The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis

Jonathan J. Ipsaro1,2*, Astrid D. Haase2*, Simon R. Knott2, Leemor Joshua-Tor1,2 & Gregory J. Hannon2

PIWI-family proteins and their associated small RNAs (piRNAs) act in an evolutionarily conserved innate immune mechanism to provide essential protection for germ-cell genomes against the activity of mobile genetic elements. piRNA populations comprise a molecular definition of transposons, which permits them to distinguish transposons from host genes and selectively silence them. piRNAs can be generated in two distinct ways, forming either primary or secondary piRNAs. Primary piRNAs come from discrete genomic loci, termed piRNA clusters, and seem to be derived from long, single-stranded precursors. The biogenesis of primary piRNAs involves at least two nucleolytic steps. An unknown enzyme cleaves piRNA cluster transcripts to generate monophosphorylated piRNA 5' ends, and piRNA 3' ends are probably formed by exonucleolytic trimming, after a piRNA precursor is loaded into its PIWI partner. Secondary piRNAs arise during the adaptive evolution of PIWIs themselves. A number of proteins have been implicated genetically in primary piRNA biogenesis. One of these, Drosophila melanogaster Zucchini, is a member of the phospholipase-D family of phosphodiesterases, which includes both phospholipases and nuclease homologues. Here we produced a dimeric, soluble fragment of the mouse Zucchini homologue (mZuc; also known as PLD6) and show that it possesses single-strand-specific nuclease activity. A crystal structure of mZuc at 1.75 Å resolution indicates greater architectural similarity to phospholipase-D family nuclease than to phosphodiesterases. Together, our data suggest that the Zucchini proteins act in primary piRNA biogenesis as nuclease, perhaps generating the 5' ends of primary piRNAs.

zucchini (zuc) was first noted as a gene essential for female fertility in Drosophila. Two independent alleles caused fully penetrant sterility and the production of eggs with dorsoventral patterning defects. Subsequent studies have traced the effects of Zucchini on germ-cell development to its function in the piRNA pathway. Animals lacking Zucchini fail to silence transposons and show a general collapse of primary piRNA populations. Mutant animals also accumulate transcripts from piRNA clusters, indicating a failure to process these precursors into small RNAs. This indicates Zucchini is an important regulator of primary piRNA biogenesis.

Zucchini belongs to the phospholipase-D (PLD) family of phosphodiesterases, which share a common biochemical mechanism and a signature H(X)K(X4)D (HKD) motif within their active site (reviewed in ref. 10). Notably, one of the zuc alleles that emerged from the original forward genetic screen was a point mutation that changed the catalytic histidine to a tyrosine. This produced a phenocopy of the presumed null allele, including similar effects on piRNA populations, indicating that the catalytic activity of Zucchini was critical for piRNA production or stability.

Biochemical and genetic studies of the mouse Zucchini homologue, mZuc, led to the conclusion that it acted as a phospholipase that affected mitochondrial fusion in a manner linked to its processing of the mitochondrial lipid cardiolipin. Thus, it was suggested that the impact of Zucchini on the piRNA pathway was indirect, through alterations in lipid levels or through changes in the structure of mitochondria. However, the proposed model of Zucchini activity required an enzyme that localized to the outer mitochondrial membrane and faced the cytoplasm to hydrolyse a lipid that is almost exclusively found on the inner mitochondrial membrane. We therefore sought to discriminate between the two divergent hypotheses for Zucchini function in piRNA biogenesis; namely that it acts indirectly through its role as a phosphodiesterase, or that it acts directly as a nuclease.

To enable biochemical analysis of mZuc, we expressed an amino-terminally truncated form of the protein in S9 cells (Fig. 1a). This produced a soluble enzyme by removal of the transmembrane domain that normally anchors it to the mitochondrial outer membrane. One of the hallmarks of the PLD family is the presence of two HKD motifs that are brought together to create the catalytic centre. This active site can arise from a single polypeptide that contains two copies of the motif, an arrangement typical of PLDs with phospholipid substrates. Alternatively, the catalytic centre can be assembled as an intermolecular dimer, a form most common with PLDs that act as nucleases. mZuc has only one HKD motif, but the protein formed a stable, ~41 kDa dimer in S9 cells, as indicated by both gel-filtration chromatography and multi-angle light scattering (Supplementary Fig. 1). For comparison, we also produced a mutant version of the protein, which lacks an intact catalytic motif (H153N).

We used two assays to test whether recombinant mZuc could function as a phospholipase, liberating phosphatic acid (PA) from cardiolipin as previously reported. Liposomes containing commercial cardiolipin were incubated with recombinant proteins, and PA was measured by thin-layer chromatography (TLC). This did not reveal the production of measurable amounts of PA by mZuc, whereas a commercially available PLD from Streptomyces chromofuscus (scPLD) completely hydrolysed the substrate (Supplementary Fig. 2a). We also used a more sensitive assay based on mass spectrometry (MS). Selected reaction monitoring MS (SRM-MS) provides an extremely accurate and sensitive method for measuring reaction components. Using this approach, no substantial changes in cardiolipin or PA were observed when comparing buffer controls to either wild-type or catalytically inactive mZuc proteins (Fig. 1b and Supplementary Fig. 2b, c); whereas scPLD again completely hydrolysed the substrate. These results failed to support mZuc acting as a phospholipase to promote piRNA biogenesis, and prompted us to examine whether it has alternative activities.

