Crystal structures of glycoprotein D of equine alphaherpesviruses reveal potential binding sites to the entry receptor MHC-I

4   Viviane Kremling[1]*, Bernhard Loll[2], Szymon Pach[3], Ismail Dahmani[4], Christoph Weise[5],

5   Gerhard Wolber[3], Salvatore Chiantia[4], Markus C. Wahl[2,6], Nikolaus Osterrieder[1,7], Walid

6   Azab[1]*

7   [1]   Institut für Virologie, Robert von Ostertag-Haus, Zentrum für Infektionsmedizin, Freie

8   Universität Berlin, Berlin, Germany,

9   [2]   Laboratory of Structural Biochemistry, Freie Universität Berlin, Berlin, Germany,

10   [3]   Institute of Pharmacy (Pharmaceutical Chemistry), Freie Universität Berlin, Berlin,

11   Germany,

12   [4]   Universität Potsdam, Institut für Biochemie und Biologie, Potsdam, Brandenburg,

13   Germany,

14   [5]   BioSupraMol Core Facility, Bio-Mass Spectrometry, Freie Universität Berlin, Berlin,

15   Germany

16   [6]   Helmholtz-Zentrum Berlin für Materialien und Energie, Macromolecular

17   Crystallography, Berlin, Germany

18   [7]   Department of Infectious Diseases and Public Health, City University of Hong Kong,

19   Kowloon Tong, Hong Kong

20

21   *Corresponding authors:

22   E-mail: walid.azab@fu-berlin.de (WA)

23   E-mail: viviane.kremling@fu-berlin.de (VK)

Abstract

25   Cell entry of most alphaherpesviruses is mediated by the binding of glycoprotein D (gD) to

26   different cell surface receptors. Equine herpesvirus type 1 (EHV-1) and EHV-4 gDs interact

27   with equine major histocompatibility complex I (MHC-I) to initiate entry into equine cells.

28   We have characterized the gD-MHC-I interaction by solving the crystal structures of EHV-1

29   and EHV-4 gDs (gD1, gD4), performing protein-protein docking simulations, surface plasmon

30   resonance (SPR) analysis, and biological assays. The structures of gD1 and gD4 revealed the

31   existence of a common V-set immunoglobulin-like (IgV-like) core comparable to those of other

32   gD homologs. Molecular modeling yielded plausible binding hypotheses and identified key

33   residues (F213 and D261) that are important for virus binding. Altering the key residues

34   resulted in impaired virus growth in cells, which highlights the important role of these residues

35   in the gD-MHC-I interaction. Taken together, our results add to our understanding of the

36   initial herpesvirus-cell interactions and will contribute to the targeted design of antiviral

37   drugs and vaccine development.

38

Author summary

40   Equine herpesvirus type 1 (EHV-1) and type 4 (EHV-4) are endemic in horses and cause great

41   suffering as well as substantial economic losses to the equine industry. Current vaccines do

42   not prevent infections and treatment is difficult. A prerequisite for vaccine and drug

43   development is an in-depth understanding of the virus replication cycle, especially the virus

44   entry process in order to block the infection at early stages. Entry of alphaherpesviruses into

45   the host cell is mediated by a set of virus envelope glycoproteins including glycoprotein D

46   (gD) that triggers the internalization of the virus particle. The structure of gD and the

47   interaction with the entry receptor equine major histocompatibility complex class I (MHC-I)

48   remains elusive. Here, we solved the crystal structures of gD1 and gD4 that allowed us to

3

49 model virus-receptor interaction and to determine the key residues for virus entry. Alterations

50 of these key residues impaired virus growth in cell culture. The overall structure of gD1 and

51 gD4 shows classical features of other alphaherpesvirus gDs making it possible to gain further

52 insights into human pathogens as well.

53

54 Introduction

55 One of the most essential steps for virus replication is the entry process into host cells. In

56 herpesviruses, more specifically alphaherpesviruses, cell entry is a complex multistep process

57 that requires a stepwise contribution of five out of twelve envelope glycoproteins, namely

58 glycoprotein B (gB), gC, gD, and the heterodimer gH/gL [1]. Of these, gD is the (main)

59 receptor-binding protein that interacts with the cell receptors and triggers the subsequent

60 fusion process with cell membrane and/or uptake by endocytosis [2–4].

61 Equine herpesvirus type 1 (EHV-1) and EHV-4 use equine major histocompatibility complex

62 class I (MHC-I) as an entry receptor, however, no details of the molecular binding mode are

63 available [5–7]. Only few other viruses are known to utilize MHC molecules as binding

64 receptors but not as entry receptors. Coxsackievirus A9 requires MHC-I and GRP78 as co-

65 receptors for virus internalization [8], Simian virus 40 (SV40) binds to cellular MHC-I,

66 however, MHC-I does not mediate virus entry [9,10]. The fiber knob of Adenovirus type 5

67 (AdV-5) binds to the α2 region of human leukocyte antigen (HLA) [11],and the functional gD

68 homolog gp42 in Epstein-Barr Virus (EBV) binds to MHC-II to activate membrane fusion

69 [12].

70 MHC-I seems to be an unlikely receptor for viral entry since it is present on all somatic

71 cells [13] and therefore does not confer tissue specificity. Additionally, it is one of the most

72 polymorphic mammalian proteins with 10 to 25% difference in the amino acid sequence

73 [14,15]. Typically, MHC-I plays a crucial role in the adaptive immunity by presenting

74 proteolytically processed intracellular proteins on the cell surface to T-cells and natural killer

75 cells [16]. In case of an infected cell, virus-derived peptides are presented and the recognition

76 by T-cell receptor (TCR) initiates an immune response [17]. Although utilized by EHV-1 and

77 EHV-4 as entry receptors, not all MHC-I genes support the entry of both viruses [5,7,18].

78 Interestingly, the residue A173 in the α2 region of MHC-I seems to be necessary but not

79 enough to trigger virus entry [7,18,19].

80 EHV-1 and EHV-4 are important pathogens that cause great suffering in *Equidae* and other

81 mammals and result in huge economic losses to the equine industry [20]. Efforts have been

82 made to find efficient vaccines against both viruses [21]. However, the protection is usually

83 limited in time and efficacy; frequent outbreaks occur also in vaccinated horses [22–25].

84 Here, we present crystal structures of free gD1 and gD4, which show a similar fold as other

85 gD proteins from related viruses such as herpes simplex virus type 1 (HSV-1) (PDB-ID 2C36,

86 [26], pseudorabies virus (PrV, PDB-ID 5X5V, 27), and bovine herpesvirus type 1 (BoHV-1,

87 PDB-ID 6LS9, [27]. We further measured dissociation constants (in a micromolar range) for

88 recombinant gD1/gD4 and C-terminally truncated gD4 binding to equine MHC-I by surface

89 plasmon resonance spectroscopy (SPR). No increased binding affinity was observed for the

90 truncated protein as was the case for gD of HSV-1, HSV-2, and PrV [28,29], suggesting a

91 structurally different mode of binding during entry into host cells. Cell culture assays showed

92 that recombinant gD1 and gD4 as well as truncated gD4 can inhibit viral replication *in vitro*,

93 where again the truncated version did not perform better than the full-length protein. The

94 crystal structures were further used for *in silico* docking analyses to equine MHC-I. Based on

95 these docking positions, viral mutants with point mutations at position F213 or D261 were

96 produced and displayed significant growth impairments which support the proposed mode of

97 binding of gD1 and gD4 to MHC-I.

98

99 ## Results

Crystal structure of unbound EHV-1 and EHV-4 gD

100 Recombinant gD1 and gD4 lacking the transmembrane region were produced in insect cells,

102 purified by affinity and size exclusion chromatography and used for crystallization

103 experiments (Figure S 1 and Figure S 2). To evaluate the correct identity, sequence, and

104 molecular mass of gD1, gD4, and equine MHC-I, mass spectrometry (MS) analysis was

105 conducted. Recombinant equine MHC-I 3.1 (Eqca-1*00101) including a peptide

106 (SDYVKVSNI) linked to the β2-microglobulin (β2m) region was produced in insect cells as

107 well and purified in the same manner as gD1 and gD4.

108 To this point, molecular masses of gD1 and gD4 were determined only by SDS-PAGE and

109 Western blotting. However, these techniques are known to often lead to an overestimation of the

110 molecular mass [30]. Here we employed matrix-assisted laser desorption ionization-time of

111 flight mass spectrometry (MALDI-TOF-MS) to analyze diluted recombinant protein. The

112 proteins gD1, gD4 and the MHC-I α-chain contain a Tobacco Etch Virus (TEV) cleavage

113 site and a His_6-tag (ENLYFQG-H_6) which contribute approximately 1675 Da to the

114 molecular weight of the molecules (calculated with https://web.expasy.org/peptide_mass).

115 Additionally, the residues EF (approximately 300 Da) originating from the *Eco-RI* restriction

116 site, were detected by in-source-decay (ISD) in the recombinant proteins. A size of

117 approximately 43078 Da for gD1, 43761 Da for gD4, 37779 Da for the α-chain of MHC-I and

118 13241 Da for β2m with its linker and attached peptide was determined (Figure 1 A-C).

119 Excluding the molecular weight of the TEV cleavage site and the His_6-tag (1675 Da), this

120 translates into a molecular weight of 41403 and 42086 Da for soluble gD1 and gD4,

121 respectively (Table 1) and implies an approximate molecular weight of 49345 Da for the

122 recombinant MHC-I molecule consisting of α-chain (36100 Da) and β2m with linker and

123 peptide (13240 Da), and 48500 Da without the linker.

124 The difference between predicted and observed molecular masses is due to post translational

125 modifications (PTMs) such as glycosylations which contribute approximately 4000-5000 Da

126   (Table 1). Further analysis of recombinant gD1, gD4, and MHC-I by ISD and tandem mass

127   spectrometry (MS/MS) of in-gel digested Coomassie-stained proteins confirmed protein

128   identities and presence of the correct N- and C-termini of gD1 and gD4 and N-terminus of

129   MHC-I α-chain (Figure S 4 A, B, C).

130

131   A 2.45 Å resolution diffraction data set was collected for a gD1 crystal, and the gD1 structure

132   (Figure 2 A) was determined using the structure coordinates of HSV-1 gD (protein data bank

133   PDB-ID 2C36) for molecular replacement and refined to an R_{work} of 20.3% and R_{free} of 25.7%

134   (Table 2, PDB-ID 6SQJ). The crystal structure of gD1 contains two gD molecules per

135   asymmetric unit. In solution only the monomeric form was observed by size exclusion

136   chromatography (SEC)-multi-angle static light scattering (MALS) (Figure S 5). Two ions

137   interpreted as magnesium are trapped between the molecules and coordinated by residues

138   E242 and D261 of both protein chains together with water molecules. The residues E32 to

139   R38 and N281 to T348 could not be modeled due to a lack of electron density. N-linked

140   glycosylations are visible at the predicted sites N20 and N28 [31] which are conserved

141   between gD1 and gD4 but not in gDs of other alphaherpesviruses.

142   GlycoproteinD4 was crystallized with only one protein per asymmetric unit. The structure

143   was determined at a resolution of 1.9 Å (Table 2, PDB-ID 6TM8, Figure 2 B) using the

144   coordinates of gD1 structure for molecular replacement and refined to an R_{work} of 17.5% and

145   R_{free} of 21.5% (Table 2). In total 244 residues could be modeled (R34 to R277).

146   In the structures of gD1 and gD4, six cysteines were found to form three disulfide bonds at

147   sites conserved in members of the gD plolypeptide family: C87/C209, C126/C223, and

148   C138/C147 [27,29,32]. The overall folds of gD1 and gD4 are very similar with a root-mean-

149   square deviation (rmsd) of 0.7 Å for 220 common Cα atoms (Figure 2 C). The cores consist of

150   a nine-stranded (A', B, C, C', C'', D, E, F, and G) β-barrel, arranged in a typical V-like Ig

151   fold, flanked by N- and C-terminal extensions with loops, α-helices (α1, α2, α3', and α3), and

152　small β-strands (str2-4). The termini in both structures point in opposite directions (Figure 2)

153　and the unresolved C-termini are predicted to be unstructured by FoldIndex

154　(https://fold.weizmann.ac.il/fldbin/findex).

155

156　Table 1: Predicted and measured molecular mass in Da of recombinant gD1, gD4 (aa 31-349),

157　and equine MHC-I 3.1 α and β2m region with the uncleaved TEV site and His$_6$-tag.

158　Theoretical masses were calculated using https://web.expasy.org/peptide_mass/ and the actual

159　mass determined by matrix-assisted laser desorption ionization-time of flight mass

160　spectrometry (MALDI-TOF-MS). Post-translational modifications like glycosylations account

161　for the discrepancies between the predictions and measurements. NA: not applicable.

