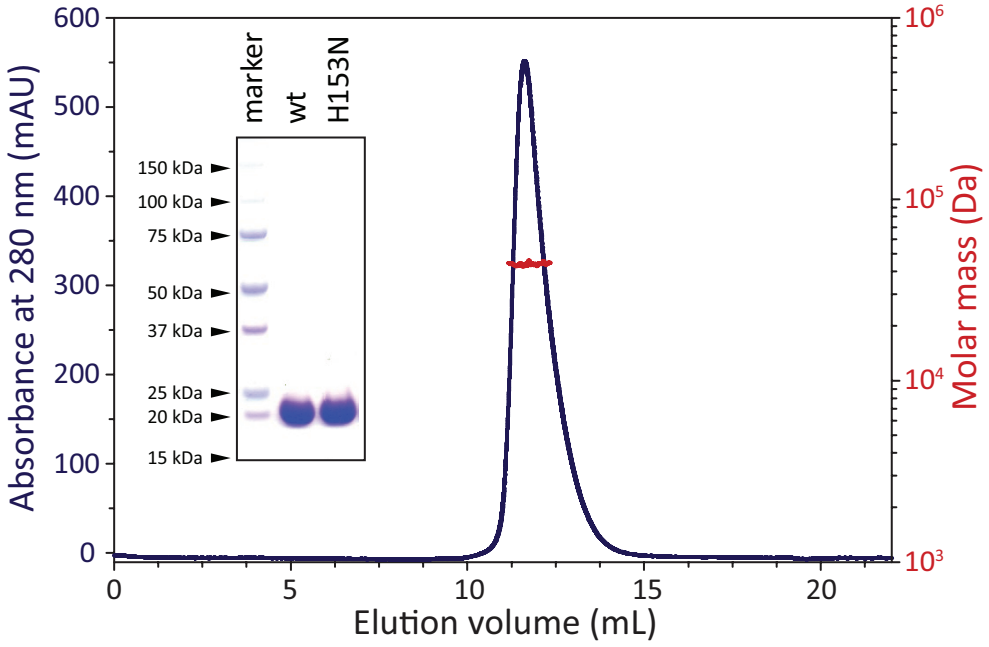
Supplementary Information
for

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

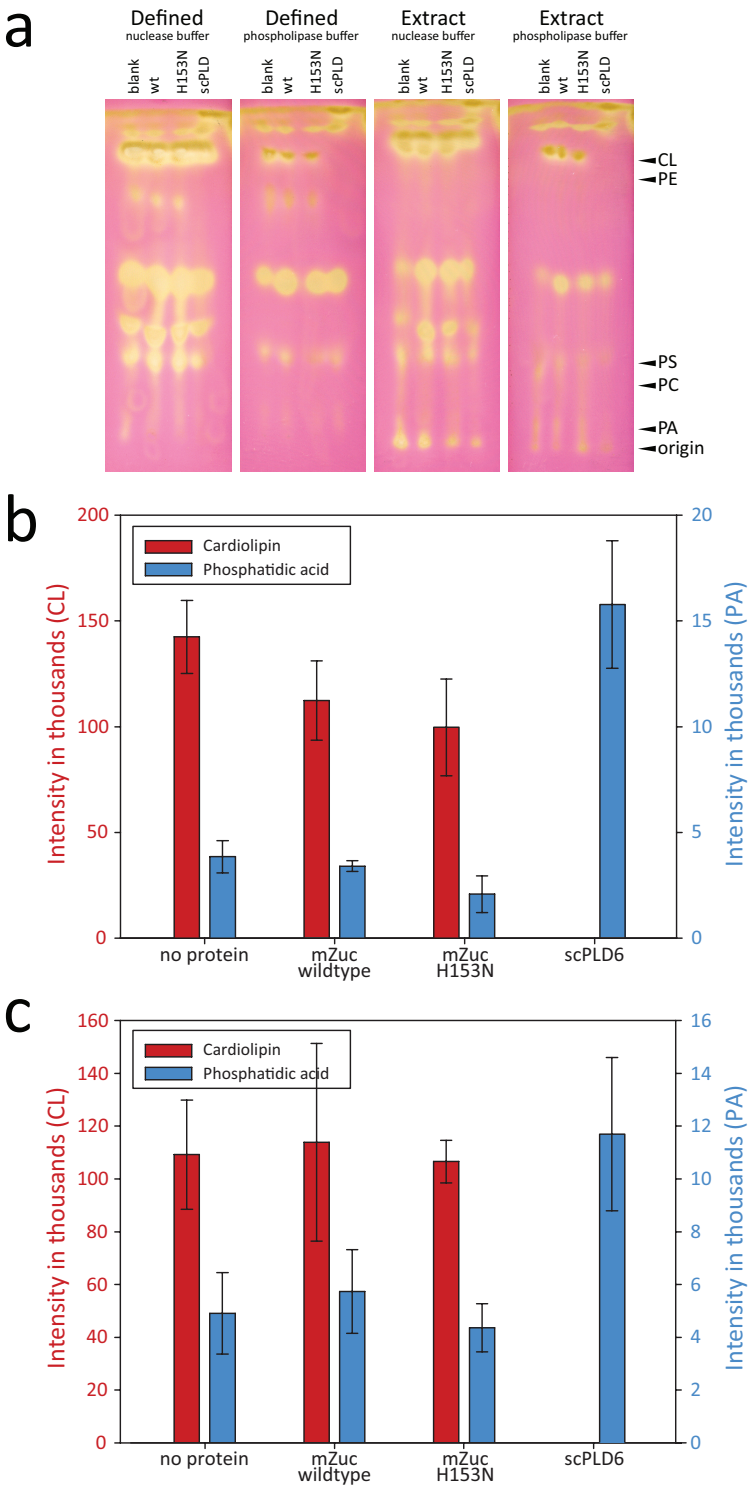
Jonathan J. Ipsaro[1,2]*, Astrid D. Haase[2]*, Simon R. Knott[2], Leemor Joshua-Tor[1,2&], and Gregory J. Hannon[2&]

