Proteome-wide cross-linking mass spectrometry to identify specific virus capsid-host interactions between tick-borne encephalitis virus and neuroblastoma cells

Sarah V. Barrass[1,2], Lauri I. A. Pulkkinen[1,2], Olli Vapalahti[3,4,5], Suvi H. Kuivanen[3] ,Maria Anastasina[1,2*], Lotta Happonen[6*], Sarah J. Butcher[1,2*]

[1] Faculty of Biological and Environmental Sciences, Molecular and Integrative Bioscience Research Programme, University of Helsinki, Helsinki, Finland

[2] Helsinki Institute of Life Sciences-Institute of Biotechnology, University of Helsinki, Helsinki, Finland

[3] Department of Virology, University of Helsinki, Helsinki, Finland

[4] Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland

[5] Division of Clinical Microbiology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[6] Lund University, Faculty of Medicine, Department of Clinical Sciences Lund, Infection Medicine, Lund, Sweden

* Corresponding authors

Emails: sarah.butcher@helsinki.fi (SB), lotta.happonen@med.lu.se (LH), maria.anastasina@helsinki.fi (MA)

Abstract

Virus-host protein-protein interactions are central to viral infection, but are challenging to identify and characterise, especially in complex systems involving intact viruses and cells. In this work, we demonstrate a proteome-wide approach to identify virus-host interactions using chemical cross-linking coupled with mass spectrometry. We adsorbed tick-borne encephalitis virus onto metabolically-stalled neuroblastoma cells, covalently cross-linked interacting virus-host proteins, and performed limited proteolysis to release primarily the surface-exposed proteins for identification by mass spectrometry. Using the intraviral protein cross-links as an internal control to assess cross-link confidence levels, we identified 22 high confidence unique intraviral cross-links and 59 high confidence unique virus-host protein-protein interactions. The identified host proteins were shown to interact with eight distinct sites on the outer surface of the virus. Notably, we identified an interaction between the substrate-binding domain of heat shock protein family A member 5, an entry receptor for four related flaviviruses, and the hinge region of the viral envelope protein. We also identified host proteins involved in endocytosis, cytoskeletal rearrangement, or located in the cytoskeleton, suggesting that entry mechanisms for tick-borne encephalitis virus could include both clathrin-mediated endocytosis and macropinocytosis. Additionally, cross-linking of the viral proteins showed that the capsid protein forms dimers within tick-borne encephalitis virus, as previously observed with purified C proteins for other flaviviruses. This method enables the identification and mapping of transient virus-host interactions, under near-physiological conditions, without the need for genetic manipulation.

Author summary

Tick-borne encephalitis virus is an important human pathogen that can cause severe infection often resulting in life-long neurological complications or even death. As with other viruses, it fully relies on the host cells, and any successful infection starts with interactions between the viral structural proteins and cellular surface proteins. Mapping these interactions is essential both for the fundamental understanding

of viral entry mechanisms, and for guiding the design of new antiviral drugs and vaccines. Here, we stabilise the interactions between tick-borne encephalitis virus and human proteins by chemical cross-linking. We then detect the interactions using mass spectrometry and analyse the data to identify protein-protein complexes. We demonstrate that we can visualise the protein interaction interfaces by mapping the cross-linked sites onto the host and viral protein structures. We reveal that there are eight distinct sites on the outer surface of the viral envelope protein that interact with host. Using this approach, we mapped interactions between the tick-borne encephalitis virus envelope protein, and 59 host proteins, identifying a possible new virus receptor. These results highlight the potential of chemical cross-linking coupled with mass spectrometry to identify and map interactions between viral and host proteins.

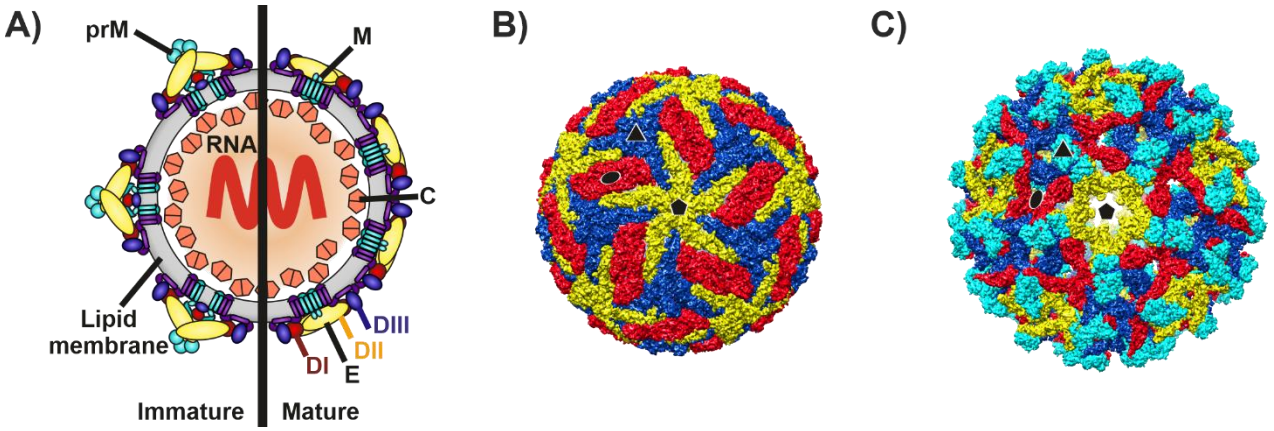
Introduction

Viruses are obligatory intracellular parasites that depend on virus-host protein-protein interactions (PPIs) to establish successful infections. The identification of these interactions and knowledge of the interaction interfaces contribute to our understanding of the initial steps of the viral life cycle, and can guide the design of antivirals and vaccines [1–6].

Advances in high throughput methods have led to the large-scale identification of virus-host interactions, but the structural characterisation of these interactions is often still limited [7]. Affinity purification coupled with mass spectrometry, yeast two-hybrid, and protein microarrays, have identified virus-host PPIs for multiple viruses including: Japanese encephalitis virus, H1N1 influenza, human immunodeficiency virus, human cytomegalovirus, and severe acute respiratory syndrome coronavirus 2 [8–12]. These methods are however limited in their applicability to detect transient interactions between wild-type viruses and cells. Alternative methods, using chemical cross-linking, or proximity labelling (BioID and TurboID), demonstrate improved detection of weak and transient interactions [13–18]. In these approaches, cells are probed with modified viral proteins conjugated to trifunctional cross-linkers or biotin ligases. Host proteins in close proximity to the viral bait are then permanently cross-linked or biotinylated, and purified using the biotin or cross-linker tag. Enriched proteins are detected by comparison of the protein signal to that in negative

68 controls using bottom-up proteomics. Alternative chemical cross-linking mass spectrometry (XL-MS)
69 workflows that directly detect the cross-linked peptides additionally provide information about the
70 interaction interfaces. Previous studies have used the finite length of the chemical cross-linker to indicate
71 the proximity of two amino acid side chains during the cross-linking reaction, and to build structural models
72 of bacteria-host PPIs [19,20]. This shows the potential of XL-MS in both the identification of PPIs and the
73 characterisation of the binding interface.

74 The flavivirus, tick-borne encephalitis virus (TBEV) is the causative agent of one of the most important
75 arbovirus-caused diseases in Europe, Russia, and Northern China [21,22]. Symptomatic infection with TBEV
76 can cause meningitis, encephalitis, and meningoencephalitis, and often results in life-long neurological
77 complications or death [23,24]. The TBEV virion has three different structural proteins, the envelope
78 protein (E protein), membrane protein (M protein) and capsid protein (C protein), in addition to a lipid
79 bilayer and an ~11 kilobase-long positive-strand RNA genome (Fig 1). The E protein forms the smooth outer
80 surface of the virion and is responsible for receptor binding [25,26]. The atomic structure of the mature
81 TBEV virion E and M proteins has been solved by cryo-electron microscopy at a resolution of 3.9 Å and the
82 crystal structure of the E protein at a resolution of 1.9 Å (Fig 1) [25,27]. Non-infectious immature and
83 partially immature viruses also egress from cells, and have a spikey surface (Fig 1) [27]. No proteome-wide
84 study of TBEV virus-host protein interactions has been published to our knowledge.



85

86 **Fig 1: TBEV mature and immature virus structures.** A) Schematic representation of TBEV with the
87 immature structure shown on the left and the mature on the right. Multiple copies of the C protein dimer

88 surround the genome forming the nucleocapsid complex, positioned beneath the lipid bilayer. The surface
89 of the immature virion is covered in 60 spikes, composed of trimers of prM (pre-membrane)-E
90 heterodimers embedded into the lipid bilayer. The surface of the mature virion is covered in 90 E-M
91 heterotetramer complexes embedded into the lipid bilayer. TBEV E protein domain I is shown in red,
92 domain III in dark blue, domain II in yellow and domain 4 in purple. B) Surface representation of the mature
93 TBEV virion (PDB accession: 5O6A)[27]. The three E proteins within each asymmetric unit are shown in
94 blue, red, and yellow. Symmetry axes are indicated by the black pentagon (five-fold), triangle (three-fold),
95 and ellipse (two-fold). C) Surface representation of the immature Spondweni virus, a related flavivirus (PDB
96 accession 6ZQW) [28]. The three E proteins within each asymmetric unit are shown in blue, red, and yellow,
97 and the prM protein is shown in cyan. Symmetry axes are indicated by a black pentagon (five-fold), triangle
98 (three-fold), and ellipse (two-fold).

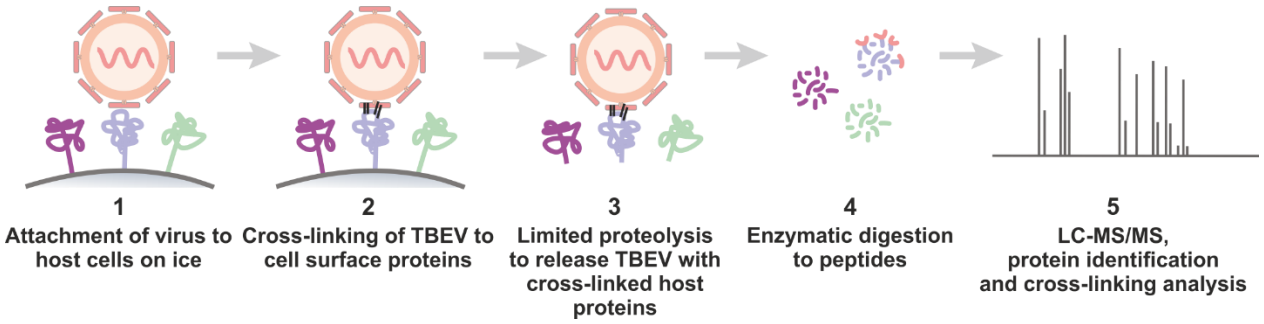
99 In this large-scale proteomics study, we used XL-MS to identify the interaction interfaces of PPIs between
100 TBEV and the surface of human neuroblastoma (SK-N-SH) cells. Here, the homobifunctional chemical cross-
101 linker disuccinimidyl suberate (DSS) was used to covalently fix PPIs by cross-linking primary amine
102 containing residues (the side chain of lysine residues or the N-terminus of the protein). The finite length of
103 DSS (11.4 Å) imposed a maximum distance between cross-linked residues and was used to validate
104 intraviral crosslinks by measuring their distances on TBEV proteins with known structures or reliable
105 homology models. The final dataset was filtered using the intraviral cross-links as an internal control,
106 leading to the identification 59 unique high confidence interactions between the TBEV E protein and
107 cellular proteins.

108 Results

109 Identification of cross-linked peptides

110 To identify interactions between the mature TBEV virion and host proteins, we incubated the virus with
111 metabolically-stalled neuroblastoma cells on ice, allowing for TBEV to bind to the cells, but preventing

112 subsequent internalization. The TBEV-host PPIs were then stabilized and fixed by chemical cross-linking
113 with DSS (Fig 2). To reduce the sample complexity and search space during data analysis, we used limited
114 proteolysis to release primarily cell-surface associated host proteins. The released proteins were digested
115 to peptides and analysed by liquid chromatography tandem mass spectrometry (LC-MS/MS) followed by
116 label-free data dependent acquisition (DDA) quantitation to determine their relative abundance (S1 Table).
117 Identified proteins were used to generate smaller, defined sets of sequences to use in the cross-linking data
118 analysis workflow.



