

Structural and Enzymatic Characterization of the Streptococcal ATP/Diadenosine Polyphosphate and Phosphodiester Hydrolase Spr1479/SapH^{*[5]}

Received for publication, February 6, 2011, and in revised form, August 4, 2011. Published, JBC Papers in Press, August 23, 2011, DOI 10.1074/jbc.M111.228585

Yong-Liang Jiang^{†1}, Jun-Wei Zhang^{†1}, Wei-Li Yu[‡], Wang Cheng[‡], Chen-Chen Zhang[‡], Cecile Frolet[§], Anne-Marie Di Guilmi[§], Thierry Vernet[§], Cong-Zhao Zhou^{‡2}, and Yuxing Chen^{†3}

From the [†]Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China and the [§]Laboratoire d'Ingénierie des Macromolécules, Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075, 41 rue Jules Horowitz, 38027 Grenoble, France

Spr1479 from *Streptococcus pneumoniae* R6 is a 33-kDa hypothetical protein of unknown function. Here, we determined the crystal structures of its apo-form at 1.90 Å and complex forms with inorganic phosphate and AMP at 2.30 and 2.20 Å, respectively. The core structure of Spr1479 adopts a four-layer $\alpha\beta\beta\alpha$ -sandwich fold, with Fe³⁺ and Mn²⁺ coordinated at the binuclear center of the active site (similar to metallophosphoesterases). Enzymatic assays showed that, in addition to phosphodiesterase activity for bis(*p*-nitrophenyl) phosphate, Spr1479 has hydrolase activity for diadenosine polyphosphate (Ap_{*n*}A) and ATP. Residues that coordinate with the two metals are indispensable for both activities. By contrast, the streptococcus-specific residue Trp-67, which binds to phosphate in the two complex structures, is indispensable for the ATP/Ap_{*n*}A hydrolase activity only. Moreover, the AMP-binding pocket is conserved exclusively in all streptococci. Therefore, we named the protein SapH for streptococcal ATP/Ap_{*n*}A and phosphodiester hydrolase.

The commensal Gram-positive bacterium *Streptococcus pneumoniae* is a major human pathogen that is responsible for many infectious diseases such as pneumoniae, sepsis, otitis media, and meningitis (1). Beyond its capsule, the virulence factors of this pathogen are not fully understood. Since the release of the complete genome sequences of virulent (TIGR4) and non-virulent (R6) isolates of *S. pneumoniae*, the identification of virulence factors and their characterization as new potential antibiotic and/or vaccine targets have been in progress (2–4). In addition to the surface-exposed proteins, a large array of hypothetical proteins have been proposed to contribute to the streptococcal pathogenesis (4, 5).

* This work was supported by Ministry of Science and Technology of China Grant 2009CB918804 and National Natural Science Foundation of China Grant 30870488.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1–S3 and Tables S1–S2.

The atomic coordinates and structure factors (codes 3QFM, 3QFN, and 3QFO) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

[†] Both authors contributed equally to this work.

[‡] To whom correspondence may be addressed. Tel./Fax: 86-551-360-0406; E-mail: zcz@ustc.edu.cn.

[§] To whom correspondence may be addressed. Tel./Fax: 86-551-360-0406; E-mail: cyxing@ustc.edu.cn.

The 33-kDa hypothetical protein Spr1479 from *S. pneumoniae* R6 was predicted to be a virulence factor. It belongs to the core genome, and its orthologs have also been found in other streptococci (5). Recently, the *Mycobacterium tuberculosis* cyclic nucleotide phosphodiesterase Rv0805, a distant homolog of Spr1479, was reported to play a key role in the pathogenicity of mycobacteria not only by hydrolyzing bacterial cAMP but also by moonlighting as a protein that can alter cell wall functioning (6). A sequence homology search against the Pfam Database (7) indicated that Spr1479 belongs to the calcineurin-like phosphoesterase (also referred to as metallophosphoesterase (MPP)⁴) superfamily. Members of this superfamily, which are widespread in all organisms, include protein phosphoserine phosphatases, nucleotidases, sphingomyelin phosphodiesterases, 2',3'-cAMP phosphodiesterases, and nucleases, as well as human VPS29 (vacuolar protein-sorting protein 29). Previous structures have shown that these members share a common $\alpha\beta\beta\alpha$ -sandwich fold, with an active site composed of two metal ions (typically manganese, iron, or zinc) that are coordinated in an octahedral geometry with the residues His, Asp, and Asn. This highly conserved active site involves a common catalytic mechanism: a hydroxide anion, activated by the binuclear metals, attacks the phosphorus atom of the substrate (8–12).

A sequence homology search (blast.ncbi.nlm.nih.gov/Blast.cgi) of Spr1479 against the Swiss-Prot Database (ExPASy) gave only hits of low identity. The top hit is an uncharacterized protein, MJ0912 (accession number Q58322) from *Methanococcus jannaschii*, which has a 27% sequence identity of over 253 residues with Spr1479. Because the next closest hits are ApaH-related symmetric diadenosine 5',5'''-P¹,P⁴-tetrakisphosphate (Ap₄A) hydrolases (EC 3.6.1.41; 21% sequence identity), which hydrolyze one Ap₄A into two ADP molecules (13), we analyzed the *in vitro* hydrolase activity of Spr1479 for Ap₄A, as well as Ap₃A and Ap₅A. Against our expectation, Spr1479 catalyzed cleavage of these molecules asymmetrically with AMP as the common product. This resembles the activity of Nudix (nucleoside diphosphate linked to X) proteins, which catalyze the hydrolysis of NDP linked to another moiety such as NTP,

⁴ The abbreviations used are: MPP, metallophosphoesterase; Ap₄A, diadenosine 5',5'''-P¹,P⁴-tetrakisphosphate; bis-pNPP, bis(*p*-nitrophenyl) phosphate; SapH-FL, full-length SapH; r.m.s.d., root mean square deviation.

Ap_nA, NDP-sugar, NADH, and coenzyme A, etc. (14). Members of this superfamily have an $\alpha\beta\alpha$ -sandwich fold (15–19) and the Nudix signature motif GX₅EX₇REUXEEXGU (where X stands for any residues, and U is a bulky aliphatic residue, usually Ile, Leu, or Val) (20). This motif forms a loop-helix-loop structure that is involved in substrate binding and catalysis (21). It has been proposed that the physiological function of Nudix proteins is “housecleaning” to eliminate potentially toxic nucleotide metabolites from cells and to regulate the concentrations of NDP derivatives (20).

