Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences

Jennifer L. Meagher[a,1], Matthew Takata[b,1], Daniel Gonçalves-Carneiro[b], Sarah C. Keane[c], Antoine Rebendenne[b], Heley Ong[b], Victoria K. Orr[a], Margaret R. MacDonald[d], Jeanne A. Stuckey[a,e], Paul D. Bieniasz[b,f,2], and Janet L. Smith[a,e,2]

[a]Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109; [b]Laboratory of Retrovirology, The Rockefeller University, New York, NY 10065; [c]Biophysics Program and Department of Chemistry, University of Michigan, Ann Arbor, MI 48109; [d]Laboratory of Virology and Infectious Diseases, The Rockefeller University, New York, NY 10065; [e]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109; and [f]Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065

Edited by John M. Coffin, Tufts University, Boston, MA, and approved October 21, 2019 (received for review July 31, 2019)

Infection of animal cells by numerous viruses is detected and countered by a variety of means, including recognition of nonself nucleic acids. The zinc finger antiviral protein (ZAP) depletes cytoplasmic RNA that is recognized as foreign in mammalian cells by virtue of its elevated CG dinucleotide content compared with endogenous mRNAs. Here, we determined a crystal structure of a protein-RNA complex containing the N-terminal, 4-zinc finger human (h) ZAP RNA-binding domain (RBD) and a CG dinucleotide-containing RNA target. The structure reveals in molecular detail how hZAP is able to bind selectively to CG-rich RNA. Specifically, the 4 zinc fingers create a basic patch on the hZAP RBD surface. The highly basic second zinc finger contains a pocket that selectively accommodates CG dinucleotide bases. Structure guided mutagenesis, cross-linking immunoprecipitation sequencing assays, and RNA affinity assays show that the structurally defined CG-binding pocket is not required for RNA binding per se in human cells. However, the pocket is a crucial determinant of high-affinity, specific binding to CG dinucleotide-containing RNA. Moreover, variations in RNA-binding specificity among a panel of CG-binding pocket mutants quantitatively predict their selective antiviral activity against a CG-enriched HIV-1 strain. Overall, the hZAP RBD RNA structure provides an atomic-level explanation for how ZAP selectively targets foreign, CG-rich RNA.

zinc finger antiviral protein | HIV-1 | RNA | innate immunity

Animal cells possess a variety of mechanisms for the detection of nonself biomolecules. Such recognition can lead to the mobilization of innate and adaptive immune mechanisms that constitute the host's defenses against infectious agents. For example, foreign nucleic acids can be recognized by endosomal Toll-like receptors or cytoplasmic sensors, based on their inappropriate location within a cell or the presence or absence of modifications that mark RNA molecules as self or nonself (1).

The zinc finger antiviral protein (ZAP) is a cytoplasmic protein that confers a cell-autonomous protective effect against a variety of RNA viruses (2). Indeed, the replication of viruses as diverse as retroviruses, alphaviruses, filoviruses, and hepadnaviruses can be inhibited by ZAP (2–5). The ZAP protein includes an N-terminal ~227-aa RNA-binding domain (RBD) containing 4 CCCH zinc fingers (ZnF1 to ZnF4), sometimes referred to as N-ZAP, that is necessary, and to some extent sufficient, for antiviral function (2, 6). The remaining portions of the protein, whose functions are unknown, include a WWE domain and a C-terminal poly(ADP-ribose) polymerase domain that is present in the ZAP-L isoform but not in the ZAP-S isoform (7) (Fig. 1*A*). Cofactors reported to be important for ZAP activity include TRIM25, whose role in antiviral activity is unknown (8, 9), KHNYN, a predicted endonuclease that is important for ZAP antiviral activity against retroviruses but not against Sindbis virus (10), and the RNA exosome (11).

We recently found that ZAP discriminates self from nonself (viral) RNAs based on their dinucleotide content (12). Specifically, HIV-1 is largely resistant to inhibition by endogenously expressed ZAP, because it mimics the CG-poor status characteristic of endogenous human mRNAs. Removal of the CG-poor bias from the HIV-1 genome via synonymous mutation results in cytoplasmic viral RNA depletion and a lethal, ZAP-dependent replication defect (12). In addition, cross-linking immunoprecipitation sequencing (CLIP-seq) assays show that in human cells, ZAP binds directly and preferentially to mRNA elements containing CG dinucleotides (12). The low level of CG dinucleotides in endogenous mRNAs probably results from the cumulative depletion of CG dinucleotides from animal genomes via DNA methylation and deamination (13). Of note, many, but not all, viruses with RNA genomes (14–16), including HIV-1 (17), appear to mimic the low CG content of their hosts, perhaps to evade ZAP. Indeed, the identities of viruses against which ZAP does and does not exhibit antiviral activity is predicted by their high and low CG content, respectively (12).

Significance

Zinc finger antiviral protein (ZAP) protects cells from infection by diverse RNA viruses through its ability to specifically detect and deplete viral RNAs that have a greater frequency of CG dinucleotides than host messenger RNAs. We solved an X-ray crystal structure of the domain of ZAP that recognizes RNA and found that a crucial component of RNA recognition by ZAP is a pocket on the protein surface that can accommodate a CG dinucleotide but no other dinucleotides. The structure explains in atomic detail how ZAP is able to selectively recognize CG-rich viral RNAs as foreign and thereby provide hosts with a defense against viral infection.

Author contributions: J.L.M., M.T., D.G.-C., J.A.S., P.D.B., and J.L.S. designed research; J.L.M., M.T., D.G.-C., A.R., H.O., and V.K.O. performed research; S.C.K. and M.R.M. contributed new reagents/analytic tools; J.L.M., M.T., D.G.-C., A.R., J.A.S., P.D.B., and J.L.S. analyzed data; and J.L.M., M.T., D.G.-C., P.D.B., and J.L.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID codes 6UEJ and 6UEI). The CLIP-seq data have been deposited in the NCBI Gene Expression Omnibus database, https://www.ncbi.nlm.nih.gov/geo/ (GSE139667).

[1]J.L.M. and M.T. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: pbieniasz@rockefeller.edu or JanetSmith@umich.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1913232116/-/DCSupplemental.

First published November 12, 2019.

