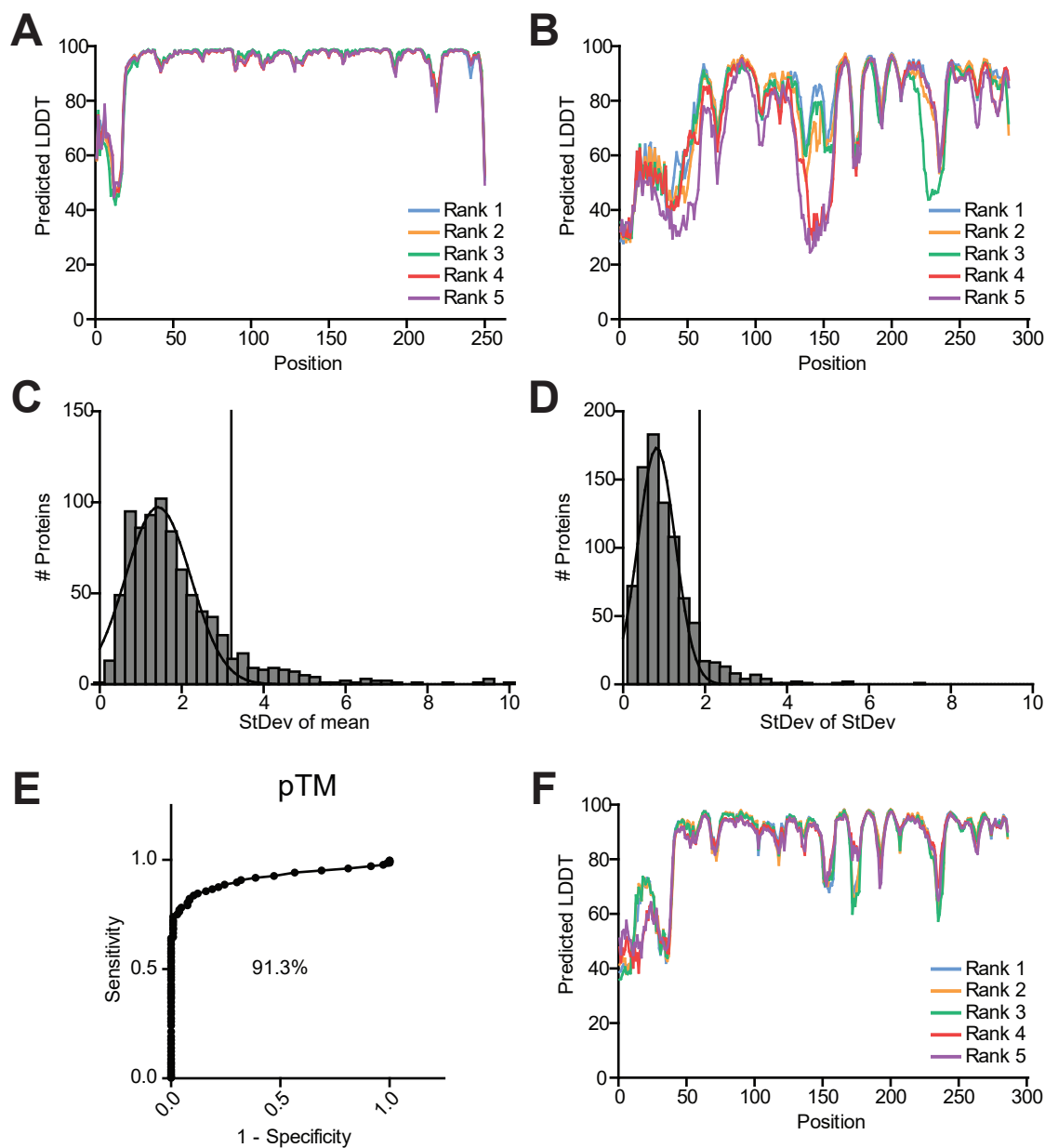
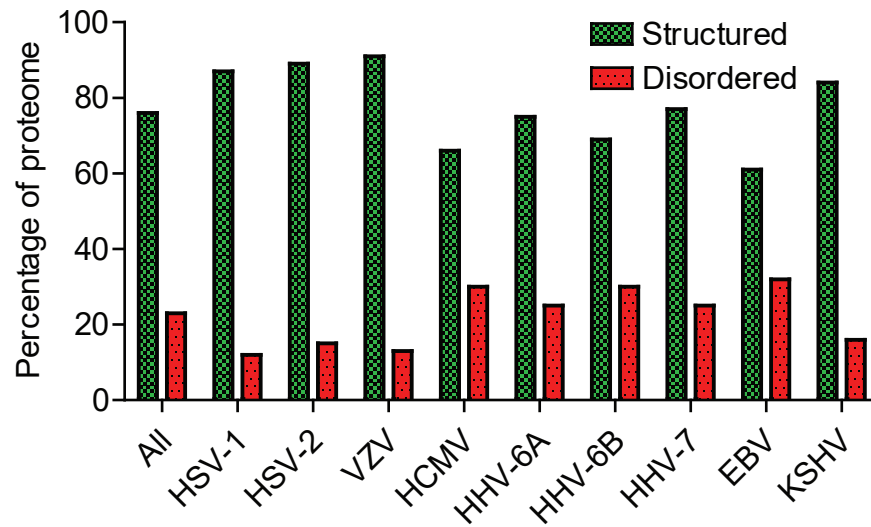
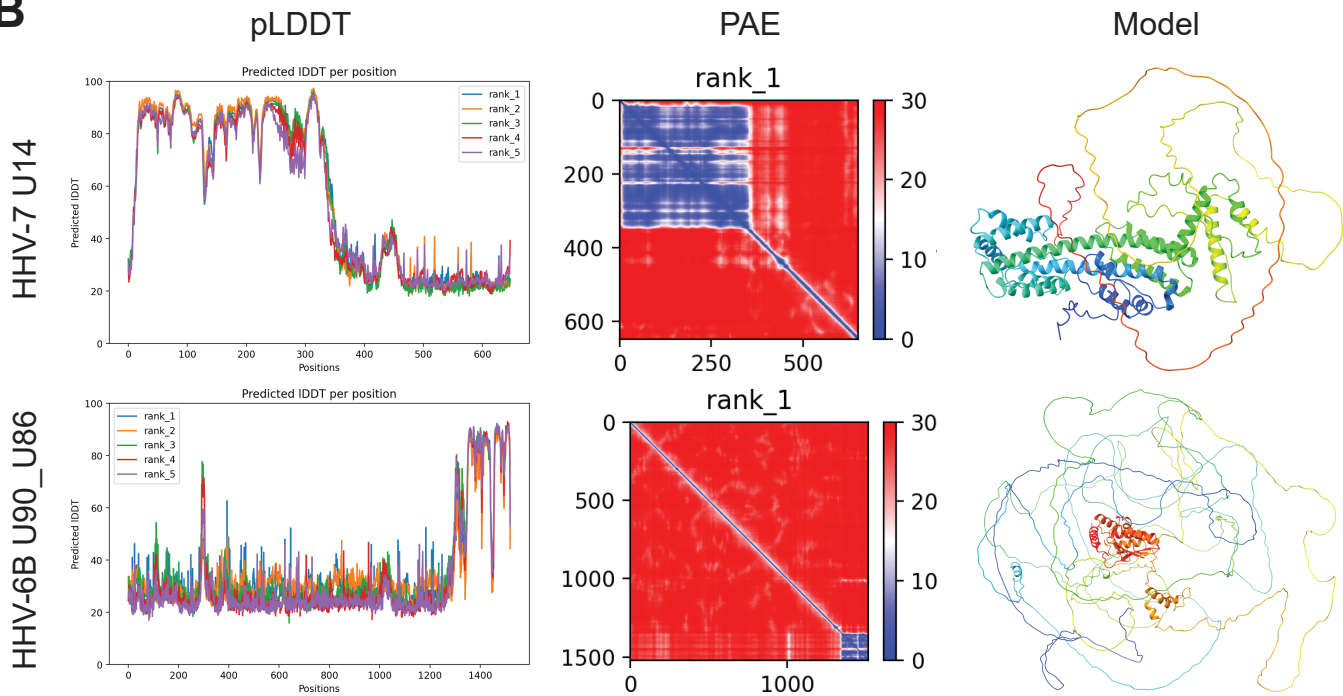


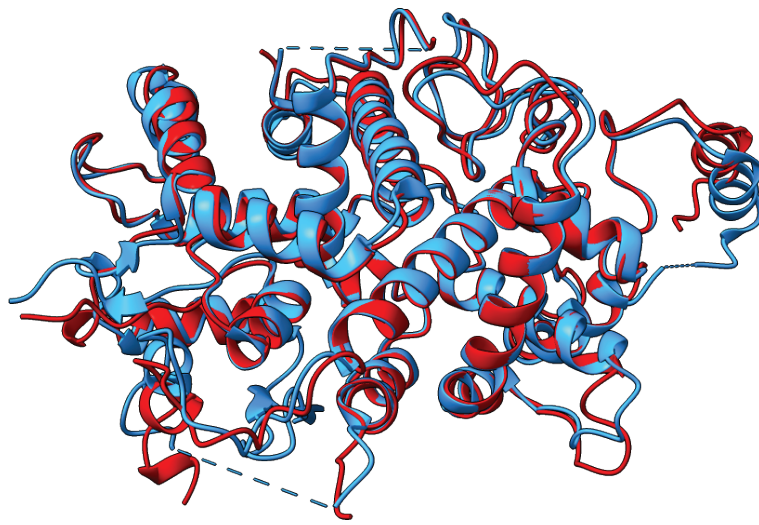
A proteome-wide structural systems approach
reveals insights into protein families
of all human herpesviruses



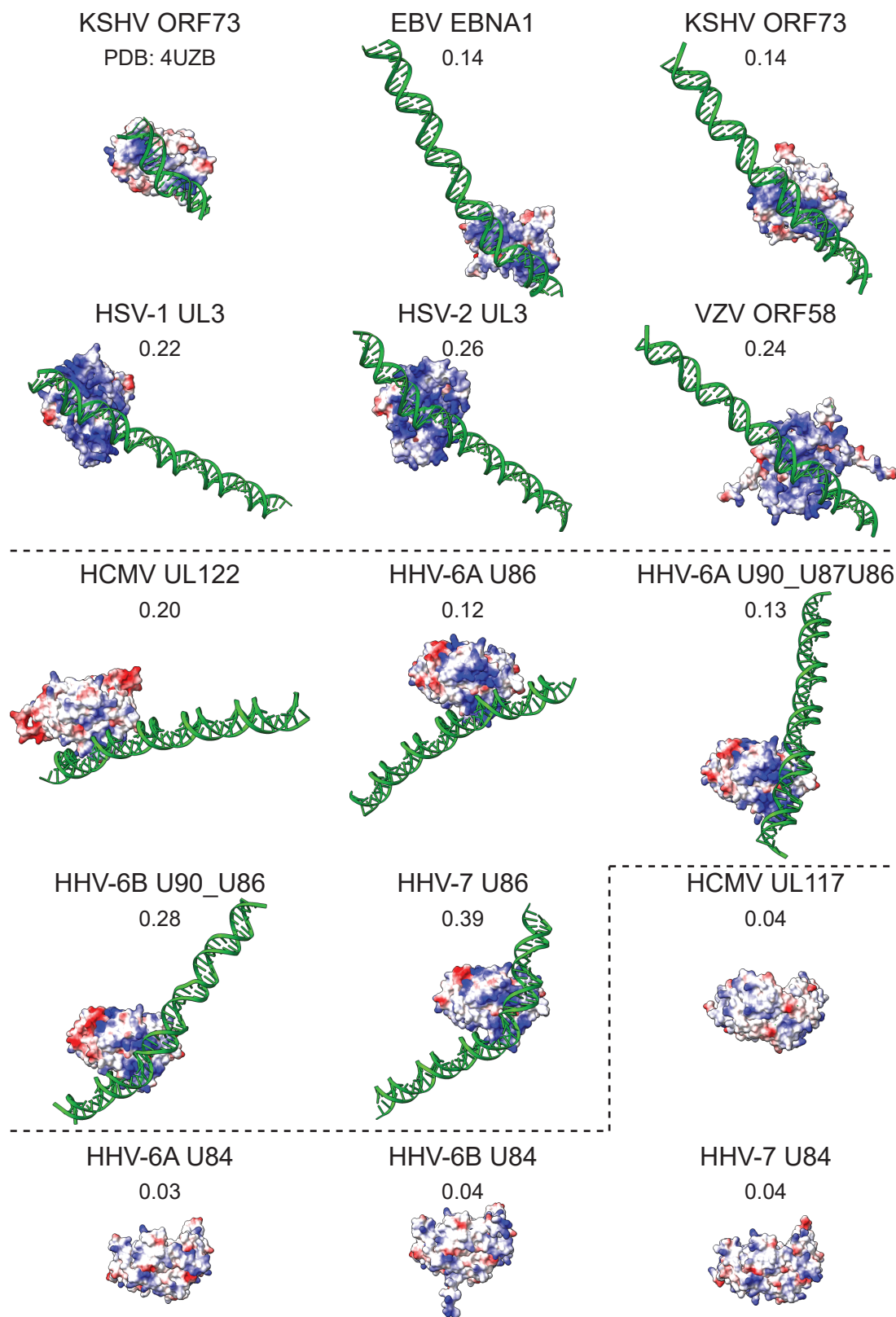
Supplementary Fig. 1. **Developing thresholds for evaluating model quality.** A. pLDDT plot of HCMV UL114 as an example of a protein where the pLDDT curves of the 5 models are consistent. B. pLDDT plot of HHV-7 U29 as an example of a protein where the pLDDT curves of the 5 models are inconsistent. C. Histogram of the standard deviation (StDev) of