Crystallographic Analysis of the Reaction Cycle of 2′,3′-Cyclic Nucleotide 3′-Phosphodiesterase, a Unique Member of the 2H Phosphoesterase Family

Matti Myllykoski[1,2]†, **Arne Raasakka**[1,2,3]†, **Mari Lehtimäki**[1], **Huijong Han**[1,2,3], **Inari Kursula**[1,3] and **Petri Kursula**[1,2,3,4]

1 - *Department of Biochemistry,* University of Oulu, FIN-90014 Oulu, Finland

2 - *Biocenter Oulu,* University of Oulu, FIN-90014 Oulu, Finland

3 - *Centre for Structural Systems Biology,* Helmholtz Centre for Infection Research, German Electron Synchrotron, D-22603 Hamburg, Germany

4 - *Department of Chemistry,* University of Hamburg, D-22603 Hamburg, Germany

Correspondence to Petri Kursula: Centre for Structural Systems Biology, Helmholtz Centre for Infection Research, German Electron Synchrotron, Notkestrasse 85, D-22603 Hamburg, Germany. petri.kursula@oulu.fi
http://dx.doi.org/10.1016/j.jmb.2013.06.012
Edited by M. Guss

Abstract

2H phosphoesterases catalyze reactions on nucleotide substrates and contain two conserved histidine residues in the active site. Very limited information is currently available on the details of the active site and substrate/product binding during the catalytic cycle of these enzymes. We performed a comprehensive X-ray crystallographic study of mouse 2′,3′-cyclic nucleotide 3′-phosphodiesterase (CNPase), a membrane-associated enzyme present at high levels in the tetrapod myelin sheath. We determined crystal structures of the CNPase phosphodiesterase domain complexed with substrate, product, and phosphorothioate analogues. The data provide detailed information on the CNPase reaction mechanism, including substrate binding mode and coordination of the nucleophilic water molecule. Linked to the reaction, an open/close motion of the β5—α7 loop is observed. The role of the N terminus of helix α7—unique for CNPase in the 2H family—during the reaction indicates that 2H phosphoesterases differ in their respective reaction mechanisms despite the conserved catalytic residues. Furthermore, based on small-angle X-ray scattering, we present a model for the full-length enzyme, indicating that the two domains of CNPase form an elongated molecule. Finally, based on our structural data and a comprehensive bioinformatics study, we discuss the conservation of CNPase in various organisms.

Introduction

The 2H phosphoesterase superfamily is characterized by the enzymatic cleavage of 2′,3′-cyclic nucleotides to 2′-nucleotides and the presence of two conserved HxT/Sx amino acid motifs (where x denotes a hydrophobic residue) in the active site [1–3]. Outside the active site, sequence conservation between different clades of the superfamily is low, although the three-dimensional fold is similar [2,4]. Structures of some 2H superfamily proteins, including archaeal 2'-5' RNA ligases [5–8], the human protein kinase A anchoring protein 18 central domain [9], *Bacillus subtilis* YjcG [10], and a hypothetical protein ATU0111 from *Agrobacterium tumefaciens*

[Protein Data Bank (PDB) ID 2fsq], have been reported. Also, structures of plant cyclic phosphodiesterases (CPDases), as well as catalytic 2H domains from mammalian 2′,3′-cyclic nucleotide 3′-phosphodiesterase (CNPase) and the regeneration-induced CNPase homologue (RICH) from fish, have been determined [4,11–14]. All the structures present two lobes, between which the active site is located. The 2H superfamily was initially discovered based on the 3′-phosphodiesterase activity of homologous enzymes toward 2′,3′-cyclic nucleotides. Of all the 2H phosphoesterases identified by primary sequence analysis, only myelin CNPase [15], RICH [16], yeast CPDase [3], yeast tRNA ligase [17], plant CPDase [18], and the *Pyrococcus furiosus* and

Escherichia coli 2'-5' RNA ligases [6] have been characterized to contain this activity, while protein kinase A anchoring protein 18 lacks it [9].

CNPase is expressed at very high levels in the non-compact regions of central nervous system myelin [19,20]. The myelin sheath is a lipid-rich, multilayered membrane structure that enables the rapid transmission of electrical impulses in vertebrate axons. The C-terminal catalytic 2H phosphoesterase domain of CNPase hydrolyzes 2′,3′-cyclic nucleotides to 2′-nucleotides, and the active site can accommodate polynucleotide substrates extended toward the 5′-end, such as spliced tRNA halves [15,21] or 2′,3′-cyclic NADP$^+$ (2′,3′-cNADP$^+$) [22]. The CNPase N-terminal domain is homologous to polynucleotide kinases (PNKs) [23], although the sequence similarity is low. CNPase interacts with the cytoskeleton and cell membranes [24–28], RNA [29], and calmodulin [30]. A genetic knockout of CNPase results in axonal degeneration, incorrect local structures in myelin, and an increase in 2′,3′-cyclic AMP (2′,3′-cAMP) in the brain after injury [31–34].

The overall reaction catalyzed by CNPase is depicted in Fig. 1. Based on structural data, mutation studies, and the well-known catalytic mechanism of RNase A [35], a mechanism for the CNPase phosphodiesterase activity has been suggested [4,11,14]. Briefly, His309 activates a water molecule, which then nucleophilically attacks the cyclic phosphate and forms a pentavalent intermediate. The 3′-oxygen is protonated by His230, and the 2′-phospho product is released. Several active-site residues and conserved water molecules are important for the reaction, and hydrogen bonds to the HxTx motifs and conserved water molecules coordinate the ligand throughout the reaction [4]. The catalytic residues of the two HxTx motifs and four water molecules between them arrange in a 2-fold pseudosymmetry. Thus, the reason for CNPase stereospecificity, involving the coordination and correct orientation of both the substrate and the nucleophilic water molecule, must therefore lie in the active-site vicinity.

Earlier, we reported the structure of the mouse CNPase catalytic domain complexed with the reaction product 2′-AMP [4], but no structural data of a 2H superfamily protein complexed with substrate exist. Here, we report the first such structure, in which a 2′,3′-cAMP molecule and the nucleophilic water are trapped in the active site of an inactive CNPase mutant. The data are complemented by product complexes and the use of phosphorothioate analogues of both substrate and product. We also constructed a model of mouse full-length CNPase in solution and analyzed the sequence conservation of CNPase in currently available animal protein sequences. Our results allow us to visualize details of the binding of substrate and product into the active site and to propose a structure-based mechanism of the CNPase enzymatic reaction.

Results

Active-site mutant variants of mouse CNPase

Our earlier work provided the crystal structure of the mouse CNPase catalytic domain, including a complex with the reaction product 2′-AMP [4]. The main components of the CNPase active site include the two HxTx motifs on β strands β2 and β5, between which lie four water molecules, forming the bottom of the active site. A number of additional residues in the vicinity are involved in binding the ribose and base moieties of the substrate. These interactions must be crucial to the stereospecificity since the core active site is 2-fold pseudosymmetric. A fifth water molecule acts as the nucleophile. The earlier liganded CNPase catalytic domain structures were very difficult to obtain and had low ligand occupancies [4]. Hence, we used mutated variants of mouse CNPase to obtain high-occupancy complexes of the CNPase catalytic domain with ligands, including the reaction substrate, bound to the active site. Instrumental for success was the use of acetate or formate buffers instead of citrate, which apparently binds to the active site tightly, preventing ligand co-crystallization and soaking (unpublished results).

For the purpose of obtaining liganded structures of the CNPase catalytic domain, we prepared mutant variants of the active-site histidines and Val321. While His309 activates the nucleophilic water molecule and His230 protonates the leaving 3′-OH group, Val321 stacks against the substrate adenine ring. The enzymatic activities of wild-type and mutant CNPase catalytic domains were determined using 2′,3′-cNADP$^+$ as substrate (Table 1). The mutation V321A did not affect k_{cat} significantly but doubled K_m, indicating a functional active site but slightly compromised substrate binding. The mutation of the active-site histidines 230 and 309 to either glutamine or serine essentially abolished activity in the assay, as expected.

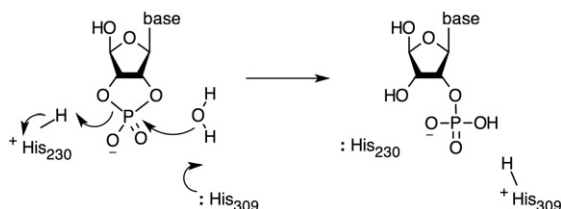The correct folding of all the mutant CNPase catalytic domains in solution was confirmed to be

Fig. 1. The overall reaction catalyzed by CNPase. Through a nucleophilic attack carried out by an activated water molecule, a 2′,3′-cyclic nucleotide is converted into a 2′-nucleotide.

Table 1. Activity and thermal stability of the CNPase catalytic domain variants

	k_{cat} (s^{-1})	K_M (µM)	k_{cat}/K_M (s^{-1} µM^{-1})	T_m (°C)
Wild type	940 ± 38	553 ± 46	1.70	58.0
H230Q	24 ± 15	1055 ± 1130	0.02	55.0
H309Q	—	—	—	52.7
H230S	14 ± 14	1305 ± 2004	0.01	56.0
H309S	21 ± 16	1192 ± 1468	0.02	57.3
H230Q/H309Q	—	—	—	49.7
V321A	732 ± 33	1045 ± 77	0.70	59.5

For all the active-site histidine mutants, the measured activities were so low that either the constants could not be calculated or their associated errors are very high.

indistinguishable from the wild-type protein by both synchrotron radiation circular dichroism spectroscopy (SRCD) and small-angle X-ray scattering (SAXS) (Supplementary Fig. 1). Thermal stability assays (Table 1) indicated that mutations of His230 and His309 lowered stability, especially the mutations to glutamine. The V321A variant, on the other hand, presented slightly increased thermal stability.

The first substrate complex for the 2H family

In order to obtain crystal structures of different steps in the catalytic cycle, we crystallized mutant variants of the catalytic domain of mouse CNPase in the presence of 2′,3′-cAMP, 2′,3′-cNADP$^+$, or 2′,3′-cAMP phosphorothioate analogues (Table 2 and Supplementary Fig. 2). The 2′,3′-cAMP substrate was co-crystallized with the CNPase H309S mutant, and the electron density clearly shows a well-defined substrate molecule in the active site of the mutant enzyme (Fig. 2 and Supplementary Fig. 2). Thus, for the first time, we have here been able to trap a substrate complex of an enzyme in the 2H family.

