

# Structural and functional characterization of the extended-diKH domain from the antiviral endoribonuclease KHNYN

Received for publication, December 23, 2024, and in revised form, February 10, 2025. Published, Papers in Press, February 19, 2025.

<https://doi.org/10.1016/j.jbc.2025.108336>

Rebecca L. Youle<sup>1,2,†</sup>, María José Lista<sup>1,†</sup>, Clement Bouton<sup>1</sup>, Simone Kunzelmann<sup>3</sup>, Harry Wilson<sup>1</sup>, Matthew A. Cottee<sup>2</sup>, Andrew G. Purkiss<sup>3</sup>, Elizabeth R. Morris<sup>2</sup>, Stuart J. D. Neil<sup>1,\*</sup>, Ian A. Taylor<sup>2,\*</sup>, and Chad M. Swanson<sup>1,\*</sup>

From the <sup>1</sup>Department of Infectious Diseases, King's College London, London, United Kingdom; <sup>2</sup>Macromolecular Structure Laboratory, The Francis Crick Institute, London, United Kingdom; and <sup>3</sup>Structural Biology Science Technology Platform, The Francis Crick Institute, London, United Kingdom.

Reviewed by members of the JBC Editorial Board. Edited by Karin Musier-Forsyth

Zinc finger antiviral protein (ZAP) binds CpG dinucleotides in viral RNA and targets them for decay. ZAP interacts with several cofactors to form the ZAP antiviral system, including KHNYN, a multidomain endoribonuclease required for ZAP-mediated RNA decay. However, it is unclear how the individual domains in KHNYN contribute to its activity. Here, we demonstrate that the KHNYN amino-terminal extended-diKH (ex-diKH) domain is required for antiviral activity and present its crystal structure. The structure belongs to a rare group of KH-containing domains, characterized by a noncanonical arrangement between two type 1 KH modules, with an additional helical bundle. N4BP1 is a KHNYN paralog with an ex-diKH domain that functionally complements the KHNYN ex-diKH domain. Interestingly, the ex-diKH domain structure is present in N4BP1-like proteins in lancelets, which are basal chordates, indicating that it is evolutionarily ancient. While many KH domains demonstrate RNA binding activity, biolayer interferometry and electrophoretic mobility shift assays indicate that the KHNYN ex-diKH domain does not bind RNA. Furthermore, residues required for canonical KH domains to bind RNA are not required for KHNYN antiviral activity. By contrast, an inter-KH domain cleft in KHNYN is a potential protein-protein interaction site, and mutations that eliminate arginine salt bridges at the edge of this cleft decrease KHNYN antiviral activity. This suggests that this domain could be a binding site for an unknown KHNYN cofactor.

Zinc finger antiviral protein (ZAP) (also called ZC3HAV1 or PARP13) is an RNA binding protein with antiviral activity against diverse virus families (1–4). While ZAP has multiple mechanisms by which it restricts viral replication, the best

characterized are translation inhibition and RNA decay (1, 3–5). The ZAP RNA binding domain (RBD) binds CpG dinucleotides with high affinity (6–8). When spaced approximately 15 to 30 nucleotides apart in an A/U-rich context, clusters of more than 15 CpGs constitute a ZAP-response element (ZRE) in viral RNA (6, 9). ZAP evolved in tetrapods and is part of a core set of evolutionarily conserved interferon-stimulated genes (10, 11). Many pathogenic human viruses have evolved to evade its activity by encoding protein countermeasures or suppressing CpG abundance in their genomes (3, 4, 6). Synonymous mutations that create CpGs impose a fitness cost for diverse viruses, including HIV-1 (12, 13). CpG abundance in HIV-1 is particularly low, likely due to evolutionary pressure from ZAP antiviral activity, and its abundance in HIV-1 *env* is linked to disease progression (14–17).

KH and NYN containing protein (KHNYN) was recently shown to be a cofactor for ZAP-directed inhibition of retroviral replication (18). KHNYN interacts with ZAP in an RNA-independent manner and KHNYN depletion increases retroviral gene expression and infectious virus production for both murine leukemia virus and HIV-1 that has CpGs introduced *via* synonymous genome recoding to create a ZRE (18, 19). Additionally, KHNYN displays antiviral activity against ZAP-sensitive strains of human cytomegalovirus (20). However, as KHNYN has only recently been identified as a ZAP cofactor and has no other identified function, it remains poorly characterized. KHNYN contains three structural domains, an N-terminal KH domain, a PiT N-terminal (PIN) nuclease domain, and a C-terminal CUE<sub>like</sub> domain (also known as a CUBAN domain) that binds ubiquitin and NEDD8 (18, 21–23). The CUE<sub>like</sub> domain also contains a C terminal nuclear export signal required for KHNYN cytoplasmic localization, its interaction with ZAP and antiviral activity (24). KHNYN has two human paralogs, NYNRIN and N4BP1 (25). NYNRIN has been proposed to regulate trophoblast invasion during placental development while N4BP1 inhibits viral replication and regulates innate immune signaling (26–32). N4BP1 is also an active nuclease that can degrade viral or cellular RNA (27, 33).

<sup>†</sup> These authors contributed equally to this work.

\* For correspondence: Stuart J. D. Neil, [stuart.neil@kcl.ac.uk](mailto:stuart.neil@kcl.ac.uk); Ian A. Taylor, [ian.taylor@crick.ac.uk](mailto:ian.taylor@crick.ac.uk); Chad M. Swanson, [chad.swanson@kcl.ac.uk](mailto:chad.swanson@kcl.ac.uk).

Present addresses for: Matthew A. Cottee, AstraZeneca, The Discovery Centre, 1 Francis Crick Avenue, Cambridge CB2 0AA, United Kingdom; Elizabeth R. Morris, Department of Biosciences, University of Durham, Durham DH1 3LE, United Kingdom.

## Structure and function of the KHNYN extended-diKH domain

We have previously shown that the catalytic residues in the KHNYN PIN domain are required for its antiviral activity and the CUE<sub>like</sub> domain regulates both KHNYN abundance and subcellular localization (18, 24). However, the role of the N-terminal KH domain remains unclear. Initially described as an unconventional KH domain (21), this domain is present only in KHNYN, N4BP1, and NYNRIN orthologs (34). Our previous evolutionary and AlphaFold2 analysis of the KHNYN gene family indicated that the N-terminal KH domain is comprised of two tandem canonical type 1 KH domains (KH1 and KH2), each with a  $\beta$ - $\alpha$ - $\alpha$ - $\beta$  core (35) with an additional extension C terminal to the second KH domain (24).

To determine the functional requirement of the KHNYN KH domains in ZAP-directed antiviral activity, we carried out a combined virological and structural study. Deletion of the KH region resulted in the loss of KHNYN antiviral activity. The crystal structure revealed an orthogonal arrangement of the KH1 and KH2 modules and extended C-terminal alpha-helical bundle not observed in other tandem KH domains, giving rise to the term extended-diKH (ex-diKH) domain. The N4BP1 ex-diKH domain has a similar structure and can replace the KHNYN ex-diKH domain without abrogating antiviral activity. While most KH domains bind short nucleic acid sequences, the GxxG motif that makes core RNA-protein interactions is absent from KHNYN KH1, and our binding studies showed no evidence of RNA binding activity by the ex-diKH. However, mutations in arginine residues that support an interdomain KH1-KH2 cleft resulted in the loss of antiviral activity. Therefore, we hypothesize that the KHNYN ex-diKH domain is a protein-protein interaction site required for KHNYN antiviral function.

## Results

### The KHNYN ex-diKH domain is required for antiviral activity

KHNYN was first annotated in the SUPERFAMILY database to contain a single KH domain (residues 58–141). Deletion of this region decreased antiviral activity against CpG-enriched HIV (HIV-CpG) (18). However, it localized the protein to cytoplasmic puncta not present for wild-type (WT) KHNYN, suggesting that this deletion negatively impacted protein folding, resulting in aggregation. Furthermore, our subsequent AlphaFold2 and evolutionary analysis predicted that the KHNYN N terminal region actually contained a structured region of approximately 200 amino acids, which included two KH modules and a further C-terminal helical bundle comprising an ex-diKH domain (Fig. 1A) (24, 36, 37). The structural homology-based tool Phyre2 (38) also predicted an ex-diKH domain similar to the AlphaFold2 model based on homology to a N4BP1 structure (Protein Data Bank (PDB) 6q3v) that shares 37% identity to the KHNYN domain.

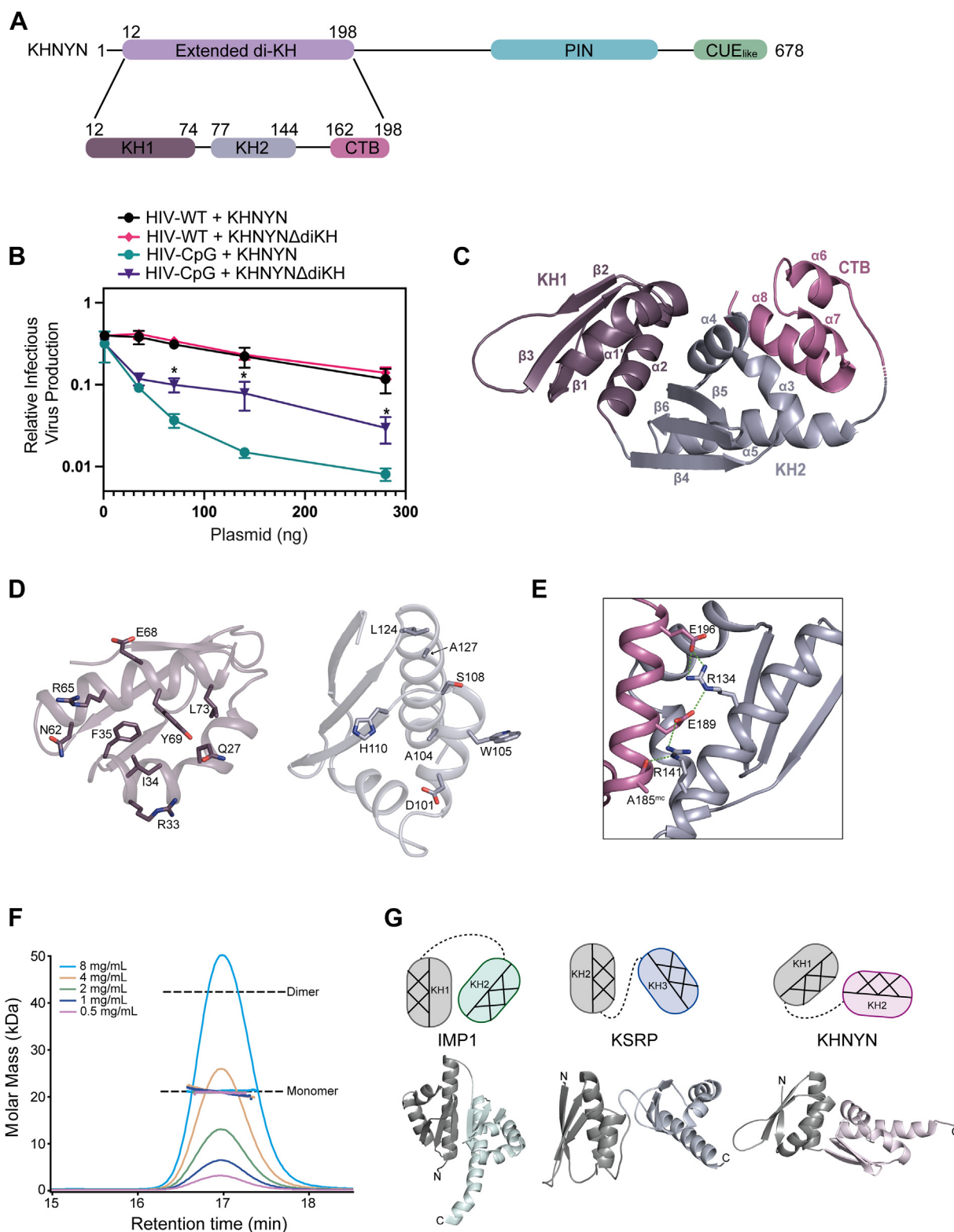
To determine if the ex-diKH domain is required for KHNYN antiviral activity, we deleted it (KHNYN $\Delta$ diKH) and analyzed whether this mutant protein restricted ZAP-sensitive HIV-CpG, which contains a ZRE introduced into *env* through synonymous mutagenesis (6, 18). As a control, we also tested both WT KHNYN and KHNYN $\Delta$ diKH on