[1] W. M. Keck Center for Structural Biology
[2] Howard Hughes Medical Institute
Cold Spring Harbor Laboratory
Cold Spring Harbor, NY 11724

*These authors contributed equally
& to whom correspondence should be addressed : hannon@cshl.edu; leemor@cshl.edu

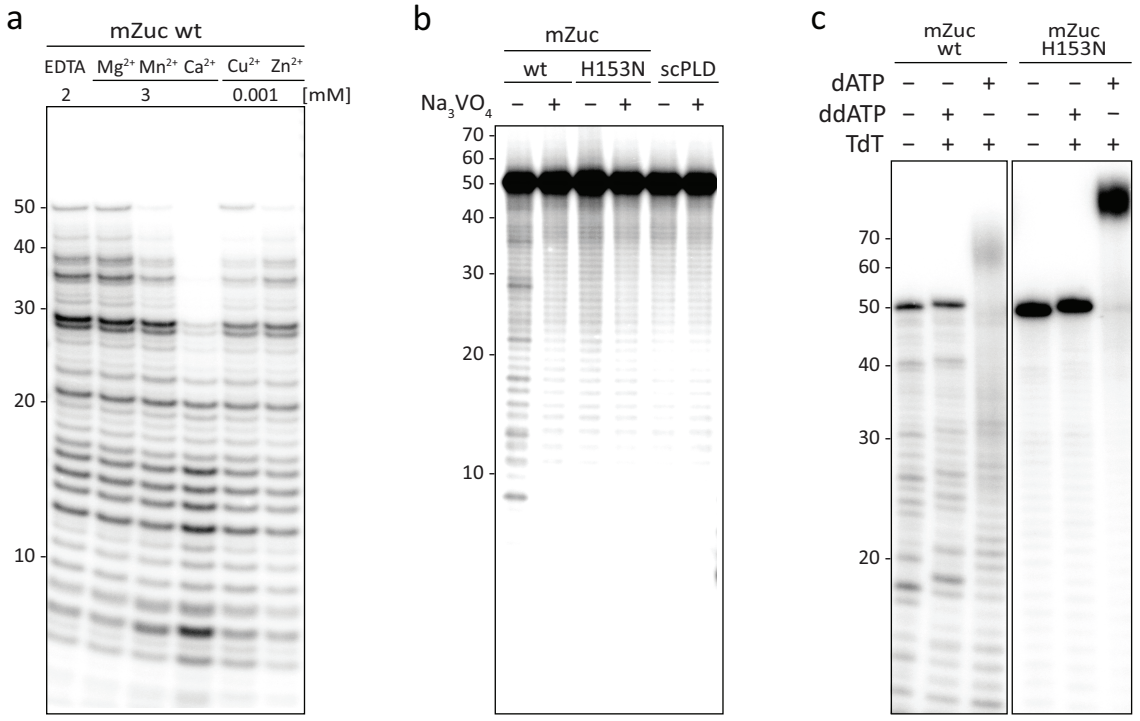


Supplementary Figure 1. Oligomeric state and purity of mZuc. Recombinant mZuc is dimeric as determined by gel filtration chromatography (blue, left axis) and multi-angle light scattering (red, right axis) and is highly purified as assessed by SDS-PAGE (inset).