The substrate is tightly bound in a groove containing the active site. The interactions of the substrate with the active site include those made by the phosphate group, the ribose sugar, and the adenine base. The base is stacked between Phe235 (at the C terminus of strand β2) and Val321 (at the N terminus of helix α7), which forms CH-π hydrogen bonds to the aromatic rings. The main interaction of the ribose moiety with the protein is also a CH-π-type interaction between the CH group at ribose C-4 and the side chain of Tyr168. Three of the four oxygen atoms of the cyclic phosphate moiety interact with the catalytic residues. The two free oxygen atoms form hydrogen bonds with the two threonine residues and two central water molecules of the active site, and the O3′ atom is coordinated by His230. This observation highlights the roles of Thr232 and Thr311, plus the conserved water structure, in positioning the substrate for catalysis.

The nucleophilic water molecule is also present in the structure, and its distance from the phosphorus

atom of the substrate is 3.7 Å (Fig. 2b). This water molecule is coordinated under the closed β5–α7 loop by the backbone NH groups of Thr323 and Gly324 at the N terminus of helix α7, the carbonyl group of Pro320, the extra water molecule replacing the side chain of His309, and another water molecule further H-bonded to Arg307.

A molecule takes the place of the His309 side chain in the H309S mutant structure, being H-bonded to the mutated Ser309, as well as to the substrate; hence, hydrogen bonding contacts for the substrate in the active site are preserved, although activity is lost. In fact, this water superimposes with the N^{ε2} atom of His309 in wild-type CNPase, which is likely to coordinate the substrate in the same way. In this respect, it is surprising that none of the His-Gln mutants could be crystallized with bound nucleotide ligands (unpublished results).

Binding mode of the reaction product

As described above, the substrate was visible in the inactive H309S mutant; however, the enzymatically active V321A variant crystallized in the presence of 2′,3′-cAMP contains the reaction product 2′-AMP (Fig. 3a). We previously determined the structure of wild-type CNPase bound to the product at 2.4 Å resolution [4]. The new structure of the V321A mutant-2′-AMP complex is of much higher resolution and has a high occupancy of the product, making it more suitable for detailed analyses.

The product sits in the active site with its 3′-OH group H-bonded to His230, and the 2′-phosphate extended toward His309 and Pro320. The oxygen atoms of the product interact with the active-site water molecules and the HxTx motifs. In contrast to the substrate complex, the flexible region 319–321 in loop β5–α7 is in an open conformation. Surprisingly, the 2′-phosphate of the product interacts directly with the N terminus of helix α7 as a result of this conformational change, possibly taking advantage of the positive helix dipole, which is hidden when the loop is closed. The amide group of Ala321 also coordinates the phosphate, and the carbonyl of Pro320 has rotated away. A superposition of the substrate and product complexes indicates that it is mainly the phosphate group, which moves during the reaction, while a binding pocket for the product is created by the opening of the β5–α7 loop (Fig. 3b and c and Supplementary Fig. 3).

Phosphorothioate complexes

To get an even better insight into the active-site events during the catalytic cycle, we used two phosphorothioate analogues of 2′,3′-cAMP, the Sp and Rp epimers of 2′,3′-cyclic adenosine monophosphorothioate (AMPS), as mimics of the substrate and product. In these analogues, one of the cyclic

Table 2. Crystallographic data processing and refinement statistics

	Dataset								
Mutation	None	None	H309S	H309S	H230S	H230S	V321A	V321A	V321A
Active-site ligand	2′-AMPS (was Sp)	2′-AMPS (was Rp)	2′,3′-Sp-cAMPS	2′,3′-cAMP	$NADP^+$	2′,3′-Rp-cAMPS	2′-AMP	2′-AMPS (was Sp)	2′-AMPS (was Rp)
PDB entry	2yoz	2yp0	2ypc	2ype	3zbr	2yph	2yq9	3zbs	3zbz
Data collection and processing									
Beamline	EMBL/ DESY X11	EMBL/ DESY X11	EMBL/ DESY P14	EMBL/ DESY X11	MAX-Lab I911-2	MAX-Lab I911-2	MAX-Lab I911-2	MAX-Lab I911-2	MAX-Lab I911-2
Wavelength (Å)	0.815	0.815	1.223	0.815	1.041	1.041	1.041	1.041	1.041
Unit cell parameters									
a (Å)	38.40	39.83	40.28	40.80	42.10	40.63	41.55	41.36	41.25
b (Å)	47.88	47.74	47.49	47.13	47.10	47.53	46.79	46.88	46.99
c (Å)	51.15	50.93	53.76	53.78	111.17	107.55	107.36	107.23	107.38
α (°)	90	90	90	90	90	90	90	90	90
β (°)	94.45	97.86	94.89	94.28	90.26	90	90	90	90
γ (°)	90	90	90	90	90	90	90	90	90
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Resolution range (Å) *	19.1–2.1 (2.15–2.10)	19.4–2.3 (2.36–2.30)	40.1–1.89 (1.94–1.89)	10.0–1.9 (1.95–1.90)	30–2.3 (2.36–2.30)	39–2.1 (2.15–2.10)	30–1.9 (1.95–1.90)	30–2.45 (2.51–2.45)	30–2.1 (2.15–2.10)
Observed/ unique reflections	39,109/ 10,896	43,123/ 8391	58,311/ 15,833	66,817/ 16,037	66,246/ 19,310	72,881/ 12,366	82,177/ 16,784	53,246/ 8049	59,100/ 12,529
R_{sym} (%)	8.5 (52.1)	12.6 (68.9)	3.8 (27.6)	9.8 (101.3)	9.2 (114.4)	9.1 (84.5)	6.9 (67.4)	13.7 (97.6)	12.9 (87.4)
$\langle I/\sigma I \rangle$	11.7 (2.0)	10.9 (2.5)	25.4 (4.3)	13.5 (1.5)	10.8 (1.3)	13.4 (2.0)	15.3 (2.5)	13.2 (2.2)	9.9 (1.9)
Completeness (%)	99.4 (98.6)	98.1 (99.8)	97.4 (82.4)	98.8 (98.5)	98.6 (91.2)	96.7 (92.1)	97.9 (93.6)	99.1 (97.6)	97.8 (93.0)
Redundancy	3.5 (2.7)	5.1 (5.1)	3.7 (3.0)	4.2 (3.8)	3.4 (3.0)	5.9 (4.8)	4.9 (4.3)	6.6 (6.0)	4.7 (4.1)
Refinement									
R_{cryst} (%)	17.5	21.1	16.5	16.7	22.0	23.0	18.7	18.4	20.8
R_{free} (%)	22.2	26.9	21.3	20.7	26.1	28.9	23.6	24.2	26.5
Wilson B factor (Å^2)	28.2	28.0	20.8	22.9	44.3	29.9	23.8	31.6	24.3
Average total B-factor (Å^2)	33.9	29.9	29.2	29.3	43.5	59.9	38.5	35.1	36.7
Average ligand B-factor (Å^2)	39.9	39.7 (active site); 31.5 (crystal contact)	18.6	18.8	55.2	84.2	26.6	31.2	36.3
Ligand occupancy	1.00	1.00	0.70	1.00	1.00	0.81	1.00	1.00	1.00
r.m.s.d. bond length (Å)	0.004	0.003	0.008	0.002	0.003	0.002	0.010	0.002	0.002
r.m.s.d. bond angle (°)	0.8	0.7	1.0	0.7	1.0	0.5	1.2	0.6	0.6
Missing residues	158–160, 206–212	158–162, 208–213	158–162, 336–337	158–161, 336–338	A: 158–160, 210–212; B: 158–160, 210–213, 294–296	158–162, 209–212	158–161, 209–212	158–162, 209–212, 293–295	158–162, 209–213, 293–295
MolProbity analysis									
Clashscore	7.0	10.9	7.0	5.8	15.6	5.0	6.9	4.2	5.7
Ramachandran favored (%)	97.1	96.1	97.7	97.7	94.0	98.1	98.1	98.6	97.6
Ramachandran disallowed (%)	0.5	0.0	0.0	0.5	0.2	0.0	0.5	0.0	0.0
Rotamer outliers (%)	1.2	0.6	0.6	0.0	1.7	0.6	1.7	0.6	0.6