WT HIV-1 (HIV-WT), which is ZAP-resistant due to low CpG abundance (6, 18). When KHNYN or KHNYN $\Delta$ diKH were expressed in KHNYN CRISPR cells, KHNYN potently inhibited HIV-CpG while KHNYN $\Delta$ diKH had reduced antiviral activity (Figs. 1B and S1A). This shows that the ex-diKH domain is required for full KHNYN antiviral activity. Importantly, we found that the ex-diKH domain was not required for the interaction of KHNYN with ZAP, KHNYN steady-state subcellular localization or trafficking to the nucleus (Fig. S1, B and C).

### Novel architecture of KHNYN ex-diKH domain

We then determined a crystal structure for the KHNYN ex-diKH domain (Figs. 1C, S1D and Table S1). The asymmetric unit contains two copies of the ex-diKH domain (Fig. S1D). In each copy, the domain spans residues 12 to 198 and is composed of an N-terminal type 1 KH module (KH1, residues 12–74), a short linker (residues 75 and 76), a second type 1 KH module (KH2, residues 77–144), a longer linker (residues 145–161) and a C-terminal bundle (CTB) extension composed of three alpha helices (residues 162–198) (Fig. 1C). The KH2 domain has the expected  $\beta$ 1- $\alpha$ 1- $\alpha$ 2- $\beta$ 2- $\beta$ 3- $\alpha$ 3 type 1 KH topology but in KH1  $\alpha$ 1 and  $\alpha$ 2 are near continuous constituting  $\alpha$ 1' that contains a pronounced proline kink surrounding P29. Inspection of the domain packing within the protein revealed a substantial interdomain interface that packs KH1 onto KH2 (Fig. 1D). This interaction is stabilized through multiple side chain interactions between residues on  $\alpha$ 1 and  $\alpha$ 2 of KH1 and  $\alpha$ 4, the  $\alpha$ 4- $\beta$ 5 linker and  $\alpha$ 5 of KH2. At the center of the interface, these include hydrophobic packing of KH1 side chains I34, F35, Y69, and L73 with A104, H110, L124, and A127 of KH2. These are accompanied by surrounding salt bridges and hydrogen bonding between R33 and D101, R65 and the mainchain carbonyl of S106 and between Y69 and H110, with R33 also  $\pi$ -stacking with W105. Additionally, the CTB extension packs onto the KH2 domain (Fig. 1E) making an interface stabilized by multiple interactions between CTB  $\alpha$ 8 and KH2  $\alpha$ 5, including a salt bridge network between E196 and R134, R134 and E189 and R141, as well as hydrogen bonding between R141 and the carbonyl of the mainchain A185.

While there were two copies of the ex-diKH in the asymmetric unit (Fig. S1D), assessment of the protein-protein interface using PDB-proteins, interfaces, structures, and assemblies interface analysis (39) revealed only minimal contacts with a buried surface area of 393 Å<sup>2</sup>. The computed Complex Formation Significance Score of 0.0 also suggested that the interface is unlikely to have biological relevance. This conclusion is supported by solution size-exclusion chromatography coupled to multiangle laser light scattering (SEC-MALLS) data (Fig. 1F) that revealed that the ex-diKH domain was exclusively monomeric over the range of concentrations tested (0.5–8 mg/ml). Therefore, the dimer observed in the crystal likely is a result of crystal packing rather than any residual assembly associated with an ex-diKH domain homooligomerization function.



**Figure 1. The KHNYN ex-diKH domain is required for antiviral activity and is a member of a novel group of diKH domain structures.** *A*, (upper) schematic representation of KHNYN. The positions of structural domains are highlighted; lilac (ex-diKH, residues 12–198), cyan (PIN), and green CUE<sub>like</sub>. (Lower) schematic of the ex-diKH domain, purple (KH1, residues 12–74), blue gray (KH2, residues 77–144), and magenta (CTB, residues 162–198). *B*, infectious virus production from KHNYN CRISPR HeLa cells cotransfected with pHIV-WT or pHIV-CpG and increasing amount of CRISPR-resistant FLAG-tagged pKHNYN or pKHNYNΔdiKH expressing plasmids. Each point shows the average value of three independent experiments normalized to the value obtained for HIV-WT in the absence of KHNYN. Error bars represent standard deviation. \* indicates that HIV-CpG + KHNYN is significantly different compared to HIV-CpG + KHNYNΔdiKH, unpaired two-tailed *t* test *p* < 0.05. *C*, crystal structure of ex-diKH. The protein backbone is shown as a cartoon representation, domains are colored as in (A), secondary structure elements are labeled from N to Ctermini. *D*, details of the KH1-KH2 interface. Cartoon representations of the protein backbone of KH1 (left) and KH2 (right) are shown, colored as in (C). The view is from one domain looking into the interface of the other. Residues with side chains that make hydrogen bonding and hydrophobic interactions that contribute to interdomain packing (KH1: Q27, R33, I34, F35, N62, R65, E68,



## Structure and function of the KHNYN extended-diKH domain

A KHNYN ex-diKH structural similarity search using the DALI server (40, 41) identified the N4BP1 ex-diKH (PDB 6q3v) as the top overall match with an RMSD (Å) of 1.8 and a Z score of 23.9 over 174 C $\alpha$  positions. In addition, the search also identified matches with single KH domains from a variety of RNA-binding proteins containing multiple KH domains with Z scores ranging from 8 to 9.6 (Table S2). Analysis of the arrangement and relative orientation of the two KH modules in the KHNYN ex-diKH structure in comparison to other diKH structures further allowed classification of three major groups of type 1 diKH arrangements (Fig. 1G). The first and most common interdomain arrangement of type 1 diKH modules observed is the side-by-side antiparallel conformation seen in IMP1 KH1-KH2, Nova-1, and PCBP2 (42–44). In this arrangement, the three  $\beta$ -strands of each KH module align to form an extended sheet made of six  $\beta$ -strands, with stabilization from interactions between the  $\alpha$ 3 helix in each KH module (35). The second group of interdomain arrangement is the orthogonal linker arrangement seen in KSRP KH2 and KH3 modules, with the relative orientation stabilized mainly by domain-linker contacts (45). Comparison with the KH module arrangement observed in KHNYN and N4BP1 now defines a new group of interdomain interaction with an orthogonal packed conformation. This differs from the orthogonal linker group because it has a different absolute rotational orientation and the arrangement is stabilized by the packing between KH modules in the KHNYN and N4BP1 group instead of domain-linker interactions.

N4BP1 is an interferon-stimulated gene (10) that has a similar domain arrangement to KHNYN (Fig. 2A). Moreover, it has previously been shown to regulate ZAP antiviral activity in a genetic screen (46). However, it has not been shown to physically interact with ZAP and is predominantly localized to the nucleus (47). Given the structural similarity of the ex-diKH domains, we determined whether KHNYN and N4BP1 have redundant or complementary activity by siRNA knockdown of N4BP1 in CRISPR control or KHNYN CRISPR KO cells in the absence (Fig. 2B) and presence (Fig. 2C) of type I interferon treatment. While there was no substantial increase in HIV-CpG infectious virus production when N4BP1 was knocked down in control cells, N4BP1 depletion increased HIV-CpG infectious virus to similar levels as HIV-WT in KHNYN CRISPR cells, regardless of interferon induction (Fig. 2, B and C and Fig. S2, A and B). This indicates that N4BP1 has a small amount of antiviral activity on HIV-CpG and there is some functional redundancy between the two proteins.

