

# **Supplementary Information**

**Molecular basis of EphA2 recognition by gHgL from gammaherpesviruses**

**Chao Su, et al**

Chao Su, Lili Wu, Yan Chai, Jianxun Qi, Shuguang Tan, George F. Gao, Hao Song, Jinghua Yan

**Supplementary Table 1. Data collection and refinement statistics**

	KSHV gHgL-LBD	EBV gHgL-LBD
<b>Data collection</b>		
Space group	<i>P212121</i>	<i>P1</i>
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	80.80, 153.20, 275.23	94.46, 113.81, 119.65
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 90.00, 90.00	90.03, 90.17, 89.89
Resolution (Å)	50.00-3.20 (3.31-3.20) <sup>a</sup>	50.00-3.00 (3.11-3.00) <sup>a</sup>
Unique reflections	57447	98531
<i>R</i> <sub>merge</sub>	0.134 (2.918) <sup>a</sup>	0.086 (1.102) <sup>a</sup>
<i>R</i> <sub>pim</sub>	0.038 (0.813) <sup>a</sup>	0.054 (0.676) <sup>a</sup>
<i>I</i> / $\sigma$ <i>I</i>	18.3 (0.989) <sup>a</sup>	12.950 (1.025) <sup>a</sup>
<i>CC</i> <sub>1/2</sub>	0.997 (0.527) <sup>a</sup>	0.994 (0.757) <sup>a</sup>
Completeness (%)	99.6 (100) <sup>a</sup>	99.0 (98.6) <sup>a</sup>
Redundancy	12.9 (13.1) <sup>a</sup>	3.5 (3.6) <sup>a</sup>
<b>Refinement</b>		
Resolution (Å)	31.65-3.20	41.22-3.00
No. reflections	45163	98231
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.2393/0.2721	0.2587/0.2888
No. atoms		
Protein	14513	29096
Ligand/ion	388	224
Water	0	0
<i>B</i> -factors		
Protein	66.67	121.51
Ligand/ion	115.18	142.95
Water		
R.m.s. deviations		
Bond lengths (Å)	0.002	0.005
Bond angles (°)	0.477	1.08
Ramachandran plot		
Favored (%)	99.89	99.40
Allowed (%)	0.11	0.54
Outliers (%)	0.00	0.05

<sup>a</sup> Values in parentheses are for highest-resolution shell.

**Supplementary Table 2. Interaction between LBD and EBV gHgL or KSHV gHgL**

			<b>Contacts<sup>a</sup></b>	<b>LBD</b>	<b>Total contacts</b>
EBV	gH	E30	1	P109	28
		G31	1, 1, 7	S107, F108, P109	
		H32	1, 10, 2, 5	V72, M73, F108, P109	
	gL	W24	2, 3, 3, 6, 1, 2	D53, M55, C70, R103 (1) <sup>b</sup> , F108, C188	149
		A25	2, 4, 3, 1	S68, C70, C188, V189	
		Y26	2	N57	
		P27	8	R103	
		T32	1	I58	
		L34	3	I58	
		H38	3, 4	I58, P63	
		T55	4	R103	
		K68	2	D53 (1)	
		N72	1, 1, 2, 4	E40, L41, G42, L54	
		Q73	2,1	Y48, L54	
		L74	1, 17, 1, 1	E40, L54, M55, Y65	
		V75	8, 8, 3	L54 (1), M55, Q56 (1)	
		I76	6	Q56	
		S77	2, 10, 10	M55, Q56 (1), N57	
		R78	4, 4, 1	Q56 (1), I58, P63	
		R130	1	V161	
A132	6, 1	M59, N60			
KSHV	gH	E52	7	R103	27
		F53	8, 3	R103, S107	
		N54	2	P109	
		FUC	7	P109	
	gL	Q20 <sup>c</sup>	1, 13	C70, N71	290
		Y21	11, 1, 3, 3, 19, 2	C70, V72, M73, R103, F108, P109	
		V22	6, 6, 2	C70, R103 (1), C188	
		A23	2, 4, 3, 5, 2, 3	S68, C70, R103, C188, V189, A190	
		L24	1, 2, 3, 1	N57, M66, T101, A190	
		P25	9	R103	
		A28	1	M59	
		Q30	11, 7, 6, 11	I58, M59, N60 (1), D61 (1)	
		A31	7	D61	
		S32	4	D61 (1)	
		K61	2	N57	
		R63	5	M55	
		D68	1	L54	
		I69	5, 2, 3, 2	L54, M55, Q56, Y65	
		T70	6, 12, 5	L54 (1), M55, Q56 (1)	
		V71	12	Q56	
		E72	5, 6, 9, 1	M55, Q56 (1), N57 (1), I58	
		H127	3	M59	
		N128	14, 15	M59, N60(2)	
		V129	8, 4	M59, L163	
		N130	8, 11, 2	M59, N60, I64	

<sup>a</sup> Numbers represent the number of atom-to-atom contacts between the gHgL residues and LBD residues, which were analyzed by the contact program in CCP4 suite (the distance cutoff is 4.5 Å).

<sup>b</sup> Numbers in the parentheses represent the number of potential hydrogen bonds between the gHgL residues and LBD residues.

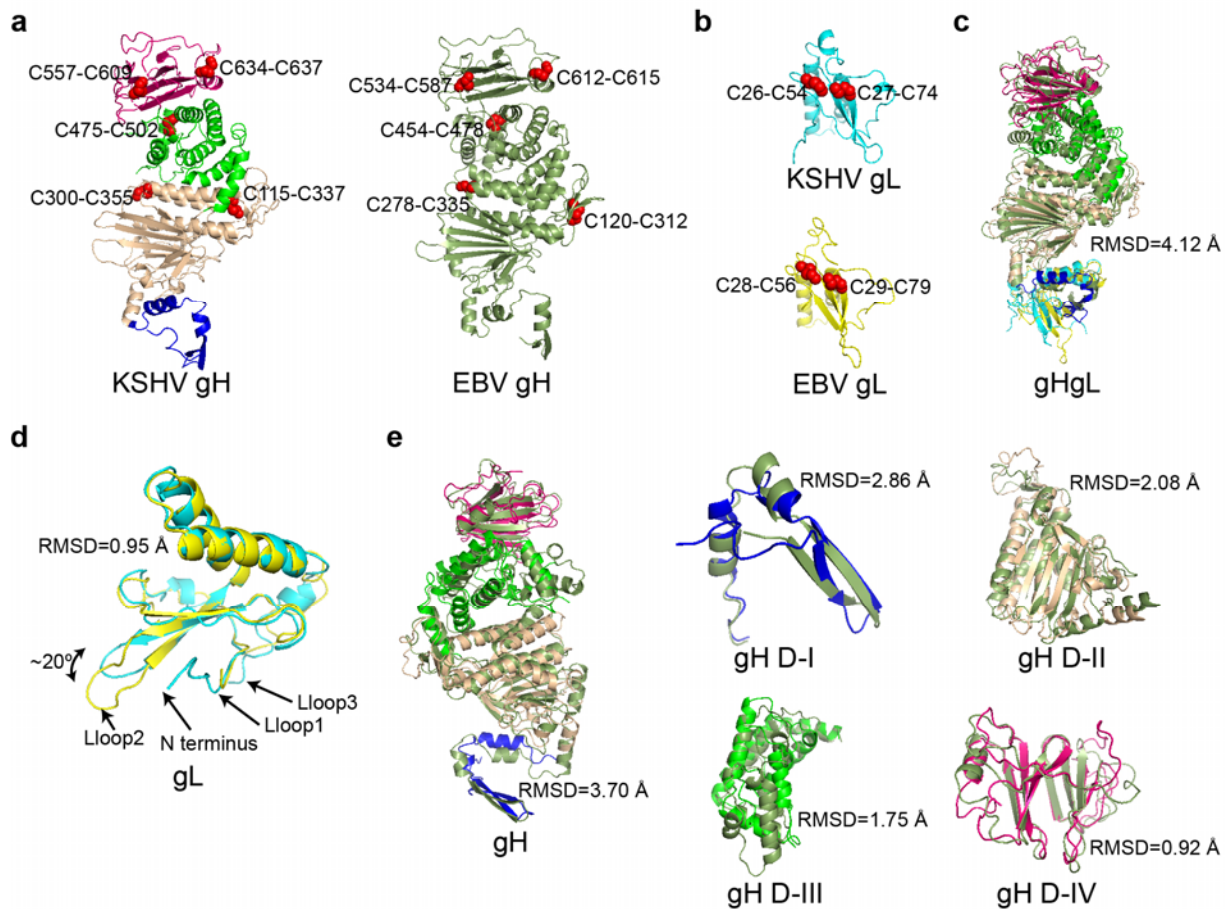
<sup>c</sup> protein-expression vector-introduced extra residue.