phosphate oxygen atoms is replaced by a sulfur. Previously, it has been shown that the Sp epimer of 2′,3′-cAMPS is a functional substrate for CNPase, while the Rp epimer is not [36]. We used both wild-type CNPase and selected mutants to determine crystal structures of CNPase in the presence of the two phosphorothioate epimers. In both wild-type and V321A CNPase crystals, both forms of 2′,3′-cAMPS had reacted to form a product, 2′-AMPS (Fig. 4), indicating at least a low level of activity also toward the Rp epimer. However, with the use of inactive CNPase, the cyclic Sp epimer could
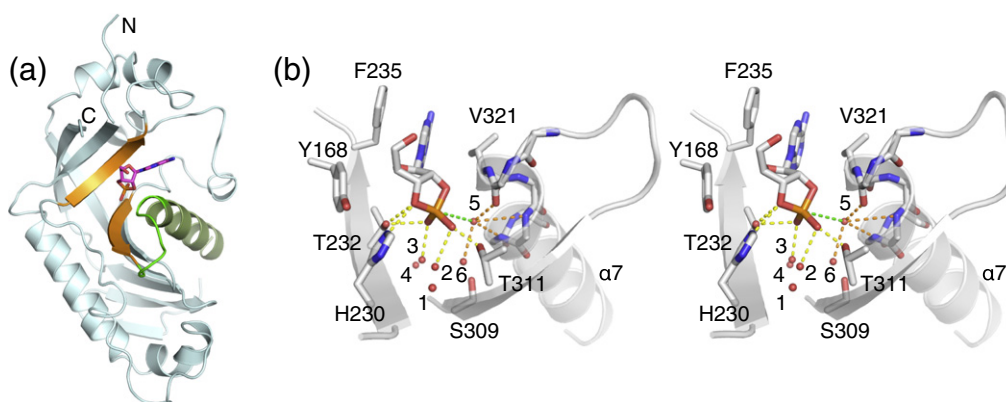
Fig. 2. The CNPase–substrate complex. (a) Overall structure of the CNPase catalytic 2H domain (residues 159–378). The substrate (magenta) is shown in the active site, and helix α7 (dark green), strands β2 and β5 harboring the HxTx motifs (orange), and the β5–α7 loop (bright green) are also highlighted. (b) A stereo view of the binding mode of 2′,3′-cAMP into the CNPase active site. Coordination of the phosphate group is indicated by yellow broken lines, and that of the catalytic water 5 is indicated by orange lines. The nucleophilic attack is shown in green. Numbering of the water molecules 1–5 corresponds to the earlier convention [4,14], while water 6 represents the one replacing the side chain of His309 in the H309S structure.
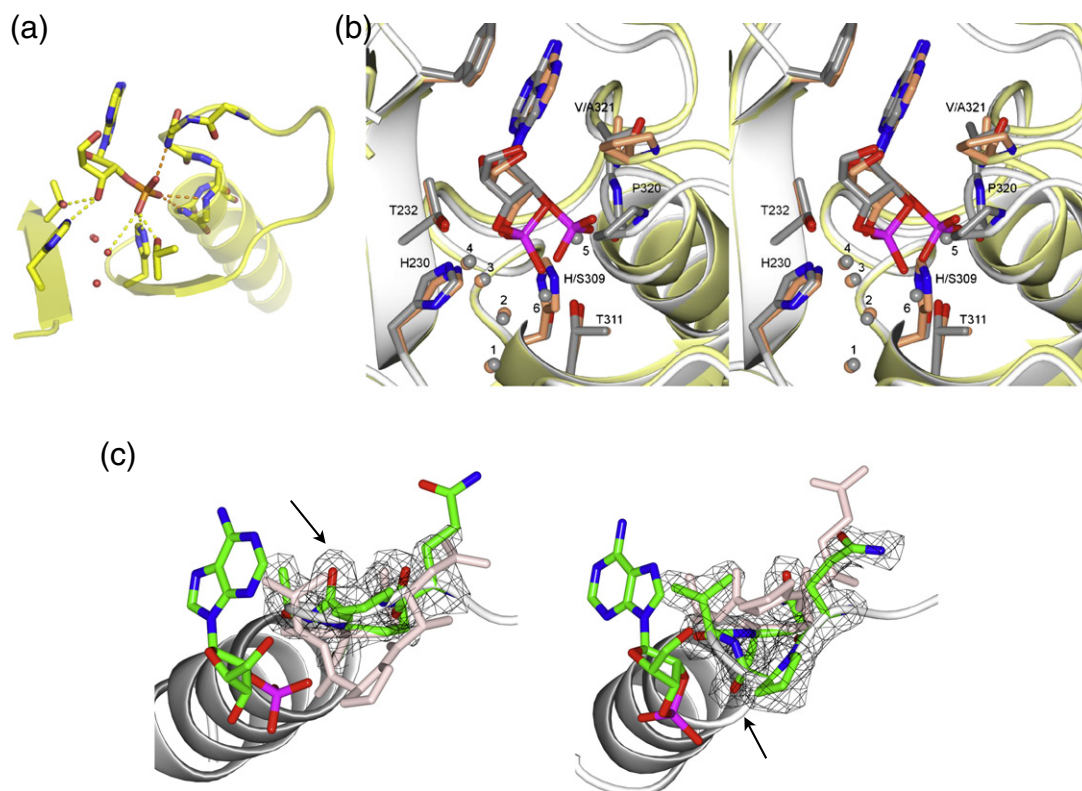

Fig. 3. Comparison between the substrate and product complexes. (a) Binding mode of the reaction product 2′-AMP. Interactions to the HxTx motifs and water molecules are shown in yellow, and interactions to the α7 helix and the β5–α7 loop are shown in orange. (b) Superposition of the substrate (H309S mutant; gray) and product (V321A mutant; yellow/pink) complexes; stereo view. Note the formation of the phosphate-binding pocket upon a conformational change in the β5–α7 loop, including the rotation of the peptide group between Pro320 and residue 321. An animation of this effect is seen in Supplementary Movies 1 and 2. Electrostatic potentials for the two conformations are shown in Supplementary Fig. 3. (c) Electron density maps for the open (left) and closed (right) conformations of the β5–α7 loop. Shown are simulated annealing omit $F_o - F_c$ maps at a contour level of 3 σ. The other conformation of the loop is shown in light pink for comparison, and the carbonyl oxygen of Pro320 is indicated (arrows).
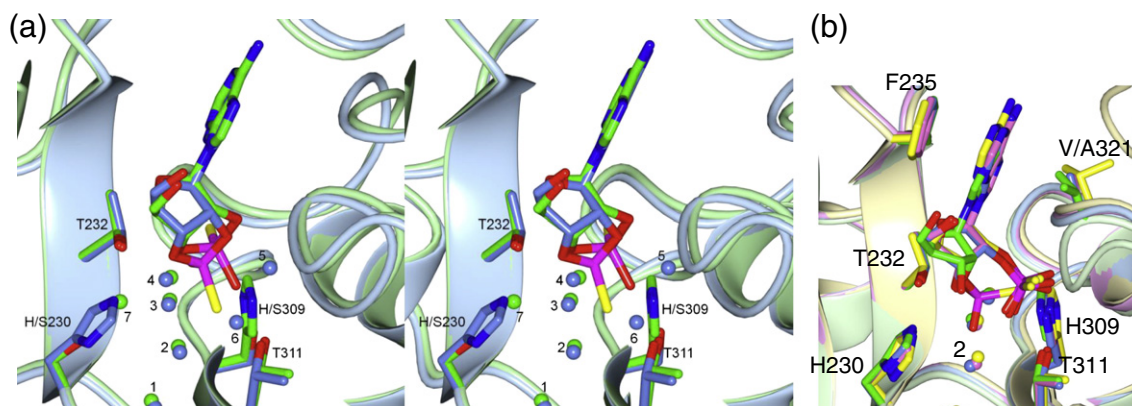
Fig. 4. Complexes of CNPase with phosphorothioate compounds. (a) A stereo view of the binding modes of 2′,3′-cAMPS isomers (H309S with Sp-2′,3′-cAMPS, blue; H230S with Rp-2′,3′-cAMPS, green). Both mutations induce the presence of an additional water molecule (waters 6 and 7). The nucleophilic water 5 is missing from the Rp isomer complex, indicating a non-productive mode of binding. (b) Binding modes of the product 2′-AMPS in crystals of wild-type and V321A CNPase. Both isomers of 2′,3′-cAMPS have been converted into product in the crystals of these active CNPase variants. Yellow, wild-type protein crystallized with Sp-2′,3′-cAMPS; green, wild-type protein crystallized with Rp-2′,3′-cAMPS; blue, V321A crystallized with Sp-2′,3′-cAMPS; pink, V321A crystallized with Rp-2′,3′-cAMPS. In all of the shown structures, except for the wild-type protein with the Rp isomer, the β5–α7 loop is open and water 2 is present.

subsequently be trapped in H309S crystals and the cyclic Rp epimer in the H230S mutant (Fig. 4).

All cyclic ligands, both 2′,3′-cAMP and 2′,3′-cAMPS, occupy fairly similar positions. However, when one of the active-site histidines is mutated, a movement of the cyclic phosphate is seen toward the non-mutated histidine (Fig. 4). This indicates a role for the His residues in the fine-tuning of the substrate position for catalysis, in addition to direct functions in catalysis. The β5–α7 loop in all these complexes is closed.

The product analogues, 2′-AMPS, occupy nearly identical positions in the different structures (Fig. 4b). In the 2′-AMP complex (Fig. 3), the phosphate is located slightly closer to His309 and Arg307, and the difference may be a steric effect due to the O-S replacement in the phosphorothioates. The electron density suggests that the phosphorothioate group rotates and takes all three possible conformations in the crystal (Supplementary Fig. 4). In the 2′-AMPS complexes, loop β5–α7 is again in the open conformation, resulting in more space for product binding.

As an exception to the above, one of the 2′-AMPS complexes presents a different binding mode, where the loop is closed and one of the conserved water molecules at the active site bottom is missing (Fig. 4b). This indicates that the product can also bind to the closed conformation, although the quality of the electron density for the product ligand suggests only weak binding. This could represent a state, where the product is leaving the CNPase active site. In this same structure, also another molecule of 2′-AMPS is seen in a crystal contact, away from the active site (data not shown).

Complex with NADP⁺

Earlier, we reported the structure of wild-type CNPase bound to NADP$^+$, but the analysis was limited due to the low occupancy of the ligand. Hence, we aimed at obtaining a high-occupancy complex with NADP$^+$ or 2′,3′-cNADP$^+$, which is also used in CNPase enzyme activity assays as the substrate. A complex structure was obtained from a crystal of the H230S mutant co-crystallized with 2′,3′-cNADP$^+$. This crystal form was twinned, with two independent monomers in the asymmetric unit.