PLD-family enzymes have been demonstrated to act as nucleases, cleaving double-stranded DNA (dsDNA) substrates. For example, prototypic PLD nucleases Nuc and Bfi1 cleave dsDNA either nonspecifically or at a defined recognition site. We therefore incubated mZuc with a selection of end-labelled DNA: single-stranded, double-stranded, or partially duplexed (Fig. 1c). The wild-type enzyme hydrolysed single-stranded DNA (ssDNA) or the single-stranded portions of partially duplexed substrates, but did not cleave dsDNA. The mutant

1W. M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. 2Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.
*These authors contributed equally to this work.
enzyme (H153N) was inactive against all substrates. As expected from studies of related PLD-family nucleases\(^\text{17}\), mZuc activity did not depend on the presence of divalent cations for substrate cleavage, but it could be stimulated by the presence of specific cations, particularly Mn\(^{2+}\), Ca\(^{2+}\) and Zn\(^{2+}\) (Supplementary Fig. 3a). The impact of Zn\(^{2+}\) may be explained by structural effects (see later), but as yet we do not understand how the other cations enhance activity.

The known catalytic mechanisms of PLD-family nucleases, including bacterial Nuc, suggest that cleavage proceeds via a two-step reaction scheme\(^\text{15}\). This includes the formation of a short-lived, covalent enzyme–substrate intermediate, joining a phosphate to the histidine in the HKD motif (Supplementary Fig. 4a). A similar mechanism for the enzyme–substrate intermediate, joining a phosphate to the histidine explains by structural effects (see later), but as yet we do not understand how the other cations enhance activity.

The known catalytic mechanisms of PLD-family nucleases, including bacterial Nuc, suggest that cleavage proceeds via a two-step reaction scheme\(^\text{15}\). This includes the formation of a short-lived, covalent enzyme–substrate intermediate, joining a phosphate to the histidine in the HKD motif (Supplementary Fig. 4a). A similar mechanism for the enzyme–substrate intermediate, joining a phosphate to the histidine explains by structural effects (see later), but as yet we do not understand how the other cations enhance activity.

Figure 1 | mZuc acts as a nuclease but not a phospholipase in vitro. a, The domain architecture of mZuc and Zuc are similar, with an N-terminal transmembrane helix (TM, red) and a cytoplasmic domain, which contains the catalytic HKD motif (gold). The construct used for crystallization is indicated as a dashed box. Residue numbers delineating each domain are indicated below each schematic. b, Phosphodiesterase activity for mZuc, its catalytic mutant (H153N), and a known phospholipase (scPLD) were monitored by SRM-MS. Levels of the cardiolipin (CL) substrate and the expected PA product are shown for each reaction. Error bars indicate standard deviation (n = 3). WT, wild type.

If mZuc functions as a nuclease in the piRNA pathway, it would probably act on RNA substrates, and the biochemical mechanism used by PLD-family enzymes is consistent with their potentially cleaving either DNA or RNA. We therefore tested the ability of mZuc to cleave a variety of single-stranded or duplexed RNA substrates. mZuc cleaved only ssRNA, and this reaction was fully inhibited by orthovanadate (Fig. 2a). The mZuc mutant bearing an alteration in its catalytic site was inactive in these assays (Fig. 2b and Supplementary Fig. 3b). These data indicate that mZuc can function as a backbone-non-specific, single-strand-specific nuclease, cleaving either RNA or DNA substrates. The binding affinity of mZuc for ssDNA or ssRNA substrates was nearly identical, measured at roughly 50 nM in each case (Supplementary Fig. 5). In vivo, the physiologically relevant targets of mZuc are probably determined by its subcellular localization and perhaps involve the help of additional proteins, which would focus its activity towards certain substrates.

HKD-family nucleases are predicted to leave 5′-phosphate and 3′-hydroxyl termini. These termini also correspond to the end polarities

Figure 2 | mZuc acts as a single-strand-specific endoribonuclease in vitro. a, mZuc was incubated with single-stranded, double-stranded, and partially dsRNA substrates in the presence of Na\(^{3+}\)VO\(_4\) as indicated. b, mZuc RNA cleavage products (from reactions shown in panel a, as indicated below) were tested for sensitivity to β-elimination and accessibility for polyadenylation.
that would be expected of a nuclease that participated in the processing of primary piRNA transcripts. We tested whether the nucleolytic products of mZuc had such termini in several ways. β-Elimination shifted mZuc cleavage products by a single base, a reaction that is diagnostic of the presence of a 3′-OH terminus (Fig. 2b). We also confirmed the presence of a 3′-OH group by extending the cleavage products with poly A polymerase (Fig. 2b). DNA cleavage also produced 5′ phosphate and 3′ hydroxyl termini (Supplementary Fig. 3c).

