3):600-12. doi: 10.1007/s13361-017-1863-0. PubMed PMID: 29318469.
131. Rahman M.M., Hiraoka K., Chen L.C. Realizing nano electrospray ionization using disposable pipette tips under super atmospheric pressure. *Analyst.* 2014;139(3):610-7. Epub 2013/12/18. doi: 10.1039/c3an01635h. PubMed PMID: 24340350.
132. Tang X., Bruce J.E., Hill H.H., Jr. Characterizing electrospray ionization using atmospheric pressure ion mobility spectrometry. *Anal Chem.* 2006;78(22):7751-60. Epub 2006/11/16. doi: 10.1021/ac0613380. PubMed PMID: 17105168; PubMed Central PMCID: PMC3633478.
133. Gale D.C., Smith R.D. Small volume and low flow-rate electrospray ionization mass spectrometry of aqueous samples. *Rapid Communications in Mass Spectrometry.* 1993;7(11):1017-21. doi: 10.1002/rcm.1290071111.
134. Banerjee S., Mazumdar S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem.* 2012;2012:282574. doi: 10.1155/2012/282574. PubMed PMID: 22611397; PubMed Central PMCID: PMC3348530.
135. Lord Rayleigh F.R.S. On the equilibrium of liquid conducting masses charged with electricity. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science.* 1882;14(87):184-6. doi: 10.1080/14786448208628425.
136. Miranker A., Robinson C.V., Radford S.E., Dobson C.M. Investigation of protein folding by mass spectrometry. *FASEB J.* 1996;10(1):93-101. Epub 1996/01/01. doi: 10.1096/fasebj.10.1.8566553. PubMed PMID: 8566553.
137. Crotti S., Seraglia R., Traldi P. Some thoughts on electrospray ionization mechanisms. *Eur J Mass Spectrom (Chichester).* 2011;17(2):85-99. doi: 10.1255/ejms.1129. PubMed PMID: 21719919.

138. Fernandez de la Mora J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta*. 2000;406(1):93-104. doi: 10.1016/s0003-2670(99)00601-7.
139. Iribarne J.V., Thomson B.A. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics*. 1976;64(6). doi: 10.1063/1.432536.
140. Marshall A.G., Blakney T.G., Chen T., Kaiser K.N., McKenna M.A., Rodgers P.R., *et al.* Mass resolution and mass accuracy: how much is enough? *Mass Spectrom (Tokyo)*. 2013;2(Spec Iss):S0009. Epub 2013/12/19. doi: 10.5702/massspectrometry.S0009. PubMed PMID: 24349928; PubMed Central PMCID: PMC3809961.
141. Dennis E.A., Gundlach-Graham A.W., Ray S.J., Enke C.G., Hieftje G.M. Distance-of-Flight Mass Spectrometry: What, Why, and How? *J Am Soc Mass Spectrom*. 2016;27(11):1772-86. Epub 2016/08/27. doi: 10.1007/s13361-016-1458-1. PubMed PMID: 27562501.
142. van den Heuvel R.H., van Duijn E., Mazon H., Synowsky S.A., Lorenzen K., Versluis C., *et al.* Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem*. 2006;78(21):7473-83. Epub 2006/11/01. doi: 10.1021/ac061039a. PubMed PMID: 17073415.
143. Mamyrin B.A., Karataev V.I., Shmikk D.V., Zagulin V.A. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Journal of Experimental and Theoretical Physics*. 1973;37:45.
144. Mamyrin B.A. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry*. 2001;206(3):251-66. doi: 10.1016/s1387-3806(00)00392-4.
145. Westmacott G., Frank M., Labov S.E., Benner W.H. Using a superconducting tunnel junction detector to measure the secondary electron emission efficiency for a microchannel plate detector bombarded by large molecular ions. *Rapid Communications in Mass Spectrometry*. 2000;14(19):1854-61. doi: 10.1002/1097-0231(20001015)14:19<1854::Aid-rcm102>3.0.Co;2-m.
146. Pinkse M.W., Maier C.S., Kim J.I., Oh B.H., Heck A.J. Macromolecular assembly of *Helicobacter pylori* urease investigated by mass spectrometry. *J Mass Spectrom*. 2003;38(3):315-20. Epub 2003/03/20. doi: 10.1002/jms.443. PubMed PMID: 12644993.
147. Heck A.J. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods*. 2008;5(11):927-33. doi: 10.1038/nmeth.1265. PubMed PMID: 18974734.
148. Covey T.R., Bonner R.F., Shushan B.I., Henion J. The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry. *Rapid Commun Mass Spectrom*. 1988;2(11):249-56. Epub 1988/11/01. doi: 10.1002/rcm.1290021111. PubMed PMID: 2577836.
149. Yost R.A., Enke C.G. Selected ion fragmentation with a tandem quadrupole mass spectrometer. *Journal of the American Chemical Society*. 1978;100(7):2274-5. doi: 10.1021/ja00475a072.
150. Steen H., Kuster B., Mann M. Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. *J Mass Spectrom*. 2001;36(7):782-90. Epub 2001/07/27. doi: 10.1002/jms.174. PubMed PMID: 11473401.

151. Welham K.J. Mass Separation. *MASS SPECTROMETRY* 2005;2:397-403. doi: 10.1016/B0-12-369397-7/00344-7.
152. Boeri Erba E., Petosa C. The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. *Protein Sci.* 2015;24(8):1176-92. doi: 10.1002/pro.2661. PubMed PMID: 25676284; PubMed Central PMCID: PMC4534170.
153. Benesch J.L. Collisional activation of protein complexes: picking up the pieces. *J Am Soc Mass Spectrom.* 2009;20(3):341-8. doi: 10.1016/j.jasms.2008.11.014. PubMed PMID: 19110440.
154. Smith R.D., Loa J.A., Barinaga C.J., Edmonds C.G., Udseth H.R. Collisional activation and collision-activated dissociation of large multiply charged polypeptides and proteins produced by electrospray ionization. *Journal of the American Society for Mass Spectrometry.* 1990;1(1):53-65. doi: 10.1016/1044-0305(90)80006-9.
155. Dülfer J., Kadek A., Kopicki J.D., Krichel B., Uetrecht C. Structural mass spectrometry goes viral. *Adv Virus Res.* 2019;105:189-238. Epub 2019/09/17. doi: 10.1016/bs.aivir.2019.07.003. PubMed PMID: 31522705.
156. Steen H., Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol.* 2004;5(9):699-711. doi: 10.1038/nrm1468. PubMed PMID: 15340378.
157. Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N., Moomaw E.W., *et al.* The conformation of hepatitis C virus NS3 proteinase with and without NS4A: a structural basis for the activation of the enzyme by its cofactor. *Clinical and diagnostic virology.* 1998;10(2):151-6.
158. Krichel B., Bylapudi G., Schmidt C., Blanchet C., Schubert R., Brings L., *et al.* Hallmarks of Alpha- and Betacoronavirus non-structural protein 7+8 complexes. *Sci Adv.* 2021;7(10). Epub 2021/03/05. doi: 10.1126/sciadv.abf1004. PubMed PMID: 33658206; PubMed Central PMCID: PMC7929516.
159. Studier W., Rosenberg A.H., Dunn J.J., Dubendorff J.W. Use of T7 RNA polymerase to direct expression of cloned genes. *Gene Expression Technology. Methods in Enzymology*1990. p. 60-89.
160. Taremi S.S., Beyer B., Maher M., Yao N., Prosis W., Weber P.C., *et al.* Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease. *Protein Sci.* 1998;7(10):2143-9. Epub 1998/10/29. doi: 10.1002/pro.5560071011. PubMed PMID: 9792101; PubMed Central PMCID: PMC2143829.
161. Sambrook J., Russell D.W. SDS-Polyacrylamide Gel Electrophoresis of Proteins. *CSH Protoc.* 2006;2006(4). Epub 2006/01/01. doi: 10.1101/pdb.prot4540. PubMed PMID: 22485921.
162. Taliani M., Bianchi E., Narjes F., Fossatelli M., Urbani A., Steinkuhler C., *et al.* A continuous assay of hepatitis C virus protease based on resonance energy transfer depsipeptide substrates. *Anal Biochem.* 1996;240(1):60-7. Epub 1996/08/15. doi: 10.1006/abio.1996.0331. PubMed PMID: 8811880.
163. Lin C., Lin K., Luong Y.P., Rao B.G., Wei Y.Y., Brennan D.L., *et al.* In vitro resistance studies of hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061: structural analysis indicates different resistance mechanisms. *The Journal of biological chemistry.* 2004;279(17):17508-14. Epub 2004/02/10. doi: 10.1074/jbc.M313020200. PubMed PMID: 14766754.

164. van den Heuvel R.H., van Duijn E., Mazon H., Synowsky S.A., Lorenzen K., Versluis C., *et al.* Improving the Performance of a Quadrupole Time-of-Flight Instrument for Macromolecular Mass Spectrometry. *Anal Chem.* 2006;78:7473-83.
165. Tahallah N., Pinkse M., Maier C.S., Heck A.J. The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun Mass Spectrom.* 2001;15(8):596-601. Epub 2001/04/20. doi: 10.1002/rcm.275. PubMed PMID: 11312509.
166. Chambers T.J., Hahn C.S., Galler R., Rice C.M. Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol.* 1990;44:649-88. doi: 10.1146/annurev.mi.44.100190.003245. PubMed PMID: 2174669.
167. Gu B., Liu C., Lin-Goerke J., Maley D.R., Gutshall L.L., Feltenberger C.A., *et al.* The RNA helicase and nucleotide triphosphatase activities of the bovine viral diarrhea virus NS3 protein are essential for viral replication. *Journal of virology.* 2000;74(4):1794-800. doi: 10.1128/jvi.74.4.1794-1800.2000. PubMed PMID: 10644352; PubMed Central PMCID: PMC111657.
168. Brecher M., Zhang J., Li H. The flavivirus protease as a target for drug discovery. *Virol Sin.* 2013;28(6):326-36. doi: 10.1007/s12250-013-3390-x. PubMed PMID: 24242363; PubMed Central PMCID: PMC111657.
169. Shiryaev S.A., Thomsen E.R., Cieplak P., Chudin E., Cheltsov A.V., Chee M.S., *et al.* New details of HCV NS3/4A proteinase functionality revealed by a high-throughput cleavage assay. *PLoS One.* 2012;7(4):e35759. Epub 2012/05/05. doi: 10.1371/journal.pone.0035759. PubMed PMID: 22558217; PubMed Central PMCID: PMC111657.
170. Cho H.S., Ha N.C., Kang L.W., Chung K.M., Back S.H., Jang S.K., *et al.* Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *The Journal of biological chemistry.* 1998;273(24):15045-52. doi: 10.1074/jbc.273.24.15045. PubMed PMID: 9614113.
171. Khu Y.L., Koh E., Lim S.P., Tan Y.H., Brenner S., Lim S.G., *et al.* Mutations that affect dimer formation and helicase activity of the hepatitis C virus helicase. *Journal of virology.* 2001;75(1):205-14. Epub 2000/12/19. doi: 10.1128/JVI.75.1.205-214.2001. PubMed PMID: 11119590; PubMed Central PMCID: PMC113914.
172. Levin M.K., Patel S.S. The helicase from hepatitis C virus is active as an oligomer. *The Journal of biological chemistry.* 1999;274(45):31839-46. Epub 1999/11/05. doi: 10.1074/jbc.274.45.31839. PubMed PMID: 10542208.
173. Wolk B., Sansonno D., Krausslich H.G., Dammacco F., Rice C.M., Blum H.E., *et al.* Subcellular localization, stability, and trans-cleavage competence of the hepatitis C virus NS3-NS4A complex expressed in tetracycline-regulated cell lines. *Journal of virology.* 2000;74(5):2293-304. Epub 2000/02/09. doi: 10.1128/jvi.74.5.2293-2304.2000. PubMed PMID: 10666260; PubMed Central PMCID: PMC111711.
174. Nomura-Takigawa Y., Nagano-Fujii M., Deng L., Kitazawa S., Ishido S., Sada K., *et al.* Non-structural protein 4A of Hepatitis C virus accumulates on mitochondria and renders the cells prone to undergoing mitochondria-mediated apoptosis. *Journal of General Virology.* 2006;87(7):1935-45.
175. Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* 1989;17(12):4713-30.

- Epub 1989/06/26. doi: 10.1093/nar/17.12.4713. PubMed PMID: 2546125; PubMed Central PMCID: PMC318027.
176. Gundry R.L., White M.Y., Murray C.I., Kane L.A., Fu Q., Stanley B.A., *et al.* Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr Protoc Mol Biol.* 2009;Chapter 10:Unit10 25. doi: 10.1002/0471142727.mb1025s88. PubMed PMID: 19816929; PubMed Central PMCID: PMC2905857.
177. Olsen J.V., Ong S.E., Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics.* 2004;3(6):608-14. doi: 10.1074/mcp.T400003-MCP200. PubMed PMID: 15034119.
178. Laskay U.A., Lobas A.A., Srzentic K., Gorshkov M.V., Tsybin Y.O. Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments. *J Proteome Res.* 2013;12(12):5558-69. doi: 10.1021/pr400522h. PubMed PMID: 24171472.
179. Wiles T.A., Saba L.M., DeLong T. Peptide-Spectrum Match Validation with Internal Standards (P-VIS): Internally-Controlled Validation of Mass Spectrometry-Based Peptide Identifications. *J Proteome Res.* 2021;20(1):236-49. doi: 10.1021/acs.jproteome.0c00355. PubMed PMID: 32924495; PubMed Central PMCID: PMC7775876.
180. Lundgren D.H., Hwang S.I., Wu L., Han D.K. Role of spectral counting in quantitative proteomics. *Expert Rev Proteomics.* 2010;7(1):39-53. doi: 10.1586/ep.09.69. PubMed PMID: 20121475.
181. Hildonen S., Halvorsen T.G., Reubsæet L. Why less is more when generating tryptic peptides in bottom-up proteomics. *Proteomics.* 2014;14(17-18):2031-41. doi: 10.1002/pmic.201300479. PubMed PMID: 25044798.
182. Bittremieux W., Tabb D.L., Impens F., Staes A., Timmerman E., Martens L., *et al.* Quality control in mass spectrometry-based proteomics. *Mass Spectrom Rev.* 2018;37(5):697-711. doi: 10.1002/mas.21544. PubMed PMID: 28802010.
183. Cech N.B., Enke C.G. Relating electrospray ionization response to nonpolar character of small peptides. *Anal Chem.* 2000;72(13):2717-23. Epub 2000/07/25. doi: 10.1021/ac9914869. PubMed PMID: 10905298.
184. Adhikari S., Nice E.C., Deutsch E.W., Lane L., Omenn G.S., Pennington S.R., *et al.* A high-stringency blueprint of the human proteome. *Nat Commun.* 2020;11(1):5301. Epub 2020/10/18. doi: 10.1038/s41467-020-19045-9. PubMed PMID: 33067450; PubMed Central PMCID: PMC7568584.
185. Gatzeva-Topalova P.Z., May A.P., Sousa M.C. Structure and mechanism of ArnA: conformational change implies ordered dehydrogenase mechanism in key enzyme for polymyxin resistance. *Structure.* 2005;13(6):929-42. Epub 2005/06/09. doi: 10.1016/j.str.2005.03.018. PubMed PMID: 15939024; PubMed Central PMCID: PMC2997725.
186. Bornhorst J.A., Falke J.J. Purification of proteins using polyhistidine affinity tags. *Applications of Chimeric Genes and Hybrid Proteins Part A: Gene Expression and Protein Purification.* 2000:245-54. doi: 10.1016/s0076-6879(00)26058-8.
187. Goyal M., Chaudhuri T.K. GroEL-GroES assisted folding of multiple recombinant proteins simultaneously over-expressed in *Escherichia coli*. *Int J Biochem Cell Biol.* 2015;64:277-86. Epub 2015/05/11. doi: 10.1016/j.biocel.2015.04.018. PubMed PMID: 25957916.

188. Hagel L. Gel filtration: size exclusion chromatography. *Methods Biochem Anal.* 2011;54:51-91. Epub 2011/10/01. doi: 10.1002/9780470939932.ch3. PubMed PMID: 21954774.
189. Tucakov A.K. Gezielte Attenuierung und Markierung von Pestiviren - Charakterisierung eines Virulenzmarkers & Kartierung eines hochkonservierten Epitops: Ludwig-Maximilians-Universität München & Friedrich-Loeffler Instituts Bundesforschungsinstitut für Tiergesundheit; 2016.
190. Gu M., Rice C.M. The Spring alpha-Helix Coordinates Multiple Modes of HCV (Hepatitis C Virus) NS3 Helicase Action. *The Journal of biological chemistry.* 2016;291(28):14499-509. Epub 2016/05/27. doi: 10.1074/jbc.M115.704379. PubMed PMID: 27226535; PubMed Central PMCID: PMC4938173.
191. Beran R.K., Pyle A.M. Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase. *The Journal of biological chemistry.* 2008;283(44):29929-37. Epub 2008/08/30. doi: 10.1074/jbc.M804065200. PubMed PMID: 18723512; PubMed Central PMCID: PMC2573085.
192. Roskoski R. Michaelis-Menten Kinetics☆: Elsevier; 2015. doi:10.1016/B978-0-12-801238-3.05143-6.
193. Sikora B., Chen Y., Lichti C.F., Harrison M.K., Jennings T.A., Tang Y., *et al.* Hepatitis C virus NS3 helicase forms oligomeric structures that exhibit optimal DNA unwinding activity in vitro. *The Journal of biological chemistry.* 2008;283(17):11516-25. Epub 2008/02/20. doi: 10.1074/jbc.M708125200. PubMed PMID: 18283103; PubMed Central PMCID: PMC2431078.
194. Wiedemann C., Kumar A., Lang A., Ohlenschläger O. Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. *Frontiers in chemistry.* 2020;8:280. Epub 2020/05/12. doi: 10.3389/fchem.2020.00280. PubMed PMID: 32391319; PubMed Central PMCID: PMC7191308.
195. Luo Y., Matejic T., Ng C.-K., Nunnally B., Porter T., Raso S., *et al.* Characterization and Analysis of Biopharmaceutical Proteins. In: Ahuja S, S. S, editors. *Handbook of Modern Pharmaceutical Analysis.* 102011. p. 283-359.
196. Christis C., Lubsen N.H., Braakman I. Protein folding includes oligomerization - examples from the endoplasmic reticulum and cytosol. *FEBS J.* 2008;275(19):4700-27. Epub 2008/08/06. doi: 10.1111/j.1742-4658.2008.06590.x. PubMed PMID: 18680510.
197. Springer S., Malkus P., Borchert B., Wellbrock U., Duden R., Schekman R. Regulated oligomerization induces uptake of a membrane protein into COPII vesicles independent of its cytosolic tail. *Traffic.* 2014;15(5):531-45. Epub 2014/02/01. doi: 10.1111/tra.12157. PubMed PMID: 24479619.
198. Preugschat F., Danger D.P., Carter L.H., 3rd, Davis R.G., Porter D.J. Kinetic analysis of the effects of mutagenesis of W501 and V432 of the hepatitis C virus NS3 helicase domain on ATPase and strand-separating activity. *Biochemistry.* 2000;39(17):5174-83. Epub 2000/05/23. doi: 10.1021/bi9923860. PubMed PMID: 10819985.
199. Tackett A.J., Chen Y., Cameron C.E., Raney K.D. Multiple full-length NS3 molecules are required for optimal unwinding of oligonucleotide DNA in vitro. *The Journal of biological chemistry.* 2005;280(11):10797-806. Epub 2005/01/07. doi: 10.1074/jbc.M407971200. PubMed PMID: 15634684.