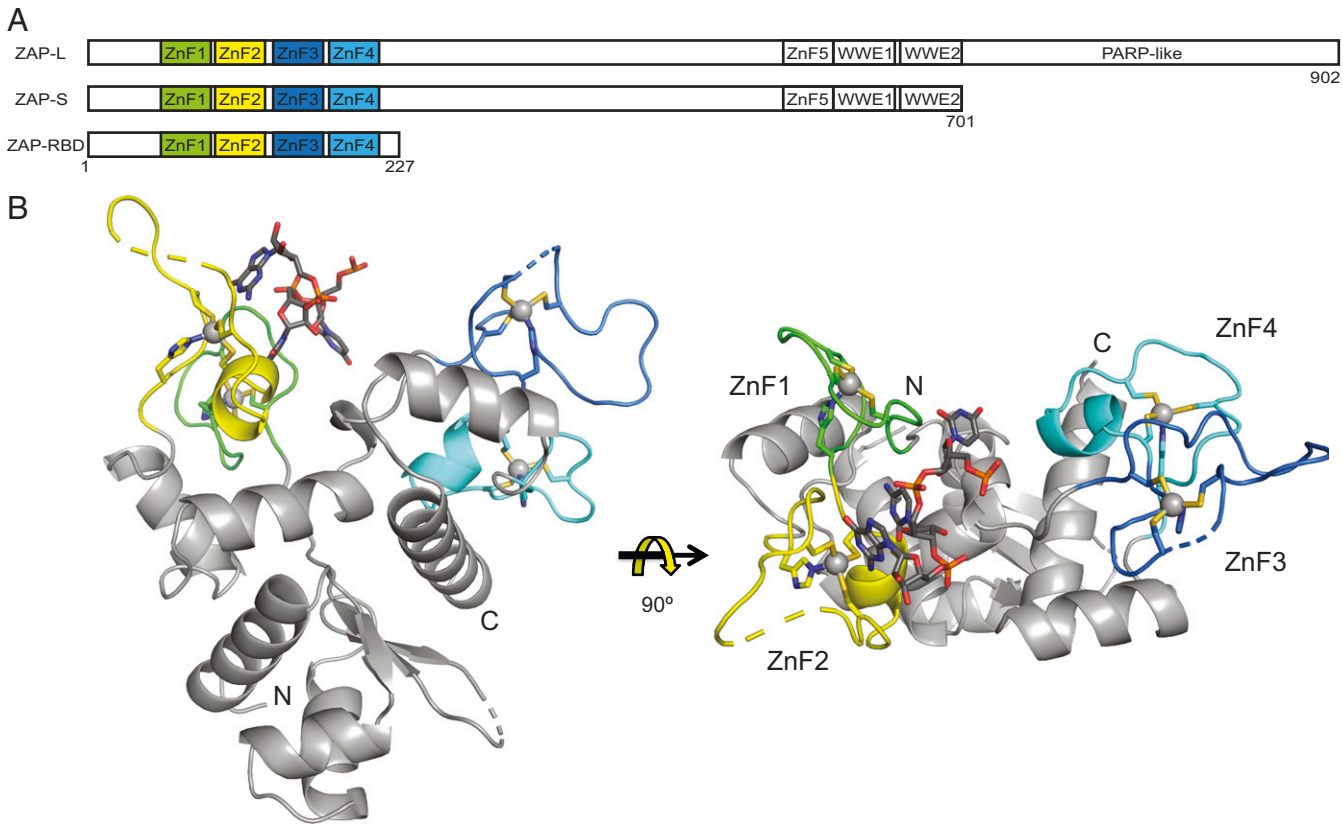

Fig. 1. hZAP RBD with bound RNA. (*A*) Schematic of hZAP showing the ZAP-L and ZAP-S isoforms and the positions of the Zn fingers in the RBD. (*B*) Structure of the hZAP RBD RNA complex. The 4 Zn fingers (ZnF) are colored separately (green, ZnF1; yellow, ZnF2; blue, ZnF3; cyan, ZnF4) in these views from the side (*Left*) and top (*Right*) of the hZAP RBD. The RNA CG dinucleotide binds to ZnF2. The UCG trinucleotide and the Zn ligands are shown in stick form with atomic colors (gray, C; red, O; blue, N; orange, P; yellow, S), and the Zn atoms are shown as gray spheres.

Here we describe the structural basis for the specific recognition of CG-containing RNA elements by ZAP. Specifically, we report an X-ray crystal structure of the human ZAP (hZAP) RBD (amino acids 2 to 227) in a complex with a CG-rich RNA target sequence. Of note, the hZAP RBD RNA complex reveals a binding pocket within a basic patch in the hZAP RBD that can accommodate a CG dinucleotide but no other dinucleotides. Structure-guided mutagenesis of the CG-binding pocket confirmed the biological relevance of the crystal structure, and we generated a panel of mutants with well correlated RNA-binding and antiviral specificity. In the most extreme case, the removal of a single oxygen atom in the CG-binding pocket (a Y to F substitution) nearly abolished the CG specificity that characterizes WT hZAP-L–RNA binding and antiviral activity. Overall, the structure of the hZAP RBD with its target explains how hZAP-L is able to specifically recognize foreign, CG-rich viral RNA.

Results

Structure of an ZAP(2-227) CG-Rich RNA Complex. To visualize RNA binding to ZAP, we determined the crystal structure of hZAP RBD with a CG-rich RNA oligomer (Fig. 1*B* and *SI Appendix*, Table S1) (18). For these experiments, we selected a 21-nt RNA sequence (5′AU**CG**ACUU**CG**AUU**CGCG**GGAA3′) from the site with the highest read count in CLIP-seq experiments with hZAP (12) and cocrystallized an equimolar mixture of hZAP RBD and the CG-rich oligonucleotide. RNA binding does not change the overall structure, as seen by comparison of hZAP RBD structures with and without RNA (rmsd 0.8 Å). Without RNA, helix 4 is shifted slightly into the CG dinucleotide site (*SI Appendix*, Fig. S1). Although the 21-nt RNA contains 4 CG dinucleotides and the hZAP RBD has 4 Zn fingers, RNA is bound only to ZnF2. Three nucleotides (5′-$U_1C_2G_3$-3′) were fit unambiguously into the electron density (*SI Appendix*, Fig. S2). No density for RNA was present in ZnF1, ZnF3, or ZnF4 even though none of these Zn fingers was blocked by a crystal contact. CG-dinucleotide binding to ZnF2 is sequence-specific, as no other dinucleotide could be accommodated in the CG dinucleotide site without steric clashes (Fig. 2*A*).