the mean of the pLDDT for each model for all LocalColabFold models. The threshold (black vertical bar) is set at the mean + 2x StDev, 3.2. D. Histogram of the StDev of the StDev of the pLDDT for each model for all LocalColabFold models. The threshold (black vertical bar) is set at the mean + 2x StDev, 1.9. E. ROC analysis of the pTM of all LocalColabFold models that passed the quality scores versus a negative control harboring scrambled amino acid sequences. The area under the curve is 91.3%. F. pLDDT curve of HHV-7 U29 as an example of a protein that passed the quality score thresholds after AlphaFold was used.

A**B**

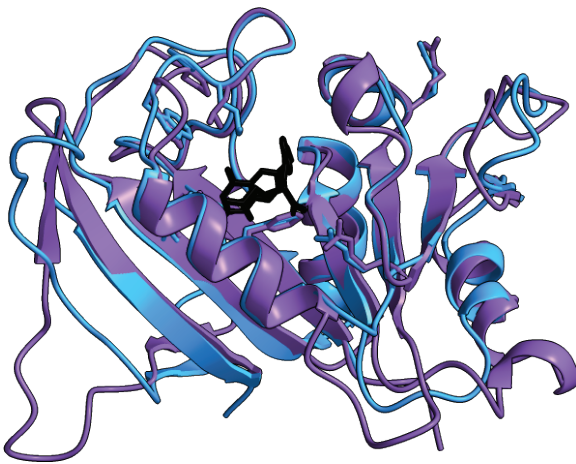