**Supplementary Table 3. The accession numbers for gH and gL proteins used for alignment**

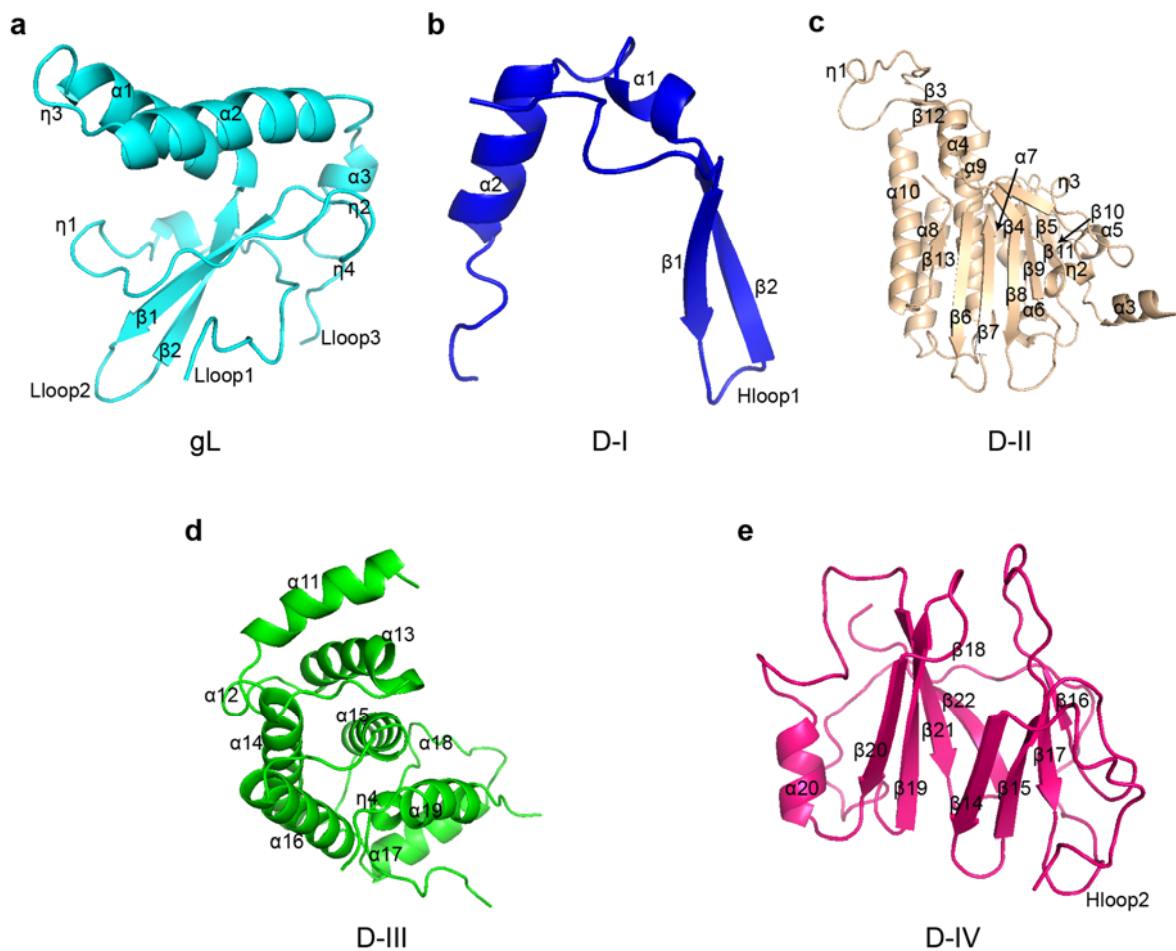
<b>Organism</b>	<b>Abbreviation</b>	<b>Accession No. (gH)</b>	<b>Accession No. (gL)</b>	<b>Accession No. (gB)</b>
<b>Kaposi's sarcoma-associated herpesvirus</b>	KSHV	YP_001129375.1	YP_001129399.1	YP_001129354.1
<b>Epstein-Barr virus</b>	EBV	YP_401700.1	YP_401678.1	YP_401713.1
<b>Rhesus monkey rhadinovirus</b>	RRV	AAF60000.1	AAF60026.1	
<b>Alcelaphine gammaherpesvirus 1</b>	AIHV-1	NP_065521.1	NP_065545.1	NP_065511.1
<b>Alcelaphine gammaherpesvirus 2</b>	AIHV-2	YP_009044405.1	YP_009044429.1	
<b>Bovine gammaherpesvirus 6</b>	BoHV-6	YP_009042001.1	YP_009042024.1	
<b>Ovine gammaherpesvirus 2</b>	OvHV-2	YP_438146.1	YP_438169.1	
<b>Porcine lymphotropic herpesvirus 1</b>	PLHV-1	YP_009505348.1	YP_009505372.1	
<b>Ateline gammaherpesvirus 3</b>	AtHV-3	NP_047993.1	NP_048020.1	
<b>Bovine gammaherpesvirus 4</b>	BoHV-4	NP_076514.1	NP_076539.1	
<b>Cricetid gammaherpesvirus 2</b>	CrGHV-2	YP_004207858	YP_004207883.1	
<b>Retroperitoneal fibromatosis-associated herpesvirus</b>	RFHVMm	AGY30705.1	AGY30728.1	
<b>Macaca fuscata rhadinovirus</b>	MFRV	AAT00015.1	AAT00055.1	
<b>Macaca nemestrina rhadinovirus 2</b>	MNRV-2	AJE29663.1	AJE29691.1	
<b>Murid gammaherpesvirus 4</b>	MuHV-4	NP_044860.1	AAF19311.1	NP_044848.3
<b>Wood mouse herpesvirus</b>	WMHV	ACY41096.1	ACY41118.1	
<b>Saimiriine gammaherpesvirus 2</b>	SaHV-2	NP_040224.1	NP_040249.1	
<b>Macacine gammaherpesvirus 5</b>	MaHV-5	NP_570762.1	NP_570788.1	
<b>Common bottlenose dolphin gammaherpesvirus 1 strain Sarasota</b>	CBDHV-1	YP_009388523.1	YP_009388547.1	
<b>Callitrichine gammaherpesvirus 3</b>	CaHV-3	NP_733867.1	NP_733891.1	
<b>Rhesus lymphocryptovirus</b>	rhLCV	YP_067997.1	YP_067974.1	
<b>Lymphocryptovirus Macaca/pfe-lcl-E3</b>	LCVMa	ALF03266.1	ALF03246.1	
<b>Equid gammaherpesvirus 2</b>	EHV-2	NP_042618.1	NP_042644.1	NP_042604.1
<b>Equid gammaherpesvirus 5</b>	EHV-5	YP_009118412.1	YP_009118436.1	
<b>Felis catus gammaherpesvirus 1</b>	FcaGHV1	YP_009173898.1	YP_009173923.1	
<b>Harp seal herpesvirus</b>	HSHV	AJG42948.1	AJG42971.1	
<b>Eptesicus fuscus gammaherpesvirus</b>	EfHV	YP_009552485.1	YP_009552510.1	
<b>Rhinolophus gammaherpesvirus 1</b>	RGHV-1	YP_009551832.1	YP_009551856.1	
<b>Myotis gammaherpesvirus 8</b>	MyHV-8	YP_009229856.1	YP_009229879.1	
<b>Human herpesvirus 1</b>	HSV-1	YP_009137096.1	YP_009137075.1	