Although the crystallization was performed with the inactive H230S mutant and the substrate 2′,3′-cNADP$^+$, the reaction product NADP$^+$ was well defined in the active site (Fig. 5). The β5–α7 loop in the NADP$^+$ complex is in the open conformation, in line with the presence of the reaction product. The most likely explanation for the presence of the product is that the timescales of the enzymatic catalysis and crystallization are different and/or that the affinity of the substrate and product to the active site of the mutant protein in the crystal state may favor the product. The 2′-AMP fragment of NADP$^+$ binds essentially as 2′-AMP in the V321A mutant, and the pyrophosphate group is well defined in electron density. The nicotinamide rings were poorly defined in difference electron density prior to inclusion of the ligand molecule in map calculation; however, the final electron density maps show the conformation of this moiety (Supplementary Fig. 2i and j). We attribute this behavior to the fact that the crystals were twinned. In each of the two monomers of the asymmetric unit, the nicotinamide/ribose unit apparently is in a slightly different
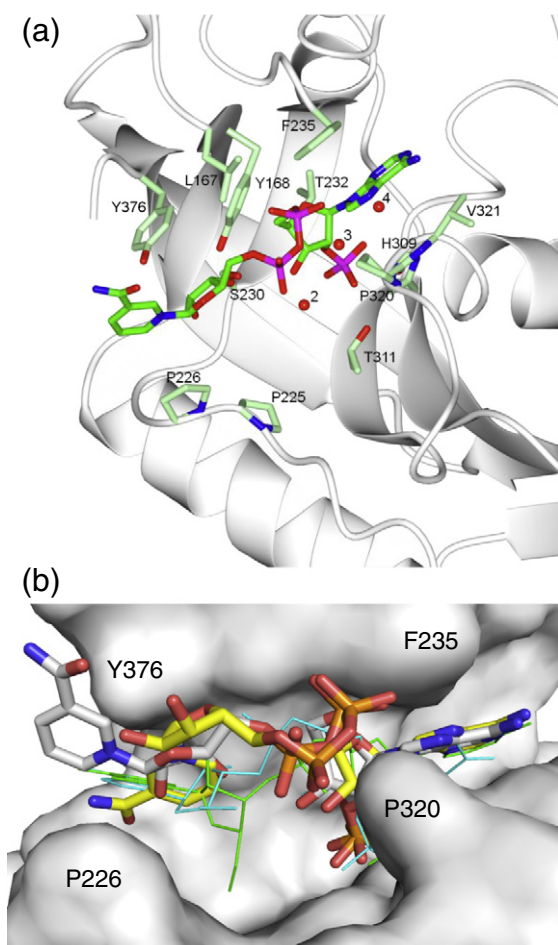
Fig. 5. Dinucleotide binding mode. (a) Binding of NADP$^+$ in chain A of the asymmetric unit. Several interacting residues are shown. (b) Comparison of the NADP$^+$ complex to docked dinucleotides indicates a potential binding groove for longer oligonucleotides. The two conformations of NADP$^+$ in the crystal are shown as sticks (white/yellow), and the docked nucleotides adenylyl-l-3′,5′-uridine-2′-monophosphate (cyan) and adenylyl-l-3′,5′-adenine-2′-monophosphate (green) are shown as thin bonds.

conformation; however, it has a tendency to bind between a loop containing Pro226 and Tyr376 close to the C terminus of the phosphodiesterase domain (Fig. 5). This groove could represent the continuation of a binding site for further nucleotide units in an RNA substrate. Furthermore, the N-terminal domain, likely to be responsible for high-affinity RNA binding [4], is localized in this direction in the full-length protein.

Although NADP$^+$ is a dinucleotide, it is not a perfect analogue for an RNA oligonucleotide. Hence, we also carried out docking of dinucleotide compounds, adenylyl-3′,5′-uridine-2′-monophosphate and adenylyl-3′,5′-adenine-2′-monophosphate, as well as their 2′,3′-cyclic derivatives, representing

substrate and product into the respective CNPase active-site conformations. This experiment indicated that, while the reactive groups were placed into the active site, the second nucleotide bound to the same groove as NADP$^+$ (Fig. 5b).

Conformation of full-length CNPase in solution

Full-length CNPase isoform 1, including the N- and C-terminal domains as well as the C-terminal tail, was purified in a homogeneous monomeric form (Supplementary Fig. 5). In our earlier work, using a slightly shorter construct without the 22-residue C-terminal tail, purified CNPase always contained a mixture of different oligomeric states [4]. SRCD was used to confirm the folding state of CNPase, and the results indicate a similar secondary structure content for the catalytic domain and full-length CNPase and, hence, also for the N-terminal domain (Supplementary Fig. 1). Based on synchrotron SAXS data (Fig. 6 and Supplementary Table 3), the full-length CNPase monomer is an elongated molecule with a maximum dimension of 8.8 nm and a radius of gyration of 2.6 nm. The determined molecular mass based on comparing the CNPase forward scattering intensity $I(0)$ to that of bovine serum albumin was 42.7 kDa, very close to the expected molar mass of a monomer (44.8 kDa), confirming the monomeric state of full-length CNPase in solution. A method not relying on a bovine serum albumin standard also gave a molecular weight consistent with a monomer. Furthermore, the estimated Porod volume (Supplementary Table 3) is well in line with the presence of monomeric CNPase.

Molecular modeling of full-length CNPase was carried out based on the SAXS data. Models were prepared both *ab initio*, without any additional structural information, and by taking advantage of the crystal structure of the CNPase catalytic domain and a homology model of the N-terminal domain, carrying out rigid-body refinement and building missing loops. Both kinds of procedures resulted in very similar elongated models of full-length CNPase, where only a small area of close contact exists between the two domains (Fig. 6). All models also gave very good fits to the raw data (Supplementary Table 3). The models indicate that the CNPase active site is accessible in the context of the full-length protein in solution, not being covered by the N-terminal domain; the latter case would result in a more globular particle. In this arrangement of the two CNPase domains, the C-terminal membrane-anchoring tail must be located in the central part of the molecule, effectively bringing also the active site close to the lipid bilayer surface.

Sequence analysis of CNPase homologues

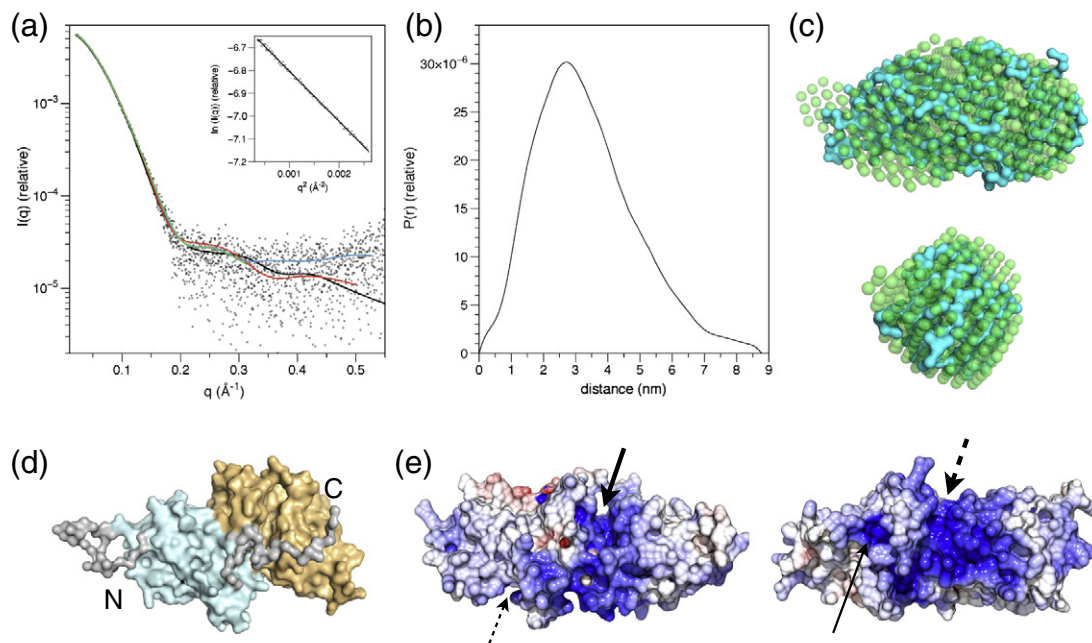More than 10 years ago, a sequence analysis of the 2H family was presented [2]. Due to the

Fig. 6. Three-dimensional model of full-length CNPase from SAXS data. (a) SAXS scattering curve. The raw data (black dots) are overlaid with fits of the *ab initio* GASBOR (blue) and DAMMIF (green) and rigid-body CORAL (black) and BUNCH (red) models. The inset shows the Guinier plot. (b) Distance distribution of full-length CNPase. (c) Chain-like *ab initio* model built by GASBOR (cyan), superimposed on an averaged dummy atom model (green), based on 10 independent runs of DAMMIF. The two views are related by a rotation of 90° about the *y*-axis. (d) Full-length CNPase model made using CORAL. The N-terminal domain is colored cyan. The C-terminal domain is orange. The N- and C-terminal tails are colored gray. During modeling, in addition to the two rigid-body domains, 30 N-terminal and 22 C-terminal residues were built, as well as a short linker of 4 residues between the domains. (e) Electrostatic potential coloring of the BUNCH model surface at two orientations, related by a 90° rotation about the *x*-axis. The coloring is from − 10 (red) to + 10 (blue) in units produced by ABPS. The CNPase active site and the location of the P-loop/ATP-binding site are located by thick and thin arrows, respectively. A broken arrow indicates the site is behind the molecule in this orientation.

significant increase of sequence data, we studied the conservation of the CNPase family in currently available protein sequences of metazoan species (Supplementary Figs. 6 and 7). CNPase is present in all complete tetrapod genomes (33), except in platypus, and CNPase or RICH is present in all complete fish genomes (9). Additionally, we detected CNPase sequences in six invertebrates. CNPase is apparently not present in most arthropods, nematodes, urochordates, or platypus. A schematic view of the domain organization in the CNPase/RICH family in different organisms is shown in Fig. 7.

The HxS/Tx motifs of the catalytic phosphodiesterase domain are conserved throughout species. The P-loop of the N-terminal domain is also present in all sequences, except the RICH proteins and the water flea CNPase, both of which only have homology to the C-terminal domain of human CNPase. The P-loop consensus residues are fully conserved, indicating functional significance of this putative ATP-binding site. Stretches of high conservation are also found at the beginning of the C-terminal domain (strand β1), around the second HxTx motif (strand β5), and in helix α7.

The C-terminal tail of CNPase contains a CTII motif, carrying the conserved (except for coelacanth that has a Phe substitution) cysteine, which is isoprenylated, and Lys379 and Gly380 that promote microtubule polymerization [26,28]. This C-terminal tail is missing in invertebrates, in which the position of the C terminus corresponds to the CNPase residue Tyr377.