200. Boeri Erba E., Barylyuk K., Yang Y., Zenobi R. Quantifying protein-protein interactions within noncovalent complexes using electrospray ionization mass spectrometry. *Anal Chem.* 2011;83(24):9251-9. doi: 10.1021/ac201576e. PubMed PMID: 22047453.
201. Riedel C., Aitkenhead H., El Omari K., Rumenapf T. Atypical Porcine Pestiviruses: Relationships and Conserved Structural Features. *Viruses.* 2021;13(5). Epub 2021/05/01. doi: 10.3390/v13050760. PubMed PMID: 33926056; PubMed Central PMCID: PMC8146772.
202. Zheng F., Lu G., Li L., Gong P., Pan Z. Uncoupling of Protease trans-Cleavage and Helicase Activities in Pestivirus NS3. *Journal of virology.* 2017;91(21). doi: 10.1128/JVI.01094-17. PubMed PMID: 28835495; PubMed Central PMCID: PMC5640827.
203. Perez S., Eichhorn P., Barcelo D., 2016 J. *Advances in Time-of-Flight Mass Spectrometry. Applications of Time-of-Flight and Orbitrap Mass Spectrometry in Environmental, Food, Doping, and Forensic Analysis: Elsevier; 2016.*
204. Donnelly D.P., Rawlins C.M., DeHart C.J., Fornelli L., Schachner L.F., Lin Z., *et al.* Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods.* 2019;16(7):587-94. doi: 10.1038/s41592-019-0457-0. PubMed PMID: 31249407; PubMed Central PMCID: PMC6719561.
205. Tolic L.P., Bruce J.E., Lei Q.P., Anderson G.A., Smith R.D. In-trap cleanup of proteins from electrospray ionization using soft sustained off-resonance irradiation with fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem.* 1998;70(2):405-8. doi: 10.1021/ac970828c. PubMed PMID: 9450366.
206. Sobott F., Robinson C.V. Characterising electrosprayed biomolecules using tandem-MS-the noncovalent GroEL chaperonin assembly. *International Journal of Mass Spectrometry.* 2004;236(1-3):25-32. doi: 10.1016/j.ijms.2004.05.010.
207. Laganowsky A., Reading E., Hopper J.T., Robinson C.V. Mass spectrometry of intact membrane protein complexes. *Nat Protoc.* 2013;8(4):639-51. doi: 10.1038/nprot.2013.024. PubMed PMID: 23471109; PubMed Central PMCID: PMC4058633.
208. Ilag L.L., Westblade L.F., Deshayes C., Kolb A., Busby S.J., Robinson C.V. Mass spectrometry of Escherichia coli RNA polymerase: interactions of the core enzyme with sigma70 and Rsd protein. *Structure.* 2004;12(2):269-75. doi: 10.1016/j.str.2004.01.007. PubMed PMID: 14962387.
209. Konermann L. Addressing a Common Misconception: Ammonium Acetate as Neutral pH "Buffer" for Native Electrospray Mass Spectrometry. *J Am Soc Mass Spectrom.* 2017;28(9):1827-35. doi: 10.1007/s13361-017-1739-3. PubMed PMID: 28710594.
210. Iavarone A.T., Udekwu O.A., Williams E.R. Buffer loading for counteracting metal salt-induced signal suppression in electrospray ionization. *Anal Chem.* 2004;76(14):3944-50. doi: 10.1021/ac049724+. PubMed PMID: 15253628; PubMed Central PMCID: PMC1343452.
211. Burley S.K., Petsko G.A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science.* 1985;229(4708):23-8. doi: 10.1126/science.3892686. PubMed PMID: 3892686.
212. Dill K.A. Dominant forces in protein folding. *Biochemistry.* 1990;29(31):7133-55. doi: 10.1021/bi00483a001. PubMed PMID: 2207096.

213. Madhusudan Makwana K., Mahalakshmi R. Implications of aromatic-aromatic interactions: From protein structures to peptide models. *Protein Sci.* 2015;24(12):1920-33. doi: 10.1002/pro.2814. PubMed PMID: 26402741; PubMed Central PMCID: PMC4815235.
214. Lioe H., O'Hair R.A. Comparison of collision-induced dissociation and electron-induced dissociation of singly protonated aromatic amino acids, cystine and related simple peptides using a hybrid linear ion trap-FT-ICR mass spectrometer. *Anal Bioanal Chem.* 2007;389(5):1429-37. doi: 10.1007/s00216-007-1535-1. PubMed PMID: 17874085.
215. El Aribi H., Orlova G., Hopkinson A.C., Siu K.W.M. Gas-Phase Fragmentation Reactions of Protonated Aromatic Amino Acids: Concomitant and Consecutive Neutral Eliminations and Radical Cation Formations. *The Journal of Physical Chemistry A.* 2004;108(17):3844-53. doi: 10.1021/jp0374915.
216. Zhang P., Chan W., Ang I.L., Wei R., Lam M.M.T., Lei K.M.K., *et al.* Revisiting Fragmentation Reactions of Protonated alpha-Amino Acids by High-Resolution Electrospray Ionization Tandem Mass Spectrometry with Collision-Induced Dissociation. *Sci Rep.* 2019;9(1):6453. doi: 10.1038/s41598-019-42777-8. PubMed PMID: 31015571; PubMed Central PMCID: PMC6478932.
217. Chattopadhyay S., Chen Y., Weller S.K. The two helicases of herpes simplex virus type 1 (HSV-1). *Front Biosci.* 2006;11:2213-23. Epub 2006/05/25. doi: 10.2741/1964. PubMed PMID: 16720308.
218. Gai D., Zhao R., Li D., Finkelstein C.V., Chen X.S. Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell.* 2004;119(1):47-60. Epub 2004/09/30. doi: 10.1016/j.cell.2004.09.017. PubMed PMID: 15454080.
219. Gai D., Wang D., Li S.X., Chen X.S. The structure of SV40 large T hexameric helicase in complex with AT-rich origin DNA. *Elife.* 2016;5. Epub 2016/12/07. doi: 10.7554/eLife.18129. PubMed PMID: 27921994; PubMed Central PMCID: PMC5140265.
220. Hill C.P., Yee J., Selsted M.E., Eisenberg D. Crystal structure of defensin HNP-3, an amphiphilic dimer: mechanisms of membrane permeabilization. *Science.* 1991;251(5000):1481-5. Epub 1991/03/22. doi: 10.1126/science.2006422. PubMed PMID: 2006422.
221. Khakshoor O., Lin A.J., Korman T.P., Sawaya M.R., Tsai S.C., Eisenberg D., *et al.* X-ray crystallographic structure of an artificial beta-sheet dimer. *J Am Chem Soc.* 2010;132(33):11622-8. Epub 2010/07/31. doi: 10.1021/ja103438w. PubMed PMID: 20669960; PubMed Central PMCID: PMC2929266.
222. Daniel J.M., Friess S.D., Rajagopalan S., Wendt S., Zenobi R. Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry. *International Journal of Mass Spectrometry.* 2002;216(1):1-27. doi: 10.1016/s1387-3806(02)00585-7.
223. Urbani A., Bianchi E., Narjes F., Tramontano A., De Francesco R., Steinkuhler C., *et al.* Substrate specificity of the hepatitis C virus serine protease NS3. *The Journal of biological chemistry.* 1997;272(14):9204-9. Epub 1997/04/04. doi: 10.1074/jbc.272.14.9204. PubMed PMID: 9083052.
224. Shoji I., Suzuki T., Chieda S., Sato M., Harada T., Chiba T., *et al.* Proteolytic activity of NS3 serine proteinase of hepatitis C virus efficiently expressed in *Escherichia coli*. *Hepatology.* 1995;22(6):1648-55. doi: 10.1002/hep.1840220606.

225. Shoji I., Suzuki T., Sato M., Aizaki H., Chiba T., Matsuura Y., *et al.* Internal processing of hepatitis C virus NS3 protein. *Virology*. 1999;254(2):315-23. Epub 1999/02/13. doi: 10.1006/viro.1998.9540. PubMed PMID: 9986797.
226. Teo K.F., Wright P.J. Internal proteolysis of the NS3 protein specified by dengue virus 2. *J Gen Virol*. 1997;78 (Pt 2):337-41. Epub 1997/02/01. doi: 10.1099/0022-1317-78-2-337. PubMed PMID: 9018055.
227. Leung D., Schroder K., White H., Fang N.X., Stoermer M.J., Abbenante G., *et al.* Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors. *The Journal of biological chemistry*. 2001;276(49):45762-71. Epub 2001/10/03. doi: 10.1074/jbc.M107360200. PubMed PMID: 11581268.
228. Kurane I., Brinton M.A., Samson A.L., Ennis F.A. Dengue virus-specific, human CD4+ CD8- cytotoxic T-cell clones: multiple patterns of virus cross-reactivity recognized by NS3-specific T-cell clones. *Journal of virology*. 1991;65(4):1823-8. Epub 1991/04/01. doi: 10.1128/JVI.65.4.1823-1828.1991. PubMed PMID: 1705990; PubMed Central PMCID: PMCPMC239991.
229. Collen T., Morrison W.I. CD4+ T-cell responses to bovine viral diarrhoea virus in cattle. *Virus Research*. 2000;67(1):67-80. doi: 10.1016/s0168-1702(00)00131-3.
230. Riitho V., Walters A.A., Somavarapu S., Lamp B., Rumenapf T., Krey T., *et al.* Design and evaluation of the immunogenicity and efficacy of a biomimetic particulate formulation of viral antigens. *Sci Rep*. 2017;7(1):13743. Epub 2017/10/25. doi: 10.1038/s41598-017-13915-x. PubMed PMID: 29062078; PubMed Central PMCID: PMCPMC5653838.
231. Riitho V., Strong R., Larska M., Graham S.P., Steinbach F. Bovine Pestivirus Heterogeneity and Its Potential Impact on Vaccination and Diagnosis. *Viruses*. 2020;12(10). Epub 2020/10/11. doi: 10.3390/v12101134. PubMed PMID: 33036281; PubMed Central PMCID: PMCPMC7601184.
232. Taylor W.R. The classification of amino acid conservation. *Journal of Theoretical Biology*. 1986;119(2):205-18. doi: 10.1016/s0022-5193(86)80075-3.
233. Di Cera E. Serine proteases. *IUBMB Life*. 2009;61(5):510-5. Epub 2009/01/31. doi: 10.1002/iub.186. PubMed PMID: 19180666; PubMed Central PMCID: PMCPMC2675663.
234. Nienaber V.L., Breddam K., Birktoft J.J. A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry*. 1993;32(43):11469-75. Epub 1993/11/02. doi: 10.1021/bi00094a001. PubMed PMID: 8105890.
235. Günther S., Reinke P.Y.A., Fernandez-Garcia Y., Lieske J., Lane T.J., Ginn H.M., *et al.* X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science*. 2021;372(6542):642-6. Epub 2021/04/04. doi: 10.1126/science.abf7945. PubMed PMID: 33811162; PubMed Central PMCID: PMCPMC8224385.
236. Zechel D.L., Konermann L., Withers S.G., Douglas D.J. Pre-steady state kinetic analysis of an enzymatic reaction monitored by time-resolved electrospray ionization mass spectrometry. *Biochemistry*. 1998;37(21):7664-9. Epub 1998/06/20. doi: 10.1021/bi980445o. PubMed PMID: 9601025.
237. Ge X., Sirich T.L., Beyer M.K., Desaire H., Leary J.A. A strategy for the determination of enzyme kinetics using electrospray ionization with an ion trap mass spectrometer. *Anal Chem*. 2001;73(21):5078-82. Epub 2001/11/28. doi: 10.1021/ac0105890. PubMed PMID: 11721902.

238. Rothan H.A., Han H.C., Ramasamy T.S., Othman S., Rahman N.A., Yusof R. Inhibition of dengue NS2B-NS3 protease and viral replication in Vero cells by recombinant retrocyclin-1. *BMC Infect Dis.* 2012;12:314. Epub 2012/11/23. doi: 10.1186/1471-2334-12-314. PubMed PMID: 23171075; PubMed Central PMCID: PMC3575220.
239. Beran R.K., Serebrov V., Pyle A.M. The serine protease domain of hepatitis C viral NS3 activates RNA helicase activity by promoting the binding of RNA substrate. *The Journal of biological chemistry.* 2007;282(48):34913-20. Epub 2007/10/09. doi: 10.1074/jbc.M707165200. PubMed PMID: 17921146.
240. Mao S.S., DiMuzio J., McHale C., Burlein C., Olsen D., Carroll S.S. A time-resolved, internally quenched fluorescence assay to characterize inhibition of hepatitis C virus nonstructural protein 3-4A protease at low enzyme concentrations. *Anal Biochem.* 2008;373(1):1-8. Epub 2007/11/21. doi: 10.1016/j.ab.2007.10.041. PubMed PMID: 18022380.

GHS information

Tab. 19 Globally Harmonized System of Classification and Labelling of Chemicals (GHS) used in the present thesis.

IUPAC Name	CAS No.	Chem. Form.	GHS (H)	GHS (P)	Coding (GHS)
Isopropyl- β -D-thiogalactopyranosid	367-93-1 7601-54-9	C ₉ H ₁₈ O ₅ S	-	-	-
Sodium phosphate	15819-50-8 (Hexahydrate)	Na ₃ PO ₄	315-319-335	261-305+351+338	GHS07
Sodium chloride	10101-89-0 (Dodecahydrate) 7647-14-5	NaCl	-	-	-
Imidazoles	288-32-4	C ₃ H ₄ N ₂	360D-302-314	201-280-301+330+331-305+351+338-308+310	GHS05 GHS08 GHS07
Ammonium acetate	631-61-8	C ₂ H ₇ NO ₂	-	-	-
XT MOPS Running Buffer	1132-61-2 151-21-3		315 319 335	261 280 304+340 305+351+338 405 501	GHS07
Acrylamide (CMR-Class)	79-06-1	C ₃ H ₅ NO	301-312-332-315-317-319-340-350-361f-372	201-280-302+352-304+340-305+351+338-308+310	GHS06 GHS08
Coomassie Brilliant Blue R-250	6104-59-2	C ₄₅ H ₄₄ N ₃ NaO ₇ S ₂	-	-	-
Ethanol	64-17-5	C ₂ H ₆ O	225-319	210-240-305+351+338-403+233	GHS02 GHS07
Acetic acid	64-19-7	C ₂ H ₄ O ₂	226-314	210-280-301+330+331-303+361+353-305+351+338	GHS02 GHS05
Methanol	67-56-1	CH ₄ O	225-331-311-301-370	210-233-280-302+352-304+340-308+310-403+235	GHS02 GHS06 GHS08
Tris(hydroxymethyl) Aminomethan	77-86-1 1185-53-1 (Hydrochloride)	C ₄ H ₁₁ NO ₃	315-319-335	261-305+351+338	GHS07

IUPAC Name	CAS No.	Chem. Form.	GHS (H)	GHS (P)	Coding (GHS)
Sodium Dodecyl Sulfate	151-21-3	C ₁₂ H ₂₅ NaO ₄ S	228-302+332-315-318-335-412	210-261-280-301+312+330-305+351+338+310-370+378	GHS02 GHS05 GHS07
Ammonium peroxodisulfate	7727-54-0	(NH ₄) ₂ S ₂ O ₈	272-302-315-317-319-334-335	220-261-280-305+351+338-342+311	GHS03 GHS07 GHS08
Tetramethylethylenediamine	110-18-9	C ₆ H ₁₆ N ₂	225-332-302-314	210-280-305+351+338-310	GHS02 GHS05 GHS07
Glycerin	56-81-5	C ₃ H ₈ O ₃	-	-	-
Mercaptoethanol	60-24-2	C ₂ H ₆ OS	301+331-310-315-317-318-361d-373-410	201-262-280-301+310+330-302+352+310-305+351+338+310	GHS05 GHS06 GHS08 GHS09
Bromphenolblue	115-39-9 (acid) 62625-28-9 (Sodium)	C ₁₉ H ₁₀ Br ₄ O ₅ S	-	-	-
Glycine	56-40-6	C ₂ H ₅ NO ₂	-	-	-
Polysorbate 20	9005-64-5	C ₅₈ H ₁₁₄ O ₂₆	-	-	-
2- (4- (2-Hydroxyethyl) -1-piperazinyl) ethanesulfonic acid	7365-45-9	C ₈ H ₁₈ N ₂ O ₄ S	-	-	-
Dithiothreitol	7634-42-6 (un-specified)	(un- C ₄ H ₁₀ O ₂ S ₂	302-315-319-335	261-305+351+338260-280-303+361+353+	GHS07
Hydrochloric acid	7647-01-0	HCL	280-331-314	315-304+340+315-305+351+338+315-403-405	GHS05 GHS07

6 Supplement

6.1 CSFV aa-sequence further information

**S. Fig. 1 Amino acid sequence of the CSFV_{WT} NS3-NS4A (single chain protease, SCP).**

Highlighted: (N-term. to C-term.) MAS-(10 x)His = 1677.71 Da, TEV (ENLYFQG) = 869.39 Da, NS4A(37aa) Cofactor = 4260.19 Da, GSGS-linker = 306.12 Da, NS3-serine-protease = 20681.35 Da, NS3-helicase: 55017.03 Da, NS4A(8aa) = 817.42 Da, (C = cysteines, lead to protein dimerization). Hypothesis: sequences in green (NS4A-kink domain) should be more exposed in the mutants comparing to the wild-type. Highlighted aa: L (Leucine : Arginine) L45A and Y (Tyrosine : Arginine) Y47A mutated in bold and double mutant LY45/47AA in NS4A cofactor region.