The C_2 pyrimidine base inserts into a hydrophobic pocket that is too small to accommodate a purine (Fig. 2*B*). The cytosine base forms 2 cytosine-specific hydrogen bonds with the protein backbone (N3 with the K89 amide and N4 with the L87 carbonyl) and 1 pyrimidine-specific hydrogen bond (O2 with the L90 amide). On one side of the pocket, the cytosine base stacks with the Y108 side chain, and on the other side, it forms a cation–π interaction with K89. A hydrophobic contact of cytosine C5 with F144, which is located in the linker connecting ZnF1-2 and ZnF3-4, completes the pocket. Interestingly, the F144 interaction would clash with any substituent at cytosine C5, precluding ZnF2 binding of a 5-methyl-cytosine nucleotide. The G_3 base is in the purine-only *syn* conformation (Fig. 2*C*) and forms 4 guanine-specific hydrogen bonds with the protein (N1 with Zn ligand C106, N2 with Zn ligand C96, and O6 with the K107 amide and amino groups) and 1 purine-specific hydrogen bond (N7 with the K107 amino). In addition, the guanine base is stacked with Y98. The 5′ nucleotide, U_1, is clearly not part of the recognition site and is more loosely bound at the periphery of the hZAP RBD, where it has hydrophobic contacts with the aliphatic portion of the R74 side chain (Fig. 2*A*). ZAP ZnF2 contacts the RNA backbone only through hydrogen bonds of the Y108 hydroxyl

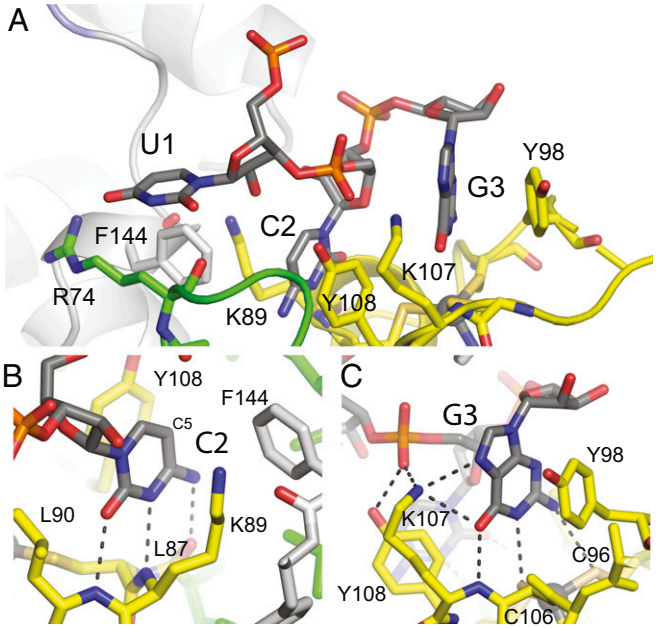

Fig. 2. CG dinucleotide binding to hZAP RBD. (*A*) The RNA C and G bases bound in pockets of ZnF2. (*B*) Detail of cytosine binding, showing C-specific hydrogen bonds to the L87 carbonyl and the K89 amide and a pyrimidine-specific hydrogen bond to the L90 amide (dashed lines). Y108 stacks on the C base. (*C*) Detail of guanine binding. The G base is in the *syn* conformation and forms G-specific hydrogen bonds with 2 Zn ligand cysteine S atoms and with the K107 amide and amino. The K107 amino is also hydrogen-bonded to the purine N7, and the K107 amino and Y108 hydroxyl are hydrogen-bonded to the 5′-phosphate of C. The U base makes no hydrogen-bonding contacts with the protein. Color-coding is as in Fig. 1.

and the K107 amino to the 5′-phosphate of the C_2 nucleotide (Fig. 2*C*).

Structure-Guided Mutagenesis Reveals Amino Acids Responsible for ZAP Antiviral Specificity. To test the biological relevance of the protein RNA contacts observed in the hZAP RBD RNA crystal structure, we designed a series of 10 mutants targeting 5 amino acids (K89, Y98, K107, Y108, and F144) in direct contact with the CG dinucleotide (Fig. 3 *A* and *B*). The mutations were either alanine substitutions or minimal, conservative substitutions (e.g., K to R, Y to F). The mutations were introduced in the context of a tagged hZAP-L–3xHA protein, such that antiviral activity and RNA-binding ability could be determined in mammalian cells.

Each mutation of the 5 amino acids that made contact with the CG dinucleotide in the hZAP RBD RNA complex resulted in hZAP-L–3xHA proteins that were well expressed (*SI Appendix*, Fig. S3). To test whether the mutant proteins could bind RNA, transfected cells were fed 4-thiouridine (4sU), which becomes incorporated into RNA, and then UV irradiated to generate protein-RNA adducts via 4sU cross-linking. Then, cell lysates were treated with RNase, and the hZAP-L–3xHA was immunoprecipitated. Polynucleotide kinase labeling of the immunoprecipitated proteins revealed that each mutant protein had an apparently undiminished ability to cross-link to RNA (*SI Appendix*, Fig. S3).

To assess antiviral activity, we used an overexpression assay in ZAP-deficient 293T cells, in which WT (HIV-1_{WT}) or CG-enriched (HIV-1_{CG}) HIV-1 proviral plasmids were cotransfected with varying amounts of WT and mutant hZAP-L expression plasmids. In these ZAP-deficient cells, the yield of infectious virions from HIV-1_{WT} or HIV-1_{CG} was indistinguishable. Moreover, the yield of infectious HIV-1_{WT} was nearly unaffected by hZAP-L coexpression (Fig. 4*A*). Conversely, the yield of infectious HIV-1_{CG} was progressively decreased as hZAP-L expression was increased, up to ∼20- to 50-fold at the highest levels of hZAP-L (Fig. 4*A*). Concordantly, HIV-1_{WT} Env protein expression was unaltered by hZAP-L expression, while HIV-1_{CG} Env protein expression was reduced to nearly undetectable levels in the presence of hZAP-L (*SI Appendix*, Fig. S4).