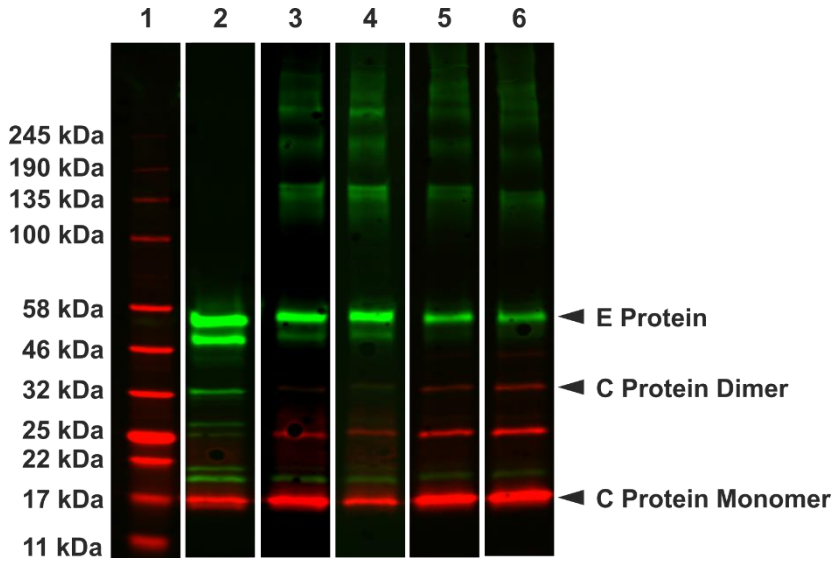
119

120 **Fig 2: Schematic representation of cross-linking workflow.** 1. TBEV was allowed to attach to metabolically-
121 stalled SK-N-SH cells. 2. TBEV-host PPIs were stabilised by chemical cross-linking with DSS. 3. Cell-surface
122 associated proteins, and cross-linked TBEV were released from the cell surface using limited proteolysis. 4.
123 The released proteins were digested to peptides. 5. Peptides were analysed by LC-MS/MS and host and
124 viral proteins identified and quantified using label-free DDA. Identified host proteins were analysed to
125 identify cross-links between the host proteins and TBEV.

126 For cross-linking, we used four different cross-linker concentrations, in addition to a negative control
127 sample to which no cross-linker was added. Each condition was repeated in triplicate, and the experiment
128 repeated three times independently with different TBEV preparations and cell line passages, yielding 9
129 replicates per cross-linker concentration. The samples were initially analysed by immunoblotting of the
130 TBEV E and C Proteins (Fig 3). The presence of higher molecular weight bands greater than 100 kDa in
131 samples treated with DSS confirms cross-linking (Fig 3). The C protein has been shown to form antiparallel
132 dimers in the crystal and NMR structures of other flaviviruses [29–31]. We identified a band with a

133 molecular weight corresponding to that of C protein dimers, indicating that the C protein dimerizes in TBEV

134 (Fig 3).



135

136 **Fig 3: Immunoblot analysis of the TBEV E and C proteins in cross-linked samples.** The TBEV C protein in

137 shown in red and the TBEV E protein in green. Lane 1- protein marker; Lane 2- negative control with 0 mM

138 DSS, Lane 3- cross-linking with 0.1 mM DSS; Lane 4- cross-linking with 0.25 mM DSS; Lane 5- cross-linking

139 with 0.5 mM DSS; Lane 6- cross-linking with 1mM DSS. Higher molecular weight bands greater than 100 kDa

140 corresponding to the cross-linking of E and C to other proteins can be seen in lanes 2-5. Lower molecular

141 weight bands less than 50 kDa corresponding to the partial cleavage of the viral proteins during the limited

142 proteolysis step are also observed in all lanes.

143 Identification of cross-linked peptides is computationally challenging as all primary amine-primary amine

144 combinations in a given sequence database need to be considered. To reduce the search space for cross-

145 linked peptide identification, proteins identified by DDA were probed for cross-links in batches (see

146 materials and methods). A total of 7167 spectral observations of cross-linked peptides were identified using

147 pLink2 at a false discovery rate (FDR) of 5%, excluding interfaces supported by cross-linked peptides

148 identified in the negative controls, which correspond to false positives likely arising from erroneous peptide

149 matches in the complex proteome background (S2 Table) [32]. Spectral observations of cross-links between

150 two peptides within the same protein (intraprotein PPIs) accounted for the majority of the observations

151 (5589), compared to 1578 for those identified between two peptides from different proteins (interprotein
152 PPIs). Intraprotein cross-links have been consistently shown to make up a higher proportion of spectral
153 observations within cross-linked datasets, as two residues within the same protein are highly likely to be in
154 close physical proximity within the cell, leading to an increased cross-linking frequency [33,34]. In total,
155 1697 different cross-linked interfaces were observed, and on average each interface was supported by 4.2
156 spectral observations. The cross-linked interfaces map to a network of 698 PPIs, consisting of 588 host
157 proteins and the 3 viral structural proteins. Overall, 66.3 % of the unique PPIs were attributed to
158 interactions between host proteins, 33.5 % to virus-host interactions and only 0.2 % to intraviral
159 interactions. We also identified intraprotein cross-links in 297 host proteins and the TBEV C and E proteins.

160 The confidence of the cross-linking dataset can be investigated by examining the spatial distances between
161 cross-linked residues for protein complexes where high-resolution structures or reliable homology models
162 are available. As intraviral protein interactions account for 62 % of the detected spectral observations, we
163 used the intraviral cross-links as an internal control to assess the confidence level for the dataset.

164 Mapping of intraviral cross-links

165 The published mature TBEV structure and homology models of the immature virus and C protein dimer
166 were used to accurately measure cross-link distances. The C-score of the homology models were -0.65 (C
167 protein), 2.00 (E protein) and 0.57 (prM protein) [35–37] . We measured the distance between cross-linked
168 residues and applied a maximum distance constraint of 30 Å between the lysine or N-terminus Cα (Table 1
169 and Figs 1 and 4) [38]. In total, 24 cross-links were mapped onto the viral structural proteins, and 22 fell
170 within the accepted distance range. Overall, 14 of the cross-links satisfied the distance constraint in both
171 the mature and immature TBEV structures, six were only acceptable in the mature structure and two in the
172 immature structure. Cross-links with acceptable distances in only the mature structure were identified by
173 2089 spectral observations compared to 31 for those only accepted in the immature structure,
174 demonstrating that the majority of the virus particles in the analysed samples were mature.

175 **Table 1: Intraviral cross-links identified, the corresponding average expectation value and SVM score of**

176 **the peptide spectrum matches, and distances in the mature and immature virus structures**

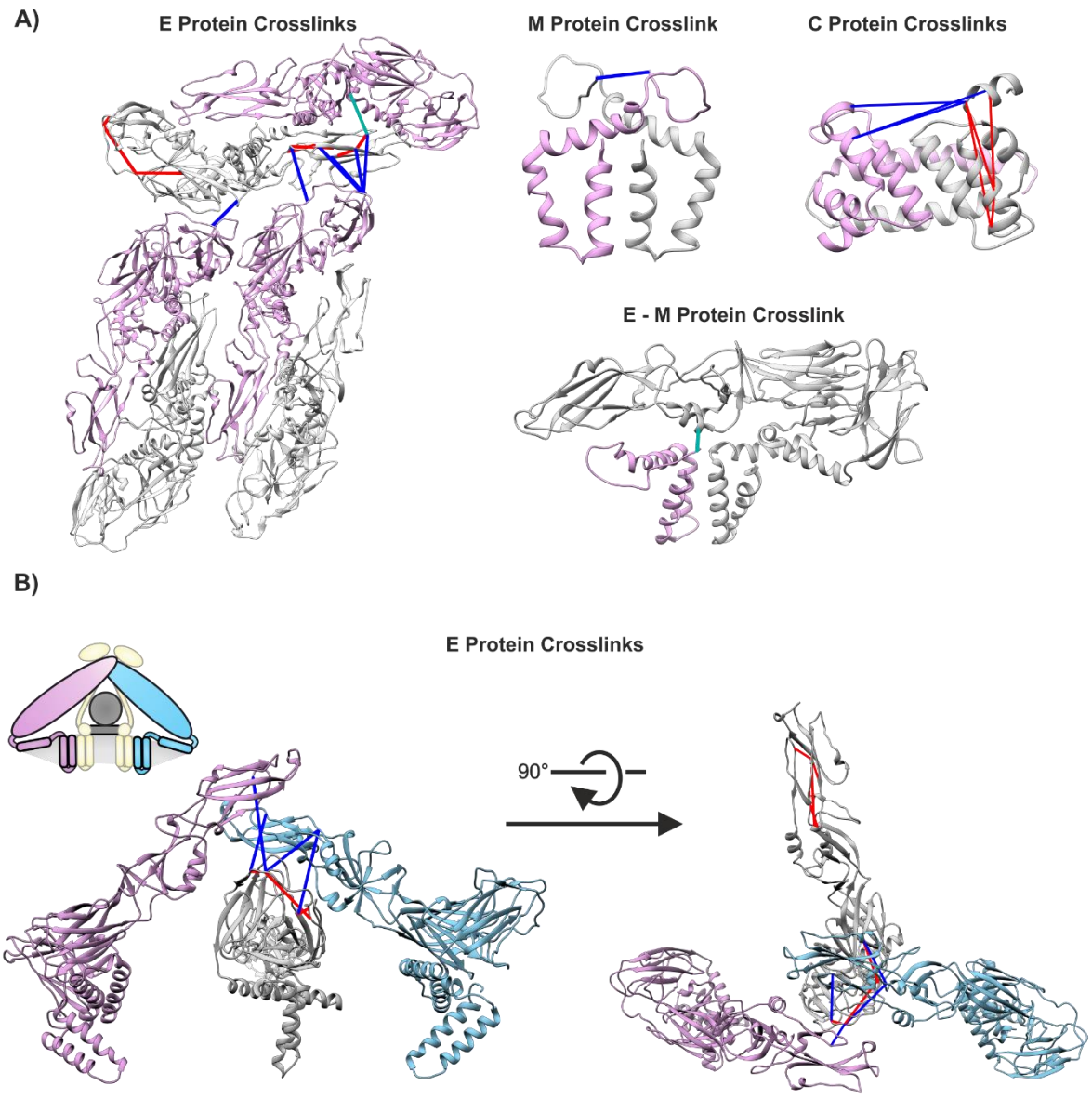
Cross-link Type	Protein 1	Protein 2	Residue 1*	Residue 2*	Mean E-Value	Mean SVM Score	Number of Spectral observations	Distance Between Cα Carbons Virion	Distance Between Cα Carbons Immature Virus
Intradimer/ Intertrimer	E Protein	E Protein	251	284	0.59	0.07	1346	16.74	47.74^
Intramonomer	E Protein	E Protein	64	126	0.25	0.08	939	12.10	11.95
Intramonomer	E Protein	E Protein	300	161	0.02	0.02	910	17.42	15.99
Intramonomer	E Protein	E Protein	118	69	0.13	0.11	315	9.47	9.08
Interdimer/ Intertrimer	E Protein	E Protein	136	309	0.48	0.10	265	17.25	32.41^
Interdimer/Intertrimer	E Protein	E Protein	309	69	0.09	0.10	172	17.88	30.87^
Intradimer	E Protein	M Protein	266	1	0.42	0.15	155	8.79	N/A
Interdimer/ Intertrimer	E Protein	E Protein	118	309	0.34	0.06	150	18.54	30.67^
Intramonomer/ Intradimer	C Protein	C Protein	10	25	0.41	0.06	63	23.18/23.68	23.18/23.68
Intramonomer	E Protein	E Protein	64	69	0.35	0.16	34	15.26	14.58
Interdimer/ Intertrimer	E Protein	E Protein	251	311	1.00	0.03	23	31.63^	28.63
Intramonomer/ Intradimer	C Protein	C Protein	20	10	0.09	0.02	19	16.45/24.22	16.45/24.22
Intramonomer/ Intradimer	C Protein	C Protein	20	8	1.00	0.35	12	19.30/28.73	19.30/28.73
Interdimer/Intertrimer	E Protein	E Protein	251	309	1.00	0.03	10	26.00	29.38
Intramonomer	E Protein	E Protein	251	69	1.00	0.03	9	9.11	8.26
Interdimer/ Intertrimer	E Protein	E Protein	64	161	1.00	0.06	8	39.39^	30.06
Interdimer/Intertrimer	E Protein	E Protein	64	309	1.00	0.30	4	25.91	27.17
Intramonomer	E Protein	E Protein	309	407	1.00	0.41	3	41.09^	43.78^
Intramonomer	C Protein	C Protein	20	25	0.69	0.22	3	7.12/32.63^	7.12/32.63^
Intramonomer	E Protein	E Protein	309	408	1.00	0.50	2	39.54^	41.45^
Intramonomer	E Protein	E Protein	69	124	1.00	0.62	1	26.85	25.98
Intramonomer	E Protein	E Protein	298	309	1.00	0.23	1	25.96	26.46
Intramonomer	E Protein	E Protein	311	309	1.00	0.22	1	6.04	5.60
Interdimer	M Protein	M Protein	1	1	1.00	0.70	1	19.70	N/A

177 *The TBEV polyprotein sequence was obtained from GenBank (accession: AWC08512.1). The C protein corresponds to

178 residues 1-96; the M protein 206-280 and the E protein 281-776. Residue numbers in the table begin from 1 at the

179 start of each protein.