Monoisotopic mass (MW) = 83521.14 Da, averaged mass (MW) = 83558.34564 Da (complete), theoretical pI= 7.11/6.66 (<https://web.expasy.org>, <http://isoelectric.org>).

6.2 CSFV protein batches 1 to 4

6.2.1 Batch 1 NS3-NS4A proteins

S. Tab. 1 MS1 based molecular weight (MW) of the batch 1 (B) monomers.

Monomer mass is based on the identified and averaged peak intensities of the charge states (18+ to 16+) in positive ESI-MS ion mode (Av. = averaged). Values are based on 10 eV energy.

Batch (1) monomers	Charge	m/z	Monomer (Da)
B1(a)	18+	4631.89	83374
	17+	4904.35	83374
	16+	5210.88	83374
Av. mass			83374
B1(b)	17+	4798.59	81576
	16+	5098.37	81574
Av. mass			81575
B1(c)	17+	4774.19	81161
	16+	5069.03	81104
Av. mass			81133

*The m/z -values were selected before, but close to the peak top of every charge state. Species: B1(a), B1(b), B1(c).

S. Tab. 2 MS1 based molecular weight (MW) of the batch 1 (B1) dimers.

Dimer mass is based on the identified and averaged peak intensities of the charge states (26+ to 22+) in positive ESI-MS ion mode. Values are based on 10 eV energy, (CM = calculated monomer mass, Av. = averaged).

Batch (1) dimers	Charge	m/z	Dimer (Da)	CM (Da)
B1d(a)	26+	6415.14	166794	83397
	25+	6671.92	166798	83399
	24+	6949.41	166786	83393
	23+	7251.57	166786	83393
	22+	7581.64	166796	83398
Av. mass			166792	83396
B1d(b)	25+	6604.38	165110	82555
	24+	6874.63	164991	82496
	23+	7170.20	164915	82457
Av. mass			165005	82503
B1d(c)	24+	6855.15	164524	82262
	23+	7150.30	164457	82228
Av. mass			164491	82245
B1d(d)	23+	7121.50	163737	81869

*The m/z -values were selected before, but close the peak top of every charge state. The calculated monomer (CM) is half the dimer mass.

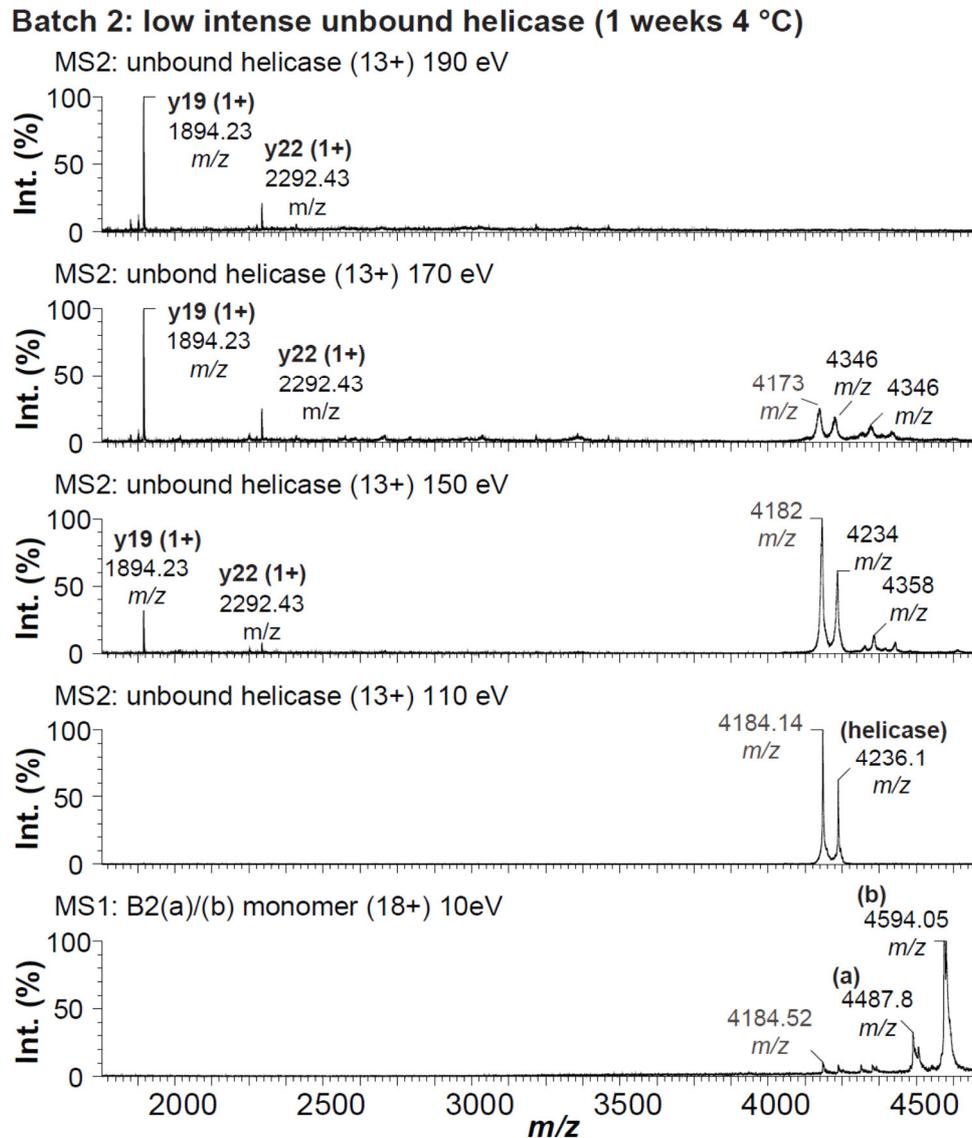
6.2.2 Batch 2 NS3-NS4A proteins

S. Tab. 3 MS1 based molecular weight (MW) of the batch 2 (B2) main monomers and dimers.

Masses are based on the identified and averaged peak intensities of the charge states (26+ to 23+) in positive ESI-MS ion mode. Values are based on 10 eV energy, CM = calculated monomer mass, Av. = averaged. MS1 energy was set to 10 eV and 195 eV, ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 8 μ M.

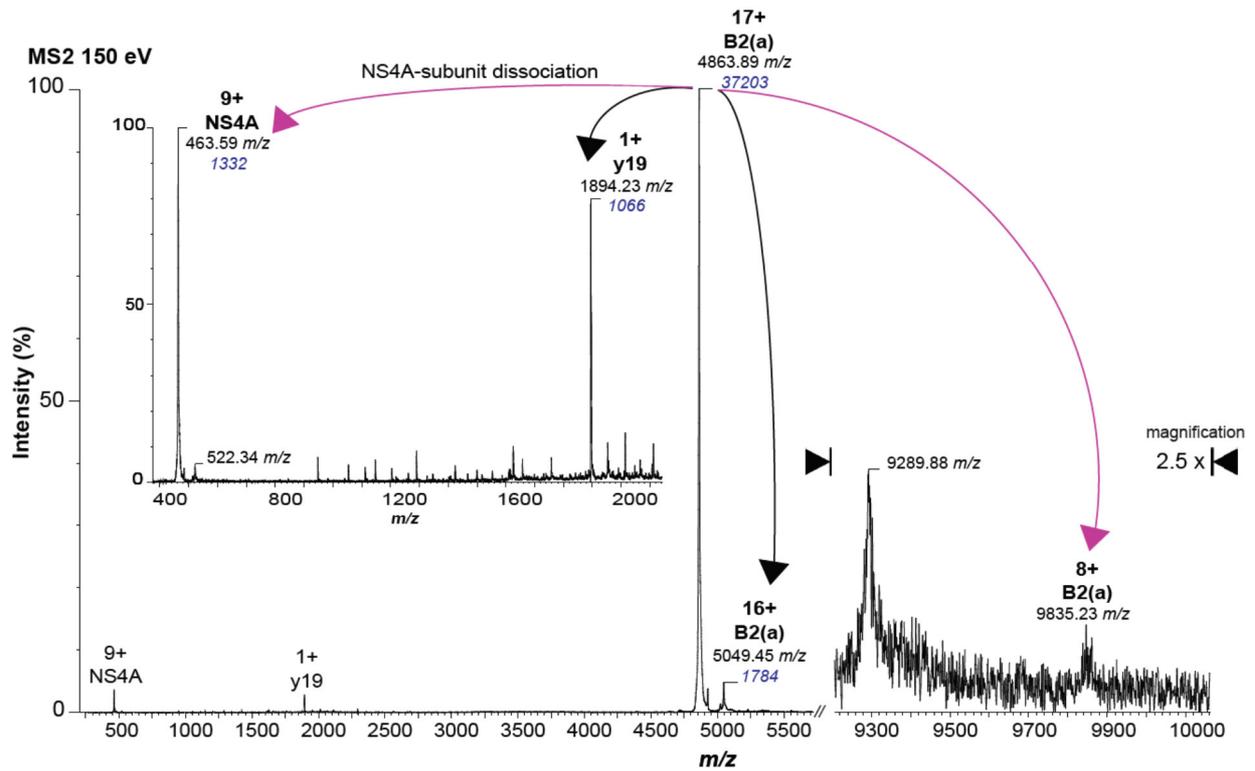
Batch (2) di- mers MS1 (10 eV)	Charge	m/z	Dimer (Da)	Calculated monomer (Da)	10 eV vs 195 eV dimer shift	
					m/z	(Da)
B2d(a)	25+	6671.42	166786	83393	19.20	480
	24+	6921.20	166109	83054	22.01	528
	23+	7192.01	165416	82708	26.91	619
Av. mass			166104	83052		
B2d(b)*	26+	6365.54	165504	82752	5.63	146
	25+	6620.10	165503	82751	7.17	180
Av. mass			165503	82751		
B2d(c)	25+	6600.49	165012	82506	16.72	418
	24+	6846.03	164305	82152	17.03	409
	23+	7111.91	163574	81787	18.35	422
Av. mass			164297	82149		
B2d(d)*	26+	6218.08	161670	80835	4.62	120.08
	25+	6464.97	161624	80812	2.85	71.25
	24+	6737.90	161710	80855	6.28	150.60
Av. mass			161667	80834		
MS1 (195 eV)	Charge	m/z	Dimer (Da) vs. Mo- nomer (Da)			
B2d(a)	25+	6652.22	166306	83153		
	24+	6899.20	165581	82790		
	23+	7165.10	164797	82399		
Av. mass			165562	82781		
B2d(b)*	26+	6359.91	165358	82679		
	25+	6612.93	165323	82662		
	24+	6889.42	165346	82673		
	23+	7190.02	165370	82685		
Av. mass			165350	82675		
B2d(c)	25+	6583.77	164594	82297		
	24+	6829.00	163896	81948		
	23+	7093.56	163152	81576		
Av. mass			163881	81941		
B2d(d)*	26+	6213.44	161550	80775		*Because the dimer peaks are too broad in shape, the m/z values were selected based on every averaged peak (peak top).
	25+	6462.13	161553	80777		
	24+	6731.63	161559	80780		
	23+	7024.36	161560	80780		
Av. mass			161558	80779		

6.3 Top-Down identification of the released NS3-helicase, the NS4A cofactor and various sub-monomers of the WT protein



S. Fig. 2 Low amount of free NS3-helicase in batch 2 (after 1 week at + 4 °C) was identified in MS1 mode at 10 eV and in MS2 mode while applying different CID energies.

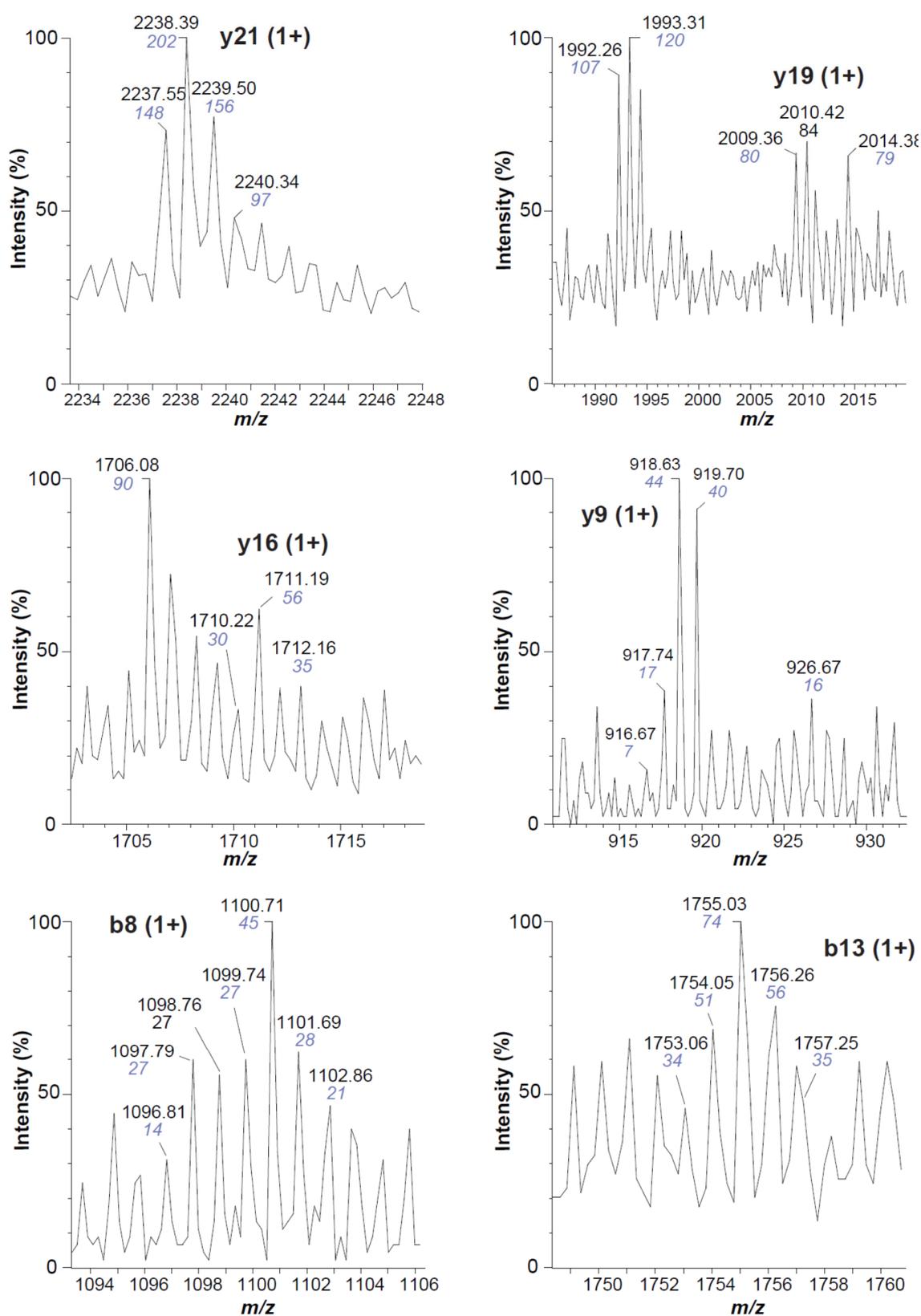
The figure is a supplement to sub-chapter 3.3.1 and 3.3.2. MS1 energy was set to 10 eV, (a) = B2(a) monomer, (b) = B2(b) monomer both indicate the charge state 18+. MS2 precursor was 13+ of the helicase, energy was set from 110 eV to 190 eV, ammonium acetate 200 mM at pH7.2, 2 mM DTT, protein concentration was 8 μ M.



S. Fig. 3 Dissociation of the NS4A-cofactor from the NS3-NS4A complex is shown by CID fragmentation of the 17+ precursor peak of B2(a) in MS2 mode.

The figure is a supplement to sub-chapter 3.3.2. MS2 precursor = 17+ of B2(a) (batch 2 monomer). Energy was set to 150 eV. The 9+ peak of NS4A (4.2 kDa) and the corresponding 8+ peak (78.7 kDa) are marked with pink-coloured arrows, the C-terminal y19-fragment ion (1+, 1894 Da) of the NS3-helicase subunit and the corresponding 16+ peak (80.8 kDa) are labelled with black arrows. Ammonium acetate 200 mM at pH7.2, 2 mM DTT, protein concentration was 8 μ M.