Despite their apparently unaltered ability to cross-link RNA, hZAP-L proteins bearing mutations at each of the 5 positions in the CG-binding pocket exhibited reduced antiviral activity against HIV-1_{CG}, enhanced activity against HIV-1_{WT}, or both, as assessed by infectious virus yield (Fig. 4 *A–F* and *SI Appendix*, Table S2) and viral protein expression (*SI Appendix*, Fig. S4) assays. For example, mutations at residue K89, which forms contacts with the C in the CG dinucleotide (K89A and K89R), reduced hZAP-L potency against HIV-1_{CG} and, notably, caused acquisition of some antiviral activity against HIV-1_{WT} (Fig. 4*B* and *SI Appendix*, Table S2 and Fig. S4). Similarly, mutations at another basic amino acid, K107, which forms contacts with the G in the CG dinucleotide (K107A and K107R), had comparable effects (Fig. 4*D* and *SI Appendix*, Table S2 and Fig. S4). Mutations at Y98 had distinct effects depending on the nature of the substitution. Specifically, removal of the stacking interaction with the G nucleotide (in the Y98A mutant) marginally diminished activity against HIV-1_{CG} and caused the acquisition of modest activity against HIV-1_{WT} (Fig. 4*C* and *SI Appendix*, Table S2). The conservative substitution Y98F, which should preserve the stacking interaction, resulted in minimal alteration of hZAP-L activity (*SI Appendix*, Table S2 and Fig. S4). Mutations at 2 further aromatic amino acids, Y108 and F144, which constrain the CG-binding pocket, had more dramatic effects. In particular, mutations at Y108, which is positioned between the C and G nucleotides (Y108A and Y108F), stacks on the C nucleotide, and packs against K107, marginally decreased activity against HIV-1_{CG}

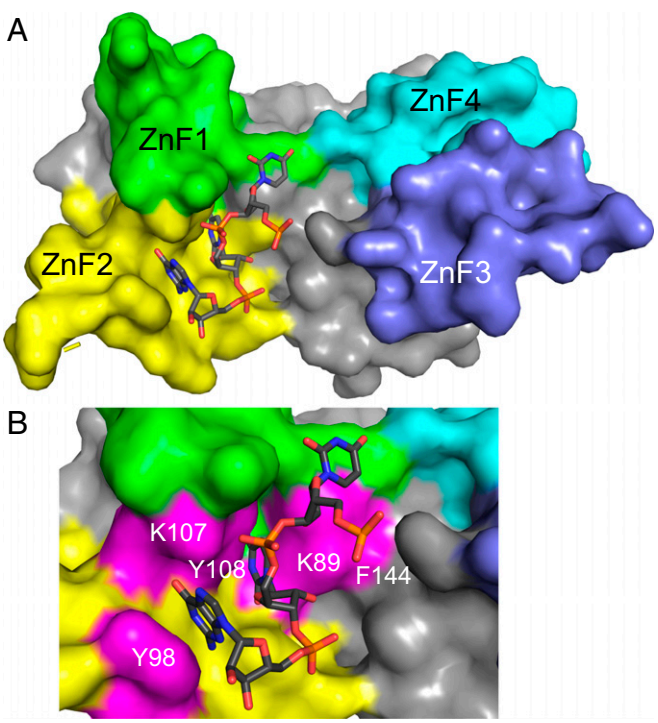

Fig. 3. The hZAP RBD-binding surface for a CG dinucleotide. (*A*) Top surface of hZAP RBD showing C and G bases bound in pockets of ZnF2. Bound ribonucleotide and surfaces are color-coded as in Fig. 1. (*B*) Surface detail of ZnF2, with mutagenized amino acids in magenta.

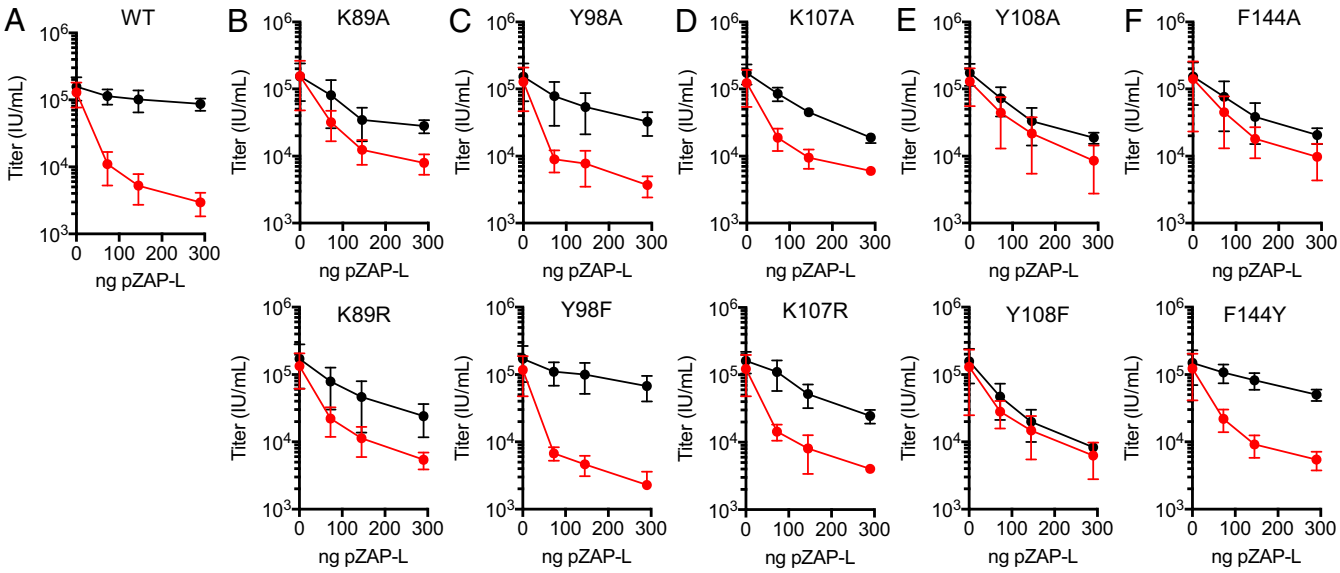

Fig. 4. Antiviral activity of structure-guided hZAP-L mutants targeting the CG-binding pocket. (*A–F*) Infectious virion titer (infectious units/mL) following transfection of ZAP-deficient 293T cell with HIV-1_{WT} (black symbols/lines) or HIV-1_{CG} (red symbols/lines) proviral plasmids and increasing amounts of the indicated WT or mutant hZAP-L expression plasmids (top row, alanine substitutions; bottom row, conservative substitutions).

and also caused acquisition of substantial activity against HIV-1_{WT} (Fig. 4*E* and *SI Appendix*, Table S2 and Fig. S4). Indeed, the Y108A and Y108F mutations caused nearly complete abolition of hZAP-L specificity for HIV-1_{CG} over HIV-1_{WT} (Fig. 4*E* and *SI Appendix*, Table S2). Finally, mutations at F144, which contacts both the C nucleotide in the RNA target and K89, resulted in substantial diminution of specificity for HIV-1_{CG} over HIV-1_{WT} (in the case of F144A) or modest effects in the case of the conservative F144Y substitution (Fig. 4*F* and *SI Appendix*, Table S2 and Fig. S4).