Some sequences in the database contain extensions at the N terminus before the point corresponding to the start of the CNPase 1 isoform. These may represent isoforms produced by alternative splicing, or artifacts. Both amphibians and all non-tetrapods lack homology to the mitochondrial targeting sequence of the CNPase 2 isoform [37]. The RICH proteins do not have a PNK-like domain; instead, they possess an N-terminal acidic domain rich in Pro, Glu, Val, Lys, and Ala. This kind of an amino acid composition is suggestive of a disordered structure, and secondary structure predictions confirm this view (unpublished results). Interestingly, salmon CNPase has the acidic domain, PNK domain, and 2H domain, combining the features of both CNPase and RICH. Schistosoma CNPase, on the other hand, has a coiled-coil domain of 80–90 residues at its N terminus.
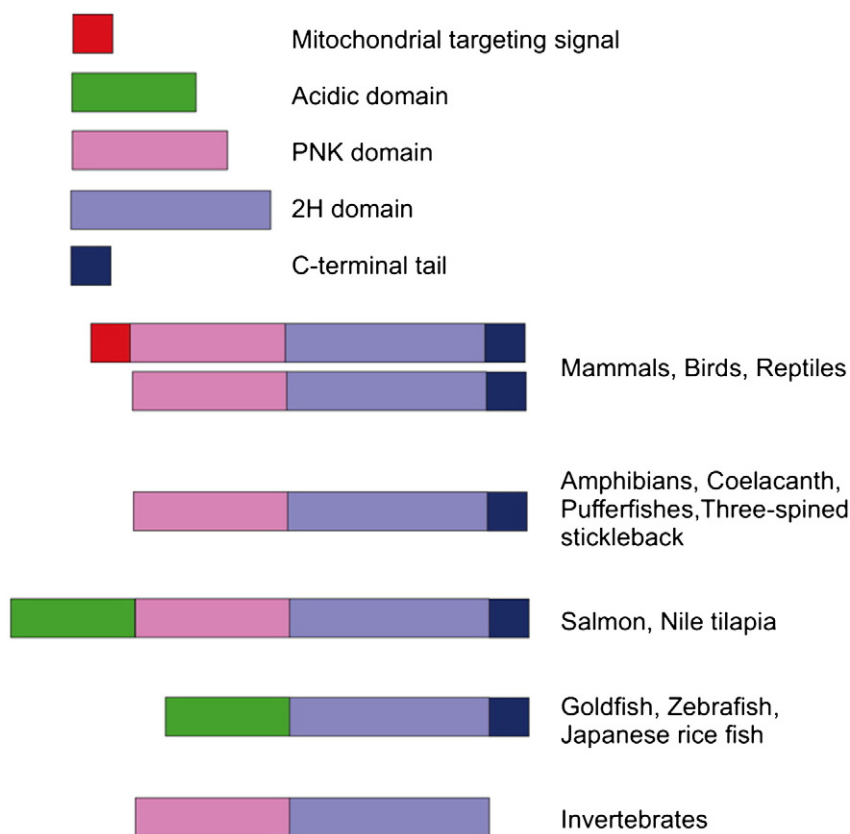
Fig. 7. Domain organization of CNPase/RICH in different organisms. Details and amino acid sequence alignments can be found in Supplementary Materials.

Conservation of ligand-binding residues

Based on the CNPase crystal structures and sequence alignments, conserved features in ligand binding can be observed. Both HxTx motifs are always present, except that the latter threonine is substituted by serine in sea urchin. Tyr168, which interacts with the ribose ring, is conserved in most species but is replaced by phenylalanine in bat, most fishes, acorn worm, and sea urchin. Pro225 and Pro226, lining the extended ligand-binding cleft, are conserved within mammals. Tyr376 on the opposite side of this cleft is conserved in most species, including mammals, birds, and amphibians. Phe235, stacking against the base of the substrate, is conserved in mammals and most other species but changed to tyrosine in birds and some invertebrates. Val321, which binds to the other face of the base through CH-π hydrogen bonds, is mostly conserved but substituted by methionine in naked mole rat, by glutamine in Atlantic salmon, and by alanine in *Schistosoma*. The residue corresponding to Pro320 is either Pro or Ala in all sequences, except for frogs and the naked mole rat.

Discussion

The catalytic cycle of CNPase

We report here the first substrate-containing structure of a 2H phosphoesterase family enzyme and present a comprehensive crystallographic analysis of the CNPase catalytic cycle, including the detailed positioning of the substrate and the reactive water molecule. The position of the cyclic nucleotide in the active site is consistent with the earlier predicted reaction mechanism [4,14]. Based on our current data, we are able to propose a much more detailed mechanism for catalysis by CNPase (Fig. 8).

Prior to catalysis, the active-site histidine residues are in the correct protonation state for catalysis: His230 is protonated and His309 is deprotonated. As the optimum pH for the CNPase catalytic activity is 6 [38], it is likely that His230 has a pK_a of over 6 and His309 has p$K_a <$ 6.

In the first step of the reaction, a 2′,3′-cyclic nucleotide substrate enters the active site. No large-scale conformational changes in the active
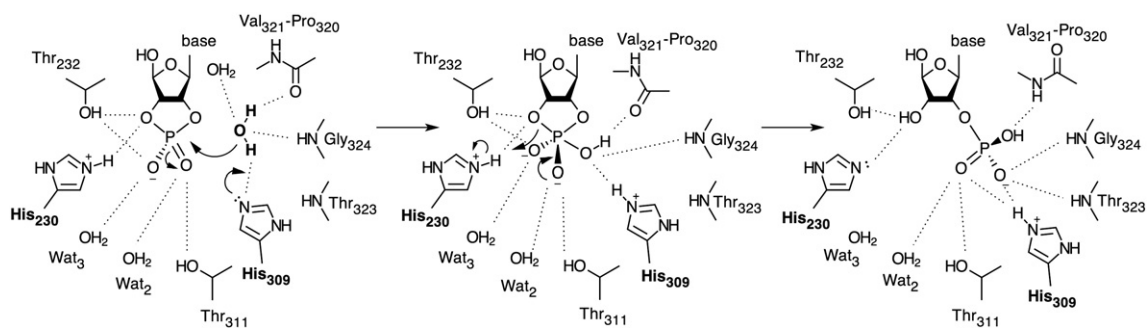
Fig. 8. The current view of the CNPase reaction mechanism. The two catalytic histidine residues and the nucleophilic water molecule are highlighted in bold. Hydrogen bonds, as seen in the crystal structures, are shown as broken lines. Note how the phosphate group changes its environment during the reaction, interacting with the amide groups of the N terminus of helix α7 and the β5–α7 loop in the product complex.

site or its vicinity are suggested by comparisons of liganded and unliganded CNPase structures. The reactive cyclic phosphate group of the substrate is accurately oriented by both of the HxTx motifs and two water molecules (waters 2 and 3) at the bottom of the active site. The nucleophilic water molecule is coordinated by hydrogen bonds to an optimal position and orientation for catalysis. The interacting moieties include the side chain of His309, the backbone carbonyl of Pro320, the backbone amide of Gly324 in the N terminus of helix α7, and a nearby water molecule. The hydrogen bond to Gly324 is in-line with the direction of nucleophilic attack.

The reaction is initiated by the abstraction of a proton from the nucleophilic water molecule by the side chain of His309. The activated water molecule then attacks the cyclic phosphate group, most likely resulting in a pentavalent intermediate. It can be predicted that most of the hydrogen bonds between the active site and the substrate are preserved during this step. It is possible that the Pro320 carbonyl group follows the nucleophilic attack and starts its rotation during this step.

In the second step of the CNPase reaction, the pentavalent intermediate is broken, and a 2′-phospho group is formed. His230 acts as a general acid and donates a proton to the 3′-oxygen. During this step, the phosphate group moves significantly, and some of the original hydrogen bonding contacts are lost, for example, to water 3. At the same time, the peptide carbonyl between Pro320 and Val321 rotates away and the backbone amide of Val321 is exposed for interaction with the reaction product. The 2′-phospho group also forms direct hydrogen bonds to the backbone amides of Thr323 and Gly324 at the N terminus of helix α7.

It is likely that the open conformation of the β5–α7 loop is strained, and product release is linked to its return to the closed conformation. This is also suggested by one of our structures, where a leaving product molecule is bound with a low occupancy and the loop is closed.

Before a new catalytic cycle can take place, the active site must be regenerated. For CNPase, this mainly involves the reversal of the protonation states of the active-site histidine side chains. This can be achieved through interactions with water and buffer molecules; in the unliganded state, both His230 and His309 interact directly with solvent in the relatively open active-site cavity. It is likely that the regeneration of the active site is impaired far away from the CNPase pH optimum.

Further insights from active-site ligand complexes

CNPase has stereospecificity toward 2′,3′-cAMPS analogues, with only the Sp epimer being a good substrate [36]. Our crystal structures with both the Sp and Rp epimers indicate subtle differences in the cyclic ligand conformation; while the Sp epimer is very similar to 2′,3′-cAMP, the Rp epimer complex structure is missing the nucleophilic water molecule. Hence, the inability of CNPase to efficiently react upon Rp-2′,3′-cAMPS could, indeed, be due to a non-productive binding mode. On the other hand, the observation of the product 2′-AMPS in the crystals grown in the presence of active CNPase and the Rp epimer suggests at least a low level of activity toward Rp-2′,3′-cAMPS.

The product NADP$^+$ was observed in active site of the inactive mutant H230S crystals, grown in the presence of the substrate 2′,3′-cNADP$^+$. This may indicate that, firstly, His230 is not absolutely required for the first step of the reaction and that, secondly, the water molecule replacing the side chain of His230 in the mutant is able to donate a proton to the leaving group in the absence of His230. On the other hand, the water molecule replacing the side chain of His309 in the crystal structure is unable to start catalysis, as shown by the fact that this mutant could be used to obtain the structure of a substrate complex.

The binding mode of NADP$^+$ in the active site hints at a continuation of a binding site for oligonucleotide

substrates, extending from the active-site groove toward the N-terminal domain (Figs. 5 and 6e). Docking experiments also showed binding of di-nucleotides into this groove. It can be envisaged that an oligonucleotide would bind along an aromatic surface, lined by a strong positive electrostatic potential, toward the N-terminal domain, which has been shown to efficiently bind RNA [4]. Kinetic parameters for CNPase toward 2′,3′-cAMP and 2′,3′-cNADP$^+$ were reported to be essentially identical [22], indicating that the kinetics are mainly determined by the nucleotide moiety in the active site; however, for longer substrates, the situation may be different. High-resolution structural data for full-length CNPase or a close homologue, preferably in complex with RNA, will be required to fully understand this process.