To decipher the molecular function of Spr1479, we solved its crystal structures in the apo-form at 1.90 Å and complex forms with inorganic phosphate and AMP at 2.30 and 2.20 Å, respectively. The core structure and binuclear active site of Spr1479 resemble those of the MPP superfamily. In addition to phosphodiesterase activity for the generic substrate bis(*p*-nitrophenyl) phosphate (bis-*p*NPP), Spr1479 has hydrolase activity for Ap_nA and ATP. Thus, Spr1479 might function not only as a phosphodiesterase but also as a housecleaning protein to keep the homeostasis of intracellular nucleotides.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification of SapH and Its Mutants—The coding regions of Spr1479/SapH (residues 1–262) and full-length SapH (SapH-FL; residues 1–282) were amplified from genomic DNA of *S. pneumoniae* R6 and individually cloned into a pET28a-derived vector and overexpressed in the *Escherichia coli* Rosetta(DE3) strain (Novagen, Madison, WI) using 2×YT culture medium (5 g of NaCl, 16 g of Bacto-Tryptone, and 10 g of yeast extract per liter). A hexahistidine tag was added to the N terminus of each of the recombinant proteins. The cells were grown at 37 °C to an A_{600 nm} of 0.6. Expression of the recombinant protein was induced with 0.2 mM isopropyl β-D-thiogalactoside for another 20 h at 16 °C before harvesting. Cells were collected by centrifugation at 4000 × *g* for 20 min and resuspended in 40 ml of lysis buffer (20 mM Tris-Cl (pH 8.0) and 200 mM NaCl). After 2.5 min of sonication and centrifugation at 12,000 × *g* for 25 min, the supernatant containing the target protein was collected and loaded onto a nickel-nitrilotriacetic acid column (GE Healthcare) equilibrated with the binding buffer (20 mM Tris-Cl (pH 8.0) and 200 mM NaCl). The target protein was eluted with 300 mM imidazole and loaded onto a Superdex 200 column (GE Healthcare) in 20 mM Tris-Cl (pH 8.0) and 100 mM NaCl. Fractions containing the target protein were combined and concentrated to 10 mg/ml for crystallization. Samples for enzymatic activity assays were collected at the highest peak fractions without concentration. The purity of protein was assessed by electrophoresis, and the protein sample was stored at –80 °C.

Site-directed mutagenesis was performed using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) with the plasmid encoding wild-type SapH as the template. The mutant proteins were expressed, purified, and stored in the same manner as used for the wild-type protein.

Crystallization, Data Collection, and Processing—Crystals of SapH were grown at 16 °C using hanging drop vapor-diffusion techniques by mixing 2 μl of 10 mg/ml protein sample with an equal volume of reservoir solution (25% isopropyl alcohol, 0.1 M

HEPES (pH 7.5), and 0.2 M NH₄Ac). Crystals appeared in 1 week. Before data collection, crystals were soaked in cryoprotectant solution (reservoir solution supplemented with 25% glycerol). The iodine derivative crystals were obtained by soaking SapH crystals in cryoprotectant solution containing 300 mM KI for ~15 s. The crystals of SapH-PO₄ and SapH-AMP were obtained by soaking the SapH crystals in 500 mM NaH₂PO₄ for ~5 s and in 30 mM AMP for ~2 min, respectively. All diffraction data were collected at 100 K in a liquid nitrogen stream using beamline 17U with an MX-225 CCD detector (Marresearch GmbH, Norderstedt, Germany) at the Shanghai Synchrotron Radiation Facility. All diffraction data were integrated and scaled with the program HKL2000 (22).

Structure Determination and Refinement—The structure of SapH was determined by the single wavelength anomalous dispersion phasing technique (23) with the iodine anomalous signal using the program phenix.solve implemented in PHENIX (24). The initial model was built automatically with the program AutoBuild in PHENIX. The resultant model was subsequently used as a search model against the 1.90 Å data of SapH. Using the SapH structure as the search model, the structures of SapH-PO₄ and SapH-AMP were determined by the molecular replacement method with the program MOLREP (25) implemented in CCP4i (26). All initial models were refined using the maximum likelihood method implemented in REFMAC5 (27) as part of the CCP4i program suite and rebuilt interactively using the program COOT (28). The final models were evaluated with the programs MOLPROBITY (29) and PROCHECK (30). Crystallographic parameters are listed in Table 1. All structure figures were prepared with PyMOL (31).

Determination of Active-site Metals—Atomic absorption spectroscopy (Atomscan Advantage, Thermo Ash Jarrell Corp.) was performed to determine the metal content. Prior to analysis, purified SapH-FL in 20 mM Tris-Cl (pH 8.5) and 150 mM NaCl was concentrated to ~8 mg/ml.

Enzymatic Assays—All enzymatic assays were performed at 37 °C in buffer containing 100 mM Tris-Cl (pH 7.0), 0.2 mM FeCl₃, and 0.2 mM MnCl₂. Using a DU800 spectrophotometer (Beckman Coulter), phosphatase and phosphodiesterase activities were measured with *p*NPP and bis-*p*NPP (Sigma) as substrates, respectively, by following the absorption increase at 405 nm. The kinetic parameters of wild-type SapH and its mutants were measured using bis-*p*NPP as the substrate to a standard curve of 4-nitrophenol. Reactions were initiated by the addition of SapH or its mutants.

The hydrolysis of nucleotides/derivatives (Sigma) was measured by HPLC (Agilent 1200 series). SapH and its mutants were incubated with a range of substrate concentrations in a volume of 25 μl. The reaction was terminated after 15 min of incubation by the addition of 50 μl of 20 mM NH₄H₂PO₄ (pH 6.2). As a control, an assay mixture without any protein added was used. NMP standards were quantified by HPLC analysis using a series of concentrations ranging from 0.1 to 5 mM. 20 mM NH₄H₂PO₄ was used for equilibration of the column (Eclipse XDB-C18 column (4.6 × 150 mm), Agilent) and separation of the components at a flow rate of 1 ml/min. Samples were injected in a volume of 10 μl. The parameters *K_m* and *k_{cat}* were calculated by nonlinear fitting to the Michaelis-Menten equation using the

