Computational and Structural Biotechnology Journal 23 (2024) 3143–3154

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Evolutionary analysis of ZAP and its cofactors identifies intrinsically disordered regions as central elements in host-pathogen interactions

Rachele Cagliani [a,*], Diego Forni [a], Alessandra Mozzi [a], Rotem Fuchs [b], Tzachi Hagai [b], Manuela Sironi [a]

[a] Scientific Institute IRCCS E. MEDEA, Computational Biology Unit, Bosisio Parini 23842, Italy
[b] Shmunis School of Biomedicine and Cancer Research, George S Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel

ARTICLE INFO

Keywords:
Zinc-finger antiviral protein
Cofactors
Mammalian evolution
Intrinsically disordered regions
Phase separation
Positive selection

ABSTRACT

The zinc-finger antiviral protein (ZAP) is an innate immunity sensor of non-self nucleic acids. Its antiviral activity is exerted through the physical interaction with different cofactors, including TRIM25, Riplet and KHNYN. Cellular proteins that interact with infectious agents are expected to be engaged in genetic conflicts that often result in their rapid evolution. To test this possibility and to identify the regions most strongly targeted by natural selection, we applied in silico molecular evolution tools to analyze the evolutionary history of ZAP and cofactors in four mammalian groups. We report evidence of positive selection in all genes and in most mammalian groups. On average, the intrinsically disordered regions (IDRs) embedded in the four proteins evolve significantly faster than folded domains and most positively selected sites fall within IDRs. In ZAP, the PARP domain also shows abundant signals of selection, and independent evolution in different mammalian groups suggests modulation of its ADP-ribose binding ability. Detailed analyses of the biophysical properties of IDRs revealed that chain compaction and conformational entropy are conserved across mammals. The IDRs in ZAP and KHNYN are particularly compact, indicating that they may promote phase separation (PS). In line with this hypothesis, we predicted several PS-promoting regions in ZAP and KHNYN, as well as in TRIM25. Positively selected sites are abundant in these regions, suggesting that PS may be important for the antiviral functions of these proteins and the evolutionary arms race with viruses. Our data shed light into the evolution of ZAP and cofactors and indicate that IDRs represent central elements in host-pathogen interactions.

1. Introduction

Organisms have evolved a plethora of strategies to detect and control viral infections. In particular, vertebrate cells encode a variety of nucleic acid sensors that detect the presence of viral RNA or DNA based on non-self features or inappropriate localization [1]. The zinc-finger antiviral protein (ZAP, also known as PARP13, and encoded by the *ZC3HAV1* gene) is one such sensor. It combines features of the nucleic acid pattern recognition receptor (PRR) with a direct antiviral activity. ZAP preferentially recognizes and directly binds viral RNAs (viral messenger RNAs or viral RNA genomes) to inhibit translation and/or to target them for degradation, eventually resulting in decreased viral protein production and suppression of virus replication [2–6].

ZAP was first identified in rat cells as a host restriction factor of Moloney murine leukemia retrovirus (MLV) [2](10.1126/science.1074276). Subsequently, it was shown to restrict a wide range of RNA and DNA viruses, including retroviruses, filoviruses, flaviviruses, coxsackieviruses, influenza viruses, coronaviruses, herpesviruses, and poxviruses [5,6]. Moreover, ZAP represents a critical regulator of transposable elements such as LINE-1 retrotransposons and Alu retroelements [7]. The preferential targets of ZAP are RNA molecules with a high proportion of CpG or UpA dinucleotides [5].

At least four ZAP isoforms are produced from the human gene through alternative splicing and polyadenylation [8]. Among these, ZAP-L and ZAP-S are the most highly expressed and their antiviral functions against different viruses have been characterized. Both isoforms contain an N-terminal domain (NTD) with four zinc-finger motifs (ZnF) directly involved in RNA-binding and protein-protein interaction,

Abbreviations: ZAP, zinc-finger antiviral protein; NTD, N-terminal domain; ZnF, zinc-finger motif; PARP, poly(ADP-ribose) polymerase; CD, central domain; SG, stress granules; IDR, intrinsically disordered region; PS, phase separation.

* Correspondence to: Rachele Cagliani, Scientific Institute IRCCS E. MEDEA, Via Don L. Monza 20, Bosisio Parini 23842, Italy.
E-mail address: rachele.cagliani@lanostrafamiglia.it (R. Cagliani).

https://doi.org/10.1016/j.csbj.2024.07.022
Received 19 June 2024; Received in revised form 30 July 2024; Accepted 30 July 2024
Available online 2 August 2024

and a central domain that consists of a fifth ZnF motif and two WWE domains (Fig. 1). ZAP-L also contains a catalytically inactive C-terminal poly(ADP-ribose) polymerase (PARP)-like domain as well as a cysteine motif that mediates S-farnesylation [9,10]. S-farnesylation is thought to partially explain the higher antiviral activity of ZAP-L compared to ZAP-S and the different cellular localization of these two isoforms: ZAP-S has a cytosolic localization, while ZAP-L localizes to the cytoplasmic endomembrane system. Upon viral infection or other stress conditions, both isoforms can be recruited to stress granules (SGs), membraneless organelles that play important roles in the response to infection [11–13]. The central domain of ZAP also encompasses a long intrinsically disordered region (IDR) (Fig. 1). IDRs, which are common in the human proteome, are protein regions that do not adopt a fixed three-dimensional structure but rather exist as conformational ensembles. IDRs are known to play different regulatory functions in the cell and are often involved in promoting phase separation (PS), a process that results in the formation of membraneless organelles, also referred to as biomolecular condensates [14]. In humans, proteins that interact with viruses were shown to have higher IDR fraction than those that do not [15].

Because it lacks nuclease activity, ZAP antiviral function is promoted and regulated by other cellular proteins, which act as cofactors. TRIM25 (with an E3 ligase activity), KHNYN (probably with a nuclease activity) and Riplet (an E3 ubiquitin ligase encoded by *RNF135*) [16] are the 3 best functionally characterized ZAP cofactors. TRIM25 is part of the tripartite motif (TRIM) family of ubiquitin ligases. It has a similar structure to Riplet, with an N-terminal RING domain and a C-terminal PRY/SPRY domain [16] (Fig. 1). Both TRIM25 and Riplet were shown to enhance ZAP antiviral activity against different viruses [5,17]. Very recently, TRIM25 was shown to undergo PS upon dsRNA binding and to co-condense with SG core proteins [18]. Finally, KHNYN contains an N-terminal extended KH-like domain, a NYN domain and a CUBAN (Cullin binding domain associating with NEDD8) domain (Fig. 1). The NYN domain likely functions as an endoribonuclease [5] and KHNYN is required for the antiviral activity of ZAP against retroviruses [19]. All the three ZAP cofactors are also characterized by the presence of IDRs of

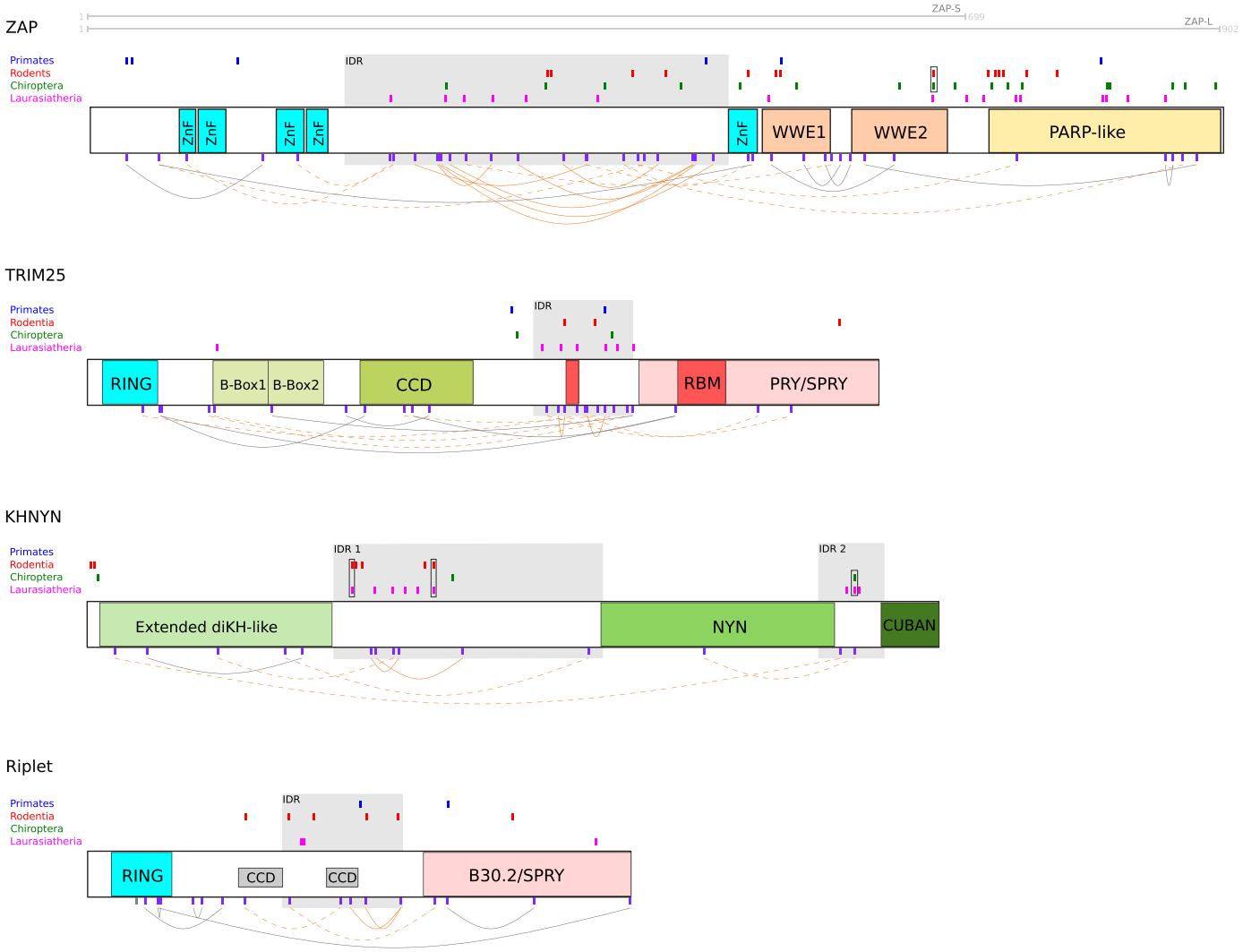

Fig. 1. Domain structures of ZAP, TRIM25, KHNYN, and Riplet. Schematic domain structures of human ZAP, TRIM25, KHNYN, and Riplet are drawn to scale. The grey shaded areas represent IDRs identified by the Metapredict tool based on human proteins. Positively selected sites identified in each mammalian group are shown above each domain structure. In blue, red, green, and magenta: positively selected sites in Primates, Rodentia, Chiroptera, and other Laurasiatheria, respectively. Positively selected sites identified in more than one mammalian phylogeny are boxed. Coevolving residue pairs inferred using the BGM method in the primate phylogeny are also reported (purple). Coevolving codon pairs are linked by different lines based on the location of the residues: orange line, both residues in IDRs; dotted orange line, only one residue in IDR; grey line, both sites outside IDRs. Abbreviation: ZnF, zinc finger motif; WWE, WWE (tryptophan, tryptophan, and glutamate) domain, CCD, coiled-coil domain; PARP: Poly (ADP-ribose) polymerase; RING: short for Really Interesting New Gene finger domain; RBM, RNA binding motif; NYN, Nedd4-BP1, YacP-like Nuclease domain; CUBAN, Cullin-binding domain associating with NEDD8; IDR, intrinsically disordered region.

varying lengths (Fig. 1).

