Molecular Biology and Evolution, 2025, **42**, 1–19
https://doi.org/10.1093/molbev/msaf256
Advance access publication 9 October 2025
Discoveries

MBE OXFORD

Functional and Structural Determinants of Long- and Short-Term Evolution of Herpesvirus Proteins

Rotem Fuchs [1,†], **Alessandra Mozzi** [2,†], **Diego Forni** [2], **Bar Naim** [1], **Ofir Schor** [1], **Manuela Sironi** [3], **Tzachi Hagai** [1,*]

[1]Shmunis School of Biomedicine and Cancer Research, George S Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel
[2]Computational Biology Unit, Scientific Institute IRCCS E. MEDEA, Bosisio Parini, 23842, Italy
[3]School of Medicine and Surgery, University of Milano-Bicocca, Monza, 20900, Italy

[†]Equal contribution.
***Corresponding author:** E-mail: tzachiha@tauex.tau.ac.il.
Associate editor: Anne-Ruxandra Carvunis

Abstract

Understanding the factors that shape viral protein evolution is a central question in evolutionary virology. However, these determinants are poorly characterized for the majority of viruses, because functional data is scarce and because most viruses encode few proteins, limiting comparisons between them. Here, we focus on the *Orthoherpesviridae* family whose related viruses, including human-infecting herpesviruses, allow evolutionary investigation at different timescales. We employ different models to estimate evolutionary rates of numerous herpesvirus proteins and residues, and assess their relationship to a set of structural, cellular and functional characteristics. Core genes whose orthologs are found in distant genera, evolve at similar rates within genera, despite their evolutionary distance and their differences in viral replication and environments. This likely stems from constraints imposed to maintain the structural fold across viruses, and is corroborated by the finding that fold complexity is a major determinant of evolutionary rates. Focusing on the evolution of specific protein regions, we show that surface and disordered regions are enriched with positively selected residues. However, motifs embedded in disordered regions, important for binding host proteins, have conserved occurrences across viruses. Additionally, viral proteins predicted to form biomolecular condensates often evolve slowly despite having high disordered content. In summary, our analyses reveal short- and long-term evolutionary constraints of herpesvirus proteins. These include constraints imposed by the protein structural fold and by elements within disordered regions important for host-virus interactions. These constraints are relevant when considering potential pathways of virus evolvability and for developing new antiviral treatments.

Keywords: Protein evolution, Virus evolution, Herpesvirus, Determinants of protein evolutionary rates, Intrinsically disordered protein regions

Introduction

Understanding protein evolution is a central goal in evolutionary biology. The rates of protein evolution, as measured by substitutions in a given amount of time, varies greatly across cellular proteins (Koonin and Wolf 2010; Zhang and Yang 2015). This variability and its underlying factors have been a source for extensive research in cellular protein evolution (Drummond et al. 2005; Chen et al. 2011; Usmanova et al. 2024). In contrast, determinants of viral protein evolution were less studied, despite their clear importance for understanding host-virus evolutionary dynamics. Furthermore, viral protein evolution is often investigated in the context of antagonistic interactions with the host immune system (Tenthorey et al. 2022; Kistler and Bedford 2023), while other functional or biophysical characteristics are rarely considered. This is, at least in part, because viruses usually encode for relatively few proteins, limiting comparisons within proteomes, and because functional data is often lacking for the majority of viral proteins within a given virus. In here, we take advantage of the structure of the *Orthoherpesviridae* family phylogeny, the large number of proteins these viruses encode, and the availability of functional data of a large set of viral proteins from human-infecting herpesviruses, to study how different determinants are linked with long- and short-term evolution of herpesvirus proteins.

Herpesviruses are enveloped, double-stranded DNA viruses that establish life-long infection in a wide range of animals, including mammals, birds, reptiles, and mollusks (Davison et al. 2009). Within the order *Herpesvirales*, the family *Orthoherpesviridae* is divided into three monophyletic subfamilies of vertebrate-infecting viruses: *Alpha-*, *Beta-*, and *Gammaherpesvirinae* (https://ictv.global/taxonomy) (Gatherer et al. 2021). Recent molecular dating analyses indicated that the three subfamilies diverged in the Jurassic period (150 to 210 million years ago) (Brito et al. 2021), and since then, have diversified into several viral species. Because of their common origin, viruses in the three subfamilies share a similar virion morphology and typically large genomes, which encode for 44 conserved genes (termed "core genes") that play key roles during virus replication. These orthologous core genes, found in viral genomes separated millions of years ago, provide the possibility to investigate how each of the orthologs evolved separately in each of the three viral subfamilies. Additional noncore, or accessory, genes are specific to individual subfamilies, genera or even species (McGeoch et al. 1995, 2006; Gatherer et al. 2021). Despite these similarities, viruses in the *Orthoherpesviridae* subfamilies also

Received: January 22, 2025. **Revised:** July 28, 2025. **Accepted:** September 15, 2025

display remarkable differences in terms of cell/tissue tropism, host range, and transmission route (Grinde 2013; Eberle and Jones-Engel 2017). For instance, members of the subfamily *Alphaherpesvirinae* establish latent infection in neurons, while members of the subfamilies *Betaherpesvirinae* and *Gammaherpesvirinae* establish their latent reservoirs in myeloid and/or lymphoid cells (Grinde 2013).

The evolution of herpesviruses has long been thought to be mainly characterized by co-speciation with their hosts (McGeoch et al. 2006). Primate herpesviruses are largely host-species-specific in natural settings, implying long-standing association and co-evolution between primates and herpesviruses (Eberle and Jones-Engel 2017; Brito et al. 2021; Wertheim et al. 2021). This evolutionary history offers us a unique opportunity to compare core and accessory herpesvirus proteins at various timescales.

Previous analyses that investigated the evolution and adaptation of primate herpesviruses over different time frames, showed that selection signatures in viral genomes often resulted from selective pressure exerted by the host immune system (Sijmons et al. 2015; Zhang et al. 2018; Mozzi et al. 2020a, 2020b, 2022; Brito et al. 2021). This is typically exemplified by several instances of fast-evolving surface-exposed residues of viral proteins that are targets of humoral responses or that interact with components of the immune system (Szpara et al. 2014; Sijmons et al. 2015; Lassalle et al. 2016; Mozzi et al. 2022; Palmer et al. 2022). However, evidence of adaptive evolution was also reported for viral core proteins that are a not exposed to immune-mediated pressure. For instance, the adaptation of human cytomegalovirus (HCMV, subfamily *Betaherpesvirinae*) to its human host was accompanied by changes in several core proteins, many of which interact with one another (Mozzi et al. 2020a).

Evolutionary patterns can also be related to biophysical features of the proteins themselves (DePristo et al. 2005; Tokuriki et al. 2009; Wilke and Drummond 2010; Echave and Wilke 2017), although relatively little is known in that regard for herpesvirus proteins and for the majority of viral proteins in general (Tokuriki et al. 2009). Data have been emerging that evolutionary changes in intrinsically disordered regions (IDRs) and low-complexity regions in herpesvirus proteins are associated with host adaptation or evasion of immune responses (Mozzi et al. 2020b; Zimmermann et al. 2024). This is in agreement with previous work suggesting that disordered regions in human and viral proteins can evolve rapidly (Nilsson et al. 2011; Gitlin et al. 2014; Afanasyeva et al. 2018). This rapid evolution can facilitate adaptation through various mechanisms (Nguyen et al. 2024; Chow and Toth-Petroczy 2025) including changes in eukaryotic linear motifs (ELMs) that are embedded in IDRs and mediate interactions with other proteins (Davey et al. 2011; Chemes et al. 2012; Hagai et al. 2014; González-Foutel et al. 2022; Glavina et al. 2022a, 2022b; Cagliani et al. 2024).

In this work, we study the evolution of herpesvirus proteins, using representatives from each of the three subfamilies within *Orthoherpesviridae*, and at three different time frames: (i) at the intra-species level, by comparing isolates of human-infecting viruses, (ii) at the inter-species level, by comparing orthologous genes within each genera, across human- and nonhuman primate (NHP) infecting viruses, and (iii) between orthologous core genes found in viruses belonging to different genera—genes separated by speciation hundreds of millions of years ago. We compare the evolutionary rates of these proteins and their residues along with various structural, cellular and functional features, providing a comprehensive analysis of a diverse set of determinants of viral protein evolution. The phylogenetic data files and structural prediction files created during this work are available for future use, as detailed in Methods.

Results

Inter- and Intra-Viral Dataset Assembly and Evolutionary Models

To obtain a global multiscale overview of herpesvirus protein evolution in primates, we focused on three viral genera—*Simplexvirus*, *Cytomegalovirus*, and *Rhadinovirus*—representatives of the three *Orthoherpesviridae* subfamilies that have diverged millions of years ago. These three genera were selected because: (i) Each has a sufficiently large number of fully or almost-fully sequenced genomes from NHP-infecting viruses; (ii) Many human clinical isolates with full sequence information are also available (Fig. 1). In humans, the three genera are represented by HSV1 and HSV2 (human herpes simplexvirus type 1 and type 2, genus *Simplexvirus*), HCMV (genus *Cytomegalovirus*), and Kaposi's sarcoma-associated herpesvirus (KSHV, genus *Rhadinovirus*). In addition to the human virus sequences and their isolates, the datasets included 4-6 viral species that infect Old World monkeys or great apes, and 1-2 that infect New World monkeys, providing a set of orthologous viral proteins across the phylogeny (Fig. 1). A detailed list of all viral species used in this work appears in Tables S1 and S2.

One-to-one orthologous viral proteins across the primate clade (human- and NHP-infecting viruses) were used for inter-species evolutionary analysis (see Materials and Methods) (Fig. 1, Left). The average nonsynonymous substitution/synonymous substitution rate ratio (dN/dS), a measure for selection, was calculated for all coding genes of the three virus genera that had a sufficient number of orthologs and were retained following additional filtering steps. For this, we used the single-likelihood ancestor counting (SLAC) method (Kosakovsky Pond and Frost 2005). These analyses provided us with long-term evolutionary rates of proteins and residues (between orthologous herpes viral proteins from viruses of the same genus that infect different primate species). In addition, we use these dN/dS values to compare core proteins between the genera, giving us a comparative analysis of proteins that were separated 150 to 210 million years ago.

For intra-species evolutionary analyses, representative genomes of human virus isolates were selected based on geographic distribution, disease status, and body compartment, as well as from strains with low-passage history in culture (Tables S3 to S5). Coding genes were aligned and a codon-wise inference of evolutionary patterns was obtained with the FEL (Fixed Effects Likelihood) method (Kosakovsky Pond and Frost 2005) (Fig. 1). FEL uses a maximum-likelihood approach to infer dN and dS at each site, and, based on the selective pattern, classifies codons into four classes: purifying selection, diversifying selection, neutral, and invariant (these are codons where $dN = dS = 0$). We thus obtained categorizations of residues based on their short-term selective patterns.

Characterizing Herpesvirus Protein Structural Features

To study the structural characteristics of herpesvirus proteins across the three viral genera, three measures were calculated