Batch 1 B1(b) and B1(c) MS2 ions at 195 eV

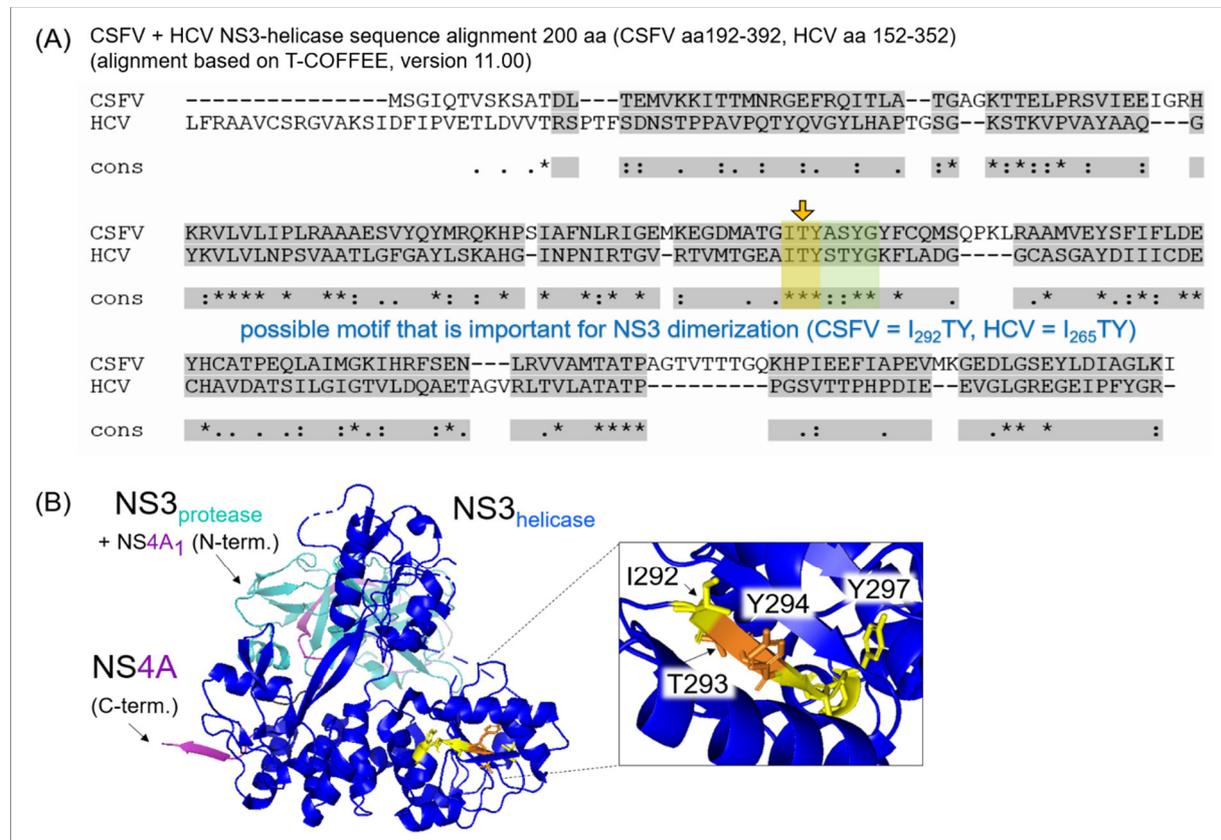


S. Fig. 4 Identification of B1(b) and B1(c) MS2 y- and b-ions at 195 eV.

The figure is a supplement to chapter 3.3.4. B1 monomers were hidden under the main peak (17+, MS2 precursor of B2(b) monomer). Ammonium acetate 200 mM at pH7.2, 2 mM DTT, protein concentration was 6 μ M.

6.4 Stoichiometry of monomers and dimers

6.4.1 Dimerization of the NS3-NS4A protein



S. Fig. 5 Possible location of the NS3 dimerization region at an exposed beta sheet structure (I₂₉₂TY).

(A) Amino acid sequence alignment of the CSFV and HCV NS3-helicase (residues starting at the N-terminal end of the NS3-helicase aa192 to aa392 CSFV and aa152 to aa352 HCV). Similar dimerization motif was found for HCV at I₂₆₅TY. (Fig. adjusted according to PDB accession code: 5MZ4, [103]).

6.4.2 Batch 1 and 2 NS3-NS4A WT protein

S. Tab. 4 MS1 based ratio between B1(a)-(c) monomers (MS1 10 eV).

The table is a supplement to chapter 3.4.1. Calculations are based on three independent measurements (M1 to M3), $\sum_{(av.)}$ = averaged values, $\sum_{(av.)} \text{int.}_{(z)}$ = av. peak intensities (int.) of charge states (z), ratios are given as (%). Ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 7 μM .

10 eV	B1(a)		B1(b)			B1(c)		
M1 Charge	Int.	$\sum_{(av.)} \text{int.}_{(z)}$	M1 Charge	Int.	$\sum_{(av.)} \text{int.}_{(z)}$	M1 Charge	Int.	$\sum_{(av.)} \text{int.}_{(z)}$
18+	161	183.667	–		24.500	–		18.000
17+	279		17+	29		17+	20	
16+	111		16+	20		16+	16	
M2 Charge			M2 Charge			M2 Charge		
18+	144	177.333	–		21.500	–		15.000
17+	272		17+	27		17+	17	
16+	116		16+	16		16+	13	
M3 Charge			M3 Charge			M3 Charge		
18+	120	140.333	–		21.000	–		13.000
17+	210		17+	24		17+	14	
16+	91		16+	18		16+	12	
$\sum_{\text{M1-M3}}$		167.111			22.333			15.333
SD		19.110			1.546			2.055
Ratio (%)		82			11			7
(a)-(b). (a)-(c)		–			71			74

*(Ratio = calculation with respect to the total amount of monomer in batch 1. SD = standard deviation).

S. Tab. 5 Ratio and distribution of the batch 1 (B1) and batch 2 (B2) monomers at 10 eV.

The table is a supplement to chapter 3.4.1. Data evaluation is based on averaged peak intensities of the presented charge states and of triplicated measurements. Intensities (Int.) are given in percentage with associated standard deviation of the values (SD). MS1 energy was set to 10 eV, ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 7 μM B1 and 8 μM B2. Averaged peak intensities (int.) of associated charge states are shown in: *Fig. 70 The ratio between the identified batch 1 and 2 monomers shows the hierarchical distribution of the different species.*

10 eV Species	Int. (%)	SD	10 eV Species	Int. (%)	SD	10 eV Species	Int. (%)	SD
B1(a)	82	19.11	B2(a)	32.24	7.84	B2(i)	3.49	0.33
B1(b)	11	1.55	B2(b)	18.14	4.75	B2(g)	1.96	0.27
B1(c)	7	2.06	B2(d)	17.19	0.12	B2(c)	1.78	0.14
–	–	–	B2(e)	13.75	1.10	B2(j)	1.69	0.30
–	–	–	B2(h)	8.11	0.85	B2(f)	1.67	0.51

S. Tab. 6 Comparison of peak intensities and peak areas of B1(a) monomer and B1d(a) dimer show similar ratios for the monomer and dimer at 10 eV in MS1 mode.

The table is a supplement to chapter 3.4.2. Data were normalized to the highest absolute peak signal (25+ protein dimer), triplicate measurements were performed. $\sum_{(av.)}$ = averaged values of M1-M3 and related standard deviation (*SD*) are shown. Ratios of monomer and dimer are indicated. Ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 7 μ M. Averaged (*av.*) peak intensities (Int.) or peak areas (PA) of associated charge states are shown in: *Fig. 71 Huge extent of dimerization of batch 1 monomers.*

B1(a) monomer	Charge	Intensity M1	Intensity M2	Intensity M3	$\sum_{(av.)}M1-M3$ (%)	SD
83.4 kDa	19+	0.016	0.013	0.021	2	0.3
	18+	0.064	0.066	0.099	8	1.6
	17+	0.218	0.182	0.242	21	2.4
	16+	0.093	0.082	0.114	10	1.3
	15+	0.01	0.012	0.017	1	0.3
$\sum_{(av.)}Int.$					8	
dimer	Charge	M1	M2	M3	$\sum_{(av.)}M1-M3$	SD
166.8 kDa	27+	0.141	0.113	0.101	12	1.7
	26+	0.559	0.505	0.499	52	2.7
	25+	1	1	1	100	0
	24+	0.862	0.928	0.971	92	4.5
	23+	0.279	0.336	0.368	33	3.7
	22+	0.033	0.043	0.054	4	0.9
$\sum_{(av.)}Int.$					49	
B1(a) monomer	Charge	PA M1	PA M2	PA M3	$\sum_{(av.)}M1-M3$ (%)	SD
83.4 kDa	19+	0.015	0.012	0.021	2	0.4
	18+	0.087	0.062	0.099	8	1.5
	17+	0.166	0.144	0.212	17	2.8
	16+	0.072	0.068	0.092	8	1.1
	15+	0.007	0.007	0.014	1	0.3
$\sum_{(av.)}PA$					7	
dimer	Charge	M1	M2	M3	$\sum_{(av.)}M1-M3$	SD
166.8 kDa	27+	0.108	0.07	0.077	9	1.7
	26+	0.452	0.421	0.441	44	1.3
	25+	1	1	1	100	0
	24+	0.882	1.00	1.105	100	9.1
	23+	0.318	0.358	0.465	38	6
	22+	0.04	0.05	0.069	5	1
$\sum_{(av.)}PA$					49	
(%)	monomer		dimer		SD _m	SD _d
Ratio _{Int.}	14		86		2.1	3.9
Ratio _{PA}	13		87		2.2	5.7

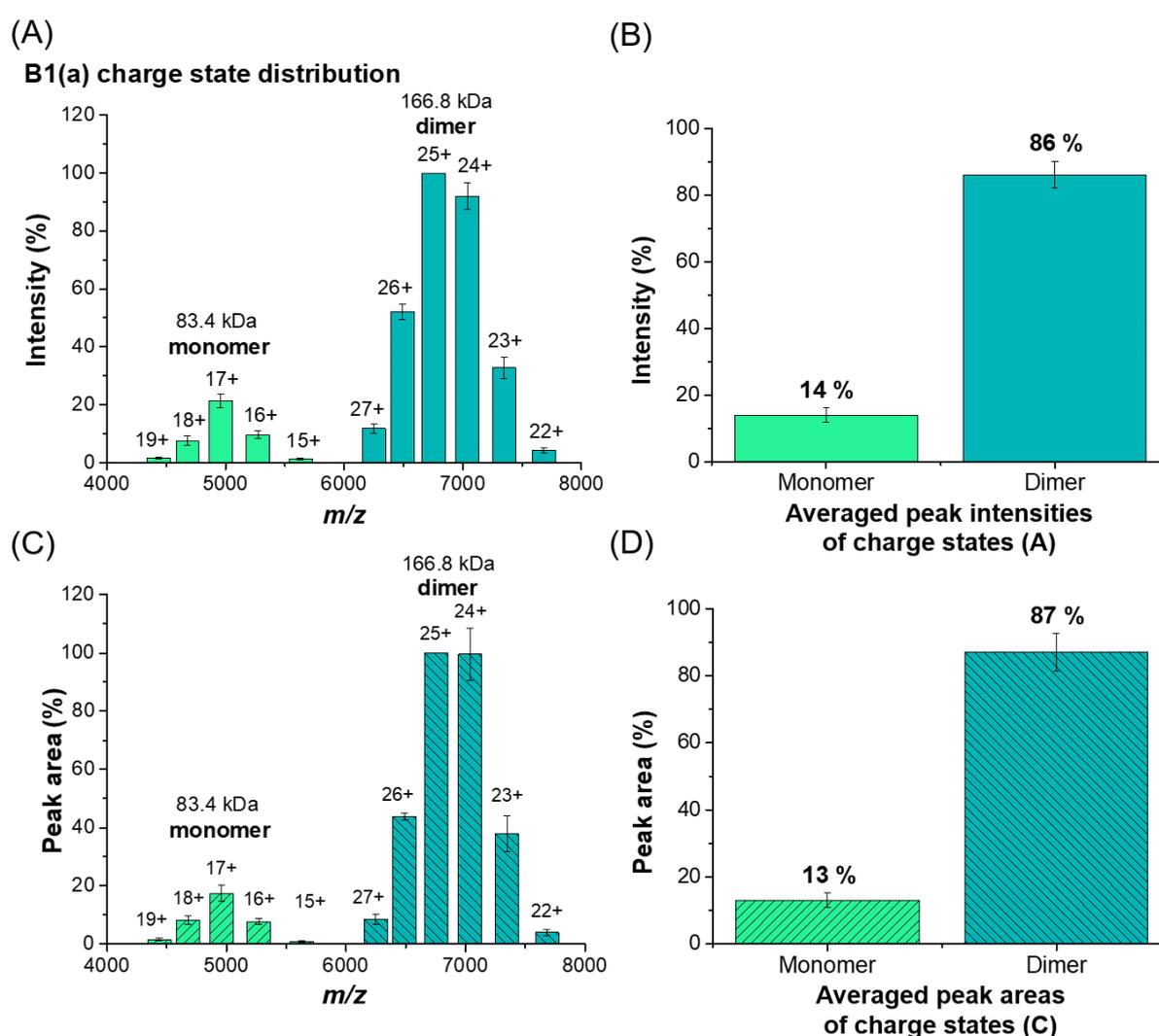
S. Tab. 7 Dimerization of B1 proteins shows a lack of different *C*- and *N*-terminal amino acids based on their specific peak distances in MS1 mode.

CM = monomer masses are based on calculations of the dimer mass (= 1/2 the dimer), aa = amino acid, lack of the aa = deviation of TM or TD with the MS measured monomer and dimer.

B1d relation	TD (kDa)	TM (kDa)	B1d(a) (kDa)	CM (a) (kDa)	B1d(b) (kDa)	CM (b) (kDa)
	167.042	83.521	166.792	83.396*	164.997	82.499
(TD:a, TD:b)	–	–	0.250	0.125**	2.045	1.022**
Lack aa (kDa)	–	–	–	0.131	–	(0.131 + 0.855)
C-terminal	–	–	–	L ₇₅₂	–	(L ₇₅₂)
N-terminal	–	–	–	–	–	aa ₁₋₇ (MASHH HH)
B1d relation	B1d(c) (kDa)	CM (c) (kDa)	B1d(d) (kDa)	CM (d) (kDa)		
	164.491	82.245	163.795	81.897		
(b:c), (c:d)	0.506	0.254**	0.696	0.348**		
Lack aa (kDa)	–	1.129 (0.855 + 0.274)	–	1.468 (0.855 + 0.274 + 0.339)	(* Variation of masses: because the dimer peaks are broad in shape a mass deviation on monomer level of = 22 Da (for B1(a) to B1d(a) CM) can be calculated.	
C-terminal	–	(L ₇₅₂)	–	(L ₇₅₂) + aa ₇₅₁₋₇₅₀ (AL)		
N-terminal	–	aa ₁₋₉ (MASHH HHHH)	–	aa ₁₋₁₀ (MASHH HHHHH)	** Differences of the masses compared with aa-sequence are based on the width of the protein peaks and their low resolution).	

To determine the ratio between the main monomer B1(a) and the main dimer B1d(a) the evaluation of data is based on the different peak intensities, in comparison to peak areas (S. Tab. 6).

The total distribution of peak intensities or peak areas, including monomer and dimer of the NS3-NS4A protein was set to 100%. This resulted in a calculated ratio between 14% monomer and 86% dimer based on peak intensity data, see S. Fig. 6 (A)/(B). As expected, the results based on the peak area approach reveal similar ratios between the protein monomer (13%) and dimer (87%) (S. Fig. 6 (C)/(D)).



S. Fig. 6 Huge extent of dimerization of batch 1 monomers.

(A)/(C) NS3-NS4A monomer and dimer at 10 eV collision energy, including masses and charge state arrangement, intensities (A) and areas (C) are normalized to the dimer peak 25+. (B)/(D) Proportion of monomer and dimer, comparison is based on averaged intensity (B) of all charge states and averaged peak areas (D), see S. Tab. 6. Measurements were performed in MS1 mode at 10 eV in 200 mM ammonium acetate, pH 7.2 + DTT, protein concentration was 7 μ M (monomer).

6.4.3 Batch 2 NS3-NS4A WT protein

S. Tab. 8 Batch 2 monomers main peak (17+) and their associated sub-peak distribution.

The table is a supplement to chapter 3.4.1. Data evaluation is based on averaged peak intensities (Int.) of charge states at 10 eV. Triplicated measurements were performed. Intensities are given in percentage with associated standard deviation of the values (SD). Ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 8 μ M. Averaged peak intensities (Int.) of associated charge states are shown in: *Fig. 72 (A)/(B) B2(a) and B2(b) main peak (17+) shows similar distribution of their associated sub-peaks at 10 eV while different activation energies revealed strong changes.*

B2(a) + sub-species	Int. (%)	SD	B2(b) + sub-species	Int. (%)	SD
B2(f)	3	0.763	B2(j)	5	0.897
B2(a)	48	11.771	B2(g)	6	0.801
B2(d)	26	0.182	B2(b)	54	14.232
B2(e)	21	1.649	B2(h)	24	2.549
B2(c)	3	0.215	B2(i)	10	0.988

S. Tab. 9 Changes of the peak intensity of batch 2 main monomers and sub-monomers at different acceleration energies in M1 mode (10 eV - 170 eV).

The table is a supplement to chapter 3.4.1. Data evaluation is based on averaged peak intensities (Int.) of charge states of triplicated measurements. Intensities are given in percentage with associated standard deviation of the values (SD). Ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 8 μ M. Averaged peak intensities (Int.) of associated charge states are shown in: *Fig. 73 (C)/(D) B2(a) and B2(b) main peak (17+) shows similar distribution of their associated sub-peaks at 10 eV while different activation energies revealed strong changes.*

Monomer species	Int. (%) 10 eV	Int. (%) 50 eV	Int. (%) 110 eV	Int. (%) 150 eV	Int. (%) 170 eV
B2(a)	2.9	12.2	10.5	6.6	2.6
SD	0.71	3.61	4.29	2.00	0.92
B2(b)	1.6	7.5	6.6	4.0	1.4
SD	0.43	2.52	2.91	1.27	0.46
B2(c)	0.2	0.7	0.2	0.3	0.3
SD	0.01	0.10	0.08	0.08	0.10
B2(d)	1.6	4.2	2.9	2.0	1.1
SD	0.01	0.77	1.01	0.58	0.35
B2(e)	1.2	3.4	2.4	1.6	0.9
SD	0.10	0.69	0.95	0.48	0.29
B2(f)	0.2	0.8	1.0	1.3	2.6
SD	0.05	0.25	0.38	0.43	0.85
B2(g)	0.2	0.7	1.0	1.9	2.4
SD	0.02	0.22	0.34	0.72	0.64
B2(h)	0.7	2.1	1.5	1.1	0.2
SD	0.08	0.49	0.67	0.39	0.07
B2(i)	0.3	0.9	0.6	0.4	0.3
SD	0.03	0.17	0.24	0.13	0.09
B2(j)	0.2	0.5	0.6	0.8	1.4
SD	0.03	0.25	0.35	0.28	0.44

S. Tab. 10 Associated charge states to the peak areas of batch 2 main monomers B2(a) and B2(b) at different acceleration energies in M1 mode (10 eV - 195 eV).