Cellular proteins that interact with infectious agents are expected to be engaged in host-pathogen conflicts that often result in their rapid evolution and in signatures of positive selection [20]. Indeed, evidence of positive selection in primates was previously reported to target the PARP-like domain of ZAP [21]. However, a more recent analysis in birds indicated that the ZAP NTD and WWE domains also evolved under positive selection [22]. Likewise, signatures of positive selection were previously reported in different domains of TRIM25 in primates [23]. Despite these findings, the evolution of ZAP and TRIM25 in other mammals has remained unexplored and this also applies to groups, such as bats and rodents, that host a large number of different viruses and are considered major vectors of zoonotic viral diseases. Furthermore, the evolution of Riplet and KHNYN has not been investigated in mammals. To fill these knowledge gaps, we analyzed the mammalian orthologs of ZAP and its cofactors in four different mammalian groups: Primates, Rodentia, Chiroptera, and other Laurasiatheria - to detect evidence of positive selection and to test whether selection signals affect specific protein regions or domains. We report a number of novel selection signals in these four genes and identify IDRs as a major target of positive selection recurrent in the four mammalian groups we study. Finally, a characterization of the biophysical features of IDRs suggests that ZAP and KHNYN promote PS and some of the selected sites modulate this process.

2. Material and methods

2.1. Sequences, alignments, and phylogenetic trees

Coding sequences of *ZC3HAV1*, *TRIM25*, *KHNYN*, and *RNF315* were retrieved from the National Center for Biotechnology Information database (http://www.ncbi.nlm.nih.gov). Sequences with stop codons or having low sequence coverage were excluded. A list of species for each gene is reported in Supplementary Table S1.

The RevTrans 2.0 utility was used to generate multiple sequence alignments (MSA) (http://www.cbs.dtu.dk/services/RevTrans/, MAFFT v6.240 as an aligner) [24]. Each resulting alignment was manually inspected and was analyzed for the presence of recombination signals using GARD (Genetic Algorithm Recombination Detection) [25]. GARD is a Genetic Algorithm implemented in the HYPHY suite (version 2.2.4), which 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, the coding alignment was split on the basis of the recombination breakpoints and sub-regions were used as the input for subsequent molecular evolution analyses. We identified 4 gene alignments showing at least one recombination event (Table 1).

Gene trees were generated with the phyML program (v.3.1) [26]; specifically, we applied a General Time Reversible (GTR) model plus gamma-distributed rates and 4 substitution rate categories with a fixed proportion of invariable sites.

2.2. Identification of intrinsically disordered regions

Intrinsically disordered regions (IDRs) were identified by the Metapredict tool [27,28]. This tool defines disordered regions by applying a deep-learning algorithm based on a consensus score calculated from eight different disorder predictors [27]. Metapredict V2 was run using default parameters and residue disorder status was determined by its label by the tool. IDRs were defined as consecutive disordered stretches longer than 30 residues.

To obtain the sequences corresponding to the IDRs in each protein for each mammalian group, we applied Metapredict tool on a representative species for Primates, Rodentia, Chiroptera, and Laurasiatheria (*Homo sapiens*, *Mus musculus*, *Pteropus vampyrus,* and *Bos taurus,* respectively) (see Supplementary Text).

2.3. Evolutionary analysis in mammals

The average nonsynonymous substitution (dN) / synonymous substitution (dS) rate ratio (dN/dS) and (dN – dS) for each residue were calculated using the single-likelihood ancestor counting (SLAC) method [29]. Inputs were the MSAs and trees generated with the phyML program.

To detect positive selection, the codon-based codeml program implemented in the PAML (Phylogenetic Analysis by Maximum

Table 1
Likelihood ratio test statistics for models of variable selective pressure among sites (F3x4).

Gene	Mammalian group	GARD	SLAC	M8 vs M8a		M8 vs M7		N. positively selected sites
		Subregions	dN/dS	-2ΔlnL[a]	p-value[b] (df=1)	-2ΔlnL[a]	p-value[b] (df=2)	
ZC3HAV1 (ZAP)	Primates	Reg1	0.377	13.08	2.99×10^{-4}	14.32	7.77×10^{-4}	6
		Reg2	0.692	31.05	2.51×10^{-8}	36.05	1.49×10^{-8}	
	Rodentia		0.53	39.67	3.01×10^{-10}	48.22	3.38×10^{-11}	14
	Chiroptera		0.641	43.56	4.12×10^{-11}	65.37	6.38×10^{-15}	17
	Other Laurasiatheria		0.557	72.24	1.91×10^{-17}	102.13	6.66×10^{-23}	16
TRIM25	Primates		0.302	21.80	3.02×10^{-6}	40.11	1.95×10^{-9}	2
	Rodentia	Reg1	0.173	0.00	1	0.00	1	3
		Reg2	0.378	32.91	9.66×10^{-9}	58.26	2.23×10^{-13}	
	Chiroptera		0.305	22.70	1.9×10^{-6}	28.99	5.08×10^{-7}	2
	Other Laurasiatheria		0.307	36.12	1.86×10^{-9}	74.23	7.62×10^{-17}	8
KHNYN	Primates		0.298	0.17	0.678	0.00	1	
	Rodentia		0.304	5.30	0.0213	14.60	6.75×10^{-4}	7
	Chiroptera		0.265	5.25	0.0219	9.15	0.0103	3
	Other Laurasiatheria		0.253	12.64	3.77×10^{-4}	41.43	1.01×10^{-9}	9
RNF135 (Riplet)	Primates	Reg1	0.335	0.00	1	0.38	0.827	2
		Reg2	0.632	7.99	0.00471	10.88	0.00433	
	Rodentia		0.525	23.05	1.58×10^{-6}	23.05	9.87×10^{-6}	6
	Chiroptera		0.569	1.68	0.196	9.36	0.00928	
	Other Laurasiatheria	Reg1	0.384	0.00	0.979	0.68	0.713	3
		Reg2	0.81	4.21	0.04	6.82	0.0331	
		Reg3	0.488	6.05	0.0139	8.39	0.015	

Note:
[a] -2ΔLnL: twice the difference of the natural logs of the maximum likelihood of the models being compared;
[b] p-value: p-value of rejecting the neutral models in favor of the positive selection model.

Likelihood) suite was applied [30]. Using F3x4 codon frequencies model (codon frequencies estimated from the nucleotide frequencies in the data at each codon site) [30,31], a model (M8, positive selection model) that allows a class of sites to evolve with dN/dS > 1 was compared to two models (M7 and M8a, neutral models) that do not allow dN/dS > 1. To assess statistical significance, twice the difference of the likelihood (ΔlnL) for the models (M8a vs M8 and M7 vs M8) was compared to a $\chi 2$ distribution (1 or 2 degrees of freedom for M8a vs M8 and M7 vs M8 comparisons, respectively).

In order to identify specific sites subject to positive selection, we applied 3 different methods: 1) the Bayes Empirical Bayes (BEB) analysis (with a posterior probability cutoff ≥ 0.90), which calculates the posterior probability that each codon is from the site class of positive selection (under model M8) [32]; 2) Fast Unbiased Bayesian AppRoximation (FUBAR) [33], an approximate hierarchical Bayesian method that generates an unconstrained distribution of selection parameters to estimate the posterior probability of positive diversifying selection at each site in a given alignment (with a cutoff ≥ 0.90); 3) the Fixed EffectsLikelihood (FEL) [29], a maximum-likelihood (ML) approach to infer dN/dS on a per-site basis, assuming that the selection pressure for each site is constant along the entire phylogeny (with a p-value cutoff < 0.1). To be conservative and to limit false positives, only sites detected using at least two methods were considered as positive selection targets. GARD, FEL, FUBAR, and SLAC analyses were run locally through the HyPhy suite [34].

2.4. Co-evolutionary analyses

Co-evolution of sites within primates coding sequence alignments were assessed using the BGM (Bayesian Graphical Models) method [35] implemented in Spidermonkey through the Datamonkey Adaptive Evolution web-based interface (http://www.datamonkey.org) [36], using default parameters. BGM infers substitution history through the use of maximum-likelihood analyses of ancestral sequences and maps these to the phylogenetic tree, which allows for the detection of correlated patterns of substitution. A significant association between two sites was defined as a posterior probability exceeding a default cutoff of 0.5. To identify coevoling sites between ZAP and its cofactors, multiple sequence alignments of ZAP and each cofactor were concatenated and used for BGM analysis.

2.5. Structural mapping of positively selected sites

Protein structures available for ZAP (PDB IDs: 6uei, 7tgq, 2x5y), TRIM25 (PDBID: 6flm) and Riplet (PDB ID: 7jl1) were obtained from the Protein Data Bank (PDB, www.rcsb.org, last accessed 29th May 2024) archive. Pymol (PyMOL(TM) Molecular Graphics System, Version 2.4.0, Schrödinger, LLC) was used to visualize, analyze, and superimpose 3D molecular structures.