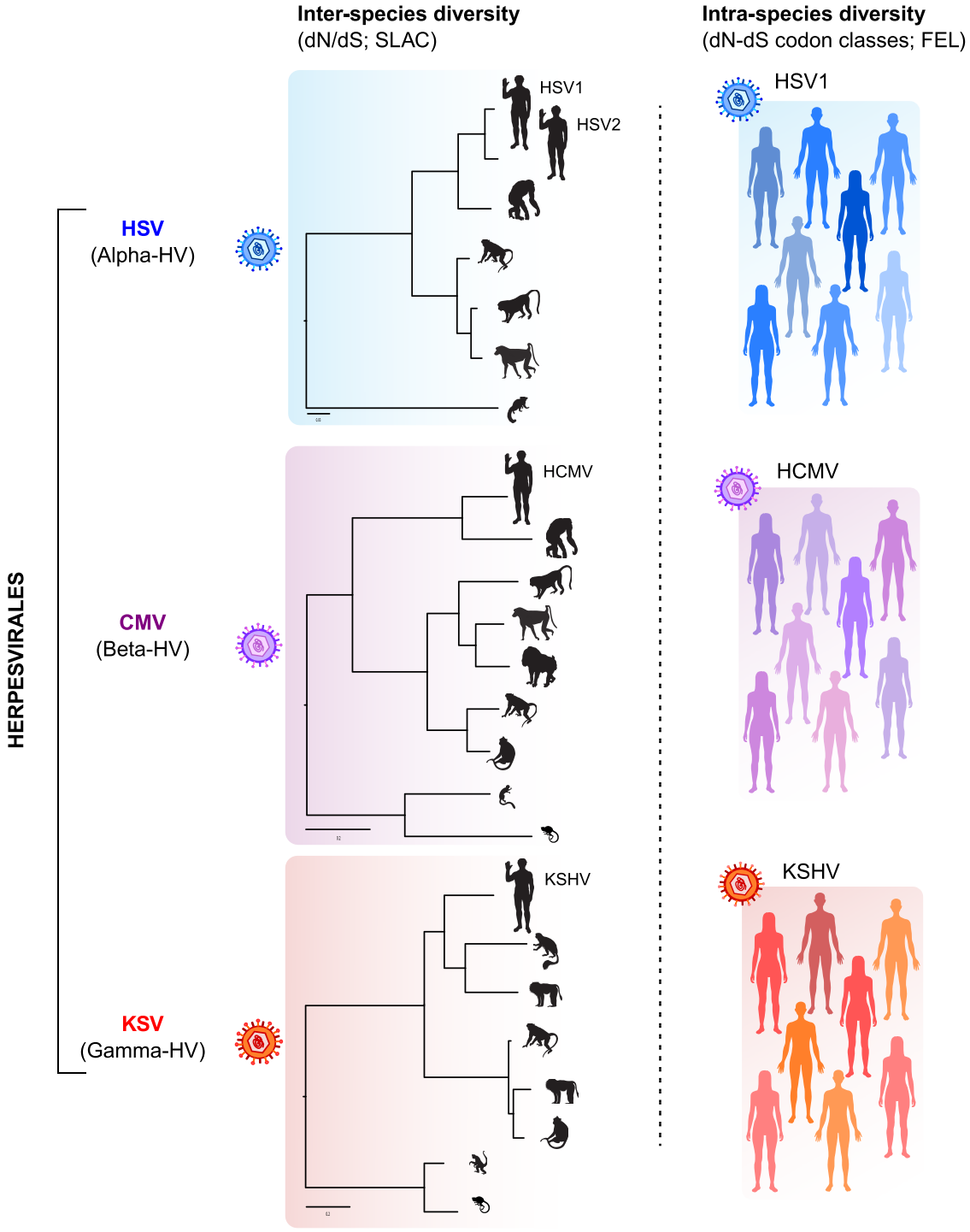


Fig. 1. Schematic representation of the herpes viral systems used in this study. Left: Within the *Orthoherpesviridae* family, we focused on human-infecting viruses, representatives of the three subfamilies, *Alpha-*, *Beta-* and *Gammaherpesvirinae*. Within each subfamily, we chose a genus (*Simplexvirus*, *Cytomegalovirus* and *Rhadinovirus*, respectively) that has one or two human-infecting viruses, and that each of these genera has several NHP-infecting viruses. Inter-species evolutionary analysis was performed for each genus separately, across orthologous proteins from the human and NHP-infecting viruses (for dN/dS-based analysis, New World monkey-infecting viruses were excluded due to the resulting MSA quality). All trees shown are based on orthologous glycoprotein B sequences (tree reconstruction is detailed in Materials and Methods). See additional trees in Figure S1. Right: Isolates of the human-infecting viruses (HSV1, HCMV, and KSHV) were used for intra-species evolutionary analysis. Functional data used in the analyses (e.g. protein type and expression time during replication) was collected based on the human-infecting viruses. Structural analyses (e.g. fold complexity and disordered content) were mostly based on the human-infecting viral proteins. Silhouette images were derived from https://www.phylopic.org/.

for each viral protein: fold complexity, surface fraction, and disorder fraction. To compute these measures, herpesvirus protein structures were first predicted using AlphaFold2 (AF) (Tunyasuvunakool et al. 2021). Overall, we predicted the structures of 73, 168, and 85 proteins of HSV1, HCMV, and KSHV, respectively. Residue structural precision accuracy (pLDDT), obtained from AF, is considered a good predictor of disorder (Akdel et al. 2022; Wilson et al. 2022; Zhao et al. 2023), and these values were thus used to determine the level of disorder for each residue and the fraction of disordered

residues per protein. Using the predicted protein structures, surface and core residues within each folded region were inferred using the contacts of structural units (CSU) method, a biophysics-based method that computes for every residue the interactions that its atoms form with atoms from other residues and the chemical nature of these interactions (Sobolev et al. 1999). The CSU method also enabled the identification of inter-connecting residues—residues that are not adjacent in primary sequence (are at least four residues apart from each other in sequence), but that their atoms interact with other in the tertiary structure. This provided us with a detailed list of all residue pairs that interact in noncovalent bonds within each protein. Using this list, a value related to fold complexity was calculated. To this end, we used a measure termed "contact order", which calculates the average distance in primary sequence between all residue pairs that form a contact in the three-dimensional structure of the protein (see scheme in Figure S2). Distant residues in primary sequence that come into contact in the three-dimensional structure increase the contact order value, while contacts between closely related residues in primary sequence decrease this value. In other words, contact order gives an estimate of how complex the folded structure is, since its value correlates with how much of the fold is composed of pairs of residues that come into contact from distant regions of the sequence. This measure has been previously shown to correlate with the kinetic rate at which proteins fold (Ivankov et al. 2003), and here we study it in relation to evolutionary rates.

In summary, for every viral protein we obtained comparable measures of (1) fold complexity, (2) the fraction of residues that are surface-exposed, and (3) the fraction of residues within disordered regions. These values appear in Table S6a to c for all proteins in each of the three viruses.

Conservation of Structural Characteristics in Herpesvirus Core Proteins across Genera

To compare the structural and evolutionary characteristics of herpesvirus proteins across the three genera, we first focused on a set of 39 core proteins whose orthologs exist as a single copy in each of the human-infecting herpesviruses used in this work (HSV1, HCMV, and KSHV) and that we obtained AF-predicted structures for each of the three orthologs. From these, 33 proteins had estimates of dN/dS in all three families. The list of core genes is based on a previous review that includes 44 proteins that exist in *Alphaherpesvirinae*, *Betaherpesvirinae*, and *Gammaherpesvirinae*, but not necessarily in all human-infecting viruses (Davison et al. 2009), and their Uniprot IDs and structural and evolutionary measures appear as Table S7. We compared the fold complexity, as measured by contact order, between the orthologs of core proteins and observed a very high and significant correlation between their contact order values. We note that in this analysis, and in all following analyses of core proteins, we compare the three viruses in a pairwise manner, each time comparing the same set of orthologous proteins in two of the three genera we study (Fig. 2a to c). The strong correlation, of $\varrho = 0.95$ or above, between contact order values of ortholog proteins suggests that the overall structure of the fold has remained highly similar in the three genera, despite the large separation in time between the three viruses (Fig. 2a to c).

Next, we compared the fraction of surface residues between orthologs and observed that it is also highly correlated, although the correlation is lower than that observed in the fold complexity comparison (Fig. 2d to f). This lower correlation is in agreement with changes in surface residues being more prevalent during protein evolution than those occurring in residues in the protein's hydrophobic core (Tóth-Petróczy and Tawfik 2011). These evolutionary changes in protein surfaces can lead to larger differences in surface exposure between orthologous proteins in comparison with the overall fold structure and core residues that are under stronger selection to be conserved, to preserve function and maintain thermodynamic stability (Tóth-Petróczy and Tawfik 2011).

We repeated the above analyses, comparing structural features (fold complexity and surface percentage) between core proteins, with proteins that their predicted structures have a relatively high prediction score (average pLDDT score > 70). The trends observed in this protein subset were similar to the overall set, but with weaker correlations (Figure S3). These weaker correlations may originate in the smaller set of proteins analyzed and in the narrower range of values compared (e.g. in the surface percentage analysis, proteins with very high surface residue fractions were excluded from the analysis, since their pLDDT values were low, likely because they possessed high fraction of disordered regions).

When considering the disorder fraction in each protein, a significant correlation between orthologs is observed (Fig. 2g to i). The correlation in disorder fraction was lower than the correlations observed when fold complexity and surface percentage were compared across this set of proteins. This is in line with the relatively fast evolution of IDRs that are known to change rapidly in both sequence composition and in overall length during evolution (Xue et al. 2012; Gitlin et al. 2014). It is also in agreement with our previous results that suggested significant differences in disorder content among herpesviruses (Mozzi et al. 2020b).

Conservation of Evolutionary Rates in Herpesvirus Core Proteins

Following the structural measure comparisons, we asked whether the evolutionary rates at which orthologous gene evolved in each of the three genera separately, are similar. In other words, we asked whether the orthologs of a core gene that evolved relatively fast in one genus across primate-infecting viruses, tend to evolve rapidly also in the other two genera, and whether a slowly-evolving gene in one genus also evolves relatively slowly in the other genera. To this end, we compared the dN/dS values obtained for the same gene, between the three genera. In each genus, the evolutionary rates were computed across primate-infecting herpesviruses (e.g. herpes simplex viruses that infect humans, great apes, and Old World monkeys). In two of the three comparisons (HCMV vs KSHV, and HSV1 vs KSHV) we observed a significant correlation (with P-value = 0.01 or below, Fig. 2j to l). This suggests that these orthologous proteins are under similar selective constraints and that selective patterns tend to be maintained over long evolutionary time frames, as expected by the high degree of structural similarity observed between orthologs of core herpesvirus genes. This is despite the differences in environments between these viruses (e.g. different cell tropism, different host proteins that act as restriction factors against these viruses, etc.) and the fact that they have split from a common ancestor millions of years ago. We note that the weak correlations observed in some of these comparisons may stem from the limited numbers of proteins analyzed, from the limited numbers of orthologous proteins used to infer evolutionary rates, and from the relatively narrow range

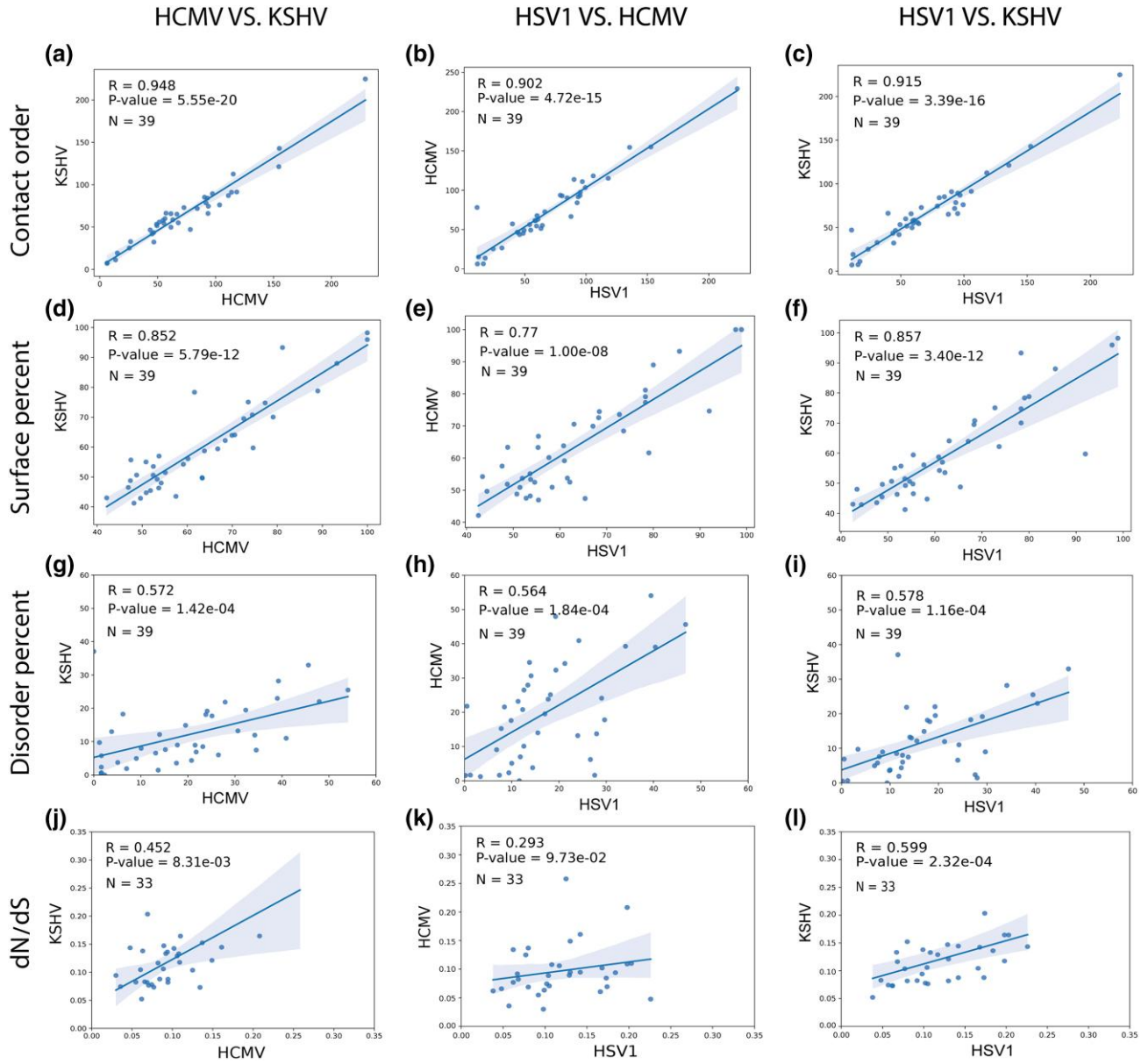


Fig. 2. Comparison of structural and evolutionary measures between core proteins belonging to the three *Orthoherpesviridae* genera. In each row, a different measure is compared between orthologs of core proteins, found in HSV1, HCMV, and KSHV (in each column, orthologs from two genera are compared). The measures are (a to c) Contact order, (d to f) Surface percent, (g to i) Disorder percent, (j to l) dN/dS within each genus (across human- and NHP-infecting viruses). Spearman's rank correlation coefficients and *P*-values are shown.

of dN/dS values across proteins in each of the viral proteomes. Thus, while we believe that the repeating trends observed across viruses can alleviate concerns regarding statistical significance, these results should still be interpreted with caution.