To gain further insight into mZuc activity and substrate specificity, we determined its structure to 1.75 Å resolution. Like other members of the HKD family, the structure of mZuc consists of a conserved area of 2,600 Å², as calculated by the Protein Interfaces, Surfaces and Assemblies (PISA) server. The resulting arrangement of the active site residues and the distance between the catalytic histidines (His 153) is consistent among both monomeric and dimeric PLD structures, as evidenced by both the apo structure and the structure of the protein in complex with tungstate, a phosphate mimic that binds to the active site (Fig. 3a and Supplementary Fig. 6). Although mZuc is clearly an HKD-family enzyme, a number of features unique to this structure support its biological activity as a nuclease and, more specifically, a single-stranded RNase. First, and most strikingly, a long, ordered loop is inserted between residues Pro 44–Ser 75. This loop extends away from the active site and contains three cysteines (Cys 49, Cys 66 and Cys 68) as well as a histidine (His 72) that form an unexpected CCCH ‘zinc wing’ (Fig. 3c and Supplementary Fig. 7). Although zinc finger motifs are widely used for nucleic-acid binding, this particular instance escaped bioinformatic detection due to the atypical primary sequence of the CCCH motif. Interestingly, the CCCH class of zinc fingers, in particular, has been implicated in the binding of ssRNA molecules such as messenger RNA and viral RNAs, in accord with ssRNA being the likely in vivo substrate for mZuc.

PLD-family members that are responsible for lipid metabolism typically have their active site in a structural ‘pocket’. In contrast, we observed that those with nucleic acid substrates typically provide a larger substrate-binding ‘groove’ (Fig. 4 and Supplementary Fig. 8). In the mZuc structure, the active site is flanked by a positively charged groove that extends to the zinc wings (Fig. 4a and Supplementary Fig. 9), probably serving as a nucleic-acid-binding interface. When compared to the structure of Nuc, a PLD nuclease that acts on dsDNA (Fig. 4b), the width of the groove in mZuc is considerably narrower, consistent with the biochemically observed single-strand specificity of the enzyme.

We constructed a hypothetical model to examine the potential interaction between the proposed substrate-binding groove of mZuc and a ssRNA. This was subjected to energy minimization using the Groningen Machine for Chemical Simulations (GROMACS) server. The resulting model (Fig. 4a and Supplementary Fig. 10) illustrates the shape and charge complementarity that a single-stranded nucleic acid substrate provides, including appropriate placement of the scissile phosphate in the active site. The surface complementarity is particularly notable; however, based on the projections of the bases away from the core of the protein, it seems unlikely that mZuc would show a strong sequence bias for binding or cleavage.

Our results indicate that Zucchini functions as a nuclease to promote primary piRNA biogenesis. Given the results of biochemical analyses performed in silkworm extracts, it is highly probable that the mature 3′ ends of piRNAs are formed by exonucleolytic trimming of precursor piRNAs that are already loaded into PIWI proteins. This leaves open the possibility that Zucchini could generate the 5′ ends of primary piRNAs. mZuc does produce products with the correct phosphate polarity. Yet, primary piRNAs show an overwhelming bias for a terminal U residue, which does not seem to be a preference of mZuc, based either upon biochemical or structural studies. Therefore, if Zucchini does generate piRNA 5′ ends, the prevalence of their characteristic

![Figure 3](image_url) **Figure 3 | Crystal structure of mZuc.** a. The overall structure of the mZuc dimer is shown as a ribbon diagram. Helices are in green, strands in blue, and loops in beige. Each monomer binds one Zn²⁺ (yellow) in an extended zinc wing. The active site histidine residues (His 153) are highlighted in red. The N- and C-termini for each protomer are indicated. b. A close-up of the zinc wing consisting of residues Cys 49, Cys 66, Cys 68 and His 72 is shown. c. A detailed view of mZuc co-crystallized with tungstate bound in the active site.
a strong binding preference for 5′-end binding within the Argonaute sub-family for proteins. The latter is a reasonable possibility as there is ample biochemical and structural support within the Argonaute sub-family for strong binding preferences for 5′-terminal nucleotides. We must also consider the possibility that Zucchini proteins could contribute an additional, as yet unanticipated activity, perhaps generating intermediate 3′ ends of precursor piRNAs that are further resected by trimming to form mature termini. Attributing a definitive role to Zucchini will rest on further studies, perhaps ultimately requiring a full biochemical reconstitution of primary piRNA biogenesis to finally resolve its function.

In addition to their implications for piRNA production, our studies also highlight some general features of the phosphodiesterases, which use HKD motifs as their active sites. A comparison of the available structures of the PLD/nuclease family proteins indicates that these enzymes define their substrate specificity by their binding properties. Whereas phospholipases within this family seem to harbour pocket-like substrate-binding structures, nucleases show extended binding grooves for nucleic acid chains. These structural motifs are nearly diagnostic of the substrate specificity of these proteins. That said, solely on the basis of our biochemical assays and existing literature we cannot exclude that mZuc can act in some circumstances as a phospholipase. We can simply argue that its structural features are more consistent with its action as a nuclease. The fact that enzymes in this family can be divided into classes based on their structural and biochemical features may also suggest that the PLD nomenclature for some of these enzymes, particularly for mZuc and Nuc, has become misleading and deserves reconsideration.

METHODS SUMMARY

Cloning, expression and purification of recombinant mZuc. A complementary DNA coding for Mus musculus Zucchini (residues 31–221; mZuc) fused to a carboxy-terminal, thrombin-protease-cleavable Strep tag was cloned into the vector pFL for expression in SF9 cells. The recombinant protein was affinity purified using StrepTactin resin followed by tag cleavage and Superdex75 gel filtration in 0.1 M MES, pH 6.5, 150 mM NaCl, 1 mM dithiothreitol (DTT).