2.6. Analysis of IDR conformational properties and PS propensity

The conformational entropy per residue (S_{conf}/N) and the Flory scaling exponent (ν) were calculated for the orthologous IDRs from the four reference species identified by Metapredict [27,28]. In particular, using a Colab notebook, S_{conf}/N and ν were estimated by a support vector regression model, which was trained on simulations performed using the CALVADOS model [37,38].

The propensity of a protein to form biomolecular condensate was analyzed using different predictors. In particular, we used PhaSePred [39], PICNIC [40], and DeePhase [41]. The first two tools accept as inputs protein names or Uniprot IDs and were run for human and mouse proteins only, as they are the only species available among our representative ones. DeepPhase accepts as input protein sequences, thus all four representative species were analyzed. All tools were run using default parameters and suggested thresholds were adopted.

PS-promoting regions were identified using the ParSe method v 2.0 [42,43]. ParSe uses sequence-based calculations of hydrophobicity, α-helix propensity, and a model of the polymer scaling exponent (ν_{model}) to predict regions prone to undergo PS. We used a model that also includes the effects of interactions between amino acids (U_{π} for π–π and cation–π interactions and U_{q} for charge-based effects) trained on c_{sat} (the saturation concentration associated with protein PS) [37,38].

3. Results

3.1. Evolutionary trends in four mammalian groups

To analyze the evolutionary history of *ZC3HAV1* and cofactor genes (*TRIM25, KHNYN,* and *RNF135*) in mammals, we obtained coding sequence information for available species from public databases (Supplementary Table S1). In particular, for each gene, we retrieved the coding sequences for four mammalian groups: primates, rodents, bats, and other Laurasiatheria (Laurasiatheria excluding bats). It is worth noting that the coding sequence of ZAP-S entirely overlaps with that of ZAP-L (Fig. 1). Thus, analyses were performed for ZAP-L, which includes all the coding sites of ZAP-S. We performed all analyses separately for the four mammalian groups for different reasons: i) to avoid large divergence among taxa that can result in poor-quality alignments and in saturation of substitution rates; ii) to highlight patterns and trajectories that may be specific to a given group(s).

Previous studies have indicated that recombination can largely inflate type I error rates when models of positive selection are applied [44]. Thus, we first screened the alignments for the presence of recombination breakpoints using GARD (Genetic Algorithm Recombination Detection) [25]. GARD detected recombination breakpoints in 4 alignments (Table 1). Taking this information into account we calculated the average non-synonymous substitution/synonymous substitution rate (dN/dS) for the four genes/gene regions (according to the recombination breakpoint) using the (SLAC) method [29]. For all genes/gene regions dN/dS was much lower than 1 (Table 1), indicating a major role for purifying selection in shaping genetic diversity of ZAP, TRIM25, KHNYN, and Riplet. However, across all groups, ZAP and Riplet had higher average dN/dS compared to TRIM25 and KHNYN (Table 1).

To gain further insight into the evolutionary pattern of the four genes, a codon-wise measure of natural selection was obtained by calculating the dN–dS parameter. This metric was preferred over the conventional dN/dS, because in cases where dS values are equal to 0, dN/dS cannot be used. Comparison of structured regions and IDRs in the four proteins showed that the latter have significantly higher dN-dS compared to the former. This was true in all mammalian groups and for all proteins (Fig. 2, Supplementary Fig. S1). Taken together, these findings are consistent with the notion that IDRs are fast evolving and with the fact that fewer constraints are imposed on their substitutions due to the lack of need to be part of a folded three-dimensional structure [45–50].

3.2. ZAP and its cofactors were targets of positive selection during mammalian evolution

While constraints on protein function and structure typically result in overall purifying selection being the primary evolutionary force acting on protein regions, diversifying selection is often limited to specific sites or domains [20]. Indeed, this was previously shown to be case in primates for *ZC3HAV1* and *TRIM25* [21,51]. To test this possibility, we applied maximum-likelihood analyses implemented in the PAML (Phylogenetic Analysis by Maximum Likelihood) package [30,31]. Specifically, we used the codeml program to compare models of gene evolution that allow (NSsite model M8, positive selection model) or disallow (NSsite models M8a and M7, null models) a class of codons to evolve with dN/dS > 1. This analysis was performed for the four genes

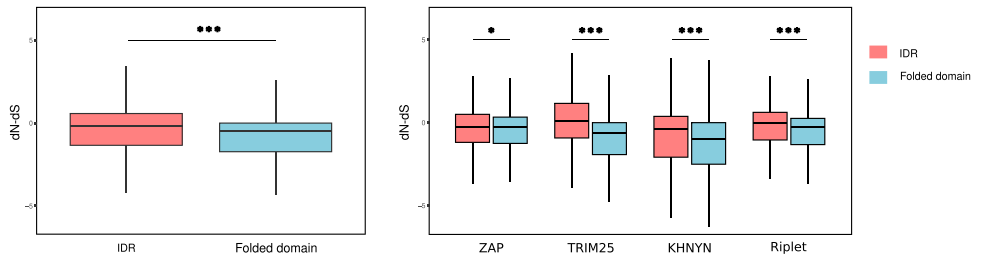

Fig. 2. Evolutionary rates in structured regions and IDRs. Codon-wise dN-dS computed for codons in structured regions and for IDRs considering all proteins and all mammalian groups together **(A),** or each protein individually **(B)**. Statistical significance was assessed by Wilcoxon rank sum test followed by False Discovery Rate (FDR) correction. * p-value < 0.05; ** p-value< 0.01; *** p-value< 0.001.

in each of the four mammalian groups. Both null models were rejected in favor of the positive selection models in all genes for each phylogeny, with the exception of *KHNYN* in primates and *RNF135* in bats (Table 1).

Overall, these analyses indicate that the four genes evolved under positive selection in most mammalian groups, suggesting that they represent a main selection target during the entire evolutionary history of placental mammals.

3.3. Intrinsically disordered regions (IDRs) tend to be enriched in positive selection signals

In order to identify specific sites subject to positive selection, we applied the Bayes Empirical Bayes (BEB) analysis from codeml model M8 [32,52] and two additional methods: FUBAR (Fast Unbiased Bayesian AppRoximation) (Murrell et al., 2013) and FEL (Fixed EffectsLikelihood) [29] (see methods). Only sites detected using at least two methods were considered as targets of positive selection. In total, we identified 98 positively selected sites: 53 in ZAP, 15 in TRIM25, 19 in KHNYN, and 11 in Riplet (Table 1, Supplementary Table 2). The positively selected sites that we identified in the region common to ZAP-S and ZAP-L necessarily impinge on both isoforms. Four sites were identified as positively selected in more than one mammalian phylogeny (Fig. 1, Supplementary Table 2). Overall, the largest number of sites was identified in Laurasiatheria, the smallest in primates. These data do not necessarily reflect the strength of selection acting on different groups, though. In fact, the power to detect positive selection is influenced by different factors, including the number of taxa and the inter-species divergence [32,53].

We next analyzed the location of positively selected sites in the four genes relative to the domain organization of the encoded proteins (Fig. 1). Positively selected sites in ZAP were distributed along the entire protein sequence. Sites selected in bats, rodents, and Laurasiatheria were particularly abundant in the central domain, the PARP-like domain, and in the IDR that links the NTD to the central domain. Conversely, three selected sites in the NTD were identified only in the primate phylogeny (Fig. 1). While these data confirm a previous description of a strong signal of positive selection in the PARP-like domain in primates, they provide wider evidence of selection that targets several additional regions within ZAP. We note that we identified most selected sites in rodents, bats, and other Laurasiatheria, which may explain why, by focusing on primates, a previous study did not detect positive selected sites outside the PARP-like domain [21].

With respect to TRIM25, KHNYN, and Riplet, most of the positively selected sites identified in all mammalian groups clustered in the IDRs (Fig. 1). The few exceptions included two sites in TRIM25 that fall into the B-box1 and PRY/SPRY domains, and 3 sites in Riplet B30.2/SPRY domain.

We thus aimed to test whether IDRs are significantly enriched of positively selected sites. We found this to be the case for TRIM25 and KHNYN (binomial test; TRIM25 p-value: $1.297x10^{-7}$; KHNYN p-value: $5.012x10^{-5}$). As for ZAP, although approximately 38 % (20 out of 53) of the positively selected sites fall into the IDR, this does not represent a statistically significant enrichment (binomial test; p-value: 0.888).

3.4. Analysis of positively selected sites in structured domains

In ZAP, most of the positively selected sites fall within structured domains, specifically the central and the PARP-like domains. The three sites that are positively selected in primates do not fall into the ZnF motifs, supporting the view that the CCCH domains of ZAP are highly conserved across mammals [21]. Analysis of the crystal structure of the NTD indicated that all of positively selected sites are surface-exposed suggesting that they might modulate interactions with cellular proteins or with viral antagonists (Fig. 3A). As mentioned above, the central domain (CD) comprises the fifth ZnF (ZnF5) and two WWE domains. We identified 11 positively selected sites in ZAP CD (2, 5 and 4 sites in ZnF5, WWE1, and WWE2, respectively) (Fig. 1). In particular, positions 671 and 672, in the WWE2 module, were independent targets of positive selection in 3 mammalian groups (Fig. 1, Supplementary Table 2). Moreover, a previous analysis of bird ZAP orthologs indicated that the position corresponding to site 672 was positively selected in these vertebrates, as well [22]. This gives a very strong indication that these residues play important functional roles. When mapped on the crystal structure of the human protein, Ile671 and Ala672 localize in the proximity of the binding pocket and of residues involved in the interaction of ZAP CD with poly(ADP-ribose) (PAR), a cellular polynucleotide [54,55] (Fig. 3A).

