Novel Z-DNA binding domains in giant viruses

Miguel F. Romero[1,‡] , Jeffrey B. Krall[2,‡] , Parker J. Nichols[2,‡] , Jillian Vantreeck[2] , Morkos A. Henen[2],
Emmanuel Dejardin[3], Frederik Schulz[1,*] , Quentin Vicens[4,§] , Beat Vögeli[2,§,*] , and
Mamadou Amadou Diallo[3,§,*]

From the [1]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA; [2]Department of
Biochemistry and Molecular Genetics, University of Colorado at Denver, Aurora, Colorado, USA; [3]GIGA I3 - Molecular Immunology
and Signal Transduction, University of Liège, Liège, Belgium; [4]Department of Biology and Biochemistry, Center for Nuclear
Receptors and Cell Signaling, University of Houston, Houston, Texas, USA

Reviewed by members of the JBC Editorial Board. Edited by Craig Cameron

Z-nucleic acid structures play vital roles in cellular processes and have implications in innate immunity due to their recognition by Zα domains containing proteins (Z-DNA/Z-RNA binding proteins, ZBPs). Although Zα domains have been identified in six proteins, including viral E3L, ORF112, and I73R, as well as, cellular ADAR1, ZBP1, and PKZ, their prevalence across living organisms remains largely unexplored. In this study, we introduce a computational approach to predict Zα domains, leading to the revelation of previously unidentified Zα domain-containing proteins in eukaryotic organisms, including non-metazoan species. Our findings encompass the discovery of new ZBPs in previously unexplored giant viruses, members of the *Nucleocytoviricota* phylum. Through experimental validation, we confirm the Zα functionality of select proteins, establishing their capability to induce the B-to-Z conversion. Additionally, we identify Zα-like domains within bacterial proteins. While these domains share certain features with Zα domains, they lack the ability to bind to Z-nucleic acids or facilitate the B-to-Z DNA conversion. Our findings significantly expand the ZBP family across a wide spectrum of organisms and raise intriguing questions about the evolutionary origins of Zα-containing proteins. Moreover, our study offers fresh perspectives on the functional significance of Zα domains in virus sensing and innate immunity and opens avenues for exploring hitherto undiscovered functions of ZBPs.

Nucleic acids adopt a variety of structures other than right-handed double helices. Both DNA and RNA can form triplexes, I-motifs, G-quadruplexes, and left-handed helices referred to as "Z-nucleic acids" (Z-DNA and Z-RNA) (1–3). These alternative structures are critical components of cellular functions, influencing a myriad of biological processes (4–6). In particular, it is now well established that the innate immune system exploits Z-nucleic acids as pathogen-associated molecular patterns and damage-associated molecular patterns (7, 8).

Z-DNA/Z-RNA biology owes its now-recognized relevance to several proteins containing Zα domains, which are responsible for recognizing Z-nucleic acids (9). Zα domains are found in metazoan proteins involved in immunity and cancer, as well as in viral proteins. The first crystal structure revealed that the Zα domain of ADAR1 (adenosine deaminase acting on RNA 1) adopts a winged-helix-turn-helix (wHTH) fold. This structure is characterized by a compact α/β architecture encompassing a three-helix bundle (α1 to α3), juxtaposed to a twisted antiparallel β-sheet (β1 to β3) (10). Specific amino acids are important for the specific recognition of Z-nucleic acids, such as N173 and Y177 within helix α3, and P192 and P193; W195 contributes to both protein stability and nucleic acid binding (10). This earlier work demonstrated that the commonly found wHTH domain across proteins only binds to Z-nucleic acids when these amino acids are present.

The only other proteins known to contain Zα domains are ZBP1 in mammals, amphibia, and reptilia (11), PKZ in salmoniform and cypriniform fish (12), E3L in poxviruses (13), ORF112 in cyprinid herpesviruses (CyHV-1-to-3) (14) and I73R in the asfarviridae African swine fever virus (ASFV) (15). Our current understanding of the biology of these proteins points to antagonistic relationships in innate immunity and a potential regulatory role in gene expression (16, 17). Briefly, the recognition of Z-nucleic acids by ADAR1 and ZBP1 is essential for the immune system's ability to detect and respond to viral threats effectively (18). Conversely, viruses have evolved Zα-containing proteins like E3L to subvert and antagonize host immune defenses, thereby enhancing their own virulence and survival within the host organism (19, 20). However, Zα domains have functional roles that need to be nuanced, as they cannot always be swapped, especially across different viruses (20, 21). This level of complexity implies an evolutionary balance between a virus and its host, where the unique characteristics of Zα domains are finely tuned to suit the particular dynamics of each virus-host relationship.

Because Z-nucleic acid biology is now viewed as a determinant for cell fate during infection and auto-immune diseases, one could ask the question of whether more proteins

---

with Zα domains could be discovered, and in what species. Recent computational efforts based on structure similarity search have unveiled the existence of 14 potential Zα domains (22). Additionally, two other potential ZBPs were identified, namely RBP7910 in *Trypanosoma brucei* (23), and DprA in *Riemerella anatipestifer* (24). However, these studies did not address the absence of the complete set of required amino acids for binding Z-nucleic acids, and no experimental validations of these putative Zα domains were carried out.

Here, we introduce a computational approach for predicting Zα domains, which combines primary sequence analysis, three-dimensional modeling, and manual examination of the resulting hits through structural analysis. This method led to the discovery of 8 putative Zα domain-containing proteins in eukaryotic organisms, and 68 in giant viruses diverse and mainly unexplored giant viruses in the viral phylum *Nucleocytoviricota*. We also report Zα-like domains in bacterial proteins, which deviate somewhat from their eukaryotic and viral counterparts in primary amino acid sequence. Finally, we have experimentally validated our predictions by assessing the Z-DNA binding capacity and ability to shift the B-Z equilibrium for two Zα and two Zα-like candidates. This work offers fresh insights into the functional role of Zα domains in virus sensing and innate immunity. Furthermore, it may shed light on previously undiscovered functions of ZBPs, expanding our understanding of their roles in biological processes.