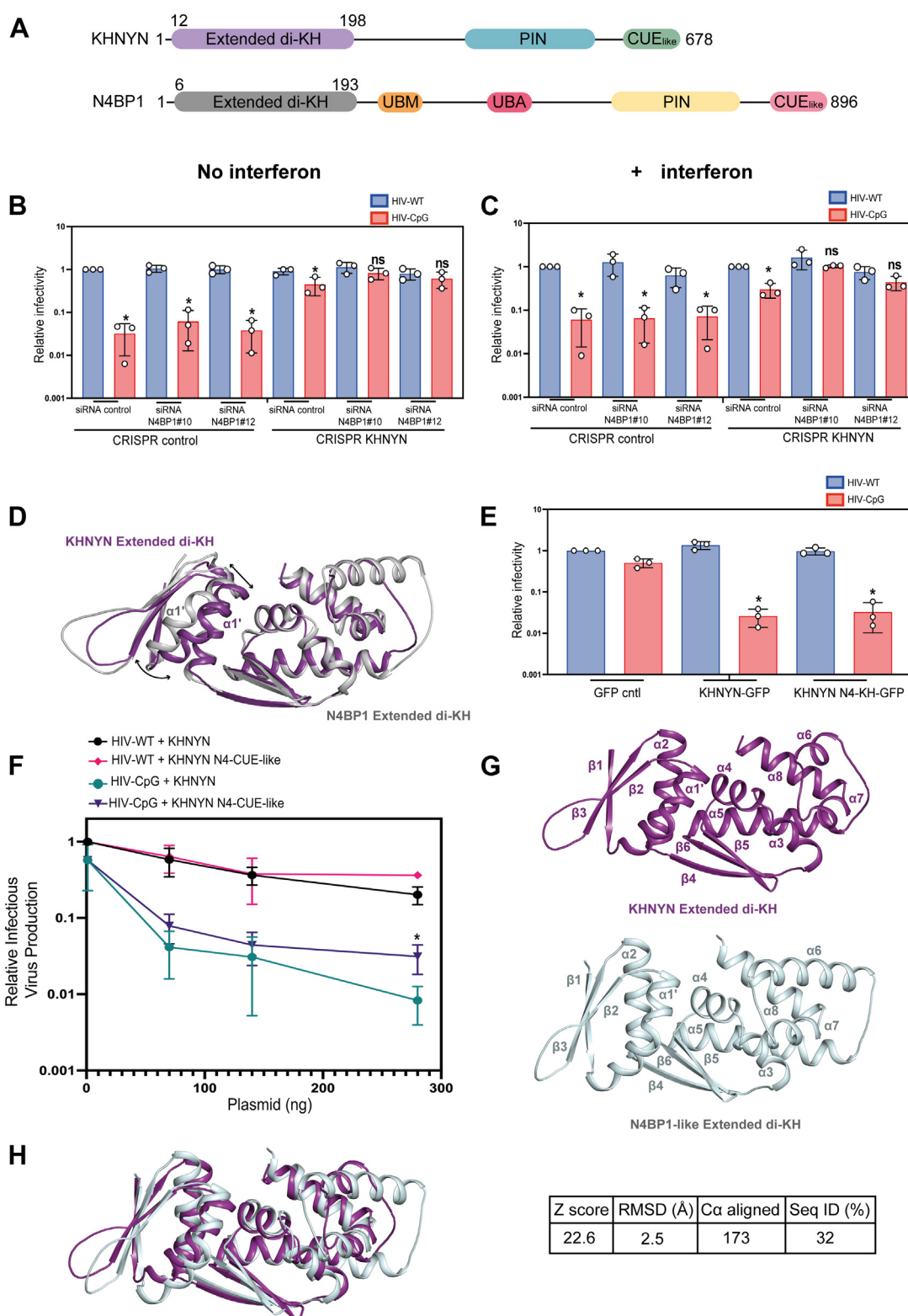
Because KHNYN displays greater antiviral activity than N4BP1, even though it shares the same orthogonal packed diKH domain arrangement as well as the CTB extension, we examined the 3D superimposition of the two ex-diKH structures in detail. This analysis showed that the only substantial difference between the two structures is a small displacement (3–4 Å) in the backbone of the KH1  $\alpha$ 1' helix of KHNYN into the KH1-KH2 interface when compared to N4BP1 (Fig. 2D). Therefore, to determine if there are functional differences between these domains, infectious virus production was measured in stable cell lines expressing WT KHNYN-GFP (24) or expressing KHNYN with the ex-diKH domain replaced with that from N4BP1 (KHNYN N4KH-GFP). These data (Fig. 2E) reveal no significant difference in infectious virus production, indicating that these domains are functionally equivalent. By comparison, our previous characterization of KHNYN had implicated the CUE<sub>like</sub> domain to be required for KHNYN function and also to have diverged evolutionary from the N4BP1 CUE<sub>like</sub> domain (24). Swapping the CUE<sub>like</sub> domains significantly decreased KHNYN activity (Figs. 2F and S2C), suggesting that KHNYN is more antiviral than N4BP1 because the CUE<sub>like</sub> domain contains a nuclear export signal present in the KHNYN tetrapod lineage and not in N4BP1 (24).

We then considered whether the ex-diKH is a recent evolutionary innovation or present in ancient N4BP1-like proteins that predate the gene duplication that led to divergent N4BP1 and KHNYN proteins in bony fish (24). Lancelets (Cephalochordata) are a basal chordate that diverged from other chordates, including vertebrates, over 500 million years ago (48) and encode one N4BP1-like protein (National Center for Biotechnology Information gene IDs: *Branchiostoma belcheri* 109480521; *Branchiostoma floridae* 118417762; and *Branchiostoma lanceolatum* 136438319). Inspection of the AlphaFold model for the *B. belcheri* N4BP1-like protein (AF-A0A6P4ZNE3-F1-v4) showed that an ex-diKH domain was present with a similar overall structure to the KHNYN ex-diKH domain (Fig. 2, G and H). This indicates that the ex-diKH domain is an evolutionary ancient domain that is present in N4BP1-like proteins throughout the chordate lineage.

### The KHNYN ex-diKH domain has no detectable RNA binding activity

Next, we examined the KHNYN ex-diKH domain potential for RNA binding. In structures of nucleic acid bound KH complexes, typically the single-stranded nucleic acid backbone

Y69, and L73) and (KH2: D101, A104, W105, S108, H110, A127, and L124) are shown as sticks, colored by atom type. E, the KH2-CTB interdomain interface. KH2 and the CTB are shown in the cartoon representation, colored as in (C). Residues R134, A185, E189, and E196 that make hydrogen bonds, represented by the green dashed lines, are shown as sticks, mc = mainchain. F, SEC-MALLS analysis of ex-diKH. The sample loading concentrations were 8 mg/ml (cyan), 4 mg/ml (wheat), 2 mg/ml (green), 1 mg/ml (blue), and 0.5 mg/ml (magenta). The differential refractive index (dRI) is plotted against column retention time, and the molar mass determined at 1 s intervals throughout the elution of each peak, is plotted as points. The monomer and dimer molecular mass for ex-diKH are indicated by the dashed lines. G, diKH domain arrangements. diKH domains from IMP1, KSRP, and KHNYN are shown in cartoon colored gray and mint, gray and light blue, and gray and pink, respectively, N and C termini are labeled. The relative orientation of KH domains in each structure is shown above schematically, color coded to the corresponding domain. The helical and  $\beta$ -sheet faces of each KH domain are shown hatched and plain, respectively. SEC-MALLS, size-exclusion chromatography coupled to multiangle laser light scattering; CTB, C-terminal bundle; ex-diKH domain, extended-diKH domain; KHNYN, KH and NYN containing protein; PIN, PiIT N-terminal.



**Figure 2. The KHNYN and N4BP1 ex-diKH domains are functionally equivalent.** *A*, schematic representation of KHNYN and N4BP1. The position of structural/functional domains are highlighted; *lilac* (KHNYN ex-diKH domain residues 12–198), *cyan* (KHNYN PIN domain), *green* (KHNYN CUE<sub>like</sub> domain), *gray* (N4BP1 ex-diKH domain residues 6–193), *orange* (N4BP1 UBM domain), *deep pink* (N4BP1 UBA domain), *yellow* (N4BP1 PIN domain), and *pink* (N4BP1 CUE<sub>like</sub> domain). *B* and *C*, infectious virus production from HeLa CRISPR control or CRISPR KHNYN cells that were either untreated (*B*) or treated with 500 U of IFN- $\beta$  (*C*) and transfected with either siRNA control or siRNAs against N4BP1 and infected with HIV-WT or HIV-CpG at MOI = 3. Each bar shows the average value of three independent experiments normalized to the value obtained for HIV-WT in the CRISPR control cells transfected with siRNA control. Error bars represent standard deviation. \* indicates that HIV-CpG is significantly different compared to HIV-WT, unpaired two-tailed *t* test *p* < 0.05. *D*, 3D DALI structural superimposition of KHNYN and N4BP1 (PDB: 6Q3V) ex-diKH domains. Structures, shown in cartoon representation, KHNYN in *purple* and N4BP1 in *gray*, were aligned over all backbone C $\alpha$  atoms (Z score 23.9, RMSD = 1.8 Å over 174 C $\alpha$ ). *E*, infectious virus production from stable HeLa CRISPR KHNYN GFP control, KHNYN WT-GFP, or KHNYN N4-diKH-GFP infected with either HIV-1 WT or HIV-1 CpG at a MOI = 3. Bar plots represent the average value of three

## Structure and function of the KHNYN extended-diKH domain

binds the GxxG motif in an extended conformation, with the Watson-Crick edge of the nucleic acid oriented toward the central beta sheet of the KH domain (35, 49, 50). This positions the four unpaired nucleic acid bases into a hydrophobic binding cleft where van der Waals, electrostatic interactions, and hydrogen bonding stabilize the binding of nucleic acid (35). KH1 does not contain the GxxG motif usually located between  $\alpha 1$  and  $\alpha 2$  in a KH module (Fig. 3A). Instead, the two helices form a continuous helix  $\alpha 1'$  connected by the proline kink link at P29, suggesting KH1 has lost the capacity to bind RNA. In KH2, an intact GxxG motif (residues 94–97, GAQG) is present and forms the canonical loop connecting the  $\alpha 1$  and  $\alpha 2$  helices ( $\alpha 3$  and  $\alpha 4$  in Fig. 1C). To determine the functional relevance of the KH2 GxxG motif, it was mutated to GDDG, which has been shown to prevent RNA binding in other KH domains (49). Surprisingly, this mutation had no effect on KHNYN antiviral activity (Figs. 3B and S3A) or subcellular localization (Fig. S3B), suggesting that KH2 RNA binding activity is not required for KHNYN antiviral activity or that KH2 does not bind RNA.

To differentiate between these two possibilities, an EMSA was performed using 33mer 5'-fluorescein amidite (FAM)-labeled U-, C-, A-, AU-, CpG-rich, and CpG-depleted RNAs (Table S3). As a control, ZAP RBD bound efficiently to the CpG-rich RNA. However, the KHNYN ex-diKH domain did not bind to any of the RNA sequences, even in the presence of excess protein (100-fold) (Fig. 3C). To further analyze RNA binding, comparative measurements with short RNAs (Table S4) were made using biolayer interferometry (BLI) experiments with KHNYN ex-diKH, ZAP-RBD, and the KH1-KH2 di-domain of IMP1 that scored as a top match in the DALI search (Table S2). These data showed that the ZAP RBD bound a short CpG containing RNA with high affinity ( $K_D = 30$  nM) and to an RNA without a CpG with a 25-fold lower affinity ( $K_D = 740$  nM) (Figs. 3D, S3C and Table S4). By comparison, the KHNYN ex-diKH showed no measurable interaction with either RNAs (Figs. 3D, S3C and Table S4). Further, BLI titrations with the IMP1 KH1-KH2 di-domain and a derivative containing a GDDG mutation in KH2 (44, 49) with CAG, CCG, and CUG containing RNAs showed both IMP1 constructs bound these RNAs with low  $\mu$ M affinities with the KH2 GDDG mutation reducing the binding affinity around twofold (Figs. 3E, S3C and Table S4). Interaction between the KHNYN ex-diKH domain and RNA was again undetectable with any of these RNAs. Overall, no detectable RNA binding activity for the KHNYN ex-diKH domain was observed using a variety of RNA oligonucleotides employing

orthogonal EMSA and BLI assays and at high protein excess. Although not every possible RNA sequence has been analyzed, these data indicate that the KHNYN ex-diKH domain does not bind RNA.