With respect to the ZAP PARP-like domain, 20 positively selected sites were identified in the 4 mammalian groups. The only selected site identified in the primate phylogeny (residue 805) overlaps with one of the 3 sites under positive selection previously reported by Kerns and colleagues [21]. Unlike other PARP proteins, the PARP-like domain of ZAP lacks poly(ADP-ribose) polymerase activity and is not NAD^+ binding competent. This is the result of the loss the HYYE (Hys--Tyr-Tyr-Glu) catalytic motif and closed cleft conformation at the canonical active site, which is thought to be physiologically relevant [56]. Indeed, structural analysis showed that the donor loop (D-loop) is forced into the dinucleotide binding pocket, closing the canonical NAD+ site [56]. Interestingly, 6 of the 20 positive selection sites fall on the D-loop and/or are involved in the bonds that contribute to stabilize the closed conformation (Fig. 3A).

As reported above, most of the positively selected sites (11/15) identified in TRIM25 fall within the IDR. The remaining 4 sites locate in the B-Box1 (1 site), in the PRY/SPRY domain (1 site) and in the non-disordered portion of the linker joining the CCD and PRY/SPRY domains (2 sites) (Fig. 1). Among these, position 599 (604 in mouse) in the SPRY domain is particularly interesting because it has recently been reported as a key residue involved in binding to the RIG-I second CARD domain [57]. The interaction between TRIM25 and RIG-I is mediated by two binding sites in the TRIM25 SPRY domain [57] and position 599 is located in the second binding site (Fig. 3B).

RIG-I also interacts with another ZAP cofactors, Riplet. Interestingly, two of the positively selected sites (human positions 287 and 338) in the Riplet SPRY domain map at the binding interface with RIG-I (Fig. 3C).

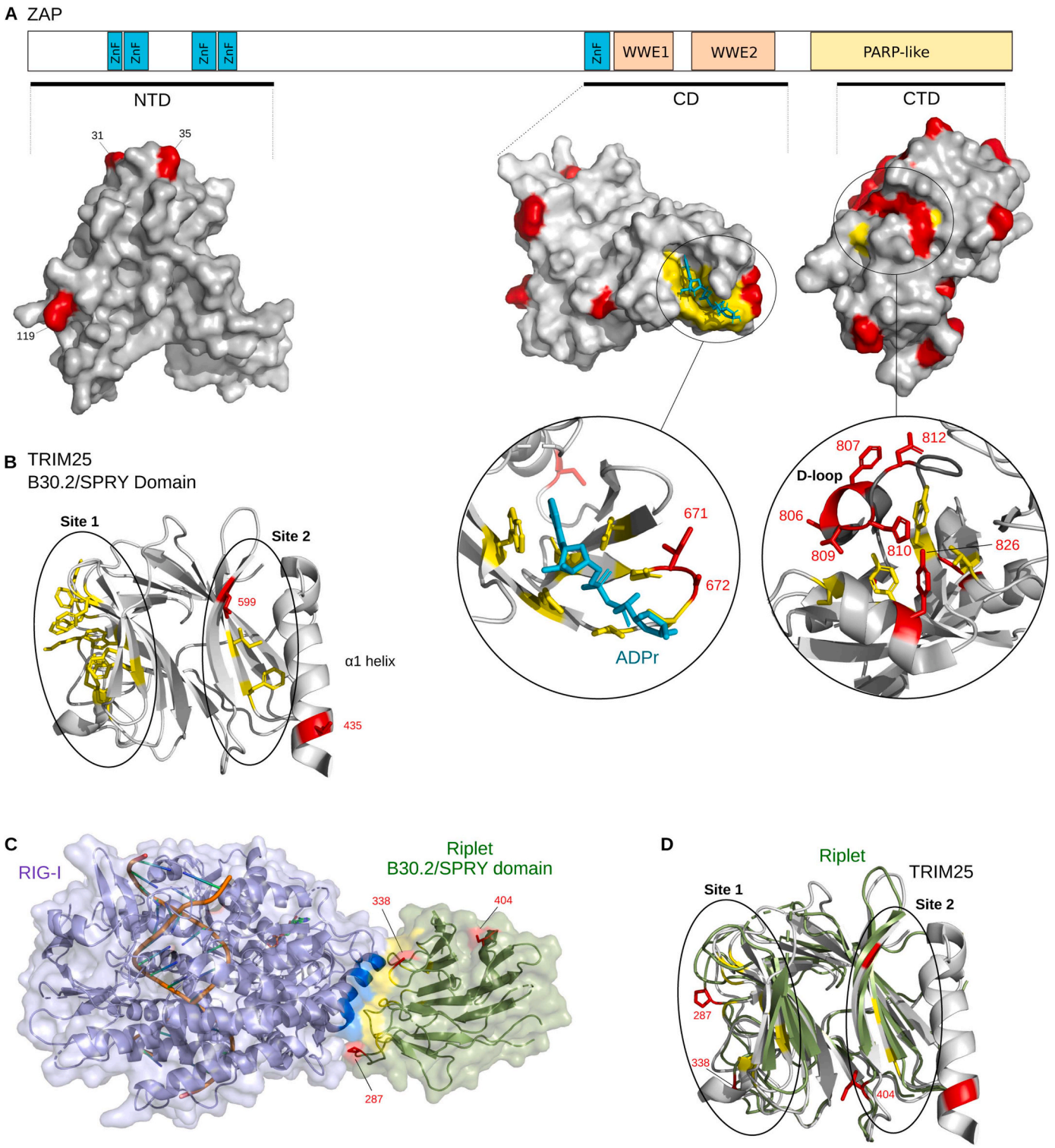

Fig. 3. Structural mapping of positively selected sites. **(A)** Molecular 3D-structures of different domains (according to the schematic domain architecture) of human ZAP (NTD, N-terminal domain, PDB ID: 6uei; CD, central domain, PDB ID: 7tgq; CTD, C-terminal domain, PDB ID:2x5y). The CD pocket involved in ADP-ribose (cyan) binding and the D-loop surrounding the HYYE (Hys-Tyr-Tyr-Glu) catalytic motif in the PARP-like domain are shown in the enlargements of each domain. Functional residues are in yellow. In all panels, positively selected sites are in red. **(B)** Human TRIM25 B30.2/SPRY domain ribbon representation (PDB ID: 6flm). Key residues defining RIG-I binding sites 1 and 2 are in yellow. Human residue 599 (604 in mouse), a target of positive selection in Rodentia, is a key residue defining RIG-I binding site 2 [57]. **(C)** Molecular 3D representation of human Riplet SPRY domain (green) in complex with RIG-I (light purple) (PDB ID:7jl1). Binding interfaces are also highlighted in yellow and blue, respectively. **(D)** Superimposition of the 3D structures of the B30.2/SPRY domains of Riplet (green, monomer from PDB ID: 7jl1) and TRIM25 (grey, PDB ID: 6flm), presented in ribbon representation. TRIM25 is color-coded as in B. Positively selected sites in Riplet are shown as sticks. The position of binding sites 1 and 2 is also reported.

The SPRY domain of Riplet has high sequence similarity to the same domain of TRIM25. Thus, to assess whether the third selected site (human position 404) in the SPRY domain of Riplet may also be involved in RIG-I binding, we superimposed the crystal structure of the Riplet and TRIM25 SPRY domains. This indicated that Riplet position 404 maps in close proximity to the beta-sheet involved in the second site-mediated interaction between TRIM25 and RIG-I (Fig. 3D). Overall, these results suggest that the positively selected sites modulate the interactions between TRIM25 and Riplet with RIG-I and, possibly, the antiviral activity of RIG-I itself.

3.5. Ensemble features of IDRs and possible involvement of selected sites in phase separation

Although IDRs do not adopt fixed 3D structures, some properties of their conformational ensembles are quantifiable and can provide information on IDR characteristics and function. These include the conformational entropy per residue (S_{conf}/N) and the Flory scaling exponent (ν), a measure of chain compactness. These features are clearly non-independent, as IDRs with low S_{conf}/N tend to be more compact [58]. We thus used predictors based on support vector regression models to calculate S_{conf}/N and ν for the IDRs in the proteins we analyzed. We first asked whether these features were conserved across orthologs in the four genes. To this end, we selected one representative species in each group (*Homo sapiens*, *Mus musculus*, *Bos taurus*, and *Pteropus vampyrus*)

