Supporting Information

Yang and Bjorkman 10.1073/pnas.0804551105

SI Text

SI Materials and Methods. A His-tagged version of the HCMV AD169 UL18 ectodomain (residues 1–284 of the mature protein) was expressed together with human β2m in baculovirus-infected insect cells (Hi5 cells) as described (1). A nonameric UL18-binding peptide derived from actin (ALPHAILRL) was added to insect cell supernatants at a concentration of 2 mg/liter before purification of UL18/β2m heterodimers by nickel-NTA and gel-filtration chromatography. The D1–D2 region of LIR-1 (residues 1–198 of the mature protein) was expressed in bacteria and refolded from insoluble inclusion bodies as described (2). Purified UL18 and LIR-1 were mixed at an ≈1:1.1 molar ratio and passed over an S200 16/60 gel filtration column in 20 mM Tris pH 7.5, 150 mM NaCl. Small crystals of UL18/LIR-1 complex were obtained, but efforts to improve them to allow collection of data to atomic resolution were unsuccessful.

A systematic effort to improve the crystals by removal of potentially heterogeneous N-linked carbohydrates was initiated by expression of individual UL18 mutants to determine which sites were critical for UL18 expression. Ten of the 13 potential N-glycosylation sites on the UL18 ectodomain were removed one at a time (the three potential sites closest to the peptide-binding groove were retained) by using site-directed mutagenesis to replace the Asn codon of each Asn-Xaa-Ser/Thr site with a Gln codon. Three of the mutants (Asn-36 to Gln, Asn-147 to Gln, Asn-220 to Gln; denoted as N36Q, N147Q, and N220Q) maintained expression of UL18 at near wild-type levels (0.5–1 mg/liter). These mutations were combined to create a UL18 heavy chain with three substitutions (N36Q/N147Q/N220Q). The resulting UL18 mutant was partially aggregated due to covalent dimerization, as revealed by nonreducing SDS/PAGE analysis (data not shown). Computational modeling of UL18 using the coordinates of HLA-A2 predicted a free cysteine, Cys-259, near one of the altered glycosylation sites (N220Q), suggesting that the N-linked carbohydrate attached to Asn-220 normally shielded Cys-259 from oxidation. Mutation of this cysteine to serine (C259S) resulted in a glycosylation-reduced UL18 mutant (N36Q/N147Q/N220Q/C259S) that was not aggregated (data not shown).

Mutant UL18 (N36Q/N147Q/N220Q/C259S)/LIR-1 D1–D2 crystals (space group $P2_1$ with two complex molecules per asymmetric unit) were grown at room temperature by hanging-drop vapor diffusion by mixing 1 μl of protein (20 mg/ml) with 1 μl of reservoir solution. Crystals were obtained in 0.1 M sodium citrate (pH 5.5), 0.4 M $Mg(NO_3)_2$, and ≈16–22% (wt/vol) PEG 3350. Crystals were prepared for data collection by sequentially transferring into drops containing 5–25% glycerol in mother liquor and then cryopreserved in liquid nitrogen.

SI Discussion. Like the grooves of class I MHC molecules, the UL18 groove is bounded by two pockets that interact with conserved portions of the peptide N and C termini (pockets A and F, respectively) and four intermediate pockets that accommodate side chains from the rest of the peptide (pockets B-E) (Fig. S3 *A* and *B* and Table S3). As also found in class I MHC molecules, both ends of the UL18 groove are occluded, effectively limiting the length of most bound peptides to a size distribution mimicking class I-associated peptides (3). Occlusion of the ends also ensures that the N termini of all UL18-associated peptides interact with pocket A rather than extend out of the groove, as occurs for peptides bound to class II MHC molecules (4). Although the end result is similar, the structural mechanisms by which the groove ends are closed differ in UL18 and class I MHC molecules. For example, UL18 lacks the conserved salt bridge that blocks the pocket A (or left) side of the class I MHC groove by linking the α1 and α2 domain helices (class I Glu-55 and Arg-170 are replaced by UL18 Tyr-59 and Ile-180), but the left side of the UL18 groove is occluded by the side chains of Tyr-59 and Trp-177 (class I residues Glu-56 and Trp-167) (Fig. S3 *C* and *D*). The right side of the UL18 groove is blocked by a hydrogen-bond network involving Tyr-150 on the α2 domain helix (a residue in a region of UL18 that has no equivalent in class I molecules), Glu-83 on the α1 domain helix (class I Thr-80), and a water molecule. In class I MHC molecules, the conserved residue Tyr-84 on the α1 domain helix (Ala-87 in UL18) acts to close off the groove in much the same way as UL18 Tyr-150 (Fig. S3 *C* and *D*).