Protein characterization. Protein was judged to be >99% pure by SDS–PAGE. The oligomeric state and mass of the protein were measured by multi-angle light scattering and liquid chromatography electrospray ionization-mass spectrometry (LC ESI-MS), respectively.

Nuclease activity assay. Recombinant proteins and 5′-32P-labelled nucleotide substrates were incubated at 37 °C in 50 mM MES, pH 6.5, 75 mM NaCl, 2 mM CaCl2, 1 mM DTT for DNase activity; 5 mM MES, pH 6.5, 7.5 mM NaCl, 2 mM CaCl2, 1 mM DTT for RNase activity. After protease K treatment, nucleic acids were extracted using phenol/chloroform and analysed by urea–PAGE.

Lipase activity assay. Liposomes were prepared using standard methods, then mixed with recombinant protein at 37 °C for 2 h. Lipids were then extracted and analysed by TLC and SRM-MS.

Crystalization and structure solution. Crystals of mZuc were grown by hanging-droplet vapour diffusion at 4 °C. Immediately before crystallization, the protein was mixed with chymotrypsin (1:100 m/m ratio). Protein solution at 3 mg ml⁻¹ was mixed with reservoir solution (50 mM Bis-Tris, pH 6.5, 18% PEG-3350 and 2% tascimate, pH 6.0). Crystals were cryoprotected in reservoir solution with 20% tasmic, 20% glycol and 10% glycerol. Crystals were grown for 4–6 weeks.

DNA coding for Zucchini (residues 31–221; mZuc) fused to a carboxy-terminal, thrombin-protease-cleavable Strep tag was cloned into the vector pFL for expression in SF9 cells. The recombinant protein was affinity purified using StrepTactin resin followed by tag cleavage and Superdex75 gel filtration in 0.1 M MES, pH 6.5, 150 mM NaCl, 1 mM dithiothreitol (DTT).

Protein characterization. Protein was judged to be >99% pure by SDS–PAGE. The oligomeric state and mass of the protein were measured by multi-angle light scattering and liquid chromatography electrospray ionization-mass spectrometry (LC ESI-MS), respectively.

Nuclease activity assay. Recombinant proteins and 5′-32P-labelled nucleotide substrates were incubated at 37 °C in 50 mM MES, pH 6.5, 75 mM NaCl, 2 mM CaCl2, 1 mM DTT for DNase activity; 5 mM MES, pH 6.5, 7.5 mM NaCl, 2 mM CaCl2, 1 mM DTT for RNase activity. After protease K treatment, nucleic acids were extracted using phenol/chloroform and analysed by urea–PAGE.

Lipase activity assay. Liposomes were prepared using standard methods, then mixed with recombinant protein at 37 °C for 2 h. Lipids were then extracted and analysed by TLC and SRM-MS.

Crystalization and structure solution. Crystals of mZuc were grown by hanging-drap vapour diffusion at 4 °C. Immediately before crystallization, the protein was mixed with chymotrypsin (1:100 m/m ratio). Protein solution at 3 mg ml⁻¹ was mixed with reservoir solution (50 mM Bis-Tris, pH 6.5, 18% PEG-3350 and 2% tascimate, pH 6.0). Crystals were cryoprotected in reservoir solution with 20% ethylene glycol. Diffraction data were collected at the National Synchrotron Light Source. The structure was solved by molecular replacement using Nuc4 as a search model.

RNA modelling. RNA modelling was performed manually, then subjected to energy minimization using GROMACS.

Full Methods and any associated references are available in the online version of the paper.

Received 7 June; accepted 10 August 2012.

Published online 14 October 2012.
7. Schupbach, T. & Wieschaus, E. Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. Genetics 129, 1119–1136 (1991).

8. Malone, C. D. et al. Specialized pIRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell 137, 522–535 (2009).

9. Olivia, D., Sylka, M. M., Sachidanandam, R., Mechtler, K. & Breneche, J. An in vivo RNAi assay identifies major genetic and cellular requirements for primary pIRNA biogenesis in Drosophila. EMBO J. 29, 3301–3317 (2010).

10. Selvy, P. E., Laviere, R. R., Lindsay, G. W. & Brown, H. A. Phospholipase D: enzymology, functionality, and chemical modulation. Chem. Rev. 111, 6064–6119 (2011).

11. Choi, S. Y. et al. A common lipid link Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis. Nature Cell Biol. 8, 1255–1262 (2006).

12. Huang, H. et al. pIRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. Dev. Cell 20, 376–387 (2011).

13. Watanebe, T. et al. MitoPLD is a mitochondrial protein essential for nuage formation and pIRNA biogenesis in the mouse germline. Dev. Cell 20, 364–375 (2011).

14. González, F. & Gottlieb, E. Cardiolipin: setting the beat of apoptosis. Apoptosis 12, 877–885 (2007).

15. Gottlin, E. B., Rudolph, A. E., Zhao, Y., Matthews, H. R. & Dixon, J. E. Catalytic mechanism of the phospholipase D superfamily proceeds via a covalent phosphohistidine intermediate. Proc. Natl Acad. Sci. USA 95, 9202–9207 (1998).

16. Lackey, D., Walker, G. C., Keng, T. & Linn, S. Characterization of an endonuclease associated with the drug resistance plasmid pKM101. J. Bacteriol. 131, 583–588 (1977).

17. Pohlman, R. F., Liu, F., Wang, L., More, M. I. & Winans, S. C. Genetic and biochemical analysis of an endonuclease encoded by the IncV plasmid pKM101. Nucleic Acids Res. 21, 4867–4872 (1993).