The β5–α7 loop and helix α7

Depending on the ligand present in the active site, two distinct conformations for the β5–α7 loop can be observed; in the substrate complexes, it is always closed, and in the product complexes, it is open, except for one structure with low product occupancy. The movement of the β5–α7 loop during catalysis, as well as the generation of a binding cavity for the 2′-phosphate group, is visualized in Supplementary Movies 1 and 2. Linked to the movement is also the conformation of Cys314, which turns away from the active site upon loop opening, contributing to the formation of a large binding pocket for the 2′-phosphate of the product. It is likely that the β5–α7 loop plays a key role both in coordinating the reactive species before and after catalysis and in releasing the product. Originally, its role was thought to reside mainly in holding the substrate nucleobase in place, but our data indicate that it also specifically co-ordinates the reactive species during the entire CNPase catalytic cycle.

When the β5–α7 loop opens and binds the product (2′-AMP, 2′-AMPS, or NADP$^+$), direct hydrogen bonds are formed between backbone amides from Val321, Thr323, and Gly324 and the product. NMR titrations with active-site ligands, including phosphate and 2′-AMP and 3′-AMP, as well as A$_6$ RNA, previously showed the largest chemical shift perturbations for Thr232, Thr311, and Gly324 [12]. Thr323 and Gly324 are fully conserved in all CNPase homologue sequences we studied here (Supplementary Figs. 6 and 7), and the mutation of Gly324 to alanine largely abolished CNPase activity [12]. Hence, we believe that the presence of the correctly positioned N terminus of loop α7 and the flexibility of the β5–α7 loop are important for positioning the reactive groups for catalysis and for product release.

Helix α7 of CNPase, while being highly conserved within CNPases, is absent from other 2H superfamily proteins with a known structure [4]. Given the importance of this helix, and especially its N-terminal loop, for the catalytic mechanism of CNPase and its conservation during CNPase evolution, the α7 helix presents a profound difference in CNPase structure–function relationships compared with the other 2H superfamily enzymes. Hence, it appears that non-CNPase members of the 2H family must have different mechanisms for substrate recognition and binding, as well as for organizing the presence of a nucleophilic group.

Structure of full-length CNPase

We further modeled the three-dimensional solution structure of full-length CNPase at low resolution. The model, based on synchrotron SAXS data and both *ab initio* and homology modeling, indicates that the PNK-like and the 2H phosphoesterase domains form an elongated assembly. This arrangement retains an open active site in the catalytic domain, with no apparent need for large-scale conformational changes prior to substrate binding. Earlier comparisons of full-length CNPase and its catalytic domain have shown that although the K_m values for the substrate are similar, the isolated catalytic domain is approximately two times faster in catalysis than the full-length protein [39,40]. It is likely that the N-terminal domain is involved in the binding of larger substrates, namely RNA molecules, the termini of which then would locate to the active site for catalysis. The electrostatic surface of the full-length CNPase model suggests the presence of a positively charged patch leading to the active site (Fig. 6e). The roles of the conserved P-loop or the binding of other ligands, such as calmodulin [30], in the regulation of RNA substrate binding and/or catalysis still remain to be determined. It is highly likely that homologous enzymes from the 2H family have similar domain arrangements between the PNK-like and phosphodiesterase domains. Indeed, in our recent experiments, we have seen similar molecular shapes for CNPase homologues (unpublished results).

CNPase as a member of the 2H superfamily

The 2H superfamily is divided into four major clades based on primary sequence (archaeobacterial LigT-like, eukaryotic-viral LigT-like, YjcG-like, and mlr3352 clades), and several enzymes, including CNPase, belong to no clades [2]. The archaeobacterial LigT-like clade is similar to *E. coli* 2′-5′ RNA ligase, and members are found in archaea, bacteria, and chloroplasts [2,41,42]. These enzymes hydrolyze 2′,3′-cyclic phosphate in (oligo)nucleotides and join the produced 2′-phosphate to a 5′-hydroxyl group of another (oligo)nucleotide to form atypical 2′,5′-linkages [6,41]. Such activity has not been reported for CNPase, and it may have been lost

upon the emergence of helix α7, which effectively blocks access to the active site from one direction, being also the direction of nucleophilic attack. In CNPase, the N terminus of helix α7 coordinates the nucleophilic water, while in some other 2H phosphoesterases, the absence of this helix allows the entrance of a larger nucleophile, such as an RNA molecule, into the active site.

Fungal tRNA ligases use PNK, CPDase, and RNA ligase activities to produce common 3′,5′-ligations during tRNA splicing [17,43]. Rat CNPase can rescue yeast with an inactive tRNA ligase 2H domain, but not yeast with an inactive PNK domain [21]. On the other hand, lancelet CNPase can rescue the PNK-inactive mutants [44]. The CNPase PNK-like domain is fairly conserved, especially within mammals. This implies that it may fulfill some conserved function, although it may have lost PNK activity during evolution. This function could be related to RNA binding [4,29,30] or even a defense mechanism against viral vectors [45].

Proteins in other major clades of the 2H superfamily have not been reported to possess CNPase activity. The murine coronavirus non-structural protein 2a (MHV-ns2) can, like the prokaryotic LigT-homologues, degrade 2′,5′-linked oligoadenylates formed by host cells during viral infection [46]. Mutating the catalytic histidine residues in this protein led to an attenuation of viral infection in the liver, but not in the brain [47]. Plant and yeast CPDases participate in tRNA splicing by degrading the side product ADP-ribose 1″,2″-cyclic phosphate to ADP-ribose 1″-phosphate [3,11,48]. They are also capable of degrading 2′,3′-cyclic phosphates to 2′-phosphates but also lack helix α7, reflecting wider substrate specificity and differences in reaction mechanism.

It has been assumed that CNPase emerged along with myelin in early vertebrates due to the lack of activity in invertebrate species such as molluscs, arthropods, and starfish [49] and a lack of homologous sequences in sequenced invertebrates [50]. However, the properties of lancelet CNPase [44] and our analysis of currently available metazoan sequences here suggest an earlier origin for the protein. CNPase is present in high amounts in oligodendrocytes and myelin of the mammalian nervous system, and CNPase activity is also present during development in chicken [51] and *Xenopus* nervous systems [52]. On the other hand, RICH, found in some actinopterygian fish, lacks the PNK-like domain of CNPase. A CNPase sequence with PNK-like and 2H domains is present in the sarcopterygian fish Coelacanth, which is more closely related to the ancestor of land-dwelling tetrapods [53]. Although no CNPase sequence from chondrichtyes is available, CNPase activity has been found in sharks and rays [49,54]. The presence of CNPase homologues in flatworms indicates its presence in a common bilaterian ancestor and the deuterostoma lineage.

Concluding remarks

We have presented a detailed structural view of the CNPase catalytic cycle and visualized a substrate complex containing 2′,3′-cAMP and the nucleophilic water molecule. The data from different active-site ligand complexes point toward an important role for the β5–α7 loop during the catalytic cycle, coordinating the reactive moieties and going through a distinct conformational change during catalysis. The large differences seen between 2H family members in the vicinity of the active site, especially concerning helix α7 and the adjacent β5–α7 loop, must result in different catalytic details in other 2H family members. While the physiological function of CNPase still remains enigmatic, the elucidation of the catalytic mechanism and our low-resolution model of full-length CNPase are important steps toward understanding the physiological function of this highly concentrated membrane-bound enzyme in the myelin sheath.

Materials and Methods

Mutagenesis, protein expression, and purification

Mutations were prepared using the QuikChange II site-directed mutagenesis kit (Agilent Technologies) and the primers listed in Supplementary Table 1. The template for the mutations was mouse CNPase catalytic domain cDNA, encoding residues 159–378 (numbering according to the CNPase 1 splice isoform), cloned into the pTH27 expression vector [40,55]. The presence of mutations was confirmed by DNA sequencing and mass spectrometric analysis of the purified protein. Mutated and wild-type CNPase catalytic domains were expressed and purified exactly as described before [4,40]. Briefly, His-tagged CNPase was expressed in *E. coli* using autoinduction [56] and purified with metal affinity, His-tag cleavage, and size-exclusion chromatography. Soluble expression and purification properties of all mutants were similar to the wild-type protein (data not shown).

Full-length mouse CNPase 1 isoform was also expressed and purified to homogeneity with the same protocol [40]. In metal affinity purification, a HisTrap Fast Flow column (GE Healthcare) was used, which had been equilibrated with washing buffer [50 mM Hepes (pH 7.0), 500 mM NaCl, 10% glycerol, 20 mM imidazole, and 1 mM tris(2-carboxyethyl) phosphine (TCEP)]. After protein binding and washing, we eluted pure protein with washing buffer containing 400 mM imidazole. Thereafter, tobacco etch virus protease was added, and cleavage was carried out simultaneously with dialysis into 50 mM Hepes (pH 7.0), 500 mM NaCl, 10% glycerol, and 1 mM TCEP. Size-exclusion chromatography was then carried out in 20 mM Hepes (pH 7.5), 300 mM NaCl, 1% glycerol, and 1 mM TCEP.

The presence of mutations was confirmed by DNA sequencing and by mass spectrometry of purified proteins. Peptide fingerprinting was carried out using in-gel trypsin digestion and matrix-assisted laser desorption/ionization/time-of-flight mass spectrometry, and accurate molecular mass determination was performed by liquid chromatography-coupled electrospray ionization/time-of-flight mass spectrometry.

Kinetics and stability assays

Enzymatic activity of the mutants was assayed as previously described [22,30]. The substrate 2′,3′-cNADP$^+$ was purchased from Sigma-Aldrich. We used 5 ng of CNPase and 0.02–2 mM 2′,3′-cNADP$^+$ in the reactions.

The effect of mutations on protein stability was determined using the thermal shift assay [40]. Briefly, 5 μg of protein was mixed with 1× SYPRO Orange (Invitrogen) in a buffer containing 20 mM 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol (pH 5.5), 0.3 M NaCl, 1% glycerol, and 1 mM TCEP. Measurements were performed using an Applied Biosciences 7500 Real-Time PCR instrument in 50-μl volumes in 96-well plates, and the plate was heated from 21 °C to 90 °C, at 1 °C/min, while fluorescence was recorded. The excitation and emission wavelengths were 490 and 575 nm, respectively.