Results and discussion

A combinatory method to predict Zα-domain and Zα-like candidates

Initially, we employed BLAST search tools in conjunction with primary sequence analysis by domain prediction software SMART and PROSITE to identify the Zα-domain candidates. Subsequently, 3D structure analysis and prediction tools were used to refine our predictions. Leveraging our existing knowledge of the essential attributes of the Zα-domain, we further refined the results derived from our *in silico* analysis (Fig. 1A). We used multiple sequence alignments to compare the candidate sequences with some representative Zα domains in all 6 known ZBPs. We checked the presence of residues crucial for the functionality of the Zα-domain by primary sequence alignment. Among these, the NxxxY motif within α3 is essential for Z-DNA binding, while the pPxW motif in the β-wing plays a significant role in both Z-DNA (10) binding and A-to-Z conversion (25). In the crystal structure of the HsZα$_{ADAR1}$, tryptophan 195 has been demonstrated to be vital not only for protein stability but also for DNA binding. The double mutation involving asparagine 173 and tyrosine 177 within the α3 helix completely abolishes Zα's capacity to bind Z-DNA, rendering it functionally inoperative for both mammalian and viral ZBPs. Additionally, proline 192 and proline 193 contribute significantly through van der Waals interactions with DNA (10). Despite the limited overall sequence similarity typical to Zα domains, the aliphatic residues (marked in red) from the three helices, along with Tryptophan 195 in the β-wing, exhibit strong conservation in

all candidates. These residues are well-established for their role in supporting the hydrophobic core of the winged helix-turn-helix structure (Figs. 1B, and S1) (10). In line with this, the predicted 3D structures, generated by AlphaFold (26), reveal that all candidates adopt the characteristic winged-HTH conformation. The 3D fold comparisons underscore a remarkable degree of structural similarity between these candidates and the crystallized ADAR1 Zα domain, PDB ID 2GXB (27) (Fig. 1, C and D). Despite these similarities, some candidates found in bacteria lack conserved residues in the corresponding positions N173 and Y177 in HsZα$_{ADAR1}$ displaying instead a serine and an asparagine respectively. As such, we classify them as Zα-like candidates (Figs. 1B and S1).

Identification of novel Zα-domain candidates in eukaryotic organisms

All predicted Zα candidates maintain the conserved residues N173 and Y177 in HsZα$_{ADAR1}$ crucial for Z-DNA binding (10). In Anthozoa, a class of marine invertebrates that includes sea anemones, stony corals, and soft corals, we predicted a new Zα containing protein. Interestingly, in some families such as the Poritidae or Actinidea, the novel Zα displays all important residues found in the known Zα. However, some families such as Acroporidae or Caryophylliidae display a Zα domain in which the Y177 is substituted by phenylalanine and W195 is replaced by a tyrosine (Fig. S1). We predicted other potential Zα domain-containing proteins in other eukaryotic organisms out of metazoan and DNA viruses, such as a protein annotated as transcription activator p15 in phytoplankton, and another annotated as acetate-CoA ligase in unicellular microalgae of the Symbiodiniaceae family (Figs. S1 and S3).

Identification of novel Zα-domain candidates in giant viruses

The most substantial contingent of novel ZBP candidates emerged within giant virus metagenome-assembled genomes (GVMAGs), revealing the presence of potential ZBPs in members of the orders pimascovirales, pandoravirales, and imitervirales. These proteins are distinct from previously documented ZBPs in fish-herpesviruses, poxviruses, and African swine fever virus (ASFV) (Fig. 2). These giant viruses were recovered in our previous metagenomic survey of samples from marine and freshwater, soil, and thermal environments (Fig. S2) (28), but also included additional sequences from deep-sea sediments (29) and permafrost (from 49 to 53,000 years old) (30). Our analysis shows that the candidate ZBPs in these giant viruses share similar domain architectures as in previously identified viral-containing proteins (13, 15, 20). The only exception is the viral ZBP candidate in which Zα is associated with tRNA metabolism domains (IMGM3300021083) (Fig. S3).