18. Stickley, J. A. & Dixon, J. E. Crystal structure of a phospholipase D family member. Nature Struct. Biol. 6, 278–284 (1999).

19. Kossi, E. E. & Hendrick, K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797 (2007).

20. Berg, J. M. & Shi, Y. The galvanization of biology: a growing appreciation for the roles of zinc. Science 271, 1081–1089 (1996).

21. Lai, W. S., Carballo, E., Thorn, J. M., Kennington, E. A. & Blackshear, P. J. Interactions of CCCH zinc finger proteins with mRNA: Binding of tristetraprolin-related zinc finger proteins to AU-rich elements and destabilization of mRNA. J. Biol. Chem. 275, 17827–17837 (2000).

22. Kelly, S. M. et al. Recognition of polyadenosine RNA by zinc finger proteins. Proc. Natl Acad. Sci. USA 104, 12306–12311 (2007).

23. Hurt, J. A. et al. A conserved CCCH-type zinc finger protein regulates mRNA nuclear adenylation and export. J. Cell Biol. 185, 265–277 (2009).

24. Gao, G., Guo, X. & Goff, S. P. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. Science 297, 1703–1706 (2002).

25. Van Der Spoel, D. et al. GROMACS: fast, flexible, and free. J. Comput. Chem. 26, 1701–1718 (2005).

26. Aravin, A. et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442, 203–207 (2006).

27. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442, 199–202 (2006).

28. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. Nature 465, 818–822 (2010).

29. Mi, S. et al. Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide. Cell 133, 116–127 (2008).

30. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. Acta Crystallogr. D 55, 938–940 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank W. Filippowicz, R. MacDonald and members of the G.J.H. and L.J. laboratories for discussions; G. Bencze, K. Rivera and D. Pappin of the Cold Spring Harbor Laboratory proteomics facility, which is funded in part by an National Cancer Institute Cancer Center Support Grant (CA045508), for support with mass spectrometry; and H. Robinson for help at the National Synchrotron Light Source, which is supported by the Department of Energy, Office of Basic Energy Sciences. J.J.I. was supported by the Harvey L. Karp award and by a Ruth L. Kirschstein National Research Service Awards National Institutes of Health (NIH) fellowship F32GM97888. This work was supported by NIH grant R01GM062534. G.J.H. and L.J. are Howard Hughes Medical Institute Investigators.

Author Contributions L.J., G.J.H., A.D.H. and J.J.I. planned studies and wrote the paper. A.D.H. and J.J.I. performed the experiments, and S.R.K. analysed datasets.

Author Information Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.edu) or L.J. (leemor@cshl.edu).
METHODS

Cloning, expression and purification of mZuc fragments. To screen for biochemically well-behaved targets, a panel of mZuc constructs was generated from M. musculus complementary DNA by sequence- and ligitation-independent cloning (SLIC) cloning. These constructs presented various N- or C-terminal tags for enhanced expression and purification using either Escherichia coli or baculovirus-induced insect cell culture systems. The sequence of each construct was verified by the Cold Spring Harbour Laboratory DNA Sequencing Facility.

The data presented here resulted from a 30-residue N-terminal deletion of mZuc to create a protein that spans residues 31–221 fused to a thrombin-cleavable Strep-tag at the C terminus (mZuc30–ThStrep2). This construct was cloned into the vector pFL for expression in Sf9 cells using the baculovirus expression system30. After expression, cells were harvested by centrifugation at 1,000g, resuspended in lysis buffer (0.1 M MES, pH 6.5, 0.15 M NaCl, 1 mM dithiothreitol (DTT)) (~20 ml per litre culture), and lysed by sonication. The cell lysate was then clarified by ultracentrifugation at 125,000g for 1 h and the supernatant was applied to a Strep-Tactin (IBA) column equilibrated with lysis buffer. The bound mZuc30–ThStrep2 was subsequently washed with lysis buffer, further washed with lysis buffer containing 2 mM ATP, and finally eluted in lysis buffer containing 5 mM d-desthiobiotin. To remove the C-terminal affinity purification tag, 15 units of thrombin protease were added per mg of purified protein and incubated overnight at 4°C. The cleaved mZuc30 (referred to as mZuc) was then further purified by gel filtration using a Superdex75 column equilibrated with lysis buffer. Once purified, the protein was concentrated to 5–10 mg ml⁻¹ and stored at 4°C for short periods or in 50% (v/v) glycerol at −20°C for extended periods. Typical yields were 1–2 mg of purified protein (>98% purity as assessed by SDS–PAGE) per litre culture. Active site mutants were constructed using SLIC cloning methods. Purification of mutant proteins was identical to that for the wild type.

For SRM–MS experiments, mZuc30–ThStrep, and mZuc30–ThHis5, H153N were co-expressed in Sf9 cells. The resulting heterodimer population (WT/ H153N) was isolated by Strep-Tactin purification followed by a second round of affinity purification with Ni-NTA resin. The purified protein was immediately concentrated and desalted into 20 mM MES, pH 6.5, 0.15 M NaH₂PO₄, 1 mM DTT for subsequent SRM-MS processing.