The table is a supplement to chapter 3.4.1. Ratio between B2(a) and B2(b) at different MS1 energies (10 to 195 eV) based on peak areas. $\sum PA$ = averaged peak areas of three independent measurements (M1 to M3), $\sum PA_{(z)}$ = Averaged PA of charge state (z), SD = standard deviation, ratios are given as (%). Ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 8 μ M. Averaged peak areas (PA) of associated charge states are shown in: Fig. 74 Entire peak width of the B2(a) and B2(b) including sub-monomers showed a contrary distribution at different energies (10 eV to 195 eV).

10 eV						50 eV					
NS3-4A kDa	z	M1-M3		$\sum PA_{(z)}$	SD	z	M1-M3		$\sum PA_{(z)}$	SD	
		$\sum PA$	SD				$\sum PA$	SD			
B2(a)	18+	3.218	0.553	6.259	1.167	18+	9.215	1.833	10.506	1.304	
82.683	17+	11.253	2.155			17+	19.019	2.242			
	16+	7.848	1.651			16+	11.146	0.650			
	15+	2.716	0.310			15+	2.644	0.491			
B2(b)	18+	0.694	0.110	2.414	0.386	18+	3.239	0.718	5.060	1.232	
80.779	17+	4.060	1.194			17+	8.230	2.245			
	16+	3.764	0.108			16+	6.833	1.798			
	15+	1.137	0.133			15+	1.940	0.168			
Ratio B2(a)				72	13.5	Ratio B2(b)				67	8.4
Ratio B2(b)				28	4.5	Difference (a)-(b)				33	7.9
Difference (a)-(b)				44		Difference (a)-(b)				35	
150 eV						195 eV					
NS3-4A kDa	z	M1-M3		$\sum PA_{(z)}$	SD	z	M1-M3		$\sum PA_{(z)}$	SD	
		$\sum PA$	SD				$\sum PA$	SD			
B2(a)	18+	8.447	0.191	10.427	0.417	18+	–	–	5.313	0.226	
82.683	17+	17.978	0.423			17+	3.600	0.233			
	16+	12.945	0.794			16+	8.473	0.365			
	15+	2.337	0.260			15+	3.866	0.082			
B2(b)	18+	2.436	0.246	5.416	0.364	18+	–	–	3.802	0.180	
80.779	17+	8.214	0.461			17+	1.737	0.142			
	16+	8.610	0.622			16+	4.244	0.302			
	15+	2.404	0.126			15+	5.425	0.095			
Ratio B2(a)				66	2.6	Ratio B2(b)				58	2.5
Ratio B2(b)				34	2.3	Difference (a)-(b)				42	2.0
Difference (a)-(b)				32		Difference (a)-(b)				17	

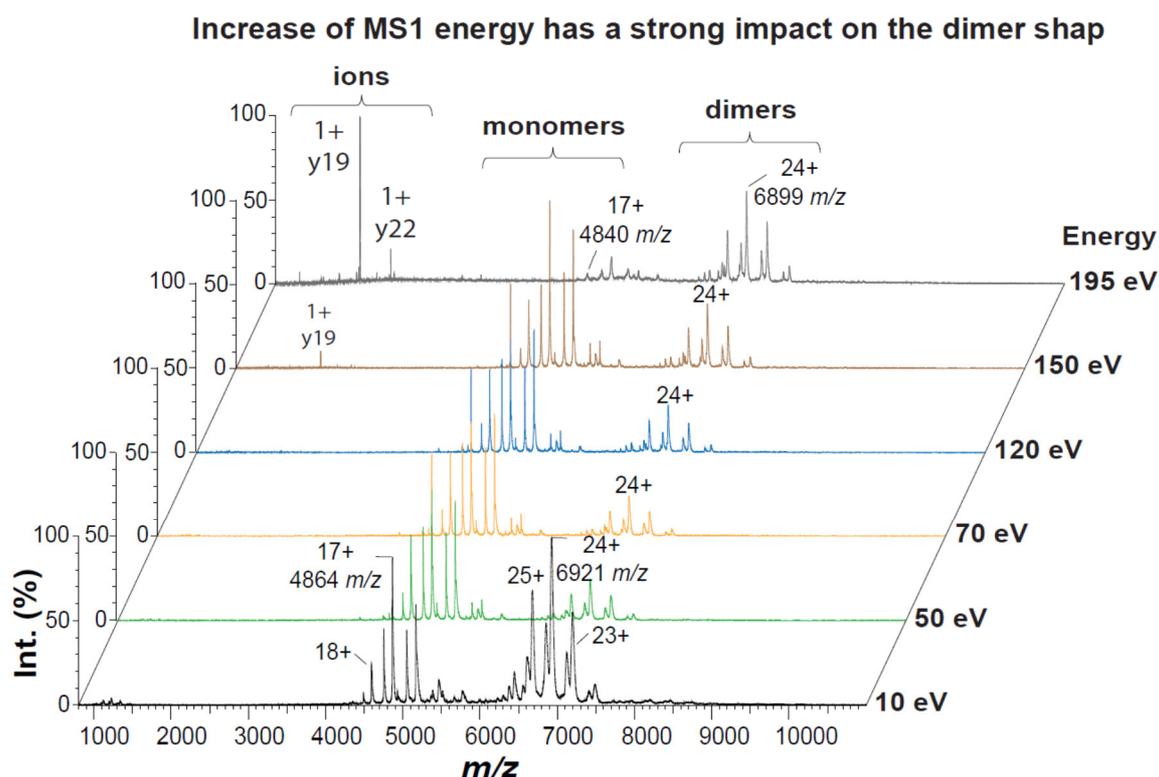
*(Ratio = calculation with respect to the total amount of monomer in batch 2. PA = peak area, $\sum PA_{(z)}$ = averaged PA of charge state (z) 18+ to 15+).

S. Tab. 11 Changes of the peak areas of batch 2 main monomers B2(a) and B2(b) at different acceleration energies in M1 mode (10 eV - 195 eV).

The table is a supplement to chapter 3.4.1. Ratio between B2(a) and B2(b) monomer (82.683 kDa, 80.779 kDa) at different MS1 energies (10 to 195 eV). PA = peak areas of three independent measurements (M1 to M3), $\sum PA_{(z)}$ = averaged PA of charge state (z), SD = standard deviation. Ammonium acetate 200 mM at pH7.2 + 2 mM DTT, protein concentration was 8 μ M. Averaged peak areas (PA) of associated charge states are shown in: Fig. 75 Entire peak width of the B2(a) and B2(b) including sub-monomers showed a contrary distribution at different energies (10 eV to 195 eV).

(%)	$\sum PA_{(z)}$	SD	$\sum PA_{(z)}$	SD
	10 eV		50 eV	
Ratio B2(a)	72.166	13.461	67.493	8.376
Ratio B2(b)	27.834	4.454	32.507	7.917
(a)-(b)	44.333		34.985	
	150 eV		195 eV	
Ratio B2(a)	65.815	2.633	58.289	2.484
Ratio B2(b)	34.185	2.295	41.711	1.974
(a)-(b)	31.630		16.578	

*(Ratio = calculation with respect to the total amount of monomer in batch 2. charge state 18+ to 15+).



S. Fig. 7 Native spectra of the batch 2 NS3-NS4A monomers and dimers reveal strong impact of MS1 energies on the dimer shape.

The figure is a supplement to chapter 3.4.1. Analyses were performed at energies between 10 eV to 195 eV. Protein concentration was 8 μ M (monomer). Sample buffer was 200 mM ammonium acetate pH 7.2 and 2 mM DTT.

S. Tab. 12 (a) Ratio between main monomer/main dimer of B2(a)/B2d(a) and B2(b)/B2d(c): Applied MS1 energies show significant changes of the distribution of peak areas.

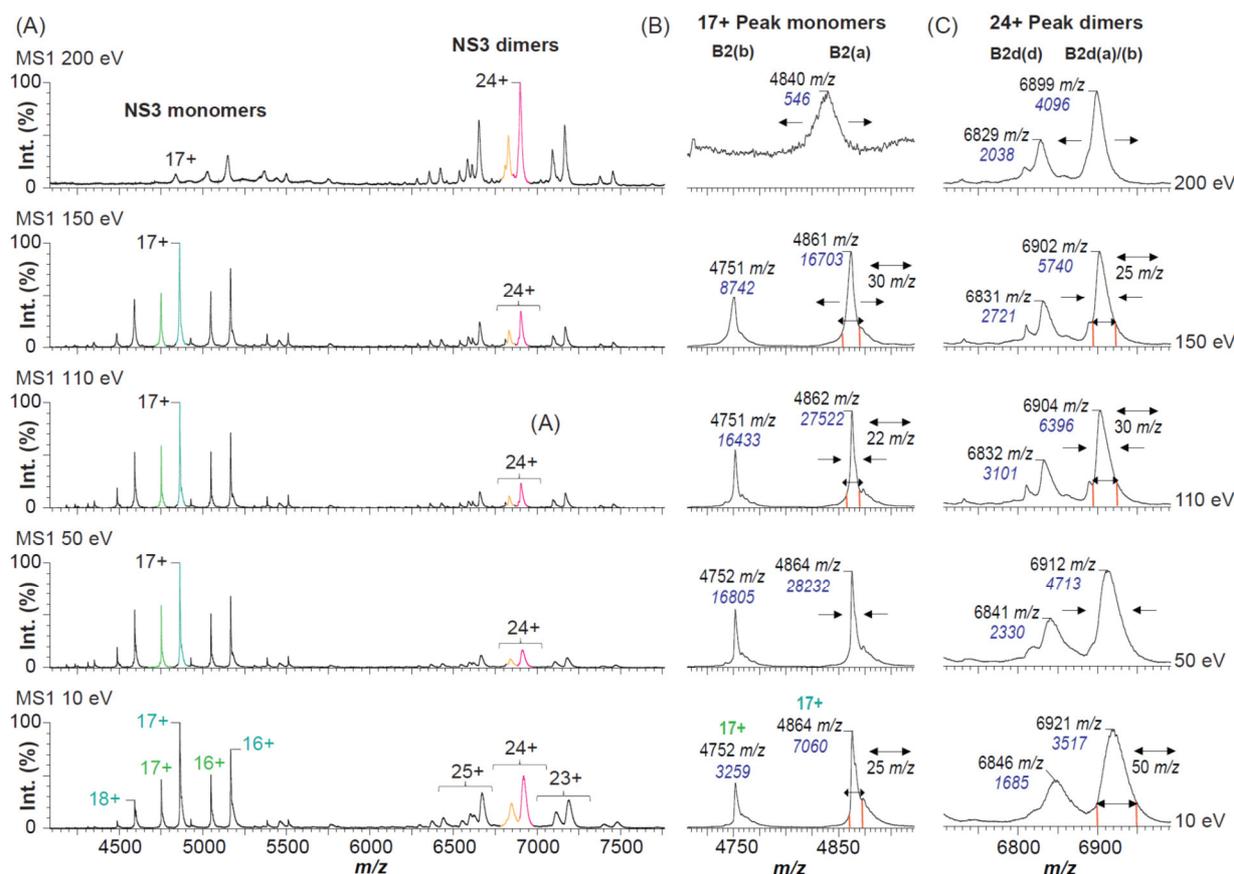
Normalized peak areas (to the highest peak 17+ monomer or 24+ protein dimer), raw and Averaged peak areas of associated charge states are shown in: Fig. 76 The ratio between the main monomer/main dimer of B2(a)/B2d(a) and B2(b)/B2d(c) show significant changes depending on MS1 energies.

CSFV _{WT} NS3-NS4A	Charge	10 eV		Ratio		50 eV		Ratio	
		\sum_{M1-M3}	<i>SD</i>	(%)	<i>SD</i>	\sum_{M1-M3}	<i>SD</i>	(%)	<i>SD</i>
B2(a) +sub-species 82.7 kDa	18+	0.203	0.052	41.772	10.710	0.480	0.050	65.721	4.846
	17+	0.716	0.215			1.000	0.000		
	16+	0.494	0.132			0.598	0.101		
	15+	0.169	0.029			0.144	0.043		
B2d(a) +sub-species 165.6 kDa	26+	0.223	0.053	56.837	3.597	0.100	0.034	30.720	5.675
	25+	0.705	0.040			0.308	0.067		
	24+	1.000	0.000			0.521	0.110		
	23+	0.622	0.073			0.292	0.058		
	22+	0.142	0.014			0.078	0.015		
NS3-heli-case 55.1 kDa	14+	–	–						
	13+	0.013	0.003	1.391	0.343	0.030	0.013	3.558	1.344
CSFV _{WT} NS3-NS4A	Charge	70 eV		Ratio		120 eV		Ratio	
		\sum_{M1-M3}	<i>SD</i>	(%)	<i>SD</i>	\sum_{M1-M3}	<i>SD</i>	(%)	<i>SD</i>
B2(a) +sub-species	18+	0.466	0.089	67.152	3.338	0.588	0.072	66.349	3.231
	17+	1.000	0.000			1.000	0.000		
	16+	0.689	0.035			0.631	0.036		
	15+	0.136	0.009			0.142	0.021		
B2d(a) +sub-species	26+	0.091	0.016	29.653	1.568	0.104	0.009	30.482	1.795
	25+	0.327	0.018			0.319	0.027		
	24+	0.488	0.007			0.553	0.022		
	23+	0.292	0.023			0.293	0.030		
	22+	0.067	0.015			0.086	0.002		
NS3-heli-case	14+	0.021	0.002	3.195	0.251	0.019	0.004	3.169	0.357
	13+	0.034	0.003			0.037	0.003		
CSFV _{WT} NS3-NS4A	Charge	150 eV		Ratio		195 eV		Ratio	
		\sum_{M1-M3}	<i>SD</i>	(%)	<i>SD</i>	\sum_{M1-M3}	<i>SD</i>	(%)	<i>SD</i>
B2(a) +sub-species	18+	0.470	0.012	66.310	2.071	0.000	0.000	28.192	1.372
	17+	1.000	0.000			0.175	0.016		
	16+	0.721	0.059			0.412	0.029		
	15+	0.130	0.012			0.188	0.009		
B2d(a) +sub-species	26+	0.090	0.012	31.391	2.686	0.124	0.019	71.808	2.133
	25+	0.364	0.046			0.558	0.009		
	24+	0.494	0.050			1.000	0.000		
	23+	0.345	0.024			0.618	0.062		
	22+	0.081	0.002			0.167	0.017		
NS3-heli-case	14+	0.011	0.001	2.300	0.109	0.000	0.000	0.000	0.000
	13+	0.029	0.001			0.000	0.000		

S. Tab. 13 (b) Ratio between main monomer/main dimer of B2(a)/B2d(a) and B2(b)/B2d(c): Applied MS1 energies show significant changes of the distribution of peak areas.

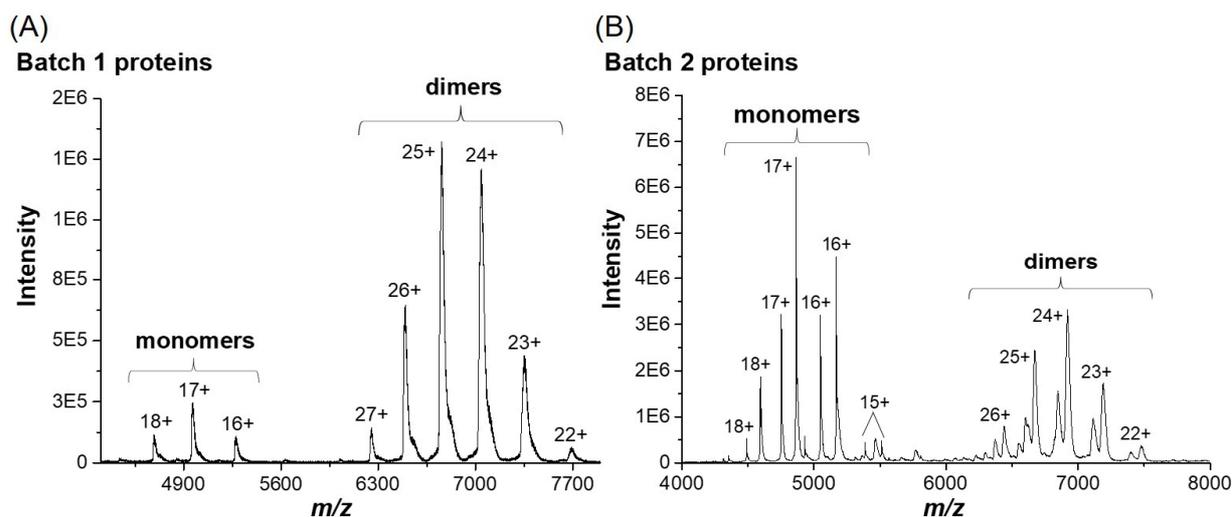
Shown are normalized peak areas (to the highest peak 17+ monomer or 24+ protein dimer) of NS3-NS4A protein and the ratios (monomer, dimer and helicase) based on triplicate measurements (\sum_{M1-M3}). Related standard deviation (*SD*) is shown. Associated charge states: monomer (18+ to 15+), dimer (26+ to 22+), helicase (14+ to 13+). Protein concentration = 8 μ M (monomer). Sample buffer with 200 mM ammonium acetate pH 7.2 and 2 mM DTT, energy: 10 eV to 195 eV. Averaged peak areas of associated charge states are shown in: Fig. 77 *The ratio between the main monomer/main dimer of B2(a)/B2d(a) and B2(b)/B2d(c) show significant changes depending on MS1 energies.*

B2	10 eV		50 eV		70 eV	
	Ratio (%)	SD	Ratio (%)	SD	Ratio (%)	SD
B2(a)	41.772	10.710	65.721	4.846	67.152	3.338
B2d(a)	56.837	3.597	30.720	5.675	29.653	1.568
helicase	1.391	0.343	3.558	1.344	3.195	0.251
B2	120 eV		150 eV		195 eV	
	Ratio (%)	SD	Ratio (%)	SD	Ratio (%)	SD
B2(a)	66.349	3.231	66.310	2.071	28.192	1.372
B2d(a)	30.482	1.795	31.391	2.686	71.808	2.133
helicase	3.169	0.357	2.300	0.109	–	–
B2	10 eV		50 eV		70 eV	
	Ratio (%)	SD	Ratio (%)	SD	Ratio (%)	SD
B2(b)	35.5	6.259	61.569	15.741	68.30	5.481
B2d(c)	61.7	5.442	31.667	11.727	25.97	3.989
helicase	3.0	0.527	6.765	3.612	5.73	0.437
B2	120 eV		150 eV		195 eV	
	Ratio (%)	SD	Ratio (%)	SD	Ratio (%)	SD
B2(b)	63.79	3.800	67.43	3.122	40.03	1.545
B2d(c)	29.99	2.950	28,05	3.355	59.97	3.774
helicase	6.22	0.629	4.52	0.494	–	–



S. Fig. 8 Changes in the peak intensity and peak width of B2 monomers and dimers based on different ion activation energies 10 to 200 eV.