180 ^Distances between cross-linked residues that are greater than 30 Å and are not considered feasible.



181

182 **Fig 4: Mapping of intraviral crosslinks.** Identified intraviral cross-links with the distances ≤ 30 Å mapped to
183 the viral structural proteins are shown above in red, blue and teal. Red cross-links corresponding to
184 intramonomer cross-links, teal intradimer cross-links, and blue interdimer or intertrimer cross-links. A)
185 Cross-links mapped to the known cryoEM structure of the TBEV virion or the C protein homology model. B)
186 Cross-links mapped to the generated homology model of the immature virus, E proteins in trimer 1 are
187 coloured in plum and light blue and the E protein in trimer 2 is coloured grey. The schematic diagram
188 additionally shows the position of the lipid bilayer (light grey) and the prM protein (light yellow).

189 Locating the cross-links on the virus structures with distance constraints applied, allowed us to distinguish
190 between three different types of E-E cross-links in the mature virion: intramonomer, intradimer and

191 interdimer cross-links, and two in the immature virus: intramonomer and intertrimer cross-links. As
192 expected, intramonomer E cross-links show similar distances in the mature and immature structures and
193 accordingly the distance constraints are satisfied equally well for both. In contrast, as the E proteins
194 rearrange from trimers to dimers upon virus maturation, only 2 cross-links satisfy the distance constraint in
195 both the intertrimer positions found in the immature virus and the intradimer or interdimer positions
196 found in the mature virus. Two types of cross-links are possible within the C protein dimer, intramonomer
197 and intradimer. In our dataset, three of the four identified C protein cross-links satisfy the distance
198 constraints for both interaction types, making it impossible for us to distinguish between these two
199 alternatives.

200 Two of the intramonomer cross-links between E protein residues 309 and 407 or 408 demonstrate cross-
201 linking lengths of ca. 40 Å in both the mature and immature structures. The cross-links were identified by a
202 low number of spectral observations, 3 for 309-407 and 2 for 309-408, and may arise from interactions
203 between disrupted virions, free E proteins, alternative viral conformations or be false positives. To
204 distinguish between these alternatives, we examined the cross-linked spectra in detail (Fig 5). Fig 5A shows
205 representative spectra for cross-links that satisfy the distance constraint and are identified by a high
206 number of spectral observations, and Fig 5B and C show representative spectra for the two questioned
207 crosslinks. As compared to the spectra shown in Fig 5A, spectra in Fig 5B and C show both a low signal-to-
208 noise ratio and a low sequence coverage for the peptide fragments, and most likely represent false positive
209 hits from background noise.

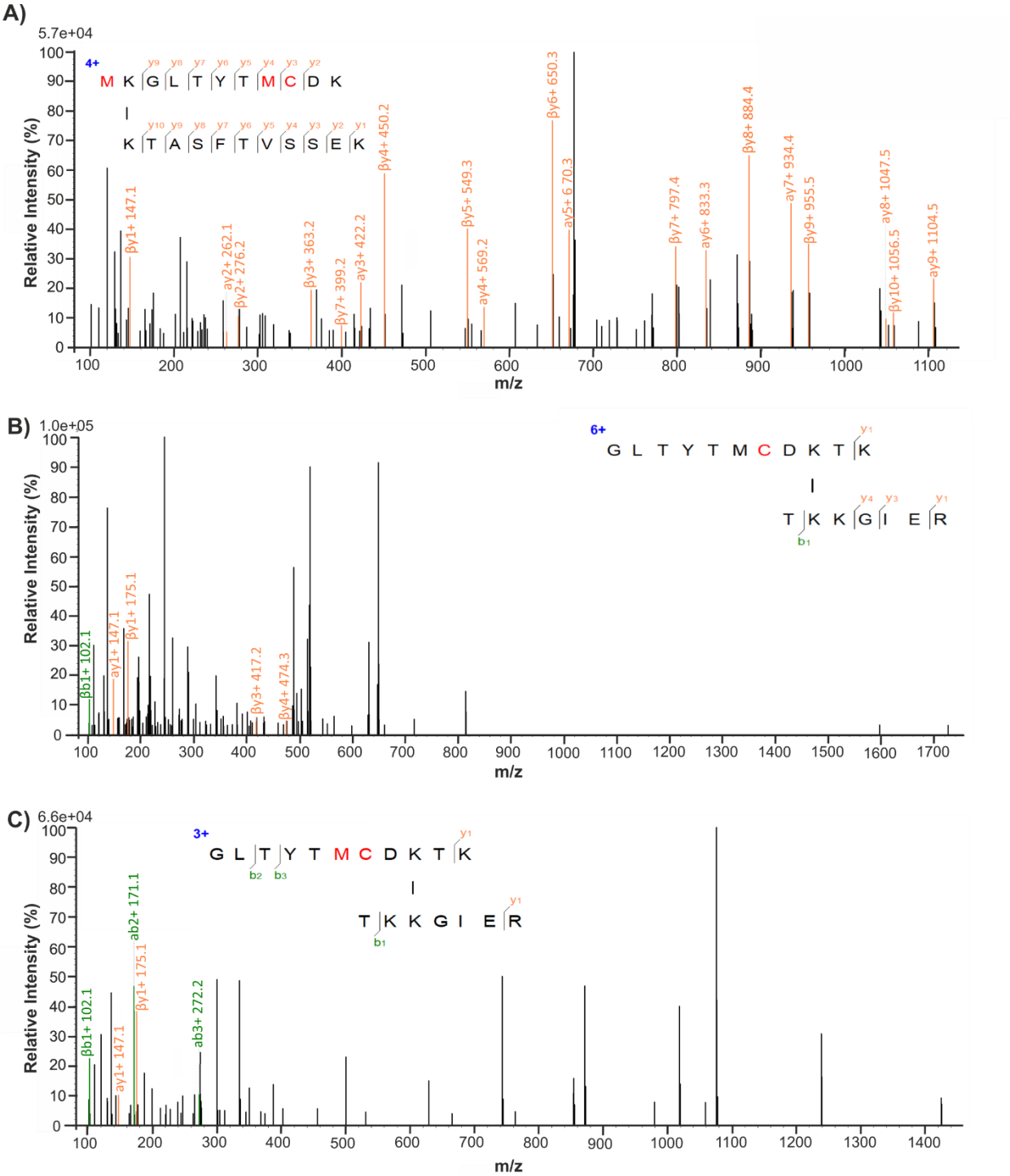


Fig 5: Representative spectra of E protein cross-links that satisfy the distance constraint (< 30 Å) and cross-links with distances > 30 Å. The cross-linked peptides are shown with the longer α peptide positioned above the shorter β peptide and a connecting line between the cross-linked residues. Modified residues in the peptides are shown in red and the detected beta and gamma fragments shown in green and orange respectively. Peaks corresponding to the detected fragments are coloured and labelled accordingly. The

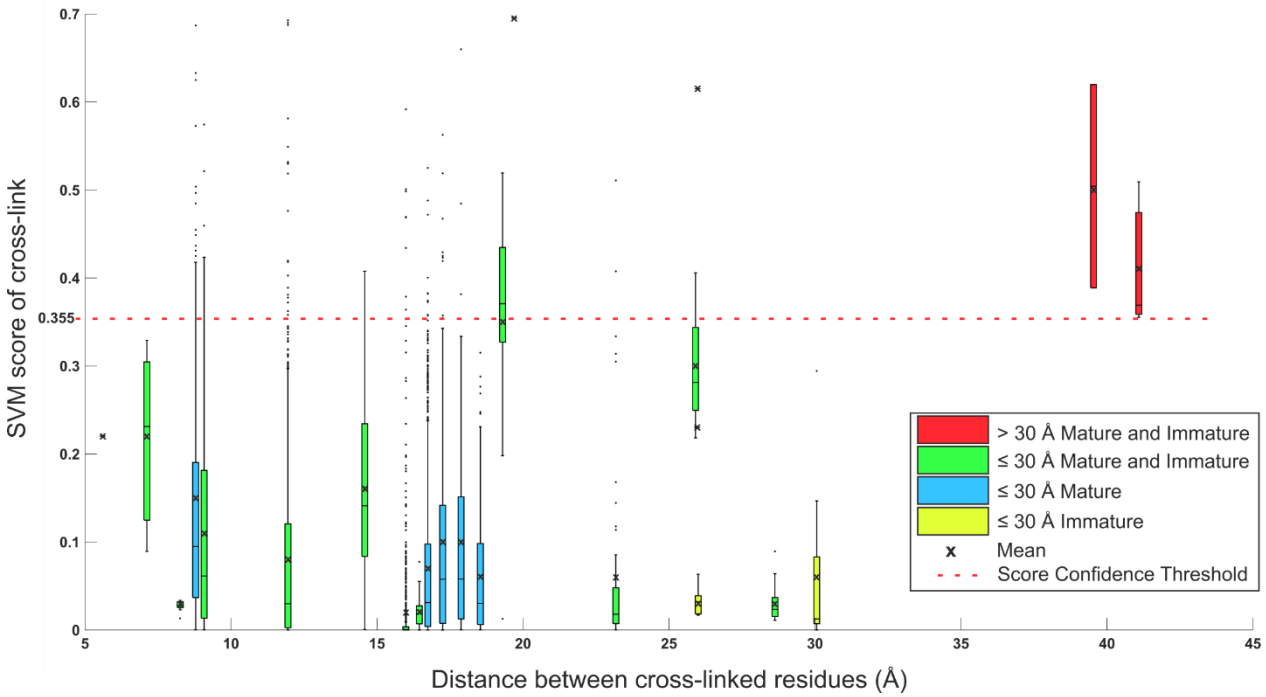
216 charge state of the cross-linked peptide is shown in blue. A) Representative spectrum of E protein cross-link
217 300-161 that are less than 30 Å and show a high number of spectral observations. B) Representative
218 spectrum for the cross-link 309-407, that has a cross-linking distance of 41.09 Å in the mature structure and
219 43.78 Å in the immature structure. C) Representative spectrum for the cross-link 309-408 that has a cross-
220 linking distance of 39.54 Å in the mature structure and 41.45 in the immature structure. Cross-links with
221 distances > 30 Å show poor fragmentation and signal to noise ratio in the cross-linked spectral (B and C),
222 compared to cross-links with distances < 30 Å (A).

223 Based on the imposed distance constraints, we calculated that 91.7 % of the intraviral cross-linked
224 interfaces are identified with high confidence, at a distance threshold of <30 Å indicating that our approach
225 allows detection of specific intra- and inter-protein cross-links with high confidence.

226 Filtering the cross-linking dataset

227 The program pLink2 provides two parameters for assessing the confidence of cross-linked peptide
228 spectrum matches (PSM), the expectation value (E-value) and the SVM score [32]. Both the E-value and
229 SVM score describe the probability of a cross-linked PSM being a random match, and have values ranging
230 from 1 to 0, where the smaller the value the more confident the PSM. The SVM score is calculated for every
231 PSM, acting as the prime measure for FDR estimation, whereas the E-value is only calculated for PSMs that
232 pass the FDR threshold [32]. In addition, cross-links may occur due to the sporadic proximity of proteins in
233 the sample, or due to specific cross-linking of interacting proteins. Cross-links identified by more spectral
234 observations have a higher probability of reflecting specific interactions. Therefore, the number of spectral
235 observations per cross-linked interface can provide information about the cross-linking confidence on the
236 protein-protein interaction level. Previous studies have used the E-value, SVM score, number of spectral
237 observations or a combination of the aforementioned to assess the confidence of the identified cross-links
238 but no standardized values have been established [32,39,40]. Here, we investigated the correlation
239 between the E-value, SVM score, number of spectral observations and the acceptable distance constraint
240 (<30 Å) as measured for our intraviral cross-links, to determine confidence parameters for the dataset.