Multi-angle light scattering (MALS) was used to determine the oligomeric state of the purified proteins. Roughly 1 mg of purified protein (at 2 mg ml⁻¹) was taken for in-line gel filtration on a Superdex75 column followed by light scattering analysis. MALS was measured with a Wyatt Dawn Helios-II and processed using the included software (ASTRA Version 5.3.4).

Biovine serum albumin (BSA) was used as a control to ensure proper calibration.

Intact mass measurements from MS. Each purified protein was diluted in water to 1 μM and applied to a Zorbax 300SB-C8 enrichment chip at 600 nl min⁻¹. In total, ~20 pmol of material was injected. A mobile phase gradient from 0.1% formic acid and 3% acetonitrile in water to 0.1% formic acid and 90% acetonitrile was used to chromatograph each sample, which was then taken for in-line electrospray ionization mass spectrometry. Mass spectra were obtained on an Agilent 6520 Accurate-Mass Quadrupole time-of-flight mass spectrometer with an accelerating voltage of 1,850 V. Deconvolution was performed with the included software, Agilent Mass Hunter Qualitative Analysis Version B04.00. Masses of all proteins were within 1.0 Da of those predicted.

Liposome preparation. To assess the phospholipase activity of the mZuc, cardiolipin-containing liposomes were prepared. Defined liposomes (phosphatidylcholine:phosphatidylethanolamine:phosphatidylserine:cardiolipin (PC:PE:PS:CL) at 2:2:1:1) or extract-based liposomes (made with bovine heart lipid extract supplemented with CL) were made using standard methods. Briefly, lipids (Avanti Polar Lipids) in chloroform were mixed and then dried using a stream of nitrogen followed by overnight evaporation. Lipid films were then resuspended in 10 mM HEPES, pH 7.4, 1 mM DTT (nuclease buffer conditions) or 50 mM HEPES, pH 7.4, 80 mM NaCl, 1 mM MgCl₂, 2 mM CaCl₂, 1 mM DTT (previously reported lipase conditions)³³. Commercially available PLD from S. chromofuscus (MP Biomedicals) was used as a positive control. Each 200 nl reaction was incubated at 37°C for 2 h, then quenched by the addition of 750 μl of methanol/chloroform (2:1).

Lipids were extracted by sequentially adding 250 μl chloroform and 250 μl 0.5 M NaCl in 0.1 M HCl (in water) with vigorous vortexing after each addition. The lower (organic) phase was removed, dried by vacuum evaporation, and resuspended in a small volume of methanol/chloroform for TLC analysis. Each of the extracts was spotted onto 2.5 × 7.5 cm, glass-backed silica 60 F₂₅₄ TLC plates (EMD). The extracts were chromatographed using a running solvent of chloroform:methanol:water:ammonium hydroxide (120:75:6:2), stained using permanganate (0.75% KMnO₄, 5% KH₂PO₄, 0.06% NaOH), and developed by heating. Extract components appeared as yellow spots on a magenta background. Lipid identification was based on standards run in parallel.

Cardiolipase activity assay using SRM–MS. To confirm the findings of the TLC-based phospholipase activity assay and to obtain higher sensitivity for detecting the presumed phosphatidic acid product, an SRM–MS assay was used. After performing the lipase reaction described above, 5 μl of each (non-extracted) defined lipidosome lipase reaction were infused into a Thermo Scientific TriPlus Triple-Stage Quadrupole (TSQ) Vantage mass spectrometer. Mass-to-charge ratio (m/z) peaks corresponding to cardiolipin (1812), species with a +1 charge state at 1448 Da (182), species with a +2 charge state at 723 Da and phosphatidic acid (1812), species with +1 charge state at 695 Da) were monitored. Injections were performed in triplicate.

Nuclease activity assays. DNA and RNA oligonucleotides were ordered from Integrated DNA Technologies (IDT) or Dharmaco. TGGCCATTGCTAGAGGACUGCAGATA (50 nucleotides DNA), TATGCGTGACTGAAAGGACCCTT (25 nucleotides DNA), TCTAATGCTGATGTCGACTGCAA (25 nucleotides DNA), GGCUCUGAAGAGAGGACCACUCUCUCUAAUGAGAGUGUGAGCAGGACA (50 nucleotides RNA), UAAUGCCAGAUCAGAGGACACUCUCUCUAAUGAGAGUGUGAGCAGGACA (50 nucleotides RNA), UAAUGCCAGAUCAGAGGACACUCUCUCUAAUGAGAGUGUGAGCAGGACA (20 nucleotides RNA), UAGCUCUGAAGAGAGGACCACUCUCUCUAAUGAGAGUGUGAGCAGGACA (20 nucleotides RNA), CGGCGAAUCAGAGGACCACUCUCUCUAAUGAGAGUGUGAGCAGGACA (20 nucleotides RNA), GGCUCUGAAGAGAGGACCACUCUCUCUAAUGAGAGUGUGAGCAGGACA (20 nucleotides RNA), CCUGCACUGAAGAGAGGACCACUCUCUCUAAUGAGAGUGUGAGCAGGACA (20 nucleotides RNA)

Oligonucleotides were 5'-labelled with 32P γ-ATP (PerkinElmer) or ATP (Sigma–Aldrich) using T4 polynucleotide kinase (NEB) according to the manufacturer’s instructions. Free ATP was removed and the buffer was exchanged using illustrate MicroSpin G-25 Columns (GE Healthcare). Double-stranded substrates were annealed by mixing equimolar amounts of the complementary strands in water, denaturing at 95°C for 1 min, then cooling slowly from 65°C to 25°C in the presence of 30 mM NaCl, 10 mM MES, pH 6.5.