The figure is a supplement to chapter 3.4.1. Analyses were performed at energies between 10 eV to 200 eV. Protein concentration was 8 μ M (monomer). Sample buffer was 200 mM ammonium acetate pH 7.2 and 2 mM DTT.



S. Fig. 9 Dimerization of batch 1 and 2 N3-NS4A monomers.

(A) Batch 1: B1(a) monomer show a charge state distribution from 18+ to 16+, B1d(a) dimer shows a distribution of 27+ to 22+, protein concentration was 7 μ M (monomer). (B) Batch 2: B2(a) monomer show a charge state distribution from 18+ to 15+, B2d(a) dimer of 26+ to 22+. Protein concentration was 8 μ M (monomer). Conditions: MS1 mode at 10 eV, 200 mM ammonium acetate, pH 7.2 + DTT.

S. Tab. 14 Identification of four different dimers in B2 by increasing energy in MS1 and MS2 studies.

Main peak dimer B2d(a) can be assigned to the B2 monomer B2(d) (82.781 kDa). Low intense dimers B2d(b) and B2d(d) belong to the monomer B2(a) (82.683 kDa) and B2(b) (80.779 kDa). The corresponding monomer to the second dominant dimer B2d(c) 163.881 kDa could not be identified. The calculated monomer mass CM = 81.941 kDa.

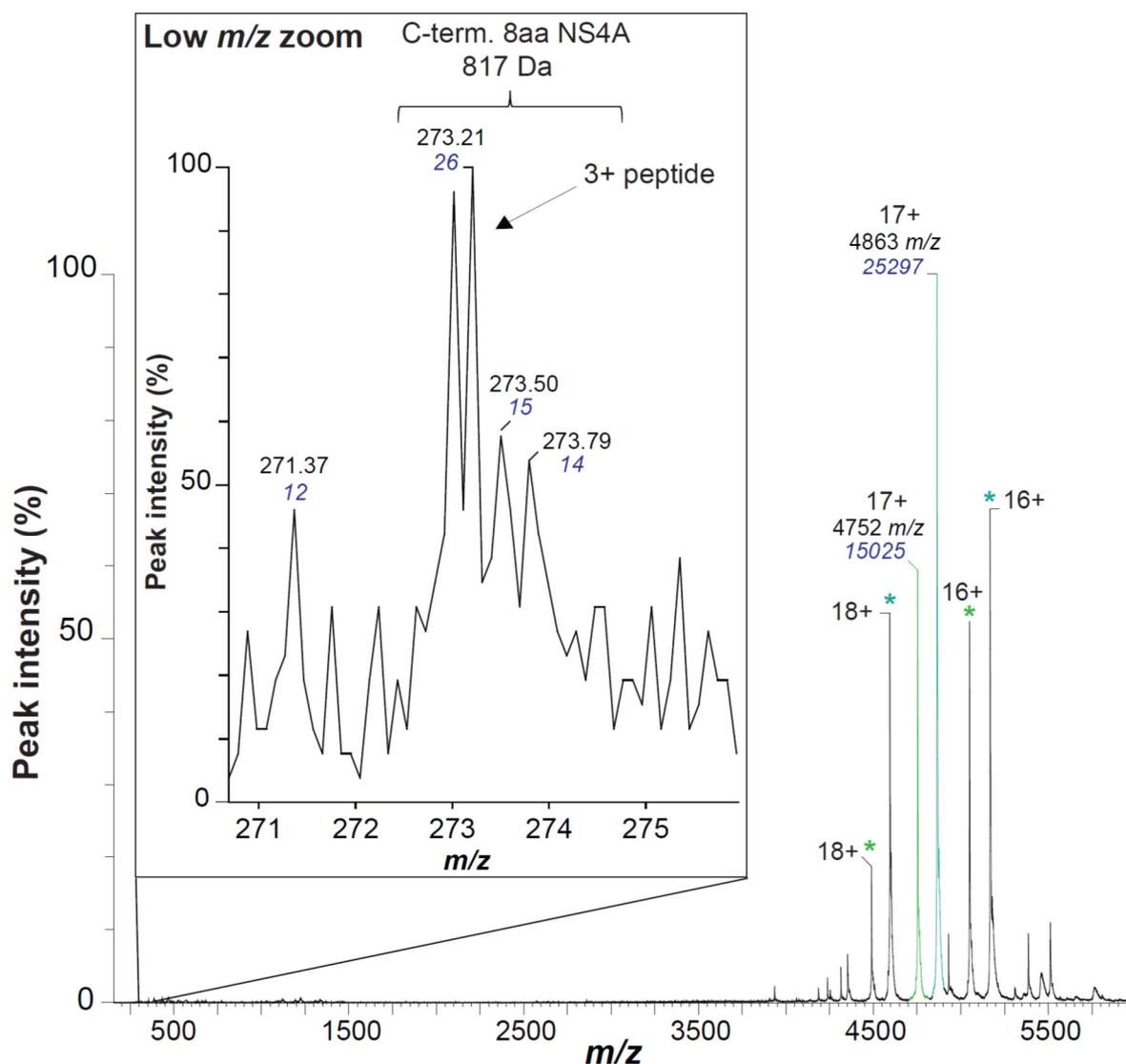
B2 relation	TD (kDa)	TM (kDa)	B2d(a) (kDa)	B2d(a) CM (kDa)	B2d(b) (kDa)	B2d(b) CM (kDa)
	167.042	83.521	165.562	82.781	165.350	82.675
(TD/TM:a), (TD/TM:b)	–	–	1.480	0.740	1.692	0.846
Lack aa (kDa)	–	–	–	0.718	–	0.818
C-terminal	–	–	–	–	–	aa ⁷⁵²⁻⁷⁴⁵ (STANALL)
N-terminal	–	–	–	aa ¹⁻⁶ (MASHHH)	–	–
B2 relation	B2d(c) (kDa)	B2d(c) CM (kDa)	B2d(c) (kDa)	B2d(c) CM (kDa)	B2d(d) (kDa)	B2d(d) CM (kDa)
	163.881	81.941 kDa	163.881	81.941	161.558	80.779
(TD/TM:c), (a:c)*, (TD/TM:d)	3.161	1.580	1.681	0.840	5.452	2.726
Lack aa (kDa)	–	1.541	–	0.840	–	(1.921 + 0.818)
C-terminal	–	–	–	–	–	aa ⁷⁵²⁻⁷⁴⁵ (STANALL)
N-terminal	–	aa ¹⁻¹² (MASHHH HHHHHH)	–	aa ⁷⁻¹² (HHHHHH)	–	aa ¹⁻¹⁵ (MASHHHHHH HHHHEN)

(TM/TD = theoretical monomer/dimer, CM = calculated monomer masses are based on calculations of the dimer mass (= 1/2 the dimer), aa = amino acid, (a:c)* = control of calculation), N-term. 6aa are marked blue.

S. Tab. 15 Course of the auto-proteolytical process of the NS3-protease.

Data are based on intensities of the proteins (pro = protease, hel = helicase, main monomer and main dimer species (a) and (b) of every batch (B1-B4). (a) The table is a supplement to Fig. 53 The auto-proteolytical activity of the NS3-protease including protein monomers and dimers leads to the release of the NS3-helicase subunit.

Week 0-3	Protease	Helicase	Monomer (17+)		Dimer (24+)
	Int.	Int.	Int. (a)	Int. (b)	Int. (a)
B1	0	0	1304	0	14900
B2	0	45	7060	3259	1080
B3	920	3934	850	4700	430
B4	0	1180	0	0	0
Av. Int.	Ratio pro	Ratio hel	Ratio (a)	Ratio (b)	Ratio dimer (a)
16204	0	0	8.047	0	91.95
11444	0	0.39	61.69	28.48	9.44
10834	8.49	36.31	7.85	43.38	3.97
1180	0	100	0	0	0



S. Fig. 10 MS1 data show the unbound C-terminal NS4A cofactor segment STAENALL (amino acids aa752-745, 817 Da) within batch 2.

Measurement was performed at 50 eV. The identified peptide carries 3 positive charges (3+). Blue italic numbers show the peak intensity. The two main monomers B2(a) (missing aa752-745, labelled in dark-green) and B2(b) (missing aa752-745 and aa1-15, labelled in light-green) Conditions: 1 week at + 4 °C, 200 mM ammonium acetate + 2 mM DTT pH 7.2, protein concentration 8 μ M.

6.5 L45A mutations of the CSFV NS3-NS4A protein

S. Tab. 16 (a) CSFV_{L45A} mutant: Absolute (Abs.) peak areas (PA) of the monomer, dimer and the free NS3-helicase of the NS3-NS4A fusion protein.

The table is a supplement to chapter 3.5.1. Calculations are based on triplicated measurements and on averaged peak areas, raw data are shown. Protein concentration = 7 μ M (monomer) in 200 mM ammonium acetate, pH 7.4 + 0.5 mM DTT, measurements: positive ion mode at 10 eV energy. Averaged peak areas of associated charge states are shown in: *Fig. 78 L45A exhibits a lower degree of dimerization.* & *Fig. 79 L45A reveals a huge amount of unbound NS3-helicase subunit.*

NS3- helicase	Charge	Abs. PA M1	Abs. PA M2	Abs. PA M3
55.8 kDa	14+	14.695	20.667	25.383
	13+	46.956	31.655	24.102
NS3-4A monomer	Charge			
83.3 kDa	19+	5.123	4.693	5.322
	18+	16.485	14.828	17.617
	17+	7.298	15.685	17.470
	16+	3.081	9.255	5.725
NS3-4A dimer	Charge			
166.8 kDa	26+	12.745	9.342	16.127
	25+	26.102	27.263	26.708
	24+	31.055	36.081	23.243
	23+	14.318	18.380	14.732
	22+	8.915	5.630	7.528

S. Tab. 17 (b) CSFV_{L45A} mutant: Relative (Rel.) peak areas (PA) of the monomer and dimer of the fusion protein.

Averaged and normalized peak areas of CSFV_{L45A} mutant. Data normalization was performed based to the most intense peak (24+ protein dimer). Averaged values (Av.) and related standard deviation (SD) are shown. Averaged peak areas of associated charge states are shown in: *Fig. 80 L45A exhibits a lower degree of dimerization.*

NS3-4A monomer	Charge	Rel. PA M1	Rel. PA M2	Rel. PA M3	Av. ^{M1-M3} (%)	SD
83.3 kDa	19+	0.165	0.130	0.229	17	4.1
	18+	0.531	0.411	0.758	57	14.4
	17+	0.235	0.435	0.752	47	21.3
	16+	0.099	0.256	0.246	20	7.2
	15+	0.165	0.130	0.229	17	4.1
NS3-4A dimer	Charge					
166.8 kDa	26+	0.410	0.259	0.694	45	18.0
	25+	0.841	0.756	1.149	92	16.9
	24+	1.000	1.000	1.000	100	0.0
	23+	0.461	0.509	0.634	53	7.3
	22+	0.287	0.156	0.324	26	7.2

S. Tab. 18 (c) CSFV_{L45A} mutant: Averaged peak areas of charge states of the monomer and dimer.

Native MS data were produced at 10 eV collision energy. Ratio of the monomer and dimer is shown. Averaged peak areas of associated charge states are shown in: *Fig. 81 L45A exhibits a lower degree of dimerization.*

NS3-4A	Ratio (%)	<i>SD</i>
monomer 83.3 kDa	36	11.7
dimer 166.8 kDa	64	9.9

S. Tab. 19 (d) CSFVL45A mutant: Relative (Rel.) peak areas (PA) of the monomer, dimer and free NS3-helicase.

Data normalization was performed based to the most intense peak (24+ protein dimer). Averaged values (*Av.*) and related standard deviation (*SD*) are shown. Averaged peak areas of associated charge states are shown in: *Fig. 82 L45A reveals a huge amount of unbound NS3-helicase subunit.*

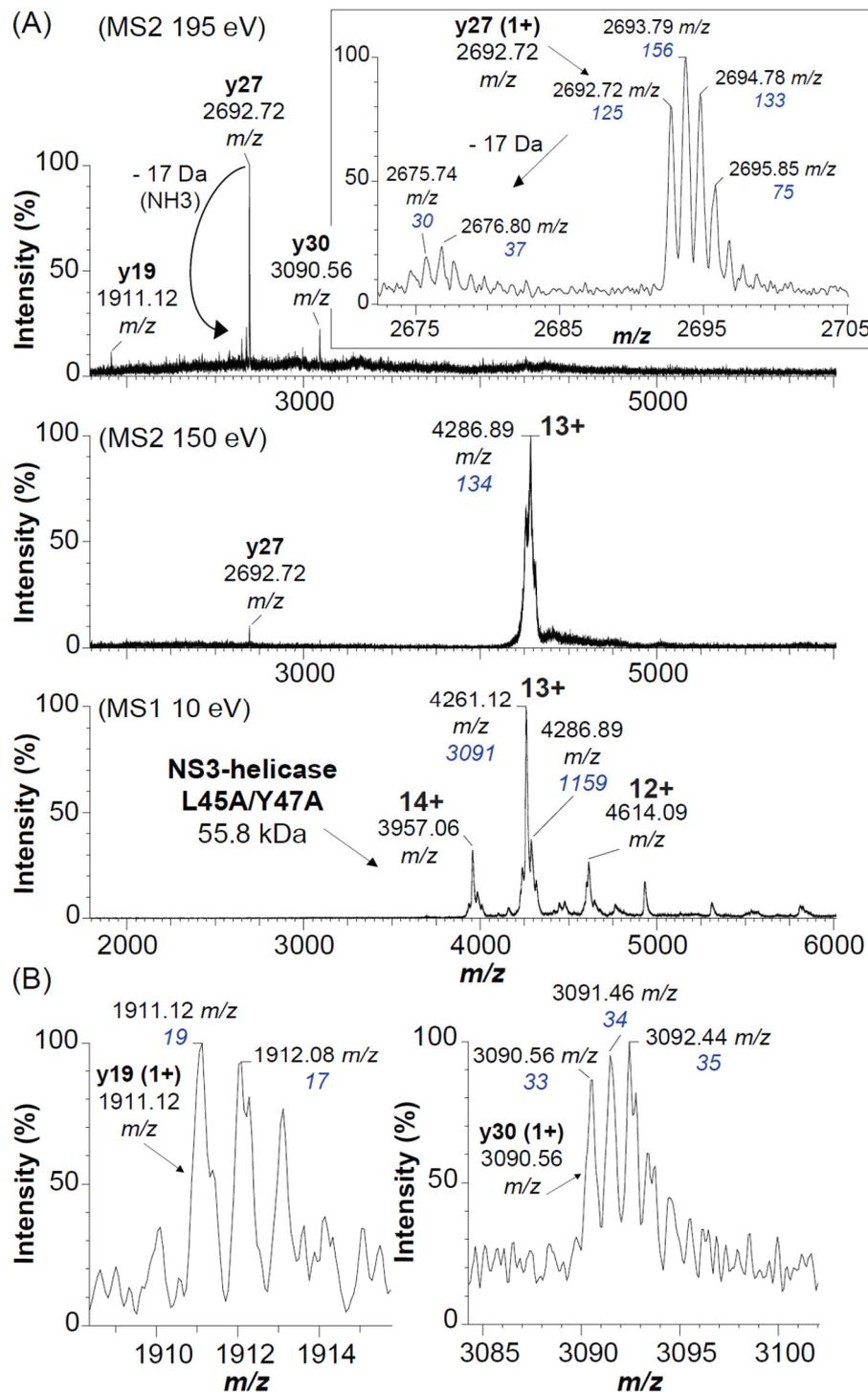
NS3- helicase	Charge	Rel. Peak Area M1	Rel. Peak Area M1	Rel. Peak Area M1
55.8 kDa	14+	0.473	0.573	1.092
	13+	1.512	0.877	1.037
NS3-4A monomer	Charge			
83.3 kDa	19+	0.165	0.130	0.229
	18+	0.531	0.411	0.758
	17+	0.235	0.435	0.752
	16+	0.099	0.256	0.246
NS3-4A dimer	Charge			
166.8 kDa	26+	0.410	0.259	0.694
	25+	0.841	0.756	1.149
	24+	1.000	1.000	1.000
	23+	0.461	0.509	0.634
	22+	0.287	0.156	0.324

S. Tab. 20 (e) CSFVL45A mutant: Average of the relative peak areas (PA) of the monomer, dimer and free NS3-helicase.

Native MS data were produced at 10 eV collision energy. Ratio of the NS3-helicase, monomer and dimer is shown. Averaged peak areas of associated charge states are shown in: *Fig. 83 L45A reveals a huge amount of unbound NS3-helicase subunit.*

NS3-4A	Ratio (%)	<i>SD</i>
NS3-helicase		
55.8 kDa	48	14.1
monomer		
83.3 kDa	19	6.1
dimer		
166.8 kDa	33	5.2

6.6 L45A/Y47A mutation of the CSFV NS3-NS4A protein



S. Fig. 11 Dissociation of the 55.8 kDa protein of the DM L45A/Y47A revealed the NS3-helicase with complete C-terminal NS4A-cofactor.