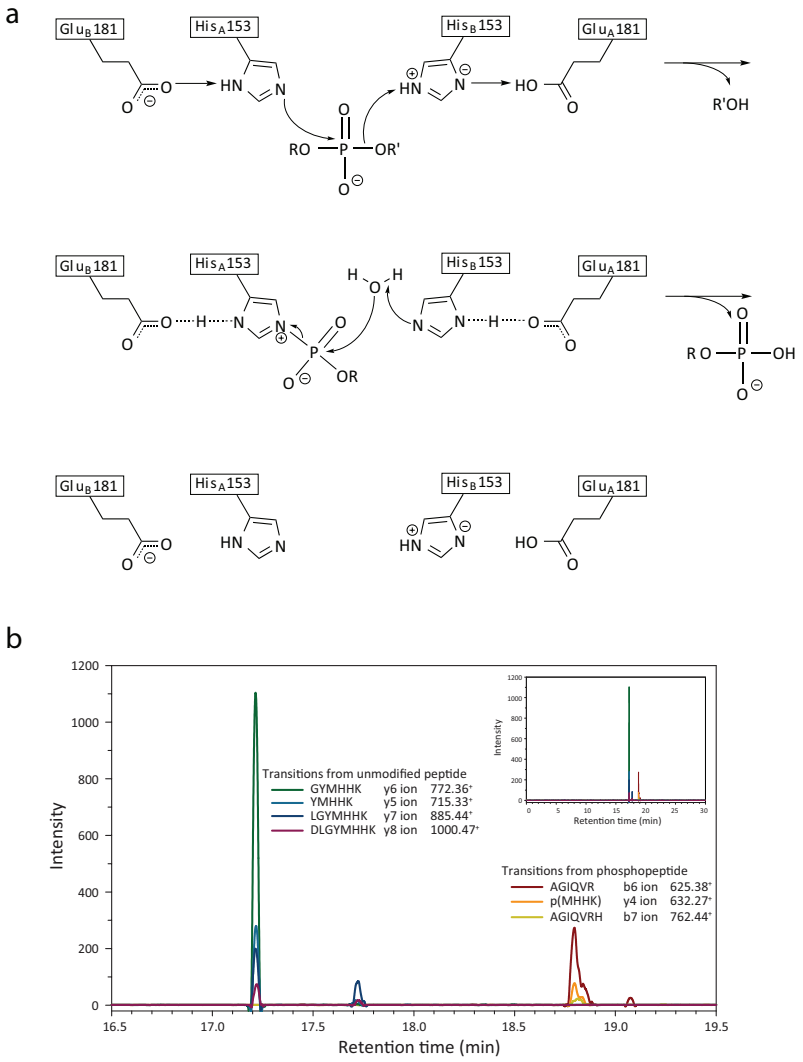


Supplementary Figure 2. Phospholipase activity was not detected for mZuc. a) A TLC-based assay was used to evaluate cardiolipin cleavage. Defined liposomes (PC:PE:PS:CL at 2:2:1:1) or extract-based liposomes (made with bovine heart lipid

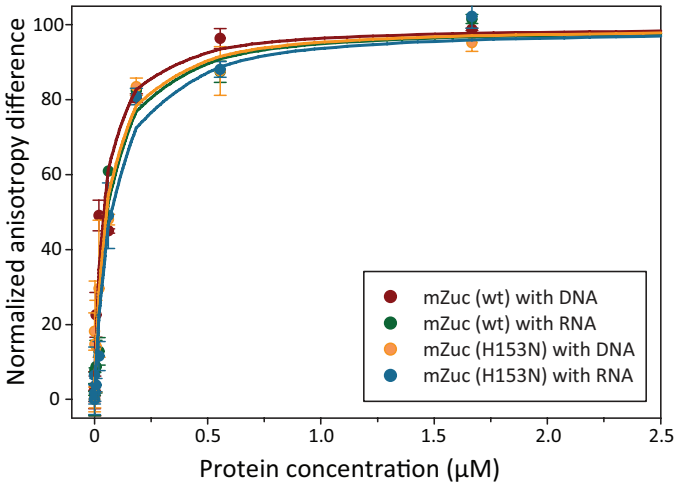
extract supplemented with CL) were incubated with purified mZuc under various conditions. In every case, wild-type mZuc and catalytically inactive mZuc H153N showed no detectable cleavage of cardiolipin or formation of PA. In contrast, a known phospholipase D from *S. chromofuscus* completely eliminated CL from each reaction. Lipid identification was based on standards run in parallel. Migration distances for the standards are indicated to the right of the TLC plates. **b-c)** SRM-MS was used to monitor the disappearance of CL and appearance of PA for defined liposome reactions under **(b)** phospholipase and **(c)** nuclease buffer conditions. No significant changes were observed for reactions containing mZuc/PLD6. The positive control showed complete hydrolysis of CL and a striking increase in the abundance of PA. Error bars indicate ± the standard deviation. *Note that panel **b** is duplicated from the main text but is reproduced here for the convenience of the reader.*



Supplementary Figure 3. mZuc shows single-strand (ss) endonuclease activity *in vitro*. a) mZuc catalysis does not require divalent cations as evidenced by ssDNase activity in the presence of EDTA. While divalents were not required, certain ions (Ca^{2+}, Mn^{2+}, and Zn^{2+}) enhanced the activity. **b)** mZuc cleaves ssRNA *in vitro*. As is the case for the ssDNase activity, addition of 4 mM Na_3VO_4 or the H153N mutation abolishes the ssRNase activity of mZuc. Phospholipase D from *Streptomyces chromofuscus* (scPLD) did not exhibit nuclease activity. Reactions were analyzed by Urea-PAGE (15%). **c)** mZuc releases DNA products with 3' OH termini. DNA fragments are extracted from mZuc cleavage reaction and incubated with Terminal deoxynucleotidyl Transferase (TdT) in the presence of ddATP or dATP (as indicated).