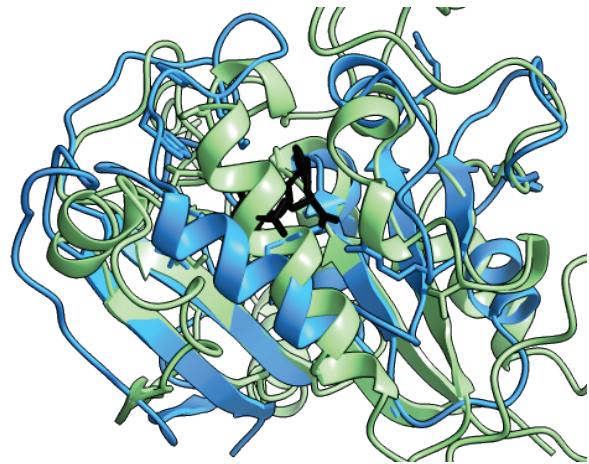
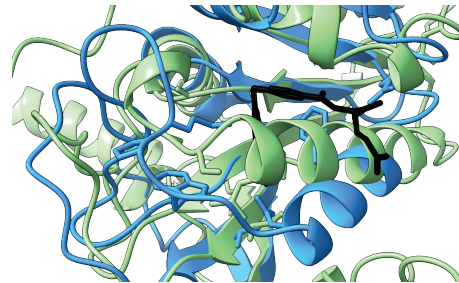
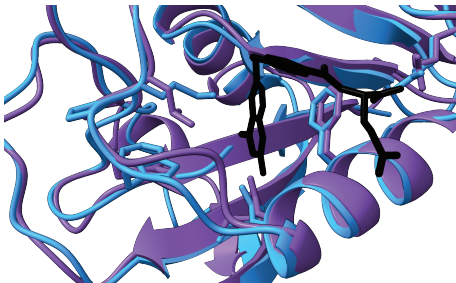
Supplementary Fig. 2. **Herpesviruses express a large fraction of proteins that contain a disordered domain.** A. The percentage of proteins that contain a structured or disordered domain are shown for each species as well as the consolidated proteome. B. HHV-7 U14 is an example of a protein with a structured N-terminus while HHV-6B U90_U86 has a structured C-terminus. The pLDDT and PAE plots are also shown to illustrate how the structured and disordered regions appear. The predicted structure is shown as a rainbow colored ribbon diagram where blue is the N-terminus and red is the C-terminus.



Supplementary Fig. 3. **Alignment of predicted HCMV UL77 structure to the experimental structure that was deposited post-AlphaFold training.** The deposited structure for HCMV UL77 (PDB 7NXP¹⁹) (blue) was aligned with the AlphaFold prediction (red), and the comparison has an RMSD of 0.509 Å against the 310 C α . Only the residues present in the deposited structure are shown in the predicted model. The corresponding UCSF ChimeraX session can be found at <https://zenodo.org/records/13284140>.



Supplementary Fig. 4. **Predicted positive surface charge and DNA binding activity of the KSHV ORF73 and EBV EBNA1 cluster.** The proteins in this cluster were structurally predicted as dimers of the conserved core with the DNA binding sequence of KSHV ORF73⁸⁰ with AlphaFold 3. The crystal structure of KSHV ORF73 with DNA (PDB: 4UZB²⁹) was included for comparison. The average pairwise ipTM between each protein and DNA strand is shown above the structure. The dashed line separates the different binding phenotypes: KSHV ORF73, HCMV UL122 (different orientation), and HCMV UL117 (no binding). Positive surface charge is displayed as blue, and negative charge is displayed as red. The corresponding UCSF ChimeraX session can be found at <https://zenodo.org/records/13284140>.