Molecule	predicted	measured	without TEV and His6
gD1	38216	43078	41403
gD4	38252	43761	42086
MHC-I α 1-3	33399	37779	36104
β2m (+linker and peptide)	13243	13241	NA

162

Table 2: Crystallographic data collection and model refinement statistics.

	gD1	gD4
PDB-ID	**6SQJ**	**6TM8**
Data collection		
Wavelength [Å]	1.0332	0.91841
Temperature [K]	100	100

8

Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters		
a, b, c [Å]	71.9; 94.5; 101.3	73.1; 59.6; 69.7
$\alpha=\beta=\gamma$ [°]	90	90
Resolution range [Å]	50.00 - 2.24 (2.38 - 2.24) [a]	50.00 - 1.90 (2.01 - 1.90)
Reflections [a]	218509 (33751) [a]	138685 (10835)
Unique reflections [a]	33402 (5140)	23671 (1810)
Completeness [%][a]	99.1 (95.8)	95.6 (78.2)
Multiplicity [a]	6.5 (6.6)	5.9 (3.5)
Data quality[a]		
$I/\sigma(I)$ [a]	11.71 (0.92)	8.96 (0.96)
Rmeas [%][a]	13.5 (199)	17.4 (126.8)
$CC_{1/2}$ [a]	99.8 (58.6)	99.5 (40.9)
Wilson B factor [Å^2]	53.3	32
Refinement gD1 gD4		
Resolution range [Å][a]	50.00 - 2.24 (2.33 - 2.24)	50.00 - 1.90 (2.01 - 1.90)
Reflections[a]	33399 (3181)	23642 (1792)
Test set (5%) [a]	1669 (159)	1182 (89)
R_{work} [%][a]	20.3 (33.8)	17.5 (30.0)
R_{free} [%][a]	25.7 (34.0)	21.5 (37.2)
Contents of Asymmetric Unit		
Molecules, residues, atoms	2; 477; 4049	1; 244; 2037
Mg^{2+}, GlcNAc molecules, glycerol	2; 5; -	-; -; 4
Water molecules	132	174
Mean Temperature factors [Å^2][b]		

All atoms	58.7	31.1
Macromolecules	58	30.4
Ligands	106.7	49.9
Water oxygens	53.5	36
RMSD from target geometry		
Bond length [Å]	0.007	0.012
Bond angles [°]	0.84	1.04
Validation statistics[c]		
Ramachandran plot		
Residues in allowed regions [%]	2.8	2.5
Residues in favored regions [%]	97.2	97.5
MOLPROBITIY clashscore[d]	3.23	3.9

163   [a] Data for the highest resolution shell in parenthesis

164   [b] Calculated with PHENIX [33]

165   [c] Calculated with MOLPROBITY [34]

166   [d] Clash score is the number of serious steric overlaps (> 0.4) per 1 000 atoms.

167

Comparison of gD1, gD4, and homolog structures

169   The amino acid sequence identity between EHV-1 and EHV-4 gDs is 76%, much higher than

170   compared to HSV-1 (25%, GenBank AAK19597), PrV (34%, GenBank AEM64108) or

171   BoHV-1 (31%, GenBank NP045370). While the global folds of gDs of these different viruses

172   are very similar (Figure 3), the number and positions of α-helices differ. Compared to EHV-1

173   and EHV-4 gDs, there is an extra helix termed α1' present in PrV and BoHV-1 gDs. PrV gD

174   has an exclusive α2' helix which cannot be observed in other gD structures elucidated so far.

175   In HSV-1 and HSV-2 gDs, the α3' helix is missing but is present in EHV-1, EHV-4, PrV,

176   and BoHV-1 gDs (Figure 3D). In HSV-1 and HSV-2 gDs, the α1 helix is split and the α3

10

177   helix is kinked in HSV-1 (Figure 3 D) which is not seen in the other known gD structures.

178   The six disulfide bonds C87/C209, C126/C223, and C138/C147 are conserved across EHV-1,

179   EHV-4, PrV, HSV-1, HSV-2, and BoHV-1 gDs, while the predicted and resolved

180   glycosylation sites in the crystal structure of gD1 and gD4 are only conserved between EHV-1

181   and EHV-4 (N52, N61, N297, N386) (Figure S 6). Between gD1 and gD4, also the

182   magnesium-coordinating residues seen in the gD1 monomer-monomer interface are

183   conserved.

184

Soluble gD1 and gD4 engage recombinant MHC-I with similar binding affinities

187   gD-binding affinities of different alphaherpesviruses to their receptors are known to differ

188   greatly. For example, PrV gD binds human nectin-1 in the nanomolar range [35]. HSV-1 gD

189   interacts more weakly, in a micromolar range, with nectin-1 and herpesvirus entry mediator

190   (HVEM) [26] similarly to BoHV-1 gD with nectin-1 [27] (Table 3). For HSV-1, HSV-2, and

191   PrV gDs, it has been demonstrated that C-terminal truncation of the proteins increases the

192   binding affinities up to 100-fold (Table 3).

193

194   To study the interaction of soluble gD1 and gD4 with recombinant equine MHC-I, a surface

195   plasmon resonance (SPR) binding assay was conducted. α-chains together with β2m with

196   linked peptide of equine MHC-I 3.1 were produced in insect cells and purified by IMAC and

197   SEC (Figure S 1 and Figure S 2 D). Additionally, the receptor affinity of a C-terminally

198   truncated EHV-4 gD, $gD4_{36-280}$, was tested. The truncated protein was produced in the same

199   manner as gD1 and gD4 originally with the goal to crystallize it because the flexible C-

200   terminus was suspected to hinder crystallization of gD1 and gD4. However, shortly after the

201   production of the truncated $gD4_{36-280}$, crystal structures were obtained for both gD1 and gD4

202   proteins. Therefore, for $gD4_{36-280}$ only receptor binding kinetics were determined instead of

11

203   crystallizing it. Another truncated version, $gD4_{45-276}$, could not be produced in insect cells,

204   suggesting that the protein failed to fold properly. Binding analyses for soluble gDs were

205   conducted by using a protein dilution series in a range of 0 to 13 950 nM.

206   Dissociation constants ($K_d{}^{app}$) of 4000 ± 800 nM and 4400 ± 1200 nM were calculated for

207   gD1 and gD4, respectively (Figure 4 A and B). The truncated gD4 version, $gD4_{36-280}$, exhibited

208   a receptor binding affinity to MHC-I in the same order of magnitude (5300 ± 1200 nM;

209   Figure 4 C-E).

210

Table 3: Comparison of dissociation constants ($K_d{}^{app}$) of alphaherpesviruses-gDs binding their respective receptors measured by SPR. SW = swine, BO = bovine, MHC-I = equine (Eqca-1*00101), nectin-1/HVEM = human. The C-terminal truncation of proteins is displayed in brackets under 'gD origin'. The upper part of the table represents full-length proteins, the bottom part truncated proteins.

gD origin	Receptor	K_d (nM)	Reference
EHV-1 (349)	MHC-I	4000 ± 800	This study
EHV-4 (349)	MHC-I	4400 ± 1200	This study
HSV-1 (306)	HVEM	3200 ± 600	[36]
HSV-1 (306)	HVEM	4000	[26]
HSV-1 (306)	nectin-1	2700 ± 200	[37]
HSV-1 (306)	nectin-1	1800	[26]
HSV-2 (306)	HVEM	1500	[36]
PrV (354)	nectin-1	130 ± 70	[35]
PrV (337)	nectin-1	191	[29]
PrV (337)	SW-nectin-1	301	[29]
BoHV-1 (301)	nectin-1	879 ± 101	[27]

BoHV-1 (301)	BO-nectin-1	341 ± 106	[27]
EHV-4 (280)	MHC-I	5300 ± 1200	This study
HSV-1 (285)	HVEM	37	[38]
HSV-1 (285)	HVEM	110	[26]
HSV-1 (285)	nectin-1	38	[39]
HSV-1 (285)	nectin-1	70	[26]
HSV-1 (285)	nectin-1	17.1	[40]
HSV-1 (285)	nectin-1	12.5	[28]
HSV-2 (285)	nectin-1	19.1	[28]
PrV (284)	nectin-1	16.1	[29]
PrV (284)	SW-nectin-1	18.4	[29]
BoHV-1 (274)	nectin-1	701 ± 68	[27]
BoHV-1 (274)	BO-nectin-1	489 ± 157	[27]

211

Recombinant gD1, gD4, gD436-280 can block cell surface MHC-I

213 To test whether recombinant gD1 and gD4 can bind to cell surface MHC-I and subsequently

214 inhibit virus entry, blocking assays were performed. Equine dermal (ED) cells were incubated

215 with the recombinant proteins ranging from 0 to 150 µg/ml (0 - 3.5 µM) for one hour on ice

216 and subsequently infected with either EHV-1 or EHV-4 at a multiplicity of infection (MOI) of

217 0.1. Viruses expressing green fluorescent protein (GFP) [41,42] during early infection were

218 used to monitor and analyze the infection levels by flow cytometry. A dose-dependent

219 reduction of infection of up to 50% and 33% on average with 150 µg/ml protein was observed

220 for gD1 and gD4, respectively (Figure 5 A and B).

221 Plaque reduction assays were performed by using a similar procedure. Here, 150 µg/ml of

222 gD1, gD4, and gD4$_{36-280}$ were used and cells were infected with 100 plaque forming units

223   (PFU) of EHV-1 or EHV-4. In the presence of soluble gD1, plaque numbers were decreased

224   on an average by 51% for EHV-1. For EHV-4, the infection was reduced by an average of

225   32% after blocking the cells with soluble gD4 recombinant protein. gD4 was also able to

226   block the entry of EHV-1 by 40%. Likewise, gD1 reduced EHV-4 infection by 29%. In

227   general, gD1 proved to be more efficient in blocking both virus infections. The gD4 variant

228   lacking the C-terminal membrane-proximal residues, $gD4_{36-280}$, could inhibit EHV-4 infection

229   more efficiently with an average of 46% and proved to be slightly more potent than the

230   extracellular domain of gD4 (32%) (Figure 5 C).

231   Taken together, all recombinant gDs compete with viral native proteins. A dose-dependent

232   reduction of infection was observed for gD1 and gD4. Notably, both recombinant gDs are able

233   to efficiently block the entry of EHV-1 and EHV-4.

In silico modeling predicts gD1 and gD4 residues F213 and D261 as hot spots for MHC-I binding

236   No structures are available for gD1 or gD4 in complex with MHC-I. Therefore, protein-

237   protein docking experiments were performed. Based on data from EHV-1 and EHV-4

238   mutational studies with diverse MHC-I genotypes it can be assumed that gD binds in close

239   proximity to MHC-I A173 since genotypes with other residues at this position are highly

240   resistant against infections [7,18]. Available crystal structures of equine MHC-I Eqca-

241   N*00602 (1.18.7–6) and Eqca-N*00601 (10.18) [43] feature a glutamic acid and a threonine

242   residue at position 173 in the α2 chain, respectively, and are known not to support EHV-1

243   and EHV-4 infection [7]. Additionally, they contain mouse instead of equine β2m. Therefore,

244   for *in silico* modeling of gD1 and gD4 binding MHC-I, a homology model of MHC-I

245   genotype 3.1 was constructed to reproduce the experimental setup from the *in vitro* assays.

246   The α-chain template (PDB ID: 4ZUU [43] and target sequences showed 85% identity and

247   88% similarity allowing the development of a high confidence model. In the next step, we

14

248   built a homology model of equine β2m to achieve a physiological MHC-I state. The β2m

249   template (PDB ID: 4ZUU) and target sequences showed 63% identity and 82% similarity and

250   were therefore also highly suitable for homology modeling. The final homology models of the

251   α-chain and β2m contained no Ramachandran outliers [44] (Figure S 7). The calculated

252   backbone rmsd to the template amounted to 0.4 Å in each structure suggesting the correct

253   global fold of the model. All positions (101-164 and 203-259 in MHC-I and 25-80 in β2m)

254   and geometries of disulfide bonds considered typical for MHC were correct, suggesting a high

255   model quality. The final model structure was obtained after assembling both chains and

256   relaxing the homology model with a molecular dynamics (MD) simulation and used directly

257   for the gD docking.

258   As described in the Materials and Methods section (*in silico* modeling suggests low impact of

259   MHC-I peptide on gD-MHC-I binding), we initially identified gD residue D261 as a plausible

260   hot spot contacting R169 in a peptide-free MHC-I docking. In order to mimic the

261   experimental setup in a more realistic way, we inserted the peptide SDYVKVSNI into the

262   MHC-I homology model for docking. This nonapeptide binds in a cleft between α2 and α3

263   helices of MHC-I and was used in the cell-based assays. Since the peptide conformation is

264   strongly dependent on the peptide length [43], the peptide CTSEEMNAF from MHC-I

265   1.18.7–6 (PDB-ID: 4ZUU) was used to build a plausible model. The modeled peptide

266   SDYVKVSNI showed no steric clashes and exhibited reasonable bond geometries with a

267   negligible deviation of backbone atom positions (calculated backbone rmsd of 0.8 Å to the

268   template) (**Error! Reference source not found.**).

269   gD1 and gD4 were docked to the MHC-I homology model with the modeled peptide.

270   Subsequently, a structure showing initially identified contacts between MHC-I R169 and gD

271   D261 were searched. Since ionic interactions are considered important for long-range binding

272   partner recognition [45], it was assumed that this contact should be present in the protein-

273   protein-interaction (PPI). In total, five complexes for gD1 and 21 complexes for gD4,

274 respectively, were found to bind to MHC-I. All structures were visually inspected according

275 to selection criteria outlined in Table 5 and one docking result was selected for each protein

276 complex. The EHV-1 gD complex was one of the 3% best scored results and the EHV-4 gD

277 complex was in the 11% top results suggesting that both docking solutions represent low-

278 energy protein complexes. Both structures showed similar gD-MHC-I orientations (Figure 6

279 A, C) and recurring comparable contacts over the trajectories of MD simulations (**Error!**

280 **Reference source not found.**).