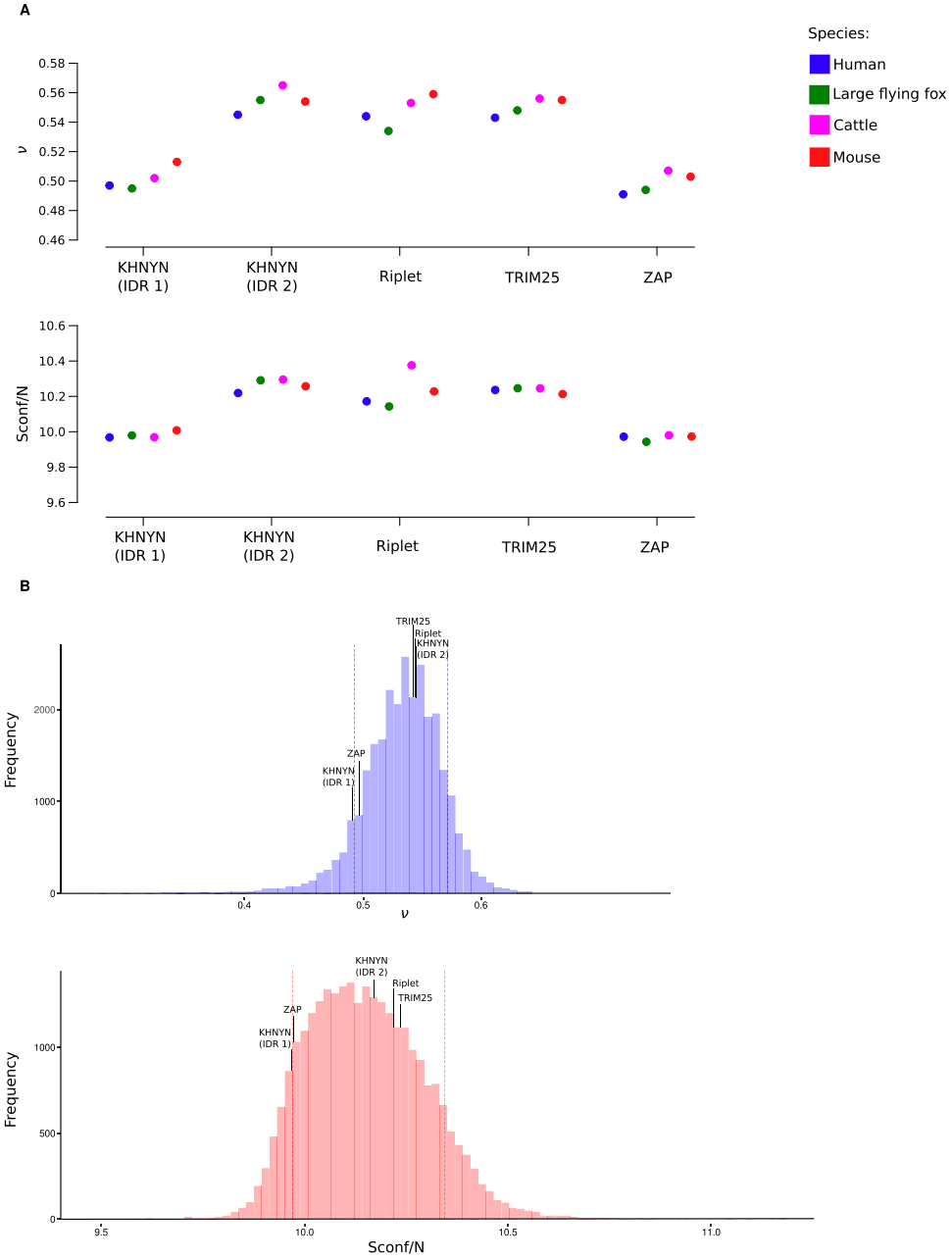

Fig. 4. IDR conformational properties. **(A)** Chain compaction (ν, upper panel) and conformational entropy (S_{conf}/N, lower panel) were calculated for orthologous IDRs in representative species of the four groups: *Homo sapiens* (human), *Mus musculus* (mouse), *Bos taurus* (cattle), and *Pteropus vampyrus* (large flying fox). Color codes for the different mammalian groups are as in Fig. 1. **(B)** Values of ν (upper panel) and S_{conf}/N (lower panel) for the IDRs of the analyzed proteins are shown in relation to the distribution of both features in IDRs in the human proteomes. Hatched lines represent the 10th and 90th percentiles.

and obtained the sequences corresponding to the IDRs in each protein. Calculation of S_{conf}/N and ν indicated that both parameters are very similar across orthologs, in line with the notion that IDRs can be divergent in sequence but ensemble features tend to be conserved [59]. The results above suggest that S_{conf}/N and ν for the human proteins can be considered representative of ensemble features in mammals. We thus used a recent analysis to compare S_{conf}/N and ν for IDRs in ZAP and cofactors with the distributions in the human proteome. Results indicated that whereas one of the IDRs in KHNYN, as well as those in Riplet and TRIM25 have average values, the one in ZAP and the longest one in KHNYN are in the low tail of the distributions of both chain compactness and entropy (around the 10th percentile) (Fig. 4). Chain compaction is often observed for the IDRs of proteins that undergo phase separation (PS) [40,57–61]. We thus reasoned that, in addition to TRIM25 [18] and ZAP [11–13], other cofactors might form biomolecular condensates. To test this hypothesis, we first queried CD-CODE (crowdsourcing condensate database and encyclopedia), an experimentally-validated database of protein condensates [60]. We found that both ZAP and TRIM25 condensate in SGs, while ZAP is also found in P-bodies (Table 2). Conversely, no experimentally validated data were available for KHNYN and Riplet. We thus exploited different tools that can predict the likelihood that a specific protein undergoes PS. For PICNIC and PhaSePred no information is available for *Bos taurus* and *Pteropus vampyrus*. Conversely, DeePhase (https://github.com/kadiliissaar/DeePhase) takes the protein sequence as the input and thus prediction were generated for representative orthologs in the four groups. All methods predicted that ZAP, TRIM25, and KHNYN are likely to form biomolecular condensates. Conversely, Riplet was not predicted to undergo PS by most methods, although the sequences of mouse and bat were identified as possibly participating to condensates by DeePhase (Table 2). We thus used ParSe (Partition Sequence) [43]to identify regions that promote PS in the orthologs from the four genes. In accordance with recent evidence, PS-promoting regions were detected in all TRIM25 orthologs [18]. Also, in line with the expectations from chain compaction and prediction methods, several PS regions were predicted in the ZAP and KHNYN proteins from all species, whereas fewer were detected in Riplet orthologs from human, mouse, and flying fox, and none in the cow ortholog (Table 3). This suggests, that, like TRIM25, ZAP and KHNYN undergo PS, whereas the PS potential of Riplet is weakly supported. Given these findings, we asked whether positive selection targeted PS regions. Combining data from the four mammalian groups, we found that positively selected sites are significantly enriched in the PS regions of TRIM25 and KHNYN (Table 3). In ZAP, 12 out of 53 sites are in PS regions, which does not represent a significant enrichment. This is most likely due to the fact that, in addition to the IDR the PARP-like domain is a major target of selection in ZAP. In fact, if we were to consider the distribution of sites in the ZAP-S isoform (which lacks the PARP domain) the p value of the binomial test for enrichment of selected sites in PS regions would drop to 0.058. With respect to Riplet, the failure to detect an enrichment of positively selected sites in PS-promoting regions is consistent with our findings that Riplet is unlikely to form biomolecular condensates.

Table 3
Positively selected sites in PS-promoting regions predicted with ParSe 2.0.

Protein	Mammalian group	Number of PS regions	N. of positively selected sites in PS regions	Binomial test
ZAP	Primates	2	0	Obs= 0.23
	Rodentia	5	5	Exp= 0.18
	Chiroptera	7	4	p-value= 0.38
	Other Laurasiatheria	4	3	
TRIM25	Primates	1	0	Obs= 0.26
	Rodentia	2	1	Exp= 0.06
	Chiroptera	1	0	p-value= 0.010
	Other Laurasiatheria	1	3	
KHNYN	Primates	4	0	Obs= 0.53
	Rodentia	4	4	Exp= 0.20
	Chiroptera	4	1	p-value= 0.0017
	Other Laurasiatheria	4	5	
Riplet	Primates	1	0	Obs= 0.18
	Rodentia	2	2	Exp= 0.07
	Chiroptera	2	0	p-value= 0.21
	Other Laurasiatheria	0	0	

3.6. Analysis of coevolving sites

Previous analyses have suggested that intra-protein coevolving residue pairs are common in IDRs [61]. Specifically, co-evolution was detected both between residues in IDRs and between IDRs and structured regions [61–64]. To examine the coevolution of codon pairs, we focused on the primate alignments of the four genes and we used the Bayesian Graphical Model (BGM) method [35]. Using this approach, we identified several coevolving residue pairs (Fig. 1, Supplementary Table S3). In ZAP, TRIM25, and KHNYN most coevolving pairs had at least one site in IDRs (Fig. 1). In particular, in the case of ZAP, several coevolving codon pairs were located within the long IDR, although signals of coevolution between the IDR and structured regions, as well as within the WWE domain, were also evident. Conversely, in the case of TRIM25 and KHNYN, most coevolving pairs involved one IDR partner and another partner located in structured domains. Finally, in Riplet, coevolution mostly involved codons in folded domains.

Table 2
Prediction of PS potential for ZAP and cofactors.

Protein	PhaSePred				PICNIC		CD-CODE	DeePhase			
								Phase Separation (score)			
	Human S-Self score[a] (Rank)	Human PS-Part score[b] (Rank)	Mouse S-Self score[a] (Rank)	Mouse PS-Part score[b] (Rank)	Human score[c]	Mouse score[c]	Human Biomolecular Condensates	Human	Mouse	Cattle	Large flying fox
ZAP	0.83 (0.95)	0.95 (0.99)	0.36 (0.78)	0.48 (0.67)	0.69	0.69	P-body, Stress granule	Yes (0.75)	Yes (0.77)	Yes (0.77)	Yes (0.83)
TRIM25	0.37 (0.81)	0.81 (0.93)	0.07 (0.50)	0.28 (0.46)	0.71	0.73	Stress granule	Yes (0.74)	Yes (0.70)	Yes (0.69)	Yes (0.68)
KHNYN	0.60 (0.87)	0.65 (0.85)	0.69 (0.92)	0.61 (0.83)	0.57	0.56		Yes (0.85)	Yes (0.84)	Yes (0.87)	Yes (0.83)
Riplet	0.05 (0.01)	0.04 (0.16)	0.07 (0.50)	0.18 (0.34)	0.17	0.23		No (0.40)	Yes (0.69)	No (0.31)	Yes (0.63)

[a] Proteins that can self-assemble to form condensates
[b] Proteins whose phase separation behaviors are regulated by protein or nucleic acid partner components
[c] a PICNIC score > 0.5 indicate the condensate-forming ability

Because ZAP and its cofactors function together in antiviral response, we sought to determine whether some evidence of coevolution was detectable. We thus applied the BGM method to search for residue pairs that coevolve at the inter-protein level. Specifically, we analyzed the coevolution of ZAP with each of its cofactors. In all cases we found evidence of coevolution, with the highest number of sites detected in ZAP and Riplet (Fig. 5). Analysis of coevolution signals again indicated that most codon pairs are either both located in IDRs or have at least one partner located in an IDR. Notably, in two cases the same sites in ZAP (positions 18 and 416) were found to coevolve with sites in two different cofactors (Fig. 5). Also, one site in ZAP (postion 451) coevolved with two distinct sites in Riplet. Overall, these data provide evidence of coevolution between ZAP and its three cofactors during primate evolution.