We next performed a similar analysis, comparing the rates of orthologous genes in the different genera, but this time focusing on short-term evolutionary timeframes. When looking at the fractions of residues under diversifying selection or purifying selection, based on the intra-species analysis of the human isolates, no significant correlation was observed between the orthologs across the three genera (Figure S4). This is in line with the notion that in a relatively short evolutionary time frame, selection may be less efficient and that some mildly deleterious mutations have accumulated and are maintained since there was insufficient time to eliminate them from the population.

Structural Characteristics of Herpesvirus Proteins and Their Link with Evolutionary Rates

We next asked how each of the structural characteristics correlates with the evolutionary rates of herpesvirus proteins, to detect potential links between them. For this, we analyzed the relationship between the three structural characteristics mentioned in Fig. 2 and their inter-species evolutionary rates for all proteins within each of the three viruses (not only core proteins, but all proteins that have a sufficient number of one-to-one orthologs within the family across NHP-infecting viruses). First, we divided proteins into groups based on their relative contact order values, corresponding to fold complexity (Fig. 3a to c). We observed a strong and significant trend with higher contact order values having slower evolutionary rates. This is in agreement with the notion that more complex folds have higher fractions of residues placed

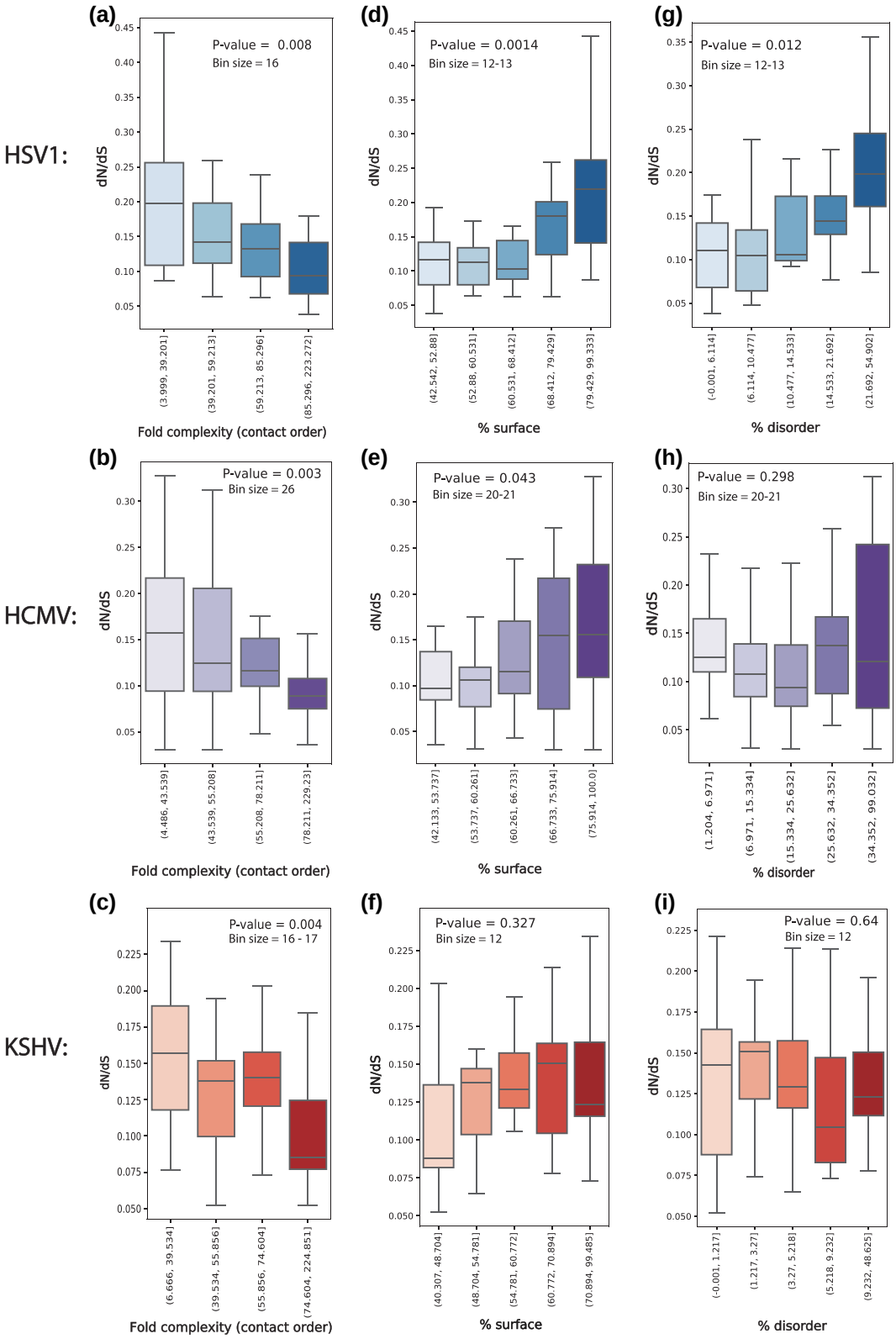


Fig. 3. Structural characteristics and inter-species evolutionary rates of herpesvirus proteins in the three genera. In each row, the evolutionary rates of proteins encoded by a different herpesvirus (HSV1, HCMV, and KSHV) with respect to NHP-infecting viruses within their genera are examined. In each column, the proteins are partitioned into bins based on a different structural characteristic: (a to c) Fold complexity, as measured by contact order, (d to f) Surface residue percent, (g to i) Disorder residue percent (taking into account only unmasked residues in the MSA). *P*-values are shown for Kruskal test. Dunn's tests for each pairwise comparison are shown in Table S8.

under constraints to preserve long-distance interactions between residues than simpler folds that are mostly formed through local interactions. We note that there is some overlap between the groups, such that some of the proteins that differ in their fold complexity (contact order values) have similar evolutionary rates (dN/dS values). Furthermore, the trends are not always consistent, as observed in the middle protein groups of KSHV. This may stem from other factors that

impact the evolutionary rates, in addition to fold complexity, and from technical factors such as noise in the data and limited number orthologs that can lead to less accurate inference of evolutionary rates.

We repeated this analysis by grouping proteins based on the fraction of surface residues, observing a trend where proteins with higher fractions of surface residues display faster evolutionary rates (Fig. 3d to f). While this trend is expected and observed across all three viruses, it is less consistent than the trends observed when looking at the relationship between fold complexity and evolutionary rates.

We repeated the two analyses above with the set of proteins with high prediction scores (pLDDT score > 70). We observe similar patterns, where proteins with higher fold complexity and lower surface residue fraction tend to be more conserved (Figure S5). However, these trends are weaker in this subset, likely because of the smaller set of proteins tested and their narrower range of values.

Finally, we compared the fraction of disordered residues with the protein's evolutionary rate (Fig. 3g to i). Surprisingly, we did not find a clear trend in HCMV and KSHV. In HSV1, we observed a trend of faster evolutionary rates in proteins that have higher disorder content, in agreement with previous findings (Mozzi et al. 2020b) and with the notion that disordered regions tend to evolve faster than ordered regions (Chow and Toth-Petroczy 2025). We expand on these findings in the context of functional constraints that may be placed on disordered regions in Discussion.

Functional Characteristics of Herpesvirus Proteins and their Associations with Evolutionary Rates

We next asked whether there are associations between protein evolutionary rates and their functions. We used a dataset of herpesvirus proteins that partitions them based on (i) their time of expression during replication [immediate early (IE), early, delayed early (DE), late, leaky-late], and (ii) the compartment to which they belong to—structural versus nonstructural, and, within the structural proteins, a further division to capsid, tegument and envelope (Kennedy et al. 2022). We compared the distributions of evolutionary rates of each of these groups to the others, in each of the three viruses. In two of the viruses, HSV1 and HCMV, we observed that IE genes are significantly more rapidly evolving, displaying higher dN/dS values than other classes. In KSHV we do not observe such trends and the three groups of proteins do show significant differences in the distribution of evolutionary rates (Fig. 4a to c).

When looking at proteins' structural function (Fig. 4d to f), we observed that capsid proteins display the lowest evolutionary rates, while envelope and tegument proteins display the highest evolutionary rates. This is observed in all three viruses, but is significant only in some of the comparisons: capsid versus tegument proteins in HSV1 and KSHV, and capsid versus envelope proteins in HSV1. In both functional and temporal classifications, we thus observed complex trends. These may stem from other factors that constrain these proteins, such as their structural characteristics. Additionally, it may be that the classifications are too crude, and include groups that have several subsets with different functions that can affect their evolution. For example, tegument proteins include both those residing in the inner and the outer tegument regions, and these are known to have different functions that may impact their evolution (Sucharita et al. 2023). Next, we asked if the evolutionary rates of latency genes differ from other classes of viral genes. We observed that both in HCMV and KSHV, the small group of latency genes [taken from a recent review on herpesvirus latency (Weidner-Glunde et al. 2020)] has higher evolutionary rates (Figure S6). However, the differences between these genes and nonlatency genes are not statistically significant, possibly due to the small size of this group of genes.

Next, we hypothesized that viral proteins localized into more than a single location would be targeted by stronger purifying selection. This is based on the assumption that their multiple locations may indicate that they are multifunctional, and having several functions can place additional constraints on the protein. Indeed, when using a dataset based on a cross-HSV1 protein localization fluorescence assay (Xing et al. 2011), we observed that proteins found in multiple cellular locations seemed to have lower dN/dS values than those localized in one cellular location (Figure S7). However, these differences are not significant because of the small numbers of proteins compared in each of the groups.

Finally, we also asked whether an additional characteristic, protein's tendency to form or be part of a biomolecular condensate, is associated with its conservation. Biomolecular condensates have recently emerged as having central roles in various cellular pathways, and many viral proteins are thought to take part in such condensates in order to modulate the host's cell and to efficiently replicate (Li et al. 2022; Zhang et al. 2023), including some herpesvirus proteins (Seyffert et al. 2021; Xu et al. 2021; Zhou et al. 2023). However, the relationship between participation in biomolecular condensates and protein conservation is not well understood and has not been explored, in either cellular or viral proteins. For this analysis, we predicted the tendency for each herpes viral protein to be part of a condensate, using PICNIC—a recent machine learning algorithm that predicts protein's participation in condensates based on sequence and structure (Hadarovich et al. 2024). We observed that in each of the three herpesviruses, the group of proteins predicted to form condensates is under stronger purifying selection than the rest of the proteins (Fig. 5a to c). We note that these comparisons are weakly significant (in the case of HCMV) or not statistically significant (in the case of HSV1 and KSHV), possibly due to the small group of proteins compared or to the small differences observed between some of the compared groups. However, the fact that these trends are consistent across the three viruses may suggest that participation in biomolecular condensates places constraints on the sequences of the viral proteins that are part of these condensates. This is especially surprising since proteins that form condensates often include segments of disordered regions known to rapidly evolve (Chow and Toth-Petroczy 2025). When testing the disorder content of the herpesvirus proteins, partitioned based on whether the proteins are predicted to participate in such condensates, we indeed observed that the group of proteins predicted to form condensates includes proteins with higher fraction of disordered regions than proteins not predicted to be part of condensates. This trend is consistent across the three viruses, and is significant in HCMV (Fig. 5d to f). To further investigate the relationship between the tendency to form condensate, disorder content and protein conservation, we used a linear model (see Methods). We observed that for two of the viruses, HCMV and HSV1, the condensate prediction scores contribute towards lower dN/dS values and the disorder content has the opposite effect (although not fully significant) (Fig. 5g).