In our phylogenetic analysis, the clade topology in the ZBPs protein tree does not align with the topology found in the *Nucleocytoviricota* species tree (Fig. 2). This finding suggests that proteins with Zα domains have been independently acquired, likely multiple times, originating from a specific host *via* gene transfer. This possibility, as proposed by Kuś *et al.*,
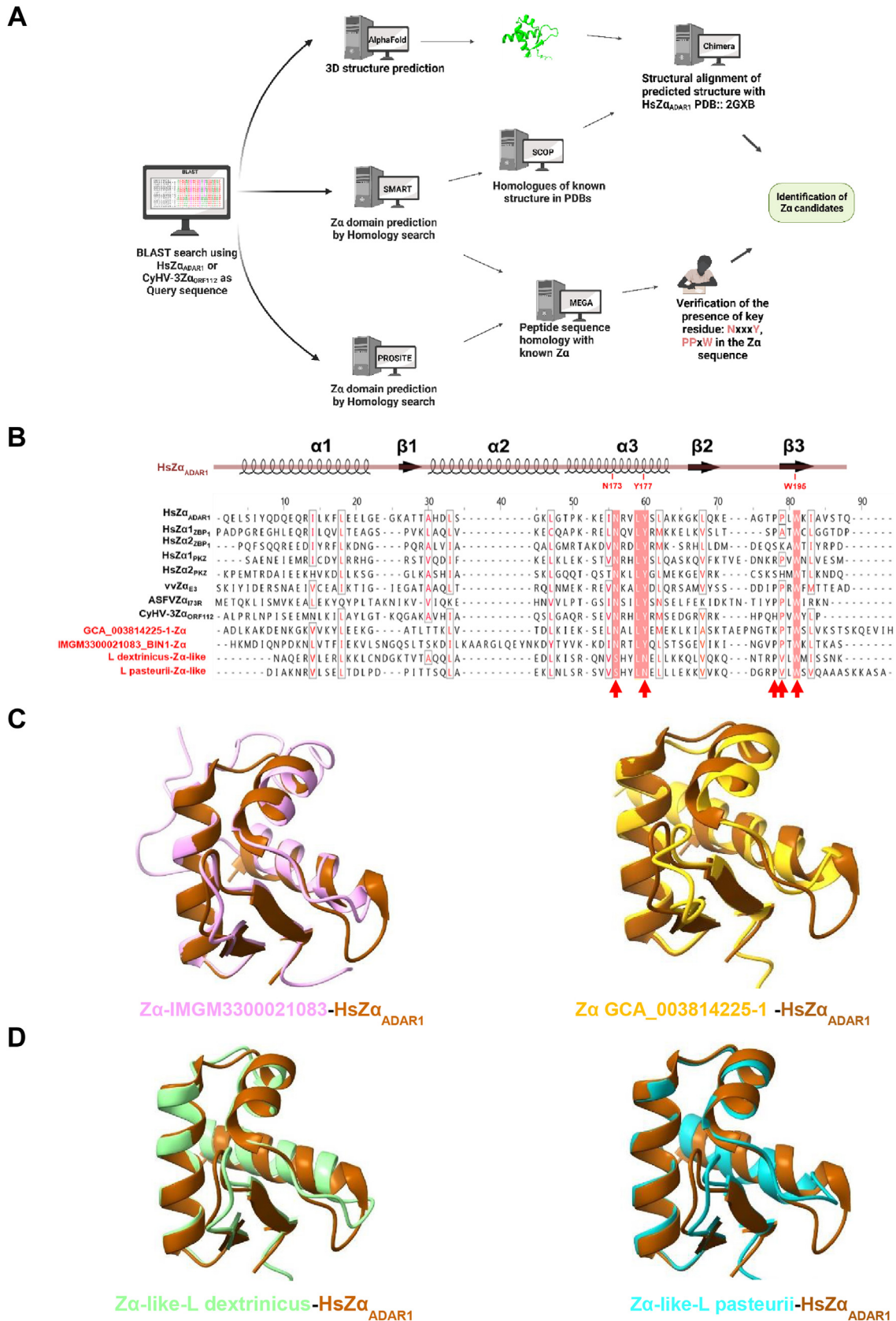
Figure 1. Prediction of new Zα candidates. *A*, the summary of the workflow for identifying Zα candidates. *B*, alignment of two predicted Zα and two Zα-like with known Zα domains from human and mouse ADAR1 (HsZα$_{ADAR1}$), human ZBP1 (HsZα$_{ZBP1}$), and viral from vaccinia virus and cyprinid-herpesvirus-3 (vvZα$_{E3}$ ASFV Zα$_{I73R}$ CyHV-3Zα$_{112}$). The consensus structure Zα domain is shown at the *top*. The key DNA-interacting residues in ADAR1 (corresponding to Asn174, Tyr177, P192, P193, and Trp195 in HsZα$_{ADAR1}$) are pointed with *red arrows*. Hydrophobic residues in the hydrophobic core are highlighted in *red front*. *C*, structural alignment of AlphaFold predicted structure of Zα candidates with HsZαADAR1 (*brown*, PDB ID 2GXB). *D*, structural alignment of AlphaFold predicted structure of Zα-like candidate with HsZα$_{ADAR1}$ (*brown*, PDB ID 2GXB).
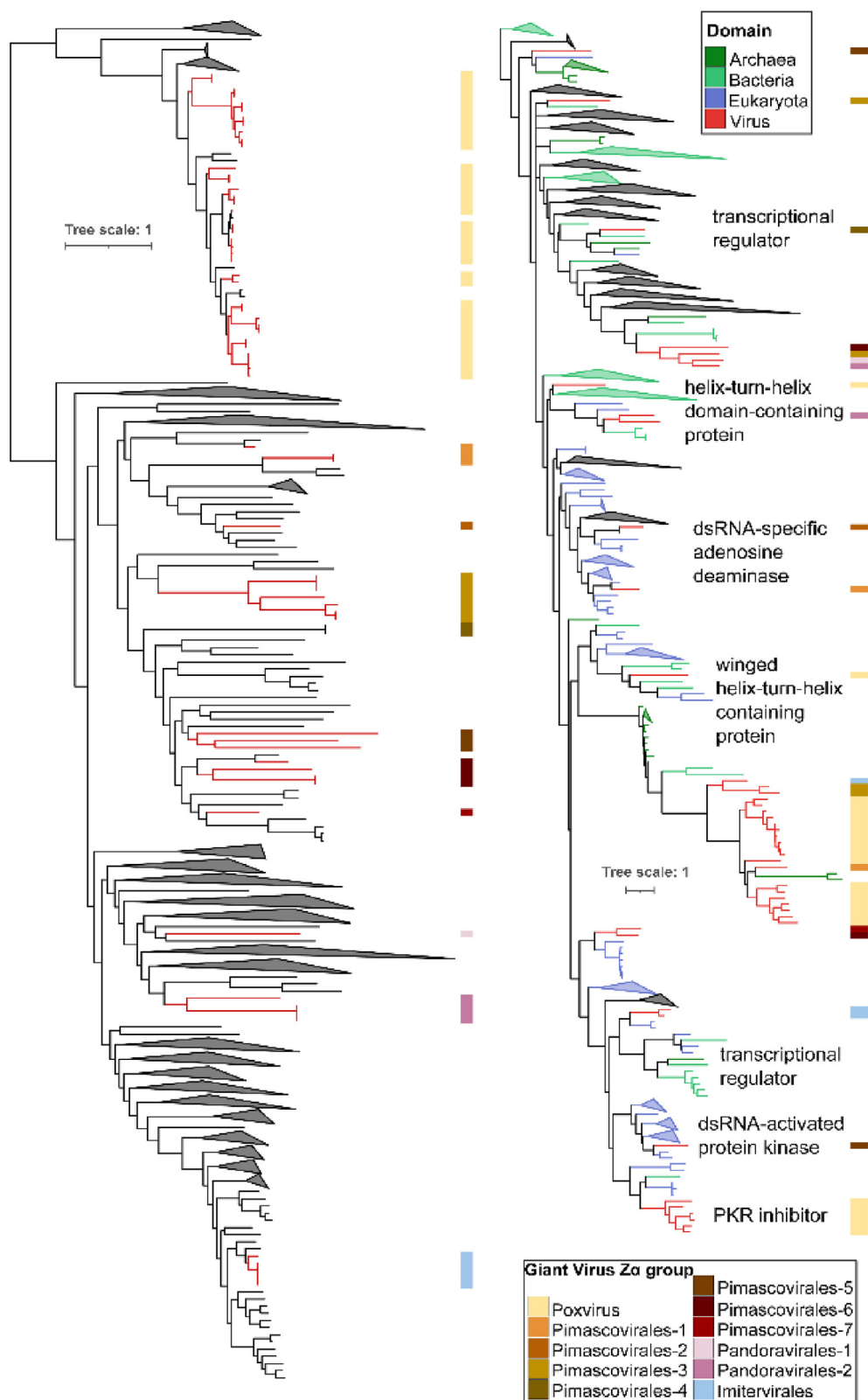
Figure 2. Phylogenetic tree of proteins that contain Zα and Zα-like domains. Proteins that were included in the analysis are giant virus ZBPs and homologs identified and extracted from the NCBI nr database.

suggests that CyHV-3 ORF112 might have been co-opted independently from the host by the common ancestor of the cyprinid herpesviruses, rather than through horizontal transfer from a poxvirus (31). Such scenarios are further supported by the fact that in our tree, proteins from different giant virus lineages group together and are simultaneously

intertwined in monophyletic clades with eukaryotic and archaeal homologs.