4. Discussion

ZAP emerged during tetrapod evolution and ZAP proteins from mammals, birds, and reptiles show antiviral activity [5,22]. It was thus suggested that the emergence of ZAP as an antiviral factor provided an opportunity for ZAP-interacting proteins to evolve as cofactors [65,66]. For instance, the nuclear export signal of KHNYN evolved after the separation of tetrapods from fish, possibly to allow its cytoplasmic localization and interaction with ZAP [66]. We thus jointly analyzed the evolution of ZAP and its cofactors in four mammalian groups, with the aim to detect convergent and divergent selection signals, and to characterize positively selected sites. In general, we found a good overlap of selection signals across groups in the four proteins. The most remarkable exception was the ZAP NTD, which was only targeted by selection in primates. Because a number of viral antagonists bind ZAP to hijack its activity [67–73], the positively selected sites, which are surface exposed, may have evolved as part of host-pathogen genetic conflicts. Alternatively, positive selection might modulate the interaction with the cofactors or other cellular partners. Two other genes showing different evolutionary patterns among groups were Riplet, which did now show evidence of positive selection in bats, and KHNYN, which was not a selection target in primates. The reasons why these genes experienced different evolutionary trajectories in these groups are difficult to envisage. One possibility is that selection signatures relate to the distribution and type of viruses that infect each group. This seems however difficult to imagine in the case of bats, as these mammals are known to represent major reservoirs of different viruses [74]. It should be mentioned, though, that, as detailed above, the power of molecular evolution methods depends on the strength of selection, and on the number and divergence of the analyzed taxa [32,53]. The sample size of bat sequences was the smallest and it is thus possible that this affected the power to detect selection in Riplet, especially if the pressure was weaker than on the other three analyzed genes. Moreover, we used a conservative test (the M8a vs M8 comparison) to define genes as positively selected. Using a less stringent test (M7 vs M8), Riplet shows statistically significant evidence of positive selection in bats, suggesting that limited power may explain our findings.

The central domain of ZAP was a target of selection in most mammalian groups and sites in close proximity were selected in more than one group, as well as in birds [22]. Such sites are located in proximity to the PAR binding region and might modulate PAR binding affinity. Although the PAR binding activity of ZAP-CD is not essential for ZAP-mediated inhibition of virus replication (at least for MLV and HIV-1), the loss of PAR binding capacity results in an appreciable reduction in ZAP antiviral activity [54]. PAR can nucleate the formation of biomolecular condensates [75]. It was thus suggested that the PAR-ZAP interaction may favor a stable association of ZAP and its cofactors to non-membranous cytoplasmic compartments (such as SGs) favoring recognition and degradation of ZAP-bound RNA [54]. In this respect, the CD may work in synergy with the central ZAP IDR, a linker between the RBD domain and the CD, to favor the localization of ZAP in non-membranous compartments (see below).

Another common target of selection in ZAP is the D-loop in the PARP-like domain, necessary to keep the closed cleft conformation of the catalytic site, which is inactive in ZAP [56]. In line with previous mutational studies, we suggest that the positively selected sites near the D-loop and the triad are involved in antiviral activity and that the natural selection signals observed are the result of host-pathogen

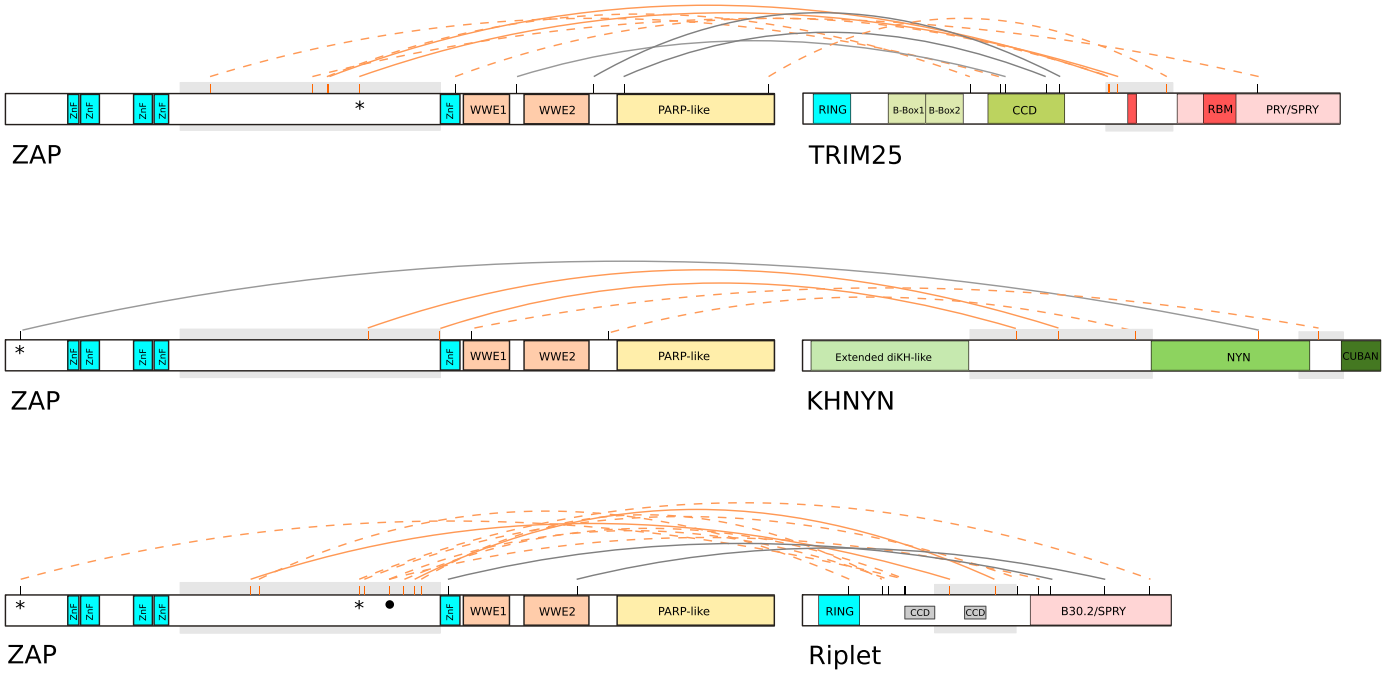

Fig. 5. Inter-protein coevolution. Coevolving residue pairs inferred using the BGM method in the primate phylogeny are reported. Sites are colored in orange if they are located in an IDR and in black when they are not located in an IDR. Coevolving codon pairs are linked by different lines based on the location of the residues: orange line, both residues in IDRs; dotted orange line, only one residue in IDR; grey line, both sites outside IDRs. Two sites in ZAP that coevolve with different sites in the cofactors are marked with asterisks. A site in ZAP that coevolves with two Riplet sites is marked with a circle.

arms-races [76]. We note that the molecular details of how the inactive triad and the D-loop modulate ZAP's activity are presently unknown and experimental analyses will be required to address this point, as well as to provide information about the functional effects of changes at the positively selected sites.

In ZAP cofactors, we detected a significant enrichment of positively selected sites within IDRs. IDRs are very common in mammalian proteomes and the majority of human proteins contain both folded protein domains and IDRs [77]. The latter were shown to evolve at a faster rate than structured regions in a number of systems [14,45,49,50,78]. This is in line with our analyses of dN-dS, whereby in all mammalian groups and in most proteins, IDRs had higher evolutionary rates than structured domains. A common explanation for the rapid evolution of IDRs is that, because they experience limited structural constraints, they are more tolerant to change [14]. While this possibility might well apply to the proteins we analyzed herein, it is worth noting that IDRs are enriched in human disease-associated proteins and it is estimated that up to 25 % of pathogenic missense mutations occur within IDRs [77]. This clearly implies that amino acid changes within these regions can have important functional consequences and that IDRs do not simply evolve by relaxation of selective constraints. Indeed, we report that the IDRs in ZAP and cofactors represent a major target of positive selection throughout mammalian evolution. Because natural selection acts on phenotypes, the positively selected sites we identified are expected to entail functional consequences, although gauging which such consequences might be is challenging and will require experimental analysis. In folded domains, the effects of amino acid changes are often inferred using structural information, as we did here for the central domain of ZAP and other folded regions. This is not possible for IDRs, which remain under-studied and poorly characterized [58]. Moreover, the fast evolutionary rates of IDRs make it difficult to reliably align distantly related orthologs, which is one of the reasons why we analyzed the four mammalian groups separately. Growing evidence however suggests that specific biophysical features are conserved in orthologous IDRs despite significant divergence in primary sequence [59]. Among such features, chain compaction and conformational entropy are important descriptors of structural ensemble properties and can be related to functional characteristics [58]. We found that both descriptors are very similar across orthologous regions of representative mammalian species, indicating that ensemble features and, possibly, functional properties are conserved. We thus leveraged data from a recent high-throughput analysis of IDRs in the human proteome [58] to contextualize the values we obtained for the IDRs of ZAP and cofactors. This showed that the ZAP IDR and the longest IDR in KHNYN are considerably compact. Chain compaction is associated with poor solvation and intra-chain interactions, which in tun can promote PS [14,43,77,79–81]. In fact, different PS prediction methods indicated that ZAP and KHNYN are likely to form biomolecular condensates. Conversely, the IDR of TRIM25 is not particularly compact. Nonetheless, TRIM25 was recently shown to undergo PS in a dsRNA-dependent manner and to co-condensate with G3BP1, a core component of SGs [18]. The predictors we used also inferred a high PS potential for this protein. Whereas the relationship between chain compaction and PS propensity will need further exploration, it is worth noting that TRIM25 binds dsRNA via the SPRY domain and a lysine-rich portion of the IDR [17]. It is thus possible that RNA binding and the RNA molecule itself dictate PS propensity and stoichiometry as a result of the ratio of RNA to protein charge units [82,83]. Interestingly, we found several positively selected sites to be located in the PS-promoting regions of ZAP, KHNYN and TRIM25. There are a number of ways in which amino acid changes can modulate the properties of membraneless organelles, including changing the threshold concentrations for PS, the material properties of the condensates (liquid, gelled or fibrillar), and their dynamics (assembling and disassembling). Again, some evidence for the relevance of amino acid substitutions for PS properties comes from human genetics, as several diseases are known to be caused by germline mutations that perturb PS [77]. Recently, PS was shown to play a central role in innate immune responses, and mammalian sensors of nucleic acids such as cGAS, MAVS, IFI16, and NLRP6 were shown to form biomolecular condensates with different material properties, ranging from droplets to fibrillar structures [84]. As a consequence, PS is yet another front where host-pathogen conflicts may play out. For instance, the tegument proteins of herpesviruses were shown to interfere with cGAS-DNA condensates [85] and diverse proteins encoded by distinct viral species interfere with SG formation [86]. It is thus possible that the selective pressure in the PS-promoting regions of ZAP and its cofactors is exerted by interacting viral proteins. Clearly, this possibility, as well as the ability of the selected sites to modulated PS properties will require future validation.