TABLE 1
Crystal parameters, data collection, and structure refinement

	<i>SapH-KI</i>	<i>SapH</i>	<i>SapH-PO₄</i>	<i>SapH-AMP</i>
Data collection				
Space group	<i>P2₁2₁2₁</i>	<i>P2₁2₁2₁</i>	<i>P2₁2₁2₁</i>	<i>P2₁2₁2₁</i>
Unit cell (a, b, c (Å); α, β, γ)	57.98, 75.11, 149.51; 90.00°	56.56, 75.10, 148.80; 90.00°	58.88, 74.94, 149.33; 90.00°	57.65, 74.94, 149.13; 90.00°
Resolution range (Å)	50.00–2.90 (3.00–2.90) ^a	30.00–1.90 (1.97–1.90)	30.00–2.30 (2.38–2.30)	50.00–2.20 (2.24–2.20)
Unique reflections	15,110 (1456)	50,589 (4756)	28,781 (2920)	33,388 (1572)
Completeness (%)	99.9 (99.1)	99.4 (94.9)	95.0 (98.5)	99.3 (95.8)
$\langle I/\sigma(I) \rangle$	29.7 (15.6)	21.8 (3.8)	22.2 (3.6)	26.1 (4.3)
R _{merge} (%) ^b	9.0 (17.2)	8.2 (39.5)	7.4 (51.7)	6.9 (28.4)
Average redundancy	11.4 (9.8)	6.7 (5.6)	7.0 (6.8)	6.7 (5.8)
Structure refinement				
Resolution range (Å)		30.00–1.90	30.00–2.30	50.00–2.20
R-factor/R _{free} (%) ^{c,d}		19.5/22.9	19.7/25.4	18.5/24.4
No. of protein atoms		4140	4046	4066
No. of water atoms		284	211	275
r.m.s.d. bond lengths (Å) ^e		0.013	0.012	0.010
r.m.s.d. bond angles		1.309°	1.324°	1.235°
Mean B-factors (Å ²)		25.4	35.4	30.5
Ramachandran plot (residues, %) ^f				
Most favored		96.2	96.4	96.6
Additionally allowed		3.4	3.2	3.0
Outliers		0.4	0.4	0.4
Protein Data Bank code		3QFM	3QFN	3QFO

^a The values in parentheses refer to statistics in the highest bin.

^b $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the intensity of an observation and $\langle I(hkl) \rangle$ is the mean value for its unique reflection. Summations are over all reflections.

^c $R\text{-factor} = \frac{\sum_h |F_o(h) - F_c(h)|}{\sum_h F_o(h)}$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.

^d R_{free} was calculated with 5% of the data excluded from the refinement.

^e r.m.s.d. from ideal values.

^f Categories were defined by MOLPROBITY.

program Origin 7.5. Three independent assays were performed to calculate the means \pm S.D. for all K_m and k_{cat} values.

RESULTS

Overall Structure of *Spr1479/SapH*—We initially found crystals of full-length *Spr1479* (*SapH-FL*) but could not optimize them to a high diffraction quality. Limited proteolysis combined with LC-MS enabled us to define a relatively stable region of residues 1–262 (termed *SapH*). We overexpressed this region and subjected it to crystallization, and the resulting crystals were diffracted to a resolution of 1.90 Å.

An asymmetric unit contains a dimer of *SapH* with an interface area of ~ 1700 Å². Met-3–Leu-260 in subunit A and Met-3–His-252 in subunit B are well fitted in the final model. The two subunits are very similar, with an overall root mean square deviation (r.m.s.d.) of 0.15 Å over 210 C α atoms (Fig. 1A). Size-exclusion chromatography also confirmed the existence of *SapH-FL* and *SapH* as a dimer in solution (data not shown). The two five-stranded β -sheets of subunit A are aligned antiparallel to their counterparts in subunit B, forming two continuous 10-stranded mixed β -sheets, which are perpendicular to the dimer-related non-crystallographic 2-fold axis (Fig. 1A). Each subunit adopts a compact core structure with a four-layer $\alpha\beta\beta\alpha$ -sandwich fold: two five-stranded β -sheets sandwiched by helices $\alpha 1$ and $\alpha 2$ on one side and helix $\alpha 6$ on the opposite side (Fig. 1B). One β -sheet is composed of strands $\beta 4$ – $\beta 8$, whereas the facing β -sheet consists of strands $\beta 1$ – $\beta 3$, $\beta 9$, and $\beta 10$ (Fig. 1B). Beyond the core structure, *SapH* has a cap of five α -helices ($\alpha 3$ – $\alpha 5$, $\alpha 7$, and $\alpha 8$) packing against the active site. Among these five α -helices, $\alpha 3$ – $\alpha 5$ (Asn-66–His-109) form one side wall of the active site, whereas the other two α -helices $\alpha 7$ and $\alpha 8$ (Val-225–Asn-249) at the C terminus might function as a scaffold. A Dali search using the five-helix subdomain

revealed no hit of a similar domain. Compared with the core domain, the secondary structural elements of this additional subdomain appear to be more variable in length and structure, which may be responsible for the versatile substrate specificities among the MPP superfamily members.

The core structures of subunits A and B are packed against each other to form a dimer interface, which consists mainly of the backbones of strands $\beta 7$ and $\beta 10$ from both subunits. Specifically, the interface is composed of hydrogen bonds between Val-214 O α and Arg-220' N α , Met-216 N α and Phe-218' O α , Met-216 O α and Phe-218' N α , Phe-218 N α and Met-216' O α , Phe-218 O α and Met-216' N α , and Arg-220 N α and Val-214' O α from strand $\beta 10$ and between Leu-170 N α and Leu-170' O α and Leu-170 O α and Leu-170' N α from strand $\beta 7$ (the residues in subunit B are labeled with a prime) (Fig. 1C). In addition to these backbone hydrogen bonds, the carboxyl group of Asp-209 (Asp-209') forms two hydrogen bonds with N $\eta 1$ and N $\eta 2$ of Arg-200' (Arg-200), further stabilizing the dimer.

Binuclear Metals at the Active Site—During the model building and refinement process, two electron density peaks were outstanding in the $F_o - F_c$ Fourier difference map (at the 30σ level), indicating the presence of metals. The brown color of concentrated protein in solution indicated the presence of iron. Atomic absorption spectroscopy showed that iron and manganese were present at molarity ratios to the protein of 0.7 and 0.8, respectively. X-ray diffraction experiments at various wavelengths were done to confirm the chemical identity of these metals. We collected data at a wavelength of 1.74 Å and found two strong anomalous peaks of 31.0 and 30.7 in the structure model-phased anomalous Fourier difference maps. This result excluded the presence of zinc, copper, nickel, and cobalt, which have a theoretical absorption K-edge below 1.61 Å. To further