### The KHNYN ex-diKH domain may be a protein binding site

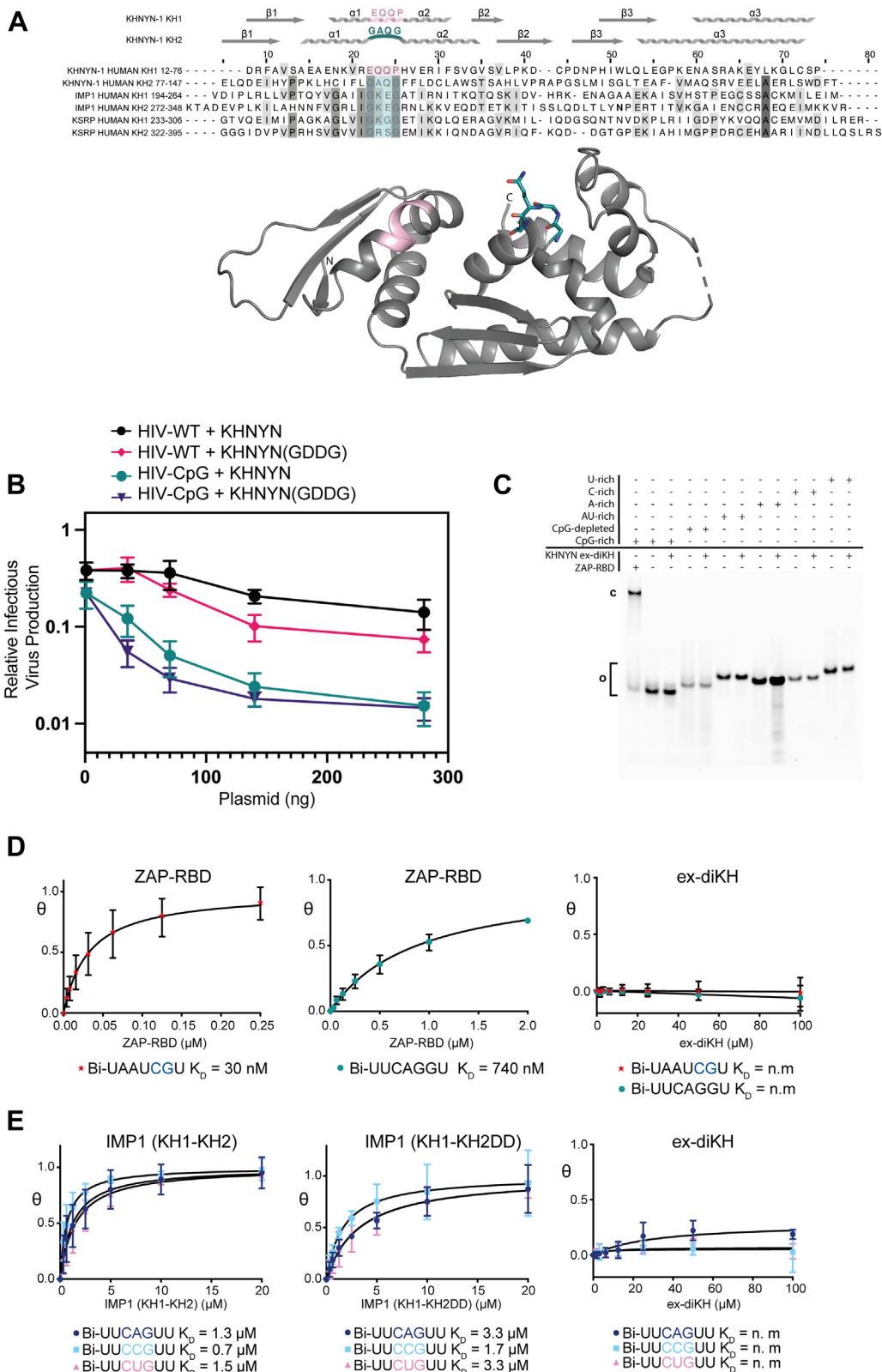
Given the lack of KHNYN ex-diKH domain RNA binding activity, an alternative hypothesis is that it is required for a protein-protein interaction. Further analysis of the inter-residue contacts in the ex-diKH crystal structure identified exposed arginine residues, R33 and R65, that make salt bridges and cation- $\pi$  stacking interactions at the edge of the KH1-KH2 cleft (Fig. 4A) that potentially could form part of a protein binding site. Mutation of these residues to abolish the interactions is likely to reduce the packing at the interdomain cleft, leading to uncoupling of KH1 and KH2 with potential functional consequences for the binding of protein interaction partners. Therefore, to test the requirement of these residues in KHNYN antiviral function, R33 and R65 were substituted to alanine either individually or in combination, and the mutant proteins were tested for antiviral activity (Fig. 4, B and C). While mutating R33 or R65 individually had no effect, and when mutated in combination they substantially reduced KHNYN antiviral activity, even though the proteins were efficiently expressed (Fig. 4D). These data suggest that the hydrogen bonding configuration of R33 and R65 and the exposed surface of the ex-diKH domain around the KH1-KH2 cleft is required for antiviral activity. Given the lack of RNA binding activity, this supports the notion that the ex-diKH domain serves as a protein-protein interaction domain with the KH1-KH2 cleft acting as a protein-binding site.

## Discussion

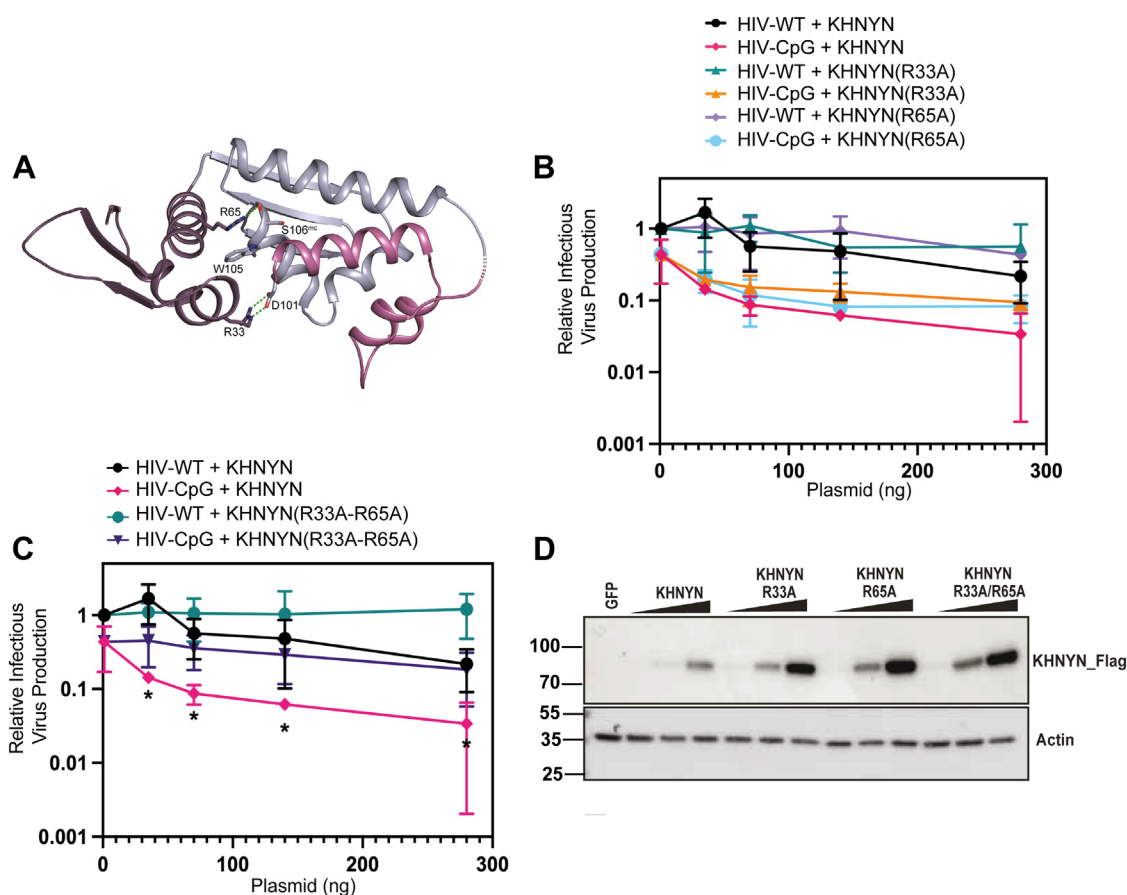
In this study, we have structurally and functionally characterized the KHNYN ex-diKH domain. The structure identifies a new group of type 1 diKH domains, orthogonally packed, and based on the relative orientation of the two KH modules. In addition, the KHNYN ex-diKH domain contains an extension to the KH modules comprising a C-terminal helical bundle that packs into KH2. This diKH arrangement and C terminal extension represent a novel domain architecture observed only in KHNYN and its paralogs and is divergent from previously observed diKH arrangements. Although the KH modules of KHNYN adopt the classic secondary structure fold of type I KH domains, KH1 lacks a GxxG motif, while it is retained in KH2. This elicits structural variation between the KH modules

independent experiments normalized to the value obtained for HIV-WT in the absence of KHNYN. Error bars represent standard deviation. \* indicates that HIV-CpG is significantly different compared to HIV-WT, unpaired two-tailed  $t$  test  $p < 0.05$ . F, infectious virus production from KHNYN CRISPR HeLa cells cotransfected with pHIV-WT or pHIV-CpG and increasing amount of CRISPR-resistant FLAG-tagged pKHNYN or pKHNYN N4-CUE-like expressing plasmids. Each point shows the average value of three independent experiments normalized to the value obtained for HIV-WT in the absence of KHNYN. Error bars represent standard deviation. \* indicates that HIV-CpG + KHNYN is significantly different compared to HIV-CpG + KHNYN N4-CUE-like, unpaired two-tailed  $t$  test  $p < 0.05$ . G, structure of KHNYN ex-diKH domain (top) and AlphaFold model of the ex-diKH domain from *Branchiostoma belcheri* NEDD4-binding protein 1-like (NBP1-like) (bottom). Protein backbones are shown in cartoon representations, secondary structure elements are labeled from N to C termini. H, 3D DALI structural superimposition of KHNYN and N4BP1-like ex-diKH domains. Protein backbones are shown in cartoon representation, KHNYN in purple and N4BP1-like in pale cyan. Structures were aligned over all backbone C $\alpha$  atoms. Alignment data are shown on the right (Z score = 22.6, rmsd = 2.5 Å over 173 C $\alpha$ , 32% Seq ID). ex-diKH domain, extended-diKH domain; IFN- $\beta$ , interferon- $\beta$ ; KHNYN, KH and NYN containing protein; MOI, multiplicity of infection; PDB, Protein Data Bank; PIN, PiIT N-terminal.