A*E. coli*
KSHV ORF2*E. coli*
HSV-1 US2**B****C**

Structure based alignment

```

E. coli      ... MISLIAALAVD RVI GMENA MPWN . L . PADLWLF KRNTL 36
KSHV_ORF2   .NDP FLICYVAVDTK LGL GDNMC LFWP AL . KGMRRFP RGLTID 41
HSV-1_US2   M.GVVVVNMTLLDQNNALPRT .SVDA.S.PALMS .F.LLRQ CRI 39
HSV-2_US2   M.GVVVVSVVTLDDQNNALPRT .SADA.S.PALMS .F.LLRQ CRI 39
          : : : : : * : : : :

E. coli      .....DKPVIMGRWTWESIG.....R..PLPGR...KWIILS... 63
KSHV_ORF2   ..C.A..PGKQNMVVMGRKTWLSI.PAGCR..PLAGR...INVVLS... 76
HSV-1_US2   .LASEPLG..TPVVVR...PAN.LRR..LAEP.LMDLPKTPPIVTRSC 79
HSV-2_US2   .LASEPLG..TPVVVR...PAN.LRR..LAEP.LMDLPKTPPIVTRSC 79
          : : : : : * : : : :

E. coli      .....SQ.PG.T.DORV.TWVKS.VDEAI.AAC..... 85
KSHV_ORF2   .....RTLETFFP..GANFIASS.LDAAL.GLAR..S.P..E 104
HSV-1_US2   RCPPTNTTGLFAE.....DSFLES.TEVVDVA.CFRL.LHRDQSPFP. 119
HSV-2_US2   RCPPTNTTGLFAE.....DSFLES.TEILDAPA.CFRL.LHQRPGPH. 119
          : : : : : * : : : :

E. coli      .GD...VPEIMVIGGGRVYEQLP.K...AQRLYLTHIDAEVGOTHF 125
KSHV_ORF2   LA.QQID.KVWVIGGGNLYREALT.G..P.NFVRLFLTRVLDFAVDVL 148
HSV-1_US2   .RL...Y.HLMVVGADLCVPFEYAGKIRLGVRFAIKTPDAWGEFW. 163
HSV-2_US2   .RL...Y.HLMVVGADLCVPFEYAGKIRLGVRFAIKTPDAWGEFW. 163
          : : : : : * : : : :

E. coli      PD..YE.PDD.WESVFS.....E.....FHADAQNS.H..S 150
KSHV_ORF2   SHSDIA..AYARV.N.KPFGQGE.....RVFQ...ERGI..F 176
HSV-1_US2   AVP.TRL..PERT.VS.....WTFPPAAPHPLE.....T...LSR 194
HSV-2_US2   PLP.DRL..PERT.VS.....WTFPPAAPHPLE.....N...LSR 194
          : : : : : * : : : :

E. coli      YCFEILERR..... 159
KSHV_ORF2   YMFETVIRV.TD..... 187
HSV-1_US2   YEYQGVVLPQT.NGRE.RDC.MRWL.RSL..IALH.KPHATP.GEL.T 235
HSV-2_US2   YEYQGVVVGDRERSCLR.MLR..SD...VAPH..NK.PPASSR.PHP 234
          * : : : :

E. coli      ..... 280