We made additional alanine point mutants at other surface-exposed, positively charged or aromatic amino acids included in the basic patch *SI Appendix*, Fig. S5*A*, based on our structure and the previously described structure of unbound rat ZAP RBD (19). The basic amino acids are positioned in ZnF1, ZnF3, and ZnF4, and some (R74A, R75A, K76A, R79A, K151A, R170A, H176A, F184A, and R189A) were previously reported to affect RNA binding and antiviral activity (19). According to a cross-linking immunoprecipitation polynucleotide kinase assay, none of these point mutations prevented RNA binding in cells (*SI Appendix*, Fig. S5*B*). Moreover, mutations outside the CG-binding pocket were largely devoid of effects on hZAP-L antiviral activity (*SI Appendix*, Fig. S5*C*). The exceptions were R74A, R75A, and particularly R76A, which caused the acquisition of modest antiviral activity against HIV-1_{WT} without diminution of activity against HIV-1_{CG}. Of note, however, R74A, R75A, and particularly R76A, unlike the other mutants in this panel, perturb amino acids immediately proximal to the CG-binding pocket. Indeed, K76 is near the 5′-phosphate of the C nucleotide in our structure. Mutant R179A, which targets the periphery of the basic binding surface, also exhibits somewhat diminished activity against HIV-1_{CG} without acquisition of activity against HIV-1_{WT}. Overall, these mutagenesis experiments show that perturbation of the structurally defined CG-binding pocket resulted in hZAP-L proteins that maintain RNA-binding activity but have altered antiviral specificity.

Antiviral Specificity of hZAP-L Mutants Is Predicted by RNA-Binding Specificity. The loss of ZAP antiviral activity against HIV-1_{CG}, coupled with a gain of activity against HIV-1_{WT}, suggested that some mutations in the CG-binding pocket could cause an alteration in the RNA-binding specificity of the hZAP RBD. To test the RNA-binding specificity of hZAP-L mutants in cells, we conducted CLIP-seq experiments, whereby RNA elements that bound hZAP-L in HIV-1_{CG}–producing cells could be identified by cross-linking and sequencing.

As we described previously (12), hZAP-L bound preferentially to viral RNA elements containing CG dinucleotides, and especially frequently to the HIV-1_{CG} genome segment (nucleotides 5,853 to 6,348) that had been deliberately CG-enriched through the addition of 37 CG dinucleotides (Fig. 5). Conversely, hZAP-L binding to the bulk of the HIV-1 genome, which is highly CG suppressed, was comparatively sparse (Fig. 5). While all of the hZAP-L mutants bound RNA (*SI Appendix*, Fig. S4), there were clear differences in their binding profiles across the HIV-1_{CG} genome compared with WT hZAP-L (Fig. 5 and *SI Appendix*, Fig. S6). Specifically, each of the mutants exhibited diminished binding to the CG-enriched portion to the HIV-1 genome and enhanced binding to the unaltered CG-poor portion of the genome. For some mutants, (e.g., Y98F), this change was minor (*SI Appendix*, Fig. S6), while for others (e.g., Y108A) it was dramatic, with clear acquisition of binding to unaltered CG-poor portions of the genome at the expense of binding to the CG-enriched portion. Indeed hZAP-L(Y108A) exhibited little apparent specificity for binding to CG-enriched sequences in HIV-1_{CG} (Fig. 5).

To complement the cell-based results in terms of direct RNA–protein interaction, we used a fluorescence polarization (FP) assay to determine the affinity of purified hZAP RBD for fluorophore-tagged versions of the 21-nt CG-rich RNA oligonucleotide from HIV-1_{CG} that was used for crystallization and 3 other RNA oligomers with one or no CG dinucleotides (*SI Appendix*, Fig. S7). The hZAP RBD_{WT} had the greatest affinity for the HIV-1_{CG} RNA with 4 CG dinucleotides ($K_d = 0.055$ μM) and 10-fold lower affinity for the RNAs with a single CG dinucleotide (0.45 μM for a 6-nt RNA and 0.57 μM for a 21-nt RNA). The hZAP RBD_{WT} bound weakly to a 23-nt RNA lacking a CG dinucleotide (5.4 μM).

We then purified and tested 4 hZAP RBDs with ZnF2 substitutions that caused reduced antiviral activity (K89A, Y98A, Y108A, and F144A). The Y108A substitution was by far the most deleterious to RNA binding, but all 4 ZnF2 substitutions had substantially