As observed for class I MHC/peptide complexes, UL18-bound peptides are tethered to the UL18 groove through interactions between main-chain atoms of the peptide N- and C-terminal residues and side chains in groove pockets A and F (Fig. S3*E*). Critical residues in UL18 pocket A are conserved with class I MHC molecules (Table S3), hence their roles in peptide binding are similar. Four UL18 tyrosines (5, 62, 169, and 181; corresponding to class I tyrosines 7, 59, 159, and 171), form hydrogen bonds directly or through a water molecule with main-chain atoms of the peptide N-terminal residues P1 and P2 (Fig. S3 *C* and *D* and Table S3). Two other conserved UL18 residues, Glu-66 (class I Glu-63) and Lys-69 (class I Lys-66), form salt bridge and hydrogen-bonding networks with each other and the amide nitrogen and carbonyl oxygen atoms of P2 (Fig. S3 *C–E*). At the other end of the groove, the components of the UL18 and HLA-A2 F pockets carry out similar functions despite sequence differences and insertions and deletions in UL18 compared with class I molecules (Fig. 2 and Table S3). In class I molecules, Tyr-84, Thr-143, Lys-146, and Trp-147 interact with the main-chain atoms of the peptide C terminus. The functionally equivalent residues in UL18 are Asn-80, which hydrogen bonds with the amide nitrogen of P9, and Lys-154, which forms a salt bridge with the carboxyl group of P9 (Fig. S3 *C* and *D*). In addition, the backbone of P7 is anchored through a hydrogen bond network involving the side chains of residues Gln-76 (to the carbonyl oxygen) and Glu-115 plus a water molecule (to the amide nitrogen).

The structure of the UL18/ALPHAILRL peptide complex can be used to rationalize UL18's peptide-binding preferences. Sequencing of peptides eluted from purified UL18 revealed predominantly leucine or methionine at P2, proline at P3, and a hydrophobic residue at P9 (3), a motif that resembles the peptide-binding motif of HLA-A2, which prefers leucine or isoleucine at P2 and a small hydrophobic residue (valine, alanine, or leucine) at P8/P9 (5). In the UL18 structure, the leucine side chain of P2 is buried in the predominantly hydrophobic pocket B, where it makes van der Waals interactions with pocket residues (Fig. S3*C*). Similar interactions were observed for P2 leucine or isoleucine residues with pocket B of HLA-A2. The P3 proline in the UL18-binding ALPHAILRL peptide is in a *trans* conformation with its ring pointing down toward the floor of the UL18 groove, where it is stabilized by interactions with UL18 Tyr-169 and His-100 (Fig. S3*C*). Similar to peptides bound to most class I MHC molecules, the side chain of the C-terminal leucine in the ALPHAILRL peptide points down into a hydrophobic pocket F. However, UL18's pocket F is deeper and larger than its counterpart in HLA-A2 and other class I MHC mole-

cules, allowing it to accommodate larger side chains, such as phenylalanine (3), and possibly an additional residue (Fig. S3 *A* and *B*). The increased size of UL18's pocket F is due in part to the smaller size of UL18 Ile-158 compared with its counterpart Trp-147 in HLA-A2 and other class I MHC molecules. In class I MHC molecules, Trp-147, a resident of pockets E and F, blocks Trp-133 (UL18 Trp-134) and Ile-124 (UL18 Met-125). The smaller size of UL18 Ile-158 as compared with class I Trp-147 allows Trp-134 to participate in pocket E and causes Met-125 to be exposed in pocket F. These substitutions increase the size of UL18 pockets E and F, largely explaining the increased surface area in the UL18 groove (900 $Å^2$) as compared with a typical class I MHC groove (≈760 $Å^2$) (6, 7).