KSHV_ORF2   TSHPVRRPCCACGMPEVDEQFTSPGRGPQETDPLIAVRGERFR..... 280
HSV-1_US2   TSHPVRRPCCACGMPEVDEQFTSPGRGPQETDPLIAVRGERFR..... 280
HSV-2_US2   ATHPTQRPCFTCMG.....RFEIP 253
          : : : : :

E. coli      ..... 291
KSHV_ORF2   .....SSDTA 192
HSV-1_US2   .....LPHI.CYPVTTL 291
HSV-2_US2   DEPSWQTGDDDPQNGPPLAVGDEMFPSSHV.CYPI.....TNL..... 291
          : : : : :

E. coli      ..... 210
KSHV_ORF2   LPDLERPRPATPPFSETS 210
HSV-1_US2   ..... 291
HSV-2_US2   ..... 291

```

CLUSTAL O(1.2.4)

```

E. coli      ---MISLIAALAVDRVIGMENAMPWNLP-ADLAWFKRN-----TLNKFVIMERHT 46
KSHV_ORF2   MDPTLYCVAVDTKLGIGDNMC.LFWPALRGMRRFPRLTIDCAP-----GKQNMVGRKT 56
HSV-1_US2   -----MGVVVVNMTLLDQNNALPRTSVDA.S.PALMS.LLRQCRILASEPLGTPVVVRPAN 55
HSV-2_US2   -----MGVVVVSVVTLDDQNNALPRTSADA.S.PALMS.LLRQCRILASEPLGTPVVVRPAN 55
          : : : : : * : : : :

E. coli      W-----ESIG---RPLGRNMIILSQPGCTDDRV---TWV-KS-----VDEA--- 81
KSHV_ORF2   W-----LSIPAGCRPLAGRINVLRTLETTPPGA---HFLASS-----LDA---L 97
HSV-1_US2   LRLAEPLMDLPKTPRTI-----VTRSCRCPPNTTGLFAEDSPLESTEVDVACFR 109
HSV-2_US2   LRLAEPLMDLPKTPRTI-----VTRSCRCPPNTTGLFAEDSPLESTEILDAPACFR 109
          : : : : : * : : : :

E. coli      -----IAACGVPEIMVIGGGRVYEQLP.KA--QKLYLTHIDA---EVEG-----DTH 124
KSHV_ORF2   GLARSPELAQQIDKVVVIGGGNLYREALTGPWPVRLFLTRVLH---DFAC-----DVF 147
HSV-1_US2   LLHRDQSPFPLVHLMVVGADLCVPFEYAGKIRLGVRFAIKTPDAWGEFWPVPTR 169
HSV-2_US2   LLHQRPGPHRLVHLMVVGADLCVPFEYAGKIRLGVRFAIKTPDAWGEFWPVPTR 169
          : : : : : * : : : :

E. coli      FPDY-----EPD-DWESVSEFH-----DADAQNSHSYCFEILERR----- 159
KSHV_ORF2   LSHDSLAAAYRVNPKPG-EQERVFG-----RGIFMFETVIRVQTSDDTA 192
HSV-1_US2   LEWNTVA-WTFPPAAPHPLETLLSRVYEQGVVLPQTNGR-ERDCMRWLSLIALH-KP 226
HSV-2_US2   LPERTVS-WTFPPAAPHPLETLLSRVYEQGVVLPQTDG-R-ERSCLRWSLVAFPHNK 225
          : : : : : * : : : :