Our results imply that giant viruses with the Zα domain likely infect a wide range of organisms ranging from dinoflagellates to marine vertebrates and arthropods as well as early diverging metazoans such as cnidarians (32). The presence of ZBP in diverse *Nucleocytoviricota* lineages raises the question of the role of Zα in the interactions of these viruses with their hosts, and how this fits into our current understanding of the role of ZBPs in animal innate immunity. Furthermore, the manifold presence of ZBPs in divergent giant virus lineages, known to infect protists and algae, sparks curiosity. This observation is particularly intriguing given that virus-protist interactions entail mechanisms distinctly different from the innate immunity found in animals (33). Could the presence of ZBP indicate that these giant viruses infect metazoan hosts? In line with this, the ASFV infects two phylogenetically distant hosts, functioning as an arthropod-borne DNA virus (34), significantly impacting pig farming due to its high virulence (19).

Experimental validation of Zα and Zα-like candidates

To validate our prediction, we experimentally studied some of these Zα and Zα-like domains by circular dichroism, an established method to probe their ability to induce B(A)-to-Z conversion (35). We found that two predicted Zα domains from two distinct members of the viral order pimascovirales are able to efficiently convert B-DNA duplex of different lengths d(CpG)$_3$ and d(CpG)$_6$ to Z-form (Fig. 3, *A* and *B*). Interestingly, we found that these two Zα domains are not able to induce A-to-Z conversion of RNA duplex as seen for the positive control HsZα$_{ADAR1}$ or in the high salt concentration condition (Fig. 3C). Consequently, our study identifies new cases of Zα capable of converting B-DNA to Z-form but not A-RNA. Indeed, the conversion of A-RNA to Z-RNA is a less favorable and energetically costly process compared with B-DNA to Z-DNA conversion (35). To test the binding capacity to sequence contexts that have a lower B-Z energy barrier or preformed Z-form nucleic acids, we used a singly methylated RNA duplex 8mG4 r(CpG)$_3$ which stabilized 50% of the population in Z-form (36, 37) (Fig. 3D). In this case, the Zα domains are capable of flipping the RNA to Z-form. This indicates that Zα domains bind to the pre-stabilized Z-RNA since they cannot flip A-RNA to the Z-form.

It is interesting to speculate on the distinction between competent A/B-to-Z flipping Zα domains, Zα domains that can only flip B-DNA, and the nonfunctional Zα-like domains, and their evolutionary/functional relevance. Numerous studies have tried to identify key residues/regions within the Zα domain that are important for binding and flipping nucleic acids to the Z-conformation. Kim *et al.* were able to restore the partial function of ADAR1's Zβ domain by mutating Ile335 back to a tyrosine as is typical for all other ZBDs (38). Additionally, residues within the β-wing play a large role in not only the stabilization of the Z-conformation but also the rate at which it converts nucleic acids to the Z-conformation (39, 40).

Furthermore, distantly related wHTH domains with no ability to convert nucleic acids to the Z-conformation can be converted into good Z-binders and converters through directed mutagenesis (41). It is likely that numerous factors contribute to the ability of any given Zα domain's ability to bind and convert nucleic acids to the Z-conformation, and the loss of one stabilizing interaction may result in the inability to convert A-RNA while still retaining the ability to convert B-DNA.

As expected, we found that for the Zα-like domain, there was no conversion of B-DNA to Z-DNA or A-RNA to Z-RNA, in contrast to our positive control HsZα$_{ADAR1}$ (Fig. 3, *E* and *F*). To test the binding capacity to sequence contexts that have a lower B-Z energy barrier or preformed Z-form nucleic acids, we used a methylated DNA duplex d(m^{5}CpG)$_3$ and the singly methylated RNA duplex 8mG4 d(CpG)$_3$ respectively (36, 37). The Zα-like domains show no ability to stabilize or shift the B-Z transition equilibrium of the Z-nucleic acid analogs (Fig. 3, *G* and *H*). The Zα-like domain is found in the bacterial protein annotated as sigma 54 (σ^{54})-interacting transcriptional regulator. σ^{54} also encompasses an ATPase AAA region, which overlaps with the Holliday junction DNA helicase RuvB. RuvB is a component of the RuvABC revolvasome, crucial for resolving Holliday junctions (42). Remarkably, both Holliday junctions and Z-DNA are found in bacterial biofilms (43, 44). Studies have shown that RuvA protein or DNABII stabilizes Holliday junctions, inducing supercoiling that reduces the energy needed for B-Z transition and promotes Z-DNA formation in the extracellular DNA biofilm matrix (43). Therefore, it would be interesting to test the B-to-Z conversion of the predicted association between the Zα-like domain and the helicase RuvB in σ^{54}. Moreover, knowing that bacteria of the genius of *Lactobacillus* are biofilm-forming bacteria (45), we hypothesize that σ^{54} could be involved in the biofilm matrix formation.