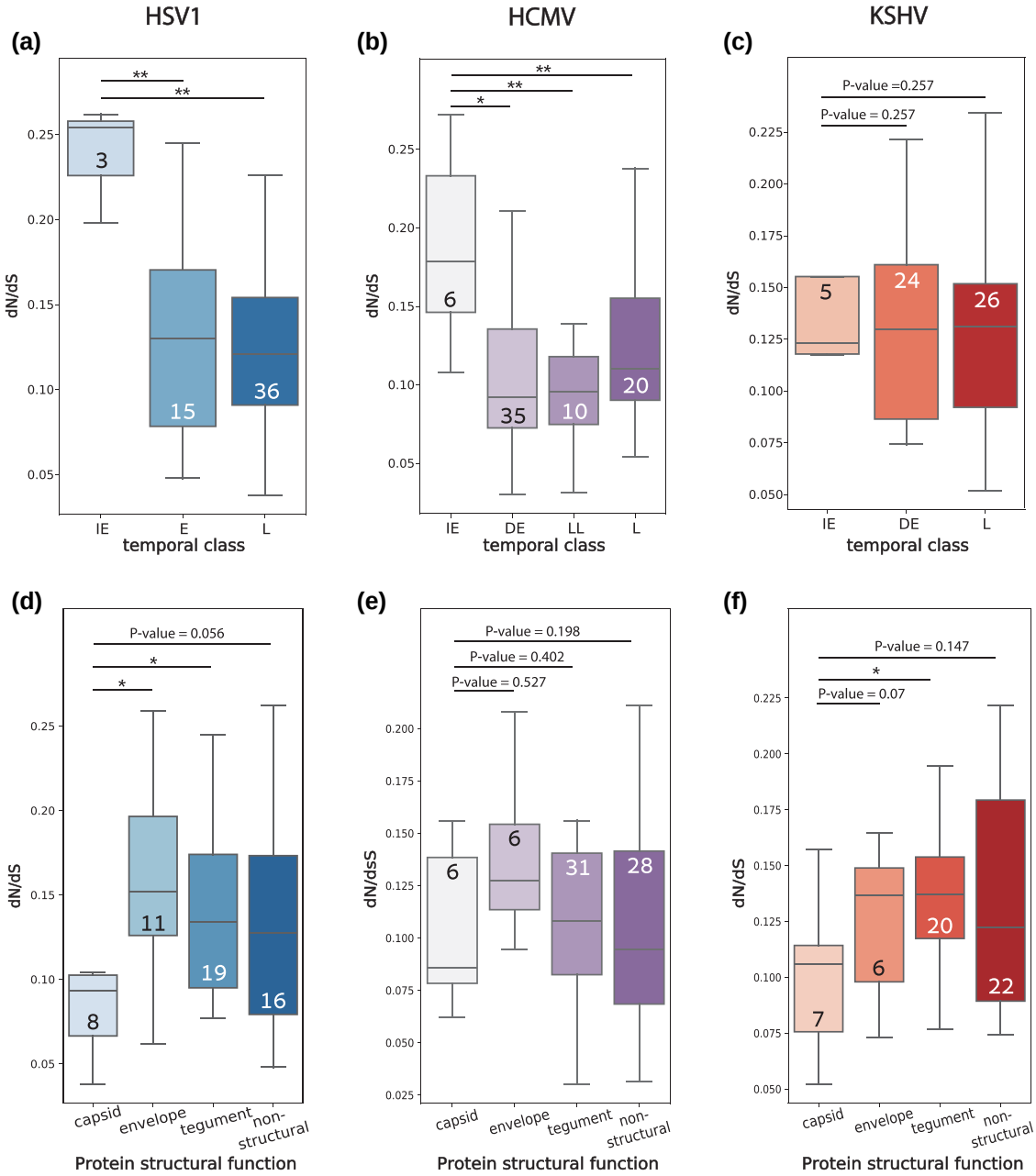
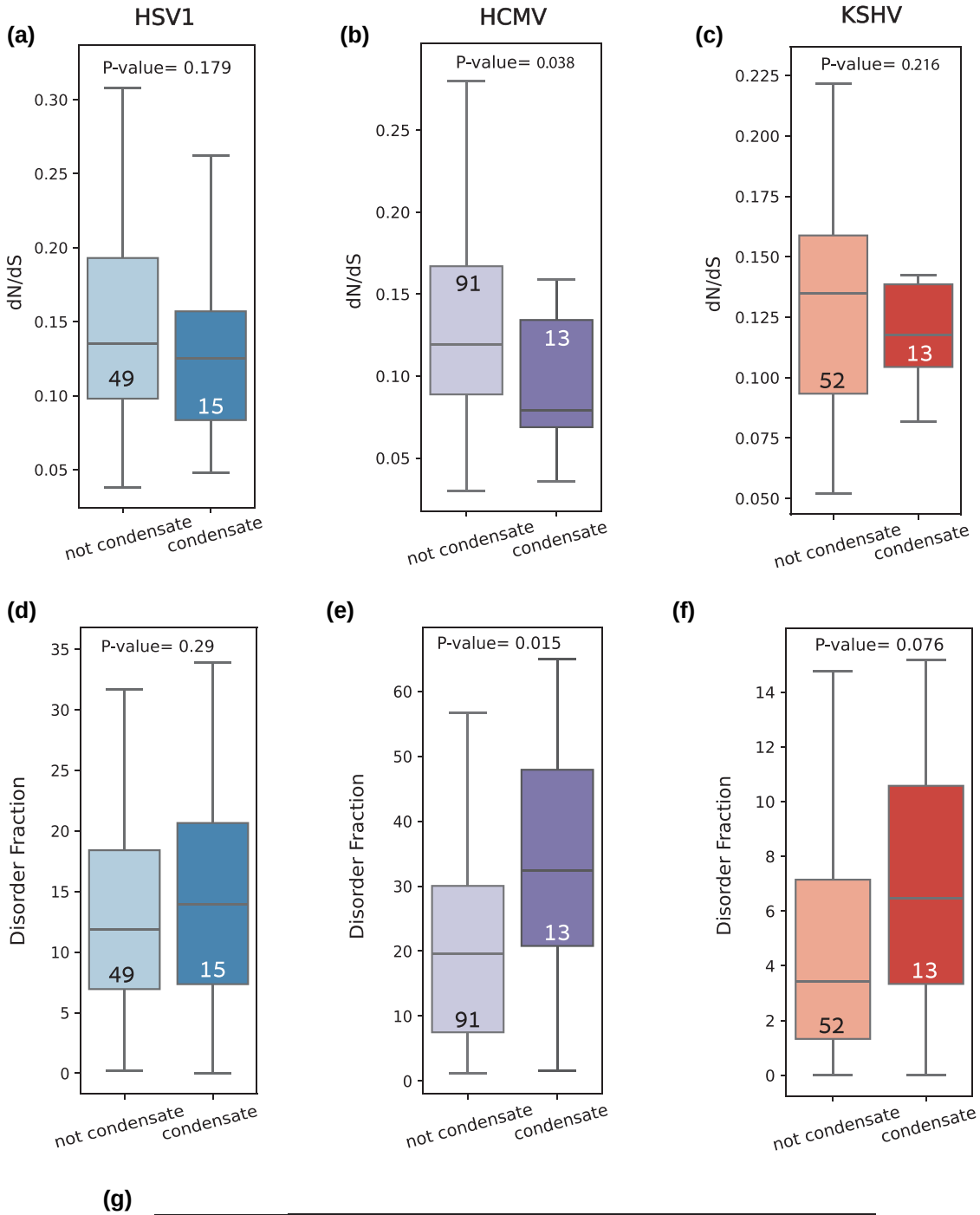


Fig. 4. Functional characteristics and inter-species evolutionary rates of herpesvirus proteins in the three genera. In each column, the evolutionary rates of proteins encoded by a different herpes virus (HSV1, HCMV, and KSHV) with respect to NHP-infecting viruses within their genera are examined. In each row, the proteins are partitioned into bins based on a different characteristic: (a to c) Temporal class, the time during infection in which the viral protein is expressed: Immediate-Early (IE), Early (E), Delayed-early (DE), Leaky-Late (LL), Late (L); (d to f) Structural function, with respect to whether the protein is structural and which part of the virion it is part of, or a nonstructural protein. FDR-corrected *P*-values are shown for Mann–Whitney test.

Residue-Level Analysis of Surface, Core and Disorder

Following our protein-level analyses, we next asked how residues with different structural characteristics differ in their evolutionary rates. We here used the intra-species analysis since in the inter-species analysis a large number of residues were masked due to recombination events or to poor alignments (see Materials and Methods). We first divided each of the proteins' residues based on their structural location: core or surface, and then studied their distribution in four classes of residues inferred to undergo different selection, as determined by FEL (Kosakovsky Pond and Frost 2005). The classes include: diversifying (positive), invariable (where the site has no synonymous or nonsynonymous substitutions in any of the species), neutral, and purifying (negative) selection. Using these classifications, we observed strong enrichment of surface residues within the diversifying class in all three viruses. In addition, in both HSV1 and HCMV, the purifying class is depleted in surface residues and enriched in core residues (Fig. 6a to c). We observed similar trends when repeating this analysis on residues from proteins with high pLDDT scores (Figure S8).

We next asked whether residue disorder is linked with its evolutionary rate. To this end, we used AF measure of accuracy in prediction as a proxy for disorder [the per-residue pLDDT (predicted local distance difference test) score], as done previously (Wilson et al. 2022). Based on this score, residues were classified as ordered, intermediate and disordered



(g)

	Disorder Fraction		Condensate Prediction Score	
	Coefficient	P-value	Coefficient	P-value
HSV1	0.0033	0	-0.1102	0.019
HCMV	0.0006	0.053	-0.1072	0.001
KSHV	2.47E-05	0.974	-0.0192	0.518

Fig. 5. Protein participation in biomolecular condensates, disorder and conservation. (a to c) Evolutionary rates of herpesvirus proteins (dN/dS), partitioned based on whether they are predicted to be part of a condensate or not, based on PICNIC scores, for HSV1, HCMV, and KSHV proteins, respectively. (d to f) Disordered residue percent of herpesvirus proteins (taking into account only unmasked residues in the MSA), partitioned based on whether they are predicted to be part of a condensate or not, based on PICNIC scores, for HSV1, HCMV, and KSHV proteins, respectively. g) Results of a linear model for each of the three viruses, with protein condensate formation prediction scores and disordered fraction as independent variables and dN/dS values as dependent variables. *P*-values in (a to f) are shown for Mann–Whitney test.

(with pLDDT values above 70, between 50 and 70, and below 50, respectively). In all three viruses, we observed that the diversifying class is highly enriched with residues that are disordered and is depleted in ordered residues, as expected based on previous findings (Brown et al. 2011; van der Lee et al. 2014) (Fig. 6d to f). In summary, as expected from

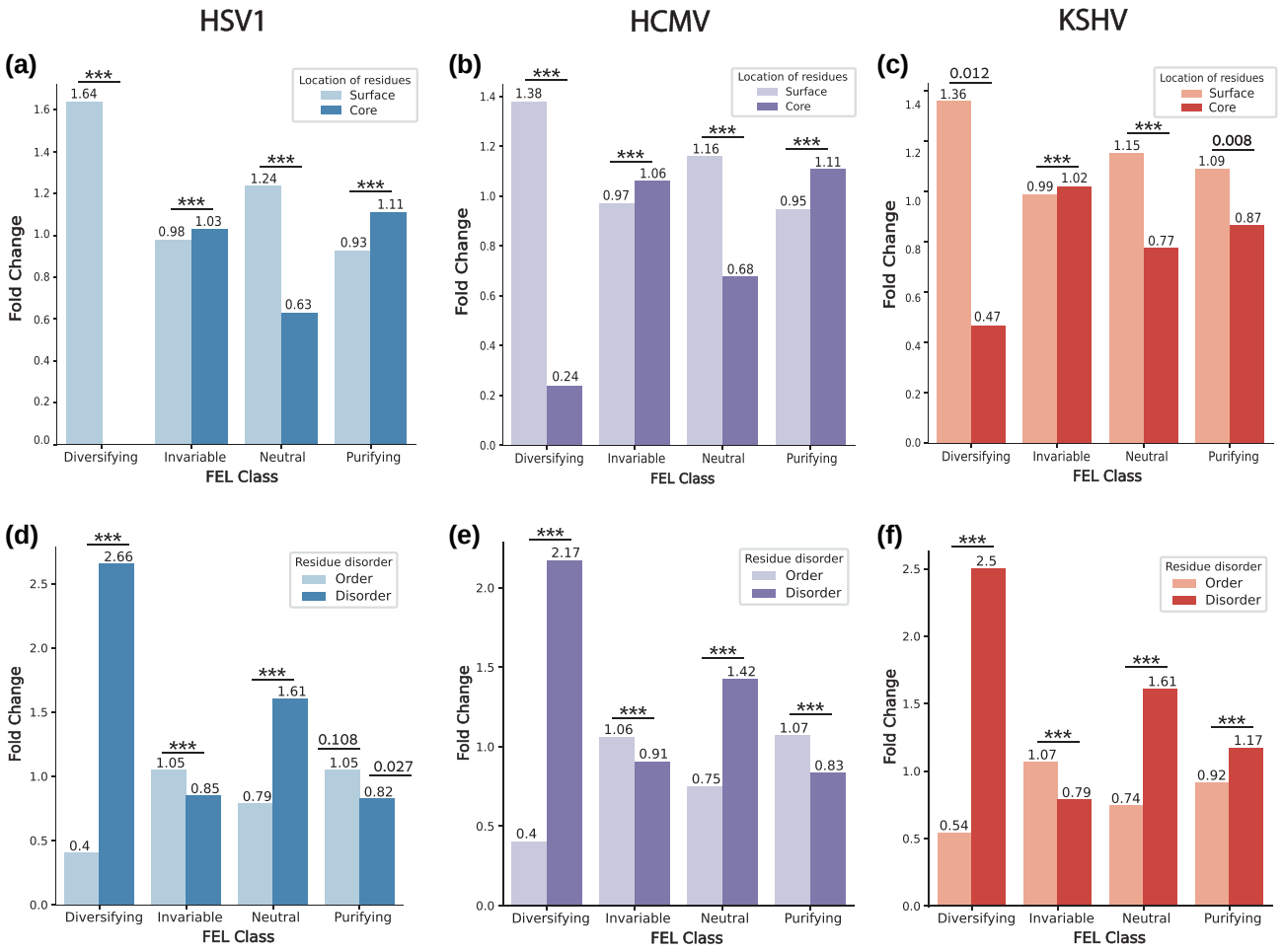


Fig. 6. Structural characteristics of residues and intra-species evolutionary rates of herpesvirus proteins in the three genera. In each column, the evolutionary rates of protein residues from a different herpes virus (HSV1, HCMV, and KSHV) are shown. Evolutionary patterns are computed based on the FEL method and using isolates from the same virus, and each residue is classified as belonging to the groups of residues that have undergone diversifying, neutral or purifying selection, or as an invariable residue. In each row, the residues are partitioned into bins (shown as fold change values of relative occurrences) based on a different characteristic: (a to c) Surface versus core residues; (d to f) Ordered versus disordered residues. Intermediate disordered residue fractions are not shown. FDR-corrected *P*-values are shown for Fisher's exact test comparing the two adjacent columns in each FEL class. In (d to f), two tests were performed (once for ordered versus all and separately for disordered versus all. Both tests yielded similar *P*-values, except for HSV1 purifying selection class, where both values are shown).

previous analyses, disordered and surface regions are enriched with positive selection, while ordered and core regions are enriched with purifying selection.