Our study has some limitations. First, although we used Metapredict, which combines different disorder predictors to generate consensus scores [27,28], our identification of IDRs relies on computational predictions. The same applies to the ability of ZAP and co-factors to form biomolecular condensates and to the PS-promoting regions. Experimental analyses will be necessary to assess the role of the regions we defined in PS and the possible effects of substitution at positively selected sites. Second, we analyzed the evolution of these proteins in mammalian groups that differ in many respects, including genetic divergence among species, lifestyle, population size, and infecting viruses. Some of these differences may affect the strength of selection acting on ZAP and cofactors, or our ability to detect it. Finally, although it is highly likely that these genes perform similar functions in all mammals, their expression levels in different tissues and organs may differ, possibly introducing a further confounding effect in the comparison among groups.

CRediT authorship contribution statement

Rachele Cagliani: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Rotem Fuchs:** Formal analysis, Investigation. **Diego Forni:** Conceptualization, Formal analysis, Investigation, Writing – review & editing. **Alessandra Mozzi:** Formal analysis, Visualization. **Tzachi Hagai:** Formal analysis, Funding acquisition, Writing – review & editing. **Manuela Sironi:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare no competing interests.

Data Availability

Multiple sequence alignments and phylogenetic trees are available from the authors upon request.

Acknowledgements

This work was supported by the Italian Ministry of Health ("Ricerca Corrente" to MS) and by the ISF, grant No. 435/20 (to TH).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.07.022.

References

[1] Chow KT, Gale M, Loo Y-M. RIG-I and other RNA sensors in antiviral immunity. Annu Rev Immunol 2018;36:667–94. https://doi.org/10.1146/annurev-immunol-042617-053309.

[2] Gao G, Guo X, Goff SP. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. Sci (N Y, NY) 2002;297:1703–6. https://doi.org/10.1126/science.1074276.

[3] Zhu Y, Wang X, Goff SP, Gao G. Translational repression precedes and is required for ZAP-mediated mRNA decay: ZAP-mediated translational repression versus mRNA decay. EMBO J 2012;31:4236–46. https://doi.org/10.1038/emboj.2012.271.

[4] Guo X, Ma J, Sun J, Gao G. The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. Proc Natl Acad Sci USA 2007;104:151–6. https://doi.org/10.1073/pnas.0607063104.

[5] Ficarelli M, Neil SJD, Swanson CM. Targeted restriction of viral gene expression and replication by the ZAP antiviral system. Annu Rev Virol 2021;8:265–83. https://doi.org/10.1146/annurev-virology-091919-104213.

[6] De Andrade KQ, Cirne-Santos CC. Antiviral activity of zinc finger antiviral protein (ZAP) in different virus families. Pathogens 2023;12:1461. https://doi.org/10.3390/pathogens12121461.

[7] Moldovan JB, Moran JV. The zinc-finger antiviral protein ZAP inhibits LINE and alu retrotransposition. PLoS Genet 2015;11:e1005121. https://doi.org/10.1371/journal.pgen.1005121.

[8] Li MMH, Aguilar EG, Michailidis E, Pabon J, Park P, Wu X, et al. Characterization of novel splice variants of zinc finger antiviral protein (ZAP). J Virol 2019;93:e00715-19. https://doi.org/10.1128/JVI.00715-19.

[9] Charron G, Li MMH, MacDonald MR, Hang HC. Prenylome profiling reveals S-farnesylation is crucial for membrane targeting and antiviral activity of ZAP long-isoform. Proc Natl Acad Sci USA 2013;110:11085–90. https://doi.org/10.1073/pnas.1302564110.

[10] Schwerk J, Soveg FW, Ryan AP, Thomas KR, Hatfield LD, Ozarkar S, et al. RNA-binding protein isoforms ZAP-S and ZAP-L have distinct antiviral and immune resolution functions. Nat Immunol 2019;20:1610–20. https://doi.org/10.1038/s41590-019-0527-6.

[11] Ly PT, Xu S, Wirawan M, Luo D, Roca X. ZAP isoforms regulate unfolded protein response and epithelial- mesenchymal transition. Proc Natl Acad Sci USA 2022;119:e2121453119. https://doi.org/10.1073/pnas.2121453119.

[12] Law LMJ, Razooky BS, Li MMH, You S, Jurado A, Rice CM, et al. ZAP's stress granule localization is correlated with its antiviral activity and induced by virus replication. PLoS Pathog 2019;15:e1007798. https://doi.org/10.1371/journal.ppat.1007798.

[13] Hirose T, Ninomiya K, Nakagawa S, Yamazaki T. A guide to membraneless organelles and their various roles in gene regulation. Nat Rev Mol Cell Biol 2023;24:288–304. https://doi.org/10.1038/s41580-022-00558-8.

[14] Holehouse AS, Kragelund BB. The molecular basis for cellular function of intrinsically disordered protein regions. Nat Rev Mol Cell Biol 2024;25:187–211. https://doi.org/10.1038/s41580-023-00673-0.

[15] Lou DI, Kim ET, Meyerson NR, Pancholi NJ, Mohni KN, Enard D, 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;20:178–88. https://doi.org/10.1016/j.chom.2016.07.003.

[16] Buckmaster MV, Goff SP. Riplet binds the zinc finger antiviral protein (ZAP) and augments ZAP-mediated restriction of HIV-1. J Virol 2022;96:e00526-22. https://doi.org/10.1128/jvi.00526-22.

[17] Sanchez JG, Sparrer KMJ, Chiang C, Reis RA, Chiang JJ, Zurenski MA, et al. TRIM25 binds RNA to modulate cellular anti-viral defense. J Mol Biol 2018;430:5280–93. https://doi.org/10.1016/j.jmb.2018.10.003.

[18] Shang Z, Zhang S, Wang J, Zhou L, Zhang X, Billadeau DD, et al. TRIM25 predominately associates with anti-viral stress granules. Nat Commun 2024;15:4127. https://doi.org/10.1038/s41467-024-48596-4.

[19] Ficarelli M, Wilson H, Galão RP, Mazzon M, Antzin-Anduetza I, Marsh M, et al. KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. eLife 2019;8. https://doi.org/10.7554/eLife.46767.

[20] Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host-pathogen interactions from mammalian sequence data. Nat Rev Genet 2015;16:224–36. https://doi.org/10.1038/nrg3905.

[21] Kerns JA, Emerman M, Malik HS. Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. PLoS Genet 2008;4:e21. https://doi.org/10.1371/journal.pgen.0040021.

[22] Odon V, Fiddaman SR, Smith AL, Simmonds P. Comparison of CpG- and UpA-mediated restriction of RNA virus replication in mammalian and avian cells and investigation of potential ZAP-mediated shaping of host transcriptome compositions. RNA 2022;28:1089–109. https://doi.org/10.1261/rna.079102.122.

[23] Malfavon-Borja R, Sawyer SL, Wu LI, Emerman M, Malik HS. An evolutionary screen highlights canonical and noncanonical candidate antiviral genes within the primate TRIM gene family. Genome Biol Evol 2013;5:2141–54. https://doi.org/10.1093/gbe/evt163.

[24] Wernersson R, Pedersen AG. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res 2003;31:3537–9.

[25] Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 2006;23:1891–901. https://doi.org/10.1093/molbev/msl051.

[26] Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol (Clifton, NJ) 2009;537:113–37. https://doi.org/10.1007/978-1-59745-251-9_6.

[27] Emenecker RJ, Griffith D, Holehouse AS. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. Biophys J 2021;120:4312–9. https://doi.org/10.1016/j.bpj.2021.08.039.

[28] Emenecker R.J., Griffith D., Holehouse A.S. Metapredict V2: An update to metapredict, a fast, accurate, and easy-to-use predictor of consensus disorder and structure 2022. https://doi.org/10.1101/2022.06.06.494887.

[29] 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;22:1208–22. https://doi.org/10.1093/molbev/msi105.

[30] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 2007;24:1586–91. https://doi.org/10.1093/molbev/msm088.

[31] Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci: CABIOS 1997;13:555–6.

[32] Anisimova M, Bielawski JP, Yang Z. Accuracy and power of bayes prediction of amino acid sites under positive selection. Mol Biol Evol 2002;19:950–8. https://doi.org/10.1093/oxfordjournals.molbev.a004152.

[33] Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLK, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol 2013;30:1196–205. https://doi.org/10.1093/molbev/mst030.

[34] Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics 2005;21:676–9. https://doi.org/10.1093/bioinformatics/bti079.

[35] Poon AFY, Lewis FI, Frost SDW, Kosakovsky Pond SL. Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models. Bioinformatics 2008;24:1949–50. https://doi.org/10.1093/bioinformatics/btn313.

[36] Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol 2018;35:773–7. https://doi.org/10.1093/molbev/msx335.

[37] Tesei G., Trolle A.I., Jonsson N., Betz J., Pesce F., Johansson K.E., et al. Conformational ensembles of the human intrinsically disordered proteome: Bridging chain compaction with function and sequence conservation 2023. ⟨https://doi.org/10.1101/2023.05.08.539815⟩.

[38] Tesei G, Lindorff-Larsen K. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. Open Res Eur 2023;2:94. https://doi.org/10.12688/openreseurope.14967.2.

[39] Chen Z, Hou C, Wang L, Yu C, Chen T, Shen B, et al. Screening membraneless organelle participants with machine-learning models that integrate multimodal features. e2115369119 Proc Natl Acad Sci USA 2022;119. https://doi.org/10.1073/pnas.2115369119.

[40] Hadarovich A., Singh H.R., Ghosh S., Rostam N., Hyman A.A., Toth-Petroczy A. PICNIC accurately predicts condensate-forming proteins regardless of their structural disorder across organisms 2023. ⟨https://doi.org/10.1101/2023.06.01.543229⟩.

[41] Saar KL, Morgunov AS, Qi R, Arter WE, Krainer G, Lee AA, et al. Learning the molecular grammar of protein condensates from sequence determinants and embeddings. Proc Natl Acad Sci USA 2021;118:e2019053118. https://doi.org/10.1073/pnas.2019053118.

[42] Wilson C, Lewis KA, Fitzkee NC, Hough LE, Whitten ST. ParSe 2.0: a web tool to identify drivers of protein phase separation at the proteome level. Protein Sci 2023;32:e4756. https://doi.org/10.1002/pro.4756.