241 Our data show a correlation between the measured distance of each unique intraviral cross-link and the
242 calculated SVM scores, but not the E-values (Fig. 6, S1 Fig). Consequently, the E-value was not considered
243 as a suitable measure of confidence for this dataset. We observed that intraviral cross-links with distances
244 > 30 Å in both the mature and immature structures (indicated as red bars in Fig. 6) had the 3rd and 4th
245 highest mean SVM scores of the intraviral cross-links (Table 1). Furthermore, cross-links with higher
246 mean SVM scores were only identified by one spectral observation (Table 1). Based on the SVM scores of
247 intraviral cross-links with distances > 30 Å in both the mature and immature structures, we imposed a SVM
248 score threshold of < 0.355, hence discarding 884 of the identified cross-linked interfaces. Interestingly, 97.8
249 % of the excluded interfaces were identified by only one spectral observation.



250

251 **Fig 6: Box and whisker plot of the SVM scores for the intraviral cross-links, plotted against the distance**
252 **between the cross-linked residues for each cross-link.** The bars are coloured based on whether the
253 distance constraint is satisfied for both the mature and immature conformations (green), only the mature
254 conformation (blue), only the immature conformation (yellow) or not satisfied in either the mature or
255 immature conformation (red). The mean SMV score is marked with a cross. A SVM score confidence
256 threshold of < 0.355 is shown by a red dashed line.

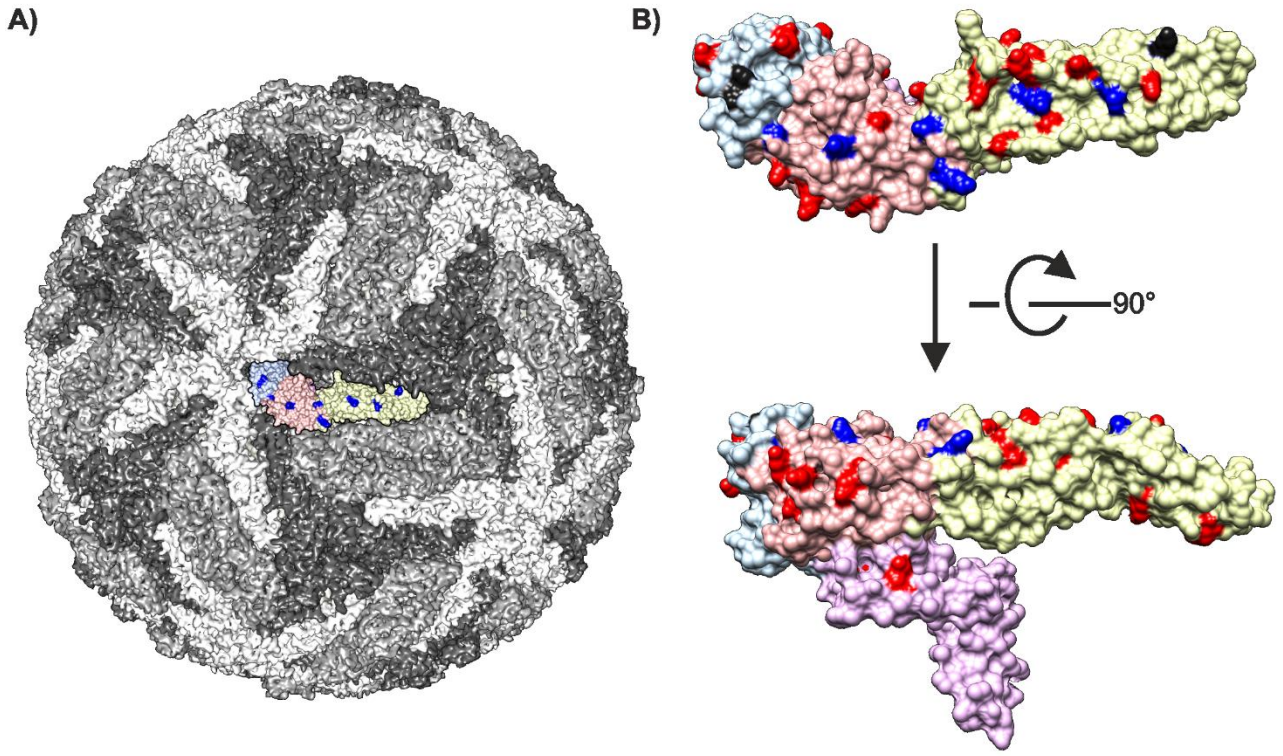
257 The number of cross-linked spectral observations for 101 randomly selected proteins in the dataset was
258 compared to the average spectral count of these proteins across all the samples in the DDA analyses (S1
259 File). The number of spectral observations for cross-linked peptide pairs for these same 101 selected
260 proteins was also compared to the number of lysine residues present within each protein (S1 File). No
261 correlation was observed in either of these cases; so, neither protein abundancy nor lysine content affect
262 the data content (S1 File). Therefore, the number of spectral observations was used in a non-biased
263 manner to assess the confidence of each unique cross-link to reduce the dataset for analysis further.

264 Here, we filtered the data in a stepwise manner, first on the PSM level using the SVM score and secondly
265 on the protein-protein interaction level using the number of spectral observations. A cross-linked interface
266 was considered to be of high confidence if it had an SVM score < 0.355 and was identified by ≥ 2 spectral
267 observations. Filtering the data in the reverse order would lead to inclusion of interactions only supported
268 by one high confidence spectral observation. The filtered cross-linking dataset consisted of 218 cross-linked
269 interfaces, 36.7 % of which were attributed to virus-host PPIs (S3 Table).

270 Virus-host protein-protein interactions

271 Virus-host PPIs were identified between TBEV and 61 host proteins in the filtered cross-linking dataset. In
272 total, 59 proteins were shown to interact with the TBEV E protein and 2 with the M protein. Cross-links with
273 the M protein form between N-terminal serine of mature M and the host proteins. As the M protein is
274 buried and not accessible for cross-linking in the mature virion, these interactions likely occur between the
275 host proteins and disrupted virions, free M protein, or conformations that have yet to be described.

276 E -host PPIs were mapped to eight different lysine residues on the outer surface of the virion (E protein
277 residues 118, 126, 136, 161, 251, 280, 300, and 336; Fig 5), indicating that the host proteins were indeed
278 interacting with assemble capsids. Only six of these residues shown in blue are also accessible on the outer
279 surface of the immature particle, as residue 336 is obscured by other E proteins and residue 251 by prM
280 (Fig 7). The cross-linked lysine residues are distributed evenly across the surface of the E protein, showing
281 no preference for domains I, II or III (Fig 7).



282

283 **Fig 7: Visualisation of E protein lysine residues that form cross-links with host proteins.** A) Surface
284 representation of the TBEV virion (PDB accession: 5O6A) [27]. The three E proteins within each asymmetric
285 unit are shown in white, grey and dark grey. One E protein monomer is shown in colour (TBEV domain I is
286 shown in peach, domain II in yellow, domain III in lilac with the cross-linked lysines shown in blue. B)
287 Mapping of the cross-linked lysines on the structure of the E protein (PDB accession 5O6A). Lysines
288 detected with cross-links are shown in blue if located on the surface of both the mature and immature virus
289 and black for those only located on the surface of the mature virus; lysines without cross-links are shown in
290 red. TBEV domain I is shown in peach, domain II in yellow, domain III in light blue and domain IV in lilac.

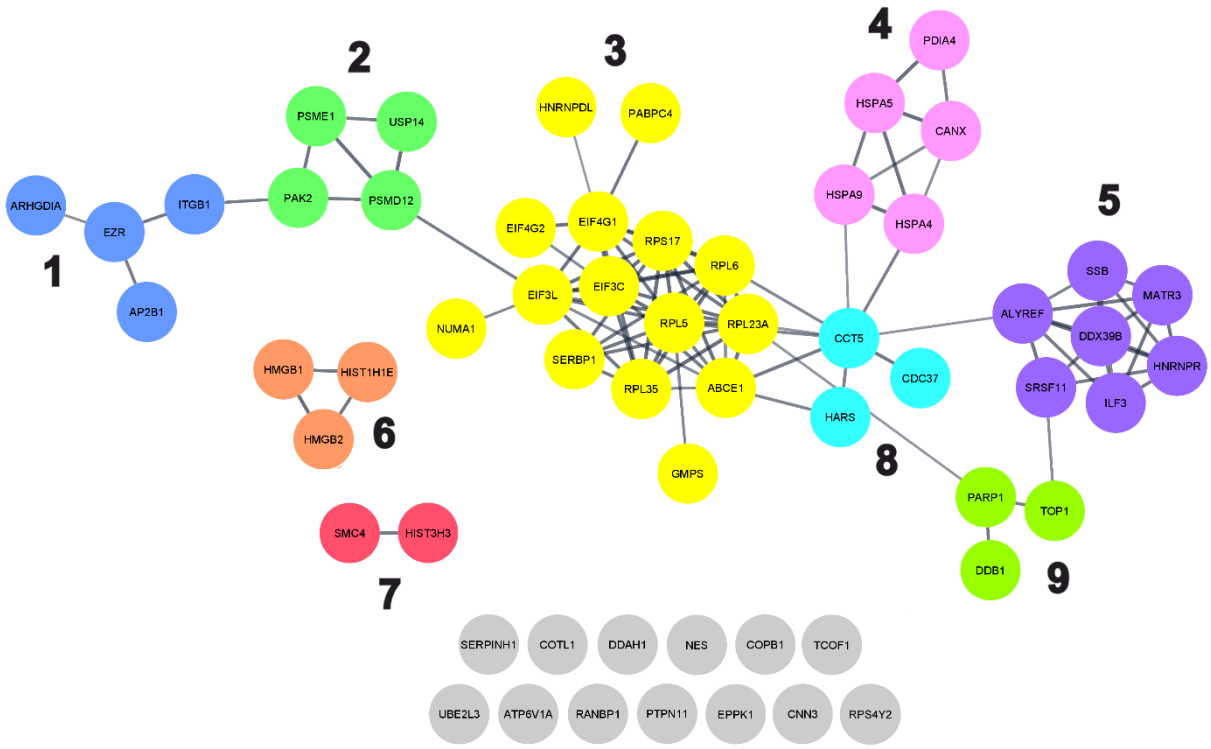


Fig 8: String network of 59 host proteins identified as interacting with the TBEV E protein in the filtered cross-linking dataset. The string-database network shows both direct and indirect protein interactions and is clustered based on the combined score associated with each interaction, using the Markov clustering algorithm [41]. Proteins are coloured based on the clustering and each cluster is numbered (1-9). Proteins that do not have any known interactions with other host proteins identified in the filtered cross-linking dataset are shown in grey.

We performed String-database analysis to identify if any of the 59 E protein-interacting host proteins also interact with each other (Fig 6) [41]. In total, 46 proteins were shown to interact with at least one other protein, and nine interaction clusters were identified. This suggests that TBEV may interact with both individual proteins and larger protein complexes. Gene ontology (GO) analysis (S4 Table) of the clusters indicates that proteins in clusters 1, 2 and 4 are located in the extracellular region (GO:0005576), at the plasma membrane (GO:0005886) or in extracellular exosomes (GO:0070062). Clusters 1 and 2 show enrichment for receptor-mediated biological processes, including receptor-mediated endocytosis (GO:0006898; cluster 1), and receptor-mediated signalling (GO:0038095, GO:0050852, GO:0002223; cluster

306 2), whereas proteins in cluster 4 are involved in protein transport (GO:0015031). Cluster 3 is the largest
307 cluster and consists of proteins found in the cytosol (GO:0005829) predominantly as part of ribosomes
308 (GO:0005840), ribonucleoprotein complexes (GO:1990904) or the eukaryotic translation initiation factor 4F
309 complex (GO:0016281). Other cytosolic proteins belong to cluster 8 (GO:0005829) and are involved in
310 protein folding (GO:0006457) or translation (GO:0006412). Finally, proteins in clusters 5, 6, 7 and 9 are
311 primarily found in the nucleus (GO:0005634) and are involved in RNA processing (GO:0006396),
312 posttranscriptional regulation of gene expression (GO:0010608; cluster 5), and chromosome organisation
313 (GO:0051276; clusters 6, 7 and 9). Importantly, many proteins primarily located in the cytosol or nucleus
314 (clusters 3, 5-9) are also found in the plasma membrane or extracellular regions where they perform
315 alternative biological functions; for example, HMGB1 (UniprotKB:P09429; cluster 6), functions as a
316 nonhistone nucleoprotein in the nucleus and an inflammatory cytokine in the extracellular region [42]. In
317 addition, GO analysis identified 19 proteins that are associated with immune system process, and 11
318 proteins associated with the cytoskeleton or cytoskeletal rearrangement.