## Structure and function of the KHNYN extended-diKH domain



**Figure 4. The KHNYN KH1-KH2 interdomain cleft is required for antiviral activity.** *A*, the KHNYN ex-diKH domain is shown in the cartoon representation. KH1, KH2, and CTB are colored purple, blue gray, and magenta respectively. Residues R33 and R65 together with D101, W105, and S106 that make salt bridge, hydrogen bond, and cation- $\pi$  interactions across the KH1-KH2 interdomain cleft are shown in sticks. Salt bridge and hydrogen bonds are represented by the green dashes, mc = mainchain. *B* and *C*, infectious virus production from KHNYN CRISPR HeLa cells cotransfected with pHIV-WT or pHIV-CpG and (B) increasing amount of CRISPR-resistant FLAG-tagged pKHNYN, pKHNYN (R33A), or pKHNYN (R65A) and (C) pKHNYN or pKHNYN(R33A/R65A) expressing plasmids. Each point shows the average value of three independent experiments normalized to the value obtained for pHIV-WT in the absence of KHNYN. HIV-WT + KHNYN and HIV-CpG + KHNYN are the same data for panel B and C. \* indicates that HIV-CpG + KHNYN is significantly different compared to HIV-CpG + KHNYN(R33A/R65A), unpaired two-tailed  $t$  test  $p < 0.05$ . *D*, representative Western blotting corresponding to the expression level of FLAG-tagged KHNYN KHNYN(R33A), KHNYN(R65A), or KHNYN(R33A-R65A) employed in the infection experiments is shown in the panels B and C. CTB, C-terminal bundle; ex-diKH domain, extended-diKH domain; KHNYN, KH and NYN containing protein.

in KHNYN, with the GxxG loop missing in KH1 replaced with a proline kink, indicating further structural divergence from classical KH domains. Notably, we detect no RNA binding activity, suggesting that the orthogonal packed diKH family may have evolved to function outside of the usual KH RNA binding activity.

While this article was in the final stages of preparation and under review, two studies were published that support its conclusions (51, 52). Similar to our findings, these articles show that N4BP1 and KHNYN have partially redundant activity (51), the KHNYN ex-diKH domain is required for antiviral activity (51) and the KHNYN ex-diKH domain does not bind RNA (52). They also report a similar structure for the

KHNYN ex-diKH domain (52). Nevertheless, the function of the ex-diKH domain in the context of KHNYN, N4BP1, or NYNRIN remains unknown. While both KHNYN and N4BP1 are active nucleases (23, 27, 33, 52), the full range of viral and cellular RNA targets for these proteins remains unclear, though N4BP1 has been reported to target specific cellular transcripts (33). Whether KHNYN and N4BP1 target the same or different cellular transcripts is also unknown. N4BP1 is predominantly nuclear (47) and KHNYN is predominantly cytoplasmic (18) so they may target transcripts in different compartments. As both KHNYN and N4BP1 ex-diKH domains have tandem KH domains, which often bind RNA (35), a potential hypothesis is that the ex-diKH domain modulates

C-rich, A-rich, AU-rich, CpG-depleted, and CpG-rich 33mer 5'-FAM-labeled RNAs (0.2  $\mu$ M) were incubated with ZAP RBD (2  $\mu$ M) and ex-diKH (20  $\mu$ M). Free (o) and bound (c) components were separated on a 10% acrylamide native TBE gel and visualized by fluorescence imaging. *D* and *E*, BLI analysis for (D) KHNYN ex-diKH and ZAP-RBD binding to CpG (Bi-UAAUCGU) and non-CpG (Bi-UUCAGGU) RNAs and (E) KHNYN ex-diKH, IMP1(KH1-KH2) and IMP1(KH1-KH2DD) RNA-binding to 5'-biotinylated CNG RNAs. Nonlinear fits of fraction bound plotted against protein concentration are shown. Data are the mean of triplicate measurements. Error bars represent the standard deviation of the mean for each triplicate measurement. The derived  $K_D$  values are shown beneath plots and presented with 95% CI values in Table S4. BLI signal response for ex-diKH interaction with CpG, non-CpG, and CNG RNAs is unobservable. BLI, biolayer interferometry; CI, confidence interval; ex-diKH domain, extended-diKH domain; FAM, fluorescein amidite; KHNYN, KH and NYN containing protein; RBD, RNA binding domain; TBE, tris-borate-EDTA; ZAP, zinc finger antiviral protein.



substrate specificity for targeted RNA decay. Nonetheless, this hypothesis does not appear correct as neither we nor the others (52) have been able to detect RNA binding activity for the KHNYN ex-diKH domain, though it is formally possible that it binds an RNA sequence that has not been tested. Therefore, the CpG binding pocket in the ZAP RBD (7, 8) likely contributes much of the targeting specificity for the ZAP-KHNYN complex, which correlates with the observation that introducing CpGs by synonymous genome recoding sensitizes several viruses to ZAP-mediated restriction (6, 53–55). It should be noted that other domains in KHNYN, such as the PIN domain, may specifically bind RNA (52, 56, 57). However, a complex of purified ZAP RBD and KHNYN PIN-CUE-like domains does not specifically degrade CpG-rich RNA (52), indicating that other components of the ZAP antiviral complex contribute to specificity.

The ex-diKH domain also does not regulate KHNYN subcellular localization and is not required for the interaction between KHNYN and ZAP. Moreover, it does not mediate homomultimerization. Inspection of the crystal structure highlighted a potential protein binding site at the KH1-KH2 interface, and mutations that disrupt salt bridges within this site substantially decreased the KHNYN antiviral activity. Therefore, we hypothesize that the KHNYN ex-diKH domain may act as a protein-protein interaction unit that binds a currently unknown cofactor in the ZAP antiviral system. Other KH domains have also been shown to mediate protein-protein interactions, supporting this hypothesis (58–60). Identifying binding partners for KHNYN may identify novel cofactors for ZAP-mediated RNA decay. Furthermore, since the N4BP1 ex-diKH domain functionally substitutes for the KHNYN ex-diKH domain, such interactions are likely conserved and could represent an ancient function predating the evolution of ZAP.

## Experimental procedures

### Plasmids and cell lines

HeLa, HEK293T, and TZM-bl cells were maintained in high glucose Dulbecco's modified Eagle's medium supplemented with GlutaMAX (Thermo Fisher Scientific), 10% fetal bovine serum, 100 U/ml penicillin, and 100 µg/ml streptomycin and incubated with 5% CO<sub>2</sub> at 37 °C. HeLa and HEK293T cells were a kind gift from Michael Malim. The TZM cells were obtained from the National Institutes of Health AIDS Research and Reference Reagent Program, Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health. Control CRISPR and KHNYN CRISPR HeLa cells were previously described (18). The CRISPR-resistant pKHNYN-FLAG plasmid has been previously described (18). Specific mutations were created using site-specific mutagenesis. The KHNYN deletion mutants were generated by site-directed mutagenesis. The KHNYN domain swap constructs were generated by HiFi assembly (NEB) using amplicons obtained using Q5 DNA polymerase (NEB). The generated plasmids were verified by Sanger sequencing. All primers are shown in Table S5, and all PCR reactions were performed

using Q5 High-Fidelity polymerase (NEB). HIV-1<sub>NL4-3</sub> (pHIV-WT) and HIV<sub>env86-561CpG</sub> (pHIV-CpG) in pGL4 were previously described (18, 61). Stable CRISPR KHNYN HeLa cells expressing KHNYN $\Delta$ ex-diKH-GFP and KHNYN-N4KH-GFP were produced by transduction with the murine leukemia virus retroviral vector MIGR1 (62) with the KHNYN-GFP constructs cloned into the multiple cloning site and GFP replaced by the Blasticidin S-resistance gene.

### Transfections and infections

HeLa cells were seeded in 24-well plates at 70% confluency. Cells were transfected according to the manufacturer's instructions using TransIT-LT2020 (Mirus) at the ratio of 3 µl TransIT-LT2020 to 1 µg DNA. Subsequently, 0.5 µg pHIV-WT or pHIV-CpG and the designated amount of KHNYN-FLAG or GFP-FLAG for a total of 1 µg DNA were transfected. Twenty-four hours post transfection, the culture media were replaced with fresh media. Supernatants and cells were then harvested for analysis after 48 h post-transfection. For infection experiments, viral stocks were produced by cotransfecting pHIV-WT or pHIV-CpG with pVSV-G (63) into HEK293T ZAP CRISPR cells (64) and titered on TZM-bl cells (65–67). HeLa cells were plated for transfection experiments and infected either with HIV-WT or HIV-CpG at multiplicity of infection = 3. Media were replaced 24 h after infection. Supernatants and cells were harvested 48 h post infection.

### TZM-bl infectivity assay

The TZM-bl indicator cell line was used to quantify the amount of infectious virus (65–67). Briefly, cells were seeded in 96-well plates and infected by incubation with virus stocks or supernatants of infected cells. Subsequently, 48 h post infection, the cells were lysed and infectivity was measured by  $\beta$ -galactosidase expression using the Galacto-Star System following the manufacturer's instructions (Applied Biosystems).  $\beta$ -galactosidase activity was quantified as relative light units per second using a PerkinElmer Luminometer.

### Analysis of protein expression by immunoblotting

Forty-eight hours post transfection, HeLa cells were lysed in Laemmli buffer at 95 °C for 10 min. The culture supernatant was also collected, and virions pelleted by centrifugation for 2 h at 20,000g through a 20% sucrose cushion in PBS. Viral pellets were resuspended in 2x Laemmli buffer. Cell and virion lysates were resolved on 8–16% Mini-PROTEAN TGX precast protein gels (Bio-Rad), transferred into nitrocellulose membranes (GE HealthCare) that were then blocked in 5% nonfat milk in PBS supplemented with 0.1% Tween 20. Primary antibodies were incubated overnight at 4 °C followed by three washes with PBS and then corresponding secondary antibody (1:5000 anti-mouse IRDye 680RD (LI-COR, 926–68070), 1:5000 anti-rabbit IRDye 800CW (LI-COR, 926–32211), 1:5000 anti-rabbit horseradish peroxidase (HRP) (Cell Signaling Technology, 7074) or 1:5000 anti-mouse HRP (Cell Signaling Technology, 7076)) was incubated for 1 h. Proteins

## Structure and function of the KHNYN extended-diKH domain

were visualized by LI-COR (Odyssey Fc) measuring secondary antibody fluorescence or using Amersham ECL Prime Western Blotting Detection reagent for HRP-linked antibodies with an ImageQuant (LAS8000 Mini). The following primary antibodies were used at dilutions of: 1:50 HIV anti-p24Gag (mouse hybridoma) (68), 1:3000 anti-HIV gp160/120 (Rabbit, ADP421; Centralized Facility for AIDS Reagents, 1:5000 anti-HSP90 (Rabbit, GeneTex, GTX109753), 1:1000 anti-FLAG (DYKDDDDK, Rabbit, Cell Signaling, 14793), 1:5000 anti-ZAP (Rabbit, Abcam, ab154680), 1:1000 anti-N4BP1 (antibodies-online GmbH, ABIN2706635), 1:2000 anti- $\beta$ -actin (Mouse, Abcam; ab6276), 1:10000 anti-GAPDH (Rabbit, Proteintech, 10494-1-AP), 1:10000 anti-GAPDH (mouse, Proteintech, 60004-1-Ig), 1:1000 anti-GFP (Mouse, Roche 11814460001). The anti-ZAP and anti-N4BP1 antibodies were validated using cells in which the protein has been depleted. The anti-HIV p24Gag and anti-HIV gp160/120 antibodies were validated by comparing untransfected to HIV transfected cells (64). The anti-FLAG antibody has been validated by comparing cells expressing GFP-FLAG or KHNYN-FLAG (18).