Evolutionary Conservation of Short Peptide Motifs in Herpesvirus Proteins that Interact with Host Protein Domains

Finally, we aimed to test whether surface regions of viral proteins that act as sites for interactions with host proteins are relatively conserved or evolve rapidly. While evolution of extracellular viral protein surfaces has been studied, especially with respect to antigenic evolution (Kistler and Bedford 2023), not much is known about how surfaces of intracellular viral proteins that are used for interactions with host proteins evolve. Since structural information on host-virus protein-protein interaction (PPI) is scarce, with relatively few experimentally solved host-viral protein complexes, we sought to find interface regions by focusing on domain-motif interactions between host and viral proteins. Several motifs found in herpesvirus proteins that interact with specific human protein domains, are known to play important roles during the virus life cycle. These motifs act as interaction modules with host proteins, and can thus be used to study the evolution of interface regions in viral proteins. When looking at the set of experimentally known motifs in proteins from human-infecting herpesvirus proteins, we can analyze their occurrence in orthologous viral proteins, to infer whether the host-virus interactions they mediate are conserved across the viral phylogeny (see Table S9, for the complete list of experimentally known motifs). For example, a motif found at the disordered C-terminus tail of HSV1 UL37 tegument protein is important for binding the human TRAF6 domain and for NF-κB activation (Liu et al. 2008). This motif is found in all orthologous proteins in NHP-infecting herpes simplex viruses (Table S9). Additionally, the HCMV tegument protein pp150 binds to the human Cyclin A2 through a specific motif (Bogdanow et al. 2013). This motif is found in orthologous proteins from Great ape- and Old World monkey-infecting CMVs, but not in orthologs from New World monkey-infecting CMVs (Table S9).

We asked whether such functional motifs in viral proteins tend to be maintained in viral orthologs such that the host-virus interactions mediated by them are conserved over long

evolutionary timescales. Since the set of experimentally validated motifs in herpesvirus proteins is relatively limited, we sought to computationally infer likely cases of such functional motifs. This can be done by (i) taking host-viral protein pairs known to experimentally interact (from host-virus PPI databases), and (ii) finding a matching pair of a viral motif and a host domain, e.g. by finding an SH3-binding motif within a herpesvirus protein's disordered region and an SH3 domain in the human protein known to interact with this viral protein (see Figure S9 for a schematic example). Such an approach was previously used by us and by other groups (Garamszegi et al. 2013; Becerra et al. 2017; Shuler and Hagai 2022; Cagliani et al. 2024), and allows the detection of the surface region in viral proteins that interacts with the host proteins.

To detect motifs in viral protein sequences, we first identified segments of HSV1, HCMV, and KSHV proteins that match known motifs, as defined in the ELM database (Kumar et al. 2024) that their residues are predicted to be disordered. Some of these motif occurrences may represent a truly functional motif that binds a matching host domain, while other occurrences can be sequences that spuriously match motif patterns due to the short and degenerate nature of motifs ("motif-matching sequences"). To distinguish between occurrences of functional motifs and spurious occurrences of motif-matching sequences, we used existing PPI data between herpesvirus proteins and human proteins [taken from several experimental studies (Nobre et al. 2019; Yang et al. 2021)]. In each pair of human-viral PPI, we considered motifs in the viral protein as "functional motifs" if they had a matching domain in the interacting human protein (see Figure S4 for details). Thus, for HSV1, HCMV, and KSHV we obtained a list of 45, 118, and 81 functional motifs, respectively. The remaining set of motif matches that lack experimental evidence are thus not known to be functional and were denoted as "motif-matching sequences".

To determine the evolutionary conservation of motifs, we searched for the presence of each motif in the orthologous sequences in NHP-infecting herpesviruses. We then defined a "conservation status" for each motif, ranging from occurring only in the human-infecting herpesvirus to occurring in all orthologous viral proteins (see Fig. 7a for an example, and Fig. 1a for detailed viral phylogenies). Importantly, since motifs usually reside in disordered regions that can be poorly aligned, and since motifs can function despite some changes in their residue composition (based on the motif patterns), we did not require that the motifs will be found in the exact aligned position across all orthologous proteins and that their residues will be highly conserved. Instead, following previous analyses (Nguyen Ba and Moses 2010; Hagai et al. 2012b, 2014), we asked whether the motif occurs or not in each such ortholog. We observed that ~80% of the functional motifs in all three herpesvirus genera occur across all primate-infecting viruses (Fig. 7b to d). Furthermore, in all three herpesviruses the set of functional motifs tend to be more conserved in occurrence across related viruses than motif-matching sequences: The fraction of motifs that occur across the viral orthologs (out of all motifs) is higher in the set of functional motifs than the corresponding fraction in motif-matching sequences. In contrast, a higher fraction of motifs that occur only in the human-infecting viral protein (HSV1, HCMV, and KSHV) is found in motif-matching sequences than the corresponding fraction in functional motifs. These observations suggest that functional interface regions in viral proteins, such as the short linear motifs we examined, occur in a relatively conserved manner during viral evolution. This conservation is observed across all viruses examined. We further discuss these findings and their relevance for host-virus interaction evolution in the Discussion section below.

Discussion

In this work, we have analyzed the short- and long-term evolution of herpesvirus proteins using three human-infecting viruses, HSV1, HCMV, and KSHV, from each of the three subfamilies of *Orthoherpesviridae*. We compared the evolutionary rates of encoded genes within each of these viruses using orthologous genes in NHP-infecting viruses from the same genus, and used these inter-species evolutionary rate estimates to test how a range of structural and functional characteristics is associated with fast or slow evolution. We further compared core proteins, whose encoding genes are found in each of the three distant genera we used. In these cross-genus analyses, we tested the similarity of orthologous core genes found in the three distant viruses in terms of their structural properties, long- and short-term evolutionary rates. Finally, we also focused on specific regions of herpesvirus proteins, examining how structural features of residues are related to their evolutionary rates and at the cross-viral-species occurrence of motifs that are important for interactions with the host. In general, our data indicate that purifying selection is the major force shaping the diversity of these proteins, and, as a consequence, the range of dN/dS over all proteins tested is limited. Thus, a shift between the median dN/dS values of two groups of genes (e.g., between capsid and tegument proteins, in Fig. 4), indicates a trend where purifying selection was stronger in the set with the lower values, and suggests that overall, this set, with its particular functional characteristics, may be more constrained.

We observed that the core proteins display strong similarity between orthologs in the three different viruses in various structural parameters we tested, in line with their conserved functions. Fold complexity was more conserved across orthologous core proteins than the fraction of disordered and surface regions (Fig. 2). This is in agreement with the requirement to maintain overall fold structure and stability and with the notions that surface and disordered regions are placed under less constraints and that their rapid evolution can support rewiring of protein regulation and interactions. When we tested the rates at which core proteins evolve within primate-infecting viruses in the three genera, we observed significant correlations in evolutionary rates in two of the three cross-genera comparisons (Fig. 2j to l). In other words, orthologs of core genes, separated by millions of years of viral evolution, show similar levels of conservation with respect to other core genes. This interesting result of relatively high conservation across genera is despite the long evolutionary time since separation of the three *Orthoherpesviridae* subfamilies and the fact that these viruses evolved to replicate in different cells and face different environments and challenges. It implies that the evolution of these genes is predominantly constrained by factors mutual to all viruses, most likely structural and biophysical factors, such as those we tested and others that place various limitations on protein evolution (DePristo et al. 2005; Echave and Wilke 2017).

Indeed, when we analyzed the association between evolutionary rates of all proteins encoded within each viral genome and various characteristics, the strongest correlations were observed when looking at structural properties (Fig. 3). When we

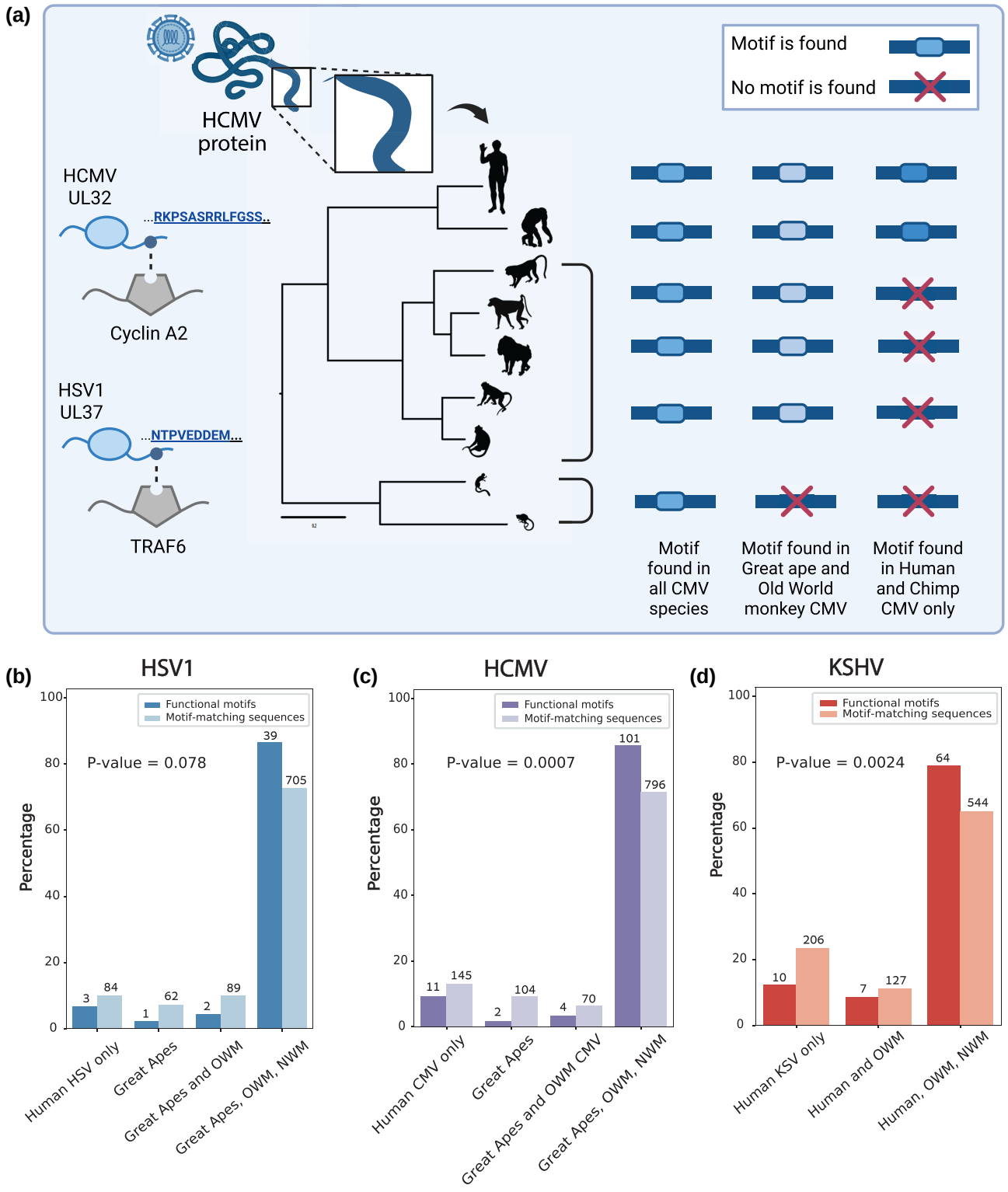


Fig. 7. Motif conservation across orthologous herpesvirus proteins in the three genera. a) Center: Schematic design of the analysis: In each of the three human-infecting viruses, we identified motif-matching sequences and inferred which motifs may be "functional" based on existing PPI data and on a method described in the text (the remaining set is denoted as "motif-matching sequences" and serves as a control). We then searched for the occurrence of these motifs in the orthologous proteins in the corresponding NHP-infecting herpes viruses. Based on their occurrences across the phylogeny, each motif was defined as found in all viral species, or only in a subset of the phylogeny (e.g. only in human and chimpanzee CMVs). Left: two examples of viral motifs and their human domain targets. The HSV1 protein motif is found across viral orthologs, whereas the HCMV protein motif is absent in New World monkey-infecting CMVs. (b to d) The fractions of functional motifs and motif-matching sequences, found in HSV1, HCMV, and KSHV, that are found (left-to-right) in only the human-infecting species, in subsets of the phylogeny, or conserved across the phylogeny. In all three viruses, functional motifs have a higher fraction of conserved occurrence across the viral lineage than the motif-matching sequences. To test this statistically, we compared the fraction of functional motifs occurring across all viruses, with the corresponding fraction of motif-matching sequences, using Fisher's exact tests and the entire sets of functional motifs and motif-matching sequences (*P*-values are shown for each of these comparisons, in HSV1, HCMV, and KSHV). Images were created in BioRender.(2025) https:// BioRender.com/1m65sw7. Silhouette images were derived from https://www.phylopic.org/.

tested the association between a range of functional characteristics with herpesvirus protein evolutionary rates, a more complex picture emerged (Figs. 4 and 5). Some functional features, such as early expression during replication and encoding for capsid and envelope proteins, showed evolutionary patterns that were often consistent and significant across the three viruses. In other cases, the results were weaker. The weak and complex patterns we observe may stem from a range of reasons including the relative strength of selection each of these characteristics effectively imposes on the protein, and the fact that these proteins may be multifunctional in a manner that can place additional or differential constraints on their evolution. In this regard, we observed that HSV1 proteins that are expressed in more than one cellular location, a potential proxy for multifunctionality, evolve under stronger purifying selection than those localized into a single location, although this was not significant, likely due to the small numbers of proteins compared. Since we do not have similar cellular expression data for the other two herpesviruses, we cannot test if the observations in HSV1 are part of a consistent trend across herpesviruses.