1. Chapman TL, Heikema AP, Bjorkman PJ (1999) The inhibitory receptor LIR-1 uses a common binding interaction to recognize class I MHC molecules and the viral MHC homolog UL18. *Immunity* 11:603–613.
2. Chapman TL, Heikema AP, West AP, Jr, Bjorkman PJ (2000) Crystal structure and ligand binding properties of the D1D2 region of the inhibitory receptor LIR-1 (ILT2). *Immunity* 13:727–736.
3. Fahnestock ML, *et al*. (1995) The MHC class I homolog encoded by human cytomegalovirus binds endogenous peptides. *Immunity* 3:583–590.
4. Madden DR (1995) The three dimensional structure of peptide-MHC complexes. *Ann Rev Immunol* 13:587–622.
5. Sudo T, *et al*. (1995) Differences in MHC class I self peptide repertoires among HLA-A2 subtypes. *J Immunol* 155:4749–4756.
6. Lebrón JA, *et al*. (1998) Crystal structure of the hemochromatosis protein HFE and characterization of its interaction with transferrin receptor. *Cell* 93:111–123.
7. Olson R, Huey-Tubman KE, Dulac C, Bjorkman PJ (2005) Structure of a pheromone receptor-associated MHC molecule with an open and empty groove. *PLoS* 3:1436–1448.