Experimental procedures

In silico prediction of Zα domain

To identify new Zα domain candidates, we developed a combinatorial approach. Using the Basic Local Alignment Search Tool, blastp algorithm of NCBI (46), we performed taxid-exclusion blast and taxid-targeted blast using the Zα peptides from CyHV-3 and HsADAR1 as the query sequences. The candidates with a minimum of three crucial residues conserved or featuring less drastic substitutions (such as in *A digitifera*-Zα, where phenylalanine replaces the tyrosine in helix-α3 and a tyrosine replaces the tryptophane in the β-wing Fig. S1), were considered as the best hits and were selected. Their primary sequences were subjected to Zα domain prediction using SMART (47) and PROSITE (48, 49) software enabling us to determine the peptide length covered by the Zα domain. Leveraging our knowledge of the essential attributes of the Zα-domain, we further refined the results derived from our *in silico* prediction. We check the presence of certain residues crucial for the functionality of the Zα-domain by primary sequence alignment using MEGA-11 (50) and UGENE (51) tools. The presence or absence of these critical motifs served as
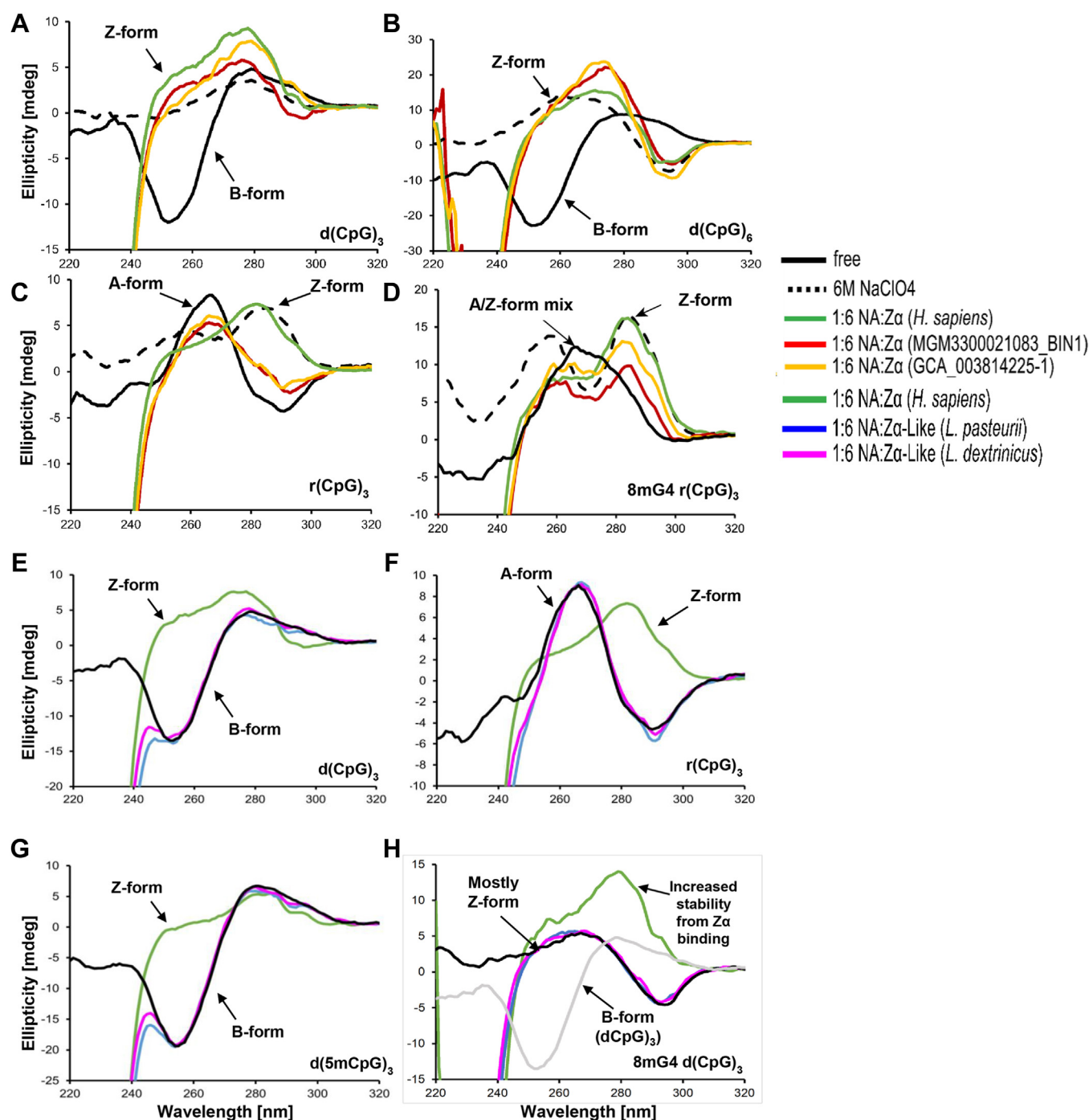
Figure 3. Experimental validation by circular dichroism. *A* and *B*, the predicted viral Zα domains induce B-to-Z conversion of DNA. *C*, viral Zα domains do not convert A-RNA to Z-RNA. *D*, viral Zα domains do convert an RNA with prestabilized Z-conformation (A:Z form ~50%/50%) to full Z-RNA. *E* and *F*, Zα –like domains are not able to induce B(A)-to-Z conversion. *G* and *H*, Zα –like domains do not bind to Z-nucleic acids. For comparison, the profile of a B-form DNA is shown in *grey* in (*H*), demonstrating the presence of a large portion of Z-form in 8mG4 d(CpG)3. The profiles overlaid with profiles from second independent measurements are provided in Fig. S4.

the final determinant in our analysis, allowing us to confidently identify Zα candidates. Any candidates lacking tyrosine and asparagine were categorized as Zα-like, given their alignment with other defining characteristics. SMART software also outputs structure primary sequence homologs of known in the protein data bank (PDB) or blast with known SCOPs (structural classification of proteins) of the Zα, for example, **d1qbja** (52).

Finally, to strengthen the Zα prediction, a structure prediction using AlphaFold (26, 53) confirms the winged helix-turn-helix fold superposable with the crystallized Zα available with the PDBs using UCSF ChimeraX (54). With the predicted 3D structures, we performed a 3D-BLAST search (55) to confirm that the best hits are entries corresponding to the Zα domain. Figures 1*A* and S3 were created with BioRender.com.

Sequence source and annotation

Most of the giant viruses used in our analysis were identified in our metagenomic analysis performed in samples from different biotopes (**S2**) (28). We also included other giant viruses identified in other studies such as Pithovirus (29) and Marseillesvirus (30). The sequencing of the full viral genome or the viral genome fragment covering the coding sequence of the ZBP allowed the complete identification of the open reading frame corresponding to the ZBP. For the other organisms, the ZBPs are annotated at the PDB genomic level and (protein and mRNA level for *Micromonas pusilla* (56). Zα domain sequence and the ID numbers of ZBPs are listed in Table S6.

Phylogenetic analysis

Novel Zα and Zα-like domain sequences were clustered at 40% identity using cd-hit (57) and the resulting cluster representative sequences were aligned using mafft-linsi (version 7.505) (58). We used the hmmbuild (version 3.3.2, hmmer.org) to create a profile hidden Markov model from the representative sequence alignment and then used it as a query for the hmmsearch program (version 3.3.2, hmmer.org) on the NCBI nr database (59). Protein sequences that had an overall e-value <0.005 and contained at least one domain with length ≥ 60 amino acids were considered as significant hits to the model. Then we de-replicated the protein sequences with significant hits with cd-hit at 90% identity and an alignment length covering at least 60% the length of the longer sequence (-aL 0.6). The representative protein sequences were aligned together with Zα containing sequences from the viral phylum Nucleocytoviricota using the mafft FFT-NS-2 method (58). The alignment was trimmed with trimal with the -gt 0.1 option (60). A protein tree was built using IQ-Tree (version 1.6.12) (61) with 1000 ultra fast bootstrap replicates (62) and the JTT+R10 model, which was selected with the ModelFinder feature in IQ-Tree (63). The tree was visualised using iToL (64).