The figure is a supplement to chapter 3.5.2. (A) MS2 spectra of the 13+ precursor (4287 m/z) at 150 and 195 eV. MS1 spectrum of the identified NS3-helicase at 10 eV. (B) MS2 spectra of the singly charged y -ions of the NS3-helicase (195 eV). Blue italic numbers show the peak intensity. Conditions: 3 weeks at + 4 °C, 200 mM ammonium acetate + 0.5 mM DTT pH 7.2, protein concentration 8 μ M. Data shown in: *Fig. 84 Crystal structure of the DM L45A/Y47A NS3-helicase (55.8 kDa) revealed the interaction with the complete C-terminal NS4A-cofactor.*

S. Tab. 21 (a) CSFVL45A/Y47A sample storage test over a four weeks period (1d, 2w and 4w) at + 4 °C.

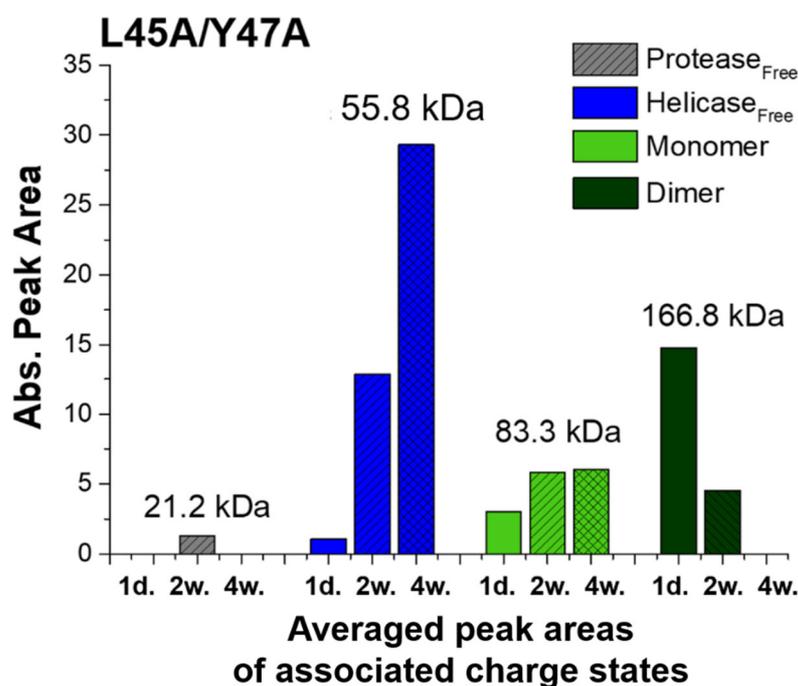
The table is a supplement to chapter 3.5.2. Shown are absolute and averaged peak areas of certain charge states of the different protein structures (protease, helicase, monomer, dimer). Calculations are based on averaged (Av.) data. Averaged peak areas of associated charge states are shown in: *Fig. 85 The double mutation (DM) LA5A/Y47A also demonstrates a high level of auto-proteolytic activity at + 4 °C.*

NS3- helicase	Av.PA 1 d.	Ratio (%)	Av.PA 2 w.	Ratio (%)	Av.PA 4 w.	Ratio (%)
55.8 kDa	1.070	5.663	12.832	52.341	29.304	82.901
DM(a)						
83.3 kDa	3.028	16.026	5.826	23.764	6.044	17.099
DMd(a)						
166.8 kDa	14.796	78.311	4.552	18.567	-	-
NS3-protease						
21.2 kDa	-	-	1.306	5.327	-	-

S. Tab. 22 (b) CSFVL45A/Y47A sample storage test over a four weeks period (1d, 2w and 4w) at + 4 °C.

The table is a supplement to chapter 3.5.2. Shown are absolute (Abs.) peak areas (PA) of all summed charge states of the different protein structures (protease, helicase, monomer, dimer). Calculations are based on averaged (Av.) data. Averaged peak areas of associated charge states are shown in: *Fig. 86 The double mutation (DM) LA5A/Y47A also demonstrates a high level of auto-proteolytic activity at + 4 °C.*

NS3- helicase	Charge	Abs. PA 1d	Abs. PA 2w	Abs. PA 4w	
55.8 kDa	15+	-	2.694	-	
	14+	-	16.147	7.969	
	13+	0.641	21.186	51.421	
	12+	1.499	11.300	28.521	
DM(a)					
83.3 kDa	Charge				
	18+	-	0.385	2.217	
	17+	2.731	10.706	9.871	
	16+	3.325	7.846	-	
DMd(a)	Charge				
	15+	3.028	4.365	-	
	166.8 kDa	26+	4.345	3.995	-
	25+	9.760	6.002	-	
	24+	19.431	6.597	-	
	23+	25.329	4.104	-	
NS3- protease	Charge				
	22+	17.451	2.061	-	
	21+	12.458	-	-	
	21.2 kDa	9+	-	1.124	-
		8+	-	1.487	-



S. Fig. 12 Auto-proteolytic activity of the double mutant (DM) L45A/Y47A at + 4 °C.

The figure is a supplement to chapter 3.5.2. Data are based on absolute peak areas. Protein concentration = 8 μ M, 200 mM ammonium acetate, pH 7.4 + 0.5 mM DTT, measurements: positive ion mode at 10 eV energy. Data are based on averaged peak areas. Associated charge states: monomer (18+ to 16+), dimer (25+ to 22+), helicase (14+ to 12+), protease (9+, 8+) and are shown in: *Fig. 87 The double mutation (DM) L45A/Y47A also demonstrates a high level of auto-proteolytic activity at + 4 °C.*

S. Tab. 23 (a) CSFVL45A/Y47A double mutant (DM) shows protease activity.

The table is a supplement to chapter 3.5.2. Triplicate measurements (M1-M3) are shown. Native MS data were produced at 10V collision energy and are denoted by absolute (Abs.) peak areas. Shown are peak areas (PA) of associated charge states. Protein concentration = 7 μ M (monomer) in 200 mM ammonium acetate, pH 7.4 + 0.5 mM DTT. Data are shown in: *Fig. 88 L45A/Y47A shows protease activity even when stored three weeks at + 4 °C.*

NS3- helicase	Charge	Abs. PA M1	Abs. PA M2	Abs. PA M3
55.8 kDa	15+	4.106	3.435	1.851
	14+	36.008	36.003	32.777
	13+	36.993	39.581	43.044
	12+	8.850	10.254	11.993
NS3-4A monomer	Charge			
83.3 kDa	18+	2.099	1.134	1.894
	17+	4.855	6.485	6.786
NS3- protease	Charge			
21.2 kDa	9+	2.736	1.024	0.444
	8+	4.353	2.084	1.212

S. Tab. 24 (b) CSFVL45A/Y47A double mutant (DM) shows protease activity.

The table is a supplement to chapter 3.5.2. Triplicate measurements (M1-M3) are shown. Native MS data were produced at 10V collision energy and are denoted by relative (Rel.) peak areas. Shown are peak areas (PA) of associated charge states and the related standard deviation (SD). Data were normalized to the most intense peak (13+) of the NS3-helicase. Protein concentration = 7 μ M (monomer) in 200 mM ammonium acetate, pH 7.4 + 0.5 mM DTT. Data are shown in: *Fig. 89 L45A/Y47A shows protease activity even when stored three weeks at + 4 °C.*

NS3-heli- case	Charge	Rel. PA M1	Rel. PA M2	Rel. PA M3	M1-M3 (%)	SD
55.8 kDa	15+	0.111	0.087	0.043	8	2.8
	14+	0.973	0.910	0.761	88	8.9
	13+	1.000	1.000	1.000	100	0.0
	12+	0.239	0.259	0.279	26	1.6
NS3-4A monomer	Charge					
83.3 kDa	18+	0.057	0.029	0.044	4	1.1
	17+	0.131	0.164	0.158	15	1.4
NS3-pro- tease	Charge					
21.2 kDa	9+	0.074	0.026	0.010	4	2.7
	8+	0.118	0.053	0.028	7	3.8

S. Tab. 25 (c) CSFVL45A/Y47A double mutant (DM) shows protease activity.

Native MS data were produced at 10 eV collision energy. Ratio of helicase, monomer and dimer is shown. Data are shown in: *Fig. 61 L45A/Y47A shows protease activity even when stored three weeks at + 4 °C.*

NS3-4A	Ratio (%)	SD
NS3- heli- case	79	4.7
55.8 kDa monomer	14	1.8
83.3 kDa NS3- pro- tease	7	4.6
21.2 kDa		

6.7 MS based activity assay

S. Tab. 26 ESI-TOF data: Basis for calculations of CSFV NS3-NS4_{WT} protease activity.

Shown are the absolute intensity data of the main substrate peak (1549 m/z), the isotopic peak 1 (1550 m/z) and the background noise signals (1544 m/z) at different time points of reaction. In addition, ratios of background corrected data are shown as signal to noise values for both, the main substrate and the isotopic peak. Standard deviation (SD) of triplicate measurements (Σ_{1-3}) was calculated. The substrate concentrations at the different time points are indicated. Substrate concentration at T0 min = 2.02 nmol, volume = 5 μ L.

Time (min)	Σ_{1-3} Substrate		Σ_{1-3} Isotope		Σ_{1-3} Background	
	1549 m/z	SD	1550 m/z	SD	1544 m/z	SD
0	1.50E+03	215.813	1.27E+03	175.353	1.66E+02	2.825
1	2.07E+03	53.787	1.75E+03	55.473	2.42E+02	32.458
3	1.51E+03	58.846	1.28E+03	75.623	2.49E+02	50.142
5	5.72E+02	15.139	4.76E+02	16.606	9.57E+01	3.247
10	5.29E+02	49.046	4.57E+02	42.697	1.73E+02	19.986
20	4.01E+02	24.595	3.83E+02	15.181	2.41E+02	14.204
60	5.40E+02	356.185	5.37E+02	354.733	5.38E+02	353.374
(min)	S/N _{sub}	SD_{sub}	S/N _{iso}	SD_{iso}	[M] _{sub}	[M] _{iso}
0	9.036	1.447	7.651	1.179	4.04E-04	4.04E-04
1	8.554	1.107	7.231	0.935	3.87E-04	3.88E-04
3	6.064	1.234	5.141	0.981	2.80E-04	2.82E-04
5	5.977	0.346	4.974	0.347	2.67E-04	2.63E-04
10	3.058	0.142	2.642	0.169	1.36E-04	1.40E-05
20	1.664	0.134	1.589	0.084	5.97E-06	8.41E-05
60	1.004	0.011	0.998	0.009	4.90E-07	5.27E-05

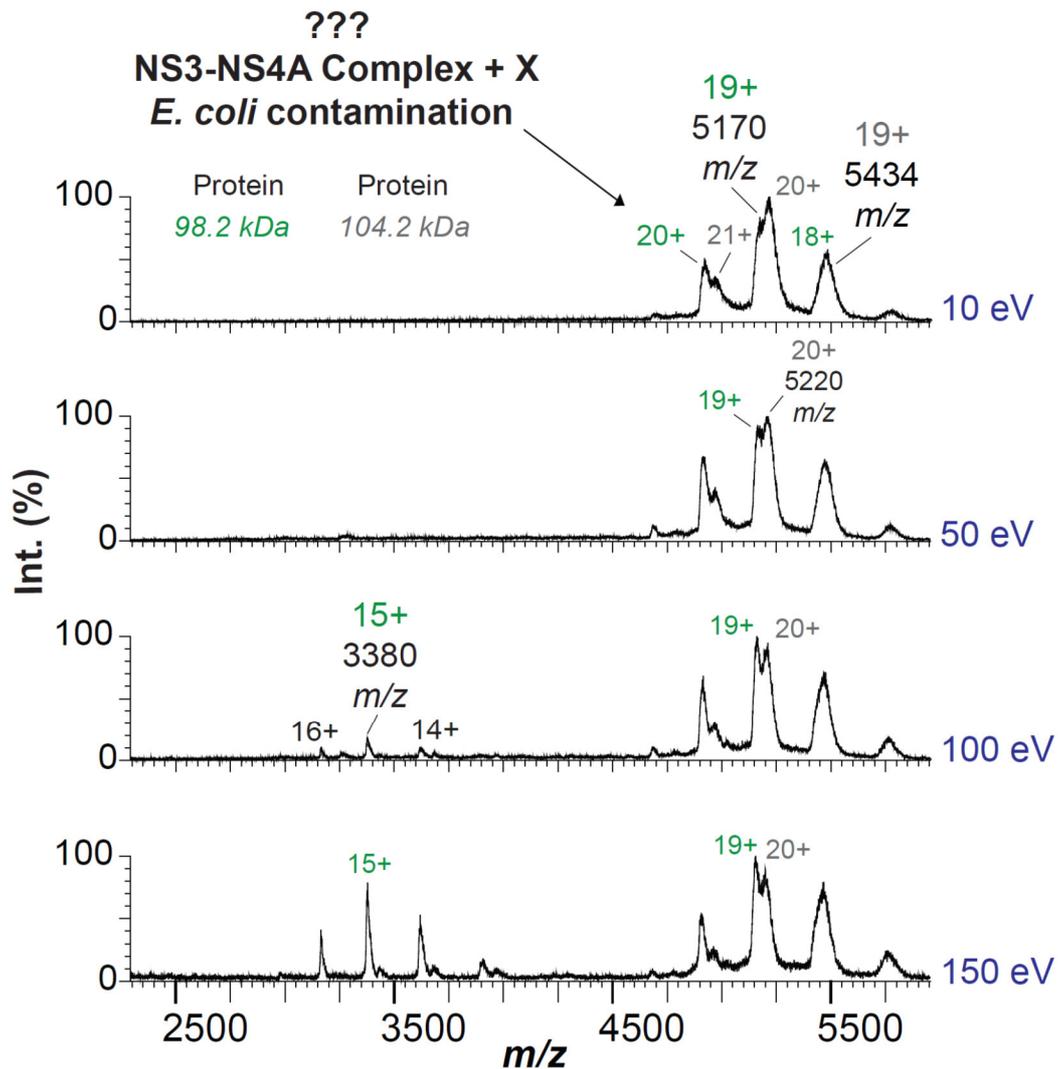
*(Σ_{1-3} = sum of triplicate absolute peak intensities, SD = standard deviation, Ratio S/N = signal/noise, sub = substrate, iso = isotope, [M] = molarity).

S. Tab. 27 ESI-TOF data: Basis for calculations of HCV NS3-NS4_{WT} protease activity.

Shown are absolute intensity data of: main substrate peak (1549 m/z), isotope peak 1 (1550 m/z), background noise signals (1544 m/z) at different time points of reaction. Ratios of background corrected data are shown as signal to noise values for both, the main substrate and the isotopic peak. Standard deviation error (SD) of triplicate measurements (Σ_{1-3}) was calculated. The substrate concentrations at the different time points are indicated. Substrate concentration at T0 min = 2.02 nmol, sample volume = 5 μ L.

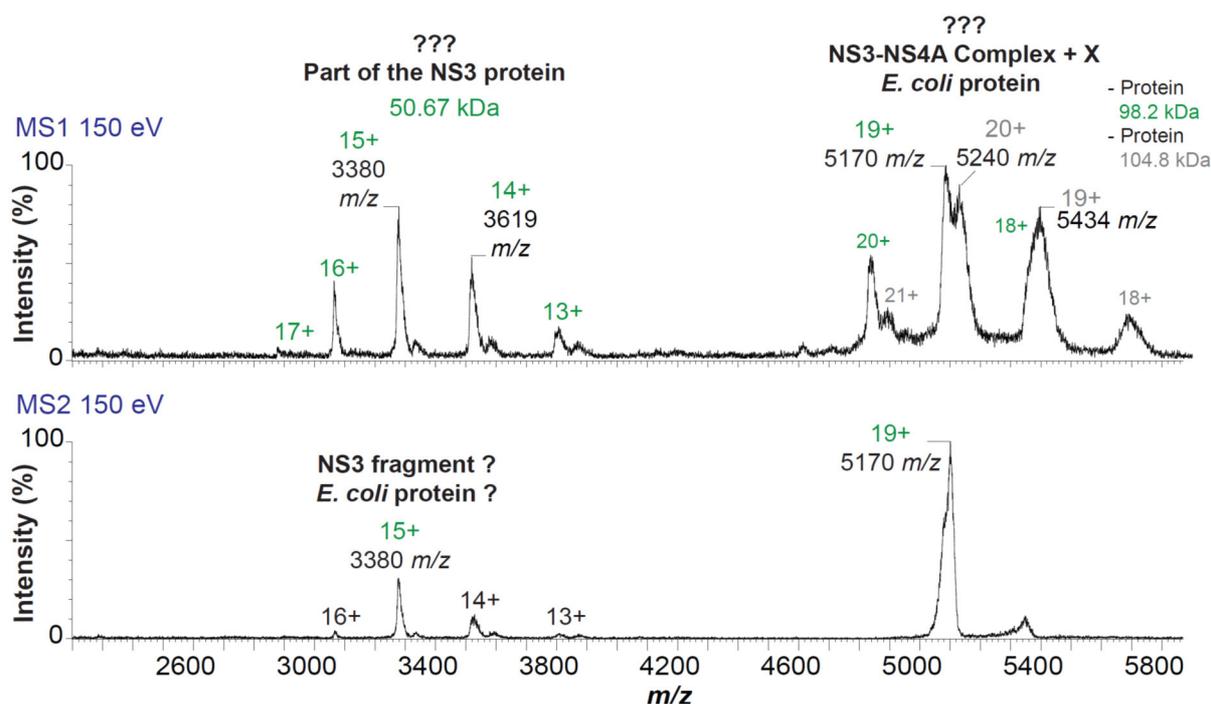
Time (min)	Σ_{1-3} Substrate		Σ_{1-3} Isotope		Σ_{1-3} Background	
	1549 m/z	SD	1550 m/z	SD	1544 m/z	SD
0	4.35E+03	76.778	3.66E+03	61.919	2.89E+02	7.837
1	3.15E+03	209.651	2.67E+03	162.672	3.83E+02	32.150
5	3.03E+03	143.442	2.54E+03	113.162	2.78E+02	27.594
15	2.77E+03	326.944	2.33E+03	270.166	5.20E+02	39.959
20	2.23E+03	390.095	1.91E+03	303.843	6.45E+02	143.634
(min)	S/N _{sub}	SD_{sub}	S/N _{iso}	SD_{iso}	[M] _{sub}	[M] _{iso}
0	1.507E+01	0.564	1.268E+01	0.432	4.04E-04	4.04E-04
1	8.216E+00	1.152	6.976E+00	0.937	2.23E-04	2.25E-04
5	1.090E+01	1.529	9.163E+00	1.273	2.96E-04	2.96E-04
15	5.323E+00	0.992	4.488E+00	0.822	1.45E-04	1.45E-04
20	3.463E+00	1.408	2.966E+00	1.157	1.00E-04	1.02E-04