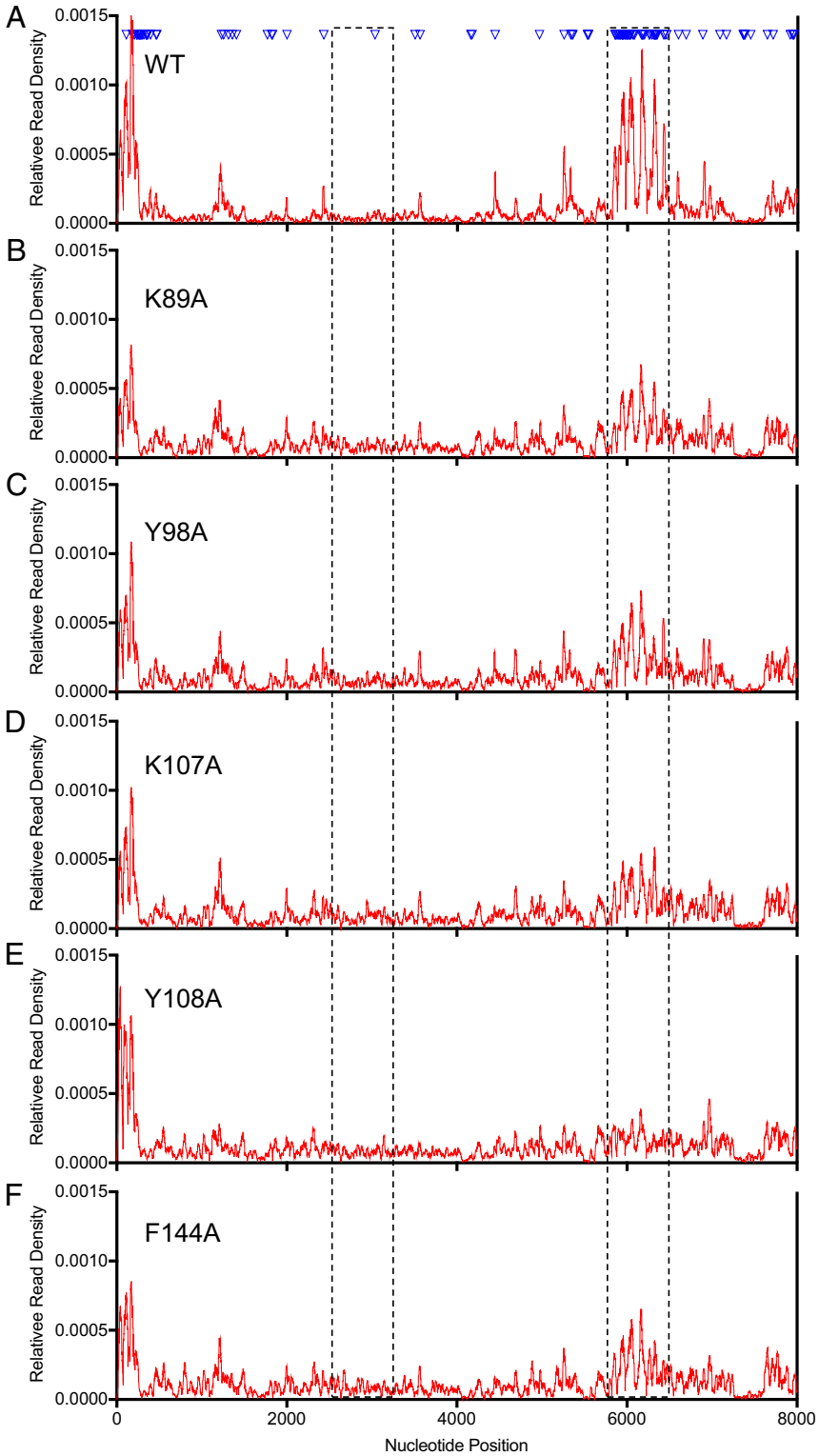

Fig. 5. CLIP-seq analysis of the RNA-binding specificity of hZAP-L mutants. (*A–F*) Reads from CLIP-seq experiments in which HIV-1_{CG} was coexpressed with each of the indicated WT or alanine mutant hZAP-L proteins and mapped to the HIV-1_{CG} genome (represented on the *x*-axis) (27). The fraction of the integrated read density at each nucleotide position is plotted on the *y*-axis. Each CLIP sample contained 4×10^6 to 10×10^6 total reads, of which 0.85% to 2.25% were mapped to the viral genome. Inverted blue triangles indicate the positions of the CG dinucleotides. Dashed lines indicate sequence blocks 2,600 to 3,300 and 5,850 to 6,550 used to assess RNA-binding specificity in Fig. 6*A*.

reduced affinity for the RNA oligonucleotides relative to hZAP RBD_{WT} (*SI Appendix*, Fig. S7). The minimal 6-nt RNA (5′-AUCGAC-3′) is an optimal probe of hZAP RBD's intrinsic affinity for a single CG dinucleotide. The Y108A substitution effectively prevented binding to the hZAP RBD (K_d >100 μM for Y108A, compared with 0.45 μM for WT). Thus, the loss of aromatic

stacking with the cytosine base and a hydrogen bond to the 5′-phosphate in ZnF2 obliterated the CG specificity of the hZAP RBD, demonstrating that ZnF1, ZnF3 ,and ZnF4 do not bind the CG dinucleotide.

To quantify the extent to which changes in RNA-binding specificity among the panel of hZAP-L mutants predicted changes in antiviral specificity, we calculated an antiviral "specificity index" for each hZAP-L mutant. This index was defined as the ratio of infectious virus yields for HIV-1_{WT} vs. HIV-1_{CG} (HIV-1_{WT} titer/HIV-1_{CG} titer) at the highest level of hZAP-L protein expression (Fig. 4). Next, as a measure of RNA-binding specificity, we calculated a ratio of the accumulated CLIP-seq read density in a 700-nt genome segment that included the CG-enriched portion of HIV-1_{CG} (nucleotides 5,850 to 6,550) vs. the cumulative read density in a 700-nt CG-poor genome segment (nucleotides 2,600 to 3,300). This measure of RNA-binding specificity was well correlated with the antiviral specificity index ($P < 0.001$; Fig. 6*A*). As an independent measure of hZAP-L RNA-binding specificity, we assessed the ability of the hZAP-L mutants to selectively cross-link to rare CG-rich RNA elements in human mRNA. Specifically, we calculated the frequency of CG dinucleotides in the 100 most preferred binding sites in cellular mRNAs for each of the hZAP-L mutants. This calculation yielded values in the range of 0.03 to 0.06 CG per nucleotide, and these values again correlated with the antiviral specificity index ($P = 0.012$; Fig. 6*B*). A similar calculation, using a different dinucleotide (UA) that is also underrepresented in mammalian mRNA, revealed no preferential binding to UA dinucleotides by hZAP-L or mutants thereof and no correlation with the antiviral specificity index (*SI Appendix*, Fig. S8). Overall, these data show that the proclivity of the hZAP-L mutant panel to preferentially bind CG-containing RNA elements predicts their ability to specifically reduce HIV-1_{CG} infectious virus yield.

Discussion

Here we present an X-ray crystal structure of the RNA-binding domain of hZAP in complex with an RNA target. The 4 Zn fingers reside on 1 face of the hZAP RBD, where they create a strongly electropositive surface, most prominently in ZnF2. The overall protein structure and details of ZnF2 are identical in hZAP RBD and the previously published structure of rat ZAP RBD (rmsd 0.6 Å) (19); however, the hZAP RBD N-terminal helix does not form a dimer in crystals, as seen in the rat ZAP RBD. More importantly, the RNA complex structure presented herein explains in molecular detail how ZAP distinguishes CG dinucleotides from other sequences. Specifically, substitution of either the C nucleotide or the G nucleotide with any other nucleotide is predicted to lead to steric clashes and would be unable to form the observed hydrogen bonds. Put another way, only a CG dinucleotide can be accommodated in the CG-binding pocket. The ZnF2 subsite for C also excludes 5-methyl-C nucleotides.