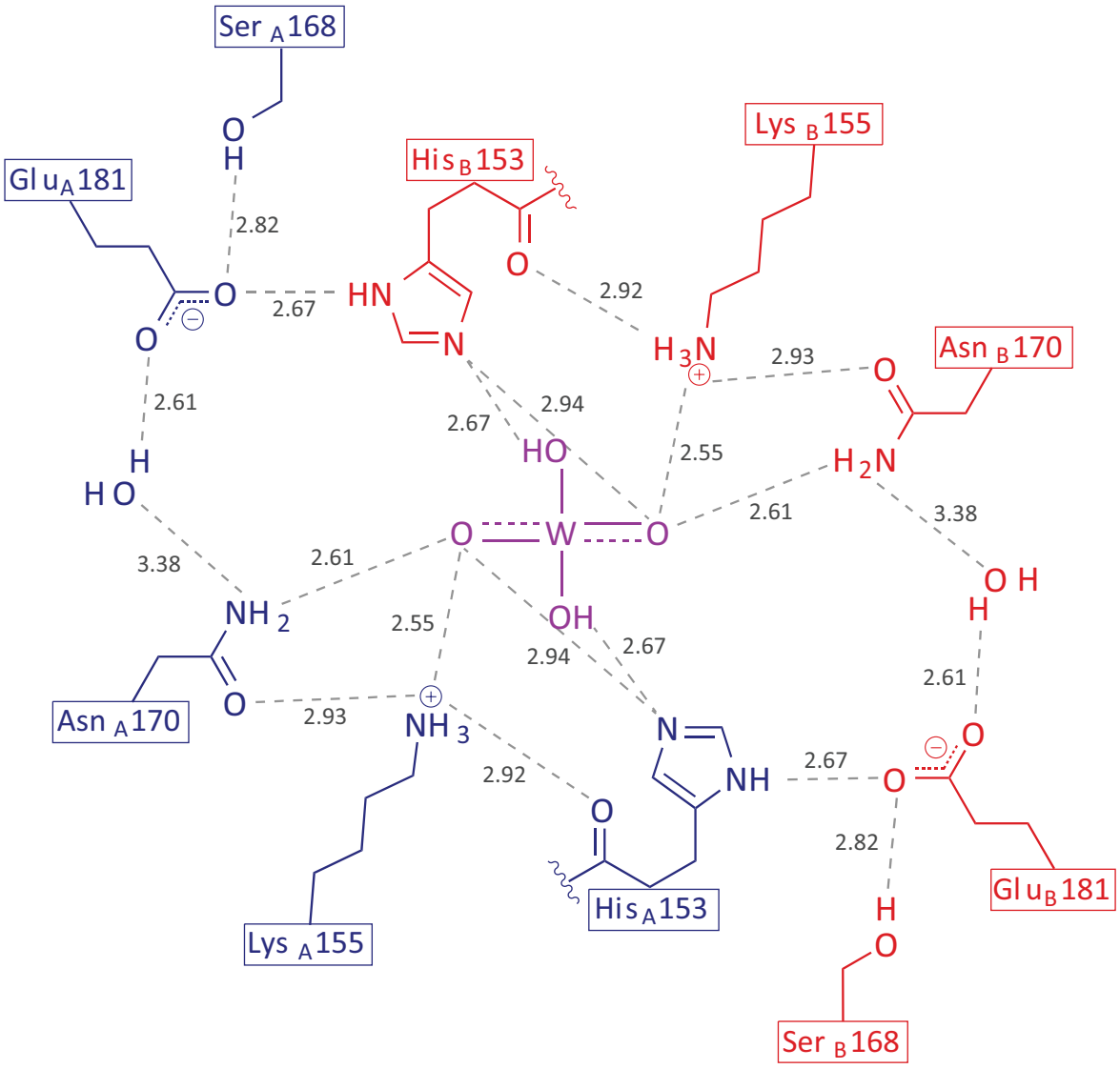
Supplementary Figure 4. mZuc proceeds through a phosphohistidine intermediate and binds nucleic acids directly. a) As proposed by Dixon and colleagues[15,18], the HKD phosphodiesterase mechanism consists of two distinct steps. In the first step, the lone pair of the imidazole nitrogen of His_A153 attacks the scissile phosphate leading to an S_N2 reaction and the formation of a covalent phosphohistidine intermediate. The leaving group then abstracts a proton from the opposing (protonated) His_B153. In the second step, a proton is abstracted from water by the deprotonated His_B153. The activated water then attacks the phosphohistidine intermediate resulting in product release. **b)** SRM-MS was used to confirm and further pinpoint the location of the phosphistidine intermediate. Chromatograms show the transition ion intensity for numerous fragments from the phosphorylated (red, orange, yellow traces) and unphosphorylated (green, blues, purple traces), +4 charge state, His 153-containing peptide precursors. Based on the ions observed, the location of the phosphorylation could be mapped to residues 152-154.