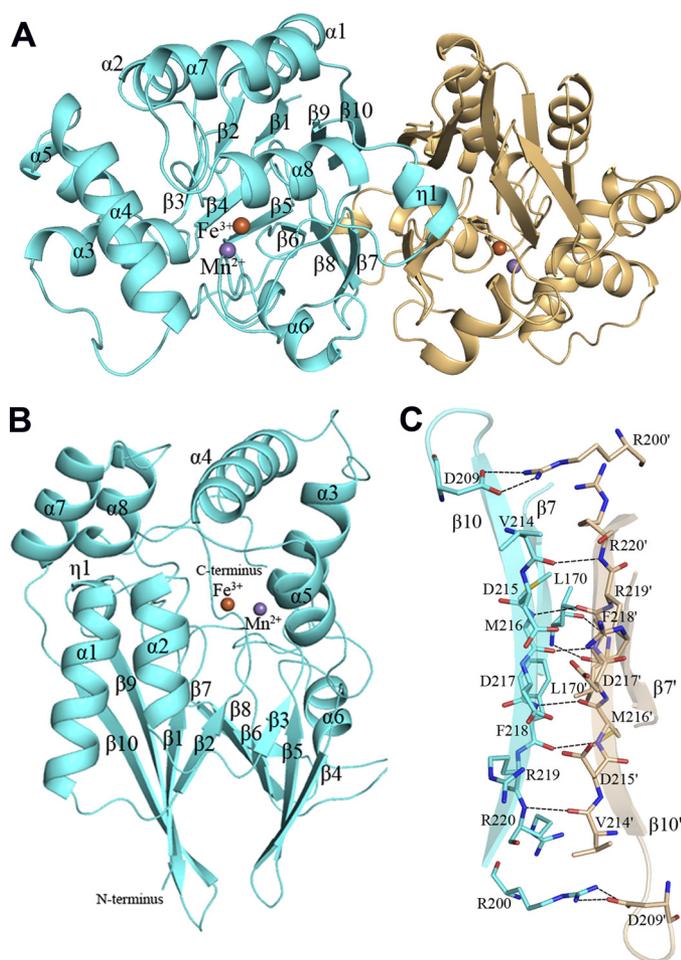


FIGURE 1. Overall structure of SapH. A and B, schematic representations of the dimer and monomer, respectively. Subunits A and B are shown in cyan and orange, respectively. The secondary structural elements are labeled sequentially. The metal ions at the active site are shown in purple for Mn^{2+} and orange for Fe^{3+} . C, dimer interface. The residues are shown as cyan and orange sticks for subunits A and B, respectively. The hydrogen bonds are denoted as dashed lines. Residues and strands from subunit B are labeled with a prime.

distinguish iron from manganese, we collected data at a wavelength of 1.80 Å and assigned peaks of 16.9 to iron and 25.0 to manganese. (The theoretical absorption K-edges for iron and manganese are 1.74 and 1.90 Å, respectively.) A significant anomalous signal of the iron site indicates that this site might be also mixed with manganese. Thus, the two sites should have mixed occupancies of Fe^{3+} and Mn^{2+} , respectively (Fig. 2A). The following activity assays also confirmed that the binuclear metals are Fe^{3+} and Mn^{2+} .

The two metals are 3.4 and 3.3 Å away from each other in subunits A and B, respectively. Fe^{3+} and coordinating residues Asp-11, His-13, Asp-39, and His-166, together with Mn^{2+} and coordinating residues Asp-39, Asn-66, His-128, and His-164, constitute an octahedral structure of the active site (Fig. 2A). A planar water molecule (designated Wat1) forms coordination bonds with Fe^{3+} and Mn^{2+} and a hydrogen bond with the carbonyl oxygen of His-164. This binuclear active site is located at the interface between the core structure and the cap of α -helices.

Comparative Structure Analysis—Overall structure comparison of SapH with structures in the Protein Data Bank using the

Dali server gave 188 hits for 41 unique proteins with a Z-score higher than 10.0; these proteins were all members of the MPP superfamily. The top hit was a hypothetical protein from *Pyrococcus furiosus* (Protein Data Bank code 1NNW; Z-score of 25.7, r.m.s.d. of 2.6 Å over 234 C α atoms), followed by *E. coli* phosphodiesterase YfcE (code 1SU1) (32) and *M. jannaschii* phosphodiesterase MJ0936 (code 1S3M) (33). The other hits included human VPS29 (code 1W24) (34) and Ser/Thr phosphatase 2B (code 1AUI) (35), symmetric Ap₄A hydrolases from *Trypanosoma brucei* (code 2QJC) and *Shigella flexneri* (code 2DFJ) (36), and cyclic nucleotide phosphodiesterase Rv0805 from *M. tuberculosis* (code 2HY1) (37). Most of these proteins have similar molecular functions (phosphodiesterase, phosphatase, nuclease, or nucleotidase), although two are symmetric Ap₄A hydrolases. Moreover, the binuclear active site of SapH can be closely superimposed onto that of *M. tuberculosis* cyclic nucleotide phosphodiesterase Rv0805 (Fig. 2B) (37), despite the fact that SapH and Rv0805 have a sequence identity of only 17% and markedly different structures, with an r.m.s.d. of 3.3 Å over 149 C α atoms. The highly conserved $\alpha\beta\alpha$ -sandwich core structure and binuclear active site strongly implied that SapH might act as a phosphodiesterase, phosphatase, or Ap₄A hydrolase.

Phosphodiesterase Activity—We first tested the enzymatic activity of SapH for generic phosphomonoesterase or phosphodiesterase substrates (*p*NPP or bis-*p*NPP) in the presence of 0.2 mM Fe^{3+} and 0.2 mM Mn^{2+} . No activity for *p*NPP was detected (data not shown). By contrast, both SapH-FL and SapH showed considerable activity for bis-*p*NPP. The K_m and k_{cat} values were 1.76 ± 0.02 mM and 5.70 ± 0.12 s⁻¹, respectively, for SapH-FL and 2.46 ± 0.15 mM and 5.51 ± 0.22 s⁻¹, respectively, for SapH (supplemental Table S1). These results suggested that SapH is a phosphodiesterase.

We further systematically screened all of the reported physiological substrates of metallophosphodiesterase. These substrates fall into three groups: cyclic nucleotides, nucleic acids, and phospholipids. We tested the following representative substrates of phospholipases: 2',3'-cAMP (Sigma), 3',5'-cAMP (Sigma), double- or single-stranded DNA and RNA, and the generic substrate *p*-nitrophenylphosphorylcholine (Sigma). However, SapH-FL showed no catalytic activity for any of the above substrates. This indicates that SapH-FL most likely performs a novel phosphodiesterase activity for a unique physiological substrate, or the activity of SapH-FL for the above-listed substrates needs the assistance of an unknown partner. Similar cases of phosphodiesterases with unknown physiological substrate have also been reported previously (32, 33).

The metal-coordinating residues have a crucial role in SapH activity. Mutation of Asp-39, His-13, or His-128 to Ala completely abolished the phosphodiesterase activity. Activity was elevated by the addition of 0.2 mM Fe^{3+} and/or Mn^{2+} and somewhat inhibited by the addition of Mg^{2+} , Ni^{2+} , Ca^{2+} , Co^{2+} , Zn^{2+} , or Fe^{2+} individually (supplemental Fig. S1). Moreover, the addition of 5 mM EDTA did not change the activity (supplemental Fig. S1), indicating that the co-purified metals have a very high affinity for SapH.