<b>Human herpesvirus 2</b>	HSV-2	YP_009137173.1	YP_009137152.1	YP_009137179.1
<b>Varicella-zoster virus</b>	VZV	NP_040160.1	NP_040182.1	
<b>Human cytomegalovirus</b>	HCMV	YP_081523.1	YP_081555.1	
<b>Human herpesvirus 6a</b>	HHV-6A	NP_042941.1	NP_042975.1	
<b>Human herpesvirus 6b</b>	HHV-6B	NP_050229.1	NP_050261.1	
<b>Human herpesvirus 7</b>	HHV-7	YP_073788.1	YP_073820.1	

## Supplementary Figure

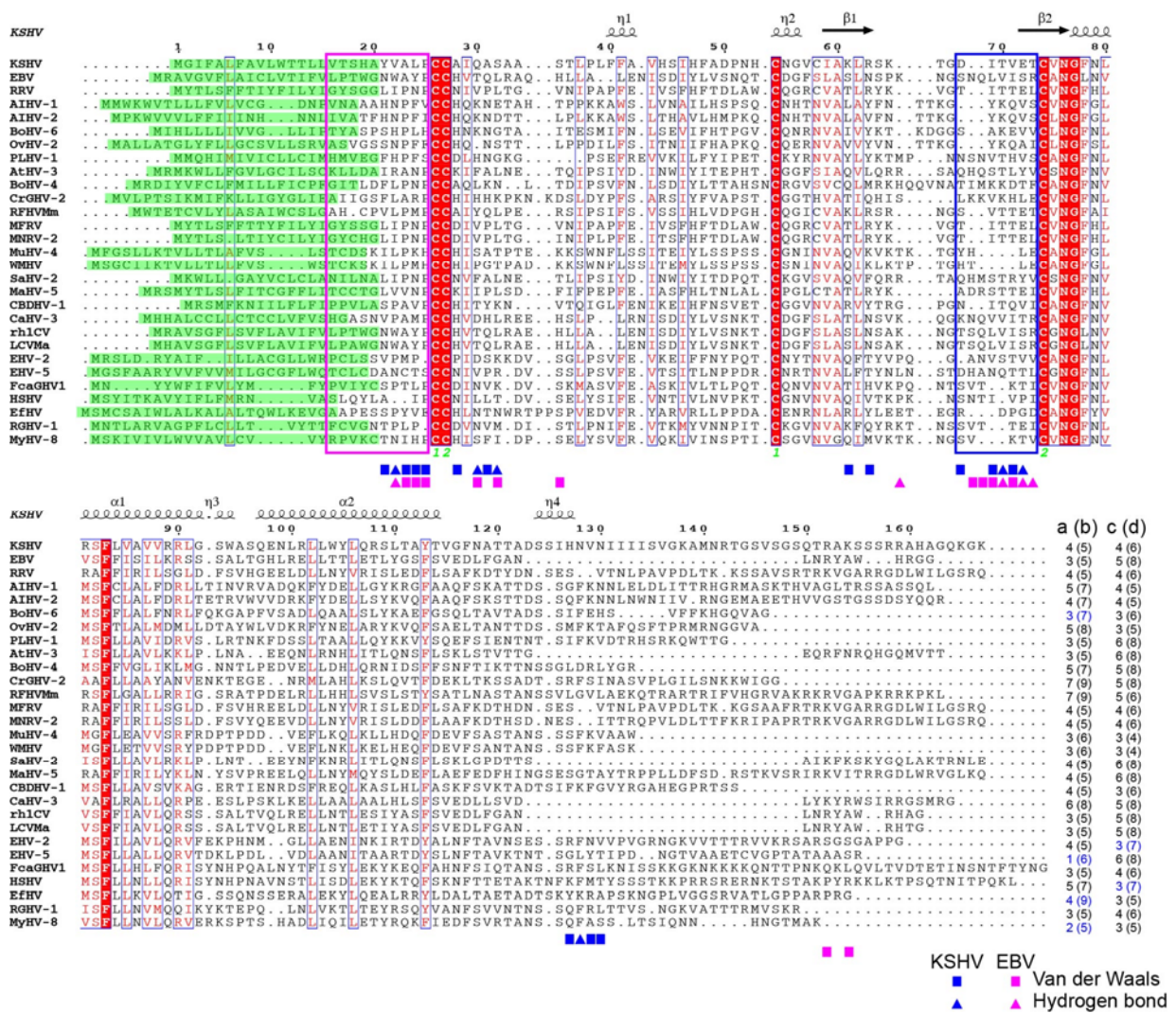


**Supplementary Figure 1. Structural comparison between KSHV gHgL and EBV gHgL.** **a,b**, Cartoon diagram of gH (**a**) or gL (**b**) of KSHV and EBV. The disulfide bonds are shown. **c-e**, Superimposition of gHgL (**c**), gL (**d**), gH and its domains (**e**) of KSHV with that of EBV. KSHV gHgL and EBV gHgL are colored as Fig. 2a. RMSD values are labeled.