We also analyzed relative evolutionary rates of residues and regions within each protein (Fig. 6). We observed that surface and disordered regions are enriched with residues that underwent diversifying selection, in agreement with the notion that these regions may evolve rapidly and adaptively to rewire viral protein interactions with the host (Hagai et al. 2014; Mozzi et al. 2020b; Dolan et al. 2021). Interestingly, it was previously shown that a mammalian protein evolves adaptively in its disordered regions in a manner that may be related to the evolutionary arms race between primates and primate-infecting herpesviruses (Lou et al. 2016).

Host-pathogen interactions are usually thought to be antagonistic since they are beneficial for one of the interacting partners and deleterious for the other side (Daugherty and Malik 2012; Duggal and Emerman 2012; Tenthorey et al. 2022). These antagonistic interactions lead to an evolutionary arms race, where both sides evolve adaptively to counteract their opponent. Indeed, signatures of positive selection are often observed in host antiviral restriction factors that are thought to be a result of direct antagonistic interaction between the host factors and viral components (Barreiro and Quintana-Murci 2010; Daugherty and Malik 2012; Duggal and Emerman 2012; Cagliani et al. 2014; Sironi et al. 2015; Enard et al. 2016). From the virus side, the evolution of interface regions in viral proteins was studied in several cases (Shen et al. 2009; Woo et al. 2010; Sauter et al. 2012; Armijos-Jaramillo et al. 2020; Guo et al. 2020; Jo et al. 2021; Patrono et al. 2022). However, overall trends regarding the links between positive diversifying selection and interactions with the host are less characterized across the complex network of host-virus PPIs. A recent study, analyzing selection patterns in several human viruses, suggested that in almost a third of these viruses surface proteins display patterns consistent with antigenic evolution to escape antibody recognition (Kistler and Bedford 2023). However, how intracellular interactions between host and viral proteins evolve and how this is related to sequence evolution in host- and viral-interacting proteins is not well understood. Previous work, based on computational analyses (Hagai et al. 2014; Glavina et al. 2022b; Schuck and Zhao 2023) and experimental manipulation of viral sequences (Gitlin et al. 2014), suggested that host-like motifs can rapidly emerge in viral sequences in a manner that supports the formation of new interface platforms to bind specific host protein domains. In here, we analyzed motifs found in herpesvirus proteins that target specific human domains and their evolution across primate-infecting herpesviruses. We showed that they occur in a highly conserved manner across related primate-infecting herpesviruses and that they are more likely to be present across orthologous viral proteins than random sequence matches to these motif patterns (Fig. 7). These results are consistently observed in all three tested herpesviruses. This suggests that while motifs can rapidly emerge in viral proteins, once they are formed, they are often maintained for long evolutionary periods to support essential interactions with host proteins. These findings of motif conservation in viruses add another perspective to the evolutionary arms race between viral motifs and the host domains they target: We previously showed that these host domains and their motif-binding pockets are highly conserved across vertebrates and that this conservation likely stems from the requirement to preserve these binding pockets for interactions with numerous host proteins (Shuler and Hagai 2022). In here, we show that viral motifs occur across viral species in a conserved manner. Thus, the interface regions from both the host and the viral sides are locked in a stasis due to constraints placed on both sides to maintain these interface regions. This is in agreement with previous suggestions regarding long-term conservation of viral proteins that may be associated with their adaptation to their host (Simmonds et al. 2019).

In that regard, we also observed that proteins predicted to be part of biomolecular condensates have lower dN/dS values than those proteins not predicted to form such condensates, although this was significant in only one of the three viruses (Fig. 5). This is despite the fact that these proteins tend to contain disordered regions, and may point to the constraints imposed on these proteins to maintain specific sequences that support condensate formation. Condensate formation can involve protein interactions through disordered regions (Woodruff et al. 2018; Villegas et al. 2022). Thus, our results regarding motif conservation and conservation of proteins predicted to form condensates, may be related to functional elements within disordered regions and how they may constrain the evolution of these regions, a concept that is gradually emerging from various studies (Chow and Toth-Petroczy 2025).

It is important to note that our analysis has several limitations and caveats: Most importantly, the number of genes in each of the analyzed groups is often relatively small (with some groups composed of a dozen or a few dozen genes). These relatively small-sized groups may yield noisier comparisons and may suffer from a few outliers that skew the distribution. Comparison between such groups can be further complicated when the values compared have a relatively small range (as in the case of the inter-species dN/dS values), or when only some of the genes within these groups have known values (such as in functional annotations of viral genes). Additionally, some of the characteristics we studied are based on measurements that may themselves be noisy (such as microscopy-based protein localization in different cellular compartments) or that their inference is complex, such as in the case of condensate formation, where the various parameters that contribute to this phenomenon are not completely understood. One approach we took to alleviate these issues is to perform these analyses across three different herpesviruses. In cases where consistent trends across the three viruses were observed, this can strengthen the possibility for an

association between the two parameters we tested, even if the results are significant in only some of the viruses. We note, however, that in all cases where the results are weakly or non-significant, their interpretation should be taken with caution. These analyses may be strengthened if similar results will be observed in future analyses of unrelated dsDNA viruses, such as poxviruses and adenoviruses.

In summary, our analysis encompasses herpesvirus protein evolution at multiple timeframes and across different lineages, to find which functional and structural characteristics are associated with evolutionary rates. Our residue- and region-level analyses further point to constraints placed on structured and core protein regions as well as to the long-term conservation of interface modules that mediate host-virus PPIs. These results delineate evolutionary constraints of viral proteins and specific regions within them that can be used as targets for designing antiviral treatments and in studies that assess virus evolvability potential.

Materials and Methods

Virus Datasets

Viral genome sequences were retrieved from the National Center for Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov/). Only complete or near-complete genome sequences were included in this study. Detailed lists of viral genomes are reported in Tables S1 to S5.

For inter-species analyses (dN/dS calculation), we used the coding sequences from all available species of viruses that infect Old World monkeys, great apes, and humans (see Table S1). The number of species used was: simplexviruses—6, cytomegaloviruses—11, and rhadinoviruses—6. For motif search analyses, we used NHP-infecting herpesviruses representative of the diversity in primates plus the reference genomes of the human-infecting species: HSV1 (NC_001806), HCMV (NC_006273), and KSHV (NC_009333) (see Table S2).

For intra-species analyses, we selected subsets of human viruses. In the case of HSV1, we included 52 viral genomes derived from clinical isolates with a low-passage number in cell culture (<3 splits). We selected sequences sampled in different countries in order to obtain a dataset of viral genomes representative of the diversity among circulating strains (Table S3). For HCMV, we retained 72 isolates derived from different body compartments (i.e. amniotic fluid, urine, and blood/plasma) and from different geographic locations. Only isolates that were directly sequenced with no in vitro passages were included (Table S4). For KSHV, we selected 53 viral genomes from clinical isolates with no or low-passage history in cell culture, selected to be representative of viral genetic diversity in terms of geographic distribution and disease status. Details of all genomes included in the analyses are reported in Table S5.

Inter-Species Analyses

The dN/dS parameter was calculated using coding sequences from viruses that infect Old World monkeys, great apes, and humans. For these analyses, New World primate-infecting viruses were excluded because high sequence diversity and unreliable alignments can inflate the false positive rate in evolutionary inference (Mozzi et al. 2020a, 2022). Whole genome alignments were obtained using Progressive MAUVE 2.3.1 (Darling et al. 2010), through the graphical user interface, using default parameters. Analysis of selective patterns was performed only for coding genes with reliable one-to-one orthologs among different viral species. It included only genes with at least four orthologs. Orthologs were inferred by MAUVE and confirmed by previously reported analyses and genome annotations.

Gene alignments were generated using the GUIDANCE2 suite (Sela et al. 2015), setting sequence type as codons and using MAFFT (Katoh and Standley 2013) as an aligner. GUIDANCE2 also allows filtering of unreliably aligned positions and we removed codons with a score lower than 0.90 (Privman et al. 2012). Only alignments with at least 150 unmasked nucleotides were retained. The final number of genes included in each set was: simplexviruses: 65 genes; cytomegaloviruses: 101 genes, rhadinoviruses: 68 genes. Each alignment was screened for the presence of recombination using GARD (Kosakovsky Pond et al. 2006), a genetic algorithm implemented in the HYPHY suite (version 2.2.4). This method uses phylogenetic incongruence among segments in the alignment to detect the best-fit number and location of recombination breakpoints. When evidence of recombination was detected (P-value < 0.01) the partitioned alignment file was used as an input for dN/dS calculation, which was performed using the SLAC method (Kosakovsky Pond and Frost 2005). In various analyses, such as those shown in Figs. 4 and 6, we used the distributions of dN/dS values between sets of proteins with different characteristics, to compare selection patterns between these groups. Phylogenetic trees used as inputs for SLAC were generated with the PhyML program (version 3.1), by applying a general time reversible model (Guindon et al. 2009). PhyML was also used in the intra-species analysis that is described below.

Intra-Species Analyses

For the intra-species analysis, gene alignments and tree reconstructions were performed as described above. Selective patterns were inferred using the FEL method (Kosakovsky Pond and Frost 2005), that calculates the rates of nonsynonymous and synonymous changes at each site. FEL estimates the probability of selection at each site in an alignment through the dN-dS metric (rate of nonsynonymous changes-rate of synonymous changes), categorizing each codon into four classes: Diversifying, Purifying, Neutral, Invariable. Statistical significance is evaluated based on the asymptotic χ^2 distribution, using default parameters (P-value < 0.1).

Core Protein Classifications

Classification of proteins into core and noncore was based on conservation of genes across viral subfamilies and was obtained according to a previous analysis (Davison 2007). In our analysis, we considered as "core" only genes that are present in all the three subfamilies of mammalian herpesviruses (*Alpha-*, *Beta-*, and *Gamma-herpesvirinae*). A table with Uniport IDs and gene names of all core proteins appears as Table S6.

Structure Prediction and Fold Complexity Analysis

To predict structures of viral proteins we ran AF2 (Tunyasuvunakool et al. 2021) with default parameters and took for each protein the highest-ranking structure (with highest pLDDT score). Predicted structures were compared with those appearing in a different database (Nomburg et al. 2024), using the TM-align program (Zhang and Skolnick 2005) (see

Figure S10). We observed that proteins with higher prediction scores tend to be better aligned between the two predictions. This may be a result of greater variation in prediction of structures more challenging to predict or in greater conformational variation between structures that include larger fractions of disordered regions (which are related to lower overall pLDDT scores). To infer complexity of protein structures, we used a previously developed measure, termed "contact order" (Ivankov et al. 2003). This measure calculates the average sequence distance (in primary sequence) between pairs of interacting residues in the three-dimensional structures. In other words, this measure takes into account all the distances of the pairs of interconnected residues, with pairs that interact from more distant sequence positions increasing its value and indicating that the fold complexity is increased. Specifically, based on the predicted structures from AF, we inferred all contacting residues using the CSU software (Sobolev et al. 1999) [as previously done (Hagai and Levy 2008; Hagai et al. 2012a)]. Using this list of contact residues, and their position in primary sequence, we computed contact order as: $\sum C|i-j|/C$, where C is the number of contacts, and $|i-j|$ is the sequence separation between residues i and j, that were found to be in contact. As in previous studies, we ignored contacts between residues that their distance in primary sequence is below three residues.

Protein Disorder Analysis

Using AF (Tunyasuvunakool et al. 2021), residues were partitioned into ordered (pLDDT > 70), intermediate (50 < pLDDT < 70), and disordered (pLDDT < 50), following previous work showing that these scores give accurate disorder predictions. Because disordered residues are often not-well aligned, masked MSAs regions (due to poor alignment quality) may be enriched with disordered regions. Because we only used unmasked residues for dN/dS analysis, we used only unmasked residues in computing disorder fraction in relevant analyses (Figs. 3 and 5).

Surface, Core, and Interface Inference

Structural analysis was performed on protein structures in order to define residues as belonging to the surface or to the core of the protein using the CSU software (Sobolev et al. 1999). In addition, CSU solvation measurements were used to calculate the solvent accessible surface area (ASA). ASA was defined as the ratio of the solvent accessible surface for a given residue within the structured protein versus in the free-state of that residue. Residues were classified as core residues for ASA < 0.15, and as surface residues for ASA > 0.15, as previously done (Tóth-Petróczy and Tawfik 2011; Shuler and Hagai 2022). The structures used are the AF-predicted structures, as described above. While this analysis and the previous disorder analyses are described at the residue level, we also used them in several sections of this work at protein level (either averaging or estimating fractions of residues with particular characteristics).