281 Two contacts frequently observed between gD1 and MHC-I were identified using PyContact

282 [46]. The first hot spot residue is D261 surrounded by the assumed O-ring including residues

283 F213 and W257 which are contacting MHC-I binding pocket residue R169. The second hot

284 spot residue is F213 surrounded by the assumed O-ring including Y108 and N110 and

285 contacting MHC-I binding pocket residue I166. Additionally, an extensive hydrogen bond

286 network between residues R103 – E242 and E113 – R238 (Figure 6, B and D) was detected

287 (MHC-I – gD residues, respectively). All contacts and their frequencies over the trajectory of

288 MD simulations are summarized in **Error! Reference source not found.**.

289 PPIs over the course of MD simulations revealed minor movements measured as backbone

290 rmsd of maximal 3.5 Å and 6.5 Å (gD1 and gD4, respectively; **Error! Reference source not**

291 **found.**C and D). Based on the optimized docking models, two gD variants were designed for

292 experimental validation in the next step: F213A and D261N. Both residue exchanges are

293 predicted to disrupt gD – MHC-I contacts and lead to inhibition of viral replication in a cell-

294 based assay.

Mutating F213A and D261N in EHV-1 and 4 gD leads to growth defects

297 The gD1/4-MHC-I binding hypotheses (Figure 6) were experimentally investigated by

298 mutating the proposed key residues F213 to alanine and D261 to asparagine in EHV-1 and

16

299   EHV-4 gDs. Two-step Red-mediated mutagenesis [47] was performed on EHV-1 and EHV-4

300   bacterial artificial chromosomes (BACs) and multi-step growth kinetics with plaque

301   reduction assays were used for virus characterization.

302   All mutant viruses were successfully reconstituted from mutated BACs and the modified gD gene

303   sequences were confirmed by Sanger sequencing. EHV-1-gD$_{F213A}$ displayed a significant 2-

304   log reduction in growth kinetics and low titers in cell supernatant compared to the parental

305   virus. Reverting the mutation rescued virus growth in cell culture (Figure 7 B). Plaque sizes of

306   wild type, mutant and revertant viruses were similar. The virus mutants EHV-1-gD$_{D261N}$,

307   EHV-4-gD$_{D261N}$ and EHV-4-gD$_{F213A}$ did not grow in cells to the extent where growth kinetics

308   could be analyzed. However, reverting the residue exchange in EHV-1-gD$_{D261N}$ rescued virus

309   growth (Figure 7 A). Taken together, the gD$_{D261N}$ and gD$_{F213A}$ variants lead to replication-

310   deficient viruses in EHV-1 and EHV-4.

Discussion

312   Although details of cell entry of alphaherpesviruses can differ greatly between virus species,

313   four common steps characterize the whole entry processes. First gB and/or gC attach in a

314   relatively unspecific and reversible manner to cell surface heparan sulfate proteoglycans

315   (HSPG) and chondroitin sulfate proteoglycans (CSPG) [48–50]. This charge-based interaction

316   is stabilized by a stronger and specific receptor-ligand interaction [51] followed by a signaling

317   cascade which is activated by gD and gH/gL [52]. The latter process leads ultimately to the

318   fusion of the viral envelope with the cell membrane or in some cases to entry via endocytosis

319   through gB [4,53], gD is the essential protein that, in case of EHV-1 and EHV-4, binds to

320   equine MHC-I [5–7]. The mode of gD binding to MHC-I remains elusive, although the

321   structural understanding of alphaherpesviral gDs binding to their putative receptors has been

322   largely extended in the last years [26,27,29,32]. Here, we present crystal structures of EHV-1

17

323 and EHV-4 gDs and propose a binding model to equine MHC-I through the key residues F213

324 and D261.

325 The crystal structures of EHV-1 and EHV-4 gDs revealed an IgV-like fold with large N- and

326 C-termini wrapping around the core which is common for members of the gD polypeptide

327 family. Despite high variability in sequence identities, the overall structure of alphaherpesviral

328 gDs is conserved with only small variations in the loop regions and number of helices

329 [26,27,29,32]. The gD termini have been shown to be important for the entry process in HSV-

330 1 [26]. To allow receptor binding, the C-terminus needs to be displaced to free the N-terminal

331 binding site. This could be a mechanism to prevent early onset of the fusion process before the

332 ligand and receptor are in proximity. Subsequently, the displacement of the C-terminus allows

333 the formation of an N-terminal hairpin loop that is crucial for HVEM binding, since

334 exclusively gD N-terminal residues (7-15 and 24-32) interact with the receptor [26,32,54]. The

335 displacement of the C-terminal tail is also needed for the complex formation with nectin-1 as

336 the binding sites overlap with those of HVEM with additional C-terminal interactions

337 (residues 35–38, 199–201, 214–217, 219–221, 223) [55]. The formation of an N-terminal loop

338 is not involved in nectin-1 binding since the deletion of residues 7-32 had little impact on the

339 interaction [56]. The N-terminus of gD1 and gD4 is, similarly to PrV gD, shorter than in

340 HSV-1 gD, suggesting that HVEM cannot function as an entry receptor in these viruses. In

341 fact, it has already been experimentally observed that HVEM is not used as entry receptors by

342 PrV [29]. Similarly, we observed that EHV-1 also does not employ the equine HVEM

343 homolog either [53]. In HSV-1, gD forms a dimer in the unbound state on the virus envelope

344 [57]. This is thought to stabilize the C-terminus since viruses with a destabilized terminus

345 could not efficiently enter cells. Although the ionic contact and high Complex Formation

346 Significance Score of the here solved EHV-1 gD dimer interface suggest a similar function, no

347 dimer was observed in SEC, SEC-MALS, and MS analysis. Therefore, we suppose that gD1

348 has no dimeric form on the virus envelope.

349   In contrast to results from C-terminally truncated gD homologs, which display a dramatic

350   increase in receptor affinity (Table 3), truncated $gD4_{36-280}$ binds MHC-I similarly as the non -

351   truncated version. The higher affinities in the truncated homologs are explained by a faster

352   interaction with the receptors, since the C- terminus that blocks the binding site is not required

353   to be displaced upon binding anymore [58]. That the C-terminal truncation had no significant

354   effect on the receptor affinity of EHV-4 gD suggests that the mode of binding differs from

355   other alphaherpesviruses. Taking into account that EHV-1 and EHV-4 bind to MHC-I instead

356   of HVEM or nectin-1, a different binding mechanism would be assumed. In line with $gD4_{36-280}$,

357   truncated BoHV-1 gD interaction with nectin-1 showed no increased receptor affinity [27]

358   (Table 3). A conformational change in the loop region between the G-strand and α2 helix is

359   needed for receptor binding [27] and might explain why the affinity of the truncated BoHV-1

360   gD does not increase.

361   SPR analysis showed binding of recombinant gD1, gD4, and $gD4_{36-280}$ to recombinant MHC-I

362   with micromolar affinities. The K_d^{app} are higher than in gD homologs of HSV-1, HSV-2, and

363   PrV binding nectin-1. However, HSV-1 gD binding to HVEM displays similar affinities

364   (Table 3 upper part). Nevertheless, there are two limitations of the SPR analysis in this study.

365   First, the MHC-I molecule Eqca-1*00101 (3.1) used here allowed lower infection rates in a

366   previous study than the molecule Eqca-16*00101 (2.16) [7]. Due to construct design reasons,

367   the gene 3.1 fitted the purpose of crystallography better, although, no crystal structure could

368   be obtained. However, the 2.16 molecule should display higher receptor affinities than the one

369   observed in this study. Second, the linker region (GGGSGGGSGGGS), inserted to tether the

370   peptide in the MHC-I complex to the β2m C-terminus, might interfere with gD receptor

371   binding. Our attempt to model the linker to the MHC-I molecule that binds gD1 in the

372   position hypothesized here support this hypothesis. Nevertheless, the results from blocking

373   assays confirm that the receptor affinities of soluble gDs are unlikely to be in the nanomolar

374   range. Furthermore, blocking assays revealed that gD1, gD4, and $gD4_{36-280}$ can block cell

19

375 surface MHC-I and thus compete with native gD in the viral envelope. It could be

376 demonstrated that gD1 can block EHV-4 infections and vice versa implying that the receptor

377 interaction is very similar in both viruses. This finding is supported by the binding models

378 presented here.

379 The finding in SPR analysis that the C-terminally truncated gD4 does not display an increased

380 receptor affinity was confirmed in blocking assays, thus suggesting that the receptor-binding

381 mode differs from HSV and PrV, which is not surprising as they enter through different

382 receptors.

383 The proposed docking position of gD1 to MHC-I explains why MHC-I Eqca-16*00101 (2.16)

384 allows higher infection rates than Eqca-1*00101 (3.1) [7]. The residue 103 in the 3.1 α1 region

385 is an arginine, which is more spacious than asparagine in 2.16, thus preventing a closer

386 interaction with gD and leading to lower receptor binding affinities. A binding hypothesis with

387 MHC-I 2.16 and a crystal structure of this molecule could confirm that theory. A173 of

388 MHC-I has been shown previously to play a major role in the entry of EHV-1 and EHV-4 by

389 two studies. First, the entry of EHV-1 into usually non-susceptible NIH3T3 cells transfected

390 with altered hamster MHC-I Q173A has been shown together with the negative effect on

391 infection rates of hydroyphilic residues at position 173 in equine MHC-I [18]. Second, it has

392 been demonstrated that not all equine MHC-I genes support entry of EHV-1 and 4 into equine

393 MHC-I transfected mouse mastocytoma (P815) cells and that MHC-I genes harboring

394 residues other than alanine at position 173 are highly resistant against EHV-1 and 4 infections

395 [7]. The gD1/4-MHC-I binding hypotheses explain the role of MHC-I A173 well by showing

396 that bulkier amino acids at that position lead to steric hindrance in the gD binding pocket.

397 This applies to MHC-I alleles 3.3 (V173), 3.4 (T173), 3.5 (E173), and 3.6 (V173), which do

398 not support EHV-1 and 4 entry [7,18]. The model can even explain why the genotype Eqca-

399 7*00201 (3.7), although harboring an alanine at position 173, does not allow entry of EHV-1

400 and 4 into P815 3.7 [7]. The glutamine residue at position 174 is assumed to hinder gD

20

401   binding sterically. The side-chain would point in the bound state into a hydrophobic residue-

402   patch (W253, F256, W257) of gD, leading to an enthalpic penalty. Strangely, the inability of

403   the viruses to enter via the MHC-I haplotype Eqca-2*00101 (3.2) which harbors A173 and

404   A174 cannot be explained by the binding model. The topology of this MHC-I molecule is

405   predicted to be very similar to those allowing virus entry. A crystal structure of the 3.2 MHC-I

406   gene might give an explanation. Mutations in the gD binding pocket R43, W253, F256, and

407   W257 could prove useful for a more detailed evaluation of the predicted interaction with

408   MHC-I A173.

409   Another observation by Sasaki et al. [18] was that the mutation W171L in equine MHC-I

410   impairs virus entry into NIH3T3 cells transfected with this MHC-I. Although the cell surface

411   expression of this mutant was reduced, this is still interesting since structural data show that

412   W171 points towards the peptide in the binding groove and should therefore not be involved

413   directly in binding gD. The tryptophan would be able to stabilize some peptides with hydrogen

414   bonds, whereas a leucine would not. A leucine at position 171 could therefore lead to a

415   more loosely bound peptide with a higher flexibility, resulting in an interference via the gD-

416   MHC-I binding. This theory would suggest that the peptide in the MHC-I binding groove

417   itself could play a role in the receptor-ligand interaction, which could be tested by using

418   different peptides bound to MHC-I in blocking assays and by testing mutated equine MHC-I

419   W171L in blocking assays with soluble recombinant gDs.

420   Considering all these results, the question arises whether EHV-1 and EHV-4 can facilitate entry

421   through, so far, unknown non-equine MHC-I molecules. Sasaki et al. [18] demonstrated that

422   mutated hamster MHC-I Q173A allowed low EHV-1 infection. Unfortunately, EHV-4 has not

423   been tested in the same manner. A computational approach could be employed to search for non-

424   equine MHC-I molecules that are similar in the binding region that is visible in the gD1/4-MHC-

425   I binding model and be used to select promising targets for transfection/infection assays.

426   Experimentally, EHV-1 and EHV-4 infections could be tested in cell lines from susceptible

427 species, e.g. bovine, rabbit, monkey, pig, cat, human [59,60], alpacas, lamas, polar bears [61],

428 and rhinoceros [61–63] cell lines, with and without inhibited MHC-I expression by using β2m

429 knockdown as in Sasaki et al. [6].

430 Taken together, the proposed docking modes of gD1 and gD4 to MHC-I can explain several

431 experimentally obtained results and are therefore plausible. Additionally, the docking models

432 are supported by EHV-1 and EHV-4 viruses with mutated gD_{F213A} and gD_{D261N} that all

433 exhibited an impaired growth. A limitation in this experiment was the difficulty of reverting

434 EHV-4 mutations to original status to confirm that the observed effect was exclusively due to

435 the gD_{F213A} and gD_{D261N} variants. Nevertheless, it could be shown that the gD residues F213

436 and D261 play a key role during entry of EHV-1 and EHV-4 providing starting points for

437 further mutational studies possibly leading to an efficient vaccine. The results presented here

438 might also be used to generate gD-based EHV-1 and EHV-4 inhibitors for reduction of clinical

439 symptoms in horses and non-definite hosts.

440

Materials and Methods

Viruses

443 EHV-1 strain RacL11 and EHV-4 strain TH20p are maintained as bacterial artificial

444 chromosome infections clones (BAC). The viruses have GFP under the control of the HCMV

445 major immediate-early (IE) promoter inserted into the Mini-F sequence to easily recognize

446 infected cells. Clones were generated as described previously in Rudolph et al. and Azab et al.

447 [41,42,64]. The viruses were grown on equine dermal (ED) cells (CCLV-RIE 1222, Federal

448 Research Institute for Animal Health, Greifswald, Germany) at 37 °C under a 5% CO2

449 atmosphere.