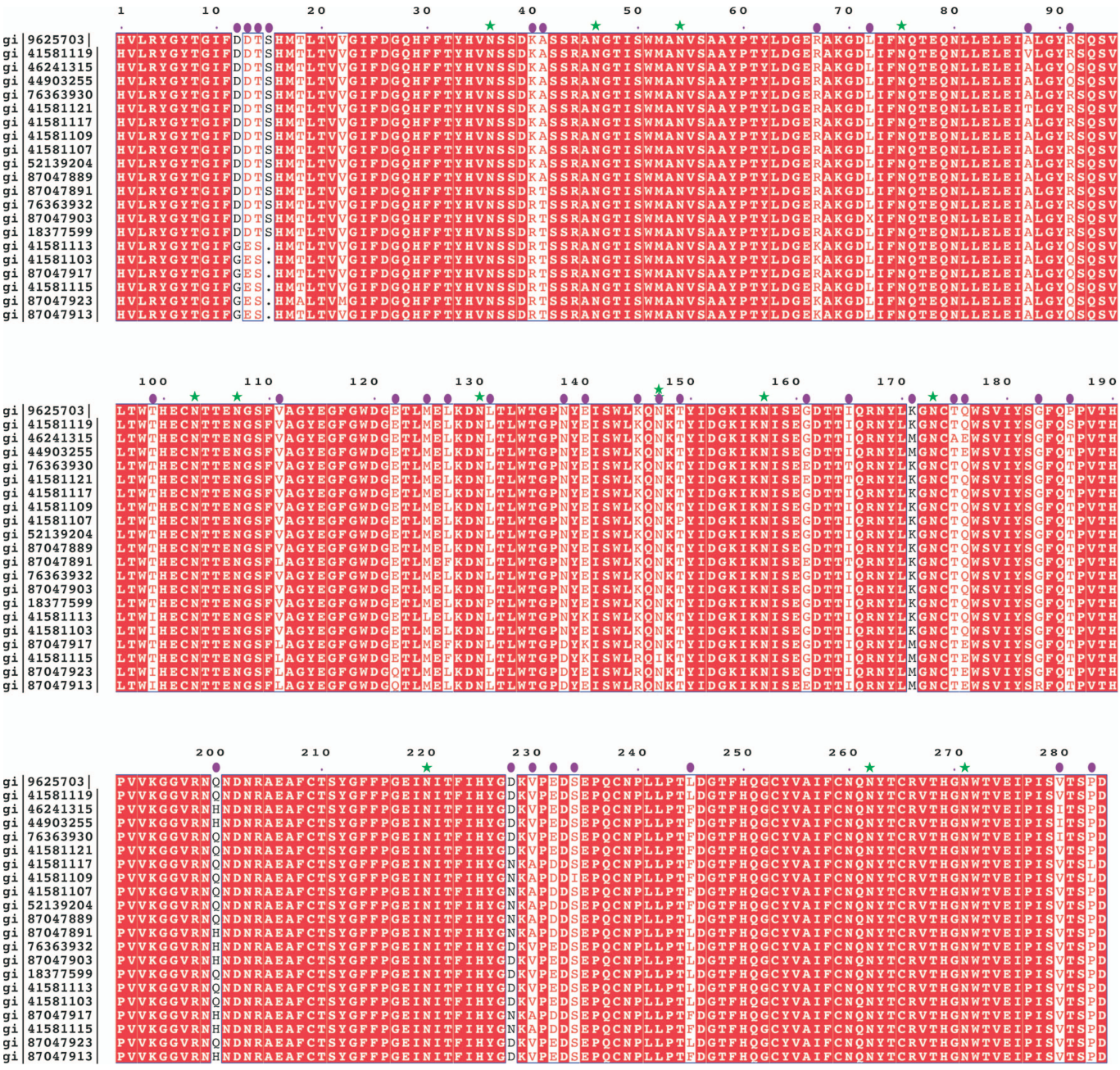


Fig. S1. Alignment of 21 unique UL18 sequences identified from the NCBI protein database (www.ncbi.nlm.nih.gov/). Conserved residues are shown as white letters in red boxes. UL18 residues that vary between clinical and laboratory isolates of HCMV are indicated with a purple dot. Potential N-glycosylation sites are marked with green stars.