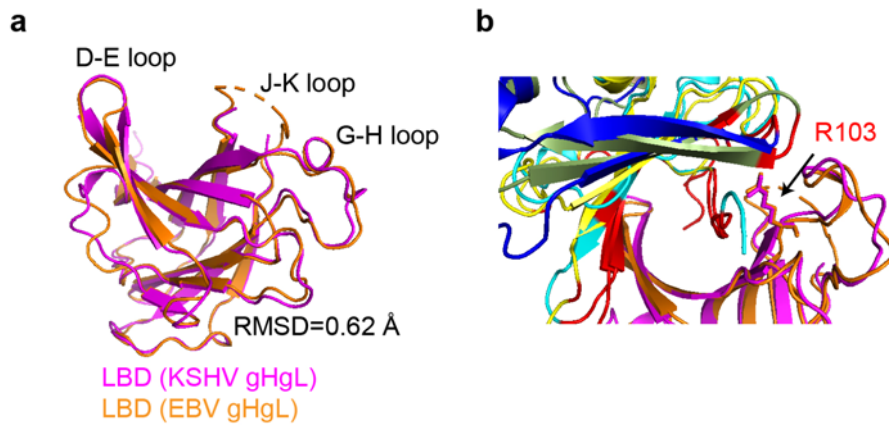
**Supplementary Figure 2. Cartoon diagram of domains of KSHV gH and gL structure. a-e,** The secondary structure elements are labeled. gH and gL loop shown are indicated with a preceding H or L, respectively. Ribbon diagram of gL (**a**), gH domain I (D-I) (**b**), gH domain II (D-II) (**c**), gH domain III (D-III) (**d**) and gH domain IV (D-IV) (**e**) are shown as in Fig. 2a.



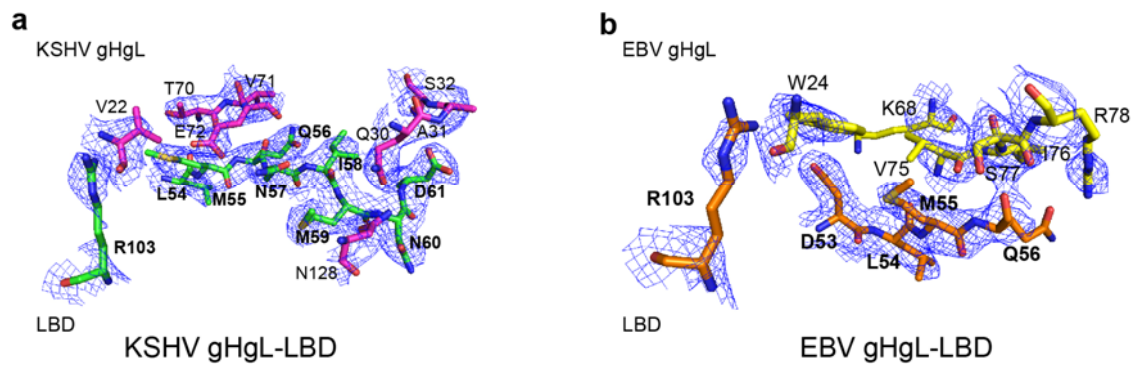


**Supplementary Figure 3. Sequence alignment of gL proteins from 29 representative  $\gamma$ -herpesviruses.** The blue or magenta triangles indicate LBD-binding amino acids via hydrogen bonds in KSHV gL or EBV gL, respectively. The blue or magenta squares indicate amino acids that contribute to binding to LBD via Van der Waals in KSHV gL or EBV gL, respectively. The predicted signal peptide produced by SignalP-5.0 Server<sup>1</sup> is shown by green rectangles. Magenta and blue rectangle represent two regions that bind to LBD, respectively. In the first region, b and a are represented the number of all residues and hydrophobic residues from the N-terminus to the first conserved cysteine residue, respectively. In the second region, d and c are represented the number of all residues and hydrophilic residues before the fourth conserved cysteine residue, respectively. The ratio of a to b or c to d less than 0.5 are highlighted in blue. Sequence alignment

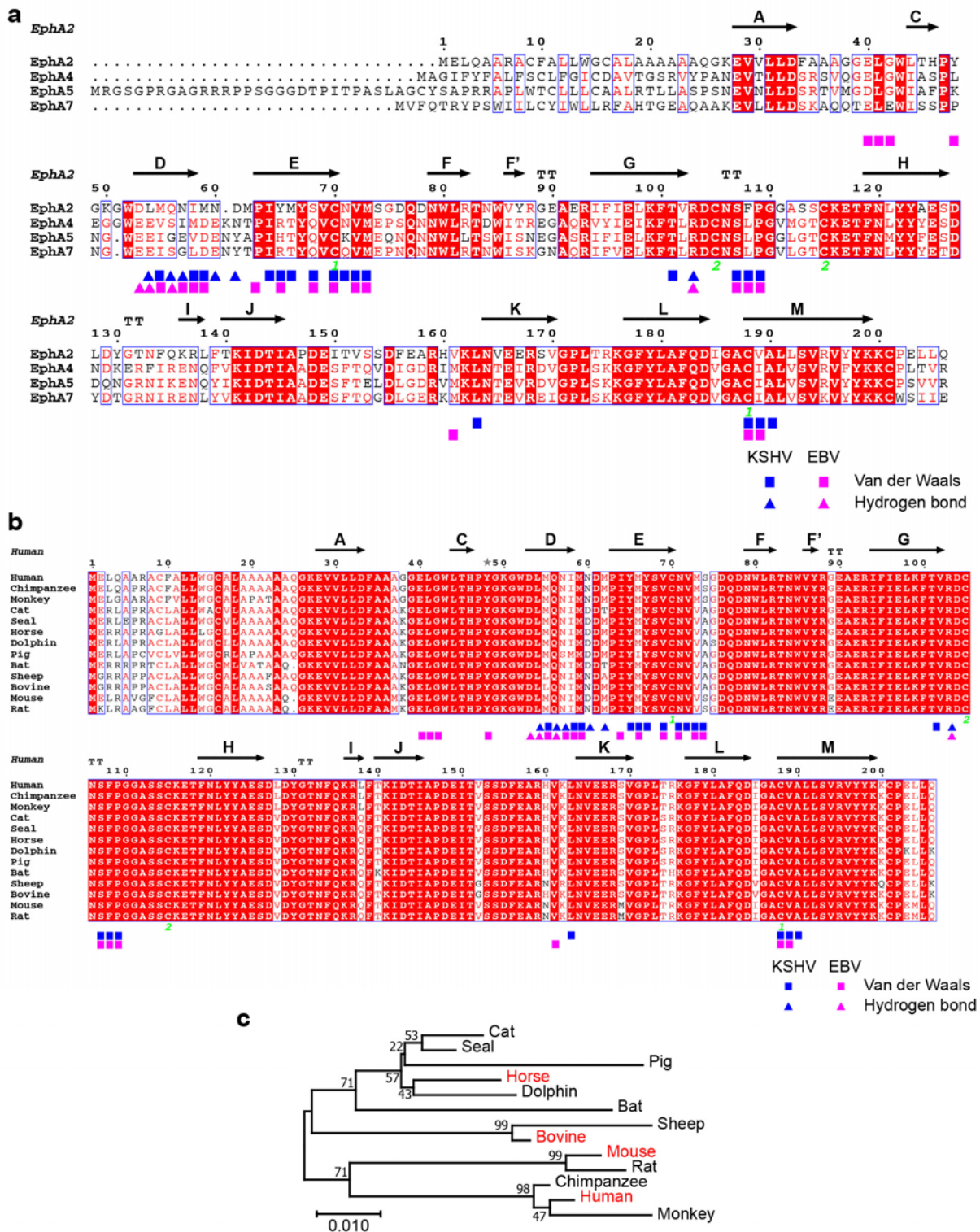
was produced by ESPript<sup>2</sup>. The GenBank accession codes are shown in the Supplementary Table 3.