Crystal structure determination

Previously, we described crystallization conditions for wild-type CNPase that contained citrate buffer and medium-sized polyethylene glycol [4]. Although we reported some successes in soaking CNPase nucleotide ligands into these crystals, most structures solved using these crystals have citrate in the active site. In the current study, when citrate was replaced by formate or acetate, crystals generally grew only in the presence of nucleotides, and the nucleotides were usually present in the active site in the refined structures. 2′,3′-Cyclic phosphorothioates were purchased from BIOLOG Life Science Institute (Bremen, Germany), and 2′,3′-cAMP, from Sigma-Aldrich. Crystallization and cryoprotection conditions are displayed in Supplementary Table 2. Crystals were briefly soaked in the cryoprotectant and flash-cooled in liquid nitrogen. Data were collected at 100 K on the I911-2 beamline at MAX-Lab (Lund, Sweden) and on the X11, X12, and P14 beamlines at EMBL (*European Molecular Biology Laboratory*)/DESY (*Deutsches Elektronen-Synchrotron*) (Hamburg, Germany). Data were processed and scaled using XDS [57]. Phases were determined by molecular replacement with a previous model (PDB entry 2xmi [4]), using Phaser [58]. phenix.refine was used for crystallographic refinement, and phenix.elbow was used for the generation of ligand restraints [59]. Manual rebuilding was performed using Coot [60], and structural figures were created using CCP4mg [61], UCSF Chimera [62], APBS [63], and PyMOL‡. MolProbity was used in structure validation [64]. The ligand omit electron density maps were calculated by setting the ligand occupancy to zero and carrying out three rounds of simulated annealing refinement in phenix.refine. The exception was the NADP$^+$ complex of the H230S mutant; since the crystal form was twinned, the density before adding in the ligand was taken as the omit map.

Small-angle X-ray scattering

Full-length CNPase was concentrated to 4–6 mg/ml in a buffer containing 10 or 50 mM Hepes (pH 7.5), 0.2 M NaCl, and 1 mM TCEP. Several dilutions of CNPase were measured at the synchrotron beamline I911-4 of MAX-Lab (Lund, Sweden), while a very small amount of aggregation was detectable in most samples, one sample, at 4.5 mg/ml in 50 mM buffer, showed no signs of aggregation and was used for further analyses. A freshly prepared sample of monomeric bovine serum albumin was used as a molecular weight standard. Molecular weights were also determined without a standard protein from the SAXS data alone, using the SAXS MoW procedure [65]. The raw data were processed with BioXTAS RAW [66], and the ATSAS package [67] was used for further analyses and structure modeling. Specifically, distance distribution functions were generated by GNOM [68], and the programs DAMMIF, GASBOR, BUNCH, and CORAL were used for *ab initio* and rigid-body modeling [67,69,70]. Ten independent DAMMIF models were further averaged using DAMAVER [71]. A homology model for the CNPase N-terminal domain, based on the crystal structure of T4 PNK [72], was obtained using the sequence of the CNPase N-terminal domain at the Phyre2 server [73]. Phyre2 uses a combination of sequence profile alignment and secondary structure prediction to find suitable models for homology modeling. This model and the crystal structure for the C-terminal domain were used as inputs for rigid-body modeling coupled to loop building in BUNCH and CORAL. SAXS data for the catalytic domain and mutants thereof {in 20 mM 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol (pH 5.5), 300 mM NaCl, 10% glycerol, and 1 mM TCEP} were similarly collected at the European Synchrotron Radiation Facility (Grenoble, France) on the ID14-3 BioSAXS beamline.

Synchrotron radiation circular dichroism spectroscopy

SRCD was used to analyze possible changes in the folding state of the mutant variants, compared to the wild-type catalytic domain. SRCD data were collected in 10 mM sodium phosphate buffer (pH 7.0) on the CD1 beamline at the ASTRID storage ring (ISA, Aarhus, Denmark). SRCD data were also collected on a sample of full-length CNPase, on the UV-CD12 beamline at ANKA (KIT, Karlsruhe, Germany). Spectral deconvolution to estimate secondary structure contents was performed at Dichroweb [74], using the CDSSTR [75] algorithm and the SP175 reference database [76].

Molecular docking

Docking studies of dinucleotide ligands to CNPase were performed with AutoDock Vina [77]. The original ligands, the substrate or the product, and water molecules in the

PDB files of H309S and V321A CNPases were removed; however, crucial water molecules in the binding site were kept. Ser309 and Ala321 of each mutant structure were exchanged to His (H309H) and Val (V321V), respectively, to mimic the wild-type structure. The ligand structures were obtained from the PDB and modified using Avogadro, if necessary [78]. The input files for proteins and ligands, PDBQT molecular structure file formats, were prepared with AutoDock Tools. The side chains of a few residues in the binding pocket were treated as flexible. The grid box was centered on the binding pocket of the protein, and the sizes were set to 22 Å × 28 Å × 24 Å for H309S and H309H and 24 Å × 22 Å × 18 Å for V321A and V321V CNPase structures. The analysis of the results was carried out using PyMOL, and the final results were chosen based on binding modes.

Sequence analysis

Sequences homologous to human CNPase (P09543) were searched from the UniProt and National Center for Biotechnology Information protein databases [79,80]. A list of sequences selected for alignment is displayed in Supplementary Table 4. Sequence alignments were prepared using MAFFT [81] and Jalview [82].

Accession numbers

The crystal structure data were deposited at the PDB with accession codes 2yoz, 2yp0, 2ypc, 2ype, 3zbr, 2yph, 2yq9, 3zbs, and 3zbz.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2013.06.012

References

[1] Ballestero RP, Dybowski JA, Levy G, Agranoff BW, Uhler MD. Cloning and characterization of zRICH, a 2′,3′-cyclic-nucleotide 3′-phosphodiesterase induced during zebrafish optic nerve regeneration. J Neurochem 1999;72:1362–71.

[2] Mazumder R, Iyer LM, Vasudevan S, Aravind L. Detection of novel members, structure–function analysis and evolutionary classification of the 2H phosphoesterase superfamily. Nucleic Acids Res 2002;30:5229–43.

[3] Nasr F, Filipowicz W. Characterization of the *Saccharomyces cerevisiae* cyclic nucleotide phosphodiesterase involved in the metabolism of ADP-ribose 1″,2″-cyclic phosphate. Nucleic Acids Res 2000;28:1676–83.

[4] Myllykoski M, Raasakka A, Han H, Kursula P. Myelin 2′,3′-cyclic nucleotide 3′-phosphodiesterase: active-site ligand binding and molecular conformation. PLoS One 2012;7:e32336.

[5] Gao YG, Yao M, Okada A, Tanaka I. The structure of *Pyrococcus horikoshii* 2′-5′ RNA ligase at 1.94 A resolution reveals a possible open form with a wider active-site cleft. Acta Crystallogr Sect F Struct Biol Cryst Commun 2006;62: 1196–200.

[6] Kanai A, Sato A, Fukuda Y, Okada K, Matsuda T, Sakamoto T, et al. Characterization of a heat-stable enzyme possessing GTP-dependent RNA ligase activity from a hyperthermophilic archaeon, *Pyrococcus furiosus*. RNA 2009;15:420–31.

[7] Kato M, Shirouzu M, Terada T, Yamaguchi H, Murayama K, Sakai H, et al. Crystal structure of the 2′-5′ RNA ligase from *Thermus thermophilus* HB8. J Mol Biol 2003;329:903–11.

[8] Rehse PH, Tahirov TH. Structure of a putative 2′-5′ RNA ligase from *Pyrococcus horikoshii*. Acta Crystallogr Sect D Biol Crystallogr 2005;61:1207–12.

[9] Gold MG, Smith FD, Scott JD, Barford D. AKAP18 contains a phosphoesterase domain that binds AMP. J Mol Biol 2008;375:1329–43.

[10] Li D, Liu C, Liang YH, Li LF, Su XD. Crystal structure of *B. subtilis* YjcG characterizing the YjcG-like group of 2H phosphoesterase superfamily. Proteins 2008;72:1071–6.

[11] Hofmann A, Zdanov A, Genschik P, Ruvinov S, Filipowicz W, Wlodawer A. Structure and mechanism of activity of the cyclic phosphodiesterase of Appr > p, a product of the tRNA splicing reaction. EMBO J 2000;19:6207–17.

[12] Kozlov G, Lee J, Elias D, Gravel M, Gutierrez P, Ekiel I, et al. Structural evidence that brain cyclic nucleotide phosphodiesterase is a member of the 2H phosphodiesterase superfamily. J Biol Chem 2003;278:46021–8.

[13] Kozlov G, Denisov AY, Pomerantseva E, Gravel M, Braun PE, Gehring K. Solution structure of the catalytic domain of RICH protein from goldfish. FEBS J 2007;274:1600–9.

[14] Sakamoto Y, Tanaka N, Ichimiya T, Kurihara T, Nakamura KT. Crystal structure of the catalytic fragment of human brain 2′,3′-cyclic-nucleotide 3′-phosphodiesterase. J Mol Biol 2005;346:789–800.

[15] Drummond GI, Iyer NT, Keith J. Hydrolysis of ribonucleoside 2′,3′-cyclic phosphates by a diesterase from brain. J Biol Chem 1962;237:3535–9.

[16] Ballestero RP, Wilmot GR, Agranoff BW, Uhler MD. gRICH68 and gRICH70 are 2′,3′-cyclic-nucleotide 3′-phosphodiesterases induced during goldfish optic nerve regeneration. J Biol Chem 1997;272:11479–86.

[17] Phizicky EM, Schwartz RC, Abelson J. *Saccharomyces cerevisiae* tRNA ligase: purification of the protein and isolation of the structural gene. J Biol Chem 1986;261:2978–86.

[18] Tyc K, Kellenberger C, Filipowicz W. Purification and characterization of wheat germ 2′,3′-cyclic nucleotide 3′-phosphodiesterase. J Biol Chem 1987;262:12994–3000.

[19] de Monasterio-Schrader P, Jahn O, Tenzer S, Wichert SP, Patzig J, Werner HB. Systematic approaches to central nervous system myelin. Cell Mol Life Sci 2012;69:2879–94.

[20] Trapp BD, Bernier L, Andrews SB, Colman DR. Cellular and subcellular distribution of 2′,3′-cyclic nucleotide 3′-phosphodiesterase and its mRNA in the rat central nervous system. J Neurochem 1988;51:859–68.

[21] Schwer B, Aronova A, Ramirez A, Braun P, Shuman S. Mammalian 2′,3′ cyclic nucleotide phosphodiesterase (CNP) can function as a tRNA splicing enzyme *in vivo*. RNA 2008;14:204–10.