To build the giant virus species tree the nsgtree pipeline was used (https://github.com/NeLLi-team/nsgtree) on representative genome of the viral phylum Nucleocytoviricota (65) and additional GVMAGs that contained proteins with a Zα domain. In brief, 7 giant virus orthologous groups (GVOGs) were identified using hmmsearch (version 3.1/b2, hmmer.org), extracted GVOGs were aligned with mafft (version 7.31) (58), trimmed with trimal (-gt 0.1, v1.4) and concatenated. A species tree was built from the supermatrix alignment using IQ-Tree (version 2.03) (66) with LG+F + I + G4 and visualized in iToL (64).

Protein expression and purification

The Zα domains from *Homo sapiens* ADAR1, L. pasteurii, *L. dextrinicus*, giant virus (IMGM3300021083 and GCA_003814225–1) were cloned into the pet-28a(+) plasmid (N-terminal 6x His-tag and thrombin cleavage site between His tag and the Zα sequence). The plasmids were transformed and expressed in BL21(DE3) *E. coli* cells. The cell cultures were grown in Luria Broth (LB) to an OD_{600} of ~0.4 and induced with IPTG to a final concentration of 1 mM and allowed to express Zα for 4 h at 37 °C, then centrifuged to collect the cell pellets. Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 10 mM imidazole, 1 mM BME) and sonicated. Lysate was centrifuged and the supernatant was applied to a His-trap column, (50 mM Tris-HCl (pH 8.0), 1 M NaCl, 10 mM imidazole, 1 mM BME), and eluted in 20 ml of elution buffer washed with 40 ml of lysis buffer, 80 ml of wash buffer (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 500 mM imidazole, 1 mM BME). The eluents were concentrated to ~2 ml and applied to a HiLoad 16/600 Superdex 200 prep grade Gel Filtration Column (GE Healthcare) and the peak corresponding to pure protein (which elutes at ~80 ml in 20 mM potassium phosphate (pH 6.4), 300 mM NaCl) was collected and concentrated using an Amicon 3 kDa cutoff centrifugal filter (Millipore-Sigma, Burlington, MA). Zα was dialyzed and concentrated into 20 mM potassium phosphate (pH 6.4), 25 mM, 0.5 mM EDTA and concentrated to ~2 mM using an Amicon 3 kDa cutoff centrifugal filter (Millipore-Sigma). Subsequent dilutions were made in 20 mM potassium phosphate (pH 6.4), 25 mM, 0.5 mM EDTA as needed for the different experiments. We confirmed that all recombinant proteins were free of contaminating nucleic acid (Fig. S5).

DNA and RNA constructs and preparation

The d(CpG)₃, d(CpG)₆, and d(5mCpG)₃ constructs were synthesized by Integrated DNA Technologies. The r(CpG)₃ construct was synthesized by Dharmacon (a part of Horizon Discovery). The 8mG4 d(CpG)₃ construct was synthesized by the Yale School of Medicine oligo synthesis resource. All nucleic acid constructs in this study were heat annealed prior to use at 95 °C for 5 minutes followed by slow cooling at room temperature for 30 min. For circular dichroism (CD) measurements, all of the constructs were 50 μM in 20 mM potassium phosphate, 25 mM NaCl (pH 6.4), 0.5 mM EDTA, or the same buffer but with 6 M NaClO₄ or a 1:6 M ratio of nucleic acid:Zα.

Circular dichroism

All CD measurements were collected using a JASCO J-815 CD spectrometer (run using Spectra Manager version 2 (JASCO)) in a 0.1 cm quartz cuvette. Z-form adoption by high-salt and Zα binding was carried out by incubating the constructs in 6M NaClO₄ or with saturating amounts of Zα at a 1:6 M ratio of nucleic acid:Zα for 1 h at 42 °C before cooling to 25 °C and measuring. Spectra were collected in 1-nm steps from 320 to 220 nm with an average of two scans. All measurements with unmodified nucleic acids as well as d(5mCpG)₃ and 8mG4 r(CpG)₃ were conducted twice, and the profiles shown in Figure 3 overlaid with profiles from second independent measurements are provided in Fig. S4. All duplicate profiles confirmed the high reproducibility of the measurements.

Conclusion

In summary, our study employed a computational combined with an experimental approach that enabled the

identification of potential Zα-containing proteins across various organisms. These ZBPs open up new avenues in studies of Zα domains and Z-nucleic acids in virus sensing and innate immunity. The domain architectures within predicted Z-DNA binding proteins imply that they potentially serve functions distinct from those currently attributed to the known ZBPs (Fig. S3). These functions could be conserved in mammals but in the form of synergy between a ZBP and another protein like the cooperative function between ADAR1$_{p150}$ and DHX9 (67). Furthermore, our experimental validation of two Zα-containing proteins confirms not only their presence in currently uncultivated giant viruses but also their functionality. Our findings open new questions regarding the role of Zα domains in these giant viruses and suggest that the interplay between ZBPs/Z-nucleic acid is an evolutionarily ancient feature as shown for the cGAS/STING pathway (68). It is tempting to speculate that the ZBPs we predicted in giant viruses and early diverging metazoan could be involved in an ancient pathogen detection mechanism that could provide functional and evolutionary insight into innate immune signaling pathways in animals. Our validation results underscore the need for heightened caution in Zα domain prediction. We demonstrate that the mere detection of a Zα domain by prediction tools or the observation of structural similarities between predicted wHTH and crystallized Zα domains fall short of substantiating experimental B(A)-to-Z conversion or Z-nucleic acid-binding. These findings emphasize that the ability to bind to Z-nucleic acid should be the definitive criterion for establishing a wHTH as a true Zα domain.

In our study, we focused on validating the Zα and Zα-like domains found mainly in multi-domain proteins. Although we demonstrate that wHTH domains we classify as Zα-like cannot bind Z-nucleic acid nor convert B-to-Z, we cannot exclude the possibility that the proteins with Zα-like domains may still be involved in Z-nucleic biology. This is because other proteins lacking Zα domains have been found to bind Z-DNA (69) or play a role in regulating Z-RNA sensing (70).