**Supplementary Figure 4. Structural comparison between KSHV gHgL-LBD and EBV gHgL-LBD.** **a**, Superimposition of LBDs in KSHV gHgL-LBD and EBV gHgL-LBD. LBDs in EBV gHgL-LBD and KSHV gHgL-LBD are colored orange and magenta, respectively. D-E, G-H and J-K loop, and RMSD are shown. **b**, R103 in the channel of LBD is highlighted in stick.

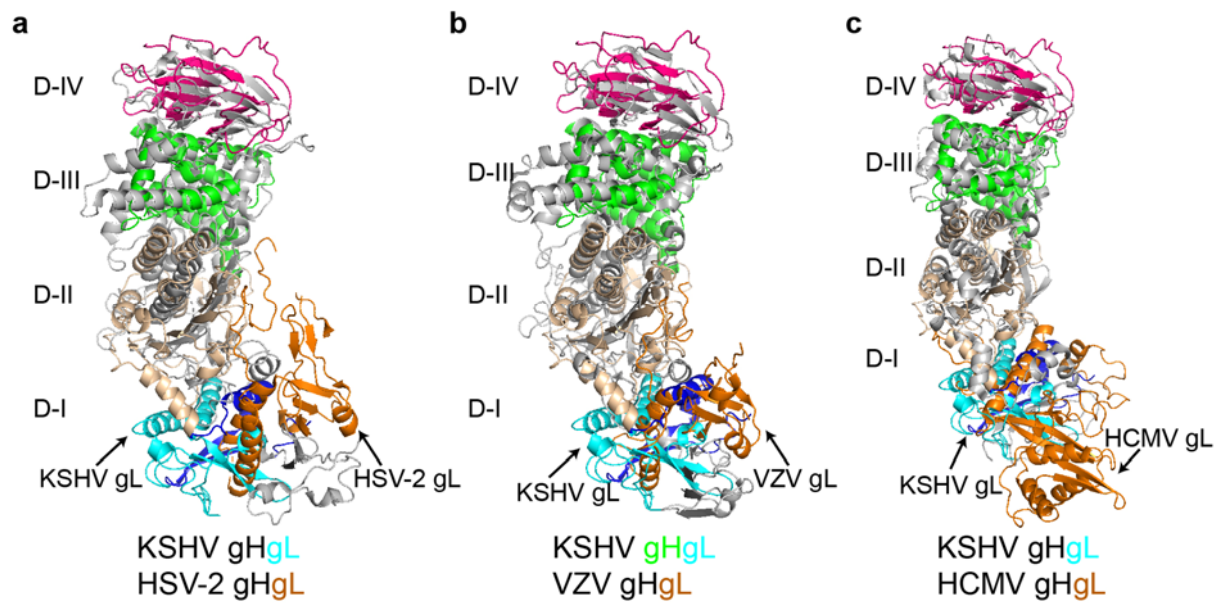


**Supplementary Figure 5. Representative electron density maps for the two complex structures. a,b,** The final 2Fo-Fc density maps of the complex structure of LBD bound to KSHV gHgL (**a**) or EBV gHgL (**b**) are drawn in blue mesh contoured at 1 sigma, respectively.



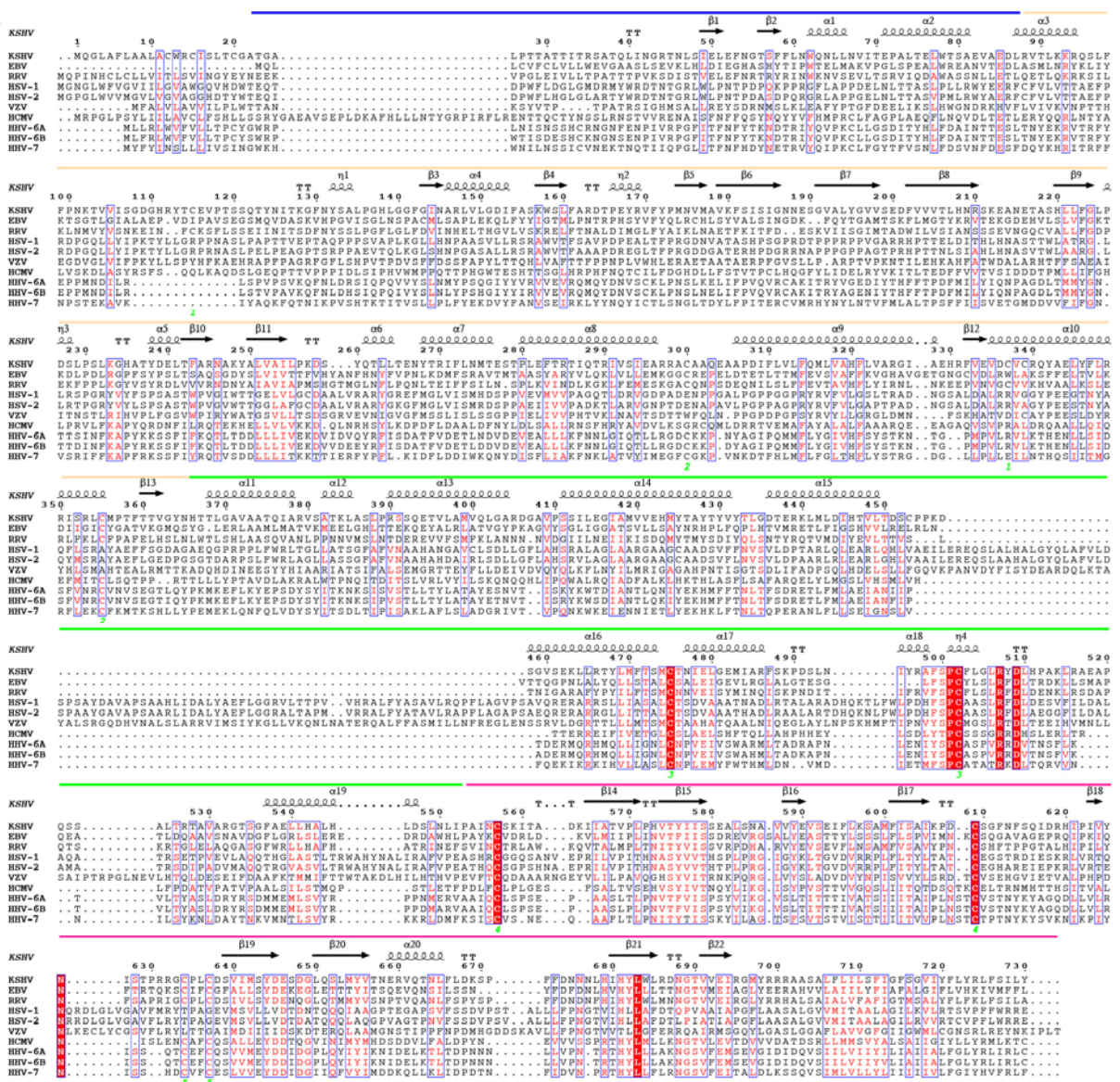
Supplementary Figure 6. Sequence alignment of EphAs. **a**, Sequence alignment of LBDs from human EphA2, EphA4, EphA5 and EphA7 proteins. The blue or magenta triangles indicate amino