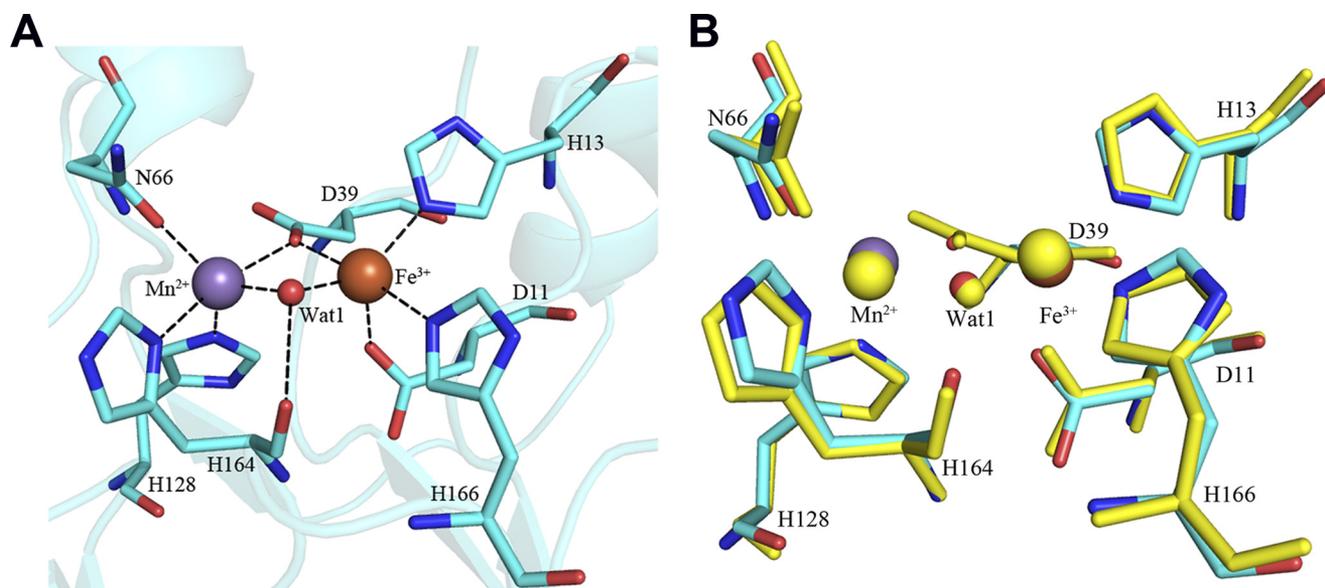


FIGURE 2. **Binuclear active site.** A, Mn^{2+} - and Fe^{3+} -coordinating residues. The residues are shown as sticks, and the planar water molecule Wat1 is shown as a red sphere. B, superposition of the active sites of SapH (cyan) and cyclic nucleotide phosphodiesterase Rv0805 (yellow).

Nucleotide/Derivative Hydrolase Activity—In addition to phosphodiesterases, the Dali search returned two hits of ApaH-related symmetric Ap_4A hydrolases from *T. brucei* and *S. flexneri* (r.m.s.d. of 3.1 Å over 177 C α atoms and 3.0 Å over 184 C α atoms and sequence identities of 21 and 20%, respectively). Thus, we tested the hydrolase activity of SapH-FL for Ap_4A , as well as Ap_3A and Ap_5A (Sigma). The results showed that SapH hydrolyzed all three molecules asymmetrically, producing ADP and AMP from Ap_3A , ATP and AMP from Ap_4A , and Ap_4 and AMP from Ap_5A . The k_{cat} values of SapH-FL for the three substrates are comparable, whereas the K_m values are 0.28 mM for Ap_3A , 2.09 mM for Ap_4A , and 1.04 mM for Ap_5A (Table 2). It is worth noting that the K_m value for Ap_3A is comparable with that of T4 bacteriophage NudE.1 (0.41 mM) and *E. coli* Orf186 (0.15 mM), both of which are Nudix proteins (38, 39).

However, the molarity ratio of the product AMP is obviously higher than that of ATP in the reaction with Ap_4A , suggesting that ATP might be further hydrolyzed to AMP. Assays showed that both SapH-FL and SapH had relatively low activity of hydrolyzing ATP to AMP (supplemental Fig. S2). This gave us the idea that SapH-FL should have a broad spectrum of substrate specificity. In consequence, several nucleotides and derivatives (Sigma), including ADP, ATP, ADP-ribose, NAD(H), and NADP(H), were tested. SapH-FL could somewhat hydrolyze ATP, NAD(H), and ADP-ribose but not ADP or NADP(H). The relative activities are 24% (ATP), 13% (ADP-ribose), 10% (NADH), and 2% (NAD⁺) of that of the activity for Ap_3A (supplemental Table S2). Furthermore, nucleotides GTP, CTP, and UTP were also tested. SapH-FL could hydrolyze GTP and CTP at a comparable rate to ATP, whereas little activity for UTP was detected (Table 2 and supplemental Table S2). Despite the relatively lower activity of SapH-FL for these substrates, the K_m value for ATP is \sim 2.37 mM, which is in the range of intracellular ATP concentrations (1–5 mM) (40). Thus, SapH might function as an ATP pyrophosphatase *in vivo*.

TABLE 2
Kinetic parameters of SapH-FL

Substrate	K_m	k_{cat}	k_{cat}/K_m
	mM	s^{-1}	$s^{-1} mM^{-1}$
bis- <i>p</i> NPP	1.76 ± 0.02	5.70 ± 0.12	3.24 ± 0.10
Ap_{3A}	0.28 ± 0.02	15.95 ± 0.90	57.94 ± 2.48
Ap_{4A}	2.09 ± 0.10	14.89 ± 0.35	7.14 ± 0.24
Ap_{5A}	1.04 ± 0.10	11.29 ± 0.27	10.96 ± 1.05
ATP	2.37 ± 0.17	9.91 ± 0.24	4.21 ± 0.28
GTP	2.42 ± 0.32	7.58 ± 0.47	3.17 ± 0.25
CTP	3.00 ± 0.18	23.74 ± 1.28	7.92 ± 0.07

AMP-binding Pocket—The electrostatic potential surface of SapH revealed an extended pocket of two clefts perpendicular to each other, with binuclear metals at the corner. To reveal the substrate-binding mode, we attempted to prepare crystals of SapH in the presence of inorganic phosphate, AMP, and ATP by either crystal soaking or co-crystallization. We also tried crystallizing the inactive W67H mutant of SapH in complex with Ap_4A . At the end, we obtained only the phosphate- and AMP-complexed structures (termed SapH- PO_4 and SapH-AMP, respectively) by soaking SapH crystals with 500 mM NaH_2PO_4 or 30 mM AMP. In SapH- PO_4 , the inorganic phosphate bridges the two metals at the active site and makes hydrogen bonds with the metal-coordinating residues Asn-66, His-164, His-13, and His-166, in addition to Trp-67 and the planar water molecule Wat1 (Fig. 3A). The binding mode is very similar to that of other enzymes in the MPP superfamily, suggesting that SapH also adopts the common catalytic mechanism (9, 12, 41).