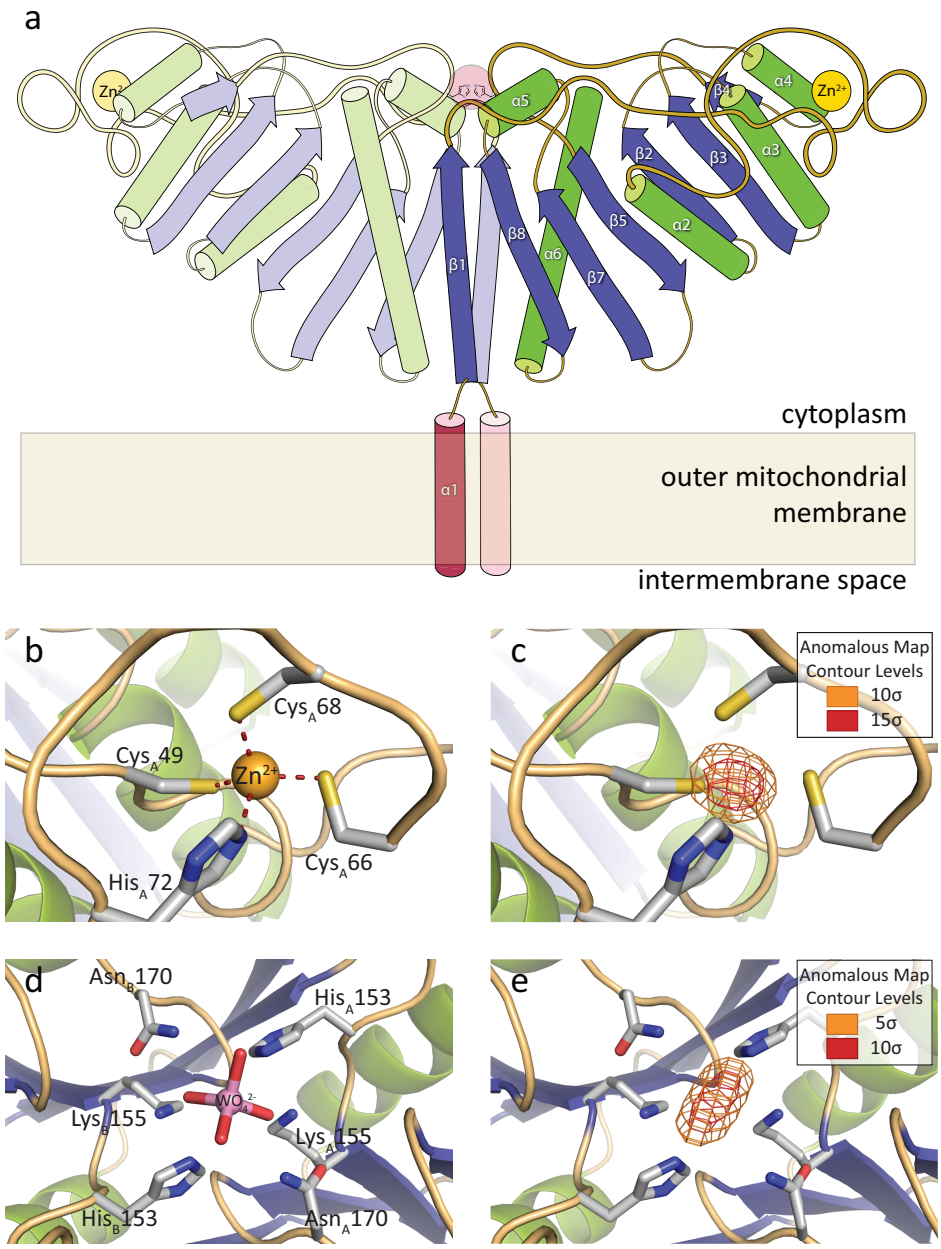
Calculated K_D values

mZuc (wt) with DNA	38.7 ± 10.4 nM
mZuc (wt) with RNA	55.7 ± 12.5 nM
mZuc (H153N) with DNA	51.2 ± 12.9 nM
mZuc (H143N) with DNA	70.5 ± 16.1 nM

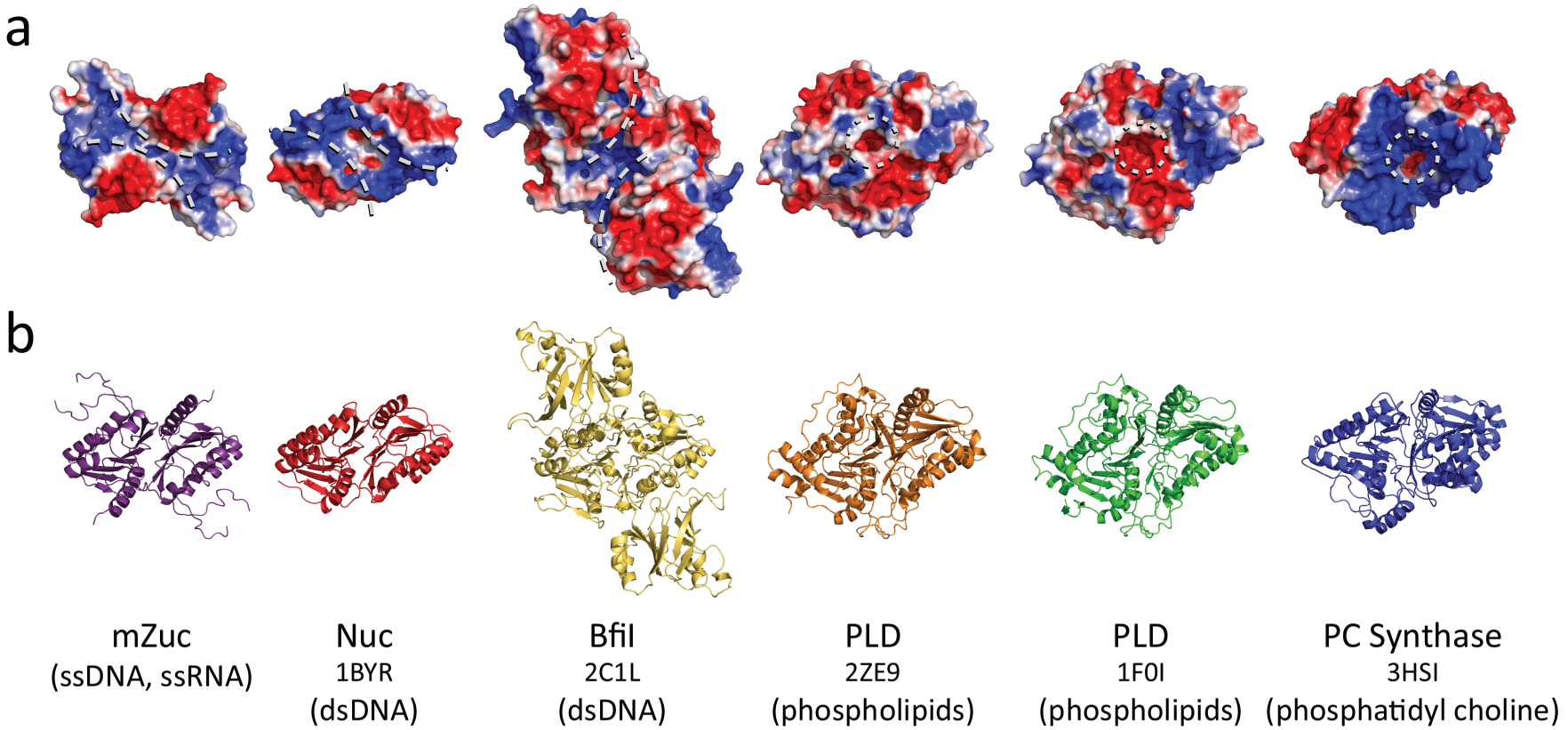
Supplementary Figure 5. mZuc binds ssDNA and ssRNA with comparable affinity. Binding affinity measurements for mZuc (wildtype and H153N mutant) with both ssDNA and ssRNA were measured using fluorescence polarization. The affinity in each case is roughly 50 nM.