### Immunoprecipitation assays

HeLa cells stably expressing WT KHNYN-GFP or mutant proteins were seeded in 6-well plates 24 h prior to immunoprecipitation. The cells were lysed on ice in 500  $\mu$ l of lysis buffer (0.5% NP-40, 150 mM KCl, 10 mM Hepes pH 7.5, and 3 mM MgCl<sub>2</sub>) supplemented with cComplete Protease Inhibitor Cocktail tablets (Roche 11873580001). The lysates were incubated on ice for 1 h and centrifugated at 20,000g for 15 min at 4 °C to pellet the nuclear fraction. Subsequently, 50  $\mu$ l of the post nuclear supernatants was saved as the input lysate and 450  $\mu$ l of the supernatants was incubated with 5  $\mu$ g of anti-GFP antibody (Roche 11814460001) for 1 h at 4 °C. Protein G Dynabeads (Invitrogen) were then added and incubated overnight at 4 °C with rotation. The lysates were washed four times with wash buffer (0.05% NP-40, 150 mM KCl, 10 mM Hepes pH 7.5, and 3 mM MgCl<sub>2</sub>) before the bound proteins were eluted with 2x Laemmli buffer and boiled for 10 min. Protein expression was analyzed by Western blotting as described above.

### Microscopy

HeLa cells were seeded in 24-well plates on coverslips pretreated with polylysine. KHNYN CRISPR HeLa cells were transfected with 250 ng of KHNYN-FLAG or KHNYN(GDDG)-FLAG. Briefly, 24 h post transfection, the cells were fixed with 4% paraformaldehyde for 20 min at room temperature, washed once with 1xPBS and once in 10 mM glycine. Cells were then permeabilized for 15 min in 1% bovine serum albumin and 0.1% Triton-X in PBS. HeLa cells stably expressing WT KHNYN-GFP or versions with specific mutations were seeded in pretreated 24-well plates 24 h prior to immunostaining and fixed and permeabilized as above. Mouse anti-FLAG (1:500) and rabbit anti-ZAP (1:500) antibodies were diluted in PBS/0.01% Triton-X and the cells were stained for

2 h at room temperature. The anti-FLAG antibody was validated on untransfected cells (18). The cells were then washed three times in PBS/0.01% Triton-X and incubated with Alexa Fluor 594 anti-mouse or anti-rabbit or Alexa Fluor 647 anti-rabbit (Molecular Probes, 1:500 in PBS/0.01% Triton-X) for 45 min in the dark. Finally, the coverslips were washed three times with 1xPBS/0.01% Triton-X and then mounted on slides using Prolong Diamond Antifade Mountant with 4',6-diamidino-2-phenylindol (Invitrogen). Imaging was performed on a Nikon Eclipse Ti Inverted Microscope, equipped with a Yokogawa CSU/X1-spinning disk unit, under 60 $\times$  objective and laser wavelengths of 405 nm, 488 nm, 561 nm, and 640 nm. Image processing and colocalization analysis was performed with ImageJ (Fiji) software (<https://imagej.net/software/fiji/>).

### Leptomycin B treatment

HeLa cells stably expressing WT KHNYN-GFP or mutant versions were seeded in pretreated 24-well plates 24 h prior to a 4-h treatment with 50 nM of Leptomycin B (Sigma-Aldrich, L2913) or dimethyl sulfoxide at 37 °C. After treatment, the cells were fixed and immunostained as described above.

### Protein expression and purification

DNA sequences encoding human KHNYN residues P8-G200 and human ZAP residues M1-N227 were PCR amplified and inserted into a pET-52b(+) expression vector (Novagen), using XmaI/SacI restriction sites (KHNYN) or ligation independent cloning (ZAP), to produce N terminally StrepII-tagged fusion proteins. Primer sequences are listed in Table S5. Insert sequences were confirmed by DNA sequencing (Eurofins). StrepII-tagged fusion proteins were expressed in *Escherichia coli* strain BL21(DE3) (NEB) incubated at 37 °C with shaking. Addition of 0.4 mM IPTG was used to induce protein expression at an A<sub>600</sub> of ~0.6, followed by 18 h incubation at 18 °C. Cells were harvested by centrifugation, resuspended in 5 ml/g of lysis buffer K (KHNYN (8–200)) or lysis buffer Z (ZAP (1–227)) and lysed by sonication (KHNYN(8–200)) or emulsiflex homogenization (ZAP (1–227)). The composition of protein purification buffers is presented in Table S6. Lysate was clarified by centrifugation for 1 h at 50,000g at 4 °C before the supernatant was applied to 10 ml of packed Strep-Tactin XT Superflow resin (IBA), followed by washing with 50 ml of wash K1 buffer followed by 50 ml wash K2 buffer (KHNYN-1 (8–200)) or 200 ml of wash Z1 buffer (ZAP (1–227)) at 4 °C. Bound protein was then step eluted with elution buffer KE (KHNYN (8–200)) or ZE (ZAP (1–227)) and highest purity fractions, identified by SDS PAGE, pooled.

For KHNYN (8–200), GST-tagged HRV 3C-protease (0.5 mg) was added to eluent overnight to remove the StrepII-tag. Following concentration of cleaved eluent, GST-tagged HRV 3C-protease was removed by affinity chromatography using a 1 ml GSTrap HP column (Cytiva) before sample application to a Superdex 75 26/600 column (Cytiva) equilibrated in KGF gel filtration buffer. Peak fractions were

identified by SDS PAGE, concentrated to ~30 mg/ml and flash frozen in liquid nitrogen.

For ZAP (1–227), GST-tagged HRV 3C-protease (0.5 mg) was added to the eluent and then dialyzed overnight against low salt buffer. The dialysate was then diluted rapidly to 125 mM NaCl before application to a 1 ml HiTrap Q HP column (GE HealthCare). The flow through was collected along with 10 ml of wash with ion exchange buffer. This sample was concentrated and any remaining GST-tagged HRV 3C-protease was removed by affinity chromatography using a 1 ml GStrap HP column (GE HealthCare) before sample application to a Superdex 75 16/600 column (GE HealthCare) equilibrated in ZAP gel filtration buffer. Peak fractions identified by SDS PAGE were then concentrated to ~2.5 mg/ml and flash frozen in liquid nitrogen. IMP1 proteins were purified as described (44) and kindly provided by Dr Robert Dagil, University of Copenhagen.

### SEC-MALLS

SEC-MALLS was performed to determine the molar mass of KHNYN(8–200) samples. In addition, 100 µl samples ranging from 0.5 to 8mg/ml were applied to a Superdex 200 10/300 Increase GL column (Cytiva) equilibrated in 50 mM Bis-tris propane pH 8.5, 150 mM NaCl, 0.5 mM tris(2-carboxyethyl) phosphine (TCEP), and 3 mM NaN<sub>3</sub> at a flow rate of 1.0 ml/min. The scattered light intensity and protein concentration of the column eluate were recorded using a DAWN-HELEOS-II laser photometer (Wyatt Technology Corp) and an OPTILAB-TrEX differential refractometer (Wyatt Technology Corp) (differential refractive index, dRI) ( $dn/dc = 0.186$ ), respectively. The weight-averaged molecular mass of the protein was then determined using the ASTRA software version 7.3.2 (Wyatt Technology Corp).

### Crystallization and structure determination

KHNYN (8–200) crystals were produced by vapor drop diffusion at 18 °C using protein at stock concentration (~30 mg/ml) in Swissci MRC two-drop 96-well plates (Molecular Dimensions), with a 75 µl reservoir of screen condition (100 mM Tris pH 7.0, 250 mM NaCl, and 35% w/v PEG 3K). A Mosquito LCP robot (TTP Labtech) was utilized to set 100 nl drops, with a 2:1 ratio of protein solution to screen condition. Crystals were harvested by looping and flash-frozen in liquid nitrogen before data collection on the i04 beamline at Diamond Light Source at a wavelength of 0.9795 Å.

An initial dataset was processed using the automatic xia2 dials Diamond pipeline (69, 70), and the structure solved by molecular replacement using Phaser ([https://phenix-online.org/documentation/reference/phaser\\_mr.html](https://phenix-online.org/documentation/reference/phaser_mr.html)) (71) employing an AlphaFold2 (<https://alphafold.ebi.ac.uk/>) model (36) of KHNYN residues 12 to 142. After initial rebuilding in COOT 0.9.6 (<https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/>) (72) and iterative rounds of refinement in Phenix refine (<https://phenix-online.org/documentation/reference/refinement.html>) (73), a higher quality and more complete dataset was obtained and processed with the autoPROC

pipeline (74) using XDS (75), POINTLESS (76) and AIMLESS (77) and anisotropic scaling applied using STARANISO (<http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>) and inheriting the FreeR flags from the initial dataset. The final model comprises two copies of KHNYN(8–200) with all residues built from 12 to 198 except for a flexible loop between residues 148 to 159 and was refined to a  $R_{work}/R_{free}$  of 0.196/0.258. MolProbity (78) and PDB\_REDO (79) were used to monitor and assess model geometry throughout the refinement process. Details of crystal parameters, data collection, and refinement statistics are shown in Table S1. The coordinates and structure factors have been deposited in the PDB under the accession code 9HTS.

### Structural and sequence alignments

Protein domain prediction was performed with AlphaFold2 (36) and Phyre2 (38). PDB files from crystal structures were viewed and rendered in PyMOL 2.5.5 (Schrodinger, LLC; <http://www.pymol.org/>). Electron density maps were visualized in COOT 0.9.6 (72). Analysis of protein interfaces within crystal structures was performed with PDB-PISA (39). The DALI server (40) was used to perform PDB25/PDB90 searches using chain A of the KHNYN (8–200) crystal structure. Primary sequences were aligned in Jalview (<http://www.jalview.org/>) (80) using Clustal default settings. Secondary structure was annotated based on the KHNYN (8–200) crystal structure.

### siRNA transfections and IFN treatment

Dharmacon ON-TARGETplus nontargeting control siRNA (D-001810-10) and siRNAs targeting N4BP1 (#10: J-020733-10-0002 and #12: J-020733-12-0002) were purchased from Horizon Discovery Biosciences Limited. The target sequences for the N4BP1 siRNAs are #10: GUUCAAGGUUUCGA-GAUA and #12: GAACCAGGGAGAACGGAUU. HeLa CRISPR control or CRISPR KHNYN cells were reverse-transfected with either siRNA control or siN4BP1 as follows: 1 µl of each siRNA at 20 nM with 1 µl of Lipofectamine RNA-iMax (Invitrogen) following manufacturing instructions. Briefly, 24 h after the first transfection, the media were replaced, and cells were transfected a second time, directly, using the same conditions. Forty-eight hours after the first transfection, cells were infected with either HIV-WT or HIV-CpG at multiplicity of infection = 3. Media were replaced 24 h post infection. Supernatants and cells were then harvested 48 h post infection and analyzed by TZM-bl assay and by immunoblotting. For the interferon-β (IFN-β) experiments, cells were treated with 500 U of IFN-β at 24 h post infection and harvested 24 h later.