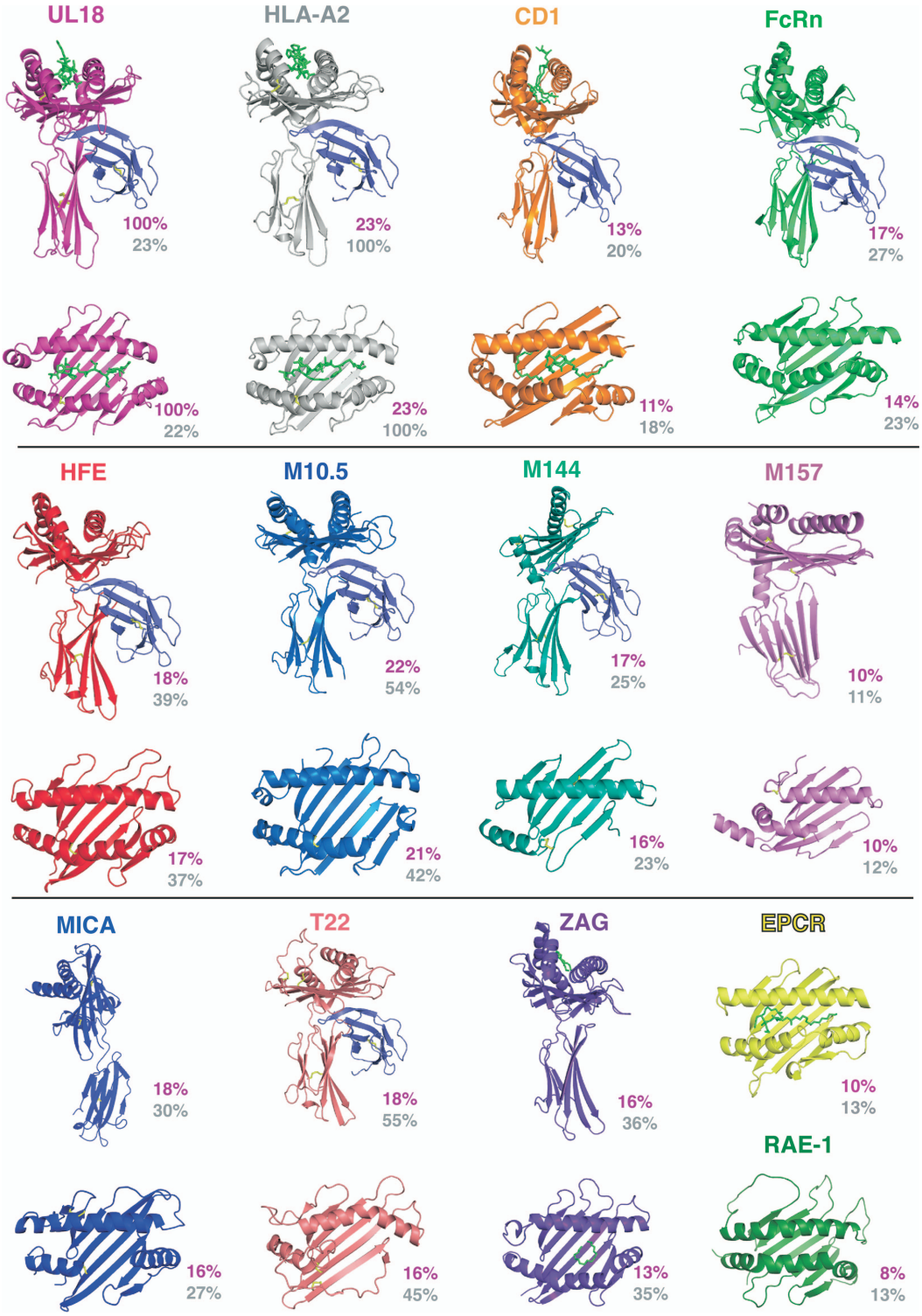
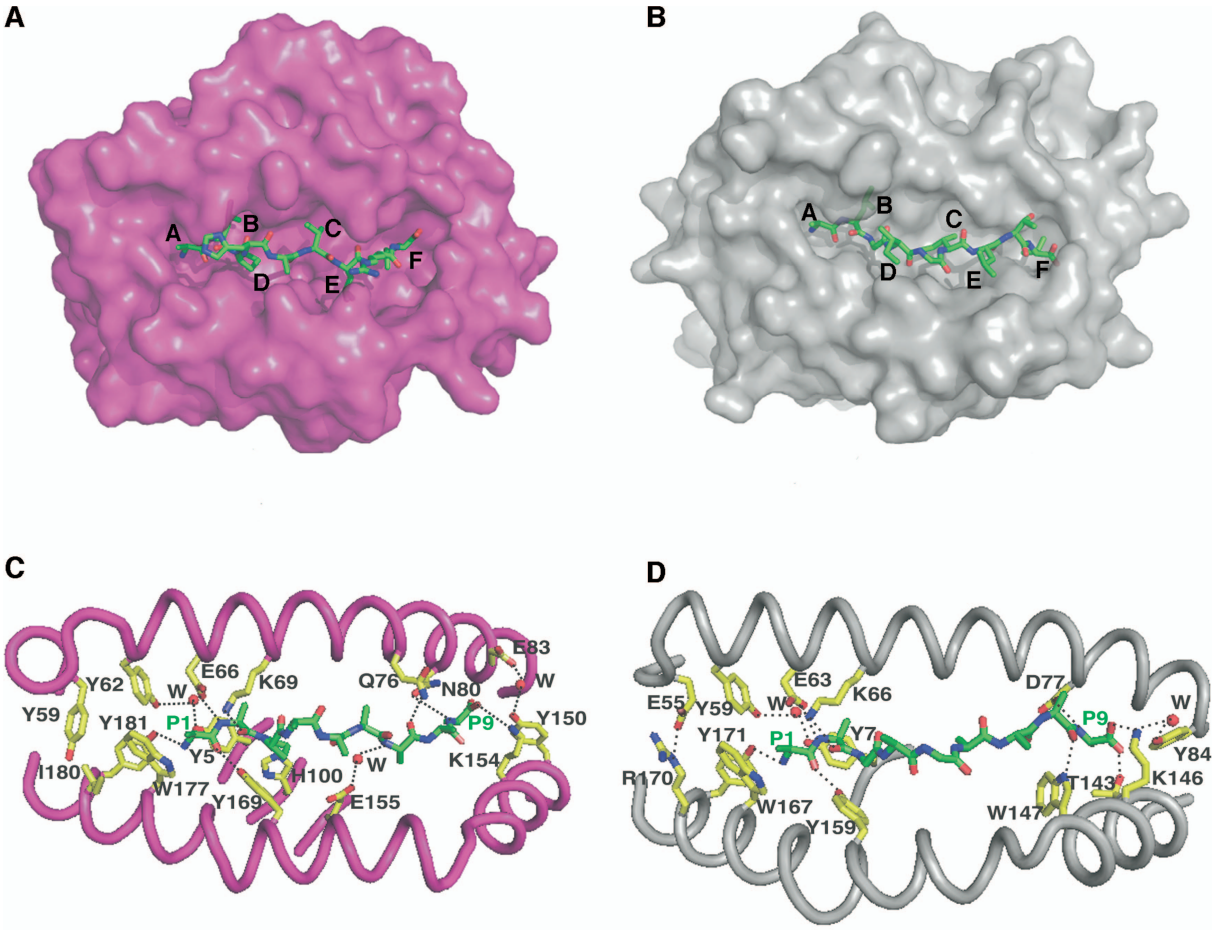