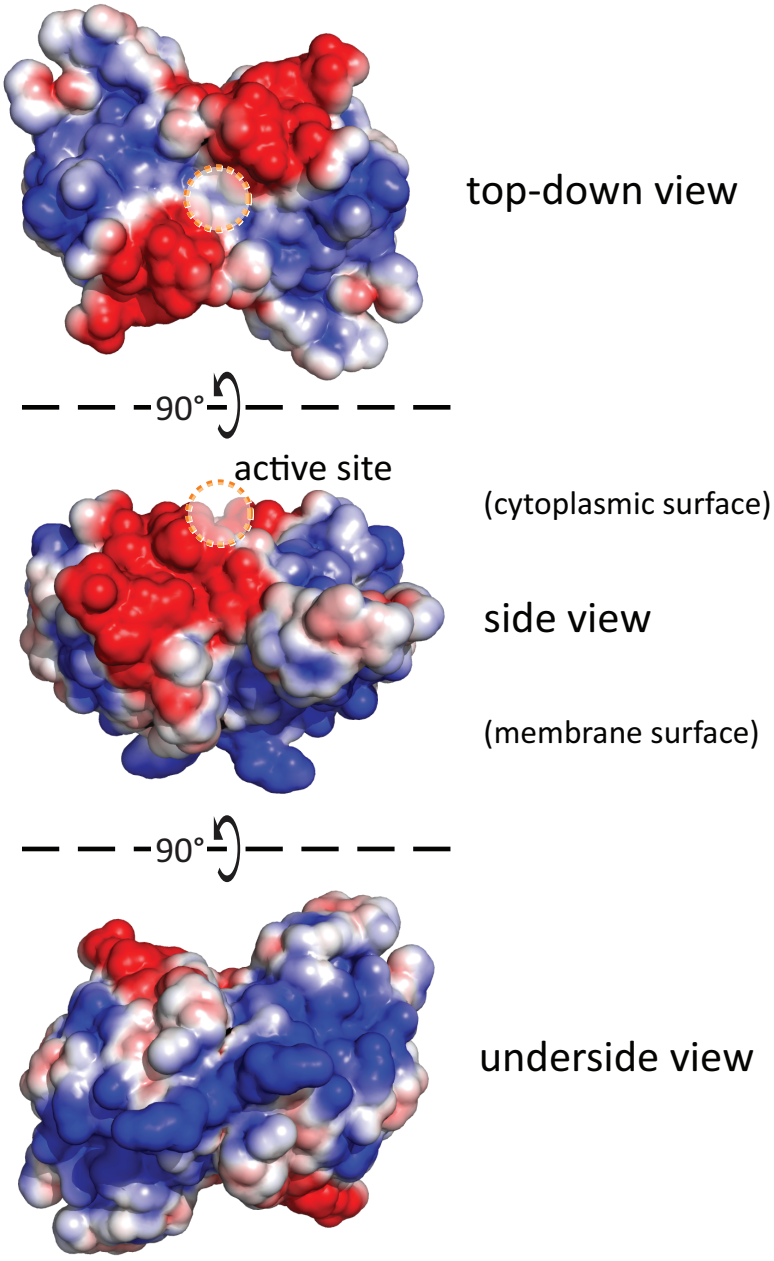
Supplementary Figure 6. The hydrogen bond network of mZuc. Similar to that observed in the structure of Nuc[18], mZuc has an extensive active site hydrogen bonding network which spans the dimerization interface. Side chains for monomer A are in red, side chains for monomer B are in blue, tungstate is in purple, and hydrogen bonds are indicated as dashed grey lines. Distances are expressed in Å and indicate the separation of the non-hydrogen nuclei. The distance from N_ε of His153 to the tungsten atom is 3.0 Å.