450

Cells

452 ED cells were grown in Iscove's modified Dulbecco's medium (IMDM; Pan, Biotech,

453 Aidenbach, Germany) containing 20% fetal bovine serum (FBS; Biochom GmBH, Berlin,

454 Germany), 1 mM sodium pyruvate (Pan Biotech, Aidenbach, Germany), 1% nonessential

455 amino acids (NEAA; Biochom GmBH, Berlin, Germany), and P-S solution (100 U/mL

456 penicillin: Panreac, AppliChem GmBH, Darmstadt, Germany; 100 µg/mL streptomycin: Alfa

457 Aesar, Thermo Fisher Scientific, Kandel, Germany (P-S) at 37°C under a 5% CO_2

458 atmosphere.

459 Human embryonic kidney (293T, ATCC CRL-11268) cells were propagated in Dulbecco's

460 modified Eagle's medium (DMEM; Biochom GmBH, Berlin, Germany), supplemented with

461 10% FBS and P-S. Sf9 cells (IPLB-Sf21-AE, Invitrogen, Germany) were propagated in serum

462 free Sf-900 III medium (Gibco, Thermo Fisher Scientific, New York, USA) and High Five™

463 cells (BTI-TN-5B1-4, Invitrogen, Germany) in serum free Express Five medium (Gibco,

464 Thermo Fisher Scientific, New York, USA) at 27°C on orbital shaker.

465

Construction of expression plasmids

467 Constructs were amplified from insect cell codon-optimized DNA fragments (Bio Basic Inc.,

468 Canada) for protein production in High Five insect cells. Synthetic truncated genes contained

469 the gene of interest (gD1 residues 32-249, gD4 residues 32-249, equine MHC-I 3.1,), a C-

470 terminal TEV protease cleavage site (ENLYFQG), and a hexa-histidine tag (His_6), all flanked

471 by *Eco*RI- and *Sca*I-restriction sites (Figure 8). Sequences of codon optimized genes can be

472 found in the supplementary data. A further truncated form of gD4 containing the residues 36-

473 280 was amplified from gD4 synthetic gene with the primer pair VK50/VK56 (**Error!**

474 **Reference source not found.**).

475 The Autographa californica nuclear polyhedrosis virus (AcNPV) baculovirus gp64 signal

476 sequence under control of the very late polyhedrin promoter was inserted into the insect cell

477 vector plasmid pACEBac1 (Addgene, LGC Standards Teddington, UK) by using another

23

478   synthetic gene (VK18, LGC Genomics, Berlin, Germany) and the primer pairs VK7/VK7

479   (**Error! Reference source not found.**). Subsequently, plasmids containing synthetic genes

480   (gD1, gD4, MHC-I) were amplified in *Escherichia coli* (E. coli), purified, and digested with

481   *Eco*RI- and *Sca*I-restriction enzymes for insertion into the transfer vectors which was digested

482   with the same restriction enzymes. After ligation, these plasmids were transformed into

483   DH10MultiBac electrocompetent cells and recombinant baculoviruses produced according to

484   manufacturer's instructions (Bac-to-Bac expression system, Invitrogen). All constructs were

485   verified by sequencing (VK8 or VK10/WA2, VK35/38; **Error! Reference source not**

486   **found.**). Recombinant BACs were isolated and used for virus production in Sf9 cells as

487   described in Santos et al. [65].

488   ## **Protein production and purification**

489   Equine MHC-I, $gD1_{32\text{-}349}$, $gD4_{32\text{-}349}$, and $gD4_{36\text{-}280}$ were expressed in HighFive cells. Cell

490   supernatant was harvested after 72 h post infection by centrifugation. The pH was adjusted to

491   7 with 1M tris(hydroxymethyl)aminomethan (Tris)-HCl buffer at pH 9 on ice and incubated

492   for at least 1 h with washed Ni^{2+}-NTA beads for affinity chromatography. Beads with bound

493   recombinant protein were collected by a gravity flow column and the proteins were eluted

494   with a buffer containing 20 mM Tris-HCl at pH 7.5 or 2-(N-morpholino)ethanesulfonic acid

495   (MES) at pH 6 for gDs and MHC-I, respectively, and 200 mM NaCl, 5% glycerol, and 200

496   mM imidazole. Concentrated protein was loaded onto a 16/600 Superdex 200 gel filtration

497   column (GE Healthcare Piscataway, NJ). The buffer conditions were the same as in Ni^{2+}-NTA

498   affinity chromatography but with 20 mM NaCl and no imidazole. Proteins collected from

499   size-exclusion chromatography were concentrated (Concentrators, Amicon Ultra, Millipore,

500   Darmstadt, Germany), aliquoted and directly used for crystallization or stored at $-80°C$.

501

Crystallization, structure determination, and refinement

Crystals of EHV-1 gD were obtained by the sitting-drop vapor-diffusion method at 18°C with a reservoir solution composed of 0.1 M Tris/HCl buffer at pH 8.5, 0.2 M MgCl, and 30% (w/v) polyethylene glycol (PEG) 4000. Crystals were cryo-protected with a solution composed of 75% mother liquor and 25% (v/v) glycerol and subsequently flash-cooled in liquid nitrogen. Synchrotron diffraction data were collected at the beamline P14 at DESY (Hamburg, Germany) and at the beamline 14-2 of the MX beamline of the BESSY II (Berlin, Germany) and processed with X-ray detector software (XDS) [66]. The structure was solved by molecular replacement with PHASER [67] using the coordinates of PDB-ID 2c36 as search model for gD1 which was then used as search model for gD4. A unique solution with two molecules in the asymmetric unit for gD1 and molecule for gD4 were subjected to the program AUTOBUILD in PHENIX [33] and manually adjusted in COOT [68]. The structures were refined by maximum-likelihood restrained refinement using PHENIX [33,69]. Model quality was evaluated with MolProbity [70] and the JCSG validation server [71]. Secondary structure elements were assigned with DSSP [72] and for displaying sequence alignments generated by ClustalOmega [73] ALSCRIPT [74] was used. Structure figures were prepared using PyMOL [75]. Coordinates and structure factors have been deposited in the PDB for gD1 with PDB-ID 6SQJ as well as for gD4 with PDB-ID 6TM8. Diffraction images have been deposited at proteindiffraction.org (gD1: DOI 10.18430/m36sqj and gD4 DOI 10.18430/m36tm8).

Mass spectrometry analysis

Intact protein mass of gD1, gD4, and MHC-I was determined by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF-MS) using an Ultraflex-II TOF/TOF instrument (Bruker Daltonics, Bremen, Germany) equipped with a 200 Hz solidstate Smart beam™ laser. Samples were spotted using the dried-droplet

25

528 technique on sinapinic acid (SA) or 2,5-dihydroxybenzoic acid (DHB) matrix (saturated

529 solution in 33% acetonitrile / 0,1% trifluoroacetic acid). The mass spectrometer was

530 operated in the positive linear mode, and spectra were acquired over an m/z range of 3,000-

531 60,000. Data was analyzed using FlexAnalysis 2.4. software provided with the instrument.

532 Protein identity was determined by tandem mass spectrometry (MS/MS) of in-gel digested

533 Coomassie stained protein with 12,5μg/ml Glu-C and trypsin, and 10μg/ml Asp-N in 25nm

534 ammonium bicarbonate.

535 N-terminal c and C-terminal (z+2) sequence ion series were generated by in-source decay

536 (ISD) with 1,5-diaminonaphthalene (1,5-DAN) as matrix (20 mg/ml 1,5-DAN in 50%

537 acetonitrile / 0,1% trifluoroacetic acid). Spectra were recorded in the positive reflector mode

538 (RP PepMix) in the mass range 800–4,000.

539 ## SEC-MALS analysis

540 For molecular mass determination of soluble, recombinant gD1, SEC-MALS [76] was

541 performed. Protein solution was run at room temperature on a Superdex 75 10/300 GL (GE

542 Healthcare, Piscataway, NJ) column with 2 mg/ml gD1 and a mobile phase composed of Tris-

543 HCl at pH 7.5, 200 mM NaCl, 5% glycerol, and 0.02% sodium azide, attached to a high-

544 performance liquid chromatography (HPLC) system (Agilent Technologies, Santa Barbara,

545 CA, USA) with a mini DAWN TREOS detector (Wyatt Technology Corp., Santa Barbara,

546 CA, USA). Data was acquired and analyzed with the ASTRA for Windows software package

547 (version 6.1.2).

548

549 ## Surface plasmon resonance

550 Binding kinetics of soluble gD1, gD4, and gD4$_{36-280}$ binding to amine-coupled recombinant,

551 equine MHC-I 3.1 were measured at 25°C on a surface plasmon resonance (SPR) GE Biacore

552 J Biomolecular Interaction Analyser instrument (Uppsala, Sweden) using a polycarboxylate

553   hydrogel sensor chip HC200M (XanTec bioanalytics GmbH). The second channel was coated

554   with poly-L-lysine and positive nanogels (size 214nm) [77] that were shown to interact only

555   weakly with gDs and used as negative control. The control sensorgrams were subtracted from

556   reaction sensorgrams and normalized. The surfaces were regenerated with buffer containing

557   200 mM NaCl and 10 mM NaOH after each cycle. Serial dilutions of gDs ranging from 0 to

558   10000 nM were injected at medium flow and the interaction with MHC-I was monitored for

559   15 min. The response curves of gDs binding to the MHC-I were fitted to the Hill-Wand

560   binding model $R_{eq} = \frac{R_{max}[A]^b}{K_D^b + [A]^b}$ [78] using Sigma plot 12.0 software.

561

Generation and analysis of gD1/4-MHC-I binding model

Protein data

564   Sequences of MHC-I and β2m were obtained from UniProt-Databank [79]. The protein

565   sequences with their respective UniProt IDs are listed in Table 4.

566

Table 4: Protein structures obtained from UniProt with their respective IDs.

Protein	UniProt ID
MHC-I gene Eqca-1*00101	Q30483
MHC-I genotype Eqca-N*00602	Q860N6
Horse β2m	P30441
Mouse β2m	P01887

567

568

Homology modeling

570   Homology models were prepared using MOE (version 2018.0101; Molecular Operating

27

571 Environment, Chemical Computing Group ULC, Montreal, Canada). The models were

572 constructed using GB/VI scoring [80] with a maximum of ten main chain models. To check

573 geometry of obtained homology models, Ramachandran (phi-psi angle) plots [44] were

574 calculated with MOE.

575 The full MHC-I gene 3.1 model was prepared based on the hybrid equine α-chain - mouse

576 β2m X-ray crystal structure with the best resolution (PDB-ID: 4ZUU [43]). The α-chain and

577 β2m homology models were superposed onto the template. All side chain clashes were

578 removed by energy minimization using the OPLS-AA force field [81], resulting in the full

579 MHC-I gene 3.1 model. The complete model was relaxed in a MD simulation on settings

580 described below.

581

582 **Protein-Protein Docking**

583 MHC-I-gD1 complex were prepared with MOE 2018 by protonation [82], modeling of

584 missing side chains, deleting water molecules and charging termini. Protein - protein docking

585 was performed using Rosetta 3 suite (version 2018-33) [83,84]. The orientations of MHC-I

586 and gD were randomized (flags -randomize 1 -randomize 2) and spun (flag -spin) to the

587 beginning of the docking process. Docking perturbation parameters were set to default: 3 Å

588 translational and 8° rotational movement (flag -dock_pert 3 8) [85]. The residue side chains of

589 both docking partners were allowed to rotate around the χ1 and the χ2 angles (flags -ex1 -

590 ex2). In total 10 000 docking runs were conducted (flag -nstruct 10 000) as recommended by

591 the Rosetta documentation [86], yielding over 7 000 poses in each docking round. A flat

592 harmonic distance constraint between the Cα of MHC-I A173 and the gD backbone was

593 added based on reported genotype studies indicating the pivotal role of MHC-I A173 [7,18].

594 This allowed us to limit the number of possible protein-protein docking complexes and

595 perform local docking as recommended by the Rosetta documentation [86]. Constraint

596 parameters were set to the default [86]: Distance 0 Å, standard deviation 1 Å and tolerance 5

597   Å to achieve the closest possible proximity between chains. In order to obtain a full MHC

598   peptide complex, peptide SDYVKVSNI, as used in cell-based assays, was manually fitted

599   into the MHC-I cleft. To fit our sequence, the peptide structure from the template (PDB-ID:

600   4ZUU) containing a nonapeptide CTSEEMNAF was superposed on the MHC-I homology

601   model. The co-crystallized nonapeptide sequence was manually mutated. The side chain

602   conformations were adjusted using MOE's rotamer tool and energy minimized using the

603   OPLS-AA force field to relax atomic clashes.

604   Finally, gD1 and gD4 were docked into the prepared MHC-peptide complex. In order to find

605   the final and most plausible docking pose of gD1 and gD4 in complex with MHC-I-peptide,

606   an in-house developed MD Analysis-based (version 0.19.2) [87,88] script was used to find

607   ionic key-contact defined as a distance of maximal 4.5 Å between Cγ atom of D261 in gD and

608   Cζ atom of R169 in MHC-I. The script was run in a Python 3.6 environment [89].

609

Filtering of Docking Poses and Classification of Residues Involved in the Protein-Protein Interface

612   In order to filter the most plausible from all best-scored docking poses, we applied three rules

613   based on reported statistical evaluation of various protein-protein interactions (PPIs) [90,91]

614   and biological function of Herpesvirus gD (Table 5) [40]. The residues involved in the

615   protein-protein binding in the obtained binding hypotheses were classified according to the O-

616   ring theory (Figure 6) [90].