The 4 Zn fingers of ZAP exist in 2 clusters (ZnF1-2 and ZnF3-4), with only 1 amino acid separating fingers within a cluster (L87 in ZnF1-2 and I173 in ZnF3-4). The 4 Zn fingers have dissimilar sequences and spacing of ligands (ZnF1: C73-X_4-C-X_3-C-X_3-H86; ZnF2: C88-X_7-C-X_9-C-X_3-H110; ZnF3: C150-X_{11}-C-X_5-C-X_3-H172; ZnF4: C174-X_7-C-X_4-C-X_3-H191). They also have critical differences in structure that preclude CG dinucleotide binding in all but ZnF2. Analogs of the ZnF2 subsites for C and G exist in ZnF1, ZnF3, and ZnF4 with backbone atoms and S ligands that could form hydrogen bonds with the C and G bases, but the relative orientations of these putative subsites are different from those in ZnF2, and the key aromatic amino acid in ZnF2 (Y108) has nonaromatic analogs in ZnF1 (N84), ZnF3 (R170), and ZnF4 (R189). Other subtle differences in the protein backbone and in side chains that would create charge clashes are incompatible with CG dinucleotide binding in ZnF1, ZnF3, and ZnF4.

The hZAP RBD-RNA complex structure is an important addition to the rather small structural database of RNA complexes of CCCH Zn finger proteins (20–23). These proteins have strikingly similar RNA-binding modes, although each recognizes a unique RNA sequence. Common features include deep pockets to bind RNA bases, hydrogen bonds of the Watson–Crick edge of the base to protein backbone atoms and to the S atoms of cysteine Zn ligands, base stacking with aromatic amino acid side chains, and a paucity of contacts with the RNA backbone. Subtle details of each structure specify the target sequence. Use of the protein backbone and Zn ligands for RNA selectivity means that the sequence specificity cannot easily be changed by mutagenesis.

The RNA-binding domain of ZAP includes a large basic surface surrounding the CG dinucleotide-binding pocket. Thus, the overall RNA-binding activity of the hZAP-L RBD consists of a less specific electrostatic component plus a highly specific, nucleotide base (CG) discriminating component. These features are reflected in the FP-binding data for hZAP RBD_{WT}. The electrostatic component likely accounts for the low-affinity binding of the longer RNA oligonucleotides and also explains why mutations that we introduced into the CG-binding pocket attenuated high-affinity binding but did not abolish the ability of hZAP-L to bind RNA. The dominant effect of the CG-binding pocket mutations was to alter hZAP-L RNA binding and antiviral specificity. Indeed, a single atom removal at Y108, whose side chain is positioned between the C and G nucleotides, caused a near-complete loss of specificity for HIV-1_{CG} and a clear gain in antiviral activity against HIV-1_{WT}, consistent with the critical role of Y108 in CG recognition. In fact, the reduction in activity against HIV-1_{CG} among the CG pocket mutants was generally accompanied by the acquisition of activity against HIV-1_{WT}, presumably due to the predominance of nonspecific RNA binding in the mutant context. The amino acids shown herein to be important for CG-specific binding are well conserved among mammalian ZAP proteins. We speculate that this conservation is driven by a requirement to maintain a restricted specificity for CG-rich RNA, as relaxation of this specificity could lead to the deleterious targeting of host RNAs or to "distraction" from viral targets through competition by host RNAs.

Multiple CG dinucleotides in a target RNA are required to confer hZAP sensitivity (12). Moreover, previous findings indicate

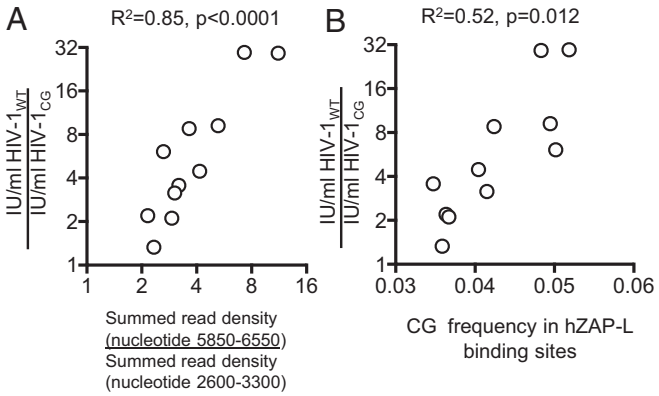

Fig. 6. CLIP-seq–determined RNA-binding specificity predicts the antiviral specificity of hZAP-L mutants. (*A*) Antiviral selectivity of hZAP-L mutants—the ratio of infectious virus yields for HIV-1_{WT} vs. HIV-1_{CG} (HIV-1_{WT}/HIV-1_{CG}) at the highest level of hZAP-L protein expression in Fig. 4—plotted against RNA-binding selectivity—the ratio of the accumulated CLIP-seq read density in a 700-nt genome segment including the CG-enriched portion of HIV-1_{CG} (nucleotides 5,850 to 6,550) vs. the cumulative read density in a CG-poor genome segment (nucleotides 2,600 to 3,300). (*B*) Antiviral selectivity of hZAP-L mutants (as in *A*) plotted against the frequency of CG dinucleotides (CG dinucleotides/total nucleotides) for the 100 most frequently bound sites in cellular mRNAs.

that ZAP acts as a multimer (24). Nevertheless, the structure described herein reveals a monovalent interaction between the hZAP RBD and a CG-containing RNA, despite the presence of 4 CG dinucleotides in the target oligonucleotide. It remains to be determined whether hZAP recognizes native CG-rich viral targets in a polyvalent manner, perhaps aided by the ZAP-binding cofactor TRIM25 (8, 9), which can form high-order multimers (25, 26). In any case, the findings described herein explain in molecular detail how hZAP achieves specific recognition of CG-rich viral RNA.

Methods

Expression and Purification of hZAP RDB and ZAP ZnF2 Mutants. An expression plasmid encoding His_6-hZAP RDB (residues 2 to 227) or mutagenized hZAP-RBD was overexpressed in *Escherichia coli* and purified in 3 steps—Ni affinity, cation exchange, and gel filtration—including proteolytic removal of the His_6 tag, as described in *SI Appendix*.

Crystallization and Structure Determination. Crystallization, data collection, and structure determination of hZAP RBD and hZAP RBD RNA were carried out as described in *SI Appendix*.