6.8 HCV native MS1 and MS2 addition



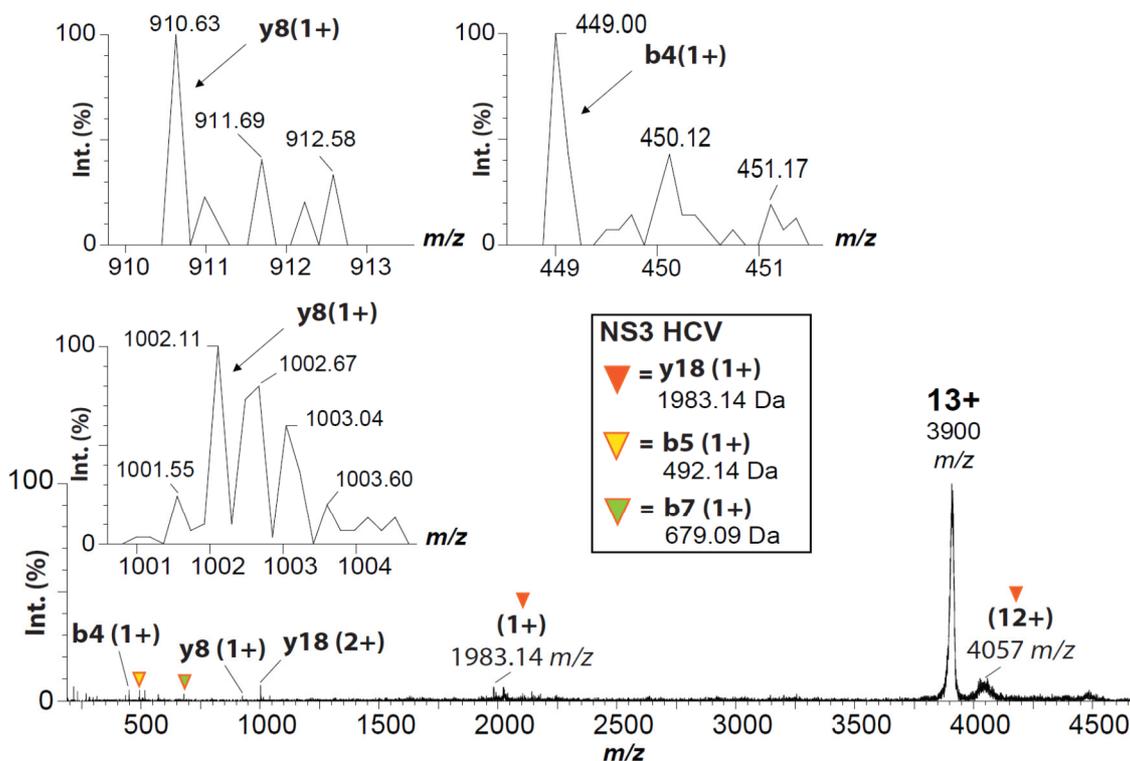
S. Fig. 13 Activation of the protein complex in the HCV sample with different energies in the MS1 mode shows the dissociation of an approx. 51 kDa protein at 100 eV.

The figure is a supplement to sub-section 3.1.1. Raw data are based on a protein concentration of 5 μ M, 200 mM ammonium acetate, pH 7.2 + 2 mM DTT. Measurements were performed in positive ion mode at 10 to 150 eV energy.



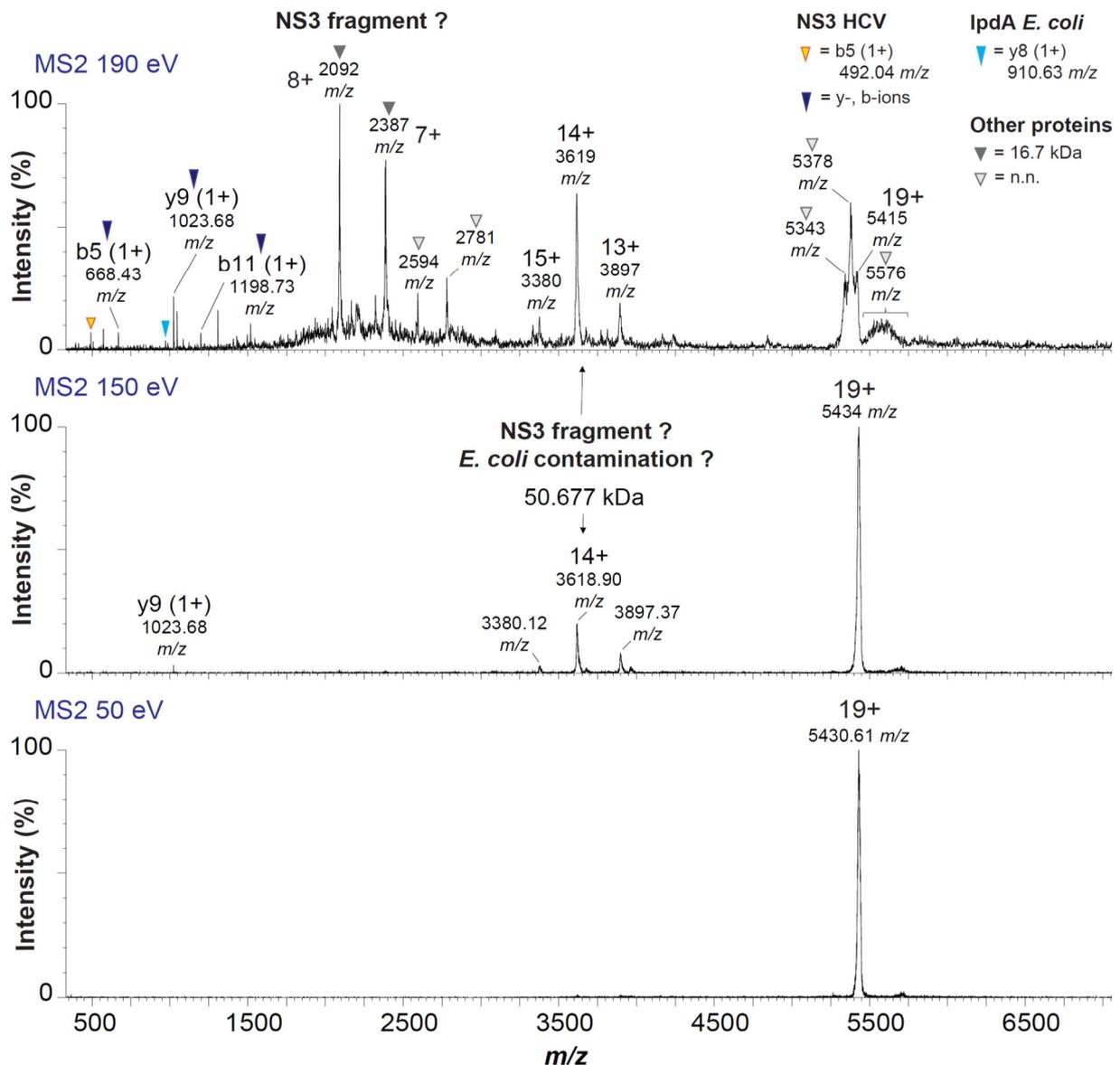
S. Fig. 14 Comparison of 150 eV in MS1modd with MS2 CID (19+ precursors peak) of the protein complex in the HCV sample shows the dissociation of the 51 kDa protein in each case.

The figure is a supplement to sub-chapter 3.1.1. Raw data are based on a protein concentration of 5 μ M, 200 mM ammonium acetate, pH 7.2 + 2 mM DTT. Measurements were performed in positive ion mode.



S. Fig. 15 Top-down dissociation of the 51 kDa protein at 170 eV (13+ 3900 *m/z*) shows specific low mass peptides for *E. coli* IpdA as well as for a sequentially shortened HCV NS3 protein.

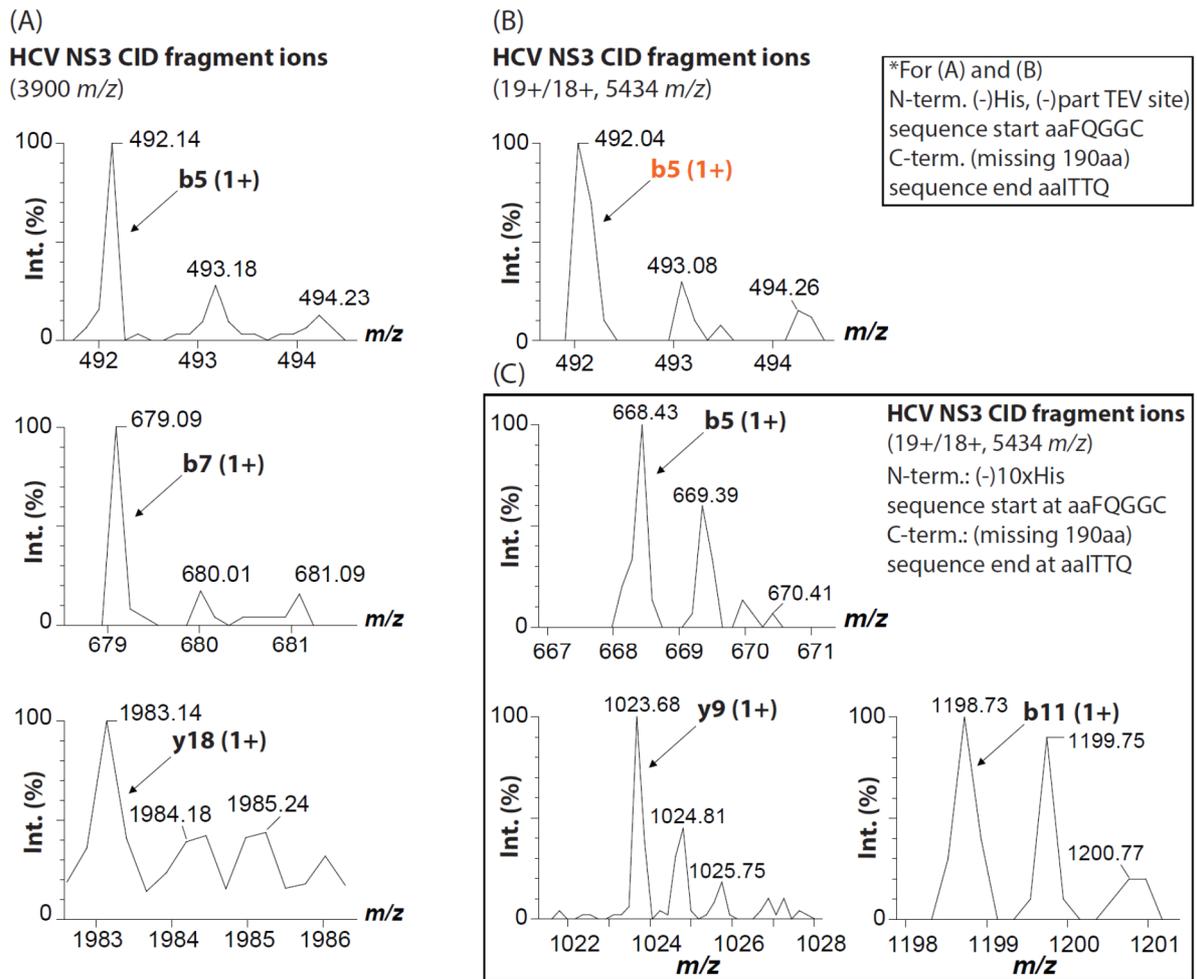
A zoom into the specific *E. coli* CID fragment ions is shown. NS3 fragment ions are marked with an orange or orange bordered triangle (missing amino acids: N-term. (-)His, (-)part TEV site, sequence start at aaFQGGC, C-term. (missing 190aa) sequence end at aaITTQ). The figure is a supplement to sub-chapter 3.1.1. Raw data are based on a protein concentration of 5 μ M, 200 mM ammonium acetate, pH 7.2 + 2 mM DTT.



S. Fig. 16 Top-Down CID fragmentation of the 19+ protein peak (5434 m/z) identified certain low mass peptides of a possibly sequentially shortened HCV NS3-NS4A protein (51 kDa) as well as an *E. coli* protein (IpdA 51 kDa), which was already identified during the proteomics analysis.

NS3 fragment ions that were also identified while dissociation of the (13+) 3900 m/z peak of the 51 kDa protein are marked with an orange bordered triangle (missing amino acids: N-term. (-)His, (-)part TEV site, sequence start at aaFQGGC, C-term. (missing 190aa) sequence end at aaITTQ). NS3 fragment ions that were identified while dissociation of the (19+/18+) 5434 m/z peak are labelled with dark-blue triangles (missing amino acids: N-term. (-)10xHis sequence start at aaFQGGC, C-term. (missing 190aa) sequence end at aaITTQ). The identified *E. coli* peptide fragment is marked with a light-blue triangle. Unidentified proteins (n.n.) are marked with gray triangles. The figure is a supplement to sub-chapter 3.1.1. Raw data are based on a protein concentration of 5 μ M, 200 mM ammonium acetate, pH 7.2 + 2 mM DTT. Measurements were performed in positive ion mode at 50 to 190 eV energy.

HCV NS3-protein: MS2 CID low mass fragment ions 150 - 190 eV



S. Fig. 17 Zoom into the low mass CID fragments of the sequentially shortened HCV NS3 protein. The figure is a supplement to sub-chapter 3.1.1 and an addition to S. Fig. 13. Raw data are based on a protein concentration of 5 μM , 200 mM ammonium acetate, pH 7.2 + 2 mM DTT. Measurements were performed in positive ion mode at 150 to 190 eV energy.

Note of Thanks

Here, I would like to express my great thanks to all the people who have supported me in my work.

I would like to thank my supervisors Dr. Charlotte Uetrecht and Prof. Dr. Julia Kehr for the continuous support during my PHD studies. I would like to especially thank you for the interesting and motivating discussions during my time at HPI as well as for the many exciting research projects in which I was able to work. I would also like to thank you in particular for the great understanding and support during my temporary suspension of work for family reasons. I couldn't have imagined better managers who made this job so much more enjoyable.

I would particularly like to thank Prof. Dr. Julia Kehr for her great support as part of the DELIGRAH graduate program at the University of Hamburg and that she was willing to take over the co-supervision of my work without any bureaucratic effort.

Further, I would also like to thank Prof. Dr. Christian Betzel from the University of Hamburg for standing by my side as the second DELIGRAH co-supervisor, as well as the other speakers who are present at my disputation on this work.

A great thank also goes to Prof. Norbert Tautz's group, as well as to Felix Rey and Alejandra Tortorici, who gave me great support in this project with advice and protein samples.

Moreover, I would like to thank all of my work colleagues for a great time and great cooperation. This is especially true of Boris Krichel, who was my greatest helper in times of need. Johannes Heidemann, who started his thesis shortly before me and finished it long before me, Jasmin Dülfer for her great support in proofreading this thesis and for a great time together at the HPI. Ronja Pogan and Janine-Deise Kopicki, without their outstanding sense of order and their cheerful and cordial manner, the work would only have been half as nice and Hao Yan, called Panda, who really rounded off this community with his Chinese flair. I would also like to thank Jürgen, Alan and Knut for an inspiring time together in AG77.

Additionally, I would like to thank Martina and my friends Elaine, Bianca and Bettina for a great study time together in Bayreuth.

Last but not least, I would like to thank my family, my parents and grandparents for their great financial support throughout my studies and through to my PhD.

I would also like to sincerely thank my two siblings for their unwavering spiritual support during this time.

Especially, I would like to thank my fiancé Pascal Steffen, who was the greatest moral and mental support for me at the end of my work and who was always happy to take care of our children.

Without you I would never have made it this far.

Thank you very much!

Declaration of Honor

I declare that I independently prepared the present thesis and that I did not use any literature or resources other than those indicated. Citations are all marked and their origin is specified. This also applies to figures, tables and illustrations, as well as to online sources. Additionally, I declare that the digital version submitted matches invariably the printed version in content and wording. I assure you that this dissertation was not submitted in a previous doctoral procedure.

J. Loeb Grause Bäumer

Hamburg, the 17.09.2021

Martina Tollgreve
English-German Teacher
Montreal, Quebec, Canada

August 27, 2021

English language certificate

Julia Lockhauserbäumers's Ph.D. thesis entitled "**Structural analysis of the *Flaviviridae* non-structural protein complex NS3-NS4A with native mass spectrometry**" is written in fluent English.

I confirm that the language is understandable and properly articulated.

Kind regards,



Prof. Mark Molloy

September 15, 2021

English language certificate

Julia Lockhauserbäumers's Ph.D. thesis titled "**Structural analysis of the *Flaviviridae* non-structural protein complex NS3-NS4A with native mass spectrometry**" is written in fluent English.

As a native English speaker, I confirm that the language is understandable and properly articulated.

Kind regards,



Mark P. Molloy, Ph.D.

Professor of Bowel Cancer Research

The University of Sydney, Australia