Data availability

The data that support the findings of this study are available from the corresponding authors upon request. The primary sequence and the id-numbers of the new ZBPs and Zα-like containing proteins are summarized in Fig. S3 and Table S6. Other sequences used in this work are available from Frederik Schulz and Mamadou Amadou Diallo upon reasonable request.

References

1. Wang, A. H. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., Van Boom, J. H., Van Der Marel, G., *et al.* (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686
2. Hall, K., Cruz, P., Tinoco, I., Jovin, T. M., and Van De Sande, J. H. (1984) Z-RNA'–a left-handed RNA double helix. *Nature* **311**, 584–586
3. Miglietta, G., Russo, M., and Capranico, G. (2020) G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders. *Nucleic Acids Res.* **48**, 11942–11957
4. Karki, R., and Kanneganti, T. D. (2023) ADAR1 and ZBP1 in innate immunity, cell death, and disease. *Trends Immunol.* **44**, 201–216
5. Nichols, P. J., Krall, J. B., Henen, M. A., Vögeli, B., and Vicens, Q. (2023) Z-RNA biology: a central role in the innate immune response? *RNA* **29**, 273–281
6. Song, J., Gooding, A. R., Hemphill, W. O., Love, B. D., Robertson, A., Yao, L., *et al.* (2023) Structural basis for inactivation of PRC2 by G-quadruplex RNA. *Science* **381**, 1331–1337
7. Thapa, R. J., Ingram, J. P., Ragan, K. B., Nogusa, S., Boyd, D. F., Benitez, A. A., *et al.* (2016) DAI senses influenza A virus genomic RNA and activates RIPK3-dependent cell death. *Cell Host Microbe* **20**, 674–681
8. Lei, Y., VanPortfliet, J. J., Chen, Y.-F., Bryant, J. D., Li, Y., Fails, D., *et al.* (2023) Cooperative sensing of mitochondrial DNA by ZBP1 and cGAS promotes cardiotoxicity. *Cell* **186**, 3013–3032.e22
9. Herbert, A. G., Spitzner, J. R., Lowenhaupt, K., and Rich, A. (1993) Z-DNA binding protein from chicken blood nuclei. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3339–3342
10. Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A., and Rich, A. (1999) Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* **284**, 1841–1845
11. Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U., and Rich, A. (2001) Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.* https://doi.org/10.1038/nsb0901-761
12. Rothenburg, S., Deigendesch, N., Dittmar, K., Koch-Nolte, F., Haag, F., Lowenhaupt, K., *et al.* (2005) A PKR-like eukaryotic initiation factor 2alpha kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1602–1607
13. Kim, Y.-G., Muralinath, M., Brandt, T., Pearcy, M., Hauns, K., Lowenhaupt, K., *et al.* (2003) A role for Z-DNA binding in vaccinia virus pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6974–6979
14. Tome, A. R., Kus, K., Correia, S., Paulo, L. M., Zacarias, S., de Rosa, M., *et al.* (2013) Crystal structure of a poxvirus-like Zalpha domain from cyprinid herpesvirus 3. *J. Virol.* **87**, 3998–4004

15. Sun, L., Miao, Y., Wang, Z., Chen, H., Dong, P., Zhang, H., *et al.* (2022) Structural insight into African swine fever virus I73R protein reveals it as a Z-DNA binding protein. *Transbound. Emerg. Dis.* **69**, e1923–e1935

16. Liddicoat, B. J., Piskol, R., Chalk, A. M., Ramaswami, G., Higuchi, M., Hartner, J. C., *et al.* (2015) RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, 1115–1120

17. Oh, D.-B., Kim, Y.-G., and Rich, A. (2002) Z-DNA-binding proteins can act as potent effectors of gene expression *in vivo*. *Proc. Nat. Acad. Sci. U. S. A.* **99**, 16666–16671

18. Koehler, H., Cotsmire, S., Zhang, T., Balachandran, S., Upton, J. W., Langland, J., *et al.* (2021) Vaccinia virus E3 prevents sensing of Z-RNA to block ZBP1-dependent necroptosis. *Cell Host Microbe* **29**, 1266–1276.e5

19. Liu, Y., Shen, Z., Xie, Z., Song, Y., Li, Y., Liang, R., *et al.* (2023) African swine fever virus I73R is a critical virulence-related gene: a potential target for attenuation. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2210808120

20. Diallo, M. A., Pirotte, S., Hu, Y., Morvan, L., Rakus, K., Suárez, N. M., *et al.* (2023) A fish herpesvirus highlights functional diversities among Zα domains related to phase separation induction and A-to-Z conversion. *Nucleic Acids Res.* **51**, 806–830

21. Vijaysri, S., Talasela, L., Mercer, A. A., Mcinnes, C. J., Jacobs, B. L., and Langland, J. O. (2003) The Orf virus E3L homologue is able to complement deletion of the vaccinia virus E3L gene *in vitro* but not *in vivo*. *Virology* **314**, 305–314

22. Bartas, M., Slychko, K., Brázda, V., Červeň, J., Beaudoin, C. A., Blundell, T. L., *et al.* (2022) Searching for new Z-DNA/Z-RNA binding proteins based on structural similarity to experimentally validated Zα domain. *Int. J. Mol. Sci.* **23**, 768

23. Nikpour, N., and Salavati, R. (2019) The RNA binding activity of the first identified trypanosome protein with Z-DNA-binding domains. *Sci. Rep.* **9**, 5904

24. Huang, L., Tian, X., Liu, M., Wang, M., Biville, F., Cheng, A., *et al.* (2019) DprA is essential for natural competence in riemerella anatipestifer and has a conserved evolutionary mechanism. *Front. Genet.* **10**, 429

25. Langeberg, C. J., Nichols, P. J., Henen, M. A., Vicens, Q., and Vögeli, B. (2023) Differential structural features of two mutant ADAR1p150 Zα domains associated with aicardi-goutières syndrome. *J. Mol. Biol.* **435**, 168040

26. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589

27. Placido, D., Brown, B. A., Lowenhaupt, K., Rich, A., and Athanasiadis, A. (2007) A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure* **15**, 395–404

28. Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D. A., Denef, V. J., *et al.* (2020) Giant virus diversity and host interactions through global metagenomics. *Nature* **578**, 432–436

29. Bäckström, D., Yutin, N., Jørgensen, S. L., Dharamshi, J., Homa, F., Zaremba-Niedwiedzka, K., *et al.* (2019) Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. *mBio* **10**, e02497-18