Inference of Domain-binding Motifs in Viral Proteins

To study the evolution of host-like linear motif occurrence in disordered regions of viral proteins along the viral phylogeny, we first searched for linear motifs within viral proteins' sequences using regular expression matching (RegEx) taken from the ELM database (Kumar et al. 2024). Motif-matching-sequences were filtered out if according to AF their pLDDT score was above 50 (which are considered a measure for well-structured region). By first finding motif-matching sequences and then computing their average disorder scores (by using the motif residues' pLDDT values), we avoid problematic cases where motifs reside at the border between ordered and disordered regions, as well as circumventing the need to deal with noncontinuous disordered regions.

We used existing PPI databases of human-viral interactions described below, in order to classify certain motifs as "inferred functional motifs": That is—motifs that likely mediate domain-motif interactions based on our domain-motif matches in PPI data. In these cases, the motif originates from a viral protein that has an experimentally verified interaction with a human protein with a compatible domain (e.g. an SH3-binding motif in the viral protein and an SH3 domain in the human-interacting protein, where both proteins are known to physically interact, based on experimental studies). All other motifs are termed "motif-sequence matches", under the assumption that many of them represent random matches to the motif sequence. In addition, we used five experimentally validated cases of motifs in herpesvirus proteins, based on their annotations in the ELM database (Kumar et al. 2024). These cases appear in Table S8.

To identify viral proteins with experimental evidence for interaction with human proteins, we used the HVIDB database (16) (Yang et al. 2021) that includes 48,643 human-virus interactions originating from different databases and additional curations. These interactions are based on a combination of studies, including high throughput approaches, such as mass spectrometry and yeast-2-hybrid and more focused studies, involving methods such as X-ray crystallography and biochemical assays. We discarded interactions with missing protein information in UniProt or that are not curated as mediated by physical or direct interaction between the human and viral proteins, leaving 48,026 human-virus PPIs. Importantly for the CMV analysis, additional Human—HCMV interactions were obtained from a study (Nobre et al. 2019). It consists of 3,704 interactions that include both within-virus (HCMV-HCMV) and human-HCMV interactions, identified during viral infection. For KSHV, we also added interactions curated by BioGrid and IntAct, since we noticed that many of them were missing in the general host-virus PPI datasets we used. In total, these resulted in 708, 3,394, and 893 interactions between human proteins and HSV1, HCMV, and KSHV proteins, respectively.

Motif Occurrence in Orthologous Proteins of NHP-infecting Viruses

We used the above-identified set of motifs in HSV1, HCMV, and KSHV proteins that may be functional and mediate experimentally-validated interactions with human proteins through domain-motif interactions ("inferred functional motifs") and contrasted them with the set of motif matches that are not known to function in interaction with host domains ("motif-matching sequences"), to study the evolutionary occurrence of motifs (i.e. assuming that the matching sequences represent the background conservation). For this, we searched for motif matches in the NHP-infecting herpesviruses orthologous proteins. We only used viral proteins that had at least one ortholog in each "class" along the phylogeny. For example, in the case of viruses in the genus *Cytomegalovirus*, in addition to HCMV, we required to have orthologs from chimpanzee-infecting CMV, at least one Old World monkey-infecting CMV ortholog, and at least one New

World monkey-infecting CMV ortholog. We then defined the conservation status of these motifs, based on their occurrence. For example, in the case of CMV, we defined motifs as either "human CMV only", "human and chimpanzee CMV only", "Great ape and Old World monkey CMVs" and "conserved across all CMVs", based on their occurrence/absence across the phylogeny. In cases of inconsistencies, we chose to be strict and preferred the less-conserved option (e.g. a motif occurring only in human CMV and in a New World monkey CMV, but not in Great apes and Old World monkeys CMVs, was defined as occurring only in human CMV). We then plotted the distribution of conservation status of functional motifs versus motif-matching sequences. To test whether functional motifs tend to be conserved in their occurrence across the viral lineage, we compared the fraction of functional motifs occurring across orthologs out of the entire set of functional motifs, with the corresponding fraction of motif-matching sequences (those occurring across the phylogeny divided by the entire set of motif-matching sequences). The statistical significance of this comparison was obtained for motifs in each of the three viruses using Fisher's exact test (Fig. 7).

Viral Protein Classification based on Temporal Expression, Structural Classes and Latency

Classification of viral proteins based on temporal classes the timing of viral gene expression in the course of replication: IE, Early (E), DE, Leaky late (LL), Late (L) and based on structural classes: capsid, tegument, envelope and nonstructural proteins were taken from a previous study (Kennedy et al. 2022). Viral genes in HCMV and KSHV were classified as latency genes based on a previous review on latency genes in herpesviruses (Weidner-Glunde et al. 2020).

Viral Protein Condensate-Formation Analysis

We used the PICNIC program (Hadarovich et al. 2024) (https://picnic.cd-code.org/) to compute condensate-forming likelihood scores for each protein in the dataset based on their structural data (predicted structures). PICNIC provided a numerical score for each protein, indicating its potential involvement in biomolecular condensates. Proteins whose scores were above 0.5 were classified as "condensate-forming proteins". To determine the relationship between protein disorder, prediction to participate in a condensate and protein evolution, we used a linear model, using OLS (ordinary least squares) Regression from the statsmodels.api library. For each virus, the values of the fraction of disorder and PICNIC condensate prediction score per protein were used as independent variables, and the value of dN/dS was used as the dependent variable.

Cellular Location Analysis

We used a previous analysis of GFP-fused herpesvirus proteins that determined the cellular localization of HSV1 protein localization in the cell (Xing et al. 2011). Based on these results, proteins that were located in a single location were labeled as being in a "single location", and all others as "multiple locations".

Statistical Analysis

Statistical analyses (Mann–Whitney test, Kruskal test, Fisher's exact test, Spearman's rank correlation and FDR-correction based on the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) were performed using the SciPy package in Python (version 3.9).

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank Sivan Friedman-Nakar for helpful discussions and technical help.

Funding

This work was supported by the Italian Ministry of Health ("Ricerca Corrente" to D.F. and A.M.), by the Israel Science Foundation (ISF, grant no. 435/20 to T.H.) and by the Joint Canada-Israel Research Program (ISF, grant no. 2930/23 to T.H.).

Data Availability

M.S., MSA and tree files and PDB files of Alphafold-predicted structures are available on Zenodo: https://zenodo.org/records/15465046. Code generated during this study and command lines are available at: https://github.com/HagaiLab/herpes_evolution.

References

Afanasyeva A, Bockwoldt M, Cooney CR, Heiland I, Gossmann TI. Human long intrinsically disordered protein regions are frequent targets of positive selection. *Genome Res*. 2018:28:975–982. https://doi.org/10.1101/gr.232645.117.

Akdel M *et al*. A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol*. 2022:29:1056–1067. https://doi.org/10.1038/s41594-022-00849-w.

Armijos-Jaramillo V, Yeager J, Muslin C, Perez-Castillo Y. SARS-CoV-2, an evolutionary perspective of interaction with human ACE2 reveals undiscovered amino acids necessary for complex stability. *Evol Appl*. 2020:13:2168–2178. https://doi.org/10.1111/eva.12980.

Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet*. 2010:11:17–30. https://doi.org/10.1038/nrg2698.

Becerra A, Bucheli VA, Moreno PA. Prediction of virus-host protein-protein interactions mediated by short linear motifs. *BMC Bioinformatics*. 2017:18:163. https://doi.org/10.1186/s12859-017-1570-7.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol*. 1995:57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bogdanow B *et al*. Human cytomegalovirus tegument protein pp150 acts as a cyclin A2-CDK-dependent sensor of the host cell cycle and differentiation state. *Proc Natl Acad Sci U S A*. 2013:110:17510–17515. https://doi.org/10.1073/pnas.1312235110.

Brito AF, Baele G, Nahata KD, Grubaugh ND, Pinney JW. Intrahost speciations and host switches played an important role in the evolution of herpesviruses. *Virus Evol*. 2021:7:veab025. https://doi.org/10.1093/ve/veab025.

Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. *Curr Opin Struct Biol*. 2011:21:441–446. https://doi.org/10.1016/j.sbi.2011.02.005.

Cagliani R *et al*. RIG-I-like receptors evolved adaptively in mammals, with parallel evolution at LGP2 and RIG-I. *J Mol Biol*. 2014:426:1351–1365. https://doi.org/10.1016/j.jmb.2013.10.040.

Cagliani R *et al.* Evolution of virus-like features and intrinsically disordered regions in retrotransposon-derived mammalian genes. *Mol Biol Evol.* 2024:41:msae154. https://doi.org/10.1093/molbev/msae154.

Chemes LB, Glavina J, Faivovich J, de Prat-Gay G, Sánchez IE. Evolution of linear motifs within the papillomavirus E7 oncoprotein. *J Mol Biol.* 2012:422:336–346. https://doi.org/10.1016/j.jmb.2012.05.036.

Chen SC-C, Chuang T-J, Li W-H. The relationships among microRNA regulation, intrinsically disordered regions, and other indicators of protein evolutionary rate. *Mol Biol Evol.* 2011:28:2513–2520. https://doi.org/10.1093/molbev/msr068.

Chow CFW, Toth-Petroczy A. The evolution and exploration of intrinsically disordered and phase-separated protein states. In: Tripathi T, Uversky VN, editors. *The three functional states of proteins*. Elsevier; 2025. p. 353–379.

Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* 2010:5:e11147. https://doi.org/10.1371/journal.pone.0011147.

Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet.* 2012:46:677–700. https://doi.org/10.1146/annurev-genet-110711-155522.

Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. *Trends Biochem Sci.* 2011:36:159–169. https://doi.org/10.1016/j.tibs.2010.10.002.

Davison AJ. Comparative analysis of the genomes. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K, editors. *Human herpesviruses: biology, therapy, and immunoprophylaxis*. Cambridge University Press; 2007.

Davison AJ *et al.* The order herpesvirales. *Arch Virol.* 2009:154:171–177. https://doi.org/10.1007/s00705-008-0278-4.

DePristo MA, Weinreich DM, Hartl DL. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet.* 2005:6:678–687. https://doi.org/10.1038/nrg1672.

Dolan PT *et al.* Principles of dengue virus evolvability derived from genotype-fitness maps in human and mosquito cells. *Elife.* 2021:10:e61921. https://doi.org/10.7554/eLife.61921.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A.* 2005:102:14338–14343. https://doi.org/10.1073/pnas.0504070102.

Duggal NK, Emerman M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol.* 2012:12:687–695. https://doi.org/10.1038/nri3295.

Eberle R, Jones-Engel L. Understanding primate herpesviruses. *J Emerg Dis Virol.* 2017:3:10.16966/2473-1846.127. https://doi.org/10.16966/2473-1846.127.

Echave J, Wilke CO. Biophysical models of protein evolution: understanding the patterns of evolutionary sequence divergence. *Annu Rev Biophys.* 2017:46:85–103. https://doi.org/10.1146/annurev-biophys-070816-033819.

Enard D, Cai L, Gwennap C, Petrov DA. Viruses are a dominant driver of protein adaptation in mammals. *Elife.* 2016:5:e12469. https://doi.org/10.7554/eLife.12469.

Garamszegi S, Franzosa EA, Xia Y. Signatures of pleiotropy, economy and convergent evolution in a domain-resolved map of human-virus protein-protein interaction networks. *PLoS Pathog.* 2013:9:e1003778. https://doi.org/10.1371/journal.ppat.1003778.

Gatherer D *et al.* ICTV virus taxonomy profile: herpesviridae 2021. *J Gen Virol.* 2021:102:001673. https://doi.org/10.1099/jgv.0.001673.

Gitlin L, Hagai T, LaBarbera A, Solovey M, Andino R. Rapid evolution of virus sequences in intrinsically disordered protein regions. *PLoS Pathog.* 2014:10:e1004529. https://doi.org/10.1371/journal.ppat.1004529.

Glavina J *et al.* Host diversification is concurrent with linear motif evolution in a Mastadenovirus hub protein. *J Mol Biol.* 2022b:434:167563. https://doi.org/10.1016/j.jmb.2022.167563.

Glavina J, Palopoli N, Chemes LB. Evolution of SLiM-mediated hijack functions in intrinsically disordered viral proteins. *Essays Biochem.* 2022a:66:945–958. https://doi.org/10.1042/EBC20220059.

González-Foutel NS *et al.* Conformational buffering underlies functional selection in intrinsically disordered protein regions. *Nat Struct Mol Biol.* 2022:29:781–790. https://doi.org/10.1038/s41594-022-00811-w.

Grinde B. Herpesviruses: latency and reactivation—viral strategies and host response. *J Oral Microbiol.* 2013:5:22766. https://doi.org/10.3402/jom.v5i0.22766.

Guindon S, Delsuc F, Dufayard J-F, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol.* 2009:537:113–137. https://doi.org/10.1007/978-1-59745-251-9_6.

Guo H *et al.* Evolutionary arms race between virus and host drives genetic diversity in bat severe acute respiratory syndrome-related coronavirus spike genes. *J Virol.* 2020:94:e00902–e00920. https://doi.org/10.1128/JVI.00902-20.