[22] Sogin DC. 2′,3′-Cyclic NADP as a substrate for 2′,3′-cyclic nucleotide 3′-phosphohydrolase. J Neurochem 1976;27:1333–7.

[23] Koonin EV, Gorbalenya AE. Related domains in yeast tRNA ligase, bacteriophage T4 polynucleotide kinase and RNA ligase, and mammalian myelin 2′,3′-cyclic nucleotide phosphohydrolase revealed by amino acid sequence comparison. FEBS Lett 1990;268:231–4.

[24] Bifulco M, Laezza C, Stingo S, Wolff J. 2′,3′-Cyclic nucleotide 3′-phosphodiesterase: a membrane-bound, microtubule-associated protein and membrane anchor for tubulin. Proc Natl Acad Sci USA 2002;99:1807–12.

[25] Braun PE, De Angelis D, Shtybel WW, Bernier L. Isoprenoid modification permits 2′,3′-cyclic nucleotide 3′-phosphodiesterase to bind to membranes. J Neurosci Res 1991;30:540–4.

[26] De Angelis DA, Braun PE. Isoprenylation of brain 2′,3′-cyclic nucleotide 3′-phosphodiesterase modulates cell morphology. J Neurosci Res 1994;39:386–97.

[27] De Angelis DA, Braun PE. 2′,3′-Cyclic nucleotide 3′-phosphodiesterase binds to actin-based cytoskeletal elements in an isoprenylation-independent manner. J Neurochem 1996;67:943–51.

[28] Lee J, Gravel M, Zhang R, Thibault P, Braun PE. Process outgrowth in oligodendrocytes is mediated by CNP, a novel microtubule assembly myelin protein. J Cell Biol 2005;170:661–73.

[29] Gravel M, Robert F, Kottis V, Gallouzi IE, Pelletier J, Braun PE. 2′,3′-Cyclic nucleotide 3′-phosphodiesterase: a novel RNA-binding protein that inhibits protein synthesis. J Neurosci Res 2009;87:1069–79.

[30] Myllykoski M, Itoh K, Kangas SM, Heape AM, Kang SU, Lubec G, et al. The N-terminal domain of the myelin enzyme 2′,3′-cyclic nucleotide 3′-phosphodiesterase: direct molecular interaction with the calcium sensor calmodulin. J Neurochem 2012;123:515–24.

[31] Edgar JM, McLaughlin M, Werner HB, McCulloch MC, Barrie JA, Brown A, et al. Early ultrastructural defects of axons and axon-glia junctions in mice lacking expression of Cnp1. Glia 2009;57:1815–24.

[32] Lappe-Siefke C, Goebbels S, Gravel M, Nicksch E, Lee J, Braun PE, et al. Disruption of Cnp1 uncouples oligodendroglial functions in axonal support and myelination. Nat Genet 2003;33:366–74.

[33] Rasband MN, Tayler J, Kaga Y, Yang Y, Lappe-Siefke C, Nave KA, et al. CNP is required for maintenance of axon-glia interactions at nodes of Ranvier in the CNS. Glia 2005;50:86–90.

[34] Verrier JD, Jackson TC, Bansal R, Kochanek PM, Puccio AM, Okonkwo DO, et al. The brain *in vivo* expresses the 2′,3′-cAMP-adenosine pathway. J Neurochem 2012;122:115–25.

[35] Raines RT. Ribonuclease A. Chem Rev 1998;98:1045–66.

[36] Heaton PA, Eckstein F. Diastereomeric specificity of 2′,3′-cyclic nucleotide 3′-phosphodiesterase. Nucleic Acids Res 1996;24:850–3.

[37] Lee J, O'Neill RC, Park MW, Gravel M, Braun PE. Mitochondrial localization of CNP2 is regulated by phosphorylation of the N-terminal targeting signal by PKC: implications of a mitochondrial function for CNP2 in glial and non-glial cells. Mol Cell Neurosci 2006;31:446–62.

[38] Hugli TE, Bustin M, Moore S. Spectrophotometric assay of 2′,3′-cyclic nucleotide 3′-phosphohydrolase: application to the enzyme in bovine brain. Brain Res 1973;58:191–203.

[39] Lee J, Gravel M, Gao E, O'Neill RC, Braun PE. Identification of essential residues in 2′,3′-cyclic nucleotide 3′-phosphodiesterase: chemical modification and site-directed mutagenesis to investigate the role of cysteine and histidine residues in enzymatic activity. J Biol Chem 2001;276:14804–13.

[40] Myllykoski M, Kursula P. Expression, purification, and initial characterization of different domains of recombinant mouse 2′,3′-cyclic nucleotide 3′-phosphodiesterase, an enigmatic enzyme from the myelin sheath. BMC Res Notes 2010;3:12.

[41] Arn EA, Abelson JN. The 2′-5′ RNA ligase of *Escherichia coli*. Purification, cloning, and genomic disruption. J Biol Chem 1996;271:31145–53.

[42] Molina-Serrano D, Marques J, Nohales MA, Flores R, Daros JA. A chloroplastic RNA ligase activity analogous to the bacterial and archaeal 2′-5′ RNA ligase. RNA Biol 2012;9:326–33.

[43] Wang LK, Shuman S. Structure–function analysis of yeast tRNA ligase. RNA 2005;11:966–75.

[44] Englert M, Sheppard K, Gundllapalli S, Beier H, Soll D. Branchiostoma floridae has separate healing and sealing enzymes for 5′-phosphate RNA ligation. Proc Natl Acad Sci USA 2010;107:16834–9.

[45] Wilson SJ, Schoggins JW, Zang T, Kutluay SB, Jouvenet N, Alim MA, et al. Inhibition of HIV-1 particle assembly by 2′,3′-cyclic-nucleotide 3′-phosphodiesterase. Cell Host Microbe 2012;12:585–97.

[46] Zhao L, Jha BK, Wu A, Elliott R, Ziebuhr J, Gorbalenya AE, et al. Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. Cell Host Microbe 2012;11:607–16.

[47] Roth-Cross JK, Stokes H, Chang G, Chua MM, Thiel V, Weiss SR, et al. Organ-specific attenuation of murine

hepatitis virus strain A59 by replacement of catalytic residues in the putative viral cyclic phosphodiesterase ns2. J Virol 2009;83:3743–53.

[48] Genschik P, Hall J, Filipowicz W. Cloning and characterization of the *Arabidopsis* cyclic phosphodiesterase which hydrolyzes ADP-ribose 1″,2″-cyclic phosphate and nucleoside 2′,3′-cyclic phosphates. J Biol Chem 1997;272:13211–9.

[49] Drummond GI, Eng DY, McIntosh CA. Ribonucleoside 2′,3′-cyclic phosphate diesterase activity and cerebroside levels in vertebrate and invertebrate nerve. Brain Res 1971;28:153–63.

[50] Gould RM, Morrison HG, Gilland E, Campbell RK. Myelin tetraspan family proteins but no non-tetraspan family proteins are present in the ascidian (*Ciona intestinalis*) genome. Biol Bull 2005;209:49–66.

[51] Kurihara T, Tsukada Y. 2′,3′-Cyclic nucleotide 3′-phosphohydrolase in the developing chick brain and spinal cord. J Neurochem 1968;15:827–32.

[52] Trapp BD, McIntyre LJ, Quarles RH, Nonaka G, Moser A, Moser HW, et al. Biochemical characterization of myelin isolated from the central nervous system of *Xenopus* tadpoles. J Neurochem 1980;34:1241–6.

[53] Takezaki N, Figueroa F, Zaleska-Rutczynska Z, Takahata N, Klein J. The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of forty-four nuclear genes. Mol Biol Evol 2004;21:1512–24.

[54] Trams EG, Brown EA. The activity of 2′,3′-cyclic adenosine monophosphate 3′-phosphoesterhydrolase in elasmobranch and teleost brain. Comp Biochem Physiol Part B Biochem Mol Biol 1974;48:185–9.

[55] Hammarstrom M, Woestenenk EA, Hellgren N, Hard T, Berglund H. Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. J Struct Funct Genomics 2006;7:1–14.

[56] Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expression Purif 2005;41:207–34.

[57] Kabsch W. XDS. Acta Crystallogr Sect D Biol Crystallogr 2010;66:125–32.

[58] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr 2007;40:658–74.

[59] Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D Biol Crystallogr 2010;66:213–21.

[60] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr Sect D Biol Crystallogr 2010;66:486–501.

[61] McNicholas S, Potterton E, Wilson KS, Noble ME. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr Sect D Biol Crystallogr 2011;67:386–94.

[62] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12.

[63] Unni S, Huang Y, Hanson RM, Tobias M, Krishnan S, Li WW, et al. Web servers and services for electrostatics calculations with APBS and PDB2PQR. J Comput Chem 2011;32:1488–91.

[64] Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure

validation for proteins and nucleic acids. Nucleic Acids Res 2007;35:W375–83.

[65] Fischer H, Oliveira Neto Md, Napolitano HB, Polikarpov I, Craievich AF. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. J Appl Crystallogr 2009;43:101–9.

[66] Nielsen SS, Toft KN, Snakenborg D, Jeppesen MG, Jacobsen JK, Vestergaard B, et al. BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. J Appl Crystallogr 2009;42:959–64.

[67] Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, et al. New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr 2012;45:342–50.

[68] Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 1992;25:495–503.

[69] Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J Appl Crystallogr 2009;42:342–6.

[70] Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. Biophys J 2001;80:2946–53.

[71] Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. J Appl Crystallogr 2003;36:860–4.

[72] Zhu H, Smith P, Wang LK, Shuman S. Structure–function analysis of the 3′ phosphatase component of T4 polynucleotide kinase/phosphatase. Virology 2007;366:126–36.

[73] Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009;4:363–71.

[74] Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res 2004;32:W668–73.

[75] Johnson WC. Analyzing protein circular dichroism spectra for accurate secondary structures. Proteins 1999;35:307–12.

[76] Lees JG, Miles AJ, Wien F, Wallace BA. A reference database for circular dichroism spectroscopy covering fold and secondary structure space. Bioinformatics 2006;22:1955–62.

[77] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455–61.

[78] Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminf 2012;4:17.

[79] Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2010;38:D5–16.

[80] The Uniprot Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 2012;40:D71–5.

[81] Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 2008;9:286–98.

[82] Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009;25:1189–91.