Fig. S2. Comparison of UL18, class I MHC, and class I MHC homolog structures. Ribbon diagrams of MHC and MHC homolog structures, each shown as a side view of the intact ectodomain (upper structures) and a top view of the α1–α2 domain platform (lower structures). Only platforms are shown for homologs that do not include α3 domains (EPCR and RAE-1). The percent sequence identities shared between each protein and UL18, or between each protein and HLA-A2, are listed beside each structure in pink (comparison with UL18) and gray (comparison with HLA-A2). The percent identities were calculated using a pairwise comparison of protein sequences as implemented in the DaliLite webserver (www.ebi.ac.uk/DaliLite/). Ribbon diagrams were prepared using coordinates from the following structures: UL18 (this paper), HLA-A2 (PDB ID code 1P7Q), CD1 (PDB ID code 2FIK), FcRn (PDB ID code 3FRU), HFE (PDB ID code 1A6Z), M10.5 (PDB ID code 1ZS8), M144 (PDB ID code 1U58), M157 (PDB ID code 2NYK), MICA (PDB ID code 1B3J), T22 (PDB ID code 1C16), ZAG (PDB ID code 1ZAG), EPCR (PDB ID code 1L8J), and RAE-1 (PDB ID code 1JFM).



E

Peptide		UL18 H bond/salt bridge partner		
Residue	Atom	Residue	Atom	Distance (Å)
P1Ala	N	Tyr5	OH	2.97
	N	Tyr181	OH	2.68
	O	Tyr169	OH	2.81
P2Leu	N	Glu66	Oε2	2.63
	N	Tyr5	OH	3.67
	O	Lys69	Nζ	2.45
P5Ala	O	Gln166	Nε2	3.24
P7Leu	N	Water	OH	3.20
	O	Gln76	Nε2	3.30
P9Leu	N	Asn80	Nδ2	2.91
	O	Asn80	Nδ2	3.39
	O	Lys154	Nζ	3.75

Fig. S3. Comparison of peptide binding by UL18 and HLA-A2. (*A* and *B*) Peptides binding to space-filling representations of the grooves of UL18 (*A*; peptide ALPHAILRL) and HLA-A2 (*B*; peptide ALGIGILTV; PDB ID code 1JHT). The relative positions of pockets A–F are indicated. (*C* and *D*) Interactions between main-chain atoms of the bound peptide and side-chain atoms of the peptide-binding groove are shown for UL18 (*C*) and HLA-A2 (*D*). The peptides are shown with main-chain atoms only. Hydrogen bonds and salt bridges are indicated by dashed lines. (*E*) Contacts between peptide main-chain and UL18 side-chain atoms.

A

	Ala	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
P1	77	5	10	0	63	0	110	25	110	10	15	20	0	0	95	60	120	0	10
P2	0	6	0	15	30	0	5	0	135	140	0	0	20	5	10	5	5	0	5
P3	63	15	13	5	37	0	0	5	0	0	50	130	5	10	10	30	10	0	0
P4	5	40	73	5	45	10	5	20	0	0	25	53	20	5	16	60	10	0	10
P5	28	13	5	5	13	18	27	23	10	0	10	13	30	13	18	10	50	0	10
P6	18	0	5	20	15	25	60	5	60	18	0	5	5	10	15	65	48	0	10
P7	20	0	0	10	0	5	88	10	28	15	10	40	10	5	10	5	38	0	5
P8	0	10	30	10	10	0	30	20	23	5	5	5	20	50	15	10	0	0	0
P9	0	0	0	50	10	10	23	0	75	0	5	0	0	0	0	20	5	0	20