### RNA

HPLC purified RNA oligonucleotides were synthesized by Horizon Discovery Biosciences Limited. RNA was deprotected according to the manufacturer's protocol, lyophilized, and then resuspended in TE buffer (10 mM Tris pH 8.0 and 0.1 mM ethylenediaminetetra-acetic acid [EDTA]). Oligonucleotide concentrations were determined from the  $A_{260}$  nm



## Structure and function of the KHNYN extended-diKH domain

derived from the UV/visible absorbance spectrum using extinction coefficients based on base sequence composition provided by the manufacturer.

### Electrophoretic mobility shift assays

To test ex-diKH RNA-binding, 10  $\mu$ l reactions containing 0.2  $\mu$ M of 33-mer 5'-FAM-labeled RNAs (Table S3) were incubated with 20  $\mu$ M KHNYN (8–200) in EMSA reaction buffer (20 mM Tris pH 8.0, 50 mM NaCl, 0.5 mM TCEP, and 5% w/v glycerol) together with a positive control sample containing 0.2  $\mu$ M 5'-FAM-labeled CpG-rich RNA and 2  $\mu$ M ZAP (1–227). Reactions were incubated at room temperature for 30 min before loading onto a 10% acrylamide tris-borate-EDTA gel prior to electrophoresis for 1 h in 1 X tris-borate-EDTA running buffer (pH 8.3) at 30 mA and 125 V on ice. Free and bound nucleic acids were imaged using a Typhoon 9500 imager with the FAM-default sample detection method ( $\lambda_{\text{laser}} = 473$  nm,  $\lambda_{\text{emission}} = 520$  nm, filter LPB) at gain between 600 and 700 V. Quantification of integrated band intensities was performed within the ImageQuant TL software (Cytiva).

### Biolayer interferometry

BLI was performed on an Octet R8 (Sartorius) instrument at 25 °C. Subsequently, 10 nM of a panel of biotinylated RNAs (Table S4) in BLI buffer (20 mM Tris pH 8.0, 50 mM NaCl, 0.5 mM TCEP, 0.01% Tween 20, and 8 units/ml RNasin Plus (Promega)) was immobilized onto streptavidin-coated biosensors. RNA-bound biosensors were then incubated with varying protein concentrations: 0 to 100  $\mu$ M of KHNYN (8–200), 0 to 20  $\mu$ M of IMP1 (V194–N369) or IMP1 (V104–N369, K294D, and E295D) and either 0 to 0.25  $\mu$ M or 0 to 2  $\mu$ M of ZAP (1–227). Between reads with biosensors, three-step regeneration was performed with BLI buffer containing 1 M NaCl. BLI interferograms were processed in Octet BLI Analysis 12.2.2.4 software (Sartorius), double referencing data with 0  $\mu$ M protein incubations and unloaded biosensors. Average responses were then calculated from the end point of association, converted to fraction bound ( $\theta$ ) based on the maximal response at saturation ( $R_{\text{max}}$ ), and plotted as a function of protein concentration.  $K_D$  values were determined in GraphPad Prism10 (GraphPad Software Inc) by fitting to a one site-specific hyperbolic binding curve ( $\theta = B_{\text{max}} * x / (K_D + x)$ ) where  $B_{\text{max}}$  is the maximal response.

### Statistical analysis

For restriction assays, statistical significance was determined using unpaired two-tailed *t* test in GraphPad Prism10 (GraphPad Software Inc.). Data are represented as mean  $\pm$  standard deviation, and significance was ascribed to *p* values *p* < 0.05.

### Data availability

All of the data except the coordinates of the KHNYN ex-diKH X-ray structure are contained within the manuscript. The coordinates have been deposited along with structure

factors at the Protein Data Bank (PDB) (<https://www.rcsb.org/>) with the accession code ID: 9HTS.

**Supporting information**—This article contains supporting information.

**Acknowledgments**—We thank members of the Neil, Taylor and Swanson laboratories for helpful discussions and Dr Robert Dagil, University of Copenhagen for the gift of IMP1 proteins. We gratefully acknowledge the Francis Crick Structural Biology Science Technology platform and Diamond Light Source, Didcot, UK (Block allocation group MX25587) and beamlines i04 for access and support from Beamline staff. The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: TZM-bl from Dr John C. Kappes, Dr Xiaoyun Wu, and Tranzyme Inc; HIV-1 p24 Hybridoma (183-H12-5C) from Dr Bruce Chesebro. The Antiserum to HIV-1 gp120 #20 (ARP421) was obtained from the NIBSC Centre for AIDS Reagents.

**Author contributions**—R. L. Y., M. J. L., C. B., S. K., H. W., M. A. C., A. G. P., E. R. M., S. J. D. N., I. A. T., and C. M. S. writing—review and editing; R. L. Y., M. J. L., C. B., S. K., H. W., M. A. C., A. G. P., and E. R. M. investigation; R. L. Y., M. J. L., C. B., S. K., H. W., M. A. C., A. G. P., E. R. M., and I. A. T. formal analysis; R. L. Y., I. A. T., and C. S. writing—original draft; S. J. D. N., I. A. T., and C. M. S. supervision; S. J. D. N., I. A. T., and C. M. S. project administration; S. J. D. N., I. A. T., and C. M. S. funding acquisition; S. J. D. N., I. A. T., and C. M. S. conceptualization.

**Funding and additional information**—This work was funded by the Wellcome Trust Senior Research Fellowship (WT098049A1A) to S. J. D. N. (<https://wellcome.org>), the Medical Research Council (<https://mrc.ukri.org>) grant MR/W018519/1 to S. J. D. N. and C. M. S., and The Francis Crick Institute to I. A. T., which receives its core funding from the Cancer Research UK (CC2029), the UK Medical Research Council (CC2029) and the Wellcome Trust (CC2029). R. L. Y. is a joint King's College London and The Francis Crick Institute PhD student. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: BLI, biolayer interferometry; CTB, C-terminal bundle; EDTA, ethylenediaminetetraacetic acid; ex-diKH domain, extended-diKH domain; FAM, fluorescein amidite; HRP, horseradish peroxidase; IFN- $\beta$ , interferon- $\beta$ ; KHNYN, KH and NYN containing protein; PDB, Protein Data Bank; PIN, PiIT N-terminal; RBD, RNA binding domain; TCEP, tris(2-carboxyethyl)phosphine; WT, wild-type; ZAP, zinc finger antiviral protein; ZRE, ZAP-response element.

### References

1. Gao, G., Guo, X., and Goff, S. P. (2002) Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* 297, 1703–1706
2. Guo, X., Carroll, J. W., Macdonald, M. R., Goff, S. P., and Gao, G. (2004) The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. *J. Virol.* 78, 12781–12787