617

Table 5: Criteria applied in our analysis of protein-protein interfaces to filter the most energetically favored docking poses.

Filtering criterion	Rationale
Discarding docking poses with the C-	Orientation unlikely to be correct because

terminus participating in the resulting PPI	the C-terminus merges with a transmembrane-helix anchoring in viral membrane [40]
Discarding docking poses without lipophilic residues in the modeled PPI	Protein-protein interfaces with lipophilic contacts are common and entropically favored [91]
Accepting poses with contact to residues R, D, H, I, K, P, W, Y buried in interface areas	These residues are statistically enriched in protein-protein binding interfaces according to the O-ring theory [90]

Table 6: Classification of residues applied in our analysis of protein-protein interface.

Residue type in the binding interface	Definition
Hot spot residues	Amino acids statistically enriched in binding sites of protein - protein complexes and contributing more than 2 kcal/mol to the binding energy [90] or form lipophilic contacts [91]
O-ring	Residues preventing solvation of binding hot spots [90]

30

Binding pocket	Counterpart of hot spot residues on the surface of the binding partner.

618

Molecular Dynamics Simulations and Protein-Protein Interaction Analysis

619

620 Molecular Dynamics (MD) simulations were prepared using Maestro (version 11.7;

621 Schrödinger, New York, USA) and carried out using Desmond 2018-3 (version 5.5) [92]. All

622 systems were simulated on water-cooled GeForce RTX 2080 Ti graphics processing units

623 (NVIDIA Corporation, Santa Clara, USA). The full MHC-I gene 3.1 homology model was

624 solvated in a cubic box with 12 Å buffering with SPC water model [93]. The system was

625 neutralized using sodium or chloride ions and osmotic pressure was adjusted with 0.15 M

626 sodium chloride to achieve an isotonic system. The subsequent system relaxation was

627 performed according to the default Desmond protocol. The MD simulation ran under periodic

628 boundary conditions and as an NPT ensemble (constant particle number, pressure and

629 temperature) using the OPLS 2005 force field [94]. The MD simulation was performed in one

630 replicate over 100 ns. Coordinates of the relaxed model were retrieved after the backbone

631 rmsd (**Error! Reference source not found.**A) had reached a stable plateau around 3 Å

632 indicating protein equilibration.

633 Docking poses were simulated under the same conditions as the homology models. The

634 movement of protein - protein complex hypotheses was observed in a single MD simulation

635 over 100 ns resulting in ca. 5000 complex conformations. MD simulations of the final

636 selected docking pose were performed in triplicates. The simulated systems contained around

637 140 000 – 168 500 atoms. The proteins were wrapped, aligned on the backbone and visually

638 inspected in VMD [95] (version 1.9.3). Protein-protein interactions were analyzed using

639 PyContact [46] (version 1.0.1) on default settings (distance cutoff 5.0 Å, angle cutoff 120.0°

640 and distance cutoff between hydrogen and hydrogen bond acceptor of 2.5 Å). The PyContact

31

641   analysis was run in a Python 2.7 environment [89].

642

In silico modeling suggests low impact of MHC-I peptide on gD-MHC-I binding

645   To test whether additional bias emerging from peptide modeling influenced docking

646   experiments, two docking rounds were performed. First, gD1 was docked to peptide-free

647   MHC-I. Second, gD1 was docked to MHC-I 3.1 homology model containing the peptide

648   SDYVKVSNI to check if docking provides comparable PPIs. Both docking rounds were

649   performed using the settings described above. As the initial filtering step, the ten highest

650   scored docking poses with the lowest Rosetta Energy were selected [96]. For further filtering,

651   the rules described in the Methods section were applied (Table 5). Four out of ten docking

652   poses fulfilled all three rules. Subsequently, single MD simulations for each docking pose

653   were performed to examine PPI stability. For all protein - protein complexes the backbone

654   rmsd was calculated to obtain an overview of the amplitude of protein movements. Only one

655   docking pose showed a nearly constant backbone rmsd value of 6 Å indicating low complex

656   movement (**Error! Reference source not found.** B). In order to characterize the obtained

657   PPI, we applied selection criteria and identified three residue patch-classes in the binding

658   surface as described in the Methods section (Figure 6). An inspection of the PPI over an MD

659   simulation trajectory with PyContact [46] (**Error! Reference source not found.**) revealed

660   two gD hot spot residues: D261 (surrounded by assumed gD O-ring T161, F213, and W257

661   and contacting binding pocket MHC-I residue R169 over whole simulation time) and W257

662   (surrounded by assumed O-ring R43, T161, and F213 and contacting binding pocket MHC-I

663   residue I166 over the whole trajectory). It can be concluded that PPIs of peptide-free and

664   peptide-bound docking poses are formed with similar residue patches (**Error! Reference**

665   **source not found.**, **Error! Reference source not found.**) suggesting that the presence of the

666   peptide in MHC-I does not influence gD binding. The key residue F213 is involved in both

32

667 PPIs indicating its importance. We observed that peptide-bound docking poses exploit larger

668 PPIs with more possible interactions than the peptide-free docking pose. We assume that

669 more contacts between gD and MHC-I are favorable for the binding. Therefore, peptide-

670 bound docking poses were chosen as the final ones.

671

BAC mutagenesis

672

673 The point mutations F213A and D261N in EHV-1 and EHV-4 gDs were introduced via a two-

674 step Red recombination [47]. In brief, polymerase chain reaction (PCR) primers (**Error!**

675 **Reference source not found.**) were designed in a way that the 50 nucleotide recombinantion

676 arms include the point mutation and sequence to amplify the kan^R gene. For construction of

677 EHV-1-gD$_{D261N}$, EHV-1gD$_{F213A}$, EHV-4-gD$_{D261N}$, and EHV-4-gD$_{F213A}$ the primer pairs

678 VK61/VK62, VK63/VK64, VK65/VK66, and VK67/VK68 were used for PCR amplification

679 respectively. After Dpn-1 digest of PCR products, fragments were electroporated into GS1783

680 containing EHV-1 or EHV-4 BACs. DNA from Kanamycin resistant colonies was extracted

681 and correct mutants were selected based on Restriction fragment length polymorphism

682 (RFLP) using the restriction enzyme Pst-I. Correct clones were subjected to another round of

683 Red recombination to remove the kan^R gene. Final clones were further analyzed by RFLP and

684 sequencing, BAC extracted, puroified and transfected into 293T cells. Cells and supernatant

685 were harvested three days post transfection and used to infect ED cells. Revertants were

686 produced from mutant clones using the same procedure with primer pairs VK69/VK70,

687 VK71/VK72, VK73/VK74, and VK75/VK76 for producing EHV-1R-gD$_{D261N}$, EHV-1R-

688 gD$_{F213A}$, EHV-4R-gD$_{D261N}$, and EHV-4-RgD$_{F213A}$, respectively. All genotypes were confirmed

689 by PCR, RFLP, and Sanger sequencing using the primer pair WA2/VK8 and WA2/VK10

690 (**Error! Reference source not found.**) for EHV-1 and EHV-4 mutants, respectively.

691

Western blotting

692

33

693   Western blot analysis was performed with soluble proteins: 50 µg/ml MHC-I, 5 µg/ml gD1,

694   and 5 µg/ml gD4. Proteins were separated by 12% SDS-PAGE, transferred to a

695   polyvinylidene difluoride (PVDF) membrane (Roth, Karlsruhe, Germany), detected with 1:1

696   000 dilution rabbit anti-His$_6$ (Sigma-Aldrich, St Louis, USA) antibody and 1:10 000 dilution

697   goat anti-rabbit-HRP antibody (Sigma-Aldrich, St Louis, USA) and visualized by enhanced

698   chemiluminescence (ECL Plus; Amersham).

699

Virus blocking assays

701   To block cell surface MHC-I, 1,5x10^5 ED cells were seeded in 24-well plates. In the next day,

702   cells were incubated with 20, 50, 100 or 150 µg/ml recombinant gD1, gD4 or gD4$_{36-280}$ for 1 h

703   on ice. Subsequently cells were infected with either EHV-1 or EHV-4 at MOI=0.1 and

704   incubated for 1 h at 37°C. To remove un-penetrated viruses, cells were washed with citrate

705   buffer, pH 3, containing 40 mM citric acid, 10 mM potassium chloride and 135 mM sodium

706   chloride, then washed twice with phosphate buffered saline (PBS) and infection allowed to

707   proceed for 24 h (EHV-1) or 48 h (EHV-4). For measurement of fluorescence intensity 10

708   000 cells were analyzed with a FACSCalibur flow cytometer (BD Biosciences) and the

709   software CytExpert (Beckman Coulter, Krefeld). The experiment was repeated three

710   independent times for each protein.

711   For plaque reduction assay the same protocol was applied for blocking surface MHC-I with

712   minor changes. Cells were initially incubated with 150 µg/ml gD1 or gD4, infected with 100

713   PFU, and overlaid with 1.5% methylcellulose (Sigma-Aldrich, Taufkirchen, Germany) in

714   Iscove's Modified Dulbecco's Medium (IMDM) after citrate treatment and washes with PBS.

715   GFP plaques were counted after 48 h with a Zeiss Axiovert.A1 fluorescent microscope (Carl

716   Zeiss AG, Jena, Germany). The experiment was repeated three independent times for each

717   protein.

718

Virus growth kinetics

719
720 Virus replication was tested using multi-step growth kinetics and plaque areas were obtained

721 as described before [50]. ED cells were grown to confluency in 24-well plates, infected with

722 an MOI of 0.1 virus and incubated for 1 h at 37°C. Viruses on the cell surface were removed

723 by washing with citrate buffer. After neutralization with IMDM, cells were washed twice with

724 PBS and finally overlaid with 500 µl IMDM. At indicated times after the citrate treatment

725 cells and supernatant were collected separately for EHV-1 and together for EHV-4 and stored

726 at -80°C. Titers were determined by plating dilution series onto ED cells and counting plaque

727 numbers after two days under a methylcellulose overlay. All plates were fixed for 10 min with

728 4% paraformaldehyde, washed with PBS and stained for 10 min with 0.1% crystal violet

729 solution in PBS which was washed away with tab water. Viral titers are expressed as PFU per

730 milliliter from three independent and blinded experiments.

731

Statistical analysis

732
733 For blocking assays, plaque numbers were normalized to infection levels without recombinant

734 proteins. Statistical analysis was done using GraphPad Prism 5 software (San Diego, CA,

735 USA) and one-way ANOVA Bonferroni's multiple comparison test, * indicates $P \leq 0.05$, **

736 indicates $P \leq 0.01$, *** indicates $P \leq 0.001$. Statistical analysis was done using an unpaired, one-

737 tailed test. $P < 0.05$ was considered significant.

Acknowledgments

738

751

References

753   [1]   Osterrieder N, Van de Walle GR. Pathogenic potential of equine alphaherpesviruses: the

754   importance of the mononuclear cell compartment in disease outcome. Veterinary

755   Microbiology 2010;143:21–8.

756   [2]   Cole NL, Grose C. Membrane fusion mediated by herpesvirus glycoproteins: the

757   paradigm of varicella-zoster virus. Reviews in Medical Virology 2003;13:207–22.

758   [3]   Frampton AR, Stolz DB, Uchida H, Goins WF, Cohen JB, Glorioso JC. Equine

759   herpesvirus 1 enters cells by two different pathways, and infection requires the activation

760   of the cellular kinase ROCK1. Journal of Virology 2007;81:10879–89.

761   [4]   Azab W, Gramatica A, Herrmann A, Osterrieder N. Binding of alphaherpesvirus

762   glycoprotein H to surface α4β1-integrins activates calcium-signaling pathways and

763   induces phosphatidylserine exposure on the plasma membrane. MBio 2015;6:e01552-15.

764   [5]   Kurtz BM, Singletary LB, Kelly SD, Frampton AR. Equus caballus major

765   histocompatibility complex class I is an entry receptor for equine herpesvirus type 1.

766   Journal of Virology 2010;84:9027–34.

767   [6]   Sasaki M, Hasebe R, Makino Y, Suzuki T, Fukushi H, Okamoto M, et al. Equine major

768   histocompatibility complex class I molecules act as entry receptors that bind to equine

769   herpesvirus-1 glycoprotein D. Genes to Cells 2011;16:343–57.

770 [7]   Azab W, Harman R, Miller D, Tallmadge R, Frampton Jr AR, Antczak DF, et al. Equid

771   herpesvirus type 4 uses a restricted set of equine major histocompatibility complex class

772   I proteins as entry receptors. J Gen Virol 2014;95:1554–63.

773 [8]   Triantafilou K, Fradelizi D, Wilson K, Triantafilou M. GRP78, a coreceptor for

774   coxsackievirus A9, interacts with major histocompatibility complex class I molecules

775   which mediate virus internalization. Journal of Virology 2002;76:633–43.

776 [9]   Atwood WJ, Norkin LC. Class I major histocompatibility proteins as cell surface

777   receptors for simian virus 40. Journal of Virology 1989;63:4474–7.

778 [10] Norkin LC. Simian virus 40 infection via MHC class I molecules and caveolae.

779   Immunological Reviews 1999;168:13–22.

780 [11] Hong SS, Karayan L, Tournier J, Curiel DT, Boulanger PA. Adenovirus type 5 fiber

781   knob binds to MHC class I α2 domain at the surface of human epithelial and B

782   lymphoblastoid cells. The EMBO Journal 1997;16:2294–306.

783 [12] Mullen MM, Haan KM, Longnecker R, Jardetzky TS. Structure of the Epstein-Barr virus

784   gp42 protein bound to the MHC class II receptor HLA-DR1. Molecular Cell

785   2002;9:375–85.