319 # Discussion

320 In this study, we present a chemical cross-linking proteomics approach to simultaneously identify TBEV-
321 neuroblastoma cell PPIs and their interaction interfaces. We used metabolically-stalled cells to adsorb and
322 cross-link virus only to the cell surface, hoping to primarily enrich proteinaceous cell surface interactions.
323 The cross-linked proteins were released by limited proteolysis. Analysis of this highly complex protein
324 mixture by LC-MS/MS generated a large database of spectra containing four different peptide species with
325 the minority being cross-linked. In addition, cross-linked peptides are the least well-fragmented in the
326 database. In the next step, the pLink2 software compares the database of spectra with all of the possible
327 theoretical cross-linker reaction outcomes. This step is a clear bottle neck in the process as in our hands,
328 only a subset proteins could be analysed at a time, requiring multiple batch runs. Here, we optimised the
329 analysis workflow in order to extract the most significant virus-host PPIs from our complex data. Firstly, we
330 reduced the cross-linking search space by only analysing proteins identified by linear peptides in the

331 samples. Secondly, we have an internal validation control in the sample. We identified high-confidence
332 intraviral crosslinks using both the known and predicted three-dimensional structures of TBEV and the
333 known length of the chemical cross-linker. Then, we correlated the high confidence cross-links with the
334 SVM score and the quality of the spectra, allowing us to use an SVM score cut-off < 0.355 for the entire
335 dataset. Finally, we imposed a spectral count cut-off ≥ 2 to select for the most specific protein interactions.
336 Using this method, we identified 22 high confidence unique intraviral cross-links and 59 high confidence
337 unique virus-host PPIs between the surface of TBEV and human neuroblastoma cells. These proteins form a
338 robust and reliable dataset that can be investigated further for their roles in the virus life cycle. They could
339 be targets for intervention.

340 Our approach presents four major advantages over alternative approaches used to identify virus-host PPIs
341 described in the literature: 1) The wild-type virus interacts with cellular proteins on the surface of
342 neuroblastoma cells. In comparison, affinity purification, yeast two hybrid and protein microarrays detect
343 interactions in artificial systems [8–12]. Therefore, the expression levels, presentation and glycosylation
344 state of the host and viral proteins may differ from that in natural infections, leading to both false positives
345 and false negatives; 2) When fully assembled viral particles are used as bait, the capsid proteins are in the
346 right biological conformation and molecular context for infection. In contrast, single recombinant bait
347 proteins used in affinity purification, yeast two hybrid, protein microarrays and BioID may not be [8–12,16].
348 For instance, the monomeric Dengue virus ancillary receptor, DC-SIGN binds across two neighbouring E
349 proteins on the capsid [43]. In our virus-based protocol, there would be 90 such sites for the DC-SIGN
350 interaction giving both the correct biological context, but also increasing the avidity of the interactions; 3)
351 The virus interacts with cellular proteins prior to crosslinking. In contrast, in the previously described
352 approaches using trifunctional cross-linkers or BioID, the viral proteins are first conjugated to the cross-
353 linker or biotin ligase prior to the interaction with cells. Conjugation may require genetic modification [13–
354 17]. Consequently, interactions may be missed if the modification sterically hinders the binding region.
355 Here we could identify PPIs, under near-physiological conditions, without the need for genetic
356 manipulation; 4) The clearest advantage in our approach as compared to other methods mapping virus host

357 PPIs in culture is the use of DSS. Despite its limitations (see below), this allows us to directly map peptide-
358 level interaction interfaces. Other approaches lacking this level of information have been used for instance
359 for identifying host receptors to Sars-CoV using crosslinking followed by immunoprecipitation and LC-
360 MS/MS from SDS-PAGE bands, identifying vimentin as a critical protein for virus entry [18]; or using an
361 affinity-enrichable crosslinker to identify NCMA1 as a receptor for Zika virus [15].

362 Although this approach shows promise for detecting a wide-range of virus-host PPIs, there are also some
363 challenges. The complete coverage of the interaction space is limited by the accessibility of surface cross-
364 linkable lysine residues and the sample complexity. In order to detect an interaction, there must be cross-
365 linkable residues on both sides of the interaction interface within 30 Å of each other. Furthermore, in
366 complex systems containing a higher number of protein species, the number of cross-links per species is
367 lower in comparison to simpler systems. Consequently, interactions that occur in lysine deficient regions or
368 with low frequency cannot be detected, leading to an incomplete picture of the interaction interface or
369 failure to detect the PPI. Performing parallel experiments using chemical cross-linkers with different lengths
370 or reactive residues such as arginine, aspartate or glutamate, could overcome this limitation [44,45]. Simple
371 cross-linking experiments including only the virus and a single interaction partner can be used to ensure
372 complete coverage of the 3D interaction space once interesting PPI have been identified.

373 Having considered the potential advantages and challenges of this protocol, we will now consider potential
374 biological implications. Laminin binding protein has previously been suggested as a TBEV receptor. It was
375 present in our dataset, but no cross-links were identified to TBEV [46,47]. Our data do not support that
376 laminin binding protein is a TBEV receptor in this cell line. However, we have identified proteins that are
377 associated with the early stages of viral infection in other viruses, including ITGB1, ATP6V1A, EZR, HSPA9,
378 and HSPA5. ITGB1 as an entry receptor for a large number of viruses including, cytomegalovirus, Epstein-
379 Barr virus, human parvovirus B19, and mammalian reovirus [48–51]. ATP6V1A directly interacts with rabies
380 viral matrix protein facilitating uncoating [52]. EZR is an essential host factor required for the entry of

381 Japanese encephalitis virus into human brain microvascular endothelial cells [53]. HSPA9 has been
382 identified as a putative receptor for Tembusu virus [54].

383 HSPA5 (Cluster 4, Fig 6) has been identified as a receptor for several flaviviruses including, Zika and
384 Japanese encephalitis where HSPA5 was shown to affinity purify with recombinant E protein domain III, but
385 the interaction context in virions has not been studied [56,57]. HSPA5 is a multifunctional regulator of
386 endoplasmic reticulum homeostasis, playing an important role in protein processing and quality control
387 [59–62]. It is located both in the endoplasmic reticulum lumen and on the outer surface of the plasma
388 membrane in many cell types including neurons [63–66]. HSPA5 consists of 2 domains, a nucleotide-binding
389 domain that binds ATP, and a substrate-binding domain that binds and stabilises partially folded or folded
390 proteins [67]. Here we identified four spectral observations mapping to the TBEV E hinge region (residue
391 136) and the HSPA5 substrate binding domain residues 521 and 516 (Fig 9). Therefore, the interaction
392 between HSPA5 and the E protein hinge region could constitute a unique binding interface, or be part of a
393 larger interface that also binds domain III consistent with other flavivirus studies [54–56,58]. We
394 hypothesize that HSPA5 could interact with both the hinge region of one E monomer and domain III of an
395 adjacent E monomer at the 3-fold axis, where the regions are in close spatial proximity (Fig 9).
396 Interestingly, Fab fragments of the TBEV neutralising antibody 19/1786 have been shown to bind across
397 this interface at the 3-fold axis potentially preventing the HSPA5 interaction[27]. Although, no cross-links
398 were detected between domain III and HSPA5 in our study, this can be partly explained by poor lysine
399 availability. Structural bioinformatics studies of the Zika virus domain III-HSPA5 interaction do not detect
400 any interacting lysine residues on HSPA5[68].

401

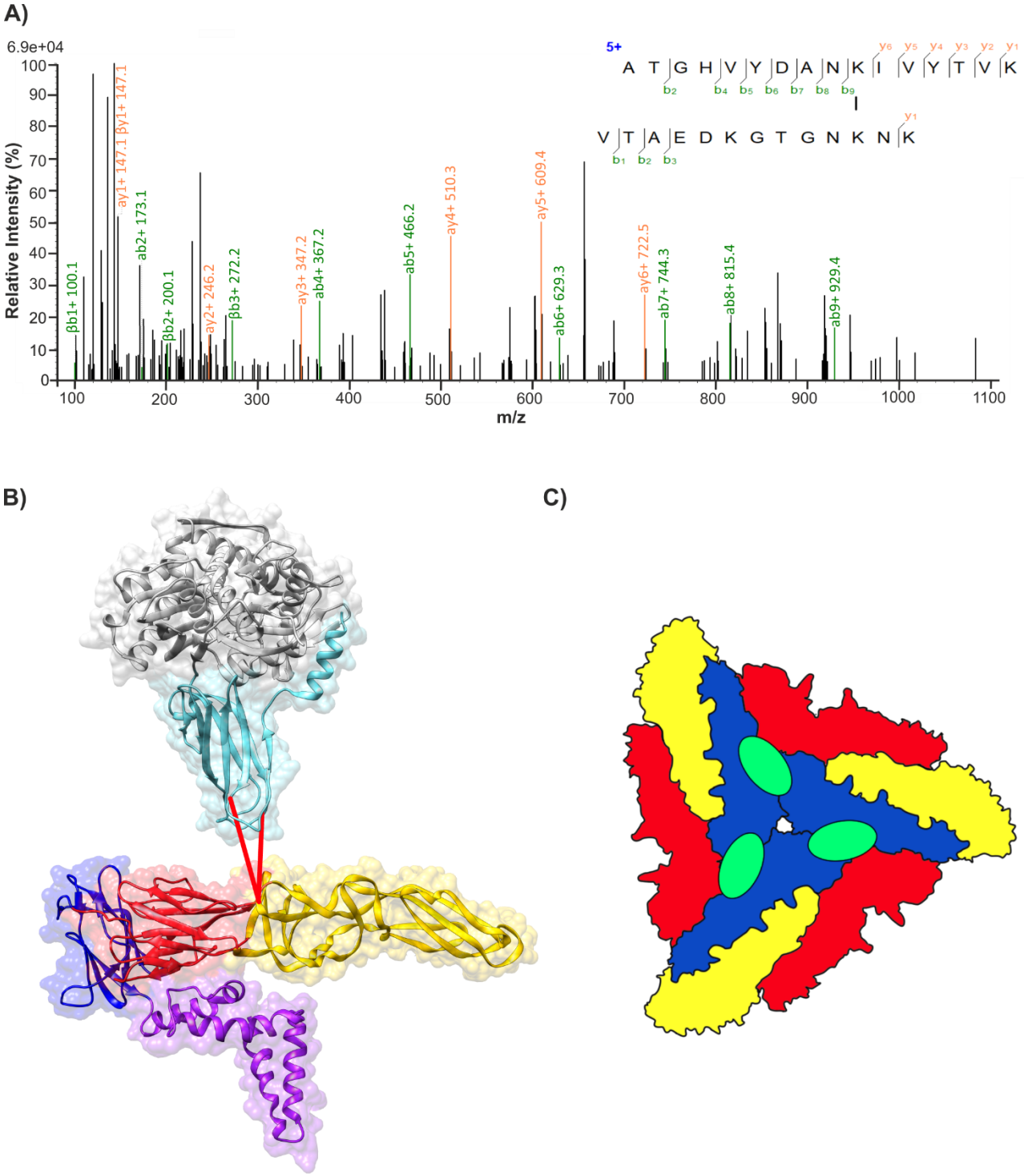


Fig 9: Cross-linking of HSPA5 and TBEV E protein A) Assigned spectrum for the peptide associated with the HSPA5-E protein interaction. B) TBEV E protein (PDB accession 5O6A) and HSPA5 (PBD accession 6ZMD) were placed in close proximity to allow for the visualisation of the cross-links, indicated here with red lines. TBEV E protein domain I is shown in red, domain III in dark blue, domain II in yellow and domain 4 in purple. HSPA5 substrate-binding domain is shown in light blue and the remainder of the protein in grey. C)

408 Schematic representation of speculated HSPA5 binding region on the mature virus. The three E proteins
409 within each asymmetric unit are shown in blue, red, and yellow. The proposed HSPA5 binding site is shown
410 in green.