In SapH-AMP, one molecule of AMP is deeply buried in the active-site pocket, with the adenine moiety in an *anti*-conformation and the ribose ring in a C₄ *exo*-conformation (Fig. 3B). The adenine and ribose ring are frozen in the inner cleft between loop 10 (between β_5 and α_6) and loop 14 (between β_8 and η_1) via hydrogen bonds and hydrophobic interactions. The O4' of ribose forms a hydrogen bond with Arg-137 N η 2, whereas O3' and O2' form two hydrogen bonds with His-252 N δ 1. The adenine is stabilized in an *anti*-conformation via a

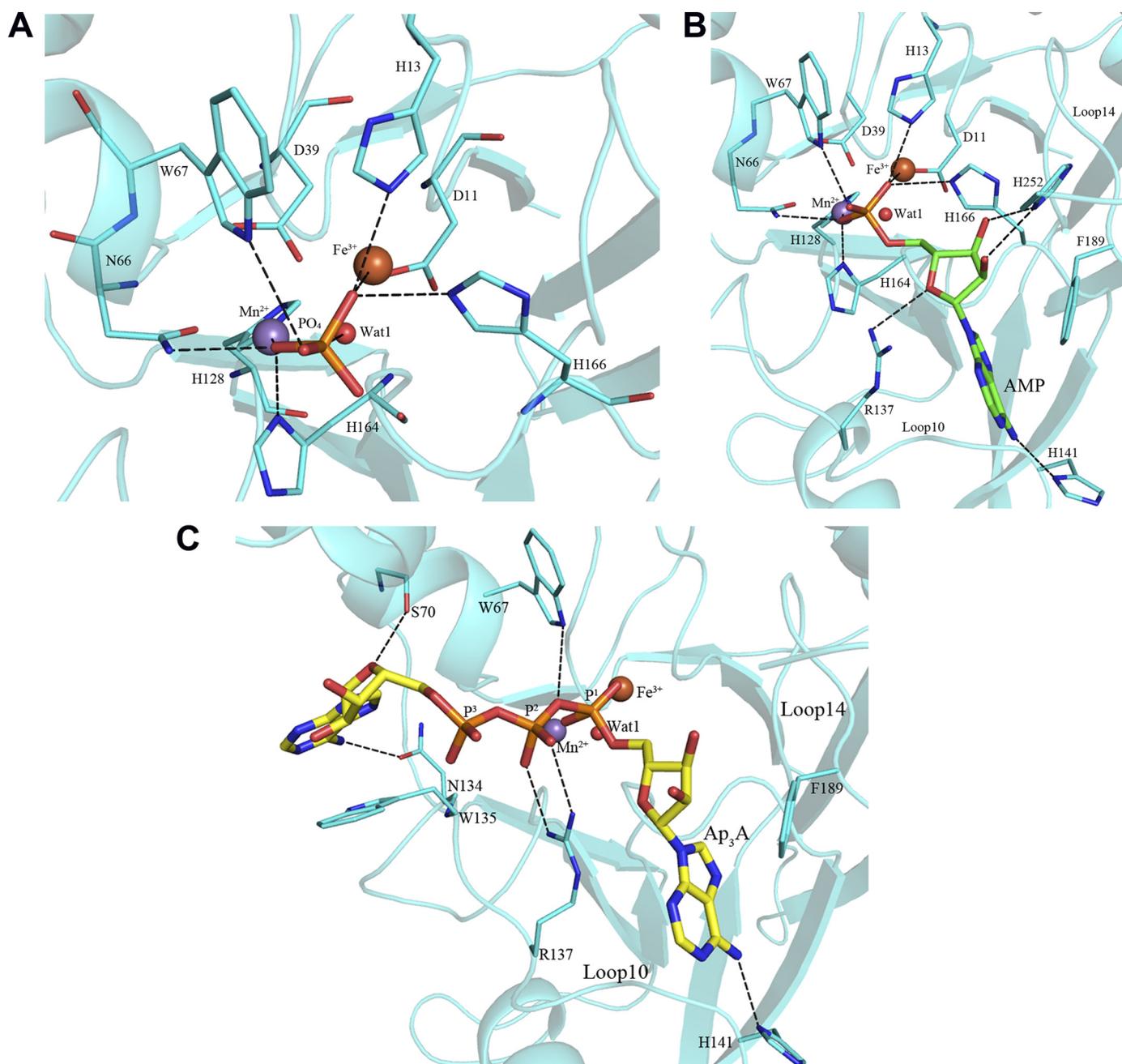


FIGURE 3. **AMP-binding pocket of SapH.** *A*, inorganic phosphate-binding site in SapH-PO₄. The residues are shown as sticks. The hydrogen bonds are shown as dashed lines. *B*, AMP-binding site. The binding pocket is between loops 10 and 14. The binding residues are shown as cyan sticks and AMP as green sticks. The metals and Wat1 are rendered as spheres. *C*, docking model of Ap₃A (yellow for its C atoms) in the active-site pocket.

π - π stacking interaction against Phe-189 and a hydrogen bond with His-141 N δ 1. The phosphate group is bound to the active site in the same manner as the inorganic phosphate in SapH (Fig. 3B).

To reveal the Ap₃A-binding mode, we docked Ap₃A onto SapH using the RosettaDock program (42). Restraints were used to fix the AMP moiety of Ap₃A in a position similar to that in the complex of SapH-AMP. In the model, the inner and outer clefts accommodate the AMP (the first adenine nucleotide and P¹-phosphate group) and ADP (the second adenine nucleotide and P²- and P³-phosphate groups) moieties of Ap₃A, respectively (Fig. 3C). The AMP moiety in the inner cleft could be well superimposed with that of SapH-AMP. The ADP moiety is sta-

bilized in the outer cleft via two hydrogen bonds between the P²-phosphate group and Arg-137. The adenosine moiety is stabilized by π - π stacking against Trp-135 and two hydrogen bonds with Asn-134 O δ 1 and Ser-70 O γ , respectively (Fig. 3C). Multiple-sequence alignment revealed that Ser-70, Asn-134, Trp-135, and Arg-137 at the outer cleft are highly conserved in streptococci. It is worth noticing that the active-site residues Asn-66, Trp-67, and Ser-70 are all from helix α 3 of the five-helix cap.