B

Peptide	Score	Accession	Position	HCMV Protein
VLALRIIRL	172	NP_039961.1	395	UL27
LMPPPVAEL	169	NP_040104.1	48	UL19
VMPPELPAL	165	NP_783807.	15	US24
QLPEVQQRL	164	NP_039982.1	765	UL48
ALIPVVIIL	164	NP_039935.1	14	UL1
ELPQLLPRL	163	NP_040011.1	307	UL77
VMAPRTLIL	160	NP_039974.1	15	UL40
LLPLLLCRL	158	NP_039947.1	14	UL13
SMPELSLTL	157	NP_040096.1	17	US11
LMNGQQIFL	155	NP_783775.1	84	UL83
PLPNPLVLL	154	NP_783775.1	84	US21A
ELPQVDARL	154	NP_040094.1	136	US9
QLPEKYIGF	154	NP_040085.1	590	UL51
ILPLFIIAF	154	NP_040012.1	290	UL78
LLPAKRSRL	154	NP_039991.1	1277	UL57
LLPSLDARL	153	NP_039985.1	89	UL51
WLPSVLSLL	152	NP_040104.1	155	US19
NLPFTVLRL	152	NP_040033.1	44	UL98
VLPHETRLL	152	NP_040018.1	34	UL83
VLPHRLEQL	150	NP_040106.1	39	US21
VLPRPLELL	150	NP_040028.1	96	UL93
LLPKVGIPT	150	NP_040021.1	9	UL86
DLPLVSSRL	149	NP_040050.1	473	UL112
IMATQLRDL	149	NP_040008.1	321	UL74
LLWWITILL	149	NP_039943.1	7	UL9
ALPTTAYLL	148	NP_040021.1	423	UL86
FLPTPSLIL	148	NP_039978.1	42	UL44
LLAEFRQRL	147	NP_783808.1	305	US28
QLRGKHIRL	147	NP_040039.1	260	UL104
SLPSLTAHL	147	NP_039982.1	1769	UL48
VLKRALVRL	147	NP_039969.1	154	UL35
PLPPYLKGL	146	NP_039989.1	703	UL55
LLDGVTASL	146	NP_039983.1	183	UL49

Fig. S4. Potential UL18-binding peptides in the AD169 HCMV genome identified by using information derived by pool sequencing of peptides eluted from purified UL18 expressed in CHO cells (1). A matrix describing the frequencies of amino acids observed at the first nine positions of UL18-binding peptides was constructed and used to query sequences of HCMV proteins. The set of all HCMV sequences annotated with National Center for Biotechnology Information (NCBI) taxonomy ID 10360 (203 sequences containing 68,075 total residues) was extracted from the NCBI nonredundant protein database. The matrix data were input into the Profit program (EMBOSS 5.0.0) (2), which was modified to allow normalization of each row of the array to a total weight of 100 and to output both the raw score and the corresponding peptide sequence at each scored position. (*A*) Frequency matrix derived from the peak heights for each amino acid at the nine sequence positions in figure 4 from ref. 1. Frequencies are expressed in arbitrary units corresponding to the peak heights in millimeters. (*B*) Top scoring 9-mer peptide sequences derived by scanning the frequency matrix represented in *A* across all positions of the HCMV sequences. The top-scoring 0.05% of peptide sequences are listed in descending order with the highest-scoring peptide listed first.

1. Fahnestock ML, *et al.* (1995) The MHC class I homolog encoded by human cytomegalovirus binds endogenous peptides. *Immunity* 3:583–590.
2. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 16:276–277.