411 After the virus has attached to the cell surface, the next step is to enter the cells either through clathrin-
412 mediated endocytosis or macropinocytosis, and we found evidence for both pathways being used. We
413 identified AP2B1 in the GO analysis that binds to the clathrin heavy chain [69]. We also identified an
414 abundance of cytoskeletal proteins and cytoskeletal remodelling proteins required in macropinocytosis
415 including, NES, NUMA1, ITGB1, EZR, PAK2, COTL1, CCT5, CNN3, EPPK1 and RANBP1. This supports the
416 hypothesis that macropinocytosis is used by TBEV as well as Dengue virus [70].

417 Conclusions

418 In this study, we present a XL-MS method to identify and map transient virus-host PPIs under near-
419 physiological conditions, without the need for genetic modification. Using this method, we identified 59
420 high confidence virus-host PPIs between TBEV and the surface of neuroblastoma cells. These proteins form
421 a robust and reliable dataset that can be investigated further for functional relevance in targeted follow-up
422 experiments. The presented methodologies are generally applicable to other virus-host systems and can
423 assist in expanding our knowledge of viral infections.

424 Materials and methods

425 Viruses and cells

426 Neuroblastoma SK-N-SH cells (ECACC 86012802) were maintained in Dulbecco's modified Eagle's medium
427 (DMEM) (Sigma) supplemented with 10 % heat inactivated FBS (Gibco), 100 µg/ml of penicillin-
428 streptomycin mix (PenStrep) (Sigma) and 2 mM L-glutamine (Sigma), at 37 °C and under a 5 % CO_2
429 atmosphere. For virus propagation, cells were infected with TBEV strain, MG569938 Kuutsalo-14 *Ixodes*
430 *ricinus* Finland-2017, at a MOI of 0.003, in Dulbecco's modified Eagle's medium (DMEM) (Sigma)

431 supplemented with 2 % heat inactivated FBS (Gibco), 100 µg/ml of penicillin-streptomycin mix (PenStrep)
432 (Sigma), 2 mM L-glutamine (Sigma) and 0.35 µM rapamycin and incubated at 37 °C for 3 days under a 5 %
433 CO_2 atmosphere. Virus-containing supernatant was aspirated and centrifuged at 4500 x g for 5 minutes to
434 remove cell debris. Cleared supernatant was aliquoted into cryopreservation vials and frozen at -80 °C. Viral
435 titers were determined using plaque-forming assay. SK-N-SH cells were grown on a 6-well plate, infected
436 with ten-fold serial dilutions of the virus and incubated for 1 hour at 37 °C, under a 5 % CO_2 atmosphere.
437 Overlay medium (Minimum Essential Medium Eagle, PenStrep and 2 mM L-glutamine, 1.2 % avicel) was
438 added to each well and the cells incubated for 4 days at 37 °C, under a 5 % CO_2 atmosphere. After 4 days
439 the cells were fixed with 10 % formaldehyde, stained with crystal violet and the plaques counted to
440 determine the number of plaque forming units per ml (PFU/ml).

441 Production of amino acid free virus stock

442 SK-N-SH cells were grown to 90% confluence, washed twice with PBS and the medium changed to amino
443 acid free DMEM (Genaxxon) supplemented with 100 µg/ml of penicillin-streptomycin mix (PenStrep)
444 (Sigma) and rapamycin 0.35 µM (selleckchem). Cells were infected with TBEV, at multiplicity of infection 1,
445 and incubated at 37 °C for 4 Days, under a 5 % CO_2 atmosphere. Virus-containing supernatant was
446 aspirated and centrifuged at 4500 x g for 5 minutes to remove cell debris. The virus was pelleted through a
447 30 % sucrose cushion in HNE buffer (20 mM HEPES pH 8.5, 150 mM NaCl, 1 mM EDTA), 2 h, 27000 rpm, 4
448 °C. The virus pellet was then resuspended in amino acid free DMEM (Genaxxon) overnight, 4 °C, with mild
449 shaking. The virus was aliquoted into cryopreservation vials and frozen at -80 °C. The stock was titered as
450 described above. The typical obtained titer was 5 x 10^9 pfu/ml.

451 Cross-linking of TBEV with SK-N-SH cells

452 SK-N-SH cells were grown in a 6 well plate to 90 % confluence, washed twice with PBS, and the medium
453 changed to amino acid free DMEM (Genaxxon) supplemented with 100 µg/ml of penicillin-streptomycin
454 mix (PenStrep) (Sigma). The cells were incubated overnight at 37 °C, under a 5 % CO_2 atmosphere. The cells
455 were infected with TBEV at a MOI of 375 for 60 min, with rocking, on ice. Heavy/light disuccinimidyl

456 suberate cross-linker (DSS-H12/D12, Creative Molecules Inc., www. creativemolecules.com) resuspended in
457 dimethylformamide (DMF) was added to final concentrations of 0, 100, 250, 500 and 1000 µM and
458 incubated for 60 min, with rocking on ice. The cross-linking reaction was quenched with a final
459 concentration of 50 mM ammonium bicarbonate with rocking, on ice. The SK-N-SH cell surface proteins
460 with attached TBEV virions were digested off with 1.25 µg trypsin (Promega) and the supernatant collected.
461 Finally, cell debris was removed via centrifugation (16,000 x g, 5 min), the supernatant recovered, and the
462 samples prepared for mass spectrometry (Fig 2).

463 Immunoblot analysis

464 Proteins were resolved in 4-20 % SDS-PAGE, transferred onto a nitrocellulose membrane, and probed using
465 anti-Langat E protein (BEI NR-40318; 1:1000 dilution) and C protein (57) (1:1000 dilution) antibodies in 5 %
466 milk, tris-buffered saline 0.1 % tween-20 (TBST) [71]. The protein bands were visualised using IR800 and
467 IR680-conjugated secondary anti-rabbit (Li-COR, 926-68071) and anti-mouse antibodies (KPL, 072071806)
468 diluted 1:10,000 in tris-buffered saline 0.1 % tween-20 (TBST). The membrane was imaged using the
469 Odyssey infrared imaging system (Li-COR).

470 Preparation of cross-linked samples for mass spectrometry

471 Samples from cross-linking were first denatured with 8 M urea-100 mM ammonium bicarbonate. The
472 cysteine bonds were then reduced with 5 mM tris(2-carboxyethyl) phosphine (37 °C, 60 min, 400 rpm) and
473 alkylated with 10 mM 2-iodoacetamide (22 °C, 30 min, in the dark). Protein digestion was then performed
474 with 0.1 µg/µl sequencing-grade lysyl endopeptidase (Wako chemicals) (37 °C, 2h, 400 rpm). Following the
475 dilution of the sample with 100 mM ammonium bicarbonate to a final urea concentration of 800 mM the
476 proteins were digested further with 0.2 µg/µl trypsin (Promega) (37 °C, 18h, 400 rpm). Digested samples
477 were then acidified with 10% formic acid to a pH of 3.0, and the peptides were subsequently purified with
478 C18 reverse-phase spin columns according to the manufactures protocol (Microspin Column, SS18V, The
479 Nest Group, Inc). Peptides were then dried in a speedvac and reconstituted in 2% acetonitrile, 0.2% formic
480 acid prior to mass spectrometric analyses

481 # Liquid chromatography tandem mass spectrometry

482 All peptide analyses were performed on a Q Exactive HFX mass spectrometer (Thermo Scientific) connected
483 to an EASY-nLC 1200 ultra-high-performance liquid chromatography system (Thermo Scientific). The
484 peptides were loaded onto an Acclaim PepMap 100 (ID 75µm x 2 cm, 3 µm, 100 Å) pre-column and
485 separated on an EASY-Spray column (Thermo Scientific; ID 75 µm × 25 cm, column temperature 45 °C)
486 operated at a constant pressure of 800 bar. A linear gradient from 4% to 45% of 0.1% formic acid in 80%
487 acetonitrile was run for 50 min at a flow rate of 300 nl/min One full MS scan (resolution 60,000@200 m/z;
488 mass range 350 to 1600 m/z) was followed by MS/MS scans (resolution 15,000@200 m/z) of the 15 most
489 abundant ion signals. The precursor ions were isolated with 2 m/z isolation width and fragmented using
490 higher-energy collisional-induced dissociation at a normalized collision energy of 30. Charge state screening
491 was enabled, and precursors with an unknown charge state and singly charged ions were excluded. The
492 dynamic exclusion window was set to 15 s and limited to 300 entries. The automatic gain control was set to
493 3×10^6 for MS and 1×10^5 for MS/MS with ion accumulation times of 110 and 60 ms, respectively. The
494 intensity threshold for precursor ion selection was set to 1.7×10^4.

495 # MS data analysis

496 Raw DDA data was converted to gzipped and Numpressed mzML [72] using MSconvert from the
497 ProteoWizard, v3.0.5930 suite [73]. All data was managed and analysed using openBIS [74]. The acquired
498 spectra were analysed using the search engine X! Tandem (2013.06.15.1-LabKey, Insilicos, ISB) [75], OMSSA
499 (version 2.1.8) [76] and COMET (version 2014.02 rev.2) [77] against an in-house compiled database
500 containing the reviewed *Homo sapiens* reference proteome (UniProt proteome ID UP000005640) and the
501 TBEV proteome (GenBank accession: AWC08512.1), yielding a total of 78121 protein entries and an equal
502 amount of reverse decoy sequences. Full tryptic digestion was used allowing two missed cleavages.
503 Carbamidomethylation (C) was set to static and oxidation (M) to variable modifications, respectively. Mass
504 tolerance for precursor ions was set to 0.2 Da, and for fragment ions to 0.02 Da. Identified peptides were
505 processed and analysed through the Trans-Proteomic Pipeline (TPP v4.7 POLAR VORTEX rev 0, Build

506 201403121010) using PeptideProphet [78]. The false discovery rate (FDR) was estimated with Mayu
507 (v1.7)[79] and peptide spectrum matches (PSMs) were filtered with protein FDR set to 1% resulting in a
508 peptide FDR >1%. Proteins were filtered to remove hits identified by only 1 unique peptide, and an average
509 spectral count of < 2 across all samples. The protein names and corresponding UniProt ID are given in
510 Supplementary information 1.

511 Cross-link identification

512 In total 874 proteins that were identified by 2 or more unique peptides, and an average spectral count of
513 ≥2 across all samples in the DDA data were probed for cross-links. Cross-links between the TBEV structural
514 proteins and host proteins were identified using the pLink2 software package [32]. In order to reduce the
515 search space for cross-link identification the data was analysed in 35 batches. The raw data dependent
516 acquisition data, and compiled FASTA file databases containing a total of 28 protein sequences, from 25
517 different host proteins and the 3 viral structural protein were used as the software input. Host protein
518 sequences were obtained from UniProt. The TBEV polyprotein sequence was obtained from GenBank
519 (accession: AWC08512.1), the C protein corresponds to residues 1-96, the M protein 206-280 and the E
520 protein 281-776. The following search parameters were used in the pLink2 software: Conventional cross-
521 linking (Higher-energy C-trap dissociation (HCD)), precursor mass tolerance of 20 ppm; fragment mass
522 tolerance of 20 ppm; peptide length of 6-60; peptide mass of 350-6000 Da, up to 3 missed cleavage sites;
523 carbamidomethylation (C) was set to static modification; and oxidation (M) to variable modification. The
524 results were then filtered using a filtering tolerance of ±10 ppm and a separate FDR >5% at the peptide
525 spectrum matches level. 7716 cross-linked spectral observations were observed across all samples, 302 of
526 these were observed in the negative control samples. In total, 247 cross-linked spectral observations in the
527 0.1-1mM DSS samples corresponded to cross-linked interfaces also identified in negative control samples
528 and were excluded from further analysis.