The phosphate in SapH-PO₄ and SapH-AMP could be closely superimposed with the P¹-phosphate group of the SapH-Ap₃A model. Moreover, this phosphate could be superimposed with that of protein analogs such as dAMP in nuclease

orescence intensity to that of ATP. The equilibrium dissociation constants (K_d) for these three substrates are $443 \pm 12 \mu\text{M}$ (ATP), $198 \pm 64 \mu\text{M}$ (GTP), and $283 \pm 10 \mu\text{M}$ (CTP). These results indicate that CTP and GTP could also bind to the AMP pocket.

DISCUSSION

SapH Is a Novel Member of the MPP Superfamily—To date, all nucleotide/derivative hydrolases of known structure belong to the Nudix hydrolase superfamily (43). However, the core structure of SapH adopts an $\alpha\beta\beta\alpha$ -sandwich fold rather than an $\alpha\beta\alpha$ -sandwich fold with a Nudix signature motif (17, 19, 43). The $\alpha\beta\beta\alpha$ -sandwich fold and binuclear active site of SapH are highly conserved in members of the MPP superfamily, which share a low sequence identity (<20%) with SapH. Among these MPP superfamily members, only *Bacillus subtilis* PrpE (44) and *E. coli* ApaH (13) were found to hydrolyze Ap₄A. However, PrpE is a Tyr-specific phosphatase and an asymmetric Ap₄A hydrolase of unknown three-dimensional structure. *E. coli* ApaH was proved to be a symmetric Ap₄A hydrolase. The structure of its 100% sequence-identical homolog from *S. flexneri* (Protein Data Bank code 2DFJ) also has an $\alpha\beta\beta\alpha$ -sandwich fold with two layers of five-stranded β -sheets sandwiched by α -helices on the two sides, respectively (36). Superposition of SapH onto *S. flexneri* ApaH revealed an r.m.s.d. of 3.0 Å over 184 C α atoms. The core $\alpha\beta\beta\alpha$ -sandwich structure of SapH is quite similar to that of ApaH. Among the helices of the additional subdomain, three helices ($\alpha 3$ – $\alpha 5$) of SapH exhibit structural similarity to the counterparts of ApaH, despite that helix $\alpha 4$ (Pro-84–Glu-99) of SapH is 10 residues longer (supplemental Fig. S3). By contrast, the other two α -helices ($\alpha 7$ and $\alpha 8$) at the C terminus of SapH are not found in ApaH. Alternatively, ApaH has an insertion of three additional α -helices ($\alpha 7$ – $\alpha 9$, Asp-128–Arg-182) located at the opposite side of the active site (supplemental Fig. S3). Moreover, the active site of ApaH is coordinated by two Mn²⁺ ions. Therefore, SapH represents the first structure of a novel nucleotide/derivative hydrolase in the MPP superfamily.

SapH Is a Bifunctional Enzyme Conserved Exclusively in Streptococci—In addition to its phosphodiesterase activity, SapH demonstrates nucleotide/derivative hydrolase activity for ATP and Ap_nA. This dual function of SapH might be a gain of function due to subtle substitutions at the active site that were proposed to explain the diverse substrate specificity of the MPP superfamily. For instance, the metal and substrate specificities of *Clostridium thermocellum* polynucleotide kinase/phosphatase can be dramatically altered by substitution of the active-site residues (45–49). Based on the analysis of the phosphate-binding residue His-98 in Rv0805, mutation of the corresponding Cys-74 to His in *E. coli* phosphodiesterase YfcE results in a gained 2',3'-cAMP phosphodiesterase activity (48). In SapH, the corresponding residue Trp-67 also forms a hydrogen bond with the phosphate in SapH-PO₄ and SapH-AMP. Compared with wild-type SapH-FL, the W67A mutant has a lower K_m and k_{cat} and comparable activity (k_{cat}/K_m) for bis-*p*NPP. By contrast, the W67H mutation increases the K_m for bis-*p*NPP by ~7-fold, resulting in activity that is about one-fifteenth that of the wild-type protein (Table 3). However, both

the W67A and W67H mutants have no detectable activity for ATP, Ap₃A, or Ap₄A (Table 3). These results indicate that Trp-67 of SapH-FL is indispensable for hydrolysis of ATP, Ap₃A, and Ap₄A.

To find the evolutionary hints of the bifunctionality of SapH, we performed a multiple-sequence alignment. The binuclear active-site residues Asp-11, His-13, Asp-39, Asn-66, His-128, His-164, and His-166 are highly conserved in Gram-positive bacteria (Fig. 4A). However, in addition to Trp-67, the other AMP-binding residues Arg-137, His-141, Phe-189, and His-252 are conserved exclusively in streptococci (Fig. 4B). Thus, we propose that the emergence of Trp-67, together with substitutions of residues at the substrate-binding pocket, resulted in a gain of nucleotide/derivative hydrolase function for streptococcal SapH, in addition to its phosphodiesterase activity.

Acknowledgments—We thank Wei Zhao (University of Science and Technology of China) for docking analysis. We are grateful to all the developers of the CCP4 Suite, ESPript, TreeView, and PyMOL. We thank the staff of the Shanghai Synchrotron Radiation Facility.

REFERENCES

- Hausdorff, W. P., Bryant, J., Paradiso, P. R., and Siber, G. R. (2000) *Clin. Infect. Dis.* **30**, 100–121
- Hoskins, J., Alborn, W. E., Jr., Arnold, J., Blaszczyk, L. C., Burgett, S., DeHoff, B. S., Estrem, S. T., Fritz, L., Fu, D. J., Fuller, W., Geringer, C., Gilmour, R., Glass, J. S., Khoja, H., Kraft, A. R., Lagace, R. E., LeBlanc, D. J., Lee, L. N., Lefkowitz, E. J., Lu, J., Matsushima, P., McAhren, S. M., McHenry, M., McLeaster, K., Mundy, C. W., Nicas, T. I., Norris, F. H., O'Gara, M., Peery, R. B., Robertson, G. T., Rockey, P., Sun, P. M., Winkler, M. E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C. A., Baltz, R. H., Jaskunas, S. R., Rosteck, P. R., Jr., Skatrud, P. L., and Glass, J. I. (2001) *J. Bacteriol.* **183**, 5709–5717
- Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., Durkin, A. S., Gwinn, M., Kolonay, J. F., Nelson, W. C., Peterson, J. D., Umayam, L. A., White, O., Salzberg, S. L., Lewis, M. R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A. M., Utterback, T. R., Hansen, C. L., McDonald, L. A., Feldblyum, T. V., Angiuoli, S., Dickinson, T., Hickey, E. K., Holt, I. E., Loftus, B. J., Yang, F., Smith, H. O., Venter, J. C., Dougherty, B. A., Morrison, D. A., Hollingshead, S. K., and Fraser, C. M. (2001) *Science* **293**, 498–506
- Wizemann, T. M., Heinrichs, J. H., Adamou, J. E., Erwin, A. L., Kunsch, C., Choi, G. H., Barash, S. C., Rosen, C. A., Masure, H. R., Tuomanen, E., Gayle, A., Brewah, Y. A., Walsh, W., Barren, P., Lathigra, R., Hanson, M., Langermann, S., Johnson, S., and Koenig, S. (2001) *Infect. Immun.* **69**, 1593–1598
- Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. (1998) *Infect. Immun.* **66**, 5620–5629
- Podobnik, M., Tyagi, R., Matange, N., Dermol, U., Gupta, A. K., Mattoo, R., Seshadri, K., and Visweswariah, S. S. (2009) *J. Biol. Chem.* **284**, 32846–32857
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) *Nucleic Acids Res.* **30**, 276–280
- Voegtli, W. C., White, D. J., Reiter, N. J., Rusnak, F., and Rosenzweig, A. C. (2000) *Biochemistry* **39**, 15365–15374
- Hopfner, K. P., Karcher, A., Craig, L., Woo, T. T., Carney, J. P., and Tainer, J. A. (2001) *Cell* **105**, 473–485
- Klabunde, T., Sträter, N., Fröhlich, R., Witzel, H., and Krebs, B. (1996) *Mol. Biol.* **259**, 737–748
- Knöfel, T., and Sträter, N. (2001) *J. Mol. Biol.* **309**, 239–254
- Schenk, G., Gahan, L. R., Carrington, L. E., Mitic, N., Valizadeh, M., Ham-