In Vitro RNA Transcription. The CG-rich oligoribonucleotide 5′-AUCGACUUCGAUUCGCGGGAA-3′ from HIV-1_{CG} was synthesized by in vitro transcription and purified on a denaturing polyacrylamide gel followed by electroelution, as described in *SI Appendix*.

FP Assay. Saturation FP assays were carried out with 5′-6-FAM (fluorescein) RNA probes (Integrated DNA Technologies) in 384-well plates. The plates were read with a SpectraMax M5 Multimode Plate Reader (Molecular Devices), and the K_d values were calculated using Prism 7 (GraphPad Software). Details are provided in *SI Appendix*.

CLIP-Seq. HA epitope-tagged ZAP-L expressed in 293T cells was cross-linked to RNA by UV irradiation and immunoprecipitated. Protein-RNA adducts were labeled with γ-[^{32}P]ATP, separated by SDS/PAGE, and digested with proteinase K. The remaining RNA oligonucleotides were ligated to adapters, amplified by PCR, and sequenced using the Illumina HiSeq 2000 sequencing system. Details are provided in *SI Appendix*.

Virion Yield Assay. ZAP knockout 293T cells were cotransfected with wild-type or CG-enriched HIV-1 proviral plasmids along with varying amounts of ZAP-L expression plasmids. At 72-h posttransfection, the culture supernatants were harvested, serially diluted, and added to MT4 cells containing a GFP reporter under control the HIV-1 LTR promoter. At 48 h after infection, MT4 cells were fixed and enumerated by FACS analysis to determine infectious viral yield, as described in *SI Appendix*.

ACKNOWLEDGMENTS. We thank members of the P.D.B. and J.L.S. labs for helpful discussions and advice. This work was supported by the Center for HIV RNA studies (Grant U54AI50470, formerly U54GM103297) and NIH Grant R01AI50111 to P.D.B. Use of the LS-CAT in Sector 21 at the Advanced Photon Source (APS) was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). The APS, a US Department of Energy (DOE) Office of Science User Facility, is operated for the DOE Office of Science by Argonne National Laboratory under Contract DE-AC02-06CH11357.

1. T. Kawasaki, T. Kawai, Discrimination between self and non-self nucleic acids by the innate immune system. *Int. Rev. Cell Mol. Biol.* **344**, 1–30 (2019).
2. G. Gao, X. Guo, S. P. Goff, Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* **297**, 1703–1706 (2002).
3. M. J. Bick *et al.*, Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *J. Virol.* **77**, 11555–11562 (2003).
4. S. Müller *et al.*, Inhibition of filovirus replication by the zinc finger antiviral protein. *J. Virol.* **81**, 2391–2400 (2007).
5. R. Mao *et al.*, Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein. *PLoS Pathog.* **9**, e1003494 (2013).
6. X. Guo, J. W. Carroll, M. R. Macdonald, S. P. Goff, G. Gao, The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. *J. Virol.* **78**, 12781–12787 (2004).
7. J. A. Kerns, M. Emerman, H. S. Malik, Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genet.* **4**, e21 (2008).
8. M. M. Li *et al.*, TRIM25 enhances the antiviral action of zinc-finger antiviral protein (ZAP). *PLoS Pathog.* **13**, e1006145 (2017).
9. X. Zheng *et al.*, TRIM25 is required for the antiviral activity of zinc finger antiviral protein. *J. Virol.* **91**, e00088-17 (2017).
10. M. Ficarelli *et al.*, KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *eLife* **8**, e46767 (2019).
11. Y. Zhu *et al.*, Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15834–15839 (2011).
12. M. A. Takata *et al.*, CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* **550**, 124–127 (2017).
13. S. Karlin, J. Mrázek, Compositional differences within and between eukaryotic genomes. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10227–10232 (1997).
14. B. K. Rima, N. V. McFerran, Dinucleotide and stop codon frequencies in single-stranded RNA viruses. *J. Gen. Virol.* **78**, 2859–2870 (1997).
15. B. D. Greenbaum, A. J. Levine, G. Bhanot, R. Rabadan, Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLoS Pathog.* **4**, e1000079 (2008).
16. X. Cheng *et al.*, CpG usage in RNA viruses: Data and hypotheses. *PLoS One* **8**, e74109 (2013).
17. F. van Hemert, A. C. van der Kuyl, B. Berkhout, On the nucleotide composition and structure of retroviral RNA genomes. *Virus Res.* **193**, 16–23 (2014).
18. J. L. Meagher, J. L. Smith, Crystal structure of human zinc finger antiviral protein bound to RNA. Protein Data Bank. https://www.rcsb.org/structure/6UEJ. Deposited 21 September 2019.
19. S. Chen *et al.*, Structure of N-terminal domain of ZAP indicates how a zinc-finger protein recognizes complex RNA. *Nat. Struct. Mol. Biol.* **19**, 430–435 (2012).
20. B. P. Hudson, M. A. Martinez-Yamout, H. J. Dyson, P. E. Wright, Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat. Struct. Mol. Biol.* **11**, 257–264 (2004).
21. M. Teplova, D. J. Patel, Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat. Struct. Mol. Biol.* **15**, 1343–1351 (2008).
22. S. I. Kuhlmann, E. Valkov, M. Stewart, Structural basis for the molecular recognition of polyadenosine RNA by Nab2 Zn fingers. *Nucleic Acids Res.* **42**, 672–680 (2014).
23. J. Murn, M. Teplova, K. Zarnack, Y. Shi, D. J. Patel, Recognition of distinct RNA motifs by the clustered CCCH zinc fingers of neuronal protein Unkempt. *Nat. Struct. Mol. Biol.* **23**, 16–23 (2016).
24. L. M. Law *et al.*, Identification of a dominant negative inhibitor of human zinc finger antiviral protein reveals a functional endogenous pool and critical homotypic interactions. *J. Virol.* **84**, 4504–4512 (2010).
25. J. G. Sanchez *et al.*, The tripartite motif coiled-coil is an elongated antiparallel hairpin dimer. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2494–2499 (2014).
26. J. G. Sanchez *et al.*, Mechanism of TRIM25 catalytic activation in the antiviral RIG-I pathway. *Cell Rep.* **16**, 1315–1325 (2016).
27. M. A. Takata, P. D. Bieniasz, Structure of a zinc-finger antiviral protein in complex with RNA reveals mechanism for selective targeting of CG-rich viral sequences. NCBI Gene Expression Omnibus database. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139667. Deposited 31 October 2019.