529 ## Homology modelling, structure visualization and measuring cross-link
530 distances

531 Homology models for the TBEV C protein, immature conformation of the E protein and the prM protein
532 were generated using the I-TASSER (Iterative Threading ASSEmbly Refinement) server [35–37]. A C-score
533 (confidence score for estimating the quality of predicted models by I-TASSER) is generated for each model
534 and can range from -5 to 2, with a higher value signifies a higher confidence and where a C-score > -1.5 is
535 considered good [35–37]. The C-protein homology model was generated using PDB accession 5OW2, and
536 the immature conformation of the E protein, and the prM protein using PDB accession 7L30 as the
537 template. The TBEV polyprotein sequence was obtained from GenBank (accession: AWC08512.1). The
538 sequence for the C protein was obtained from residues 1-96, the E protein residue 281-776, and the prM
539 residues 113-280 in the polyprotein. Generation of the assembled immature virus homology model, and
540 the C-protein dimer was performed in UCSF Chimera [80]. The homology model for the immature virus was
541 generated by superimposing the models for the immature E protein and prM protein onto the assembled
542 immature Spondweni virus (PDB accession 6ZQW) structure, using the MatchMaker function. The
543 homology model for the C protein dimer was generated by superimposing two models of the C protein
544 onto the Zika C protein dimer (PDB accession 5YGH), using MatchMaker. The atomic models were used to
545 position both intraprotein and interprotein cross-links by choosing the distance between Cα atoms in
546 Chimera [25,27].

547 ## Networking and clustering analysis

548 String data were obtained from string database [41] and imported into Cytoscape 3.4 [81] .The following
549 interaction sources were considered: experiments, databases, co-expression, co-occurrence and gene
550 fusion, and interactions with a minimum interaction score of 0.7 are shown. Clusters were generated using
551 the Markov clustering algorithm and gene ontology annotations for each cluster were obtained using
552 GOnet [82].

553 ## Data Availability Statement

554 The datasets generated during and/or analysed in the current study are available in the repositories with
555 the persistent web links: ftp://massive.ucsd.edu/MSV000088272/ [76].

556 The mass spectrometry data has been deposited to the ProteomeXchange consortium via the MassIVE
557 partner repository https://massive.ucsd.edu/ with the dataset identifier PXD029384 [76].

558 # Acknowledgments

559 We thank Markku Varjosalo and Johan Malmström for helpful discussions. The work was carried out in the
560 Instruct Centre Finland supported by grants to SJB from the Swedish Research Council reference number
561 2018-05851; the Academy of Finland grant number 336471, the Sigrid Juselius Foundation grant number
562 95-7202-38. This project has also received funding from the European Union's Horizon 2020 Research and
563 Innovation Programme under the Marie Skłodowska-Curie grant agreements number 765042 (ViBRANT)
564 and number 799929 (MA). SVB is a fellow of the Doctoral Programme in Integrative Life Science, University
565 of Helsinki. LIAP is a fellow of the Doctoral Programme in Microbiology and Biotechnology, University of
566 Helsinki. We gratefully acknowledge support from the Swedish National Infrastructure for Biological Mass
567 Spectrometry.

568 # Conflict of interests

569 The authors declare that they have no conflict of interest.

570 # References

571 1. Kumar N, Sharma S, Kumar R, Tripathi BN, Barua S, Ly H, et al. Host-Directed Antiviral Therapy. Clin
572 Microbiol Rev. 2020 May 13;33(3):e00168-19, /cmr/33/3/CMR.00168-19.atom.

573 2. Anasir MI, Zarif F, Poh CL. Antivirals blocking entry of enteroviruses and therapeutic potential. J
574 Biomed Sci. 2021 Dec;28(1):10.

575 3. Plavec Z, Pöhner I, Poso A, Butcher SJ. Virus structure and structure-based antivirals. Curr Opin
576 Virol. 2021 Dec;51:16–24.

577 4. Tahir ul Qamar M, Shokat Z, Muneer I, Ashfaq UA, Javed H, Anwar F, et al. Multiepitope-Based
578 Subunit Vaccine Design and Evaluation against Respiratory Syncytial Virus Using Reverse Vaccinology
579 Approach. Vaccines. 2020 Jun 8;8(2):288.

580 5. Wang N, Shang J, Jiang S, Du L. Subunit Vaccines Against Emerging Pathogenic Human
581 Coronaviruses. Front Microbiol. 2020 Feb 28;11:298.

582 6. Roulin PS, Murer LP, Greber UF. A Single Point Mutation in the Rhinovirus 2B Protein Reduces the
583 Requirement for Phosphatidylinositol 4-Kinase Class III Beta in Viral Replication. Pfeiffer JK, editor. J Virol.
584 2018 Sep 12;92(23):e01462-18, /jvi/92/23/e01462-18.atom.

585 7. Barrass SV, Butcher SJ. Advances in high-throughput methods for the identification of virus
586 receptors. Med Microbiol Immunol (Berl). 2020 Jun;209(3):309–23.

587 8. Mukherjee S, Sengupta N, Chaudhuri A, Akbar I, Singh N, Chakraborty S, et al. PLVAP and GKN3 Are
588 Two Critical Host Cell Receptors Which Facilitate Japanese Encephalitis Virus Entry Into Neurons. Sci Rep.
589 2018 Dec;8(1):11784.

590 9. Shapira SD, Gat-Viks I, Shum BOV, Dricot A, de Grace MM, Wu L, et al. A Physical and Regulatory
591 Map of Host-Influenza Interactions Reveals Pathways in H1N1 Infection. Cell. 2009 Dec;139(7):1255–67.

592 10. Jäger S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, et al. Global landscape of
593 HIV–human protein complexes. Nature. 2012 Jan 19;481(7381):365–70.

594 11. Martinez-Martin N, Marcandalli J, Huang CS, Arthur CP, Perotti M, Foglierini M, et al. An Unbiased
595 Screen for Human Cytomegalovirus Identifies Neuropilin-2 as a Central Viral Receptor. Cell. 2018
596 Aug;174(5):1158-1171.e19.

597 12. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein
598 interaction map reveals targets for drug repurposing. Nature. 2020 Jul 16;583(7816):459–68.

599 13. Frei AP, Jeon O-Y, Kilcher S, Moest H, Henning LM, Jost C, et al. Direct identification of ligand-
600 receptor interactions on living cells and tissues. Nat Biotechnol. 2012 Oct;30(10):997–1001.

601 14. Sobotzki N, Schafroth MA, Rudnicka A, Koetemann A, Marty F, Goetze S, et al. HATRIC-based
602 identification of receptors for orphan ligands. Nat Commun. 2018 Dec;9(1):1519.

603 15. Srivastava M, Zhang Y, Chen J, Sirohi D, Miller A, Zhang Y, et al. Chemical proteomics tracks virus
604 entry and uncovers NCAM1 as Zika virus receptor. Nat Commun. 2020 Dec;11(1):3896.

605 16. Terracciano R, Preianò M, Fregola A, Pelaia C, Montalcini T, Savino R. Mapping the SARS-CoV-2–
606 Host Protein–Protein Interactome by Affinity Purification Mass Spectrometry and Proximity-Dependent
607 Biotin Labeling: A Rational and Straightforward Route to Discover Host-Directed Anti-SARS-CoV-2
608 Therapeutics. Int J Mol Sci. 2021 Jan 7;22(2):532.

609 17. Coyaud E, Ranadheera C, Cheng D, Gonçalves J, Dyakov BJA, Laurent EMN, et al. Global
610 Interactomics Uncovers Extensive Organellar Targeting by Zika Virus. Mol Cell Proteomics. 2018
611 Nov;17(11):2242–55.

612 18. Yu YT-C, Chien S-C, Chen I-Y, Lai C-T, Tsay Y-G, Chang SC, et al. Surface vimentin is critical for the cell
613 entry of SARS-CoV. J Biomed Sci. 2016 Dec;23(1):14.

614 19. Hauri S, Khakzad H, Happonen L, Teleman J, Malmström J, Malmström L. Rapid determination of
615 quaternary protein structures in complex biological samples. Nat Commun. 2019 Dec;10(1):192.

616 20. Chowdhury S, Khakzad H, Bergdahl GE, Lood R, Ekstrom S, Linke D, et al. Streptococcus pyogenes
617 Forms Serotype- and Local Environment-Dependent Interspecies Protein Complexes. mSystems. 2021 Oct
618 26;6(5):e0027121.

619 21. Kupča AM, Essbauer S, Zoeller G, de Mendonça PG, Brey R, Rinder M, et al. Isolation and molecular
620 characterization of a tick-borne encephalitis virus strain from a new tick-borne encephalitis focus with
621 severe cases in Bavaria, Germany. Ticks Tick-Borne Dis. 2010 Mar;1(1):44–51.

622 22. Han X, Juceviciene A, Uzcategui NY, Brummer-Korvenkontio H, Zygutiene M, Jääskeläinen A, et al.
623 Molecular epidemiology of tick-borne encephalitis virus inIxodes ricinus ticks in Lithuania. J Med Virol. 2005
624 Oct;77(2):249–56.

625 23. Bogovic P. Tick-borne encephalitis: A review of epidemiology, clinical characteristics, and
626 management. World J Clin Cases. 2015;3(5):430.

627 24. Taba P, Schmutzhard E, Forsberg P, Lutsar I, Ljøstad U, Mygland Å, et al. EAN consensus review on
628 prevention, diagnosis and management of tick-borne encephalitis. Eur J Neurol. 2017 Oct;24(10):1214-e61.

629 25. Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC. The envelope glycoprotein from tick-borne
630 encephalitis virus at 2 Å resolution. Nature. 1995 May;375(6529):291–8.

631 26. Mandl CW, Allison SL, Holzmann H, Meixner T, Heinz FX. Attenuation of Tick-Borne Encephalitis
632 Virus by Structure-Based Site-Specific Mutagenesis of a Putative Flavivirus Receptor Binding Site. J Virol.
633 2000 Oct 15;74(20):9601–9.

634 27. Füzik T, Formanová P, Růžek D, Yoshii K, Niedrig M, Plevka P. Structure of tick-borne encephalitis
635 virus and its neutralization by a monoclonal antibody. Nat Commun. 2018 Dec;9(1):436.

636 28. Renner M, Dejnirattisai W, Carrique L, Martin IS, Karia D, Ilca SL, et al. Flavivirus maturation leads to
637 the formation of an occupied lipid pocket in the surface glycoproteins. Nat Commun. 2021 Dec;12(1):1238.

638 29. Ma L, Jones CT, Groesch TD, Kuhn RJ, Post CB. Solution structure of dengue virus capsid protein
639 reveals another fold. Proc Natl Acad Sci. 2004 Mar 9;101(10):3414–9.

640 30. Dokland T, Walsh M, Mackenzie JM, Khromykh AA, Ee K-H, Wang S. West Nile Virus Core Protein.
641 Structure. 2004 Jul;12(7):1157–63.

642 31. Shang Z, Song H, Shi Y, Qi J, Gao GF. Crystal Structure of the Capsid Protein from Zika Virus. J Mol
643 Biol. 2018 Mar;430(7):948–62.

644 32. Chen Z-L, Meng J-M, Cao Y, Yin J-L, Fang R-Q, Fan S-B, et al. A high-speed search engine pLink 2 with
645 systematic evaluation for proteome-scale identification of cross-linked peptides. Nat Commun. 2019
646 Dec;10(1):3404.

647 33. Zybailov BL. Large Scale Chemical Cross-linking Mass Spectrometry Perspectives. J Proteomics
648 Bioinform [Internet]. 2013 [cited 2021 Jun 10];01(S2). Available from: https://www.omicsonline.org/large-
649 scale-chemical-cross-linking-mass-spectrometry-perspectives-jpb.S2-001.php?aid=12235

650 34. Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, et al. Identification of cross-
651 linked peptides from large sequence databases. Nat Methods. 2008 Apr;5(4):315–8.

652 35. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and
653 function prediction. Nat Protoc. 2010 Apr;5(4):725–38.

654 36. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function
655 prediction. Nat Methods. 2015 Jan;12(1):7–8.

656 37. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions.
657 Nucleic Acids Res. 2015 Jul 1;43(W1):W174–81.

658 38. Merkley ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN. Distance restraints from
659 crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine-lysine
660 distances: Evaluating Lysine-Lysine Distances by MD for XL-MS. Protein Sci. 2014 Jun;23(6):747–59.

661 39. Yang B, Wu Y-J, Zhu M, Fan S-B, Lin J, Zhang K, et al. Identification of cross-linked peptides from
662 complex samples. Nat Methods. 2012 Sep;9(9):904–6.