For enzymatic assays, mZuc was stored in 50% (v/v) glycerol at ~20°C. scPLD (EMD) was freshly prepared from the lyophilized stock according to the manufacturer’s protocol. To assess nuclease activity, substrates were incubated with recombinant protein in 50 mM MES, pH 6.5, 75 mM NaCl, 2 mM CaCl₂, 1 mM DTT for DNase activity, or 5 mM MES, pH 6.5, 75 mM NaCl, 2 mM CaCl₂, 1 mM DTT for RNase activity at 37°C for 3 h (Fig. 1c and Supplementary Fig. 3c) or 6 h (Figs 1e, 2 and Supplementary Fig. 2b). DNA and RNA substrates were used at a final concentration of 5 μM (Fig. 1c and Supplementary Fig. 3c), or 50 μM (Fig. 2). Recombinant mZuc dimer was used at a final concentration of 5 μM (Fig. 1f) or 50 μM (Fig. 2 and Supplementary Fig. 2b). When RNA was used as a substrate, RNasin Plus RNase Inhibitor (Promega) was added. Protein was removed from the cleavage reactions by proteinase K treatment (Roche) at a final concentration of 2 mg ml⁻¹ (in 50 mM Tris–HCl, pH 7.5, 75 mM NaCl, 6 mM EDTA, 1% (v/v) SDS) for 30 min. Nucleic acids were extracted using phenol/chloroform (Ambion), precipitated with sodium acetate/ethanol and separated by 15% urea–PAGE. Low molecular weight marker 10–100 nucleotides (Affilimetric) was labelled with 32P γ-ATP and used as a size ladder. Visualization was accomplished using a Storm PhosphorImager.

To assess the chemical properties of mZuc cleavage products, nucleic acids were recovered from the cleavage reactions using proteinase K and phenol/chloroform extraction followed by an additional chloroform extraction and sodium acetate/ethanol precipitation. For DNA products, terminal deoxynucleotidyl transferase (TdT) (NEB) was incubated with the extracted oligonucleotides in the presence of ddATP or dATP according to the manufacturer’s protocol. For RNA products, β-elimination was performed as previously described³³. Poly-A tailing was performed using Poly(A) tailing kit (Ambion) at 37°C for 15 min.

Detection of the phosphohistidine intermediate using 32P-disodium phosphate. mZuc wild-type and H153N (at a dimer concentration of 50 μM) were incubated with 0.4 mM 32P-disodium phosphate (PerkinElmer) in 50 mM MES, pH 6.5, 75 mM NaCl, 1 mM DTT at 37°C for 3 h. Proteins were separated by SDS–PAGE on a 10% NuPAGE gel (Innogen) and transferred to a nitrocellulose...
membrane. Proteins bound to the membrane were visualized by Poncense S stain-
ing (0.1% Poncense S, 5% acetic acid), and $^{32}$P was detected by phosphoimaging.

**Detection of the phosphohistidine intermediate using SRM-MS.** To confirm the findings of the $^{32}$PO$_4^{3-}$ labelling assay and further resolve the phosphorylation site, SRM-MS was used. mZuc wild-type/H153N heterodimer was incubated in 50 mM MES, pH 6.5, 0.15 M NaH$_2$PO$_4$, 1 mM DTT for 1 h at 4 °C. The protein (20 µg) was then denatured and reduced by adding 0.1% Protease-Max surfactant (Promega) and 5 mM TCEP in 200 mM triethylammonium bicarbonate, pH 10, and incubated at 55 °C for 20 min. The sample was then treated with 5 mM methyl methanethiosulphonate and proteolyzed with 2 µg Lys-C overnight at 37 °C. After digestion, 1 µg of the proteolytic fragments were applied to a home-packed C18 column. Peptides were chromatographed with a mobile phase gradient from 0% formic acid and 3% acetonitrile in water to 0.1% formic acid and 90% acetonitrile followed by in-line electrospray ionization mass spectrometry.

Mass spectra were collected on a Thermo Scientific TSQ Vantage mass spectrometer. In the first round of MS, m/z peaks corresponding to a Lys-C generated, His-153-containing peptide (AGIQVRHDQDLGYMHHK) were selected (+4 charge state masses of 522.0 Da and 502.0 Da for the phosphorylated and non-phosphorylated precursors, respectively). These precursors were then fragmented while monitoring for several y and b ion transitions.

**Fluorescence polarization.** The affinity of mZuc for ssDNA and ssRNA was assessed using fluorescence polarization. Fluorescecin-labeled 20-oligonucleotide probes (5'-ACAGAGTGGCCTTCTTCAAC–Fluorescin-3' DNA, Sigma; 5'-ACAGAGUCGCCACUUCAAC–Fluorescin-3' RNA, Pharmaco) at 10 nM were mixed with mZuc at various concentrations in a binding buffer of 0.1 M MES, pH 6.5, 75 mM NaCl, 1 mM DTT and incubated at room temperature (27 °C) for 10 min. Fluorescence polarization was then measured at room temperature (27 °C) on a BioTek Synergy4 microplate reader using excitation and emission wavelengths of 485 and 528 nm, respectively. Readings were performed in duplicate.