786 [13] David-Watine B, Israël A, Kourilsky P. The regulation and expression of MHC class I

787   genes. Immunology Today 1990;11:286–92.

788 [14] Tallmadge RL, Campbell JA, Miller DC, Antczak DF. Analysis of MHC class I genes

789   across horse MHC haplotypes. Immunogenetics 2010;62:159–72.

790 [15] Gilcrease MZ. Integrin signaling in epithelial cells. Cancer Letters 2007;247:1–25.

791 [16] Bjorkman PJ, Parham P. Structure, function, and diversity of class I major

792   histocompatibility complex molecules. Annual Review of Biochemistry 1990;59:253–

793   88.

794 [17] Germain RN, Margulies DH. The biochemistry and cell biology of antigen processing

795   and presentation. Annual Review of Immunology 1993;11:403–50.

796   [18] Sasaki M, Kim E, Igarashi M, Ito K, Hasebe R, Fukushi H, et al. Single amino acid

797   residue in the A2 domain of major histocompatibility complex class I is involved in the

798   efficiency of equine herpesvirus-1 entry. Journal of Biological Chemistry

799   2011;286:39370–8.

800   [19] Ellis S, Martin A, Holmes E, Morrison W. At least four MHC class I genes are

801   transcribed in the horse: phylogenetic analysis suggests an unusual evolutionary history

802   for the MHC in this species. International Journal of Immunogenetics 1995;22:249–60.

803   [20] Patel J, Heldens J. Equine herpesviruses 1 (EHV-1) and 4 (EHV-4)–epidemiology,

804   disease and immunoprophylaxis: a brief review. The Veterinary Journal 2005;170:14–

805   23.

806   [21] Kydd JH, Townsend HG, Hannant D. The equine immune response to equine

807   herpesvirus-1: the virus and its vaccines. Veterinary Immunology and Immunopathology

808   2006;111:15–30.

809   [22] Allen G. Molecular epizootiology, pathogenesis, and prophylaxis of equine herpesvirus-

810   1 infections. Prog Vet Microbiol Immunol 1986;2:78–144.

811   [23] Burrows R, Goodridge D. In vivo and in vitro studies of equine rhinopneumonitis virus

812   strains. Equine Infectious Diseases, Karger Publishers; 1974, p. 306–21.

813   [24] Goehring L, Wagner B, Bigbie R, Hussey S, Rao S, Morley P, et al. Control of EHV-1

814   viremia and nasal shedding by commercial vaccines. Vaccine 2010;28:5203–11.

815   [25] Goodman LB, Wimer C, Dubovi EJ, Gold C, Wagner B. Immunological correlates of

816   vaccination and infection for equine herpesvirus 1. Clin Vaccine Immunol 2012;19:235–

817   41.

818   [26] Krummenacher C, Supekar VM, Whitbeck JC, Lazear E, Connolly SA, Eisenberg RJ, et

819   al. Structure of unliganded HSV gD reveals a mechanism for receptor-mediated

820   activation of virus entry. The EMBO Journal 2005;24:4144–53.

821 [27] Yue D, Chen Z, Yang F, Ye F, Lin S, He B, et al. Crystal structure of bovine herpesvirus

822   1 glycoprotein D bound to nectin-1 reveals the basis for its low-affinity binding to the

823   receptor. Science Advances 2020;6:5147.

824 [28] Lu G, Zhang N, Qi J, Li Y, Chen Z, Zheng C, et al. Crystal structure of herpes simplex

825   virus 2 gD bound to nectin-1 reveals a conserved mode of receptor recognition. Journal

826   of Virology 2014;88:13678–88.

827 [29] Li A, Lu G, Qi J, Wu L, Tian K, Luo T, et al. Structural basis of nectin-1 recognition by

828   pseudorabies virus glycoprotein D. PLoS Pathogens 2017;13:e1006314.

829 [30] Matsumoto H, Haniu H, Komori N. Determination of protein molecular weights on

830   SDS-PAGE. Electrophoretic Separation of Proteins, Springer; 2019, p. 101–5.

831 [31] Flowers CC, Eastman EM, O'Callaghan DJ. Sequence analysis of a glycoprotein D gene

832   homolog within the unique short segment of the EHV-1 genome. Virology

833   1991;180:175–84.

834 [32] Carfi A, Willis SH, Whitbeck JC, Krummenacher C, Cohen GH, Eisenberg RJ, et al.

835   Herpes simplex virus glycoprotein D bound to the human receptor HveA. Molecular Cell

836   2001;8:169–79.

837 [33] Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a

838   comprehensive Python-based system for macromolecular structure solution. Acta

839   Crystallographica Section D: Biological Crystallography 2010;66:213–21.

840 [34] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al.

841   MolProbity: all-atom structure validation for macromolecular crystallography. Acta

842   Crystallographica Section D: Biological Crystallography 2010;66:12–21.

843 [35] Connolly SA, Whitbeck JC, Rux AH, Krummenacher C, Cohen GH, Eisenberg RJ, et al.

844   Glycoprotein D homologs in herpes simplex virus type 1, pseudorabies virus, and bovine

845   herpes virus type 1 bind directly to human HveC (nectin-1) with different affinities.

846   Virology 2001;280:7–18.

847   [36] Willis SH, Rux AH, Peng C, Whitbeck JC, Nicola AV, Lou H, et al. Examination of the

848   kinetics of herpes simplex virus glycoprotein D binding to the herpesvirus entry

849   mediator, using surface plasmon resonance. Journal of Virology 1998;72:5937–47.

850   [37] Whitbeck JC, Muggeridge MI, Rux AH, Hou W, Krummenacher C, Lou H, et al. The

851   major neutralizing antigenic site on herpes simplex virus glycoprotein D overlaps a

852   receptor-binding domain. Journal of Virology 1999;73:9879–90.

853   [38] Rux AH, Willis SH, Nicola AV, Hou W, Peng C, Lou H, et al. Functional region IV of

854   glycoprotein D from herpes simplex virus modulates glycoprotein binding to the

855   herpesvirus entry mediator. Journal of Virology 1998;72:7091–8.

856   [39] Krummenacher C, Rux AH, Whitbeck JC, Ponce-de-Leon M, Lou H, Baribaud I, et al.

857   The first immunoglobulin-like domain of HveC is sufficient to bind herpes simplex virus

858   gD with full affinity, while the third domain is involved in oligomerization of HveC.

859   Journal of Virology 1999;73:8127–37.

860   [40] Zhang N, Yan J, Lu G, Guo Z, Fan Z, Wang J, et al. Binding of herpes simplex virus

861   glycoprotein D to nectin-1 exploits host cell adhesion. Nature Communications

862   2011;2:577.

863   [41] Rudolph J, O'CALLAGHAN D, Osterrieder N. Cloning of the genomes of equine

864   herpesvirus type 1 (EHV-1) strains KyA and RacL11 as bacterial artificial chromosomes

865   (BAC). Journal of Veterinary Medicine, Series B 2002;49:31–6.

866   [42] Azab W, Kato K, Arii J, Tsujimura K, Yamane D, Tohya Y, et al. Cloning of the

867   genome of equine herpesvirus 4 strain TH20p as an infectious bacterial artificial

868   chromosome. Archives of Virology 2009;154:833–42.

869   [43] Yao S, Liu J, Qi J, Chen R, Zhang N, Liu Y, et al. Structural illumination of equine

870   MHC class I molecules highlights unconventional epitope presentation manner that is

871   evolved in equine leukocyte antigen alleles. The Journal of Immunology

872   2016;196:1943–54.

873   [44] Ramachandran GN. Stereochemistry of polypeptide chain configurations. J Mol Biol

874   1963;7:95–9.

875   [45] Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science

876   1995;268:1144–9.

877   [46] Scheurer M, Rodenkirch P, Siggel M, Bernardi RC, Schulten K, Tajkhorshid E, et al.

878   PyContact: Rapid, customizable, and visual analysis of noncovalent interactions in MD

879   simulations. Biophysical Journal 2018;114:577–83.

880   [47] Tischer KB, von Einem J, Kaufer B, Osterrieder N. Two-step red-mediated

881   recombination for versatile high-efficiency markerless DNA manipulation in Escherichia

882   coli. Biotechniques 2006;40:191–7.

883   [48] Osterrieder N. Construction and characterization of an equine herpesvirus 1 glycoprotein

884   C negative mutant. Virus Research 1999;59:165–77.

885   [49] Spear PG, Longnecker R. Herpesvirus entry: an update. Journal of Virology

886   2003;77:10179–85.

887   [50] Azab W, Tsujimura K, Maeda K, Kobayashi K, Mohamed YM, Kato K, et al.

888   Glycoprotein C of equine herpesvirus 4 plays a role in viral binding to cell surface

889   heparan sulfate. Virus Research 2010;151:1–9.

890   [51] Csellner H, Walker C, Wellington J, McLure L, Love D, Whalley J. EHV-1 glycoprotein

891   D (EHV-1 gD) is required for virus entry and cell-cell fusion, and an EHV-1 gD deletion

892   mutant induces a protective immune response in mice. Archives of Virology

893   2000;145:2371–85.

894   [52] Azab W, Zajic L, Osterrieder N. The role of glycoprotein H of equine herpesviruses 1

895   and 4 (EHV-1 and EHV-4) in cellular host range and integrin binding. Veterinary

896   Research 2012;43:61.

897   [53] Azab W, Osterrieder K. Initial contact: the first steps in herpesvirus entry. Cell Biology

898   of Herpes Viruses, Springer; 2017, p. 1–27.

41

[54] Lazear E, Carfi A, Whitbeck JC, Cairns TM, Krummenacher C, Cohen GH, et al. Engineered disulfide bonds in herpes simplex virus type 1 gD separate receptor binding from fusion initiation and viral entry. Journal of Virology 2008;82:700–9.

[55] Di Giovine P, Settembre EC, Bhargava AK, Luftig MA, Lou H, Cohen GH, et al. Structure of herpes simplex virus glycoprotein D bound to the human receptor nectin-1. PLoS Pathogens 2011;7:e1002277.

[56] Manoj S, Jogger CR, Myscofski D, Yoon M, Spear PG. Mutations in herpes simplex virus glycoprotein D that prevent cell entry via nectins and alter cell tropism. Proceedings of the National Academy of Sciences 2004;101:12414–21.

[57] Handler CG, Cohen GH, Eisenberg RJ. Cross-linking of glycoprotein oligomers during herpes simplex virus type 1 entry. Journal of Virology 1996;70:6076–82.

[58] Krummenacher C, Carfí A, Eisenberg RJ, Cohen GH. Entry of herpesviruses into cells: the enigma variations. Viral entry into host cells, Springer; 2013, p. 178–95.

[59] Studdert M, Blackney M. Equine herpesviruses: on the differentiation of respiratory from foetal strains of type 1. Australian Veterinary Journal 1979;55:488–92.

[60] Ahn BC, Zhang Y, O'Callaghan DJ. The equine herpesvirus-1 (EHV-1) IR3 transcript downregulates expression of the IE gene and the absence of IR3 gene expression alters EHV-1 biological properties and virulence. Virology 2010;402:327–37.

[61] Greenwood AD, Tsangaras K, Ho SY, Szentiks CA, Nikolin VM, Ma G, et al. A potentially fatal mix of herpes in zoos. Current Biology 2012;22:1727–31.

[62] Abdelgawad A, Hermes R, Damiani A, Lamglait B, Czirják GÁ, East M, et al. Comprehensive serology based on a peptide ELISA to assess the prevalence of closely related equine herpesviruses in zoo and wild animals. PLoS One 2015;10:e0138370.

[63] Abdelgawad A, Azab W, Damiani AM, Baumgartner K, Will H, Osterrieder N, et al. Zebra-borne equine herpesvirus type 1 (EHV-1) infection in non-African captive mammals. Veterinary Microbiology 2014;169:102–6.

42

925 [64] Azab W, Kato K, Abdel-Gawad A, Tohya Y, Akashi H. Equine herpesvirus 4: Recent

926 advances using BAC technology. Veterinary Microbiology 2011;150:1–14.

927 [65] Santos KF, Jovin SM, Weber G, Pena V, Lührmann R, Wahl MC. Structural basis for

928 functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling

929 of the spliceosome. Proceedings of the National Academy of Sciences 2012;109:17418–

930 23.

931 [66] Kabsch W. Xds. Acta Crystallographica Section D: Biological Crystallography

932 2010;66:125–32.

933 [67] Bunkóczi G, Echols N, McCoy AJ, Oeffner RD, Adams PD, Read RJ. Phaser. MRage:

934 automated molecular replacement. Acta Crystallographica Section D: Biological

935 Crystallography 2013;69:2276–86.

936 [68] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta

937 Crystallographica Section D: Biological Crystallography 2010;66:486–501.

938 [69] Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov

939 M, et al. Towards automated crystallographic structure refinement with phenix. refine.

940 Acta Crystallographica Section D: Biological Crystallography 2012;68:352–67.

941 [70] Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al.

942 MolProbity: More and better reference data for improved all-atom structure validation.

943 Protein Science 2018;27:293–315.

944 [71] Yang H, Guranovic V, Dutta S, Feng Z, Berman HM, Westbrook JD. Automated and

945 accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank.

946 Acta Crystallographica Section D: Biological Crystallography 2004;60:1833–9.

947 [72] Kabsch W, Sander C. DSSP: definition of secondary structure of proteins given a set of

948 3D coordinates. Biopolymers 1983;22:2577–637.

949 [73] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable

950 generation of high-quality protein multiple sequence alignments using Clustal Omega.

951   Molecular Systems Biology 2011;7.

952 [74] Barton GJ, others. ALSCRIPT: a tool to format multiple sequence alignments. Protein

953   Engineering Design and Selection 1993;6:37–40.

954 [75] DeLano WL. PyMOL 2002.