3. Ficarella, M., Neil, S. J. D., and Swanson, C. M. (2021) Targeted restriction of viral gene expression and replication by the ZAP antiviral system. *Annu. Rev. Virol.* **8**, 265–283
4. Shao, R., Visser, I., Fros, J. J., and Yin, X. (2024) Versatility of the zinc-finger antiviral protein (ZAP) as a modulator of viral infections. *Int. J. Biol. Sci.* **20**, 4585–4600
5. Bick, M. J., Carroll, J. W., Gao, G., Goff, S. P., Rice, C. M., and MacDonald, M. R. (2003) Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *J. Virol.* **77**, 11555–11562
6. Takata, M. A., Goncalves-Carneiro, D., Zang, T. M., Soll, S. J., York, A., Blanco-Melo, D., *et al.* (2017) CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* **550**, 124–127
7. Meagher, J. L., Takata, M., Goncalves-Carneiro, D., Keane, S. C., Rebendenne, A., Ong, H., *et al.* (2019) Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24303–24309
8. Luo, X., Wang, X., Gao, Y., Zhu, J., Liu, S., Gao, G., *et al.* (2020) Molecular mechanism of RNA recognition by zinc-finger antiviral protein. *Cell Rep.* **30**, 46–52.e4
9. Goncalves-Carneiro, D., Mastrocola, E., Lei, X., DaSilva, J., Chan, Y. F., and Bieniasz, P. D. (2022) Rational attenuation of RNA viruses with zinc finger antiviral protein. *Nat. Microbiol.* **7**, 1558–1567
10. Shaw, A. E., Hughes, J., Gu, Q., Behdenna, A., Singer, J. B., Dennis, T., *et al.* (2017) Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol.* **15**, e2004086
11. Goncalves-Carneiro, D., Takata, M. A., Ong, H., Shilton, A., and Bieniasz, P. D. (2021) Origin and evolution of the zinc finger antiviral protein. *PLoS Pathog.* **17**, e1009545
12. Theys, K., Feder, A. F., Gelbart, M., Hartl, M., Stern, A., and Pennings, P. S. (2018) Within-patient mutation frequencies reveal fitness costs of CpG dinucleotides and drastic amino acid changes in HIV. *PLoS Genet.* **14**, e1007420
13. Caudill, V. R., Qin, S., Winstead, R., Kaur, J., Tisthammer, K., Pineda, E. G., *et al.* (2020) CpG-creating mutations are costly in many human viruses. *Evol. Ecol.* **34**, 339–359
14. Wasson, M. K., Borkakoti, J., Kumar, A., Biswas, B., and Vivekanandan, P. (2017) The CpG dinucleotide content of the HIV-1 envelope gene may predict disease progression. *Sci. Rep.* **7**, 8162
15. Kmiec, D., Nchioua, R., Sherrill-Mix, S., Sturzel, C. M., Heusinger, E., Braun, E., *et al.* (2020) CpG frequency in the 5' third of the env gene determines sensitivity of primary HIV-1 strains to the zinc-finger antiviral protein. *mBio* **11**, e02903-02919
16. Kypr, J., Mrazek, J., and Reich, J. (1989) Nucleotide composition bias and CpG dinucleotide content in the genomes of HIV and HTLV 1/2. *Biochim. Biophys. Acta* **1009**, 280–282
17. Shpaer, E. G., and Mullins, J. I. (1990) Selection against CpG dinucleotides in lentiviral genes: a possible role of methylation in regulation of viral expression. *Nucleic Acids Res.* **18**, 5793–5797
18. Ficarella, M., Wilson, H., Pedro Galao, R., Mazzon, M., Antzin-Anduetza, I., Marsh, M., *et al.* (2019) KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *Elife* **8**, e46767
19. Ficarella, M., Antzin-Anduetza, I., Hugh-White, R., Firth, A. E., Sertkaya, H., Wilson, H., *et al.* (2020) CpG dinucleotides inhibit HIV-1 replication through zinc finger antiviral protein (ZAP)-Dependent and -independent mechanisms. *J. Virol.* **94**, e01337-19
20. Lista, M. J., Witney, A. A., Nichols, J., Davison, A. J., Wilson, H., Latham, K. A., *et al.* (2023) Strain-dependent restriction of human cytomegalovirus by zinc finger antiviral proteins. *J. Virol.* **97**, e0184622
21. Anantharaman, V., and Aravind, L. (2006) The NYN domains: novel predicted RNases with a PIN domain-like fold. *RNA Biol.* **3**, 18–27
22. Castagnoli, L., Mandaliti, W., Nepravishta, R., Valentini, E., Mattioni, A., Procopio, R., *et al.* (2019) Selectivity of the CUBAN domain in the recognition of ubiquitin and NEDD8. *FEBS J.* **286**, 653–677
23. Hong, S., and Choe, J. (2024) Crystallization and biochemical studies of the NYN domain of human KHNYN. *Acta Crystallogr. F Struct. Biol. Commun.* **80**, 67–72
24. Lista, M. J., Ficarella, M., Wilson, H., Kmiec, D., Youle, R. L., Wanford, J., *et al.* (2023) A nuclear export signal in KHNYN required for its antiviral activity evolved as ZAP emerged in tetrapods. *J. Virol.* **97**, e0087222
25. Mino, T., and Takeuchi, O. (2021) Regnase-1-related endoribonucleases in health and immunological diseases. *Immunological Rev.* **304**, 97–110
26. Plianachaisuk, A., Kusama, K., Kato, K., Sriswasdi, S., Tamura, K., and Iwasaki, W. (2022) Origination of LTR retroelement-derived NYNRIN coincides with therian placental emergence. *Mol. Biol. Evol.* **9**, msac176
27. Yamasoba, D., Sato, K., Ichinose, T., Imamura, T., Koepke, L., Joas, S., *et al.* (2019) N4BP1 restricts HIV-1 and its inactivation by MALT1 promotes viral reactivation. *Nat. Microbiol.* **4**, 1532–1544
28. Gitlin, A. D., Heger, K., Schubert, A. F., Reja, R., Yan, D., Pham, V. C., *et al.* (2020) Integration of innate immune signalling by caspase-8 cleavage of N4BP1. *Nature* **587**, 275–280
29. Shi, H., Sun, L., Wang, Y., Liu, A., Zhan, X., Li, X., *et al.* (2021) N4BP1 negatively regulates NF-kappaB by binding and inhibiting NEMO oligomerization. *Nat. Commun.* **12**, 1379
30. Gu, H., Zheng, S., Han, G., Yang, H., Deng, Z., Liu, Z., *et al.* (2022) Porcine reproductive and respiratory syndrome virus adapts antiviral innate immunity via manipulating MALT1. *mBio* **13**, e0066422
31. Gitlin, A. D., Maltzman, A., Kanno, Y., Heger, K., Reja, R., Schubert, A. F., *et al.* (2024) N4BP1 coordinates ubiquitin-dependent cross-talk within the IkappaB kinase family to limit Toll-like receptor signaling and inflammation. *Immunity* **57**, 973–986.e7
32. Kliza, K. W., Song, W., Pinzuti, I., Schaubeck, S., Kunzelmann, S., Kuntin, D., *et al.* (2024) N4BP1 functions as a dimerization-dependent linear ubiquitin reader which regulates TNF signalling. *Cell Death Discov.* **10**, 183
33. Zheng, W., Guo, J., Ma, S., Sun, R., Song, Y., Chen, Y., *et al.* (2024) The NEDD4-binding protein N4BP1 degrades mRNA substrates through the coding sequence independent of nonsense-mediated decay. *J. Biol. Chem.* **300**, 107954
34. Marco, A., and Marin, I. (2009) CGIN1: a retroviral contribution to mammalian genomes. *Mol. Biol. Evol.* **26**, 2167–2170
35. Nicastro, G., Taylor, I. A., and Ramos, A. (2015) KH-RNA interactions: back in the groove. *Curr. Opin. Struct. Biol.* **30**, 63–70
36. 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
37. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., *et al.* (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444
38. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858
39. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
40. Holm, L., Laiho, A., Toronen, P., and Salgado, M. (2023) DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci.* **32**, e4519
41. Holm, L., and Sander, C. (1997) Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res.* **25**, 231–234
42. Teplova, M., Malinina, L., Darnell, J. C., Song, J., Lu, M., Abagyan, R., *et al.* (2011) Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. *Structure* **19**, 930–944
43. Du, Z., Fenn, S., Tjhen, R., and James, T. L. (2008) Structure of a construct of a human poly(C)-binding protein containing the first and second KH domains reveals insights into its regulatory mechanisms. *J. Biol. Chem.* **283**, 28757–28766
44. Dagil, R., Ball, N. J., Ogrodowicz, R. W., Hobor, F., Purkiss, A. G., Kelly, G., *et al.* (2019) IMP1 KH1 and KH2 domains create a structural platform with unique RNA recognition and re-modelling properties. *Nucleic Acids Res.* **47**, 4334–4348
45. Diaz-Moreno, I., Hollingworth, D., Kelly, G., Martin, S., Garcia-Mayoral, M., Briata, P., *et al.* (2010) Orientation of the central domains of KSRP and its implications for the interaction with the RNA targets. *Nucleic Acids Res.* **38**, 5193–5205

46. OhAinle, M., Helms, L., Vermeire, J., Roesch, F., Humes, D., Basom, R., *et al.* (2018) A virus-packageable CRISPR screen identifies host factors mediating interferon inhibition of HIV. *Elife* **7**, e39823
47. Sharma, P., Murillas, R., Zhang, H., and Kuehn, M. R. (2010) N4BP1 is a newly identified nucleolar protein that undergoes SUMO-regulated polyubiquitylation and proteasomal turnover at promyelocytic leukemia nuclear bodies. *J. Cell Sci.* **123**, 1227–1234
48. Huang, S., Chen, Z., Yan, X., Yu, T., Huang, G., Yan, Q., *et al.* (2014) Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat. Commun.* **5**, 5896
49. Hollingworth, D., Candel, A. M., Nicastro, G., Martin, S. R., Briata, P., Gherzi, R., *et al.* (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res.* **40**, 6873–6886
50. Garcia-Mayoral, M. F., Hollingworth, D., Masino, L., Diaz-Moreno, I., Kelly, G., Gherzi, R., *et al.* (2007) The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation. *Structure* **15**, 485–498
51. Bohn, J. A., Meagher, J. L., Takata, M. A., Goncalves-Carneiro, D., Yeoh, Z. C., Ohi, M. D., *et al.* (2024) Functional anatomy of zinc finger antiviral protein complexes. *Nat. Commun.* **15**, 10834
52. Yeoh, Z. C., Meagher, J. L., Kang, C. Y., Bieniasz, P. D., Smith, J. L., and Ohi, M. D. (2024) A minimal complex of KHNYN and zinc-finger antiviral protein binds and degrades single-stranded RNA. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2415048121
53. Odon, V., Fros, J. J., Goonawardane, N., Dietrich, I., Ibrahim, A., Alshaikhahmed, K., *et al.* (2019) The role of ZAP and OAS3/RNaseL pathways in the attenuation of an RNA virus with elevated frequencies of CpG and UpA dinucleotides. *Nucleic Acids Res.* **47**, 8061–8083
54. Fros, J. J., Visser, I., Tang, B., Yan, K., Nakayama, E., Visser, T. M., *et al.* (2021) The dinucleotide composition of the Zika virus genome is shaped by conflicting evolutionary pressures in mammalian hosts and mosquito vectors. *PLoS Biol.* **19**, e3001201
55. Sharp, C. P., Thompson, B. H., Nash, T. J., Diebold, O., Pinto, R. M., Thorley, L., *et al.* (2023) CpG dinucleotide enrichment in the influenza A virus genome as a live attenuated vaccine development strategy. *PLoS Pathog.* **19**, e1011357
56. Hong, S., and Choe, J. (2024) Crystal structure of NYN domain of Human KHNYN in complex with single strand RNA. *Biochem. Biophys. Res. Commun.* **738**, 150545
57. Jolma, A., Zhang, J., Mondragon, E., Morgunova, E., Kivioja, T., Laverty, K. U., *et al.* (2020) Binding specificities of human RNA-binding proteins toward structured and linear RNA sequences. *Genome Res.* **30**, 962–973
58. Nakel, K., Hartung, S. A., Bonneau, F., Eckmann, C. R., and Conti, E. (2010) Four KH domains of the *C. elegans* Bicaudal-C ortholog GLD-3 form a globular structural platform. *RNA* **16**, 2058–2067
59. Saleh, H., Liloglou, T., Rigden, D. J., Parsons, J. L., and Grundy, G. J. (2024) KH-Like domains in PARP9/DTX3L and PARP14 coordinate protein-protein interactions to promote cancer cell survival. *J. Mol. Biol.* **436**, 168434
60. Oddone, A., Lorentzen, E., Basquin, J., Gasch, A., Rybin, V., Conti, E., *et al.* (2007) Structural and biochemical characterization of the yeast exosome component Rps40. *EMBO Rep.* **8**, 63–69
61. Antzin-Anduetza, I., Mahiet, C., Granger, L. A., Odendall, C., and Swanson, C. M. (2017) Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication. *Retrovirology* **14**, 49
62. Pear, W. S., Miller, J. P., Xu, L., Pui, J. C., Soffer, B., Quackenbush, R. C., *et al.* (1998) Efficient and rapid induction of a chronic myelogenous leukemia-like myeloproliferative disease in mice receiving P210 bcr/abl-transduced bone marrow. *Blood* **92**, 3780–3792
63. Fouchier, R. A., Meyer, B. E., Simon, J. H., Fischer, U., and Malim, M. H. (1997) HIV-1 infection of non-dividing cells: evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import. *EMBO J.* **16**, 4531–4539
64. Kmiec, D., Lista, M. J., Ficarella, M., Swanson, C. M., and Neil, S. J. D. (2021) S-farnesylation is essential for antiviral activity of the long ZAP isoform against RNA viruses with diverse replication strategies. *PLoS Pathog.* **17**, e1009726
65. Derdeyn, C. A., Decker, J. M., Sfakianos, J. N., Wu, X., O'Brien, W. A., Ratner, L., *et al.* (2000) Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. *J. Virol.* **74**, 8358–8367
66. Wei, X., Decker, J. M., Liu, H., Zhang, Z., Arani, R. B., Kilby, J. M., *et al.* (2002) Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother.* **46**, 1896–1905
67. Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B., and Kabat, D. (1998) Effects of CCR5 and CD4 cell surface concentrations on infections by macrophage-tropic isolates of human immunodeficiency virus type 1. *J. Virol.* **72**, 2855–2864
68. Chesebro, B., Wehrly, K., Nishio, J., and Perryman, S. (1992) Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. *J. Virol.* **66**, 6547–6554
69. Winter, G. (2010) An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190
70. Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., *et al.* (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D Struct. Biol.* **74**, 85–97
71. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
72. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
73. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., *et al.* (2008) Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435
74. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., *et al.* (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302
75. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
76. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82
77. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214
78. Chen, V. B., Wedell, J. R., Wenger, R. K., Ulrich, E. L., and Markley, J. L. (2015) MolProbity for the masses-of data. *J. Biomol. NMR* **63**, 77–83
79. Joosten, R. P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A. C., Blanchet, C., *et al.* (2009) PDB\_REDO: automated re-refinement of X-ray structure models in the PDB. *J. Appl. Crystallogr.* **42**, 376–384
80. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191