Hadarovich A *et al.* PICNIC accurately predicts condensate-forming proteins regardless of their structural disorder across organisms. *Nat Commun.* 2024:15:10668. https://doi.org/10.1038/s41467-024-55089-x.

Hagai T, Azia A, Babu MM, Andino R. Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell Rep.* 2014:7:1729–1739. https://doi.org/10.1016/j.celrep.2014.04.052.

Hagai T, Azia A, Trizac E, Levy Y. Modulation of folding kinetics of repeat proteins: interplay between intra- and interdomain interactions. *Biophys J.* 2012a:103:1555–1565. https://doi.org/10.1016/j.bpj.2012.08.018.

Hagai T, Levy Y. Folding of elongated proteins: conventional or anomalous? *J Am Chem Soc.* 2008:130:14253–14262. https://doi.org/10.1021/ja804280p.

Hagai T, Tóth-Petróczy Á, Azia A, Levy Y. The origins and evolution of ubiquitination sites. *Mol Biosyst.* 2012b:8:1865–1877. https://doi.org/10.1039/c2mb25052g.

Ivankov DN *et al.* Contact order revisited: influence of protein size on the folding rate. *Protein Sci.* 2003:12:2057–2062. https://doi.org/10.1110/ps.0302503.

Jo WK, Drosten C, Drexler JF. The evolutionary dynamics of endemic human coronaviruses. *Virus Evol.* 2021:7:veab020. https://doi.org/10.1093/ve/veab020.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013:30:772–780. https://doi.org/10.1093/molbev/mst010.

Kennedy MA *et al.* A TRUSTED targeted mass spectrometry assay for pan-herpesvirus protein detection. *Cell Rep.* 2022:39:110810. https://doi.org/10.1016/j.celrep.2022.110810.

Kistler KE, Bedford T. An atlas of continuous adaptive evolution in endemic human viruses. *Cell Host Microbe.* 2023:31:1898–1909.e3. https://doi.org/10.1016/j.chom.2023.09.012.

Koonin EV, Wolf YI. Constraints and plasticity in genome and molecular-phenome evolution. *Nat Rev Genet.* 2010:11:487–498. https://doi.org/10.1038/nrg2810.

Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005:22:1208–1222. https://doi.org/10.1093/molbev/msi105.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol.* 2006:23:1891–1901. https://doi.org/10.1093/molbev/msl051.

Kumar M *et al.* ELM-the Eukaryotic linear motif resource-2024 update. *Nucleic Acids Res.* 2024:52:D442–D455. https://doi.org/10.1093/nar/gkad1058.

Lassalle F *et al.* Islands of linkage in an ocean of pervasive recombination reveals two-speed evolution of human cytomegalovirus genomes. *Virus Evol.* 2016:2:vew017. https://doi.org/10.1093/ve/vew017.

Li H *et al.* Phase separation in viral infections. *Trends Microbiol.* 2022:30:1217–1231. https://doi.org/10.1016/j.tim.2022.06.005.

Liu X, Fitzgerald K, Kurt-Jones E, Finberg R, Knipe DM. Herpesvirus tegument protein activates NF-kappaB signaling through the

TRAF6 adaptor protein. *Proc Natl Acad Sci U S A*. 2008:105:11335–11339. https://doi.org/10.1073/pnas.0801617105.
Lou DI *et al*. An intrinsically disordered region of the DNA repair protein Nbs1 is a species-specific barrier to herpes simplex virus 1 in primates. *Cell Host Microbe*. 2016:20:178–188. https://doi.org/10.1016/j.chom.2016.07.003.
McGeoch DJ, Cook S, Dolan A, Jamieson FE, Telford EA. Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *J Mol Biol*. 1995:247:443–458. https://doi.org/10.1006/jmbi.1995.0152.
McGeoch DJ, Rixon FJ, Davison AJ. Topics in herpesvirus genomics and evolution. *Virus Res*. 2006:117:90–104. https://doi.org/10.1016/j.virusres.2006.01.002.
Mozzi A *et al*. Past and ongoing adaptation of human cytomegalovirus to its host. *PLoS Pathog*. 2020a:16:e1008476. https://doi.org/10.1371/journal.ppat.1008476.
Mozzi A *et al*. Intrinsically disordered regions are abundant in simplexvirus proteomes and display signatures of positive selection. *Virus Evol*. 2020b:6:veaa028. https://doi.org/10.1093/ve/veaa028.
Mozzi A *et al*. Simplexviruses successfully adapt to their host by fine-tuning immune responses. *Mol Biol Evol*. 2022:39:msac142. https://doi.org/10.1093/molbev/msac142.
Nguyen A *et al*. Modulation of biophysical properties of nucleocapsid protein in the mutant spectrum of SARS-CoV-2. *Elife*. 2024:13:RP94836. https://doi.org/10.7554/eLife.94836.3.
Nguyen Ba AN, Moses AM. Evolution of characterized phosphorylation sites in budding yeast. *Mol Biol Evol*. 2010:27:2027–2037. https://doi.org/10.1093/molbev/msq090.
Nilsson J, Grahn M, Wright APH. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. *Genome Biol*. 2011:12:R65. https://doi.org/10.1186/gb-2011-12-7-r65.
Nobre LV *et al*. Human cytomegalovirus interactome analysis identifies degradation hubs, domain associations and viral protein functions. *Elife*. 2019:8:e49894. https://doi.org/10.7554/eLife.49894.
Nomburg J *et al*. Birth of protein folds and functions in the virome. *Nature*. 2024:633:710–717. https://doi.org/10.1038/s41586-024-07809-y.
Palmer WH, Telford M, Navarro A, Santpere G, Norman PJ. Human herpesvirus diversity is altered in HLA class I binding peptides. *Proc Natl Acad Sci U S A*. 2022:119:e2123248119. https://doi.org/10.1073/pnas.2123248119.
Patrono LV *et al*. Archival influenza virus genomes from Europe reveal genomic variability during the 1918 pandemic. *Nat Commun*. 2022:13:2314. https://doi.org/10.1038/s41467-022-29614-9.
Privman E, Penn O, Pupko T. Improving the performance of positive selection inference by filtering unreliable alignment regions. *Mol Biol Evol*. 2012:29:1–5. https://doi.org/10.1093/molbev/msr177.
Sauter D *et al*. Human tetherin exerts strong selection pressure on the HIV-1 group N Vpu protein. *PLoS Pathog*. 2012:8:e1003093. https://doi.org/10.1371/journal.ppat.1003093.
Schuck P, Zhao H. Diversity of short linear interaction motifs in SARS-CoV-2 nucleocapsid protein. *MBio*. 2023:14:e0238823. https://doi.org/10.1128/mbio.02388-23.
Sela I, Ashkenazy H, Katoh K, Pupko T. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res*. 2015:43:W7–W14. https://doi.org/10.1093/nar/gkv318.
Seyffert M *et al*. The HSV-1 transcription factor ICP4 confers liquid-like properties to viral replication compartments. *Int J Mol Sci*. 2021:22:4447. https://doi.org/10.3390/ijms22094447.
Shen J, Kirk BD, Ma J, Wang Q. Diversifying selective pressure on influenza B virus hemagglutinin. *J Med Virol*. 2009:81:114–124. https://doi.org/10.1002/jmv.21335.
Shuler G, Hagai T. Rapidly evolving viral motifs mostly target biophysically constrained binding pockets of host proteins. *Cell Rep*. 2022:40:111212. https://doi.org/10.1016/j.celrep.2022.111212.
Sijmons S *et al*. High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. *J Virol*. 2015:89:7673–7695. https://doi.org/10.1128/JVI.00578-15.
Simmonds P, Aiewsakun P, Katzourakis A. Prisoners of war—host adaptation and its constraints on virus evolution. *Nat Rev Micro*. 2019:17:321–328. https://doi.org/10.1038/s41579-018-0120-2.
Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat Rev Genet*. 2015:16:224–236. https://doi.org/10.1038/nrg3905.
Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. *Bioinformatics*. 1999:15:327–332. https://doi.org/10.1093/bioinformatics/15.4.327.
Sucharita S, Krishnagopal A, van Drunen Littel-van den Hurk S. Comprehensive analysis of the tegument proteins involved in capsid transport and virion morphogenesis of alpha, beta and gamma herpesviruses. *Viruses*. 2023:15:2058. https://doi.org/10.3390/v15102058.
Szpara ML *et al*. Evolution and diversity in human herpes simplex virus genomes. *J Virol*. 2014:88:1209–1227. https://doi.org/10.1128/JVI.01987-13.
Tenthorey JL, Emerman M, Malik HS. Evolutionary landscapes of host-virus arms races. *Annu Rev Immunol*. 2022:40:271–294. https://doi.org/10.1146/annurev-immunol-072621-084422.
Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. Do viral proteins possess unique biophysical features? *Trends Biochem Sci*. 2009:34:53–59. https://doi.org/10.1016/j.tibs.2008.10.009.
Tóth-Petróczy A, Tawfik DS. Slow protein evolutionary rates are dictated by surface-core association. *Proc Natl Acad Sci U S A*. 2011:108:11151–11156. https://doi.org/10.1073/pnas.1015994108.
Tunyasuvunakool K *et al*. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021:596:590–596. https://doi.org/10.1038/s41586-021-03828-1.
Usmanova DR, Plata G, Vitkup D. Functional optimization in distinct tissues and conditions constrains the rate of protein evolution. *Mol Biol Evol*. 2024:41:msae200. https://doi.org/10.1093/molbev/msae200.
van der Lee R *et al*. Classification of intrinsically disordered regions and proteins. *Chem Rev*. 2014:114:6589–6631. https://doi.org/10.1021/cr400525m.
Villegas JA, Heidenreich M, Levy ED. Molecular and environmental determinants of biomolecular condensate formation. *Nat Chem Biol*. 2022:18:1319–1329. https://doi.org/10.1038/s41589-022-01175-4.
Weidner-Glunde M, Kruminis-Kaszkiel E, Savanagouder M. Herpesviral latency-common themes. *Pathogens*. 2020:9:125. https://doi.org/10.3390/pathogens9020125.
Wertheim JO *et al*. Discovery of novel herpes simplexviruses in wild gorillas, bonobos, and chimpanzees supports zoonotic origin of HSV-2. *Mol Biol Evol*. 2021:38:2818–2830. https://doi.org/10.1093/molbev/msab072.
Wilke CO, Drummond DA. Signatures of protein biophysics in coding sequence evolution. *Curr Opin Struct Biol*. 2010:20:385–389. https://doi.org/10.1016/j.sbi.2010.03.004.
Wilson CJ, Choy W-Y, Karttunen M. Alphafold2: a role for disordered protein/region prediction? *Int J Mol Sci*. 2022:23:4591. https://doi.org/10.3390/ijms23094591.
Woo J, Robertson DL, Lovell SC. Constraints on HIV-1 diversity from protein structure. *J Virol*. 2010:84:12995–13003. https://doi.org/10.1128/JVI.00702-10.
Woodruff JB, Hyman AA, Boke E. Organization and function of non-dynamic biomolecular condensates. *Trends Biochem Sci*. 2018:43:81–94. https://doi.org/10.1016/j.tibs.2017.11.005.
Xing J *et al*. Characterization of the subcellular localization of herpes simplex virus type 1 proteins in living cells. *Med Microbiol Immunol*. 2011:200:61–68. https://doi.org/10.1007/s00430-010-0175-9.
Xu G *et al*. Viral tegument proteins restrict cGAS-DNA phase separation to mediate immune evasion. *Mol Cell*. 2021:81:2823–2837.e9. https://doi.org/10.1016/j.molcel.2021.05.002.
Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the

three domains of life. *J Biomol Struct Dyn*. 2012:30:137–149. https://doi.org/10.1080/07391102.2012.675145.

Yang X *et al*. HVIDB: a comprehensive database for human-virus protein-protein interactions. *Brief. Bioinformatics*. 2021:22: 832–844. https://doi.org/10.1093/bib/bbaa425.

Zhang J *et al*. Species-specific deamidation of cGAS by herpes simplex virus UL37 protein facilitates viral replication. *Cell Host Microbe*. 2018:24:234–248.e5. https://doi.org/10.1016/j.chom.2018.07.004.

Zhang J, Yang J-R. Determinants of the rate of protein sequence evolution. *Nat Rev Genet*. 2015:16:409–420. https://doi.org/10.1038/nrg3950.

Zhang X, Zheng R, Li Z, Ma J. Liquid-liquid phase separation in viral function. *J Mol Biol*. 2023:435:167955. https://doi.org/10.1016/j.jmb.2023.167955.

Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 2005:33:2302–2309. https://doi.org/10.1093/nar/gki524.

Zhao B, Ghadermarzi S, Kurgan L. Comparative evaluation of AlphaFold2 and disorder predictors for prediction of intrinsic disorder, disorder content and fully disordered proteins. *Comput Struct Biotechnol J*. 2023:21:3248–3258. https://doi.org/10.1016/j.csbj.2023.06.001.

Zhou S *et al*. Liquid-liquid phase separation mediates the formation of herpesvirus assembly compartments. *J Cell Biol*. 2023:222: e202201088. https://doi.org/10.1083/jcb.202201088.

Zimmermann C *et al*. Diverse cytomegalovirus US11 antagonism and MHC-A evasion strategies reveal a tit-for-tat coevolutionary arms race in hominids. *Proc Natl Acad Sci U S A*. 2024:121: e2315985121. https://doi.org/10.1073/pnas.2315985121.