955 [76] Tarazona MP, Saiz E. Combination of SEC/MALS experimental procedures and

956   theoretical analysis for studying the solution properties of macromolecules. Journal of

957   Biochemical and Biophysical Methods 2003;56:95–116.

958 [77] Dey P, Bergmann T, Cuellar-Camacho JL, Ehrmann S, Chowdhury MS, Zhang M, et al.

959   Multivalent Flexible Nanogels Exhibit Broad-Spectrum Antiviral Activity by Blocking

960   Virus Entry. ACS Nano 2018;12:6429–42.

961 [78] Schasfoort RB. Handbook of surface plasmon resonance. Royal Society of Chemistry;

962   2017.

963 [79] Consortium U. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research

964   2019;47:D506–15.

965 [80] Labute P. The generalized Born/volume integral implicit solvent model: estimation of

966   the free energy of hydration using London dispersion instead of atomic surface area.

967   Journal of Computational Chemistry 2008;29:1693–8.

968 [81] Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and

969   reparametrization of the OPLS-AA force field for proteins via comparison with accurate

970   quantum chemical calculations on peptides. The Journal of Physical Chemistry B

971   2001;105:6474–87.

972 [82] Labute P. Protonate3D: assignment of ionization states and hydrogen coordinates to

973   macromolecular structures. Proteins: Structure, Function, and Bioinformatics

974   2009;75:187–205.

975 [83] Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, et al.

976   Protein–protein docking with simultaneous optimization of rigid-body displacement and

977   side-chain conformations. Journal of Molecular Biology 2003;331:281–99.

978 [84] Sircar A, Chaudhury S, Kilambi KP, Berrondo M, Gray JJ. A generalized approach to

979   sampling backbone conformations with RosettaDock for CAPRI rounds 13–19. Proteins:

980   Structure, Function, and Bioinformatics 2010;78:3115–23.

981 [85] Chaudhury S, Gray JJ. Conformer selection and induced fit in flexible backbone

982   protein–protein docking using computational and NMR ensembles. Journal of Molecular

983   Biology 2008;381:1068–87.

984 [86] Weitzner J BB,M; Kilambi, K; Thottungal, R; Chaudhury, S; Wang, C; Gray. Docking

985   Protocol (RosettaDock). 2017.

986 [87] Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, et al. MDAnalysis: a

987   Python package for the rapid analysis of molecular dynamics simulations. Los Alamos

988   National Lab.(LANL), Los Alamos, NM (United States); 2019.

989 [88] Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for

990   the analysis of molecular dynamics simulations. Journal of Computational Chemistry

991   2011;32:2319–27.

992 [89] van Rossum G. Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde

993   en Informatica (CWI), Amsterdam." 1995.

994 [90] Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. Journal of Molecular

995   Biology 1998;280:1–9.

996 [91] Liu Q, Li J. Protein binding hot spots and the residue-residue pairing preference: a water

997   exclusion perspective. BMC Bioinformatics 2010;11:244.

998 [92] Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, et al. Scalable

999   algorithms for molecular dynamics simulations on commodity clusters. SC'06:

1000   Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, IEEE; 2006, p.

1001   43–43.

1002 [93] Toukan K, Rahman A. Molecular-dynamics study of atomic motions in water. Physical

1003   Review B 1985;31:2643.

1004 [94] Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, et al. Integrated modeling

1005   program, applied chemical theory (IMPACT). Journal of Computational Chemistry

1006   2005;26:1752–80.

1007 [95] Humphrey W, Dalke A, Schulten K, others. VMD: visual molecular dynamics. Journal

1008   of Molecular Graphics 1996;14:33–8.

1009 [96] Chaudhury S, Berrondo M, Weitzner BD, Muthu P, Bergman H, Gray JJ. Benchmarking

1010   and analysis of protein docking performance in Rosetta v3. 2. PloS One 2011;6:e22477.

1011

Figures
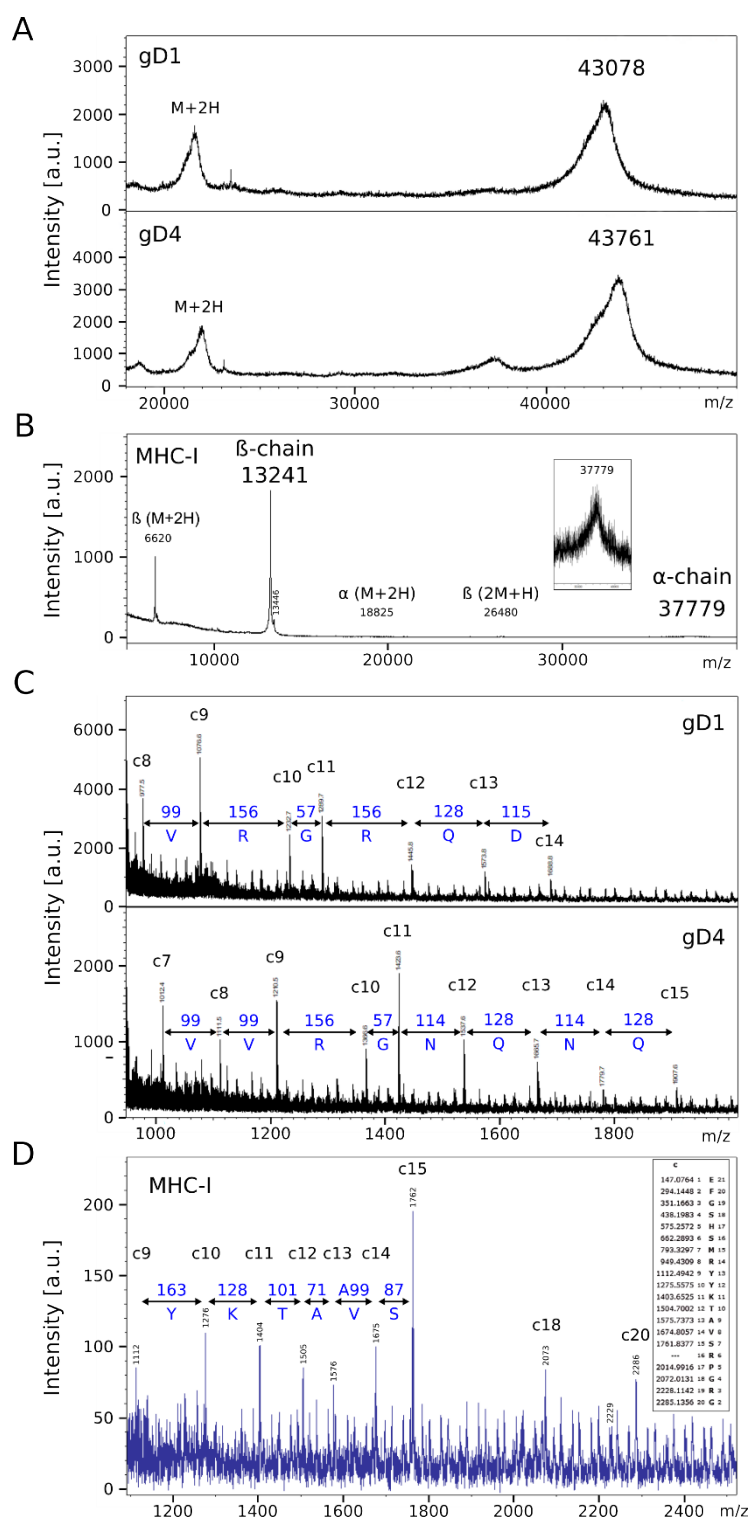
Figure 1: Mass spectrometric analysis of gD1 and gD4.

(A) Intact protein mass analysis of recombinant gD1 (top) and gD4 (bottom) including N-

terminal residues (EF: glutamic acid and phenylalanine) from *Eco*RI restriction site, TEV

1017   cleavage site and His_6-tag on sinapinic acid (SA) matrix. (B) Intact protein mass analysis of

1018   recombinant MHC-I complex comprised of β2m and MHC-I-α-chain (insert zoom) including

1019   the same additional residues as gD1 and gD4. (C) In-source decay (ISD) spectra of

1020   recombinant gD1 (top), gD4 (bottom), and (D) MHC-I to ascertain the correct N-termini,

1021   insert: Theoretical c-ion series including the N-terminal EF extension.

1022

48

Figure 2: Crystal structure comparison of gD1 and gD4.

Cartoon representation of (A) gD1 (2.45 Å resolution, PDB-ID: 6SQJ) and (B) gD4 (1.9 Å resolution, PDB-ID: 6TM8) monomer crystal structures. Molecule orientation is identical and secondary structures were assigned with dssp [72]. Helices are displayed in red, sheets in blue, and loops in green. N-acetylglucosamine (GlcNAc) and glycerol (GOL) molecules are shown in stick representation in beige. (C) Superposition of the crystal structures of gD1 (blue, PDB-ID 6SQJ) and gD4 (red, PDB-ID 6TM8). GlcNAc and glycerol molecules are not shown.

1032

Figure 3: Glycoprotein D from alphaherpesviruses have a similar crystal and secondary structure.

Superposition of crystal structures in cartoon representation of gD from EHV-1 (blue, PDB-ID 6SQJ) with (A) HSV-1 (orange, PDB-ID 2C36), (B) PrV (green, PDB-ID 5X5V), and (C) BoHV-1 (brown, PDB-ID 6LS9) gD. Main differences in global fold are highlighted. (D) Comparison of secondary structure elements of EHV-1/4, HSV-1/2, PrV, and BoHV-1 gD. Main differences in global fold are encircled.

1033

1034
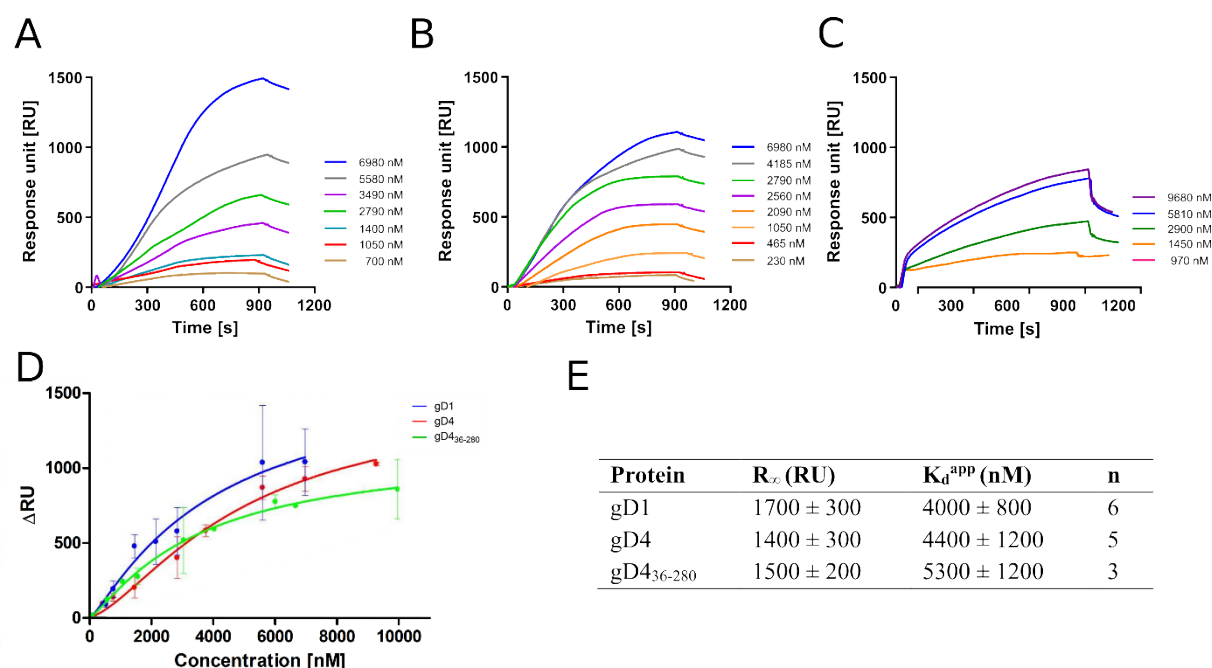
Figure 4: Binding affininties of gD1, gD4 and gD4$_{36-280}$ to the entry receptor MHC-I are in micro molar range.

(A, B, C) Representative SPR sensorgram profiles of recombinant gDs binding to amine-coupled recombinant MHC-I. Data were collected for several independent experiments [(A) gD1 n=6, (B) gD4 n=5, (C) gD4$_{36-280}$ n=3). (D) Binding curves for different gD concentrations from at least three independent experiments. Displayed are means with standard deviation (SD). The solid lines represent a fit of a Hill-Wand model to the data. (E) Parameters obtained from SPR binding curves of gD1, gD4, and gD436-280. R∞ is the maximum signal obtained from the bound protein; K$_d^{app}$ is the apparent equilibrium dissociation constant, n corresponds to the number of independent experiments.