[43] Ibrahim AY, Khaodeuanepheng NP, Amarasekara DL, Correia JJ, Lewis KA, Fitzkee NC, et al. Intrinsically disordered regions that drive phase separation form a robustly distinct protein class. J Biol Chem 2023;299:102801. https://doi.org/10.1016/j.jbc.2022.102801.

[44] Anisimova M, Nielsen R, Yang Z. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 2003;164:1229–36.

[45] 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;28:975–82. https://doi.org/10.1101/gr.232645.117.

[46] Brown CJ, Johnson AK, Daughdrill GW. Comparing models of evolution for ordered and disordered proteins. Mol Biol Evol 2010;27:609–21. https://doi.org/10.1093/molbev/msp277.

[47] Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197–208. https://doi.org/10.1038/nrm1589.

[48] 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;7:1729–39. https://doi.org/10.1016/j.celrep.2014.04.052.

[49] Mozzi A, Forni D, Cagliani R, Clerici M, Pozzoli U, Sironi M. Intrinsically disordered regions are abundant in simplexvirus proteomes and display signatures of positive selection. Virus Evol 2020;6:veaa028. https://doi.org/10.1093/ve/veaa028.

[50] Molteni C, Forni D, Cagliani R, Mozzi A, Clerici M, Sironi M. Evolution of the orthopoxvirus core genome. Virus Res 2023;323:198975. https://doi.org/10.1016/j.virusres.2022.198975.

[51] Daugherty MD, Young JM, Kerns JA, Malik HS. Rapid evolution of PARP genes suggests a broad role for ADP-ribosylation in host-virus conflicts. PLoS Genet 2014;10:e1004403. https://doi.org/10.1371/journal.pgen.1004403.

[52] Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 2005;22:1107–18. https://doi.org/10.1093/molbev/msi097.

[53] Wong WSW, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 2004;168:1041–51. https://doi.org/10.1534/genetics.104.031153.

[54] Xue G, Braczyk K, Gonçalves-Carneiro D, Dawidziak DM, Sanchez K, Ong H, et al. Poly(ADP-ribose) potentiates ZAP antiviral activity. PLoS Pathog 2022;18:e1009202. https://doi.org/10.1371/journal.ppat.1009202.

[55] Kuttiyatveetil JRA, Soufari H, Dasovich M, Uribe IR, Mirhasan M, Cheng S-J, et al. Crystal structures and functional analysis of the ZnF5-WWE1-WWE2 region of PARP13/ZAP define a distinctive mode of engaging poly(ADP-ribose). Cell Rep 2022;41:111529. https://doi.org/10.1016/j.celrep.2022.111529.
[56] Karlberg T, Klepsch M, Thorsell A-G, Andersson CD, Linusson A, Schüler H. Structural basis for lack of ADP-ribosyltransferase activity in Poly(ADP-ribose) Polymerase-13/Zinc finger antiviral protein. J Biol Chem 2015;290:7336–44. https://doi.org/10.1074/jbc.M114.630160.
[57] D'Cruz AA, Kershaw NJ, Hayman TJ, Linossi EM, Chiang JJ, Wang MK, et al. Identification of a second binding site on the TRIM25 B30.2 domain. Biochem J 2018;475:429–40. https://doi.org/10.1042/BCJ20170427.
[58] Tesei G, Trolle AI, Jonsson N, Betz J, Knudsen FE, Pesce F, et al. Conformational ensembles of the human intrinsically disordered proteome. Nature 2024;626:897–904. https://doi.org/10.1038/s41586-023-07004-5.
[59] Zarin T, Strome B, Nguyen Ba AN, Alberti S, Forman-Kay JD, Moses AM. Proteome-wide signatures of function in highly diverged intrinsically disordered regions. eLife 2019;8:e46883. https://doi.org/10.7554/eLife.46883.
[60] Rostam N, Ghosh S, Chow CFW, Hadarovich A, Landerer C, Ghosh R, et al. CD-CODE: crowdsourcing condensate database and encyclopedia. Nat Methods 2023;20:673–6. https://doi.org/10.1038/s41592-023-01831-0.
[61] Basu S, Bahadur RP. Conservation and coevolution determine evolvability of different classes of disordered residues in human intrinsically disordered proteins. Proteins 2022;90:632–44. https://doi.org/10.1002/prot.26261.
[62] González-Foutel NS, Glavina J, Borcherds WM, Safranchik M, Barrera-Vilarmau S, Sagar A, et al. Conformational buffering underlies functional selection in intrinsically disordered protein regions. Nat Struct Mol Biol 2022;29:781–90. https://doi.org/10.1038/s41594-022-00811-w.
[63] Shafee T, Bacic A, Johnson K. Evolution of sequence-diverse disordered regions in a protein family: order within the chaos. Mol Biol Evol 2020;37:2155–72. https://doi.org/10.1093/molbev/msaa096.
[64] Karamanos TK. Chasing long-range evolutionary couplings in the AlphaFold era. Biopolymers 2023;114:e23530. https://doi.org/10.1002/bip.23530.
[65] Gonçalves-Carneiro D, Takata MA, Ong H, Shilton A, Bieniasz PD. Origin and evolution of the zinc finger antiviral protein. PLoS Pathog 2021;17:e1009545. https://doi.org/10.1371/journal.ppat.1009545.
[66] Lista MJ, Ficarelli M, Wilson H, Kmiec D, Youle RL, Wanford J, et al. A nuclear export signal in KHNYN required for its antiviral activity evolved as ZAP emerged in tetrapods. J Virol 2023;97:e00872-22. https://doi.org/10.1128/jvi.00872-22.
[67] Fehr AR, Singh SA, Kerr CM, Mukai S, Higashi H, Aikawa M. The impact of PARPs and ADP-ribosylation on inflammation and host–pathogen interactions. Genes Dev 2020;34:341–59. https://doi.org/10.1101/gad.334425.119.
[68] Zhao Y, Song Z, Bai J, Liu X, Nauwynck H, Jiang P. ZAP, a CCCH-type zinc finger protein, inhibits porcine reproductive and respiratory syndrome virus replication and interacts with viral Nsp9. J Virol 2019;93:e00001-19. https://doi.org/10.1128/JVI.00001-19.
[69] Chuenchat J, Kardkarnklai S, Narkpuk J, Liwnaree B, Jongkaewwattana A, Jaru-Ampornpan P, et al. PEDV nucleocapsid antagonizes zinc-finger antiviral protein by disrupting the interaction with its obligate co-factor, TRIM25. Vet Microbiol 2024;291:110033. https://doi.org/10.1016/j.vetmic.2024.110033.
[70] Peng C, Wyatt LS, Glushakow-Smith SG, Lal-Nag M, Weisberg AS, Moss B. Zinc-finger antiviral protein (ZAP) is a restriction factor for replication of modified vaccinia virus Ankara (MVA) in human cells. PLoS Pathog 2020;16:e1008845. https://doi.org/10.1371/journal.ppat.1008845.
[71] Xie L, Lu B, Zheng Z, Miao Y, Liu Y, Zhang Y, et al. The 3C protease of enterovirus A71 counteracts the activity of host zinc-finger antiviral protein (ZAP). J Gen Virol 2018;99:73–85. https://doi.org/10.1099/jgv.0.000982.
[72] Tang Q, Wang X, Gao G. The short form of the zinc finger antiviral protein inhibits influenza A virus protein expression and is antagonized by the virus-encoded NS1. J Virol 2017;91:e01909-16. https://doi.org/10.1128/JVI.01909-16.
[73] Xuan Y, Gong D, Qi J, Han C, Deng H, Gao G. ZAP inhibits murine gammaherpesvirus 68 ORF64 expression and is antagonized by RTA. J Virol 2013;87:2735–43. https://doi.org/10.1128/JVI.03015-12.
[74] Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. Nature 2017;546:646–50. https://doi.org/10.1038/nature22975.
[75] Rhine K, Odeh HM, Shorter J, Myong S. Regulation of biomolecular condensates by Poly(ADP-ribose). Chem Rev 2023;123:9065–93. https://doi.org/10.1021/acs.chemrev.2c00851.
[76] Gläsker S, Töller M, Kümmerer BM. The alternate triad motif of the poly(ADP-ribose) polymerase-like domain of the human zinc finger antiviral protein is essential for its antiviral activity. J Gen Virol 2014;95:816–22. https://doi.org/10.1099/vir.0.060988-0.
[77] Tsang B, Pritišanac I, Scherer SW, Moses AM, Forman-Kay JD. Phase separation as a missing mechanism for interpretation of disease mutations. Cell 2020;183:1742–56. https://doi.org/10.1016/j.cell.2020.11.050.
[78] Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. Curr Opin Struct Biol 2011;21:441–6. https://doi.org/10.1016/j.sbi.2011.02.005.
[79] Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, et al. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science 2020;367:694–9. https://doi.org/10.1126/science.aaw8653.
[80] Lin Y-H, Chan HS. Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. Biophys J 2017;112:2043–6. https://doi.org/10.1016/j.bpj.2017.04.021.
[81] Polyansky AA, Gallego LD, Efremov RG, Köhler A, Zagrovic B. Protein compactness and interaction valency define the architecture of a biomolecular condensate across scales. eLife 2023;12:e80038. https://doi.org/10.7554/eLife.80038.
[82] Ranganathan S, Shakhnovich E. Effect of RNA on morphology and dynamics of membraneless organelles. J Phys Chem B 2021;125:5035–44. https://doi.org/10.1021/acs.jpcb.1c02286.
[83] Laghmach R, Alshareedah I, Pham M, Raju M, Banerjee PR, Potoyan DA. RNA chain length and stoichiometry govern surface tension and stability of protein-RNA condensates. iScience 2022;25:104105. https://doi.org/10.1016/j.isci.2022.104105.
[84] Liu D, Yang J, Cristea IM. Liquid–liquid phase separation in innate immunity. Trends Immunol 2024. https://doi.org/10.1016/j.it.2024.04.009. S1471490624000978.
[85] Xu G, Liu C, Zhou S, Li Q, Feng Y, Sun P, 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.
[86] Li H, Ernst C, Kolonko-Adamska M, Greb-Markiewicz B, Man J, Parissi V, et al. Phase separation in viral infections. Trends Microbiol 2022;30:1217–31. https://doi.org/10.1016/j.tim.2022.06.005.