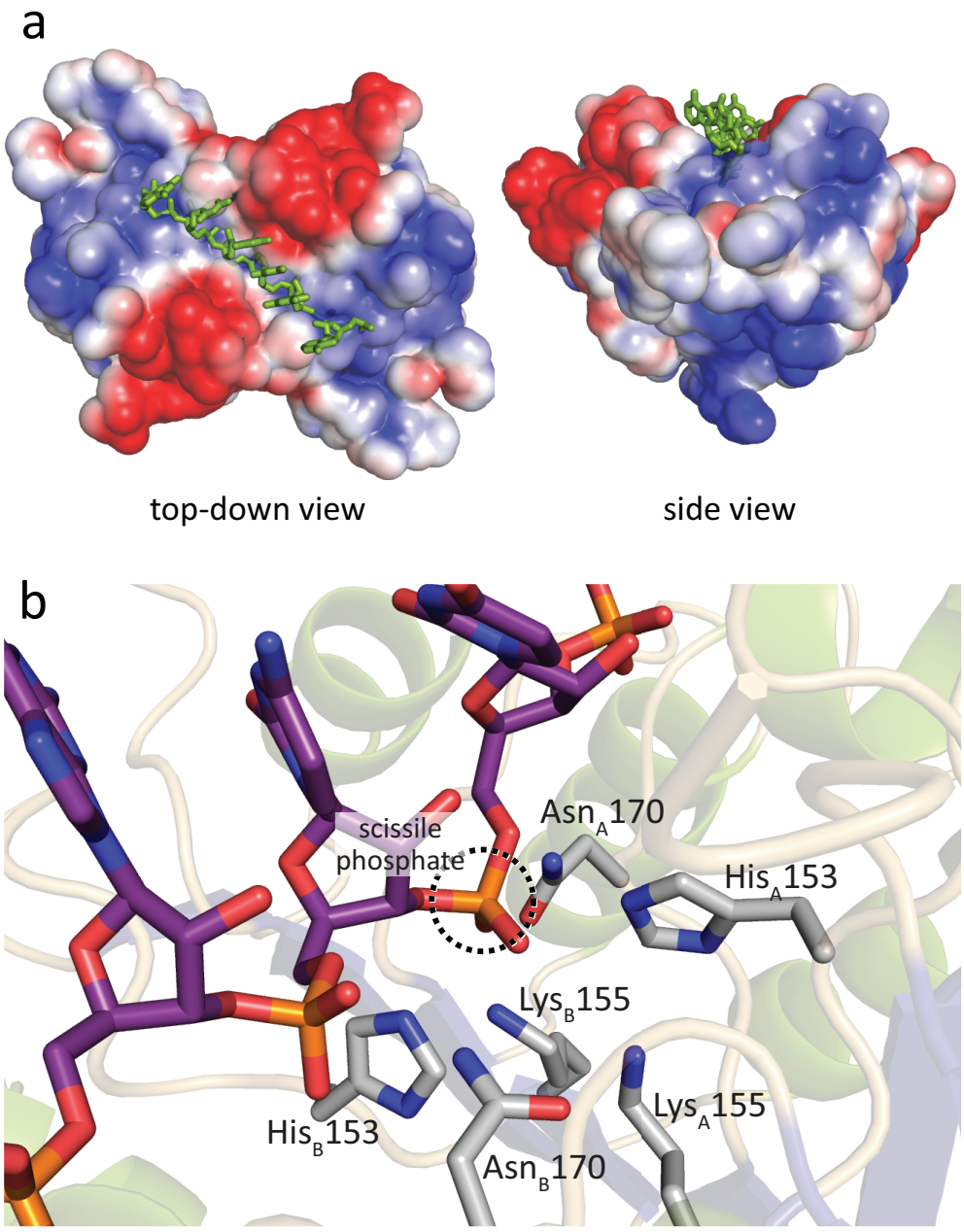
Supplementary Figure 7. Coordination geometries of bound ligands. a) Stylized drawing of the mZuc structure. The active site is formed by the dimerization interface with the zinc wings extending away from the center of the homodimer. **b)** An unexpected CCCH zinc wing was found in the protein consisting of residues Cys49, Cys66, Cys68, and His72. **c)** The identity of the Zn^{2+} was confirmed by calculating anomalous difference maps for datasets collected from the same crystal above (contoured at 10 and 15σ) and below the Zn^{2+} K edge (no signal detectable above the noise). **d)** mZuc co-crystallized with tungstate bound tungstate exclusively in the active site. **e)** The anomalous difference map (contoured at 5 and 10σ) indicates the presence an anomalous scatterer in the active site only for the tunstate-containing co-crystals.