acids in LBD that bind to KSHV gHgL or EBV gHgL via hydrogen bonds. The blue or magenta squares indicate amino acids in LBD that contribute to bind to KSHV gHgL or EBV gHgL via Van der Waals. Sequence alignment was produced by ESPript<sup>2</sup>. The GenBank accession codes: EphA4, AAH26327.1; EphA5, AAI43428.1; EphA7, NP\_001363394.1. **b**, Sequence alignment of EphA2 LBD from different mammals. The GenBank accession codes: Chimpanzee, XP\_016810290.1; Monkey, NP\_001035768.1; Cat, XP\_023113942.1; Seal, XP\_021539729.1; Horse, XP\_001488789.1; Dolphin, XP\_026965844.1; Pig, XP\_005665094.1; Bat, XP\_006105814.2; Sheep, XP\_027821486.1; Bovine, NP\_001192660.1; Mouse, NP\_034269.2; Rat, NP\_001102447.1. **c**, Phylogenetic tree analysis of EphA2 LBD from different mammals. The host species infected by the selected viruses including AIHV-1, EHV-2 and MuHV-4 for experiment are highlighted.

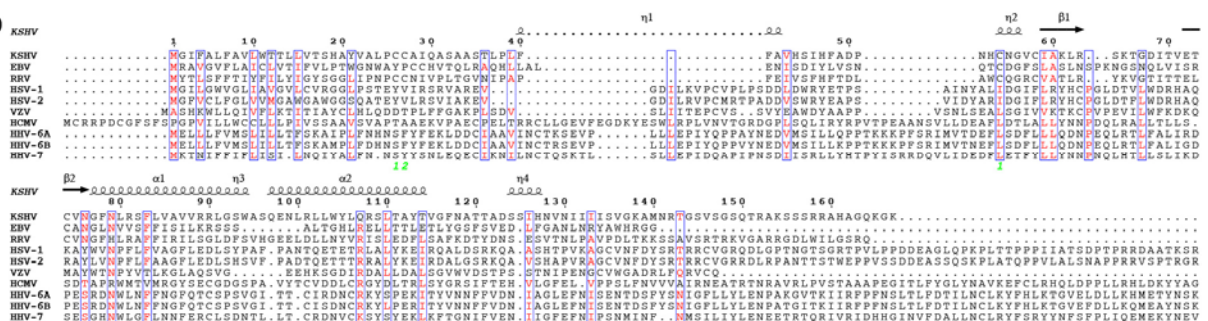


**Supplementary Figure 7. Structural comparison between KSHV gHgL and gHgL from  $\alpha$ - and  $\beta$ -herpesvirus.** a-c, Superimposition of KSHV gHgL and HSV-2 gHgL (PDB: 3M1C) (a) or VZV gHgL (PDB: 4XHJ) (b) or HCMV gHgL (PDB: 5VOB) (c). KSHV gH and gL are colored as in Fig. 2a. gH and gL of HSV-2, VZV and HCMV are colored gray and orange, respectively.

a



b





**Supplementary Figure 8. Sequence alignment of gHgL proteins from 10 representative herpesviruses.** **a,b**, gH (**a**) and gL (**b**) proteins from representative  $\alpha$ -,  $\beta$ - and  $\gamma$ -herpesvirus are included for alignment. Different colored lines are corresponding to the different domains of gH. Sequence alignment was produced by ESPript<sup>2</sup>. The GenBank accession codes are shown in the Supplementary Table 3.

**KSHV** 1 10 20 30 40

KSHV .....MOGLAPLADACRCISL.....TCGATGALPTTATTITRAGTALINGRTNLS  
 RFPVNm .....MIFVAMVLTALFIL.....GCQGE.....ONATNLP\_YENLS  
 RRV .....MQPINHCCLLVITLSVING.....YEYNEKVPGLIEVLLTTPATTPVKSIDST  
 MFRV .....MQPINHCCLLVITTSVING.....YEYNEKVPGLIEVLLTTPATTPVKSIDST  
 MNRV-2 .....MARISFIFPFTIIRCSVTDK.....VYVEEK.....SN  
 ALRV-3 .....MMIFSLFVHLILEA.....IDSVQLPRPLNEVPPADOLQMRNGYNTDL  
 SaHV-2 .....MTLQLFVHLNILEA.....LCDQLPKRINKPKAEERLRLNGYNTL  
 MNRV-4 .....MVTCLRNFLVLCIFVSM .....VHLVHGVDGRDSKEGATIKVRKNDNGVLS  
 MNRV .....MKTCLRNFLVLCIFVSM .....VCLVYGLLDKASKGATIKVRKNDNGVLS  
 CcGRV-2 .....MAPRCFVPLVFGAC .....FGLIGNLKAARKMPELNTLDS\_VK  
 ERV-2 .....MNRVAKLCPGIALFLCL .....CNMSFC .....QTRDAPHYATISKPPDRKDFALIKDQPMIT  
 ERV-5 .....MMKYLKACVTLVWVAG .....SVGTS .....PTTKSP .....PPTRTFTALIKNPSLT  
 RGHV-1 .....MMLKQPKLRLLSLLCLQICSVYGNATSTPTPPASPSPTPTSGNNTQTNVST .....TAKTLKTTLRPDPNPFNFQILKMPME  
 MNRV .....MNRVAKLCPGIALFLCL .....CNMSFC .....QTRDAPHYATISKPPDRKDFALIKDQPMIT  
 PcaGHV1 .....MSNIMNRTSPGICNLISMLVVMYLLS .....STLWCTVGTKTTTKR .....SPKATKNEKEMGALRHXYNFTELLARPERE  
 ALRV-1 .....MLFLLCLVCGAQA .....ITTPAPFRPATTFRRGVTS .....PLIVPASSSELI  
 ALRV-2 .....MLLLEDCAGATA .....VTPSPRPPPTFF .....VRLASSTSEI  
 OVRV-2 .....MPSVPHAPLLMHNPPAPGGL .....LLCFPTLISLVNWNAGORRTIS .....EHYAPSSALV  
 BoRV-6 .....MNSLMMPLALVLLAVTSNA .....FAPPGPVDSYTPSPRK .....LLPRSVTEK  
 MNRV .....MNRVAKLCPGIALFLCL .....CNMSFC .....QTRDAPHYATISKPPDRKDFALIKDQPMIT  
 rILCV .....MCLLVCVLCVLLCERGLASLS .....EYKLVH  
 LCVm .....MCLLVCVLCVLLDGGGLASLS .....EYKLVH  
 ERV-5 .....MNHFLPLANFLVFWGNDLES .....LSCVLSYCNVRGDLGKLEENLNK .....LT  
 CaHV-3 .....MKLFLIFLT .....PTLPLLLLLLSHAAALPPSPRPPTGGDGLV  
 EHV .....MGRSAPGLCGAPLTPSLVSTRTSPF .....PVVALRSAPFVSRRAVRGDKGLKLD  
 CBDRV-1 .....MEINKIQGKNCVACPSLCLAVFLAWS .....PVVALRSAPFVSRRAVRGDKGLKLD