**Crystalization.** Crystals of mZuc were grown by hanging-drop vapour diffusion on siliconized glass coverslips (Hampton Research). Immediately before crystalization, the purified protein was mixed in a 1:100 (m/m) ratio with chymotrypsin. Two microlitres of the protein solution at 3 mg mL$^{-1}$ were then mixed with an equal volume of reservoir solution containing 50 mM Bis-Tris, pH 6.5, 18% PEG-3350, and 2% tascimate, pH 6.0, and suspended over 0.5 ml of reservoir solution. Rectangular prismatic crystals (~150 µm × 50 µm × 50 µm) grew at 4 °C within 24 h. Crystals were harvested and frozen in liquid nitrogen after serial transfer to 50 mM Bis-Tris, pH 6.5, 18% PEG-3350, and 2% tascimate, pH 6.0, and 20% ethylene glycol in increasing steps of 5% ethylene glycol. Tungstate-derivatized crystals were obtained by co-crystalization with 10 mM Na$_2$WO$_4$ added to the crystallization drop.

**Crystallographic data collection and structure determination.** Data were collected at 100 K using synchrotron radiation at the X29 beamline at the National Synchrotron Light Source at Brookhaven National Laboratory. Data were processed with XDS$^{39}$ and scaled with SCALA$^{35}$. Additional processing was performed with programs from the CCP4 suite$^{35}$. The crystals belong to the space group $P4_1_2_1_2$ with one molecule per asymmetric unit and a solvent content of ~35%. The structure was solved by molecular replacement with the program Phaser$^{40}$ using the structure of Nuc, a bacterial PLD-family nuclease (PDB accession 1BYR) as a search model$^{16}$. Automatic model building was performed using ARP/wARP$^{37}$, which correctly built approximately 140 residues of the 158 in the final structure. Subsequent model building and inspection were carried out using the program Coot$^{38}$. Model refinement was performed using REFMAC5$^{36}$. TLS refinement was carried out using one TLS group. The final native model consists of residues 35–58, 66–126 and 130–209 in addition to one coordinated Zn$^{2+}$ and 106 water molecules. The final models were validated by Molprobity$^{41}$. All residues in

the final structures were in the allowed regions of the Ramachandran plot with 152/158 residues being in favoured regions. Data collection and refinement statistics are listed in Supplementary Table 1.

In addition to native crystals, tungstate-derivatized crystals were also produced. These diffracted to similar resolution and had nearly identical unit cell parameters. Initial processing was performed as described earlier. The native structure was used as an isomorphous replacement search model and then refined accordingly. To avoid over-fitting, refinement of the tungstate derivative used the same $R_{free}$ set as the native. To confirm the identity of the Zn$^{2+}$, two data sets were collected from a single tungstate-derivatized crystal at ±100 eV from the Zinc K edge. Anomalous maps for each data set confirmed the identity of the metal. The root mean squared deviation for all protein atoms in the native and tungstate derivative was 0.49 Å. Crystallographic statistics for the tungstate-derivatized crystals are also presented in Supplementary Table 1.

**Figures.** Figures of molecular models were generated using PyMOL$^{42}$. Electrostatic surface calculations were performed with APBS$^{30}$ with a solvent ion concentration of 0.15 M using the PARSE force field. Superpositioning of structural homologues was performed by the DALI server$^{42}$. The refined mZuc structure was used as a template for manual model building of a short ssRNA using Coot$^{38}$. The initial ssRNA model placed the phosphate backbone along the observed positively charged active site groove (with the scissile phosphate positioned at the tungstate position in the derivative structure) while maintaining approximate two-fold symmetry of phosphates with respect to the homodimer (the RNA breaks the strict two-fold symmetry due to its polarity). Energy minimization was performed using GROMACS v.4.5.5$^{25}$ with the AMBER-99SB-ILDN force field$^{43}$ and particle-mesh Ewald (PME) long-range electrostatic modelling$^{42}$ with a conjugate gradient energy minimization algorithm, keeping the protein model (including Zn$^{2+}$) fixed while allowing minimization of the RNA molecule and solvent model. A nearly indistinguishable model was also generated using the steepest descent algorithm.

31. Bierniosek, C., Richmond, T. J. & Berger, I., MultiBac: multigene baculovirus-based eukaryotic protein complex production. Curr. Protoc. Protein Sci. Chapter 5, Unit 5.2 (2008).
32. Pozharski, E. V., McWilliams, L. & MacDonald, R. C. Relationship between turbidity of lipid vesicle suspensions and particle size. Anal. Biochem. 291, 158–162 (2001).
33. Vagin, V. V. et al. A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313, 320–324 (2006).
34. Kalbach, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. 26, 795–800 (1993).
35. Collaborative Computation Project, Number 4, The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D 50, 760–763 (1994).
36. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).
37. Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. ARP/wARP and molecular replacement. Acta Crystallogr. D 57, 1445–1450 (2001).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).
39. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D 53, 240–255 (1997).
40. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D 66, 12–21 (2010).
41. Delano, W. L., The PyMOL Molecular Graphics System. (Schrodinger, LLC, 2002).
42. Holm, L. & Rosenstrom, P. Dalil server: conservation mapping in 3D. Nucleic Acids Res. 38, W545–W549 (2010).
43. Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78, 1950–1958 (2010).
44. Essmann, U. et al. A smooth particle mesh Ewald method. J. Chem. Phys. 103, 8577–8593 (1995).