Table S1. Data collection and refinement statistics for the LIR-1/UL18 complex

	Data set 1	Data set 2
Data collection		
Wavelength, Å	1.0	1.0
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c*, Å	52.15, 97.44, 172.86	52.36, 98.11, 172.15
α, β, γ, °	90.0, 99.35, 90.0	90.0, 98.46, 90.0
Resolution, Å	2.6	2.2
R_{merge}, %	5.1 (24.7)	4.7 (23)
$I/\sigma I$	27.0 (5.5)	24.9 (6.6)
Completeness, %	99.9 (99.9)	97.0 (91.0)
Redundancy	3.8 (3.6)	3.6 (3.5)
Refinement		
Resolution, Å	50–2.6	50–2.2
No. reflections	52,515	85,873
R_{work}/R_{free}, %	22.2/25.1	24.1/25.9
No. atoms		
Protein	8,983	9,003
Peptide	0	142
Carbohydrate	278	214
Water	276	362
B factors		
Protein	53.3	42.2
Peptide	—	59.2
Carbohydrate	103.2	65.3
Water	55.3	46.1
rms deviations		
Bond lengths, Å	0.008	0.009
Bond angles, °	1.7	1.7
Coordinate error*, Å	0.35	0.31
Ramachandran statistics		
Residues in most favored regions, %	86.4	88.3
Residues in additional allowed regions, %	11.7	10.7
Residues in generously allowed regions, %	1.7	0.8
Residues in disallowed regions, %	0.2†	0.2†

*Estimate for the upper limit of error in the atomic coordinates determined by a Luzzati plot.
†Accounted for by one residue, LIR-1 residue Lys-41, in both copies of the LIR-1/UL18 complex in the asymmetric unit. Lys-41 is at the interface with UL18 and forms water-mediated hydrogen bonds with UL18 residues Ala-256 and Glu-235 (Table S2).

Table S2. Water-mediated hydrogen bonds in the UL18/LIR-1 interface

LIR-1			UL18 heavy chain	
Residue	Atom	Water	Residue	Atom
Tyr38	OH	W179	Asp233	Oδ1
Lys41	O	W2	Ala256	N
Lys41	O	W2	Glu235	N
Lys42	Nζ	W63	Glu206	Oε1, Oε2
Ala44	O	W166	Asn201	Nδ2
Tyr76	OH	W37	Asn203	Oδ1
LIR-1			β2m	
Residue	Atom	Water	Residue	Atom
Ala127	O	W102	Thr4	N
Trp67	Nε1	W123	Pro90	O

Table S3. Comparison of residues in the peptide-binding grooves of UL18 and HLA-A2

UL18 residue	Pocket*	HLA-A2 residue	Pocket*
Leu-3†	A	Met-5†	A
Tyr-5	A, B	Tyr-7	A, B
Tyr-7	B, C	Phe-9	B, C
Ala-45	B	Met-45	B
Tyr-62	A	Tyr-59	A
Glu-66	A, B	Glu-63	A, B
Lys-69‡	A	Lys-66‡	A
Gly-70†	B	Val-67†	B
Ile-73	B, C	His-70	B, C
Gln-76‡	C	Thr-73‡	C
Asn-80	F	Asp-77	F
Glu-83‡	F	Thr80‡	F
Leu-84	F	Leu-81	F
Leu-96	F	Val-95	Buried§
Trp-98	C, E	Arg-97	C, E
His-100	A, B, C, D	Tyr-99	A, B, C, D
Glu-115	C, D, E	His-114†	C, D, E
Phe-117	C	Tyr-116	C, F
Trp-119	Buried§	Tyr-118	F
Leu-124	F	Tyr-123	F
Met-125	F	Ile-124†	F
Trp-134	E	Trp-133	Buried§
Lys-154	Exposed¶	Thr-143	F
Ile-158	E, F	Trp-147	E, F
Asp-162	Exposed¶	Val-152	E
Gln-166	D, E	Leu-156	D, E
Tyr-169	A, D	Tyr-159	A, D
Trp-177‡	A	Trp-167‡	A
Tyr-181†	A	Tyr-171†	A

*Pocket residues defined as having $\geq$5.0 Å^2 accessible surface area to a 1.4-Å probe, but $<$ 5.0 Å^2 accessible surface area to a 5.0-Å radius probe. Surface areas were calculated excluding bound peptide and water molecules by using the coordinates of UL18 or HLA-A2 (PDB code 1JHT).

†Residues that contact the bound peptide with $<$5.0 Å^2 of accessible surface area using a 1.4-Å probe.

‡Residues with $\geq$5.0 Å^2 of accessible surface area using a 5.0-Å probe.

§Buried residues defined as those with $<$5.0 Å^2 of accessible surface area using a 1.4-Å probe and not involved in interactions with peptide.

¶Exposed residues defined as those with $\geq$50.0 Å^2 of accessible surface area using a 5.0-Å probe and that contact the bound peptide. These residues were not included in the groove surface area calculation.