**KSHV** β1 β2 α1 α2 α3 T...T η1 η2 β3 α4 β4 TT η2

50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200

KSHV .....RLEPNQTSPPFNQNLNVTEPAATLTSVSAEVAQDQVLRKLRKLPFN.K.TVVISGDRHYRCEVPSSTQYINIK.....GPNYSALPDRGGGNNARLLDGLPASKMSFARD.TPEYR  
 RFPVNm .....VALVNGTCSLNRSLRNCVTEPVTLEKTAHVQSGVTLQKKNLWVST.QSPLGNGLOYCGWTRQLKFNLA .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.QREY  
 RRV .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 MFRV .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 MNRV-2 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 MNRV-5 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 ALRV-3 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 SaHV-2 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 MNRV-4 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 MNRV .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 MNRV .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 CcGRV-2 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 ERV-2 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 ERV-5 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 RGHV-1 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 MNRV .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 PcaGHV1 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 ALRV-1 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 ALRV-2 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 OVRV-2 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 BoRV-6 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 FLRV-1 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 rILCV .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 LCVm .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 ERV-5 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 CaHV-3 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 EHV .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG  
 CBDRV-1 .....VLEFNRTNRKNNKNNSEVTSRVIQDAWASSKLLQDQELQKRRGLKLN.....MAYGNKINF.CKFSSEIINIS .....DNYSALPDRGGGNNARLLDGLPASKMSFARD.PATH.LDMG

**KSHV** β5 β6 β7 β8 β9 η3 TT α5 β10 β11 TT α6

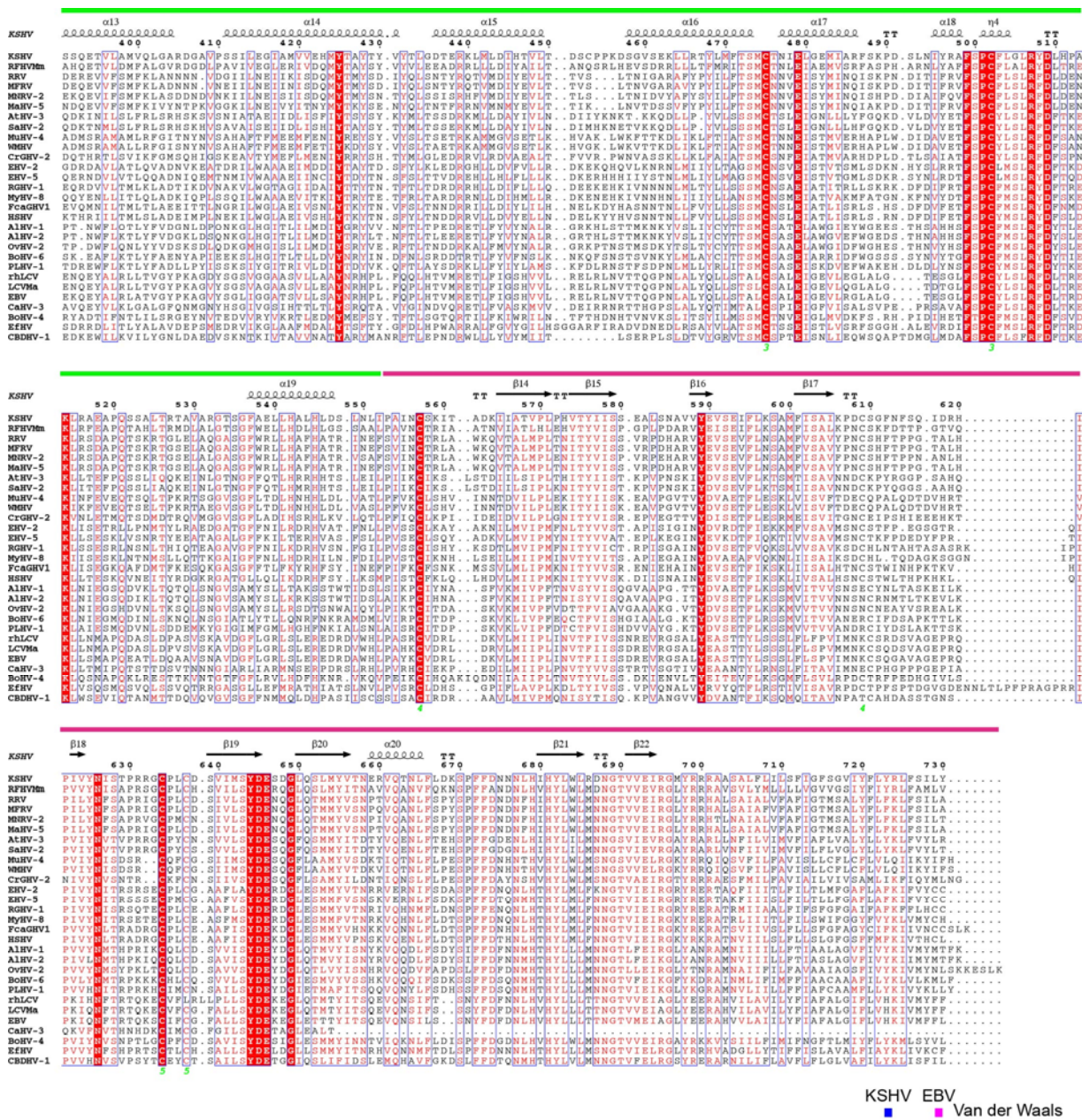
170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390

KSHV .....FFPNMVMVAKFSISIGNSSVALVYVSEVYVTHHRSKEA.....NETASHLLFLQPLDSLKHVHYD.EHTVAKAKLVAIILKDS.....YOTLLT.....E  
 RFPVNm .....AFYSNNLSALRMPAALGNEAPGRLGVVSEELVTLVNOTATG.....NDTALHMLFQPLRALSLAAGVLYG.EVTVLDANALVSLVAIPASA.FDQLG.....E  
 RRV .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 MFRV .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 MNRV-2 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 MNRV-5 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 ALRV-3 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 SaHV-2 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 MNRV-4 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 MNRV .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 MNRV .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 CcGRV-2 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 ERV-2 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 ERV-5 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 RGHV-1 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 MNRV .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 PcaGHV1 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 ALRV-1 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 ALRV-2 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 OVRV-2 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 BoRV-6 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 FLRV-1 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 rILCV .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 LCVm .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 ERV-5 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 CaHV-3 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 EHV .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q  
 CBDRV-1 .....LFYAKLNAETFKITFDE.SKVIISGIMTADMVLSIANSSEV.....NGQCVALLFDGPEPKLQVYSYR.DLVVVRNDNVAIAVAPMSHGTMLNPLP.....Q