Streptococcal Nucleotide and Phosphodiester Hydrolase SapH

- ilton, S. E., de Jersey, J., and Guddat, L. W. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102**, 273–278
13. Guranowski, A., Jakubowski, H., and Holler, E. (1983) *J. Biol. Chem.* **258**, 14784–14789
14. Guranowski, A. (2000) *Pharmacol. Ther.* **87**, 117–139
15. Jeyakanthan, J., Kanaujia, S. P., Nishida, Y., Nakagawa, N., Praveen, S., Shinkai, A., Kuramitsu, S., Yokoyama, S., and Sekar, K. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 116–124
16. Swarbrick, J. D., Bashtannyk, T., Maksel, D., Zhang, X. R., Blackburn, G. M., Gayler, K. R., and Gooley, P. R. (2000) *J. Mol. Biol.* **302**, 1165–1177
17. Swarbrick, J. D., Buyya, S., Gunawardana, D., Gayler, K. R., McLennan, A. G., and Gooley, P. R. (2005) *J. Biol. Chem.* **280**, 8471–8481
18. Bailey, S., Sedelnikova, S. E., Blackburn, G. M., Abdelghany, H. M., Baker, P. J., McLennan, A. G., and Rafferty, J. B. (2002) *Structure* **10**, 589–600
19. Iwai, T., Kuramitsu, S., and Masui, R. (2004) *J. Biol. Chem.* **279**, 21732–21739
20. Bessman, M. J., Frick, D. N., and O’Handley, S. F. (1996) *J. Biol. Chem.* **271**, 25059–25062
21. Gabelli, S. B., Bianchet, M. A., Bessman, M. J., and Amzel, L. M. (2001) *Nat. Struct. Biol.* **8**, 467–472
22. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
23. Brodersen, D. E., de La Fortelle, E., Vornrhein, C., Bricogne, G., Nyborg, J., and Kjeldgaard, M. (2000) *Acta Crystallogr. D Biol. Crystallogr.* **56**, 431–441
24. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
25. Vagin, A., and Teplyakov, A. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25
26. Collaborative Computational Project No. 4. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
27. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
28. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
29. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucleic Acids Res.* **35**, W375–W383
30. Laskowski, R., Macarthur, M., Moss, D., and Thornton, J. (1993) *J. Appl. Crystallogr.* **26**, 283–291
31. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific LLC, San Carlos, CA
32. Miller, D. J., Shuvalova, L., Evdokimova, E., Savchenko, A., Yakunin, A. F., and Anderson, W. F. (2007) *Protein Sci.* **16**, 1338–1348
33. Chen, S., Yakunin, A. F., Kuznetsova, E., Busso, D., Pufan, R., Proudfoot, M., Kim, R., and Kim, S. H. (2004) *J. Biol. Chem.* **279**, 31854–31862
34. Wang, D., Guo, M., Liang, Z., Fan, J., Zhu, Z., Zang, J., Zhu, Z., Li, X., Teng, M., Niu, L., Dong, Y., and Liu, P. (2005) *J. Biol. Chem.* **280**, 22962–22967
35. Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., and Moomaw, E. W. (1995) *Nature* **378**, 641–644
36. Wang, Q. H., Hu, W. X., Gao, W., and Bi, R. C. (2006) *Proteins* **65**, 1032–1035
37. Shenoy, A. R., Capuder, M., Draskovic, P., Lamba, D., Visweswariah, S. S., and Podobnik, M. (2007) *J. Mol. Biol.* **365**, 211–225
38. O’Handley, S. F., Frick, D. N., Dunn, C. A., and Bessman, M. J. (1998) *J. Biol. Chem.* **273**, 3192–3197
39. Xu, W., Gauss, P., Shen, J., Dunn, C. A., and Bessman, M. J. (2002) *J. Biol. Chem.* **277**, 23181–23185
40. Traut, T. W. (1994) *Mol. Cell. Biochem.* **140**, 1–22
41. Bellinzoni, M., Wehenkel, A., Shepard, W., and Alzari, P. M. (2007) *Structure* **15**, 863–872
42. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., and Baker, D. (2003) *J. Mol. Biol.* **331**, 281–299
43. Mildvan, A. S., Xia, Z., Azurmendi, H. F., Saraswat, V., Legler, P. M., Massiah, M. A., Gabelli, S. B., Bianchet, M. A., Kang, L. W., and Amzel, L. M. (2005) *Arch. Biochem. Biophys.* **433**, 129–143
44. Iwanicki, A., Herman-Antosiewicz, A., Pierechod, M., Séror, S. J., and Obuchowski, M. (2002) *Biochem. J.* **366**, 929–936
45. Martins, A., and Shuman, S. (2005) *RNA* **11**, 1271–1280
46. Keppetipola, N., and Shuman, S. (2006) *RNA* **12**, 73–82
47. Keppetipola, N., and Shuman, S. (2006) *J. Biol. Chem.* **281**, 19251–19259
48. Keppetipola, N., and Shuman, S. (2007) *Nucleic Acids Res.* **35**, 7721–7732
49. Keppetipola, N., and Shuman, S. (2008) *J. Biol. Chem.* **283**, 30942–30949