Supplementary Figure 8. Comparative analysis of PLD family members with HKD motifs. a) Electrostatic surfaces of an array of PLD family members in the same orientation illustrate the utilization of specific structural contexts to convey substrate specificity. In the case of the nucleic acid phosphodiesterases (PDB IDs: 1BYR[18], 2C1L[31], and this structure), a long, positively-charged groove runs through the active site. For the phospholipases (PDB IDs: 2ZE9, 1F0I[32], 3HSI), a deep pocket accommodates the substrate. Each surface depicts the solvent-accessible surface contoured at ± 2 k_BT/e. Surfaces were calculated using APBS[30] with a solvent ion concentration of 0.15 M. **b)** Ribbon diagrams of each molecule in the same orientation as in (**a**). Note that the overall fold and position of the active site remains consistent in each case, with additional structural elements decorating the core.



Supplementary Figure 9. Various views of the mZuc electrostatic surface. In addition to the positively-charged groove which spans the active site, two negative patches are present on the top surface of the dimer. Another positively charged patch is also present on the underside of the protein, which may serve to strengthen the attachment of the protein to the negatively-charged phospholipid head groups on the surfaces of the underlying mitochondrial membrane. Each surface depicts the solvent-accessible surface contoured at ± 2 k_BT/e. Surfaces were calculated using APBS[30] with a solvent ion concentration of 0.15 M. The active site region is indicated with a dashed circle in the top-down and side views.



Supplementary Figure 10. Views of RNA in the active site groove. a) Using the structure of the mZuc, a short RNA molecule was manually built into the model then subjected to energy minimization using GROMACS[25]. The minimized model (RNA in green) shows the phosphates of the RNA backbone positioned in the most positively charged areas of the groove with the bases extending away from the dimer core. **b)** A close-up view of the active site places the RNA (purple) in the active site groove with the scissile phosphate (circled) situated directly between the catalytic histidines. Conserved active site residues are shown as sticks (grey) overlaid on the ribbon diagram (colored as in Fig. 3, faded).

Supplementary Table 1. Data collection and refinement statistics

	mZuc	mZuc + Na_2WO_4	
Data collection			
Space group	$P4_32_12$	$P4_32_12$	
Cell dimensions			
a, *b*, *c* (Å)	38.7, 38.7, 214.5	38.9, 38.9, 213.3	
α, *β*, *γ* (°)	90, 90, 90	90, 90, 90	
		Above Zn^{2+} peak	*Below Zn^{2+} edge*
Wavelength	1.0750 Å	1.2716 Å	1.2983 Å
Resolution (Å)	40.0-1.75 (1.85-1.75)	40.0-2.10 (2.23-2.10)	40.0-2.20 (2.32-2.20)
R_{sym} or R_{merge}	4.8 (27.8)	5.9 (29.7)	5.7 (32.1)
$I/\sigma I$	21.3 (5.9)	24.0 (7.5)	24.8 (7.1)
Completeness (%)	98.1 (98.8)	99.7 (98.5)	99.9 (100.0)
Multiplicity	6.8 (6.9)	10.2 (10.5)	10.2 (10.6)
Refinement			
Resolution (Å)	40.0-1.75 (1.80-1.75)	40.0-2.10 (2.16-2.10)	
No. reflections	16154 / 862 (1143 / 65)	9676 / 514 (565 / 35)	
R_{work}/ R_{free}	17.5 / 19.7 (24.9 / 25.2)	20.9 / 25.1 (22.6 / 29.4)	
No. atoms			
Protein	1336	1316	
Ligand/ion	1 (Zn^{2+})	6 (1 Zn^{2+}; 1 WO_4^{2-})	
Water	105	35	
B-factors**			
Protein	28.3	41.6	
Ligand/ion	33.4 (Zn^{2+})	64.1 (Zn^{2+}) 49.3 (WO_4^{2-})	
Water	23.1	24.5	
R.m.s deviations			
Bond lengths (Å)	0.011	0.013	
Bond angles (°)	1.281	1.382	

*Highest resolution shell is shown in parenthesis.
**The TLS contribution is included in the presented average B-factors.

Supplementary Literature Cited

31 Grazulis, S. *et al.* Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. *Proc Natl Acad Sci U S A* **102**, 15797-15802, doi:10.1073/pnas.0507949102 (2005).

32 Leiros, I., Secundo, F., Zambonelli, C., Servi, S. & Hough, E. The first crystal structure of a phospholipase D. *Structure* **8**, 655-667 (2000).

33 Bieniossek, C., Richmond, T. J. & Berger, I. MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Curr Protoc Protein Sci* **Chapter 5**, Unit 5 20, doi:10.1002/0471140864.ps0520s51 (2008).

34 Pozharski, E. V., McWilliams, L. & MacDonald, R. C. Relationship between turbidity of lipid vesicle suspensions and particle size. *Anal Biochem* **291**, 158-162, doi:10.1006/abio.2001.5012 (2001).

35 Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320-324, doi:10.1126/science.1129333 (2006).

36 Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J Appl Crystallogr* **26**, 795-800 (1993).

37 Collaborative. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763, doi:10.1107/S0907444994003112 (1994).

38 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
39 Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr* **57**, 1445-1450 (2001).
40 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
41 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255, doi:10.1107/S0907444996012255 (1997).
42 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).
43 DeLano, W. L. The PyMOL Molecular Graphics System. (2002).
44 Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549, doi:10.1093/nar/gkq366 (2010).
45 Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958, doi:10.1002/prot.22711 (2010).
46 Essmann, U. *et al.* A Smooth Particle Mesh Ewald Method. *J Chem Phys* **103**, 8577-8593 (1995).