30. Rigou, S., Christo-Foroux, E., Santini, S., Goncharov, A., Strauss, J., Grosse, G., *et al.* (2022) Metagenomic survey of the microbiome of ancient Siberian permafrost and modern Kamchatkan cryosols. *microLife* **3**, uqac003

31. Kuś, K., Rakus, K., Boutier, M., Tsigkri, T., Gabriel, L., Vanderplasschen, A., *et al.* (2015) The structure of the cyprinid herpesvirus 3 ORF112-Zα·Z-DNA complex reveals a mechanism of nucleic acids recognition conserved with E3L, a poxvirus inhibitor of interferon response. *J. Biol. Chem.* **290**, 30713–30725

32. Schulz, F., Abergel, C., and Woyke, T. (2022) Giant virus biology and diversity in the era of genome-resolved metagenomics. *Nat. Rev. Microbiol.* **20**, 721–736

33. Queiroz, V. F., Tatara, J. M., Botelho, B. B., Rodrigues, R. A. L., Almeida, G. M. F., and Abrahao, J. S. (2024) The consequences of viral infection on protists. *Commun. Biol.* **7**, 306

34. Basto, A. P., Nix, R. J., Boinas, F., Mendes, S., Silva, M. J., Cartaxeiro, C., *et al.* (2006) Kinetics of African swine fever virus infection in Ornithodoros erraticus ticks. *J. Gen. Virol.* **87**, 1863–1871

35. Brown, B. A., Lowenhaupt, K., Wilbert, C. M., Hanlon, E. B., and Rich, A. (2000) The Zα domain of the editing enzyme dsRNA adenosine deaminase binds left-handed Z-RNA as well as Z-DNA. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13532–13536

36. Krall, J. B., Nichols, P. J., Henen, M. A., Vicens, Q., and Vögeli, B. (2023) Structure and formation of Z-DNA and Z-RNA. *Molecules* **28**, 843

37. Nichols, P. J., Krall, J. B., Henen, M. A., Welty, R., Macfadden, A., Vicens, Q., *et al.* (2023) Z-form adoption of nucleic acid is a multi-step process which proceeds through a melted intermediate. *J. Am. Chem. Soc.* **146**, 694

38. Kim, Y. G., Lowenhaupt, K., Oh, D. B., Kim, K. K., and Rich, A. (2004) Evidence that vaccinia virulence factor E3L binds to Z-DNA *in vivo*: implications for development of a therapy for poxvirus infection. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1514–1518

39. Van quyen, D., Ha, S. C., Lowenhaupt, K., Rich, A., Kim, K. K., and Kim, Y. G. (2007) Characterization of DNA-binding activity of Zα domains from poxviruses and the importance of the β-wing regions in converting B-DNA to Z-DNA. *Nucleic Acids Res.* **35**, 7714–7720

40. Kim, D., Hur, J., Park, K., Bae, S., Shin, D., Ha, S. C., *et al.* (2014) Distinct Z-DNA binding mode of a PKR-like protein kinase containing a Z-DNA binding domain (PKZ). *Nucleic Acids Res.* **42**, 5937–5948

41. Park, C., Zheng, X., Park, C. Y., Kim, J., Lee, S. K., Won, H., *et al.* (2020) Dual conformational recognition by Z-DNA binding protein is important for the B-Z transition process. *Nucleic Acids Res.* **48**, 12957–12971

42. Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K. (2000) Crystal structure of the holliday junction DNA in complex with a single RuvA tetramer. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8257–8262

43. Devaraj, A., Buzzo, J. R., Mashburn-Warren, L., Gloag, E. S., Novotny, L. A., Stoodley, P., *et al.* (2019) The extracellular DNA lattice of bacterial biofilms is structurally related to Holliday junction recombination intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 25068–25077

44. Buzzo, J. R., Devaraj, A., Gloag, E. S., Stoodley, P., Bakaletz, L. O., Goodman Correspondence, S. D., *et al.* (2021) Z-form extracellular DNA is a structural component of the bacterial biofilm matrix. *Cell* **184**, 5740–5758.e17

45. Salas-Jara, M. J., Ilabaca, A., Vega, M., and García, A. (2016) Biofilm forming Lactobacillus: new challenges for the development of probiotics. *Microorganisms* **4**, 35

46. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410

47. Letunic, I., Khedkar, S., and Bork, P. (2021) SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* **49**, D458–D460

48. de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., *et al.* (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365

49. Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., *et al.* (2013) New and continuing developments at PROSITE. *Nucleic Acids Res.* **41**, D344–D347

50. Tamura, K., Stecher, G., and Kumar, S. (2021) MEGA11: molecular evolutionary Genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027

51. Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., *et al.* (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167

52. Tung, C. H., and Yang, J. M. (2007) FastSCOP: a fast web server for recognizing protein structural domains and SCOP superfamilies. *Nucleic Acids Res.* **35**, W438–W443

53. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682

54. Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., *et al.* (2018) UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25

55. Yang, J. M., and Tung, C. H. (2006) Protein structure database search and evolutionary classification. *Nucleic Acids Res.* **34**, 3646–3659

56. TSA: Noctiluca scintillans, MMETSP0253-doi:10.5281-zenodo.249982-Transcript-1000, transcribed RNA Sequence.National Library of Medicine Bethesda, MD 20894

ASBMB

J. Biol. Chem. (2024) 300(8) 107504 **9**

57. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152

58. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780

59. Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., *et al.* (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26

60. Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973

61. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274

62. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522

63. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589

64. Letunic, I., and Bork, P. (2021) Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296

65. Aylward, F. O., Moniruzzaman, M., Ha, A. D., and Koonin, E. V. (2021) A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol.* **19**, e3001430

66. Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., *et al.* (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534

67. Aktaş, T., Ilik, I. A., Maticzka, D., Bhardwaj, V., Pessoa Rodrigues, C., Mittler, G., *et al.* (2017) DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* **544**, 115–119

68. Li, Y., Slavik, K. M., Toyoda, H. C., Morehouse, B. R., de Oliveira Mann, C. C., Elek, A., *et al.* (2023) cGLRs are a diverse family of pattern recognition receptors in innate immunity. *Cell* **186**, 3261–3276.e20

69. Krishna, P., Morgan, A. R., and Van de Sande, J. H. (1991) Interaction of recA protein with left-handed Z-DNA. *Biochem. J.* **275**, 711–719

70. Herbert, A., and Poptsova, M. (2022) Z-RNA and the flipside of the SARS Nsp13 helicase: is there a role for flipons in coronavirus-induced pathology? *Front. Immunol.* **13**, 912717