E. coli      ..... 159
KSHV_ORF2   LPDLERPRPATPPFSETS----- 210
HSV-1_US2   WPATPGILTSTSPVRRCCACGMPEVDEQFTSPGRGPQETDPLIAVRGERFR 285
HSV-2_US2   RPASSRPHPATHTQRCFTCMGRPEIPDEPSWQTGDDDPQNGPPLAVGDEMFPSSHV 285
          : : : : :

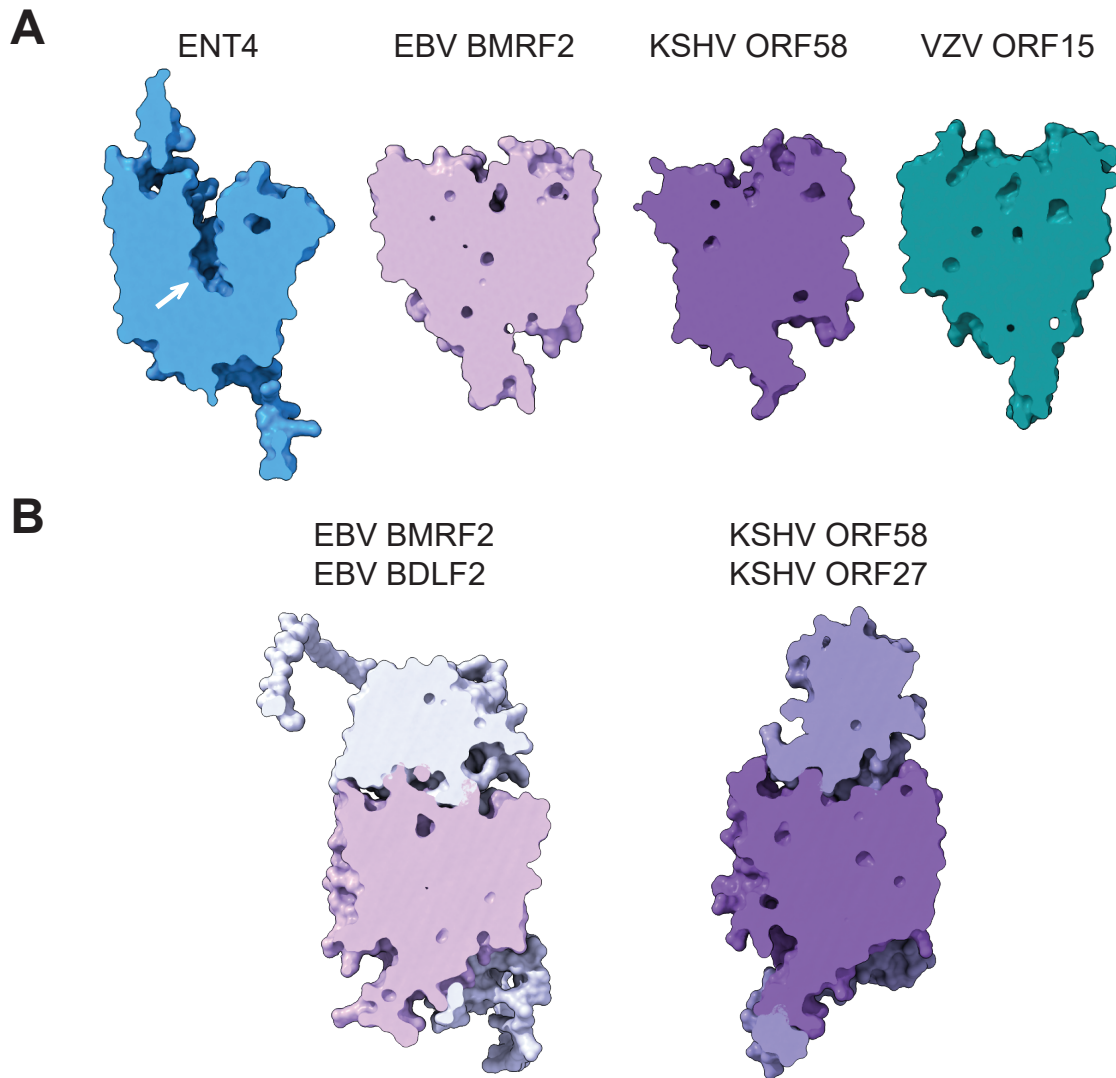
```

```

E. coli      ..... 159
KSHV_ORF2   ..... 210
HSV-1_US2   YPVTTL 291
HSV-2_US2   YPITNL 291

```

Supplementary Fig. 5. **Alignment of *E. coli* DHFR to KSHV ORF2 and HSV-1 US2.** A. The *E. coli* DHFR (PDB 1RX2⁴⁸) (blue) was aligned to KSHV ORF2 (purple) and HSV-1 US2 (green), and the cofactor folic acid is shown in black. The atomic structure of the catalytic residues of the *E. coli* DHFR (M20, P21, W22, D27, F31, R44, R57, G95, G96, T113) and the potential catalytic residues that are in similar positions in ORF2 and US2 are shown. B. An enlarged view of the catalytic site is shown. C. The protein sequences were aligned based on the structure-based alignment or with Clustal Omega. The *E. coli* catalytic residues are marked with red asterisks. The potential catalytic residues in the herpesvirus proteins are from the structured based alignment. The corresponding UCSF ChimeraX session can be found at <https://zenodo.org/records/13284140>.



Supplementary Fig. 6. **The herpesvirus ENTs are unlikely to have transporter activity.** A. Cross-sections of surface plots of the monomer predictions of ENT4 (AF-Q8R139-F1-v4) and the structurally similar viral proteins. The white arrow points to the metabolite channel in ENT4. B. Structure predictions of the known EBV (BMRF2 pink and BDLF2 white) and KSHV (ORF58 purple and ORF27 light purple) heterodimers. The corresponding UCSF ChimeraX sessions can be found at <https://zenodo.org/records/13284140>.