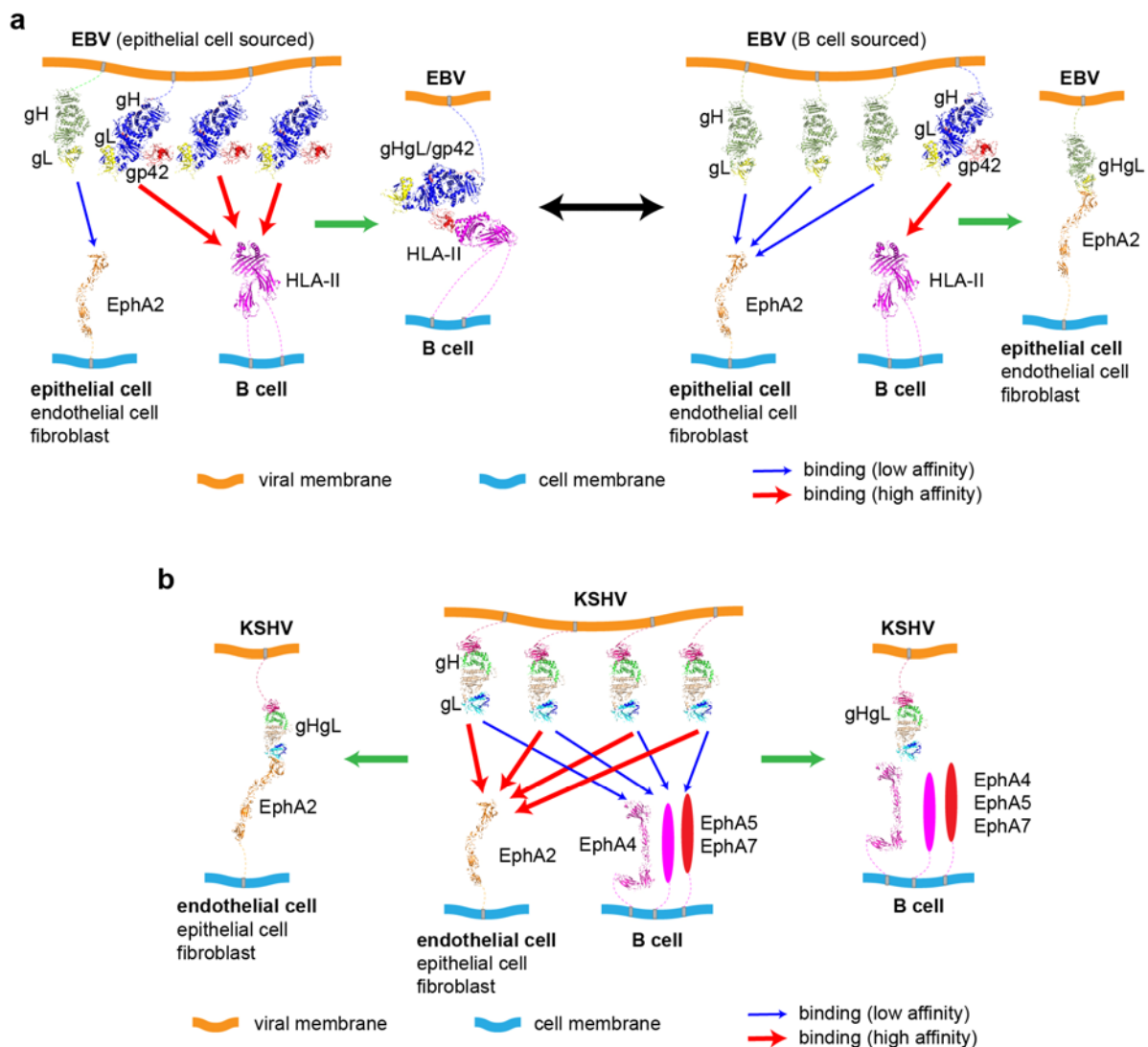
**KSHV** α7 α8 α9 β12 α10 β13 α11 α12

270 280 290 300 310 320 330 340 350 360 370 380 390

KSHV .....NYTRPLNVESTPLETETOTRIISARRCAAAQ.EAAPDIFLVLQMLVAHIAEHR.FVIVOCVQVAVLDFRRISRLXPTFTVGNYSHTLGAVALATDARVSAKTAASPR  
 RFPVNm .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 RRV .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 MFRV .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 MNRV-2 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 MNRV-5 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 ALRV-3 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 SaHV-2 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 MNRV-4 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 MNRV .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 MNRV .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 CcGRV-2 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 ERV-2 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 ERV-5 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 RGHV-1 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 MNRV .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 PcaGHV1 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 ALRV-1 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 ALRV-2 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 OVRV-2 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 BoRV-6 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 FLRV-1 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 rILCV .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 LCVm .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 ERV-5 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 CaHV-3 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 EHV .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL  
 CBDRV-1 .....NLTHPLRADDADDITRAGVGRMFGLSARRCAASAP.AAAMHILFFLFAALYVLPVGRREQA.FVNAVTCVQVAVLDFLRLGLDPLTSLVGOYGRSALRCLAAQIRSLPAGALATSL



Supplementary Figure 9. Sequence alignment of gH proteins from 29 representative  $\gamma$ -herpesviruses. The blue or magenta squares indicate the amino acids contributing to LBD interaction via Van der Waals from KSHV or EBV, respectively. Different colored lines are corresponding to the different domains of gH. Sequence alignment was produced by ESPrnt<sup>2</sup>. The GenBank accession codes are shown in the Supplementary Table 3.



**Supplementary Figure 10. Modes of EBV and KSHV binding to high- and low-affinity receptors for entry. a**, Epithelial cell-derived EBV virions contain more gHgL-gp42 complex than gHgL complex. Although gHgL complex can bind to low-affinity receptor EphA2 in the epithelial cell, the higher binding of gHgL-gp42 to high-affinity receptor HLA-II in the B cell make virions better infect B cells. Virions originating in B cells contain more gHgL complex due to the degradation of gp42 in the cells. Enough gHgL proteins binding to low-affinity receptor EphA2 enable virions to infect epithelial cells effectively. Therefore, EBV uses this strategy to alter and switch the tropism of virus<sup>3</sup>. The structural views of HLA-II (PDB: 1KG0), EBV gHgL-gp42 (PDB: 5T1D), EphA2 (PDB: 2X10) and the complexes of EBV gHgL-gp42-HLA-II and EBV gHgL-EphA2 were generated using PyMOL software. **b**, KSHV virions bind to the high-

affinity receptor EphA2 for entry into endothelial cells. In addition, KSHV could bind to the low-affinity receptors in B cell, such as EphA4, EphA5 and EphA7. Multiple types of low-affinity receptors in B cell make sure the infection efficiency of KSHV. The structural views of EphA2 (PDB: 2X10), EphA4 (PDB: 4M4P) and the complexes of KSHV gHgL-EphA2 were generated using PyMOL software.

#### **Supplementary Reference**

- 1 Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423, doi:10.1038/s41587-019-0036-z (2019).
- 2 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324, doi:10.1093/nar/gku316 (2014).
- 3 Borza, C. M. & Hutt-Fletcher, L. M. Alternate replication in B cells and epithelial cells switches tropism of Epstein-Barr virus. *Nature medicine* **8**, 594-599, doi:10.1038/nm0602-594 (2002).