51

1048
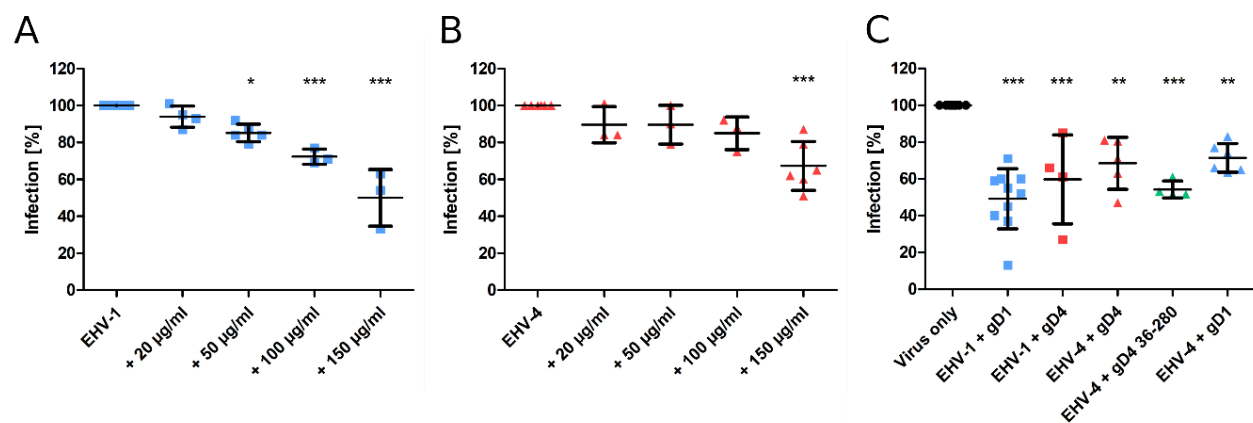
Figure 5: Recombinant gD1, gD4 and gD4$_{36-280}$ blocks EHV-1 and EHV-4 infection in ED cells.

(A) EHV-1 and (B) EHV-4 virus entry into ED cells blocked by different concentrations of gD1 and 4, respectively, and analyzed by flow cytometry. Cells were incubated with soluble proteins for 1 h on ice and infected with either EHV-1 or EHV-4 at MOI = 0.1. After 1 h, viruses on the cell surface were removed with citrate buffer and GFP levels were analyzed after 24-48 h by flow cytometry. (C) Plaque reduction assay of EHV-1 and EHV-4 with recombinant protein. ED cells were incubated for 1 h on ice with 150 μg/ml gD1, gD4 or gD4$_{36-280}$ and infected with 100 PFU of each virus. After 1 h, viruses on the cell surface were removed with citrate buffer and cells were overlaid with methylcellulose. GFP plaques were counted after 48 h. The experiment was repeated independently three times for each protein. Plaque numbers were normalized to infection levels without recombinant proteins. Statistical analysis was done using one-way ANOVA Bonferroni's multiple comparison test, * indicates P≤0.05, ** indicates P≤0.01, *** indicates P ≤0.001. Error bars represent mean with SD.
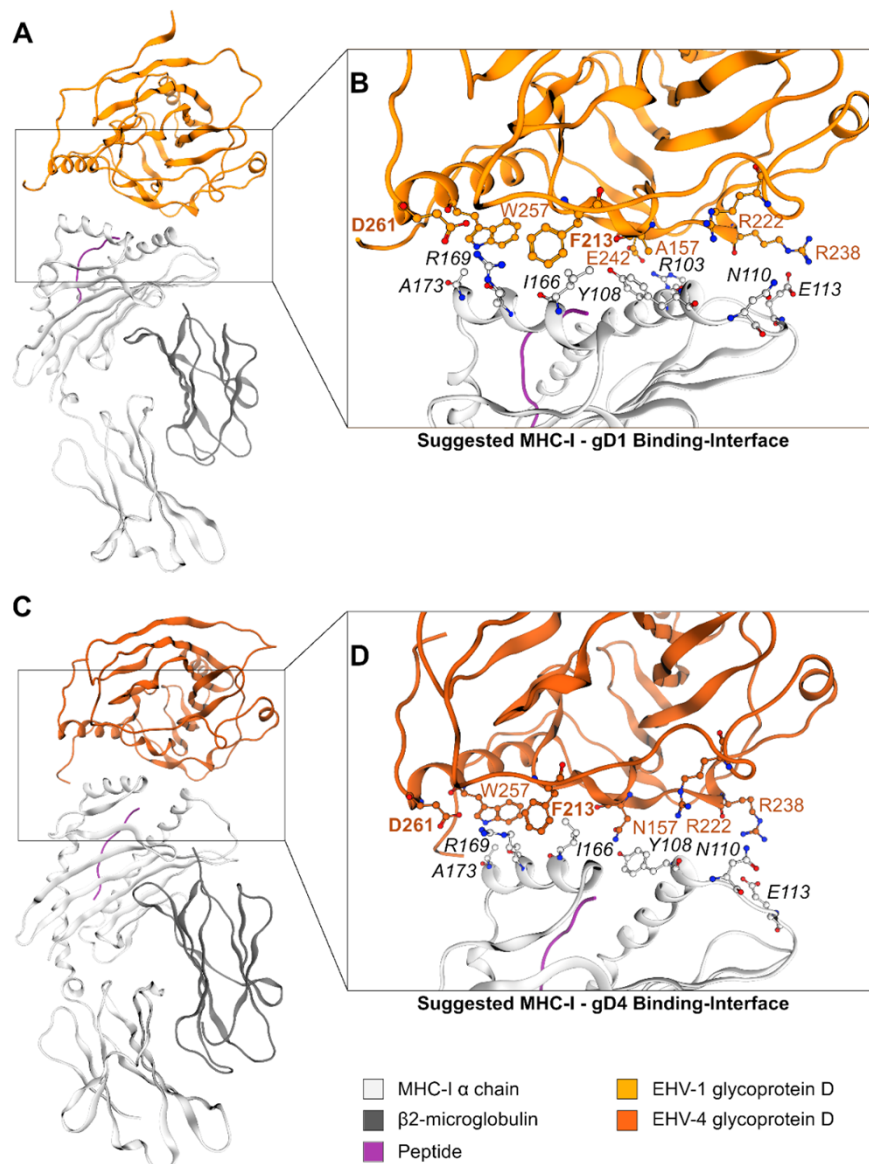
1049

1050

Figure 6: The gD1- and gD4-MHC-I interface.

(A) Suggested model of the MHC-I – gD1 complex and (B) detailed view on the hypothesized binding interface. (C) Suggested model of the MHC-I – gD4 complex and (D) detailed view on the hypothesized binding interface. EHV gD residues are highlighted in orange and hot spot residues additionally in bold font. Color-code: grey ribbon – MHC-I, dark grey ribbon – β2m, orange ribbon – gD1, dark orange ribbon – gD4, purple ribbon – peptide, grey/orange balls – carbon atoms, blue balls – nitrogen atoms, red balls – oxygen atoms.
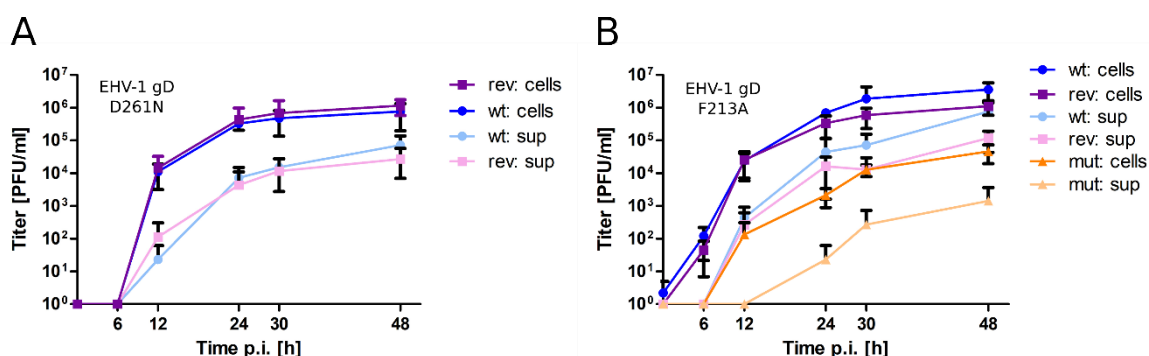
1051

Figure 7: Mutating gD1 residues D261N and F213A impairs EHV-1 growth in ED cells.

Multi-step growth kinetics of EHV-1 parental virus and gD mutants. ED cells were infected with an MOI of 0.01, cells and supernatant were collected separately at indicated time points post infection and titrated on ED cells. Shown are means with standard deviation (SD) of three independent experiments. (A) EHV-1 parental virus (blue colors) and EHV-1-gD$_{D261N}$ (violet colors). (B) EHV-1 parental virus (blue colors), EHV-1-gD$_{F213A}$ (orange colors) and EHV-1-gD-$_{F213A}$ (violet colors).
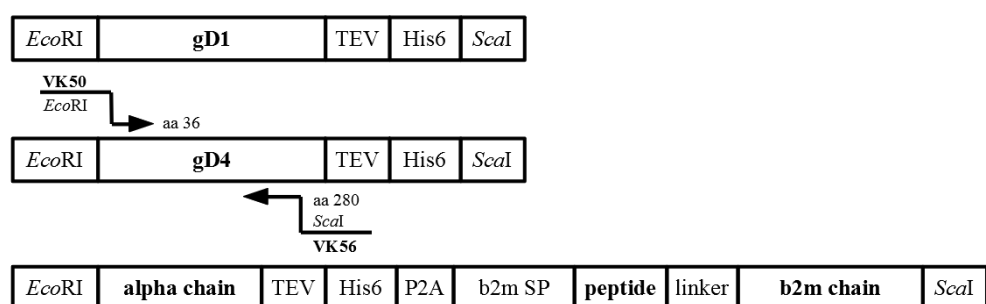
Figure 8: Synthetic genes used for cloning.

Schematic representation of synthetic genes for protein production of gD1, gD4, equine MHC-I 3.1 with cloning strategy for gD436-280.

1055

Supporting information

Figure S 1: SDS-PAGE and western blot of MHC-I, gD1, and gD4. (A) Coomassie stained SDS-PAGE on 12% gel. For MHC-I, only the α-chain is visible in (A) and (B). M = marker. (B) Western blot: MHC-I (50 µg/ml), gD1 (5 µg/ml), and gD4 (5 µg/ml) detected with 1:1000 rabbit anti-His$_6$ antibody and 1:10000 goat anti-rabbit-HRP as secondary antibody.

Figure S 2: Protein purification by size exclusion chromatography (SEC). Representative SEC curves of concentrated (A) gD1, (B) gD4, (C) gD4$_{36-280}$, and (D) MHC-I run on Superdex 200 16/600 after purification through immobilized metal ion affinity chromatography (IMAC). Solid curves show UV absorbance at 280nm and the dotted curves at 260nm.

Figure S 3: SDS-PAGE of protein purification by size exclusion chromatography (SEC). Representative size exclusion chromatography (SEC) fractions of proteins produced in insect cells on Coomassie stained 12% sodium dodecyl sulfate (SDS) gels. (A) gD1 (approximately 43 kDa), (B) gD4 (approximately 43 kDa), (C) gD4$_{36-280}$ (approximately 30 kDa), and (D) MHC-I (comprised of α-chain with an approximate size of 38 kDa and β2m (with the linker and peptide) with an approximate size of 13 kDa). M = marker, FT = flow through affinity chromatography, E = elution affinity chromatography, L = loaded on SEC column.

Figure S 4: Tandem mass spectrometry (MS/MS) of in-gel digested gD1 and gD4. Exemplary spectra are shown which confirm the identity of the analyzed proteins and the integrity of their termini. The inserts display the theoretical b and y fragments ions. (A) N-terminal peptide of gD1 generated by cleavage with Asp-N endoproteinase (M+H= 1574.90, pos. 001-013, sequence EFEKAKRAVRGRQ.D); (B) N-terminal peptide of gD4 obtained by trypsin cleavage (M+H=1013.52, pos. 001-007, sequence EFENYRR); (C) C-terminal peptide of gD4 including the His$_6$-tag, generated by Glu-C endoproteinase (M+H=1563.74, pos. 323-334, sequence ENLYFQG-H$_6$).

1081 **Figure S 5: *S*ize exclusion chromatography (SEC) combined with multi-angle static light**

1082 **scattering (MALS) analysis of gD1.**

1083 The gD1 crystal structure consists of a homodimer with two ions interpreted as magnesium

1084 originating from the crystallization solution, trapped between them. The ionic interaction

1085 together with a high Complex Formation Significance Score of 0.765 (PDB Proteins,

1086 Interfaces, Structures and Assemblies (PISA) server www.ebi.ac.uk/pdbe/pisa/) suggested that

1087 gD1 might form a dimer on the virus envelope as has been proposed for HSV-1 gD [26]. To

1088 evaluate whether recombinant gD of EHV-1 has a homodimeric and/or monomeric form in

1089 solution, molecular mass calculation based on SEC-MALS analysis was performed for gD1.

1090 Green curve represents the normalized refractive index trace (intensity, right y-axis) for gD1

1091 eluted from a Superdex 200 10/300 column. Blue line under the peak corresponds to the

1092 averaged molecular mass distribution (left y axis) across the peak. Exclusively the monomeric

1093 form with an approximate molecular weight of 44 kDa was detected. It can be concluded that

1094 gD1 is a monomer in solution.

1095 **Figure S 6: Sequence alignment of EHV-1 with HSV-1 and PrV.** Sequence alignment

1096 based on secondary structures of gD1 with (A) HSV-1 (PDB ID 2C3A) and (B) PrV (PDB ID

1097 5X5V) gD according to dssp [72]. Sheets are indicated as pink arrows, helices as blue

1098 cylinder, disulfide bonds as yellow boxes, glycosylation sites in gD1 as green dots, and

1099 magnesium coordinating residues in gD1 as purple dots. Labels correspond to the naming

1100 scheme presented by Li et al. [29].

1101 **Figure S 7: Ramachandran plots for modeled MHC-I.** Ramachandran plots for (A) MHC-I

1102 (gene 3.1) and (B) equine β2m. Symbol code: green point-residue with favorable geometry,

1103 yellow point- residue with allowed geometry.