663 40. Debelyy MO, Waridel P, Quadroni M, Schneiter R, Conzelmann A. Chemical crosslinking and mass
664 spectrometry to elucidate the topology of integral membrane proteins. Gasset M, editor. PLOS ONE. 2017
665 Oct 26;12(10):e0186840.

666 41. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–
667 protein association networks with increased coverage, supporting functional discovery in genome-wide
668 experimental datasets. Nucleic Acids Res. 2019 Jan 8;47(D1):D607–13.

669 42. Yuan S, Liu Z, Xu Z, Liu J, Zhang J. High mobility group box 1 (HMGB1): a pivotal regulator of
670 hematopoietic malignancies. J Hematol OncolJ Hematol Oncol. 2020 Dec;13(1):91.

671 43. Pokidysheva E, Zhang Y, Battisti AJ, Bator-Kelly CM, Chipman PR, Xiao C, et al. Cryo-EM
672 Reconstruction of Dengue Virus in Complex with the Carbohydrate Recognition Domain of DC-SIGN. Cell.
673 2006 Feb;124(3):485–93.

674 44. Jones AX, Cao Y, Tang Y-L, Wang J-H, Ding Y-H, Tan H, et al. Improving mass spectrometry analysis
675 of protein structures with arginine-selective chemical cross-linkers. Nat Commun. 2019 Dec;10(1):3911.

676 45. Leitner A, Joachimiak LA, Unverdorben P, Walzthoeni T, Frydman J, Förster F, et al. Chemical cross-
677 linking/mass spectrometry targeting acidic residues in proteins and protein complexes. Proc Natl Acad Sci.
678 2014 Jul 1;111(26):9455–60.

679 46. Malygin AA, Bondarenko EI, Ivanisenko VA, Protopopova EV, Karpova GG, Loktev VB. C-terminal
680 fragment of human laminin-binding protein contains a receptor domain for Venezuelan equine encephalitis
681 and tick-borne encephalitis viruses. Biochem Mosc. 2009 Dec;74(12):1328–36.

682 47. Protopopova EV, Sorokin AV, Konovalova SN, Kachko AV, Netesov SV, Loktev VB. Human laminin
683 binding protein as a cell receptor for the tick-borne encephalitis virus. Zentralblatt Für Bakteriol. 1999
684 Dec;289(5–7):632–8.

685 48. Maginnis MS, Forrest JC, Kopecky-Bromberg SA, Dickeson SK, Santoro SA, Zutter MM, et al. β1
686 Integrin Mediates Internalization of Mammalian Reovirus. J Virol. 2006 Mar 15;80(6):2760–70.

687 49. Feire AL, Roy RM, Manley K, Compton T. The Glycoprotein B Disintegrin-Like Domain Binds Beta 1
688 Integrin To Mediate Cytomegalovirus Entry. J Virol. 2010 Oct;84(19):10026–37.

689 50. Xiao J, Palefsky JM, Herrera R, Berline J, Tugizov SM. The Epstein–Barr virus BMRF-2 protein
690 facilitates virus attachment to oral epithelial cells. Virology. 2008 Jan;370(2):430–42.

691 51. Weigel-Kelley KA, Yoder MC, Srivastava A. α5β1 integrin as a cellular coreceptor for human
692 parvovirus B19: requirement of functional activation of β1 integrin for viral entry. Blood. 2003 Dec
693 1;102(12):3927–33.

694 52. Liu X, Li F, Zhang J, Wang L, Wang J, Wen Z, et al. The ATPase ATP6V1A facilitates rabies virus
695 replication by promoting virion uncoating and interacting with the viral matrix protein. J Biol Chem. 2021
696 Jan;296:100096.

697 53. Liu Y-G, Chen Y, Wang X, Zhao P, Zhu Y, Qi Z. Ezrin is essential for the entry of Japanese encephalitis
698 virus into the human brain microvascular endothelial cells. Emerg Microbes Infect. 2020 Jan 1;9(1):1330–
699 41.

700 54. Liu Q, Huang X, Zhao D, Han K, Liu Y, Yang J, et al. Identification of heat shock protein A9 as a
701 Tembusu virus binding protein on DF-1 cells. Virus Res. 2017 Jan;227:110–4.

702 55. Jindadamrongwech S, Thepparit C, Smith DR. Identification of GRP 78 (BiP) as a liver cell expressed
703 receptor element for dengue virus serotype 2. Arch Virol. 2004 May;149(5):915–27.

704 56. Khongwichit S, Sornjai W, Jitobaom K, Greenwood M, Greenwood MP, Hitakarun A, et al. A
705 functional interaction between GRP78 and Zika virus E protein. Sci Rep. 2021 Jan 11;11(1):393.

706 57. Nain M, Mukherjee S, Karmakar SP, Paton AW, Paton JC, Abdin MZ, et al. GRP78 Is an Important
707 Host Factor for Japanese Encephalitis Virus Entry and Replication in Mammalian Cells. J Virol. 2017 Mar
708 15;91(6):e02274-16.

709 58. Zhao D, Liu Q, Han K, Wang H, Yang J, Bi K, et al. Identification of Glucose-Regulated Protein 78
710 (GRP78) as a Receptor in BHK-21 Cells for Duck Tembusu Virus Infection. Front Microbiol. 2018 Apr 9;9:694.

711 59. Dana RC, Welch WJ, Deftos LJ. Heat shock proteins bind calcitonin. Endocrinology. 1990
712 Jan;126(1):672–4.

713 60. Oka OBV, Pringle MA, Schopp IM, Braakman I, Bulleid NJ. ERdj5 is the ER reductase that catalyzes
714 the removal of non-native disulfides and correct folding of the LDL receptor. Mol Cell. 2013 Jun
715 27;50(6):793–804.

716 61. Evensen NA, Kuscu C, Nguyen H-L, Zarrabi K, Dufour A, Kadam P, et al. Unraveling the role of
717 KIAA1199, a novel endoplasmic reticulum protein, in cancer cell migration. J Natl Cancer Inst. 2013 Sep
718 18;105(18):1402–16.

719 62. Cuevas EP, Eraso P, Mazón MJ, Santos V, Moreno-Bueno G, Cano A, et al. LOXL2 drives epithelial-
720 mesenchymal transition via activation of IRE1-XBP1 signalling pathway. Sci Rep. 2017 Mar 23;7:44988.

721 63. Bellani S, Mescola A, Ronzitti G, Tsushima H, Tilve S, Canale C, et al. GRP78 clustering at the cell
722 surface of neurons transduces the action of exogenous alpha-synuclein. Cell Death Differ. 2014
723 Dec;21(12):1971–83.

724 64. Bhattacharjee G, Ahamed J, Pedersen B, El-Sheikh A, Mackman N, Ruf W, et al. Regulation of tissue
725 factor--mediated initiation of the coagulation cascade by cell surface grp78. Arterioscler Thromb Vasc Biol.
726 2005 Aug;25(8):1737–43.

727 65. Philippova M, Ivanov D, Joshi MB, Kyriakakis E, Rupp K, Afonyushkin T, et al. Identification of
728 Proteins Associating with Glycosylphosphatidylinositol- Anchored T-Cadherin on the Surface of Vascular
729 Endothelial Cells: Role for Grp78/BiP in T-Cadherin-Dependent Cell Survival. Mol Cell Biol. 2008 Jun
730 15;28(12):4004–17.

731 66. Honda T, Horie M, Daito T, Ikuta K, Tomonaga K. Molecular Chaperone BiP Interacts with Borna
732 Disease Virus Glycoprotein at the Cell Surface. J Virol. 2009 Dec;83(23):12622–5.

733 67. Mayer MP, Gierasch LM. Recent advances in the structural and mechanistic aspects of Hsp70
734 molecular chaperones. J Biol Chem. 2019 Feb;294(6):2085–97.

735 68. Elfiky AA, Ibrahim IM. Zika virus envelope – heat shock protein A5 (GRP78) binding site prediction. J
736 Biomol Struct Dyn. 2021 Sep 22;39(14):5248–60.

737 69. van der Schaar HM, Rust MJ, Chen C, van der Ende-Metselaar H, Wilschut J, Zhuang X, et al.
738 Dissecting the Cell Entry Pathway of Dengue Virus by Single-Particle Tracking in Living Cells. Farzan M,
739 editor. PLoS Pathog. 2008 Dec 19;4(12):e1000244.

740 70. Suksanpaisan L, Susantad T, Smith DR. Characterization of dengue virus entry into HepG2 cells. J
741 Biomed Sci. 2009;16(1):17.

742 71. Iacono-Connors LC, Smith JF, Ksiazek TG, Kelley CL, Schmaljohn CS. Characterization of Langat virus
743 antigenic determinants defined by monoclonal antibodies to E, NS1 and preM and identification of a
744 protective, non-neutralizing preM-specific monoclonal antibody. Virus Res. 1996 Aug;43(2):125–36.

745 72. Teleman J, Dowsey AW, Gonzalez-Galarza FF, Perkins S, Pratt B, Röst HL, et al. Numerical
746 Compression Schemes for Proteomics Mass Spectrometry Data. Mol Cell Proteomics. 2014 Jun;13(6):1537–
747 42.

748 73. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform
749 toolkit for mass spectrometry and proteomics. Nat Biotechnol. 2012 Oct;30(10):918–20.

750 74. Bauch A, Adamczyk I, Buczek P, Elmer F-J, Enimanev K, Glyzewski P, et al. openBIS: a flexible
751 framework for managing and analyzing complex data in biology research. BMC Bioinformatics.
752 2011;12(1):468.

753 75. Craig R, Beavis RC. A method for reducing the time required to match protein sequences with
754 tandem mass spectra. Rapid Commun Mass Spectrom. 2003 Oct 30;17(20):2310–6.

755 76. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open Mass Spectrometry
756 Search Algorithm. J Proteome Res. 2004 Oct;3(5):958–64.

757 77. Eng JK, Jahan TA, Hoopmann MR. Comet: An open-source MS/MS sequence database search tool.
758 PROTEOMICS. 2013 Jan;13(1):22–4.

759 78. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical Statistical Model To Estimate the Accuracy
760 of Peptide Identifications Made by MS/MS and Database Search. Anal Chem. 2002 Oct 1;74(20):5383–92.

761 79. Reiter L, Claassen M, Schrimpf SP, Jovanovic M, Schmidt A, Buhmann JM, et al. Protein
762 Identification False Discovery Rates for Very Large Proteomics Data Sets Generated by Tandem Mass
763 Spectrometry. Mol Cell Proteomics. 2009 Nov;8(11):2405–17.

764 80. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera?A
765 visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605–12.

766 81. Shannon P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction
767 Networks. Genome Res. 2003 Nov 1;13(11):2498–504.

768 82. Pomaznoy M, Ha B, Peters B. GOnet: a tool for interactive Gene Ontology analysis. BMC
769 Bioinformatics. 2018 Dec;19(1):470.

770 # Supporting information captions

771 **S1 Table: DDA analysis of viral and host protein in all samples** (XLSX).

772 **S2 Table: Cross-linking dataset** Identified cross-linked peptides, corresponding proteins, E-values and SVM
773 scores for each cross-linked spectral observation (XLSX).

774 **S3 Table: Filtered cross-linking dataset** Cross-linking dataset (S2 Table) filtered to remove spectral
775 observations corresponding to unique cross-links identified by less than 2 spectral observations with SVM
776 scores < 0.355 (XLSX).

777 **S4 Table: GO analysis of host proteins** GO analysis of the 59 host proteins shown to interact with the
778 surface of TBEV in the filtered cross-linking dataset (S3 Table) (XLSX).

779 **S1 File: Correlation between number of cross-linked spectral observations, protein abundance and**
780 **protein lysine content** Table showing the number of cross-linked spectral observations, the average protein
781 abundance calculated in the DDA analysis (S1 Table) and the number of lysines for 101 randomly selected
782 proteins. Scatter graphs show no correlation between the number of cross-linked spectral observations and
783 the number of lysines for the proteins, and no correlation between the number of cross-linked spectral
784 observations and the protein abundance (XLSX).

785 **S1 Fig: Box and whisker plot of the E-value for the intraviral cross-links, plotted against the distance**
786 **between the cross-linked residues for each cross-link.** The smallest distance out of the mature and
787 immature calculated distances, and the C protein dimer or monomer distances is plotted. The mean E-value
788 is marked with a cross.